A Novel Hybrid Approach to Improve Performance of Frequency Division Duplex Systems with Linear Precoding

Paula M. Castro, José A. García-Naya, Daniel Iglesia, and Adriana Dapena

Department of Electronics and Systems
University of A Coruña, Spain
{pcastro,jagarcia,dani,adriana}@udc.es

Abstract. Linear precoding is an attractive technique to combat interference in multiple-input multiple-output systems because it reduces costs and power consumption in the receiver equipment. Most of the frequency division duplex systems with linear precoding acquire the channel state information at the receiver by using supervised algorithms. Such algorithms make use of pilot symbols periodically sent by the transmitter. In a later step, the channel state information is sent to the transmitter side through a limited feedback channel.

In order to reduce the overhead inherent to the periodical transmission of training data, we propose to acquire the channel state information by combining supervised and unsupervised algorithms, leading to a hybrid and more efficient approach. Simulation results show that the performance achieved with the proposed scheme is clearly better than that with standard algorithms.

Keywords: Linear Precoding, MIMO Systems.

1 Introduction

The increased demand of multimedia contents has produced a continuous development of new techniques to try to improve the throughput of digital communication systems. For instance, current transmission standards for Multiple-Input Multiple-Output (MIMO) systems include the so-called precoders in order to guarantee that the link throughput be maximized [1, 2]. Precoding algorithms for MIMO are classified into linear and nonlinear precoding types. In the sequel, we consider Linear Precoding (LP) approaches because they achieve reasonable throughput with a complexity lower than that required by non linear precoding approaches.

In order to be able to implement precoding schemes, the base station must know the Channel State Information (CSI). However, in most of the Frequency Division Duplex (FDD) systems the transmitter (TX) cannot obtain the CSI from the received signals —even under the assumption of perfect calibration— because the channels are not reciprocal. The CSI is thus estimated at the receiver (RX) side and transmitted back through a limited feedback channel. Usually,

E.S. Corchado Rodriguez et al. (Eds.): HAIS 2010, Part II, LNAI 6077, pp. 248–255, 2010.

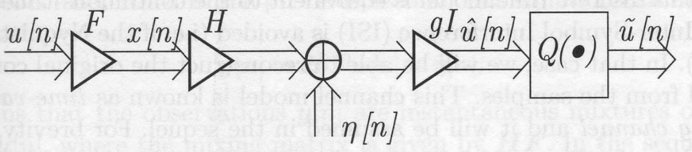

Fig. 1. MIMO System with linear transmit filter (linear precoding)

current standards perform the channel estimation by using supervised algorithms that make use of pilot symbols. Such pilot symbols do not convey information, and therefore, both system throughput and spectral efficiency are penalized.

In this paper, we propose to combine two important paradigms of Neural Networks: supervised and unsupervised learning. The kind of learning to be used is decided based on a simple criterion that determines the time instant when the channel has suffered a significant variation. In such moment, a supervised algorithm is employed to re-estimate the channel making use of pilot symbols. The rest of the time, when the channel variation is not significant enough, the unsupervised algorithm known as Infomax [3] is utilized.

2 System Model

We consider a MIMO system with N_t transmit antennas and N_r receive antennas. The precoder generates the transmit signal $\boldsymbol{x}$ from all data symbols $\boldsymbol{u} = [u_1, \ldots, u_{N_\mathrm{r}}]$ corresponding to the different receive antennas $1, \ldots, N_\mathrm{r}$.We denote the equivalent low-pass channel impulse response between the j–th transmit antenna and the i–th receive antenna as $h_{i,j}(\tau, t)$. For flat fading channels, the channel matrix $\mathbf{H}(t)$ is given by

$$\mathbf{H}(t) = \begin{pmatrix} h_{1,1}(t) & \cdots & h_{1,N_\mathrm{t}}(t) \\ \vdots & & \vdots \\ h_{N_\mathrm{r},1}(t) & \cdots & h_{N_\mathrm{r},N_\mathrm{t}}(t) \end{pmatrix},$$

and the received signal is

$$y_j(t) = \sum_{i=1}^{N_\mathrm{t}} h_{ji}(t)x_i(t) + \eta_j(t), \quad \boldsymbol{y}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t) + \eta(t), \tag{1}$$

where $\eta_j(t)$ is the additive noise, $\boldsymbol{x}(t) = [x_1(t), \ldots, x_{N_\mathrm{t}}(t)]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{t}}$, $\boldsymbol{y}(t) = [y_1(t), \ldots, y_{N_\mathrm{r}}(t)]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{r}}$, and $\eta(t) = [\eta_1(t), \ldots, \eta_{N_\mathrm{r}}(t)]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{r}}$.

