

Asymptotic Optimality of the Maximum-likelihood Filter for Bayesian Tracking in Sensor Networks

Damián Marelli^{1,2} and Minyue Fu^{1,3}

¹School of Electrical Engineering and Computer Science, University of Newcastle,
Australia.

²Acoustics Research Institute, Austrian Academy of Sciences.

³Department of Control Science and Engineering, Zhejiang University, China.

Abstract: A recently proposed Bayesian tracking procedure for sensor networks approximates the update equation, which involves non-linear measurements, with a simple equation using the maximum likelihood (ML) estimate of the unknown state. This approach permits a numerically efficient implementation of the tracking procedure, and is suitable for a distributed implementation. In this paper we study the extent to which this approach approximates the theoretical Bayesian solution. We provide conditions to guarantee that the approximation becomes asymptotically exact, as the number of nodes becomes large. This result is relevant in applications where each sensor obtains a measurement with limited information about the state, but a large number of sensors is available. We apply our result to a case study, and present numerical simulations, showing that the approximation error becomes negligible for a relatively small number of nodes.

Keywords: sensor networks, Bayesian tracking, distributed estimation.

1. INTRODUCTION

A sensor network consists of a large number of sensing and computing devices, with limited computational and communication capabilities, connected via a communication network [Dargie and Poellabauer, 2010]. Sensor networks find applications in environmental, health-care and weather monitoring, industrial process monitoring and control, surveillance, smart grids, etc. They are classified according to the architecture used for data fusion. In a centralized architecture, a central node collects the data from all sensor nodes to perform a given task. A variant of this architecture is the hierarchical one, in which the set of nodes is partitioned into clusters, forming hierarchical levels, and each cluster behaves as a centralized network. On the other hand, in a decentralized architecture, the desired task is executed via a distributed algorithm carrying out a cooperative strategy over all the nodes of the network.

The development of algorithms for Bayesian tracking in sensor networks has recently attracted a great deal of attention [Carli et al., 2008, Khan and Moura, 2008, Ribeiro et al., 2010, Hlinka et al., 2012]. In this application, the measurements from all nodes are combined to track the evolution of a set of stochastic parameters (state vector) for which a dynamic model is available. In a number of applications, the model describing the dynamic evolution of the state vector (*state equation*) is not known, and it is therefore approximated by a linear equation. On the other

hand, the measurement model (*measurement equation*) of each sensor node can be accurately described, and it leads in general to a non-linear equation. This is particularly true in sensors using coarse quantization, as often occurs in practice. This yields a Wiener dynamical model, i.e., in which the state evolution is linear but the measurements are non-linear in the state. There is a number of methods which can be used for addressing the Bayesian tracking problem in general non-linear models [Chang et al., 2013, Arasaratnam and Haykin, 2009, Arulampalam et al., 2002, Kotecha and Djuric, 2003, Simandl et al., 2006]. These methods alternate between two steps called *prediction*, which involves the state equation, and *update*, which involves the measurement equation. A drawback of these methods is that they are either computationally expensive, or, in the case of the extended Kalman filter [Anderson and Moore, 1979], inaccurate to the extent that they can lead to instability. Motivated by this, the authors of [Wang et al., 2012] proposed a Bayesian tracking technique which is particularly suitable for Wiener dynamical models. In this technique, the update step (resulting from the non-linear measurement equation) is replaced by an alternative step obtained using a maximum likelihood (ML) estimation of the state, based on the sensor measurements. In this way, the Bayesian tracking procedure can be executed in a way similar to that of a Kalman filter, with the extra cost of executing a ML estimation at each update step. In this paper we refer to this Bayesian tracking strategy as the *ML filter*. The ML filter offers an efficient Bayesian tracking method for sensor networks with either centralized and decentralized architecture. The latter can be achieved by complementing the ML filter with a distributed ML algorithm [Hlinka et al., 2012].

¹ This work was partially supported by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (No. ICT1414) and by the Austrian Science Fund (FWF project M1230-N13).

In this paper we provide conditions under which the ML filter becomes asymptotically optimal, in the sense of approaching the theoretical Bayesian tracking solution, as the number of sensor nodes tends to infinity. This theoretical result is particularly relevant in applications where each sensor provides limited information about the state vector to be estimated, but the number of sensors is large. To illustrate the application of our result, we show the asymptotic optimality of the ML filter for tracking a parameter vector using nodes, each of which obtaining its measurement from a quantizer. Numerical simulations show that, even for a one-bit quantizer, the difference between the ML filter and an approximation to the theoretically optimal Bayesian tracker obtained using particle filtering, becomes negligible for a relatively small number of nodes.

The rest of the paper is organized as follows: In Section 2 we describe the ML filter, and the addressed problem. In Section 3 we introduce our main result. In Section 4 we use this result to show the asymptotic optimality of the ML filter in a sensor network application described above, and in Section 5 we provide the simulation results. We give concluding remarks in Section 6. Due to the limited space, the proof of our main result is omitted and will appear in a journal version.

