A SEQUENTIAL BAYESIAN INFERENCE FRAMEWORK FOR BLIND FREQUENCY OFFSET ESTIMATION

Theodoros Tsiligkaridis, Keith W. Forsythe

MIT Lincoln Laboratory
Lexington, MA 02421 USA
{ttsili,forsythe}@ll.mit.edu

ABSTRACT

Precise estimation of synchronization parameters is essential for reliable data detection in digital communications and phase errors can result in significant performance degradation. The literature on estimation of synchronization parameters, including the carrier frequency offset, are based on approximations or heuristics because the optimal estimation problem is analytically intractable for most cases of interest. We develop an online Bayesian inference procedure for blind estimation of the frequency offset, for arbitrary signal constellations. Our unified approach is built on a sequential inference procedure that leverages a novel result on conjugacy of the von Mises and Gaussian distributions. This conjugacy allows for an easily computable, closed form parametric expression for the posterior distribution of the parameters given the streaming data, in which hyperparameters are recursively updated, making the optimal sequential estimation problem mathematically tractable. Our algorithm is computationally efficient and can be implemented in real-time with very low memory requirements. Numerical experiments are also provided and show that our methods outperform approximate sequential maximum-likelihood carrier frequency offset estimators.

***Index Terms*—** Sequential Estimation, Blind Synchronization, Frequency Offset Estimation, Bayesian Inference, Phase tracking.

1. INTRODUCTION

Synchronization plays a critical role in attaining reliable digital transmission through wireless channels. Timing and frequency offsets need to be estimated well in order to align the received signal appropriately at the receiver before data detection. Optimal estimators for synchronization parameters do not exist and as a result, most of the existing literature uses approximate maximum-likelihood techniques and heuristics [1, 2]. Even in the simpler setting, where the amplitude is known and constant, optimal estimation of the frequency offset is not known in closed-form and only approximate maximum-likelihood (ML) and maximum a posterior (MAP) approaches are tractable [3].

In digital communications, there is often a mismatch between the local oscillator of the transmitter and the receiver. This translates to a carrier frequency error when downconverting at the receiver, which effectively rotates the signal constellation from sample-to-sample. For small carrier frequency errors, Δf, the constellation rotates slowly and tracking its rotation rate is required for successful data detection. Popular methods for carrier synchronization include the use of a phase-lock loop (PLL), which often requires hardware implementation or waveform-level block-based processing in software, both of which can be expensive. Furthermore, the PLL approach often requires pilot data for higher order modulations. A synchronization approach based on model-based sequential Monte Carlo (SMC) techniques was proposed in [4] to estimate the timing offset and the data, but is heavily dependent on known state-space dynamic models governing the evolution of the parameters, several approximations are made to make the SMC algorithm tractable, and is computationally expensive as the complexity of the SMC grows exponentially fast [4].

Phase synchronization often relies on known features of the signal. This may involve training data [5], modulation [6], or higher-order moments [7]. In contrast, our approach only assumes baud synchronization, is not data-aided (i.e., blind), works for arbitrary signal modulations, operates at the sample-level, is simple to implement, and has very low latency and storage requirement as it proceeds on-the-fly in an online fashion. In this paper, we propose a novel sequential Bayesian inference framework to estimate the unknown carrier frequency offset and the parameters of the signal modulation.

We leverage ideas from the sequential inference approach of [8] to derive a sequential Bayesian algorithm to estimate the frequency offset and the constellation parameters given streaming non-i.i.d. received samples. This is a challenging machine learning problem primarily because of the temporal dynamics of the data generation process. Our method exploits the temporal relation from sample-to-sample and uses the online clustering and parameter estimation framework developed in [8]. Solid empirical performance is observed for slow enough rotation rates in our experiments.

2. SEQUENTIAL BAYESIAN INFERENCE FRAMEWORK

Here, we review the sequential Bayesian inference framework of [8] for online clustering and parameter estimation. Define the unknown parameters $\theta = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K, \delta)$, where $\delta \in [0, 2\pi)$ is the unknown rotation offset and $\boldsymbol{\mu}_h$ is the mean of each class (i.e., cluster center). Let the observations be given by $\mathbf{y}_i \in \mathbb{R}^d$, and γ_i to denote the class label of the ith observation (a latent variable). We define the available information at time i as $\mathbf{y}^{(i)} = \{\mathbf{y}_1, \dots, \mathbf{y}_i\}$ and $\gamma^{(i-1)} = \{\gamma_1, \dots, \gamma_{i-1}\}$. The algorithm proceeds as follows. For $i = 1, 2, \dots,$

This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract #FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

1. Choose best class label for $\mathbf{y}_i$:

$$\gamma_i \sim \{q_h^{(i)}\} = \left\{ \frac{L_{i,h}(\mathbf{y}_i)\pi_{i,h}}{\sum_{h'} L_{i,h'}(\mathbf{y}_i)\pi_{i,h'}} \right\}. \tag{1}$$

2. Update the posterior distribution using $\mathbf{y}_i, \gamma_i$:

$$\pi(\theta_{\gamma_i}|\mathbf{y}^{(i)}, \gamma^{(i)}) \propto f(\mathbf{y}_i|\theta_{\gamma_i})\pi(\theta_{\gamma_i}|\mathbf{y}^{(i-1)}, \gamma^{(i-1)}). \tag{2}$$

where θ_h are the parameters of class h, $f(\mathbf{y}_i|\theta_h)$ is the observation density conditioned on class h. The conditional likelihood $P(\mathbf{y}_i|\gamma_i = h, \mathbf{y}^{(i-1)}, \gamma^{(i-1)})$ is denoted by $L_{i,h}(\mathbf{y}_i)$ and $\pi_{i,h}$ denotes the class priors. The algorithm sequentially allocates observations $\mathbf{y}_i$ to classes by sampling the conditional posterior probability distribution $\{q_h^{(i)}\}$.

