An Optimal Error Nonlinearity for Robust Adaptation Against Impulsive Noise

Sara Al-Sayed and Abdelhak M. Zoubir
Signal Processing Group
Technische Universität Darmstadt, 64283 Darmstadt, Germany
Email: {salsayed,zoubir}@spg.tu-darmstadt.de

Ali H. Sayed
Department of Electrical Engineering
University of California, Los Angeles, CA 90095
Email: sayed@ee.ucla.edu

***Abstract*—The least-mean squares algorithm is non-robust against impulsive noise. Incorporating an error nonlinearity into the update equation is one useful way to mitigate the effects of impulsive noise. This work develops an adaptive structure that parametrically estimates the optimal error-nonlinearity jointly with the parameter of interest, thus obviating the need for a priori knowledge of the noise probability density function. The superior performance of the algorithm is established both analytically and experimentally.**

I. Introduction

We consider the problem of adaptive estimation of an unknown deterministic parameter when the measurements are corrupted by impulsive noise. An impulsive noise process can be described as one whose realizations contain sparse, random samples of amplitude considerably larger than nominally accounted for, and, hence, the process is best modeled by heavy-tailed distributions [1], [2]. As far as the popular least-mean squares (LMS) algorithm [3] is concerned, the presence of impulsive noise in the measurements degrades the adaptive filter's performance, in terms of stability and steady-state behavior [4]. Several LMS-type algorithms have been developed that are robust against impulsive noise [5]–[8]. A recurrent feature in these algorithms is that their updates are nonlinear functions of the error signal. The problem of optimal nonlinearity design was addressed in [9], [10] and some references therein. Optimal design techniques, however, are hampered by their prerequisite of exact knowledge of the noise probability density function (pdf), which is rarely available in practice. It is therefore the aim of this work to develop an adaptive filtering algorithm that parametrically estimates the optimal error-nonlinearity *jointly* with the parameter of interest for improved mean stability and steady-state performance in impulsive noise environments.

II. Robust Adaptive Filtering

The goal is to adaptively estimate an unknown deterministic real-valued $M \times 1$ parameter w^o from available data $\{d(i), u_i\}$ for $i \geq 0$. The data are related to w^o via the linear regression model:

$$d\left(i\right) = u_i w^o + v\left(i\right) \tag{1}$$

where the $\{d\left(i\right)\}$ are real-valued scalar measurements, and the $\{u_i\}$ are real-valued row regression vectors of size M. Both measurements arise from realizations of jointly wide-sense stationary zero-mean random processes $\{\boldsymbol{d}\left(i\right), \boldsymbol{u}_i\}$, where the boldface font notation is used to denote random variables. The regressors have covariance matrix $R_u = E\boldsymbol{u}_i^T \boldsymbol{u}_i > 0$, where E denotes expectation. The noise-process $\{\boldsymbol{v}\left(i\right)\}$ is a real-valued zero-mean white process with variance σ_v^2. The random variables $\boldsymbol{u}_i$ and $\boldsymbol{v}\left(j\right)$ are assumed to be independent for all i and j.

The LMS filter is a stochastic gradient algorithm based on minimizing the mean-square error (MSE) cost function:

$$J\left(w\right) \triangleq E\left(\boldsymbol{d}\left(i\right) - \boldsymbol{u}_i w\right)^2. \tag{2}$$

With the output error defined as $e\left(i\right) \triangleq d\left(i\right) - u_i w_{i-1}$, the LMS recursion reads

$$w_i = w_{i-1} + \mu u_i^T e\left(i\right), \quad i \geq 0 \tag{3}$$

where $\left(\cdot\right)^T$ denotes transposition, and μ is a positive step-size parameter chosen to ensure stability.

In robust adaptive filtering [6]–[8], the cost function (2) is modified to

$$J^M\left(w\right) \triangleq E\rho\left(\boldsymbol{d}\left(i\right) - \boldsymbol{u}_i w\right) \tag{4}$$

where $\rho\left(x\right)$ is some bounded and continuous M-estimate function, which ensures qualitative robustness [11]. Assuming $\rho\left(x\right)$ is differentiable, the steepest-descent recursion that attempts to minimize (4), subject to a suitable choice of the initial condition, takes the form:

$$w_i = w_{i-1} - \mu\left(\nabla_w J^M\left(w_{i-1}\right)\right)^T. \tag{5}$$

By forgoing the expectation, and with $\psi\left(x\right) \triangleq \frac{d\rho(x)}{dx}$ referred to as the score function, the resulting stochastic instantaneous approximation of (5) is

