Optimized Confidence Weights for Localization Algorithms with Scarce Information

Giuseppe Destino and Giuseppe Thadeu Freitas de Abreu
Centre for Wireless Communications
University of Oulu, Finland
E-mails: destino@ee.oulu.fi, giuseppe@ee.oulu.fi

***Abstract*—A method to derive weights to be used in distance-based *multi-dimensional scaling* (MDS) source localization algorithms under scarce information is discussed. In particular, a family of weighing function is derived with basis on small-scale statistics and the parameter that drives the choice of a particular weighing function out of such a family is optimized with basis on an information-theoretical criterion. It is found that, under the condition of scarce information, the proposed weighing strategy outperforms the alternative of utilizing an approximation of the *maximum-likelihood* (ML) weighing strategy.**

I. Introduction

Source localization has recently emerged as an important research area in wireless communications [1]. Due to a number of factors, distance-based localization algorithms are particularly attractive alternative to solve such problems. Pairwise distance estimates can be obtained from either passive or active ranging techniques, which offer diversity of choice in terms of cost, robustness, etc. From a purely algorithmic point of view, however, one attractive feature of distance-based localization is availability of a powerful mathematical framework that allows the processing of distance estimates onto the coordinates of associated vertices, which is generally referred to as *multi-dimensional scaling* (MDS) [2].

MDS-based localization techniques vary from simple [3] (typically less accurate) to highly sophisticated [4] (usually more accurate). Regardless of the formulation, however, it has been found that one of the most powerful tools to improve flexibility and accuracy of MDS-based localization methods is the weigh that can be applied onto the input data in order to shape the cost-function associated with the problem, reducing the influence of bad data and enhancing the contribution of good data [4]–[7].

In this paper we discuss a concrete method to derive weights to be used in distance-based MDS source localization algorithms under the assumption that distance estimates are scarce. First, a family of weighing function is derived with basis on small-scale statistics (confidence bounds). Then, an information-theoretical criterion is derived and utilized to optimize the parameter that drives the selection of a particular weighing function out of such a family of curves. It is found that, under the condition of scarce information, the proposed weighing strategy outperforms the alternative of utilizing an approximation of the *maximum-likelihood* (ML) weighing strategy [4], [8]

II. Problem Formulation

Consider a wireless network of N devices (nodes), deployed in an $\eta-$dimensional space (2 or 3). Let N_a and $N_t = N - N_a$ be the number of anchor and target nodes, respectively. An anchor is a node whose location is known a priori, while a target is a node whose position is yet to be determined. Nodes are labeled with a unique number i from 1 to N and specifically, labels from 1 to N_a and from $N_a + 1$ to N are used for anchors and targets, respectively.

Let $\mathbf{x}_i \in \mathbb{R}^\eta$ be the i-th row-vector of the coordinate matrix $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ that contains the Euclidean coordinates of all nodes, and $\mathbf{D}$ denote the *Euclidean distance matrix* (EDM) of $\mathbf{X}$ whose (i,j)-th element d_{ij} is given by the Euclidean distance between the pair of nodes (i,j), $i.e.$,

$$\mathbf{D} = \mathcal{D}(\mathbf{X}) \triangleq [d_{ij}]\,, \tag{1}$$

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2, \tag{2}$$

where $\|\cdot\|_2$ denotes Euclidean norm.

The k-th measurement of d_{ij} can be modeled by

$$\tilde{d}_{ij,k} = f_{ij}(d_{ij} + \rho_{ij}, n_{ij,k}), \tag{3}$$

where $n_{ij,k}$ is typically a random number that captures uncertainty due to noise, while ρ_{ij} is an unknown constant describing bias due to deterministic channel conditions.

The function $f()$ shown above can take a number of different shapes. Most typically, a purely linear additive model $f(d + \rho, n) = d + \rho + n$ with a Gaussian n is employed [9], although other models have also been considered in some references [10], [11]. Even in the latter case, however, $d + \rho$ is typically significantly larger than the variance of n that $\tilde{d}$ has an "almost-Gaussian" distribution.

