



Article DoA Estimation Using Neural Tangent Kernel under Electromagnetic Mutual Coupling

Qifeng Wang^{1,*}, Xiaolin Hu¹, Xiaobao Deng¹ and Nicholas E. Buris^{2,†}

- ¹ School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China; xiaolinhu@shu.edu.cn (X.H.); Xiaobaodeng@i.shu.edu.cn (X.D.)
- ² Amazon, Sunnyvale, CA 94089, USA; nick.buris@ieee.org
- * Correspondence: wangqifeng77@shu.edu.cn
- + He was with the School of Communication and Information Engineering, Shanghai University.

Abstract: Antenna element mutual coupling degrades the performance of Direction of Arrival (DoA) estimation significantly. In this paper, a novel machine learning-based method via Neural Tangent Kernel (NTK) is employed to address the DoA estimation problem under the effect of electromagnetic mutual coupling. NTK originates from Deep Neural Network (DNN) considerations, based on the limiting case of an infinite number of neurons in each layer, which ultimately leads to very efficient estimators. With the help of the Polynomial Root Finding (PRF) technique, an advanced method, NTK-PRF, is proposed. The method adapts well to multiple-signal scenarios when sources are far apart. Numerical simulations are carried out to demonstrate that this NTK-PRF approach can handle, accurately and very efficiently, multiple-signal DoA estimation problems with realistic mutual coupling.

Keywords: direction of arrival; mutual coupling; neural tangent kernel; polynomial root finding

1. Introduction

Direction of arrival estimation has been at the forefront of research activity for several decades in a variety of areas including radar, sonar, wireless communications, and radio astronomy [1]. DoA estimation is a long-lasting research topic in array signal processing and various methods have been proposed, such as the conventional beamforming (CBF) method [2], multiple signal classification (MUSIC) method [3], estimating signal parameters via rotational invariant techniques (ESPRIT) [4], maximum likelihood (ML) method [5] and compressive sensing (CS)-based method [6,7]. CS is a promising technique that uses the output of received sensors to determine the DoAs directly, thus the computation of the covariance matrix is avoided [8]. However, the application of these traditional methods to real systems has been very limited, since they critically rely on exact knowledge of the array manifold. The array manifold is often affected by various perturbations in practical applications, which seriously degrade the performance of high-resolution DoA estimations.

Mutual coupling among the elements of an array is inevitable for real antennas. In the presence of mutual coupling, the performance of DoA estimation may degrade substantially without array manifold calibration. Taking account of this issue, Friedlander proposes an approximate representation of the coupling via a simple Mutual Coupling Matrix (MCM) in the form of a banded symmetric Toeplitz matrix [9]. In the model, the mutual coupling coefficients between sensors are inversely related to their distance. This direction-independent model has been used by many investigators [10–14]. However, the exact electromagnetic coupling is more complicated than the model captured by a Toeplitz matrix. Direction-dependent mutual coupling is taken into account [15–18], where the mutual coupling effect will vary with the source direction. In order to overcome the effect of mutual coupling, there are many techniques developed, which can be roughly classified into three categories: Auxiliary array or subarray [10,18], sparse reconstruction



Citation: Wang, Q.; Hu, X.; Deng, X.; Buris, N.E. DoA Estimation Using Neural Tangent Kernel under Electromagnetic Mutual Coupling. *Electronics* **2021**, *10*, 1057. https:// doi.org/10.3390/electronics10091057

Academic Editor: Rafael González Ayestarán

Received: 22 March 2021 Accepted: 25 April 2021 Published: 29 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). algorithms [11,13] and subspace-based algorithms [14–17]. However, these studies are based on simplified models, and, to a certain extent, these assumptions do not hold in practice.

Another array "imperfection" scenario which degrades the efficacy of traditional DoA estimation algorithms is the sometimes-unavoidable situation of having to use an irregular grid array [19]. In some scenarios, however, irregular grid arrays can provide decisive advantages for wireless communication compared to the regular, i.e., uniform, grid arrays [20]. The MUSIC algorithm can be universally applied to irregular arrays. However, it has high computational cost owing to the exhaustive search. ESPRIT is a computational efficient method, but is only applicable to uniform arrays. A conventional idea is to transform the irregular array into a ULA. The array interpolation method [21] is proposed to approximate an arbitrary array by a virtual ULA. However, the drawback of this method is that the interpolation is not accurate enough. Another prevalent array transformation method, called the Manifold Separation Technique (MST) [22], models the arbitrary array steering vector as a product of a sampling matrix and a Vandermonde structured coefficients vector that has the same structure as the ULA steering vector. However, the performance of MST is rather poor in low-SNR scenarios. Furthermore, it requires a large number of virtual arrays to achieve satisfactory performance, which demands increased complexity. Moreover, in the presence of combined effects of multiple types of array "imperfections", parametric methods may encounter great difficulties. Although the NTK-PRF can handle both the mutual coupling as well as the irregular grid arrays very well, the latter are left for a future paper.

