A Unifying Bayesian Optimization Framework for Radio Frequency Localization

Nachikethas A. Jagadeesan and Bhaskar Krishnamachari

***Abstract*—We consider the problem of estimating an RF-device's location based on observations, such as received signal strength, from a set of transmitters with known locations. We survey the literature on this problem, showing that previous authors have considered implicitly or explicitly various metrics. We present a Bayesian optimization framework that unifies these works and shows how to optimize the location estimation with respect to a given metric. We demonstrate how the framework can incorporate a general class of algorithms, including both model-based methods and data-driven algorithms such finger-printing. This is illustrated by re-deriving the most popular algorithms within this framework. Furthermore, we propose using the error-CDF as a unified way of comparing algorithms based on two methods: (i) stochastic dominance, and (ii) an upper bound on error-CDFs. We prove that an algorithm that optimizes any distance based cost function is not strictly stochastically dominated by any other algorithm. This suggests that in lieu of the search for a universally best localization algorithm, the community should focus on finding the best algorithm for a given well-defined objective.**

***Index Terms*—Indoor environments, Algorithm design and analysis, Estimation, Optimization, Bayes methods.**

I. INTRODUCTION

THE ability to locate a wireless device based on received signal strength from known-location transmitters is of great utility for many applications, including indoor location-based mobile apps and services, interactive media, emergency search and rescue, asset tracking, etc. A significant number of researchers have tackled this fundamental problem and proposed various algorithms for radio signal strength (RSS) based localization. Many works adopt standard algorithms from signal processing, specifically estimation theory, such as Maximum Likelihood Estimation [1], Minimum Mean Squared Error Estimation [2], Best Linear Unbiased Estimator [3], etc., while other techniques such as fingerprinting [4], and sequence-based localization [5], are somewhat more heuristically derived. These algorithms are typically evaluated using numerical and trace-based simulations, using varied metrics such as the mean squared position error, the absolute distance error, etc.

We contend that the literature is disconnected and disorganized and that it is hard to decipher any unified theory that fairly evaluates these algorithms across different metrics of interest. We argue that the state-of-the-art approach to localization in the literature — which typically involves first presenting an algorithm and then evaluating its performance according to a particular metric or a set of metrics — is akin to putting the proverbial cart before the horse. For instance, it is not uncommon for algorithms to be evaluated on metrics for which they are not explicitly or implicitly optimized.

We advocate a systematic way of designing location estimation algorithms which we refer to as the "optimization-based approach to localization". In this approach, first the localization metric is defined in terms of a suitable cost function, then an appropriate estimation algorithm is derived to minimize that cost function. In addition, our optimization framework is applicable to any deterministic or stochastic model of the observations (assumed to be known) and can accommodate any prior distribution for location. Our framework also applies to data-driven approaches such as fingerprinting. We show that, in such data-driven settings, our framework makes better use of the available data compared to traditional methods. Fundamentally a Bayesian approach, this framework is also compatible with Bayesian filtering for location tracking over time [6].

As an illustration of our framework, we consider first a common metric used in the evaluation of localization algorithms, the absolute distance error, and derive an algorithm which yields location estimates so as to minimize the expected distance error (MEDE). For a second illustration, we also consider as another metric the probability that the location estimate is within a given radius of the true location ($P(d)$) and derive an algorithm which maximizes this probability. Furthermore, we show that standard algorithms such as MLE and MMSE can be derived similarly from optimizing the corresponding metrics (likelihood, mean squared error respectively).

In conjunction with our framework for deriving algorithms to optimize a specified metric, we also consider the problem of comparing different localization algorithms with each other; for which we make use of the error CDF. For an important class of cost functions that can be expressed as non-negative monotonically increasing functions of distance error, we prove that there is in effect a partial ordering among various estimation algorithms. Certain algorithms dominate other algorithms for all such cost functions, and we show the necessary condition for this to happen. But there could also be two algorithms, A_1 and A_2, and two metrics, M_1 and M_2, such that A_1 is better than A_2 in terms of M_1 while the reverse is true for M_2. Thus we show that there is, in general, no single-best localization algorithm, but rather a "Pareto Set" of algorithms that are each optimized for different cost functions.

We evaluate the optimization-based framework for location estimation using both numerical simulations, traces [7], and data obtained from experiments in an indoor office environment. We illustrate how our framework can incorporate a variety of

The authors are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089. Email: {nanantha, bkrishna}@usc.edu

localization algorithms, including fingerprinting based methods. Our evaluation confirms what is predicted by the theory — no single algorithm outperforms others with respect to all metrics, thus underlining the need for an optimization based approach such as the one we propose.

A. Contributions

- We describe an optimization-based Bayesian framework that unifies and puts in context various previously proposed techniques for localization and provides a systematic basis for developing new algorithms.
- We introduce a partial ordering over the set of algorithms by considering a stochastic dominance relationship between their error CDFs. We prove that any algorithm that optimizes a distance based cost function is not stochastically dominated by any other algorithm.
- We also present how algorithms may be compared based on how 'close' an algorithm gets to the upper bound on error CDFs. We propose one such measure of closeness (area of difference) and identify MEDE as the optimal algorithm over that measure.
- We illustrate how our framework encompasses both model-based approaches and data-driven methods such as fingerprinting, through simulations and real-world experiments.

The rest of the paper is organized as follows: A survey of the existing literature on RSS based localization algorithms is given in Section II. We introduce our optimization based localization framework in Section III. In Section IV we introduce the concept of stochastic dominance and prove how it leads to a partial ordering over the set of localization algorithms. We also introduce the upper bound on error CDFs and illustrate the evaluation of algorithms using the same. In Section V we evaluate our framework using simulations and trace data. We also show how fingerprinting methods fit into our framework. We conclude in Section VI.

II. Literature on RSS based Localization

In this section, we survey the existing literature on RSS based localization algorithms with the intention of comparing, across papers, the metrics used to evaluate the algorithms. The results of the survey are summarized in Table I. We identify certain metrics that are commonly used across the literature:

- *MSE*: Mean Squared Error is the expected value of the square of the Euclidean distance between our estimate and the true location. Often the square root of this quantity Root Mean Squared Error (RMSE) is given instead of MSE. As RMSE may be derived using MSE, we shall only use MSE in our discussions in this paper. The minimum mean squared error (MMSE) algorithm returns an estimate that minimizes the MSE.
- *EDE*: The Expected Distance Error (EDE) is the expected value of the Euclidean distance between our estimate and the true location. The minimum expected distance error (MEDE) algorithm returns an estimate that minimizes the EDE.

TABLE I
LITERATURE SURVEY ON LOCALIZATION METHODS

Study	*Algorithm*	*Model*	*Metric*
WLAN location determination via clustering and probability distributions [8]	MLE	Fingerprinting	$P(d)$
Indoor Localization Without the Pain [9]	Genetic Algorithm	Log-Normal	$D(p)$
RADAR: An In-Building RF-based user location and tracking system [4]	Clustering	Fingerprinting	EDE
The Horus WLAN location determination system [10]	MLE	Fingerprinting	EDE
Locating in fingerprint space: Wireless indoor localization with little human intervention [11]	Clustering	Fingerprinting	EDE
Weighted centroid localization in Zigbee-based sensor networks [12]	RSSI weighted position	Free Space Path Loss	EDE
Sequence based localization in wireless sensor networks [5]	SBL	Free Space Path Loss	EDE
Best linear unbiased estimator algorithm for RSS based localization [3]	Best linear unbiased estimate	Log-Normal	MSE
Cooperative Received Signal Strength-Based Sensor Localization With Unknown Transmit Powers [13]	MLE	Log-Normal	MSE
Relative location estimation in wireless sensor networks [14]	MLE	Log-Normal	MSE
RSS-Based Wireless Localization via SDP: Noncooperative and Cooperative Schemes [15]	MLE	Log-Normal	MSE
A Study of Localization Accuracy Using Multiple Frequencies and Powers [16]	M$P(d)$	Log-Normal	MSE
Maximum likelihood localization estimation based on received signal strength [1]	MLE	Log-Normal	MSE
Distance Estimation from RSS under log-normal shadowing [17]	Best unbiased estimate	Log-Normal	MSE

- $P(d)$: $P(d)$ indicates the probability that the receiver location is within a distance of d from our location estimate. $P(d)$ is closely related to the metric $D(p)$ which gives the radius at which an open ball around our location estimate yields a probability of at least p. The M$P(d)$ algorithm returns an estimate that minimizes the $P(d)$.

