IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 20, 2019, accepted September 15, 2019, date of publication September 18, 2019, date of current version October 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2942088

Transient Performance Analysis of Zero-Attracting Gaussian Kernel LMS Algorithm With Pre-Tuned Dictionary

WEI GAO[1], (Member, IEEE), AND JIE CHEN[2], (Senior Member, IEEE)
[1]School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China
[2]Center of Intelligent Acoustics and Immersive Communications (CIAIC), School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Wei Gao (wei_gao@ujs.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701200 and Grant 61671382, and in part by the Research Foundation for Talented Scholars of Jiangsu University under Grant 17JDG015.

ABSTRACT Although the sparse kernel adaptive filtering algorithms have been proposed to address the problem of redundant dictionary in non-stationary environments, there is few attempt of analyzing their stochastic convergence behaviors. In this paper, we briefly review the zero-attracting kernel least-mean-square (ZA-KLMS) algorithm with ℓ_1-norm regularization from the perspective of nonlinear sparse system. Then, the theoretical transient convergence performance of ZA-KLMS algorithm using Gaussian kernel function with pre-tuned dictionary is analyzed in the mean and mean-square senses. The simulation results illustrate the accuracy of derived analytical models by the excellent consistency between the Monte Carlo simulations and the theoretical predictions, and the ZA-KLMS algorithm has better convergence performance than the KLMS algorithm for nonlinear sparse systems in stationary environment.

INDEX TERMS Nonlinear sparse system identification, zero-attracting, kernel least-mean-square, transient performance analysis.

I. INTRODUCTION

There are considerable research interests in the sparse system identification based on the sparse-aware adaptive filters [1], [2], which has been applied in a wide range of fields such as image processing [3], [4], communications [5], [6], acoustic signal processing [7], etc. The common objective of diverse scenarios is to find a satisfactory sparse solution that involves only a small number of nonzero significant coefficients to enhance the performance and reduce the computational complexity of algorithm. Conventional linear zero-attracting least-mean-square (ZA-LMS) with ℓ_1-norm regularization was proposed for the linear sparse systems in [8], and then due to substantial research interests its stochastic convergence behaviors were extensively analyzed in [9]–[11]. In fact the nonlinear sparse systems are frequently encountered in many practical applications ranging from satellite channel estimation [12], biomedical engineering [13], to adaptive echo cancellation [14], [15]. Since the schemes of several typical linear adaptive filters were reformulated in the high or infinite dimensional reproducing kernel Hilbert spaces (RKHS) [16]–[18], hence, the sparse-aware kernel adaptive filtering (KAF) algorithm has become a powerful tool of solving the nonlinear sparse system identification problem [19].

The surprise criterion was proposed to design the sparse kernel adaptive filters taking into account the sparsity of online dictionary [20]. A recurrent kernel algorithm was developed to accelerate the convergence speed by introducing sparse updates when the estimated errors do not fall into the given interval [21]. The sparse quantized kernel least-mean-square (KLMS) with ℓ_1-norm regularization and KLMS with forward-backward splitting (FOBOS-KLMS) algorithms were independently proposed to attempt to improve the performance in terms of convergence speed and steady-state mean-square error (MSE) by automatically eliminating the obsoleted elements in the dynamic dictionary within context of time-varying environments [22]–[24]. Meanwhile, in order to alleviate the adverse impact of obsolete dictionary-elements, the multikernel LMS (MKLMS) algorithm with weighted block ℓ_1-norm penalty was proposed to

The associate editor coordinating the review of this manuscript and approving it for publication was Roberto Gomez-Garcia.

dramatically reduce computational complexity and memory storage requirements [25]–[27]. The kernel online sequential extreme learning machine (KOS-ELM) with the approximate linear dependency and the fixed-budget criteria was proposed to obtain both sparse filters [28]. Recently, a framework based on an eigenvalue analysis was proposed to study the sparsity measures and sparsification criteria of KAF algorithms [29].

The KLMS-tpye algorithm as a representative of kernel adaptive filters has been extensively studied due to its simplicity and robustness. The zero-attracting kernel least-mean-square (ZA-KLMS) algorithm can be regarded as a particular case of FOBOS-KLMS algorithm for nonlinear sparse system in stationary environment. Before proceeding, it should be emphasized that the sparse KLMS-type algorithms have twofold aspects: the negligible weight coefficients and the obsolete elements of dictionary. When the online dictionary is assumed to be statistically stable in the stationary environment, the zero weight coefficients of KLMS-type algorithms indicate the sparse characteristic of the nonlinear model. On the contrary, in the non-stationary environment the obsolete elements in online dictionary will also lead to the trivial weight coefficients even for nonlinear systems without sparse feature. Since the stochastic behavior of FOBOS-KLMS algorithm with sparsity-promoting of recursive weight vector was investigated in the latter case, we thus only consider the former case in the following of analysis. Moreover, there is few theoretical study of concerning on the transient analysis for nonlinear ZA-KLMS algorithm relative to the various theoretical analyzes of conventional linear ZA-LMS.

In this paper, we briefly review the ZA-KLMS algorithm from the viewpoint of the nonlinear sparse system in stationary environment. Then, the transient stochastic behaviors of the ZA-KLMS algorithm using Gaussian kernel function with pre-tuned dictionary are derived in the mean and mean-square senses under the necessary assumptions and reasonable approximations. Finally, we illustrate the accuracy of derived analytical models of ZA-KLMS algorithm and its better convergence performance with the simulation results.

The rest of this paper is organized as follows. Section II presents some basic knowledge of the KAF and briefly reviews the ZA-KLMS algorithm. In Section III, we introduce the preliminaries and some useful statistical assumptions in the following theoretical analysis. The transient mean and mean-square weight behaviors of the ZA-KLMS algorithm using Gaussian kernel with fixed dictionary are studied in Section IV. In Section V, simulations are performed to demonstrate the correctness and the accuracy of the derived analytical models, as well as the better performance of ZA-KLMS compared to KLMS for the nonlinear sparse system. Finally, the paper is concluded in Section VI.

Notation: Normal font letters x denote scalars, boldface small letters $\boldsymbol{x}$ denote column vectors, and boldface capital letters $\boldsymbol{X}$ denote matrices. $[\boldsymbol{x}]_i$ and $[\boldsymbol{X}]_{ij}$ denote the i-th entry of $\boldsymbol{x}$ and the (i, j)-th entry of $\boldsymbol{X}$, respectively. Identity matrix of size $M \times M$ is denoted by $\boldsymbol{I}_M$, and all-zero vector of length N is denoted by $\boldsymbol{0}_N$. The superscript $(\cdot)^\top$ represents the transpose of a matrix and a vector, and $\text{tr}\{\cdot\}$ denotes the trace of its matrix argument. The operator $\text{vec}\{\cdot\}$ stacks a matrix of column vectors on top of each other to generate a connected vector. The notations $\otimes$ and $\odot$ denote the Kronecker product and the Hadamard product, respectively. The operator $\text{sgn}(\cdot)$ is the sign function. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. In the multivariate case, the corresponding notation will become $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The cumulative distribution function (CDF) of the standard Gaussian distribution is denoted by $\phi(x)$. The CDF of the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is denoted by $\Phi(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

II. ZERO-ATTRACTING KERNEL LMS ALGORITHM

Let $\mathcal{U}$ be a compact domain in Euclidean space $\mathbb{R}^L$ and $\mathcal{D} = \mathbb{R}$. Let $\rho_\mathcal{Y}$ be a Borel probability measure on $\mathcal{Y} = \mathcal{U} \times \mathcal{D}$ whose regularity properties will be assumed as needed. $\mathcal{H}$ is defined as a reproducing kernel Hilbert space with kernel $\kappa : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$. Given the input-output sequence $y \in \mathcal{Y}^N$, $y = \{(\boldsymbol{u}_n, d_n)\}_{n=1}^N$, we aim at estimating a optimum regression function $\psi^\star$ that minimizes the regularized least-square error

$$\min_{\psi \in \mathcal{H}} \sum_{n=1}^{N} |d_n - \psi(\boldsymbol{u}_n)|^2 + \gamma \, \|\psi\|_\mathcal{H}^2 \tag{1}$$

with $\gamma \geq 0$ a regularization constant. By virtue of the representer theorem [30], the function $\psi(\cdot)$ can be written as

$$\psi^\star(\cdot) = \sum_{n=1}^{N} w_n \, \kappa(\cdot, \boldsymbol{u}_n). \tag{2}$$

