On Information, Estimation and Lookahead

Kartik Venkat  
EE, Stanford University  
kvenkat@stanford.edu

Tsachy Weissman  
EE, Stanford University  
tsachy@stanford.edu

Yair Carmon  
EE, Technion-IIT  
yairc@tx.technion.ac.il

Shlomo Shamai  
EE, Technion-IIT  
sshlomo@ee.technion.ac.il

Abstract—We consider mean squared estimation with lookahead of a continuous-time signal corrupted by additive white Gaussian noise. We investigate the connections between lookahead in estimation, and information under this model. We show that the mutual information rate function, i.e., the mutual information rate as function of the signal-to-noise ratio (SNR) does not in general determine the mean squared error with fixed finite lookahead, in contrast to the special cases with 0 and infinite lookahead (filtering and smoothing errors), respectively, which were previously established in the literature. Independently, we define the notion of information utility of finite lookahead, and characterize properties of the same. We also establish a new expectation identity under a generalized observation model (where the Gaussian channel has an SNR jump at \( t = 0 \)), capturing the tradeoff between lookahead and SNR.

I. INTRODUCTION

Mean squared estimation of a signal in the presence of Gaussian noise has been a topic of considerable importance to the communities of information theory and estimation theory. There have further been discoveries tying the two fields together, through identities between fundamental informational quantities and the squared estimation loss.

Consider the continuous-time Gaussian channel. In [1], Duncan established the equivalence between the input-output mutual information and the integral of half the mean squared error in estimating the signal based on the observed process. In [2], Guo et al. present what is now known as the I-MMSE relationship, which equates the derivative of the mutual information to half the average non-causal squared error. These results were extended to mismatched estimation for the scalar Gaussian channel by Verdú in [3], and for the continuous Gaussian channel by Weissman in [4]. In the latter, it was shown that when the decoder assumes an incorrect law for the input process, the difference in mismatched causal estimation loss is simply given by the relative entropy between the true and incorrect output distributions. Recently, in [5], pointwise analogues of these and related results were developed, characterizing the original relationships as identities involving expectations of random objects with information-estimation significance of their own.

Let \( \mathbf{X} = \{ X_t : -\infty < t < \infty \} \) denote a continuous-time signal, which is the channel input process. The output process \( \mathbf{Y} \) of the continuous time Gaussian channel with input \( \mathbf{X} \) is given by

\[
dY_t = \sqrt{\text{snr}} X_t dt + dW_t,
\]

for all \( t \), where \( \text{snr} > 0 \) is the signal-to-noise ratio (SNR) and \( W \) is a standard Brownian Motion [6] independent of \( \mathbf{X} \). For simplicity let us assume the input process to be stationary. Let \( \gamma(\text{snr}) \) denote the mutual information rate. Let \( \text{mmse}(\text{snr}) \) denote the smoothing squared error which, in our setting of stationarity, can equivalently be defined as

\[
\text{mmse}(\text{snr}) = \mathbb{E}[(X_0 - \mathbb{E}[X_0|Y_{<\infty}])^2].
\]

Similarily, the causal squared error is given by

\[
\text{cmmse}(\text{snr}) = \mathbb{E}[(X_0 - \mathbb{E}[X_0|Y_{\infty}])^2].
\]

From [1] and [2], we know that, under some benign regularity conditions, the above quantities are related according to

\[
\frac{2\gamma(\text{snr})}{\text{snr}} = \frac{\text{cmmse}(\text{snr})}{\text{mmse}(\text{snr})} = \frac{1}{\text{snr}} \int_0^{\text{snr}} \text{mmse}(\gamma) d\gamma
\]

for all signal-to-noise ratios, \( \text{snr} > 0 \), regardless of the distribution of \( \mathbf{X} \). Thus, the mutual information, the causal error and the non-causal error are linked together in a distribution independent fashion for the Gaussian channel under mean squared loss. Having understood how the time averaged causal and non-causal errors behave, and in particular the relation (4) implying that they are completely characterized by the mutual information rate function - the current work seeks to address the question “what happens for lookahead between 0 and \( \infty \)?”.