Recently, machine learning-based methods have proved powerful for DoA estimation. For example, Support Vector Regression (SVR) is able to estimate DoA by establishing training sets of the possible configurations of the impinging sources, and then deriving a mapping from array outputs to signal directions [23,24]. The derived mapping is then used on test data to estimate DoA. However, the performance of these methods degrades rapidly when dealing with unknown configurations, e.g., a varying number of elements in the array or number of sources. To solve this problem, a novel framework based on a deep neural network that partitions the search area into several sectors is proposed [25]. Liu [26] proposes an end-to-end DNN to address the DoA estimation problem. The proposed DNNbased framework consists of a multitask auto-encoder and a series of parallel multilayer classifiers, which improve the ability in adaptation to multiple signals. However, these networks need to be trained offline and it costs a lot of time. The networks critically rely on the training data. When the SNR changes, in order to ensure the accuracy of the estimation, the network needs to be retrained. Therefore, these methods are computationally expensive for real-time applications. In the infinite-width limit, it has been demonstrated that DNNs have a Gaussian distribution described by a kernel that yields results comparable to DNNs trained with gradient descent [27]. In the same limit, the behavior of DNNs during training is described by a related kernel [28], which has low computation burden compared with DNN training.

In this paper, we solve the problem of DoA estimation of multiple incoming waves on a ULA in the presence inter-element coupling. The main contributions of this paper are:

- We leverage recent progress in modeling and deriving DNNs using kernel regression with a NTK to estimate DoA. In the infinite width limit, instead of training the network, the exact dynamics of its outputs are characterized by a differential equation (through gradient descent training). Numerical experiments show that the developed method has good performance in term of the estimation accuracy and computational time under the effect of electromagnetic mutual coupling.
- We incorporate the PRF feature into NTK to estimate multiple DoAs, which largely
 enhances the generalization of the proposed method to multiple arriving signals
 scenarios. This method performs well even though the training dataset represents
 single incoming waves.

• We use signal models with electromagnetically exact active electric fields of Multi Element Antenna (MEA) systems, including mutual coupling (captured by full S-parameter matrices).

For the convenience of readers, Table 1 lists the acronyms appeared in the paper.

Table 1. '	The acro	onyms.
------------	----------	--------

Acronyms	Description
DoA	Direction of Arrival
NTK	Neural Tangent Kernel
PRF	Polynomial Root Finding
DNN	Deep Neural Network
SVR	Support Vector Regression
CBF	Conventional Beamforming
MUSIC	Multiple Signal Classification
ESPRIT	Estimating Signal Parameters via Rotational Invariant Techniques
CS	Compressive Sensing
ML	Maximum Likelihood
MCM	Mutual Coupling Matrix
MST	Manifold Separation Technique
ULA	Uniform Linear Array
MEA	Multi Element Antenna
LoS	Line of Sight
MIMO	Multi Input Multi Output
UAT	Universal Approximation Theorem
ReLU	Rectified Linear Unit
SNR	Signal-to-Noise Ratio
RMSE	Root Mean Squared Error

Notations: Matrices and vectors are denoted by upper case and lower case bold characters respectively. $[\cdot]_{ij}$ denotes the (i, j)-th element in a matrix. A $M \times N$ matrix with complex entries is denoted by $\mathbb{C}^{M \times N}$. $[\cdot]^T$ and $[\cdot]^H$, respectively, stand for the transpose and conjugate transpose. $\|\cdot\|$ denotes the Euclidean norm. $\mathbb{E}[\cdot]$ is the expectation of a random variable.

The rest of this paper is organized as follows. Section 2 introduces our system model, and proposes developed algorithm. The baseline methods are introduced in this section. Section 3 carries out numerical simulations to validate the effectiveness of the proposed method. Section 4 states our conclusions.

2. System Model and Methodology

This section describes our proposed method of solving the DoA estimation problem by a general Rx antenna array. The method employs a Neural Tangent Kernel approach to estimate the DoA function from a set of induced voltages on the Rx array elements.

2.1. System Model and the Creation of Realistic Sensor Data

Without loss of generality, we consider N 1-port antenna Tx radios forming $N M \times 1$ MIMO wireless links in a free-space Line of Sight (LoS) propagation environment with an M-port Rx radio. The Rx M-port antenna system is general and need not be, for example, a regular uniform array. Ignoring noise and interference, the Rx voltages induced by the incident wave(s) from the Tx radio(s) are related by

$$V_m = \sum_{n}^{N} H_{mn} X_n, \tag{1}$$

where $m = 1, 2, \dots, M$ is the port index of the Rx, and $n = 1, 2, \dots, N$ is the index of the incoming plane wave. The columns of the H_{mn} matrix represent the $N M \times 1$ vector

