A Truth Discovery Approach with Theoretical Guarantee

Houping Xiao[1], Jing Gao[1], Zhaoran Wang[2], Shiyu Wang[3], Lu Su[1], and Han Liu[2]

[1]SUNY Buffalo, Buffalo, NY USA [2]Princeton University, Princeton, NJ USA
[3]University of Illinois, Urbana-Champaign, IL USA
{houpingx,jing}@buffalo.edu, zhaoran@princeton.edu
swang86@illinois.edu, lusu@buffalo.edu, hanliu@princeton.edu

ABSTRACT

In the information age, people can easily collect information about the same set of entities from multiple sources, among which conflicts are inevitable. This leads to an important task, *truth discovery*, i.e., to identify true facts (truths) via iteratively updating truths and source reliability. However, the convergence to the truths is never discussed in existing work, and thus there is no theoretical guarantee in the results of these truth discovery approaches. In contrast, in this paper we propose a truth discovery approach with theoretical guarantee. We propose a randomized gaussian mixture model (RGMM) to represent multi-source data, where truths are model parameters. We incorporate source bias which captures its reliability degree into RGMM formulation. The truth discovery task is then modeled as seeking the maximum likelihood estimate (MLE) of the truths. Based on expectation-maximization (EM) techniques, we propose population-based (i.e., on the limit of infinite data) and sample-based (i.e., on a finite set of samples) solutions for the MLE. Theoretically, we prove that both solutions are contractive to an ϵ-ball around the MLE, under certain conditions. Experimentally, we evaluate our method on both simulated and real-world datasets. Experimental results show that our method achieves high accuracy in identifying truths with convergence guarantee.

CCS Concepts

•Information systems → Data mining;

Keywords

Truth Discovery; Mixture Model; Asymptotic Consistency

1. INTRODUCTION

With the increase in our capabilities in collecting data from the physical world, an important feature of the data collection is its wide variety, i.e., data about the same object can be obtained from various sources. For example, cus-tomer information can be found from multiple databases in a company, a patient's medical records may be scattered at different hospitals, and product specifications are typically listed at different websits (e.g., Amazon.com, ShopZilla.com). Conflicts among information from different sources are commonly observed. Therefore, it is an important task to discover the truth (i.e., correct information) out of conflicting multi-source data, which is referred to as *truth discovery*.

A trivial approach of accomplishing the truth discovery task is to treat the average of the data as the truth. The drawback of this simple averaging approach is that it treats each source equally reliable, which is often violated in real practice. Usually there exist sources with low-quality information, such as faulty sensors that keep emanating erroneous information, or spammers who propagate false information on the Web. To address this challenge, researchers propose a variety of approaches that infer both source reliability and truths from multi-source data. These approaches, referred to as *truth discovery* approaches, are developed based on heuristic principles [3,5–8,13,16,24], optimization [9,10,14,27], or probabilistic models [2,4,12,15,17,19–21,25,26]. Despite the difference in the techniques, the underlying principle is the same: The sources which often provide truths should be reliable, and the information from reliable sources usually represent the truths. Based on this principle, existing truth discovery approaches usually start with an initialization of source reliability, and then conduct the following two steps iteratively until convergence:

(i) Based on source reliability, obtain the truth by assigning high weights to reliable sources in the aggregation of sources' inputs;

(ii) Based on the truths, calculate the reliability degree of each source by measuring the deviation between truths and sources' inputs.

Although these algorithms differ in the specific ways to compute truths or source reliability, they typically follow this iterative procedure. However, the convergence of such an iterative procedure has not been discussed in existing work, and there is no theoretical guarantee in the results of existing truth discovery approaches.

In contrast, in this paper we propose an effective truth discovery approach with theoretical guarantee in convergence. Compared with existing approaches, the proposed approach follows the same principle that reliable sources provide truths and truths are stated by reliable sources, but the major advantage of the proposed approach is its theoretical guarantee. Specifically, we introduce the bias of each source which is a

random variable measuring its reliability degree. To represent multi-source data with various reliability degrees, we propose a **R**andomized **G**aussian **M**ixture **M**odel (RGMM) formulation, which consists of a gaussian mixture model [18] with sources' biases incorporated. We cast the truth discovery problem as inferring a maximum likelihood estimator (MLE) of the unknown parameters in RGMM. To solve the problem, we derive both population-based and sample-based Expectation-Maximization (EM) solutions [4] for the limit of infinite data and a finite set of samples respectively. We name the two approaches as *population*- and *sample*-EM$_{\text{RGMM}}$ (i.e., EM for RGMM), respectively.

Theoretically, we prove that the output of the proposed EM$_{\text{RGMM}}$ approaches converge to a MLE of the truth as the number of sources increases. The proof is first derived for *population*-EM$_{\text{RGMM}}$. We first bound the distance between the *population*-EM$_{\text{RGMM}}$ and the MLE. In Property 1, we present the conditions under which *population*-EM$_{\text{RGMM}}$ estimate is contractive to the MLE, in an ϵ-ball around the MLE. Then to prove that the output of *sample*-EM$_{\text{RGMM}}$ converges to an ϵ-ball around MLE, we bound the deviation from *sample*-EM$_{\text{RGMM}}$ to *population*-EM$_{\text{RGMM}}$. To achieve this goal, we introduce the definition of the covering number [22] of a metric space, and show that the distance between *population*-EM$_{\text{RGMM}}$ and *sample*-EM$_{\text{RGMM}}$ is upper bounded by the number of sources (Corollary 1). Combining these theoretical results and applying triangle inequality, we prove that the distance between *sample*-EM$_{\text{RGMM}}$ and the MLE is bounded, under certain conditions.

Experimentally, we evaluate the effectiveness of the proposed *sample*-EM$_{\text{RGMM}}$ and verify the theoretical results on several simulated data as well as three real-world application datasets: Weather Forecast, Indoor Floorplan, and Stock data. Experimental results show that the proposed *sample*-EM$_{\text{RGMM}}$ approach are able to estimate truths from conflicting multi-source data. Compared with the state-of-the-art truth discovery algorithms, the proposed *sample*-EM$_{\text{RGMM}}$ can achieve comparable performance. Moreover, the estimate is asymptotically consistent to an ϵ-ball around the MLE under the stated conditions.,

The remainder of the paper is organized as follows. We discuss related work in Section 2. In Section 3, we mathematically formulate the task setting, introduce the randomized Gaussian mixture model, and derive the closed form EM algorithms for MLE. Then we present two important theoretical results for the convergence of the proposed approaches (Peoperties 1 and 2) in Section 4. The detailed proofs are presented in Appendix. In Section 5, we report experimental results on both simulated and real-world datasets. We conclude the paper in Section 6.

2. RELATED WORK

The problem of *truth discovery* has attracted much attention recently. People develop various approaches to extract true information from multiple sources of conflicting data. Initial studies [6, 24] were motivated by the observation that source reliability and truths are highly relevant–Truths are often stated by many reliable sources and reliable sources tend to tell truths more often. Based on this principle, approaches were developed to iteratively update source reliability and true facts. Later, various approaches were further developed to capture various factors that affect truth discovery [3, 5, 7, 8, 13, 16].

In recent work, this principle is formulated as an optimization framework [10]. The objective is to minimize the overall distance between source observations and truths in which sources are weighted by their reliability degrees. Then this framework was extended to handle data with long-tail distributions by calculating the confidence of source reliability estimates [9]. In [27], an optimization framework was developed based on min-max entropy. The solutions to these optimization formulations usually involve an iterative update of truths and source reliability as well.

In addition, probabilistic approaches [2, 4, 15, 19–21, 25, 26] were developed to tackle the truth discovery task. The basic idea is to formulate multi-source data as certain mixture of distributions and incorporate source reliability as some random variable into the probabilistic models. The approaches differ in the way of selecting proper distributions and capture source reliability to handle various scenarios in truth discovery, such as the existence of multiple truths, various difficulty levels of the task, and various data types. To obtain the truths, these approaches try to maximize likelihood or posterior distributions, which leads to iteratively updates in model parameters and truth inference. In particular, Dawid & Skene were the first to develop a maximum likelihood formulation and an expectation-maximization based approach to solve the problem. This approach was then adapted to social sensing scenarios in [19]. In [20], the authors developed a maximum likelihood estimator for source reliability and approximately quantified confidence in its estimation based on an asymptotic Cramer-Rao lower bound, but the convergence to the truth is not analyzed and cannot be guaranteed.

In summary, although the topic of *truth discovery* has been widely studied and most approaches achieve success in real-world applications, there is no theoretical guarantee that the results of these approaches converge to the truths. In this paper, we propose a novel truth discovery approach which can not only achieve comparable effectiveness in identifying the truths but also has theoretical guaranteed in the convergency to the truths.

3. METHODOLOGY

In this section, we mathematically formulate the truth discovery task, propose a randomized gaussian mixture model to represent multi-source data with various reliability degrees, derive both population-based and sample-based expectation-maximization solutions (i.e., *population*-EM$_{\text{RGMM}}$ and *sample*-EM$_{\text{RGMM}}$), and finally present the complete algorithm of *sample*-EM$_{\text{RGMM}}$.

3.1 Problem Formulation

We first introduce the notations that will be used throughout the paper and then state the target problem.

Input. Consider a set of entities $\mathcal{N} := \{n\}_{n=1}^{N}$ that we are interested in, and there are $\mathcal{S} := \{s\}_{1}^{S}$ sources which provide information about all $\mathcal{N}$ entities. For the s-th source, denote its claims on $\mathcal{N}$ entities as $\boldsymbol{X}^s = (x_1^s, \cdots, x_N^s)$, where x_n^s represents its claim on the n-th entity. Then, $\mathcal{X} = \{\boldsymbol{X}^s\}_{1}^{S}$ represents the whole set of claims over the sources $\mathcal{S}$.

Output. The truths for all entities are denoted as $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$, which is unknown a priori. Let's also denote the estimator of the truth obtained from a truth discovery approach as $\widehat{\boldsymbol{\mu}}$.

Truth Discovery Task. The truth discovery task is formally defined as follows: Given the data collection $\mathcal{X}$, the

goal of a truth discovery method is to obtain an estimate $\widehat{\boldsymbol{\mu}}$ for all entities' truths as close to $\boldsymbol{\mu}$ as possible.

