Asymptotic Optimality of the Maximum-Likelihood Kalman Filter for Bayesian Tracking With Multiple Nonlinear Sensors

Damián Marelli, Minyue Fu, *Fellow, IEEE*, and Brett Ninness, *Senior Member, IEEE*

***Abstract*—*Bayesian tracking* is a general technique for state estimation of nonlinear dynamic systems, but it suffers from the drawback of computational complexity. This paper is concerned with a class of Wiener systems with multiple nonlinear sensors. Such a system consists of a linear dynamic system followed by a set of static nonlinear measurements. We study a *maximum-likelihood Kalman filtering* (MLKF) technique which involves maximum-likelihood estimation of the nonlinear measurements followed by classical Kalman filtering. This technique permits a distributed implementation of the Bayesian tracker and guarantees the boundedness of the estimation error. The focus of this paper is to study the extent to which the MLKF technique approximates the theoretically optimal Bayesian tracker. We provide conditions to guarantee that this approximation becomes asymptotically exact as the number of sensors becomes large. Two case studies are analyzed in detail.**

***Index Terms*—Bayesian tracking, distributed estimation, maximum likelihood, sensor networks, Wiener systems.**

I. Introduction

SENSOR networks find applications in environmental, health-care and weather monitoring, industrial process monitoring and control, surveillance, smart grids and so on [1]. The development of algorithms for Bayesian tracking in sensor networks has recently attracted a great deal of attention [2]–[5]. In a standard Bayesian tracking procedure, the measurements from all sensors are combined to track the evolution of a set of stochastic parameters (i.e., the state vector) for which a dynamic model is available. In many applications, the model describing the dynamic evolution of the state vector can be represented or approximated by a linear *state equation*, but the *measurement equation* of each sensor is often nonlinear. Such a system is known as a Wiener system [6], [7]. Examples of Wiener systems include: measurements with coarse quantization (including on-off detectors as a special case) [8]–[10]; target tracking using distance sensing [11], [12], sensor network localization using distance and/or angle measurements [13], [14], the traditional state estimation for power systems using nonlinear measurements, known as supervisory data acquisition (SCADA) measurements [15], [16], and many others.

Many Bayesian tracking algorithms are available for general nonlinear models [17]–[21]. 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 [22], inaccurate to the extent that they can lead to instability. Motivated by this, the authors of [11], [12] proposed a sub-optimal Bayesian tracking technique which is particularly suitable for Wiener dynamical models. In this technique, which is based on intuition and empirical observation, the prediction step is the same as the one used in the Kalman filter, but the update step is replaced with a maximum-likelihood (ML) estimator of the state, or a partial state, using the nonlinear measurements. In this paper we refer to this Bayesian tracking algorithm as the *maximum-likelihood Kalman filter (MLKF)*. A particular advantage of the MLKF is its guaranteed stability, i.e., the state estimation error is guaranteed to be bounded. A computational drawback of this method is that it requires solving a ML estimation problem at each update step. However, the complexity of this task is largely simplified by initializing the associated numerical optimization algorithm with the outcome of the previous prediction step. In this way, the MLKF offers a numerically efficient Bayesian tracking implementation. In addition, in contrast to other Bayesian tracking methods, the MLKF is suitable not only for centralized implementation but also for distributed implementation where the update step is done using a distributed ML algorithm [5].

In this paper we provide conditions under which the MLKF becomes asymptotically optimal in the sense of approaching the theoretical Bayesian tracking solution, as the number of sensors becomes large. This theoretical result provides a rigorous justification of the intuition-motivated MLKF. Its implication is to turn the MLKF into an advantageous alternative to other Bayesian tracking algorithms, for applications using a large number of sensors (each of which providing possibly limited information about the state vector to be estimated). To illustrate the application of our result, we show the asymptotic optimality of the MLKF for two case studies. The first one is the tracking of a moving target by a sensor network using

Manuscript received October 02, 2014; revised February 21, 2015; accepted May 13, 2015. Date of publication June 04, 2015; date of current version July 21, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Y.-W. Peter Hong. This work was supported in part by the Austrian Science Fund (FWF, Project M1230-N13).

D. Marelli is with the School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia, and also with the Acoustics Research Institute, Austrian Academy of Sciences, Austria (e-mail: Damian.Marelli@newcastle.edu.au).

M. Fu is with the School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia, and also with the Department of Control Science and Engineering, Zhejiang University, China (e-mail: Minyue.Fu@newcastle.edu.au).

B. Ninness is with the School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: Brett.Ninness@newcastle.edu.au).

Digital Object Identifier 10.1109/TSP.2015.2440220

the distance measurements from each sensor to the target. The second case study is state estimation for a linear dynamic system using measurements with coarse quantization. In both cases, numerical experiments show that the difference between the MLKF and an approximation to the theoretically optimal Bayesian tracker (obtained using particle filtering), becomes negligible even for a relatively small number of sensors.

The rest of the paper is organized as follows: In Section II we describe the standard Bayesian tracker and the MLKF, followed by the problem formulation. In Section III, we introduce our main result on asymptotic optimality of the MLKF. In Sections IV–V, we apply the result above to the aforementioned two case studies. We give concluding remarks in Section VI. The proof of our main result is contained in Appendix. A shortened version of this work appears in the conference paper [23].

Notation: The symbols $\mathbb{N}$ and $\mathbb{R}$ denote the sets of natural and real numbers, respectively. Also, $L_p(\mathbb{R}^d)$ denotes the set of functions whose p-th power is absolutely integrable, $C(\mathbb{R}^d)$ denotes the space of continuous functions with domain in $\mathbb{R}^d$, and $C_n(\mathbb{R}^d)$ denotes the space of functions with n-th order continuous derivatives. 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 distribution function of mean μ and covariance matrix Σ is denoted by $g_{\mu,\Sigma}$, and its associated cumulative distribution function by $G_{\mu,\Sigma}$, i.e.,

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

For a function $f \in C_n(\mathbb{R})$, $\partial^n f$ denotes its n-th order derivative. Finally, for $f \in C_2(\mathbb{R}^d)$, $\nabla f(x)$ denotes its gradient and $\nabla^2 f(x)$ its Hessian.

II. Problem Description

Consider the following Wiener system with multiple measurement sensors:

$$s(t+1) = As(t) + w(t), \quad t \in \mathbb{N}, \tag{1}$$
$$x(t) = Cs(t), \tag{2}$$
$$y_n(t) = d_n(x(t)), n = 1, \ldots, N, \tag{3}$$

where $s(t) \in \mathbb{R}^p$ is the state with initial value $s(1) \sim \mathcal{N}(0, P)$, $x(t) \in \mathbb{R}^d$ is the linear output, $y_n(t) \in \mathbb{R}^{r_n}$ is the measurement of the n-th sensor and $w(t) \in \mathbb{R}^p$ is the process noise with distribution $\mathcal{N}(0, R)$. The vector $x(t)$ denotes the components of $s(t)$ which are involved in the non-linear measurements. Hence, without loss of generality, it is assumed that C has full row rank. It is also assumed that $R > 0$ and $P > 0$. To account for measurement noises, the measuring functions d_n are assumed to be *stochastic functions* (i.e., for each x, $d_n(x)$ is a random vector). The properties of the measuring functions will be discussed later. In the sequel, we will denote

$$Y_N(t) = [y_1^T(t), \ldots, y_N^T(t)]^T,$$
$$\mathbf{Y}_{N,t} = [Y_N^T(1), \ldots, Y_N^T(t)]^T.$$

1) Bayesian Tracker: The well-known Bayesian tracker is a general technique for optimal state estimation; see, e.g., [19]. It is a recursive procedure for computing (or tracking) the conditional probability density function $\mathcal{P}(s(t)|\mathbf{Y}_{N,t})$, i.e., the probability density function of $s(t)$ conditioned on the prior distribution of $s(1)$ and measurements $\mathbf{Y}_{N,t}$. The procedure is as follows:

Bayesian Tracking Procedure: Set $\mathcal{P}(s(1)|\mathbf{Y}_{N,0}) = g_{0,P}(s(1))$. Then, for each $t \in \mathbb{N}$, do the following:

---

1) **Update:**

$$\mathcal{P}(s(t)|\mathbf{Y}_{N,t}) = \frac{\mathcal{P}(Y_N(t)|s(t))}{\mathcal{P}(Y_N(t)|\mathbf{Y}_{N,t-1})}\mathcal{P}(s(t)|\mathbf{Y}_{N,t-1}). \tag{4}$$

2) **Prediction:**

$$\mathcal{P}(s(t+1)|\mathbf{Y}_{N,t}) = \int_{\mathbb{R}^p} \mathcal{P}(s(t+1)|s(t))\mathcal{P}(s(t)|\mathbf{Y}_{N,t})ds(t). \tag{5}$$

In the Update step, the term $\mathcal{P}(s(t)|\mathbf{Y}_{N,t-1})$ is given by the previous iteration and the term $\mathcal{P}(Y_N(t)|\mathbf{Y}_{N,t-1})$ can be computed using the fact that $\mathcal{P}(s(t)|\mathbf{Y}_{N,t})$ integrates to one, i.e.,

$$\mathcal{P}(Y_N(t)|\mathbf{Y}_{N,t-1}) = \int_{\mathbb{R}^p} \mathcal{P}(Y_N(t)|s)\mathcal{P}(s|\mathbf{Y}_{N,t-1})ds. \tag{6}$$

The key term to be studied is $\mathcal{P}(Y_N(t)|s(t))$, which is computed using the stochastic characteristics of the measuring functions. Let

$$H_{N,t}(s) = \mathcal{P}(Y_N(t)|s(t) = s)$$

be the *likelihood function* of $s(t) = s$, given $Y_N(t)$. In the event that the measurements $y_n(t)$ are linear on $x(t)$, with additive Gaussian noises, it is clear that, $H_{N,t}(s)$ is Gaussian. This leads to the well-known Kalman filter [22]. In the general case, the functions d_n may not be Gaussian and linear in $x(t)$, and consequently, the Bayesian tracker is difficult to evaluate. The purpose of this paper is to study the conditions under which the functions $H_{N,t}(s)$ are approximately Gaussian, i.e.,

$$H_{N,t}(s) \simeq a_N(t) \times \exp\left(-\frac{1}{2}(s-\nu_N(t))^T\Upsilon_N(t)(s-\nu_N(t))\right), \tag{7}$$

for some $a_N(t) > 0$, $\nu_N(t) \in \mathbb{R}^p$ and $0 < \Upsilon_N(t) \in \mathbb{R}^{p\times p}$, which depend on $Y_N(t)$ (Notice that the value of $a_N(t)$ can be inferred from (6)). If so, the Bayesian tracking procedure can be approximated by one similar to a Kalman filter. More precisely, from (4) and (7), it follows that

$$\mathcal{P}(s(t)|\mathbf{Y}_{N,t}) = g_{\nu_{t|t},\Psi_{t|t}}(s(t)), \tag{8}$$

with

$$\nu_{t|t} = \Psi_{t|t}\left(\Psi_{t|t-1}^{-1}\nu_{t|t-1} + \Upsilon_N(t)\nu_N(t)\right), \tag{9}$$
$$\Psi_{t|t} = \left(\Psi_{t|t-1}^{-1} + \Upsilon_N(t)\right)^{-1}, \tag{10}$$

and

$$\mathcal{P}(s(t+1)|\mathbf{Y}_{N,t}) = g_{\nu_{t+1|t},\Psi_{t+1|t}}(s(t+1)), \quad (11)$$

with

$$\nu_{t+1|t} = A\nu_{t|t}, \quad (12)$$
$$\Psi_{t+1|t} = A\Psi_{t|t}A^T + R. \quad (13)$$

In principle, we would adopt a Gaussian approximation in which $\nu_N(t)$ is taken as the argument that maximizes $H_{N,t}(s)$. However, depending on the matrix C, it may occur that $H_{N,t}(s)$ is maximized at an infinite set of points. To go around this, we define the likelihood function of $x(t) = x$, given $Y_N(t)$, i.e.,

$$L_{N,t}(x) = \mathcal{P}(Y_N(t)|x(t) = x), \quad (14)$$

and its (normalized) *log-likelihood function*

$$\Xi_{N,t}(x) = \frac{1}{N}\log L_{N,t}(x). \quad (15)$$

Suppose that

$$L_{N,t}(x) \simeq a_N(t)\times \exp\left(-\frac{1}{2}(x-\mu_N(t))^T\Sigma_N^{-1}(t)(x-\mu_N(t))\right), \quad (16)$$

for some $\mu_N(t) \in \mathbb{R}^d$ and $0 < \Sigma_N(t) \in \mathbb{R}^{d\times d}$. Then, (7) holds

$$\nu_N(t) = C^\dagger\mu_N(t), \quad (17)$$
$$\Upsilon_N(t) = (C^\dagger\Sigma_N(t)C^{\dagger T})^\dagger = C^T\Sigma_N^{-1}(t)C, \quad (18)$$

where $C^\dagger$ denotes the Moore-Penrose pseudoinverse of C [24], and the last equality follows because C has full row rank.

