Calibration-Free Network Localization using Non-Line-of-Sight Ultra-Wideband Measurements

Carmelo Di Franco[†], Amanda Prorok[⋆], Nikolay Atanasov[⋆], Benjamin Kempke[‡], Prabal Dutta[‡], Vijay Kumar[⋆], and George J. Pappas[⋆]

[†]ReTiS Laboratory
Scuola Superiore Sant'Anna
Pisa, Italy
c.difranco@santannapisa.it

[‡]EECS Department
University of Michigan
Ann Arbor, MI 48109, USA
{bpkempke,prabal}@umich.edu

[⋆]GRASP Laboratory
University of Pennsylvania
Philadelphia, PA 19104, USA
{prorok,atanasov,kumar,pappasg}@seas.upenn.edu

ABSTRACT

We present a method for calibration-free, infrastructure-free localization in sensor networks. Our strategy is to estimate node positions and noise distributions of all links in the network *simultaneously* — a strategy that has not been attempted thus far. In particular, we account for biased, non-line-of-sight (NLOS) range measurements from ultra-wideband (UWB) devices that lead to multi-modal noise distributions, for which few solutions exist to date. Our approach circumvents cumbersome a-priori calibration, allows for rapid deployment in unknown environments, and facilitates adaptation to changing conditions. Our first contribution is a generalization of the classical multidimensional scaling algorithm to account for measurements that have multi-modal error distributions. Our second contribution is an online approach that iterates between node localization and noise parameter estimation. We validate our method in 3-dimensional networks, *(i)* through simulation to test the sensitivity of the algorithm on its design parameters, and *(ii)* through physical experimentation in a NLOS environment. Our setup uses UWB devices that provide time-of-flight measurements, which can lead to positively biased distance measurements in NLOS conditions. We show that our algorithm converges to accurate position estimates, even when initial position estimates are very uncertain, initial error models are unknown, and a significant proportion of the network links are in NLOS.

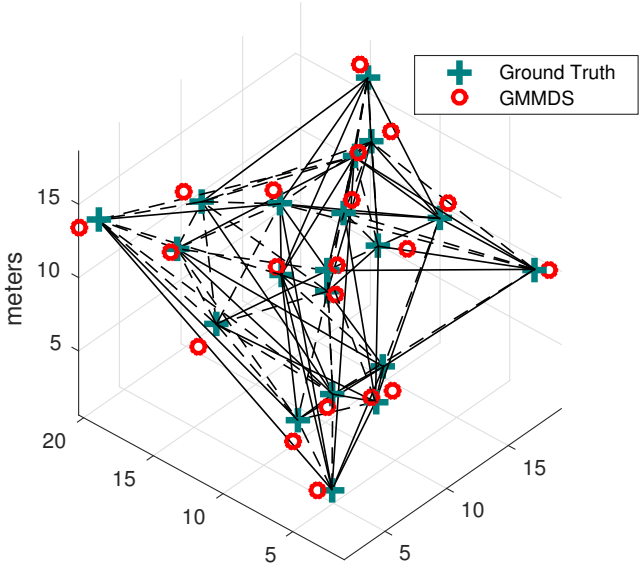


Figure 1: Actual and estimated node coordinates for a network with 20 nodes, 50% NLOS links (indicated by dashed lines) and average node degree 9.

CCS CONCEPTS

•Mathematics of computing →Maximum likelihood estimation; •Information systems →Location based services; Sensor networks; •Hardware →*Modeling and parameter extraction;*

KEYWORDS

sensor networks, localization, range measurement, non-line-of-sight, Gaussian mixture model, multidimensional scaling, expectation maximization

ACM Reference format:
Carmelo Di Franco[†], Amanda Prorok[⋆], Nikolay Atanasov[⋆], Benjamin Kempke[‡], and Prabal Dutta[‡], Vijay Kumar[⋆], and George J. Pappas[⋆]. 2016. Calibration-Free Network Localization using Non-Line-of-Sight Ultra-Wideband Measurements. In *Proceedings of The 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, Pittsburgh, PA USA, April 2017 (IPSN 2017),* 12 pages.
DOI: http://dx.doi.org/10.1145/3055031.3055091

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IPSN 2017, Pittsburgh, PA USA
© 2017 ACM. 978-1-4503-4890-4/17/04…$$15.00
DOI: http://dx.doi.org/10.1145/3055031.3055091

1 INTRODUCTION

Localizing the nodes of a sensor network is an essential technology with applications spanning the commercial, public, and military sectors. The success of such technologies, however, hinges on

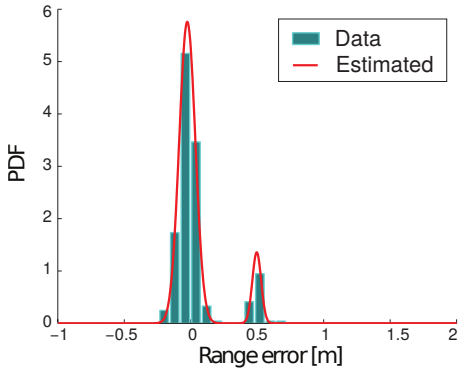


Figure 2: Range error between two SurePoint [21] devices based on a total of 303 time-of-flight measurements, collected inside a room with NLOS conditions. We superimposed the density of a multi-modal Gaussian (fitted on the underlying data).

the robustness of the underlying localization algorithm to arbitrary measurement noise distributions. Indeed, while we aim at engineering localization hardware with predictable measurement uncertainty, external environmental factors can have significant impact on signal propagation and, thus, affect measurement characteristics. Since these external factors are very hard to model, our goal is to develop methods that can handle arbitrary error distributions. In particular, we focus on time-of-flight based range-only localization in multi-path environments. In spite of sophisticated signal processing methods, the resolution of multi-path signals is still a hard problem [34]: when hardware-level algorithms fail to detect first signal paths, overestimated (positively biased) distance measures are returned. When aggregated, these biased range measurements lead to multi-modal error distributions. This is clearly seen in Fig. 2, which shows an example of experimental data. Consequently, we leverage statistical models that are able to capture NLOS measurements, allowing us to compute accurate and precise position estimates.

Formally, we consider the problem as an r-dimensional graph realization problem in a network of n sensors whose positions are unknown. Communication links are established bidirectionally between selected nodes, for which pairwise distance measurements, characterized by unique, unknown error distributions, are obtained. The problem of simultaneously estimating the node positions (and finding the unique graph realization), as well as estimating the parameters of the underlying error distributions is challenging — recent approaches assume either *(i)* knowledge of the underlying statistical data model, to then successfully localize the nodes [27, 30], or *(ii)* knowledge of the node positions to then successfully estimate the model parameters [20, 31]. Moreover, the joint localization and parameter estimation problem is compounded by distance measurements that are biased and have multi-modal error distributions.

This paper is the first to propose an approach that tackles network localization in absence of both *(i)* and *(ii)*, and, as a result, allows for rapid deployments in unknown and changing environments. Moreover, by considering multi-modal error models, we are capable of capturing NLOS behavior. This implicitly allows us to distinguish between line-of-sight (LOS) and NLOS links in our localization algorithm. Overall, our strategy is to progressively reduce the error on both the node coordinates and the noise distribution parameters for all links in the network by alternating online between the two estimation sub-problems, ultimately returning accurate node coordinates.

2 RELATED WORK

The literature on network localization algorithms is vast — comprehensive, recent summaries can be found in [19, 24]. Here, we only review selected methods that are most closely related to our work.

2.1 Graph Realization

The problem of finding Euclidean positions for the vertices of a graph given a set of edge lengths is known as the *graph realization problem*. This is a difficult problem for two main reasons: *(i)* the existence of a unique realization depends on the topology of the graph [16], and *(ii)* noise on the distance measurements compounds the difficulty of finding a correct realization. In range-based multilateration, the Euclidean distances between pairs of nodes are used to estimate the node coordinates. Aspen et al. [3] provide a theoretical foundation for network localization in terms of graph rigidity theory. The authors show that the network has a unique localization if and only if it is generically globally rigid. However, under the presence of noise in the range measurements, certain ambiguities cannot be resolved, which means that even when the graph is rigid, alternative incorrect realizations may result. Moore et al. [30] propose an algorithm that addresses this problem, and is capable of localizing a sensor network in the presence of range measurement noise. The authors assume an unbiased Gaussian noise distribution with known variance that is uniform all links in the network. They introduce the probabilistic notion of robust quadrilaterals as a way to avoid ambiguities that would otherwise corrupt localization computations. Similarly, several works present methods that exploit geometric relations between node-to-node-distances [8, 15, 23, 24]. Finally, the work in [38] proposes a method to jointly estimate target locations and received signal strength indicator (RSSI) model parameters. The approach differs from our work, however, since it makes use of fixed anchor nodes to estimate both locations and channel parameters, and refers to the specific case of RSSI.

