Correntropy Induced Metric Penalized Sparse RLS Algorithm to Improve Adaptive System Identification

Guan Gui[1], Linglong Dai[2], Baoyu Zheng[1], Li Xu[3], and Fumiyuki Adachi[4]
1. College of Communication and Information Engineering, NUPT, Nanjing, China
2. Tsinghua National Laboratory for Information Science and Technology (TNlist), Tsinghua University, China
3. Department of Electronics and Information Systems, Akita Prefectural University, Yurihonjo, Japan
4. Department of Communication Engineering, Graduate School of Engineering, Tohoku University, Sendai, Japan
Email: guiguan@outlook.jp, daill@tsinghua.edu.cn, adachi@ecei.tohoku.ac.jp

***Abstract*—Sparse adaptive filtering algorithms are utilized to exploit potential sparse structure information as well as to mitigate noises in many unknown sparse systems. Sparse recursive least square (RLS) algorithms have been attracted intensely attentions due to their low-complexity and easy-implementation. Basically, these algorithms are constructed by standard RLS algorithm and sparse penalty functions (e.g., ℓ_1-norm). However, existing sparse RLS algorithms do not exploit the sparsity efficiently. In this paper, an improved adaptive filtering algorithm is proposed by incorporating a novel correntropy induced metric (CIM) constraint into RLS, which is termed as RLS-CIM algorithm. Specifically, we adopt a well-known Gaussian kernel in CIM and further devise a novel variable kernel width to control the sparse penalty in different transient-error scenarios. Numerical simulation results are given to corroborate the proposed algorithm via mean square deviation (MSD).**

Keywords—sparse adaptive filtering; correntropy induced metric (CIM); recursive least square (RLS); adaptive system identification.

I. INTRODUCTION

Adaptive filtering algorithms have been successfully applied in many system identifications such as sparse channel estimation [1]–[3] and acoustic echo cancellation [4]. In particular, recursive least square (RLS) algorithm and its variants are considered as one of most important techniques due to their robust performance and fast convergence [5]. In particular, the convergence rate is even faster than the well-known least mean square (LMS) algorithm [5]. These advances are also attractive to develop sparse RLS algorithms with application to system identifications. By incorporating sparse penalty function (e.g., ℓ_1-norm and ℓ_0-norm) into standard RLS algorithm, recently, several sparse regularization RLS algorithms have been developed. In [6], expectation-maximization (EM) approach based sparse recursive ℓ_1-norm reguxlarized least square (SPARLS) is proposed to identify sparse systems. In [7], the authors propose the application of an online coordinate descent algorithm together with the least-squares cost function penalized by an ℓ_1-norm. In [8], the authors proposes two sparse RLS algorithms by using convex regularization (i.e., ℓ_1-norm and approximate ℓ_0-norm), which are termed as RLS-L1 and RLS-L0. Indeed, these state-of-the-art algorithms have been successfully applied into adaptive sparse system identification.

However, existing algorithms can be further improved if we can find the more accurate sparse approximation function in accordance of compressive sensing [9], [10]. In other words, the identification performance can be further improved if we can exploited more accurate system sparsity by using stronger sparse penalty function.

It is shown that the correntropy induced metric (CIM) [11] as a nonlinear metric in the input space can provide a good approximation for the ℓ_0-norm function. Motivated by this fact, in this paper, we propose an improved sparse adaptive filtering algorithm by incorporating CIM into RLS, which is termed as RLS-CIM algorithm. Our contribution of this paper can be summarized as follows. *Firstly*, we introduce concept of the correntropy and adopt general Gaussian kernel for CIM. In order to exploit the sparsity more efficiently, we propose a novel variable kernel width (VKW) to control CIM so that it can adaptively approach to optimal ℓ_0-norm in different transient-state mean square deviation (MSD) scenarios. *Secondly*, we derive the basic update equation of the proposed RLS-CIM algorithm as well as propose an adaptive method to select suitable regularization parameter which can balance the transient error and the system sparsity. *Finally*, several representative simulation results are provided to validate our proposed RLS-CIM algorithm in contrast to existing sparse RLS algorithms.

