© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Signal Amplitude Estimation and Detection from Unlabeled Binary Quantized Samples

Guanyu Wang, *Student member, IEEE*, Jiang Zhu, *member, IEEE*, Rick S. Blum, *Fellow, IEEE*, Peter Willett, *Fellow, IEEE*, Stefano Marano, Vincenzo Matta, Paolo Braca *Senior member, IEEE*

***Abstract*—Signal amplitude estimation and detection from unlabeled quantized binary samples are studied, assuming that the order of the time indexes is completely unknown. First, maximum likelihood (ML) estimators are utilized to estimate both the permutation matrix and unknown signal amplitude under arbitrary, but known signal shape and quantizer thresholds. Sufficient conditions are provided under which a ML estimator can be found in polynomial time and an alternating maximization algorithm is proposed to solve the general problem via a good initializations scheme. In addition, the statistical identifiability of the model is studied.**

Furthermore, an approximation of the generalized likelihood ratio test (GLRT) detector is adopted to detect the presence of signal. In addition, an accurate approximation of the probability of successful permutation matrix recovery is derived, and explicit expressions are provided to reveal the relationship between the signal length and the number of quantizers. Finally, numerical simulations are performed to verify the theoretical results.

***Index Terms*—Estimation, detection, permutation, unlabeled sensing, quantization, identifiability, alternating maximization.**

I. Introduction

In many systems, the data is transmitted with time information, which may sometimes be imprecise [1], [2], [3], [4], [5], [6], [7]. One example is the global positioning system (GPS) spoofing attack which can alter the time stamps on electric grid measurements [1] and make them useless so that the data must be processed without time stamps. Since the exact form of civilian GPS signals is publicly known and the elements needed are inexpensive, building a circuit to generate signals to spoof the GPS is easy. In [2], a refined assessment of the spoofing threat is provided. In addition, the detailed information of receiver-spoofer architecture, its implementation and performance, and spoofing countermeasures are introduced. As a case study in [3], the impact of the GPS spoofing attack on the wireless communication networks, more specifically, the frequency hopping code division multiple access (FH-CDMA) based ad hoc network, is investigated. A timing synchronization attack (TSA) has been coined to the wide area monitoring systems (WAMSs), and its effectiveness is demonstrated to three applications of a phasor measurement unit (PMU) [4]. In [5], the out-of-sequence measurement (OOSM) problem where the observations produced by sensors are sent to a fusion center (FC) over communication networks with random delays is studied, and a Bayesian solution is provided. The problem of random delay and packet loss in networked control systems (NCSs) is studied in [6]. In addition, a minimum error covariance estimator is derived, and two alternative estimator architectures are presented for efficient computation. In [7], the effect of an unknown timestamp delay in Automatic Identification System (AIS) is studied, and a method based on adaptive filtering is proposed.

In the above examples, the relative order of the data is unknown, i.e., the samples are unlabeled. Estimation and detection from unlabeled samples have drawn a great deal of attention recently [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. In [8], it is shown that the convex relaxation based on a Birkhoff polytope approach does not recover the permutation matrix, and a global branch and bound algorithm is proposed instead. In the noiseless case with a random linear sensing matrix, it is shown that the permutation matrix can be recovered correctly with probability 1, given that the number of measurements is twice the number of unknowns [9], [18]. In [10], [19], the noise is taken into account, and a condition under which the permutation matrix can be recovered with high probability are provided. In addition, a polynomial time algorithm is proposed for the scalar parameter case. Denoising linear regression model with shuffled data and additive Gaussian noise is studied in [11]. The characterization of minimax error rate is given. An algorithm for the noiseless problem is also proposed, and its performance is demonstrated on an image point-cloud matching task [11]. In [12], several estimators are compared in recovering the weights of the

Manuscript received November 18, 2017; revised May 11, 2018; accepted June 13, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tay Wee Peng. G. Wang and J. Zhu are supported by Zhejiang Provincial Natural Science Foundation of China under Grant LQ18F010001 and the Fundamental Research Funds for the Central Universities under Grant 2017QNA4042. R. S. Blum is partially supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-17-1-0331 and by the National Science Foundation under grants ECCS-1744129, and ECCS-1405579. P. Braca is supported by NATO ACT under project SAC000808-"Data Knowledge Operational Effectiveness."(*Corresponding Author: Jiang Zhu.*)

G. Wang and J. Zhu are with the key laboratory of ocean observation-imaging testbed of Zhejiang Province, Ocean College, Zhejiang University, No.1 Zheda Road, Zhoushan, China, 316021 (e-mail: guanyuwang@zju.edu.cnjiangzhu16@zju.edu.cn).

R. S. Blum is with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: rb0f@lehigh.edu).

P. Willet is with the Department of ECE, University of Connecticut, Storrs, CT 06269, USA (e-mail: willett@engr.uconn.edu).

S. Marano and V. Matta are with the Department of Electrical and Information Engineering and Applied Mathematics, University of Salerno, Fisciano I-84084, Italy (e-mail: marano@unisa.it; vmatta@unisa.it).

P. Braca is with the NATO STO Centre for Maritime Research and Experimentation, 19126 La Spezia, Italy (e-mail: paolo.braca@cmre.nato.int).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material provides a brief proof of Claim 1 (page 5). This material is 136KB in size.

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

noisy linear model from shuffled labels, and an estimator based on the self-moments of the input features and labels is introduced. For the unlabeled ordered sampling problem [13], an alternating maximization algorithm combined with dynamic programming is proposed. In [15], a signal detection problem where the known signal is permuted in an unknown way is studied.

Compared to the location parameter estimation problem ($x_i = \theta + w_i$) in [17, eq. (1)], the model in this paper is a scale parameter estimation problem ($x_i = h_i\theta + w_i$) where $h_i,\ i = 1, \cdots, K$ is the shape of a signal, and θ is the amplitude of signal. As a result, the scale parameter estimation problem is much more difficult than the location estimation in several aspects, and the scale parameter is especially relevant in relation to the mislabeling/permutation issue. First, the model in [17] is always identifiable, while our model may be unidentifiable, as shown later. Second, the problem in [17] can be solved efficiently via simple sorting, while we can only prove that the problem in this paper can be solved efficiently under certain conditions. Third, good initial points are proposed to improve the performance of alternating maximization algorithm. Furthermore, we provide an approximation to the probability of successful permutation matrix recovery, which reveals the relationship between the length of signal and the number of quantizers.

In this paper, we focus on the problems of scale estimation and signal detection from unlabeled quantized samples. In the first part of this paper we consider the estimation of θ from noisy measurements, and in the second part we address the detection problem of deciding between the null hypothesis of observing only noise, against a composite alternative in which the signal values $\{h_i\}_{i=1}^K$ are known except for a common multiplicative factor θ. Both problems have very great practical relevance and have been widely addressed in the literature of distributed inference, even under the assumption that data arrives at the fusion center after quantization and is possibly flipped by noisy links. The main contribution of this paper is to revise these classical problems under the emerging paradigm of unlabeled data [1], [3], [4], [7], in which the observations at the fusion center lack a timing reference. To be specific, we first provide a sufficient condition for the existence of a polynomial time algorithm for the unlabeled estimation problem, and the model is shown to be unidentifiable in some special cases. Second, good initial points are provided to improve the performance of an alternating maximization algorithm. And third, we provide analytic approximations on probability of permutation matrix recovery in the case of known signal amplitude, which can be used to predict when the permutation matrix can be correctly recovered.

The organization of this paper is as follows. In Section II, the problem is described. Background on ML estimation and generalized likelihood ratio test (GLRT) detection from labeled data is presented in Section III. In Section IV, the model identifiability is studied, and the estimation problem from unlabeled data is studied. Section V extends the detection work to unlabeled data, and derives an approximate analytic formula for permutation matrix recovery probability. Finally, the numerical results are presented in Section VI, and the conclusion follows in Section VII.

Notation: The $K \times 1$ vector of ones is $\mathbf{1}_K$. For an unknown deterministic parameter θ, θ_0 denotes its true value. For an unknown permutation matrix $\mathbf{\Pi}$, $\mathbf{\Pi}_0$ denotes its true value. For a random vector $\mathbf{y}$, $p(\mathbf{y};\theta)$ denotes the probability density function (PDF) of $\mathbf{y}$ parameterized by θ, and $\mathrm{E}_{\mathbf{y}}[\cdot]$ denotes the expectation taken with respect to $\mathbf{y}$. Let $\mathcal{N}(\mu, \sigma^2)$ denote a Gaussian distribution with mean μ and variance σ^2. Let $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the cumulative distribution function (CDF) and probability density function (PDF) of a standard Gaussian random variable respectively. Let $\mathcal{U}(a, b)$ denote an uniform distribution, whose minimum and maximum values are a and b. Let $\mathcal{B}(N, p)$ denote a binomial distribution, where N and p denote the number of trials and the probability of event, respectively.

II. Problem Setup

Consider a signal amplitude estimation and detection problem where a collection of N binary quantizers generates binary quantized samples which will be utilized to estimate the unknown scaling factor θ of a K length signal and detect the presence of the signal, as shown in Fig. 1. The binary quantized samples b_{ij} are obtained via

$$b_{ij} = Q_i(h_i\theta + w_{ij}), \quad i = 1, \cdots, K, \quad j = 1, \cdots, N, \tag{1}$$

and the corresponding hypothesis problem can be formulated as

$$\begin{cases} \mathcal{H}_0 : b_{ij} = Q_i(w_{ij}),\ i = 1, \cdots, K,\ j = 1, \cdots, N, \\ \mathcal{H}_1 : b_{ij} = Q_i(h_i\theta + w_{ij}),\ i = 1, \cdots, K,\ j = 1, \cdots, N, \end{cases}$$

where i and j respectively denote one of the K time indexes and one of the N quantizers, $h_i,\ i = 1, \cdots, K$, are the coefficients characterizing the signal shape and are a priori known, w_{ij} is the i.i.d. noise drawn from the σ_w^2-variance distribution whose PDF is $f_w(x/\sigma_w)/\sigma_w$ and CDF is $F_w(x/\sigma_w)$, where $f_w(x)$ and $F_w(x)$ are the corresponding unit-variance PDF and CDF, and $Q_i(\cdot)$ implies a binary quantizer which produces 1 if the argument is larger than a scalar threshold τ_i and 0 otherwise. The thresholds of N quantizers are identical given any time index[1]. We assume that the PDF $f_w(w)$ is log-concave, which is often met in practice such as Gaussian distributions.

The quantized data $\{b_{ij}\}$ is transmitted over a binary channel with flipping probabilities q_0 and q_1 which are defined as $\Pr(u_{ij} = 1|b_{ij} = 0) = q_0$ and $\Pr(u_{ij} = 0|b_{ij} = 1) = q_1$, where u_{ij} is the sample received at the output of the channel, which we call the FC [21].