In contrast to these methods, our approach is anchor-free, does not assume a-priori knowledge of the error distributions and allows different *multi-modal* distributions for each link in the network. Also, it is unclear how prior methods deal with biased, multi-modal error distributions that add ambiguities to their geometric assumptions. Many existing approaches model range errors as having a Gaussian distribution, whereas in practice ranging errors do not follow this behavior. In order to achieve accurate localization, we use a multi-modal distribution to capture the measurement characteristics accurately.

2.2 Multi-Dimensional Scaling

An alternative centralized distance-based approach builds on the theory of Multidimensional Scaling (MDS) [25] and formulates network localization as a least squares problem. MDS is a general technique that represents a set of elements in r-dimensional space

using a similarity measure between pairs of elements. When used for localization, similarity is measured via the Euclidean distance [7]. There exist several implementations of MDS, such as classical MDS, metric MDS, non-metric MDS, depending on the characteristics of the distance information. Variations consider a weighted version of distributed MDS, termed distributed weighted Multidimensional Scaling (dwMDS) [12], and a version that accounts for missing distances and local information [35, 36]. Overall, the collection of cited works operate under the assumption that the range measurements are sampled from unbiased Gaussian distributions with known variance. This can be problematic in environments that induce biased measurements; furthermore, even if the methods were to account for biased measurements, the problem of accurately calibrating the error models as a function of the underlying data is not only cumbersome, but near to impossible in changing environments. We tackle this problem by proposing an online method that improves nodes estimates while autonomously reducing the uncertainty of the error model parameters.

2.3 Ultrawideband Localization

UWB is an attractive sensing modality for distance-based localization. The main difference between UWB and other radio frequency signals is that UWB transmits its signal over multiple bands of frequencies simultaneously, in the interval of 3.1 to 10.6 GHz, exceeding a bandwidth of 500 MHz or 20% of the center frequency. In the time-domain, this is typically achieved through the transmission of short pulses (with a duration on the order of nanoseconds) which have large characteristics. Additionally, UWB systems typically run on very low duty cycles, and, thus, are very low power. In the context of localization, UWB features high positioning accuracy (due to the fine time resolution of the emitted signals) and high material penetrability (due to the large bandwidth). However, despite these desirable traits, UWB localization is challenged by NLOS scenarios that induce multi-path signal components [31, 34].

Methods that handle NLOS measurements can be classified into two groups: those that detect and discard NLOS measurements, using only LOS measurements [9, 11], and those that use NLOS measurements in addition to the LOS measurements [10, 37]. The approaches in the second group employ means of identifying which measurements are NLOS, and adapt the computations accordingly, e.g. by weighting or scaling those measurements. Our approach belongs to the second group, allowing us to take advantage of a larger amount of information, in particular when the majority of links in the network is in NLOS. Yet, we differ from the latter approaches in terms of our underlying assumptions. Whereas Chen et al. [10] assume that NLOS measurements are always significantly larger in magnitude than LOS measurements, we do not. In fact, our model captures arbitrary error behavior, and uses the estimated distribution parameters to scale the importance of network links. Venkatraman et al. [37] assume knowledge of the geometric layout of certain base station nodes, to infer bounds on the NLOS range errors that are obtained. This can be troublesome when the material properties of the nearby environment has non-negligible affects on the signal propagation. Similarly, in [26], the multi-path reflections of the signals are computed geometrically by knowing the locations of the anchors nodes and the floor plans of the building. In contrast to the latter approach, we assume no a-priori knowledge of the layout of the sensor nodes nor of the floor plan. We are, thus, capable of rapidly deploying sensor networks in unknown environments. Moreover, we fully rely on our procedures to perform system identification and calibration in real-time.

Previous work analyzes UWB error characteristics in indoor environments, and indicates that *(i)* the error behavior is multimodal, *(ii)* the multi-modal characteristic is preserved over different spatial scales, and *(iii)* the error behavior is different at different locations in space (and depends on the embedding of the UWB devices in a given environment) [32]. In spite of the complexity of the error patterns, past work discusses the suitability of a variety of statistical models with exponential behavior, supported on the semi-infinite interval $(0, \infty)$ [1, 33]. In particular, it was shown in several comprehensive measurement campaigns that error characteristics are well captured by a log-normal distribution [1, 31]. The proposed models are compact and efficient, as they capture the entirety of NLOS measurements in one statistical mode with a heavy tail. In [32], a real-time calibration of the error distribution parameters is performed. The approach uses the knowledge of the node and anchor locations to estimate a Gaussian-Lognormal mixture model. The latter work focuses on the calibration procedure, and does not address the real-time estimation of node positions. Furthermore, the disadvantage of the Gaussian-Lognormal mixture model is that it fails to capture multiple *precise* NLOS modes, when present. Building on this insight, we propose a more general approach that does not depend on a-priori information, and that performs a joint estimation of the node locations and error distributions. Indeed, in this work, we make use of a mixture of Gaussians to model the error distributions, which can effectively capture multiple, potentially narrow modes over biased NLOS measurements (representing multi-path propagation). Later, in Section 6.2, we show how Gaussian mixtures are well-suited models, producing good fits on experimental data (cf. Fig. 9).

2.4 Contributions

The novelty of our approach is summarized as follows: *(i)* we assume no prior knowledge of the noise characteristics of the UWB measurements (beyond the fact that NLOS measurements have positive bias), *(ii)* we estimate statistical noise models on a per link basis (since NLOS errors are location dependent [11]), *(iii)* we assume no prior knowledge of the layout of the nodes in our network, and finally, *(iv)* we use all collected measurements (NLOS as well as LOS), and assume no upper bound on the ratio of NLOS to LOS links. In light of the aforementioned characteristics, our work makes the following contributions: *(i)* a generalization of Multidimensional Scaling (MDS) that enables localization with Gaussian-mixture-distributed range measurements, typical in NLOS settings, *(ii)* an online algorithm that iterates between node localization and model parameter estimation and hence enables rapid calibration-free network deployment, and *(iii)* experimentation on a setup composed of 13 UWB radio nodes in an indoor environment with multipath and NLOS characteristics.

3 PROBLEM FORMULATION

Consider a collection of n sensors with positions $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]^\top \in \mathbb{R}^{n \times r}$. Suppose that the sensors can communicate with one another and represent the communication network by an undirected graph $G = (\mathcal{V}, \mathcal{E})$ with vertices $\mathcal{V} := \{1, \ldots, n\}$ and $|\mathcal{E}| = m$ edges. An edge $(i, j) \in \mathcal{E}$ from sensor i to sensor j exists if the two sensors can communicate. Let $d_{ij}(\mathbf{X}) = \|x_i - x_j\|$ be the distance between nodes i and j. We consider the problem of estimating $\mathbf{X}$ using noisy distance measurements:

$$\hat{d}_{ij}(t) := d_{ij}(\mathbf{X}) + v_{ij}(t) \tag{1}$$

collected at discrete times $t = 1, \ldots, T$ between pairs (i, j) of nodes that are in communication. The measurement noise is modeled using a random variable $v_{ij}(t)$ with a Gaussian-mixture probability density function (pdf) with K mixture components (see Fig. 2):

$$p(v_{ij}(t); \theta_{ij}) := \sum_{k=1}^{K} \alpha_{ij}^k \, \phi\left(v_{ij}(t); \mu_{ij}^k, \sigma_{ij}^k\right), \tag{2}$$

where α_{ij}^k are the mixture weights and $\phi\left(\,\cdot\,; \mu_{ij}^k, \sigma_{ij}^k\right)$ is a Gaussian pdf with mean μ_{ij}^k and standard deviation σ_{ij}^k. We denote the whole set of parameters associated with node pair (i, j) by $\boldsymbol{\theta}_{ij} = \left[\alpha_{ij}^1 \cdots \alpha_{ij}^K, \mu_{ij}^1 \cdots \mu_{ij}^K, \sigma_{ij}^1 \cdots \sigma_{ij}^K\right]$ and assume that the measurement noise v_{ij} is independent for any pair of times and across different sensors. To avoid any calibration steps needed to determine the noise parameters during network deployment, especially in unknown environments, the goal is to estimate the parameters $\boldsymbol{\theta} := \{\boldsymbol{\theta}_{ij}\}_{ij \in \mathcal{E}}$ online, concurrently with the sensor positions.

Problem (Joint Localization and Parameter Estimation). *Given range measurements* $\{\hat{d}_{ij}(t) \mid ij \in \mathcal{E}, t = 1, \ldots, T\}$, *determine the maximum likelihood estimates:*

$$\hat{\mathbf{X}}, \hat{\boldsymbol{\theta}} = \arg\max_{\mathbf{X}, \boldsymbol{\theta}} \sum_{t=1}^{T} \sum_{(i,j) \in \mathcal{E}} \log p\left(\hat{d}_{ij}(t) \mid d_{ij}(\mathbf{X}), \boldsymbol{\theta}_{ij}\right) \tag{3}$$

of the sensor positions $\mathbf{X}$ *and noise parameters* $\boldsymbol{\theta}$.