3. PROBLEM FORMULATION AND PROBABILISTIC MODEL

Consider a digital communication system where symbols from a discrete unknown alphabet, $x_m \in \mathcal{A} \subset \mathbb{C}$ are transmitted through a channel. The received signal at the front-end of the receiver consisting of a matched filter can be modeled as:

$$y(t) = e^{j2\pi\Delta f t} \sum_m x_m g(t - mT) + w(t)$$

where T is the symbol period, $g(\cdot)$ is the raised-cosine pulse waveform, and $w(t)$ is additive white Gaussian noise (AWGN) with power spectral density $N_0/2$. Here, Δf is the carrier frequency error. Sampling the output of the matched filter at a rate $1/T$, we obtain the discrete-time signal:

$$y_k = e^{j2\pi\Delta f kT} x_k + w_k \tag{3}$$

where $y_k = y(kT), w_k = w(kT)$, and $k = 0, 1, \ldots$ is the discrete-time index. Note the phase rotation offset $\delta = 2\pi\Delta f T$ is proportional to the carrier frequency error Δf.

We consider a Bayesian framework for estimating the unknown mean for each class and unknown phase rotation offset δ. This framework can be extended to include unknown covariances as well (corresponding to unknown SNR), but is not considered here for simplicity (see [8]).

The observation model (3) in vector form is given by:

$$\mathbf{y}_i = \mathbf{R}(\theta_i)\mathbf{x}_i + \mathbf{w}_i$$
$$\theta_{i+1} = \theta_i + \delta$$

where $\mathbf{y}_i \overset{\text{def}}{=} [\mathrm{Re}(y_i), \mathrm{Im}(y_i)]^T$, $\mathbf{x}_i \overset{\text{def}}{=} [\mathrm{Re}(x_i), \mathrm{Im}(x_i)]^T$ are symbols from a constellation, $\mathbf{w}_i$ is additive Gaussian noise with covariance $\sigma^2\mathbf{I}_2$. [1] Here, $\mathbf{R}(\theta)$ is the rotation matrix

$$\mathbf{R}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Without loss of generality, we assume $\theta_1 = 0$. The unknown parameters in the model are the constellation symbols $\mathbf{x}$, and the phase offset δ. Let $K = |\mathcal{A}|$ denote the size of the symbol alphabet.

The probabilistic model for the unknown parameters is given as:

$$\begin{aligned} \mathbf{y}|\boldsymbol{\mu}, \delta &\sim \mathcal{N}(\cdot|\mathbf{R}(\delta)\boldsymbol{\mu}, \sigma^2\mathbf{I}) \\ \mu &\sim \mathcal{N}(\cdot|\boldsymbol{\mu}_0, \sigma_0^2\mathbf{I}) \\ \delta &\sim \mathcal{V}(\cdot|\delta_0, \kappa_0) \end{aligned} \tag{4}$$

[1]The data dimensions is $d = 2$.

where $\mathcal{N}(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathcal{V}(\cdot|\delta_0, \kappa_0)$ denotes the von Mises distribution with direction parameter δ_0 and concentration parameter κ_0. The parameters $\theta = (\boldsymbol{\mu}, \delta) \in \mathbb{R}^d \times [-\pi, \pi)$ follow a normal-von Mises joint distribution. The corresponding graphical model is shown in Fig. 1.

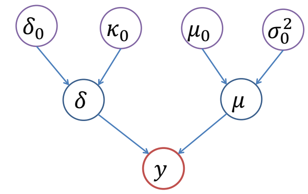

Fig. 1. Graphical model for simultaneous phase tracking and parameter estimation problem.

For concreteness, let us write the distribution in (4):

$$f(\mathbf{y}_i|\theta) = p(\mathbf{y}_i|\boldsymbol{\mu}, \delta) = \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}_i - \mathbf{R}(\delta)\boldsymbol{\mu}\|_2^2}$$
$$p(\boldsymbol{\mu}) = \frac{1}{(2\pi\sigma_0^2)^{d/2}} e^{-\frac{1}{2\sigma_0^2}\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|_2^2}$$
$$p(\delta) = \frac{1}{2\pi I_0(\kappa_0)} e^{\kappa_0 \cos(\delta - \delta_0)}$$

where $I_0(\cdot)$ is the modified Bessel function of order 0.

Due to the conjugacy of the distributions, the posterior distribution $\pi(\theta_h|\mathbf{y}^{(i-1)}, \gamma^{(i-1)})$ always has the form:

$$\begin{aligned} \pi(\theta_h|\mathbf{y}^{(i-1)}, \gamma^{(i-1)}) = \mathcal{N}(&\boldsymbol{\mu}_h|\boldsymbol{\mu}_h^{(i-1)}, (\sigma_h^{(i-1)})^2\mathbf{I}) \\ &\times \mathcal{V}(\delta|\delta^{(i-1)}, \kappa^{(i-1)}) \end{aligned} \tag{5}$$

where $\boldsymbol{\mu}_h^{(i-1)}, \sigma_h^{(i-1)}, \delta^{(i-1)}, \kappa^{(i-1)}$ are hyperparameters that can be recursively computed as new samples come in. This factorization is proven analytically in Section 4.

We will show in Section 4 that the model (4) leads to closed-form expressions for hyperparameter updates due to conjugacy.

