Technical University of Crete

Machine Learning: Final Report

Panos N. Alevizos

Course Instructor: Michail G. Lagoudakis

Thursday 3rd March, 2016

1 Introduction

Scatter radio technology has emerged as a potential key-enabling technology for low-power, cost-effective communication systems. Radio frequency identification (RFID) applications [1] as well as bistatic scatter radio wireless sensor networking [2] are the most notorious examples of scatter radio technology. The rapid advent of Internet-of-Things (IoT) idea leverages the adoption of scatter radio technology imperative in future applications

Location-awareness has become an essential aspect of ubiquitous sensing applications, for instance, environmental and agricultural monitoring [3], health-care monitoring [4], and target tracking scenarios [5]. Localization becomes more challenging in harsh indoor environments where the use of global position systems (GPS) technology becomes prohibitory due to signal blockage or/and severe signal attenuation.

Consequently, in harsh indoor environments consisting of massively deployed IoT devices localization may be conducted by the use of scatter radio technology [6]. However, as classic RFID can offer reliable communication in a limited range of 5 to 8 meters [7], the vision of establishing accurate localization over ranges of more than 10 meters is hindered. With more than 140 meters reliable communication range [2], bistatic scatter radio architecture with semi-passive radio frequency (RF) tags has become the imminent low-cost and low-power communication technology that can achieve the goal of almost-zero power, increased range indoor localization. In contrast to classic monostatic architecture, encountered in RFID, in bistatic architecture, the antenna of transmitter (also called carrier emitter) and receiver (also called reader) are dislocated.

Unfortunately, in harsh indoor environments the presence of walls, doors, and other obstacles signifies severe signal attenuation and shadowing, which in turn offers biased range estimates. Thus, the solution of the following two critical problems becomes the core of indoors localization:

- Distinguish between line-of-sight (LoS) and non-line-of-sight (NLoS) signals.
- Mitigate the effects of NLoS signals which usually add a positive bias on range estimates.

The former is known as NLoS identification, and the latter is called NLoS mitigation.

Prior art has utilized plethora of machine learning, as well as, estimation and detection theory techniques to approach the above two problems. For NLoS identification work in [8] has utilized the Neyman-Pearson test for angle-of-arrival (AoA), time-of-arrival (ToA), and received signal strength (RSS) measurements and offered closed form expressions for the probability of detection and false alarm. Work in [9] utilizes support vector machine (SVM) framework for classification and regression, as well as hypothesis testing framework for both NLoS identification and mitigation in the presence of RSS measurements. The choice among machine learning and hypothesis testing is based on different user requirements and information available.

Related work [10, 11] exploits the power efficiency, fine delay resolution, and robust operation of ultra wide-band (UWB) technology in harsh environments. Work [10] applies SVM classification NLoS identification and SVM regression NLoS mitigation, while work in [11] uses relevance vector machine (RVM) framework for NLoS identification and mitigation. In both works the choice of features relies directly on functions that depend explicitly on received signal, e.g., energy and maximum amplitude of the received signal, rise time, mean excess delay, to name a few.

Due to the limited range of passive RFID, NLoS effects cannot even identified in a tag-to-software defined radio (SDR) reader distance of 5 meters. Thus, without the limited range as bottleneck, bistatic scatter radio architecture with almost zero-power semi-passive RF tags is the

most promising communication technology for indoor localization. In this work a three-device bistatic scatter radio architecture is employed, that consists of a carrier emitter, a sensor tag and a SDR reader (see for example Fig. 1). Adopting the aforementioned bistatic architecture, this work applies state-of-the-art machine learning algorithms for NLoS identification signals in scatter radio systems. In contrast to [9], where the moments of RSS measurements are utilized as features, this work applies directly the machine learning classification algorithm in RSS measurements.

Real experimental RSS ranging measurements are conducted indoors at different distance setups. Using directly the gathered RSS measurements, this work aims to (a) provide an estimate path-loss exponents (PLEs) for the nontrivial scatter radio signal model, (b) compare different machine learning classification algorithms in terms of how good can distinguish LoS and NLoS measurements. Although, the proposed PLE estimation framework as well as machine learning classification algorithms will be applied on the context bistatic architecture, both of them can be trivially applied to the monostatic RFID architecture, as a simpler case of the bistatic one.

In this work four classification algorithms are studied and implemented for the purposes of NLoS identification; generative Gaussian classification (GGC), logistic regression (LR), SVM classification and RVM classification. Several type of feature mappings (also known as basis functions) on the input data are tested on all algorithms. In addition to the above, for SVM and RVM classification framework, the corresponding kernelized versions are also studied and implemented. After several simulations, it is deduced that only Gaussian kernels can offer very small misclassification error. It is concluded that SVM classification method offers the best error performance acrosss the majority of different feature mapping and kernel simulation scenarios.

1.1 Notation

The set of real numbers is denoted as $\mathbb{R}$, the binary field is denoted as $\mathbb{B} \triangleq \{0,1\}$. Nonbold lower-case letters (e.g., x or $\mathbf{x}$) will stand for variables. Vectors and matrices will be denoted by lower-case (e.g., $\mathbf{x}$) and capital (e.g., $\mathbf{A}$), respectively, bold characters. Symbol $(\cdot)^\top$ denotes the transpose a of a vector or matrix. The inverse of an invertible matrix $\mathbf{A}$ is denoted $\mathbf{A}^{-1}$. For a vector $\mathbf{x} = \begin{bmatrix} x_{[1]} & x_{[2]} & \dots & x_{[L]} \end{bmatrix}^\top \in \mathbb{R}^L$, we will adopt notation $x_{[l]}$ to denote its lth element, while notation $\mathbf{x}_{[l_1:l_2]}$ will stand for the sub-vector of $\mathbf{x}$ that consists of the l_1th up to l_2th elements of it. Similarly, for a matrix $\mathbf{X}$, $X_{[l_1,l_2]}$ will denote its (l_1,l_2)th entry. The all zeros (ones) vector of size L will be denoted by $\mathbf{0}_L$ ($\mathbf{1}_L$). The all zeros (ones) matrix of size $L_1 \times L_2$ will be denoted by $\mathbf{0}_{L_1\times L_2}$ ($\mathbf{1}_{L_1\times L2}$). The $L \times L$ identity matrix is denoted by $\mathbf{I}_L$. Notation $\mathsf{tr}\{\mathbf{A}\}$ is the trace operator of square matrix $\mathbf{A}$. Symbol $\otimes$ denotes the Kronecker product, which for two matrices $\mathbf{A} \in \mathbb{R}^{L_1\times L_2}$ and $\mathbf{B} \in \mathbb{R}^{L_3\times L_4}$ is defined as

$$\mathbb{R}^{(L_1\cdot L_3)\times(L_2\cdot L_4)} \ni \mathbf{C} = \mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} A_{[1,1]}\mathbf{B} & A_{[1,2]}\mathbf{B} & \cdots & A_{[1,L_2]}\mathbf{B} \\ A_{[2,1]}\mathbf{B} & A_{[2,2]}\mathbf{B} & \cdots & A_{[2,L_2]}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A_{[L_1,1]}\mathbf{B} & A_{[L_1,2]}\mathbf{B} & \cdots & A_{[L_1,L_2]}\mathbf{B} \end{bmatrix} \tag{1.1}$$

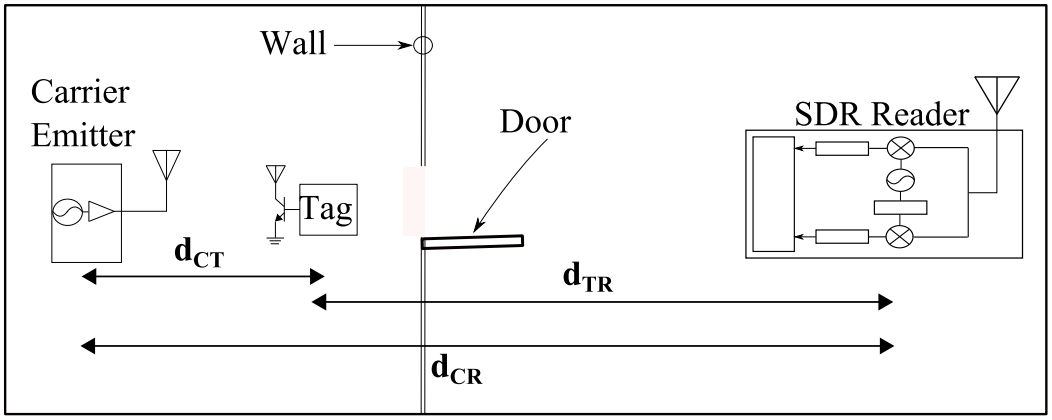

Figure 1: LoS setup.

Symbol $\odot$ stands for the Hadamard (or element-wise product); for two matrices $\mathbf{A} \in \mathbb{R}^{L_1 \times L_2}$ and $\mathbf{B} \in \mathbb{R}^{L_1 \times L_2}$ it is defined as

$$\mathbb{R}^{L_1 \times L_2} \ni \mathbf{C} = \mathbf{A} \odot \mathbf{B} = \begin{bmatrix} A_{[1,1]}B_{[1,1]} & A_{[1,2]}B_{[1,2]} & \cdots & A_{[1,L_2]}B_{[1,L_2]} \\ A_{[2,1]}B_{[2,1]} & A_{[2,2]}B_{[2,2]} & \cdots & A_{[2,L_2]}B_{[2,L_2]} \\ \vdots & \vdots & \ddots & \vdots \\ A_{[L_1,1]}B_{[L_1,1]} & A_{[L_1,2]}B_{[L_1,2]} & \cdots & A_{[L_1,L_2]}B_{[L_1,L_2]} \end{bmatrix} \tag{1.2}$$

The sign function is denoted $\mathrm{sgn}(x)$ and returns the sign of $x \in \mathbb{R}$. Notation $\mathbf{1}\{x\}$ is the indicator function of the statement x, that returns 1, if x is true, and zero, otherwise. Gradient and Hessian operators of a function $\mathsf{f}(\mathbf{x})$ are denoted $\nabla_{\mathbf{x}}\mathsf{f}(\mathbf{x})$ and $\nabla^2_{\mathbf{x}}\mathsf{f}(\mathbf{x})$, respectively. The probability density function (PDF) (probability mass function (PMF)) of a continuous (discrete) random vector $\mathbf{x}$ is denoted as $\mathsf{p}(\mathbf{x})$. The PDF or PMF of a random vector $\mathbf{x}$ parametrized by a nonrandom parameter $\boldsymbol{\theta}$ is denoted by $\mathsf{p}(\mathbf{x};\boldsymbol{\theta})$. The conditional PDF or PMF of a random vector $\mathbf{x}$ parametrized by a random parameter $\boldsymbol{\theta}$ is denoted by $\mathsf{p}(\mathbf{x}|\boldsymbol{\theta})$. The expectation operator is denoted as $\mathbb{E}[\cdot]$. The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of a random vector $\mathbf{x}$ is denoted by $\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$.

2 Data Acquisition

Indoors RSS ranging measurements are conducted for $K+1$ different distance setups for both LoS and NLoS environments. The LoS measurement setup can be seen in Fig. 1, whereas the

Table 1: Distance setups for RSS measurements

	Setup 0	Setup 1	Setup 2	Setup 3	Setup 4	Setup 5	Setup 6
$\mathbf{s}_{\mathrm{R}}$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
$\mathbf{s}_{k,\mathrm{C}}$	$(1,0.95)$	$(6,0)$	$(7,0)$	$(8,0)$	$(9,0)$	$(10,0)$	$(11,-0.94)$
$\mathbf{s}_{k,\mathrm{T}}$	$(1,0)$	$(7,0)$	$(8,0)$	$(9,0)$	$(10,0)$	$(11,0)$	(12,0)
$d_{k,\mathrm{CT}}$	0.95	1	1	1	1	1	1.3724
$d_{k,\mathrm{TR}}$	1	6	7	8	9	10	11.04
$d_{k,\mathrm{CR}}$	1.3793	7	8	9	10	11	12

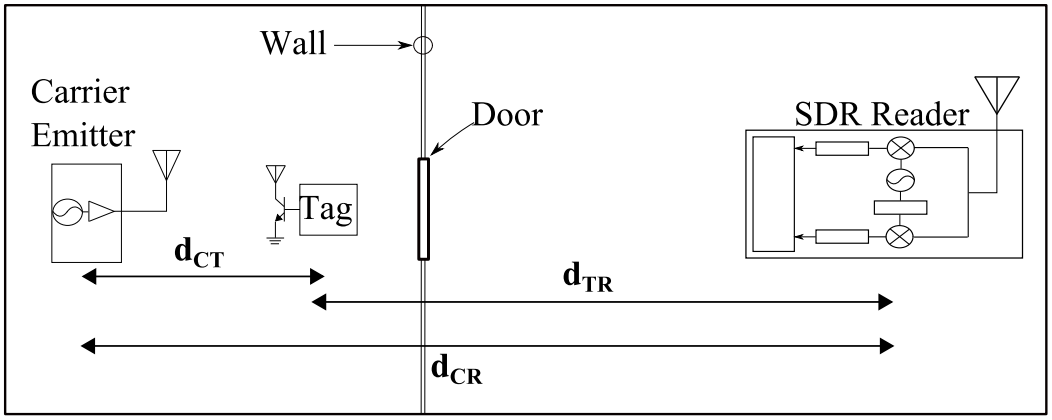

Figure 2: NLoS setup.

NLoS setup is depicted in Fig. 2. For each setup $k \in \{0, 1, 2, \ldots, K\}$, the position of tag, $\mathbf{s}_{k,\mathrm{T}}$, and carrier emitter (CE), $\mathbf{s}_{k,\mathrm{C}}$, was altered, while SDR reader's position, $\mathbf{s}_{\mathrm{R}}$, was not changing. Table 1 shows the 2D positions of tag, carrier emitter (CE) and SDR reader, as well as the Euclidean distances of the corresponding links

$$d_{k,\mathrm{CT}} = \|\mathbf{s}_{k,\mathrm{C}} - \mathbf{s}_{k,\mathrm{T}}\|_2 \tag{2.1}$$

$$d_{k,\mathrm{TR}} = \|\mathbf{s}_{k,\mathrm{T}} - \mathbf{s}_{\mathrm{R}}\|_2 \qquad k \in \{0, 1, 2, \ldots, K\}, \tag{2.2}$$

$$d_{k,\mathrm{CR}} = \|\mathbf{s}_{k,\mathrm{C}} - \mathbf{s}_{\mathrm{R}}\|_2, \tag{2.3}$$

for $K = 6$. For each distance setup, SDR reader obtained N_{meas} LoS and N_{meas} NLoS RSS measurements, with $N_{\mathrm{meas}} = 199$, i.e., $2N_{\mathrm{meas}}$ RSS measurements per distance setup. Each measurement has 2 values, the RSS value of CE-to-SDR reader (CR) link and the RSS value of the compound CE-to-tag-to-SDR reader (CTR) link (this will be explained in more detail at the next section). As we will explain subsequently, the measurements of setup 0 are utilized only for the estimation of the path-loss exponents (PLEs) and won't be utilized for the classification procedure. Thus our machine learning algorithm utilize $2KN_{\mathrm{meas}}$ for classification.