In *line-of-sight* (LOS) conditions ρ_{ij} is obviously 0 and in *non-line-of-sight* (NLOS) conditions ρ_{ij} is strictly positive. For example, in *time-of-arrival* (ToA) based techniques bias is caused by blockage of the direct path and the presence of a stronger secondary path, whereas in systems employing *received signal strength index* (RSSI) ranging approaches, usually indicates that the path-loss coefficient of the environment is higher than predicted by the model.

We employ a radio coverage disc model [12], where nodes have a limited transmission power P_{MAX}, or equivalently, a limited radio range R_{MAX}, such that the i-th node is connected to the j-th node if and only if (iff) $d_{ij} \le R_{\text{MAX}}$. In that case, we assume that a communication link e_{ij} between nodes i and j exists, and an estimate of the distance d_{ij} is available.

The set of measurable links is denoted by E and $\tilde{d}_{ij,k}$ will be arbitrarily set to zero for all pairs $(i,j) \notin E$, at no expense of accuracy. The *EDM sample* $\widetilde{\mathbf{D}}_k$ is the matrix that carries the collection of the k-th imperfect and incomplete pairwise distance samples.

The set of independent EDM samples is denoted by $\{\widetilde{\mathbf{D}}_k\}$. Given $\{\widetilde{\mathbf{D}}_k\}$, the localization problem can be defined by

$$\min_{\hat{\mathbf{X}} \in \mathbb{R}^{N\times\eta}} \left\| \mathbf{W} \circ \left(\mathcal{A}(\{\widetilde{\mathbf{D}}_k\})^{\circ 2} - \mathcal{D}(\hat{\mathbf{X}})^{\circ 2} \right) \right\|_{\mathrm{F}}^2, \qquad \text{(P – 1)}$$

where $\mathbf{W}$ are the weights, $\mathcal{A}(\cdot)$ is a function that aggregates the EDM samples, $\|\cdot\|_{\mathrm{F}}^2$ denotes Frobenius norm, and $\circ$ indicate the point-wise (Hadamard) power or product, respectively.

III. The Weighing Strategy

The problem summarized by equation (P – 1) is essentially a metric *multi-dimensional scaling* (MDS) problem [2], where dissimilarities are given by the Euclidean distances. A large number of techniques can be used to solve (P – 1), ranging from algebraic [3], to *semi-definite programming* (SDP) [4], to non-linear Newtonian methods [10]. In this article, however, we shall not focus on optimization strategies, but rather on the analytical derivation of the weighing matrix $\mathbf{W}$.

The role of the weighing matrix $\mathbf{W}$ is to shape the cost-function so as to bias the result of the minimization towards a solution that dependents more on better data. In other words, $\mathbf{W}$ should be such that more reliable samples influence the solution $\hat{\mathbf{X}}$ more than less reliable ones [4]–[6].

More precisely, if we define $\hat{d}_{ij} \triangleq \|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|_2$, then each weight w_{ij} must be such that the cost $\|\hat{\mathbf{d}}_{ij}^2 - \mathcal{A}(\{\tilde{d}_{ij,k}\})^2\|_{\mathrm{F}}^2$ is inversely proportional to the "quality" of the aggregated sample $\mathcal{A}(\{\tilde{d}_{ij,k}\})$. The questions that naturally arise at this point are: a) what the aggregation function $\mathcal{A}(\cdot)$ must be, and b) how can the quality of $\mathcal{A}(\{\tilde{d}_{ij,k}\})$ be quantified in a tangible manner. Although these questions are actually interconnected, a sufficiently independent answer to the former emerges from fundamental constraints of the problem.