We summarize the notations in Table 1. Some of the notations will be introduced later in the paper.

Table 1: Notations

Notation	Definition
$\mathcal{N}$	the set of entities
$\mathcal{S}$	the set of sources
$\boldsymbol{\mu}$	the vector of entities' truths
$\boldsymbol{\mu}^*$	the maximum likelihood estimator of $\boldsymbol{\mu}$
$\widehat{\boldsymbol{\mu}}$	the estimator of truths
$\boldsymbol{\eta}^s$	the bias for the source s
$\boldsymbol{X}^s$	the data collected from source s
$\mathcal{X}$	the whole claims over S sources
σ^2	the variance of data collection in the model

3.2 Randomized Gaussian Mixture Model

Now we propose a probabilistic model with parameters $\boldsymbol{\mu}$ whose estimate will be inferred via maximum likelihood estimator technique. In truth discovery, different sources are treated differently depending on the quality of their data. To do this, we assume that the reliability of the s-th source is captured by a random variable $\boldsymbol{\eta}^s$ that measures bias (i.e., deviation from the truth). The smaller absolute value the bias $\boldsymbol{\eta}^s$, the more reliable the s-th source. Considering all the $\mathcal{S}$ sources together, their reliability degrees $\{\boldsymbol{\eta}^s\}_1^S$ can be assumed to follow certain distributions, for example, a uniform distribution $\boldsymbol{\eta}^s \sim \mathrm{Uniform}(-C, C)^N$. This distribution models the overall quality of the collection of sources.

For the s-th source, we assume that the claims made by this source follow a multi-variate gaussian distribution with variance $\sigma^2 I_N$, where I_N is an identity matrix of size n. This assumption is typically adopted in many existing truth discovery work [9,10,27]. The mean of the gaussian distribution is $\boldsymbol{\mu} - \boldsymbol{\eta}^s$ in which $\boldsymbol{\mu}$ is the truth and $\boldsymbol{\eta}^s$ is the source bias. Therefore, a more reliable source's mean is closer to the truth and thus its claims are more likely to be close to the truth. With these notations, we can write the distribution of the claims made by the s-th source as:

$$\boldsymbol{X}^s|\boldsymbol{\eta}^s \sim \mathrm{Normal}(\boldsymbol{\mu} - \boldsymbol{\eta}^s, \sigma^2 I_N). \quad (1)$$

We further model the whole data collection using the following mixture model [23]:

$$p_{\boldsymbol{\mu}}(\boldsymbol{X}) = \frac{1}{S} \sum_{s=1}^{S} \mathrm{Normal}(\boldsymbol{\mu} - \boldsymbol{\eta}^s, \sigma^2 I_N), \quad (2)$$

where we assume equal weights among components. Different from traditional gaussian mixture model with fixed mean in each component, the proposed model's mean is random because we incorporate the random variables η^s as the source bias. Thus, we name the proposed model as **R**andomized **G**aussian Mixture **M**odel (RGMM). We introduce a latent variable $Y \in \{s\}_1^S$ as a source indicator, which is also an indicator of the underlying mixture component. Namely, the claims from the s-th component of RGMM is denoted as:

$$\boldsymbol{X}|Y = s \sim \mathrm{Normal}(\boldsymbol{\mu} - \boldsymbol{\eta}^s, \sigma^2 I_N). \quad (3)$$

We assume that the pair $(\boldsymbol{X}, Y)$ are random variables in the sample spaces $\mathcal{X} \times \mathcal{Y}$. Based on the RGMM, we formulate the problem of truth discovery as the task of estimating the model parameter $\boldsymbol{\mu}$. Specifically, the objective is to obtain

the estimate of $\boldsymbol{\mu}$ which maximizes the likelihood of observing the multi-source input. We define $\boldsymbol{\mu}^*$ as the maximum likelihood estimator of $\boldsymbol{\mu}$ in the proposed RGMM formulation. In the following sections, we propose effective solutions to estimate $\boldsymbol{\mu}^*$ and then demonstrate the approaches' convergence guarantee in Section 4.

3.3 EM Solutions for RGMM

The objective of the proposed approach is to obtain an estimate of the unknown parameter $\boldsymbol{\mu}^*$ which maximizes the likelihood of RGMM on the data. An effective approach for deriving MLE is expectation-maximization (EM) method. We develop EM solutions for two versions of MLE estimates in this task: population-based MLE (assuming the limit infinite data) and sample-based MLE (assuming a finite set of samples). We name these solutions as *population-* and *sample-* EMRGMM, respectively. Data is finite in real practice, so we should use *sample*-EMRGMM to identify entities' truths, but the introduction of *population*-EMRGMM enables us to conduct the convergence analysis for both solutions. In this subsection, we first introduce the general EM procedure, and then derive the EM updates of both *population-* and *sample*-EMRGMM.

First, let us briefly review the EM algorithm. Given the lower bounds on the log likelihood $Q(\cdot|\cdot)$, EM algorithm successively maximizes the lower bound and then reevaluates the lower bound at the new parameter value. The update procedure is as follows.

EM updates: Given $\boldsymbol{\mu}^{t-1}$ obtained at the $(t-1)$-th iteration, the t-th iteration of EM algorithm can be summarized in the following two steps:

- E-Step. Calculate the lower bound of the log likelihood $Q(\cdot|\boldsymbol{\mu}^{t-1})$.
- M-Step. Compute the maximizer as follows:

$$\boldsymbol{\mu}^t = \arg\max_{\boldsymbol{\mu}' \in \Omega} \quad Q(\boldsymbol{\mu}'|\boldsymbol{\mu}^{t-1}).$$

An EM-based method iteratively conduct these steps until some conditions are satisfied. Next, we introduce both the *population-* and *sample-* EMRGMM.

population-EMRGMM. When deriving the E-step and M-step for population-based MLE of RGMM, we assume that there is an infinite set of samples. The population-based Q-function $Q(\cdot|\boldsymbol{\mu})$ takes the form that

$$Q(\boldsymbol{\mu}'|\boldsymbol{\mu}) = \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} f_{\boldsymbol{\mu}}(y|\boldsymbol{X}) \log p_{\boldsymbol{\mu}'}(\boldsymbol{X}, y) dy \right) g_{\boldsymbol{\mu}}(\boldsymbol{X}) d\boldsymbol{X}, \quad (4)$$

where $f_{\boldsymbol{\mu}}(y|\boldsymbol{X})$ denotes the conditional density of Y given $\boldsymbol{X}$ and $g_{\boldsymbol{\mu}}(\boldsymbol{X})$ is the density function of the observed variable $\boldsymbol{X}$. The population-based EM operator $M : \mathbb{R}^N \to \mathbb{R}^N$ is defined as follows:

$$M(\boldsymbol{\mu}) := \arg\max_{\boldsymbol{\mu}' \in \mathbb{R}^N} Q(\boldsymbol{\mu}'|\boldsymbol{\mu}). \quad (5)$$

$M(\cdot)$ is to find the maximizer of the Q-function given the parameters $\boldsymbol{\mu}$ obtained in previous step.

sample-EMRGMM. *sample*-EMRGMM is derived on a finite set of the claims, i.e., $\mathcal{X} = \{\boldsymbol{X}^s\}_1^S$. We assume that each sample given by every source is drawn i.i.d. from the mixture density Eq. (2). Under this assumption, we define the sample-based Q-function as Q_S, which is shown as follows:

$$Q_S(\boldsymbol{\mu}'|\boldsymbol{\mu}) = \frac{1}{S} \sum_{s=1}^{S} \left(\sum_{s'}^{S} \mathbb{P}\left[y = s'|\boldsymbol{X}^s\right] \log p_{\boldsymbol{\mu}'}(\boldsymbol{X}^s, y) \right). \quad (6)$$

$\mathbb{P}[y = s'|\boldsymbol{X}^s]$ is the probability that a sample $\boldsymbol{X}$ belongs to source s', and its value is defined by the following function:

$$\mathbb{P}[y = s|\boldsymbol{X}] := e^{-\frac{\|\boldsymbol{X}-\boldsymbol{\mu}-\boldsymbol{\eta}^s\|_2^2}{2\sigma^2}} \left(\sum_{s'=1}^{S} e^{-\frac{\|\boldsymbol{X}-\boldsymbol{\mu}-\boldsymbol{\eta}^{s'}\|_2^2}{2\sigma^2}}\right)^{-1}. \quad (7)$$

To simplify the notation, we denote Eq. (7) as $\omega_{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{\eta}^s)$. Substituting $\omega_{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{\eta}^s)$ (Eq. (7)) into Eq. (6) and ignoring terms that do not contain $\boldsymbol{\mu}$, we show that the sample-based function $Q_S(\boldsymbol{\mu}'|\boldsymbol{\mu})$ takes the form:

$$Q_S(\boldsymbol{\mu}'|\boldsymbol{\mu}) = -\frac{1}{S}\sum_{s=1}^{S}\sum_{s'=1}^{S}\omega_{\boldsymbol{\mu}}(\boldsymbol{X}^s, \boldsymbol{\eta}^{s'})\frac{\|\boldsymbol{X}^s - \boldsymbol{\mu}' - \boldsymbol{\eta}^{s'}\|_2^2}{2\sigma^2}. \quad (8)$$

We denote the sample-based EM operator as $M_n : \mathbb{R}^N \to \mathbb{R}^N$, which is to maximize the sample-based Q-function. Namely, $M_S(\boldsymbol{\mu}) := \arg\max_{\boldsymbol{\mu}' \in \mathbb{R}^N} Q_S(\boldsymbol{\mu}'|\boldsymbol{\mu})$. According to Eq. (8), we have that

$$M_S(\boldsymbol{\mu}) = \frac{1}{S}\sum_{s=1}^{S}\sum_{s'=1}^{S}\omega_{\boldsymbol{\mu}}(\boldsymbol{X}^s, \boldsymbol{\eta}^{s'})(\boldsymbol{X}^s - \boldsymbol{\eta}^{s'}). \quad (9)$$

Moreover, Eq. (9) implies that $M(\boldsymbol{\mu}) = \mathbb{E}[\omega(\boldsymbol{X}, \boldsymbol{\eta})(\boldsymbol{X} - \boldsymbol{\eta})]$, where the expectation is taken over $\boldsymbol{X} \times \boldsymbol{\eta}$.

