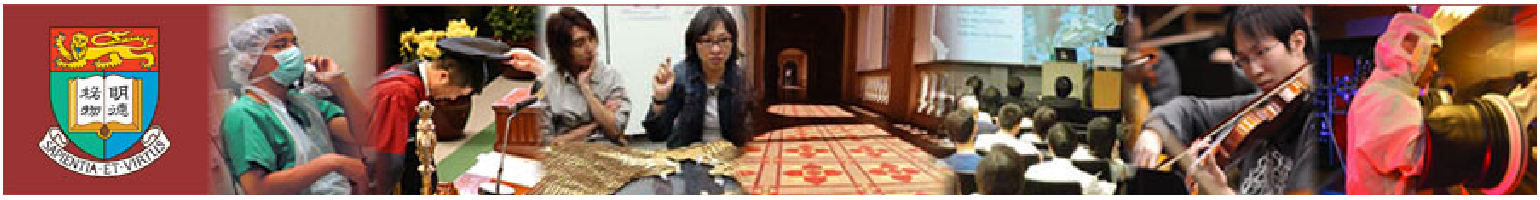

A New Transform-Domain Regularized Recursive Least M-Estimate Algorithm for a Robust Linear Estimation

S. C. Chan, *Member, IEEE*, Z. G. Zhang, *Member, IEEE*, and Y. J. Chu

***Abstract*—This brief proposes a new transform-domain (TD) regularized M-estimation (TD-R-ME) algorithm for a robust linear estimation in an impulsive noise environment and develops an efficient QR-decomposition-based algorithm for recursive implementation. By formulating the robust regularized linear estimation in transformed regression coefficients, the proposed TD-R-ME algorithm was found to offer better estimation accuracy than direct application of regularization techniques to estimate system coefficients when they are correlated. Furthermore, a QR-based algorithm and an effective adaptive method for selecting regularization parameters are developed for recursive implementation of the TD-R-ME algorithm. Simulation results show that the proposed TD regularized QR recursive least M-estimate (TD-R-QRRLM) algorithm offers improved performance over its least squares counterpart in an impulsive noise environment. Moreover, a TD smoothly clipped absolute deviation R-QRRLM was found to give a better steady-state excess mean square error than other QRRLM-related methods when regression coefficients are correlated.**

***Index Terms*—QR decomposition (QRD), recursive linear estimation and filtering, regularization, smoothly clipped absolute deviation (SCAD), system identification, transformed M-estimation (ME).**

I. Introduction

ESTIMATIONS of unknown coefficients of a linear regression model is a fundamental problem in statistical and signal processing communities [1], [2]. A least squares (LS) estimator, which is obtained by minimizing the sum of squared errors, is a popular method for addressing linear estimation problems. The LS estimator implicitly assumes that additive noise is Gaussian, and hence, its performance will be considerably degraded in an impulsive noise environment. This problem can be tackled by robust statistical techniques such as an M-estimator (ME) [3], [4], which employs an ME cost function instead of an LS cost function to suppress the adverse effect of impulsive noise. Recently, regularized LS estimators have attracted much interest as a valuable approach to overcome fundamental limitations of LS estimators when regression inputs are collinear or when a small number of observations are available. By incorporating a regularization term of regression coefficients to an LS cost function, an estimation variance of a regularized LS estimator can be significantly reduced. It has also been found in [5] that a linear absolute shrinkage and selection operator method, which is based on L_1 regularization, tends to produce sparse solutions. Therefore, appropriate regularization can also serve as a powerful technique for automatic variable or model selection. Because of these reasons, regularized LS estimations have gained popularity in a wide variety of fields such as image processing [6], audio signal processing [7], compressive sensing [8], [9], etc. Among numerous regularization techniques, smoothly clipped absolute deviation (SCAD) regularization is particularly attractive because of its asymptotically unbiased property [10]. Direct applications of SCAD regularization, however, may not fully explore its advantages such as sparsity and unbiasedness since regression coefficients may not be sparse in nature.

In this brief, we will: 1) introduce a new transform-domain (TD) regularized ME (TD-R-ME) algorithm for a robust linear estimation in an impulsive noise environment; 2) develop an efficient QR-decomposition (QRD)-based algorithm for recursive implementation of the TD-R-ME algorithm; and 3) propose an adaptive regularization parameter selection approach for the recursive implementation of the TD-R-ME. The basic idea of the TD-R-ME algorithm is to apply appropriate sparsity-enhancing orthogonal transformations such as a discrete cosine transformation (DCT) [11] and a discrete wavelet transform (DWT) [12] to regression coefficients and formulate a robust regularized linear estimation in a TD. Simulation results on system identification to be presented later show that the TD approach improves considerably sparsity of transform coefficients and hence leads to better estimation accuracy than direct application of regularization techniques. Moreover, it was found that a TD-SCAD-ME (i.e., a TD-R-ME with SCAD regularization) performs slightly better than its TD L_1-regularized counterpart (TD-L_1-ME).

