MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Order-Extended Sparse RLS Algorithm for Doubly-Selective MIMO Channel Estimation

Koike-Akino, T.; Molisch, A.F.; Annnavajjala, R.;, Orlik, P.; Pun, M-O.

TR2011-030 June 2011

Abstract

We develop a recursive least squares (RLS) algorithm which employs L1-Lq regularized sparse regressions to estimate a sparse channel matrix in frequency time selective fading for multi output (MIMO) wireless communications. We propose an improved sparse RLS by using an order extension technique for rapid fading channels. Simulation results demonstrate that the proposed sparse RLS algorithm offers a significant improvement over the conventional RLS algorithm.

IEEE International Conference on Communications (ICC)

Order–Extended Sparse RLS Algorithm for Doubly–Selective MIMO Channel Estimation

Toshiaki Koike-Akino†, Andreas F. Molisch‡, Ramesh Annavajjala†, Philip Orlik†, and Man-On Pun†
† Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, U.S.A.
‡ Depart. of Electrical Engineering, University of Southern California, 3740 McClintock Av., Los Angeles, CA 90089, U.S.A.
Email: {koike, annavajjala, porlik}@merl.com, molisch@usc.edu

***Abstract*—We develop a recursive least–squares (RLS) algorithm which employs $\mathcal{L}_1$–$\mathcal{L}_q$ regularized sparse regressions to estimate a sparse channel matrix in frequency–and–time selective fading for multi–input multi–output (MIMO) wireless communications. We propose an improved sparse RLS by using an order extension technique for rapid fading channels. Simulation results demonstrate that the proposed sparse RLS algorithm offers a significant improvement over the conventional RLS algorithm.**

I. Introduction

There has emerged a lot of focuses on sparse signal regressions among the researchers in the fields of signal processing and information theory (e.g., [1–6]). Although those works contain theoretical fundamentals, most of the algorithms are not tailored to time–varying environments with real–time requirements. Recently, Bajwa *et al.* [2] used the Dantzig selector [3] and least–squares (LS) for sparse channel sensing. Choi *et al.* [7] investigated an expectation–maximization (EM) method for doubly–selective multi–input multi–output (MIMO) sparse channels. Although those methods offer good estimates with improved mean–square error (MSE), the underlying sparsity is not fully exploited to reduce the complexity.

In [8, 9], an $\mathcal{L}_1$–regularized recursive least–squares (RLS) algorithm was introduced for adaptive filtering. The sparse RLS algorithm is based on an EM algorithm proposed in [10]. It was shown that the sparse RLS algorithm significantly outperforms the conventional RLS algorithm both in terms of MSE and computational complexity for time–varying sparse channels. In this paper, we extend it to MIMO systems. In addition, we improve the algorithm by introducing order extension techniques (or, basis expansion model) [11–13] to enhance the ability of channel tacking. For order extensions, we use an $\mathcal{L}_1$–$\mathcal{L}_q$ regularized sparse regressions [14, 15]. The order extension technique enjoys a significant gain in high SNR regimes and for fast fading, whereas it can degrade in low SNRs and for slow fading in general. The key idea behind the use of the $\mathcal{L}_1$–$\mathcal{L}_q$ regularized sparse regressions for high–order estimation lies in the fact that the higher–order channel matrix becomes highly sparse. Therefore, we can expect that the sparse regression can automatically deal with the drawback of the high–order estimation schemes in low SNR regimes. Through computer simulations, we demonstrate that the proposed sparse high–order RLS algorithm significantly outperforms conventional schemes in the sense of MSE and achievable data rate.

Notations: Throughout the paper, we describe matrices and vectors by bold–face italic letters in capital cases and small cases, respectively. Let $\boldsymbol{X} \in \mathbb{C}^{m\times n}$ be a complex–valued $(m \times n)$–dimensional matrix, where $\mathbb{C}$ denotes the complex field. The notations $\boldsymbol{X}^*$, $\boldsymbol{X}^{\mathrm{T}}$, $\boldsymbol{X}^{\dagger}$, $\boldsymbol{X}^{-1}$, $\mathrm{tr}[\boldsymbol{X}]$, $\det[\boldsymbol{X}]$, and $\|\boldsymbol{X}\|_{\mathrm{F}}$ represent the complex conjugate, the transpose, the Hermite transpose, the inverse, the trace, the determinant, and the Frobenius norm of $\boldsymbol{X}$, respectively. The operator $\mathrm{vec}[\boldsymbol{X}]$ denotes the vector–operation which stacks all columns of $\boldsymbol{X}$ into a single column vector in a left–to–right fashion, and the operator $\otimes$ stands for the Kronecker product. The set of real numbers and integers are denoted by $\mathbb{R}$ and $\mathbb{N}$. The non–negative sets in real numbers and integers are $\mathbb{R}_+$ and $\mathbb{N}_+$. A positive ring from 1 to m is represented by $\mathbb{N}_m$. A matrix $\boldsymbol{I}_m$ denotes an m–dimensional identity matrix. The $\mathcal{L}_q$ vector norm is defined as $\|\boldsymbol{x}\|_q = \left(\sum_i |x_i|^q\right)^{1/q}$, the $\mathcal{L}_0$–quasi vector norm $\|\boldsymbol{x}\|_0$ is the number of non–zero entries, and $\mathcal{L}_\infty$ vector norm is given as $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$. A multivariate complex–valued Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. A positive operator is written by $(r)_+ \triangleq \max(0, r)$ for a real value $r \in \mathbb{R}$. The expectation is represented by $\mathbb{E}[\cdot]$.

II. MIMO Channel Estimation Scheme

A. Channel Model

We consider $M \times N$ MIMO systems in which M antennas are used at the transmitter and N antennas are used at the receiver. Let $x_m(k) \in \mathbb{C}$ be a transmitting signal from the m–th antenna ($m \in \mathbb{N}_M$) at the k–th symbol instance. We assume that no signal is transmitted before the symbol index $k = 1$, i.e., $x_m(k) = 0$ for any $k < 1$, and that the average power of the transmission signal is normalized as $\mathbb{E}[|x_m(k)|^2] = 1$. The frequency–selective channel is modeled by a sample–space tapped delay line. The p–th tap is denoted by $h_{n,m,p}(k) \in \mathbb{C}$ for the channel between the m–th transmitting antenna and the n–th receiving antenna at the k–th time instance. We suppose that the maximum delay in symbol is at most P.