The mean squared error with finite lookahead \( d \geq 0 \) is defined as

\[
\text{Immse}(d, \text{snr}) = \mathbb{E}[(X_0 - \mathbb{E}[X_0|Y_{-\infty}^d])^2],
\]
where it is instructive to note that the special cases $d = 0$ and $d = \infty$ in (5) yield the filtering (3) and smoothing (2) errors respectively. Let us be slightly adventurous, and rewrite (4) as
\[ \frac{2I(\text{snr})}{\text{snr}} = \mathbb{E}[\text{lmmse}(0, \Gamma_0)] = \mathbb{E}[\text{lmmse}(\infty, \Gamma_\infty)], \]
where the random variables \(\Gamma_0, \Gamma_\infty\) are distributed according to \(\Gamma_0 = \text{snr}\) a.s., and \(\Gamma_\infty \sim U[0, \text{snr}]\), where \(U[a, b]\) denotes the uniform distribution on \([a, b]\). This characterization of the mutual information rate may lead one to conjecture the existence of a random variable \(\Gamma_d\), whose distribution depends only on \(d, \text{snr}\), and basic features of the process (as its bandwidth), and satisfies the identity
\[ \frac{2I(\text{snr})}{\text{snr}} = \mathbb{E}[\text{lmmse}(d, \Gamma_d)], \]
regardless of the distribution of \(X\). I.e., there exists a family of distributions \(\Gamma_d\) ranging from one extreme of a point mass at \(\text{snr}\) for \(d = 0\), to the uniform distribution over \([0, \text{snr}]\) when \(d = \infty\). One of the corollaries of our results in this work is that such a relation does not hold in general. In fact, in Section IV we show that in general, the mutual information rate of a process is not sufficient to characterize the estimation error with finite lookahead. Motivated by this question, however, in Section III we establish an identity which relates the filtering error to a double integral over lookahead and SNR of the estimation error, for the Gaussian channel with an SNR jump at \(t = 0\).

The rest of the paper is organized as follows. We begin in Section II, by introducing the information utility of finite lookahead, and studying properties of the same. For example, we show that the information utility of infinitesimal lookahead is proportional to the filtering error of the underlying process. In Section III, we introduce a generalized observation model for a stationary continuous-time signal where the channel has a non-stationary SNR jump at \(t = 0\). For this model, we establish an identity relating the squared error with finite lookahead and the causal filtering error. Finally, in Section IV, we show by means of an example that a distribution-independent identity of the form in (7) cannot hold in general. We summarize our findings and outline future directions in Section V.

II. INFORMATION UTILITY OF LOOKAHEAD

A. Definition and relation to finite lookahead MMSE

Consider a stationary process \(X_t\) observed through the continuous-time Gaussian channel (1) at SNR level \(\text{snr}\). We now address the question, how much information does lookahead provide in general, and whether this quantity has any interesting connections with classical estimation theoretic objects.

For \(t > 0\), let us define the Information Utility \(U(\cdot)\) as a function of lookahead \(t\), to be
\[ U(t) = I(X_0; Y^t_0 | Y^-\infty_0). \]

When the input process is Gaussian,
\[ U(t) = h(X_0 | Y^-\infty_0) - h(X_0 | Y^t_\infty) \]
\[ = \frac{1}{2} \log \left( \frac{2\pi e \text{Var}(X_0 | Y^-\infty_0)}{2 \pi e \text{Var}(X_0 | Y^t_\infty)} \right) \]
\[ = \frac{1}{2} \log \left( \frac{\text{cmmse}(\text{snr})}{\text{lmmse}(t, \text{snr})} \right). \]

Rearranging the terms, we have
\[ \text{lmmse}(t, \text{snr}) = \text{cmmse}(\text{snr}) e^{-2U(t)}. \]

Furthermore, when the input is non-Gaussian but \(h(X_0 | Y^t_\infty)\) is well-defined for every \(t \geq 0\),
\[ h(X_0 | Y^-\infty_0) \leq \mathbb{E}_{y^-\infty} \left[ \frac{1}{2} \log \left( \frac{2\pi e \text{Var}(X_0 | Y^t_\infty = y^t_\infty)}{2 \pi e \text{Var}(X_0 | Y^-\infty_0)} \right) \right] \]
\[ \leq \frac{1}{2} \log \left( \frac{2\pi e \text{lmmse}(t, \text{snr})}{2 \pi e \text{lmmse}(\text{snr})} \right). \]

The first inequality is due to the fact that the Gaussian distribution has maximum entropy under a variance constraint, and the second inequality stems from Jensen’s inequality. Rearranging the terms once more, we have that for every stationary input process,
\[ \text{lmmse}(t, \text{snr}) \geq N(X_0 | Y^-\infty_0) e^{-2U(t)}, \]
where \(N(Z) = \frac{1}{2\pi e^2} e^{2h(Z)}\) is the entropy power functional.