$$w_i = w_{i-1} + \mu u_i^T \psi\left(e\left(i\right)\right). \tag{6}$$

The LMS recursion (3) is recovered when $\rho\left(x\right) = \frac{x^2}{2}$. It was shown in [9] that the optimal score function that minimizes the steady-state MSE is

$$\psi_1^{\text{opt}}\left(x\right) = -\frac{f'_{\boldsymbol{v}}\left(x\right)}{f_{\boldsymbol{v}}\left(x\right)} \tag{7}$$

where $f_{\boldsymbol{v}}\left(x\right)$ is the noise pdf, and $g'\left(x\right) \triangleq \frac{dg(x)}{dx}$. In this case, the LMS algorithm is MSE-optimal when $\{\boldsymbol{v}\left(i\right)\}$ is Gaussian, with the $\frac{1}{\sigma_v^2}$ proportionality constant absorbed into the step-size parameter, μ. However, the LMS algorithm is suboptimal when the noise is non-Gaussian [3], [12]. Yet, in order to design the filter optimally, the noise pdf must

This work was performed within the LOEWE Priority Program Cocoon (http://www.cocoon.tu-darmstadt.de). The work of Ali H. Sayed was supported in part by NSF grant CCF-1011918.

be known exactly, which does not always hold in practice. Under less restrictive assumptions, the authors in [10] derived an optimal score function that holds over a wider range of adaptation and not only at steady-state, leading to the choice:

$$\psi_{2,i}^{\text{opt}}(x) = -\frac{f'_{\boldsymbol{e}(i)}(x)}{f_{\boldsymbol{e}(i)}(x)} \tag{8}$$

in terms of the pdf of the error signal, $\boldsymbol{e}(i)$. This function is more intuitive in an adaptive setting, and reduces to $\psi_1^{\text{opt}}(x)$ at steady-state.

In [13] and [14], in the context of offline robust estimation, the optimal score function $\psi_{2,i}^{\text{opt}}(e(i))$ is approximated by a function, $\varphi(e(i))$, that is a linear combination of preselected basis score functions:

$$\varphi(e(i)) \triangleq \Phi_i^T \alpha_i \tag{9}$$

where $\alpha_i \triangleq (\alpha_i(1), \ldots, \alpha_i(B))^T$ is the vector of combination weights, and $\Phi_i \triangleq (\phi_1(e(i)), \ldots, \phi_B(e(i)))^T$ is the vector of basis score functions evaluated at the output error. The vector α_i is chosen to minimize the MSE between the true and approximate score functions:

$$\alpha_i^{\text{opt}} \triangleq \arg\min_{\alpha_i} E\left(\varphi(\boldsymbol{e}(i)) - \psi_{2,i}^{\text{opt}}(\boldsymbol{e}(i))\right)^2. \tag{10}$$

In the online adaptive context pertinent to this article, it is imperative to compute α_i adaptively and jointly with w_i. This is treated in Subsection II-A, where in the process of deriving the adaptive update for α_i we shall exploit the condition

$$E\left[\phi_b(x)\,\psi_{2,i}^{\text{opt}}(x)\right] = E\left[\phi'_b(x)\right], \tag{11}$$

which follows from the assumption

$$\lim_{x\to\pm\infty} \phi_b(x)\, f_{\boldsymbol{e}(i)}(x) = 0. \tag{12}$$

We consider herein B basis score functions that arise from zero-mean Gaussian pdfs with distinct variances, i.e., we select:

$$\phi_b\left(x;\sigma_b^2\right) = \frac{x}{\sigma_b^2}, \quad b \in \{1,\ldots,B\}. \tag{13}$$

Let

$$s \triangleq \Phi'_i = \left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_B^2}\right)^T. \tag{14}$$

The selection of $\{\sigma_b^2\}$ is addressed in Subsection II-B. Replacing $\psi(x)$ in (6) by its approximation (9), and using (13) and (14), the recursion now reads

$$w_i = w_{i-1} + \mu\left(\alpha_i^T s\right) u_i^T e(i). \tag{15}$$

This recursion is one of LMS with variable step-size (VSS-LMS), with $\mu(i) \triangleq \mu\left(\alpha_i^T s\right)$. Several VSS-LMS variants have been developed in the literature in order to improve the tradeoff between misadjustment and convergence rate compared to LMS [15]–[18]. Here, on the other hand, we shall design the variable step-size with the intent of enhancing the robustness of the LMS filter against impulsive noise.

