Distributionally Robust Radio Frequency Localization

Nachikethas A. Jagadeesan and Bhaskar Krishnamachari

***Abstract*—We consider the problem of estimating the location of an RF-device using observations such as received signal strengths, generated according to an uncertain distribution from a set of transmitters with known locations. We present a distributionally robust formulation of the localization problem that explicitly takes into account the uncertainty in the distribution that generates the observations. We identify the structure of the robust solution and demonstrate how to construct the optimization problem so that it is easily computed, and always yields the optimal solution. We show that the robust estimate outperforms traditional methods in the presence of modeling errors, while remaining close to the traditional estimate when the modeling is exact. This suggests that the formulation presented here is an attractive option in applications where we use a model that may not be an exact fit to our environment or if changes in our environment have induced errors in an empirically derived model.**

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

I. Introduction

The problem of consistently and accurately estimating the location of a wireless device in an indoor environment, merits significant scientific attention. In addition to the rich set of research challenges that this problem affords, there exists a plethora of applications that benefit from advances in solving this important problem [1]–[3]. These include indoor location based smartphone applications and services, search and rescue operations, geo-fencing, and asset tracking, to name a few. Localization services that utilize received signal strength (RSS) measurements are particularly attractive, since they impose no additional hardware requirements on the wireless devices, while remaining sufficiently accurate for many applications. Consequently, advances in localization algorithms based on RSS measurements can be rolled out easily as a software upgrade for a variety of wireless devices.

The literature on RSS based localization is rich and varied. Many works focus on the software systems and infrastructure required to enable localization services, while others focus on the design of algorithms that estimate the device location [4]–[10]. Underlying each of these works is a model that explains how the RSS observations vary with the device location. These models may be either empirically derived, as in the case of fingerprinting based methods, or an analytical model such as the log-normal path loss model may be used. In general, these models can be represented as a distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$, where $\boldsymbol{o}$ is the vector of observations and $\boldsymbol{r}$ is the device location. Adopting a Bayesian view, this distribution of observations may be used to derive a posterior belief of the device location, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$. Most localization algorithms may be viewed as methods that derive a location estimate from this posterior distribution [11]. However, the accuracy of the location estimates returned by these methods, as measured by various metrics such as the mean squared error (MSE), expected distance error (EDE), likelihood, etc., depends crucially on how well the chosen model approximates the actual distribution of the RSS observations.

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

While most works employ the services of a model to obtain the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$, not enough attention is given to the issue of how well the location estimate performs when the actual distribution deviates from that predicted by the model. This issue persists even when this distribution is empirically derived. Empirically derived distributions are sensitive to small changes in the environment and moreover, such distributions are often non-stationary, and hence change depending on the time of the day. This mismatch, between the distribution used to obtain the location estimate and the actual distribution generating the observations, results in subpar performance of localization algorithms. The performance guarantees of a localization algorithm is only valid as long as the model, upon which the algorithm is based, accurately tracks reality. Consequently, there is a need for an approach that explicitly takes into account the inherent ambiguity in modeling the environment, specifically the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$.

In this paper, we address these deficiencies by explicitly specifying the ambiguity in the distribution of observations, and deriving a location estimate that is resilient to such uncertainties. Our approach is applicable to both model-based methods, where the distribution $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$ is taken from a known family such as the log-normal path-loss model [12], and data-driven approaches such as fingerprinting [4], [7], [13], where the distribution $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$ is empirically constructed using observation data from each location. Uncertainty in the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$, results in an uncertain posterior, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$. We demonstrate how to construct an uncertainty set that contains all possible posterior distributions that we may wish to consider, and further show how to derive the robust location estimate. This robust formulation demonstrates better performance compared to the traditional approaches that do not take into account ambiguities in the underlying distribution. Moreover, even in the case where the distribution predicted by the model is accurate, the robust formulation tracks the performance of traditional localization methods very closely. In other words, the robust formulation presented here

gains increased resilience to inaccuracies in the model, while giving up very little performance compared to when the model is exactly true. This makes the robust formulation a very useful tool to deploy whenever we do not have perfect knowledge of the environment. Furthermore, the uncertainty set construction takes into account the confidence that we have in the model. The uncertainty set can be made more or less stringent in the distributions that it admits, depending on how close we think the model is to the true distribution. Moreover, the proposed robust formulation is computationally feasible and, as a consequence of the data parallelism inherent in the construction, its run time performance improves considerably when run in multi-core or multi-processor environments. In summary, our main contributions are:

- We describe a distributionally robust formulation of the indoor localization problem, that takes into account uncertainty in the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$.
- We derive and illustrate the structure of the robust solution, and further demonstrate how to construct the optimization problem in form that is computationally feasible and easily computed using standard software tools.
- We demonstrate that the robust formulation performs better than traditional methods when there are errors in the distribution of observations given by the model. In case when the model distribution is accurate, we show that the robust estimate closely tracks the performance of traditional methods.

The rest of this paper is organized as follows. In Section II, we review the existing literature on robust indoor localization, and place the current formulation in the context of various robustness targets and existing works in this direction. We introduce the formal problem statement in Section III. The construction of the uncertainty set is discussed in Section IV. Section V explores the structure of the robust solution. We formulate the problem in an easily computable form in Section VI, and discuss guidelines on choosing the various parameters of the robust formulation. We compare the robust formulation with traditional methods in Section VII. We discuss some future directions in Section VIII, and we conclude in Section IX.

II. Related Work

In this section, we give a brief overview of robust optimization methods and review the literature on robust localization. In general, robust methods aim to maintain satisfactory performance in the face of small variations from the model assumptions [14]. In this approach, we first specify an appropriate uncertainty set that captures the possible model variations that we wish to consider. The robust solution maximizes its performance with respect to the worst possible model contained within this uncertainty set [15]. The heart of the problem lies in choosing a set that sufficiently models the complexities of the problem at hand while remaining computationally tractable.

Recent years saw a dramatic increase in both the availability of data and the computational capability needed to process that data. Robust optimization methods have adapted to this trend by designing uncertainty sets that better utilize the data at hand [16], [17]. In this paper, we make use of these methods to develop data-driven distributionally robust solutions for the problem of indoor localization. While not addressing the problem of indoor localization, robust optimization methods have been employed in a similar setting [18], where objectives such as minimizing the energy consumption was considered under distance uncertainty.

Many indoor positioning systems base their location estimation on a model-based probabilistic description of location-dependent observations such as received signal strength (RSS) measurements obtained from transmitters located in the building [2]–[10], [13]. To estimate the receiver location using RSS measurements, we make use of a distribution that explains how the observation vector changes with location. Such a distribution may be derived using a model, such as the commonly used log-normal path loss model [5], [8], [19]–[23], or it may be empirically estimated using signal strength data collected at each location [4], [7], [13], [24]. Invariably, such a distribution makes assumptions about the observations and the environment that may not strictly hold in practice, giving rise to the need for robust solutions. We may develop different robust solutions depending on the uncertainties we wish to be robust against. We list below three common robustness targets that are most relevant to the indoor localization problem:

Robustness to Outliers: In this regime, we wish to be robust to arbitrary variation of a small subset of our observations. Such variation may be intentional as in the case of an attack mounted against the localization infrastructure. Thus this notion also confers some security properties. Specifically, this robustness target protects the localization system against arbitrary tampering of a subset of observations. The literature on robust indoor localization has focused mostly on this robustness target [25]–[33].

Robustness to Parameter Uncertainty: In this regime, we wish to be robust to small variations in the model parameters. In the context of indoor localization, this paradigm commonly assumes that the family that the distribution of the observation vectors belongs to is known. For instance, it is often assumed that the indoor received signal strength measurements have a log-normal distribution [34]. Consequently, we may wish to be robust to perturbations to the mean and variance of these observations. Another network parameter that is often a source of uncertainty is the location of the transmitters or anchors. In a model based approach to the indoor localization problem, it is assumed that we know the location of the devices that transmit a signal which is detected by the receiver. Some works [35]–[40] tackle the problem of developing algorithms that take into account the uncertainty in the location of some of these transmitters, also referred to as anchor nodes. The robust algorithms proposed in these works differ depending on the assumptions on the origin of the uncertainty in the transmitter locations. Thus, they address a specific source of uncertainty in contrast to the more general notion of distributional robustness presented in this paper.

Distributional Robustness: In this regime, we aim to be

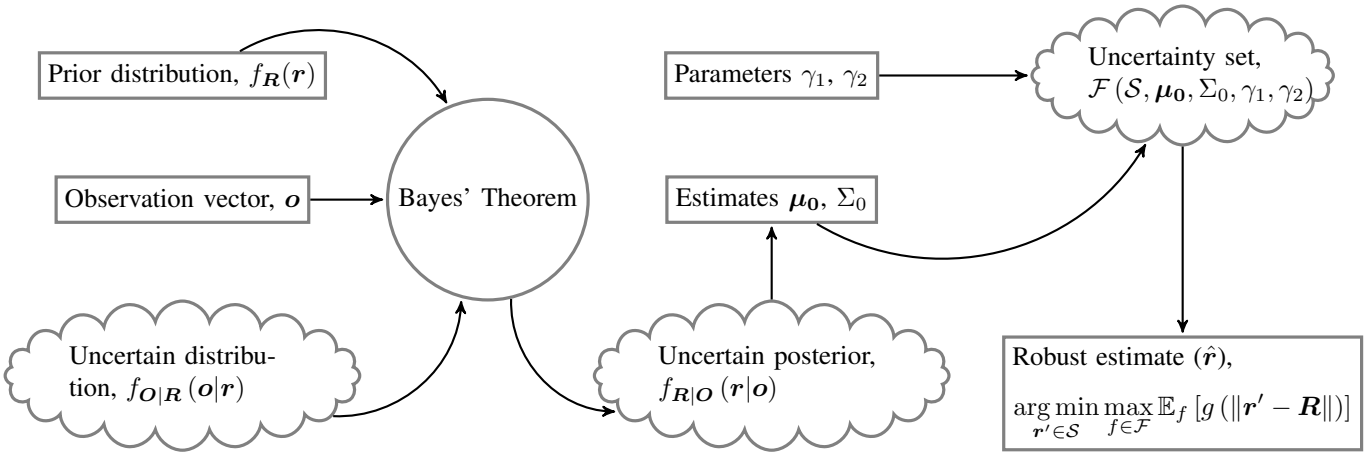


Fig. 1. Illustration of key steps involved in the construction of the uncertainty set and the computation of the robust estimate. The uncertainty in the posterior distribution, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$, stems from the uncertainty about the distribution of the observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. This uncertain posterior is used to derive the initial estimates $\boldsymbol{\mu_0}$ and Σ_0. These estimates, along with the parameters γ_1 and γ_2, determine the uncertainty set used to derive the robust estimate.

robust to changes in the underlying distribution of our observation vector. We specify an appropriate class of distributions for our posterior belief of the location, and then optimize over the worst case distribution from that class. Note that this is a generalization of robustness to parameter uncertainty. Indeed, being robust to parameter uncertainty is equivalent to specifying that our distribution set includes only distributions of a specified form (say, the Normal distribution), but with the distribution parameters (say, the mean and variance) taking values from within a specified set.

