Decision Directed Iterative Channel Estimation for MIMO Systems

Xinmin Deng, Alexander M. Haimovich
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, NJ 07102, USA
Email: xinmin.deng@ieee.org, haimovic@njit.edu

Javier Garcia-Frias
Department of Electrical and Computer Engineering
University of Delaware
Newark, DE 19716, USA
Email: jgarcia@mail.eecis.udel.edu

***Abstract*— Decision directed channel estimation is investigated in this paper. For systems with multiple transmit antennas, high computation complexity of the inversion of a data-dependent matrix hinders the application of optimal channel estimation. An iterative method is introduced to avoid the matrix inversion. Performance of the proposed method is demonstrated by simulation results.**

I. Introduction

The use of multiple antennas at both the transmitter and receiver in wireless communications has been shown to provide significant increase in capacity [6], [16]. In particular, if the fading between pairs of transmit and receive antenna are independently Rayleigh distributed, it is known that for high enough signal-to-noise ratio (SNR), the average capacity grows linearly with the smaller of the number of transmit or receive antennas, provided that perfect channel state information (CSI) is available at the receiver. However in real world perfect CSI is never known *a priori*. In practice an estimate of CSI is obtained from known pilot symbols and subsequently used for decoding as if it were exact. Therefore the performance depends on the quality of channel estimate and hence the number of pilot symbols. Estimation of the multiple-input-multiple-output (MIMO) fading channel is a major challenge for multiple antenna systems. When the number of antennas increases, accurate channel estimation becomes more difficult because of the increase in the number of parameters to be estimated. In [2], [10], effect of availability of pilot symbols on MIMO fading channels is determined by information theoretic approaches.

Recently there is increasing interest in joint channel and data decoding [3], [4], [5], [7], [11], [12], [14], [17], where data decision obtained from the decoding, either hard or soft, is used as additional training to refine the channel estimate. In our discussion the channel is assumed to be time varying flat Rayleigh fading and the channel statistics is known. For single transmit antenna systems the optimal maximum *a posteriori* (MAP) channel estimate can be obtained by applying a data-independent Wiener filter to the maximum likelihood (ML) estimate. For multiple transmit antenna systems, however, the ML channel estimate does not exist, and it can be shown the computation of the optimal channel estimate involves the inversion of a data-dependent matrix [3], [5], [12]. Due to the high computation complexity of matrix inversion and the fact that the matrix inversion has to be calculated in each iteration, optimal decision directed channel estimation for MIMO channels is difficult to implement in practice. In this paper, we propose an iterative method for decision directed channel estimation in MIMO systems. The resulting estimator can be interpreted as applying a data-independent filter to the weighted sum of the minimum norm least square (MNLS) estimate and the estimate obtained from the previous iteration. When the weight parameter is chosen appropriately, for single transmit antenna systems the proposed iterative method reduces to the optimal estimator.

This paper is organized as follows. The system and channel models are described in Section II. Section III contains the development of the optimal MAP estimate and an iterative alogrithm for its evaluation. Simulation results are provided in section IV to illustrate the performance of the proposed method. Conclusions are given in Section V.

II. System Model

The MIMO system of interest is equipped with N transmit antennas and M receive antennas. The channel between each transmit and receive antenna pair is modeled as flat fading process. In the sequel, lower- and upper-case boldface letters are used to denote column vectors and matrices, respectively. We use $\mathbf{0}$ to denote the all-zero vector. Matrices $\mathbf{I}$ and $\mathbf{O}$ stand for the identity matrix and all-zero matrix, respectively. The discrete time received signal at the mth receive antenna at time t, $r_{m,t}$, is given by

$$r_{m,t} = \mathbf{s}_t^T \mathbf{h}_{m,t} + n_{m,t} \tag{1}$$

where $\mathbf{h}_{m,t} = [h_{1m,t}\ h_{2m,t}\ \cdots\ h_{Nm,t}]^T$ is the $N \times 1$ vector whose nth component represents the complex fading gain between the nth transmit antenna and the mth receive antenna at time t, $\mathbf{s}_t = [s_{1,t}\ s_{2,t}\ \cdots\ s_{N,t}]^T$ is the $N \times 1$ vector of symbols simultaneously transmitted by the N transmit antennas at time t, $n_{m,t}$ is the complex noise sample with variance $N_0/2$ per dimension.