B. Relation to MMSE with initial conditions

Consider an infinitesimal increment \(dt\) in \(U(t)\),
\[ U(t + dt) - U(t) = I(X_0; Y^t_{t+dt} | Y^-\infty_0) \]
\[ = I(X^t_{t+dt}; Y^t_{t+dt} | Y^-\infty_0) - I(X^t_{t+dt}; Y^t_{t+dt} | Y^t_{-\infty}, X_0) \]
\[ = I(X^t_{t+dt}; Y^t_{t+dt} | Y^-\infty_0) - \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
\[ = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t | Y^-\infty_0) + o(dt), \]
and
\[ I(X_t \mid Y_t) = \frac{1}{2} dt \cdot \text{snr} \cdot \text{Var}(X_t \mid Y_t) + o(dt). \] (15)

The expression for the time derivative of the information utility function is therefore,
\[ U'(t) = \frac{\text{snr}}{2} \left( \text{Var}(X_t \mid Y_t) - \text{Var}(X_t \mid Y_{-\infty}, X_0) \right). \] (16)

Since the input is assumed stationary,
\[ \text{Var}(X_t \mid Y_{-\infty}) = \text{Var}(X_0 \mid Y_{-\infty}) = \text{cmmse}(\text{snr}). \] (17)

We notice that \( \text{Var}(X_t \mid Y_{-\infty}, X_0) \) is the causal MMSE in estimating \( X_t \) when \( X_0 \) is known. The value of \( U'(t) \) is therefore intimately connected to the effect of initial conditions on the causal MMSE estimation of the process. In particular,
\[ U'(0) = \frac{\text{snr}}{2} \text{cmmse}(\text{snr}), \] (18)
meaning that the added value in information of an infinitesimal lookahead is proportional to the causal MMSE.

Notice that \( U(0) = 0 \), we may integrate (16) to obtain,
\[ U(t) = \text{snr} \left[ \text{cmmse}(\text{snr}) - \int_0^t ds \text{Var}(X_s \mid Y_{-\infty}, X_0) \right]. \] (19)

Specializing to stationary Gaussian input processes, and joining equations (10) and (19), we obtain
\[ \text{mmse}(t, \text{snr}) = \text{cmmse}(\text{snr}) \times \exp \left[ -\text{snr} \left( \text{cmmse}(\text{snr}) - \int_0^t ds \text{Var}(X_s \mid Y_{-\infty}, X_0) \right) \right] \] (20)

Thus, we are able to obtain a characterization of the information utility for stationary processes, in terms of the filtering error and the causal MMSE with known initial condition. In addition, we are also able to show via (18), that the benefit of infinitesimal lookahead, characterized by the derivative at 0 of the information utility function, \( U'(0) \) is proportional to the filtering error.

III. GENERALIZED OBSERVATION MODEL

In this section, we present a new observation model to understand the behavior of estimation error with lookahead. Consider a stationary continuous-time stochastic process \( X_t \). The observation model is still additive Gaussian noise, where the SNR level of the channel has a jump at \( t = 0 \). Letting \( Y_t \) denote the channel output, we describe the channel as given below.
\[ dY_t = \begin{cases} \sqrt{\text{snr}} X_t dt + dW_t & t \leq 0 \\ \sqrt{\gamma} X_t dt + dW_t & t > 0 \end{cases} \] (21)

where, as usual, \( W \) is a standard Brownian motion independent of \( X \). Note that for \( \gamma \neq \text{snr} \), the \((X_t, Y_t)\) process is not jointly stationary. Letting \( d, l \geq 0 \), we define the finite lookahead estimation error at time \( d \) with lookahead \( l \) as
\[ f(\text{snr}, \gamma, d, l) = \text{Var}(X_d \mid Y_{-\infty}^{l+d}). \] (22)

We call this a generalized observation model, as for \( \gamma = \text{snr} \), we recover the usual time-invariant Gaussian channel. For instance, we note that the error \( f \) reduces to the filtering, smoothing or finite lookahead errors, depending on the parameters \( \gamma, l \) and \( d \) as:
\[ \text{cmmse}(\text{snr}) = f(\text{snr}, \text{snr}, d, 0) \] (23)
\[ \text{mmse}(\text{snr}) = f(\text{snr}, \text{snr}, d, \infty) \] (24)
\[ \text{immse}(l, \text{snr}) = f(\text{snr}, \text{snr}, t, l) \] (25)

In the following we relate the estimation error with finite lookahead for the observation model described in (21), with the original filtering error.

