2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Real-Time Tracking of Linear Processes over an AWGN Channel without Feedback: the MMSE Optimality of the Innovation Encoder

Ching-Ling Huang, Raja Sengupta, Jiangchuan Huang

***Abstract*— We first present a formulation for the real-time tracking of a scalar continuous-time linear process over an Additive White Gaussian Noise (AWGN) channel without channel feedback and prove several results for minimum mean-squared error (MMSE) tracking. For the one-to-one channel case, we give an optimal encoder-decoder pair along with the optimal tracking performance. With the Gaussian distributed source innovation, one optimal form of the encoder is shown to be a linear innovation encoder, which scales the source innovation to match with the power constraint of the channel input. We then extend the formulation to the case where multiple source processes are tracked via a shared AWGN channel.**

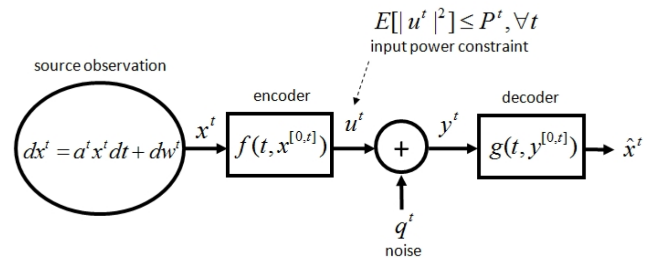

Fig. 1. Analyzed one-to-one channel formulation: the real-time tracking of a scalar linear process over an AWGN channel *without* feedback

I. Introduction

In this paper, a formulation of real-time tracking *without* channel feedback is studied and several results are shown for the MMSE optimality. It has been noted (e.g., by [2,4]) that Shannon's classical information theorems [1] cannot be applied directly to real-time control/tracking problems. For example, many channels that have the same Shannon's capacity do not behave the same way when the encoded sequence is short; they only *function* similarly when the encoded sequence is long enough to achieve decoding typicality and thus channel capacity [1]. Since real-time tracking cannot tolerate such a long encoding delay, many different real-time formulations have been proposed to approach the question of how to track a source process over an unreliable channel, e.g., see [2-9].

For real-time tracking *with* perfect channel feedback, Walrand and Varaiya [2] give one optimal information structure of the encoder. In [2], a discrete-time discrete-state Markov source is tracked over an unreliable discrete-time channel with perfect feedback, i.e., the encoder knows exactly what has been received by the decoder. One of the optimal encoders is shown to be a function of the current state of the Markov source. The formulation is extended by Teneketzis [9] with the channel feedback removed. The optimal encoder in [9] turns out to be a function of current state and the probability distribution of the decoder memory (as a sufficient statistic at the encoder for the decoder's knowledge of the source). The question of how to use minimum information bit rate to control/stabilize a system through feedback is first introduced in [15,27]. Tatikonda [3], Sahai [4], and Mitter specifically treat the case of the unstable Linear Time-Invariant (LTI) source and extend Shannon's results to real-time information measures, namely 1) the *sequential rate distortion* [3] for the bit rate at the source encoder required to achieve stable tracking; and 2) the *anytime capacity* [4] which the channel input bit rate cannot exceed so that the channel bit errors can be corrected exponentially fast and thus enabling uses of the channel for stable real-time tracking of an unstable LTI source.

For real-time tracking *without* channel feedback, the literature also has the packet formulation (e.g., see [5-8]). These papers assume that the source state information (a real number) can be transferred by the packet without distortion. Once the latest measurement arrives at the decoder, any past tracking error is reset as that in a renewal process. In [5], Seiler and Sengupta assume a discrete-time, continuous-state Markov source and derive the necessary and sufficient condition as a linear matrix inequality (LMI) to achieve stable tracking. Sinopoli *et. al.* [6] use a time-varying Kalman filter as the decoder and derive a bound for the channel error probability limiting stable tracking. Xu and Hespanha [7] derive the minimum required packet rate to achieve a stable mean-squared error (MSE) when tracking an unstable LTI source. In [7], the error-dependent transmission is shown to guarantee the stability of tracking MSE. Gupta and Murray [8,29] derive the optimal encoder from a special information structure of the Kalman filter. A packet contains a real number and carries the accumulated source state innovation. The arrival of the latest packet gives all the past source innovation and *washes away* all previous channel errors. A good survey of related studies can be found in [28].

Different from previous real-time tracking formulations (e.g., [2-9]), we start with a continuous-time, continuous-state assumption to formulate the MMSE real-time tracking problem. Like [9], we formulate *without* channel feedback. We also focus only on the scalar AWGN channel. Our work can be considered as the continuous-time, continuous-state counterpart of the formulation in [9]. We show that an MMSE encoding of the innovation of the source process

C. L. Huang, Dr. R. Sengupta, and J. Huang are with Systems Engineering Group, Dept. of Civil and Environmental Engineering, University of California at Berkeley. Contact: chingling@berkeley.edu

at each time epoch gives the MMSE state estimates for all time. For the case of tracking one source, we derive an optimal encoder-decoder pair and the achievable MMSE. Our one-to-one channel formulation is then extended to a multiple-access case with multiple source processes (see Fig. 2). Again, the optimality of innovation encoders is shown for the multiple access formulation. As a corollary, we show that it is impossible to achieve finite asymptotic MSE for real-time tracking an unstable process *without* feedback.

This paper is organized as follows: The formulation of tracking one source over an AWGN channel is presented in Section II along with our main results. The multiple-access AWGN channel formulation is presented in Section III. Section IV summarizes this paper.

