Group Sparse RLS Algorithms

Ender M. Eksioglu
Electronics and Communications Engineering Department,
Istanbul Technical University, Istanbul, Turkey

Abstract

Group sparsity is one of the important signal priors for regularization of inverse problems. Sparsity with group structure is encountered in numerous applications. However, despite the abundance of sparsity based adaptive algorithms, attempts at group sparse adaptive methods are very scarce. In this paper we introduce novel Recursive Least Squares (RLS) adaptive algorithms regularized via penalty functions which promote group sparsity. We present a new analytic approximation for $\ell_{p,0}$ norm to utilize it as a group sparse regularizer. Simulation results confirm the improved performance of the new group sparse algorithms over regular and sparse RLS algorithms when group sparse structure is present.
Keywords: Adaptive filter; RLS; sparsity; group sparsity; block structure; mixed norm.

1 Introduction

Recursive Least Squares (RLS) algorithm is an important adaptive filtering technique. The computational complexity of RLS is quadratic in filter length per iteration, and the complexity is higher than the first order adaptive methods such as the Least Mean Square (LMS) algorithm. However, the relatively faster convergence speed of RLS compared to the first order methods makes RLS still an intriguing adaptive paradigm. Sparsity has been exploited as a significant prior condition in the LMS setting extensively. The proportionate adaptation concept was introduced with the proportionate normalized least-mean-squares (PNLMS) algorithm in the context of acoustic echo cancellation, where echo-path impulse response is assumed to be sparse [1]. Proportionate adaptation has been extended to selective partial update subband adaptive filters in [2]. In [3] the authors propose LMS variants exploiting sparsity under a framework of natural gradient algorithms. A modified proportionate updating method has been developed in [4]. In [5], the authors develop a robust NLMS algorithm for sparse adaptive filtering. Only recently there have been attempts at developing RLS variants which exploit the sparsity of the system or signal to be estimated. Time-weighted and time-and norm-weighted schemes for sparsity-aware recursive real-time filtering are presented in [6]. In [7] the authors develop a recursive ℓ_1-regularized least squares algorithm (SPARLS) for sparse system identification. SPARLS utilizes an expectation-maximization (EM) type algorithm to minimize the ℓ_1-regularized least squares cost function for streaming data. Nonlinear system identification using Volterra models where the Volterra kernels are assumed to be of a sparse nature is considered in [8]. An approximation of the ℓ_0-norm penalty is added to the RLS cost function for the Volterra system, and a recursive estimator based on coordinate descent approach is developed. An algorithm for the application of general convex constraints onto RLS was presented in [9] under the rubric of Convex Regularized RLS (CR-RLS). Sparsity aware variants of CR-RLS regularized with ℓ_1 norm and an analytic approximation to ℓ_0 norm have been proposed in [9, 10].

Another useful prior condition related to the sparsity is group or block sparsity. In the traditional notion of sparsity, the few nonzero coefficients occur randomly in a non-structured manner along the signal support. In group sparsity, these few nonzero (or significant) coefficients assemble in clusters in a structured manner. Group sparsity finds usage in a wide range of applications including source localization [11], image denoising [12], image classification [13], network inference [14] and cognitive spectrum sensing [15]. Literature on adaptive algorithms benefiting from group sparsity on the other hand is very scarce. A group sparse LMS algorithm is developed in [16] using mixed $\ell_{2,1}$ and reweighted $\ell_{2,1}$ norms as the convex penalties. An online, homotophy based solution for the minimization of the RLS cost function penalized by the $\ell_{\infty,1}$ norm is developed in [17].

In this paper, we develop novel group sparsity cognizant adaptive algorithms by using mixed $\ell_{p,1}$ norms and also analytic approximations to the mixed $\ell_{p,0}$ pseudo-norm. These group sparsity enhancing functions are utilized in the CR-RLS framework to develop the group sparse adaptive algorithms. This paper also proposes a novel exponential approximation to the mixed $\ell_{p,0}$ pseudo-norm. As part of the developed algorithms, the subgradients for the various group sparsity promoting functions are calculated. These group sparsity promoting penalties for which the subgradients are calculated include the $\ell_{2,1}$ and $\ell_{\infty,1}$ norms and the novel exponential approximations for the $\ell_{p,0}$ pseudo-norm. Simulations show that the resulting novel group sparse RLS algorithms are effective and robust online solvers for the group sparse system identification problem. The paper is organized as follows. Section 2 presents a recap on the CR-RLS scheme. In Section 3, first we introduce the group sparsity concept and some necessary tools. Afterwards, we develop the novel approximation for $\ell_{p,0}$ norm. In Section 4 we introduce the new group sparse RLS algorithms after calculating the subgradients for the various group sparsity based regularizing functions. Section 5 presents the simulation results which verify the superiority of the group sparse adaptive algorithms when applied in a setting where group sparse signal structure is present.

2 CR-RLS Algorithm Prior Art

We consider the supervised system identification setting with the linear input-output relation.

$$y_n = \widehat{\mathbf{h}}^T \mathbf{x}_n + v_n \tag{1}$$

y_n is the noisy output signal, $\widehat{\mathbf{h}} = [h_0, h_1, \ldots, h_{N-1}]^T \in \mathbb{R}^N$ is the actual system impulse response for the FIR system to be identified. $\mathbf{x}_n = [x_n, x_{n-1}, \ldots x_{n-N+1}]^T \in \mathbb{R}^N$ is the input vector at time n. v_n is the additive noise term corrupting the output of the system. Adaptive system identification methods estimate the system response $\widehat{\mathbf{h}}$ sequentially as the input and output signals stream in. In [9], the least-squares cost with exponential forgetting factor λ was regularized by a convex penalty, and the result was considered as the instantaneous cost function $J_n(\mathbf{h})$ which should be minimized as a function of the variable $\mathbf{h}$.

$$J_n(\mathbf{h}) = \frac{1}{2} \sum_{m=0}^{n} \lambda^{n-m} |y_m - \mathbf{h}^T \mathbf{x}_m|^2 + \gamma_n f(\mathbf{h}) \tag{2}$$

Here $f : \mathbb{R}^N \to \mathbb{R}$ is a convex, not necessarily differentiable function. $\gamma_n \geqslant 0$ is a possibly time-varying regularization parameter, which governs a tradeoff between the approximation error and the penalty function. Let $\widehat{\mathbf{h}}_n$ denote the instantaneous minimizer of the regularized cost function $J_n(\mathbf{h})$, that is $\widehat{\mathbf{h}}_n = \underset{\mathbf{h}}{\operatorname{argmin}} J_n(\mathbf{h})$. A recursive algorithm in the same vein as RLS for the approximate calculation of $\widehat{\mathbf{h}}_n$ was developed in [9] under the title of CR-RLS. The coefficient vector update equation is given as

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \xi_n \mathbf{P}_n \mathbf{x}_n - \gamma_{n-1}(1-\lambda)\mathbf{P}_n \nabla^S f(\mathbf{h}_{n-1}). \tag{3}$$

$\mathbf{P}_n$ is the inverse of the deterministic autocorrelation matrix estimate for the input signal. $\nabla^S f(\boldsymbol{\nu})$ denotes a subgradient of f at $\boldsymbol{\nu}$ [18]. $\xi_n = y_n - \mathbf{h}_{n-1}^T \mathbf{x}_n$ is the a priori estimation error. Let us assume that an upper bound for the convex constraint is given in the form $\rho \geqslant f(\widehat{\mathbf{h}})$. ρ is a constant which quantifies the a priori knowledge about the true system impulse response $\widehat{\mathbf{h}}$. In [9] a closed form formula for selecting the regularization parameter in the case of white input is given as follows:

$$\gamma_n = (\gamma'_n)_+ \tag{4}$$

where

$$\gamma'_n = 2 \frac{\frac{\mathrm{tr}(\mathbf{P}_n)}{N}\big(f(\mathbf{h}_n) - \rho\big) + \nabla^S f(\mathbf{h}_n)^T \mathbf{P}_n \boldsymbol{\epsilon}'_n}{\|\mathbf{P}_n \nabla^S f(\mathbf{h}_n)\|_2^2}. \tag{5}$$

Here, $\mathrm{tr}(\cdot)$ is the trace operator for the argument matrix. In (4) $(x)_+ = \max(0, x)$, and in (5) $\boldsymbol{\epsilon}'_n = \mathbf{h}_n - \widetilde{\mathbf{h}}_n$. Furthermore, $\widetilde{\mathbf{h}}_n$ is the instantaneous system estimate produced by the standard, non-regularized RLS algorithm [9]. The term $\widetilde{\mathbf{h}}_n$ is updated by the standard RLS equation given as $\widetilde{\mathbf{h}}_n = \widetilde{\mathbf{h}}_{n-1} + \xi_n \mathbf{P}_n \mathbf{x}_n$.