**Theorem 1:** Let \( X_t \) be any finite variance continuous time stationary process which is corrupted by the Gaussian channel in (21). Let \( f \) be as defined in (22). For fixed \( \text{snr} > 0 \) and \( T > 0 \) let \( \Gamma \sim U[0, \text{snr}] \) and \( L \sim U[0, T] \) be independent random variables. Then
\[ \text{cmmse}(\text{snr}) = \mathbb{E}_{\Gamma, L}[f(\text{snr}, \Gamma, T - L, L)] \] (26)

**Proof:** Before proving this result, we take a detour to consider a stationary continuous-time stochastic process \( X_t \) which is governed according to law \( P_{X,T} \) and is observed in the window \( t \in [0, T] \) through the continuous-time Gaussian channel (1) at SNR level snr. We define the operators which evaluate the filtering and smoothing estimation errors for this process model, as follows:
\[ \text{cmmse}(P_{X,T}, \text{snr}) = \int_0^T \mathbb{E}[(X_t - \mathbb{E}[X_t \mid Y_T]^2)] dt \] (27)
\[ \text{mmse}(P_{X,T}, \text{snr}) = \int_0^T \mathbb{E}[(X_t - \mathbb{E}[X_t \mid Y_T]^2)] dt \] (28)

Further, [2] gives us the celebrated relationship between the causal and non-causal estimation errors:
\[ \text{cmmse}(P_{X,T}, \text{snr}) = \frac{1}{\text{snr}} \int_0^{\text{snr}} \text{mmse}(P_{X,T}, \gamma) d\gamma. \] (29)
Lookahead and Mutual Information

New Observation Model

Information Utility of Lookahead

Theorem

In other words, we get

\[
c_{\text{mmse}}(\text{snr}) = \frac{1}{T} \mathbb{E}_s \left[ c_{\text{mmse}}(P_{X_t^T} | Y_0^{\infty}, \text{snr}) \right]
\]

\[
m(\text{snr}) = \frac{1}{T} \mathbb{E}_s \left[ \mathbb{M}(P_{X_t^T} | Y_0^{\infty}, \text{snr}) \right] .
\]

where \( \mathbb{E}_s \) denotes expectation over \( Y_0^{\infty} \). And, now employing the relation between the two quantities above, we get

\[
\frac{1}{T} \mathbb{E}_s \left[ c_{\text{mmse}}(P_{X_t^T} | Y_0^{\infty}, \text{snr}) \right] = \\
\frac{1}{T} \mathbb{E}_s \left[ \frac{1}{\text{snr}} \int_0^{\text{snr}} \mathbb{M}(P_{X_t^T} | Y_0^{\infty}, \gamma, \text{snr}) d\gamma \right] .
\]

Using the definition of \( f \) and inserting above, we get the following intergral equation which holds for every stationary process \( X_t \) observed through the continuous Gaussian channel as described in (21)

\[
f(\text{snr}, \text{snr}, x, 0) = \\
\frac{1}{T} \cdot \frac{1}{\text{snr}} \int_0^{\text{snr}} \int_0^T f(\text{snr}, \gamma, t, T - t) dt d\gamma,
\]

for all \( T \) and \( x \). Note that the LHS is nothing but the causal squared error at SNR level \( \text{snr} \). Note in particular that for independent random variables \( \Gamma \sim U[0, \text{snr}] \) and \( L \sim U[0, T] \), (33) can be expressed as

\[
c_{\text{mmse}}(\text{snr}) = \mathbb{E}_{\Gamma, L} [f(\text{snr}, \Gamma, T - L, L)] \quad \forall T .
\]

Fig. 1: Region in the Lookahead-SNR plane over which the finite lookahead MMSE given by (22) is averaged over in Theorem 1 to yield the filtering error at level \( \text{snr} \).

