AN INFORMATION GEOMETRIC APPROACH TO CHANNEL IDENTIFICATION

Amin Zia, James P. Reilly, Shahram Shirani

Dept. of ECE, McMaster University, Hamilton, Ontario, Canada
email: amin@grads.ece.mcmaster.ca; { reilly, shirani}@mail.ece.mcmaster.ca

ABSTRACT

In this paper, the semi-blind MIMO channel identification problem is modelled as a *stochastic* maximum likelihood estimation problem and an iterative method, called information geometric identification (IGID), for channel identification and tracking is presented. The method is developed based on the results from information geometry; specifically, the *alternating projections theorem* first proved by Csiszar [1]. It is demonstrated that the proposed method has similar performance compared to a recently reported method based on the expectation maximization (EM) algorithm [2]. Since the IGID method has an analytical solution, the proposed algorithm can be implemented much faster while while having a similar performance. The method can be considered as a generalization of all the methods developed based on the EM algorithm.

1. INTRODUCTION

It is known that in a MIMO wireless system with OFDM modulation, the output of each of N subchannels with M users can be modelled by:

$$\boldsymbol{y}(n) = H\boldsymbol{x}(n) + \boldsymbol{w}(n) \tag{1}$$

where $\boldsymbol{w}(n) \in \Re^N$ is independent white Gaussian noise with covariance Ψ, $H \in \Re^{N\times M}$ is the channel gain matrix and $\boldsymbol{x}(n) \in \Re^M$ and $\boldsymbol{y}(n) \in \Re^N$ are input and output signals, respectively. Due to the orthogonality of the subchannels, the channel identification methods can be applied to each tone independently. Therefore, it is enough to solve the identification problem for each individual tone. Interested readers are encouraged to read [2] for further details.

Stochastic maximum likelihood (ML) estimation for blind channel identification and tracking is the subject of a great deal of research within the last two decades. Rather than assuming a deterministic input to the channel, the main assumption in this approach is to consider a random input space with a certain probability density function (*pdf*). The stochastic ML estimation problem for identifying the channel gains is defined as:

$$\hat{H}_{ML} = \arg\max f(\boldsymbol{y}) \tag{2}$$

where $f(\boldsymbol{y})$ is the likelihood function of the output observations $\boldsymbol{y}$. If a training data set consisting of the input signal and output observation pairs are available, and assuming a Gaussian noise, the so called *complete-data* maximum likelihood estimation of the channel $\hat{H}$ can be solved easily by the pseudo-inverse method [2].

However, in blind identification, since there is no training set and the input is assumed to be unknown, the observations are not sufficient statistics for estimating the channel parameters. Therefore, considering a distribution $p(\boldsymbol{x})$ for the input signal, one can solve the following equivalent *incomplete-data* problem:

$$\hat{H} = \arg\max \int_X f(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}. \tag{3}$$

It is easy to observe that the maximization is over all possible values of the unknown input signal. Therefore, solving this problem usually is very complex. The EM algorithm provides a general prescription to the maximization problem of (3) and is the core of almost all the iterative methods proposed so far [2]. These methods are mainly based on the observation that instead of maximizing the unavailable "complete-data" likelihood it is sufficient to maximize the *expectation* of the likelihood over the input probability distribution, in two main iterations: *Expectation* and *Maximization*. Based on this algorithm, Aldana [2] has applied the EM algorithm to channel identification and tracking in a multiuser MIMO system. The method requires computation of expectations over all possible constellation points. This is a limitation of the method when the number of constellation points is large. This limitation is more cumbersome when the input space is not discrete. In this case, Monte Carlo integration methods are necessary for computing the expectations inherently embedded in the EM-type algorithms.

