Parameter Estimation in Sensor Networks under Probabilistic Censoring

Ido Nevat1, Pengfei Zhang2, Gad Frenkel3 and Gareth W. Peters4
1. TUMCREATE, Singapore
2. Department of Engineering Science, University of Oxford, UK
3. School of Engineering, Ruppin Academic Center, Emek-Hefer 40250, Israel
4. Department of Statistical Sciences, University College London (UCL), UK

Abstract—We consider the problem of estimating an unknown parameter which is observed indirectly by a sensor network. Each of the sensors exhibits probabilistic non-linear behavior causing censoring (clipping) of the signals at random. The signals then get transmitted over fading wireless channels to the Fusion Centre (FC). We develop three algorithms for estimating the unknown parameter: the first is based on Pearson’s Method of Moments which involves the calculation of the statistical moments of the model which can be derive exactly; the second is based on the Maximum Likelihood Estimator (MLE) which results in an intractable likelihood function. To evaluate the likelihood function, we develop a novel approximation via a non-parametric probability density estimator that is based on series expansions of the Gram-Charlier family of basis functions. The third is the Marginalized Least Squares (M-LS) in which we solve the Least Squares problem by marginalizing out the random unknown nuisance parameters of the model. All three algorithms enjoy a low computational complexity, and only involve solving a one-dimensional optimization problem which is simple to implement in practice. We compare the performance of the proposed algorithms under various system configurations and show various trade-offs between the algorithms as a function of system parameters (e.g. number of sensors, frame length, signal-to-noise-ratio, sensing quality etc.).

Keywords: Wireless sensor networks, Maximum likelihood estimation, Method of Moments, Least Squares, Askey polynomials series expansion.

I. INTRODUCTION

Wireless Sensor Networks (WSN) have attracted considerable attention due to the large number of applications, such as environmental monitoring, weather forecasts [1], [2], surveillance, health care, and home automation [2], [3]. WSN consists of a set of spatially distributed sensors which monitor a spatial physical phenomenon containing some desired attribute (e.g. pressure, temperature, concentrations of substance, sound intensity, radiation levels, pollution concentrations etc.), and regularly communicate their observations to a Fusion Center (FC) [4]. A fundamental problem in WSN is the estimation of an unknown parameter and has many practical applications [5]–[7]. A common assumption in the literature on sensor networks regarding the operation of the sensors is that they employ some linear operation on the received signal (e.g. power amplifiers, local processing, cooperative processing), and provide an output signal which is a linear function of the input with constant gain, known as Amplify-and-Forward (AF) transmission [8]–[10]. However, in practice the sensors may exhibit nonlinear behavior for a wide range of different reasons, which include internal failure of sensors hardware [11]–[14]; natural degradation over time of operation [15]; power saving mechanisms which limit the amplification of signals [14]; Byzantine attack on the system which causes changes to the operation mode of the sensors [16]; and heterogeneous sensor networks where low quality sensors have a nonlinear behavior and high quality sensors have a linear behavior [16], [17].

Under all of these models, the characteristics of the amplifiers change across operational mode, manufacturing process variation, temperature, operating voltage and aging, depending on the particular application. Since the exact operational state of the sensors is unknown to the FC and may change over time, we model its operational state (i.e. linear and non-linear states) as a binary random variable which indicates whether the sensor is in the linear amplification state or in a non-linear state, due to one or more of the aforementioned reasons.

In many practical cases, the physical phenomenon to be detected by WSN is not directly observed, but instead is observed indirectly after being attenuated by the effects of unknown and random physical medium (we denote this generically as propagation medium) [18], [19]. In addition, the sensors transmit their noisy and distorted observations over a set of independent wireless channels to the FC [18]–[20]. These three aspects of random propagation medium, non-linear amplifiers and random wireless channels make the estimation problem very challenging.

The most common approach to solving the estimation of unknown parameter in stochastic models is the Maximum Likelihood Estimator (MLE), developed by Fisher [21], as it enjoys many desirable theoretical properties, such as consistency, efficiency and unbiasedness under certain conditions, see [22]. The MLE is a special case of the well known family of estimators, known as M-estimators [23] and can be easily derived as long as the likelihood function can be evaluated analytically. However, as we will present in Section IV, the likelihood function in our model contains intractable integrals. As such, we are not able to directly derive the MLE. To overcome this difficulty we approximate the likelihood function via a low-complexity algorithm which is based on a basis expansion density estimation using Askey-orthogonal polynomial expansion. The series expansion we utilise is based on a kernel density multiplied by polynomials, known as Askey polynomials [24]. In this paper we use a Gaussian
density basis, although other basis could be used (eg. Gamma and Beta density basis). This series expansion will allow us to obtain an analytic form of the likelihood function and the series expansion can be truncated according to the required accuracy. It’s important to remember that our goal is to locate the maximum of the likelihood function rather than reconstruct the shape of the function over the entire domain, and therefore this will not require as much precision on the entire state-space, which in turn requires fewer coefficients, thus keeping the computational complexity low. Since the MLE we derive is an approximation, we also derive two other competing algorithms which can be calculated exactly. The first is based on the Method of Moments of Pearson [25], in which one matches the statistical moments with the empirical ones, and then solves a set of (possibly non-linear) equations. Finally, we also develop an algorithm which is based on the well known Least Squares (LS) approach of Gauss [26] and Legendre [27]. We show that the classic LS cannot be applied directly, because our model contains nuisance parameters. To overcome this difficulty, we extend the classical LS solution and develop the Marginalized Least Squares (MLS) in which we integrate out the effects of the nuisance parameters.