In our particular context of large scale heterogeneous networks with scarce information, it is reasonable to assume that the ranging model (3) is unknown and may differ for each pair (i,j) – since networks typically spread across different environments and contain terminals of variable specifications – and sets $\{\tilde{d}_{ij,k}\}$ not only not only are small but also may vary in size. In the face of these conditions little choice for $\mathcal{A}(\cdot)$ exists beyond the simple average, *i.e.*

$$\mathcal{A}(\{\tilde{d}_{ij,k}\}) = \bar{d}_{ij} \triangleq \frac{1}{K_{ij}} \sum_{k=1}^{K_{ij}} \tilde{d}_{ij,k}, \qquad (4)$$

where K_{ij} is the cardinality of $\{\tilde{d}_{ij,k}\}$.

The aggregation strategy shown above has two major consequences of interest. The first is that it imposes undemanding assumptions on the data sets $\{\tilde{d}_{ij,k}\}$, namely, that the corresponding ranging models are stationary[1] for each $\{\tilde{d}_{ij,k}\}$, and that $\mathrm{E}[\tilde{d}_{ij}] = d_{ij} + \rho_{ij}$, where $\mathrm{E}[\cdot]$ denotes expectation.

The second is that it suggest an answer to question on the mechanism to quantify the quality of $\mathcal{A}(\{\tilde{d}_{ij,k}\}) = \bar{d}_{ij}$. Indeed, the quality of a sample mean $\bar{d}_{ij}$, given a small data $\{\tilde{d}_{ij,k}\}$, an unbiased estimator under LOS conditions $\mathcal{A}(\cdot)$, and no knowledge (but an estimate) of the channel state, amounts to the confidence that the true distance d_{ij} is within $\pm\gamma$ of the estimate $\bar{d}_{ij}$ *and* the confidence that the samples $\{\tilde{d}_{ij,k}\}$ are taken under LOS conditions. Thus,

$$\begin{aligned} w_{ij} &= \Pr\left\{\bar{d}_{ij} - \gamma \le d_{ij} \le \bar{d}_{ij} + \gamma \middle| \rho_{ij} = 0\right\} \cdot \Pr\left\{\rho_{ij} = 0\right\}, \\ &= \Pr\left\{-\gamma \le \varepsilon_{ij} \le \gamma\right\} \cdot \Pr\left\{\rho_{ij} = 0\right\}, \end{aligned} \qquad (5)$$

where $\varepsilon_{ij} \triangleq d_{ij} - \bar{d}_{ij}$ and the last equality results from a simple change of variables and the fact that the absence of bias (or existence of LOS conditions) is an event independent on the deviations of $\bar{d}_{ij}$ due to noise.

While weights computed via the principle summarized by equation (5) emphasize measurements taken under LOS conditions over those taken under NLOS, it should be clear that the approach does not preclude the possibility of bias compensation. In fact, one could easily iterate any particular localization (optimization) algorithm of choice, estimating bias from the previous solution and compensating the samples $\{\tilde{d}_{ij,k}\}$ at each iteration [10].

For the sake of mathematical convenience, we shall hereafter use w_{ij}^{L} and w_{ij}^{C} in reference to the two probabilities on the last equality of equation (5), such that $w_{ij} = w_{ij}^{\mathrm{L}} \cdot w_{ij}^{\mathrm{C}}$. Due to reasons to be clarified in the text to follow, w_{ij}^{L} and w_{ij}^{C} will also be dubbed the *dispersion weights* and *distortion weights*, respectively.

The focus of this article is on w_{ij}^{L}. In particular on the optimization of the parameter γ that drives the choice of weighing function out of the family derived with basis on small scale statistics method described above. Regarding distortion weights, we limit our discussion the following remarks.

In some cases, distortion weights can be computed utilizing special physical-layer capabilities of the nodes [1], [13] or the existence of sufficiently many samples ($K \gg 1$) [1].

