Proportionate Adaptive Filtering Algorithms Derived Using an Iterative Reweighting Framework

Ching-Hua Lee ⬤ , *Student Member, IEEE*, Bhaskar D. Rao, *Fellow, IEEE*, and Harinath Garudadri, *Member, IEEE*

Abstract—In this paper, based on sparsity-promoting regularization techniques from the sparse signal recovery (SSR) area, least mean square (LMS)-type sparse adaptive filtering algorithms are derived. The approach mimics the iterative reweighted ℓ_2 and ℓ_1 SSR methods that majorize the regularized objective function during the optimization process. We show that introducing the majorizers leads to the same algorithm as simply using the gradient update of the regularized objective function, as is done in existing approaches. Different from the past works, the reweighting formulation naturally leads to an affine scaling transformation (AST) strategy, which effectively introduces a diagonal weighting on the gradient, giving rise to new algorithms that demonstrate improved convergence properties. Interestingly, setting the regularization coefficient to zero in the proposed AST-based framework leads to the Sparsity-promoting LMS (SLMS) and Sparsity-promoting Normalized LMS (SNLMS) algorithms, which exploit but do not strictly enforce the sparsity of the system response if it already exists. The SLMS and SNLMS realize proportionate adaptation for convergence speedup should sparsity be present in the underlying system response. In this manner, we develop a new way for rigorously deriving a large class of proportionate algorithms, and also explain why they are useful in applications where the underlying systems admit certain sparsity, e.g., in acoustic echo and feedback cancellation.

Index Terms—Affine scaling, iterative reweighted, proportionate adaptation, sparse adaptive filter, sparse signal recovery.

I. INTRODUCTION

ADAPTIVE filters [1]–[4] have been an active research area over the past few decades for their capabilities of estimating and tracking time-varying systems. In several applications, the impulse responses (IRs) of the underlying systems to be identified are often sparse or compressible (quasi-sparse), i.e., only a small percentage of the IR components have a significant magnitude while the rest are zero or small. Examples include network and acoustic echo cancellation [5]–[7], hands-free mobile telephony [8], acoustic feedback reduction in

hearing aids [9], and underwater acoustic communications [10], to mention a few. Designing adaptive filters that can exploit the sparse structure of the underlying system response for performance improvement over the conventional approaches, e.g., the least mean square (LMS) and normalized LMS (NLMS), is of great interest and importance especially for acoustic and speech applications. In this paper, we utilize the iterative reweighted ℓ_2 and ℓ_1 algorithms that have been developed in the sparse signal recovery (SSR) area to minimize diversity measures as a starting point [11]. By incorporating an affine scaling transformation (AST) strategy [12], [13] into the algorithm design process, a new methodology for developing a large class of adaptive filters is presented that leverage the sparse nature of the system responses.

A. Related Work

An early and influential work on identifying sparse IRs is the proportionate NLMS (PNLMS) algorithm proposed by Duttweiler [5] for acoustic echo cancellation. The main idea behind the approach is to update each filter coefficient using a step size proportional to the magnitude of the estimated coefficient, as opposed to the NLMS which assigns a uniform adaptation gain to all coefficients. Consequently, when the system is sparse, larger coefficients are adapted using relatively large steps compared to the smaller ones with PNLMS. The overall convergence can thus be sped up by focusing on adjusting the significant coefficients, rather then treating them all equally as in NLMS. Although PNLMS was developed in an intuitive way, i.e., the equations used to calculate the proportionate factors that realize step-size control were not based on any optimization criterion but were based on good heuristics, it has motivated many new proportionate variants for sparse system identification. The proportionate class of algorithms represent an important subset among sparsity-aware adaptive filters.

The recent progress on SSR has led to a number of computational algorithms, e.g., [14]–[19], among others. This makes available a plethora of approaches for systematically designing sparsity-aware adaptive algorithms that are a natural complement to the SSR batch estimation techniques. As a result, different from the proportionate approaches, another class of sparse adaptive filters have been introduced by utilizing sparsity-inducing regularization to speed up the adaptation of near-zero coefficients in sparse systems. This has led to several sparse adaptive filtering algorithms and even obtaining a

general framework of adaptive filters that incorporate sparsity. SSR-motivated adaptive algorithms represent another important class of sparsity-aware adaptive filters. We now discuss a few works on the proportionate class followed by the SSR variants.

Several variations of the PNLMS have been proposed and [20] provides a good summary. Examples include the improved PNLMS (IPNLMS) [21], IPNLMS based on the ℓ_0 "norm"[1] (IPNLMS-ℓ_0) [22], etc. In [23], Martin *et al.* utilized a natural gradient framework to deduce adaptive filters having similar features to the PNLMS that can exploit the sparse structure. Rao and Song [24] and Jin [25] proposed a framework for promoting sparsity in adaptive filters based on minimizing diversity measures. The framework is quite general and encompasses a broad range of adaptive filtering algorithms having similarity with the PNLMS algorithm. Benesty *et al.* [26] derived the PNLMS from a different perspective by using a basis pursuit [17] formulation. Following them, Liu and Grant [27] proposed a general framework of proportionate adaptive filters based on convex optimization and sparseness measures, which covers many traditional proportionate algorithms.

Several SSR-inspired algorithms have been introduced by integrating a sparsity-inducing regularizer into the original LMS objective function to accelerate the convergence of near-zero coefficients in sparse systems. For example, Chen *et al.* [28] proposed the zero-attracting LMS (ZA-LMS) derived by including the ℓ_1 norm penalty in the objective function. They also proposed the reweighted ZA-LMS (RZA-LMS) obtained by incorporating the log-sum penalty. Later, using the approximation of the ℓ_0 "norm" as a sparsity-inducing term, Gu *et al.* [29] proposed the ℓ_0-LMS that is capable of better estimating sparse systems. In [30], the authors utilized the p-norm-like penalty and considered the quantitative learning of the regularizer. Another work in this area is the new reweighted ℓ_1 norm penalized LMS algorithm proposed and studied in [31] for improving the ZA-LMS and RZA-LMS.

Recently, some works have considered both proportionate adaptation and sparsity-inducing regularization together. For example, [32] presents a modified PNLMS update equation with a zero attractor as in the ZA-LMS for all the taps, derived by introducing a carefully constructed ℓ_1 norm penalty in the PNLMS objective function. Other than the ℓ_1 norm, [33], [34] apply the ℓ_p norm penalty to the PNLMS cost function and derive ℓ_p-norm-constrained proportionate algorithms for improved broadband multipath channel estimation and active noise control. [35] encompasses a number of sparsity-aware adaptive filtering algorithms that go beyond the LMS and NLMS, including proportionate and regularization-based approaches. [36], [37] provide a general framework to combine proportionate updates and sparsity-inducing regularizers. In Section III, we will derive algorithms whose update rules also consist of a proportionate term and another term due to regularization. However, our derivation follows a very different path from these previous works.

B. Contributions of the Paper

In this paper, inspired by the conceptual similarity with SSR,[2] our goal is to add to this interesting body of work on adaptive filtering and sparsity. The contributions of the paper are the following:

1) The sparsity-aware adaptive filters developed lie at the intersection of the proportionate-class and SSR-inspired adaptive algorithms and provide an interesting bridge. We start with the rigorous formulation of a regularization framework and derive novel sparse adaptive filtering algorithms. Specifically, based on diversity measure minimization in SSR, we adopt the iterative reweighted ℓ_2 and ℓ_1 approaches [11] and utilize an AST methodology [12], [13] in the algorithm development, naturally leading to a general class of proportionate adaptive filters. This is an unique feature of this work. The combination of AST and the reweighting frameworks contribute to the main innovation of our adaptive algorithm development framework.

2) Under the proposed framework, we introduce Sparsity-promoting LMS (SLMS) and Sparsity-promoting NLMS (SNLMS) algorithms that promote sparsity without having to bias the adaptation process by adopting $\lambda = 0$, where λ is the regularization coefficient associated with the sparsity penalty. This is not possible for the class of algorithms currently in existence that utilize a sparsity-inducing regularization penalty.[3] The SLMS and SNLMS can be viewed as realizing proportionate adaptation like the PNLMS class of algorithms [5]. Therefore, our framework provides theoretical support to existing proportionate algorithms which were mostly developed based on good heuristics rather than on optimization criteria, and paves the way for explaining why they are useful in circumstances where the channels to be estimated admit certain sparsity. More importantly, unlike most of them that design the proportionate factors heuristically, our SSR-motivated framework leads to a more systematic way of designing the factors, and permits incorporation of a broad class of diversity measures that have proved effective for SSR in our algorithms.

3) Compared to existing derivations of proportionate-type algorithms, using the proposed framework we derive the algorithms in a more natural way in terms of incorporating sparsity using a regularization framework. For instance, in some of the existing works modified objective functions were proposed that impose sparsity on the "change" of the filter rather than on the filter itself, e.g., [24]–[27], [32]. However, since the assumption is that the filter itself is sparse, the motivation for enforcing sparsity on the "change" rather than on the filter is not clear and at best indirect. In contrast, we work with the general mean

---

[1] The ℓ_0 "norm" of a vector is defined as the number of its nonzero entries. The quotation marks are used to warn that it is not a proper norm.

[2] This similarity has been noticed in [38] where sparse adaptive filtering techniques were utilized for solving the SSR problem. Here we take the opposite direction as we are interested in utilizing SSR techniques for assisting the adaptive filtering algorithms. Both cases exploit the connections between SSR and adaptive filtering but the objectives are different.

[3] The algorithms usually reduce to the standard LMS or NLMS algorithm if the regularization coefficient λ is set to zero.

squared error (MSE) criterion in which sparsity can be directly imposed via regularization on the filter.

4) Steady-state analysis of the proposed algorithms is conducted and simulation results are provided to demonstrate the effectiveness of the proposed algorithms compared to existing approaches. Examples with the acoustic channel response measured on a real-world hearing aid device using speech input are also presented.

A portion of this work has been previously published as a conference paper [39]. To a fuller extent, the current paper provides a general framework for incorporating flexible diversity measures into sparse adaptive filters, together with theoretical discussion and supporting simulated results.

C. Organization of the Paper

The rest of the paper is organized as follows. Section II provides background on adaptive filters and iterative reweighting SSR algorithms. Section III derives adaptive filters that incorporate sparsity based on diversity measure minimization by utilizing the reweighted ℓ_2 and ℓ_1 frameworks together with the AST methodology. Section IV introduces the SLMS and SNLMS that adopt $\lambda = 0$. Section V discusses the steady-state analysis. Section VI presents simulation results. Section VII concludes the paper.

D. Notation

Let $\mathbb{R}^M$ denote the M-dimensional real Euclidean space. $\mathbb{R}^{N \times M}$ denotes the set of $N \times M$ real matrices. $\mathbb{R}_+$ denotes the set of non-negative real numbers. Superscript ^{T} denotes the transpose of a vector or matrix. $\mathrm{E}[\cdot]$ denotes the mathematical expectation. Vectors and matrices are denoted by boldface lowercase and uppercase letters, respectively. Scalars are denoted by italics. For a vector $\mathbf{x} = [x_0, x_1, \ldots, x_{M-1}]^T \in \mathbb{R}^M$, the ℓ_p norm[4] (where $p > 0$) is defined as: $\|\mathbf{x}\|_p = (\sum_{i=0}^{M-1} |x_i|^p)^{1/p}$. We use $\mathrm{diag}\{x_i\}$ to denote the M-by-M diagonal matrix whose i-th diagonal element is x_i. We use $\mathrm{sgn}(\cdot)$ to denote the component-wise sign function. $\nabla_{\mathbf{x}}$ denotes the gradient operator[5] w.r.t. $\mathbf{x}$. d denotes the differential operator. $\mathrm{tr}(\mathbf{X})$ denotes the trace of a square matrix $\mathbf{X} \in \mathbb{R}^{M \times M}$. $\mathbf{I}$ denotes the identity matrix. $\mathbf{1}$ denotes the vector of all ones. $\mathbf{0}$ denotes the vector of all zeros. We use $\mathcal{N}(\cdot, \cdot)$ to denote the normal distribution with the first and second arguments being the mean and (co)variance, respectively.

II. BACKGROUND ON ADAPTIVE FILTERING AND SSR

We provide some preliminaries of adaptive filters in the context of system identification and present several examples of existing sparsity-aware adaptive filtering algorithms. We also discuss the iterative reweighting SSR frameworks for developing our adaptive algorithms in later sections.

---

[4]Note that $\|\mathbf{x}\|_p$ for $0 < p < 1$ does not satisfy the required axioms for a norm and therefore it is not technically a norm. For exposition simplicity, since the range of p considered is from 0 to 2, we use the norm terminology to cover this range.

[5]By abusing the notation we use $\nabla_{\mathbf{x}}$ also for the subgradient operator without explicitly noting it.

