Stochastic Behavior Analysis of the Gaussian Kernel Least-Mean-Square Algorithm

Wemerson D. Parreira, José Carlos M. Bermudez, *Senior Member, IEEE*, Cédric Richard, *Senior Member, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—The kernel least-mean-square (KLMS) algorithm is a popular algorithm in nonlinear adaptive filtering due to its simplicity and robustness. In kernel adaptive filters, the statistics of the input to the linear filter depends on the parameters of the kernel employed. Moreover, practical implementations require a finite nonlinearity model order. A Gaussian KLMS has two design parameters, the step size and the Gaussian kernel bandwidth. Thus, its design requires analytical models for the algorithm behavior as a function of these two parameters. This paper studies the steady-state behavior and the transient behavior of the Gaussian KLMS algorithm for Gaussian inputs and a finite order nonlinearity model. In particular, we derive recursive expressions for the mean-weight-error vector and the mean-square-error. The model predictions show excellent agreement with Monte Carlo simulations in transient and steady state. This allows the explicit analytical determination of stability limits, and gives opportunity to choose the algorithm parameters *a priori* in order to achieve prescribed convergence speed and quality of the estimate. Design examples are presented which validate the theoretical analysis and illustrates its application.**

***Index Terms*—Adaptive filtering, kernel least-mean-square (KLMS), convergence analysis, nonlinear system, reproducing kernel.**

I. Introduction

MANY practical applications (e.g., in communications and bioengineering) require nonlinear signal processing. Nonlinear systems can be characterized by representations ranging from higher-order statistics to series expansion methods [1]. Nonlinear system identification methods based on reproducing kernel Hilbert spaces (RKHS) have gained popularity over the last decades [2], [3]. More recently, kernel adaptive filtering has been recognized as an appealing solution to the nonlinear adaptive filtering problem, as working in RKHS allows the use of linear structures to solve nonlinear estimation problems. See [4] for an overview. The block diagram of a kernel-based adaptive system identification problem is shown in Fig. 1. Here, $\mathcal{U}$ is a compact subspace of $\mathbb{R}^q$, $\kappa\colon \mathcal{U}\times\mathcal{U}\to\mathbb{R}$ is a reproducing kernel, $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is the induced RKHS with its inner product and $z(n)$ is a zero-mean additive noise uncorrelated with any other signal. The representer theorem [2] states that the function $\psi(\,\cdot\,)$ which minimizes the squared estimation error $\sum_{n=1}^{N}(d(n) - \psi(\boldsymbol{u}(n)))^2$, given N input vectors $\boldsymbol{u}(n)$ and desired outputs $d(n)$, can be written as the kernel expansion $\psi(\,\cdot\,) = \sum_{n=1}^{N}\alpha_n\,\kappa(\cdot,\boldsymbol{u}(n))$. This reduces the problem to determining $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_N]^\top$ that minimizes $\|\boldsymbol{d} - \boldsymbol{K}\boldsymbol{\alpha}\|^2$, where $\boldsymbol{K}$ is the Gram matrix with (n,ℓ)th entry $\kappa(\boldsymbol{u}(n),\boldsymbol{u}(\ell))$, and $\boldsymbol{d} = [d(1),\ldots,d(N)]^\top$. Since the order of the model is equal to the number N of available data $\boldsymbol{u}(n)$, this approach cannot be considered for online applications. To overcome this barrier, one can focus on finite-order models

$$\psi(\,\cdot\,) = \sum_{j=1}^{M}\alpha_j\,\kappa(\cdot,\boldsymbol{u}(\omega_j)) \tag{1}$$

where ω_j, $j = 1,\ldots,M$ form a subset of $\{1,2,\ldots,N\}$ corresponding to the time indexes of the $M < N$ input vectors chosen to build the Mth-order model (1). The M kernel functions $\kappa(\cdot,\boldsymbol{u}(\omega_j))$ form the *dictionary*. In [4], the authors present an overview of the existing techniques to select the kernel functions in (1), an example of which is the approximate linear dependence (ALD) criterion [5]. It consists of including a kernel function $\kappa(\cdot,\boldsymbol{u}(\ell))$ in the dictionary if it satisfies

$$\min_{\boldsymbol{\gamma}}\|\kappa(\cdot,\boldsymbol{u}(\ell)) - \sum_{j}\gamma_j\,\kappa(\cdot,\boldsymbol{u}(\omega_j))\|_{\mathcal{H}}^2 > \epsilon_0 \tag{2}$$

where ϵ_0 is a parameter determining model sparsity level. To control the model order with reduced computational complexity, the coherence-based sparsification rule has also been considered [6], [7]. According to this rule, kernel $\kappa(\cdot,\boldsymbol{u}(\ell))$ is inserted into the dictionary if

$$\max_{j}|\kappa(\boldsymbol{u}(\ell),\boldsymbol{u}(\omega_j))| \le \varepsilon_0 \tag{3}$$

with ε_0 a parameter determining the dictionary coherence. It was shown in [7] that the dictionary dimension determined under rule (3) is finite. For the rest of the paper, we shall assume that the dictionary size M is known, fixed and finite.

It is well known that a nonlinear adaptive filtering problem with input signal in $\mathcal{U}$ can be solved using a linear adaptive filter [4]. The linear adaptive filter input is a nonlinear mapping of $\mathcal{U}$ to an Hilbert space $\mathcal{H}$ possessing a reproducing kernel. The theory outlined above shows that the order of the linear adaptive filter can be finite if a proper input sparsification rule is employed, even if the dimensionality of the transformed input in

W. D. Parreira J. C. M. Bermudezis with the Federal University of Santa Catarina, Florianópolis, SC, Brazil (e-mail: wemerson@lpds.ufsc.br; j.bermudez@ieee.org).

C. Richard is the Université de Nice Sophia-Antipolis, Institut Universitaire de France, France (e-mail: cedric.richard@unice.fr).

J.-Y. Tourneret is with the Université de Toulouse, CNRS, France (e-mail: jean-yves.tourneret@enseeiht.fr).

Digital Object Identifier 10.1109/TSP.2012.2186132

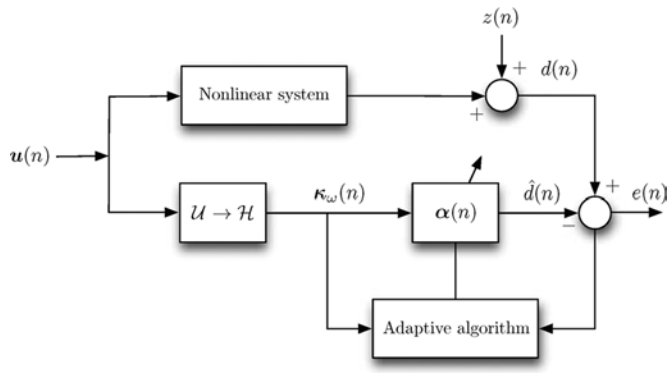


Fig. 1. Kernel-based adaptive system identification.

$\mathcal{H}$ is infinite as in the case of Gaussian kernel. Algorithms developed using these ideas include the kernel least-mean-square (KLMS) algorithm [8], [9], the kernel recursive-least-square (KRLS) algorithm [5], the kernel-based normalized least-mean-square (KNLMS) algorithm and the affine projection (KAPA) algorithm [6], [10], [7]. See also the monograph [4]. In addition to the choice of the usual linear adaptive filter parameters, designing kernel adaptive filters requires the choice of the kernel and its parameters. Moreover, using the finite order model in (1) implies that the adaptive algorithm behavior cannot be studied as the behavior of the algorithm presented in [4, (2.17)], which is a regular LMS algorithm operating in the RKHS. Choosing the algorithm and nonlinearity model parameters to achieve a prescribed performance is still an open issue, and requires an extensive analysis of the algorithm stochastic behavior. Our work brings a new contribution to the discussion about kernel-based adaptive filtering by providing the first convergence analysis of the KLMS algorithm with Gaussian kernel.

The paper is structured as follows. In Section II, we derive recursive expressions for the mean weight error vector and the mean-square error (MSE) for Gaussian inputs. In Section III, we define analytical models for the transient behavior of the first and second-order moments of the adaptive weights. Section IV studies the algorithm convergence properties. Stability conditions and a steady-state behavior model are derived which allow the algorithm design for prescribed convergence speed and quality of estimate. In Section V, we use the analysis results to establish design guidelines. Section VI presents design examples which validate the theoretical analysis and illustrate its application. The model predictions show excellent agreement with Monte Carlo simulations both in transient and steady state.

II. Mean Square Error Analysis

Consider the nonlinear system identification problem shown in Fig. 1, and the finite-order model (1) based on the Gaussian kernel

$$\kappa(\boldsymbol{u}, \boldsymbol{u}') = \exp\left(\frac{-\|\boldsymbol{u}-\boldsymbol{u}'\|^2}{2\xi^2}\right) \tag{4}$$

where ξ is the kernel bandwidth. The environment is assumed stationary, meaning that $\psi(\boldsymbol{u}(n))$ is stationary for $\boldsymbol{u}(n)$ stationary. This assumption is satisfied by several nonlinear systems used to model practical situations, such as memoryless, Wiener and Hammerstein systems. System inputs are zero-mean, independent, and identically distributed Gaussian $(q \times 1)$ vectors $\boldsymbol{u}(n)$ so that $E\{\boldsymbol{u}(n-i)\,\boldsymbol{u}^\top(n-j)\} = \boldsymbol{0}$ for $i \neq j$. The components of the input vector $\boldsymbol{u}(n)$ can, however, be correlated. Let $\boldsymbol{R}_{\boldsymbol{uu}} = E\{\boldsymbol{u}(n)\,\boldsymbol{u}^\top(n)\}$ denote their autocorrelation matrix.

For a dictionary of size M, let $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$ be the vector of kernels at time $n > M$,[1] that is

$$\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n) = [\kappa(\boldsymbol{u}(n), \boldsymbol{u}(\omega_1)), \ldots, \kappa(\boldsymbol{u}(n), \boldsymbol{u}(\omega_M))]^\top \tag{5}$$

where $\boldsymbol{u}(\omega_i)$ is the ith element of the dictionary, with $\boldsymbol{u}(\omega_i) \neq \boldsymbol{u}(n)$ for $i = 1, \ldots, M$. Here we consider that the vectors $\boldsymbol{u}(\omega_i)$, $i = 1, \ldots, M$ may change at each iteration following some dictionary updating schedule. The only limitation imposed in the following analysis is that $\boldsymbol{u}(\omega_i(n)) \neq \boldsymbol{u}(\omega_j(n))$ for $i \neq j$ so that the dictionary vectors which are arguments of different entries of $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$ are statistically independent. To keep the notation simple, however, we will not show explicitly the dependence of ω_i on n and represent $\boldsymbol{u}(\omega_i(n))$ as $\boldsymbol{u}(\omega_i)$ for all i.

From Fig. 1 and model (1), the estimated system output is

$$\hat{d}(n) = \boldsymbol{\alpha}^\top(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n) \tag{6}$$

with $\boldsymbol{\alpha}(n) = [\alpha_1(n), \ldots, \alpha_M(n)]^\top$. The corresponding estimation error is defined as

$$e(n) = d(n) - \hat{d}(n). \tag{7}$$

Squaring both sides of (7) and taking the expected value leads to the MSE

$$J_{ms}(n) = E\{e^2(n)\} = E\{d^2(n)\} - 2\boldsymbol{p}_{\kappa d}^\top\,\boldsymbol{\alpha}(n) + \boldsymbol{\alpha}^\top(n)\,\boldsymbol{R}_{\kappa\kappa}\,\boldsymbol{\alpha}(n) \tag{8}$$

where $\boldsymbol{R}_{\kappa\kappa} = E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\}$ is the correlation matrix of the kernelized input, and $\boldsymbol{p}_{\kappa d} = E\{d(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\}$ is the cross-correlation vector between $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$ and $d(n)$. It is shown in Appendix A that $\boldsymbol{R}_{\kappa\kappa}$ is positive definite. Thus, the optimum weight vector is given by

$$\boldsymbol{\alpha}_{\mathrm{opt}} = \boldsymbol{R}_{\kappa\kappa}^{-1}\,\boldsymbol{p}_{\kappa d} \tag{9}$$

and the corresponding minimum MSE is

$$J_{\min} = E\{d^2(n)\} - \boldsymbol{p}_{\kappa d}^\top\,\boldsymbol{R}_{\kappa\kappa}^{-1}\,\boldsymbol{p}_{\kappa d}. \tag{10}$$

These are the well-known expressions of the Wiener solution and minimum MSE, where the input signal vector has been replaced by the kernelized input vector. Determining the optimum $\boldsymbol{\alpha}_{\mathrm{opt}}$ requires the determination of the covariance matrix

[1] If the dictionary size M is adapted online, assume that n is sufficiently large so that the size M does not increase anymore.

$\boldsymbol{R}_{\kappa\kappa}$, given the statistical properties of $\boldsymbol{u}(n)$ and the reproducing kernel.