3.4 *Sample*-EMRGMM **Algorithm**

As discussed, *sample*-EMRGMM which deals with finite samples, is typically adopted in real practice. We summarize this algorithm in Algorithm 1. We will show its performance on both simulated and real-world datasets in Section 5.

Algorithm 1 *sample*-EMRGMM

Input: Entities $\mathcal{N} = \{n\}_1^N$, Sources $\mathcal{S} = \{s\}_1^S$, and data collection $\mathcal{X} = \{x_n^s\}_{s=1,n=1}^{S,N}$,
Output: Truth estimates $\widehat{\boldsymbol{\mu}}$

1: **for** entity n ($n = 1, \cdots, N$) **do**
2:    Calculate its variance $\hat{\sigma}_n^2$ over S sources;
3:    Initialize μ_n^{old} using the mean of claims over sources;
4: **end for**
5: Estimate model variance: $\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}\hat{\sigma}_n^2$;
6: Estimate upper bound of biases C using the maximum absolute value of $\mathcal{X}$;
7: **while** convergence criterion is not satisfied **do**
8:    For each source s, generate $\boldsymbol{\eta}^s \sim \text{Uniform}(-C, C)^N$;
9:    Update truth estimator according to Eq. (9):

$$\boldsymbol{\mu}^{new} = M_n(\boldsymbol{\mu}^{old});$$

10: **end while**
11: **return** $\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu}^{new}$

---

4. THEORETICAL ANALYSIS

In this section, we theoretically present convergence analysis for the proposed solutions: *population*- and *sample*- EMRGMM. The outline of this section is: (1) In Property 1, we first provide conditions under which the distance between the *population*-EMRGMM's result and the MLE is bounded; (2) Based on the concept of covering number of a metric space, we bound the distance between the result of *population*- and that of *sample*-EMRGMM (Corollary 1); (3) Based on Property 1 and Corollary 1, applying triangle inequality, the error between the *sample*-EMRGMM's result and MLE is upper bounded by the number of samples (Property 2).

*Convergence of population-*EMRGMM

Let's first introduce the convergence property of the *population*-EMRGMM. Recall that $\boldsymbol{\mu}^*$ represents the maximizer of the population likelihood. [1] introduces the self-consistency property for the maximum likelihood estimator, that is,

$$\boldsymbol{\mu}^* = \arg\max_{\boldsymbol{\mu}'} \quad Q(\boldsymbol{\mu}'|\boldsymbol{\mu}^*). \quad (10)$$

Eq. (10) implies that the maximum likelihood estimator should maximize the population-based Q-function. Combining with Eq. (5), it is obvious that $\boldsymbol{\mu}^* = M(\boldsymbol{\mu}^*)$. For the proposed RGMM (Eq. (2)), we have the following property.

PROPERTY 1. *Given the RGMM with a sufficiently small bias-to-mean ratio $\frac{C}{\|\boldsymbol{\mu}^*\|_2}$ and a sufficiently large signal-to-noise ratio $\frac{\|\boldsymbol{\mu}^*\|_2}{\sigma^2}$, there is a universal constant $c > 0$ and a constant $\lambda \in (0, 1)$ with $\lambda \leq \exp(-c\frac{\|\boldsymbol{\mu}^*\|_2}{\sigma^2})$, such that*

$$\|M(\boldsymbol{\mu}) - \boldsymbol{\mu}^*\|_2^2 \leq \lambda\|\boldsymbol{\mu}^* - \boldsymbol{\mu}\|_2^2, \quad (11)$$

holds for all $\boldsymbol{\mu}$ if $\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2 \leq \frac{\|\boldsymbol{\mu}^*\|_2}{4}$.

PROOF. Please refer to §A for a detailed proof. $\square$

The idea of the proof for Property 1 is adopted from [1]. However, two major differences are: (1) we incorporate a random variable $\boldsymbol{\eta}$ for each latent component whose mean share the same sign, and (2) we consider arbitrary number of latent components more than 2. Both differences make the proof more complicated comparing with Corollary 1 in [1]. The detailed proof is deferred in Appendix A.

Property 1 establishes the conditions under which the convergence of the *population*-EMRGMM $M(\boldsymbol{\mu})$ is guaranteed. Namely, the proposed $M(\boldsymbol{\mu})$ is contractive over a small ball around $\boldsymbol{\mu}^*$, a maximum likelihood estimator of the truths. Given an initial $\boldsymbol{\mu}^0$, an immediate result from Property 1 is:

$$\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^*\|_2^2 = \|M(\boldsymbol{\mu}^{t-1}) - \boldsymbol{\mu}^*\|_2^2 \leq \lambda\|\boldsymbol{\mu}^{t-1} - \boldsymbol{\mu}^*\|_2^2$$
$$\leq \lambda\|M(\boldsymbol{\mu}^{t-2}) - \boldsymbol{\mu}^*\|_2^2 \leq \cdots \leq \lambda^t\|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2^2. \quad (12)$$

Eq. (12) implies that $M(\boldsymbol{\mu})$ is linear convergence. Moreover, It shows that given any initialization, the proposed $M(\boldsymbol{\mu})$ is able to modify it as the iteration increases.

*Convergence of sample-*EMRGMM

To prove the convergence property of the *sample*-EMRGMM, we first measure the deviation of its result from that of *population*-EMRGMM. As the covering number is used in the proof, we formally introduce its definition from [22].

DEFINITION 1. *Let $\mathcal{G}$ be a subset of a metric space. $\forall\epsilon > 0$, the covering number $N(\mathcal{G}, \epsilon)$ is defined to be the minimal integer $n \in \mathbb{N}$ such that these n balls with radius ϵ cover $\mathcal{G}$.*

Based on Definition 1, the difference between *population*-EMRGMM ($M(\boldsymbol{\mu})$) and *sample*-EMRGMM ($M_n(\boldsymbol{\mu})$) is upper bounded as follows.

COROLLARY 1. *Given the population- and sample- based EM operator M_n and M, there exists a constant c_δ such that*

$$\sup_{\boldsymbol{\mu}\in\Omega}\|M_n(\boldsymbol{\mu}) - M(\boldsymbol{\mu})\|_2 \leq c_{\delta, N(\Omega, \delta)}S^{-\frac{1}{2}} \quad (13)$$

holds with probability at least $1 - \delta$, where Ω is the parameter space.

PROOF. Please refer to §B for a detailed proof. $\square$

Based on Property 1 and Corollary 1, we can bound the distance between the result of *sample*-EMRGMM $M_n(\boldsymbol{\mu})$ and the MLE $\boldsymbol{\mu}^*$, which is stated in the following property.

PROPERTY 2. *Under the conditions of Property 1, $\forall \boldsymbol{\mu}^0$ such that $\|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2 \leq r$, if there are enough sources, then $\{\boldsymbol{\mu}^t\}_{t=0}^{\infty}$ obtained by the sample-EM algorithm satisfies that*

$$\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^*\|_2 \leq \lambda^t \|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2 + \frac{1}{1-\lambda} c_{\delta, N(\Omega, \delta)} S^{-\frac{1}{2}} \quad (14)$$

holds with probability of at least $1 - \delta$, where $\lambda \in (0, 1)$.

PROOF. Please refer to §C for a detailed proof. $\square$

After conduct enough iterations, Eq. (14) shows that the main component of the upper bound is the second terms (i.e., $\frac{1}{1-\lambda} c_{\delta, N(\Omega, \delta)} S^{-\frac{1}{2}}$). Namely, Property 2 states that the performance of the proposed *sample*-EMRGMM is upper bounded by $\Omega(S^{-1/2})$. Intuitively, in truth discovery tasks the more the sources, the better the performance of methods. In Property 2, we theoretically present that the convergence rate of the *sample*-EMRGMM is $\Omega(S^{-1/2})$.

5. EXPERIMENTS

Note that, only finite number of samples can be obtained in real-world applications, which fits the *sample*-EMRGMM setting. Therefore, all experiments are conducted using *sample*-EMRGMM[1]. In this section, we test the *sample*-EMRGMM on both simulated and real-world data sets. The experimental results show the effectiveness of the proposed *sample*-EMRGMM in identifying truths as well as its convergence. We first introduce baselines and performance measures in Subsections 5.1 and 5.2, respectively. Experimental results on simulated data are presented in Subsection 5.3. In Subsection 5.4, we show experimental results on three real-world application datasets: Weather Forecast, Indoor Floorplan, and Stock Data.

5.1 Baselines

A variety of truth discovery methods have been developed to identify each object's truth. As we consider applications of continuous data in this paper, we compare the *sample*-EMrgmm with three state-of-the-art truth discovery methods CRH, CATD, GTM, and two naive methods: Mean and Median. Details of the baselines are shown as follows:

- CRH: In [10], truth discovery task is formulated as an optimization problem, seeking the optimal truth estimators and weights to minimize the weighted distance between claims and the truths. CRH is proposed to iteratively update the truths and source weights.
- CATD: CATD [9] is a statistical method that has been proposed for long-tail phenomenon in truth discovery, where confidence interval is incorporated in source weight estimation.
- GTM: [25] proposes a probabilistic graphical model based method to solve the truth discovery task. In their framework, source reliability and truths are model parameters to estimate.
- Mean: The average of claims provided by multiple sources is treated as the final estimator of the truth.
- Median: The median of claims provided by multiple sources is defined as the final estimator of the truth.

[1] We will use RGMM and *sample*-EMRGMM interchangeably in the experiment section

5.2 Performance Measures

In the experiments, we have continuous input obtained from multiple sources. Although the ground truths are available, we conduct all methods in an unsupervised manner and the ground truths will only be used in evaluation. To evaluate the performance of sample-EMRGMM as well as baselines, we adopt the following measures:

- MAE: It measures the mean of absolute error between the output of methods and the ground truths. As L^1-norm is applied, MAE penalizes more on small errors.
- $RMSE$: It measures the root of mean squared error between output and groundtruth. $RMSE$ penalizes more on big errors because of the involved L^2-norm.
- $ErrorRate(\epsilon)$: It is defined as the percentage of the estimated truths falling outside an ϵ-ball[2] of the ground truth.