A. Adaptive Filtering for System Identification

Let $\mathbf{h}_n = [h_{0,n}, h_{1,n}, \ldots, h_{M-1,n}]^T$ denote the adaptive filter of length M at discrete time instant n. Assume the IR of the underlying system is $\mathbf{h}^o = [h_0^o, h_1^o, \ldots, h_{M-1}^o]^T$, and the model for the observed or desired signal is $d_n = \mathbf{u}_n^T \mathbf{h}^o + v_n$. $\mathbf{u}_n = [u_n, u_{n-1}, \ldots, u_{n-M+1}]^T$ is the vector containing the M most recent samples of the input signal u_n and v_n is an additive noise signal. The output of the adaptive filter $\mathbf{u}_n^T \mathbf{h}_n$ is subtracted from d_n to obtain the error signal $e_n = d_n - \mathbf{u}_n^T \mathbf{h}_n$. The goal in general is to sequentially update the coefficients of $\mathbf{h}_n$ upon the arrival of a new data pair $(\mathbf{u}_n, d_n)$, such that eventually $\mathbf{h}_n = \mathbf{h}^o$; i.e., to identify the unknown system.

The most classic adaptive filtering algorithms are the LMS and NLMS [1]–[3], which can be derived based on minimizing the MSE objective function:

$$\min_{\mathbf{h}} \quad J(\mathbf{h}) \triangleq \mathrm{E}\left[e_n^2\right] = \mathrm{E}\left[\left(d_n - \mathbf{u}_n^T \mathbf{h}\right)^2\right]. \tag{1}$$

The method of steepest descent (gradient descent) for optimizing (1) suggests the following recursion for updating the filter coefficients [2]:

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \frac{\mu}{2} \nabla_{\mathbf{h}} J(\mathbf{h}_n), \tag{2}$$

where $\mu > 0$ is the step size. To develop adaptive algorithms, in practice the gradient $\nabla_{\mathbf{h}} J(\mathbf{h}_n) = -2\mathrm{E}[\mathbf{u}_n e_n]$ is replaced by the instantaneous estimate $-2\mathbf{u}_n e_n$, i.e., the stochastic gradient [2], [3], leading to the standard LMS algorithm:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{u}_n e_n. \tag{3}$$

The normalized version of (3), i.e., the NLMS algorithm, can be derived based on the *principle of minimum disturbance* [2]. Alternatively, it can be obtained by performing *exact line search* for the optimal step size for each iteration [40]. Then, practically, a scaling factor $\tilde{\mu} > 0$ is introduced to exercise control over the adaptation[6] and a small regularization constant $\delta > 0$ is also employed to avoid division by zero [2], leading to the standard NLMS algorithm:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \frac{\tilde{\mu} \mathbf{u}_n e_n}{\mathbf{u}_n^T \mathbf{u}_n + \delta}. \tag{4}$$

1) Sparsity-Aware Adaptive Filtering Algorithms: When the underlying system response is sparse, a class of algorithms realizing proportionate adaptation [20] are able to take advantage of the structural sparsity. A typical update rule with proportionate adaptation is:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{\Gamma}_n \mathbf{u}_n e_n, \tag{5}$$

or the normalized version:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \frac{\tilde{\mu} \mathbf{\Gamma}_n \mathbf{u}_n e_n}{\mathbf{u}_n^T \mathbf{\Gamma}_n \mathbf{u}_n + \delta}, \tag{6}$$

where

$$\mathbf{\Gamma}_n = \mathrm{diag}\{\gamma_{0,n}, \gamma_{1,n}, \ldots, \gamma_{M-1,n}\} \tag{7}$$

---

[6]Formally, $\tilde{\mu}$ is called the *normalized* step size. For brevity, we still refer to it as the *step size* but keep in mind that it does not have the same significance as the μ in (3). Note that it is also common in the literature that the same notation of the step size is shared for both LMS and NLMS without explicit distinction.

is an M-by-M diagonal matrix assigning different weights to the step sizes for different filter taps, referred to as the *proportionate matrix*. It redistributes the adaptation gains among all coefficients and emphasizes the large ones in order to speed up their convergence. Typically, at the n-th iteration the diagonal entries are computed as:

$$\gamma_{i,n} = \frac{\pi_{i,n}}{\sum_{j=0}^{M-1} \pi_{j,n}}, \tag{8}$$

$\forall i = 0, 1, \ldots, M - 1$, where $\pi_{i,n}$ is algorithm-dependent and examples of such algorithms include the PNLMS [5], IPNLMS [21], IPNLMS-ℓ_0 [22], etc. In general, if the estimated filter coefficients $h_{i,n}$ are sparse, the resulting $\pi_{i,n}$ (thus $\gamma_{i,n}$) will also tend to be sparsely distributed (with positive values).

Another class of algorithms, inspired by developments in the SSR area, take sparsity into account using a regularization-based approach, e.g., [28]–[31]. The algorithms are obtained by adding a sparsity-inducing term $G(\mathbf{h})$ to the MSE objective function:

$$\min_{\mathbf{h}} \quad J^G(\mathbf{h}) \triangleq J(\mathbf{h}) + \lambda G(\mathbf{h}), \tag{9}$$

where $\lambda > 0$ is the regularization coefficient. By simply applying (stochastic) gradient descent[7] on (9):

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{u}_n e_n - \frac{\mu \lambda}{2} \nabla_{\mathbf{h}} G(\mathbf{h}_n), \tag{10}$$

various algorithms can be obtained with different sparsity-inducing functions $G(\cdot)$. Examples include the ZA-LMS [28], RZA-LMS [28], and ℓ_0-LMS [29], [41].

B. Iterative Reweighting Algorithms in SSR

The optimization of (9) is actually an SSR problem. The sparsity regularization term $G(\cdot)$ represents the *general diversity measure* that when minimized encourages sparsity in its argument. A *separable* function of the form $G(\mathbf{h}) = \sum_{i=0}^{M-1} g(h_i)$ is commonly used, where $g(\cdot)$ has the following properties [11]:

Property 1: $g(z)$ is symmetric, i.e., $g(z) = g(-z) = g(|z|)$;
Property 2: $g(|z|)$ is monotonically increasing with $|z|$;
Property 3: $g(0)$ is finite;
Property 4: $g(z)$ is concave in $|z|$ or z^2.

Any function that holds the above properties is a candidate for effective SSR algorithm development.

The concave nature of the regularization penalty $G(\mathbf{h})$ poses challenges to the diversity measure minimization problem (9). The iterative reweighted ℓ_2 [15], [18] and ℓ_1 [19] methods are popular batch estimation algorithms for solving such minimization problems in SSR. By introducing a weighted ℓ_2 or ℓ_1 norm term as an upper bound for the diversity measure term in each iteration, they form and solve a convex optimization problem at each step to approach the optimal solution [11]. Specifically, instead of (9), at iteration n the reweighted ℓ_2 framework suggests solving:

$$\min_{\mathbf{h}} \quad J_n^{\ell_2}(\mathbf{h}) \triangleq J(\mathbf{h}) + \lambda \|\mathbf{W}_n^{-1}\mathbf{h}\|_2^2, \tag{11}$$

and the reweighted ℓ_1 framework suggests solving:

$$\min_{\mathbf{h}} \quad J_n^{\ell_1}(\mathbf{h}) \triangleq J(\mathbf{h}) + \lambda \|\mathbf{W}_n^{-1}\mathbf{h}\|_1, \tag{12}$$

where $\mathbf{W}_n = \mathrm{diag}\{w_{i,n}\}$ is positive definite[8] and each $w_{i,n}$ is computed based on the current estimate $h_{i,n}$, depending on which framework (reweighted ℓ_2 or ℓ_1) and diversity measure (choice of $G(\cdot)$) are used.

To elaborate, for using the reweighted ℓ_2 (11), the diversity measure function $g(z)$ has to be concave in z^2 for Property 4; i.e., it satisfies $g(z) = f(z^2)$, where $f(z)$ is concave for $z \in \mathbb{R}_+$. Based on [11], we have $w_{i,n}$ given as:

$$w_{i,n} = \left(\left. \frac{\mathrm{d}f(z)}{\mathrm{d}z} \right|_{z=h_{i,n}^2} \right)^{-\frac{1}{2}}. \tag{13}$$

For using the reweighted ℓ_1 (12), $g(z)$ has to be concave in $|z|$ for Property 4; i.e., it satisfies $g(z) = f(|z|)$, where $f(z)$ is concave for $z \in \mathbb{R}_+$. In this case, $w_{i,n}$ is given as:

$$w_{i,n} = \left(\left. \frac{\mathrm{d}f(z)}{\mathrm{d}z} \right|_{z=|h_{i,n}|} \right)^{-1}, \tag{14}$$

To utilize the reweighting frameworks, we first choose an appropriate diversity measure $G(\mathbf{h})$ and then use (13) or (14) to obtain the corresponding update form of $\mathbf{W}_n$. Several examples will be presented is Section IV-B.

III. PROPOSED FRAMEWORK FOR INCORPORATING SPARSITY IN ADAPTIVE FILTERS

Our framework for developing sparse adaptive filters is also based on (9). However, we will be deriving algorithms in a different way rather than using a simple gradient descent as is typically done in existing regularization-based adaptive filtering approaches, e.g., (10). Our novel derivation consists of two stages: i) adapting the iterative reweighting frameworks [11] popular in SSR to the adaptive filtering setting, followed by ii) incorporating the AST strategy [12], [13] from the optimization literature to obtain new adaptive filtering algorithms.

A. Reweighting Methods for Adaptive Filtering

The reweighting methods introduced in Section II-B actually belong to the more general class of majorization-minimization (MM) algorithms [42]. In each iteration n, the weighted ℓ_2 or ℓ_1 norm term majorizes $G(\mathbf{h})$ at the current estimate $\mathbf{h}_n$, thereby providing a surrogate function (or majorizer) $J_n^{\ell_2}(\mathbf{h})$ or $J_n^{\ell_1}(\mathbf{h})$ for the regularized objective function $J^G(\mathbf{h})$. Sequentially minimizing the surrogate functions allows the algorithm to produce more focal estimates as optimization progresses. Hopefully when the number of iterations is large enough, the optimal solution can be well approached or even achieved [11].

In SSR, it is typical that the surrogate function is *exactly* minimized in each iteration n. For the purpose of developing

---

[7]By abusing the terminology we implicitly use "gradient" also for subgradient whenever appropriate.

[8]The positive definiteness can be shown to hold for a wide variety of diversity measures used in SSR. In cases where it is not, the positive definiteness can still be ensured by utilizing some small regularization constant.

adaptive filtering algorithms, here we consider performing *only one step of gradient descent per iteration*. In this sense, it corresponds to the *generalized* MM [43] where one does not need to minimize the majorizer but only to assure that it decreases in every iteration. Indeed, the MM viewpoint provides an interesting observation of using gradient descent for optimizing (9) and the reweighting formulations (11) and (12), as stated in the following proposition:

Proposition 1: For any point $\mathbf{h}_n$ at which $G(\mathbf{h})$ is differentiable, the gradient vector of the surrogate function $J_n^{\ell_2}(\mathbf{h})$ or $J_n^{\ell_1}(\mathbf{h})$ evaluated at $\mathbf{h}_n$ coincides with that of the regularized objective function $J^G(\mathbf{h})$, i.e., $\nabla_{\mathbf{h}} J_n^{\ell_2}(\mathbf{h}_n) = \nabla_{\mathbf{h}} J^G(\mathbf{h}_n)$ for the reweighted ℓ_2 case and $\nabla_{\mathbf{h}} J_n^{\ell_1}(\mathbf{h}_n) = \nabla_{\mathbf{h}} J^G(\mathbf{h}_n)$ for the reweighted ℓ_1 case.

Proof: Since the surrogate function majorizes $J^G(\mathbf{h})$ at $\mathbf{h}_n$, the tangent plane (supporting hyperplane) of the majorizer coincides with that of $J^G(\mathbf{h})$ at $\mathbf{h}_n$. Consequently, the gradient vectors are the same at $\mathbf{h}_n$.

The observation in Proposition 1 implies that, if the gradient descent (when using a fixed step size) is utilized for optimization,[9] then adopting the reweighting frameworks (11) and (12) will be equivalent to directly working on (9) and lead to the existing regularization-based algorithms such as the ZA-LMS. In the following, we introduce the AST strategy naturally suggested by the reweighting frameworks, leading to new algorithms markedly different from those obtained by directly optimizing (9) with gradient descent.

B. AST-Based Adaptive Filtering Algorithms

The reweighting frameworks (11) and (12) naturally suggest the following reparameterization in terms of the (affinely) scaled variable $\mathbf{q}$:

$$\mathbf{q} \triangleq \mathbf{W}_n^{-1}\mathbf{h}. \tag{15}$$

This step can be interpreted as the AST commonly employed by the interior point approach to solving linear and nonlinear programming problems [12], where $\mathbf{W}_n$ is used as the *scaling matrix*. It is pre-calculated and treated as a given matrix at iteration n to perform a change of coordinates (variables) [44] from $\mathbf{h}$ to $\mathbf{q}$, acting as a scaling technique in gradient descent methods [45]. In the optimization literature, AST-based methods transform the original problem into an equivalent one, favorably positioning the current point at the center of the feasible region for expediting the optimization process [13]. While we do not claim this argument is rigorous in the context of adaptive filtering, where the convergence behavior is hard to characterize due to the nonlinear nature of the update equations and the long term dependency on the data, the numerical results appear to support this observation of enjoying the benefits of AST for convergence speedup.

