UNCLASSIFIED

Defense Technical Information Center Compilation Part Notice

ADP014209

TITLE: One-vs-One Multiclass Least Squares Support Vector Machines for Direction of Arrival Estimation

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Applied Computational Electromagnetics Society Journal. Volume 18, Number 2. Special Issue on Neural Network Applications in Electromagnetics.

To order the complete compilation report, use: ADA417719

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP014206 thru ADP014212

UNCLASSIFIED

One-vs-One Multiclass Least Squares Support Vector Machines for Direction of Arrival Estimation

Judd A. Rohwer[1] and Chaouki T. Abdallah[2]

[1]Sandia National Laboratories, P.O. Box 5800 MS-0986, Albuquerque, NM, 87185-0986 Email: jarohwe@sandia.gov
[2]Department of Electrical and Computer Engineering, MSC01 1100
University of New Mexico, Albuquerque, NM, 87131-001 Email: chaouki@eece.unm.edu

Abstract—**This paper presents a multiclass, multilabel implementation of Least Squares Support Vector Machines (LS-SVM) for DOA estimation in a CDMA system. For any estimation or classification system the algorithm's capabilities and performance must be evaluated. This paper includes a vast ensemble of data supporting the machine learning based DOA estimation algorithm. Accurate performance characterization of the algorithm is required to justify the results and prove that multiclass machine learning methods can be successfully applied to wireless communication problems. Thel earning algorithm presented in this paper includes steps for generating statistics on the multiclass evaluation path. The error statistics provide a confidence level of the classification accuracy.**

I. Introduction

Machine learningr esearch has largely been devoted to binary and multiclass problems relating to data mining, text categorization, and pattern recognition. Recently, machine learning techniques have been applied to various problems relating to cellular communications, notably spread spectrum receiver design, channel equalization, and adaptive beamforming with direction of arrival estimation (DOA). In our research we present a machine learning based approach for DOA estimation in a CDMA communication system [1]. The DOA estimates are used in adaptive beamforming for interference suppression, a critical component in cellular systems.I nterference suppression reduces the multiple access interference (MAI) which lowers the required transmit power. The interference suppression capability directly influences the cellular system capacity, i.e.,t he number of active mobile subscribers per cell.

Beamforming, tracking, and DOA estimation are current research topics with various technical approaches. Least mean square estimation, Kalman filtering, and neural networks [2],[3],[4], have been successfully applied to these problems. Many approaches have been developed forc alculating the DOA; three techniques based on signal subspace decomposition are ESPRIT, MUSIC, and Root-MUSIC [1].

Neural networks have been successfully applied to the problem of DOA estimation and adaptiveb eamforming in [4], [5], [6]. New machine learning techniques, such as support vector machines(SVM) and boosting [7], perform exceptionally well in multiclass problems and new optimization techniques are published regularly. These new machine learning techniques have the potential to exceed the performance of then eural network algorithms relating to communication applications.

The machine learning methods presented in this paper include subspace based estimation applied to the sample covariance matrix of the received signal. The one-vs-one multiclass LS-SVM algorithm uses both training data and received data to generate the DOA estimates. The end result is an efficient approach for estimating the DOAs in CDMA cellular architecture [1].

This paper is organized as follows. Section II presents the system models for an adaptive antenna array CDMA systems. A review of binary and multiclass machine learning methods is presented in Section III, along with background information on the LS-SVM algorithm. Section IV includes a brief review of classic DOA estimation algorithms and the elements of a machine learning based DOA estimation algorithm. Section V presents a one-vs-onem ulticlass LS-SVM algorithm for DOA estimation and simulation results are presented in Section VI. Section VII includes a comparison between standard DOA estimation algorithmsa nd our machine learning based algorithm.

II. System Models

This section includes an overview of system models for the received signal and adaptive antenna arrays designs. All notation is described below and is consistently used throughout the paper.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for theU nited States Department of Energy under Contract DE-AC04-94AL85000.

A. Received Signal at Antenna Array output

The baseband signal, $\mathbf{r}_A(t)$, from the antenna array is

$$\begin{aligned}
\mathbf{r}_A(t) &= \mathbf{A}\mathbf{s}(t) + \mathbf{n}_r(t). && (1)\\
\mathbf{A} &= \left[\ \mathbf{a}(\theta_1)\ \ \mathbf{a}(\theta_2)\ \ \ldots\ \ \mathbf{a}(\theta_L)\ \right] && (2)\\
\mathbf{a}(\theta_l) &= \left[\ 1\ \ e^{-jk_l}\ \ e^{-j2k_l}\ \ \ldots\ \ e^{-j(D-1)k_l}\ \right]^T && (3)\\
\mathbf{s}(t) &= \left[\ s_1(n)\ \ s_2(n)\ \ \ldots\ \ s_L(n)\ \right]^T && (4)\\
s_l(t) &= \sqrt{p_{t_i}(t)}q^l b_i(t),\ \text{for path } l, && (5)
\end{aligned}$$

where $\mathbf{r}_A(t)$ is the received signal of mobile i, $\mathbf{A}$ is a $D \times L$ array steering vector for D antenna elements and L transmission paths, $\mathbf{s}(t)$ is the $L \times 1$ received basebands ignal at the output of the matched filter, $\mathbf{a}(\theta_l) = \left[\ 1\ \ e^{-jk_l}\ \ \ldots\ \ e^{-j(D-1)k_l}\ \right]^T$ is the $D \times 1$ steering vector, $k_l = \frac{\nu w_c}{c}\sin\theta_l$, v is the spacing between antenna elements, w_c is the carrier frequency, c is the velocity of propagation, $\theta_l$is the direction of arrival of the l signal, $p_{t_i}(t)$ is the transmit signal power from mobile i, q_i^l is the attenuation due to shadowing from path l, $b_i(t)$ is the data stream of mobile i, and $\mathbf{n}_r(t)$ is the additive noise vector.

To ease the complexity of the notation the terms relative to the multiple paths are combined as

$$\mathbf{z}_i = \sum_{l=1}^{L} \mathbf{a}(\theta_l)\, q_i^l. \tag{6}$$

In [8] $\mathbf{z}_i$ is defined as the spatial signature of the antenna array to the ith mobile.

III. Support Vector Machines - Background

A major machine learning application, pattern classification, observes input data and applies classification rules to generate a binary or multiclass labels. In the binary case, a classification function is estimated using input/output training pairs,$(\mathbf{x}_i,y_i)$ $i = 1\ldots n$, with unknown probability distribution, $P(\mathbf{x}, y)$,

$$\begin{aligned}
(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n) &\in \mathbb{R}^N \times Y, && (7)\\
y_i &= \{-1,+1\}. && (8)
\end{aligned}$$

The estimated classification function maps the input to a binary output, $f : \mathbb{R}^N \rightarrow \{-1,+1\}$. The system is first trained with the given input/output data pairs then the test data, taken from the same probability distribution $P(\mathbf{x}, y)$, is applied to the classification function. For the multiclass case $Y \in \mathbb{R}^G$ where Y is a finite set of real numbers and G is the size of the multiclass label set.I n multiclass classification the objective ist o estimate the function which maps the input data to a finite set of output labels $f : \mathbb{R}^N \rightarrow S\left(\mathbb{R}^N\right) \in \mathbb{R}^G$

Support Vector Machines (SVMs) were originally designed for the binary classification problem. Much like all machine learning algorithmsS VMs find a classification function that separates data classes, with the largest margin, using a hyperplane . The data points near the optimal hyperplane are the "support vectors". SVMs are a nonparametric machine learning algorithm with the capability of controlling the capacity through the support vectors.