In general, if $f[n] = f(nT_\mathrm{s} + \Delta)$ denote the samples of $f(t)$ every T_s seconds with Δ being the sampling delay and T_s the symbol time, then sampling $\boldsymbol{y}(t)$ every T_s seconds yields to the discrete time signal $\boldsymbol{y}[n] = \boldsymbol{y}(nT_\mathrm{s} + \Delta)$ given by

$$\boldsymbol{y}[n] = \boldsymbol{H}[q]\boldsymbol{x}[n] + \eta(n), \tag{2}$$

where $n = 0, 1, 2, \ldots$ corresponds to samples spaced T_s seconds, and q denotes the time slot. The channel remains stationary during a block of N_B symbols.

Note that this discrete time model is equivalent to the continuous-time model in (1) only if Inter-Symbol Interference (ISI) is avoided (i.e. if the Nyquist criterion is satisfied). In that case, we will be able to reconstruct the original continuous-time signal from the samples. This channel model is known as *time-varying flat block-fading channel* and it will be assumed in the sequel. For brevity, we omit the slot index q in the sequel.

At the TX side, a way to carry out the pre–equalizer (or precoding) step consists in including a transmit filter matrix $\boldsymbol{F} \in \mathbb{C}^{N_\mathrm{t} \times N_\mathrm{r}}$, and a RX filter matrix $\boldsymbol{G} = g\boldsymbol{I} \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{r}}$, leading to N_r scalar data streams. Figure 1 shows the resulting communications system in which the data symbols $\boldsymbol{u}[n]$ are passed through the transmit filter $\boldsymbol{F}$ to form the transmit signal $\boldsymbol{x}[n] = \boldsymbol{F}\boldsymbol{u}[n] \in \mathbb{C}^{N_\mathrm{t}}$. Note that the constraint for the transmit energy must be fulfilled. Therefore, the received signal is given by

$$\boldsymbol{y}[n] = \boldsymbol{H}\boldsymbol{F}\boldsymbol{u}[n] + \eta[n] \in \mathbb{C}^{N_\mathrm{r}},$$

where $\boldsymbol{H} \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$, and $\eta[n] \in \mathbb{C}^{N_\mathrm{r}}$ is the Additive White Gaussian Noise (AWGN). After multiplying by the receive gain g, we get the estimated symbols

$$\hat{\boldsymbol{u}}[n] = g\boldsymbol{H}\boldsymbol{F}\boldsymbol{u}[n] + g\eta[n] \in \mathbb{C}^{N_\mathrm{r}}. \tag{3}$$

Clearly, the restriction that all the receivers apply the same scalar weight g is not necessary for decentralized receivers. Replacing $\boldsymbol{G}$ by a diagonal matrix suffices (e.g. [4]). However, usually no closed form can be obtained for the precoder if $\boldsymbol{G}$ is diagonal. Fortunately, $\boldsymbol{F}$ can be found in closed form for $\boldsymbol{G} = g\mathbf{I}$. Thus, we use $\boldsymbol{G} = g\mathbf{I}$ in the following.

Although Wiener filtering for precoding has only been considered by a few authors [5] in comparison with other criteria for precoding, it is a very powerful transmit optimization that minimizes the Mean Square Error (MSE) with a transmit energy constraint [6, 7, 8, 2], i.e.

$$\{\boldsymbol{F}_{\mathrm{WF}}, g_{\mathrm{WF}}\} = \underset{\{\boldsymbol{F},g\}}{\operatorname{argmin}}\, \mathrm{E}\left[\|\boldsymbol{u}[n] - \hat{\boldsymbol{u}}[n]\|_2^2\right], \text{ s.t.: } \operatorname{tr}(\boldsymbol{F}\boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{F}^{\mathrm{H}}) \leq E_{\mathrm{tx}}, \tag{4}$$

where $\boldsymbol{C}_{\boldsymbol{u}} = \mathrm{E}\left[\boldsymbol{u}[n]\boldsymbol{u}^{\mathrm{H}}[n]\right]$. It has been demonstrated in [5] that (4) leads to a unique solution if we restrict g to be positive real. Then, the solution for the Wiener filter is given by

$$\boldsymbol{F}_{\mathrm{WF}} = g_{\mathrm{WF}}^{-1}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \xi\mathbf{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}},\; g_{\mathrm{WF}} = \sqrt{\frac{\operatorname{tr}\left(\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}+\xi\mathbf{I}\right)^{-2}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{H}\right)}{E_{\mathrm{tx}}}},\; \xi = \operatorname{tr}\left(\frac{\boldsymbol{C}_\eta}{E_{\mathrm{tx}}}\right) \tag{5}$$