Now we apply (15) to reparameterize the objective functions $J_n^{\ell_2}(\mathbf{h})$ and $J_n^{\ell_1}(\mathbf{h})$ and perform minimization w.r.t. $\mathbf{q}$, that is:

$$\min_{\mathbf{q}} \quad \tilde{J}_n^{\ell_2}(\mathbf{q}) \triangleq J_n^{\ell_2}(\mathbf{W}_n\mathbf{q}) = J(\mathbf{W}_n\mathbf{q}) + \lambda\|\mathbf{q}\|_2^2 \tag{16}$$

and

$$\min_{\mathbf{q}} \quad \tilde{J}_n^{\ell_1}(\mathbf{q}) \triangleq J_n^{\ell_1}(\mathbf{W}_n\mathbf{q}) = J(\mathbf{W}_n\mathbf{q}) + \lambda\|\mathbf{q}\|_1, \tag{17}$$

for the reweighted ℓ_2 and ℓ_1 cases, respectively. The overall update process conceptually can be summarized as follows: i) given an $\mathbf{h}$ compute $\mathbf{W}_n$ followed by $\mathbf{q}$. ii) Update $\mathbf{q}$ using a gradient descent algorithm. iii) Use this new $\mathbf{q}$ to obtain the updated $\mathbf{h}$. iv) Repeat Steps i)–iii) till convergence.

More formally, to proceed with gradient-based updates, following [40] we define the *a posteriori* AST variable at time n:

$$\mathbf{q}_{n|n} \triangleq \mathbf{W}_n^{-1}\mathbf{h}_n \tag{18}$$

and the *a priori* AST variable at time n:

$$\mathbf{q}_{n+1|n} \triangleq \mathbf{W}_n^{-1}\mathbf{h}_{n+1}. \tag{19}$$

The recursive update by using gradient descent in the $\mathbf{q}$ domain can be formulated as:

$$\mathbf{q}_{n+1|n} = \mathbf{q}_{n|n} - \frac{\mu}{2}\nabla_{\mathbf{q}}\tilde{J}_n^{\ell_2}(\mathbf{q}_{n|n}) \tag{20}$$

and

$$\mathbf{q}_{n+1|n} = \mathbf{q}_{n|n} - \frac{\mu}{2}\nabla_{\mathbf{q}}\tilde{J}_n^{\ell_1}(\mathbf{q}_{n|n}), \tag{21}$$

for optimizing (16) and (17), respectively.

Using the chain rule[10] and the AST relationships (15) and (18), we can write (20) and (21) respectively as:

$$\mathbf{q}_{n+1|n} = \mathbf{q}_{n|n} - \frac{\mu}{2}\mathbf{W}_n\nabla_{\mathbf{h}}J_n^{\ell_2}(\mathbf{h}_n) \tag{22}$$

and

$$\mathbf{q}_{n+1|n} = \mathbf{q}_{n|n} - \frac{\mu}{2}\mathbf{W}_n\nabla_{\mathbf{h}}J_n^{\ell_1}(\mathbf{h}_n). \tag{23}$$

Premultiplying $\mathbf{W}_n$ on both sides of (22) and (23) and noting the relationships (18) and (19), we transform the $\mathbf{q}$ domain updates (22) and (23) back to the $\mathbf{h}$ domain respectively as:

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \frac{\mu}{2}\mathbf{W}_n^2\nabla_{\mathbf{h}}J_n^{\ell_2}(\mathbf{h}_n) \tag{24}$$

and

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \frac{\mu}{2}\mathbf{W}_n^2\nabla_{\mathbf{h}}J_n^{\ell_1}(\mathbf{h}_n). \tag{25}$$

By Proposition 1, we can replace $\nabla_{\mathbf{h}}J_n^{\ell_2}(\mathbf{h}_n)$ and $\nabla_{\mathbf{h}}J_n^{\ell_1}(\mathbf{h}_n)$ with $\nabla_{\mathbf{h}}J^G(\mathbf{h}_n)$. Thus, (24) and (25) can both be written as:

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \frac{\mu}{2}\mathbf{W}_n^2\nabla_{\mathbf{h}}J^G(\mathbf{h}_n). \tag{26}$$

Note that based on the aforementioned update process i)-iv), we can in fact directly apply (15) to reparameterize $J^G(\mathbf{h})$ to obtain (26) without going through the reweighting formulation, as long as the scaling matrix $\mathbf{W}_n$ is specified. In this sense, the reweighing methods essentially play the role of suggesting a suitable $\mathbf{W}_n$ that eventually becomes a diagonal *weighting matrix* $\mathbf{W}_n^2$ on the gradient $\nabla_{\mathbf{h}}J^G(\mathbf{h}_n)$ in the update rule. Hopefully, it alters the ordinary descent direction in such a way that leads to convergence improvement. We should also

---

[9]For a point at which $G(\mathbf{h})$ is non-differentiable, this can still hold by properly choosing the subgradients.

[10]Note that the chain rule here is basically $\nabla_{\mathbf{q}} = \mathbf{W}_n\nabla_{\mathbf{h}}$ as a result of the change of variables (15) for a given $\mathbf{W}_n$ at iteration n.

emphasize that the scaling matrix $\mathbf{W}_n$ suggested by (24) and (25) will in general be different for a given $G(\mathbf{h})$ despite the fact that both can be expressed as (26).

In practice, the following update rule is suggested over (26) for avoiding instability and slow convergence issues:

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \frac{\mu}{2}\mathbf{S}_n \nabla_{\mathbf{h}} J^G(\mathbf{h}_n), \qquad (27)$$

where

$$\mathbf{S}_n = \frac{\mathbf{W}_n^2}{\frac{1}{M}\mathrm{tr}\left(\mathbf{W}_n^2\right)}, \qquad (28)$$

referred to as the *sparsity-promoting matrix*, is the normalized version of $\mathbf{W}_n^2$. As a fixed step size μ is used, performing normalization of the weighting matrix compensates for any arbitrary scaling inherent in $\mathbf{W}_n^2$ that might cause instability (scaling too large) or slow convergence (scaling too small). Note that by (28) we always have $\mathrm{tr}(\mathbf{S}_n) = M$, aligned with the non-AST case (i.e., using the ordinary gradient descent) which essentially has $\mathbf{S}_n = \mathbf{I}$ whose trace is also M.

Finally, to obtain the adaptive algorithm, we follow the standard procedure of replacing $\nabla_{\mathbf{h}} J^G(\mathbf{h}_n) = -2\mathrm{E}[\mathbf{u}_n e_n] + \lambda \nabla_{\mathbf{h}} G(\mathbf{h}_n)$ in (27) with its instantaneous estimate $-2\mathbf{u}_n e_n + \lambda \nabla_{\mathbf{h}} G(\mathbf{h}_n)$, leading to:

$$\boxed{\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{S}_n \mathbf{u}_n e_n - \frac{\mu \lambda}{2}\mathbf{S}_n \nabla_{\mathbf{h}} G(\mathbf{h}_n).} \qquad (29)$$

We see that there is a term with a diagonal weighting $\mathbf{S}_n$ on the LMS update vector $\mathbf{u}_n e_n$, similar to that in proportionate algorithms (5) and (6). We also see another term weighted by λ which is due to the introduction of the regularizer, like that of (10). Therefore, the AST framework leads to a more general algorithm comprised of proportionate adaptation and sparsity-inducing regularization. We thus refer to (29) as the *generalized sparse LMS* algorithm.

C. Discussions

It may seem at the first glance that applying the reweighting techniques to (9) straightforwardly leads to our algorithm. We stress that it is not true. If the AST (15) was not considered, adopting the reweighting schemes would still end up with an update rule like (10) according to Proposition 1, rather than the proposed (29). It is also worth mentioning that there is considerable difference between the proposed algorithm (29) and existing SSR algorithms based on (11) and (12) – the conventional SSR techniques are batch estimation methods for recovering the underlying sparse representation, while the proposed algorithm is specifically tailored for the adaptive filtering scenario. That being said, as gradient descent is adopted for optimization, we actually perform a gradual update of the filter coefficients in each iteration n, rather than looking for an exact minimizer of the surrogate function as is typically pursued in SSR. This enables the algorithm to track temporal variations and environmental changes. Certainly, considering the gradient noise in real scenarios, it may post the issue of whether the algorithm is convergent. However, even the standard LMS and NLMS that are based on gradual updates, work well in many

practical situations with gradient noise. In Section VI, experimental results will demonstrate that the proposed algorithm, like the LMS and NLMS, also behaves well when certain level of environmental noise is present.

Finally, the following theorem establishes the convergence of the $\mathbf{q}$ domain recursions (20) and (21) and their relationships to (9) to shed light on the convergence of the adaptive algorithm (29) developed based on them:

Theorem 1: For the objective function $J^G(\mathbf{h})$ in (9) with the general diversity measure $G(\mathbf{h})$ satisfying Properties 1-4 in Section II-B,[11] there exists a step size sequence $\{\mu_n\}_{n=0}^{\infty}$ such that each of the update recursions (20) and (21) monotonically converges to a local minimum (or saddle point) of (9) under a wide-sense stationary (WSS) environment, i.e., $\mathbf{u}_n$ and d_n are jointly WSS.

Proof: See Appendix A.

IV. SPARSITY-PROMOTING ALGORITHMS ADOPTING $\lambda = 0$

An interesting situation arises when we consider the limiting case of $\lambda \to 0^+$ for the proposed framework. By setting $\lambda = 0$ in (29), we see the λ-weighted term due to regularization vanishes, leading to a simpler equation:

$$\boxed{\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{S}_n \mathbf{u}_n e_n.} \qquad (30)$$

The main feature of (30) is that it is able to *promote* sparsity of the system (through $\mathbf{S}_n$) if it already exists while *not strictly enforcing* it (as $\lambda = 0$). This property shall become clearer in later discussions. We refer to the algorithm (30) as the Sparsity-promoting LMS (SLMS).

The normalized version of (30) can also be developed by performing exact line search for the optimal step size at iteration n just like that when deriving the NLMS:

$$\mu_n = \arg\min_{\boldsymbol{\mu}} \left(d_n - \mathbf{u}_n^T\left(\mathbf{h}_n + \mu \mathbf{S}_n \mathbf{u}_n e_n\right)\right)^2 = \frac{1}{\mathbf{u}_n^T \mathbf{S}_n \mathbf{u}_n}. \qquad (31)$$

Similar to the NLMS, we introduce $\tilde{\mu} > 0$ to exercise control over the adaptation and $\delta > 0$ to avoid division by zero, resulting in:

$$\boxed{\mathbf{h}_{n+1} = \mathbf{h}_n + \frac{\tilde{\mu}\,\mathbf{S}_n \mathbf{u}_n e_n}{\mathbf{u}_n^T \mathbf{S}_n \mathbf{u}_n + \delta}.} \qquad (32)$$

We refer to the algorithm (32) as the Sparsity-promoting NLMS (SNLMS).

An obvious benefit of adopting $\lambda = 0$ is that the computation for the term due to regularization is no longer needed, and we do not have to tweak this coefficient anymore (which is typically not a trivial task in practice). Still, the SLMS and SNLMS have the ability to leverage sparsity owing to the diagonal weighting $\mathbf{S}_n$, which is similar to the proportionate matrix $\mathbf{\Gamma}_n$ in (5) and (6). Again, this is made possible due to the use of the AST (15), wherein the gradient descent update is performed w.r.t. the $\mathbf{q}$ variable rather than in the original $\mathbf{h}$ domain. Otherwise, we

---

[11]Note that for Property 4, Theorem 1 holds for (20) of the reweigthed ℓ_2 framework if $g(z)$ is concave in z^2. On the other hand, it holds for (21) of the reweighted ℓ_1 framework if $g(z)$ is concave in $|z|$.

will end up with algorithms like (10) that reduce to the ordinary LMS/NLMS when using $\lambda = 0$.

The SLMS and SNLMS can in fact be viewed as a broader class of proportionate algorithms. Actually, with certain choices of diversity measures and corresponding parameters, we can have the PNLMS (approximately) as a special case. For example, as we will see in Section IV-B, using $p = 1$ in (34) for $\mathbf{W}_n$, the sparsity-promoting matrix $\mathbf{S}_n$ approximates the proportionate matrix $\mathbf{\Gamma}_n$ of the PNLMS. Indeed, one of the main advantages of the SLMS and SNLMS is their ability to incorporate flexible diversity measures. It allows the algorithms to fit the sparsity level of the system response by optimizing corresponding sparsity control parameters in a more informed manner due to the underlying connections to SSR. Furthermore, the derivations provide theoretical support to the class of proportionate algorithms that were mostly motivated based on heuristics, explaining why they are useful in practical identification tasks with sparse channels, e.g., in acoustic echo/feedback cancellation, from an SSR viewpoint.

A. Interpretation of $\lambda = 0$ From Optimization Perspective

We further discuss the interpretation of using $\lambda = 0$ in our framework from an optimization perspective. Recall that the AST reparameterization (15) results in the optimization problems (16) and (17). Setting $\lambda = 0$ leads both to:

$$\min_{\mathbf{q}} J(\mathbf{W}_n \mathbf{q}). \tag{33}$$

This actually applies a change of coordinates to the unregularized problem (1) via (15). Since $\mathbf{W}_n$ is invertible, the problem of finding the $\mathbf{h}$ that minimizes $J(\mathbf{h})$ is equivalent to finding the $\mathbf{q}$ which minimizes $J(\mathbf{W}_n \mathbf{q})$. Therefore, the advantage of solving (33) is that the solution is guaranteed to also be a solution of (1), which is not true for (9) with $\lambda > 0$. Thus, the optimization is *unbiased* while promoting sparsity – it is able to take advantage of sparsity whereas without having to supplement a sparsity penalty that incurs bias to the MSE objective. As noted in [45], the performance of gradient-based methods is dependent on the parameterization – a new choice may substantially alter convergence characteristics. Introducing variable scalings may speed up convergence by altering the descent direction toward the optimum. In our case, solving (33) with appropriately selected $\mathbf{W}_n$ can expedite the adaptation procedure toward the optimum of (1).