An additional observation which was not considered in [9] is that the second term in the numerator of (5) is in general much smaller than the first term in magnitude. Hence, ${\gamma'}_n$ in (5) can be approximated as

$$ {\gamma'}_n \approx 2 \frac{\frac{\mathrm{tr}(\mathbf{P}_n)}{N}\big(f(\mathbf{h}_n) - \rho\big)}{\|\mathbf{P}_n \nabla^S f(\mathbf{h}_n)\|_2^2}. \tag{6} $$

For the white input case, the selection of γ_n as given in (4) results in at least as good performance as the regular RLS. This formula also avoids the ad hoc regularization parameter selection process for different experimental setups. In the convex-regularized RLS algorithm as developed in [9] and refined in this paper, the degree of regularization is determined by the time-dependent variable γ_n. This update procedure calculates the regularization term from the data on-the-fly. Hence, there is no need for cross-validation to pre-determine a regularization parameter suitable for the data in hand.

3 Group Sparsity and Mixed Norms

In some applications it is the case that there is a group or block structure to the coefficient vector sought. The group structure means that the nonzero coefficient values occur in groups rather than being distributed randomly over the whole vector. Convex optimization algorithms such as Group Lasso [19] and Group Adaptive Lasso [20], and greedy algorithms such as Block Orthogonal Matching Pursuit (BOMP) [21] solve the group sparse representation problem. The superiority of using group Lasso over standard Lasso for group structured sparse coding in batch mode has been established [22, 23]. Encouraged by the batch group sparse representation algorithms, in this paper we develop group structured algorithms for block-sparsity aware adaptive filtering.

Some group sparsity inducing penalty functions utilized in the literature include mixed $\ell_{2,1}$ [19, 16] and reweighted $\ell_{2,1}$ norms [16], and smoothed $\ell_{2,0}$ pseudo-norm [11]. Here, we will formalize a novel group structure penalty, based on the exponential approximation of ℓ_0 pseudo-norm [24]. We assume that the group structure for the vector $\mathbf{h} \in \mathbb{R}^N$ is given by a group partition $\{\mathcal{G}_i\}_{i=1}^G$ of the index set $\mathcal{N} = \{0, 1, \ldots, N-1\}$ fullfilling the following [16]:

$$ \bigcup_{i=1}^{G} \mathcal{G}_i = \mathcal{N}, \mathcal{G}_i \bigcap \mathcal{G}_j = \varnothing \text{ for } i \neq j. \tag{7} $$

Hence, there are a total of G groups, and we do not allow group overlaps. We assume that the group structure is a priori known. We establish a different notation compared to [16] in presenting the group structure of vectors. We denote with $\mathbf{h}_{\mathcal{G}_i} \in \mathbb{R}^N$ the vector which is obtained by zeroing all values of the coefficient vector $\mathbf{h}$ except at the positions included in the set $\mathcal{G}_i$

$$ \{\mathbf{h}_{\mathcal{G}_i}\}_k = \begin{cases} h_k & \text{if } k \in \mathcal{G}_i; \\ 0 & \text{if } k \notin \mathcal{G}_i. \end{cases} \forall k = 0, \ldots, N-1 \tag{8} $$

Hence, the overall vector $\mathbf{h}$ is decomposed in the following manner:

$$\mathbf{h} = \sum_{i=1}^{G} \mathbf{h}_{\mathcal{G}_i}. \tag{9}$$

The mixed $\ell_{p,q}$ (also called as the ℓ_p/ℓ_q) norm corresponding to this group structure can be calculated as

$$\|\mathbf{h}\|_{p,q} = \Big(\sum_{i=1}^{G}(\|\mathbf{h}_{\mathcal{G}_i}\|_p)^q\Big)^{\frac{1}{q}} = \Bigg(\sum_{i=1}^{G}\Big(\sum_{k\in\mathcal{G}_i} h_k^p\Big)^{\frac{q}{p}}\Bigg)^{\frac{1}{q}}. \tag{10}$$

The $\ell_{p,q}$ norm is known to be convex when $p, q \geqslant 1$ [25].

The actual count for group sparsity is the mixed pseudo-norm $\|\mathbf{h}\|_{p,0}$ defined as

$$\|\mathbf{h}\|_{p,0} = \sum_{i=1}^{G} \mathcal{I}\big(\|\mathbf{h}_{\mathcal{G}_i}\|_p\big). \tag{11}$$

$\mathcal{I}(\cdot)$ is the indicator function producing one if the argument is non-zero and zero otherwise. However, this exact marker of group sparse structure is non-convex, and hence subgradient analysis is not applicable. One idea is to replace this non-convex penalty function $\|\mathbf{h}\|_{p,0}$ in (11) with its convex relaxation $\|\mathbf{h}\|_{p,1}$, just as the ℓ_0 norm is replaced with the ℓ_1 norm in the Basis Pursuit [26] (or Lasso [27]) formulations. In $\ell_{p,1}$, if p is chosen as $p = 1$ the mixed-norm reduces to the regular ℓ_1 norm. The two most popular $\ell_{p,1}$ mixed-norms used in practice to enforce structured sparsity are $\ell_{2,1}$ and $\ell_{\infty,1}$ [28].

In [16], the $\ell_{2,1}$ and the reweighted $\ell_{2,1}$ norms have been utilized as promoters of block sparsity. From (10), the $\ell_{2,1}$ norm can be calculated as follows:

$$\|\mathbf{h}\|_{2,1} = \sum_{i=1}^{G}\|\mathbf{h}_{\mathcal{G}_i}\|_2 = \sum_{i=1}^{G}\Big(\sum_{k\in\mathcal{G}_i} h_k^2\Big)^{\frac{1}{2}}. \tag{12}$$

The reweighted $\ell_{2,1}$ norm utilized in [16] is formulated in an analogous fashion to the reweighted ℓ_1 norm employed by Candès et al. in [29].

$$\|\mathbf{h}\|_{2,1_r} = \sum_{i=1}^{G} r_i \|\mathbf{h}_{\mathcal{G}_i}\|_2 \tag{13}$$

The positive valued reweighting parameters r_i are chosen as $r_i = \dfrac{1}{\|\mathbf{h}_{\mathcal{G}_i}\|_2 + \delta}$ [16], where δ is a small positive constant.

The $\ell_{\infty,1}$ norm has been used in [25] as a penalty function for simultaneous sparse signal approximation. The $\ell_{\infty,1}$ norm is called as the relaxed norm, and it is utilized as a convex substitute for the row-ℓ_0 quasi-norm of a matrix. An online homotophy based algorithm for solving the $\ell_{\infty,1}$ regularized Lasso problem is developed in [17]. Gradient methods for optimization of $\ell_{\infty,1}$ regularized loss functions have been developed in [30, 31]. In both papers the introduced algorithms have been applied to multi-task sparse learning problems. Using the notation developed in this section, $\ell_{\infty,1}$ norm is calculated in the following manner.

$$\|\mathbf{h}\|_{\infty,1} = \sum_{i=1}^{G} \|\mathbf{h}_{\mathcal{G}_i}\|_\infty = \sum_{i=1}^{G} \max_{k\in\mathcal{G}_i} |h_k| \tag{14}$$

Another way to replace the non-convex group sparsity count $\ell_{p,0}$ of (11) would be by introducing convex approximations to this function. In [11], Gaussian functions were utilized to approximate the non-convex $\|\mathbf{h}\|_{2,0}$. In this paper, we propose instead a novel approximation $\|\mathbf{h}\|_{p,0^\beta}$ based on the exponential function. An exponential approximation to the ℓ_0 pseudo-norm was formulated in [24]. Here, we introduce the below given novel approximation for the $\ell_{p,0}$.

$$\|\mathbf{h}\|_{p,0} \approx \|\mathbf{h}\|_{p,0^\beta} = G - \sum_{i=1}^{G} e^{-\beta\|\mathbf{h}_{\mathcal{G}_i}\|_p} \tag{15}$$

In (15), β is a proper positive constant, and $\lim_{\beta\to\infty}\|\mathbf{h}\|_{p,0^\beta} = \|\mathbf{h}\|_{p,0}$.