3 Adaptive Algorithms

The model explained in Section 2 states that the observations are linear and instantaneous mixtures of the transmitted signals $\boldsymbol{x}[n]$ of (2). For the case of the linear precoder described in the previous section, this equation can be rewritten as follows

$$\boldsymbol{y}[n] = \boldsymbol{HFu}[n] + \eta[n]. \tag{6}$$

This means that the observations $\boldsymbol{y}[n]$ are instantaneous mixtures of the data symbols $\boldsymbol{u}[n]$, where the mixing matrix is given by $\boldsymbol{HF}$. In the sequel, we will denote this mixing matrix as $\boldsymbol{A}$, so the observations $\boldsymbol{y}[n]$ can be obtained as

$$\boldsymbol{y}[n] = \boldsymbol{Ad}[n] + \eta[n]. \tag{7}$$

According to our target, $\boldsymbol{A}$ may represent the channel matrix (see (2)), or the whole coding–channel matrix $\boldsymbol{HF}$ (see (6)). In the first case, $\boldsymbol{d}[n]$ represents the coded signal $\boldsymbol{x}[n] = \boldsymbol{Fu}[n]$ and, in the second case, the user data signal $\boldsymbol{u}[n]$. We assume that the mixing matrix is unknown but full rank. Without any loss of generality, we can suppose that the source data have a normalized power equal to one since possible differences in power may be included into the mixing matrix $\boldsymbol{A}$.

In order to recover the source data, we will use a linear system of which output is a combination of the observations, expressed as

$$z[n] = \boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n], \boldsymbol{W} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{r}}}, \tag{8}$$

by combining both (7) and (8), the output $\boldsymbol{z}[n]$ can be rewritten as a linear combination of the desired signal

$$\boldsymbol{z}[n] = \boldsymbol{\Gamma}[n]\boldsymbol{d}[n], \tag{9}$$

where $\boldsymbol{\Gamma}[n] = \boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{A}$ represents the overall mixing/separating system. Sources are optimally recovered when the matrix $\boldsymbol{W}[n]$ is selected such as every output extracts a different single source. This occurs when the matrix $\boldsymbol{\Gamma}[n]$ has the form

$$\boldsymbol{\Gamma}[n] = \boldsymbol{D}[n]\boldsymbol{P}[n], \tag{10}$$

where $\boldsymbol{D}[n]$ is a diagonal invertible matrix, and $\boldsymbol{P}[\boldsymbol{n}]$ is a permutation matrix.

In this paper, we consider two types of Neural Network paradigms: supervised and unsupervised approaches.

Supervised Approach. A way to estimate the channel matrix, **H**, consists in minimizing the Mean Square Error MSE between the outputs $\boldsymbol{y}[n]$ and the code signals $\boldsymbol{x}[n]$. In particular, by considering only one sample, we obtain the Least Mean Squares (LMS) algorithm,

$$\boldsymbol{W}[n+1] = \boldsymbol{W}[n] - \mu\boldsymbol{y}[n](\boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n] - \boldsymbol{d}[n])^{\mathrm{H}}.$$

This algorithm is also called *delta rule* of *Widrow-Hopf* [9] in the context of Artificial Neural Networks. It is easy to prove that the stationary points of this rule are

$$\boldsymbol{W}[n] = \boldsymbol{C_y}^{-1}\boldsymbol{C_{yd}}, \tag{11}$$

where $\boldsymbol{C_y} = \mathrm{E}[\boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n]]$ is the autocorrelation of the observations and $\boldsymbol{C_{yd}} = \mathrm{E}[\boldsymbol{y}[n]\boldsymbol{d}^{\mathrm{H}}[n]]$ is the cross–correlation between the observations and the desired signals. In practice, the desired signal is considered known only during a finite number of instants (pilot symbols) and the expectations are estimated by averaging samples.