Since most regularized LS or related algorithms are designed for batch processing and their direct implementation to a time series will lead to excessive computational complexity, we shall extend in this brief the QRD-based recursive regularized ME algorithms that we developed in [4] and [7] to a TD for online implementation and propose a new method for choosing regularization parameters. A resultant TD regularized QRD recursive least M-estimate (TD-R-QRRLM) algorithm requires only $O(M^2)$ complexity per iteration. Conventionally, a regularization parameter is chosen using cross validation [10] or by trial and error in tracking applications [4]. For online linear

Manuscript received July 15, 2010; revised October 20, 2010; accepted December 14, 2010. Date of current version February 24, 2011. This work was supported in part by The University of Hong Kong Committee on Research and Conference Grants Small Project Funding. This paper was recommended by Associate Editor Z. Lin.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong (e-mail: scchan@eee.hku.hk; zgzhang@eee.hku.hk; yjchu@eee.hku.hk).

Digital Object Identifier 10.1109/TCSII.2011.2106314

modeling, cross validation is unsuitable as signal statistics usually change over time. The regularization parameter proposed in this brief is derived from the theoretical analysis of an R-RLM algorithm for Gaussian inputs in [7]. It adapts automatically to an input signal-to-noise ratio (SNR), and simulation results show that it works well in practice. Moreover, simulation results show that the TD SCAD-regularized QRRLM (TD-SCAD-QRRLM) can achieve more accurate estimates for slowly varying systems than its L_2- or L_1-regularized counterparts (TD-L_2-QRRLM or TD-L_1-QRRLM, respectively).

The rest of this brief is organized as follows. In Section II, the proposed TD-R-ME algorithm is introduced. Section III is devoted to its recursive version called the TD-R-QRRLM algorithm. Simulation results and comparisons are presented in Section IV. Finally, conclusions are drawn in Section V.

II. TD-R-ME

Consider the following linear regression model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{y} = [y(1), y(2), \ldots, y(N)]^T \in \boldsymbol{R}^N$ is the observation vector, $\boldsymbol{X} = [\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(N)]^T \in \boldsymbol{R}^{N\times M}$ is the design matrix, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_M]^T \in \boldsymbol{R}^M$ is the regression coefficient vector, and $\boldsymbol{e} = [e(1), e(2), \ldots, e(N)]^T \in \boldsymbol{R}^N$ is the additive noise, which is generally assumed to be independent and an identically distributed Gaussian random variable with zero mean and a variance σ^2, i.e., $e(i) \sim \mathcal{N}(0, \sigma^2)$. N and M are, respectively, the number of observations and unknown coefficients.

In many applications, a coefficient vector $\boldsymbol{\beta}$ may not be sparse, and direct incorporation of a regularization term of $\boldsymbol{\beta}$ may not be desirable. On the other hand, coefficients of $\boldsymbol{\beta}$ may be correlated, and it is advantageous to represent it in a TD by means of an appropriate orthogonal linear transformation as follows:

$$\boldsymbol{c} = \boldsymbol{\Phi}\boldsymbol{\beta} \tag{2}$$

where $\boldsymbol{c}$ is the transformed coefficient vector, and $\boldsymbol{\Phi} \in \boldsymbol{R}^{M\times M}$ is the orthogonal transformation matrix satisfying $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \boldsymbol{I}$. If the coefficients of $\boldsymbol{\beta}$ exhibit certain periodicity, then a DCT will usually lead to a sparse representation and, hence, a sparse coefficient vector $\boldsymbol{c}$ [11]. For piecewise smooth coefficients, a DWT is usually more suitable in obtaining a sparse presentation [12].

With $\boldsymbol{\beta}$ expressed in (2) using the orthogonal transformation, the linear model of (1) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\beta} + \boldsymbol{e} = \boldsymbol{Z}\boldsymbol{c} + \boldsymbol{e} \tag{3}$$

where $\boldsymbol{Z} = [\boldsymbol{z}(1), \ldots, \boldsymbol{z}(N)]^T = [\boldsymbol{\Phi}^T\boldsymbol{x}(1), \ldots, \boldsymbol{\Phi}^T\boldsymbol{x}(N)]^T$ is the transformed design matrix. As most elements in $\boldsymbol{c}$ are usually very small, sparsity-promoting regularization such as L_1 and SCAD can be applied to reduce estimation variance and obtain sparse estimates, particularly when few samples are available. Similar argument holds for system identification during tracking of a slowly varying channel because there are very few samples corresponding to a channel response at a particular time. The estimated coefficient $\hat{\boldsymbol{\beta}}$ can be obtained by taking an inverse transformation of the estimated transform coefficient $\hat{\boldsymbol{c}}$ as $\hat{\boldsymbol{\beta}} = \boldsymbol{\Phi}^T\hat{\boldsymbol{c}}$. We shall refer to this procedure as the "TD estimation" hereinafter, in contrast with the "direct estimation" of $\boldsymbol{\beta}$ using (1) and regularization.