Notation: The symbols \mathbb{N} and \mathbb{R} denote the sets of natural and real numbers, respectively. Also, $C_n(\mathbb{R}^d)$ denotes the space of functions with n -th order continuous derivatives, and $L_p(\mathbb{R}^d)$ the set of functions whose p -th power is absolutely integrable. For a vector x , $\|x\|$ denotes its 2-norm, and for a matrix A , $\|A\|$ denotes its operator (induced) norm. Convergence with probability one is denoted by w.p.1. The multivariate normal probability density function of mean μ and covariance matrix Σ is denoted by $g_{\mu,\Sigma}$, i.e.,

$$g_{\mu,\Sigma}(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}\right).$$

Finally, $\nabla f(x)$ denotes the gradient of the function $f(x)$ and $\nabla^2 f(x)$ its Hessian.

2. PROBLEM DESCRIPTION

Consider the following state equation

$$x(t+1) = Ax(t) + w(t), \quad (1)$$

where $w(t) \sim \mathcal{N}(0, R)$, for all $t \in \mathbb{N}$, and $x(0) \sim \mathcal{N}(0, P)$, with $0 < R, P \in \mathbb{R}^{d \times d}$. There are N sensor nodes. At time $t \in \mathbb{N}$, for each $n = 1, \dots, N$, the measurement $y_n(t)$ of Node n is a random vector given by

$$y_n(t) = f_n(x(t)), \quad (2)$$

with $f_n(\cdot)$ being a stochastic function (i.e., for each x , $f_n(x)$ is a random vector). Let $Y_N(t) = [y_1^T(t), \dots, y_N^T(t)]^T$ and $\mathbf{Y}_{N,t} = [Y_N(1)^T, \dots, Y_N(t)^T]^T$.

A Bayesian tracking procedure for estimating $x(t)$ is obtained by iterating the following two steps [Arulampalam et al., 2002]:

Bayesian tracker:

Update:

$$\mathcal{P}(x(t)|\mathbf{Y}_{N,t}) = \frac{\mathcal{P}(Y_N(t)|x(t))}{\mathcal{P}(Y_N(t)|\mathbf{Y}_{N,t-1})} \mathcal{P}(x(t)|\mathbf{Y}_{N,t-1}). \quad (3)$$

Prediction:

$$\begin{aligned} \mathcal{P}(x(t+1)|\mathbf{Y}_{N,t}) &= \\ &= \int \mathcal{P}(x(t+1)|x(t)) \mathcal{P}(x(t)|\mathbf{Y}_{N,t}) dx(t). \end{aligned} \quad (4)$$

The iterations above can be approximated by analytic expressions, similar to those used in a Kalman filter, if for any given value of $Y_N(t)$, $\mathcal{P}(Y_N(t)|x(t))$ can be approximated by a Gaussian function of $x(t)$, i.e.,

$$\mathcal{P}(Y_N(t)|x(t)) \simeq a_N(t) e^{-\frac{1}{2}(x(t)-\mu_N(t))^T \Sigma_N(t)^{-1} (x(t)-\mu_N(t))}, \quad (5)$$

for some $a_N(t) > 0$, $\mu_N(t) \in \mathbb{R}^d$ and $0 < \Sigma_N(t) \in \mathbb{R}^{d \times d}$, which depend on Y_N (notice that the value of $a_N(t)$ is not required, because its value can be inferred from the fact that the right-hand side of (3) has unit integral with respect to $x(t)$). More precisely, If $\mathcal{P}(x(t)|\mathbf{Y}_{N,t-1}) = g_{\mu_{t|t-1}, \Sigma_{t|t-1}}(x(t))$, then, from [Petersen and Pedersen, 2006, Sec. 8.1.8], (3) becomes

$$\mathcal{P}(x(t)|\mathbf{Y}_{N,t}) \simeq g_{\mu_{t|t}, \Sigma_{t|t}}(x(t)), \quad (6)$$

where $\mu_{t|t}$ and $\Sigma_{t|t}$ are given by (10) and (11), respectively. Let

$$L_{N,t}(x) = \mathcal{P}(Y_N(t)|x(t) = x) \quad (7)$$

be the likelihood function (LF) of $x(t)$, given $Y_N(t)$. Then, a particular choice for the pair $(\mu_N(t), \Sigma_N(t))$ is

$$\mu_N(t) = \arg \max_{x \in \mathbb{R}^d} L_{N,t}(x), \quad (8)$$

$$\Sigma_N(t) = -L_{N,t}(\mu_N(t)) \nabla^2 L_{N,t}(\mu_N(t))^{-1}. \quad (9)$$

The ML filter is the Bayesian procedure obtained by using (1) in (4) and (6) in place of (3).