In this paper, the semi-blind MIMO channel identification problem is modelled as a *stochastic* maximum likelihood estimation problem and an iterative method for channel identification and tracking is presented. Since the algorithm works in batch mode on blocks of input data, it is assumed that the channel remains constant over each block. This assumption holds for high data rate wireless applications. The method is developed based on the results from information geometry; specifically, the *alternating projections theorem* first proved by Csiszar [1] which provides an iterative method for minimizing the distance between two sets of probability distributions. It is demonstrated that IGID has similar performance compared to a recently reported method based on the expectation maximization (EM) algorithm [2] when applied to a multi-input single-output (MISO) channel identification problem. The IGID algorithm provides similar performance; however, it admits an analytical solution, which provides a faster recursive solution compared to the previous EM-type algorithm, in which complex multidimensional integrations are usually necessary. This characteristic provides a very fast implementation relative to previous algorithms.

2. MATHEMATICAL BACKGROUND

2.1. A Brief Look into Information Geometry

From information geometry, the Kullback-Leibler distance, a.k.a *information divergence (I-div)*, between two probability distributions p and q is defined as:

$$D(p||q) = \int_X p(x) \log \frac{p(x)}{q(x)} d(\boldsymbol{x}). \tag{4}$$

Given that $\mathcal{P}$ as a convex set of probability distributions (PD) then:

$$p^* = \min_{p \in \mathcal{P}} D(p||q) \tag{5}$$

is called the I-projection of q on $\mathcal{P}$. The convexity of $\mathcal{P}$ guarantees the existence and the uniqueness of the I-projection. The proposed method in this paper is based on the following theorem proved originally by Csiszar [1]: Let $\mathcal{P}$ and $\mathcal{Q}$ be convex sets of PD measures on $\mathcal{X}$. Also let

$$p^* = \min_{p \in \mathcal{P}} D(p||q) \tag{6}$$

$$q^* = \min_{q \in \mathcal{Q}} D(p||q) \tag{7}$$

be the I-projection of p on $\mathcal{Q}$ and q on $\mathcal{P}$, respectively. Then if $\{p_n\}_{n=0}^{\infty}$and $\{q_n\}_{n=0}^{\infty}$ are the sequences obtained by alternately minimizing $D(p||q)$ using (6) and (7), starting from some $p_0 \in \mathcal{P}$, then we have:

$$\lim_{n \to \infty} D(p_n||q_n) = \inf_{q \in \mathcal{Q}, p \in \mathcal{P}} D(\mathcal{P}||\mathcal{Q}). \tag{8}$$

Furthermore, if $\mathcal{X}$ is finite and $\mathcal{P}$ and $\mathcal{Q}$ are closed in the topology of point-wise convergence, then:

$$p_n \to p^* \qquad \text{such that} \qquad D(p^*||\mathcal{Q}) = \min D(\mathcal{P}||\mathcal{Q}). \tag{9}$$

Under certain conditions [1], this theorem guarantees that alternating projections between the two convex sets $\mathcal{P}$ and $\mathcal{Q}$ converges to the minimum distance of the two sets. According to this theorem, iteratively projecting a PD onto the convex sets of PD's converges to the I-projection of the PD on the intersection of the sets.

We now assume that $\mathcal{Q}$ is the set of all possible PD's defining the complete-data likelihood function specified by the MIMO channel model (1), parameterized by H and Ψ. Assume also that $\mathcal{P}$ is the set of all possible PD's known as *empirical distributions* whose marginal distributions (with respect to the input data $\boldsymbol{x}(n)$) are equal to the observed output likelihood, i.e. $f(\boldsymbol{y})$.

It is shown [3] that the maximum likelihood estimation problem using incomplete data can be considered as an iterative minimization of the distance between the two probability distribution sets $\mathcal{Q}$ and $\mathcal{P}$:

$$\hat{q} = \arg\max_{q \in \mathcal{Q}} f(\boldsymbol{y}) = \arg\min_{q \in \mathcal{Q}} \min_{p \in \mathcal{P}} D(\mathcal{P}||\mathcal{Q}) \tag{10}$$

Therefore, alternatively minimizing the KL distance between $\mathcal{P}$ and $\mathcal{Q}$ will converge to a solution for the original ML estimation problem, i.e. estimating the model parameters H and Ψ, defined in (1).

