ON SINGULAR ESTIMATION PROBLEMS IN SENSOR LOCALIZATION SYSTEMS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Joshua N. Ash, B.S.E.E., B.S. Physics, M.S.E.E.

* * * * *

The Ohio State University

2007

Dissertation Committee:

Randolph L. Moses, Adviser

Lee C. Potter

Jose B. Cruz

Alfred O. Hero III

Approved by

Adviser

Graduate Program in
Electrical and Computer
Engineering

ABSTRACT

Distributed sensor networks are growing in popularity for a large number of sensing applications ranging from environmental monitoring to military target classification and tracking. However, knowledge of the individual sensor positions is a prerequisite to obtaining meaningful information from measurements made by the sensors. With the scale of sensor networks rapidly increasing due to advances in communications and MEMS technology, an automatic localization service based on inter-sensor measurements is becoming an essential element in modern networks. This dissertation studies fundamental aspects of localization performance while deriving general results for singular estimation problems.

Because inter-sensor measurements, such as distances or angles-of-arrival (AOA), are invariant to absolute positioning of the sensor scene, localizing sensors with an absolute reference, e.g., latitude and longitude, is inherently a singular estimation problem suffering from non-identifiability of the absolute location parameters. This results in a corresponding singular Fisher information matrix.

We consider performance characterizations of self-localization algorithms and show that the location parameters have a natural decomposition into relative configuration components along with centroid transformation components that give rise to the singularity of the estimation problem. A linear representation of the non-identifiable transformation manifold, which includes representations for rotation, translation,

and scaling, is used for decomposition of general localization error covariance matrices. The unified statistical framework presented—which naturally generalizes to non-localization problems—allows us to quantify and bound performance in the relative and transformation domains. The decomposition may be applied to a large class of localization algorithms and measurement types, to the posterior Cramér-Rao bound (CRB) in a Bayesian setting, or to a traditional CRB. Along with the CRB itself, this decomposition provides both geometric insight and a quantitative analysis tool for understanding how external inputs affect absolute localization performance.

The relative-transformation error decomposition also allows us to compute CRBs for relative localization algorithms which would otherwise not exist due to singular Fisher information matrices. We apply this analysis to a novel, AOA-based, relative localization algorithm that is presented in the dissertation. In this context, we show that localization may be interpreted as a subspace identification problem which is solved, in closed-form, by singular value decomposition (SVD).

Finally, we consider anchor nodes as a means to regularize the absolute localization problem and address optimal anchor selection and placement strategies for minimum mean-square localization error. We present a novel sensor placement heuristic based on minimizing principal angles between the anchor-induced constraint subspace and the non-identifiable transformation subspace. This work provides analytical justification for the frequent, but empirical, observation that perimeter-placement of anchors is desirable.

To Mom, Dad, and Katie...

ACKNOWLEDGMENTS

First and foremost, I wish to thank my advisor, Professor Randy Moses, for his guidance and support throughout my Ph.D. program. I also with to thank Professor Lee Potter for his valuable advice and collaboration. Finally, I am grateful to the Army Research Lab (ARL) and MIT Lincoln Lab for providing funding which supported my graduate education and research.

VITA

August 23, 1975	Born - Louisville, KY
1998	B.S. Electrical Engineering, Washington University, St. Louis, MO
1998	B.S. Physics, Washington University, St. Louis, MO
2000–2002	Member Technical Staff, Celox Networks, St. Louis, MO
2003	M.S. Electrical Engineering, Washington University, St. Louis, MO

PUBLICATIONS

Research Publications

J. N. Ash and R. L. Moses, "Sensor localization error decomposition: Theory and applications," in *Proc. IEEE Statistical Signal Processing Workshop*, Aug. 2007 (invited), pp. 660–664.

J. N. Ash and L. C. Potter, "Robust system multiangulation using subspace methods," in *Proc. Information Processing in Sensor Networks*, April 2007, pp. 61–68.

J. N. Ash and R. L. Moses, "Relative and absolute errors in sensor network localization," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1033–1036, April 15-20, 2007.

N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, July 2005.

J. N. Ash and R. L. Moses, "Outlier compensation in sensor network self-localization via the EM algorithm," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 749–752, March 19-23, 2005.

J. N. Ash and R. L. Moses, "Acoustic time delay estimation and sensor network self-localization: Experimental results," *Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 841–850, 2005.

J. Ash and L. Potter, "Sensor network localization via received signal strength measurements with directional antennas," *Proc.* 42^{nd} *Annual Allerton Conference on Communication, Control, and Computing*, pp. 1861–1870, Monticello, IL, Sep 2004.

J. Wilden, J. Agniel, R. L. Moses, and J. N. Ash, "Position and orientation for distributed sensors: the PODIS network," *Proceedings of the 2004 MSS Battlefield Acoustics Symposium*, Laurel, MD, Aug 2004.

J. N. Ash and R. L. Moses, "Acoustic sensor network self-localization: Experimental results," *2003 Military Sensing Symposia (MSS) Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors*, Laurel, MD, Oct 2003.

FIELDS OF STUDY

Major Field: Electrical and Computer Engineering

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Sensor Networks

1.1.1 Sensor Network Applications

Sensor networks have proven useful in a number of distributed monitoring and control applications. Examples from precision agriculture include sensors that monitor temperature, relative humidity, and air quality within the crop canopy as well as soil properties and crop conditions at the ground level. These microclimate sensings may drive optimal crop irrigation, fertilization, and disease control measures, thereby reducing environmental impacts and improving crop yield [1, 2]. Other environmental applications monitor air and water pollution for urban planning and assist enforcement of environmental regulations [3]. Sensor networks also enable habitat monitoring at scales and resolutions not possible via traditional methods. In these applications, where a human observer would impact natural animal behavior, wireless sensors are ideally suited to provide non-intrusive observation [4, 5, 6].

Sensor networks, which may have rapid and ad-hoc deployment, have also found significant utility in military applications where unmanned sensors can reduce a soldier's exposure to hostile environments and provide improved surveillance of a target

area. In this setting, some functions of sensor networks include intrusion detection in cleared areas, classification of enemy targets, tracking of enemy forces, battlefield surveillance to monitor strike effectiveness, and monitoring supply and ammunition levels among friendly forces [7, 8, 9]. Many other military and commercial applications, such as forest fire monitoring, inventory control, and structural monitoring are given in the survey papers [9, 10].

1.1.2 The Self-localization Problem

In most sensor network applications, including those previously mentioned, knowledge of the individual sensor locations is necessary in order to meaningfully interpret data obtained by the sensors. With advances in micro-electro-mechanical systems (MEMS) and wireless communications, the size of sensor networks—as measured by both the number of nodes and size of deployment area—is rapidly increasing, with some current networks exceeding 1000 nodes [11]. Due to the large scale of such networks and ad-hoc deployment methodologies, an automated self-localization mechanism is a key enabling technology for modern sensor networks. This dissertation will study the underpinnings and fundamental performance limits of such sensor localization technologies and will develop novel algorithms to perform localization.

Sensor network self-localization (also called self-calibration and sensor localization) typically utilizes a set of inter-node measurements based on distance, time-of-arrival (TOA), time-difference-of-arrival (TDOA), received signal strength (RSS), or angle-of-arrival (AOA) observations of transmitted calibration signals. As illustrated in Figure 1.1, cooperative localization systems combine inter-node measurements, collected in a measurement vector $\boldsymbol{z}$, with any available prior information in order to

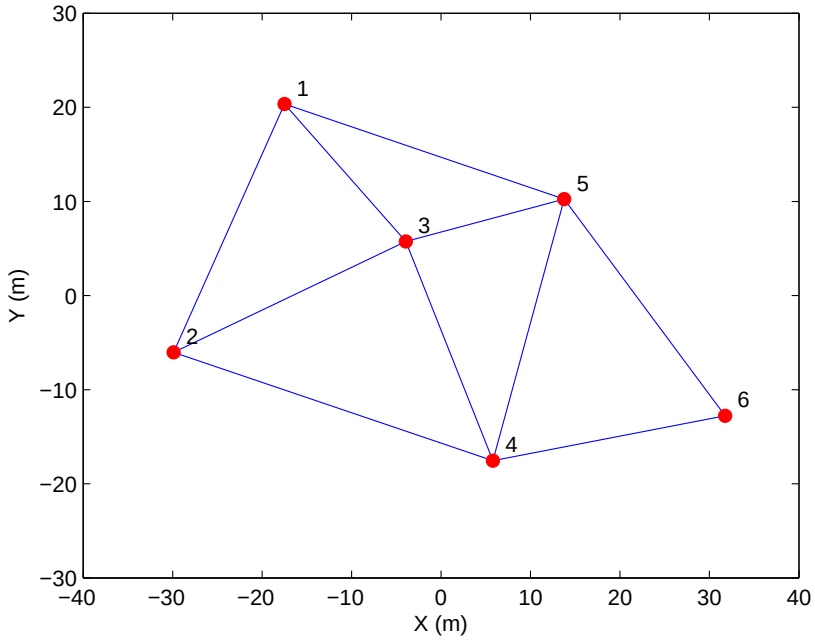


Figure 1.1: Example of a network to be localized. Edges between sensor nodes indicate availability of an inter-node measurements, such as distances or angles-of-arrival. The measurement set, which need not contain all possible pairs, is combined with prior information in order to obtain coordinate estimates $(\hat{x}_i, \hat{y}_i)$ of each node $i \in \{1, \ldots, N\}$ in the network.

obtain coordinate estimates $\{(\hat{x}_i, \hat{y}_i) : i \in (1, \ldots, N)\}$ of the N constituent nodes of the network. We refer to this problem as absolute localization, because the absolute locations of the nodes are sought. The prior information may be *a priori* knowledge of the locations of a subset of the sensors in the network, or it could be a more general constraint on the sensor positions, such as knowledge of the scene centroid. Probabilistic priors on node locations are also possible.

One common application of sensor networks is *source localization* (also called target localization), where the position of a foreign target is to be estimated. By considering the target as an additional unknown-location sensor and treating the positions of sensors making measurements of the target as known priors, we see that source localization is a specific instance of the more general self-localization problem. As such, the theoretical results obtained for self-localization apply equally to source localization.

1.1.3 Algorithms and Measurement Types

Sensor network localization algorithms may be classified in a number of ways, including:

- *Centralized vs. distributed.* In centralized processing, all of the data are transmitted to a single node, called the fusion center, which computes the sensor locations for the entire network. While this simplifies processing, it introduces a single point of failure in the system and does not scale well with network size. In distributed algorithms, the estimation task is distributed over the network.

- *Relative vs. absolute.* Relative localization algorithms only provide an estimate of a sensor network's *shape*; that is, the (x, y) locations of the sensor nodes

relative to one another but not anchored in an absolute reference frame. Absolute localization algorithms provide sensor position estimates with an absolute reference, latitude and longitude for example.

- *Statistical basis.* Many localization algorithms have no statistical basis but produce correct estimates in the noiseless case. Statistically based algorithms consider the type of measurement noise and provide a tailored estimate. Classic techniques such as maximum likelihood (ML) estimation and maximum *a posteriori* (MAP) estimation fall into this category.

- *Iterative vs. closed-form.* Computational complexity plays an important part in localization algorithms because the number of sensors may be very large and the algorithms may be implemented on resource-constrained sensors. Iterative techniques are typically employed in algorithms requiring optimization of a complex non-linear cost function. However, the high dimensionality of sensor localization can make these algorithms difficult to initialize and prone to local convergence problems. Closed-form algorithms do not suffer from these problems but rarely match the performance of iterative routines.

- *Measurement type.* Finally, most localization algorithms are specific to a particular type of measurements, such as inter-sensor distances or angles, and may be classified on that basis as well.

All localization algorithms work by converting a set of position-dependent measurements $\boldsymbol{z}$ into position estimates $\hat{\theta}$. Time-of-arrival (TOA) is one type of measurement commonly considered in localization literature and corresponds to the emission time of a signal (RF, acoustic, seismic, etc.) plus a propagation-induced time delay.

When the emission times and propagation velocity are known, TOA measurements are effectively the same as a distance measurements. When the emission times are unknown, the measurement model is dubbed time-difference-of-arrival (TDOA) because it is the differences in arrival times which bear position information.

When a complete set of inter-node distance or TOA measurements are available, classical multidimensional scaling (MDS) provides a robust subspace-based location estimate [12]. The Isomap algorithm [13] generalizes MDS to incomplete measurement sets by replacing missing measurements with shortest-path distances, while [14] considers alternative distance matrix completions. MDS and ISOMAP are general tools for describing dissimilarity measures and were first applied to sensor localization in [15, 16]. Other closed form solutions for localization using range differences and time differences are presented in [17]. Iterative methods based on maximum likelihood estimates were derived in [18, 19], while other time- and distance-based localization algorithms are considered in [20, 21, 22].

Received signal strength (RSS) is an alternative measurement type that is attractive for inexpensive sensors because it does not require time synchronization between sensors. When the modality is RF, measurements can be made in the course of normal communication activities, which decreases the energy requirements of localization. ML estimation of sensor positions from RSS was considered in [19, 23] and the effect of random unknown transmit power in [24]. Estimation by spherical intersection, based on energy ratios, was considered in [25]. Additional RSS-based methods are described in [26, 27]. Because RSS measurements are prone to large noise, some researchers have considered localization based on communication connectivity [28, 29]—essentially exploiting the idea that sensors within communication

range must be in the same geographic proximity. Connectivity is effectively a one-bit quantization of RSS; this idea has been generalized to n-bit quantization in [27].

Angle-of-arrival (AOA) measurements may also be used for self-localization in a process that is a slight generalization of triangulation. When the orientations of the sensors making AOA measurements are unknown, the measurement type is known as angle-difference-of-arrival (ADOA) because—analogously to TDOA—the location-bearing information is in the difference of arrival angles. A distributed AOA-based localization algorithm is presented in [30] where sensors first estimate their bearing to known-location beacons and then triangulate themselves. A centralized approach was taken in [28] utilizing semidefinite programming to estimate sensor locations from the intersection of AOA-derived constraint sets. Simultaneous maximum-likelihood estimation of all sensor positions from AOA measurements was considered in [31].

A summary of popular measurement models is given in Table 1.1, and additional localization algorithms may be found in the survey papers [32, 33].

Measurement type	Observation model	Typical measurement units	Unknown parameters
distance	$z_{t,r} = \|\|p_t - p_r\|\|_2 + n_{t,r}$	meters	$\{p_t\}$
TOA	$z_{t,r} = \tau_t + \|\|p_t - p_r\|\|_2/c + n_{t,r}$	milliseconds	$\{p_t\}$
TDOA	$z_{t,r} = \tau_t + \|\|p_t - p_r\|\|_2/c + n_{t,r}$	milliseconds	$\{p_t\}, \{\tau_t\}$
AOA	$z_{t,r} = \angle(p_t, p_r) - \gamma_r + n_{t,r}$	degrees	$\{p_t\}$
ADOA	$z_{t,r} = \angle(p_t, p_r) - \gamma_r + n_{t,r}$	degrees	$\{p_t\}, \{\gamma_r\}$
RSS	$z_{t,r} = P_t - 10\alpha \log_{10} \frac{\|\|p_t - p_r\|\|_2}{d_0} + n_{t,r}$	dBm	$\{p_t\}$

Table 1.1: Common measurement models used in localization. Notation: $p_i = (x_i, y_i)$ denotes the position of sensor i, τ_t is the emission time of a calibration signal emanating from sensor t, γ_r is the orientation of sensor r, c is signal propagation velocity, P_t is signal transmit power, α is a propagation pathloss exponent, and $n_{t,r}$ is measurement noise.

1.1.4 Cramér-Rao Bounds in Localization

The quality of a self-localization solution depends on a number of elements, including the type of measurements used (AOA, TDOA, etc.), the measurement noise distribution, the geometry of the true sensor positions, the connectivity of the measurement graph, prior information on sensor locations, and the location-estimation algorithm itself. By interpreting sensor localization as a parameter estimation problem, we may use Cramér-Rao bounds (CRBs) to evaluate localization performance bounds in an estimator-independent way. This provides a benchmark for localization algorithms and allows us to explore the sensitivity of localization solutions to various network characteristics, such as the noise level and measurement connectivity.

The CRB formalism also allows us to evaluate the utility of the measurement types themselves with respect to one another. Clearly, for a given measurement type, lower noise results in improved location estimates. However, the comparison is less straightforward across measurement types. For example, given the alternatives of an acoustic TOA system that can measure arrival times with a standard deviation $\sigma_t = 1\,\text{ms}$, or a radio-frequency based AOA system with angular measurement errors of $\sigma_\theta = 3^\circ$, it is not obvious which system provides better measurements for self-localization.

Under the assumption of Gaussian noise, we have derived Fisher's information and the corresponding CRB for all of the measurement models given in Table 1.1. An example application is illustrated in Figure 1.2 where we examine the performance of TDOA versus TOA. The only difference between these two measurement types is whether the estimator has knowledge (TOA) of the emission times $\{\tau_t\}$ or does not (TDOA). In practice, it can be difficult for a node to precisely estimate the time of its

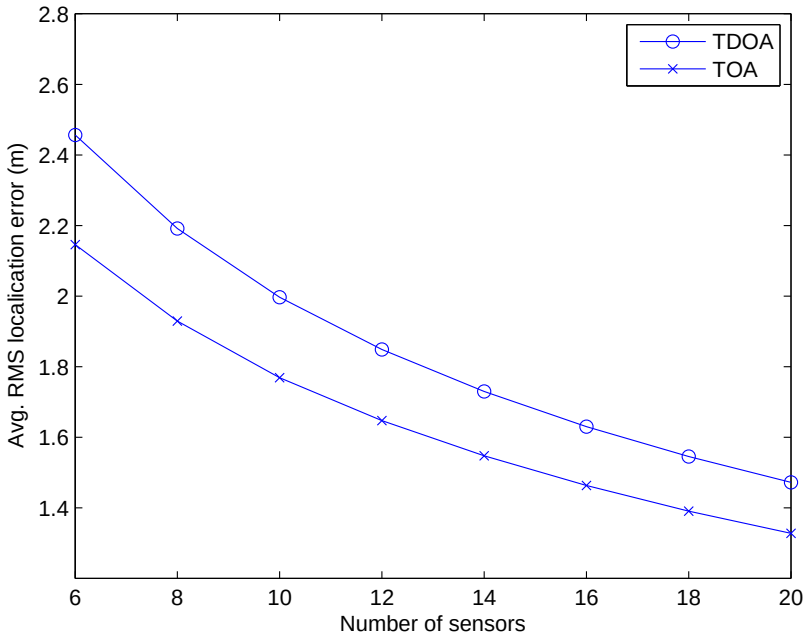


Figure 1.2: Cramér-Rao bound analysis illustrating average localization improvement of TOA over TDOA versus the number of deployed sensors.

own emission due to various random delays through command and communication queues. TOA also requires the communication overhead of transmitting the emission times from each source to the estimation center. As such, it is useful to understand whether the effort involved in obtaining emission times is worthwhile with respect to localization performance. In the example in Figure 1.2, which considers uniformly deployed sensors in a $100\,\mathrm{m} \times 100\,\mathrm{m}$ area with acoustic signaling and zero-mean Gaussian error with standard deviation $\sigma_t = 10\,\mathrm{ms}$, the average root-mean-square positioning error of TDOA is about 12% greater than TOA.

Cramér-Rao bound analysis also allows us to answer questions regarding the utility of simultaneously using multiple types of measurements. Clearly the information

present in two types of measurements is at least as high as that of one, and estimation performance must be improved; however, it is not clear that the performance improvement justifies the additional hardware necessary for an additional measurement type and the necessarily greater communication and estimation complexity. By considering the Fisher information present in the joint statistics of multiple measurement types, we are able to assess the utility of combining different forms of measurements. This is a way to evaluate the optimal fusion performance of such disparate forms of information.

An example is provided in Figure 1.3 where we consider the fusion of distance and angle measurements for localization. For 16 nodes randomly deployed in a $100\,\mathrm{m} \times 100\,\mathrm{m}$ area, equal-error contours are plotted as a function of the independent variables of the figure: the standard deviation of angle measurements σ_θ, and the standard deviation of available distance measurements σ_d. By considering large errors in one type of measurement, the performance of the other measurement modality alone can be inferred from the asymptotic nature of the contours. For example, without angle measurements distance measurements with $\sigma_d = 2\,\mathrm{m}$ result in RMS localization error of approximately $0.8\,\mathrm{m}$. Following this contour into the region of large distance errors and low angular errors, we see that AOA measurements with $\sigma_\theta \approx 2.8^\circ$ are required to achieve an equivalent level of localization performance.

The type of CRB analysis leading to Figure 1.3 is also useful in evaluating the utility of one type of measurement in the presence of another. For example, consider two vertical cuts in Figure 1.3 at $\sigma_\theta = 2^\circ$ and $\sigma_\theta = 15^\circ$. When $\sigma_\theta = 15^\circ$, any reduction in σ_d improves the localization estimates substantially. However, when $\sigma_\theta = 2^\circ$, the quality angle measurements dominate, and improved distance measurements have

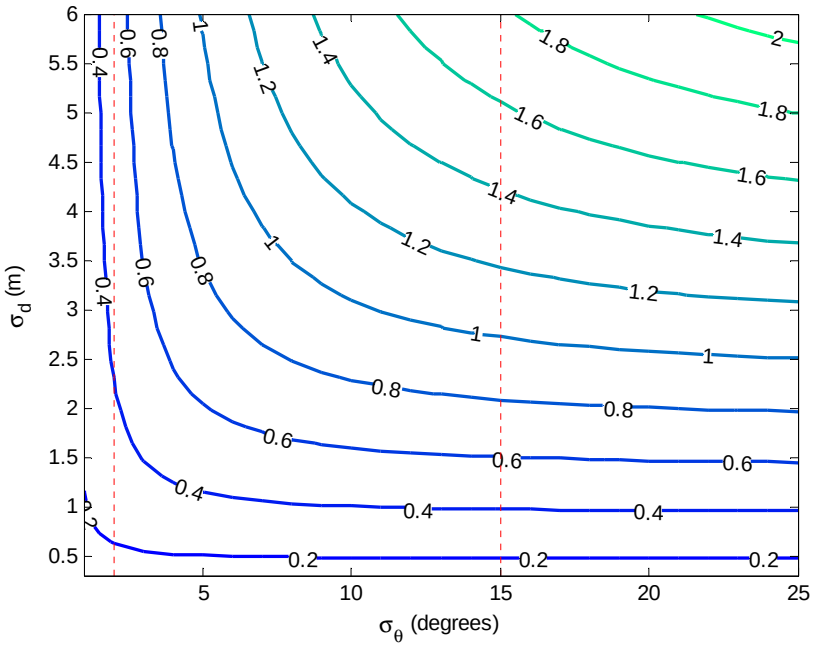


Figure 1.3: Cramér-Rao bound analysis illustrating the utility of simultaneous distance and AOA measurements. Contours of equal RMS localization error (in meters) are plotted as a function of the measurement error standard deviation of distance measurements (σ_d) and the error standard deviation of angle measurements (σ_θ).

little improvement on overall scene estimation error—until σ_d is approximately less than 2 m.

1.2 Singular Estimation Problems

Singular estimation problems are those where the parameter vector θ cannot be uniquely identified from the data. In order of increasing generality, it may be that (1) certain elements of the parameter vector θ cannot be estimated from the data, (2) certain linear combinations of parameters cannot be estimated, or (3) the post-measurement parameter estimate can only be confined to a nonlinear manifold in the complete parameter space. These problems are all characterized by having singular Fisher information matrices (FIMs) which preclude the existence of any unbiased estimator with finite error variance [34]. Singular estimation problems are found in many applications areas. Examples include digital communications, where blind symbol and channel identification are inherently singular [35, 36]; in overparameterized signal models [37, 38]; and in many learning problems with hidden variables [39]. In general, when measurements are invariant to smooth changes of the parameter vector, as in underdetermined linear systems for example, local identifiability problems arise and the estimation problem becomes singular.

Singular problems may be regularized by supplying additional information. For example, deterministic constraints may be imposed on the parameter vector θ such that the effective dimension of the problem is reduced and the resulting FIM becomes non-singular. By reinterpreting the problem in a Bayesian setting, a prior distribution on θ may be used for regularization. In this setting, the prior can be used to supply information about the components of θ not informed upon by the data. The total

Fisher information in this case [40, p. 84] is $J_T = J_D + J_P$, where J_D represents the information from the data and is equal to the non-random FIM averaged over the prior, and J_P represents *a priori* information derived solely from the prior. For suitable priors, the addition of J_P will eliminate the singularity of the total FIM.

Absolute localization of sensors based only on internode measurements, such as TDOA or AOA, is inherently a singular problem [18, 41, 42]. This is because internode measurements only depend on the relative configuration of sensors, not their absolute positions. For example, the structural variability in TDOA measurements depends only on the distances between sensors (see Table 1.1), therefore, any rigid translations and rotations of the sensor scene will produce data with identical statistics—making it impossible to infer the true translation and rotation in any absolute frame of reference.

The singularity of absolute localization will play a major theme throughout this dissertation.

1.3 Summary of Chapters and Contributions

In the second chapter we present a novel closed-form algorithm to self-localize and orient sensors from AOA and ADOA measurements. We demonstrate that the AOA localization problem has the interpretation of a subspace identification problem that can be solved with singular value decomposition (SVD). The algorithm is similar to classical multidimensional scaling (MDS) [12] and Isomap [13] in that it finds generative point configurations from partial measurement sets, except that the measurements are now angles instead of distances or general measures of dissimilarity. As singularity problems arrise in absolute positioning, we demonstrate how the

relative estimator proceeds by making arbitrary choices about the non-identifiable parameters.

We show that the computational complexity of the algorithm is $\mathcal{O}(mn^2)$, where m is the number of measurements and n is the total number of sensors. Simulation results demonstrate that the error of the proposed subspace algorithm is only marginally greater than an iterative maximum-likelihood estimator (MLE), while the computational complexity is two orders of magnitude less. Additionally, the iterative MLE is prone to converge to local maxima in the likelihood function without accurate initialization. We illustrate that the proposed subspace method can be used to initialize the MLE and obtain near–Cramér-Rao performance for sensor localization. Finally, the scalability of the subspace algorithm is illustrated by demonstrating how clusters within a large network may be individually localized and then merged.

In the third chapter, we consider the accuracy of sensor node location estimates from self-calibration in sensor networks. The total parameter space is shown to have a natural decomposition into *relative* and centroid *transformation* components. The singular nature of absolute localization results in a singular Fisher information matrix (FIM). We show that a linear representation of the transformation parameter space coincides with the nullspace of the FIM. The centroid transformation subspace—which includes representations of rotation, translation, and scaling—is characterized for a number of measurement models including distances, TOA, TDOA, AOA, and ADOA measurements. Along with the CRB itself, the relative-transformation decomposition provides insight into how external inputs effect absolute localization performance. The error decomposition may be applied to any localization algorithm in order to better understand its performance characteristics in these domains, or it

may be applied to a traditional CRB or posterior CRB in a Bayesian setting. Sensor localization bounds are also considered for arbitrary functional constraints ($f(\theta) = 0$) on the location parameters θ using a constrained CRB [43]. Geometric interpretations of the constrained CRB are provided based on the principal angles between the measurement subspace and the constraint subspace.

One benefit of this analysis is that it provides a direct bound on the performance of relative localization algorithms. Previously, arbitrary constraints were imposed to regularize the relative estimator and produce a non-singular Fisher information matrix. The drawback of this latter approach is that the error bound depends on the particular constraint employed and does not directly measure relative performance. The relative–transformation error partitioning is also useful to higher level applications in a sensor network that utilize results of the localization service and must account for its uncertainty. Examples are presented and an application demonstrates the utility of relative error decomposition to the problem of angle-of-arrival estimation with sensor location uncertainty.

In the fourth chapter, we consider using a select number of sensors with known locations, called anchor nodes, as a means of removing the singularity of absolute localization. We show that absolute localization performance depends on the locations of the anchor nodes relative to the remaining unknown-location nodes, and we consider means of optimally selecting and placing anchor nodes in order to minimize mean-square localization error. A heuristic approach is taken to minimize total estimation error by forcing constraints to be maximally informative in the singular domain of the problem. In this case, the singular domain corresponds to the space of

unknown rigid transformations and scalings of the sensors. By minimizing the principal angles between the anchor-induced constraint subspace and the non-identifiable subspace, we are able to develop anchor placement heuristics that do not depend on knowledge of the true sensor locations.

We demonstrate empirically that the subspace alignment heuristic is a strong indicator of mean-square estimation performance and that anchors placed in this way yield total localization error similar to that of oracle-placement, where perfect knowledge of the unknow-location sensors is assumed. The analysis of this chapter also provides analytical justification for the frequent empirical observation [44, 45, 46] that anchor nodes should be positioned around the perimeter of a sensor network.

The appendix contains a derivation of Fisher's information for problems with missing data. The cases where the probability of a missing measurement depends, and does not depend, on the parameter vector are considered separately. In the second case, we find an explicit expression for the amount of information in a missing measurement.

CHAPTER 2

CLOSED-FORM SYSTEM MULTIANGULATION

2.1 Introduction

Multiangulation (or simply angulation) represents a generalization of the familiar concept of triangulation and denotes the task of localizing a source point from a set of known locations and the angles from those locations to the source. In the noiseless case, two known-location points are sufficient to localize a third point—hence the prevalence of triangulation—however, larger numbers of known-location points are beneficial when the angles contain noise. In this chapter, we consider a novel technique for the *simultaneous* relative localization of all nodes in a sensor network from a set of angle-of-arrival (AOA) measurements. For this reason, we refer to the method as "system" multiangulation. No known-location points (called anchors) are required, however, if present, they may be used to subsequently provide the relative scene estimate an absolute frame of reference.

Many sensors support AOA measurements as part of their sensing or communication subsystems. For example, sensors tracking people, wildlife, and vehicles are often equipped with phased microphone arrays which allow enhanced observations in the direction of the target of interest through beamforming [47, 6]. These arrays can also be

used for AOA estimation via algorithms like MUSIC and ESPRIT [48]. Sensors may also be equipped with arrays of directional antennas in order to focus communication transmissions toward the intended destination and limit interference to neighboring sensors. With knowledge of the antenna beam patterns, a multi-antenna system with a single radio receiver may also be used for accurate AOA estimation of any in-band RF source [49]. RF phased arrays, as used in current multiple-input multiple-output (MIMO) communication systems, will likely make their way into sensor networks and could also be used for AOA estimation, albeit with increased circuit complexity.

The aim of this chapter is to develop a robust and low-complexity algorithm for network localization from AOA measurements. We illustrate that the system multi-angulation problem may be interpreted as a subspace identification problem. When a complete set of inter-node *distance* measurements are available, classical multidimensional scaling (MDS) provides a robust subspace-based location estimate [12]. The Isomap algorithm generalizes MDS to incomplete measurement sets by replacing missing measurements with shortest-path distances [13]. Because time-of-arrival (TOA) and received-signal-strength (RSS) measurements may be converted to distance estimates, MDS and Isomap apply to these measurement modalities as well. Subspace methods were used in [17] for closed-form source localization using range differences and time differences. To the best of our knowledge, subspace-based methods have not been previously applied to AOA. As such, this chapter fills a gap in existing literature by providing subspace-based localization from AOA or angle-difference-of-arrival (ADOA) measurements. The algorithm is non-iterative and provides a good trade-off between complexity and performance.

Previous work on AOA-based localization considered a distributed approach where bearing-to-anchor information is propagated throughout the network allowing nodes to triangulate themselves when they obtain enough anchor bearings [30]. The strength of this approach is that it is distributed; the weakness is that bearing error accumulates as the estimates are propagated through the network. In [28], a centralized approach was taken where semidefinite programming was employed to estimate sensor locations from the intersection of AOA-derived constraint sets. Maximum-likelihood estimates of sensor positions from AOA measurements were considered in [31].

