Indirect NRDF for Partially Observable Gauss-Markov Processes with MSE Distortion: Characterizations and Optimal Solutions

Photios A. Stavrou, *Senior Member, IEEE* and Mikael Skoglund *Fellow, IEEE*

***Abstract*—We study the problem of characterizing and computing the Gaussian nonanticipative rate distortion function (NRDF) of partially observable multivariate Gauss-Markov processes with mean squared error (MSE) distortion constraints. First, we extend Witsenhausen's "tensorization" approach originally used for single-letter random variables to causal processes, to obtain a new modified representation of the NRDF for the specific problem. For time-varying vector processes, we prove conditions so that the new modified NRDF is achieved and study its implications when it is not achievable. For both cases (which correspond to different bounds), we derive the characterization and the optimal realization, whereas we give the optimal numerical solution using semidefinite programming (SDP) algorithm. Interestingly, the realization (for both bounds) is shown to be a linear functional of the current time sufficient statistic of the past and current observations signals. For the infinite time horizon, we give conditions to ensure existence of a time-invariant characterization from the finite-time horizon problems and a numerical solution using the SDP algorithm. For the time-invariant characterization, we also give strong structural properties that enable an optimal and an approximate solution via a reverse-waterfilling algorithm implemented via an iterative scheme which executes much faster than the SDP algorithm. For both finite and infinite time horizons, we study the special case of scalar processes. Our results are corroborated with various simulation studies and are also compared with existing results in the literature.**

***Index Terms*—indirect NRDF, partially observable Gaussian process, sufficient statistic, optimization, algorithmic analysis**

I. INTRODUCTION

Nonanticipatory $\epsilon-$entropy was introduced in [1], [2] motivated by applications where real-time communication with minimal encoding and decoding delays is essential. This entity is shown to be a tight lower bound on causal codes for scalar processes [3] whereas for vector processes it provides a tight lower bound at high rates on causal codes and on the average length of all causal prefix free codes [4] (also termed zero-delay coding).

Inspired by the usefulness of nonanticipatory-ϵ entropy in real-time communication, Tatikonda *et al.* in [5] reinvented the same information measure under the name sequential rate distortion function (RDF)[1] to study a linear fully observable Gaussian closed-loop control system over a memoryless communication channel subject to rate constraints. In particular, [5] used the sequential RDF subject to a pointwise MSE distortion constraint to describe a lower bound on the minimum cost of control for scalar-valued Gaussian processes and a suboptimal lower bound for the multivariate case obtained by means of a reverse-waterfilling algorithm [7, 10.3.3].[2]

Tanaka *et al.* in [10] revisited the estimation/communication part of the problem introduced by Tatikonda *et al.* and showed that the specific description of the sequential RDF is semidefinite representable. Around the same time, Stavrou *et al.* in [11] solved the general KKT conditions that correspond to the rate distortion characterization of the optimal estimation problem in [5] and proposed an adaptive reverse-waterfilling characterization (for both pointwise and total MSE distortions) that computes optimally the KKT conditions as long as all dimensions of the multidimensional setup are active, which is the case at high rates regime. In addition, in [11] they found the optimal linear coding policies (by means of a linear forward test-channel realization) that achieve the specific rate distortion characterization thus filling a gap created in [1, Theorem 5]. Recently, the optimal realization therein was used in [12] to derive bounds on a zero delay multiple description source coding problem with feedback for scalar Gaussian processes.

Kostina and Hassibi in [8] revisited the framework of [5] and derived bounds on the optimal rate-cost tradeoffs in control for time-invariant fully observable multivariate Markov processes under the assumption of uniform cost (or distortion) allocation. Recently, Charalambous *et al.* in [13] used a state augmentation technique to extend the characterization of the Gaussian nonanticipatory $\epsilon-$entropy derived in [2] to nonstationary multivariate Gaussian autoregressive models of any finite order (see also [14] for a similar result).

The extension of the framework of [5] to stochastic linear partially observable Gaussian control systems under noisy or noiseless communication channels was initially studied in [15] whereas a variation of the uncontrolled problem is studied in [16]. Particularly, Tanaka in [16] considered the estimation/communication part of the problem and derived performance limitations by minimizing a sequential RDF with *soft weighted pointwise MSE distortion constraints*. To deal

P. A. Stavrou was partially supported by a Huawei France-EURECOM Chair on Future Wireless Networks and by the KAW Foundation and the Swedish Foundation for Strategic Research. M. Skoglund received funding from the KAW Foundation and the Swedish Foundation for Strategic Research.

P. A. Stavrou is with the Communication Systems Department, EURECOM, Sophia-Antipolis, France *email:* fotios.stavrou@eurecom.fr

M. Skoglund is with the Division of Information Science and Engineering, KTH Royal Institute of Technology, Sweden *email:* skoglund@kth.se.

[1]In the literature this information measure is also encountered as nonanticipative RDF (NRDF) [6].

[2]The suboptimality of the lower bound obtained in [5] for multivariate Gaussian processes was recently identified in [8], [9].

with this problem, he first reduced the time-varying partially observable Gaussian system into a fully observable one by employing a pre-Kalman filtering (pre-KF) algorithm. Then, he assumed *á priori* a structural result on its observations process to claim causal invertibility of the pre-KF algorithm and hence to guarantee that the á posteriori state estimate between the state process and the observations process computed by the pre-KF is an information lossless operation of the true observations process at each instant of time. Armed with this result and a modified MSE distortion constraint he then claimed that the resulting problem can be equivalently reformulated as fully observable multi-letter optimization for which a cascade realization was proposed via the connection of a pre-KF, a covariance scheduling SDP algorithm, an additive white Gaussian noise (AWGN) channel and a post-KF algorithm. The stationary case of the specific optimization problem is also briefly discussed.

Despite the interesting analysis of [15], [16], there are *several important open questions* still unanswered even for the estimation/communication problem. For instance, in [16] (see *mutatis mutandis* [15]) it is not clear if the proposed lower bound (obtained by means of a sequential RDF) is tight for the proposed physical system model, or what is the characterization that needs to be solved similar to what is already known for example when the input data are modeled via a linear fully-observable multidimensional system driven by additive white Gaussian noise (see, e.g., [11, Eq. (5.22)]). Moreover, the (minimum) realization of the optimal test-channel distribution including the identification of the reverse-waterfilling parameters that achieve the specific characterization is also missing. Another important question has to do with the conditions that are needed to ensure (strict) feasibility of the optimization problem in both finite and infinite time horizons. Equally important questions include the derivation of optimal or suboptimal (numerical or analytical) solutions for this problem for both scalar or beyond scalar processes as well as the analysis of the problem for high dimensional systems that necessitates scalable optimization algorithms (an issue already known from the analysis of [17]).

Kostina and Hassibi in [8] considered some of the previous questions and derived analytical bounds on the exact solutions of the estimation and control problems for time-invariant multivariate jointly Gaussian processes *again under the assumption of uniform distortion allocation*. Hence, a natural open question related to the bounds in [8] is their tightness for multidimensional systems. This question is also related to the fact that *no insightful examples appeared in the literature so far to compute optimally partially observable multivariate Gauss-Markov processes* and compare with any of the closed form bounds proposed in [8].

A. Contributions

In this work, we study the problem of characterizing and computing a lower bound (using a modified version of NRDF) on a zero-delay source coding problem when a partially observable multivariate Gauss-Markov process is quantized and transmitted subject to a hard MSE distortion constraint in both finite and infinite time horizons. We obtain the following results.

(R1) A new modified version of the NRDF (also called indirect or remote NRDF) which is a lower bound on the optimal rates of our system model depicted in Fig. 1. The bound is obtained by extending Witsenhausen's "tensorization" approach [18] to causal processes with memory (see Section III, (22)).

(R2) Necessary and sufficient structural conditions that guarantee the tightness of the proposed indirect NRDF (see Proposition 1, Lemma 3) for jointly Gaussian processes.

(R3) For the finite time horizon, we derive the characterization and the optimal test-channel realization of the Gaussian indirect NRDF. Remarkably, the optimal realization is shown to be a linear functional of the current sufficient statistic of the past and present observation signals (see Theorem 1).

(R4) For the infinite time horizon, we identify necessary and sufficient conditions (i.e., detectability and stabilizability of appropriate pair of matrices) to ensure a steady state solution of the error covariance matrices of the sufficient statistic process (see Lemma 5) and give conditions that allow for a time-invariant characterization in the asymptotic limit (see Theorem 4).

(R5) For both finite and infinite time horizons, we give the numerical solutions of the proposed lower bound characterizations (assuming the solution is finite) by showing that they are semidefinite representable (see Theorem 2, Corollary 2). For time-varying scalar processes with average total MSE distortion constraints, we derive the optimal closed form solution via a dynamic reverse-waterfilling algorithm (see Theorem 3) that we implement in Algorithm 1 whereas for pointwise MSE distortion constraints we derive the optimal closed form solution (see Corollary 1). Under certain strong structural properties on the time-invariant characterization of the problem (see Proposition 5) we derive an optimal scalable reverse-waterfilling solution (see Theorem 5) with its algorithmic embodiment (see Algorithm 2).

(R6) We supplement our major results with numerical validations including connections with [8] (see Section VI).

B. Comparison to prior art

The derivation of the modified NRDF in **(R1)** is new and possess similar properties to the classical NRDF [1], [19], i.e., convexity, lower-semicontinuity etc. Based on the structural properties derived in **(R2)**, we claim that the corresponding structural conditions assumed á priori in the system model of [16, Equations (1a), (1b)] are not sufficient to ensure the tightness of the proposed lower bound but instead, these correspond to a conservative lower bound compared to the original lower bound that we prove in this paper. The implications of this conservative lower bound are also studied in our paper (see Propositions 2-6). The characterizations obtained in **(R3)**, **(R4)**, are different compared to [16] because they are obtained with *hard average total* MSE distortion constraints (and global Lagrange multipliers) instead of soft pointwise MSE distortion constraints (and given á priori multiple Lagrange multipliers) that are assumed in [16]. The structural simplification of the multi-letter optimization problem of Definition 2 obtained via

[11, Theorem 4.1] is also new (see Equation (37)). The computational complexity of the algorithmic approaches derived in **(R5)** compared to the prior research studies is discussed in Remarks 4, 6, 9. Note that Theorem 5 and its implementation in Algorithm 2 are extremely important for two reasons; first, we can gain better insights of the problem in the infinite time horizon (for instance it paves the way to derive optimal closed form solutions beyond scalar processes, hence generalizing the results obtained for the fully observable time-invariant multivariate Gauss-Markov processes, see e.g., [17, Section IV]) and, second, Algorithm 2 as Table I suggests can operate much faster than the SDP algorithm in high dimensional systems (it is scalable). Our numerical simulation in Example 2, apart from verifying numerically that both Corollary 2 and Theorem 5 coincide under certain structural properties, it also shows that the corresponding analytical lower bound obtained for partially observable time-invariant multidimensional Gauss-Markov processes via [8, Corollary 1, Theorem 9] is not tight in general but a fairly tight performance (not exact) can be observed at a very low distortion. Consequently, its utility to controlled processes in [8, Theorem 5] should be seen under this consideration. Example 3, shows the utility of Algorithm 1 when we restrict our system to time-invariant scalar processes, namely, it recovers the steady-state closed form solution of Corollary 3 (or [20, eq. (103)]). Finally, for every result in this paper we recover or explain how to recover as a special case known results in the literature.

Notation. We let $\mathbb{R}=(-\infty,\infty)$, $\mathbb{Z}=\{\ldots,-1,0,1,\ldots\}$, $\mathbb{N}_0=\{0,1,\ldots\}$, $\mathbb{N}_0^n=\{0,1,\ldots,n\}$, $n\in\mathbb{N}_0$. Let $\mathcal{X}$ be a finite dimensional Euclidean space and $\mathcal{B}(\mathcal{X})$ the Borel σ-field of $\mathcal{X}$. A random variable (RV) defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ is a map $\mathbf{x}:\Omega\longmapsto\mathcal{X}$, where $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ is a measurable space. We denote a sequence of RVs by $\mathbf{x}_r^t\triangleq(\mathbf{x}_r,\mathbf{x}_{r+1},\ldots,\mathbf{x}_t),(r,t)\in\mathbb{Z}\times\mathbb{Z},t\geq r$, and their realizations by $x_r^t\in\mathcal{X}_r^t\triangleq\times_{k=r}^t\mathcal{X}_k$, for simplicity. If $r=-\infty$ and $t=-1$, we use the notation $\mathbf{x}_{-\infty}^{-1}=\mathbf{x}^{-1}$, and if $r=0$, we use the notation $\mathbf{x}_0^t=\mathbf{x}^t$. The distribution of the RV $\mathbf{x}$ on $\mathcal{X}$ is denoted by $\mathbf{P}(dx)$. The conditional distribution of a RV $\mathbf{y}$ given $\mathbf{x}=x$ is denoted by $\mathbf{P}(dy|x)$. The transpose and covariance of a random vector $\mathbf{x}$ are denoted by $\mathbf{x}^{\mathrm{T}}$ and $\Sigma_{\mathbf{x}}$. We denote the determinant, trace, rank, diagonal, diagonal elements, and the eigenvalues of a square matrix $S\in\mathbb{R}^{p\times p}$ by $|S|$, $\mathrm{trace}(S)$, $\mathrm{rank}(S)$, $\mathrm{diag}(S)$, $[\cdot]_{ii}$ and $\{\mu_{S,i}\}_{i=1}^p$. We denote the transpose of a real (rectangular) matrix $F\in\mathbb{R}^{i\times j}$ by F^{T}. The notation $\Sigma\succ 0$ (resp. $\Sigma\succeq 0$) denotes a positive definite (resp. positive semi-definite) matrix. The notation $A\succ B$ (resp. $A\succeq B$) means $A-B\succ 0$ (resp. $A-B\succeq 0$). We denote a $p\times p$ identity matrix by I_p. $R^G(D)$ denotes the Gaussian version of the RDF. The expectation operator is denoted by $\mathbb{E}\{\cdot\}$; $||\cdot||$ denotes Euclidean norm; $[\cdot]^+\triangleq\max\{0,\cdot\}$. We denote by $\mathrm{abs}(|\cdot|)$ the absolute value of a determinant.

II. Problem Statement

We consider the causal zero-delay source coding setup of Fig. 1. In this setting, the "hidden" $\mathbb{R}^p$-valued source is modeled by a discrete-time time-varying partially observable Gauss-Markov process as follows

$$\mathbf{x}_{t+1}=A_t\mathbf{x}_t+\mathbf{w}_t,\ \mathbf{x}_0=\bar{x}, \tag{1}$$
$$\mathbf{z}_t=C_t\mathbf{x}_t+\mathbf{n}_t,\ t\in\mathbb{N}_0, \tag{2}$$

where $A_t\in\mathbb{R}^{p\times p}$ is a square non-random matrix, $C_t\in\mathbb{R}^{m\times p}$ is a possibly rectangular non-random fat matrix ($m\leq p$), $\mathbf{x}_0\in\mathbb{R}^p\sim(0;\Sigma_{\mathbf{x}_0})$, $\Sigma_{\mathbf{x}_0}\succ 0$ is the initial state, $\mathbf{w}_t\in\mathbb{R}^p\sim\mathcal{N}(0;\Sigma_{\mathbf{w}_t})$, $\Sigma_{\mathbf{w}_t}\succ 0$ is an independent sequence, $\mathbf{n}_t\in\mathbb{R}^m\sim\mathcal{N}(0;\Sigma_{\mathbf{n}_t})$, $\Sigma_{\mathbf{n}_t}\succeq 0$, is an independent sequence, independent of $\{\mathbf{w}_t:\ t\in\mathbb{N}_0\}$, whereas $\mathbf{x}_0$ is independent of $\{(\mathbf{w}_t,\mathbf{n}_t):\ t\in\mathbb{N}_0\}$.