II. One-to-One Channel Formulation

Fig. 1 shows the analyzed framework. We consider a finite time horizon problem formulation for time $t \in [0, 1]$. Let C be the Banach space of all continuous functions $z : [0, 1] \to \mathbf{R}$ with norm $\|z\| = \max\{|z(t)||0 \le t \le 1\}$, where $|r|$ is the Euclidean norm of $r \in \mathbf{R}$. Let Γ_t be the smallest σ-field of subsets of C which contains all sets of the form $\{z|z(\tau) \in \beta\}$ where $\tau \in [0, t]$ and β is a Borel subset of $\mathbf{R}$. Let $\Gamma = \Gamma_1$. Let B be the set of the Borel measurable subsets of $[0, 1]$.

The source is a specific type of *Itô processes* (e.g., see [16]) with the differential structure in eq. (1). More specifically, the source process x^t is a scalar linear continuous-time process described by the stochastic differential equation:

$$dx^t = a^t x^t dt + dw^t, x^0 \sim \mathbf{N}(0, \Lambda^0) \quad (1)$$

where $t \in [0, 1]$ is the time, $x^t \in \mathbf{R}$ is the process state, $a^t \in \mathbf{R}, a^t > 0$ is the amplification factor, and $w^0 = 0, w^t \in \mathbf{R}$ is a Wiener process [19,21] that steers the source process. For $0 \le s < t \le 1$, $w^t - w^s \sim \mathbf{N}(0, V(t-s))$ for $V > 0$. The dw^t in eq. (1) is defined as $dw^t \equiv w^{t+dt} - w^t$ and thus $dw^t \sim \mathbf{N}(0, V dt)$. The trajectory of a Wiener process is known to be continuous but not differentiable *almost everywhere* [19,21]. However, a generalized sense of derivative (defined with the integration by parts) for Wiener process can be shown to be a white Gaussian noise [21]. This can be understood as that a Wiener process can be viewed as the limiting behavior of a random walk consisting of small independent increments. In our formulation, this Wiener process w^t is equipped with a generalized sense of derivative $\frac{d}{dt}w^t$, denoted as $v^t \equiv \frac{d}{dt}w^t$ for convenience. This white Gaussian process v^t has $E[v^t] = 0$ and $E[v^t v^s] = V\delta(t-s)$, for $t, s \in (0, 1]$. Throughout this paper, this process v^t is referred as the *innovation* of the source process. The initial condition x^0 is assumed to be zero-mean and Gaussian distributed with variance $\Lambda^0 > 0$. In addition, the Wiener process $w^{[0,1]}$ (and thus the innovation process $v^{(0,1]}$) is independent of the initial state x^0.

The process x^t in eq. (1) is a Gaussian process [21]. The probability distribution of state trajectory $x^{[0,t]}$ depends on the probability distribution of the initial state x^0 and the Wiener measure of the trajectory $w^{[0,t]}$. The probability measure on $x^{[0,t]}$ is *absolutely continuous* with respect to the Wiener measure on the space C of all continuous functions from $[0, 1]$ into $\mathbf{R}$ (e.g., see Lemma 2 and Corollary 2 of [20]). The source process x^t is observed perfectly by the causal measurable encoder function at time t, denoted as f, which produces u^t as the input to the AWGN channel at time t. Let $f : [0, 1] \times C \to \mathbf{R}$ be causal, i.e., f is not only $B \otimes \Gamma$ measurable but also measurable with respect to Γ_t for each fixed $t \in [0, 1)$. The encoder f can be parameterized as:

$$u^t = f(t, x^{[0,t]}) \quad (2)$$

where $x^{[0,t]}$ represents the trajectory of the source process as the full history up to time t. The channel input $u^t \in \mathbf{R}$ is chosen to deliver information to the decoder so that the decoder can produce an estimate of the state of the source process. The encoder function f can be time-varying and hence has the parameter t (i.e., one can specify a different encoder function f for each t). With the given trajectory of $x^{[0,t]}$, the trajectory of $w^{[0,t]}$ can be perfectly reconstructed for each t (e.g., see the discussion on pp. 355-358 and Theorem 1 of [20]). Therefore, both the trajectory of $x^{[0,t]}$ and the trajectory of $w^{[0,t]}$ (and thus $v^{(0,t]}$) are measurable by the encoder function $f(t, \cdot)$.

The AWGN channel output is the superposition of channel input u^t and the channel noise q^t:

$$y^t = u^t + q^t \quad (3)$$

where $y^t \in \mathbf{R}$ is the channel output at time t, and $q^t \in \mathbf{R}$ is a white Gaussian noise process independent of the source initial state x^0, the Wiener process $w^{[0,1]}$, and the innovation process $v^{(0,1]}$. The noise process q^t has $E[q^t] = 0$ and $E[q^t q^s] = Q^t \delta(t-s)$, $Q^t > 0$, for $t, s \in [0, 1]$. We choose the formulation in eq. (3) instead of a stochastic differential form (e.g., eq. (1)) to match with most definitions of an AWGN channel [13,14]. The channel output y^t is observed by the decoder, denoted as g, which produces a real-time estimate of current state of the source process x^t. Let $g : [0, 1] \times C \to \mathbf{R}$ be causal, i.e., g is not only $B \otimes \Gamma$ measurable but also measurable with respect to Γ_t for each fixed $t \in [0, 1)$. The decoder g can be parameterized as:

$$\hat{x}^t = g(t, y^{[0,t]}) \quad (4)$$

where $\hat{x}^t \in \mathbf{R}$ is the estimate of x^t and $y^{[0,t]}$ is the full history of observed channel outputs up to time t. In this real-time tracking formulation, causality is imposed on the encoder f and decoder g. Once this estimate $\hat{x}^t$ is produced at time t, it is final and can not be improved later based on future channel outputs $y^{(t,1]}$. This decoder function g can be time-varying and hence has the parameter t (i.e., one can specify a different encoder function g for each t).