Notation: Throughout this chapter capital letters represent matrices and bold lowercase letters denote column vectors. $(\cdot)^T$ and $(\cdot)^*$ represent matrix transpose and conjugate-transpose respectively, while $\mathcal{R}(A)$ and $\mathcal{N}(A)$ represent the range space and null space of matrix A, respectively.

2.2 System Multiangulation

The localization problem is to estimate sensor node positions $\boldsymbol{p}_i = [x_i \ y_i]^T, \ i \in \{1, ..., n\}$ from a partial set of angle-of-arrival measurements $\{\phi_{ij}\}$. In order to obtain absolute position estimates, a small subset $\mathcal{K} \subseteq \{1, ..., n\}$ of the nodes, called anchors, are assumed to have known positions and are used to resolve ambiguities in translation, rotation, and scale of a relative solution. As illustrated in Figure 2.1, each node r makes AOA measurements in a local coordinate system which is offset by an angle α_r from a global reference. In the global coordinate system, the AOA at receiving node r, of a transmission from node t is

$$\theta_{rt} = \phi_{rt} + \alpha_r, \tag{2.1}$$

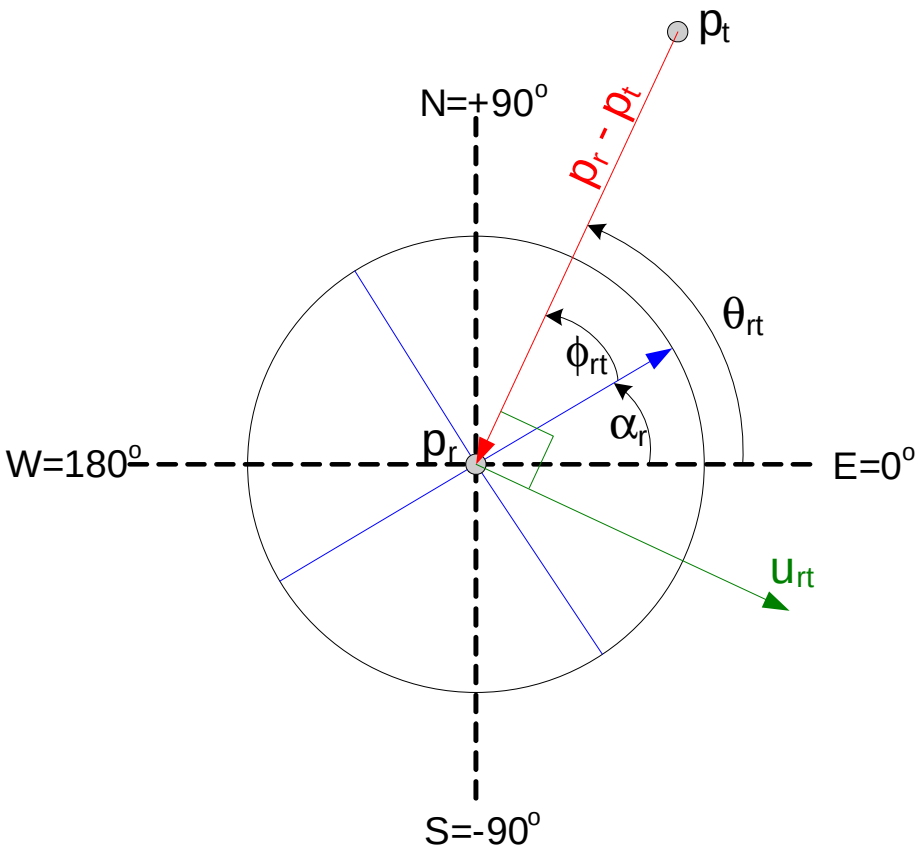


Figure 2.1: Illustration of global and local coordinate systems for AOA measurements. Each node r makes measurements in a local coordinate system which is offset by an amount α_r from the global coordinate system.

where ϕ_{rt} is the measurement in r's local system. We first consider localization when the orientation angles $\{\alpha_r\}$ are known (as provided by a digital compass on each sensor, for example) and then address the more difficult unknown orientation case.

2.2.1 Known Orientations

When the sensor orientations $\{\alpha_r\}$ are known, we can immediately map local measurements $\{\phi_{rt}\}$ into their global frame counterparts $\{\theta_{rt}\}$ via (2.1). For each angle θ_{rt}, we form a unit vector

$$\boldsymbol{u}_{rt} = \begin{bmatrix} \sin\theta_{rt} \\ -\cos\theta_{rt} \end{bmatrix} \tag{2.2}$$

which, as illustrated in Figure 2.1, is orthogonal to the difference of the position vectors

$$\boldsymbol{u}_{rt}^T(\boldsymbol{p}_t - \boldsymbol{p}_r) = 0. \tag{2.3}$$

This construction allows us to determine a system of equations which can be solved for the node positions.

Let $\mathcal{M}$ denote the set of ordered measurement pairs; that is, $(r,t) \in \mathcal{M}$ if node r makes an AOA measurement of a transmission from node t, and let $\mathcal{M}_r(i)$ and $\mathcal{M}_t(i)$ denote the receiver and transmitter of the ith element of $\mathcal{M}$, respectively. Let $U = \{U_{ij}\}$ be an $|\mathcal{M}| \times |\mathcal{M}|$ block diagonal matrix, where each block is an element of $\mathbb{R}^{2\times 1}$ and $|\mathcal{M}|$ is the total number of measurements. We populate the diagonal (block) entries of U with the previously defined unit vectors as $U_{ii} = \boldsymbol{u}_{r't'}$, where $r' = \mathcal{M}_r(i),\ t' = \mathcal{M}_t(i)$. Let $K = \{K_{ij}\}$ be an $|\mathcal{M}| \times n$ block matrix with common block size 2×2. The non-zero elements of the ith row of K are populated as $K_{it'} = I_2$ and $K_{ir'} = -I_2$, where I_2 is the 2×2 identity matrix and again $r' = \mathcal{M}_r(i),\ t' = \mathcal{M}_t(i)$.

If we let $\boldsymbol{p} = [\boldsymbol{p}_1^T\ \boldsymbol{p}_2^T\ \dots\ \boldsymbol{p}_n^T]^T$ be the stacking of the n position vectors, then from (2.3) we can form the homogeneous linear system

$$U^T K \boldsymbol{p} = 0. \tag{2.4}$$

The function of the matrix K is to form the necessary position vector differences corresponding to each AOA measurement. An example of a three-node system ($n = 3$) with four measurements ($|\mathcal{M}| = 4$) may look like

$$\underbrace{\begin{bmatrix} \boldsymbol{u}_{21} & 0 & 0 & 0 \\ 0 & \boldsymbol{u}_{31} & 0 & 0 \\ 0 & 0 & \boldsymbol{u}_{32} & 0 \\ 0 & 0 & 0 & \boldsymbol{u}_{13} \end{bmatrix}^T}_{U^T,\ |\mathcal{M}|\times 2|\mathcal{M}|} \underbrace{\begin{bmatrix} I_2 & -I_2 & 0_2 \\ I_2 & 0_2 & -I_2 \\ 0_2 & I_2 & -I_2 \\ -I_2 & 0_2 & I_2 \end{bmatrix}}_{K,\ 2|\mathcal{M}|\times 2n} \underbrace{\begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \boldsymbol{p}_3 \end{bmatrix}}_{p,\ 2n\times 1} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{|\mathcal{M}|\times 1}. \tag{2.5}$$

The relative localization solution sought is a particular solution to the linear homogeneous system (2.4) above. In general, the dimension of the nullspace of $U^T K$ is three. One dimension corresponds to scalings of the point configuration $\boldsymbol{p}$, another corresponds to translations in the x-direction, and the third corresponds to translations along the y-direction; that is $\mathcal{N}(U^T K) = \mathcal{R}([\boldsymbol{p}\ \boldsymbol{v}_x\ \boldsymbol{v}_y])$, where $\boldsymbol{v}_x = [1\ 0\ 1\ 0 \dots]^T$ and $\boldsymbol{v}_y = [0\ 1\ 0\ 1 \dots]^T$. The vectors $\boldsymbol{v}_x$ and $\boldsymbol{v}_y$ come directly from the matrix K, ($K\boldsymbol{v}_x = K\boldsymbol{v}_y = 0$) and reflect the fact that AOA measurements only depend on the relative point configuration and are invariant to scene translations. The third basis vector is, by construction, the desired point configuration $\boldsymbol{p}$ up to scale and translation. If there are an insufficient number of measurements, the dimension of $\mathcal{N}(U^T K)$ will be greater than three.

Because the translation vectors $\boldsymbol{v}_x$ and $\boldsymbol{v}_y$ do not depend on the coordinates being estimated, we can reduce the dimension of the null space by augmenting the rows of

$U^T K$ with $\boldsymbol{v}_x^T$ and $\boldsymbol{v}_y^T$. Let

$$A = \begin{bmatrix} U^T K \\ \boldsymbol{v}_x^T \\ \boldsymbol{v}_y^T \end{bmatrix} \in \mathbb{R}^{(|\mathcal{M}|+2)\times(2n)}. \tag{2.6}$$

Then, $\mathcal{N}(A) = \text{span}(\boldsymbol{p})$, and we can solve

$$A\boldsymbol{p} = 0 \tag{2.7}$$

using the singular value decomposition $A = U_A \Sigma_A V_A^T$. The (unit-norm) minimizing solution $\boldsymbol{p}$ of $\|A\boldsymbol{p}\|$ is

$$\hat{\boldsymbol{p}} = V_A^{(2n)}, \tag{2.8}$$

where $V_A^{(2n)}$ is the right singular vector corresponding to the minimum singular value of A.

The point configuration so obtained will be centered at the origin and, in the noiseless case, will be a equal to the generating configuration up to translation and scale. The final estimate of $\boldsymbol{p}$ is obtained by scaling and translating $\hat{\boldsymbol{p}}$ for maximal agreement with prior information. We use the Procrustes algorithm for this purpose as described in the next subsection.

Transforming from Relative to Absolute Point Configurations

In order to transform our relative solution obtained above into an absolute solution, we we find a suitable transformation mapping the estimated values of the anchor nodes $\mathcal{K}$ to their *a priori* known positions.

Let the $|\mathcal{K}| \times 2$ matrix $P = [\boldsymbol{p}_{\mathcal{K}(1)}, \boldsymbol{p}_{\mathcal{K}(2)}, \dots]^T$ denote the matrix of *a priori* known positions, and similarly let $\hat{P} = [\hat{\boldsymbol{p}}_{\mathcal{K}(1)}, \hat{\boldsymbol{p}}_{\mathcal{K}(2)}, \dots]^T$ form the previously obtained estimates of this subset of nodes. We use the Procrustes algorithm to find the squared-error minimizing translation, rotation, and scale of the points $\hat{P}$ into P. That is,

we seek a scale factor s, an orthogonal rotation matrix R, and a translation vector $\boldsymbol{t} = [\boldsymbol{t}_x \ \boldsymbol{t}_y]^T$ that minimizes

$$\left\| P - (s\hat{P}R + \mathbf{1}\boldsymbol{t}^T) \right\|_F, \tag{2.9}$$

where the norm is in the Frobenius sense and $\mathbf{1}$ is a vector of all ones. We then apply these transformations to all of the estimated points in order to obtain the complete scene estimate.

The solution to this problem is a slight generalization of the orthogonal Procrustes problem and is given in [50]. Let $J = (I - |\mathcal{K}|^{-1}\mathbf{1}\mathbf{1}^T)$, and compute the SVD $U_p \Sigma_p V_p^T = P^T J \hat{P}$. The desired transformation parameters are

$$R = V_p U_p^T \tag{2.10}$$

$$s = (tr\ P^T J \hat{P} R)/(tr\ \hat{P}^T J \hat{P}) \tag{2.11}$$

$$\boldsymbol{t} = |\mathcal{K}|^{-1}(P - s\hat{P}R)^T \mathbf{1}. \tag{2.12}$$

For the known orientation case, the scene rotation is not ambiguous. That is, the rotation R above will always be identity (or a reflection matrix), however, rotation will be needed below for the case of unknown orientations, so we include it in (2.10)–(2.12).

The complexity of determining rotation, translation, and scale using the Procrustes method is very low because the SVD only involves a 2×2 matrix, for any number of anchors and total sensors. The complexity of obtaining the relative estimate $\hat{\boldsymbol{p}}$ is addressed in Section 2.4 after considering the unknown orientation case.

2.2.2 Unknown Orientations

Unknown orientations $\{\alpha_r\}$ are nuisance parameters in the localization problem, however their estimates are useful quantities in many sensor applications. In our approach, we first estimate the orientations up to a common offset and then apply the point estimation procedure above for known orientations. We begin by describing the procedure for the case of complete measurements where all pairs of sensors make AOA measurements in both directions. This is then extended to the partial measurement case.

Complete Measurements

Using the fact that $\theta_{rt} = \theta_{tr} + \pi$ we have

$$\alpha_r - \alpha_t = \phi_{tr} - \phi_{rt} + \pi \qquad \text{for all } r \neq t. \tag{2.13}$$

In order to simplify the modulo 2π arithmetic involved in angular calculations, we represent the sensor orientation angles as the phase of points in the complex plane, $\boldsymbol{a}(\boldsymbol{\alpha}) = [e^{i\alpha_1},\ ...,\ e^{i\alpha_n}]^T$, where $\boldsymbol{\alpha} = [\alpha_1,\ ...,\ \alpha_n]^T$ is the vector of unknown orientations. Let $\Phi = \{\phi_{rt}\}$ denote the matrix of AOA measurements, from which we form

$$\Psi = \Phi^T - \Phi + \pi \mathbf{1}_n \mathbf{1}_n^T, \quad \text{and} \tag{2.14}$$

$$B = e^{i\Psi} + 2\boldsymbol{I}, \tag{2.15}$$

where $e^{i\Psi}$ denotes elementwise exponentiation of Ψ, not the conventional matrix exponential.

We now have the fundamental relation

$$\boldsymbol{a}(\boldsymbol{\alpha})\ \boldsymbol{a}(\boldsymbol{\alpha})^* = B \tag{2.16}$$

which holds in the noiseless case. Let $V_B \Lambda_B V_B^*$ denote the eigen-decomposition of B. The closest rank-one approximation (in the Frobenius norm sense) to B is $\lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^*$, where λ_n is the largest eigenvalue of B and $\boldsymbol{v}_n$ is the associated eigenvector. We may then estimate $\boldsymbol{a}(\boldsymbol{\alpha})$ as

$$\hat{\boldsymbol{a}} = \boldsymbol{v}_n \tag{2.17}$$

and compute the orientation estimates as the phase of each complex element of $\hat{\boldsymbol{a}}$

$$\hat{\alpha}_i = \angle \hat{a}_i. \tag{2.18}$$

In the noiseless case, the orientation angles so obtained will equal the true angles up to a constant, that is $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}} + \alpha_0 \mathbf{1}$. The ambiguity of α_0 is due to the fact that the elements of B only depend on the differences between elements of $\boldsymbol{\alpha}$ as seen in (2.13) and (2.15). The scalar α_0 represents a common orientation offset which results in an overall unknown rotation of the relative scene estimate. This last fact is seen by noting that if $\boldsymbol{p}_r$ and $\boldsymbol{p}_t$ satisfy (2.3), then $\tilde{p}_r = R_{\alpha_0} \boldsymbol{p}_r$ and $\tilde{p}_t = R_{\alpha_0} \boldsymbol{p}_t$ satisfy

$$\begin{bmatrix} \sin(\theta_{rt} + \alpha_0) \\ -\cos(\theta_{rt} + \alpha_0) \end{bmatrix}^T (\tilde{p}_r - \tilde{p}_t) = 0, \tag{2.19}$$

where

$$R_{\alpha_0} = \begin{bmatrix} \cos\alpha_0 & -\sin\alpha_0 \\ \sin\alpha_0 & \cos\alpha_0 \end{bmatrix} \tag{2.20}$$

is a counterclockwise rotation matrix by α_0.

Thus, the unknown α_0 corresponds to an overall unknown scene rotation which can be disambiguated with prior knowledge – along with scale and translation. Therefore, we can use the $\hat{\boldsymbol{\alpha}}$ estimates (2.18) to determine the unit vectors $\{\boldsymbol{u}_{rt}\}$ and proceed with localization as described earlier for the known-orientation case.

Partial Measurements

The problem is complicated slightly in practice where finite communication ranges makes it unrealistic to assume that each node obtains AOA measurements from all other nodes—that is, some elements of Φ, and consequently B, may be unknown.

An unknown element $B_{rt} = e^{i(\alpha_r - \alpha_t)}$ may be derived from any known path from r to t. Consider a path

$$\mathcal{P} = \{(r, n_1), (n_1, n_2), ...(n_k, t)\} \tag{2.21}$$

with k intermediate nodes. We then have that

$$\begin{aligned} B_{rt} &= e^{i(\alpha_r - \alpha_t)} \\ &= e^{i(\alpha_r - \alpha_{n_1})} \cdot e^{i(\alpha_{n_1} - \alpha_{n_2})} \cdot \ldots \cdot e^{i(\alpha_{n_k} - \alpha_t)} \\ &= \prod_{(m,n)\in\mathcal{P}} B_{mn} \end{aligned} \tag{2.22}$$

and see that B_{rt} may be estimated as the product of entries along any path from r to t for which measurements are known. In the case of noisy measurements a minimum-length path should be used in order to decrease the variance of the resulting estimate. Conventional shortest-path algorithms, such as Floyd-Warshall and Dijkstra, are useful here [51]. As an extension to this technique, it may be desirable to perform a *weighted* rank-one approximation of B which minimizes the weighted Frobenius norm

$$J(a) = \sum_{(r,t)} W_{rt}((aa^*)_{rt} - B_{rt})^2, \tag{2.23}$$

where the weights $W_{rt} = 1/h_{rt}$ and h_{rt} is equal to the number of hops used in establishing entry B_{rt}. Unfortunately, the known methods for computing weighted

low-rank approximations are all iterative [52], unlike the unweighted case which is determined via a single eigen-decomposition. Although not considered here, missing entries in B could also be constructed from a weighted average among multiple paths. Our implementation is based on the shortest path using Dijkstra's algorithm and the (unweighted) eigen-approximation (2.17).

The entire algorithm, termed Robust Angulation using Subspace Techniques (RAST), is summarized in Table 2.1. In the noiseless case, RAST yields the exact solution. Performance in noise is considered in the next section.

- If sensor orientations α are unknown
  - Calculate B using (2.15)
  - If B has missing entries from incomplete measurements
    - Use shortest path construction (2.22) to complete entries
  - Compute eigen-decomposition $B = V_B \Lambda_B V_B^*$
  - Estimate orientations as the phase of the elements of the eigenvector corresponding to the largest eigenvalue.
- Calculate global-frame arrival angles $\{\theta_{rt}\}$ from (2.1) and compute unit vectors as in (2.2)
- Form matrix A as in (2.6) and compute its SVD $A = U_A \Sigma_A V_A^T$
- Estimate the relative point configuration from the rightmost singular vector of A, $\hat{\boldsymbol{p}} = V_A^{(2n)}$
- If absolute positioning (scale, translation, rotation) are required, use Procrustes equations (2.10), (2.11), (2.12) to estimate these parameters from $\hat{\boldsymbol{p}}$ and anchor points

Table 2.1: Robust Angulation using Subspace Techniques (RAST), algorithm summary.

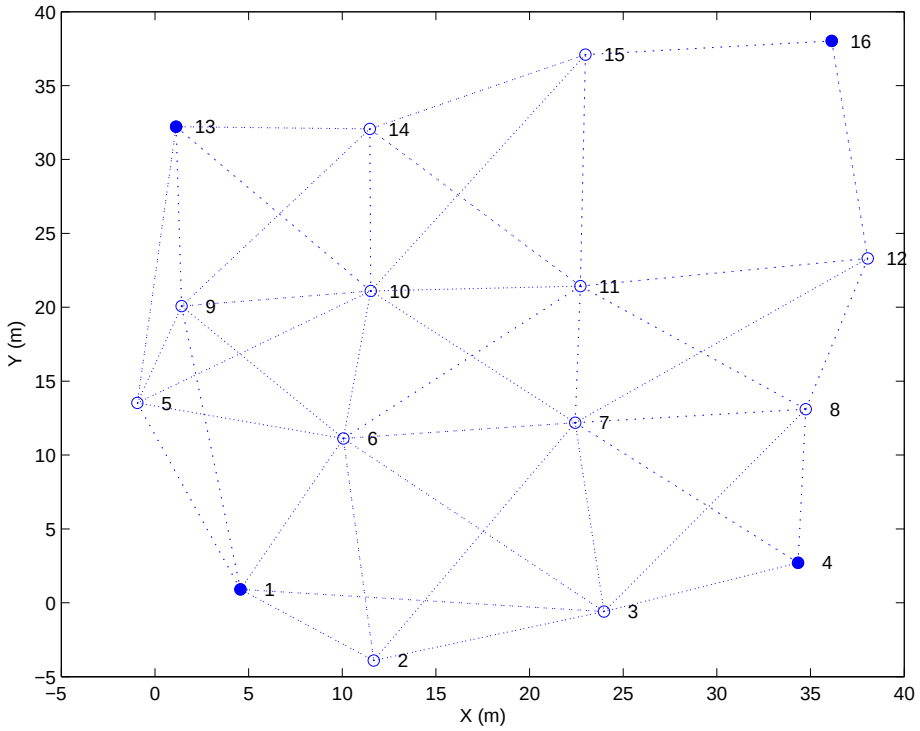


Figure 2.2: Sample sparse network to be localized with anchor nodes on the four corners.

2.3 Performance in Noise and the CRB

Through simulation, we evaluate the performance of our algorithm on the sparse network depicted in Figure 2.2 and compare the localization performance to the Cramér-Rao bound (CRB) and the maximum-likelihood estimator (MLE). In the figure, the measurement radius has been limited to 20 m making the network relatively sparse. The measurements $\{\tilde{\phi}_{rt}\}$ are Gaussian perturbations of the true values: $\tilde{\phi}_{rt} = \phi_{rt} + n_{rt}$, $n_{rt} \sim N(0, \sigma_\phi^2)$, where σ_ϕ^2 is the common observation noise variance. The four corner nodes of the network serve as anchors.

Let $\mathbf{\Theta} = \{x_i \; y_i \; \alpha_i\}_{i=1:n}$ denote the parameter vector to be estimated, and let $\phi(\mathbf{\Theta})$ be the resulting AOA measurements that would result from a proposed value of $\mathbf{\Theta}$. The MLE for this problem is equal to the (non-linear) least squares estimate [31]

$$\hat{\mathbf{\Theta}}_{mle} = \arg\min_{\mathbf{\Theta}} J(\mathbf{\Theta}), \tag{2.24}$$

where $J(\mathbf{\Theta})$ is the cost function

$$J(\mathbf{\Theta}) = \sum_{(r,t)\in\mathcal{M}} \left(\tilde{\phi}_{rt} - \phi(\mathbf{\Theta})_{rt}\right)^2. \tag{2.25}$$

As a performance metric we consider the scene RMS error defined as

$$E_{rms} = \left(\frac{1}{n - |\mathcal{K}|} \sum_{i\notin\mathcal{K}} E[\hat{d}_i^2]\right)^{\frac{1}{2}}, \tag{2.26}$$

where $E[\hat{d}_i^2]$ is the expected value of the squared distance between an estimate and the true position of node i. For the MLE and RAST estimators, this expected quantity is derived from a large number of simulations. A lower bound on the minimum possible error achievable by any unbiased estimator is given by the Cramér-Rao bound (CRB) which is derived from the inverse of the Fisher information matrix (FIM) [53]. The FIM for this problem is derived in [49, 32].

2.3.1 Simulations

In Figure 2.3 we plot the localization performance of the RAST algorithm as a function of the AOA measurement error σ_ϕ for the network in Figure 2.2—based on 1000 Monte Carlo measurement realizations. We consider a wide range in the quality of AOA estimates available; with errors ranging from good ($\sigma_\phi = 2.5°$) to very poor ($\sigma_\phi = 20°$). As a point of reference, AOA errors of approximately $3° - 5°$ were experimentally observed for an RF-based system in [49] and for an ultrasound-based

system in [54]. From Figure 2.3, we see that the estimation error is approximately 60% greater than the minimum possible CRB-predicted error when $\sigma_\phi = 20^\circ$ and sensor orientations are known. When the measurement noise decreases to $\sigma_\phi = 2.5^\circ$, the estimation error exceedes the CRB by only 21%. Similarly, for the unknown-orientation case, the estimation error ranges from 29% to 55% above the CRB as σ_ϕ varies from 2.5° to 20°.

Figures 2.4 and 2.5 present simulation results comparing RAST performance with maximum-likelihood performance in localizing the sample network of Figure 2.2. The MLE was obtained as an iterative solution to the nonlinear least-squares problem (2.24) using Matlab's implementation of the interior-reflective Newton method [55]. In the unknown orientation case, MLE convergence is very sensitive to the initial value given to the algorithm and is prone to converge to local maxima. These figures consider three different initialization mechanisms for the MLE: *(i)* the true point positions (used as a benchmark), *(ii)* random initial points drawn uniformly from the 40m $\times$ 40m scene, and *(iii)* the output from our subspace estimator.

When the sensor orientations are known, we see from Figure 2.4 that all of the ML estimates have nearly identical error and that they outperform RAST errorwise. However, as illustrated in Table 2.2, the closed-form nature of RAST results in a runtime that is nearly two orders of magnitude less than the iterative MLE. Algorithm runtimes were measured with Matlab's `cputime` function.

When orientations are unknown, the cost function (2.25) becomes plagued with local maxima and the MLE optimization has a very difficult time converging to the global maximum. In the case of random initialization the MLE always converged to distant local maxima resulting in errors significantly outside the plot region of

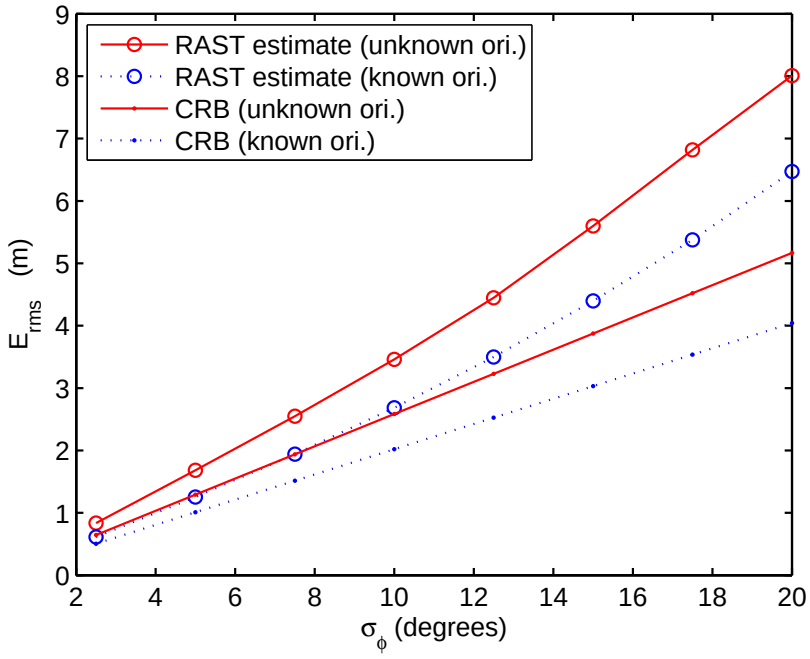


Figure 2.3: Performance of the RAST localization method for known and unknown sensor orientations compared to the Cramér-Rao lower bound

Figure 2.5. However, the RAST algorithm provided estimates sufficiently close to the attraction region of the optimal solution that MLE initialized from RAST was nearly equivalent to initialization with true values. As such, RAST-initialized maximum-likelihood estimates nearly achieve CRB performance. For $2.5° \leq \sigma_\phi \leq 12°$, the amount that these estimates exceed the CRB by ranges from only 1% to 4%, for both known and unknown orientations. Runtimes for the unknown orientation case are given in Table 2.3.

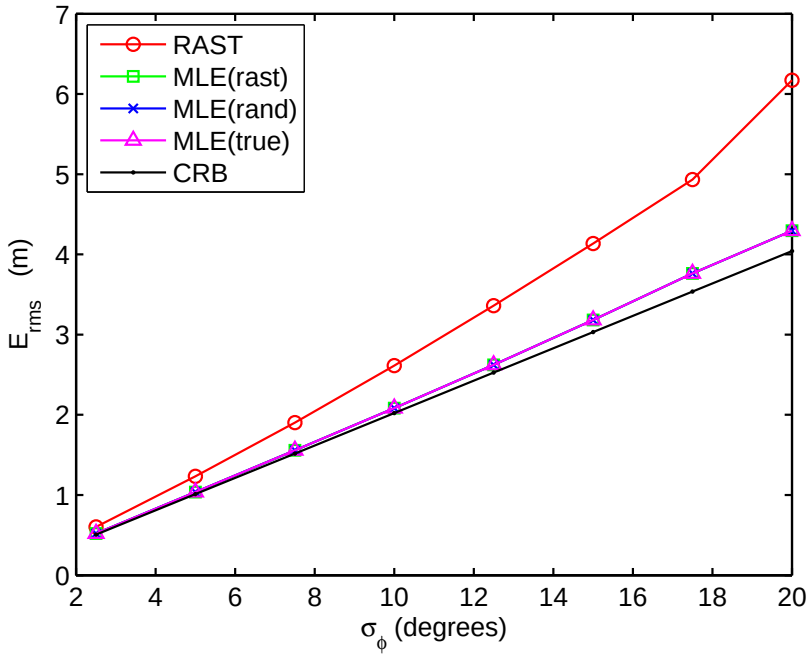


Figure 2.4: *Orientations known* – Comparison of the RAST estimator to the iterative maximum-likelihood estimator initialized in different ways. All maximum-likelihood estimates are nearly identical for this case.

2.4 Complexity

The dimensions of the matrix A in (2.6) are $(|\mathcal{M}|+2)\times 2n$, and for practical cases the total number of measurements $|\mathcal{M}| > 2n$, making the complexity of evaluating the SVD $\mathcal{O}\left((|\mathcal{M}|+2)(2n)^2\right)$. For unknown orientations, the matrix B in (2.15) is $n \times n$ and its eigen-decomposition $\mathcal{O}(n^3)$. Using the Dijkstra or Floyd-Warshall algorithms for shortest path construction to fill in missing entries in the matrix B has worst-case complexity $\mathcal{O}(n^3)$ [51, Ch. 5]. Thus, the overall complexity of the algorithm is $\mathcal{O}(|\mathcal{M}|\, n^2)$ for both known and unknown orientation angles.

Estimator	RMS Error (m)	CPU time per estimate (ms)
RAST	1.23	9
ML(true)	1.04	384
ML(rast)	1.04	377
ML(rand)	1.04	823
CRB	1.01	*n.a.*

Table 2.2: *Orientations known* – Performance and CPU time characteristics of each estimator, $\sigma_\phi = 5°$.

Estimator	RMS Error (m)	CPU time per estimate (ms)
RAST	1.60	48
ML(true)	1.30	670
ML(rast)	1.30	662
ML(rand)	3418	19258
CRB	1.29	*n.a.*

Table 2.3: *Orientations unknown* – Performance and CPU time characteristics of each estimator, $\sigma_\phi = 5°$.

There are a number of numerical linear algebra techniques that can reduce this complexity. The matrix A is sparse and we need only the least dominant right singular vector, not the entire SVD; thus using Lanczos method reduces the storage and compute time required [56, Ch. 9]. Using Lanczos method and the RAST algorithm, we observe that a 2 GHz desktop can simultaneously localize a 2000 node network in approximately 15 sec when each node makes 10 AOA measurements to its neighbors.