The goal of the above optimization is to determine the sensor positions $\mathbf{X}$ and noise parameters $\boldsymbol{\theta}$ that maximize the likelihood of the given noisy measurements using the noise model in (2). Note that once a graph realization is found (i.e., a solution to (3)), the network still needs to be embedded into a global coordinate system. In other words, the solution $\hat{\mathbf{X}}$ is isometry-invariant, and the node locations can be recovered only up to an unknown translation, rotation, and reflection. To remove this ambiguity in practice, it is necessary to deploy $r + 1$ anchors (nodes with known positions) in r-dimensions. In the rest of the paper, we do not explicitly assume the presence of anchors and instead evaluate our estimation results via a scale-free orthogonal Procrustes analysis (e.g., via the Kabsch algorithm [17, 18]). Besides needing $r + 1$ anchors to fix a global coordinate frame, we make no assumptions on the dimensionality of the problem and hence the approach we present below is applicable in arbitrary r dimensions.

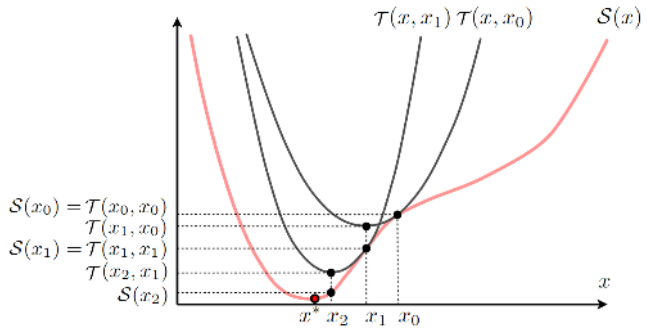


Figure 3: Uni-dimensional illustration of the SMACOF algorithm [7].

4 GENERALIZATION OF MULTIDIMENSIONAL SCALING

In this section, we begin with the simpler localization-only case in which the noise distribution parameters $\boldsymbol{\theta}$ are known. We present a non-trivial generalization of the classical MDS technique [7] for range-only network localization for the case of Gaussian-mixture-distributed noise as in (2). Rewriting (3) using the noise distribution in (2) and known $\boldsymbol{\theta}$ results in:

$$\max_{\mathbf{X}} \sum_{t=1}^{T} \sum_{(i,j) \in \mathcal{E}} \log \sum_{k=1}^{K} \alpha_{ij}^k \phi\Big(\hat{d}_{ij}(t) - d_{ij}(\mathbf{X}); \mu_{ij}^k, \sigma_{ij}^k\Big). \tag{4}$$

Note that with only one measurement per edge ($T = 1$), error modeled as a Gaussian distribution ($K = 1$) with zero mean ($\mu_{ij}^K = 0$), and measurement weight defined as $w_{ij} := (2\sigma_{ij}^K \sigma_{ij}^K)^{-1}$, the above is equivalent to the problem considered by weighted MDS:

$$\min_{\mathbf{X}} \mathcal{S}(\mathbf{X}) := \sum_{(i,j) \in \mathcal{E}} w_{ij} \left(\hat{d}_{ij} - d_{ij}(\mathbf{X})\right)^2. \tag{5}$$

In other words, (4) is a generalization of the MDS objective that captures multi-modal noise.

4.1 Review of Multidimensional Scaling

MDS is a general technique that estimates the coordinates of a set of elements in r-dimensional space by minimizing the mismatch between measured distances and the distances corresponding with the coordinate estimates. The mismatch is measured using a square error function (Eq. (5)), which is also called the *stress function*. Among the most successful approaches in terms of guarantees and convergence rate for optimizing (5) is an iterative method called Scaling by MAjorizing a COmplicated Function (SMACOF) [13]. Instead of minimizing the non-convex objective $\mathcal{S}(\mathbf{X})$ in (5), SMACOF minimizes a convex function $\mathcal{T}(\mathbf{X}, \mathbf{Z})$ that majorizes the stress function $\mathcal{S}(\mathbf{X})$ ($\mathcal{T}$ bounds $\mathcal{S}$ from above and touches its surface at a supporting point Z). The SMACOF algorithm is shown in Alg. 1 below and illustrated in the scalar case in Fig. 3. The following proposition establishes a majorizing function for the stress function $\mathcal{S}(\mathbf{X})$.

Algorithm 1 SMACOF

input: initial position estimate $\mathbf{X}^{(0)}$
stress function $\mathcal{S}$
majorizing function $\mathcal{T}$
repeat
$\mathbf{X}^k \leftarrow \min_{\mathbf{X}} \mathcal{T}(\mathbf{X}, \mathbf{X}^{k-1})$
until $\mathcal{S}(\mathbf{X}^{k-1}) - \mathcal{S}(\mathbf{X}^k) < \epsilon$

Proposition 4.1 ([7, Ch.8]). *For $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{n\times r}$, the function:*

$$\mathcal{T}(\mathbf{X}, \mathbf{Z}) := \sum_{(i,j)\in\mathcal{E}} w_{ij}\left(\hat{d}_{ij}^2 + d_{ij}^2(\mathbf{X}) - 2\hat{d}_{ij}\frac{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{z}_i - \mathbf{z}_j)}{\|\mathbf{z}_i - \mathbf{z}_j\|}\right)$$

is quadratic in $\mathbf{X}$ and $\mathcal{S}(\mathbf{X}) \le \mathcal{T}(\mathbf{X}, \mathbf{Z})$ for any $\mathbf{Z}$.

Proof. See Appendix A. □

4.2 Gaussian Mixture MDS

Now, consider the more general problem in (4) whose stress function is not convex:

$$\mathcal{F}(\mathbf{X}) := -\sum_{t=1}^{T}\sum_{(i,j)\in\mathcal{E}} \log \sum_{k=1}^{K} \alpha_{ij}^k \, \phi\left(\hat{d}_{ij}(t) - d_{ij}(\mathbf{X}); \mu_{ij}^k, \sigma_{ij}^k\right). \quad (6)$$

Similarly to SMACOF, we want to find a convex majorizing function which can be used to optimize the non-convex stress function $\mathcal{F}$ in (6) iteratively. The following proposition establishes a majorizing function for the Gaussian-mixture stress function above.

Proposition 4.2. *For $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{n\times r}$, the function:*

$$\overline{\mathcal{T}}(\mathbf{X}, \mathbf{Z}) := \sum_{t=1}^{T}\sum_{(i,j)\in\mathcal{E}}\sum_{k=1}^{K} \frac{\alpha_{ij}^k}{2\sigma_{ij}^k\sigma_{ij}^k}\left(\left(\hat{d}_{ij}(t) - \mu_{ij}^k\right)^2 + d_{ij}^2(\mathbf{X}) - 2\left(\hat{d}_{ij}(t) - \mu_{ij}^k\right)\frac{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{z}_i - \mathbf{z}_j)}{\|\mathbf{z}_i - \mathbf{z}_j\|}\right) + \sum_{t=1}^{T}\sum_{(i,j)\in\mathcal{E}}\sum_{k=1}^{K}\alpha_{ij}^k \log\left(\sigma_{ij}^k\sqrt{2\pi}\right)$$

is quadratic in $\mathbf{X}$ and $\mathcal{F}(\mathbf{X}) \le \overline{\mathcal{T}}(\mathbf{X}, \mathbf{Z})$ for any $\mathbf{Z}$.

Proof. See Appendix B. □

The main utility of the majorizing function $\overline{\mathcal{T}}(\mathbf{X}, \mathbf{Z})$ in Prop. 4.2 is that it allows us to apply the SMACOF algorithm to the multi-modal range-only localization problem in (4). The ability to handle multi-modal noise enables applications in much more general settings than those amenable to the classical MDS algorithm. In particular, as we demonstrate in Sec. 6, the new Gaussian mixture algorithm allows us to address UWB radio localization, in which LOS measurements have (uni-modal) Gaussian noise while NLOS measurements may have multi-modal noise [31]. Note that the derivative of the majorizing function $\overline{\mathcal{T}}$, which is provided in Appendix C, can be used to optimize the GM-MDS objective via an existing MDS implementation.