channels of the aforementioned MIMO links. In this paper, we use MIMObit [29] in order to evaluate channel matrix **H** precisely and generate realistic sensor data (training datasets). The H_{mn} are calculated assuming LoS propagation and the exact knowledge of the Tx and the Rx antennas via an electromagnetics exact approach rendering the full S-par matrix plus the active E-field gains of each port [30]. The latter capture all the mutual coupling including the skewness of the radiation patterns and even active impedances and the termination effects of the antennas in the Rx antenna system. In that sense, the training data generated here is reasonably realistic for the purposes of evaluating and comparing the performance of DoA estimation algorithms. Adding noise to Equation (1), concatenation of all the *T* times and *M* ports renders the sensor response of the Rx as

$$\mathbf{V} = \mathbf{H}(\boldsymbol{\theta})\mathbf{X} + \mathbf{N},\tag{2}$$

where **V** is the $M \times T$ received signal matrix, **X** is the $N \times T$ source signal matrix, **N** represents the received $M \times T$ Gaussian noise matrix with zero mean and variance σ_n^2 , and **H**(θ) is a $M \times N$ channel matrix (*N* here pertains to the different DoA angles and not number of antennas in the same Tx).

The sensor response, **V**, in Equation (2) is a function of the angles of incidence, θ , of the incoming wave(s), i.e., **V** = F(θ), where F will be considered unknown for the purposes of the DoA estimation. Using *f* to denote the inverse function of F, we have

$$\boldsymbol{\theta} = f(\mathbf{V}),\tag{3}$$

Equation (3) implies that the DoA estimation problem can be reduced to a function fitting problem. However, the analytical expression of f is difficult to obtain, since f is a highly nonlinear function.

2.2. Neural Tangent Kernel Based DoA Estimation

The Universal Approximation Theorem (UAT) [31] guarantees that a deep neural network is able to represent any function. Let f in Equation (3) denote a fully-connected over-parameterized feedforward network with L hidden layers. Indicating with $d_h|_{h=1}^L$, the dimensionality of the h-th hidden layer, there are $d_0 = M$ inputs and $d_{L+1} = 1$ output. We initialize all the weights W and biases b to be identically distributed (i.i.d.) Gaussians $\mathcal{N}(0,1)$, and consider the limit of hidden widths $d_h \to \infty$. Given a training dataset $(\mathbf{V}, \boldsymbol{\theta}) = \{(v_i, \theta_i)\}_{i=1}^N$, where $v_i \in \mathbb{C}^M$ is an input vector and θ_i denotes the corresponding scalar output of a nonlinear function $\theta_i = f(v_i)$, the values of $\boldsymbol{\theta}_* = \{\theta_{*i}\}_{i=1}^p$ at a test dataset of P input points, $\mathbf{V}_* = \{v_{*i}\}_{i=1}^p$, can be estimated via the following kernel regression [32]:

$$\hat{\boldsymbol{\theta}}_* \approx \mathbf{K}_* \mathbf{K}^{-1} \boldsymbol{\theta},\tag{4}$$

where $\mathbf{K}_* \in \mathbb{R}^{P \times N}$ with $(\mathbf{K}_*)_{ij} = k(v_{*i}, v_j)$ is the NTK matrix between the *P* input points, \mathbf{V}_* , and the *N* points in the training dataset \mathbf{V} . $\mathbf{K} \in \mathbb{R}^{N \times N}$ with $\mathbf{K}_{ij} = k(v_i, v_j)$ is the NTK matrix between all the pairs of training points in \mathbf{V} . The scalar kernel function, k(x, x'), between two points, x and x', is defined by [28]

$$k(\mathbf{x}, \mathbf{x}') = \sum_{h=1}^{L+1} \left(\Sigma^{(h-1)}(\mathbf{x}, \mathbf{x}') \cdot \prod_{h'=h}^{L+1} \dot{\Sigma}^{(h')}(\mathbf{x}, \mathbf{x}') \right),$$
(5)

where $\Sigma^{(h)}(x, x')$ is a scalar covariance defined recursively as

$$\Sigma^{(h)}(\mathbf{x}, \mathbf{x}') = \mathbb{E}_{g \sim \mathcal{N}(0, \Sigma^{(h-1)})} \left[\sigma(g(\mathbf{x})) \sigma(g(\mathbf{x}')) \right] + 1,$$
(6)

$$\Sigma^{(0)}(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{M} \boldsymbol{x}^T \cdot \boldsymbol{x}' + 1,$$
(7)

with \mathbb{E} indicating the expected value operation and σ being an "activation function". Of all the options, the Rectified Linear Unit function (ReLU), defined as $f_{\text{ReLU}}(y) = max(y,0)$, is used as activation function throughout this paper. The derivative of $\Sigma^{(h)}(x, x')$ is also recursively calculated as

$$\dot{\boldsymbol{\Sigma}}^{(h)}(\boldsymbol{x},\boldsymbol{x}') = \mathbb{E}_{g \sim \mathcal{N}(0,\boldsymbol{\Sigma}^{(h-1)})} [\dot{\boldsymbol{\sigma}}(g(\boldsymbol{x}))\dot{\boldsymbol{\sigma}}(g(\boldsymbol{x}'))], \tag{8}$$