To overcome the problem of linearly increasing amount N of input candidates as new data is collected, the fixed-order model with a promoting sparsity criterion is often adopted

$$\psi(\cdot) = \sum_{m=1}^{M} w_m \, \kappa(\cdot, \boldsymbol{u}_m). \tag{3}$$

The data set $\boldsymbol{\omega} = \{\boldsymbol{u}_{\omega_m}\}_{m=1}^M$ is the so-called dictionary. Based on the parametric vector model (3), we consider the minimization problem with ℓ_1-norm regularization

$$\boldsymbol{w}^\star = \arg \min_{\boldsymbol{w} \in \mathbb{R}^N} \left\{ \|\boldsymbol{d} - \boldsymbol{K}\boldsymbol{w}\|^2 + \lambda \|\boldsymbol{w}\|_1 \right\} \tag{4}$$

where $\boldsymbol{K}$ is the $(M \times M)$ Gram matrix with the (i, j)-th entry $\kappa(\boldsymbol{u}_i, \boldsymbol{u}_j)$, $\boldsymbol{d} = [d_1, \ldots, d_M]^\top$ is the desired output vector, and λ is a positive regularization constant compromising between the convergence speed and the estimation error. Without manipulating the dictionary separately as in [24], and using subgradient of (4), the recursive update equation of ZA-KLMS algorithm is given by

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \eta \, e_n \boldsymbol{\kappa}_{\omega,n} - \rho \, \text{sgn}\{\boldsymbol{w}_n\} \tag{5}$$

with the positive step-size η, the kernelized input vector $\boldsymbol{\kappa}_{\omega,n} = \left[\kappa(\boldsymbol{x}_n, \boldsymbol{x}_{\omega,1}), \ldots, \kappa(\boldsymbol{x}_n, \boldsymbol{x}_{\omega,M})\right]^\top$, and the shrinkage parameter $\rho = \lambda\eta$. Here, the instantaneous estimation error e_n is given by

$$e_n = d_n - \boldsymbol{w}_n^\top \boldsymbol{\kappa}_{\omega,n}. \tag{6}$$

Note that although the zero-attracting term $\rho\,\mathrm{sgn}\{\boldsymbol{w}_n\}$ can effectively promoting zero-valued of weight vector, this simultaneously resulting in the bias for large entries.

The FOBOS-KLMS algorithm was proposed to reduce the convergence performance degradation cased by the redundant dictionary in non-stationary environments, which is implicitly to achieve the sparsity of nonlinear system by discarding the obsolete elements out of online dictionary. It should be pointed out that the ZA-KLMS algorithm removing the redundant elements corresponding with zero weight coefficients from the fixed dimensional dictionary at each time instant, is equivalent to the so-called FOBOS-KLMS algorithm with the adaptively compressing online dictionary.

Consider the mean-square error criterion, which is given by

$$E\{e_n^2\} = \int_{\Omega}\int_{\mathcal{U}\times\mathcal{D}} e_n^2\, d\varphi_{\mathcal{Y}}(\boldsymbol{u}, \boldsymbol{d}|\boldsymbol{\omega})\, d\varphi_{\Omega} \tag{7}$$

with φ_{Ω} a Borel probability measure on the dictionary space Ω. Except of the simplified assumptions applied in [31], we consider the dictionary as a part of the filter parameters to be set. The objective is to characterize the transient of the mean-square criterion condition on dictionary $\boldsymbol{\omega}$, that is

$$E_{\mathcal{Y}}\{e_n^2|\boldsymbol{\omega}\} = \int_{\mathcal{U}\times\mathcal{D}} e_n^2\, d\varphi(\boldsymbol{u}, \boldsymbol{d}|\boldsymbol{\omega}). \tag{8}$$

Note that we shall use the subscript ω for quantities conditioned on the dictionary $\boldsymbol{\omega}$, and $\mathcal{Y}$ for expectation with respect to input data distribution in the following.

III. PRELIMINARIES AND STATISTICAL ASSUMPTIONS

Studying the statistical behavior analysis of ZA-KLMS algorithm is a challenging task, which requires some preliminaries and necessary statistical assumptions for the mathematical derivations.

A. PRELIMINARIES

Given the fixed dictionary $\boldsymbol{\omega}$ with length M, the estimation error at instant n is given by

$$\begin{aligned} e_{\omega,n} &= d_n - \phi_{\omega}(\boldsymbol{u}_n) \\ &= d_n - \boldsymbol{w}_n^{\top}\boldsymbol{\kappa}_{\omega,n} \end{aligned} \tag{9}$$

with $\phi_{\omega}(\boldsymbol{u}_n) = \phi(\boldsymbol{u}_n)|\boldsymbol{\omega}$. Note that the theoretical convergence performance of ZA-KLMS depends on the dictionary setting which can be regarded as a part of the filter parameters to be set. Squaring both sides of (9), and taking the expected value, leads to the MSE criterion

$$\begin{aligned} J_{\mathrm{MSE},\omega} &= E\{e_{\omega,n}^2\} \\ &= E\{d_n^2\} - 2\boldsymbol{p}_{\kappa d,\omega}\boldsymbol{w}_n + \boldsymbol{w}_n^{\top}\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}\boldsymbol{w}_n \end{aligned} \tag{10}$$

where

$$\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega} = E\{\boldsymbol{\kappa}_{\omega,n}\boldsymbol{\kappa}_{\omega,n}^{\top}|\boldsymbol{\omega}\} \tag{11}$$

is the correlation matrix of the kernelized input, and

$$\boldsymbol{p}_{\kappa d,\omega} = E\{d_n\,\boldsymbol{\kappa}_{\omega,n}|\boldsymbol{\omega}\} \tag{12}$$

is the cross-correlation vector between d_n and $\boldsymbol{\kappa}_{\omega,n}$.

In next section, we will derive the mean and mean-square weight behaviors of ZA-KLMS algorithm with Gaussian kernel defined as

$$\kappa(\boldsymbol{u}, \boldsymbol{u}') = \exp\Big(-\frac{\|\boldsymbol{u}-\boldsymbol{u}'\|^2}{2\xi^2}\Big) \tag{13}$$

where $\xi > 0$ is the kernel bandwidth. Inputs $\boldsymbol{u}_n$ are assumed to be statistically independent zero-mean Gaussian random vectors with the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{uu}} = E\{\boldsymbol{u}_n\boldsymbol{u}_n^{\top}\}$.

Let $\boldsymbol{v}_{\omega,n}$ be the weight error vector as the difference between the estimated weight vector $\boldsymbol{w}_n$ and the optimum weight vector $\boldsymbol{w}^{\star}$, namely

$$\boldsymbol{v}_{\omega,n} = \boldsymbol{w}_n - \boldsymbol{w}^{\star}. \tag{14}$$

The stochastic convergence analysis of ZA-KLMS is to study the evolution of the first- and second-order moments of $\boldsymbol{v}_{\omega,n}$ over time.

Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_L]^{\top} \in \mathbb{R}^L$ be a random vector following Gaussian distribution with the zero-mean vector

$$E\{\boldsymbol{x}\} = \boldsymbol{0}_L \tag{15}$$

and the covariance matrix defined by

$$\boldsymbol{R}_{\boldsymbol{xx}} = E\{\boldsymbol{x}\boldsymbol{x}^{\top}\}. \tag{16}$$

The random variable y is defined as the quadratic form of

$$y = \boldsymbol{x}^{\top}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}^{\top}\boldsymbol{x}. \tag{17}$$

The moment generating function of y is given by [32]

$$\begin{aligned} \Psi_y(s) =& |\boldsymbol{I} - 2s\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{xx}}|^{-\frac{1}{2}} \\ &\times \exp\Big(\frac{s^2}{2}\boldsymbol{b}^{\top}\boldsymbol{R}_{\boldsymbol{xx}}\big(\boldsymbol{I} - 2s\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{xx}}\big)^{-1}\boldsymbol{b}\Big). \end{aligned} \tag{18}$$

The above result is very useful in the following derivation of theoretical transient performance analysis.

B. STATISTICAL ASSUMPTIONS

Three simplified assumptions are required to make the derivation of stochastic behavior of $\boldsymbol{v}_{\omega,n}$ mathematically tractable. The statistical assumptions in the analysis are listed as follows:

(A.1) The weight error vector $\boldsymbol{v}_{\omega,n}$ is statistically independent of the kernelized input vector $\boldsymbol{\kappa}_{\omega,n}$.