We assume that all the sets of data $\{u_{ij}\}_{j=1}^N$ are transmitted to the FC with permuted time indexes. Accordingly, the FC receives the set of data, say $\{\tilde{u}_{ij}\}_{j=1}^N$, whose time reference (represented by the index i) is invalid. The FC does not know

[1]Here we have thresholds fixed across quantizers and varying with time, with permutation across time. We could, equivalently, have fixed thresholds of quantizers across time but varying across sensors and permuted across sensors. Mathematically, it is the same problem and the formulation could as easily encompass it.

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

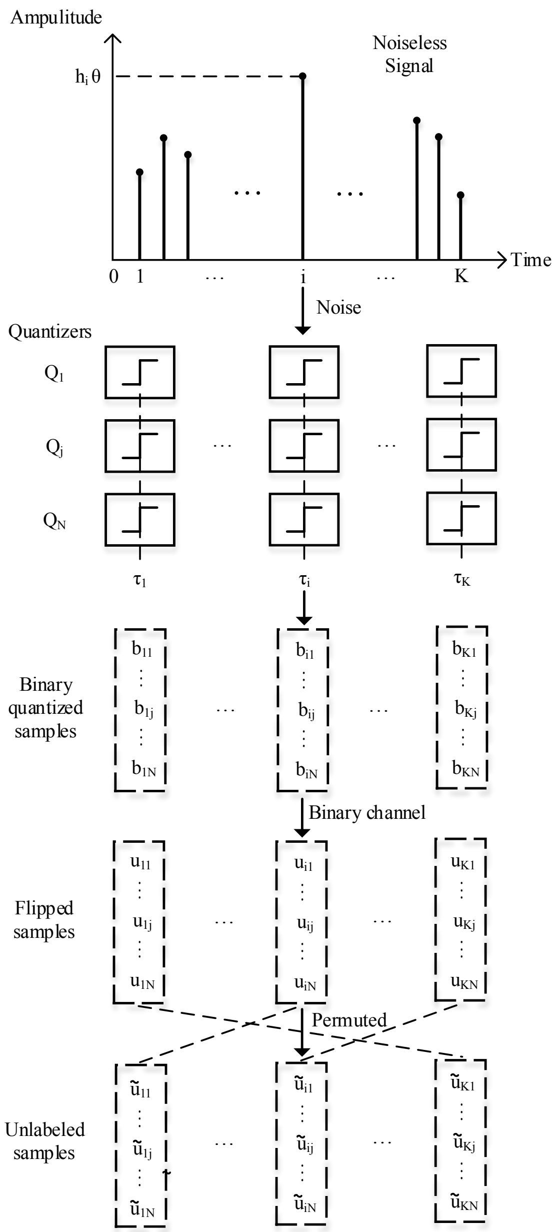

Fig. 1: System diagram of unlabeled binary quantized samples generation.

which time index the data $\{\tilde{u}_{ij}\}_{j=1}^N$ belongs to, but knows that $\{\tilde{u}_{ij}\}_{j=1}^N$ belongs to one of the K time indexes. Let us introduce the matrix $\mathbf{U}$ whose $(i,j)-$th entry is u_{ij}. Then, the unlabeled samples can be collected in a matrix $\widetilde{\mathbf{U}}$, as follows:

$$\widetilde{\mathbf{U}} = \mathbf{\Pi}\mathbf{U}, \tag{2}$$

where $\mathbf{\Pi} \in \mathbb{R}^{K\times K}$ is an unknown permutation matrix, i.e., a matrix of $\{0,1\}$ entries in which each row or each column sums to unity. We assume that θ is constrained to an interval $[-\Delta, \Delta]$, for algorithm and theoretical reasons [22, p. 4].

It is worth mentioning that GPS spoofing attack on time synchronization in smart grid networks and wireless networks can be abstracted to the above model [1], [3]. In smart grid networks, θ can be viewed as the nodal voltage magnitude which can be unobservable by PMUs at all the generator nodes in the network. h_i is the cosine function value of the voltage phase describing the shape of voltage signal. The synchronization delay attack changes the model by adding an extra factor to the phase, and the attacked signal can be regarded as a permutation $\mathbf{\Pi}$ of the original signal [1]. In FH-CDMA based ad-hoc network, h_i is the known GPS signal shape. The transmission of the GPS signal is under the Rayleigh fading channel, whose amplitude gain θ is unknown. The time offsets caused by the GPS spoofing attack impact the time sequences of received signals and can be modeled by a permutation matrix $\mathbf{\Pi}$ [3].

III. Preliminaries

In this section, standard materials of parameter estimation and signal detection using labeled data are presented.

A. Maximum likelihood estimation

The probability mass function (PMF) of u_{ij} can be calculated as

$$\begin{aligned}&\Pr(u_{ij}=1) = q_0 + (1-q_0-q_1)F_w\left(\frac{h_i\theta-\tau_i}{\sigma_w}\right) \triangleq p_i,\\ &\Pr(u_{ij}=0) = 1-p_i.\end{aligned} \tag{3}$$

The PMF of $\mathbf{U}$ is

$$p(\mathbf{U};\theta) = \prod_{i=1}^K\prod_{j=1}^N \Pr(u_{ij}=1)^{u_{ij}}\Pr(u_{ij}=0)^{(1-u_{ij})}. \tag{4}$$

Let η_i denote the fraction of $u_{ij}=1$ in $\{u_{ij}\}_{j=1}^N$, i.e.,

$$\eta_i = \sum_{j=1}^N u_{ij}/N. \tag{5}$$

Consequently, the log-likelihood function $l(\boldsymbol{\eta};\theta)$ is

$$l(\boldsymbol{\eta};\theta) = N\sum_{i=1}^K(\eta_i\log p_i + (1-\eta_i)\log(1-p_i)), \tag{6}$$

where p_i is given in (3). In the error free binary symmetric channel scenario, i.e., $q_0=q_1=0$ or $q_0=q_1=1$, the CDF $F_w(x)$ is log-concave because it is the integral of a log-concave PDF $f_w(x)$. Therefore maximizing the log-likelihood function is a convex optimization problem, which can be solved efficiently via numerical algorithms [23], [24], [25, pp. 7-8]. For $0<q_0+q_1<2$, it is difficult to determine the convexity of the negative log-likelihood function. In this case a local optimum is guaranteed. As we show in numerical experiments, we found that the ML estimator using gradient descent algorithm works well and approaches the Cramér Rao lower bound (CRLB).

In addition, the Fisher Information (FI) $I(\theta)$ is the expectation of the negative second derivative of the log-likelihood function $l(\boldsymbol{\eta};\theta)$ (6) with respect to θ, i.e. (7). The expectation $\mathrm{E}_{\boldsymbol{\eta}}[\eta_i/p_i-(1-\eta_i)/(1-p_i)]=0$ is used in (7). Consequently, the CRLB is

$$\mathrm{CRLB}(\theta) = 1/I(\theta), \tag{8}$$

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

$$I(\theta) = -\frac{N(1-q_0-q_1)}{\sigma_w}\sum_{i=1}^{K} h_i \left\{ f_w\left(\frac{h_i\theta-\tau_i}{\sigma_w}\right) \mathrm{E}_{\boldsymbol{\eta}}\left[\frac{\partial}{\partial\theta}\left(\frac{\eta_i}{p_i} - \frac{1-\eta_i}{1-p_i}\right)\right] + \frac{\partial}{\partial\theta} f_w\left(\frac{h_i\theta-\tau_i}{\sigma_w}\right) \mathrm{E}_{\boldsymbol{\eta}}\left[\frac{\eta_i}{p_i} - \frac{1-\eta_i}{1-p_i}\right] \right\} = \frac{N(1-q_0-q_1)^2}{\sigma_w^2}\sum_{i=1}^{K} \frac{h_i^2 f_w^2\left(\frac{h_i\theta-\tau_i}{\sigma_w}\right)}{p_i(1-p_i)} \quad (7)$$

which is later used as a benchmark performance for ML estimation from labeled data in Section VI.

B. GLRT detection

Given θ is known under the alternative hypothesis $\mathcal{H}_1$, the optimal detector according to the NP criterion is the log-likelihood ratio test [26, p. 65, Th. 3.1]. Without the knowledge of amplitude, the GLRT is usually adopted. Although there is no optimality associated with the GLRT, it appears to work well in many scenarios of practical interest [26, p. 200]. The GLRT replaces the unknown parameter by its maximum likelihood estimate (MLE) and decides $\mathcal{H}_1$ if

$$T_1(\boldsymbol{\eta}) = \max_{\theta\in[-\Delta,\Delta]} l(\boldsymbol{\eta};\theta) - l(\boldsymbol{\eta};0) > \gamma, \quad (9)$$

where γ is a threshold determined by the given false alarm probability P_{FA}.

IV. Estimation from unlabeled data

In this section, we study the estimation problem from unlabeled data, namely we suppose that $\mathcal{H}_1$ is in force, that the FC receives the unlabeled set of data $\{\tilde{u}_{ij}\}_{j=1}^N, i = 1,\ldots,K$, and the problem is to estimate θ. In the following sections we first introduce the ML estimation from unlabeled data. Next we decompose the original problem into two subproblems, i.e., permutation matrix recovery problem with knowledge of amplitude and θ-estimation problem from labeled data. The first subproblem is important because $\mathbf{\Pi}$ plays the role of a nuisance parameter both in the estimation case addressed here and in the detection problem considered in Section V. We study the subproblems separately and then address the original θ-estimation problem. We also pay attention to the identifiability of the estimation problem and to the initializations scheme of the estimation algorithm.

A. Maximum likelihood estimation

Introduce the function $\pi(\cdot)$ such that $m = \pi(i)$ if the permutation matrix $\mathbf{\Pi}$ in (2) maps the ith row of $\mathbf{U}$ to the mth row of $\tilde{\mathbf{U}}$. The PMF of $\tilde{\mathbf{U}}$ is

$$p(\widetilde{\mathbf{U}};\theta,\mathbf{\Pi}) = \prod_{m=1}^{K}\prod_{j=1}^{N} \Pr(\tilde{u}_{mj}=1)^{\tilde{u}_{mj}} \Pr(\tilde{u}_{mj}=0)^{(1-\tilde{u}_{mj})}$$
$$= \prod_{i=1}^{K}\prod_{j=1}^{N} \Pr(\tilde{u}_{\pi(i)j}=1)^{\tilde{u}_{\pi(i)j}} \Pr(\tilde{u}_{\pi(i)j}=0)^{(1-\tilde{u}_{\pi(i)j})}, \quad (10)$$

where $(\Pr(\tilde{u}_{ij}=1), \Pr(\tilde{u}_{ij}=0))$ is the PMF of $\tilde{u}_{ij}$. The corresponding log-likelihood function $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})$ is

$$l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi}) = N\sum_{i=1}^{K}\left(\tilde{\eta}_{\pi(i)}\log p_i + (1-\tilde{\eta}_{\pi(i)})\log(1-p_i)\right), \quad (11)$$

where $\tilde{\eta}_{\pi(i)} = \sum_{j=1}^N \tilde{u}_{\pi(i)j}/N = \sum_{j=1}^N \tilde{u}_{mj}/N$. The ML estimation problem can be formulated as

$$\max_{\theta\in[\Delta,\Delta],\mathbf{\Pi}\in\mathcal{P}_K} l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi}), \quad (12)$$

where $\mathcal{P}_K$ denotes the set of all possible $K\times K$ permutation matrices. Please, note that θ is the unknown parameter we are interested in, while $\mathbf{\Pi}$ acts as a nuisance parameter.