Maximum likelihood filter:

Update: $\mathcal{P}(x(t)|\mathbf{Y}_{N,t}) = g_{\mu_{t|t}, \Sigma_{t|t}}(x(t))$, with

$$\mu_{t|t} = \Sigma_{t|t} \left(\Sigma_{t|t-1}^{-1} \mu_{t|t-1} + \Sigma_N^{-1}(t) \mu_N(t) \right), \quad (10)$$

$$\Sigma_{t|t} = \left(\Sigma_{t|t-1}^{-1} + \Sigma_N^{-1}(t) \right)^{-1}, \quad (11)$$

and $\mu_N(t)$ and $\Sigma_N(t)$ given by (8) and (9), respectively.

Prediction: From [Anderson and Moore, 1979, p. 40], $\mathcal{P}(x(t+1)|\mathbf{Y}_{N,t}) = g_{\mu_{t+1|t}, \Sigma_{t+1|t}}(x(t+1))$, with

$$\mu_{t+1|t} = A \mu_{t|t}, \quad (12)$$

$$\Sigma_{t+1|t} = A \Sigma_{t|t} A^T + R. \quad (13)$$

In this paper we provide conditions under which the approximation (5) with (8)-(9) becomes asymptotically exact, as the number N of measurements tends to infinity.

3. MAIN RESULT

Notation. Since the study of the approximation (5) is independent of the time index t , in this section we drop this index to simplify the notation.

Consider the LF L_N of x , given in (7), as well as μ_N and Σ_N , given in (8) and (9), respectively. Notice that, since $L_N(x) \geq 0$, for all $x \in \mathbb{R}^d$, (??) implies that $L_N(\mu_N) > 0$. Hence the matrix Σ_N is positive definite. In stating the regularity conditions required by our main result, we use the concept of strong uniform convergence [Davidson, 1994, Sec. 21.2]. Its definition is given below.

Definition 1. Let $\mathcal{D} \subseteq \mathbb{R}^d$. We say that a sequence of stochastic functions $f_n : \mathcal{D} \rightarrow \mathbb{R}^q$, $n \in \mathbb{N}$, is continuous and strongly uniformly convergent (CSUC) to $f : \mathcal{D} \rightarrow \mathbb{R}^q$, if each f_n is continuous, and

$$\lim_{N \rightarrow \infty} \sup_{x \in \mathcal{D}} \|f_n(x) - f(x)\| \stackrel{\text{w.p.1}}{=} 0.$$

We now state our main result, whose proof is omitted and will appear in a journal version.

Theorem 2. Let $L_N \in C_2(\mathbb{R}^d)$, $\mu_N \in \mathbb{R}^d$, $\Sigma_N \in \mathbb{R}^{d \times d}$, $N \in \mathbb{N}$, be defined as above. Let also

$$\Xi_N(x) = \frac{1}{N} \log L_N(x), \quad (14)$$

be the normalized logarithmic likelihood function and

$$\bar{\Xi}(x) = \lim_{N \rightarrow \infty} \mathcal{E} \{\Xi_N(x)\},$$

be its asymptotic expectation. If

(G1) $\arg \max_{x \in \mathbb{R}^d} \bar{\Xi}(x) = \{x_\star\}$ (i.e., x_\star is the unique maximizer for $\bar{\Xi}(x)$);

(G2) $\lim_{N \rightarrow \infty} \mu_N \stackrel{\text{w.p.1}}{=} x_\star$;

(G3) there exists a compact set $\mathcal{D} \subset \mathbb{R}^d$, with $x_\star \in \text{int}(\mathcal{D})$ (i.e., the interior of \mathcal{D}), such that $\Xi_N : \mathcal{D} \rightarrow \mathbb{R}$ and $\nabla \Xi_N : \mathcal{D} \rightarrow \mathbb{R}^d$ converge w.p.1 and $\nabla^2 \Xi_N : \mathcal{D} \rightarrow \mathbb{R}^{d \times d}$ is CSUC;

(G4) We have, w.p.1.,

$$\nabla^2 L_N^{1/N} \in L_2^{d \times d}(\mathbb{R}^d), \quad \forall N \in \mathbb{N}, \quad (15)$$

$$\sup_{N \in \mathbb{N}, x \in \mathbb{R}} \left\| \nabla^2 L_N^{1/N}(x) \right\| < \infty. \quad (16)$$

then

$$\lim_{N \rightarrow \infty} \frac{1}{L_N(\mu_N)} L_N \left(\Sigma_N^{1/2} x + \mu_N \right) \stackrel{\text{w.p.1}}{=} \exp \left(-\frac{x^T x}{2} \right). \quad (17)$$

Theorem 2 states that, for large N ,

$$\mathcal{P}(Y_N|x) \stackrel{\text{w.p.1}}{\simeq} L_N(\mu_N) e^{-\frac{1}{2}(x-\mu_N)^T \Sigma_N^{-1}(x-\mu_N)}, \quad (18)$$

where μ_N is the ML estimate of x , and Σ_N is obtained from the Hessian $\nabla^2 L_N(\mu_N)$ of $L_N(x)$ at μ_N , using (??). Hence, the theorem asserts that the approximation (5)-(9), and therefore the ML filter, becomes asymptotically exact, as the number N of measurements tends to infinity.