A. Kernel Functions

The kernel based SVM maps the input space into a higher dimensional feature, F, space via a nonlinearm apping

$$\begin{aligned}
\Gamma &: \mathbb{R}^N \rightarrow F && (9)\\
\mathbf{x} &\mapsto \Gamma(\mathbf{x}). && (10)
\end{aligned}$$

The data does not have the same dimensionality as the feature space since the mapping process is to a non-unique generalized surface [9]. The dimension of the feature space is not as important as the complexity of the classification functions. For example, in the input space, separating the input/output pairsm ay require a nonlinear separating function, but in a higher dimension feature space the input/output pairs may be separated with a linear hyperplane. The nonlinear mapping function $\mathbf{\Gamma}(\mathbf{x}_i)$ is related to kernel, $\mathbb{k}(\mathbf{x}, \mathbf{x}_i)$ by

$$\Gamma(\mathbf{x}) \cdot \mathbf{\Gamma}(\mathbf{x}_i) = \mathbb{k}(\mathbf{x}, \mathbf{x}_i). \tag{11}$$

Four popular kernel functions are the linear kernel, polynomial kernel, radial basis function (RBF), and multilayer perceptrons (MLP).

$$\begin{aligned}
\text{linear, } \mathbb{k}(\mathbf{x}, \mathbf{x}_i) &= \mathbf{x} \cdot \mathbf{x}_i && (12)\\
\text{polynomial, } \mathbb{k}(\mathbf{x}, \mathbf{x}_i) &= ((\mathbf{x} \cdot \mathbf{x}_i) + \theta)^{\mathbf{d}} && (13)\\
\text{RBF, } \mathbb{k}(\mathbf{x}, \mathbf{x}_i) &= \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_i\|^2}{\sigma^2}\right) && (14)\\
\text{MLP, } \mathbb{k}(\mathbf{x}, \mathbf{x}_i) &= \tanh(\kappa(\mathbf{x} \cdot \mathbf{x}_i) + \theta) && (15)
\end{aligned}$$

The performance of each kernel function varies with the characteristics of the input data. Refer to [10] for more information on feature spaces and kernel methods.

B. Binary Classification

In binary classification systems the machine learning algorithm generate the output labels with a hyperplane separation where $y_i \in [-1, 1]$ represents the classification "label" of the input vector $\mathbf{x}$. The input sequence and a set of training labels are represented as $\{\mathbf{x}_i,y_i\}_{i=1}^n$, $y_i = \{-1,+1\}$. If the two classes are linearly separable in the input space then the hyperplane is defined as $\mathbf{w}^T\mathbf{x}+b = 0$, $\mathbf{w}$ is a weight vector perpendicular to the separating hyperplane, b isa biast hat shifts the hyperplane parallel to itself. If the input space is projected into a higher dimensional feature spacet hen the hyperplane becomes $\mathbf{w}^T\mathbf{\Gamma}(\mathbf{x}) + b = 0$.

The SVM algorithm is based on the hyperplane definition [11],

$$y_i\left[\mathbf{w}^T\mathbf{\Gamma}(\mathbf{x}_i) + b\right] \geq 1, i = 1,\ldots,N. \tag{16}$$

Given the training sets in (7) the binary support vector machine classifieri s defined as

$$y(\mathbf{x}) = sign\left[\sum_{i=1}^{N} \alpha_i y_i \mathbb{k}(\mathbf{x}, \mathbf{x}_i) + b\right]. \quad (17)$$

The non-zero $\alpha_i's$ are "support values" and the corresponding data points, $\mathbf{x}_i$, are the "support vectors". Quadratic programming is one method ofs olving for the $\alpha_i's$ and b in the standard SVM algorithm.

C. Multiclass Classification

For the multiclass problem the machine learning algorithm produces estimates with multiple hyperplane separations. The set of input vectors and training labels is defined as $\{\mathbf{x}_i, y_i^c\}_{i=1,c=1}^{i=n,c=C}$, $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \{1, \ldots, G\}$, n is the index of the training pattern and C is the number of classes. There exist many SVM approaches to multiclass classification problem. Two primary multiclass techniques are one-vs-one and one-vs-rest. One-vs-onea pplies SVMs to selected pairs of classes. For C distinct classes there are $\frac{C(C-1)}{2}$ hyperplanes that separate the classes. The one-vs-rest SVM technique generates C hyperplanes that separate each distinct class from the ensemble of the rest. In this paper we only consider the one-vs-one multiclass SVM.

Platt, et.al., [12] introduced the decision directed acyclic graph (DDAG) and a Vapnik-Chervonenkis (VC) analysis of the margins. The DDAG technique isb ased on $\frac{C(C-1)}{2}$ classifiers fora C class problem, one node for each pair of classes. In [12] it is proved that maximizing the margins at each node of the DDAG will minimize the generalization error. The performance benefit of the DDAG architecture is realized when the ith classifier is selected at the ith/j^{th} node and the jth class is eliminated. Refer to Figure 1 for a diagram of a fourc lass DDAG.

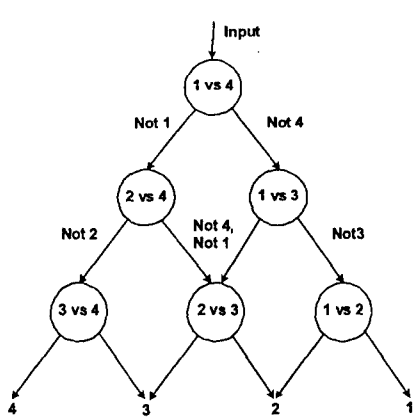


Fig. 1. Fourc lass DDAG for one-vs-one multiclass LS-SVM based DOA estimation.

D. Least Squares SVM

Suykens, et.al., [13] introduced the LS-SVM which is based on the SVM classifier, refer to equation (17). The LS-SVM classifier is generated from the optimization problem:

$$\min_{\mathbf{w},b,\phi} \mathcal{L}_{LS}(\mathbf{w},\phi) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2}\gamma \sum_{i=1}^{n} \phi_i^2, \quad (18)$$

γ and ϕ_i are the regularization and error variables, respectively. The minimization in (18)includes the constraints

$$y_i\left[\mathbf{w}^T \mathbf{\Gamma}(\mathbf{x}_i) + b\right] \geq 1 - \phi_i, \quad i = 1, \ldots, n, \quad (19)$$

The LS-SVM includes one universal parameter, γ, that regulates the complexity of the machine learning model.T his parameter is applied to the data in the feature space, the output of the kernel function. A small value of γ minimizes the model complexity, while a large value of γ promotes exact fitting to the training points. The error variable ϕ_i allows misclassifications foro verlapping distributions [14].

The Lagrangian ofe quation (18) is defined as

$$\begin{aligned} \mathcal{Z}_{LS}(\mathbf{w},b,\phi,\alpha) &= \mathcal{L}_{LS}(\mathbf{w},b,\phi) - \\ &\quad \sum_{i=1}^{n} \alpha_i \left\{y_i\left[\mathbf{w}^T \mathbf{\Gamma}(\mathbf{x}_i) + b\right] - 1 + \phi_i\right\} \end{aligned} \quad (20)$$

where α_i are Lagrangian multipliers that can either be positive or negative. The conditions of optimality are

$$\frac{d\mathcal{Z}_{LS}}{d\mathbf{w}} = 0, \ \mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{\Gamma}(\mathbf{x}_i) \quad (21)$$

$$\frac{d\mathcal{Z}_{LS}}{db} = 0, \ \sum_{i=1}^{n} \alpha_i y_i = 0 \quad (22)$$

$$\frac{d\mathcal{Z}_{LS}}{d\phi} = 0, \ \alpha_i = \gamma \phi_i \quad (23)$$

$$\frac{d\mathcal{Z}_{LS}}{d\alpha_i} = 0, \ y_i\left[\mathbf{w}^T \mathbf{\Gamma}(\mathbf{x}_i) + b\right] - 1 + \phi_i = 0 \quad (24)$$

A linear system can be constructed from equations (21) – (24) [13],

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ Z & Y & I & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ b \\ \phi \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{1} \end{bmatrix} \quad (25)$$

$$Z = \left[\mathbf{\Gamma}(\mathbf{x}_1)^T y_1, \ldots, \mathbf{\Gamma}(\mathbf{x}_n)^T y_n\right] \quad (26)$$

$$Y = [y_1, \ldots, y_n], \ \vec{1} = [1, \ldots, 1] \quad (27)$$

$$\phi = [\phi_1, \ldots, \phi_n], \ \alpha = [\alpha_1, \ldots, \alpha_n] \quad (28)$$

By eliminating weight vector $\mathbf{w}$ and the error variable ϕ, the linear system is reduced to:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \quad (29)$$

In the linear systems defined in (25)−(29) the support values α_i are proportional to the errors at thed ata points. In the standard SVM case many of these support values are zero, but most of the least squares support values are non-zero. In [13] a conjugate gradient method isp roposed for finding b and α, which are required for the SVM classifier in equation (17).