We then adopt the Gaussian approximation yield by (19)–(20), and obtain the following algorithm:

Maximum Likelihood Kalman Filter (MLKF): Set $\nu_{1|0} = 0$ and $\Psi_{1|0} = P$. Then, for each $t \in \mathbb{N}$, do the following:

1) **ML estimation:** Choose

$$\mu_N(t) = \arg\max_{x\in\mathbb{R}^d}\Xi_{N,t}(x), \quad (19)$$
$$\Sigma_N(t) = -\frac{1}{N}(\nabla^2\Xi_{N,t}(\mu_N(t)))^{-1}. \quad (20)$$

The numerical optimization algorithm used to solve **(19)** is initialized by taking $x = C\nu_{t|t-1}$.

2) **Update:**

$$\mathcal{P}(s(t)|\mathbf{Y}_{N,t}) = g_{\nu_{t|t},\Psi_{t|t}}(s(t))$$

with **(9)–(10)**.

3) **Prediction:**

$$\mathcal{P}(s(t+1)|\mathbf{Y}_{N,t}) = g_{\nu_{t+1|t},\Psi_{t+1|t}}(s(t+1))$$

with **(12)–(13)**.

Remark 1: Notice that, in view of (17)–(18), the update step (9)–(10) can be written as

$$\nu_{t|t} = \Psi_{t|t}\left(\Psi_{t|t-1}^{-1}\nu_{t|t-1} + C^T\Sigma_N^{-1}(t)\mu_N(t)\right),$$
$$\Psi_{t|t} = (\Psi_{t|t-1}^{-1} + C^T\Sigma_N^{-1}(t)C)^{-1}.$$

Hence, we can interpret it as the Kalman filter update, in information form [22, Section 6.3] when the output equation is given by

$$\mu_N(t) = Cs(t) + v(t),$$

where $v \sim \mathcal{N}(0, \Sigma_N(t))$. Hence, the same step can be expressed as a standard (i.e., covariance form) Kalman filter step, as follows

$$\nu_{t|t} = \nu_{t|t-1} + K_t(\mu_N(t) - C\nu_{t|t-1}),$$
$$\Psi_{t|t} = (I - K_tC)\Psi_{t|t-1},$$

with

$$K_t = \Psi_{t|t-1}C^T(C\Psi_{t|t-1}C^T + \Sigma_N(t))^{-1}.$$

2) Problem Formulation: The question we try to answer in this paper is when the MLKF approaches to the Bayesian tracker as the number of sensors becomes large. This is equivalent to ask whether the approximate equality (16) becomes asymptotically exact as $N \to \infty$. In the remainder of this work we will be concerned in finding conditions under which this property holds.

III. Main Result

In this section, we provide conditions under which the approximation (16) with (19)–(20) becomes asymptotically exact, as the number N of measurements tends to infinity. Since the study of the approximation (16) is independent of the time index t, we drop this index to simplify the notation. Also, we denote the true value of $x(t)$ by $x_\star \in \mathbb{R}^d$.

A. Main Theorem

Our aim is to show that the sequence of likelihood functions L_N, $N \in \mathbb{N}$, given by (14), converges to a Gaussian function. We remark that this property does not follow from the well-known central limit theorem, although it may be tempting to draw a parallel between them. Notice that L_N depends on the random observations $Y_N = \left[y_1^T(t), \ldots, y_N^T(t)\right]^T$. Hence, $L_N(x)$, $N \in \mathbb{N}$, is a sequence of random functions, and our goal is to show that the desired convergence holds w.p.1. The first step towards this end is given in Lemma 2, which provides conditions to guarantee that the point-wise limit of a sequence of deterministic functions of one variable (i.e., from $\mathbb{R}$ to $\mathbb{R}$) equals a normalized Gaussian function. Its proof is partially based on the proof of the central limit theorem of Lindeberg-Feller [25, Th. 15.43]. Then, Corollary 3 generalizes this result for sequences of deterministic functions of many variables (i.e., from $\mathbb{R}^d$ to $\mathbb{R}$). Finally, the desired result is stated in Theorem 6.

Lemma 2: Let $\phi_N \in C_2(\mathbb{R})$, $N \in \mathbb{N}$, and $\psi_N = -N\partial^2\phi_N^{1/N}$. If

1) $\phi_N(0) = 1$ and $\partial\phi_N(0) = 0$;
2) $\psi_N \in L_2(\mathbb{R})$ *for all* $N \in \mathbb{N}$;
3) $\sup_{N\in\mathbb{N}}\|\psi_N\|_\infty < \infty$;

4) $\lim_{N\to\infty}\psi_N(x) = 1$ uniformly on any compact set $\mathcal{K} \subset \mathbb{R}$;

then, for any $t \in \mathbb{R}$,

$$\lim_{N\to\infty}\phi_N(t) = e^{-\frac{t^2}{2}}.$$

Corollary 3: Let $\Phi_N : \mathbb{R}^d \to \mathbb{R}$, $N \in \mathbb{N}$. If, for each $y \in \mathbb{R}^d$, with $\|y\| = 1$, the sequence of functions $\phi_{y,N} : \mathbb{R} \to \mathbb{R}$, $N \in \mathbb{N}$, defined by $\phi_{y,N}(t) = \Phi_N(ty)$, satisfies the conditions of Lemma 2, then, for any $x \in \mathbb{R}^d$,

$$\lim_{N\to\infty}\Phi_N(x) = e^{-\frac{x^Tx}{2}}.$$

We now introduce the definitions of *strong convergence* and *strong uniform convergence* [26, S 21.2]. They are used in the statement of Theorem 6.

Definition 4: A sequence v_n, $n \in \mathbb{N}$, of random variables is said to be *strongly convergent* (SC) if $\lim_{n\to\infty}\mathcal{E}\{v_n\}$ exists and

$$\lim_{n\to\infty} v_n \overset{\text{w.p.1}}{=} \lim_{n\to\infty}\mathcal{E}\{v_n\}. \tag{21}$$

Definition 5: Let $\mathcal{D} \subseteq \mathbb{R}^d$. A sequence of stochastic functions $f_n : \mathcal{D} \to \mathbb{R}^q$, $n \in \mathbb{N}$, is said to be *continuous and strongly uniformly convergent* (CSUC) in $\mathcal{D}$ if, w.p.1, every f_n is a continuous function on $\mathcal{D}$, and

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

for some deterministic function $f : \mathcal{D} \to \mathbb{R}^q$.

We are now ready to state our main result.

Theorem 6: Let $\Xi_N : \mathbb{R}^d \to \mathbb{R}$, $N \in \mathbb{N}$, be a sequence of stochastic functions, and μ_N and Σ_N be defined as in (19) and (20), respectively. Suppose that the following conditions hold:

(G1) there exists a compact connected set $\mathcal{D} \subset \mathbb{R}^d$, such that Ξ_N is twice continuously differentiable on $\mathcal{D}$ and $\nabla^2\Xi_N$ is CSUC in $\mathcal{D}$;

(G2) $\Xi_N(x)$, $\nabla\Xi_N(x)$ and $\nabla^2\Xi_N(x)$ are SC, for all $x \in \mathcal{D}$;

(G3) there exists $x_\star \in \text{int}(\mathcal{D})$ (i.e., the interior of $\mathcal{D}$ excluding its boundary), such that $\lim_{N\to\infty}\mu_N \overset{\text{w.p.1}}{=} x_\star$;

(G4) $\nabla^2\overline{\Xi}(x_\star) < 0$, where[1]

$$\overline{\Xi}(x) = \lim_{N\to\infty}\mathcal{E}\{\Xi_N(x)\}. \tag{22}$$

Then,

$$\lim_{N\to\infty}\frac{1}{L_N(\mu_N)}L_N\left(\Sigma_N^{1/2}x + \mu_N\right) \overset{\text{w.p.1}}{=} \exp\left(-\frac{x^Tx}{2}\right). \tag{23}$$

Remark 7: Theorem 6 states that, under the conditions (G1)-(G4),

$$L_N(x) \overset{\text{w.p.1}}{\simeq} L_N(\mu_N)\exp\left(-\frac{1}{2}(x-\mu_N)^T\Sigma_N^{-1}(x-\mu_N)\right) \tag{24}$$

for large N. That is, the approximation (16), with (19)–(20) becomes valid as N becomes large, and therefore, so does the MLKF.

The result in Theorem 6 requires the technical conditions (G1)-(G4). These conditions follow the style used in other classical results in asymptotic statistics (e.g., conditions to guarantee strong consistency [27, Property 24.2], [28, Theorem 3.3], [29, Theorem 2.1] and asymptotic normality [27, Property 24.16], [28, Theorem 5.1], [29, Theorem 3.1] of extremum estimators). The use of these conditions permit stating our result in a very general form, making it applicable to a large class of measurement functions, modeling nonlinearities and measurement errors. For this reason, the conditions are quite technically involved. The disadvantage of this generality is that some technical effort may be required to verify these conditions for a given problem setting. In the next three subsections, we describe some technical tools which can be used to help verify these conditions. We end this section by introducing, in Subsection III-E , an example where the convergence in (23) fails, because Conditions (G1)-(G4) are not satisfied.

B. About Condition (G1)

The purpose here is of twofold. We first state in Lemma (8) a key property of CSUC sequences. We then introduce in Lemma 10 a more verifiable condition to guarantee the CSUC condition, which is required by condition (G1).

Lemma 8: Let $\mathcal{D} \subset \mathbb{R}^d$ be compact and connected, and $f_n : \mathcal{D} \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of differentiable stochastic functions. If $f_n(x)$ and $\nabla f_n(x)$ are SC for every $x \in \mathcal{D}$, and ∇f_n is CSUC, then

1) f_n is CSUC;
2) $\nabla\lim_{n\to\infty}\mathcal{E}\{f_n\} = \lim_{n\to\infty}\mathcal{E}\{\nabla f_n\}$.

The proof of Lemma 8 requires the following lemma.