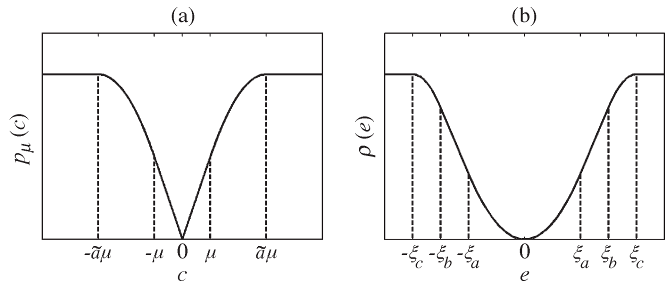

Fig. 1. (a) SCAD regularization function. (b) Hampel's M-estimation function.

The TD-R-ME $\hat{\boldsymbol{c}}_{R-M}$ can be obtained by minimizing the cost function as [10]:

$$J_{R-M}(\boldsymbol{c}) = \sum_{n=1}^{N} \rho\left(e(n)\right) + \sum_{m=1}^{M} p_\mu(c_m) \tag{4}$$

where $e(n) = y(n) - \boldsymbol{x}^T(n)\boldsymbol{\beta} = y(n) - \boldsymbol{z}^T(n)\boldsymbol{c}$, $\rho(e)$ is the M-estimate function, $p_\mu(\cdot)$ is the regularization function, and μ is the regularization parameter. The first and second terms on the right-hand size of (4) represent the robust error measure of $\boldsymbol{c}$ and the regularization term, respectively. If $\rho(e) = e^2$, we obtain the regularized LS solution. In terms of a Bayesian estimation, it is equivalent to assuming that the error is Gaussian distributed with a prior distribution on $\boldsymbol{c}$ being measured by the regularization function $p_\mu(\cdot)$. Note that the regularization term in (4) is imposed on $\boldsymbol{c}$ instead of $\boldsymbol{\beta}$, which usually results in a better performance of the TD estimation over its direct counterpart.

The most commonly used regularization functions include L_2 regularization $p_\mu(c) = \mu c^2$ and L_1 regularization $p_\mu(c) = \mu|c|$ [5]. SCAD regularization was proposed in the statistical community, and it possesses the desirable properties of sparsity and unbiasedness [10]. The SCAD function is based on the following function [see Fig. 1(a)]:

$$p_\mu(c) = \begin{cases} \mu|c|, & \text{for } |c| \le \mu, \\ -\frac{(|c|-\tilde{a}\mu)^2}{2(\tilde{a}-1)} + \frac{(\tilde{a}+1)\mu^2}{2}, & \text{for } \mu < |c| \le \tilde{a}\mu, \\ (\tilde{a}+1)\mu^2/2, & \text{for } |c| > \tilde{a}\mu \end{cases} \tag{5}$$

where the parameter $\tilde{a} > 2$ is, in general, selected as 3.7 [10].

It is well known that an LS cost function is sensitive to outliers, and a more robust approach is to employ an ME, which refers to a "generalized maximum-likelihood (ML) estimation" [3]. In the ML estimation, the objective function $\rho(e)$ to be minimized is $-\ln[P(e)]$, where $P(e)$ is the probability density of noise. Since it is difficult to estimate $P(e)$ accurately, $\rho(e)$ in the ME is chosen as an appropriate function to reduce the sensitivity of estimators to outliers. An effective ME function is the Hampel's function [see Fig. 1(b)] given by

$$\rho(e) = \begin{cases} e^2/2 & 0 \le |e| < \xi_a, \\ \xi_a|e| - \xi_a^2/2 & \xi_a \le |e| < \xi_b, \\ \xi_a\left[(\xi_b + \xi_c) - \xi_a \right. & \\ \qquad \left. + \left(|e| - \xi_c\right)^2/(\xi_b - \xi_c)\right]/2 & \xi_b \le |e| < \xi_c, \\ \xi_a(\xi_b + \xi_c)/2 - \xi_a^2/2 & \xi_c \le |e| \end{cases} \tag{6}$$

where ξ_a, ξ_b, and ξ_c are the thresholds to control the degree of outlier suppression. The Hampel's function behaves like a quadratic function when an error is below the threshold ξ_a. The effect of errors with large amplitudes will be reduced substantially beyond the thresholds. The thresholds can be estimated based on the variance of the "impulse-free" estimation error $\hat{\sigma}^2$, which is estimated as [4]

$$\hat{\sigma} = \{\text{median}\left(|y(n) - y(n-1)|\right)\} / (\sqrt{2} \cdot 0.6745), \quad n = 2, \ldots, N. \tag{7}$$

For the Hampel's function, these thresholds can be set as $\xi_a = 1.96 \cdot \hat{\sigma}$, $\xi_b = 2.24 \cdot \hat{\sigma}$, and $\xi_c = 2.58 \cdot \hat{\sigma}$. Alternatively, other well-known ME functions can also be used, but for simplicity, we shall only focus on the Hampel's function in this brief because of its good performance [4], [13].