Some remarks are in order before proceeding with the development of the robust algorithm, which we refer to henceforth as RLMS. First, since $\mu(i) \triangleq \mu\left(\alpha_i^T s\right)$ is constrained to be positive, the entries of α_i must be positive. Moreover, the authors of [14] imposed a convexity constraint on the entries of α_i in the offline estimation context, illustrating performance gains. We shall consider a similar convexity constraint on the entries of α_i.

A. Joint Parameter Adaptation

In this section, we apply a technique from [19] to solve (10) adaptively, subject to the aforementioned constraints on α_i. Let $\Omega_+ \triangleq \left\{\alpha \in \mathbb{R}_+^B | \alpha^T \mathbb{1} = 1\right\}$ where $\mathbb{R}_+^B$ is the set of $B \times 1$ vectors on the set of non-negative real numbers $\mathbb{R}_+$, and $\mathbb{1}$ is the all-one $B \times 1$ vector. We seek the solution of the following convex optimization problem:

$$\min_{\alpha\in\Omega_+} E\left(\Phi_i^T \alpha - \psi_{2,i}^{\text{opt}}(\boldsymbol{e}(i))\right)^2. \tag{16}$$

We would like to transform (16) in such a way that eliminates the constraints. First, we suppress the non-negativity constraint on the entries of α and later appropriately transform the solution to accommodate this requirement. Hence, let $\Omega \triangleq \left\{\alpha \in \mathbb{R}^B | \alpha^T \mathbb{1} = 1\right\}$. Next, we introduce the projection $\mathcal{P}_\Omega$ from $\mathbb{R}^B$ onto Ω:

$$\mathcal{P}_\Omega(\beta) = \left(I_B - \frac{\mathbb{1}\mathbb{1}^T}{B}\right)\beta + \frac{\mathbb{1}}{B}, \quad \forall\, \beta \in \mathbb{R}^B$$

where I_B is the identity matrix of size B. Every $\alpha \in \Omega$ can be represented as $\alpha = \mathcal{P}_\Omega(\beta)$ for some $\beta \in \mathbb{R}^B$. We are therefore motivated to introduce the unconstrained optimization problem:

$$\min_{\beta} E\left(\Phi_i^T \mathcal{P}_\Omega(\beta) - \psi_{2,i}^{\text{opt}}(\boldsymbol{e}(i))\right)^2. \tag{17}$$

With

$$P \triangleq I_B - \frac{\mathbb{1}\mathbb{1}^T}{B}, \quad S \triangleq ss^T,$$

and using (13), the steepest-descent recursion that solves (17) is of the form:

$$\begin{cases} \beta_i &= \beta_{i-1} - 2\gamma(i) P\left\{E\left[\boldsymbol{e}^2(i)\right] S \mathcal{P}_\Omega(\beta_{i-1}) - s\right\} \\ \alpha_i &= \mathcal{P}_\Omega(\beta_i) \end{cases} \tag{18}$$

where we have appealed to (11); and $\gamma(i)$ is a non-negative step-size sequence, the computation of which is discussed further ahead. Note that if β_{-1} is chosen from Ω, then it is ensured that $\beta_i \in \Omega$ for all i. The recursion in (18) then becomes

$$\alpha_i = \alpha_{i-1} - 2\gamma(i) P\left\{E\left[\boldsymbol{e}^2(i)\right] S \alpha_{i-1} - s\right\},\ \alpha_{-1} \in \Omega. \tag{19}$$

At each time index i, $E\left[\boldsymbol{e}^2(i)\right]$ in (19) may be estimated by means of the following smoothing operation:

$$E\left[\boldsymbol{e}^2(i)\right] \approx \hat{\sigma}_e^2(i) = \nu \hat{\sigma}_e^2(i-1) + (1-\nu)\, e^2(i) \tag{20}$$

with $\nu \in (0,1)$ and usually close to 1. In this case, we can replace (19) by

$$\alpha_i = \alpha_{i-1} - 2\gamma(i) P\left[\hat{\sigma}_e^2(i) S \alpha_{i-1} - s\right], \quad \alpha_{-1} \in \Omega. \tag{21}$$

We are now in a position to incorporate the non-negativity constraint on the entries of α_i at each iteration. One way to accomplish this task is to start from an initial condition $\alpha_{-1} \in \Omega_+$ and construct the step-size sequence in (21) as follows:

$$\gamma(i) \triangleq \gamma \frac{\min\{\alpha_{i-1}(b)\,|1 \le b \le B\}}{\left\|2P\left[\hat{\sigma}_e^2(i)\, S\alpha_{i-1} - s\right]\right\|_\infty + \epsilon} \tag{22}$$

where $\gamma \in (0,1)$ and $\epsilon > 0$ are constants, with the latter chosen very small to prevent division by zero; and $\|.\|_\infty$ denotes the maximum absolute entry of its vector argument. Using the α_i from (21), we can then compute the step-size as $\mu(i) = \mu\left(\alpha_i^T s\right)$ in the update (15) for w_i.

