

Directed Information and Causal Estimation in Continuous Time

Young-Han Kim
University of California, San Diego
La Jolla, CA, USA
yhk@ucsd.edu

Haim H. Permuter
Ben Gurion University of the Negev
Beer-Sheva, Israel
haimp@ee.bgu.ac.il

Tsachy Weissman
Stanford University/Technion
Stanford, CA, USA/Haifa, Israel
tsachy@ee.technion.ac.il

Abstract—The notion of directed information is introduced for stochastic processes in continuous time. Properties and operational interpretations are presented for this notion of directed information, which generalizes mutual information between stochastic processes in a similar manner as Massey’s original notion of directed information generalizes Shannon’s mutual information in the discrete-time setting. As a key application, Duncan’s theorem is generalized to estimation problems in which the evolution of the target signal is affected by the past channel noise, and the causal minimum mean squared error estimation is related to directed information from the target signal to the observation corrupted by additive white Gaussian noise. An analogous relationship holds for the Poisson channel.

I. INTRODUCTION

Directed information $I(X^n \rightarrow Y^n)$ between two random n -sequences $X^n = (X_1, \dots, X_n)$ and $Y^n = (Y_1, \dots, Y_n)$ is a natural generalization of Shannon’s mutual information to random objects with causal structures. Introduced by Massey [1], this notion of directed information has been shown to arise as the canonical answer to a variety of problems with causally dependent components. For example, it plays a pivotal role in characterizing the capacity C_{FB} of communication channels with feedback. Massey [1] showed that the feedback capacity is upper bounded by

$$C_{\text{FB}} \leq \lim_{n \rightarrow \infty} \max_{p(x^n || y^{n-1})} \frac{1}{n} I(X^n \rightarrow Y^n),$$

where the definition of directed information $I(X^n \rightarrow Y^n)$ is given in Section II and $p(x^n || y^{n-1}) = \prod_{i=1}^n p(x_i | x^{i-1}, y^{i-1})$ is the causal conditioning notation streamlined by Kramer [2], [3]. This upper bound is tight for a certain class of ergodic channels [4]–[6], paving the road to a computable characterization of feedback capacity; see [7], [8] for examples.

Directed information and its variants also characterize (via multi-letter expressions) the capacity of two-way channels and multiple access channels with feedback [2], [9], the sequential rate distortion function [10], and the rate distortion function with feedforward [11], [12]. In another context, directed information also captures the difference in growth rates of wealth in horse race gambling due to *causal* side information [13]. This provides a natural interpretation of $I(X^n \rightarrow Y^n)$ as the amount of information about Y^n *causally* provided by X^n on the fly. A similar conclusion can be drawn for other engineering and

science problems, in which directed information measures the value of causal side information [14].

In this paper, we extend the notion of directed information to continuous-time random processes. The contribution of this paper is twofold. First, the definition we give for directed information in continuous time is valuable in itself. Just as in the discrete-time setting, directed information in continuous time generalizes mutual information between two stochastic processes. Indeed, when two processes do not have any causal dependence among them, the two notions become identical. Directed information in continuous time is also a generalization of its discrete time counterpart.

Second, we demonstrate the utility of this notion of directed information by generalizing classical results on the relationship between mutual information and causal estimation in continuous time. In particular, we generalize Duncan’s theorem which relates the minimum mean squared error (MMSE) of a target signal based on an observation through an additive white Gaussian channel to directed information between the target signal and the observation. We similarly generalize the Poisson analogue of Duncan’s theorem.

The rest of the paper is organized as follows. Section II is devoted to the definitions of directed information and directed information density in continuous time, which is followed by key properties of continuous-time directed information in Section III. Section IV presents a generalization of Duncan’s theorem, and of its Poisson counterpart, for target signals that depend on the past noise. We conclude with a few remarks in Section V. More details, proofs of the stated results, and additional related results are given in [15].