The solution of (4) can be obtained by setting its derivative with respect to $\boldsymbol{c}$ to 0. This gives the following:

$$-\sum_{n=1}^{N} \rho'\left(e(n)\right) \boldsymbol{z}(n) + \boldsymbol{P}(\boldsymbol{c})\boldsymbol{1} = 0 \tag{8}$$

where $\boldsymbol{P}(\boldsymbol{c}) = \text{diag}\{p'_\mu(c_1), \ldots, p'_\mu(c_M)\}$ and $\boldsymbol{1}$ is a vector of all 1. By writing $\rho'(e(n))$ as $\rho'(e(n)) = q(e(n))e(n)$ and $p'_\mu(c_m)$ as $p'_\mu(c_m) = \lambda_\mu(c_m)c_m$ and approximating $q(e(n))$ and $\lambda_\mu(c_m)$ with the coefficient estimate at lth iteration (i.e., $q(e^{(l)}(n)) = \rho'(e^{(l)}(n))/e^{(l)}(n)$ and $\lambda_\mu(c_m^{(l)}) = p'_\mu(c_m^{(l)})/c_m^{(l)}$, where $e^{(l)}(n) = y(n) - \boldsymbol{z}(n)\hat{\boldsymbol{c}}^{(l)}$), one gets the following iteratively reweighted LS (IRLS) estimator of $\hat{\boldsymbol{c}}$ [10]:

$$\hat{\boldsymbol{c}}^{(l+1)} = \left(\boldsymbol{Z}^T\boldsymbol{W}^{(l)}\boldsymbol{Z} + \boldsymbol{\Lambda}^{(l)}\right)^{-1}\left(\boldsymbol{Z}^T\boldsymbol{W}^{(l)}\boldsymbol{y}\right) \tag{9}$$

where $\boldsymbol{W}^{(l)} = \text{diag}\{q(e^{(l)}(1)), \ldots, q(e^{(l)}(N))\}$, and $\boldsymbol{\Lambda}^{(l)} = \text{diag}\{\lambda_\mu(\hat{c}_1^{(l)}), \ldots, \lambda_\mu(\hat{c}_M^{(l)})\}$. The IRLS estimator can start with $\boldsymbol{W}^{(0)} = \boldsymbol{I}_N$ and $\boldsymbol{\Lambda}^{(0)} = \mu\boldsymbol{I}_M$ and will stop when a maximum number of iterations is reached or when the difference between two successive iterations is small enough. To yield the solution of quadratic minimization as in (9), the nonquadratic SCAD function is usually approximated locally at $\hat{\boldsymbol{c}}^{(l)}$ as [10]

$$p_\mu(c_m) = p_\mu(\hat{c}_m^{(l)}) + \frac{1}{2}p'_\mu\left(\hat{c}_m^{(l)}\right)\Big/\hat{c}_m^{(l)} \cdot \left[c_m^2 - \left(\hat{c}_m^{(l)}\right)^2\right] \tag{10}$$

for $\hat{c}_m^{(l)} \neq 0$, and $p_\mu(c_m) = 0$ for $\hat{c}_m^{(l)} = 0$.

III. TD-R-QRRLM Algorithm

The SCAD-regularized ME works on a batch of observations. For online application and a slowly time-varying parameter estimation, the QRD-based time-recursive algorithm is more attractive owing to its low arithmetic complexity, high numerical stability, and efficient hardware implementation [1], [4]. We now extend the time-recursive QRD algorithm in [7] to the TD-R-ME. The resulting algorithm is called the TD-R-QRRLM algorithm (see Table I). More precisely, the cost function of the TD-R-QRRLM is

$$J_{\text{R-QRRLM}}\left(\boldsymbol{c}(n)\right) = \sum_{i=1}^{n} \lambda^{n-i}\left\{\rho\left(e(i)\right) + \sum_{m=1}^{M} p_\mu\left(|c_m(i)|\right)\right\} \tag{11}$$

TABLE I
TD-R-QRRLM Algorithm

Initialization:

$\tilde{\boldsymbol{R}}_{\rho,p}(0) = \sqrt{\delta}\boldsymbol{I}$, δ is a small positive constant and $\boldsymbol{I}$ is an $M \times M$ identity matrix; $\tilde{\boldsymbol{P}}_\rho(0) = \boldsymbol{0}$; $\hat{\boldsymbol{c}}(0) = \boldsymbol{0}$.

