Regularized Blind Detection for MIMO Communications

Yuejie Chi, Yiyue Wu and Robert Calderbank
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: {ychi, yiyuewu, calderbk}@princeton.edu

***Abstract*—Multiple-Input Multiple-Output (MIMO) systems improve the throughput and reliability of wireless communications. Perfect Channel State Information (CSI) is needed at the receiver to perform coherent detection and achieve the optimal gain of the system. In fast fading and low SNR regimes, it is hard or impossible to obtain perfect CSI, which leads the receiver to operate without knowledge of the CSI and perform blind detection. In reality CSI may be available to the receiver but this CSI may be insufficient to support coherent detection. In this paper, we fill the gap between coherent and blind detection by considering a more realistic model where the receiver knows the statistics of the channel, that is Channel Distribution Information (CDI). We propose a new detection algorithm, called Regularized Blind Detection (RBD), where coherent and blind detection can be viewed as special cases in our model. The algorithm estimates CDI from any training symbols that are available and maximizes performance given the estimated CDI. Simulations demonstrate significant improvement in performance over blind detection. Our work can be viewed as a systematic exploration of space between coherent and blind detection with a strong Bayesian statistic flavor.**

I. Introduction

Multiple-Input Multiple-Output (MIMO) systems [1], [2] enable higher rate transmission on fading channels through higher spectral efficiency while maintaining reliability through spatial diversity. The use of space-time codes, introduced by Tarokh et al. [3], [4], has further improved the reliability of communication over fading channel by correlating signals across different transmit antennas. However, from the perspective of Shannon capacity [5], [6], full realization of system potential is heavily dependent on knowledge of Channel State Information (CSI) both at the transmitter and the receiver.

To perform coherent detection, the channel is usually estimated at the receiver through the transmission of a sufficiently long training/pilot sequence.This system overhead reduces the available data rate. Hassibi et al. [7] explored the value of training by analyzing channel capacity when the channel statistics are known and training is sufficient to obtain an estimate of the CSI. In fast fading scenarios and low SNR regimes, it requires more resources and bandwidth to acquire perfect CSI, and it may be impractical to introduce the necessary overhead. In this case, the transmitter will transmit few training/pilot sequences (to resolve phase ambiguity), and the receiver will perform blind detection with around 3dB SNR degradation in performance. One of the most important approaches is joint maximum likelihood data detection and channel estimation, which has been extensively studied in [8]-[14]. The sphere decoding algorithm [15] can be used to find the lattice point in the signal constellation to minimize the target norm; in particular, it is shown in [16] that for a wide range of SNR the complexity of the sphere decoding algorithm is polynomial, making it feasible in many applications.

In practice, requiring a choice between coherent and blind detection is too restrictive. In systems where coherent detection is impractical, i.e. not enough resources can be allocated to estimate exact CSI, it is usually possible, and much easier to obtain partial knowledge of the channel information, for example, the Channel Distribution Information (CDI). When the channel is assumed to be Gaussian, it can be described by its mean and covariance matrix. We provide a graceful Bayesian approach to estimate CDI that does not require the transmission of a minimum number of training symbols. Once perfect CDI is assumed known at the receiver, we propose a new detection algorithm based on a Bayesian framework for joint data detection and channel estimation, called Regularized Blind Detection (RBD), and we describe two RBD variants. Conventional coherent and blind detection are special cases of choosing the regularization parameter. Maximum likelihood detection can be realized using a modified version of sphere decoding, so the complexity is the same as blind detection. In the simulation, our algorithm performs very close to coherent detection in the low SNR regime, and still much better than blind detection in the high SNR regime. Our work can be viewed as a systematic exploration of space between coherent and blind detection with a strong Bayesian statistic flavor. It is worth mentioning that Bayesian detection is also considered in [17] for interference cancellation in MIMO systems using Alamouti signaling.

The paper is organized as follows. In Section II we describe the MIMO model used in this paper and explain both coherent and blind detection. In Section III we assume perfect CDI at the receiver and present two variants of the RBD algorithm within a Bayesian framework. In Section IV we analyze how to extract CDI from training data within our Bayesian framework and interpret the classical use of training data as selection of regularization parameters. Section V provides numerical results. Finally, Section VI draws the conclusion.