This observation can also be illustrated by looking at (9) which indicates a trade-off between estimation quality, as reflected in the MSE objective function, and solution sparsity as controlled by λ. In the limiting case of $\lambda \to 0^+$, the objective function exerts diminishing impact on enforcing sparsity on the solution, meaning that eventually no sparse solution is favored over other possible solutions. To elaborate, with $\lambda = 0$ and under a WSS environment, all the algorithms derived from (9) minimize the MSE and converge toward the Wiener-Hopf solution. However, not surprisingly, the path they take is different and depends on how the iterations are developed. If the Wiener-Hopf solution is sparse, then all will converge toward the same sparse solution asymptotically. Interestingly, the SLMS and SNLMS, because

of their proportionate nature similar to the PNLMS-type algorithms, can take advantage of the sparsity and are capable of speeding up convergence without compromising estimation quality should sparsity be present. This observation will later be supported by experimental results in Section VI-B.

B. Example Diversity Measures and Corresponding $\mathbf{W}_n$

To illustrate the flexibility of the proposed framework, we provide example algorithms instantiated with popular diversity measures that have proved effective in SSR.

Consider the p-norm-like diversity measure with $g(h_i) = |h_i|^p$, $0 < p \leq 2$ for the reweighted ℓ_2 framework [12], [15]. Using (13) leads to the update form of $\mathbf{W}_n$:

$$w_{i,n} = \left(\frac{2}{p} \left(|h_{i,n}| + c \right)^{2-p} \right)^{\frac{1}{2}}. \tag{34}$$

Note that we have empirically added a small regularization constant $c > 0$ for avoiding algorithm stagnation and instability,[12] which also ensures the positive definiteness of $\mathbf{W}_n$ [39]. The parameter $p \in (0, 2]$ in (34) is responsible for controlling the sparsity degree, as the p-norm-like diversity measure is associated with super-Gaussian prior distributions. In general, a smaller p corresponds to a heavier-tailed distribution, encouraging stronger sparsity in the parameters. It is worth noting that using $p \to 1$ in (34) results in a proportionate factor close to that of the PNLMS. On the other hand, letting $p = 2$ recovers the standard LMS/NLMS.

The p-norm-like diversity measure can also be adopted in the reweighted ℓ_1 framework if $0 < p \leq 1$. Applying (14), we obtain the update form of $\mathbf{W}_n$ in this case:

$$w_{i,n} = \frac{1}{p} \left(|h_{i,n}| + c \right)^{1-p}. \tag{35}$$

Again, a small constant $c > 0$ is added. The sparsity control parameter of (35) is now $p \in (0, 1]$. In this case, using $p \to 0.5$ in (35) results in a proportionate factor close to that of the PNLMS, whereas letting $p = 1$ recovers the standard LMS/NLMS.

We can also consider the log-sum penalty with $g(h_i) = \log(h_i^2 + \epsilon)$, $\epsilon > 0$ for the reweighted ℓ_2 framework [18]. The function is readily amenable to the use of (13) to obtain the update form of $\mathbf{W}_n$ as:

$$w_{i,n} = \left(h_{i,n}^2 + \epsilon \right)^{\frac{1}{2}}. \tag{36}$$

Or consider the log-sum penalty with $g(h_i) = \log(|h_i| + \epsilon)$, $\epsilon > 0$ for the reweighted ℓ_1 framework [19]. Using (14), the update form of $\mathbf{W}_n$ becomes:

$$w_{i,n} = |h_{i,n}| + \epsilon. \tag{37}$$

The sparsity control parameter is $\epsilon > 0$ for the two log-sum penalty cases. From (36) and (37) we can see that ϵ controls how much proportionate adaptation is encouraged: as ϵ becomes smaller, the term $h_{i,n}^2$ or $|h_{i,n}|$ becomes more dominant. Consequently, they exhibit a stronger proportionate adaptation

---

[12]We suggest that c be kept relatively small as compared to the amplitude of the filter coefficients so that it would not affect the convergence significantly.

characteristic. On the contrary, as ϵ becomes larger, the influence of $h_{i,n}^2$ or $|h_{i,n}|$ reduces. Thus, the algorithm will approach the standard LMS/NLMS when $\epsilon \gg h_{i,n}^2$ or $\epsilon \gg |h_{i,n}|$. In practice, one can start from a large ϵ and reduce it to find a suitable value.

More example functions can be found in [27], [46], including $g(h_i) = \arctan(|h_i|/\epsilon)$, $\epsilon > 0$ also suggested in [19], which works for both the reweighted ℓ_2 and ℓ_1 frameworks. Note that different diversity measures can result in different computational complexity for calculating $\mathbf{W}_n$. Notably, for example, the p-norm-like function resulting in (34) or (35) might incur extra computation for calculating the quantity to the power $2 - p$ or $1 - p$ for some p value (e.g., non-integer power).

Algorithm 1 summarizes the proposed SLMS and SNLMS. A MATLAB implementation of the algorithms is available at https://github.com/chinghualee/SLMS_SNLMS.

Algorithm 1: SLMS and SNLMS.

1: **Input:** step size $\mu > 0$ (or $\tilde{\mu} > 0$), regularization constant $\delta > 0$, input signal $\mathbf{u}_n$, desired signal d_n, and the choice of the diversity measure
2: **Output:** estimated filter $\mathbf{h}_n$
Initialize: $\mathbf{h}_0$
3:   **for** $n = 0, 1, 2, \ldots$ **do**
4:     Compute error signal: $e_n = d_n - \mathbf{u}_n^T \mathbf{h}_n$
5:     Compute scaling matrix: $\mathbf{W}_n$ according to the specified diversity measure (e.g., using (34), (35), (36), or (37))
6:     Compute sparsity-promoting matrix: $\mathbf{S}_n$ by (28)
7:     Update adaptive filter coefficients:
8:     * SLMS: $\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{S}_n \mathbf{u}_n e_n$
9:     * SNLMS: $\mathbf{h}_{n+1} = \mathbf{h}_n + \dfrac{\tilde{\mu}\,\mathbf{S}_n \mathbf{u}_n e_n}{\mathbf{u}_n^T \mathbf{S}_n \mathbf{u}_n + \delta}$
10:   **end for**

C. Comparison to Existing Work on PNLMS-Type Algorithms

Note that in IPNLMS [21] and IPNLMS-ℓ_0 [22] there is also a parameter for fitting the sparsity degree, which was heuristically introduced to weight between proportionate and non-proportionate updates. However, this empirical parameter does not reflect the sparsity level of the underlying system directly. In our algorithms, we have the sparsity control parameters that play a similar role for fitting different sparsity levels. However, based on diversity measures in SSR, they have direct connections to the system sparsity, thereby offering a more intuitive parameter selection procedure. Our algorithms thus have the advantages of enjoying theoretical support and leveraging sparsity more straightforwardly.

In terms of algorithm derivations, PNLMS-type algorithms were mostly developed from a constrained optimization problem following the principle of minimal disturbance, e.g., [24]–[27], [32], in which modified objective functions have been proposed that impose sparsity on the "change" of the filter rather than on the filter itself. For example, [24], [25], [32] considered enforcing sparsity on the difference between the current and updated filters; [26], [27] imposed sparsity on the so-called *correctness*

component as defined in [26] which also represents the change in the filter coefficients. However, since the assumption is that the filter itself is sparse rather than the difference between successive updates, the motivation to enforce sparsity on the "change" of the filter is less clear. Sparsity, in turn, does not seem to fit in straightforwardly under the commonly adopted constrained optimization framework. In contrast, we work with the general MSE criterion in which filter sparsity can be directly imposed via regularization, which is more straightforward and also makes intuitive sense.

V. STEADY-STATE PERFORMANCE ANALYSIS

The signal model of system identification described in Section II-A is employed for performance analysis. We further assume the noise v_n is i.i.d. according to $\mathcal{N}(0, \sigma_v^2)$. We also introduce several other assumptions useful for simplifying analysis. Although these assumptions may seem restrictive, they make meaningful analysis possible without significant loss of insight and are also commonly adopted in the literature. We shall later see that these assumptions lead to theoretical results that are supported by experiments.

Assumption 1: The input data vector $\mathbf{u}_n$ is independent of $\mathbf{u}_k$ for $n \neq k$. Furthermore, $\mathbf{u}_n$ is independent of v_k for all n and k. In practice and from past experience in adaptive filters, this assumption simplifies the analysis and does lead to useful insights [2], [3], despite the fact that it does not in general hold true.

Assumption 2: The input data vector obeys $\mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ for all n. This technical assumption facilitates the analysis by taking advantage of the useful results on Gaussian random variables [4].

Assumption 3: At steady-state, the diagonal matrix $\mathbf{W}_n$ in the update equations can be view as a fixed matrix. As suggested in [5], [23], when the system is at steady-state and when the step size is sufficiently small, the coefficients converge in both mean and mean squared senses. Thus, replacing $\mathbf{W}_n$ by a fixed matrix becomes reasonable and convenient.

For convenience we shall consider the algorithm of the following form for performance analysis:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{S} \mathbf{u}_n e_n, \tag{38}$$

where $\mathbf{S} = \mathrm{diag}\{s_i\}$ with $s_i > 0$, $\forall i = 0, 1, \ldots, M - 1$.

For a fixed underlying system $\mathbf{h}^o$, define the steady-state excess MSE [4]:

$$J_{\mathrm{ex}} \triangleq \lim_{n \to \infty} \mathrm{E}\left[\left(\mathbf{u}_n^T \left(\mathbf{h}^o - \mathbf{h}_n\right)\right)^2\right]. \tag{39}$$

Under Assumption 1, we have the steady-state MSE:

$$J \triangleq \lim_{n \to \infty} \mathrm{E}\left[e_n^2\right] = \sigma_v^2 + J_{\mathrm{ex}}. \tag{40}$$

The following theorems characterize the steady-state behavior of (38):

Theorem 2 (Steady-state excess MSE): Under Assumptions 1-2 with a sufficiently small μ and assume $\mathbf{R} = \sigma_u^2 \mathbf{I}$, for the

adaptive filter (38), the steady-state excess MSE is given by:

$$J_{\text{ex}} = \frac{\mu \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2-2\mu\sigma_u^2 s_i}}{1 - \mu \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2-2\mu\sigma_u^2 s_i}} \sigma_v^2. \tag{41}$$

Proof: See Appendix B.

Theorem 3 (Convergence conditions): Under Assumptions 1-2 with a sufficiently small μ and assume $\mathbf{R} = \sigma_u^2 \mathbf{I}$, *for the adaptive filter (38):*

i) It converges in the mean sense if:

$$|\lambda_{\max}\{\mathbf{I} - \mu\sigma_u^2 \mathbf{S}\}| < 1, \tag{42}$$

where $\lambda_{\max}\{\mathbf{X}\}$ *denotes the largest eigenvalue of a square matrix* $\mathbf{X}$ *in magnitude.*

ii) It converges in the mean squared sense if:

$$0 < \mu < \left(\sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2 - 2\mu\sigma_u^2 s_i} \right)^{-1}. \tag{43}$$

Proof: See Appendix C.

A. Steady-State Performance of SLMS

Consider the case where Assumptions 1–3 are in position and $\mathbf{R} = \sigma_u^2 \mathbf{I}$. For analyzing the proposed SLMS (30), first we need to recognize an appropriate $\mathbf{S}$ with regard to Assumption 3. A useful approximation at steady-state is to replace the occurrence of $\mathbf{h}_n$ by the true system $\mathbf{h}^o$; that is, to use $\mathbf{S} = \frac{\mathbf{W}^2}{\frac{1}{M}\text{tr}(\mathbf{W}^2)}$, where $\mathbf{W} = \text{diag}\{w_i\}$ with w_i given by (13) for the reweighted ℓ_2 case, or by (14) for the reweighted ℓ_1 case, both computed based on the corresponding true coefficient h_i^o. Now, since $\text{tr}(\mathbf{S}) = M$, the excess MSE (41) can be approximated as:

$$J_{\text{ex}} \approx \frac{\mu \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2}}{1 - \mu \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2}} \sigma_v^2 = \frac{\mu \frac{\sigma_u^2}{2} \text{tr}(\mathbf{S})}{1 - \mu \frac{\sigma_u^2}{2} \text{tr}(\mathbf{S})} \sigma_v^2$$

$$= \frac{\mu}{\frac{2}{M\sigma_u^2} - \mu} \sigma_v^2, \tag{44}$$

where for the approximation we assume a sufficiently small step size μ such that $2\mu\sigma_u^2 s_i \ll 2, \forall i = 0, 1, \dots, M - 1$.