If new AOA measurements are subsequently made after an initial localization solution obtained, the matrix A can be augmented with new rows and the original nullspace basis $\boldsymbol{p}$ can be *updated* in $\mathcal{O}(n^2)$ operations without having to compute the SVD of the augmented matrix from scratch [56, §12.5.5], [57]. The situation of a new

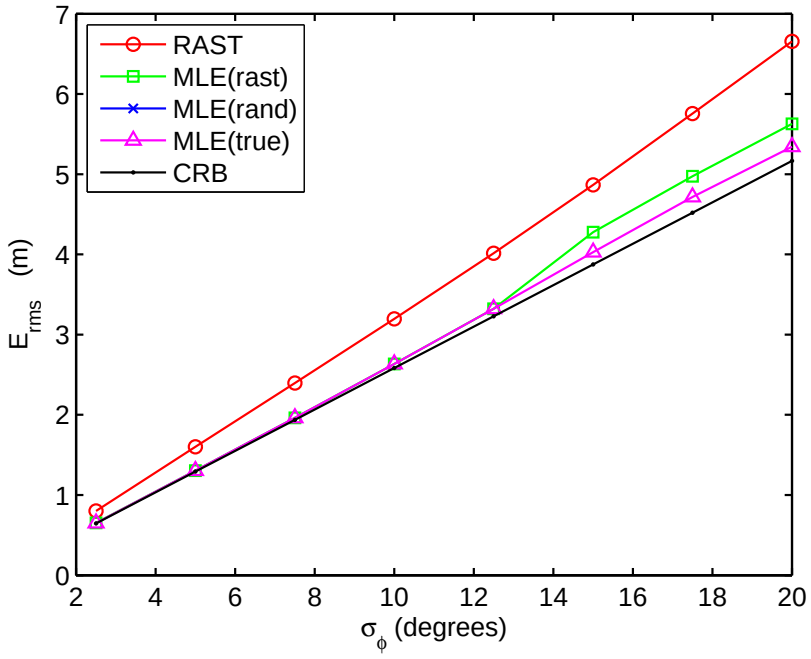


Figure 2.5: *Orientations unknown* – For this case, random initialize of the MLE failed to achieve the global maximum and errors are significantly outside of the plot region. MLE initialization from RAST produces results nearly as good as true-value initialization (used as a benchmark).

sensor node joining a previously calibrated network is more complicated as both new rows and columns must be appended to A. In this case, a previously computed SVD of A can be updated in two steps as row augmentations to A and A^T as described in [58]. This update can be performed in $\mathcal{O}\left((|\mathcal{M}| + n)n\right)$ operations.

2.5 Merging Subgraphs

The subspace-based algorithm presented in this chapter is centralized, however, it naturally accommodates the merging of localization solutions from multiple subgraphs into one complete solution. This fits the popular hierarchical model for large sensor

networks where the entire network is divided into clusters, and each cluster is managed by a cluster-head. Using the methods described in this section, each cluster may perform localization of its constituent nodes using the previously described algorithm. All of the clusters may then be "stitched" together into a complete network. For clarity of presentation, we describe the procedure for a 2-cluster system, as illustrated in Figure 2.6.

2.5.1 Known Orientations

Let $\boldsymbol{p}^{(1)}$ denote a relative localization solution for Cluster 1. That is, $U_1 K_1 \boldsymbol{p}^{(1)} = 0$, where U_1 is derived solely from measurements between nodes in Cluster 1, and K_1 forms the necessary differences between positions of nodes in Cluster 1. Similarly, let the solution for the second cluster be $U_2 K_2 \boldsymbol{p}^{(2)} = 0$. Without loss of generality, we will assume the position of $\boldsymbol{p}^{(1)}$ to be fixed and determine the transformation of $\boldsymbol{p}^{(2)}$ necessary to merge the two solutions in a manner that is consistent with the measurements made *between* the two clusters. When sensor orientations are known, this amounts to determining a suitable translation and scaling of $\boldsymbol{p}^{(2)}$. In particular, we seek

$$\tilde{\boldsymbol{p}}^{(2)} = s\,\boldsymbol{p}^{(2)} + x\,\boldsymbol{v}_x + y\,\boldsymbol{v}_y = Q\,\boldsymbol{\delta}, \tag{2.27}$$

where $Q = [\boldsymbol{p}^{(2)}\ \boldsymbol{v}_x\ \boldsymbol{v}_y]$, and the elements of $\boldsymbol{\delta} = [s\ x\ y]^T$ represent scale, x-translation, and y-translation, respectively.

The linear system (2.4) for the entire network may be written

$$\begin{bmatrix} U_1^T & 0 & 0 \\ 0 & U_x^T & 0 \\ 0 & 0 & U_2^T \end{bmatrix} \begin{bmatrix} K_1 & 0 \\ K_x^{(1)} & K_x^{(2)} \\ 0 & K_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{p}^{(1)} \\ \boldsymbol{p}^{(2)} \end{bmatrix} = 0, \tag{2.28}$$

where U_x is formed from cross-cluster measurements (transmitter in one cluster, receiver in another) and $[K_x^{(1)} K_x^{(2)}]$ forms the necessary vector differences between cross-cluster pairs. Rewriting (2.28) as

$$\begin{bmatrix} U_1^T K_1 & 0 \\ U_x^T K_x^{(1)} & U_x^T K_x^{(2)} \\ 0 & U_2^T K_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{p}^{(1)} \\ \boldsymbol{p}^{(2)} \end{bmatrix} = 0, \tag{2.29}$$

we see that the total network gives three systems of equations (i) $U_1^T K_1 \boldsymbol{p}^{(1)} = 0$, (ii) $U_x^T K_x^{(1)} \boldsymbol{p}^{(1)} + U_x^T K_x^{(2)} \boldsymbol{p}^{(2)} = 0$, and (iii) $U_2^T K_2 \boldsymbol{p}^{(2)} = 0$. Equations (i) and (iii) are solved by performing relative localization within each cluster (as already assumed), and equation (ii) provides the constraints necessary to determine how the clusters themselves are positioned relative to one another. Combining (ii) and (2.27) we can estimate the translation and scale parameters as

$$\hat{\boldsymbol{\delta}} = -(U_x^T K_x^{(2)} Q)^\dagger U_x^T K_x^{(1)} \boldsymbol{p}^{(1)}, \tag{2.30}$$

where $(\cdot)^\dagger$ denotes pseudo-inverse.

2.5.2 Unknown Orientations

When the orientations of the sensors $\{\alpha_i\}$ are unknown, the relative cluster estimates $\boldsymbol{p}^{(1)}$ and $\boldsymbol{p}^{(2)}$ have arbitrary orientations. Thus, in addition to translation and scale, we also seek a rotation angle Δ of $\boldsymbol{p}^{(2)}$.

Let $\hat{\boldsymbol{a}}^{(i)}$ denote the complex orientation estimates from the relative localization of Cluster i (see (2.17)), and let $\tilde{\boldsymbol{a}}^{(i)} \in \mathbb{C}^{n^{(i)}}$ denote the subset of $\hat{\boldsymbol{a}}^{(i)}$ corresponding to the $n^{(i)}$ sensors of Cluster i involved in cross-cluster measurements. For the example in Figure 2.6, $n^{(1)} = 2$ and $\tilde{\boldsymbol{a}}^{(1)} = [e^{i\alpha_3}\ e^{i\alpha_4}]^T$. Form the $(n^{(1)} + n^{(2)}) \times (n^{(1)} + n^{(2)})$ matrix B_x as in (2.15) using the cross-cluster measurements and compute its rank-1

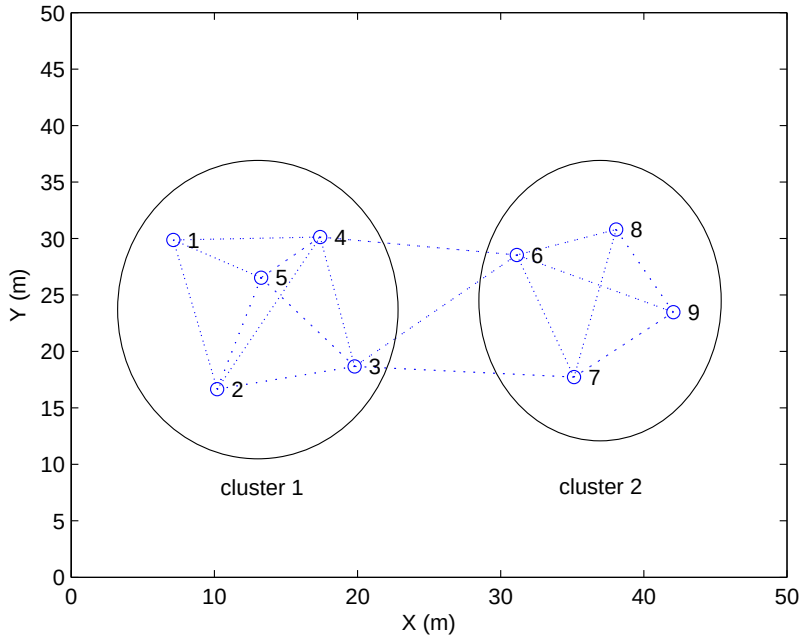


Figure 2.6: A large network can be localized in clusters which may subsequently be merged together using cross-cluster measurements.

approximation using eigen-decomposition as in (2.17)

$$\hat{\boldsymbol{a}}_x \, \hat{\boldsymbol{a}}_x^* \approx B_x. \tag{2.31}$$

As before, the elements of $\hat{\boldsymbol{a}}_x$ are partitioned as $\hat{\boldsymbol{a}}_x^{(1)}$ and $\hat{\boldsymbol{a}}_x^{(2)}$, representing the elements from Cluster 1 and Cluster 2, respectively. Finally, rotational offsets $\Delta^{(i)}$ need to be estimated for each cluster in order for the previous orientation estimates to coincide with those of the cross-cluster measurements. Specifically, we solve for real scalars $\Delta^{(1)}$ and $\Delta^{(2)}$ such that

$$\left\| \begin{bmatrix} e^{i\Delta^{(1)}} \, \tilde{\boldsymbol{a}}^{(1)} \\ e^{i\Delta^{(2)}} \, \tilde{\boldsymbol{a}}^{(2)} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{a}}_x^{(1)} \\ \hat{\boldsymbol{a}}_x^{(2)} \end{bmatrix} \right\|_2^2 \tag{2.32}$$

is minimized:

$$\Delta^{(i)} = \angle \left((\tilde{\boldsymbol{a}}^{(i)*} \hat{\boldsymbol{a}}_x^{(i)}) / (\tilde{\boldsymbol{a}}^{(i)*} \tilde{\boldsymbol{a}}^{(i)}) \right), \quad i = 1, 2. \tag{2.33}$$

Since, as shown in Section 2.2.2, uniformly adjusting the orientation angles by Δ corresponds to a rotation of estimates by Δ, we can apply a rotation of $\Delta = \Delta^{(2)} - \Delta^{(1)}$ to the relative cluster estimate $\boldsymbol{p}^{(2)}$. Translation and scale may then be determined as in Section 2.5.1.

Finally, we note that the process of aligning two clusters only utilizes cross-cluster measurements and relative cluster-position estimates for sensors involved in these measurements[1]. Therefore, the transmission and computation requirements for merging are proportional to the number of cross-cluster measurements—which is expected to be much smaller than the number of intra-cluster measurements. Networks of more than two clusters may be localized by merging one cluster at a time with a

[1]From (2.30), $\hat{\boldsymbol{\delta}}$ appears to depend on $\boldsymbol{p}^{(1)}$ and $\boldsymbol{p}^{(2)}$, however $K_x^{(i)}$ only has non-zero columns corresponding to sensors involved in cross-cluster measurements.

growing "base-cluster." However, in the presence of noise, this solution will depend on the order in which the clusters are added. Future work will consider methods to simultaneously merge multiple clusters.

2.6 Conclusions

In this chapter we presented a new localization algorithm that provides both sensor position and orientation estimates from angle-of-arrival measurements. The algorithm is subspace based in that it relies on singular-value and eigen-decompositions to identify subspaces associated with position and orientation vectors. The algorithm compared favorably to the Cramér-Rao bound on estimation error variance and was shown to have complexity $\mathcal{O}(|\mathcal{M}|\, n^2)$, where $|\mathcal{M}|$ is the number of measurements and n is the total number of sensors. In Monte Carlo simulations, the closed-form nature of the algorithm resulted in a runtime nearly two orders of magnitude less than an iterative maximum-likelihood estimator. The fast and robust nature of the algorithm makes it well-suited to initialize high fidelity routines, such as the MLE. Finally, while the algorithm is centralized, it was demonstrated how disjoint subnetwork clusters could be localized independently and subsequently merged together with minimal overhead.

CHAPTER 3

RELATIVE AND ABSOLUTE POSITIONING ERRORS IN SELF-LOCALIZATION SYSTEMS

3.1 Introduction

In order to better understand how noise, deployment geometry, and measurement type effect fundamental location estimation performance, a number of authors have considered the Cramér-Rao bound (CRB) for these scenarios. The CRB has been derived for distance (range) measurements in [59, 60], for TOA measurements in [32, 19, 59, 31], for TDOA measurements in [61], for RSS measurements in [32, 19, 61], and for AOA measurements in [49, 32, 31, 60].

As illustrated in Figure 1.1, the absolute self-localization problem combines inter-node measurements collected in a measurement vector $\boldsymbol{z}$ with prior information in order to obtain coordinate estimates $\{(\hat{x}_i, \hat{y}_i) : i \in (1, \ldots, N)\}$ of the N constituent nodes of the network. In this chapter we establish a general partitioning of localization error that may be applied to all of the previously mentioned measurement types and their associated CRBs. In particular, we decompose the total localization error into a relative portion representing error in the estimated node locations relative to one another (network "shape") and a transformation component representing error in the

absolute position of the relative scene. This decomposition is motivated by the fact that relative information is derived from both measurements and prior information, while transformation information comes solely from prior information. In this chapter we also relate, for a number of different measurement types, the nullspace of the Fisher information matrix (FIM) to transformation parameters not informed upon by measurements.

Because inter-node calibration measurements provide no information about the location and orientation of the entire network, estimation of the absolute node locations from inter-node measurements alone is an ill-conditioned problem. However, inter-node measurements provide significant information on the node locations *relative* to one another. In order to regularize the absolute localization problem, generic constraints on the parameter set are considered. These constraints represent a generalization of the use of anchor (also called beacon) nodes typically used in self-localization.

One of the main contributions of this work is an analysis illustrating how the constraint subspace interacts with the measurement subspace to effect total localization performance. Along with the CRB itself, the relative–transformation decomposition presented here gives insight into how external inputs effect absolute localization. This partitioning of error is also useful to higher level applications in a sensor network that utilize output and uncertainty of the localization service.

This chapter is organized as follows. Definitions of relative and transformation error in a deterministic setting are given in Section 3.2. Randomness is introduced in Section 3.3 where these concepts are extended to expected error and related to the FIM. A general measurement model, along with several specific measurement types, is also given in this section. In Section 3.4 we introduce constrained estimation

and prove generic results illustrating how the interaction between the constrained and measurement subspaces effects total estimation performance. Some examples utilizing the relative–transformation error decomposition are given in Section 3.5, and conclusions are given in Section 3.6.

Notation: Throughout this chapter capital letters represent matrices and bold lowercase letters denote column vectors. $(\cdot)^T$ and $\mathrm{tr}(\cdot)$ represent matrix transpose and trace respectively. $||\cdot||$ denotes the vector 2-norm (Euclidean norm), $\mathcal{R}(A)$ represents the column-span of matrix A, $\mathcal{R}(A)^\perp$ represents the orthogonal complement of $\mathcal{R}(A)$, and $A \geq B$ implies that $(A - B)$ is positive semidefinite. The construction $\mathrm{diag}([A_1 \ldots A_n])$ represents a block diagonal matrix with matrices $\{A_i\}$ on the main diagonal, and $A^\dagger$ denotes the Moore-Penrose pseudo-inverse of matrix A.

3.2 Relative and Transformation Error

3.2.1 Definitions

Absolute location estimates are derived from two sources of information: (1) some form of inter-node measurements (such as distances or time-difference-of-arrival) and (2) prior information in the form of constraints on the parameters. Probabilistic priors may also be considered, and can also be decomposed into relative and transformation components as considered below. The relative configuration of the sensors represents the "shape" of the network without regard to absolute location, orientation, and in some cases scale. The measurements only depend on (and thus only inform upon) the relative configuration of the nodes, whereas the prior information may inform upon both relative configuration and global transformation information. Thus, it is natural to partition the absolute coordinate estimation error into relative localization

error and transformation estimation error. Let $\boldsymbol{\theta} = [x_1 \ y_1 \ \ldots \ x_N \ y_N]^T$ be the absolute location parameter vector of an N-node system with parameter estimate $\hat{\boldsymbol{\theta}} = [\hat{x}_1 \ \hat{y}_1 \ \ldots \ \hat{x}_N \ \hat{y}_N]^T$ given by some estimator. As a performance metric for this estimator we consider the sum of the squared distances between the true node locations and their estimates

$$\begin{aligned} \epsilon &= \sum_{i=1}^{N} d_i^2 \\ &= ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}||^2, \end{aligned} \tag{3.1}$$

where $d_i^2 = (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2$. If the estimator did not yield the optimal transformation parameters (translation, rotation, and potentially scale), the error in (3.1) can be further reduced by applying a planar transformation to the previous estimates. Consider translation of the N node locations in $\hat{\boldsymbol{\theta}}$ along the x- and y-axis by x and y respectively, rotation counter clockwise by an angle ϕ, and scaling by a factor s—both about the point $p_c = [x_c \ y_c]^T$. Letting $\alpha = \{s, \phi, x, y\}$, we write this combined transformation of a location estimate vector $\hat{\boldsymbol{\theta}}$ as

$$T_\alpha(\hat{\boldsymbol{\theta}}) = sR_\phi(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}} + x\boldsymbol{v}_x + y\boldsymbol{v}_y, \tag{3.2}$$

where $\boldsymbol{v}_x = [1\,0\,1\,0\,\ldots]^T \in \mathbb{R}^{2N\times 1}$, $\boldsymbol{v}_y = [0\,1\,0\,1\,\ldots]^T \in \mathbb{R}^{2N\times 1}$, $\bar{\boldsymbol{\theta}} = [x_c\, y_c\, x_c\, y_c \ldots]^T \in \mathbb{R}^{2N\times 1}$, and the total rotation matrix R_ϕ is composed of N 2×2 equivalent rotation matrices, $R_\phi = \text{diag}([\Gamma_\phi \ \Gamma_\phi \ \ldots \ \Gamma_\phi]) \in \mathbb{R}^{2N\times 2N}$, where

$$\Gamma_\phi = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}. \tag{3.3}$$

In practice, measurements provide information about a strict subset of α, which may be the empty subset; we denote this subset as $\alpha_m \subset \alpha$. The remaining transformation parameters $\overline{\alpha_m} = \alpha \setminus \alpha_m$ are completely non-estimable from the measurements. With this partitioning, we write $\alpha = \{\alpha_m, \overline{\alpha_m}\}$. For example, when

the measurements are inter-node distances, $\alpha_m = \{s\}$ and $\overline{\alpha_m} = \{\phi, x, y\}$. Other measurement types are explored in detail in Section 3.3.2.

We denote by $\overline{\alpha_{m_0}}$ the optimal transformation parameters over $\overline{\alpha_m}$ for a particular estimate $\hat{\boldsymbol{\theta}}$

$$\overline{\alpha_{m_0}} = \arg\min_{\overline{\alpha_m}} ||\boldsymbol{\theta} - T_{\alpha_0}(\hat{\boldsymbol{\theta}})||^2. \tag{3.4}$$

In (3.4), $\alpha_0 = \{s_0, \phi_0, x_0, y_0\} = \{\alpha_{m_0}, \overline{\alpha_{m_0}}\}$ and $\alpha_{m_0} \subset \bar{\alpha} = \{\bar{s}, \bar{\phi}, \bar{x}, \bar{y}\} = \{1, 0, 0, 0\}$, with the appropriate elements chosen as necessary. Thus, a transformation based on α_0 optimally corrects components not informed upon by the measurements, but does not alter the other components. In the distance measurement example, $\alpha_{m_0} = \{\bar{s}\} = \{1\}$ and the optimization (3.4) does not include scale.

We denote the optimally transformed estimate by

$$\hat{\boldsymbol{\theta}}_r = T_{\alpha_0}(\hat{\boldsymbol{\theta}}) \tag{3.5}$$

and define the *relative* localization error to be

$$\epsilon_r = ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_r||^2. \tag{3.6}$$

The *transformation* error is the portion of the total error due to the estimator not correctly estimating the transformation parameters and is given by

$$\begin{aligned} \epsilon_t &= \epsilon - \epsilon_r \\ &= ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}||^2 - ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_r||^2. \end{aligned} \tag{3.7}$$

The optimal transformation parameters are determined by singular-value decomposition (SVD) and given as the solution to the extended orthogonal Procrustes problem, first solved by Schönemann and Carroll [50], and repeated here for completeness.

Let $A = [x_1\, y_1; x_2\, y_2; ...] \in \mathbb{R}^{N\times 2}$ represent a rearrangement of the coordinates of $\boldsymbol{\theta}$, and let B represent a similar rearrangement of $\hat{\boldsymbol{\theta}}$. Let $C = (I - N^{-1}\mathbf{1}_N\mathbf{1}_N^1)$, where $\mathbf{1}_N$ is a column vector of N ones, and compute the SVD $U_p\Lambda_p V_p^T = A^TCB$. The optimal transformation parameters are

$$\Gamma_{\phi_0} = U_pV_p^T \tag{3.8}$$

$$s_0 = (\text{tr } A^TCB\Gamma_{\phi_0}^T)/(\text{tr } B^TCB) \tag{3.9}$$

$$[x_0\ y_0]^T = N^{-1}(A - s_0B\Gamma_{\phi_0}^T)^T\mathbf{1}_N. \tag{3.10}$$

If modifications to scale are not required, (3.9) may be ignored and $s_0 = 1$ substituted into (3.10) for optimal translation under this condition. Because the Procrustes method determines $\hat{\boldsymbol{\theta}}_r$, $\sqrt{\epsilon_r}$ is sometimes referred to as the Procrustes distance.

3.2.2 Linear Subspace Approximation

For any scene estimate $\hat{\boldsymbol{\theta}}$, we can define an equivalence class $S(\hat{\boldsymbol{\theta}})$ which represents all scalings and rigid transformations of $\hat{\boldsymbol{\theta}}$ such that the Procrustes distance between any two elements of $S(\hat{\boldsymbol{\theta}})$ is zero. The equivalence class $S(\hat{\boldsymbol{\theta}})$ is a four-dimensional non-linear manifold in $\mathbb{R}^{2N}$ parameterized by $s, \phi, x,$ and y. The remainder of this section is devoted to developing a linear approximation of the transformation operator which will give us a linear subspace representation of $S(\hat{\boldsymbol{\theta}})$ and allow us to simplify the expressions for ϵ_r and ϵ_t.

We linearize the transformation operator by considering a first order Taylor approximation of T_α about $\bar{\alpha}$

$$T_\alpha(\boldsymbol{\theta}) \approx T_{\bar{\alpha}}(\boldsymbol{\theta}) + \left.\frac{\partial\, T_\alpha(\boldsymbol{\theta})}{\partial\, \alpha^T}\right|_{\alpha=\bar{\alpha}} (\alpha - \bar{\alpha}), \tag{3.11}$$

where

$$\begin{aligned}\left.\frac{\partial T_\alpha(\boldsymbol{\theta})}{\partial \alpha^T}\right|_{\alpha=\bar{\alpha}} &= \left[R_\phi(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}), s\frac{d R_\phi}{d\phi}(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}), \boldsymbol{v}_x, \boldsymbol{v}_y\right]\Big|_{\alpha=\bar{\alpha}} \\ &= \left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}), M_0(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}), \boldsymbol{v}_x, \boldsymbol{v}_y\right], \end{aligned} \tag{3.12}$$

and where $M_0 = \left.\frac{d R_\phi}{d\phi}\right|_{\phi=0} = \text{diag}([\Phi_0\ \Phi_0\ \ldots \Phi_0]) \in \mathbb{R}^{2N\times 2N}$, and

$$\Phi_0 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{3.13}$$

In $\mathbb{R}^2$ we wish to consider scaling and rotation of each point (x_i, y_i) about the true centroid $(\bar{x}, \bar{y})$, where $\bar{x} = N^{-1}\sum_i x_i$ and $\bar{y} = N^{-1}\sum_i y_i$. In $\mathbb{R}^{2N}$ this corresponds to a single operation on $\boldsymbol{\theta}$ about $\bar{\boldsymbol{\theta}} = \bar{x}\boldsymbol{v_x} + \bar{y}\boldsymbol{v_y}$. Substituting this value for $\bar{\boldsymbol{\theta}}$ in (3.12) and using (3.11) we obtain

$$T_\alpha(\boldsymbol{\theta}) \approx \boldsymbol{\theta} + (s-1)\boldsymbol{v}_s + \phi\boldsymbol{v}_\phi + x\boldsymbol{v}_x + y\boldsymbol{v}_y, \tag{3.14}$$

where

$$\boldsymbol{v}_s = \begin{bmatrix} (x_1-\bar{x}) \\ (y_1-\bar{y}) \\ (x_2-\bar{x}) \\ (y_2-\bar{y}) \\ \vdots \end{bmatrix} \text{ and } \boldsymbol{v}_\phi = \begin{bmatrix} -(y_1-\bar{y}) \\ (x_1-\bar{x}) \\ -(y_2-\bar{y}) \\ (x_2-\bar{x}) \\ \vdots \end{bmatrix}. \tag{3.15}$$

Note that $\boldsymbol{v}_s$ and $\boldsymbol{v}_\phi$ depend on $\boldsymbol{\theta}$, the point being transformed. However, the approximation (3.14) remains valid for transforming any other point $\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}$ when $\alpha \approx \bar{\alpha}$. We use this fact in subsequent error analysis to construct the scaling vector v_s and rotation vector v_ϕ from the true parameter vector $\boldsymbol{\theta}$. These may then be applied, as in (3.14), to perform transformations on estimates $\hat{\boldsymbol{\theta}}$ which are close to $\boldsymbol{\theta}$.

When $\alpha_m = \{\}$, we can approximate ϵ_r as

$$\epsilon_r \approx ||\hat{\boldsymbol{\theta}} + (s_0-1)\boldsymbol{v}_s + \phi_0\boldsymbol{v}_\phi + x_0\boldsymbol{v}_x + y_0\boldsymbol{v}_y - \boldsymbol{\theta}||^2, \tag{3.16}$$

and we define the approximation to the relative error as

$$\tilde{\epsilon}_r = \min_{\boldsymbol{\beta}} ||\hat{\boldsymbol{\theta}} + \beta_s \tilde{v}_s + \beta_\phi \tilde{v}_\phi + \beta_x \tilde{v}_x + \beta_y \tilde{v}_y - \boldsymbol{\theta}||^2, \tag{3.17}$$

where $\boldsymbol{\beta} = [\beta_s\, \beta_\phi\, \beta_x\, \beta_y]^T$ and, for convenience, we used normalized versions of the already orthogonal transformation vectors

$$\tilde{v}_x = \boldsymbol{v}_x/||\boldsymbol{v}_x||,\ \tilde{v}_y = \boldsymbol{v}_y/||\boldsymbol{v}_y||,\ \tilde{v}_\phi = \boldsymbol{v}_\phi/||\boldsymbol{v}_\phi||,\ \tilde{v}_s = \boldsymbol{v}_s/||\boldsymbol{v}_s|| \tag{3.18}$$

with $||\boldsymbol{v}_x||^2 = ||\boldsymbol{v}_y||^2 = N$ and $||\boldsymbol{v}_s||^2 = ||\boldsymbol{v}_\phi||^2 = \sum \tilde{d}_i^2$, where $\tilde{d}_i$ is the distance of node i to the scene center $(\bar{x}, \bar{y})$. Let $W = [\tilde{v}_s\, \tilde{v}_\phi\, \tilde{v}_x\, \tilde{v}_y]$ and denote by

$$\begin{aligned} \hat{\boldsymbol{\beta}} &\triangleq W^T(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \\ &= [\hat{\beta}_s\, \hat{\beta}_\phi\, \hat{\beta}_x\, \hat{\beta}_y]^T \end{aligned} \tag{3.19}$$

the minimizing value of (3.17). Then we may define the transformation correction to $\hat{\boldsymbol{\theta}}$ as $\tilde{\boldsymbol{w}}_t = W\hat{\boldsymbol{\beta}}$ and can represent the linear approximation to $\hat{\boldsymbol{\theta}}_r$ as $\tilde{\boldsymbol{\theta}}_r = \hat{\boldsymbol{\theta}} + \tilde{\boldsymbol{w}}_t$. See Figure 3.1 for graphical definitions of these and other vectors of interest.

Thus, the linear approximation $\tilde{\epsilon}_r$ of the relative error ϵ_r is given by

$$\begin{aligned} \tilde{\epsilon}_r &= ||\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_r||^2 \\ &= ||\tilde{\boldsymbol{w}}_r||^2 \end{aligned} \tag{3.20}$$

and the corresponding linear approximation $\tilde{\epsilon}_t$ of the transformation error is approximated as

$$\begin{aligned} \tilde{\epsilon}_t &= \epsilon - \tilde{\epsilon}_r \\ &= ||\boldsymbol{\xi}||^2 - ||\tilde{\boldsymbol{w}}_r||^2 \\ &= ||\tilde{\boldsymbol{w}}_t||^2. \end{aligned} \tag{3.21}$$

The vectors $\tilde{\boldsymbol{w}}_t$ and $\tilde{\boldsymbol{w}}_r$ are easily expressed as projections of the error vector $\boldsymbol{\xi} = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ onto the subspaces $\mathcal{R}(W)$ and $\mathcal{R}(W)^{\perp}$

$$\tilde{\boldsymbol{w}}_t = P_W \, \boldsymbol{\xi} = WW^T \, \boldsymbol{\xi} \tag{3.22}$$

$$\tilde{\boldsymbol{w}}_r = P_W^{\perp} \, \boldsymbol{\xi} = (I - WW^T) \, \boldsymbol{\xi}. \tag{3.23}$$

When $\alpha_m \neq \{\}$, a similar derivation to (3.16)–(3.23) holds, where now we simply limit the columns of W to those corresponding to elements of $\overline{\alpha_m}$ and adjust the size of $\boldsymbol{\beta}$ accordingly.

The transformation vectors $(\tilde{v}_s, \tilde{v}_\phi, \tilde{v}_x, \tilde{v}_y)$ can also be interpreted as a meaningful orthonormal basis for the tangent plane of the manifold $S(\boldsymbol{\theta})$ at the point $\boldsymbol{\theta}$. As such, we are approximating general movements on the manifold by movements on this tangent plane.

3.3 Relation to the FIM

In the previous section we saw that rigid transformations and scaling of $\boldsymbol{\theta}$ can be approximated by $\boldsymbol{\theta} + W\boldsymbol{\beta}$. Because inter-node measurements do not generally inform about these bulk properties, we are typically only able to estimate $\boldsymbol{\theta}$ up to its equivalence class $S(\boldsymbol{\theta})$ using inter-node measurements. In terms of the linear approximation, we cannot estimate the components of $\boldsymbol{\theta}$ in the subspace spanned by $\boldsymbol{v}_s, \boldsymbol{v}_\phi, \boldsymbol{v}_x$, and $\boldsymbol{v}_y$. As the measurements provide no information about $\boldsymbol{\theta}$ in this subspace, the resulting Fisher information matrix will be singular. In this section we consider a number of different measurement models and illustrate the relationship between the FIM's nullspace and non-estimable transformation parameters.