A note on notation: We use capital boldface to denote matrices and vectors, and use $\|\cdot\|_F$ for the Frobenious norm. For a matrix $\mathcal{A}$, $\mathcal{A}^\dagger$ denotes its Penrose-Moore pseudo-inverse, $\mathcal{A}^H$ denotes its conjugate transpose, and $\text{Tr}(\mathcal{A})$ denotes its trace. $\mathbf{I}$ denotes the identity matrix.

II. System Model

We consider a general MIMO wireless communication system with N_t transmit antennas and N_r receive antennas in a block fading channel model, where the channel is constant over T consecutive blocks, after which it changes to an independent constant for another T consecutive blocks. The received signals $\mathcal{Y} \in \mathbb{C}^{N_r \times T}$ over T consecutive blocks at the receiver is given by

$$\mathcal{Y} = \mathcal{H}\mathcal{X} + \mathcal{N}, \tag{1}$$

where $\mathcal{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, $\mathcal{N}$ is the additive noise matrix with i.i.d. complex Gaussian random variable entries, and $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_T\} \in \mathbb{C}^{N_t \times T}$ is the transmitted symbol matrix with $\mathbf{X}_i$, $i = 1, \cdots, T$ as one coding block whose entries are taken from a signal constellation Λ such as QPSK and QAM.

A. Coherent Detection

Coherent detection handles the case when $\mathcal{H}$ is perfectly known at the receiver. The coherent maximum likelihood (ML) decoding rule are given as

$$\hat{\mathcal{X}} = \underset{\mathcal{X} \in \Lambda^{KT}}{\text{argmin}} \|\mathcal{Y} - \mathcal{H}\mathcal{X}\|_F^2 \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm and K is the number of symbols in one coding block. Since the noise is independent from one coding block to another, the decoding scheme in equation (2) can be decomposed into single coding blocks as

$$\hat{\mathcal{X}}_i = \underset{\mathcal{X}_i \in \Lambda^{K}}{\text{argmin}} \|\mathcal{Y}_i - \mathcal{H}\mathcal{X}_i\|_F^2 \tag{3}$$

where $\mathcal{Y}_i = \mathcal{H}\mathcal{X}_i + \mathcal{N}_i$ with $\mathcal{N}_i$ as the corresponding noise for coding block i.

B. Blind Detection

Blind detection handles the case when $\mathcal{H}$ is unknown at the receiver. In conventional blind detection, the decoding rule is

$$\begin{aligned}\hat{\mathcal{X}} &= \underset{\mathcal{X} \in \Lambda^{KT}, \mathcal{H} \in \mathbb{C}^{N_r \times N_t}}{\text{argmin}} \|\mathcal{Y} - \mathcal{H}\mathcal{X}\|_F^2 \\ &= \underset{\mathcal{X} \in \Lambda^{KT}}{\text{argmin}} \left\{ \min_{\mathcal{H} \in \mathbb{C}^{N_r \times N_t}} \|\mathcal{Y} - \mathcal{H}\mathcal{X}\|_F^2 \right\}. \end{aligned} \tag{4}$$

In equation (4), the inner minimization is a least square problem given $\mathcal{X}$, so the close form for the estimate $\hat{\mathcal{H}}$ is given by [14]

$$\hat{\mathcal{H}} = \mathcal{Y}\mathcal{X}^H[\mathcal{X}\mathcal{X}^H]^{-1}. \tag{5}$$

In the case of orthogonal codes where $\mathbf{X}_i\mathbf{X}_i^H = K\mathbf{I}$, equation (5) can be reduced to

$$\hat{\mathcal{H}} = \frac{1}{KT}\mathcal{Y}\mathcal{X}^H. \tag{6}$$

Substituting equation (6) into equation (4), we have

$$\begin{aligned}\hat{\mathcal{X}} &= \underset{\mathcal{X} \in \Lambda^{KT}}{\text{argmin}} \|\mathcal{Y}\|_F^2 - \frac{1}{KT}\|\mathcal{Y}\mathcal{X}^H\|_F^2 \\ &= \underset{\mathcal{X} \in \Lambda^{KT}}{\text{argmax}}\, \text{Tr}\{\mathcal{Y}\mathcal{X}^H\mathcal{X}\mathcal{Y}^H\}. \end{aligned} \tag{7}$$

Remark: Blind detection is actually not totally blind. In order to solve the ambiguity, the first coding block is normally assumed known. In another word, blind detection still requires the use of limited piloting signals.