The rest of the paper is organized as follows. In Section II, system model is introduced and also general sparse RLS algorithms are briefly reviewed. In Section III, an improved RLS-CIM algorithm is proposed and corresponding variable kernel width (VKD) is also devised to adaptively exploit system sparsity. In Section IV, numerical simulations are conducted against channel sparsity as well as forgetting factor. Finally, this paper is concluded in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

The unknown coefficients to be identified and the input signal at time instant n are denoted, respectively, by $\mathbf{w}=[w_1,w_2,\cdots,w_N]^T$ and $\mathbf{x}_n=[x_n,x_{n-1},\cdots,x_{n-N+1}]^T$, where both $\mathbf{w}$ and $\mathbf{x}_n$ are real-valued vectors, N is the length of the unknown system, and $(\cdot)^T$ represents the transposition operation. The observed output signal d_n is

$$d_n = \mathbf{w}^T \mathbf{x}_n + z_n, \tag{1}$$

where z_n denotes additive random noise variable. The identification error between the output of the unknown system and the adaptive filter can be expressed as

$$e_n = d_n - \mathbf{w}_n^T \mathbf{x}_n = z_n - (\mathbf{w}_n - \mathbf{w})^T \mathbf{x}_n, \tag{2}$$

where $\mathbf{w}_n = [w_{1,n}, w_{2,n}, \cdots, w_{N,n}]^T$ denotes the adaptive filtering vector. Then the regular RLS cost function with exponential forgetting factor λ is defined as

$$\xi_n = \sum_{m=0}^{n} \lambda^{n-m} e_m^2. \tag{3}$$

To take advantage of system sparsity, we revise the RLS cost function (3) by combing a convex penalty function, which can be chose to exploit any potential sparse prior knowledge about the true system. Hence, cost function of the sparse RLS algorithms can be constructed as

$$\mathcal{G}(\mathbf{w}_n) \triangleq \frac{1}{2}\xi_n + \gamma_n \mathcal{S}(\mathbf{w}_n), \tag{4}$$

where $\mathcal{S}(\mathbf{w}_n)$ is a sparsity-aware convex function; $\gamma_n \geq 0$ is the possibly time-varying regularization parameter which balances between sparsity penalty term and the estimation error. We expect that to find the possible optimal steady-state identification vector $\mathbf{w}_n$ which minimizes the regularized cost function $\mathcal{G}(\mathbf{w}_n)$. It is worth noting that convex and non-differentiable functions sub-gradient analysis can provide a substitute for the gradient when finding this minimum. That is to say, even if the convex function $\mathcal{S}(\cdot)$ fails to be differentiable, there still exist possibly valid sub-gradient vectors $\mathbf{w}_n$. In this case, all the sub-gradients set are called as the sub-differential of $\mathcal{S}(\cdot)$ and is denoted by $\partial\mathcal{S}(\mathbf{w}_n)$. Here we denote a sub-gradient vector $\mathcal{S}(\cdot)$ at $\mathbf{w}_n$ with $\nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n) \in \partial\mathcal{S}(\mathbf{w}_n)$. Hence, a valid sub-gradient vector of $\mathcal{S}(\cdot)$ with respect to $\mathbf{w}_n$ can be written as follows,

$$\nabla_{\mathbf{w}}\mathcal{G}(\mathbf{w}_n) = \frac{1}{2}\nabla_{\mathbf{w}}\xi_n + \gamma_n \nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n). \tag{5}$$

One theorem from the sub-differential calculus states that a transient point $\mathbf{w}_n$ minimizes a convex function $\mathcal{G}(\cdot)$ if and only if $0 \in \partial\mathcal{G}(\mathbf{w}_n)$. That is to say, finding the optimal solution $\mathbf{w}_n$ is equivalent to minimizes $\mathcal{G}(\cdot)$. After evaluating the gradient $\nabla_{\mathbf{w}}\xi_n$ and setting the sub-gradient $\nabla_{\mathbf{w}}\mathcal{G}(\mathbf{w}_n)$ equal to 0, the relation for the i-th term can be formulated as

$$\sum_{m=0}^{n} \lambda^{n-m}\left\{y_m - \sum_{k=1}^{N} w_{k,i} x_{m-k}\right\} x_{m-i} = \gamma_n \left\{\nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n)\right\}_i. \tag{6}$$