II. DEFINITION OF DIRECTED INFORMATION IN CONTINUOUS TIME

Let (X^n, Y^n) be a pair of random n -sequences. Directed information (from X^n to Y^n) is defined as

$$I(X^n \rightarrow Y^n) := \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}).$$

Note that unlike mutual information, directed information is asymmetric in its arguments, so $I(X^n \rightarrow Y^n) \neq I(Y^n \rightarrow X^n)$.

For a continuous-time process $\{X_t\}$, let $X_a^b = \{X_s : a \leq s \leq b\}$ denote the process in the interval $[a, b]$ when $a \leq b$ and the empty set otherwise. Let $X_a^{b-} = \{X_s : a \leq s < b\}$

denote the process in the interval $[a, b]$ if $a < b$ and the empty set otherwise. Similarly, let $X_{a+}^b = \{X_s : a < s \leq b\}$ denote the process in the interval $(a, b]$ if $a < b$ and the empty set otherwise. Throughout this section, equalities and inequalities between random objects, when not explicitly indicated, are to be understood in the sure sense (i.e., hold for all sample paths). Functions of random objects are assumed measurable even though not explicitly indicated.

We now develop the notion of directed information between two continuous-time stochastic processes on the time interval $[0, T]$. Let $\mathbf{t} = (t_1, t_2, \dots, t_n)$ denote an n -dimensional vector with components satisfying

$$0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq T. \quad (1)$$

Let $X_0^{T, \mathbf{t}}$ denote the sequence of length $n + 1$ resulting from “chopping up” the continuous-time signal X_0^T into consecutive segments as follows

$$X_0^{T, \mathbf{t}} = (X_0^{t_1-}, X_{t_1}^{t_2-}, \dots, X_{t_{n-1}}^{t_n-}, X_{t_n}^T). \quad (2)$$

Note that each sequence component is a continuous-time stochastic process. Define now

$$\begin{aligned} I_{\mathbf{t}}(X_0^T \rightarrow Y_0^T) &:= I(X_0^{T, \mathbf{t}} \rightarrow Y_0^{T, \mathbf{t}}) \\ &= \left[\sum_{i=1}^n I(Y_{t_i-}^{t_i-}; X_0^{t_i-} | Y_0^{t_{i-1}-}) \right] + I(Y_{t_n}^T; X_0^T | Y_0^{t_n-}), \end{aligned} \quad (3)$$

where on the right side of (3) is the directed information between two sequences of length $n + 1$ defined in previous sections, in (4) we take $t_0 = 0$, and the mutual information terms between two continuous time processes, conditioned on a third, are well defined objects, as developed in [16], [17]. $I_{\mathbf{t}}(X_0^T \rightarrow Y_0^T)$ is monotone in \mathbf{t} in the following sense:

Proposition 1. *If \mathbf{t}' is a refinement of \mathbf{t} , then $I_{\mathbf{t}'}(X_0^T \rightarrow Y_0^T) \leq I_{\mathbf{t}}(X_0^T \rightarrow Y_0^T)$.*

The following definition is now natural:

Definition 1. *Directed information between X_0^T and Y_0^T is defined as*

$$I(X_0^T \rightarrow Y_0^T) := \inf_{\mathbf{t}} I_{\mathbf{t}}(X_0^T \rightarrow Y_0^T), \quad (5)$$

where the infimum is over all n and \mathbf{t} as in (1).

Note, in light of Proposition 1, that

$$I(X_0^T \rightarrow Y_0^T) = \lim_{\varepsilon \rightarrow 0^+} \inf_{\{\mathbf{t}: t_i - t_{i-1} \leq \varepsilon\}} I_{\mathbf{t}}(X_0^T \rightarrow Y_0^T). \quad (6)$$

Directed information can be given an integral representation via the following notion of a density.