Lemma 9: (Immediate consequence of [30, Th. 7.17]). Let $\mathcal{D} \subset \mathbb{R}$ be compact and connected, and $f_n : \mathcal{D} \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of differentiable functions. If $f_n(x_0)$ converges at some point $x_0 \in \mathcal{D}$ and $\nabla f_n(x)$ converges uniformly on $\mathcal{D}$, then $f_n(x)$ converges uniformly on $\mathcal{D}$, and

$$\nabla\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty}\nabla f_n(x).$$

Proof of Lemma 8: Fix $x \in \mathcal{D}$. Let A_x be the event (i.e., the set on the underlying probability space) where the SC condition on $f_n(x)$ holds, and B be the event where the CSUC condition on ∇f_n holds. Claim 1 follows since the conditions of Lemma 9 hold on the event $C_x = A_x \cap B$, which has probability one. Also, since $\nabla f_n(x)$ is SC, there exists an event D_x, having probability one, on which $\lim_{n\to\infty}\nabla f_n(x) = \lim_{n\to\infty}\mathcal{E}\{\nabla f_n(x)\}$, for any $x \in \mathcal{D}$. Hence, the same condition must hold on the event $E_x = B \cap D_x$. Then, on the event $C_x \cap E_x$, we have

$$\begin{aligned}\lim_{n\to\infty}\mathcal{E}\{\nabla f_n(x)\} &= \lim_{n\to\infty}\nabla f_n(x)\\ &\overset{\text{(a)}}{=} \nabla\lim_{n\to\infty} f_n(x)\\ &= \nabla\lim_{n\to\infty}\mathcal{E}\{f_n(x)\},\end{aligned}$$

where (a) follows from Lemma 9. Then, Claim 2 follows since the event $C_x \cap E_x$ has probability one. ■

[1]Notice that (G2) guarantees the existence of the limit in (22), and (G1), together with Lemma 8-2) guarantees that $\overline{\Xi}$ is twice differentiable on $x_\star$.

The following result can be used for guaranteeing the CSUC condition.

Lemma 10: Let $\mathcal{D} \subset \mathbb{R}$ be compact and $f_n : \mathcal{D} \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of differentiable functions. If $f_n(x)$ is SC for every $x \in \mathcal{D}$, and there exists $M > 0$ such that

$$\sup_{n\in\mathbb{N},x\in\mathcal{D}} \|\partial f_n(x)\| \overset{\text{w.p.1}}{\leq} M, \tag{25}$$

then, f_n, $n \in \mathbb{N}$, is CSUC on $\mathcal{D}$.

Proof: Let $f(x) = \lim_{n\to\infty} \mathcal{E}\{f_n(x)\}$ and $\tilde{f}_n(x) = f_n(x) - f(x)$. We have

$$\left|\tilde{f}_n(x) - \tilde{f}_n(y)\right| \leq |f_n(x) - f_n(y)| + |f(x) - f(y)|.$$

Choose $\delta > 0$ and a realization of f_n satisfying (25). We have that

$$\begin{aligned}&\sup_{n\in\mathbb{N},x\in\mathcal{D}} \sup_{y\in\mathcal{D},\|y-x\|<\delta} |f_n(x) - f_n(y)|\\ &\leq \delta \sup_{n\in\mathbb{N},x\in\mathcal{D}} \|\nabla f_n(x)\|\\ &\leq \delta M.\end{aligned}$$

Also, for all $x, y \in \mathcal{D}$,

$$\begin{aligned}|f(x) - f(y)| &\overset{\text{w.p.1}}{=} \lim_{n\to\infty} |f_n(x) - f_n(y)|\\ &\leq \|x - y\|M.\end{aligned}$$

Hence

$$\sup_{x\in\mathcal{D}} \sup_{y\in\mathcal{D},\|y-x\|<\delta} |f(x) - f(y)| \leq \delta M.$$

Then,

$$\begin{aligned}&\sup_{n\in\mathbb{N},x\in\mathcal{D}} \sup_{y\in\mathcal{D},\|y-x\|<\delta} \left|\tilde{f}_n(x) - \tilde{f}_n(y)\right|\\ &\leq \sup_{n\in\mathbb{N},x\in\mathcal{D}} \sup_{y\in\mathcal{D},\|y-x\|<\delta} |f_n(x) - f_n(y)|\\ &\quad + \sup_{x\in\mathcal{D}} \sup_{y\in\mathcal{D},\|y-x\|<\delta} |f(x) - f(y)|\\ &\leq 2\delta M.\end{aligned} \tag{26}$$

Since the event formed by realizations satisfying (26) has probability one, the sequence $\tilde{f}_n$, $n \in \mathbb{N}$, is strongly stochastically equicontinuous [26, eq. (21.43)]. Also, $\lim_{n\to\infty} \tilde{f}_n(x) \overset{\text{w.p.1}}{=} 0$. Hence, from [26, Th. 21.8], $\tilde{f}_n$ it is strongly uniformly convergent, and therefore so is f_n. Finally, the existence of the gradient implies that f_n is continuous on $\mathcal{D}$, for each $n \in \mathbb{N}$, and the result follows. ∎

C. About Condition (G4)

Definition 11: The Kullback-Leibler distance between the probability distributions p and q is defined by [31]

$$D_{\mathrm{KL}}(p\|q) = \int \log \frac{p(\Xi)}{q(\Xi)} p(\Xi) d\Xi.$$

We then have the following result.

Lemma 12: If Condition *(G1)* holds, and

$$\lim_{N\to\infty} \frac{1}{N} D_{\mathrm{KL}}(\mathcal{P}(Y_N|x_\star)\|\mathcal{P}(Y_N|x)) > 0, \tag{27}$$

for all $x \in \mathcal{D} \setminus \{x_\star\}$, then Condition *(G4)* holds.

Proof: Let $F(x) = \overline{\Xi}(x_\star) - \overline{\Xi}(x)$. We have

$$\begin{aligned}F(x) &= \lim_{N\to\infty} \frac{1}{N} \mathcal{E}\left\{\log \frac{L_N(x_\star)}{L_N(x)}\right\}\\ &= \lim_{N\to\infty} \frac{1}{N} D_{\mathrm{KL}}(\mathcal{P}(Y_N|x_\star)\|\mathcal{P}(Y_N|x)).\end{aligned}$$

From [31, Th. 2.6.3], $F(x) \geq 0$, and clearly $F(x_\star) = 0$. Then, from (27), it follows that $\overline{\Xi}(x)$ has a unique maximum in $\mathcal{D}$ at $x_\star$. Also, from Condition (G1), two applications of Lemma 8 gives that $\overline{\Xi}(x)$ is twice continuously differential, and Condition (G4) follows. ∎

D. About Condition (G3)

Condition (G3) of Theorem 6 requires that μ_N converges w.p.1 to the true value $x_\star$. This kind of convergence is called strong consistency. Conditions to guarantee strong consistency typically require that the maximization problem used to find μ_N is constrained to a known compact set [27, Property 24.2], [29, Th. 2.1]. However, this assumption is unsuitable in our context because μ_N can be anywhere in $\mathbb{R}^d$. To go around this difficulty, we introduce the following variant of the strong consistency results cited above.

Lemma 13: Suppose that the following condition holds w.p.1: there exists a compact set $\mathcal{K} \subset \mathbb{R}^d$ (which may depend on the realization of Ξ_N), with $x_\star \in \mathrm{int}(\mathcal{K})$, such that, on $\mathcal{K}$, Ξ_N is continuous and converges uniformly, and

$$\arg\max_{x\in\mathcal{K}} \overline{\Xi}(x) = \{x_\star\}, \tag{28}$$

$$\limsup_{N\to\infty} \sup_{x\in\mathcal{K}^{\complement}} \Xi_N(x) < \overline{\Xi}(x_\star), \tag{29}$$

then, condition *(G3)* holds. In the above, $\mathcal{K}^{\complement}$ is the complement set of $\mathcal{K}$ in $\mathbb{R}^d$.

Proof: Consider a realization of Ξ_N, together with its associated set $\mathcal{K}$, satisfying (28)–(29). Since Ξ_N is continuous and converges uniformly on $\mathcal{K}$, we have that $\overline{\Xi}$ is continuous in $\mathcal{K}$. Let $\mathcal{L} \subseteq \mathbb{R}^d$ be an open set with $x_\star \in \mathcal{L}$. We have that $\mathcal{L}^{\complement} \cap \mathcal{K}$ is closed and contained in $\mathcal{K}$. Hence, there exists $\tilde{x} \in \mathcal{L}^{\complement} \cap \mathcal{K}$ such that,

$$\sup_{x\in\mathcal{L}^{\complement}\cap\mathcal{K}} \overline{\Xi}(x) = \overline{\Xi}(\tilde{x}). \tag{30}$$

Let $\eta = \overline{\Xi}(x_\star) - \overline{\Xi}(\tilde{x})$ and $\varepsilon = \overline{\Xi}(x_\star) - \limsup_{N\to\infty} \sup_{x\in\mathcal{K}^{\complement}} \Xi_N(x)$. From (28)–(29), we have that $\eta, \varepsilon > 0$. Choose $0 < \epsilon < \min(\eta, \varepsilon)$. From the uniform convergence of Ξ_N, we can choose $N_0 \in \mathbb{N}$ such that, for all $N \geq N_0$,

$$\begin{aligned}&\sup_{x\in\mathcal{K}} \left|\Xi_N(x) - \overline{\Xi}(x)\right| < \frac{\epsilon}{2},\\ &\sup_{x\in\mathcal{K}^{\complement}} \overline{\Xi}(x_\star) - \Xi_N(x) > \frac{\epsilon}{2}.\end{aligned} \tag{31}$$

Then, for all $N \geq N_0$, $\mu_N \in \mathcal{K}$. Hence, it follows from (31) that, for all $N \geq N_0$, see the equation at the bottom of the next page. Then, in view of (30), $\mu_N \in \mathcal{L}$, for all $N \geq N_0$. Since this holds for any open set $\mathcal{L}$ containing $x_\star$, we have that, for all realization of Ξ_N satisfying (10), $\lim_{N\to\infty} \mu_N = x_\star$. And the result follows since the event formed by all those realizations has probability one. ∎

E. Example on the Necessity of the Conditions on Theorem 6

Suppose that, for each $n \in \mathbb{N}$,

$$y_n = c_n^T x + v_n,$$

where $x \in \mathbb{R}^2$ and $c_n \in \mathbb{R}^2$, $n \in \mathbb{N}$. Also, v_n, $n \in \mathbb{N}$, are independent and identically distributed random variables with $v_n \sim \mathcal{U}(a,b)$ ($\mathcal{U}(a,b)$ denotes the uniform distribution on the interval $[a,b]$), i.e., the probability distribution of v_n is

$$\mathcal{P}(v_n = v) = \begin{cases} (b-a)^{-1} & a \le v \le b, \\ 0 & \text{otherwise.} \end{cases}$$

For each N, we have

$$L_N(x) = \prod_{n=1}^{N} \mathcal{P}(v_n = y_n - c_n^T x).$$

Hence, the function $\Xi_N(x) = \frac{1}{N}\log L_N(x)$, evaluates to $(b-a)^{-1}$, inside a polygon (once N is large enough so that $\text{span}\{c_n : n = 1, \dots N\} = \mathbb{R}^2$) and to zero outside it. Let $\mathcal{S}_N$ denote the support of $\Xi_N(x)$, determined by the interior of the aforementioned polygon. Different realizations of $\Xi_N(x)$ have their supports on different $\mathcal{S}_N$, but it is clear that no scaled version of $\Xi_N(x)$ converges to a Gaussian function. This due to the fact that $\Xi_N(x)$ does not satisfy the conditions of Theorem 6. More precisely, for each realization $\Xi_N(x)$, $N \in \mathbb{N}$, the sequence $\mathcal{S}_N$, $n \in \mathbb{N}$ is decreasing (i.e., $\mathcal{S}_{N+1} \subseteq \mathcal{S}_N$). Also, it is straightforward to see that

$$\bigcap_{N=1}^{\infty} \mathcal{S}_N \overset{\text{w.p.1}}{=} \{x_\star\},$$

where as above, $x_\star$ denotes the true value of x. It follows that it is not possible to choose a set $\mathcal{D}$, including a neighborhood of $x_\star$, and on which $\Xi_N(x)$ is twice continuously differentiable, for all $N \in \mathbb{N}$, and w.p.1. Hence, Conditions (G1) and G3 cannot be simultaneously satisfied.