Remark 3. Conditions (G2) and (G1) together require that the sequence μ_N converges to the maximum of the limit logarithmic likelihood function $\bar{\Xi}$. Also, (G3) and (G4) are regularity conditions which are not difficult to satisfy. Hence, from a practical point of view, the conditions of Theorem 2 are rather general.

4. CASE STUDY: SENSOR NETWORK WITH QUANTIZED MEASUREMENTS

To illustrate the use of Theorem 2, we consider a network of sensors measuring quantized linear combinations of a

common random vector. Consider the system (1)-(2), with

$$f_n(x(t)) = \mathcal{Q} [c_n^T x(t) + v_n],$$

where $c_n \in \mathbb{R}^d$ is a vector and $v_n \sim \mathcal{N}(0, \sigma^2)$. The map $\mathcal{Q} : \mathbb{R} \rightarrow \{q_1, \dots, q_K\}$ is a quantizer defined by $\mathcal{Q}^{-1}[q_k] = [b_{k-1}, b_k]$, with $-\infty = b_0 < b_1 < \dots < b_K = \infty$. We also let v_n and v_m be statistically independent whenever $n \neq m$.

Using Theorem 2, we can show the following result, which implies, via (18), that the ML filter becomes asymptotically exact, as the number N of measurements tends to infinity.

Theorem 4. In the system described above, if $\sup_{n \in \mathbb{N}} \|c_n\| < \infty$ and

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N c_n c_n^T > 0, \quad (19)$$

then (17) holds for each $t \in \mathbb{N}$.

Before proving Theorem 4, we introduce the definition of strong asymptotic equicontinuity [Davidson, 1994, eq. (21.43)]. This is a technical condition which is used to guarantee the CSUC property.

Definition 5. Let $\mathcal{D} \subseteq \mathbb{R}^d$. A sequence of stochastic functions $g_n : \mathcal{D} \rightarrow \mathbb{R}^q$, $n \in \mathbb{N}$, is *strongly asymptotically equicontinuous* (SAE) if, for every $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^d} \sup_{y: \|y-x\| < \delta_\epsilon} \|g_n(y) - g_n(x)\| \stackrel{\text{w.p.1}}{<} \epsilon.$$

In order to use Theorem 2 to show Theorem 4, we need to show that, for each $t \in \mathbb{N}$, $\mu_N(t)$, obtained from (8), converges w.p.1 to some local maximum $x_\star(t) \in \mathbb{R}^d$ of $\bar{\Xi}(x)$. While not necessary for Theorem 2, we go a step further and show that $x_\star(t)$ is actually the true value $x(t)$ of the state to be estimated at time t . The convergence in this case is called strong consistency. Conditions to guarantee strong consistency typically require that the maximization problem used to find $\mu_N(t)$ is constrained to a compact set [Gourieroux and Monfort, 1996, Property 24.2], [Newey and McFadden, 1994, Th. 2.1]. However, this assumption is inappropriate in our context. This is because we are estimating the state of a linear dynamic model, which is not guaranteed to be included in any compact set. To go around this issue, we introduce the following variant of the strong consistency results cited above.

Theorem 6. Let $x_\star \in \mathbb{R}^d$, for each $N \in \mathbb{N}$, $\Xi_N : \mathbb{R}^d \rightarrow \mathbb{R}$ and

$$\mu_N = \arg \max_{x \in \mathbb{R}^d} \Xi_N(x), \quad (20)$$

$$\bar{\Xi} = \lim_{N \rightarrow \infty} \mathcal{E} \{\Xi_N(x)\}. \quad (21)$$

If

(C1) $\arg \max_{x \in \mathbb{R}^d} \bar{\Xi}(x) = \{x_\star\}$;

and there exists a compact set $\mathcal{K} \subset \mathbb{R}^d$ with $x_\star \in \mathcal{K}$ such that

(C2) Ξ_N is CSUC on \mathcal{K} ;

(C3) $\limsup_{N \rightarrow \infty} \sup_{x \in \mathcal{K}^c} \Xi_N(x) \stackrel{\text{w.p.1}}{<} \bar{\Xi}(x_\star)$ (\mathcal{K}^c denotes the complement of the set \mathcal{K});

then

$$\lim_{N \rightarrow \infty} \mu_N \stackrel{\text{w.p.1}}{=} x_*$$

Proof. Let $\epsilon = \bar{\Xi}(x_*) - \limsup_{N \rightarrow \infty} \sup_{x \in \mathcal{K}^c} \Xi_N(x)$. We have $\epsilon \stackrel{\text{w.p.1}}{>} 0$. Let $N_0 \in \mathbb{N}$, such that, for all $N \geq N_0$,

$$\begin{aligned} \sup_{x \in \mathcal{K}} |\Xi_N(x) - \bar{\Xi}(x)| &< \frac{\epsilon}{2}, \\ \sup_{x \in \mathcal{K}^c} \bar{\Xi}(x_*) - \Xi_N(x) &> \frac{\epsilon}{2}. \end{aligned}$$

Then, for all $N \geq N_0$, $\mu_N \in \mathcal{K}$. Then, the result follows from [Gourieroux and Monfort, 1996, Property 24.2].