The relations for all $i = 1, 2, \cdots, N$ can be written together in a matrix-vector forms as a set of modified normal equations

$$\mathbf{\Phi}_n \mathbf{w}_n = \mathbf{r}_n - \gamma_n \nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n), \tag{7}$$

where $\mathbf{\Phi}_n$ is the deterministic autocorrelation matrix which be estimated from the input signal $\mathbf{x}_n$. Hence, $\mathbf{\Phi}_n$ can be adaptively updated as

$$\mathbf{\Phi}_n = \sum_{m=0}^{n} \lambda^{n-m} \mathbf{x}_m \mathbf{x}_m^T = \lambda \mathbf{\Phi}_{n-1} + \mathbf{x}_n \mathbf{x}_n^T, \tag{8}$$

and $\mathbf{r}_n$ is the deterministic cross-correlation estimate vector between y_n and x_n

$$\mathbf{r}_n = \sum_{m=1}^{n} \lambda^{n-m} y_m \mathbf{x}_m = \lambda \mathbf{r}_{n-1} + y_n \mathbf{x}_n. \tag{9}$$

Please note that both $\mathbf{\Phi}_n$ and $\mathbf{r}_n$ have rank-one update equations associated with them. For the right hand side of (7), a new variable $\mathbf{\theta}_n$ can be defined as

$$\mathbf{\theta}_n = \mathbf{r}_n - \gamma_n \nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n). \tag{10}$$

The update equation (9) and definition (10) together lead to an update equation for $\mathbf{\theta}_n$. Assuming that γ_n and $\nabla_{\mathbf{w}}\mathcal{G}(\mathbf{w}_n)$ do not change considerably over a single time step, this update equation can be approximately written as

$$\mathbf{\theta}_n \approx \lambda \mathbf{\theta}_{n-1} + y_n \mathbf{x}_n - \gamma_{n-1}(1-\lambda)\nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_{n-1}). \tag{11}$$

We define the inverse of the autocorrelation matrix by $\mathbf{P}_n = \mathbf{\Phi}_n^{-1}$. According to the matrix inverse lemma and Eq. (8), one can obtain the update equation for $\mathbf{P}_n$ as

$$\mathbf{P}_n = \lambda^{-1}\left\{\mathbf{P}_{n-1} - \mathbf{k}_n \mathbf{x}_n^T \mathbf{P}_{n-1}\right\}, \tag{12}$$

where $\mathbf{k}_n$ is the gain vector defined as

$$\mathbf{k}_n = \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{\lambda + \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n}. \tag{13}$$

According to Eq. (12), Eq. (7) can be rewritten as

$$\mathbf{w}_n = \mathbf{P}_n \mathbf{\theta}_n. \tag{14}$$

After evaluating (14) using the recursions (11) and (12), we can obtain the update equation for $\mathbf{w}_n$ as

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{k}_n \xi_n - \gamma_{n-1}(1-\lambda)\mathbf{P}_n \nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_{n-1}), \tag{15}$$

where $\xi_n = y_n - \mathbf{w}_{n-1}^T \mathbf{x}_n$ denotes the a prior estimation error. Then the update equation for the standard a prior RLS algorithm can be written as

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{k}_n \xi_n = \mathbf{w}_{n-1} + \mathbf{k}_n (y_n - \mathbf{w}_{n-1}^T \mathbf{x}_n). \tag{16}$$

III. THE PROPOSED CIM-RLS ALGORITHM

In this paper, we introduce CIM as a sparsity penalty function to develop the improved sparse RLS algorithm. We first review the correntropy [12] which is novel nonlinear similarity between two random variables and it can induce a metric in the sample space. Given two vectors,