3. SEMI-BLIND CHANNEL IDENTIFICATION AND TRACKING USING ML ESTIMATION

We now apply the proposed information geometry method to the semi-blind channel identification and tracking problem. The method is *semi-blind* due to the fact that the initial point of the algorithm is obtained by training the algorithm with a small number of training data in each block. As explained in Section 2, we need to define two convex probability distributions $\mathcal{P}$ and $\mathcal{Q}$. Assume the source $\boldsymbol{x}(n)$ in (1) is distributed as $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x} - \boldsymbol{\mu}, \Phi)$, where $\mathcal{N}(\boldsymbol{x} - \boldsymbol{\gamma}, \boldsymbol{\Sigma})$ denotes a normal distribution in the random variable $\boldsymbol{x}$, with mean $\boldsymbol{\gamma}$ and covariance $\boldsymbol{\Sigma}$. Then:

$$\mathcal{Q} = \{q | q(\boldsymbol{y}, \boldsymbol{x}) = f(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x})\} \tag{11}$$

where $f(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y} - H\boldsymbol{x}, \Psi)$ is the likelihood function. Therefore $\mathcal{Q}$ is the set of all the likelihood distributions with normal distribution:

$$q(\boldsymbol{y}, \boldsymbol{x}) = p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z} - \hat{\boldsymbol{z}}, Q) \tag{12}$$

where:

$$\hat{\boldsymbol{z}} = \begin{pmatrix} H\boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix} \tag{13}$$

$$Q = \begin{pmatrix} H\Phi H^T + \Psi & H\Phi \\ \Phi H^T & \Phi \end{pmatrix} \tag{14}$$

where we have used $\Phi^T = \Phi$.
The expression for Q^{-1} is given as: [3]:

$$Q^{-1} = \begin{pmatrix} \Psi^{-1} & -\Psi^{-1}H \\ -H^T\Psi^{-1} & \Phi^{-1} + H^T\Psi^{-1}H \end{pmatrix}. \quad (15)$$

Also assume we model the observations as an empirical Normal distribution $\tilde{p}(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y} - \boldsymbol{r}, S)$. We define:

$$\mathcal{P} = \{p(\boldsymbol{y}, \boldsymbol{x}) | \int_X p(\boldsymbol{y}, \boldsymbol{x}) d\boldsymbol{x} = \tilde{p}(\boldsymbol{y})\}. \quad (16)$$

Obviously $\mathcal{P}$ is the set of all empirical distributions with marginal equal to the observation distribution $\hat{p}(\boldsymbol{y})$.

3.1. The First Projection: Computing the Best Complete-Data Distribution

Having an initial distribution Q_0, the first projection provides the closest empirical distribution in $\mathcal{P}$ to Q_0 whose marginal coincides with the observation distribution. It is shown that the first projection (6) is solved according to [3]:

$$p^* = \arg\min_{p \in \mathcal{P}} D(p||q) = q(\boldsymbol{x}|\boldsymbol{y})\tilde{p}(\boldsymbol{y}). \quad (17)$$

Therefore, the optimum distribution that minimizes the objective function is the posterior distribution of complete-data likelihood distribution. In the Gaussian case, a closed form solution exists [3]:

$$p^* = q(\boldsymbol{x}|\boldsymbol{y})\tilde{p}(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{z} - \boldsymbol{m}, \boldsymbol{P}^*). \quad (18)$$

where:

$$\boldsymbol{m} = \hat{\boldsymbol{z}} + P^* S^{(-1)} \begin{pmatrix} H\boldsymbol{\mu} - \boldsymbol{r} \\ 0 \end{pmatrix} \quad (19)$$

$$(P^*)^{-1} = \begin{pmatrix} \Psi^{-1} - (H\Phi H^T)^{-1} + S^{-1} & -\Psi^{-1}H \\ -H^T\Psi^{-1} & \Phi^{-1} + H^T\Psi^{-1}H \end{pmatrix} \quad (20)$$

where $S^{(-1)}$ is the covariance matrix S^{-1} properly augmented with zero blocks. Therefore, to solve the first projection, i.e., to calculate $(P^*)^{-1}$, it is sufficient only to add the necessary terms to the first element of the inverse covariance matrix (15). The closed-form solution (20) avoids multi-dimensional integrations usually necessary in EM-type algorithms [2].