IV. Algorithms for DOA Estimation

Two primary, classic methods for subspace based DOA estimation exist in literature, Multiple Signal Classification (MUSIC) [15] and Estimation ofS ignal Parameters Via Rotational Invariance Techniques (ESPRIT) [16]. The MUSIC algorithm is based on the noise subspace and ESPRIT is based on the signal subspace.

Many computational techniques exist for working through limitationso f DOA estimation techniques, but currently no techniques exist for as ystem level approach to accurately estimating the DOAs at the base station. A number of limitations relating to popular DOA estimation techniques are: 1) the signals ubspace dimension is not known, many papers assume that it is.T he differences between the covariance matrix and the sample covariance matrix add to the uncertainty, 2) searching all possible angles to determine the maximum response of the MUSIC algorithm, 3) evaluating the Root-MUSIC polynomial on the unit circle, 4) multiple eigen decompositions for ESPRIT, 5) computational complexity for maximum likelihood method. The capabilities, in terms of resolution and computational requirements, of these standard DOA estimation algorithms serve as the benchmark for the machinel earning based DOA estimation. Refer to Section VII for a comparison between standard DOA estimation algorithms and the one-vs-one multiclass LS-SVM DOA estimation algorithm.

A. Machine Learning for DOA Estimation

To estimate the antenna array response, $\mathbf{z}_j = \sum_{l=1}^{L} \mathbf{a}(\theta_l) q_j^l$, we must know $\mathbf{a}(\theta_l)$ and q_i^l. The continuous pilot signal, includedi n cdma2000, can be used in estimating q_i^l. This must be done for each resolvable path, i.e., $q_i = \left[\, q_i^1, \; q_i^2, \; \ldots \; , q_i^L \,\right]$. Estimating $\mathbf{A}(\theta) = \left[\, \mathbf{a}(\theta_1), \; \mathbf{a}(\theta_2), \; \ldots \; , \mathbf{a}(\theta_L) \,\right]$ requires information on the DOA.

The process of DOA estimation is to monitor the outputs of D antenna elements and predict the angle of arrival of L signals, $L<D$. The output matrix from the antenna elements is

$$\begin{aligned} \mathbf{A} &= \left[\, \mathbf{a}(\theta_1) \;\; \mathbf{a}(\theta_2) \;\; \ldots \;\; \mathbf{a}(\theta_L) \,\right] \qquad (30) \\ \mathbf{a}(\theta_l) &= \left[\, 1 \;\; e^{-jk_l} \;\; e^{-j2k_l} \;\; \ldots \;\; e^{-j(D-1)k_l} \,\right]^T, \end{aligned}$$

and the vector of incident signals is $\boldsymbol{\theta}_r = \left[\, \theta_1, \; \theta_2, \; \ldots \; , \theta_L \,\right]$. With a training process, the learning algorithms generate DOA estimates, $\widehat{\boldsymbol{\theta}}_r = \left[\, \widehat{\theta}_1, \; \widehat{\theta}_2, \; \ldots \; , \widehat{\theta}_L \,\right]$, based on the responses from the antenna elements, $\mathbf{a}(\theta_l)$.

For thep roposed machinel earning technique there is a trade-off between the accuracy of the DOA estimation and antennaa rray beamwidth. An increase in DOA estimation accuracy translates into a smaller beamwidth and a reduction in MAI. Therefore the accuracy in DOA estimation directly influences the minimum required power transmitted by the mobile. There should be a balance between computing effort and reduction in MAI.

V. LS-SVM DDAG based DOA Estimation Algorithm

In this paper we propose a multiclass SVM algorithm trained with projection vectors generated from the signal subspace eigenvectors and the sample covariance matrix.T he output labels from the SVM system are the DOA estimates.

The one-vs-one multiclass LS-SVM DDAG technique for DOA estimation is trained for C DOA classes. The DDAG tree is initialized with $\frac{C(C-1)}{2}$ nodes. Therefore $\frac{C(C-1)}{2}$ one-vs-one LS-SVMs are trained to generated the hyperplanes with maximum margin. For each class thet raining vectors, $\mathbf{x}_n$, areg enerated from thee igenvectors spanning the signal subspace. The number of classes is dependent upon on the antenna sectoring and required resolution. For a CDMA system the desired interference suppression dictates the fixed beamwidth. CDMA offers this flexibility since the all mobiles use the same carrier frequency. For FDMA systems a narrow beamwidth is desired, since frequency reuse determines the capacity of a cellular system.

Thes ignal subspace eigenvectors of the received signal covariance matrix are required for accurate DOA estimation. For a CDMA system with adaptive antenna arrays the covariance matrix of the received signal is

$$\mathbf{R}_{rr} = \mathbb{E}\left[\mathbf{r}_A \mathbf{r}_A^H\right]. \qquad (31)$$

In our machine learning based DOA estimation algorithm the principal eigenvectors must be calculated. Eigen decomposition (ED) is the standard computational approach for calculating the eigenvalues and eigenvectors of a the covariance matrix. ED is a computationally intense technique, faster algorithmss uch asP ASTd [17] have been developed forr eal-time processing applications.

For the LS-SVM based approach to DOA estimation the output of the receiver is used to calculate the sample covariance matrix $\widehat{\mathbf{R}}_{rr}$ of the input data signal $\mathbf{r}_A(k)$,

$$\widehat{\mathbf{R}}_{rr} = \frac{1}{M} \sum_{k=K-M+1}^{K} \mathbf{r}_A(k) \mathbf{r}_A^H(k). \qquad (32)$$

The dimension of the observation matrix is $D \times M$, M is ideal sample size (window length), and the dimension of the

TABLE I
PROJECTION COEFFICIENTS FOR MACHINE LEARNING BASED POWER CONTROL

	Projection Coefficients		
	25°	30°	40°
1	0.17+i·0.86	-0.20-i·0.54	0.00+i·0.86
2	0.66+i·0.05	-0.82+i·0.14	0.73-i·0.55
3	0.04-i·0.73	0.28+i·0.96	-1.01-i·0.58
4	-1.08-i·0.50	1.04-i·0.37	0.06+i·1.05
5	-0.60+i·0.92	-0.56-i·1.01	0.72-i·0.61
6	0.60+i·0.74	-0.87+i·0.64	-0.92-i·0.51
7	0.72-i·0.56	0.63+i·0.62	-0.03+i·0.76
8	-0.52-i·0.78	0.51-i·0.44	0.45-i·0.42

sample covariance matrix is $D \times D$. The principal eigenvectors, $\mathbf{v}_1, \ldots, \mathbf{v}_D$, are calculated via eigen decomposition (ED) or subspacet racking techniques. Each eigenvector is used to calculate a covariance matrix, $\widehat{\mathbf{R}}_{vv_1}, \ldots, \widehat{\mathbf{R}}_{vv_D}$.