Before closing this section, let us evaluate the correlation matrix $\boldsymbol{R}_{\kappa\kappa}$. Its entries are given by

$$[\boldsymbol{R}_{\kappa\kappa}]_{ij} = \begin{cases} E\{\kappa^2(\boldsymbol{u}(n),\boldsymbol{u}(\omega_i))\}, & i=j \\ E\{\kappa(\boldsymbol{u}(n),\boldsymbol{u}(\omega_i))\,\kappa(\boldsymbol{u}(n),\boldsymbol{u}(\omega_j))\}, & i\neq j \end{cases} \quad (11)$$

with $1 \leq i,\, j \leq M$. Note that $\boldsymbol{R}_{\kappa\kappa}$ remains time-invariant even if the dictionary is updated at each iteration, as $\boldsymbol{u}(n)$ is stationary and $\boldsymbol{u}(\omega_i)$ and $\boldsymbol{u}(\omega_j)$ are statistically independent for $i \neq j$.

Let us introduce the following notations

$$\begin{aligned} &\|\boldsymbol{u}(n)-\boldsymbol{u}(\omega_i)\|^2 = \boldsymbol{y}_2^\top\,\boldsymbol{Q}_2\,\boldsymbol{y}_2 \\ &\|\boldsymbol{u}(n)-\boldsymbol{u}(\omega_i)\|^2 + \|\boldsymbol{u}(n)-\boldsymbol{u}(\omega_j)\|^2 = \boldsymbol{y}_3^\top\,\boldsymbol{Q}_3\,\boldsymbol{y}_3, \quad i\neq j \end{aligned} \quad (12)$$

where $\|\cdot\|$ is the ℓ_2 norm and

$$\begin{aligned} \boldsymbol{y}_2 &= (\boldsymbol{u}^\top(n)\;\boldsymbol{u}^\top(\omega_i))^\top \\ \boldsymbol{y}_3 &= (\boldsymbol{u}^\top(n)\;\boldsymbol{u}^\top(\omega_i)\;\boldsymbol{u}^\top(\omega_j))^\top \end{aligned} \quad (13)$$

and

$$\boldsymbol{Q}_2 = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{I} \\ -\boldsymbol{I} & \boldsymbol{I} \end{pmatrix} \quad \boldsymbol{Q}_3 = \begin{pmatrix} 2\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{I} \\ -\boldsymbol{I} & \boldsymbol{I} & \boldsymbol{O} \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{I} \end{pmatrix} \quad (14)$$

where $\boldsymbol{I}$ is the $(q\times q)$ identity matrix and $\boldsymbol{O}$ is the $(q\times q)$ null matrix. From [11, p. 100], we know that the moment generating function of a quadratic form $z=\boldsymbol{y}^\top\boldsymbol{Q}\,\boldsymbol{y}$, where $\boldsymbol{y}$ is a zero-mean Gaussian vector with covariance matrix $\boldsymbol{R}_y$, is given by

$$\psi_z(s) = E\{e^{sz}\} = \det\{\boldsymbol{I}-2\,s\boldsymbol{Q}\,\boldsymbol{R}_y\}^{-1/2}. \quad (15)$$

Making $s=-1/(2\xi^2)$ in (15), we find that the (i,j)th element of $\boldsymbol{R}_{\kappa\kappa}$ is given by

$$[\boldsymbol{R}_{\kappa\kappa}]_{ij} = \begin{cases} r_{\rm md} = \det\{\boldsymbol{I}_2 + 2\,\boldsymbol{Q}_2\,\boldsymbol{R}_2/\xi^2\}^{-1/2}, & i=j \\ r_{\rm od} = \det\{\boldsymbol{I}_3 + \boldsymbol{Q}_3\,\boldsymbol{R}_3/\xi^2\}^{-1/2}, & i\neq j \end{cases} \quad (16)$$

with $1\leq i,\,j\leq M$. The main diagonal entries $[\boldsymbol{R}_{\kappa\kappa}]_{ii}$ are all equal to $r_{\rm md}$ and the off-diagonal entries $[\boldsymbol{R}_{\kappa\kappa}]_{ij}$ are all equal to $r_{\rm od}$ because $\boldsymbol{u}(\omega_i)$ and $\boldsymbol{u}(\omega_j)$ are i.i.d. In (16), $\boldsymbol{R}_\ell$ is the $(\ell q\times \ell q)$ correlation matrix of vector $\boldsymbol{y}_\ell$, $\boldsymbol{I}_\ell$ is the $(\ell q\times \ell q)$ identity matrix, and $\det\{\cdot\}$ denotes the determinant of a matrix. Finally, note that matrix $\boldsymbol{R}_\ell$ is block-diagonal with $\boldsymbol{R}_{\boldsymbol{uu}}$ along its diagonal.

III. Gaussian KLMS Algorithm: Transient Behavior Analysis

The KLMS weight-update equation for the system presented in Fig. 1 is [4]

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \eta\,e(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n). \quad (17)$$

Defining the weight-error vector $\boldsymbol{v}(n) = \boldsymbol{\alpha}(n) - \boldsymbol{\alpha}_{\rm opt}$ leads to the weight-error vector update equation

$$\boldsymbol{v}(n+1) = \boldsymbol{v}(n) \;+\eta\,e(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n). \quad (18)$$

From (6) and (7), and the definition of $\boldsymbol{v}(n)$, the error equation is given by

$$e(n) = d(n) - \boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\,\boldsymbol{v}(n) - \boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\,\boldsymbol{\alpha}_{\rm opt} \quad (19)$$

and the optimal estimation error is

$$e_0(n) = d(n) - \boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\,\boldsymbol{\alpha}_{\rm opt}. \quad (20)$$

Substituting (19) into (18) yields

$$\begin{aligned} \boldsymbol{v}(n+1) = \boldsymbol{v}(n) &+ \eta\,d(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n) \\ &-\eta\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\,\boldsymbol{v}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n) - \eta\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\,\boldsymbol{\alpha}_{\rm opt}\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n). \end{aligned} \quad (21)$$

A. Simplifying Statistical Assumptions

Simplifying assumptions are required in order to make the study of the stochastic behavior of $\boldsymbol{v}(n)$ mathematically feasible. The statistical assumptions required in different parts of the analysis are the following:

A1: $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)$ is statistically independent of $\boldsymbol{v}(n)$. This assumption is justified in detail in [12] and has been successfully employed in several adaptive filter analyses. It is called here for further reference "modified independence assumption" (MIA). This assumption has been shown in [12] to be less restrictive than the classical independence assumption [13].

A2: The finite-order model provides a close enough approximation to the infinite-order model with minimum MSE, so that $E[e_0(n)] \approx 0$.

A3: $e_0(n)$ and $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)$ are uncorrelated. This assumption is also supported by the arguments supporting the MIA (**A1**) [12].

B. Mean Weight Behavior

Taking the expected value of both sides of (21) and using the MIA (**A1**) yields

$$E\{\boldsymbol{v}(n+1)\} = (\boldsymbol{I}-\eta\,\boldsymbol{R}_{\kappa\kappa})\,E\{\boldsymbol{v}(n)\} \quad (22)$$

which is the LMS mean weight behavior for an input vector $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$.

C. Mean-Square Error

Using (19) and the MIA (**A1**), the second-order moments of the weights are related to the MSE through [13]

$$J_{ms}(n) = J_{\rm min} + \mathrm{trace}\{\boldsymbol{R}_{\kappa\kappa}\,\boldsymbol{C}_v(n)\} \quad (23)$$

where $\boldsymbol{C}_v(n) = E\{\boldsymbol{v}(n)\,\boldsymbol{v}^\top(n)\}$ is the autocorrelation matrix of $\boldsymbol{v}(n)$ and $J_{\rm min} = E\{e_0^2(n)\}$ the minimum MSE. The study of the MSE behavior (23) requires a model for $\boldsymbol{C}_v(n)$. This model is highly affected by the transformation imposed on the input signal $\boldsymbol{u}(n)$ by the kernel. An analytical model for the behavior of $\boldsymbol{C}_v(n)$ is derived in the next subsection.

D. Second-Order Moment Behavior

Using (20) and (21), the weight-error vector update becomes

$$\begin{aligned} \boldsymbol{v}(n+1) = \boldsymbol{v}(n) &+ \eta\,e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n) \\ &-\eta\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\,\boldsymbol{v}(n). \end{aligned} \quad (24)$$

Post-multiplying (24) by its transpose, and taking the expected value, leads to

$$\begin{aligned}
&\boldsymbol{C}_v(n+1)\\
&\quad= \boldsymbol{C}_v(n) - \eta\, E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\,\boldsymbol{v}(n)\,\boldsymbol{v}^{\top}(n)\}\\
&\qquad - \eta\, E\{\boldsymbol{v}(n)\,\boldsymbol{v}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}\\
&\qquad + \eta^2\, E\left\{e_0^2(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\right\}\\
&\qquad + \eta\, E\{e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{v}^{\top}(n)\}\\
&\qquad - \eta^2\, E\{e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{v}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}\\
&\qquad + \eta\, E\{e_0(n)\,[\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{v}^{\top}(n)]^{\top}\}\\
&\qquad - \eta^2\, E\{e_0(n)\,[\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{v}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)]^{\top}\}\\
&\qquad + \eta^2\, E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\,\boldsymbol{v}(n)\,\boldsymbol{v}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}.
\end{aligned} \tag{25}$$

Using the MIA (**A1**), the first two expected values are given by

$$\begin{aligned}
E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\,\boldsymbol{v}(n)\,\boldsymbol{v}^{\top}(n)\} &= \boldsymbol{R}_{\kappa\kappa}\,\boldsymbol{C}_v(n)\\
E\{\boldsymbol{v}(n)\,\boldsymbol{v}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\} &= \boldsymbol{C}_v(n)\,\boldsymbol{R}_{\kappa\kappa}.
\end{aligned} \tag{26}$$

Using assumptions **A2** and **A3** the third expected value is given by

$$\begin{aligned}
&E\left\{e_0^2(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\right\}\\
&\quad\approx E\{e_0^2(n)\}\,E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}\\
&\quad= \boldsymbol{R}_{\kappa\kappa}\,J_{\min}.
\end{aligned} \tag{27}$$

The fourth and the sixth expected values can be approximated using the MIA (**A1**), that is,

$$\begin{aligned}
E\{e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{v}^{\top}(n)\} &\approx E\{e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\}\,E\{\boldsymbol{v}^{\top}(n)\}\\
&= 0
\end{aligned} \tag{28}$$

since $E\{e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\} = 0$ by the orthogonality principle [13].

Evaluation of the fifth and seventh expected values requires further simplifications for mathematical tractability. A reasonable approximation that preserves the effect of $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$ up to its second-order moments is to assume that $e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$ and $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)$ are uncorrelated[2]. Under both this approximation and MIA (**A1**),

$$\begin{aligned}
&E\{e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{v}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}\\
&\quad= E\{\boldsymbol{v}^{\top}(n)\,e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}\\
&\quad\approx E\{\boldsymbol{v}^{\top}(n)\}\,E\{e_0(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\}\,E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\} = 0
\end{aligned} \tag{29}$$

where the equality to zero is due to the orthogonality principle.

Using (26)–(29) in (25) yields

$$\begin{aligned}
\boldsymbol{C}_v(n+1) \approx \boldsymbol{C}_v(n) &- \eta\,(\boldsymbol{R}_{\kappa\kappa}\,\boldsymbol{C}_v(n) + \boldsymbol{C}_v(n)\,\boldsymbol{R}_{\kappa\kappa})\\
&+\eta^2\,\boldsymbol{T}(n) + \eta^2\,\boldsymbol{R}_{\kappa\kappa}\,J_{\min}
\end{aligned} \tag{30a}$$

with

$$\boldsymbol{T}(n) = E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\,\boldsymbol{v}(n)\,\boldsymbol{v}^{\top}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}. \tag{30b}$$

[2]Using this approximation we are basically neglecting the fluctuations of $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)$ about its mean $E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^{\top}(n)\}$.

Evaluation of expectation (30b) is an important step in the analysis. In the classical LMS analysis [14], the input signal is assumed zero-mean Gaussian. Then the expectation in (30b) can be approximated using the moment factoring theorem for Gaussian variates. In the present analysis, as $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$ is a nonlinear transformation of a quadratic function of the Gaussian input vector $\boldsymbol{u}(n)$, it is neither zero-mean nor Gaussian.

Using the MIA (**A1**) to determine the (i,j)th element of $\boldsymbol{T}(n)$ in (30b) yields

$$\begin{aligned}
[\boldsymbol{T}(n)]_{ij} \approx \sum_{\ell=1}^{M}\sum_{p=1}^{M} &E\left\{\kappa_{\omega_i}(n)\,\kappa_{\omega_\ell}(n)\,\kappa_{\omega_p}(n)\,\kappa_{\omega_j}(n)\right\}\\
&\times [\boldsymbol{c}_v(n)]_{\ell p}
\end{aligned} \tag{31}$$

where $\kappa_{\omega_q}(n) = \kappa(\boldsymbol{u}(n), \boldsymbol{u}(\omega_q))$. Each of the moments in (31) can be determined using (15). Depending on i, j, ℓ and p, we have five different moments μ_k, $k = 1, \ldots, 5$ to evaluate.