In both cases, sphere decoding [15] is a general technique which can efficiently reduce the average computational complexity of maximum likelihood decoding. Some other optimal decoding algorithms (cf [18] etc) achieve low complexity by taking advantage of coding structures. Note that introducing correlation at the transmitter through space-time codes leads to correlation at the receiver which is a form of CDI.

III. Regularized Blind Detection

In this section, we formulate the decoding problem in a Bayesian probabilistic model, where we assume perfect CDI (namely, the distribution and corresponding parameters of the channel) is known at the receiver. In practice, the estimation of CSI always involves uncertainty of the true value, so that it is reasonable to assume that the true channel follows certain distribution with the estimated CSI as the mean.

From the Bayesian statistical viewpoint, the coherent detection algorithm in equation (2) is equivalent to maximizing the probability of receiving $\mathcal{Y}$ given channel $\mathcal{H}$ and data $\mathcal{X}$ with known $\mathcal{H}$, i.e.

$$\hat{\mathcal{X}} = \underset{\mathcal{X} \in \Lambda^{KT}}{\text{argmax}}\, \text{Pr}(\mathcal{Y}|\mathcal{X}, \mathcal{H}); \tag{8}$$

similarly, the blind detection algorithm in equation (4) is equivalent to maximizing the same probability given channel $\mathcal{H}$ and data $\mathcal{X}$ without assuming any prior on $\mathcal{H}$, which is equivalent to

$$\hat{\mathcal{X}} = \underset{\mathcal{X} \in \Lambda^{KT}}{\text{argmax}} \left\{ \max_{\mathcal{H} \in \mathbb{C}^{N_r \times N_t}} \text{Pr}(\mathcal{Y}|\mathcal{X}, \mathcal{H}) \right\}, \tag{9}$$

where $\text{Pr}(\mathcal{Y}|\mathcal{X}, \mathcal{H}) = \text{Pr}(\mathcal{H}\mathcal{X} + \mathcal{N}) \sim \mathcal{N}(\mathcal{H}\mathcal{X}, \sigma_n^2\mathbf{I})$ is the pdf of the noise matrix $\mathcal{N}$.

When $\mathcal{H}$ is assumed to have a prior distribution, namely a Gaussian prior on $\mathcal{H}$, $\mathcal{H} \sim \mathcal{N}(\mathbf{\Theta}, \mathbf{\Sigma}_\mathcal{H})$, where $\mathbf{\Theta}$ and $\mathbf{\Sigma}_\mathcal{H}$ are assumed known, we propose two variations of algorithms for data detection:

A. Joint ML estimation of channel and data

Joint maximum likelihood estimation of channel $\mathcal{H}$ and data $\mathcal{X}$ with Gaussian prior on $\mathcal{H}$ is given by:

$$\hat{\mathcal{X}} = \underset{\mathcal{X} \in \Lambda^{KT}}{\text{argmax}} \left\{ \max_{\mathcal{H} \in \mathbb{C}^{N_r \times N_t}} \text{Pr}(\mathcal{Y}|\mathcal{X}, \mathcal{H})\text{Pr}(\mathcal{H}) \right\}. \tag{10}$$

With $\Pr(\mathcal{Y}|\mathcal{H},\mathcal{X}) \sim \mathcal{N}(\mathcal{H}\mathcal{X}, \sigma_n^2\mathbf{I})$, and $\mathcal{H} \sim \mathcal{N}(\mathbf{\Theta}, \mathbf{\Sigma}_\mathcal{H})$,

$$\Pr(\mathcal{Y}|\mathcal{X},\mathcal{H})\Pr(\mathcal{H}) \propto \exp\left\{-\frac{1}{2\sigma_n^2}\|\mathcal{Y}-\mathcal{H}\mathcal{X}\|_F^2\right\} \cdot \exp\left\{-\tfrac{1}{2}\left(\mathcal{H}-\mathbf{\Theta}\right)^H \mathbf{\Sigma}_\mathcal{H}^{-1}\left(\mathcal{H}-\mathbf{\Theta}\right)\right\}. \quad (11)$$

Using the Cholesky decomposition $\mathbf{\Sigma}_\mathcal{H}^{-1} = \mathbf{C}^H\mathbf{C}$, the minimization problem in equation (10) is equivalent to the following regularized estimation:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmin}}\left\{\min_{\mathcal{H}\in\mathbb{C}^{N_r\times N_t}} \|\mathcal{Y}-\mathcal{H}\mathcal{X}\|_F^2 + \lambda\|\mathbf{C}(\mathcal{H}-\mathbf{\Theta})\|_F^2\right\} \quad (12)$$

where $\|\mathbf{C}(\mathcal{H}-\mathbf{\Theta})\|_F^2$ is the regularization term and $\lambda = \sigma_n^2$ is the regularization factor. This is the reason the new algorithm is named by *Regularized Blind Detection*.