In this paper we aim for distributional robustness, and to the best of our knowledge, this work is the first to address distributionally robust indoor localization. Various calibration-free localization systems employ an alternate approach to deal with an uncertain distribution of observations. Crowd-sourcing based systems [41], [42] build and maintain a database of crowd-sourced observations which can be used to obtain an empirical estimate of the distribution of observations at each location. Other approaches [5], [43], [44], which may combine data from multiple sensors [45], are similar in spirit in that they incrementally build and update a signal map or model of the environment, which may be then used to estimate the unknown model parameters or unknown node locations. For instance, a recent work [46] uses a Gaussian-mixture distribution to model the observations and develops algorithms to simultaneously estimate the unknown sensor locations and noise parameters in an online manner. These works are also related to simultaneous location and mapping (SLAM) algorithms [47], [48], which rely on the knowledge of the motion of a robot to map the environment. Another class of works [49]–[51] often make use of a small transmitter (tag) on the device whose location is to be estimated, which can communicate with a set of anchors whose locations are known. Recently, such designs have been shown [52] to perform very well, being able to track even small fast moving objects.

The distributional robustness paradigm presented in this paper is complementary to these efforts. While we present this work in the context of localization using RSS observations for ease of exposition, it is applicable to any localization system that utilizes a statistical model, either empirically constructed or model-based, for the observational data, and it is agnostic to the manner in which the data is collected or maintained.

The robustness paradigm presented in this paper is related to the notion of misspecified models [53]. The fields of robust statistics and estimation theory under model misspecification both share the same goal of tackling the uncertain nature of the true distribution underlying the observational data. Under misspecified estimation theory, the distribution of observations is derived from a parametric family of distributions that may differ from the true family. The misspecified model may be chosen because of its analytical or computational tractability [53], and the impact of this mismatch is analyzed, including the derivation of bounds on the mean squared error, similar to the Cramér–Rao Bound [54]. In contrast, robust estimation methods are explicitly designed to give good performance over a large class of distributions for the observational data, at the cost of a slightly increased computational complexity. They minimize the error incurred by the estimate for all distributions within the specified uncertainty set, thereby explicitly bounding the performance impact of modelling errors.

III. Problem Statement

We focus our attention in this paper to the commonly considered case of 2D localization. However, the robust formulation presented here and the theoretical results carry over to the case when the space of interest is three dimensional. We begin by formulating indoor localization as a Bayesian optimization problem [11], which allows us to introduce our robustness requirements in a simple and natural manner.

Let $\mathcal{S} \subseteq \mathbb{R}^d$, where $d \in \{2, 3\}$ be the space of interest in which localization is to be performed. We assume that $\mathcal{S}$ is convex, closed, and bounded. Let the location of the receiver (the node whose location is to be estimated) be denoted as $\boldsymbol{r} \in \mathcal{S}$. Using a Bayesian viewpoint, we assume that this location is a random vector 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 distribution that depends on the receiver location $\boldsymbol{r}$: $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. 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 [34].

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

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

Traditionally, algorithms for localization are methods that derive a location estimate from the above posterior distribution. In this view, a 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 $\boldsymbol{o}$, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$,

to a location estimate $\hat{\boldsymbol{r}}$. Consequently, the usefulness of this location estimate is intimately tied to the validity of the derived posterior distribution. Our objective in this paper is to obtain a location estimate that accounts for ambiguity in this posterior. Such ambiguity might stem from either uncertainty about the prior or from uncertainty about the conditional distribution of the observation vector, or both. In this paper, our focus is on dealing with uncertainty about the conditional distribution of the observation vector, leaving the investigation of other cases for future work.

A. Distributionally Robust Formulation

Our uncertainty about the posterior distribution is specified by constructing a set of possible posterior distributions, denoted by $\mathcal{F}$. Given an estimate, say $\hat{\boldsymbol{r}}$, of the true location, we incur a cost that is assumed to depend only of the Euclidean distance between $\hat{\boldsymbol{r}}$ and the true location $\boldsymbol{r}$. Denote the cost function as $g(\|\boldsymbol{r} - \hat{\boldsymbol{r}}\|)$. We assume that $g : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ is a non-decreasing continuous function.

In the classical non-robust formulation, we construct a single posterior distribution function using Bayes' rule as given in equation (1). This posterior, say $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$, is used to derive a location estimate $\hat{\boldsymbol{r}}$ that solves the optimization problem

$$\hat{\boldsymbol{r}} = \arg\min_{\boldsymbol{r}'\in\mathcal{S}} \mathbb{E}_f \left[g\left(\|\boldsymbol{r}' - \boldsymbol{R}\| \right) \right], \tag{2}$$

where the expectation is over $\boldsymbol{R} \sim f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$ and $\mathcal{S}$ is our space of interest.

In the distributionally robust formulation, we minimize the cost of our estimate over the worst possible posterior distribution in $\mathcal{F}$. The worst possible distribution in $\mathcal{F}$ is the distribution in $\mathcal{F}$ which yields the highest cost in expectation. In other words, it is the distribution in $\mathcal{F}$ that maximizes $\mathbb{E}_f \left[g\left(\|\boldsymbol{r}' - \boldsymbol{R}\| \right) \right]$. Throughout this paper, we shall also refer to this as the worst case distribution. This robust formulation can be expressed as

$$\hat{\boldsymbol{r}} = \arg\min_{\boldsymbol{r}'\in\mathcal{S}} \max_{f\in\mathcal{F}} \mathbb{E}_f \left[g\left(\|\boldsymbol{r}' - \boldsymbol{R}\| \right) \right]. \tag{3}$$

The key to solving this robust formulation efficiently is to find an appropriate uncertainty set $\mathcal{F}$ that is not overly conservative while still providing satsifactory robustness guarantees. We discuss this issue in the following section.

IV. Uncertainty Set Construction

We define our uncertainty set $\mathcal{F}$ to be the class of distributions that have the mean and the covariance matrix to be close to our best estimate of our mean and covariance of $\boldsymbol{R}$. Denote this estimate of the mean $\boldsymbol{\mu_0}$, and the estimate of the covariance Σ_0. This estimate of the mean and covariance may be derived by utilizing the collected observation data. For a choice of tunable parameters, $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$, that express our confidence about the estimated mean and covariance, consider the set of distributions, $\mathcal{F}(\mathcal{S}, \boldsymbol{\mu_0}, \Sigma_0, \gamma_1, \gamma_2)$, such that each distribution function $f \in \mathcal{F}$ satisfies

$$P(\boldsymbol{R} \in \mathcal{S}) = 1, \tag{4a}$$

$$\left(\mathbb{E}_f[\boldsymbol{R}] - \boldsymbol{\mu_0}\right)^\top \Sigma_0^{-1} \left(\mathbb{E}_f[\boldsymbol{R}] - \boldsymbol{\mu_0}\right) \leq \gamma_1, \tag{4b}$$

$$\mathbb{E}_f\left[\left(\boldsymbol{R} - \boldsymbol{\mu_0}\right)\left(\boldsymbol{R} - \boldsymbol{\mu_0}\right)^\top \right] \preceq \gamma_2 \Sigma_0. \tag{4c}$$

These constraints define a set of distributions that satisfy the following properties: (i) The mean lies within an ellipsoid of size γ_1 centered at $\boldsymbol{\mu_0}$, and (ii) the covariance matrix lies within a positive semi-definite cone defined by the matrix inequality. It has been shown that such a construction of the uncertainty set often results in computationally tractable optimization problems that can be solved using mature, widely available software packages [17], [55]. Moreover, this construction is conceptually simple while still encompassing a rich collection of interesting uncertainty sets [55] and naturally incorporating the use of all the available data [16].

The initial estimate of the mean, $\boldsymbol{\mu_0}$, and the covariance, Σ_0, is crucial to the above formulation. These may be derived using the available observation data. In the following section we illustrate how to incorporate our observation data in the uncertainty set construction. In addition, we identify the properties of the robust solution and simplify the above problem to a form that is easily solved using a software solver.

A. Initial Estimate for Uncertainty Sets

Typically localization systems collect observations that depend on location and then describe this relationship either using an analytical model or using an empirically derived distribution. Such descriptions are used to construct the distribution of the observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. Thus the source of uncertainty in the posterior, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$, may be viewed as stemming from

uncertainty about the conditional distribution of the observation vector, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$.

Given an uncertainty set containing possible distributions for $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$, we can conceptually derive an uncertainty set for the posterior by applying Bayes' rule for each candidate distribution for $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. However, such a naïve approach is not computationally feasible since the uncertainty sets are not necessarily finite. Thus, we use a model-derived or empirically estimated distribution for $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$ to derive the estimates $\boldsymbol{\mu_0}$ and Σ_0. For RSS data, the log-normal model may be used. Denote this model-derived or empirically estimated distribution as $\hat{f}_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. Using this distribution, we can derive a candidate posterior using Bayes' rule. Denote this derived posterior distribution as $\hat{f}_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$. Then,

$$\boldsymbol{\mu_0} = \mathbb{E}_{\hat{f}}[\boldsymbol{R}], \tag{5}$$

$$\Sigma_0 = \mathbb{E}_{\hat{f}}\left[(\boldsymbol{R}-\boldsymbol{\mu_0})(\boldsymbol{R}-\boldsymbol{\mu_0})^\top\right]. \tag{6}$$

The estimates given above, along with a choice of the parameters $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$, completes the description of our uncertainty set, $\mathcal{F}(\mathcal{S}, \boldsymbol{\mu_0}, \Sigma_0, \gamma_1, \gamma_2)$, for the posterior. Note that the initial estimates presented in equations (5),(6) are the MMSE estimate and its covariance, under the derived posterior distribution, $\hat{f}_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$. Thus the uncertainty set can be viewed as a representation of our confidence in the assumptions under which the MMSE estimate was derived. Specifically, it represents our confidence in the derived posterior distribution, $\hat{f}_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$, and hence in the MMSE estimate derived from this distribution. The parameters $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$, allow us to constrain the distributions admitted inside the uncertainty set, thereby limiting how far the robust estimate is allowed to stray from the MMSE estimate.

V. Solution Structure

In this section we investigate the structure of the solution to the inner moment problem $\max_{f\in\mathcal{F}} \mathbb{E}_f[g(\|\boldsymbol{r}' - \boldsymbol{R}\|)]$, for sufficiently large values of the parameters γ_1 and γ_2, leaving the discussion on the impact of γ_1 and γ_2 for the subsequent section. The following formulation is inspired by Scarf's classical result in inventory theory [56]. We show that for any given candidate location $\boldsymbol{r}'$, the distribution that yields the maximum cost, in other words the worst case distribution, will have positive support only on the boundary of our space, $\partial\mathcal{S}$. This is formalized in the following theorem.