SDR reader measures the instantaneous received power from the CE and the tag using periodogram-based techniques [12]. The power estimation procedure proposed in [12] is tailored to the bistatic scatter radio architecture and respects the signal model of scatter radio, that differs from classic Marconi radio architectures.

3 Determining Path-Loss Exponent

The measurements for setup 0 will be utilized are reference measurements in order to estimate the path-loss exponents (PLEs) of links CR, carrier emitter-to-tag (CT), and tag-to-SDR reader (TR), denoted as ν_{CR}, ν_{CT}, and ν_{TR}, respectively All devices (carrier emitter, tag, and SDR reader) are equipped with dipole antennas and during the measurements all of them are in the same height. Thus, in the sequel it is assumed that all devices have the same transmit and receive antenna gains, i.e., $\mathtt{G}_{\mathrm{C}} = \mathtt{G}_{\mathrm{T}} = \mathtt{G}_{\mathrm{R}} = \mathtt{G} = 10^{2.15/10}$.

3.1 Bistatic Scatter Radio RSS Signal Model

In contrast to conventional Marconi radios, tags employ scatter radio and do not posses receiver capabilities [2], thus ν_{CT} cannot be directly estimated at tag. While the estimation of PLE associated with CR link is trivial, the estimation of ν_{CT} and ν_{TR} requires extra machinery that needs to account the bistatic scatter radio signal model.

The distances of setup 0 are considered as reference distances [13], thus we conduct only LoS measurements. Accordingly with [13], the received powers associated with link CR and the compound link CTR at reference distances are assumed to obey the Frii's law

$$[\text{miliWatts}] \quad \overline{\mathtt{P}}_{0,\mathrm{CR}} = \frac{\mathtt{P}_{\mathrm{C}}\ \mathtt{G}^2\ \lambda}{(4\pi)^2\ (d_{0,\mathrm{CR}})^2}, \tag{3.1}$$

$$[\text{miliWatts}] \quad \overline{\mathtt{P}}_{0,\mathrm{CTR}} = \mathtt{P}_{\mathrm{C}} \frac{\mathtt{G}^2\ \lambda}{(4\pi)^2\ (d_{0,\mathrm{CT}})^2} \frac{\mathtt{s}\ \mathtt{G}^2\ \lambda}{(4\pi)^2\ (d_{0,\mathrm{TR}})^2}, \tag{3.2}$$

where $\overline{\mathtt{P}}_{0,\mathrm{CR}}$ and $\overline{\mathtt{P}}_{0,\mathrm{CTR}}$ are the received power of link CR and compound link CTR, respectively, for setup 0, $\mathtt{P}_{\mathrm{C}}$ is the of carrier emitter transmit power, given in miliWatt units, λ is the wavelength, defined as the ratio of the carrier frequency F_{car}, and the propagation velocity $\mathtt{c}$, i.e., $\lambda = \mathtt{c}/F_{\mathrm{car}}$, and parameter $\mathtt{s}$ incorporates microwave- and antenna-related tag parameters, that change very slowly over time.

For the rest K distance setups, the received power does not necessarily obey the Frii's law. In this case the received power for the distance setup $k \in \{1, 2, \ldots, K\}$ can be written as [13]

$$[\text{miliWatts}] \quad \overline{\mathtt{P}}^{(\mathrm{s})}_{k,\mathrm{CR}} = \overline{\mathtt{P}}_{0,\mathrm{CR}} \left(\frac{d_{0,\mathrm{CR}}}{d_{k,\mathrm{CR}}} \right)^{\nu^{(\mathrm{s})}_{\mathrm{CR}}}, \tag{3.3}$$

$$[\text{miliWatts}] \quad \overline{\mathtt{P}}^{(\mathrm{s})}_{k,\mathrm{CTR}} = \overline{\mathtt{P}}_{0,\mathrm{CTR}} \left(\frac{d_{0,\mathrm{CT}}}{d_{k,\mathrm{CT}}} \right)^{\nu^{(\mathrm{s})}_{\mathrm{CT}}} \left(\frac{d_{0,\mathrm{TR}}}{d_{k,\mathrm{TR}}} \right)^{\nu^{(\mathrm{s})}_{\mathrm{TR}}}, \tag{3.4}$$

where $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$. It is known that PLE tends to have larger values when there is no direct path between transmitter and receiver due to the larger attenuation. The value of PLE is usually affected by the reflection, diffusion, and scattering characteristics of the environment [13]. Common values for PLE are between 2 and 6.

In practice, there major sources of randomness in the received power of Eqs. (3.3) and (3.4). These sources come from large-scaling as well as small-scale effects. Large-scale effects depend the surrounding environment of both transmitter and receiver and vary very slowly with respect to wavelength λ. Whereas, small-scale effects depend on fading that in turn vary rapidly with the wavelength.

In this work we focus on the large-scale effects that will cause random variations of the received power at a given distance due to blockage from objects in the signal path, also known as shadowing, or even changes in reflecting surfaces and scattering objects. The log-normal shadowing is employed, according to which, the received power is random variable (RV), given by

$$[\text{miliWatts}] \quad P^{(\mathrm{s})}_{k,\mathrm{CR}} = \overline{\mathtt{P}}^{(\mathrm{s})}_{k,\mathrm{CR}} 10^{\frac{w^{(\mathrm{s})}_{k,\mathrm{CR}}}{10}} \tag{3.5}$$

$$[\text{miliWatts}] \quad P^{(\mathrm{s})}_{k,\mathrm{CTR}} = \overline{\mathtt{P}}^{(\mathrm{s})}_{k,\mathrm{CTR}} 10^{\frac{w^{(\mathrm{s})}_{k,\mathrm{CTR}}}{10}} \tag{3.6}$$

or, equivalently,

$$[\text{dBm}] \quad g_{k,\text{CR}}^{(\text{s})} = \overline{\text{g}}_{k,\text{CR}}^{(\text{s})} + w_{k,\text{CR}}^{(\text{s})} \tag{3.7}$$

$$[\text{dBm}] \quad g_{k,\text{CTR}}^{(\text{s})} = \overline{\text{g}}_{k,\text{CTR}}^{(\text{s})} + w_{k,\text{CTR}}^{(\text{s})}, \tag{3.8}$$

for $k \in \{1, 2, \ldots, K\}$, $\text{s} \in \{\text{LoS}, \text{NLoS}\}$, while for $k = 0$, we have $\text{s} = \text{LoS}$. Random variables $w_{k,\text{l}}^{(\text{s})}$ are modeled as zero-mean additive Gaussian RV, with standard deviation $\sigma_{k,\text{l}}^{(\text{s})}$. Clearly, $g_{k,\text{l}}^{(\text{s})} = 10\mathsf{log}_{10}\left(P_{k,\text{l}}^{(\text{s})}\right)$, $\overline{\text{g}}_{k,\text{l}}^{(\text{s})} = 10\mathsf{log}_{10}\left(\overline{\text{P}}_{k,\text{l}}^{(\text{s})}\right)$, for $\text{l} \in \{\text{CR}, \text{CTR}\}$, $\text{s} \in \{\text{LoS}, \text{NLoS}\}$, and $\forall k \in \{1, 2, \ldots, K\}$; whereas, for the reference setup, we have $\text{s} = \text{LoS}$. Standard deviation has dB units. It is assumed that $w_{k,\text{l}}^{(\text{s})}$ are independent across different links $\text{l} \in \{\text{CR}, \text{CTR}\}$ and different distance setups $k \in \{0, 1, 2, \ldots, K\}$. Due to the independence of the measurements the problem can be decoupled in 2 subproblems: (i) estimate $\nu_{\text{CR}}^{(\text{s})}$ using the probabilistic model of Eq. (3.7) for both $\text{s} \in \{\text{LoS}, \text{NLoS}\}$ and (ii) estimate $\{\nu_{\text{CT}}^{(\text{s})}, \nu_{\text{TR}}^{(\text{s})}\}$ using the probabilistic model of Eq. (3.8) for both $\text{s} \in \{\text{LoS}, \text{NLoS}\}$. With this background we now turn our attention on the estimation of PLEs of links $\{\text{CR}, \text{CT}, \text{TR}\}$ for the bistatic scatter radio signal model.

3.2 Probabilistic Model for PLEs over Log-Normal Shadowing

At distances $d_{k,\text{CT}}, d_{k,\text{CT}}, d_{k,\text{TR}}$, for both LoS and NLoS setups, it is assumed all N_{meas} measurements collected by the SDR reader are independent and identically distributed (IID) and adopt the probabilistic model of Eqs. (3.7) and (3.8) for link CR and compound link CTR, respectively.[1] Let the measurements associated with link CR and compound link CTR be denoted as $\mathbf{g}_{k,\text{CR}}^{(\text{s})} \triangleq \left\{g_{k,\text{CR}}^{(n),(\text{s})}\right\}_{n=1}^{N_{\text{meas}}}$, and $\mathbf{g}_{k,\text{CTR}}^{(\text{s})} \triangleq \left\{g_{k,\text{CTR}}^{(n),(\text{s})}\right\}_{n=1}^{N_{\text{meas}}}$, respectively, where according to the assumptions in Section 3.1, for $k = 0$ we have $\text{s} = \text{LoS}$, while $\forall k \in \{1, 2, \ldots, K\}$, we have $\text{s} \in \{\text{LoS}, \text{NLoS}\}$. Hence, for log-normal shadowing, and for known parameters $\overline{\text{g}}_{0,\text{CR}}$, $\overline{\text{g}}_{0,\text{CTR}}$, $\sigma_{0,\text{CR}}^{(\text{LoS})}, \sigma_{0,\text{CTR}}^{(\text{LoS})}$, and $\left\{\sigma_{k,\text{CR}}^{(\text{s})}, \sigma_{k,\text{CTR}}^{(\text{s})}\right\}_{k=1}^{K}$, $\text{s} \in \{\text{LoS}, \text{NLoS}\}$, the IID measurements assumption along with Eqs. (3.7) and (3.8) for $k = 0$, imply

$$\mathbf{g}_{0,\text{CR}}^{(\text{LoS})} \sim \mathcal{N}\left(\mathbf{g}_{0,\text{CR}}^{(\text{LoS})}; \overline{\text{g}}_{0,\text{CR}}\mathbf{1}_{N_{\text{meas}}}, \left(\sigma_{0,\text{CR}}^{(\text{LoS})}\right)^2 \mathbf{I}_{N_{\text{meas}}}\right), \tag{3.9}$$

$$\mathbf{g}_{0,\text{CTR}}^{(\text{LoS})} \sim \mathcal{N}\left(\mathbf{g}_{0,\text{CTR}}^{(\text{LoS})}; \overline{\text{g}}_{0,\text{CTR}}\mathbf{1}_{N_{\text{meas}}}, \left(\sigma_{0,\text{CTR}}^{(\text{LoS})}\right)^2 \mathbf{I}_{N_{\text{meas}}}\right), \tag{3.10}$$

while $\forall k \in \{1, 2, \ldots, K\}, \forall \text{s} \in \{\text{LoS}, \text{NLoS}\}$, imply

$$\mathbf{g}_{k,\text{CR}}^{(\text{s})} \sim \mathcal{N}\left(\mathbf{g}_{k,\text{CR}}^{(\text{s})}; \left(\overline{\text{g}}_{0,\text{CR}} - \nu_{\text{CR}}^{(\text{s})} 10\mathsf{log}_{10}\left(\frac{d_{k,\text{CR}}}{d_{0,\text{CR}}}\right)\right)\mathbf{1}_{N_{\text{meas}}}, \left(\sigma_{0,\text{CR}}^{(\text{s})}\right)^2 \mathbf{I}_{N_{\text{meas}}}\right), \tag{3.11}$$

$$\mathbf{g}_{k,\text{CTR}}^{(\text{s})} \sim \mathcal{N}\left(\mathbf{g}_{k,\text{CTR}}^{(\text{s})}; \left(\overline{\text{g}}_{0,\text{CTR}} - 10\mathsf{log}_{10}\left[\left(\frac{d_{k,\text{CT}}}{d_{0,\text{CT}}}\right)^{\nu_{\text{CT}}^{(\text{s})}}\left(\frac{d_{k,\text{TR}}}{d_{0,\text{TR}}}\right)^{\nu_{\text{TR}}^{(\text{s})}}\right]\right)\mathbf{1}_{N_{\text{meas}}}, \left(\sigma_{0,\text{CTR}}^{(\text{s})}\right)^2 \mathbf{I}_{N_{\text{meas}}}\right). \tag{3.12}$$

[1] It is remarked that for each distance setup, SDR reader collects the measurements of CR links and CTR compound link, concurrently.