Fig. 2: Characteristic behavior of the minimum mean squared error with lookahead for any process.

It is interesting to note that the double integral over lookahead and SNR are conserved in such a manner by the filtering error, for any arbitrary underlying stationary process. Note that one way of interpreting this result, is to take the average of the finite lookahead mmse (under the given observation model), over a rectangular region in the Lookahead vs. Signal-to-Noise Ratio plane, as depicted in Fig. 1. Theorem 1 tells us that for all underlying processes, this quantity is always the filtering error at level \( \text{snr} \). Thus, we find the classical estimation theoretic quantity described by the causal error to emerge as a bridge between the effects of varying lookahead and the signal to noise ratio.

Given Theorem 1, it is natural to investigate whether the relation (26) can be inverted to determine the function \( f \), from the causal mmse. If that were the case, then, in particular, the filtering error would completely determine the estimation loss with finite lookahead. To pose an equivalent question, let us recall Duncan’s result in [1] establishing the equivalence of the mutual information rate and the filtering error. On the other hand, by [2], [7] the mutual information rate function \( I(\cdot) \), determines the non-causal mmse. Might the same mutual information rate function \( I(\text{snr}) \) completely determine the estimation loss for a fixed finite lookahead \( d \) as well? This question is addressed in the following section.

IV. CAN I(\cdot) RECOVER FINITE-LOOKAHEAD MMSE?

In Fig. 2, we present a characteristic plot of the mmse with lookahead for an arbitrary continuous time stationary process, corrupted by the Gaussian channel (1) at SNR level \( \text{snr} \). In particular, we note three important features of the process, namely (i) the asymptote at \( d = \infty \) is mmse(\text{snr}), (ii) lmmse(0, \text{snr}) = c_{\text{mmse}}(\text{snr})
and (iii) the asymptote at \( d = -\infty \) is the variance of the stationary process, \( \text{Var}(X_0) \). Further, the curve is non-decreasing for all \( d \).

From [1] and [2], we know that the mutual information, the causal, and the non-causal \( \text{mmse} \) determine each other according to

\[
2I(\text{snr}) \frac{\text{snr}}{\text{snr}} = \operatorname{cmmsn}(\text{snr}) = \frac{1}{\text{snr}} \int_0^{\text{snr}} \text{mmse}(\gamma) d\gamma. \tag{35}
\]

In other words, the mutual information rate function is sufficient to characterize the behavior of the causal and smoothing errors as functions of \( \text{snr} \). In particular, (35) also determines the three important features of the \( \text{mmse}(\cdot, \text{snr}) \) curve discussed above, i.e. the value \(^1\) at \( d = 0 \), and the asymptotes at \( d = \pm \infty \). It is natural to ask whether it can actually characterize the entire curve \( \text{mmse}(\cdot, \text{snr}) \). The following theorem implies that the answer is negative.

**Theorem 2:** For any finite \( d < 0 \) there exist stationary continuous-time processes which have the same mutual information rate \( I(\text{snr}) \) for all \( \text{snr} \), but have different minimum mean squared errors with (negative) lookahead \( d \).

One of the corollaries of Theorem 2, which we demonstrate by means of an example, is that the mutual information rate as a function of \( \text{snr} \) does not in general determine \( \text{mmse}(d, \text{snr}) \) for any finite non-zero \( d \).

Thus, if the triangular relationship between mutual information, the filtering, and the smoothing errors is to be extended to accommodate finite lookahead, one will have to resort to distributional properties that go beyond mutual information. To see why this must be true in general, we note that the functional \( \text{mmse}(\cdot, \text{snr}) \) may depend on the time-dynamical features of the process, to which the mutual information rate function is invariant. For example, we note the following.

**Observation 3:** Let \( X_t \) be a stationary continuous time process. Define \( X_t^{(a)} = X_{at} \) for a fixed constant \( a > 0 \). Let \( Y_t \) and \( Y_t^{(a)} \), denote respectively, the outputs of the Gaussian channel (1) with \( X_t \) and \( X_t^{(a)} \) as inputs. Then, for all \( d \) and \( \text{snr} > 0 \),

\[
\text{mmse}_{X_t}^{(a)}(d, \text{snr}) = \text{mmse}_X(ad, \text{snr}/a), \tag{36}
\]

where the subscript makes the process under consideration explicit. Note that for the special cases when \( d \in \{0, \pm \infty\} \), the error of the scaled process results in a scaling of just the \( \text{SNR} \) parameter, i.e.

\[
\text{mmse}_{X_t}^{(0)}(\text{snr}) = \text{mmse}_X(\text{snr}/a), \tag{37}
\]

\(^1\)Note that \( \text{Var}(X_0) = \text{cmmsn}(0) \), which \( I(\cdot) \) determines.