We consider the transmission of a block (a *frame*) of L symbol vectors, $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_L$. The received sequence at the mth receive antenna, denoted by an $L \times 1$ vector $\mathbf{r}_m =$

0-7803-7802-4/03/$17.00 © 2003 IEEE

$[r_{m,1}\ r_{m,2}\ \cdots\ r_{m,L}]^T$, is of the form

$$\mathbf{r}_m = \mathbf{S}\mathbf{h}_m + \mathbf{n}_m \tag{2}$$

where $\mathbf{n}_m$ is an $L \times 1$ (possibly colored) complex Gaussian noise vector with zero mean and covariance matrix $\mathbf{R_n} = E[\mathbf{n}_m \mathbf{n}_m^H]$ (the superscript H denotes the Hermitian operation). For white noise we have $\mathbf{R_n} = N_0\mathbf{I}$. The $NL \times 1$ vector $\mathbf{h}_m = [\mathbf{h}_{m,1}^T\ \mathbf{h}_{m,2}^T\ \cdots\ \mathbf{h}_{m,L}^T]^T$ is the vector of channel parameters. The $L \times NL$ data matrix $\mathbf{S}$ is defined as

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1^T & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{0}^T & \mathbf{s}_2^T & \cdots & \mathbf{0}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^T & \mathbf{0}^T & \cdots & \mathbf{s}_L^T \end{bmatrix}. \tag{3}$$

For Rayleigh fading the channel coefficients $h_{nm,t}$ can be modeled as zero mean complex Gaussian random variables. If the antennas are spaced sufficiently far apart, the fading is assumed to be uncorrelated across antennas. For two-dimensional isotropic scattering (Clarke's model), the correlation function can be expressed as

$$E[h_{n_1m_1,t_1} h^*_{n_2m_2,t_2}] = \delta_{n_1n_2}\delta_{m_1m_2} J_0(2\pi f_D(t_1 - t_2)) \tag{4}$$

where $\delta_{n_1n_2} = 1$ if $n_1 = n_2$ and $\delta_{n_1n_2} = 0$ otherwise, J_0 is the zeroth order modified Bessel function of the first kind, f_D is the normalized fading rate.

Since the fading is assumed to be uncorrelated across antennas, the observation $\mathbf{r}_m$ provides no information for the channel parameter $\mathbf{h}_{m'}$ if $m \neq m'$. In the sequel we suppress the receive antenna index m and (2) becomes

$$\mathbf{r} = \mathbf{S}\mathbf{h} + \mathbf{n}. \tag{5}$$

Given the decision on $\mathbf{S}$ (from detector or decoder), we are interested in the estimation of the channel parameter $\mathbf{h}$ based on the observation $\mathbf{r}$.

Conditioned on $\mathbf{S}$ and $\mathbf{h}$, we have $\mathbf{r} \sim \mathcal{CN}(\mathbf{Sh}, \mathbf{R_n})$, where $\mathcal{CN}$ denotes the complex-valued Gaussian distribution. Therefore the conditional probability density function (pdf) for $\mathbf{r}$ is

$$p(\mathbf{r}|\mathbf{h},\mathbf{S}) = \pi^{-L}|\mathbf{R_n}|^{-1} \exp(-(\mathbf{r}-\mathbf{Sh})^H \mathbf{R_n}^{-1}(\mathbf{r}-\mathbf{Sh})). \tag{6}$$

For Rayleigh fading channel, the *a priori* pdf for $\mathbf{h}$ is

$$p(\mathbf{h}) = \pi^{-NL}|\mathbf{R_h}|^{-1} \exp(-\mathbf{h}^H \mathbf{R_h}^{-1}\mathbf{h}) \tag{7}$$

where the covariance matrix $\mathbf{R_h}$ is determined by the normalized fading rate via (4).