The receiving signal at the k–th symbol is modeled as

$$y_n(k) = \sum_{p=0}^{P}\sum_{m=1}^{M} h_{n,m,p}(k)x_m(k-p) + z_n(k), \qquad (1)$$

with $y_n(k) \in \mathbb{C}$ and $z_n(k) \in \mathbb{C}$ being the receiving signal and the additive noise at the n–th receiving antenna ($n \in \mathbb{N}_N$). When we define

$$\boldsymbol{y}(k) \triangleq \begin{bmatrix} y_1(k) & \cdots & y_N(k) \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}, \quad (2)$$

$$\boldsymbol{H}_p(k) \triangleq \begin{bmatrix} h_{1,1,p}(k) & \cdots & h_{1,M,p}(k) \\ \vdots & & \vdots \\ h_{N,1,p}(k) & \cdots & h_{N,M,p}(k) \end{bmatrix} \in \mathbb{C}^{N\times M}, \quad (3)$$

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix} x_1(k) & \cdots & x_M(k) \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{M\times 1}, \quad (4)$$

$$\boldsymbol{z}(k) \triangleq \begin{bmatrix} z_1(k) & \cdots & z_N(k) \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}, \quad (5)$$

$$\boldsymbol{H}(k) \triangleq \begin{bmatrix} \boldsymbol{H}_0(k) & \cdots & \boldsymbol{H}_P(k) \end{bmatrix} \in \mathbb{C}^{N\times M'}, \quad (6)$$

$$\boldsymbol{\chi}(k) \triangleq \begin{bmatrix} \boldsymbol{x}^{\mathrm{T}}(k) & \cdots & \boldsymbol{x}^{\mathrm{T}}(k-P) \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{M'\times 1}, \quad (7)$$

with $M' = M(P+1)$, we can rewrite the receiving signals:

$$\begin{aligned} \boldsymbol{y}(k) &= \sum_{p=0}^{P} \boldsymbol{H}_p(k)\boldsymbol{x}(k-p) + \boldsymbol{z}(k) \\ &= \boldsymbol{H}(k)\boldsymbol{\chi}(k) + \boldsymbol{z}(k). \end{aligned} \quad (8)$$

The vector $\boldsymbol{\chi}(k)$ stands for the transmitted signal which stacks the past P symbols as well as the k–th signal $\boldsymbol{x}(k)$. Upon the time instance k, we wish to estimate the time–varying channel matrix $\boldsymbol{H}(k)$ by using the known transmitted signal $\boldsymbol{x}(l)$ and the received signal $\boldsymbol{y}(l)$ obtained in the past for any $l \in \mathbb{N}_k$.

B. Conventional RLS Algorithm

Here, we first review one of the most widely used channel estimation algorithms; a recursive least–squares (RLS) [16]. In the RLS algorithm, an exponentially–weighted MSE $\mathcal{E}(k)$ is minimized as follows

$$\begin{aligned} \min_{\hat{\boldsymbol{H}}(k)} \quad & \left\{ \mathcal{E}(k) \triangleq \sum_{l=1}^{k} \lambda^{k-l} \left\| \boldsymbol{y}(l) - \hat{\boldsymbol{H}}(k)\boldsymbol{\chi}(l) \right\|_2^2 \right. \\ & \left. = \mathrm{tr}\Big[\big(\boldsymbol{Y}_k - \hat{\boldsymbol{H}}(k)\boldsymbol{X}_k\big)\boldsymbol{\Lambda}_k\big(\boldsymbol{Y}_k - \hat{\boldsymbol{H}}(k)\boldsymbol{X}_k\big)^{\dagger} \Big] \right\}, \end{aligned} \quad (9)$$

where we defined

$$\boldsymbol{Y}_k \triangleq \begin{bmatrix} \boldsymbol{y}(1) & \cdots & \boldsymbol{y}(k) \end{bmatrix} \in \mathbb{C}^{N\times k}, \quad (10)$$

$$\boldsymbol{\Lambda}_k \triangleq \mathrm{diag}\begin{bmatrix} \lambda^{k-1} & \cdots & \lambda^0 \end{bmatrix} \in \mathbb{R}^{k\times k}, \quad (11)$$

$$\boldsymbol{X}_k \triangleq \begin{bmatrix} \boldsymbol{\chi}(1) & \cdots & \boldsymbol{\chi}(k) \end{bmatrix} \in \mathbb{C}^{M'\times k}. \quad (12)$$

The parameter $\lambda \in \mathbb{R}_+$ is termed forgetting factor which controls the tradeoff between the tracking ability and the noise tolerance in time–varying channels. The forgetting factor should be adjusted according to the channel condition; typically, we use $\lambda \simeq 0.98$. In [17], a near–optimal forgetting factor has been reported for frequency–flat Rayleigh fading channels in single antenna systems; $\lambda_{\mathrm{opt}} \simeq 1 - \big(2(2\pi f_{\mathrm{D}} T_{\mathrm{s}})^2/\sigma^2\big)^{1/3}$, where f_{D}, T_{s} and σ^2 denote the maximum Doppler frequency, the symbol duration, and the noise variance, respectively.

The LS solution at the k–th sample is given as

$$\hat{\boldsymbol{H}}(k) = \boldsymbol{Y}_k\boldsymbol{\Lambda}\boldsymbol{X}_k^{\dagger} \underbrace{\left(\boldsymbol{X}_k\boldsymbol{\Lambda}\boldsymbol{X}_k^{\dagger}\right)^{-1}}_{\boldsymbol{\Phi}_k \in \mathbb{C}^{M'\times M'}}. \quad (13)$$

The exponential weighting enables a low–complexity implementation with rank–one update instead of computing a direct matrix inverse as follows:

$$\boldsymbol{\zeta}_k = \frac{1}{\lambda}\boldsymbol{\Phi}_{k-1}\boldsymbol{\chi}(k), \quad \boldsymbol{e}_k = \boldsymbol{y}(k) - \hat{\boldsymbol{H}}(k-1)\boldsymbol{\chi}(k), \quad (14)$$

$$\boldsymbol{\Phi}_k = \frac{1}{\lambda}\boldsymbol{\Phi}_{k-1} - \frac{\boldsymbol{\zeta}_k\boldsymbol{\zeta}_k^{\dagger}}{1+\boldsymbol{\zeta}_k^{\dagger}\boldsymbol{\chi}(k)}, \quad (15)$$

$$\hat{\boldsymbol{H}}(k) = \hat{\boldsymbol{H}}(k-1) + \frac{\boldsymbol{e}_k\boldsymbol{\zeta}_k^{\dagger}}{1+\boldsymbol{\zeta}_k^{\dagger}\boldsymbol{\chi}(k)}. \quad (16)$$

It reduces the arithmetic complexity from $\mathcal{O}[M'^3]$ to $\mathcal{O}[M'^2]$.

