Non-Parametric Non-Line-of-Sight Identification[1]

Sinan Gezici, Hisashi Kobayashi and H. Vincent Poor
Department of Electrical Engineering
School of Engineering and Applied Science
Princeton University, Princeton, NJ 08544
{sgezici,hisashi,poor}@princeton.edu

Abstract— **Recently, there has been much interest in accurate determination of mobile user locations in cellular environments. A general approach to this geolocation problem is to gather time-of-arrival measurements from a number of base stations (BSs) and to estimate user locations using the traditional least square approach. However, in non-line-of-sight (NLOS) situations, measurements are significantly biased. Hence, very large errors in location estimation may be introduced when traditional techniques are adopted. For this reason, before employing an algorithm for location estimation, it is useful to know which BS's are in line-of-sight (LOS) and which are in NLOS of the mobile station. In this paper, a non-parametric approach to this NLOS identification problem is proposed. Since the statistics of NLOS errors are usually unknown, a non-parametric probability density estimation technique is employed to approximate the distribution of the measurements. Then, an appropriate metric is used to determine the distance between the distribution of the measurements and the distribution of the measurement noise. Depending on the closeness of the distributions, the propagation environment is classified as LOS or NLOS. In a situation where reliability of measurements from a BS is to be quantified, the distance can be used to represent the reliability of the measurements as well as to classify the station.**

I. Introduction

Recently, the subject of mobile positioning in wireless communication systems has drawn considerable attention. With accurate location estimation, a variety of new applications and services such as Enhanced-911, location sensitive billing, improved fraud detection, intelligent transport system (ITS) and improved traffic management will become feasible [2].

Mobile positioning using radiolocation techniques usually involves time of arrival (TOA), time difference of arrival (TDOA), angle of arrival (AOA), or signal strength (SS) measurements, or some combination of these methods. Multipath, non-line-of-sight (NLOS) propagation and multiple access interference are often the main sources of errors in geolocation, and make mobile positioning challenging. Among these error sources, NLOS is perhaps the most crucial one.

In an NLOS situation, TOA measurements that are used to estimate the distance between a mobile station (MS) and a base station (BS) are severely biased. In this case, using traditional location algorithms may result in large errors in location estimation [1]. However, if it is known that the MS is in NLOS with respect to a BS, then some special methods can be applied depending on the scenario. For example, if at least three BS's are in LOS of the MS, then the measurements from the NLOS BS's can be discarded in obtaining a two-dimensional location estimate. Alternatively, when the Recursive Weighting Algorithm [2] is employed to reduce NLOS propagation errors, knowledge of LOS/NLOS BS's becomes important. In other words, identification of NLOS BS's can help considerably to improve the location estimation.

The problem of NLOS identification is essentially a detection problem. It compares the LOS hypothesis to the NLOS hypothesis. The probability distribution of the measurements under the LOS hypothesis is usually known except for its mean. If the distribution under NLOS hypothesis is also assumed to be known, then the problem can be solved by the conventional hypothesis testing method [3]. However, the probability distribution of NLOS errors, hence that of the measurements under the NLOS hypothesis, is usually unknown. Therefore, a technique which does not assume the knowledge of NLOS error statistics should be developed.

In this paper, a non-parametric NLOS identification technique is proposed. Since the statistics of TOA delays due to NLOS are not known exactly, a non-parametric approach is adopted to approximate the probability density function of the measurements. Then, a suitable distance metric between a known measurement error distribution and a non-parametrically estimated distance measurement distribution is defined to determine whether a given BS is within LOS or NLOS of the MS. The distance between these two distributions can also be used as a reliability measure for the measurements from the given BS.

The remainder of the paper is organized as follows. Section II formulates the problem of NLOS identification and describes the non-parametric NLOS identification algorithm. The performance of the algorithm is evaluated in Section III by simulation studies. Finally, some concluding remarks are made in Section IV.

II. Non-Parametric NLOS Identification

Consider a situation in which m independent identically distributed (iid) range measurements (obtained from TOA measurements multiplied by the speed of light) between an MS and a BS are taken. Assume that the change in the location of the MS during these measurements can be ignored. Hence the distance between the MS and the BS can be considered approximately constant for the geolocation purpose. Then, for

[1]This research was supported in part by the U.S. National Science Foundation under Grant CCR-99-79361, and in part by the New Jersey Center for Wireless Telecommunications.

the ith measurement, the hypotheses can be expressed as:

$$\begin{aligned} H_0 &: \quad r_i = d + n_i \\ H_1 &: \quad r_i = d + n_i + e_i, \end{aligned} \tag{1}$$

for $i = 1, ..., m$, where H_0 is the LOS hypothesis and H_1 is the NLOS hypothesis. In the former case, the measurement is modelled as the summation of the true distance d and a measurement noise, n_i, while in the latter case, the NLOS error e_i is also present, which is modelled by a positive random variable.