Remark 1 *The function (11) can also be written as the KL divergence between the empirical $\tilde{p}_i = \tilde{\eta}_{\pi(i)}$ and p_i, where $\tilde{\eta}_{\pi(i)}$ denotes the fraction of $\tilde{u}_{\pi(i)j} = 1$ in $\{\tilde{u}_{\pi(i)j}\}_{j=1}^N$. The KL divergence between the empirical $\tilde{p}_i = \tilde{\eta}_{\pi(i)}$ and p_i for each i can be calculated and turns out to be*

$$D(\tilde{\boldsymbol{\eta}}||\mathbf{p}) = \sum_{i=1}^{K}\left(\tilde{\eta}_{\pi(i)}\log\frac{\tilde{\eta}_{\pi(i)}}{p_i} + (1-\tilde{\eta}_{\pi(i)})\log\frac{1-\tilde{\eta}_{\pi(i)}}{1-p_i}\right)$$
$$= -\sum_{i=1}^{K}\left(\tilde{\eta}_{\pi(i)}\log p_i + (1-\tilde{\eta}_{\pi(i)})\log(1-p_i) + \mathrm{const}\right)$$
$$= -l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi}) + \mathrm{const},$$

where const *denotes the constant terms. As a consequence, maximizing the log-likelihood function is equivalent to minimize the KL divergence $D(\tilde{\boldsymbol{\eta}}||\mathbf{p})$, where one looks for the optimal permutations such that the corresponding empirical probability distribution is the most similar to the true distribution $\mathbf{p}$ [27].*

B. Signal amplitude estimation from permuted data

For problem (12), both the permutation matrix $\mathbf{\Pi}$ and the desired parameter θ under $\mathcal{H}_1$ are unknown. In order to estimate the unknown θ, the permutation matrix plays the role of a nuisance parameter. Thus, the θ-estimation problem can be formalized as the joint estimation of θ and $\mathbf{\Pi}$. However, finding the best permutation matrix is very challenging in most problems due to non-convexity. One method is to enumerate all the possible permutation matrices, which leads to a computation complexity of $O(K!)$. To reduce the computation complexity, we decompose the joint optimization problem into the following two subproblems and optimize them alternately.

1) Permutation matrix recovery with knowledge of amplitude: The first subproblem is to estimate $\mathbf{\Pi}$ with the knowledge of amplitude, i.e., the amplitude θ is fixed under the alternative hypothesis $\mathcal{H}_1$. With reference to the smart grid example mentioned at the end of Section II, the case of known

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

θ and $\{h_i\}_{i=1}^{K}$ corresponds to the situation in which the nodal voltage magnitudes and phases are observable by PMUs at all the generator nodes [1]. The following proposition shows that the permutation matrix can be recovered efficiently by simple sorting, which costs $O(K \log K)$.

Proposition 1 *Assume $\mathcal{H}_1$ true. Given that the amplitude θ is known, the ML estimation problem (12) reduces to a permutation matrix recovery problem. The computation of optimal permutation matrix $\mathbf{\Pi}$ will reorder the rows of $\widetilde{\mathbf{U}}$, equivalently the elements of $\tilde{\boldsymbol{\eta}}$, to have the same relative order as the elements of $(1-q_0-q_1)(\mathbf{h}\theta-\boldsymbol{\tau})$.*

Proof: The objective function $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})$ (11) can be decomposed as

$$l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi}) = K\sum_{i=1}^{N} \tilde{\eta}_{\pi(i)} s_i + K\sum_{i=1}^{N} \log(1-p_i), \qquad (13)$$

where $s_i = \log(p_i/(1-p_i))$. From (13), we see that the computation of the optimal permutation matrix amounts to reordering the rows of $\mathbf{U}$, equivalently the elements of $\tilde{\boldsymbol{\eta}}$, to have the same relative order as the elements of $\mathbf{s}$ [15, Lemma 1], [17, Proposition 1]. Because s_i is monotonically increasing with respect to $(1-q_0-q_1)(h_i\theta-\tau_i)$, the elements of $\tilde{\boldsymbol{\eta}}$ should be reordered by the permutation matrix to have the same relative order as the elements of $(1-q_0-q_1)(\mathbf{h}\theta-\boldsymbol{\tau})$ to maximize the likelihood. ■

If $\tau_i = c_0 h_i$, changing $\theta - c_0$ to $-(\theta-c_0)$ would reverse the ordering. This might help to explain why two solutions appear in the subsequent Proposition 2 when θ is unknown under $\mathcal{H}_1$.

2) Estimation of θ from labeled data: The second subproblem is to estimate θ with given $\mathbf{\Pi}$, which is equivalent to estimation from labeled data, as discussed in Section IV. In this settings, one obtains the ML estimate of θ via numerical algorithms and achieves global optimum, provided that $q_0 = q_1 = 0$ or 1. Under the assumption $0 < q_0 + q_1 < 2$, we do not know whether the negative log-likelihood function is convex, and only a local optimum is guaranteed.

3) Alternating maximization algorithm: Now we optimize the two subproblems alternately as shown in Algorithm 1. The alternating maximization in Algorithm 1 can be viewed

Algorithm 1 Alternating Maximization

1: Initialize $t=1$ and $\hat{\theta}_{t-1}$;
2: Fix $\theta=\hat{\theta}_{t-1}$, reorder $\tilde{\boldsymbol{\eta}}$ according to $(1-q_0-q_1)(\mathbf{h}\theta-\boldsymbol{\tau})$ and obtain the corresponding permutation matrix $\hat{\mathbf{\Pi}}_{t-1}$;
3: Solve $\max_{\theta} l(\tilde{\boldsymbol{\eta}};\theta,\hat{\mathbf{\Pi}}_{t-1})$ and obtain $\hat{\theta}_t$;
4: Set $t=t+1$ and return to step 2 until a sufficient number of iterations has been performed or $|\hat{\theta}_t-\hat{\theta}_{t-1}| \le \epsilon$, where ϵ is a tolerance parameter.

as the alternating projection with respect to θ and $\mathbf{\Pi}$. The objective function is $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})$. In step 2, given $\hat{\theta}_{t-1}$, we update the permutation matrix as $\hat{\mathbf{\Pi}}_{t-1}$, and the objective value is $l(\tilde{\boldsymbol{\eta}};\hat{\theta}_{t-1},\hat{\mathbf{\Pi}}_{t-1})$. Given $\hat{\mathbf{\Pi}}_{t-1}$, we obtain ML estimation of θ as $\hat{\theta}_t$, and the objective value is $l(\tilde{\boldsymbol{\eta}};\hat{\theta}_t,\hat{\mathbf{\Pi}}_{t-1})$ satisfying $l(\tilde{\boldsymbol{\eta}};\hat{\theta}_t,\hat{\mathbf{\Pi}}_{t-1}) \ge l(\tilde{\boldsymbol{\eta}};\hat{\theta}_{t-1},\hat{\mathbf{\Pi}}_{t-1})$. Given $\hat{\theta}_t$, we update the permutation matrix as $\hat{\mathbf{\Pi}}_t$, and the objective value is $l(\tilde{\boldsymbol{\eta}};\hat{\theta}_t,\hat{\mathbf{\Pi}}_t)$ satisfying $l(\tilde{\boldsymbol{\eta}};\hat{\theta}_t,\hat{\mathbf{\Pi}}_t) \ge l(\tilde{\boldsymbol{\eta}};\hat{\theta}_t,\hat{\mathbf{\Pi}}_{t-1})$. Consequently, we have

$$l(\tilde{\boldsymbol{\eta}};\hat{\theta}_t,\hat{\mathbf{\Pi}}_t) \ge l(\tilde{\boldsymbol{\eta}};\hat{\theta}_{t-1},\hat{\mathbf{\Pi}}_{t-1}) \qquad (14)$$

and the following claim.

Claim 1 *Suppose that the maximum with respect to each θ and $\mathbf{\Pi}$ is unique, and consider the relaxed version of the optimization problem (12) in which the set $\mathcal{P}_K$ is replaced by the set of the doubly stochastic matrices with entries in $[0,1]$. Then, it can be shown that problem (12) and its relaxed version have the same solutions, and any limit point produced by the alternating maximization algorithm for the relaxed problem is a stationary point.*

Proof: The proof is based on [28, pp. 268-269] and is provided as supplemental material to this paper. ■

Furthermore, we find a special case in which signal amplitude estimation problem (12) can be efficiently solved. From the special case, it can be seen that the optimal solution can be obtained under some circumstances. The special case is detailed in Appendix VIII-A, and the corresponding Algorithm 2 is also provided.

C. An example for model unidentifiability

Algorithm 2 may generate two solutions $(\hat{\theta}_{s1},\hat{\mathbf{\Pi}}_{s1})$ and $(\hat{\theta}_{s2},\hat{\mathbf{\Pi}}_{s2})$. Given system parameters $\mathbf{h}$ and $\boldsymbol{\tau}$, it is important to determine whether the two solutions $(\hat{\theta}_{s1},\hat{\mathbf{\Pi}}_{s1})$ and $(\hat{\theta}_{s2},\hat{\mathbf{\Pi}}_{s2})$ will yield the same log-likelihood $l(\tilde{\boldsymbol{\eta}};\hat{\theta}_{s1},\hat{\mathbf{\Pi}}_{s1}) = l(\tilde{\boldsymbol{\eta}};\hat{\theta}_{s2},\hat{\mathbf{\Pi}}_{s2})$. If $l(\tilde{\boldsymbol{\eta}};\hat{\theta}_{s1},\hat{\mathbf{\Pi}}_{s1}) = l(\tilde{\boldsymbol{\eta}};\hat{\theta}_{s2},\hat{\mathbf{\Pi}}_{s2})$, two parameter values lead to the same maximum likelihood. In this situation, θ clearly can not be estimated consistently since $\tilde{\boldsymbol{\eta}}$ provides no information as to whether the true value is $\hat{\theta}_{s1}$ or $\hat{\theta}_{s2}$. This phenomenon motivates us delving into the identifiability of the model. Statistical identifiability is a property of a statistical model which describes one-to-one correspondence between parameters and probability distributions [30, pp. 456-457]. In this subsection, we provide the following proposition which justifies that there exist cases in which the model is unidentifiable, i.e., there exist two different parameter values θ_{s1} and θ_{s2} leading to the same distribution of the observations $\tilde{\boldsymbol{\eta}}$ [30].