As evidenced by Table I, it is with striking regularity that one encounters a mismatch between an algorithm and the metric used for its evaluation. While there is hardly anything amiss in checking how an algorithm performs on a metric that it is not optimized for, it is shortsighted to draw a conclusion as to the efficacy of the said algorithm based on such an evaluation. An awareness of the metric that an algorithm is implicitly or explicitly optimized for, is essential to its fair assessment. We believe that such of notion of *consistent* evaluation of algorithms across *all* important metrics of interest has been absent in the community so far. In addition, while the literature on localization abounds in algorithms that yield a location estimate, there is no unifying theory that relates them to each other with appropriate context.

For instance, [18] picks four algorithms for evaluation, independent of the metrics used to evaluate the algorithms. Such an approach makes it unclear if an algorithm is optimal with respect with any of the given metrics. In this case, we can only make (empirical) inferences regarding the relative ordering of the *chosen* algorithms among the *chosen* metrics. Consequently, there are no theoretical guarantees on algorithm performance and it becomes hard, if not impossible, to accurately predict how a chosen algorithm will behave when evaluated with a metric that was not considered.

Moreover, while error CDFs have been used earlier to evaluate localization algorithms [16], [19]–[21], they are typically used to derive inferences about algorithm performance with respect to the Euclidean distance and $D(p)$ metrics. In the absence of the unifying theory presented in this proposal, it is unclear how one may draw meaningful conclusions regarding the relative performance of algorithms across various metrics based on their error CDF. Our proposed unifying framework places the commonly employed subjective reading of error CDFs on a firm theoretical footing and enables a better understanding of algorithm performance than what was previously possible. Moreover, our framework is computationally tractable as the optimization is typically done over a reasonably sized discrete set of possible locations.

Table I also indicates that there is considerable interest in the community for the EDE metric. However, it is interesting to note that *none* of the algorithms evaluated using that metric are explicitly optimized for it. In the following sections we show how such metrics fit into our framework. More importantly, it is our hope that thinking in terms of the framework below shall lead to a clearer understanding of the trade-offs involved in choosing an algorithm and a better specification of the criterion necessary for its adoption.

The optimization-based framework for localization presented in this paper is inspired in part by the optimization-based approach to networking developed since the late 90's [22], which has shown successfully that efficient medium access, routing, and congestion control algorithms, protocols, and architectures all can be derived from suitably specified network utility maximization problems [23]. Moreover, Bayesian optimization is increasingly gaining in popularity in the recent years [24], including applications in cognitive radio networks [25], [26], largely due to the increased availability of both abundant data and the computational power needed to process that data. Our proposed framework is poised to leverage both these trends.

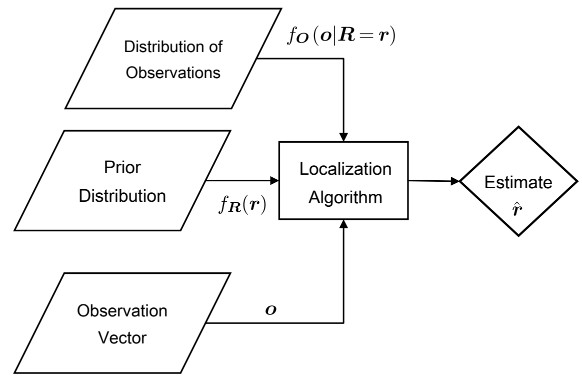


Fig. 1. Localization Algorithms

The Bayesian structure of the localization problem, as presented in this paper, bears similarities to the formulation of the simultaneous localization and mapping (SLAM) problem commonly employed by the robotics community [27], [28]. In general, SLAM algorithms tackle the problem of incrementally building a map of the environment by a mobile robot, in addition to estimating the location of the robot within this constructed map. SLAM algorithms are particularly suited for tracking a mobile robot over time. However this imposes additional assumptions such as knowledge of a model that describes the motion of the robot. While SLAM algorithms may be used to provide simple localization services, for instance by assuming that the robot is not mobile, the focus there is more on maintaining a consistent view of the posterior belief of the location over time, rather than deriving an estimate of the location from such a belief. In contrast, in this paper we mainly focus on the derivation and evaluation of indoor localization algorithms, employing an optimization-based approach. Incorporating posterior belief updates, using Bayesian filters [29] or algorithms based on SLAM, to enable location tracking services within an optimization framework remains an area of future work.

III. A Unifying Framework for Deriving Localization Algorithms

In this section, we present a unifying optimization based approach to localization. We base our approach on a Bayesian view of parameter estimation that can be found in classical statistics and signal processing texts [30], [31]. We adapt the general theory of Bayesian estimation to the indoor localization setting, pointing out the constraints and advantages this entails. We show how existing algorithms can be derived in this

TABLE II
RECOVERING EXISTING ALGORITHMS IN OUR FRAMEWORK

Algorithm	*Cost function*	*Optimization*
MMSE	$C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o}) = (\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2)^2$	$\boldsymbol{r}_{MMSE} = \arg\min_{\tilde{\boldsymbol{r}}} \mathrm{E}\left[(\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2)^2\right]$
MEDE	$C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o}) = \|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2$	$\boldsymbol{r}_{MEDE} = \arg\min_{\tilde{\boldsymbol{r}}} \mathrm{E}\left[\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2\right]$
M$P(d)$	$C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o}) = -P\left(\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2 \leq d\right)$	$\boldsymbol{r}_{MP(d)} = \arg\max_{\tilde{\boldsymbol{r}}} \mathrm{E}\left[P\left(\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2 \leq d\right)\right]$
MLE	$C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o}) = -P\left(\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2 \leq \epsilon\right)$	$\boldsymbol{r}_{MLE} = \lim_{\epsilon \to 0} \arg\max_{\tilde{\boldsymbol{r}}} \mathrm{E}\left[P\left(\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_2 \leq \epsilon\right)\right]$

framework and point out how alternate algorithms may be derived.

Let $\mathcal{S} \subseteq \mathbb{R}^2$ be the two-dimensional space of interest in which localization is to be performed[1]. We assume that $\mathcal{S}$ is closed and bounded. Let the location of the receiver (the node whose location is to be estimated) be denoted as $\boldsymbol{r} = [x_r, y_r]$. Using a Bayesian viewpoint [30], [31], we assume that this location is a random variable with some prior distribution $f_{\boldsymbol{R}}(\boldsymbol{r})$. This prior distribution is used to represent knowledge about the possible position, obtained, for instance from previous location estimates or knowledge of the corresponding user's mobility characteristics in the space; in the absence of any prior knowledge, it could be set to be uniform over $\mathcal{S}$. Let $\boldsymbol{o} \in \mathbb{R}^N$ represent the location dependent observation data that was collected. As an example, $\boldsymbol{o}$ could represent the received signal strength values from transmitters whose locations are known. Mathematically, we only require that the observation vector is drawn from a known distribution that depends on the receiver location $\boldsymbol{r}$: $f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R} = \boldsymbol{r})$. In case of RSS measurements, this distribution characterizes the stochastic radio propagation characteristics of the environment and the location of the transmitters. Note that this distribution could be expressed in the form of a standard fading model whose parameters are fitted with observed data, such as the well-known simple path loss model with log-normal fading [32]. The distribution $f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R} = \boldsymbol{r})$ is general enough to incorporate more data-driven approaches such as the well-known fingerprinting procedure. In fingerprinting, there is a training phase in which statistical measurements are obtained at the receiver at various known locations and used to estimate the distribution of received signal strengths at each location.[2] Fundamentally, the data-driven approach constructs $f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R} = \boldsymbol{r})$ empirically, while model-dependent approaches take the distribution over observations directly from the model.