Recursion:

Given $\tilde{\boldsymbol{R}}_{\rho,p}(n-1)$, $\tilde{\boldsymbol{P}}_\rho(n-1)$, $\hat{\boldsymbol{c}}(n-1)$, $\boldsymbol{z}(n)$ and $y(n)$, compute at n:

(i). the 1st QRD

$$\begin{bmatrix} \tilde{\boldsymbol{R}}_{\rho,p}'(n) & \tilde{\boldsymbol{P}}_\rho'(n) \\ \boldsymbol{0} & \varepsilon'(n) \end{bmatrix} = \boldsymbol{Q}'(n)\begin{bmatrix} \sqrt{\lambda_e(n)}\tilde{\boldsymbol{R}}_{\rho,p}(n)(n-1) & \sqrt{\lambda_e(n)}\tilde{\boldsymbol{P}}_\rho(n-1) \\ \sqrt{q(e(n))}\boldsymbol{z}^T(n) & \sqrt{q(e(n))}y(n) \end{bmatrix}$$

(ii). the 2nd QRD

$$\begin{bmatrix} \tilde{\boldsymbol{R}}_{\rho,p}(n) & \tilde{\boldsymbol{P}}_\rho(n) \\ \boldsymbol{0} & \varepsilon(n) \end{bmatrix} = \boldsymbol{Q}(n)\begin{bmatrix} \tilde{\boldsymbol{R}}_{\rho,p}'(n) & \tilde{\boldsymbol{P}}_\rho'(n) \\ \sqrt{q(e(n))M\Lambda_m(n)} & 0 \end{bmatrix}$$

where the QRD can be implemented by Givens rotation [1], $\boldsymbol{Q}'(n)$, $\boldsymbol{Q}(n)$, $\varepsilon'(n)$, and $\varepsilon(n)$ are produced by QRD, $\Lambda_m(n)$ is a row vector randomly or sequentially taken from $\boldsymbol{\Lambda}(n)$, and $\boldsymbol{\Lambda}(n) = diag\{\lambda_\mu(\hat{c}_1(n-1)), \ldots, \lambda_\mu(\hat{c}_M(n-1))\}$.

(iii). Solve $\boldsymbol{c}(n)$ from $\tilde{\boldsymbol{R}}_{\rho,p}(n)\boldsymbol{c}(n) = \tilde{\boldsymbol{P}}_\rho(n)$ by back-substitution [1].

Inverse transformation:

$\hat{\boldsymbol{\beta}}(n) = \boldsymbol{\Phi}^T\hat{\boldsymbol{c}}(n)$.

where λ is the forgetting factor close to but smaller than 1 and $e(i) = y(i) - \boldsymbol{z}^T(i)\boldsymbol{c}(n)$. By setting the partial derivative of $J_{\text{R-QRRLM}}(\boldsymbol{c}(n))$, with respect to $\boldsymbol{c}(n)$ and 0, we obtain the following necessary optimal condition called the regularized M-estimate normal equation:

$$\tilde{\boldsymbol{R}}_{\rho,p}(n)\boldsymbol{c}(n) = \tilde{\boldsymbol{P}}_\rho(n) \tag{12}$$

where

$$\tilde{\boldsymbol{R}}_{\rho,p}(n) = \lambda\tilde{\boldsymbol{R}}_{\rho,p}(n-1) + q\left(e(n)\right)\boldsymbol{z}(n)\boldsymbol{z}^T(n) + \boldsymbol{\Lambda}(n) \tag{13}$$

$$\tilde{\boldsymbol{P}}_\rho(n) = \lambda\tilde{\boldsymbol{P}}_\rho(n-1) + q\left(e(n)\right)y(n)\boldsymbol{z}(n). \tag{14}$$

Note that, for a long duration of impulses, $\tilde{\boldsymbol{R}}_{\rho,p}(n)$ will be dominated by $\boldsymbol{\Lambda}(n)$, and $\tilde{\boldsymbol{P}}_\rho(n)$ will approach to zero. Thus, it is better to update $\tilde{\boldsymbol{R}}_{\rho,p}(n)$ and $\tilde{\boldsymbol{P}}_\rho(n)$, respectively, as

$$\tilde{\boldsymbol{R}}_{\rho,p}(n) = \lambda_e(n)\tilde{\boldsymbol{R}}_{\rho,p}(n-1) + q\left(e(n)\right)\left[\boldsymbol{z}(n)\boldsymbol{z}^T(n) + \boldsymbol{\Lambda}(n)\right] \tag{15}$$

$$\tilde{\boldsymbol{P}}_\rho(n) = \lambda_e(n)\tilde{\boldsymbol{P}}_\rho(n-1) + q\left(e(n)\right)y(n)\boldsymbol{z}(n) \tag{16}$$

where $\lambda_e(n) = (\lambda - 1)q(e(n)) + 1$.