For all other values of \( d \), we see that the error depends on the error at a scaled lookahead, in addition to a scaled \( \text{SNR} \) level. This indicates, the general dependence of the \( \text{mmse} \) with finite, non-zero lookahead on the time-dynamics of the underlying process.

One of the consequences of Duncan’s result in [1] is that the causal and anti-causal\(^2\) errors as functions of \( \text{snr} \) are the same (due to the mutual information acting as a bridge, which is invariant to the direction of time). Let \( X_t \) be a continuous-time stationary stochastic process and \( Y_t \) be the output process of the Gaussian channel (1) with \( X_t \) as input. We can write,

\[
2I(\gamma) = \text{cmmsn}(\gamma) = \text{acmmse}(\gamma), \tag{39}
\]
or, writing the rightside equality explicitly,

\[
\text{Var}(X_0|Y_0^{-\infty}) = \text{Var}(X_0|Y_0^\infty). \tag{40}
\]

It is now natural to wonder whether (40) carries over to the presence of lookahead, which would have to be the case if the associated conditional variances are to be determined by the mutual information function, which is invariant to the direction of the flow of time. In the following we present an explicit construction of a process for which

\[
\text{Var}(X_0|Y_0^{-d}) \neq \text{Var}(X_0|Y_0^d) \tag{41}
\]

for some values of \( d \). Note that the left and right sides of (41) are the \( \text{mmse} \)’s with lookahead \( d \) associated with the original process, and its time reversed version, respectively. Thus, mutual information alone does not characterize these objects.

**A. Construction of a continuous-time process**

In this section, we construct a stationary continuous time process from a stationary discrete time process. This process will be the input to the continuous time Gaussian channel in (1).

Let \( \tilde{X} = \{X_i\}_{i=-\infty}^{+\infty} \) be a discrete time stationary process following a certain law \( P_{\tilde{X}} \). Let us define a piecewise constant continuous-time process \( X_t \) such that

\[
X_t = \tilde{X}_{i}, \quad t \in (i-1, i] \tag{42}
\]

We now apply a random shift \( \Delta \sim U[0,1] \) to the \( \{X_t\} \) process to transform the non-stationary continuous time process into a stationary one. Let us denote this stationary process by \( X \). The process \( X \) is observed

\(^2\)the anti-causal error denoted as \( \text{acmmse}(\text{snr}) \) denotes the filtering error for the time-reversed input process.
through the Gaussian channel in (1) at snr = 1, with Y denoting the channel output process. This procedure is illustrated in Fig. 3

1) Time Reversed Process: Consider the discrete-time process \( \tilde{X}(R) \), denoting the time reversed version of the stationary process \( \tilde{X} \). The reversed process will, in general, not have the same law as the forward process (though it will of course inherit its stationarity from that of \( \tilde{X} \)). Let us construct an equivalent continuous time process corresponding to \( \tilde{X}(R) \), using the procedure described above for the process \( \tilde{X} \), and label the resulting stationary continuous-time process by \( X(R) \). In the following example, we compare the minimum mean square errors with finite lookahead for the processes \( X \) and \( X(R) \), for a certain underlying discrete-time Markov process.

B. Proof of Theorem 2, Examples and Discussions

Define a stationary discrete time Markov chain \( \tilde{X} \), on a finite alphabet \( \mathcal{X} \), characterized by the transition probability matrix \( P \).

The intuition behind the choice of alphabet and transition probabilities, is that we would like the Discrete Time Markov Chain (DTMC) to be highly predictable (under mean square error criterion) in the forward direction compared to the time reversed version of the chain. Note that the time reversed Markov chain will have the same stationary distribution, but different transition probabilities. We transform this discrete time Markov chain to a stationary continuous time process by the transformation described above. Because of the difference in the state predictability of the forward and reverse time processes in the DTMC, the continuous time process thus created will have predictability that behaves differently for the forward and the time reversed processes and, in turn, the MSE with finite lookahead will also depend on whether the process or its time-reversed version is considered.

