

# Efficient Parametric Signal Estimation from Samples with Location Errors

Sumeet Kumar\*, *Student Member, IEEE*, Vivek K. Goyal, *Senior Member, IEEE*,  
and Sanjay E. Sarma

## Abstract

We introduce an iterative linear estimator (ILE) for estimating a signal from samples having location errors and additive noise. We assume that the signals lie in the span of a finite basis and the location errors and noise are mutually independent and normally distributed. The parameter estimation problem is formulated as obtaining a maximum likelihood (ML) estimate given the observations and an observation model. Using a linearized observation model we derive an approximation to the likelihood function. We then adopt an iterative strategy to develop a computationally efficient estimator, which captures the first order effect of sample location errors on signal estimation. Through numerical simulations we establish the efficacy of the proposed estimator for one-dimensional and two-dimensional parametric signals, comparing the mean squared estimation error against a basic linear estimator. We develop a numerical approximation of the Cramér-Rao lower bound (CRB) and the Expectation-Maximization (EM) algorithm, and for a one-dimensional signal compare our algorithm against them. We show that for high location error variance and small noise variance the mean squared error (MSE) with ILE is significantly lower when compared to the baseline linear estimator. When compared to EM, our algorithm provides comparable MSE with a significant reduction in computational time.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

S. Kumar and S. E. Sarma are with the Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA e-mail: [sumeetkr@mit.edu](mailto:sumeetkr@mit.edu); [sesarma@mit.edu](mailto:sesarma@mit.edu).

V. K. Goyal is with the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA e-mail: [vgoyal@mit.edu](mailto:vgoyal@mit.edu).

Manuscript received February 22, 2013; revised July 3, 2013; accepted July 12, 2013.

### Index Terms

Parameter estimation, iterative estimation, jitter, unknown location, maximum likelihood estimator, Expectation Maximization, Cramér-Rao lower bound.

## I. INTRODUCTION

Pervasive sensing through wireless sensor networks (WSNs) has the potential to transform the way we monitor our built and natural environments by providing measurements at spatial and temporal scales that were not possible a few years ago [1], [2]. Furthermore, the compactness of today's sensors allow them to be easily attached to mobile platforms such as robots, cars, buoys, humans, and animals to achieve wide-area coverage [3]. For example, air quality sensors attached to taxicabs can provide emissions data in urban areas via smartphones. One key application of such mobile WSNs is to reconstruct a spatio-temporal field such as temperature or chemical concentration. Several approaches have been explored in the literature for reconstructing such fields in the presence of measurement/modeling errors [4]–[8].

Researchers have studied and developed various localization schemes for mobile platforms. GPS, which can be used to obtain position information of mobile sensing platforms in outdoor environments like taxicabs, has an accuracy of few meters [9], [10]. Commercially available differential GPS can further achieve sub-meter accuracy [11]. Localization of autonomous robots has been studied extensively in the literature [12]–[16]. Though significant advances have been made in the localization of autonomous robots, location errors exist and their effect on the reconstruction of a spatio-temporal field monitored by a system of mobile agents has not received much attention. As the interest in real-world deployment of mobile WSNs is increasing, the aforementioned problem is expected to be of utmost practical interest. Tolerating higher sample location errors would in turn allow the use of simpler and less expensive localization/positioning systems, increasing the feasibility of real-world deployment of mobile sensor nodes.

Location errors in sampling have been studied in the signal processing community as random time jitter in analog-to-digital converters (ADC) [17]–[21]. Typically the power consumption of an ADC goes up with the speed and the accuracy of the clock signals [20]. Hence tolerating higher time jitter through post processing presents an opportunity for lowering the power of acquisition subsystems. In [17]–[19], [22] the authors characterize error due to time jitter and study estimation error of bandlimited signals in the presence of jitter. In [17] the author develops a linear interpolator for processing bandlimited signals with correlated and uncorrelated time jitter. In [20] the authors implement a nonlinear maximum

likelihood estimator through Expectation Maximization to mitigate the effect of sample location error in the reconstruction of bandlimited signals. Likewise, Bayesian approaches through Gibbs sampling and slice sampling have been explored to nonlinearly post-process the jitter affected samples, showing success in reducing the effect of time jitter [21]. Nonlinear estimators like the polynomial least squares estimator has also been explored in the literature [23]. Though nonlinear estimators exhibit improved performance over linear estimators, in general they suffer from high computational cost and limit the possibility of developing real-time embedded applications.

In this paper, we develop a computationally efficient estimator to mitigate the effect of sample location errors in reconstructing signals that lie in a finite-dimensional subspace. The parameter estimation problem is formulated as obtaining a maximum likelihood estimate given the observations and an observation model. Using the linear term from a Taylor series expansion, we derive an approximation to the nonlinear observation model. We then adopt an iterative approach to develop a computationally efficient estimator, the *iterative linear estimator* (ILE), which captures the first order effect of sample location errors on signal estimation. We highlight practical issues in computing the numerical approximation of the complete likelihood function and show the suitability of Monte Carlo integration over Gauss-Hermite quadrature for our problem. Using Monte Carlo integration we derive a numerical scheme for finding the Cramér-Rao lower bound (CRB) and an Expectation Maximization (EM) algorithm to implement the maximum likelihood estimation.

We present simulations to show the efficacy of our proposed algorithm in improving the estimation error at low computational cost. When compared with a baseline linear estimator, ILE gives a lower estimation error and the improvement increases with the variance of the sample location error but decreases with the number of sampling points. We present statistical properties of our estimator, like contribution of bias versus variance and the mean computation time. The mean computational time with ILE for the simulated one-dimensional signals is of the order of  $10^{-2}$  s which increases to 1 s for the simulated two-dimensional signals. For a one-dimensional signal, we show that for low additive noise and a reasonable level of oversampling ILE outperforms EM (in terms of estimation error) while at higher noise and higher number of samples EM algorithm gives a lower estimation error. We also show that the performance of EM worsens with decreasing additive noise due to high numerical errors associated with computing integrals that involve narrow Gaussian distributions, while ILE does not suffer from such issues. Furthermore, implementation of ILE is around five orders of magnitude faster when compared to EM, making it an appealing candidate for real-time embedded applications.

Our approach is related to the Taylor-series based vehicle position estimation [24], where the known

functional relationships between the unknown vehicle position and the known observation stations are linearized to obtain a linear system of equations which is subsequently solved iteratively until convergence. The key difference between our work and [24] is that in our case the Taylor-series approximation does not lead to a linear system of equations as both the function and the location errors are unknown.

### A. Notation

Let  $\{\phi_k(\mathbf{x})\}_{k=1}^M$  be an orthonormal set of continuous and differentiable functions defined on the  $D$ -dimensional domain  $Q \triangleq [-1, 1]^D$ , meaning

$$\langle \phi_k(\mathbf{x}), \phi_l(\mathbf{x}) \rangle = \int_Q \phi_k(\mathbf{x}) \phi_l(\mathbf{x}) d\mathbf{x} = \begin{cases} 1, & \text{if } k = l; \\ 0, & \text{if } k \neq l. \end{cases} \quad (1)$$

We consider signals in the span of  $\{\phi_k(\mathbf{x})\}_{k=1}^M$ :

$$f(\mathbf{x}) = \sum_{k=1}^M a_k \phi_k(\mathbf{x}). \quad (2)$$

We denote the  $\mathcal{L}_2$  norm of the signals as  $\|f(\mathbf{x})\| = \sqrt{\langle f(\mathbf{x}), f(\mathbf{x}) \rangle}$ . For a vector in  $D$  dimensional space we use the regular  $\mathcal{L}_2$  norm as  $\|\mathbf{x}\| = \sqrt{\sum_{j=1}^D x_j^2}$ . The determinant of a matrix  $M$  is denoted as  $|M|$ .

The signal estimation problem can be defined as obtaining an estimate of  $a_k$ 's by appropriate sampling of the signal and employing an estimation algorithm. Note that here we do not restrict our analysis to bandlimited functions, making our framework suitable for a wide range of spatio-temporal signals frequently encountered in the real-world application of pervasive sensing. We assume uniform arrangements of the nominal/expected sampling locations with equal spacing between the samples in every dimension. We assume the location errors to be independent, identically distributed (i.i.d.) Gaussian noise and the measurement/modeling error to be additive i.i.d. Gaussian noise. We then have the following observation model: For each  $i = 1, 2, \dots, N$ ,

$$y_i = f(\mathbf{x}_i + \mathbf{z}_i) + \eta_i, \quad (3)$$

where  $\mathbf{x}_i$  is the nominal sensor location,  $\mathbf{z}_i \sim \mathcal{N}(0, \sigma_z^2 I)$  is the location error (with per-component standard deviation  $\sigma_z$ ),  $\eta_i \sim \mathcal{N}(0, \sigma^2)$  is the additive noise (with standard deviation  $\sigma$ ) and  $y_i$  is the observation. The noise matrix  $\{\mathbf{z}_i^T\}_{i=1}^N$  and the noise vector  $\{\eta_i\}_{i=1}^N$  are independent. In theory, the Gaussian distribution assumption of location error may lead the sampling points to be outside  $Q$ . In simulations, we ensure the points are in  $Q$  by considering location errors to be a fraction of the spacing between the nominal sampling points. The assumption of Gaussian distribution enables an analytical treatment of the problem and subsequent derivation of the estimator.

In matrix and vector notation the observations can be described in the following semilinear form:

$$\mathbf{y} = \Phi(X, Z)\mathbf{a} + \boldsymbol{\eta}, \quad (4)$$

where  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ ,  $\mathbf{a} = [a_1, a_2, \dots, a_M]^T$ ,  $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_N]^T$ ,  $X = \{\mathbf{x}_i | i = 1 \dots N\}$ ,  $Z = \{\mathbf{z}_i | i = 1 \dots N\}$  and

$$\Phi(X, Z) = \begin{bmatrix} \phi_1(\mathbf{x}_1 + \mathbf{z}_1) & \cdots & \phi_M(\mathbf{x}_1 + \mathbf{z}_1) \\ \phi_1(\mathbf{x}_2 + \mathbf{z}_2) & \cdots & \phi_M(\mathbf{x}_2 + \mathbf{z}_2) \\ \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_N + \mathbf{z}_N) & \cdots & \phi_M(\mathbf{x}_N + \mathbf{z}_N) \end{bmatrix}.$$

For convenience we denote the  $j^{\text{th}}$  row of  $\Phi(X, Z)$  as  $\phi_j(\mathbf{z}_j)^T = [\phi_1(\mathbf{x}_j + \mathbf{z}_j), \dots, \phi_M(\mathbf{x}_j + \mathbf{z}_j)]$ . From the observation model we can clearly see that the effect of  $\mathbf{z}_i$  on  $y_i$  is nonlinear and non-trivial to characterize. We denote the probability density function (pdf) of  $\mathbf{x}$  by  $p(\mathbf{x})$ , the pdf of  $\mathbf{x}$  parametrized by a nonrandom vector  $\mathbf{a}$  by  $p(\mathbf{x}; \mathbf{a})$ , and the pdf of  $\mathbf{x}$  conditioned on a random variable  $\mathbf{y}$  by  $p(\mathbf{x}|\mathbf{y})$ . We define expectations similarly. We denote the multivariate normal distribution of a  $D$  dimensional vector  $\mathbf{x}$  by

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = |(2\pi)^D \Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad (5)$$

where  $\boldsymbol{\mu}$  is the mean vector and  $\Sigma$  is the covariance matrix.

