Analysis of mutual information for informative forecasting using mobile sensors

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ICARCV.2010.5707788">http://dx.doi.org/10.1109/ICARCV.2010.5707788</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Tue Feb 12 08:51:30 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/65860">http://hdl.handle.net/1721.1/65860</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Analysis of Mutual Information for Informative Forecasting Using Mobile Sensors

Han-Lim Choi  
Div. of Aerospace Engineering  
KAIST (Korea Advanced Institute of Science and Technology)  
Daejeon, Rep. of Korea  
ahanlmc@kaist.ac.kr

Jonathan P. How  
Dept. of Aeronautics and Astronautics  
Massachusetts Institute of Technology  
Cambridge, MA, USA  
jhow@mit.edu

Abstract—This paper presents several analysis of mutual information that is often used to define the objective function for trajectory planning (and scheduling) of sensor networks, when the goal is to improve the forecast accuracy of some quantities of interest. The approach extends the present author’s prior work in order to consider more general notion of verification entities and to enable more robust decision with potential uncertainty in the mission specifications. The expression of mutual information for windowed forecasting, in which the verification entities are defined by a finite time window instead of a single time instance, is derived and quantified without adding significant computational cost. It is also demonstrated that the sensitivity of mutual information to the variation of verification time can be calculated in the same process of computing the mutual information. Simple numerical examples are presented for preliminary validation of the applicability of the proposed analysis.

Index Terms—Mutual Information, Forecasting, Mobile Sensor Networks

I. INTRODUCTION

One key decision for sensor networks is to design plans for maneuvering/locating/scheduling sensing resources in order to gather information from the environment. In this decision making, the plans are often generated to reduce uncertainty in some quantity of interest (e.g., position and velocity of targets [1]–[10], the pose of the sensor platform itself, the forecast of weather over some region of interest [11]–[14], physical quantities under the ocean [15,16], or the distribution of radioactive materials [17]) at some point in time – called the verification time.

Among these, the informative forecasting problem specifically focuses on the case when the verification time is further in the future than the mission horizon. For example, adaptive sampling in the context of numerical weather prediction considers design of sensor networks deployed in the near future (e.g., in 24 hours) while the goal is to improve forecast in the far future (e.g., 3-5 days later). The authors’ prior work [10]–[12,18] have presented methods to efficiently and correctly quantify the value of information in this context of informative forecasting. The mutual information has been used to define the uncertainty reduction of the quantity of interest (in the far future), and information-theoretic properties of mutual information are exploited to yield alternative formula for mutual information to a straightforward extension of the state-of-the-art. In particular, the smoother approach in [18] efficiently and correctly quantifies the information value for continuous-time planning problems together with the rate of information accumulation.

This paper extends the approach in [18] in that a more general notion of verification quantities are introduced. For some applications, it might make more sense to reduce uncertainty in the entities of interest over some finite window of time instead of a single particular time instance (for example, weather forecast over the weekend). In addition, in complicated missions of sensor networks, sometimes the mission characterization can be uncertain so that the question of which entities should be considered as the figure-of-merit might be an important issue to consider. The smoother form in [18] cannot directly be used for the windowed forecasting case, because the mutual information between two continuous random processes (as opposed to one finite-dimensional random vector and one random process) needs to be calculated. This paper presents a formula for the mutual information for this windowed forecasting that is indeed quite similar to the form in [18]. Also, the expression of sensitivity of the mutual information to variation of the verification time is derived, and is shown to be quantifiable in the process of computing the mutual information.

II. INFORMATIVE FORECASTING PROBLEMS

A. Continuous-Time Linear System Model

Consider the dynamics of objects/environment with a finite dimensional state vector $X_t \in \mathbb{R}^n$ that is described by the following linear (time-varying) system:

$$\dot{X}_t = A(t)X_t + W_t \quad (1)$$

where $W_t \in \mathbb{R}^n$ is a zero-mean Gaussian process noise with $\mathbb{E}[W_tW_t'] = \Sigma_W \delta(t-s)$, $\Sigma_W \succ 0$, which is independent of $X_t$. The prime ’ denotes the transpose of a matrix. The initial condition of the state, $X_0$, is normally distributed as $X_0 \sim \mathcal{N}(\mu_0, P_0)$, $P_0 \succ 0$.