In [10] we studied an interesting alternative to the above, which can be briefly described as follows. First, recall that bias due to NLOS manifests itself in the form of an increase on distance estimates. Also, due to the geometry of the problem, the occurrence NLOS condition on a link e_{ij} is independent on the occurrence NLOS on the links e_{iq} and e_{jq} between a third node q and the pair of nodes (i,j). Consequently, the occurrence of bias on a link e_{ij} tends to result in obtuse triangles $\triangle_{iqj}$. With that in mind, the fourth property [14] of the EDM of a meshy network can be used to indirectly estimate the likelihood that each of its links is in LOS condition. This is done employing Hypothesis Testing on the triangularity consistency of the sets of estimates $\{\tilde{d}_{ij}, \tilde{d}_{iq}, \tilde{d}_{qj}\}$, where $\{e_{ij}, e_{iq}, e_{qj}\} \in E$, and leads to an entire gradation of distortion weights[2] for all links in the network, as required by equation (5).

[1] There are techniques to eliminate biased samples from sets of mixed biased and unbiased samples [1, Sec. IV]. Our approach does not preclude the use of such techniques, which can be understood as a pre-processing filter.

[2] A related idea can be found in [15]. In the latter work, however, a hard decision on the LOS/NLOS is produced, which is inadequate in our context as it would imply binary weights

IV. Dispersion Weights

For simplicity, we shall in this section temporarily omit the subscript $_{ij}$ from the notation, with no sacrifice of clarity.

Under the assumption that $\rho = 0$ and that $\tilde{d}_k$ are independent, the dispersion weights w^{L} can be rewritten in the form

$$w^{\mathrm{L}} = 2 \cdot \Pr\{\varepsilon \leq \gamma\} - 1. \tag{6}$$

Given that the number of samples $\tilde{d}_k$ are typically small, and that the distribution of $\tilde{d}$ is almost Gaussian [9]–[11], ε approaches a t-Student distribution with $K-1$ degrees of freedom [16], [17]. Mathematically, $\varepsilon \sim p_{\mathrm{T}}(t; K-1)$, where the t-Student probability density function is shown below

$$p_{\mathrm{T}}(t;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})}. \tag{7}$$

Thus, equation (6) reduces to

$$w^{\mathrm{L}}(S, K; \gamma) = -1 + 2 \cdot \int_{-\infty}^{T} p_{\mathrm{T}}(t; K-1)\, dt, \tag{8}$$

where $T = \gamma \cdot \sqrt{K/S^2}$ is the so-called *T-score* [16], [17], with S^2 denoting the sample variance

$$S^2 = \frac{1}{K-1}\sum_{k=1}^{K}\left(\tilde{d}_k - \bar{d}\right)^2. \tag{9}$$

In equation (8), the dependence of w^{L} on S, K and γ are intentionally emphasized for reasons to become clear soon.

The recipe for w^{L} given above requires a final ingredient to be complete, namely, the value of γ. The remainder of this section is dedicated to the criterion and method to optimize γ.

First, recall the "rule of thumb" regarding the desirable values of w^{L} in association with the observations $\{\tilde{d}_k\}$. In the words of Costa *et al.* [6, Sec. 4.3], "the weighing function w^{L} should reflect the accuracy of measured dissimilarities." Biswas *et al.* [18, Sec. II] put the same principe as follows: "we [assign] higher weights to the distance [...] corresponding to the more reliable measures." Similar notions are invoked by numerous authors, $e.g.$, [3], [5], [19].

Next, notice that w^{L} according to equation (8) is a function of the sample variance S^2 and the number of samples K of the distance under consideration. Since each of these parameters carry a different type of information about the true value of d, it is unsurprising that both impact on the weight w^{L}. Specifically, w^{L} grows with the inverse of S^2 (for fixed K), and with K (for fixed S^2). These results are in accordance with the common argument outlined in the preceding paragraphs, since S^2 is proportional to the uncertainty of $\bar{d}$, as a measure of d, while K relates to the quality of $\bar{d}$ and S^2 as measures of d and its dispersion, respectively.

In order to move beyond this general intuition towards a concrete criterion for the optimization of γ, we recognize that the underlaying principle at hand is one of *diversity*. More than simply distinguishing reliable and unreliable measurements (more for better, less for worse), it is desirable that w^{L} enables a gradation of *all* sets of measurements according to their confidences *relative* to one another such that the rate of increase of w^{L} as a function of the relative confidence is optimum.