and $\dot{\Sigma}^{(L+1)}(\mathbf{x}, \mathbf{x}') = 1$ for convenience. The derivative of the activation function, σ , is indicated by $\dot{\sigma}$. In the derivation outlined in [32] with infinite number of neurons in each layer and infinite training iterations, the argument variables, $g(\mathbf{x})$, of the activation function, σ , are the output functions of corresponding layers of the DNN and are independent zero-mean Gaussian processes with covariance $\Sigma^{(h)}(\mathbf{x}, \mathbf{x}')$. That is $g \sim \mathcal{N}(0, \Sigma^{(h)})$.

However general this NTK-based DoA estimation approach is (i.e., it also applies to irregular grid arrays), it does not scale well when there are multiple incoming waves, unless we have and use additional information about the Rx antennas and/or a lot of training data for all the combinations of number of incoming signals, their strength and angles of arrival.

In the following sections, we will use a ULA at the Rx and, with that information (i.e., antenna topology) and the steering vector it implies, we will extend the applicability of NTK DoA estimation to multiple incoming waves.

2.3. NTK-Augmented Polynomial Root Finder DoA Estimation

Consider the signals an incoming wave induces on an *M*-element ULA Rx array which is oversimplified by ignoring coupling and by assuming isotropic radiation pattern for each antenna. Using the traditional MUSIC method steering vector,

$$\boldsymbol{a}(\theta) = \left[1, e^{-jkd\cos(\theta)}, \cdots, e^{-j(M-1)kd\cos(\theta)}\right]^{T},$$
(9)

yields the following model of the received signals:

$$\mathbf{V} = \boldsymbol{a}(\theta) \cdot \mathbf{E} + \mathbf{N}. \tag{10}$$

The autocorrelation of the Rx signal, $\mathbb{E}[\mathbf{V}\cdot\mathbf{V}^H]$ forms a symmetric polynomial in z and 1/z, where $z = e^{-jkd\cos(\theta)}$. I.e., if z_0 is a root of this polynomial, so is $1/z_0$. In the PRF method [32], the autocorrelation of the real data is used and the phase of the complex root pairs of the aforementioned polynomial provides estimates of the multiple DoAs. Thus the PRF enables the estimation of multiple DoA as they are the multiple complex root pairs, z and 1/z, of the autocorrelation polynomial.

In the present work, we generate realistic data using Equation (2) as mentioned earlier in Section 2.1. However, the steering vector in Equation (9) is not accurate and, therefore, the roots of the polynomial do not yield accurate DoA results. Therefore, in order to capture the effects of the coupling on the roots of the autocorrelation polynomial, we employ the NTK kernel regression method described in Section 2.2 to derive the mapping from polynomial roots to signal directions.

We generate realistic **V** data from Equation (2) for *N* known DoA angles. This data constitutes the training dataset $(\mathbf{V}, \boldsymbol{\theta}) = \{(\boldsymbol{v}_i, \theta_i)\}_{i=1}^N$. We then form the signal autocorrelation and proceed with the PRF method to obtain two complex roots for each DoA, $\boldsymbol{\theta}$. The training dataset then is transformed into a new training dataset, $(\mathbf{Z}, \boldsymbol{\theta}) = \{(\boldsymbol{z}_i, \theta_i)\}_{i=1}^N$, where $\boldsymbol{z}_i \in \mathbb{R}^4$ are the real and imaginary parts of the two complex roots corresponding to $\theta_i \in \mathbb{R}^1$.

The NTK-PRF method we propose for ULAs is shown in Algorithm 1.

Algorithm 1: NTK-PRF for multiple DoAs estimation

■Input: Training dataset $(\mathbf{V}, \boldsymbol{\theta}) = \{(v_i, \theta_i)\}_{i=1}^{N}$, where $v_i \in \mathbb{C}^M, \theta_i \in \mathbb{R}^1$ Test inputs \mathbf{V}_* with *P* signals, where $\mathbf{V}_* \in \mathbb{C}^{M \times T}$, P < M**Output**: estimated DoAs θ_* 1: for i = 1, *N* do 2: $\mathbf{U}_n \leftarrow \frac{1}{T} \sum_{j=1}^{T} v_i(t_j) v_i^H(t_j) = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^H + \sigma_n^2 \mathbf{U}_n \mathbf{U}_s^H$ 3: $z_i \in \mathbb{R}^4 \leftarrow \mathbf{p}^H(z) \mathbf{U}_n \mathbf{U}_n^H \mathbf{p}(z) = 0$, where $\mathbf{p}(z) = [1, z, \dots, z^{M-1}]^T$ 4: end 5: obtain new training data pairs $(\mathbf{Z}, \boldsymbol{\theta}) = \{(z_i, \theta_i)\}_{i=1}^{N}$ 6: $\mathbf{Z}_* \in \mathbb{R}^{P \times 4} \leftarrow \mathbf{V}_*$ is substituted for $v_i(t_j)$ and redo 2–3 without iteration. 7: initialize the NTK-parameterized networks 8: compute \mathbf{K}_* and \mathbf{K} using Equations (5) and (8) with \mathbf{Z} and \mathbf{Z}_* . 9: compute θ_* using Equation (4) 10: return θ_*