Equipped with these results, we can define various group sparsity promoting functions which can be utilized as the $f(\mathbf{h}_n)$ of (2), and we can also calculate their subgradients $\nabla^S f(\mathbf{h}_n)$. Some viable choices for $\nabla^S f(\mathbf{h}_n)$ are discussed in the following section.

4 Group Sparse RLS Algorithms

In this section we develop new group sparsity cognizant RLS algorithms by employing the group sparsity inducing mixed norms introduced in Section 3. We insert these mixed norms as the regularizing penalty functions into the cost function (2) and into the update equation (3).

4.1 $\ell_{2,1}$ Norm

One choice for $f(\cdot)$ to be used in (2) is the $\ell_{2,1}$ norm with $f(\mathbf{h}) = \|\mathbf{h}\|_{2,1}$. The function $\|\mathbf{h}\|_{2,1}$ as given in (12) is differentiable everywhere, except when $\|\mathbf{h}_{\mathcal{G}_i}\|_2 = 0$ or equivalently when $\mathbf{h}_{\mathcal{G}_i}$ is the all zero vector $\mathbf{0} \in \mathbb{R}^N$. At the point $\mathbf{h}_{\mathcal{G}_i} = \mathbf{0}$, the subdifferential includes the all zero vector as a legitimate subgradient, that is $\mathbf{0} \in \partial\|\mathbf{h}_{\mathcal{G}_i}\|_2$ when $\mathbf{h}_{\mathcal{G}_i} = \mathbf{0}$. Using these results, a subgradient for $f(\mathbf{h}_n) = \|\mathbf{h}_n\|_{2,1}$ can be written as follows.

$$\nabla^S\{\|\mathbf{h}_n\|_{2,1}\} = \sum_{i=1}^{G} \nabla^S \|\mathbf{h}_{n,\mathcal{G}_i}\|_2 \tag{16}$$

where,

$$\nabla^S \|\mathbf{h}_{n,\mathcal{G}_i}\|_2 = \begin{cases} \dfrac{\mathbf{h}_{n,\mathcal{G}_i}}{\|\mathbf{h}_{n,\mathcal{G}_i}\|_2}, & \mathbf{h}_{\mathcal{G}_i} \neq \mathbf{0}; \\ \mathbf{0}, & \mathbf{h}_{\mathcal{G}_i} = \mathbf{0}. \end{cases} \tag{17}$$

In [16], not this exact expression but an approximation for the subgradient is used. This approximation is given as,

$$\nabla^S\big\{\|\mathbf{h}_n\|_{2,1}\big\} \approx \sum_{i=1}^{G} \frac{\mathbf{h}_{n,\mathcal{G}_i}}{\|\mathbf{h}_{n,\mathcal{G}_i}\|_2+\delta} \tag{18}$$

where again δ is a small, positive constant.

A similar choice for $f(\mathbf{h})$ is the reweighted norm $f(\mathbf{h}) = \|\mathbf{h}\|_{2,1_r}$ (13). Using the results for $\ell_{2,1}$ norm, a subgradient of the reweighted norm can be written as follows:

$$\nabla^S\big\{\|\mathbf{h}_n\|_{2,1_r}\big\} = \sum_{i=1}^{G} r_{n,i}\nabla^S\|\mathbf{h}_{n,\mathcal{G}_i}\|_2. \tag{19}$$

Here, $\nabla^S\|\mathbf{h}_{n,\mathcal{G}_i}\|_2$ is as defined in (17). The approximation for this subgradient used in [16] is,

$$\nabla^S\big\{\|\mathbf{h}_n\|_{2,1_r}\big\} \approx \sum_{i=1}^{G} r_{n,i}\frac{\mathbf{h}_{n,\mathcal{G}_i}}{\|\mathbf{h}_{n,\mathcal{G}_i}\|_2+\delta}. \tag{20}$$

4.2 $\ell_{\infty,1}$ Norm

One other popular $\ell_{p,1}$ norm is the $\ell_{\infty,1}$ norm with $p=\infty$ [25, 28, 17, 31, 30]. In [25] the subdifferential for the row-ℓ_1 matrix norm is calculated. We follow a similar procedure to determine the subdifferential of $\ell_{\infty,1}$ norm for vectors. Let us first consider the simpler ℓ_∞ norm which does not entail any group structure. The subdiffential of ℓ_∞ is known to be [25]

$$\partial\|\mathbf{h}\|_\infty = \begin{cases} \{\mathbf{g} : \|\mathbf{g}\|_1 \leqslant 1\} & \text{if } \mathbf{h}=\mathbf{0} \\ \text{conv}\{\text{sgn}(h_k)\mathbf{e}_k : |h_k| = \max_j |h_j|\} & \text{otherwise.} \end{cases} \tag{21}$$

Here, conv denotes the convex hull of the union of the argument vectors. Furthermore, $\mathbf{e}_k \in \mathbb{R}^N, k = 0 \ldots N-1$ denote the canonical basis vectors, and $\text{sgn}(\cdot)$ possibly applied onto a vector denotes the elementwise sign function.

The $\ell_{\infty,1}$ norm as defined in (14) is separable with $\|\mathbf{h}\|_{\infty,1} = \sum_{i=1}^{G}\|\mathbf{h}_{\mathcal{G}_i}\|_\infty$. The inner product is also separable on the group partition $\{\mathcal{G}_i\}_{i=1}^{G}$. Hence, the subdifferential $\partial\|\mathbf{h}\|_{\infty,1}$ is simply found by applying the ℓ_∞ subdifferential $\partial\|\mathbf{h}\|_\infty$ in (21) to each group partition vector $\mathbf{h}_{\mathcal{G}_i}$ separately. Let $\nabla^S\big\{\|\mathbf{h}\|_{\infty,1}\big\} \in \partial\|\mathbf{h}\|_{\infty,1}$ denote a subgradient of the $\ell_{\infty,1}$ norm at $\mathbf{h}$. The subgradient vector should satisfy the following:

$$\nabla^S\big\{\|\mathbf{h}\|_{\infty,1}\big\}_{\mathcal{G}_i} \in \begin{cases} \{\mathbf{g} : \|\mathbf{g}\|_1 \leqslant 1\} & \text{if } \mathbf{h}_{\mathcal{G}_i}=\mathbf{0} \\ \text{conv}\{\text{sgn}(h_{\mathcal{G}_i,k})\mathbf{e}_k : |h_{\mathcal{G}_i,k}| = \max_j |h_{\mathcal{G}_i,j}|\} & \text{otherwise.} \end{cases} \tag{22}$$

4.3 $\ell_{p,0^\beta}$ Approximations

We can also use the convex approximations $\|\mathbf{h}\|_{p,0^\beta}$ as the group sparse penalty $f(\mathbf{h})$. A subgradient for $\|\mathbf{h}\|_{p,0^\beta}$ in (15) can be calculated using the chain rule for subgradient [18]

$$\nabla^S\big\{\|\mathbf{h}\|_{p,0^\beta}\big\} = \sum_{i=1}^{G} \beta \nabla^S\big\{\|\mathbf{h}_{\mathcal{G}_i}\|_p\big\} e^{-\beta\|\mathbf{h}_{\mathcal{G}_i}\|_p}. \tag{23}$$

The product of the two column vectors, $\nabla^S\big\{\|\mathbf{h}_{\mathcal{G}_i}\|_p\big\}$ and $e^{-\beta\|\mathbf{h}_{\mathcal{G}_i}\|_p}$ in (23), is just an elementwise multiplication. The exponential term $e^{-\beta\|\mathbf{h}_{\mathcal{G}_i}\|_p}$ in (23) can be approximated by its first order Taylor series expansion as given below:

$$e^{-\beta|x|} \approx \begin{cases} 1-\beta|x|, & |x| \leqslant \dfrac{1}{\beta}; \\ 0, & \text{elsewhere.} \end{cases} \tag{24}$$