Now, for the mean squared convergence condition, although the upper bound in (43) of Theorem 3 contains μ itself, after some inspection it is clear that the lowest stability limit on μ occurs when $\mathbf{S}$ has its diagonal elements nonzero at one tap position (with a value of M) and zero at all others [5]. With such an $\mathbf{S}$, it leads to:

$$0 < \mu < \frac{2}{3M\sigma_u^2}. \tag{45}$$

On the other hand, the largest stability limit is associated with a proportionate matrix assigning equal gains at each position [5], i.e., $\mathbf{S} = \text{diag}\{s_i\}$ with $s_i = 1, \forall i = 0, 1, \dots, M - 1$. With such an $\mathbf{S}$ we have:

$$0 < \mu < \frac{2}{(2 + M)\sigma_u^2}. \tag{46}$$

For a large M, the largest stability limit can be approximated as $\frac{2}{M\sigma_u^2} = \frac{2}{\text{tr}(\mathbf{R})}$, which is also the stability limit of the LMS [4].

This result is not surprising since using an $\mathbf{S}$ that assigns uniform gains essentially becomes the LMS.

B. Steady-State Performance of SNLMS

Consider the case where Assumptions 1-3 are in position and $\mathbf{R} = \sigma_u^2 \mathbf{I}$. For analyzing the proposed SNLMS (32), first we must identify a fixed $\mathbf{S}$ to approximate the term $\frac{\mathbf{S}_n}{\mathbf{u}_n^T \mathbf{S}_n \mathbf{u}_n}$ (where we have ignored δ), for which an exact characterization seems difficult, if at all possible, to obtain. However, if we fix $\mathbf{W}_n = \mathbf{W}$ at steady-state by Assumption 3, where $\mathbf{W}$ is again computed based on the true system $\mathbf{h}^o$, then we have:

$$\mathbf{S} = \frac{\frac{\mathbf{W}^2}{\frac{1}{M}\text{tr}(\mathbf{W}^2)}}{\mathbf{u}_n^T \left(\frac{\mathbf{W}^2}{\frac{1}{M}\text{tr}(\mathbf{W}^2)} \right) \mathbf{u}_n} = \frac{\mathbf{W}^2}{\mathbf{u}_n^T \mathbf{W}^2 \mathbf{u}_n} \approx \frac{\mathbf{W}^2}{\sigma_u^2 \text{tr}(\mathbf{W}^2)}, \tag{47}$$

with the approximation $\mathbf{u}_n^T \mathbf{W}^2 \mathbf{u}_n \approx \text{E}[\mathbf{u}_n^T \mathbf{W}^2 \mathbf{u}_n] = \sigma_u^2 \text{tr}(\mathbf{W}^2)$ utilized. A useful fact of (47) is that $\text{tr}(\mathbf{S}) = (\sigma_u^2)^{-1}$. We can thus use the following approximation for (41) to express the excess MSE (and replace μ by $\tilde{\mu}$):

$$J_{\text{ex}} \approx \frac{\tilde{\mu} \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2}}{1 - \tilde{\mu} \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2}} \sigma_v^2 = \frac{\tilde{\mu}\sigma_u^2 \sum_{i=0}^{M-1} s_i}{2 - \tilde{\mu}\sigma_u^2 \sum_{i=0}^{M-1} s_i} \sigma_v^2$$

$$= \frac{\tilde{\mu}\sigma_u^2 \text{tr}(\mathbf{S})}{2 - \tilde{\mu}\sigma_u^2 \text{tr}(\mathbf{S})} \sigma_v^2 = \frac{\tilde{\mu}\sigma_u^2 (\sigma_u^2)^{-1}}{2 - \tilde{\mu}\sigma_u^2 (\sigma_u^2)^{-1}} \sigma_v^2 = \frac{\tilde{\mu}}{2 - \tilde{\mu}} \sigma_v^2, \tag{48}$$

for $\tilde{\mu}$ sufficiently small such that $2\tilde{\mu}\sigma_u^2 s_i \ll 2, \forall i = 0, 1, \dots, M - 1$,

For the mean squared convergence condition, using the same argument as in the SLMS case for (45) and (46), we can obtain the lowest stability limit as:

$$0 < \tilde{\mu} < \frac{2}{3} \tag{49}$$

and the largest stability limit as:

$$0 < \tilde{\mu} < \frac{2}{1 + \frac{2}{M}}. \tag{50}$$

For a large M, we have (50) approximately as $0 < \tilde{\mu} < 2$, which is the classic result of the NLMS [4].

VI. SIMULATION RESULTS

The proposed algorithms are evaluated using computer simulations in MATLAB. We consider three system IRs as shown in Fig. 1 which represent different sparsity levels: quasi-sparse, sparse, and dispersive systems. The IR of the quasi-sparse system is an acoustic feedback path between the microphone and the loudspeaker of a hearing aid that was measured from a real-world scenario. It represents a typical IR of many practical system identification problems where certain degree of structural sparsity exists. The sparse and dispersive IRs were artificially generated. Each of these IRs has 256 taps. We conducted experiments to obtain the MSE learning curves (i.e., the ensemble average of e_n^2 as a function of iteration n) for performance comparison. The ensemble averaging was performed over 1000 independent Monte Carlo runs for obtaining each curve. In all experiments, the adaptive filter coefficients were initialized with
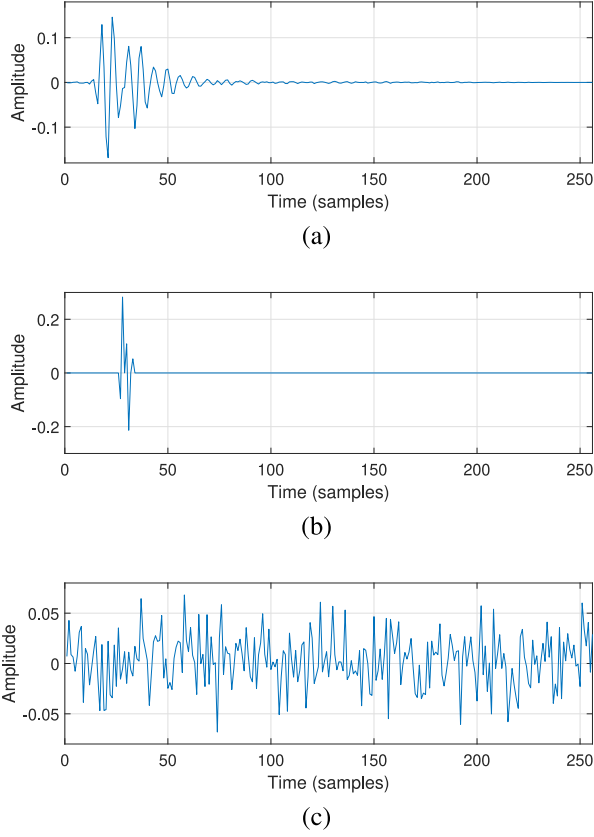
Fig. 1.   IRs of (a) quasi-sparse, (b) sparse, and (c) dispersive systems. The quasi-sparse IR is an acoustic feedback path of a hearing aid that was measured from a real-world scenario. The sparse and dispersive IRs were artificially generated.
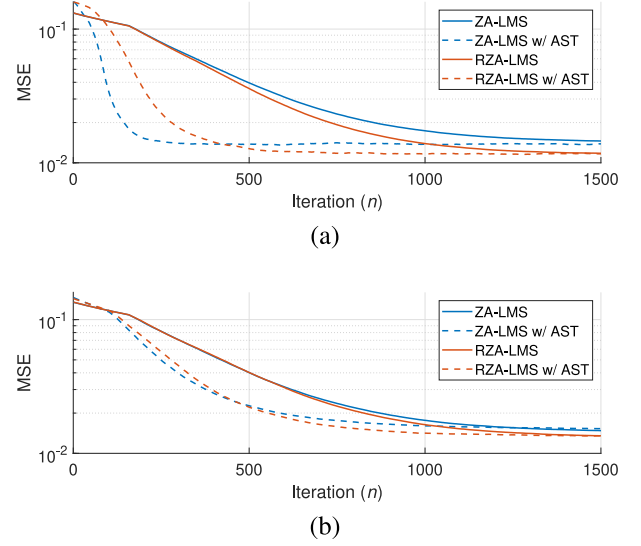

Fig. 2.   Comparison of algorithms with and without AST for identifying (a) sparse and (b) quasi-sparse IRs with white Gaussian process input. Solid lines are existing approaches as given by (10). Dotted lines are their corresponding AST-based algorithms given by (29). It can be seen that AST leads to improved performance.

all zeros. For the input signal, we mainly consider two types of u_n for theoretical analysis: i) a zero mean, unit variance white Gaussian process and ii) a first order autoregressive (AR) process according to $u_n = \rho u_{n-1} + \eta_n$, where $\rho = 0.8$ and η_n is i.i.d. according to $\mathcal{N}(0,1)$. We also include results of speech inputs for demonstrating the algorithm performance with non-stationary signals. The system noise v_n is i.i.d. according to $\mathcal{N}(0, 0.01)$. Regarding the algorithms, when using (34) for updating $\mathbf{W}_n$, a small positive constant $c = 0.001$ was always used.

A. Comparison of Algorithms With and Without AST

Fig. 2 compares the proposed generalized sparse LMS (29), i.e., the AST-based approach, to some existing regularization-based algorithms of (10), i.e., the regular gradient descent without AST. Specifically, we use the p-norm-like penalty $\|\mathbf{h}\|_p^p$ with $p = 1$ and the log-sum penalty $\sum_{i=0}^{M-1} \log(|h_i| + \epsilon)$ with $\epsilon = 0.1$ as examples. These two choices of the sparsity-inducing function $G(\mathbf{h})$ in (10) result in the ZA-LMS and RZA-LMS [28], respectively. We compare them with the corresponding AST-based algorithms obtained from (29), also adopting the two penalty functions for $G(\mathbf{h})$ that lead to (34) and (37) for computing $\mathbf{W}_n$, respectively. We set $\mu = 0.0025$ and $\lambda = 0.001$ in all cases and used the white Gaussian process input. Fig. 2 (a)

shows the results of identifying the sparse IR and Fig. 2 (b) is the case of estimating the quasi-sparse IR. From the results we see that the AST strategy leads to algorithms (dotted lines) that demonstrate faster convergence than the existing approaches (solid lines).

B. Effect of Sparsity Control Parameter on SLMS and SNLMS

In this experiment we investigate the effect of the sparsity control parameter on the convergence of SLMS (30) and SNLMS (32). We use the p-norm-like diversity measure $\|\mathbf{h}\|_p^p$ within the reweighted ℓ_2 framework, i.e., using (34) for updating $\mathbf{W}_n$, for demonstration purposes. We study the cases of the sparsity control parameter $p = 1, 1.2, 1.5, 1.8, 2$. We also include the LMS (3) and NLMS (4) performance curves for reference. For LMS and SLMS we used $\mu = 0.0025$. For NLMS and SNLMS we used $\tilde{\mu} = 0.5$ and $\delta = 0.01$.

Fig. 3 and Fig. 4 show the resulting MSE curves for SLMS using the white Gaussian noise input and SNLMS using the AR process input, respectively. Recall that the proportionate factors of SLMS/SNLMS using (34) for $\mathbf{W}_n$ approximate that of the PNLMS when $p \to 1$, and regenerate the LMS/NLMS when $p = 2$, as has been discussed in Section IV-B. Therefore, the parameter p plays the role for fitting different sparsity levels and the selection of p can be crucial for obtaining optimal performance for IRs with different sparsity degrees. The results in both Fig. 3 and Fig. 4 suggest that for the quasi-sparse case, the fastest convergence is given by $p \in [1.2, 1.5]$, which seems reasonable in terms of finding a balance between PNLMS ($p \to 1$) and LMS/NLMS ($p = 2$). On the other hand, for the sparse system, $p \in [1, 1.2]$ gives the best results, which is also reasonable since as the sparsity level increases, a more PNLMS-like algorithm can be more favorable. Finally, for the dispersive system we see that
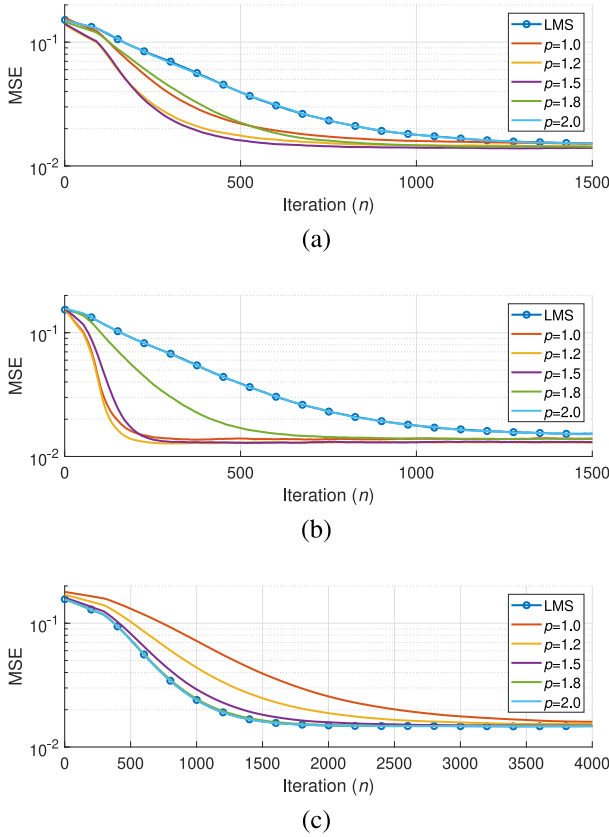
(a)

(b)

(c)

Fig. 3. Effect of sparsity control parameter p on convergence of SLMS for (a) quasi-sparse, (b) sparse, and (c) dispersive IRs with white Gaussian process input. It can be seen that the optimal p value varies with the sparsity degree.