5 JOINT LOCALIZATION AND PARAMETER ESTIMATION

Before returning to the general problem in (3), we consider the other subproblem – that of estimating the parameters $\boldsymbol{\theta}$ when the sensor node positions $\mathbf{X}$ are known. As commonly done in Gaussian-mixture parameter estimation, our idea is to introduce a discrete latent variable that specifies which of the mixture components $k = 1, \ldots, K$ generated a given range measurement $\hat{d}_{ij}(t)$ between nodes (i, j). The probability mass function $q_{ij}(t, k)$ of this latent variable specifies the likelihood that measurement $\hat{d}_{ij}(t)$ came from mixture component k and is commonly refereed to as the *membership* probability $q_{ij}(t, k)$. The expectation-maximization (EM) algorithm [5, 14] is an efficient way to estimate the parameters $\boldsymbol{\theta}_{ij}$ by iteratively computing the membership probabilities $q_{ij}(t, k)$ and updating the parameter estimates based on $q_{ij}(t, k)$. More precisely, starting from an initial guess $\hat{\boldsymbol{\theta}}_{ij}$, the EM algorithm iterates the following two steps for each node pair:

(1) **E step**: compute the membership probabilities:

$$q_{ij}(t,k) = \frac{\hat{\alpha}_{ij}^k\phi\left(\hat{d}_{ij}(t) - d_{ij}(\mathbf{X}); \hat{\mu}_{ij}^k, \hat{\sigma}_{ij}^k\right)}{\sum_l \hat{\alpha}_{ij}^l\phi\left(\hat{d}_{ij}(t) - d_{ij}(\mathbf{X}); \hat{\mu}_{ij}^l, \hat{\sigma}_{ij}^l\right)} \quad (7)$$

(2) **M step**: update the parameter estimates using the soft mixture component assignments $q_{ij}(t, k)$:

$$\begin{aligned}\hat{\alpha}_{ij}^k &\leftarrow \frac{1}{T}\sum_{t=1}^{T} q_{ij}(t,k)\\ \hat{\mu}_{ij}^k &\leftarrow \frac{\sum_{t=1}^{T}\left(\hat{d}_{ij}(t) - d_{ij}(\mathbf{X})\right)q_{ij}(t,k)}{\sum_{t=1}^{T} q_{ij}(t,k)}\\ \hat{\sigma}_{ij}^k &\leftarrow \frac{\sum_{t=1}^{T}\left(\hat{d}_{ij}(t) - d_{ij}(\mathbf{X}) - \hat{\mu}_{ij}^k\right)^2 q_{ij}(t,k)}{\sum_{t=1}^{T} q_{ij}(t,k)}\end{aligned} \quad (8)$$

The iterations in (7) and (8) allow us to refine an initial parameter estimate $\boldsymbol{\theta}_{ij}$ and determine an accurate model fit to a measurement set that may contain biased, NLOS measurements. Returning to the original problem in (3), we note that the multi-variable objective function in $\mathbf{X}$ and $\boldsymbol{\theta}$ can be optimized via *coordinate descent*, which maximizes it along a single direction at a time. More precisely, starting with initial estimates $\mathbf{X}^{(0)}$, $\boldsymbol{\theta}^{(0)}$, we can fix $\mathbf{X}^{(0)}$ and optimize over $\boldsymbol{\theta}$ and then fix $\boldsymbol{\theta}^{(1)}$ and optimize over $\mathbf{X}$ and so on. The first step results in the the parameter estimation problem that can be addressed via EM as discussed above, while the second step results in the localization problem in (4), which we can address via the Gaussian mixture MDS algorithm developed in Sec. 4. This procedure is summarized in Alg. 2.

One aspect that we left vague is the initialization phase of Alg. 2. Since both the GM-MDS and the EM algorithms provide only local convergence guarantees, the choice of $\mathbf{X}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$ determines the performance of Alg. 2. There exist several reliable approaches for initializing the EM algorithm [4, 6, 29]. We adopted the method proposed by Blömer and Bujna [6], which first selects candidate means via the k-means++ algorithm [2] and then converts them into Gaussian mixture parameters via [6, Alg.1]. Note that in Alg. 2, the EM is re-initialized at each coordinate descent step. After a GM-MDS step, the optimal parameter choice might be significantly

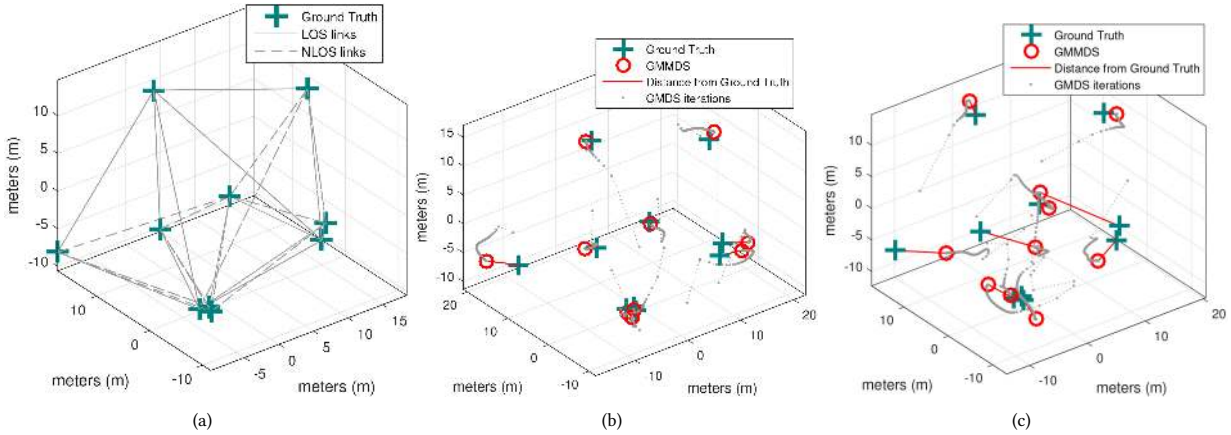


Figure 4: An example showing how multiple initialization points for the GM-MDS algorithm lead to different local minima: (a) shows a network of $n = 10$ nodes with average connectivity $\bar{D} = 5$ and LOS percentage 0.5, (b) shows the trajectories (gray) taken by the GM-MDS node estimates for an initial condition that leads to accurate localization, (c) shows an unlucky initialization of GM-MDS that leads to a local minimum.

Algorithm 2 Coordinate descent for joint localization and parameter estimation

input: measurements $\{\hat{d}_{ij}(t)\}$ for $ij \in \mathcal{E}$, $t = 1, \dots, T$
initial estimates $\mathbf{X}^{(0)}$
number of mixture components K
number of iterations L

for $l = 0, \dots, L-1$ **do**

$\boldsymbol{\theta}^{(l)} \leftarrow InitEM\left(\{\hat{d}_{ij}(t)\}, \mathbf{X}^{(l)}, K\right)$ via [4, 6, 29]

$\boldsymbol{\theta}^{(l+1)} \leftarrow EM\left(\{\hat{d}_{ij}(t)\}, \mathbf{X}^{(l)}, \boldsymbol{\theta}^{(l)}\right)$ via (7),(8)

$\mathbf{X}^{(l+1)} \leftarrow GM\text{-}MDS\left(\{\hat{d}_{ij}(t)\}, \mathbf{X}^{(l)}, \boldsymbol{\theta}^{(l+1)}\right)$ via Alg.1 and Prop.4.2

return $\{\mathbf{X}^{(L)}, \boldsymbol{\theta}^{(L)}\}$

different from the previous iteration and hence to avoid getting stuck in a local minimum in parameter space, we reinitialize the EM algorithm. If the GM-MDS step did not change the position estimate $\hat{\mathbf{X}}$ by much, this reinitialization of EM will cause a temporary worse estimate but will be optimized by the subsequent EM steps (see Fig. 5(c) for an example). However, if the GM-MDS changes the position estimate $\hat{\mathbf{X}}$ significantly, the re-initialization of EM may improve the performance.

Besides reinitializing the EM algorithm, it is also important to try several choices of $\mathbf{X}^{(0)}$. Fig. 4 illustrates this. If any prior information about the network deployment is available, it should be used to select a reasonable guess for $\mathbf{X}^{(0)}$. Otherwise, one could repeat the whole coordinate descent (Alg. 2) several times with different random initializations of $\mathbf{X}^{(0)}$ and select the the final estimate that obtains the minimum value of the stress function in (6). A slightly more efficient approach of re-initialization would be to repeat the GM-MDS algorithm several times only in the first coordinate descent step ($l = 0$ in Alg. 2) and to choose the estimate $\mathbf{X}^{(1)}$ with the minimum value of the stress function $\mathcal{F}$.

6 EVALUATION

To evaluate the practical feasibility of the proposed approach, we implemented it on an UWB system that relies on time-of-flight measurements to return distance values. First, we present simulations that test the performance of our algorithm and its sensitivity to varying conditions (network connectivity, sample size T, and percentage of links in LOS). Second, we present experimental results obtained using range measurements from UWB devices in an indoor environment.

6.1 Simulations

Our approach does not rely on prior knowledge of the node coordinates, nor of the error distributions for any links in the network. We ran simulations to test its sensitivity to: *(i)* the number of nodes n, *(ii)* the average degree of connectivity of the network $\bar{D}$, and *(iii)* the percentage of links in LOS, p_{LOS}.