The cost function $J_{(1)}$ is defined over a finite time horizon $t \in [0, 1]$:

$$J_{(1)} \equiv \int_0^1 E[|x^t - \hat{x}^t|^2] dt \quad (5)$$

of which the expectation is taken with respect to the probability distribution of the initial condition x^0, the Wiener process $w^{[0,1]}$, and the white Gaussian channel noise process $q^{[0,1]}$.

With a given pair of encoder-decoder, the MSE $E[|x^t - \hat{x}^t|^2]$ can be calculated for each t. This cost function $J_{(1)} \in \mathbf{R}, J_{(1)} \geq 0$ is essentially the Lebesgue integral of the real-time MSE over the unit time interval $[0, 1]$.

We define a finite time horizon optimization problem based on the cost function $J_{(1)}$ in eq. (5) to find an encoder-decoder pair, denoted as $\{f, g\}$, such that the cost function $J_{(1)}$ can be minimized:

$$\{f^*, g^*\} = \underset{E[|u^t|^2] \leq P^t, \forall t \in [0,1]}{\operatorname{argmin}} J_{(1)} \qquad (6)$$

where $\{f^*, g^*\}$ denotes the optimal encoder-decoder pair and $P^t < \infty$ is the finite power constraint on the channel input at time t. The total transmission energy allowed is also finite over the same time horizon $t \in [0, 1]$, i.e., $\int_0^1 P^t dt < \infty$. This power constraint P^t is meant to avoid the degenerate formulation in which the encoder can amplify the channel input signal to an arbitrarily large value to *overwhelm* the channel noise q^t and the decoder can then restore the signal with an arbitrarily small error.

A. Preliminaries for One-to-One Channel Formulation

Let $\bar{g} : [0, 1] \times C \to \mathbf{R}$ be causal, i.e., $\bar{g}$ is not only $B \otimes \Gamma$ measurable but also measurable with respect to Γ_t for each fixed $t \in [0, 1)$. We define a differential decoder for $\hat{x}^t$ (the real-time estimate of the source process) in the stochastic differential equation:

$$d\hat{x}^t = a^t \hat{x}^t dt + \bar{g}(t, y^{[0,t]})dt, \hat{x}^0 = \bar{g}(0, y^0) \qquad (7)$$

where a^t is exactly the same amplification factor as that in eq. (1) and the function $\bar{g}(t, y^{[0,t]})$ steers the evolution of $\hat{x}^t$. The similarity between eq. (1) and eq. (7) acknowledges the rationale that the decoder output $\hat{x}^t$ in eq. (7) incorporates the knowledge of the model of the source process x^t.

Lemma 1: Given the real-time tracking formulation in (6), there is no loss of optimality (in the MMSE sense for each $t \in [0, 1]$) by assuming the differential decoder in eq. (7).

Proof: The form of the differential decoder in eq. (7) is MMSE optimal according to the nonlinear filtering analysis by Clark [11], Frost and Kailath [12], and Lo [17] (e.g., see Theorem 3 in [11], Theorem 3-5 in [12], and Theorem 1 in [17]). The finite channel input energy and Gaussian channel noise are the key elements to guarantee the optimality of the differential decoder form in eq. (7). See specifically the discussion of Gauss-Markov Process (Linear Case) on pp. 222 in [12]. ■

Theorem 2: Let $\Sigma^t \equiv E[(x^t - \hat{x}^t)^2]$ denote the tracking MSE at time t. With the optimal differential decoder (defined in eq. (7)), denoted as $\bar{g}^*(t, y^{[0,t]})$, the MMSE at time t, denoted as Σ^{*t}, can be described by the differential equation: for $t \in (0, 1]$,

$$d\Sigma^{*t} = 2a^t \Sigma^{*t} dt + E[(dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2] \qquad (8)$$

with the initial condition, $t = 0$: $\Sigma^{*0} = E[(x^0 - \bar{g}^*(0, y^0))^2]$.

Proof: Let $e^t \equiv x^t - \hat{x}^t$. We get the differential equation of e^t based on eq. (1) and eq. (7):

$$de^t = a^t e^t dt + dw^t - \bar{g}(t, y^{[0,t]})dt. \qquad (9)$$

With eq. (9) and the definition $\Sigma^t = E[(e^t)^2]$, the differential of MMSE Σ^{*t} with the optimal differential decoder $\bar{g}^*$ can be derived as follows:

$d\Sigma^{*t} = \Sigma^{*(t+dt)} - \Sigma^{*t} = E[(e^{*t} + de^{*t})^2] - E[(e^{*t})^2]$
$= 2E[e^{*t} de^{*t}] + E[(de^{*t})^2]$
$= 2a^t E[(e^{*t})^2]dt + 2E[e^{*t}(dw^t - \bar{g}^*(t, y^{[0,t]})dt)]$
$+E[(a^t e^{*t} dt + dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2]$
(since $de^{*t} = a^t e^{*t} dt + dw^t - \bar{g}^*(t, y^{[0,t]})dt$ from eq. (9))
$= 2a^t E[(e^{*t})^2]dt - 2E[e^t \bar{g}^*(t, y^{[0,t]})dt]$
$+(a^t)^2 E[(e^{*t})^2](dt)^2 + E[(dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2]$
$-2a^t E[e^{*t} \bar{g}^*(t, y^{[0,t]})dt]dt$
(since the Wiener process has independent increments, dw^t is independent of e^{*t})
$= 2a^t E[(e^{*t})^2]dt + (a^t)^2 E[(e^{*t})^2](dt)^2$
$+E[(dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2]$
(due to a general form of Orthogonality Principle: the optimal tracking error is orthogonal to any function of all the past observations, e.g., see pp. 268 in [18])
$= 2a^t E[(e^{*t})^2]dt + E[(dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2]$
(neglect the higher order term associated with $(dt)^2$)

which leads to the differential equation in (8). ■

Corollary 3: Given the real-time tracking formulation in (6), the MMSE optimal differential decoder $\bar{g}^*$ can be expressed, for $t \in (0, 1]$, as

$$\bar{g}^*(t, y^{[0,t]}) = E[v^t | y^{[0,t]}] \qquad (10)$$

with the initial condition, $t = 0$: $\bar{g}^*(0, y^0) = E[x^0 | y^0]$.