C. Extended–Order RLS Algorithm

In [12], the optimally–weighted LS channel estimation and the high–order RLS are studied. In general, the conventional RLS algorithm can suffer from a severe performance degradation in rapid fading channels even if we use an optimized forgetting factor. To deal with channel variation, high–order RLS algorithm which extends a polynomial order for regressions was introduced. It offers a significant performance improvement in rapid fading and high SNR regimes. The high–order estimation scheme is extended for non–coherent MIMO communications with Grassmann space–time coding in [13].

Here, we describe the high–order RLS channel estimation algorithm, in which we model the channel variation in the past samples (for $l \in \mathbb{N}_k$ at the k–th estimation symbol) by higher–order polynomial as follows:

$$\boldsymbol{H}(l) = \sum_{d=0}^{D} l^d \hat{\boldsymbol{H}}^{[d]}(k) = \hat{\boldsymbol{\mathcal{H}}}(k)\boldsymbol{\mathcal{D}}(l), \quad (17)$$

where the order–extended channel matrix $\hat{\boldsymbol{\mathcal{H}}}(k)$ and the order extension matrix $\boldsymbol{\mathcal{D}}(l)$ are defined as

$$\hat{\boldsymbol{\mathcal{H}}}(k) \triangleq \begin{bmatrix} \hat{\boldsymbol{H}}^{[0]}(k) & \cdots & \hat{\boldsymbol{H}}^{[D]}(k) \end{bmatrix} \in \mathbb{C}^{N\times M'(D+1)}, \quad (18)$$

$$\boldsymbol{\mathcal{D}}(l) \triangleq \begin{bmatrix} l^0\boldsymbol{I}_{M'} & \cdots & l^D\boldsymbol{I}_{M'} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{M'(D+1)\times M'}. \quad (19)$$

The channel matrix $\hat{\boldsymbol{H}}^{[d]}(k) \in \mathbb{C}^{N\times M'}$ corresponds to the d–th order term of the time–varying channel polynomial model. We assume that the order extension is up to $D \in \mathbb{N}_+$ which should be optimized according to the channel condition.

With the aforementioned channel model, the optimization problem for the high–order channel estimation is reformulated:

$$\begin{aligned} \min_{\hat{\boldsymbol{\mathcal{H}}}(k)} \quad & \left\{ \mathcal{E}'(k) \triangleq \sum_{l=1}^{k} \lambda^{k-l} \left\| \boldsymbol{y}(l) - \hat{\boldsymbol{\mathcal{H}}}(k)\boldsymbol{\mathcal{D}}(l)\boldsymbol{\chi}(l) \right\|_2^2 \right. \\ & \left. = \mathrm{tr}\Big[\big(\boldsymbol{Y}_k - \hat{\boldsymbol{\mathcal{H}}}(k)\boldsymbol{\mathcal{X}}_k\big)\boldsymbol{\Lambda}_k\big(\boldsymbol{Y}_k - \hat{\boldsymbol{\mathcal{H}}}(k)\boldsymbol{\mathcal{X}}_k\big)^{\dagger} \Big] \right\}, \end{aligned} \quad (20)$$

where

$$\boldsymbol{\mathcal{X}}_k \triangleq \begin{bmatrix} \boldsymbol{\mathcal{D}}(1)\boldsymbol{\chi}(1) & \cdots & \boldsymbol{\mathcal{D}}(k)\boldsymbol{\chi}(k) \end{bmatrix} \in \mathbb{C}^{M'(D+1)\times k}. \quad (21)$$

Hence, substituting $\boldsymbol{\mathcal{D}}(k)\boldsymbol{\chi}(k)$ for $\boldsymbol{\chi}(k)$ and $\hat{\boldsymbol{\mathcal{H}}}(k)$ for $\hat{\boldsymbol{H}}(k)$ in (15) and (16), we can implement the high–order estimation in a recursive manner. With an RLS estimation of high–order channel $\hat{\boldsymbol{\mathcal{H}}}(k)$, the desired estimate of the MIMO channel is obtained as $\hat{\boldsymbol{H}}(k) = \hat{\boldsymbol{\mathcal{H}}}(k)\boldsymbol{\mathcal{D}}(k)$.

The higher–order RLS channel estimation offers more accurate estimates for fast fading channels in high SNR regimes, while it can degrade the performance in low SNRs and for slow fading channels due to over–fitting loss. In this paper, we introduce $\mathcal{L}_1$–$\mathcal{L}_q$ regularized regression scheme to deal with the over–fitting loss. The key idea lies in the fact that the channel gains of higher–order terms tend to be small in slow fading and the order–extended channel $\hat{\boldsymbol{\mathcal{H}}}(k)$ becomes sparse.

III. $\mathcal{L}_1$–$\mathcal{L}_q$ Regularized Sparse High–Order RLS

In this section, we first introduce $\mathcal{L}_1$–regularized sparse RLS algorithm, proposed in [8, 9], which is based on the lasso sparse matrix regression [19, 20]. We then generalize it to the group lasso sparse regression [14, 15], i.e., $\mathcal{L}_1$–$\mathcal{L}_q$ regularized RLS algorithm, for order–extended channel estimations.

A. Sparse Channel Matrix Regression

The frequency–selective wireless channel usually has a few significant multi–path components, that leads to a sparse matrix for $\boldsymbol{H}(k) \in \mathbb{C}^{N\times M'}$. More importantly, the tap position for those principal paths is not known prior to the estimation in general. A joint estimation of tap positions and tap weights can be performed by novel approaches of sparse matrix regression techniques in compressive sensing, such as $\mathcal{L}_0$ matching pursuit, $\mathcal{L}_1$ lasso [19, 20], $\mathcal{L}_1$–$\mathcal{L}_2$ group lasso [21], and $\mathcal{L}_1$–$\mathcal{L}_\infty$ logistic regression methods [14].