(A.2) $\boldsymbol{\kappa}_{\omega,n}\boldsymbol{\kappa}_{\omega,n}^{\top}$ is independent of $\boldsymbol{v}_{\omega,n}$.

(A.3) Any pair of entries $[\boldsymbol{v}_{\omega,n}]_i$ and $[\boldsymbol{v}_{\omega,n}]_j$ with $i \neq j$ is jointly Gaussian.

Assumption (A.1), called conditioned independence assumption (CIA), is originated from the well known independence assumption (IA) widely used in the analysis of adaptive filters [33], [34]. Assumption (A.2), called modified independence assumption (MIA), is justified in detail in [35], which has been shown to be less restrictive than the well known independence assumption IA (A.1). For further reference we named MIA as conditioned MIA (CMIA) which has been successfully employed in Gaussian kernel LMS analyses [36], [37]. Assumption (A.3) is consistent with the

Gaussian assumptions in [8], [10], [36]. The effectiveness of (A.3) has been verified by the histograms of bivariate vector $\left[[\boldsymbol{v}_{\omega,n}]_i, [\boldsymbol{v}_{\omega,n}]_j\right]^\top$ in [11]. Accordingly, a more accurate model can be provided by making the calculation of the nonlinear terms tractable without further approximations.

IV. TRANSIENT CONVERGENCE ANALYSIS OF ZA-KLMS ALGORITHM

We shall now analyze the transient stochastic behaviors of ZA-KLMS algorithm consisting of the weight error vector in both the mean and the mean-square error senses.

A. MEAN WEIGHT BEHAVIOR MODEL

The instantaneous estimation error in (9) can be expressed using the definition of $\boldsymbol{v}_{\omega,n}$ as

$$e_{\omega,n} = z_n - \boldsymbol{v}_{\omega,n}^\top \boldsymbol{\kappa}_{\omega,n} \tag{19}$$

where z_n is the stationary, independent and identically distributed (i.i.d.) additive Gaussian noise with zero-mean and variance σ_z^2.

Subtracting $\boldsymbol{w}^\star$ from both sides of (5), and using (14) and (19), yields the recursive update equation of $\boldsymbol{v}_{\omega,n}$

$$\begin{aligned}\boldsymbol{v}_{\omega,n+1} &= \boldsymbol{v}_{\omega,n} - \eta\boldsymbol{\kappa}_{\omega,n}\boldsymbol{\kappa}_{\omega,n}^\top\boldsymbol{v}_{\omega,n} + \eta z_n\boldsymbol{\kappa}_{\omega,n}\\ &\quad -\rho\,\mathrm{sgn}\left\{\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\right\}.\end{aligned} \tag{20}$$

Taking the expected values of both sides of (20) and using assumption (A.2) CMIA, we have

$$\begin{aligned}E\left\{\boldsymbol{v}_{\omega,n+1}\right\} &= (\boldsymbol{I} - \eta\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega})E\left\{\boldsymbol{v}_{\omega,n}\right\}\\ &\quad -\rho\, E\left\{\mathrm{sgn}\{\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\}\right\}.\end{aligned} \tag{21}$$

The last term on the right hand side (RHS) of (21) introduced by the zero-attracting term is the remarkable difference from the mean weight behavior model of KLMS [36]. With the Gaussian kernel function, the (i,j)-th entry of matrix $\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}$ is given by

$$\begin{aligned}&[\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}]_{ij}\\ &= E_U\left\{\exp\left(-\frac{1}{2\xi^2}\left[\|\boldsymbol{u}_n - \boldsymbol{u}_{\omega,i}\|^2 + \|\boldsymbol{u}_n - \boldsymbol{u}_{\omega,j}\|^2\right]\right)\right\}\\ &= \exp\left(-\frac{1}{2\xi^2}\left[\|\boldsymbol{u}_{\omega,i}\|^2 + \|\boldsymbol{u}_{\omega,j}\|^2\right]\right)\\ &\quad\times E_U\left\{\exp\left(-\frac{1}{\xi^2}\left[\|\boldsymbol{u}_n\|^2 - (\boldsymbol{u}_{\omega,i} + \boldsymbol{u}_{\omega,j})^\top\boldsymbol{u}_n\right]\right)\right\}.\end{aligned} \tag{22}$$

Comparing the second term of the RHS of (22) with (17), using the substitutions $\boldsymbol{H} = \boldsymbol{I}$, $\boldsymbol{b} = -(\boldsymbol{u}_{\omega,i} + \boldsymbol{u}_{\omega,j})$ and $s = -\frac{1}{\xi^2}$, then we have (23), as shown at the top of the next page. By using the identity property $(\boldsymbol{I} + \boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}(\boldsymbol{A} + \boldsymbol{I})^{-1}$ for the component $\boldsymbol{R}_{\boldsymbol{uu}}\left(\boldsymbol{I} + \frac{2}{\xi^2}\boldsymbol{R}_{\boldsymbol{uu}}\right)^{-1}$, yields (24), as shown at the top of the next page. Note that $\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}$ is the symmetric matrix. In order to characterize the evolution of $E\left\{\boldsymbol{v}_{\omega,n+1}\right\}$, it is necessary to evaluate the last term of the RHS of (21) according to the following lemma .

Lemma 1. Consider a random variable $x \sim \mathcal{N}(\mu, \sigma^2)$. The expectation of its sign value is given by [11]

$$E\left\{\mathrm{sgn}\{x\}\right\} = 1 - 2\phi(-\mu/\sigma). \tag{25}$$

Consequently, the entries of $E\left\{\mathrm{sgn}\{\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\}\right\}$ can be obtained by making the following identification:

$$x \triangleq [\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}]_i \tag{26}$$

with

$$[\boldsymbol{w}^\star]_i + E\left\{[\boldsymbol{v}_{\omega,n}]_i\right\} \rightarrow \mu \tag{27}$$

$$E\left\{[\boldsymbol{v}_{\omega,n}]_i^2\right\} - E\left\{[\boldsymbol{v}_{\omega,n}]_i\right\}^2 \rightarrow \sigma \tag{28}$$

where $E\left\{[\boldsymbol{v}_{\omega,n}]_i^2\right\}$ can be extracted from the diagonal of the correlation matrix of the weight error, i.e., $\boldsymbol{C}_{\omega,n} = E\left\{\boldsymbol{v}_{\omega,n}\boldsymbol{v}_{\omega,n}^\top\right\}$ that will be calculated in the next subsection. Hence, the last term of the RHS of (21) can be calculated based on (25) and the above relations. Therefore, the analytical model (21) with (24)–(28) can be used to characterize the first-order mean weight behavior of the ZA-KLMS algorithm using Gaussian kernel with pre-tuned dictionary.

B. MEAN-SQUARE ERROR BEHAVIOR MODEL

The objective of this subsection is to derive the analytical model of the transient MSE for ZA-KLMS. Using (19) and the assumption (A.1), the MSE can be expressed in terms of the second-order moments of the weight error vector [34]

$$\begin{aligned}J_{\mathrm{MSE},\omega} &\approx \sigma_z^2 + \mathrm{tr}\left\{\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}\,\boldsymbol{C}_{\omega,n}\right\}\\ &= \sigma_z^2 + J_{\mathrm{EMSE},\omega}\end{aligned} \tag{29}$$

where $J_{\mathrm{EMSE},\omega}$ is the excess mean-square error (EMSE). Hence, the evaluation of the MSE (or EMSE) requires the calculation model of the recursive update equation for $\boldsymbol{C}_{\omega,n}$.