For each simulation, we iterate over a collection of random 3-D networks with n nodes, for which the connectivity is defined by the average edge degree $\bar{D}$. Each edge (i, j) samples T range measurements from a unique tri-modal distribution with true parameters $\boldsymbol{\theta}_{ij}$ (unknown to our algorithm except that $K = 3$ is known), which are assigned as follows. First, we define a fixed percentage of LOS links, p_{LOS}. The true distribution for LOS edges was generated according the following values:

$$\alpha_{ij} = [1, 0, 0],\ \mu_{ij} = [0, 0, 0],\ \sigma_{ij} = [0.2, 0, 0],$$

whose choice was motivated by the empirical distribution of real UWB range measurements. We assumed that distributions of all the LOS links in the network are the same. For NLOS edges, the mean values were sampled uniformly at random in a neighborhood of ± 0.5 of $\mu_{ij} = [0, 2, 2.4]$. The standard deviations were sampled uniformly at random in the interval [0,1], and are then additionally biased by $\sigma_{ij} = [0.2, 0.5, 0.6]$. The mixture weights were sampled uniformly at random in the interval [0,1], and normalized such that the sum equals 1. The choice of standard deviations for the

components of the NLOS edge noise was based on average values observed in real UWB data, while the bias μ_{ij} introduced by the different noise modes was inflated compared to real data[1] to test the robustness of our algorithm.

We evaluated the performance of our approach by computing the the Root Mean Square Error (RMSE) between true and estimated node positions,

$$\mathrm{RMSE}\left(\mathbf{X}, \hat{\mathbf{X}}\right) := \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2}. \quad (9)$$

We also evaluated the accuracy of our estimated model parameters $\hat{\boldsymbol{\theta}}$ by comparing them to the true parameters $\boldsymbol{\theta}$ using the Kullback-Leibler Divergence,

$$D_{\mathrm{KL}}\left(\boldsymbol{\theta}\middle\|\hat{\boldsymbol{\theta}}\right) := \frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}} \int p(v;\boldsymbol{\theta}_{ij}) \log \frac{p(v;\boldsymbol{\theta}_{ij})}{p(v;\hat{\boldsymbol{\theta}}_{ij})} \mathrm{d}v, \quad (10)$$

where the probability densities $p(\cdot;\boldsymbol{\theta}_{ij})$ are defined in (2). Unless otherwise specified, our default simulation setup employs 30 nodes located in a 3-D work-space of dimension $30 \times 30 \times 30$ m^3, with average connectivity $\bar{D} = 15$, $T = 200$ measurements, and $p_{LOS} = 0.5$. Initial node coordinates $\mathbf{X}^{(0)}$ were sampled randomly from within this work-space. For each simulation case, the coordinate descent (Alg. 2) was run 20 times with different initial $\mathbf{X}^{(0)}$ and, among those, the final position estimate that minimizes the stress function $\mathcal{F}$ in (6) was chosen. Note that, since our approach is calibration-free, it does not require an initial estimate of the parameters $\boldsymbol{\theta}^{(l)}$. However, since only local convergence is guaranteed by GM-MDS for the node positions, better initializations $\mathbf{X}^{(0)}$ will lead to better overall performance. This is illustrated in Fig. 4 where two different starting points were used to perform localization in a network with $n = 10$ nodes, average connectivty of $\bar{D} = 5$, and LOS percentage of 0.5. In Fig 4 (c), the initial location estimates lead the GM-MDS algorithm to a local minimum, while a luckier choice in Fig 4 (b) leads to more accurate estimates. For this reason, localization techniques with local convergence guarantees such as MDS and GM-MDS rely on repeated initializations with different starting points to achieve a better estimate. The stress function in (6) can be used to judge the quality of the different initializations. To bound the computation time of our algorithm we used a maximum number of 400 GM-MDS iterations and 50 EM iterations.

Fig. 5 shows the performance of our algorithm for a varying proportion of links in LOS. Fig. 5(a) shows the convergence properties over 3 iterations of the coordinate descent algorithm. The error is shown as a function of GM-MDS iterations (with executions of EM at iterations 0, 400, 800, and 1200). The RMSE demonstrates successful localization, despite initially high uncertainty and significant NLOS — for 50% of links in LOS, and considering a modest connectivity of $\bar{D} = 15$, our algorithm performs robustly, with a mean error below 1 m. Fig. 5(b) evaluates the sensitivity of our approach to average node connectivity[2]. As expected, the higher the connectivity, the lower the error. This is due to the reduction of ambiguities present when the graphs are not rigid. It is noteworthy that the proportion of LOS links only has a marginal affect on performance. Finally, Fig. 5(c) demonstrates the ability of the EM algorithm to produce good parameter estimates. The Kullback-Leibler divergence is shown as a function of EM iterations, for four executions of GM-MDS (at iterations 50, 100, 150). Its performance is affected by the LOS proportion — less NLOS allows for better initial positioning, which subsequently allows EM to more accurately calibrate the noise models. The initial ripple at every group of 50 iterations is caused by the re-initialization of the distribution parameters, as suggested in Alg. 2 and discussed in Sec. 5.

The simulations were implemented in MATLAB® on a Core-I7-4770K at 3.50GHz. The complete localization procedure (Alg. 2) with 3 coordinate descent steps, each involving 400 GM-MDS and 50 EM iterations, takes on the order of a few seconds to compute the node positions and estimate the noise parameters. In detail, the 1200 GM-MDS iterations take approximately 30 milliseconds per node, while the 150 EM iterations take approximately 10 milliseconds per edge. For instance, in a network with 50 nodes and average node connectivity of $\bar{D} = 5$, the complete localization procedure takes about 2.2 sec. of which half is due to the GM-MDS iterations and the other half to the EM iterations.

Finally, we ran a simulation to understand the effect of the measurement sample size T on the accuracy of the localization system. The results are summarized in Fig. 6 and suggest that networks with networks with larger percentages of NLOS links benefit from larger measurement sample sizes. However, in all scenarios the improvement in performance due to an increasing number of measurements becomes less evident after about 100 samples.

6.2 Experiments

Our experimental setup uses SurePoint localization devices [21]. SurePoint nodes leverage the DecaWave DW1000 UWB transceiver in conjunction with multiple antennas and RF channels to achieve high ranging accuracy in the presence of heavy multi-path interference. Prior work [21] has shown that 50% of ranging estimates are within 8 cm and 95% of ranging estimates are within 31 cm in predominantly-LOS environments. All SurePoint devices are equivalent in terms of their hardware and can execute the role of either an anchor or a tag, where 'tag' is defined as those nodes which initiate ranging operations with nearby anchors. One 'coordinator' node provides global time synchronization and scheduling functionality through the use of UWB floods. In order to produce range measurements between all nodes of the system, we program each node to request a time-slot for ranging operations, effectively rotating tag functionality through each node participating in the network.

Experiment 1. Our first experiment was conducted in an approximately rectangular $4.6 \times 7.2 \times 2.7$ m^3 room in a commercial building. Ground truth node positions were measured using a laser rangefinder [3]. We deployed 8 UWB devices, as shown in Fig. 7, and schematized in Fig. 8(a), considering only the subset of nodes that are placed in the closed space.

Experiment 2. Our second experiment was an extension of Experiment 1, with a total of 13 UWB devices. The additional 4 nodes were installed in locations with heavy multi-path characteristics,

[1] We observed 2 to 3 modes in real UWB distributions with biases around 0, 0.5, and 1 meters, respectively. See Fig. 9 for details.

[2] This experiment was performed on random 2-D networks.

[3] https://www.boschtools.com/us/en/boschtools-ocs/measuring-tools-and-surveying-equipment-23413-c/

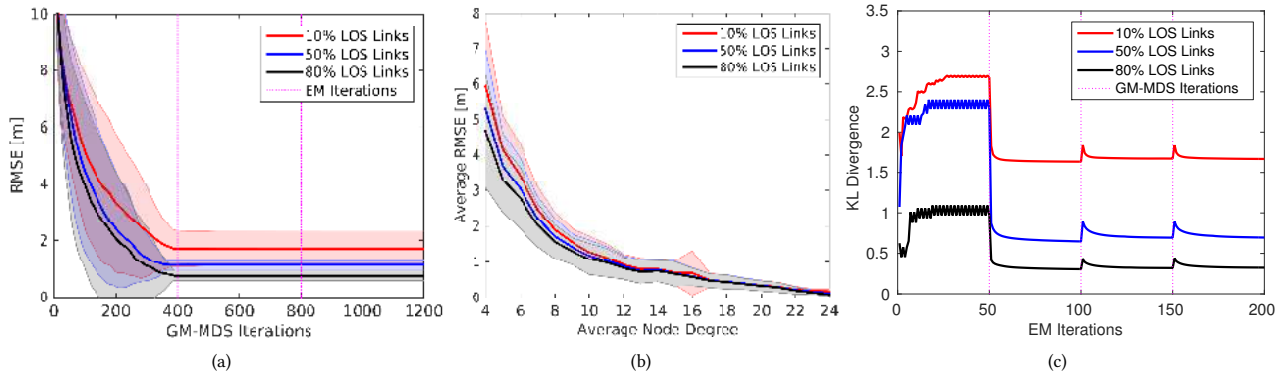


Figure 5: Performance of GM-MDS, for a varying percentage of links in LOS, $p_{LOS} \in \{10\%, 50\%, 80\%\}$. Panels (a) and (b) show the root-mean-square displacement between true and estimated positions (after application of the Procrustes transformation). The results are averaged over 50 random networks, with the standard deviation shown by the shaded area. (a) Results as a function of GM-MDS iterations. (b) Results as a function of average node degree. The RMSE demonstrates low sensitivity of our approach to varying numbers of NLOS edges and robust performance even for low network connectivities. Panel (c) shows the Kullback-Leibler divergence between estimated and true error distributions, for a single run (one random network). The spikes at every 50 iterations are caused by the re-initialization of EM (see Sec. 5 for details).