The algorithm requires only the set of estimated eigenvectorsf rom the sample covariance matrix, which are used to generate projection coefficients for the classification process. The projection vectors are generated from the projection of $\widehat{\mathbf{R}}_{vv_d}$, $1 \leq d \leq D$, onto the primary eigenvector of the signal subspace. In the training phaset he hyperplanes at each DDAG node are constructed with thesep rojection vectors. In the testing phase $\widehat{\mathbf{R}}_{vv_d}$ is generated from the received signal $\mathbf{r}_A(k)$ and the principal eigenvectors. Then the projection coefficients for the ith/j^{th} node of the DDAG are computed with dot products of $\widehat{\mathbf{R}}_{vv_d}$ and the ith/j^{th} training eigenvectors. Thisn ew set of projection vectorsi s testing with the ith/j^{th} hyperplane generated during the training phase. The DOA labels are then assigned based on the DDAG evaluation path. A similar projection coefficient technique has been successfully applied to a multiclass SVM facialr ecognition problem presented in [18].T able I includes threes ets of projection vectors, each set corresponds to a different DOA. From a review of the data it is evident that the classes are not linearly separable. The data must be projected to a higher dimension feature space and tested against the separating hyperplane.

The following algorithm for the one-vs-one multiclass LS-SVM implementation for DOA estimation includesp reprocessing, training, and testing steps.S pecifically, the algorithm requires two sets of projection vectors for each DDAG node. This allows for automatic MSE calculations at each step of the DDAG evaluation path, thus providing a unique method fore rror control and validation.

- Preprocessing for SVM Training
  1) Generate the $D \times N$ training signal vectors for the C LS-SVM classes, D is the number of antenna elements, N is the number of samples.
  2) Generate the C sample covariance matrices, $\mathbf{U}$,with M samples from the $D \times N$ data vector.
  3) Calculate the signal eigenvector, $\mathbf{S}$, from each of the C sample covariance matrices.
  4) Calculate the $D \times 1$ projection vectors, $\mathbf{U} \cdot \mathbf{S}$, for each of the C classes. The ensemble of projection vectors consists of $\frac{N}{M}$ samples.
  5) Store the projection vectors for the training phase and the eigenvectors for the testing phase.
- LS-SVM Training
  1) With the C projection vectors train the $\frac{C(C-1)}{2}$ nodes with the one-vs-one LS-SVM algorithm.
  2) Store the LS-SVM variables, α_i and b from equation (17), which define the hyperplane separation fore ach DDAG node.
- Preprocessing for SVM Testing
  1) Acquire $D \times N$ input signal from antenna array, this signal has unknown DOAs.
  2) Generate the sample covariance matrix with M samples from the $D \times N$ data vector.
  3) Calculate the eigenvectors for the signal subspace and the noise subspace.
  4) Generate the covariance matricesf or each eigenvector.
- LS-SVM Testing for the i/j DDAG Node
  1) Calculate **TWO** $D \times 1$ projection vectors with the desired eigenvector covariance matrix and the ith and j^{th} eigenvectors from the training phase.
  2) Test both projection vectorsa gainst the LS-SVM hyperplane for the i/j node. This requires two separate LS-SVM testing cycles, one with the projection vector from the ith eigenvector and one with the projection vector from the jth eigenvector.
  3) Calculate the mean value of the two LS-SVM output vectors (labels). Select the mean value that is closest to a decision boundary, 0 or 1. Compare this value to the label definition at the node, then select the proper label.
  4) Repeat process for the next DDAG node in the evaluation path or declare the final DOA label.
- Error Control
  1) Review the MSE calculations for the DDAG evaluation path.
  2) Apply error control and validation measures to classify the label as either an accurate **DOA** estimate or as **NOISE**.

VI. SIMULATION RESULTS

Two simulation plots are included below. Each simulation consists of af our class LS-SVM DDAG system. Figure 2 shows results for a ten degree range per class. Figure 3 shows results for a one degree range per class.

The antenna array includes eight elements, therefore the training and test signals were 8×1 vectors. The training

and test signals are the complex outputs from the antenna array. The received complex signal ism odeled with a zero mean normal distribution with unit variance; the additive noise includes a zero mean distribution with a 0.2 variance. This combination of signal and noise power translates into a $7dB\ SIR$.

The system training consists of six DDAG nodes for the four DOA classes. Both the training and test signals consisted of 1500 samples and the window length of the sample covariance matrix was set to five. Therefore the training and test sets were composed of 300 samples of each 8×1 projection vector.

To completely test the LS-SVM DDAG system's capabilities the simulation were automated to test a wide range of DOAs. The DOA test set consisting of signals ranging from three degrees before the first DOA class to three degrees after the last DOA class. Thus there were forty-six test signals for Figure 2 and fourteen test signals for Figure 3. As can been seen from the two plots the LS-SVM DDAG DOA estimation algorithm is extremely accurate. No misclassifications were logged. Testing shows that the LS-SVM DDAG system accurately classifies the DOAs fora ny desired number of classes and DOA separations from one degree to twenty degrees.

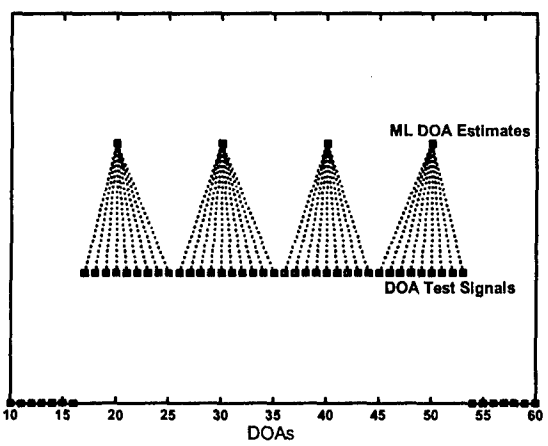


Fig. 2. LS-SVM for DOA estimation, four classes with ten degree separation between each.

A. Decision Grids

The decision grid (DG) technique was developed to track the DDAG evaluation path and generate statistics to characterize the confidence level of the DOA classifications. The theoretical DG (T-DG) is a technique we developed to quantify errors and add insighti nto the robustness of the LS-SVM DDAG architecture. The T-DG is a deterministic $2D$ grid for DDAGs with a relatively small number of classes and small DOA range between classes.T he elements of the T-DG

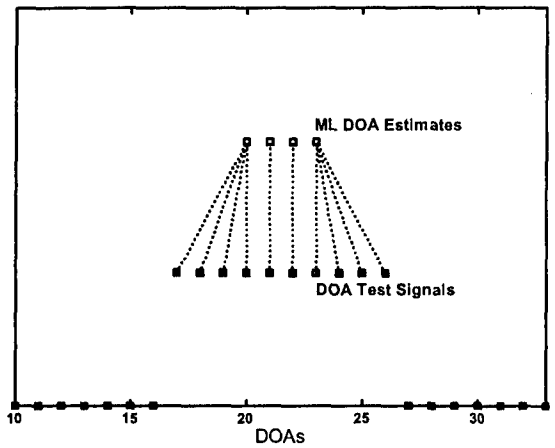


Fig. 3. LS-SVM for DOA estimation, four classes with one degree separation between each.

represent the deterministic values of the two LS-SVM labels at each DDAG level, thed eterministic values are referred to as" theoretical decision statistics". Designing T-DGs for DDAGs with threet o five classes and DOA ranges up to five degrees between classes is straight forward. The T-DGs are not deterministic for large DOA ranges, i.e. for a DOA range of ten degrees between classes empirical results show that the DDAG evaluation path isu npredictable. The large DOA ranges lead to uncertainty in the evaluation path, even though the test DOA is classified correctly.

Empirical decision grids (E-DG) are automatically generated in the LS-SVM DDAG DOA estimation algorithm. The E-DGs tabulate the mean of the LS-SVM output label vectors at each DDAG node and level, the mean values are referred to as "decision statistics". The unique design of this algorithm includes testing the input data against two hyperplanes at the ith/j^{th} node. With this approach the two output vectors at each node are compared to one another. In a noise-free environment,w ith perfect classification, the two label vectors would be binary opposites, i.e. one label vector would be all $0's$ and the other label vector would be all $1's$. This technique enables computation of theoretical mean square errors and empirical mean square errors, refer to Section VI-B.