Definition 2. For $0 \leq t < T$, define $i_{t+}(X_0^T \rightarrow Y_0^T)$

$$i_{t+}(X_0^T \rightarrow Y_0^T) = \lim_{\delta \rightarrow 0^+} \frac{1}{\delta} I(Y_t^{t+\delta}; X_0^{t+\delta} | Y_0^t), \quad (7)$$

and for $0 < t \leq T$, define $i_{t-}(X_0^T \rightarrow Y_0^T)$ by

$$i_{t-}(X_0^T \rightarrow Y_0^T) = \lim_{\delta \rightarrow 0^+} \frac{1}{\delta} I(Y_{t-\delta}^t; X_0^t | Y_0^{t-\delta}) \quad (8)$$

whenever the limits exist. When $i_{t+}(X_0^T \rightarrow Y_0^T)$ and $i_{t-}(X_0^T \rightarrow Y_0^T)$ exist and are equal, we denote them by $i_t(X_0^T \rightarrow Y_0^T)$, which we refer to as the *directed information density*.

Proposition 2. *If the directed information density $i_t(X_0^T \rightarrow Y_0^T)$ is finite for all $0 < t < T$, then*

$$I(X_0^T \rightarrow Y_0^T) = \int_0^T i_t(X_0^T \rightarrow Y_0^T) dt \quad (9)$$

and, for every $0 < t < T$,

$$\frac{d}{dt} I(X_0^t \rightarrow Y_0^t) = i_t(X_0^T \rightarrow Y_0^T). \quad (10)$$

Example 1. Let $\{B_t\}$ be the standard Brownian motion and let $A \sim N(0, 1)$ be independent of $\{B_t\}$. Let $X_t \equiv A$ for all t and $dY_t = X_t dt + dB_t$. Letting $J(\sigma_X^2, \sigma_N^2) = \frac{1}{2} \ln \frac{\sigma_X^2 + \sigma_N^2}{\sigma_N^2}$ denote the mutual information between a Gaussian random variable of variance σ_X^2 and itself corrupted by an independent Gaussian of variance σ_N^2 , we have for every $t \in [0, T)$

$$I(Y_t^{t+\delta}; X_0^{t+\delta} | Y_0^t) = J\left(\frac{1/t}{1 + 1/t}, \frac{1}{\delta}\right) = \frac{1}{2} \ln \left[1 + \frac{\delta}{t + 1}\right]$$

and, similarly, for every $t \in (0, T]$,

$$\begin{aligned} I(Y_{t-\delta}^t; X_0^t | Y_0^{t-\delta}) &= J\left(\frac{1/(t-\delta)}{1 + 1/(t-\delta)}, \frac{1}{\delta}\right) \\ &= \frac{1}{2} \ln \left[1 + \frac{\delta}{t - \delta + 1}\right]. \end{aligned}$$

Evidently, for all $t \in (0, T)$,

$$\begin{aligned} i_t(X_0^T \rightarrow Y_0^T) &= i_{t+}(X_0^T \rightarrow Y_0^T) \\ &= i_{t-}(X_0^T \rightarrow Y_0^T) \\ &= \lim_{\delta \rightarrow 0} \frac{1}{2\delta} \ln \left[1 + \frac{\delta}{t + 1}\right] \\ &= \lim_{\delta \rightarrow 0} \frac{1}{2\delta} \ln \left[1 + \frac{\delta}{t - \delta + 1}\right] = \frac{1}{2(t + 1)}. \end{aligned}$$

We can now compute the directed information by applying Proposition 2:

$$\begin{aligned} I(X_0^T \rightarrow Y_0^T) &= \int_0^T i_t(X_0^T \rightarrow Y_0^T) dt \\ &= \int_0^T \frac{1}{2(t + 1)} dt = \frac{1}{2} \ln(1 + T). \end{aligned} \quad (11)$$

Note that in this example $I(X_0^T; Y_0^T) = J(1, 1/T) = \frac{1}{2} \ln(1 + T)$ and thus, by (11), we have $I(X_0^T \rightarrow Y_0^T) = I(X_0^T; Y_0^T)$. This equality between mutual and directed information holds in more general situations, as elaborated in the next section.

The directed information we have just defined is between two processes on $[0, T]$. We extend this definition to processes on any other closed and bounded interval, and to the conditional directed information $I(X_0^T \rightarrow Y_0^T | V)$, where V is a random object jointly distributed with (X_0^T, Y_0^T) , in the obvious way.