Since a direct estimation of $\boldsymbol{c}(n)$ in (12) requires $O(M^3)$ arithmetic complexity, efficient QRD implementation with complexity of $O(M^2)$ will be proposed. We note that the update of $\tilde{\boldsymbol{R}}_{\rho,p}(n)$ in (15) involves the two terms $q(e(n))\boldsymbol{z}(n)\boldsymbol{z}^T(n)$ and $q(e(n))\boldsymbol{\Lambda}(n)$, which can be achieved by means of two successive QRD operations. The update of the term $q(e(n))\boldsymbol{z}(n)\boldsymbol{z}^T(n)$ in $\tilde{\boldsymbol{R}}_{\rho,p}(n)$ requires QRD with complexity of $O(M^2)$. The second term $q(e(n))\boldsymbol{\Lambda}(n)$ is slightly complicated since it is of a full rank, and its QRD updating has complexity of $O(M^3)$. To reduce the complexity, we only update a row vector $M\Lambda_m(n)$ at each iteration, where $\Lambda_m(n)$ is a row randomly or sequentially taken from $\boldsymbol{\Lambda}(n)$, instead of the whole matrix $\Lambda(n)$, and such implementation can reduce the complexity of the second QRD to $O(M^2)$. The idea is to embed the regularization progressively over time, which approximates the updating with $\boldsymbol{\Lambda}(n)$. More precisely, at each time instant, QRD is executed once for the data vector $\sqrt{q(e(n))}[\boldsymbol{z}^T(n),\, y(n)]$ and once for the regularization vector $\sqrt{q(e(n))}[\sqrt{M\Lambda_m(n)},\, 0]$. As a result, the TD-R-QRRLM method only increases the complexity of the conventional QR-RLM by about two times. Compared with a batch processing algorithm, which has $O(M^3)$ complexity at each time instant, the TD-R-QRRLM method has much lower arithmetic complexity of $O(M^2)$.

In [7], the mean square convergence analysis of the L_2-QRRLM for Gaussian inputs and contaminated Gaussian (CG) noise was performed, and the regularization parameter $\mu_{L_2}(n) = (1-\lambda)M\overline{\sigma}_z^2(n)\sqrt{\hat{\sigma}^2(n)/[Tr(\boldsymbol{R}_{\rho\_zz}(n))\boldsymbol{c}(n-1)^T\boldsymbol{c}(n-1)]}$ was found to give good performance in minimizing the steady-state excess mean square error (EMSE). Here, $\boldsymbol{R}_{\rho\_zz}(n)$ is the robust covariance matrix of the input, $\overline{\sigma}_z^2(n)$ is the long-term estimate of input signal power, and $\hat{\sigma}^2(n)$ is the variance of additive noise. We now extend μ_{L_2} to the cases of L_1 and SCAD regularization, which can be viewed as a weighted L_2 regularization with the regularization parameter $p'_\mu(|c_m|)/|c_m|$ [10]. Due to the sparsity property, the steady-state EMSE of the L_1- or SCAD-QRRLM is only contributed by nonzero coefficients. Thus, for the L_1 regularization, we have $\mu_{L_1}/|\overline{c}| \approx \mu_{L_2}$, where $\overline{c}$ is the mean of nonzero coefficients. Using the approximation $\boldsymbol{c}^T\boldsymbol{c} \approx M_1|\overline{c}|^2$, where M_1 is the number of nonzero coefficients, we obtain the variable regularization parameter for the L_1-TD-QRRLM as

$$\mu_{L_1}(n) = (1-\lambda)M\overline{\sigma}_z^2(n)\sqrt{\hat{\sigma}^2(n)/\{[Tr\left(\boldsymbol{R}_{\rho\_zz}(n)\right)+\varepsilon]\, M_1\}} \tag{17}$$

where ε is the small positive constant included to prevent $\mu_{L_1}(n)$ from having an infinity value when the input is very small, i.e., when $Tr(\boldsymbol{R}_{\rho\_zz}(n))$ approaches zero. The selection of regularization parameter for a SCAD is more complicated because $p'_\mu(|c_m|)$ is a piecewise function. Based on the facts that SCAD regularization is identical to L_1 regularization for small coefficients and its regularization parameter does not affect the estimates of large coefficients (for which $p_\mu(|c_m|)$ is a constant), the variable regularization parameter $\mu_{SCAD}(n)$ for the TD-SCAD-QRRLM can also be approximated as in (17). The good performance of above regularization parameter selection method is substantiated by simulation results in the next section.

In (17), $\boldsymbol{R}_{\rho\_zz}(n)$ can be recursively estimated as

$$\boldsymbol{R}_{\rho\_zz}(n) = \lambda_e(n)\boldsymbol{R}_{\rho\_zz}(n-1) + (1-\lambda_e(n))\, q\left(e(n)\right)\boldsymbol{z}(n)\boldsymbol{z}^T(n) \tag{18}$$

whereas $\overline{\sigma}_z^2(n) = 1/n\sum_{i=1}^{n} Tr(\boldsymbol{R}_{\rho\_zz}(i))$ is estimated as the long-term averaged trace of $\boldsymbol{R}_{\rho\_zz}(i)$, and M_1 can be approximated as the number of coefficients with absolute values larger than 1% of the maximum absolute value in the coefficients. $\hat{\sigma}^2(n)$ can be recursively estimated as [13]

$$\hat{\sigma}^2(n) = \lambda_\sigma\hat{\sigma}^2(n-1) + c_1(1-\lambda_\sigma)\text{median}\left[A_e(n)\right] \tag{19}$$