Table II includes a standard T-DG and Tables III and IV include E-DGsf or a three classD DAG with a two degree DOA range per class. The two levels of a three class DDAG are equivalent to the first two levels of a four class DDAG, refer to Figure 1. Table II includes the possible evaluation paths of thist hree class DDAG. The nodesf or each DOA evaluation path are included for the first and second DDAG level. For example, DOA 1 has an evaluation path of Node 1 vs 3 at Level 1 and Node 1 vs 2 at Level 2. In Table III E-DG presents the decision statistics for a signal subspace eigenvector; in Table IV the second E-DG presents the

TABLE II

THEORETIC DECISION GRID FOR A DDAG SYSTEM WITH 3 CLASSES AND A 2 DEGREE DOA RANGE.

	DOAs				
	Class1		Class 2		Class 3
T-DG, Level 1	1	2	3	4	5
Node	1vs3	1vs3	1vs3	1vs3	1vs3
Label 0	0	0	0.5	1	1
Label 1	1	1	0.5	0	0
T-DG, Level 2					
Node	1vs2	1vs2	1vs2\2vs3	2vs3	2vs3
Label 0	0	0.5	0\1	0.5	1
Label 1	1	0.5	1\0	0.5	0

TABLE III

EMPIRICAL DECISION GRID FOR A SIGNAL EIGENVECTOR

	DOAs				
Signal Data	Class 1		Class 2		Class 3
E-DG, Level 1	1	2	3	4	5
Node	1vs3	1vs3	1vs3	1vs3	1vs3
Label 0	0	0	0.032	0.952	1
Label 1	1	1	0.576	0	0
E-DG, Level 2					
Node	1vs2	1vs2	1vs2\2vs3	2vs3	2vs3
Label 0	0	0.176	1	0.808	1
Label 1	1	0.816	0	0.496	0

TABLE IV

EMPIRICAL DECISION GRID FOR A NOISE EIGENVECTOR

	DOAs				
Noise Data	Class 1		Class 2		Class 3
E-DG, Level 1	1	2	3	4	5
Node	1vs3	1vs3	1vs3	1vs3	1vs3
Label 0	0.328	0.376	0.304	0.352	0.384
Label 1	0.752	0.744	0.712	0.768	0.776
E-DG, Level 2					
Node	1vs2	1vs2	1vs2\2vs3	2vs3	2vs3
Label 0	0.232	0.256	0.144	0.136	0.184
Label 1	0.896	0.904	0.952	0.944	0.944

decision statistics for a noise subspace eigenvector.

B. Theoretical and Empirical MSEs

The difficulty in tracking the performance of the LS-SVM DDAG DOA estimation algorithm is duet o the numerous DDAG evaluation paths. For many DDAGs thee valuation paths can be determined based on the input data and the class definitions. How can decision statistics be applied to performance characterization?

The two primary performance measures for the LS-SVM DDAG are the theoretical MSE (T-MSE) and the empirical MSE (E-MSE). Both MSE performance measures are based on MSE calculations with T-DGs and E-DGs. The T-MSE is a MSE calculation between the corresponding elements of the T-DG and the E-DG. This is a measure of the algorithm's empirical decision statistics in relation to the" theoretical" decision statistics. For example, the T-MSE for a 3 class DDAG is calculated with the T-DG and E-DG presented in Tables IIa nd III. The T-MSE for Class 2 is calculated as

Level 1		Level 2	
Label 0	Label 1	Label 0	Label 1
$(0.5-0.032)^2$	$(0.5-0.576)^2$	$(1-1)^2$	$(0-0)^2$

Unlike the T-MSE, the E-MSE is at echnique that allows for real-time error tracking with only the empirical decision statistics. The E-MSE uses onlyt he E-DGs and the differences between the two LS-SVM decision statistics at each node in the evaluation path. Thisi sa measure of the empirical classification accuracy achieved at each DDAG node. The E-MSE for a3 class DDAG is calculated with only the E-DG presented in Table III. The MSE for Class 2, Level 1 is $(|0.032-0.576|-1)^2 = 0.208$ and the MSE for Class 2, Level 2 is $(|1-0|-1)^2 = 0$.

C. Misclassifications vs. Gross Errors

Two secondary performance measures for the LS-SVM DDAG are misclassifications and grosse rrors. These measures are used for performance characterization of the multi-class LS-SVM DDAG DOA estimation algorithm and for tracking variations in performance for various algorithm parameters. Misclassifications and gross errors can not be used in real time implementation because knowledge of the test DOAs is required.

Misclassifications measure "small shifts" in DOA classifications. If a DOA is located near a border between labelst he machine learning processc ould classify the data to an adjacent label, not the closest label. Therefore, a misclassification is a shift related error where a signal is detected, but classified to a spatially adjacent label. This type of error still gives an indication of the received DOA. The region of misclassifications is defined as $\frac{1}{2}$ of the DOA range applied to both sides of a DOA class.

Gross errors measure significant errors in DOA classifications.I f a DOA is classified into a specific class,b ut spatially located at least one entire class away, then the error is due to a breakdown in the machine learning process. This type of error assigns false/misleading information to a received DOA. The region of gross errors is defined as the magnitude of the DOA range applied to both sides of the DOA class.

Figure 4 displays the DOA regions for correct classifications, misclassifications and gross errors. This specific example is for a DDAG class centered at 0° with a 5° DOA range, i.e., any DOA in the range $[-2, 2]$ is correctly classified to the 0° class. The region enclosed by the dashed brackets includes all DOAs that are correctly classified at the 0° class. If any DOAso utside the dashed bracketsb ut inside the solid brackets are assigned the 0° class, then that DOA would be a misclassification. If any DOAs outside the solid bracketsa re assigned to the 0° class, then that DOA

would be a gross error. The misclassification region, for a DOA classified at 0°, is $DOA \in [-4,-3],[3,4]$. The gross error region, for a DOA classified at 0°, is $DOA \notin [-4,4]$.

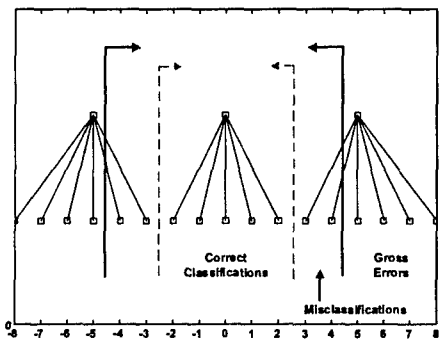


Fig. 4. Diagram of regions definingD OA misclassifications and gross errors.

D. Kernel Parameters

Simulation results show that kernel selection has the greatest effect, out of all tunable variables, in thec lassification process. The four kernels discussed in Section III-A are tested with the LS-SVM DDAG DOA estimation algorithm. The performances of each kernel function and the associated parameters are characterized with in terms of MSE, misclassifications, and gross errors. In addition, the LS-SVM regularization parameter, γ, is varied to show the influence of the LS-SVM complexity.

1) Polynomial Kernel: The polynomial kernel provides the best results, in relation to the RBF, MLP, and linear kernels. Figure 5 displays the T-MSE in terms of the polynomial degree, $\mathbf{d}$, and constant, $\boldsymbol{\theta}$. The simulation is based on af our class DDAG with a 5° DOA rangea nd a fixed LS-SVM variable, $\gamma = 2$. The results show that the degree of the polynomial kernel affects the DOA estimation; the best values are $\mathbf{d} = 2$ and $\mathbf{d} = 4$. For $\mathbf{d} = 1$ the polynomial kerneli s equivalent to the linear kernel. The MSE is constant for $1 \leq \gamma \leq 6$, and the polynomial constant, $\boldsymbol{\theta}$, does not influence the performance. The rate of misclassifications is 1.2% with zero gross errors. The degree of the polynomial is the only factor affecting the computational time for system training.