System's operation: At every time instant, the *encoder (E)* observes the impaired measurement $\mathbf{z}_t$ (provided $\mathbf{z}^{t-1}$ are already observed) and generates the data packet $\mathbf{m}_t\in\mathcal{M}_t\subset\{0,1\}^{\ell_t}$ of instantaneous expected rate $R_t=\mathbb{E}|\boldsymbol{\ell}_t|$, where $|\boldsymbol{\ell}_t|$ denotes the binary sequence of $\boldsymbol{\ell}_t$. At time t, $\mathbf{m}_t$ is transmitted across a noiseless channel with rate R_t. Upon receiving $\mathbf{m}^t$, a *minimum MSE (MMSE) decoder (D)* immediately produces an estimate $\mathbf{y}_t$ of the source sample $\mathbf{x}_t$, under the assumption that $\mathbf{y}^{t-1}$ are already reproduced. We assume that at time $t=0$ there is no prior information whereas the clocks of the encoder and the decoder are synchronized. Formally, the *(ED)* pair is specified by the sequence of measurable functions $\{(f_t,g_t)\ :\ t\in\mathbb{N}_0\}$ with $f_t:\mathcal{M}^{t-1}\times\mathcal{Z}^t\mapsto\mathcal{M}_t$ and $g_t:\ \mathcal{M}^t\mapsto\mathcal{Y}_t$, $t\in\mathbb{N}_0$, such that

$$\begin{aligned}&(\mathcal{E}):\ m_t=f_t(m^{t-1},z^t),\ m^{-1}=\emptyset,z^{-1}=\emptyset,\\&(\mathcal{D}):\ y_t=g_t(m^t).\end{aligned} \tag{3}$$

Distortion Constraint. The distortion constraint is described

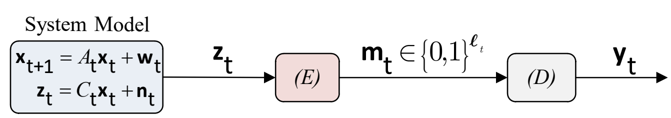

Fig. 1: System model.

by the *average total* MSE *distortion* constraint given by

$$\frac{1}{n+1}\sum_{t=0}^n\mathbb{E}\left\{||\mathbf{x}_t-\mathbf{y}_t||^2\right\}\leq D, \tag{4}$$

and its asymptotic (upper) limit by

$$\limsup_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^n\mathbb{E}\left\{||\mathbf{x}_t-\mathbf{y}_t||^2\right\}\leq D. \tag{5}$$

Performance. The performance of the multidimensional system in Fig. 1 for some finite n can be cast as follows

$$R^c_{[0,n],\mathrm{in}}(D)\triangleq\inf_{\substack{\text{eq. (3)},\ t\in\mathbb{N}_0^n\\ \text{eq. (4)}}}\frac{1}{n+1}\sum_{t=0}^n R_t. \tag{6}$$

The asymptotic limit of (6) is given as follows

$$R^c_{\mathrm{in}}(D)\triangleq\inf_{\substack{\text{eq. (3)},\ t\in\mathbb{N}_0,\\ \text{eq. (5)}}}\limsup_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^n R_t. \tag{7}$$

III. Lower bounds

In this section, we first prove a lower bound that corresponds to the operational rates given in (6), (7).

From data processing properties of the system in Fig. 1, we have that the joint distribution induced by the joint process $\{(\mathbf{z}_t,\mathbf{m}_t,\mathbf{y}_t):\ t\in\mathbb{N}_0^n\}$ admits the following decomposition:

$$\begin{aligned}
&\mathbf{P}(dy^n,dm^n,dz^n)=\otimes_{t=0}^n\mathbf{P}(dy_t,dm_t,dz_t|y^{t-1},m^{t-1},z^{t-1})\\
&=\otimes_{t=0}^n\mathbf{P}(dy_t|y^{t-1},z^t,m^t)\otimes\mathbf{P}(dm_t|m^{t-1},z^t,y^{t-1})\\
&\quad\otimes\mathbf{P}(dz_t|z^{t-1},y^{t-1},m^{t-1})\\
&\overset{(a)}{=}\otimes_{t=0}^n\mathbf{P}(dy_t|y^{t-1},m^t)\otimes\mathbf{P}(dm_t|m^{t-1},z^t,y^{t-1})\\
&\quad\otimes\mathbf{P}(dz_t|z^{t-1}),
\end{aligned}\tag{8}$$

where (a) stems from the fact that we assume in our system the following natural conditional independence constraints

$$\mathbf{P}(dz_t|z^{t-1},y^{t-1},m^{t-1})=\mathbf{P}(dz_t|z^{t-1}),\tag{9}$$
$$\mathbf{P}(dy_t|y^{t-1},z^t,m^t)=\mathbf{P}(dy_t|y^{t-1},m^t).\tag{10}$$

Based on the setup of Fig. 1, in (8) we assume that the joint distribution $\mathbf{P}(dz^{-1},dm^{-1},dy^{-1})$ generates zero information. The following data processing result, provides the appropriate information measure that can be used to compute a lower bound on (6).

Lemma 1. *(Data processing inequalities) Under the decomposition of the joint distribution in (8), the communication system in Fig. 1 admits the following data processing inequalities*

$$I(\mathbf{z}^n;\mathbf{y}^n)\overset{\textbf{(ii)}}{\leq}I(\mathbf{z}^n;\mathbf{m}^n||\mathbf{y}^{n-1})\overset{\textbf{(i)}}{\leq}\sum_{t=0}^n R_t,\tag{11}$$

where $I(\mathbf{z}^n;\mathbf{y}^n)=\sum_{t=0}^n I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1})$, *and* $I(\mathbf{z}^n;\mathbf{m}^n||\mathbf{y}^{n-1})=\sum_{t=0}^n I(\mathbf{z}^t;\mathbf{m}_t|\mathbf{m}^{t-1},\mathbf{y}^{t-1})$, *assuming* $I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1})<\infty$, $I(\mathbf{z}^t;\mathbf{m}_t|\mathbf{m}^{t-1},\mathbf{y}^{t-1})<\infty$, $\forall t$.

Proof: The proof is similar to [21, Theorem 1] thus we omit it. ∎

Next, we show how to formally construct the information measure $I(\mathbf{z}^n;\mathbf{y}^n)$.

Observations Process. The observations process $\{\mathbf{z}_t:\ t\in\mathbb{N}_0^n\}$ induces the sequence of conditional distributions $\mathbf{P}(dz_t|z^{t-1})$, $t\in\mathbb{N}_0^n$. At $t=0$ we assume that $\mathbf{P}(dz_0|z^{-1})=\mathbf{P}(dz_0)$ and by Bayes' rule we obtain

$$\mathbf{P}(dz^n)=\otimes_{t=0}^n\mathbf{P}(dz_t|z^{t-1}).\tag{12}$$

It should be noted that for the system model (1), (2), at each instant of time, the conditional distribution of $\mathbf{P}(dz_t|z^{t-1})$ depends on the distribution of the hidden data $\mathbf{x}_t$ given all the past observation symbols $\mathbf{z}^{t-1}$ via

$$\mathbf{P}(dz_t|z^{t-1})=\int_{\mathcal{X}_t}\mathbf{P}(dz_t|x_t)\mathbf{P}(dx_t|z^{t-1}).\tag{13}$$

Reproduction or "test-channel". The reproduction process $\mathbf{y}_t$ parametrized by $\mathcal{Y}^{t-1}\times\mathcal{Z}^t$ induces the sequence of conditional distributions or as test-channels $\mathbf{P}(dy_t|y^{t-1},z^t)$, $t\in\mathbb{N}_0^n$. At $t=0$, no initial information is assumed, hence $\mathbf{P}(dy_0|y^{-1},z^0)=\mathbf{P}(dy_0|z_0)$. The sequence of conditional distributions $\{\mathbf{P}(dy_t|y^{t-1},z^t):\ t\in\mathbb{N}_0\}$ uniquely defines the family of conditional distributions on $\mathcal{Y}^n$ parametrized by $z^n\in\mathcal{Z}^n$, given by

$$\mathbf{Q}(dy^n|z^n)\triangleq\otimes_{t=0}^n\mathbf{P}(dy_t|y^{t-1},z^t),\tag{14}$$

and vice-versa. From (12) and (14), we can uniquely define the joint distribution of $\{(\mathbf{z}_t,\mathbf{y}_t):\ t\in\mathbb{N}_0^n\}$ by

$$\mathbf{P}(dy^n,dz^n)=\mathbf{P}(dz^n)\otimes\mathbf{Q}(dy^n|z^n).\tag{15}$$

In addition, from (15), we can define the $\mathcal{Y}^n-$marginal distribution $\mathbf{P}(dy^n)\triangleq\otimes_{t=0}^n\mathbf{P}(dy_t|y^{t-1})$, where

$$\mathbf{P}(dy_t|y^{t-1})=\int_{\mathcal{Z}^t}\mathbf{P}(dy_t|y^{t-1},z^t)\otimes\mathbf{P}(dz^t|y^{t-1}).\tag{16}$$

Given the above construction of distributions we obtain the following variant of directed information [22]

$$\begin{aligned}
I(\mathbf{z}^n;\mathbf{y}^n)&\overset{(a)}{=}\sum_{t=0}^n\mathbb{E}\left\{\log\left(\frac{d\mathbf{P}(\cdot|\mathbf{y}^{t-1},\mathbf{z}^t)}{d\mathbf{P}(\cdot|\mathbf{y}^{t-1})}(\mathbf{y}_t)\right)\right\}\\
&\overset{(b)}{\triangleq}\sum_{t=0}^n I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1}),
\end{aligned}\tag{17}$$

where (a) is due to chain rule of relative entropy using the Radon-Nykodym derivative [23]; (b) follows by definition.

Definition 1. *(Lower bounds on (6), (7)) For a given processes $\{\mathbf{z}_t:\ t\in\mathbb{N}_0^n\}$ that induces the conditional distribution (13), a lower bound on (6), hereinafter called remote or indirect NRDF, subject to (4) is defined as follows*

$$R_{[0,n],\text{in}}(D)\triangleq\inf_{\substack{\mathbf{P}(dy_t|y^{t-1},z^t):\ t\in\mathbb{N}_0^n\\ eq.\ (4)}}I(\mathbf{z}^n;\mathbf{y}^n).\tag{18}$$

Moreover, its asymptotic (upper) limit expression that corresponds to a lower bound on (7) is given by

$$R_{\text{in}}(D)\triangleq\inf_{\substack{\mathbf{P}(dy_t|y^{t-1},z^t):\ t\in\mathbb{N}_0\\ eq.\ (5)}}\limsup_{n\longrightarrow\infty}\frac{1}{n+1}I(\mathbf{z}^n;\mathbf{y}^n).\tag{19}$$

Next, we further analyze (18) and discuss some of its most important properties. Before we do it, we remark that the name indirect or remote NRDF is adopted because (18) can be seen as an extension to causal processes (with memory) of the remote or indirect RDF defined for i.i.d. memoryless processes $\{(\mathbf{x}_t,\mathbf{z}_t,\mathbf{y}_t):\ t\in\mathbb{N}_0^n\}$ or RVs $(\mathbf{x},\mathbf{z},\mathbf{y})$ in the context of non-causal coding see, e.g., [24], [25], [26, Chapters 3.5, 4.5]. In the sequel, we generalize Witsenhausen's "tensorization" approach (see e.g., [18]) to time-varying causal processes to prove an equivalent expression to (18) with a modified MSE distortion.

$$\begin{aligned}
&\sum_{t=0}^{n}\mathbb{E}\{||\mathbf{x}_t-\mathbf{y}_t||^2\}=\sum_{t=0}^{n}\int_{\mathcal{X}^t\times\mathcal{Y}^t}||x_t-y_t||^2\mathbf{P}(dx^t,dy^t)\\
&=\sum_{t=0}^{n}\int_{\mathcal{X}^t\times\mathcal{Z}^t\times\mathcal{Y}^t}||x_t-y_t||^2\mathbf{P}(dx^t,dz^t,dy^t)\\
&\stackrel{(\star)}{=}\sum_{t=0}^{n}\int_{\mathcal{Z}^t\times\mathcal{Y}^t}\mathbf{P}(dz^t,dy^t)\int_{\mathcal{X}^t}||x_t-y_t||^2\mathbf{P}(dx^t|z^t)\\
&\stackrel{(\star\star)}{=}\sum_{t=0}^{n}\int_{\mathcal{Z}^t\times\mathcal{Y}^t}\mathbf{P}(dz^t,dy^t)\int_{\mathcal{X}_t}||x_t-y_t||^2\mathbf{P}(dx_t|z^t)\\
&\stackrel{(\star\star\star)}{=}\sum_{t=0}^{n}\mathbb{E}\{\hat{d}(\mathbf{z}^t,\mathbf{y}_t)\},
\end{aligned}\tag{20}$$

where $(\star)$ follows due to the conditional independence constraint $\mathbf{P}(dy^t|z^t,x^t)=\mathbf{P}(dy^t|z^t)$, for any $t=0,1,\ldots,n$; $(\star\star)$ follows because the integration concerning $x^t\in\mathcal{X}^t$ in $(\star)$, due to the squared-error distortion, affects only the current realization x_t in the *posterior distribution*, which in turn admits the following recursion

$$\mathbf{P}(dx_t|z^t)=\frac{\mathbf{P}(dz_t|x_t)\mathbf{P}(dx_t|z^{t-1})}{\int_{\mathcal{X}_t}\mathbf{P}(dz_t|x_t)\mathbf{P}(dx_t|z^{t-1})},\tag{21}$$

$(\star\star\star)$ follows if we define $\hat{d}(\mathbf{z}^t,\mathbf{y}_t)\triangleq\int_{\mathcal{X}_t}||x_t-y_t||^2\mathbf{P}(dx_t|z^t)$. Hence, (18) can be equivalently reformulated as follows

$$R_{[0,n],\text{in}}(D)\triangleq\inf_{\substack{\mathbf{P}(dy_t|y^{t-1},z^t):\ t\in\mathbb{N}_0^n\\ \frac{1}{n+1}\sum_{t=0}^{n}\mathbb{E}\left\{\hat{d}(\mathbf{z}^t,\mathbf{y}_t)\right\}\leq D}}I(\mathbf{z}^n;\mathbf{y}^n),\tag{22}$$

which corresponds precisely to a "modified" direct NRDF. Further, it is easy to show that (22) is convex with respect to the minimizer $\{\mathbf{P}(dy_t|y^{t-1},z^t):\ t\in\mathbb{N}_0^n\}$ following for instance [19]. In addition, $R_{[0,n],\text{in}}(D)$ is monotonically non-increasing, convex with respect to D, continuous in $D\in(D^{\min},\infty)$ and if $R_{[0,n],\text{in}}(D^{\min})<\infty$, then, it is continuous in $D\in[D^{\min},\infty)$. It is also well known that, $R_{[0,n],\text{in}}(D)$ achieves smaller rates if in addition to $\{(\mathbf{x}_t,\mathbf{z}_t):\ t\in\mathbb{N}_0^n\}$ being a jointly Gaussian process with the linear evolution of (1), (2), the joint process $\{(\mathbf{x}_t,\mathbf{z}_t,\mathbf{y}_t):\ t\in\mathbb{N}_0^n\}$ is also Gaussian because then $I(\mathbf{z}^n;\mathbf{y}^n)\geq I^G(\mathbf{z}^n;\mathbf{y}^n)$ (that is, the Gaussian version of $I(\mathbf{z}^n;\mathbf{y}^n)$) which in turn implies that $R_{[0,n],\text{in}}(D)\geq R^G_{[0,n],\text{in}}(D)$ (see, e.g., [27, Theorem 1.8.6]).

IV. Finite Time Horizon Problems

In this section, we assume that the end-to-end system in Fig. 1 is jointly Gaussian and we study, in finite time, the characterization, the realization and computation of (22).

To characterize the problem we use a two-step approach comprised of a pre-KF step followed by two structural results.