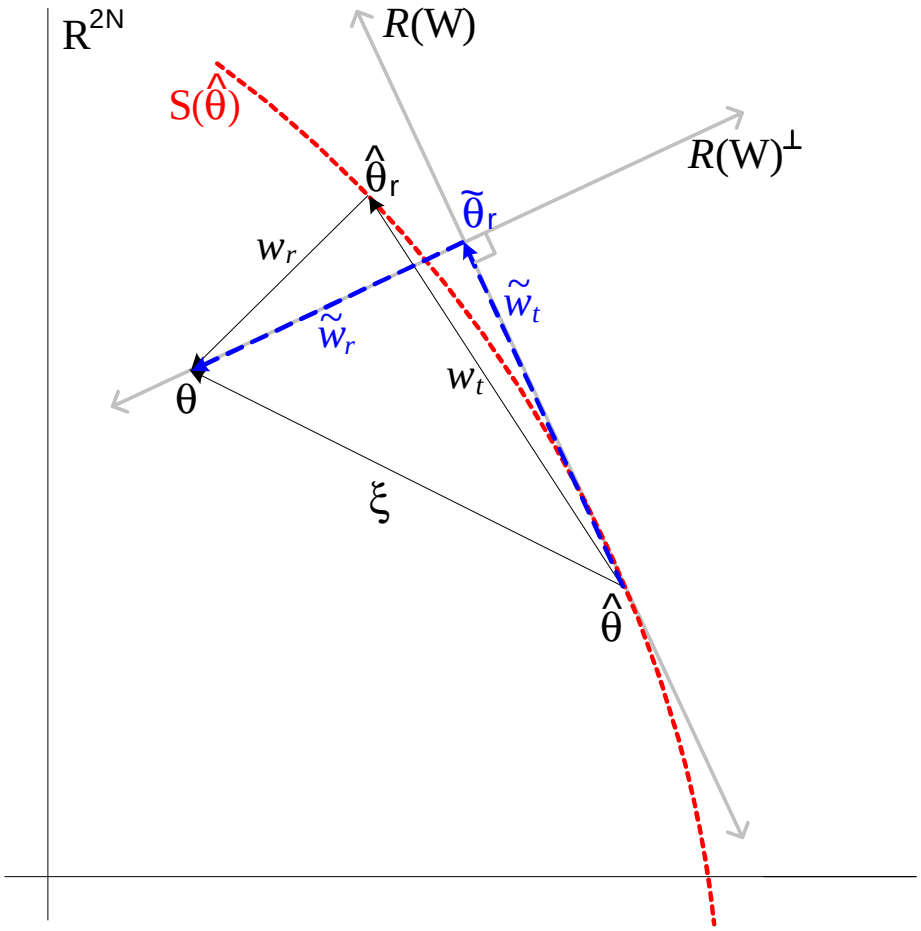


Figure 3.1: Geometric illustration of relative and transformation errors in the location parameter vector $\boldsymbol{\theta}$. The manifold $S(\hat{\boldsymbol{\theta}})$ represents rigid transformations and scalings of the coordinate estimates $\hat{\boldsymbol{\theta}}$, and the point on $S(\hat{\boldsymbol{\theta}})$ closest to $\boldsymbol{\theta}$ represents the optimally transformed estimate, $\hat{\boldsymbol{\theta}}_r$. The error vector $\xi = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ may be decomposed into $\xi = \boldsymbol{w}_r + \boldsymbol{w}_t$, where $\boldsymbol{w}_r = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_r$ is the relative error vector and $\boldsymbol{w}_t = \hat{\boldsymbol{\theta}}_r - \hat{\boldsymbol{\theta}}$. $\boldsymbol{w}_t$ and $\boldsymbol{w}_r$ may be approximated, respectively, by $\tilde{\boldsymbol{w}}_t$ and $\tilde{\boldsymbol{w}}_r$, the projections of the error vector ξ onto the transformation subspace $\mathcal{R}(W)$ and the relative subspace $\mathcal{R}(W)^{\perp}$.

3.3.1 Measurement Model and Fisher Information

We consider a measurement model of the general form

$$\boldsymbol{z} = \boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{\eta} \in \mathbb{R}^M, \tag{3.24}$$

where $\boldsymbol{z}$ is a vector of M measurements, $\boldsymbol{\mu}$ is the mean of the observation which is structured by the true coordinate parameters $\boldsymbol{\theta}$, and $\boldsymbol{\eta}$ is zero-mean noise with density $f_\eta(\boldsymbol{\eta})$. While the Fisher information matrix, J, for estimating $\boldsymbol{\theta}$ from $\boldsymbol{z}$ depends on the distribution of $\boldsymbol{\eta}$, its nullspace does not.

The Fisher information matrix is

$$J = E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_Z(\boldsymbol{z};\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_Z(\boldsymbol{z};\boldsymbol{\theta})\right]^T. \tag{3.25}$$

and

$$\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_Z(\boldsymbol{z};\boldsymbol{\theta})\right]^T = \frac{-1}{f_\eta(\boldsymbol{\eta})}\frac{\partial f_\eta(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}, \tag{3.26}$$

yielding

$$J = \int f_Z(\boldsymbol{z};\boldsymbol{\theta})\frac{1}{f_\eta^2(\boldsymbol{\eta})}\frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial f_\eta(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\frac{\partial f_\eta(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\, d\boldsymbol{z}. \tag{3.27}$$

Letting $G = \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$, and noting that $d\boldsymbol{z} = d\boldsymbol{\eta}$ we have

$$J = G^T \int \frac{1}{f_\eta(\boldsymbol{\eta})}\left[\frac{\partial f_\eta(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right]\left[\frac{\partial f_\eta(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right]^T d\boldsymbol{\eta}\; G \tag{3.28}$$

$$= G^T R G, \tag{3.29}$$

where

$$R = \int \frac{1}{f_\eta(\boldsymbol{\eta})}\left[\frac{\partial f_\eta(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right]\left[\frac{\partial f_\eta(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right]^T d\boldsymbol{\eta}. \tag{3.30}$$

Thus, $\text{null}(G) \subseteq \text{null}(J)$. Under the mild assumption that R is non-singular, the nullspace of J is fully determined by G, $\text{null}(J) = \text{null}(G)$. We further assume

that a sufficient number of measurements have been made to uniquely determine the network's shape. We note that for Gaussian noise, $\boldsymbol{\eta} \sim \mathcal{N}(0, \Sigma_\eta)$, we have $R = \Sigma_\eta^{-1}$ in (3.29).

In the following subsection we consider a number of different types of inter-node measurements and illustrate that, in each case, null(J) is spanned by a subset of the transformation vectors $\{\boldsymbol{v}_s, \boldsymbol{v}_\phi, \boldsymbol{v}_x, \boldsymbol{v}_y\}$. This provides a physically-based meaningful basis for null$(J) = \mathcal{R}(W)$ and allows us construct the matrix W needed in order to calculate relative and absolute errors for each measurement type.

3.3.2 Measurement Types

Let $\mathcal{M}$ denote the set of M ordered measurement pairs; that is $(r, t) \in \mathcal{M}$ if node r makes a measurement from a transmission originating at node t, and let $\mathcal{M}(k) = (r, t)_k$ denote the k_{th} such pair. Denote by $\boldsymbol{p}_i = [x_i \ y_i]^T$ the coordinates of the i_{th} sensor.

Distance Measurements

Let $\boldsymbol{\mu}(\boldsymbol{\theta})_k$ denote the k_{th} element of $\boldsymbol{\mu}(\boldsymbol{\theta})$, then for distance measurements

$$\boldsymbol{\mu}(\boldsymbol{\theta})_k = ||\boldsymbol{p}_r - \boldsymbol{p}_t||, \tag{3.31}$$

for $(r, t) = \mathcal{M}(k)$. In this case $\boldsymbol{\mu}(\boldsymbol{\theta})$ is unchanged by translations and rotations to $\boldsymbol{\theta}$ and we find that null(J) = null$(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T})$ = span$(\boldsymbol{v}_x, \boldsymbol{v}_y, \boldsymbol{v}_\phi)$, as expected. Thus, $W = [\tilde{\boldsymbol{v}}_x \ \tilde{\boldsymbol{v}}_y \ \tilde{\boldsymbol{v}}_\phi]$.

Time-of-arrival (TOA)

This case is essentially the same as distance measurements. The arrival time of a signal at sensor node r from a transmission by node t is equal to the distance between

them divided by the propagation velocity, c, plus the time of emission at node t, τ_t

$$\boldsymbol{\mu}(\boldsymbol{\theta})_k = ||\boldsymbol{p}_r - \boldsymbol{p}_t||/c + \tau_t. \tag{3.32}$$

Here c and τ_t are known, and again $\text{null}(J) = \text{span}(\boldsymbol{v}_x, \boldsymbol{v}_y, \boldsymbol{v}_\phi)$.

Angle-of-arrival (AOA)

As described in Chapter 2, each sensor node r makes angle measurements in a local coordinate system which is offset by an angle γ_r from a global reference. In the global coordinate system, the AOA at receiving node r, of a transmission from sensor node t is $\omega_{rt} = \psi_{rt} + \gamma_r$, where ψ_{rt} is the measurement in node r's local system

$$\boldsymbol{\mu}(\boldsymbol{\theta})_k = \psi_{rt} = \omega_{rt} - \gamma_r. \tag{3.33}$$

In the AOA measurement model, the orientation angles $\{\gamma_r : r \in (1, \ldots, N)\}$ are assumed to be known and we effectively measure arrival angles with an absolute reference. AOA measurements are invariant to translation and scaling of $\boldsymbol{\theta}$, and we find $\text{null}(J) = \text{span}(\boldsymbol{v}_x, \boldsymbol{v}_y, \boldsymbol{v}_s)$. Orientation is informed upon by the measurements $(\alpha_m = \{\phi\})$.

Time-difference-of-arrival (TDOA)

In this model, the receivers measure signal arrival times as in TOA,

$$\boldsymbol{\mu}(\boldsymbol{\theta})_k = ||\boldsymbol{p}_r - \boldsymbol{p}_t||/c + \tau_t, \tag{3.34}$$

except that the emission times $\{\tau_j : j \in (1, \ldots, N)\}$ are unknown and must be inserted into the parameter vector to be estimated, so $\boldsymbol{\theta} = [x_1\ y_1\ \tau_1\ x_2\ y_2\ \tau_2\ \ldots\ ]^T \in \mathbb{R}^{3N}$. In this model, a single arrival-time measurement is not informative without knowing the emission time, and the useful quantity is the difference between the

arrival times at two distinct receiving sensors corresponding to the same source signal. For this measurement model $\boldsymbol{\mu}(\boldsymbol{\theta})$ remains invariant to translations and rotations of the sensors. However, the corresponding transformation vectors spanning null(J) must account for the addition of the emission times in $\boldsymbol{\theta}$. Thus, the basis becomes

$$\boldsymbol{v}_x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \boldsymbol{v}_y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}, \boldsymbol{v}_\phi = \begin{bmatrix} -(y_1 - \bar{y}) \\ (x_1 - \bar{x}) \\ 0 \\ -(y_2 - \bar{y}) \\ (x_2 - \bar{x}) \\ 0 \\ \vdots \end{bmatrix}. \tag{3.35}$$

Angle-difference-of-arrival (ADOA)

The measurements for ADOA are the same as AOA,

$$\boldsymbol{\mu}(\boldsymbol{\theta})_k = \psi_{rt} = \angle(\boldsymbol{p}_r, \boldsymbol{p}_t) - \gamma_r, \tag{3.36}$$

except that the orientations of the individual sensors $\{\gamma_j : j \in (1, \ldots, N)\}$ are unknown and must be inserted into the parameter vector for estimation, so $\boldsymbol{\theta} = [x_1\ y_1\ \gamma_1\ x_2\ y_2\ \gamma_2\ \ldots\ ]^T \in \mathbb{R}^{3N}$. The measurements in the ADOA model are insensitive to translations, rotations, and scaling of the sensor coordinates. Accounting for the augmented parameter vector, the corresponding basis of null(J) is

$$\boldsymbol{v}_x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \boldsymbol{v}_y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}, \boldsymbol{v}_\phi = \begin{bmatrix} -(y_1 - \bar{y}) \\ x_1 - \bar{x} \\ 1 \\ -(y_2 - \bar{y}) \\ x_2 - \bar{x} \\ 1 \\ \vdots \end{bmatrix}, \boldsymbol{v}_s = \begin{bmatrix} x_1 - \bar{x} \\ y_1 - \bar{y} \\ 0 \\ x_2 - \bar{x} \\ y_2 - \bar{y} \\ 0 \\ \vdots \end{bmatrix} \tag{3.37}$$

3.3.3 Expected Error

For an unbiased estimator $\hat{\boldsymbol{\theta}}$, we may express the expected values of the three estimation errors ϵ, $\tilde{\epsilon}_r$ and $\tilde{\epsilon}_t$ in terms of the estimator covariance matrix $\Sigma_{\hat{\boldsymbol{\theta}}} = E[\boldsymbol{\xi}\boldsymbol{\xi}^T]$. Let

$$\Sigma_t = E[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T] = W^T \Sigma_{\hat{\boldsymbol{\theta}}} W \tag{3.38}$$

denote the covariance matrix of the transformation coefficients, and let

$$\Sigma_r = E[\tilde{\boldsymbol{w}}_r \tilde{\boldsymbol{w}}_r^T] = (\widetilde{W}\widetilde{W}^T)\Sigma_{\hat{\boldsymbol{\theta}}}(\widetilde{W}\widetilde{W}^T) \tag{3.39}$$

denote the covariance matrix of the error in the relative subspace $\mathcal{R}(W)^\perp$, where the columns of $\widetilde{W}$ form an orthonormal basis for $\mathcal{R}(W)^\perp$. The expected errors are

$$e \triangleq E[\epsilon] = E[\boldsymbol{\xi}^T\boldsymbol{\xi}] = \operatorname{tr}\Sigma_{\hat{\boldsymbol{\theta}}} \tag{3.40}$$

$$e_r \triangleq E[\tilde{\epsilon}_r] = E[\tilde{\boldsymbol{w}}_r^T\tilde{\boldsymbol{w}}_r] = \operatorname{tr}\Sigma_r \tag{3.41}$$

$$e_t \triangleq E[\tilde{\epsilon}_t] = E[\tilde{\boldsymbol{w}}_t^T\tilde{\boldsymbol{w}}_t] = \operatorname{tr}\Sigma_t, \tag{3.42}$$

and, as desired, the sum of the mean component errors equals the total: $e = \operatorname{tr}[\widetilde{W}W]^T\ \Sigma_{\hat{\boldsymbol{\theta}}}\ [\widetilde{W}W] = e_r + e_t$. Lower bounds on the expected errors e, e_r, and e_t maybe be obtained from the constrained CRB as described in the next section.

3.4 Constrained Estimation

3.4.1 Constrained CRB

Because the FIM J in (3.25) is singular, the Cramér-Rao bound based on measurements alone does not exist and additional information must be supplied in order to regularize the self-localization problem. This information may be in the form of

a prior distribution on the parameters $p(\boldsymbol{\theta})$, or, as we consider, a set of parametric constraints

$$\boldsymbol{f}(\boldsymbol{\theta}) = \mathbf{0}. \tag{3.43}$$

In general, $\boldsymbol{f}(\boldsymbol{\theta})$ is a k-vector representing a system of k constraints. For example, to constrain the centroid ($k = 2$) to the origin $(0, 0)$, we have

$$\boldsymbol{f}(\boldsymbol{\theta}) = \frac{1}{N}[\boldsymbol{v}_x \ \boldsymbol{v}_y]^T\boldsymbol{\theta}. \tag{3.44}$$

The constrained CRB Σ_c bounds the covariance matrix of an unbiased estimator $\hat{\boldsymbol{\theta}}$ satisfying a constraint $\boldsymbol{f}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$,

$$\Sigma_{\hat{\boldsymbol{\theta}}} = E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \geq \Sigma_c, \tag{3.45}$$

and is given in [43] as

$$\Sigma_c = U_c(U_c^T J U_c)^{-1}U_c^T, \tag{3.46}$$

where U_c is a semiunitary matrix whose columns form an orthonormal basis for the nullspace of the Jacobian matrix $F = \frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$, $FU_c = 0$. We assume that the constraints $\boldsymbol{f}(\boldsymbol{\theta})$ are non-degenerate such that the inverse in (3.46) exists. For a given constraint function $\boldsymbol{f}(\boldsymbol{\theta})$ and noise distribution $p_\eta(\boldsymbol{\eta})$, lower bounds on the expected errors e, e_r and e_t may be obtained by substituting the constrained CRB Σ_c in (3.46) for $\Sigma_{\hat{\theta}}$ in (3.40), (3.41), and (3.42), respectively.

Next, we derive upper and lower bounds on the minimum total error, $e = \operatorname{tr}\Sigma_c$, for general estimation of an n-vector under parametric constraints (3.43) and degenerate measurements. In the localization context $n = 2N$ for TOA, AOA, and distance measurements; and $n = 3N$ for TDOA and ADOA. As the measurements are assumed

degenerate, $r = \text{rank}(J) < n$, and the FIM may be written

$$J = [U_J \widetilde{U}_J] \begin{bmatrix} \Lambda_J & 0 \\ 0 & 0 \end{bmatrix} [U_J \widetilde{U}_J]^T \tag{3.47}$$

$$= U_J \Lambda_J U_J^T. \tag{3.48}$$

Let $d = n - r$ denote the rank deficiency of J. For localization, we have already defined a meaningful basis of $\mathcal{R}(\widetilde{U}_J)$ given by the d columns of W.

3.4.2 Error Bounds for a k-constrained System, $k \geq d$

In this section we bound the minimum of the total error e in the constrained estimation problem using information from the unconstrained FIM, J, and properties relating the two subspaces $\mathcal{R}(U_J)$ and $\mathcal{R}(U_c)$. To do so, we first present two properties concerning the singular values of the product of two matricies which will be used later.

Let $A, B \in \mathbb{R}^{n \times n}$ be positive semidefinite Hermitian matrices with ordered singular values $\sigma_1(A) \geq \cdots \geq \sigma_n(A)$, $\sigma_1(B) \geq \cdots \geq \sigma_n(B)$. The following are true:

Property 1 ([62, Th. 3.3.14])**.**

$$\sum_{i=1}^{j} \sigma_i(AB) \leq \sum_{i=1}^{j} \sigma_i(A)\sigma_i(B), \quad 1 \leq j \leq n \tag{3.49}$$

Property 2 ([63])**.**

$$\sum_{i=1}^{j} \sigma_i(AB) \geq \sum_{i=1}^{j} \sigma_i(A)\sigma_{n-i+1}(B), \quad 1 \leq j \leq n \tag{3.50}$$

Letting $j = n$ in the above properties, we obtain the following upper and lower bounds on the trace of AB

$$\sum_{i=1}^{n} \sigma_i(A)\sigma_{n-i+1}(B) \leq \text{tr}\, AB \leq \sum_{i=1}^{n} \sigma_i(A)\sigma_i(B). \tag{3.51}$$

Now we write Σ_c as

$$\Sigma_c = U_c(A^T \Lambda_J A)^{-1} U_c^T, \tag{3.52}$$

where

$$A = U_J^T U_c, \tag{3.53}$$

and consider its SVD

$$A = [Y_1 Y_2] \begin{bmatrix} \Lambda_A \\ 0 \end{bmatrix} Z^T, \tag{3.54}$$

where $Y_1 \in \mathbb{R}^{(n-d)\times(n-k)}$, $Y_2 \in \mathbb{R}^{(n-d)\times(k-d)}$, and $\Lambda_A = \mathrm{diag}([\sigma_1(A), \ldots, \sigma_{n-k}(A)])$. The singular values, $\sigma_1(A) \geq \cdots \geq \sigma_{n-k}(A)$, correspond to the cosines of the principal angles, $0 \leq \phi_1 \leq \cdots \leq \phi_{n-k} \leq \pi/2$, between the two subspaces $\mathcal{R}(U_c)$ and $\mathcal{R}(U_J)$, [56, Ch. 12]

$$\cos \phi_i = \sigma_i(A). \tag{3.55}$$

See also Section 4.2.1 of Chapter 4 for more information about principal angles.

It follows that the minimum total localization error is given by:

$$\begin{aligned} \mathrm{tr}\, \Sigma_c &= \mathrm{tr}(A^T \Lambda_J A)^{-1} \\ &= \mathrm{tr}(Z \Lambda_A Y_1^T \Lambda_J Y_1 \Lambda_A Z^T)^{-1} \\ &= \mathrm{tr}\, Z^{-1} Z (\Lambda_A Y_1^T \Lambda_J Y_1 \Lambda_A)^{-1} \quad (Z \text{ is unitary}) \\ &= \mathrm{tr}(\Lambda_A Y_1^T \Lambda_J Y_1 \Lambda_A)^{-1} \\ &= \mathrm{tr}\, \Lambda_A^{-2} (Y_1^T \Lambda_J Y_1)^{-1}. \end{aligned} \tag{3.56}$$

Let $X = (Y_1^T \Lambda_J Y_1)$ and assume the singular values are ordered $\sigma_1(X) \geq \cdots \geq \sigma_{n-k}(X)$. Because Y_1 is semiunitary, the singular values of X are bounded according

to the Poincaré separation theorem [64, Corollary 4.3.16]:

$$\sigma_i(\Lambda_J) = \sigma_i(J) \leq \sigma_i(X) \leq \sigma_{i+k-d}(J), \quad i = 1, \ldots, (n-k). \tag{3.57}$$

Combining (3.51), (3.55), (3.56), and (3.57) we have the following bound on total error.

Theorem 1 (Bounds on minimum total error)**.**

$$\sum_{i=1}^{n-k} \frac{1}{\cos^2 \phi_i} \frac{1}{\sigma_{n-k+1-i}(J)} \leq \operatorname{tr} \Sigma_c \leq \sum_{i=1}^{n-k} \frac{1}{\cos^2 \phi_i} \frac{1}{\sigma_{k-d+i}(J)}. \tag{3.58}$$

The lower bound consists of a weighted sum of the reciprocals of the largest $(n-k)$ singular values of J, and the upper bound uses the reciprocals of the $(n-k)$ smallest nonzero singular values of J. The weightings are based on the principal angles between $\mathcal{R}(U_c)$ and $\mathcal{R}(U_J)$. A geometric interpretation of (3.58) is provided in Section 3.4.3 below. Note that (3.58) suggests infinite estimation error when any of the principal angles are 90°, although, in practice sensor locations may be known to lie in a bounded region. This discrepancy is present in all CRB analyses and arises because the CRB is inherently a local bound—ignoring global structure information that may be available. Local and global bounds are further considered in Section 3.4.5.

Owing to the minimum and maximum principal angles, (3.58) may be simplified as the following *weaker* bounds

$$\frac{1}{\cos^2 \phi_1} \sum_{i=1}^{n-k} \frac{1}{\sigma_{n-k+1-i}(J)} \leq \operatorname{tr} \Sigma_c \leq \frac{1}{\cos^2 \phi_{n-k}} \sum_{i=1}^{n-k} \frac{1}{\sigma_{k-d+i}(J)}. \tag{3.59}$$

A simple corollary of Theorem 1 follows for a minimally constrained ($k = d$) system.

Corollary 1 (Minimally constrained system, total error)**.**

$$\frac{1}{\cos^2 \phi_1} \operatorname{tr} J^\dagger \leq \operatorname{tr} \Sigma_c \leq \frac{1}{\cos^2 \phi_{n-d}} \operatorname{tr} J^\dagger. \tag{3.60}$$

Proof. The proof follows immediately from (3.59) after noting that $\sum_{i=1}^{n-d} 1/\sigma_i(J) = \operatorname{tr} J^\dagger$. □

Further, when $k = d$ and the subspaces $\mathcal{R}(U_c)$ and $\mathcal{R}(U_J)$ are identical, then $\phi_i = 0, \forall\, i \in \{1, \ldots, n-d\}$, and the upper and lower bounds converge yielding

$$\operatorname{tr} \Sigma_c = \operatorname{tr} J^\dagger. \tag{3.61}$$

Theorem 2 (Bound on relative error)**.** *For any system of constraints, we have the following bound for the covariance matrix of the relative error*

$$\Sigma_r \leq J^\dagger. \tag{3.62}$$

Proof. To show (3.62), we must show

$$U_J U_J^T U_c (U_c^T J U_c)^{-1} U_c^T U_J U_J^T \leq U_J \Lambda_J^{-1} U_J^T. \tag{3.63}$$

After substituting $J = U_J \Lambda_J U_J^T$, $U_J^T U_c = Y_1 \Lambda_A Z^T$ and canceling appropriate factors, the claim (3.62) is equivalently reduced to

$$U_J Y_1 (Y_1^T \Lambda_J Y_1)^{-1} Y_1^T U_J^T \leq U_J \Lambda_J^{-1} U_J^T. \tag{3.64}$$

Let $Q = [Y_1 Y_2]^T \Lambda_J [Y_1 Y_2]$, which has $Q_{11} = Y_1^T \Lambda_J Y_1$ as a principal submatrix. Using the fact [64, Th. 7.7.8] that $Q_{11}^{-1} \leq (Q^{-1})_{11}$, where the right-hand side is the equivalent principal submatrix of Q^{-1}, we have

$$(Y_1^T \Lambda_J Y_1)^{-1} \leq Y_1^T \Lambda_J^{-1} Y_1 \tag{3.65}$$

and therefore

$$Y_1(Y_1\Lambda_J Y_1^T)^{-1}Y_1^T \leq Y_1 Y_1^T \Lambda_J^{-1} Y_1 Y_1^T \tag{3.66}$$

$$\leq \Lambda_J^{-1}, \tag{3.67}$$

where (3.67) follows because $Y_1 Y_1^T$ is a projection matrix. Multiplying (3.67) on the left by U_J and on the right by U_J^T yields (3.64) and completes the proof of (3.62). □

Finally, we observe in the following corollary to Theorem 2 that the upper bound $J^\dagger$ on relative error is achieved for any minimally constrained system.

Corollary 2 (Minimally constrained system, relative error)**.** *For a minimally constrained ($k = d$) system, we have for the relative error*

$$\Sigma_r = J^\dagger, \tag{3.68}$$

and therefore $e_r = \operatorname{tr} J^\dagger$.

Proof. We write $\Sigma_r = U_J A(A^T\Lambda_J A)^{-1}A^T U_J^T$, with A as in (3.53). When $k = d$, A is square and invertible yielding

$$\Sigma_r = U_J(AA^{-1})\Lambda_J^{-1}(A^{T^{-1}}A^T)U_J^T \tag{3.69}$$

$$= U_J\Lambda_J^{-1}U_J^T = J^\dagger \tag{3.70}$$

for any system of $k = d$ non-degenerate constraints. □

The interpretation of Theorem 2 is that measurements alone provide a certain amount of information about the relative scene configuration. This establishes the upper bound $J^\dagger$ on the covariance matrix of the relative error. Any additional information, in the form of parametric constraints, can only reduce the relative error

below $J^\dagger$. Corollary 2 says that when we have the minimum number of constraints, equal to the rank deficiency d of J, then the relative error is still equal to $J^\dagger$. In effect, a minimally constrained system only serves to provide information about unknown transformation components which are not informed upon by measurements; relative error is not decreased. We use this fact in Section 3.5.1 in order to construct a relative estimation algorithm.

3.4.3 Geometric Interpretations

Examining (3.58) we can see how the interplay between the two information sources – constraints and measurements – influences the total estimation error. Considering a linearization of the constraint $\boldsymbol{f}(\hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}$,

$$\boldsymbol{f}(\hat{\boldsymbol{\theta}}) \approx \boldsymbol{f}(\boldsymbol{\theta}) + F\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0} \tag{3.71}$$

$$\implies F\hat{\boldsymbol{\theta}} \approx F\boldsymbol{\theta} - \boldsymbol{f}(\boldsymbol{\theta}), \tag{3.72}$$

we see that, for $\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}$, the constraint function precisely determines $\hat{\boldsymbol{\theta}}$ in the k-dimensional subspace $\mathcal{R}(F^T)$ but says nothing about the components of $\hat{\boldsymbol{\theta}}$ in $\mathcal{R}(F^T)^\perp = \text{null}(F) \triangleq \mathcal{R}(U_c)$. We call $\mathcal{R}(F^T)$ and $\mathcal{R}(U_c)$ the constrained and unconstrained subspaces under constraint $\boldsymbol{f}$, respectively.

The parameter space $\mathbb{R}^n$ may also be partitioned from the measurements into $\mathcal{R}(U_J)$ and $\mathcal{R}(\widetilde{U}_J)$, where $\mathcal{R}(U_J)$ represents the subspace informed upon by measurements, and $\mathcal{R}(\widetilde{U}_J) = \mathcal{R}(W)$ represents the transformation subspace which is not estimable from measurements.

When the unconstrained subspace $\mathcal{R}(U_c)$ is closely aligned with the measurement subspace $\mathcal{R}(U_J)$, all of the principal angles (3.55) are small (i.e., $\cos\phi_i \approx 1$) and the localization estimation performance, from (3.58), is good.

3.4.4 The Pseudoinverse Bound, $J^\dagger$

If, for a minimally constrained system, the $k = d$ constraints precisely determine the components of $\hat{\boldsymbol{\theta}}$ in the transformation subspace $\mathcal{R}(\widetilde{U}_J)$, then $\mathcal{R}(U_c) = \mathcal{R}(U_J)$ and we may write $U_c = U_J B$, for some non-singular matrix B. As $\mathcal{R}(\widetilde{U}_J) = \mathcal{R}(W)$, this corresponds to a constraint that fully specifies the unknown transformation parameters. (For example, with distance measurements, the constraint would uniquely specify the scene centroid and rotation.) In this case, the total error CRB from (3.46) may be rewritten using the pseudo-inverse

$$\begin{aligned} \Sigma_c &= U_c (B^T \Lambda_J B)^{-1} U_c^T \\ &= U_J \Lambda_J^{-1} U_J^T = J^\dagger. \end{aligned} \tag{3.73}$$

Because we have assumed precise specification of the transformation parameters, the transformation portion of the total error is zero, meaning that all of $E_c = J^\dagger$ comes from relative error. This is consistent with Corollary 2.

In the localization context, the pseudo-inverse $J^\dagger$ was considered in [65] and referred to as the relative CRB, and later in [41] being called the anchor-free CRB. The interpretation of $J^\dagger$ is that it bounds localization error in the relative domain due to measurements alone; absolute positioning error in the transformation subspace is ignored. This chapter generalizes the relative CRB concept to the case of general constraints, and provides a geometric understanding of the subspaces involved.

In Section 3.5 we demonstrate the use of the relative CRB in evaluating relative estimators and apply the decomposition ideas to absolute localization algorithms.

3.4.5 Local and Global Error Bounds

Bounds on parameter estimation error may be classified as either local or global. Local bounds only capture small deviations about the true parameter value $\boldsymbol{\theta}$. The CRB is a local bound as it is based solely on the local structure of the likelihood function. Global bounds, which consider estimation performance over the entire parameter range, may be larger or smaller than a local bound. For example, multi-modal likelihood functions—such as those encountered in time-delay-estimation [66, 67]—frequently yield global uncertainty much greater than local uncertainty, especially at low SNRs; while problems with bounded parameter ranges may have less global (total) uncertainty than that predicted by a local bound, which ignores the finite parameter range. Although sometimes difficult to evaluate, global bounds, such as the Weiss-Weinstein bound [68], the Ziv-Zakai bound [69] or the Bayesian Cramér-Rao bound [40, p. 84] [70], may be computed for localization problems and would represent a generalization of the deterministic Cramér-Rao analyses considered in this dissertation.