A well-known metric of the diversity of a population is *Shannon's diversity index* [20], defined by

$$H = -\int_{\mathcal{Z}} p_{\mathrm{Z}}(z) \cdot \ln(p_{\mathrm{Z}}(z))\, dz, \tag{10}$$

where $\mathcal{Z}$ is the support of $p_{\mathrm{Z}}(z)$, which is defined by the pairs $(K_{\min}, K_{\max})$ and $(S_{\min}, S_{\max})$, indicating the minimum and maximum number and standard deviation of samples typically observed across all data sets $\{d_{ij,k}\}\, \forall (i,j) \in E$.

Shannon's diversity index is commonly used as a measure of diversity in categorical data. In that context, the diversity notion is one of relative abundance of a category within the entire population, $i.e.$, which categories are more frequent than the others and by how much. In other words, the relevance of a category is measured by its occurrence in comparison to the others. Here, however, our "population" is the set of all possible pairs (S, K), the notion of diversity is one reliability (which distance estimates associated to a given pair (S, K) is more reliable than the others and by much), and the relevance of each pair is its weight w^{L}.

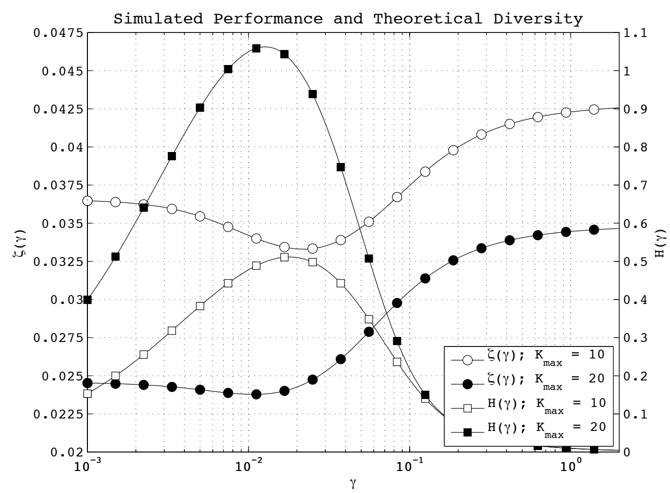

Fig. 1. Simulated rmse's and theoretical diversity indexes as function of γ.

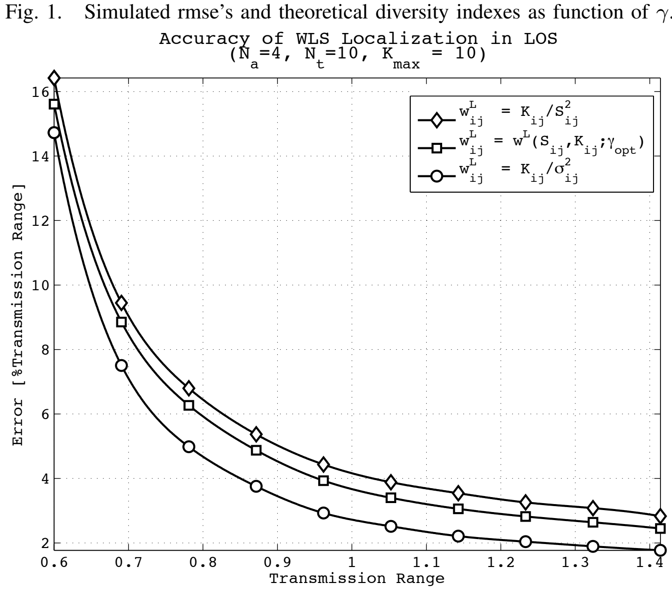

Fig. 2. Performance of WLS localization with different weighing strategies.

Notice also that the function in the integrand of equation (10) is the well-known *information entropy function*, and although this function is commonly used in signal processing, communications and information theory communities in association with probability measures, its notion can also be used in association with non-probabilistic measures [21].