We assume that the measurement noise statistics are completely known and is modelled by a zero mean Gaussian random variable. However, neither d nor the probability density function of the NLOS error are known. Therefore, it is not possible to invoke conventional hypothesis testing techniques like generalized likelihood ratios.

Let the probability density function (pdf) of the measurement noise be $p_n(x)$, which is completely known. Then, the pdf of the measurements in the LOS hypothesis case is given by $p_n(x-d)$. Note that this distribution is completely known except for one parameter, d, which affects only the mean of the distribution. The main idea in the non-parametric NLOS identification test is to compare the closeness of this pdf to the pdf of range measurements. Thus we first approximate the pdf of the range measurements non-parametrically, compare the closeness of this pdf to the LOS pdf by defining a distance metric, and then decide LOS/NLOS after a threshold test. This test can be summarized as follows:

1) Estimate the pdf of the distance from m iid range measurements. Let this estimate be denoted by $\hat{p}_r(x)$.

2) Calculate the distance between $p_n(x-d)$ and $\hat{p}_r(x)$ for all possible d values and find the minimum distance.

3) Compare this minimum distance to a threshold: Decide H_0 if the minimum distance is smaller than the threshold, and decide H_1 otherwise.

We will discuss these steps in more details in the following subsections.

A. Parzen Window Density Estimation

In order to estimate the pdf of the distance, a non-parametric density estimation technique, called Parzen window density estimation, is employed, which approximates the pdf using some window functions around the samples. The reason for employing this technique is its flexibility in choosing density estimation parameters depending on the sample size.

Given iid distance measurements $r_1, ..., r_m$, the distance pdf can be estimated by the following formula [4]

$$\hat{p}_r(x) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{h_m} \phi(\frac{x - r_i}{h_m}), \tag{2}$$

where $\phi(.)$ is the window function and h_m is a scaling parameter. The window function must be a probability density function in order for $\hat{p}_r(x)$ to be a valid pdf. In other words, it is always non-negative and integrates to one. Commonly used window functions include Gaussian and rectangular windows [4].

B. Distance Function

After obtaining the approximate pdf of the distance, our next step is to determine whether these distance measurements are coming from $p_n(x-d)$ or the pdf under the NLOS hypothesis. Since the pdf under the NLOS hypothesis is unknown, it is reasonable to compare the distance between $p_n(x-d)$ and $\hat{p}_r(x)$ and accept the LOS hypothesis if the distance is smaller than a threshold, that is, if the two distributions are sufficiently close. Since the true distance, d, is unknown, the minimum distance between $p_n(x-d)$ and $\hat{p}_r(x)$ must be calculated among all possible d's.

The Kullback-Leibler (KL) distance [5] can be used to calculate the distance between two probability distributions. For given pdf's p_1 and p_2, the KL distance between them is given by

$$D(p_1 \| p_2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} \, dx. \tag{3}$$

C. Decision Criterion

The decision criterion to determine LOS or NLOS hypothesis becomes the following test:

$$\inf_{d} \; \{D(\hat{p}_r(x) \| p_n(x-d))\} \; \mathop{\lesseqgtr}_{H_1}^{H_0} \; \delta, \tag{4}$$

where δ is the threshold.

If the value of the d minimizing the decision variable can be found, the test can be expressed simply as

$$D(\hat{p}_r(x) \| p_n(x-\hat{d})) \; \mathop{\lesseqgtr}_{H_1}^{H_0} \; \delta. \tag{5}$$

We assume that the measurement noise is a zero mean Gaussian random variable, which is a valid approximation when the TOA's are acquired with a matched filter approach at high signal-to-noise ratio (SNR) [6]. Then, $p_n(x-d)$ is expressed as

$$p_n(x-d) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-d)^2/(2\sigma^2)}. \tag{6}$$

In this case, the following result indicates the simplification of the decision test.

Proposition 2.1 *For a zero mean Gaussian measurement error and a symmetric window function, i.e.,* $\phi(x) = \phi(-x)$ *for all* x*, the value of* d *minimizing the distance function of (4) is the sample mean of the measurements, that is,* $\hat{d} = \frac{1}{m}\sum_{i=1}^{m} r_i$.