A. Related works

The estimation of unknown parameter in sensor networks is a fundamental problem and has been widely studied. Most of these works usually make the simplifying assumptions of either a linear observation model or utilizing a linear estimator (or combination of both). In [6], [7] the authors developed linear estimation algorithms of an unknown signal by a sensor network with a fusion center over non-orthogonal channels via coherent combination of sensor messages at the fusion center. In [10] the authors designed the optimal sensor collaboration strategy for the estimation of parameters in a linear model. In [28] the authors studied the properties of linear and nonlinear distributed parameter estimation problems in general settings. In [29] the authors developed algorithms for data reduction in linear models and quantized observations. In [15] the authors developed an estimation algorithm in the presence of non-linear relay functions which relay of matching the theoretical and empirical moments.

In [30] the authors introduced a non-linear transmission function in order to control the power consumption of the sensors. They then developed a distributed average consensus algorithm in the presence of impulsive noise, and showed that the use of bounded transmission makes their algorithm robust to a wide range of channel noise distributions. In [31] the authors considered a binary hypotheses testing scheme which uses bounded transmission functions over Gaussian multiple access channels. In [32] the authors consider distributed average consensus when the topology is random and the communication in the channels is corrupted by additive noise.

The main difference between these works and ours lies in the fact that in this work the operation of the sensors is not deterministic, but probabilistic, in the sense that their operation may or may not be linear, and is not known at the FC. The combination of probabilistic sensor operation with the non-linear model results in a stochastic mixture model which induces new challenges on the estimation of the unknown parameter, and requires the development of novel statistical techniques. Another aspect which has not been addressed before is that the likelihood function in our model involves intractable integrals and as such, no likelihood based inference can be directly applied. We therefore develop a novel approximation of the likelihood function via non-parametric probability density estimator that is based on series expansions of a specific type of Askey bases representation, known as the the Gram-Charlier family of basis functions [33], [34].

B. Contributions:

1) In Section II we develop a statistical model to account for both linear and non-linear operational modes of the sensors. In addition, we incorporate the practical scenario of sensing over random propagation medium as well as transmission to the FC over wireless channels which are all considered random and unknown.

2) In Section III we develop an estimation algorithm which is based on Method of Moments. We show that the moments can be calculated exactly and therefore the estimator is exact.

3) In Section IV we develop the approximated Maximum Likelihood Estimator. We first show that the likelihood function involves intractable integrals. To overcome this difficulty, we approximate it via a low-complexity algorithm which is based on a basis expansion density estimation using Askey-orthogonal polynomial expansion

4) In Section V we develop the Marginalized Least Squares estimator. This estimator is based on the Least Squares (LS) method, where we marginalize over all the nuisance random variables in the model.

We denote throughout the paper the Gaussian density function by \( N(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi} \sigma} \exp \left( -\frac{1}{2\sigma^2} (x - \mu)^2 \right) \), and the standard Gaussian density and distribution functions by \( \phi(x) := N(x; 0, 1) \), and \( \Phi(x) := \int_{-\infty}^{x} N(x; 0, 1) \, dx \).

II. WIRELESS SENSOR NETWORK SYSTEM MODEL

We consider a wireless sensor network consisting of \( M \) sensors which are spatially deployed. The sensors observe
indirectly a deterministic variable \( \theta \) over a random propagation medium. Each sensor amplifies its signal and transmits it to the FC over wireless channels. The FC then estimates the value of \( \theta \) based on the aggregate inputs. We now present the sensor network model (see Fig. 1):

A1 The unknown parameter to be estimated \( \theta \in \Theta \) (e.g., temperature, pressure, concentration etc.) is constant throughout a frame of \( L \) samples.

A2 For each frame of \( L \) samples, the observed signal at the \( n^{th} \) sensor, \( R^{(n)}_m \), is given by:

\[
R^{(n)}_m = H^{(n)}_m \theta + V^{(n)}_m,
\]

where \( H^{(n)}_m \sim N (\mathbf{T}, \sigma^2_H) \) denotes the random propagation medium between the source and the \( n^{th} \) sensor, and \( V^{(n)}_m \sim N (0, \sigma^2_V) \) is the independent and identically distributed (i.i.d.) Gaussian noise at the \( n^{th} \) sensor.

A3 Probabilistic sensor processing: each sensor measures the physical value \( R^{(n)}_m \) and converts it into an electrical signal. The physical mechanism of the transformation can exhibit linear (denoted \( F_l \)) or non-linear (denoted \( F_{nl} \)) behavior as well as additive noise. This electrical signal is amplified and may or may not be linear depending on the quality of the sensor, its battery state etc. Finally the global conversion can be seen as a random and unknown event. We model this behaviour probabilistically according to the following mixture model:

\[
F_k (x) = \begin{cases} 
F_l (x), & \text{w.p. } \pi_k, \\
F_{nl} (x), & \text{w.p. } (1 - \pi_k), 
\end{cases}
\]

where \( F_l (x) \) and \( F_{nl} (x) \) are the linear and non-linear relay amplification functions, respectively, and \( \pi_k \) is the a-priori probability of the \( n^{th} \) sensor’s amplifier being linear. The linear AF function is given by \( F_l (x) = \alpha x \), where \( \alpha \) is a known amplification factor. The non-linear amplifier function is defined as the widely used Ideal Soft-Limiter Amplifier (ISLA) model [35], which introduces clipping of the signal as follows:

\[
F_{nl} (x) = \begin{cases} 
\alpha P_{max}, & x \geq P_{max}, \\
\alpha x, & P_{min} \leq x \leq P_{max}, \\
\alpha P_{min}, & x \leq P_{min}, 
\end{cases}
\]

where \( P_{max} \) and \( P_{min} \) are known.