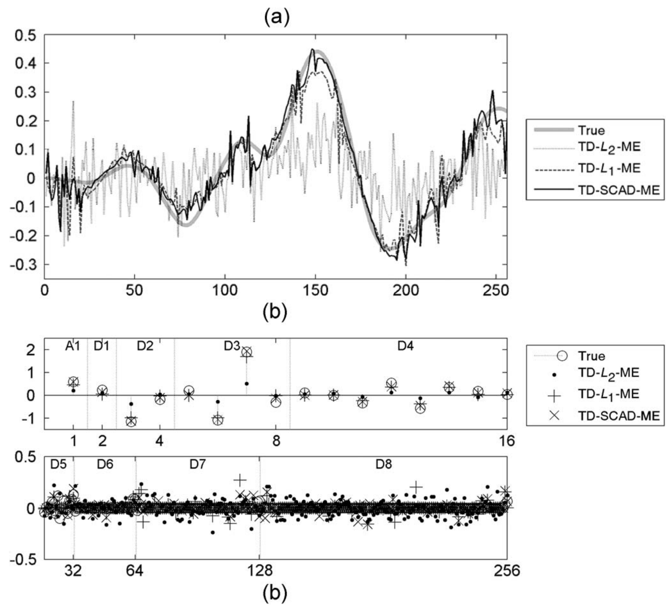

Fig. 2. Comparison of the correlated coefficient estimates using various linear estimation methods.

where λ_σ is the forgetting factor, $A_e(n) = \{e^2(n), \ldots, e^2(n - N_w + 1)\}$, $c_1 = 1.483(1 + 5/(N_w - 1))$ is the finite sample correction factor, and N_w is the length of an estimation window.

IV. Simulation Results

The performances of the TD-R-ME and TD-R-QRRLM algorithms are evaluated using computer simulations of the system identification problem in an impulsive noise environment.

A. TD-R-ME for Time Invariant Coefficients

The system coefficients are generated by passing a length-M random sequence $\widetilde{\boldsymbol{\beta}} \sim \mathcal{N}(0,1)$, i.e., a Gaussian process with a mean value of 0 and a variance value of 1 through a low-pass filter. Since the impulse response consists of significant low-frequency components, the wavelet-transformed coefficients exhibit a sparse representation. In this simulation, $M = 256$, and the low-pass filter used to generate the coefficients has a length of 4 and a cutoff frequency of $8/M$. Daubechies-5 wavelet is used to decompose the coefficients with eight levels. As seen from Fig. 2(b), most of the DWT coefficients are very small, and thus, they can be considered as sparse. The number of observations $y(n)$ is $N = 128$, and the design vectors are generated from $\boldsymbol{x}(n) = [s(n),\, \ldots,\, s(n-M+1)]^T$ with $s(n) \sim \mathcal{N}(0,\, 4)$. The noise is simulated using the CG noise model as follows:

$$P(e) \sim (1-\eta)(0, \sigma_g) + \eta\mathcal{N}(0, \sigma_{im}) \tag{20}$$

where σ_g^2 and σ_{im}^2 are set to make the overall SNR equal to 10 and -20 dB, respectively, and η denotes the occurrence probability of the impulsive component with the variance σ_{im}^2 ($\sigma_{\text{im}}^2 \gg \sigma_g^2$). Here, η is set to 0.05, and the generated impulses are randomly located in the observations. The L_2-, L_1-, and SCAD-regularized LS estimators and MEs are tested using both the direct and TD estimation approaches. The regularization parameters are selected by the generalized cross validation proposed in [10], and the maximum iteration number

TABLE II
MSE COMPARISON OF VARIOUS LINEAR ESTIMATION METHODS

		L_2	L_1	SCAD
Direct Estimation	LS estimator	37.28	51.01	51.13
	M-estimator	6.06	6.66	6.65
Transform-domain Estimation	LS estimator	37.28	49.30	49.22
	M-estimator	6.06	1.33	1.02

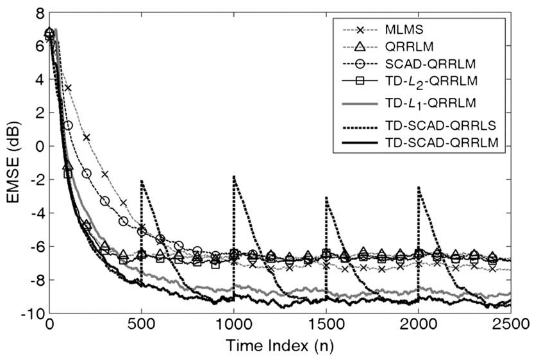

Fig. 3. EMSE performance of various algorithms for tracking the time-varying correlated coefficients in impulsive noise.

for IRLS is 20. Because of the orthogonal transformation, it can be easily verified that the L_2-regularized estimators in the direct and TD estimations actually offer the same results.