Using the conditional distribution of the observed vector and the prior over $\boldsymbol{R}$, we obtain the conditional distribution over the receiver locations using Bayes' rule:

$$f_{\boldsymbol{R}}(\boldsymbol{r}|\boldsymbol{O} = \boldsymbol{o}) = \frac{f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R} = \boldsymbol{r}) f_{\boldsymbol{R}}(\boldsymbol{r})}{\int_{\boldsymbol{r} \in \mathcal{S}} f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R} = \boldsymbol{r})\, f_{\boldsymbol{R}}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}}. \tag{1}$$

[1]It is trivial to extend the framework to 3-D localization, for simplicity, we focus on the more commonly considered case of 2-D localization here.

[2]We note that in many implementations of fingerprinting, only the mean received signal strength from each transmitter is used, which of course is a special case, equivalent to assuming a deterministic signal strength measurement with a unit step function cumulative distribution function.

Algorithms for localization are essentially methods that derive a location estimate from the above posterior distribution. In fact, any localization algorithm A is a mapping from

- the observation vector $\boldsymbol{o}$
- the prior distribution over the location, $f_{\boldsymbol{R}}(\boldsymbol{r})$
- the conditional distribution over $\boldsymbol{o}$, $f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R} = \boldsymbol{r})$

to a location estimate $\hat{\boldsymbol{r}}$, as illustrated in Figure 1. A visualization[3] of the posterior distribution for the popular simple path loss model with log-normal fading is given in Figure 3.

A. Optimization based approach to Localization

The starting point for estimating the receiver location is a cost function that must be defined *a priori*. In the most general terms, the cost function is modeled as $C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o})$, i.e., a function of the true location $\boldsymbol{r}$, a given proposed location estimate $\tilde{\boldsymbol{r}}$, and the observation vector $\boldsymbol{o}$. We define the expected cost function given an observation vector as follows:

$$\mathrm{E}[C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o})] = \int_{\boldsymbol{r} \in \mathcal{S}} C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o})\, f_{\boldsymbol{R}}(\boldsymbol{r}|\boldsymbol{O} = \boldsymbol{o})\, \mathrm{d}\boldsymbol{r}. \tag{2}$$

Given any cost function C, the optimal location estimation algorithm can be obtained in a unified manner by solving the following optimization for any given observation vector to obtain the optimal estimate $\hat{\boldsymbol{r}}$:

$$\hat{\boldsymbol{r}} = \arg\min_{\tilde{\boldsymbol{r}}} \mathrm{E}[C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o})]. \tag{3}$$

Note that this optimization may be performed to obtain an arbitrarily near-optimal solution by numerically computing $\mathrm{E}[C(\boldsymbol{r}, \tilde{\boldsymbol{r}}, \boldsymbol{o})]$ over the discretization of a two or three dimensional search space. Given recent gains in computing power, the optimization is feasible for typical indoor localization problems. Moreover, the optimization naturally lends itself to parallel execution since the computation of the expected cost at all candidate locations are independent of each other. Assuming uniform coverage, the solution will improve upon increasing the number of points in our search space. In practice, for RSS localization, these points could be spaced apart on the order of 10's of centimeters.

Existing algorithms such as MLE, MMSE, MEDE and M$P(d)$ can be recovered in this framework using suitable choices of the cost function C. For instance, it is straightforward to verify that minimizing the expected distance error yields MEDE. Perhaps more interestingly, the MLE estimate can also be recovered

[3]The code used to generate the figures in this paper is available online [33].

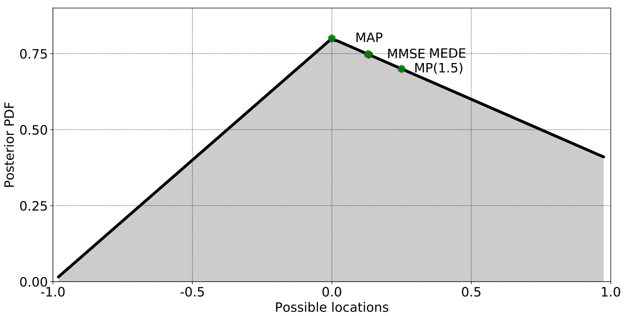


Fig. 2.
An Illustration of the distribution of true location given the observations and the locations that correspond to different optimizations. In this example, we show the different estimates returned by the various localization algorithms for a unimodal, asymmetric posterior probability density function of the form

$$f(x) = \begin{cases} \frac{4}{5}(1+x) & \text{if } -1 \leq x < 0 \\ \frac{4}{5}(1-\frac{x}{2}) & \text{if } 0 \leq x \leq 1. \end{cases}$$

The asymmetry of the distribution function pulls estimates other than MAP from the mode, with M$P(d)$ being the most affected. In this example, it may be shown that for $d \leq 1.5$, the M$P(d)$ estimate is $\hat{x}_{MP(d)} = \frac{d}{6}$. Thus we see that the M$P(d)$ estimate moves closer to the MAP estimate with decreasing d. This example serves to illustrate how differing optimization objectives can yield very different estimates for the same posterior distribution, thereby underlining the importance of deciding on an optimization objective upfront.

using an appropriate *distance based* cost function, as can be seen by employing Theorem 1 with a uniform prior. Table II lists the choice of cost functions and the optimization problem to be solved that results in each of these algorithms. Figure 2 provides an example of how these different optimizations can yield very different location estimates.

The most compelling aspect of this unified optimization-based approach to localization is its generality. Being Bayesian in nature, it can incorporate both model and data-driven approaches to characterizing the radio environment in a given space, and can accommodate prior information in a natural way (as such, it is also highly compatible with location tracking approaches that use Bayesian filtering). In addition, the framework gets better over time as more observations or inputs help improve the prior. While we present and evaluate our framework using RSS measurements for ease of exposition, it is not limited to such measurements. Other modalities such as ToA, TDoA and AoA [34], [35] are easily incorporated as well.

IV. A Unifying Framework for Evaluating Localization Algorithms

In addition to the previously defined unifying framework, we also propose the use of the distance *error cdf* as a unified way of evaluating localization algorithms. For a localization algorithm, say A, the L_2 (Euclidean) distance between the estimate ($\hat{\boldsymbol{r}}_A$) and the true location ($\boldsymbol{r}$) is represented by the random variable D_A. Note that

$$D_A = \|\hat{\boldsymbol{r}}_A - \boldsymbol{r}\|_2. \tag{4}$$

The CDF of D_A, also termed the error cdf of algorithm A, may be characterized by averaging the probability that the true location lies within a certain distance, say d, of our estimate, over the all possible receiver locations. This notion is defined below.

Definition 1. Let $A \in \mathcal{A}$, where $\mathcal{A}$ denotes the set of all localization algorithms. Denote by $\hat{\boldsymbol{r}}_A$ a location estimate returned by the algorithm A. Then, the *error cdf* of A is a monotonically increasing function $F_A : \mathcal{Q} \subseteq \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ such that

$$F_A(d) = \int\limits_{\boldsymbol{r}\in\mathcal{S}} P\left[D_A \leq d\right] f_{\boldsymbol{R}}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}. \tag{5}$$

Let d^* be the maximum distance between any two points in $\mathcal{S}$. Then $\mathcal{Q}$ is the closed interval $[0, d^*]$. Using the error cdf, we may meaningfully define an ordering over the class of localization algorithms using the concept of *stochastic dominance*.