IV. Case Study I: Target Tracking in Sensor Networks Using Distance Measurements

To illustrate the use of Theorem 6, we consider a network of sensors tracking the position of a moving target. The tracking is done by measuring the distance between the target and each sensor.

A. Problem Description

We have a set of N sensor nodes, located at the positions $x_n = [x_{1,n}, x_{2,n}]^T$, $n = 1, \dots, N$, inside a bounded region $\mathcal{R}$, and a moving target located at $x(t) = [x_1(t), x_2(t)]^T$. The system is described by (1)–(3), with $C = I$ and

$$d_n(x(t)) = r_n(x(t)) + v_n,$$

where

$$r_n(x(t)) = (x_1(t) - x_{1,n})^2 + (x_2(t) - x_{2,n})^2, \tag{32}$$

is the squared distance between Node n and the target, and $v_n \sim \mathcal{N}(0, \sigma^2)$. Also, v_n and v_m are statistically independent whenever $n \neq m$. Again, for notational simplicity, we drop the dependence on t.

We then have the following result, which implies, via Remark 7, that the MLKF becomes asymptotically exact, as the number N of measurements tends to infinity.

Theorem 14: If

$$R(x) \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (r_n(x) - r_n(x_\star))^2 \tag{33}$$

is minimized at the single point $x_\star$, then (23) holds.

Proof: We need to show that the conditions in Theorem 6 are satisfied. Due to the statistical independence of measurements from different sensors, we have

$$L_N(x) = \prod_{n=1}^{N} l_n(x),$$

with

$$l_n(x) = g_{r_n(x),\sigma^2}(y_n). \tag{34}.$$

Let $\Xi_N(x)$ and $\overline{\Xi}(x)$ be defined as in (15) and (22), respectively. Let $x_\star$ denote the true value of x, and $\mathcal{D} \subset \mathbb{R}^d$ be a compact set with $x_\star \in \text{int}(\mathcal{D})$. We split the rest of the argument in five steps:

Step 1: Step 1: We have

$$\log l_n(x) = -\frac{1}{2}\left(\log 2\pi\sigma^2 + \frac{(y_n - r_n(x))^2}{\sigma^2}\right). \tag{35}$$

Then,

$$\begin{aligned} &\mathcal{E}\{\log l_n(x)\} \\ &= -\frac{1}{2}\left(\log 2\pi\sigma^2 + \frac{\mathcal{E}\{(y_n - r_n(x))^2\}}{\sigma^2}\right) \\ &= -\frac{1}{2}\left(\log 2\pi\sigma^2 + \frac{(r_n(x) - r_n(x_\star))^2 + \sigma^2}{\sigma^2}\right). \end{aligned}$$

Hence,

$$\overline{\Xi}(x) = -\frac{1}{2}\left(\log 2\pi\sigma^2 + \frac{1}{\sigma^2}R(x) + 1\right), \tag{36}$$

$$\begin{aligned} \overline{\Xi}(x_\star) - \overline{\Xi}(\mu_N) =& \overline{\Xi}(x_\star) - \Xi_N(x_\star) + \Xi_N(x_\star) \\ &- \Xi_N(\mu_N) + \Xi_N(\mu_N) - \overline{\Xi}(\mu_N) \\ \le& \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon < \eta. \end{aligned}$$

and Condition (G4) follows from (33), since $R(x)$ is twice continuously differentiable.

Step 2: Step 2: From (35), we easily obtain

$$\nabla \log l_n(x) = \frac{2}{\sigma^2}(y_n - r_n(x))(x - x_n)$$

and

$$\nabla^2 \log l_n(x) = -\frac{2}{\sigma^2}[2(x - x_n)(x - x_n)^T - (y_n - r_n(x))I]. \quad (37)$$

Since y_n has a normal distribution, it is straightforward to verify that, for a all $x \in \mathbb{R}^2$, $\log l_n(x)$, $\nabla \log l_n(x)$ and $\nabla^2 \log l_n(x)$ have uniformly bounded second moments. Then, from Rajchman's strong law of large numbers [32, Th. 5.1.2], $\Xi_N(x)$, $\nabla \Xi_N(x)$ and $\nabla^2 \Xi_N(x)$ are SC for every $x \in \mathbb{R}^d$, and (G2) follows.

Step 3: Step 3: From (37), we obtain

$$\nabla[\nabla^2 \log l_n(x)]_{i,j} = = -\frac{4}{\sigma^2}[\nabla((x_i - x_{i,n})(x_j - x_{j,n})) + \delta(i-j)(x - x_n)], \quad (38)$$

with $\delta(i) = 1$ if $i = 0$ and 0 otherwise. It follows from (38) that there exists $M > 0$ such that, $\sup_{n\in\mathbb{N}, x\in\mathcal{D}} \left\|\nabla[\nabla^2 \log l_n(x)]_{i,j}\right\| \overset{\text{w.p.1}}{\leq} M$. Hence, $\sup_{n\in\mathbb{N}, x\in\mathcal{D}} \left\|\nabla[\nabla^2 \Xi_N(x)]_{i,j}\right\| \overset{\text{w.p.1}}{\leq} M$. Then, Condition (G1) follows Lemma 10.

Step 4: Step 4: from (33) and (36), the function $\overline{\Xi}(x)$ is maximized at the single point $x_\star$. Let $\mathcal{C} = \{x \in \mathbb{R}^2 : \|x - p_\mathcal{C}\| \leq r_\mathcal{C}\}$ be a disc with center $p_\mathcal{C}$ and radius $r_\mathcal{C}$, such that $\mathcal{R} \cup \{x_\star\} \subset \text{int}(\mathcal{E})$. Let $M > 1$ and $\mathcal{E} = \{x \in \mathbb{R}^2 : \|x - p_\mathcal{C}\| \leq r_{M\mathcal{C}}\}$ be a disc concentric with and larger than $\mathcal{C}$. From Conditions (G1) and (G2), it follows that $\Xi_N(x)$ is CSUC with limit $\overline{\Xi}(x)$. Hence, for every $\epsilon > 0$, w.p.1., we can choose $N_0(\epsilon)$ such that $|\Xi_N(x) - \overline{\Xi}(x)| < \epsilon$, for all $x \in \mathcal{E}$ and $N > N_0(\epsilon)$. Then, in order to satisfy the conditions of Lemma 13, we need to show that we can choose M and ϵ such that $\sup_{x\in\mathcal{E}^\complement} \Xi_N(x) < \overline{\Xi}(x_\star) - \epsilon$, for all $N > N_0(\epsilon)$.

Fix ϵ. Suppose that there exists $y \in \mathcal{C}^\complement$ such that

$$\Xi_N(y) \geq \overline{\Xi}(x_\star) - \epsilon. \quad (39)$$

Then, there must exist $n \in \{1, \ldots, N\}$ such that

$$\log l_n(y) \geq \overline{\Xi}(x_\star) - \epsilon. \quad (40)$$

Let $\kappa = -\frac{1}{2}\log 2\pi\sigma^2$. For any $L \in \mathbb{N}$, we can then choose M sufficiently large so that $\log l_n(x_\star) \leq \kappa - L(\frac{1}{2} + \epsilon)$. Since $\overline{\Xi}(x_\star) = \kappa - \frac{1}{2}$ and $\max_{x\in\mathbb{R}^2} \log l_n(x) = \kappa$, it follows that there must exist at least L values n_l, $l \in \{1, \ldots, L\}$, such that $\log l_n(x_\star) > \kappa - \frac{1}{2} - \epsilon$, for otherwise $|\Xi_N(x) - \overline{\Xi}(x)| < \epsilon$ cannot be satisfied. Since all nodes are inside $\mathcal{C}$, this implies that there exists $\upsilon < \overline{\Xi}(x_\star) - \epsilon$ such that $\log l_{n_l}(y) < \upsilon$, for all $l \in \{1, \ldots, L\}$. Then, we can choose L such that $\log l_n(y) + \sum_{l=1}^{L} \log l_{n_l}(y) < (L+1)\overline{\Xi}(x_\star) - \epsilon$. Since the same holds for every n satisfying (40), then (39) cannot be satisfied. Hence we can use Lemma 13 and Condition (G3) follows. ■

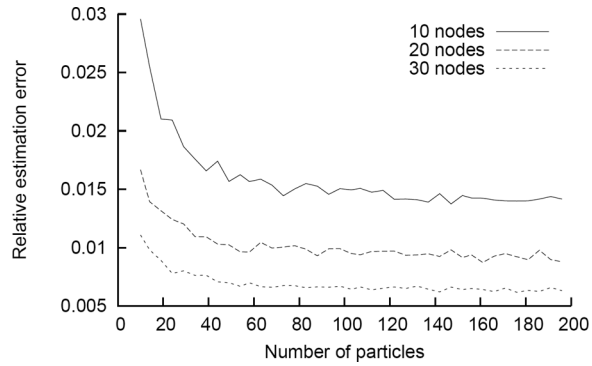

Fig. 1. Tracking error of the particle filter vs. number of particles.

B. Numerical Experiments

In this section we compare the performance of the MLKF, with the one obtained using the Bayesian tracker (4)–(5), when applied to the target tracking system described in Section IV-A. Here, $d = 2$, and we use $R = P = 0.1 \times I$, $A = 0.999 \times I$ and $\sigma^2 = 10$. Each sensor is placed at a uniform distributed randomly chosen location within the square region $[-10, 10] \times [-10, 10]$.

To obtain an approximation of the performance of the Bayesian tracker, we use an importance-sampling particle filter [19]. 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 Fig. 1 the relative estimation error e_{rel} obtained using the particle filter with different number of particles. This error is defined by

$$e_{\text{rel}} = \frac{\sum_{t=1}^{T} \|x(t) - \hat{x}(t)\|^2}{\sum_{t=1}^{T} \|x(t)\|^2}, \quad (41)$$

where $\hat{x}(t)$ denotes the estimated state, and we use $T = 10^5$ samples. We see that the improvement obtained using more than 150 particles is rather small. Hence, we use this value in our experiments.

In Fig. 2 we show the relative estimation errors obtained using the particle filter and the MLKF, as a function of the number of nodes. Again, we use $T = 10^5$ samples. We see that the difference in performance between both filters becomes negligible when the number N of nodes is greater than or equal to 20. This is due to the fact that, for these values of N, the difference between the LF (14) and its Gaussian approximation (24) is very small.