(a)

(b)

(c)

Fig. 4. Effect of sparsity control parameter p on convergence of SNLMS for (a) quasi-sparse, (b) sparse, and (c) dispersive IRs with AR process input. In the colored input case here we have similar observations to the white input case of Fig. 3.

$p \in [1.8, 2]$ results in the fastest convergence and is comparable to, if not better than, the LMS and NLMS. This indicates that a more LMS/NLMS-like algorithm is preferable when the system IR is far from sparse. To conclude, the results show that the algorithms exploit the underlying system structure in the way we expect.

C. Effect of Step Size on SLMS and SNLMS

Fig. 5 studies the effect of the step size on the convergence behavior of the SLMS and SNLMS. We again used (34) for updating $\mathbf{W}_n$. Fig. 5 (a) shows the resulting MSE curves obtained by running the SLMS with $p = 1.2$ on the sparse IR with various μ values, using the white Gaussian noise input. Fig. 5 (b) shows the resulting MSE curves obtained by running the SNLMS with $p = 1.5$ on the quasi-sparse IR with various $\tilde{\mu}$ values, using the AR process input. The dotted lines indicate the theoretical steady-state MSE levels computed from (40) using (44) and (48) for SLMS and SNLMS, respectively. We can see that similar to the well-known trade-off in LMS and NLMS, a larger step size results in faster convergence while at the expense of steady-state performance. We also see that as the step size increases the theoretical prediction becomes less accurate; this is probably due to the approximation made based on the small step size assumption for arriving at (44) and (48). Nevertheless, the
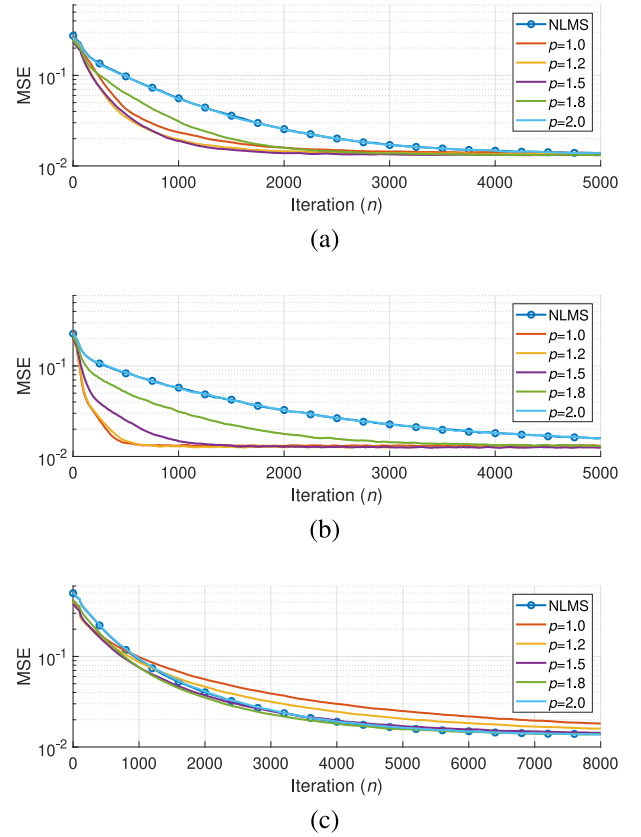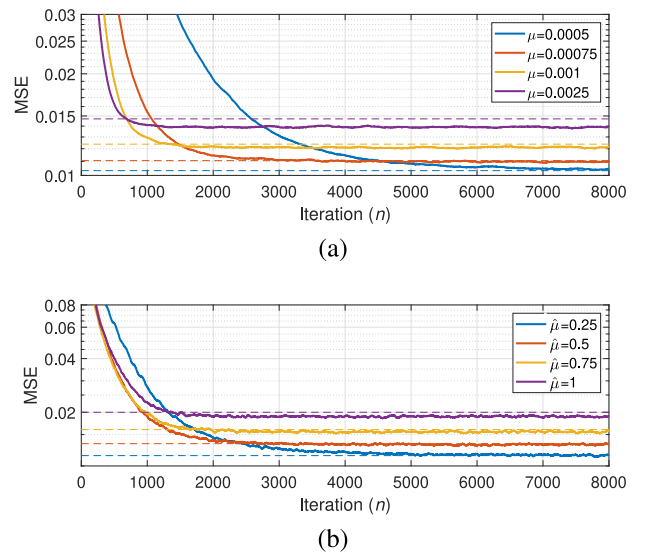

(a)

(b)

Fig. 5. Effect of step size μ or $\hat{\mu}$ on convergence of (a) SLMS for the sparse IR with white Gaussian process input and (b) SNLMS for the quasi-sparse IR with AR process input. Dotted lines indicate the theoretical steady-state MSE levels. It can be seen that the theoretical prediction agrees reasonably well with the experimental results especially for a small step size.
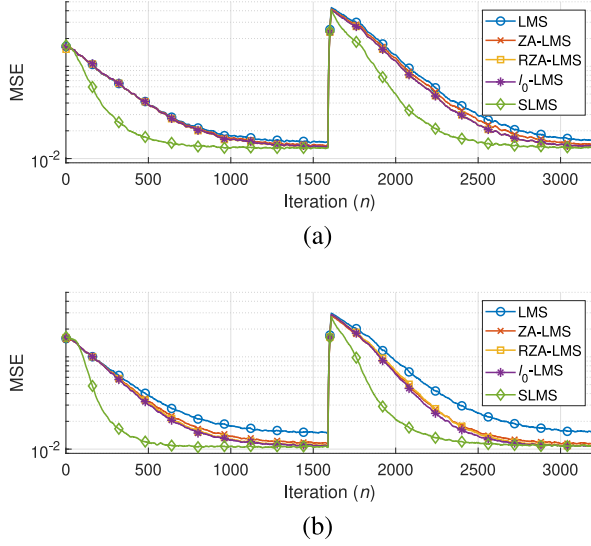
Fig. 6. Comparison of LMS-type algorithms with white Gaussian process input on (a) quasi-sparse and (b) sparse IRs. One can see that the proposed SLMS outperforms all the other approaches in both cases.
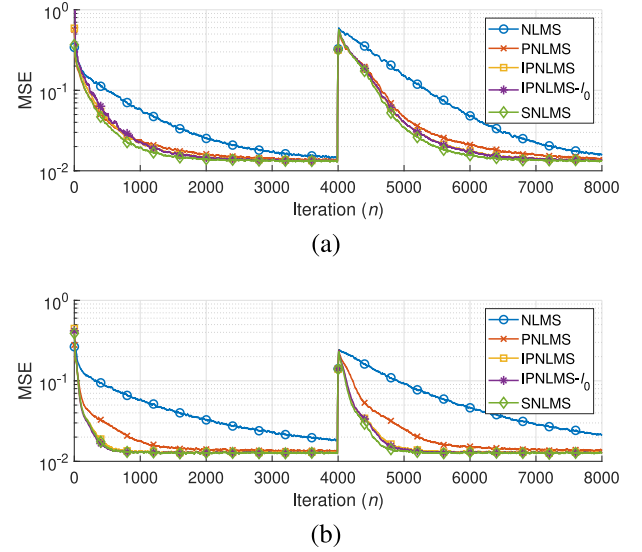

Fig. 7. Comparison of NLMS-type algorithms with AR process input on (a) quasi-sparse and (b) sparse IRs. One can see that the proposed SNLMS performs better than all the other approaches in both cases.

prediction agrees well with the stead-state MSE in most cases for a small step size. In addition, though several assumptions have been made to arrive at (40), (44) and (48), the results show that they predict reasonably well in the case of white input and even for correlated input.

D. Comparison With Existing Algorithms

We compare the proposed SLMS and SNLMS using (34) for $\mathbf{W}_n$ with existing LMS-type and NLMS-type algorithms. To see how the algorithms behave in a changing environment, in each of the following experiments, a change in the underlying system was introduced by shifting the IR to the right by 16 samples in the middle of the adaptation process [47].

Fig. 6 compares the LMS-type algorithms using the white Gaussian process input. Fig. 6 (a) and Fig. 6 (b) show the MSE curves obtained with the quasi-sparse and sparse IRs, respectively. For LMS we used $\mu = 0.0025$. For ZA-LMS, RZA-LMS, and ℓ_0-LMS we fixed $\mu = 0.0025$ and then experimentally optimized the remaining parameters to obtain the best performance. For SLMS we used $p = 1.5$ and $\mu = 0.002$ in the quasi-sparse case and $p = 1.2$ and $\mu = 0.0005$ in the sparse case. The results show that all the sparsity-aware algorithms outperform the LMS, with SLMS demonstrating the best result. Comparing Fig. 6 (a) and Fig. 6 (b), we also see that the benefit brought by existing sparsity-aware algorithms becomes limited when the system is less sparse, while the SLMS still provides significant improvement.

Fig. 7 compares the NLMS-type algorithms using the AR process input. Fig. 7 (a) and Fig. 7 (b) show the MSE curves obtained with the quasi-sparse and sparse IRs, respectively. For all the algorithms we used $\tilde{\mu} = 0.5$. For NLMS we used $\delta = 0.01$. For PNLMS, IPNLMS, and IPNLMS-ℓ_0 we set $\delta = 0.01/M$ according to [47], and experimentally optimized the remaining parameters to obtain the best performance in each

case. For SNLMS we used $p = 1.5$ in the quasi-sparse case and $p = 1.2$ in the sparse case. We used $\delta = 0.01$ for SNLMS, same as NLMS.[13] From the results we again observe the benefit of using sparsity-aware adaptation. In addition, the SNLMS demonstrates performance as good as, if not better than, the other proportionate algorithms.

Fig. 8 considers a more practical scenario where we used a speech signal as the input and the quasi-sparse IR, which represents an acoustic channel of practical interest, as the underlying system. The input signal-to-noise ration (SNR) was set to 20 dB using white Gaussian noise. For the SLMS and SNLMS we used $p = 1.5$ which is a suitable choice for quasi-sparse systems. For evaluation we compare the normalized misalignment $\|\mathbf{h}^o - \mathbf{h}_n\|_2^2 / \|\mathbf{h}^o\|_2^2$. In Fig. 8 (a) we see that SLMS performs much better than the LMS, while the ℓ_0-LMS fails to provide any improvement. This may be due to the fact that existing regularization-based algorithms tend to enforce sparsity in a more aggressive manner as they work with $\lambda > 0$, and this may not be beneficial, if not harmful, when the underlying system is not truly sparse. In Fig. 8 (b) we see that SNLMS demonstrates superior convergence behavior than the NLMS, and is also better than the IPNLMS and IPNLMS-ℓ_0.

Fig. 9 shows the results for a noisier environment, i.e., 0 dB input SNR, for the same experimental setting of Fig. 8 (only the step size parameters were further tuned due to the stronger noise). We see that in Fig. 9 (a) the SLMS significantly outperforms the LMS, while the ℓ_0-LMS performs worse. The SNLMS in Fig. 9 (b), on the other hands, still performs better than the NLMS, and is comparable to other proportionate algorithms. This indicates that our observation on the SLMS and SNLMS superiority may be robust to the noise condition.

---

[13]Due to the division by $\frac{1}{M}$ in (28) which is not present in (8) of existing PNLMS-type algorithms, the division by M is not needed for δ in SNLMS.
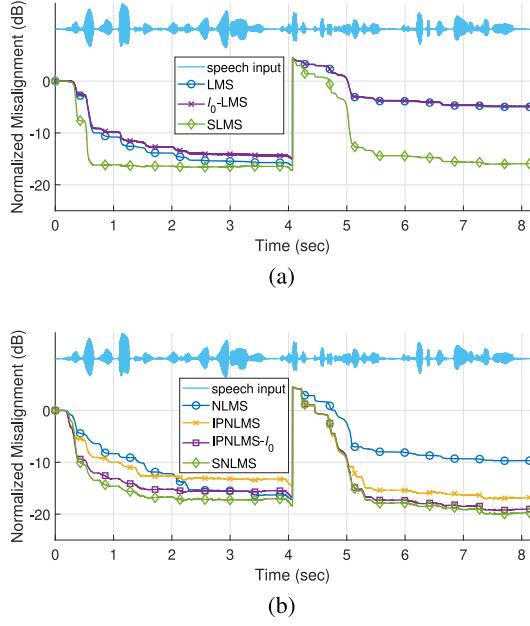
Fig. 8. Comparison of (a) LMS-type and (b) NLMS-type algorithms for identifying the quasi-sparse acoustic channel response with speech input at 20 dB SNR. In can be seen that the SLMS and SNLMS perform the best in both cases.
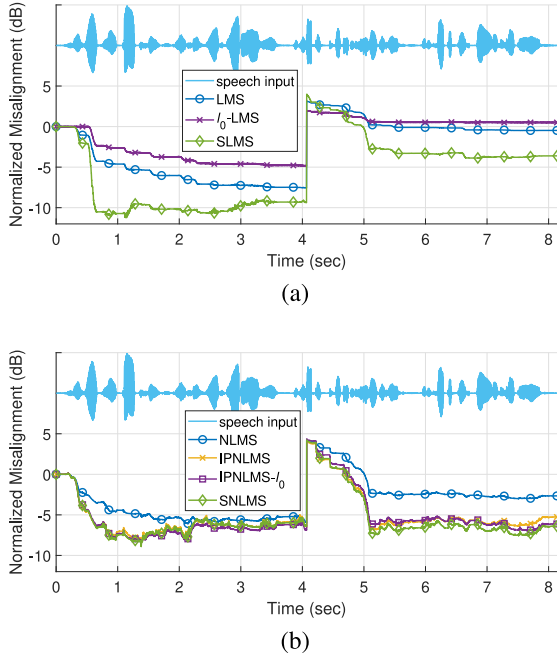

Fig. 9. Comparison of (a) LMS-type and (b) NLMS-type algorithms for identifying the quasi-sparse acoustic channel response with speech input at 0 dB SNR. In the noisier setting here the SLMS and SNLMS perform comparably well, if not better than, the competing algorithms.