Definition 2. Let $A_1, A_2 \in \mathcal{A}$. We say that A_1 *stochastically dominates* A_2 if

$$F_{A_1}(d) \geq F_{A_2}(d) \quad \forall d \in \mathcal{Q}. \tag{6}$$

Definition 3. Let $A_1, A_2 \in \mathcal{A}$. We say that A_1 *strictly stochastically dominates* A_2 if in addition to equation (6), there exists $d_1, d_2 \in \mathcal{Q}$ such that $d_1 < d_2$ and

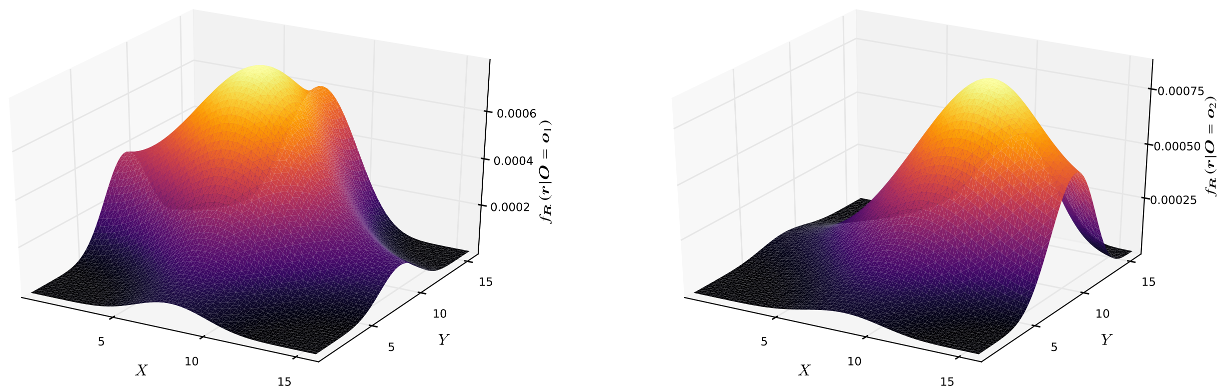
$$F_{A_1}(d) > F_{A_2}(d) \quad \forall d \in [d_1, d_2]. \tag{7}$$

A. Distance Based Cost Functions

We now restrict our attention to an important class of metrics, those cost functions that can be specified to be monotonically increasing with respect to the distance between the true and estimated positions. We show that in fact localization algorithms form a partially-ordered set with respect to this important class of metrics. We also show that localization algorithms derived using the optimization-based approach for these metrics lie essentially on a "Pareto Boundary" of the set of all localization algorithms. The localization cost function, formally defined below, generalizes most metrics commonly used in the localization literature.

Definition 4. Let $g : \mathcal{Q} \subseteq \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ be a monotonically increasing function. Denote the set of all such functions by $\mathcal{G}$. For a localization algorithm A, $g(D_A)$ is the *distance error localization cost function*. $\mathrm{E}\left[g(D_A)\right]$ is the *expected cost* of the algorithm A.

We use the above notion of expected cost as a metric to compare different localization algorithms. Note that this cost function is a special case of the more general cost function introduced in the previous section. Here we are only interested in cost functions that depend on the distance between the true location and our estimate. Many localization algorithms of interest try to optimize for some distance based cost function, either explicitly or implicitly. We have seen already that MMSE, MEDE and M$P(d)$ have distance based cost functions. Although perhaps not immediately apparent, MAP may also be computed using a distance based cost function. Specifically, we may retrieve the MAP estimate using M$P(d)$ with an adequately small radius d as shown in the theorem below and also borne out by our evaluation results.



(a) (b)

Fig. 3. Illustration of the posterior distribution of $\boldsymbol{R}$ using observations taken from a log-normal distribution. For this illustration, four transmitters were placed in a 16 m × 16 m area. A log-normal path loss model was used to to determine the signal strengths. Each subplot above shows the posterior distribution of $\boldsymbol{R}$ constructed by the receiver upon receiving a different vector of observations.

Theorem 1. *If the posterior distribution $f_{\boldsymbol{R}|\boldsymbol{O}}$ is continuous and the MAP estimate lies in the interior of $\mathcal{S}$, then for any $\delta > 0$ there exists $\epsilon > 0$ such that,*

$$|P_1(d) - P_2(d)| \leq \delta \quad \forall\, 0 < d < \epsilon, \tag{8}$$

where $P_1(d)$ and $P_2(d)$ give the probability of the receiver location being within distance d of the MAP and the $MP(d)$ estimate respectively.

Proof. See [36]. □

B. Comparing Algorithms using Stochastic Dominance

In this section, we explore how we may meaningfully compare algorithms using our optimization based framework. If we are interested in a particular cost function, then comparing two algorithms is straightforward. Compute their expected cost and the algorithm with the lower cost is better. However, with stochastic dominance we can deduce something more powerful. For any two localization algorithms A_1 and A_2, if A_1 stochastically dominates A_2, then the expected cost of A_1 does not exceed that of A_2 for *any* distance based cost function. More formally,

Theorem 2. *For any two localization algorithms $A_1, A_2 \in \mathcal{A}$, if A_1 stochastically dominates A_2, then*

$$\mathrm{E}\left[g(D_{A_1})\right] \leq \mathrm{E}\left[g(D_{A_2})\right] \quad \forall g \in \mathcal{G}. \tag{9}$$

If A_1 strictly stochastically dominates A_2, then

$$\mathrm{E}\left[g(D_{A_1})\right] < \mathrm{E}\left[g(D_{A_2})\right] \quad \forall g \in \mathcal{G}. \tag{10}$$

Proof. See [36]. □

Theorem 2 is the first step towards ranking algorithms based on stochastic dominance. It also gives us a first glimpse of what an optimal algorithm might look like. From Theorem 2, an algorithm A^* that stochastically dominates every other algorithm is clearly optimal for the entire set of distance based cost functions. However, it is not obvious that such an algorithm need even exist. On the other hand, we can compute algorithms that are optimal with respect to a particular cost function. As given in the following theorem, such optimality implies that the algorithm isn't strictly dominated by any other algorithm. In other words, if algorithm A is optimal with respect to a distance based cost function g, then A is not strictly stochastically dominated by *any* other algorithm B.

Theorem 3. *For a localization algorithm $A \in \mathcal{A}$, if there exists a distance based cost function $g \in G$ such that for any other localization algorithm $B \in \mathcal{A}$*

$$\mathrm{E}\left[g\left(D_A\right)\right] \leq \mathrm{E}\left[g\left(D_B\right)\right], \tag{11}$$

then for all algorithms $B \in \mathcal{A}$, there exists a distance $d \in Q$ such that

$$F_A(d) \geq F_B(d). \tag{12}$$

Proof. See [36]. □

Theorems 2 & 3 establish the utility of ranking algorithms based on stochastic dominance. However, if we are given two algorithms, it is not necessary that one should dominate the other. As Theorem 4 shows, if they do not conform to a stochastic dominance ordering, the algorithms are incomparable.

Theorem 4. *For any two localization algorithms A_1 and A_2, if A_2 does not stochastically dominate A_1 and vice versa, then there exits distance based cost functions $g_1, g_2 \in \mathcal{G}$ such that*

$$\mathrm{E}\left[g_1(D_{A_1})\right] < \mathrm{E}\left[g_1(D_{A_2})\right], \tag{13}$$

and

$$\mathrm{E}\left[g_2(D_{A_2})\right] < \mathrm{E}\left[g_2(D_{A_1})\right]. \tag{14}$$

Proof. See [36]. □

Theorem 4 establishes the existence of a "Pareto Boundary" of the set of all localization algorithms. Choosing an algorithm from within this set depends on additional considerations such as its performance on specific cost functions of interest.