Proof *See Appendix A.*

Proposition 2.1 states that for a symmetric window function, the minimum distance to be used in the decision criterion can be computed by simply shifting the Gaussian measurement error pdf by the sample mean of the measurements and calculating the KL distance between this shifted pdf and the estimated pdf, $\hat{p}_r(x)$.

Another important issue is the appropriate choice of the threshold value, δ. Since the pdf's are not known exactly under either hypothesis, it does not seem possible to set the "false alarm" (i.e. misinterpret a LOS situation as NLOS) and "miss detection" (i.e. misinterpret an NLOS situation as LOS) probabilities. However, the following result states that in some situations the false alarm probability can be set even though the true distance d is not known, that is, without any information about the mean of the random variable under H_0.

Proposition 2.2 *For a zero mean Gaussian measurement error and a symmetric window function, the false alarm probability can be set independently of the true distance between the mobile and the base station.*

Proof *See Appendix B*

Proposition 2.2 states that under suitable conditions the distance function is independent of the true distance under the LOS hypothesis and it is therefore theoretically possible to set the false alarm rate.

Under the conditions stated in the above two propositions, the decision test can be expressed as follows (Appendix C):

$$\int \hat{p}_r(x) \log(\sqrt{2\pi}\sigma\hat{p}_r(x))dx + \frac{\hat{\sigma}^2}{2\sigma^2} \underset{H_1}{\overset{H_0}{\lesseqgtr}} \delta', \qquad (7)$$

where $\hat{\sigma}^2$ is the sample variance[2] of the range measurements, that is $\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(r_i - \hat{d})^2$.

Depending on the technique to locate the mobile user, the classification of BS's may not be necessary. Instead some reliability information about the measurements from each BS might be required. In this case, the distance value between the LOS and NLOS pdf's can be used as a reliability information, which can help us to locate the mobile more accurately.

III. Simulation Results

In this section, the performance of the non-parametric NLOS BS identification technique is evaluated.

Figure 1 shows the false alarm probability, P_{FA}, of the non-parametric technique for different numbers of measurements ($m = 3, 5, 10, 15$) when the miss detection probability, P_{MD}, is set to 0.05. The measurement noise is modelled by $\mathcal{N}(0, 100m^2)$ and the NLOS error is modelled by an exponential random variable with mean $25m$. The unit variance Gaussian window is used and the scaling parameter h_m is set to 20. In this scenario, we see that we do not need many samples to have a reliable decision.

In order to evaluate the performance of the technique for different NLOS errors, we plot the false alarm probability for different NLOS errors. Specifically, we change the mean of the exponential random variable representing the NLOS error and plot P_{FA} when $P_{MD} = 0.10$ and the number of samples m is 10. The Gaussian measurement noise and the parameters of the Parzen window density estimation technique are kept the same as in the previous case. Figure 2 shows the performance of the technique for different NLOS noise levels. Obviously, the test becomes more successful as the mean (hence the variance) of the exponential random variable increases since it becomes easier to distinguish between the two hypotheses.

[2]The sample variance is often defined as $s^2 = \frac{1}{m-1}\sum_{i=1}^{m}(r_i - \bar{r})^2$, where $\bar{r}$ is the sample mean. This definition makes s^2 an unbiased estimate of the population variance.

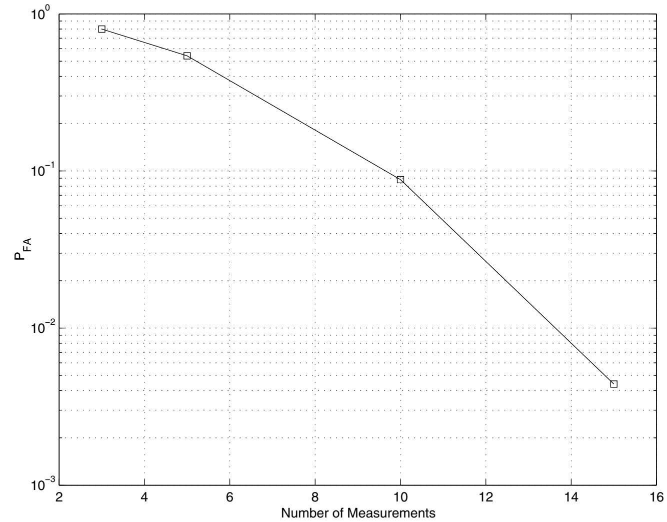

Fig. 1. False alarm probability versus the number of measurements when $P_{MD} = 0.05$.