Note that a lower measure value means that the truth estimates are closer to the ground truths. Thus, for all measures, the lower the value, the better the method's performance.

5.3 Simulated Data

The advantage of using simulated data is that we can simulate different truth discovery scenarios to compare the performance of the proposed *sample*-EMRGMM with that of the baselines. In this section, we first introduce the procedure of generating simulated data. Then, we show the performance of the *sample*-EMRGMM as well as the comparison with baselines in terms of MAE, $RMSE$, and $ErrorRate(\epsilon)$.

Data Generation. In each experiment, we generate $N = 200$ entities and $S = 100$ sources. For each source, its bias $(\boldsymbol{\eta}^s)$ is drawn i.i.d. from a distribution F. We assume that the ground truth for all entities are 0s. Thus, the s-th component (i.e., source) of the mixture model follows a multivariate normal distribution, Normal$(-\boldsymbol{\eta}^s, \sigma^2 I_N)$. To generate the sample of claims $\{\boldsymbol{X}^s\}_1^S$, we first randomly generate a source index s from $[1, \cdots, S]$, and then $\boldsymbol{X}^s$ is drawn from Normal$(-\boldsymbol{\eta}^s, \sigma^2 I_N)$, where $\sigma^2 = 1$. We use MAE and $RMSE$, and $Error\ Rate$ (0.1) for evaluation. We simulate three different scenarios involving different distributions of source biases: Uniform, Normal, and Student's t-distribution, and then evaluate the performance of all truth discovery methods.

Scenario 1: $\eta^s \sim$ Uniform$(-c, c)$. In this scenario, sources' biases are drawn from a uniform distribution with $c = 2$. The source reliability degrees are uniformly distributed. We report the results on experiments with different source number $S = \{10, 20, \cdots, 100\}$ in terms of all methods in Figure 1. In Figure 1, the solid and dark line represents the value of function $f(S) = S^{-\frac{1}{2}}$, which is the dominated term in upper bound of the proposed *sample*-EMRGMM (Eq. (14)). From Figures 1(a), 1(b) and 1(c), we can see that the convergence of the proposed *sample*-EMRGMM is similar to the $S^{-\frac{1}{2}}$, which confirms the result in Property 2. Moreover, we can see that the performance of the proposed RGMM is better when comparing with baselines in terms of $RMSE$ and $Error\ Rate$. It means that most of truth estimates from the *sample*-EMRGMM have smaller errors compared with that of baselines. For MAE, all truth discovery methods have the same performance.

Scenario 2: $\eta^s \sim$ Normal$(0, \delta^2)$. In this scenario, sources' biases are drawn from a Normal distribution with variance

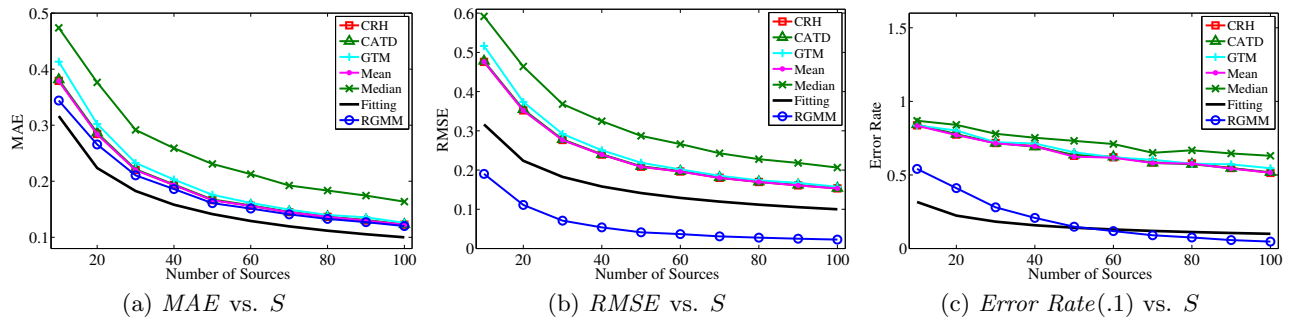[2] ϵ is chosen based on the scale of the multi-source data.

(a) *MAE* vs. *S*   (b) *RMSE* vs. *S*   (c) *Error Rate*(.1) vs. *S*

Figure 1: *Simulated data on Scenario* 1*: Performance with respect to the Number of Sources (S).*

(a) *MAE* vs. *S*   (b) *RMSE* vs. *S*   (c) *Error Rate*(.1) vs. *S*
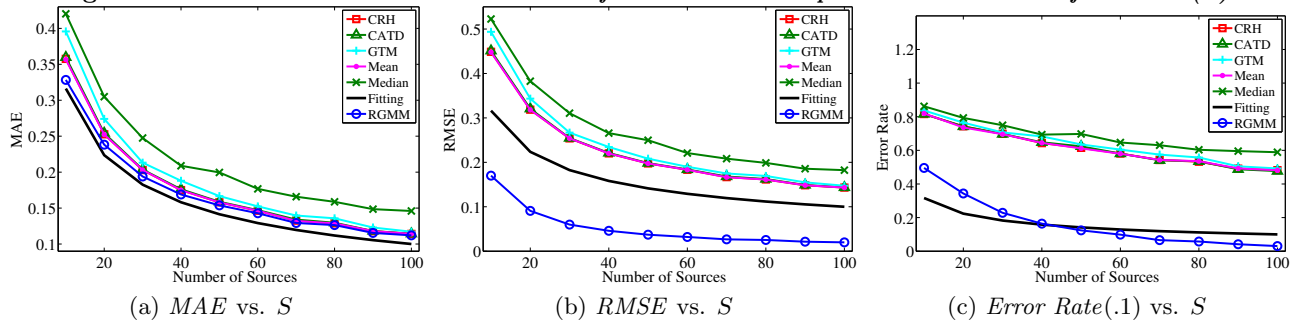
Figure 2: *Simulated data on Scenario* 2*: Performance with respect to the Number of Sources (S).*

(a) *MAE* vs. *S*   (b) *RMSE* vs. *S*   (c) *Error Rate*(.1) vs. *S*
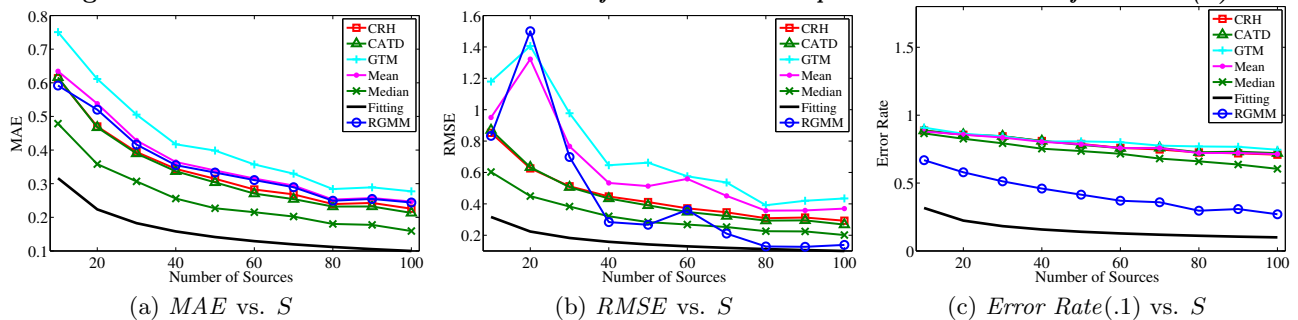
Figure 3: *Simulated data on Scenario* 3*: Performance with respect to the Number of Sources (S).*

$\delta^2 = 0.5$. Based on three-sigma rule of thumb, sources biases fall into the interval $[-\frac{3}{\sqrt{2}}, \frac{3}{\sqrt{2}}]$ with probability of 99.73%. Meanwhile, as the variance of the normal distribution is small, there are many samples closer to the mean 0. Therefore, there are many reliable sources than unreliable sources. The results are shown in Figure 2. The convergence performance of the *sample*-EMRGMM with respect to the number of sources is similar to that in Scenario 1.

Scenario 3: $\eta^s \sim$ Student's t-(ν). In this scenario, sources' biases are drawn from a student's t-distribution with freedom $\nu = 2$. Compared with previous scenarios, Student's t-distribution has heavier tails, i.e., it is more prone to producing values that fall far from its mean 0. Consequently, there are more unreliable sources. The results in terms of *MAE*, *RMSE*, and *Error Rate*(.1) are shown in Figure 3.

Result Analysis. Comparing the performance on Scenarios 1 ∼ 3, we can see that all methods perform best in Scenario 3 and worst in Scenario 3. In Property 1, the conditions show that the performance is better if the upper bound of data C is smaller and the variance σ^2 is larger. In Scenario 1, $\widehat{C} = 6.3476$ and $\hat{\sigma}^2 = 2.4502$, while $\widehat{C} = 5.74$, and $\hat{\sigma}^2 = 2.0886$ in Scenario 2. As student t-distribution is long-tail, there are some sources which have very large bias. Thus, in Scenario 3, the upper bound of original simulated data is 161.4553. The claims provided these sources can be treat-

ed as outliers. After removing them, we have $\widehat{C} = 19.9335$ and $\hat{\sigma}^2 = 5.8458$. Based on Property 11, the performance in Scenario 2 should be the best, as shown in Figure 2.

The convergence rate of the *sample*-EMRGMM is $\Omega(S^{-\frac{1}{2}})$ as shown in the Section 4. Experimentally, the convergence rate in Scenarios 1 and 2 is nicely fit to $\Omega(S^{-\frac{1}{2}})$, as shown in Figures 1 and 2. In Scenario 3, the convergence is not as clear as that in Scenarios 1 and 2, as there are more outliers.

5.4 Real-World Data

Data Description. We test the proposed *sample*-RGMM and baselines on real-world data. The detailed description of each dataset and their tasks are shown as follows:

- Weather Forecast [10]: Temperature forecasts information for 88 cities in US are collected from three websites: HAM weather[3], Wundergound[4], and World Weather Online[5]. Besides, the real temperature for all cities are also crawled as ground truths for evaluation. The goal is to estimate the true temperature for each city from the conflicting data provided by different sources.
- Indoor Floorplan [9]: An Andriod App is designed for smart phone users to collect their estimates of hallway

---

[3]http://www.hamweather.com
[4]http://www.wunderground.com
[5]http://www.worldweatheronline.com

distance. There are totally 308 claims from 44 users on 7 indoor hallways. The ground truths are obtained via manually measuring the hallways by tape. The goal is to estimate the distance of indoor hallways from the data provided by a crowd of users.