### B. Parameter Estimation

In parameter estimation, we seek to find an estimate  $\hat{\mathbf{a}}$  of  $\mathbf{a}$  which minimizes a certain cost function. Typically the cost function is chosen to be the MSE:

$$C(\hat{\mathbf{a}}(\mathbf{y}); \mathbf{a}) = E_{\mathbf{y}} \left[ \|\hat{\mathbf{a}} - \mathbf{a}\|^2 \right]. \quad (6)$$

Using the estimate of the parameters we reconstruct the signal using

$$\hat{f}(\mathbf{x}) = \sum_{k=1}^M \hat{a}_k \phi_k(\mathbf{x}) \quad (7)$$

As discussed in [25] it is impossible to compute the minimum MSE (MMSE) estimator without prior knowledge about  $\mathbf{a}$ . From an application perspective it is important to derive an estimator based only on the observation model and the actual observations. The maximum likelihood (ML) estimator is one such estimator which uses the observations and the observations model to find an estimate that maximizes the

likelihood function  $l(\mathbf{a}; \mathbf{y}) \triangleq p(\mathbf{y}; \mathbf{a})$ . The likelihood function corresponding to the observation model in (3) can be written as

$$l(\mathbf{a}; \mathbf{y}) = \int_Q p(\mathbf{y}|\mathbf{z}; \mathbf{a})p(\mathbf{z})d\mathbf{z} = \int_Q \mathcal{N}(\mathbf{y}; \Phi(X, Z)\mathbf{a}, \sigma^2 I)\mathcal{N}(\mathbf{z}; 0, \sigma_z^2 I)d\mathbf{z}. \quad (8)$$

We now have the following definition of the ML estimator:

$$\hat{\mathbf{a}}_{ML} = \arg \max_{\mathbf{a}} l(\mathbf{a}; \mathbf{y}). \quad (9)$$

Maximizing (8) is not straightforward and does not lead to an analytical solution. In [20] the EM algorithm is employed to obtain the ML estimate. EM is computationally expensive and may not be suitable for real-time embedded applications. In this paper, we present an approximation to the observation model and the subsequent development of ILE and discuss its performance.

### C. Outline

In Section II, we discuss the development of ILE. Section III first presents some background and practical issues related to numerical integration. We then develop a numerical approximation of the Cramér-Rao lower bound on the variance of any unbiased estimator and an EM approximation for ML estimation. We present simulations providing comparisons and discussions in Section IV. In conclusion we summarize the results and our contributions and introduce future research directions.

## II. ITERATIVE LINEAR ESTIMATOR (ILE)

Using the Taylor series expansion and keeping only the first order term, we obtain the following linearization of the observation model:

$$y_i = f(\mathbf{x}_i) + \mathbf{z}_i^T \nabla f(\mathbf{x}_i) + \eta_i, \quad (10)$$

where  $\nabla f(\mathbf{x}_i)$  is the gradient of the function calculated at  $\mathbf{x}_i$ . Note that for one-dimensional signals the linearized observation model takes the following simplified form

$$y_i = f(x_i) + z_i f'(x_i) + \eta_i. \quad (11)$$

Since the sum of two independent Gaussian random variables is Gaussian, we can write the marginal pdf of  $y_i$  as

$$p(y_i; \mathbf{a}) = \frac{1}{\sqrt{2\pi(\sigma^2 + \|\nabla f(\mathbf{x}_i)\|^2 \sigma_z^2)}} \exp \left\{ -\frac{(y_i - f(\mathbf{x}_i))^2}{2(\sigma^2 + \|\nabla f(\mathbf{x}_i)\|^2 \sigma_z^2)} \right\}. \quad (12)$$

For one-dimensional signals we replace  $\|\nabla f(\mathbf{x}_i)\|^2$  with  $f'(x_i)^2$ . Note that higher order Taylor expansion, for example  $y_i = f(x_i) + z_i f'(x_i) + z_i^2 f''(x_i) + \eta_i$ , will lead to a complicated pdf of  $y_i$  (square of Gaussian

random variable is not Gaussian) and hence it will be difficult to obtain an analytical expression for the marginal pdf of  $y_i$ . Since all the observations are independent we have  $p(\mathbf{y}; \mathbf{a}) = \prod_{i=1}^N p(y_i; \mathbf{a})$  and hence the approximated likelihood function  $\hat{l}(\mathbf{a}; \mathbf{y})$  can be written as

$$\hat{l}(\mathbf{a}; \mathbf{y}) = \frac{1}{\sqrt{(2\pi)^N \prod_{i=1}^N (\sigma^2 + \|\nabla f(\mathbf{x}_i)\|^2 \sigma_z^2)}} \exp \left\{ - \sum_{i=1}^N \frac{(y_i - f(\mathbf{x}_i))^2}{2(\sigma^2 + \|\nabla f(\mathbf{x}_i)\|^2 \sigma_z^2)} \right\}. \quad (13)$$

Our aim is to find an estimator  $\hat{\mathbf{a}}$  that maximizes  $\hat{l}(\mathbf{a}|\mathbf{y})$ . Note that  $f(\mathbf{x}) = \sum_{k=1}^M a_k \phi_k(\mathbf{x})$  and  $\nabla f(\mathbf{x}) = \sum_{k=1}^M a_k \nabla \phi_k(\mathbf{x})$ . The derivative of  $\hat{l}(\mathbf{a}; \mathbf{y})$  with respect to  $\mathbf{a}$  will still result in a nonlinear set of equations making the estimation problem difficult.

We now present an iterative strategy to obtain an estimate of  $\mathbf{a}$  using the linearized likelihood function, (13). In the absence of sample location error the resulting observation model is linear with additive i.i.d. Gaussian noise, i.e.  $\mathbf{y} = \Phi(X, 0)\mathbf{a} + \boldsymbol{\eta}$ . The minimum variance unbiased estimator (MVUE) with the above observation model is

$$\mathbf{a}_b = (\Phi(X, 0)^T \Phi(X, 0))^{-1} \Phi(X, 0)^T \mathbf{y}. \quad (14)$$

We call the estimator defined in (14) the *blind* estimator (BE). Using  $\mathbf{a}_b$  we can obtain an initial estimate of the unknown signal as  $f_b(\mathbf{x})$  using (7) and its gradient  $\nabla f_b(\mathbf{x})$ . Using the approximation for the gradient we write (13) as

$$\tilde{l}(\mathbf{a}; \mathbf{y}) = \frac{1}{\sqrt{(2\pi)^N \prod_{i=1}^N (\sigma^2 + \|\nabla f_b(\mathbf{x}_i)\|^2 \sigma_z^2)}} \exp \left\{ - \sum_{i=1}^N \frac{(y_i - f(\mathbf{x}_i))^2}{2(\sigma^2 + \|\nabla f_b(\mathbf{x}_i)\|^2 \sigma_z^2)} \right\}. \quad (15)$$

We take the derivative of the approximated likelihood function (15) to obtain the following linear systems of equations:

$$\begin{aligned} \frac{\partial \tilde{l}(\mathbf{a}; \mathbf{y})}{\partial a_j} &= 0 \\ \Rightarrow \sum_{i=1}^N \frac{y_i \phi_j(\mathbf{x}_i)}{\sigma^2 + \|\nabla f_b(\mathbf{x}_i)\|^2 \sigma_z^2} &= \sum_{k=1}^M a_k \sum_{i=1}^N \frac{\phi_k(\mathbf{x}_i) \phi_j(\mathbf{x}_i)}{\sigma^2 + \|\nabla f_b(\mathbf{x}_i)\|^2 \sigma_z^2} \\ \forall j &= 1 \dots M. \end{aligned} \quad (16)$$

We define

$$\alpha_i = \sigma^2 + \|\nabla f_b(\mathbf{x}_i)\|^2 \sigma_z^2, \quad (17)$$

and

$$C = \text{diag}(1/\alpha_1, 1/\alpha_2, \dots, 1/\alpha_N). \quad (18)$$

Solving the above system of linear equations we obtain the following estimator:

$$\mathbf{a}_{ILE} = (\Phi(X, 0)^T C \Phi(X, 0))^{-1} \Phi(X, 0)^T C \mathbf{y}. \quad (19)$$

We implement the estimator in (19) iteratively by using the estimate  $\mathbf{a}_{ILE}^t$  obtained above to calculate  $\nabla f_b(\mathbf{x}_i)$  using  $\nabla f(\mathbf{x}) = \sum_{k=1}^M a_{ILE,k}^t \nabla \phi_k(\mathbf{x})$ , where  $t$  is the current iteration index.  $\alpha_i$ 's are subsequently recalculated which are then used to obtain a new estimate of  $\mathbf{a}$  using (19) as  $\mathbf{a}_{ILE}^{(t+1)}$ . The procedure is repeated until convergence, i.e.,  $\|\mathbf{a}_{ILE}^{(t+1)} - \mathbf{a}_{ILE}^t\| \leq \epsilon$ , where  $\epsilon$  is the tolerance for convergence or a maximum number of iterations have been reached. Algorithm 1 summarizes the steps involved, and the performance of the estimator is presented in Section IV.

---

**Algorithm 1** Iterative Linear Estimator (ILE) for signal reconstruction under measurement and sampling location errors

---

**Require:**  $\mathbf{y}, X, \phi_k(\mathbf{x})'s \forall k = 1 \dots M, \epsilon$  (convergence tolerance),  $P$  (maximum number of iterations)

Use (14) to find  $\mathbf{a}_b$

Find  $f_b^1(\mathbf{x})$  using (7) and compute  $\nabla f_b^1(\mathbf{x})$

$t \leftarrow 0$

$\hat{\mathbf{a}}^t = \mathbf{a}_b$

**repeat**

$t \leftarrow t + 1$

Compute  $\alpha_i = \sigma^2 + \|\nabla f_b^t(\mathbf{x}_i)\|^2 \sigma_z^2 \forall i = 1 \dots N$

Create  $C^t = \text{diag}(1/\alpha_1, 1/\alpha_2, \dots, 1/\alpha_N)$

Compute  $\hat{\mathbf{a}}^t$  from (19)

Using  $\hat{\mathbf{a}}^t$  find  $f_b^{(t+1)}(\mathbf{x})$  using (7) and compute  $\nabla f_b^{(t+1)}(\mathbf{x})$

**until**  $\|\hat{\mathbf{a}}^t - \hat{\mathbf{a}}^{(t-1)}\| \leq \epsilon$  or  $t \geq P$

**return**  $\hat{\mathbf{a}}^t$

---

### III. CRAMÉR-RAO LOWER BOUND AND EXPECTATION MAXIMIZATION ALGORITHM

Since for a general choice of basis functions, no closed form solution exists to the likelihood function given by (8), a numerical method is needed to approximate (8). This numerical approximation is a

precursor for either obtaining the CRB or implementing an EM approximation to the ML estimator (9) as discussed in [20]. In this section we first discuss techniques and issues involved in obtaining an approximation to (8). On the basis of our numerical approximation scheme we then present an approximation to the CRB and an EM algorithm for ML estimation. We present the formulation and simulation results for one-dimensional signals and highlight the challenges associated with higher dimensional signals.