III. Channel Estimation

A. Optimal Estimate

The optimal minimum mean square error (MMSE) estimate of $\mathbf{h}$ based on the received signal vector $\mathbf{r}$ and transmitted data matrix $\mathbf{S}$, is the conditional mean, $E[\mathbf{h}|\mathbf{r},\mathbf{S}]$. Since $\mathbf{r}$ and $\mathbf{h}$ are jointly Gaussian, the conditional mean is equivalent to the maximum *a posteriori* (MAP) estimate of $\mathbf{h}$, $\hat{\mathbf{h}}_{MAP}(\mathbf{S})$, which maximizes the *a posteriori* pdf $p(\mathbf{h}|\mathbf{r},\mathbf{S})$ with respect to $\mathbf{h}$. A necessary condition for this maximization is

$$\frac{\partial}{\partial \mathbf{h}} \ln p(\mathbf{h}|\mathbf{r},\mathbf{S})\big|_{\mathbf{h}=\hat{\mathbf{h}}_{MAP}(\mathbf{S})} = \mathbf{0}. \tag{8}$$

The *a posteriori* pdf for $\mathbf{h}$ is

$$p(\mathbf{h}|\mathbf{r},\mathbf{S}) = \frac{p(\mathbf{r}|\mathbf{h},\mathbf{S})p(\mathbf{h}|\mathbf{S})}{p(\mathbf{r}|\mathbf{S})} \tag{9}$$

where $p(\mathbf{r}|\mathbf{h},\mathbf{S})$ and $p(\mathbf{h}|\mathbf{S}) = p(\mathbf{h})$ are given in (6) and (7), respectively. The pdf $p(\mathbf{r}|\mathbf{S})$ does not affect the maximization over $\mathbf{h}$. From (6), (7) and (9), we have

$$\frac{\partial}{\partial \mathbf{h}} \ln p(\mathbf{h}|\mathbf{r},\mathbf{S}) = \underbrace{(\mathbf{S}^H\mathbf{R_n}^{-1}\mathbf{S} + \mathbf{R_h}^{-1})}_{\mathbf{A}(\mathbf{S})}\mathbf{h} - \underbrace{\mathbf{S}^H\mathbf{R_n}^{-1}\mathbf{r}}_{\mathbf{b}}. \tag{10}$$

The MAP estimate of $\mathbf{h}$ is found by setting the derivative in (10) equal to zero. Thus, we obtain

$$\hat{\mathbf{h}}_{MAP}(\mathbf{S}) = (\mathbf{S}^H\mathbf{R_n}^{-1}\mathbf{S} + \mathbf{R_h}^{-1})^{-1}\mathbf{S}^H\mathbf{R_n}^{-1}\mathbf{r}. \tag{11}$$

The computation of (11) involves the inversion of an $NL \times NL$ matrix $\mathbf{A}(\mathbf{S})$. By application of the matrix inversion lemma, it can be shown that the cannel estimate can be expressed as

$$\hat{\mathbf{h}}_{MAP}(\mathbf{S}) = \mathbf{R_h}\mathbf{S}^H(\mathbf{S}\mathbf{R_h}\mathbf{S}^H + \mathbf{R_n})^{-1}\mathbf{r} \tag{12}$$

where the dimension of the matrix to be inverted is $L \times L$. Expressions (11) and (12) have been employed as the optimal channel estimator in iterative channel estimation and decoding of space time coded systems [3], [5], [12]. Since computation of (12) requires the inversion of an $L \times L$ matrix of complexity $O(L^3)$, the computational cost is very high for practical frame sizes L. Furthermore, the matrix to be inverted is in general dependent on transmitted data $\mathbf{S}$, which is highly undesirable for practical implementation.