2.4. Baseline DoA Estimation Methods

We compare the performance of our proposed NTK-PRF method to the following baseline methods:

- Root-MUSIC [33]. This is one of the most common parametric methods suitable for uniform arrays. Here we use the steering vector in Equation (9), which ignores mutual coupling, but we generate realistic signal data with coupling using MIMObit (see Equation (1) and Reference [29]).
- ESPRIT [4]. This is another parameter estimation technique, based on the fact that the signal at one element is a constant phase shift from the earlier element in the steering vector. In the simulation, the signal data is same as the case for Root-MUSIC.
- Auxiliary array [10]. We add 6 auxiliary sensors on each side of the array to form a new ULA with the same element spacing compared to the original. In this way, the effect of mutual coupling between the middle sensors is reduced. Then the Root-MUSIC as mentioned above is used to estimate DoA.
- SVR. The support vector regression is trained using LIBSVM [34] with Gaussian kernel, defined as $\mathbf{K}(\mathbf{Z}_i, \mathbf{Z}) = exp(-\gamma || \mathbf{Z} \mathbf{Z}_i ||^2)$. A parameter $v \in (0, 1]$ is introduced to control the number of support vectors. The upper diagonal elements of the covariance matrix of the array output are utilized as features since they contain sufficient information about the received signals [24].
- DNN [26]. The framework of DNN consists of two parts, a multitask auto-encoder and a series of parallel multilayer classifiers. The multitask auto-encoder is introduced before the multilayer classifiers to decompose the inputs into 10 spatial subregions. For the training of the auto-encoder, the network is designed to be linear. The covariance matrix of the array outputs is computed and the upper diagonal elements are used as the input. The $[0^\circ, 180^\circ)$ space is sampled with an interval of 0.9° to obtain the training data in single-signal scenarios. For the training of classifiers, another data set is collected in the two-signal scenarios with inter-signal angle sampled from the set of $\{1.8^\circ, 3.6^\circ, \ldots, 36.04^\circ\}$. The network of each classifier has L = 3 hidden layers with ReLU as activation function.
- DNN-PRF. The neural network with ReLU as activation function is designed using TensorFlow [35], and we use the Adam optimizer to train the network with 50,000 iterations [36]. The number of the hidden layers is chosen to be L = 3, i.e., the same as with the NTK-PRF proposed here. The widths (i.e., number of neurons) of the three hidden layers are 64, 32, and 16, respectively. In order to make a more reasonable comparison with NTK-PRF, the polynomial root of the covariance matrix is calculated to train the neural network, which is similar to NTK-PRF.

Table 2 gives a comparison of the proposed method with baseline methods used in the simulation. Root-MUSIC, ESPRIT, and Auxiliary array are parametric methods, thus there

is no need to calibration data. For Toeplitz structured mutual coupling model, auxiliary sensors can reduce the effect of mutual coupling. However, the estimation accuracy of these three methods is poor for the signal data generated by MIMObit. The remaining four are learning-based methods, which require training (or calibration) data. The performance comparison of these methods will be shown in Section 3.

Algorithms	Need Calibration Data	Real Time	Multiple DoAs Estimation	Suitable for Mutual Coupling
Root-MUSIC	No	Yes	Yes	No
ESPRIT	No	Yes	Yes	No
Auxiliary Array	No	Yes	Yes	Yes
SVR	Yes	No	No	Yes
DNN	Yes	No	Yes	Yes
DNN-PRF	Yes	No	Yes	Yes
NTK-PRF	Yes	Yes	Yes	Yes

Table 2. Main distinctive characteristics between different methods.

3. Simulation Experiments and Discussion

In this paper, we examine DoA with a 6-port ULA of half-wave V-pol dipoles as a receiver (Figure 1a). We assume that the propagation environment is Line of Sight (LoS). The element spacing is $d = 0.3\lambda$ at the frequency of this study (f = 3.5 GHz). The antennas are also designed at the frequency of 3.5 GHz and the magnitude of all the S-par matrix elements of the ULA is below -8 dB (Figure 1b). The azimuth of signals is from the angular range of $[0^{\circ}, 180^{\circ})$. The elevation angle is fixed at 90° in all simulations and we simply estimate the azimuthal angle.



Figure 1. (a) 6-port ULA of V-pol half-wave dipoles. (b) The magnitude of all the EM-exact S-parameters of a 6-port ULA of dipoles used in the present work.