### A. Numerical Integration

Since the observations and the location errors are independent, (8) can be written as

$$l(\mathbf{a}; \mathbf{y}) = \prod_{i=1}^N \int_{\mathbf{z}_i} \mathcal{N}(y_i; \phi_i(\mathbf{z}_i)^T \mathbf{a}, \sigma^2) \mathcal{N}(\mathbf{z}_i; 0, \sigma_z^2 I) d\mathbf{z}_i = \prod_{i=1}^N p(y_i; \mathbf{a}). \quad (20)$$

Numerical approximation of the marginal pdf of the observations  $y_i$ ,  $p(y_i; \mathbf{a})$ , involves estimating the  $D$  dimensional integral

$$p(y_i; \mathbf{a}) = \int \cdots \int_{z_{i,1} \cdots z_{i,D}} \mathcal{N}(y_i; \phi_i(\mathbf{z}_i)^T \mathbf{a}, \sigma^2) \mathcal{N}(z_{i,1}; 0, \sigma_z^2) \cdots \mathcal{N}(z_{i,D}; 0, \sigma_z^2) dz_{i,1} \cdots dz_{i,D}, \quad (21)$$

where  $\mathbf{z}_i = [z_{i,1}, z_{i,2}, \dots, z_{i,D}]^T$  is the sample location error at the sampling point  $\mathbf{x}_i$ . Clearly the complexity of estimation grows with the dimension of the signal. For a one-dimensional signal the marginal has a simpler expression:

$$p(y_i; \mathbf{a}) = \int_{z_i} \mathcal{N}(y_i; \phi_i(z_i)^T \mathbf{a}, \sigma^2) \mathcal{N}(z_i; 0, \sigma_z^2) dz_i = E_{z_i}[\mathcal{N}(y_i; \phi_i(z_i)^T \mathbf{a}, \sigma^2)]. \quad (22)$$

There are two major approaches to obtain a numerical approximation to the marginal pdf (22). The first method, called quadrature integration, approximates the weighted integral with a weighted sum [26], i.e.,

$$p(y_i; \mathbf{a}) \approx \sum_{j=1}^J p(y_i | z_{i,j}, \mathbf{a}) w_j, \quad (23)$$

where  $z_{i,j}$ 's are the abscissas,  $w_j$ 's are the weights,  $J$  is the number of points for the quadrature integration and  $p(y_i | z_{i,j}, \mathbf{a}) = \mathcal{N}(y_i; \phi_i(z_{i,j})^T \mathbf{a}, \sigma^2)$ . Since  $z_i$  is normally distributed, the appropriate choice of abscissas and weights are derived through the Gauss-Hermite quadrature (GHQ) rule [26]. Using a Cartesian product of the one-dimensional quadrature grid for evaluating multidimensional integrals leads to an exponential increase in the computational complexity (number of function evaluations  $\sim J^D$ ). Though use of sparse grids provides improved performance, the complexity still increases significantly with the dimension [27].

Monte Carlo integration (MCI) approximates the integration by first generating  $S$  samples of  $z_i$  from its distribution and uses the samples  $(z_{i,s})$  to approximate the integral as

$$p(y_i; \mathbf{a}) \approx \frac{1}{S} \sum_{s=1}^S p(y_i | z_{i,s}, \mathbf{a}). \quad (24)$$

It is known that the approximation error is proportional to  $\text{var}(f)/\sqrt{S}$ , where  $\text{var}(f)$  is the variance of the function to be integrated. Though the error is independent of the dimension of the problem [28], the variance ( $\text{var}(f)$ ) can grow exponentially with the dimension, requiring an exponential increase in the number of samples for achieving a desired accuracy. Use of quasi-Monte Carlo integration can further improve the performance for higher dimensional integrals [28].

We present an example of comparing the performance of GHQ and MCI in computing the numerical approximation of the marginal pdf (22) for a one-dimensional signal. We chose  $1$ ,  $\sin(kx)$  and  $\cos(kx)$  as the choice of basis functions for a specific frequency,  $k$ , and create an orthonormal set  $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$  from them [8]. We chose  $\mathbf{a} = [0.5, 1, -1]^T$  and  $f(x) = 0.5\phi_1(x) + \phi_2(x) - \phi_3(x)$ . For a specific sampling location, say  $x_i = 0$  and  $k = \{1, 5, 10\}$  we evaluate the performance of GHQ and MCI. We first generate  $10^5$  samples of  $y_i$  according to (3) and generate histograms to obtain an empirical estimate of the pdf of  $y_i$ . We use an open source code [29] to implement GHQ (23) with  $J = 128$ . For MCI, we first generate  $10^4$  samples of  $z_i$  by sampling from a zero mean normal distribution and given standard deviation ( $\sigma_z$ ) and estimate the pdf by (24). Fig. 1a,b,c shows that as the frequency of the signal increases the MCI outperforms GHQ. Similarly, Fig. 1a,d shows that for the same frequency,  $k = 1$ , as the sample location error ( $\sigma_z$ ) increases, MCI provides a better approximation of the pdf compared to GHQ. In the rest of the paper we employ MCI for approximating marginal pdfs and expectations.

### B. CRB and its Numerical Approximation

The CRB provides a lower bound on the variance of any unbiased estimator. The  $\text{CRB}(\mathbf{a})$  is defined as the trace of the inverse of the Fisher information matrix  $I_{\mathbf{y}}(\mathbf{a})$ . Under certain regularity conditions [30],  $I_{\mathbf{y}}(\mathbf{a})$  satisfies the following:

$$I_{\mathbf{y}}(\mathbf{a}) \triangleq E \left[ \left( \frac{\partial \log l(\mathbf{a}; \mathbf{y})}{\partial \mathbf{a}} \right) \left( \frac{\partial \log l(\mathbf{a}; \mathbf{y})}{\partial \mathbf{a}} \right)^T \right] = -E \left[ \frac{\partial^2 \log l(\mathbf{a}; \mathbf{y})}{\partial \mathbf{a} \partial \mathbf{a}^T} \right], \quad (25)$$

where the expectation is taken with respect to  $p(\mathbf{y}; \mathbf{a})$  (20). Using (20),  $\log l(\mathbf{a}; \mathbf{y}) = \sum_{i=1}^N \log p(y_i; \mathbf{a})$ , hence we can write (25) as

$$I_{\mathbf{y}}(\mathbf{a}) = \sum_{i=1}^N E \left[ \left( \frac{\partial \log p(y_i; \mathbf{a})}{\partial \mathbf{a}} \right) \left( \frac{\partial \log p(y_i; \mathbf{a})}{\partial \mathbf{a}} \right)^T \right]. \quad (26)$$

Now  $\partial \log p(y_i; \mathbf{a}) / \partial \mathbf{a} = (1/p(y_i; \mathbf{a}))(\partial p(y_i; \mathbf{a}) / \partial \mathbf{a})$  where we compute  $p(y_i, \mathbf{a})$  from (24). We can write

$$\frac{\partial p(y_i; \mathbf{a})}{\partial \mathbf{a}} \approx \frac{1}{S} \sum_{s=1}^S \left\{ \frac{(y_i - \phi_i(z_{i,s})^T \mathbf{a}) \phi_i(z_{i,s})}{\sigma^2} \mathcal{N}(y_i; \phi_i(z_{i,s})^T \mathbf{a}, \sigma^2) \right\}, \quad (27)$$

$$\begin{aligned} \Rightarrow \frac{\partial \log p(y_i; \mathbf{a})}{\partial \mathbf{a}} &\approx \frac{1}{\sum_{s=1}^S p(y_i | z_{i,s}, \mathbf{a})} \sum_{s=1}^S \left\{ \frac{(y_i - \phi_i(z_{i,s})^T \mathbf{a}) \phi_i(z_{i,s})}{\sigma^2} \mathcal{N}(y_i; \phi_i(z_{i,s})^T \mathbf{a}, \sigma^2) \right\} \\ &\triangleq \mathbf{d}(y_i). \end{aligned} \quad (28)$$

Using (26), (28) and MCI we have

$$I_{\mathbf{y}}(\mathbf{a}) \approx \sum_{i=1}^N E [\mathbf{d}(y_i) \mathbf{d}(y_i)^T] \approx \frac{1}{S} \sum_{i=1}^N \sum_{s=1}^S \mathbf{d}(y_{i,s}) \mathbf{d}(y_{i,s})^T, \quad (29)$$

where  $y_{i,s}$  are generated from  $p(y_i; \mathbf{a})$ .

Hence for any unbiased estimator  $\hat{\mathbf{a}}$  we have

$$\text{var}(\hat{\mathbf{a}}) \geq \text{CRB}(\mathbf{a}) = \text{trace}(I_{\mathbf{y}}(\mathbf{a})^{-1}). \quad (30)$$

### C. An EM algorithm for ML Estimation

EM is a classical approach to approximating ML estimators in the presence of incomplete or unobserved data [31]. The observation vector  $\mathbf{y}$  in (8) depends on the unobserved location errors,  $\mathbf{z}$ . Instead of implementing (9), EM maximizes the following lower bound of the likelihood function at the current estimate  $\mathbf{a}^{(t-1)}$ :

$$Q(\mathbf{a}; \mathbf{a}^{(t-1)}) = E \left[ \log p(\mathbf{y}, \mathbf{z}; \mathbf{a}) | \mathbf{y}; \mathbf{a}^{(t-1)} \right]. \quad (31)$$

Under certain regularity conditions the EM algorithm is guaranteed to converge to a local maximum of the likelihood function (8) [20], [31]. Using (8) we can write

$$\log p(\mathbf{y}, \mathbf{z}; \mathbf{a}) = -\frac{1}{2\sigma^2} \|\mathbf{y} - \Phi(X, Z)\mathbf{a}\|^2 - \frac{1}{2\sigma_z^2} \|\mathbf{z}\|^2 + K, \quad (32)$$

where  $K$  is a constant terms not relevant to the optimization problem. Substituting (32) into (31) gives

$$\begin{aligned} Q(\mathbf{a}; \mathbf{a}^{(t-1)}) &= -\frac{1}{2\sigma^2} \left( \|\mathbf{y}\|^2 - 2\mathbf{y}^T E \left[ \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right] \mathbf{a} + \mathbf{a}^T E \left[ \Phi(X, Z)^T \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right] \mathbf{a} \right) \\ &\quad - \frac{1}{2\sigma_z^2} E \left[ \|\mathbf{z}\|^2 | \mathbf{y}; \mathbf{a}^{(t-1)} \right] + K. \end{aligned} \quad (33)$$

Since (33) is quadratic in  $\mathbf{a}$ , taking its derivative with respect to  $\mathbf{a}$  and equating it to 0 gives the following system of linear equations:

$$E \left[ \Phi(X, Z)^T \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right] \mathbf{a} = E \left[ \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right]^T \mathbf{y}. \quad (34)$$

Hence the EM estimate of  $\mathbf{a}$  at  $t$  is

$$\hat{\mathbf{a}}^t = E \left[ \Phi(X, Z)^T \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right]^{-1} E \left[ \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right]^T \mathbf{y}. \quad (35)$$

As discussed in [20], we can separate the expectations in (34) into univariate expectations:

$$E \left[ \Phi(X, Z)^T \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right] = \sum_{i=1}^N E \left[ \phi_i(z_i) \phi_i(z_i)^T | \mathbf{y}; \mathbf{a}^{(t-1)} \right], \quad (36)$$

$$E \left[ \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right]^T \mathbf{y} = \sum_{i=1}^N E \left[ \phi_i(z_i) | \mathbf{y}; \mathbf{a}^{(t-1)} \right] y_i. \quad (37)$$