For systems with single transmit antenna ($N = 1$), in the case of mutiple phase shift keying (MPSK) modulation ($s^*_{n,t}s_{n,t} = 1$, $n = 1, 2, \cdots, N$, $t = 1, 2, \cdots, L$) and white noise ($\mathbf{R_n} = N_0\mathbf{I}$), we have in (11) $\mathbf{A}(\mathbf{S}) = \frac{1}{N_0}\mathbf{I} + \mathbf{R_h}^{-1}$. Therefore $\mathbf{A}(\mathbf{S})$ becomes independent of $\mathbf{S}$ and the estimate can be simplified to

$$\begin{aligned} \hat{\mathbf{h}}_{MAP}(\mathbf{S}) &= (\tfrac{1}{N_0}\mathbf{I} + \mathbf{R_h}^{-1})^{-1}\mathbf{S}^H\tfrac{1}{N_0}\mathbf{r} \\ &= \underbrace{\mathbf{R_h}(\mathbf{R_h} + N_0\mathbf{I})^{-1}}_{\mathbf{F_W}} \underbrace{\mathbf{S}^H\mathbf{r}}_{\hat{\mathbf{h}}_{ML}(\mathbf{S})}. \end{aligned} \tag{13}$$

Equation (13) can be interpreted as a Wiener filter $\mathbf{F_W}$ applied to the maximum likelihood (ML) estimate $\hat{\mathbf{h}}_{ML}(\mathbf{S})$ [13]. It has been employed as the optimal channel estimator in iterative channel estimation and decoding for systems with single transmit antenna [14], [17].

B. Iterative solution

Instead of carrying out the direct matrix inversion in (11), we can use iterative methods to find an approximate solution of the MAP equation

$$\mathbf{A}(\mathbf{S})\mathbf{h} = \mathbf{b}. \tag{14}$$

Let $\mathbf{A} = \mathbf{Q} - \mathbf{P}$ be any decomposition of $\mathbf{A}$ such that $\mathbf{Q}$ is nonsingular. Let $\hat{\mathbf{h}}^{(0)}$ be an arbitrary initial vector, then the vector sequence $\hat{\mathbf{h}}^{(0)}, \hat{\mathbf{h}}^{(1)}, \hat{\mathbf{h}}^{(2)}, \cdots$ generated by the following iteration

$$\mathbf{Q}\hat{\mathbf{h}}^{(k+1)} = \mathbf{P}\hat{\mathbf{h}}^{(k)} + \mathbf{b} \tag{15}$$

converges to the true solution if and only if the spectral radius of the iteration matrix $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{P}$ satisfies $\rho(\mathbf{B}) < 1$, where the spectral radius $\rho(\mathbf{B})$ is defined as the modulus of the largest eigenvalue of $\mathbf{B}$.

Different decompositions of $\mathbf{A}$ result in iterative methods given by different names, such as Jacobi, Gauss-Seidel, and successive overrelaxation (SOR) iteration. Details of convergence and computation aspects of these methods can be found in mathematical literature [8] [9]. In order to avoid calculating the inverse of a data dependent matrix in decision directed channel estimation, we can decompose $\mathbf{A}(\mathbf{S})$ into data independent part $\mathbf{Q}$ and data dependent part $-\mathbf{P}(\mathbf{S})$, i.e., $\mathbf{A}(\mathbf{S}) = \mathbf{Q} - \mathbf{P}(\mathbf{S})$. One straightforward way to do so is

$$\mathbf{Q} = C\mathbf{I} + \mathbf{R}_{\mathbf{h}}^{-1}, \tag{16}$$

$$\mathbf{P}(\mathbf{S}) = C\mathbf{I} - \mathbf{S}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{S}, \tag{17}$$

where C is some constant.