The system (1) is observed by sensors with additive Gaussian noise and admits the following measurement model for $Z_t \in \mathbb{R}^m$:

$$Z_t = C(t)X_t + N_t \quad (2)$$

where $N_t \in \mathbb{R}^m$ is zero-mean Gaussian with $\mathbb{E}[N_tN_t'] = \Sigma_N \delta(t-s)$, $\Sigma_N \succ 0$, which is independent of $X_t$ and $W_s$, $\forall s$. 

978-1-4244-7815-6/10/$26.00 \copyright 2010 IEEE 1245 ICARCV2010
Also, a measurement history over the time window \([t_1,t_2]\) is defined as
\[
\mathcal{Z}_{[t_1,t_2]} = \{ Z_t : t \in [t_1,t_2] \}. \tag{3}
\]
The verification variables are a subset of the state variables that define the performance measure:
\[
V_t = M_V X_t \in \mathbb{R}^p \tag{4}
\]
where \(M_V \in \{0,1\}^{p \times n}, p < n\) with every row-sum of \(M_V\) being unity. Although this work is specifically interested in the case entries of \(M_V\) are zero or one, the results can be easily extended to a general \(M_V \in \mathbb{R}^{p \times n}\). Without loss of generality, this work assumes that
\[
M_V = [0_{p \times (n-p)} \quad I_p]
\]
for notational simplicity in the analysis in the later sections.

B. Formulations and Challenges

The planning problem for informative forecasting addresses design of sensing paths for mobile sensors (or equivalently scheduling sequence of distributed sensors) over some given time window \([0, \tau]\) in the near future to reduce the uncertainty in some verification entities in the far future. One popular notion to quantify the expected amount of uncertainty reduction of the verification entities by the measurements taken along the sensing paths is the mutual information (between the verification entities and the measurement sequence).

1) Informative Point-Wise Forecasting: One case of interest is when the verification entities are the verification variables, i.e., some subset of state variables defined by (4), at some fixed verification time, \(T > \tau\). In this case, the informative forecasting problem can be written as the following optimization:
\[
\max_{\mathcal{Z}_{[0,\tau]}} \mathcal{I}(V_T; \mathcal{Z}_{[0,\tau]}) \tag{IPF}
\]
where \(\mathcal{I}(Y_1; Y_2)\) represents the mutual information between two random quantities \(Y_1\) and \(Y_2\) (e.g., random variables, random processes, random functions). In other words, (IPF) finds the (continuous) measurement sequence over \([0, \tau]\) that represents the largest mutual information between the verification variables at \(T\). Previous work by the present authors [18] has developed methods to efficiently and correctly quantify the mutual information for this first type of informative forecasting. An example of this type of planning is to design sensing paths for UAV sensor platforms operating in the near future (e.g., within 24 hours) to improve weather forecast over some verification region at a given time of interest in the future (e.g., in 5 days).

2) Informative Windowed Forecasting: This paper also considers a more generalized version of the informative forecasting problem where the entities of interest are the verification variables over some time window \([T_i, T_f]\) rather than a single time instance \(T\). This generalized problem can be written as:
\[
\max_{\mathcal{Z}_{[0,\tau]}} \mathcal{I}(\mathcal{V}_{[T_i, T_f]}; \mathcal{Z}_{[0,\tau]}) \tag{IWF}
\]
where
\[
\mathcal{V}_{[T_i, T_f]} \triangleq \{ V_t : t \in [T_i, T_f] \}.
\]
This generalization allows for handling more diverse types of sensing missions like improving weather forecast over the coming weekend, better predicting behaviors of some targets of interest between 9am and noon, and so on. In addition, the formulation in (IWF) can also be used to address more robust decision making for (IWF), specifically with setting \(T_f = T_i + \epsilon\) with some small \(\epsilon\). For sensor networks operating in dynamic environments, the mission specifications might be unclear and there might be some uncertainty in when the verification time should be; then, (IWF) can be posed to account for such uncertainty in the verification time.