- Stock Data: The stock data in [11] contains the price information for 1000 stocks from 55 sources over 21 days. To fit our scenarios, we preprocess the data and obtain a full dense dataset which contains the price information of 300 stocks from 36 sources over 19 days. In this task, we only focus on the open price, so the goal is to estimate the true open price for each stock.

Note that we have a different task setting compared with [9,10] on the real-world datasets reported in this paper. We consider a scenario where all entities are claimed by all sources while CATD [9] and CRH [10] were applied to entities that are observed by a subset of sources. To fit the full observation scenario, we preprocess the data used in [9,10] by deleting those entities which have not been claimed by all sources. In addition, our model in this paper tackles continuous data only. Therefore, we select continuous attributes in the Weather Forecast dataset, and the "Price" attribute (i.e., a continuous attribute) in the Stock dataset ([11]), which also differs from the setting in [10] (i.e., using both categorical and continuous attributes). The statistics of three real-world datasets are summarized in Table 2.

Table 2: Statistics of real-world datasets

Dataset	sources	objects	claims
Weather Forecast	30	88	3,344
Indoor Floorplan	44	7	308
Stock Data	36	$300 * 19$	$108,000 * 19$

Result Analysis. On the Weather Forecast data, we reduce the scale of the observations in preprocess step. For example, the original 77 Fahrenheit is changed to .77. We evaluate all methods on different scenarios in which the number of sources increases from 4 to 10 by the stepsize of 2. In Figure 4, we report the experimental results in terms of MAE and $RMSE$. We can see that the performance of the proposed RGMM improves as the number of sources increases. When the number of sources is relatively small, the performance of the proposed RGMM is worse than baselines. however, given a plenty of sources, RGMM converges to other baselines. The experimental results with respect to $Error\ Rate(\epsilon)$ is presented in Table 3. Table 3 shows that the performance of RGMM is comparable to that of baselines.
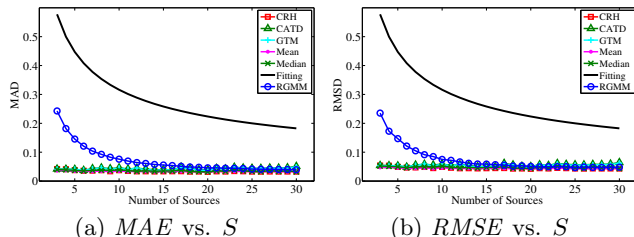

(a) MAE vs. S (b) $RMSE$ vs. S

Figure 4: *Weather Forecast dataset: Performance with respect to the Number of Sources (S).*

On Indoor Floorplan dataset, we test all truth discovery methods on different scenarios where the number of sources increases from 6 to 44 by the stepsize of 1. In each scenario, we randomly choose the pre-fixed number of sources. To

Table 3: *Error Rate(ϵ) on Weather Forecasts*

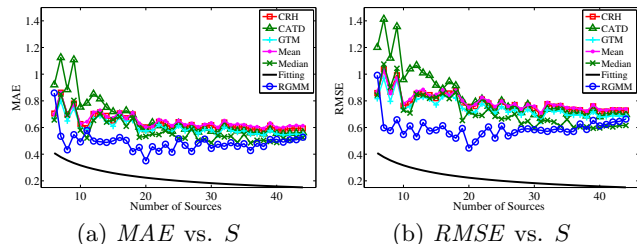| Method	Error Rate(ϵ)					
	.05	.06	.07	.08	.09	.10
RGMM	.2159	.1591	.1477	.1023	.0795	.0455
CRH	.2045	.1477	.1023	.0682	.0568	.0341
CATD	.4205	.3523	.2955	.2273	.1705	.1364
GTM	.4091	.3409	.2386	.1932	.1591	.1023
Mean	.2045	.1477	.1023	.0682	.0568	.0341
Median	.2045	.1477	.1023	.0682	.0455	.0341

(a) MAE vs. S (b) $RMSE$ vs. S

Figure 5: *Indoor Floorplan dataset: Performance with respect to the Number of Sources (S).*

reduce the randomness, we repeat experiments 20 times and report the average of evaluation measures. The performance comparison between the proposed RGMM and baselines in terms of MAE and $RMSE$ are presented in Figures 5(a) and 5(b), respectively. Note that the decrease rate is not exactly fit to the exponential function with power to $-\frac{1}{2}$. For both measures, although the performance of all methods decrease as the number of sources increases, the proposed RGMM is the best. Namely, the truth estimates obtained by RGMM is closer to the truths. We also report experimental results in terms of $Error\ Rate(\epsilon)$. We change ϵ from .5 to 1.0 and show the results in Table 4. We can see that the performance of the proposed RGMM is the best on all scenarios exept $\epsilon = .6$.

Table 4: *Error Rate(ϵ) on Indoor Floorplan*

Method	Error Rate(ϵ)					
	.5	.6	.7	.8	0.9	1.0
RGMM	.5714	.4286	.2857	.2857	.1429	.1429
CRH	.4286	.4286	.2857	.2857	.2857	.2857
CATD	.4286	.4286	.2857	.2857	.2857	.2857
GTM	.4286	.2857	.2857	.2857	.2857	.2857
Mean	.5714	.4286	.4286	.2857	.2857	.2857
Median	.7143	.2857	.2857	.2857	.1429	.1429

On Stock data, we follow the similar experiment design. We report the results in terms of MAE and $RMSE$ measures in Table 5. Table 5 shows that the performance of RGMM comparable with baselines. We also test the performance with respect to the number of sources. Due to the page limit, we only show the results on the data collected on day 1 in Figure 6. To better confirm the theoretical results obtained in Section 4, we report the convergence of the RGMM with respect to the number of sources on all days' Stock data in Figure 7. Each blue line represents a experiment conducted in a single day Stock data. We also fit each line into a function $a * S^{-\frac{1}{2}} + c$ where a and c are coefficients. The red line is plot using the average of a and c over 19 days. From Figures 6 and 7, we can see that the convergence of the RGMM's performance is indeed $\Omega(S^{-\frac{1}{2}})$, which perfectly confirms our theoretical results.

Table 5: Performance comparison on Stock data

Data	MAE						RMSE					
	RGMM	CRH	CATD	GTM	Mean	Median	RGMM	CRH	CATD	GTM	Mean	Median
Day1	.0162	.0144	.0145	.0145	.0138	.0145	.0200	.0175	.0175	.0175	.0170	.0175
Day2	.0161	.0149	.0149	.0149	.0144	.0150	.0200	.0199	.0199	.0199	.0189	.0199
Day3	.0155	.0238	.0238	.0238	.0225	.0238	.0200	.0303	.0303	.0303	.0288	.0303
Day4	.0175	.0221	.0222	.0222	.0220	.0222	.0245	.0287	.0287	.0287	.0279	.0288
Day5	.0227	.0234	.0235	.0235	.0231	.0235	.0316	.0316	.0316	.0316	.0304	.0316
Day6	.0308	.0240	.0240	.0240	.0233	.0240	.0337	.0337	.0337	.0337	.0330	.0337
Day7	.0351	.0273	.0273	.0273	.0268	.0273	.0469	.0399	.0400	.0400	.0393	.0400
Day8	.0394	.0303	.0304	.0304	.0296	.0304	.0414	.0414	.0414	.0414	.0407	.0414
Day9	.0505	.0370	.0370	.0370	.0365	.0370	.0525	.0524	.0525	.0525	.0520	.0525
Day10	.0522	.0402	.0403	.0403	.0397	.0403	.0532	.0548	.0548	.0548	.0542	.0548
Day11	.0543	.0425	.0427	.0427	.0412	.0427	.0663	.0558	.0560	.0560	.0545	.0560
Day12	.0479	.0436	.0437	.0437	.0419	.0437	.0608	.0561	.0562	.0562	.0546	.0562
Day13	.0576	.0461	.0462	.0462	.0440	.0462	.0600	.0600	.0601	.0600	.0585	.0600
Day14	.0472	.0438	.0439	.0439	.0422	.0439	.0612	.0585	.0585	.0585	.0575	.0586
Day15	.0506	.0453	.0454	.0454	.0439	.0454	.0600	.0598	.0598	.0598	.0590	.0598
Day16	.0539	.0500	.0501	.0501	.0486	.0501	.0652	.0642	.0643	.0642	.0632	.0643
Day17	.0605	.0544	.0546	.0545	.0525	.0546	.0723	.0703	.0705	.0705	.0690	.0705
Day18	.0761	.0659	.0660	.0660	.0645	.0660	.0856	.0849	.0849	.0849	.0833	.0850
Day19	.0794	.0678	.0679	.0678	.0666	.0679	.0901	.0856	.0857	.0857	.0842	.0857

(a) *MAE* vs. *S*   (b) *RMSE* vs. *S*
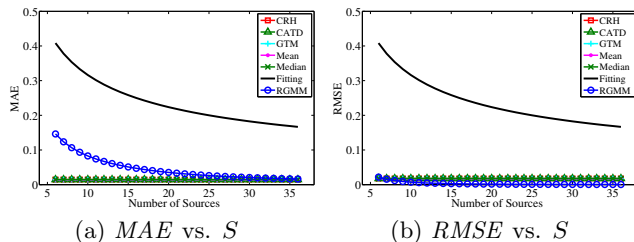
Figure 6: *Stock data on Day* 1*: Performance with respect to the number of sources (S).*
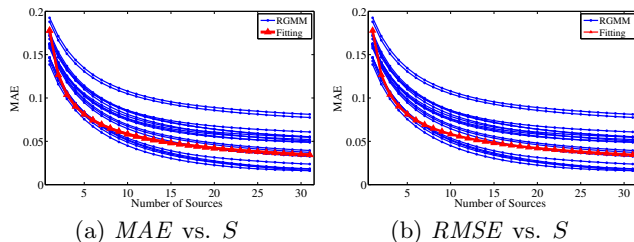

(a) *MAE* vs. *S*   (b) *RMSE* vs. *S*

Figure 7: *Convergence w.r.t. the number of sources on all Stock data*

In summary, the experimental results on both simulated and real-world datasets demonstrate that: (1) the performance of the proposed *sample*-EMRGMM is comparable compared with several state-of-the-art methods, and (2) the convergence rate of its performance is $\Omega(S^{-\frac{1}{2}})$ as proved in Section 4 (Property 2).