The expectations are with respect to  $p(\mathbf{z} | \mathbf{y}; \mathbf{a}^{(t-1)}) = \prod_{i=1}^N p(z_i | y_i; \mathbf{a}^{(t-1)})$ . Hence each univariate expectation is calculated with respect to  $p(z_i | y_i; \mathbf{a}^{(t-1)}) = p(y_i | z_i; \mathbf{a}^{(t-1)}) p(z_i) / p(y_i; \mathbf{a}^{(t-1)})$  where  $p(y_i | z_i; \mathbf{a}^{(t-1)}) = \mathcal{N}(y_i; \phi_i(z_i)^T \mathbf{a}^{(t-1)}, \sigma^2)$ . Given  $\mathbf{a}^{(t-1)}$  we compute  $p(y_i; \mathbf{a}^{(t-1)})$  from (24). Each univariate expectation is of the form  $\int_{z_i} g(z_i) \mathcal{N}(y_i; \phi_i(z_i)^T \mathbf{a}^{(t-1)}, \sigma^2) \mathcal{N}(z_i; 0, \sigma_z^2) dz_i$ , which can be approximated by MCI. We have the following approximations to the expectations in (36) and (37):

$$E \left[ \Phi(X, Z)^T \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right] \approx \sum_{i=1}^N \frac{1}{Sp(y_i; \mathbf{a}^{(t-1)})} \sum_{s=1}^S \phi_i(z_{i,s}) \phi_i(z_{i,s})^T p(y_i | z_{i,s}; \mathbf{a}^{(t-1)}), \quad (38)$$

$$E \left[ \Phi(X, Z) | \mathbf{y}; \mathbf{a}^{(t-1)} \right]^T \mathbf{y} \approx \sum_{i=1}^N \frac{y_i}{Sp(y_i; \mathbf{a}^{(t-1)})} \sum_{s=1}^S \phi_i(z_{i,s}) p(y_i | z_{i,s}; \mathbf{a}^{(t-1)}). \quad (39)$$

Eqn. 35 is then employed to obtain the current estimate  $\mathbf{a}^t$ . The aforementioned steps are repeated until convergence, i.e.,  $\|\mathbf{a}^t - \mathbf{a}^{(t-1)}\| \leq \epsilon$  or a maximum number of iterations have been reached. Algorithm 2 summarizes the steps involved in the EM implementation of ML estimation.

#### IV. NUMERICAL SIMULATIONS

In this section we present simulation results on the performance of ILE. It is important to note that ILE is not an unbiased estimator, so we study its bias and variance apart from the MSE. We perform simulations with both one- and two-dimensional basis functions and present performance in terms of the MSE and computational time. We also present a detailed argument for the stability of ILE. We then present comparisons of the ILE with the CRB for unbiased estimators and the EM approximation of the ML estimator for one-dimensional basis functions highlighting the computational advantage of our algorithm. Note that Matlab was used for all the simulations and the parallel processing toolbox was used when multiple simulations were needed for a given set of parameters.

---

**Algorithm 2** Expectation Maximization for maximum likelihood estimation of signal coefficients under measurement and sampling location errors

---

**Require:**  $\mathbf{y}, X, \phi_k(\mathbf{x})^t \forall k = 1 \dots M, \epsilon$  (convergence tolerance),  $P$  (maximum number of iterations)

Use (14) to find  $\mathbf{a}_b$

$t \leftarrow 0$

$\hat{\mathbf{a}}^t = \mathbf{a}_b$

**repeat**

$t \leftarrow t + 1$

Compute  $p(y_i; \hat{\mathbf{a}}^{(t-1)})$  by (24)

Approximate  $E [\Phi(X, Z)^T \Phi(X, Z) | \mathbf{y}; \hat{\mathbf{a}}^{(t-1)}]$  by (38)

Approximate  $E [\Phi(X, Z) | \mathbf{y}; \hat{\mathbf{a}}^{(t-1)}]^T \mathbf{y}$  by (39)

Find  $\hat{\mathbf{a}}^t$  using (35)

**until**  $\|\hat{\mathbf{a}}^t - \hat{\mathbf{a}}^{(t-1)}\| \leq \epsilon$  or  $t \geq P$

**return**  $\hat{\mathbf{a}}^t$

---

### A. Performance of ILE

We implement numerical simulations with two classes of basis functions. Eqn. 14 presents the baseline *blind* linear estimator against which we compare the performance of ILE. In all the simulations we generate  $\mathbf{x}_i$ 's on a uniform grid. For one-dimensional signals we place  $N$  samples at an equidistant spacing of  $\Delta x = 2/N$  and  $x_i = (i - 0.5)\Delta x$  for  $i = 1, \dots, N$ . Similarly for two-dimensional simulations we have samples at  $[(i - 0.5)\Delta x, (j - 0.5)\Delta y]^T$  for  $i = 1, \dots, N, j = 1, \dots, N, \Delta x = \Delta y = 2/N$  and total number of samples =  $N^2$ .

We now describe the parameters used for simulation studies. As mentioned before, the measurement errors are assumed to be i.i.d Gaussian variables with known variance. Here we assume the variance to be a certain fraction of the maximum value of the function, i.e.  $\sigma = (\gamma/100) \max |f(\mathbf{x})|$ . Similarly we assume the variance of the position error to be a fraction of the spacing between the sampling points, i.e.  $\sigma_z = (p/100)\Delta x$ . Though the sample location error was assumed to have a Gaussian distribution, in simulations we used the following hard thresholding on the values of  $z_i$ 's:  $z_i = \text{sign}(z_i) \min\{|z_i|, \Delta x/2\}$ . This ensures that the sample location, say,  $x_i + z_i$  is lower than  $x_{i+1} + z_{i+1}$  and we have an ordered set of samples. Hence, in simulations, a higher  $\sigma_z$  would lead to a distribution close to a uniform distribution on  $\left[-\frac{\Delta x}{2}, \frac{\Delta x}{2}\right]$ .

For one-dimensional simulations we first have  $\psi_0(x) = 1$ ,  $\psi_k(x) = \sin(kx) \forall k = 1$  to 5 and  $\psi_k(x) = \cos((k-5)x) \forall k = 6$  to 10. We use Gram-Schmidt orthonormalization [8] to generate 11 orthonormal basis functions  $\phi_k(x)$ 's and call it the set of trigonometric basis functions. Similarly, we create an orthonormal set of 10 polynomial basis functions from  $\psi_k(x) = x^k \forall k = 0$  to 9. For each combination of the parameters,  $p$ ,  $\gamma$  and  $N$ , 1000 simulations were carried out in a parallel fashion using Matlab's `parfor`. For trigonometric basis functions  $\|\mathbf{a}\| = 2.917$ , for polynomial basis functions  $\|\mathbf{a}\| = 4.283$  and we chose  $\epsilon = 10^{-4}$ ,  $P = 200$ . We compute the mean and the standard deviation of the squared estimation error ( $\|\hat{\mathbf{a}} - \mathbf{a}\|^2$ ). We compare the squared estimation error with ILE, the *blind* estimator (14) and the following least square estimator with known sample location error:

$$\hat{\mathbf{a}}_{KLE} = (\Phi(X, Z)^T \Phi(X, Z))^{-1} \Phi(X, Z)^T \mathbf{y}. \quad (40)$$

Fig. 2 presents the performance of the ILE for trigonometric basis functions and Fig. 3 for polynomial basis functions. Fig. 2a and Fig. 3a show that as the sample location error,  $\sigma_z$ , is increased the ILE performs better compared to the *blind* estimator. The estimation error with ILE is comparable to BE for low values of  $p(\sigma_z)$ , but with increase in  $p$ , ILE outperforms BE. As expected, Fig. 2b and Fig. 3b show that the estimation error decreases with increase in  $N$ . Interestingly, the difference between estimation error with ILE and BE is higher for lower  $N$  and at higher  $N$  the difference between them is negligible. As expected, the estimate from (40) does not depend on sample location error ( $\sigma_z$ ).

We have the following relationship for the MSE:

$$\begin{aligned} E[\|\hat{\mathbf{a}} - \mathbf{a}\|^2] &= \sum_{k=1}^M E[(\hat{a}_k - a_k)^2] = \sum_{k=1}^M \left( E[(\hat{a}_k - E[\hat{a}_k])^2] + (E[\hat{a}_k] - a_k)^2 \right) \\ &= \sum_{k=1}^M (\text{var}(\hat{\mathbf{a}}_k) + \text{bias}_k^2) = \text{variance} + \text{bias}^2. \end{aligned} \quad (41)$$

We use the 1000 simulations for each parameter set and compute the bias and the variance of ILE. We define  $\Lambda = \text{bias}/\sqrt{\text{variance}}$ . Fig. 4 plots the variation of  $\Lambda$  for trigonometric basis functions with the variation in  $p$  for various combinations of  $\gamma$  and  $N$ . We observe that the contribution of bias to the overall estimation error increases with  $p$  and the effect is more pronounced at higher  $N$  and at lower  $\gamma$ . Recall that  $\sigma = (\gamma/100) \max |f(\mathbf{x})|$  and hence higher  $\gamma$  implies higher additive noise. We observe a similar variation of  $\Lambda$  with polynomial basis functions.

For polynomial basis functions, Fig. 5 shows the variation of the mean computational time for our algorithm with the variation in simulation parameters ( $p$ ,  $\gamma$  and  $N$ ). We observe that the mean computational time increases with  $p$  and are similar for different values of  $\gamma$ . Furthermore, we observe an

interesting trend with  $N = 50$ , where there is a sharp increase in computational time with  $p$ . We observe a similar trend with trigonometric basis functions. Note that the computational time is of the order of  $10^{-2}$  s. Typically the number of iterations for convergence is around 5 except for  $N = 50$  where at high  $p$  the number of iterations required is around 30.

Similarly, we create a set of two-dimensional orthonormal polynomial basis functions from the following set of basis functions:  $\psi = \{1, x_1, x_2, x_1x_2, x_1^2, x_2^2, x_1^2x_2, x_1x_2^2, x_1^2x_2^2, x_1^3, x_2^3\}$ . We have  $\|a\| = 2.398$ . From Fig. 6, we observe a trend similar to the one-dimensional case where the ILE outperforms BE with increase in  $p(\sigma_z)$  and the difference between ILE and BE decreases with  $N$ . The mean computational time is of the order of 1s and shows a trend similar to the one-dimensional signals. Though the computational time increased by two orders of magnitude, our method is still suitable for estimating a slowly varying higher dimensional field.

### B. Convergence and Stability of ILE

In this section we show the numerical stability of ILE, which ensures that the estimate does not diverge with iterations. We first assume that the  $N \times M$  matrix  $\Phi(X, 0)$  is of rank  $M$ . This is a reasonable assumption given we have orthonormal basis functions, we sample at distinct locations, and  $N \geq M$ . Note that if the aforementioned assumption does not hold, then the *blind* estimator (14) does not exist and ILE terminates at the initialization (Algorithm 1).