$\mathbf{w}=[w_1,w_2,...,w_N]^T$ and $\mathbf{u}=[u_1,u_2,...,u_N]^T$ in the sample space, the CIM is defined by

$$\text{CIM}(\mathbf{w},\mathbf{u})\triangleq\sqrt{\kappa(0)-\mathbf{v}(\mathbf{w},\mathbf{u})}, \quad (17)$$

where $\kappa(0)=1/(\sigma\sqrt{2\pi})$ and $\mathbf{v}(\mathbf{w},\mathbf{u})=(1/N)\sum_{i=1}^{N}\kappa(w_i,u_i)$ is the sample estimation of the correntropy. Without loss of generality, the most popular Gaussian kernel in correntropy is

$$\kappa(\mathbf{w},\mathbf{u})=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\mathbf{s}^2/2\sigma^2\right), \quad (18)$$

where $\mathbf{s}=\mathbf{w}-\mathbf{u}$, and σ denotes the width of Gaussian kernel. The CIM provides a close approximation for the ℓ_0-norm function. Given a vector $\mathbf{w}=[w_1,w_2,...,w_N]^T$, it can be well approximated by [16]

$$\|\mathbf{w}\|_0\sim\text{CIM}^2(\mathbf{w},0)=\frac{\kappa(0)}{N}\sum_{i=1}^{N}\left(1-\exp\left(-w_i^2/2\sigma^2\right)\right). \quad (19)$$

where $\kappa(0)=1/(\sigma\sqrt{2\pi})$. It can be shown that if $|w_i|>\delta$, $\forall x_i\neq 0$, then as $\sigma\to 0$, one can get a sparse solution arbitrarily close to that of the ℓ_0-norm, where δ is a small positive number. A proper value of the kernel width makes CIM a close approximation to ℓ_0-norm [11]. As an approximation of the ℓ_0-norm, the CIM favors sparsity and can be used as a penalty term in sparse channel estimation. Hence, CIM sparse penalty function can be denoted as

$$\mathcal{S}(w_n)=\frac{\kappa(0)}{N}\sum_{i=1}^{N}\left(1-\exp(-w_{n,i}^2/2\sigma^2)\right), \quad (20)$$

and its corresponding gradient descend can be derived as

$$\nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n)=\frac{\kappa(0)\mathbf{w}_n}{N\sigma^2}\exp\left(-\mathbf{w}_n^2/2\sigma^2\right). \quad (21)$$

Accurately, the fixed kernel width σ does not efficiently exploit the sparsity. To fully take advantage of the sparsity, this paper proposes an iteration-promoting exponential-decay VKD σ_n [2] as

$$\sigma_n=\sigma_{\max}\exp\left(-20n/N_{iter}\right)+\sigma_{\min}, \quad (22)$$

where $\sigma_{\max}$ and $\sigma_{\min}$ denote maximal and minimal kernel width, respectively; N_{iter} stands for number of iteration times.

Remark: To evaluate an adaptive filtering algorithm, convergence speed, steady-state MSD performance, and computational complexity are the three basic index. That is to say, at the cost of possible lowest computational complexity, the fast convergence speed is to seek in the initial large identification error scenarios while the good identification accuracy is dominant. In addition, it is very hard to exploit accurate sparsity in the case of large error. Hence, a suitable scheme is to devise a VKD which can relax the sparse penalty in the large error while enhance the sparse penalty in the small error. In such situation, one may devise a VKD σ_n which depends on the transient error ξ_n. But it will generate heavy computational burden for the proposed RLS-CIM algorithm. Hence, we proposed the novel low-complexity VKD in (22), which can reduce kernel width via iteration-promoting exponential-decay as the MSD performance improvement. Hence, the proposed RLS-CIM can be exploited efficiently by using our proposed VKD.