A4 The received signal at the FC from the \( n^{th} \) sensor over wireless channels at epoch \( l \), denoted \( y^{(l)}_m \), is given by the following model:

\[
y^{(l)}_m = G^{(l)}_m F_k \left( R^{(l)}_m \right) + W^{(l)}, \quad m = \{1, \ldots, M\},
\]

where the random variable \( G^{(l)}_m \sim N (\mathbf{G}, \sigma^2_G) \) is the wireless channel between the \( m^{th} \) sensor and the FC, and \( W \sim N (0, \sigma^2_W) \) is the additive noise at the FC.

Our goal is to develop estimators for \( \theta \) based on the available observations \( y^{(1:L)}_1 \). To this end we derive in the following sections three different estimators for \( \theta \): Method of Moments estimator due to Pearson [25], the Maximum Likelihood Estimator (MLE) due to Fisher [21], and the Least Squares estimator due to Gauss [26] and Legendre [27].

### III. Estimator I: Method of Moments Estimation

In this Section we develop the method of moments estimator for \( \theta \). This method dates back to Karl Pearson’s solution for identifying the parameters of a mixture of two univariate Gaussians [25]. The method of moments is semi-parametric since the parameter of interest \( \theta \) is finite-dimensional, but the full shape of the distribution function of the observations \( p \left( y^{(1:L)}_1 | \theta \right) \) is not known. This method yields consistent estimators under very weak assumptions [36], meaning that as the number of observations increases indefinitely (i.e. \( M, L \to \infty \)), the resulting sequence of estimates converges in probability to \( \theta \). The Method of Moments works by equating the population moments with the empirical moments. In our case, since \( \theta \) is considered as a scalar, we only need to equate the first moment, as follows:

\[
E \left[ y^{(1:L)}_1 | \theta \right] = \bar{y}^{(1:L)}_1 \left( \theta \right),
\]

where \( \bar{y}^{(1:L)}_1 (\theta) \) denotes the empirical first moment given by

\[
\bar{y}^{(1:L)}_1 (\theta) = \frac{1}{M L} \sum_{m=1}^{M} \sum_{l=1}^{L} y^{(l)}_m.
\]

Then, the Method of Moments estimator, denoted \( \hat{\theta}_{MM} \) is given by the solution to the following non-linear equation:

\[
E \left[ y^{(1:L)}_1 | \theta \right] = \bar{y}^{(1:L)}_1 \left( \theta_{MM} \right).
\]

To derive the estimator we need to calculate the statistical moments of the model. The first moment of the population is given by:

\[
E \left[ y^{(1:L)}_1 | \theta \right] = \frac{1}{M} \frac{1}{L} \sum_{m=1}^{M} \sum_{l=1}^{L} E \left[ y^{(l)}_m | \theta \right] = \frac{1}{M} \frac{1}{L} \sum_{m=1}^{M} \sum_{l=1}^{L} \left( \pi_l \int_{-\infty}^{\infty} y^{(l)}_m P^{F_l} \left( y^{(l)}_m | \theta \right) dy^{(l)}_m \right),
\]

where we define the likelihood function under linear transformation \( p^{F_l} \left( y^{(l)}_m | \theta \right) = p \left( y^{(l)}_m | \theta, F_k = F_l \right) \), and under non-linear transformation \( p^{F_{nl}} \left( y^{(l)}_m | \theta \right) = p \left( y^{(l)}_m | \theta, F_k = F_{nl} \right) \). This expression involves the calculation of the statistical moments of the likelihood function under the linear and non-linear probabilistic models. We first present in Lemma 1 a generic result for the moments of non-central Normal random variables, then derive in Lemma 2 the moments of the linear component, \( p^{F_l} \left( y^{(l)}_m | \theta \right) \) followed in Lemma 3 by the moments of the non-linear component, \( p^{F_{nl}} \left( y^{(l)}_m | \theta \right) \).
Lemma 1 (Moments of Normal random variables [37]).
The p-th moment of \( X \sim N(\mu_x, \sigma_x^2) \) is given by
\[
E [x^p] := \mu^{(p)} (\mu_x, \sigma_x^2) = \sigma_x^p \left( -i \sqrt{2} \right)^p U \left( \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \left( \frac{\mu_x}{\sigma_x} \right)^2 \right),
\]
where \( U (a, b, z) \) is the Tricomi's confluent function [38].

We now derive the statistical moments of the linear component \( p^T \mathbf{y} (\theta) \).

Lemma 2. The p-th moment of \( p^T \mathbf{y} (\theta) \) is given by
\[
\mathbb{E}_{p^T \mathbf{y} (\theta)} [p^T] := \mu^{(p)} \left( \Theta, \sigma_y^2 \right) \mu^{(p)} \left( \alpha \Theta, \alpha^2 \left( \Theta^2 \Theta \sigma_y^2 + \sigma_y^2 \right) \right)
\]

Proof. See Appendix A.

Next, we derive the statistical moments of the non-linear component \( p^T \mathbf{y}_n (\theta) \), where we first derive the Moment Generating Function of \( p^T \mathbf{y}_n (\theta) \) and then obtain the moments.

Lemma 3. The Moment Generating Function (MGF) of \( p^T \mathbf{y}_n (\theta) \) is given by
\[
M_T (t) = \exp \left( \Pi_{\Theta t} \phi \left( \frac{P_{\max} - \Pi \Theta}{\sqrt{\sigma_H^2 \Theta^2 + \sigma_V^2}} \right) - \phi \left( \frac{P_{\min} - \Pi \Theta}{\sqrt{\sigma_H^2 \Theta^2 + \sigma_V^2}} \right) \right) \times \exp \left( \frac{1}{2} \sigma_V^2 t^2 \right)
\]

Proof. See Appendix B.