Unsupervised Approach. The inclusion of pilot symbols reduces the system throughput (or equivalently, the spectral efficiency of the system) and wastes transmission energy because pilot sequences do not convey user data. This limitation can be avoided by using Blind Source Separation (BSS) algorithms, which simultaneously estimate the mixing matrix $\boldsymbol{A}$, and the realizations of the source vector $\boldsymbol{u}[n]$ from the corresponding realizations of the observations $\boldsymbol{y}[n]$.

One of the best known BSS algorithms has been proposed by Bell and Sejnowski in [3]. Given an activation function $h(\cdot)$, the idea is to obtain the weighted coefficients of a Neural Network $\boldsymbol{W}[n]$ in order to maximize the mutual information between the outputs before the activation function $\boldsymbol{h}(\boldsymbol{z}[n]) = \boldsymbol{h}(\boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n])$, and its inputs $\boldsymbol{y}[n]$, which is given by

$$J_{\mathrm{MI}}(\boldsymbol{W}[n]) = \ln(\det(\boldsymbol{W}^{\mathrm{H}}[n])) + \sum_{i=1}^{N_{\mathrm{B}}} \mathrm{E}[\ln(h_i'(z_i[n]))], \tag{12}$$

where h_i is the i–th element of the vector $\boldsymbol{h}(\boldsymbol{z}[n])$ and $'$ denotes the first derivative. The resulting algorithm, named *Infomax*, has the following form

$$\begin{aligned}\boldsymbol{W}[n+1] &= \boldsymbol{W}[n] + \mu\boldsymbol{W}[n]\boldsymbol{W}^{\mathrm{H}}[n]\cdot\left(\boldsymbol{y}[n]\ \boldsymbol{g}^{\mathrm{H}}(\boldsymbol{z}[n]) - \boldsymbol{W}^{-\mathrm{H}}[n]\right)\\ &= \boldsymbol{W}[n] + \mu\boldsymbol{W}[n]\left(\boldsymbol{z}[n]\boldsymbol{g}^{\mathrm{H}}(\boldsymbol{z}[n]) - \boldsymbol{I}\right).\end{aligned} \tag{13}$$

The expression in (12) admits an interesting interpretation when the non–linear function $g(z) = z^*(1 - |z|^2)$ is utilized. In this case, Castedo and Macchi [12] have shown that the Bell and Sejnowski rules are equivalent to the *Constant Modulus Algorithm* (CMA) proposed by Godard in [13].

4 Hybrid Approach

One of advantages of adaptive unsupervised algorithms is their ability to track low variations of the channel. On the contrary, supervised solutions provide a fast channel estimation for low or high variations at the cost of using pilot symbols. In this section, we combine this two paradigms in order to obtain a performance similar to that offered by supervised approaches, but using lower number of pilot symbols.

We will denote by $\boldsymbol{W}_u[n]$ and $\boldsymbol{W}_s[n]$ the matrices for the unsupervised and the supervised modules, respectively. We start with an initial estimation of the channel matrix obtained using the Widrow-Hopf solution (11). This estimation is used at the TX in order to obtain the optimum coding matrix $\boldsymbol{F}$, and at the RX with the goal of initializing the unsupervised algorithm to $\boldsymbol{W}_u[n] = (\boldsymbol{FH})^{-\mathrm{H}}$.

While the channel does not suffer from a significant variation, the matrix $\boldsymbol{W}_u[n]$ is adapted (unsupervised mode) and the data symbols $\boldsymbol{u}[n]$ are recovered using $\boldsymbol{z}[n] = \boldsymbol{W}_u^{\mathrm{H}}[n]\boldsymbol{y}[n]$. However, when a significant variation is detected, the RX sends an *alarm* to the TX through the feedback channel. Next, a pilot sequence is transmitted. Then, at the RX a supervised algorithm estimates the channel from the pilot symbols (channel estimation update). In particular, we make use of Widrow-Hopf solution (11) by considering that $\boldsymbol{u}[n]$ are the coded signals at the output of the linear precoder. This solution provides us the channel matrix estimation, which is sent to the TX in order to adapt the coding matrix. The RX also computes the coding matrix $\boldsymbol{F}$, the reference matrix $\hat{\boldsymbol{H}}\boldsymbol{F}$, and initializes the unsupervised algorithms as $\boldsymbol{W}_u[n] = \hat{\boldsymbol{H}}\boldsymbol{F}^{-1}$.