Proposition 2 *Let $\mathbf{h}_a$ and $\mathbf{h}_d$ denote the ascending and descending ordered versions of $\mathbf{h}$, and $\mathbf{\Pi}_a\mathbf{h} = \mathbf{h}_a$ and $\mathbf{\Pi}_d\mathbf{h} = \mathbf{h}_d$, where $\mathbf{\Pi}_a$ and $\mathbf{\Pi}_d$ are permutation matrices. Given $\boldsymbol{\tau} = c_0\mathbf{h}$ and $\mathbf{h}_a = -\mathbf{h}_d$, the model is unidentifiable, i.e., $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s1},\mathbf{\Pi}=\mathbf{\Pi}_{s1}} = l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s2},\mathbf{\Pi}=\mathbf{\Pi}_{s2}}$, where $\theta_{s2} = 2c_0 - \theta_{s1}$ and $\mathbf{\Pi}_{s2} = \mathbf{\Pi}_{s1}\mathbf{\Pi}_a^{\mathrm{T}}\mathbf{\Pi}_d$.*

Proof: Let $\mathbf{\Pi}_{s1}$ be a permutation matrix such that $\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}$ has the same relative order as $\mathbf{h}$. Now we prove that $\mathbf{\Pi}_{s2}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}$ has the same relative order as $-\mathbf{h}$. Utilizing $\mathbf{h}_a = -\mathbf{h}_d = -\mathbf{\Pi}_d\mathbf{h}$ and $\mathbf{\Pi}_d\mathbf{\Pi}_d^{\mathrm{T}} = \mathbf{I}$, we obtain $\mathbf{\Pi}_d^{\mathrm{T}}\mathbf{h}_a = -\mathbf{h}$. Note that $\mathbf{\Pi}_{s2}^{\mathrm{T}}\tilde{\boldsymbol{\eta}} = \mathbf{\Pi}_d^{\mathrm{T}}\mathbf{\Pi}_a\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}$. Because $\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}$ has the same relative order as $\mathbf{h}$, $\mathbf{\Pi}_a\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}$ has the same relative order as $\mathbf{\Pi}_a\mathbf{h} = \mathbf{h}_a$, and $\mathbf{\Pi}_d^{\mathrm{T}}\mathbf{\Pi}_a\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}$ has the same relative order as $\mathbf{\Pi}_d^{\mathrm{T}}\mathbf{h}_a = -\mathbf{h}$.

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Next we prove that $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s1},\mathbf{\Pi}=\mathbf{\Pi}_{s1}} = l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s2},\mathbf{\Pi}=\mathbf{\Pi}_{s2}}$ holds. Because $\theta_{s2} = 2c_0 - \theta_{s1}$, we have

$$h_i\theta_{s1}-\tau_i = h_i(\theta_{s1}-c_0), \quad h_i\theta_{s2}-\tau_i = -h_i(\theta_{s1}-c_0). \quad (15)$$

By examining $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})$ (13) and utilizing $\mathbf{h}_a = -\mathbf{h}_d$, the second term of $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s1},\mathbf{\Pi}=\mathbf{\Pi}_{s1}}$ is equal to that of $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s2},\mathbf{\Pi}=\mathbf{\Pi}_{s2}}$. For the first term, note that given θ_{s1} and θ_{s2}, the corresponding $\mathbf{s}_1$ and $\mathbf{s}_2$ in (13) can be viewed as evaluating at $\mathbf{h}$ and $-\mathbf{h}$ according to (15), respectively. Because $\mathbf{h}_a = -\mathbf{h}_d$, we can conclude that $\mathbf{s}_1$ is a permutated version of $\mathbf{s}_2$. The first term of (13) can be expressed as either $(\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}})^{\mathrm{T}}\mathbf{s}_1$ or $(\mathbf{\Pi}_{s2}^{\mathrm{T}}\tilde{\boldsymbol{\eta}})^{\mathrm{T}}\mathbf{s}_2$. Because $(\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}})^{\mathrm{T}}$ and $\mathbf{s}_1$ have the same relative order as $\mathbf{h}$, and $(\mathbf{\Pi}_{s2}^{\mathrm{T}}\tilde{\boldsymbol{\eta}})^{\mathrm{T}}$ and $\mathbf{s}_2$ have the same relative order as $-\mathbf{h}$, one has $(\mathbf{\Pi}_{s1}^{\mathrm{T}}\tilde{\boldsymbol{\eta}})^{\mathrm{T}}\mathbf{s}_1 = (\mathbf{\Pi}_{s2}^{\mathrm{T}}\tilde{\boldsymbol{\eta}})^{\mathrm{T}}\mathbf{s}_2$. Thus $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s1},\mathbf{\Pi}=\mathbf{\Pi}_{s1}} = l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=\theta_{s2},\mathbf{\Pi}=\mathbf{\Pi}_{s2}}$. ■

Please notice that Proposition 2 provides an example instead of rigorous conditions on identifiability. For the odd signal such as the sinusoidal and sawtooth signals with proper fixed sampling frequencies, the ascending order $\mathbf{h}_a$ and descending order $\mathbf{h}_d$ of the signal shape $\mathbf{h}$ satisfies $\mathbf{h}_a = -\mathbf{h}_d$, and the model is unidentifiable in this scenario.

Now an example is presented to substantiate the above proposition. Let $c_0 = 0.5$, the true value $\theta_0 = 1$, $\mathbf{h} = [2,-1,-2,1]^{\mathrm{T}}$, $\boldsymbol{\eta} = [\eta_1,\eta_2,\eta_3,\eta_4]^{\mathrm{T}}$ and $\mathbf{\Pi}_0 = [0\ 0\ 1\ 0; 0\ 1\ 0\ 0; 0\ 0\ 0\ 1; 1\ 0\ 0\ 0]$. Then $\tilde{\boldsymbol{\eta}} = [\eta_3,\eta_2,\eta_4,\eta_1]^{\mathrm{T}}$, $\mathbf{h}_a = [-2,-1,1,2]^{\mathrm{T}}$, $\mathbf{h}_d = [2,1,-1,-2]^{\mathrm{T}}$ and $\mathbf{h}_a = -\mathbf{h}_d$. We can conclude that $l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=1,\mathbf{\Pi}=\mathbf{\Pi}_0} = l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi})|_{\theta=0,\mathbf{\Pi}=\mathbf{\Pi}'}$, where $\mathbf{\Pi}' = [0\ 0\ 0\ 1; 1\ 0\ 0\ 0; 0\ 0\ 1\ 0; 0\ 1\ 0\ 0]$.

Proposition 2 shows that two different values of θ may give the same value of the likelihood, which implies that the estimation of θ may not be consistent. In terms of detection, if $\theta = 0$ and $\theta = \widetilde{\theta} \neq 0$ give the same GLRT metric, then the detection performance may degrade significantly. In addition, if $|c_0| \geq \Delta$, only one of $\{\theta_{s1},\theta_{s2}\}$ lies in the interval $[-\Delta,\Delta]$, and the model is identifiable.

D. Good initial points

For the alternating maximization algorithm dealing with nonconvex optimization problems, an initializations scheme is important for the algorithm to converge to the global optimum. In the following text, we provide good initial points for the alternating maximization algorithm. The key idea is to obtain a coarse estimate of θ via matching the expected and actual number of ones in observations, and utilizing the orthogonal property of permutation matrix.

Suppose that the number of measurements K is large. Consequently, as the number of measurements tends to infinity, the law of large numbers (LLN) implies

$$\eta_i \xrightarrow{\mathrm{p}} q_0 + (1-q_0-q_1)F_w\left((h_i\theta-\tau_i)/\sigma_w\right), \quad (16)$$

where $\xrightarrow{\mathrm{p}}$ denotes convergence in probability. Given $\theta \in [-\Delta,\Delta]$, $-|h_i|\Delta - \tau_i \leq h_i\theta - \tau_i \leq |h_i|\Delta - \tau_i$. In the following text, we only deal with $q_0+q_1<1$ case. The case that $q_0+q_1>1$ is very similar and is omitted here. Define $l = \min_{i\in[1,\cdots,N]} (q_0 + (1-q_0-q_1)F_w((-|h_i|\Delta-\tau_i)/\sigma_w))$ and $u = \max_{i\in[1,\cdots,N]} (q_0 + (1-q_0-q_1)F_w((|h_i|\Delta-\tau_i)/\sigma_w))$. Then η_i should satisfy $l \leq \eta_i \leq u$. Let $\mathcal{I}_{l,u}(\tilde{\eta}_i)$ denotes the projection of $\tilde{\eta}_i$ onto the interval $[l,u]$. Note that this projection operation is needed because (16) is valid in the limit as K goes to infinity. From (16) one obtains

$$\mathbf{m} \triangleq \sigma_w F_w^{-1}\left((\mathcal{I}_{l,u}(\tilde{\boldsymbol{\eta}}) - q_0\mathbf{1}_N)/(1-q_0-q_1)\right) \xrightarrow{\mathrm{p}} \mathbf{\Pi}(\mathbf{h}\theta - \boldsymbol{\tau}).$$

Utilizing $\mathbf{\Pi}\mathbf{\Pi}^{\mathrm{T}} = \mathbf{I}$ yields

$$\mathbf{m}^{\mathrm{T}}\mathbf{m} \xrightarrow{\mathrm{p}} \mathbf{h}^{\mathrm{T}}\mathbf{h}\theta^2 - 2\boldsymbol{\tau}^{\mathrm{T}}\mathbf{h}\theta + \boldsymbol{\tau}^{\mathrm{T}}\boldsymbol{\tau}, \quad (17)$$

which is a quadratic equation in θ. Accordingly, using the asymptotic properties of $\mathbf{m}^{\mathrm{T}}\mathbf{m}$, one obtains (18) via inverting (17).

$$\theta_{1,2} = \frac{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{h}}{\mathbf{h}^{\mathrm{T}}\mathbf{h}} \pm \sqrt{\frac{\mathbf{m}^{\mathrm{T}}\mathbf{m} - \boldsymbol{\tau}^{\mathrm{T}}\boldsymbol{\tau}}{\mathbf{h}^{\mathrm{T}}\mathbf{h}} + \left(\frac{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{h}}{\mathbf{h}^{\mathrm{T}}\mathbf{h}}\right)^2}. \quad (18)$$

The above two solutions can be used for the alternating maximization algorithm as initial points. Finally, the optimum with larger likelihood is chosen as the MLE. In Section VI, to provide a fair comparison of the alternating maximization algorithm with good initial points, $-\Delta$ and Δ are used as two initial points, and we choose the solution whose likelihood is larger as the MLE.