However, our primary interest lies in the decomposition of parameter estimation error into the transformation (singular) and relative (non-singular) domains, and this decomposition—which relies on the tangent plane approximation of the transformation manifold (Section 3.2.2)—applies generally, to both local and global bounds.

3.5 Examples

3.5.1 Relative Estimators

Relative-only localization algorithms only estimate the relative configuration of the sensors from inter-node measurements; absolute position information is not estimated, but rather assigned arbitrarily. Examples include classical multidimensional scaling (MDS) [12, Ch. 12] and Isomap [13] which place the centroid of relative scene estimates at the origin. As previously noted, the FIM based on measurements alone is singular and cannot directly be used to bound location estimation performance. However, relative localization algorithms may be compared to the relative CRB based on the pseudo-inverse of J (3.73). This bound ignores absolute positioning information which is arbitrarily specified by the relative localization algorithm. An alternative to using the relative CRB is to choose a subset of anchor nodes in the network, and then to translate, rotate, and scale the relative solution for maximal agreement with the anchor subset. These resulting estimates may then be compared to a constrained CRB on total estimation error. The drawback of this latter approach is that the total error depends not only on the relative estimates, but also on the alignment process and the particular anchors selected. Hence, the advantage of the relative CRB is that we may directly evaluate the quality of relative estimates without resorting to arbitrary anchor node selection. We illustrate the procedure for both distance and angle measurements.

Distance Measurements

We consider localization of the network depicted in Figure 3.2 where we assume all nodes make distance measurements to one another according to (3.24) and (3.31) with

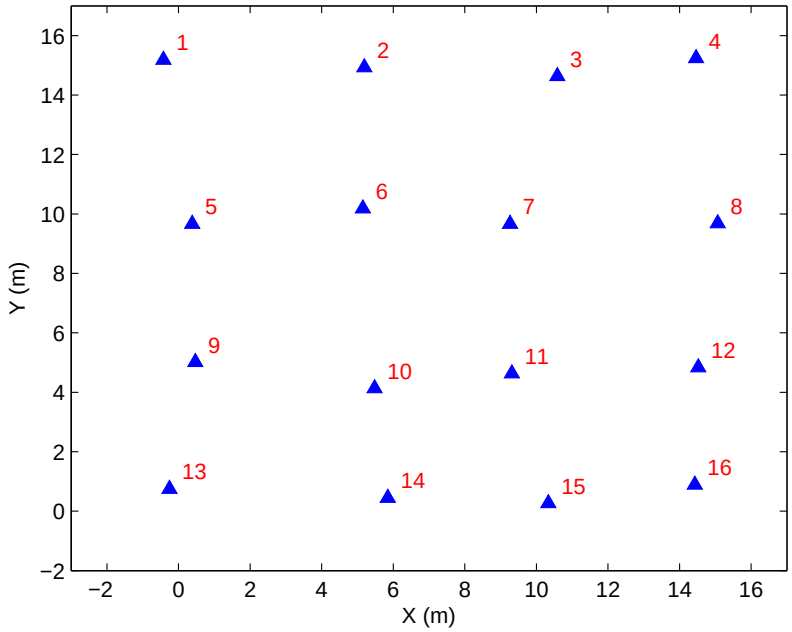


Figure 3.2: Sample network in a crude grid configuration used for localization examples.

Gaussian measurement noise $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_d^2 I)$. Let $d_{i,j}$ denote the measured distance between nodes i and j, and denote by $\hat{d}_{i,j}(\boldsymbol{\theta})$ the resulting inter-node distance for potential parameter vector $\boldsymbol{\theta} = [x_1 \ y_1 \ \dots \ x_N \ y_N]^T$. We consider estimators of the form

$$\hat{\boldsymbol{\theta}}_p = \arg\min_{\boldsymbol{\theta} \setminus \mathcal{C}} \sum_{i \neq j} [d_{i,j}^p - \hat{d}_{i,j}^p(\boldsymbol{\theta})]^2, \quad p = 1, 2. \tag{3.74}$$

When $p = 1$, (3.74) is the maximum likelihood estimator, and we consider $p = 2$ because this cost function affords closed-form expressions for its gradient and Hessian, thereby simplifying the optimization process [71, §8.5.3].

Without restriction, the optimization (3.74) will not converge and we must, therefore, "pin down" the relative scene. We arbitrarily assign $x_1 = y_1 = x_2 = 0$; therefore, $\mathcal{C} = \{x_1, y_1, x_2\}$ and the optimization is over $\boldsymbol{\theta} \setminus \mathcal{C}$. From Corollary 2, we know that utilizing the minimum number of constraints (three in this case) does not introduce any relative information into the problem. Therefore, while (3.74) could be used for absolute localization, we may also use it to derive relative scene estimates. We also consider relative estimates derived from MDS.

Figure 3.3 demonstrates the relative localization error of these three algorithms as a function of σ_d and compares them to the relative CRB (3.73). The relative errors for each estimator were calculated by averaging 200 simulation values of ϵ_r (3.6) and the relative CRB error is the trace of $J^\dagger$. The performance of the relative MLE, $\hat{\boldsymbol{\theta}}_1$, is well-described by the relative CRB for the noise range considered. The performance of $\hat{\boldsymbol{\theta}}_2$ is also reasonably close the relative CRB, while MDS, which has much lower computational complexity, has significantly larger error.

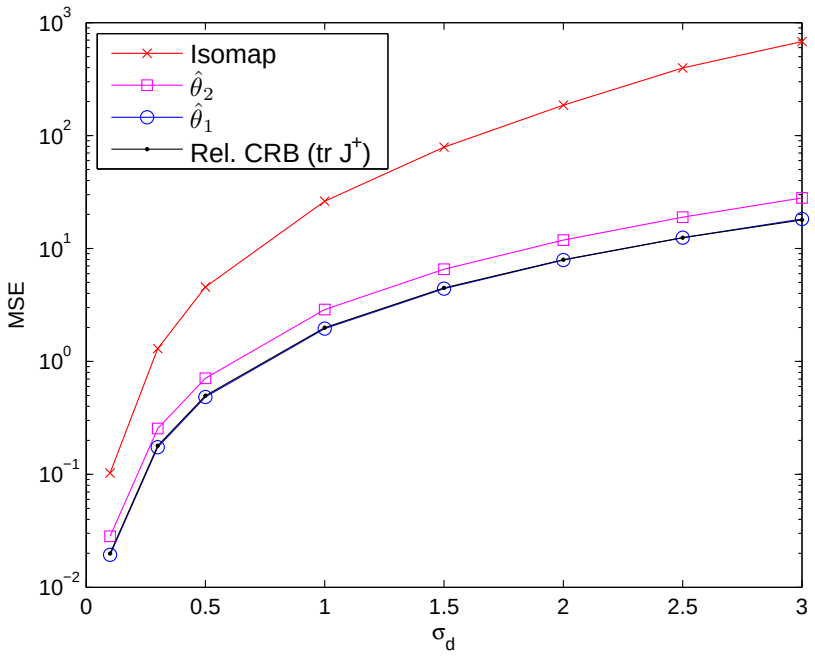


Figure 3.3: Performance evaluation of distance-based relative localization algorithms compared to the relative CRB. The relative MLE $\hat{\boldsymbol{\theta}}_1$ achieves the relative CRB for the noise range considered.

Angle Measurements

We again consider the localization of the network in Figure 3.2, although now based on AOA measurements. We assume that all sensors make measurements to one another and that the measurements are equal to true arrival angles corrupted by Gaussian measurement noise $\mathcal{N}(0, \sigma_{aoa}^2)$. Similar to the distance measurement case, we consider a relative MLE, $\hat{\boldsymbol{\theta}}_{MLE}$, obtained from a minimal constraint $x_1 = y_1 = x_2 = 0$. We also consider a closed-form low-complexity relative algorithm called RAST, $\hat{\boldsymbol{\theta}}_{RAST}$, [72]. The singular FIM J for angle measurements was computed (as in [32]), and the relative CRB $J^\dagger$ calculated. The performance of the two relative estimators is compared to the relative error bound, $\operatorname{tr} J^\dagger$, in Figure 3.4. The relative MLE nearly achieves the relative CRB for the range of σ_{aoa} considered while the localization error of the RAST algorithm is somewhat greater.

3.5.2 Decomposition of Absolute Estimators

In this subsection we consider absolute localization algorithms and the decomposition of their performance in the relative and transformation domains. To achieve absolute localization we utilize anchor nodes, which provide a particular, but convenient, type of the more general constraint (3.43). We let $\mathcal{A}$ denote the set of anchor nodes which have *a priori* known locations, and we let the vector $\boldsymbol{\theta}_{\mathcal{A}}$ denote the true values of this subset of parameters. The anchor-based constraint function takes the linear form

$$\boldsymbol{f}(\boldsymbol{\theta}) = H\,\boldsymbol{\theta} - \boldsymbol{\theta}_{\mathcal{A}} = \mathbf{0}, \tag{3.75}$$

where the rows of H correspond to appropriate rows of the identity matrix in order to extract known coordinates from $\boldsymbol{\theta}$.

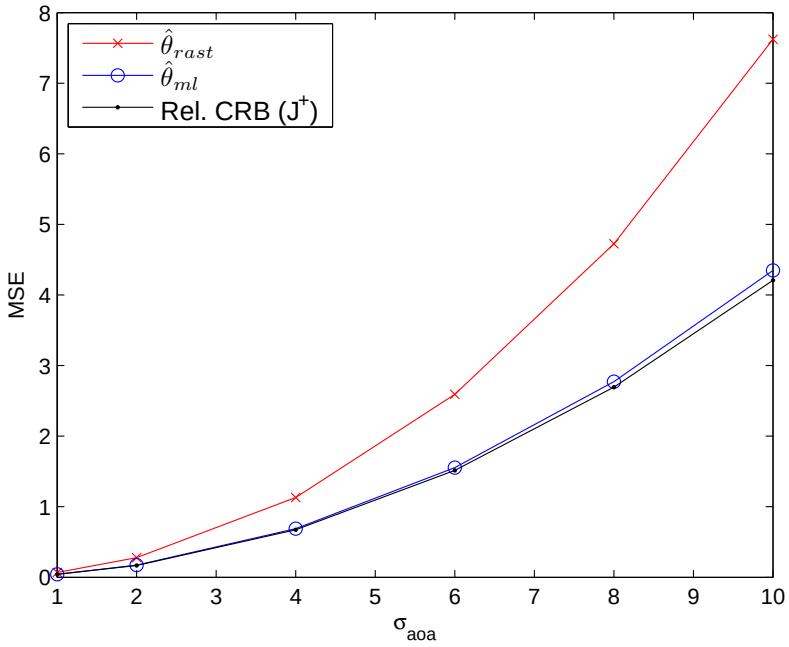


Figure 3.4: Performance evaluation of AOA-based relative localization algorithms compared to the relative CRB. The relative MLE $\hat{\boldsymbol{\theta}}_{MLE}$ nearly achieves the relative CRB for the given noise range.

For anchor set $\mathcal{A} = \{9, 13, 14\}$, Figure 3.5 illustrates a scatter plot of 500 maximum likelihood estimates of absolute positions, where we have again assumed pairwise distance measurements with Gaussian measurement noise ($\sigma_d = 1\,\mathrm{m}$). The elliptical shape of the point clusters indicates obvious correlation between the x and y coordinate estimates for a given node, however, there is also significant correlation across nodes. To demonstrate this, each of the 500 16-node estimates is plotted with a different greyscale intensity. For a given scene estimate, the shading of all 16 nodes was determined by the position of the estimate of node 3 (chosen arbitrarily) relative to the principle axis of cluster 3. This results in the smooth shading seen for cluster 3 in the figure. If the estimates of the other node locations were uncorrelated, their cluster estimates would appear randomly colored. However, the general trend of the shading is seen in the other clusters as well. This suggests that the variability in the *shape* of the estimated scene is actually less than what is implied by the size of the absolute scatter plots. We quantify this using the decomposition ideas of this chapter.

The average empirical total error ϵ (see (3.1)), calculated as the average of ϵ over the 500 simulations, was equal to $5.54\,\mathrm{m}^2$. This equals the CRB-predicted bound on total error given by the constrained CRB, $\mathrm{tr}\,\Sigma_c = 5.54\,\mathrm{m}^2$. The 3-$\sigma$ uncertainty ellipses from Σ_c are also plotted on Figure 3.5. While these ellipses are good predictors of the total error distribution, they do not give a complete picture of estimation error. In particular, they fail to capture the relative error and correlation structure observed above.

For each scene estimate, we determine the optimally transformed relative estimate $\hat{\boldsymbol{\theta}}_r$ as in (3.5) and show the relative scatter plots of $\{\hat{\boldsymbol{\theta}}_r\}$ in Figure 3.6. The average empirical relative error ϵ_r (see (3.6)) was calculated to be $1.78\,\mathrm{m}^2$ which compares

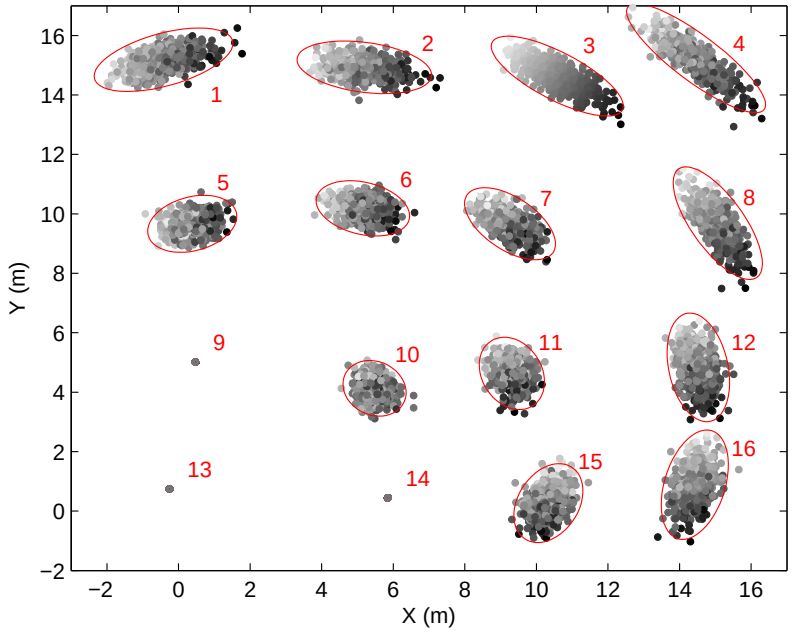


Figure 3.5: Total error: Scatter plots of ML estimates of absolute positions exhibit large rotational uncertainty, as predicted by the 3-σ ellipses of the constrained CRB (–). Color coding of estimates illustrates high correlation between sensors.

favorably to the relative portion of the constrained CRB, $\operatorname{tr} P_W^{\perp} \Sigma_c P_W^{\perp} = 1.77\,\mathrm{m}^2$, where the projection operator $P_W^{\perp}$ projects onto the relative subspace $\mathcal{R}(W)^{\perp}$. We also see in Figure 3.6 that the shape of the relative estimates is well-described by the relative portion of the constrained CRB, $\Sigma_r = P_W^{\perp} \Sigma_c P_W^{\perp}$, and that the relative error is significantly less than the total error – as expected from the shading arguments above. In addition, there is much less correlation of localization error across nodes, as seen by the lack of shading structure in the relative estimates of Figure 3.6. Note that $\Sigma_r \neq J^{\dagger}$ in this case because the number of actual constraints (six) is greater than the minimum number of three.

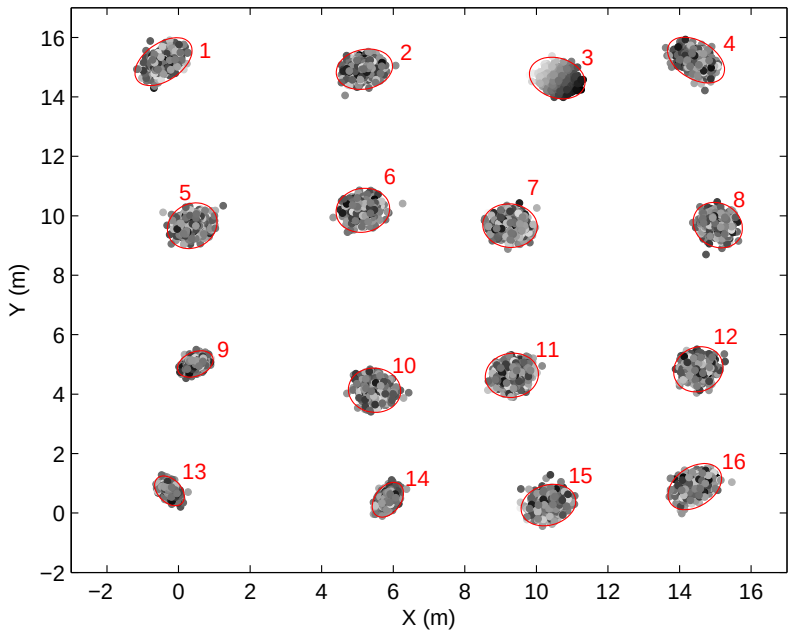


Figure 3.6: Relative error: The large rotational uncertainty of Figure 3.5 is not seen in the optimally transformed relative estimates, $\{\hat{\boldsymbol{\theta}}_r\}$. The 3-$\sigma$ uncertainty ellipses (–) of the relative bound Σ_r accurately describe the empirical relative error.

3.5.3 Anchor Selection

The relatively poor total localization performance in the previous example is due to poor selection of anchor nodes. In this subsection we consider how absolute and relative localization error depend on anchor selection. We assume distance measurements with Gaussian noise ($\sigma_d = 2\,\mathrm{m}$) and consider the selection of three anchor nodes.

While the relative scene estimates in the previous example were good, the total estimation error was poor because of a large rotational uncertainly in the transformation components. Qualitatively, we might expect that the more the anchor nodes are "spread out", the better. Two heuristic selection mechanisms, then, are to choose anchors that 1) cover a maximal area, or 2) have a maximal perimeter. In Figure 3.7 we plot the transformation error e_t, and the relative error e_r, as a function of all possible $\binom{16}{3}$ anchor sets – sorted by decreasing e_t. From the figure we see that different anchor sets have little effect on the relative error but have a dramatic effect on the transformation error. The total absolute localization error is the sum of the relative and transformation components shown in the figure. The optimal anchor set, in the sense of minimum total error, is $\mathcal{A} = \{1, 4, 13\}$ and is illustrated by the vertical bar in Figure 3.7. For this case, the maximum perimeter heuristic yielded the optimal anchor set, and the maximum area heuristic yielded $\mathcal{A} = \{1, 8, 13\}$ with an error only 0.5% greater than the optimal. In general, neither heuristic gives the lowest total error, but both heuristcs result in localization estimates very close to optimal for a large number of example networks that have been considered.

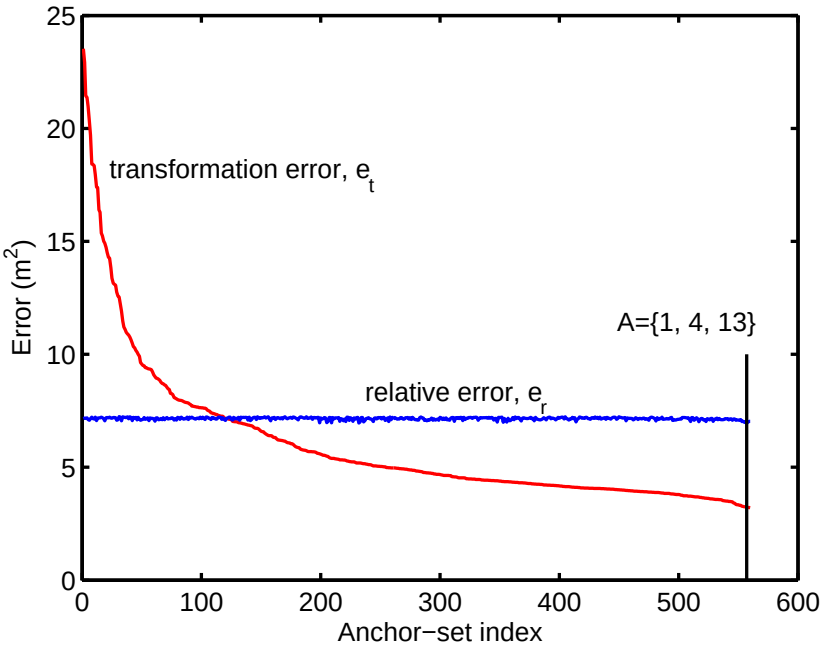


Figure 3.7: The transformation error e_t and relative error e_r exhibit very different responses to alterations in the constraint function—which is achieved by selecting different anchor triplets of the network in Figure 3.2. The total error is the sum of these two curves, $e = e_t + e_r$, and is minimized by the anchor set $\mathcal{A} = \{1, 4, 13\}$.

3.5.4 Application: AOA Estimation with Sensor Uncertainty

In addition to providing insight for understand and improving localization algorithms, relative-absolute error decomposition can also be useful to higher level applications that make use of sensors with position uncertainty. In this example, we demonstrate how transformation uncertainty may be easily incorporated into a Bayesian framework for angle-of-arrival estimation. We assume that the sensor positions are described by the random vector $\boldsymbol{X} = [x_1\, y_1\, \ldots\, x_N\, y_N]^T$ which, as a result of sensor localization, is known to have distribution $p_{\boldsymbol{X}}(\boldsymbol{x})$ with mean $\boldsymbol{X}_0$ and covariance matrix Σ_X. From this sensor array, the goal is to estimate the angle-of-arrival (AOA) ω of a far-field source from a set of time-of-arrival measurements $\boldsymbol{\tau}$ of a signal emanating from that source and measured by the sensors in the network.

This problem is naturally posed in a Bayesian setting where, after the measurement $\boldsymbol{\tau}$, we have the posterior distribution $p(\omega, \boldsymbol{X}|\boldsymbol{\tau})$ from which we wish to obtain the posterior marginal

$$p(\omega|\boldsymbol{\tau}) = \int_{\mathbb{R}^{2N}} p(\omega|\boldsymbol{\tau}, \boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{3.76}$$

as a complete representation of our post-measurement knowledge of the AOA ω. For an N-sensor array, the integral in (3.76) is $2N$-dimensional. Even if the localization algorithm is nearly statistically efficient and the locations $\boldsymbol{X}$ are well-described by the Gaussian $\mathcal{N}(\boldsymbol{X}_0, \Sigma_X)$, the integral remains computationally complex and its computation impractical for a resource constrained sensor network. However, in situations where the transformation uncertainty dominates the relative uncertainty, we may approximate $p(\omega|\boldsymbol{\tau})$ by neglecting relative errors. Focusing on transformation

error alone, we may also neglect unknown array translations as they do not affect far-field AOA estimates.

What remains is the rotational uncertainty of the sensor array. As such, we may approximate the array positions by their nominal value $\boldsymbol{X}_0$ along with a random rotation ϕ. Further, rotating the entire array by an angle ϕ only shifts the posterior distribution; that is, $p(\omega|\boldsymbol{\tau}, \phi) = p(\omega - \phi|\boldsymbol{\tau}, \boldsymbol{X} = \boldsymbol{X}_0)$. Therefore, the marginal in (3.76) may be approximated as

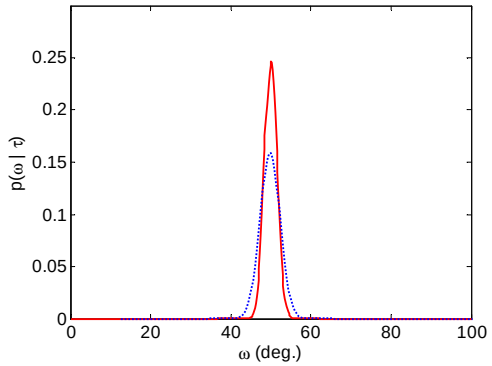
$$p(\omega|\boldsymbol{\tau}) \approx \int_{\mathbb{R}} p(\omega - \phi|\boldsymbol{\tau}, \boldsymbol{X} = \boldsymbol{X}_0) p(\phi) \, \mathrm{d}\phi. \tag{3.77}$$

Comparing the approximation (3.77) to (3.76), we see that the $2N$-dimensional integral has been reduced to a single scalar convolution.

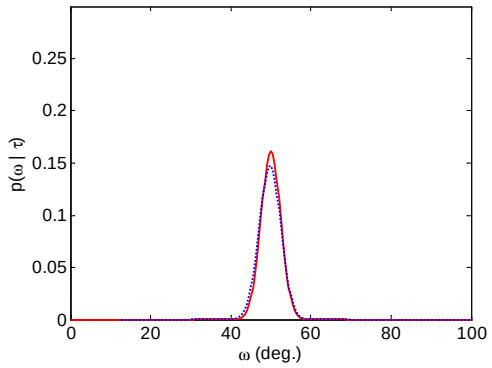
When only $\boldsymbol{X}_0$ and Σ_X are known, we can approximate $p(\phi)$ as a Normal distribution $\phi \sim \mathcal{N}(0, \sigma_\phi^2)$, with variance σ_ϕ^2 easily calculated from the upper left element of the covariance matrix Σ_t of the transformation parameters, $(\Sigma_t)_{1,1} = E[\hat{\beta}_\phi^2]$,

$$\sigma_\phi^2 = \frac{E[\hat{\beta}_\phi^2]}{||\boldsymbol{v}_\phi||^2}. \tag{3.78}$$

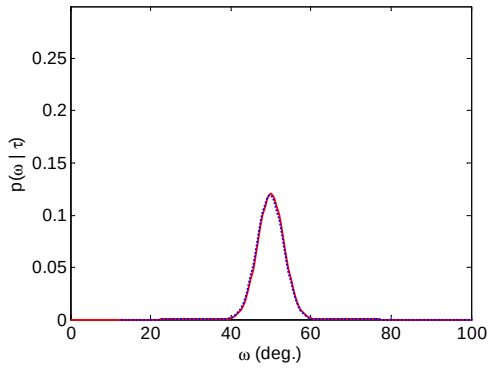
We illustrate the technique using the sensor array of Figure 3.2 with Σ_X given by the CRB described in the previous section and illustrated graphically in Figure 3.5. In this case, (3.78) yields $\sigma_\phi = 2.46°$. We demonstrate using a true source AOA of 50° and assume that the six sensors each measure the arrival time of an acoustic signal with independent Gaussian measurement errors $\mathcal{N}(0, \sigma_t^2)$. In Figure 3.8 we plot the true marginal $p(\omega|\boldsymbol{\tau})$ from (3.76) and compare it to the convolution approximation (3.77) resulting from only considering rotational uncertainty. Three different measurement qualities are considered, $\sigma_t \in \{2\,\mathrm{ms}, 5\,\mathrm{ms}, 10\,\mathrm{ms}\}$. For the $\sigma_t = 10\,\mathrm{ms}$ case (and greater values of σ_t), the approximation is indistinguishable from the true



(a)



(b)



(c)

Figure 3.8: Angle-of-arrival estimation with sensor uncertainty. The true post-measurement distribution $p(\omega|\boldsymbol{\tau})$ of the AOA computed from the $2N$-dimensional integral (3.76) is plotted (—) and compared to the approximation (- -) obtained from the 1-D convolution integral (3.77). Three different measurement qualities are considered; $\sigma_t = 2\,\text{ms}$ (a), $\sigma_t = 5\,\text{ms}$ (b), $\sigma_t = 10\,\text{ms}$ (c). The width of $p(\omega|\boldsymbol{\tau})$ increases as the measurements grow worse, and, for the last case, the approximation is nearly indistinguishable from the true distribution.

marginal. As the measurements improve and σ_t decreases, the approximation begins to break down because, relative to the measurements, sensor uncertainty plays a larger role in the marginal, and the consequences of the approximation are more evident.

3.6 Conclusions

This chapter presented a decomposition of localization error into relative and transformation components. This natural partitioning arose by considering how different sources of information influenced different portions of an absolute localization estimate. In particular, transformation information, which represents the translation, rotation, and scaling in an absolute localization solution, is only informed upon by prior information, such as constraints. Relative information, which represents the "shape" or relative configuration of the sensors, is derived from both measurements and prior information. By considering a linearization of the rigid transformation operator, we demonstrated how a localization error covariance matrix may be decomposed into relative and transformation components. This decomposition may be applied to the error covariance matrix of a particular localization algorithm, the posterior CRB in a Bayesian setting, or a traditional CRB with constraints, as primarily considered in this chapter.

The nullspace of a localization Fisher information matrix was shown to provide a linear subspace approximation of the transformation manifold, where the localization components are non-estimable from measurements alone. For use in the error decomposition, we derived a meaningful basis of the FIM nullspace for a number of measurement types including distance, TOA, AOA, TDOA, and ADOA.

In order to obtain absolute localization estimates, parametric constraints of the from $\boldsymbol{f}(\boldsymbol{\theta}) = \mathbf{0}$ were considered, and in Section 3.4.2 we derived general results (not specific to localization) relating total estimation error in a constrained system to the unconstrained FIM and the principal angles between the measurement subspace $\mathcal{R}(U_J)$ and the unconstrained subspace $\mathcal{R}(U_c)$. The use of generic constraints generalizes the typical use of anchor nodes localization problems. We also proved that localization error in the relative domain is bounded above by the pseudoinverse of the FIM.

The results of this chapter provide additional insight into how different information sources impact different parts of the final localization estimate. There are three general areas which can benefit from this work. The first area involves analytic performance bounds for any relative-only estimators, such as Isomap and RAST, as considered in Example 3.5.1. In this case, the relative CRB (3.73) is the appropriate benchmark. The second area includes the development of new absolute localization algorithms whose performance and design may be analyzed with respect to relative and transformation components independently. In Example 3.5.3 we illustrated how these components behaved very differently with respect to anchor selection. The final area includes applications relying on localization results and their associated uncertainty. Here it may be beneficial to dissect a localization error covariance matrix into transformation error Σ_t (as in (3.38)), and relative error Σ_r (as in (3.39)). For example, if transformation error dominates relative error, a source tracking algorithm may wish to initially ignore sensor uncertainty in order to obtain an initial track relative to the sensors at reduced complexity. Transformation uncertainty Σ_t could subsequently be applied to prescribe translation and rotation variability to the estimated track.

CHAPTER 4

REGULARIZATION OF ABSOLUTE LOCALIZATION USING OPTIMIZATION OF SUBSPACE PRINCIPAL ANGLES

4.1 Introduction

The determination of sensor positions with an absolute reference (e.g., latitude and longitude) is an inherently ill-posed problem when based solely on inter-sensor measurements, such as distances or angles-of-arrival. This is because inter-sensor measurements only depend on the relative configuration of sensors and are invariant to absolute positioning. This results in a singular estimation problem with an accompanying singular Fisher information matrix (FIM) [18, 41, 42].

In order to regularize the absolute localization problem, additional information or assumptions are needed about the network. We could, for example, specify the location of the scene centroid and the angle from the centroid to one of the sensors. In a Bayesian setting, prior distributions on a subset of sensor positions may be used to regularize the problem [65]. As the variance of these priors goes to zero, this is equivalent to precisely specifying the location of a subset of the sensors. Nodes with *a priori* known locations are called anchor nodes (or beacon nodes), and they are commonly used in localization because of their relative ease of implementation.

Depending on how localization is performed, anchor nodes can serve different purposes in the network. In the so-called one-hop localization algorithms [33, 73], unknown-location sensors are in direct contact with anchors and compute their locations based on measurements to multiple anchor nodes. For this method to be effective, it requires a high density of anchor nodes or long-range measurement capabilities—neither of which are common in typical sensor network deployments. In contrast, cooperative networks [32] allow all sensors, non-anchors and anchors, to make measurements to one another. Cooperative networks typically only have a few anchor nodes whose purpose is to "anchor" relative position estimates to an absolute frame of reference.