Suppose that there are few non–zero taps in the channel, i.e., $\big\|\mathrm{vec}[\boldsymbol{H}(k)]\big\|_0 = NMQ \ll NM(P+1)$. A sparse regression is obtained by solving the following optimization problem:

$$\min_{\hat{\boldsymbol{H}}(k)} \quad \big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}(k)\big]\big\|_0, \qquad \text{s.t.} \quad \mathcal{E}(k) \le \epsilon, \tag{22}$$

where ϵ is a positive constant controlling the allowable estimation error level. The above optimization problem is computationally intractable due to its non–convexity. A convex relaxation provides a viable alternative, whereby the $\mathcal{L}_0$ quasi–norm, $\big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}(k)\big]\big\|_0$, is replaced by the $\mathcal{L}_1$ norm, $\big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}(k)\big]\big\|_1$. The Lagrangian formulation shows that the solution can be equivalently derived from the problem:

$$\min_{\hat{\boldsymbol{H}}(k)} \quad \frac{1}{\sigma^2}\mathcal{E}(k) + \gamma\big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}(k)\big]\big\|_1, \tag{23}$$

where a Lagrangian multiplier $\gamma \in \mathbb{R}_+$ represents a tradeoff between error and sparsity. The $\mathcal{L}_1$–regularized regression can be solved in a computationally efficient way as in [22]. Note that the transmitting signal $\boldsymbol{x}(l)$ should be properly chosen so that there exists a unique solution [2, 18].

B. $\mathcal{L}_1$–Regularized Sparse RLS Algorithm

In [8, 9], $\mathcal{L}_1$–regularized sparse RLS regression (termed SPARLS) was introduced for the application of the zero–th order RLS channel estimation in single–antenna systems. After reviewing SPARLS with MIMO extensions in this section, we will further extend it to high–order RLS algorithm by means of $\mathcal{L}_1$–$\mathcal{L}_q$ regularized estimation method.

In appendix, we present an algorithm derivation of SPARLS with a slight modification from [8–10] for MIMO channels. The algorithm is summarized below

Initialize $\boldsymbol{A}(0) = \boldsymbol{0}$, $\boldsymbol{B}(0) = \boldsymbol{0}$, $\hat{\boldsymbol{H}}(0) = \boldsymbol{0}$
for all input $\boldsymbol{\chi}(k)$ and observation $\boldsymbol{y}(k)$ **do**
  $\boldsymbol{A}(k) = \lambda\boldsymbol{A}(k-1) + \frac{\alpha^2}{\sigma^2}\boldsymbol{\chi}(k)\boldsymbol{\chi}^\dagger(k)$
  $\boldsymbol{B}(k) = \lambda\boldsymbol{B}(k-1) + \frac{\alpha^2}{\sigma^2}\boldsymbol{y}(k)\boldsymbol{\chi}^\dagger(k)$
  repeat
    $\hat{\boldsymbol{G}}(k) = \hat{\boldsymbol{H}}(k)\big(\boldsymbol{I}_{M'} - \boldsymbol{A}(k)\big) + \boldsymbol{B}(k)$
    $\hat{\boldsymbol{H}}(k) = \mathcal{F}_{\mathrm{th}}\big(\hat{\boldsymbol{G}}(k), \alpha^2\gamma\big)$
  until $\hat{\boldsymbol{H}}(k)$ converges
end for

The function $\mathcal{F}_{\mathrm{th}}(x, \beta)$ for $x \in \mathbb{C}$ and $\beta \in \mathbb{R}_+$ is referred to as a *soft–threshold* function [10], which is defined as

$$\mathcal{F}_{\mathrm{th}}(x, \beta) \triangleq x\Big(1 - \frac{\beta}{2|x|}\Big)_+. \tag{24}$$

For a complex–valued matrix argument $\hat{\boldsymbol{G}}(k) \in \mathbb{C}^{N\times M'}$, the function $\mathcal{F}_{\mathrm{th}}(\hat{\boldsymbol{G}}(k), \beta)$ generates element–wise thresholding for each complex–valued entry. The algorithm described above can be further simplified by considering zero–valued entries within thresholding (for more detail description, see [8, 9]). It has been reported that a couple of iterations are sufficient for convergence. The computational complexity becomes $\mathcal{O}[QM']$ which is significantly lower than that of the conventional RLS, $\mathcal{O}[M'^2]$, for a sparse matrix $Q \ll M'$.

C. $\mathcal{L}_1$–$\mathcal{L}_q$ Regularized Sparse High–Order RLS Algorithm

Since the sparsity of the high–order channel matrix $\boldsymbol{\mathcal{H}}(k)$ is thought to be more significant when we extend polynomial order D, the sparse regression becomes more important in high–order RLS estimation even if the original channel matrix $\boldsymbol{H}(k)$ is not sparse. To the best of authors' knowledge, there is no literature investigating the high–order sparse regression. We now propose a high–order RLS algorithm which employs $\mathcal{L}_1$–$\mathcal{L}_q$ regularized sparse matrix regression for order extensions.

We focus on $\mathcal{L}_1$–$\mathcal{L}_q$ regularized regression method for the application of high–order RLS channel estimation as follows:

$$\min_{\hat{\boldsymbol{\mathcal{H}}}(k)} \quad \frac{1}{\sigma^2}\mathcal{E}'(k) + \gamma\big\|\mathrm{vec}\big[\hat{\boldsymbol{\mathcal{H}}}(k)\big]\big\|_1 + \sum_{d=0}^{D}\mu_d\big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}^{[d]}(k)\big]\big\|_q. \tag{25}$$

The first term is the weighted MSE for high–order RLS defined in (20), the next term is an $\mathcal{L}_1$–norm penalty for the matrix sparsity, and the last term denotes a summation of $\mathcal{L}_q$–norm penalties for high–order channel matrices. For the d–th order polynomial, we introduced individual penalty parameters $\mu_d \in \mathbb{R}_+$. If we set $\mu_d = 0$, it reduces to the lasso regression for order–extended sparse RLS.