C. Comparison based on Upper Bound of Error CDFs

In the previous section, we focused on using stochastic dominance to rank and compare algorithms without paying much attention to what an ideal algorithm might look like. In this section, we explore this topic more detail. To begin, we ask if there exists an algorithm that dominates every other algorithm? From Theorem 2 we know that such an algorithm, if it exists, will be the best possible algorithm for the class of distance based cost functions. Moreover the error CDF of such an algorithm will be an upper bound on the error CDFs of all algorithms $A \in \mathcal{A}$.

Definition 5. We denote the upper envelope of error CDFs for all possible algorithms $A \in \mathcal{A}$ by F^*.

We now turn our attention to formally defining the error bound F^*. Our definition also provides us with a way to compute F^*. Let D_A represent the distance error for algorithm A. Consider the following class of M$P(d)$ cost functions. For each $d \in Q$, let

$$g_d(D) = \begin{cases} 0 & \text{if } D \leq d \\ 1 & \text{if } D > d. \end{cases} \tag{15}$$

Then, the value of F^* at any distance $d \in Q$ may be computed using the M$P(d)$ cost function at that distance. More formally,

Definition 6. The upper envelope of error CDFs for all possible algorithms $A \in \mathcal{A}$, F^* is defined as

$$F^*(d) = \sup_{A \in \mathcal{A}} \{1 - \mathrm{E}[g_d(D_A)]\}, \quad \forall d \in Q. \tag{16}$$

The upper envelope of error CDFs, F^*, satisfies the following properties:

1) F^* stochastically dominates every algorithm $A \in \mathcal{A}$,
2) F^* is monotonically increasing in $[0, d^*]$,
3) F^* is Riemann integrable over $[0, d^*]$.

The monotonicity of F^* is direct consequence of the monotonicity of CDFs. Moreover, since F^* is monotonic, it is also Riemann integrable [37, p. 126]. In general, F^* may not be attainable by any other algorithm. However, as we show below, it is achievable under certain circumstances, which lends credence to its claim as a useful upper bound that may be used as a basis of comparison of localization algorithms.

Given the ideal performance characteristics of F^*, it is worthwhile to investigate if it is ever attained by an algorithm. A trivial case is when the M$P(d)$ algorithm yields the same estimate for all distances of interest in the domain. In this particular case, MAP and M$P(d)$ are optimal since the error CDF of the MAP or M$P(d)$ estimate traces F^*. As an illustration consider a continuous symmetric unimodal posterior distribution over a circular space with the mode located on the center of the circle. Clearly, the MAP estimate is given by the center. Moreover, the MP(d) estimate is the same at all distances, namely the center of the circle. Thus we immediately have that both the MAP and M$P(d)$ estimates have attained F^*. An extensive discussion on the attainability of F^* can be found in [36].

Thus we see that there exist conditions under which F^* is attained by an algorithm. Consequently, it is worthwhile to search for algorithms that are close to this bound or even attain it under more general settings. This leads us directly to the second method of comparing algorithms. We identify how close the error CDFs of our algorithms get to the upper bound F^*.

Consider algorithms $A, B \in \mathcal{A}$. Intuitively, if the error CDF of A is closer to F^* than that of B, then it seems reasonable to expect A to perform better. To make this idea precise, we need to define our measure of closeness to F^*. In the following paragraph, we propose one such measure of how close the error CDF of an algorithm A is to F^*. Our proposal satisfies a nice property. Namely, searching for an algorithm that is optimal over this measure is equivalent to searching for an algorithm that minimizes a particular distance based cost function. Consequently, to specify the algorithm we only need to identify this cost function.

Definition 7. The area between the error CDF of algorithm A and the upper envelope of error CDFs is given by

$$\Theta_A = \int_0^{d^*} (F^*(x) - F_A(x))\,\mathrm{d}x. \tag{17}$$

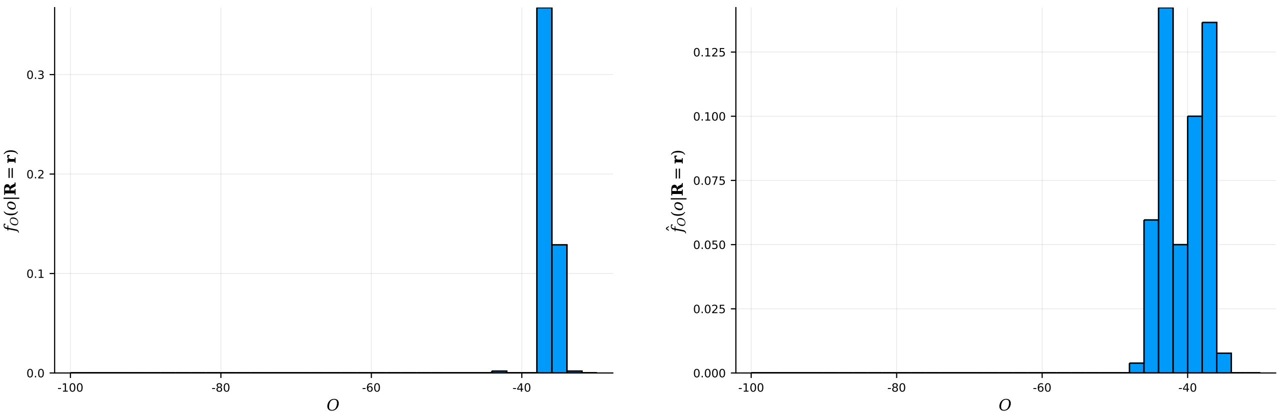
The intuition behind our measure is can be summarized easily. We seek to find an algorithm A that minimizes the "area enclosed" by F^* and the error CDF of A. Note that $\Theta_A \geq 0$ for all $A \in \mathcal{A}$. In general, it is not clear if every measure of closeness between F^* and F_A will yield a cost function for us to minimize. However, if we do find such a cost function, then we have the advantage of not needing to explicit know F^* in the execution of our algorithm. This is the case for Θ_A as proved in the theorem below.

Theorem 5. *The algorithm that minimizes the area between its error CDF and the upper envelope of error CDFs for all possible algorithms is the MEDE algorithm.*

Proof. See [36]. □

As consequence of Theorem 5, we note that if F^* is attainable by any localization algorithm, then it is attained by MEDE. This in turn yields a simple test for ruling out the existence of an algorithm that attains F^*. On plotting the error CDF plots of different algorithms, if we find an algorithm that is not dominated by MEDE, then we may conclude that F^* is unattainable. Thus it is relatively easy to identify cases where there is a gap between F^* and MEDE. However, the issue of confirming that MEDE has attained F^* is more difficult as it involves a search over the set of all algorithms.

In summary, the utility of F^* lies in its ability to pin point the strengths and weaknesses of a proposed algorithm. As we have seen, some algorithms such as M$P(d)$ is designed to do well at specific distances while others such as MEDE aims for satisfactory performance at all distances. Other algorithms lie somewhere in between. Consequently, choosing one algorithm over the other depends on the needs of the application utilizing the localization algorithm. Therein lies the strength of our proposed framework. It allows us to effectively reason about the applicability of an algorithm for the use case at hand.



(a) (b)

Fig. 4. Illustration of the empirically estimated distribution of O using signal strength measurements taken at different locations. Each subplot refers to the distribution of signal strength from a unique access point. The left subplots refer to measurements taken at night, while the right subplots refer to measurements taken during daytime. Our framework makes better use of the higher variance data.

D. Comparison with the Cramér–Rao Bound

The Cramér–Rao Bound (CRB) is often used as an aid in evaluating localization algorithms [14], [17], [38], [39]. However, the CRB is not necessarily a good choice as an absolute measure of performance for every localization algorithm.