Lemma 2. *(Classical KF) For the jointly Gaussian system model of (1), (2), define the á priori and á posteriori state estimates as $\widehat{\mathbf{x}}_{t|t-1}\triangleq\mathbb{E}\{\mathbf{x}_t|\mathbf{z}^{t-1}\}$ and $\widehat{\mathbf{x}}_{t|t}\triangleq\mathbb{E}\{\mathbf{x}_t|\mathbf{z}^t\}$, respectively, and their corresponding error covariance matrices by*[3]

$$\Sigma^{\mathbf{x}}_{t|t-1}\triangleq\mathbb{E}\{(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t-1})^{T}\}\tag{23}$$

$$\Sigma^{\mathbf{x}}_{t|t}\triangleq\mathbb{E}\{(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t})(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t})^{T}\}.\tag{24}$$

Then, the optimal values of $\{(\widehat{\mathbf{x}}_{t|t-1},\widehat{\mathbf{x}}_{t|t},\Sigma^{\mathbf{x}}_{t|t-1},\Sigma^{\mathbf{x}}_{t|t}):\ t\in\mathbb{N}_0^n\}$ are computed recursively forward in time as follows

$$\begin{aligned}
&\widehat{\mathbf{x}}_{t|t}=\widehat{\mathbf{x}}_{t|t-1}+\mathbf{k}^{\mathbf{z}}_t\mathbf{I}^{\mathbf{z}}_t,\ \ \widehat{\mathbf{x}}_{0|-1}=\bar{x}_0,\\
&\widehat{\mathbf{x}}_{t|t-1}=A_{t-1}\widehat{\mathbf{x}}_{t-1|t-1},\\
&\Sigma^{\mathbf{x}}_{t|t-1}=A_{t-1}\Sigma^{\mathbf{x}}_{t-1|t-1}A^{T}_{t-1}+\Sigma_{\mathbf{w}_{t-1}},\ \ \Sigma^{\mathbf{x}}_{0|-1}=\Sigma_{\mathbf{x}_0},\\
&\mathbf{I}^{\mathbf{z}}_t=\mathbf{z}_t-\mathbb{E}\{\mathbf{z}_t|\mathbf{z}^{t-1}\}=C_t(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t-1})+\mathbf{n}_t,\\
&\Sigma_{\mathbf{I}^{\mathbf{z}}_t}=C_t\Sigma^{\mathbf{x}}_{t|t-1}C^{T}_t+\Sigma_{\mathbf{n}_t}\\
&\mathbf{k}^{\mathbf{z}}_t=\Sigma^{\mathbf{x}}_{t|t-1}C^{T}_t\Sigma^{-1}_{\mathbf{I}^{\mathbf{z}}_t}\ \textit{(Kalman Gain)},\\
&\Sigma^{\mathbf{x}}_{t|t}=\Sigma^{\mathbf{x}}_{t|t-1}-\Sigma^{\mathbf{x}}_{t|t-1}C^{T}_t\Sigma^{-1}_{\mathbf{I}^{\mathbf{z}}_t}C_t\Sigma^{\mathbf{x}}_{t|t-1},
\end{aligned}\tag{25}$$

where $\mathbf{I}^{\mathbf{z}}_t$ is the innovations process that is independent of $(\widehat{\mathbf{x}}_{t|t-1},\mathbf{x}^{t-1},\mathbf{z}^{t-1},\mathbf{y}^{t-1})$ and $\Sigma^{\mathbf{x}}_{t|t}\succeq 0$ and $\Sigma^{\mathbf{x}}_{t|t-1}\succ 0$.

Proof: The proof is known, see e.g., [28]–[31]. ■

Next, we prove a proposition where we extend [7, Theorem 2.8.1] to causal processes with memory.

Proposition 1. *(Data processing inequality) Suppose that for the joint process $\{(\mathbf{x}_t,\mathbf{z}_t,\mathbf{y}_t):\ t\in\mathbb{N}_0^n\}$ we have that $\mathbf{P}(dy_t|y^{t-1},\xi^t,z^t)=\mathbf{P}(dy_t|y^{t-1},z^t),\ \forall t\in\mathbb{N}_0^n$. Moreover, let the statistic $\boldsymbol{\xi}_t=f_t(\mathbf{z}^t),\forall t\in\mathbb{N}_0^n$. Then,*

$$\sum_{t=0}^{n}I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1})\leq\sum_{t=0}^{n}I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1}),\tag{26}$$

for any n, assuming $I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1})<\infty$, $I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1})<\infty,\forall t$. Additionally, (26) holds with equality if and only if (iff)

$$\mathbf{P}(dy_t|y^{t-1},\xi^t,z^t)=\mathbf{P}(dy_t|y^{t-1},\xi^t),\ \forall t\in\mathbb{N}_0^n.\tag{27}$$

Proof: By the chain rule, we can expand conditional mutual information in two different ways, i.e.,

$$\begin{aligned}
&\sum_{t=0}^{n}I(\mathbf{z}^t,\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1})\\
&=\sum_{t=0}^{n}\Big[I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1})+\underbrace{I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1},\mathbf{z}^t)}_{=0}\Big] &(28)\\
&=\sum_{t=0}^{n}\left[I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1})+I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1},\boldsymbol{\xi}^t)\right] &(29)\\
&\geq\sum_{t=0}^{n}I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1}), &(30)
\end{aligned}$$

where in (28) $I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1},\mathbf{z}^t)=0,\forall t$ because of the natural conditional independence constraint of the proposition and (30) holds because $I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1},\boldsymbol{\xi}^t)\geq 0,\forall t$. From (28) and (30) we obtain (26). Clearly, the inequality holds with equality

[3]For jointly Gaussian systems, the conditional covariance is equal to its unconditional version [28].

iff $I(\mathbf{z}^t;\mathbf{y}_t|\mathbf{y}^{t-1},\boldsymbol{\xi}^t) = 0, \forall t$, i.e., when (27) holds. This completes the proof. ■

If (27) holds, then, we say that the statistic $\boldsymbol{\xi}^t$ is *sufficient* because it contains all the information of $\mathbf{z}^t$ about $\mathbf{y}_t$ parametrized by $\mathcal{Y}^{t-1}$ at each instant of time.

The following lemma is a main result of this paper. It derives a sufficient condition such that (27) in Proposition 1 holds with equality, hence ensuring that (26) holds with equality for jointly Gaussian multivariate processes.

Lemma 3. *(Structural conditions for equality of (26)) Suppose that $\{(\mathbf{x}_t,\mathbf{z}_t,\mathbf{y}_t) : t\in\mathbb{N}_0^n\}$ is a jointly Gaussian multivariate process with a source model given by (1), (2). Moreover, let the optimal estimator in Lemma 2 be denoted by $\boldsymbol{\xi}_t=\mathbb{E}\{\mathbf{x}_t|\mathbf{z}^t\}$. Then, the conditional independence (27) holds if in Lemma 2, $\mathbf{k}_t^{\mathbf{z}}\in\mathbb{R}^{p\times m}$ is square ($p=m$) and invertible at each t.*

Proof: The idea follows similar (but not identical) arguments to [32]. Observe that the following hold

$$\begin{aligned}\mathbf{P}(dy_t|y^{t-1},z^t) &\stackrel{(a)}{=} \mathbf{P}(dy_t|y^{t-1},z^t,\xi^t)\\ &\stackrel{(b)}{=} \mathbf{P}(dy_t|y^{t-1},I^{\mathbf{z}^t},\xi^t)\\ &\stackrel{(c)}{=} \mathbf{P}(dy_t|y^{t-1},\xi^t),\end{aligned} \tag{31}$$

where (a) follows from Proposition 1 by setting $\boldsymbol{\xi}_t=\mathbb{E}\{\mathbf{x}_t|\mathbf{z}^t\}$; (b) follows from Lemma 2 because from the innovations process we have $\mathbf{z}_t = \mathbf{I}_t^{\mathbf{z}}+\mathbb{E}\{\mathbf{z}_t|\mathbf{z}^{t-1}\} = C_tA_{t-1}\boldsymbol{\xi}_{t-1}+\mathbf{I}_t^{\mathbf{z}}$; (c) follows by ensuring that the estimator $\boldsymbol{\xi}_t$ and the innovations $\mathbf{I}_t^{\mathbf{z}}$ generate the same information at each t (a standard argument to ensure an optimal KF algorithm [28]). Indeed, (c) can be guaranteed as follows. From Lemma 2 and the expression $\boldsymbol{\xi}_t = A_{t-1}\boldsymbol{\xi}_{t-1}+\mathbf{k}_t^{\mathbf{z}}\mathbf{I}_t^{\mathbf{z}}$ we can let $T(\boldsymbol{\xi}_t,\boldsymbol{\xi}_{t-1})=\boldsymbol{\xi}_t-A_{t-1}\boldsymbol{\xi}_{t-1}$ for specified $\boldsymbol{\xi}_{t-1}$, and observe that the information between $(\boldsymbol{\xi}_t,\mathbf{I}_t^{\mathbf{z}})$ at each t is preserved if the solution of both linear equations $T(\boldsymbol{\xi}_t,\boldsymbol{\xi}_{t-1})=\mathbf{k}_t^{\mathbf{z}}\mathbf{I}_t^{\mathbf{z}}$ and $\mathbf{I}_t^{\mathbf{z}}=\widehat{\mathbf{k}}_t^{\mathbf{z}}T(\boldsymbol{\xi}_t,\boldsymbol{\xi}_{t-1})$ for some matrix $\widehat{\mathbf{k}}_t^{\mathbf{z}}$, is concurrently unique (forming a bijective linear transformation), which is the case if $\mathbf{k}_t^{\mathbf{z}}$ is invertible, i.e., $|\mathbf{k}_t^{\mathbf{z}}|\neq 0$ and $\widehat{\mathbf{k}}_t^{\mathbf{z}}=(\mathbf{k}_t^{\mathbf{z}})^{-1}$. ■

To put it simply, Lemma 3 claims that for the system model (1), (2) and jointly Gaussian processes, the information between $\boldsymbol{\xi}^t$ and $\mathbf{z}^t$ is preserved *if we can uniquely obtain $\boldsymbol{\xi}^t$ from $\mathbf{z}^t$ and vice versa*, for any t. Based on this observation, we state the following remark.

Remark 1. *(On Lemma 3 and connections to [16]) From properties of the rank of a matrix (see e.g., [33, Corollary 8.3.3]), the structural condition in Lemma 3 holds iff $C_t\in\mathbb{R}^{m\times p}$ in (2) is square (i.e., $m=p$) and full rank at each t. It should be emphasized that the structural condition derived in Lemma 3 is not the same as the one considered in [16, Lemma 2] which claims a different structural property of matrix $\mathbf{k}_t^{\mathbf{z}}$ (i.e., $\mathbf{k}_t^{\mathbf{z}}$ is full column rank as a result of matrix C_t being full row rank throughout that paper). Indeed, therein the author claimed that the KF algorithm which corresponds exactly to our Lemma 2, forms a causally invertible operation for any t, if it is possible to recover $\mathbf{z}^t$ from $\boldsymbol{\xi}^t$. Provided that causally invertible operation means invertible (or bijective) operation for each time instant t, the result of [16, Lemma 2] should be seen with caution, because it does not guarantee a unique reconstruction of $\mathbf{z}^t$ as a function of $\boldsymbol{\xi}^t$ and vice versa. Indeed, one can easily verify our claim by taking as an example the initial time instant of the á posteriori conditional mean in (25) (for $\boldsymbol{\xi}_t=\widehat{\mathbf{x}}_{t|t}$, take $t=0$ with $\boldsymbol{\xi}_{-1}=0$) and check the conditions for invertibility of the resulting linear matrix equation, i.e., $\boldsymbol{\xi}_0=\mathbf{k}_0^{\mathbf{z}}\mathbf{z}_0$. This implies that the statement of [16, Lemma 2] does not suffice, in general, to ensure the conditional independence (27), hence under the specific structural condition, the inequality in (26) is strict for all t, and as a result (22) is not achievable. In the sequel, we will also discuss the implications of $C_t\in\mathbb{R}^{m\times p}$ being a full row rank with $m<p$.*

Next, we study the structure of the amended distortion constraint in the convex optimization problem of (22) obtained for jointly Gaussian processes. Specifically, following [25] and using the fact that $\boldsymbol{\xi}_t=\mathbb{E}\{\mathbf{x}_t|\mathbf{z}^t\}$ we obtain

$$\begin{aligned}\hat{d}(z^t,y_t) =&\mathbb{E}_{\mathbf{x}_t|\mathbf{z}^t=z^t}\{||\mathbf{x}_t-\mathbf{y}_t||^2\}\\ =&\mathbb{E}_{\mathbf{x}_t|\mathbf{z}^t=z^t}\{||\mathbf{x}_t-\boldsymbol{\xi}_t+\boldsymbol{\xi}_t-\mathbf{y}_t||^2\}\\ \stackrel{(i)}{=}&\mathbb{E}_{\mathbf{x}_t|\mathbf{z}^t=z^t}\{||\mathbf{x}_t-\boldsymbol{\xi}_t||^2\}+||\boldsymbol{\xi}_t-\mathbf{y}_t||^2\\ \stackrel{(ii)}{=}&\operatorname{trace}(\Sigma_{t|t}^{\mathbf{x}})+||\boldsymbol{\xi}_t-\mathbf{y}_t||^2,\end{aligned} \tag{32}$$

where (i) follows because for jointly Gaussian processes $\boldsymbol{\xi}_t$ is the optimal MMSE estimator of $\mathbf{x}_t$ given $\mathbf{z}^t$ and from the orthogonality principle; (ii) follows by definition of the *á posteriori* error covariance of the optimal MMSE obtained from the KF recursions in Lemma 2. Finally, the amended distortion constraint in (22) is obtained by taking the expectation concerning the joint distribution of $\{(\mathbf{z}_t,\mathbf{y}_t) : t\in\mathbb{N}_0^n\}$ in (32) and then the summation which will give

$$\sum_{t=0}^n \operatorname{trace}(\Sigma_{t|t}^{\mathbf{x}})+\sum_{t=0}^n\mathbb{E}\{||\boldsymbol{\xi}_t-\mathbf{y}_t||^2\}. \tag{33}$$

Putting all the pieces together, we can reformulate (22) (and its asymptotic (upper) limit) as follows.

Definition 2. *(Indirect Gaussian NRDF) Suppose that the process $\{(\mathbf{x}_t,\mathbf{z}_t,\mathbf{y}_t) : t\in\mathbb{N}_0^n\}$ is jointly Gaussian and $C_t\in\mathbb{R}^{m\times p}$ in (2) is square and full rank, i.e., $m=p$. Then, (18) and (19), respectively, can be reformulated as follows*

$$\begin{aligned}&R^G_{[0,n],\text{in}}(D-D^{\min}_{[0,n]})\\ &=\inf_{\substack{\mathbf{P}(dy_t|y^{t-1},\xi^t):t\in\mathbb{N}_0^n\\ \frac{1}{n+1}\sum_{t=0}^n\mathbb{E}\{||\boldsymbol{\xi}_t-\mathbf{y}_t||^2\}\leq D-D^{\min}_{[0,n]}}}\sum_{t=0}^n I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1}),\end{aligned} \tag{34}$$

$$\begin{aligned}&R^G_{\text{in}}(D-D^{\min}_{[0,\infty]})\\ &=\inf_{\substack{\mathbf{P}(dy_t|y^{t-1},\xi^t):t=0,1,\ldots,\infty\\ \limsup_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^n\mathbb{E}\{||\boldsymbol{\xi}_t-\mathbf{y}_t||^2\}\leq D-D^{\min}_{[0,\infty]}}}\bar{R},\end{aligned} \tag{35}$$

where in (34) $(D-D^{\min}_{[0,n]})\in[0,\infty]$, $D^{\min}_{[0,n]}=\frac{1}{n+1}\sum_{t=0}^n\operatorname{trace}(\Sigma^{\mathbf{x}}_{t|t})$, in (35) $D^{\min}_{[0,\infty]}=\limsup_{n\longrightarrow\infty}D^{\min}_{[0,n]}$ and $\bar{R}\triangleq\limsup_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^n I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1})$.

Clearly, if in Definition 2 we assume that $C_t \in \mathbb{R}^{m\times p}$ is full row rank ($m<p$), then, a consequence of Lemma 3 is that (18) > (34) and (19) > (35).

Next, we remark some technical comments on Definition 2.