Rewriting the terms inside the bracket in equation (12) as

$$\left\|\begin{bmatrix}\Delta\mathcal{Y}^H \\ 0\end{bmatrix} - \begin{bmatrix}\mathcal{X}^H \\ \sqrt{\lambda}\mathbf{C}\end{bmatrix}\Delta\mathcal{H}^H\right\|_F^2 \quad (13)$$

where $\Delta\mathcal{Y} = \mathcal{Y} - \mathbf{\Theta}\mathcal{X}$ and $\Delta\mathcal{H} = \mathcal{H} - \mathbf{\Theta}$. Then the channel matrix that minimizes equation (12) is:

$$\hat{\mathcal{H}} = \mathbf{\Theta} + (\mathcal{Y}-\mathbf{\Theta}\mathcal{X})\mathcal{X}^H(\mathcal{X}\mathcal{X}^H + \lambda\mathbf{\Sigma}_\mathcal{H}^{-1})^{-1}. \quad (14)$$

Substituting this back into equation (12), we have the detection rule for data $\mathcal{X}$:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmin}}\left\{\|\mathcal{Y}-\hat{\mathcal{H}}\mathcal{X}\|_F^2 + \lambda\|\mathbf{C}(\hat{\mathcal{H}}-\mathbf{\Theta})\|_F^2\right\} \quad (15)$$

$$= \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmin}} \operatorname{Tr}\{(\mathcal{Y}-\mathbf{\Theta}\mathcal{X})(\mathbf{I}-\mathcal{X}^H(\mathcal{X}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1})^{-1}\mathcal{X}) \cdot (\mathcal{Y}-\mathbf{\Theta}\mathcal{X})^H\}. \quad (16)$$

B. ML estimation of data

Maximum likelihood estimation of data $\mathcal{X}$ can be written as:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}} \int_\mathcal{H} \Pr(\mathcal{Y}|\mathcal{X},\mathcal{H})\Pr(\mathcal{H})d\mathcal{H} \quad (17)$$

$$= \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}} \mathbb{E}_\mathcal{H}\Pr(\mathcal{Y}|\mathcal{X},\mathcal{H}) \quad (18)$$

$$= \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}} \Pr(\mathcal{Y}|\mathcal{X}). \quad (19)$$

When $\mathcal{H} \sim \mathcal{N}(\mathbf{\Theta}, \mathbf{\Sigma}_\mathcal{H})$, this is computable since

$$\begin{aligned}&\mathbb{E}_\mathcal{H}\Pr(\mathcal{Y}|\mathcal{X},\mathcal{H}) \\ \propto &\int_\mathcal{H}\exp\left\{-\frac{1}{2\sigma_n^2}\|\mathcal{Y}-\mathcal{H}\mathcal{X}\|_F^2\right\} \\ &\cdot\exp\left\{-\frac{1}{2}\left(\mathcal{H}-\mathbf{\Theta}\right)^H\mathbf{\Sigma}_\mathcal{H}^{-1}\left(\mathcal{H}-\mathbf{\Theta}\right)\right\}d\mathcal{H} \\ \propto &|\mathcal{X}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1}|^{-1/2} \\ &\cdot\exp\left\{\frac{1}{2\lambda}\operatorname{Tr}\{(\mathcal{Y}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1}\mathbf{\Theta})(\mathcal{X}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1})^{-1} \right.\\ &\left.\cdot(\mathcal{Y}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1}\mathbf{\Theta})^H\}\right\}\end{aligned} \quad (20)$$

Taking the logarithm of the above formula, the ML estimate is equivalent to

$$\begin{aligned}\hat{\mathcal{X}} = \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}} &-\frac{1}{2}\log|\mathcal{X}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}|+ \\ &\frac{1}{2\lambda}\operatorname{Tr}\{(\mathcal{Y}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1}\mathbf{\Theta})(\mathcal{X}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1})^{-1} \\ &\cdot(\mathcal{Y}\mathcal{X}^H+\lambda\mathbf{\Sigma}_\mathcal{H}^{-1}\mathbf{\Theta})^H\}.\end{aligned} \quad (21)$$