With $\mathbf{Q}$ and $\mathbf{P}$ difined in (16) and (17), the iteration becomes

$$\begin{aligned}\hat{\mathbf{h}}^{(k+1)} &= \underbrace{(C\mathbf{I} + \mathbf{R}_{\mathbf{h}}^{-1})^{-1}}_{\mathbf{Q}^{-1}} \\ &\quad (\underbrace{(C\mathbf{I} - \mathbf{S}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{S})}_{\mathbf{P}(\mathbf{S})}\hat{\mathbf{h}}^{(k)} + \mathbf{S}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{r}).\end{aligned} \tag{18}$$

With regard to the convergence of (18), we have the following proposition:

Proposition 1: If the noise is white, the iteration (18) converge to the optimal estimate if

$$C \geq \frac{N}{2N_0}. \tag{19}$$

Proof: Since $\mathbf{R}_{\mathbf{h}}$ is positive definite, all the eigenvalues of $\mathbf{Q}$ are greater than C, hence $\rho(\mathbf{Q}^{-1}) < \frac{1}{C}$. Let λ be any eigenvalue of $\mathbf{P}$, we have $C - \frac{N}{N_0} \leq \lambda \leq C$. If $C \geq \frac{N}{2N_0}$, then $-C \leq C - \frac{N}{N_0} \leq \lambda \leq C$ so that $\rho(\mathbf{P}) \leq C$. Therefore

$$\rho(\mathbf{Q}^{-1}\mathbf{P}) \leq \|\mathbf{Q}^{-1}\mathbf{P}\|_2 \tag{20}$$

$$\leq \|\mathbf{Q}^{-1}\|_2\|\mathbf{P}\|_2 \tag{21}$$

$$= \rho(\mathbf{Q}^{-1})\rho(\mathbf{P}) \tag{22}$$

$$< 1, \tag{23}$$

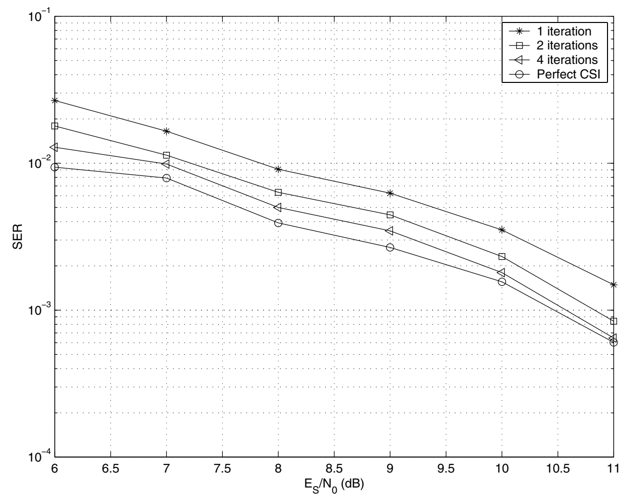

Fig. 1. SER versus E_S/N_0 ($f_D = 0.001$)

where $\|\cdot\|_2$ denotes spectral norm, inequalities (20), (21) and (23) are immediately clear, equality (22) holds because $\mathbf{Q}^{-1}$ is symmetric and $\mathbf{P}$ is Hermitian. ■

From the signal model (5), we can formulate an underdetermined least-squares problem for the channel estimate, which affords the following minimum norm solution

$$\hat{\mathbf{h}}_{LS}(\mathbf{S}) = \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}\mathbf{r}. \tag{24}$$