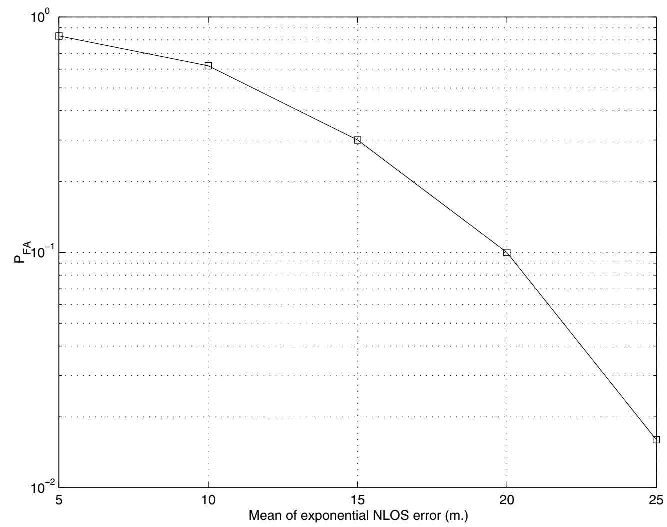

Fig. 2. False alarm probability versus NLOS errors when $P_{MD} = 0.10$ and $m = 10$.

IV. Conclusion

A non-parametric test to determine whether a given BS is in LOS or NLOS of the MS in question has been proposed. A suitable distance metric between a known measurement error distribution and a non-parametrically estimated distance distribution has been defined to determine whether a given BS is within LOS or NLOS of the MS. The performance of the algorithm has been evaluated by simulation experiments. Future work includes more detailed simulation studies under different NLOS scenarios and comparisons to other techniques [3].

REFERENCES

[1] J. Caffery, Jr. and G. Stuber, "Subscriber location in CDMA cellular networks," *IEEE Trans. Vehicular Technology*, vol. 47, pp. 406-416, May 1998.

[2] J. Caffery, Jr., *Wireless location in CDMA cellular radio systems,* Kluwer Academic Publishers, Boston, 2000.

[3] J. Borras, P. Hatrack and N. B. Mandayam, "Decision theoretic framework for NLOS identification," *Proc. IEEE Vehicular Technology Conference, Spring 1998*, vol. 2, Ottawa, Canada, pp. 1583-1587.

[4] R. O. Duda, P. E. Hart and D. D. Stork, *Pattern classification,* Wiley-Interscience, New York, 2001.

[5] S. Kullback, *Information theory and statistics,* Wiley, New York, 1959.

[6] C. E. Cook and M. Bernfeld, *Radar signal: an introduction to theory and application,* Academic Press, 1970.

APPENDIX

A. Proof of Proposition 2.1

Using (3), the KL distance between $\hat{p}_r(x)$ and $p_n(x-d)$ can be expressed as

$$D(\hat{p}_r(x)\|p_n(x-d)) = \int \hat{p}_r(x)\,\log\frac{\hat{p}_r(x)}{p_n(x-d)}\,dx. \qquad (8)$$

After inserting (6) and some manipulations, we get

$$\begin{aligned} D(\hat{p}_r(x)\|p_n(x-d)) &= \int \hat{p}_r(x)\,\log(\sqrt{2\pi}\sigma\,\hat{p}_r(x))\,dx \\ &+ \frac{1}{2\sigma^2}\int (x-d)^2\hat{p}_r(x)\,dx. \qquad (9)\end{aligned}$$

When we differentiate (9) with respect to d and equate to zero, we get

$$\hat{d} = \frac{\int x\,\hat{p}_r(x)\,dx}{\int \hat{p}_r(x)\,dx}. \qquad (10)$$

Since $\hat{p}_r(x)$ is a probability density function, the denominator is equal to unity. For the numerator, if we consider a symmetric window function,

$$\begin{aligned} \int x\,\hat{p}_r(x)\,dx &= \frac{1}{mh_m}\sum_{i=1}^{m}\int x\phi\left(\frac{x-r_i}{h_m}\right)dx \\ &= \frac{1}{m}\sum_{i=1}^{m} r_i \qquad (11)\end{aligned}$$

where we use the fact that window functions integrate to unity.

So the optimal value of d is the sample mean of the measurements,

$$\hat{d} = \frac{1}{m}\sum_{i=1}^{m} r_i, \qquad (12)$$

which minimizes (8).