The question now is how to determine when the channel has suffered a significant change. An interesting consequence of using a linear precoder is that the permutation indeterminacy (see (10)) associated to unsupervised algorithms is avoided because of the following initialization $\boldsymbol{W}_u[n] = (\boldsymbol{F}\boldsymbol{H})^{-\mathrm{H}}$. This means that the sources are recovered in the same order as they were transmitted. (10) implies that the optimum separation matrix produces a diagonal matrix $\boldsymbol{\Gamma}[n]$, and therefore, the mismatch of $\boldsymbol{\Gamma}[n]$ with respect to a diagonal matrix allows us to measure the variations of the channel.

Although the channel matrix is unknown, we can use the estimation $\hat{\boldsymbol{H}}\boldsymbol{F}$ computed by the supervised approach as a reference. This means that in each iteration we can compute $\boldsymbol{\Gamma}[n] = \boldsymbol{W}_u^{\mathrm{H}}[n]\hat{\boldsymbol{H}}\boldsymbol{F}$. Consequently, the difference with respect to a diagonal matrix can be obtained using the following error criterion

$$\mathrm{Error}(n) = \sum_{i=1}^{N_t} \sum_{j=1, j\neq i}^{N_t} \left(\frac{|\gamma_{ij}[n]|^2}{|\gamma_{ii}[n]|^2} + \frac{|\gamma_{ji}[n]|^2}{|\gamma_{ii}[n]|^2} \right), \tag{14}$$

where $\gamma_{ii}[n]$ denotes the i–th element of its diagonal. The decision rule consists in comparing with some threshold t, i.e.

$$\begin{cases} \mathrm{Error}(n) > t \rightarrow \text{Use supervised approach} \\ \mathrm{Error}(n) \leq t \rightarrow \text{Use unsupervised approach} \end{cases} \tag{15}$$

5 Experimental Results

We evaluate the performance of the proposed combined schemes by simulations. We transmit 8 000 pixels of the image *cameraman* (in TIF format with 256 gray levels) using a QPSK and a 4×4 MIMO system. The channel matrix is updated each 2 000 symbols using the following model: $\boldsymbol{H} = (1-\alpha)\boldsymbol{H} + \alpha\boldsymbol{H}_{\mathrm{new}}$, where $\boldsymbol{H}_{\mathrm{new}}$ is a 4×4 matrix randomly generated according to a Gaussian distribution. The SNR has been fixed to 20 dB.

We compare the performance of three different schemes (see Figure 2): the Widrow-Hopf solution (11) with 200 pilot symbols transmitted every 2 000 symbols (supervised approach); the Infomax algorithm (13) with the non linear