Similar to (13), (18) can be written as

$$\begin{aligned}\hat{\mathbf{h}}^{(k+1)} &= \mathbf{R}_{\mathbf{h}}(\mathbf{R}_{\mathbf{h}} + \tfrac{1}{C}\mathbf{I})^{-1} \\ &\quad ((\mathbf{I} - \tfrac{1}{C}\mathbf{S}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{S})\hat{\mathbf{h}}^{(k)} + \tfrac{1}{C}\mathbf{S}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{r}) \\ &= \underbrace{\mathbf{R}_{\mathbf{h}}(\mathbf{R}_{\mathbf{h}} + \frac{1}{C}\mathbf{I})^{-1}}_{\mathbf{F}}((\mathbf{I} - \tfrac{1}{C}\mathbf{S}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{S})\hat{\mathbf{h}}^{(k)} \\ &\quad + \underbrace{\frac{1}{C}\mathbf{S}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{S}}_{\mathbf{W}_{LS}(\mathbf{S})}\underbrace{\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}\mathbf{r}}_{\hat{\mathbf{h}}_{LS}(\mathbf{S})}) \\ &= \mathbf{F}((\mathbf{I} - \mathbf{W}_{LS}(\mathbf{S}))\hat{\mathbf{h}}^{(k)} + \mathbf{W}_{LS}(\mathbf{S})\hat{\mathbf{h}}_{LS}(\mathbf{S})).\end{aligned} \tag{25}$$

It can be seen from (25) that the next estimate, $\hat{\mathbf{h}}^{(k+1)}$, is obtained by applying a filter, $\mathbf{F}$, to the weighted sum of the current estimate, $\hat{\mathbf{h}}^{(k)}$, and the minimum norm least square estimate, $\hat{\mathbf{h}}_{LS}(\mathbf{S})$.

For systems with single transmit antenna, MPSK modulation and white noise, we have $\mathbf{W}_{LS}(\mathbf{S}) = \frac{1}{C}\frac{1}{N_0}\mathbf{I}$. In particular, if we choose $C = \frac{1}{N_0}$, then $\mathbf{W}_{LS}(\mathbf{S}) = \mathbf{I}$ and $\mathbf{I} - \mathbf{W}_{LS}(\mathbf{S}) = \mathbf{O}$, so that (25) reduces to (13).

IV. Simulation Results

The performance of the proposed iterative decision channel estimation is demonstrated by a space time code system with joint channel estimation and data decoding. The system structure is described in [3]. Numerical results were obtained for the 4PSK 8 state space time code presented in [15] with 2 transmit and 2 receive antennas. Each frame consists of 14 pilot symbols and 116 data symbols. The pilot symbols are

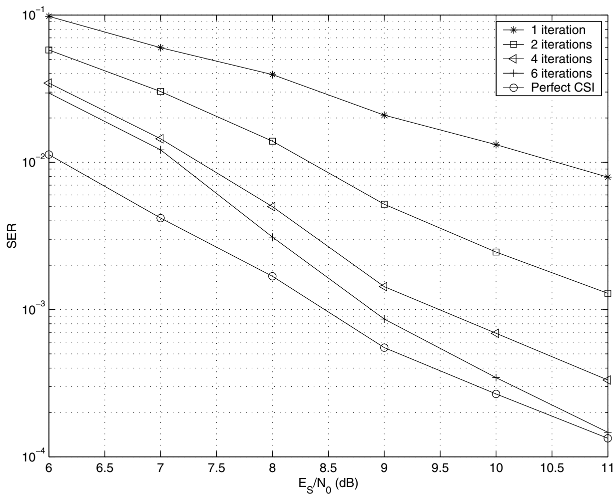

Fig. 2. SER versus E_S/N_0 ($f_D = 0.01$)

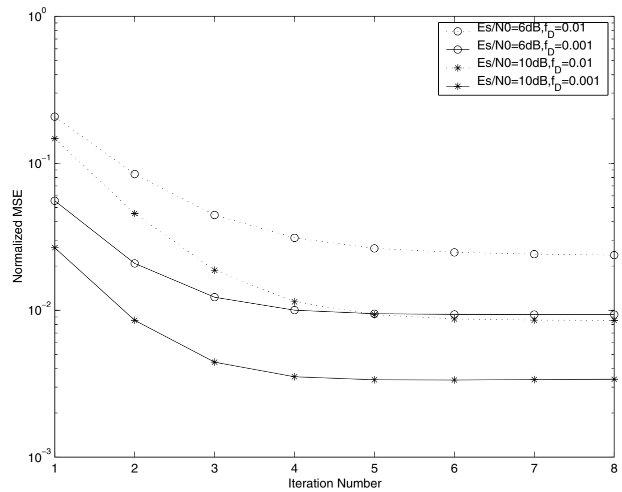

Fig. 3. Normalized MSE versus iteration number

used to obtain the initial estimate of channel. Soft decision feedback is used to improve the channel after each iteration. The soft decision of each symbol is obtained by averaging with respect to its *a posteriori* probability from the BCJR decoding algorithm [1].