Our first goal is to estimate parameters $\left\{\nu_{\mathrm{CR}}^{(\mathrm{s})}, \nu_{\mathrm{CT}}^{(\mathrm{s})}, \nu_{\mathrm{TR}}^{(\mathrm{s})}\right\}$, $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$, using the maximum-likelihood framework. However, there exist an obstacle, because the parameters $\overline{\mathsf{g}}_{0,\mathrm{CR}}$, $\overline{\mathsf{g}}_{0,\mathrm{CTR}}$, $\sigma_{0,\mathrm{CR}}^{(\mathrm{LoS})}, \sigma_{0,\mathrm{CTR}}^{(\mathrm{LoS})}$, and $\left\{\sigma_{k,\mathrm{CR}}^{(\mathrm{s})}, \sigma_{k,\mathrm{CTR}}^{(\mathrm{s})}\right\}_{k=1}^{K}$, $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$, are also unknown. Due to Eqs. (3.9) to (3.12), measurements independence across difference distance setups and across links CR and CTR, and the fact that the parameters are deterministic, the maximum likelihood estimate (MLE) of all the parameters can be expressed as the following two (decoupled) problems:

$$
\begin{aligned}
&\left\{\breve{\nu}_{\mathrm{CR}}^{(\mathrm{s})}, \breve{\overline{\mathsf{g}}}_{0,\mathrm{CR}}, \breve{\sigma}_{0,\mathrm{CR}}^{(\mathrm{LoS})}, \left\{\breve{\sigma}_{k,\mathrm{CR}}^{(\mathrm{s})}\right\}_{k=1}^{K}\right\} \\
&= \arg \max_{\substack{\nu_{\mathrm{CR}}^{(\mathrm{s})}, \overline{\mathsf{g}}_{0,\mathrm{CTR}}, \\ \sigma_{0,\mathrm{CR}}^{(\mathrm{LoS})}, \left\{\sigma_{k,\mathrm{CR}}^{(\mathrm{s})}\right\}_{k=1}^{K}}} \left\{\mathsf{p}\left(\mathbf{g}_{0,\mathrm{CR}}^{(\mathrm{LoS})}; \overline{\mathsf{g}}_{0,\mathrm{CR}}, \sigma_{0,\mathrm{CR}}^{(\mathrm{LoS})}\right) \prod_{k=1}^{K} \mathsf{p}\left(\mathbf{g}_{k,\mathrm{CR}}^{(\mathrm{s})}; \overline{\mathsf{g}}_{0,\mathrm{CR}}, \sigma_{k,\mathrm{CR}}^{(\mathrm{s})}, \nu_{\mathrm{CR}}^{(\mathrm{s})}\right)\right\} \\
&= \arg \min_{\substack{\nu_{\mathrm{CR}}^{(\mathrm{s})}, \overline{\mathsf{g}}_{0,\mathrm{CR}}, \\ \sigma_{0,\mathrm{CR}}^{(\mathrm{LoS})}, \left\{\sigma_{k,\mathrm{CR}}^{(\mathrm{s})}\right\}_{k=1}^{K}}} \left\{\left(N_{\mathrm{meas}} \mathsf{ln}\left(\sigma_{0,\mathrm{CR}}^{(\mathrm{LoS})}\right) + \sum_{n=1}^{N_{\mathrm{meas}}} \frac{\left(g_{0,\mathrm{CR}}^{(n),(\mathrm{LoS})} - \overline{\mathsf{g}}_{0,\mathrm{CR}}\right)^2}{2\left(\sigma_{0,\mathrm{CR}}^{(\mathrm{LoS})}\right)^2}\right)\right. \\
&\left. + \left(\sum_{k=1}^{K}\left[N_{\mathrm{meas}} \mathsf{ln}\left(\sigma_{k,\mathrm{CR}}^{(\mathrm{s})}\right) + \sum_{n=1}^{N_{\mathrm{meas}}} \frac{\left[g_{k,\mathrm{CR}}^{(n),(\mathrm{s})} - \left(\overline{\mathsf{g}}_{0,\mathrm{CR}} - \nu_{\mathrm{CR}}^{(\mathrm{s})} 10 \mathsf{log}_{10}\left(\frac{d_{k,\mathrm{CR}}}{d_{0,\mathrm{CR}}}\right)\right)\right]^2}{2\left(\sigma_{k,\mathrm{CR}}^{(\mathrm{s})}\right)^2}\right]\right)\right\} \qquad (3.13)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left\{\breve{\nu}_{\mathrm{CT}}^{(\mathrm{s})}, \breve{\nu}_{\mathrm{TR}}^{(\mathrm{s})}, \breve{\overline{\mathsf{g}}}_{0,\mathrm{CTR}}, \breve{\sigma}_{0,\mathrm{CTR}}^{(\mathrm{LoS})}, \left\{\breve{\sigma}_{k,\mathrm{CTR}}^{(\mathrm{s})}\right\}_{k=1}^{K}\right\} \\
&= \arg \max_{\substack{\nu_{\mathrm{CT}}^{(\mathrm{s})}, \nu_{\mathrm{TR}}^{(\mathrm{s})}, \overline{\mathsf{g}}_{0,\mathrm{CTR}}, \\ \sigma_{0,\mathrm{CTR}}^{(\mathrm{LoS})}, \left\{\sigma_{k,\mathrm{CTR}}^{(\mathrm{s})}\right\}_{k=1}^{K}}} \left\{\mathsf{p}\left(\mathbf{g}_{0,\mathrm{CTR}}^{(\mathrm{LoS})}; \overline{\mathsf{g}}_{0,\mathrm{CTR}}, \sigma_{0,\mathrm{CTR}}^{(\mathrm{LoS})}\right) \prod_{k=1}^{K} \mathsf{p}\left(\mathbf{g}_{k,\mathrm{CTR}}^{(\mathrm{s})}; \overline{\mathsf{g}}_{0,\mathrm{CTR}}, \sigma_{k,\mathrm{CTR}}^{(\mathrm{s})}, \nu_{\mathrm{CT}}^{(\mathrm{s})}, \nu_{\mathrm{TR}}^{(\mathrm{s})}\right)\right\} \\
&= \arg \min_{\substack{\nu_{\mathrm{CT}}^{(\mathrm{s})}, \nu_{\mathrm{TR}}^{(\mathrm{s})}, \overline{\mathsf{g}}_{0,\mathrm{CTR}}, \\ \sigma_{0,\mathrm{CTR}}^{(\mathrm{LoS})}, \left\{\sigma_{k,\mathrm{CTR}}^{(\mathrm{s})}\right\}_{k=1}^{K}}} \left\{\left(N_{\mathrm{meas}} \mathsf{ln}\left(\sigma_{0,\mathrm{CTR}}^{(\mathrm{LoS})}\right) + \sum_{n=1}^{N_{\mathrm{meas}}} \frac{\left(g_{0,\mathrm{CTR}}^{(n),(\mathrm{LoS})} - \overline{\mathsf{g}}_{0,\mathrm{CTR}}\right)^2}{2\left(\sigma_{0,\mathrm{CTR}}^{(\mathrm{LoS})}\right)^2}\right) + \left(\sum_{k=1}^{K}\left[N_{\mathrm{meas}} \mathsf{ln}\left(\sigma_{k,\mathrm{CTR}}^{(\mathrm{s})}\right)\right.\right.\right. \\
&\left.\left.\left. + \sum_{n=1}^{N_{\mathrm{meas}}} \frac{\left[g_{k,\mathrm{CTR}}^{(n),(\mathrm{s})} - \left(\overline{\mathsf{g}}_{0,\mathrm{CTR}} - \nu_{\mathrm{CT}}^{(\mathrm{s})} 10 \mathsf{log}_{10}\left(\frac{d_{k,\mathrm{CT}}}{d_{0,\mathrm{CT}}}\right) - \nu_{\mathrm{TR}}^{(\mathrm{s})} 10 \mathsf{log}_{10}\left(\frac{d_{k,\mathrm{TR}}}{d_{0,\mathrm{TR}}}\right)\right)\right]^2}{2\left(\sigma_{k,\mathrm{CTR}}^{(\mathrm{s})}\right)^2}\right]\right)\right\}, \qquad (3.14)
\end{aligned}
$$

for $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$. Unfortunately, the above two problems do not admit closed form solution. Thus, we will reside to sub-optimal methods on estimating PLEs $\left\{\nu_{\mathrm{CR}}^{(\mathrm{s})}, \nu_{\mathrm{CT}}^{(\mathrm{s})}, \nu_{\mathrm{TR}}^{(\mathrm{s})}\right\}$, $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$, which will offer simpler formulas and comparable root-mean-squared error with respect to Cramér-Rao lower bound, discussed below.

3.3 PLE Estimation for CR Link

In order to find a tractable sub-optimal solution for the optimization problem in (3.13), we will further decouple the minimizations in (3.13). Firstly we estimate $\overline{\mathsf{g}}_{0,\mathrm{CR}}$ using only the measurements $\mathbf{g}_{0,\mathrm{CR}}^{(\mathrm{LoS})}$ (ignoring the rest measurements). After obtaining an estimation for $\overline{\mathsf{g}}_{0,\mathrm{CR}}$, $\widehat{\overline{\mathsf{g}}}_{0,\mathrm{CR}}$, each $\sigma_{k,\mathrm{CR}}^{(\mathrm{s})}$ is estimated separately using only the measurements $\mathbf{g}_{k,\mathrm{CR}}^{(\mathrm{s})}$, $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$ (ignoring the rest measurements). In doing so, we obtain estimates $\left\{\widehat{\sigma}_{k,\mathrm{CR}}^{(\mathrm{s})}\right\}_{k=1}^{K}$, for $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$. As a final step, PLE $\nu_{\mathrm{CR}}^{(\mathrm{s})}$ is estimated using measurements $\mathbf{g}_{1,\mathrm{CR}}^{(\mathrm{s})}, \mathbf{g}_{2,\mathrm{CR}}^{(\mathrm{s})}, \ldots, \mathbf{g}_{K,\mathrm{CR}}^{(\mathrm{s})}$, $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$.

More specifically, $\overline{\mathsf{g}}_{0,\mathrm{CR}}$ is estimated as the sample mean of $\mathbf{g}_{0,\mathrm{CR}}^{(\mathrm{LoS})}$ as

$$\widehat{\overline{\mathsf{g}}}_{0,\mathrm{CR}} = \frac{1}{N_{\mathrm{meas}}} \sum_{n=1}^{N_{\mathrm{meas}}} g_{0,\mathrm{CR}}^{(n),(\mathrm{LoS})}. \tag{3.15}$$

It is noted that after estimating $\widehat{\overline{\mathsf{g}}}_{0,\mathrm{CR}}$, variance $\sigma_{0,\mathrm{CTR}}^{(\mathrm{LoS})}$ does not affect the rest unknown parameter, and thus its calculation can be omitted.

Afterwards, for each distance setup $k \in \{1, 2, \ldots, K\}$, and each $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$, the standard deviation $\sigma_{k,\mathrm{CR}}^{(\mathrm{s})}$ is computed as the maximum likelihood variance estimator of measurements $\mathbf{g}_{k,\mathrm{CR}}^{(\mathrm{s})}$ [14] (assuming unknown mean)

$$\widehat{\sigma}_{k,\mathrm{CR}}^{(\mathrm{s})} = \sqrt{\frac{1}{N_{\mathrm{meas}} - 1} \sum_{n=1}^{N_{\mathrm{meas}}} \left(g_{k,\mathrm{CR}}^{(n),(\mathrm{s})} - \left(\frac{1}{N_{\mathrm{meas}}} \sum_{j=1}^{N_{\mathrm{meas}}} g_{k,\mathrm{CR}}^{(j),(\mathrm{s})} \right) \right)^2}. \tag{3.16}$$

After estimating the parameters $\widehat{\overline{\mathsf{g}}}_{0,\mathrm{CR}}$ and $\left\{\widehat{\sigma}_{k,\mathrm{CR}}^{(\mathrm{s})}\right\}_{k=1}^{K}$, the following abbreviation are utilized for $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$

$$\widehat{\mathbf{q}}_{0,\mathrm{CR}} = \widehat{\overline{\mathsf{g}}}_{0,\mathrm{CR}} \mathbf{1}_{K N_{\mathrm{meas}}} \tag{3.17}$$

$$\widehat{\mathbf{\Sigma}}_{\mathrm{CR}}^{(\mathrm{s})} = \begin{bmatrix} \left(\widehat{\sigma}_{1,\mathrm{CR}}^{(\mathrm{s})}\right)^2 & 0 & \cdots & 0 \\ 0 & \left(\widehat{\sigma}_{2,\mathrm{CR}}^{(\mathrm{s})}\right)^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\widehat{\sigma}_{K,\mathrm{CR}}^{(\mathrm{s})}\right)^2 \end{bmatrix} \otimes \mathbf{I}_{N_{\mathrm{meas}}} \tag{3.18}$$

$$\mathbf{b}_{\mathrm{CR}} = \left[10\mathsf{log}_{10}\left(\frac{d_{1,\mathrm{CR}}}{d_{0,\mathrm{CR}}}\right) \;\; 10\mathsf{log}_{10}\left(\frac{d_{2,\mathrm{CR}}}{d_{0,\mathrm{CR}}}\right) \;\; \cdots \;\; 10\mathsf{log}_{10}\left(\frac{d_{K,\mathrm{CR}}}{d_{0,\mathrm{CR}}}\right) \right]^{\top} \otimes \mathbf{1}_{N_{\mathrm{meas}}}, \tag{3.19}$$

and, according to Eq. (3.11) and the fact that the measurements at different distances are independent, we observe that, given $\widehat{\mathbf{q}}_{0,\mathrm{CR}}$ and $\widehat{\mathbf{\Sigma}}_{\mathrm{CR}}^{(\mathrm{s})}$, the random vector

$$\mathbf{g}_{\mathrm{CR}}^{(\mathrm{s})} = \left[\left(\mathbf{g}_{1,\mathrm{CR}}^{(\mathrm{s})}\right)^{\top} \;\; \left(\mathbf{g}_{2,\mathrm{CR}}^{(\mathrm{s})}\right)^{\top} \;\; \cdots \;\; \left(\mathbf{g}_{K,\mathrm{CR}}^{(\mathrm{s})}\right)^{\top} \right]^{\top}, \tag{3.20}$$

follows multivariate normal distribution with mean $\widehat{\mathbf{q}}_{0,\mathrm{CR}} - \nu_{\mathrm{CR}}^{(\mathrm{s})} \mathbf{b}_{\mathrm{CR}}$ and covariance matrix $\widehat{\mathbf{\Sigma}}_{\mathrm{CR}}^{(\mathrm{s})}$, i.e.,

$$\mathbf{g}_{\mathrm{CR}}^{(\mathrm{s})} \sim \mathcal{N}\left(\mathbf{g}_{\mathrm{CR}}^{(\mathrm{s})} \; ; \; \widehat{\mathbf{q}}_{0,\mathrm{CR}} - \nu_{\mathrm{CR}}^{(\mathrm{s})} \mathbf{b}_{\mathrm{CR}}, \widehat{\mathbf{\Sigma}}_{\mathrm{CR}}^{(\mathrm{s})} \right), \tag{3.21}$$

with $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$. Thus, a convenient estimator for $\nu_{\mathrm{CR}}^{(\mathrm{s})}$, would be the MLE of $\nu_{\mathrm{CR}}^{(\mathrm{s})}$ given the parameters $\left\{\widehat{\overline{\mathsf{g}}}_{0,\mathrm{CR}},\ \left\{\widehat{\sigma}_{k,\mathrm{CR}}^{(\mathrm{s})}\right\}_{k=1}^{K}\right\}$ expressed as

$$\widehat{\nu}_{\mathrm{CR}}^{(\mathrm{s})} = \left((\mathbf{b}_{\mathrm{CR}})^{\top}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{CR}}^{(\mathrm{s})}\right)^{-1}\mathbf{b}_{\mathrm{CR}}\right)^{-1}(\mathbf{b}_{\mathrm{CR}})^{\top}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{CR}}^{(\mathrm{s})}\right)^{-1}\left(\widehat{\mathbf{q}}_{0,\mathrm{CR}} - \mathbf{g}_{\mathrm{CR}}^{(\mathrm{s})}\right),\ \mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}. \quad (3.22)$$