2) Radial Basis Function Kernel: The performance of the RBF kernel is characterized in terms of the LS-SVM regularization variable, γ, and the smoothing parameter, σ^2. The simulation is based on a four class DDAG with a 5° DOA range.T he results show that the MSE is constant for $\gamma \geq 1.5$, and $\sigma^2 \geq 0.5$. The rate of misclassifications is 0.4% with zero gross errors. The training time increases with the value of γ and for small values of σ^2. The performance of

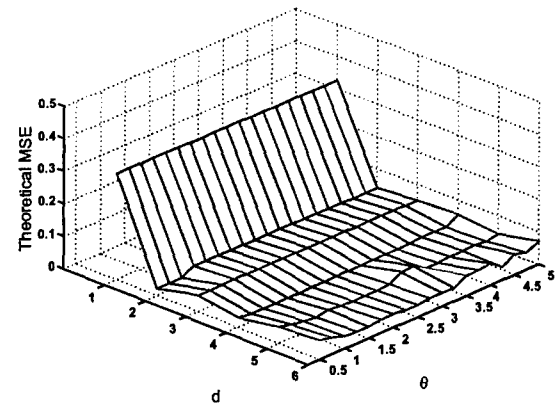


Fig. 5. Theoretical MSE for a the polynomial kernel, the DOA range is 5° and spans the DDAG classes at 30°, 35°, 40°, 45°, the LS-SVM parameter, γ, is set at 2.

the RBF kernel matches the performance of the polynomial kernel for DOAs in the range of 15° to 60°. The performance of the polynomial kernel exceeds that of the RBF kernel for DOAs < 15° and > 60°.

3) Multilayer Perceptron Kernel: Results show that the MLP kernel is ineffective in maintaining a low MSE for the range of parameters tested. The rate of misclassifications is 42.5% and the rate of gross errors is 17.2%. Overall the performance of the MLP kernel is inferior to the polynomial and RBF kernels.

4) Linear Kernel: The linear kernel ise quivalent to the polynomial kernel with $\mathbf{d} = \mathbf{1}$. Large MSE values show that the linear kernel is not effective in the LS-SVM DOA estimation algorithm. The average T-MSE is 27.8% and the average E-MSE is 61.1%.

E. Training and Test Vectors

The design of training sequences is an important factor in machine learning applications. For adaptive antennaa rrays the training sequences represent the array outputs for the C DOA classes. Three specific elements of the training sequences are noise variance, training vector length, and length of thes amplec ovariance window. The requirement is to design training sequences that minimize both the training error and generalization error. Empirical analysis of the multiclass LS-SVM based DOA estimation algorithm shows that training error is effectively zero; the hyperplane separation of the data in the feature space is welld efined and separable. In this papert he generalization error is expressed in terms of MSEs, misclassifications and gross errors.

The primary method for training LS-SVM DDAG systems for DOA estimation is based on synthetic training vectors generated with known noise power and preselected vector lengths. In practice, the training vectors would be stored in the memory of the receiver that employs the DOA estimation

algorithm. This approach allows for offline training of the binary LS-SVM algorithms.

Simulation results show that the LS-SVM DOA estimation algorithm is robust, in terms of MSE, when analyzed for a range ofS IRs in the training vectors and the test signals. In general, the noise power of the training vectors doesn't have a dramatic effect on the generalization error. Simulations were conducted with training vectors that included SIRs in the range of 20 dB to 7 dB. Review of them isclassification and gross error statistics show that training vectors with noise variances of 0.04 and 0.12, which correspond to SIRs of 13 dB and 10 dB, provide the best performance.

1) Length of Training and Testing Vectors: Figure 6 includes two plots of average theoretical MSE versus training vector length. The data is specific to a four class LS-SVM DDAG system with a four degree polynomial kernel. The two plots show that the window length of the sample covariance matrix does not impact the performance.L ikewise there is no correlation between the length of the training vector and the MSE. The results in Figure 6 are based on test vectors with size equivalent to the training vectors. Figure 7 is a 3D plot of the theoretical MSE as a function of vector dimensions; the dimensions of the training vectors and input data vectors. The length of the input data vector ranges from 0.5 to 2 times the length of the training vectors. The data shows that range of input data vectors has no effect on the MSE statistics.

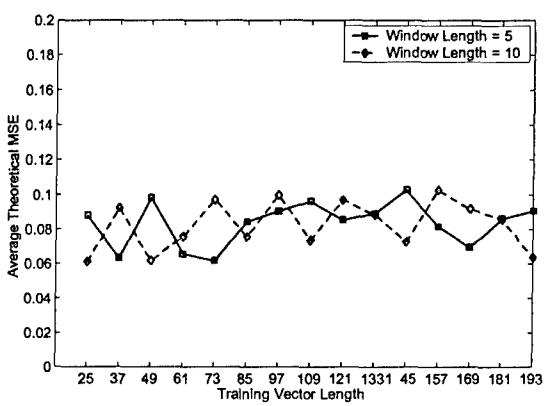


Fig. 6. Average theoretical MSE as a function of training vector length. Two data plots are included; one plot is for a sample covariance matrix with a five sample window, one plot is for a sample covariance matrix with a ten sample window.

Table V shows the processing times, in seconds, required for training a four class LS-SVM DDAG system with a four degreep olynomial kernel. and testing the input data. The results Data is included for training and test vectors that range from 25 samples to 200 samples. The simulations were conducted with a Pentium 4 running at 2.5 GHz. The processing times are relative to the computer system and the level of optimization applied to the programming, but serve

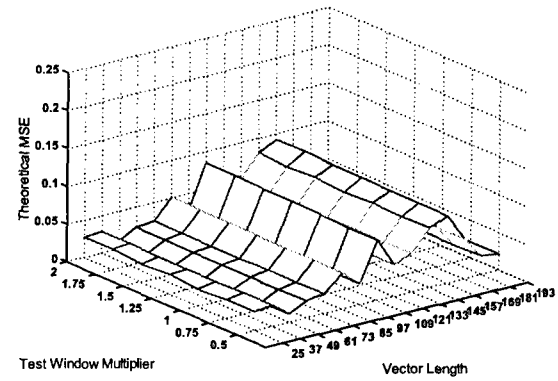


Fig. 7. Theoretical MSE as a function of training vector length and input vector length. The LS-SVM DDAG system includes four class and a four degree polynomial kernel. The test window multiplier defines the input vector length, i.e. the input vector length ranges between 0.5 to 2 times the training vectorl ength.

TABLE V

PROCESSING TIMES, IN SECONDS, FOR ONE-VS-ONE MULTICLASS LS-SVM FOR DOA ESTIMATION.

	Vector Size							
	25	50	75	100	125	150	175	200
Train	0.30	0.94	2.25	4.49	7.39	11.27	15.23	20.38
Test	0.20	0.23	0.31	0.47	0.56	0.66	0.72	0.91

as a basic indicator for possible hardware implementation and real-time applications.

The data in this section shows that the design of the training vectors is important, but there is a tolerance in the selection of noise power and training vector length. The available tolerance in choosing parameters of the training vectors validates the design of the LS-SVM DOA estimation algorithm. This characteristic allows flexibility in the system design and provides a high confidence level in the DOA estimates. In addition,w hen considering real-time implementation of the algorithm, the dimensions of the training vector must be carefully reviewed. Shorter training vectors offer high performance, in terms of MSE, and fast training times.

F. Range of DDAG Parameters for DOA Estimation

The exceptional performance of the LS-SVM DDAG DOA estimation algorithm has been proved in thep revious sections. Most thep revious simulation results wereb ased on three and four class DDAGs. To cover the desired span of the antenna array sector the algorithm must be flexible in the numbero f DDAG classes and DOA ranges. Different applications require different DDAG architectures. Many times the application will require fast training and high accuracy. Training a LS-SVM DDAG system can be performed offline. But covering a large antenna sector with high resolution would require either:

TABLE VI
PERCENTAGE OF MISCLASSIFICATIONS VERSUS DDAG CLASSES (3-6) AND DOA RANGES (1-10).