The result of (18) is consistent with that of Proposition 2. Given that the conditions in Proposition 2 are satisfied, and substituting $\boldsymbol{\tau} = c_0\mathbf{h}$ into (18), the solutions are $\theta_1 = \theta$ and $\theta_2 = 2c_0 - \theta$.

V. Detection from unlabeled data

In this section, we study the detection problem from unlabeled data. In addition, we investigate the permutation matrix recovery probability.

A. Signal detection from permuted data with knowledge of amplitude

With the knowledge of amplitude θ, the statistical test can be formulated as

$$T_2(\tilde{\boldsymbol{\eta}}) = \max_{\mathbf{\Pi}\in\mathcal{P}_N} l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi}) - \max_{\mathbf{\Pi}\in\mathcal{P}_N} l(\tilde{\boldsymbol{\eta}};0,\mathbf{\Pi}) > \gamma. \quad (19)$$

As shown in Proposition 1, the computation of optimal permutation matrix corresponding to the first term in (19) corresponds to reordering the elements of $\tilde{\boldsymbol{\eta}}$ to have the same relative order as the elements of $(1-q_0-q_1)(\mathbf{h}\theta-\boldsymbol{\tau})$. Similarly, for the computation corresponding to the second term in (19), we reorder the elements of $\tilde{\boldsymbol{\eta}}$ to have the same order as that of $-(1-q_0-q_1)\boldsymbol{\tau}$.

B. Signal detection from permuted data without knowledge of amplitude

Without the knowledge of amplitude θ, an approximation of the GLRT decides $\mathcal{H}_1$ if

$$T_3(\tilde{\boldsymbol{\eta}}) = \max_{\theta,\mathbf{\Pi}\in\mathcal{P}_N} l(\tilde{\boldsymbol{\eta}};\theta,\mathbf{\Pi}) - \max_{\mathbf{\Pi}\in\mathcal{P}_N} l(\tilde{\boldsymbol{\eta}};0,\mathbf{\Pi}) > \gamma. \quad (20)$$

If $\mathbf{\Pi}$ were known, this statistics would be the GLRT. The qualification *approximate* refers to the fact that the maximization

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

involves also the nuisance parameter $\mathbf{\Pi}$. Algorithm 1 for joint estimation of θ and $\mathbf{\Pi}$ has been described in Section IV-B3. The performance of the test (20) will be evaluated by using Algorithm 1 for joint estimation of θ and $\mathbf{\Pi}$, see Section IV-B3 and Section V-A.

C. Approximations on permutation matrix recovery probability

In this subsection, we investigate the permutation matrix recovery probability problem. Since errors in permutation matrix recovery are more likely to happen in the relatively indistinguishable cases, the performances in terms of signal detection or estimation tasks may not be closely related to the recovery of permutation matrix. However, it is meaningful to extract the accurate timestamp information or sensors' identity information which corresponds to recovery of permutation matrix, as presented in the following.

It is difficult to obtain the permutation matrix recovery probability without the knowledge of amplitude. Instead, we assume that θ is known under the alternative hypothesis $\mathcal{H}_1$, and analyze the permutation matrix recovery probability in terms of the recovery algorithm provided in Proposition 1. Without loss of generality, we also assume that $q_0 + q_1 < 1$ in the following analysis. The case that $q_0 + q_1 > 1$ is similar and is omitted here. First, let p_i be ordered such that $p_{(1)} > p_{(2)} > \cdots > p_{(K)}$. From (3) we have $(h_i\theta - \tau_i)_{(1)} > (h_i\theta - \tau_i)_{(2)} > \cdots > (h_i\theta - \tau_i)_{(K)}$. Provided $q_0 + q_1 < 1$, the elements of $\tilde{\boldsymbol{\eta}}$ should be reordered according to the order of the elements of $\mathbf{h}\theta - \boldsymbol{\tau}$ in Proposition 1. Therefore the permutation matrix will be correctly recovered if and only if $\eta_{(1)} > \eta_{(2)} > \cdots > \eta_{(K)}$. Note that the subscripts of $(h_i\theta - \tau_i)_{(\cdot)}$ and $\eta_{(\cdot)}$ also correspond to the order of p_i, instead of the order of $h_i\theta - \tau_i$ or η_i.

Define E_i as the event such that $\eta_{(i)} > \eta_{(i+1)}$ and $\bar{E}_i$ as the corresponding complement event of E_i, namely $\eta_{(i)} \le \eta_{(i+1)}$. The probability that permutation matrix is recovered correctly can be written as

$$\Pr(\hat{\mathbf{\Pi}}_{\rm ML} = \mathbf{\Pi}_0) = \Pr(\eta_{(1)} > \cdots > \eta_{(K)}) = \Pr\left(\bigcap_{i=1}^{K-1} E_i\right)$$
$$=1 - \Pr\left(\bigcup_{i=1}^{K-1} \bar{E}_i\right) \ge 1 - \sum_{i=1}^{K-1} \Pr(\bar{E}_i),$$

where union bound $\Pr\left(\bigcup_{i=1}^{K-1} \bar{E}_i\right) \le \sum_{i=1}^{K-1} \Pr(\bar{E}_i)$ is utilized in (21). From (3), we have $u_{ij} \sim \mathcal{B}(1, p_i)$ and $N\eta_i = \sum_{j=1}^{N} u_{ij} \sim \mathcal{B}(N, p_i)$. When N is large, the De Moivre-Laplace theorem [29, pp. 49, equation (3-27)] implies that the distribution of η_i can be approximated by $\mathcal{N}(p_i, p_i(1-p_i)/N)$. As a consequence, $\eta_{(i)} - \eta_{(i+1)}$ is approximately distributed as $\mathcal{N}(p_{(i)} - p_{(i+1)}, p_{(i)}(1-p_{(i)})/N + p_{(i+1)}(1-p_{(i+1)})/N)$, and

$$\begin{aligned}
\Pr(\hat{\mathbf{\Pi}}_{\rm ML} = \mathbf{\Pi}_0) &\ge 1 - \sum_{i=1}^{K-1} \Pr(\bar{E}_i) \\
&= 1 - \sum_{i=1}^{K-1} \Pr(\eta_{(i)} - \eta_{(i+1)} \le 0) \\
&\approx 1 - \sum_{i=1}^{K-1} \Phi\left(\frac{-(p_{(i)} - p_{(i+1)})\sqrt{N}}{\sqrt{p_{(i)}(1-p_{(i)}) + p_{(i+1)}(1-p_{(i+1)})}}\right) \\
&\ge 1 - (K-1)\Phi\left(-t\sqrt{N}\right) \\
&\approx 1 - (K-1)\frac{1}{\sqrt{2\pi}t\sqrt{N}} e^{-t^2 N/2} \\
&= 1 - \frac{1}{\sqrt{2\pi}} e^{\ln(K-1) - \ln t - \frac{1}{2}\ln N - \frac{t^2}{2}N} \triangleq \Pr(K, N), \quad (21)
\end{aligned}$$

where $t = \min_{i=1,\cdots,K-1} \frac{v_i}{\sqrt{p_{(i)}(1-p_{(i)}) + p_{(i+1)}(1-p_{(i+1)})}}$, $v_i = p_{(i)} - p_{(i+1)}$ and the approximation $\Phi(-x) \approx \frac{1}{\sqrt{2\pi}x}\mathrm{e}^{-\frac{x^2}{2}}$ $(x \gg 0)$ is utilized.

Utilizing $p_{(i)}(1-p_{(i)}) + p_{(i+1)}(1-p_{(i+1)}) \le 1/2$, we define $\tilde{t}$ as

$$\tilde{t} = \min_{i=1,\cdots,K-1} v_i \le \frac{\sqrt{2}}{2}t. \quad (22)$$

We conjecture that $\tilde{t}$ is on the order of $K^{-\alpha}$, i.e., $\tilde{t} = O(K^{-\alpha})$, which means that there exists constant c_t such that

$$\tilde{t} \approx c_t K^{-\alpha}. \quad (23)$$

We will prove that $\mathbf{h}$ can be constructed such that $\tilde{t} = O(K^{-1})$ and $\tilde{t} = O(K^{-2})$ later. According to (22) and (23), the approximation $\Pr(K, N)$ (21) can be further simplified and relaxed as

$$\widetilde{\Pr}(K, N) = 1 - \frac{1}{2\sqrt{\pi}} e^{\ln(K-1) - \ln\tilde{t} - \frac{1}{2}\ln N - \tilde{t}^2 N} \quad (24)$$
$$\approx 1 - \frac{1}{2\sqrt{\pi}c_t} e^{(1+\alpha)\ln K - \frac{1}{2}\ln N - \frac{c_t^2}{K^{2\alpha}}N}. \quad (25)$$

From (25), the exponent $(1+\alpha)\ln K - \frac{1}{2}\ln N - \frac{c_t^2}{K^{2\alpha}}N$ of (25) must be far less than 0 for the recovery of permutation matrix. Given N is large, the term $-\frac{1}{2}\ln N$ is small compared to N. Thus $(1+\alpha)\ln K - \frac{c_t^2}{K^{2\alpha}}N < 0$ will ensure that the permutation matrix can be recovered in high probability. Simplifying $(1+\alpha)\ln K - \frac{c_t^2}{K^{2\alpha}}N < 0$ yields

$$N > \frac{(1+\alpha)}{c_t^2} K^{2\alpha} \ln K. \quad (26)$$

According to the definition of p_i (3), equations (22) and (23), $c_t \propto 1 - q_0 - q_1$. From (26), the number of quantizers $N_{\rm req}$ required for permutation matrix recovery probability satisfies

$$N_{\rm req} \propto 1/(1 - q_0 - q_1)^2. \quad (27)$$

From (27), one can conclude that the number of quantizers for permutation matrix recovery with high probability is $1/(1-q_0-q_1)^2$ times that of unflipped case where $q_0 = q_1 = 0$.

Now we give examples to illustrate $\tilde{t} = O(K^{-\alpha})$. In

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

these cases, the noise is assumed Gaussian such that $w_{ij} \sim \mathcal{N}(0, \sigma_w^2)$, and different $\mathbf{h}$ correspond to the shapes of ramp signal, Gaussian signal and sinusoidal signal respectively. For simplicity, we assume $\boldsymbol{\tau} = c\mathbf{h} (c < \theta)$ and

$$a \triangleq (\theta - c)/\sigma_w > 0. \tag{28}$$

First, let $\mathbf{h}$ be the shape of a ramp signal, i.e., $h_i = u - \frac{(u-l)(i-1)}{K-1} (u > |l|)$. We prove that $\tilde{t} = O(K^{-1})$ in Appendix VIII-B1. Second, let $\mathbf{h}$ be the shape of a random signal from a standard Gaussian PDF, i.e., $h_i \sim \mathcal{N}(0, 1)$. The numerical results under different a are shown in Fig. 2. Under $a = 1$, we prove that $\tilde{t} = O(K^{-2})$ in Appendix VIII-B2. Third, let $\mathbf{h}$ be the shape of a sinusoidal signal, i.e, $h_i = \sin(2\pi x_i)$ and $x_i \sim \mathcal{U}(0, 1)$. The numerical results under different a are shown in Fig. 3. Based on the proofs and the intuitive numerical results, we conjecture that $\tilde{t} = O(K^{-\alpha})$.