In appendix, the extension for high–order RLS with $\mathcal{L}_1$–$\mathcal{L}_q$ regularized sparse regressions is described. We summarize the high–order sparse RLS as follows:

Initialize $\boldsymbol{\mathcal{A}}(0) = \boldsymbol{0}$, $\boldsymbol{\mathcal{B}}(0) = \boldsymbol{0}$, $\hat{\boldsymbol{\mathcal{H}}}(0) = \boldsymbol{0}$
for all input $\boldsymbol{\chi}(k)$ and observation $\boldsymbol{y}(k)$ **do**
  $\boldsymbol{\mathcal{A}}(k) = \lambda\boldsymbol{\mathcal{A}}(k-1) + \frac{\alpha^2}{\sigma^2}\boldsymbol{\mathcal{D}}(k)\boldsymbol{\chi}(k)\boldsymbol{\chi}^\dagger(k)\boldsymbol{\mathcal{D}}^\dagger(k)$

$\boldsymbol{\mathcal{B}}(k) = \lambda \boldsymbol{\mathcal{B}}(k-1) + \frac{\alpha^2}{\sigma^2} \boldsymbol{y}(k) \boldsymbol{\chi}^\dagger(k) \boldsymbol{\mathcal{D}}^\dagger(k)$
repeat
  $\hat{\boldsymbol{\mathcal{G}}}(k) = \hat{\boldsymbol{\mathcal{H}}}(k)\big(\boldsymbol{I}_{M'(D+1)} - \boldsymbol{\mathcal{A}}(k)\big) + \boldsymbol{\mathcal{B}}(k)$
  $\hat{\boldsymbol{H}}^{[d]}(k) = \mathcal{F}_{\text{th}}\big(\hat{\boldsymbol{G}}^{[d]}(k), \alpha^2(\gamma + \mu_d \omega_d)\big)$
until $\hat{\boldsymbol{\mathcal{H}}}(k)$ converges
$\hat{\boldsymbol{H}}(k) = \hat{\boldsymbol{\mathcal{H}}}(k) \boldsymbol{\mathcal{D}}(k)$
end for

The parameter $\omega_d \in \mathbb{R}_+$ is dependent on q, as addressed in the appendix.

It should be noted that there are many choices to make a partition for $\hat{\boldsymbol{\mathcal{H}}}(k)$ when we use the group lasso regression. We introduce two strategies; i) *order–domain grouping* which makes $D+1$ groups of $\{\hat{\boldsymbol{H}}_0^{[d]}(k), \ldots, \hat{\boldsymbol{H}}_P^{[d]}(k)\}$ for each order d as in (25) and ii) *tap–domain grouping* which makes $P+1$ groups of $\{\hat{\boldsymbol{H}}_p^{[0]}(k), \ldots, \hat{\boldsymbol{H}}_p^{[D]}(k)\}$ for each tap p. This paper focuses on the order–domain grouping for MSE evaluations.

IV. Performance Evaluations

A. Channel Estimation Error

We evaluate the MSE between the estimated channel matrix $\hat{\boldsymbol{H}}(k)$ and the desired channel matrix $\boldsymbol{H}(k)$:

$$\varepsilon_k^2 \triangleq \frac{1}{N} \mathbb{E}\Big[\big\|\boldsymbol{H}(k) - \hat{\boldsymbol{H}}(k)\big\|_{\text{F}}^2\Big]. \tag{26}$$

We define the average SNR by $\rho \triangleq 1/\sigma^2$, where we assume $\mathbb{E}\big[\boldsymbol{z}(k)\boldsymbol{z}^\dagger(k)\big] = \sigma^2 \boldsymbol{I}_N$, $\mathbb{E}\big[\boldsymbol{x}(k)\boldsymbol{x}^\dagger(k)\big] = \boldsymbol{I}_M$, and $\mathbb{E}\big[\boldsymbol{H}(k)\boldsymbol{H}^\dagger(k)\big] = \boldsymbol{I}_N$. We use $P = 20$ taps for channel estimations. The number of arriving principal paths is $Q = 6$, whose tap positions are randomly selected out of 20 taps. Each principal path has an identical gain on average, and it is generated by 32 equal–gain subpaths based on the Jakes model with the maximum Doppler frequency f_{D}. We consider $M = 2$ transmitting antennas and $N = 2$ receiving antennas. We generate an *i.i.d.* random signal following a Gaussian distribution for the transmitting signal, i.e., $x_m(l) \sim \mathcal{CN}(0, 1)$ for all $l \in \mathbb{N}_+$ and $m \in \mathbb{N}_M$. Note that such a Gaussian signal is not optimal for a training sequence but for an information sequence. The MSE is evaluated at the symbol instance of $k = 128$. We suppose that the receiver knows the transmitted signal for channel tracking. For high–order sparse RLS, we use the $\mathcal{L}_1$–$\mathcal{L}_2$ regularization.

In Fig. 1, we plot the MSE performance as a function of normalized maximum Doppler frequency $f_{\text{D}}T_{\text{s}}$ for an average SNR of $\rho = 40$ dB. For reference, we present the performance of the conventional RLS algorithm using the D–th order polynomial (for $D \in \{0, 1, 2\}$) with and without the genie–aided ideal knowledge of the non–zero tap positions. For each simulation point, the forgetting factor λ is manually optimized to obtain the minimum MSE.

From Fig. 1, it is observed that the MSE performance is severely degraded especially for the conventional zero–th order RLS algorithm as the fading speed ($f_{\text{D}}T_{\text{s}}$) increases. The 1–st order RLS improves MSE for fast fading around $f_{\text{D}}T_{\text{s}} \simeq 0.001$, whereas it has a poor performance in slow fading due to the over–fitting loss. Compared with the genie–aided case,

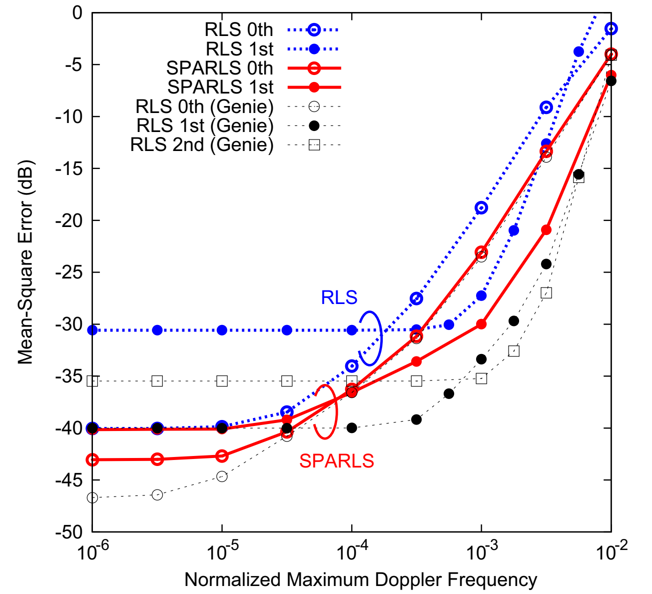

Fig. 1. MSE performance as a function of normalized maximum Doppler frequency $f_{\text{D}}T_{\text{s}}$ for an average SNR of 40 dB (2×2 MIMO, $P = 20$ taps, $Q = 6$ paths).