Theorem 1. *For any $\boldsymbol{r}' \in \mathcal{S}$ and a non-decreasing continuous cost function $g : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$, there exists $\boldsymbol{r}^* \in \partial\mathcal{S}$ such that*

$$\max_{f\in\mathcal{F}} \mathbb{E}_f[g(\|\boldsymbol{r}' - \boldsymbol{R}\|)] \leq g(\|\boldsymbol{r}' - \boldsymbol{r}^*\|). \tag{7}$$

Proof. See Appendix B □

Theorem 1 indicates that attempting to maximize our expected cost, $\mathbb{E}_f[g(\|\boldsymbol{r}' - \boldsymbol{R}\|)]$, pushes the support of the resulting distribution f closer to the boundary of our space $\mathcal{S}$. Furthermore, since our space of interest $\mathcal{S}$ is convex and in $\mathbb{R}^2$ or $\mathbb{R}^3$, we can efficiently approximate $\mathcal{S}$ using a convex polyhedron [57], [58]. Using such an approximation allows us to further simplify the structure of the worst case distribution. Let $\hat{\mathcal{S}}$ be our convex polyhedron approximation of $\mathcal{S}$. For any point $\boldsymbol{r}' \in \hat{\mathcal{S}}$, the point in $\hat{\mathcal{S}}$ that is farthest from $\boldsymbol{r}'$ is one of the vertices of $\hat{\mathcal{S}}$. Consequently, for a convex polyhedron $\hat{\mathcal{S}}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, we have the following refinement of Theorem 1.

Theorem 2. *If $\hat{\mathcal{S}}$ is a convex polyhedron in $\mathbb{R}^2$ or $\mathbb{R}^3$, then for each $\boldsymbol{r}' \in \hat{\mathcal{S}}$, there exists a vertex $\boldsymbol{v} \in \hat{\mathcal{S}}$ such that*

$$\max_{f\in\mathcal{F}} \mathbb{E}_f[g(\|\boldsymbol{r}' - \boldsymbol{R}\|)] \leq g(\|\boldsymbol{r}' - \boldsymbol{v}\|). \tag{8}$$

Proof. See Appendix C □

Theorem 1 indicates that the worst case distribution has support only on the boundary of our space. Moreover, in case we employ a convex polyhedron approximation for our space of interest, Theorem 2 suggests that we only need to consider distributions that have support on the vertices. However, it is worth noting that is not guaranteed that these bounds are attained by a distribution within $\mathcal{F}$ for all values of γ_1 and γ_2.

For instance, consider the scenario where we attempt to simplify the inner moment problem, $\max_{f\in\mathcal{F}} \mathbb{E}_f[g(\|\boldsymbol{r}' - \boldsymbol{R}\|)]$, by reducing our search space to include only those distributions within $\mathcal{F}$ that have support only on the vertices of $\hat{\mathcal{S}}$. There always exists such distributions that can satisfy constraint (4b) for any choice of μ_0 and γ_1. However, the same cannot be said for constraint (4c). Considering only distributions with support on the vertices of $\hat{\mathcal{S}}$ effectively imposes a lower bound on the covariance matrix of $\boldsymbol{R}$, and hence the parameters γ_2 and Σ_0 needs to be chosen with some care. We will revisit this issue of parameter selection in the following section.

VI. Simplified Formulation

We now turn our attention to the issue of simplifying our optimization problem (3) to a form that is easily computed using standard software tools. Assume that $\mathcal{S}$ is a convex polygon. If the original space is convex but not a polygon, we can always find a convex polygon approximation with a desired level of accuracy [57]. We then discretize the distributions within $\mathcal{F}$ by allowing them to have support only on a discrete grid-like set of locations, say V, within $\mathcal{S}$. This imposes nearly no compromises since we can get arbitrarily close to the continuous setting by making our grid progressively finer. The vertices of $\mathcal{S}$ are always included in V. This ensures that the convex hull of V always returns $\mathcal{S}$, $\mathrm{conv}(V) = \mathcal{S}$.

This reduction of the uncertainty set to include only discrete distributions allows for us to write our optimization problem in a simpler manner. Let $V = \{\boldsymbol{r}_i\}_{i=1}^n$ represent the locations of the grid points within $\mathcal{S}$. The vector $\boldsymbol{p}$ represents a distribution over this set of grid points. Define

$$A_i = (\boldsymbol{r}_i - \boldsymbol{\mu}_0)(\boldsymbol{r}_i - \boldsymbol{\mu}_0)^\top \tag{9}$$

$$B_i = \begin{bmatrix} \Sigma_0 & (\boldsymbol{r}_i - \boldsymbol{\mu}_0) \\ (\boldsymbol{r}_i - \boldsymbol{\mu}_0)^\top & \gamma_1 \end{bmatrix}, \tag{10}$$

for all $i \in \{1, 2, \ldots, n\}$. For each candidate location $\boldsymbol{r} \in \mathcal{S}$, let $\boldsymbol{g_r}$ represent the cost vector for that location, $(\boldsymbol{g_r})_i = g(\|\boldsymbol{r} - \boldsymbol{r}_i\|)$ for all $i \in \{1, 2, \ldots, n\}$. Then for each candidate location

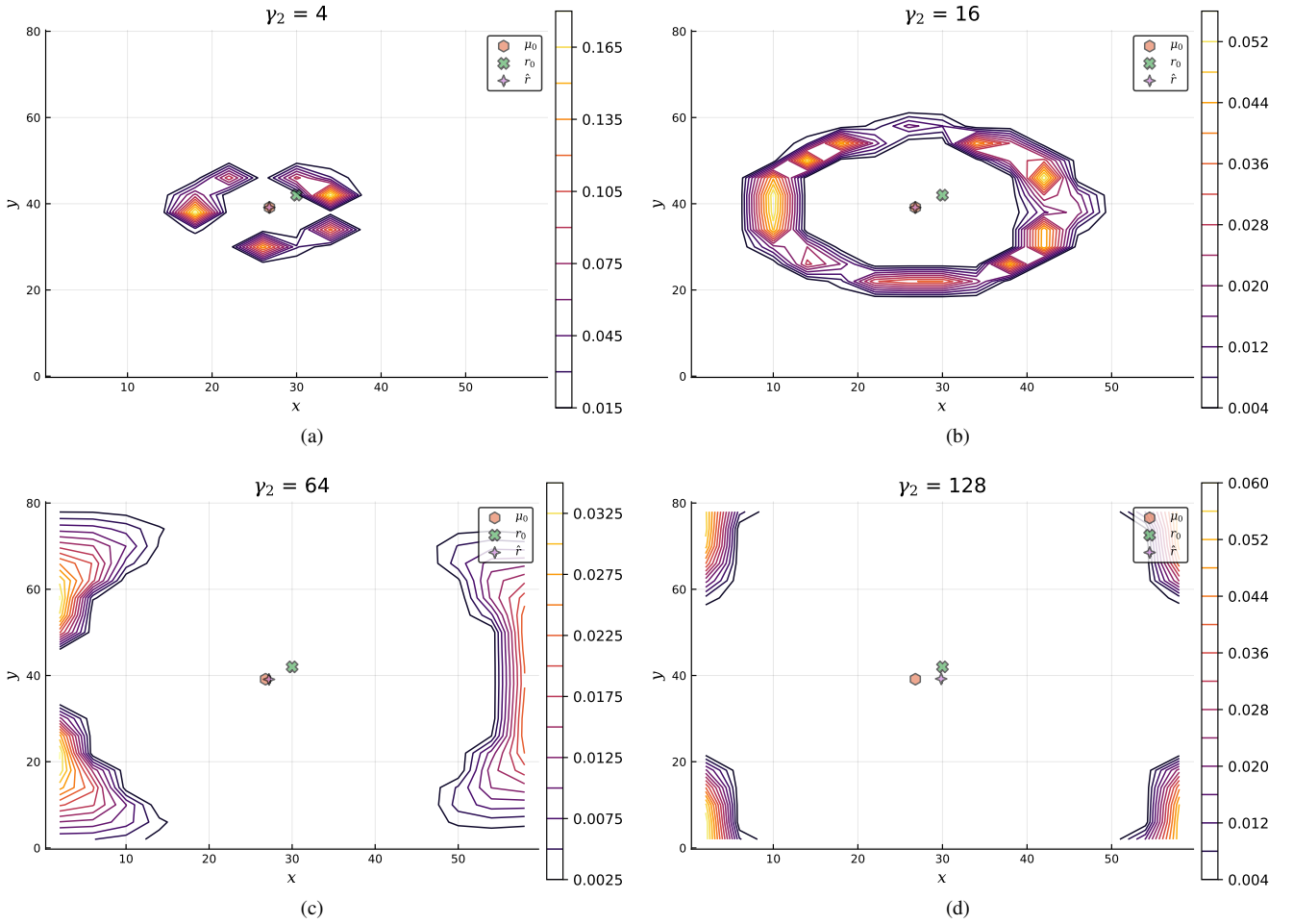


Fig. 2. An illustration of the posterior distribution corresponding to the robust estimate ($\hat{\boldsymbol{r}}$), for various values of the parameter γ_2. The MMSE estimate ($\boldsymbol{\mu_0}$) and the actual receiver location ($\boldsymbol{r}_0$) are shown as well. As predicted by Theorems 1 and 2, the worst-case distribution is pushed towards the boundary and vertices of the rectangular 60 m by 80 m space.

$\boldsymbol{r}$, the inner moment problem in (3) can be represented by the following semi-definite program (SDP) [59]:

$$\begin{aligned}
\text{maximize} \quad & \boldsymbol{g}_{\boldsymbol{r}}^{\top} \boldsymbol{p} && \text{(11a)}\\
\text{subject to} \quad & \sum_{i=1}^{n} p_i A_i \preceq \gamma_2 \Sigma_0, && \text{(11b)}\\
& \sum_{i=1}^{n} p_i B_i \succeq 0, && \text{(11c)}\\
& \sum_{i=1}^{n} p_i = 1, && \text{(11d)}\\
& p_i \geq 0 \quad \text{for all } i \in \{1, 2, \ldots, n\}. && \text{(11e)}
\end{aligned}$$

The solution to the above optimization problem gives us the worst-case distribution corresponding to the candidate location $\boldsymbol{r}$. Note that, for any given $\boldsymbol{r}$, the above inner moment problem (11) is a SDP that can be solved efficiently both in theory and practice [60]. As we typically deal with a finite number of possible location estimates, we can potentially enumerate the solution of problem (11) for all possible location estimates and then choose the best among them. Thus the problem remains computationally feasible in its current form. Moreover, as the optimization at the candidate locations are independent of each other, we can further accelerate the process using parallel or distributed computing.