1) Infinite SNR scenario: We now concentrate on the case when the signal-to-noise ratio of the channel is infinite. The input to the Gaussian channel is the continuous time process \( X \), which is constructed based on \( \tilde{X} \) as described above. Let us consider negative lookahead \( d = -1 \). For infinite SNR, the filter will know what the underlying process is exactly, so that

\[
\text{lmmse}(-1, \infty) = \text{Var}(X_0|Y_{-\infty}^{-1})
\]

where (45) follows from the Markovity of \( \tilde{X} \). Note that the quantity in (45) is the prediction variance of the DTMC \( \tilde{X} \). Let \( \nu : \mathcal{X} \rightarrow \mathbb{R} \) be any probability measure on the finite alphabet \( \mathcal{X} \) and \( V[\nu] \) be the variance of this distribution. Let \( \mu, P(\tilde{X}_{t+1}|\tilde{X}_t) \) denote respectively, the stationary distribution and the probability transition matrix of the Markov chain \( \tilde{X} \). Then the prediction variance is given by

\[
\text{Var}(\tilde{X}_0|\tilde{X}_{-1}) = \sum_{x \in \mathcal{X}} \mu(x) V[P(\cdot|x)].
\]

In the infinite SNR setting, it is straightforward to see that

\[
\text{lmmse}(d, \infty) = 0, \quad \forall d \geq 0.
\]

Since the process \( X \) is constructed by a uniformly random shift according to a \( U[0,1] \) law, for each \(-1 \leq d \leq 0 \), we have

\[
\text{lmmse}(d, \infty) = |d| \text{Var}(X_0|X_{-1}), \quad -1 \leq d \leq 0.
\]

For fixed \( d \in [-1,0] \), with probability \( 1 - |d| \), the process \( Y_{\infty}^d \) will sample \( X_0 \) in which case the resulting mean squared error is 0 at infinite SNR. Alternately, with probability \( |d| \) the error will be given by the prediction variance, i.e \( \text{Var}(X_0|X_{-1}) \), which gives rise to (48). A similar analysis can be performed for the time-reversed DTMC \( \tilde{X}(R) \). Having found analytic expressions for the mmse with lookahead for the infinite SNR case, we show a characteristic plot of the same in Fig. 4. Note the difference in the curves for negative lookahead arises simply due to a difference in prediction variance for the forward and time-reversed DTMC’s \( X \) and \( \tilde{X}(R) \). Since the mmse with lookahead are different for \( d < 0 \) at infinite SNR, they would also be different at a large enough, finite signal-to-noise ratio. This completes the proof of Theorem 2 as stated for \( d < 0 \).
To further argue the existence of processes where $\text{lmmse}(d, \text{snr})$ are different, also for positive $d$, we provide the following explicit construction of the underlying processes. We then provide plots of the mmse with lookahead for this process, based on Markov Chain Monte Carlo simulations. The associated plots make it clear that $\text{lmmse}(d, \cdot)$ are distinct for both processes when $d$ is finite and non-zero.

2) Simulations: Let $\tilde{X}$ be a Discrete Time Markov Chain with the following specifications: The alphabet is $\mathcal{X} = \{5, 0, -5\}$. The probability transition matrix $\mathcal{P}$ is given by:

$$
\mathcal{P} = \begin{bmatrix}
0.6 & 0.4 & 0 \\
0 & 0.2 & 0.8 \\
0.875 & 0 & 0.125 \\
\end{bmatrix}
$$

where $\mathcal{P}_{ij} = P(\tilde{X}_{k+1} = x_j | \tilde{X}_k = x_i)$. Note that $x_1 = 5, x_2 = 0, x_3 = -5$, in the above example. For this markov chain, we can compute the prediction variance according to (46) to obtain $\text{Var}(X_0 | X_{-1}) = 6.6423$. The stationary prior for this DTMC is $\mu = (0.5109, 0.2555, 0.2336)$. For the reversed process $\tilde{X}^{(R)}$, the probability transition matrix is given by

$$
\mathcal{P}^{(R)} = \begin{bmatrix}
0.6 & 0 & 0.4 \\
0.8 & 0.2 & 0 \\
0 & 0.875 & 0.125 \\
\end{bmatrix}
$$

and the prediction variance is 13.9234.