3.2. The Second Projection: the Complete-Data ML Estimation

The second projection is a complete-data maximum likelihood estimation. The computed distribution P^* is in fact the best complete-data empirical distribution. Therefore, the problem in the second projection is to find the best distribution in the likelihood model family $\mathcal{Q}$ that minimizes the *I-divergence*. Since the $\mathcal{Q}$ family is parameterized by H and Ψ, the optimization is performed with respect to these parameters. For the second projection it is necessary to solve the following minimization problem:

$$q^* = \arg\min_{q \in \mathcal{Q}} D(p^*||q). \quad (21)$$

Assuming that P^* has the following block matrix form:

$$P^* = \begin{pmatrix} P_{11} & P_{12} \\ {P_{12}}^T & P_{22} \end{pmatrix}, \quad (22)$$

and since q^* is parameterized by H and Ψ the second projection is equivalent to the following minimization [3]:

$$\begin{aligned} \{H^*, \Psi^*\} &= \arg\min_{\{H,\Psi\}} trace(\Psi^{-1}P_{11}) & (23)\\ &- 2trace(\Psi^{-1}HP_{12}^T) & (24)\\ &+ trace(\Phi^{-1}P_{22} + H^T\Psi^{-1}HP_{22}) & (25)\\ &- \log\det\Psi^{-1} - \log\det\Phi^{-1} & (26)\\ &- \log\det P^* + d, & (27)\end{aligned}$$

where $d = M + N$ is the dimension of the complete-data $\boldsymbol{z} = [\ \boldsymbol{y} \quad \boldsymbol{x}]^T$, $p^*$and $P^* \in \Re^{d\times d}$ are the probability distribution and the corresponding covariance matrix computed from the first projection, respectively, and Q is the covariance matrix for the likelihood distribution q. The minimization of the objective with respect to the parameters H and Ψ gives [3]:

$$H^* = P_{12}P_{22}^{-1} \quad (28)$$

$$\Psi^* = P_{11} - P_{12}P_{22}^{-1}P_{12}^T. \quad (29)$$

Therefore the iterative application of Equation (20), (28) and (29) gives a recursive update for the model parameters $\hat{H}$ and Ψ.

3.3. Convergence

It is easy to show that $\mathcal{P}$ and $\mathcal{Q}$ are convex sets of probability distributions. Therefore, when $\mathcal{P}$ and $\mathcal{Q}$ are disjoint, and assuming a bounded KL distance measure, the final convergent point does not depend on the initial condition [1]. However, an arbitrarily chosen initial point can cause a very slow convergence. Also, special conditions must be imposed on the projections to avoid the sets $\mathcal{P}$ and $\mathcal{Q}$ colliding. Numerical inaccuracies can cause the failure of these assumptions to hold, which can result in problems with convergence [[3]]. Therefore, in this paper we consider a semi-blind application of the algorithm in which an initial point is chosen using the training set in each data block.

4. SIMULATION RESULTS

The IGID method is used for semi-blind channel identification and tracking in an ISI-free, flat-fading MISO wireless system with OFDM modulation. For the purpose of comparison with results provided in [2], it is assumed that there are two users, ($M = 2$), and one receiving antenna ($N = 1$). The results can be extended to the general case with slight modifications.

It is assumed that blocks of $L = 1000$ received symbols are used. In each block, the algorithm is initialized using training by a set of 20 symbols. Each user transmits ± 1 in each subchannel and the channel gains are chosen from an iid Gaussian random variable with $\|\boldsymbol{H}\| = 1$.

Figures 1 and 2 show the root mean square error (RMSE) for estimation of the channel gains and the noise variance computed for 30 Monte Carlo runs, comparing the results from the IGID, the EM algorithm [2] and the ideal complete-data ML estimation. The normalized channel gain mean-squared error is defined as $MSE_H = \frac{\|H - \hat{H}\|^2}{M}$ and the noise variance MSE as $MSE_{\sigma^2} = |\sigma^2 - \hat{\sigma}^2|^2$ [2].