4. CONJUGACY AND HYPERPARAMETER UPDATES

In this section, we derive the hyperparameter updates. For simplicity, let $\mathbf{y}$ be a generic observation of the form:

$$\mathbf{y} = \mathbf{R}(\delta)\mathbf{x} + \mathbf{w} \tag{6}$$

The conditional distribution $p(\boldsymbol{\mu}|\delta, \mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\mu}, \delta)p(\mu)$ is multivariate normal with mean and covariance given by:

$$\mathbb{E}[\boldsymbol{\mu}|\delta, \mathbf{y}] = \frac{\sigma_0^2}{\sigma^2 + \sigma_0^2}\mathbf{R}(\delta)^T\mathbf{y} + \frac{\sigma^2}{\sigma^2 + \sigma_0^2}\boldsymbol{\mu}_0$$
$$\mathrm{Cov}(\boldsymbol{\mu}|\delta, \mathbf{y}) = \frac{\sigma^2\sigma_0^2}{\sigma^2 + \sigma_0^2}\mathbf{I}$$

Recall the von Mises prior distribution on δ, i.e., $p(\delta) \propto e^{\kappa_0 \cos(\delta - \delta_0)}$. The posterior is given by

$$\begin{aligned} p(\delta|\mathbf{y}) &\propto p(\mathbf{y}|\delta)p(\delta) \\ &\propto \exp\left(\frac{-\|\mathbf{y} - \mathbf{R}(\delta)\boldsymbol{\mu}_0\|_2^2}{2(\sigma^2 + \sigma_0^2)} + \kappa_0 \cos(\delta - \delta_0) \right) \\ &\propto \exp\left(\frac{\mathbf{y}^T\mathbf{R}(\delta)\boldsymbol{\mu}_0}{\sigma^2 + \sigma_0^2} + \kappa_0 \cos(\delta - \delta_0) \right) \end{aligned} \tag{7}$$

We will show that the posterior is also a von Mises distribution, i.e., $p(\delta|\mathbf{y}) \propto \exp(\kappa_{new}\cos(\delta - \delta_{new}))$. Let $\boldsymbol{\mu}_0 = [\mu_{0,R}, \mu_{0,I}]^T$. Then, direct calculations yield:

$$\mathbf{y}^T\mathbf{R}(\delta)\boldsymbol{\mu}_0 = (\mathbf{y}^T\boldsymbol{\mu}_0)\cos\delta + (\mathbf{y}^T\tilde{\boldsymbol{\mu}}_0)\sin\delta \tag{8}$$

where $\tilde{\boldsymbol{\mu}}_0 \stackrel{\text{def}}{=} [-\mu_{0,I}, \mu_{0,R}]^T$. Furthermore, from a trigonometric identity, we obtain:

$$\kappa_0\cos(\delta - \delta_0) = \kappa_0\cos(\delta_0)\cos\delta + \kappa_0\sin(\delta_0)\sin\delta \tag{9}$$

Using (8) and (9) into (7), we obtain:

$$p(\delta|\mathbf{y}) \propto \exp\left(\left(\kappa_0\cos(\delta_0) + \frac{\mathbf{y}^T\boldsymbol{\mu}_0}{\sigma^2+\sigma_0^2}\right)\cos\delta + \left(\kappa_0\sin(\delta_0) + \frac{\mathbf{y}^T\tilde{\boldsymbol{\mu}}_0}{\sigma^2+\sigma_0^2}\right)\sin\delta\right)$$

This equals $\exp(\kappa_{new}\cos(\delta - \delta_{new}))$ for all $\delta \in [-\pi, \pi)$ iff:

$$\kappa_0\cos(\delta_0) + \frac{\mathbf{y}^T\boldsymbol{\mu}_0}{\sigma^2+\sigma_0^2} = \kappa_{new}\cos(\delta_{new})$$
$$\kappa_0\sin(\delta_0) + \frac{\mathbf{y}^T\tilde{\boldsymbol{\mu}}_0}{\sigma^2+\sigma_0^2} = \kappa_{new}\sin(\delta_{new})$$

Solving for κ_{new} and δ_{new}, we obtain:

$$\kappa_{new}^2 = \left(\kappa_0\cos(\delta_0) + \frac{\mathbf{y}^T\boldsymbol{\mu}_0}{\sigma^2+\sigma_0^2}\right)^2 + \left(\kappa_0\sin(\delta_0) + \frac{\mathbf{y}^T\tilde{\boldsymbol{\mu}}_0}{\sigma^2+\sigma_0^2}\right)^2$$
$$\delta_{new} = \tan^{-1}\left(\frac{\kappa_0\sin(\delta_0) + \frac{\mathbf{y}^T\tilde{\boldsymbol{\mu}}_0}{\sigma^2+\sigma_0^2}}{\kappa_0\cos(\delta_0) + \frac{\mathbf{y}^T\boldsymbol{\mu}_0}{\sigma^2+\sigma_0^2}}\right)$$

Returning to the model at the ith time instant, pre-multiplying with the rotation matrix $\mathbf{R}(\theta_{i-1})^T$:

$$\begin{aligned}\mathbf{R}(\theta_{i-1})^T\mathbf{y}_i &= \mathbf{R}(\theta_{i-1})^T\mathbf{R}(\theta_i)\mathbf{x}_i + \mathbf{R}(\theta_{i-1})^T\mathbf{w}_i \\ &= \mathbf{R}(\theta_{i-1})^{-1}\mathbf{R}(\theta_{i-1})\mathbf{R}(\delta)\mathbf{x}_i + \mathbf{R}(\theta_{i-1})^T\mathbf{w}_i \\ &= \mathbf{R}(\delta)\mathbf{x}_i + \mathbf{R}(\theta_{i-1})^T\mathbf{w}_i\end{aligned}$$

where $\mathbf{R}(\theta_{i-1})^T\mathbf{w}_i$ is Gaussian noise with zero mean and covariance $\sigma^2\mathbf{I}$, due to rotation invariance. Thus, the model (6) becomes applicable when applied to $\mathbf{R}(\theta_{i-1})^T\mathbf{y}_i$. This essentially amounts to rotating the vector $\mathbf{y}_i$ by an angle θ_{i-1}. A graphical illustration of this rotation is shown in Fig. 2.

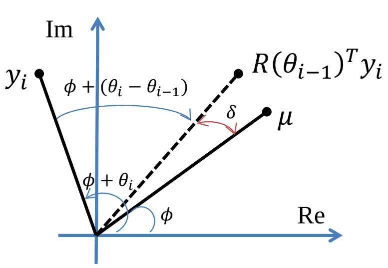

Fig. 2. Illustration of pre-rotation of observation $\mathbf{y}_i$ to reduce to model (6).