6. CONCLUSIONS

With the increasing possibilities to collect data from multiple sources in the real world, it is critical to identify true facts from conflicting data. To solve this problem, many algorithms were developed based on heuristic principles, optimizations, or probabilistic model. However, in existing literature on truth discovery, the convergence analysis is missing, and thus there is no theoretical guarantee that the results of these algorithms converge to the truths. In this paper, we proposed an effective truth discovery approach with theoretical guarantee. We first introduced the randomized biases of sources to measure their reliability degrees. Then we proposed a novel model (RGMM) to represent multi-source data with various reliability degrees, which consists of a Gaussian mixture model with the randomized biases incorporated. The parameters of interests in this model are the truths to be identified. We then derive both *population*- and *sample*- EMRGMM to the MLE of the truth parameter of RGMM. Theoretically, we prove that *population*-EMRGMM converges in probability to an ϵ-ball around the MLE with the increasing number of sources, under certain conditions. Moreover, we prove that *sample*-EMRGMM also converges to the ϵ-ball around the MLE with more iterations. In the experiments, we evaluate the effectiveness of the proposed *sample*-EMRGMM on both simulated and real-world datasets. Experimental results demonstrate that the proposed *sample*-EMRGMM is able to identify reliable information from multi-source data, and the estimator converges to an ϵ-ball around the MLE under the stated conditions.

8. REFERENCES

[1] S. Balakrishnan, M. J. Wainwright, and B. Yu. Statistical guarantees for the EM algorithm: From population to sample-based analysis. *arXiv:1408.2156*, 2014.

[2] L. Blanco, V. Crescenzi, P. Merialdo, and P. Papotti. Probabilistic models to reconcile complex data from inaccurate data sources. In *Proc. of CAiSE*, pages 83–97, 2010.

[3] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu. An approach to evaluate data trustworthiness based on data provenance. In *Proc. of SDM*, pages 82–98, 2008.

[4] A. P. Dawid and A. M. Skene. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Appl. Stat.*, pages 20–28, 1979.

[5] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proc. of KDD*, pages 601–610, 2014.

[6] X. L. Dong, L. Berti-Equille, and D. Srivastava. Integrating conflicting data: the role of source dependence. *PVLDB*, pages 550–561, 2009.

[7] A. Galland, S. Abiteboul, A. Marian, and P. Senellart. Corroborating information from disagreeing views. In *Proc. of WSDM*, pages 131–140, 2010.

[8] F. Li, M. L. Lee, and W. Hsu. Entity profiling with varying source reliabilities. In *Proc. of KDD*, pages 1146–1155, 2014.

[9] Q. Li, Y. Li, J. Gao, L. Su, B. Zhao, M. Demirbas, W. Fan, and J. Han. A confidence-aware approach for truth discovery on long-tail data. *PVLDB*, 2014.

[10] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han. Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In *Proc. of SIGMOD*, pages 1187–1198, 2014.

[11] X. Li, X. L. Dong, K. Lyons, W. Meng, and D. Srivastava. Truth finding on the deep web: is the problem solved? *PVLDB*, pages 97–108, 2012.

[12] F. Ma, Y. Li, Q. Li, M. Qiu, J. Gao, S. Zhi, L. Su, B. Zhao, H. Ji, and J. Han. Faitcrowd: Fine grained truth discovery for crowdsourced data aggregation. In *Proc. of KDD*, pages 745–754, 2015.

[13] A. Marian and M. Wu. Corroborating information from web sources. *Data Eng. Bull.*, pages 11–17, 2011.

[14] C. Meng, W. Jiang, Y. Li, J. Gao, L. Su, H. Ding, and Y. Cheng. Truth discovery on crowd sensing of correlated entities. In *Proc. of SenSys*, pages 169–182, 2015.

[15] S. Mukherjee, G. Weikum, and C. Danescu-Niculescu-Mizil. People on drugs: credibility of user statements in health communities. In *Proc. of KDD*, pages 65–74, 2014.

[16] J. Pasternack and D. Roth. Making better informed trust decisions with generalized fact-finding. In *Proc. of IJCAI*, pages 2324–2329, 2011.

[17] G.-J. Qi, C. C. Aggarwal, J. Han, and T. Huang. Mining collective intelligence in diverse groups. In *Proc. of WWW*, pages 1041–1052, 2013.

[18] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, pages 195–239, 1984.

[19] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher. On truth discovery in social sensing: A maximum likelihood estimation approach. In *Proc. of IPSN*, pages 233–244, 2012.

[20] D. Wang, L. M. Kaplan, T. F. Abdelzaher, and C. C. Aggarwal. On scalability and robustness limitations of real and asymptotic confidence bounds in social sensing. In *Proc. of SECON*, pages 506–514, 2012.

[21] P. Welinder, S. Branson, P. Perona, and S. J. Belongie. The multidimensional wisdom of crowds. In *Proc. of NIPS*, pages 2424–2432, 2010.

[22] Q. Wu and D.-X. Zhou. SVM soft margin classifiers: linear programming versus quadratic programming. *Neural Comput.*, pages 1160–1187, 2005.

[23] J. Yin and J. Wang. A dirichlet multinomial mixture model-based approach for short text clustering. In *Proc. of KDD*, pages 233–242, 2014.

[24] X. Yin, J. Han, and P. S. Yu. Truth discovery with multiple conflicting information providers on the web. *TKDE*, pages 796–808, 2008.

[25] B. Zhao and J. Han. A probabilistic model for estimating real-valued truth from conflicting sources. *Proc. of QDB*, 2012.

[26] B. Zhao, B. I. P. Rubinstein, J. Gemmell, and J. Han. A bayesian approach to discovering truth from conflicting sources for data integration. *PVLDB*, pages 550–561, 2012.

[27] D. Zhou, J. C. Platt, S. Basu, and Y. Mao. Learning from the wisdom of crowds by minimax entropy. In *Proc. of NIPS*, pages 2204–2212, 2012.

APPENDIX

A. PROOF OF PROPERTY 1

We follow the similar procedure of the proof in [1]: Decompose $M(\boldsymbol{\mu}) - \boldsymbol{\mu}^*$ into two separate functions: $\mathbb{E}[\Gamma_\delta^1(\boldsymbol{X})]$ and $\Gamma_\delta^2(\boldsymbol{X})$, and then bound $\Gamma_\delta^1(\boldsymbol{X})$ and $\Gamma_\delta^2(\boldsymbol{X})$.

Define $\boldsymbol{\mu}_u = \boldsymbol{\mu}^* + \delta\Delta$ where $\Delta := \boldsymbol{\mu} - \boldsymbol{\mu}^*$, $\forall \delta \in [0,1]$. Before applying Taylor's property to the function $\boldsymbol{\mu} \to \omega_{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{\eta})$, we first take a look at its derivative and upper bound it as follows: $\frac{\partial \omega_{\boldsymbol{\mu}}^s(\boldsymbol{X}, \boldsymbol{\eta})}{\partial \boldsymbol{\mu}} \leq \alpha \frac{\exp(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle + \langle \boldsymbol{\mu}_\delta, \boldsymbol{\mu}_\delta \rangle}{\sigma^2})}{\exp(\frac{2\langle \boldsymbol{X}, \boldsymbol{\mu}_\delta \rangle}{\sigma^2})}$ where $\alpha \triangleq \frac{2C}{S\sigma^2} e^{\frac{C^2}{2\sigma^2}}$. We first apply Taylor property to the function $\boldsymbol{\mu} \to \omega_{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{\eta})$ and take the expectation over $\boldsymbol{X}$. Combining with the upper bound, we have that

$$
\mathbb{E}\left[(\omega_{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{\eta}) - \omega_{\boldsymbol{\mu}^*}(\boldsymbol{X}, \boldsymbol{\eta}))(\boldsymbol{X} - \boldsymbol{\eta})\right]
$$
$$
\leq \underbrace{\int_0^1 \mathbb{E}\left[\alpha \Psi(\boldsymbol{X}, \boldsymbol{\mu}_\delta)\boldsymbol{X}^T\right]\Delta d\delta}_{\Gamma_\delta^1(\boldsymbol{X})} + \underbrace{\int_0^1 \mathbb{E}\left[\alpha \Psi(\boldsymbol{X}, \boldsymbol{\mu}_\delta)\boldsymbol{\eta}\right]\Delta d\delta}_{\Gamma_\delta^2(\boldsymbol{X})}, \quad (15)
$$

where $\Psi(\boldsymbol{X}, \boldsymbol{\mu}_\delta) = \frac{\exp(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle + \langle \boldsymbol{\mu}_\delta, \boldsymbol{\mu}_\delta \rangle}{\sigma^2})}{\exp(\frac{2\langle \boldsymbol{X}, \boldsymbol{\mu}_\delta \rangle}{\sigma^2})}$. Thus, we have that

$$
\|\mathbb{E}\left[(\omega_{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{\eta}) - \omega_{\boldsymbol{\mu}^*}(\boldsymbol{X}, \boldsymbol{\eta}))(\boldsymbol{X} - \boldsymbol{\eta})\right]\|_2
$$
$$
\leq \left[\sup_{\delta \in [0,1]} \|\Gamma_\delta^2(\boldsymbol{X})\|_{op} + \sup_{\delta \in [0,1]} \|\Gamma_\delta^1(\boldsymbol{X})\|_{op}\right] \|\Delta\|_2. \quad (16)
$$

The following two Lemmas provide the upper bounds for $\mathbb{E}[\Gamma_\delta^2(\boldsymbol{X})]$ and $\mathbb{E}[\Gamma_\delta^2(\boldsymbol{X})]$, respectively. The proofs are deferred in A.1 and A.2.