higher–order RLS is more susceptive to the estimation error caused by the channel sparsity in slow fading; 10 dB and 7 dB loss in MSE are seen for 1–st and 2–nd order RLS algorithms, respectively, at $f_{\text{D}}T_{\text{s}} = 10^{-5}$. We can see that the performance of the zero–th order sparse RLS algorithm is comparable to the ideal performance of the genie–aided zero–th order RLS for fast fading regimes. The 1–st order sparse RLS further improves the MSE in fast fading, and its performance in slow fading is close to the genie–aided 1–st order RLS.

Fig. 2 shows the MSE performance as a function of average SNR ρ for a normalized maximum Doppler frequency of $f_{\text{D}}T_{\text{s}} = 0.001$. It is found that an optimized parameter γ is well approximated by $\gamma \simeq a\rho^b$ with proper constants a and b over the whole SNR regimes. As shown in Fig. 2, the performance curve of the conventional zero–th order RLS saturates as an average SNR increases, and the 1–st order RLS outperforms the zero–th order RLS for higher SNRs than 25 dB. The zero–th order SPARLS significantly improves the MSE of the conventional zero–th order RLS. More remarkably, it offers good performance close to the ideal case where the tap positions for non–zero entries are known at the receiver. It suggests that the SPARLS successfully estimates joint tap coefficients and positions with the use of $\mathcal{L}_1$–regularized sparse regressions. One can see that the high–order RLS degrades the MSE performance considerably for low SNR regimes for frequency–selective sparse channels, although it achieves good performance for high SNRs. This drawback is solved by the proposed high–order SPARLS, which can achieve near–optimal performance of genie–aided zero–th order RLS in low SNRs while its performance in high SNRs approaches to the genie–aided 1–st order RLS. The 2–nd order regression is not of use for less–than 35 dB SNRs in this channel condition.

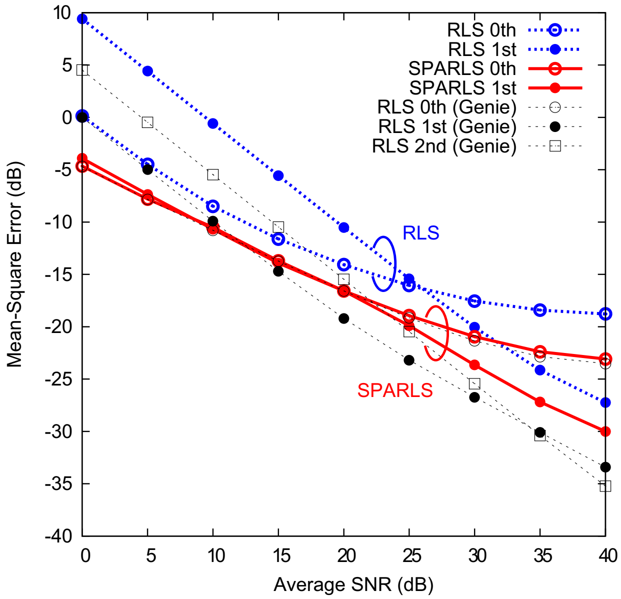

Fig. 2. MSE performance as a function of average SNR for $f_{\mathrm{D}}T_{\mathrm{s}} = 0.001$ (2×2 MIMO, $P = 20$ taps, $Q = 6$ paths).

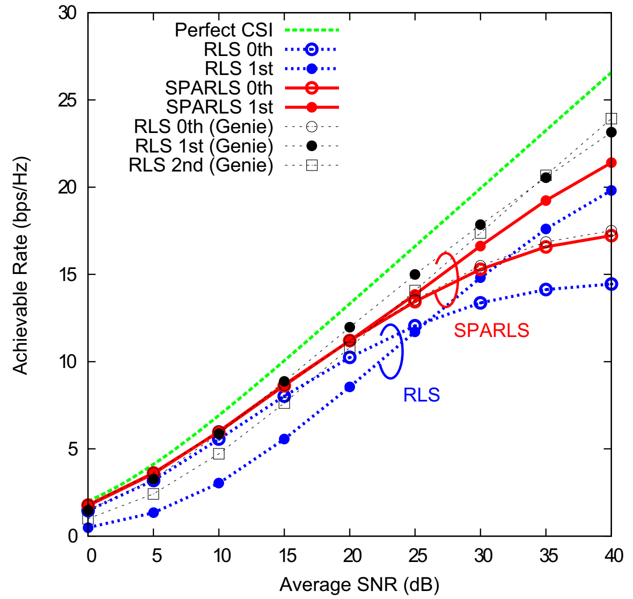

Fig. 3. Achievable rate performance as a function of average SNR for $f_{\mathrm{D}}T_{\mathrm{s}} = 0.001$ (2×2 MIMO, $P = 20$ taps, $Q = 6$ paths).

B. Capacity Upper-Bound

The channel estimation error in MSE has a great impact on the equalization performance (e.g., in bit error rate) and the achievable data rate. We show the effect of MSE on the data rate in Fig. 3, where we model the estimation error as an additional Gaussian noise which degrades the rate as follows:

$$R_k \leq \mathbb{E}\left[\log_2 \det\left[\boldsymbol{I}_N + \frac{1}{\sigma^2 + \varepsilon_k^2}\boldsymbol{H}(k)\boldsymbol{H}^\dagger(k)\right]\right] \leq \min(N, M') \log_2\left(1 + \frac{1}{\sigma^2 + \varepsilon_k^2}\right), \tag{27}$$

where we use Jensen's inequality. In this figure, we also present the ideal rate performance when perfect channel state information (CSI) is available at receivers ($\varepsilon_k^2 = 0$). Shown in this figure, the conventional zero–th order RLS do not perform well, especially in the high SNR regimes; approximately 12 bps/Hz (over the ideal rate with perfect CSI) is lost at an average SNR of 40 dB. It is demonstrated that the high–order sparse RLS algorithm offers the best performance over all the SNR regimes in the sense of achievable rate as well as MSE.