To summarize, once the γ_ith component is chosen, the parameter updates for the γ_ith class become:

$$\delta^{(i)} = \tan^{-1}\left(\frac{\kappa^{(i-1)}\sin\delta^{(i-1)} + \frac{\left\langle\mathbf{R}(\theta^{(i-1)})^T\mathbf{y}_i, \tilde{\boldsymbol{\mu}}_{\gamma_i}^{(i-1)}\right\rangle}{\sigma^2+(\sigma_{\gamma_i}^{(i-1)})^2}}{\kappa^{(i-1)}\cos\delta^{(i-1)} + \frac{\left\langle\mathbf{R}(\theta^{(i-1)})^T\mathbf{y}_i, \boldsymbol{\mu}_{\gamma_i}^{(i-1)}\right\rangle}{\sigma^2+(\sigma_{\gamma_i}^{(i-1)})^2}}\right) \tag{10}$$

$$(\kappa^{(i)})^2 = \left(\kappa^{(i-1)}\sin\delta^{(i-1)} + \frac{\left\langle\mathbf{R}(\theta^{(i-1)})^T\mathbf{y}_i, \tilde{\boldsymbol{\mu}}_{\gamma_i}^{(i-1)}\right\rangle}{\sigma^2+(\sigma_{\gamma_i}^{(i-1)})^2}\right)^2 + \left(\kappa^{(i-1)}\cos\delta^{(i-1)} + \frac{\left\langle\mathbf{R}(\theta^{(i-1)})^T\mathbf{y}_i, \boldsymbol{\mu}_{\gamma_i}^{(i-1)}\right\rangle}{\sigma^2+(\sigma_{\gamma_i}^{(i-1)})^2}\right)^2 \tag{11}$$

$$\theta^{(i)} = \theta^{(i-1)} + \delta^{(i)} \tag{12}$$

$$\gamma^{(i)} \sim q_h^{(i)} = \frac{L_{i,h}(\mathbf{R}(\theta^{(i)})^T\mathbf{y}_i)\pi_{i,h}}{\sum_h L_{i,h}(\mathbf{R}(\theta^{(i)})^T\mathbf{y}_i)\pi_{i,h}}$$

$$\boldsymbol{\mu}_{\gamma_i}^{(i)} = \frac{(\sigma_{\gamma_i}^{(i-1)})^2}{\sigma^2+(\sigma_{\gamma_i}^{(i-1)})^2}\mathbf{R}(\theta^{(i)})^T\mathbf{y}_i + \frac{\sigma^2}{\sigma^2+(\sigma_{\gamma_i}^{(i-1)})^2}\boldsymbol{\mu}_{\gamma_i}^{(i-1)} \tag{13}$$

$$(\sigma_{\gamma_i}^{(i)})^2 = (\sigma_{\gamma_i}^{(i-1)})^2\left(\frac{\sigma^2}{\sigma^2+(\sigma_{\gamma_i}^{(i-1)})^2}\right) \tag{14}$$

For exactly known means (i.e., constellation parameters), one can set $\sigma_h^{(i)} = 0$ and $\boldsymbol{\mu}_h^{(i)} = \boldsymbol{\mu}_h$ for all h, i. In this case, the updates (13)-(14) become superfluous. In practice, an early-late gate timing recovery method [9] can be coupled into the parameter updates (10)-(14) to track timing. Since early-late gate methods are based on magnitude only, they are not affected by the phase offset estimation.

5. CALCULATION OF CONDITIONAL LIKELIHOOD

To calculate the class posteriors $\{q_h^{(i)}\}$, the conditional likelihoods of $\mathbf{y}_i$ given assignment to class h and the previous class assignments need to be calculated first. The conditional likelihood of $\mathbf{y}_i$ given assignment to class h and the history $(\mathbf{y}^{(i-1)}, \gamma^{(i-1)})$ is given by:

$$L_{i,h}(\mathbf{y}_i) = \int f(\mathbf{y}_i|\theta_h)\pi(\theta_h|\mathbf{y}^{(i-1)}, \gamma^{(i-1)})d\theta_h \tag{15}$$

We recall from (5) that the posterior distribution has the product form:

$$\pi(\theta_h|\mathbf{y}^{(i-1)}, \gamma^{(i-1)}) = \mathcal{N}(\boldsymbol{\mu}_h|\boldsymbol{\mu}_h^{(i-1)}, (\sigma_h^{(i-1)})^2\mathbf{I})\mathcal{V}(\delta|\delta^{(i-1)}, \kappa^{(i-1)})$$

5.1. Exact Calculation

Using (5) into (15), we obtain:

$$L_{i,h}(\mathbf{y}_i) = \int_{-\pi}^{\pi} f_h^{(i-1)}(\delta; \mathbf{y}_i)\mathcal{V}(\delta|\delta^{(i-1)}, \kappa^{(i-1)})d\delta \tag{16}$$

where

$$f_h^{(i-1)}(\delta;\mathbf{y}_i) \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} \mathcal{N}(\mathbf{y}_i|\mathbf{R}(\theta^{(i-1)}+\delta)\boldsymbol{\mu},\sigma^2\mathbf{I}) \times \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_h^{(i-1)},(\sigma_h^{(i-1)})^2\mathbf{I})d\boldsymbol{\mu} = \frac{1}{(2\pi\sigma^2)^{d/2}}\frac{1}{(2\pi(\sigma_h^{(i-1)})^2)^{d/2}} \times \int e^{-\frac{1}{2}\left(\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_h^{(i-1)}\|_2^2}{(\sigma_h^{(i-1)})^2}+\frac{\|\boldsymbol{\mu}-R(\theta^{(i-1)}+\delta)^T\mathbf{y}_i\|_2^2}{\sigma^2}\right)}d\boldsymbol{\mu}$$