Let D_A denote the distance error corresponding to a algorithm A. Define $d_A = \mathrm{E}[D_A]$. Since the ideal distance error is 0, d_A is the *bias* of algorithm A. The expected mean squared error of algorithm A may be then expressed as

$$\mathrm{E}\left[D_A^2\right] = \mathrm{E}\left[(D_A - d_A)^2\right] + d_A^2 \tag{18}$$

$$= \mathrm{Var}[D_A] + (\mathrm{Bias}[D_A])^2. \tag{19}$$

In our setting, the CRB is a lower bound on $\mathrm{Var}[D_A]$. Thus an algorithm that attains the CRB is optimal only if (i) our objective is to minimize the mean squared error, and (ii) the class of algorithms under consideration is unbiased. In our framework, the MMSE algorithm given in Table II shares the same objective as that of an algorithm evaluated against the CRB. However, the MMSE algorithm considers both variance and bias simultaneously, allowing for estimates that have a small, non-zero bias combined with a small variance.

V. Evaluation

We evaluate the proposed framework using simulations, traces and real world experiments. In Section V-A, we provide an illustration of how fingerprinting methods fit in to the framework presented in Section III, using real-world data collected from an indoor office environment. We show that while many implementations of fingerprinting use only the mean signal strength from each transmitter, we are able to better utilize the collected data by building an empirical distribution of the received observations.

We also evaluate the performance of the MLE, M$P(d)$, MMSE and MEDE using simulations as well as using traces [7]. In both cases the signal propagation was modelled using a simplified path loss model with log-normal shadowing [32]. We assume that the prior distribution ($f_{\boldsymbol{R}}$) is uniform over $\mathcal{S}$.

A. Fingerprinting Methods

Model-based methods assume that the distribution of observations given a receiver location is known. In contrast, fingerprinting methods avoid the need to model the distribution of observations by noting that one only needs to identify the change in distribution of observations from one location to another. Most implementations simplify matters even further by *assuming* that mean of the observations is distinct across different locations in our space of interest. The estimated mean is thus said to 'fingerprint' the location.

This approach works well only in cases when the distribution of signal strengths is mostly concentrated around the mean. In this case, the approach of using only the mean signal strength amounts to approximating the signal strength distribution with a normal distribution centered at the estimated mean signal strength and variance approaching zero. However, if the distribution has significant variance this approach is likely to fail. Indeed, in the regime of significantly varying signal strengths, keeping only the mean amounts to throwing away much of the information that one has already taken pains to collect.

As already indicated in Section III, we make better use of the collected data by empirically constructing the distribution of observations $f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R}=\boldsymbol{r})$. This formulation allows us to the use the same algorithms as in the model-based approach, as the only difference here is in the construction of $f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R}=\boldsymbol{r})$. This is in contrast to many existing implementations where one resorts to heuristics such as clustering. Indeed, under mild assumptions it is well known that the empirical distribution converges with probability one to true distribution [40] which gives our approach the nice property that it can always do better given more data.

As a proof-of-concept, we compare the performance of our approach with that of traditional fingerprinting methods in two different settings. The data was collected from a $4\,\mathrm{m} \times 2\,\mathrm{m}$ space inside an office environment. The space was divided into eight $1\,\mathrm{m} \times 1\,\mathrm{m}$ squares and signal strength samples were collected from the center of each square. Two hundred and fifty signal

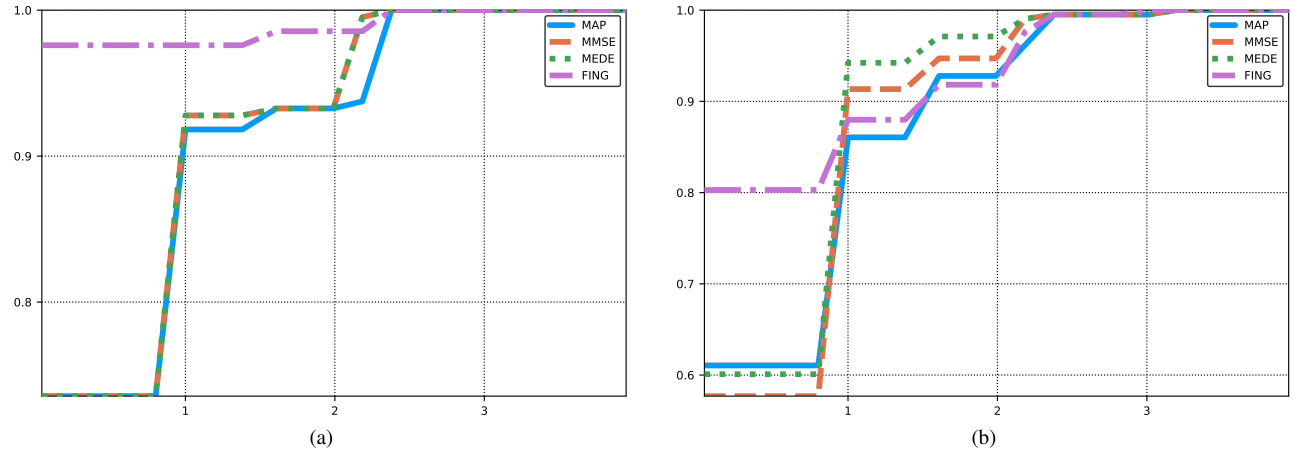


Fig. 5. Comparison of error CDFs for different localization algorithms. Traditional fingerprinting based on matching the mean signal strengths is indicated by the moniker 'FING'. Each subplot indicates the algorithm performance for a different set of test data. The left subplots refer to the case when the signal strengths are tightly clustered around the mean, while the right subplots refer to the measurements with more variance.

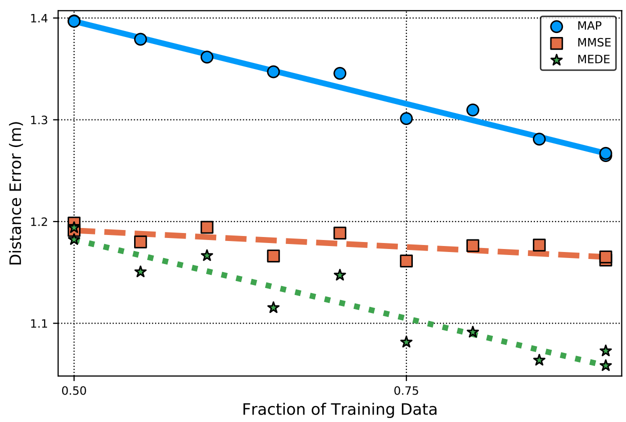


Fig. 6. Variation of distance error with the size of the training data set. For each fraction of the original data set, we compute the distance error for 100 random choices of the data points. The regression line for each algorithm is also shown. The results indicate that our empirical estimates of the prior improves with increasing training data, which results in better algorithm performance.

strength readings were collected for the ten strongest access points detected using the WiFi card on a laptop running Linux. The beacon interval for each access point was approximately 100 ms. The signal strength measurements were taken 400 ms apart. Two sets of data were collected, one at night time and the other during the day. The measurements taken at night show that the observed signal strengths are highly concentrated around the mean, as can be seen from the left subplots of Figure 4. The measurements taken during daytime show slightly more variability as can be seen in the right subplots of Figure 4.

Ten percent of the collected data is randomly chosen for evaluating algorithm performance. The remaining data is used to construct the empirical distribution $\hat{f}_{\boldsymbol{O}}\left(\boldsymbol{o}|\boldsymbol{R}=\boldsymbol{r}\right)$ from which the MMSE, MAP and MEDE estimates are derived. It is also used to compute the mean signal strength vector or fingerprint for each location. For the algorithm denoted as 'FING' in Figure 5, the fingerprint closest to the test observation vector (in terms of Euclidean distance) is used to predict the location.

From the performance results given in Figure 5, we see that, as expected, the traditional fingerprinting approach works very well when the variability in the signal strength data is low. On the other hand, even with slight variability in the data, the estimates derived using our Bayesian framework outperforms traditional fingerprinting.