Remark 2. *(On Definition 2)* **(1)** *The information measure (34) has a finite solution if we ensure that $D-D^{\min}_{[0,n]}\in(0,\infty]$ with $D^{\min}_{[0,n]}<\infty$;* **(2)** *One can take the more stringent pointwise MSE distortion constraint in (34) in which case the problem particularizes to*

$$R^G_{[0,n],\text{in}}(\{D_t-D_t^{\min}\}_{t=0}^n) = \inf_{\substack{\mathbf{P}(dy_t|y^{t-1},\xi^t):\ t\in\mathbb{N}_0^n\\ \mathbb{E}\{||\boldsymbol{\xi}_t-\mathbf{y}_t||^2\}\le D_t-D_t^{\min},\ \forall t}} \sum_{t=0}^n I(\boldsymbol{\xi}^t;\mathbf{y}_t|\mathbf{y}^{t-1}), \tag{36}$$

where $D_t^{\min}=\text{trace}(\Sigma^{\mathbf{x}}_{t|t})$, and $D_t-D_t^{\min}\in(0,\infty]$ with $D_t^{\min}<\infty$, for any t. **(3)** *The lower bound (34) shows an interesting resemblance to the classical remote RDF obtained for* i.i.d. *memoryless Gaussian processes or RVs using non-causal coding [25]. In particular, similar to that case, the distortion constraint in (34) consists of two parts of which only one affects the rates. As a result the other part can be essentially subtracted from the given distortion level.* **(4)** *Definition 2 corresponds to a different optimization problem compared to [16, Eq. (16)]. That paper deals with a lower bound with a soft pointwise MSE distortion constraint and multiple "Lagrange multipliers" (denoted therein by $\{\alpha_t:\ t=0,\ldots,n\}$) that are chosen á priori, whereas we consider a lower bound subject to a hard average total MSE distortion constraint for which we look for a global Lagrange multiplier.* **(5)** *To the best of the authors' knowledge, [8] has never proved the information measure in Definition 2 for jointly Gaussian processes either in finite time or in the asymptotic limit.*

Since the process $\{\boldsymbol{\xi}_t:\ t\in\mathbb{N}_0^n\}$ admits the Markov realization obtained from the KF recursions of (25), then, one can leverage the implicit recursions of [11, Theorem 4.1] (obtained via dynamic programming [34]) to simplify (34) to

$$R^G_{[0,n],\text{in}}(D-D^{\min}_{[0,n]}) = \inf_{\substack{\mathbf{P}(dy_t|y^{t-1},\xi_t):t\in\mathbb{N}_0^n\\ \frac{1}{n+1}\sum_{t=0}^n\mathbb{E}\{||\boldsymbol{\xi}_t-\mathbf{y}_t||^2\}\le D-D^{\min}_{[0,n]}}} \sum_{t=0}^n I(\boldsymbol{\xi}_t;\mathbf{y}_t|\mathbf{y}^{t-1}). \tag{37}$$

Next we use (37) to provide *for the first time*, the *optimal characterization in finite time* of the indirect Gaussian NRDF obtained in (22). To do it, we need a lemma which is an extension of [11] to partially observable Gaussian processes.

Lemma 4. *(Realization of $\{\mathbf{P}^*(dy_t|y^{t-1},\xi_t):\ t\in\mathbb{N}_0^n\}$) For the system model in (1), (2), suppose that the joint process $\{(\mathbf{x}_t,\ \mathbf{y}_t,\ \mathbf{z}_t):\ t\in\mathbb{N}_0^n\}$ is jointly Gaussian and $C_t\in\mathbb{R}^{m\times p}$ in (2) is square and full rank. Then, the following statements hold.*

(1) *Any $\{\mathbf{P}^*(dy_t|y^{t-1},\xi_t):\ t\in\mathbb{N}_0^n\}$ is realized by*

$$\mathbf{y}_t = H_t\left(\boldsymbol{\xi}_t-\widehat{\boldsymbol{\xi}}_{t|t-1}\right)+\widehat{\boldsymbol{\xi}}_{t|t-1}+\mathbf{v}_t,\quad t\in\mathbb{N}_0^n, \tag{38}$$

where $\widehat{\boldsymbol{\xi}}_{t|t-1}\triangleq\mathbb{E}\{\boldsymbol{\xi}_t|\mathbf{y}^{t-1}\}$, $\{\mathbf{v}_t\in\mathbb{R}^p\sim\mathcal{N}(0;\Sigma_{\mathbf{v}_t}):\ t\in\mathbb{N}_0^n\}$ is an independent Gaussian process independent of $\{(\mathbf{w}_t,\ \mathbf{n}_t):\ t\in\mathbb{N}_0^n\}$ and $\mathbf{x}_0$, and $\{H_t\in\mathbb{R}^{p\times p}:\ t\in\mathbb{N}_0^n\}$ are time-varying deterministic matrices (to be designed). Moreover, the innovations process $\{\mathbf{I}_t^{\boldsymbol{\xi}}\in\mathbb{R}^p:\ t\in\mathbb{N}_0^n\}$ of (38) is the orthogonal process given by

$$\mathbf{I}_t^{\boldsymbol{\xi}}=\mathbf{y}_t-\mathbb{E}\{\mathbf{y}_t|\mathbf{y}^{t-1}\}=H_t\left(\boldsymbol{\xi}_t-\widehat{\boldsymbol{\xi}}_{t|t-1}\right)+\mathbf{v}_t, \tag{39}$$

where $\mathbf{I}_t^{\boldsymbol{\xi}}\sim\mathcal{N}(0;\Sigma_{\mathbf{I}_t^{\boldsymbol{\xi}}})$, $\Sigma_{\mathbf{I}_t^{\boldsymbol{\xi}}}=H_t\Sigma^{\boldsymbol{\xi}}_{t|t-1}H_t^T+\Sigma_{\mathbf{v}_t}\succ 0$, with $\Sigma^{\boldsymbol{\xi}}_{t|t-1}\triangleq\mathbb{E}\left\{(\boldsymbol{\xi}_t-\widehat{\boldsymbol{\xi}}_{t|t-1})(\boldsymbol{\xi}_t-\widehat{\boldsymbol{\xi}}_{t|t-1})^T\right\}$.

(2) *Let $\widehat{\boldsymbol{\xi}}_{t|t}\triangleq\mathbb{E}\{\boldsymbol{\xi}_t|\mathbf{y}^t\}$, $\Sigma^{\boldsymbol{\xi}}_{t|t}\triangleq\mathbb{E}\{(\boldsymbol{\xi}_t-\widehat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_t-\widehat{\boldsymbol{\xi}}_{t|t})^T\}$. Then, $\{(\widehat{\boldsymbol{\xi}}_{t|t-1},\ \Sigma^{\boldsymbol{\xi}}_{t|t-1},\widehat{\boldsymbol{\xi}}_{t|t},\ \Sigma^{\boldsymbol{\xi}}_{t|t}):\ t\in\mathbb{N}_0^n\}$ satisfy the following generalized discrete-time forward KF recursions:*

$$\begin{aligned}
&\widehat{\boldsymbol{\xi}}_{t|t}=\widehat{\boldsymbol{\xi}}_{t|t-1}+\mathbf{k}_t^{\boldsymbol{\xi}}\mathbf{I}_t^{\boldsymbol{\xi}},\\
&\widehat{\boldsymbol{\xi}}_{t|t-1}=A_{t-1}\widehat{\boldsymbol{\xi}}_{t-1|t-1},\ \ \widehat{\boldsymbol{\xi}}_{0|-1}=\bar{\boldsymbol{\xi}}_0\\
&\Sigma^{\boldsymbol{\xi}}_{t|t-1}=A_{t-1}\Sigma^{\boldsymbol{\xi}}_{t-1|t-1}A_{t-1}^T+\mathbf{k}_t^{\mathbf{z}}\Sigma_{\mathbf{I}_t^{\mathbf{z}}}\mathbf{k}_t^{\mathbf{z}T},\\
&\mathbf{k}_t^{\boldsymbol{\xi}}=\Sigma^{\boldsymbol{\xi}}_{t|t-1}H_t^T\Sigma^{-1}_{\mathbf{I}_t^{\boldsymbol{\xi}}}\ \textit{(Kalman Gain)},\\
&\Sigma^{\boldsymbol{\xi}}_{t|t}=\Sigma^{\boldsymbol{\xi}}_{t|t-1}-\Sigma^{\boldsymbol{\xi}}_{t|t-1}H_t^T\Sigma^{-1}_{\mathbf{I}_t^{\boldsymbol{\xi}}}H_t\Sigma^{\boldsymbol{\xi}}_{t|t-1},
\end{aligned} \tag{40}$$

where $\Sigma^{\boldsymbol{\xi}}_{0|-1}=\Sigma_{\mathbf{k}_0^{\mathbf{z}}\Sigma_{\mathbf{I}_0^{\mathbf{z}}}\mathbf{k}_0^{\mathbf{z}T}}$, $\Sigma^{\boldsymbol{\xi}}_{t|t}\succeq 0$ and $\Sigma^{\boldsymbol{\xi}}_{t|t-1}\succeq 0$.

(3) *The characterization of $R^G_{[0,n],\text{in}}(D-D^{\min}_{[0,n],\text{in}})$ achieved by (38) is given by*

$$R^G_{[0,n],\text{in}}(D-D^{\min}_{[0,n]}) = \inf_{\substack{H_t\in\mathbb{R}^{p\times p},\ \Sigma_{\mathbf{v}_t}\succeq 0\\ \Sigma^{\boldsymbol{\xi}}_{t|t}\succeq 0,\ \Sigma^{\boldsymbol{\xi}}_{t|t-1}\succeq 0\\ \frac{1}{n+1}\sum_{t=0}^n\text{trace}(\mathcal{G})\le D-D^{\min}_{[0,n]}}} \frac{1}{2}\sum_{t=0}^n\left[\log\frac{|\Sigma^{\boldsymbol{\xi}}_{t|t-1}|}{|\Sigma^{\boldsymbol{\xi}}_{t|t}|}\right]^+, \tag{41}$$

where $\mathcal{G}=(I_p-H_t)\Sigma^{\boldsymbol{\xi}}_{t|t-1}(I_p-H_t)^T+\Sigma_{\mathbf{v}_t}$, for some $(D-D^{\min}_{[0,n]})\in[0,\infty]$.

Proof: **(1)** Since the joint process $\{(\mathbf{x}_t,\mathbf{z}_t,\mathbf{y}_t):\ t\in\mathbb{N}_0^n\}$ is assumed to be jointly Gaussian, then, $\{\mathbf{P}^*(dy_t|y^{t-1},\xi_t):\ t\in\mathbb{N}_0^n\}$ is conditionally Gaussian, and we can obtain the orthogonal realization

$$\mathbf{y}_t=H_t\boldsymbol{\xi}_t+R_t(\mathbf{y}^{t-1})+\mathbf{v}_t,\ t\in\mathbb{N}_0^n, \tag{42}$$

where $R_t(\mathbf{y}^{t-1})\triangleq\Gamma_{t-1}\mathbf{y}^{t-1}$, $\mathbf{P}^*(\cdot|y^{t-1},\xi_t)\sim\mathcal{N}(H_t\boldsymbol{\xi}_t+\Gamma_{t-1}\mathbf{y}^{t-1};\Sigma_{\mathbf{v}_t})$, with $\{(H_t,\Gamma_{t-1}):\ t\in\mathbb{N}_0^n\}$ being deterministic matrices of appropriate dimensions. For such realization, $I(\boldsymbol{\xi}_t;\mathbf{y}_t|\mathbf{y}^{t-1})$ does not depend on $R_t(\cdot)$, $\forall t\in\mathbb{N}_0^n$. Moreover,

$$\begin{aligned}
&\mathbb{E}\left\{||\boldsymbol{\xi}_t-\mathbf{y}_t||^2\right\}=\mathbb{E}\left\{||(I_p-H_t)\boldsymbol{\xi}_t-R_t(\mathbf{y}^{t-1})||^2\right\}\\
&+\text{trace}\left(\Sigma_{\mathbf{v}_t}\right)\\
&\overset{(\star)}{\ge}\mathbb{E}\left\{||(I_p-H_t)\boldsymbol{\xi}_t-R_t^*(\mathbf{y}^{t-1})||^2\right\}+\text{trace}\left(\Sigma_{\mathbf{v}_t}\right),
\end{aligned}$$

where $(\star)$ holds with equality if $R_t(\cdot)=R_t^*(\cdot)=(I_p-H_t)\widehat{\boldsymbol{\xi}}_{t|t-1}$, $\forall t\in\mathbb{N}_0^n$. **(2)** This follows from the discrete-time KF equations. **(3)** The characterization that achieves (38) is obtained from **(1)**, **(2)** and the definition of conditional mutual information $I(\boldsymbol{\xi}_t;\mathbf{y}_t|\mathbf{y}^{t-1})$ at each time instant t. This completes the proof. ■

It should be noted that the characterization in (41) is general and at this point we did not give conditions to ensure existence of a finite solution neither we specify the design of the decision variables $(H_t, \Sigma_{\mathbf{v}_t})$. Moreover, if in the system model (1), (2) we set $C_t = I_p$ and $\mathbf{n}_t = 0, \forall t$, then, in Lemma 4 we obtain $\boldsymbol{\xi}_t = \mathbf{x}_t, \ \forall t, \ \{\mathbf{P}^*(dy_t|y^{t-1}, \xi_t) \equiv \mathbf{P}^*(dy_t|y^{t-1}, x_t) : t \in \mathbb{N}_0^n\}$ and $\widehat{\boldsymbol{\xi}}_{t|t-1} = \mathbb{E}\{\mathbf{x}_t|\mathbf{y}^{t-1}\}$. The analysis will recover as a special case [11, Lemma 5.2].

Proposition 2. *(Full row rank matrix C_t) If in Lemma 4 we assume that $C_t \in \mathbb{R}^{m\times p}$ is full row rank with $m < p$ and $\Sigma_{\mathbf{I}_t^{\boldsymbol{\xi}}} \succ 0$, then, the KF recursions will hold for $\mathbf{k}_t^{\mathbf{z}}\Sigma_{\mathbf{I}_t^{\mathbf{z}}}\mathbf{k}_t^{\mathbf{z}^T} \succeq 0$. Moreover, (22) > (41).*

Proof: This is a consequence of Lemma 3. ■

The next theorem achieves the minimum of (37) parametrized by $(H_t, \Sigma_{\mathbf{v}_t})$.

Theorem 1. *(Characterization of (37))* **(1)** *The optimal minimizer $\{\mathbf{P}^*(dy_t|y^{t-1}, \xi_t) : t \in \mathbb{N}_0^n\}$ that achieves (37) is induced by the parametric realization (38) such that $(H_t, \Sigma_{\mathbf{v}_t})$ are given by*

$$H_t \triangleq \Sigma^{\boldsymbol{\xi}}_{t|t-1} - \Sigma^{\boldsymbol{\xi}}_{t|t}\Sigma^{\boldsymbol{\xi}^{-1}}_{t|t-1}, \ \Sigma_{\mathbf{v}_t} \triangleq \Sigma^{\boldsymbol{\xi}}_{t|t}H_t^T \succeq 0. \tag{43}$$

The choice of $(H_t, \Sigma_{\mathbf{v}_t})$ if $\Sigma_{\mathbf{I}_t^{\boldsymbol{\xi}}} \succ 0$, implies that the optimal minimizer $\{\mathbf{P}^*(dy_t|y^{t-1}, \xi_t) : t \in \mathbb{N}_0^n\}$ is realized by

$$\mathbf{y}_t = H_t\boldsymbol{\xi}_t + (I_p - H_t)A\mathbf{y}_{t-1} + \mathbf{v}_t, \ t \in \mathbb{N}_0^n, \tag{44}$$

where $\mathbf{y}_t \in \mathbb{R}^p$, with $\mathbf{y}_{-1} = 0$.
(2) *Moreover, $R^G_{[0,n],\mathrm{in}}(D - D^{\min}_{[0,n]})$ in (37) parametrized by $(H_t, \Sigma_{\mathbf{v}_t})$ is achieved by the following optimization problem*

$$R^G_{[0,n],\mathrm{in}}(D - D^{\min}_{[0,n]}) = \min_{\substack{0 \preceq \Sigma^{\boldsymbol{\xi}}_{t|t} \preceq \Sigma^{\boldsymbol{\xi}}_{t|t-1} \\ \frac{1}{n+1}\sum_{t=0}^n \mathrm{trace}\left(\Sigma^{\boldsymbol{\xi}}_{t|t}\right) \leq D - D^{\min}_{[0,n]}}} \frac{1}{2}\sum_{t=0}^n \left[\log\frac{|\Sigma^{\boldsymbol{\xi}}_{t|t-1}|}{|\Sigma^{\boldsymbol{\xi}}_{t|t}|}\right]^+, \tag{45}$$

for some $D - D^{\min}_{[0,n]} \in [D_{\min}, D_{\max}] \subset [0, D_{\max}]$.