After completing the square, using the Gaussian normalization and simplifying, we obtain:

$$f_h^{(i-1)}(\delta;\mathbf{y}_i) = \frac{\exp\left(-\frac{\|\mathbf{R}(\theta^{(i-1)}+\delta)^T\mathbf{y}_i-\boldsymbol{\mu}_h^{(i-1)}\|_2^2}{2(\sigma^2+(\sigma_h^{(i-1)})^2)}\right)}{(2\pi(\sigma^2+(\sigma_h^{(i-1)})^2))^{d/2}}$$

Define $\tilde{\mathbf{y}}_i = \mathbf{R}(\theta^{(i-1)})^T\mathbf{y}_i$. Plugging this expression into (16), we obtain after some algebra:

$$\begin{aligned} L_{i,h}(\mathbf{y}_i) &= \frac{1}{2\pi I_0(\kappa^{(i-1)})}\frac{\exp\left(-\frac{\|\mathbf{y}_i\|_2^2+\|\boldsymbol{\mu}_h^{(i-1)}\|_2^2}{2(\sigma^2+(\sigma_h^{(i-1)})^2)}\right)}{(2\pi(\sigma^2+(\sigma_h^{(i-1)})^2))^{d/2}} \\ &\times \int_{-\pi}^{\pi}\exp\left(\kappa^{(i-1)}\cos(\delta-\delta^{(i-1)}) + \frac{1}{\sigma^2+(\sigma_h^{(i-1)})^2}\tilde{\mathbf{y}}_i^T\mathbf{R}(\delta)\boldsymbol{\mu}_h^{(i-1)}\right)d\delta \\ &= \frac{1}{2\pi I_0(\kappa^{(i-1)})}\frac{\exp\left(-\frac{\|\mathbf{y}_i\|_2^2+\|\boldsymbol{\mu}_h^{(i-1)}\|_2^2}{2(\sigma^2+(\sigma_h^{(i-1)})^2)}\right)}{(2\pi(\sigma^2+(\sigma_h^{(i-1)})^2))^{d/2}} \\ &\times \int_{-\pi}^{\pi}\exp\left(\kappa_{h,i-1}\cos(\delta-\delta_{h,i-1})\right)d\delta \\ &= \frac{I_0(\kappa_{h,i-1})}{I_0(\kappa^{(i-1)})}\frac{\exp\left(-\frac{\|\mathbf{y}_i\|_2^2+\|\boldsymbol{\mu}_h^{(i-1)}\|_2^2}{2(\sigma^2+(\sigma_h^{(i-1)})^2)}\right)}{(2\pi(\sigma^2+(\sigma_h^{(i-1)})^2))^{d/2}} \end{aligned} \tag{17}$$

where the parameters $\delta_{h,i-1}, \kappa_{h,i-1}$ are given by:

$$\begin{aligned} \delta_{h,i-1} &= \tan^{-1}\left(\frac{\kappa^{(i-1)}\sin\delta^{(i-1)}+\frac{\langle\tilde{\mathbf{y}}_i,\tilde{\boldsymbol{\mu}}_h^{(i-1)}\rangle}{\sigma^2+(\sigma_h^{(i-1)})^2}}{\kappa^{(i-1)}\cos\delta^{(i-1)}+\frac{\langle\tilde{\mathbf{y}}_i,\boldsymbol{\mu}_h^{(i-1)}\rangle}{\sigma^2+(\sigma_h^{(i-1)})^2}}\right) \\ (\kappa_{h,i-1})^2 &= \left(\kappa^{(i-1)}\sin\delta^{(i-1)}+\frac{\langle\tilde{\mathbf{y}}_i,\tilde{\boldsymbol{\mu}}_h^{(i-1)}\rangle}{\sigma^2+(\sigma_h^{(i-1)})^2}\right)^2 \\ &+ \left(\kappa^{(i-1)}\cos\delta^{(i-1)}+\frac{\langle\tilde{\mathbf{y}}_i,\boldsymbol{\mu}_h^{(i-1)}\rangle}{\sigma^2+(\sigma_h^{(i-1)})^2}\right)^2 \end{aligned} \tag{18}$$

This computation becomes difficult to compute numerically due to the ratio of Bessel functions, for large $\kappa^{(i-1)}$.

5.2. Asymptotic Approximation

As the number of iterations grow, the hyperparameter $\kappa^{(i-1)}$ gets large, corresponding to high precision during learning. Expanding the expression in (18), we obtain:

$$\begin{aligned} (\kappa_{h,i-1})^2 &= (\kappa^{(i-1)})^2 + \frac{(\langle\tilde{\mathbf{y}}_i,\tilde{\boldsymbol{\mu}}_h^{(i-1)}\rangle)^2+(\langle\tilde{\mathbf{y}}_i,\boldsymbol{\mu}_h^{(i-1)}\rangle)^2}{(\sigma^2+(\sigma_h^{(i-1)})^2)^2} \\ &+ \frac{2\kappa^{(i-1)}}{\sigma^2+(\sigma_h^{(i-1)})^2}\left(\langle\tilde{\mathbf{y}}_i,\tilde{\boldsymbol{\mu}}_h^{(i-1)}\rangle\sin\delta^{(i-1)}+\langle\tilde{\mathbf{y}}_i,\boldsymbol{\mu}_h^{(i-1)}\rangle\cos\delta^{(i-1)}\right) \\ &= (\kappa^{(i-1)})^2 + \frac{\|\tilde{\mathbf{y}}_i\|_2^2\|\boldsymbol{\mu}_h^{(i-1)}\|_2^2}{(\sigma^2+(\sigma_h^{(i-1)})^2)^2} + \frac{2\kappa^{(i-1)}\tilde{\mathbf{y}}_i^T\mathbf{R}(\delta^{(i-1)})\boldsymbol{\mu}_h^{(i-1)}}{\sigma^2+(\sigma_h^{(i-1)})^2} \end{aligned}$$