This CIM constraint imposes a zero attraction of the filter coefficients according to the relative value of each coefficient among all the entries which in turn leads to an improved performance when the system is sparse. It is well known the regularization parameter term $\gamma_n(1-\lambda)\mathbf{P}_n\nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n)$ in Eq. (15) is one of important steps to decide the identification performance. In [8], an sophisticated regularization selection approach is proposed. For the proposed RLS-CIM algorithm, we also consider the selection method to calculate γ_n as

$$\gamma_n'=2\frac{\left(\text{tr}(\mathbf{P}_n)/N\right)\left(\mathcal{S}(\mathbf{w}_n)-\rho\right)+\nabla_{\mathbf{w}}\mathcal{S}^T(\mathbf{w}_n)\mathbf{P}_n\boldsymbol{\varepsilon}_n'}{\|\mathbf{P}_n\nabla_{\mathbf{w}}\mathcal{S}(\mathbf{w}_n)\|_2^2}, \quad (23)$$

where $\text{tr}(\cdot)$ denotes trace operator. The pseudocode of the proposed RLS-CIM algorithm is summarized in **Algorithm 1**.

Algorithm 1 The proposed RLS-CIM algorithm

Input:

Received signal: d_n;
Training signal: $\mathbf{x_n}=[x_n,x_{n-1},\dots,x_{n-N+1}]^T$;
Kernel width: $\sigma_{\max}$ and $\sigma_{\min}$;
Initial forgetting factor: λ;
Initial autocorrelation matrix: $\mathbf{P}_0=\delta^{-1}\mathbf{I}_N$;
Initial system vector: $\mathbf{w}_0=\mathbf{0}$;
Threshold parameter: $\rho>0$;

Output:

Estimated system vector: $\mathbf{w}_n$.

1: **for** $n\coloneqq 1,2,\dots N_{Iter}$ **do**
2: $e_n=d_n-\mathbf{w}_n^T\mathbf{x}_n$.
3: $\xi_n=\sum_{m=0}^{n}\lambda^{n-m}e_m^2$.
4: $\mathbf{P}_n=(1/\lambda)(\mathbf{P}_{n-1}-\mathbf{k}_n\mathbf{x}_n^T\mathbf{P}_{n-1})$.
5: $\mathbf{k}_n=\mathbf{P}_{n-1}\mathbf{x}_n/(\lambda+\mathbf{x}_n^T\mathbf{P}_{n-1}\mathbf{x}_n)$.
6: $\sigma_n=\sigma_{max}\cdot\exp(-20n/N_{Iter})+\sigma_{\min}$.
7: $\kappa_0=1/(\sqrt{2\pi}\sigma_n)$.
8: $\nabla_{\mathbf{w}}S(\boldsymbol{w}_n)=\kappa_0\mathbf{w}_n\exp(-\mathbf{w}_n^2/2\sigma_n^2)/(N\sigma_n^2)$.
8: $\gamma_n'=2\frac{(\text{tr}(\mathbf{P}_n)/N)(S(\mathbf{w}_n)-\rho)+\nabla_{\mathbf{w}}S^T(\mathbf{w}_n)\mathbf{P}_n\varepsilon_n}{\|\mathbf{P}_n\nabla_{\mathbf{w}}S^T(\mathbf{w}_n)\|_2^2}$.
9: $\gamma_n\in[0,\max(\gamma_n',0)]$.
8: $\mathbf{w}_n=\mathbf{w}_{n-1}+\xi_n\mathbf{k}_n-\gamma_{n-1}(1-\lambda)\mathbf{P}_n\nabla_{\mathbf{w}}S(\boldsymbol{w}_{n-1})$.
10: **end for**