Post-multiplying (20) by its transpose, taking the expected value, and using assumption (A.1), leads to

$$\begin{aligned}\boldsymbol{C}_{\omega,n+1} &\approx \boldsymbol{C}_{\omega,n} - \eta\left(\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}\boldsymbol{C}_{\omega,n} + \boldsymbol{C}_{\omega,n}\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}\right)\\ &\quad+\eta^2\sigma_z^2\,\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega} + \eta^2\,\boldsymbol{Q}_{\omega,1} + \rho^2\,\boldsymbol{Q}_{\omega,2}\\ &\quad-\rho\left(\boldsymbol{Q}_{\omega,3} + \boldsymbol{Q}_{\omega,3}^\top\right) + \eta\rho\left(\boldsymbol{Q}_{\omega,4} + \boldsymbol{Q}_{\omega,4}^\top\right)\end{aligned} \tag{30}$$

where

$$\boldsymbol{Q}_{\omega,1} = E\left\{\boldsymbol{\kappa}_{\omega,n}\boldsymbol{\kappa}_{\omega,n}^\top\boldsymbol{v}_{\omega,n}\boldsymbol{v}_{\omega,n}^\top\boldsymbol{\kappa}_{\omega,n}\boldsymbol{\kappa}_{\omega,n}^\top\right\} \tag{31}$$

$$\boldsymbol{Q}_{\omega,2} = E\left\{\mathrm{sgn}\{\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\}\mathrm{sgn}^\top\{\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\}\right\} \tag{32}$$

$$\boldsymbol{Q}_{\omega,3} = E\left\{\boldsymbol{v}_{\omega,n}\,\mathrm{sgn}^\top\{\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\}\right\} \tag{33}$$

$$\boldsymbol{Q}_{\omega,4} = E\left\{\boldsymbol{\kappa}_{\omega,n}\boldsymbol{\kappa}_{\omega,n}^\top\boldsymbol{v}_{\omega,n}\,\mathrm{sgn}^\top\{\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\}\right\}. \tag{34}$$

Note that the three matrices defined in (32)–(34) are introduced by the zero-attracting term. We shall sequentially calculate the four terms $\boldsymbol{Q}_{\omega,1}$ to $\boldsymbol{Q}_{\omega,4}$. To determine the expected value of (31), assuming the assumption (A.2) holds, the (i,j)-th entry of (31) can be approximated as:

$$\begin{aligned}[\boldsymbol{Q}_{\omega,1}]_{ij} &\approx \sum_{p=1}^{M}\sum_{q=1}^{M} E_U\left\{\kappa_{\omega,i}(n)\kappa_{\omega,j}(n)\kappa_{\omega,p}(n)\kappa_{\omega,q}(n)\right\}\\ &\qquad\times[\boldsymbol{C}_{\omega,n}]_{pq}\end{aligned} \tag{35}$$

$$[\boldsymbol{R}_{\boldsymbol{\kappa}\boldsymbol{\kappa},\omega}]_{ij} = \exp\left(-\frac{1}{2\xi^2}\left[\|\boldsymbol{u}_{\omega,i}\|^2 + \|\boldsymbol{u}_{\omega,j}\|^2\right]\right) |\boldsymbol{I} + \frac{2}{\xi^2}\boldsymbol{R}_{\boldsymbol{uu}}|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2\xi^4}\big(\boldsymbol{u}_{\omega,i} + \boldsymbol{u}_{\omega,j}\big)^\top \boldsymbol{R}_{\boldsymbol{uu}}\big(\boldsymbol{I} + \frac{2}{\xi^2}\boldsymbol{R}_{\boldsymbol{uu}}\big)^{-1}\big(\boldsymbol{u}_{\omega,i} + \boldsymbol{u}_{\omega,j}\big)\right) \tag{23}$$

$$[\boldsymbol{R}_{\boldsymbol{\kappa}\boldsymbol{\kappa},\omega}]_{ij} = |\boldsymbol{I} + \frac{2}{\xi^2}\boldsymbol{R}_{\boldsymbol{uu}}|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{4\xi^2}\left[2\big(\|\boldsymbol{u}_{\omega,i}\|^2 + \|\boldsymbol{u}_{\omega,j}\|^2\big) - \|\boldsymbol{u}_{\omega,i} + \boldsymbol{u}_{\omega,j}\|^2_{(\boldsymbol{I}+\xi^2\boldsymbol{R}_{\boldsymbol{uu}}^{-1}/2)^{-1}}\right]\right) \tag{24}$$

---

with $\kappa_{\omega,i}(n) = \kappa(\boldsymbol{u}_n, \boldsymbol{u}_{\omega,i})$. In addition, let us define the matrix $\boldsymbol{T}^{(i,j)}_{\omega,n}$ with its (p,q)-th entry given by

$$\big[\boldsymbol{T}^{(i,j)}_{\omega,n}\big]_{pq} = E_U\Big\{\kappa_{\omega,i}(n)\kappa_{\omega,j}(n)\kappa_{\omega,p}(n)\kappa_{\omega,q}(n)\Big\}. \tag{36}$$

Based on (36), thus (35) can be rewritten as

$$[\boldsymbol{Q}_{\omega,1}]_{ij} \approx \operatorname{tr}\Big\{\boldsymbol{T}^{(i,j)}_{\omega,n}\boldsymbol{C}_{\omega,n}\Big\}. \tag{37}$$

In order to determine the value of $[\boldsymbol{Q}_{\omega,1}]_{ij}$ in (37), we need to evaluate the expectation of matrix $\boldsymbol{T}^{(i,j)}_{\omega,n}$. Then, (36) can be rewritten as

$$\begin{aligned}
&\big[\boldsymbol{T}^{(i,j)}_{\omega,n}\big]_{pq} \\
&= E_U\Big\{\kappa_{\omega,i}(n)\kappa_{\omega,j}(n)\kappa_{\omega,p}(n)\kappa_{\omega,q}(n)\Big\} \\
&= E_U\Big\{\exp\Big(-\frac{1}{2\xi^2}\sum_{k=\{i,j,p,q\}}\|\boldsymbol{u}_n - \boldsymbol{u}_{\omega,k}\|^2\Big)\Big\} \\
&= \exp\Big(-\frac{1}{2\xi^2}\sum_{k=\{i,j,p,q\}}\|\boldsymbol{u}_{\omega,k}\|^2\Big) \\
&\quad\times E_U\Big\{\exp\Big(-\frac{1}{\xi^2}\Big[2\|\boldsymbol{u}_n\|^2 - \Big(\sum_{k=\{i,j,p,q\}}\boldsymbol{u}_{\omega,k}\Big)^\top\boldsymbol{u}_n\Big]\Big)\Big\}.
\end{aligned} \tag{38}$$

Likewise, by setting $\boldsymbol{H} = 2\boldsymbol{I}$, $\boldsymbol{b} = -\sum_{k=\{i,j,p,q\}}\boldsymbol{u}_{\omega,k}$ and $s = -\frac{1}{\xi^2}$ in (17), we can obtain (39), as shown at the top of the next page.

The rest of analysis involving the calculation of the three terms $\boldsymbol{Q}_{\omega,2}$ to $\boldsymbol{Q}_{\omega,4}$ in (32)–(34) remains valid even if the kernelized input vector $\boldsymbol{\kappa}_{\omega,n}$ is not Gaussian distribution. To evaluate the term $\boldsymbol{Q}_{\omega,2}$, we introduce the following lemma.

Lemma 2. Consider two random variables x and y which are jointly Gaussian, namely

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu} := \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \boldsymbol{\Sigma}_{xy} := \begin{bmatrix} \sigma_x^2 & \rho_{xy} \\ \rho_{xy} & \sigma_y^2 \end{bmatrix}\right) \tag{40}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_{xy}$ are their mean vector and covariance matrix respectively, defined as the previous descriptions. The expectation $E\{\operatorname{sgn}\{x\}\operatorname{sgn}\{y\}\}$ can be given by [11]

$$\begin{aligned}
&E\big\{\operatorname{sgn}\{x\}\operatorname{sgn}\{y\}\big\} \\
&= \Phi\big(\boldsymbol{0}_2, [\mu_x, \mu_y]^\top, \boldsymbol{\Sigma}_{xy}\big) + \Phi\big(\boldsymbol{0}_2, -[\mu_x, \mu_y]^\top, \boldsymbol{\Sigma}_{xy}\big) \\
&\quad -\Phi\big(\boldsymbol{0}_2, [\mu_x, -\mu_y]^\top, \overline{\boldsymbol{\Sigma}}_{xy}\big) - \Phi\big(\boldsymbol{0}_2, [-\mu_x, \mu_y]^\top, \overline{\boldsymbol{\Sigma}}_{xy}\big)
\end{aligned} \tag{41}$$

with

$$\overline{\boldsymbol{\Sigma}}_{xy} = \boldsymbol{\Sigma}_{xy} \odot \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{42}$$