where, in the last step, we used the fact that $\left\{\frac{\tilde{\boldsymbol{\mu}}_h^{(i-1)}}{\|\boldsymbol{\mu}_h^{(i-1)}\|_2}, \frac{\boldsymbol{\mu}_h^{(i-1)}}{\|\boldsymbol{\mu}_h^{(i-1)}\|_2}\right\}$ is an orthonormal basis. We next argue that the norm of the random vector $\tilde{\mathbf{y}}_i$ cannot be too large with high probability. The Gaussian random vector $\tilde{\mathbf{y}}_i$ can be written as $\tilde{\mathbf{y}}_i = \mathbf{R}(\theta^{(i-1)})^T\mathbf{R}(\theta_i)\mathbf{x}_i + \mathbf{e}_i$, where $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0},\sigma^2\mathbf{I})$ and $\mathbf{x}_i$ is the ith symbol. Since $\|\mathbf{e}_i\|_2^2 = \sum_{j=1}^2([\mathbf{e}_i]_j)^2$ is a sum of independent chi-squared random variables, it follows that $\|\mathbf{e}_i\|_2^2$ is chi-squared distributed with 2 degrees of freedom, and as a result:

$$\mathbb{P}(\|\mathbf{e}_i\|_2^2 > t) = \int_t^{\infty}\frac{e^{-u/(2\sigma^2)}}{2}du = e^{-\frac{t}{2\sigma^2}}$$

which implies that $\|\mathbf{e}_i\|_2^2 \leq 2\sigma^2\ln(1/\delta)$ with probability at least $1-\delta$. Factoring out $(\kappa^{(i-1)})^2$:

$$(\kappa_{h,i-1})^2 = (\kappa^{(i-1)})^2\left(1+\frac{2\tilde{\mathbf{y}}_i^T\mathbf{R}(\delta^{(i-1)})\boldsymbol{\mu}_h^{(i-1)}}{\kappa^{(i-1)}\left(\sigma^2+(\sigma_h^{(i-1)})^2\right)}+O\left(\frac{1}{\kappa^{(i-1)}}\right)^2\right).$$

Thus, after taking square roots and using the Taylor approximation $\sqrt{1+x} \approx 1+\frac{x}{2}$ for small x, we obtain

$$\kappa_{h,i-1} = \kappa^{(i-1)} + \frac{\tilde{\mathbf{y}}_i^T\mathbf{R}(\delta^{(i-1)})\boldsymbol{\mu}_h^{(i-1)}}{\sigma^2+(\sigma_h^{(i-1)})^2} + O\left(\frac{1}{\kappa^{(i-1)}}\right). \tag{19}$$

Recall the large-z approximation of the first-kind zero-order Bessel function $I_0(z)$ [10]:

$$I_0(z) \approx \frac{e^z}{\sqrt{2\pi z}}$$

For large $\kappa^{(i-1)}$, in virtue of (19), the ratio of Bessel functions in (17) can be approximated as:

$$\begin{aligned} \frac{I_0(\kappa_{h,i-1})}{I_0(\kappa^{(i-1)})} &\approx \sqrt{\frac{\kappa^{(i-1)}}{\kappa_{h,i-1}}}e^{\kappa_{h,i-1}-\kappa^{(i-1)}} \\ &= \frac{\exp\left(\frac{\tilde{\mathbf{y}}_i^T\mathbf{R}(\delta^{(i-1)})\boldsymbol{\mu}_h^{(i-1)}}{\sigma^2+(\sigma_h^{(i-1)})^2}+O\left(\frac{1}{\kappa^{(i-1)}}\right)\right)}{\sqrt{1+\frac{1}{\kappa^{(i-1)}}\frac{\tilde{\mathbf{y}}_i^T\mathbf{R}(\delta^{(i-1)})\boldsymbol{\mu}_h^{(i-1)}}{\sigma^2+(\sigma_h^{(i-1)})^2}+O\left(\frac{1}{(\kappa^{(i-1)})^2}\right)}} \\ &\longrightarrow \exp\left(\frac{\tilde{\mathbf{y}}_i^T\mathbf{R}(\delta^{(i-1)})\boldsymbol{\mu}_h^{(i-1)}}{\sigma^2+(\sigma_h^{(i-1)})^2}\right) \end{aligned}$$

as $\kappa^{(i-1)} \to \infty$. Using this approximation in (17), we obtain after some algebra:

$$L_{i,h}(\mathbf{y}_i) \approx \frac{\exp\left(-\frac{\|\mathbf{R}(\theta^{(i-1)}+\delta^{(i-1)})^T\mathbf{y}_i-\boldsymbol{\mu}_h^{(i-1)}\|_2^2}{2(\sigma^2+(\sigma_h^{(i-1)})^2)}\right)}{(2\pi(\sigma^2+(\sigma_h^{(i-1)})^2))^{d/2}} \tag{20}$$

Thus, the large-sample approximation to the conditional likelihood is a normal distribution, i.e., $\mathbf{y}_i \sim \mathcal{N}(\cdot|\mathbf{R}(\theta^{(i-1)}+\delta^{(i-1)})\boldsymbol{\mu}_h^{(i-1)}, (\sigma^2+(\sigma_h^{(i-1)})^2)\mathbf{I})$. We remark that a similar asymptotic normality result was proven in [8] for the conditional likelihood under the Gaussian-Wishart conjugate model. In the simulations, we use the logarithm of (20) for numerical stability to implement the selection step (1).

6. SIMULATIONS

In this section, we perform a simulation experiment on detecting and estimating the constellation parameters of a 16-QAM modulation. The phase rotation angle per-symbol is $\delta = 3°$. The correct number of classes for this data set is 16, each one corresponding to 4 information bits.

Data symbols $\{\mathbf{x}_i\}$ are randomly (uniformly) chosen from the 16-QAM constellation and corrupted with additive noise at SNR of 15dB. The data is plotted in Fig. 3, along with the original signaling locations arranged in a rectangular grid.