Proof: It has been noted that conditional expectation given observations minimizes the MSE (e.g., see the discussion on pp. 218 in [12]). From the differential form in eq. (8), observe that the MMSE Σ^{*t} can be minimized if all the previous $E[(dw^\tau - \bar{g}^*(\tau, y^{[0,\tau]})d\tau)^2]$ are minimized for all $0 < \tau \leq t$ and $E[(x^0 - \bar{g}^*(0, y^0))^2]$ is minimized for the initial condition. The differential decoder function $\bar{g}(t, \cdot)$ has all the channel observations $y^{[0,t]}$ and can be chosen for each t independently. This $\bar{g}(t, \cdot)$ does not depend on previous decoder function $\bar{g}(\tau, \cdot), 0 \leq \tau < t$.

Because the conditional expectation $E[dw^\tau | y^{[0,\tau]}]$ minimizes this expectation for each τ, we get $\bar{g}^*(\tau, y^{[0,\tau]})d\tau = E[dw^\tau | y^{[0,\tau]}]$ and thus $\bar{g}^*(\tau, y^{[0,\tau]}) = E[\frac{d}{d\tau} w^\tau | y^{[0,\tau]}]$. Since the innovation process v^t is defined as the generalized sense of derivative of w^t, i.e., $v^t \equiv \frac{d}{dt} w^t$, we get eq. (10). For the initial condition, $t = 0$, the decoder is the conditional expectation of the source initial state x^0 given the available channel observation y^0. ■

Corollary 4: Given the real-time tracking formulation in (6) and $a^t \geq 0.5, \forall t$, it is impossible to have asymptotic stable MSE, i.e., $\lim_{t \to \infty} \Sigma^t \to \infty$.

Proof: First partition the infinite time line into segments of unit time length as in our formulation. In each unit time segment, the same differential equation (8) for Σ^{*t} still applies but with different initial conditions. Define the variable Ψ_1^t by

$$d\Psi_1^t = 2a^t \Psi_1^t dt, \Psi_1^0 = \Sigma^{*0} \qquad (11)$$

and thus $\Psi_1^t \leq \Sigma^{*t}, \forall t > 0$ by comparing eq. (11) and eq. (8) and the fact that $E[(dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2] > 0$ for nontrivial channel noise q^t, $\forall t > 0$. Now define another variable Ψ_2^t by

$$d\Psi_2^t = \Psi_2^t dt, \Psi_2^0 = \Sigma^{*0} \quad (12)$$

and thus, if $2a^t \geq 1, \forall t$, $\Psi_2^t \leq \Psi_1^t \leq \Sigma^{*t}, \forall t > 0$. The solution of the differential equation (12) can be expressed as $\Psi_2^t = \exp(t)\Psi_2^0$. Since $\Psi_2^0 = E[(x^0 - \bar{g}^*(0, y^0))^2] > 0$ for nontrivial channel noise q^0,

$$\infty = \lim_{t\to\infty} \Psi_2^t \leq \lim_{t\to\infty} \Psi_1^t \leq \lim_{t\to\infty} \Sigma^{*t}.$$

Therefore, in our formulation *without* channel feedback, if $2a^t \geq 1, \forall t$, no encoder-decoder pair can achieve stable MSE tracking asymptotically. ■

B. Main Results for One-to-One Channel Formulation

In this subsection, we state an optimal encoder-decoder pair for the one-to-one channel formulation and the optimal tracking performance. With the Gaussian distributed source innovation v^t, the encoder is shown to be a linear innovation encoder and the associated decoder is a differential decoder that steers the state estimate $\hat{x}^t$ according to eq. (7).

Lemma 5: Given the real-time tracking formulation in (6), one optimal form of the encoder function and the differential decoder function (defined in eq. (7)) can be expressed as follows: for each $t \in (0, 1]$, the encoder is a function of v^t, denoted as $\bar{f}^*(t, v^t)$, and the associated differential decoder is denoted as $\bar{g}^*(t, y^t) = E[v^t|y^t]$. For the initial condition, $t = 0$, the optimal encoder can be expressed as $\bar{f}^*(0, x^0)$ with the associated optimal decoder $\bar{g}^*(0, y^0) = E[x^0|y^0]$.

Proof: Assume a pair of optimal encoder-decoder at t: $f^*(t, \cdot)$ and $\bar{g}^*(t, \cdot)$. The optimal differential decoder form $\bar{g}^*$ follows from Corollary 3:

$\bar{g}^*(t, y^{[0,t]}) = E[v^t|y^{[0,t]}] = E[v^t|y^t]$
(since v^t is independent of x^0, $v^{(0,t)}$, past channel outputs $y^{[0,t)}$ do not contain any information of v^t due to causality)
$= E[v^t|q^t + u^t]$
$= E[v^t|q^t + f^*(t, \{x^0, v^{(0,t)}, v^t\})]$
(the source, encoder f^*, channel, and decoder $\bar{g}^*$ form a Markov chain: $\{x^0, v^{(0,t)}, v^t\} \to u^t \to y^t \to E[v^t|y^t]$)
$= E[v^t|q^t + \bar{f}^*(t, v^t)]$ for some function $\bar{f}^*(t, \cdot)$
(since v^t is independent of x^0, $v^{(0,t)}$, the set $\{v^t\}$ is the same *informative* as the set $\{x^0, v^{(0,t)}, v^t\}$ for the decoder to estimate v^t, e.g., see Theorem 3 and 4 in [24]; thus one can design another function $\bar{f}^*(t, \cdot)$ to focus on delivering the sufficient statistic $\{v^t\}$ without loss of optimality)