We denote the singular values of  $\Phi(X, 0)$  by  $\{s_k\}_{k=1}^M$  with  $s_1 \geq s_2 \geq \dots \geq s_M > 0$ . At the current iteration,  $t$ , we have  $C^t$  given by (18). We obtain  $\nabla f_b^t(\mathbf{x}_i)$  from the previous estimate  $\hat{\mathbf{a}}^{(t-1)}$  by  $\nabla f_b^t(\mathbf{x}_i) = \sum_{k=1}^M \hat{a}_k^{(t-1)} \nabla \phi_k(\mathbf{x}_i)$ . We define  $\beta_k = \max_{\mathbf{x} \in Q} \|\nabla \phi_k(\mathbf{x})\|$ . We have  $\|\beta\| = \sum_{k=1}^M \beta_k^2$  and  $\|\beta\| < \infty$  as the basis functions are continuous and differentiable and the domain  $Q$  is closed and bounded. Using Cauchy-Schwarz inequality we have:

$$\begin{aligned}
\|\nabla f_b^t(\mathbf{x})\| &= \sqrt{\sum_{j=1}^D \left( \sum_{k=1}^M \hat{a}_k^{(t-1)} \frac{\partial \phi_k(\mathbf{x})}{\partial x_j} \right)^2} \leq \sqrt{\sum_{j=1}^D \left( \sum_{k=1}^M (\hat{a}_k^{(t-1)})^2 \sum_{k=1}^M \left( \frac{\partial \phi_k(\mathbf{x})}{\partial x_j} \right)^2 \right)} \\
&= \sqrt{\sum_{j=1}^D \left( \|\hat{\mathbf{a}}^{(t-1)}\|^2 \sum_{k=1}^M \left( \frac{\partial \phi_k(\mathbf{x})}{\partial x_j} \right)^2 \right)} = \|\hat{\mathbf{a}}^{(t-1)}\| \sqrt{\sum_{j=1}^D \sum_{k=1}^M \left( \frac{\partial \phi_k(\mathbf{x})}{\partial x_j} \right)^2} \\
&= \|\hat{\mathbf{a}}^{(t-1)}\| \sqrt{\sum_{k=1}^M \sum_{j=1}^D \left( \frac{\partial \phi_k(\mathbf{x})}{\partial x_j} \right)^2} = \|\hat{\mathbf{a}}^{(t-1)}\| \sqrt{\sum_{k=1}^M \|\nabla \phi_k(\mathbf{x})\|^2} \\
&\leq \|\hat{\mathbf{a}}^{(t-1)}\| \|\beta\|.
\end{aligned} \tag{42}$$

Hence for  $i = 1, \dots, N$ ,  $\|\nabla f_b^t(\mathbf{x}_i)\| \leq \|\hat{\mathbf{a}}^{(t-1)}\| \|\boldsymbol{\beta}\|$ . According to (14), we initialize ILE with  $\hat{\mathbf{a}}^0 = \mathbf{a}_b$ , so we have:

$$\begin{aligned} \|\hat{\mathbf{a}}^0\| &= \|(\Phi(X, 0)^T \Phi(X, 0))^{-1} \Phi^T(X, 0) \mathbf{y}\| \\ &\leq \|(\Phi(X, 0)^T \Phi(X, 0))^{-1}\| \|\Phi^T(X, 0)\| \|\mathbf{y}\| = \frac{1}{s_M^2} s_1 \|\mathbf{y}\| < \infty. \end{aligned} \quad (43)$$

For the first iteration of ILE ( $t = 1$ ), every diagonal entry of  $C^1$  satisfies:

$$\begin{aligned} C_{i,i}^1 &= \frac{1}{\sigma^2 + \|\nabla f_b^1(\mathbf{x}_i)\|^2 \sigma_z^2} \\ \implies \frac{1}{\sigma^2 + \|\hat{\mathbf{a}}^0\| \|\boldsymbol{\beta}\| \sigma_z^2} &\leq C_{i,i}^1 \leq \frac{1}{\sigma^2}. \end{aligned} \quad (44)$$

Using (43) we get:

$$0 < \frac{s_M^2}{\sigma^2 s_M^2 + s_1 \|\mathbf{y}\| \|\boldsymbol{\beta}\| \sigma_z^2} \leq C_{i,i}^1 \leq \frac{1}{\sigma^2}. \quad (45)$$

Note that the above lower bound of  $C_{i,i}^1$  holds true for any finite norm initialization of ILE, i.e.,  $\|\hat{\mathbf{a}}^0\| < \infty$ . Since  $C^1$  is a diagonal matrix with all its diagonal entries greater than zero,  $C^1$  is positive definite, i.e.,  $C^1 \succ 0$ . We now make use of the following lemma [32]:

**Lemma.** *Let  $A$  be an  $m \times n$  matrix of full row rank, and let  $\mathcal{D}_+$  denote the set of positive definite diagonal  $n \times n$  matrices. Then,*

$$\sup_{D \in \mathcal{D}_+} \|(ADA^T)^{-1} AD\| = \max_{J \in \mathcal{J}(A)} \|A_J^{-T}\|,$$

where  $\mathcal{J}(A)$  is the collection of sets of column indices associated with nonsingular  $m \times m$  submatrices ( $A_J$ ) of  $A$ .

After the first iteration of ILE, we have:

$$\begin{aligned} \hat{\mathbf{a}}^1 &= (\Phi(X, 0)^T C^1 \Phi(X, 0))^{-1} \Phi(X, 0)^T C^1 \mathbf{y} \\ \implies \|\hat{\mathbf{a}}^1\| &\leq \|(\Phi(X, 0)^T C^1 \Phi(X, 0))^{-1} \Phi(X, 0)^T C^1\| \|\mathbf{y}\| \\ &\leq \max_{J \in \mathcal{J}(\Phi(X, 0))} \|\Phi(X, 0)_J^{-1}\| \|\mathbf{y}\| = G \|\mathbf{y}\| < \infty, \end{aligned} \quad (46)$$

where  $\mathcal{J}(\Phi(X, 0))$  is the collection of sets of row indices associated with nonsingular  $M \times M$  submatrices of  $\Phi(X, 0)$  and  $G = \max_{J \in \mathcal{J}(\Phi(X, 0))} \|\Phi(X, 0)_J^{-1}\|$  and is finite.

Since  $\|\hat{\mathbf{a}}^1\|$  is finite, from (42), (44), (46) and  $t = 2$ , we have  $\|\nabla f_b^2(\mathbf{x}_i)\| \leq G \|\mathbf{y}\| \|\boldsymbol{\beta}\|$ ,  $C^2 \succ 0$  and hence  $\|\hat{\mathbf{a}}^2\| \leq G \|\mathbf{y}\|$ . Extending the argument in (44) and (46) to future iterations, we can claim that

$C^t \succ 0$  and  $\|\hat{\mathbf{a}}^t\| \leq G\|\mathbf{y}\| \forall t$ . Hence, ILE is a stable estimator and the estimate is guaranteed to have a finite norm.

ILE may however converge to different  $\hat{\mathbf{a}}\mathbf{s}$  with different initialization. Numerically, we studied the variance in the convergence of ILE by random initialization of  $\mathbf{a}$  from a zero mean Gaussian distribution with standard deviation of 100. Fig. 7 shows the standard deviation of the converged  $\|\hat{\mathbf{a}}\|$  with 1000 different random initializations for different values of  $p$ ,  $\gamma$  and  $N = \{100, 200\}$ . We observe that the standard deviation is acceptable and is within 6% of  $\|\mathbf{a}\|$  even at high levels of noise ( $\gamma = 10, p = 100$ ).

### C. Comparison of ILE with CRB and EM algorithm

We perform simulations with the one-dimensional trigonometric basis functions as described before to compare the performance of ILE with the CRB and the EM approximation of the ML estimator. The CRB gives the minimum achievable variance of an unbiased estimator. On the other hand EM is a state of the art approach to implement ML estimation for problems with missing data [31].

Since EM involves approximation of integrals through MCI (Section III-C), computational times are high. As discussed in Section III-A, the number of computations and hence computational time may scale up exponentially with the dimension of the signal. Here we only present comparison with a one-dimensional signal and we further perform 250 simulations (as opposed to 1000) for each combination of the simulation parameters for both ILE and EM. Fig. 8 shows the variation of the moving average of the squared estimation error with the iteration number for seven randomly chosen simulation settings. In all the cases we observe that the moving average reaches close to its final value after around 250 simulations implying that 250 iterations are sufficient to quantify the mean performance of ILE.

Fig. 9 compares the squared estimation error with ILE, BE (14) and EM (Section III-C) with the variation in  $p$ . For all the simulations, we chose trigonometric basis functions as described in IV-A and  $\|\mathbf{a}\| = 5.823$ . We further compare the squared estimation errors with the CRB (Section III-B) which is the minimum achievable variance of any unbiased estimator. We observe that for low additive noise ( $\gamma = 1$ ), ILE outperforms EM while at higher noise ( $\gamma = 5$ ), the EM algorithm gives a lower estimation error. Interestingly, Fig. 9a shows that as  $p$  increases, EM performs worse when compared to BE. The EM algorithm uses the marginal and conditional probabilities of the observations  $y_i\mathbf{s}$  to iteratively calculate an estimate of  $\mathbf{a}$ . The conditional probability  $p(y_i|z_i; \mathbf{a}^{(t-1)}) = \mathcal{N}(y_i; \phi_i(z_i)^T \mathbf{a}^{(t-1)}, \sigma^2)$  is a function of the noise level ( $\sigma = (\gamma/100) \max |f(\mathbf{x})|$ ). The normally distributed conditional probability  $p(y_i|z_i; \mathbf{a}^{(t-1)})$  gets narrower and eventually assumes a Kronecker  $\delta$  function as  $\gamma \rightarrow 0$ . This causes higher numerical errors when conditional expectations are calculated (Section III-C). Fig. 10 shows the marginal distribution

of  $y_i$  calculated through MCI (Eqn. 24) with different values of  $\gamma$  for one-dimensional trigonometric basis functions with  $\|\mathbf{a}\| = 3.4528$  and  $x_i = 0$ . Since EM involves numerical evaluation of expectations that involve  $\mathcal{N}(y_i; \phi_i(z_i)^T \mathbf{a}^{(t-1)}, \sigma^2)$ , the numerical error in the calculation of these expectations increases with decreasing  $\gamma$ , worsening the performance of EM with decreasing noise. On the other hand, ILE does not suffer from such issues making it a more suitable choice for signal estimation under low noise levels or high SNR.

The mean computational time for EM is of the order of  $10^3$  s while for ILE is of the order of  $10^{-2}$  s. Fig. 11 shows the variation in the squared estimation error with ILE and EM for  $N = 50$  and  $N = 100$ . It also shows the comparison with CRB. We observe that for a lower number of samples ILE outperforms EM with increase in  $p$  and for a higher number of samples their performance is comparable. Note that here the unknown vector ( $\mathbf{a}$ ) is 11 dimensional, so  $N = 50$  refers to a reasonable level of oversampling while  $N = 100$  implies high oversampling. Hence for low additive noise and a reasonable level of oversampling, ILE clearly outperforms EM in both the MSE and the computational time. Even for higher additive noise and number of samples, ILE is comparable to EM with a significant improvement in the computational time making ILE an appealing candidate for real-time embedded applications like mobile sensor networks for environmental monitoring.