To illustrate that our framework performs better given more observations over time, we investigate the variation of distance error with increasing size of the training data set. The data set with more variability was chosen for the purposes of this illustration. For each fraction of the original data set, we compute the distance error for 100 random choices of the data points, for MAP, MEDE and MMSE algorithms. Figure 6 shows how the average of these distance errors varies on increasing the size of the training data set. As can be seen from Figure 6, with increasing data we are able to better estimate the empirical distribution $\hat{f}_{\boldsymbol{O}}\left(\boldsymbol{o}|\boldsymbol{R}=\boldsymbol{r}\right)$ from which the MMSE, MAP and MEDE estimates are derived, thereby resulting in better performance.

B. Simulation Model

Say $\{\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_N\}$ $(N > 2)$ are the known positions of (N) wireless transmitters. We assume each transmitter is located on a planar surface given by $\mathcal{S} = [0, l] \times [0, b]$ where $l, b \in \mathbb{R}_{>\mathbf{0}}$. The locations of the transmitters are given by the two dimensional vector $\boldsymbol{l}_i = (x_i, y_i) \in \mathcal{S}\ \forall i \in \{1, 2, \ldots, N\}$. We wish to estimate the receiver locations, given by the vector $\boldsymbol{r} = (x, y)$, from the received signal strengths. For a given transmitter-receiver pair, say i, the relationship between the received signal power (P_r^i) and the transmitted signal power (P_t^i) may be modelled by the simplified path loss model: $P_r^i = P_t^i K \left[\frac{d_0}{d_i}\right]^\eta W_i$, where the distance between the receiver and the i^{th} transmitter is given by $d_i\left(\boldsymbol{r}\right) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$, and W_i represents our noise that is log-normally distributed with zero mean and variance σ^2. In log scale, the path loss model is given by

$$P_r^i|_{\text{dBm}} = P_t^i|_{\text{dBm}} + K|_{\text{dB}} - 10\eta \log_{10}\left[\frac{d_i}{d_0}\right] + W_i|_{\text{dB}}, \qquad (20)$$

where K is a constant given by the gains of the receiver and transmit antennas and possibly the frequency of transmission. d_0 is a reference distance, taken to be 1m. In this setting, our estimation problem may be restated as follows. We are given measurements of the receiver signal strengths $\left\{P_r^1, P_r^2, \ldots, P_r^N\right\}$ from which we are to estimate the receiver location $\boldsymbol{r}$. Thus, our observation vector $\boldsymbol{O}$ may be written as

$$O_i = P_r^i|_{\mathrm{dBm}} - P_t^i|_{\mathrm{dBm}} - K|_{\mathrm{dB}} = W_i|_{\mathrm{dB}} - 10\eta \log_{10}\left[\frac{d_i}{d_0}\right],$$

for all $i \in \{1, \ldots, N\}$. In other words, the distribution of each observation is given by $O_i \sim \mathcal{N}\left(-10\eta \ln\left[d_i(\boldsymbol{r})\right], \sigma^2\right)$. Finally, the distribution of the observation vector $f_{\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{R} = \boldsymbol{r})$ can be obtained from the above by taking the product of all the individual observation pdfs.

C. Simulation and Trace Results

TABLE III
NORMALIZED PERFORMANCE RESULTS

	Simulations				
	Likelihood	$P(\epsilon)$	$P(d)$	MSE	EDE
MLE	1.0000	0.9222	0.8994	1.4359	1.1240
M$P(\epsilon)$	0.9808	1.0000	0.9165	1.3172	1.0997
M$P(d)$	0.6963	0.7573	1.0000	1.2860	1.0857
MMSE	0.6806	0.6980	0.8737	1.0000	1.0643
MEDE	0.7247	0.7455	0.9080	1.1272	1.0000
	Traces				
	Likelihood	$P(\epsilon)$	$P(d)$	MSE	EDE
MLE	1.0000	0.9989	0.9865	1.1583	1.1808
M$P(\epsilon)$	0.9976	1.0000	0.9881	1.1522	1.1785
M$P(d)$	0.8213	0.8596	1.0000	1.2605	1.3760
MMSE	0.9013	0.9171	0.9797	1.0000	1.1101
MEDE	0.8529	0.8685	0.9517	1.1569	1.0000

The parameters for the simulation were chosen to be identical as that of the traces. The dimensions of the area of interest ($\mathcal{S}$) was $50\,\mathrm{m} \times 70\,\mathrm{m}$. Sixteen transmitters were chosen randomly and 100 RSSI readings were taken for each transmitter at 5 distinct receiver locations. The transmit power was kept constant at 16 dBm. The estimated model parameters were a path loss of $K = 39.13\,\mathrm{dB}$ at reference distance $d_0 = 1\,\mathrm{m}$, fading deviation $\sigma = 16.16$ and path loss exponent $\eta = 3.93$. We used two distances for the M$P(d)$ algorithm: (i) $\epsilon = 0.5\,\mathrm{m}$ was relatively small while (ii) $d = 3\,\mathrm{m}$ covers a more sizable area. The normalized performance results are presented in Table III. Each row evaluates the performance of the indicated algorithm across different metrics, while each column demonstrates how different algorithms perform under the given metric. The performance value for each metric is normalized by the performance of the best algorithm for that metric. Thus the fact that each algorithm performs best in the metric that it is optimized for, is reflected in the occurrence of ones as the diagonal entries in the table.

As the algorithms presented here are each optimal for a specific cost function, the theory predicts that none of them are strictly stochastically dominated by any other algorithm. The results confirm this theoretical prediction. Moreover, choosing ϵ to be small results in the performance of M$P(\epsilon)$ being near identical to that of MLE, which is in line with what we expect from Theorem 1. Intuitively, M$P(d)$ tries to identify the region with given radius that 'captures' most of the *a posteriori* pdf $f_{\boldsymbol{R}}(\boldsymbol{r}|\boldsymbol{O} = \boldsymbol{o})$. Consequently for a sufficiently small radius d, M$P(d)$ will return a region that contains the MLE estimate. As a result, we are justified in thinking of the M$P(d)$ algorithm as a generalization of MLE (or MAP, in case our prior is non-uniform). In practice, the value of d to be used will be dictated by the needs of the application that makes use of the localization service.

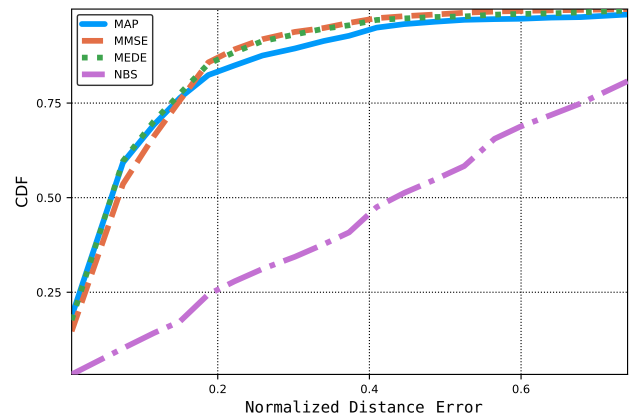


Fig. 7. An illustration of stochastic dominance. We plot the error CDFs of MAP, MMSE, MEDE, and a naïve baseline algorithm (NBS) that returns the location of the base station with the highest signal strength. For this illustration, three transmitters were placed evenly on a line and log-normal fading was assumed. The baseline algorithm is outperformed by MAP, MMSE, and MEDE, as demonstrated by their strict stochastic dominance over the baseline.

VI. CONCLUSION

We have introduced an optimization based approach to localization using a general unified framework that also allows for a fair and consistent comparison of algorithms across various metrics of interest. We have demonstrated how this framework may be used to derive localization algorithms, including fingerprinting methods. We have shown the existence of a partial ordering over the set of localization algorithms using the concept of stochastic dominance, showing further that the optimality of an algorithm over a particular distance based cost function implies that the algorithm is not stochastically dominated by another. We have identified key properties of an "ideal" localization algorithm whose performance corresponds to the upper bound on error CDFs, and highlighted how we may compare different algorithms relative to this performance. Specifically, we have shown that MEDE minimizes the area between its own error CDF and the performance of such an ideal algorithm. We believe that the framework presented here goes a long way towards unifying the localization literature. The optimization based approach places the localization algorithm desideratum at the forefront, where we believe it belongs.