Therefore, without loss of optimality, the encoder at time $t \in (0, 1]$ can focus on delivering v^t to help the decoder better estimate the innovation v^t to steer the state estimate $\hat{x}^t$ according to eq. (7). The decoder can be a function of current channel observation y^t mainly due to causality. ■

Theorem 6: Given the real-time tracking formulation in (6), one pair of optimal encoder function $\bar{f}^*$ and associated differential decoder $\bar{g}^*$ can be expressed, for $t \in (0, 1]$, as:

$$\bar{f}^*(t, v^t) = v^t \sqrt{\frac{P^t}{V}} \quad (13)$$

and

$$\bar{g}^*(t, y^t) = y^t \frac{\sqrt{VP^t}}{P^t + Q^t}. \quad (14)$$

For the initial condition, $t = 0$, $\bar{f}^{*0}(x^0) = x^0\sqrt{\frac{P^0}{\Lambda^0}}$ and $\bar{g}^{*0}(y^0) = y^0 \frac{\sqrt{\Lambda^0 P^0}}{P^0 + Q^0}$.

Proof: Based on Lemma 5, the optimal encoder scales the Gaussian distributed innovation v^t to match with the channel input power constraint P^t (see pp. 561-562, 564 in Goblick [13] and pp. 1153 in Gastpar [14]), and this direct transmission can minimize the term $E[(dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2]$ in eq. (8) for each $t \in (0, 1]$ and achieves MMSE optimal. The optimal differential decoder in eq. (14) is the conditional expectation of the source innovation v^t given y^t and the linear innovation encoder in eq. (13). The initial condition for $t = 0$ follows similarly. ■

Corollary 7: Given the real-time tracking formulation in (6), the optimal cost $J^*_{(1)}$ achievable is given by

$$J^*_{(1)} = \int_0^1 \Sigma^{*t} dt \quad (15)$$

where Σ^{*t} can be described by the differential equation:

$$\frac{d}{dt}\Sigma^{*t} = 2a^t\Sigma^{*t} + V(\frac{Q^t}{P^t + Q^t})^2, \Sigma^{*0} = \frac{\Lambda^0 Q^0}{P^0 + Q^0}. \quad (16)$$

Proof: Based on the optimal encoder and differential decoder in Theorem 6,

$E[(dw^t - \bar{g}^*(t, y^{[0,t]})dt)^2]$
$= E[(dw^t - (v^t\sqrt{\frac{P^t}{V}} + q^t)\frac{\sqrt{VP^t}}{P^t+Q^t}dt)^2]$
$= E[(dw^t - v^t dt\frac{P^t}{P^t+Q^t} - q^t dt\frac{\sqrt{VP^t}}{P^t+Q^t})^2]$
$= E[(dw^t\frac{Q^t}{P^t+Q^t} - q^t dt\frac{\sqrt{VP^t}}{P^t+Q^t})^2]$
$= E[(dw^t\frac{Q^t}{P^t+Q^t})^2] + E[(q^t dt\frac{\sqrt{VP^t}}{P^t+Q^t})^2]$
(since dw^t is independent of the channel noise q^t)
$= (\frac{Q^t}{P^t+Q^t})^2 E[(dw^t)^2] + \frac{VP^tQ^t}{(P^t+Q^t)^2}(dt)^2 = (\frac{Q^t}{P^t+Q^t})^2 V dt$
(neglect the higher order term associated with $(dt)^2$; apply the property of the Wiener process $E[(dw^t)^2] = Vdt$)

Together with eq. (8), we get the differential form in eq. (16). The initial condition, $t = 0$, can be derived as follows:

$E[(x^0 - \bar{g}^*(0, y^0))^2] = E[(x^0 - (x^0\sqrt{\frac{P^0}{\Lambda^0}} + q^0)\frac{\sqrt{\Lambda^0 P^0}}{P^0+Q^0})^2]$
$= E[(x^0\frac{Q^0}{P^0+Q^0} - q^0\frac{\sqrt{\Lambda^0 P^0}}{P^0+Q^0})^2]$
$= E[(x^0\frac{Q^0}{P^0+Q^0})^2] + E[(q^0\frac{\sqrt{\Lambda^0 P^0}}{P^0+Q^0})^2]$
(since x^0 is independent of the channel noise q^0)
$= (\frac{Q^0}{P^0+Q^0})^2\Lambda^0 + \frac{\Lambda^0 P^0}{(P^0+Q^0)^2}Q^0 = \frac{\Lambda^0 Q^0}{P^0+Q^0}$

which leads to the Σ^{*0} in eq. (16). The simple form of the optimal encoder-decoder pair in eq. (13) and eq. (14) is mainly due to the fact that the Gaussian distributed source innovation v^t and Gaussian channel noise q^t are *matched* (see pp. 1152-1153 in Gastpar [14]). ■