The symbol error rate (SER) performance versus SNR per symbol (E_S/N_0) is evaluated for normalized fading rates $f_D = 0.001$ and $f_D = 0.01$ in Fig. 1 and Fig. 2, respectively. Performance of the same code with perfectly known channel state information at the receiver is also included for comparison. Significant performance gain from the decision feedback from the decoding can be observed. In both cases, the performance with the decision directed channel estimate at high SNR is within 0.5 dB of that with known CSI.

Fig. 3 shows the normalized mean square error (MSE) versus the number of iterations. It can be seen that the performance improves with the increase in the number of iterations, but the improvement is negligible after 4 and 6 iterations for $f_D = 0.001$ and $f_D = 0.01$, respectively.

V. Conclusion

A simple iterative method for decision directed channel estimation for MIMO systems is introduced in this paper. Its application to the joint channel estimation and data decoding for space time coded system is illustrated. Simulation results suggest that near optimal performance can be achieved with reasonable number of iterations.

References

[1] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory,* vol. IT-20, pp. 284-287, March 1974.

[2] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.,* vol. 19, no. 12, pp. 2358-2368, Dec. 2001.

[3] Z. B. Baranski, A. M. Haimovich, and J. Garcia-Frias, "EM-based iterative receiver for space-time coded modulation with noise covariance estimation," *IEEE Globecom,* Nov. 2002.

[4] E. Chiavaccini and G. M. Vitetta, "MAP symbol estimation on frequency-flat Rayleigh faidng channels via a Bayesian EM algorithm," *IEEE Trans. Commun.,* vol. 49, no. 11, pp. 1869-1872, Nov. 2001.

[5] C. Cozzo and B. L. Hughes, "An iterative receiver for space-time communications," *2000 Conference on Information Sciences and Systems,* 2000.

[6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications,* vol. 6, no. 3, pp. 311-335, March 1998.

[7] A. Grant, "Joint decoding and channel estimation for space-time codes," *IEEE Vehicular Technology Conference,* pp. 416-420, 2000.

[8] A. Greenbaum, *Iterative Methods for Solving Linear Systems,* SIAM, 1997.

[9] L. A. Hageman and D. M. Young, *Applied Iterative Methods,* Academic Press, 1981.

[10] B. Hassibi, B. Hochwald, "Optimal training in space time systems," *34th Asilomar Conference on Signals, Systems and Computers*, pp. 743-747, 2000.

[11] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE J. Select. Areas Commun.,* vol. 19, no. 9, pp. 1706-1717, Sept. 2001.

[12] Y. Li, C. N. Georghiades, and G. Huang, "Iterative maximum-likelihood sequence estimation for space-time coded systems," *IEEE Trans. Commun.,* vol. 49, no. 6, pp. 948-951, June 2001.

[13] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing,* John Wiley & Sons, 1998.

[14] H.-J. Su and E. Geraniotis, "Low-complexity joint iterative channel estimation and decoding for pilot symbol-assisted modulation and multiple differential detection systems with correlated Rayleigh fading," *IEEE Trans. Commun.,* vol. 50, no. 2, pp. 249-261, Feb. 2002.

[15] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory,* vol. 44, no. 2, pp. 744-765, March 1998.

[16] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585-595, Nov.-Dec. 1999.

[17] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE J. Select. Areas Commun.,* vol. 19, no. 9, pp. 1697-1705, Sept. 2001.