To show how the accuracy of the Gaussian approximation to the LF increases with N, we show in Fig. 3 the quadratic error e_{q} between the LF and its Gaussian approximation, for different valaues of N. This error is defined by

$$e_{\text{q}} = \frac{\int \left|L_N(x) - \tilde{L}_N(\mu_N)\right|^2 dx}{\int |L_N(x)|^2 dx}, \quad (42)$$

with

$$\tilde{L}_N(x) = L_N(\mu_N) \exp\left(-\frac{1}{2}(x - \mu_N)^T \Sigma_N^{-1}(x - \mu_N)\right).$$

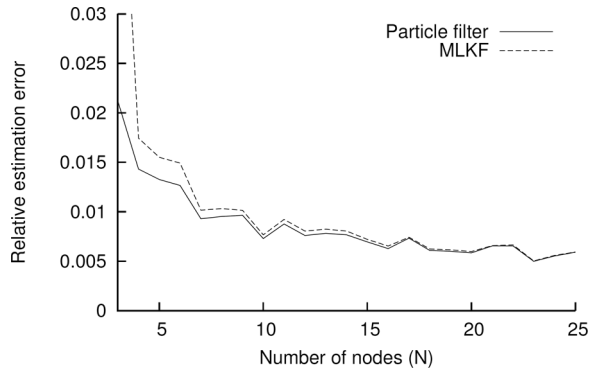

Fig. 2. Tracking error of the particle filter and the MLKF.

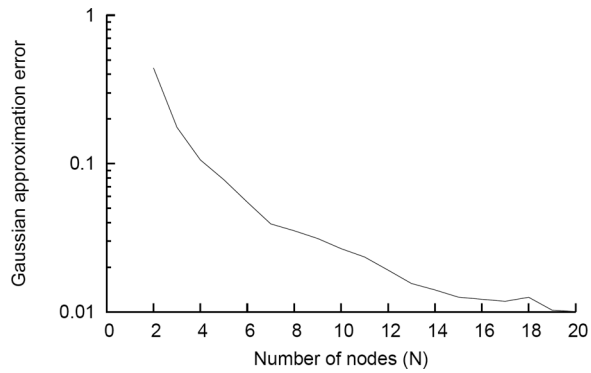

Fig. 3. Quadratic error between the LF and its Gaussian approximation for different number of nodes.

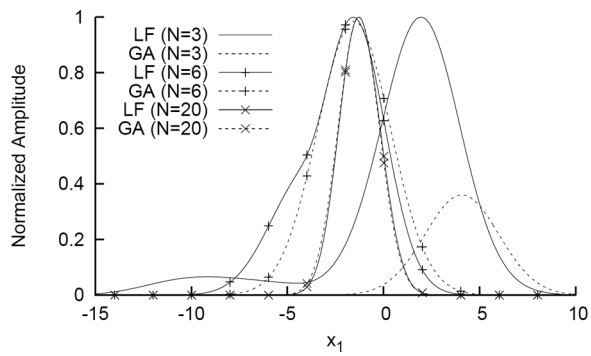

Fig. 4. LF and its Gaussian approximation (GA) for different number of nodes.

In evaluating this error, we use the true value $x_\star = 0$, and for each N, we use 1000 Monte Carlo runs. Fig. 4 shows examples of both functions, for N equal to 3, 6 and 20.

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 MLKF, is shown in Fig. 5. In this comparison we use $N = 20$ nodes. We see how both estimates closely resemble each other.

C. Complexity Analysis

As mentioned in Section I, one of the advantages of the MLKF over particle filters is that it offers major computational savings. To illustrate this point, we do a complexity comparison based on the number of multiplications per time-step. We assume that solving an $n \times n$ linear system of equations, via Gaussian elimination, requires $n^3/3$ multiplications. Also, we use F to denote the number of multiplications required to

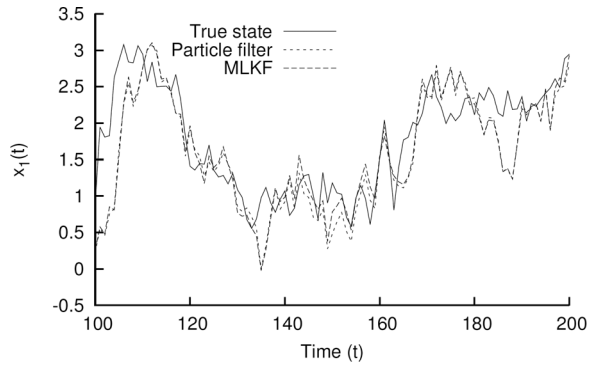

Fig. 5. Evolution of the first component $x_1(t)$ of the state $x(t)$, and its two estimates obtained using particle filtering and MLKF, for $N = 20$ nodes.

evaluate an elementary function (i.e., log, exp, etc.) and R to denote those required for generating a pseudo-random number. We use the setting in Fig. 5, i.e., $N = 20$ nodes and 150 particles. In order to solve the optimization problem required to find the ML estimate (19), we use the NLopt [33] implementation (NLOPT_LD_VAR2) of the shifted limited-memory variable metric method described in [34]. This requires, on average, 3.781 function and gradient evaluations per time-step. Then, the MLKF requires 463.6 multiplications per time-step, which are calculated as follows:

- ML estimate $=$ 306.26 (3.781 iterations on average, 60 mult. per likelihood function evaluation, 21 mult. per gradient evaluation).
- Hessian evaluation $=$ 120.
- Prediction step $=$ 20 (4 mult. for computing the mean and 16 for the covariance).
- Update step $=$ 17.333 (5.333 for two Gaussian elimination steps, 4 for computing the mean and 8 for the covariance).

On the other hand, the particle filter requires $9750 + 150 \times F + 300 \times R$ multiplications, which are calculated as follows:

- Update $= 9000 + 150F$ (150 particles, each particle 60 mult. $+F$ for the likelihood function).
- Prediction $= 600 + 150R$ (150 particles, 4 mult. per particle $+R$ for random number generation).
- Normalization $=$ 150.
- Resampling $= 150R$.

V. Case Study II: Bayesian Tracking in Sensor Networks Using Quantized Measurements

As a second example we consider an array of sensors measuring quantized linear combinations of a common random vector.

A. Problem Description

Consider the system (1)–(3) with $C = I$ and

$$d_n(x) = \mathcal{Q}[c_n^T x + v_n], \tag{43}$$

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

As in Section IV, we provide the following result, stating the asymptotic optimality of the MLKF.

Theorem 15: For the system (1)–(3) with (43), if $\sup_{n\in\mathbb{N}} \|c_n\| < \infty$ and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} c_n c_n^T > 0, \tag{44}$$

then (23) holds.

Proof: As before, we use $x_\star$ to denote the true value of x. Again, we need to show that $L_N(x) = \mathcal{P}(Y_N|x)$ satisfies the conditions of Theorem 6. We have

$$L_N(x) = \prod_{n=1}^{N} l_n(x),$$

with

$$l_n(x) = \int_{\mathcal{Q}^{-1}[y_n]} g_{c_n^T x, \sigma^2}(\Xi) d\Xi. \tag{45}$$

We also define $\Xi_N(x)$ and $\overline{\Xi}(x)$ as in the proof of Theorem 15. We split the rest of the argument in five steps:

Step 1: Step 1: Let $z_n = c_n^T x$. For each $k = 1, \ldots, K$, define

$$\begin{aligned} h_k(z) &= \log \mathcal{P}(y_n = q_k + k | z_n = z) \\ &= \log(G_{0,\sigma^2}(b_k - z) - G_{0,\sigma^2}(b_{k-1} - z)), \end{aligned} \tag{46}$$

which is clearly independent of n. We then have

$$\partial h_k(z) = -\frac{g_{0,\sigma^2}(b_k - z) - g_{0,\sigma^2}(b_{k-1} - z)}{G_{0,\sigma^2}(b_k - z) - G_{0,\sigma^2}(b_{k-1} - z)},$$

and

$$\begin{aligned} \partial^2 h_k(z) = &\frac{\partial g_{0,\sigma^2}(b_k - z) - \partial g_{0,\sigma^2}(b_{k-1} - z)}{G_{0,\sigma^2}(b_k - z) - G_{0,\sigma^2}(b_{k-1} - z)} \\ &- \left(\frac{g_{0,\sigma^2}(b_k - z) - g_{0,\sigma^2}(b_{k-1} - z)}{G_{0,\sigma^2}(b_k - z) - G_{0,\sigma^2}(b_{k-1} - z)} \right)^2. \end{aligned} \tag{47}$$

Let

$$f_n(z) = D_{\mathrm{KL}} \left(\mathcal{P}\left(y_n | z_n = c_n^T x_\star\right) \| \mathcal{P}\left(y_n | z_n = z\right) \right).$$

From (47) we obtain

$$\partial^2 f_n(z) = -\sum_{k=1}^{K} \mathcal{P}(y_n = q_k | z_n = c_n^T x_\star) \partial^2 h_k(z). \tag{48}$$

Now, it is easy to check that $-\partial^2 h_k(z) \geq 0$, for all $z \in \mathbb{R}$. Also, for every $M > 0$, there exists $\varepsilon > 0$ such that, $h(z) > \varepsilon$ whenever $|z| < M$. Let $M = 2\|x_\star\| \sup_{n\in\mathbb{N}} \|c_n\|$. In view of (48), it follows that $\partial^2 f_n(z) \geq \epsilon$, for all $|z| \leq 2a$. We also have that $f_n(z) \geq 0$ and $f_n(c_n^T x_\star) = 0$. Then, since f_n is differentiable, we have $\partial f_n(c_n^T x_\star) = 0$. We thus obtain that, for $|z| \leq M$,

$$f_n(z) \geq \epsilon (z - c_n^T x_\star)^2.$$

Let

$$F(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} D_{\mathrm{KL}}(\mathcal{P}(y_n|x_\star) \| \mathcal{P}(y_n|x)).$$

For all $x \in \mathcal{C}$ where $\mathcal{C} = \{x \in \mathbb{R}^d : \|x\| \leq 2\|x_\star\|\}$, we have $|c_n^T x| \leq M$. Then, for all $x \in \mathcal{C}$,

$$\begin{aligned} F(x) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(c_n^T x) \\ &\geq \epsilon \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (c_n^T (x - x_\star))^2 \\ &= \epsilon (x - x_\star)^T \left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} c_n c_n^T \right) (x - x_\star). \end{aligned}$$

Also, since $\partial^2 f_n(z) \geq 0$, $F(x_\star + \lambda_1 y) \leq F(x_\star + \lambda_2 y)$ for any $y \in \mathbb{R}^d$ and $0 \leq \lambda_1 \leq \lambda_2$ (i.e., it is monotonously increasing along rays departing from $x_\star$). Hence, Condition (G4) follows from (44) and Lemma 12.

Step 2: Step 2: Fix $x \in \mathbb{R}^d$. We have that $l_n(x)$, $n \in \mathbb{N}$, are independent random variables. Hence, so are $\log l_n(x)$, $\nabla \log l_n(x)$ and $\nabla^2 \log l_n(x)$. If $y_n = q_k$, from (46), we have

$$\begin{aligned} \log l_n(x) &= h_k(c_n^T x), \\ \nabla \log l_n(x) &= \partial h_k(c_n^T x) c_n, \\ \nabla^2 \log l_n(x) &= \partial^2 h_k(c_n^T x) c_n c_n^T. \end{aligned} \tag{49}$$

Then, it follows immediately that $\log l_n(x)$, $\nabla \log l_n(x)$ and $\nabla^2 \log l_n(x)$ have uniformly bounded second moments. Then, Condition (G2) follow by the same argument as the one in Theorem 15.