Note that the diagonal entries of $\boldsymbol{Q}_{\omega,2}$ are simply to be one, i.e., $[\boldsymbol{Q}_{\omega,2}]_{ii} = 1$. Then, the off-diagonal entries $[\boldsymbol{Q}_{\omega,2}]_{ij}$ with $i \neq j$ are obtained by making the following identifications:

$$x \triangleq \big[\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\big]_i \tag{43}$$

$$y \triangleq \big[\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}\big]_j \tag{44}$$

with

$$E\big\{[\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}]_i\big\} \to \mu_x \tag{45}$$

$$E\big\{[\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}]_j\big\} \to \mu_y \tag{46}$$

$$E\big\{[\boldsymbol{v}_{\omega,n}]_i^2\big\} - E\big\{[\boldsymbol{v}_{\omega,n}]_i\big\}^2 \to \sigma_x^2 \tag{47}$$

$$E\big\{[\boldsymbol{v}_{\omega,n}]_j^2\big\} - E\big\{[\boldsymbol{v}_{\omega,n}]_j\big\}^2 \to \sigma_y^2 \tag{48}$$

$$E\big\{[\boldsymbol{v}_{\omega,n}]_i[\boldsymbol{v}_{\omega,n}]_j\big\} - E\big\{[\boldsymbol{v}_{\omega,n}]_i\big\}E\big\{[\boldsymbol{v}_{\omega,n}]_j\big\} \to \rho_{xy} \tag{49}$$

where $E\big\{[\boldsymbol{v}_{\omega,n}]_i[\boldsymbol{v}_{\omega,n}]_j\big\}$ can be extracted from $[\boldsymbol{C}_{\omega,n}]_{ij}$. Consequently, the term $\boldsymbol{Q}_{\omega,2}$ can be computed using (41) and the corresponding relations.

In order to calculate the term $\boldsymbol{Q}_{\omega,3}$, we need to introduce the following lemma.

Lemma 3. Consider two random variables x and y, which are jointly Gaussian, namely

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu} := \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \boldsymbol{\Sigma}_{xy} := \begin{bmatrix} \sigma_x^2 & \rho_{xy} \\ \rho_{xy} & \sigma_y^2 \end{bmatrix}\right) \tag{50}$$

with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_{xy}$ their mean vector and covariance matrix respectively, as the previous definitions. The expression $E\{x\operatorname{sgn}\{y\}\}$ is given by

$$\begin{aligned}
&E\big\{x\operatorname{sgn}\{y\}\big\} \\
&= \int_{-\infty}^{+\infty}\operatorname{sgn}\{y\}\left(\int_{-\infty}^{+\infty} x\,\mathcal{N}\big([x,y]^\top, \boldsymbol{\Sigma}_{xy}\big)dx\right)dy \\
&= \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}_{xy}|}}\int_{-\infty}^{+\infty}\Big\{\operatorname{sgn}\{y\}\exp\Big(-\frac{1}{2}\big(b - \frac{c^2}{a}\big)(y-\mu_y)^2\Big) \\
&\quad\times\left(\frac{2\pi}{a}\int_{-\infty}^{+\infty} x\,\mathcal{N}\Big(\mu_x - \frac{c}{a}(y-\mu_y), \frac{1}{a}\Big)dx\right)\Big\}dy \\
&= \frac{1}{\sqrt{2\pi a|\boldsymbol{\Sigma}_{xy}|}}\Big\{\sqrt{\frac{2\pi}{\delta}}\big(\mu_x + \frac{c}{a}\mu_y\big)\big[1 - 2\phi(-\mu_y\sqrt{\delta})\big] \\
&\quad -\frac{c}{a}\sqrt{\frac{2\pi}{\delta}}\Big[\sqrt{\frac{2}{\pi\delta}}\exp\Big(-\frac{1}{2}\mu_y^2\delta\Big) + \mu_y\big(1 - 2\phi(-\mu_y\sqrt{\delta})\big)\Big]\Big\}
\end{aligned} \tag{51}$$

$$[\boldsymbol{T}^{(i,j)}_{\omega,n}]_{pq} = |\boldsymbol{I} + \frac{4}{\xi^2}\boldsymbol{R_{uu}}|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{8\xi^2}\left[4\sum_{k=\{i,j,p,q\}} \|\boldsymbol{u}_{\omega,k}\|^2 - \|\sum_{k=\{i,j,p,q\}} \boldsymbol{u}_{\omega,k}\|^2_{(\boldsymbol{I}+\xi^2\boldsymbol{R}^{-1}_{\boldsymbol{uu}}/4)^{-1}}\right]\right) \tag{39}$$

with the positive definite matrix $\boldsymbol{\Sigma}_{xy}$ of which inverse is defined as

$$\boldsymbol{\Sigma}^{-1}_{xy} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \tag{52}$$

and $\delta = b - c^2/a > 0$. The proof is referred to [11]. It should be noted that, in Lemmas 2 and 3, the random variables x and y are not necessarily zero-mean. Consequently, both of expectations $E\{\text{sgn}\{x\}\text{sgn}\{y\}\}$ and $E\{x\,\text{sgn}\{y\}\}$ are not able to be calculated via Price's theorem as in [34], [38]. Likewise, we should compute $\boldsymbol{Q}_{\omega,3}$ by making the following identifications:

$$x \triangleq [\boldsymbol{v}_{\omega,n}]_i \tag{53}$$
$$y \triangleq [\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}]_j \tag{54}$$

with

$$E\{[\boldsymbol{v}_{\omega,n}]_i\} \rightarrow \mu_x \tag{55}$$
$$E\{[\boldsymbol{w}^\star + \boldsymbol{v}_{\omega,n}]_j\} \rightarrow \mu_y. \tag{56}$$

The variances σ_x^2, σ_y^2 and covariance ρ_{xy} have the same expressions as in (47)–(49) because they are invariant to a constant shift. Similarly to $\boldsymbol{Q}_{\omega,2}$, the term $\boldsymbol{Q}_{\omega,3}$ is available to be calculated using (51).

Finally, the term $\boldsymbol{Q}_{\omega,4}$ can be expressed as

$$\boldsymbol{Q}_{\omega,4} = \boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}\,\boldsymbol{Q}_{\omega,3}. \tag{57}$$

With these expressions of expectation matrices $\boldsymbol{Q}_{\omega,1}$ to $\boldsymbol{Q}_{\omega,4}$, in the lexicographic form, i.e., the column vectors of a matrix are stacked on top of each other into a vector, the recursive update equation of (30) can be reformulated in the vector form as

$$\begin{aligned}\boldsymbol{c}_{\omega,n+1} &= (\boldsymbol{G}_{\omega,0} + \boldsymbol{G}_{\omega,1})\boldsymbol{c}_{\omega,n} + \eta^2\sigma_z^2\,\boldsymbol{r}_{\boldsymbol{\kappa\kappa},\omega} \\ &\quad + \rho^2\boldsymbol{q}_{\omega,2} + \rho\,\boldsymbol{G}_{\omega,2}\boldsymbol{q}_{\omega,3} + \rho\,\overline{\boldsymbol{G}}_{\omega,2}\bar{\boldsymbol{q}}_{\omega,3}\end{aligned} \tag{58}$$

with the definitions of variables

$$\boldsymbol{c}_{\omega,n} = \text{vec}\{\boldsymbol{C}_{\omega,n}\} \tag{59}$$
$$\boldsymbol{r}_{\boldsymbol{\kappa\kappa},\omega} = \text{vec}\{\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}\} \tag{60}$$
$$\boldsymbol{q}_{\omega,2} = \text{vec}\{\boldsymbol{Q}_{\omega,2}\} \tag{61}$$
$$\boldsymbol{q}_{\omega,3} = \text{vec}\{\boldsymbol{Q}_{\omega,3}\} \tag{62}$$
$$\bar{\boldsymbol{q}}_{\omega,3} = \text{vec}\{\boldsymbol{Q}^\top_{\omega,3}\} \tag{63}$$

and the expectations of matrices

$$\boldsymbol{G}_{\omega,0} = \boldsymbol{I}_{M^2} - \eta\left(\boldsymbol{I}_M \otimes \boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega} + \boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega} \otimes \boldsymbol{I}_M\right) \tag{64}$$
$$\boldsymbol{G}_{\omega,1} = \eta^2\,\overline{\boldsymbol{Q}}_{\omega,1} \tag{65}$$
$$\boldsymbol{G}_{\omega,2} = \eta\,\boldsymbol{I}_M \otimes \boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega} - \boldsymbol{I}_{M^2} \tag{66}$$
$$\overline{\boldsymbol{G}}_{\omega,2} = \eta\,\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega} \otimes \boldsymbol{I}_M - \boldsymbol{I}_{M^2}. \tag{67}$$