	DOA Range between Classes, Degrees									
Classes	1	2	3	4	5	6	7	8	9	10
3	0	0	0	0	6.7	0	4.8	4.2	0	0
4	0	0	0	0	0	0	3.6	3.1	0	0
5	0	0	0	0	4.0	0	2.9	0	6.7	0
6	0	0	0	0	0	0	4.8	0	5.6	0

1) A DDAG with a large number of classes and a small DOA range,
2) A two stage system where the antenna sector is partitioned into a set number of classes with a wide DOA range. First, the signal is detected in a specific partition, then a DDAG structure for high resolution can classify the DOA with high accuracy

Whatever the desired approach is, the LS-SVM DDAG algorithm must be flexible in design and robust in performance.

The data in this section proves the performance for a wide range of DDAG structures. Simulations were conducted for three to ten classes with DOA ranges between 1° and 20°. With these classes and DOA ranges the LS-SVM DDAG algorithms is able to span antenna sectors of 3° to 90°. Table VI lists the number of misclassifications. Seventy-five percent of the DDAG structures with DOA ranges between 1° and 10° have zero misclassifications; the average rate of misclassifications for the set of DDAG structures is 1.2%. The largest percentage of misclassifications is 6.7% and occurs with a five class DDAG with a nine degree DOA range.

G. Multilabel Capability for Multiple DOAs

In DOA estimation for cellular systems, there can be multiple DOAs for a given signal. This results from multipath effects induced by the communication channel. The machine learning system must be able to discriminate between a small number of independent DOAs that include signal components with similar time delays. With this constraint the machine learning algorithm then mustb e a multiclass system and able to process multiple labels.

The machine learning algorithm must generate multiclass labels, $y_i \epsilon C$, where $C \epsilon [-90, 90]$ is a set of real numbers that represent an appropriate range of expected DOA values, and multiple labels $y_i, i = 1 \ldots L$ for L dominant signal paths. If antenna sectoring is used in the cellular system the multiclass labels are from the set $C \epsilon [\mathbb{S}_i]$, where $\mathbb{S}_i$ is field of view for the ith sector.

Multilabel classification is possible with the LS-SVM DDAG algorithm presented in Section V. The machine learning algorithm for DOA estimation assigns DOA labels to each eigenvector in the signal subspace. By repeating the DDAG cycle for each eigenvector the multiclass algorithm has the capability of assigning multiple labelst o the input signal.

VII. COMPARISON TO STANDARD DOA ESTIMATION ALGORITHMS

Thep erformanceo f the one-vs-one multiclass LS-SVM algorithm for DOE estimation is described, in detail, in the previous section. The results show that the multiclass classification approach to DOA estimation providesu nique benefits,i n terms of computational complexity and flexibility. Each algorithm is trained for C DOA classes. The number of classes is dependent upon on the antenna sectoring and required resolution. The ideal application of this technique is CDMA cellular systems. For a CDMA system the desired interference suppression dictates the fixed beamwidth. A reduction in beamwidth corresponds to a reduction in MAI, thus reducing the required transmit power at the mobile subscriber. CDMA offers this flexibility since the all mobiles use the same carrier frequency. For Frequency Division Multiple Access (FDMA) systems a narrow beamwidth is desired, since frequency reuse factors into the capacity of a cellular system, thus requiring accurate DOA estimates with high resolution.

A. Computational Complexity

Conventional subspace based DOA estimation algorithms, such as MUSIC and ESPRIT, are computationally complex. The algorithms require accurate knowledge of the signal subspace dimension and accurate estimates of the signal and noise subspace eigenvectors. Additionally, the MUSIC algorithm requires a precise characterization of the antenna array and the ESPRIT algorithm requires multiple eigen decompositions.

Theo ne-vs-onem ulticlass LS-SVM algorithm for DOA estimation is flexible, with respect to computationally requirements. The training cycle for the LS-SVM based DDAG iss traight forward and can be completed offline with simulated data. The only information required is the size of the antenna array and the number of DDAG nodes, which corresponds to DOA classes.F or accurate DOA estimates the only information required, for the LS-SVM DDAG testing cycle, is the dimension of thea ntenna array and accurate eigenvector estimates of the sample covariance matrix. The dimension of the signal subspace is not required, nor is accurate characterization of the antenna array.

B. Simulation Results

Figure 8 compares the one-vs-one multiclass LS-SVM DOA estimation algorithm and the MUSIC algorithm. The top window shows perfect DOA estimation for the machine learning method presented in this paper. The multiclass

algorithm includes an eight class DDAG and a one degree DOA range per class. Note that multiclass LS-SVM algorithm classifies signalso utside the DOA classest o the nearest class, as shown with theD OAs at 12° – 14° and 23° – 25°. The bottom window displays the DOA estimation with the MUSIC algorithm,1 00 DOA estimates are averaged for each received signal and the amplitudes are normalized to the largest estimate. Thep lots show that ther esolution capabilities one-vs-one multiclass LS-SVM DOA estimation algorithm equal that of the MUSIC algorithm.One drawback of the MUSIC algorithm is the broad width of the DOA estimate; a level detection step is required to accurately select the maximum response.

Figure 9 compares the errors and DOA estimates ofe ach algorithm. For this simulation the one-vs-one multiclass LS-SVM algorithm includes a seventeen class DDAG and a five degree DOA range per class. The top window plots the errorsi n the DOA estimates for ninety degree antenna sector and one DOA sample per degree. The definitionso f an error are specific to the two algorithms. For the machine learning based algorithm, an error is defined as a DOA that is classified into a wrong DOA class. For the MUSIC algorithm an error is the difference between the estimated DOA and the actual DOA. As shown in the top window, the only errors associated with theL S-SVM based algorithm occur for DOAs greater than 82°. The DOAs in error are classified into the spatially adjacent DOA class at 80°. Likewise, the errors associated with the MUSIC algorithm, that are greater than 1°,o ccur for DOAs greater than 70°. The plots in Figure 9 prove the robust performance of the one-vs-one multiclass LS-SVM algorithm for DOA estimation.

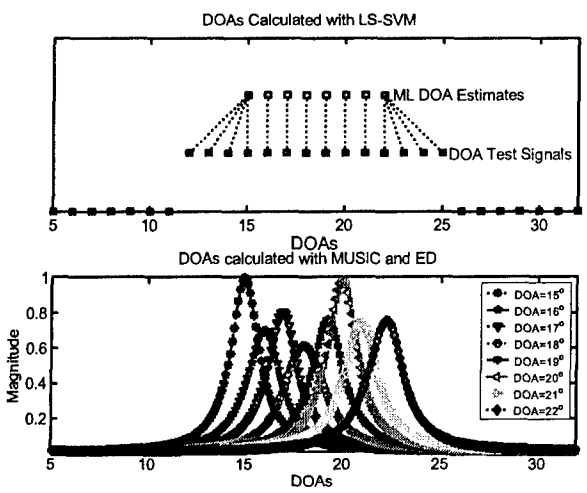


Fig. 8. Comparision between the LS-SVM based DOA estimation algorithm and the MUSIC algorithm. The one-vs-one multiclass LS-SVM DOA estimation algortihm includes eight classes and a one degree DOA range.

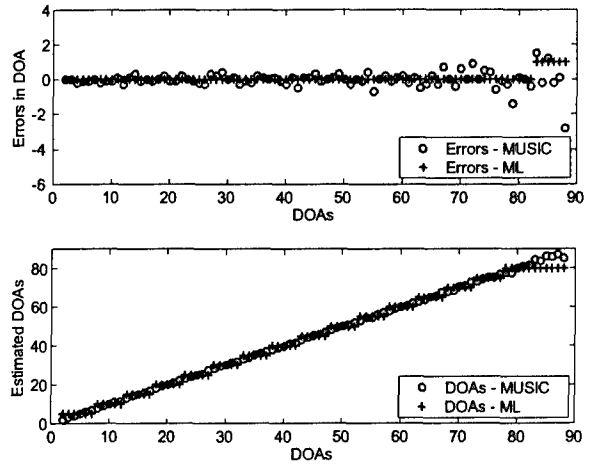


Fig. 9. Comparision of errors and estimated DOAs for the LS-SVM based DOA estimation algorithm and the MUSIC algorithm. The one-vs-one multiclass LS-SVM DOA estimation algortihm includes seventeen classes and a five degree DOA range.