B. The RLMS Algorithm

1) *Initializations*: w_{-1}, μ, ν, i_o, B, $\alpha_{i_o-1} \in \Omega_+$, P, γ
2) *Estimation of* s*:*
   a) For a few iterations until some time index i_o, run the regular LMS algorithm (3) and the smoothing operation (20) to estimate $\hat{\sigma}_e^2(i)$, i.e., for each time index $0 \leq i < i_o$, compute:
$$\begin{cases} e(i) &= d(i) - u_i w_{i-1} \\ w_i &= w_{i-1} + \mu u_i^T e(i) \\ \hat{\sigma}_e^2(i) &= \nu \hat{\sigma}_e^2(i-1) + (1-\nu) e^2(i) \end{cases}$$
   b) Select the entries of the vector $s = \text{col}\left\{\frac{1}{\sigma_b^2}\right\}$ as the inverse of B equi-spaced points on the interval $\hat{\sigma}_e^2(i_o-1)[I_{\min}, I_{\max}]$, for some positive scalars $I_{\min}$ and $I_{\max}$, forming $\hat{s}$ and the matrix $\hat{S} = \hat{s}\hat{s}^T$.
3) For subsequent time instants $i \geq i_o$, compute:
$$\begin{cases} e(i) &= d(i) - u_i w_{i-1} \\ \hat{\sigma}_e^2(i) &= \nu \hat{\sigma}_e^2(i-1) + (1-\nu) e^2(i) \\ \delta_i &= 2P\left[\hat{\sigma}_e^2(i)\hat{S}\alpha_{i-1} - \hat{s}\right] \\ \gamma(i) &= \gamma \frac{\min\{\alpha_{i-1}(b), 1\leq b\leq B\}}{\|\delta_i\|_\infty + \epsilon} \\ \alpha_i &= \alpha_{i-1} - \gamma(i)\delta_i \\ \mu(i) &= \mu\left(\alpha_i^T \hat{s}\right) \\ w_i &= w_{i-1} + \mu(i) u_i^T e(i) \end{cases}$$

III. Performance Analysis

In this section, we analyze briefly the mean convergence behavior and steady-state performance of the RLMS algorithm under the data model introduced in Section II. Moreover, the variable s is treated as a deterministic quantity for analytical tractability. Let $\tilde{\boldsymbol{w}}_i \triangleq w^o - \boldsymbol{w}_i$.

A. Mean Stability

We make the following additional assumptions:

(*A1*) The regressors, $\boldsymbol{u}_i$, are independently and identically distributed (i.i.d.), which implies that $\boldsymbol{u}_i$ and $\tilde{\boldsymbol{w}}_{i-1}$ are independent of each other.

(*A2*) The variables $\boldsymbol{u}_i$ and $\tilde{\boldsymbol{w}}_{i-1}$ are also independent of $\boldsymbol{\alpha}_i$.

Both assumptions are reasonable near steady-state and for small step-size parameter, μ [3], [15]. Subtracting both sides of (15) from w^o and referring to (1), we get

$$\tilde{\boldsymbol{w}}_i = \left[I_M - \boldsymbol{\mu}(i)\boldsymbol{u}_i^T\boldsymbol{u}_i\right]\tilde{\boldsymbol{w}}_{i-1} - \boldsymbol{\mu}(i)\boldsymbol{u}_i^T\boldsymbol{v}(i), \qquad (23)$$

which, under expectation, gives

$$E\tilde{\boldsymbol{w}}_i = \{I_M - E[\boldsymbol{\mu}(i)]R_u\}E\tilde{\boldsymbol{w}}_{i-1}. \qquad (24)$$