In this chapter, we consider the use of a small number of anchor nodes as a particular type of parametric constraint in order to regularize absolute localization problems. However, the performance of absolute localization is sensitive to the position of the anchor nodes relative to the remainder of unknown-location nodes. Therefore, we consider means of both selecting and placing anchor nodes in order to minimize mean-square localization error. Conventional wisdom in the localization literature is that anchor nodes should be uniformly spread around the perimeter of the sensor network—observations made empirically by several researchers [44, 45, 46] for multiple algorithms and measurement types. In this chapter we provide analytical justification of this strategy and attempt to elucidate the mechanisms affecting total localization performance under anchor-constraints.

The remainder of this chapter is organized as follows. Before addressing anchor-based regularization, in Section 4.2 we consider optimal parameter selection strategies for estimation problems whose non-identifiable parameters lie in a linear subspace.

This is shown to have combinatoric complexity, and we present a reduced-complexity algorithm for parameter selection called the greedy gap algorithm. In Section 4.3 we apply parameter selection constraints to absolute sensor localization which has the interpretation of anchor node selection. In Section 4.4 we consider the situation where we have control over the placement of anchor nodes and consider strategies for optimal placement. Conclusions and future work are described in Section 4.5

4.2 Parameter Selection Constraints for Linear Non-identifiable Subspaces

In this section we consider a general estimation problem with an n-dimensional parameter vector $\boldsymbol{\theta} \in \mathbb{R}^n$ that has a linear non-identifiable subspace $\mathcal{V} \subset \mathbb{R}^n$; that is, the parameter space $\mathbb{R}^n$ may be partitioned into orthogonal subspaces $\mathcal{V}^\perp$ and $\mathcal{V}$, where the components of $\boldsymbol{\theta}$ in $\mathcal{V}^\perp$ are estimable from data, and those in $\mathcal{V}$ are indeterminate. In problems where the non-identifiable manifold is non-linear, these subspaces may be linear approximations of the manifold taken as tangent planes at particular points of interest. We consider this in the localization context in Sections 4.3 and 4.4.

Because measurements are completely non-informative in the subspace $\mathcal{V}$, it is desirable to impose constraints which are maximally informative in this subspace. Although arbitrary regularizing constraints of the form $\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{c}$, may be possible, we are specifically interested in parameter selection constraints because of their simple nature and applicability to large classes of problems. Parameter selection constraints simply fix (constrain) a subset of the parameters; i.e., for a subset $\Omega \subset \{1 \dots n\}$ constraints take the form $\theta_i = c_i,\ i \in \Omega$, where $\{c_i\}$ are scalar constants. Letting $k = |\Omega|$ denote the number of constraints, we may write the system of constraints in

matrix-vector form

$$C_\Omega^T \boldsymbol{\theta} = \boldsymbol{c}_\Omega, \tag{4.1}$$

where $C_\Omega \in \{0,1\}^{n\times k}$ extracts the constrained parameters from $\boldsymbol{\theta}$, and $\boldsymbol{c}_\Omega$ contains the constraint constants $\{c_i\}$. The i^{th} column of C_Ω is equal to the Ω_i^{th} column of the n–dimensional identity matrix I_n. Because the components of $\boldsymbol{\theta}$ in the k–dimensional subspace $\mathcal{R}(C_\Omega)$ are fully determined by the constraints, we call $\mathcal{C}_\Omega = \mathcal{R}(C_\Omega)$ the *constraint subspace*.

In order to quantify the value of different constraints on estimation performance, we consider the Cramér-Rao bound (CRB) for the remaining unconstrained parameters. As noted in Section 3.4.5, the CRB is a local bound; however, our interest here is in the case where there is no *a priori* global information. All regularization will be supplied by the selection constraints, and we further assume that SNR is sufficiently high for the CRB to be a good predictor of estimation performance.

Let Ω and $\bar{\Omega}$ denote the indicies of the constrained and unconstrained parameters, respectively, and let $\boldsymbol{\theta}_\Omega$ denote the elements of the parameter vector corresponding to index set Ω. The Fisher information matrix J_θ for the entire parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_\Omega, \boldsymbol{\theta}_{\bar{\Omega}}]$ is, by assumption, singular. The Fisher information matrix $J_{\theta_{\bar{\Omega}}}$ for $\boldsymbol{\theta}_{\bar{\Omega}}$ is obtained by eliminating the rows and columns of J_θ corresponding to the indicies Ω.

In a given problem, the FIM $J_{\theta_{\bar{\Omega}}}$ of the unconstrained parameters depends on the model relating parameters to measurements, the values of all parameters $[\boldsymbol{\theta}_\Omega, \boldsymbol{\theta}_{\bar{\Omega}}]$, and the distribution of measurement noise. If all of these quantities were known, we could find the CRB-optimal parameter constraint set Ω^* as

$$\Omega^* = \arg\min_\Omega \operatorname{tr} J_{\theta_{\bar{\Omega}}}^{-1}(\boldsymbol{\theta}_\Omega, \boldsymbol{\theta}_{\bar{\Omega}}). \tag{4.2}$$

Obviously the dependence of Ω^* on the true parameter values makes this approach impractical because the solution to (4.2) requires knowledge of the parameters we seek to estimate. In absence of the true parameter values, we utilize a heuristic technique based on the idea that it is desirable for the constrains to provide as much information as possible about the non-identifiable subspace. A 2D example is shown in Figure 4.1 where measurements $\boldsymbol{y} = A\boldsymbol{\theta} + \boldsymbol{e}$ are taken from an underdetermined system, and e is additive Gaussian noise. In Figure 4.1(a) we assume that the noise is zero, in which case we cannot determine the component of $\boldsymbol{\theta}$ in the nullspace $\mathcal{N}(A)$—illustrated by the diagonal line of Figure 4.1(a). If however, we specify (constrain) $\theta_1 = c_1$, we may precisely determine the value of θ_2; the same holds for specifying $\theta_2 = c_2$ and determining θ_1. When the noise has non-zero variance, however, it is more advantageous to specify θ_1 than θ_2. In Figures 4.1(b) and 4.1(c) we plot a 1σ uncertainty band about the position of the nullspace of A due to the noisy measurement. Clearly, the uncertainty in θ_1 after specifying $\theta_2 = c_2$ (Figure 4.1(b)) is greater than the uncertainty of θ_2 after specifying $\theta_1 = c_1$ (Figure 4.1(c)). When the angle ϕ between the constraint direction and $\mathcal{N}(A)$ is small, the resulting uncertainty in $\boldsymbol{\theta}$ is small. In the example, $\phi_1 < \phi_2$ and constraining θ_1 results in lower final uncertainty in $\boldsymbol{\theta}$.

In general, when the constraint subspace $\mathcal{C}_\Omega$ and non-identifiable subspace $\mathcal{V}$ have larger dimensions, there will be many different angles between their members. If any vector $v \in \mathcal{V}$ is orthogonal to all elements of $\mathcal{C}_\Omega$, then the component of $\boldsymbol{\theta}$ in the direction v remains completely unconstrained and unspecified by measurements. Choosing different parameter sets Ω allows us to control the constraint space $\mathcal{C}_\Omega$. As a

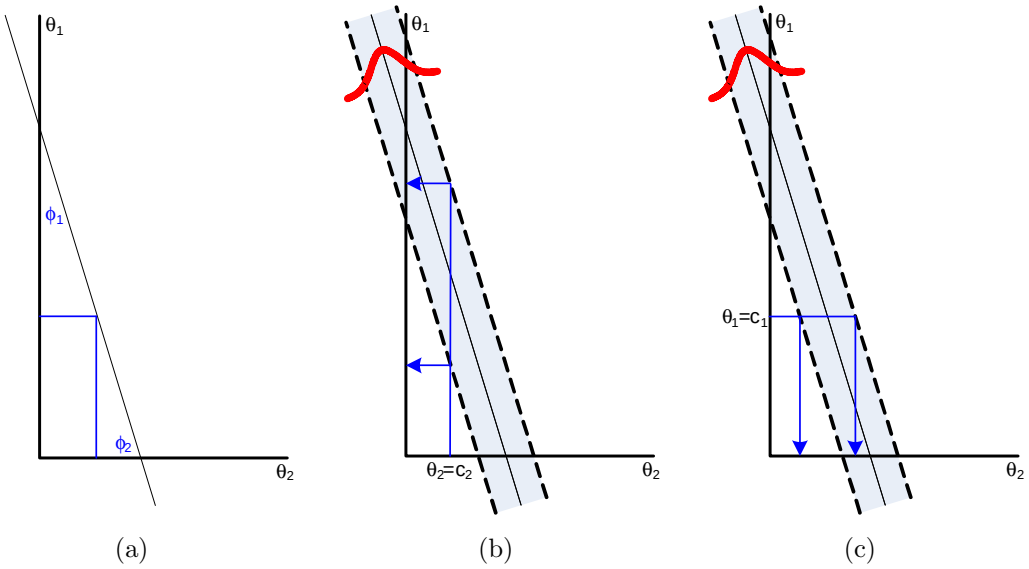


Figure 4.1: Constraint alignment example. (a) In the noiseless case, the underdetermined system $\boldsymbol{y} = A\boldsymbol{\theta}$ only determines $\boldsymbol{\theta}$ to within the nullspace of A, although specifying θ_1 or θ_2 precisely determines the other variable. (b) In the presence of noise, specifying $\theta_2 = c_2$ leaves large uncertainty about θ_1. (c) Constraining $\theta_1 = c_1$ results in lower uncertainty about θ_2 because the constraint axis (θ_1–axis) is more closely aligned with the nullspace of A.

surrogate to the subset optimization problem in Eq. (4.2), we consider the parameter subset which minimizes the maximum angle between $\mathcal{C}_\Omega$ and $\mathcal{V}$.

4.2.1 Principal Angles Between Subspaces

In this subsection we summarize the definition, some properties, and uses of principal angles between subspaces.

Let $\mathcal{A}$ and $\mathcal{B}$ be two subspaces of dimension d_a and d_b respectively, $d_b \geq d_a$. The principal angles $\phi_1, \ldots, \phi_{d_a} \in [0, \pi/2]$ between $\mathcal{A}$ and $\mathcal{B}$ are defined recursively [56,

Ch. 12] as

$$\cos\phi_i = \max_{\substack{\boldsymbol{a}\in\mathcal{A} \\ ||\boldsymbol{a}||=1}} \max_{\substack{\boldsymbol{b}\in\mathcal{B} \\ ||\boldsymbol{b}||=1}} \boldsymbol{a}^T\boldsymbol{b} = \boldsymbol{a}_i^T\boldsymbol{b}_i \tag{4.3}$$

subject to

$$\boldsymbol{a}^T\boldsymbol{b}_j = 0 \quad j \in \{1,\ldots,i-1\} \tag{4.4}$$

$$\boldsymbol{b}^T\boldsymbol{a}_j = 0 \quad j \in \{1,\ldots,i-1\}, \tag{4.5}$$

where $\{a_i\},\{b_i\}$ are known as the principal vectors of $\mathcal{A},\mathcal{B}$. The principal angles satisfy $0 \leq \phi_1 \leq \cdots \leq \phi_{d_a} \leq \pi/2$. Let $\tilde{\mathcal{B}} = \text{span}(\boldsymbol{b}_1 \ \ldots \ \boldsymbol{b}_{d_a})$ denote the d_a-dimensional subspace subspace of $\mathcal{B}$ closest to $\mathcal{A}$, then ϕ_{d_a} is the largest angle between any vector in $\mathcal{A}$ and any vector in $\tilde{\mathcal{B}}$. Further, $\sin\phi_{d_a}$ is known as the "gap" between $\mathcal{A}$ and $\mathcal{B}$ and may alternatively be expressed as [74]

$$\sin\phi_{d_a} = ||P_{\mathcal{A}} - P_{\tilde{\mathcal{B}}}||_2, \tag{4.6}$$

where $P_{\mathcal{X}}$ denotes a projection operator onto subspace $\mathcal{X}$.

The principal angles may be computed as

$$\cos\phi_i = \sigma_i \quad i \in \{1,\ldots,d_a\}, \tag{4.7}$$

where σ_i is the i^{th} largest singular value of A^TB, with orthonormal matrices A and B having column spans equal to $\mathcal{A}$ and $\mathcal{B}$, respectively.

Principal angles and principal vectors have been previously used in statistics in order to determine canonical correlations and canonical variables, respectively, in canonical correlation analysis [75, 76]. Here, our interest is in using principal angles as a measure of alignment between the constraint subspace and the non-identifiable

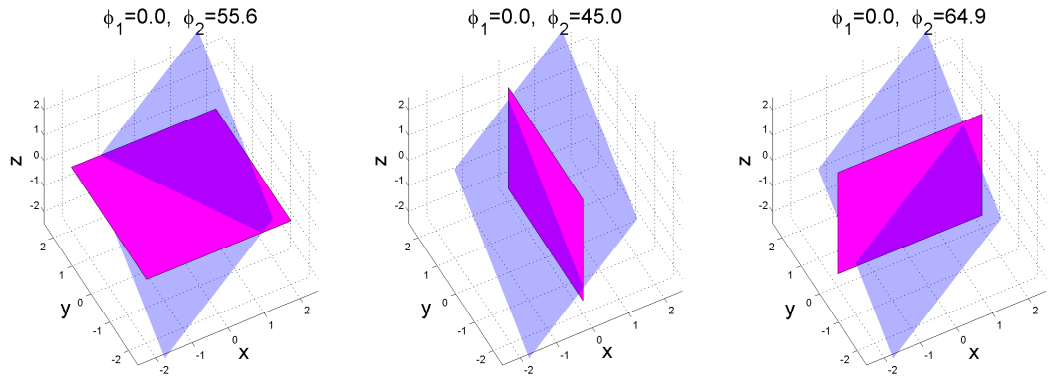
subspace. Let ϕ_Ω denote the set of principal angles between $\mathcal{C}_\Omega$ and $\mathcal{V}$, then as an alternative to (4.2), we take the constrained parameter set to be

$$\hat{\Omega} = \arg \min_{\Omega \in \Pi} \max \phi_\Omega, \tag{4.8}$$

where the minimization is over Π—the set of all possible $\binom{n}{k}$ size-k subsets of $\{1, \ldots, n\}$—and the maximization is taken over the set of principal angles. Importantly, (4.8), unlike (4.2), does *not* depend on the true parameter vectors, so it does not assume *a priori* knowledge of the answer to the problem we wish to solve. Minimizing the maximum principal angle finds the constraint space $\mathcal{C}_{\hat{\Omega}}$ most closely aligned with the indeterminate subspace $\mathcal{V}$. When the maximum angle is $0°$, $\mathcal{V}$ is completely contained in the constraint space. When the maximum angle is $90°$, then at least one component of $\boldsymbol{\theta}$ is indeterminate and the problem remains singular.

Figure 4.2 illustrates the geometric interpretation of this problem: given a d–dimensional subspace $\mathcal{V}$ of interest in $\mathbb{R}^n$, we wish to find the closest k–dimensional ($k \geq d$) elementary subspace. We define an elementary subspace to be one spanned by the canonical basis elements $\{\boldsymbol{e}_i\}$, where $\boldsymbol{e}_i$ is the i^{th} column of I_n. In the example in Figure 4.2, we have $n = 3$, $k = d = 2$, and the candidate elementary subspaces are the xy–plane, the xz–plane, and the yz–plane. The yz–plane has the minimum maximum principal angle in the example.

The optimization (4.8) is an improvement over (4.2) in that it does not depend on the true parameter values, however it remains a discrete optimization problem with combinatoric complexity. In the next section we highlight some similarities of this problem with the problem of sparse signal reconstruction, then in Section 4.2.3 we present a greedy algorithm to find approximate solutions to (4.8) in linear time.



(a) (b) (c)

Figure 4.2: Example of the elementary subspace selection problem in $\mathbb{R}^3$, $n = 3$, $k = m = 2$. Subspace of interest (blue) is compared with the xy–plane (a), yz–plane (b), and xz–plane (c). The principle angles between the two 2D subspaces are given in the plot titles. The closest subspace, with the minimum maximum principle angle, is the yz–plane where $\phi_{max} = \phi_2 = 45.0°$.

4.2.2 Connection to Compressive Sampling

In this section we highlight similarities and dissimilarities between the selection problem (4.8) above and compressive sampling (CS); also known variously as compressive sensing, compressed sensing, sparse reconstruction, and basis pursuit [77, 78]. CS attempts to find the input vector $\boldsymbol{x}$ of the noisy linear system $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{e}$ for the case when the matrix A has more columns than rows. Ordinarily, this is an ill-posed problem; when $\boldsymbol{x}$ is known to be sparse the problem becomes regularized and a unique solution may exist. In the signal processing context, the columns of A typically represent an overcomplete set of signal basis vectors, such as those produced by sinusoids and wavelets. Because $\boldsymbol{x}$ is sparse, the measured signal $\boldsymbol{y}$ consists of a superposition of only a few of these basis elements with additive noise. If $\boldsymbol{x} \in \mathbb{R}^n$ is k-sparse, the difficult part of reconstructing $\boldsymbol{x}$ from $\boldsymbol{y}$ is selecting which k out of n

elements of $\boldsymbol{x}$ are non-zero. Substantial theory and algorithms have been developed to address this problem, see, e.g. [78, 79, 80, 81], and references therein.

In CS, we select a few columns from A whose span is close to a *single* vector $\boldsymbol{y}$. In the parameter selection problem (4.8), we seek a small number of columns from A whose span is close, in some sense, to the span of several vectors $V = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d\}$. Therefore, the parameter selection problem may be interpreted as having multiple right-hand sides: $AX = V$, where the matrix X has sparse columns. A solution $\hat{X}$ minimizes a distance measure $dist(AX, V)$ between AX and V, under a given sparsity constraint on $\hat{X}$.

In an extension to CS, Tropp *et al.* consider simultaneous sparse approximations for multiple right-hand sides using a greedy approach [82] and a convex relaxation approach [83]. However, this work assumes that the right-hand side vectors (V) all approximately span the same 1D space, which is not the case in the parameter selection problem. Tropp also assumes a Frobenius distance $dist(AX, V) = ||AX - V||_F$, whereas (4.8) requires minimization of a spectral norm, as seen from (4.6)

$$\max \phi_\Omega = \sin^{-1} ||P_{\mathcal{C}_\Omega} - P_{\mathcal{V}}||_2 \tag{4.9}$$

and noting that $\sin^{-1}$ does not alter the minimization in (4.8).

4.2.3 The Greedy Gap Algorithm

The minimax optimization (4.8) has combinatoric complexity because the optimization is over a discrete set Π of size $\binom{n}{k}$. In this section we present a greedy algorithm capable of identifying a near-optimal subspace in linear time. The sine of the maximum principal angle between subspaces is sometimes referred to as the gap metric [84], hence we call this algorithm the greedy gap algorithm because it greedily

minimizes the gap between the space of interest and the constructed space. Details of the algorithm follow.

Inputs:

- an $n \times d$ matrix V with columns spanning the subspace of interest
- an $n \times n$ matrix D with orthogonal columns spanning $\mathbb{R}^n$
- a number k, $k \geq d$, indicating the number of columns of D to select in approximating $\mathcal{R}(V)$

Output:

- a set Ω_k, with k indicies corresponding to the columns of D

Procedure:

1. Initialize $\Omega_0 = \emptyset$ as the empty set, set the iteration counter $i = 1$, and let $N = \{1, \ldots, n\}$
2. Find an index ω_i that solves the optimization problem

$$\omega_i = \arg_{\omega} \min_{\omega \in (N \setminus \Omega_{i-1})} \phi_m\Big(\mathcal{R}(V), \mathcal{R}(D_{\Omega_{i-1} \cup \omega})\Big), \tag{4.10}$$

   where $m = \min(d, i)$ is the index of the maximum principal angle between the subspaces considered, and matrix D_β represents the columns of D determined by the set β. Note that (4.10) has low complexity, requiring approximately n computations.

3. Let $\Omega_i = \Omega_{i-1} \cup \{\omega_i\}$
4. Increment i and repeat steps 2 and 3 until $i = k$.

At each step of the algorithm, the method chooses to augment the approximation space $\mathcal{R}(D_{\Omega_i})$ with one of the remaining basis vectors such that the resulting maximum principal angle is minimized. Note that when $i < d$ the size of the set of principal angles is growing with each iteration. This method does not guarantee that the maximum principal angle of step $i+1$ will always be less than the maximum angle of step i. However, a *particular* principal angle will never increase between steps for which it exists.

In the parameter selection application we are interested in canonical subspaces, i.e., $D = I_n$, the identity matrix; however, the greedy gap algorithm does not require this and applies equally well for more complicated subspaces.

4.3 Anchor Node Selection for Absolute Localization

We now turn to regularization of absolute sensor locations as an application of subset constraints as described above. The problem in anchor selection is to identify which sensors should be anchor nodes in order to minimize the resultant mean-square error in estimates of the position vector $\boldsymbol{\theta} = [x_1\, y_1\, \ldots\, x_n\, y_n]$, where n is the total number of sensors, and (x_i, y_i) are the coordinates of the i^{th} sensor. The scenario we consider is that the sensors have been deployed, measurements between a subset of all pairs of sensors have been taken, and a relative localization algorithm has determined the relative shape $\hat{\boldsymbol{\theta}}_r$ of the network. The relative shape estimate is an estimate of the node locations relative to one another, but arbitrarily specifies information about absolute network parameters, such as location, rotation, and scale of the entire sensor scene. From $\hat{\boldsymbol{\theta}}_r$, we wish to select an optimal set of k anchor nodes in order to convert the relative map into absolute positions with minimal error. The actual positions of

the identified anchors are then measured in an absolute coordinate system (by a person with a GPS receiver, for example) and then used in positioning the remaining sensor nodes in an absolute frame of reference. The final absolute scene estimate may use the anchors to prescribe necessary translation, rotation, and scaling to $\hat{\boldsymbol{\theta}}_r$, or it may jointly consider the measurements and anchor information to derive a new final estimate.

The anchor selection problem is more complicated than the problem in the previous section because the non-identifiable space is not linear. As described in Chapter 3, internode measurements only determine sensor positions to within a non-linear manifold representing all possible rotations, translations, and possibly scalings of a relative sensor configuration. We may approximate the non-identifiable manifold at a point by its tangent plane at that point and then apply the parameter selection results of the previous section. However, this implies that the non-identifiable subspace $\mathcal{V}$ is now a function of the manifold parameterization: translation, rotation, and scaling.

For clarity of explanation, we adopt distances as a concrete example for inter-node measurements. At the end of the section we will show that the results hold for other measurement types as well. Distance measurements are invariant to rigid translations and rotations of the sensors, and in Chapter 3 we derived a linear approximation of the space of all translations and rotations of $\boldsymbol{\theta}$ as the span of the three $2n$–vectors

$$\boldsymbol{v}_x = \frac{1}{c_1}\begin{bmatrix}1\\0\\1\\0\\\vdots\end{bmatrix}, \boldsymbol{v}_y = \frac{1}{c_1}\begin{bmatrix}0\\1\\0\\1\\\vdots\end{bmatrix}, \boldsymbol{v}_\phi = \frac{1}{c_2}\begin{bmatrix}-(y_1-\bar{y})\\(x_1-\bar{x})\\-(y_2-\bar{y})\\(x_2-\bar{x})\\\vdots\end{bmatrix}, \tag{4.11}$$

where n is the number of sensors, $\boldsymbol{v}_x$ and $\boldsymbol{v}_y$ represent x– and y–translations, $\boldsymbol{v}_\phi$ corresponds to rotation about the centroid $(\bar{x}, \bar{y})$, and $c_1 = n^{\frac{1}{2}}$ and $c_2 = (\sum_{i=1}^{n}(x_i - \bar{x})^2 + (y_i - \bar{y})^2)^{\frac{1}{2}}$ are normalization constants.

We assume that a relative localization algorithm has supplied us with a relative estimate $\hat{\boldsymbol{\theta}}_r$ of the sensor positions—with the transformation and rotation components arbitrarily specified. Using elements of $\hat{\boldsymbol{\theta}}_r$ and (4.11), we construct $V = [\boldsymbol{v}_x\, \boldsymbol{v}_y\, \boldsymbol{v}_\phi]$ whose span $\mathcal{V}$ forms the non-identifiable subspace approximating the transformation manifold. Note that V does not change with translations of $\hat{\boldsymbol{\theta}}_r$ but does depend on how the points in $\hat{\boldsymbol{\theta}}_r$ are rotated. If we attempt to identify an optimal subset of k parameters (not nodes) using (4.2) or (4.8), we find that the selected parameters depend on the arbitrary rotation of $\hat{\boldsymbol{\theta}}_r$, meaning that the optimal parameter subset depends on the scene's true orientation, which is unknown.

However, if we perform optimal node selection, where parameters x_i and y_i must be selected together as a pair, we find that the optimal set of k nodes is invariant to rotations of $\hat{\boldsymbol{\theta}}_r$. This is not a restrictive condition as most means of measuring a sensor's absolute position give both x and y coordinates. To show this invariance, we first explicitly derive the principal angles between the constraint space and the transformation space and then show that the principal angles are invariant to arbitrary rotations of $\hat{\boldsymbol{\theta}}_r$, for a given set of anchor nodes.

4.3.1 Derivation of Principal Angles Between Constraint Space and Transformation Space

We assume, without loss of generality, that the centroid $(\bar{x}, \bar{y})$ is at the origin $(0, 0)$, that the anchor nodes are identified by the index set A, and that the number of anchors $n_a = |A| \geq 3$, since this is sufficient to disambiguate translations, rotations,

scalings, and mirror images. As such, the dimension $k = 2n_a$ of the anchor-induced constraint space $\mathcal{C}_A$ is at least 6; the dimension of $\mathcal{V}$ is 3; and therefore, the number of principal angles is 3. Let $C_A \in \{0,1\}^{2n \times 2n_a}$ denote the constraint matrix with columns drawn from I_{2n} corresponding to the x- and y-coordinates of the anchor set A. The matrices V and C_A have orthonormal columns spanning $\mathcal{V}$ and $\mathcal{C}_A$, and we seek the singular values of the $3 \times 2n_a$ matrix $Q = V^T C_A$, or the eigenvalues of QQ^T. The matrix Q is evaluated as

$$Q = \begin{bmatrix} \frac{1}{c_1} & 0 & \frac{1}{c_1} & 0 & \dots & \frac{1}{c_1} & 0 \\ 0 & \frac{1}{c_1} & 0 & \frac{1}{c_1} & \dots & 0 & \frac{1}{c_1} \\ \frac{-y_{A_1}}{c_2} & \frac{x_{A_1}}{c_2} & \frac{-y_{A_2}}{c_2} & \frac{x_{A_2}}{c_2} & \dots & \frac{-y_{A_{n_a}}}{c_2} & \frac{x_{A_{n_a}}}{c_2} \end{bmatrix}, \tag{4.12}$$

and

$$QQ^T = \begin{bmatrix} q & 0 & a \\ 0 & q & b \\ a & b & c \end{bmatrix}, \tag{4.13}$$

where

$$q = \frac{n_a}{c_1^2}, \tag{4.14}$$

$$a = \frac{-1}{c_1 c_2} \sum_{i=1}^{n_a} y_{A_i}, \tag{4.15}$$

$$b = \frac{1}{c_1 c_2} \sum_{i=1}^{n_a} x_{A_i}, \tag{4.16}$$

$$c = \frac{1}{c_2^2} \sum_{i=1}^{n_a} (x_{A_i}^2 + y_{A_i}^2). \tag{4.17}$$

The eigenvalues of QQ^T, in order from largest to smallest, are

$$\lambda_1 = \frac{1}{2}(c+q) + \frac{1}{2}\sqrt{(c-q)^2 + 4(a^2+b^2)} \tag{4.18}$$

$$\lambda_2 = q \tag{4.19}$$

$$\lambda_3 = \frac{1}{2}(c+q) - \frac{1}{2}\sqrt{(c-q)^2 + 4(a^2+b^2)}. \tag{4.20}$$

and the principal angles between $\mathcal{V}$ and $\mathcal{C}_A$ are

$$\phi_i = \cos^{-1}\sqrt{\lambda_i} \quad i = 1, 2, 3. \tag{4.21}$$

To show that optimal anchor node selection is rotationally invariant, we show that the eigenvalues $\{\lambda_i\}$, and therefore the principal angles, are invariant to rigid scene rotations. The quantities c, c_1, c_2, and q are clearly invariant to rigid rotations of the points $\{x_i, y_i\}$. To show that all of the eigenvalues are rotationally invariant, it remains to show that $a^2 + b^2$ does not depend on point rotations. Let

$$\boldsymbol{p} = \begin{bmatrix} a \\ b \end{bmatrix} \tag{4.22}$$

$$= \frac{1}{c_1 c_2} R \begin{bmatrix} -y_{A_1} & \cdots & -y_{A_{n_a}} \\ x_{A_1} & \cdots & x_{A_{n_a}} \end{bmatrix} \mathbf{1}_{n_a}, \tag{4.23}$$

where R is a 2×2 rotation matrix and $\mathbf{1}_{n_a}$ is an n_a-vector of ones. Rotations are norm-preserving, hence $a^2+b^2 = ||\boldsymbol{p}||^2$ is constant for any rotation matrix R. Therefore, the eigenvalues—and the principal angles between $\mathcal{C}_A$ and $\mathcal{V}$—do not depend on arbitrary scene orientations.

4.3.2 Other Measurement Types

In Chapter 3 we showed that distances, time-of-arrival, and time-difference-of-arrival measurements all had non-identifiable transformation spaces corresponding to translations and rotations. The structural variations in received-signal-strength measurements are governed by distances as well, so they too have translation and rotation uncertainty. Hence, the results above are applicable to all four of these measurement types.

Angle-of-arrival measurements have a non-identifiable space corresponding to translations and scalings, which from Chapter 3, has a linear approximation given by the

range of $V = [\boldsymbol{v}_x\, \boldsymbol{v}_y\, \boldsymbol{v}_s]$, where

$$\boldsymbol{v}_s = \frac{1}{c_2} \begin{bmatrix} x_1 - \bar{x} \\ y_1 - \bar{y} \\ x_2 - \bar{x} \\ y_2 - \bar{y} \\ \vdots \end{bmatrix} \tag{4.24}$$

and the other quantities are as in (4.11). The eigenvalues for this case are the same as in (4.18)–(4.20) for distance measurements, however now we must show that the principal angles between $\mathcal{C}_A$ and $\mathcal{V}$ do not depend on arbitrary scaling provided by a relative AOA localization algorithm. We consider an arbitrary scaling s which maps $(x_i, y_i) \rightarrow (sx_i, sy_i)$, for all nodes i. It can be shown that all variables (a, b, c, q) making up the eigenvalues are independent of the value s. Hence, node selection based on minimizing principal angles is both translation and scale invariant.

Angle-difference-of-arrival measurements are unchanged with translations, rotations, and scalings, giving us a transformation subspace spanned by $V = [\boldsymbol{v}_x\, \boldsymbol{v}_y\, \boldsymbol{v}_\phi\, \boldsymbol{v}_s]$. In this case, the eigenvalues of QQ^T are

$$\lambda_1, \lambda_2 = \frac{1}{2}(c+q) + \frac{1}{2}\sqrt{(c-q)^2 + 4(a^2+b^2)} \quad \text{(double root)} \tag{4.25}$$

$$\lambda_3, \lambda_4 = \frac{1}{2}(c+q) + \frac{1}{2}\sqrt{(c-q)^2 + 4(a^2+b^2)} \quad \text{(double root)}, \tag{4.26}$$

which, from arguments above, are invariant to arbitrary translation, rotations, and scalings.