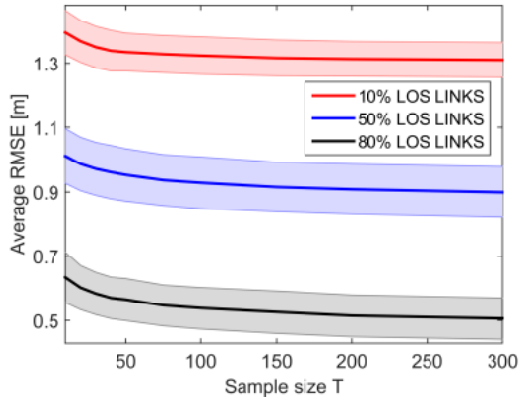


Figure 6: Performance of GM-MDS for varying percentage of LOS nodes, $p_{LOS} \in \{10\%, 50\%, 80\%\}$, and varying measurement sample size T. The improvement in performance due to increasing sample size becomes less evident after 100 samples.

causing NLOS signal propagation. The average node degree of the network was $\bar{D} = 8$. The layout is schematized in Fig. 8(a), where we see that additional nodes are placed on the opposite wall of the hallway, as well as on the far left wall of the office space (additional node IDs are 37, 1F, 38, and 2C).

Fig. 8(b) shows the quantitative localization performance after 3 coordinate descent iterations with 50 EM and 400 GM-MDS iterations, respectively, per coordinate descent step. The final RMSE of our approach, GM-MDS, in Experiment 1 and 2 is 0.32 m and 0.34 m, respectively. We compared the performance against a standard weighted MDS approach (eq. (5) and Alg. 1 using a (uni-modal) Gaussian distribution), whose weights were computed based on the empirical variance of the range measurements, i.e., $w_{ij} := (2\mathbf{var}(\{\hat{d}_{ij}\}))^{-1}$. Although this weighted MDS approach does not employ accurate parameter estimation, its choice of weights discounts the effect of NLOS edges with large measurement variance.

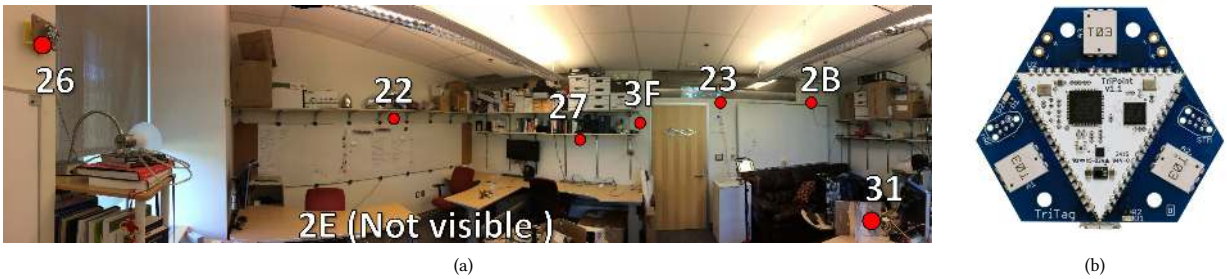
The weighted MDS algorithm performs as well as our coordinate descent approach in Experiment 1 due to the prevalence of LOS links. However, notice that its performance degrades significantly in Experiment 2 due to the presence of the NLOS measurements. This highlights the benefit of using careful parameter estimation and hence taking advantage of biased NLOS measurements in environments that are prone to induce multi-path fading. Both experiments demonstrate the ability of our method to localize nodes accurately, with no overhead in terms of a-priori system identification and calibration. To illustrate the effect of lower network connectivity on the localization accuracy, experiment 2 was repeated for several choices of average node degree. Fig. 8(c) shows that the performance of weighted MDS degrades significantly as the average node degree decreases, while GM-MDS remains robust due to its ability to incorporate information from NLOS measurements.

Finally, Fig. 9 shows actual data gathered during Experiment 2. We superimposed the density of a multi-modal Gaussian estimated by Expectation Maximization. The modes of the empirical distribution are captured well by this error model. In fact, the high fidelity of this representation is key to refining the position estimates of nodes providing NLOS measurements. In contrast, unbiased uni-modal models will tend to overestimate the inter-node distances, ultimately leading to alternate (and incorrect) graph realizations.

7 DISCUSSION

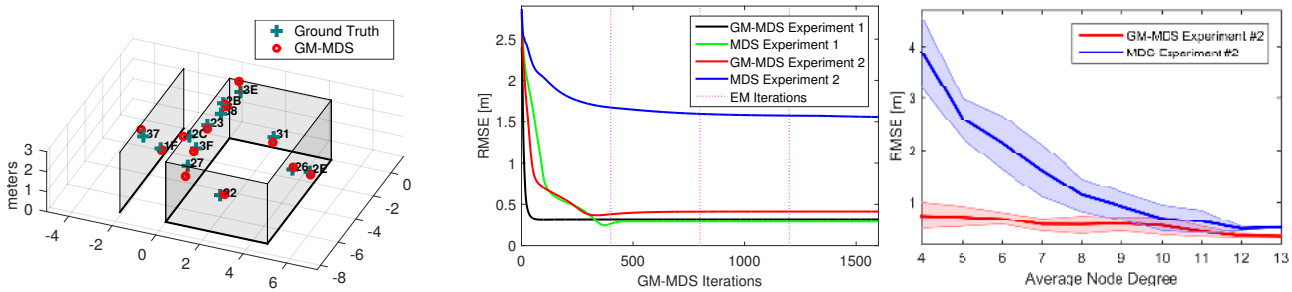
7.1 Partial Connectivity & Decentralization

An important question is how to implement the Gaussian mixture MDS in networks that are not fully-connected, or in a distributed setting, where sensor nodes receive range measurements from their one-hop neighbors only. Partial connectivity arises in networks for which *(i)* distance measurements $\hat{d}_{ij}$ are not available for some



(a) (b)

Figure 7: (a) Photo of experimental space with UWB devices. (b) SurePoint Hardware includes a DW1000 UWB transceiver and 3 antennas. It is composed of a TriPoint module (that implements the SurePoint system) and a Tritag Carrier Board that includes a BLE interface.



(a) (b) (c)

Figure 8: Experiment 2: We deployed 13 UWB SurePoint devices in an indoor environment with heavy multi-path characteristics. The nodes covered the space of one office and a partial hallway area. (a) Illustration of localization performance. The red lines indicate data associations between ground truth and estimates. (b) RMSE for Experiment 1 and Experiment 2, implemented with our approach, GM-MDS, and classical MDS as a benchmark. The final RMSE of GM-MDS (Alg. 2) in Experiment 1 and 2 is 0.32 m and 0.34 m, respectively. The final RMSE of weighted MDS in Experiment 1 and 2 is 0.3 m and 0.55 m, respectively. The cusps in the GM-MDS RMSE curve are due to the EM iterations altering the noise parameter estimates. (c) Performance as a function of average node degree. The performance of weighted MDS degrades significantly as the average node degree decreases, while GM-MDS remains robust due to its ability to incorporate information from NLOS measurements.

node pairs (i, j), or *(ii)* in strongly connected topologies with *uni-directional* links (e.g., a node i may receive measurements for the link (i, j), but not node j). The basic solution, as presented here, is to simply ignore the missing distances and sum the log-likelihood in (4) only over the available edges $\mathcal{E}$. Another possibility is to obtain a rough estimate for the missing measurements, e.g., by using the topology of the network or the number of hops. A more complex approach is map stitching [35, 36], which builds a local map at each node of the immediate neighbors and then merges maps together to form a global map. Finally, we note that our approach is compatible with variant MDS solutions that allow the integration of the majorizing function we derived in Prop. 4.2. In particular, we can integrate our method with the decentralized dwMDS algorithm [12] to produce a fully-distributed solution for the localization problem with Gaussian mixture noise.