VII. CONCLUSION

In this paper, we developed a mathematical framework for rigorously deriving adaptive filters that exploit the sparse structure of the underlying system response. We started with the regularized objective framework of SSR and developed algorithms that are of the proportionate type. As a result, the adaptive algorithms

are quite general and can accommodate a range of regularization functions. The framework utilizes the AST methodology within the iterative reweighted ℓ_2 and ℓ_1 schemes, which is shown to be crucial for obtaining improved adaptive filtering performance over existing algorithms when gradient descent is concerned. We further introduced the SLMS and SNLMS by adopting a zero regularization coefficient, which take advantage of, though do not strictly enforce, the sparsity of the underlying system if it already exists. Note that the proposed framework is not limited to the algorithms that we have presented so far. Any other penalty function that satisfies the conditions imposed on the diversity measure can potentially be a good candidate for obtaining effective adaptive algorithms by utilizing the framework.

APPENDIX A
PROOF OF THEOREM 1

The proof follows the idea in [48]. We wish to show that the regularized objective function $J^G(\mathbf{h})$ in (9) is decreased in each iteration when optimized via (20) and (21). Before proceeding, we need the following lemmas:

Lemma 1: For the general diversity measure $G(\mathbf{h}) = \sum_{i=0}^{M-1} g(h_i)$ that satisfies Properties 1-4 in Section II-B, with $g(z)$ being strictly concave in z^2 for Property 4, we have:

$$G(\mathbf{h}_{n+1}) - G(\mathbf{h}_n) < \|\mathbf{W}_n^{-1}\mathbf{h}_{n+1}\|_2^2 - \|\mathbf{W}_n^{-1}\mathbf{h}_n\|_2^2, \quad (51)$$

where $\mathbf{W}_n = \text{diag}\{w_{i,n}\}$ with $w_{i,n}$ given by (13).

Proof: Since $g(z)$ is strictly concave in z^2, it satisfies $g(z) = f(z^2)$ where $f(z)$ is concave for $z \in \mathbb{R}_+$. Due to the concavity, we have the following inequality:

$$f(z_2) - f(z_1) < f'(z_1)(z_2 - z_1) \quad (52)$$

hold for some $z_1, z_2 \in \mathbb{R}_+$. Note that we use $f'(z_1)$ to denote the first order derivative of $f(z)$ w.r.t. z evaluated at $z = z_1$.

Substituting $z_1 = h_{i,n}^2$ and $z_2 = h_{i,n+1}^2$ into (52) gives:

$$f(h_{i,n+1}^2) - f(h_{i,n}^2) < f'(h_{i,n}^2)(h_{i,n+1}^2 - h_{i,n}^2). \quad (53)$$

Noting that $f(h_{i,n+1}^2) = g(h_{i,n+1})$ and $f(h_{i,n}^2) = g(h_{i,n})$, we have:

$$g(h_{i,n+1}) - g(h_{i,n}) < f'(h_{i,n}^2)(h_{i,n+1}^2 - h_{i,n}^2). \quad (54)$$

From (13) we have $f'(h_{i,n}^2) = w_{i,n}^{-2}$. Therefore,

$$g(h_{i,n+1}) - g(h_{i,n}) < w_{i,n}^{-2}(h_{i,n+1}^2 - h_{i,n}^2). \quad (55)$$

Summing over $i = 0, 1, \ldots, M-1$ on both sides of (55) justifies (51) of Lemma 1.

Lemma 2: For the general diversity measure $G(\mathbf{h}) = \sum_{i=0}^{M-1} g(h_i)$ that satisfies Properties 1-4 in Section II-B, with $g(z)$ being strictly concave in $|z|$ for Property 4, we have:

$$G(\mathbf{h}_{n+1}) - G(\mathbf{h}_n) < \|\mathbf{W}_n^{-1}\mathbf{h}_{n+1}\|_1 - \|\mathbf{W}_n^{-1}\mathbf{h}_n\|_1, \quad (56)$$

where $\mathbf{W}_n = \text{diag}\{w_{i,n}\}$ with $w_{i,n}$ given by (14).

Proof: Since $g(z)$ is strictly concave in $|z|$, it satisfies $g(z) = f(|z|)$ where $f(z)$ is concave for $z \in \mathbb{R}_+$. Again, the inequality (52) holds due to the concavity of $f(z)$.

Substituting $z_1 = |h_{i,n}|$ and $z_2 = |h_{i,n+1}|$ into (52) gives:

$$f(|h_{i,n+1}|) - f(|h_{i,n}|) < f'(|h_{i,n}|)(|h_{i,n+1}| - |h_{i,n}|). \quad (57)$$

Noting that $f(|h_{i,n+1}|) = g(h_{i,n+1})$ and $f(|h_{i,n}|) = g(h_{i,n})$, we have:

$$g(h_{i,n+1}) - g(h_{i,n}) < f'(|h_{i,n}|)(|h_{i,n+1}| - |h_{i,n}|). \quad (58)$$

From (14) we have $f'(|h_{i,n}|) = w_{i,n}^{-1}$. Therefore,

$$g(h_{i,n+1}) - g(h_{i,n}) < w_{i,n}^{-1}(|h_{i,n+1}| - |h_{i,n}|). \quad (59)$$

Summing over $i = 0, 1, \dots, M - 1$ on both sides of (59), we have (56) of Lemma 2 justified.

Now we are ready to show that $J^G(\mathbf{h})$ decreases in each iteration by using the update recursions (20) and (21).

First, for the reweighted ℓ_2 framework with $J_n^{\ell_2}(\mathbf{q})$ in (16), we have:

$$
\begin{aligned}
J^G(\mathbf{h}_{n+1}) - J^G(\mathbf{h}_n) &= [J(\mathbf{h}_{n+1}) + \lambda G(\mathbf{h}_{n+1})] \\
&\quad - [J(\mathbf{h}_n) + \lambda G(\mathbf{h}_n)] \\
&< [J(\mathbf{h}_{n+1}) + \lambda \|\mathbf{W}_n^{-1}\mathbf{h}_{n+1}\|_2^2] \\
&\quad - [J(\mathbf{h}_n) + \lambda \|\mathbf{W}_n^{-1}\mathbf{h}_n\|_2^2] \\
&= [J(\mathbf{W}_n\mathbf{q}_{n+1|n}) + \lambda \|\mathbf{q}_{n+1|n}\|_2^2] \\
&\quad - [J(\mathbf{W}_n\mathbf{q}_{n|n}) + \lambda \|\mathbf{q}_{n|n}\|_2^2] \\
&= J_n^{\ell_2}(\mathbf{q}_{n+1|n}) - J_n^{\ell_2}(\mathbf{q}_{n|n}).
\end{aligned}
$$
$$(60)$$

The inequality follows from Lemma 1. The AST relationships (18) and (19) are also utilized. As we perform optimization of (16) with gradient descent, we can have $J_n^{\ell_2}(\mathbf{q})$ decrease in each iteration n, i.e., $J_n^{\ell_2}(\mathbf{q}_{n+1|n}) - J_n^{\ell_2}(\mathbf{q}_{n|n}) < 0$, using some μ_n. Therefore, the choice of $\{\mu_n\}_{n=0}^{\infty}$ ensures the decrease in $J^G(\mathbf{h})$ according to (60), and the update recursion (20) monotonically converges to a local minimum (or saddle point) of (9) under a WSS environment.

On the other hand, for the reweighted ℓ_1 framework with $J_n^{\ell_1}(\mathbf{q})$ in (17), we have:

$$
\begin{aligned}
J^G(\mathbf{h}_{n+1}) - J^G(\mathbf{h}_n) &= [J(\mathbf{h}_{n+1}) + \lambda G(\mathbf{h}_{n+1})] \\
&\quad - [J(\mathbf{h}_n) + \lambda G(\mathbf{h}_n)] \\
&< [J(\mathbf{h}_{n+1}) + \lambda \|\mathbf{W}_n^{-1}\mathbf{h}_{n+1}\|_1] \\
&\quad - [J(\mathbf{h}_n) + \lambda \|\mathbf{W}_n^{-1}\mathbf{h}_n\|_1] \\
&= [J(\mathbf{W}_n\mathbf{q}_{n+1|n}) + \lambda \|\mathbf{q}_{n+1|n}\|_1] \\
&\quad - [J(\mathbf{W}_n\mathbf{q}_{n|n}) + \lambda \|\mathbf{q}_{n|n}\|_1] \\
&= J_n^{\ell_1}(\mathbf{q}_{n+1|n}) - J_n^{\ell_1}(\mathbf{q}_{n|n}).
\end{aligned}
$$
$$(61)$$

The inequality follows from Lemma 2. The AST relationships (18) and (19) are also utilized. Similar to the above argument of the reweighted ℓ_2 case, there exists a choice of $\{\mu_n\}_{n=0}^{\infty}$ that ensures the decrease in $J^G(\mathbf{h})$ according to (61), and the update recursion (21) monotonically converges to a local minimum (or saddle point) of (9) under a WSS environment.

APPENDIX B
PROOF OF THEOREM 2

The proof follows the discussion in [4], [5], [25]. Substituting $e_n = d_n - \mathbf{u}_n^T\mathbf{h}_n$ into (38) we have:

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \mu\mathbf{S}\mathbf{u}_n\mathbf{u}_n^T\mathbf{h}_n + \mu\mathbf{S}\mathbf{u}_n d_n. \quad (62)$$

Using the fact that $d_n = \mathbf{u}_n^T\mathbf{h}^o + v_n$, we have:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \mu\mathbf{S}\mathbf{u}_n\mathbf{u}_n^T(\mathbf{h}^o - \mathbf{h}_n) + \mu\mathbf{S}\mathbf{u}_n v_n. \quad (63)$$

Define the misalignment vector ε_n as:

$$\varepsilon_n = \mathbf{h}^o - \mathbf{h}_n. \quad (64)$$

Then from (63) we have:

$$\varepsilon_{n+1} = (\mathbf{I} - \mu\mathbf{S}\mathbf{u}_n\mathbf{u}_n^T)\varepsilon_n - \mu\mathbf{S}\mathbf{u}_n v_n. \quad (65)$$

Next, based on (65) we have:

$$
\begin{aligned}
\varepsilon_{n+1}\varepsilon_{n+1}^T = {}& (\mathbf{I} - \mu\mathbf{S}\mathbf{u}_n\mathbf{u}_n^T)\varepsilon_n\varepsilon_n^T(\mathbf{I} - \mu\mathbf{u}_n\mathbf{u}_n^T\mathbf{S}) \\
& + \mu^2 v_n^2\mathbf{S}\mathbf{u}_n\mathbf{u}_n^T\mathbf{S} + \boldsymbol{\Xi},
\end{aligned}
$$
$$(66)$$

where $\boldsymbol{\Xi}$ represents the remaining cross terms whose expectations are zero.