In particular, the first moment of the non-linear component is given by:
\[
\mathbb{E}_{p^T \mathbf{y}_n} \left[ \mathbf{y}_n \right] = \Pi_{\Theta t} \phi \left( \frac{P_{\max} - \Pi \Theta}{\sqrt{\sigma_H^2 \Theta^2 + \sigma_V^2}} \right) - \phi \left( \frac{P_{\min} - \Pi \Theta}{\sqrt{\sigma_H^2 \Theta^2 + \sigma_V^2}} \right) \right) \times \exp \left( \frac{\mu^2}{2 \sigma_V^2} t^2 \right)
\]

IV. ESTIMATOR II: MAXIMUM LIKELIHOOD ESTIMATION VIA ASKEY POLYNOMIALS SERIES EXPANSION

In this section we develop the Maximum Likelihood Estimator (MLE) for \( \theta [21] \), which is given by:
\[
\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} p \left( \mathbf{y}^{(1:L)} \mid \theta \right),
\]
and is presented in the following Proposition.

Proposition 2 (Maximum Likelihood Estimator).
The Maximum Likelihood Estimator (MLE) of the state \( \theta \), denoted \( \hat{\theta}_{ML} \), is given by the solution to the optimization problem in (1), where \( f (x, \mu, \sigma^2) := \frac{1}{2} \phi \left( \frac{x - \mu}{\sigma} \right) \), and \( \Omega_{max} := \frac{P_{max} - \Pi \Theta}{\sqrt{\sigma_H^2 \Theta^2 + \sigma_V^2}} \).

Proof. See Appendix C.

Unfortunately, each of the integrals in Proposition 2 do not admit analytic solution. As such, the likelihood function cannot be calculated analytically, prohibiting the implementation of the MLE. We therefore develop a principled approach to approximating the likelihood function \( p \left( \mathbf{y}^{(1:L)} \mid \theta \right) \) via a basis expansion density estimation using Askey-orthogonal polynomial expansion. These types of density estimators enjoy a low computational complexity that can be controlled via truncation of the series and have many desirable properties [39], [40]. They have the generic form of a kernel density, \( g (y) \), multiplied by polynomials, \( P_n (y) \), known as Askey polynomials [24]. An important property of this family of series expansions is that it only requires the calculation of the moments of \( \mathbf{y}^{(1:L)} \). These moments can be easily obtained using Lemma 3, Lemma 3 and Eq. (2). The series expansions for the scalar case can be generically expressed as follows:
\[
p_Y (y) = g (y) \left( 1 + \sum_{n=1}^{\infty} \alpha_n P_n (y) \right),
\]
where \( \alpha_n \) is the n-th weight associated with the n-th polynomial. Many different choices for \( g (y) \) and \( P_n (y) \) exist, and in this work we concentrate on a particular choice that is based on the Gaussian kernel density and Hermite polynomials, known as the Gram-Charlier Series Expansion.

A. Gram-Charlier Series Expansion for the likelihood function

We now develop an approximation for the intractable likelihood function \( p \left( \mathbf{y}^{(1:L)} \mid \theta \right) \) via Gram-Charlier series expansion. The Gram-Charlier is an infinite series expansion that approximates a probability distribution, say \( g (y) \), in terms of its cumulants [33], [34]. The Gram-Charlier series expansion utilises a Gaussian kernel, \( g (y) \), given by
\[
g_{GC} (y) := \frac{1}{\sqrt{2 \pi} \kappa_2} \exp \left( -\frac{(y - \kappa_1)^2}{2 \kappa_2} \right),
\]
where \( \kappa_i \) is the i-th cumulant of \( g (y) \). The basis functions are defined as Hermite polynomials, \( H_n (y) \). These polynomials are defined in terms of the derivatives of \( g_{GC} (y) \), and can be obtained by the Rodrigues’ formula:
\[
H_n (y) = (-1)^n \frac{d^n g_{GC} (y)}{d^n y}.
\]
The Gram-Charlier series expansion is presented in the following Theorem.

**Theorem 1** (Gram-Charlier density series expansion [33, 34]).
The Gram-Charlier series expansion taking the generic form of (8) involves the kernel $g_{GC}(x)$ and the coefficients $a_n$. The Gram-Charlier series expansion is given by:

$$p_Y(y) = g_{GC}(y) \left(1 + \sum_{n=3}^{\infty} \frac{\kappa_n}{n!} (g_{GC}(y))^n \right),$$  \hspace{1cm} (11)

where the $n$-th cumulant of $Y$, $\kappa_n$ is given by:

$$\kappa_n = \frac{d^n \log M_Y(t)}{d^nt} |_{t=0},$$

where $M_Y(t)$ is the Generating Moment function of $Y$.

If we include only the first two correction terms, we obtain the Gram-Charlier A series expansion next.

**Proposition 3** (Special case: Fourth Order Expansion.).
The fourth order approximation of a probability distribution, $p_Y(y)$, via the Gram-Charlier series is given by

$$p_Y(y) \approx \frac{1}{\sqrt{2\pi \kappa_2}} \exp \left( -\frac{(y - \kappa_1)^2}{2\kappa_2} \right) \times \left(1 + \frac{\kappa_3}{6\kappa_2} H_3 \left( \frac{y - \kappa_1}{\sqrt{\kappa_2}} \right) + \frac{\kappa_4}{24\kappa_2} H_4 \left( \frac{y - \kappa_1}{\sqrt{\kappa_2}} \right) \right),$$

where $H_3(y) = y^3 - 3y$ and $H_4(y) = y^4 - 6y^2 + 3$ are the Hermite polynomials.

The Gram-Charlier series expansion, being a polynomial approximation, is not guaranteed to be positive, and may not be a valid probability distribution. Characterizing the model parameters that produce the “envelope” for the density approximation in which it will remain positive is therefore important [41, 42].