Let $\{\lambda_m\}$, $m \in \{1, \ldots, M\}$, denote the eigenvalues of R_u. From [3] and [20], one sufficient condition for the asymptotic unbiasedness of (23), i.e., $\lim_{i\to\infty} E\tilde{\boldsymbol{w}}_i = 0$ irrespective of the initial condition, w_{-1}, is for there to exist a time index i^* and a number $0 < a < 1$, such that $|1 - E[\boldsymbol{\mu}(i)]\lambda_m| \leq a < 1$, for all $i > i^*$ and all $m \in \{1, \ldots, M\}$. This translates into the requirement:

$$\frac{1-a}{\lambda_{\min}} \leq E[\boldsymbol{\mu}(i)] \leq \frac{1+a}{\lambda_{\max}} \qquad (25)$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues of R_u, respectively. For example, if $R_u = \sigma_u^2 I_M$, then this condition requires selecting μ small enough to ensure $\frac{1-a}{\sigma_u^2} \leq E[\boldsymbol{\mu}(i)] \leq \frac{1+a}{\sigma_u^2}$. An alternative *approximate* argument to deduce asymptotic unbiasedness is to assume that, as $i \to \infty$, the recursion (21) for α_i approaches a fixed point $\bar{\alpha}$, which would need to satisfy $\hat{\sigma}_e^2 S\bar{\alpha} = s$, where $\hat{\sigma}_e^2$ denotes the steady-state variance of $\boldsymbol{e}(i)$. It follows that $\bar{\alpha}$ satisfies $s^T\bar{\alpha} = \frac{1}{\hat{\sigma}_e^2}$. If we adopt the approximation that $\mu\left(\alpha_i^T s\right) \to \frac{\mu}{\hat{\sigma}_e^2}$, as $i \to \infty$, then recursion (23) can be approximated in the limit by

$$\tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i-1} = \left[I_M - \frac{\mu}{\hat{\sigma}_e^2}\boldsymbol{u}_i^T\boldsymbol{u}_i\right]\tilde{\boldsymbol{w}}_{i-1} - \frac{\mu}{\hat{\sigma}_e^2}\boldsymbol{u}_i^T\boldsymbol{v}(i), \quad (26)$$

so that we may replace (24) by the following approximate mean recursion in the limit

$$E\tilde{\boldsymbol{w}}_i = \left(I_M - \frac{\mu}{\hat{\sigma}_e^2}R_u\right)E\tilde{\boldsymbol{w}}_{i-1}, \quad i \to \infty. \qquad (27)$$

From this relation we conclude that sufficiently small step-sizes μ ensure asymptotic mean stability such as selecting μ to satisfy $\mu < \frac{2\hat{\sigma}_e^2}{\lambda_{\max}}$.

B. Steady-State Performance

From model (1), it holds that $\boldsymbol{e}(i) = \boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1} + \boldsymbol{v}(i) = \boldsymbol{e}_a(i) + \boldsymbol{v}(i)$, where $\boldsymbol{e}_a(i)$ is the a priori estimation error. Under the data model assumptions in Section II, it holds that $E\left[\boldsymbol{e}^2(i)\right] = E\left[\boldsymbol{e}_a^2(i)\right] + \sigma_v^2$. Let MSE $\triangleq \lim_{i\to\infty} E\left[\boldsymbol{e}^2(i)\right]$, and EMSE $\triangleq \lim_{i\to\infty} E\left[\boldsymbol{e}_a^2(i)\right]$, yielding MSE $=$ EMSE $+ \sigma_v^2$, where EMSE stands for excess mean-square error. We replace (*A1*) and (*A2*) by the following generally weaker assumptions, which are required to hold only at steady-state as $i \to \infty$:

(*A1**) $\|\boldsymbol{u}_i\|^2$ is asymptotically independent of $\boldsymbol{\alpha}_i$ and $\boldsymbol{e}_a(i)$.

(*A2**) $\boldsymbol{\alpha}_i$ is asymptotically independent of $\boldsymbol{e}(i)$.

These assumptions are reasonable under small μ, and the second assumption more so when ν is close to 1 [3], [15]. Let ζ^{RLMS} denote the EMSE of RLMS. Following the energy conservation framework of [3], the following variance relation holds at steady-state:

$$\mu E\left[\left(s^T\boldsymbol{\alpha}_i\right)^2\|\boldsymbol{u}_i\|^2\boldsymbol{e}^2(i)\right] = 2E\left[\left(s^T\boldsymbol{\alpha}_i\right)\boldsymbol{e}_a(i)\boldsymbol{e}(i)\right]. \qquad (28)$$