We now provide the proof of Theorem 4.

Proof. We split the argument in steps:

Step 1: We need to show that, for each $t \in \mathbb{N}$, $L_{N,t}(x) = \mathcal{P}(Y_N(t)|x(t) = x)$ satisfies the conditions of Theorem 2. Due to the statistical independence of measurements from different sensors, we have

$$\begin{aligned} L_{N,t}(x) &= \prod_{n=1}^N l_{n,t}(x), \\ l_{n,t}(x) &= \int_{\mathcal{Q}^{-1}[y_n(t)]} g_{c_n^T x, \sigma^2}(\xi) d\xi. \end{aligned}$$

Let

$$\begin{aligned} \Xi_{N,t}(x) &= \frac{1}{N} \sum_{n=1}^N \log l_{n,t}(x), \\ \bar{\Xi}_t(x) &= \lim_{N \rightarrow \infty} \mathcal{E} \{ \Xi_{N,t}(x) \}. \end{aligned}$$

Step 2: We have

$$\begin{aligned} &\bar{\Xi}_t(x(t)) - \bar{\Xi}_t(x) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathcal{E} \left\{ \log \frac{l_{n,t}(x(t))}{l_{n,t}(x)} \right\} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N D_{\text{KL}}(\mathcal{P}(y_n(t)|x(t)) \parallel \mathcal{P}(y_n(t)|x)), \end{aligned}$$

where $D_{\text{KL}}(p||q)$ denotes the Kullback-Leibler distance between the probability distributions p and q [Cover and Thomas, 2006]. Let $z_n = c_n^T x$, and

$f_n(z) = D_{\text{KL}}(\mathcal{P}(y_n(t)|z_n = c_n^T x(t)) \parallel \mathcal{P}(y_n(t)|z_n = z))$. From [Cover and Thomas, 2006, Th. 2.6.3], $f_n(z) \geq 0$ with equality if and only if $z = c_n^T x(t)$. Then, $\partial f_n(c_n^T x(t)) = 0$. Also, it is easy to verify that there exists $\epsilon > 0$ such that $\partial^2 f_n(z) \geq \epsilon$. Thus,

$$f_n(z) \geq \epsilon (z - c_n^T x(t))^2.$$

Then,

$$\begin{aligned} &\bar{\Xi}_t(x(t)) - \bar{\Xi}_t(x) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f_n(c_n^T x) \\ &\geq \epsilon (x - x(t))^T \left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N c_n c_n^T \right) (x - x(t)). \end{aligned}$$

Hence, from (19),

$$\arg \max_{x \in \mathbb{R}^d} \bar{\Xi}(x) = \{x(t)\}, \quad (22)$$

and Condition (G1) follows.

Step 3: Fix $t \in \mathbb{N}$ and $x \in \mathbb{R}^d$. Then, $l_{n,t}(x)$, $n \in \mathbb{N}$, are independent random variables. Hence, so are $\log l_{n,t}(x)$, $\nabla \log l_{n,t}(x)$ and $\nabla^2 \log l_{n,t}(x)$. It is straightforward to verify that $\log l_{n,t}(x)$, $\nabla \log l_{n,t}(x)$ and $\nabla^2 \log l_{n,t}(x)$ have uniformly bounded second moments. Then, from Rajchman's strong law of large numbers [Chung, 2000, Th. 5.1.2], $\Xi_{N,t} : \mathbb{R}^d \rightarrow \mathbb{R}$, $\nabla \Xi_{N,t} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\nabla^2 \Xi_{N,t} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ converge with probability one. Also it is straightforward to verify that,

$$\sup_{i \in \{1, \dots, d\}, n \in \mathbb{N}, x \in \mathbb{R}^d} \left\| \partial_i \nabla^2 \log l_{n,t}(x) \right\| \stackrel{\text{w.p.1}}{<} M < \infty.$$

Then, the sequence $\nabla^2 \log l_{n,t}$ is SAE on \mathbb{R}^d , and from Lemma 9, so is $\nabla^2 \Xi_{N,t} = \frac{1}{N} \sum_{n=1}^N \nabla^2 \log l_{n,t}$. Then, from Lemma 8,

$$\nabla^2 \Xi_{N,t} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d} \text{ is CSUC}, \quad (23)$$

and Condition (G3) follows.