Proof: From MSE estimation theory we know that the MSE inequality $\sum_{t=0}^n \mathbb{E}\left\{||\boldsymbol{\xi}_t - \mathbf{y}_t||^2\right\} \geq \sum_{t=0}^n \mathbb{E}\left\{||\boldsymbol{\xi}_t - \widehat{\boldsymbol{\xi}}_{t|t}||^2\right\}$ holds for all $(H_t, \ \Sigma_{\mathbf{v}_t})$, $t \in \mathbb{N}_0^n$, and it is achieved if $\widehat{\boldsymbol{\xi}}_{t|t} = \mathbf{y}_t$. The choice of (43) for $\Sigma^{\xi}_{\mathbf{I}} \succ 0$ ensures in (40) that $\mathbf{k}_t^{\boldsymbol{\xi}} = I_p$ and $\widehat{\mathbf{x}}_{t|t-1} = A\mathbf{y}_{t-1}$ and hence via (38) we obtain (44). This means that if in (41) we substitute the scalings in (43), we obtain (45) making sure that the distortion obtained from the optimal MSE estimator can be achieved for the specific rate (objective function). ■

If the choice of $(H_t, \Sigma_{\mathbf{v}_t})$ in Theorem 1 generates $\Sigma_{\mathbf{I}_t^{\boldsymbol{\xi}}} \succeq 0$ with $\mathrm{rank}(\Sigma_{\mathbf{I}_t^{\boldsymbol{\xi}}}) = l < p$, then, (45) will still be achieved by the linear Gaussian "test channel" (44) with reduced dimension $\mathbf{y}_t \in \mathbb{R}^l$ $(l < p)$. This is because by finding the decision variable $\{\Sigma^{\boldsymbol{\xi}}_{t|t} : t \in \mathbb{N}_0^n\}$ one can further compute the rank deficient matrices $(H_t, \Sigma_{\mathbf{v}_t})$ and then discard the $(p - l)$ "inactive" dimensions using singular value decomposition.

Remark 3. *(Existence of solution in Theorem 1) An optimal solution with finite value in (45) exists if (i) $D^{\min}_{[0,n]} < \infty$ for any finite n; (ii) $D - D^{\min}_{[0,n]} > 0$ (non-zero distortion) which implies the strict linear matrix inequality (LMI) constraint $0 \prec \Sigma^{\boldsymbol{\xi}}_{t|t} \preceq \Sigma^{\boldsymbol{\xi}}_{t|t-1}, \ \forall t$.*

One can easily verify via Lemma 2 that if in (2) we set $C_t = I_p$ and $\mathbf{n}_t = 0, \forall t$, then, $D^{\min}_{[0,n]} = 0$ and Theorem 1 recovers as a special case the optimization problem of the classical NRDF for time-varying fully observable Gauss-Markov processes with hard average total MSE distortion, see e.g., [11].

Proposition 3. *(Characterization for full row rank C_t) Suppose that the conditions of Proposition 2 hold and $A_t \in \mathbb{R}^{p\times p}$ in (1) is full rank for any t. Then, the statements of Theorem 1 hold with (22) > (45).*

Proof: This is immediate from Proposition 2. ■

Next, we state the optimal solution of the characterization in Theorem 1 under the conditions of Remark 3.

Theorem 2. *(Optimal numerical solution of (45)) Compute forward in time via (25) $\{(\Sigma^{\mathbf{x}}_{t|t}, \Sigma^{\mathbf{x}}_{t|t-1}) : t \in \mathbb{N}_0^n\}$ such that the conditions of Remark 3 hold. Moreover, introduce the decision variable $\Gamma_t \succ 0$. Then, the optimal solution of (45) for $D > D^{\min}_{[0,n]}$ is semidefinite representable as follows*

$$R^G_{[0,n],\mathrm{in}}(D - D^{\min}_{[0,n]}) = \min_{\substack{\{\Sigma^{\boldsymbol{\xi}}_{t|t} \succ 0, \Gamma_t \succ 0\}_{t=0}^n \\ 0 \prec \Sigma^{\boldsymbol{\xi}}_{0|0} \preceq \Sigma^{\boldsymbol{\xi}}_{0|-1} \\ 0 \prec \Sigma^{\boldsymbol{\xi}}_{t+1|t+1} \preceq \Sigma^{\boldsymbol{\xi}}_{t+1|t}, \ t \in \mathbb{N}_0^{n-1} \\ \Sigma^{\boldsymbol{\xi}}_{n|n} = \Gamma_n}} \frac{1}{2}\sum_{t=0}^n \log|\Gamma_t|^{-1} + c, \tag{46}$$

$$s.\ t. \quad \frac{1}{n+1}\sum_{t=0}^n \mathrm{trace}\left(\Sigma^{\boldsymbol{\xi}}_{t|t}\right) \leq D - D^{\min}_{[0,n]}$$

$$\begin{bmatrix} \Sigma^{\boldsymbol{\xi}}_{t|t} - \Gamma_t & \Sigma^{\boldsymbol{\xi}}_{t|t}A_t^T \\ A_t\Sigma^{\boldsymbol{\xi}}_{t|t} & \Sigma^{\boldsymbol{\xi}}_{t+1|t} \end{bmatrix} \succeq 0, \ t \in \mathbb{N}_0^{n-1} \tag{47}$$

where $c = \frac{1}{2}\log|\Sigma^{\boldsymbol{\xi}}_{0|-1}| + \frac{1}{2}\sum_{t=0}^{n-1}\log|\mathbf{k}^{\mathbf{z}}_{t+1}\Sigma_{\mathbf{I}^{\mathbf{z}}_{t+1}}\mathbf{k}^{\mathbf{z}T}_{t+1}|$, with $\mathbf{k}^{\mathbf{z}}_{t+1}\Sigma_{\mathbf{I}^{\mathbf{z}}_{t+1}}\mathbf{k}^{\mathbf{z}T}_{t+1} \succ 0, \ \forall t$.

Proof: The derivation is similar to [10, Theorem 1]. ■

Next, we stress some technical comments on Theorem 2.

Remark 4. *(On Theorem 2)* **(1)** *To compute the optimal numerical solutions in Theorem 2 is computationally very expensive. First we need to compute $\{(\Sigma^{\mathbf{x}}_{t|t-1}, \Sigma^{\mathbf{x}}_{t|t}) : t \in \mathbb{N}_0^n\}$ of Lemma 2 both of dimension $p \times p$, which correspond to approximately $\mathcal{O}(p^{2.376})$ operations for each time instant t, then, to engage SDP algorithm of which the most computationally expensive step is the Cholesky factorization that requires, in general, approximately $\mathcal{O}(p^3)$ operations at each time instant t. Some additional analysis on the arithmetic complexity of the SDP algorithm is provided in [10, Sec. IV-C]. In fact as we demonstrate in the sequel (see Table I) even for the single stage case at high dimensional problems, the SDP algorithm operates extremely slow. Hence finding alternative optimal or near-optimal algorithmic approaches with reasonable computational complexity aligned with the state of the art large scale networks that operate using computationally limited resources remains an intriguing open problem.* **(2)** *Theorem 2 continues to hold with appropriate changes if we consider the stronger pointwise distortion constraint, i.e., $\mathrm{trace}\left(\Sigma^{\boldsymbol{\xi}}_{t|t}\right) \leq D_t - D_t^{\min}, \ D_t > D_t^{\min}, \ D_t^{\min} < \infty, \ \forall t$.*

Proposition 4. *(Computation of (45) via Proposition 3) Compute forward in time via (25) $\{(\Sigma^{\mathbf{x}}_{t|t}, \Sigma^{\mathbf{x}}_{t|t-1}):\ t \in \mathbb{N}_0^n\}$ such that the conditions of Remark 3 hold. Moreover, introduce the decision variable $\Gamma^1_t \succ 0$ and let the factorization of the singular matrix $\mathbf{k}^{\mathbf{z}}_{t+1}\Sigma_{\mathbf{I}^{\mathbf{z}}_{t+1}}\mathbf{k}^{\mathbf{z}^T}_{t+1} \triangleq B_{t+1}B^T_{t+1}$. Then, the optimal solution of (45) for $D > D^{\min}_{[0,n]}$ is semidefinite representable as follows*

$$
\begin{aligned}
&R^G_{[0,n],\text{in}}(D - D^{\min}_{[0,n]}) = \\
&\min_{\substack{\{\Sigma^{\boldsymbol{\xi}}_{t|t}\succ 0, \Gamma^1_t \succ 0\}^n_{t=0}\\ 0\prec\Sigma^{\boldsymbol{\xi}}_{0|0}\preceq\Sigma^{\boldsymbol{\xi}}_{0|-1}\\ 0\prec\Sigma^{\boldsymbol{\xi}}_{t+1|t+1}\preceq\Sigma^{\boldsymbol{\xi}}_{t+1|t},\ t\in\mathbb{N}_0^{n-1}\\ \Sigma^{\boldsymbol{\xi}}_{n|n}=\Gamma^1_n}} \quad \frac{1}{2}\sum_{t=0}^n \log|\Gamma^1_t|^{-1} + c_1, \qquad (48)\\
&s.\ t. \qquad \frac{1}{n+1}\sum_{t=0}^n \operatorname{trace}\left(\Sigma^{\boldsymbol{\xi}}_{t|t}\right) \leq D - D^{\min}_{[0,n]}\\
&\qquad\quad \begin{bmatrix} I_p - \Gamma^1_t & B^T_{t+1}\\ B_{t+1} & \Sigma^{\boldsymbol{\xi}}_{t+1|t}\end{bmatrix} \succeq 0,\ t\in\mathbb{N}_0^{n-1}
\end{aligned}
$$

where $c_1 = \frac{1}{2}\log|\Sigma^{\boldsymbol{\xi}}_{0|-1}| + \sum_{t=0}^{n-1}\log\operatorname{abs}(|A_t|)$.

Proof: The steps of the proof are known, but the proof for the finite time horizon has not appeared elsewhere, hence we give it for completeness in [35, Appendix B]. ■

Closed-form solutions for scalar time-varying processes: To gain further insights into the solution to this problem, we also study time-varying scalar processes. Specifically, we propose a solution (via a dynamic reverse waterfilling algorithm) under an average total MSE distortion constraint and a closed-form solution under a pointwise MSE distortion. Consider the scalar-valued system model of (1), (2) of the form

$$
\begin{aligned}
\mathbf{x}_{t+1} &= \alpha_t\mathbf{x}_t + \mathbf{w}_t,\ \mathbf{x}_0 = \bar{x},\\
\mathbf{z}_t &= c_t\mathbf{x}_t + \mathbf{n}_t,\quad t\in\mathbb{N}_0^n,
\end{aligned} \qquad (49)
$$

where $\alpha_t \in \mathbb{R}$ and $c_t \in \mathbb{R}\setminus\{0\}$ are non-random, $\mathbf{x}_0 \in \mathbb{R} \sim \mathcal{N}(0;\sigma^2_{\mathbf{x}_0})$ is the initial state, $\mathbf{w}_t \in \mathbb{R} \sim \mathcal{N}(0;\sigma^2_{\mathbf{w}_t})$, $\sigma^2_{\mathbf{w}_t} > 0$ is an independent sequence, $\mathbf{n}_t \in \mathbb{R} \sim \mathcal{N}(0;\sigma^2_{\mathbf{n}_t})$, $\sigma^2_{\mathbf{n}_t} \geq 0$, is an independent sequence, independent of $\{\mathbf{w}_t:\ t\in\mathbb{N}_0^n\}$, whereas $\mathbf{x}_0$ is independent of $\{(\mathbf{w}_t,\mathbf{n}_t):\ t\in\mathbb{N}_0^n\}$. Before we proceed, we denote $\Sigma^{\mathbf{x}}_{t|t} \equiv \sigma^2_{\mathbf{x}_{t|t}}$, $\Sigma^{\mathbf{x}}_{t|t-1} \equiv \sigma^2_{\mathbf{x}_{t|t-1}}$, $\Sigma^{\boldsymbol{\xi}}_{t|t} \equiv \sigma^2_{\boldsymbol{\xi}_{t|t}}$, $\Sigma^{\boldsymbol{\xi}}_{t|t-1} \equiv \sigma^2_{\boldsymbol{\xi}_{t|t-1}}$, $\mathbf{k}^{\mathbf{z}}_t\Sigma_{\mathbf{I}^{\mathbf{z}}_t}\mathbf{k}^{\mathbf{z}^{\text{T}}}_t = \frac{c_t^2\sigma^4_{\mathbf{x}_{t|t-1}}}{c_t^2\sigma^2_{\mathbf{x}_{t|t-1}}+\sigma^2_{\mathbf{n}_t}} \equiv \sigma^2_{\upsilon_t}$, for any $t\in\mathbb{N}_0^n$. Additionally, we simplify the characterization of Theorem 1 for scalar processes under the assumption that the total rates yield a finite solution, i.e.,

$$
R^G_{[0,n],\text{in}}(D - D^{\min}_{[0,n]}) = \inf_{\substack{0<\sigma^2_{\boldsymbol{\xi}_{t|t}}\leq\sigma^2_{\boldsymbol{\xi}_{t|t-1}}\\ \frac{1}{n+1}\sum_{t=0}^n\left(\sigma^2_{\boldsymbol{\xi}_{t|t}}\right)\leq D-D^{\min}_{[0,n]}}} \frac{1}{2}\sum_{t=0}^n\log\left(\frac{\sigma^2_{\boldsymbol{\xi}_{t|t-1}}}{\sigma^2_{\boldsymbol{\xi}_{t|t}}}\right), \qquad (50)
$$

where $D - D^{\min}_{[0,n]} > 0$, $D^{\min}_{[0,n]} = \frac{1}{n+1}\sum_{t=0}^n\sigma^2_{\mathbf{x}_{t|t}} < \infty$. In the next theorem, we give the optimal solution of (50) via a dynamic reverse-waterfilling algorithm.

Theorem 3. *(Optimal solution of (50)) The optimal parametric solution of (50) can be computed as follows:*

$$
R^G_{[0,n],\text{in}}(D - D^{\min}_{[0,n]}) = \frac{1}{2}\sum_{t=0}^n\log\left(\frac{\sigma^2_{\boldsymbol{\xi}_{t|t-1}}}{\sigma^2_{\boldsymbol{\xi}_{t|t}}}\right), \qquad (51)
$$

such that $\sigma^2_{\boldsymbol{\xi}_{t|t}} > 0$ is computed at each time instant as follows:

$$
\sigma^2_{\boldsymbol{\xi}_{t|t}} = \begin{cases}\sigma^{2,*}_{\boldsymbol{\xi}_{t|t}} & \text{if } \sigma^{2,*}_{\boldsymbol{\xi}_{t|t}} < \sigma^2_{\boldsymbol{\xi}_{t|t-1}}\\ \sigma^2_{\boldsymbol{\xi}_{t|t-1}} & \text{if } \sigma^{2,*}_{\boldsymbol{\xi}_{t|t}} \geq \sigma^2_{\boldsymbol{\xi}_{t|t-1}}\end{cases},\ \forall t, \qquad (52)
$$

with $\sum_{t=0}^n\sigma^2_{\boldsymbol{\xi}_{t|t}} = (n+1)(D - D^{\min}_{[0,n]})$ and

$$
\sigma^{2,*}_{\boldsymbol{\xi}_{t|t}} = \begin{cases}\frac{1}{\beta_{t,t+1}}\left(\sqrt{1+\frac{\beta_{t,t+1}}{\theta^*}}-1\right),\ \forall t\in\mathbb{N}_0^{n-1}\\ \frac{1}{2\theta^*},\ t=n,\end{cases}, \qquad (53)
$$

where $\theta^* > 0$, $\beta_{t,t+1} \triangleq \frac{2\alpha_t^2}{\sigma^2_{\upsilon_{t+1}}}$, and $D > D^{\min}_{[0,n]}$ with $D^{\min}_{[0,n]} < \infty$.

Proof: The proof is based on KKT conditions [36, Chapter 5.5.3] and can be obtained following [17, Theorem 2]. ■

Remark 5. *(On Theorem 3) Suppose that in (49) we set $c_t = 1$ and $\mathbf{n}_t = 0$, $\forall t$. Then, using Lemma 2 it can be easily shown that $\sigma^2_{\upsilon_t} = \sigma^2_{\mathbf{w}_{t-1}}$, $\beta_{t,t+1} = \frac{2\alpha_t^2}{\sigma^2_{\mathbf{w}_t}}$ and $D^{\min}_{[0,n]} = 0$, $\forall t$, and we recover [37, Theorem 1].*

In Algorithm 1, we implement the optimal solution of Theorem 3.

Remark 6. *(On Algorithm 1) Algorithm 1 ensures linear convergence in finite time via a bisection method for a given error tolerance ϵ by picking as starting points appropriate nominal range of values for θ (i.e., $\theta^{\min}$ and $\theta^{\max}$). The convergence of bisection method implies that θ converges, hence $\frac{1}{n+1}\sum_{t=0}^n\sigma^2_{\boldsymbol{\xi}_{t|t}} \longrightarrow (D - D^{\min}_{[0,n]})$ within the error tolerance ϵ. We note that the nominal values of $\theta^{\min}$ and $\theta^{\max}$ vary depending on the system model (49). The most computationally expensive operation in Algorithm 1 is the for loop and the bisection method that yield a time complexity of approximately $\mathcal{O}(n\log(n))$ (linearithmic time complexity). In Fig. 2 we illustrate a numerical simulation of the average running time needed for Algorithm 1 to execute (vs) the time horizon n when the error tolerance is $\epsilon = 10^{-9}$. We consider that each n is the mean of 10000 time instants.*

Next we give the analytical expression of (50) under pointwise MSE distortion constraints.