$\mu_1 := E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_\ell}(n)\,\kappa_{\boldsymbol{\omega}_p}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\}$ with $i = j = p = \ell$.
Denoting $\boldsymbol{y}_2 = (\boldsymbol{u}^{\top}(n)\,\boldsymbol{u}^{\top}(\omega_i))^{\top}$, yields

$$\mu_1 = [\det\{\boldsymbol{I}_2 + 4\,\boldsymbol{Q}_2\,\boldsymbol{R}_2/\xi^2\}]^{-1/2} \tag{32}$$

$\mu_2 := E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_\ell}(n)\,\kappa_{\boldsymbol{\omega}_p}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\}$ with $i = j = p \neq \ell$.
Denoting $\boldsymbol{y}_3 = (\boldsymbol{u}^{\top}(n)\,\boldsymbol{u}^{\top}(\omega_i)\,\boldsymbol{u}^{\top}(\omega_\ell))^{\top}$, yields

$$\mu_2 = [\det\{\boldsymbol{I}_3 + \boldsymbol{Q}_{3'}\,\boldsymbol{R}_3/\xi^2\}]^{-1/2} \tag{33}$$

where

$$\boldsymbol{Q}_{3'} = \begin{pmatrix} 4\boldsymbol{I} & -3\boldsymbol{I} & -\boldsymbol{I} \\ -3\boldsymbol{I} & 3\boldsymbol{I} & \boldsymbol{O} \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{I} \end{pmatrix}. \tag{34}$$

$\mu_3 := E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_\ell}(n)\,\kappa_{\boldsymbol{\omega}_p}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\}$ with $i = j \neq p = \ell$.
Denoting $\boldsymbol{y}_3 = (\boldsymbol{u}^{\top}(n)\,\boldsymbol{u}^{\top}(\omega_i)\,\boldsymbol{u}^{\top}(\omega_p))^{\top}$, yields

$$\mu_3 = [\det\{\boldsymbol{I}_3 + 2\,\boldsymbol{Q}_3\,\boldsymbol{R}_3/\xi^2\}]^{-1/2} \tag{35}$$

$\mu_4 := E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_\ell}(n)\,\kappa_{\boldsymbol{\omega}_p}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\}$ with $i = j \neq p \neq \ell$.
Denoting $\boldsymbol{y}_4 = (\boldsymbol{u}^{\top}(n)\,\boldsymbol{u}^{\top}(\omega_i)\,\boldsymbol{u}^{\top}(\omega_\ell)\,\boldsymbol{u}^{\top}(\omega_p))^{\top}$, yields

$$\mu_4 = [\det\{\boldsymbol{I}_4 + \boldsymbol{Q}_4\,\boldsymbol{R}_4/\xi^2\}]^{-1/2} \tag{36}$$

where

$$\boldsymbol{Q}_4 = \begin{pmatrix} 4\boldsymbol{I} & -2\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{I} \\ -2\boldsymbol{I} & 2\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{O} \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{I} & \boldsymbol{O} \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{I} \end{pmatrix}. \tag{37}$$

$\mu_5 := E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_\ell}(n)\,\kappa_{\boldsymbol{\omega}_p}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\}$ with $i \neq j \neq p \neq \ell$.
Denoting
$\boldsymbol{y}_5 = (\boldsymbol{u}^{\top}(n)\,\boldsymbol{u}^{\top}(\omega_i)\,\boldsymbol{u}^{\top}(\omega_j)\,\boldsymbol{u}^{\top}(\omega_\ell)\,\boldsymbol{u}^{\top}(\omega_p))^{\top}$,

$$\mu_5 = [\det\{\boldsymbol{I}_5 + \boldsymbol{Q}_5\,\boldsymbol{R}_5/\xi^2\}]^{-1/2} \tag{38}$$

where

$$\boldsymbol{Q}_5 = \begin{pmatrix} 4\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{I} \\ -\boldsymbol{I} & \boldsymbol{I} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{I} & \boldsymbol{O} & \boldsymbol{O} \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{I} & \boldsymbol{O} \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{I} \end{pmatrix}. \tag{39}$$

Using these moments, the elements of $\boldsymbol{T}(n)$ are finally given by

$$\begin{aligned}[\boldsymbol{T}(n)]_{ii} &= \mu_1[\boldsymbol{C}_v(n)]_{ii} \\ &\quad + \sum_{\substack{\ell=1 \\ \ell\neq i}}^{M} \Bigg\{ 2\mu_2[\boldsymbol{C}_v(n)]_{i\ell} + \mu_3[\boldsymbol{C}_v(n)]_{\ell\ell} \\ &\quad + \mu_4 \sum_{\substack{p=1 \\ p\neq\{i,\ell\}}}^{M} [\boldsymbol{C}_v(n)]_{\ell p} \Bigg\}\end{aligned} \tag{40}$$

and, for $j \neq i$,

$$\begin{aligned}[\boldsymbol{T}(n)]_{ij} &= \mu_2([\boldsymbol{C}_v(n)]_{ii} + [\boldsymbol{C}_v(n)]_{jj}) + 2\mu_3[\boldsymbol{C}_v(n)]_{ij} \\ &\quad + \sum_{\substack{\ell=1 \\ \ell\neq\{i,j\}}}^{M} \Bigg\{ 2\mu_4[\boldsymbol{C}_v(n)]_{j\ell} + 2\mu_4[\boldsymbol{C}_v(n)]_{i\ell} \\ &\quad + \mu_4[\boldsymbol{C}_v(n)]_{\ell\ell} + \mu_5 \sum_{\substack{p=1 \\ p\neq\{i,j,\ell\}}}^{M} [\boldsymbol{C}_v(n)]_{\ell p} \Bigg\}\end{aligned} \tag{41}$$

which completes the evaluation of $\boldsymbol{T}(n)$ in (30b). Substituting this result into (30a) yields the following recursive expressions for the entries of the autocorrelation matrix $\boldsymbol{C}_v(n)$:

$$\begin{aligned}[\boldsymbol{C}_v(n+1)]_{ii} &= (1 - 2\eta r_{\mathrm{md}} + \eta^2\mu_1)\,[\boldsymbol{C}_v(n)]_{ii} \\ &\quad + \eta^2\mu_3 \sum_{\substack{\ell=1 \\ \ell\neq i}}^{M} [\boldsymbol{C}_v(n)]_{\ell\ell} \\ &\quad + (2\eta^2\mu_2 - 2\eta r_{\mathrm{od}}) \sum_{\substack{\ell=1 \\ \ell\neq i}}^{M} [\boldsymbol{C}_v(n)]_{i\ell} \\ &\quad + \eta^2\mu_4 \sum_{\substack{\ell=1 \\ \ell\neq i}}^{M} \sum_{\substack{p=1 \\ p\neq\{i,\ell\}}}^{M} [\boldsymbol{C}_v(n)]_{\ell p} + \eta^2\, r_{\mathrm{md}}\, J_{\min}\end{aligned} \tag{42}$$

and, for $j \neq i$

$$\begin{aligned}&[\boldsymbol{C}_v(n+1)]_{ij} \\ &\quad = (1 - 2\eta r_{\mathrm{md}} + 2\eta^2\mu_3)\,[\boldsymbol{C}_v(n)]_{ij} \\ &\qquad + \eta^2\mu_4 \sum_{\substack{\ell=1 \\ \ell\neq\{i,j\}}}^{M} [\boldsymbol{C}_v(n)]_{\ell\ell} \\ &\qquad + (\eta^2\mu_2 - \eta r_{\mathrm{od}})([\boldsymbol{C}_v(n)]_{ii} + [\boldsymbol{C}_v(n)]_{jj}) \\ &\qquad + (2\eta^2\mu_4 - \eta r_{\mathrm{od}}) \sum_{\substack{\ell=1 \\ \ell\neq\{i,j\}}}^{M} ([\boldsymbol{C}_v(n)]_{i\ell} + [\boldsymbol{C}_v(n)]_{j\ell}) \\ &\qquad + \eta^2\mu_5 \sum_{\substack{\ell=1 \\ \ell\neq\{i,j\}}}^{M} \sum_{\substack{p=1 \\ p\neq\{i,j,\ell\}}}^{M} [\boldsymbol{C}_v(n)]_{\ell p} + \eta^2\, r_{\mathrm{od}}\, J_{\min}\end{aligned} \tag{43}$$

where $r_{\mathrm{md}} = [\boldsymbol{R}_{\kappa\kappa}]_{ii}$ and $r_{\mathrm{od}} = [\boldsymbol{R}_{\kappa\kappa}]_{ij}$ with $j \neq i$, as defined in (16).

E. Some Useful Inequalities

Before concluding this section, let us derive inequalities relating the fourth-order moments μ_i and the entries of the covariance matrix $\boldsymbol{R}_{\kappa\kappa}$. For real random variables X and Y, remember that Hölder's inequality says [15]

$$E\{|XY|\} \leq E\{|X|^p\}^{\frac{1}{p}}\, E\{|Y|^q\}^{\frac{1}{q}} \tag{44}$$

where p and q are in $(1, +\infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. As shown hereafter, this inequality yields

$$\mu_5 \overset{(d)}{\leq} \mu_4 \overset{(c)}{\leq} \mu_3 \overset{(b)}{\leq} \mu_2 \overset{(a)}{\leq} \mu_1. \tag{45}$$

Inequality (a) is obtained by using (44) with $X = \kappa^3_{\boldsymbol{\omega}_i}(n)$, $Y = \kappa_{\boldsymbol{\omega}_\ell}(n)$, $p = \frac{4}{3}$, and $q = 4$. In order to prove inequality (b), we first need to observe that Hölder's inequality yields[3]

$$E\{X^2Y^2\} \leq \sqrt{E\{|X^3\,Y|\}\,E\{|X\,Y^3|\}}. \tag{46}$$

The result (b) directly follows from (46) for $X = \kappa_{\boldsymbol{\omega}_i}(n)$ and $Y = \kappa_{\boldsymbol{\omega}_\ell}(n)$. The inequality (c) can be proved using (44) with $p = q = 2$, $X = \kappa_{\omega_i}(n)\kappa_{\omega_\ell}(n)$ and $Y = \kappa_{\omega_i}(n)\kappa_{\omega_p}(n)$,

$$\begin{aligned}&E\left\{\kappa^2_{\omega_i}(n)\kappa_{\omega_\ell}(n)\kappa_{\omega_p}(n)\right\} \\ &\quad \leq E\left\{\kappa^2_{\omega_i}(n)\kappa^2_{\omega_\ell}(n)\right\}^{1/2} E\left\{\kappa^2_{\omega_i}(n)\kappa^2_{\omega_p}(n)\right\}^{1/2} \\ &\quad = E\left\{\kappa^2_{\omega_i}(n)\kappa^2_{\omega_\ell}(n)\right\}\end{aligned}$$

where the equality is due to stationarity of $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$. Now, for $p = q = 2$, $X = \kappa_{\omega_i}(n)[\kappa_{\omega_p}(n)\kappa_{\omega_\ell}(n)]^{1/2}$ and $Y = \kappa_{\omega_j}(n)[\kappa_{\omega_p}(n)\kappa_{\omega_\ell}(n)]^{1/2}$, (44) yields

$$\begin{aligned}&E\left\{\kappa_{\omega_i}(n)\kappa_{\omega_\ell}(n)\kappa_{\omega_p}(n)\kappa_{\omega_j}(n)\right\} \\ &\quad \leq E\left\{\kappa^2_{\omega_i}(n)\kappa_{\omega_\ell}(n)\kappa_{\omega_p}(n)\right\}^{1/2} \\ &\qquad \times E\left\{\kappa^2_{\omega_j}(n)\kappa_{\omega_\ell}(n)\kappa_{\omega_p}(n)\right\}^{1/2} \\ &\quad = E\left\{\kappa^2_{\omega_i}(n)\kappa_{\omega_\ell}(n)\kappa_{\omega_p}(n)\right\}\end{aligned}$$

where the equality is again due to stationarity of $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$. This last relation proves inequality (d) and completes the proof of (45).

Finally, let us state an inequality involving the main diagonal entry r_{md} of the covariance matrix $\boldsymbol{R}_{\kappa\kappa}$. By virtue of Chebyshev's sum inequality, we have

$$r^2_{\mathrm{md}} \leq \mu_3. \tag{47}$$

[3]Replace X with $X^{3/2}Y^{1/2}$ and Y with $X^{1/2}Y^{3/2}$ in (44) with $p = q = 2$.

In the next section, we shall use the recursive expressions (42) and (43) of the autocorrelation matrix $\boldsymbol{C}_v(n)$, and (45) and (47) to study the steady-state behavior of the Gaussian KLMS algorithm.