**B. Maximum Likelihood Estimator via Gram-Charlier Series Expansion**

Now that we presented the Gram-Charlier series expansion, we express the Gram-Charlier based MLE approximation as follows:

**Theorem 2** (Maximum Likelihood Estimation via Gram-Charlier Series Expansion).

The Gram-Charlier based Maximum Likelihood Estimator (MLE-GC) of the state $\theta$, denoted $\hat{\theta}_{ML-GC}$, is given by the solution for the optimization problem in (12), where $\Omega := \frac{y^{(l)} - \kappa_l^{2l}}{\kappa_l^{l}}$, and $\kappa_l^T$ and $\kappa_l^{2l}$ can be found by using Lemmas 2 and 3.

**Proof.** See Appendix D

In order to calculate the coefficients $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ for the Gram-Charlier density series expansion, we use the results we derived in Section III. For example $\kappa_1$ is given by:

$$\kappa_1 = \pi L_2^{(1)}(G, \sigma_G^2) \mu(1) \left( \alpha \bar{H} \theta, \alpha^2 \left( \bar{\sigma}_H^2 + \bar{\sigma}_W^2 \right) \right) + (1 - \pi_L) \left( \bar{H} \theta \alpha \left( \Phi(\Omega_{\max}) - \Phi(\Omega_{\min}) \right) \right) + \sqrt{\bar{\sigma}_H^2 \theta_2^2 + \bar{\sigma}_W^2 \alpha \left( \phi(\Omega_{\min}) - \phi(\Omega_{\max}) \right) \right).$$

**V. ESTIMATOR III: MARGINALIZED LEAST SQUARES**

In this Section we develop an estimator for $\theta$ which is based on the Least Squares (LS) method, developed independently by Gauss [43] and Legendre [27]. The LS solution is defined as follows:

$$\hat{\theta}_{LS} = \arg \min_{\theta \in \Theta} \sum_{m=1}^{M} \sum_{l=1}^{L} \left( y_{m}^{(l)} - G_{\min}^{(l)} \Phi \left( R_{\min}^{(l)} \right) \right)^2.$$
Proof.

Appendix E.

Proposition 4 (Marginalized Least Squares Estimator)
The M-LS estimator is presented in the following Proposition.

The Marginalized Least Squares (M-LS) Estimator of the state \( \theta \) is given by:

\[
\hat{\theta}_{\text{MLE-GC}} \approx \arg \max_{\theta \in \Theta} \prod_{l=1}^{L} \prod_{m=1}^{M} \left( \pi_l \frac{1}{\sqrt{2\pi \kappa_l^2}} \exp \left( -\frac{\Omega_l^2}{2} \right) \left( 1 + \frac{\kappa_l^3}{6} \right) H_3(\Omega) + \frac{\kappa_l^4}{24} H_4(\Omega) \right)
\]

\[
+ (1 - \pi_l) \left( \frac{1}{\sqrt{\sigma_y^2 + \sigma_G^2 (\alpha P_{\text{max}})^2}} \left( \frac{y_m^{(l)} - G_l \theta_{\text{min}}}{\sigma_y^2 + \sigma_G^2 (\alpha P_{\text{max}})^2} \right) \left( 1 - \Phi \left( \frac{P_{\text{max}} - \theta_{\text{max}}}{\sigma_z} \right) \right) \right)
\]

\[
+ \left( \frac{1}{\sqrt{2\pi \kappa_{\text{nl}}^2}} \exp \left( -\frac{\Omega_y^2}{2} \right) \left( 1 + \frac{\kappa_{\text{nl}}^3}{6} \right) H_3(\Omega) + \frac{\kappa_{\text{nl}}^4}{24} H_4(\Omega) \right)
\]

The M-LS estimator is presented in the following Proposition.

Proposition 4 (Marginalized Least Squares Estimator).
The Marginalized Least Squares (M-LS) Estimator of the state \( \theta \), denoted \( \hat{\theta}_{\text{M-LS}} \), is given by the solution to the optimization problem in (13), where \( \mathbb{E}_{H_{\text{m}}^{(1)},S_{\text{m}}^{(1)},V_{\text{m}}^{(1)}} \left[ F_{\text{R}} \left( H_{\text{m}}^{(1)} \right) \right] \) and \( \mathbb{E}_{H_{\text{m}}^{(1)},S_{\text{m}}^{(1)},V_{\text{m}}^{(1)}} \left[ \left( F_{\text{R}} \left( H_{\text{m}}^{(1)} \right) \right)^2 \right] \) are given in (15-16) in Appendix E.

Proof. See Appendix E. \( \square \)

VI. SIMULATION RESULTS

In this section we evaluate the performance of the three algorithms we developed for various system parameters via Monte Carlo simulations. We begin by evaluating the accuracy of the intractable likelihood function in Section IV-A via the Gram-Charlier density estimator. Following that, we present the estimation accuracy of the three algorithms.

A. Algorithms implementation and Computational Complexity

The three estimators involve solving a uni-variate optimization problem, where the objective functions are non-convex and non-linear. To have a simple algorithm, we implemented the solution using a Golden section search. Since the objective functions are non-convex, we used multiple starting points to avoid getting trapped in local minima. Using a single Golden section search, we start with an interval of length \( L \), in which \( \theta \) is likely to be in. In order to reach a final interval which is smaller than a pre-defined value \( \epsilon \) we need \( N \) iterations, where \( N \) is the first integer such that \( L(1 - \rho)^N \leq \epsilon \), where \( \rho \approx 0.382 \), is the Golden section parameter. Therefore we obtain that \( N \log (1 - \rho) \leq \log \frac{\epsilon}{\tau} \), and finally we have that the number of iterations for a single Golden section search is given by:

\[
N = \left\lfloor \frac{\log \frac{\epsilon}{\tau}}{\log (1 - \rho)} \right\rfloor.
\]

The computational complexity is therefore \( O \left( \log \frac{\epsilon}{\tau} \right) \) and is the same for all our algorithms, and depends of course on the accuracy resolution required, which is determined by \( \epsilon \).