## V. CONCLUSION

In this paper we addressed the problem of estimating a parametric signal from samples with location errors and independent additive noise. Such problems are encountered as time jitter in ADCs where tolerating a higher jitter leads to a lower-power hardware. As interest in the use of mobile sensor platforms for environmental monitoring increases, addressing the effect of localization errors on the estimation of spatio-temporal fields will become increasingly important. We formulate the signal estimation problem as of finding deterministic parameters by representing the signal in the span of a finite basis. The parameter estimation is then formulated as finding the maximum likelihood estimator. We derive an approximation of the likelihood function through linearizing the observation model and then employ an iterative strategy to develop a computationally efficient iterative linear estimator. We highlight practical issues in computing the numerical approximation of the complete likelihood function and show the suitability of Monte Carlo integration over Gauss-Hermite quadrature for our problem. Using MCI we derive a numerical scheme for finding the Cramér-Rao lower bound and an Expectation Maximization algorithm to implement the maximum likelihood estimation. We present simulations to show the suitability of our proposed algorithm in improving the estimation error at low computational cost and also present a detailed argument for

the stability of ILE. When compared with the estimator with no information (BE), ILE gives a lower estimation error and its performance increases with the variance of the sample location error ( $p$ ) but decreases with the number of sampling points ( $N$ ). We numerically demonstrate that the contribution of the bias in the overall estimation error increases with  $p$  and  $N$  but decreases with the variance of the additive Gaussian noise ( $\gamma$ ). The mean computation time for ILE for the one-dimensional signal is of the order of  $10^{-2}$  s which increases to 1 s for the two-dimensional signal. For a one-dimensional signal, we show that for low additive noise and a reasonable level of oversampling, ILE outperforms EM while at higher additive noise and higher number of samples EM algorithm gives a lower estimation error. We further show that the performance of EM worsens with decreasing additive noise due to high numerical errors associated with computing integrals that involve narrow Gaussian distributions, while ILE does not suffer from such issues. Further, ILE is around five orders of magnitude faster when compared to EM making it an appealing candidate for real-time embedded applications.

In our future work, we would like to use the posterior estimate of the location errors,  $\mathbf{z}_i$ 's, through our linearized observation model to explore further improvement of the estimation error. We would also like to study how the algorithm presented here can be incorporated in the frameworks for adaptive path planning of mobile sensor nodes for the application of spatio-temporal field mapping. Further, we aim to conduct experimental studies with a set of mobile platforms equipped with a sensor (for example light sensor) to understand the applicability and suitability of our algorithm. We hope that such computationally efficient estimators can become an integral component of mobile WSNs, allowing real time improvement of field estimates, improving path planning and reducing sampling at less important locations.

#### ACKNOWLEDGMENT

The work of S. Kumar was supported by a research grant from Ferrovial Servicios, Madrid, Spain. The authors would also like to thank members of the Field Intelligence Laboratory, MIT for useful discussions and suggestions on the manuscript.

#### REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.
- [2] D. Culler, D. Estrin, and M. Srivastava, "Guest editors' introduction: Overview of sensor networks," *Computers*, vol. 37, no. 8, pp. 41–49, Aug. 2004.
- [3] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: Participatory urban sensing using a network of handheld air quality monitors," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys'09)*, New York, NY, USA, 2009, pp. 349–350.

- [4] A. Deshpande and S. E. Sarma, "Error tolerant arrangements of sensors for sampling fields," in *American Control Conference, 2008*, Jun. 2008, pp. 2401–2408.
- [5] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, Jun. 2008.
- [6] Y. M. Lu and M. Vetterli, "Spatial super-resolution of a diffusion field by temporal oversampling in sensor networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009*, Apr. 2009, pp. 2249–2252.
- [7] A. Deshpande, S. E. Sarma, and V. K. Goyal, "Generalized regular sampling of trigonometric polynomials and optimal sensor arrangement," *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 379–382, Apr. 2010.
- [8] S. Kumar, A. Deshpande, and S. E. Sarma, "Stable arrangements of mobile sensors for sampling physical fields," in *American Control Conference, 2012*, Jun. 2012.
- [9] L. L. Arnold and P. A. Zandbergen, "Positional accuracy of the wide area augmentation system in consumer-grade GPS units," *Computers & Geosciences*, vol. 37, no. 7, pp. 883–892, Jul. 2011.
- [10] P. A. Zandbergen, "Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning," *Transactions in GIS*, vol. 13, pp. 5–25, Jun. 2009.
- [11] L. S. Monteiro, T. Moore, and C. Hill, "What is the accuracy of DGPS?" *The Journal of Navigation*, vol. 58, no. 02, pp. 207–225, 2005.
- [12] D. Fox, W. Burgard, H. Kruppa, and S. Thrun, "A probabilistic approach to collaborative multi-robot localization," *Autonomous Robots*, vol. 8, no. 3, pp. 325–344, 2000.
- [13] D. Fox, W. Burgard, and S. Thrun, "Markov localization for mobile robots in dynamic environments," *Journal of Artificial Intelligence Research*, vol. 11, pp. 391–427, 1999.
- [14] J. E. Guivant and E. M. Nebot, "Optimization of the simultaneous localization and map-building algorithm for real-time implementation," *IEEE Trans. Robot. Automat.*, vol. 17, no. 3, pp. 242–257, Jun. 2001.
- [15] J. J. Leonard and H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons," *IEEE Trans. Robot. Automat.*, vol. 7, no. 3, pp. 376–382, Jun. 1991.
- [16] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo localization for mobile robots," *Artificial Intelligence*, vol. 128, no. 12, pp. 99–141, May 2001.
- [17] A. Balakrishnan, "On the problem of time jitter in sampling," *IRE Trans. Inf. Theory*, vol. 8, no. 3, pp. 226–236, Apr. 1962.
- [18] B. Liu and T. Stanley, "Error bounds for jittered sampling," *IEEE Trans. Autom. Control*, vol. 10, no. 4, pp. 449–454, Oct. 1965.
- [19] A. Nordio, C. F. Chiasserini, and E. Viterbo, "Signal reconstruction errors in jittered sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4711–4718, 2009.
- [20] D. S. Weller and V. K. Goyal, "On the estimation of nonrandom signal coefficients from jittered samples," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 587–597, Feb. 2011.
- [21] —, "Bayesian post-processing methods for jitter mitigation in sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2112–2123, May 2011.
- [22] E. Masry, "Analysis of signal reconstruction with jittered sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 27–34, Jan. 2011.
- [23] D. S. Weller and V. K. Goyal, "Jitter compensation in sampling via polynomial least squares estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009*, Apr. 2009, pp. 3341–3344.

- [24] W. H. Foy, "Position-location solutions by Taylor-series estimation," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-12, no. 2, pp. 187–194, 1976.
- [25] D. S. Weller, "Mitigating timing noise in adcs through digital post-processing," Master's thesis, Electrical Eng. Comput. Sci. Dept., Massachusetts Inst. of Technology, Jun. 2008.
- [26] G. H. Golub and J. H. Welsch, "Calculation of Gauss quadrature rules," Stanford University, Stanford, CA, USA, Tech. Rep., 1967.
- [27] T. Gerstner and M. Griebel, "Numerical integration using sparse grids," *Numerical Algorithms*, vol. 18, no. 3, pp. 209–232, 1998.
- [28] X. Wang and K. Fang, "The effective dimension and quasi-Monte Carlo integration," *Journal of Complexity*, vol. 19, no. 2, pp. 101–124, Apr. 2003.
- [29] J. Burkardt. (2010, Feb.) [http://people.sc.fsu.edu/~jburkardt/m\\_src/gen\\_hermite\\_rule/gen\\_hermite\\_rule.html](http://people.sc.fsu.edu/~jburkardt/m_src/gen_hermite_rule/gen_hermite_rule.html).
- [30] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.
- [31] F. Dellaert, "The expectation maximization algorithm," Georgia Institute of Technology, Technical Report, 2002.
- [32] A. Forsgren and G. Sporre, "On weighted linear least-squares problems related to interior methods for convex quadratic programming," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 42–56, 2001.



**Sumeet Kumar** (S'11) received the B.Tech. degree in mechanical engineering from the Indian Institute of Technology Kanpur in 07 where he received the Academic Excellence Awards in 05 and 06 for top academic performance among all the undergraduates. He received the M.S. in mechanical engineering from the Massachusetts Institute of Technology in 09 where he received the Pappalardo Fellowship. His M.S. research was in the area of microfluidic design and fabrication and thermal modeling.

Currently he is pursuing Ph.D. in mechanical engineering at the Massachusetts Institute of Technology. His current research interests include sensor networks, signal processing, image processing and machine learning.



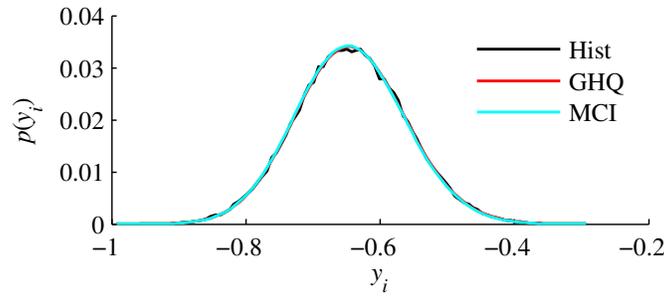
**Vivek K Goyal** (S'92-M'98-SM'03) received the B.S. degree in mathematics and the B.S.E. degree in electrical engineering from the University of Iowa, Iowa City, where he received the John Briggs Memorial Award for the top undergraduate across all colleges. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, where he received the Eliahu Jury Award for outstanding achievement in systems, communications, control, or signal processing.

He was a Member of Technical Staff in the Mathematics of Communications Research Department of Bell Laboratories, Lucent Technologies, 1998–2001; and a Senior Research Engineer for Digital Fountain, Inc., 2001–2003. He is currently Esther and Harold E. Edgerton Associate Professor of electrical engineering at the Massachusetts Institute of Technology. His research interests include computational imaging, sampling, quantization, and source coding theory.

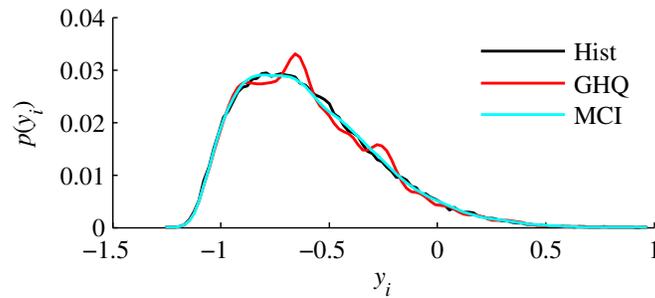
Dr. Goyal is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, Eta Kappa Nu and SIAM. He was awarded the 2002 IEEE Signal Processing Society Magazine Award and an NSF CAREER Award. He served on the IEEE Signal Processing Society's Image and Multiple Dimensional Signal Processing Technical Committee. He is a Technical Program Committee Co-chair of IEEE ICIP 2016 and a permanent Conference Co-chair of the SPIE Wavelets and Sparsity conference series.



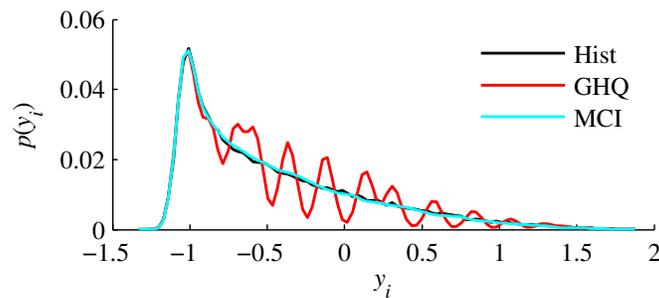
**Sanjay E. Sarma** Sanjay Sarma is a Professor of Mechanical Engineering at MIT. He founded the Auto-ID Center at MIT and developed many of the key technologies behind the EPC suite of RFID standards now used worldwide. He was also the the founder and CTO of OATSystems, which was acquired by Checkpoint Systems (NYSE: CKP) in 2008. He several on the boards of GS1, EPCglobal and several startup companies. Dr. Sarma received his Bachelors from the Indian Institute of Technology, his Masters from Carnegie Mellon University and his PhD from the University of California at Berkeley. Sarma also worked at Schlumberger Oilfield Services in Aberdeen, UK, and at the Lawrence Berkeley Laboratories in Berkeley, California. His current research interests are street scanning, sensing, RFID, logistics and manufacturing.