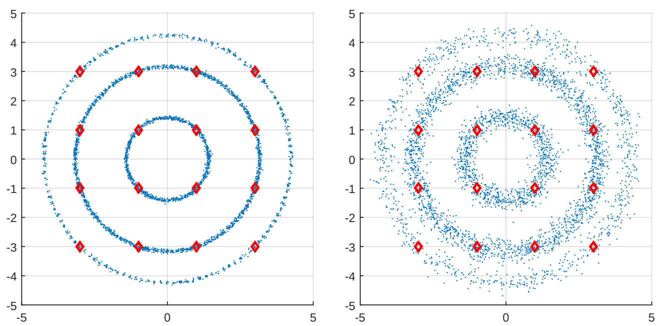

Fig. 3. Noisy constellation data from 16-QAM constellation with phase rotation rate of $\delta = 3°$ at 30dB (left) and 15dB (right).

For comparison, we consider the approximate sequential maximum-likelihood (ML) phase estimator (see Appendix A for derivation):

$$\theta^{(i)} = \angle \left\{ \frac{y_i}{P_{\mathcal{A}}(e^{-j\theta^{(i-1)}} y_i)} \right\} \tag{21}$$

where $y_i = y_i^R + jy_i^I$, $P_{\mathcal{A}}(\mathbf{z}) \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{\mu}\in\mathcal{A}} \|\mathbf{z} - \boldsymbol{\mu}\|_2^2$ is the minimum norm projection on the constellation $\mathcal{A}$. Fig. 4 shows the clustering performance of this approximate ML method for 30dB and 15dB SNR. For low SNR, the clusters have a high circular spread and are not appropriately estimated. As shown in Fig. 4, the phase offset δ is not learned as the number of samples grow; although the mean is around $3°$, the variance is quite high (SNR-dependent) and is not decaying as a function of iteration.

To alleviate this problem and improve the estimation performance at low SNR, we use the online Bayesian estimation algorithm developed to estimate the frequency offset. Fig. 5 shows the clustering performance and the estimation performance of the Bayesian algorithm. In our simulations, we implemented the parameter updates (10)-(12), and set $\boldsymbol{\mu}_h^{(i)} = \boldsymbol{\mu}_h, \sigma_h^{(i)} = 0$ for all h, i because no uncertainty in the constellation parameters was assumed. The algorithm was initialized with $\delta^{(1)} = 0$. With the Bayesian learning algorithm based on the von Mises prior, asymptotic learning occurs for the phase offset δ as more samples are processed. This leads to a significantly more stable performance when compared to the approximate ML phase estimator (see Fig. 4). Fig. 5 implies that the rate of learning is SNR-dependent; the higher the SNR, the faster the frequency offset parameter is learned.

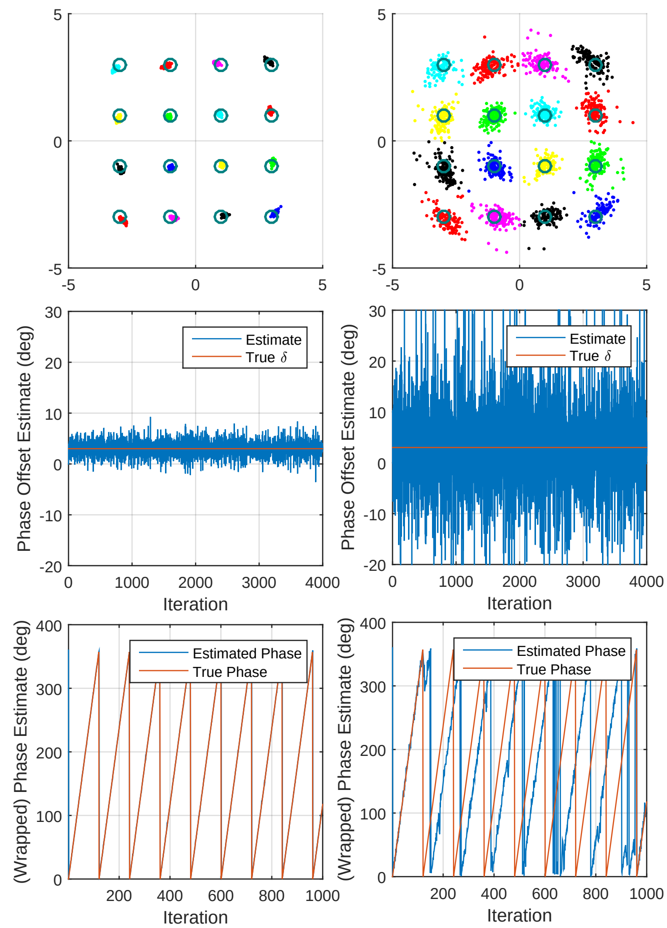

Fig. 4. Resulting compensated constellation (top), frequency offset estimate (middle) and phase estimate (bottom), for approximate sequential ML phase estimator (21) for 30dB (left) and 15dB SNR (right), respectively.

7. CONCLUSION

We have proposed an online Bayesian framework for blind estimation of the frequency offset and the parameters of an arbitrary signal constellation. Our approach leveraged novel conjugate prior distribution theory for the von Mises and Gaussian distribution, which allowed us to derive closed-form updates of the hyperparameters of the posterior distribution of the unknown parameters given streaming data. Asymptotic normality of the conditional likelihood was proven, and led to numerically stable algorithms. Simulations showed that our estimation algorithm results in fast convergence and learning of the frequency offset, and significantly outperformed the approximate sequential maximum-likelihood frequency offset estimators.