Numerical simulation results of the DoA estimation illustrate the performance of the proposed method against the baselines. In the NTK algorithm, we use a DNN with L = 3 hidden layers and the ReLU as activation function of each hidden layer. We also consider 1000 different positions of the Tx (i.e., 1000 different DoAs), uniformly distributed in $[0^{\circ}, 180^{\circ})$ space. The sensor data is then generated using Equation (2), and the covariance vectors of data sets are obtained from N = 1000 snapshots with random noise. A small number of uniformly selected DoAs are chosen as the training (or calibration) dataset while all the DoAs are used as the "test" dataset in order to test the performance of each algorithm. The Root Mean Square Error (RMSE) between the estimated DoA angles and

the true DoA angles is used as the performance metric for each of the DoA estimation algorithms considered, which is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{\theta}_i - \theta_i)^2}{N}},$$
(11)

where N is the number of total DoAs for testing, and $\hat{\theta}_i$ and θ_i denote the estimated DoA and true DoA of the *i*-th signal source respectively.

Table 3 lists the parameters of system setting.

Table 3. System setting.

-	-	Settings	
	Array	ULA, $d = 0.3\lambda$	
Rx	Antenna	6-port V-pol half-wavelength dipole	
	Frequency	3.5 GHz	
Tx	Antenna	1-port V-pol half-wavelength dipole	
Coometry	Tx-Rx distance	400 m	
Geometry	Spatial range	$[0^\circ, 180^\circ)$	
Propagation model	-	LoS	

Figure 2 depicts the performance comparison of different methods versus SNR. Root-MUSIC and ESPRIT can achieve high-resolution angle estimation with known array manifold. In the presence of unknown mutual coupling, however, their performances are severely degraded. The auxiliary array method can effectively reduce the effect of mutual coupling in the Toeplitz structure. However, the realistic electromagnetic mutual coupling included in the data here does not follow the Toeplitz structure. Thus, the performance of the auxiliary array method is not that well suited for this problem. SVR and DNN are very time consuming as they update parameters such as weight and bias at each iteration. The SVR method maps the upper triangle of the covariance matrix to the estimated angle. Although it has good estimation performance, its generalization to incoming signals from multiple angles is weak. In the DNN method, an auto-encoder preprocesses the original array outputs to reduce the distribution divergences of the input data first. Then, multilayer classifiers following a one-versus-all classification guideline are introduced to estimate multiple incoming signals. However, the estimation accuracy is reduced in the case of multiple incoming signals. The estimation error is relatively large especially when the angle of the signal is close to the axis of the array (i.e., 0 degrees) or close to the boundary of the training data angle space. The DNN-PRF here uses 50,000 iterations and only 64, 32 and 16 neurons in layers 1, 2 and 3, respectively. The NTK is derived as the limit of a neural network with infinite layer width and iterations, and is not an iterative algorithm. Hence, not unexpectedly, the NTK-PRF is more accurate than the DNN-PRF. The discrepancy is actually higher at lower SNR values. NTK-PRF method proposed here has better performance than each of the baseline methods in the presence of mutual coupling.



Figure 2. RMSE of a single signal DoA estimation with different methods vs. SNR for a 6-port dipole ULA Rx. Training data is at every 0.9°, and test data is at every 0.18°.

SNR(dB)

In the case of multiple arriving signals, the estimated DoAs of the proposed NTK-PRF method has higher accuracy compared with the baseline methods as shown in Table 4. The DNN method achieves satisfactory estimation accuracy in the case of two signals, since the classifiers of DNN are trained in two-signal scenarios. However, when more signals impinge onto the array simultaneously, this method results in high estimation error. Note that the proposed NTK-PRF method can detect multiple incoming waves even though it has been derived with training data from a single-angle of arrival (albeit said training data were 0.9° apart).

 Table 4. Multiple signal DoA estimation.

-	<i>N</i> = 2		N = 3		
SNR (dB)		5		5	
True DoAs (°)	30.09	134.95	30.09	74.95	134.95
Root-MUSIC (°)	21.27	133.64	20.16	74.59	133.32
ESPRIT (°)	12.90	138.18	15.10	74.88	136.92
DNN (°)	32.14	135.87	30.57	59.56	159.72
NTK-PRF (°)	30.50	134.41	29.84	74.62	134.13

The roots, *z* (see Section 2.3), are shown in Figure 3. Three sources are considered at 5 dB SNR. The multiple complex root pairs, *z* and 1/z, have the same phase and reciprocal magnitude, one zero is within the unit circle and the other outside. As the detailed insert in Figure 3 shows, NTK-PRF method has higher estimation accuracy when sources are far apart.



Figure 3. Three DoAs estimation in the presence of mutual coupling. Training data is at every 0.9°, and three DoAs are 30.09°, 74.95° and 134.95° respectively.