B. Gram-Charlier series expansion likelihood function approximation accuracy

We now evaluate the accuracy of the Gram-Charlier series expansion in estimating the likelihood function \( p \left( y^{(1:M)} | \theta \right) \), presented in Section IV-A. The accuracy of this approximation is important and controls how well the MLE will perform. We compare our estimator with a Monte Carlo simulation in which we draw 10,000 samples from the model and plot its density. In Fig. 2 we present two different scenarios which illustrate the performance of the Gram-Charlier series expansion estimator. In the first, the system parameters are \( \{ \theta = 1, \sigma_G = \sigma_H = 0.1, \sigma_y = \sigma_v = 0.3, P_{\text{min}} = 3, P_{\text{max}} = 10, \pi_l = 0.5 \} \). We observe that the likelihood function is unimodal with slight skew to the right; in the second we use the same system parameters, but this time \( P_{\text{min}} = 7 \). We observe that in this case the likelihood function is bimodal and is captured by the estimator. These results illustrate that the Gram-Charlier series expansion provides high accuracy in recovering the likelihood function in both cases, making it suitable for our MLE algorithm.

C. Comparison of the estimation accuracy of the algorithms

In this section we evaluate and compare the accuracy of the three algorithms we developed. Our performance metric is the Mean Squared Error (MSE) defined as \( \text{MSE} = \frac{1}{N} \sum_{n=1}^{N} \left( \theta - \hat{\theta} \right)^2 \), where \( N = 5000 \) is the number of Monte Carlo simulations for each system parameter.

D. Effect of Non-linear Sensors

In Fig. 3 we study the effect of linear/non-linear sensors on the MSE of the three algorithms, as a function of the variance
\[ \hat{\theta}_{\text{MLS}} = \arg \min_{\theta \in \Theta} \sum_{m=1}^{M} \left( -2b_m^{(l)} \mathcal{G} (p_l \alpha H \theta + (1 - p_l) \mathbb{E}_{H_m^{(l)} W_m^{(l)}} \left[ F_m \left( R_m^{(l)} \right) \right]) + (\sigma_G^2 + \mathcal{G} p_l \mathbb{E}_{H_m^{(l)} W_m^{(l)}} \left[ F_m \left( R_m^{(l)} \right) \right])^2 \right) + (1 - p_l) \mathbb{E}_{H_m^{(l)} W_m^{(l)}} \left[ F_m \left( R_m^{(l)} \right) \right]^2 \right) \]

\[ (13) \]

(a) Unimodal likelihood function  
(b) Bimodal likelihood function

Fig. 2: Probability density estimation via Gram-Charlier series expansion estimation

of the additive noise \( V \) and \( W \). For simplicity we assume that their variance values are the same (\( \sigma^2_W = \sigma^2_V \)). The other system parameters are \( \{ M = 1, L = 10, \theta = 1, \sigma_G = \sigma_H = 0.1, P_{\text{min}} = -0.5, P_{\text{max}} = 0.5 \} \). We plot the MSE for the case where all the sensors are linear (\( \pi_l = 1 \)), for the case that \( \pi_l = 0.5 \) and for the case where all sensors are non-linear (\( \pi_l = 0 \)). We observe that the non-linear effect causes a small performance degradation for all three algorithms. It is interesting that the performance of the MLS algorithm is the best for low SNR in the cases that \( \pi_l = \{ 1, 0.5 \} \) but becomes the worst when only non-linear sensors are deployed \( \pi_l = 0 \) as well as in high SNR region.

The reason that the MLS algorithm outperforms the MLE and MOM estimators in low SNR when there is no censoring (\( \pi_l = 1 \)) is that in the low SNR region the values of the additive noise at the sensors may be very high. This has a detrimental effect on the MLE and MOM estimators, but less on the MLS algorithm which is a non-parametric approach, thus less sensitive to outliers in the data. In contrast, when there is censoring \( \pi_l = 0 \), the outliers in the additive noise get truncated and have a much smaller effect, thus the moments estimation is much more accurate, resulting in the MOM outperforming the MLS algorithm. In the case where the probability of censoring is \( \pi_l = 0.5 \), there is an averaging effect, which explains why all three estimators are compatible.

E. Effect of Frame Length

In Fig. 4 we study the effect of the frame length \( L \) has on the MSE of the three algorithms, as a function of the variance of the additive noise \( V \) and \( W \). For low SNR we observe an ordering of the performance as the frame length increases. However, in high SNR it seems that the MLS algorithm starts to perform poorly and encounters an error floor.

F. Effect of Number of Sensors

We now turn our attention to evaluating the performance of the proposed algorithms when a full sensor network is deployed, by varying the number of sensors \( M = \{ 10, 20, 30 \} \). The MSE results are presented in Fig. 5 for frame length \( L = 10 \). We observe that there is a big performance gap between the MLS and the other two algorithms, which means that the MLS algorithm is sensitive to the number of sensors deployed.