Step 3: Step 3: If $y_n = q_k$, from (49) we obtain

$$\partial_i \nabla^2 \log l_n(x) = [c_n]_i \partial^3 h_k(c_n^T x) c_n c_n^T.$$

Let $\mathcal{D} \subset \mathbb{R}^d$ be a compact set. Since $\sup_{n\in\mathbb{N}} \|c_n\| < \infty$, it is straightforward to verify that there exists $M > 0$ such that

$$\sup_{n\in\mathbb{N}, i\in\{1,\ldots,d\}, x\in\mathcal{D}} \left\| \partial_i \nabla^2 \log l_n(x) \right\| \overset{\text{w.p.1}}{<} M. \tag{50}$$

Then, Condition (G1) follows from Lemma 10, and is valid for any compact connected set $\mathcal{D}$ with $x_\star \in \mathrm{int}(\mathcal{D})$.

Step 4: Step 4: Let $\mathcal{D} = \{x_\star + \alpha y : \|y\| \leq \alpha\}$. Since Ξ_N is CSUC on $\mathcal{D}$, from (G4), there exist $\epsilon > 0$ and $N_0 \in \mathbb{N}$, such that, w.p.1, and for all $N \geq N_0$ and all x in the boundary of $\mathcal{D}$, $\Xi_N(x) \leq \overline{\Xi}(x_\star) - \epsilon$. Also, from (45), the functions $\log l_n(x)$ are concave, hence so is $\Xi_N(x)$. Then $\Xi_N(x)$ satisfies (29) and (G3) follows from Lemma 13.

■

B. Numerical Experiments

As in Section IV-B, we compare the performances of the MLKF and a particle filter approximating the Bayesian tracking recursions (4)–(5), when tracking the state of the quantized system described in Section V-A. We use $d = 3$, $R = P$

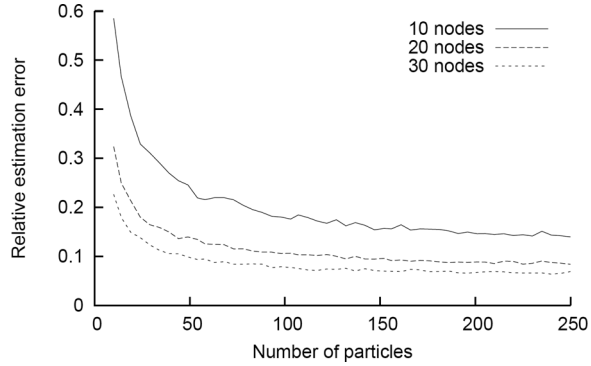

Fig. 6. State estimation error of the particle filter vs. number of particles.

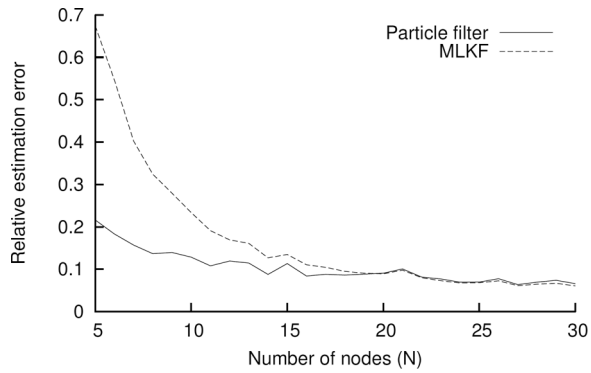

Fig. 7. State estimation error of the particle filter and the MLKF.

$= 0.01 \times I$, $A = 0.999 \times I$ 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 follows

$$c_n = \frac{\tilde{c}_n}{\|\tilde{c}_n\|},$$
$$\tilde{c}_n = [\tilde{c}_{n,1}, \ldots, \tilde{c}_{n,d}]^T,$$

with $\tilde{c}_{n,i}$, $n \in \{1, \ldots, N\}$, $i \in \{1, \ldots, d\}$, being drawn from the distribution $\mathcal{N}(0, 1)$, such that $\tilde{c}_n \neq 0$.

As in Section IV-B, we use an importance-sampling particle filter [19]. Its relative estimation error e_{rel}, evaluated as in (41), as a function of the number of particles, is shown in Fig. 6. In view of this plot, we choose 200 particles for our experiments.

The tracking errors obtained using the particle filter and the MLKF, as a function of the number of nodes, are shown in Fig. 7. We see that the difference in performance becomes negligible when there are more than 20 nodes.

The quadratic difference, measured as in (42), for different values of N, between the LF and its Gaussian approximation, is shown in Fig. 8. Again, we use $x_\star = 0$ as the true value and 1000 Monte Carlo runs. Examples of both functions for N equal to 6, 12 and 18 are given in Fig. 9.

An example of the evolution of the first component $x_1(t)$ of the target, together with its estimates, is shown in Fig. 10.

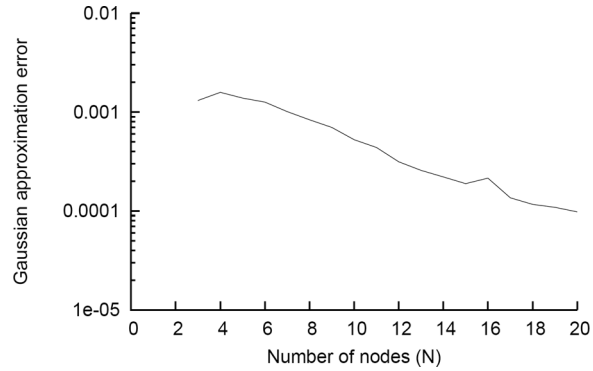

Fig. 8. Quadratic error between the LF and its Gaussian approximation for different number of nodes.

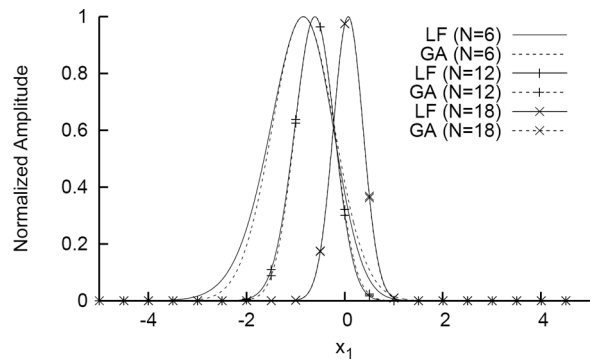

Fig. 9. LF and its Gaussian approximation (GA) for different number of nodes.

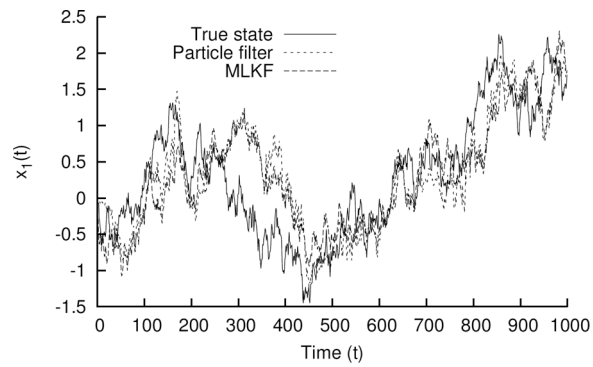

Fig. 10. Evolution of the first component $x_1(t)$ of the state $x(t)$, and its two estimates obtained using particle filtering and MLKF, for $N = 20$ nodes.

C. Complexity Analysis

In this example, and using the algorithm mentioned in Section IV-B, solving the optimization problem (19) requires 5.386 function and gradient evaluations per time-step. Then, repeating the complexity analysis done in Section IV-B, we obtain that the MLKF requires $1563 + 874 \times F$ multiplications per time-step, calculated as follows:

- ML estimate $= 1077 + 754F$ (5.386 iterations on average, $60 + 60F$ mult. per likelihood function evaluation, $140 + 80F$ mult. per gradient evaluation).
- Hessian evaluation $= 360 + 120F$.
- Prediction step $= 63$ (9 mult. for computing the mean and 54 for the covariance).
- Update step $= 63$ (27 for three Gaussian elimination steps, 9 for computing the mean and 27 for the covariance).

The particle filter instead requires $14e3 + 8e3 \times F + 400 \times R$ multiplications, calculated as follows:

- Update $= 12\times10^3+8\times10^3F$ (200 particles, each particle 60 mult. $+40F$ for the likelihood function).
- Prediction $= 1800 + 200R$ (200 particles, 9 mult. per particle $+R$ for random number generation).
- Normalization $= 200$.
- Resampling $= 200R$.

VI. Conclusion

We have provided a set of conditions to guarantee that the MLKF converges to the theoretically optimal Bayesian tracking solution, as the number of sensors becomes large. The implication of this result is that, in tracking applications using a large number of sensors, the computational advantages offered by the MLKF, as well as the guarantee for the stability of the tracker, and the possibility of a distributed implementation, come without noticeable performance detriment. We have applied our result to two case studies, and presented experimental results confirming our theoretical claim. We remark that although the asymptotic optimality conditions require the number of sensors approaching to infinity, in practice only a few sensors are sufficient to ensure that the difference between the MLKF and the Bayesian tracker becomes negligible, as we have demonstrated via simulations in the two case studies.

Appendix

We need the following additional notation.

Notation: The Fourier transform of the function f is denoted by $\hat{f}$. To simplify the notation, we use the following unconventional definition of the Fourier transform pair

$$\hat{f}(\omega) = \frac{1}{2\pi}\int f(x)e^{-jx^T\omega}dx,$$
$$f(x) = \int \hat{f}(\omega)e^{jx^T\omega}d\omega.$$

Notice that the $\frac{1}{2\pi}$ factor usually appears in the inverse Fourier transform formula.

Proof of Lemma 2: We split the proof in steps:

Step 1: Step 1: Let $f \in L_2(\mathbb{R})\cap C(\mathbb{R})$ and $g_N = \psi_N\hat{f}$. Choose $\epsilon > 0$ and $\mathcal{K}_\epsilon \subset \mathbb{R}$ such that $\left\|\hat{f}\chi_{\mathcal{K}_\epsilon^{\complement}}\right\|_2 < \epsilon$, where $\chi_{\mathcal{S}}$ denotes the indicator function of the set $\mathcal{S}$, and $\mathcal{S}^{\complement}$ denotes the complement of the set $\mathcal{S}$. In view of Condition 4, we can choose $N_\epsilon \in \mathbb{N}$ such that $\sup_{N>N_\epsilon}\|(\psi_N-1)\chi_{\mathcal{K}_\epsilon}\|_\infty < \epsilon$. Let $C = \sup_{N\in\mathbb{N}}\|\psi_N\|_\infty$. Then, for any $N \geq N_0$,

$$\begin{aligned}\left\|g_N-\hat{f}\right\|_2^2 &= \left\|(\psi_N-1)\hat{f}\right\|_2^2\\ &= \left\|(\psi_N-1)\hat{f}\chi_{\mathcal{K}_\epsilon}\right\|_2^2 + \left\|(\psi_N-1)\hat{f}\chi_{\mathcal{K}_\epsilon^{\complement}}\right\|_2^2\\ &< \left(\|f\|_2^2+(C+1)^2\right)\epsilon^2.\end{aligned}$$

Hence,

$$\lim_{N\to\infty}\left\|g_N-\hat{f}\right\|_2 = 0.$$

Now, since $f, \psi_N \in L_2(\mathbb{R})$, from Parseval's identity, $\int\hat{\psi}_N f = \int\psi_N\hat{f}$. Also, $g_N \in L_2(\mathbb{R})$, for all $N \in \mathbb{N}$. Then,

$$\begin{aligned}\lim_{N\to\infty}\int\hat{\psi}_N f &= \lim_{N\to\infty}\int\psi_N\hat{f} = \lim_{N\to\infty}\mathcal{F}^{-1}\{g_N\}(0)\\ &= \mathcal{F}^{-1}\left\{\lim_{N\to\infty}g_N\right\}(0) = \mathcal{F}^{-1}\left\{\hat{f}\right\}(0) = f(0). \quad (51)\end{aligned}$$

Step 2: Step 2: We have

$$e^{j\lambda t} = \lambda^2 h_t(\lambda) + jt\lambda + 1, \quad (52)$$

where

$$h_t(\lambda) = f_t(\lambda)+g_t(\lambda),$$
$$f_t(\lambda) = \begin{cases}\frac{1}{\lambda^2}\left(e^{j\lambda t}-1-\frac{j\lambda t}{1+\lambda^2}\right), & \lambda\neq 0,\\ -\frac{t^2}{2}, & \lambda = 0.\end{cases}$$
$$g_t(\lambda) = \frac{-j\lambda t}{1+\lambda^2}.$$

From [25, Lemma 15.47], for every $t \in \mathbb{R}$, $(1+\lambda^2)f_t(\lambda)$ is continuous and bounded, hence $f_t \in L_2(\mathbb{R})\cap C(\mathbb{R})$. Also $g_t \in L_2(\mathbb{R})\cap C(\mathbb{R})$, hence $h_t \in L_2(\mathbb{R})\cap C(\mathbb{R})$.