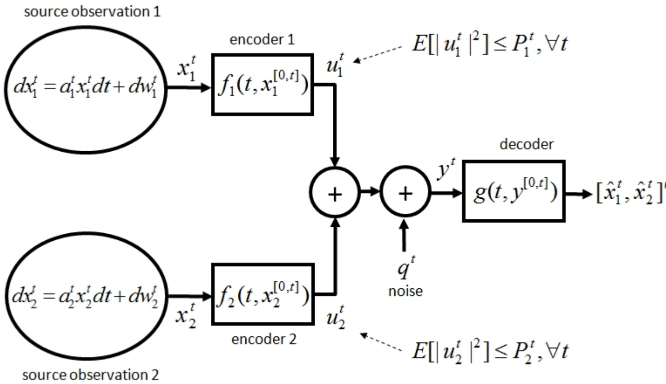

Fig. 2. Analyzed multiple-access channel formulation: the real-time tracking of $n = 2$ linear processes over an AWGN channel *without* feedback

III. Multiple-Access Channel Formulation

In this section, we extend the one-to-one channel formulation in Section II to a multiple-access channel formulation with n source processes, $n \ge 2$. Fig. 2 shows one example of the analyzed formulation with $n = 2$ sources. Similar to the *Itô process* defined in eq. (1), each source process $x_i^t, i = 1, ..., n,$ is a scalar linear continuous-time process described by the stochastic differential equation:

$$dx_i^t = a_i^t x_i^t dt + dw_i^t, x_i^0 \sim \mathbf{N}(0, \Lambda_i^0) \tag{17}$$

where $t \in [0, 1]$ is the time, $x_i^t \in \mathbf{R}$ is the process state, $a_i^t \in \mathbf{R}$ is the amplification factor, and $w_i^t \in \mathbf{R}$ is a Wiener process that steers the i-th source process. For $0 \le s < t \le 1$, $w_i^t - w_i^s \sim \mathbf{N}(0, V_i(t-s))$ for $V_i > 0$. This Wiener process w_i^t is equipped with a generalized sense of derivative $\frac{d}{dt}w_i^t$, denoted as the i-th innovation process $v_i^t \equiv \frac{d}{dt}w_i^t$. This white Gaussian process v_i^t has $E[v_i^t] = 0$ and $E[v_i^t v_i^s] = V_i \delta(t-s)$, for $t, s \in (0, 1]$. The initial condition x_i^0 is assumed to be zero-mean and Gaussian distributed with variance $\Lambda_i^0 > 0$. Among n sources, the initial condition x_i^0 and the innovation v_i^t are assumed to be mutually independent for all t and i.

Each source process $x_i^t, i = 1, ..., n,$ is observed perfectly by the i-th causal measurable encoder function f_i which produces $u_i^t \in \mathbf{R}$ as the i-th input to the AWGN channel. Let $f_i : [0, 1] \times C \to \mathbf{R}$ be causal, i.e., f_i is not only $B \otimes \Gamma$ measurable but also measurable with respect to Γ_t for each fixed $t \in [0, 1)$. The encoder f_i can be parameterized as:

$$u_i^t = f_i(t, x_i^{[0,t]}) \tag{18}$$

where $x_i^{[0,t]}$ represents the trajectory of the i-th source process as the full history up to time t. Note that the information structure of our formulation only allows the i-th encoder to observe the i-th source process.

The AWGN channel output is the superposition of the sum of the channel inputs $\sum_{i=1}^n u_i^t$ and the channel noise q^t:

$$y^t = \sum_{i=1}^n u_i^t + q^t \tag{19}$$

where $y^t \in \mathbf{R}$ is the channel output at time t, and $q^t \in \mathbf{R}$ is a white Gaussian noise process independent of $x_i^0, i = 1, ..., n,$ $w_i^{[0,1]}, i = 1, ..., n,$ and $v_i^{(0,1]}, i = 1, ..., n.$ The noise process q^t has $E[q^t] = 0$ and $E[q^t q^s] = Q^t \delta(t-s)$, $Q^t > 0$, for $t, s \in [0, 1]$. The channel output y^t is observed by the decoder g to produce a vector of estimates of $x_i^t, i = 1, ..., n.$ Let $g : [0, 1] \times C \to \mathbf{R}_{n \times 1}$ be causal, i.e., g is not only $B \otimes \Gamma$ measurable but also measurable with respect to Γ_t for each fixed $t \in [0, 1)$. The decoder g can be parameterized as:

$$[\hat{x}_1^t, ..., \hat{x}_n^t]' = g(t, y^{[0,t]}) \tag{20}$$

where each $\hat{x}_i^t \in \mathbf{R}$ is the estimate of x_i^t and $y^{[0,t]}$ is the full history of observed channel outputs up to time t.

The cost function $J_{(n)}$ is defined over a finite time horizon $t \in [0, 1]$:

$$J_{(n)} \equiv \int_0^1 \sum_{i=1}^n E[|x_i^t - \hat{x}_i^t|^2] dt \tag{21}$$

where n is the number of source processes and the expectation is taken with respect to the probability distribution of the initial condition $x_i^0, i = 1, ..., n,$ the Wiener processes $w_i^{[0,1]}, i = 1, ..., n,$ and the channel noise process $q^{[0,1]}$. With a given set of encoders-decoder, the MSE $E[|x_i^t - \hat{x}_i^t|^2], \forall i$ can be calculated for each t. This cost function $J_{(n)} \in \mathbf{R}, J_{(n)} \ge 0$ is essentially the Lebesgue integral of the sum of all the real-time tracking MSE over the unit time interval $[0, 1]$. The subscript of $J_{(n)}$ means that this multiple-access formulation considers tracking of n multiple sources, which is different from the cost function in eq. (5).

We define an optimization problem based on the cost function $J_{(n)}$ in eq. (21) to find the set of encoders and decoder for $t \in [0, 1]$, denoted as $\{f_1, ..., f_n, g\}$, such that the cost function $J_{(n)}$ can be minimized:

$$\{f_1^*, ..., f_n^*, g^*\} = \underset{E[|u_i^t|^2] \le P_i^t, \forall i, \forall t \in [0,1]}{\operatorname{argmin}} J_{(n)} \tag{22}$$

where f_i^* is the optimal encoder for the i-th source, g^* is the optimal decoder, and $P_i^t < \infty$ is the finite power constraint on u_i^t, the channel input from the i-th encoder, at time $t \in [0, 1]$, and the total energy is also finite: $\sum_{i=1}^n \int_0^1 P_i^t dt < \infty$. Again, this finite power/energy constraint is meant to avoid the degenerate formulation mentioned in Section II.