IV. SIMULATION STUDIES

In this section, the proposed RLS-CIM algorithm is evaluated in different scenarios, i.e., SNR, system sparsity K as well as forgetting factor λ. For achieving average performance, 1000 independent Monte-Carlo runs are adopted in all of simulation experiments. The maximum length of FIR system vector $\mathbf{w}$ is $N=128$ and its number of nonzero coefficients is set as $K \in \{4,16,128\}$. Each nonzero tap is set as satisfying random Gaussian distribution $\mathcal{CN}(0,1)$ and all of the nonzero taps are randomly allocated in $\mathbf{w}$. To evaluate the performance of the proposed RLS-CIM algorithm, we compare the average MSD metric, which is defined as

$$\mathrm{MSE}\{\mathbf{w}_n\} \triangleq 10\log_{10} E\{\mathbf{w}_n - \mathbf{w}\}, \tag{24}$$

where $\mathbf{w}$ and $\mathbf{w}_n$ are the actual system vector and identification vector, respectively. The received SNR is defined as P_0/σ_n^2, where P_0 is the power of system output. Detailed parameters for computer simulations are set in in Tab. I.

TAB. I. SIMULATION PARAMETERS.

Parameters	Values
Input signal	Pseudo-random binary sequence
Additive noise distribution	$\mathcal{CN}(0,1)$
Length of FIR system	$N = 128$
No. of nonzero taps	$K \in \{4,16,128\}$
Nonzero coefficient	Random Gaussian $\mathcal{CN}(0,1)$
Received SNR	SNR $\in \{10\text{dB}, 20\text{dB}\}$
Forgetting factor	$\lambda \in \{0.991, 0.995, 0.999\}$
Kernel width of RLS-CIM	$\sigma_{\max} = 200$ and $\sigma_{\min} = 0.005$
Threshold parameter of RLS-L0	$\beta = 50$

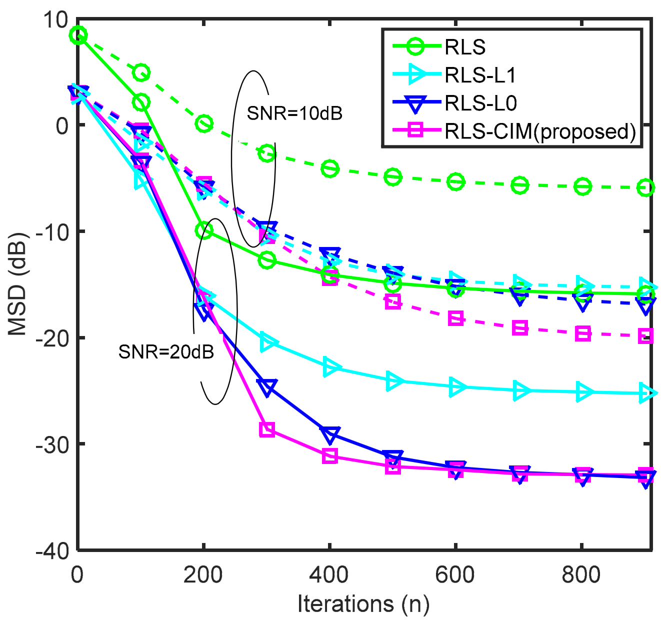

Fig. 1. MSD comparisons in $K = 4$ and $\lambda = 0.995$.

In the first experiment, MSD performance of the proposed RLS-CIM algorithm is evaluated against different number of nonzero taps in the case of $\lambda = 0.995$. Three state-of-the-art algorithms, i.e., RLS, RLS-L1, and RLS-L0, are considered as comparison benchmark. In Fig. 1, MSD curves of these algorithms are depicted in $K = 4$ and SNR $\in \{10\text{dB}, 20\text{dB}\}$. Under the two SNR regimes, one can find that the proposed RLS-CIM algorithm can achieve better MSD performance than existing algorithms. It is worth noting that the RLS-L0 algorithm can get almost same steady-state MSD performance as the proposed RLS-CIM algorithm in SNR = 20dB, but the convergence speed of the former is slower than latter. To verify the stability of the proposed algorithm, we evaluate the proposed RLS-CIM algorithm in Fig. 2 ($K = 16$) and Fig. 3 ($K = 128$). We can observe that the proposed algorithm can still better performance gain than existing algorithms even if the system does not satisfying sparse structure as shown in Fig. 3. Hence, the validation of the proposed RLS-CIM is verified in different sparse systems scenarios.