Step 3: Step 3: From Condition 2, $\partial^2\phi_N^{1/N} \in L_2(\mathbb{R})$. Then, the function $\widehat{\partial^2\phi_N^{1/N}}(\lambda) = -\lambda^2\widehat{\phi_N^{1/N}}(\lambda)$ is in $L_2(\mathbb{R})$, hence so are $\lambda\widehat{\phi_N^{1/N}}$, $\widehat{\partial\phi_N^{1/N}}$, $\widehat{\phi_N^{1/N}}$ and $\phi_N^{1/N}$. Then, using Condition 1, we obtain,

$$\begin{aligned}\phi_N^{1/N}(t) &= \int\widehat{\phi_N^{1/N}}(\lambda)e^{j\lambda t}d\lambda\\ &= \int\lambda^2\widehat{\phi_N^{1/N}}h_t + t\int j\lambda\widehat{\phi_N^{1/N}} + \int\widehat{\phi_N^{1/N}}\\ &= -\int\widehat{\partial^2\phi_N^{1/N}}h_t + t\partial\phi_N^{1/N}(0) + \phi_N^{1/N}(0)\\ &= -\int\widehat{\partial^2\phi_N^{1/N}}h_t + 1.\end{aligned}$$

Then, since $h_t \in L_2(\mathbb{R})\cap C(\mathbb{R})$, from (51),

$$\begin{aligned}\lim_{N\to\infty}N\left(\phi_N^{1/N}(t)-1\right) &= -\lim_{N\to\infty}N\int\widehat{\partial^2\phi_N^{1/N}}h_t\\ &= \lim_{N\to\infty}\int\hat{\psi}_N h_t = h_t(0) = -\frac{t^2}{2}. \quad (53)\end{aligned}$$

Step 4: Step 4: For all $z \in \mathbb{C}$, with $|z| \leq \frac{1}{2}$, we have

$$|\log(1+z)-z| \leq z^2.$$

From (53), $\lim_{N\to\infty}\phi_N^{1/N}(t)-1 = 0$. Then,

$$\begin{aligned}&\limsup_{N\to\infty}\left|\log\phi_N(t) - N(\phi_N^{1/N}(t)-1)\right|\\ &= \limsup_{N\to\infty}N\left|\log\phi_N^{1/N}(t) - \left(\phi_N^{1/N}(t)-1\right)\right|\\ &= \limsup_{N\to\infty}N\left|\log\left[1+\left(\phi_N^{1/N}(t)-1\right)\right] - \left(\phi_N^{1/N}(t)-1\right)\right|\\ &\leq \limsup_{N\to\infty}N\left(\phi_N^{1/N}(t)-1\right)^2\\ &\leq \left(\limsup_{N\to\infty}\sqrt{N}\left(\phi_N^{1/N}(t)-1\right)\right)^2 = 0,\end{aligned}$$

and the result follows from (53). ∎

Proof of Corollary (3): Since $\phi_{y,N}$, $N \in \mathbb{N}$, satisfies the conditions of Lemma 2, we have

$$\lim_{N\to\infty}\Phi_N(ty) = \lim_{N\to\infty}\phi_{y,N}(t) = e^{-\frac{t^2}{2}}.$$

Now, since the above holds for any $\|y\| = 1$, it follows that, for any $x \in \mathbb{R}^d$,

$$\lim_{N\to\infty} \Phi_N(x) = e^{-\frac{\|x\|_2^2}{2}} = e^{-\frac{x^T x}{2}}.$$

■

In order to show Theorem 6, we need the following two lemmas.

Lemma 16: Let $\mathcal{D} \subset \mathbb{R}^d$ and $f_n : \mathcal{D} \to \mathbb{R}$, $n \in \mathbb{N}$, be CSUC to $f : \mathcal{D} \to \mathbb{R}$. If $\lim_{n\to\infty} x_n \overset{w.p.1}{=} x$, then

$$\lim_{n\to\infty} f_n(x_n) \overset{w.p.1}{=} f(x).$$

Proof: The result follows since the following holds in an event with probability one. From [30, Th. 7.12], f is continuous. Then,

$$\begin{aligned}
&\limsup_{n\to\infty} |f_n(x_n) - f(x)| \\
&\le \limsup_{n\to\infty} |f_n(x_n) - f(x_n) + f(x_n) - f(x)| \\
&\le \limsup_{n\to\infty} |f_n(x_n) - f(x_n)| + \limsup_{n\to\infty} |f(x_n) - f(x)| \\
&= 0.
\end{aligned}$$

■

Lemma 17: Let $f_n : \mathbb{R}^d \to \mathbb{R}$. Then

$$\nabla^2 f_n^{1/n} = f_n^{1/n}\left(\frac{1}{n}\nabla^2 \log f_n + \frac{1}{n^2}\nabla \log f_n \nabla^T \log f_n\right).$$

Proof: It follows immediately from

$$\nabla^2 f_n^{1/n} = \frac{f_n^{1/n}}{n}\left(\left(\frac{1}{n} - 1\right)\frac{\nabla f_n}{f_n}\frac{\nabla^T f_n}{f_n} + \frac{\nabla^2 f_n}{f_n}\right).$$

■

We can now state the proof of Theorem 6.

Proof of Theorem 6: We split the proof in steps:

Step 1: Step 1: Condition (G1) states that

$$\nabla^2 \Xi_N \text{ is CSUC on } \mathcal{D}. \tag{54}$$

Then, (G2) and two applications of Lemma 8-1) give

$$\nabla \Xi_N \text{ is CSUC on } \mathcal{D}, \tag{55}$$

$$\Xi_N \text{ is CSUC on } \mathcal{D}. \tag{56}$$

Step 2: Step 2: From (G3), (54), Lemma 16 and Lemma 8-(2),

$$\begin{aligned}
\lim_{N\to\infty} (N\Sigma_N)^{-1} &= -\lim_{N\to\infty} \nabla^2 \Xi_N(\mu_N) \\
&\overset{\text{w.p.1}}{=} -\nabla^2 \overline{\Xi}(x_\star).
\end{aligned} \tag{57}$$

Then, from (G4),

$$\lim_{N\to\infty} \|N\Sigma_N\| \overset{\text{w.p.1}}{<} \infty, \tag{58}$$

and therefore, for any $x \in \mathbb{R}^d$,

$$\lim_{N\to\infty} \Sigma_N^{1/2} x \overset{\text{w.p.1}}{=} 0. \tag{59}$$

Step 3: Step 3: From (G3), (59) and (56) satisfy the conditions of Lemma 16. Then,

$$\lim_{N\to\infty} \Xi_N(\mu_N) \overset{\text{w.p.1}}{=} \overline{\Xi}(x_\star), \tag{60}$$

$$\lim_{N\to\infty} \Xi_N\left(\Sigma_N^{1/2} x + \mu_N\right) \overset{\text{w.p.1}}{=} \overline{\Xi}(x_\star). \tag{61}$$

Replacing (56) by (55) in the argument above we obtain

$$\lim_{N\to\infty} \nabla \Xi_N\left(\Sigma_N^{1/2} x + \mu_N\right) \overset{\text{w.p.1}}{=} \nabla \overline{\Xi}(x_\star) = 0, \tag{62}$$

and using (54) instead

$$\lim_{N\to\infty} \nabla^2 \Xi_N\left(\Sigma_N^{1/2} x + \mu_N\right) \overset{\text{w.p.1}}{=} \nabla^2 \overline{\Xi}(x_\star). \tag{63}$$

Step 4: Step 4: A consequence of the continuous extension theorem [35, Th. 5.1] is that we can choose a map $T_N : \mathbb{R}^d \setminus \mathcal{D} \to \mathbb{R}^{d\times d}$ such that the following map

$$H_N(x) = \begin{cases} \nabla^2 L_N^{1/N}(x), & x \in \mathcal{D}, \\ T_N(x), & x \notin \mathcal{D}, \end{cases}$$

is continuous and

$$\begin{aligned}
\sup_{x\in\mathbb{R}^d} \|H_N(x)\| &= \sup_{x\in\mathcal{D}} \left\|\nabla^2 L_N^{1/N}(x)\right\|, \\
H_N(x) &\in L_2(\mathbb{R}^d).
\end{aligned} \tag{64}$$

Let $\tilde{L}_N : \mathbb{R}^d \to \mathbb{R}^d$ be the unique map satisfying

$$\begin{aligned}
\nabla^2 \tilde{L}_N^{1/N}(x) &= H_N(x), \\
\tilde{L}_N(x) &= L_N(x), \text{ for all } x \in \mathcal{D},
\end{aligned} \tag{65}$$

and

$$\begin{aligned}
\tilde{\Xi}_N(x) &= \frac{1}{N}\log \tilde{L}_N(x), \\
\overline{\tilde{\Xi}}(x) &= \lim_{N\to\infty} \mathcal{E}\left\{\tilde{\Xi}_N(x)\right\}.
\end{aligned}$$

Let also

$$\Phi_N(x) = \frac{1}{\tilde{L}_N(\mu_N)} \tilde{L}_N\left(\Sigma_N^{1/2} x + \mu_N\right),$$

and $\Psi_N = -N\nabla^2 \Phi_N^{1/N}$. From Lemma (17), we have

$$\Psi_N = -\Phi_N^{1/N}\left(\nabla^2 \log \Phi_N + \frac{1}{N}\nabla \log \Phi_N \nabla^T \log \Phi_N\right), \tag{66}$$

with

$$\Phi_N^{1/N}(x) = \exp\left(\tilde{\Xi}_N\left(\Sigma_N^{1/2} x + \mu_N\right) - \tilde{\Xi}_N(\mu_N)\right), \tag{67}$$

and

$$\frac{1}{\sqrt{N}}\nabla \log \Phi_N(x) = (N\Sigma_N)^{1/2} \nabla \tilde{\Xi}_N(\Sigma_N^{1/2} x + \mu_N), \tag{68}$$

and

$$\begin{aligned}
\nabla^2 \log \Phi_N(x) = (N\Sigma_N)^{1/2} \times \\
\times \nabla^2 \tilde{\Xi}_N\left(\Sigma_N^{1/2} x + \mu_N\right)(N\Sigma_N)^{1/2}.
\end{aligned} \tag{69}$$