We now define the notion of directed information between a process on $[0, T)$ and a process on $[0, T]$. Let $X_0^{[\delta], T}$ denote the process on $[0, T]$ formed by shifting X_0^T by δ to the right and filling the gap with 0, i.e.: $X_t^{[\delta]} = X_{t-\delta}$ for $t \in [\delta, T]$ and $X_t^{[\delta]} \equiv 0$ for $t \in [0, \delta)$. Define now

$$\bar{I}(X_0^{T-} \rightarrow Y_0^T) := \limsup_{\delta \rightarrow 0^+} I(X_0^{[\delta], T} \rightarrow Y_0^T) \quad (12)$$

and

$$\underline{I}(X_0^{T-} \rightarrow Y_0^T) := \liminf_{\delta \rightarrow 0^+} I(X_0^{[\delta], T} \rightarrow Y_0^T), \quad (13)$$

where the directed information expressions on the right sides of (12) and (13) are according to the definition we already have for directed information between two processes on $[0, T]$. Finally, define the directed information $I(X_0^{T-} \rightarrow Y_0^T)$ by

$$I(X_0^{T-} \rightarrow Y_0^T) = \lim_{\delta \rightarrow 0^+} I(X_0^{[\delta], T} \rightarrow Y_0^T) \quad (14)$$

whenever the limit exists or, equivalently, when $\bar{I}(X_0^{T-} \rightarrow Y_0^T) = \underline{I}(X_0^{T-} \rightarrow Y_0^T)$. The fifth part of Proposition 3 below provides, among other implications, a regularity condition that suffices to ensure the existence of $I(X_0^{T-} \rightarrow Y_0^T)$. In some senses, $\bar{I}(X_0^{T-} \rightarrow Y_0^T)$, $\underline{I}(X_0^{T-} \rightarrow Y_0^T)$ and, when it exists, $I(X_0^{T-} \rightarrow Y_0^T)$ are continuous-time analogues of $I(X^{n-1} \rightarrow Y^n)$. One such sense is the conservation law provided in Proposition 3.

III. PROPERTIES OF THE DIRECTED INFORMATION IN CONTINUOUS TIME

The following proposition collects some properties of directed information in continuous time:

Proposition 3. *Directed information $I(X_0^T \rightarrow Y_0^T)$ has the following properties:*

- 1) *Monotonicity: $I(X_0^t \rightarrow Y_0^t)$ is monotone nondecreasing in t .*
- 2) *Invariance to time dilation: For $\alpha > 0$, if $\tilde{X}_t = X_{t\alpha}$ and $\tilde{Y}_t = Y_{t\alpha}$ then $I(\tilde{X}_0^{T/\alpha} \rightarrow \tilde{Y}_0^{T/\alpha}) = I(X_0^T \rightarrow Y_0^T)$. More generally, if ϕ is monotone strictly increasing and continuous, and $(\tilde{X}_{\phi(t)}, \tilde{Y}_{\phi(t)}) = (X_t, Y_t)$, then*

$$I(X_0^T \rightarrow Y_0^T) = I(\tilde{X}_{\phi(0)}^{\phi(T)} \rightarrow \tilde{Y}_{\phi(0)}^{\phi(T)}). \quad (15)$$

- 3) *Coincidence of directed and mutual information: If the Markov relation $Y_0^{t-} - X_0^{t-} - X_t^T$ holds for all $0 \leq t \leq T$ then*

$$I(X_0^T \rightarrow Y_0^T) = I(X_0^T; Y_0^T). \quad (16)$$

- 4) *Equivalence between discrete-time and piecewise constancy in continuous-time: Let U^n, V^n be an arbitrarily*

jointly distributed pair of n -tuples and let t_0, t_1, \dots, t_n be a sequence of numbers satisfying $t_0 = 0$, $t_n = T$, and $t_{i-1} < t_i$ for $1 \leq i \leq n$. Let the pair X_0^T, Y_0^T be formed as the piecewise-constant process satisfying