Here, the matrix $\boldsymbol{G}_{\omega,0}$ can be obtained referring to the results in [36], [37]. On account of the trace of product of two matrices in (37) and using (39), the any entry of matrix $\overline{\boldsymbol{Q}}_{\omega,1}$ in (65) can be given by

$$[\overline{\boldsymbol{Q}}_{\omega,1}]_{i+(j-1)M,p+(q-1)M} = [\boldsymbol{T}^{(i,j)}_{\omega,n}]_{pq} \tag{68}$$

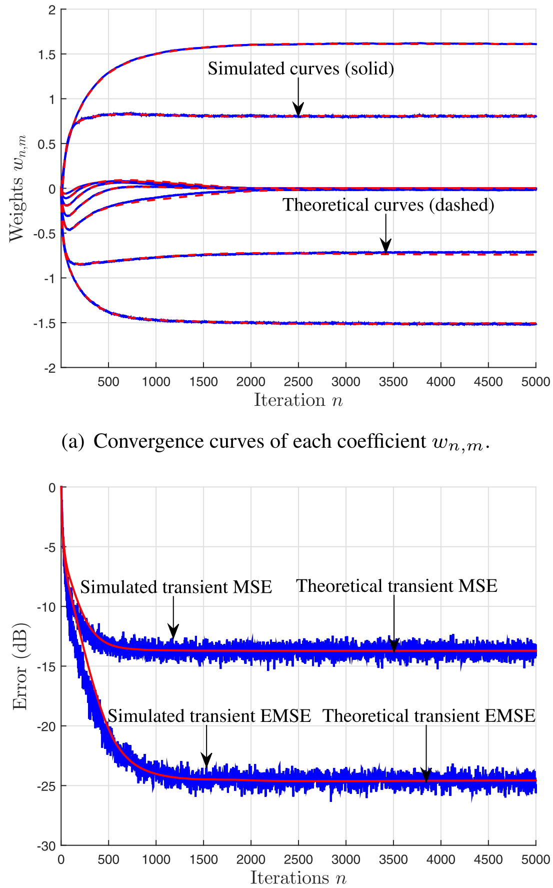
with $1 \leq i, j, p, q \leq M$. Since the relation between $\boldsymbol{Q}_{\omega,3}$ and $\boldsymbol{Q}_{\omega,4}$ is given by (57), using the product of two matrices in the lexicographic form, the last four terms of the RHS of (30) can be reformulated as (66) and (67). It should be pointed out that the first terms of the RHS of (66) and (67) are corresponding to the terms $\boldsymbol{Q}_{\omega,4}$ and $\boldsymbol{Q}^\top_{\omega,4}$, respectively. Hence, the second terms of the RHS of (66) and (67) are corresponding to the terms $\boldsymbol{Q}_{\omega,3}$ and $\boldsymbol{Q}^\top_{\omega,3}$, respectively.

Substituting (64)–(68) into (58) and using the expectations of $\boldsymbol{Q}_{\omega,1}$ to $\boldsymbol{Q}_{\omega,4}$, we can obtain the update equation of $\boldsymbol{c}_{\omega,n}$. Based on (58), then we can readily obtain the recursive analytical model (30) for the behavior of $\boldsymbol{C}_{\omega,n}$ by arranging $\boldsymbol{c}_{\omega,n}$ into the matrix form. Replacing (30) into (29), we finally can study the transient convergence behavior of the ZA-KLMS algorithm using Gaussian kernel with pre-tuned dictionary. The next section illustrates the derived models accuracy of predicting the transient performance of ZA-KLMS algorithm.

Note that $\boldsymbol{R}_{\boldsymbol{\kappa\kappa},\omega}$ and $\overline{\boldsymbol{Q}}_{\omega,1}$ are symmetric matrices, which imply that the matrices $\boldsymbol{G}_{\omega,0}$ and $\boldsymbol{G}_{\omega,1}$ are also symmetric. Assume that the assumption A.2 holds. For any initial condition, given a dictionary $\boldsymbol{\omega}$, the ZA-KLMS algorithm using Gaussian kernel (5) is mean-square stable if and only if the matrices $\boldsymbol{G}_{\omega,0}$ and $\boldsymbol{G}_{\omega,1}$ are stable following from (58), (64) and (65).

V. SIMULATION RESULTS

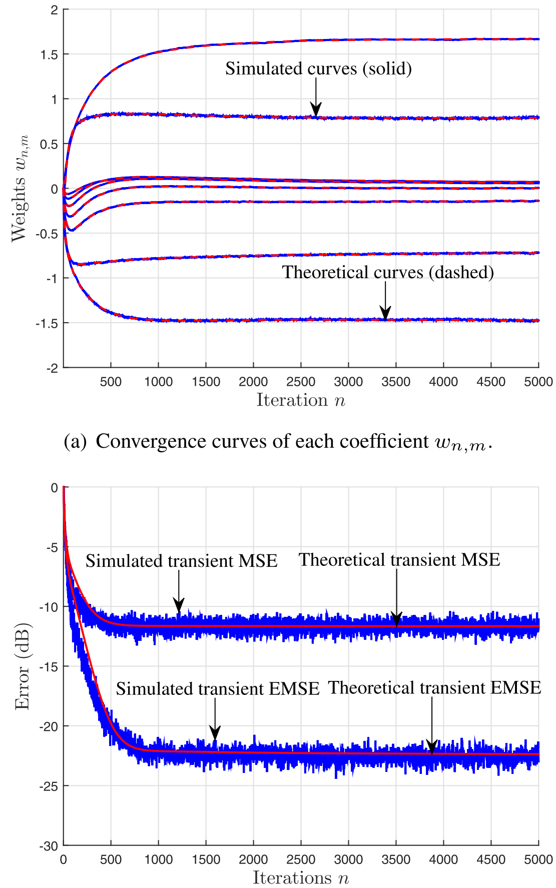
In this section, we present the simulation results of two examples to confirm the usefulness and accuracy of our derived analytical models. In order to show the superiority of ZA-KLMS algorithm for the nonlinear sparse system in stationary environment, the corresponding simulation results of the classical KLMS algorithm are also provided for performance comparison. It should be pointed out that the optimal weight vectors for most KLMS-type algorithms are unknown in advance and depend on the specific elements of dynamic dictionary in online manner. In order to reliably investigate the transient analysis of stochastic behavior of ZA-KLMS algorithms, we design the nonlinear system characterized by an optimum weight coefficients vector and the pre-tuned dictionary *a priori*. All the learning curves are obtained by averaging over 200 independent Monte Carlo trials.



(a) Convergence curves of each coefficient $w_{n,m}$.

(b) Learning curves of MSE and EMSE.

FIGURE 1. Simulation results of Example I for theoretical predictions and Monte Carlo simulation of ZA-KLMS.



(a) Convergence curves of each coefficient $w_{n,m}$.

(b) Learning curves of MSE and EMSE.

FIGURE 2. Simulation results of Example I for theoretical predictions and Monte Carlo simulation of KLMS.

A. EXAMPLE I

Consider an unknown nonlinear sparse system with the optimum weight coefficients vector defined by

$$\boldsymbol{w}^{\star} = \begin{bmatrix}1.65, & 0.8, & 0, & 0, & 0, & 0, & 0, & -0.75, & -1.5\end{bmatrix}^{\top} \tag{69}$$

and the fixed dictionary with 9-elements given by

$$\begin{aligned}\boldsymbol{\omega} &= \{u_{\omega,1}, u_{\omega,2}, \ldots, u_{\omega,9}\}\\ &= \{0.85, 0.2, -1.1, -1, -0.9, -0.8, -0.7, -0.5, -0.25\}\end{aligned}$$

where the dictionary elements $u_{\omega,m} \in \mathbb{R}^L$ with $L = 1$. As a consequence, the desired response of nonlinear system is defined as follow:

$$d_n = \sum_{m=1}^{M} w_m^{\star} \exp\left(-\frac{(u_n - u_{\omega,m})^2}{2\xi^2}\right) + z_n$$

where the input signal u_n is a zero-mean i.i.d. Gaussian random process with standard deviation $\sigma_u = 0.25$, and the additive noise z_n is a zero-mean i.i.d. Gaussian distribution with SNR $= -5$dB. The Gaussian kernel bandwidth ξ and the step-size η were set to 0.35 and 0.1, respectively. The regularization parameter λ was set to 1×10^{-3}.