Alternately, we can formulate the inner moment problem (11) in its dual form and use the fact that the minimization operations may be performed jointly. This yields the following SDP formulation to obtain the robust location estimate:

$$\begin{aligned}
\underset{Z_1, Z_2, \nu, \boldsymbol{r}, \alpha}{\text{minimize}} \quad & \alpha && \text{(12a)}\\
\text{subject to} \quad & \boldsymbol{r} \in \mathcal{S}, && \text{(12b)}\\
& Z_1, Z_2 \succeq 0, && \text{(12c)}\\
& \gamma_2 \operatorname{tr}(\Sigma_0 Z_1) + \nu \leq \alpha, && \text{(12d)}\\
& \operatorname{tr}(A_i Z_1) - \operatorname{tr}(B_i Z_2) + \nu - (\boldsymbol{g}_{\boldsymbol{r}})_i \geq 0, && \text{(12e)}
\end{aligned}$$

for all $i \in \{1, 2, \ldots, n\}$. The SDP formulation given above (12) can be solved easily using a standard solver such as CVXPY [61] or Convex.jl [62]. This dual form has the advantage that we can obtain both the robust estimate and the corresponding posterior distribution, through the dual variables of SDP (12), using a single optimization program. An illustration of the same is given in Figure 2 for various values of the parameter γ_2. As can be seen in the aforementioned

figure, different choices of the parameter γ_2 yields very different solutions. For large values of γ_2, the support of the posterior get progressively closer to the vertices of $\mathcal{S}$. This raises the question of how we should choose these parameters for a particular problem at hand. More fundamentally, we need to know the range of parameters for which we are guaranteed that an optimal solution exists. We explore this topic below.

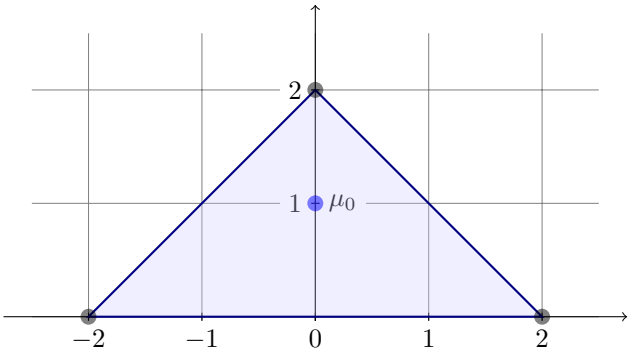


Fig. 3. An illustration of valid parameter selection. Assume that we restrict $\boldsymbol{p}$ to have support only on the vertices. Say we have an initial estimate $\boldsymbol{\mu_0} = (0, 1)$. Then the covariance matrix $\mathbb{E}_{\boldsymbol{p}}\left[(\boldsymbol{R}-\boldsymbol{\mu_0})(\boldsymbol{R}-\boldsymbol{\mu_0})^\top\right]$ is lower bounded by $\left[\begin{smallmatrix}2 & 0\\ 0 & 1\end{smallmatrix}\right]$ for small values of γ_1. In particular, the problem (3) is infeasible for the choice of parameters $\gamma_1 = 1$, $\gamma_2 = 1.5$, and $\Sigma_0 = \frac{I}{2}$, where I is the identity matrix. The optimization becomes feasible if $\Sigma_0 = 2I$.

A. Parameter Selection

In this section, we investigate the issue of how to choose the parameters involved in the construction of our uncertainty set $\mathcal{F}(\mathcal{S}, \boldsymbol{\mu_0}, \Sigma_0, \gamma_1, \gamma_2)$, which consequently determines the optimization problem (12) that generates the robust estimate. As indicated in Section IV-A, the mean $\boldsymbol{\mu_0}$ and the covariance matrix Σ_0 of the posterior distribution $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{o}|\boldsymbol{r})$ is obtained either using a model such as the log-distance path loss model [12] or using an empirically estimated distribution $\hat{f}_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. It remains to be shown that the low complexity formulation (12) always yields an optimal solution for any choice of $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$.

Recall that each distribution in our uncertainty set $\mathcal{F}$ must satisfy the constraints (4a), (4b), and (4c). In the low complexity formulations (11) and (12), our search space involves distributions that have support on a discrete set of locations within $\mathcal{S}$. Since all locations with positive support lie in $\mathcal{S}$, constraint (4a) is always satisfied. Since $\mathcal{S}$ is a closed, and convex polygon in $\mathbb{R}^2$ or $\mathbb{R}^3$, any point in its interior can be represented as a convex combination of its vertices [58]. In other words, for any choice of $\boldsymbol{\mu_0}$ in the interior of $\mathcal{S}$, there exists a distribution $\boldsymbol{p}$ such that $\mathbb{E}_{\boldsymbol{p}}[\boldsymbol{R}] = \boldsymbol{\mu_0}$. Thus the constraint (4b) is satisfied strictly for any choice of $\gamma_1 > 0$. In addition, algorithms to compute convex hulls of a finite set of points in $\mathbb{R}^2$ and $\mathbb{R}^3$ are well known in the literature [63], [64], so this distribution can be explicitly computed if need be.

The satisfiability of the final constraint (4c) is sensitive to our choice of Σ_0 and our choice of V, the discretization of $\mathcal{S}$. To illustrate this, consider the example given in Figure 3. Assume that we force the distribution $\boldsymbol{p}$ to have positive support only on the vertices of $\mathcal{S}$. Then for a sufficiently small value of γ_1, constraint (4b) determines $\boldsymbol{p}$, which in turn determines the covariance matrix $\hat{\Sigma} = \mathbb{E}_{\boldsymbol{p}}\left[(\boldsymbol{R}-\boldsymbol{\mu_0})(\boldsymbol{R}-\boldsymbol{\mu_0})^\top\right]$. Thus, if we happen to choose an initial estimate $\Sigma_0 \prec \hat{\Sigma}$, then our optimization problem (12) is infeasible for some values of $\gamma_2 \geq 1$. In summary, restricting $\boldsymbol{p}$ to have support only on the vertices effectively imposes a lower bound on the covariance matrix which needs to be taken into consideration while choosing the γ_2 and Σ_0.

A similar situation holds, albeit to a significantly lesser degree, when we discretize $\mathcal{S}$ to V. As might be expected, the lower bound on the covariance matrix is now determined by the points in V closest to $\boldsymbol{\mu_0}$. Indeed, as we make V finer the lower bound becomes progressively loose and it disappears entirely when there exists a location in V that coincides with $\boldsymbol{\mu_0}$. This can always be ensured while constructing V. Hence the constraint (4c) is satisfied strictly for any choice of $\gamma_2 > 1$. Thus, by this construction of the uncertainty set, we see that Slater's condition holds for the primal problem (11), and hence strong duality holds [59]. Consequently, for any choice of $\boldsymbol{\mu_0}$ in the interior of $\mathcal{S}$, the optimum of the dual problem (12) is attained.

Algorithm: robust location estimation

input : prior distribution $f_{\boldsymbol{R}}(\boldsymbol{r})$
observation vector $\boldsymbol{o}$
distribution of observations $\hat{f}_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$
cost function g
space of interest $\mathcal{S}$
parameter $\gamma_1 > 0$
parameter $\gamma_2 > 1$
number of grid points n

output : robust location estimate $\hat{\boldsymbol{r}}$

begin
- compute $\hat{f}_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$ ▷ via (1)
- compute $\boldsymbol{\mu_0}$ ▷ via (5)
- compute Σ_0 ▷ via (6)
- $V \leftarrow Discretize(\mathcal{S}, n-1) \cup \boldsymbol{\mu_0}$
- compute $\{A_i\}_{i=1}^n$ ▷ via (9)
- compute $\{B_i\}_{i=1}^n$ ▷ via (10)
- construct and solve dual SDP ▷ via (12)
- **return** $\hat{\boldsymbol{r}}$

end

Fig. 4. A summary of the steps required to compute the robust estimate using the simplified dual formulation

B. Computational Complexity

In our formulation, the robust location estimate is obtained as the solution to a semi-definite program. Consequently, our formulation is computationally tractable, both in theory and in practice [60], [65]. The complexity of the SDP is determined by its size [66], and in our formulation the size of the SDPs (11) and (12) is proportional to the size of the uncertainty set of distributions. This size of the uncertainty set is determined by the size of our space $\mathcal{S}$. In our simplified formulation,

the size of our space can be measured by the number of grid points within $\mathcal{S}$, namely n. Thus, as we make our grid progressively finer by increasing n, we increase the number of distributions supported by our uncertainty set. This has the effect of increasing the time required to compute the robust estimate. To capture this scaling behaviour, we only need to specify the computational complexity of solving the SDP formulation (12). Note that the size of the matrices A_i, B_i for $i \in \{1, \ldots, n\}$, Z_1, and Z_2 is determined by the dimension of our space. Since our space $\mathcal{S}$ will lie in $\mathbb{R}^2$ or $\mathbb{R}^3$, this is a constant. Recall that we assume $\mathcal{S}$ to be a convex polygon. Let t denote the number of faces of $\mathcal{S}$. Then the SDP (12) can be solved in $\mathcal{O}(n^{1.5}t^3)$ operations [66]. To obtain the robust estimate using formulation (11), note that we need to solve n such inner moment problems, each corresponding to a candidate location. Each of these inner moment problems can be solved in $\mathcal{O}(n^{1.5})$ time, independent of each other. The solution to each such problem returns the cost, $\boldsymbol{g_r}^\top \boldsymbol{p}$, and the robust estimate is the location that yields the least cost. This approach lends itself very well to a distributed implementation, such as MapReduce [67]. In a distributed implementation, the mobile device might communicate its observation data to a localization server which computes the location estimate using a pool of worker nodes and then communicates the estimate back to the device. Such an implementation reduces the computation time while incurring a communication overhead.

So far we have discussed the complexity of solving the SDPs (11) and (12) to obtain the robust estimate. In constructing the uncertainty set used in these SDPs, we need to obtain initial estimates for the location, $\boldsymbol{\mu_0}$, and the covariance matrix, Σ_0. As indicated in Section IV, these estimates are computed using a model-derived or an empirically estimated distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. We may use the services of estimators, such as MMSE, to obtain these initial estimates. The computational complexity of the robust estimate, presented in the preceding paragraph, is incurred in addition to this cost of finding the initial estimates. This presents us with a trade-off between robustness and computational complexity, as one could simply use the initial estimates if robustness was not a concern. In this regard, the choice of implementing the robust formulation presented here should be made depending on the application needs, and the level of uncertainty in modelling the environment.

VII. Evaluation

In this section, we investigate the performance of the robust estimate compared to the MMSE estimator under simulations and real world experiments.

Firstly, we consider the wireless setting that corresponds to large-scale fading or shadowing. In this setting, the observations, corresponding to the average received power level at a given location, are drawn from a log-normal distribution [12]. In this case, we set the MMSE estimator to correctly assume the distribution of observations. The observations arriving from different transmitters may be correlated with each other. We consider both settings corresponding to correlated observations, and when the observations are independent and identically distributed (IID).

Secondly, we consider the wireless setting that corresponds to small-scale fading. In this setting, the signal strength at a location fluctuates about the average value according to a Rayleigh distribution [12]. In this case, the MMSE estimator ignores the fluctuations about the mean power level.

Thirdly, we incorporate the non-stationary behavior exhibited by signal strength measurements [68]. In this setting, the variance of the observations changes depending on the time of the day. A higher variance is used to model observations generated during working hours (daytime), while a lower variance is used for nighttime observations.

Finally, we evaluate the performance of the robust estimate under a real world setting. We collect 802.11 received signal strength observations in an indoor office environment under two scenarios, namely, daytime and nighttime. The daytime observations correspond to a setting with relatively high variance, while the nighttime observations correspond to a relatively low variance setting. In all of these cases, the initial estimates for the uncertainty set are obtained using the model assumed by the MMSE estimator.