The expansion in (24) can be more succinctly stated as $e^{-\beta|x|} \approx \big(1-\beta|x|\big)_+$. Substituting this expression into (23), we obtain the following result for the subgradient:

$$\nabla^S\big\{\|\mathbf{h}\|_{p,0^\beta}\big\} \approx \sum_{i=1}^{G} \beta \nabla^S\big\{\|\mathbf{h}_{\mathcal{G}_i}\|_p\big\}\big(1-\beta\|\mathbf{h}_{\mathcal{G}_i}\|_p\big)_+. \tag{25}$$

In this paper we will consider the two cases with $p = 1$ and $p = 2$. The subgradients for the corresponding penalty functions are calculated from (25). When $p = 1$, that is when $f(\mathbf{h}_n) = \|\mathbf{h}_n\|_{1,0^\beta}$, a corresponding subgradient is calculated in the following manner:

$$\nabla^S\big\{\|\mathbf{h}_n\|_{1,0^\beta}\big\} = \sum_{i=1}^{G} \beta\ \mathrm{sgn}\big(\mathbf{h}_{n,\mathcal{G}_i}\big)\big(1-\beta\|\mathbf{h}_{n,\mathcal{G}_i}\|_1\big)_+. \tag{26}$$

For $p = 2$, $f(\mathbf{h}_n) = \|\mathbf{h}_n\|_{2,0^\beta}$ and a matching subgradient can be computed as follows:

$$\begin{aligned} \nabla^S\big\{\|\mathbf{h}_n\|_{2,0^\beta}\big\} &= \sum_{i=1}^{G} \beta \nabla^S \|\mathbf{h}_{n,\mathcal{G}_i}\|_2\big(1-\beta\|\mathbf{h}_{n,\mathcal{G}_i}\|_2\big)_+ \\ &\approx \sum_{i=1}^{G} \frac{\beta}{\|\mathbf{h}_{n,\mathcal{G}_i}\|_2+\delta}\mathbf{h}_{n,\mathcal{G}_i}\big(1-\beta\|\mathbf{h}_{n,\mathcal{G}_i}\|_2\big)_+. \end{aligned} \tag{27}$$

4.4 Group Sparse Adaptive Algorithms

We will employ the group sparsity inducing penalty functions with their corresponding subgradients as given in (16), (22), (26) and (27) inside the update equation (3). We will call the resulting group sparsifying RLS algorithms as

Algorithm 1 $\ell_{1,0}$-Group RLS algorithm for adaptive filtering with structured sparsity.

$\lambda,\ x_n,\ y_n,\ \mathbf{h}_{-1}=\mathbf{0},\ \mathbf{P}_{-1}=\delta^{-1}\mathbf{I}_N$, group partition $\{\mathcal{G}_i\}_{i=1}^{G}$, ρ such that $\|\mathbf{h}\|_{1,0}\leqslant\rho$ ▷ inputs

1: **for** $n := 0, 1, 2, \ldots$ **do** ▷ time recursion

2: $\mathbf{k}_n = \dfrac{\mathbf{P}_{n-1}\mathbf{x}_n}{\lambda + \mathbf{x}_n^T\mathbf{P}_{n-1}\mathbf{x}_n}$ ▷ gain vector

3: $\xi_n = y_n - \mathbf{h}_{n-1}^T\mathbf{x}_n$ ▷ a priori error

4: $\mathbf{P}_n = \dfrac{1}{\lambda}\Big[\mathbf{P}_{n-1} - \mathbf{k}_n\mathbf{x}_n^T\mathbf{P}_{n-1}\Big]$

5: $\gamma_{n-1} = \left(\dfrac{2\ \mathrm{tr}(\mathbf{P}_{n-1})\big(\|\mathbf{h}_{n-1}\|_{1,0}-\rho\big)}{N\big\|\mathbf{P}_{n-1}\sum_{i=1}^{G}\beta\ \mathrm{sgn}\big(\mathbf{h}_{n-1,\mathcal{G}_i}\big)\big(1-\beta\|\mathbf{h}_{n-1,\mathcal{G}_i}\|_1\big)_+\big\|_2^2}\right)_+$ ▷ regularization parameter

6: $\mathbf{h}_n = \mathbf{h}_{n-1} + \xi_n\mathbf{k}_n - \gamma_{n-1}(1-\lambda)\mathbf{P}_n\Big(\sum_{i=1}^{G}\beta\ \ \mathrm{sgn}(\mathbf{h}_{n-1,\mathcal{G}_i})(1-\beta\|\mathbf{h}_{n-1,\mathcal{G}_i}\|_1)_+\Big)$

7: **end for** ▷ end of recursion

$\ell_{2,1}$-Group RLS ($\ell_{2,1}$-GRLS), $\ell_{\infty,1}$-GRLS, $\ell_{1,0}$-GRLS and $\ell_{2,0}$-GRLS, respectively. The $\ell_{1,0}$-GRLS algorithm resulting from the use of (26) is outlined in Algorithm 1. The realization for the other algorithms are obtained by replacing the subgradient vector with the corresponding subgradient function of the respective algorithm.

We have also implemented the group sparse adaptive algorithm using (19) as the subgradient vector. This variant can be appropriately called as $\ell_{2,1^r}$-GRLS. When γ_n is found by repeated trials as an optimized value, the algorithm with (19) performs on par with the $\ell_{p,0}$ methods. However, when (4) is used to calculate γ_n with (19) as the subgradient vector, the performance of $\ell_{2,1^r}$-GRLS falls in between RLS and $\ell_{2,1}$-GRLS. The γ_n values calculated using (4) are too conservative in this scenario, and they are small compared to the optimal γ value which would result in best steady-state performance. Hence, we have omitted the results of the $\ell_{2,1^r}$-GRLS which uses (19).

5 Simulation Results

We present results depicting the performance of the introduced algorithms in an adaptive group sparse system identification setting. The system impulse response $\mathbf{h}$ has a total of $N = 64$ coefficients. The impulse response is assumed to have a known group structure of $G = 16$ blocks with 4 coefficients in each block. The impulse response is generated as to have only K active blocks out of the total of $G = 16$. Hence, the total number of nonzero coefficients is $S = 4K$.

The K active groups are chosen randomly, and the coefficients in the active groups assume values from an i.i.d. $\mathcal{N}(0, \frac{1}{S})$ distribution. The input signal x_n takes its values from a zero-mean, unit variance normal distribution $\mathcal{N}(0, 1)$. A representative snapshot of the impulse response for the group sparse system with $K = 4$ is presented in Fig.1.

The observation noise is $v_n \sim \mathcal{N}(0, \sigma^2)$, where σ^2 takes differing values. Signal-to-noise ratio (SNR) is defined as SNR:$=\mathbb{E}\{\|\mathbf{y}_\text{o}\|_2^2\}/\sigma^2$, where $\mathbf{y}_\text{o}$ denotes the noise-free output signal. The main performance measure is the average mean squared deviation, which can be defined as MSD:$=\frac{\mathbb{E}\{\|\mathbf{h}_n - \widehat{\mathbf{h}}\|_2^2\}}{\mathbb{E}\{\|\widehat{\mathbf{h}}\|_2^2\}}$. The realized novel group sparse RLS algorithms include $\ell_{2,1}$-GRLS (16), $\ell_{\infty,1}$-GRLS (22), $\ell_{1,0}$-GRLS (26) and $\ell_{2,0}$-GRLS (27). We have also implemented the sparsity based CR-RLS algorithms ℓ_1-RLS and ℓ_0-RLS as introduced in [9]. All of these CR-RLS based algorithms are implemented with automated γ_n selection employing (4) and (6), hence there is no need to tweak the regularization parameter γ. For all algorithms $\lambda = 0.995$, and the results for each setting is averaged over 500 independent realizations. The β values to use are chosen after repeated simulations, as to find the values which result in minimum steady-state MSD for their respective algorithms. The utilized β values in the case of $\sigma^2 = 0.01$ are, $\beta = 50$ for ℓ_0-RLS, $\beta = 4$ for $\ell_{1,0}$-GRLS and $\beta = 7$ for $\ell_{2,0}$-GRLS. For $\sigma^2 = 0.1$ we chose, $\beta = 20$ for ℓ_0-RLS, $\beta = 3$ for $\ell_{1,0}$-GRLS and $\beta = 5$ for $\ell_{2,0}$-GRLS. Other parameters are $\delta = 10^{-5}$ for $\ell_{2,1}$-GRLS and $\ell_{2,0}$-GRLS.