However, the generalization (IWF) can introduce challenges in quantifying the mutual information in the objective function. For the discrete-time representation, in which \(\mathcal{Z}_{[0,\tau]}\) and \(\mathcal{V}_{[T_i, T_f]}\) can be represented by finite dimensional random vectors, the similar generalization of verification entities as in (IWF) does not incur significant difficulty other than computational cost due to the increased dimension of the verification entities. Thus, quantification and optimization methods developed for the discrete-time counterpart of (IPF) can trivially extended for the discrete-time counterpart of (IWF).

But, for the continuous-time representation in this work, the objective term in (IWF) is a mutual information between two continuous-time random processes, while that in (IPF) is a mutual information between a finite-dimensional random vector and a continuous-time random process. The difficulty is caused by the possible need to calculating an entropy of a continuous random process. Although there have been researches on the calculation of entropy of a continuous random process [19,20], these approaches were to estimate the entropy value from experimentally-obtained time series, and thus are not suitable for quantifying the information by a future measurement that is not taken yet at the decision time.

As the authors presented in [18], the mutual information in (IPF), \(\mathcal{I}(V_T; \mathcal{Z}_{[0,\tau]}),\) can be expressed in terms of unconditional and conditional entropies of finite-dimensional random vectors, exploiting the popular conditional independence between the future and the past given the present. For linear Gaussian cases, these entropy terms are represented by functions of covariance matrices. Inspired by this observation, this paper provides an expression for the mutual information in (IWF) as a function of covariance matrices for some finite-dimensional random vectors in section III-B.

3) Sensitivity Analysis: In addition to this new quantification, this paper will also investigate the sensitivity of the mutual information in (IPF) to the verification time \(T\). This could allow for more robust decision of information forecasting when there is uncertainty in the mission specification, by trying to maximize the information while maintaining (or minimizing) a certain level of sensitivity measure. In addition, for large-scale decisions, sensitivity calculation can facilitate easy analysis of solution variability without repeatedly solving complicated problems.
III. MAIN RESULTS

A. Summary of Mutual Information for Point-Wise Forecasting

Exploiting conditional independence, the authors have proposed an expression for the mutual information, \( I(V_T; Z_{[0,T]}), \) as the difference between the unconditioned and the conditioned mutual information for a filtering problem [18]:

\[
I(V_T; Z_{[0,T]}) = I(X_T; Z_{[0,T]}) - I(X_T; Z_{[0,T]}|V_T) \tag{5}
\]

With (5), the smoother form of the mutual information for forecasting is derived as

\[
I(V_T; Z_{[0,T]}) = I(X_T; Z_{[0,T]}) - I(X_T; Z_{[0,T]}|V_T)
= \mathcal{J}_0(\tau) - \frac{1}{2} \text{ldet}(I + Q_X(\tau) \Delta_S(\tau)) \tag{6}
\]

with \( \mathcal{J}_0 \triangleq \frac{1}{2} \text{ldet} S_{X|V} - \frac{1}{2} \text{ldet} S_X \) and \( \Delta_S \triangleq S_{X|V} - S_X, \) where \( \text{ldet} \) stands for log det of a positive definite matrix. The matrices \( S_X(\tau) \triangleq \text{Cov}^{-1}(X_T), \) \( S_{X|V}(\tau) \triangleq \text{Cov}^{-1}(X_T|V_T), \) and \( Q_X(\tau) \triangleq \text{Cov}(X_T|Z_{[0,T]}), \) are determined by the following matrix differential equations:

\[
\begin{align*}
\dot{S}_X &= -S_X A' S_X - S_X \Sigma_W S_X \tag{7} \\
\dot{S}_{X|V} &= -S_{X|V}(A + \Sigma_W S_X) - (A + \Sigma_W S_X)' S_{X|V} \tag{8} \\
\dot{Q}_X &= A Q_X + Q_X A' + \Sigma_W - Q_X C^{-1} \Sigma_N C Q_X \tag{9}
\end{align*}
\]