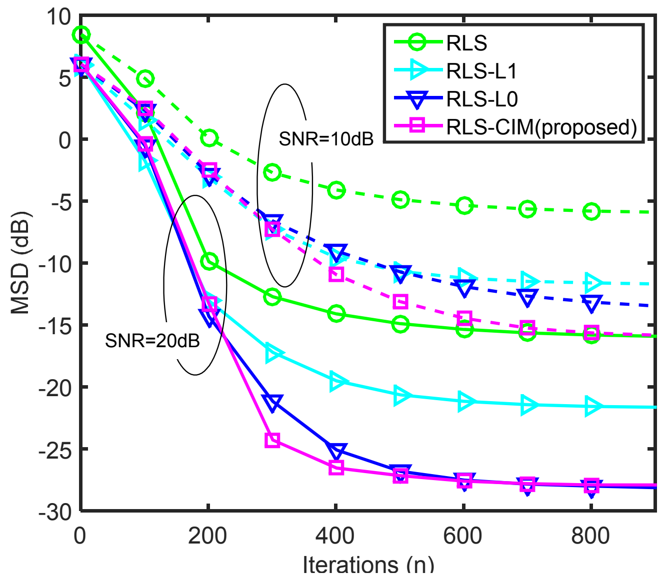

Fig. 2. MSD comparisons in $K = 16$ and $\lambda = 0.995$.

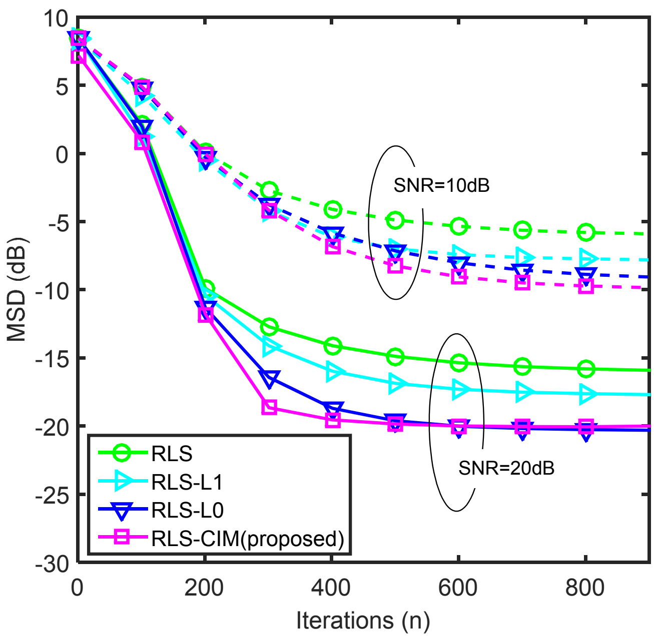

Fig. 3. MSD comparisons in $K = 128$ and $\lambda = 0.995$.

In the second experiment, MSD performance of the proposed RLS-CIM algorithm is evaluated against forgetting factor λ as shown in Fig. 4. It is well known the stability of the RLS-type algorithms is decided by forgetting factor λ. In this experiment, forgetting factor is selected in range [0.991,0.999] and RLS algorithm is considered as benchmark. One can find that stability of the proposed RLS-CIM algorithm is highly depends on the forgetting factor, smaller forgetting factor can bring it potential lower MSD performance but simultaneously suffers instability problem, while larger forgetting factor can ensure algorithm stability but sacrificing MSD performance. Hence, suitable forgetting factor is necessary selected for the proposed RLS-CIM algorithm. Fig. 4 shows that the forgetting factor $\lambda = 0.995$ is proper for the proposed CIM-RLS algorithm.