3.4 PLE Estimation for CTR Link

The same estimation procedure with Section 3.3 is followed to obtain a tractable estimator for $\widehat{\nu}_{\mathrm{CT}}^{(\mathrm{s})}$ and $\widehat{\nu}_{\mathrm{TR}}^{(\mathrm{s})}$ for both $\mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}$. Thus, accordingly with Eq. (3.15), parameter $\overline{\mathsf{g}}_{0,\mathrm{CTR}}$ is estimated as

$$\widehat{\overline{\mathsf{g}}}_{0,\mathrm{CTR}} = \frac{1}{N_{\mathrm{meas}}}\sum_{n=1}^{N_{\mathrm{meas}}} g_{0,\mathrm{CTR}}^{(n),(\mathrm{LoS})}. \quad (3.23)$$

Next, following exactly the same reasoning with Section 3.3, for each distance setup the standard deviation $\sigma_{k,\mathrm{CTR}}^{(\mathrm{s})}$ is estimated as

$$\widehat{\sigma}_{k,\mathrm{CTR}}^{(\mathrm{s})} = \sqrt{\frac{1}{N_{\mathrm{meas}}-1}\sum_{n=1}^{N_{\mathrm{meas}}}\left(g_{k,\mathrm{CTR}}^{(n),(\mathrm{s})} - \left(\frac{1}{N_{\mathrm{meas}}}\sum_{j=1}^{N_{\mathrm{meas}}} g_{k,\mathrm{CTR}}^{(j),(\mathrm{s})}\right)\right)^2}, \quad (3.24)$$

As a final step, the following abbreviations are utilized

$$\widehat{\mathbf{q}}_{0,\mathrm{CTR}} = \widehat{\overline{\mathsf{g}}}_{0,\mathrm{CTR}}\mathbf{1}_{KN_{\mathrm{meas}}} \quad (3.25)$$

$$\widehat{\mathbf{\Sigma}}_{\mathrm{CTR}}^{(\mathrm{s})} = \begin{bmatrix} \left(\widehat{\sigma}_{1,\mathrm{CTR}}^{(\mathrm{s})}\right)^2 & 0 & \cdots & 0 \\ 0 & \left(\widehat{\sigma}_{2,\mathrm{CTR}}^{(\mathrm{s})}\right)^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\widehat{\sigma}_{K,\mathrm{CTR}}^{(\mathrm{s})}\right)^2 \end{bmatrix} \otimes \mathbf{I}_{N_{\mathrm{meas}}} \quad (3.26)$$

$$\mathbf{b}_{\mathrm{CT}} = \left[10\mathsf{log}_{10}\left(\frac{d_{1,\mathrm{CT}}}{d_{0,\mathrm{CT}}}\right)\ 10\mathsf{log}_{10}\left(\frac{d_{2,\mathrm{CT}}}{d_{0,\mathrm{CT}}}\right)\ \cdots\ 10\mathsf{log}_{10}\left(\frac{d_{K,\mathrm{CT}}}{d_{0,\mathrm{CT}}}\right)\right]^{\top} \otimes \mathbf{1}_{N_{\mathrm{meas}}} \quad (3.27)$$

$$\mathbf{b}_{\mathrm{TR}} = \left[10\mathsf{log}_{10}\left(\frac{d_{1,\mathrm{TR}}}{d_{0,\mathrm{TR}}}\right)\ 10\mathsf{log}_{10}\left(\frac{d_{2,\mathrm{TR}}}{d_{0,\mathrm{TR}}}\right)\ \cdots\ 10\mathsf{log}_{10}\left(\frac{d_{K,\mathrm{TR}}}{d_{0,\mathrm{TR}}}\right)\right]^{\top} \otimes \mathbf{1}_{N_{\mathrm{meas}}} \quad (3.28)$$

$$\mathbf{B}_{\mathrm{CTR}} = [\mathbf{b}_{\mathrm{CT}}\ \ \mathbf{b}_{\mathrm{TR}}]. \quad (3.29)$$

so that, for given $\widehat{\overline{\mathsf{g}}}_{0,\mathrm{CTR}}$ and $\left\{\widehat{\sigma}_{k,\mathrm{CTR}}^{(\mathrm{s})}\right\}_{k=1}^{K}$,

$$\mathbf{g}_{\mathrm{CTR}}^{(\mathrm{s})} = \left[\left(\mathbf{g}_{1,\mathrm{CTR}}^{(\mathrm{s})}\right)^{\top}\ \left(\mathbf{g}_{2,\mathrm{CTR}}^{(\mathrm{s})}\right)^{\top}\ \cdots\ \left(\mathbf{g}_{K,\mathrm{CTR}}^{(\mathrm{s})}\right)^{\top}\right]^{\top}, \quad (3.30)$$

is a multivariate Gaussian vector, with mean $\widehat{\mathbf{q}}_{0,\mathrm{CTR}} - \mathbf{B}_{\mathrm{CTR}}\left[\nu_{\mathrm{CT}}^{(\mathrm{s})}\ \nu_{\mathrm{TR}}^{(\mathrm{s})}\right]^{\top}$ and covariance matrix $\widehat{\mathbf{\Sigma}}_{\mathrm{CTR}}^{(\mathrm{s})}$, i.e.,

$$\mathbf{g}_{\mathrm{CTR}}^{(\mathrm{s})} \sim \mathcal{N}\left(\mathbf{g}_{\mathrm{CTR}}^{(\mathrm{s})}\ ;\ \widehat{\mathbf{q}}_{0,\mathrm{CTR}} - \mathbf{B}_{\mathrm{CTR}}\left[\nu_{\mathrm{CT}}^{(\mathrm{s})}\ \nu_{\mathrm{TR}}^{(\mathrm{s})}\right]^{\top}, \widehat{\mathbf{\Sigma}}_{\mathrm{CTR}}^{(\mathrm{s})}\right),\ \mathrm{s} \in \{\mathrm{LoS}, \mathrm{NLoS}\}. \quad (3.31)$$

Thus, the MLE for PLEs $\nu_{\rm CT}^{(\rm s)}$ and $\nu_{\rm TR}^{(\rm s)}$, given the parameters $\left\{\widehat{\overline{\mathsf{g}}}_{0,\rm CTR},\ \left\{\widehat{\sigma}_{k,\rm CTR}^{(\rm s)}\right\}_{k=1}^{K}\right\}$ (associated with the measurements $\mathbf{g}_{\rm CTR}^{(\rm s)}$) is given by

$$\begin{bmatrix}\widehat{\nu}_{\rm CT}^{(\rm s)}\\ \widehat{\nu}_{\rm TR}^{(\rm s)}\end{bmatrix} = \left((\mathbf{B}_{\rm CTR})^{\top}\left(\widehat{\mathbf{\Sigma}}_{\rm CTR}^{(\rm s)}\right)^{-1}\mathbf{B}_{\rm CTR}\right)^{-1}(\mathbf{B}_{\rm CTR})^{\top}\left(\widehat{\mathbf{\Sigma}}_{\rm CTR}^{(\rm s)}\right)^{-1}\left(\widehat{\mathbf{q}}_{0,\rm CTR}-\mathbf{g}_{\rm CTR}^{(\rm s)}\right),\ {\rm s}\in\{\rm LoS,NLoS\}. \tag{3.32}$$

3.5 Comparison with Cramér-Rao Lower Bound with Known Parameters

Suppose that the parameters $\overline{\mathsf{g}}_{0,\rm CR}$, $\overline{\mathsf{g}}_{0,\rm CTR}$, $\sigma_{0,\rm CR}^{(\rm LoS)}$, $\sigma_{0,\rm CTR}^{(\rm LoS)}$, and $\left\{\sigma_{k,\rm CR}^{(\rm s)},\sigma_{k,\rm CTR}^{(\rm s)}\right\}_{k=1}^{K}$, ${\rm s}\in\{\rm LoS,NLoS\}$, are known. Abbreviating

$$\mathbf{\Sigma}_{\rm l}^{(\rm s)} = \begin{bmatrix}\left(\sigma_{1,\rm l}^{(\rm s)}\right)^2 & 0 & \cdots & 0\\ 0 & \left(\sigma_{2,\rm l}^{(\rm s)}\right)^2 & \cdots & 0\\ \vdots & \cdots & \ddots & \vdots\\ 0 & 0 & \cdots & \left(\sigma_{K,\rm l}^{(\rm s)}\right)^2\end{bmatrix}\otimes\mathbf{I}_{N_{\rm meas}},\ {\rm l}\in\{\rm CR,CTR\},\ {\rm s}\in\{\rm LoS,NLoS\}, \tag{3.33}$$

$$\mathbf{S}^{(\rm s)} = \begin{bmatrix}\mathbf{\Sigma}_{\rm CR}^{(\rm s)} & \mathbf{0}_{N_{\rm meas}\times N_{\rm meas}}\\ \mathbf{0}_{N_{\rm meas}\times N_{\rm meas}} & \mathbf{\Sigma}_{\rm CTR}^{(\rm s)}\end{bmatrix},\ {\rm s}\in\{\rm LoS,NLoS\} \tag{3.34}$$

$$\mathbf{B} = \begin{bmatrix}\mathbf{b}_{\rm CR} & \mathbf{0}_{KN_{\rm meas}\times 2}\\ \mathbf{0}_{KN_{\rm meas}\times 1} & \mathbf{B}_{\rm CTR}\end{bmatrix} \tag{3.35}$$

$$\mathbf{q} = \begin{bmatrix}\overline{\mathsf{g}}_{0,\rm CR}\mathbf{1}_{KN_{\rm meas}}\\ \overline{\mathsf{g}}_{0,\rm CTR}\mathbf{1}_{KN_{\rm meas}}\end{bmatrix}, \tag{3.36}$$

$$\boldsymbol{\nu}^{(\rm s)} = \begin{bmatrix}\nu_{\rm CR}^{(\rm s)} & \nu_{\rm CT}^{(\rm s)} & \nu_{\rm TR}^{(\rm s)}\end{bmatrix}^{\top},\ {\rm s}\in\{\rm LoS,NLoS\} \tag{3.37}$$

and using the independence of measurements across different links, we deduce that for given $\overline{\mathsf{g}}_{0,\rm CR}$, $\overline{\mathsf{g}}_{0,\rm CTR}$, $\sigma_{0,\rm CR}^{(\rm LoS)}$, $\sigma_{0,\rm CTR}^{(\rm LoS)}$, and $\left\{\sigma_{k,\rm CR}^{(\rm s)},\sigma_{k,\rm CTR}^{(\rm s)}\right\}_{k=1}^{K}$, ${\rm s}\in\{\rm LoS,NLoS\}$, vector $\mathbf{g}^{(\rm s)}\triangleq\begin{bmatrix}\mathbf{g}_{\rm CT}^{(\rm s)}\\ \mathbf{g}_{\rm CTR}^{(\rm s)}\end{bmatrix}$

$$\mathbf{g}^{(\rm s)}\sim\mathcal{N}\left(\mathbf{g}^{(\rm s)}\ ;\ \mathbf{q}-\mathbf{B}\boldsymbol{\nu}^{(\rm s)},\mathbf{S}^{(\rm s)}\right),\ {\rm s}\in\{\rm LoS,NLoS\}. \tag{3.38}$$

Hence, applying [15, Eq. (3.31)] we conclude that the Fisher information matrix for parameter vector $\boldsymbol{\nu}^{(\rm s)}$ is given by

$$\mathbf{F}^{(\rm s)} = \left(\mathbf{B}^{\top}\left(\mathbf{S}^{(\rm s)}\right)^{-1}\mathbf{B}\right),\ {\rm s}\in\{\rm LoS,NLoS\}. \tag{3.39}$$

Thus according to Cramér-Rao bound (CRB), the mean squared-error of any unbiased estimator of nonrandom parameter $\boldsymbol{\nu}^{(\rm s)}$, say $\widehat{\boldsymbol{\nu}}^{(\rm s)}$, is lower bounded as

$$\mathbb{E}_{\mathbf{g}^{(\rm s)};\boldsymbol{\nu}^{(\rm s)}}\left[\left\|\widehat{\boldsymbol{\nu}}^{(\rm s)}-\boldsymbol{\nu}^{(\rm s)}\right\|_2^2\right]\geq\mathsf{tr}\left\{\left(\mathbf{F}^{(\rm s)}\right)^{-1}\right\},\ {\rm s}\in\{\rm LoS,NLoS\}. \tag{3.40}$$

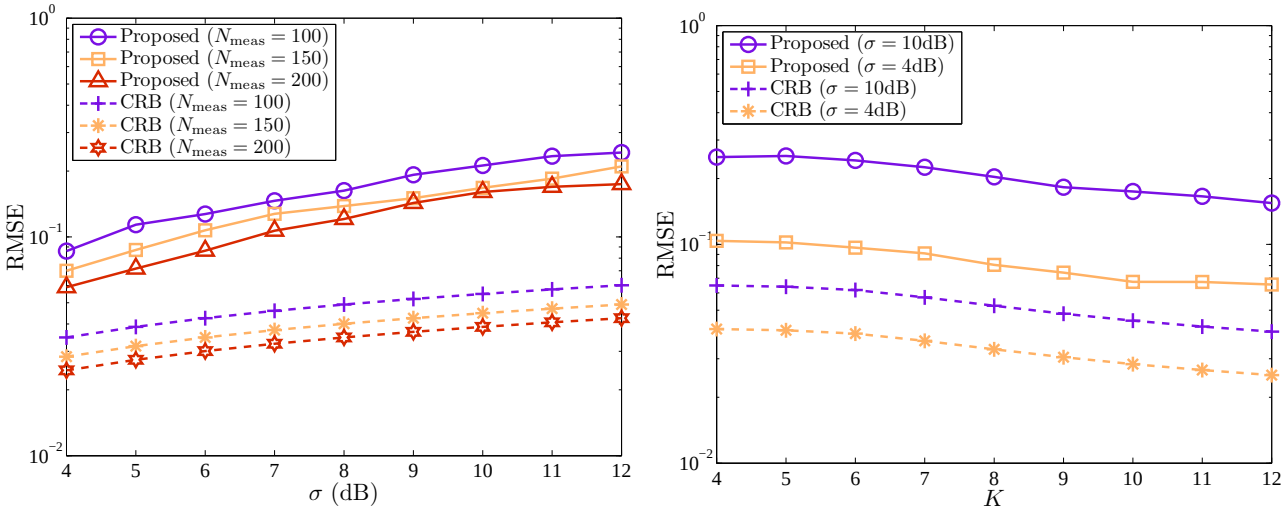

Figure 3: Left: RMSE of the proposed PLE vector estimator and CRB versus σ for different values of N_{meas} and $K = 10$. Left: RMSE of the proposed PLE estimator and CRB versus K for different values of σ and $N_{\text{meas}} = 150$.