A. Large-Scale Fading

Say $\{\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_m\}$ $(m > 2)$ are the known positions of (m) 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 $\boldsymbol{l}_i = (x_i, y_i) \in \mathcal{S}$ $\forall i \in \{1, 2, \ldots, m\}$. 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

$$P_r^i = P_t^i K \left[\frac{d_0}{d_i}\right]^\eta W_i, \tag{13}$$

where the distance between the receiver and the i^{th} transmitter is given by $d_i(\boldsymbol{r}) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$, 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

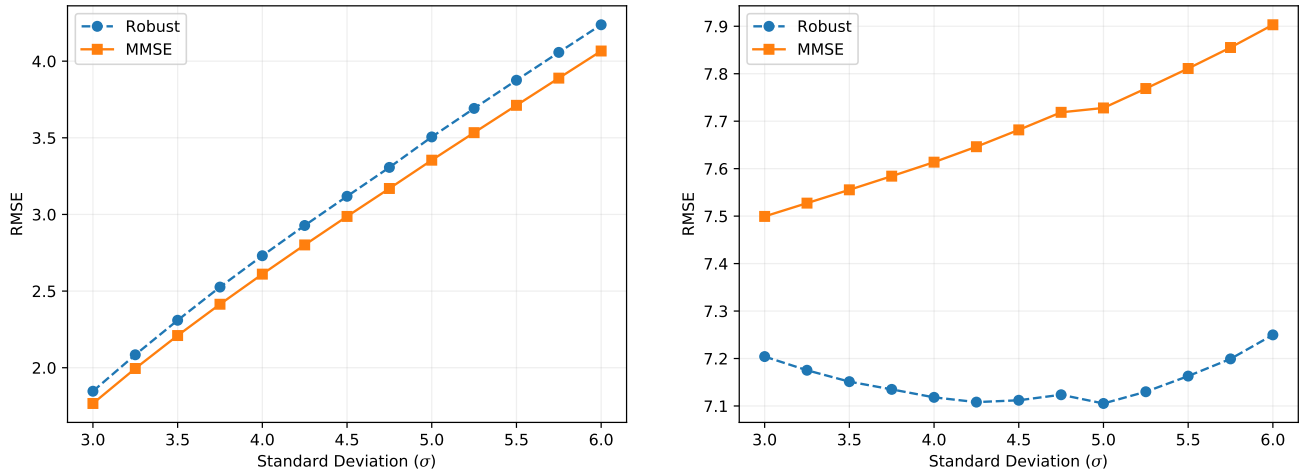
$$P_r^i|_{\mathrm{dBm}} = P_t^i|_{\mathrm{dBm}} + K|_{\mathrm{dB}} - 10\eta \log_{10}\left[\frac{d_i}{d_0}\right] + W_i|_{\mathrm{dB}}, \tag{14}$$

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 $1\mathrm{m}$. In this setting, our estimation problem may be restated as follows. We are given measurements of the receiver signal strengths $\left\{P_r^1, P_r^2, \ldots, P_r^m\right\}$ 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 $O_i \sim \mathcal{N}\left(-10\eta \ln\left[d_i(\boldsymbol{r})\right], \sigma^2\right)$.

For the setting where the received observations are IID, the distribution of the observation vector $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$ can be obtained from the above by taking the product of all the



(a) (b)

Fig. 5. Performance of the MMSE and Robust estimates on increasing noise variance. For this illustration, a log-normal fading model was used to generate the observations in a rectangular 60 m by 80 m space using identical parameters as that used in Figure 2. In Plot 5a, the observations are drawn in an IID manner. This corresponds to the best case scenario for MMSE. Note that in this case the performance of the robust estimate tracks that of the MMSE estimator very closely. In Plot 5b, the observations are correlated with each other. We see that the robust estimate performs much better than MMSE in this case.

marginal observation pdfs. To evaluate the setting where the observations may be correlated, we transform the observation vector that was obtained above. Under the IID assumption, the covariance matrix of $\boldsymbol{O}|\boldsymbol{R}$ is given by

$$\tilde{\Sigma} = \mathbb{E}_{f_{\boldsymbol{O}|\boldsymbol{R}}}\left[(\boldsymbol{O}-\bar{\boldsymbol{o}})(\boldsymbol{O}-\bar{\boldsymbol{o}})^{\top}\right] \tag{15}$$

$$= \sigma^2 I, \tag{16}$$

where I is the identity matrix, and $\bar{\boldsymbol{o}} = \mathbb{E}_{f_{\boldsymbol{O}|\boldsymbol{R}}}[\boldsymbol{O}]$. Using a transformation matrix C, we can transform the observation vector as $\boldsymbol{O}^* = C\boldsymbol{O}$. This yields the updated covariance matrix, $\tilde{\Sigma}^* = \sigma^2 CC^{\top}$. The transformed observations are now jointly normal with covariance $\tilde{\Sigma}^*$. We would like the transformation matrix to encode the property that the observations coming from transmitters that are closer to each other are likely to be more correlated compared to a pair that is spaced farther apart. This can be achieved by setting $C_{ij} = \frac{1}{1+\beta d_{ij}}$, where d_{ij} is the distance between the transmitters i and j, and β is a parameter that can be used to tune the strength of the induced correlations. In this manner, we can evaluate the robust formulation under the setting of correlated observations. The specific values of the parameters used in the evaluation are given below.

The dimensions of the area of interest ($\mathcal{S}$) was $60\,\mathrm{m} \times 80\,\mathrm{m}$. Sixteen transmitters were chosen randomly and 64 RSSI readings were taken for each transmitter at 300 distinct receiver locations. The transmit power was kept constant at $16\,\mathrm{dBm}$. The model parameters are a path loss of $K = 39.13\,\mathrm{dB}$ at reference distance $d_0 = 1\,\mathrm{m}$, and path loss exponent $\eta = 3.93$. The robust estimator uses the cost function corresponding to the MSE, $g(\|\boldsymbol{r}-\hat{\boldsymbol{r}}\|) = \|\boldsymbol{r}-\hat{\boldsymbol{r}}\|^2$. The parameters of the robust estimator are, $\gamma_1 = 8$ and $\gamma_2 = 8$. The value of β was chosen to be 4. The prior distribution was assumed to be uniform over $\mathcal{S}$ which corresponds to the case where we have no prior knowledge of device location.

In Figure 5, the MMSE estimator correctly fits a log-normal distribution over the received power levels, which is used to generate the initial estimates, $\boldsymbol{\mu_0}, \Sigma_0$, used by the robust estimator. In this setting, the MMSE estimator is expected to be the best performing estimator as measured by the RMSE metric. This behavior is shown in Figure 5. However, it is notable that the robust estimate does not deviate too far from the MMSE estimate, with the mean RMSE of the robust estimate being at most $0.2\,\mathrm{m}$ above that of the MMSE estimate.

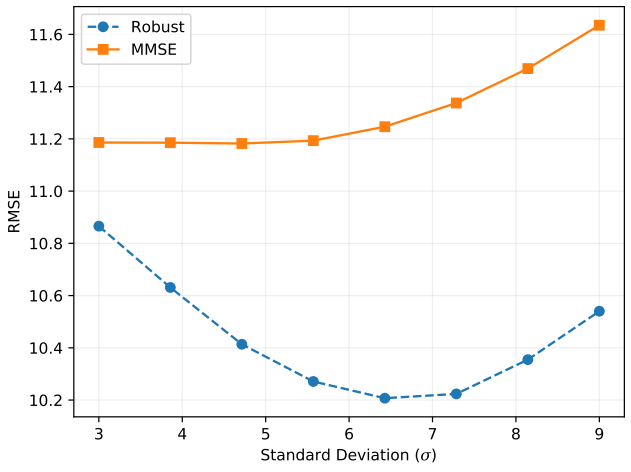


Fig. 6. Performance of the MMSE and Robust estimates on increasing noise variance. For this illustration, a Rayleigh fading model was used to generate the observations in a rectangular 60 m by 80 m space using identical parameters as that used in Figure 5. The observations are correlated. In this plot, the MMSE estimator incorrectly assumes that the observations are drawn from a log-normal distribution. The robust estimate consistently performs better than the MMSE estimator here.

B. *Small-Scale Fading*

Consider the setting where the signal strength at any given location is the result of superposition of different multi path components, none of which is dominant. Then, the in-phase

and the quadrature-phase component of the received signal is the sum of many random components and is well approximated by a normal distribution using the central limit theorem [12]. Consequently, the absolute amplitude of the received signal follows a Rayleigh distribution. The squared amplitude, and hence the power, follows an exponential distribution.

Let the mean received signal power, for a given transmitter-receiver pair, say i, be given by P_r^i. Note that this local mean, P_r^i, follows a log-normal distribution as given in equation (13). The received signal power, say $\tilde{P}_r^i$, fluctuates around this average value according to a exponential distribution,

$$\tilde{P}_r^i = P_r^i Z_i \tag{17}$$

where $Z_i \sim \text{Exp}(1)$. The distribution of the received power, $\tilde{P}_r^i$, is a product of a log-normal random variable, corresponding to the mean power level, and an exponential random variable, corresponding to fluctuations about this mean.

In Figure 6, the MMSE estimator effectively ignores these fluctuations about the mean value. The MMSE estimator fits a log-normal distribution over the received power levels, ignoring the fluctuations due to small scale fading. This incorrect distribution used by the MMSE estimator is used to generate the initial estimates, $\boldsymbol{\mu_0}, \Sigma_0$, used by the robust estimator. As can be seen in Figure 6, the robust estimator outperforms the MMSE estimator in this case, which demonstrates its usefulness in the presence of modeling errors.

C. Non-Stationary Behavior

It is well known that signal strength measurements in indoor environments exhibit non-stationary behavior [68]. Specifically, the variability exhibited by the signal strength measurements varies depending on the time of the day. During daytime, the presence of an increased number of people indoors creates a dynamic environment that results in increased variability for the signal strength measurements. This effect is reduced during the evening and nighttime.

To evaluate the robust formulation under this setting, we generate two sets of observations using the small-scale fading model. The ratio of the standard deviation of the two sets of observations is given by $\frac{\sigma_1}{\sigma_2} = \theta$, where θ is a parameter. For the illustration in Figure 7, we choose $\theta = 2$. The smaller standard deviation is given by $\sigma_2 = \sigma\sqrt{\frac{2}{1+\theta^2}}$, where σ is a parameter we can use to vary the overall variance of the combined set of observations. This parameter σ is used as the standard deviation of the log-normal model used by the MMSE estimator and thus to generate the initial estimates for the robust formulation. The performance of the robust formulation, and the MMSE estimator is shown in Figure 7. The robust formulation performs better in this setting as well, further strengthening the case for its adoption when the model is not known exactly.