G. Effect of Sensing Quality and Wireless Channels

We also analyze the effect of sensing variance (\( \sigma_H \)) and wireless channel variance (\( \sigma_G \)) as shown in Figs. 6 and 7 respectively. The other system parameters are \( \{ M = 1, L = 10, \theta = 1, \sigma_w = \sigma_v = 0.01, P_{\text{min}} = -0.5, P_{\text{max}} = 0.5 \} \).
We developed three low computational complexity algorithms for parameter estimation in wireless sensor networks where the sensors observations exhibit random clipping of the signals due to internal failure or degradation: the Method of Moments estimator; the Maximum Likelihood Estimator; and the Marginalized Least Squares (M-LS) Estimator. We showed that the likelihood function can be estimated accurately using a non-parametric probability density estimator that is based on series expansions of the Gram-Charlier family of basis functions. Simulation results demonstrate the effectiveness of the proposed algorithms and also show that no single estimation technique dominates the other ones, and should be used according to the specific system parameters.

APPENDIX A

PROOF OF LEMMA 2

We derive the $p$-th moment of the linear component, according to the following steps:

1) Calculate the $p^{th}$ moment of $G \alpha (H \theta + V)$.
2) Calculate MGF of $Q := G \alpha (H \theta + V)$. 
3) Calculate MGF of $Y = Z + W$.
4) Calculate the moments of $Y$.

The $p$-th moment of the $Q$ is given as:


where we denote the following change of variables $Z := \alpha (H \theta + V)$. We obtain that $Z \sim N (\alpha H \theta, \alpha^2 (\theta^2 \sigma_H^2 + \sigma_V^2))$. Using Lemma 1 we have that

$$E_p^{\mathcal{G}_{(q)}[\theta]}[Q^p] := \mu^{(p)}(\mu, \sigma_H^2, \sigma_V^2) = \mu^{(p)}(\alpha H \theta, \alpha^2 (\theta^2 \sigma_H^2 + \sigma_V^2)).$$

Next, the MGF of $Q$ is calculated:

$$M_Q(q) = 1 + \sum_{p=1}^{\infty} \frac{q^pE[Z^p | \theta]}{p!}.$$ 

Due to the independence of $Q$ and $W$, the moment generating function for $Y$ is given by:

$$M_Y(t) = M_Q(t) M_W(t)$$

$$= \left(1 + \sum_{p=1}^{\infty} \frac{q^pE[Z^p | \theta]}{p!}\right) e^\frac{1}{2} \sigma_W^2 t^2.$$ 

Finally, the $p$-th moment of $p^{\mathcal{G}_{(q)}[\theta]}$ is given by

$$\mu^{(p)}(\theta) = \frac{d^pM_Y(t)}{dt^p} |_{t=0}$$

APPENDIX B

PROOF OF LEMMA 3

The $p$-th moment of the observation $y$ under non-linear transformation is given by:

$$E_p^{\mathcal{G}_{(y)}[\theta]}[y^p] = E_{G,H,V,W}[(G \alpha (H \theta + V) 1 (P_{min} < H \theta + V < P_{max}) + W)^p | \theta],$$

where we define $Z := H \theta + V$. We then repeat the same procedure as in Lemma 2.

Fig. 5: Effect of number of sensors on the estimation accuracy

Fig. 6: Effect of sensing uncertainty $\sigma_H$ on the estimation accuracy

Fig. 7: Effect of wireless channels uncertainty $\sigma_G$ on the estimation accuracy
The MLE is given by:

\[
\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} \left\{ y_{1:M}^{(l)} | \theta \right\}
\]

(14)

The first likelihood term is given by:

\[
p(y | \theta, F_R = \mathcal{F}_i) = \int_{-\infty}^{\infty} p(G \alpha (H \theta + V) + W | G, \theta) \, dG
\]

The second likelihood term is given by:

\[
p(y | \theta, F_R = \mathcal{F}_i) = \int_{-\infty}^{\infty} p(G \alpha (H \theta + V) + W | G, \theta) \, dG
\]

\[
= \int_{-\infty}^{\infty} \mathcal{N}(y | G \alpha \mathcal{P} \theta, (G \alpha)^2 (\theta^2 \sigma_H^2 + \sigma_V^2) + \sigma_V^2) \, dG
\]

\[
= \int_{-\infty}^{\infty} \mathcal{N}(y | G \alpha \mathcal{P} \theta, (G \alpha)^2 (\theta^2 \sigma_H^2 + \sigma_V^2) + \sigma_V^2) \, dG.
\]

The expectation with respect to the linear component is given by:

\[
\mathbb{E}_{G \alpha m_{Hm}, V_m} \left[ F_l \left( R^{(l)}_m \right) \right] = \alpha \mathbb{H} \theta
\]

The expectation with respect to the non-linear component is given by:

\[
\mathbb{E}_{G \alpha m_{Hm}, V_m} \left[ F_m \left( R^{(l)}_m \right) \right]
\]

\[
= \int_{-\infty}^{\infty} p(z_{1:m}) \left\{ \alpha z_{1:m} \mathbb{I} \left( P_{\text{min}} \leq z_{1:m} \leq P_{\text{max}} \right) \right\} \, dz_{1:m}
\]

\[
= \alpha \mathbb{H} \theta
\]

where we define \( C_{\text{min}} := \frac{P_{\text{min}} - \alpha H \theta}{\alpha \sqrt{\theta^2 \sigma_H^2 + \sigma_V^2}} \) and \( C_{\text{max}} := \frac{P_{\text{max}} - \alpha H \theta}{\alpha \sqrt{\theta^2 \sigma_H^2 + \sigma_V^2}} \). Next, we obtain:

\[
\mathbb{E}_{G \alpha m_{Hm}, V_m} \left[ \left( \mathcal{F}_l \left( R^{(l)}_m \right) \right)^2 \right] = p_l \cdot \mathbb{E}_{G \alpha m_{Hm}, V_m} \left[ \left( F_l \left( R^{(l)}_m \right) \right)^2 \right]
\]

\[
+ (1 - p_l) \cdot \mathbb{E}_{G \alpha m_{Hm}, V_m} \left[ \left( F_{\text{ne}} \left( R^{(l)}_m \right) \right)^2 \right].
\]