The MSE is calculated as $MSE = \|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|_2^2$ for quantitative comparison, Table II lists the MSE values averaged over 100 independent runs, and Fig. 2 shows the results of one realization. It can be seen that the results of the LS estimators are seriously degraded by impulsive noise, whereas the proposed ME methods give significantly better performance. Among the various regularization methods, the SCAD- and L_1-regularized MEs are slightly inferior to the L_2-regularized ME when the nonsparse coefficients are estimated directly. However, in the TD estimation, the SCAD-regularized ME considerably outperforms the L_2-regularized ME due to its sparsity property, and it also attains a small improvement over the L_1-regularized ME due to its unbiasedness. For clarity, Fig. 2 shows only the ME-based results in the TD estimation obtained in one realization. It can be seen that the SCAD estimator is able to offer a sparse solution while avoiding the bias brought by L_2 or L_1 regularization to large wavelet coefficients.

B. TD-R-QRRLM for Time-Varying Coefficients

We now consider the performance of the TD-R-QRRLM algorithms for tracking time-varying linear systems. The number of observations is $N = 2500$, and the number of coefficients is $M = 64$. The time-varying coefficients $\boldsymbol{\beta}(n)$ are generated by passing a length-M random sequence $\widetilde{\boldsymbol{\beta}}(n)$ through a low-pass filter with a cutoff frequency $10/M$, where $\widetilde{\boldsymbol{\beta}}(n)$ is given by $\widetilde{\beta}_m(n) = \widetilde{\beta}_m(n-1) + \widetilde{\delta}_m(n)$ with $\widetilde{\beta}_m(0) \sim \mathcal{N}(0, 4)$ and $\widetilde{\delta}_m(n) \sim \mathcal{N}(0, 4 \times 10^{-4})$. The forgetting factor in Table I is $\lambda = 0.99$, and the parameters in (19) for estimating the noise variance are $\lambda_\sigma = 0.95$ and $N_w = 9$. The impulsive noise is added at time instants 500, 1000, 1500, and 2000 for better visualization, and their amplitudes are CG distributed. All the other testing parameters are the same as in the previous example. Fig. 3 compares the tracking results of several LS-based and ME-based recursive algorithms in a TD or a direct estimation. Since the good performance and comparison of the RLM algorithm with other conventional algorithms had been reported in [13], we will compare the proposed method with the RLM algorithm and a related median least mean squares (MLMS) algorithm [14] with step size chosen by trial and error to minimize the steady-state EMSE. It can be seen that: 1) the regularization-based QRRLM has better performance than the conventional QRRLM and the MLMS; 2) the TD-SCAD-QRRLM has a smaller EMSE than other TD-QRRLM algorithms; 3) the TD-SCAD-QRRLM is more robust than its LS counterpart (i.e., TD-SCAD-QRRLS); and 4) the TD-SCAD-QRRLM has better performance than its direct counterpart (i.e., SCAD-QRRLM).

V. CONCLUSION

A new robust linear estimation method using a transformed ME and regularization is presented. For correlated regression coefficients, the TD-R-ME algorithm gives better performance than its direct estimation and LS counterparts in an impulsive noise environment. Recursive QRD-based algorithm for online implementation of the TD-R-ME algorithms and a new variable regularization parameter are also presented. Simulations showed that the TD-SCAD-QRRLM method has better performance than other QRRLM-related methods for tracking of correlated regression coefficients.

REFERENCES

[1] J. Proakis, C.M. Rader, F. Ling, M. Moonen, I. K. Proudler, and C. L. Nikias, *Algorithms for Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[2] Y. J. Zheng and Z. P. Lin, "Recursive adaptive algorithms for fast and rapidly time-varying systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 50, no. 9, pp. 602–614, Sep. 2003.

[3] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[4] Z. G. Zhang, S. C. Chan, Y. Zhou, and Y. Hu, "Robust linear estimation using M-estimation and weighted L_1 regularization: Model selection and recursive implementation," in *Proc. ISCAS*, Taipei, Taiwan, May 24-27, 2009, vol. 5, pp. 1193–1196.

[5] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[6] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization–minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[7] S. C. Chan, Y. J. Chu, and Z. G. Zhang, "A new regularized QRD recursive least M-estimate algorithm: Performance analysis and application," in *Proc. ICGCS*, Shanghai, China, Jun. 21–23, 2010, pp. 190–195.

[8] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

[9] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[10] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.

[11] B. Jafarpour, V. K. Goyal, W. T. Freeman, and D. B. McLaughlin, "Compressed history matching: Exploiting transform-domain sparsity for regularization of nonlinear dynamic data integration problems," *Math. Geosci.*, vol. 42, no. 1, pp. 1–27, Jan. 2010.

[12] A. Robertson, K. C. Park, and K. F. Alvin, "Extraction of impulse response data via wavelet transforms for structural system identification," *J. Vib. Acoust.*, vol. 120, no. 1, pp. 252–260, Jan. 1998.

[13] S. C. Chan and Y. X. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulse noise: Fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.

[14] G. A. Williamson, P. Clarkson, and W. A. Sethares, "Performance characteristics of the adaptive median LMS filter," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 667–680, Feb. 1993.