IV. Gaussian KLMS Algorithm: Convergence Analysis

We now determine convergence conditions for the Gaussian KLMS algorithm using the analytical model derived in Section III. Let $\boldsymbol{c}_v(n)$ be the lexicographic representation of $\boldsymbol{C}_v(n)$, i.e., the matrix $\boldsymbol{C}_v(n)$ is stacked column-wise into a single vector $\boldsymbol{c}_v(n)$. Consider the family of $(M \times M)$ matrices $\boldsymbol{H}^{ij}$, $1 \le i, j \le M$, whose elements are given by (48)–(49), shown at the bottom of the page. Finally, we define the $(M^2 \times M^2)$ matrix $\boldsymbol{G}$ as follows:

$$\boldsymbol{G} = [\boldsymbol{h}^{11}\ \boldsymbol{h}^{12}\ \ldots\ \boldsymbol{h}^{1M}\ \ldots\ \boldsymbol{h}^{\mathrm{MM}}] \tag{50}$$

with $\boldsymbol{h}^{\ell p}$ the $(M^2 \times 1)$ lexicographic representation of $\boldsymbol{H}^{\ell p}$. Using these definitions, it can be shown that the lexicographic representation of the recursion (30a) can be written as

$$\boldsymbol{c}_v(n+1) = \boldsymbol{G}\,\boldsymbol{c}_v(n) + \eta^2 J_{\min}\,\boldsymbol{r}_{\kappa\kappa} \tag{51}$$

where $\boldsymbol{r}_{\kappa\kappa}$ is the lexicographic representation of $\boldsymbol{R}_{\kappa\kappa}$. We shall use this expression to derive stability conditions for the Gaussian KLMS algorithm. But before closing this subsection, let us remark that matrix $\boldsymbol{G}$ is symmetric. This can be shown from (48)–(49), using $[\boldsymbol{H}^{ij}]_{\ell p} = [\boldsymbol{G}]_{(i-1)M+j,(\ell-1)M+p}$, and observing that $[\boldsymbol{H}^{ij}]_{\ell p} = [\boldsymbol{H}^{\ell p}]_{ij}$. This implies that $\boldsymbol{G}$ can be diagonalized by a unitary transformation, and all its eigenvalues are real-valued.

A. Convergence Conditions

A necessary and sufficient condition for convergence of $\boldsymbol{c}_v(n)$ in (51) is that all the eigenvalues of $\boldsymbol{G}$ lie inside the open interval $(-1,\, 1)$ ([16], Section 5.9). Thus, the stability limit for η can be numerically determined for given values of M and ξ. In the following we derive a set of analytical sufficient conditions that can also be used for design purposes.

It is well known that all the eigenvalues of $\boldsymbol{G}$ lie inside the union of Gerschgorin disks [17]. Each of these disks is centered at a diagonal element of $\boldsymbol{G}$ and has a radius given by the sum of the absolute values of the remaining elements of the same row. A sufficient condition for stability of (51) is thus given by

$$|[\boldsymbol{G}]_{ii}| + \sum_{\substack{\ell=1 \\ \ell \ne i}}^{M^2} |[\boldsymbol{G}]_{i\ell}| < 1, \quad \text{for } i = 1, \ldots, M^2. \tag{52}$$

Equations (48)–(50) show that the rows of $\boldsymbol{G}$ have only two distinct forms, in the sense that each row of $\boldsymbol{G}$ has the same entries as one of these two distinct rows, up to a permutation. This implies that only two distinct Gerschgorin disks can be defined. Also, all the entries of $\boldsymbol{G}$ are positive except possibly for $[\boldsymbol{G}]_{i\ell} = \eta^2\,\mu_2 - \eta\,r_{\mathrm{od}}$ and $[\boldsymbol{G}]_{i\ell} = \frac{1}{2}(2\,\eta^2\,\mu_4 - \eta\,r_{\mathrm{od}})$. See Appendix B for proof. Expression (52) thus leads to only two inequalities, defined as follows for $M \ge 3$,

$$\begin{aligned}
&(1 - 2\eta\,r_{\mathrm{md}} + \eta^2\,\mu_1) + (M-1)\,\eta^2\,\mu_3 \\
&\quad + 2\,(M-1)\,|\eta^2\,\mu_2 - \eta\,r_{\mathrm{od}}| \\
&\quad + (M-1)(M-2)\,\eta^2\mu_4 < 1 \qquad \text{(53a)} \\
&(1 - 2\,\eta\,r_{\mathrm{md}} + 2\,\eta^2\,\mu_3) \\
&\quad + 2|\eta^2\,\mu_2 - \eta\,r_{\mathrm{od}}| + (M-2)\,\eta^2\,\mu_4 \\
&\quad + 2\,(M-2)\,|2\,\eta^2\,\mu_4 - \eta\,r_{\mathrm{od}}| \\
&\quad + (M-2)(M-3)\,\eta^2\,\mu_5 < 1. \qquad \text{(53b)}
\end{aligned}$$

The study of the limiting values of η that satisfy both (53a) and (53b) yields bounds on the maximum step size. The intersection of these bounds yields the sufficient stability limits. In what follows, we present the analysis for $M \ge 3$. The results for $M = 1$ and $M = 2$ are presented in Appendix C.

To determine the bounds imposed by (53a), we write it for $\eta > 0$ as

$$\begin{aligned}
2\,(M-1)\,|\eta\,\mu_2 - r_{\mathrm{od}}| < {}& 2\,r_{\mathrm{md}} \\
& -\eta\,\mu_1 - (M-1)\,\eta\,\mu_3 - (M-1)(M-2)\,\eta\,\mu_4. \qquad \text{(54)}
\end{aligned}$$

Thus, the following two conditions must be satisfied

$$\begin{aligned}
2\,(M-1)\,(\eta\,\mu_2 - r_{\mathrm{od}}) < {}& 2\,r_{\mathrm{md}} - \eta\,\mu_1 - (M-1)\,\eta\,\mu_3 \\
& - (M-1)(M-2)\,\eta\,\mu_4 \qquad \text{(55a)} \\
2\,(M-1)\,(\eta\,\mu_2 - r_{\mathrm{od}}) > {}& -2\,r_{\mathrm{md}} + \eta\,\mu_1 + (M-1)\,\eta\,\mu_3 \\
& + (M-1)(M-2)\,\eta\,\mu_4. \qquad \text{(55b)}
\end{aligned}$$

$$\text{if } (i=j): \begin{cases}
[\boldsymbol{H}^{ii}]_{ii} = 1 - 2\eta r_{\mathrm{md}} + \eta^2\mu_1, & \\
[\boldsymbol{H}^{ii}]_{pp} = \eta^2\mu_3, & p \ne i \\
[\boldsymbol{H}^{ii}]_{ip} = \eta^2\mu_2 - \eta r_{\mathrm{od}} = [\boldsymbol{H}^{ii}]_{pi}, & p \ne i \\
[\boldsymbol{H}^{ii}]_{p\ell} = \eta^2\mu_4, & \text{otherwise,}
\end{cases} \tag{48}$$

$$\text{if } (i \ne j): \begin{cases}
[\boldsymbol{H}^{ij}]_{ij} = [\boldsymbol{H}^{ij}]_{ji} = \frac{1}{2}(1 - 2\eta r_{\mathrm{md}} + 2\eta^2\mu_3) & \\
[\boldsymbol{H}^{ij}]_{pp} = \eta^2\mu_4, & p \ne i,j \\
[\boldsymbol{H}^{ij}]_{ii} = [\boldsymbol{H}^{ij}]_{jj} = \eta^2\mu_2 - \eta r_{\mathrm{od}}, & \\
[\boldsymbol{H}^{ij}]_{ip} = [\boldsymbol{H}^{ij}]_{pi} = \frac{1}{2}(2\eta^2\mu_4 - \eta r_{\mathrm{od}}), & p \ne i,j \\
[\boldsymbol{H}^{ij}]_{pj} = [\boldsymbol{H}^{ij}]_{jp} = \frac{1}{2}(2\eta^2\mu_4 - \eta r_{\mathrm{od}}), & p \ne i,j \\
[\boldsymbol{H}^{ij}]_{p\ell} = \eta^2\mu_5, & \text{otherwise.}
\end{cases} \tag{49}$$

TABLE I
CONDITIONS ON η DERIVED FROM (55B)

Case	Sign of θ_1	Sign of θ_2	Condition	Observation
(i)	+	+	$0 \leq \eta < \eta_2$	
(ii)	+	–	$0 < \eta$	No additional restriction
(iii)	–	+	$\eta < \eta_2 < 0$	No $\eta > 0$ satisfies the condition
(iv)	–	–	$0 < \eta_2 < \eta$	

TABLE II
CONDITIONS ON η DERIVED FROM (62B)

Case	Sign of θ_1	Sign of θ_4	Condition	Observation
(v)	+	+	$0 \leq \eta < \eta_4$	
(vi)	+	–	$0 < \eta$	No additional restriction
(vii)	–	+	$\eta < \eta_4 < 0$	No $\eta > 0$ satisfies the condition
$(viii)$	–	–	$0 < \eta_4 < \eta$	

Condition (55a) yields (56), shown at the bottom of the page, because the numerator and denominator of η_1 are positive. On the other hand, (55b) leads us to a condition of the form $\theta_2\,\eta < \theta_1$ with

$$\theta_1 = 2\,r_{\rm md} - 2\,(M-1)\,r_{\rm od} \tag{57}$$

$$\theta_2 = \mu_1 - 2\,(M-1)\,\mu_2 + (M-1)\,\mu_3 + (M-1)(M-2)\,\mu_4. \tag{58}$$

Hereafter, we shall denote by $\eta_2 := \frac{\theta_1}{\theta_2}$ the resulting bound on η. Then, solving (55b) yields the four possible cases shown in Table I, depending on the signs of θ_1 and θ_2. In case (ii), note that there exists no $\eta > 0$ that satisfies (55b) when $\theta_1 < 0$ and $\theta_2 > 0$. This situation arises because condition (53a) is a sufficient condition imposing that the Gerschgorin disks defined by (52) are completely inside the unit circle in the z-plane. This condition is obviously not necessary for having all the eigenvalues of $\boldsymbol{G}$ inside the unit circle. The lower bound in case (iv) of Table I is also due to this excessive restriction, as the algorithm is certainly stable for $\eta = 0$. It is kept here for completeness, but it should be disregarded in practice.

Combining the possible solutions of (55a) and (55b) and disregarding lower bounds yields the following stability bounds for η

$$\begin{cases} \eta < \min\{\eta_1, \eta_2\}, & \text{for case (i) in Table I} \\ \eta < \eta_1, & \text{for cases (ii) and } (iv) \text{ in Table I.} \end{cases} \tag{59}$$

If case (iii) happens, which should be tested right away, the stability conditions should be determined numerically through the eigenvalues of $\boldsymbol{G}$.

Having determined limits for (53a) such that $\eta > 0$, and assuming that case (iii) in Table I did not happen, we proceed to determine the extra restrictions imposed on η by (53b). First, we multiply (53b) by $(M-1)$ and divide by η, rewriting it as

$$\begin{aligned} &2(M-1)\,|\mu_2\,\eta - r_{\rm od}| \\ &\quad < 2(M-1)\,r_{\rm md} - 2(M-1)\,\mu_3\,\eta \\ &\qquad - (M-1)(M-2)\,\mu_4\,\eta \\ &\qquad - 2\,(M-1)(M-2)\,|2\,\mu_4\,\eta - r_{\rm od}| \\ &\qquad - (M-1)(M-2)(M-3)\,\eta\,\mu_5. \end{aligned} \tag{60}$$

Now, given that (54) has already been satisfied, we replace the left-hand side (LHS) of (60) with the right-hand side (RHS) of (54). After rearranging the terms, we have the new condition

$$\begin{aligned} 2(M-1)(M-2)\,|2\,\mu_4\,\eta - r_{\rm od}| &< 2(M-2)\,r_{\rm md} + \mu_1\,\eta \\ &\quad -(M-1)\,\mu_3\,\eta - (M-1)(M-2)(M-3)\,\mu_5\,\eta \end{aligned} \tag{61}$$

which leads to the following two conditions:

$$\begin{aligned} &4(M-1)(M-2)\,\mu_4\,\eta - 2(M-1)(M-2)\,r_{\rm od} \\ &\quad < 2(M-2)\,r_{\rm md} + \mu_1\,\eta \\ &\qquad - (M-1)\,\mu_3\,\eta - (M-1)(M-2)(M-3)\,\mu_5\,\eta \end{aligned} \tag{62a}$$

$$\begin{aligned} &-2(M-2)\,r_{\rm md} - \mu_1\,\eta + (M-1)\,\mu_3\,\eta \\ &\quad + (M-1)(M-2)(M-3)\,\mu_5\,\eta \\ &\qquad < 4(M-1)(M-2)\,\mu_4\,\eta - 2(M-1)(M-2)\,r_{\rm od}. \end{aligned} \tag{62b}$$

On the one hand, (62a) yields two different conditions defined by

$$\begin{cases} \eta < \dfrac{2(M-2)(r_{\rm md} + (M-1)\,r_{\rm od})}{\theta_3} := \eta_3, & \text{if } \theta_3 > 0 \\ \eta > 0, & \text{if } \theta_3 < 0 \end{cases} \tag{63}$$

where

$$\begin{aligned} \theta_3 = -\mu_1 + (M-1)\,\mu_3 + 4(M-1)(M-2)\,\mu_4 \\ +(M-1)(M-2)(M-3)\,\mu_5. \end{aligned} \tag{64}$$

On the other hand, (62b) leads us to a condition of the form $\theta_4\,\eta < (M-2)\,\theta_1$, where θ_1 has already been defined in (57) and

$$\begin{aligned} \theta_4 = -\mu_1 + (M-1)\,\mu_3 - 4(M-1)(M-2)\,\mu_4 \\ +(M-1)(M-2)(M-3)\,\mu_5. \end{aligned} \tag{65}$$

Solving (62b) thus leaves four possible cases to consider, shown in Table II, where $\eta_4 := \frac{(M-2)\,\theta_1}{\theta_4}$.

$$\eta < \frac{2\,r_{\rm md} + 2\,(M-1)\,r_{\rm od}}{\mu_1 + 2\,(M-1)\,\mu_2 + (M-1)\,\mu_3 + (M-1)(M-2)\,\mu_4} := \eta_1 \tag{56}$$

Combing the possible solutions of (62a) and (62b), and again disregarding lower bounds, yields the following stability bounds on η [see (66) at the bottom of the page].