Figures 6 and 7 show that the robust estimator initially improves its performance on increasing the noise variance. The performance degrades on continuing to increase the variance. In the above simulations the initial estimates, $\boldsymbol{\mu_0}, \Sigma_0$, were obtained assuming a log-normal fading model for the observations. At low variances, the covariance matrix is small, and hence the robust estimator is constrained by the covariance

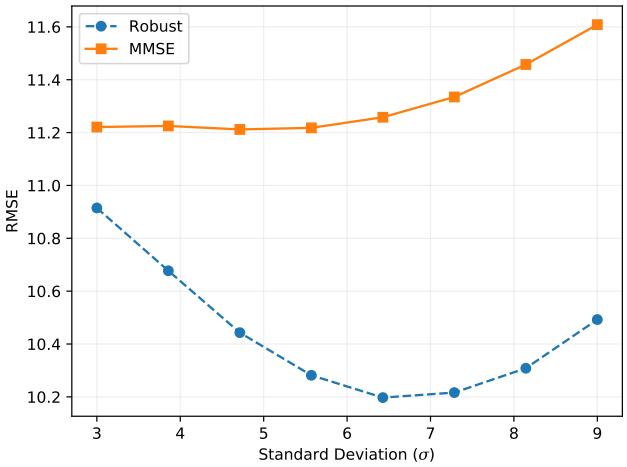


Fig. 7. Performance of the MMSE and Robust estimates on increasing noise variance. For this illustration, a Rayleigh fading model with different noise variances for the mean power was used to simulate the non-stationary behavior of signal strength measurements. The observations were generated in a rectangular $60\,\text{m}$ by $80\,\text{m}$ space using identical parameters as that used in Figure 6. The observations are correlated. In this plot, the MMSE estimator incorrectly assumes that the observations are drawn from a log-normal distribution with standard deviation σ. The robust estimate consistently performs better than the MMSE estimator here.

matrix Σ_0 which forms a tight upper bound on the covariance of the worst case distribution. While the robust estimate still performs better than MMSE in this regime, the uncertainty set (4) is limited by the upper bound (4c). As the variance increases, the upper bound becomes progressively loose which results in the uncertainty set admitting more distributions, which in turn increases the performance of the estimator. This behaviour continues until it is counteracted by the fact that the increasing noise variance distorts the observations too much and the initial estimate $\boldsymbol{\mu_0}$ degrades, resulting in an increase in the error of the robust estimate.

D. Real World Experiments

We evaluated our robust formulation using real world data. The data was collected from a $4\,\text{m} \times 2\,\text{m}$ space inside an office environment. The space was divided into eight $1\,\text{m} \times 1\,\text{m}$ squares and signal strength samples were collected from the center of each square. Two hundred and fifty signal strength readings were collected for the sixteen strongest access points detected using the WiFi card on a laptop running Linux. Not all access points were in line of sight of the WiFi card, and the closest access point is about six to eight meters away. The beacon interval for each access point was approximately $100\,\text{ms}$. The signal strength measurements were taken $400\,\text{ms}$ apart. Two sets of data were collected, one at night time, when the office was empty, and the other during the day, when there was regular movement of people within the office. As shown in Figure 9, the measurements taken at night show that the observed signal strengths are highly concentrated around the mean. The measurements taken during daytime show more variability.

Thirty percent of the collected data was randomly chosen for evaluating algorithm performance. The remaining data was

used to compute the mean signal strength and its standard deviation for each access point at each measurement location. The computed mean and standard deviation were used to fit a normal distribution for the observations (in log-scale) at each measurement location, from which the MMSE estimates and the initial estimates for the robust formulation were derived. This approach to obtaining the MMSE estimates is similar to data-driven methods such as fingerprinting [4], [7], [13], where the mean signal strength at each location is used to obtain the location estimate. The parameters of the robust estimator are, $\gamma_1 = 1$ and $\gamma_2 = 1$. The performance of the robust estimator on this real dataset is shown in Figure 8. The results shown that the robust estimate outperforms the MMSE estimate in both the daytime and the nighttime datasets. The performance increase is more pronounced in the daytime, which is expected since the observations display more variability during daytime. This echos the trend seen in the simulation results.

VIII. Discussion

The robust formulation presented in this paper defines an uncertainty set of posterior distributions, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$, from which we derive the robust estimate. As illustrated in Figure 1, the uncertainty in the posterior distribution fundamentally stems from uncertainty about the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. This suggests an alternate approach. Namely, we may consider defining an uncertainty set of the distribution of observations, and then derive a robust estimate using this new uncertainty set. Note that while both of these approaches aim to quantify the uncertainty stemming from the distribution of observations, they do so in different ways. In the approach adopted hitherto, we derive a candidate posterior distribution which is used to construct an uncertainty set of posterior distributions, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$. On the other hand, in this section we explore an alternate formulation that attempts to directly specify an uncertainty set for the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. The key challenge in this approach lies in translating the uncertainty constraints to the posterior distribution.

Specifically, consider the situation where we define a set of distributions similar to the definition in equation (4). Assume that we have collected a set of observations, $\{\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_m\}$, from a particular unknown location in $\mathcal{S}$. Then empirical mean and covariance matrix of the observations may be computed for that location,

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{o}_i, \tag{18}$$

$$\hat{\Sigma} = \frac{1}{m-1}\sum_{i=1}^{m} (\boldsymbol{o}_i - \hat{\mu})(\boldsymbol{o}_i - \hat{\mu})^\top. \tag{19}$$

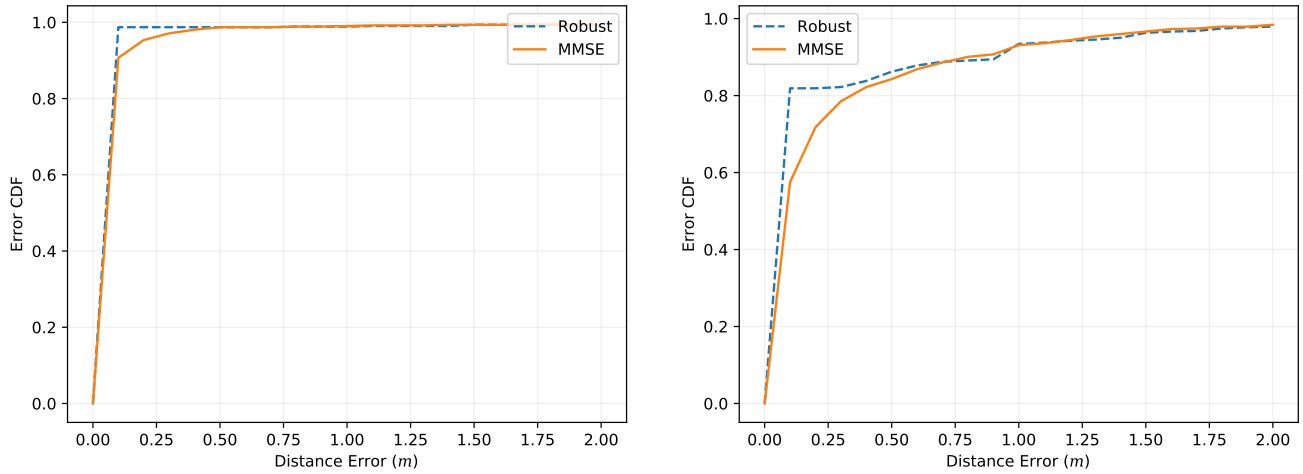
This empirical mean and covariance matrix may be used as the initial estimates to construct an uncertainty set similar to equation (4), but for the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. However, without further assumptions, it is unclear how these uncertainty constraints for $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$ can be translated to define an uncertainty set for the posterior distribution, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$. One possible way forward is to further assume that the relevant distributions have a certain functional form. Specifically, assume that the observations are generated from a certain family of distributions, such as the log-normal family. By choosing an appropriate conjugate prior, we can fix the family of the posterior distributions. The constraints on the mean and covariance matrix of the distribution of observations can be translated to constraints on the mean and covariance matrix of the posterior.

This approach has the disadvantage that we are fixing the family of distributions of the prior, and hence that of the posterior, a priori. Depending on our prior knowledge of the location, or lack thereof, such an assumption may not be appropriate. More work needs to be done to investigate the feasibility and performance of such an approach, and to compare it with the robust formulation presented in this paper. The uncertainty set construction in Section IV has the appealing feature of being conceptually close to how localization algorithms are designed and implemented currently. Current methods specify a single model for observation vector and do not impose any restriction on the choice of the prior. These features are retained in the current robust formulation.

As we have seen in Section II, there are many works in the localization literature that have attempted to account for uncertainty in various parameters of the localization system. Consider the works that account for uncertainty in location of the transmitters or anchor nodes [35]–[40]. The distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$, at any location depends on the location of the transmitters. Consequently, if there is uncertainty in the location of the transmitters, then the distribution of observations is also uncertain. As we have seen already, we can use the formulation presented here in this case to obtain a robust estimate. Another way of accounting for this uncertainty is to model the uncertain parameter, in this case the transmitter location, as a random vector following some known distribution. This is the approach adopted by almost all of the works dealing with uncertain transmitter locations in localization systems [35]–[40]. This is a highly common and valid approach to modelling uncertainty that is used beyond transmitter location uncertainty [69] or even indoor localization [70].

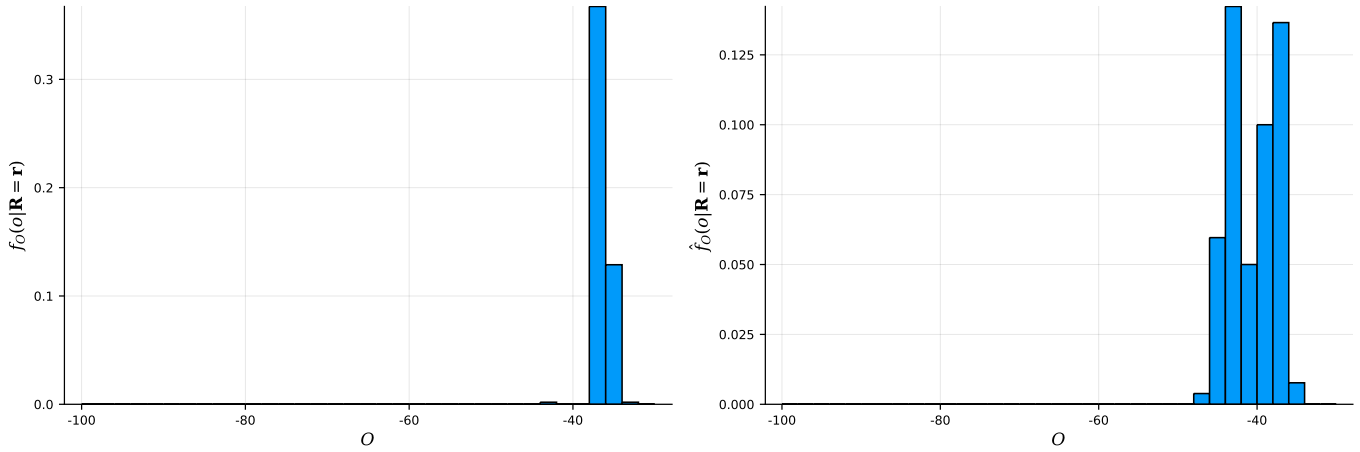
The key difference between the aforementioned works and the formulation presented in this paper is that, while these works aim to appropriately *model* the uncertainty, in this paper we aim to be *robust* to the uncertainty. These aims are complementary. Explicitly designing for robustness can help reduce the complexity of the model, while a well thought out model aids in designing the uncertainty set for robustness. Moreover, the robustness formulation presented here is general and is applicable to any model.