$$(X_t, Y_t) = (U_i, V_i) \quad \text{if } t_{i-1} \leq t < t_i \quad (17)$$

and $(X_T, Y_T) = (U_n, V_n)$. Then

$$I(X_0^T \rightarrow Y_0^T) = I(U^n \rightarrow V^n).$$

- 5) *Conservation law: For all $0 < \delta \leq T$ we have*

$$I(X_0^{\delta-}; Y_0^{\delta-}) + I(X_\delta^T \rightarrow Y_\delta^T | Y_0^{\delta-}) + I(Y_0^{[\delta]T} \rightarrow X_0^T) = I(X_0^T; Y_0^T).$$

In particular,

- a) $\limsup_{\delta \rightarrow 0^+} [I(X_0^{\delta-}; Y_0^{\delta-}) + I(X_\delta^T \rightarrow Y_\delta^T | Y_0^{\delta-})] = I(X_0^T; Y_0^T) - \underline{I}(Y_0^{T-} \rightarrow X_0^T)$.
- b) $\liminf_{\delta \rightarrow 0^+} [I(X_0^{\delta-}; Y_0^{\delta-}) + I(X_\delta^T \rightarrow Y_\delta^T | Y_0^{\delta-})] = I(X_0^T; Y_0^T) - \bar{I}(Y_0^{T-} \rightarrow X_0^T)$.
- c) *If the continuity condition*

$$\lim_{\delta \rightarrow 0^+} [I(X_0^{\delta-}; Y_0^{\delta-}) + I(X_\delta^T \rightarrow Y_\delta^T | Y_0^{\delta-})] = I(X_0^T \rightarrow Y_0^T)$$

holds, then the directed information $I(Y_0^{T-} \rightarrow X_0^T)$ exists and

$$I(X_0^T \rightarrow Y_0^T) + I(Y_0^{T-} \rightarrow X_0^T) = I(X_0^T; Y_0^T). \quad (18)$$

Remarks. 1) The first, second and fourth items in the above proposition present properties that are known to hold for mutual information (i.e., when all the directed information expressions in those items are replaced by the corresponding mutual information), and that follow immediately from the data processing inequality and from the invariance of mutual information to one-to-one transformations of its arguments. That these properties hold also for directed information is not as obvious in view of the fact that directed information is, in general, not invariant to one-to-one transformations nor does it satisfy a data processing inequality in its second argument.

- 2) The third part of the proposition is the natural analogue of the fact that $I(X^n; Y^n) = I(X^n \rightarrow Y^n)$ whenever $Y^i - X^i - X_{i+1}^n$ for all $1 \leq i \leq n$. It covers, in particular, any scenario where X_0^T and Y_0^T are the input and output of any channel of the form $Y_t = g_t(X_0^t, W_0^T)$, where the process W_0^T (which can be thought of as the internal channel noise) is independent of the channel input process X_0^T . To see this note that in this case we (trivially) have $(X_0^{t-}, W_0^T) - X_0^{t-} - X_t^T$ for all $0 \leq t \leq T$, implying $Y_0^{t-} - X_0^{t-} - X_t^T$ since Y_0^{t-} is determined by the pair (X_0^{t-}, W_0^T) .

- 3) Particularizing even further, we get $I(X_0^T \rightarrow Y_0^T) = I(X_0^T; Y_0^T)$ whenever Y_0^T is the result of corrupting

X_0^T with additive noise, i.e., $Y_t = X_t + W_t$ where X_0^T and W_0^T are independent.

- 4) The fifth part of the proposition can be considered the continuous-time analogue of the discrete-time conservation law

$$I(U^n \rightarrow V^n) + I(V^{n-1} \rightarrow U^n) = I(U^n; V^n). \quad (19)$$

It is consistent with, and in fact generalizes, the third part. Indeed, if the Markov relation $Y_0^{t-} - X_0^{t-} - X_t^T$ holds for all $0 \leq t \leq T$ then our definition of directed information is readily seen to imply that $I(Y_0^{[\delta]T} \rightarrow X_0^T) = 0$ for all $\delta > 0$ and therefore that $I(Y_0^{T-} \rightarrow X_0^T)$ exists and equals zero. Thus (18) in this case collapses to (16).