C. Benefits over Standard Techniques

Evaluation of the performance statistics, Section VI, proves that the one-vs-one multiclass LS-SVM algorithm for DOA estimation is reliable with a high degree of accuracy. In terms of performance our new algorithm provides the same capabilities as the standard DOA estimation methods.S pecifically, accurate DOA estimates, to a one degree resolution, can be achieved with the standard subspace based algorithms and our machine learning based algorithm. The primary benefits of our LS-SVM based DOA estimation algorithm are the reduced computational complexity, described above, and the flexibility, in terms of DOA classes versus requirements. The specific application dictatest he desired resolution and therefore the number of DOA classes. For example, one application may include a sixty degree antenna sector and a desired resolution of ten degrees. These requirements would translate into a seven class system. Another application may include a twenty degree sector and a desired resolution of two degrees; this would translate into a eleven class system. An additional option is to place two DDAG systems in series, as described in Section VI-F, that allows for a high resolution with a small number of classes. In general, the one-vs-one multiclass LS-SVM algorithm for DOA estimation can be adapted to specific requirements, as influenced by system capacity, channel conditions, and available computational resources. The MUSIC and ESPRIT algorithms offer no flexibility, in terms of DOA resolution and computational resources.

VIII. Conclusion

In this paper we presented a machine learning architecture for DOA estimation as applied to a CDMA cellular system.

The broad range of our research in machine learning based DOA estimation includesm ulticlass and multilabel classification, classification accuracy, error control and validation, kernel selection, estimation of signal subspace dimension, and overall performance characterization. Wep resented an overview of a multiclass SVM learning method ands uccessful implementation of a one-vs-one multiclass LS-SVM DDAG system forD OA estimation.

The LS-SVM DOA estimation algorithm is superior to standard techniquesd ue to the robust design that is insensitive to received SIR, Doppler shift, size of the antenna array, and the computational requirementsa re adaptable to the desired applications. The algorithm was designed with a multiclass, multilabel capability and includes an error control and validation process. In addition, there are many limitations of standard DOA estimation algorithms, ESPRIT and MUSIC, that do not exist with the LS-SVM DOA estimation algorithm.

The LS-SVM algorithm for DOA estimation assigns DOA labels to each eigenvector in the signal subspace. By repeating the DDAG cycle for each eigenvector the multiclass algorithm hast he capability of assigning multiple labelst o the input signal. Simulation results show a high degree of accuracy and prove that the LS-SVM DDAG system has a wide range of performance capabilities. The results show that the algorithm is accurate for a large range of DDAG performance independent of DDAG class or DOA range per class. The LS-SVM DDAG system accurately classifies the DOAs for three to ten classes and DOA ranges from one degree to twenty degrees.

REFERENCES

[1] J.C. Liberti, Jr. and T.S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Prentice Hall, Upper Saddle River, NJ, 1999.

[2] J.H. Winters, "Signal Acquisition and Tracking with Adaptive Arrays in the Digital Mobile Radio System IS-54 with Flat Fading," IEEE Transactions on Vehicular Technology, Vol. 42, No. 4, 377-384, November 1993.

[3] Z. Raida, "Steeringa n Adaptive Antenna Arrayb yt he Simplified Kalman Filter," IEEE Transactions on Antennas and Propagation, Vol. 43, No. 6, 627-629, June 1995.

[4] A.H. El Zooghby, C.G. Christodoulou, and M. Georgiopoulos, "A NeuralN etwork-Based SmartA ntenna For Multiple Source Tracking", IEEE Transactions On Antennas and Propagation, vol. 48, no. 5, pp. 768-776, May 2000

[5] A.H. El Zooghby, C.G. Christodoulou, and M. Georgiopoulos, "Performance of Radial-Basis Function Networks for Direction of Arrival Estimationw ith Antenna Arrays", IEEE Transactions OnA ntennas and Propagation, vol. 45, no. 11, pp. 1611-1617, November 1997

[6] Ahmed H. El Zooghby, Christos G. Christodoulou, and Michael Georgiopoulos, "Neural Network-BasedA daptive Beamformingf or One-a nd Two- Dimensional Antenna Arrays", IEEE Transactions On Antennas and Propagation, vol. 46, no. 12, pp. 1891-1893, December 1998

[7] Richard O. Duda, Perter E. Hart, and David G. Stork, *Pattern Classification, Second Edition* John Wiley & Sons, New York, NY, 2001.

[8] F. Rashid-Farrokhi, L. Tassiulas, K.J. Ray Liu, "Joint Optimum Power Control and Beamforming in Wireless Networks Using Antenna Arrays", IEEE Transactions On Communications, vol. 46, no. 10, pp. 1313-1324, October 1998.

[9] D.J. Sebald and J.A. Bucklew, "Support Vector Machine Techniques for Nonlinear Equalization",I EEE Transactions On SignalP rocessing, vol. 48, no. 11, pp. 3217-3226, November 2000.

[10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines* , Cambridge University Press, New York, *2000.*

[11] J.A.K. Suykens, "Support Vector Machines: A Nonlinear Modelling and ControlP erspective",E uropean Journalo f Control,v ol 7,p p.3 11-327, 2001

[12] J.C. Platt, N.Christianini, and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification", in *Advances in Neural Information Processing Systems*,v ol.1 2,p p.5 47-553,C ambridge,M A, MIT Press, 2000.

[13] J.A.K. Suykens, L. Lukas, P. Van Dooren, B. DeMoor, and J. Vandewalle, "Least Squares Support Vector Machine Classifiers: a Large Scale Algorithm", ECCTD'99 European Conf. on Circuit Theory and Design, pp. 839-842, August 1999.

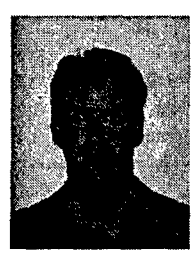
[14] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines* World Scientific, New Jersey, 2002.

[15] R.O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", IEEE Transactions on Antennas and Propagation, AP-34, pp. 276-280, March 1986.

[16] R.H. Roy, and T. Kailath, "ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques", IEEE Transactions On Acoustics, Speech, and Signal Processing, vol. 37, no. 7, pp. 984-995, July 1989.

[17] B.Y ang, "Projection Approximation Subspace Tracking", IEEE Transactions on Signal Processing,v ol.4 3,n o. 1, pp.9 5-107,J anuary 1995.

[18] G. Guo, S.Z. Li, and K.L. Chan, "Support Vector Machines for Face Recognition", Image and Vision Computing, vol 19, pp. 631-638, 2001.



Judd A. Rohwer received his BS degree in electrical engineering in 1996 from GMI Engineering & Management Institute, Flint, MI and his MS and Ph.D. degrees in electrical engineering from University of New Mexico in 2000 and 2003 respectively.U pon his graduation from GMI he accepted a technical staff position atS andia National Laboratories, where he has been involved with various system engineeringa ndR &D programs. Dr. Rohwer is a Member of the IEEE and his researchi nterests include signal andi mage processing,w ireless communications,w ireless networks,a nd machine learning.

Chaouki T. Abdallah obtained his MS and Ph.D. in Electrical Engineering from GA Tech in 1982, and 1988 respectively. He joined the Electrical and Computer Engineering department at the University of New Mexico where he is currently professor, associate chair, and the director of the graduate program. Dr. Abdallah was Exhibit Chair of the 1990 International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and the Local Arrangements Chairf or the 1997 American Control Conference. He is currently serving as the Program Chair for the 2003 Conference on Decision and Control. His research interests are in the area of wireless communications, robust control, and adaptive and nonlinear systems. He is a co-editor of the IEEE Press book Robot Control: Dynamics, Motion Planning, and Analysis; co-author of the book Control of Robot Manipulators, published by Macmillan, and of Linear Quadratic Control: An Introduction, published by Prentice Hall. Dr. Abdallah is a senior member of IEEE anda recipient of the IEEE Millennium medal.