From equations (8) and (10) we finally reach an information-theoretical optimization criterion for γ, namele

$$H(\gamma) = \sum_{k=K_{\text{min}}}^{K_{\text{max}}} \int_{S_{\text{min}}}^{S_{\text{max}}} w^{\text{L}}(S,k;\gamma) \cdot \ln\left(w^{\text{L}}(S,k;\gamma)\right)\, dS, \quad (11)$$

where the relationship with γ is made explicit for clarity.

Thus, the optimum γ is given by

$$\gamma_{\text{opt}} = \arg\max_{\gamma \in \mathbb{R}^+} \; H(\gamma). \qquad (\text{P} - 2)$$

In order to assess the performance of the weight optimization strategy proposed above, let us first define the *root-mean-square error* (rmse) of the node location estimates

$$\zeta(\gamma) \triangleq \frac{\|\mathbf{X}_t - \hat{\mathbf{X}}_t\|_{\text{F}}}{\sqrt{N_t}} \qquad (12)$$

where $\mathbf{X}_t$ and $\hat{\mathbf{X}}_t$ are the sub-matrices of $\mathbf{X}$ and $\hat{\mathbf{X}}$ containing only target-node coordinates and $\hat{\mathbf{X}}$ is the solution of (P – 1) with a particular γ used in the weighing functions, which justifies the notation of ζ as a function γ.

As an illustration, consider a 2-dimensional scenario with $N_a = 4$ and $N_t = 10$ deployed in a square with unitary sides, with anchors coordinates $\mathbf{X}_a = [(0,0);(1,0);(1,1);(0,1)]$. Full connectivity is assumed ($R_{\text{MAX}} \geq \sqrt{2}$) and distance measurements are generated using the additive error model described in section II with $\rho_{ij} = 0$, $n_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$ and $\sigma_{ij} \sim \mathcal{U}(0.01, 0.2)$, that is, LOS conditions with uneven Gaussian disturbance on distance estimates. The number of measurements of each distance d_{ij} is a uniformly distributed random $K_{ij} \sim \mathcal{U}(2, K_{\text{max}})$ with $K_{\text{max}} \in \{10, 20\}$ (two cases).

Plots of the rmse $\zeta(\gamma)$ and diversity index $H(\gamma)$ as functions of γ are shown in figure 1. It is found that the values of γ for which $H(\gamma)$ is maximized are very close to those for which $\zeta(\gamma)$ is minimized, which validates the analytical weight optimization method described earlier.

In figure 2 the performance of the WLS localization algorithm [10] with the optimized deviation weighing function here proposed is compared to those obtained using $w_{ij}^{\text{L}} = K_{ij}/\sigma_{ij}^2$ and $w_{ij}^{\text{L}} = K_{ij}/S_{ij}^2$. The latter weights are referred to as the ML weights with perfect and estimated variances, respectively. This weighing strategy is inspired by [4], where it was shown that under an additive zero-mean Gaussian error model and with a ML formulation of the localization problem, the optimum weights are given by the inverse of the variance of the distance estimates. In our context, the distance estimates are the average $\bar{d}_{ij}$, taken over K_{ij} samples, which not only justifies the scaling factors K_{ij}, but more importantly imply that the variances σ_{ij}^2 *cannot* be known *a priori*. In other words, the results with $w_{ij}^{\text{L}} = K_{ij}/\sigma_{ij}^2$ must be interpreted as a theoretical (unachievable) lower-bound, in contrast to those with $w_{ij}^{\text{L}} = K_{ij}/S_{ij}^2$, which are the results one would obtain in practice.

It is noticeable that our weighing strategy outperforms the "practical" ML. This is a consequence of the fact that the latter alternative does not make use sufficient of the "information" that K_{ij} provides on the quality of S_{ij}^2 as an estimate of σ_{ij}^2.

References

[1] B. F. Guoqiang Mao and B. D. O. Anderson, "Wireless sensor network localization techniques," 2006, accepted.

[2] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*, 2nd ed. Chapman & Hall/CRC, 2000.