Corollary 1. *(Analytical solution) Find forward in time $\{(\sigma^2_{\mathbf{x}_{t|t}},\sigma^2_{\mathbf{x}_{t|t-1}}):\ t\in\mathbb{N}_0^n\}$ via (25) and let $D_t > D^{\min}_t = \sigma^2_{\mathbf{x}_{t|t}}$, $\forall t$. Then, the closed form solution of (50) under a pointwise MSE distortion constraint is given as follows*

$$
R^G_{[0,n],\text{in}}(\{D_t - D^{\min}_t\}^n_{t=0}) = \frac{1}{2}\sum_{t=0}^n\left[\log\left(\frac{\sigma^2_{\boldsymbol{\xi}_{t|t-1}}}{D_t - D^{\min}_t}\right)\right]^+,
$$

where $\sigma^2_{\boldsymbol{\xi}_{t|t-1}} = \alpha^2_{t-1}(D_{t-1} - D^{\min}_{t-1}) + \sigma^2_{\upsilon_t}$.

Proof: The proof is similar to Theorem 2 by employing KKT conditions hence it is omitted. ■

Algorithm 1 Implementation of Theorem 3

Initialize: number of time-steps n; error tolerance ϵ; nominal minimum and maximum value of θ, denoted by $\theta^{\min}$ and $\theta^{\max}$; initial variance $\sigma^2_{\mathbf{x}_{0|-1}}=\sigma^2_{\mathbf{x}_0}$; set values for $\{(\alpha_t,\sigma^2_{\mathbf{w}_t},c_t,\sigma^2_{\mathbf{n}_t}):~t\in\mathbb{N}_0^n\}$ of (49).
for $t=0:n$ **do**
  Compute $(\sigma^2_{\mathbf{x}_{t|t}},\sigma^2_{\mathbf{x}_{t|t-1}})$ via (25).
end for
Compute $D^{\min}_{[0,n]}=\frac{1}{n+1}\sum_{t=0}^n\sigma^2_{\mathbf{x}_{t|t}}<\infty$; set the distortion level $D>D^{\min}_{[0,n]}$; Pick some $\theta\in[\theta^{\min},\theta^{\max}]$; flag $=0$.
while flag $=0$ **do**
  for $t=0:n$ **do**
    Compute $\sigma^{2,*}_{\boldsymbol{\xi}_{t|t}}$ according to (53).
    Compute $\sigma^2_{\boldsymbol{\xi}_{t|t}}$ according to (52).
    if $t<n$ **then**
      Compute $\sigma^2_{\boldsymbol{\xi}_{t+1|t}}$ according to $\sigma^2_{\boldsymbol{\xi}_{t+1|t}}\triangleq\alpha_t^2\sigma^2_{\boldsymbol{\xi}_{t|t}}+\sigma^2_{\upsilon_{t+1}}$.
    end if
  end for
  if $\frac{1}{n+1}\sum_{t=0}^n\sigma^2_{\boldsymbol{\xi}_{t|t}}-(D-D^{\min}_{[0,n]})\geq\epsilon$ **then**
    Set $\theta^{\min}=\theta$.
  else
    Set $\theta^{\max}=\theta$.
  end if
  if $\theta^{\max}-\theta^{\min}\geq\frac{\epsilon}{n+1}$ **then**
    Compute $\theta=\frac{(\theta^{\min}+\theta^{\max})}{2}$.
  else
    flag $\leftarrow 1$
  end if
end while
Output: $\{\sigma^2_{\boldsymbol{\xi}_{t|t}}~:~t\in\mathbb{N}_0^n\}$, $\{\sigma^2_{\boldsymbol{\xi}_{t|t-1}}~:~t\in\mathbb{N}_0^n\}$, for a given distortion level $D-D^{\min}_{[0,n]}$.

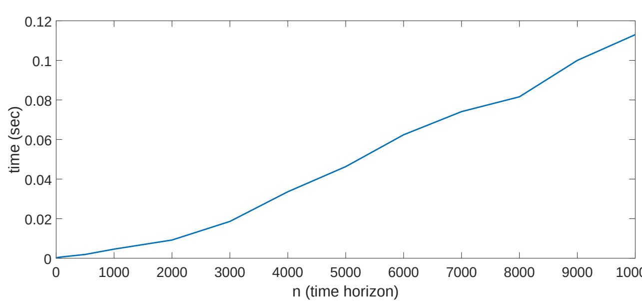

Fig. 2: Demonstration of the average running time needed for Algorithm 1 to execute for 10000 instances. Simulations were performed in MATLAB and tested on a single CPU with an Intel Core i7 processor at 2.6 GHz and 16 GB RAM.

For the special case of time-varying fully-observable Gauss-Markov processes it can be easily seen following precisely Remark 5 that we can recover [38, Corollary 2].

V. Infinite Time Horizon Problems

In this section, we analyze the asymptotic limit of (37). To do it, we restrict our system model (1), (2) to time-invariant processes, i.e., $A_t=A$, $\Sigma_{\mathbf{w}_t}=\Sigma_{\mathbf{w}}$, $C_t=C$, $\Sigma_{\mathbf{n}_t}=\Sigma_{\mathbf{n}}$, $\forall t$. We apply known results for the convergence of the discrete time Riccati equation (DRE) of Lemma 2. These results can be found for instance in [30, Chapter 7.3], [29, Appendix E] or [31]. Before we state a lemma, we note that in the sequel, we adopt for simplicity the following notation

$$\begin{aligned}\Sigma_t&=\Sigma^{\mathbf{x}}_{t|t},~\text{and}~\Sigma=\lim_{t\longrightarrow\infty}\Sigma_t\\ \Pi_t&=\Sigma^{\mathbf{x}}_{t|t-1},~\text{and}~\Pi=\lim_{t\longrightarrow\infty}\Pi_t\\ \bar{\Sigma}_t&=\mathbf{k}^{\mathbf{z}}_t\Sigma_{\mathbf{I}^{\mathbf{z}}_t}\mathbf{k}^{\mathbf{z}^{\mathrm{T}}}_t,~\text{and}~\bar{\Sigma}=\lim_{t\longrightarrow\infty}\bar{\Sigma}_t.\end{aligned}\tag{54}$$

Lemma 5. *[29], [30] (Necessary and sufficient conditions for convergence of the time-invariant DRE of Lemma 2 to a unique stabilizing solution) Let $(A,\Sigma_{\mathbf{w}},C,\Sigma_{\mathbf{n}})\in\mathbb{R}^{p\times p}\times\mathbb{R}^{p\times p}\times\mathbb{R}^{m\times p}\times\mathbb{R}^{m\times m}$. Then, the DRE that corresponds to Lemma 2 is the following*

$$\begin{aligned}\Pi_t=&A\Pi_{t-1}A^{\mathrm{T}}-A\Pi_{t-1}C^{\mathrm{T}}(C\Pi_{t-1}C^{\mathrm{T}}+\Sigma_{\mathbf{n}})^{-1}C\Pi_{t-1}A^{\mathrm{T}}\\&+\Sigma_{\mathbf{w}},~t\in\mathbb{N}_0,\end{aligned}\tag{55}$$

where $\Pi_0\succ 0$ (always positive definite). Moreover, the corresponding discrete time algebraic Riccati equation (DARE) is as follows

$$\Pi=A\Pi A^{\mathrm{T}}-A\Pi C^{\mathrm{T}}(C\Pi C^{\mathrm{T}}+\Sigma_{\mathbf{n}})^{-1}C\Pi A^{\mathrm{T}}+\Sigma_{\mathbf{w}}.\tag{56}$$

Then, the following statement holds. Let the pair (A,C) to be detectable and the pair $(A,\Sigma_{\mathbf{w}}^{\frac{1}{2}})$ to be stabilizable (or controllable on and outside the unit circle). Then, any solution of (55), i.e, $\{\Pi_t~:~t\in\mathbb{N}_0\}$, is such that $\lim_{t\longrightarrow\infty}\Pi_t=\Pi$, $\Pi\succeq 0$ for any $\Pi_0\succeq 0$ which corresponds to the maximal unique stabilizing solution of (56). This further means that the steady-state KF, i.e., the limiting expression of $\widehat{\mathbf{x}}_{t|t}\equiv\boldsymbol{\xi}_t$ in (25) is asymptotically stable.

Next, we provide an example applied to scalar processes, to illustrate the concept of Lemma 5.

Example 1. *(Solution of DARE for scalar processes) Consider the time-invariant version of the system model in (49), i.e., $\alpha_t=\alpha\in\mathbb{R}$, $\sigma^2_{\mathbf{w}_t}=\sigma^2_{\mathbf{w}}>0$, $c_t=c\in\mathbb{R}\setminus\{0\}$, $\sigma^2_{\mathbf{n}_t}=\sigma^2_{\mathbf{n}}\geq 0$, $\forall t$. Then, the time-invariant scalar-valued DRE of (55) is*

$$\Pi_t=\alpha^2\Pi_{t-1}+\sigma^2_{\mathbf{w}}-\frac{\alpha^2c^2\Pi^2_{t-1}}{c^2\Pi_{t-1}+\sigma^2_{\mathbf{n}}},~t\in\mathbb{N}_0,\tag{57}$$

where $\Pi_0>0$. The corresponding scalar-valued DARE of (56) is as follows

$$\Pi=\alpha^2\Pi+\sigma^2_{\mathbf{w}}-\frac{\alpha^2c^2\Pi^2}{c^2\Pi+\sigma^2_{\mathbf{n}}}.\tag{58}$$

Moreover, introduce the pairs (a,c) and $(\alpha,(\sigma^2_{\mathbf{w}})^{\frac{1}{2}})$. Then, by definition, the pair (α,c) is always detectable and the pair $(\alpha,(\sigma^2_{\mathbf{w}})^{\frac{1}{2}})$ is always stabilizable (because $\sigma^2_{\mathbf{w}}>0$). Hence, from Lemma 5 any solution of (57) is such that $\lim_{t\longrightarrow\infty}\Pi_t=\Pi$, with $\Pi\geq 0$ that corresponds to the unique stabilizing solution of (58). In what follows, we compute the closed form solution of $\Pi\geq 0$. Note that (58) can be reformulated to *the quadratic equation $c^2\Pi^2+\gamma\Pi-\sigma^2_{\mathbf{w}}\sigma^2_{\mathbf{n}}=0$, where $\gamma\triangleq((1-\alpha)^2\sigma^2_{\mathbf{n}}-c^2\sigma^2_{\mathbf{w}})$, that gives the following two solutions*

$$\Pi=\begin{cases}\frac{\sqrt{\gamma^2+4c^2\sigma^2_{\mathbf{w}}\sigma^2_{\mathbf{n}}}-\gamma}{2c^2}>0\\ \frac{-\sqrt{\gamma^2+4c^2\sigma^2_{\mathbf{w}}\sigma^2_{\mathbf{n}}}-\gamma}{2c^2}<0.\end{cases} \tag{59}$$

Clearly, by conditions the negative solution of Π is rejected. Hence from (59) we have that the unique stabilizing solution is not only non-negative but also positive. This is because from by definition of our system model ($\Pi_0>0$ because $\sigma^2_{\mathbf{w}}>0$).

Special cases: (i) Suppose that in (57) we let $\sigma^2_{\mathbf{n}}=0$. Then, from (58) we obtain $\Pi=\sigma^2_{\mathbf{w}}>0$. (ii) Suppose that in (57) we let $\alpha=0$. Then, from (58) we obtain $\Pi=\sigma^2_{\mathbf{w}}>0$. (iii) Suppose that in (57) we let $\alpha=0$, $\sigma^2_{\mathbf{n}}=0$. Then, from (57) we obtain $\Pi=\sigma^2_{\mathbf{w}}>0$.

Remark 7. *(On Lemma 5) In Lemma 5 we gave necessary and sufficient conditions for the DARE of the á priori error covariance in Lemma 2 to converge to its steady state. We observed via Example 1 that this value is always positive for scalar processes. Clearly, what we observe for scalar processes holds for multidimensional processes because $\Sigma_{\mathbf{w}}\succ 0$ with $\Pi_t\succ 0,~\forall t$ (from Lemma 2). Moreover, the steady state of the á priori error covariance in Lemma 2 implies the convergence of the á posteriori error covariance as well. That case however is slightly different because we can allow an initial condition $\Sigma_0\succeq 0$ (see Lemma 2). In fact if in our time-invariant system model we assume $\Sigma_{\mathbf{n}}=0$ and $C=I_p$, then $\Sigma^{\mathbf{x}}_{t|t}=0,~\forall t$, and as a result $\Sigma=\lim_{t\longrightarrow\infty}\Sigma_t=0$.*

Next, we derive necessary and sufficient conditions for the pre-KF algorithm to converge to its steady-state values and sufficient conditions to ensure a time-invariant solution of the characterization in (45).

Theorem 4. *(Asymptotic characterization of (45)) Suppose that the system (1), (2) is restricted to time-invariant processes with the pair (A,C) detectable and the pair $(A,\Sigma^{\frac{1}{2}}_{\mathbf{w}})$ stabilizable. Moreover restrict the test-channel distribution $\mathbf{P}(dy_t|y_{t-1},\xi_t)$ to be time-invariant and the output distribution $\mathbf{P}(dy_t|y_{t-1})$ to be time-invariant with a unique invariant distribution. Then, if $R^G_{[0,\infty],\text{in}}(D-D^{\min}_{[0,\infty]})<\infty$, for $D-D^{\min}_{[0,\infty]}\in(0,\infty]$, this is given by*

$$R^G_{[0,\infty],\text{in}}(D-D^{\min}_{[0,\infty]})=\inf_{\substack{0\prec\Sigma^{\boldsymbol{\xi}}\preceq\Pi^{\boldsymbol{\xi}}\\ \text{trace}\left(\Sigma^{\boldsymbol{\xi}}\right)\leq D-D^{\min}_{[0,\infty]}}}\frac{1}{2}\log\frac{|\Pi^{\boldsymbol{\xi}}|}{|\Sigma^{\boldsymbol{\xi}}|}, \tag{60}$$

where $\Sigma^{\boldsymbol{\xi}}\succ 0$ and $\Pi^{\boldsymbol{\xi}}\succ 0$ are the time-invariant values of $\Sigma^{\boldsymbol{\xi}}_{t|t}$, and $\Sigma^{\boldsymbol{\xi}}_{t|t-1}$, respectively. Moreover,

$$\Pi^{\boldsymbol{\xi}}=A\Sigma^{\boldsymbol{\xi}}A^{\mathrm{T}}+\bar{\Sigma} \tag{61}$$

$$D^{\min}_{[0,\infty]}=\text{trace}(\Sigma). \tag{62}$$

Finally, (60) is achieved by a time-invariant linear Gaussian "test channel" $\mathbf{P}^*(dy_t|y_{t-1},\xi_t)$ of the form

$$\mathbf{y}_t=H\boldsymbol{\xi}_t+(I_p-H)A\mathbf{y}_{t-1}+\mathbf{v}_t, \tag{63}$$

where $H=I_p-\Sigma^{\boldsymbol{\xi}}(\Pi^{\boldsymbol{\xi}})^{-1}$ and $\mathbf{v}_t\sim\mathcal{N}(0;\Sigma_{\mathbf{v}})$, $\Sigma_{\mathbf{v}}=\Sigma^{\boldsymbol{\xi}}H^{\mathrm{T}}\succeq 0$.

Proof: See Appendix A. ■

Remark 8. *(On Theorem 4)* **(1)** *If in the time-variant version of the system model (1), (2) we let $C=I_p$ and $\Sigma_{\mathbf{n}}=0$, then Theorem 4 we can easily recover the known result obtained for the infinite time horizon of the time-invariant fully observable Gauss-Markov processes (see, for instance [10, Eq. (27)], [4, Theorem 3]) because $\bar{\Sigma}=\Sigma_{\mathbf{w}}\succ 0$ and $D^{\min}_{[0,\infty]}=0$.* **(2)** *Clearly, Theorem 4 continues to hold if Propositions 2-4 hold.*

In what follows, we give the optimal numerical solution of the problem in Theorem 4.