V. Conclusion

We proposed an improved channel estimation scheme which efficiently estimates a sparse channel matrix in doubly–selective fading MIMO channels by means of $\mathcal{L}_1$–$\mathcal{L}_q$ regularized high–order RLS algorithm. Through computer simulations, we demonstrated that the proposed high–order sparse RLS algorithm significantly improve MSE, and the achievable data rate can be considerably increased compared with the conventional RLS algorithm. The proposed algorithm can be implemented for multi-carrier transmissions as well as single-carrier transmissions. Comparisons to the Fourier basis expansion model, fractional-sample-space tap delay line and tap-position change remain as future works.

Appendix

A. Penalized ML Problem

Let us consider the exponentially–weighted noise: $\boldsymbol{Z}'_k = \boldsymbol{Z}_k \boldsymbol{\Lambda}_k^{-1/2}$, which follows the Gaussian distribution, $\mathrm{vec}\left[\boldsymbol{Z}'_k\right] \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma^2 \boldsymbol{\Lambda}_k^{-1} \otimes \boldsymbol{I}_N\right)$. When we assume the observation $\boldsymbol{Y}_k$ is modeled as

$$\boldsymbol{Y}_k = \hat{\boldsymbol{H}}(k)\boldsymbol{X}_k + \boldsymbol{Z}'_k, \tag{28}$$

the conditional probability of $\boldsymbol{Y}_k$ given $\boldsymbol{X}_k$ and $\hat{\boldsymbol{H}}(k)$ is given

$$\Pr\left(\boldsymbol{Y}_k \mid \boldsymbol{X}_k, \hat{\boldsymbol{H}}(k)\right) = \frac{\exp\left(-\frac{1}{\sigma^2}\mathcal{E}(k)\right)}{\det\left[\pi\sigma^2\boldsymbol{\Lambda}_k^{-1}\right]^N}, \tag{29}$$

where $\mathcal{E}(k)$ is a weighted MSE defined in (9). It implies that the $\mathcal{L}_1$–regularized problem in (23) is identified as a maximum–likelihood (ML) estimation problem with a penalty:

$$\max_{\hat{\boldsymbol{H}}(k)} \quad \log \Pr\left(\boldsymbol{Y}_k \mid \boldsymbol{X}_k, \hat{\boldsymbol{H}}(k)\right) - \gamma \left\|\mathrm{vec}\left[\hat{\boldsymbol{H}}(k)\right]\right\|_1. \tag{30}$$

This penalized ML problem is efficiently solved by an expectation–maximization (EM) approach with the noise decomposition, proposed in [8–10].

B. Noise Decomposition

We introduce a noise decomposition as $\boldsymbol{Z}'_k = \alpha \boldsymbol{\mathcal{N}}_k \boldsymbol{X}_k + \boldsymbol{\Xi}_k$, where $\boldsymbol{\mathcal{N}}_k \in \mathbb{C}^{N \times M'}$ and $\boldsymbol{\Xi}_k \in \mathbb{C}^{N \times k}$ are decomposed noise matrices of the Gaussian distributions:

$$\mathrm{vec}[\boldsymbol{\mathcal{N}}_k] \sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{I}_{M'} \otimes \boldsymbol{I}_N\right), \tag{31}$$

$$\mathrm{vec}[\boldsymbol{\Xi}_k] \sim \mathcal{CN}\left(\boldsymbol{0}, (\sigma^2\boldsymbol{\Lambda}_k^{-1} - \alpha^2 \boldsymbol{X}_k^\dagger \boldsymbol{X}_k)^{\mathrm{T}} \otimes \boldsymbol{I}_N\right), \tag{32}$$

where $\alpha \in \mathbb{R}_+$ is a decomposition constant which must fulfil $\alpha^2 \leq \sigma^2/\lambda_{\max}[\boldsymbol{X}_k^\dagger \boldsymbol{X}_k]$ with $\lambda_{\max}[\cdot]$ being the maximum eigenvalue. Since $\lambda_{\max}[\boldsymbol{X}_k^\dagger \boldsymbol{X}_k] \simeq M'$ for large k and for independent input, $\alpha^2 = \sigma^2/5M'$ can satisfy its requirement condition with a high probability.

With the decomposed noise matrices $\boldsymbol{\mathcal{N}}_k$ and $\boldsymbol{\Xi}_k$, we can rewrite the received signal model in (28) as follows:

$$\boldsymbol{Y}_k = \boldsymbol{G}(k)\boldsymbol{X}_k + \boldsymbol{\Xi}_k, \qquad \boldsymbol{G}(k) = \hat{\boldsymbol{H}}(k) + \alpha \boldsymbol{\mathcal{N}}_k, \quad (33)$$

where $\boldsymbol{G}(k) \in \mathbb{C}^{N \times M'}$ is a noisy channel matrix.

C. Expectation Step

The conditional probability of $\boldsymbol{G}(k)$ given an estimate $\hat{\boldsymbol{H}}(k)$ and an observation $\boldsymbol{Y}_k$ is a Gaussian distribution with mean

$$\hat{\boldsymbol{G}}(k) \triangleq \hat{\boldsymbol{H}}(k)\Big(\boldsymbol{I}_{M'} - \frac{\alpha^2}{\sigma^2}\boldsymbol{X}_k \boldsymbol{\Lambda}_k \boldsymbol{X}_k^\dagger\Big) + \frac{\alpha^2}{\sigma^2}\boldsymbol{Y}_k \boldsymbol{\Lambda}_k \boldsymbol{X}_k^\dagger. \quad (34)$$

This is the expectation step (E–step) of the EM algorithm to obtain the expectation $\hat{\boldsymbol{G}}(k)$ given an estimate $\hat{\boldsymbol{H}}(k)$. Note that if $\hat{\boldsymbol{G}}(k)$ converges to $\hat{\boldsymbol{H}}(k)$ such that $\hat{\boldsymbol{G}}(k) = \hat{\boldsymbol{H}}(k)$, the above equation reduces into the well–known normal equation, $\hat{\boldsymbol{H}}(k)\boldsymbol{X}_k \boldsymbol{\Lambda}_k \boldsymbol{X}_k^\dagger = \boldsymbol{Y}_k \boldsymbol{\Lambda}_k \boldsymbol{X}_k^\dagger$ for the conventional RLS estimation derived in (13). It implies that using smaller α approaches the performance of the conventional RLS estimation, while larger α offers a higher gain of the EM algorithm for sparse matrices. For faster convergence, we can use an estimated channel by the conventional RLS algorithm as an initial expectation of $\hat{\boldsymbol{G}}(k)$ in the E–step.