Finally, except for cases (*iii*) in Table I and (*vii*) in Table II, which should be tested right away, the sufficient stability conditions will be given by the intersection of (59) and (66). These conditions can be slightly simplified by observing that $\eta_4 \leq \eta_1$, as can be easily proved using (45).

In the next section, we shall derive the expression of the weight-error correlation matrix $\boldsymbol{C}(n)$ in steady state. This will allow us to calculate the MSE and the excess MSE.

B. Steady-State Behavior

The closed-form solution of (51) can be written as [16]

$$\boldsymbol{c}_v(n) = \boldsymbol{G}^n \left[\boldsymbol{c}_v(0) - \boldsymbol{c}_v(\infty)\right] + \boldsymbol{c}_v(\infty) \tag{67}$$

where $\boldsymbol{c}_v(\infty)$ denotes the vector $\boldsymbol{c}_v(n)$ in steady state, and is given by

$$\boldsymbol{c}_v(\infty) = \eta^2 \, J_{\min} \, (\boldsymbol{I} - \boldsymbol{G})^{-1} \, \boldsymbol{r}_{\kappa\kappa}. \tag{68}$$

Assuming convergence, we define the time for convergence as the number n_ϵ of iterations required for (67) to reach

$$\|\boldsymbol{c}_v(n) - \boldsymbol{c}_v(n_\epsilon)\| \leq \epsilon \tag{69}$$

where ϵ is a design parameter to be chosen by the user.

At this point, it is important to note that $\boldsymbol{c}_v(\infty)$ is unique if the system under consideration is stable. Indeed, the matrix $(\boldsymbol{G} - \boldsymbol{I})$ has only nonzero eigenvalues because it satisfies conditions (53a)–(53b), and can thus be inverted. Let $\boldsymbol{C}_v(\infty)$ be the matrix whose lexicographic representation is $\boldsymbol{c}_v(\infty)$. It is also unique and satisfies the following expression derived from (30a) for $n \to \infty$:

$$\boldsymbol{R}_{\kappa\kappa} \, \boldsymbol{C}_v(\infty) + \boldsymbol{C}_v(\infty) \, \boldsymbol{R}_{\kappa\kappa} - \eta \, \boldsymbol{T}(\infty) = \eta \, \boldsymbol{R}_{\kappa\kappa} \, J_{\min}. \tag{70}$$

In (70), $\boldsymbol{T}(\infty)$ denotes the matrix $\boldsymbol{T}(n)$ in steady state. It is interesting to note that (68) is the lexicographic counterpart of (70), as (51) is the lexicographic representation of (30a).

From (23), the steady-state MSE is given by

$$J_{ms}(\infty) = J_{\min} + \mathrm{trace}\{\boldsymbol{R}_{\kappa\kappa} \, \boldsymbol{C}_v(\infty)\} \tag{71}$$

where $\mathrm{trace}\{\boldsymbol{R}_{\kappa\kappa} \, \boldsymbol{C}_v(\infty)\}$ is the steady-state excess MSE, denoted by $J_{ex}(\infty)$. To evaluate these quantities, we shall now compute the entries of $\boldsymbol{C}_v(\infty)$. In order to achieve this, we first justify that $\boldsymbol{C}_v(\infty)$ is a matrix with the same structural properties as $\boldsymbol{R}_{\kappa\kappa}$, namely, all its main diagonal entries are equal to each other, having a value denoted by c_{md}, and all its off-diagonal entries are also equal to each other, having a value denoted by c_{od}. Then, we determine c_{md} and c_{od} so that $\boldsymbol{C}_v(\infty)$ is the solution to (70), which we know is unique. It is straightforward to see that the LHS of (70) is also a matrix with the same structural properties as $\boldsymbol{R}_{\kappa\kappa}$, as it is equal to $\eta \, \boldsymbol{R}_{\kappa\kappa} \, J_{\min}$.

One way for $\boldsymbol{C}_v(\infty)$ to have the proposed structure would be that $\boldsymbol{R}_{\kappa\kappa} \, \boldsymbol{C}_v(\infty)$, $\boldsymbol{C}_v(\infty) \, \boldsymbol{R}_{\kappa\kappa}$ and $\boldsymbol{T}(\infty)$ all have the same structure. It is straightforward to show that matrices $\boldsymbol{R}_{\kappa\kappa} \, \boldsymbol{C}_v(\infty)$ and $\boldsymbol{C}_v(\infty) \, \boldsymbol{R}_{\kappa\kappa}$ have this structure if, and only if, $\boldsymbol{C}_v(\infty)$ has that structure. If this is the case, a direct consequence, through (70), would be that $\boldsymbol{T}(\infty)$ also has this structure. Using c_{md} and c_{od}, respectively, for the main-diagonal and off-diagonal entries of $\boldsymbol{C}_v(\infty)$ in (40) and (41) yields, for $M > 1$ [see (72) at the bottom of the page]. For $M = 1$, $\boldsymbol{T}(\infty) = \mu_1 \, c_{\mathrm{md}}$. Writing, for ease of notation, $[\boldsymbol{T}(\infty)]_{ii} = t_1 \, c_{\mathrm{md}} + t_2 \, c_{\mathrm{od}}$ and $[\boldsymbol{T}(\infty)]_{ij} = t_3 \, c_{\mathrm{md}} + t_4 \, c_{\mathrm{od}}$ for all $i \neq j$, and solving (70) for c_{md} and c_{od} yields (73)-(74), shown at the bottom of the page, with r_{md} and r_{od} defined in (16).

It can be verified that using (72)–(74) in the LHS of (70) yields the correct RHS. Also, we know that this solution is unique since $(\boldsymbol{I} - \boldsymbol{G})$ can be inverted. Then, going back to (71), we obtain the following desired result

$$\begin{aligned} J_{ms}(\infty) &= J_{\min} + J_{ex}(\infty) \\ &= J_{\min} + M(r_{\mathrm{md}} c_{\mathrm{md}} + (M-1) r_{\mathrm{od}} c_{\mathrm{od}}). \end{aligned} \tag{75}$$

Based on the results presented in the previous sections, we shall now propose design guidelines to set the parameters of

$$\begin{aligned} &\text{If } \theta_3 > 0, \begin{cases} \eta < \min\{\eta_3, \eta_4\}, & \text{for case } (v) \text{ in Table II} \\ \eta < \eta_3, & \text{for cases } (vi) \text{ and } (viii) \text{ in Table II} \end{cases} \\ &\text{If } \theta_3 < 0, \begin{cases} \eta < \eta_4, & \text{for case } (v) \text{ in Table II} \\ \eta > 0, & \text{for cases } (vi) \text{ and } (viii) \text{ in Table II.} \end{cases} \end{aligned} \tag{66}$$

$$[\boldsymbol{T}(\infty)]_{ij} = \begin{cases} (\mu_1 + (M-1)\,\mu_3) c_{\mathrm{md}} + (2(M-1)\,\mu_2 + (M-1)(M-2)\,\mu_4) c_{\mathrm{od}}, & i = j \\ (2\,\mu_2 + (M-2)\,\mu_4) c_{\mathrm{md}} + \big(4(M-2)\,\mu_4 + 2\,\mu_3 + (M-3)(M-2)\,\mu_5\big) c_{\mathrm{od}}, & i \neq j. \end{cases} \tag{72}$$

$$c_{\mathrm{md}} = \frac{\eta J_{\min}}{2r_{\mathrm{md}} - \eta t_1} \left(\frac{((2r_{\mathrm{md}} - \eta t_1) r_{\mathrm{od}} - (2r_{\mathrm{od}} - \eta t_3) r_{\mathrm{md}})(\eta t_2 - 2(M-1) r_{\mathrm{od}})}{(\eta t_2 - 2(M-1) r_{\mathrm{od}})(2r_{\mathrm{od}} - \eta t_3) - (\eta t_4 - 2(r_{\mathrm{md}} + (M-2) r_{\mathrm{od}}))(2r_{\mathrm{md}} - \eta t_1)} + r_{\mathrm{md}} \right) \tag{73}$$

and

$$c_{\mathrm{od}} = \frac{\eta J_{\min}((2r_{\mathrm{md}} - \eta t_1) r_{\mathrm{od}} - (2r_{\mathrm{od}} - \eta t_3) r_{\mathrm{md}})}{(\eta t_2 - 2(M-1) r_{\mathrm{od}})(2r_{\mathrm{od}} - \eta t_3) - (\eta t_4 - 2(r_{\mathrm{md}} + (M-2) r_{\mathrm{od}}))(2r_{\mathrm{md}} - \eta t_1)} \tag{74}$$

the Gaussian KLMS algorithm in order to achieve a prescribed performance.

V. Design Guidelines

The analysis results are now used to establish design guidelines. Without limiting generality, the coherence-based sparsification rule (3) is considered hereafter to design the dictionaries. It is however implicit that any other existing technique to select the kernel functions $\kappa(\cdot, \boldsymbol{u}(\omega_j))$ could be used. These design guidelines assume the availability of multiple realizations of the sequences $\boldsymbol{u}(n)$ and $d(n)$.

Suppose a design goal is to obtain an MSE which is less than a specified value $J_{\max}$. The following procedure could be applied.

1) Set a coherence threshold ε_0 and define a set of kernel bandwidths ξ to be tested. For the following design examples, a set of values for ξ was chosen as equally spaced points in (0, 1). Then, ε_0 was chosen to yield reasonable values of M for the chosen set of ξ values. The value of M is determined for each pair (ξ, ε_0) by training the dictionary with the input signal $\boldsymbol{u}(n)$ until its size stabilizes. The training is repeated several times. A value M_i is determined for the ith realization. The value of M associated with the pair (ξ, ε_0) is the average of the dictionary sizes for all realizations, rounded to the nearest integer. This is the value of M to be used in the theoretical model.
2) Using the system input $\boldsymbol{u}(n)$, determine the desired output $d(n)$ and estimate $E\{d^2(n)\}$ and $\boldsymbol{p}_{\kappa d}$ over several runs. In a practical setting $d(n)$ should be measured at the unknown system output.
3) From (10), determine the minimum MSE $J_{\min}$ for each set of parameters (M, ξ). If none of the pairs (M, ξ) leads to $J_{\max} > J_{\min}$, return to step 1 to choose a new set of parameters ε_0 and ξ.
4) Determine $\eta_{\max}$ using rules (59) and (66) for each value of ξ and choose the largest possible value of η so that $\eta < \eta_{\max}$ and $J_{ms}(\infty) < J_{\max}$ (note that $J_{ms}(\infty)$ is computed using (75)).
5) Given M, $\boldsymbol{R}_{\kappa\kappa}$, $\boldsymbol{p}_{\kappa d}$, $J_{\min}$ and η determine $\boldsymbol{G}$ using (50) and then $\boldsymbol{c}_v(n)$ from (51).
6) Determine $\boldsymbol{C}_v(\infty)$ from (73)–(74), and then $J_{ms}(\infty)$ from (75).
7) Choose ϵ in criterion (69) and find n_ϵ from simulations using the parameters determined in steps 1 to 8.

Repeat steps 1 to 9 for all kernel parameters in the chosen set and use the one that leads to the smallest n_ϵ.

VI. Simulation Results

This section presents examples to illustrate the proposed design procedure and to verify the theoretical results. The simulation conditions which are common to all examples are described in Example 1. Only the changes to these conditions are described in the remaining examples.

A. Example 1

Consider the problem studied in [18] and [19], for which

$$\begin{cases} y(n) = \dfrac{y(n-1)}{1+y^2(n-1)} + u^3(n-1) \\ d(n) = y(n) + z(n) \end{cases} \tag{76}$$

TABLE III
Stability Results for Example 1

ξ	M	$\eta_{\max}^{EG}$	$\eta_{\max}^{GD}$
0.0075	17	1.70	0.29
0.01	13	1.70	0.30
0.025	6	1.66	0.22
0.05	3	1.80	1.47

TABLE IV
Summary of Simulation Results for Example 1

ξ	M	η	$J_{\min}$ [dB]	$J_{ms}(\infty)$ [dB]	$J_{ex}(\infty)$ [dB]	n_ϵ
0.0075	17	0.143	-22.19	-22.04	-36.85	1271
0.01	13	0.152	-21.84	-21.69	-36.27	914
0.025	6	0.007	-21.53	-21.52	-49.15	7746
0.05	3	0.011	-21.52	-21.51	-47.00	2648

where the output signal $y(n)$ was corrupted by a zero-mean white Gaussian noise $z(n)$ with variance $\sigma_z^2 = 10^{-4}$. The input sequence $u(n)$ is zero-mean i.i.d. Gaussian with standard deviation $\sigma_u = 0.15$.