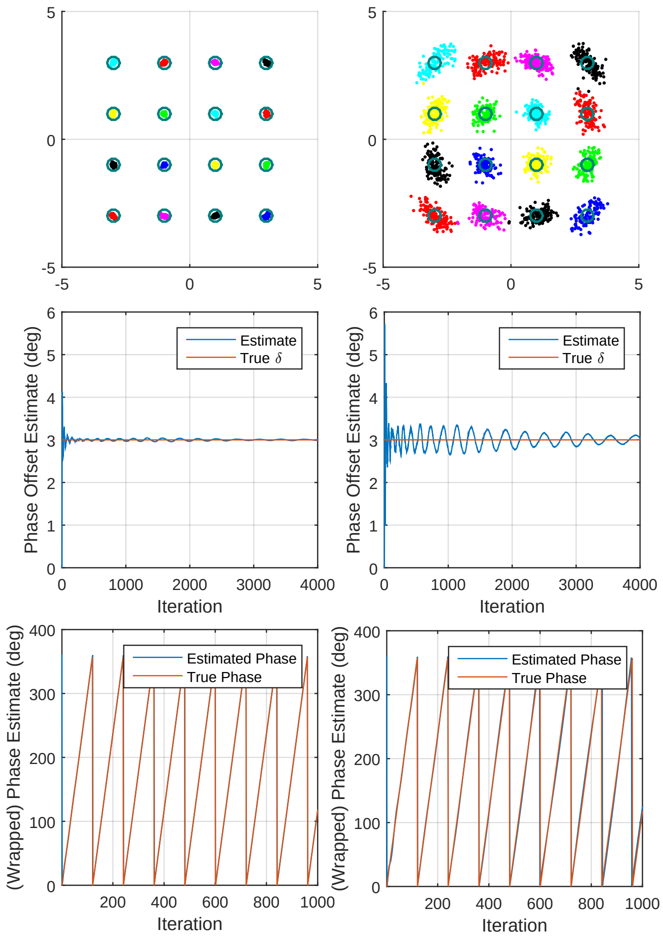

Fig. 5. Resulting compensated constellation (top), frequency offset estimate (middle) and phase estimate (bottom), for sequential Bayesian estimator (10)-(14) for 30dB (left) and 15dB SNR (right), respectively.

Appendix A: Derivation of approximate sequential ML phase estimator

In this appendix, we derive the approximate sequential ML phase estimator. The ML estimator for θ maximizes the likelihood:

$$p(y_k|\theta) = \frac{1}{(\pi\sigma^2)^d} \exp\left(-\frac{|y_k - x_k e^{j\theta_k}|^2}{\sigma^2}\right)$$

$$= \frac{\exp\left(-\frac{|y_k|^2+|x_k|^2}{\sigma^2}\right)}{(\pi\sigma^2)^d} \exp\left(\frac{2|x_k||y_k|}{\sigma^2}\cos(\angle\{y_k\} - (\angle\{x_k\} + \theta_k))\right)$$

Taking the logarithm of $p(y_k|\theta)$, the necessary condition for the ML estimator becomes:

$$\frac{\partial \log p(y_k|\theta)}{\partial\theta} = \frac{2|x_k||y_k|}{\sigma^2}\sin(\angle\{y_k\} - (\angle\{x_k\} + \theta_k)) = 0$$

For high SNR, the phase error $e_k = \angle\{y_k\} - (\angle\{x_k\} + \theta_k)$ is expected to be small. Using the small-angle approximation $\sin(x) \approx x$, the optimality condition becomes $\angle\{y_k\} - (\angle\{x_k\} + \theta_k) \approx 0$, which leads to the approximate ML phase estimator:

$$\theta_k^{(ML)} = \angle\left\{\frac{y_k}{x_k}\right\} \tag{22}$$

In practice, x_k is unknown so (22) cannot be implemented directly. Let $\theta^{(k-1)}$ denote the phase estimator corresponding to the previous sample. For small phase rotation δ, the symbol x_k can be approximated as:

$$\begin{aligned}
P_{\mathcal{A}}(e^{-j\theta^{(k-1)}}y_k) &= P_{\mathcal{A}}(e^{-j\theta^{(k-1)}}(x_k e^{j\theta_k} + w_k)) \\
&\approx P_{\mathcal{A}}(x_k e^{j(\theta_k - \theta^{(k-1)})}) && \text{(high SNR)} \\
&\approx P_{\mathcal{A}}(x_k e^{j\delta}) && (\theta^{(k-1)} \approx \theta_{k-1}) \\
&\approx x_k && (\delta \approx 0)
\end{aligned}$$

Thus, substituting $P_{\mathcal{A}}(e^{-j\theta^{(k-1)}}y_k)$ into the denominator of the angle argument in (22), we arrive at (21).

8. REFERENCES

[1] U. Mengali and A. N. D'Andrea, *Synchronization techniques for digital receivers*, Applications of Communications Theory. Plenum Press, New York, London, 1997.

[2] G. Vazquez and J. Riba, "Non data-aided digital synchronization," in *Trends in Single and Multi-User Systems (Vol. II)*, G. B. Giannakis and Y. Hua, Eds. Prentice-Hall, Upper Saddle River, NJ, 2000.

[3] H. Fu and P. Y. Kam, "MAP/ML Estimation of the Frequency and Phase of a Single Sinusoid in Noise," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 834–845, March 2007.

[4] T. Ghirmai, M. F. Bugallo, J. Miguez, and P. M. Djuric, "A Sequential Monte Carlo Method for Adaptive Blind Timing Estimation and Data Detection," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2855–2865, August 2005.

[5] J.-J. van de Beek, M. Sandell, and P. O. Borjesson, "ML Estimation of Time and Frequency Offset in OFDM Systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1800–1805, July 1997.

[6] M. P. Fitz and W. C. Lindsey, "Decision-Directed Burst-Mode Carrier Synchronization Techniques," *IEEE Transactions on Communications*, vol. 40, no. 10, pp. 1644–1653, October 1992.

[7] D. P. Taylor, "Introduction to "Synchronous Communications"," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1459–1460, August 2002.

[8] T. Tsiligkaridis and K. W. Forsythe, "Adaptive Low-Complexity Sequential Inference for Dirichlet Process Mixture Models," *Preprint, arXiv: 1409.8185*, September 2014.

[9] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, Wiley, 2013.

[10] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1995.