Step 4: Fix $t \in \mathbb{N}$ and $x \in \mathbb{R}^d$, and choose a compact set $\mathcal{D}_x \subset \mathbb{R}^d$ with $x \in \text{int}(\mathcal{D}_x)$. From (23) and two applications of Lemma 7, $\bar{\Xi}_t(x)$ is twice continuously differentiable at x . Then,

$$\bar{\Xi}_t(x) \text{ is twice continuously differentiable on } \mathbb{R}^d. \quad (24)$$

Step 5: Our next step is to show that (G2) holds, for fix $t \in \mathbb{N}$. To this end, we need to show that the conditions of Theorem 6 are satisfied. Condition (C1) follows from (22). Choose a compact set $\mathcal{K} \subset \mathbb{R}^d$ such that $x(t) \in \text{int}(\mathcal{D}_t)$. Then, Condition (C2) follows from (23) and two applications of Lemma 7. Also, Condition (C3) follows since, in view of (24), $\bar{\Xi}_t(x)$ is continuous, and, for all $0 \neq x \in \mathbb{R}^d$,

$$\lim_{\alpha \rightarrow \infty} \bar{\Xi}_t(\alpha x) = -\infty.$$

Step 6: It remains to be shown that (G4) holds. We have

$$\begin{aligned} &\nabla^2 L_{N,t}^{1/N} \\ &= L_{N,t}^{1/N} \left(\frac{1}{N} \nabla^2 \log L_{N,t} + \frac{1}{N^2} \nabla \log L_{N,t} \nabla^T \log L_{N,t} \right) \\ &= L_{N,t}^{1/N} (\nabla^2 \Xi_{N,t} + \nabla \Xi_{N,t} \nabla^T \Xi_{N,t}). \end{aligned}$$

From (19), w.p.1, there exists $N_0 \in \mathbb{N}$ such that, for all $N \geq N_0$, $\prod_{n=1}^N l_{n,t}(x)$ is bounded by a multivariate Gaussian function, and therefore, so is $L_{N,t}^{1/N}$. Also, $\nabla \log l_n$ has linear growth, and $\nabla^2 \log l_n$ is bounded, hence $\nabla^2 \Xi_{N,t} + \nabla \Xi_{N,t} \nabla^T \Xi_{N,t}$ exhibits quadratic growth. Then, the first part of (G4) holds. For the second part, we have

$$\begin{aligned} \left[\nabla^2 L_{N,t}^{1/N} \right]_{i,j} &= \left[\nabla^2 \exp(\Xi_N) \right]_{i,j} \\ &\leq \max_{x \in \mathbb{R}^d} \exp(\Xi_N(x)) \left[\nabla^2 \Xi_N \right]_{i,j} \\ &\leq \exp(\Xi_N(x_*)) \max_{n \in \mathbb{N}, x \in \mathbb{R}^d} \left[\nabla^2 \log l_n(x) \right]_{i,j} \\ &< \infty. \end{aligned}$$

5. NUMERICAL SIMULATIONS

In this section we compare the tracking performance of the ML filter, with the one obtained using the Bayesian

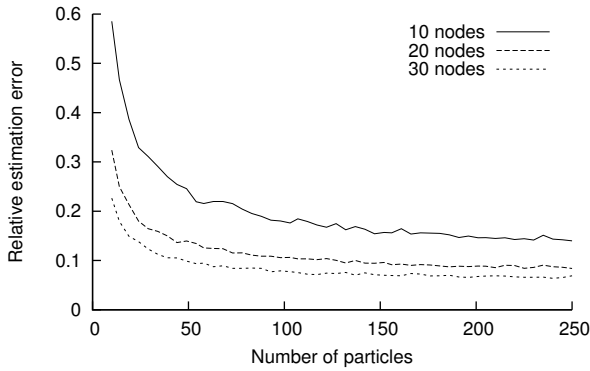


Fig. 1. State estimation error of the particle filter vs. number of particles.

tracker (3)-(4). This comparison is done using the quantized system described in Section 4, with $d = 3$, $R = P = 0.01$, $A = 0.999$ and $\Sigma = 50$. Also, $\mathcal{Q}[\cdot]$ is a one-bit quantizer, i.e.,

$$\mathcal{Q}[\xi] = \begin{cases} q_1, & \xi < 0 \\ q_2, & \xi \geq 0 \end{cases},$$

and each d -dimensional row vector c_n is randomly chosen as $c_n = \tilde{c}_n / \|\tilde{c}_n\|$, with $\tilde{c}_n = [\tilde{c}_{n,1}, \dots, \tilde{c}_{n,d}]^T$ and $\tilde{c}_{n,i}$, $n \in \{1, \dots, N\}$, $i \in \{1, \dots, d\}$, being drawn from the distribution $\mathcal{N}(0, 1)$.

To obtain an approximation of the performance of the Bayesian tracker, we use an importance-sampling particle filter [Arulampalam et al., 2002]. The accuracy of this filter depends on the number of particles used. In order to choose a number of particles such that the approximation error is negligible, we plot in Figure 1 the relative estimation error e_{rel} obtained using the particle filter with different number of particles. This error is defined by

$$e_{\text{rel}} = \frac{\sum_{t=1}^T \|x(t) - \hat{x}(t)\|^2}{\sum_{t=1}^T \|x(t)\|^2},$$

where $\hat{x}(t)$ denotes the estimated state, and we use $T = 10^5$ samples. We see that the improvement obtained using more than 200 particles is rather small. Hence, we use this value in our simulations.