The expectation with respect to the non-linear component is given by:

\[
\mathbb{E}_{G \alpha m_{Hm}, V_m} \left[ \left( F_l \left( R^{(l)}_m \right) \right)^2 \right] = \alpha^2 \left( H^{(l)} \theta + V_m \right)^2
\]

\[
= \alpha^2 \left( \sigma_H^2 + \mathbb{H} \theta \right) + \alpha^2 \sigma_V^2.
\]
The expectation with respect to the non-linear component is given by:

$$E_{H^{(l)} \theta, Z^{(l)}} \left[ \left( \mathbf{F}_{m} \left( \mathbf{R}_{m}^{(l)} \right) \right)^{2} \right]$$

$$= E_{H^{(l)} \theta, Z^{(l)}} \left[ \left( \alpha (H\theta + V) \right)^{2} 1 \left( P_{\text{min}} < H\theta + V < P_{\text{max}} \right) \right]$$

$$+ E_{H^{(l)} \theta, Z^{(l)}} \left[ \left( \alpha P_{\text{min}} \right)^{2} 1 \left( P_{\text{min}} > H\theta + V \right) \right]$$

$$+ E_{H^{(l)} \theta, Z^{(l)}} \left[ \left( \alpha P_{\text{max}} \right)^{2} 1 \left( P_{\text{max}} < H\theta + V \right) \right]$$

$$= \alpha \sigma_{\theta}^{2} \sum_{m} \left( \frac{\mu_{Z}^{2}}{\sigma_{Z}^{2}} + 1 \right) \left( \Phi(C_{\text{max}}) - \Phi(C_{\text{min}}) \right)$$

$$- \left( \frac{\mu_{Z} + P_{\text{max}}}{\sigma_{Z}} \Phi(C_{\text{max}}) - \frac{\mu_{Z} + P_{\text{min}}}{\sigma_{Z}} \Phi(C_{\text{min}}) \right)$$

$$+ \alpha^{2} \sigma_{\theta}^{2} \sigma_{Z}^{2} \Phi(C_{\text{min}}) \left( 1 - \Phi(C_{\text{max}}) \right),$$

where we define $C_{\text{min}} := \frac{P_{\text{min}} - \alpha \theta}{\sqrt{\alpha^{2} (\sigma_{\theta}^{2} + \sigma_{Z}^{2})}}$, and use the following change of variable:

$$Z := \alpha (H\theta + V),$$

and obtain that:

$$Z \sim N \left( \alpha \overline{H}\theta, \sigma^{2} \left( \overline{\theta}^{2} \sigma_{H}^{2} + \sigma_{Z}^{2} \right) \right),$$

with $\mu_{Z} = \alpha \overline{H}\theta$, $\sigma_{Z} = \alpha \sqrt{\overline{\theta}^{2} \sigma_{H}^{2} + \sigma_{Z}^{2}}$.


Ido Nevat received the B.Sc. degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel in 1998 and the Ph.D. degree in Electrical Engineering from the University of NSW, Sydney, Australia, in 2010. Between 2010-2013 he was a postdoctoral research fellow with the Wireless and Networking Technologies Laboratory at CSIRO, Australia. Between 2013-2016 he was a scientist at the Institute for Infocomm Research (I2R), Singapore. Since 2017 Ido has been a team leader at TUMCREATE in Singapore. His main areas of interests include statistical signal processing, machine learning and Bayesian statistics.

Pingfei Zhang received his Bachelor of Engineering in Electrical & Electronic Engineering and Ph.D from Nanyang Technological University in 2010 and 2015 respectively. He is now working in department of engineering science in Oxford University as postdoctoral researcher. The topic he is working on is machine learning in wireless sensor networks. Before that, he worked as a research scientist in the Institute for Infocomm Research (I2R), Singapore. Since 2017 Pingfei has been a team leader at TUMCREATE in Singapore. His main areas of interests include energy efficient clustering algorithms, energy harvesting wireless sensor networks (EH-WSN), and statistical modeling in WSN.

Gad Frenkel received the B.Sc. degree in Physics and Mathematics from the Tel Aviv University - Israel in 1997 and the Ph.D. degree in Condensed Matter Physics from the Tel Aviv University in 2004. In 2005-2008 he worked as a Postdoctoral Research Associate, at Imperial College London - UK, on Statistical mechanics of granular porous media. During 2009 he was a lecturer at the Holon Institute of Technology. Since 2010 he is a lecturer at the school of engineering of the Ruppin Academic Center, Israel. His main areas of interest are statistical mechanics and its application to noisy systems such as granular flows, complex fluids and motion of tracers in noisy flows.

Gareth W. Peters is an Assistant Professor in the Department of Statistical Science in University College London. He is a Principle Investigator in CSML, University College London (UCL) and an Academic Member of the UK PhD Center in Financial Computing (UCL). He has published in excess of 100 peer reviewed articles on risk and insurance modelling, 2 research text books on Operational Risk and Insurance as well as being the editor and contributor to 3 edited text books on spatial statistics and Monte Carlo methods. He holds positions as an Adjunct Scientist in the Mathematics, Informatics and Statistics, Commonwealth Scientific and Industrial Research Organisation (CSIRO) since 2009 as well being an Associate Member Oxford-Man Institute in Oxford University since 2012, an Associate Member Systemic Risk Center in London School of Economics since 2014; an Affiliated Prof. School of Earth and Space Sciences, Peking University PKU, Beijing since 2015 and a Visiting Prof. in the Institute of Statistical Mathematics, Tokyo, Japan each year since 2010.