Automatica 81 (2017) 403–408

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Joint maximum *a posteriori* state path and parameter estimation in stochastic differential equations☆

Dimas Abreu Archanjo Dutra[a], Bruno Otávio Soares Teixeira[b], Luis Antonio Aguirre[b]

[a] Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

[b] Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

ARTICLE INFO

Article history:
Received 6 December 2014
Received in revised form
6 February 2017
Accepted 6 March 2017

Keywords:
Estimation theory
Parameter estimation
State estimation
Optimization problems
Stochastic systems

ABSTRACT

In this article, we introduce the joint maximum *a posteriori* state path and parameter estimator (JME) for continuous-time systems described by stochastic differential equations (SDEs). This estimator can be applied to nonlinear systems with discrete-time (sampled) measurements with a wide range of measurement distributions. We also show that the minimum-energy state path and parameter estimator (MEE) obtains the joint maximum *a posteriori* noise path, initial conditions, and parameters. These estimators are demonstrated in simulated experiments, in which they are compared to the prediction error method (PEM) using the unscented Kalman filter and smoother. The experiments show that the MEE is biased for the damping parameters of the drift function. Furthermore, for robust estimation in the presence of outliers, the JME attains lower state estimation errors than the PEM.

1. Introduction

In the context of discrete-time systems, maximum *a posteriori* (MAP) state path estimators have recently emerged as a powerful alternative to Kalman filters and smoothers due to their robustness properties and applicability to a larger class of models (Aravkin, Bell, Burke, & Pillonetto, 2011; Aravkin, Burke, & Pillonetto, 2012b,c, 2013; Bell, Burke, & Pillonetto, 2009; Dutra, Teixeira, & Aguirre, 2012; Farahmand, Giannakis, & Angelosante, 2011; Monin, 2013). A wide variety of phenomena of engineering interest are continuous-time in nature and can be modeled by stochastic differential equations (SDEs). For this class of models, the MAP state-path estimator is built upon the Onsager–Machlup functional and is the solution to an optimal control problem (Aihara & Bagchi, 1999a,b; Zeitouni & Dembo, 1987).

To evaluate if a discretization of a variational optimization problem is *consistent*, the concept of hypographical convergence is used (cf. Polak, 2011). If a sequence of discretized problems hypo-converges to a variational problem, then the discretized optima converge to the variational optima. In a previous work (Dutra, Teixeira, & Aguirre, 2014), we showed that the discrete-time MAP state path estimator applied to trapezoidally discretized continuous-time systems converges hypographically to the MAP state path estimator of the continuous systems, as the discretization step vanishes. However, when the Euler discretization is used instead – the most widespread approach – the discretized estimator hypo-converges to the minimum-energy estimator, whose estimates were proved to be MAP *noise* paths. This implies that the discretized MAP estimates have a different interpretation depending on the discretization scheme used.

In this work, we present the extension of the estimators of Dutra et al. (2014) for joint state path and parameter estimation. We introduce the joint MAP state path and parameter estimator (JME) for continuous-time systems and also show that the joint minimum-energy state path and parameter estimator (MEE) corresponds to the joint MAP *noise* path, initial state and parameter estimator. The JME and MEE are also the hypographical limits of the trapezoidally-and Euler-discretized joint state path and parameter estimators (Dutra, 2014, Chap. 3), respectively.

The fact that the parameter is estimated as a single vector instead of a time-varying augmented state places the JME and MEE in a similar niche to the Kalman-filter-based prediction error method (PEM) (Kristensen, Madsen, & Jørgensen, 2004), to which it is compared. Furthermore, if the joint state path and parameter posterior distribution is unimodal and approximately symmetric, the JME estimates should be close to the marginal MAP parameter

☆ This work has been financially supported by the Brazilian agencies CAPES, CNPq (grants 309122/2014-7, 140102/2013-2, 302079/2011-4, 302127/2011-9, 470452/2011-0) and FAPEMIG. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Valery Ugrinovskii under the direction of Editor Ian R. Petersen.