5.1 Initial results and dependence on block-sparsity

In the first experiment, we realize the algorithms for a block-sparsity value $K = 1$ and noise variance values $\sigma^2 = 0.01$ and $\sigma^2 = 0.1$. $\sigma^2 = 0.01$ translates to an SNR of roughly 20 dB, whereas $\sigma^2 = 0.1$ results in an SNR of about 10 dB. The ρ value is assumed to be the true value of $f(\widehat{\mathbf{h}})$, $\rho = f(\widehat{\mathbf{h}})$. Hence, for $\ell_{2,1}$-GRLS $\rho = \|\widehat{\mathbf{h}}\|_{2,1}$, for $\ell_{\infty,1}$-GRLS $\rho = \|\widehat{\mathbf{h}}\|_{\infty,1}$, for $\ell_{1,0}$-GRLS and $\ell_{2,0}$-GRLS $\rho = \|\widehat{\mathbf{h}}\|_{p,0} = K = 1$, for ℓ_1-RLS $\rho = \|\widehat{\mathbf{h}}\|_1$ and for ℓ_0-RLS $\rho = \|\widehat{\mathbf{h}}\|_0 = S = 4$. The MSD curves for the two different SNR values are presented in Figs. 2 and

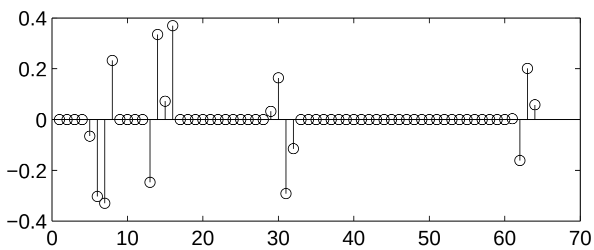

Figure 1: Typical group sparse impulse response used in the simulations with $K = 4$.

3. Secondly, we repeat the setup of the first experiment, but this time for a block sparsity value $K = 2$. All the other parameters are exactly the same as the earlier setup. We did not change the β values depending on this change of the block sparsity value K. The MSD curves pertaining to the two SNR values are given in Figs. 4 and 5. Figs. 2, 3, 4 and 5 demonstrate that the novel block-sparsity cognizant algorithms provide improved performance

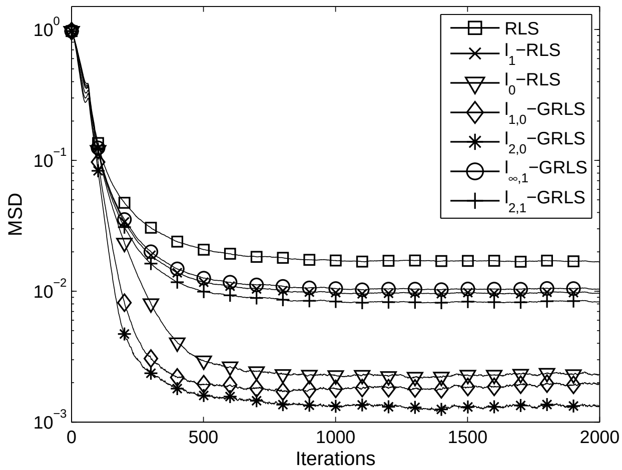

Figure 2: Performance of different algorithms for $K = 1$ and $\sigma^2 = 0.1$.

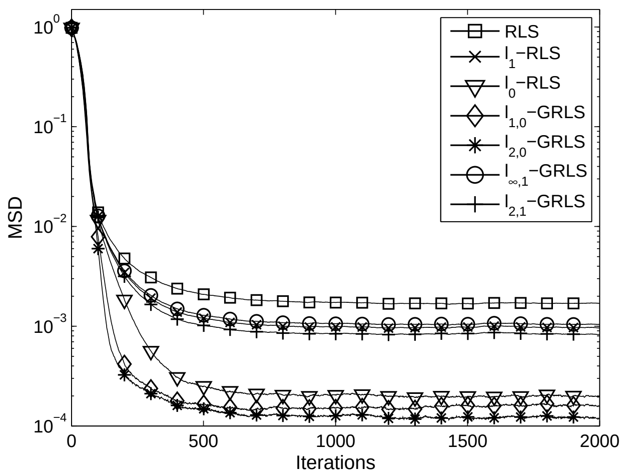

Figure 3: Performance of different algorithms for $K = 1$ and $\sigma^2 = 0.01$.

when compared to the algorithms which only exploit sparsity. The $\ell_{2,1}$-GRLS algorithm presents performance superior to the ℓ_1-RLS algorithm. The $\ell_{\infty,1}$-GRLS on the other hand performs very close but slightly worse than the ℓ_1-RLS. All three algorithms are inferior to the ℓ_0-RLS. The $\ell_{1,0}$-GRLS and $\ell_{2,0}$-GRLS algorithms which introduce a novel approximation for the $\ell_{p,0}$ norm, have the best performance. For all K and SNR values the $\ell_{p,0}$-GRLS algorithms outperform the ℓ_0-RLS which only utilizes the sparsity prior.

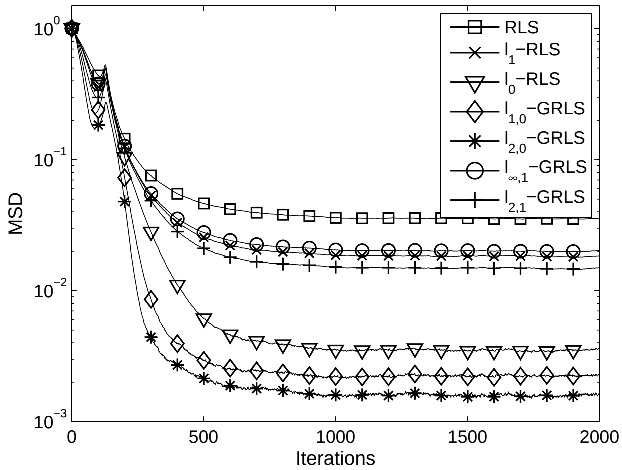

Figure 4: Performance of different algorithms for $K = 2$ and $\sigma^2 = 0.1$.

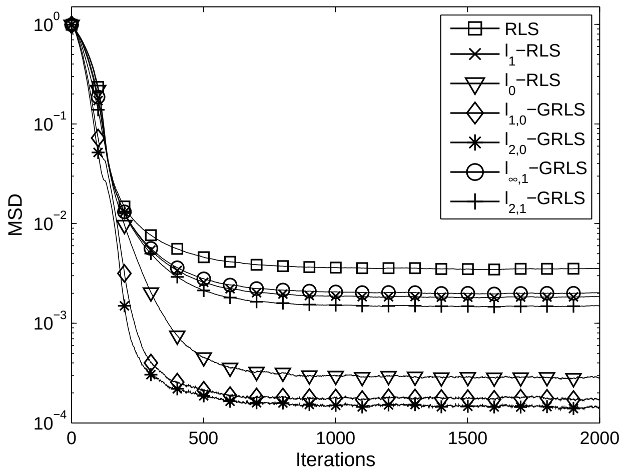

Figure 5: Performance of different algorithms for $K = 2$ and $\sigma^2 = 0.01$.

We have also implemented the previous experiments for additional block sparsity values. Tables 1 and 2 detail the average final MSD values for the various algorithms at the end of 2000 iterations. The group sparsity values vary as $K = 1, 2, 4$ and 16, where $K = 16$ corresponds to a system with fully active (non-sparse) impulse response. The results indicate that just like the sparse algorithms, e.g. ℓ_0-RLS, the novel group sparse algorithms do also naturally approach the standard RLS algorithm as the group sparsity vanishes. The results indicate that the novel $\ell_{p,0}$ approximations exploit the group structure better than the $\ell_{2,1}$ and $\ell_{\infty,1}$ norms.

5.2 Behavior of regularization parameter γ_n

The degree of regularization in the studied CR-RLS algorithms is determined by the time-varying variable γ_n. Parameter γ_n is updated on-the-fly as the data flows in. Figs. 6 and 7 present the time evolution of the regularization parameter for block-sparsity value $K = 1$ and noise variance values $\sigma^2 = 0.01$ and $\sigma^2 = 0.1$. These figures show that the regularization parameters γ_n converge to particular steady-state constants γ with time. These constants are suitable regularization parameters for the stationary noisy data and for the particular algorithm under consideration. We have also realized a select few of the CR-RLS algorithms using these steady-state γ values. Fig. 8 presents the MSD curves for ℓ_1-RLS, ℓ_0-RLS and $\ell_{2,0}$-GRLS algorithms using two approaches. First, the algorithms are run with time-varying, learned regularization parameter, γ_n. Secondly, the three algorithms are run using a constant regularization parameter, where the constant parameters γ are acquired from Fig. 6 as the steady-state values of the γ_n curves. The results in Fig. 8 suggest that the γ_n learning scheme works well, and the adaptive γ_n values converge to a constant γ which is suitable for the noisy data under consideration.