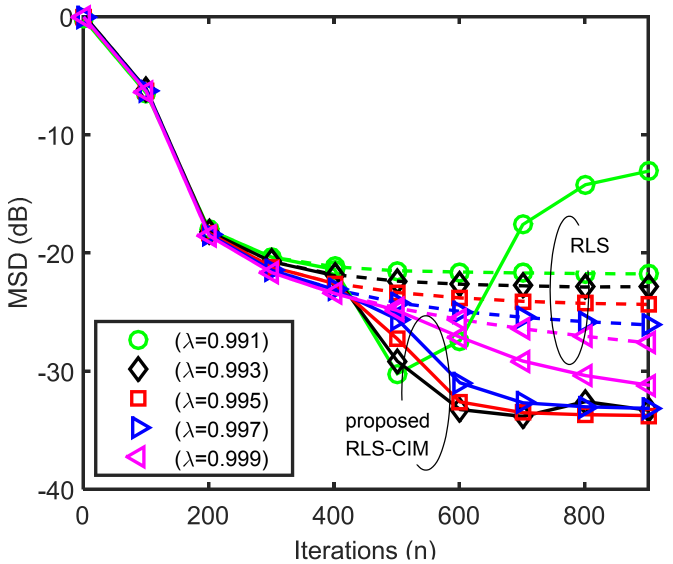

Fig. 4. MSD comparisons against forgetting factor λ in $K = 4$ and SNR = 20dB, where the dashed line and solid line denote RLS and the proposed RLS-CIM algorithm, respectively.

V. Conclusions

In this paper, we have proposed an improved RLS-CIM algorithm by using more efficient sparse constraint function, i.e., CIM. The proposed algorithm can exploit system sparsity more efficiently even if in the non-sparse system identification scenarios. Moreover, adaptive selection of forgetting factor has also been investigate to trade off the convergence speed and the stability. Representative simulation results have been shown to confirm the effectiveness of the proposed RLS-CIM against sparsity as well as forgetting factors, where existing sparse RLS algorithms are considered as the performance benchmarks.

Acknowledgment

This work was supported by National Key Basic Research Program of China (Grant No. 2013CB329203), National Natural Science Foundation of China grants (No. 61401069, No. 61271240, 61411130156) as well as Japan Society for the Promotion of Science (JSPS) research grants (No. 15K06072).

References

[1] G. Gui, L. Xu, and S. Matsushita, “Improved adaptive sparse channel estimation using mixed square/fourth error criterion,” *J. Franklin Inst.*, vol. 352, no. 10, pp. 4579–4594, 2015.

[2] W. Ma, H. Qu, G. Gui, L. Xu, J. Zhao, and B. Chen, “Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments,” *J. Franklin Inst.*, vol. 352, no. 7, pp. 2708–2727, 2015.

[3] G. Gui, L. Dai, S. Kumagai, and F. Adachi, “Variable earns profit: Improved adaptive channel estimation using sparse VSS-NLMS algorithms,” *IEEE International Conference on Communications (ICC),* Sydney, Australia, 10-14 June, 2014, pp. 1–5.

[4] Y. Tsao, S. Fang, S. Member, and Y. Shiao, “Acoustic echo cancellation using a vector-space-based adaptive filtering algorithm,” *IEEE Signal Process. Lett.*, vol. 22, no. 3, pp. 351–355, 2015.

[5] A. H. Sayed, *Adaptive filters, John Wiley & Sons,* 2008.

[6] B. Babadi, N. Kalouptsidis, and V. Tarokh, “SPARLS: The sparse RLS algorithm,” *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4013–4025, 2010.

[7] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, “Online adaptive estimation of sparse signals: Where RLS meets the l1-norm,” *IEEE Trans. Signal Process.*, vol. 58, no. 7, p. 3436–3447, 2010.

[8] E. M. Eksioglu and a. K. Tanc, “RLS algorithm with convex regularization,” *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 470–473, 2011.

[9] E. J. Candès, J. Romberg, and T. Tao, “Robust Uncertainty Principles : Exact Signal Frequency Information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[10] D. L. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[11] S. Seth and J. C. Príncipe, “Compressed signal reconstruction using the correntropy induced metric,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, March 30 – April 4, 2008, pp. 3845–3848.

[12] W. Liu, P. Pokharel, and J. Príncipe, “Correntropy: properties and applications in non-Gaussian signal processing,” *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, 2007.