3.5.1 Numerical Results for PLE Estimation

In this section we compare the root-mean-squared error (RMSE) performance of the PLEs estimator described in Sections 3.3 and 3.4 with the CRB associated with perfect knowledge of the unknown parameters given in (3.40). We simulate the proposed estimator using computer simulations and study the impact of two parameters:

- The impact of parameter N_{meas} for fixed K.
- The impact of parameter K for fixed N_{meas}.

We consider a bistatic scatter radio topology parametrized by number K, as follows

$$\mathbf{s}_{\text{R}} = \mathbf{0}_2 \tag{3.41}$$

$$\mathbf{s}_{k,\text{T}} = \begin{bmatrix} 7+k \\ 6+k \end{bmatrix}, \quad k = 1, 2, \ldots, K, \tag{3.42}$$

$$\mathbf{s}_{k,\text{C}} = \begin{bmatrix} 14-k \\ 10 \end{bmatrix}, \quad k = 1, 2, \ldots, K, \tag{3.43}$$

with reference distance $d_{0,\text{CR}} = d_{0,\text{CT}} = d_{0,\text{TR}} = 1$ meters. Otherwise stated the following parameter values are considered: $\mathtt{s} = 10^{-2}$, $F_{\text{c}} = 868$MHz, $\mathtt{c} = 3 \cdot 10^8$m/s, $\mathtt{P}_{\text{C}} = 20$mWatt (or -13dBm). We implicitly assume a scenario where carrier emitter and tag have LoS among each other, and both of them have NLoS with SDR reader, i.e., s = NLoS. Thus, we choose as PLE values $\nu_{\text{CR}}^{(\text{NLoS})} = 2.75$, $\nu_{\text{CT}}^{(\text{NLoS})} = 2.08$, and $\nu_{\text{TR}}^{(\text{NLoS})} = 2.6$. We further assume that for reference distance setup (setup 0) $\sigma_{0,\text{CTR}}^{(\text{LoS})} = \sigma_{0,\text{CR}}^{(\text{LoS})} = 1$dB. For simplicity it is also assumed that the shadowing variance remains the same across different distance setups with values $\sigma_{k,\text{CTR}}^{(\text{NLoS})} = 1.5\sigma_{k,\text{CR}}^{(\text{NLoS})}$ and $\sigma_{k,\text{CR}}^{(\text{NLoS})} = \sigma$, $\forall k \in \{1, 2, \ldots, K\}$. The range of values for shadowing variance considered in this work are taken from realistic RSS measurement testbeds [16].

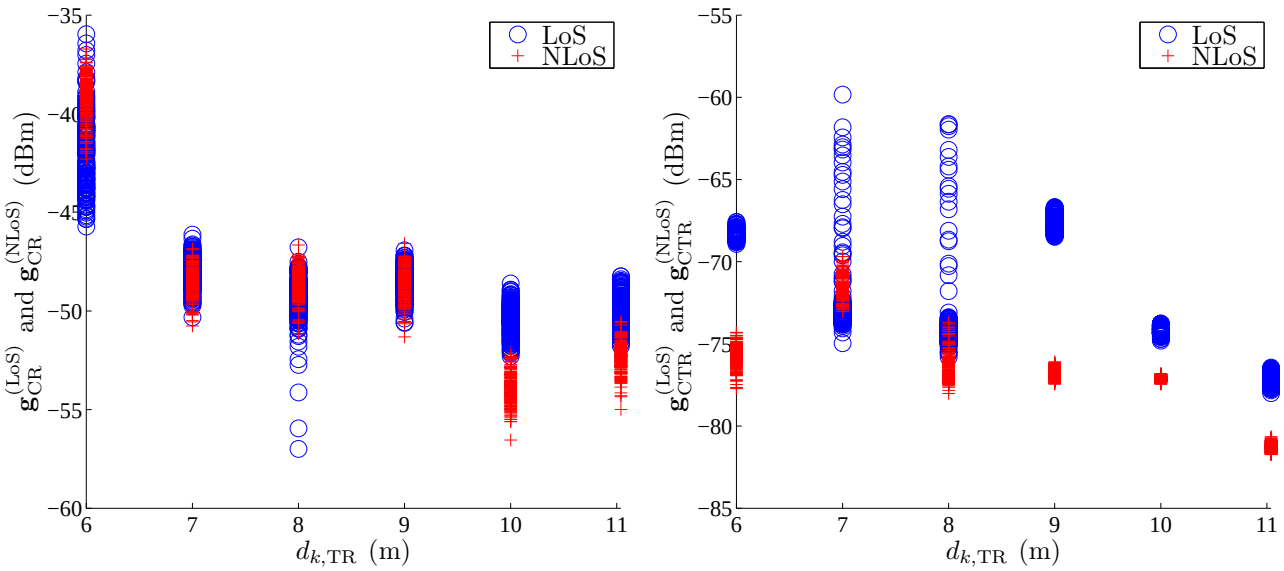

Figure 4: Left (Right): Experimental RSS measurements values in dBm for CR (CTR) link as a function of $d_{k,\mathrm{CR}}$.

Fig. 3-Left shows the RMSE of the proposed PLE vector estimator and CRB as a function of RSS ranging noise standard deviation, σ, for different values of N_{meas} and fixed value of $K = 10$. The RMSE of CRB is calculated by taking the square root of the right-hand side (RHS) in (3.40). It is noted that for fixed number of distance setups, K, as we increase the value of parameter σ both PLE estimator's and CRB RMSE decrease. In addition, as we increase the number of collected measurements per distance setup, N_{meas}, the RMSE also decreases. For fixed σ, the RMSE performance gap between CRB and the proposed estimator is approximately 5-7dB, which quite remarkable from the estimation accuracy perspective, since the proposed estimator estimates all unknown parameters first, while the CRB in Eq. (3.40) assumes full knowledge for the unknown parameters. It is noted that as σ increases, the RMSE performance gap decreases.

In Fig. 3-Right the RMSE performance of the proposed estimator, as well as the CRB are plotted as function of K for two different standard deviation values and fixed value for $N_{\mathrm{iter}} = 150$. It can be seen that for fixed N_{iter} and σ the RMSE decreases slowly. As in the previous case, we note that for smaller values of ranging noise standard deviation, σ, the RMSE performance gap between the proposed estimator and CRB is diminished.

3.6 Estimated PLEs for Real Data

If we apply the estimation procedure described in Sections 3.3 and 3.4, to our collected measurements, depicted in Fig. 4, we obtain

$$\widehat{\boldsymbol{\nu}}^{(\mathrm{LoS})} = [2.6084 \quad 2.0705 \quad 2.1838]^{\top}, \tag{3.44}$$

$$\widehat{\boldsymbol{\nu}}^{(\mathrm{NLoS})} = [2.7558 \quad 2.0666 \quad 2.5797]^{\top}, \tag{3.45}$$

which corroborate the efficacy of the proposed PLE estimation framework.

4 Machine Learning for NLoS Identification

4.1 Creating Input Data and Labels

The input data and the corresponding labels are denoted $\mathbf{x}^{(i)} \in \mathbb{R}^M$ and $y^{(i)} \in \mathbb{B}$, respectively. In this work we used the following input data

$$\mathbf{x}^{(i)} = \begin{bmatrix} g_{k,\mathrm{CR}}^{(n),(\mathrm{LoS})} & g_{k,\mathrm{CTR}}^{(n),(\mathrm{LoS})} & d_{k,\mathrm{TR}} \end{bmatrix}^\top, \quad i = n + (k-1)N_{\mathrm{meas}}, \tag{4.1}$$

for $n = 1, 2, \ldots, N_{\mathrm{meas}}$, and $k = 1, 2, \ldots, K$, associated with LoS measurements and

$$\mathbf{x}^{(i)} = \begin{bmatrix} g_{k,\mathrm{CR}}^{(n),(\mathrm{NLoS})} & g_{k,\mathrm{CTR}}^{(n),(\mathrm{NLoS})} & d_{k,\mathrm{TR}} \end{bmatrix}^\top, \quad i = KN_{\mathrm{meas}} + n + (k-1)N_{\mathrm{meas}}, \tag{4.2}$$

for $n = 1, 2, \ldots, N_{\mathrm{meas}}$ and $k = 1, 2, \ldots, K$, associated with NLoS measurements. Clearly the specific choice of input data offers $M = 3$, however, more elements can be added in the input data $\mathbf{x}^{(i)}$. After experimentation over different classification algorithms, we deduced that, adding more elements in each $\mathbf{x}^{(i)}$ increases the computational cost, without improving further the error performance. The corresponding labels are

$$y^{(i)} = \begin{cases} 1, & \text{if} \quad 1 \le i \le KN_{\mathrm{meas}} \\ 0, & \text{if} \quad KN_{\mathrm{meas}} + 1 \le i \le 2KN_{\mathrm{meas}}, \end{cases} \tag{4.3}$$

i.e., the labeling is the "1" for LoS measurements and "0" for NLoS measurements. We conclude that we have at total $N_{\mathrm{tot}} = 2KN_{\mathrm{meas}}$ data.

4.2 Cross-Validation

Recall that $N_{\mathrm{tot}} = 2KN_{\mathrm{meas}}$ is the total number of data. In this work we adopt a Mont Carlo (MC)-based cross-validation technique. We define N_{tr} and N_{test} the number of training samples and testing samples, respectively, assuming that both of them are divided by $2K$. We also define $\mathcal{N}_{\mathrm{tr}}$ and $\mathcal{N}_{\mathrm{test}}$ the empty sets.

The idea the following: for each random experiment do the following procedure. For each distance setup $k \in \{1, 2, \ldots, K\}$ choose at random (without replacement) $N_{\mathrm{tr}}/(2K)$ elements from the set $\mathcal{N}_k^{\mathrm{LoS}} = \{(k-1)N_{\mathrm{meas}} + 1, (k-1)N_{\mathrm{meas}} + 2, \ldots, kN_{\mathrm{meas}}\}$ and another $N_{\mathrm{tr}}/(2K)$ elements from the set $\mathcal{N}_k^{\mathrm{NLoS}} = \{(K+k-1)N_{\mathrm{meas}} + 1, (K+k-1)N_{\mathrm{meas}} + 2, \ldots, (K+k)N_{\mathrm{meas}}\}$. Move the chosen elements to the set $\mathcal{N}_{\mathrm{tr}}$, and the rest elements are moved the testing set. Continue for the rest k. The above procedure is illustrated in the pseudocode in Algorithm 1. The function ChooseWithoutReplacement$(\mathcal{A}, b)$ return b elements chosen without replacement from set $\mathcal{A}$.

4.3 Feature Mappings

Sometimes, it is more convenient to consider functions applied on the original data points $\left\{\mathbf{x}^{(i)}\right\}_{i=1}^{N_{\mathrm{tot}}}$. The most common example arises when the data $\left\{\mathbf{x}^{(i)}\right\}_{i=1}^{N_{\mathrm{tot}}}$ are not linearly separable, but if we apply a suitable feature mapping (or basis functions) they may be.

For any $\mathbf{x} \in \mathbb{R}^M$ we define

$$\mathbf{l}(\mathbf{x}) \triangleq \begin{bmatrix} \mathsf{ln}\left(\left|x_{[1]}\right|\right) & \mathsf{ln}\left(\left|x_{[2]}\right|\right) & \ldots & \mathsf{ln}\left(\left|x_{[M]}\right|\right) \end{bmatrix}^\top \in \mathbb{R}^M. \tag{4.4}$$

Algorithm 1 Monte Carlo-based Cross-Validation

Input: $\left\{\mathbf{x}^{(i)}, y^{(i)}\right\}_{i=1}^{N_{\text{tot}}}$, $L_{\text{tr}}^{\text{LoS}}, L_{\text{tr}}^{\text{NLoS}} \in \{1, 2, \ldots, N_{\text{meas}}\}$ such that $L_{\text{tr}}^{\text{NLoS}} + L_{\text{tr}}^{\text{LoS}}$ is even natural.

1: $N_{\text{tr}} := \left(L_{\text{tr}}^{\text{NLoS}} + L_{\text{tr}}^{\text{LoS}}\right) K$; $N_{\text{test}} := N_{\text{tot}} - N_{\text{tr}}$; $\mathcal{N}_{\text{tr}} = \mathcal{N}_{\text{test}} = \emptyset$;
2: **for** $k = 1 : K$ **do**
3: $\quad \mathcal{N}_k^{\text{LoS}} := \{(k-1)N_{\text{meas}} + 1, (k-1)N_{\text{meas}} + 2, \ldots, kN_{\text{meas}}\}$;
4: $\quad \mathcal{N}_k^{\text{NLoS}} := \{(K+k-1)N_{\text{meas}} + 1, (K+k-1)N_{\text{meas}} + 2, \ldots, (K+k)N_{\text{meas}}\}$;
5: $\quad \mathcal{A}_k^{\text{LoS}} := \left\{i_1, i_2, \ldots, i_{L_{\text{tr}}^{\text{LoS}}}\right\} = \text{ChooseWithoutReplacement}\left(\mathcal{N}_k^{\text{LoS}}, L_{\text{tr}}^{\text{LoS}}\right)$
6: $\quad \mathcal{A}_k^{\text{NLoS}} := \left\{i'_1, i'_2, \ldots, i'_{L_{\text{tr}}^{\text{NLoS}}}\right\} = \text{ChooseWithoutReplacement}\left(\mathcal{N}_k^{\text{NLoS}}, L_{\text{tr}}^{\text{LoS}}\right)$
7: $\quad \mathcal{N}_{\text{tr}} := \mathcal{N}_{\text{tr}} \cup \mathcal{A}_k^{\text{LoS}} \cup \mathcal{A}_k^{\text{NLoS}}$;
8: **end for**
9: $\mathcal{N}_{\text{test}} := \{1, 2, \ldots, N_{\text{tot}}\} \backslash \mathcal{N}_{\text{tr}}$;

Output: Training set: $\left\{\mathbf{x}^{(i)}, y^{(i)}\right\}_{i \in \mathcal{N}_{\text{tr}}}$, testing set: $\left\{\mathbf{x}^{(i)}, y^{(i)}\right\}_{i \in \mathcal{N}_{\text{test}}}$