Under (*A1**) and (*A2**), expression (28) reduces to

$$\mu\,\text{Tr}(R_u)\left(\zeta^{\text{RLMS}} + \sigma_v^2\right)\lim_{i\to\infty} E\left(s^T\boldsymbol{\alpha}_i\right)^2 = 2\zeta^{\text{RLMS}}\lim_{i\to\infty} E\left(s^T\boldsymbol{\alpha}_i\right). \qquad (29)$$

It remains to approximate the first- and second-order moments of $\left(s^T\boldsymbol{\alpha}_i\right)$ at steady-state. First, note that $\lim_{i\to\infty}\hat{\boldsymbol{\sigma}}_e^2(i) \approx \lim_{i\to\infty} E\left[\boldsymbol{e}^2(i)\right] = \zeta^{\text{RLMS}} + \sigma_v^2$, so that $\lim_{i\to\infty}\hat{\boldsymbol{\sigma}}_e^2(i) \approx \sigma_v^2$ for small step-size μ. We hereby introduce the following assumption:

(*A3**) $\boldsymbol{\gamma}(i)$, $\hat{\boldsymbol{\sigma}}_e^2(i)$ and $\boldsymbol{\alpha}_{i-1}$ are asymptotically mutually independent.

From (21) and under (*A3**), it follows that

$$\lim_{i\to\infty} E\left(s^T\boldsymbol{\alpha}_i\right) \approx \frac{1}{\sigma_v^2}, \quad \lim_{i\to\infty} E\left(s^T\boldsymbol{\alpha}_i\right)^2 \approx \frac{1}{\sigma_v^4}. \tag{30}$$

Substituting into (29) yields

$$\zeta^{\text{RLMS}} \approx \frac{\mu\,\text{Tr}\left(R_u\right)}{2-\mu\frac{\text{Tr}(R_u)}{\sigma_v^2}}. \tag{31}$$

It is well-known that the EMSE of the traditional LMS algorithm is approximated by [3]:

$$\zeta^{\text{LMS}} \approx \frac{\mu\sigma_v^2\,\text{Tr}\left(R_u\right)}{2-\mu\,\text{Tr}\left(R_u\right)}. \tag{32}$$

Comparing (31) and (32), it is obvious that under small step-size μ, the RLMS algorithm is less sensitive to the noise variance. Had we instead used the approximation $E\left(s^T\boldsymbol{\alpha}_i\right) \to \zeta^{\text{RLMS}} + \sigma_v^2$, then (31) and (32) would be replaced by $\zeta^{\text{RLMS}} \approx \mu\,\text{Tr}(R_u)/2$ and $\zeta^{\text{LMS}} \approx \mu\sigma_v^2\,\text{Tr}(R_u)/2$. Note that the expression for ζ^{RLMS} in this case does not depend on σ_v^2.

Another performance metric of interest is the mean-square deviation (MSD), defined as MSD $\triangleq \lim_{i\to\infty} E\left\|\tilde{\boldsymbol{w}}_i\right\|^2$. It was shown in [21], in the context of LMS filters with error nonlinearities, that for long enough filters, the MSD can be approximated by

$$\text{MSD} \approx \frac{M\zeta}{\text{Tr}\left(R_u\right)} \tag{33}$$

where ζ denotes the EMSE of the filter.

IV. Simulations

We compare the transient and steady-state performance of the LMS and RLMS algorithms, and verify expressions (31) through (33) for the EMSE and MSD. We consider a system identification setup, where the aim is to estimate a randomly initialized unit-norm w^o of size $M = 3$. The regressors $\{\boldsymbol{u}_i\}$ are i.i.d. zero-mean Gaussian vectors with covariance $\sigma_u^2 I_M$. The noise samples $\{\boldsymbol{v}(i)\}$ are drawn independently of the regressors and are i.i.d. according to an ε-contaminated Gaussian mixture model with pdf

$$f_{\boldsymbol{v}}(v) = (1-\varepsilon)\,\mathcal{N}\left(0, \bar{\sigma}_v^2\right) + \varepsilon\mathcal{N}\left(0, \kappa\bar{\sigma}_v^2\right)$$

where $\bar{\sigma}_v^2$ is the nominal noise variance, ε is the contamination ratio, and $\kappa \gg 1$. The signal-to-noise ratio (SNR) is defined as the ratio $\frac{\sigma_u^2}{\bar{\sigma}_v^2}$, while the effective noise variance is given by $\sigma_v^2 = (1-\varepsilon)\,\bar{\sigma}_v^2 + \varepsilon\kappa\bar{\sigma}_v^2$, with $\bar{\sigma}_v^2$ set to 1 throughout. We consider the case with two basis score functions, i.e., $B = 2$. The initial estimate for w^o is set to $w_{-1} = 0$. For the RLMS algorithm, the second stage following LMS is triggered at time index $i_o = 100$. For the smoothing operation (20), ν is set to 0.9 and $\hat{\sigma}_e^2(-1)$ to 0. For the estimation of s, we set I_{min} and I_{max} to 1 and 10, respectively. The initial estimate of α, α_{i_o-1}, is set to $\frac{1}{B}\mathbb{1}$. The step-size parameters, μ and γ, are both set to 0.01 unless mentioned otherwise. All simulation results are obtained by averaging over 1000 experiments.

