Systems & Control Letters 121 (2018) 24–30

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Sensor placement for optimal estimation of vector-valued diffusion processes

Mohamed-Ali Belabbas [a,*], Xudong Chen [b]

[a] ECE department and Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, United States
[b] ECEE department, University of Colorado at Boulder, United States

ARTICLE INFO

Article history:
Received 7 November 2017
Received in revised form 30 April 2018
Accepted 13 September 2018

Keywords:
Sensor design
Estimation
Wiener process

ABSTRACT

Diffusion processes are commonplace in many scientific disciplines, as they describe a broad range of physical phenomenon. Consider a diffusion process observed through linear sensors with additive white noise. We derive the optimal placement of these sensors for estimating this process, where optimality is defined in terms of the mean squared estimation error (MSE) of the state given past observations. We consider two cases. First, we assume the sensors to be orthogonal. We show in this case that the minimum MSE is related to the nuclear norm of the system matrix of the process. Second, we remove the orthogonality constraint and show that the MSE is related to the Schatten p-norm of the system matrix of the process and the optimal sensors are proportional its matrix cube root. We present simulation results illustrating the fact that the gain afforded by optimizing the choice of sensors depends on the ratio p/n, where n is the dimension of the system and p the dimension of the Wiener processes driving it, and this gain is in general large, especially when p/n is small.

1. Introduction

Diffusion processes are widely used to model stochastic, uncertain or noisy dynamics in engineering, economic, financial and biological systems [1–4]. We refer [5,6] for an introductory exposition of their range of applications and basic properties. As a consequence, estimating the state of such process from noisy observations is a very commonly encountered problem in a wide range of disciplines. We provide in this paper a complete solution to this outstanding problem.

The type of processes we consider are described by the stochastic differential equation [7]:

$$\begin{cases} dx_t = B\,dw_t \\ dy_t = C^\top x_t\,dt + dv_t, \end{cases} \tag{1}$$

where $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{n\times m}$ and v_t, w_t are independent, vector-valued standard Wiener processes, and $x_0 = Bw$ for $w \in \mathbb{R}^p$ a Gaussian random variable. It is well-known that the density $\rho(x)$ of x obeys the generalized heat equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\sum_{j=1}^{p}\sum_{i,i'=1}^{n} b_{ij}b_{i'j}\frac{\partial^2\rho}{\partial x_i \partial x_{i'}}. \tag{2}$$

Eq. (1) can be interpreted as the sample path description of the generalized diffusion described by Eq. (2), which is well-known to arise in the modeling of, e.g., solvent diffusion or heat diffusions [8].

Denote by span$\{B\}$ the vector subspace of $\mathbb{R}^n$ spanned by the columns of B. It is easy to see from (1) that $x_t \in$ span$\{B\}$. Without loss of generality, we assume throughout the paper that B is of full column rank. This implies in particular that $p \le n$. We call x_t the *state process*, and y_t the *measurement process*. We denote by $y_{[0,t)}$ the signal y_s for $s \in [0, t)$. We refer to C as the *sensing matrix* of the system.

The well-known Kalman–Bucy filter [9] provides the minimum mean-squared error estimate of the state x_t given the past measurements $y_{[0,t)}$. We describe the optimal estimation procedure in Section 2.2. The question we answer here is the following:

Q: Which sensing matrix C *minimizes the estimation error*? Equivalently, which sensing matrix minimizes the trace of the covariance of the *conditional* density $\rho(x \mid y)$.

The estimate of the state provided by the Kalman–Bucy filter has bounded mean squared error provided that $B^\top C$ is of full row rank (we prove this in Lemma 2). But within the class of sensing matrices C yielding a bounded estimation error, the error afforded by the Kalman–Bucy filter is easily seen to depend on the *placement or design* of sensors C. One such type of dependence can be easily dealt with: the estimation error can be shown to decrease when the norm of C increases, all other things equal; said more precisely, for

* Corresponding author.
E-mail addresses: belabbas@illinois.edu (M.-A. Belabbas), xudong.chen@colorado.edu (X. Chen).

https://doi.org/10.1016/j.sysconle.2018.09.004
0167-6911/

$0 < r_1 < r_2$ real numbers, and C so that $C^\top B$ in non-singular, the estimation error afforded by $r_1 C$ can be shown to be larger than the estimation error afforded by $r_2 C$. In physical terms, this behavior reflects that a higher signal-to-noise ratio at the sensors yields a better estimate. We investigate here the complementary problem of finding an optimal sensor C of fixed norm.

Related work. The problem of optimally designing sensors to estimate the state of a dynamic process has been investigated in a variety of scenarios. We provide a brief overview of the recent relevant literature. There is a large literature exploring the optimal sensor design, or the dual problem of optimal actuator design, in the infinite-dimensional case [10–12]. The main objective in this setting is to obtain conditions under which the existence of an optimal design is guaranteed, and exhibit conditions under which a sequence of finite-dimensional optimal designs converge to an optimal design. In the finite-dimensional case, choosing a set of sensors c_i out of a *finite* family of potential sensors, has been investigated by several authors. In [13], the authors assign a cost to each sensor and show that optimally choosing a subset of sensors meeting cost constraints is an NP-hard problem, and furthermore exhibit a class of dynamics for which greedy algorithms yield a provably good approximation to the optimal selection. In [14], the authors look at a "relaxed" selection problem, where sensors are selected with a weight w_i to be optimized and propose a convex optimization algorithm. A different type of methods, based on L_1 optimization as a proxy for sensor selection, has been investigated in [15]. Similar scenarios have also been investigated in the statistics literature in the subfield of experiment design, see [16] for a start to the relevant literature.