E-mail addresses: dimasad@ufmg.br (D.A.A. Dutra), brunoot@ufmg.br (B.O.S. Teixeira), aguirre@cpdee.ufmg.br (L.A. Aguirre).

http://dx.doi.org/10.1016/j.automatica.2017.03.035
0005-1098/

estimates. Similarly, even when the parameters are not of interest, the JME can be used as a state-path estimator under parametric uncertainty.

The merit function of these estimators admits a tractable expression for a wide range of nonlinear systems, lending them a wider applicability than Kalman-filter-based estimators. In particular, it is possible to use heavy-tailed measurement distributions which confer robustness against outliers (Aravkin et al., 2011; Aravkin, Burke, & Pillonetto, 2012a; Aravkin et al., 2012b,c, 2013; Dutra et al., 2012; Farahmand et al., 2011). The resulting estimators can be seen as an extension of Huber's M-estimators (Huber & Ronchetti, 2009, Sec. 3.2) to the smoothing problem. M-estimates of location parameters using heavy-tailed distributions can be interpreted as implicit weighted means, with low weights assigned to outlying observations. This approach "combines conceptual simplicity with generality, since it can be applied to a wide range of settings" (Lange, Little, & Taylor, 1989, p. 882). A competing approach to robust estimation is to consider a family of distributions in the neighborhood of a nominal guess and design filters or smoothers which guarantee the best behavior in the worst-case scenario, i.e., minimax estimators (Levy & Nikoukhah, 2013; Zorzi, 2016).

The remainder of this article is organized as follows. In Section 2 we define the problem being tackled and common notation and variables. In Section 3 the fictitious densities and the MAP estimators are defined and presented. The simulated example applications are presented in Section 4 and conclusions and future work in Section 5.

2. Problem definition

In what follows, $(\Omega, \mathcal{F}, P)$ is a standard probability space on which all random variables are defined. Random variables will be denoted by uppercase letters and their values by lowercase, so that if $Y : \Omega \to \mathcal{Y}$ is a $\mathcal{Y}$-valued random variable, $y \in \mathcal{Y}$ will denote specific values it might take. The same applies to stochastic processes. The dependency on the random outcome $\omega \in \Omega$ will be omitted when unambiguous, to simplify the notation. For a random variable Θ, $\operatorname{supp}(P_\Theta)$ denotes the topological support of its induced measure. The time argument of functions may also be written as subscripts for compactness.

Let X and Z be $\mathbb{R}^m$- and $\mathbb{R}^n$-valued stochastic processes, respectively, representing the state of a system over the experiment interval $\mathcal{T} := [0, T]$ and satisfying the following system of SDEs:

$$\mathrm{d}X_t = f(t, X_t, Z_t, \Theta)\,\mathrm{d}t + G\,\mathrm{d}W_t, \tag{1a}$$

$$\mathrm{d}Z_t = h(t, X_t, Z_t, \Theta)\,\mathrm{d}t, \tag{1b}$$

where f and h are the drift functions, the $\mathbb{R}^q$-valued random variable Θ is the unknown parameter vector, the full rank $G \in \mathbb{R}^{m\times m}$ is the diffusion matrix, and W is an m-dimensional Wiener process. This division of the state in two parts, X and Z, is done to cover systems in which the evolution of some state variables is not under directly influence of noise.

Consider, in addition, that some $\mathcal{Y}$-valued random variable Y is observed. We assume that the conditional distribution of Y, given $X = x$, $Z = z$ and $\Theta = \theta$, is absolutely continuous and admits a density ψ with respect to a measure ν on the measurable space $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$, i.e., for all $\mathbb{B} \in \mathcal{B}_\mathcal{Y}$,

$$P_Y(\mathbb{B} \mid X = x, Z = z, \Theta = \theta) = \int_\mathbb{B} \psi\,(y \mid x, z, \theta)\,\mathrm{d}\nu(y).$$