To illustrate the difference between these goals, let us consider an example, namely [35]. In this work, the authors consider the case when all the transmitter locations are modelled by a Gaussian distribution with known mean and variance, and the observations are assumed to follow a normal distribution with a known mean value and a random variance. Specifically, the inverse variance of the observations is assumed follow a known Wishart distribution. Under these assumptions, the authors derive the posterior distribution of the receiver and describe an algorithm to navigate the complexities of the posterior distribution to obtain the MMSE or the maximum a posteriori (MAP) estimate. This approach is correct, up



(a) (b)

Fig. 8. Performance of the MMSE and Robust estimates on a real dataset. In this experiment, two sets of observations were collected from a a rectangular $4\,\mathrm{m}$ by $2\,\mathrm{m}$ space in an indoor office environment. Plot 8a shows the performance of the robust estimate on observations taken during the night when there was little movement in the environment. Plot 8b shows the performance of the robust estimate on observations taken during the daytime, when there were regular movement of people inside the office. The parameters of the robust estimator was unchanged between both cases. We see that the robust estimate performs better than MMSE both cases. The increase in performance improves during daytime.



(a) (b)

Fig. 9. Illustration of the empirically estimated distribution of O using signal strength measurements (in dBm) taken at different locations. Each subplot refers to the distribution of signal strength from a unique access point. The left subplot refer to measurements taken at night, while the right subplot refer to measurements taken during daytime.

to a modelling error. For instance, it is unclear how the performance would change if the observations followed a multimodal distribution instead of a Gaussian. It is precisely this error in modelling that we tackle using the robust formulation presented here. By specifying the uncertainty set as given in (4), we weaken our dependence on the derived posterior distribution. Note that the dependence on a model is not completely absent. Specifically, we rely only on the first and second moments of the resulting posterior distribution, while the guarantees presented in [35] and similar works [36]–[40] depends on the correctness of the full posterior distribution. Consequently, the robust formulation presented here complements these modelling efforts, aiming to cushion the impact of errors in modelling.

In Section VI, we discussed how to select the parameters, γ_1 and γ_2, such that the resulting optimization problem always returns an optimal solution. Ideally, we would like to derive our parameters from the received observations in a manner that yields a probabilistic guarantee that our solution is robust with respect to the true posterior distribution. For instance, if we construct the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$, empirically from a large data set, then we are reasonably justified in choosing γ_1 close to 0, and γ_2 close to 1. However, further work is needed to fully characterize this dependence. To this end, a first step is to obtain a confidence region for the mean and the covariance matrix of $\boldsymbol{R}$, based solely on the received set of observations. Finally, we aim to choose our initial estimates $\boldsymbol{\mu_0}$ and Σ_0, and the parameters γ_1 and γ_2, using the received set of observations in a manner that will guarantee that the true posterior distribution lies within the

uncertainty set $\mathcal{F}(\mathcal{S}, \boldsymbol{\mu_0}, \Sigma_0, \gamma_1, \gamma_2)$, with high probability.

In [17], under a different problem setting, the authors specify the how to obtain the parameters of the uncertainty set from historical data, such that the true distribution lies within the constructed set with high probability. Specifically, the authors consider problems of the form

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad \max_{f\in\mathcal{F}} \mathbb{E}_f[h(\boldsymbol{x}, \boldsymbol{\xi})] \tag{20}$$

where $\mathcal{X}$ is a convex set of feasible solutions, $\boldsymbol{\xi}$ is some random vector of parameters, $h(\boldsymbol{x}, \boldsymbol{\xi})$ is a cost function that is convex in $\boldsymbol{x}$ and concave in $\boldsymbol{\xi}$, and $\mathcal{F}$ is the set of distributions of $\boldsymbol{\xi}$. In this setting, it is assumed that one has access to an independent set of samples $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_m\}$ generated according to an unknown distribution. This set of samples is used to construct the set $\mathcal{F}$ such that it contains the unknown distribution of $\boldsymbol{\xi}$ with high probability.

The current work differs from the above primarily in that we construct the uncertainty set $\mathcal{F}$ for the posterior distribution, $f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})$, and not the distribution of observations, $f_{\boldsymbol{O}|\boldsymbol{R}}(\boldsymbol{o}|\boldsymbol{r})$. Consequently, as illustrated in Figure 1, our estimates of the mean and covariance matrix involved in the construction of $\mathcal{F}$ are not the empirical mean and covariance matrix of the received set of observations, but that of $\boldsymbol{R}$ derived from the observations with the help of a model. The problem of obtaining performance guarantees for the robust estimate under this setting remains an area of future work.

IX. Conclusion

We have introduced a distributionally robust approach to the problem of indoor localization based on RSS observations, that explicitly takes into account the inherent uncertainty in the distribution of observations. We have identified the structure of the robust solution and illustrated how the solution changes on varying the parameters. We have demonstrated how to construct the problem such that it is easily computed using standard software tools and always returns an optimal solution. We have evaluated our robust solution under realistic channel fading models. Our results show that the robust solution outperforms the traditional approach in the presence of modeling errors, while remaining close to the traditional estimate when the modeling is exact. Our results suggests that the robust formulation presented here is well suited for settings where the environment is highly variable, such as an office building or an indoor mall. To the best of our knowledge, this is the first work that addresses distributionally robust indoor localization.

Appendix A
Deriving the Dual

Note that the matrices A_i and B_i are symmetric for all $i \in \{1, 2, \ldots, n\}$. The Lagrangian associated with the primal problem (11) is given by

$$\begin{aligned} L(\boldsymbol{p}, Z_1, Z_2, \nu, \boldsymbol{\lambda}) = &\sum_{i=1}^{n} p_i \operatorname{tr}(A_i Z_1) - \gamma_2 \operatorname{tr}(\Sigma_0 Z_1) \\ &- \sum_{i=1}^{n} p_i \operatorname{tr}(B_i Z_2) + \sum_{i=1}^{n} \nu p_i \\ &- \sum_{i=1}^{n} p_i (\boldsymbol{g_r})_i - \sum_{i=1}^{n} \lambda_i p_i - \nu. \end{aligned} \tag{21}$$

Grouping together the terms with the variable coefficients yields

$$\begin{aligned} L(\boldsymbol{p}, Z_1, Z_2, \nu, \boldsymbol{\lambda}) = &\sum_{i=1}^{n} p_i \left[\operatorname{tr}(A_i Z_1) - \operatorname{tr}(B_i Z_2) + \nu - \lambda_i \right. \\ &\left. - (\boldsymbol{g_r})_i\right] - \gamma_2 \operatorname{tr}(\Sigma_0 Z_1) - \nu. \end{aligned} \tag{22}$$

The dual function is given by

$$h(Z_1, Z_2, \nu, \boldsymbol{\lambda}) = \inf_{\boldsymbol{p}} L(\boldsymbol{p}, Z_1, Z_2, \nu, \boldsymbol{\lambda}). \tag{23}$$

Consequently we need

$$\operatorname{tr}(A_i Z_1) - \operatorname{tr}(B_i Z_2) + \nu - \lambda_i - (\boldsymbol{g_r})_i = 0, \tag{24}$$

for all $i \in \{1, 2, \ldots, n\}$, to ensure that the dual function lies above $-\infty$. In addition, the dual variables Z_1, Z_2 are symmetric, and $\lambda_i \geq 0$ for all $i \in \{1, 2, \ldots, n\}$. Thus the dual program can be expressed as

$$\begin{aligned} \text{minimize} \quad & \alpha && \text{(25a)} \\ \text{subject to} \quad & Z_1, Z_2 \succeq 0, && \text{(25b)} \\ & \gamma_2 \operatorname{tr}(\Sigma_0 Z_1) + \nu \leq \alpha, && \text{(25c)} \\ & \operatorname{tr}(A_i Z_1) - \operatorname{tr}(B_i Z_2) + \nu - (\boldsymbol{g_r})_i \geq 0, && \text{(25d)} \end{aligned}$$

for all $i \in \{1, 2, \ldots, n\}$.

Appendix B
Proof of Theorem 1

Let $\boldsymbol{r}^* = \arg\max_{\boldsymbol{r}\in\mathcal{S}} g(\|\boldsymbol{r}' - \boldsymbol{r}\|)$. Since $\mathcal{S}$ is closed and bounded, $\boldsymbol{r}^*$ exists. Assume that no point on the boundary attains the maximum cost $c^* = g(\|\boldsymbol{r}^* - \boldsymbol{r}'\|)$. Fix a value of $\boldsymbol{r}^*$ that lies in the interior of $\mathcal{S}$ and attains c^*. Consider the ray $\boldsymbol{z}(\lambda) = \boldsymbol{r}' + \lambda(\boldsymbol{r}^* - \boldsymbol{r}')$ where $\lambda \geq 0$. Since $\mathcal{S}$ is convex, $\boldsymbol{z}(\lambda) \in \mathcal{S}$ for all $\lambda \in [0, 1]$. Since $\boldsymbol{r}^*$ lies in the interior of $\mathcal{S}$ there exists a $\lambda' > 1$ such that $\boldsymbol{z}(\lambda')$ lies in the interior of $\mathcal{S}$. Since $\mathcal{S}$ is closed and bounded, the ray must intersect the boundary for some value of λ greater than λ'. In other words, we can find a $\hat{\lambda} > \lambda' > 1$ such that the ray $\boldsymbol{z}(\lambda)$ intersects the boundary $\partial\mathcal{S}$ at $\hat{\boldsymbol{r}} = \boldsymbol{z}(\hat{\lambda}) \in \partial\mathcal{S}$. Then,

$$\|\hat{\boldsymbol{r}} - \boldsymbol{r}'\| = \hat{\lambda}\|\boldsymbol{r}^* - \boldsymbol{r}'\| > \|\boldsymbol{r}^* - \boldsymbol{r}'\|. \tag{26}$$

Since the cost function g is non-decreasing,

$$g(\|\hat{\boldsymbol{r}} - \boldsymbol{r}'\|) \geq g(\|\boldsymbol{r}^* - \boldsymbol{r}'\|) = c^*. \tag{27}$$

Since c^* is the maximum cost by definition, we must have $g(\|\hat{\boldsymbol{r}} - \boldsymbol{r}'\|) = c^*$, which contradicts our initial assumption that no point on the boundary attains the maximum cost c^*. Thus there exists $\boldsymbol{r}^* \in \partial\mathcal{S}$ such that $\boldsymbol{r}^* = \arg\max_{\boldsymbol{r}\in\mathcal{S}} g(\|\boldsymbol{r}' - \boldsymbol{r}\|)$.