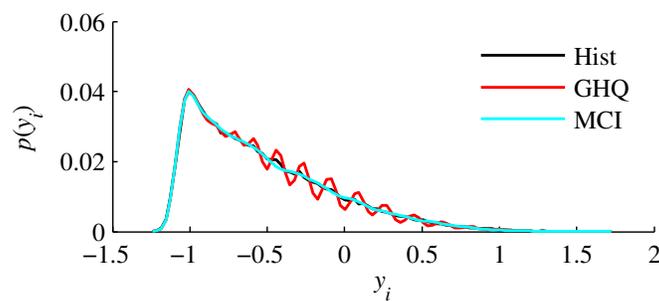
(a)



(b)

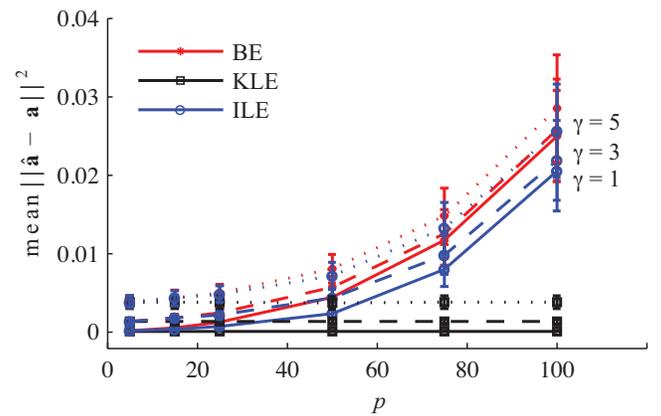


(c)

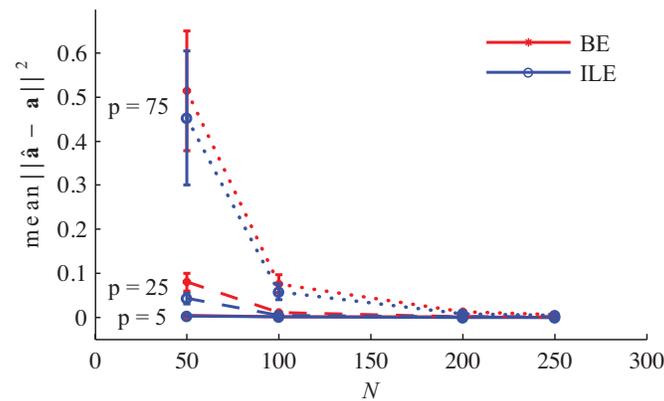


(d)

Fig. 1. Comparison of the histogram of the sample  $y_i$  (Hist) with numerical approximation of the pdf through Gauss Hermite quadrature (GHQ) and Monte Carlo integration (MCI);  $\sigma = 0.053$ , (a)  $k = 1, \sigma_z = 0.01$ ; (b)  $k = 5, \sigma_z = 0.01$ ; (c)  $k = 10, \sigma_z = 0.01$ ; (d)  $k = 1, \sigma_z = 0.075$ .

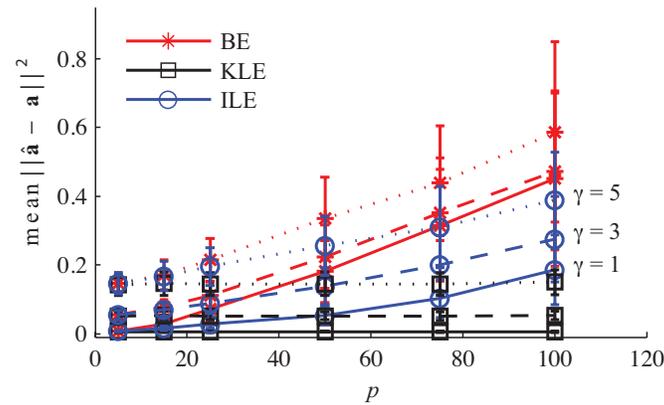


(a)

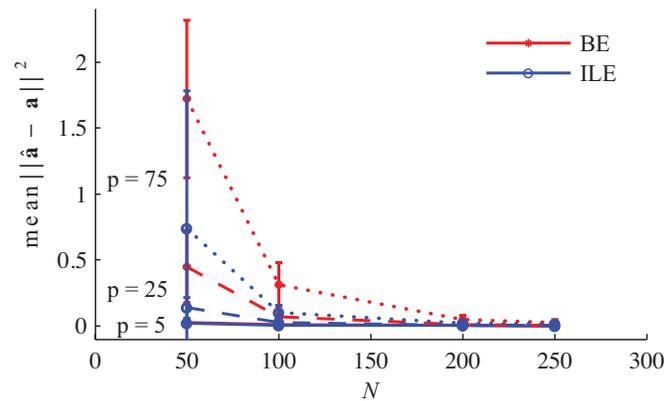


(b)

Fig. 2. Comparison of the mean of the squared estimation error  $\|\hat{\mathbf{a}} - \mathbf{a}\|^2$  with ILE, BE and KLE for trigonometric basis functions. The error bars indicate 1/2 standard deviations. (a) Comparison of the performance with the variation in  $p$ . Solid line corresponds to  $\gamma = 1$ , dashed line  $\gamma = 3$ , dotted line  $\gamma = 5$  and for all the simulations  $N = 200$ . (b) Comparison of the performance of ILE against BE with the variation in  $N$ . The percentage sample location error ( $p$ ) is indicated on the left of each set of plots and  $\gamma = 1$ .

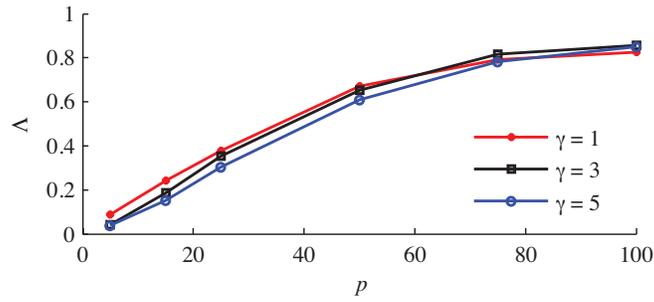


(a)

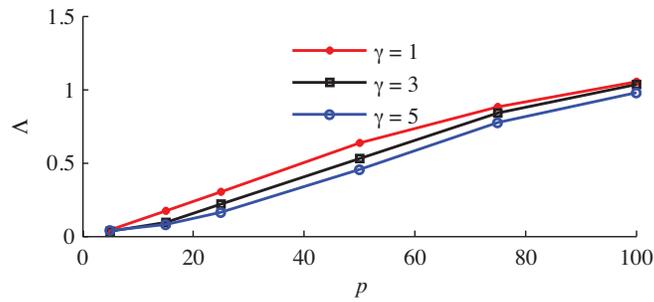


(b)

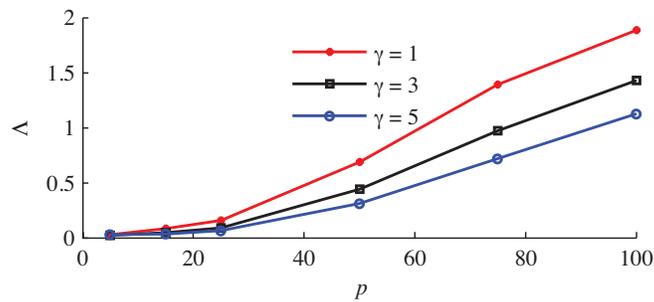
Fig. 3. Comparison of the mean of the squared estimation error  $\|\hat{\mathbf{a}} - \mathbf{a}\|^2$  with ILE, BE and KLE for polynomial basis functions. The error bars indicate 1/2 standard deviations. (a) Comparison of the performance with the variation in  $p$ . Solid line corresponds to  $\gamma = 1$ , dashed line  $\gamma = 3$ , dotted line  $\gamma = 5$  and for all the simulations  $N = 100$ . (b) Comparison of the performance of ILE against BE with the variation in  $N$ . The percentage sample location error ( $p$ ) is indicated on the left of each set of plots and  $\gamma = 1$ .



(a)



(b)



(c)

Fig. 4. Comparison of  $\Lambda$  with the variation in  $p$  for (a)  $N = 50$ ; (b)  $N = 100$ ; (c)  $N = 250$ .

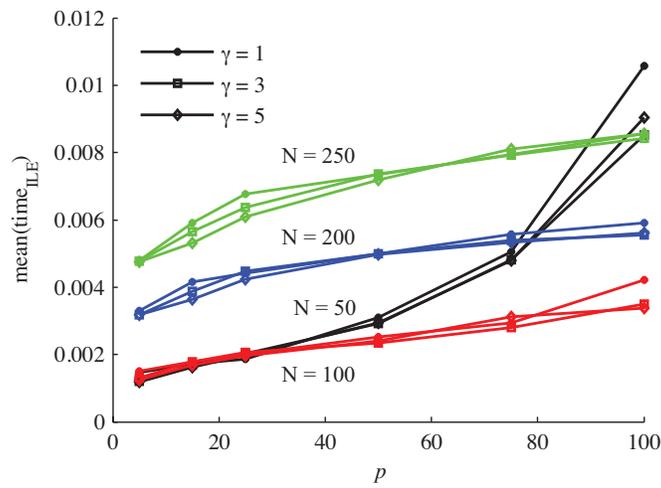
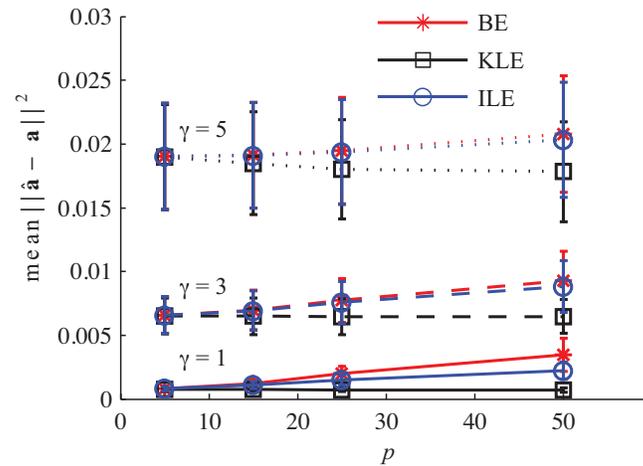
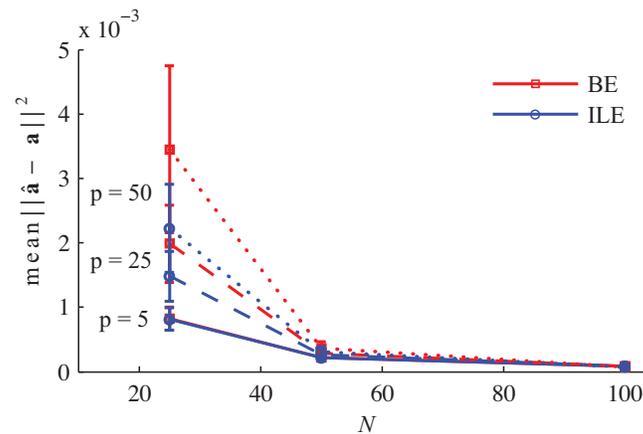


Fig. 5. Variation of the mean of the computational time for ILE with % sample location error ( $p$ ) for different % additive noise ( $\gamma$ ) and the number of sampling points ( $N$ ).