After much experimentation over different feature mappings, we deduced that the following families feature mappings work very well for the classification algorithms studied in bistatic scatter radio NLoS identification,

$$\boldsymbol{\phi}_{\text{P},D}(\mathbf{x}) \triangleq \left[1 \quad (\mathbf{x})^\top \quad (\mathbf{x} \otimes \mathbf{x})^\top \quad \ldots \quad \underbrace{(\mathbf{x} \otimes \mathbf{x} \otimes \cdots \otimes \mathbf{x})^\top}_{D \text{ times}}\right]^\top \in \mathbb{R}^{R_{D,M}} \tag{4.5}$$

$$\boldsymbol{\phi}_{\text{L},D}(\mathbf{x}) \triangleq \left[1 \quad (\mathbf{l}(\mathbf{x}))^\top \quad (\mathbf{l}(\mathbf{x}) \odot \mathbf{l}(\mathbf{x}))^\top \quad \ldots \quad \underbrace{(\mathbf{l}(\mathbf{x}) \odot \mathbf{l}(\mathbf{x}) \odot \cdots \odot \mathbf{l}(\mathbf{x}))^\top}_{D \text{ times}}\right]^\top \in \mathbb{R}^{DM+1}, \tag{4.6}$$

where $R_{D,M} = \sum_{d=0}^{D} M^d$. It is worth noting that the feature mapping in (4.5) employees all the possible coefficients of a polynomial of M variables and degree D. On the other hand, the feature mapping in (4.6) uses only the powers up to degree D of individual variables. Finally, the identity feature mapping is given by

$$\boldsymbol{\phi}_{\text{I}}(\mathbf{x}) \triangleq [1 \quad \mathbf{x}]^\top. \tag{4.7}$$

4.4 Kernels

For a given feature mapping $\boldsymbol{\phi} : \mathbb{R}^M \longrightarrow \mathbb{R}^L$,[2] we define an equivalent kernel $\mathsf{k}_{\boldsymbol{\phi}} : \mathbb{R}^M \times \mathbb{R}^M \longrightarrow \mathbb{R}$ associated with $\boldsymbol{\phi}$ as

$$\mathsf{k}_{\boldsymbol{\phi}}(\mathbf{x}, \mathbf{y}) \triangleq \boldsymbol{\phi}_{[2:L]}(\mathbf{x})^\top \boldsymbol{\phi}_{[2:L]}(\mathbf{y}). \tag{4.8}$$

The concept of formulating the kernels as an inner product in a feature space allows us to extend machine learning algorithms by making use of kernel substitution. The general idea is that, many machine learning algorithms can be formulated in such a way that the feature mapping $\boldsymbol{\phi}$ enters only in the form of scalar products, and thus the scalar product can be replaced with the equivalent kernel in (4.8). In this work the Gaussian kernel is utilized, defined as

$$\mathsf{k}_{\text{G}}\left(\mathbf{x}, \mathbf{y}; \sigma_{\text{g}}^2\right) \triangleq \mathsf{exp}\left\{-\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{2\sigma_{\text{g}}^2}\right\}, \tag{4.9}$$

[2] In this work we make the convention that for any $\mathbf{x} \in \mathbb{R}^M$, $\phi_{[1]}(\mathbf{x}) = 1$.

where tuning parameter σ_{g}^2 determines the width of the function around its mode.

4.5 Classification Algorithms

The specific NLoS identification problem requires two classes (or equivalently in detection theory terms, two hypothesis) in order to be formalized in a proper probabilistic way. Let us the denote $\mathcal{C}_1$ and $\mathcal{C}_0$ the classes associated with LoS and NLoS data points, respectively.

The derivation of classification algorithms will take place in the feature space, for generality. We consider an arbitrary feature mapping on $\boldsymbol{\phi} : \mathbb{R}^M \longrightarrow \mathbb{R}^L$ and we are given a set of feature vectors along with the corresponding labels (the training data), i.e., $\left\{\boldsymbol{\phi}\left(\mathbf{x}^{(i)}\right), y^{(i)}\right\}_{i\in\mathcal{N}_{\mathrm{tr}}} \equiv \left\{\boldsymbol{\phi}^{(i)}, y^{(i)}\right\}_{i\in\mathcal{N}_{\mathrm{tr}}}$, and our goal is to to detect the labels of the rest data $\left\{\boldsymbol{\phi}^{(i)}\right\}_{i\in\mathcal{N}_{\mathrm{test}}}$. Without any loss in generality, it is assumed that the training index set, $\mathcal{N}_{\mathrm{tr}}$, consists of all natural numbers that are less or equal to N_{tr}, i.e.,

$$\mathcal{N}_{\mathrm{tr}} = \{1, 2, \ldots, N_{\mathrm{tr}}\}. \tag{4.10}$$

This implies that $\mathcal{N}_{\mathrm{test}} = \{N_{\mathrm{tr}}+1, N_{\mathrm{tr}}+2, \ldots, N_{\mathrm{tot}}\}$. In order to follow a unified algorithmic presentation framework, it is assumed that the first element of any $\boldsymbol{\phi}^{(i)}, i = 1, 2, \ldots, N_{\mathrm{tot}}$, is equal to unity, i.e., $\phi_{[1]}^{(i)} = 1$, $\forall i = 1, 2, \ldots, N_{\mathrm{tot}}$.

4.5.1 Generative Gaussian Classification

In generative Gaussian classification the class-conditional densities are assumed to be Gaussian and our goal is to determine the posterior probabilities. Specifically it is assumed that for any input feature data point

$$\mathsf{p}\Big(\boldsymbol{\phi}^{(i)}|\mathcal{C}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}\Big) \equiv \mathcal{N}\Big(\boldsymbol{\phi}_{[2:L]}^{(i)}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}\Big) \tag{4.11}$$

$$\mathsf{p}\Big(\boldsymbol{\phi}^{(i)}|\mathcal{C}_0; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}\Big) \equiv \mathcal{N}\Big(\boldsymbol{\phi}_{[2:L]}^{(i)}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}\Big). \tag{4.12}$$

If we denote $\mathsf{p}(\mathcal{C}_1; p) = p$, then $\mathsf{p}(\mathcal{C}_0; p) = 1 - p$; it tuns out that the posterior probability of class of $\mathcal{C}_1$ given any input feature point $\boldsymbol{\phi}^{(i)}$ takes the following form [17, Eqs. (4.65)–(4.67)].

$$\mathsf{p}\Big(\mathcal{C}_1|\boldsymbol{\phi}^{(i)}; \boldsymbol{\theta}\Big) = \mathsf{h}\left((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}_{[2:L]}^{(i)} - \frac{1}{2}\boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0 + \mathsf{ln}\left(\frac{p}{1-p}\right)\right), \tag{4.13}$$

where $\boldsymbol{\theta} = \{p, \boldsymbol{\mu}_1, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}\}$ squeezes all the (nonrandom) parameters of the problem and function h is the logistic sigmoid function, defined as

$$\mathsf{h}(x) \triangleq \frac{1}{1+\mathsf{e}^{-x}} = \frac{\mathsf{e}^x}{1+\mathsf{e}^x}. \tag{4.14}$$

Thus the posterior of any data point $\boldsymbol{\phi}^{(i)}$ can be determined if we knew the parameters $p, \boldsymbol{\mu}_1, \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ in Eq. (4.13). Using the training data points $\left\{\boldsymbol{\phi}^{(i)}, y^{(i)}\right\}_{i\in\mathcal{N}_{\mathrm{tr}}}$, our goal is to find the most probable parameters $p, \boldsymbol{\mu}_1, \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ that maximize the following [17, Eq. (4.71)]

$$\mathsf{p}\left(\left\{\boldsymbol{\phi}^{(i)}, y^{(i)}\right\}_{i\in\mathcal{N}_{\mathrm{tr}}}; p, \boldsymbol{\mu}_1, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}\right) = \prod_{i=1}^{N_{\mathrm{tr}}} \Big(p \cdot \mathsf{p}\Big(\boldsymbol{\phi}^{(i)}|\mathcal{C}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}\Big)\Big)^{y^{(i)}} \Big((1-p) \cdot \mathsf{p}\Big(\boldsymbol{\phi}^{(i)}|\mathcal{C}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}\Big)\Big)^{1-y^{(i)}}, \tag{4.15}$$

where we have assumed that $\left\{\boldsymbol{\phi}^{(i)}, y^{(i)}\right\}_{i\in\mathcal{N}_{\rm tr}}$ are independent of each other given the parameters. There exists closed form solution for the parameters in the optimization problem of Eq. (4.15) [17, pp. 201-202]:

$$\widehat{p} = \frac{\sum_{i\in\mathcal{N}_{\rm tr}} y^{(i)}}{N_{\rm tr}} \tag{4.16}$$

$$\widehat{\boldsymbol{\mu}}_1 = \frac{\sum_{i\in\mathcal{N}_{\rm tr}} y^{(i)}\boldsymbol{\phi}^{(i)}_{[2:L]}}{\sum_{i\in\mathcal{N}_{\rm tr}} y^{(i)}}, \qquad \widehat{\boldsymbol{\mu}}_0 = \frac{\sum_{i\in\mathcal{N}_{\rm tr}} \left(1-y^{(i)}\right)\boldsymbol{\phi}^{(i)}_{[2:L]}}{\sum_{i\in\mathcal{N}_{\rm tr}} \left(1-y^{(i)}\right)} \tag{4.17}$$

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}} = &\frac{\sum_{i\in\mathcal{N}_{\rm tr}} y^{(i)}\left(\boldsymbol{\phi}^{(i)}_{[2:L]}-\widehat{\boldsymbol{\mu}}_1\right)\left(\boldsymbol{\phi}^{(i)}_{[2:L]}-\widehat{\boldsymbol{\mu}}_1\right)^\top}{\sum_{i\in\mathcal{N}_{\rm tr}} y^{(i)}} \\ &+ \frac{\sum_{i\in\mathcal{N}_{\rm tr}} \left(1-y^{(i)}\right)\left(\boldsymbol{\phi}^{(i)}_{[2:L]}-\widehat{\boldsymbol{\mu}}_0\right)\left(\boldsymbol{\phi}^{(i)}_{[2:L]}-\widehat{\boldsymbol{\mu}}_0\right)^\top}{\sum_{i\in\mathcal{N}_{\rm tr}} \left(1-y^{(i)}\right)}.\end{aligned} \tag{4.18}$$

The posterior of any input feature test point $\boldsymbol{\phi}^{(i)}, i\in\mathcal{N}_{\rm test}$, can be calculated by substituting Eqs. (4.16)–(4.18) in Eq. (4.13) and evaluating it at the specific $\boldsymbol{\phi}^{(i)}$. It is noted that, when matrix $\widehat{\boldsymbol{\Sigma}}$ is non-invertible, then its pseudoinverse, $\widehat{\boldsymbol{\Sigma}}^\dagger$, can be utilized instead in Eq. (4.13).

4.5.2 Logistic Regression

Logistic regression (LR) method does not assume any specific form for the likelihood, but instead, assumes that posterior distribution of any input feature data point has the following form

$$\mathsf{p}\Big(\mathcal{C}_1|\boldsymbol{\phi}^{(i)};\boldsymbol{\theta}\Big) = \mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big), \quad i=1,2,\ldots,N_{\rm tot}. \tag{4.19}$$

The goal of LR classification method is to estimate parameter $\boldsymbol{\theta}\in\mathbb{R}^L$, using the training data $\left\{\boldsymbol{\phi}^{(i)}, y^{(i)}\right\}_{i\in\mathcal{N}_{\rm tr}}$. Assuming that training data $\left\{\boldsymbol{\phi}^{(i)}, y^{(i)}\right\}_{i\in\mathcal{N}_{\rm tr}}$ are independent for any $\boldsymbol{\theta}$ and using that $\mathsf{p}\Big(\mathcal{C}_1|\boldsymbol{\phi}^{(i)};\boldsymbol{\theta}\Big) = 1-\mathsf{p}\Big(\mathcal{C}_0|\boldsymbol{\phi}^{(i)};\boldsymbol{\theta}\Big)$, the goal of LR method reverts on maximizing (w.r.t. $\boldsymbol{\theta}$) the following objective function

$$\mathsf{ln}\left(\mathsf{p}\left(\left\{y^{(i)}\right\}_{i\in\mathcal{N}_{\rm tr}}\,\middle|\,\left\{\boldsymbol{\phi}^{(i)}\right\}_{i\in\mathcal{N}_{\rm tr}};\boldsymbol{\theta}\right)\right) = \mathsf{ln}\left(\prod_{i=1}^{N_{\rm tr}}\mathsf{p}\Big(y^{(i)}\Big|\boldsymbol{\phi}^{(i)};\boldsymbol{\theta}\Big)\right), \tag{4.20}$$

which is equivalent to

$$\min_{\boldsymbol{\theta}}\left\{\mathsf{f}_{\rm LR}(\boldsymbol{\theta}) \triangleq -\sum_{i=1}^{N_{\rm tr}} y^{(i)}\mathsf{ln}\Big(\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big)\Big) + \left(1-y^{(i)}\right)\mathsf{ln}\Big(1-\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big)\Big)\right\}. \tag{4.21}$$

The optimization problem in (4.21) is convex in $\boldsymbol{\theta}$ and can be solved with a gradient-based or Newton-based optimization method. In this work the optimization problem in (4.21) is solved using the Newton method, that requires the gradient and the Hessian matrix of $\mathsf{f}_{\rm LR}(\boldsymbol{\theta})$. After

some elementary algebra we obtain [17]

$$\nabla_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) = \sum_{i=1}^{N_{\mathrm{tr}}} \left(\mathsf{h}\left(\boldsymbol{\theta}^{\top} \boldsymbol{\phi}^{(i)} \right) - y^{(i)} \right) \boldsymbol{\phi}^{(i)} \tag{4.22}$$

$$\nabla^2_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) = \sum_{i=1}^{N_{\mathrm{tr}}} \left[\left(1 - \mathsf{h}\left(\boldsymbol{\theta}^{\top} \boldsymbol{\phi}^{(i)} \right) \right) \cdot \mathsf{h}\left(\boldsymbol{\theta}^{\top} \boldsymbol{\phi}^{(i)} \right) \right] \boldsymbol{\phi}^{(i)} \left(\boldsymbol{\phi}^{(i)} \right)^{\top}. \tag{4.23}$$

Newton method uses the following iterative fixed point equation

$$\boldsymbol{\theta} := \boldsymbol{\theta} - \left(\nabla^2_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) \right)^{-1} \nabla_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}). \tag{4.24}$$

However, in many cases Hessian matrix in (4.23) is not invertible. To overcome this problem, the following regularized optimization problem is solved instead of (4.21),

$$\min_{\boldsymbol{\theta}} \left\{ \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) + \frac{\zeta_{\mathrm{LR}}}{2} \|\boldsymbol{\theta}\|_2^2 \right\} \tag{4.25}$$

Parameter ζ_{LR} is a regularization parameter and its value is determined by the structure of feature mapping $\boldsymbol{\phi}$. The optimization problem in (4.25) remains convex and, in a similar fashion with the problem in (4.21), iterative Newton-based method can be employed to find the optimal $\boldsymbol{\theta}$. The update fixed point equation for the problem in (4.25) takes the following form

$$\boldsymbol{\theta} := \boldsymbol{\theta} - \left(\nabla^2_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) + \zeta_{\mathrm{LR}} \mathbf{I}_L \right)^{-1} \left(\nabla_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) + \zeta_{\mathrm{LR}} \boldsymbol{\theta} \right). \tag{4.26}$$

The above fixed point iteration stops when

$$d_{\mathrm{LR}}(\boldsymbol{\theta}) \triangleq \frac{1}{2} \left(\nabla_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) + \zeta_{\mathrm{LR}} \boldsymbol{\theta} \right)^{\top} \left(\nabla^2_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) + \zeta_{\mathrm{LR}} \mathbf{I}_L \right)^{-1} \left(\nabla_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{LR}}(\boldsymbol{\theta}) + \zeta_{\mathrm{LR}} \boldsymbol{\theta} \right) < \varepsilon_{\mathrm{LR}}, \tag{4.27}$$

where $\sqrt{2 d_{\mathrm{LR}}(\boldsymbol{\theta})}$ is called Newton decrement at point $\boldsymbol{\theta}$ for the objective in (4.25), and $\varepsilon_{\mathrm{LR}}$ is a predetermined constant to control how close to the optimal objective value the fixed point iteration stops.