[3] Y. Shang and W. Ruml, "Improved mds-based localization," in *Proc. 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom'04)*, vol. 4, Hong-Kong, China, March, 7-11 2004, pp. 2640 – 2651.

[4] P. Biswas, T.-C. Liang, K.-C. Toh, and T.-C. Wang, "Semidefinite programming based algorithms for sensor network localization with noisy distance measurements," Dept. of Management Science Engineering, Stanford University, Tech. Rep., 2005.

[5] A. Y. Alfakih, H. Wolkowicz, and A. Khandani, "Solving euclidean distance matrix completion problems via semidefinite programming," *Journal on Computational Optimization and Applications*, vol. 12, no. 1, pp. 13 – 30, 1999. [Online]. Available: http://portal.acm.org/citation.cfm?id=316254\#

[6] J. A. Costa, N. Patwari, and A. O. H. III, "Distributed weighted-multidimensional scaling for node localization in sensor networks," pp. 1–26, June 2005. [Online]. Available: http://www-personal.engin.umich.edu/∼npatwari/wmds.pdf

[7] G. Destino and G. T. F. de Abreu, "Localization from imperfect and incomplete ranging," in *Proc. IEEE* 17^{th} *International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06)*, Septermber 11-14 2006.

[8] B. Denis, L. He, and L. Ouvry, "A flexible distributed maximum log-likelihood scheme for UWB indoor positioning," in *IEEE 4th Workshop on Positioning, Navigation and Communication*, 2007.

[9] M. Wylie and J. Holtzmann, "The non-line-of-sight problem in mobile location estimation," in *5th IEEE International Conference on Universal Personal Communications*, vol. 2, 1996, pp. 827–831.

[10] G. Destino, D. Macagnano, and G. Abreu, "Hypothesis testing and iterative wls minimization for wsn localization under los/nlos conditions," in *IEEE Asilomar Conference on Signals, Systems, and Computers*, Nov. 4-7 2007.

[11] S. Gezici, "A survey on wireless position estimation," *Wireless Personal Communincations*, vol. 44, no. 3, pp. 263–282, Feb. 2008.

[12] R. D'Souza, D. Galvin, C. Moore, and D. Randall, "Global connectivity from local geometric constraints for sensor networks with various wireless footprint," in *Proc. IEEE/ACM International Conference on Information Processing in Sensor Networks (IPSN'06)*, Apr. 19 - 21 2006.

[13] J. Schroeder, S. Galler, K. Kyamakya, and K. Jobmann, "Nlos detection algorithms for ultra-wideband localization," in *4th Workshop on Positioning, Navigation and Communication*, March 2007, pp. 159–166.

[14] J. Dattorro, *Convex Optimization and Euclidean Distance Geometry*. Meboo Publishing, 2005. [Online]. Available: http://www.stanford.edu/∼dattorro/mybook.html

[15] L. Girod and D. Estrin, "Robust range estimation for localization in ad-hoc sensor networks," November 2000.

[16] C. Walck, *Handbook on Statistical Distributions for Experimentalists*. Stockholm, Sweden: University of Stockholm Press, 2000.

[17] J. D. Gibbons, *Nonparametric Statistical Inference*. Marcel Dekker, 1992.

[18] P. Biswas, T.-C. Liang, K.-C. Toh, T.-C. Wang, and Y. Ye, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," in *IEEE Transactions on Automation Science and Engineering*, ser. 1545-5955, vol. 3, Oct. 2006, pp. 360–371.

[19] T. Graepel, "Kernel matrix completion by semidefinite programming," in *International Conference on Artificial Neural Networks*, J. R. Dorronsoro, Ed., vol. Vol. 2415. Springer-Verlag, August 28 - 30 2002, pp. 694 – 699.

[20] C. A. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, vol. 27, pp. 379 – 423, Jul. 1948.

[21] C. Bertoluzza and C. Poggi, "Non-probabilistic uncertainty measures," in *Proc. IEEE* 1^{st} *International Symposium on Uncertainty Modeling and Analysis*, Dec. 3 - 5 1990, pp. 647–651.