IV. DIRECTED INFORMATION AND CAUSAL ESTIMATION

A. The Gaussian Channel

In [18], Duncan discovered the following fundamental relationship between the minimum mean squared error in causal estimation of a target signal corrupted by an additive white Gaussian noise in continuous time and the mutual information between the clean and noise-corrupted signal:

Theorem 1 (Duncan [18]). *Let X_0^T be a signal of finite average power $\int_0^T E[X_t^2]dt < \infty$, independent of the standard Brownian motion $\{B_t\}$, and let Y_0^T satisfy $dY_t = X_t dt + dB_t$. Then*

$$\frac{1}{2} \int_0^T E[(X_t - E[X_t|Y_0^t])^2] dt = I(X_0^T; Y_0^T). \quad (20)$$

A remarkable aspect of Duncan's theorem is that the relationship (20) holds regardless of the distribution of X_0^T . Among its ramifications is the invariance of the causal MMSE (minimum mean squared error) to the flow of time, or indeed to any way of reordering time [19], [20].

A key stipulation in Duncan's theorem is the independence between the noise-free signal X_0^T and the channel noise $\{B_t\}$, which excludes scenarios in which the evolution of X_t is affected by the channel noise, as is often the case in signal processing (e.g., target tracking) and in communications (e.g., in the presence of feedback). Indeed, (20) does not hold in the absence of such a stipulation.

As an extreme example, consider the case where the channel input is simply the channel output with some delay, i.e., $X_{t+\varepsilon} = Y_t$ for some $\varepsilon > 0$ (and say $X_t \equiv 0$ for $t \in [0, \varepsilon)$). In this case the causal MMSE on the left side of (20) is clearly 0, while the mutual information on its right side is infinite. On the other hand, in this case the directed information $I(X_0^T \rightarrow Y_0^T) = 0$, as can be seen by noting that $I_t(X_0^T \rightarrow Y_0^T) = 0$ for all t satisfying $\max_i t_i - t_{i-1} \leq \varepsilon$ (since for such t , $X_0^{t_i-}$ is determined by $Y_0^{t_{i-1}-}$ for all i).

The third comment following Proposition 3 implies that Theorem 1 could equivalently be stated with $I(X_0^T; Y_0^T)$ on the right side of (20) replaced by $I(X_0^T \rightarrow Y_0^T)$. Further, such a modified equality would be valid in the extreme example

just given. This is no coincidence, and is a consequence of the following result that generalizes Duncan's theorem.

Theorem 2. *Let $\{B_t\}$ be a standard Brownian motion, let $\{W_t\}$ be independent of $\{B_t\}$, and let $\{(X_t, Y_t)\}$ satisfy*

$$X_t = a_t(X_0^{t-\delta}, Y_0^{t-\delta}, W_0^T) \quad (21)$$

(for deterministic mappings a_t and some $\delta > 0$) such that $\{X_t\}$ has finite average power $\int_0^T E[X_t^2]dt < \infty$ and

$$dY_t = X_t dt + dB_t. \quad (22)$$

Then

$$\frac{1}{2} \int_0^T E[(X_t - E[X_t|Y_0^t])^2] dt = I(X_0^T \rightarrow Y_0^T). \quad (23)$$

Note that since, in general, $I(X_0^T \rightarrow Y_0^T)$ is not invariant to the direction of the flow of time, Theorem 2 implies, as should be expected, that neither is the causal MMSE for processes evolving in the generality afforded by (21) and (22).