After obtaining the optimal solution for the problem in (4.21) or (4.25), we substitute the optimal $\boldsymbol{\theta}^{\star}$ in Eq. (4.19) and we calculate the posterior probabilities for the testing set data. Finally we apply the following rule

$$\widehat{y}^{(i)} = \begin{cases} 1, & \text{if } \mathsf{h}\left((\boldsymbol{\theta}^{\star})^{\top} \boldsymbol{\phi}^{(i)} \right) \geq \frac{1}{2}, \\ 0, & \text{if } \mathsf{h}\left((\boldsymbol{\theta}^{\star})^{\top} \boldsymbol{\phi}^{(i)} \right) < \frac{1}{2}, \end{cases} \qquad i \in \mathcal{N}_{\mathrm{test}}. \tag{4.28}$$

4.5.3 SVM

The support vector machine (SVM) approaches the classification problem through the concept of the margin, which is defined to be the smallest distance between the decision boundary and any of the input feature data samples in the dataset. SVM classify the feature data points in a non-probabilistic manner, using the following class decision rule

$$u^{(i)} = \mathsf{u}\left(\boldsymbol{\phi}^{(i)} \right) \triangleq \mathrm{sgn}\left(\boldsymbol{\theta}^{\top} \boldsymbol{\phi}^{(i)} \right) = \mathrm{sgn}\left(\theta_{[1]} + (\boldsymbol{\theta}_{[2:L]})^{\top} \boldsymbol{\phi}^{(i)}_{[2:L]} \right), \quad i = 1, 2, \ldots, N_{\mathrm{tot}}, \tag{4.29}$$

and parameter vector $\boldsymbol{\theta} \in \mathbb{R}^L$ needs to be estimated so as to maximize the smallest distance between the decision boundary and any of the input feature data samples in the training dataset. Let $\mathbf{u}_{\text{tr}} = \begin{bmatrix} u^{(1)} & u^{(2)} & \dots & u^{(N_{\text{tr}})} \end{bmatrix}^\top$ and $\mathbf{y}_{\text{tr}} = \begin{bmatrix} y^{(1)} & y^{(2)} & \dots & y^{(N_{\text{tr}})} \end{bmatrix}^\top$ denote, respectively, the column-wise concatenation of all $u^{(i)}$ and output labels $y^{(i)}$ associated with the training set. It turns out that the solution for the SVM problem is the optimal solution of the following convex optimization problem

$$\begin{aligned} \underset{\boldsymbol{\theta},\boldsymbol{\xi}}{\text{minimize}} \quad & \zeta \boldsymbol{\xi}^\top \mathbf{1}_{N_{\text{tr}}} + \frac{1}{2} \left\| \boldsymbol{\theta}_{[2:L]} \right\|_2^2 \\ \text{subject to} \quad & \boldsymbol{\xi} \succeq \mathbf{0}_{N_{\text{tr}}} \\ & \mathbf{u}_{\text{tr}} \odot (2\mathbf{y}_{\text{tr}} - \mathbf{1}_{N_{\text{tr}}}) \succeq \mathbf{1}_{N_{\text{tr}}} - \boldsymbol{\xi}, \end{aligned} \tag{4.30}$$

where $\boldsymbol{\xi}$ is a slack vector that allows some of the training points to be misclassified and ζ is a regularization parameter that models the trade-off between minimizing training errors and controlling model complexity. After obtaining the optimal parameter $\boldsymbol{\theta}^\star$ of (4.30), a new testing data $\boldsymbol{\phi}^{(i)}$, $i \in \mathcal{N}_{\text{test}}$, is classified as LoS or NLoS according to

$$\widehat{y}^{(i)} = \frac{1}{2} \text{sgn}\Big(\left(\boldsymbol{\theta}^\star\right)^\top \boldsymbol{\phi}^{(i)} \Big) + \frac{1}{2}, \quad i \in \mathcal{N}_{\text{test}}. \tag{4.31}$$

Alternatively, the dual version of (4.30) can be employed, which makes use of the equivalent kernel associated with feature mapping $\boldsymbol{\phi}$, $\mathsf{k}_{\boldsymbol{\phi}}(\mathbf{a}, \mathbf{b}) = \boldsymbol{\phi}_{[2:L]}(\mathbf{a})^\top \boldsymbol{\phi}_{[2:L]}(\mathbf{b})$, with $\mathbf{a}, \mathbf{b} \in \mathbb{R}^M$. Let $\mathbf{K}_{\text{tr}}$ be the kernel matrix associated with the training features, having elements $K_{\text{tr}[i,j]} = \mathsf{k}_{\boldsymbol{\phi}}\big(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\big) = \Big(\boldsymbol{\phi}_{[2:L]}^{(i)}\Big)^\top \boldsymbol{\phi}_{[2:L]}^{(j)}$, with $i, j \in \mathcal{N}_{\text{tr}}$. The dual of convex program in (4.30) is also a convex program, expressed as

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{maximize}} \quad & \boldsymbol{\alpha}^\top \mathbf{1}_{N_{\text{tr}}} - \frac{1}{2} \boldsymbol{\alpha}^\top \Big(\Big((2\mathbf{y}_{\text{tr}} - \mathbf{1}_{N_{\text{tr}}}) (2\mathbf{y}_{\text{tr}} - \mathbf{1}_{N_{\text{tr}}})^\top \Big) \odot \mathbf{K}_{\text{tr}} \Big) \boldsymbol{\alpha} \\ \text{subject to} \quad & \zeta \mathbf{1}_{N_{\text{tr}}} \succeq \boldsymbol{\alpha} \succeq \mathbf{0}_{N_{\text{tr}}} \\ & (2\mathbf{y}_{\text{tr}} - \mathbf{1}_{N_{\text{tr}}})^\top \boldsymbol{\alpha} = 0, \end{aligned} \tag{4.32}$$

where ζ is the same regularization parameter with the one in convex program of (4.30). Let $\boldsymbol{\alpha}^\star$ be denoting the optimal solution of the dual problem in (4.32). Let

$$\mathcal{S}(\boldsymbol{\alpha}^\star) \triangleq \big\{ i \in \{1, 2, \dots, N_{\text{tr}}\} : \alpha_{[i]}^\star > 0 \big\} \tag{4.33}$$

$$\mathcal{M}(\boldsymbol{\alpha}^\star, \zeta) \triangleq \big\{ i \in \{1, 2, \dots, N_{\text{tr}}\} : 0 < \alpha_{[i]}^\star < \zeta \big\}, \tag{4.34}$$

be the set of support vectors indexes that may have been misclassified and the set of support vectors indexes that have been classified correctly, respectively. Then the optimal parameter associated with the constant feature term is given by [17, Eq. (7.37)]

$$\theta_{[1]}^\star = \frac{1}{|\mathcal{M}(\boldsymbol{\alpha}^\star, \zeta)|} \sum_{j \in \mathcal{M}(\boldsymbol{\alpha}^\star, \zeta)} \left(\big(2y^{(j)} - 1\big) - \sum_{i \in \mathcal{S}(\boldsymbol{\alpha}^\star)} \big(2y^{(i)} - 1\big) \cdot \alpha_{[j]}^\star \cdot K_{\text{tr}[j,i]} \right). \tag{4.35}$$

A test feature $\boldsymbol{\phi}^{(i)}$, $i \in \mathcal{N}_{\text{test}}$, is classified as LoS or NLoS according to the following rule [17, Eq. (7.13)]

$$\widehat{y}^{(i)} = \frac{1}{2} \text{sgn}\left(\theta_{[1]}^\star + \sum_{j=1}^{N_{\text{tr}}} \alpha_{[j]}^\star \cdot \big(2y^{(j)} - 1\big) \cdot \mathsf{k}_{\boldsymbol{\phi}}\big(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\big) \right) + \frac{1}{2}, \quad i \in \mathcal{N}_{\text{test}}. \tag{4.36}$$

4.5.4 RVM

The relevance vector machine (RVM) method uses a probabilistic framework, quite similar to the LR. The main difference is that the parameter $\boldsymbol{\theta}$ is treated as Gaussian RV. More specifically, accordingly with LR method, RVM method models the LoS class posterior as

$$\mathsf{p}\Big(\mathcal{C}_1\big|\boldsymbol{\phi}^{(i)},\boldsymbol{\theta}\Big)=\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big),\quad i=1,2,\ldots,N_{\text{tot}}. \tag{4.37}$$

Parameter $\boldsymbol{\theta}$ is modeled as a Gaussian vector with zero mean and diagonal covariance, i.e.,

$$\mathsf{p}(\boldsymbol{\theta};\mathbf{w})=\mathcal{N}(\boldsymbol{\theta};\mathbf{0}_L,\mathbf{W}^{-1})=\mathcal{N}\left(\boldsymbol{\theta};\mathbf{0}_L,\begin{bmatrix}1/w_{[1]} & 0 & \cdots & 0\\ 0 & 1/w_{[2]} & \cdots & 0\\ \vdots & \cdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1/w_{[L]}\end{bmatrix}\right). \tag{4.38}$$

For a fixed value of hyper-parameter $\mathbf{w}$, RVM method uses the training data $\Big\{\boldsymbol{\phi}^{(i)},y^{(i)}\Big\}_{i\in\mathcal{N}_{\text{tr}}}$ to maximize w.r.t. $\boldsymbol{\theta}$ the posterior $\mathsf{p}\Bigg(\boldsymbol{\theta}\Big|\Big\{\boldsymbol{\phi}^{(i)},y^{(i)}\Big\}_{i\in\mathcal{N}_{\text{tr}}};\mathbf{w}\Bigg)$. It turns out that the latter problem is equivalent to minimizing w.r.t. $\boldsymbol{\theta}$ the following [17, Eq. (7.109)]

$$\min_{\boldsymbol{\theta}}\left\{\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta};\mathbf{w})\triangleq-\sum_{i=1}^{N_{\text{tr}}}y^{(i)}\ln\Big(\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big)\Big)+\big(1-y^{(i)}\big)\ln\Big(1-\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big)\Big)+\frac{1}{2}\boldsymbol{\theta}^\top\mathbf{W}\boldsymbol{\theta}\right\}. \tag{4.39}$$

Accordingly with LR case, after some algebra, the gradient and the Hessian can be obtained as

$$\nabla_{\boldsymbol{\theta}}\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta})=\left(\sum_{i=1}^{N_{\text{tr}}}\Big(\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big)-y^{(i)}\Big)\boldsymbol{\phi}^{(i)}\right)+\mathbf{W}\boldsymbol{\theta} \tag{4.40}$$

$$\nabla^2_{\boldsymbol{\theta}}\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta})=\left(\sum_{i=1}^{N_{\text{tr}}}\Big[\Big(1-\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big)\Big)\cdot\mathsf{h}\Big(\boldsymbol{\theta}^\top\boldsymbol{\phi}^{(i)}\Big)\Big]\boldsymbol{\phi}^{(i)}\Big(\boldsymbol{\phi}^{(i)}\Big)^\top\right)+\mathbf{W}. \tag{4.41}$$

Since the problem in Eq. (4.39) is convex can be solved by the Newton method. The fixed point iteration of Newton method for RVM takes the following form

$$\boldsymbol{\theta}:=\boldsymbol{\theta}-\big(\nabla^2_{\boldsymbol{\theta}}\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta})\big)^{-1}\nabla_{\boldsymbol{\theta}}\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta}). \tag{4.42}$$

The above fixed point iteration stops when

$$d_{\text{RVM}}(\boldsymbol{\theta})\triangleq\frac{1}{2}\left(\nabla_{\boldsymbol{\theta}}\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta})\right)^\top\big(\nabla^2_{\boldsymbol{\theta}}\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta})\big)^{-1}\nabla_{\boldsymbol{\theta}}\mathsf{f}_{\text{RVM}}(\boldsymbol{\theta})<\varepsilon_{\text{RVM}}. \tag{4.43}$$

The optimization problem in (4.39) is solved assuming fixed value for hyper-parameter $\mathbf{w}$. If we want to optimize with respect to both $\boldsymbol{\theta}$ and $\mathbf{w}$ we apply the iterative rules found in [17, p. (355)]. First parameter $\boldsymbol{\theta}$ and hyper-parameter $\mathbf{w}$ are initialized, and then the following coupled equations are solved iteratively until the variation on the parameters is less than a

Table 2: Comparison of Classification Algorithms: Feature Mapping Simulation Cases

Scenario No.	Feature Mapping	Feature Parameter Values	Size of $\boldsymbol{\theta}$
No. 1	$\boldsymbol{\phi}_{\mathrm{I}}$	No parameters	4
No. 2	$\boldsymbol{\phi}_{\mathrm{P},D}$	$D = 2$	13
No. 3	$\boldsymbol{\phi}_{\mathrm{P},D}$	$D = 3$	40
No. 4	$\boldsymbol{\phi}_{\mathrm{P},D}$	$D = 4$	121
No. 5	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 2$	7
No. 6	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 3$	10
No. 7	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 4$	13
No. 8	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 5$	16
No. 9	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 6$	19
No. 10	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 7$	22
No. 11	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 8$	25
No. 12	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 9$	28
No. 13	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 10$	31
No. 14	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 11$	34
No. 15	$\boldsymbol{\phi}_{\mathrm{L},D}$	$D = 12$	37

predetermined threshold ϵ_{RVM}

$$\boldsymbol{\theta} := \arg\min_{\boldsymbol{\theta}'} \{\mathsf{f}_{\mathrm{RVM}}(\boldsymbol{\theta}'; \mathbf{w})\} \tag{4.44}$$

$$\boldsymbol{\Sigma} := \left(\nabla^2_{\boldsymbol{\theta}} \mathsf{f}_{\mathrm{RVM}}(\boldsymbol{\theta})\right)^{-1} \tag{4.45}$$

$$\gamma_{[i]} := 1 - \Sigma_{[i,i]} w_{[i]}, \quad i = 1, 2, \ldots, N_{\mathrm{tr}} \tag{4.46}$$

$$w_{[i]} := \frac{\gamma_{[i]}}{\left(\theta_{[i]}\right)^2}, \qquad i = 1, 2, \ldots, N_{\mathrm{tr}}. \tag{4.47}$$

After finding the optimal $\boldsymbol{\theta}$ using Eqs. (4.44)–(4.47), we plug it in Eq. (4.37), and we apply the rule in Eq. (4.28) for the test data.