In Figure 2 we show the relative estimation errors obtained using the particle filter and the ML filter, as a function of the number of sensors. Again, we use $T = 10^5$ samples. We see that the difference in performance between both filters is negligible when the number N of sensors is greater than or equal to 20. This is due to the fact that, for these values of N , the difference between the LF (7) and its Gaussian approximation (18) is very small.

To show how the accuracy of the Gaussian approximation to the LF increases with N , we show in Figure 3 the quadratic error e_q between the LF and its Gaussian approximation, defined by

$$e_q^2 = \int \left| L_N(x) - L_N(\mu_N) e^{-\frac{1}{2}(x-\mu_N)^T \Sigma_N^{-1}(x-\mu_N)} \right|^2 dx,$$

for different values of N , when the true value is $x_* = 0$. To evaluate the error, for each point, we use 1000 Monte Carlo runs. Also, Figure 4 shows examples of both functions for N equal to 6, 12 and 18.

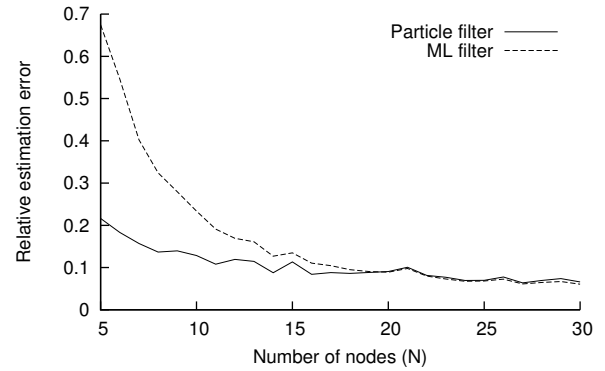


Fig. 2. State estimation error of the particle filter and the ML filter.

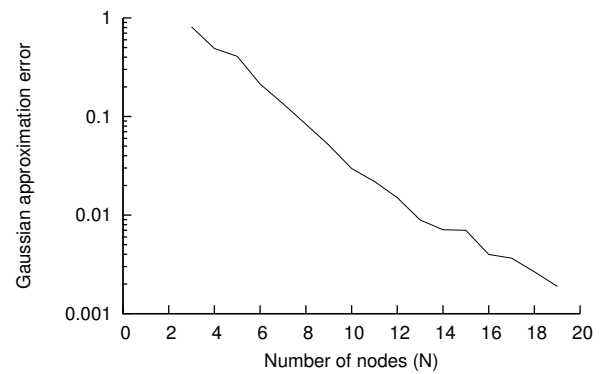


Fig. 3. Quadratic error between the LF and its Gaussian approximation for different numbers of sensors.

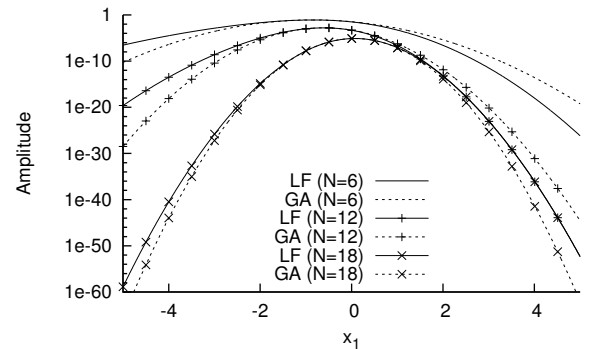


Fig. 4. LF and its Gaussian approximation (GA) for different numbers of sensors.

Finally, an example of the evolution of the first component $x_1(t)$ of the state $x(t)$, together with its estimates obtained using particle filtering and ML filtering, is shown in Figure 5. In this comparison we use $N = 20$ sensors. We see how both estimates closely resemble each other.

6. CONCLUSION

We provide a number of technical conditions to guarantee that the ML filter converges to the theoretical Bayesian tracking solution, as the number of sensors tends to infinity. The implication of this result is that, in tracking applications using a large number of sensors, the computational advantages offered by the ML filter, a guarantee

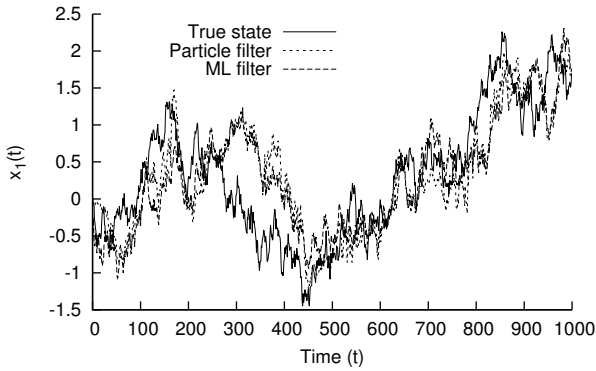


Fig. 5. Evolution of the first component $x_1(t)$ of the state $x(t)$, and its two estimates obtained using particle filtering and ML filtering, for $N = 20$ nodes.

for the stability of the tracker, as well as the possibility of a distributed implementation, come without noticeable performance detriment. We used our result in a case study, and presented simulation results confirming our theoretical claim.