We performed monte carlo simulations to obtain the MSE with finite lookahead (and lookback) for the forward and time reversed continuous-time processes. A brief description of the simulation is provided in Appendix A.

3) Discussion: In Fig. 5, we present the MSE with finite lookahead and lookback for the continuous time process $X$ and $X^{(R)}$ denoting the forward and time-reversed stationary noise-free processes, respectively. From Duncan’s result, we know that the causal and the anti-causal errors must coincide. This is observed (and highlighted in Fig. 6) by the same values for the MSE with 0 lookahead for the forward and reversed processes. Indeed, for both positive and negative lookahead, the MSE’s are different.

Note that we know the asymptotic behavior of the MSE’s with lookahead. As $d \to -\infty$, the forward (and reverse) MSE will converge to the variance $\text{Var}(\tilde{X}_0)$ of the corresponding underlying DTMC (similarly for the time-reversed chain). For $d \to \infty$, the MSE’s converge to the non-causal errors respectively (which are naturally equal for the forward and reversed processes).

This construction illustrates the complicated nature of lookahead in its role as a link between estimation and information for the Gaussian channel. While answering several important questions, it also raises new ones - Do there exist other informational measures of the input-output laws that are sufficient to characterize the estimation error with lookahead? Such directions remain for future exploration.

V. CONCLUSIONS

This work can be viewed as a step towards understanding the role of lookahead in information and estimation. We investigate the benefit of finite lookahead in mean squared signal estimation under additive white noise.
Let \(X\) be a stationary discrete time Markov chain, with known probability transition matrix \(P\). Let \(Y_i(\gamma)\) denote the noisy observation of \(X_i\) corrupted via independent Gaussian noise with the signal to noise ratio of the measurement being \(\gamma\). i.e.,

\[
Y_i(\gamma) = \gamma X_i + W_i^{(i)},
\]

where \(\{W_i^{(i)}\}_i\) are independent standard Brownian motions, which are further independent of the \(X_i\)'s. Let \(X\) denote the corresponding stationary continuous time process generated by the process described in Section IV-A, and let \(Y\) denote the noisy process generated via the continuous time Gaussian channel (1) with input \(X\), i.e.

\[
dY_t = X_t dt + dW_t,
\]

where \(W\) is a standard Brownian motion independent of the \(X\) process.

Define

\[
h(\gamma_1, \gamma_2) = \text{Var}(\tilde{X}_0 | Y_{-\infty}^{-1}(1), Y_0(\gamma_1), Y_1(\gamma_2))
\]

Let, as before

\[
\text{lmmse}(d) = \text{Var}(X_0 | Y_{-\infty}^{d}).
\]

For \(0 < d < 1\), note that

\[
\text{lmmse}(d) = \int_0^d h(1, d - u) du + \int_d^1 h(1 + d - u, 0) du.
\]

A. MCMC approach to estimating \(h(\cdot, \cdot)\)

Note that \(\{\tilde{X}_i, \tilde{Y}_i\}_{i \geq 1}\) for a Hidden Markov process. By using state estimation for HMP's, the following computation can be performed for any \(i\),

\[
\tilde{X}_i = E[\tilde{X}_i | \tilde{Y}_i^{-1}(1), \tilde{Y}_i(\gamma_1), \tilde{Y}_{i+1}(\gamma_2)]
\]

Also, one can observe that

\[
\frac{1}{M} \sum_{i=1}^M (\tilde{X}_i - \tilde{X}_i)^2 \rightarrow h(\gamma_1, \gamma_2),
\]

as \(M \rightarrow \infty\). In the simulations, we chose \(M = 10000\). For this value of \(M\), the quantity in the l.h.s. of (55) was used to approximate \(h(\cdot, \cdot)\), which was then used in the expression for \(\text{lmmse}(d)\) in (53) to obtain the desired values of MSE with finite lookahead, via a monte carlo approach.

Based on a similar approach, it is also possible to compute via MCMC, \(\text{lmmse}(d)\) for \(-1 < d < 0\).

ACKNOWLEDGMENT

This work has been supported under a Stanford Graduate Fellowship, the National Science Foundation (NSF) under grant agreement CCF-0939370, and the Israel Science Foundation (ISF).

REFERENCES