with initial conditions \( S_X(0) = P_0^{-1}, \) \( S_{X|V}(0) = P_{0|V}, \) and \( Q_X(0) = P_0; \) \( P_{0|V} \triangleq \text{Cov}(X_0|V_T) > 0 \) can be calculated in advance by a fixed-point smoothing process, which is equivalent to:

\[
P_{0|V} = P_0 - P_0 \Phi(T_0,0) M_V [M_V P_X(T) M_V]^{-1} M_V \Phi(T_0,0) P_0 \tag{10}
\]

where \( \Phi(T_1,T_2) \) is the transition matrix for the linear system in (1) from \( t_2 \) to \( t_1, \) which becomes \( e^{A(t_1-t_2)} \) for the time-invariance case.

It was demonstrated in [18] that the smoother form is preferred to the filter form in terms of the computational efficiency and accessibility of the on-the-fly knowledge of the accumulated information. For comparison, the filter form is given by:

\[
I(V_T; Z_{[0,T]}) = \frac{1}{2} \text{ldet}(M_V P_X(T) M_V') - \frac{1}{2} \text{ldet}(M_V Q_X(T) M_V') \tag{11}
\]

where \( P_X(T) \triangleq \text{Cov}(X_T) \) and \( Q_X(T) \triangleq \text{Cov}(X_T|Z_{[0,T]}), \) are obtained by

\[
\begin{align*}
\dot{P}_X(t) &= A(t) P_X(t) + P_X(t) A'(t) + \Sigma_W \\
\dot{Q}_X &= AQ_X + Q_X A' + \Sigma_W - \mathbb{I}_{[0,\tau]} Q_X C^{-1} \Sigma_N C Q_X
\end{align*}
\]

with initial conditions \( P_X(0) = Q_X(0) = P_0, \) \( \mathbb{I}_{[0,\tau]}(t) : \mathbb{R}_+ \to \{0,1\} \) is the indicator function that is unity for \( t \in [0,\tau] \) and zero elsewhere. Note that for the filter form, the Riccati equation (of \( Q_X \)) needs to be integrated over \([0,T]\), while it is integrated only over \([0,\tau]\) for the smoother form; this yields computational saving of a factor of \( T/\tau \) of the use of the smoother form for informative planning problems.

B. Mutual Information for Windowed Forecasting

First note that the mutual information for (IWF) cannot be readily quantified in the filtering framework that computes the difference between the prior and posterior entropies of the entities of interest. In (IWF), the verification entity is a continuous-time random process \( V_{[T_1,T_2]} \) and there is no direct way to calculate the entropy of some Gaussian random process over an arbitrary time window without the actual time series of that signal. For the informative forecasting problem in this work, such time series are not available at the decision time, and the decision should be made based on the statistics of the random quantities rather than their actual values.

The similar conditional independence exploited to derive the smoother form for (IPF) can also be used for (IWF), because the verification variables of a future time window \([T_1,T_2]\) is conditionally independent of the (past) measurement sequence \( Z_{[0,T]} \), conditioned on the (current) state variables \( X_T \). Thus, we have:

\[
I(V_{[T_1,T_2]}; Z_{[0,T]}) = I(X_T; Z_{[0,T]}) - I(X_T; Z_{[0,T]}|V_{[T_1,T_2]}), \tag{12}
\]

which is derived using the fact that \( I(V_{[T_1,T_2]}; Z_{[0,T]}|X_T) = 0 \) due to the conditional independence. Notice that the first term in the left-hand side is identical to that in the expression for (IPF). The second term represents the difference between two conditional entropies, \( H(X_T|V_{[T_1,T_2]}|Z_{[0,T]}) \) and \( H(X_T|V_{[T_1,T_2]}, Z_{[0,T]}). \) Since the conditional distribution of a Gaussian vector conditioned on some Gaussian random process is still Gaussian, these two entropy expressions can be represented by log det of the corresponding covariance matrices:

\[
I(X_T; Z_{[0,T]}|V_{[T_1,T_2]}) = \frac{1}{2} (\text{ldet} P_X|V|) - \text{ldet} Q_X|V| \tag{13}
\]

where \( P_X|V| \triangleq \text{Cov}(X_T|V_{[T_1,T_2]}) \) and \( Q_X|V| \triangleq \text{Cov}(X_T|V_{[T_1,T_2]}, Z_{[0,T]}). \) Notice that the smoother form for (IPF) utilized the (symmetric) two-filter approach to fixed-interval smoothing in [21] to express the conditional covariance \( Q_X|V| \) in terms of \( P_X(\tau), Q_X(\tau), \) and \( P_X|V| \) of \( X_T \) in the context, minus the double-counted information from the underlying dynamics, i.e., \( P_X^{-1}, \) thus yielding \( Q_X^{-1}|V| = Q_X^{-1}(\tau) + P_X^{-1}|V| - P_X^{-1}(\tau). \) Notice that this key insight in fixed-interval smoothing still holds even in case the future measurement (of \( V_i \) is taken over a finite window \([T_1,T_2]). \) Thus, we have

\[
Q_X^{-1}|V| = Q_X^{-1}(\tau) + P_X^{-1}|V| - P_X^{-1}(\tau),
\]

and therefore the only term that does not appear in (IPF) is \( P_X|V \) (or equivalently, \( S_X|V \equiv P_X^{-1}|V \)). Notice that the dynamics of \( S_X|V \) needs to be integrated only over \([0,\tau]\); within this time period, its dynamics has the same form as that of
state variables, i.e., \( M \).

Remark 1 In case the verification variables are the whole state variables, i.e., \( M = I_n \), the mutual information for (IWF) is reduced to that for (IPF):

\[
\mathcal{I}(X_{[T,T_f]}, Z_{[0,\tau]}) = \mathcal{I}(X_T; Z_{[0,\tau]})
\]

where \( X_{[T,T_f]} = \{ X_t : t \in [T, T_f] \} \). This can be shown as follows. The state history over \([T_i, T_f]\) can be decomposed as \( X_{[T_i, T_f]} = X_T \cup X_{(T_i, T_f)} \). Notice that for any \( X_t, t > T_f \), it is conditionally independent of \( Z_{[0,\tau]} \) conditioned on \( X_T \); thus, \( \mathcal{I}(X_{(T_i, T_f)}; Z_{[0,\tau]} | X_T) = 0 \). Together with the chain rule of mutual information [22], this yields

\[
\mathcal{I}(X_{[T,T_f]}, Z_{[0,\tau]}) = \mathcal{I}(X_T; Z_{[0,\tau]}) - \mathcal{I}(X_{(T_i, T_f)}; Z_{[0,\tau]} | X_T)
\]

\[
= \mathcal{I}(X_T; Z_{[0,\tau]}).
\]

1) Calculation of Initial Conditioned Covariance \( P_{0|V} \):

The key difference between the mutual information for the windowed forecasting is that the initial conditional covariance \( P_{0|V} \) needs to be calculated conditioned on a random process rather than a finite-dimensional random vector. A typical way of calculating this type of initial covariance is to pose a fixed-point smoothing problem, in which the fixed-point is rather than a finite-dimensional random vector. A typical way to transcribe the smoothing problem into a filtering problem by augmenting the model is identical to (4) that does not contain any noise in the measurement process. This prevents direct implementation of the Riccati equation in the conventional Kalman-Bucy filter in this case. Alternatively, the modification technique presented by [23] can be utilized for this case.