A. Preliminaries for Multiple-Access Channel Formulation

Let $\mathbf{x}^t \equiv [x_1^t, ..., x_n^t]'$ denote the vector of the source states and similarly $\mathbf{w}^t \equiv [w_1^t, ..., w_n^t]'$. The source processes can then be described in the stochastic differential equation:

$$d\mathbf{x}^t = \mathbf{a}^t \mathbf{x}^t dt + d\mathbf{w}^t \tag{23}$$

with $\mathbf{a}^t \equiv diag(a_1^t, ..., a_n^t)$ where $diag$ represents a diagonal matrix with indicated diagonal elements of the same order.

Let $\bar{g} : [0, 1] \times C \to \mathbf{R}_{n \times 1}$ be causal, i.e., $\bar{g}$ is not only $B \otimes \Gamma$ measurable but also measurable with respect to Γ_t for each fixed $t \in [0, 1)$. We define a differential decoder for real-time estimate $\hat{\mathbf{x}}^t \equiv [\hat{x}_1^t, ..., \hat{x}_n^t]'$ in the stochastic differential equation:

$$d\hat{\mathbf{x}}^t = \mathbf{a}^t \hat{\mathbf{x}}^t dt + \bar{g}(t, y^{[0,t]}) dt, \hat{\mathbf{x}}^0 = \bar{g}(0, y^0) \tag{24}$$

where $\mathbf{a}^t$ is exactly the same amplification factors as that in eq. (23) and the function $\bar{g}(t, y^{[0,t]})$ produces the $n \times 1$ vector that steers the evolution of $\hat{\mathbf{x}}^t$, the decoder output.

Lemma 8: Given the multiple-access real-time tracking formulation in (22), there is no loss of optimality by assuming the differential decoder in eq. (24).

Proof: Similar to the proof of Lemma 1. ■

Theorem 9: Let $\mathbf{\Sigma}^t \equiv E[(\mathbf{x}^t - \hat{\mathbf{x}}^t)(\mathbf{x}^t - \hat{\mathbf{x}}^t)']$ denote the tracking error covariance matrix at time t. The optimal error covariance $\mathbf{\Sigma}^{*t}$ can be described by the differential equation:

$$d\mathbf{\Sigma}^{*t} = 2\mathbf{a}^t \mathbf{\Sigma}^{*t} dt + E[\Psi^t (\Psi^t)'] \tag{25}$$

where $\Psi^t \equiv d\mathbf{w}^t - \bar{g}^*(t, y^{[0,t]})dt$. The vector $\mathbf{w}^t$ is the vector of Wiener processes as defined in eq. (23). The function $\bar{g}^*(t, y^{[0,t]})$ is the optimal differential decoder in eq. (24).

Proof: Let $\mathbf{e}^t \equiv \mathbf{x}^t - \hat{\mathbf{x}}^t$ denote the error vector. We get the differential equation of $\mathbf{e}^t$ based on eq. (23) and eq. (24):

$$d\mathbf{e}^t = \mathbf{a}^t \mathbf{e}^t dt + d\mathbf{w}^t - \bar{g}(t, y^{[0,t]})dt. \tag{26}$$

With eq. (26) and $\mathbf{\Sigma}^t = E[\mathbf{e}^t (\mathbf{e}^t)']$, $d\mathbf{\Sigma}^{*t}$ with the optimal differential decoder $\bar{g}^*$ can be expressed as eq. (25) with similar arguments as in the proof of Theorem 2. ■

Theorem 10: Let $tr(\cdot)$ denote the trace of the input matrix. Given the multiple-access real-time tracking formulation in (22), the optimal differential decoder $\bar{g}^*$ that achieves the Minimum Sum of MSE (MSMSE) $tr(\mathbf{\Sigma}^{*t})$ can be expressed, for $t \in (0, 1]$, as:

$$\bar{g}^*(t, y^t) = E[\mathbf{v}^t | y^t] \tag{27}$$

where the innovation vector $\mathbf{v}^t \equiv [v_1^t, ..., v_n^t]'$. One optimal form of encoders is the innovation encoders: for $t \in (0, 1]$, $\bar{f}_i^*(t, v_i^t), i = 1, ..., n$. For the initial condition, $t = 0$, optimal encoders can be expressed as $\bar{f}_i^*(0, x^0), i = 1, ..., n$, with the associated optimal decoder $\bar{g}^*(0, y^0) = E[\mathbf{x}^0 | y^0]$.

Proof: Based on the differential form of $\mathbf{\Sigma}^{*t}$ in eq. (25) and similar arguments as in the proof of Corollary 3, to achieve MSMSE $tr(\mathbf{\Sigma}^{*t})$, $\bar{g}^*(t, y^{[0,t]})dt = E[d\mathbf{w}^t | y^{[0,t]}]$ and thus $\bar{g}^*(t, y^{[0,t]}) = E[\frac{d}{dt}\mathbf{w}^t | y^{[0,t]}]$. Since $v_i^t \equiv \frac{d}{dt} w_i^t, \forall i$, we get $\bar{g}^*(t, y^{[0,t]}) = E[\mathbf{v}^t | y^{[0,t]}]$. Based on similar arguments in the proof of Lemma 5, $\bar{g}^*(t, y^{[0,t]}) = E[\mathbf{v}^t | y^t]$ due to causality. Since the i-th source innovation can only be observed by the corresponding i-th encoder, one optimal form of encoders is to focus on delivering the innovation v_i^t at each time t similar to those arguments in the proof of Lemma 5. The initial condition $t = 0$ follows similarly. ■

Theorem 10 says that, at each t, the i-th encoder can focus on communicating v_i^t to the decoder. However, how to design optimal encoders to communicate v_i^t over a shared channel to the decoder is an on-going research [25,26]. We will extend this multiple-access formulation in our future work.