5.3 Inexact knowledge of $f(\widehat{\mathbf{h}})$

In this subsection, we study the non-ideal case $\rho > f(\widehat{\mathbf{h}})$, where again the parameter ρ quantifies our prior knowledge about the true system impulse response $\widehat{\mathbf{h}}$. In this subsection we assume the prior knowledge of the $f(\widehat{\mathbf{h}})$ value is inexact, and a coarse approximation is used. First, we present the results when the parameter ρ is non-ideally chosen as twice the value of the exact parameter $f(\widehat{\mathbf{h}})$. Figs. 9 and 10 show the MSD curves for the case $\rho = 2 \times f(\widehat{\mathbf{h}})$. From these figures it is seen that the ℓ_1-RLS, $\ell_{\infty,1}$-GRLS and $\ell_{2,1}$-GRLS algorithms are very sensitive to the use of inexact estimates for $f(\widehat{\mathbf{h}})$. At twice the exact value, the steady-state MSD for these variants deteriorate and converge to the result for the non-regularized standard RLS. On the other hand, the ℓ_0-RLS $\ell_{1,0}$-GRLS and $\ell_{2,0}$-GRLS algorithms are quite insensitive to the use of imperfect estimates for $f(\widehat{\mathbf{h}})$. In both figures the performances of the $\ell_{1,0}$-GRLS and $\ell_{2,0}$-GRLS for the case $\rho = 2 \times f(\widehat{\mathbf{h}})$, are quite similar to their performance when the exact parameter value was used. We have also realized simulations for other unideal cases of ρ, and the corresponding steady-state MSD results are listed in Tables 3 and 4. Table 3 is for SNR=10dB, and Table 4 is for SNR=20dB. The unideal choices for ρ range from $\rho = 2 \times f(\widehat{\mathbf{h}})$ up to $\rho = 16 \times f(\widehat{\mathbf{h}})$. The results in Tables 3 and 4 further affirm that the ℓ_0-RLS $\ell_{1,0}$-GRLS and $\ell_{2,0}$-GRLS algorithms are

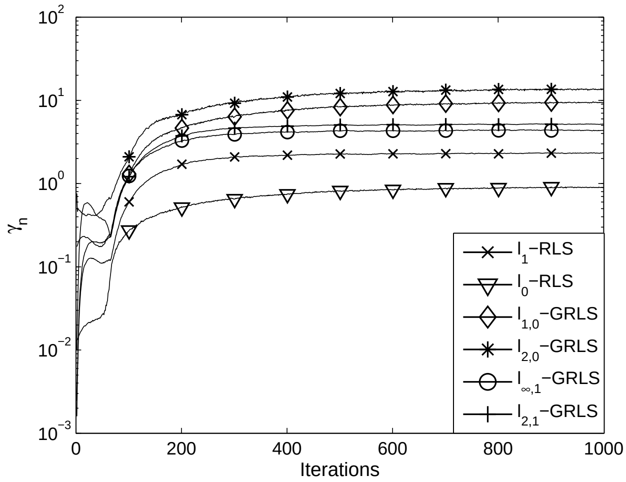

Figure 6: Time dependence of the regularization parameter γ_n for different algorithms, $K = 1$ and $\sigma^2 = 0.1$.

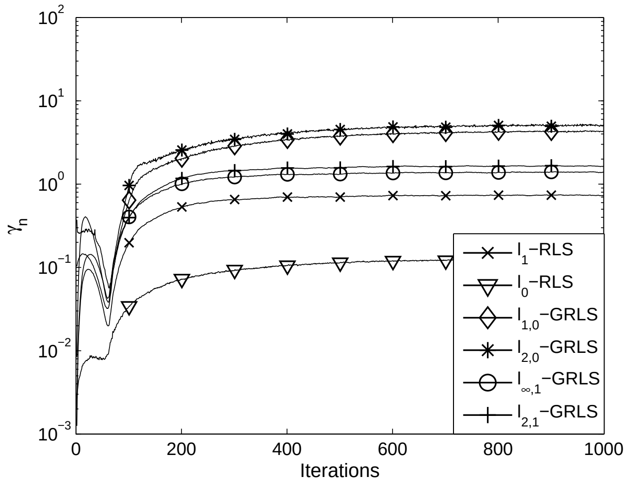

Figure 7: Time dependence of the regularization parameter γ_n for different algorithms, $K = 1$ and $\sigma^2 = 0.01$.

much more robust to inexact estimates of $f(\widehat{\mathbf{h}})$ when compared to the ℓ_1-RLS, $\ell_{\infty,1}$-GRLS and $\ell_{2,1}$-GRLS algorithms.

5.4 Varying length of impulse response $\widehat{\mathbf{h}}$

Here, we study the performance of the group sparse adaptive algorithms with respect to the length of the underlying true impulse response $\widehat{\mathbf{h}}$. Simulations for impulse response lengths ranging from 64 to 512 have been realized. The results for these simulations are listed in the Tables 5 and 6. In these tables, steady-state MSD for different impulse response length values N for constant sparsity ($K = 1$) are studied. Tables 5 and 6 show that the group sparse RLS algorithms exploit the group sparsity prior effectively for these longer impulse response lengths.

6 Conclusions

We have developed RLS algorithms which employ a priori knowledge of group sparsity for the underlying system to be identified. We have introduced a novel convex approximation for the $\ell_{p,0}$ pseudo-norms based on a similar approximation of the ℓ_0 pseudo-norm. The introduced group sparse RLS algorithms utilize these $\ell_{p,0}$ approximations and the convex $\ell_{2,1}$ and $\ell_{\infty,1}$ norms as regularizing functions. The simulations show that the group sparse algorithms, equipped with the automated regularizing parameter selection, exploit the group sparse prior efficiently. The group sparse algorithms in general perform better than the

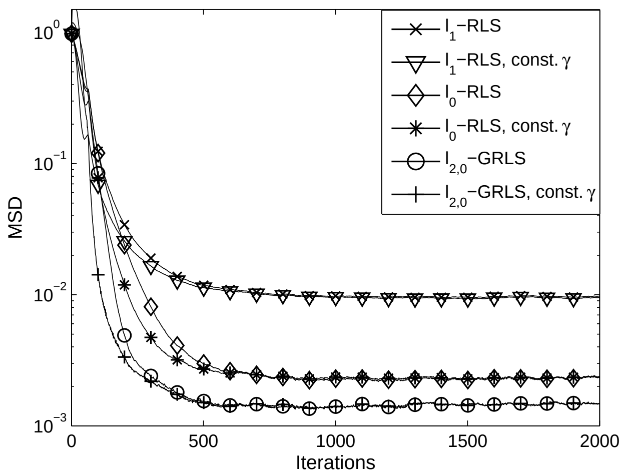

Figure 8: Performance of different algorithms with constant γ and time-varying γ_n, $K = 1$ and $\sigma^2 = 0.1$.

regular and sparse RLS algorithms under the group sparse setting. The group sparse algorithms which utilize the novel approximation for $\ell_{p,0}$ norm perform better than the convex relaxation based group sparse algorithms, which utilize the $\ell_{2,1}$ and $\ell_{\infty,1}$ norms. The $\ell_{p,0}$ based RLS algorithms are also much more robust to inexact estimates of the group sparsity prior when compared to the $\ell_{p,1}$ based algorithms. Additionally, the performance of all the group spare algorithms converge naturally to the regular RLS as the group sparsity prior vanishes.

References

[1] Duttweiler DL. Proportionate normalized least-mean-squares adaptation in echo cancelers. *IEEE Trans. Speech Audio Process.* Sep 2000; **8**(5):508–518, doi:10.1109/89.861368.

[2] Abadi MSE. Selective partial update and set-membership improved proportionate normalized subband adaptive filter algorithms. *International Journal of Adaptive Control and Signal Processing* 2010; **24**(9):786–804, doi:10.1002/acs.1172.