Note that continuous measurement of a certain portion of the state over some finite period of time without noise is equivalent to measurement of that quantity at the initial time and the derivative of that quantity during the period. Based on this observation, [23] derived the expression of the Kalman-Bucy filter in case there is no measurement noise. A popular approach to fixed-point smoothing is to transcribe the smoothing problem into a filtering problem by augmenting the initial state value to the state variables. Let define the augmented state \( Y_t = [X_0, \dot{X}_0]^T \). For notational convenience, \( X_t \) is decomposed into two parts - the verification variables \( V_t \) and the other variables \( V_t' \). Then, the dynamics of the augmented state can be written as:

\[
Y_t \triangleq \frac{d}{dt} \begin{bmatrix} X_0 \\ V_t' \\ \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & A_{11} & A_{12} \\ 0 & A_{21} & A_{22} \end{bmatrix} Y_t + \begin{bmatrix} 0 \\ W_{t,1} \\ W_{t,2} \end{bmatrix}
\]

where the process noise terms are independent, white, Gaussian, with covariance of \( \Sigma_1 \) and \( \Sigma_2 \) such that \( \Sigma_W = \text{diag}(\Sigma_1, \Sigma_2) \). The goal of the smoothing is to find the covariance matrix for \( X_0 \) after incorporating all the (fictitious) measurement over the window \([0, T_f] \). Notice that the only measurement within this window is perfect sensing of the verification variables \( V_t \) over the time window \([T_i, T_f]\).

Bryson [23] pointed out that a perfect measurement of some quantity over some finite time window is equivalent to a perfect discrete-time measurement of that quantity at the initial time, i.e., \( V_{T_i} \) in our formulation, and the time derivative of that quantity over the remaining window, i.e., \( \dot{V}_t \) over \([T_i, T_f]\).

In order to incorporate these measurements, the covariance for the augmented state needs to be propagated using the following Lyapunov equation:

\[
\dot{P}_{aug} = A_{aug} P_{aug} + P_{aug} A_{aug}' + \text{diag}(0_{n \times n}, \Sigma_W)
\]

from the initial condition \( P_{aug}(0) = P_0 \otimes 1_{2 \times 2} \) where \( \otimes \) denotes the Kronecker product. After integrating this equation until \( T_i \), we need to incorporate a discrete-time measurement of \( V_{T_i} \) without any noise. The updated covariance becomes:

\[
P_{aug|V_{T_i}(T_i)} = \begin{bmatrix} P_{11}(T_i) & 0_{(n-2p) \times p} \\ 0_{p \times (n-2p)} & 0_{p \times p} \end{bmatrix}
\]

where

\[
P_{11}(T_i) = \begin{bmatrix} P_{X_0,V(T_i)} \\ P_{V',V}(T_i) \end{bmatrix} P_V^{-1}(T_i) \begin{bmatrix} P_{X_0,V(T_i)}' \\ P_{V',V}(T_i) \end{bmatrix}'
\]

Then, the continuous measurement of \( \dot{V}_t, t \in [T_i, T_f] \) follows; [23] showed that under the assumption of the positive definiteness of \( \Sigma_W \) the covariance propagation through this measurement is represented by the following Riccati equation for \( P_{11} \in \mathbb{R}^{(2n-p) \times (2n-p)} \), while other entries of \( P_{aug} \) remain all zero.

\[
\dot{P}_{11} = \begin{bmatrix} 0 & 0 \\ 0 & A_{11} \end{bmatrix} P_{11} + P_{11} \begin{bmatrix} 0 & 0 \\ 0 & A_{11}' \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_1 \end{bmatrix} - P_{11} \begin{bmatrix} 0 & 0 \\ 0 & A_{21}' \Sigma_2^{-1} A_{21} \end{bmatrix}.
\]

After integrating this equation until \( T_f \), the first \( n \times n \) submatrix of \( P_{11} \) becomes the conditional initial covariance conditioned on the fictitious measurement of \( V_t \) over time window \([T_i, T_f]\). Simply integrating the equations with this initial condition over \([0, \tau]\), the mutual information value for (IWF) is obtained.
C. Sensitivity Analysis of Mutual Information

The verification time might be sometimes unclear, and taking this into account can be useful for more robust decision making for informative forecasting. Thus, the sensitivity measure of interest in this section is the variation of the mutual information value with respect to the verification time. The analysis will focus on the (IPF) case while putting forward extension to (IWF) as future work.