C. Special Case: Orthogonal Codes

As a special case, we consider the orthogonal codes $\mathcal{X}\mathcal{X}^H = KT\mathbf{I}$ and the channel covariance matrix $\mathbf{\Sigma}_\mathcal{H} = \sigma_h^2\mathbf{I}$. Let $\mu = \lambda/\sigma_h^2 = \sigma_n^2/\sigma_h^2$; the first variation of the algorithm in equation (16) can be simplified as

$$\begin{aligned}\hat{\mathcal{X}} &= \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}}\operatorname{Tr}\left\{\mathcal{Y}\mathcal{X}^H\mathcal{X}\mathcal{Y}^H + \mu\left(\mathcal{Y}\mathcal{X}^H\mathbf{\Theta}^H + \mathbf{\Theta}\mathcal{X}\mathcal{Y}^H\right)\right\} \\ &= \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}}\operatorname{Tr}\left\{\mathcal{Y}\mathcal{X}^H\mathcal{X}\mathcal{Y}^H + 2\mu\Re\left[\mathcal{Y}\mathcal{X}^H\mathbf{\Theta}^H\right]\right\}.\end{aligned} \quad (22)$$

Let $\mathcal{W} = \mathbf{\Theta}^H(\mathcal{Y}\mathcal{Y}^H)^{-1}\mathcal{Y}$, we have

$$\hat{\mathcal{X}} = \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}}\operatorname{Tr}\left\{(\mathcal{X}+\mu\mathcal{W})(\mathcal{Y}^H\mathcal{Y})(\mathcal{X}+\mu\mathcal{W})^H\right\}. \quad (23)$$

Equation (23) can be further rewritten as

$$\hat{\mathcal{X}} = \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmin}}\operatorname{Tr}\left\{(\mathcal{X}+\mu\mathcal{W})(\rho\mathbf{I}-\mathcal{Y}^H\mathcal{Y})(\mathcal{X}+\mu\mathcal{W})^H\right\} \quad (24)$$

where ρ is a real constant greater than all the eigenvalues of $\mathcal{Y}^H\mathcal{Y}$.

Note that $\mathcal{Y}^H\mathcal{Y}$ is positive semidefinite, therefore we can perform Cholesky decomposition of $(\rho\mathbf{I}-\mathcal{Y}^H\mathcal{Y})$ such that

$$\mathbf{B}^H\mathbf{B} = \rho\mathbf{I} - \mathcal{Y}^H\mathcal{Y}. \quad (25)$$

Therefore, equation (24) can be reformulated as a standard sphere decoding problem with a shift $\mu\mathcal{W}$.

$$\hat{\mathcal{X}} = \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmin}}\|\mathbf{B}(\mathcal{X}+\mu\mathcal{W})^H\|_F^2. \quad (26)$$

Similarly, the second variation of the algorithm in equation (21) becomes

$$\begin{aligned}\hat{\mathcal{X}} &= \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmax}}\operatorname{Tr}\left\{\mathcal{Y}\mathcal{X}^H\mathcal{X}\mathcal{Y}^H + \mu\left(\mathcal{Y}\mathcal{X}^H\mathbf{\Theta}^H + \mathbf{\Theta}\mathcal{X}\mathcal{Y}^H\right)\right\} \\ &= \underset{\mathcal{X}\in\Lambda^{KT}}{\operatorname{argmin}}\|\mathbf{B}(\mathcal{X}+\mu\mathcal{W})^H\|_F^2.\end{aligned} \quad (27)$$

which coincides with the first variation of RBD. Notice that the variance of the channel plays a role in choosing the shift in sphere decoding.

IV. Extracting CDI from Limited Training

We now suppose some training bits are transmitted to obtain partial information about the channel and we rewrite the received signals $\mathcal{Y}$ over T consecutive blocks at the receiver as

$$\mathcal{Y} = [\mathcal{Y}_\tau\ \mathcal{Y}_d] = \mathcal{H}[\mathcal{X}_\tau\ \mathcal{X}_d] + \mathcal{N} \quad (28)$$

where $\mathcal{X}_\tau \in \mathbb{C}^{N_t\times T_\tau}$ is the training symbol matrix, and $\mathcal{X}_d \in \mathbb{C}^{N_t\times T_d}$ is the data symbol matrix; respectively $\mathcal{Y}_\tau \in \mathbb{C}^{N_t\times T_\tau}$ and $\mathcal{Y}_d \in \mathbb{C}^{N_t\times T_d}$ are the received symbol matrix.