Appendix A. LEMMAS

Lemma 7. Let $\mathcal{D} \subset \mathbb{R}^d$ be compact and $f_n : \mathcal{D} \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of stochastic differentiable functions. If $\lim_{n \rightarrow \infty} f_n(x)$ converges for every $x \in \mathcal{D}$, w.p.1, and ∂f_n is CSUC, then:

- (1) f_n is CSUC;
- (2) $\partial \lim_{n \rightarrow \infty} \mathcal{E} \{f_n\} = \lim_{n \rightarrow \infty} \mathcal{E} \{\partial f_n\}$.

Proof. Claim 1 follows by applying [Rudin, 1976, Th. 7.17] on each event where the pointwise convergence of $\lim_{n \rightarrow \infty} f_n$ and the uniform convergence of ∂f_n occur. On the same event we have

$$\lim_{n \rightarrow \infty} \mathcal{E} \left\{ \frac{\partial}{\partial x} f_n(x) \right\} = \lim_{n \rightarrow \infty} \frac{\partial}{\partial x} f_n(x) \stackrel{(a)}{=} \frac{\partial}{\partial x} \lim_{n \rightarrow \infty} f_n(x) = \frac{\partial}{\partial x} \lim_{n \rightarrow \infty} \mathcal{E} \{f_n(x)\},$$

where (a) follows from [Rudin, 1976, Th. 7.17], and Claim 2 follows.

Lemma 8. If $f_n : \mathbb{R}^d \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, is SAE, and converges w.p.1, then it is CSUC.

Proof. Clearly, the SAE condition implies that f_n is continuous, for each $n \in \mathbb{N}$. Then, the result follows from [Davidson, 1994, Th. 21.8].

Lemma 9. If $f_n : \mathbb{R}^d \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, is SAE, then so is $F_N = \frac{1}{N} \sum_{n=1}^N f_n$.

Proof. Choose $\epsilon > 0$. We have

$$\begin{aligned} & \sup_{x \in \mathbb{R}^d} \sup_{y: \|y-x\| < \delta_{\epsilon/2}} \|F_N(y) - F_N(x)\| \\ & \leq \frac{1}{N} \sum_{n=1}^N \sup_{x \in \mathbb{R}^d} \sup_{y: \|y-x\| < \delta_{\epsilon/2}} \|f_n(y) - f_n(x)\|. \quad (\text{A.1}) \end{aligned}$$

Let $N_\epsilon \in \mathbb{N}$ be such that

$$\sup_{x \in \mathbb{R}^d} \sup_{y: \|y-x\| < \delta_{\epsilon/2}} \|f_n(y) - f_n(x)\| < \epsilon,$$

for all $n > N_\epsilon$. Then, the result follows by taking lim sup on both sides of (A.1).

REFERENCES

- BDO Anderson and JB Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.
- I. Arasaratnam and S. Haykin. Cubature kalman filters. *IEEE Transactions on Automatic Control*, 54(6):1254–1269, june 2009.
- M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188, 2002.
- R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications*, 26(4):622–633, 2008.
- Lubin Chang, Baiqing Hu, An Li, and Fangjun Qin. Transformed unscented kalman filter. *IEEE Transactions on Automatic Control*, 58(1):252–257, jan. 2013.
- Kai Lai Chung. *A Course in Probability Theory*. Academic Press, 3 edition, 2000.
- Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.
- Waltenegus Dargie and Christian Poellabauer. *Fundamentals of Wireless Sensor Networks: Theory and Practice*. Wiley, 2010.
- James Davidson. *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford university press, 1994.
- C. Gourieroux and A. Monfort. *Statistics and Econometric Models, Volume 2*. Cambridge University Press, 1996.
- O. Hlinka, O. Sluciak, F. Hlawatsch, P.M. Djuric, and M. Rupp. Likelihood consensus and its application to distributed particle filtering. *IEEE Transactions on Signal Processing*, 60(8):4334–4349, 2012.
- U.A. Khan and J.M.F. Moura. Distributing the kalman filter for large-scale systems. *IEEE Transactions on Signal Processing*, 56(10):4919–4935, 2008.
- J.H. Kotecha and P.M. Djuric. Gaussian sum particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2602–2612, 2003.
- Whitney K Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.
- Kaare Brandt Petersen and Michael Syskind Pedersen. *The matrix cookbook*, 2006.
- A. Ribeiro, I. Schizas, S. Roumeliotis, and G. Giannakis. Kalman filtering in wireless sensor networks. *IEEE Control Systems*, 30(2):66–86, 2010.
- Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Science/Engineering/Math, 3rd edition, 1976.
- Miroslav Šimandl, Jakub Královec, and Torsten Söderström. Advanced point-mass method for nonlinear state estimation. *Automatica*, 42(7):1133–1145, 2006.
- Xingbo Wang, Minyue Fu, and Huanshui Zhang. Target tracking in wireless sensor networks based on the combination of kf and mle using distance measurements. *Mobile Computing, IEEE Transactions on*, 11(4):567–576, 2012.