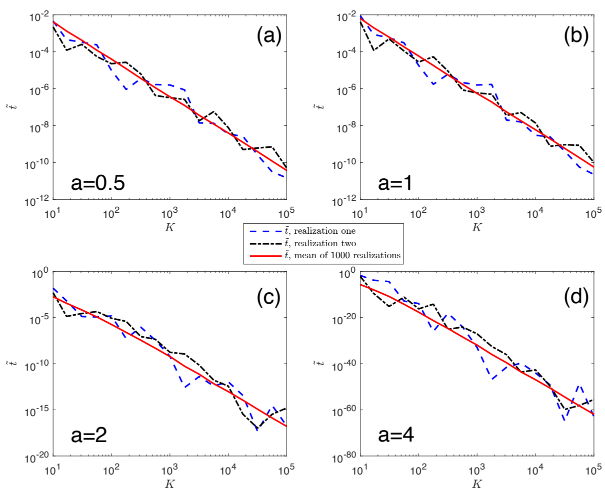

Fig. 2: The relationship of $\tilde{t}$ and K under different a. Note that $q_0 = q_1 = 0$, $h_i \sim \mathcal{N}(0, 1)$ and $w_{ij} \sim \mathcal{N}(0, \sigma_w^2)$.

VI. Numerical Simulations

In this section, numerical experiments are conducted to evaluate the theoretical results. For simplicity, the Gaussian distribution $\mathcal{N}(0, \sigma_w^2)$ is selected as the distribution of noise w_{ij}.

A. Parameter estimation

For the first two experiments, we evaluate the performance of the ML estimators proposed in Section IV. Parameters are set as follows: $K = 20$, $\theta = 1$, $\sigma_w^2 = 1$, $\Delta = 2$, $q_0 = 0.05$, $q_1 = 0.05$ and the tolerance parameter ϵ in Algorithm 1 is 10^{-7}. The number of Monte Carlo trials is 5000.

For the first experiment, the MSE performance of Algorithm 2 is evaluated in Fig. 4. We let $\boldsymbol{\tau} = 0.5\mathbf{h}$, which is a special case mentioned in Proposition 3. The coefficients $\mathbf{h}$ are equispaced with $\mathbf{h} = [-1.50, -1.29, -1.08, \cdots, 2.50]^{\mathrm{T}}$, which correspond to a ramp signal. It can be seen that $\mathbf{h}$ do not satisfy the condition in Proposition 2, thus the model may be identifiable. It can be seen that the ML estimator

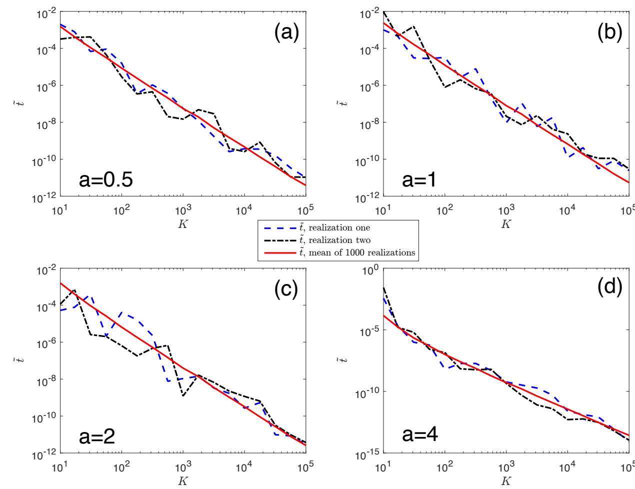

Fig. 3: The relationship of $\tilde{t}$ and K under different a. Note that $q_0 = q_1 = 0$, $h_i = \sin(2\pi x_i)$, $x_i \sim \mathcal{U}(0, 1)$ and $w_{ij} \sim \mathcal{N}(0, \sigma_w^2)$.

from labeled data always works well. Given limited number of quantizers, there is an obvious gap between the MSEs of two estimators. As the number of quantizers increases, the performance of the estimator from unlabeled data approaches that from labeled data. In addition, we provide a further comparison. We consider the extrema of the observations as an estimator of the unknown θ. Fig. 4 shows that Algorithm 2 performs better than the algorithm utilizing the extrema, especially when the number of quantizers N is less than 60, which is of interest in many realistic settings. As the number of quantizers increases, the algorithm utilizing the extrema becomes effective but still its MSE performance does not approach the CRLB.

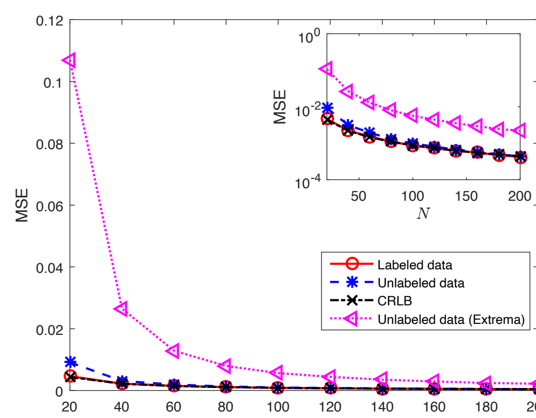

Fig. 4: MSE of θ vs. number of quantizers for the ML estimators from labeled and unlabeled data, compared with the CRLB (8) and the estimator using the extrema for a ramp signal.

For the second experiment, the MSE performance of Algorithm 1 (for the general case) is evaluated in Fig. 5. The

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

elements of the vector $\mathbf{h}$ describe the shape of a sinusoidal signal such that $h_i = \sin(2\pi x_i)$, where x_i is drawn independently and randomly from the uniform distribution $\mathcal{U}(0,1)$ and is then sorted in ascending order. The elements of the vector $\boldsymbol{\tau}$ are drawn independently and randomly from the uniform distribution $\mathcal{U}(-\Delta,\Delta)$. It can be seen that when $N<80$, good initial points improve the MSE performance of the alternating maximization algorithm from unlabeled data. As N increases to 80, the MSE performances of both unlabeled ML estimators approach a common level which is larger than that achieved by the labeled data. Finally, the MSEs of both estimators from unlabeled data approach to that from labeled data around $N=3\times 10^4$.

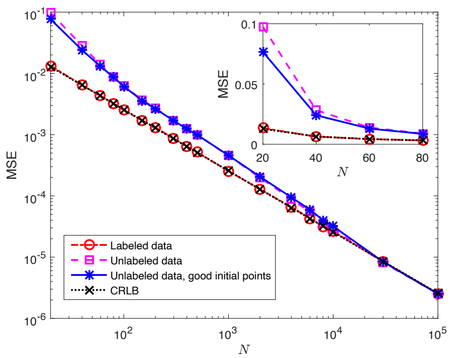

Fig. 5: MSE of θ vs. number of quantizers for the three ML estimators from labeled data, unlabeled data via initial points $\pm\Delta$ and unlabeled data via good initial points (18), compared with the CRLB (8) for a sinusoidal signal.

B. Signal detection

In Fig. 6, the relationship between P_D and the number of quantizers N is employed. Parameters are consistent with the first experiment, except that $\sigma_w^2=9$ and $P_{FA}=0.05$.

In subgraph (a), $\mathbf{h}$ and $\boldsymbol{\tau}$ are the same as those in the first experiment. It can be seen that the number of quantizers has a significant effect on the detection probability. As N increases, the performances of all the detectors increase, and the detection performance of the approximation of the GLRT (20) approaches the GLRT (9). In subgraph (b), $\mathbf{h}$ and $\boldsymbol{\tau}$ are the same as those in the second experiment, and the similar phenomena are observed. It seems that in this case little is gained by the good initializations scheme.

C. Permutation matrix recovery

In this subsection, the approximations for permutation matrix recovery are verified. Parameters are set as follows: $K=20$, $\theta=1.5$, $\Delta=2$, $q_0=0$, $q_1=0$ and $\sigma_w^2=1$. The number of Monte Carlo trials is 1000.

First, the relationship of t and $\tilde{t}$ (22) and the conjecture of $\tilde{t}$ (23) are illustrated in three cases. From Fig. 7, one obtains that t can be approximated as $\sqrt{2\tilde{t}}$ in practice. For a ramp signal, $\mathbf{h}=[-0.800,-0.705,-0.610,\cdots,1.000]^{\mathrm{T}}$ and $\boldsymbol{\tau}=0.5\mathbf{h}$. $t\approx\sqrt{2}c_e/K$ where $c_t=c_e=0.4355$ is evaluated via (30).

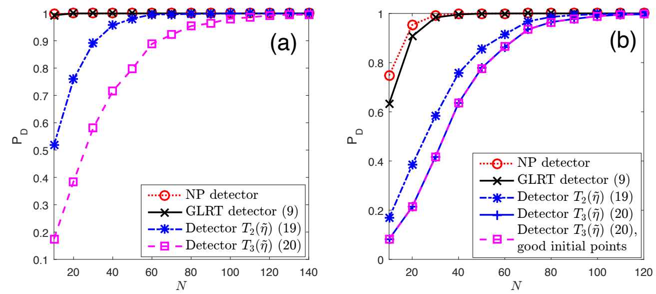

Fig. 6: P_d vs. number of quantizers N for a ramp signal in subgraph (a) and a sinusoidal signal in subgraph (b).

Because of the gap between t and $\sqrt{2}c_e/K$, we use linear regression to fit t and obtain $c_{ea}=0.6717$, which is much more accurate than c_e and will be utilized later to predict the number of quantizers for permutation matrix recovery. For random generated $\mathbf{h}$, $\mathbf{h}$ is drawn from standard normal distribution and $\boldsymbol{\tau}=0.5\mathbf{h}$. It can be seen that t can be approximated by $1/K^2$. For a sinusoidal signal, $\mathbf{h}$ and $\boldsymbol{\tau}$ are drawn in the same way as the second experiment. We use linear regression and obtain $t\approx 0.71/K^{2.23}\approx\sqrt{2\tilde{t}}=\sqrt{2}c_{t,s}/K^{\alpha_{t,s}}$, $c_{t,s}=0.5020$ and $\alpha_{t,s}=2.23$.