LEMMA 1. *There exist* $\beta_1, \beta_2, \beta_3$ *and* β_4, *such that*

$$
\Gamma_\delta^1(\boldsymbol{X}) \leq \alpha \left[\beta_1 \frac{\sigma^2}{\|\boldsymbol{\mu}^*\|_2} + \beta_2 \|\boldsymbol{\mu}^*\|_2 + \beta_3 \frac{\sigma^2}{\|\boldsymbol{\mu}^*\|_2^3} + \beta_4 \|\boldsymbol{\mu}^*\|_2^3\right] e^{-\frac{\|\boldsymbol{\mu}^*\|_2^2}{8\sigma^2}}.
$$

LEMMA 2. *There exist* λ_1, λ_2 *and* λ_3, *such that*

$$
\Gamma_\delta^2(\boldsymbol{X}) \leq \alpha \left[\lambda_1 + \lambda_2 \frac{\|\boldsymbol{\mu}^*\|_2^2}{\sigma^2} + \lambda_3 \frac{\sigma^2}{\|\boldsymbol{\mu}^*\|_2^2}\right] e^{-\frac{\|\boldsymbol{\mu}^*\|_2^2}{8\sigma^2}}.
$$

Applying Lemmas 1 and 2, it is easy to prove Theorem 1. Specifically, substituting Equations (20), (21), (23), and (24) into (16), we have that

$$
\mathbb{E}\left[\omega_{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{\eta}) - \omega_{\boldsymbol{\mu}^*}(\boldsymbol{X}, \boldsymbol{\eta})(\boldsymbol{X} - \boldsymbol{\eta})\right]
$$
$$
\leq c_1(1 + \phi + \phi\rho + \phi\rho^2 + \frac{1}{\rho})\frac{e^{-c_2\rho^2}}{S}\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2, \quad (17)
$$

whenever $\frac{\|\boldsymbol{\mu}^*\|_2^2}{\sigma^2} \geq 16/3$. Based on this fact, the bound (11) holds provided that the single-to-noise ratio $\frac{\|\boldsymbol{\mu}^*\|_2}{\sigma^2}$ is sufficiently large, and the bias-to-mean ratio $\frac{C}{\|\boldsymbol{\mu}^*\|_2}$, is sufficiently small. So far, we have finished the proof for Property 1.

A.1 Proof of Lemma 1

We first apply Taylor's property to the function $\boldsymbol{X} \to \exp(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle + \langle \boldsymbol{\mu}_\delta, \boldsymbol{\mu}_\delta \rangle}{\sigma^2})$, which yields

$$\Gamma^1_\delta(\boldsymbol{X}) = \mathbb{E}\underbrace{\left[\frac{\alpha(1+\frac{\|\boldsymbol{\mu}_\delta\|_2^2}{\sigma^2})\boldsymbol{X}^T}{\exp(\frac{2\langle\boldsymbol{X},\boldsymbol{\mu}_\delta\rangle}{\sigma^2})}\right]}_{\gamma_1^1(\boldsymbol{X})} + \mathbb{E}\underbrace{\left[\frac{\alpha\langle\boldsymbol{X},\boldsymbol{X}\rangle\boldsymbol{X}^T}{\exp(\frac{2\langle\boldsymbol{X},\boldsymbol{\mu}_\delta\rangle}{\sigma^2})}\right]}_{\gamma_2^1(\boldsymbol{X})}. \quad (18)$$

Based on (18), it is easy to obtain that

$$\sup_{\delta\in[0,1]}\|\Gamma^1_\delta(\boldsymbol{X})\|_{op} \leq \sup_{\delta\in[0,1]}\|\mathbb{E}(\gamma_1^1(\boldsymbol{X}))\|_{op} + \sup_{\delta\in[0,1]}\|\mathbb{E}(\gamma_2^1(\boldsymbol{X}))\|_{op}.$$

The remainder of the proof is to show a sufficient uniform upper bound of $\|\mathbb{E}(\gamma_1^1(\boldsymbol{X}))\|_{op}$ and $\|\mathbb{E}(\gamma_2^1(\boldsymbol{X}))\|_{op}$ over $\delta \in [0,1]$. Based on the discussion before, the distribution of $\boldsymbol{X}$ is symmetric around $\|\boldsymbol{\mu}^*\|_2$. Let us define $A = \{X \leq \frac{\|\boldsymbol{\mu}^*\|}{4}\}$. Note that $\|\boldsymbol{\mu}_\delta\|_2 \leq (1+r)\|\boldsymbol{\mu}*\|_2$. For $\|\mathbb{E}(\gamma_1^1(\boldsymbol{X}))\|_{op}$, therefore, we have that

$$\mathbb{E}(\gamma_1^1(\boldsymbol{X}))$$
$$\leq \mathbb{E}[\frac{\alpha(1+\frac{\|\boldsymbol{\mu}_\delta\|_2^2}{\sigma^2})\boldsymbol{X}^T}{\exp(\frac{2\langle\boldsymbol{X},\boldsymbol{\mu}_\delta\rangle}{\sigma^2})}|A]\mathbb{P}[A] + \mathbb{E}[\frac{\alpha(1+\frac{\|\boldsymbol{\mu}_\delta\|_2^2}{\sigma^2})\boldsymbol{X}^T}{\exp(\frac{2\langle\boldsymbol{X},\boldsymbol{\mu}_\delta\rangle}{\sigma^2})}|A^c]$$
$$\leq \mathbb{E}[\frac{\alpha(1+\frac{1+r}{\sigma^2}\|\boldsymbol{\mu}^*\|_2^2)\boldsymbol{X}^T}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{X}}{\sigma^2})}|A]\mathbb{P}[A] + \mathbb{E}[\frac{\alpha(1+\frac{1+r}{\sigma^2}\|\boldsymbol{\mu}^*\|_2^2)\boldsymbol{X}^T}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{X}}{\sigma^2})}|A^c] \quad (19)$$
$$\leq \frac{\alpha\sigma^2(1+\frac{1+r}{\sigma^2}\|\boldsymbol{\mu}^*\|_2^2)}{2e\|\boldsymbol{\mu}_\delta\|_2}\mathbb{P}[A] + \frac{\alpha(1+\frac{1+r}{\sigma^2}\|\boldsymbol{\mu}^*\|_2^2)\|\boldsymbol{\mu}^*\|_2}{4\exp(\frac{\|\boldsymbol{\mu}_\delta\|_2\|\boldsymbol{\mu}^*\|_2}{2\sigma^2})}.$$

Based on the fact that $\|\boldsymbol{\mu}_\delta\|_2 = \|\boldsymbol{\mu}^* + \delta(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\|_2 \geq \|\boldsymbol{\mu}^*\|_2 - \frac{1}{4}\|\boldsymbol{\mu}^*\|_2$, and the standard Gaussian tail bounds, $\mathbb{P}[A] \leq \exp(-\frac{\|\boldsymbol{\mu}^*\|_2^2}{32\sigma^2})$, we can narrow the upper bound (19) further:

$$\mathbb{E}(\gamma_1^1(\boldsymbol{X})) \leq \frac{2\alpha\sigma^2(1+\frac{1+r}{\sigma^2}\|\boldsymbol{\mu}^*\|_2^2)}{e\|\boldsymbol{\mu}^*\|_2\exp(\frac{\|\boldsymbol{\mu}^*\|_2^2}{32\sigma^2})} + \frac{\alpha(1+\frac{1+r}{\sigma^2}\|\boldsymbol{\mu}^*\|_2^2)\|\boldsymbol{\mu}^*\|_2}{4\exp(\frac{\|\boldsymbol{\mu}^*\|_2^2}{8\sigma^2})}. \quad (20)$$

Similarly, we have that

$$\mathbb{E}(\gamma_2^1(\boldsymbol{X})) \leq \mathbb{E}[\frac{\alpha\langle\boldsymbol{X},\boldsymbol{X}\rangle\boldsymbol{X}^T}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{X}}{\sigma^2})}|A]\mathbb{P}[A] + \mathbb{E}[\frac{\alpha\langle\boldsymbol{X},\boldsymbol{X}\rangle\boldsymbol{X}^T}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{X}}{\sigma^2})}|A^c]$$
$$\leq \frac{27\alpha\sigma^6}{4e^3\|\boldsymbol{\mu}^*\|_2^3}e^{-\frac{\|\boldsymbol{\mu}^*\|_2^2}{32\sigma^2}} + \frac{\alpha\|\boldsymbol{\mu}^*\|_2^3}{4^3}e^{-\frac{\|\boldsymbol{\mu}^*\|_2^2}{8\sigma^2}}$$
$$(21)$$

Therefore, Lemma 1 holds based on (20) and (21), where $\beta_1, \beta_2, \beta_3$ and β_4 are chosen properly.