The probability of resolution is the probability that the algorithm can identify the two closely separated sources. Assuming that the true angles of two incident signals are θ_1 and θ_2 , the estimated angles are $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. If both $|\hat{\theta}_1 - \theta_1|$ and $|\hat{\theta}_2 - \theta_2|$ are less than $|\theta_1 - \theta_2|/2$, the two signals are considered successfully resolved. When the two sources are very close together, mutual coupling results in high estimation error. In Figure 4, we plot the probability of resolution versus SNR for the different methods. The two incident signals with $\theta_1 = 70.09^\circ$ and $\theta_2 = 85.05^\circ$ are relatively apart, in order to obtain the successful estimation. The statistical performance is evaluated by performing 300 Monte Carlo runs for method. The results show that the proposed NTK-PRF has relatively low resolution probability compared to baseline methods, when two sources are close.

Table 5 shows the computation time required to train the model with 200 points in the training dataset and estimate DoA. The number of iterations can be adjusted to make a tradeoff between the computation time and the performance of the network for DNN-based method. Furthermore, the trained models differ a lot between widely varying SNR. It is intractable to tune the parameters of DNN to obtain a satisfactory model. Therefore, the idea of offline training and on-line estimation may not be practical. In addition to the DoA estimation accuracy, the NTK-PRF also outperforms the baseline methods in terms of computational time. For some applications, the proposed method may be able to yield real-time DoA estimation with the help of a set of calibration sources at known locations.

Table 5. Training and test time of different methods.

-	NTK-PRF	DNN-PRF	DNN	SVR
Training time (s)	0.255	152.764	2059.968	1515.466
Test time (s)	0.252	0.316	0.295	0.199





Figure 4. Probability of resolution with different methods versus SNR. Training data is at every 0.9° , and two DoAs are 70.09° and 85.05° respectively.

Finally, we examine the effect of the density of training data. Figure 5 shows the RMSE (i.e., DoA estimation error) versus the number of training points used. As expected, the larger the amount of training data the better the performance of the NTK-PRF method. However, even a small amount of training data is enough to provide a reasonable RMSE in the DoA estimation.



Figure 5. NTK-PRF RMSE vs. training data. Test data at every 0.18° has been used.

4. Conclusions

This paper investigates an advanced machine learning based method, NTK-PRF, which is capable of DoA estimation under array element mutual coupling conditions and in multiple incoming signal scenarios. Instead of using the Toeplitz structured mutual

coupling model, realistic electromagnetic coupling, i.e., including the skewness of the radiation patterns and even active impedances, is taken into consideration. Numerical simulations are carried out to validate the effectiveness of the proposed method compared with the existing methods (baselines), both in estimation accuracy and computational time. Results show that even a small amount of training data is enough to obtain acceptable estimation accuracy. The proposed method is an attractive option for applications with high real-time requirements and can handle realistic mutual coupling as well as irregular grid arrays (although the latter is not addressed here). Further work will improve the estimation accuracy of two closely separated sources under the effect of mutual coupling.

Author Contributions: Conceptualization, Q.W.; methodology, Q.W. and X.H.; software, Q.W. and X.D.; validation, Q.W.; formal analysis, Q.W.; investigation, Q.W.; resources, Q.W.; data curation, Q.W.; writing—original draft preparation, Q.W.; technical discussions, writing, reviewing and editing, N.E.B.; visualization, Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a start-up fund from Shanghai University.