Corollary 2. *(Optimal numerical solution of (60)) The optimal numerical solution of (60) is semidefinite representable as follows. Introduce the decision variable $\Gamma\succ 0$ with $\Sigma^{\boldsymbol{\xi}}\succ 0$ and $\bar{\Sigma}\succ 0$. Then, for $D>D^{\min}_{[0,\infty]}$ we obtain*

$$\begin{aligned}R^G_{[0,\infty],\text{in}}(D-D^{\min}_{[0,\infty]})=&\min_{\substack{\Gamma\succ 0\\ 0\prec\Sigma^{\boldsymbol{\xi}}\preceq\Pi^{\boldsymbol{\xi}}}}\frac{1}{2}\log|\Gamma|^{-1}+\frac{1}{2}\log|\bar{\Sigma}|,\\ s.~t.\qquad &\text{trace}\left(\Sigma^{\boldsymbol{\xi}}\right)\leq D-D^{\min}_{[0,\infty]}\\ &\begin{bmatrix}\Sigma^{\boldsymbol{\xi}}-\Gamma & \Sigma^{\boldsymbol{\xi}}A^{\mathrm{T}}\\ A\Sigma^{\boldsymbol{\xi}} & \Pi^{\boldsymbol{\xi}}\end{bmatrix}\succeq 0.\end{aligned} \tag{64}$$

Proof: The proof is a special case of Theorem 2 hence we omit it. ■

We note that the numerical solution of (60) under the structural conditions of Proposition 2 can be derived as a special case of Proposition 4. We will not include this SDP representation as it follows similar to Corollary 2.

In the sequel, we derive strong structural properties on (60) that allow for a simplified optimization problem that can be optimally solved via a reverse-waterfilling algorithm.

Proposition 5. *(Strong structural properties on (60)) Suppose that in the characterization of (60) one of the following structures between $(A,\bar{\Sigma})$ hold.*

(i) *Suppose that $A=\alpha I_p$ (scalar matrix) and $\bar{\Sigma}\succeq 0$;*
(ii) *Suppose that A is real symmetric and $\bar{\Sigma}=\sigma^2_{\bar{\Sigma}}I_p$ (scalar matrix);*
(iii) *Suppose that $A=\bar{\Sigma}\succ 0$;*

Then $(A,\Sigma^{\boldsymbol{\xi}},\bar{\Sigma})$ commute by pairs[4] *and consequently $(\Sigma^{\boldsymbol{\xi}},\Pi^{\boldsymbol{\xi}})$ commute.*

Proof: The proof is similar to [17, Proposition 1] thus we omit it. ■

Theorem 5. *(Optimal numerical solution of (60)) Suppose that in the time-invariant version of (2), C is full rank ($m=p$). Then, if any of the structural conditions of Proposition 5 is satisfied, the following reverse-waterfilling solution holds*

$$R^G_{[0,\infty],\text{in}}(D-D^{\min}_{[0,\infty]})=\min_{\substack{0<\mu_{\Sigma^{\boldsymbol{\xi}},i}\leq\mu_{\Pi^{\boldsymbol{\xi}},i}\\ \sum_{i=1}^p\mu_{\Sigma^{\boldsymbol{\xi}},i}\leq D-D^{\min}_{[0,\infty]}}}\frac{1}{2}\sum_{i=1}^p\log\left(\frac{\mu_{\Pi^{\boldsymbol{\xi}},i}}{\mu_{\Sigma^{\boldsymbol{\xi}},i}}\right). \tag{65}$$

[4] Details on this concept can be found in e.g., [33, Theorem 21.13.1].

Moreover, the optimal parametric solution of (65) *can be computed for* $\mu_{\Sigma\boldsymbol{\xi},i} > 0$*, and any* i *as follows:*

$$\mu_{\Sigma\boldsymbol{\xi},i} = \begin{cases} \mu^*_{\Sigma\boldsymbol{\xi},i} & \text{if } \mu^*_{\Sigma\boldsymbol{\xi},i} < \mu_{\Pi\boldsymbol{\xi},i} \\ \mu_{\Pi\boldsymbol{\xi},i} & \text{if } \mu^*_{\Sigma\boldsymbol{\xi},i} \geq \mu_{\Pi\boldsymbol{\xi},i} \end{cases}, \ \forall i, \tag{66}$$

with $\sum_{i=1}^{p} \mu_{\Sigma\boldsymbol{\xi},i} = (D - D^{\min}_{[0,\infty]})$ *and*

$$\mu^*_{\Sigma\boldsymbol{\xi},i} = \begin{cases} \frac{1}{\mu_{\Upsilon,i}}\left(\sqrt{1+\frac{\mu_{\Upsilon,i}}{\theta^*}}-1\right), & \mu_{\Upsilon,i} > 0 \text{ for some } i \\ \frac{1}{2\theta^*}, & \mu_{\Upsilon,i} = 0 \end{cases}, \tag{67}$$

where $\theta^* > 0$*,* $\mu_{\Upsilon,i} \triangleq \frac{2\mu_{A^2,i}}{\mu_{\bar{\Sigma},i}} > 0$ *and* $D > D^{\min}_{[0,\infty]}$.

Proof: The proof follows similar steps to the derivation of [17, Theorem 2] hence we omit it. ∎

Proposition 6. *(The case of full row rank matrix* C*) Suppose that in the time-invariant version of* (2)*,* $C \in \mathbb{R}^{m\times p}$ *is full row rank with* $m < p$*. Then, under Proposition* 5*, (i), an approximate reverse-waterfilling solution is obtained via Theorem* 5 *with* $\bar{\Sigma}$ *replaced by* $\bar{\Sigma}_\epsilon = \bar{\Sigma} + \epsilon I_p$*.*

Proof: If C is full row rank with $m < p$, we have $\bar{\Sigma} \succeq 0$. Moreover, if Proposition 5, (i), holds, then, we use a standard continuity argument, that is, there exists a $\delta > 0$ such that $\bar{\Sigma}_\epsilon = \bar{\Sigma} + \epsilon I_p$ is nonsingular for all $\epsilon \in (0, \delta)$ (see, e.g., [39, Theorem 2.9]). In other words, we create a $\bar{\Sigma}_\epsilon \succ 0$, and follow similar steps to the derivation of [17, Theorem 2]. Then, by taking in the computations that $\lim_{\epsilon \longrightarrow 0^+} \bar{\Sigma}_\epsilon$ the result follows. This completes the proof. ∎

An implementation of the reverse-waterfilling solution of Theorem 5 (or Proposition 6) is provided in Algorithm 2.

Algorithm 2 Implementation of Theorem 5

Initialize: error tolerance ϵ; nominal minimum and maximum value of θ, i.e., $\theta^{\min}$ and $\theta^{\max}$; set values for $(A, \Sigma_{\mathbf{w}}, C, \Sigma_{\mathbf{n}})$ of (49) so that the par (A, C) is detectable and the pair $(A, \Sigma_{\mathbf{w}}^{\frac{1}{2}})$ is stabilizable.
Find the unique stabilizing solution Π and the steady-state value of Σ via (56) and compute $D^{\min}_{[0,\infty]} = \text{trace}(\Sigma) < \infty$; choose distortion level $D > \text{trace}(\Sigma)$; Pick $\theta \in [\theta^{\min}, \theta^{\max}]$; find the eigenvalues of $(A, \bar{\Sigma})$, i.e., $\{\mu_{A,i} : i \in \mathbb{N}_1^p\}$, $\{\mu_{\bar{\Sigma},i} : i \in \mathbb{N}_1^p\}$ (in decreasing order); flag $= 0$.
while flag $= 0$ **do**
  Compute $\mu_{\Sigma\boldsymbol{\xi},i}$, $\forall i$, as follows:
  for $i = 1 : p$ **do**
    Compute $\mu^*_{\Sigma\boldsymbol{\xi},i}$ according to (67).
    Compute $\mu_{\Sigma\boldsymbol{\xi},i}$ according to (66).
  end for
  if $\theta^{\max} - \theta^{\min} \geq \epsilon$ **then**
    Compute $\theta = \frac{(\theta^{\min}+\theta^{\max})}{2}$.
  else
    flag $\leftarrow 1$
  end if
end while
Output: $\{\mu_{\Sigma\boldsymbol{\xi},i} : i \in \mathbb{N}_1^p\}$, $\{\mu_{\Pi\boldsymbol{\xi},i} : i \in \mathbb{N}_1^p\}$, for a given distortion level $D - \text{trace}(\Sigma)$.

Remark 9. *(Complexity of Algorithm* 2*) The convergence of Algorithm* 2 *is guaranteed for finite dimensional matrices due to the bisection method, similar to Algorithm* 1*. The most computationally expensive parts in Algorithm* 2 *are the matrix multiplications in the computation of the DARE of the steady-state pre-KF recursions which can have a time complexity of approximately* $\mathcal{O}(p^3)$ *followed by the for loop and the bisection method with approximately linearithmic time complexity similar to Algorithm* 1*, i.e.,* $\mathcal{O}(p\log(p))$*. Hence the overall time complexity is approximately* $\mathcal{O}(p^3 + p\log(p))$*. Nevertheless, if we optimize matrix multiplication using for example the current state of the art computing approaches that allow time complexity of around* $\mathcal{O}(p^{2.37286})$*)* [40] *the complexity can further reduce to* $\mathcal{O}(p^{2.37286} + p\log(p))$*. In Table* I *we compare the general optimal solution obtained via SDP in Corollary* 2 *with the structural optimal solution obtained in Theorem* 5 *and implemented in Algorithm* 2 *for the same input data and distortion level. For low dimensional vector systems (i.e.,* $p = 10$*) we compute the average computational time needed for* 1000 *instances using both computational methods for an error tolerance of* $\epsilon = 10^{-9}$*. We see that Algorithm* 2 *is approximately* 550 *times faster than SDP. For medium size vector systems (i.e.,* $p = 100$*) we perform the same experiment for* 100 *instances with* $\epsilon = 10^{-7}$*. The results show that Algorithm* 2 *is approximately* 17500 *times faster than SDP. We note that to obtain a result from SDP for* 1000 *instances would require days therefore we did not attempt with the specific computer such experiment. In addition, it is likely that the result for both SDP and Algorithm* 2 *would not change much. For high dimensional vector systems (i.e.,* $p = 500$*) the result is not-conclusive because SDP would take many days to give a relatively fair result even for* 100 *instances. In contrast Algorithm* 2 *operates fine as illustrated in Table* I*. The results clearly demonstrate that Algorithm* 2 *is much more appealing choice to use when solving problems with certain structure or systems with computationally limited resources as opposed to the SDP algorithm.*

Solver (Numb. dimens. $p = 10$**)**	**Mean** (sec)	**Numb. inst.**
SDP (by default $\epsilon = 10^{-9}$)	0.7134	1000
Algorithm 2 ($\epsilon = 10^{-9}$)	0.0013	1000
Solver (Numb. dimens. $p = 100$**)**	**Mean** (sec)	**Numb. inst.**
SDP (by default $\epsilon = 10^{-7}$)	725.0770	100
Algorithm 2 ($\epsilon = 10^{-7}$)	0.0412	100
Solver (Numb. dimens. $p = 500$**)**	**Mean** (sec)	**Numb. inst.**
SDP	non-conclusive	insufficient
Algorithm 2 ($\epsilon = 10^{-9}$)	6.8997	1000

TABLE I: Comparison of the computational time needed between SDP in Corollary 2 and Algorithm 2. Simulations were performed in MATLAB and tested on a single CPU with an Intel Core i7 processor at 2.6 GHz and 16 GB RAM.

We conclude this section, by finding the closed-form solution of the time-invariant system model of (49).

Corollary 3. *(Closed form solution: time-invariant scalar processes) Consider the characterization of Theorem* 4 *restricted to time-invariant scalar Gaussian processes. Then for*

$D > D^{\min}_{[0,\infty]} = \Sigma$, the closed form solution of $R^G_{\rm in}(D-\Sigma)$ is as follows

$$R^G_{\rm in}(D-\Sigma) = \frac{1}{2}\log\left(\alpha^2 + \frac{\bar{\Sigma}}{D-\Sigma}\right) \tag{68}$$

where

$$\bar{\Sigma} = \frac{c^2\Pi^2}{c^2\Pi + \sigma^2_{\mathbf{n}}}, \tag{69}$$

with $\Pi > 0$ given by the unique stabilizing solution of (58) whereas $\Sigma \geq 0$ is given by the non-negative solution of the quadratic equation

$$\alpha^2 c^2 \Sigma^2 + \bar{\gamma}\Sigma - \sigma^2_{\mathbf{w}}\sigma^2_{\mathbf{n}} = 0, \tag{70}$$

where $\bar{\gamma} = (1-\alpha^2)\sigma^2_{\mathbf{n}} + c^2\sigma^2_{\mathbf{w}}$.

Proof: For scalar processes, the characterization in Theorem 4 simplifies to

$$R^G_{\rm in}(D - D^{\min}_{[0,\infty]}) = \min_{\substack{0<\Sigma^{\boldsymbol{\xi}}\leq\Pi^{\boldsymbol{\xi}} \\ \Sigma^{\boldsymbol{\xi}}\leq D-D^{\min}_{[0,\infty]}}} \frac{1}{2}\log\left(\frac{\Pi^{\boldsymbol{\xi}}}{\Sigma^{\boldsymbol{\xi}}}\right). \tag{71}$$

where $\Pi^{\boldsymbol{\xi}} = \alpha^2\Sigma^{\boldsymbol{\xi}} + \bar{\Sigma}$, $\bar{\Sigma}$ is given by (69) and $D^{\min}_{[0,\infty]} = \Sigma \geq 0$, i.e., the unique stabilizing solution obtained for scalar processes given by (70). The problem in (71) is convex concerning $\Sigma^{\boldsymbol{\xi}}$ and the optimal solution follows by employing KKT conditions similar to Theorem 3, and Theorem 5. It easy to see that the solution ensures $\Sigma^{\boldsymbol{\xi}} = D - D^{\min}_{[0,\infty]} = D - \Sigma$. Substituting the latter in $\Pi^{\boldsymbol{\xi}}$ and then substituting both $\Sigma^{\boldsymbol{\xi}}$ and $\Pi^{\boldsymbol{\xi}}$ in (71) we obtain (68) and the result follows. ■

Equivalent expressions and special cases for scalar processes: (i) We note that our closed form expression (68) coincides with the closed-form solution obtained via [8, Corollary 1, Theorem 9] (see also [20, eq. (103)]) because the steady-state counterpart of the á posteriori error variance equation (25) implies the equality $\bar{\Sigma} = \Pi - \Sigma > 0$; (ii) Consider in Corollary 3 $c = 1$, $\sigma^2_{\mathbf{n}} = 0$. Then, using Example 1 we obtain from (58) that $\Pi = \sigma^2_{\mathbf{w}} > 0$, from (70) the steady state solution is $\Sigma = 0$ and from (69) $\bar{\Sigma} = \sigma^2_{\mathbf{w}} > 0$. By substituting these in (68) we recover the known result obtained for time-invariant or stationary fully observable Gauss-Markov processes, see e.g., [5, eq. (14)], [2, eq. (1.43)].

VI. Numerical Simulations

In this section, we provide two examples with numerical simulations to illustrate some of the major results of this paper.