D. Maximization Step

In the maximization step (M–step) of the EM algorithm, we solve the penalized ML problem given an expected $\hat{\boldsymbol{G}}(k)$. We have

$$\log \Pr\big(\boldsymbol{Y}_k \mid \boldsymbol{X}_k, \hat{\boldsymbol{H}}(k), \hat{\boldsymbol{G}}(k)\big) = -\frac{\big\|\hat{\boldsymbol{H}}(k) - \hat{\boldsymbol{G}}(k)\big\|_{\mathrm{F}}^2}{\alpha^2} - \kappa,$$

where $\kappa \triangleq NM' \log(\alpha^2)$ is a constant. The optimization problem is rewritten as

$$\min_{\hat{\boldsymbol{H}}(k)} \quad \big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}(k) - \hat{\boldsymbol{G}}(k)\big]\big\|_2^2 + \alpha^2\gamma \big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}(k)\big]\big\|_1. \quad (35)$$

The maximized solution is obtained as

$$\hat{\boldsymbol{H}}(k) = \mathcal{F}_{\mathrm{th}}\Big(\hat{\boldsymbol{G}}(k), \alpha^2\gamma\Big). \quad (36)$$

E. Order Extensions

For high–order RLS estimation, we can use a similar EM algorithm to solve $\mathcal{L}_1$–$\mathcal{L}_q$ regularized sparse regressions, by replacing $\boldsymbol{X}_k$ with $\boldsymbol{\mathcal{X}}_k$ and $\hat{\boldsymbol{H}}(k)$ with $\hat{\boldsymbol{\mathcal{H}}}(k)$. For the E–step, we just need to use a proper α such that

$$\alpha^2 \ll \frac{\sigma^2}{\lambda_{\max}[\boldsymbol{\mathcal{X}}_k^\dagger \boldsymbol{\mathcal{X}}_k]} \simeq \frac{\sigma^2}{M'}\frac{k^2 - 1}{k^{2(D+1)} - 1} \simeq \frac{\sigma^2}{M' k^{2D}}. \quad (37)$$

For the M–step, we have the problem for each order:

$$\min_{\hat{\boldsymbol{H}}^{[d]}(k)} \quad \frac{1}{\alpha^2}\big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}^{[d]}(k) - \hat{\boldsymbol{G}}^{[d]}(k)\big]\big\|_2^2 + \gamma\big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}^{[d]}(k)\big]\big\|_1 + \mu_d \big\|\mathrm{vec}\big[\hat{\boldsymbol{H}}^{[d]}(k)\big]\big\|_q. \quad (38)$$

For $q \in \{1, 2\}$, the optimal solution is given as

$$\hat{\boldsymbol{H}}^{[d]}(k) = \mathcal{F}_{\mathrm{th}}\Big(\hat{\boldsymbol{G}}^{[d]}(k), \alpha^2\big(\gamma + \mu_d\omega_d\big)\Big), \quad (39)$$

where ω_d is chosen such that $\omega_d = 1/\big\|\mathrm{vec}[\hat{\boldsymbol{H}}^{[d]}(k)]\big\|_1$ for $q = 2$ and $\omega_d = 1$ for $q = 1$. For more general q, see [15].

References

[1] M. Akçakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *IEEE Trans. IT*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[2] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," *IEEE CISS*, pp. 5–10, Mar. 2008.

[3] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. IT*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *J. Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[5] Y. Chen, Y. Gu, and A. O. Hero III, "Sparse LMS for system identification," *IEEE ICASSP*, pp. 3125–3128, Apr. 2009.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. IT*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] J. W. Choi, K. Kim, T. J. Riedl, and A. C. Singer, "Iterative estimation of sparse and doubly-selective multi-input multi-output (MIMO) channel," *Asilomar Conf.*, pp. 620–624, Nov. 2009.

[8] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm," *IEEE Trans. SP*, vol. 58, no. 8, pp. 4013–4025, Aug. 2010.

[9] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Comparison of SPARLS and RLS algorithms for adaptive filtering," *IEEE SARNOFF*, pp. 1–5, Princeton, Apr. 2009.

[10] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[11] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, Oct. 1998.

[12] T. K. Akino, "Optimum-weighted RLS channel estimation for rapid fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4248–4260, Nov. 2008.

[13] T. Koike-Akino and P. Orlik, "High–order super–block GLRT for non–coherent Grassmann codes in MIMO–OFDM systems," *IEEE GLOBECOM*, Miami, Dec. 2010.

[14] P. Zhao, G. Rocha, and B. Yu, "Grouped and hierarchical model selection through composite absolute penalties," *J. Ann. Statist.*, vol. 37, no. 6A, pp. 3468–3497, 2009.

[15] H. Liu and J. Zhang, "On the l_1-l_q regularized regression," *Tech. Rep. Carnegie Mellon Univ.*, Feb. 2008.

[16] S. Haykin, *Adaptive Filter Theory, 3rd Ed.*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1996

[17] J. Lin, J. G. Proakis, F. Ling, and L. A. Hanoch, "Optimal tracking of time-varying channel: A frequency domain approach for known and new algorithms," *IEEE JSAC*, vol. 13, pp. 141–154, Jan. 1995.

[18] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. IT*, vol. 51, no. 3, pp. 1030–1051, Mar. 2006.

[19] R. Tibshirani, "Regression shrinkage and selection via the lasso," *JSTOR J. Royal Statist. Society*, Series B (Methodological), vol. 58, pp. 267–288, 1996.

[20] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *JSTOR SIAM Review*, vol. 43, no. 1, pp. 129–159, Feb. 2001.

[21] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statist. Society*, Series B (Methodology), vol. 68, no. 1, pp. 49–67, Feb. 2006.

[22] M. R. Osborne, B. Presnell, and B. A. Turlach, "On the lasso and its dual," *JSTOR J. Computational Graphical Statist.*, vol. 9, no. 2, pp. 319–337, 2000.