Data Availability Statement: All the simulation/experimental data and codes will be made available on request to the correspondent author's email with appropriate justification.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Johnson, D.H.; Dudgeon, D.E. Array Signal Processing: Concepts and Techniques; PTR Prentice Hall: Englewood Cliffs, NJ, USA, 1992.
- Krim, H.; Viberg, M. Two decades of array signal processing research: The parametric approach. *IEEE Signal Process. Mag.* 1996, 13, 67–94. [CrossRef]
- 3. Schmidt, R. Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. 1986, 34, 276–280. [CrossRef]
- Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* 1989, 37, 984–995. [CrossRef]
- Stoica, P.; Nehorai, A. MUSIC, maximum likelihood, and Cramer-Rao bound. *IEEE Trans. Acoust. Speech Signal Process.* 1989, 37, 720–741. [CrossRef]
- Malioutov, D.; Cetin, M.; Willsky, A.S. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* 2005, 53, 3010–3022. [CrossRef]
- Yang, Z.; Xie, L.; Zhang, C. Off-grid direction of arrival estimation using sparse Bayesian inference. *IEEE Trans. Signal Process.* 2012, 61, 38–43. [CrossRef]
- Carlin, M.; Rocca, P.; Oliveri, G.; Viani, F.; Massa, A. Directions-of-arrival estimation through Bayesian compressive sensing strategies. *IEEE Trans. Antennas Propag.* 2013, 61, 3828–3838. [CrossRef]
- 9. Friedlander, B.; Weiss, A.J. Direction finding in the presence of mutual coupling. *IEEE Trans. Antennas Propag.* **1991**, *39*, 273–284. [CrossRef]
- 10. Ye, Z.; Liu, C. On the resiliency of MUSIC direction finding against antenna sensor coupling. *IEEE Trans. Antennas Propag.* 2008, 56, 371–380. [CrossRef]
- 11. Dai, J.; Zhao, D.; Ji, X. A sparse representation method for DOA estimation with unknown mutual coupling. *IEEE Antennas Wirel. Propag. Lett.* **2012**, *11*, 1210–1213. [CrossRef]
- Wang, Q.; Dou, T.; Chen, H.; Yan, W.; Liu, W. Effective block sparse representation algorithm for DOA estimation with unknown mutual coupling. *IEEE Commun. Lett.* 2017, 21, 2622–2625. [CrossRef]
- 13. Wang, Y.; Wang, L.; Xie, J.; Trinkle, M.; Ng, B.W.-H. DOA estimation under mutual coupling of uniform linear arrays using sparse reconstruction. *IEEE Wirel. Commun.* **2019**, *8*, 1004–1007. [CrossRef]
- 14. Wang, Q.; Wang, X.; Chen, H. DOA estimation algorithm for strictly noncircular sources with unknown mutual coupling. *IEEE Commun. Lett.* **2019**, *23*, 2215–2218. [CrossRef]
- 15. Wang, B.H.; Hui, H.T.; Leong, M.S. Decoupled 2D direction of arrival estimation using compact uniform circular arrays in the presence of elevation-dependent mutual coupling. *IEEE Trans. Antennas Propag.* **2009**, *58*, 747–755. [CrossRef]
- 16. Elbir, A.M. Direction finding in the presence of direction-dependent mutual coupling. *IEEE Antennas Wirel. Propag. Lett.* **2017**, *16*, 1541–1544. [CrossRef]
- 17. Ge, Q.; Zhang, Y.; Wang, Y. A low complexity algorithm for direction of arrival estimation with direction-dependent mutual coupling. *IEEE Commun. Lett.* **2019**, *24*, 90–94. [CrossRef]
- Zheng, Z.; Yang, C. Direction-of-Arrival Estimation of Coherent Signals under Direction-Dependent Mutual Coupling. *IEEE Commun. Lett.* 2020, 25, 147–151. [CrossRef]
- 19. Ge, X.; Zi, R.; Wang, H.; Zhang, J.; Jo, M. Multi-user massive MIMO communication systems based on irregular antenna arrays. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 5287–5301. [CrossRef]

- 20. Rocca, P.; Oliveri, G.; Mailloux, R.J.; Massa, A. Unconventional phased array architectures and design methodologies—A review. *Proc. IEEE* 2016, *104*, 544–560. [CrossRef]
- 21. Friedlander, B.; Weiss, A.J. Direction finding for wide-band signals using an interpolated array. *IEEE Trans. Signal Process.* **1993**, 41, 1618–1634. [CrossRef]
- 22. Belloni, F.; Richter, A.; Koivunen, V. DoA estimation via manifold separation for arbitrary array structures. *IEEE Trans. Signal Process.* 2007, 55, 4800–4810. [CrossRef]
- 23. Pastorino, M.; Randazzo, A. A smart antenna system for direction of arrival estimation based on a support vector regression. *IEEE Trans. Antennas Propag.* 2005, 53, 2161–2168. [CrossRef]
- 24. Wu, L.-L.; Huang, Z.-T. Coherent SVR learning for wideband direction-of-arrival estimation. *IEEE Signal Process. Lett.* **2019**, *26*, 642–646. [CrossRef]
- Chen, M.; Gong, Y.; Mao, X. Deep Neural Network for Estimation of Direction of Arrival with Antenna Array. *IEEE Access* 2020, 8, 140688–140698. [CrossRef]
- Liu, Z.-M.; Zhang, C.; Philip, S.Y. Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections. *IEEE Trans. Antennas Propag.* 2018, 66, 7315–7327. [CrossRef]
- Lee, J.; Bahri, Y.; Novak, R.; Schoenholz, S.S.; Pennington, J.; Sohl-Dickstein, J. Deep Neural Networks as Gaussian Processes. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
- Jacot, A.; Gabriel, F.; Hongler, C. Neural tangent kernel: Convergence and generalization in neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 2–8 December 2018; pp. 8571–8580.
- 29. MIMObit. Available online: www.nebens.com (accessed on 27 April 2021).
- Buris, N.E. Active E-field gain: Toward a standard description of MEAs. In Proceedings of the 2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, California, CA, USA, 9–14 July 2017; pp. 2061–2062.
- 31. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
- 32. Lee, J.; Xiao, L.; Schoenholz, S.S.; Bahri, Y.; Novak, R.; Sohl-Dickstein, J.; Pennington, J. Wide neural networks of any depth evolve as linear models under gradient descent. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.
- 33. Rao, B.D.; Hari, K.S. Performance analysis of root-MUSIC. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, 37, 1939–1949. [CrossRef]
- 34. Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2011, 2, 1–27. [CrossRef]
- 35. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467.
- Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.