A. Traditional Decoding Schemes with Training.

There are two baseline approaches to using training symbols that do not make use of prior information about channel statistics. The first scheme is to obtain the maximum likelihood detection decoding rule using the formula below

$$\hat{\mathcal{X}}_d = \underset{\mathcal{X}_d \in \Lambda^{KT_d}}{\operatorname{argmin}} \left\{ \min_{\mathcal{H} \in \mathbb{C}^{N_r \times N_t}} \|\mathcal{Y} - \mathcal{H}\mathcal{X}\|_F^2 \right\}. \tag{29}$$

This is equivalent to

$$\hat{\mathcal{X}}_d = \underset{\mathcal{X}_d \in \Lambda^{KT_d}}{\operatorname{argmin}} \left\{ \min_{\mathcal{H} \in \mathbb{C}^{N_r \times N_t}} \|\mathcal{Y}_d - \mathcal{H}\mathcal{X}_d\|_F^2 + \|\mathcal{Y}_\tau - \mathcal{H}\mathcal{X}_\tau\|_F^2 \right\}. \tag{30}$$

The second scheme is to first use the training symbols to get an estimated CSI, where

$$\begin{aligned}\hat{\mathcal{H}} &= \underset{\mathcal{H} \in \mathbb{C}^{N_r \times N_t}}{\operatorname{argmin}} \|\mathcal{Y}_\tau - \mathcal{H}\mathcal{X}_\tau\|_F^2 \\ &= \mathcal{Y}_\tau \mathcal{X}_\tau^H (\mathcal{X}_\tau \mathcal{X}_\tau^H)^{-1},\end{aligned} \tag{31}$$

then use $\hat{\mathcal{H}}$ to decode $\mathcal{X}_d$, i.e.

$$\hat{\mathcal{X}}_d = \underset{\mathcal{X}_d \in \Lambda^{KT_d}}{\operatorname{argmin}} \|\mathcal{Y}_d - \hat{\mathcal{H}}\mathcal{X}_d\|_F^2. \tag{32}$$

The least square channel estimation $\hat{\mathcal{H}}$ from the training data lies in the span of $\mathcal{Y}_\tau$, therefore it is necessary to transmit enough training symbols in order to get a reasonable estimate, i.e. greater than the number of transmit antennas.

B. Unified Approach

It is possible to unify these two approaches by introducing a penalty weight λ to the second training term in equation (30), given as

$$\hat{\mathcal{X}}_d = \underset{\mathcal{X}_d \in \Lambda^{KT_d}}{\operatorname{argmin}} \left\{ \min_{\mathcal{H} \in \mathbb{C}^{N_r \times N_t}} \|\mathcal{Y}_d - \mathcal{H}\mathcal{X}_d\|_F^2 + \lambda\|\mathcal{Y}_\tau - \mathcal{H}\mathcal{X}_\tau\|_F^2 \right\}, \tag{33}$$

then the first scheme is equivalent to choosing $\lambda = 1$, and the second scheme is equivalent to the limit solution when λ goes to $+\infty$.

The general solution to equation (33) is essentially the same as the regularized blind detection scheme discussed earlier under certain parameterizations. In fact, in the special case where the code is orthogonal and the channel variance matrix $\mathbf{\Sigma}_\mathcal{H} = \sigma_h^2 \mathbf{I}$ is a scalar of identity matrix, let $\mathbf{\Theta} = \mathcal{Y}_\tau \mathcal{X}_\tau^H (\mathcal{X}_\tau \mathcal{X}_\tau^H)^{-1} = \mathcal{Y}_\tau \mathcal{X}_\tau^H / KT_\tau$, it can be restated as equation (15).

C. Estimation of CDI through training symbols.

When taking advantage of the training data, the above decoding schemes essentially only obtain a rough least-square estimate of the CSI, and use it either with full confidence ($\lambda \to +\infty$), or no confidence ($\lambda = 1$) without distinguishing training symbols from data symbols. When there is limited training available, the CSI can be very inaccurate. In fact we shall demonstrate that it is more effective to use low-rate training data over L blocks to estimate CDI and to then use this estimate within our proposed RBD algorithms.