[3] Martin RK, Sethares WA, Williamson RC, Johnson J C R. Exploiting sparsity in adaptive filters. *IEEE Trans. Signal Process.* Aug 2002; **50**(8):1883–1894, doi:10.1109/TSP.2002.800414.

[4] Naylor PA, Cui J, Brookes M. Adaptive algorithms for sparse echo cancellation. *Signal Process.* 2006; **86**(6):1182–1192, doi: http://dx.doi.org/10.1016/j.sigpro.2005.09.015.

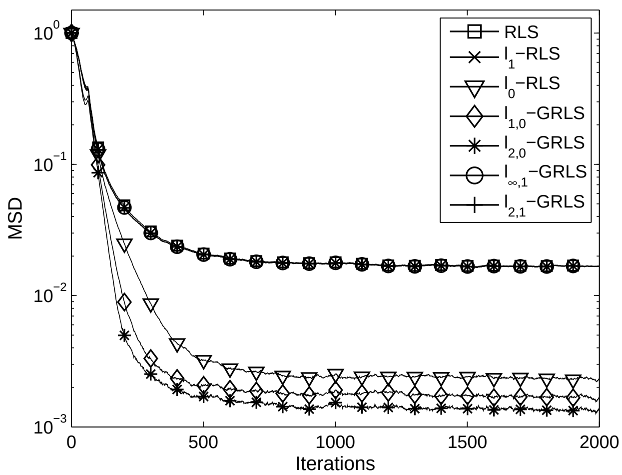

Figure 9: Performance of different algorithms for $\rho = 2 \times f(\widehat{\mathbf{h}})$ and $\sigma^2 = 0.1$.

[5] Vega LR, Rey H, Benesty J, Tressens S. A family of robust algorithms exploiting sparsity in adaptive filters. *IEEE Transactions on Audio, Speech, and Language Processing* May 2009; **17**(4):572–581, doi: 10.1109/TASL.2008.2010156.

[6] Angelosante D, Bazerque J, Giannakis G. Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1-norm. *IEEE Trans. Signal Process.* Jul 2010; **58**(7):3436–3447, doi:10.1109/TSP.2010.2046897.

[7] Babadi B, Kalouptsidis N, Tarokh V. SPARLS: The sparse RLS algorithm. *IEEE Trans. Signal Process.* Aug 2010; **58**(8):4013–4025, doi: 10.1109/TSP.2010.2048103.

[8] Shi K, Shi P. Adaptive sparse Volterra system identification with ℓ_0 norm penalty. *Signal Processing* 2011; **91**(10):2432–2436, doi: 10.1016/j.sigpro.2011.04.028.

[9] Eksioglu EM, Tanc AK. RLS algorithm with convex regularization. *IEEE Signal Process. Lett.* Aug 2011; **18**(8):470–473, doi: 10.1109/LSP.2011.2159373.

[10] Eksioglu EM. Sparsity regularised recursive least squares adaptive filtering. *IET Signal Processing* Aug 2011; **5**(5):480–487, doi:10.1049/iet-spr.2010.0083.

[11] Hyder M, Mahata K. Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation. *IEEE Trans. Signal Process.* Sep 2010; **58**(9):4646–4655, doi:10.1109/TSP.2010.2050477.

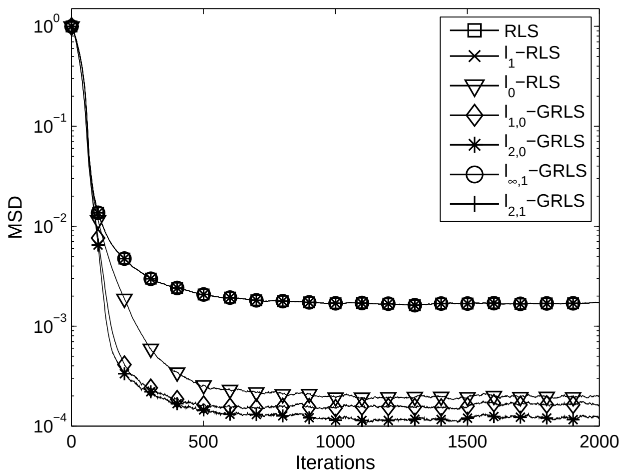

Figure 10: Performance of different algorithms for $\rho = 2 \times f(\widehat{\mathbf{h}})$ and $\sigma^2 = 0.01$.

[12] Peyre G, Fadili K. Group sparsity with overlapping partition functions. *Proc. EUSIPCO 2011*, 2011; 303–307.

[13] Bengio S, Pereira F, Singer Y, Strelow D. Group sparse coding. *Advances in Neural Information Processing Systems*, 2009.

[14] Bolstad A, Van Veen B, Nowak R. Causal network inference via group sparse regularization. *IEEE Trans. Signal Process.* Jun 2011; **59**(6):2628–2641, doi:10.1109/TSP.2011.2129515.

[15] Dall'Anese E, Bazerque J, Zhu H, Giannakis G. Group sparse total least-squares for cognitive spectrum sensing. *Signal Processing Advances in Wireless Communications (SPAWC), 2011 IEEE 12th International Workshop on*, 2011; 96–100, doi:10.1109/SPAWC.2011.5990487.

[16] Chen Y, Gu Y, Hero AO. Regularized Least-Mean-Square Algorithms. *ArXiv e-prints* Dec 2010; [Online]. Available: http://arxiv.org/abs/1012.5066v2.

[17] Chen Y, Hero A. Recursive $\ell_{1,\infty}$ group lasso. *IEEE Trans. Signal Process.* Aug 2012; **60**(8):3978 –3987.

[18] Bertsekas D, Nedic A, Ozdaglar A. *Convex Analysis and Optimization.* Athena Scientific, Cambridge, Massachusetts, 2003.

[19] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* 2006; **68**(Part 1):49–67.

[20] Wang H, Leng C. A note on adaptive group lasso. *Computational Statistics & Data Analysis* 2008; **52**(12):5277–5286, doi:10.1016/j.csda.2008.05.006.

[21] Eldar YC, Kuppinger P, Bölcskei H. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.* Jun 2010; **58**(6):3042–3054.

[22] Huang J, Zhang T. The benefit of group sparsity. *Annals of Statistics* 2010; **38**(4):1978–2004.

[23] Lv X, Bi G, Wan C. The group lasso for stable recovery of block-sparse signal representations. *IEEE Trans. Signal Process.* Apr 2011; **59**(4):1371 –1382, doi:10.1109/TSP.2011.2105478.

[24] Gu Y, Jin J, Mei S. l_0 norm constraint LMS algorithm for sparse system identification. *IEEE Signal Processing Letters* Sep 2009; **16**(9):774–777, doi:10.1109/LSP.2009.2024736.

[25] Tropp JA. Algorithms for simultaneous sparse approximation. part i-i: Convex relaxation. *Signal Processing* 2006; **86**(3):589–602, doi: 10.1016/j.sigpro.2005.05.031.

[26] Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* 1998; **20**(1):33–61.

[27] Tibshirani R. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* 1996; **58**(1):267–288.

[28] Bach F, Jenatton R, Mairal J, Obozinski G. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning* 2012; **4**(1):1–106.

[29] Candès EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications* 2008; **14**:877–905.

[30] Chen X, Pan W, Kwok J, Carbonell J. Accelerated gradient method for multi-task sparse learning problem. *Data Mining, 2009. ICDM '09. Ninth IEEE International Conference on*, 2009; 746 –751, doi: 10.1109/ICDM.2009.128.

[31] Quattoni A, Collins M, Carreras X, Darrell T. An efficient projection for l1 infinity regularization. *Proc. 26th International Conference on Machine Learning ICML '09*, 2009.

Table 1: Steady-state MSD for different group sparsity values with $\sigma^2 = 0.1$.