Fig. 1(a) illustrates the mean convergence behavior of the estimated weight coefficient vector of ZA-KLMS algorithm. One can observe that the averaged curves of five negligible coefficients are rapidly attracted to zero due to the promoting-sparsity. As shown in Fig. 1(a), we can also see that the theoretical curves calculated by the analytical model consistently match with all the simulated curves of coefficients $w_{n,m}$. Fig. 1(b) shows that the theoretical MSE and EMSE curves coincide with their respective Mont Carlo simulated curves very well during the transient phase. Therefore, the simulation results of Example I demonstrate the effectiveness and accuracy of our derived analytical models for ZA-KLMS under the necessary assumptions and approximations.

One can obviously see from Fig. 2(a) that the five negligible weight coefficients of KLMS algorithm did not converge towards the optimal value zero resulting in the performance degradation. Fig. 2(b) shows that the simulated and theoretical learning curves of MSE and EMSE of KLMS are all much higher than those of ZA-KLMS shown in Fig. 1(b). Therefore, the ZA-KLMS algorithm exhibits better transient performance than the KLMS algorithm for the nonlinear sparse system verified by both of the simulated and theoretical results.

B. EXAMPLE II

For the second example, consider another unknown nonlinear sparse system whose the optimum weight coefficients vector is given by

$$\boldsymbol{w}^{\star} = \left[1.5,\ 1,\ 0.5,\ 0,\ 0,\ 0,\ 0,\ 0,\ -0.7,\ -1,\ -1.5\right]^{\top} \tag{70}$$

and the fixed dictionary with 11-elements given by

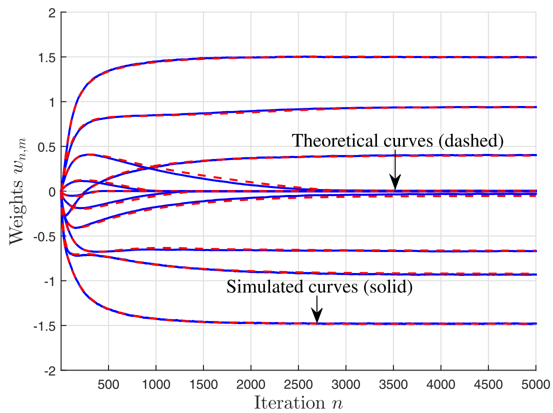
$$\begin{aligned}\boldsymbol{\omega} &= \left\{\boldsymbol{u}_{\omega,1}, \boldsymbol{u}_{\omega,2}, \ldots, \boldsymbol{u}_{\omega,11}\right\}\\ &= \left\{\begin{bmatrix}0.72\\1.44\end{bmatrix}, \begin{bmatrix}2.31\\1.28\end{bmatrix}, \begin{bmatrix}-1.54\\-0.29\end{bmatrix}, \begin{bmatrix}3.43\\1.81\end{bmatrix}, \begin{bmatrix}1.81\\-2.43\end{bmatrix}, \begin{bmatrix}2.01\\2.69\end{bmatrix},\right.\\ &\quad \left.\begin{bmatrix}-0.1\\-2.2\end{bmatrix}, \begin{bmatrix}2.13\\-1.16\end{bmatrix}, \begin{bmatrix}1.48\\-1.06\end{bmatrix}, \begin{bmatrix}-1.28\\-0.92\end{bmatrix}, \begin{bmatrix}0.32\\0.15\end{bmatrix}\right\}.\end{aligned}$$

Subsequently, the measured outputs of nonlinear system can be defined as follow:

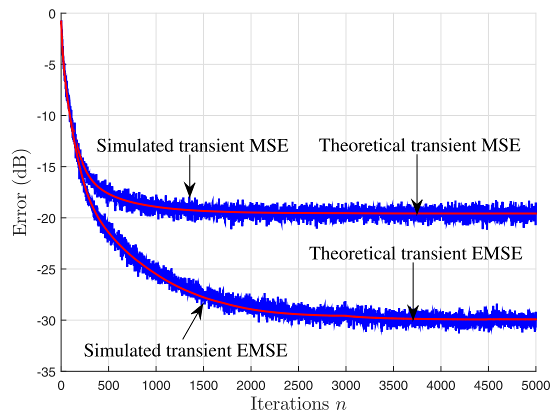
$$d_n = \sum_{m=1}^{M} w_m^{\star} \exp\left(-\frac{\|\boldsymbol{u}_n - \boldsymbol{u}_{\omega,m}\|^2}{2\xi^2}\right) + z_n$$

where the additive noise z_n follows zero-mean i.i.d. Gaussian distribution with SNR = 10dB. Furthermore, the input signal $\boldsymbol{u}_n$ was assumed to be a sequence of statistically independent vectors $\boldsymbol{u}_n = [u_{1,n}, u_{2,n}]^{\top}$ with correlated samples satisfying $u_{1,n} = 0.5\,u_{2,n} + v_{u,n}$, where $u_{2,n}$ is a white Gaussian random sequence with variance $\sigma_{u_2}^2 = 1$ and $v_{u,n}$ is also a white Gaussian random sequence with $\sigma_v^2 = 0.75$ so that $u_{1,n}$ has variance $\sigma_{u_1}^2 = 1$. The Gaussian kernel bandwidth ξ and the step-size η were set to 0.95 and 0.1, respectively. The regularization parameter λ was set to 2×10^{-3}.

Fig. 3(a) shows the mean convergence behavior of weight coefficients vector of ZA-KLMS, where five curves of negligible coefficients are enforced to rapidly approach to zero. In addition, it can be observed from Fig. 3(a) that the analytical model predictions and the Monte Carlo simulation curves of all coefficients $w_{n,m}$ are consistently superimposed. Likewise, Fig. 3(b) shows that the theoretical MSE and EMSE curves predicted by analytical models perfectly match with their respective simulated learning curves during the entire transient period. Simulation results of Example II illustrate the accuracy of the second order stochastic convergence model of ZA-KLMS with promoting-sparsity, and also validate the reasonableness of assumptions and approximations again. Our derived analytical models can allow to investigate

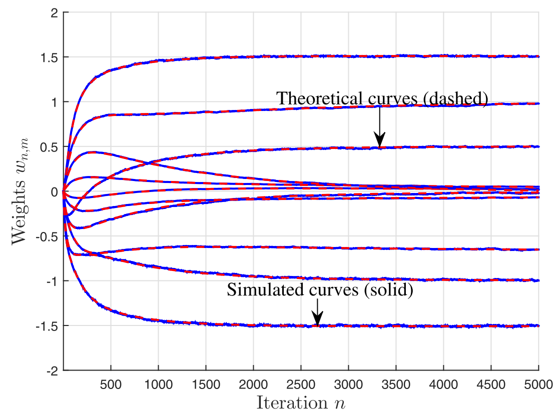


(a) Convergence curves of each coefficient $w_{n,m}$.

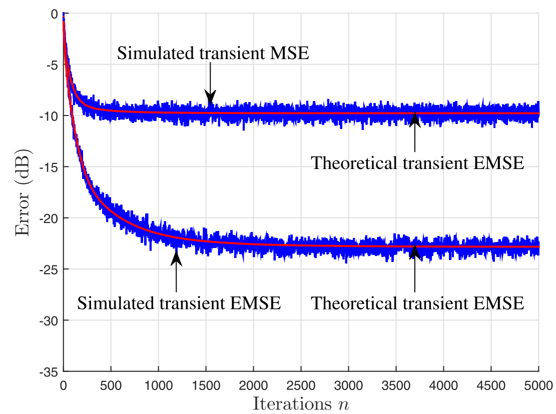


(b) Learning curves of MSE and EMSE.

FIGURE 3. Simulation results of Example II for theoretical predictions and Monte Carlo simulation of ZA-KLMS.



(a) Convergence curves of each coefficient $w_{n,m}$.



(b) Learning curves of MSE and EMSE.

FIGURE 4. Simulation results of Example II for theoretical predictions and Monte Carlo simulation of KLMS.

the transient convergence behaviors of ZA-KLMS algorithm with different parameters setting and compare the convergence performances among the distinct types of ZA-KLMS algorithms.

Fig 4(a) clearly shows that the five trivial weight coefficients of KLMS rapidly decrease but not to zero, which are different with ZA-KLMS as shown in Fig. 3(a). As comparison of Fig. 3(b) and Fig 4(b), we can obviously see that as a consequence of the nonzero weight coefficients the KLMS suffers from the severe performance degradation for the nonlinear sparse system. Consequently, both of simulated and theoretical transient simulation results demonstrate that ZA-KLMS algorithm significantly outperforms KLMS algorithm in terms of convergence speed and steady-state EMSE when applied to the nonlinear systems with sparse features.