The proposed method was tested with a maximum MSE $J_{\max} = -21.5$ dB, a coherence level $\varepsilon_0 = 10^{-5}$ and a set of kernel bandwidths $\xi \in \{0.0075, 0.01, 0.025, 0.05\}$[4]. For each value of ξ, 500 dictionary dimensions $M_i, i = 1, \ldots, 500$, were determined using 500 realizations of the input process. The length of each realization was 500 samples. Each M_i was determined as the minimum dictionary length required to achieve the coherence level ε_0. The value $M(\xi)$ was determined as the average of all M_i, rounded to the nearest integer. The values of $J_{\min}(\xi)$ were calculated from (10) for each pair (ξ, M). To this end, second-order moments $\boldsymbol{p}_{\kappa d}$ and $E\{d^2(n)\}$ were estimated by averaging over 500 runs.

Before searching for a value of η, such that $J_{ms}(\infty) < J_{\max}$, we have to define $\eta_{\max}$. There are two possibilities for defining $\eta_{\max}$: 1) use the Gerschgorin disk analysis yielding a value of $\eta_{\max}$ denoted as $\eta_{\max}^{GD}$, 2) compute and test the eigenvalues of $\boldsymbol{G}$ yielding a value of $\eta_{\max}$ denoted as $\eta_{\max}^{EG}$. Table III shows that, as expected, the condition imposed by the Gerschgorin disks is more restrictive than that imposed by the eigenvalues of $\boldsymbol{G}$. However, note that choosing η from $\eta_{\max} = \eta_{\max}^{GD}$ is simpler and usually yields good design results.

Table IV presents the obtained results for the chosen values of ξ. For each pair (ξ, M), the step-size η was chosen so that the algorithm was stable (η less than $\eta_{\max}$) and $J_{ms}(\infty) < -21.5$ dB. The values of $J_{ms}(\infty)$ and $J_{ex}(\infty)$ were determined from (75) and n_ϵ was obtained from (69) for $\epsilon = 10^{-3}$. Note that $J_{\min}(\xi) < -21.5$ dB in all cases. It clearly appears that $(M, \xi, \eta) = (13, 0.01, 0.152)$ is a good design choice, as it satisfies all design conditions after only $n_\epsilon = 914$ iterations. Note that the value of η chosen is about 1/10 of $\eta_{\max}$. This is due to the small value of $J_{\max}$ imposed by the design. This is the same phenomenon that happens when designing the regular LMS algorithm for practical specifications in linear estimation problems.

For each simulation, the order M of the dictionary remained fixed. It was initialized for each realization by generating input

[4]These values of ξ are samples within a range of values experimentally verified to be adequate for the application.

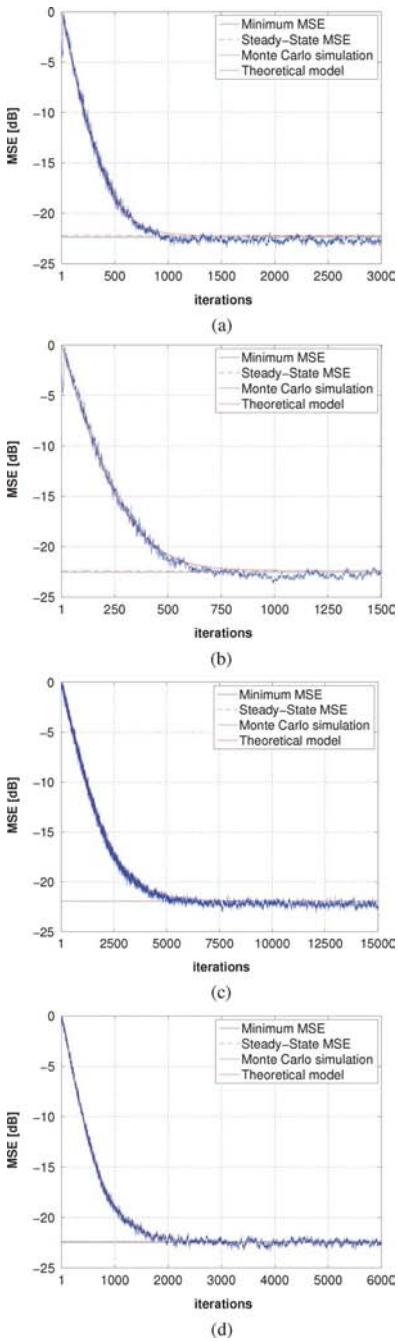


Fig. 2. Theoretical model and Monte Carlo simulation of KLMS for different kernel bandwidths. Ragged curves (blue): simulation results averaged over 500 runs. Continuous curves (red): Theory using (23) and (67). Continuous horizontal lines (blue): Steady-state MSE predicted by theory. Dashed horizontal lines (red): Steady-state MSE from simulations. (a) $\xi = 0.0075$ and $M = 17$. (b) $\xi = 0.01$ and $M = 13$. (c) $\xi = 0.025$ and $M = 6$. (d) $\xi = 0.05$ and $M = 3$.

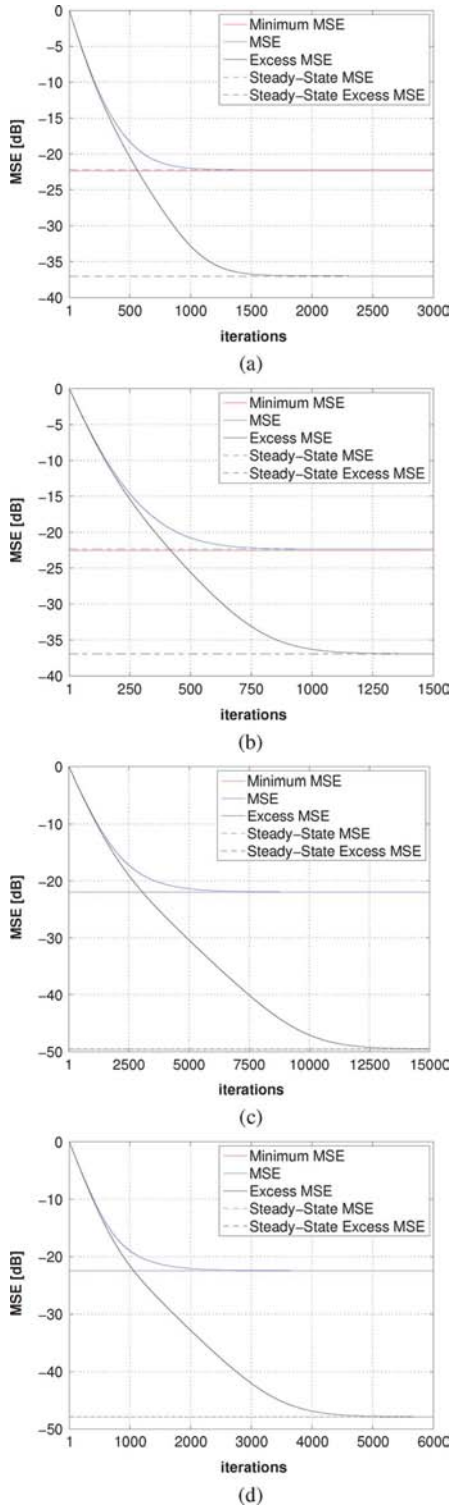


Fig. 3. Steady-State results. Dashed horizontal lines: MSE and Excess MSE averaged over 500 realizations. Continuous horizontal lines (red): Minimum MSE predicted by theory. Continuous decaying lines: Theoretical MSE (blue) and Excess MSE (black). (a) $\xi = 0.0075$ and $M = 17$. (b) $\xi = 0.01$ and $M = 13$. (c) $\xi = 0.025$ and $M = 6$. (d) $\xi = 0.05$ and $M = 3$.

vectors in $\mathcal{U}$ and filling the M positions with vectors that satisfy the desired coherence level. Thus, the initial dictionary is different for each realization. During each realization, the dictionary elements were updated at each iteration n so that the least recently added element is replaced with $\boldsymbol{u}(n-1)$.

Figs. 2 and 3 illustrate the accuracy of the analytical model for the four cases presented in Table IV. Fig. 2 shows an excellent agreement between Monte Carlo simulations, averaged over 500 runs, and the theoretical predictions made by using (23) and (67). Fig. 3 compares simulated steady-state results (dashed horizontal lines) with theoretical predictions using (68). There is again an excellent agreement between theory and simulations.

TABLE V
STABILITY RESULTS FOR EXAMPLE 2

ξ	M	$\eta_{\max}^{EG}$	$\eta_{\max}^{GD}$
0.05	7	2.33	*w.s*
0.065	4	2.50	0.68
0.075	3	2.60	1.92
0.125	2	2.34	2.32

TABLE VI
SUMMARY OF SIMULATION RESULTS FOR EXAMPLE 2

ξ	M	η	$J_{\min}$ [dB]	$J_{ms}(\infty)$ [dB]	$J_{ex}(\infty)$ [dB]	n_ϵ
0.05	7	0.072	-20.32	-20.27	-39.73	1544
0.065	4	0.239	-20.41	-20.25	-34.63	314
0.075	3	0.340	-20.68	-20.45	-33.33	184
0.125	2	0.073	-20.54	-20.46	-38.39	502

B. Example 2

As a second design example, consider the nonlinear dynamic system identification problem studied in [20] where the input signal was a sequence of statistically independent vectors

$$\boldsymbol{u}(n) = [u_1(n)\ u_2(n)]^\top \tag{77}$$

with correlated samples satisfying $u_1(n) = 0.5u_2(n) + \eta_u(n)$. The second component of $\boldsymbol{u}(n)$ in (77) is an i.i.d. Gaussian noise sequence with variance $\sigma_{u_2}^2 = 0.0156$ and $\eta_u(n)$ is a white Gaussian noise with variance $\sigma_{u_1}^2 = 0.0156$. Consider a linear system with memory defined by

$$y(n) = \boldsymbol{a}^\top \boldsymbol{u}(n) - 0.2\, y(n-1) + 0.35\, y(n-2) \tag{78}$$

where $\boldsymbol{a} = [1\ 0.5]^\top$ and a nonlinear Wiener function

$$\varphi(y(n)) = \begin{cases} \dfrac{y(n)}{3[0.1 + 0.9\, y^2(n)]^{1/2}} & \text{for } y(n) \geq 0 \\ \dfrac{-y^2(n)[1 - \exp(0.7y(n))]}{3} & \text{for } y(n) < 0 \end{cases} \tag{79}$$

$$d(n) = \varphi(y(n)) + z(n) \tag{80}$$

where $d(n)$ is the output signal, corrupted by a zero-mean white Gaussian noise $z(n)$ with variance $\sigma_z^2 = 10^{-6}$. The initial condition $y(1) = 0$ was considered in this example.

The proposed method was tested with a maximum MSE $J_{\max} = -20.25$ dB, a coherence level $\varepsilon_0 = 10^{-4}$ and a set of kernel bandwidths $\xi \in \{0.05,\ 0.065,\ 0.075,\ 0.125\}$. The other simulation conditions were similar to the first example.

Table V shows the estimated values of $\eta_{\max}$ obtained with the Gerschgorin disk analysis and from the eigenvalues of $\boldsymbol{G}$. The expression without solution (w.s.) indicates that the intersection of solutions provided by (59) and (66) is empty. In this case, we need to use the limit obtained from the eigenvalues of $\boldsymbol{G}$, i.e., $\eta_{\max} = \eta_{\max}^{EG}$.

Table VI presents the obtained results for the chosen values of ξ. For each pair (ξ, M), the step-size η was chosen in order to ensure algorithm stability (η less than $\eta_{\max}$) and $J_{ms}(\infty) < -20.25$ dB. The values of $J_{ms}(\infty)$ and $J_{ex}(\infty)$ were determined from (75) and n_ϵ was obtained from (69) for $\epsilon = 10^{-3}$. Note that $J_{\min}(\xi) < -20.25$ dB in all cases. It clearly appears that $(M, \xi, \eta) = (3,\ 0.075,\ 0.340)$ is a good design choice as it satisfies all design conditions after $n_\epsilon = 184$ iterations.

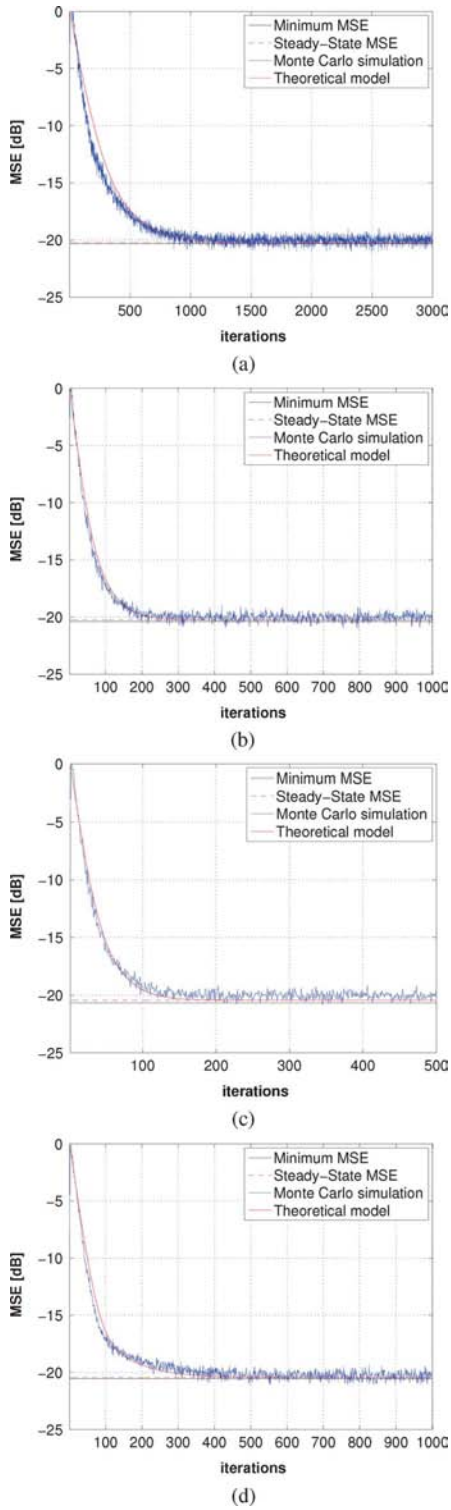


Fig. 4. Theoretical model and Monte Carlo simulation of KLMS for different kernel bandwidths. Ragged curves (blue): simulation results averaged over 500 runs. Continuous curves (red): Theory using (23) and (67). Continuous horizontal lines (blue): Steady-state MSE predicted by theory. Dashed horizontal lines (red): Steady-state MSE from simulations. (a) $\xi = 0.05$ and $M = 7$. (b) $\xi = 0.065$ and $M = 4$. (c) $\xi = 0.075$ and $M = 3$. (d) $\xi = 0.125$ and $M = 2$.