IV. Summary

We first analyze the real-time tracking MMSE of a scalar linear continuous-time source over a scalar AGWN channel *without* channel feedback. With the Gaussian distributed source innovation, the optimality of the linear innovation encoder and associated optimal tracking performance are shown for the one-to-one channel case. We then extend the one-to-one channel formulation to the case of tracking multiple sources over a shared AWGN channel.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol.27, July 1948.
[2] J. C. Walrand, P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Info. Theory*, vol.29, no.6, Nov. 1983.
[3] S. Tatikonda, A. Sahai, S. K. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. on Automat. Contr.*, vol.49, no.9, Sept. 2004.
[4] A. Sahai, Anytime Information Theory, Ph.D. dissertation, Dept. EECS, MIT, Cambridge, MA, Feb. 2001.
[5] P. Seiler, R. Sengupta, "An H_∞ approach to networked control," *IEEE Trans. on Automat. Contr.*, vol.50, no.3, March 2005.
[6] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, S. S. Sastry, "Kalman Filtering with Intermittent Observations," *IEEE Trans. Automat. Contr.*, vol.49, no.9, Sept. 2004.
[7] Y. Xu, J. Hespanha, "Estimation under uncontrolled and controlled communication in Networked Control Systems," *Proc. of the IEEE Conf. Decision and Contr.*, Dec. 2005.
[8] V. Gupta, Distributed Estimation and Control in Networked Systems, Ph.D. dissertation, California Institute of Technology, CA, June 2006.
[9] D. Teneketzis, "On the Structure of Optimal Real-Time Encoders and Decoders in Noisy Communication," *IEEE Trans. Info. Theory*, vol.52, no.9, Sept. 2006.
[10] A. Zygmund, Trigonometric series, 2nd Edition, Cambridge University Press, 1988.
[11] J. M. C. Clark, "Conditions for one-to-one correspondence between an observation process and its innovation," *Center for Computing and Automation, Imperial College, London, Tech. Rep. 1*, 1969.
[12] P. A. Frost, T. Kailath, "An Innovations Approach to Least-Squares Estimation - Part III: Nonlinear Estimation in White Gaussian Noise," *IEEE Trans. Automat. Contr.*, vol.16, no.3, June 1971.
[13] T. J. Goblick, "Theoretical Limitation on Transmission of Data from Analog Sources," *IEEE Trans. Info. Theory*, vol.11, no.4, Oct. 1965.
[14] M. Gastpar, B. Rimoldi, M. Vetterli, "To Code, or Not to Code: Lossy Source-Channel Communication Revisited," *IEEE Trans. Info. Theory*, vol.49, no.5, May 2003.
[15] J. K. Yook, D. M. Tilbury, N. R. Soparkar, "Trading computation for bandwidth: Reducing communication in distributed control systems using state estimators," *IEEE Trans. Contr. Syst. Technol*, vol.10, no.4, July 2002.
[16] N. Ikeda, S. Watanabe, Stochastic Differential Equations and Diffusion Processes, North-Holland Publishing Company, 1981.
[17] J. T. Lo, "On optimal nonlinear estimation-Part I: continuous observation," *Proc. 8th Allerton Conf. on Cir. and Sys. Theory*, 1970.
[18] A. Papoulis, S. U. Phillai, Probability, Random Variables and Stochastic Processes, McGrawHill, 2002.
[19] R. Durrett, Probability: Theory and Examples, Brooks/Cole, 2002.
[20] T. Duncan, P. Varaiya, "On the solutions of a stochastic control systems," *SIAM J. Control*, vol.9, no.3, Aug. 1971.
[21] R. S. Lipster, Lecture Notes #10 for the course "Stochastic Process": http://www.eng.tau.ac.il/~liptser/lectures/lect_new10.pdf.
[22] J. Galambos, Advanced Probability Theory, CRC Press, 1988.
[23] D. Bertsekas, R. Gallager, Data Networks, Prentice Hall, 1992.
[24] C. Striebel, "Sufficient Statistics in the Optimal Control of Stochastic Systems," *J. of Math. Analysis and App.*, vol.12, pp. 576-592, 1965.
[25] A. Lapidoth, S. Tinguely, "Sending a Bivariate Gaussian Source over a Gaussian MAC," *IEEE Trans. Info. Theory*, vol.56, no.6, June 2010.
[26] P. A. Floor, A. N. Kim, N. Wernersson, T. A. Ramstad, M. Skoglund, I. Balasingham, "On Transmission of Two Correlated Gaussian Memoryless Sources over a Gaussian MAC using Delay-Free Mappings," submitted to *IEEE Transactions on Communications*, 2011.
[27] G. N. Nair, R. J. Evans, "Exponential stabilisability of finite dimensional linear systems with limited data rates," *Automatica*, vol.39, no.4, Apr. 2003.
[28] J. Hespanha, P. Naghshtabrizi, Y Xu, "A Survey of Recent Results in Networked Control Systems," *Proc. of IEEE, Special Issue on Networked Control Systems*, Jan. 2007.
[29] V. Gupta, A. Dana, J. Hespanha, R. Murray, "Data Transmission over Networks for Estimation," *Proc. of International Symp. on the Math Theory of Networks and Systems*, 2006.