For any posterior distribution $f \in \mathcal{F}$,

$$\mathbb{E}_f\left[g\left(\|\boldsymbol{r}' - \boldsymbol{R}\|\right)\right] = \int\limits_{\boldsymbol{r}\in\mathcal{S}} g\left(\|\boldsymbol{r}' - \boldsymbol{r}\|\right) f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})\,\mathrm{d}\boldsymbol{r} \quad (28)$$

$$\leq g\left(\|\boldsymbol{r}' - \boldsymbol{r}^*\|\right) \int\limits_{\boldsymbol{r}\in\mathcal{S}} f_{\boldsymbol{R}|\boldsymbol{O}}(\boldsymbol{r}|\boldsymbol{o})\,\mathrm{d}\boldsymbol{r} \quad (29)$$

$$= g\left(\|\boldsymbol{r}' - \boldsymbol{r}^*\|\right). \quad (30)$$

□

APPENDIX C
PROOF OF THEOREM 2

It follows from Theorem 1 that the upper bound is attained by a point $\boldsymbol{r}^* \in \partial\hat{\mathcal{S}}$ such that $\boldsymbol{r}^* = \arg\max_{\boldsymbol{r}\in\hat{\mathcal{S}}} g\left(\|\boldsymbol{r}' - \boldsymbol{r}\|\right)$. Since the function g is non-decreasing, $\boldsymbol{r}^* = \arg\max_{\boldsymbol{r}\in\hat{\mathcal{S}}} \|\boldsymbol{r}' - \boldsymbol{r}\|$. Hence, it remains to be shown that the maximum of $\|\boldsymbol{r}' - \boldsymbol{r}\|$ over $\boldsymbol{r} \in \hat{\mathcal{S}}$ is attained by a vertex of $\hat{\mathcal{S}}$.

Let $\hat{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be the set of vertices of $\hat{\mathcal{S}}$. Then $\hat{\mathcal{S}}$ may be represented as the convex hull of its vertices, $\mathrm{conv}(\hat{V}) = \hat{\mathcal{S}}$. In other words, for any $\boldsymbol{r} \in \hat{\mathcal{S}}$ there exists $\{\lambda_i\}_{i=1}^n$ such that $\boldsymbol{r} = \sum_{i=1}^n \lambda_i \boldsymbol{v}_i$, where $\lambda_i \in [0,1]$ for all $i \in \{1, 2, \ldots, n\}$ and $\sum_{i=1}^n \lambda_i = 1$. Fix some $\boldsymbol{r}' \in \hat{\mathcal{S}}$. Then for any $\boldsymbol{r} \in \hat{\mathcal{S}}$, we have

$$\|\boldsymbol{r}' - \boldsymbol{r}\| = \left\|\boldsymbol{r}' - \sum_{i=1}^n \lambda_i \boldsymbol{v}_i\right\| \quad (31a)$$

$$= \left\|\sum_{i=1}^n \lambda_i \left(\boldsymbol{r}' - \boldsymbol{v}_i\right)\right\| \quad (31b)$$

$$\leq \sum_{i=1}^n \lambda_i \left\|\boldsymbol{r}' - \boldsymbol{v}_i\right\| \quad (31c)$$

$$\leq \left\|\boldsymbol{r}' - \boldsymbol{v}_j\right\|, \quad (31d)$$

where (31c) follows from the Cauchy-Schwarz inequality, and $\boldsymbol{v}_j = \arg\max_{1\leq i\leq n} \|\boldsymbol{r}' - \boldsymbol{v}_i\|$. Clearly the upper bound in (31d) is attained by setting $\lambda_j = 1$ and $\lambda_i = 0$ for all $i \neq j$, $i \in \{1, \ldots, n\}$. In other words, the maximum of $\|\boldsymbol{r}' - \boldsymbol{r}\|$ over $\boldsymbol{r} \in \hat{\mathcal{S}}$ is attained by setting $\boldsymbol{r}$ to be an appropriate vertex of $\hat{\mathcal{S}}$. □

REFERENCES

[1] R. Zekavat and R. Buehrer, *Handbook of Position Location: Theory, Practice and Advances*, ser. IEEE Series on Digital & Mobile Communication. Wiley, 2011.

[2] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, Nov 2007.

[3] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 25:1–25:31, Jun. 2016.

[4] 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.

[5] 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.

[6] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proc. ACM MobiCom'00*, 2000, pp. 32–43.

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

[8] 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.

[9] 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.

[10] 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.

[11] N. A. Jagadeesan and B. Krishnamachari, "A Unifying Bayesian Optimization Framework for Radio Frequency Localization," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 1, pp. 135–145, Mar. 2018.

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

[13] 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.

[14] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. John Wiley & Sons, 2011.

[15] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Math. Oper. Res.*, vol. 23, no. 4, pp. 769–805, Nov. 1998.

[16] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," *Mathematical Programming*, pp. 1–58, 2017.

[17] E. Delage and Y. Ye, "Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, Jan. 2010.

[18] W. Ye and F. Ordonez, "Robust optimization models for energy-limited wireless sensor networks under distance uncertainty," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2161–2169, June 2008.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] L. Lazos and R. Poovendran, "Hirloc: high-resolution robust localization for wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 233–246, Feb 2006.

[26] L. Lazos, R. Poovendran, and S. Capkun, "Rope: robust position estimation in wireless sensor networks," in *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.*, April 2005, pp. 324–331.

[27] S. Capkun and J. P. Hubaux, "Secure positioning of wireless devices with application to sensor networks," in *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 3, March 2005, pp. 1917–1928 vol. 3.

[28] L. Lazos and R. Poovendran, "Serloc: Secure range-independent localization for wireless sensor networks," in *Proceedings of the 3rd ACM Workshop on Wireless Security*, ser. WiSe '04, 2004, pp. 21–30.

[29] Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.*, April 2005, pp. 91–98.

[30] D. Liu, P. Ning, and W. K. Du, "Attack-resistant location estimation in sensor networks," in *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.*, April 2005, pp. 99–106.

[31] B. Krishnamachari and K. Yedavalli, "Secure sequence-based localization for wireless networks," in *Secure Localization and Time Synchronization for Wireless Sensor and Ad Hoc Networks*, ser. Advances in Information Security, 2007, no. 30, pp. 237–247.

[32] R. Casas, D. Cuartielles, A. Marco, H. J. Gracia, and J. L. Falco, "Hidden issues in deploying an indoor location system," *IEEE Pervasive Computing*, vol. 6, no. 2, pp. 62–69, April 2007.

[33] P. Li and X. Ma, "Robust acoustic source localization with tdoa based ransac algorithm," in *Emerging Intelligent Computing Technology and Applications: 5th International Conference on Intelligent Computing, ICIC 2009, Ulsan, South Korea, September 16-19, 2009. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 222–227.

[34] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[35] B. Zhou, Q. Chen, H. Wymeersch, P. Xiao, and L. Zhao, "Variational inference-based positioning with nondeterministic measurement accuracies and reference location errors," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2955–2969, Oct 2017.

[36] 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.

[37] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations–part i: Far-field sources," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 3, pp. 286–299, March 1987.

[38] M. Vemula, M. F. Bugallo, and P. M. Djurić, "Sensor self-localization with beacon position uncertainty," *Signal Processing*, vol. 89, no. 6, pp. 1144 – 1154, 2009.

[39] P. Zhang and Q. Wang, "Anchor selection with anchor location uncertainty in wireless sensor network localization," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 4172–4175.

[40] K. W. K. Lui, W. Ma, H. C. So, and F. K. W. Chan, "Semi-definite programming algorithms for sensor network node localization with uncertainties in anchor positions and/or propagation speed," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 752–763, Feb 2009.

[41] S. Yang, P. Dessai, M. Verma, and M. Gerla, "Freeloc: Calibration-free crowdsourced indoor localization," in *2013 Proceedings IEEE INFOCOM*, April 2013, pp. 2481–2489.

[42] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 293–304.

[43] A. M. Hossain and W.-S. Soh, "A survey of calibration-free indoor positioning systems," *Computer Communications*, vol. 66, pp. 1 – 13, 2015.

[44] H. Lim, L. . Kung, J. C. Hou, and H. Luo, "Zero-configuration, robust indoor localization: Theory and experimentation," in *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, April 2006, pp. 1–12.

[45] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: Unsupervised indoor localization," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '12. New York, NY, USA: ACM, 2012, pp. 197–210.

[46] C. Di Franco, A. Prorok, N. Atanasov, B. Kempke, P. Dutta, V. Kumar, and G. J. Pappas, "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*, ser. IPSN '17. New York, NY, USA: ACM, 2017, pp. 235–246.

[47] 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.

[48] 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.

[49] L. M. Ni, D. Zhang, and M. R. Souryal, "Rfid-based localization and tracking technologies," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 45–51, April 2011.

[50] T. Sathyan, D. Humphrey, and M. Hedley, "Wasp: A system and algorithms for accurate radio localization using low-cost hardware," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 2, pp. 211–222, March 2011.

[51] B. Kempke, P. Pannuto, B. Campbell, and P. Dutta, "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*, ser. SenSys '16. New York, NY, USA: ACM, 2016, pp. 137–149.

[52] P. Pannuto, B. Kempke, L.-X. Chuo, D. Blaauw, and P. Dutta, "Harmonium: Ultra wideband pulse generation with bandstitched recovery for fast, accurate, and robust indoor localization," *ACM Trans. Sen. Netw.*, vol. 14, no. 2, pp. 11:1–11:29, Jun. 2018.

[53] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 142–157, Nov 2017.

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

[55] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Operations Research*, vol. 62, no. 6, pp. 1358–1376, 2014.

[56] H. E. Scarf, *A min-max solution of an inventory problem*. Santa Monica, Calif.: Rand Corp., 1957.

[57] E. M. Bronstein, "Approximation of convex sets by polytopes," *Journal of Mathematical Sciences*, vol. 153, no. 6, pp. 727–762, 2008.

[58] R. Rockafellar, *Convex Analysis*, ser. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 2015.

[59] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[60] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.

[61] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[62] M. Udell, K. Mohan, D. Zeng, J. Hong, S. Diamond, and S. Boyd, "Convex optimization in Julia," *SC14 Workshop on High Performance Technical Computing in Dynamic Languages*, 2014.

[63] M. d. Berg, O. Cheong, M. v. Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, 3rd ed. Santa Clara, CA, USA: Springer-Verlag TELOS, 2008.

[64] F. P. Preparata and S. J. Hong, "Convex hulls of finite sets of points in two and three dimensions," *Commun. ACM*, vol. 20, no. 2, pp. 87–93, Feb. 1977.

[65] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[66] A. Ben-Tal and A. S. Nemirovskiaei, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2001.

[67] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, Jan. 2008.

[68] K. Kaemarungsi and P. Krishnamurthy, "Analysis of WLAN's received signal strength indication for indoor location fingerprinting," *Pervasive and Mobile Computing*, vol. 8, no. 2, pp. 292 – 316, 2012, special Issue: Wide-Scale Vehicular Sensor Networks and Mobile Sensing.

[69] A. Coluccia and F. Ricciato, "Rss-based localization via bayesian ranging and iterative least squares positioning," *IEEE Communications Letters*, vol. 18, no. 5, pp. 873–876, May 2014.

[70] M. Fröhle, T. Charalambous, I. Nevat, and H. Wymeersch, "Channel prediction with location uncertainty for ad hoc networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 2, pp. 349–361, June 2018.