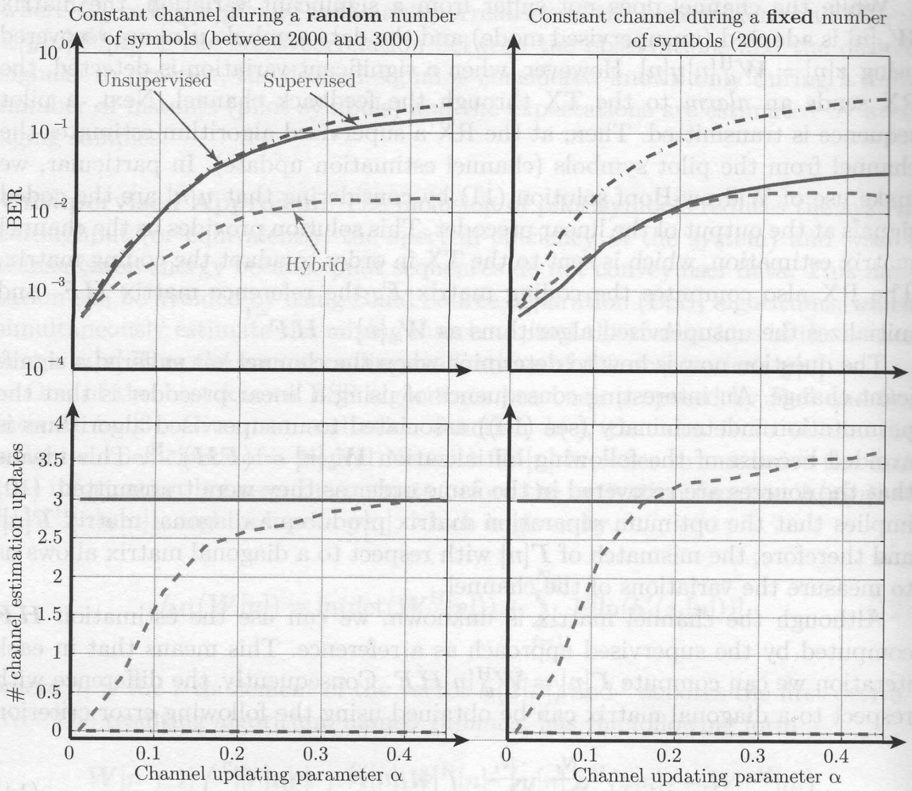

Fig. 2. Performance results (see Section 5)

function $g(z) = z^*(1 - |z|^2)$, and $\mu = 0.001$ (unsupervised approach); and the hybrid approach with a threshold $t = 0.7$. The left-hand side of Figure 2 shows the results when the channel remains constant during a random number of symbols between 2 000 and 3 000. The right-hand side of Figure 2 plots the results when the channel remains constant during $N_B = 2\,000$ symbols. The top side of Figure 2 shows the Bit Error Ratio (BER) obtained for all approaches. The bottom side shows the number of times the mixing matrix has been estimated, and updated, using the supervised approach. Comparing the curves in Figure 2, we observe that the BER offered by the hybrid system is invariant to the number of symbols in which the channel remains constant.

6 Conclusion

In order to reduce the overhead due to the transmission of pilot symbols we have proposed to combine supervised and unsupervised algorithms. The algorithm selection was done by using a simple decision rule to determine a significant variation in the channel. This information was sent to the TX using a limited feedback channel. The experimental results showed that the hybrid approach is an attractive solution because it provides an adequate BER with a reduced number of pilot symbols.

Acknowledgment

This work been supported by Xunta de Galicia, Ministerio de Ciencia e Innovación of Spain, and FEDER funds under the grants 09TIC008105PR, TEC2007-68020-C04-01, CSD2008-00010, and TIN2009-05736-E/TIN.

References

[1] Fischer, R.F.H.: Precoding and Signal Shaping for Digital Transmission. John Wiley & Sons, Chichester (2002)
[2] Joham, M.: Optimization of Linear and Nonlinear Transmit Signal Processing. PhD dissertation, Munich University of Technology (2004)
[3] Bell, A., Sejnowski, T.: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation 7(6), 1129–1159 (1995)
[4] Hunger, R., Joham, M., Utschick, W.: Extension of Linear and Nonlinear Transmit Filters for Decentralized Receivers. In: European Wireless 2005, vol. 1, pp. 40–46 (2005)
[5] Joham, M., Kusume, M., Gzara, M.H., Utschick, W., Nossek, J.A.: Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems. In: Proc. ISSSTA, vol. 1, pp. 9–13 (2002)
[6] Choi, R.L., Murch, R.D.: New Transmit Schemes and Simplified Receiver for MIMO Wireless Communication Systems. IEEE Transactions on Wireless Communications 2(6), 1217–1230 (2003)
[7] Karimi, H.R., Sandell, M., Salz, J.: Comparison between Transmitter and Receiver Array Processing to Achieve Interference Nulling and Diversity. In: Proc. PIMRC 1999, vol. 3, pp. 997–1001 (1999)
[8] Nossek, J.A., Joham, M., Utschick, W.: Transmit Processing in MIMO Wireless Systems. In: Proc. 6th IEEE Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication, Shanghai, pp. 1–18 (2004)
[9] Haykin, S.: Neural Networks. A Comprehensive Foundation. Macmillan College Publishing Company, New York (1994)
[10] Amari, S.-I.: Gradient Learning in Structured Parameter Spaces: Adaptive Blind Separation of Signal Sources. In: Proc. WCNN 1996, pp. 951–956 (1996)
[11] Mejuto, C., Castedo, L.: A Neural Network Approach to Blind Source Separation. In: Proc. Neural Networks for Signal Processing VII, pp. 486–595 (1997)
[12] Castedo, L., Macchi, O.: Maximizing the Information Transfer for Adaptive Unsupervised Source Separation. In: Proc. SPAWC 1997, pp. 65–68 (1997)
[13] Godard, D.N.: Self-Recovering Equalization Carrier Tracking in Two-Dimensional DataCommunications Systems. IEEE Transactions on Communications (1980)