Another interesting attribute of RVM method is that it can be "kernelized". Specifically, the kernelized RVM models the posterior of class LoS as

$$\mathsf{p}\Big(\mathcal{C}_1 \big| \boldsymbol{\phi}^{(i)}, \boldsymbol{\theta}\Big) = \mathsf{h}\left(\theta_{[1]} + \left(\sum_{j=1}^{N_{\mathrm{tr}}} \theta_{[j+1]} \boldsymbol{\phi}^{(j)}_{[2:N]}\right)^{\top} \boldsymbol{\phi}^{(i)}_{2:N}\right) = \mathsf{h}\left(\theta_{[1]} + \sum_{j=1}^{N_{\mathrm{tr}}} \theta_{[j+1]} \mathsf{k}_{\boldsymbol{\phi}}\big(\mathbf{x}^{(j)}, \mathbf{x}^{(i)}\big)\right), \tag{4.48}$$

$i = 1, 2, \ldots, N_{\mathrm{tot}}$, with $\mathsf{k}_{\boldsymbol{\phi}}$ defined as in (4.8) and $\boldsymbol{\theta} \in \mathbb{R}^{N_{\mathrm{tr}}+1}$. We can apply the same reasoning with non-kernelized RVM method described through Eqs. (4.38)–(4.47) and obtain similar results.

5 Numerical Results: Comparison of Classification Algorithms

The following machine learning algorithms are compared for the NLoS identification problem in bistatic scatter radio:

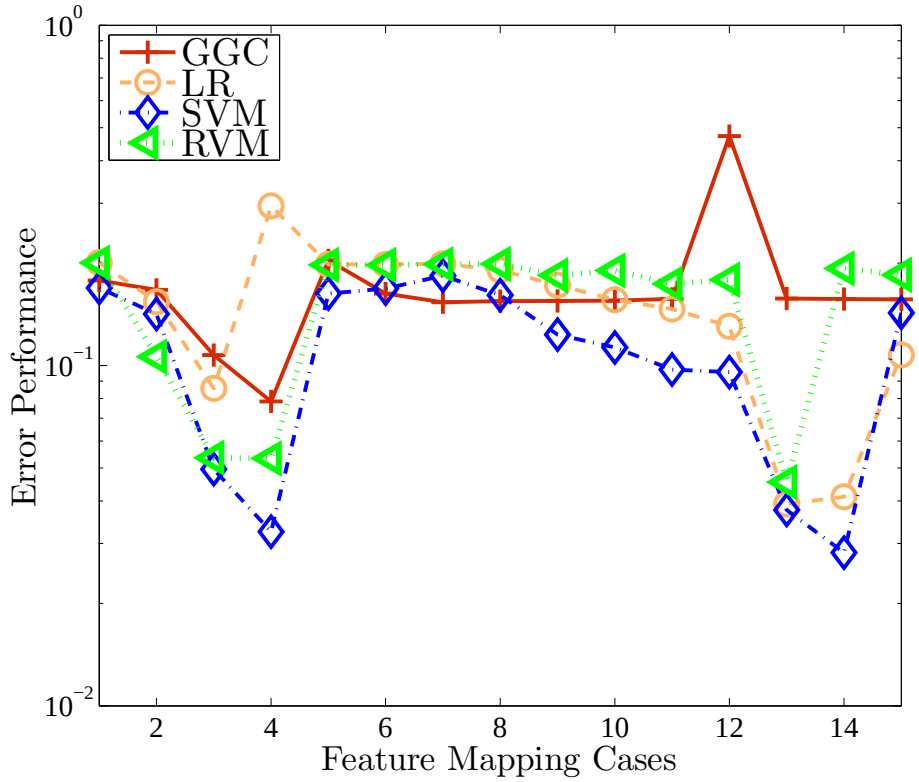

Figure 5: Classification algorithms error comparison for the feature mapping scenarios of Table 2.

- generative Gaussian classification (GGC)
- logistic regression (LR)
- support vector machines (SVMs), and
- relevance vector machines (RVMs).

We obtained $N_{\text{meas}} = 199$ LoS and NLoS measurements for $K = 6$ different distance setups. Thus, we collected $N_{\text{tot}} = 2388$ data points and the corresponding labels. Each original input data point is described by Eqs. (4.1) and (4.2). The feature mappings of Section 4.3 are applied on input data. We use $L_{\text{tr}}^{\text{LoS}} = L_{\text{tr}}^{\text{NLoS}} = 50$, thus $N_{\text{tr}} = 600$ training features are utilized for training, and the rest, for testing.

Using the cross-validation method described in Section 4.2, we run 1000 Monte Carlo (MC) experiments. For the mth MC experiment, let $\left\{\widehat{y}_{\text{a},m}^{(i)}\right\}_{i\in\mathcal{N}_{\text{test}}}$ be denoting the estimated labels on testing data for algorithm $\text{a} \in \{\text{GGC}, \text{LR}, \text{SVM}, \text{RVM}\}$ at the mth experiment (out of 1000), while let $\left\{y^{(i)}\right\}_{i\in\mathcal{N}_{\text{test}}}$ be denoting the true labels of the testing data. Then, the average misclassification error for algorithm a after 1000 experiments is given by

$$e_{\text{a}} = \frac{1}{1000}\sum_{m=1}^{1000}\left(\frac{1}{N_{\text{test}}}\sum_{i\in\mathcal{N}_{\text{test}}} 1\{\widehat{y}_{\text{a},m}^{(i)} \neq y^{(i)}\}\right), \quad \text{a} \in \{\text{GGC}, \text{LR}, \text{SVM}, \text{RVM}\}, \tag{5.1}$$

It is noted that, the choice of regularization values for feature mapping cases for LR and SVM and kernel cases for SVM, as well as the choice of the initialization algorithmic hyperparameter values for LR and RVM, were chosen empirically after conducting several simulation results.

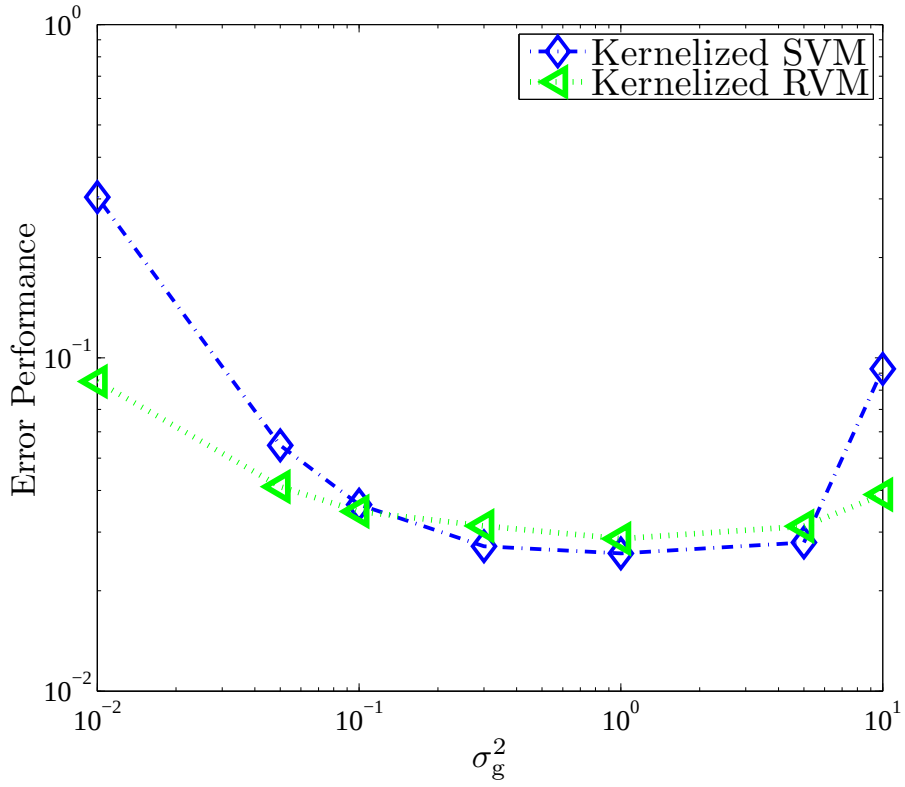

Figure 6: Kernelized SVM and RVM error performance as a function of Gaussian kernel parameter σ_{g}^2.

Table 2 illustrates the different feature mapping scenarios considered in this work. Regarding the polynomial feature mapping $\boldsymbol{\phi}_{\mathrm{P},D}$ in Eq. (4.5), increasing the value of polynomial degree D, the size of parameter vector θ increases as $\mathcal{O}\big(M^D\big)$, which is exponential on the input space size. Hence, for our case where $M = 3$, the maximum studied degree for $\boldsymbol{\phi}_{\mathrm{P},D}$ was $D = 4$. On the other hand, the size of parameter vector $\boldsymbol{\theta}$ for the log-polynomial feature mapping $\boldsymbol{\phi}_{\mathrm{L},D}$ in Eq. (4.6) is $\mathcal{O}(MD)$, which is linear in both M and D. In that case the maximum degree D that we considered was $D = 12$.

Fig. 5 compares average error performance in (5.1) for a $\in \{\mathrm{GGC}, \mathrm{LR}, \mathrm{SVM}, \mathrm{RVM}\}$ for the feature mappings scenarios given in Table 2. The average error performance for GGC is very poor over a large number of feature mappings scenarios. The main reason of that poor performance, is that GGC algorithm assumes Gaussian likelihoods in Eqs. (4.11)-(4.12), which is a strong assumption, and in our case, this strong assumption does not hold. We observe that SVM outperforms the other classification algorithms in terms of misclassification error for a wide range of feature mapping scenarios. The best performance overall schemes and overall feature mapping scenarios is achieved by SVM in conjunction with log-polynomial feature mapping of degree $D = 11$. Very similar error performance is achieved by SVM classification for the polynomial feature mapping with degree $D = 4$. For the log-polynomial feature mapping we observe that the best degree for LR and RVM in terms of average error is achieved for degree $D = 10$. We also observe that LR offers very good error performance; in many cases comparable with SVM. It is noted that GGC and LR methods have quite lower computational cost compared to SVM and RVM. Thus, LR offers a good error performance-complexity trade-off.

Next, we compare the kernelized version of SVM method with RVM method. Both algorithms utilize the Gaussian kernel defined in (4.9). We consider 6 different simulation scenarios; each scenario utilizes different Gaussian kernel parameter value $\sigma_{\rm g}^2$. In contrast to feature mapping cases, both kernelized algorithms offer size of parameter vector θ equal to $N_{\rm tr} = 601$. Observing Table 2, the above size is quite larger compared to the one offered by any feature mapping scenario. However, as we noted after much experimentation, this increase in complexity, offers improved and more stable average error performance for both algorithms. Fig. 6 illustrates the average misclassification error performance of the algorithms as a function of $\sigma_{\rm g}^2$. It is noted that the best performance of both algorithms is achieved for $\sigma_{\rm g}^2 = 1$. It is remarked that SVM slightly outperforms RVM for $\sigma_{\rm g}^2 \in (0.1, 5]$, while otherwise, RVM exhibits a large average error performance gap compared compared to SVM.

References

[1] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*, 2nd ed. New York, NY, USA: John Wiley & Sons, Inc., 2003.

[2] N. Fasarakis-Hilliard, P. N. Alevizos, and A. Bletsas, "Coherent detection and channel coding for bistatic scatter radio sensor networking," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1798–1810, May 2015.

[3] P. Corke, T. Wark, R. Jurdak, P. Valencia, and D. Moore, "Environmental wireless sensor networks," *Proc. IEEE*, vol. 98, no. 11, pp. 1903–1917, Oct. 2010.

[4] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," *Proc. IEEE*, vol. 98, no. 11, pp. 1947–1960, Sep. 2010.

[5] F. Zhao and L. Guibas, *Wireless Sensor Networks: An Information Processing Approach*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004.

[6] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[7] N. Kargas, F. Mavromatis, and A. Bletsas, "Fully-coherent reader with commodity SDR for Gen2 FM0 and computational RFID," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 617–620, 2015.

[8] K. Yu and Y. J. Guo, "Statistical NLOS identification based on AOA, TOA, and signal strength," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 274–286, 2009.

[9] Z. Xiao, H. Wen, A. Markham, N. Trigoni, P. Blunsom, , and J. Frolik, "Identification and mitigation of non-line-of-sight conditions using received signal strength," in *Proceedings of the 9th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2013)*, Lyon, France, October 2013.

[10] S. Marano, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE J. Select. Areas Commun.*, vol. 28, no. 7, pp. 1026–1035, 2010.

[11] T. V. Nguyen, Y. Jeong, H. Shin, and M. Z. Win, "Machine learning for wideband localization," *IEEE J. Select. Areas Commun.*, vol. 33, no. 7, pp. 1357–1380, 2015.

[12] E. Alimpertis, "Smart sensors of RF and backscatter signals with localization," Master's thesis, Technical University of Crete, Aug. 2014, supervisor A. Bletsas.

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

[14] B. C. Levy, *Principles of Signal Detection and Parameter Estimation.* New York: Springer, 2008.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

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

[17] C. M. Bishop, *Pattern Recognition and Machine Learning.* New York, NY: Springer, 2006.