VI. CONCLUSION

In this paper, we briefly reviewed the ZA-KLMS algorithm with ℓ_1-norm regularization for the nonlinear sparse system in stationary environment. Then, we studied the transient performance analysis of ZA-KLMS using Gaussian kernel function with fixed dictionary in the mean and mean-square senses. The accuracy of the derived analytical models was validated by the excellent agreement between the Monte Carlo simulations and the theoretical predictions. The simulations results also demonstrated that the ZA-KLMS outperforms the KLMS for nonlinear sparse system. In the future work, we will investigate the stochastic convergence behavior of sparse-ware KLMS-type algorithm in the presence of non-Gaussian impulsive noise.

REFERENCES

[1] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[2] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[3] L. B. Montefusco, D. Lazzaro, and S. Papi, "Fast sparse image reconstruction using adaptive nonlinear filtering," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 534–544, Feb. 2011.

[4] Q. Zhao, Q. Du, X. Gong, and Y. Chen, "Signal-preserving erratic noise attenuation via iterative robust sparsity-promoting filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3547–3560, Jun. 2018.

[5] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[6] K. Pelekanakis and M. Chitre, "Adaptive sparse channel estimation under symmetric alpha-stable noise," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3183–3195, Jun. 2014.

[7] 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.

[8] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE ICASSP*, Apr. 2009, pp. 3125–3128.

[9] S. Zhang and J. Zhang, "Transient analysis of zero attracting NLMS algorithm without Gaussian inputs assumption," *Signal Process.*, vol. 97, pp. 100–109, Apr. 2014.

[10] K. Shi and P. Shi, "Convergence analysis of sparse LMS algorithms with ℓ_1-norm penalty based on white input signal," *Signal Process.*, vol. 90, pp. 3289–3293, Dec. 2010.

[11] J. Chen, C. Richard, Y. Song, and D. Brie, "Transient performance analysis of zero-attracting LMS," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1786–1790, Dec. 2016.

[12] N. Kalouptsidis, G. Mileounis, B. Babadi, and V. Tarokh, "Adaptive algorithms for sparse nonlinear channel estimation," in *Proc. IEEE/SP 15th Workshop Stat. Signal Process.*, Aug./Sep. 2009, pp. 221–224.

[13] X. Yu, M. Rashid, J. Feng, N. Hobbs, I. Hajizadeh, S. Samadi, M. Sevil, C. Lazaro, Z. Maloney, E. Littlejohn, L. Quinn, and A. Cinar, "Online glucose prediction using computationally efficient sparse kernel filtering algorithms in type-1 diabetes," *IEEE Trans. Control Syst. Technol.*, to be published.

[14] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, and A. Uncini, "Nonlinear acoustic echo cancellation based on sparse functional link representations," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 7, pp. 1172–1183, Jul. 2014.

[15] S. Zhang and W. X. Zheng, "Recursive adaptive sparse exponential functional link neural network for nonlinear AEC in impulsive noise environment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4314–4323, Sep. 2018.

[16] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2275–2285, Aug. 2004.

[17] W. Liu, P. P. Pokharel, and J. C. Príncipe, "The kernel least-mean-square algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 543–554, Feb. 2008.

[18] B. Chen, S. Zhao, P. Zhu, and J. C. Príncipe, "Quantized kernel least mean square algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 22–32, Jan. 2012.

[19] W. Liu, J. C. Príncipe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*. New York, NY, USA: Wiley, 2010.

[20] W. Liu, I. Park, and J. C. Principe, "An information theoretic approach of designing sparse kernel adaptive filters," *IEEE Trans. Neural Netw.*, vol. 20, no. 12, pp. 1950–1961, Dec. 2009.

[21] H. Fan, Q. Song, and Z. Xu, "An information theoretic sparse kernel algorithm for online learning," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4349–4359, 2014.

[22] B. Chen, S. Zhao, S. Seth, and J. C. Príncipe, "Online efficient learning with quantized KLMS and L_1 regularization," in *Proc. IEEE IJCNN*, Brisbane, QLD, Australia, Jun. 2012, pp. 1–6.

[23] W. Gao, J. Chen, C. Richard, J. Huang, and R. Flamary, "Kernel LMS algorithm with forward-backward splitting for dictionary learning," in *Proc. IEEE ICASSP*, Vancouver, BC, Canada, May 2013, pp. 5735–5739.

[24] W. Gao, J. Chen, C. Richard, and J. Huang, "Online dictionary learning for kernel LMS," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2765–2777, Jun. 2014.

[25] M. Yukawa, "Multikernel adaptive filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4672–4682, Sep. 2012.

[26] M. Yukawa and R.-I. Ishii, "Online model selection and learning by multikernel adaptive filtering," in *Proc. EUSIPCO*, Marrakech, Morocco, Sep. 2013, pp. 1–5.

[27] M.-A. Takizawa and M. Yukawa, "Efficient dictionary-refining kernel adaptive filter with fundamental insights," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4337–4350, Aug. 2016.

[28] S. Scardapane, D. Comminiello, M. Scarpiniti, and A. Uncini, "Online sequential extreme learning machine with kernels," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 2214–2220, Sep. 2015.

[29] P. Honeine, "Analyzing sparse dictionaries for online learning with kernels," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6343–6353, Dec. 2015.

[30] B. Schölkopf, R. Herbrich, and R. Williamson, "A generalized representer theorem," NeuroCOLT, Roy. Holloway College, Univ. London, London, U.K., Tech. Rep. NC2-TR-2000-81, 2000.

[31] W. D. Parreira, J. C. M. Bermudez, C. Richard, and J. Y. Tourneret, "Stochastic behavior analysis of the Gaussian kernel least-mean-square algorithm," *IEEE Trans. Signal Process*, vol. 60, no. 5, pp. 2208–2222, May 2012.

[32] J. Omura and T. Kailath, "Some useful probability distributions," Stanford Electron. Lab., Stanford Univ., Stanford, CA, USA, USA, Tech. Rep. 7050-6, 1965.

[33] S. O. Haykin, *Adaptive Filter Theory*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1991.

[34] A. H. Sayed, *Adaptive Filters*. Hoboken, NJ, USA: Wiley, 2008.

[35] J. Minkoff, "Comment on the 'Unnecessary assumption of statistical independence between reference signal and filter weights in feedforward adaptive systems,"' *IEEE Trans. Signal Process.*, vol. 49, no. 5, p. 1109, May 2001.

[36] J. Chen, W. Gao, C. Richard, and J.-C. M. Bermudez, ''Convergence analysis of kernel LMS algorithm with pre-tuned dictionary,'' in *Proc. IEEE ICASSP*, Florence, Italy, May 2014, pp. 7243–7247.

[37] W. Gao, J. Chen, C. Richard, J.-C. M. Bermudez, and J. Huang, ''Convergence analysis of the augmented complex KLMS algorithm with pre-tuned dictionary,'' in *Proc. IEEE ICASSP*, Brisbane, QLD, Australia, Apr. 2015, pp. 2006–2010.

[38] J. Chen, J. C. M. Bermudez, and C. Richard, ''Steady-state performance of non-negative least-mean-square algorithm and its variants,'' *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 928–932, Aug. 2014.

WEI GAO (S'11–M'17) received the M.S. degree in signal and information processing and the Ph.D. degree in information and communication engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 2010 and 2014, respectively, and the Ph.D. degree in engineering science from the University of Nice Sophia Antipolis, Nice, France, in 2016.

He is currently an Associate Professor with Jiangsu University, Zhenjiang, China. His current research interests include nonlinear adaptive filtering, adaptive noise cancellation, and array signal processing.

JIE CHEN (S'12–M'14–SM'18) received the Dipl.Ing. degree in information and telecommunication engineering from the Troyes University of Technology (UTT), Troyes, France, in 2009, the M.S. degree in information and telecommunication engineering from Xi'an Jiaotong University, Xi'an, China, and the Ph.D. degree from UTT, in 2013.

From 2013 to 2014, he was with Lagrange Laboratory, University of Nice Sophia Antipolis, France. From 2014 to 2015, he was with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA. He is currently a Professor with the Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an. His current research interests include adaptive signal processing, distributed optimization with applications in hyperspectral image analysis, acoustic signal processing, and bioinformatics. He was a recipient of the Thousand Talents Plan (Youth Program) Award in China. He was the Technical Co-Chair of IWAENC 2016, Xi'an.