Consider the partial derivative of mutual information in (6):

\[
\frac{\partial}{\partial \tau} \mathcal{I}(V_T; Z_{[0,\tau]}) = \frac{\partial}{\partial \tau} \mathcal{I}(X; Z_{[0,\tau]}|V_T) = -\mathcal{I}(X; Z_{[0,\tau]}|V_T) \\
= -\frac{\partial}{\partial \tau} \mathcal{I}(X; Z_{[0,\tau]}|V_T) - \mathcal{I}(X; Z_{[0,\tau]}|V_T) \\
= -\frac{\partial}{\partial \tau} \mathcal{I}(X; Z_{[0,\tau]}|V_T) \\
= -\frac{\partial}{\partial \tau} \left[ \frac{1}{2} \text{ldet} P_X|V(\tau) - \frac{1}{2} \text{ldet} Q_X|V(\tau) \right] \\
= \frac{1}{2} \text{tr} \left\{ \left[ S_X^{-1}(\tau) - (S_X(V(\tau)) + Q_X(\tau) - S_X(\tau))^{-1} \right] \\
\times \frac{\partial}{\partial \tau} S_X(\tau) \right\}
\]

Observe that all the terms other than \( \frac{\partial}{\partial \tau} (S_X(V(\tau))) \) are computed to calculate the mutual information value. Now focus on the calculation of the partial derivative of \( S_X(\tau) \) with respect to \( \tau \):

\[
\frac{\partial}{\partial \tau} S_X(\tau) = -\frac{\partial}{\partial \tau} \left( \frac{\partial}{\partial \tau} S_X(V(\tau)) \right) S_X(\tau).
\]

Since \( P_X(V(\tau)) \) is the conditional covariance of the state at \( \tau \), conditioned on \( V_T \), it can be written as:

\[
P_X(V(\tau)) = P_X(\tau) - P_X(\tau) \Phi(T,\tau) M_Y P_V^{-1}(T) M_Y \Phi(T,\tau) P_X(\tau)
\]

where \( P_V(T) \) = \( M_Y P_X(T) M_Y \) and \( \Phi(T,\tau) \) is the state transition matrix from \( \tau \) to \( T \). By taking the partial derivative over \( T \), we have:

\[
\frac{\partial}{\partial \tau} P_X(V(\tau)) = -P_X(\tau) \lambda e^A(T-\tau) M_Y P_V^{-1}(T) M_Y e^A(T-\tau) P_X(\tau) \\
- P_X(\tau) e^A(T-\tau) M_Y P_V^{-1}(T) M_Y e^A(T-\tau) P_X(\tau) \\
+ P_X(\tau) e^A(T-\tau) M_Y P_V^{-1}(T) \left( \frac{\partial}{\partial \tau} P_V(T) \right) \\
\times P_V^{-1}(T) M_Y e^A(T-\tau) P_X(\tau)
\]

The derivative term in the middle becomes:

\[
\frac{\partial}{\partial \tau} P_V(T) = M_Y \left( \frac{\partial}{\partial \tau} P_X(T) \right) M_Y' \\
= M_Y (A P_X(T) + P_X(T) A' + \Sigma_W) M_Y'
\]

Thus, the sensitivity term can be calculated by the value of \( P_X(\tau) \) at \( \tau \) and \( T \). Note that \( P_X(T) \) has already been calculated to compute \( P_{0|V} \), and \( P_X(\tau) = S_X^{-1}(\tau) \) that can be calculated along with the mutual information. Therefore, no significant further information is needed to quantify the sensitivity of the mutual information to the verification time. Moreover, the only term that depends on the measurement is \( Q_X(\tau) \) while other terms can have been computed in advance of posing the optimal planning problem that considers many different options of \( Z_{[0,\tau]} \).

1) Rate of Sensitivity Change: As derived above, the sensitivity of mutual information at time \( \tau \) to the variation in \( T \) is expressed as a function of covariance matrices (equivalently, inverse covariance matrices) at time \( \tau \) and the Lyapunov matrix \( P_X \) at \( T \). Notice that the time argument \( \tau \) can be replaced by any arbitrary time \( t \); thus, the sensitivity value can be tracked on the fly at any time \( t \) \( \in [0,\tau] \), once \( P_X(T) \) has been computed in advance. Now, think of the time derivative of the sensitivity, \( \frac{d}{dt} \mathcal{I}(V_T; Z_{[0,t]}) \), which represents the rate of change of sensitivity over time as measurements are being taken. Therefore, it is expected that some type of sensitivity field can be obtained by evaluating the sensitivity rate at every possible sensor locations, as a similar approach is taken to build the information potential field representing the rate of change of mutual information [5,18].