The transient MSD and EMSE for LMS and RLMS are compared at SNR $= 10$ dB in the presence of uncontaminated Gaussian noise ($\varepsilon = 0$) in Fig. 1, and in the presence of Gaussian mixture noise, with $\varepsilon = 0.1$ and $\kappa = 10$, in Fig. 2. Also plotted are the theoretical steady-state MSD and EMSE. Both algorithms exhibit the same performance under nominal conditions, albeit at the expense of additional complexity in the case of RLMS. However, RLMS exhibits a performance gain of about 3 dB when the noise deviates from Gaussianity.

In Fig. 3, LMS and RLMS are compared in terms of steady-state EMSE for increasing ε while $\kappa = 10$, at SNR $= 10$ dB. The simulated steady-state values are obtained by averaging 100 samples after convergence. It is evident that RLMS is insensitive to increasingly impulsive noise.

In Fig. 4, the SNR performance of the algorithms is compared in terms of their steady-state EMSE, with $\varepsilon = 0.1$ and $\kappa = 10$. RLMS exhibits an SNR gain of about 2 dB.

Finally, in Fig. 5, the algorithms are compared in terms of steady-state EMSE for increasing μ, at SNR $= 5$ dB with $\varepsilon = 0.1$ and $\kappa = 10$. The parameter γ is set equal to μ throughout. RLMS displays the same superiority with respect to mean stability.

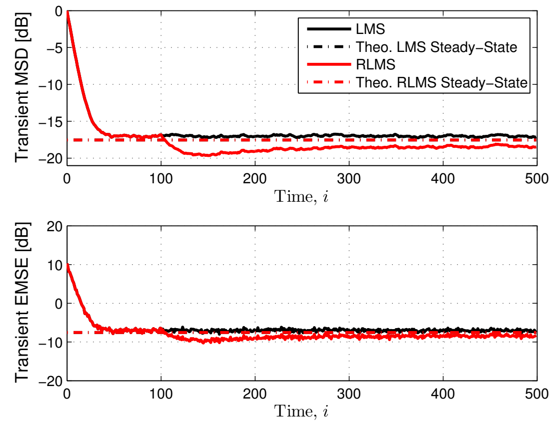

Fig. 1. Simulated MSD and EMSE learning curves (solid lines) and theoretical steady-state MSD and EMSE (dashed lines) for LMS (black) and RLMS (red) algorithms at SNR $= 10$ dB under uncontaminated Gaussian noise ($\varepsilon = 0$).

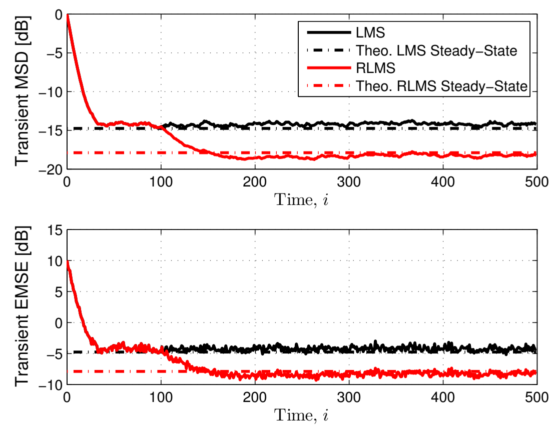

Fig. 2. Simulated MSD and EMSE learning curves (solid lines) and theoretical steady-state MSD and EMSE (dashed lines) for LMS (black) and RLMS (red) algorithms at SNR $= 10$ dB under Gaussian mixture noise, with $\varepsilon = 0.1$ and $\kappa = 10$.