A.2 Proof of Lemma 2

Similar to the proof in Corollaries 1, we have that

$$\Gamma^2_\delta(\boldsymbol{X}) = \mathbb{E}\underbrace{\left[\alpha\frac{1+\frac{\|\boldsymbol{\mu}_\delta\|_2^2}{\sigma^2}}{\exp(\frac{2\langle\boldsymbol{X},\boldsymbol{\mu}_\delta\rangle}{\sigma^2})}\right]}_{\gamma_1^2(\boldsymbol{X})} + \mathbb{E}\underbrace{\left[\frac{\alpha\langle\boldsymbol{X},\boldsymbol{X}\rangle}{\sigma^2\exp(\frac{2\langle\boldsymbol{X},\boldsymbol{\mu}_\delta\rangle}{\sigma^2})}\right]}_{\gamma_2^2(\boldsymbol{X})}. \quad (22)$$

To derive the uniform upper bounds of $\|\mathbb{E}(\gamma_1^2(\boldsymbol{X}))\|_{op}$ and $\|\mathbb{E}(\gamma_2^2(\boldsymbol{X}))\|_{op}$, let $\delta \in [0,1]$ be arbitrarily given and $e_1 \in \mathbb{R}^N$ denotes the first canonical basis vector. We can construct an orthonormal matrix, Q, such that $Q\boldsymbol{\mu}_\delta = \|\boldsymbol{\mu}_\delta\|_2 e_1$. Assume that $\boldsymbol{Y} = Q\boldsymbol{X}$, which makes $\boldsymbol{Y} \sim \mathcal{N}(Q\boldsymbol{\mu}^*, \sigma^2 I_N)$. Note that $\|\boldsymbol{\mu}_\delta\|_2 \leq (1+r)\|\boldsymbol{\mu}*\|_2$. Thus, we have $\|\mathbb{E}(\gamma_1^2(\boldsymbol{X}))\|_{op} \leq \mathbb{E}[\alpha\frac{1+(1+r)\frac{\|\boldsymbol{\mu}^*\|_2^2}{\sigma^2}}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{X}}{\sigma^2})}]$. Conditioned on $A := \{\boldsymbol{Y}_1 \leq \frac{\|\boldsymbol{\mu}^*\|_2}{4}\}$,

$$\|\mathbb{E}(\gamma_1^2(\boldsymbol{X}))\|_{op}$$
$$\leq \mathbb{E}[\alpha\frac{1+(1+r)\frac{\|\boldsymbol{\mu}^*\|_2^2}{\sigma^2}}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{X}}{\sigma^2})}|A]\mathbb{P}[A] + \mathbb{E}[\alpha\frac{1+(1+r)\frac{\|\boldsymbol{\mu}^*\|_2^2}{\sigma^2}}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{X}}{\sigma^2})}|A^c]$$
$$\leq \alpha(1+(1+r)\frac{\|\boldsymbol{\mu}^*\|_2^2}{\sigma^2})[\mathbb{P}[A] + \exp(-\frac{\|\boldsymbol{\mu}_\delta\|_2\|\boldsymbol{\mu}^*\|_2}{2\sigma^2})] \quad (23)$$
$$\leq \alpha(1+(1+r)\frac{\|\boldsymbol{\mu}^*\|_2^2}{\sigma^2})[e^{-\frac{\|\boldsymbol{\mu}^*\|_2^2}{32\sigma^2}} + e^{-\frac{\|\boldsymbol{\mu}^*\|_2^2}{8\sigma^2}}].$$

Based on the constructed orthonormal projection matrix, the operator norm of the matrix is shown as follows $\mathbb{E}(\gamma_2^2(\boldsymbol{X})) = \mathbb{E}\left[\frac{\alpha\langle\boldsymbol{X},\boldsymbol{X}\rangle}{\sigma^2\exp(\frac{2\langle\boldsymbol{X},\boldsymbol{\mu}_\delta\rangle}{\sigma^2})}\right] =: E$. Based on [1], it is easy to obtain that $E_{11} \leq \alpha\sigma^2(\frac{16\sigma^2}{9e^2\|\boldsymbol{\mu}^*\|_2^2}e^{-\frac{\|\boldsymbol{\mu}^*\|_2^2}{32\sigma^2}} + \frac{\|\boldsymbol{\mu}^*\|_2^2}{16\sigma^2}e^{-\frac{3\|\boldsymbol{\mu}^*\|_2^2}{8\sigma^2}})$ whenever $\|\boldsymbol{\mu}^*\|_2^2 \geq 16\sigma^2/3$. Moreover, for any index $j \neq 1$, we have $E_{jj} = \mathbb{E}[\frac{\alpha}{\exp(\frac{2\|\boldsymbol{\mu}_\delta\|_2\boldsymbol{Y}_1}{\sigma^2})}]$. Similar to the derivation of the uniform upper bound of $\|\mathbb{E}(\gamma_1^2(\boldsymbol{X}))\|_{op}$, we have $E_{jj} \leq 2\alpha\exp(-\frac{\|\boldsymbol{\mu}^*\|_2^2}{32\sigma^2})$. Combining the results on E_{11} and E_{jj}, we have that

$$\|\mathbb{E}(\gamma_2^2(\boldsymbol{X}))\|_{op} \leq \alpha(2 + \frac{16\sigma^2}{9e^2\|\boldsymbol{\mu}^*\|_2^2} + \frac{\|\boldsymbol{\mu}^*\|_2^2}{16\sigma^2})\exp(-\frac{\|\boldsymbol{\mu}^*\|_2^2}{32\sigma^2}). \quad (24)$$

Therefore, the corollary holds based on Equations (23) and (24), where λ_1, λ_2 and λ_3 are chosen properly.

B. PROOF OF COROLLARY 1

Define $g_\mu(\boldsymbol{X},\boldsymbol{\eta}) = \omega_\mu(\boldsymbol{X},\boldsymbol{\eta})(\boldsymbol{X} - \boldsymbol{\eta})$. Therefore, $M_n(\boldsymbol{\mu}) = \frac{1}{S}\sum_{s=1}^S g_\mu(\boldsymbol{X}^s,\boldsymbol{\eta}^s)$ and $M(\boldsymbol{\mu}) = \mathbb{E}(g_\mu)$. Thus, $\mathcal{G} = \{g_\mu|\boldsymbol{\mu} \in \Omega\}$. Define $X \triangleq \sup_{s\in[S]}\|\boldsymbol{X}^s\|_2$. Based on Theorem 1, we can show that $\|g - \mathbb{E}(g)\|_2 \leq X + C + \lambda r\|\boldsymbol{\mu}^*\|_2$, and $\mathbb{E}(g^2) \leq (X + C)^2$. As for the special case with $\tau = 0$ in Lemma 5.1 in [22], for $\epsilon > 0$ we have that

$$\mathbb{P}[\sup_{\boldsymbol{\mu}\in\Omega}\|M_n(\boldsymbol{\mu}) - M(\boldsymbol{\mu})\|_2 > 4\epsilon]$$
$$= \mathbb{P}[\sup_{g\in\mathcal{G}}\|\frac{1}{S}\sum_{s=1}^S g_\mu(\boldsymbol{X}^s,\boldsymbol{\eta}^s) - \mathbb{E}(g)\|_2 > 4\epsilon] \quad (25)$$
$$\leq cN(\Omega,\epsilon)e^{\frac{-S\epsilon^2}{2((C+X)^2+\frac{1}{3}(C+X+\lambda r\|\boldsymbol{\mu}^*\|_2)\epsilon)}},$$

where c is a positive constant. Define the right hand side as δ. Then we can derive that $\epsilon = c_{\delta,N(\Omega,\delta)}S^{-\frac{1}{2}}$ for some constant $c_{\delta,N(\Omega,\delta)}$.

C. PROOF OF PROPERTY 2

Define event $A_t = \{\|M_n(\boldsymbol{\mu}^t) - M(\boldsymbol{\mu}^t)\|_2 \leq c_{\delta,N(\Omega,\delta)}S^{-\frac{1}{2}}\}$ and $\mathcal{A} = \bigcap_{i=0}^t A_i$. Corollary 1 states that for each t we have that $\mathbb{P}[A_t \leq c_{\frac{\delta}{t},N(\Omega,\frac{\delta}{t})}S^{-\frac{1}{2}}] \geq 1-\delta$. Therefore, for any $\epsilon > 0$, we have that

$$\mathbb{P}[\|\boldsymbol{\mu}^{t+1} - \boldsymbol{\mu}^*\|_2 > \epsilon]$$
$$= \mathbb{P}[\|M_n(\boldsymbol{\mu}^t) - M(\boldsymbol{\mu}^t)\|_2 + \|M(\boldsymbol{\mu}^t) - \boldsymbol{\mu}^*\|_2 > \epsilon|\mathcal{A}]\mathbb{P}[\mathcal{A}]$$
$$+ \mathbb{P}[\|M_n(\boldsymbol{\mu}^t) - M(\boldsymbol{\mu}^t)\|_2 + \|M(\boldsymbol{\mu}^t) - \boldsymbol{\mu}^*\|_2 > \epsilon|\mathcal{A}^c]\mathbb{P}[\mathcal{A}^c]$$
$$\leq \mathbb{P}[\|M_n(\boldsymbol{\mu}^t) - M(\boldsymbol{\mu}^t)\|_2 + \|M(\boldsymbol{\mu}^t) - \boldsymbol{\mu}^*\|_2 > \epsilon|\mathcal{A}] + \mathbb{P}[\mathcal{A}^c]$$
$$\leq \mathbb{P}[\lambda^{t+1}\|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2 + \frac{1}{1-\lambda}c_{\frac{\delta}{t},N(\Omega,\frac{\delta}{t})}S^{-\frac{1}{2}} > \epsilon] + \sum_{i=0}^t\mathbb{P}[A_i^c]$$
$$\leq \mathbb{P}[\lambda^{t+1}\|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2 + \frac{1}{1-\lambda}c_{\frac{\delta}{t},N(\Omega,\frac{\delta}{t})}S^{-\frac{1}{2}} > \epsilon] + t\frac{\delta}{t} \leq \delta.$$

The first inequality holds because $\mathbb{P}(A) \leq 1, \forall A$. To obtain the second inequality, we first show that

$$\|\boldsymbol{\mu}^{t+1} - \boldsymbol{\mu}^*\|_2 = \|M_n(\boldsymbol{\mu}^t) - \boldsymbol{\mu}^*\|_2$$
$$\leq \|M_n(\boldsymbol{\mu}^t) - M(\boldsymbol{\mu}^t)\|_2 + \|M(\boldsymbol{\mu}^t) - \boldsymbol{\mu}^*\|_2$$

As $A_t \subseteq \bigcap_{i=0}^t A_i$, we have that $\|M_n(\boldsymbol{\mu}^t) - M(\boldsymbol{\mu}^t)\|_2 \leq \delta$. Applying the same procedure on $\|M(\boldsymbol{\mu}^t) - \boldsymbol{\mu}^*\|_2$, we can obtain the first part. The second part is easy to obtain by using the Boole's Inequality. The last inequality holds provided that $\epsilon \geq \lambda^{t+1}\|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2 + \frac{1}{1-\lambda}c_{\frac{\delta}{t},N(\Omega,\frac{\delta}{t})}S^{-\frac{1}{2}}$.

Therefore, in probability at least $1 - \delta$, we have that

$$\|\boldsymbol{\mu}^{t+1} - \boldsymbol{\mu}^*\|_2 \leq \lambda^{t+1}\|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2 + \frac{1}{1-\lambda}c_{\frac{\delta}{t},N(\Omega,\frac{\delta}{t})}S^{-\frac{1}{2}}. \quad (26)$$

which completes the proof.