Example 2. *(Optimal numerical solutions and comparison with [8]) Consider the time-invariant version of (1), (2) with*

$$A = \mathrm{diag}(1.2, 1.2, 1.2),\ C = \begin{bmatrix} 0.8147 & 0.9134 & 0.2785 \\ 0.9058 & 0.6324 & 0.5469 \\ 0.1270 & 0.0975 & 0.9575 \end{bmatrix},$$
$$\Sigma_{\mathbf{w}} = \begin{bmatrix} 0.8895 & 1.1744 & 0.2309 \\ 1.1744 & 1.8616 & 0.2953 \\ 0.2309 & 0.2953 & 0.0614 \end{bmatrix},\ \Sigma_{\mathbf{n}} = \mathrm{diag}(1,1,0). \tag{72}$$

Clearly, from Lemma 5, the pair (A, C) is detectable and the pair $(A, \Sigma^{\frac{1}{2}}_{\mathbf{w}})$ is stabilizable. Hence the filter $\boldsymbol{\xi}_t$ is asymptotically stable, with

$$\bar{\Sigma} = \begin{bmatrix} 2.6928 & -0.7211 & 0.1847 \\ -0.7211 & 4.0349 & 0.3254 \\ 0.1847 & 0.3254 & 0.0645 \end{bmatrix}, \tag{73}$$

and from (56) we obtain $\Pi \succ 0$ which further implies the steady-state solution of $\Sigma \succeq 0$ both given as follows

$$\Pi = \begin{bmatrix} 6.7910 & -5.0291 & 0.0798 \\ -5.0291 & 8.9742 & 0.3939 \\ 0.0798 & 0.3939 & 0.0714 \end{bmatrix}, \tag{74}$$

$$\Sigma = \begin{bmatrix} 4.0983 & -4.3080 & -0.1049 \\ -4.3080 & 4.9393 & 0.0684 \\ -0.1049 & 0.0684 & 0.0069 \end{bmatrix}. \tag{75}$$

We recall using [8, Corollary 1, Theorem 9], that the closed form solution of the sum-rate therein under the assumption of uniform rate-distortion allocation is given by

$$R^{G,KH}_{[0,\infty],\rm in}(D - \mathrm{trace}(\Sigma)) = \frac{p}{2}\log\left(\bar{a}^2 + \frac{|\bar{\Sigma}|^{\frac{1}{p}}p}{D - \mathrm{trace}(\Sigma)}\right), \tag{76}$$

where $\bar{a} \triangleq \mathrm{abs}(|A|)^{\frac{1}{p}}$, $\bar{\Sigma} = \Pi - \Sigma$, with $D > \mathrm{trace}(\Sigma)$. In Fig. 3, we give the optimal numerical solution obtained via Corollary 2, (2) using the CVX platform [41] and the reverse-waterfilling solution of Theorem 5 using Algorithm 2 (because the input data in (72) satisfy the strong structural properties of Proposition 5, (i)). We compare the optimal sum-rate with the closed-form solution of (76). We observe that the latter is in general highly suboptimal with respect to the optimal numerical solution with the maximum rate-loss (RL), which for this example is approximately 1.05 bits/vector source, to be observed at moderate to low rates. A good performance of (76) in the sense that it almost coincides with the exact optimal solution can be observed at very high rates. This means that Corollary 2 and Theorem 5 that allow non-uniform distortion allocation may achieve significant performance gains compared to (76) that only allows uniform distortion allocation.

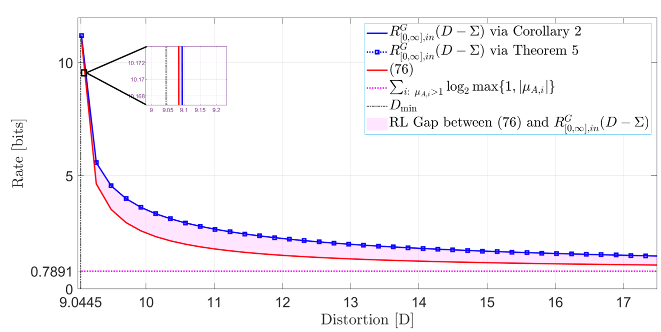

Fig. 3: Comparison of the optimal sum-rates obtained via Corollary 2 and Theorem 5 with the analytical expression of (76).

Example 3. *(Convergence to steady-state solution) Consider the time-invariant version of (49) with $(\alpha, c, \sigma^2_{\mathbf{w}}, \sigma^2_{\mathbf{n}}) =$*

$(1.1, 0.5, 1, 1)$, $\forall t$. Clearly, from Example 1, the pair (α, c) is detectable and the pair $(\alpha, (\sigma^2_{\mathbf{w}})^{\frac{1}{2}})$ is stabilizable. Hence the filter $\boldsymbol{\xi}_t = \mathbb{E}\{\mathbf{x}_t|\mathbf{z}^t\}$ is asymptotically stable and from (58) we obtain $\Pi = 3.1215 > 0$ whereas from (70) the non-negative solution is $\Sigma = 1.7532$. For a given distortion level $D = 2.7532 > \Sigma$ we obtain via (68) $R^G_{\text{in}}(D - \Sigma) = 0.6832$ (bits/source sample). Using Algorithm 1, we compute (51) (normalized over the time horizon $(n + 1)$) for sufficiently large time horizon, i.e., $n \longrightarrow 10^5$. In Fig. 4, we illustrate the asymptotic behavior of Algorithm 1 versus (vs) the steady-state solution (68) in a semi-logarithmic scale. The two lines are met really fast but do not coincide. In fact, depending on the precision error of Algorithm 1 (a reasonable error tolerance is $\epsilon = 10^{-9}$) one can also infer about the discrepancy of the two lines. We note that Algorithm 1 also gives $D^{\min}_{[0,n]} \approx \Sigma$ and $\lim_{t \longrightarrow 10^5} \sigma^2_{\mathbf{x}_{t|t-1}} \approx \Pi$. Moreover, the starting point of the plot obtained from Algorithm 1 depends on $\sigma^2_{\mathbf{x}_0}$.

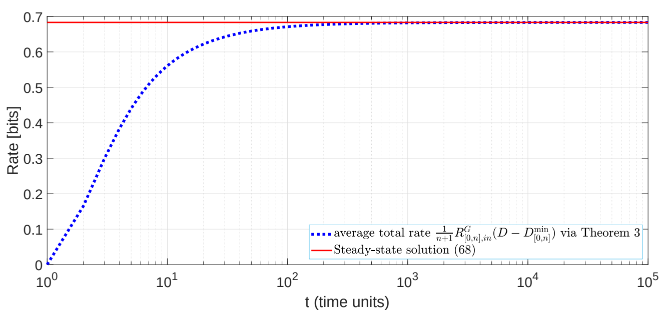

Fig. 4: Comparison of Theorem 3 vs the steady-state solution of Corollary 3 for time-invariant scalar-valued processes.

VII. Conclusions and Ongoing Research

In this paper we revisited the problem of characterizing and computing the indirect NRDF for partially observable multivariate Gauss-Markov processes with hard MSE distortion constraints. Our major results include a new formulation of the indirect NRDF, structural conditions that allow this formulation to be achieved, as well as the characterization and the corresponding optimal test channel realization for jointly Gaussian processes in both finite and infinite time horizon. Moreover, we obtained optimal numerical and closed form solutions for vector and scalar systems under either average total or pointwise MSE distortion constraints.

The results and observations of this paper can pave the way to many new problems concerning communication for (Gaussian) controlled systems see, for instance, [42]. One particular question that we do not directly address herein but can be answered from our results, is the relaxation of the Gaussian noise process that drives the state of the system model in (1), (2) to positive semidefinite covariance matrices. Another important question is the extension of Theorem 5 to time-varying processes which will require strong time-varying structural properties in the spirit of Proposition 5. Finally, the extension of this problem to controlled processes is also of major importance.

Appendix A
Proof of Theorem 4

First note that under the conditions of the theorem, we have the unique stabilizing solution $\lim_{t \longrightarrow \infty} \Pi_t = \Pi \succ 0$ and consequently $\lim_{t \longrightarrow \infty} \Sigma_t = \Sigma \succeq 0$. This in turn implies via (25) of Lemma 2 that $\lim_{t \longrightarrow \infty} \bar{\Sigma}_t = \bar{\Sigma}$. The specific steady state solution corresponds to an asymptotically stable filter. Then, the objective function in (60) is obtained as follows

$$\limsup_{n \longrightarrow \infty} \frac{1}{n+1} \sum_{t=0}^{n} \log \frac{|\Sigma^{\boldsymbol{\xi}}_{t|t-1}|}{|\Sigma^{\boldsymbol{\xi}}_{t|t}|}$$
$$\stackrel{(a)}{=} \limsup_{n \longrightarrow \infty} \frac{1}{n+1} \sum_{t=0}^{n} \log \frac{|A\Sigma^{\boldsymbol{\xi}}_{t-1|t-1}A^{\mathrm{T}} + \bar{\Sigma}_t|}{|\Sigma^{\boldsymbol{\xi}}_{t|t}|} \stackrel{(b)}{=} (60),$$

where (a) follows because $\Sigma^{\boldsymbol{\xi}}_{t|t-1} = A\Sigma^{\boldsymbol{\xi}}_{t-1|t-1}A^{\mathrm{T}} + \bar{\Sigma}_t$; (b) follows because we restrict the numerator and denominator in (a) to be have a time invariant value (because we impose the optimal minimizer to be time invariant and the corresponding output distribution to be time-invariant with a unique invariant distribution). Note that Π^{ξ} is given by (61) and $\{\bar{\Sigma}_n : n \in \mathbb{N}_0\}$ is a convergent sequence (by the conditions of the theorem) and its steady-state (time invariant) solution is $\bar{\Sigma} = \lim_{n \longrightarrow \infty} \Sigma_n$. The constraint set in (60) is obtained because via Remark 3 we ensure a finite solution to the optimization problem if we impose the strict LMI $0 \prec \Sigma^{\boldsymbol{\xi}} \preceq \Pi^{\boldsymbol{\xi}}$ which implies that $\Sigma^{\boldsymbol{\xi}} \succ 0$ and $\Pi^{\boldsymbol{\xi}} \succ 0$. From the conditions of the theorem, we have a convergent sequence $\{\Sigma_n : n \in \mathbb{N}_0\}$, i.e., $\lim_{n \longrightarrow \infty} \Sigma_n = \Sigma$ which further means that $\{\text{trace}(\Sigma_n) : n \in \mathbb{N}_0\}$ is also convergent. This in turn implies that $\frac{1}{n+1}\sum_{t=0}^{n} \text{trace}(\Sigma_t) = \text{trace}(\Sigma)$ as $n \longrightarrow \infty$ which is precisely (62). This completes the characterization of (60). The optimal time-invariant test channel realization (63) follows easily from the conditions of the theorem. This completes the derivation.

Acknowledgement

The authors wish to thank the Associate Editor and the anonymous reviewers for their valuable comments and suggestions.

References

[1] A. K. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problems Inf. Transmiss.*, vol. 9, no. 3, pp. 184–191, July-Sept. 1972.

[2] ——, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problems Inf. Transmiss.*, vol. 10, no. 2, pp. 93–109, Apr.-June 1972, translation from Problemy Peredachi Informatsii, vol. 10, no. 2, pp. 5-–25, April-June 1974.

[3] 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.

[4] P. A. Stavrou, J. Østergaard, and C. D. Charalambous, "Zero-delay rate distortion via filtering for vector-valued Gaussian sources," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 841–856, Oct 2018.

[5] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, pp. 1549 – 1561, 2004.

[6] C. D. Charalambous, P. A. Stavrou, and N. U. Ahmed, "Nonanticipative rate distortion function and relations to filtering theory," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 937–952, 2014.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[8] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4525–4540, Nov 2019.

[9] P. A. Stavrou, T. Tanaka, and S. Tatikonda, "The time-invariant multidimensional Gaussian sequential rate-distortion problem revisited," *IEEE Transactions on Automatic Control*, vol. 65, no. 5, pp. 2245–2249, 2020.

[10] 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.

[11] P. A. Stavrou, T. Charalambous, C. D. Charalambous, and S. Loyka, "Optimal estimation via nonanticipative rate distortion function and applications to time-varying Gauss-Markov processes," *SIAM J. on Control Optim.*, vol. 56, no. 5, pp. 3731–3765, 2018.

[12] A. J. Fuglsig and J. Østergaard, "Zero-delay multiple descriptions of stationary scalar Gauss-Markov sources," *Entropy*, vol. 21, no. 12, 2019.

[13] C. D. Charalambous, C. Kourtellaris, T. Charalambous, and J. H. van Schuppen, "Generalizations of nonanticipative rate distortion function to multivariate nonstationary Gaussian autoregressive processes," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 8190–8195.

[14] C. D. Charalambous, T. Charalambous, C. Kourtellaris, and J. H. van Schuppen, "Complete characterization of Gorbunov and Pinsker nonanticipatory epsilon entropy of multivariate Gaussian sources: Structural properties," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 1440–1464, 2022.

[15] T. Tanaka, P. M. Esfahani, and S. K. Mitter, "LQG control with minimum directed information: Semidefinite programming approach," *IEEE Trans. Autom. Control*, vol. 63, no. 1, pp. 37–52, Jan 2018.

[16] T. Tanaka, "Zero-delay rate-distortion optimization for partially observable Gauss-Markov processes," in *Proc. IEEE Conf. Decision Control*, Dec 2015, pp. 5725–5730.

[17] P. A. Stavrou and M. Skoglund, "Asymptotic reverse waterfilling algorithm of nrdf for certain classes of vector Gauss–Markov processes," *IEEE Transactions on Automatic Control*, vol. 67, no. 6, pp. 3196–3203, 2022.

[18] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.

[19] 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.

[20] V. Kostina and B. Hassibi, "Fundamental limits of distributed tracking," in *Proc. IEEE Int. Symp. Inf. Theory*, 2020, pp. 2438–2443.

[21] P. A. Stavrou, J. Østergaard, and M. Skoglund, "Bounds on the sum-rate of MIMO causal source coding systems with memory under spatio-temporal distortion constraints," *Entropy*, vol. 22, no. 8, 2020.

[22] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl.*, Nov. 27-30 1990, pp. 303–305.

[23] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, Inc., New York, 1997.

[24] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Info. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[25] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, pp. 406–411, 1970.

[26] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[27] S. Ihara, *Information theory - for Continuous Systems*. World Scientific, 1993.

[28] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[29] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, New Jersey: Prentice Hall, 200.

[30] D. Simon, *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[31] J. V. Schuppen, *Control and Systems Theory of Discrete-Time Stochastic Systems*, ser. Communications and Control Engineering. Springer, 2021.

[32] C. D. Charalambous, C. Kourtellaris, and S. Louka, "New formulas of feedback capacity for AGN channels with memory: A time-domain sufficient statistic approach," 2020.

[33] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. Springer-Verlag, 1997.

[34] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific, 2005.

[35] P. A. Stavrou and M. Skoglund, "Indirect nrdf for partially observable gauss-markov processes with mse distortion: Complete characterizations and optimal solutions," *arxiv.org*, 2021, https://arxiv.org/abs/1912.07640v4.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[37] P. A. Stavrou, M. Skoglund, and T. Tanaka, "Sequential source coding for stochastic systems subject to finite rate constraints," *IEEE Trans. Autom. Control*, vol. 67, no. 8, pp. 3822–3835, 2022.

[38] P. A. Stavrou, T. Charalambous, and C. D. Charalambous, "Finite-time nonanticipative rate distortion function for time-varying scalar-valued Gauss-Markov sources," *IEEE Control Syst. Lett.*, vol. 2, no. 1, pp. 175–180, Jan 2018.

[39] F. Zhang, *Matrix Theory: Basic results and techniques*, 2nd ed. Springer-Verlag New York, 2011.

[40] J. Alman and V. V. Williams, *A Refined Laser Method and Faster Matrix Multiplication*, ser. Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA), 2021, pp. 522–539.

[41] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[42] C. D. Charalambous and T. Charalambous, "State-dependent generalizations of nonanticipatory epsilon entropy of partially observable processes," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 7415–7422.

PLACE PHOTO HERE

Photios A. Stavrou (S'10-M'16-SM'22) received his D. Eng in 2008 from the Department of Electrical and Computer Engineering (ECE) of the Faculty of Engineering at Aristotle University of Thessaloniki, Greece and his Ph.D degree in 2016 from the Department of ECE of the Faculty of Engineering at University of Cyprus, Cyprus. From November 2016 to September 2022, he held post-doctoral positions at Aalborg University in Denmark, at KTH Royal Institute of Technology in Sweden, and EURECOM in France. As of September 2022, Dr Stavrou is an Assistant Professor in the Communication Systems Department at EURECOM, Sophia-Antipolis, France. His research interests span information and communication theories, networked control systems, goal-oriented communication, feedback and privacy in communication, optimization, and game theory.

PLACE PHOTO HERE

Mikael Skoglund (S'93-M'97-SM'04-F'19) received the Ph.D. in 1997 from Chalmers University of Technology, Sweden. In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed to the Chair of Communication Theory in 2003. At KTH he heads the Division of Information Science and Engineering and the Department of Intelligent Systems. Dr. Skoglund has worked on problems in source-channel coding, coding and transmission for wireless communications, Shannon theory, information and control, and statistical signal processing. He has authored and co-authored some 160 journals and 380 conference papers. Dr. Skoglund is a Fellow of the IEEE. During 2003–08 he was an associate editor for the IEEE Transactions on Communications and during 2008–12 he was on the editorial board for the IEEE Transactions on Information Theory. He has served on numerous technical program committees for IEEE-sponsored conferences, he was general co-chair for IEEE ITW 2019, and he will serve as TPC co-chair for IEEE ISIT 2022.