Proof of Theorem 2: For every $t \geq 0$ and $\varepsilon \leq \delta$,

$$\frac{1}{2} \int_t^{t+\varepsilon-} E[(X_s - E[X_s|Y_0^s])^2] ds \quad (24)$$

$$= \frac{1}{2} \int_t^{t+\varepsilon-} E[E[(X_s - E[X_s|Y_0^s])^2|Y_0^{t-}]] ds \quad (25)$$

$$= E\left[\frac{1}{2} \int_t^{t+\varepsilon-} E[(X_s - E[X_s|Y_0^s])^2|Y_0^{t-}] ds\right] \quad (26)$$

$$= \int \left[\frac{1}{2} \int_t^{t+\varepsilon-} E[(X_s - E[X_s|Y_0^s])^2|y_0^{t-}] ds\right] dP(y_0^{t-}) \quad (27)$$

$$\stackrel{(a)}{=} \int I(X_t^{t+\varepsilon-}; Y_t^{t+\varepsilon-}|y_0^{t-}) dP(y_0^{t-}) \quad (28)$$

$$= I(X_t^{t+\varepsilon-}; Y_t^{t+\varepsilon-}|Y_0^{t-}), \quad (29)$$

where (a) is due to the following: since $Y_0^{t-}, X_0^{t-}, W_0^T$ is independent of $B_t^{t+\varepsilon-}$, and $X_t^{t+\varepsilon-}$ is determined by $Y_0^{t-}, X_0^{t-}, W_0^T$, it follows that Y_0^{t-}, X_0^{t-} is independent of $B_t^{t+\varepsilon-}$. Thus, (a) is nothing but an application of Duncan's theorem on the conditional distribution of $(X_t^{t+\varepsilon-}, B_t^{t+\varepsilon-})$ given y_0^{t-} to get equality between the integrand of (27) and that of (28). Fixing now an arbitrary t that satisfies $\max_i t_i - t_{i-1} \leq \delta$ gives

$$\frac{1}{2} \int_0^T E[(X_t - E[X_t|Y_0^t])^2] dt \quad (30)$$

$$= \sum_{i=1}^n \frac{1}{2} \int_{t_{i-1}}^{t_i-} E[(X_t - E[X_t|Y_0^t])^2] dt \quad (31)$$

$$+ \frac{1}{2} \int_{t_n}^T E[(X_t - E[X_t|Y_0^t])^2] dt \quad (32)$$

$$\stackrel{(a)}{=} \left[\sum_{i=1}^n I(Y_{t_{i-1}-}^{t_i-}; X_0^{t_i-}|Y_0^{t_{i-1}-}) \right] + I(Y_{t_n}^T; X_0^T|Y_0^{t_n-}) \quad (33)$$

$$= I_t(X_0^T \rightarrow Y_0^T), \quad (34)$$

where (a) follows by applying (29) on each of the summands in (32) with the associations $t_{i-1} \leftrightarrow t$ and $t_i - t_{i-1} \leftrightarrow \varepsilon$. Finally, since (34) holds for arbitrary t satisfying $\max_i t_i - t_{i-1} \leq \delta$, by (6) we obtain

$$\frac{1}{2} \int_0^T E [(X_t - E[X_t|Y_0^t])^2] dt = I(X_0^T \rightarrow Y_0^T). \quad (35)$$

The evolution of the noise-free process in the above theorem (equation (21)), which assumes a nonzero delay in the feedback loop, is the standard evolution arising in signal processing and communications (cf., e.g., [21]). The result, however, can be extended to accommodate a more general model with zero delay, as introduced and developed in [22], by combining our arguments above with those used in [22].

B. The Poisson Channel

The following result can be considered an analogue of Duncan's theorem for the case of Poisson noise.

Theorem 3 ([23]). *Let Y_0^T be a doubly stochastic Poisson process and let X_0^T be its intensity process. Then, provided $E \int_0^T |X_t \log X_t| dt < \infty$,*

$$\int_0^T E [\phi(X_t) - \phi(E[X_t|Y_0^t])] dt = I(X_0^T; Y_0^T), \quad (36)$$

where $\phi(\alpha) = \alpha \log \alpha$.

It is easy to verify that the condition stipulated in the third item of Proposition 3 holds when Y_0^T is a doubly stochastic Poisson process and X_0^T is its intensity process. Thus, the above theorem could equivalently be stated with directed rather than mutual information on the right hand side of (36). Indeed, with continuous-time directed information replacing mutual information, this relationship remains true in much wider generality, as the next theorem shows. In the statement of the theorem, we use the notions of a point process and its predictable intensity, as developed in detail in [24, Chapter II].