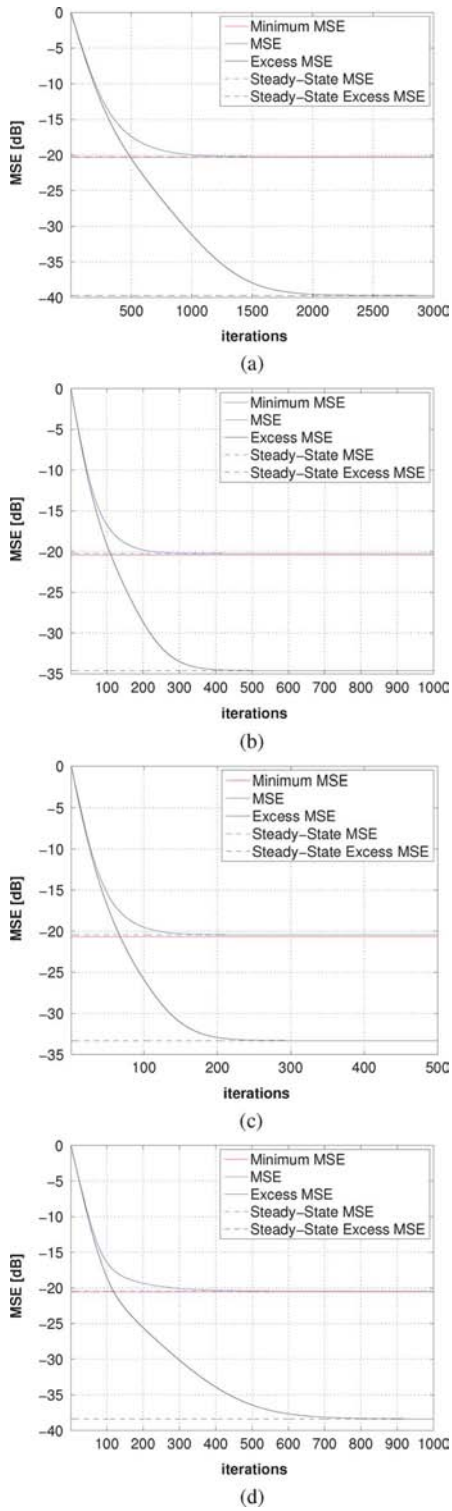


Fig. 5. Steady-State results. Dashed horizontal lines: MSE and Excess MSE averaged over 500 realizations. Continuous horizontal lines (red): Minimum MSE predicted by theory. Continuous decaying lines: Theoretical MSE (blue) and Excess MSE (black). (a) $\xi = 0.05$ and $M = 7$. (b) $\xi = 0.065$ and $M = 4$. (c) $\xi = 0.075$ and $M = 3$. (d) $\xi = 0.125$ and $M = 2$.

Figs. 4 and 5 illustrate the accuracy of the analytical model for the four cases presented in Table VI. The agreement between theory and simulations is excellent as in the first example.

TABLE VII
STABILITY RESULTS FOR EXAMPLE 3

ξ	M	$\eta_{\max}^{EG}$	$\eta_{\max}^{GD}$
0.15	11	1.17	*w.s.*
0.20	7	1.19	*w.s.*
0.25	5	1.24	*w.s.*
0.30	3	1.60	*w.s.*

TABLE VIII
SUMMARY OF SIMULATION RESULTS FOR EXAMPLE 3

ξ	M	η	$J_{\min}$ [dB]	$J_{ms}(\infty)$ [dB]	$J_{ex}(\infty)$ [dB]	n_ϵ
0.15	11	0.147	-20.58	-20.20	-31.23	517
0.20	7	0.149	-20.41	-20.03	-30.81	382
0.25	5	0.039	-20.07	-19.97	-36.67	1298
0.30	3	0.013	-19.78	-19.75	-42.48	4052

C. Example 3

As a third design example, consider the fluid-flow control problem studied in [21] and [22] whose input signal was a sequence of statistically independent vectors

$$\boldsymbol{u}(n) = [u_1(n)\; u_2(n)]^\top \tag{81}$$

with correlated samples satisfying $u_1(n) = 0.5\, u_2(n) + \eta_u(n)$, where $u_2(n)$ is a white Gaussian noise sequence with variance $\sigma_{u_2}^2 = 0.0625$ and $\eta_u(n)$ is a white Gaussian so that $u_1(n)$ has variance $\sigma_{u_1}^2 = 0.0625$. Consider a linear system with memory defined by

$$y(n) = \boldsymbol{a}^\top \boldsymbol{u}(n) + 1.4138\, y(n-1) - 0.6065\, y(n-2) \tag{82}$$

with $\boldsymbol{a} = [0.1044\; 0.0883]^\top$ and a nonlinear Wiener function

$$d(n) = \frac{0.3163\, y(n)}{\sqrt{0.10 + 0.90\, y^2(n)}} + z(n) \tag{83}$$

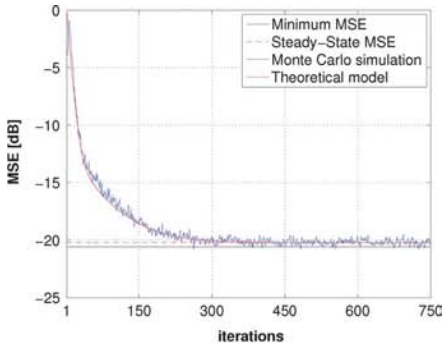
where the output signal $d(n)$ was corrupted by a zero-mean white Gaussian noise $z(n)$ with variance $\sigma_z^2 = 10^{-4}$.

The design paramenters were set to $J_{\max} = -19.75$ dB (maximum MSE), $\varepsilon_0 = 10^{-1}$ (coherence level) and $\xi \in \{0.15, 0.20, 0.25, 0.30\}$ (set of possible kernel bandwidths).

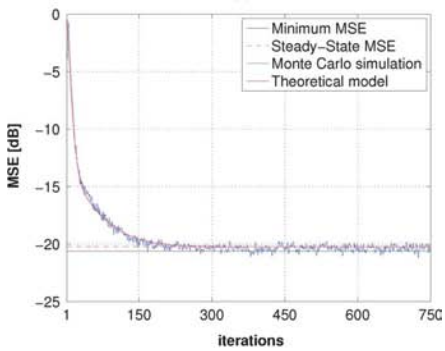
Table VII shows the values of $\eta_{\max}$ obtained by testing the eigenvalues of $\boldsymbol{G}$. Note that the Gerschgorin disk conditions were too strict for all cases making the computation of $\eta_{\max}^{GD}$ impossible.

Table VIII presents the obtained results for the chosen values of ξ. For each pair (ξ, M), the step-size η was chosen to ensure stability (η less than $\eta_{\max}$) and $J_{ms}(\infty) < -19.75$ dB. The values of $J_{ms}(\infty)$ and $J_{ex}(\infty)$ were determined from (75) and n_ϵ was obtained from (69) for $\epsilon = 10^{-4}$. Note that $J_{\min}(\xi) < -19.75$ dB in all cases. It clearly appears that $(M, \xi, \eta) = (7,\, 0.2,\, 0.149)$ is a good design choice, as it satisfies the design conditions after $n_\epsilon = 382$ iterations.

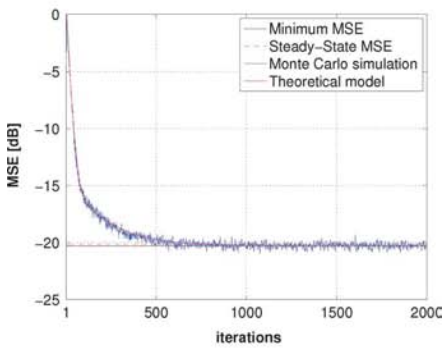
Figs. 6 and 7 illustrate the accuracy of the analytical model for the four cases presented in Table VIII. Again, the results are very promising for this example.



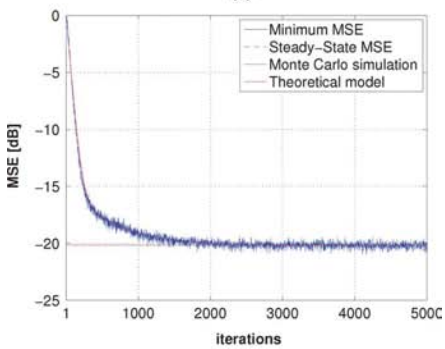
(a)



(b)

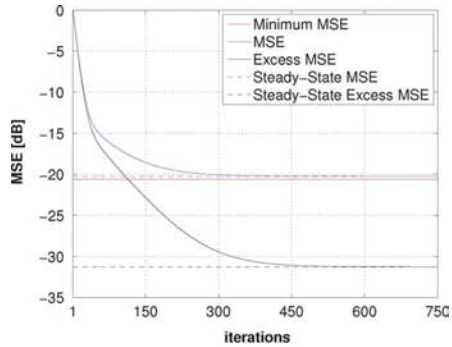


(c)

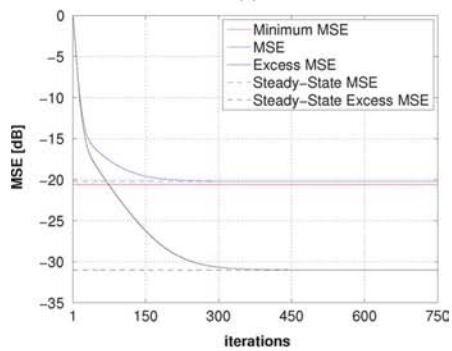


(d)

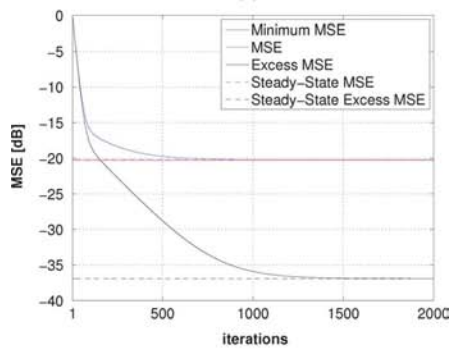
Fig. 6. Theoretical model and Monte Carlo simulation of KLMS for different kernel bandwidths. Ragged curves (blue): simulation results averaged over 500 runs. Continuous curves (red): Theory using (23) and (67). Continuous horizontal lines (blue): Steady-state MSE predicted by theory. Dashed horizontal lines (red): Steady-state MSE from simulations. (a) $\xi = 0.15$ and $M = 11$. (b) $\xi = 0.20$ and $M = 7$. (c) $\xi = 0.25$ and $M = 5$. (d)$\xi = 0.30$ and $M = 3$.



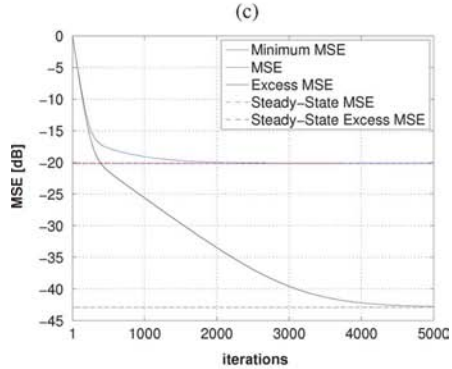
(a)



(b)



(c)



(d)

Fig. 7. Steady-State results. Dashed horizontal lines: MSE and Excess MSE averaged over 500 realizations. Continuous horizontal lines (red): Minimum MSE predicted by theory. Continuous decaying lines: Theoretical MSE (blue) and Excess MSE (black). (a) $\xi = 0.15$ and $M = 11$. (b) $\xi = 0.20$ and $M = 7$. (c) $\xi = 0.25$ and $M = 5$. (d) $\xi = 0.30$ and $M = 3$.