In this paper, we derive the joint MAP estimator for X, Z_0 and Θ, given $y \in \mathcal{Y}$. Note that, conditioned on that, the whole Z path is also uniquely defined. We also prove that the minimum-energy estimator is the joint MAP estimator for W, X_0, Z_0 and Θ. We use the abstract definitions of mode and the MAP estimator of Dutra et al. (2014, Defns. 1 and 2), which cover random variables over infinite-dimensional spaces such as state paths of continuous-time systems. These definitions can be better understood using the concept of a *fictitious density*, which we introduce formally in the definition below. Similar terminology was applied to the Onsager–Machlup functional in this context, it was described as an *ideal* density with respect to a *fictitious* uniform measure by Takahashi and Watanabe (1981, p. 433) and as a *fictitious density* by Zeitouni (1989, p. 1037).

Definition 1 (*Dutra, 2014, Defn. 2.4*)**.** Let A be an $\mathcal{A}$-valued random variable, where $(\mathcal{A}, d)$ is a metric space. The function $\zeta : \mathcal{A} \to \mathbb{R}$ is a *fictitious density* if $\zeta(a) > 0$ for at least one $a \in \mathcal{A}$ and there exists $\xi : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that

$$\lim_{\epsilon\downarrow 0} \frac{P\,(d(A, a) < \epsilon)}{\xi\,(\epsilon)} = \zeta\,(a) \quad \text{for all } a \in \mathcal{A}.$$

The fictitious density can be understood as a density with respect to a metric. It quantifies the concentration of probability in the neighborhood of a point. When, for some $a', a'' \in \mathcal{A}$, the fictitious density $\zeta(a') > \zeta(a'')$, then the ϵ-balls around a' have a larger probability than those around a'', for all sufficiently small ϵ. This means that the MAP estimates according to Dutra et al. (2014, Defn. 2) are the maxima of the posterior fictitious density. It should be noted that for Euclidean spaces any fictitious density is proportional to the probability density function in the usual sense. We now show the application of these concepts to the state paths and parameters of the system described by (1).

3. MAP and minimum energy estimators

The following assumptions will be made on the system's probabilistic and dynamical model.

Assumption 2. a. The initial states X_0, Z_0 and the parameter vector Θ are $\mathcal{F}_0$-measurable and admit a continuous joint prior density π with respect to the Lebesgue measure.

b. The functions f and h are uniformly continuous with respect to all their arguments for $\theta \in \operatorname{supp}(P_\Theta)$.

c. For all fixed $\theta \in \operatorname{supp}(P_\Theta)$, the functions f and h are Lipschitz continuous with respect to their second and third arguments x and z, uniformly over their first argument t.

d. For all fixed $\theta \in \operatorname{supp}(P_\Theta)$, the function f is twice differentiable with respect to its second argument x and differentiable with respect to its first and third arguments t and z. Furthermore, its first and second derivatives mentioned above are continuous with respect to all arguments, for all $\theta \in \operatorname{supp}(P_\Theta)$.

e. The system is such that

$$\mathrm{E}\left[\exp\left(\int_0^T \left\|G^{-1}f(t, X_t, Z_t, \Theta)\right\|^2 \mathrm{d}t\right)\right] < \infty.$$

f. The measurement likelihood ψ is continuous with respect to the given x, z and θ.

g. The observed y value has a positive prior predictive density, i.e., $\mathrm{E}\,[\psi\,(y|X, Z, \Theta)] > 0$.

In what follows, we will denote by $\mathcal{H}^d$ the space of absolutely continuous $x : \mathcal{T} \to \mathbb{R}^d$ with square integrable weak derivatives $\dot{x}$. For $x \in \mathbb{R}^d$, $\|x\|$ will denote its Euclidean norm. Furthermore, $|\!|\!|\cdot|\!|\!|$ will denote the supremum norm of continuous functions from $\mathcal{T}$ to $\mathbb{R}^d$:

$$|\!|\!|w|\!|\!| := \sup_{t\in\mathcal{T}} \|w(t)\|\,.$$

The divergence of a vector field function f, with respect to a variable x is denoted $\operatorname{div}_x f$, i.e., $\operatorname{div}_x f = \sum_k \frac{\partial f_k}{\partial x_k}$.