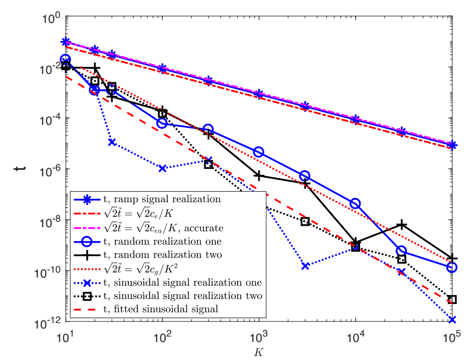

Fig. 7: The relationship of t and K, including equispaced, randomly generated and sinusoidal $\mathbf{h}$ cases.

Next, the empirical permutation recovery probability $\Pr(\hat{\boldsymbol{\Pi}}=\boldsymbol{\Pi}_0)$ versus N or K are presented in Fig. 8, and the theoretical approximations $\Pr(K,N)$ (21) and $\widetilde{\Pr}(K,N)$ (24) are plotted for comparison. In subgraph (a), (b) and (c), we set $K=20$. While in subgraph (d), we set $N=10^4$. All $\{h_i\}_{i=1}^K$ are drawn in the same way as the second experiment. We also evaluate the empirical permutation matrix recovery probability from permuted data without the knowledge of amplitude, which has negligible difference compared to that with the knowledge of amplitude.

In subgraph(a), it can be seen that the permutation matrix of the ramp signal can be recovered with high probability given $N\geq 5000$. From $N>\frac{1+\alpha}{c_t^2}K^{2\alpha}\ln K$ (26) where $c_t=c_e=0.4355$ and $\alpha=1$, one can conclude that $N>\frac{2}{0.4355^2}K^2\ln K|_{K=20}\approx 12636$, which is more than twice of 5000. Utilizing the fitted parameter c_{ea}, one obtains a more accurate result that $N>\frac{2}{0.6717^2}K^2\ln K|_{K=20}\approx 5312$ ensures permutation matrix recovery with high probability.

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

For random $\mathbf{h}$, $N > 3K^4 \ln K|_{K=20} \approx 1.438 \times 10^6$ ensures recovery with high probability, which is not accurate enough, as subgraph(b) shows that $N \approx 10^5$ is enough for recovery of permutation matrix. In subgraph(c), it is shown that $N \approx 10^6$ is enough for recovery of permutation matrix, which is also inaccurate compared to the fitted results of the sinusoidal signal $N > \frac{3.23}{0.5020^2} K^{4.46} \ln K|_{K=20} \approx 2.437 \times 10^7$. The numerical results show that the theoretical bound $\Pr(K,N)$ is accurate in predicting N with high probability in permutation matrix recovery, which demonstrates that $\widetilde{\Pr}(K,N)$ may be too conservative in predicting the number of quantizers ensuring perfect permutation matrix recovery. In subgraph(d), $10000 = N > \frac{2}{0.6717^2} K^2 \ln K|_{K=26} \approx 9763$. Thus $K \leq 26$ will ensure permutation matrix recovery with high probability, which is consistent with the numerical results.

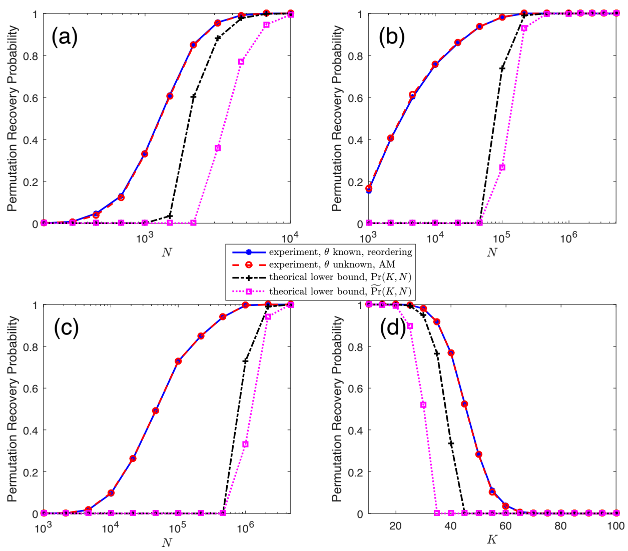

Fig. 8: $\Pr(\hat{\mathbf{\Pi}} = \mathbf{\Pi}_0)$ vs. N or K for the ramp signal in subgraph $(a)(d)$, random generated $\mathbf{h}$ in subgraph (b) and the sinusoidal signal in subgraph (c). $\Pr(K,N)$ and $\widetilde{\Pr}(K,N)$ are evaluated via (21) and (24), respectively.

In Fig. 9, the relationship between flipping probabilities (q_0, q_1) and number of quantizers N_{req} (27) required for permutation matrix recovery with high probability is verified. Parameters are the same as those in Fig. 8-(a) except for the values of (q_0, q_1). We use the results of the experiment in which $q_0 = q_1 = 0$ to predict those in which $q_0 = q_1 = 0.05$, $q_0 = q_1 = 0.1$ and $q_0 = q_1 = 0.15$, and plot the experimental results for comparison. It can be seen that the predictions are basically consistent with the experimental results, which verifies (27).

VII. Conclusion

We study a scale parameter estimation and signal detection problem from unlabeled quantized data for a canonical (known signal shape) sensing model. A sufficient condition under which the signal amplitude estimation problem can be solved efficiently is provided. It is also shown that in some settings the model can even be unidentifiable. Given that the number of quantizers is limited, the performance of the unlabeled estimator via reordering and alternating maximization algorithms is good, although there is a gap between the performances of labeled and unlabeled ML estimators. In addition, good initial points are provided to improve the performance of an alternating maximization algorithm for general estimation problems. As the number of quantizers increases, the performance of the unlabeled estimator approaches that of the labeled estimator due to the recovery of permutation matrix.

Furthermore, the performance of the approximation of GLRT detector is evaluated, and the numerical results show that the performance degradation of the approximation of GLRT detector is significant in noisy environments, compared to the GLRT given that the number of quantizers is small. As the number of quantizers increases, the performance of the approximation of GLRT approaches the GLRT. The explicit approximated permutation matrix recovery probability predicts that in order to find the true label of K time indexes, the number of quantizers N should be on the order of $K^{2\alpha} \log K$, where α is a constant depending on the signal shape and the distribution of noise.

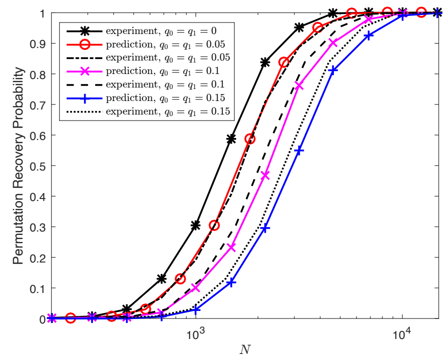

Fig. 9: $\Pr(\hat{\mathbf{\Pi}} = \mathbf{\Pi}_0)$ vs. number of quantizers N for a ramp signal under different flipping probabilities (q_0, q_1).

VIII. Appendix

A. Special cases for efficient recovery of $\mathbf{\Pi}$

Proposition 3 *In the problem (12), if there exist constants $c, d, e \in \mathbb{R}$ such that $c\boldsymbol{\tau} + d\mathbf{h} = e\mathbf{1}$, the elements of $\tilde{\boldsymbol{\eta}}$ should be reordered according to the order of the elements of $(q_0 + q_1 - 1)\boldsymbol{\tau}$ if $c = 0$, otherwise reordered according to $\mathbf{h}$ or $-\mathbf{h}$.*

Proof: We separately address the cases $c = 0$ and $c \neq 0$. In the case of $c = 0$, $\mathbf{h}$ must be a constant vector. Reordering according to $(1 - q_0 - q_1)(\mathbf{h}\theta - \boldsymbol{\tau})$ is equivalent to reordering according to $(q_0 + q_1 - 1)\boldsymbol{\tau}$. Since (q_0, q_1) are known in this problem, $\tilde{\boldsymbol{\eta}}$ should be reordered according to $\boldsymbol{\tau}$ if $q_0 + q_1 > 1$ or $-\boldsymbol{\tau}$ if $q_0 + q_1 < 1$. In the case of $c \neq 0$, we have $\boldsymbol{\tau} = (e/c)\mathbf{1} - (d/c)\mathbf{h}$. Consequently, $\mathbf{h}\theta - \boldsymbol{\tau} = (\theta + d/c)\mathbf{h} - (e/c)\mathbf{1}$, and $\tilde{\boldsymbol{\eta}}$ is reordered according to $\mathbf{h}$ or $-\mathbf{h}$. ■

The above proposition deals with four cases, i.e., $\mathbf{h}$ is a constant vector ($c = 0$), $\boldsymbol{\tau}$ is a constant vector ($d = 0$), $\mathbf{h}$ is a multiple of $\boldsymbol{\tau}$ ($e = 0$) and each pair of components of $\mathbf{h}$ and $\boldsymbol{\tau}$ lies in the same line $c\tau_i + dh_i = e$ ($cde \neq 0$). In [17] it is shown that reordering yields the optimal MLE

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

given $\mathbf{h} = \mathbf{1}$. Proposition 3 extends the special case in [17] to more general cases. Consequently, we propose Algorithm 2, an efficient algorithm for parameter estimation.

Algorithm 2 Reordering algorithm

1: If $c = 0$, reorder the elements of $\tilde{\boldsymbol{\eta}}$ according to the elements of $(q_0 + q_1 - 1)\boldsymbol{\tau}$. The corresponding permutation matrix is $\hat{\mathbf{\Pi}}_{s0}$. Solve the parameter estimation problem by numerical algorithm and obtain $\hat{\theta}_{\mathrm{ML}} = \underset{\theta}{\operatorname{argmax}}\ l(\tilde{\boldsymbol{\eta}}; \theta, \hat{\mathbf{\Pi}}_{s0})$;

2: If $c \neq 0$, reorder the elements of $\tilde{\boldsymbol{\eta}}$ according to the elements of $\mathbf{h}$ and $-\mathbf{h}$. The corresponding permutation matrices are $\hat{\mathbf{\Pi}}_{s1}$ and $\hat{\mathbf{\Pi}}_{s2}$;

3: Solve the single variable optimization problems and obtain $\hat{\theta}_{s1} = \underset{\theta}{\operatorname{argmax}}\ l(\tilde{\boldsymbol{\eta}}; \theta, \hat{\mathbf{\Pi}}_{s1})$ and $\hat{\theta}_{s2} = \underset{\theta}{\operatorname{argmax}}\ l(\tilde{\boldsymbol{\eta}}; \theta, \hat{\mathbf{\Pi}}_{s2})$. Choose $\hat{\theta}_{\mathrm{ML}} = \hat{\theta}_{s1}$ given that $l(\tilde{\boldsymbol{\eta}}; \hat{\theta}_{s1}, \hat{\mathbf{\Pi}}_{s1}) \geq l(\tilde{\boldsymbol{\eta}}; \hat{\theta}_{s2}, \hat{\mathbf{\Pi}}_{s2})$, otherwise $\hat{\theta}_{\mathrm{ML}} = \hat{\theta}_{s2}$.