One way to estimate the mean $\mathbf{\Theta}$ and variance $\mathbf{\Sigma}_\mathcal{H}$ of the channel is via the following maximum likelihood estimation, using available training symbols from L blocks $(\mathcal{Y}_{\tau,i}, \mathcal{X}_{\tau,i})$, $i = 1, \cdots, L$:

$$\begin{aligned}\mathbf{\Theta}, \mathbf{\Sigma}_\mathcal{H} &= \underset{\hat{\mathbf{\Theta}}, \hat{\mathbf{\Sigma}}_\mathcal{H}}{\operatorname{argmax}} \prod_{i=1}^{L} \mathbb{E}_\mathcal{H} \Pr(Y_{\tau,i}, X_{\tau,i} | \mathcal{H}) \\ &= \underset{\hat{\mathbf{\Theta}}, \hat{\mathbf{\Sigma}}_\mathcal{H}}{\operatorname{argmax}} -\frac{1}{2} \sum_{i=1}^{L} \log |\mathcal{X}_{\tau,i}\mathcal{X}_{\tau,i}^H + \lambda \hat{\mathbf{\Sigma}}_\mathcal{H}^{-1}| + \\ &\frac{1}{2\lambda} \sum_{i=1}^{L} \operatorname{Tr}\big\{ (\mathcal{Y}_{\tau,i}\mathcal{X}_{\tau,i}^H + \lambda \hat{\mathbf{\Sigma}}_\mathcal{H}^{-1} \hat{\mathbf{\Theta}}) (\mathcal{X}_{\tau,i}\mathcal{X}_{\tau,i}^H + \lambda \hat{\mathbf{\Sigma}}_\mathcal{H}^{-1})^{-1} \\ &\cdot (\mathcal{Y}_{\tau,i}\mathcal{X}_{\tau,i}^H + \lambda \hat{\mathbf{\Sigma}}_\mathcal{H}^{-1} \hat{\mathbf{\Theta}})^H - \lambda \hat{\mathbf{\Theta}}^H \hat{\mathbf{\Sigma}}_\mathcal{H}^{-1} \hat{\mathbf{\Theta}} \big\}.\end{aligned} \tag{34}$$

This can be further simplified in the special case when $\mathcal{X}_{\tau,i}\mathcal{X}_{\tau,i}^H = KT_\tau \mathbf{I}$ and $\mathbf{\Sigma}_\mathcal{H} = \sigma_h^2 \mathbf{I}$ as (let $\hat{\mu} = \sigma_n^2 / \hat{\sigma}_h^2$):

$$\begin{aligned}\mathbf{\Theta}, \sigma_h^2 &= \underset{\hat{\mathbf{\Theta}}, \hat{\sigma}_h^2}{\operatorname{argmax}} -\frac{L}{2} \log |KT_\tau + \hat{\mu}| + \\ &\frac{1}{2\lambda} \left\{ \frac{1}{KT_\tau + \hat{\mu}} \sum_{i=1}^{L} \|\mathcal{Y}_{\tau,i}\mathcal{X}_{\tau,i}^H + \hat{\mu}\hat{\mathbf{\Theta}}\|_F^2 - \hat{\mu} L \|\hat{\mathbf{\Theta}}\|_F^2 \right\}.\end{aligned} \tag{35}$$

The choice of $\mathbf{\Theta}$ does not depend on μ, and is given as

$$\mathbf{\Theta} = \frac{1}{KT_\tau L} \sum_{i=1}^{L} \mathcal{Y}_{\tau,i}\mathcal{X}_{\tau,i}^H. \tag{36}$$

Plugging this back to equation (35), and setting its derivatives with respect to $\hat{\mu}$ to zero, we obtain the maximum likelihood estimate of σ_h^2.

V. Numerical results

In our simulation, we consider a wireless communication system with two transmit antennas and a single receive antenna which adopts Alamouti signaling [4]. We consider the decoding of Alamouti signals during 3 consecutive coherent code blocks (namely, 6 consecutive time slots). We assume that the channel matrix $\mathcal{H}$ is complex Gaussian with mean

$$\mathbf{\Theta} = (1 + i, 1 + i)$$

and covariance matrix

$$\mathbf{\Sigma}_\mathcal{H} = \begin{pmatrix} 0.08 & 0 \\ 0 & 0.08 \end{pmatrix}.$$

Fig. 1 shows the comparison of different detection algorithms. For blind detection, the phase ambiguity is assumed to have been solved; for regularized blind detection, CDI is assumed known at the receiver and for coherent detection, CSI is assumed perfectly known at the receiver. The first variation of the proposed RBD algorithm is used. In the simulation, our algorithm performs very close to coherent detection in the low SNR regime, and still much better than blind detection in the high SNR regime.