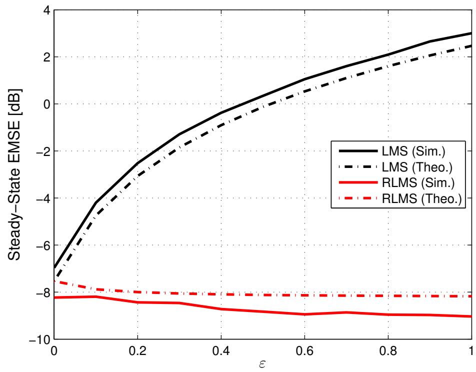

Fig. 3. Simulated and theoretical steady-state EMSE (solid and dashed lines, respectively) for LMS (black) and RLMS (red) algorithms, for increasing ε, with $\kappa = 10$, at SNR $= 10$ dB.

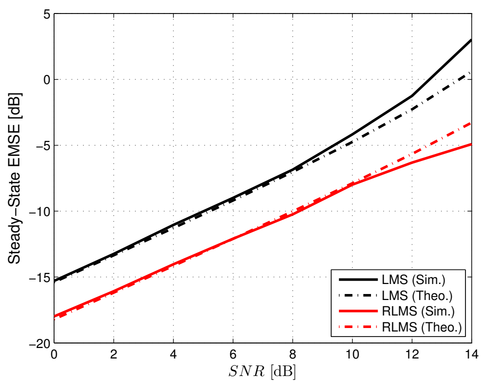

Fig. 4. Simulated and theoretical steady-state EMSE (solid and dashed lines, respectively) for LMS (black) and RLMS (red) algorithms, for increasing SNR, with $\varepsilon = 0.1$ and $\kappa = 10$.

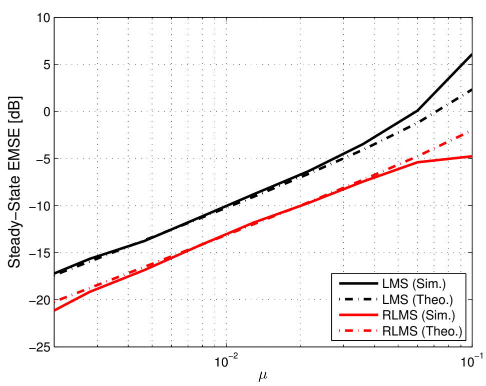

Fig. 5. Simulated and theoretical steady-state EMSE (solid and dashed lines, respectively) for LMS (black) and RLMS (red) algorithms, for increasing μ, at SNR $= 5$ dB with $\varepsilon = 0.1$ and $\kappa = 10$.

REFERENCES

[1] A. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, Jul. 2012.

[2] A. Zoubir and R. Brcich, "Multiuser detection in heavy tailed noise," *Digital Signal Processing*, vol. 12, no. 2-3, pp. 262–273, 2002.

[3] A. H. Sayed, *Adaptive Filters*. NJ: Wiley, 2008.

[4] N. Bershad, "On error saturation nonlinearities for LMS adaptation in impulsive noise," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4526–4530, Sep. 2008.

[5] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 46–48, Feb. 1997.

[6] Y. Zou, S.-C. Chan, and T.-S. Ng, "Least mean M-estimate algorithms for robust adaptive filtering in impulse noise," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 12, pp. 1564–1569, Dec. 2000.

[7] T. Gänsler, S. Gay, M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

[8] L. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, May 2008.

[9] S. Douglas and T.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1335–1351, Jun. 1994.

[10] T. Y. Al-Naffouri and A. H. Sayed, "Adaptive filters with error nonlinearities: Mean-square analysis and optimum design," *EURASIP Journal on Advances in Signal Processing*, vol. 2001, no. 1, pp. 192–205, Jan. 2001.

[11] P. Huber and E. Ronchetti, *Robust Statistics*. NJ: Wiley, 2009.

[12] G. Shevlyakov and N. Vilchevski, *Robustness in Data Analysis: Criteria and Methods*. The Netherlands: VSP, 2002.

[13] A. Taleb, R. Brcich, and M. Green, "Suboptimal robust estimation for signal plus noise models," in *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 2, Nov. 2000, pp. 837–841.

[14] R. Brcich and A. Zoubir, "Robust estimation with parametric score function estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, May 2002, pp. 1149–1152.

[15] R. Kwong and E. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.

[16] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Mar. 1997.

[17] D. Pazaitis and A. Constantinides, "A novel kurtosis driven variable step-size adaptive algorithm," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 864–872, Mar. 1999.

[18] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132 – 135, Feb. 2004.

[19] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4795–4810, Sep. 2010.

[20] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.

[21] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653–663, Mar. 2003.