VII. Conclusion

This paper studied the stochastic behavior of the Gaussian KLMS adaptive algorithm for Gaussian inputs and nonlinearities which preserve the stationarity of their inputs. The study resulted in analytical models that predict the behavior of the algorithm as a function of the design parameters. In particular, the new models clarify the joint contribution of the kernel bandwidth and the step-size for the algorithm performance, both during the transient adaptation phase and in steady state. The algorithm convergence was studied and analytical expressions were derived which provide sufficient

stability conditions. Design guidelines were proposed using the theoretical model. These guidelines were applied to different nonlinear system identification problems. Simulation results have illustrated the accuracy of the theoretical models and their usefulness for design purposes. The extension of this analysis to the case of time-varying systems will be the subject of future investigations.

APPENDIX A
POSITIVE-DEFINITENESS OF $\boldsymbol{R}_{\kappa\kappa}$

To prove that $\boldsymbol{R}_{\kappa\kappa}$ is positive definite we prove that all its eigenvalues are positive [17]. Let $\lambda_i \in \mathbb{R}$ and $\boldsymbol{\nu}_i$, $i = 1, \ldots, M$ be, respectively, the ith eigenvalue and the corresponding eigenvector of $\boldsymbol{R}_{\kappa\kappa}$. Hence,

$$\boldsymbol{R}_{\kappa\kappa}\,\boldsymbol{\nu}_i = \lambda_i\,\boldsymbol{\nu}_i, \quad i = 1, \ldots, M. \tag{84}$$

From (16), we can write $\boldsymbol{R}_{\kappa\kappa}$ as

$$\boldsymbol{R}_{\kappa\kappa} = (r_{\rm md} - r_{\rm od})\boldsymbol{I} + r_{\rm od}\mathbf{1}\mathbf{1}^\top \tag{85}$$

where $\mathbf{1} = [1, 1, \ldots, 1]^\top_{M\times 1}$ and $\boldsymbol{I}$ is the $(M \times M)$ identity matrix. Using (85) in (84) yields

$$[(r_{\rm md} - r_{\rm od})\boldsymbol{I} + r_{\rm od}\mathbf{1}\mathbf{1}^\top]\,\boldsymbol{\nu}_i = \lambda_i\,\boldsymbol{\nu}_i, \qquad i = 1, \ldots, M. \tag{86}$$

Noting that $\boldsymbol{\nu}_1 = \beta_1\,\mathbf{1}$, with $\beta_1 \in \mathbb{R}$ is an eigenvector of $\boldsymbol{R}_{\kappa\kappa}$ we have

$$\beta_1(r_{\rm md} - r_{\rm od})\mathbf{1} + \beta_1 r_{\rm od}\mathbf{1}\mathbf{1}^\top\mathbf{1} = \lambda_1\beta_1\,\mathbf{1} \tag{87}$$

which yields

$$\lambda_1 = r_{\rm md} + (M-1)\,r_{\rm od} \tag{88}$$

which is positive since $r_{\rm od} > 0$ ($r_{\rm od}$ is real and is the result of a square root in (16)) and $M \geq 1$.

The remaining eigenvectors are orthogonal to $\boldsymbol{\nu}_1$. Thus, $\mathbf{1}^\top\boldsymbol{\nu}_j = 0$ for all $j \neq 1$ and thus

$$(r_{\rm md} - r_{\rm od})\,\boldsymbol{\nu}_j = \lambda_j\boldsymbol{\nu}_j, \quad j = 2, \ldots, M \tag{89}$$

which yields

$$\lambda_j = r_{\rm md} - r_{\rm od}, \quad j = 2, \ldots, M. \tag{90}$$

which are also positive since $r_{\rm md} > r_{\rm od}$. This concludes the proof.

APPENDIX B
SIGN ANALYSIS OF MATRIX $\boldsymbol{G}$ ENTRIES

We shall now analyze the sign of the entries of matrix $\boldsymbol{G}$ defined in (48)–(49). On the one hand, we know that $\eta^2\mu_3$, $\eta^2\mu_4$ and $\eta^2\mu_5$ are strictly positive. On the other hand, $\eta^2\mu_2 - \eta r_{\rm od}$ and $\frac{1}{2}(2\eta^2\mu_4 - \eta r_{\rm od})$ can be either positive or negative depending on η. The analysis of the diagonal entries of $\boldsymbol{G}$ needs more attention.

Consider first $1 - 2\eta r_{\rm md} + \eta^2\mu_1$. This is a second-degree polynomial with respect to the parameter η whose minimum value is $1 - \frac{r_{\rm md}^2}{\mu_1}$. Using (45) and (47), we know that $\mu_1 > r_{\rm md}^2$. This implies that $1 - 2\eta r_{\rm md} + \eta^2\mu_1 > 0$ for all η. Let us focus now on $\frac{1}{2}(1 - 2\eta r_{\rm md} + 2\eta^2\mu_3)$, whose minimum is equal to $\frac{1}{2}(1 - \frac{r_{\rm md}^2}{2\mu_3})$. Similarly as above, we know that $\mu_3 > r_{\rm md}^2$, which means that $\frac{1}{2}(1 - 2\eta r_{\rm md} + 2\eta^2\mu_3) > 0$ for all η.

As a conclusion, all the diagonal entries of the matrix $\boldsymbol{G}$ are strictly positive, which greatly simplifies the analysis of (52) for stability.

APPENDIX C
STABILITY ANALYSIS IN THE CASES $M = 1$ AND $M = 2$

Let us first derive the condition for stability of the system (51) in the case $M = 1$. The matrix $\boldsymbol{G}$ reduces to the entry $1 - 2\eta r_{\rm md} + \eta^2\mu_1$. This directly implies that the system (51) is stable if

$$\eta < \frac{2\,r_{\rm md}}{\mu_1}. \tag{91}$$

Consider now the case $M = 2$. Expression (52) leads us to the following two inequalities:

$$(1 - 2\eta\,r_{\rm md} + \eta^2\,\mu_1) + \eta^2\,\mu_3 + 2\,|\eta^2\,\mu_2 - \eta\,r_{\rm od}| < 1, \tag{92a}$$

$$(1 - 2\,\eta\,r_{\rm md} + 2\,\eta^2\,\mu_3) + 2\,|\eta^2\,\mu_2 - \eta\,r_{\rm od}| < 1. \tag{92b}$$

We observe that the LHS of (92a) is larger than the LHS of (92b) for all η because $\mu_1 \geq \mu_3$, as shown by (45). Proceeding the same way as we did for $M \geq 3$ in Section IV-A, we conclude that (51) is stable if the following conditions are satisfied:

$$\begin{cases} \eta < \min\{\eta_1, \eta_2\}, & \text{for case } (i) \text{ in Table I} \\ \eta < \eta_1, & \text{for cases } (ii) \text{ and } (iv) \text{ in Table I.} \end{cases} \tag{93}$$

where

$$\eta_1 = \frac{2\,r_{\rm md} + 2\,r_{\rm od}}{\mu_1 + 2\,\mu_2 + \mu_3}$$
$$\eta_2 = \frac{\theta_1}{\theta_2} = \frac{2\,r_{\rm md} - 2\,r_{\rm od}}{\mu_1 - 2\,\mu_2 + \mu_3}. \tag{94}$$

REFERENCES

[1] M. Schetzen, *The Volterra and Wiener Theory of the Nonlinear Systems*. New York: Wiley, 1980.

[2] G. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *J. Math. Anal. Appl.*, vol. 33, pp. 82–95, 1971.

[3] D. L. Duttweiler and T. Kailath, "An RKHS approach to detection and estimation theory: Some parameter estimation problems (Part V)," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 29–37, 1973.

[4] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering*. New York: Wiley, 2010.

[5] Y. Engel, S. Mannor, and R. Meir, "Kernel recursive least squares," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2275–2285, 2004.

[6] P. Honeine, C. Richard, and J. C. M. Bermudez, "On-line nonlinear sparse approximation of functions," in *Proc. IEEE ISIT'07*, Nice, France, Jun. 2007, pp. 956–960.

[7] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1058–1067, Mar. 2009.

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

[9] P. Bouboulis and S. Theodoridis, "Extension of Wirtinger's calculus in reproducing kernel Hilbert spaces and the complex kernel LMS," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 964–978, Mar. 2011.

[10] K. Slavakis and S. Theodoridis, "Sliding window generalized kernel affine projection algorithm using projection mappings," *EURASIP J. Adv. Signal Process.*, vol. 2008, 2008.
[11] J. Omura and T. Kailath, "Some Useful Probability Distributions," Stanford Electron. Lab., Stanford Univ., Stanford, CA, Tech. Rep. 7050-6, 1965.
[12] 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.
[13] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ: Wiley, 2003.
[14] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
[15] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. New York: Oxford Univ. Press, 2001.
[16] D. G. Luenberger, *Introduction Dynamic Systems: Theory, Models & Applications*. New York: Wiley, 1979.
[17] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[18] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 3–27, Mar. 1990.
[19] D. P. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Process. Lett.*, vol. 2, pp. 115–118, Feb. 2004.
[20] J. Vörös, "Modeling and identification of wiener systems with two-segment nonlinearities," *IEEE Trans. Contr. Syst. Technol.*, vol. 11, no. 2, pp. 253–257, Mar. 2003.
[21] J.-S. Wang and Y.-L. Hsu, "Dynamic nonlinear system identification using a Wiener-type recurrent network with okid algorithm," *J. Inf. Sci. Eng.*, vol. 24, pp. 891–905, 2008.
[22] H. Al-Duwaish, M. N. Karim, and V. Chandrasekar, "Use of multilayer feedforward neural networks in identification and control of wiener model," in *Proc. Contr. Theory Appl.*, May 1996, vol. 143, no. 3.

Wemerson D. Parreira was born in Uberaba, Brazil, on March 15, 1978. He received the bachelor's degree in mathematics in 2002 and the M.Sc. degree in electrical engineering from the Federal University of Uberlândia in 2005.

He is currently a doctoral student in electrical engineering at the Federal University of Santa Catarina, Florianópolis, Brazil. He worked as a professor in the Department of Informatics and Statistics, the Federal University of Santa Catarina (2006–2007). His experience is in mathematical modeling, with emphasis on Electrical Engineering. His current research interests include reproducing kernels, nonlinear filtering, adaptive filtering and statistical analysis.

José Carlos M. Bermudez (M'85–SM'02) received the B.E.E. degree from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, the M.Sc. degree in electrical engineering from COPPE/UFRJ, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, Canada, in 1978, 1981, and 1985, respectively.

He joined the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 1985. He is currently a Professor of electrical engineering. In the winter of 1992, he was a Visiting Researcher with the Department of Electrical Engineering, Concordia University. In 1994, he was a Visiting Researcher with the Department of Electrical Engineering and Computer Science, University of California, Irvine (UCI). His research interests have involved analog signal processing using continuous-time and sampled-data systems. His recent research interests are in digital signal processing, including linear and nonlinear adaptive filtering, active noise and vibration control, echo cancellation, image processing, and speech processing.

Prof. Bermudez served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING in the area of adaptive filtering from 1994 to 1996 and from 1999 to 2001. He also served as an Associate Editor of the EURASIP *Journal of Advances on Signal Processing* from 2006 to 2010. He was a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society from 1998 to 2004.

Cédric Richard (S'98–M'01–SM'07) was born on January 24, 1970, in Sarrebourg, France. He received the Dipl.-Ing. and the M.S. degrees in 1994 and the Ph.D. degree in 1998 from the University of Technology of Compiègne (UTC), France, all in electrical and computer engineering.

He joined the Côte d'Azur Observatory, University of Nice Sophia-Antipolis, France, in 2009. He is currently a Professor of electrical engineering. From 1999 to 2003, he was an Associate Professor at the University of Technology of Troyes (UTT), France. From 2003 to 2009, he was a Professor at the UTT, and the supervisor of a group consisting of 60 researchers and Ph.D. students. In winter 2009 and autumn 2010, he was a Visiting Researcher with the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianòpolis, Brazil. He has been a junior member of the Institut Universitaire de France since October 2010. He is the author of more than 120 papers. His current research interests include statistical signal processing and machine learning.

Dr. Richard was the General Chair of the XXIth Francophone conference GRETSI on Signal and Image Processing that was held in Troyes, France, in 2007, and of the IEEE Statistical Signal Processing Workshop (IEEE SSP'11) that was held in Nice, France, in 2011. Since 2005, he has been a member of the board of the Federative CNRS research group ISIS on Information, Signal, Images, and Vision. He is a member of the GRETSI association board and of the EURASIP society.

Cédric Richard served as an Associate Editor of the IEEE Transactions on Signal Processing from 2006 to 2010, and of Elsevier Signal Processing since 2009. In 2009, he was nominated liaison local officer for EURASIP, and member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society. Paul Honeine and Cédric Richard received Best Paper Award for "Solving the pre-image problem in kernel machines: a direct method" at the 2009 IEEE Workshop on Machine Learning for Signal Processing (IEEE MLSP'09).

Jean-Yves Tourneret (SM'08) received the ingnieur degree in electrical engineering from Ecole Nationale Suprieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique in Toulouse (ENSEEIHT) in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992.

He is currently a professor with the University of Toulouse, France (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal processing with a particular interest to Bayesian and Markov Chain Monte Carlo methods.

Dr. Tourneret has been involved in the organization of several conferences including the European Conference on Signal Processing (EUSIPCO'2002) (as the Program Chair), the international Conference ICASSP'06 (in charge of plenaries) and the Statistical Signal Processing Workshop (SSP 2012) (for international liaisons). He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001–2007, 2010–present). He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2008–2011).