4.3.3 Anchor Selection Using the Greedy Gap Algorithm

Although the transformation subspace $\mathcal{V}$ depends on the true parameter vector, we have demonstrated above that if we constrain by nodes—not merely parameters—then a *given* constraint subspace is equi-distant (in the sense of principal angles)

from all rigid and scaled transformations that can be generated from a given relative configuration $\hat{\boldsymbol{\theta}}_r$. As such, we may use the output of a relative localization algorithm, which prescribes arbitrary scale and absolute positioning, to form V and then seek a set of anchor nodes A whose constraint subspace $\mathcal{R}(C_A)$ is most closely aligned with $\mathcal{R}(V)$. The columns of V are a subset of $\{\boldsymbol{v}_x, \boldsymbol{v}_y, \boldsymbol{v}_\phi, \boldsymbol{v}_s\}$, as appropriate for the type of measurements. After forming V, the greedy gap algorithm may be employed to identify a set of anchor nodes. The greedy gap algorithm for *node selection* is essentially the same as in Section 4.2.3 for *parameter selection*. The only modification needed is in Step 2, where, instead of considering the best single parameter to add, we find the best node (with corresponding x and y parameters) to add.

Obtaining the CRB-optimal solution (4.2) or the gap-optimal solution (4.8) has combinatoric computational complexity because the number of possible anchor sets is $\binom{n}{n_a}$, whereas, the complexity of the greedy gap approach is only approximately nn_a.

4.3.4 Examples

As a performance metric for different anchor sets A, we consider the root-mean-square (RMS) localization error

$$\Big(\frac{1}{|U|}\sum_{i\in U} E[d_i^2]\Big)^{1/2}, \tag{4.27}$$

where $E[d_i^2]$ is the expected squared distance of sensor i from its estimated position, and $U = \{1 \dots n\} \setminus A$ is the set of unknown-location nodes. Rather than considering the RMS error for a particular estimator, we utilize the lower bound provided by the CRB and define that as the RMS localization error for a particular anchor set A

$$e_{rms}(\boldsymbol{\theta}, A) = \Big(\operatorname{tr} J^{-1}_{\boldsymbol{\theta}_U}(\boldsymbol{\theta}_A, \boldsymbol{\theta}_U)/(n-n_a)\Big)^{1/2}, \tag{4.28}$$

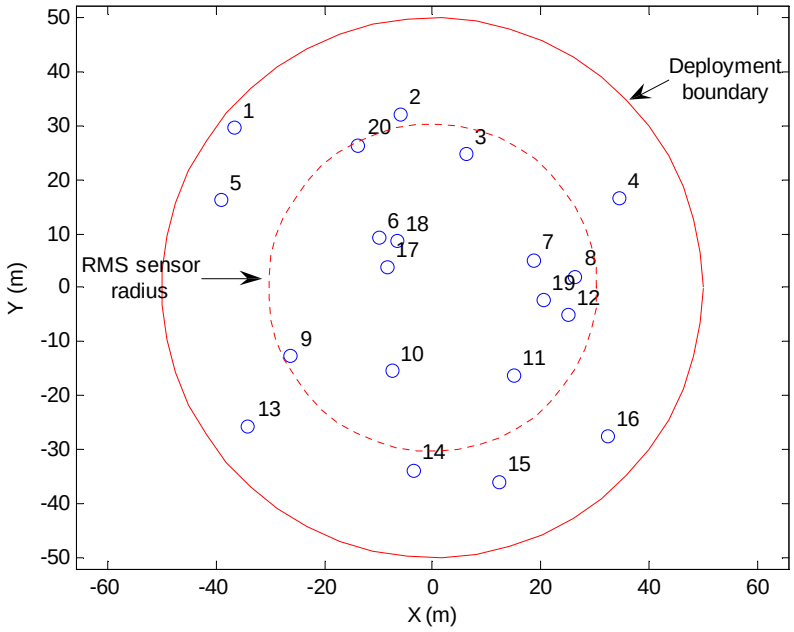


Figure 4.3: Sample 20 node sensor network used in anchor selection and anchor placement examples.

where $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_U$ represent the elements of the location parameter vector $\boldsymbol{\theta}$ corresponding to the anchor nodes and unknown-locations nodes, respectively. The RMS error (4.28) is clearly minimized by the node version of (4.2)

$$A^* = \arg\min_A \operatorname{tr} J^{-1}_{\boldsymbol{\theta}_U}(\boldsymbol{\theta}_A, \boldsymbol{\theta}_U). \tag{4.29}$$

As an example, we consider the sensor network depicted in Figure 4.3 consisting of 20 sensors uniformly deployed in a circular region with 50 m radius. For this network, we plot in Figure 4.4 the CRB RMS localization error corresponding to all possible $\binom{20}{3} = 1140$ sets of three anchors. The error is plotted versus the maximum principal angle ϕ_3 between the anchor-induced constraint space $\mathcal{C}_A$ and the non-identifiable

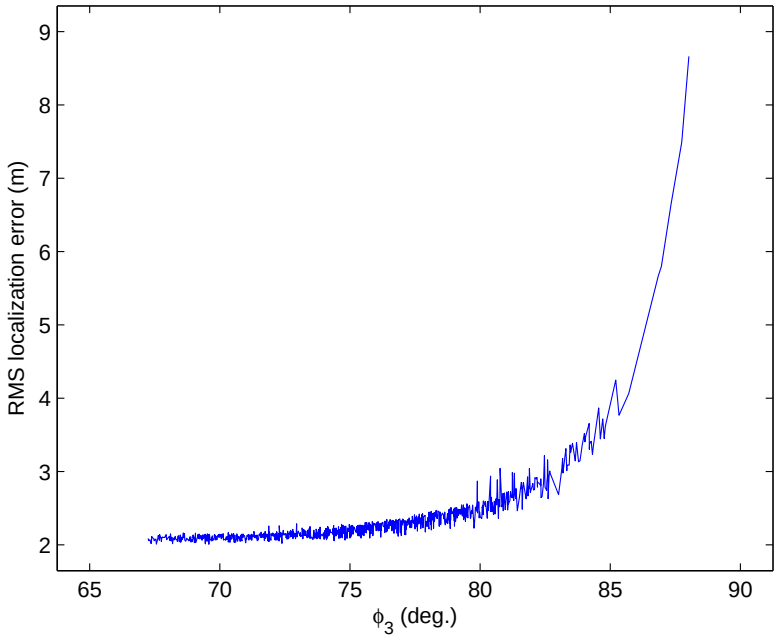


Figure 4.4: RMS localization error versus maximum principal angle ϕ_3 for the network in Figure 4.3 using all possible selections of 3 anchor nodes.

transformation subspace $\mathcal{V}$. In this example we assume that all sensors (unknown-location nodes and anchors) make pairwise distance measurements to one another and that the distance measurements are independently corrupted by zero-mean Gaussian noise with standard deviation $\sigma = 5\,\mathrm{m}$.

As the maximum angle ϕ_3 approaches 90° at least one dimension becomes uninformed by either constraints or measurements, and the estimation error rapidly increases. The general decreasing nature of localization error with maximum principal angle ϕ_3 supports the assertion that aligning the constraint and transformation subspaces provides a viable heuristic for anchor node placement. The error is not monotonic with ϕ_3 because this single measure does not capture all of the interactions between the constraints and measurements, however minimizing the maximum subspace angle is a good surrogate performance metric for minimizing RMS localization error.

In Figure 4.5(a) we plot the average RMS localization error of different anchor selection methods as a function of the number of sensors deployed. The average is taken over 50 random network deployments. The deployment region, deployment distribution, and measurements remain as before. Three anchor nodes were identified using the CRB-optimal approach (4.29), the gap-optimal approach (4.8), and the greedy gap algorithm. The methods have similar average RMS localization error, however the computational complexities, as shown in Figure 4.5(b), vary drastically. These timing results were obtained on a 2.4 GHz desktop PC. For the case of 50 deployed sensors, the CRB-optimal approach took an average of 33.3 s to identify three anchors. The gap-optimal and greedy gap algorithms took 7.0 s and 0.05 s, respectively. Random anchor selection is also a viable alternative which typically has

good average performance, but risks obtaining large errors a small fraction of the time—as seen by the right-hand side of Figure 4.4. The results in the next section about anchor placement also provide good heuristics to drive sensor selection.

4.4 Anchor Node Placement for Absolute Localization

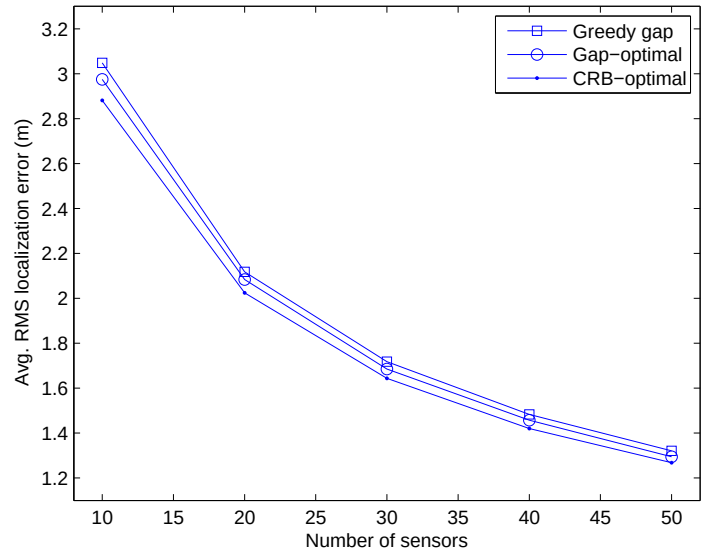
In the previous section we assumed that relative estimates of the network were available in order to aid in the selection of a subset of sensors to become anchor nodes. In this section we assume that we have control over the placement of a fixed number of anchor nodes during sensor deployment, but we do not assume any specific knowledge about the relative shape of the network. This problem differs from the anchor selection problem because the non-identifiable transformation subspace $\mathcal{V}$ changes with the anchor positions. In anchor selection, $\mathcal{V}$ remained fixed and we chose anchors A such that the resultant constraint subspace $\mathcal{C}_A$ was closely aligned with $\mathcal{V}$. Our interest in anchor placement continues to be the alignment of the subspaces, however now the constraint space is fixed and we influence $\mathcal{V}$ by positioning the anchors.

When the anchors are free to be placed arbitrarily, an additional complicating factor arises because some anchor locations significantly degrade the relative scene estimates. Before addressing this, we first consider how anchors should be positioned for maximal alignment of constraint and transformation subspaces.

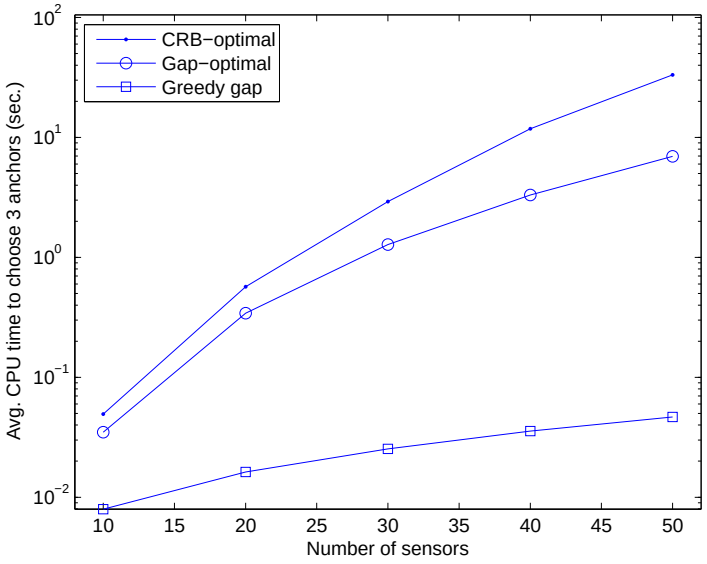
4.4.1 Anchor Positioning for Subspace Alignment

Angular Positioning of Anchors

In Section 4.3.2 we showed for distances, RSS, TOA, TDOA, and AOA that the principal angles between $\mathcal{C}_A$ and $\mathcal{V}$ had corresponding eigenvalues described by λ_1, λ_2, and λ_3 in equations (4.18)–(4.20). The second eigenvalue $\lambda_2 = q = n_a/n$ is



(a)



(b)

Figure 4.5: Average RMS localization error (a) of unknown-location sensors resulting from different anchor selection methods and the average runtime (b) of each method. Three anchor nodes where selected from the sensor population which was uniformly deployed in a circular region with radius 50 m.

constant with respect to the sensor positions. Therefore, in all of these measurement models, we may minimize the angular differences between $\mathcal{C}_A$ and $\mathcal{V}$ by maximizing the eigenvalues λ_1 and λ_3. In anchor positioning, the set A of anchors is fixed and the maximization is taken over the anchor positions $\boldsymbol{\theta}_A$; that is, we perform the optimization by choosing $2n_a$ real-valued parameters, namely the (x, y) locations of the n_a anchor nodes. We begin by maximizing the minimum eigenvalue, λ_3.

Observe, from (4.15) and (4.16), that

$$a^2 + b^2 = \frac{(\bar{x}_A)^2 + (\bar{y}_A)^2}{(n/n_a^2)\ ||\boldsymbol{\theta}||^2}, \tag{4.30}$$

where $\bar{x}_A$ is the average x-coordinate of the anchor nodes and $\bar{y}_A$ is the average y-coordinate. With respect to a and b, λ_3 is maximized when $a^2 + b^2 = 0$, which occurs when the mean x and mean y coordinates of the anchors are zero. One configuration that achieves this condition is when all anchors are uniformly distributed around a circle of any radius.

With respect to c, we see from (4.20), that λ_3 achieves its maximum value of q whenever $c \geq q$.

Radial Positioning of Anchors

The variable c, in (4.17), may be written

$$c = \frac{||\boldsymbol{\theta}_A||^2}{||\boldsymbol{\theta}_A||^2 + ||\boldsymbol{\theta}_U||^2}. \tag{4.31}$$

Therefore, c may be made arbitrarily large by placing the anchors progressively farther from the scene center. When $c \geq q$, we have $\lambda_2 = \lambda_3 = n_a/n$ and $\lambda_1 = c$. Hence, having maximized the two smallest eigenvalues we could proceed to maximize λ_1, the largest eigenvalue. λ_1 is maximized by maximizing $||\boldsymbol{\theta}_A||$, which is achieved by pushing

all of the anchor nodes as far out from the centroid as possible (given the physical constraints of the deployment). However, because the smallest eigenvalue (largest principal angle) dominates estimation performance, we expect negligible performance improvement in extending λ_1 significantly beyond the other two eigenvalues $\lambda_2 = \lambda_3 = n_a/n$. Therefore, we attempt to equate all three eigenvalues.

Setting $\lambda_1 = c = n_a/n$, we solve

$$\frac{||\boldsymbol{\theta}_A||^2}{||\boldsymbol{\theta}_A||^2 + ||\boldsymbol{\theta}_U||^2} = \frac{n_a}{n} \tag{4.32}$$

for $||\boldsymbol{\theta}_A||^2$ and obtain

$$||\boldsymbol{\theta}_A||^2 = \frac{n_a}{n - n_a}||\boldsymbol{\theta}_U||^2. \tag{4.33}$$

If we assume that all n_a anchors have a common radius r_0, then $||\boldsymbol{\theta}_A||^2 = n_a r_0^2$. Substituting this into (4.33) and solving for r_0 we find

$$r_0 = \frac{||\boldsymbol{\theta}_U||}{\sqrt{n - n_a}}. \tag{4.34}$$

The difference $n - n_a$ is the number of unknown-location sensors, and therefore, the quantity r_0 may be interpreted as the root-mean-square (RMS) distance of the unknown-location sensors from the scene centroid. When the anchors are all uniformly spaced around a circle of radius r_0, all three principal angles will be equal to

$$\phi_0 = \cos^{-1}(\sqrt{n_a/n}), \tag{4.35}$$

and when the common anchor radius exceeds r_0, only the smallest principal angle continues to decrease beyond (4.35).

For the sample network in Figure 4.3, the RMS sensor radius r_0 is illustrated by the inner dotted circle.

4.4.2 Total Error Considerations in Anchor Positioning

The previous subsection demonstrated that optimal subspace alignment is achieved by circular placement of anchor nodes around the sensor centroid and by extending them as far from the scene center as possible. However, extending the anchors too far radially has a detrimental effect on the network's ability to estimate the anchor locations relative to the unknown-location nodes. The subspace alignment criterion was developed under the assumption that absolute error was dominated by error in the positioning of the relative shape, i.e., that the relative error was not the dominant factor. When anchor locations are too far removed from the remaining nodes, though, the relative error begins to dominate. It is the correspondence between estimated anchor locations and the true anchor locations that determines the transformations needed to convert a relative scene estimate into an absolute one. In terms of the subspaces, $\mathcal{V}$ is the completely non-estimable transformation space that we have been focusing on. The orthogonal complement $\mathcal{V}^{\perp}$ represents the relative shape, and it is this portion that is degraded if the anchor nodes are poorly positioned relative to the rest of the network. In general, anchors should be sufficiently close to other unknown-location nodes such that several anchor-to-unknown measurements can be made in order to provide a good estimate of each anchor's relative position. Anchor-unknown measurements can also improve the relative error in an unknown-location nodes, however, as we assume these nodes are already well-connected this is a secondary effect.

To summarize, there are two competing effects related to anchor positioning which govern total localization error:

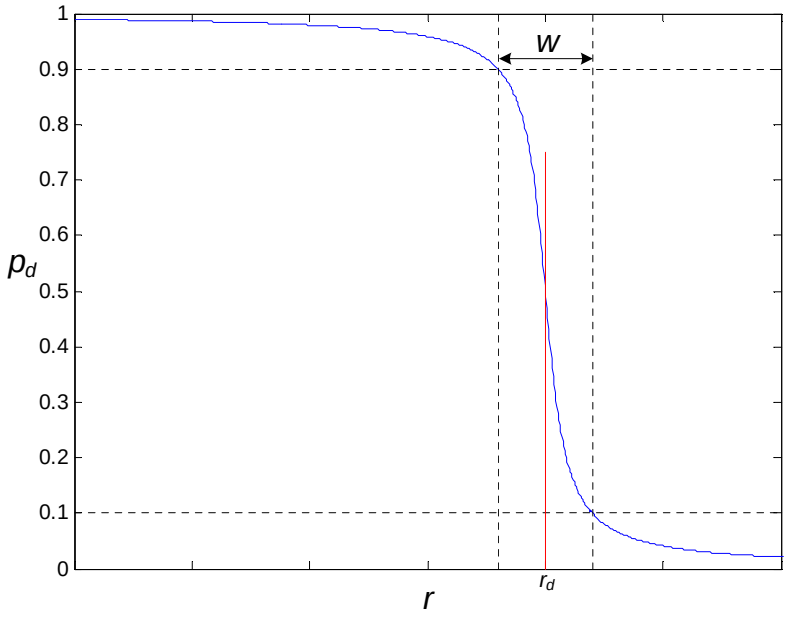


Figure 4.6: Limited measurement ranges are modeled using this soft disk model to control the probability of a measurement as a function of sensor separation, r.

1. Anchor nodes should be positioned as far from the scene center as possible in order to align the constraint and transformation subspaces, however positioning beyond the RMS sensor radius r_0 is expected to bring little improvement.

2. Anchor nodes should be positioned in the midst of unknown-location nodes in order for their relative positions to be well-estimated. This means that they should not be extended significantly beyond the network edge when measurement range is limited.

We illustrate these points quantitatively in the example considered below.

Limited-range Measurement Model

In order to realistically model finite measurement ranges, we consider a soft variation of classic hard detection boundaries. We model the probability $p_d(r)$ of obtaining a measurement between two sensors as a function of the distance r between them

$$p_d(r) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\big(\alpha(r - r_d)\big), \tag{4.36}$$

where r_d is the 50% detection range, $\alpha = \frac{-2}{w}\tan(-0.4\pi)$, and w is the 10%–90% transition width as illustrated in Figure 4.6. We refer to (4.36) as the disk model.

As we show in Appendix A, Fisher's information for a single measurement with limited probability of detection is

$$J = p_d(\boldsymbol{\theta})J_\theta + \frac{1}{p_d(\boldsymbol{\theta})\big(1 - p_d(\boldsymbol{\theta})\big)}\Big(\frac{\partial p_d(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)\Big(\frac{\partial p_d(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)^T. \tag{4.37}$$

where the first term is the information from the measurement itself, and the second term is the information derived from the observability status and the known observation model $p_d(\boldsymbol{\theta})$. For example, if two sensors do not make a measurement to one another, they are most likely separated by a distance greater than r_d. In this work, we assume that estimators only utilize the available measurements and that they do not try to derive information from the presence of missing data. Therefore, only the first term of (4.37) is used in calculating the information matrix of a single measurement m_{ij} between sensors i and j

$$J_{ij} = p_d(||\boldsymbol{p_i} - \boldsymbol{p_j}||)\; E\left[\left(\frac{\partial \ln f(m_{ij}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln f(m_{ij}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right], \tag{4.38}$$

where $\boldsymbol{p_i} = [x_i\, y_i]$ is the position vector of sensor i, and $f(m_{ij}|\boldsymbol{\theta})$ is the probability density function of measurement m_{ij}. For independent measurements between the

various sensor pairs, the Fisher information adds, giving

$$J_T(\boldsymbol{\theta}) = \sum_{i \neq j} J_{ij} \tag{4.39}$$

as the total Fisher information matrix.

When the measurements m_{ij} have their mean $\mu_{ij}(\boldsymbol{\theta})$ structured by the parameter vector $\boldsymbol{\theta}$ and have i.i.d. zero-mean Gaussian measurement noise $N(0, \sigma^2)$, $J_T(\boldsymbol{\theta})$ takes the form

$$J_T(\boldsymbol{\theta}) = \sum_{i \neq j} \frac{p_d(||\boldsymbol{p_i} - \boldsymbol{p_j}||)}{\sigma^2} \left(\frac{\partial \mu_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \mu_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T. \tag{4.40}$$

Note that the presence of p_d in (4.40) has the interpretation of increasing the effective noise variance with distance. In the examples that follow, we consider distance measurements with independent Gaussian noise. Hence, (4.40) applies with $\mu_{ij} = ||\boldsymbol{p_i} - \boldsymbol{p_j}||$.

4.4.3 Results

Subspace Alignment Versus Relative Error

In Figure 4.7 we plot the RMS localization error (4.28) for the 20 unknown-location sensors in Figure 4.3 corresponding to three additional anchor nodes which have been added to the scene. The anchors are uniformly distributed around a circle (at 30°, 150°, and 270° counter-clockwise from the x–axis), and performance is plotted as a function of the common anchor radius r. Distance measurements are used with Gaussian measurement noise with $\sigma = 5\,\mathrm{m}$, and detection parameters are $r_d = 25\,\mathrm{m}$, $w = 5\,\mathrm{m}$.

On the left side of Figure 4.7 ($0 < r < 30$, approximately), we a observe a large decrease in localization error as r increases due to improving subspace alignment

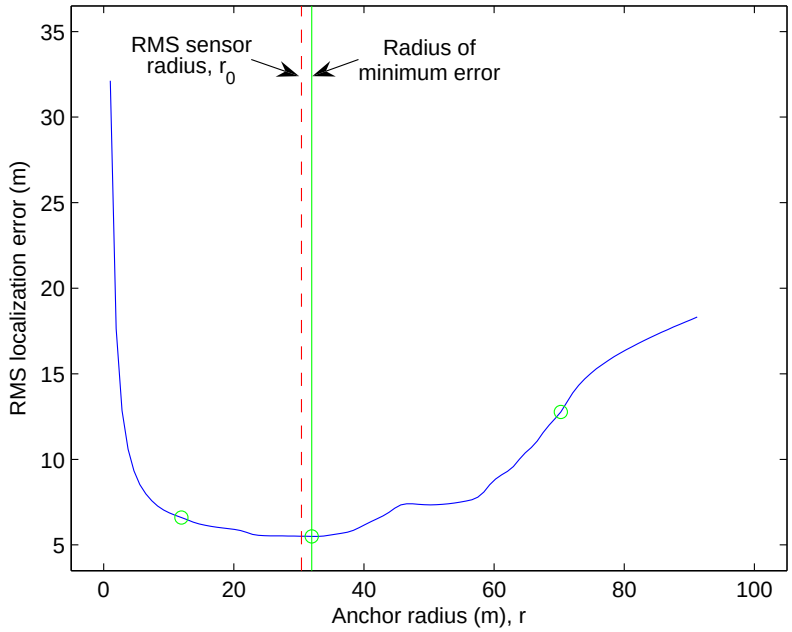


Figure 4.7: RMS localization error of the sensors in Figure 4.3 versus the common anchor radius of 3 additional anchor nodes with uniform circular deployment.

(Item 1, Page 107). The minimum error occurs at $r = 32.0\,\mathrm{m}$ which is only slightly larger than the RMS sensor radius $r_0 = 30.4\,\mathrm{m}$. As r continues to increase, the limited detection range degrades the anchor estimates and total RMS localization increases; this latter region quantitatively demonstrates the error increase that was qualitatively discussed in Item 2, Page 107.

These concepts are further illustrated in Figure 4.8 where we explicitly consider the effects at three radii $r \in \{12.0\,\mathrm{m}, 32.0\,\mathrm{m}, 70.0\,\mathrm{m}\}$, identified by the circles in Figure 4.7. In the bottom row of figures, 4.8(d)–4.8(f), we plot the 20 unknown-location sensors as well as the three anchor nodes for each of the three radii. Around each of the anchor nodes we also plot the 3σ uncertainty ellipse derived from the relative CRB $J^\dagger$ (see Chapter 3, Section 3.4.4). In general, the anchor uncertainty is growing with r. However, looking at Figure 4.8(f) we see that the anchor in the NW quadrant has lower uncertainty than the other two anchors due to a closer proximity of unknown-location nodes. The error plateau around $r = 50\,\mathrm{m}$ seen in Figure 4.7 is due to the improved relative localizability of this particular anchor as it passes between sensors 1 and 5. The relative error ellipses of the 20 unknown-location sensors are not shown in order to simplify the plots in Figure 4.8.

In the top row of figures, 4.8(a)–4.8(c), we give a 2D depiction of the subspaces involved, their relative orientation, and the size of relative uncertainty. When the anchor radius is small, as in Figure 4.8(a), the angle between the constraint space $\mathcal{C}_A$ and the non-identifiable transformation subspace $\mathcal{V}$ is large, 81° in this case. As r is increased, this maximum angle decreases until $r \geq r_0$. Because the anchor radius in Figures 4.8(b) and 4.8(c) is greater than r_0, they both achieve the minimum maximum subspace angle predicted by Eq. (4.35), $\phi_0 = 69°$ ($n_a = 3$, $n = 23$).

The relative uncertainty is described by the relative CRB $J^\dagger$, which we may interpret as a hyperellipsoid in $\mathbb{R}^{46}$ (there are 23 total nodes, with an x– and y–coordinate for each). Three of the ellipsoid axes corresponding to $\mathcal{V}$, will extend to infinity. The other $46 - 3$ axes correspond to the relative configuration $\mathcal{V}^\perp$ and have finite error radii. To encode the total relative error into a single number e_r we report the RMS error of $J^\dagger$, $e_r = [\mathrm{tr}\, J^\dagger/n]^{1/2}$. In Figures 4.8(a)–4.8(c), the relative error e_r is depicted as the "thickness" of the relative subspace, and we see that, in concert with the anchor uncertainty ellipses in the bottom row of plots, the relative error e_r grows large when r is large.

The total error is equal to the intersection of the constraint (which is perpendicular to the constraint axis) with the uncertainty ellipsoid (thin strip in the 2D rendering). When r is small the relative uncertainty band is thin, as in Figure 4.8(a), but the the large angle results in a large intersection with high resultant error—illustrated by the thick line intersecting the band in Figures 4.8(a)–4.8(c). When r is large, as in Figure 4.8(c), the angle is smaller but the band becomes thick and the total error remains large. An optimal comprimise between subspace angle and relative uncertainty is achieved for some intermediate value of r, as illustrated in Figure 4.8(b).

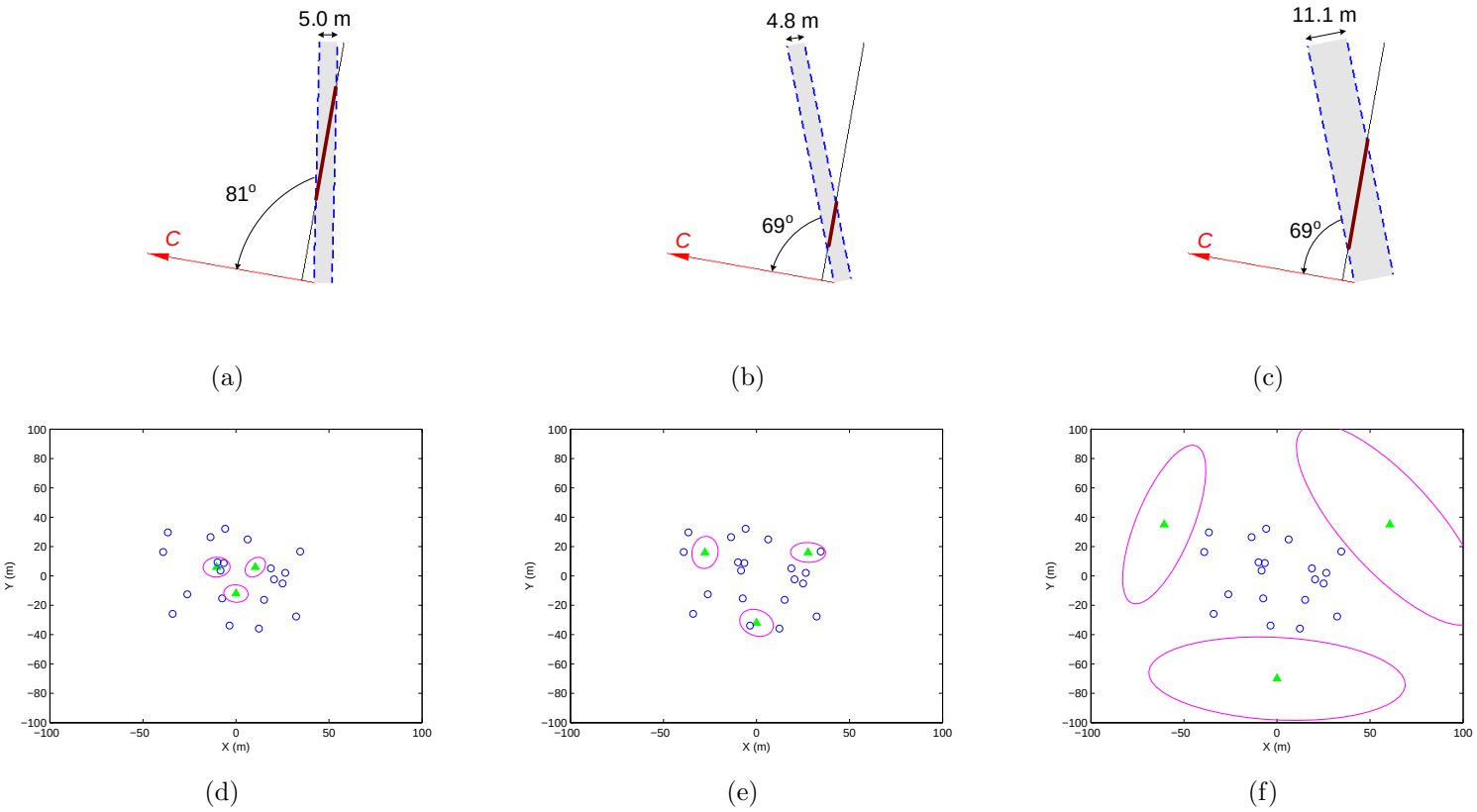


Figure 4.8: Example of the trade-off between subspace alignment and relative localizability. Top row: The subspace $\mathcal{C}$ denotes the constraint space induced by the anchors. The non-identifiable transformation subspace is represented by the infinitely long strip. The thickness of the strip encodes the relative localization error, e_r. Going from (a) to (b) to (c), the anchor radius is increasing and the angle between $\mathcal{C}$ and $\mathcal{V}$ decreases, although the relative error generally increases. The final total error is represented by the intersection of a constraint, which is perpendicular to the $\mathcal{C}$–direction, with the strip. The optimal radius (b) represents a compromise between minimizing subspace angle and minimizing relative error. Bottom row: The sensor field with unknown-location sensors (○) and three anchor nodes (▲). Each sensor-anchor configuration induces the subspace diagram above it. The *relative* uncertainty of the anchors nodes is shown by the 3σ ellipses drawn around them. (a),(d): radius=12.0 m, (b),(e): radius=32.0 m, (c),(f): radius=70.0 m

Comparison to Oracle-placement

In Figure 4.9 we consider CRB-optimal anchor locations, as given by an oracle with perfect sensor location knowledge after solving Eq. 4.29. Figures 4.9(a)–4.9(c) present histograms of the optimal anchor radii and angular separation between three oracle-placed anchors in 1000 randomly generated network configurations. The random networks consisted of 20 sensors uniformly deployed in a circular region with radius 50 m. We again use distance measurements and the disk model ($r_d = 25\,\text{m}$, $w = 5\,\text{m}$). The measurement noise was $\sigma = 10\,\text{m}$, although this does not influence optimal placement (see Eq. 4.40).