7.2 Number of Mixture Components

Although Gaussian mixtures allow arbitrary modeling precision, using large values for the number of modes, K, may lead to a number of pitfalls. Firstly, models with large K are prone to over-fitting. Previous results compare the performance of localization algorithms using *(i)* histogram distributions and *(ii)* bi-modal distributions, and show that the former model leads to poor performance due to over-fitting [32]. Secondly, models with large K require more data to ensure that the model captures true environmental traits. Generally, the value of K depends on the environment's multipath conditions and the amount of available measurement data, and hence, must be determined by in-field measurements. One approach is to start with a small number of modes K, and to gradually increase this number until over-fitting becomes apparent. For instance, from the data collected in our experiments (See Fig. 9, as an example) it is possible to see that UWB measurements can be modeled with enough precision with $K = 3$.

7.3 Dynamic Environments

The validity of our model still needs to be investigated for highly dynamic environments. On the one hand, one could assume that moderately busy environments would only produce very sporadic

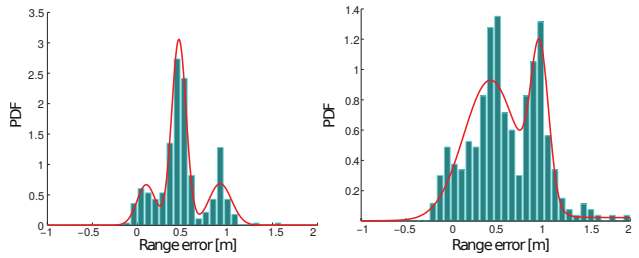


Figure 9: Selected error distributions from Experiment 2. Gaussian mixture probability densities estimated by EM algorithm are superimposed onto gathered data, represented by histograms. (a) Range between devices 2C and 38, based on a total of 352 measurements. (b) Range error between devices 27 and 2B, based on a total of 333 measurements. (c) Range error between devices 23 and 27, based on a total of 303 measurements. These data clearly demonstrate that using a single Gaussian distribution to model UWB range measurements might not accurately capture the noise characteristics and will tend to overestimate the inter-node distances, ultimately leading to inaccurate localization results.

anomalies, and that the overall localization quality would be maintained for infrequent model updates. On the other hand, since our iterative solution is capable of calibrating the measurement model continuously, we can readily produce a dynamic sequence of models so that the network remains faithful to time-varying error statistics. Moreover, the initialization of subsequent coordinate descent runs can be bootstrapped by using the final parameter estimates of the current coordinate descent run.

8 CONCLUSION

In this work, we introduced an approach for network localization that uses UWB range measurements to simultaneously estimate the node positions and the parameters of the measurement distributions. In order to account for biased, NLOS measurements, we modeled the noise distributions via Gaussian mixtures and generalized the classical MDS algorithm to handle such distributions. Coordinate descent was used to combine the resulting GM-MDS localization algorithm with an EM algorithm for online estimation of the mixture parameters. We demonstrated through simulations and physical experiments with UWB radios that the proposed approach provides robust calibration-free network localization in the presence of partial connectivity and NLOS measurements. We compare our approach to an alternative implementation of our coordinate descent algorithm that employs simple Gaussian error models, and hence, does not capture the multi-modality of error distributions. Our results highlight the benefit of our approach in environments with heavy multi-path characteristics. In summary, our method allows for rapid deployment of network nodes in environments that are unknown a-priori and that do not necessarily guarantee LOS conditions.

In future work, we plan on applying our algorithms to the tracking of multiple mobile targets in potentially dense, cluttered and dynamic environments, extending the results presented in [22]. We also plan on evaluating the applicability of our method to Received Signal Strength (RSS) based measurement techniques, such as in [28].

9 ACKNOWLEDGMENTS

We are grateful to Gian Pietro Picco for helping us improve the readability and quality of this manuscript.

This work was supported in part by the TerraSwarm Research Center, one of six centers supported by the STARnet phase of the Focus Center Research Program (FCRP) a Semiconductor Research Corporation program sponsored by MARCO and DARPA.

APPENDIX A: PROOF OF PROPOSITION 4.1

Expanding the stress function as follows:

$$\mathcal{S}(\mathbf{X}) = \sum_{(i,j)\in\mathcal{E}} w_{ij}\left(\hat{d}_{ij}^2 + d_{ij}^2(\mathbf{X}) - 2\hat{d}_{ij}d_{ij}(\mathbf{X})\right),$$

we notice that the first term is constant and the second term is quadratic in $\mathbf{X}$. For any $\mathbf{Z} \in \mathbb{R}^{n\times r}$, the problematic non-convex third term can be bounded via the Cauchy-Schwarz inequality as follows:

$$2\hat{d}_{ij}d_{ij}(\mathbf{X})\frac{\|\mathbf{z}_i - \mathbf{z}_j\|}{\|\mathbf{z}_i - \mathbf{z}_j\|} \geq 2\hat{d}_{ij}\frac{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{z}_i - \mathbf{z}_j)}{\|\mathbf{z}_i - \mathbf{z}_j\|}$$

APPENDIX B: PROOF OF PROPOSITION 4.2

To find a convex function that majorizes the Gaussian-mixture stress function in (6), we exploit the concavity of log to exchange the summation over the mixture components and the log function:

$$\begin{aligned}\mathcal{F}(\mathbf{X}) \leq &-\sum_{t=1}^{T}\sum_{(i,j)\in\mathcal{E}}\sum_{k=1}^{K}\alpha_{ij}^k \log\phi\left(\hat{d}_{ij}(t) - d_{ij}(\mathbf{X}); \mu_{ij}^k, \sigma_{ij}^k\right)\\ &= \sum_{t=1}^{T}\sum_{(i,j)\in\mathcal{E}}\sum_{k=1}^{K}\frac{\alpha_{ij}^k}{2\sigma_{ij}^k\sigma_{ij}^k}\left(\hat{d}_{ij}(t) - \mu_{ij}^k - d_{ij}(\mathbf{X})\right)^2\\ &\quad + \sum_{t=1}^{T}\sum_{(i,j)\in\mathcal{E}}\sum_{k=1}^{K}\alpha_{ij}^k\log\left(\sigma_{ij}^k\sqrt{2\pi}\right).\end{aligned}$$

The above transformation results in a simpler function that upper-bounds $\mathcal{F}$ but is unfortunately still non-convex. Fortunately, the problematic non-convex term is exactly the same as in the case of the MDS stress function majorization in Prop. 4.1. Proceeding as in the proof of Prop. 4.1, we expand the term:

$$\begin{aligned}\left(\hat{d}_{ij}(t) - \mu_{ij}^k - d_{ij}(\mathbf{X})\right)^2 = &\left(\hat{d}_{ij}(t) - \mu_{ij}^k\right)^2 + d_{ij}^2(\mathbf{X})\\ &- 2\left(\hat{d}_{ij}(t) - \mu_{ij}^k\right)d_{ij}(\mathbf{X})\end{aligned}$$

and notice that the first term is constant in $\mathbf{X}$, the second term is quadratic in $\mathbf{X}$, and the third term is a problematic non-convex term. For any $\mathbf{Z} \in \mathbb{R}^{n\times r}$, the third term can be bounded via the Cauchy-Schwarz inequality as before:

$$d_{ij}(\mathbf{X}) \geq \frac{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{z}_i - \mathbf{z}_j)}{\|\mathbf{z}_i - \mathbf{z}_j\|}$$

which establishes the upper bound in Prop. 4.2.

APPENDIX C: DERIVATIVE OF THE GM-MDS MAJORIZING FUNCTION

The Gaussian-mixture majorizing function in Prop. 4.2 can be written in matrix form as follows:

$$\overline{\mathcal{T}}(\mathbf{X},\mathbf{Z}) = \operatorname{tr}\left(\mathbf{X}^T V \mathbf{X}\right) - 2\operatorname{tr}\left(\mathbf{X}^T B(\mathbf{Z})\mathbf{Z}\right) + \eta$$

where:

$$V_{ij} := \begin{cases} \sum_{k=1}^{K} \frac{-\alpha_{ij}^k T}{2\sigma_{ij}^k \sigma_{ij}^k}, & (i,j) \in \mathcal{E} \\ 0, & (i,j) \notin \mathcal{E}, i \neq j \\ -\sum_{j \neq i} V_{ij}, & i = j \end{cases}$$

$$B_{ij}(\mathbf{Z}) := \begin{cases} \sum_{t=1}^{T} \sum_{k=1}^{K} \frac{-\alpha_{ij}^k \left(\hat{d}_{ij}(t) - \mu_{ij}^k\right)}{2\sigma_{ij}^k \sigma_{ij}^k d_{ij}(\mathbf{Z})}, & (i,j) \in \mathcal{E} \\ 0, & (i,j) \notin \mathcal{E}, i \neq j \\ -\sum_{j \neq i} B_{ij}(\mathbf{Z}), & i = j \end{cases}$$

$$\eta := \sum_{t=1}^{T} \sum_{(i,j)\in\mathcal{E}} \sum_{k=1}^{K} \left(\frac{\alpha_{ij}^k}{2\sigma_{ij}^k \sigma_{ij}^k} \left(\hat{d}_{ij}(t) - \mu_{ij}^k\right)^2 + \alpha_{ij}^k \log\left(\sigma_{ij}^k \sqrt{2\pi}\right) \right).$$

Relying on the linearity of the trace operator, the gradient of $\overline{\mathcal{T}}(\mathbf{X},\mathbf{Z})$, necessary to implement Alg. 1 (SMACOF), can be computed as follows:

$$\nabla_{\mathbf{X}} \overline{\mathcal{T}}(\mathbf{X},\mathbf{Z}) = 2V\mathbf{X} - 2B(\mathbf{Z})\mathbf{Z}$$

and vanishes at $X^* := V^\dagger B(\mathbf{Z})\mathbf{Z}$, where

$$V^\dagger = \left(V + \mathbf{1}\mathbf{1}^T\right)^{-1} - \frac{1}{n^2}\mathbf{1}\mathbf{1}^T$$

is the Moore-Penrose pseudoinverse of V.