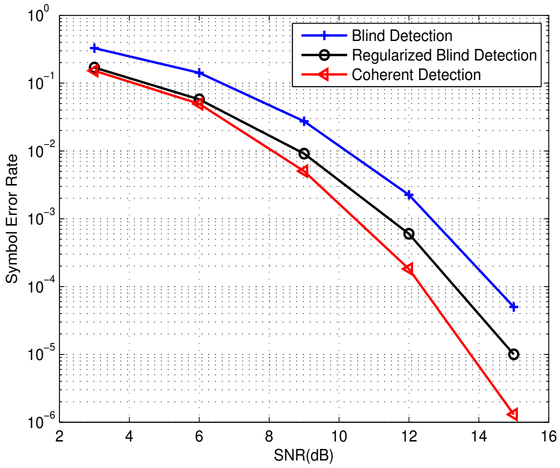

Fig. 1. Comparison of different detection algorithms with decoding block length as 3 coding blocks.

VI. Conclusions

We have introduced and analyzed the performance of Regularized Blind Detection, a new algorithm that provides a systematic way of interpolating between coherent and blind detection. Significant performance benefits are possible in environments where it may be impractical to obtain full CSI. The algorithm requires knowledge of CDI at the receiver and we have described how CDI may be estimated from any available training symbols. Coherent and blind detection are special cases within our Bayesian framework and it is CDI that parameterizes the space in between. Simulations demonstrate that our algorithm performs very close to coherent detection in the low SNR regime, and still much better than blind detection in the high SNR regime.

Acknowledgements

The authors would like to thank Vaneet Aggarval and Waheed Bajwa for helpful discussions.

References

[1] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. New York, NY: Cambridge University Press, 2005.
[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, 1996.
[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication:performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, 1998.
[4] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, 1998.
[5] E. Telatar, "Capacity of multiple-antenna gaussian channels," *European Transactions on Communications*, vol. 10, pp. 585–596, 1999.
[6] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas In Communications*, vol. 21, pp. 684–702, 2003.
[7] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, pp. 951–963, 2003.
[8] H. Vikalo, B. Hassibi, and P. Stoica, "Efficient joint maximum likelihood channel estimation and signal detection," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 1838–1845, 2006.
[9] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Transactions on Information Theory*, vol. 40, pp. 340–349, 1994.
[10] N. Seshadri, "Joint data and channel estimation using fast blind trellis search," *US Patent 5263033*, 1993.
[11] H. Bolcskei, J. R. W. Heath, and A. J. Paulraj, "Blind channel estimation in spatial multiplexing systems using nonredundant antenna precoding," *Proc. of the 33rd Annual IEEE Asilomar Conf. on Signals, Systems, and Computers*, vol. 2, pp. 1127 –1132, 1999.
[12] W.-K. Ma, B.-N. Vo, T. Davidson, and P. Ching, "Blind ML detection of orthogonal space-time block codes: High performance, efficient implementations," *IEEE Transactions on Signal Processing*, vol. 54, pp. 738–751, 2006.
[13] W. Xu, M. Stojnic, and B. Hassibi, "On exact maximum-likelihood detection for non-coherent mimo wireless systems: A branch-estimate-bound optimization framework," *Proceedings of the International Symposium on Information Theory*, pp. 2017–2021, 2008.
[14] E. G. Larsson, P. Stoica, and J. Li, "On maximum-likelihood detection and decoding for space-time coding systems," *IEEE Transactions on Signal Processing*, vol. 50, pp. 937–944, 2002.
[15] M. Pohst, "On the computation of lattice vectors of minimal length, successive minima and reduced basis with applications," *ACM SIGSAM*, vol. 15, pp. 37–44, 1981.
[16] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm: Part I, the expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2806–2818, 2005.
[17] S. Sirianunpiboon, A. R. Calderbank, and S. D. Howard, "Space-polarization-time codes for diversity gains across line of sight and rich scattering environments," *Proc. IEEE ITW*, May 2007.
[18] S. Sirianunpiboon, Y. Wu, A. R. Calderbank, and S. D. Howard, "Fast optimal decoding of multiplexed orthogonal designs by conditional optimization," *to appear, IEEE Transactions on Information Theory*, 2010.