Theorem 4. *Let Y_t be a point process and let X_t be its \mathcal{F}_t^Y -predictable intensity, where $\mathcal{F}_t^Y = \sigma(Y_0^t)$ (the σ -field generated by Y_0^t). Then, provided $E \int_0^T |X_t \log X_t| dt < \infty$,*

$$\int_0^T E [\phi(X_t) - \phi(E[X_t|Y_0^t])] dt = I(X_0^T \rightarrow Y_0^T). \quad (37)$$

V. CONCLUDING REMARKS

The machinery developed here for directed information between continuous-time stochastic processes appears to have several powerful applications. In [15], we provide further substantiation of the significance of directed information in continuous time through connections to Kalman–Bucy filtering theory. We also show that this notion of directed information emerges naturally in characterizing the capacity of a wide class of continuous-time channels with feedback.

REFERENCES

- [1] J. Massey, "Causality, feedback and directed information," *Proc. Int. Symp. Inf. Theory Applic. (ISITA-90)*, pp. 303–305, 1990.
- [2] G. Kramer, "Directed information for channels with feedback," Ph.D. Dissertation, Swiss Federal Institute of Technology (ETH) Zurich, 1998.
- [3] —, "Feedback strategies for a class of two-user multiple-access channels," *IEEE Trans. Inf. Theory*, vol. IT-45, no. 6, pp. 2054–2059, 1999.
- [4] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *Information Theory, IEEE Transactions on*, vol. 55, no. 1, pp. 323–349, Jan. 2009.
- [5] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory*, vol. 25, pp. 1488–1499, April, 2008.
- [6] H. Permuter, T. Weissman, and A. Goldsmith, "Capacity of finite-state channels with time-invariant deterministic feedback," July 2006.
- [7] Y. H. Kim, "Feedback capacity of stationary Gaussian channels," 2006, submitted to *IEEE Trans. Inform. Theory*. Available at arxiv.org/pdf/cs.IT/0602091.
- [8] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-54, no. 7, pp. 3150–3165, July 2008.
- [9] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 4–21, 2003.
- [10] S. Tatikonda, "Control under communication constraints," *Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA*, 2000.
- [11] S. Pradhan, "Source coding with feedforward: Gaussian sources," in *Proceedings 2004 International Symposium on Information Theory*, 2004, p. 212.
- [12] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion function for general sources," in *IEEE Information theory workshop (ITW)*, 2004.
- [13] H. H. Permuter, Y.-H. Kim, and T. Weissman, "On directed information and gambling," in *Proc. International Symposium on Information Theory (ISIT)*, Toronto, Canada, 2008.
- [14] Y.-H. Kim, H. H. Permuter, and T. Weissman, "An interpretation of directed information in gambling, portfolio theory and data compression," 2009, in preparation.
- [15] —, "Directed information in continuous time," 2009, in preparation.
- [16] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *Information Theory, IRE Transactions on*, vol. 2, no. 4, pp. 102–108, December 1956.
- [17] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Moskva: Izv. Akad. Nauk, 1960, in Russian.
- [18] T. E. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.*, vol. 19, pp. 215–220, 1970.
- [19] A. Cohen, N. Merhav, and T. Weissman, "Scanning and sequential decision making for multidimensional data, Part II: noisy data," *IEEE Trans. Inf. Theory*, vol. IT-54, no. 12, pp. 5609–5631, Dec. 2008.
- [20] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. IT-51, no. 4, pp. 1261–1283, Apr. 2005.
- [21] T. T. Kadota, M. Zakai, and J. Ziv, "Capacity of a continuous memoryless channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-17, pp. 372–378, 1971.
- [22] —, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. IT-17, pp. 368–371, 1971.
- [23] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes II: Applications*, 2nd ed. Springer, 2001.
- [24] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*, 1st ed. New York: Springer-Verlag, 1981.