B. Proof of two special cases $\tilde{t} = O(K^{-\alpha})$

1) $\tilde{t} = O(K^{-1})$: Let $\mathbf{h}$ be the shape of a ramp signal such that $h_i = u - \frac{(u-l)(i-1)}{K-1}(u > |l|)$, and $w_{ij} \sim \mathcal{N}(0, \sigma_w^2)$. Then the ordered sequence $p_{(i)} = p_i$, and $\tilde{t}$ can be approximated as

$$\begin{aligned}\tilde{t} =& \min_{i=1,\cdots,K-1} p_i - p_{i+1} \\ =& \frac{a(1-q_0-q_1)(u-l)}{K-1} \min_{i=1,\cdots,K-1} f_w(a\xi_i) \\ \approx& \frac{a(1-q_0-q_1)(u-l)f_w(au)}{K-1} \\ \approx& c_t K^{-1},\end{aligned} \tag{29}$$

where mean value theorem is utilized for $\xi_i \in (h_{i+1}, h_i)$, $\xi_1 \approx h_1 = u$ is utilized when K is large, and

$$c_t = a(1-q_0-q_1)(u-l)f_w(au). \tag{30}$$

Therefore $\tilde{t}$ can be reshaped in the form of (23).

2) $\tilde{t} = O(K^{-2})$: Let h_i be independently drawn from the same distribution of w_{ij}/σ_w. The CDF of p_i is

$$\begin{aligned}F_{p_i}(x) &= \Pr(p_i \leq x) \\ &= \Pr(q_0 + (1-q_0-q_1)F_w(ah_i) \leq x) \\ &= \Pr\left(h_i \leq \frac{1}{a}F_w^{-1}\left(\frac{x-q_0}{1-q_0-q_1}\right)\right) \\ &= F_w\left(\frac{1}{a}F_w^{-1}\left(\frac{x-q_0}{1-q_0-q_1}\right)\right).\end{aligned} \tag{31}$$

Now we prove that $\tilde{t} = O(K^{-2})$ under certain conditions. Given that h_i and w_{ij}/σ_w are i.i.d. random variables and $a = 1$, the CDF $F_{p_i}(x) = (x-q_0)/(1-q_0-q_1)$, and the PDF of p_i is

$$f_{p_i}(x) = \begin{cases} \frac{1}{1-q_0-q_1}, & q_0 \leq x \leq 1-q_1, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

Then the variates $p_{(1)}, p_{(2)} \cdots, p_{(K)}$ are distributed as K descending ordered statistics from an uniform $(q_0, 1-q_1)$ parent. For $x \leq (1-q_0-q_1)/(K-1)$, the CDF of $\tilde{t}$ can be derived as [31, p. 135, equation (6.4.3)]

$$\begin{aligned}F_{\tilde{t}}(x) =& \Pr\left(\min_{i=1,\cdots,K-1} v_i \leq x\right) \\ =& 1 - \Pr(v_1 > x, v_2 > x, \cdots, v_{K-1} > x) \\ =& 1 - \left[1 - \frac{(K-1)x}{1-q_0-q_1}\right]^K.\end{aligned} \tag{33}$$

For $x \geq (1-q_0-q_1)/(K-1)$, $F_{\tilde{t}}(x) = 1$. Then the PDF of $\tilde{t}$ is

$$f_{\tilde{t}}(x) = \begin{cases} \frac{K(K-1)}{1-q_0-q_1}\left[1 - \frac{(K-1)x}{1-q_0-q_1}\right]^{K-1}, & 0 \leq x \leq \frac{1-q_0-q_1}{K-1}, \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

The expectation of $\tilde{t}$ is

$$\begin{aligned}\mathrm{E}_{\tilde{t}}[\tilde{t}] =& \int_0^1 x f_{\tilde{t}}(x)\,\mathrm{d}x = \int_0^{\frac{1-q_0-q_1}{K-1}} \tilde{t} f_{\tilde{t}}(x)\,\mathrm{d}x \\ =& \frac{K(K-1)}{1-q_0-q_1}\int_0^{\frac{1-q_0-q_1}{K-1}} x\left[1 - \frac{(K-1)x}{1-q_0-q_1}\right]^{K-1} \mathrm{d}x \\ =& \frac{1-q_0-q_1}{K^2-1}.\end{aligned} \tag{35}$$

Hence the probability that $\tilde{t}$ falls into $[c_1/K^2, c_2/K^2]$ is

$$\begin{aligned}&\Pr(c_1/K^2 \leq \tilde{t} \leq c_2/K^2) \\ =& F_{\tilde{t}}(c_2/K^2) - F_{\tilde{t}}(c_1/K^2) \\ =& \left[1 - \frac{c_1(K-1)}{(1-q_0-q_1)K^2}\right]^K - \left[1 - \frac{c_2(K-1)}{(1-q_0-q_1)K^2}\right]^K.\end{aligned} \tag{36}$$

When K is large, $(K-1)/K \approx 1$ and $(1 - 1/(c'K))^{c'K} \approx 1/e (c' > 0)$. Equation (36) can be approximated as

$$\Pr(c_1/K^2 \leq \tilde{t} \leq c_2/K^2) \approx e^{-\frac{c_1}{1-q_0-q_1}} - e^{-\frac{c_2}{1-q_0-q_1}}. \tag{37}$$

Provided that $q_0 = q_1 = 0$, when $c_1 = 0.1$ and $c_2 = 10$, $\Pr(0.1/K^2 \leq \tilde{t} \leq 10/K^2) \approx 0.94$; when $c_1 = 0.01$ and $c_2 = 100$, $\Pr(0.01/K^2 \leq \tilde{t} \leq 100/K^2) \approx 0.99$. It can be seen that $\tilde{t}$ falls near the order of magnitude of K^{-2} with high probabilities. Thus it is reasonable that $\tilde{t} = O(K^{-2})$.

References

[1] P. Pradhan, K. Nagananda, P. Venkitasubramaniam, S. Kishore and R. S. Blum, "GPS spoofing attack characterization and detection in smart grids," *Communications and Network Security (CNS), 2016 IEEE Conference on*, pp. 391-395, 2016.

[2] T. E. Humphreys, B. M. Ledvina, M. L. Psiaki, B. W. O'hanlon and P. M. Kintner, "Assessing the spoofing threat: development of a portable GPS civilian spoofer," in *Proc. Int. Tech. Meet. Satellite Div. The Ins. Navigation*, pp. 2314-2325, 2008.

[3] Q. Zeng, H. Li and L. Qian, "GPS spoofing attack on time synchronization in wireless networks and detection scheme design," in *MILCOM'12*, pp. 1-5, 2012.

[4] Z. Zhang, S. Gong, A. D. Dimitrovski and H. Li, "Time synchronization attack in smart grid: Impact and analysis," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 87-98, 2013.

[5] S. Challa, R. J. Evans and X. Wang, "A Bayesian solution and its approximations to out-of-sequence measurement problems," *Information Fusion*, vol. 4, no. 3, pp. 185-199, 2003.

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

[6] L. Schenato, "Optimal estimation in networked control systems subject to random delay and packet drop," *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 1311-1317, 2008.
[7] L. M. Millefiori, P. Braca, K. Bryan and P. Willett, "Adaptive filtering of imprecisely time-stamped measurements with application to AIS networks," in *Proc. of the 18th Intern. Conf. on Inform. Fusion (FUSION)*, pp. 359-365, 2015.
[8] V. Emiya, A. Bonnefoy, L. Daudet and R. Gribonval, "Compressed sensing with unknown sensor permutation," *ICASSP*, pp. 1040-1044, 2014.
[9] J. Unnikrishnan, S. Haghighatshoar and M. Vetterli, "Unlabeled sensing: Solving a linear system with unordered measurements," *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on. IEEE*, pp. 786-793, 2015.
[10] A. Pananjady, M. J. Wainwright and T. A. Courtade, "Linear regression with an unknown permutation: statistical and computational limits,"*Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on. IEEE*, pp. 417-424, 2016.
[11] A. Pananjady, M. J. Wainwright and T. A. Courtade, "Denoising linear models with permutated data," http://arxiv.org/abs/1704.07461, 2017.
[12] A. Abid, A. Poon and J. Zou, "Linear regression with shuffled labels," http://arxiv.org/abs/1705.01342, 2017.
[13] S. Haghighatshoar and G. Caire, "Signal recovery from unlabeled samples," *ISIT*, pp. 451-455, 2017.
[14] L. Keller, M. J. Siavoshani, C. Fragouli and K. Argyraki, "Identity aware sensor networks," *Proceedings - IEEE INFOCOM*, pp. 2177-2185, 2009.
[15] S. Marano, V. Matta, P. Willett, P. Braca and R. S. Blum, "Hypothesis testing in the presence of Maxwell's daemon: Signal detection by unlabeled observations," *ICASSP*, pp. 3286-3290, 2017.
[16] P. Braca, S. Marano, V. Matta, P. Willett, "Asymptotic efficiency of the PHD in multitarget/multisensor estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 553-564, 2013.
[17] J. Zhu, H. Cao, C. Song and Z. Xu, "Parameter estimation via unlabeled sensing using distributed sensors," *IEEE commun. lett.*, vol. 21, no. 10, pp. 2130-2133, 2017.
[18] J. Unnikrishnan, S. Haghighatshoar, M. Vetterli, "Unlabeled sensing with random linear measurements," to appear in *IEEE Trans. Inf. Theory*, 2018.
[19] A. Pananjady, M. J. Wainwright and T. A. Courtade, "Linear regression with shuffled data: statistical and computational limits of permutation recovery," to appear in *IEEE Trans. Inf. Theory*, 2017.
[20] S. Haghighatshoar and G. Caire, "Signal recovery from unlabeled samples," *IEEE Trans. Signal Process.*, vol. 66, no.5, pp. 1242-1257, 2018.
[21] O. Ozdemir and P. K. Varshney, "Channel aware target location with quantized data in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 57, pp. 1190-1202, 2009.
[22] H. C. Papadopoulos, G. W. Wornell and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978-1002, Mar. 2001.
[23] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case,"*IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131-1143, 2006.
[24] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part II: unknown probability density function,"*IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2784-2796, 2006.
[25] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
[26] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, Englewood Cliffs, NJ: Prentice Hall, 1993.
[27] J. Unnikrishnan, D. Huang, S. P. Meyn, A. Surana and V. V. Veeravalli, "Universal and composite hypothesis testing via mismatched divergence," vol. 57, no. 3, pp. 1587-1603, Mar. 2011.
[28] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 1999.
[29] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes, Fourth Edition*, 2002.
[30] E. L. Lehmann, *Elements of large-sample theory*, Springer-Verlag New York, inc., 1999.
[31] H. A. David and H. N. Nagaraja, *Order Statistics: Third Edition*, John Wiley, 2003.

View publication stats