References

[1] A. Waadt, C. Kocks, S. Wang, G. Bruck, and P. Jung, "Maximum likelihood localization estimation based on received signal strength," in *Proc. IEEE ISABEL'10*, Nov. 2010, pp. 1–5.

[2] Y.-F. Huang, Y.-T. Jheng, and H.-C. Chen, "Performance of an MMSE based indoor localization with wireless sensor networks," in *Proc. IEEE NCM'10*, Aug. 2010, pp. 671–675.

[3] L. Lin and H. So, "Best linear unbiased estimator algorithm for received signal strength based localization," in *Proc. IEEE EUSIPCO'11*, Aug. 2011, pp. 1989–1993.

[4] P. Bahl and V. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM'00*, vol. 2, Mar. 2000, pp. 775–784.

[5] K. Yedavalli and B. Krishnamachari, "Sequence-based localization in wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 81–94, Jan. 2008.

[6] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello, "Bayesian filtering for location estimation," *IEEE Pervasive Comput.*, vol. 2, no. 3, pp. 24–33, Jul. 2003.

[7] K. Bauer, E. W. Anderson, D. McCoy, D. Grunwald, and D. C. Sicker, "CRAWDAD data set cu/rssi (v. 2009-05-28)," Downloaded from http://crawdad.org/cu/rssi/, May 2009.

[8] M. Youssef, A. Agrawala, and A. Udaya Shankar, "WLAN location determination via clustering and probability distributions," in *Proc. IEEE PerCom'03*, Mar. 2003, pp. 143–150.

[9] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '10. New York, NY, USA: ACM, 2010, pp. 173–184.

[10] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proc. ACM MobiSys'05*, 2005, pp. 205–218.

[11] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. ACM MobiCom'12*, 2012, pp. 269–280.

[12] J. Blumenthal, R. Grossmann, F. Golatowski, and D. Timmermann, "Weighted centroid localization in zigbee-based sensor networks," in *Proc. IEEE WISP'07*, Oct. 2007, pp. 1–6.

[13] R. M. Vaghefi, M. R. Gholami, R. M. Buehrer, and E. G. Strom, "Cooperative received signal strength-based sensor localization with unknown transmit powers," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1389–1403, March 2013.

[14] N. Patwari, A. Hero, M. Perkins, N. Correal, and R. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.

[15] R. W. Ouyang, A. K. S. Wong, and C. T. Lea, "Received signal strength-based wireless localization via semidefinite programming: Noncooperative and cooperative schemes," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1307–1318, March 2010.

[16] X. Zheng, H. Liu, J. Yang, Y. Chen, R. P. Martin, and X. Li, "A study of localization accuracy using multiple frequencies and powers," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 1955–1965, Aug. 2014.

[17] S. Chitte, S. Dasgupta, and Z. Ding, "Distance estimation from received signal strength under log-normal shadowing: Bias and variance," *IEEE Signal Process. Lett.*, vol. 16, no. 3, pp. 216–218, Mar. 2009.

[18] H. Aksu, D. Aksoy, and I. Korpeoglu, "A study of localization metrics: Evaluation of position errors in wireless sensor networks," *Comput. Netw.*, vol. 55, no. 15, pp. 3562–3577, Oct. 2011.

[19] E. Elnahrawy, X. Li, and R. Martin, "The limits of localization using signal strength: a comparative study," in *Proc. IEEE SECON'04*, Oct. 2004, pp. 406–414.

[20] R. M. Vaghefi and R. M. Buehrer, "Received signal strength-based sensor localization in spatially correlated shadowing," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 4076–4080.

[21] B. Wang, S. Zhou, W. Liu, and Y. Mo, "Indoor localization based on curve fitting and location search using received signal strength," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 572–582, Jan. 2015.

[22] S. H. Low and D. E. Lapsley, "Optimization flow control i: Basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1999.

[23] S. Shakkottai and R. Srikant, *Network Optimization and Control, in Foundations and Trends in Networking*. Boston - Delft: Now Publishers Inc, Jan. 2008.

[24] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.

[25] J. Jacob, B. R. Jose, and J. Mathew, "Spectrum prediction in cognitive radio networks: A bayesian approach," in *2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*, Sep. 2014, pp. 203–208.

[26] X. Xing, T. Jing, Y. Huo, H. Li, and X. Cheng, "Channel quality prediction based on bayesian inference in cognitive radio networks," in *2013 Proceedings IEEE INFOCOM*, April 2013, pp. 1465–1473.

[27] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics Automation Magazine*, vol. 13, no. 2, pp. 99–110, June 2006.

[28] S. Thrun, W. Burgard, and D. Fox, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Mach. Learn.*, vol. 31, no. 1-3, pp. 29–53, Apr. 1998.

[29] S. Särkkä, *Bayesian Filtering and Smoothing*. New York, NY, USA: Cambridge University Press, 2013.

[30] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Cengage Learning, 2001.

[31] H. Van Trees, K. Bell, and Z. Tian, *Detection Estimation and Modulation Theory, Part I*. Wiley, 2013.

[32] A. F. Molisch, *Wireless Communications*, 2nd ed. Chichester, West Sussex, U.K: Wiley, 2010.

[33] N. A. Jagadeesan. (2017) Lczn.jl. [Online]. Available: https://github.com/ANRGUSC/Lczn.jl

[34] D. Niculescu and B. Nath, "Ad hoc positioning system (APS) using AoA," in *Proc. IEEE INFOCOM'03*, vol. 3, Mar. 2003, pp. 1734–1743.

[35] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proc. ACM MobiCom'01*, 2001, pp. 166–179.

[36] N. A. Jagadeesan and B. Krishnamachari, "A unifying bayesian optimization framework for radio frequency localization," USC ANRG Technical Report, ANRG-2017-03, http://anrg.usc.edu/www/papers/tr_201703.pdf.

[37] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill Science/Engineering/Math, Jan. 1976.

[38] J. Wang, J. Chen, and D. Cabric, "Cramer-Rao bounds for joint RSS/DoA-based primary-user localization in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1363–1375, Mar. 2013.

[39] M. Angjelichinoski, D. Denkovski, V. Atanasovski, and L. Gavrilovska, "Cramér–Rao lower bounds of RSS-based localization with anchor position uncertainty," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2807–2834, May 2015.

[40] H. G. Tucker, "A generalization of the glivenko-cantelli theorem," *Ann. Math. Statist.*, vol. 30, no. 3, pp. 828–830, Sep. 1959.

Nachikethas A. Jagadeesan is a Ph.D. student in Electrical Engineering at University of Southern California. He received his B.Tech. in Engineering Physics from the Indian Institute of Technology, Madras, in 2011, and his M.S. degrees in Electrical Engineering and Statistics from University of Southern California in 2015 and 2017 respectively. His research interests broadly lie in wireless networks, mathematical statistics, and machine learning.

Bhaskar Krishnamachari received his B.E. in Electrical Engineering at The Cooper Union, New York, in 1998, and his M.S. and Ph.D. degrees from Cornell University in 1999 and 2002 respectively. He is a Professor in the Department of Electrical Engineering at the University of Southern California's Viterbi School of Engineering. His primary research interest is in the design, theoretical analysis and experimental evaluation of algorithms and protocols for next-generation wireless networks including low power wireless sensor networks. He has co-authored more than 200 publications on this topic, including best paper awards at Mobicom (2010), IPSN (2004, 2010), and MSWiM (2006), that have been collectively cited more than 20000 times per Google Scholar.