It can be shown that:

\[
\frac{d}{dt} \mathcal{I}(V_T; Z_{[0,t]}) = \frac{1}{2} \text{tr} \left\{ Q_X|V(\frac{d}{dt} Q_X^{-1}) Q_X|V(\frac{d}{dt} S_X|V) \right\} + g_1(t) \\
= -\frac{1}{2} \text{tr} \left\{ Q_X|V Q_X^{-1} Q_X|V(\frac{d}{dt} S_X|V) \right\} + g_1(t) \\
= \frac{1}{2} \text{tr} \left\{ Q_X|V C' \Sigma N^1 C Q_X|V(\frac{d}{dt} S_X|V) \right\} + g_2(t) \\
= \frac{1}{2} \text{tr} \left\{ C Q_X|V(\frac{d}{dt} S_X|V) Q_X|V C' \right\} + g_2(t),
\]

where \( g_1(t) \) and \( g_2(t) \) are collections of terms that are not dependent on the observation matrix \( C \), and \( \frac{d}{dt} S_X|V \) is calculated as described in the earlier part of this section. Note that the first term of (15) can be evaluated as a function of the sensor location, and it has of very similar form, i.e., quadratic function of \( C \) with some covariance-like weighting matrix, as the information potential field developed in [18].

IV. Numerical Examples

For preliminary validation of the proposed formula, an example of a simple linear system with two states is considered. The system matrices are:

\[
A = \begin{bmatrix} 0.1 & 1 \\ -1 & -0.5 \end{bmatrix}, \quad \Sigma_W = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}, \quad P_0 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad M_Y = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \Sigma_N = 0.01,
\]

and the observation matrix is expressed in terms of a scalar parameter \( \lambda \) such that:

\[
C(t) = [\lambda(t) - 1 - \lambda(t)].
\]

Also, it is assumed that the dynamics of \( \lambda(t) \) is given by

\[
\dot{\lambda} = -20\lambda + 20u,
\]

with \( \lambda(0) = 0.5 \) and \( u \in [0,1] \).

Three optimization problems with the decision variables of \( u(t), t \in [0,\tau] \) are posed with the objective function: (a)
metric and the sensitivity metric drives the system in a different direction. Fig. 1 illustrates the optimized history of $\tau_I(t)$ for the three optimization problems. It is noted that the mutual information metric and the sensitivity metric drives the system in a different direction.

The sensitivity information can be used to approximate the true mutual information value with slight changes in $T$. 

$$I(V_T;Z_{[0,T]}) = \frac{\partial}{\partial T} I(V_T;Z_{[0,T]}) + \Delta T \times \frac{\partial^2}{\partial T^2} I(V_T;Z_{[0,T]}).$$

Fig. 2 compares the true and the approximate value of the accumulated information, defined by $I(V_{T+\Delta T}; Z_{[0,T]})$ for some $t \in [0,T]$. For three different observation parameter values $\lambda = 1, 0.5, 0$, the approximation is obtained with $\Delta T = 0.1$ when $T = 5$ for differing values of $\Delta T$. Observe that the sensitivity-based approximation is very close to the true value.

V. CONCLUSIONS

This paper presented analysis of mutual information in the context of informative forecasting. The expression of mutual information for windowed forecasting was derived as an extension to the point-wise forecasting, and shown to be readily computed without incurring significant additional computation cost. The sensitivity of mutual information to the variation of the verification time was derived, highlighting the capability of constructing the sensitivity field. Simple numerical examples preliminarily verified applicability of the presented analysis, while more realistic examples on mobile sensor planning will be included in the future work.

REFERENCES