B. Proof of Proposition 2.2

The false alarm probability is the probability that the KL distance between $\hat{p}_r(x)$ and $p_n(x-d)$ exceeds the threshold given that H_0 is the true hypothesis. That is,

$$\mathrm{P_{FA}} = \Pr\{D(\hat{p}_r(x)\|p_n(x-\hat{d})) > \delta \mid H_0\}. \qquad (13)$$

Under H_0, each measurement is equal to sum of the true distance and measurement noise, which is a zero mean Gaussian random variable. In other words, $r_i = d + n_i$ for $i = 1, ..., m$. From Proposition 2.1, the optimal value of d minimizing the KL distance of (4) is given by $\hat{d} = \frac{1}{m}\sum_{i=1}^{m} r_i$.

Inserting (2) and (6) in (5), we get

$$\begin{aligned} &D(\hat{p}_r(x)\|p_n(x-\hat{d})) = \\ &\frac{1}{mh_m}\sum_{i=1}^{m}\int \phi(\frac{x-r_i}{h_m})\,\log\left(\frac{1}{mh_m}\sum_{i=1}^{m}\phi(\frac{x-r_i}{h_m})\right)dx - \\ &\frac{1}{mh_m}\sum_{i=1}^{m}\int \phi(\frac{x-r_i}{h_m})\log\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\hat{d})/(2\sigma^2)}\right)dx \qquad (14)\end{aligned}$$

Note that $x-\hat{d} = x - \frac{1}{m}\sum_{i=1}^{m} r_i = x - d - \frac{1}{m}\sum_{i=1}^{m} n_i$, since $r_i = d + n_i$ under H_0. So defining a new dummy variable for the integrals as $u = x - d$, we obtain

$$\begin{aligned} &D(\hat{p}_r(x)\|p_n(x-\hat{d})) = \\ &\frac{1}{mh_m}\sum_{i=1}^{m}\int \phi(\frac{u-n_i}{h_m})\,\log\left(\frac{1}{mh_m}\sum_{i=1}^{m}\phi(\frac{u-n_i}{h_m})\right)du - \\ &\frac{1}{mh_m}\sum_{i=1}^{m}\int \phi(\frac{u-n_i}{h_m})\,\log\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-(u-\frac{1}{m}\sum_{i=1}^{m} n_i)/(2\sigma^2)}\right)du,\end{aligned}$$

which solely depends on $n_1, ..., n_m$. Therefore, it is possible to set the false alarm rate without any information about the true distance, d.

C. Derivation of Equation (7)

The decision test in the Gaussian measurement noise case can be expressed as

$$D(\hat{p}_r(x)\|p_n(x-\hat{d})) \underset{H_1}{\overset{H_0}{\lesseqgtr}} \delta, \qquad (15)$$

where $\hat{d} = \frac{1}{m}\sum_{i=1}^{m} r_i$.

Similar to (9), the distance can be expressed as

$$\begin{aligned} D(\hat{p}_r(x)\|p_n(x-\hat{d})) &= \int \hat{p}_r(x)\,\log(\sqrt{2\pi}\sigma\,\hat{p}_r(x))\,dx \\ &+ \frac{1}{2\sigma^2}\int (x-\hat{d})^2\hat{p}_r(x)\,dx. \qquad (16)\end{aligned}$$

The integral in the second term can be expressed as follows:

$$\int (x-\hat{d})^2\hat{p}_r(x)\,dx = \frac{1}{mh_m}\sum_{i=1}^{m}\int (x^2-2\hat{d}x+\hat{d}^2)\phi(\frac{x-r_i}{h_m})dx, \qquad (17)$$

which can be shown equal to

$$\int (x-\hat{d})^2\hat{p}_r(x)\,dx = \frac{1}{mh_m}\sum_{i=1}^{m}\{h_m[h_m^2\sigma_w^2+r_i^2]-2\hat{d}r_ih_m+\hat{d}^2h_m\}, \qquad (18)$$

where $\sigma_w^2 = \int x^2\phi(x)dx$. After combining the terms we get

$$\int (x-\hat{d})^2\hat{p}_r(x)\,dx = h_m^2\sigma_w^2 + \hat{\sigma}^2, \qquad (19)$$

where $\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(r_i - \hat{d})^2$ is the sample variance of the measurements. Inserting (19) in (16), the final decision test can be expressed as in (7) where $\delta^{'} = \delta - h_m^2\sigma_w^2/(2\sigma^2)$.