Let $\boldsymbol{\Omega}_n = \mathrm{E}[\varepsilon_n\varepsilon_n^T]$. Taking expectation on both sides of (66) we have:

$$
\boldsymbol{\Omega}_{n+1} = \underbrace{\mathrm{E}\left[(\mathbf{I} - \mu\mathbf{S}\mathbf{u}_n\mathbf{u}_n^T)\varepsilon_n\varepsilon_n^T(\mathbf{I} - \mu\mathbf{u}_n\mathbf{u}_n^T\mathbf{S})\right]}_{\triangleq\boldsymbol{\Theta}}
$$
$$+ \mu^2\sigma_v^2\mathbf{S}\mathbf{R}\mathbf{S}. \quad (67)$$

Note that:

$$
\begin{aligned}
\boldsymbol{\Theta} = {}& \boldsymbol{\Omega}_n - \mu\mathbf{S}\mathbf{R}\boldsymbol{\Omega}_n - \mu\boldsymbol{\Omega}_n\mathbf{R}\mathbf{S} \\
& + \mu^2\mathbf{S}\mathrm{E}\left[\mathbf{u}_n\mathbf{u}_n^T\varepsilon_n\varepsilon_n^T\mathbf{u}_n\mathbf{u}_n^T\right]\mathbf{S}.
\end{aligned}
$$
$$(68)$$

With Assumptions 1 and 2 it can be shown that [4]:

$$\mathrm{E}\left[\mathbf{u}_n\mathbf{u}_n^T\varepsilon_n\varepsilon_n^T\mathbf{u}_n\mathbf{u}_n^T\right] = 2\mathbf{R}\boldsymbol{\Omega}_n\mathbf{R} + \mathbf{R}\mathrm{tr}(\mathbf{R}\boldsymbol{\Omega}_n). \quad (69)$$

Thus,

$$
\begin{aligned}
\boldsymbol{\Theta} = {}& \boldsymbol{\Omega}_n - \mu\mathbf{S}\mathbf{R}\boldsymbol{\Omega}_n - \mu\boldsymbol{\Omega}_n\mathbf{R}\mathbf{S} + 2\mu^2\mathbf{S}\mathbf{R}\boldsymbol{\Omega}_n\mathbf{R}\mathbf{S} \\
& + \mu^2\mathbf{S}\mathbf{R}\mathrm{tr}(\mathbf{R}\boldsymbol{\Omega}_n)\mathbf{S}.
\end{aligned}
$$
$$(70)$$

Then, with $\mathbf{R} = \sigma_u^2\mathbf{I}$, in steady-state, i.e., $n \to \infty$,

$$
\begin{aligned}
\boldsymbol{\Omega}_\infty = {}& \boldsymbol{\Omega}_\infty - \mu\sigma_u^2\mathbf{S}\boldsymbol{\Omega}_\infty - \mu\sigma_u^2\boldsymbol{\Omega}_\infty\mathbf{S} + 2\mu^2\sigma_u^4\mathbf{S}\boldsymbol{\Omega}_\infty\mathbf{S} \\
& + \mu^2\sigma_u^4\mathbf{S}\mathrm{tr}(\boldsymbol{\Omega}_\infty)\mathbf{S} + \mu^2\sigma_u^2\sigma_v^2\mathbf{S}^2.
\end{aligned}
$$
$$(71)$$

This implies:

$$
\begin{aligned}
\boldsymbol{\omega}_\infty = {}& \boldsymbol{\omega}_\infty - 2\mu\sigma_u^2\mathbf{S}\boldsymbol{\omega}_\infty + 2\mu^2\sigma_u^4\mathbf{S}^2\boldsymbol{\omega}_\infty + \mu^2\sigma_u^4\mathbf{s}^2\mathbf{1}^T\boldsymbol{\omega}_\infty \\
& + \mu^2\sigma_u^2\sigma_v^2\mathbf{s}^2,
\end{aligned}
$$
$$(72)$$

where $\boldsymbol{\omega}_\infty$ and $\mathbf{s}$ are the vectors consisting of the diagonal elements of $\boldsymbol{\Omega}_\infty$ and $\mathbf{S}$, respectively, and $\mathbf{s}^2$ denotes the element-wise squares of the vector $\mathbf{s}$.

The steady-state excess MSE is then:

$$J_{ex} \triangleq \lim_{n \to \infty} E\left[\left(\mathbf{u}_n^T(\mathbf{h}^o - \mathbf{h}_n)\right)^2\right] = \lim_{n \to \infty} E\left[\left(\mathbf{u}_n^T \boldsymbol{\varepsilon}_n\right)^2\right]$$

$$= \lim_{n \to \infty} \text{tr}\left(\boldsymbol{\Omega}_n E\left[\mathbf{u}_n \mathbf{u}_n^T\right]\right) = \sigma_u^2 \text{tr}\left(\boldsymbol{\Omega}_\infty\right) = \sigma_u^2 \mathbf{1}^T \boldsymbol{\omega}_\infty. \quad (73)$$

Using (73) for (72) and rearrange the equation, we have:

$$\omega_{i,\infty} = \frac{\mu^2 \sigma_u^2 s_i^2 J_{ex} + \mu^2 \sigma_u^2 \sigma_v^2 s_i^2}{2\mu \sigma_u^2 s_i - 2\mu^2 \sigma_u^4 s_i^2}. \quad (74)$$

Then it leads to:

$$J_{ex} = \sigma_u^2 \sum_{i=0}^{M-1} \omega_{i,\infty} = \sum_{i=0}^{M-1} \frac{\mu^2 \sigma_u^2 s_i^2 J_{ex} + \mu^2 \sigma_u^2 \sigma_v^2 s_i^2}{2\mu s_i - 2\mu^2 \sigma_u^2 s_i^2}, \quad (75)$$

which yields:

$$J_{ex} = \frac{\mu \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2 - 2\mu \sigma_u^2 s_i}}{1 - \mu \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2 - 2\mu \sigma_u^2 s_i}} \sigma_v^2. \quad (76)$$

This justifies Theorem 2.

APPENDIX C
PROOF OF THEOREM 3

Note that Assumption 1 ensures that $\mathbf{h}_n$, $\mathbf{u}_n$, and v_n are mutually independent. Thus, taking expectation of both sides of (65) gives:

$$E[\boldsymbol{\varepsilon}_{n+1}] = (\mathbf{I} - \mu \mathbf{S} \mathbf{R}) E[\boldsymbol{\varepsilon}_n]. \quad (77)$$

Therefore, the following condition is sufficient for convergence in the mean sense [4]:

$$|\lambda_{\max}\{\mathbf{I} - \mu \mathbf{S} \mathbf{R}\}| < 1. \quad (78)$$

With $\mathbf{R} = \sigma_u^2 \mathbf{I}$, Theorem 3-i) is justified.

From (76) we see, by requiring:

$$1 - \mu \sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2 - 2\mu \sigma_u^2 s_i} > 0, \quad (79)$$

we obtain the stability bound for μ as:

$$0 < \mu < \left(\sum_{i=0}^{M-1} \frac{\sigma_u^2 s_i}{2 - 2\mu \sigma_u^2 s_i}\right)^{-1}, \quad (80)$$

which justifies Theorem 3-ii).

ACKNOWLEDGMENT

REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Upper Saddle River, NJ, USA: Pearson, 1985.

[2] S. Haykin, *Adaptive Filter Theory*, 5th ed., Upper Saddle River, NJ, USA: Pearson, 2013.

[3] A. H. Sayed, *Adaptive Filters*, Hoboken, NJ, USA: John Wiley & Sons, 2011.

[4] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing*, New York, NY, USA: McGraw-Hill, 2000.

[5] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[6] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer, 2001.

[7] C. Paleologu, J. Benesty, and S. Ciochină, "Sparse adaptive filters for echo cancellation," *Synth. Lect. Speech Audio Process.*, vol. 6, no. 1, pp. 1–124, 2010.

[8] E. Hänsler and G. Schmidt, *Topics in Acoustic Echo and Noise Control: Selected Methods for the Cancellation of Acoustical Echoes, the Reduction of Background Noise, and Speech Processing*. Berlin, Germany: Springer, 2006.

[9] C.-H. Lee, B. D. Rao, and H. Garudadri, "Sparsity promoting LMS for adaptive feedback cancellation," in *Proc. Eur. Signal Process. Conf.*, 2017, pp. 226–230.

[10] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proc. MTS/IEEE OCEANS*, 1995, vol. 3, pp. 1417–1422.

[11] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

[12] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[13] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, New York, NY, USA: McGraw-Hill, 1996.

[14] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, pp. 267–288, 1996.

[15] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[16] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, no. Jun, pp. 211–244, 2001.

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

[18] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 3869–3872.

[19] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

[20] K. Wagner and M. Doroslovački, *Proportionate-Type Normalized Least Mean Square Algorithms*. Hoboken, NJ, USA: Wiley , 2013.

[21] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, pp. 1881–1884.

[22] C. Paleologu, J. Benesty, and S. Ciochină, "An improved proportionate NLMS algorithm based on the l_0 norm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 309–312.

[23] R. K. Martin, W. A. Sethares, R. C. Williamson, and C. R. Johnson, "Exploiting sparsity in adaptive filters," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1883–1894, Aug. 2002.

[24] B. D. Rao and B. Song, "Adaptive filtering algorithms for promoting sparsity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, pp. 361–364.

[25] Y. Jin, "*Algorithm Development for Sparse Signal Recovery and Performance Limits Using Multiple-User Information Theory*," Ph.D. dissertation, Univ. California, San Diego, 2011.

[26] J. Benesty, C. Paleologu, and S. Ciochină, "Proportionate adaptive filters from a basis pursuit perspective," *IEEE Signal Process. Lett.*, vol. 17, no. 12, pp. 985–988, Dec. 2010.

[27] J. Liu and S. L. Grant, "A generalized proportionate adaptive algorithm based on convex optimization," in *Proc. IEEE China Summit Int. Conf. Signal Inform. Process.*, 2014, pp. 748–752.

[28] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 3125–3128.

[29] Y. Gu, J. Jin, and S. Mei, "l_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 774–777, 2009.

[30] F. Y. Wu and F. Tong, "Gradient optimization p-norm-like constraint LMS algorithm for sparse system estimation," *Signal Process.*, vol. 93, no. 4, pp. 967–971, 2013.

[31] O. Taheri and S. A. Vorobyov, "Reweighted l_1-norm penalized LMS for sparse channel estimation and its analysis," *Signal Process.*, vol. 104, pp. 70–79, 2014.

[32] R. L. Das and M. Chakraborty, "Improving the performance of the PNLMS algorithm using l_1 norm regularization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 7, pp. 1280–1290, Jul. 2016.

[33] Y. Li and M. Hamamura, "An improved proportionate normalized least-mean-square algorithm for broadband multipath channel estimation," *Sci. World J.*, vol. 2014, Art. no. 572969. [Online]. Available: https://doi.org/10.1155/2014/572969

[34] F. Albu, I. Caciula, Y. Li, and Y. Wang, "The l_p-norm proportionate normalized least mean square algorithm for active noise control," in *Proc. Int. Conf. Syst. Theory, Control, Comput.*, 2017, pp. 401–405.

[35] M. V. S. Lima, T. N. Ferreira, W. A. Martins, and P. S. R. Diniz, "Sparsity-aware data-selective adaptive filters," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4557–4572, Sep. 2014.

[36] K. Pelekanakis and M. Chitre, "New sparse adaptive algorithms based on the natural gradient and the l_0-norm," *IEEE J. Ocean. Eng.*, vol. 38, no. 2, pp. 323–332, 2013.

[37] T. N. Ferreira, M. V. S. Lima, P. S. R. Diniz, and W. A. Martins, "Low-complexity proportionate algorithms with sparsity-promoting penalties," in *Proc. IEEE Int. Symp. Circuits. Syst.*, 2016, pp. 253–256.

[38] J. Jin, Y. Gu, and S. Mei, "A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 409–420, Apr. 2010.

[39] C.-H. Lee, B. D. Rao, and H. Garudadri, "Proportionate adaptive filters based on minimizing diversity measures for promoting sparsity," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 769–773.

[40] T. Variddhisaï and D. P. Mandic, "On an RLS-like LMS adaptive filter," in *Proc. IEEE Int. Conf. Digit. Signal Process.*, 2017, pp. 1–5.

[41] G. Su, J. Jin, Y. Gu, and J. Wang, "Performance analysis of l_0 norm constraint least mean square algorithm," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2223–2235, May 2012.

[42] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[43] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: A majorization–minimization approach," *Signal Process.*, vol. 89, no. 9, pp. 1683–1693, 2009.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, U.K.: Cambridge Univ. Press, 2004.

[45] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 4th ed., Berlin, Germany: Springer, 2016.

[46] L. Chen and Y. Gu, "From least squares to sparse: A non-convex approach with guarantee," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 5875–5879.

[47] J. Benesty, C. Paleologu, and S. Ciochină, "On regularization in adaptive filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1734–1742, Aug. 2011.

[48] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 760–770, Mar. 2003.
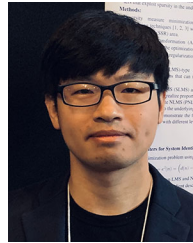
Ching-Hua Lee (Student Member, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2013. He is currently working toward the Ph.D. degree in electrical and computer engineering with the University of California, San Diego, CA, USA. His main research interests include adaptive filtering, sparse signal recovery, machine learning, and optimization, with applications to speech and audio signal processing.

Bhaskar D. Rao (Fellow, IEEE) is currently a Distinguished Professor with the Department of Electrical and Computer Engineering and the holder of the Ericsson Endowed Chair in wireless access networks with the University of California, San Diego, CA, USA. His research interests include digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and biomedical signal processing. He was the recipient of the 2016 IEEE Signal Processing Society Technical Achievement Award.

Harinath Garudadri (Member, IEEE) is currently a Research Scientist with the Qualcomm Institute, University of California, San Diego, San Diego, CA, USA. He moved to academia, in November 2013, after 26 years in the industry to work on technologies that will improve healthcare delivery beyond hospital walls. His research interests include signal processing applications in diverse fields, such as speech recognition, machine learning, speech, audio and video compression, multimedia delivery in 3G/4G networks, low-power sensing and telemetry of physiological data, reliable body area networks, noise cancellation, and artifacts mitigation, among other areas. He holds more than 52 granted patents and more than 20 pending patents in these areas.