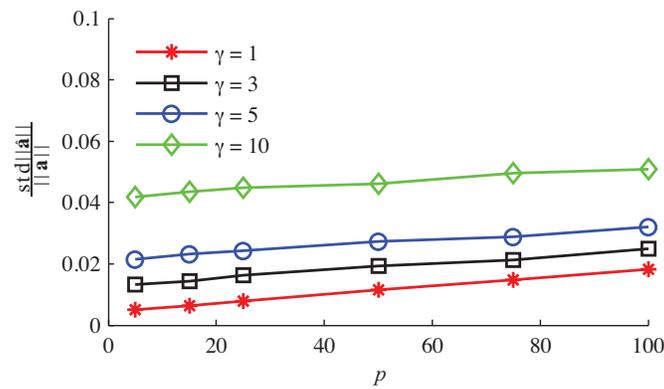


(a)

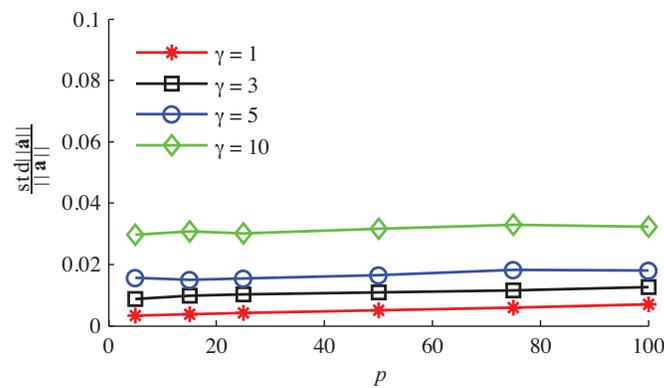


(b)

Fig. 6. Comparison of the mean of the squared estimation error  $\|\hat{\mathbf{a}} - \mathbf{a}\|^2$  with ILE, BE and KLE for two-dimensional polynomial basis functions. The error bars indicate 1/2 standard deviations. (a) Comparison of the performance with the variation in  $p$ . Solid line corresponds to  $\gamma = 1$ , dashed line  $\gamma = 3$ , dotted line  $\gamma = 5$  and for all the simulations  $N = 25$ . (b) Comparison of the performance of ILE against BE with the variation in  $N$ . The percentage sample location error ( $p$ ) is indicated on the left of each set of plots and  $\gamma = 1$ .



(a)



(b)

Fig. 7. Variation of the normalized standard deviation of the norm of the converged parameter estimates ( $\text{std}\|\hat{\mathbf{a}}\|/\|\mathbf{a}\|$ ) with  $p$  for one-dimensional polynomial basis functions and for (a)  $N = 100$  and (b)  $N = 200$ .

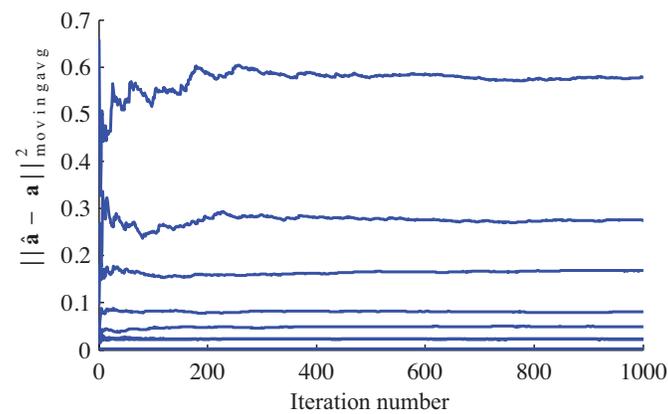
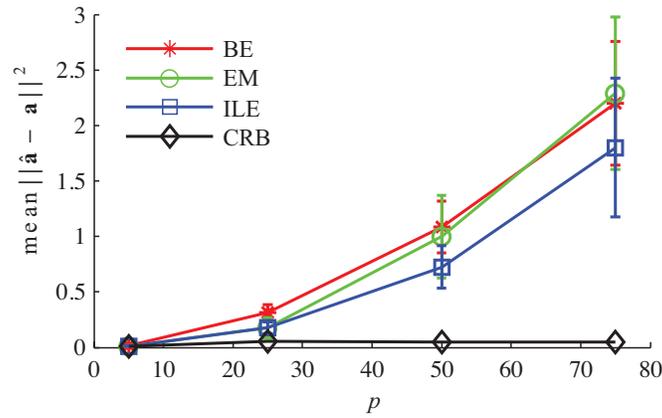
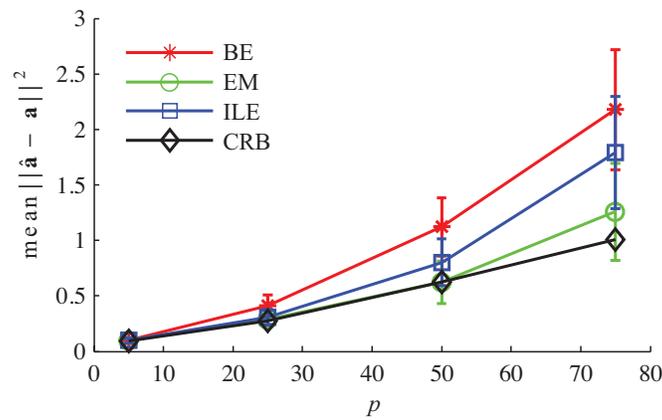


Fig. 8. Variation of the moving average of the squared estimation error with the iteration number for seven randomly chosen simulation settings with the one-dimensional polynomial basis functions.



(a)



(b)

Fig. 9. Comparison of the mean of the squared estimation error  $\|\hat{\mathbf{a}} - \mathbf{a}\|^2$  with ILE, BE and EM with the variation in  $p$ . The error bars indicate  $1/2$  standard deviations. The approximated CRB is presented by the black line. For all the simulations  $N = 50$  and (a)  $\gamma = 1$ ; (b)  $\gamma = 5$ .

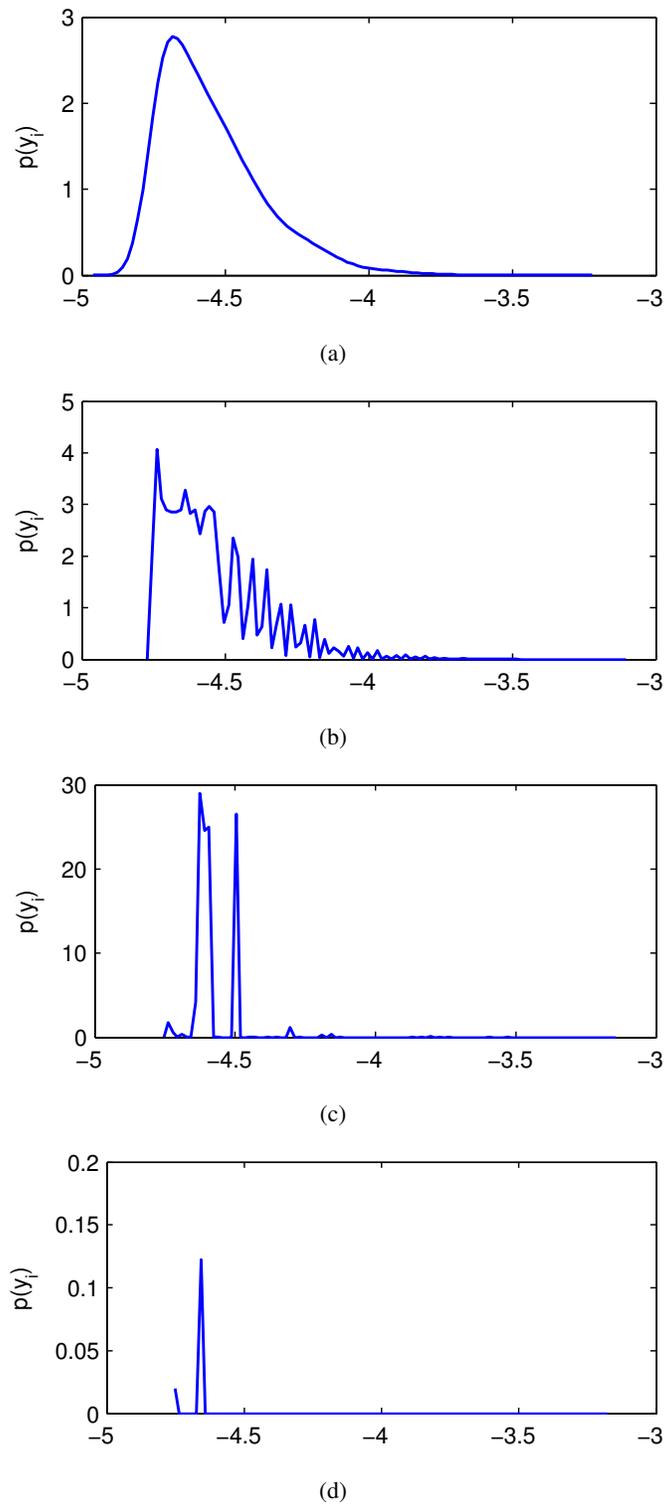
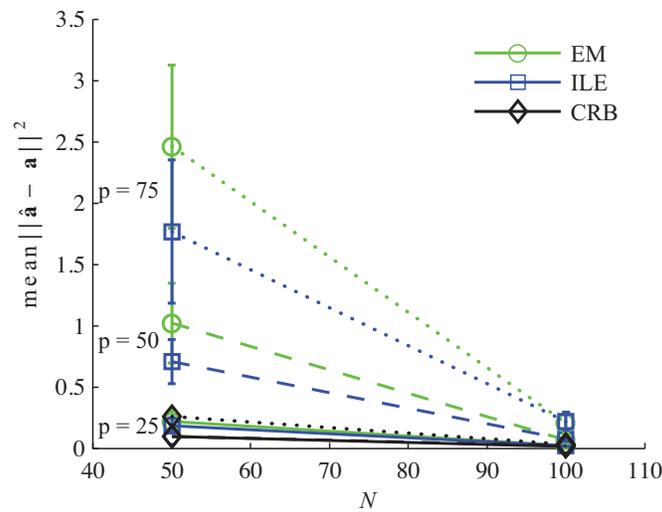


Fig. 10. Marginal probability distribution of  $y_i$  for one-dimensional trigonometric basis functions with  $\|\mathbf{a}\| = 3.4528$  and  $x_i = 0$ . (a)  $\gamma = 1$ , (b)  $\gamma = 0.1$ , (c)  $\gamma = 0.01$  and (d)  $\gamma = 0.001$



(a)

Fig. 11. Comparison of the mean of the squared estimation error  $\|\hat{\mathbf{a}} - \mathbf{a}\|^2$  with ILE and EM with the variation in  $N$ . The error bars indicate 1/2 standard deviations. The approximated CRB is presented by the black line. For all the simulations  $\gamma = 1$ .