Step 5: Step 5: From (60), (61) and (67), since $\tilde{\Xi}_N = \Xi_N$, on $\mathcal{D} \ni x_\star$,

$$\lim_{N\to\infty} \log \Phi_N^{1/N}(x) \overset{\text{w.p.1}}{=} 0.$$

Hence,

$$\lim_{N\to\infty} \Phi_N^{1/N}(x) \overset{\text{w.p.1}}{=} 1. \tag{70}$$

Also, from (62) and (68)

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \nabla \log \Phi_N \stackrel{\text{w.p.1}}{=} 0, \tag{71}$$

and from (63) and (69)

$$\begin{aligned}&\lim_{N\to\infty} \nabla^2 \log \Phi_N(x)\\ &\stackrel{\text{w.p.1}}{=} -\nabla^2\overline{\Xi}(x_\star)^{1/2}\nabla^2\overline{\Xi}(x_\star)\nabla^2\overline{\Xi}(x_\star)^{1/2}\\ &= -I.\end{aligned} \tag{72}$$

Hence, from (66), (70)–(72),

$$\lim_{N\to\infty} \Psi_N(x) \stackrel{\text{w.p.1}}{=} I. \tag{73}$$

Step 6: Step 6: From Lemma (17) we have

$$\nabla^2 L_N^{1/N} = \exp(\Xi_N)(\nabla^2\Xi_N + \nabla\Xi_N\nabla^T\Xi_N).$$

Then, from (54)–(56), it follows that, w.p.1,

$$\sup_{x\in\mathbb{R}^d} \|H_N(x)\| < \infty. \tag{74}$$

We then have,

$$\Psi_N(x) = -\frac{1}{\exp\left(\tilde{\Xi}_N(\mu_N)\right)}(N\Sigma_N)^{1/2}\times H_N\left(\Sigma_N^{1/2}x+\mu_N\right)(N\Sigma_N)^{1/2}.$$

From (60), (57), (64) and (74) w.p.1, there exists $N_0 \in \mathbb{N}$ such that, for all $N \geq N_0$,

$$\Psi_N \in L_2(\mathbb{R}^d), \tag{75}$$

$$\sup_{x\in\mathbb{R}} \|\Psi_N(x)\| < \infty. \tag{76}$$

Step 7: Step 7: Let $\mathcal{K} \subset \mathbb{R}^d$ be compact and $\mathcal{D}_N = \left\{\Sigma_N^{1/2}x+\mu_N : x\in\mathcal{K}\right\}$. From (G3) and (58), w.p.1, there exists $N_0 \in \mathbb{N}$ such that, $\mathcal{D}_N \subset \mathcal{D}$, for all $N > N_0$. Then, from (54)–(56) and (66),

$$\Psi_N \text{ is CSUC on } \mathcal{K}. \tag{77}$$

Step 8: Step 8: Our next step is to show that $\phi_{y,N}$, $N \in \mathbb{N}$, satisfies the conditions of Lemma 2. We have that $\phi_{y,N} \in C_2(\mathbb{R})$, and in view of (65),

$$\begin{aligned}\phi_{y,N}(0) &= \Phi_N(0) = 1,\\ \partial\phi_{y,N}(0) &= y^T\nabla\Phi_N(0) = 0.\end{aligned}$$

Then, $\phi_{y,N}$ satisfies Condition 1 of Lemma 2. So we need to show that $\psi_{y,N} = -N\partial^2\phi_{y,N}^{1/N}$ satisfies Conditions 2–4. We have

$$\begin{aligned}\psi_{y,N}(t) &= -N\frac{\partial^2}{\partial t}\Phi_N^{1/N}(ty)\\ &= -y^T N\nabla^2\Phi_N^{1/N}(ty)y = y^T\Psi_N(ty)y.\end{aligned}$$

Hence, from (75), $\psi_{y,N} \in L_2(\mathbb{R})$, for all $N \geq N_0$, and Condition 2 of Lemma 2 follows by a simple substitution making N_0 become the initial term in the sequence. Also, Condition 3 follows from (76) and Condition 4 from (73) and (77).

Step 9: Step 9: Since $\phi_{y,N}$, $N \in \mathbb{N}$, satisfies the conditions of Lemma 2, from Corollary 3 it follows that, for any $x \in \mathbb{R}^d$,

$$\lim_{N\to\infty} \Phi_N(x) = e^{-\frac{x^T x}{2}},$$

from where the result follows by substituting (65) in the equation above.

■

REFERENCES

[1] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks: Theory and Practice*. New York, NY, USA: Wiley, 2010.

[2] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 622–633, May 2008.

[3] U. Khan and J. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4919–4935, Oct. 2008.

[4] A. Ribeiro, I. Schizas, S. Roumeliotis, and G. Giannakis, "Kalman filtering in wireless sensor networks," *IEEE Control Syst. Mag.*, vol. 30, no. 2, pp. 66–86, Apr. 2010.

[5] O. Hlinka, O. Sluciak, F. Hlawatsch, P. Djuric, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, Aug. 2012.

[6] S. Billings, "Identification of nonlinear systems: a survey," in *Proc. IEE P-Contr. Theory AP. IET*, 1980, vol. 127, pp. 272–285.

[7] N. Bershad, S. Bouchired, and F. Castanie, "Stochastic analysis of adaptive gradient identification of Wiener-hammerstein systems for Gaussian inputs," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 557–560, Feb. 2000.

[8] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2617–2624, Aug. 2005.

[9] K. You, L. Xie, S. Sun, and W. Xiao, "Multiple-level quantized innovation Kalman filter," in *Proc. IFAC World Cong.*, 2008, vol. 17, pp. 1420–1425.

[10] M. Fu and C. E. de Souza, "State estimation for linear discrete-time systems using quantized measurements," *Automatica*, vol. 45, no. 12, pp. 2937–2945, 2009.

[11] X. Wang, M. Fu, and H. Zhang, "Target tracking in wireless sensor networks based on the combination of kf and mle using distance measurements," *IEEE Trans. Mobile Comput.*, vol. 11, no. 4, pp. 567–576, Apr. 2012.

[12] X. Wang, H. Zhang, and M. Fu, "Collaborative target tracking in wsns using the combination of maximum likelihood estimation and Kalman filtering," *J. Contr. Theory Appl.*, vol. 11, no. 1, pp. 27–34, 2013.

[13] U. A. Khan, S. Kar, and J. M. Moura, "Distributed sensor localization in random environments using minimal number of anchor nodes," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 2000–2016, May 2009.

[14] Y. Diao, M. Fu, and H. Zhang, "New results on Node localizable conditions for sensor networks," in *Proc. IEEE CDC*, 2012, pp. 2726–2731.

[15] S. A. Boyer, *SCADA: Supervisory Control and Data Acquisition*. Research Triangle Park, NC, USA: Instrumentation, Systems, and Automation Society, 2009.

[16] X. Tai, D. Marelli, E. Rohr, and M. Fu, "Optimal pmu placement for power system state estimation with random component outages," *Int J. Electron. Power*, vol. 51, pp. 35–42, 2013.

[17] L. Chang, B. Hu, A. Li, and F. Qin, "Transformed unscented Kalman filter," *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 252–257, Jan. 2013.

[18] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[19] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[20] J. Kotecha and P. Djuric, "Gaussian sum particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.

[21] M. Simandl, J. Královec, and T. Söderström, "Advanced point-mass method for nonlinear state estimation," *Automatica*, vol. 42, no. 7, pp. 1133–1145, 2006.

[22] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.
[23] D. Marelli and M. Fu, "Asymptotic optimality of the maximum-likelihood filter for Bayesian tracking in sensor networks," presented at the IFAC World Cong., 2014.
[24] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses*, Second ed. Berlin, Germany: Springer-Verlag, 2003.
[25] A. Klenke, *Probability Theory: A Comprehensive Course (Universitext)*, 2008 ed. Berlin, Germany: Springer-Verlag, 2010, vol. 6.
[26] J. Davidson, *Stochastic Limit Theory: An Introduction for Econometricians*. London, U.K.: Oxford Univ. Press, 1994.
[27] C. Gourieroux and A. Monfort, *Statistics and Econometric Models, Volume 2*. Cambridge, U.K.: Cambridge Univ. Press, 1996.
[28] A. R. Gallant and H. White, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, 1st ed. New York, NY, USA: Blackwell, 1988.
[29] W. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook Economet.*, vol. 4, pp. 2111–2245, 1994.
[30] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1976.
[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*, 2nd ed. New York, NY, USA: Wiley-Interscience, 2006.
[32] K. L. Chung, *A Course in Probability Theory, Third Edition*, 2nd ed. Waltham, MA, USA: Academic Press, 2000, vol. 10.
[33] S. G. Johnson, "The NLopt nonlinear-optimization package," [Online]. Available: http://ab-initio.mit.edu/nlopt
[34] J. Vlcek and L. Luksan, "Shifted limited-memory variable metric methods for large-scale unconstrained optimization," *J. Comput. Appl. Math.*, vol. 186, pp. 365–390, 2006.
[35] J. Dugundji, *Topology (Allyn and Bacon Series in Advanced Mathematics)*, 1st ed. Boston, MA, USA: Allyn and Bacon, 1966.

Damián Marelli received his Bachelors Degree in Electronics Engineering from the Universidad Nacional de Rosario, Argentina in 1995, a Ph.D. degree in Electrical Engineering and a Bachelor (Honours) degree in Mathematics from the University of Newcastle, Australia in 2003.

From 2004 to 2005 he held a postdoctoral position at the Laboratoire d'Analyse Topologie et Probabilités, CNRS/Université de Provence, France. Since 2005 he is Research Academic at the School of Electrical Engineering and Computer Science at the University of Newcastle, Australia. In 2007 he received a Marie Curie Postdoctoral Fellowship, hosted at the Faculty of Mathematics, University of Vienna, Austria, and in 2011 he received a Lise Meitner Senior Fellowship, hosted at the Acoustics Research Institute of the Austrian Academy of Sciences. His main research interests include time-frequency analysis, system identification, statistical signal processing and sensor networks.

Minyue Fu (S'84–M'87–SM'94–F'04) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 1982, and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, in 1983 and 1987, respectively.

From 1983 to 1987, he held a Teaching Assistantship and a Research Assistantship at the University of Wisconsin-Madison. He worked as a Computer Engineering Consultant at Nicolet Instruments, Inc., Madison, during 1987. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI. He joined the Department of Electrical and Computer Engineering, University of Newcastle, Australia, in 1989, where he is currently Chair Professor in Electrical Engineering and Head of School of Electrical Engineering and Computer Science. In addition, he was a Visiting Associate Professor at the University of Iowa during 1995/1996, and a Senior Fellow/Visiting Professor at Nanyang Technological University, Singapore, 2002. He holds a Qian-ren Professorship at Zhejiang University, China. He was an Associate Editor for Automatica and the Journal of Optimization and Engineering. His main research interests include control systems and signal processing.

Dr. Fu was an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Brett Ninness (S'88–M'89–SM'14) was born in 1963 in Singleton, Australia and received his BE, ME, and Ph.D. degrees in Electrical Engineering from the University of Newcastle, Australia in 1986, 1991 and 1994 respectively.

His research interests are in the areas of system identification and stochastic signal processing, in which he has authored approximately one hundred papers in journals and conference proceedings. He has served on the editorial boards of Automatica, IEEE TRANSACTIONS ON AUTOMATIC CONTROL and has served as Editor in Chief for IET Control Theory and Applications.

He is currently Pro-Vice Chancellor of the Faculty of Engineering and Built Environment at the University of Newcastle, Australia.