It can be seen from the figures that the performance of the IGID algorithm is close to optimum. Further, it is shown that the IGID has better performance for some values of noise variance compared to the EM algorithm, with less computational complexity. Since the IGID has a rigorous foundation in statistics and information theory, it can be considered as a generalized framework for iterative estimation problems.

5. DISCUSSION

In this paper, a general solution to the maximum likelihood estimation of channel parameters and noise covariance matrix, called information geometric identification (IGID) is provided and it is shown that the method has analytical recursive update equations when the noise is assumed to be Gaussian. The results for the IGID algorithm are compared to the results of a recently reported method based on the expectation maximization (EM) algorithm [2]. It is shown that the IGID method has a similar performance while benefitting from an analytical solution. Thus, complex multidimensional integrations usually necessary in similar EM-type methods are avoided. This characteristic provides very fast computation times relative to previous algorithms. The proposed method can be used for any other estimation problem which has a similar signal model.

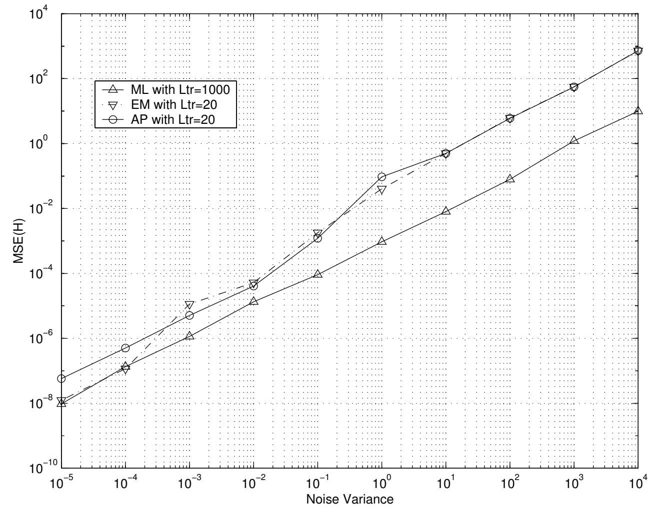

Fig. 1. Channel Gain RMSE (30 Monte Carlo) for 2 users.

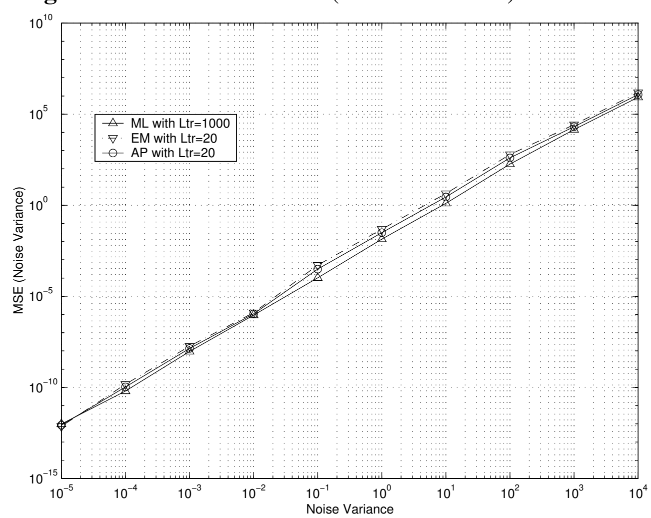

Fig. 2. Noise Variance RMSE (30 Monte Carlo) for 2 users.

6. REFERENCES

[1] I. Csiszar and G. Tusnady, "Information Geometry and Alternative Minimization Procedures," *Statistics and Decisions, Suppl. Issue, No. 1, 1984, Page(s) 205-237.*

[2] C. H. Aldana and J. Cioffi, "Channel Tracking for MIMO Systesm using EM algotihms," *IEEE International Conference on Communications, 2001.*

[3] Amin Zia, "Maximum Likelihood Estimation of Incomplete Data using Divergence Minimization," Technical Report, Statistical Signal Processing Lab. (http://www.ece.mcmaster.ca/∼reilly/), McMaster University, 2003.