REFERENCES

[1] Nayef A Alsindi, Bardia Alavi, and Kaveh Pahlavan. 2009. Measurement and modeling of ultrawideband TOA-based ranging in indoor multipath environments. *IEEE Transactions on Vehicular Technology* 58, 3 (2009), 1046–1058.

[2] David Arthur and Sergei Vassilvitskii. 2007. k-means++: The Advantages of Careful Seeding. In *ACM-SIAM Symposium on Discrete Algorithms*. 1027–1035.

[3] James Aspnes, Tolga Eren, David K Goldenberg, A Stephen Morse, Walter Whiteley, Y Richard Yang, Brian Anderson, and Peter N Belhumeur. 2006. A theory of network localization. *Mobile Computing, IEEE Transactions on* 5, 12 (2006), 1663–1678.

[4] Christophe Biernacki, Gilles Celeux, and Gerard Govaert. 2003. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis* 41, 3-4 (2003), 561–575. DOI: http://dx.doi.org/10.1016/S0167-9473(02)00163-9

[5] Jeff Bilmes. 1998. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute* 4, 510 (1998), 126.

[6] Johannes Blömer and Kathrin Bujna. 2016. Adaptive Seeding for Gaussian Mixture Models. In *Advances in Knowledge Discovery and Data Mining: Pacific-Asia Conference*. Springer International Publishing, 296–308. DOI: http://dx.doi.org/10.1007/978-3-319-31750-2_24

[7] I. Borg and P.J.F. Groenen. 2005. *Modern Multidimensional Scaling: Theory and Applications*. Springer.

[8] Ming Cao, Brian DO Anderson, and A Stephen Morse. 2006. Sensor network localization with imprecise distances. *Systems & control letters* 55, 11 (2006), 887–893.

[9] Yiu-Tong Chan, Wing-Yue Tsui, Hing-Cheung So, and Pak-chung Ching. 2006. Time-of-arrival based localization under NLOS conditions. *IEEE Transactions on Vehicular Technology* 55, 1 (2006), 17–24.

[10] Hongyang Chen, Gang Wang, Zizhuo Wang, Hing-Cheung So, and H Vincent Poor. 2012. Non-line-of-sight node localization based on semi-definite programming in wireless sensor networks. *IEEE Transactions on Wireless Communications* 11, 1 (2012), 108–116.

[11] Li Cong and Weihua Zhuang. 2005. Nonline-of-sight error mitigation in mobile location. *IEEE Transactions on Wireless Communications* 4, 2 (2005), 560–573.

[12] Jose Costa, Neal Patwari, and Alfred Hero III. 2006. Distributed weighted-multidimensional scaling for node localization in sensor networks. *ACM Transactions on Sensor Networks (TOSN)* 2, 1 (2006), 39–64.

[13] Jan De Leeuw. 2005. Applications of convex analysis to multidimensional scaling. *Department of Statistics, UCLA* (2005).

[14] Arthur Dempster, Nan Laird, and Donald Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B* (1977), 1–38.

[15] Y. Diao, Z. Lin, and M. Fu. 2014. A Barycentric Coordinate Based Distributed Localization Algorithm for Sensor Networks. *IEEE Trans. on Signal Processing* 62, 18 (2014), 4760–4771. DOI: http://dx.doi.org/10.1109/TSP.2014.2339797

[16] Bruce Hendrickson. 1992. Conditions for unique graph realizations. *SIAM journal on computing* 21, 1 (1992), 65–84.

[17] W. Kabsch. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* 32 (1976), 922–923. DOI: http://dx.doi.org/10.1107/S0567739476001873

[18] W. Kabsch. 1978. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* 34 (1978), 827–828. DOI: http://dx.doi.org/10.1107/S0567739478001680

[19] Anushiya Kannan, Baris Fidan, Guoqiang Mao, and Brian Anderson. 2007. Analysis of flip ambiguities in distributed network localization. *Information, Decision and Control (IDC)* (2007), 193–198.

[20] Soummya Kar, J. Moura, and Kavita Ramanan. 2012. Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication. *Trans. on Information Theory* 58, 6 (2012), 3575–3605. DOI: http://dx.doi.org/10.1109/TIT.2012.2191450

[21] Benjamin Kempke, Pat Pannuto, Bradford Campbell, and Prabal Dutta. 2016. SurePoint: Exploiting Ultra Wideband Flooding and Diversity to Provide Robust, Scalable, High-Fidelity Indoor Localization. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. ACM, 137–149.

[22] Benjamin Kempke, Pat Pannuto, and Prabal Dutta. 2016. Harmonium: Asymmetric, Bandstitched UWB for Fast, Accurate, and Robust Indoor Localization. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 1–12.

[23] U. Khan, S. Kar, and J. Moura. 2010. DILAND: An Algorithm for Distributed Sensor Localization With Noisy Distance Measurements. *IEEE Transactions on Signal Processing* 58, 3 (2010), 1940–1947. DOI: http://dx.doi.org/10.1109/TSP.2009.2038423

[24] Usman A Khan, Soummya Kar, and José MF Moura. 2009. Distributed sensor localization in random environments using minimal number of anchor nodes. *IEEE Transactions on Signal Processing* 57, 5 (2009), 2000–2016.

[25] Joseph B Kruskal. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1 (1964), 1–27.

[26] Josef Kulmer, Erik Leitinger, Paul Meissner, and Klaus Witrisal. 2015. Cooperative Multipath-assisted Navigation and Tracking: A Low-Complexity Approach. In *Future Access Enablers of Ubiquitous and Intelligent Infrastructures*. Springer, 159–165.

[27] Duke Lee. 2005. *Localization using multidimensional scaling (LMDS)*. Ph.D. Dissertation. UNIVERSITY OF CALIFORNIA.

[28] Julia Letchner, Dieter Fox, and Anthony LaMarca. 2005. Large-scale Localization from Wireless Signal Strength. In *National Conference on Artificial Intelligence*, Vol. 1. AAAI Press, 15–20.

[29] Volodymyr Melnykov and Igor Melnykov. 2012. Initializing the EM algorithm in Gaussian mixture models with an unknown number of components. *Computational Statistics and Data Analysis* 56, 6 (2012), 1381–1395. DOI: http://dx.doi.org/10.1016/j.csda.2011.11.002

[30] David Moore, John Leonard, Daniela Rus, and Seth Teller. 2004. Robust distributed network localization with noisy range measurements. In *Proceedings of the 2nd ACM International Conference on Embedded Networked Sensor Systems*. 50–61.

[31] Amanda Prorok. 2013. *Models and Algorithms for Ultra-Wideband Localization in Single- and Multi-Robot Systems*. Ph.D. Dissertation. École Polytechnique Fédérale de Lausanne.

[32] Amanda Prorok and Alcherio Martinoli. 2014. Accurate indoor localization with ultra-wideband using spatial models and collaboration. *The International Journal of Robotics Research* 33, 4 (2014), 547–568.

[33] Yihong Qi. 2003. *Wireless geolocation in a non-line-of-sight environment*. Ph.D. Dissertation. Ph. D. dissertation, Princeton Univ., Princeton, NJ.

[34] Zafer Sahinoglu, Sinan Gezici, and Ismail Guvenc. 2008. Ultra-wideband positioning systems. *Cambridge, New York* (2008).

[35] Yi Shang, W Rumi, Ying Zhang, and Markus Fromherz. 2004. Localization from connectivity in sensor networks. *Parallel and Distributed Systems, IEEE Transactions on* 15, 11 (2004), 961–974.

[36] Yi Shang and Wheeler Ruml. 2004. Improved MDS-based localization. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, Vol. 4. IEEE, 2640–2651.

[37] Saipradeep Venkatraman, James Caffery, and Heung-Ryeol You. 2004. A novel TOA location algorithm using LOS range estimation for NLOS environments. *IEEE Transactions on Vehicular Technology* 53, 5 (2004), 1515–1524.

[38] Radim Zemek, Shinsuke Hara, Kentaro Yanagihara, and Ken-ichi Kitayama. 2007. A joint estimation of target location and channel model parameters in an IEEE 802.15. 4-based wireless sensor network. In *IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. 1–5.