$\sigma^2 = 0.1$	$K = 1$	$K = 2$	$K = 4$	$K = 16$
RLS	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}
ℓ_1-RLS	8.5×10^{-3}	1.1×10^{-2}	1.3×10^{-2}	1.7×10^{-2}
$\ell_{2,1}$-GRLS	7.2×10^{-3}	9.8×10^{-3}	1.2×10^{-2}	1.7×10^{-2}
$\ell_{\infty,1}$-GRLS	9.1×10^{-3}	1.2×10^{-3}	1.4×10^{-2}	1.7×10^{-2}
ℓ_0-RLS	2.1×10^{-3}	3.8×10^{-3}	7.1×10^{-3}	1.7×10^{-2}
$\ell_{1,0}$-GRLS	1.5×10^{-3}	2.3×10^{-3}	5.1×10^{-3}	1.7×10^{-2}
$\ell_{2,0}$-GRLS	1.2×10^{-3}	1.8×10^{-3}	5.7×10^{-3}	1.7×10^{-2}

Table 2: Steady-state MSD for different group sparsity values with $\sigma^2 = 0.01$.

$\sigma^2 = 0.01$	$K = 1$	$K = 2$	$K = 4$	$K = 16$
RLS	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}
ℓ_1-RLS	1.0×10^{-3}	1.1×10^{-3}	1.4×10^{-3}	1.7×10^{-3}
$\ell_{2,1}$-GRLS	8.5×10^{-4}	1.0×10^{-3}	1.3×10^{-3}	1.7×10^{-3}
$\ell_{\infty,1}$-GRLS	1.1×10^{-3}	1.1×10^{-3}	1.4×10^{-3}	1.7×10^{-3}
ℓ_0-RLS	2.0×10^{-4}	3.1×10^{-4}	6.0×10^{-4}	1.7×10^{-3}
$\ell_{1,0}$-GRLS	1.5×10^{-4}	1.7×10^{-4}	7.1×10^{-4}	1.7×10^{-3}
$\ell_{2,0}$-GRLS	1.1×10^{-4}	1.4×10^{-4}	7.6×10^{-4}	1.7×10^{-3}

Table 3: Steady-state MSD for the case when there is inexact knowledge of $f(\widehat{\mathbf{h}})$, $\sigma^2 = 0.1$.

$\sigma^2 = 0.1$	$\rho = f(\widehat{\mathbf{h}})$	$\rho = 2 \times f(\widehat{\mathbf{h}})$	$\rho = 4 \times f(\widehat{\mathbf{h}})$	$\rho = 8 \times f(\widehat{\mathbf{h}})$	$\rho = 12 \times f(\widehat{\mathbf{h}})$	$\rho = 16 \times f(\widehat{\mathbf{h}})$
RLS	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}
ℓ_1-RLS	8.5×10^{-3}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}
$\ell_{2,1}$-GRLS	7.2×10^{-3}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}
$\ell_{\infty,1}$-GRLS	9.1×10^{-3}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}	1.7×10^{-2}
ℓ_0-RLS	2.1×10^{-3}	2.4×10^{-3}	2.7×10^{-3}	3.6×10^{-3}	6.7×10^{-3}	1.7×10^{-2}
$\ell_{1,0}$-GRLS	1.5×10^{-3}	1.7×10^{-3}	1.9×10^{-3}	2.0×10^{-3}	3.5×10^{-3}	1.7×10^{-2}
$\ell_{2,0}$-GRLS	1.2×10^{-3}	1.3×10^{-3}	1.4×10^{-3}	1.5×10^{-3}	2.2×10^{-3}	1.7×10^{-2}

Table 4: Steady-state MSD for the case when there is inexact knowledge of $f(\widehat{\mathbf{h}})$, $\sigma^2 = 0.01$.

$\sigma^2 = 0.1$	$\rho = f(\widehat{\mathbf{h}})$	$\rho = 2 \times f(\widehat{\mathbf{h}})$	$\rho = 4 \times f(\widehat{\mathbf{h}})$	$\rho = 8 \times f(\widehat{\mathbf{h}})$	$\rho = 12 \times f(\widehat{\mathbf{h}})$	$\rho = 16 \times f(\widehat{\mathbf{h}})$
RLS	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}
ℓ_1-RLS	1.0×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}
$\ell_{2,1}$-GRLS	8.5×10^{-4}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}
$\ell_{\infty,1}$-GRLS	1.1×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}
ℓ_0-RLS	2.0×10^{-4}	2.1×10^{-4}	2.3×10^{-4}	3.1×10^{-4}	6.1×10^{-4}	1.7×10^{-3}
$\ell_{1,0}$-GRLS	1.5×10^{-4}	1.5×10^{-4}	1.6×10^{-4}	1.6×10^{-4}	1.8×10^{-4}	1.7×10^{-3}
$\ell_{2,0}$-GRLS	1.1×10^{-4}	1.4×10^{-4}	1.5×10^{-4}	1.6×10^{-4}	1.8×10^{-4}	1.7×10^{-3}

Table 5: Steady-state MSD for different impulse response length values N, $\sigma^2 = 0.1$.

$\sigma^2 = 0.1$	$N = 64$	$N = 128$	$N = 192$	$N = 256$	$N = 320$	$N = 384$	$N = 448$	$N = 5$
RLS	1.7×10^{-2}	3.6×10^{-2}	5.5×10^{-2}	7.5×10^{-2}	9.7×10^{-2}	1.2×10^{-1}	1.4×10^{-1}	1.5×10
ℓ_1-RLS	8.5×10^{-3}	1.8×10^{-2}	2.8×10^{-2}	3.7×10^{-2}	4.8×10^{-2}	6.0×10^{-2}	7.0×10^{-2}	8.0×10
$\ell_{2,1}$-GRLS	7.2×10^{-3}	1.5×10^{-2}	2.1×10^{-2}	2.8×10^{-2}	3.5×10^{-2}	4.3×10^{-2}	4.9×10^{-2}	5.6×10
$\ell_{\infty,1}$-GRLS	9.1×10^{-3}	2.0×10^{-2}	3.0×10^{-2}	4.1×10^{-2}	5.4×10^{-2}	6.7×10^{-2}	7.8×10^{-2}	9.0×10
ℓ_0-RLS	2.1×10^{-3}	2.8×10^{-3}	3.8×10^{-3}	5.6×10^{-3}	8.2×10^{-3}	1.0×10^{-2}	1.5×10^{-2}	2.0×10
$\ell_{1,0}$-GRLS	1.5×10^{-3}	1.8×10^{-3}	1.9×10^{-3}	2.6×10^{-3}	3.8×10^{-3}	4.8×10^{-3}	6.8×10^{-3}	1.0×10
$\ell_{2,0}$-GRLS	1.2×10^{-3}	1.5×10^{-3}	1.6×10^{-3}	2.0×10^{-3}	2.7×10^{-3}	3.6×10^{-3}	4.6×10^{-3}	5.7×10

Table 6: Steady-state MSD for different impulse response length values N, $\sigma^2 = 0.01$.

$\sigma^2 = 0.01$	$N = 64$	$N = 128$	$N = 192$	$N = 256$	$N = 320$	$N = 384$	$N = 448$	$N = 5$
RLS	1.7×10^{-3}	3.6×10^{-3}	5.4×10^{-3}	7.6×10^{-3}	9.9×10^{-3}	1.2×10^{-2}	1.4×10^{-2}	1.6×10
ℓ_1-RLS	1.0×10^{-3}	1.8×10^{-3}	2.7×10^{-3}	3.7×10^{-3}	4.9×10^{-3}	6.0×10^{-3}	7.0×10^{-3}	8.1×10
$\ell_{2,1}$-GRLS	8.5×10^{-4}	1.5×10^{-3}	2.1×10^{-3}	2.7×10^{-3}	3.6×10^{-3}	4.2×10^{-3}	4.9×10^{-3}	5.5×10
$\ell_{\infty,1}$-GRLS	1.1×10^{-3}	2.0×10^{-3}	3.0×10^{-3}	4.1×10^{-3}	5.5×10^{-3}	6.7×10^{-3}	7.8×10^{-3}	9.1×10
ℓ_0-RLS	2.0×10^{-4}	2.9×10^{-4}	3.8×10^{-4}	5.5×10^{-4}	8.3×10^{-4}	1.1×10^{-3}	1.5×10^{-3}	1.9×10
$\ell_{1,0}$-GRLS	1.5×10^{-4}	1.8×10^{-4}	2.0×10^{-4}	2.6×10^{-4}	3.7×10^{-4}	4.7×10^{-4}	6.8×10^{-4}	8.9×10
$\ell_{2,0}$-GRLS	1.1×10^{-4}	1.5×10^{-4}	1.7×10^{-4}	2.1×10^{-4}	2.7×10^{-4}	3.5×10^{-4}	4.5×10^{-4}	5.6×10