Optimal design for continuous-time dynamics has been investigated in our recent work [17], where the case of a continuous-time Gauss–Markov process:

$$\begin{cases} dx_t = Ax_t dt + Bdw_t \\ dy_t = C^\top x_t dt + dv_t, \end{cases} \tag{3}$$

where $A \in \mathbb{R}^{n\times n}$ was considered. It was shown there that if A is Hurwitz, i.e. the real part of its eigenvalues are strictly negative, and $\|C\|$ is not too large, then the optimal sensor design problem admits an essentially *unique* (i.e. up to some symmetry) optimal sensor. Furthermore, we have provided a gradient descent algorithm which was proved to converge to this optimal solution generically. The assumption of A stable was required for the result of [17] to hold, which precludes their application to system (1).

Contributions of the paper. We give in this paper an *explicit* solution to the optimal sensor design problem for system (1). We will see that the solution is reminiscent of the *matched filters* that appear often in signal detection [18]. This is not surprising, as one can think of the matched filters as a way to maximize the signal-to-noise ratio of the signal that will be fed to a detection algorithm or, in the present case, to the Kalman–Bucy filter.

We now describe the contents of the paper in more detail. We think of the columns of the sensing matrix C as individual sensors. We investigate in the paper two scenarios. First, we show that if the sensors are orthogonal to each other and are of the same norm, then a solution C^{opt} to the optimal sensor design problem is such that the column space of C^{opt} contains the column space of B (Theorem 5). This optimal performance measure is exactly captured by the nuclear norm of B.

In the second scenario, we relax the constraint that the angles between the column vectors of C are fixed. In other words, we constrain only the sensors to have a given norm (which is related to the signal-to-noise ratio of the measurements they provide, as argued above), but their orientations and number are arbitrary. This second scenario requires a more involved analysis than the first one, but we nevertheless can still exhibit an explicit solution. We show in Theorem 6 that the optimal sensors are proportional to the *cube root* of B, in a sense made precise below. The optimal performance measure is in this case related to the Schatten $2/3$-norm of B.

2. Mathematical background and problem formulation

2.1. Background and notation

We denote by I_n the $n \times n$ identity matrix, by $0_{n\times p}$ the $n \times p$ zero matrix and by $\mathrm{diag}(s_1, \ldots, s_p)$ the $p \times p$ diagonal matrix with diagonal entries s_i, $i = 1, \ldots, p$. With a slight abuse of notation, we also denote by $\mathrm{diag}(s)$, for $s \in \mathbb{R}^{p\times p}$ the vector of $\mathbb{R}^p$ with entries the diagonal entries of s: $\mathrm{diag}(s) = (s_{11}, \ldots, s_{pp})^\top$. We let $GL(n)$ be the set of invertible $n \times n$ matrices. We let $O(n)$ be the set of $n \times n$ orthogonal matrices: $O(n) = \{\Theta \in GL(n) \mid \Theta^\top \Theta = I_n\}$, and $SO(n)$ be the $n \times n$ orthogonal matrices with unit determinant:

$$SO(n) := \{\Theta \in O(n) \mid \det(\Theta) = 1\}.$$

Given a matrix $B \in \mathbb{R}^{n\times p}$, we denote its *singular value decomposition* [19] by

$$B = usv$$

where $u \in SO(n)$, $s \in \mathbb{R}^{n\times p}$ and $v \in O(p)$. The matrix s can have non-zero elements only at the entries $s_{11}, \ldots, s_{pp}$. With a slight abuse of terminology, we call such a matrix s diagonal, The values s_{ii}, for $i = 1, \ldots, p$ are called the *singular values* of B and are always non-negative. Note that the s_{ii}^2 are also the *eigenvalues* of $B^\top B$. We refer to the first p columns of u as the *principal part of* u, and denote it by u_p. Thus, $u_p \in \mathbb{R}^{n\times p}$ is such that $u_p^\top u_p = I_p$. We let $\mathrm{St}(n, p)$ be the Stiefel manifold of p-frames in $\mathbb{R}^n$ [20]:

$$\mathrm{St}(n, p) := \{C \in \mathbb{R}^{n\times p} \mid C^\top C = I_p\},$$

and u_p is an element of $\mathrm{St}(n, p)$. Similarly, we denote by s_p the matrix formed by the first p rows of s. Thus s_p is a diagonal $p \times p$ matrix with the singular values of B on the diagonal. The singular value decomposition is then reduced to

$$B = u_p s_p v.$$

with $u_p \in \mathrm{St}(n, p)$, $s_p \in \mathbb{R}^{p\times p}$ and $v \in O(p)$.

Schatten p-norms and nuclear norm of matrices.

Let $r > 0$ be a positive real number, let $B \in \mathbb{R}^{n\times p}$ and set $B = usv$ to be the singular value decomposition of B. The *Schatten–von Neumann* r-quasi-norm of B is defined as

$$\|B\|_r = \left(\sum_{i=1}^{p} s_i^r\right)^{1/r}.$$

Said otherwise, it is the r-quasi-norm of the vector of singular values of B [21]. If $r \geq 1$, it is a norm and if $r = 1$, it is the nuclear norm (also known as the trace norm) of B. We have the following inequality:

Lemma 1. *Let $0 < q < r$ be two real numbers and $B \in \mathbb{R}^{n\times p}$. Then*

$$\|B\|_q \leq p^{\frac{1}{q}-\frac{1}{r}} \|B\|_r.$$

This inequality has certainly been observed before; we reproduce a proof in the Appendix for the sake of completeness.

2.2. Problem formulation

The minimum mean squared error (MSE) estimator of x_t given the past measurements $y_{[0,t)}$ is the conditional expectation:

$$\hat{x}_t := \mathbb{E}(x_t \mid y_{[0,t)}).$$