From Figure 4.9(a), the average anchor radius in optimal placement was 31.8 m. If we calculate the expected value of the RMS sensor radius r_0 for this deployment scenario (not based on the realizations of the experiment), we obtain $E[r_0] = 35.3\,\text{m}$, which is reasonably close to the empirical mean optimal. The average angular separation between adjacent anchors was, from Figures 4.9(b) and 4.9(c), 116° and 113°. These values are close to 120° which would be obtained in uniform circular placement.

These empirical results and the theoretical results above suggest that, in absence of any detailed information about the relative deployment of sensors, anchor nodes should be uniformly distributed around a circle with radius approximately equal to r_0. In Figure 4.10 we compare the performance of circular–r_0 placement with oracle placement. Let $\boldsymbol{\theta}_A^*$ denote the positions of the oracle-placed anchors, and let $\boldsymbol{\theta}_A^c$ denote the positions in circular–r_0 placement. Further, let $e(\boldsymbol{\theta}_A^*)$ and $e(\boldsymbol{\theta}_A^c)$ denote the RMS localization errors for optimal and circular–r_0 placement, respectively. Using the previously described 1000 network realizations, we plot in Figure 4.10 a histogram

of the ratio $\gamma = e(\boldsymbol{\theta}_A^c)/e(\boldsymbol{\theta}_A^*)$. From the figure we see that the performance of optimal anchor placement, utilizing perfect knowledge of the sensor locations, exceeds circular–r_0 placement by only 23%, on average.

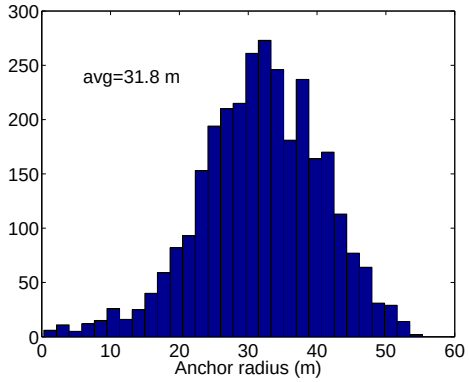
Effects of Measurement Range

Finally, we consider the effects of the detection range r_d on average localization performance. Let $e_{rms}(\boldsymbol{\theta}_A^c(r), \boldsymbol{\theta}_U)$ denote the RMS localization error of circularly placed anchors at radius r for sensors described by positions $\boldsymbol{\theta}_U$. We are interested in the expected error $E[e_{rms}(\boldsymbol{\theta}_A^c(r), \boldsymbol{\theta}_U)]$, where the expectation is taken over realizations of $\boldsymbol{\theta}_U$. We evaluate the expected error for uniform deployment by taking an empirical average over the 1000 random networks described in the previous subsection. In Figure 4.11 we evaluate $E[e_{rms}(\boldsymbol{\theta}_A^c(r), \boldsymbol{\theta}_U)]$ as a function of anchor radius r for three anchor nodes and for three different detection ranges $r_d \in \{25\,\mathrm{m}, 35\,\mathrm{m}, 45\,\mathrm{m}\}$; in all cases, w in equation (4.36) is set to 5 m.

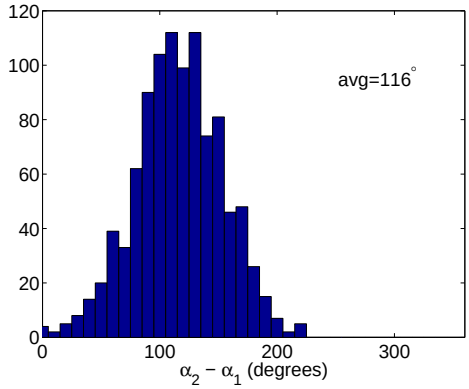
Figure 4.11 is similar to Figure 4.7 with smoother trends due to averaging. Clearly increasing the measurement range improves localization performance as expected. Also, a larger measurement range r_d increases the width of the valley around the optimal minimum-error anchor radius. This is because the larger measurement range allows the system to obtain quality relative estimates of the anchor positions for a larger range of anchor radii.

4.5 Conclusions

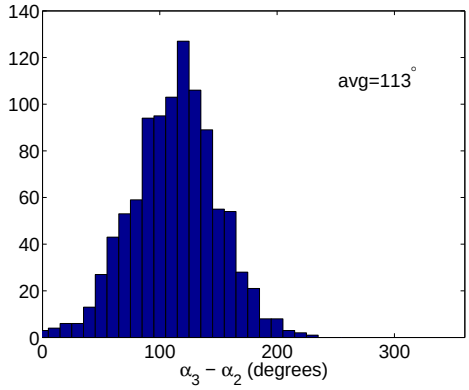
Performing absolute sensor localization from inter-sensor measurements is an inherently ill-posed problem because measurements are invariant to global transformations, such as translations and rotations, of the network. In this chapter we



(a)



(b)



(c)

Figure 4.9: Statistics of CRB-optimal anchor positions utilizing perfect knowledge of the unknown sensor locations. The results are for 3 anchors and 1000 uniform deployments of 20 sensors in a circular region with 50 m radius. Detection model parameters were $r_d = 25\,\mathrm{m}$ and $w = 5\,\mathrm{m}$.

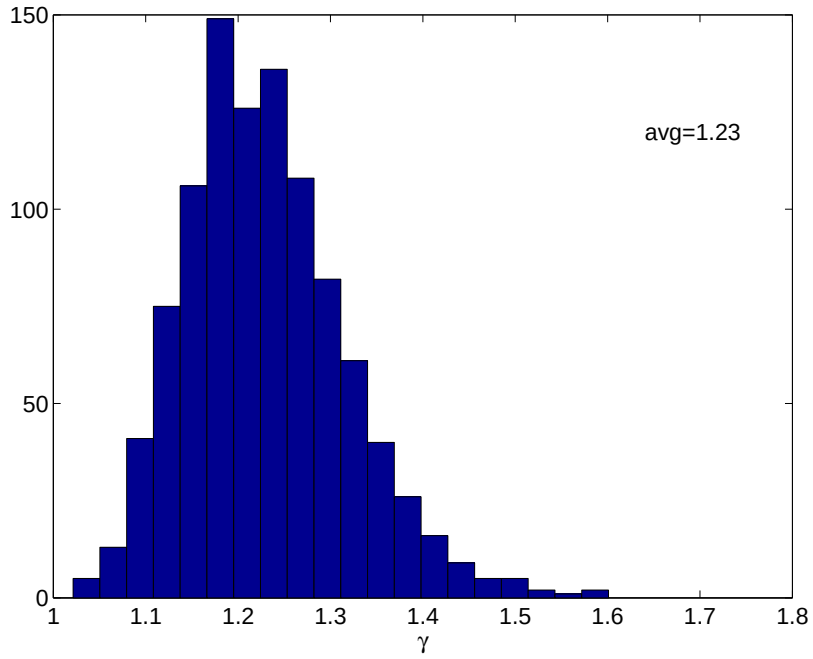


Figure 4.10: Histogram of $\gamma = e(\boldsymbol{\theta}_A^c)/e(\boldsymbol{\theta}_A^*)$ indicating that the localization error of circular-r_0 placement exceeds the oracle-placement bound by only 23%, on average.

addressed the use of anchor nodes to regularize the absolute localization problem and considered how the anchor nodes should be positioned, relative to the remaining unknown-location sensors, in order to maximize total localization performance.

The primary idea behind our approach was that the subspace $\mathcal{C}_A$ of constrained location parameters, induced by the anchor nodes, should be closely aligned with non-identifiable subspace of rigid transformations $\mathcal{V}$. In the anchor selection problem, $\mathcal{V}$ is fixed and $\mathcal{C}_A$ is discretely controlled by selecting different anchor sets. In anchor placement, $\mathcal{C}_A$ is fixed and $\mathcal{V}$ is smoothly controlled by moving the anchor nodes. In both cases, selection and placement, we vary the control such that the constrained subspace and non-identifiable subspace are brought into close alignment, as measured by the principal angles between them.

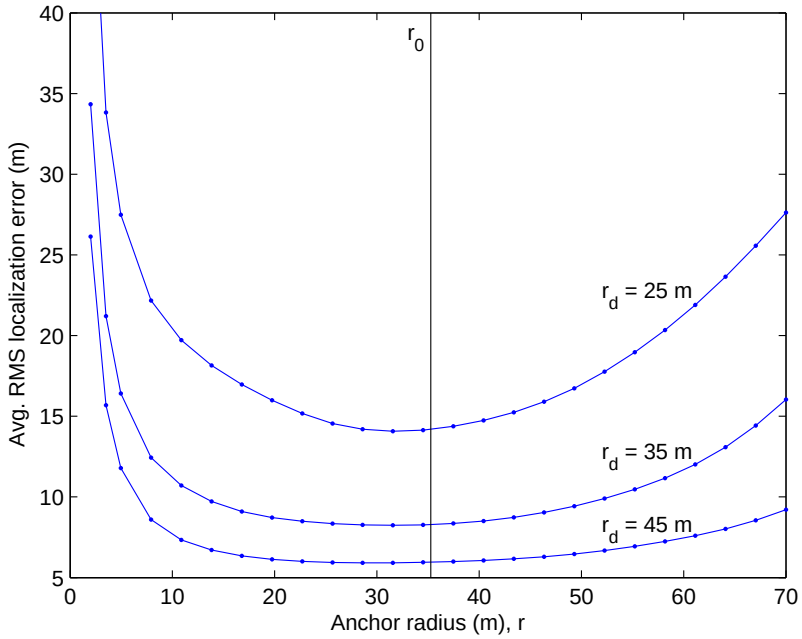


Figure 4.11: Average RMS localization error as a function of common anchor radius and measurement range, r_d.

We demonstrated that subspace alignment is a strong indicator of localization performance, and that, through uniform circular placement of anchor nodes, it can be optimized without requiring specific details about the network's configuration. This strategy holds irregardless of measurement type (distances, RSS, TOA, TDOA, AOA, or ADOA) and establishes a theoretical basis for the frequent, but empirical, observation that anchors should be placed on the network boundary. Although, we also showed that boundary placement is sub-optimal in the presence of finite measurement ranges.

Finally, we considered the scenario where only a few anchors were available and focused on positioning them such that they were maximally informative in the non-identifiable space of rigid transformations—which would otherwise have infinite estimation error. However, when large numbers of anchors are available for deployment, it may not be necessary to use all anchor nodes to resolve ambiguities in absolute position, and a fraction could be placed with the aim of reducing relative error in the network's estimated shape. Identification and placement of anchors in this scenario is a topic of future research.

APPENDIX A

FISHER INFORMATION FOR MISSING-DATA PROBLEMS

In this appendix we evaluate Fisher's information matrix for cases of missing data. We consider separately (i) the case where the mechanism controlling the missing data is independent of the parameter vector $\boldsymbol{\theta}$, and (ii) the case where the missing data mechanism depends on $\boldsymbol{\theta}$. In the language of missing data process classifications in statistics, the first case corresponds to missing-completely-at-random (MCAR), and the second case corresponds to missing-not-at-random (MNAR) or non-ignorable (NI) [85, 86].

A.1 Parameter-independent Case, p_d

Let $Y \sim f(y;\boldsymbol{\theta})$ be a random variable—scalar or vector—that we observe with probability p_d (probability of detection) in a given experiment. Therefore, a realization of the experiment results in a measurement that consists of the pair (Y, I), where I is an indicator; $I = 1$ if Y is observed, $I = 0$ if Y is not observed. We use the convention $Y = y_0$ when Y is not observed. Hence, the two possible observation cases

are

$$(Y, I) = (y, 1) \qquad \text{observation case} \tag{A.1}$$

$$(Y, I) = (y_0, 0) \qquad \text{missed observation case.} \tag{A.2}$$

The density function of (Y, I) is

$$p(y, i; \boldsymbol{\theta}) = p_d f(y; \boldsymbol{\theta}) i + (1 - p_d)(1 - i)\delta_{y_0}(y), \tag{A.3}$$

where $\delta_{y_0}(y)$ is the Dirac delta function centered at y_0. The (j, k) element of Fisher's information matrix is

$$J_1(j,k) = E_{y,i}\Big[\frac{\partial \ln p(y,i;\boldsymbol{\theta})}{\partial \theta_j}\frac{\partial \ln p(y,i;\boldsymbol{\theta})}{\partial \theta_k}\Big] \tag{A.4}$$

$$= E_{y,i}\Big[\frac{1}{(p(y,i;\boldsymbol{\theta}))^2}\frac{\partial p(y,i;\boldsymbol{\theta})}{\partial \theta_j}\frac{\partial p(y,i;\boldsymbol{\theta})}{\partial \theta_k}\Big] \tag{A.5}$$

$$= E_{y,i}\Big[\Big(\frac{p_d\, i}{p_d f(y;\boldsymbol{\theta}) i + (1-p_d)(1-i)\delta_{y_0}(y)}\Big)^2 \frac{\partial f(y;\boldsymbol{\theta})}{\partial \theta_j}\frac{\partial f(y;\boldsymbol{\theta})}{\partial \theta_k}\Big]. \tag{A.6}$$

Performing the expectation over i and then y, we have

$$J_1(j,k) = p_d E_{y|i=1}\Big[\frac{1}{(f(y;\boldsymbol{\theta}))^2}\frac{\partial f(y;\boldsymbol{\theta})}{\partial \theta_j}\frac{\partial f(y;\boldsymbol{\theta})}{\partial \theta_k}\Big] + (1-p_d)0 \tag{A.7}$$

$$= p_d J_{\boldsymbol{\theta}}(j,k), \tag{A.8}$$

where $J_{\boldsymbol{\theta}}(j,k)$ is the (j,k) element of the standard Fisher information matrix $J_{\boldsymbol{\theta}}$ given a single measurement. Hence, the total Fisher information matrix for a single experiment with probability p_d of missing a measurement is

$$J_1 = p_d J_{\boldsymbol{\theta}}. \tag{A.9}$$

If there are n independent measurements of (Y, I), then on average we have $n\, p_d$ measurements of Y. So, we would expect the Fisher information to be $(n\, p_d) J_{\boldsymbol{\theta}}$. This is consistent with (A.9).

A.2 Parameter-dependent Case, $p_d(\boldsymbol{\theta})$

In the parameter-independent case, the probability of missing a measurement did not depend on the parameter being estimated; hence, missed data were essentially ignored and the resulting Fisher information depended only on the expected number of actual observations of the data Y. However, when the probability of missing a measurement depends on the parameter to be estimated, there is information about $\boldsymbol{\theta}$ in a missed measurement itself. This will increase the Fisher information above the parameter-independent case.

We now write the probability of obtaining a measurement as a function of $\boldsymbol{\theta}$, $p_d(\boldsymbol{\theta})$. Recall that $p_d(\boldsymbol{\theta}) \in [0,1]$ is not a prior or even a density on $\boldsymbol{\theta}$ and need not integrate to one. The distribution of (Y, I) is otherwise the same as (A.3), and, omitting the arguments of $p(y, i; \boldsymbol{\theta})$, $f(y; \boldsymbol{\theta})$, and $p_d(\boldsymbol{\theta})$ for notational simplicity, we evaluate the (j,k) element of Fisher's information matrix as

$$
\begin{aligned}
J_2(j,k) &= E_{y,i}\Big[\frac{1}{p^2}\Big(\frac{\partial p_d f i}{\partial \theta_j} + \frac{\partial (1-p_d)(1-i)\delta_{y_0}(y)}{\partial \theta_j}\Big)\Big(\frac{\partial p_d f i}{\partial \theta_k} + \frac{\partial (1-p_d)(1-i)\delta_{y_0}(y)}{\partial \theta_k}\Big)\Big] \\
&= p_d E_{y|i=1}\Big[\frac{1}{(p_d f)^2}\frac{\partial p_d f}{\partial \theta_j}\frac{\partial p_d f}{\partial \theta_k}\Big] + \\
&\qquad (1-p_d)E_{y|i=0}\Big[\frac{1}{((1-p_d)\delta_{y_0}(y))^2}\frac{\partial (1-p_d)\delta_{y_0}(y)}{\partial \theta_j}\frac{\partial (1-p_d)\delta_{y_0}(y)}{\partial \theta_k}\Big] \\
&= p_d E_{y|i=1}\Big[\Big(\frac{1}{f}\frac{\partial f}{\partial \theta_j} + \frac{1}{p_d}\frac{\partial p_d}{\partial \theta_j}\Big)\Big(\frac{1}{f}\frac{\partial f}{\partial \theta_k} + \frac{1}{p_d}\frac{\partial p_d}{\partial \theta_k}\Big)\Big] + \frac{1}{1-p_d}\frac{\partial p_d}{\partial \theta_j}\frac{\partial p_d}{\partial \theta_k} \\
&= p_d E_{y|i=1}\Big[\frac{1}{f^2}\frac{\partial f}{\partial \theta_j}\frac{\partial f}{\partial \theta_k}\Big] + \frac{1}{p_d(1-p_d)}\frac{\partial p_d}{\partial \theta_j}\frac{\partial p_d}{\partial \theta_k} + \\
&\qquad \frac{\partial p_d}{\partial \theta_j}E_{y|i=1}\Big[\frac{1}{f}\frac{\partial f}{\partial \theta_k}\Big] + \frac{\partial p_d}{\partial \theta_k}E_{y|i=1}\Big[\frac{1}{f}\frac{\partial f}{\partial \theta_j}\Big] \qquad \text{(A.10)}
\end{aligned}
$$

The expectation in the first term of (A.10) is equal to the (j,k) element of Fisher's information matrix $J_{\boldsymbol{\theta}}$ for a single observation of Y, and the expectations in the last

two terms of (A.10) are of elements of the score vector of Y, which always has zero expectation. Hence, Fisher's information matrix for the parameter-dependent case is

$$J_2 = p_d(\boldsymbol{\theta}) J_{\boldsymbol{\theta}} + \frac{1}{p_d(\boldsymbol{\theta})\big(1 - p_d(\boldsymbol{\theta})\big)} \Big(\frac{\partial p_d(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big) \Big(\frac{\partial p_d(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)^T. \tag{A.11}$$

When p_d is independent of $\boldsymbol{\theta}$, (A.11) clearly reduces to (A.9). Note also that the second term in (A.11) is always positive semi-definite, so, for equivalent values of p_d, Fisher's information in the parameter-dependent case is never less than the independent case

$$J_2 \geq J_1. \tag{A.12}$$

The second term of (A.11) quantifies the information present in a missed observation.

BIBLIOGRAPHY

[1] D. Goense, J. Thelen, and K. Langendoen, "Wireless sensor networks for precise phytophthora decision support," 5^{th} *European Conference on Precision Agriculture (5ECPA)*, June 2005.

[2] P. C. Robert, Ed., *Int. Journal on Advances in the Science of Precision Agriculture*, 1998-2006, http://www.kluweronline.com/issn/1385-2256.

[3] B. Hansen, "Wireless sensor network helps prevent water, soil pollution," *Civil Engineering*, vol. 75, no. 7, pp. 34–35, 2005.

[4] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. New York, NY, USA: ACM Press, 2002, pp. 88–97.

[5] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao, "Habitat monitoring: application driver for wireless communications technology," *SIGCOMM Comput. Commun. Rev.*, vol. 31, no. 2 supplement, pp. 20–41, 2001.

[6] C. Chen, A. M. Ali, and H. Wang, "Design and testing of robust acoustic arrays for localization and enhancement of several bird sources," *Proc. Information Processing in Sensor Networks*, pp. 268–275, 2006.

[7] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, 2002.

[8] A. Arora *et al.*, "A line in the sand: a wireless sensor network for target detection, classification, and tracking," *Computer Networks*, vol. 46, pp. 605–634, 2004.

[9] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Journal of Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[10] M. Perkins, N. Correal, and B. O'Dea, "Emergent wireless sensor network limitations: a plea for advancement in core technologies," in *Sensors, 2002. Proceedings of IEEE*, vol. 2, 12-14 June 2002, pp. 1505–1509.

[11] A. Arora *et al.*, "Exscal: elements of an extreme scale wireless sensor network," in *Embedded and Real-Time Computing Systems and Applications, 2005. Proceedings. 11th IEEE International Conference on*, 17-19 Aug. 2005, pp. 102–108.

[12] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Springer, 1997.

[13] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[14] P. Drineas, M. Magdon-Ismail, G. Pandurangan, R. Virrankoski, and A. Savvides, "Distance matrix reconstruction from incomplete distance information for sensor network localization," *Proc. 3rd Annual IEEE Comm. Society on Sensor and Ad Hoc Communications and Networks (SECON)*, vol. 2, pp. 536–544, 2006.

[15] Y. Shang, W. Ruml, Y. Zhang, and M. P. Fromherz, "Localization from mere connectivity," *Mobihoc*, pp. 201–212, June 2003.

[16] X. Ji and H. Zha, "Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling," *Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 4, pp. 2652 – 2661, 2004.

[17] J. Smith and J. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 12, 1987.

[18] R. L. Moses, D. Krishnamurthy, and R. Patterson, "A self-localization method for wireless sensor networks," *Eurasip Journal on Applied Signal Processing, Special Issue on Sensor Networks*, vol. 2003, no. 4, pp. 348–358, March 2003.

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

[20] P. Biswas and Y. Ye, "Semidefinite programming for ad hoc wireless sensor network localization," in *IPSN '04: Proceedings of the Third International Symposium on Information Processing in Sensor Networks*. New York, NY, USA: ACM Press, 2004, pp. 46–54.

[21] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *MobiCom '01: Proceedings of the 7th Annual International Conference on Mobile Computing and Networking.* New York, NY, USA: ACM Press, 2001, pp. 166–179.

[22] D. Moore, J. Leonard, D. Rus, and S. Teller, "Robust distributed network localization with noisy range measurements," in *SenSys '04: Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems.* New York, NY, USA: ACM Press, 2004, pp. 50–61.

[23] X. Sheng and Y.-H. Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 44–53, 2005.

[24] N. Patwari and A. Hero, "Signal strength localization bounds in ad hoc and sensor networks when transmit powers are random," in *IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 2006.

[25] D. Li and Y. H. Hu, "Energy-based collaborative source localization using acoustic microsensor array," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 4, pp. 321–337, 2003.

[26] K. Yedavalli, B. Krishnamachari, S. Ravula, and B. Srinivasan, "Ecolocation: a sequence based technique for RF localization in wireless sensor networks," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, 15 April 2005, pp. 285–292.

[27] N. Patwari and A. Hero III, "Using proximity and quantized RSS for sensor localization in wireless networks," *Proc. 2nd International ACM Workshop on Wireless Sensor Networks and App.*, San Diego, CA, Sept. 19, 2003.

[28] L. Doherty, L. El Ghaoui, and K. Pister, "Convex position estimation in wireless sensor networks," *Proc. INFOCOM*, vol. 3, pp. 1655–1663, 2001.

[29] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking.* New York, NY, USA: ACM Press, 2003, pp. 81–95.

[30] D. Niculescu and B. Nath, "Ad hoc positioning system (APS) using AOA," *Proc. INFOCOM*, vol. 3, pp. 1734 – 1743, 2003.

[31] R. L. Moses, D. Krishnamurthy, and R. Patterson, "An auto-calibration method for unattended ground sensors," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, no. 4, pp. 2941–2944, May 13-17, 2002.

[32] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, July 2005.

[33] G. Mao, B. Fidan, and B. D. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, pp. 2529–2553, 2007.

[34] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Transactions on Signal Processing*, vol. 49, pp. 87–90, January 2001.

[35] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEETrans. on Sig. Proc.*, vol. 43, no. 11, pp. 2714–2723, 1995.

[36] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel fir filters," *IEEE Trans. on Sig. Proc.*, vol. 43, no. 2, pp. 516–525, 1995.

[37] S. Veres, "On the overparametrized maximum likelihood estimation of ARMA processes," *Proc.* 5^{th} *Pannonian Symp. on Math. Stat.*, pp. 381–400, 1985.

[38] A. Klein and P. Spreij, "On fisher's information matrix of an ARMAX process and sylvester's resultant matrix," *Lin. Alg. Appl.*, vol. 237, pp. 579–590, 1996.

[39] S. Watanabe, "Almost all learning machines are singular," *Proc. IEEE Symp. on Foundations of Comp. Intelligence*, pp. 383–388, 2007.

[40] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968.

[41] C. Chang and A. Sahai, "Cramér-Rao type bounds for localization," *EURASIP Journal on Applied Signal Processing*, pp. 1–13, 2006.

[42] J. N. Ash and R. L. Moses, "Relative and absolute errors in sensor network localization," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1033–1036, April 15–20 2007.

[43] P. Stoica and B. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Processing Lett.*, vol. 5, no. 7, pp. 177–179, 1998.

[44] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *ACM Trans. on Sensor Networks*, vol. 2, no. 2, pp. 188–220, 2006.

[45] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Comput. Networks*, vol. 43, no. 4, pp. 499–518, 2003.

[46] A. Savvides, H. Park, and M. B. Srivastava, "The bits and flops of the n-hop multilateration primitive for node localization problems," *Intl. Workshop on Sensor Nets. & Apps.*, pp. 112–121, Sep 2002.

[47] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 30–39, 2002.

[48] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[49] J. Ash and L. Potter, "Sensor network localization via received signal strength measurements with directional antennas," *Proc.* 42^{nd} *Annual Allerton Conference on Communication, Control, and Computing*, pp. 1861–1870, Sept. 2004.

[50] P. H. Schönemann and R. M. Carroll, "Fitting one matrix to another under choice of a central dilation and a rigid motion," *Psychometrika*, vol. 35, pp. 245–255, 1970.

[51] D. Bertsekas and R. Gallager, *Data Networks*, 2^{nd} ed. Prentice-Hall, 1992.

[52] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," *Proc. of the* 20^{th} *International Conf. on Machine Learning*, 2003.

[53] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2^{nd} ed. New York: Springer, 1994.

[54] N. Priyantha, A. Miu, H. Balakrishnan, and S. Teller, "The cricket compass for context-aware mobile applications," *Proc. of the 6th ACM MOBICOM*, 2001.

[55] T. F. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on Optimization*, vol. 6, pp. 418–445, 1996.

[56] G. Golub and C. Van Loan, *Matrix Computations*, 3^{rd} ed. Johns Hopkins University Press, 1996.

[57] G. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. on Signal Processing*, vol. 40, no. 6, pp. 1535–1541, June 1992.

[58] M. Gu and S. C. Eisenstat, "A stable and fast algorithm for updating the singular value decomposition," New Haven, CT, Tech. Rep. YALEU/DCS/RR-966, 1993.

[59] E. G. Larsson, "Cramér-Rao bound analysis of distributed positioning in sensor networks," *IEEE Signal Processing Lett.*, vol. 11, no. 3, pp. 334–337, 2004.

[60] A. Savvides, W. L. Garber, R. L. Moses, and M. B. Srivastava, "An analysis of error inducing parameters in multihop sensor node localization," *IEEE Trans. Mobile Computing*, vol. 4, no. 6, pp. 567–577, 2005.

[61] A. Catovic and Z. Sahinoglu, "The Cramér-Rao bounds of hybrid TOA/RSS and TDOA/RSS location estimation schemes," *IEEE Commun. Lett.*, vol. 8, no. 10, pp. 626–628, 2004.

[62] R. Horn and C. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[63] B. Wang and F. Zhang, "Some inequalities for the eigenvalues of the product of positive semidefinite hermitian matricies," *Linear algebra and its applications*, vol. 160, pp. 113–118, 1992.

[64] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, New York, 1985.

[65] R. Moses and R. Patterson, "Self-calibration of sensor networks," *Unattended Ground Sensor Technologies and Applications IV, Proc. SPIE vol. 4743*, pp. 108–119, 2002.

[66] J. N. Ash and R. L. Moses, "Acoustic time delay estimation and sensor network self-localization: Experimental results," *Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 841–850, 2005.

[67] A. J. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation, Part I: Narrow-band systems," *IEEE Transactions on Acoustics, Speech, and Signal-Processing*, vol. 31, no. 2, pp. 472–485, April 1983.

[68] ——, "A lower bound on the mean square error in random parameter estimation," *IEEE Transactions on Information Theory*, vol. IT-31, no. 5, pp. 680–682, September 1985.

[69] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. on Information Theory*, vol. 15, no. 3, pp. 386–391, 1969.

[70] D. Nicholson and A. Vecchio, "Bayesian bounds on parameter estimation accuracy for compact coalescing binary gravitational wave signals," *Phys. Rev. D*, vol. 57, no. 8, pp. 4588–4599, Apr 1998.

[71] M. Chu and G. Golub, *Inverse Eigenvalue Problems: Theory, Algorithms, and Applications*. Oxford University Press, 2005.

[72] J. N. Ash and L. C. Potter, "Robust system multiangulation using subspace methods," in *Proc. Information Processing in Sensor Networks*, April 2007, pp. 61–68.

[73] C. Savarese, J. M. Rabaey, and J. Beutel, "Locationing in distributed ad-hoc wireless sensor networks," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2037–2040., May 2001.

[74] A. Knyazev and M. Argentati, "Principal angles between subspaces in an a-based scalar product: Algorithms and perturbation estimates," *SIAM J. Sci. Comput.*, vol. 23, no. 6, pp. 2009–2041, 2002.

[75] H. Hotelling, "Relations between two sets of variables," *Biometrika*, vol. 28, pp. 321–377, 1936.

[76] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*, 5^{th} ed. Pearson, 2002.

[77] E. Candès, "Compressive sampling," *Proc. of the Int. Congress of Mathematicians, Madrid, Spain*, 2006.

[78] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[79] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2005.

[80] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[81] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[82] J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[83] J. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[84] C. He and J. Moura, "Robust detection with the gap metric," *IEEE Trans. Sig. Proc.*, vol. 45, no. 6, pp. 1591–1604, 1997.

[85] D. Rubin, "Inference and missing data," *Biometrika*, vol. 63, pp. 581–592, 1976.

[86] G. Verbeke and G. Molenberghs, *Linear mixed models for longitudinal data*. Springer-Verlag, 2000.