Closed-form conditions for convergence of the Gaussian kernel-least-mean-square algorithm

Cédric Richard(1), Jose-Carlos M. Bermudez(2)
(1)Université de Nice Sophia-Antipolis, France
(2)Department of Electrical Engineering, Federal University of Santa Catarina, Florianòpolis, SC, Brazil

***Abstract*—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. One of our recent works has brought a new contribution to the discussion about kernel-based adaptive filtering by providing the first convergence analysis of the kernel-LMS algorithm with Gaussian kernel. A necessary and sufficient condition for convergence has been clearly established. Checking the stability of the algorithm can, unfortunately, be computationally expensive because one needs to calculate the extreme eigenvalues of a large matrix, for each set of candidate tuning parameters. The aim of this paper is to circumvent this drawback by examining two easy-to-handle conditions that allow to examine how the stability limit varies as a function of the step-size, the kernel bandwidth, and the filter length. One of them is a conjectured necessary and sufficient condition for convergence that allows to greatly simplify calculations.**

I. Introduction

Many practical applications require nonlinear signal processing. Nonlinear system identification methods based on reproducing kernel Hilbert spaces (RKHS) have gained popularity over the last decades [2], [6]. 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. For an overview, see [8]. The block diagram of a kernel-based adaptive system identification problem is presented in Figure 1. Here, $\mathcal{U}$ is a compact subspace of $\mathbb{R}^q$, $\kappa : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ 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 [6] states that the function $\psi(\cdot)$ which minimizes the cost function $\sum_{n=1}^{N}(\psi(\boldsymbol{u}(n)) - d(n))^2$, given N input vectors $\boldsymbol{u}(n)$ and desired outputs $d(n)$, can be written as $\psi(\cdot) = \sum_{n=1}^{N} \alpha_n \, \kappa(\cdot, \boldsymbol{u}(n))$. 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, authors in the field have focused on finite-order models

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

In [8], the authors present an overview of the existing techniques to select the M kernel functions in (1) that form the so-called *dictionary*, an example of which is the coherence criterion [12]. The algorithms developed using these ideas include the kernel least-mean-square (KLMS) algorithm [7], the kernel recursive-least-square (KRLS) algorithm [3], the kernel normalized least-mean-square (KNLMS) algorithm and the kernel affine projection (KAPA) algorithm [5], [12], [13]. 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. Choosing the algorithm and nonlinear model parameters to achieve a prescribed performance is a difficult task, and requires an extensive analysis of the algorithm stochastic behavior. Our work [11] has recently brought 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 filtering process is defined by

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \eta \, e(n) \, \boldsymbol{\kappa}_{\boldsymbol{\omega}}(n). \tag{2}$$

where $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n) = [\kappa(\boldsymbol{u}(n), \boldsymbol{u}(\omega_1)), \ldots, \kappa(\boldsymbol{u}(n), \boldsymbol{u}(\omega_M))]^\top$, and $\kappa(\boldsymbol{u}, \boldsymbol{u}')$ the Gaussian kernel

$$\kappa(\boldsymbol{u}, \boldsymbol{u}') = \exp\left(\frac{-\|\boldsymbol{u} - \boldsymbol{u}'\|^2}{2\xi^2}\right) \tag{3}$$

with kernel bandwidth ξ. In [11], we derived expressions for the mean-weight-error vector and the mean-square-error. These models give engineers the opportunity to choose the algorithm parameters *a priori* in order to achieve prescribed convergence speed and quality of the estimate, and allow the determination of stability limits. Checking the stability of the algorithm (2) can be computationally expensive as it needs to calculate the extreme eigenvalues of an $(M^2 \times M^2)$ matrix, say $\boldsymbol{G}$, for each candidate tuning parameters η, M and ξ.

The aim of this paper is to circumvent this drawback by examining two easy-to-handle conditions that allow to examine how the stability limit varies as a function of the step-size, the kernel bandwidth, and the filter length. The first one is a sufficient condition based on the Gerschgorin disk theorem, which has already been derived in [11]. The second one is a conjectured necessary and sufficient condition for convergence. It allows to greatly simplify calculations, and to examine how the stability limits vary as a function of the step-size η, the kernel bandwidth ξ, and the filter length M.

II. Convergence analysis

Let $\boldsymbol{v}(n) = \boldsymbol{\alpha}(n) - \boldsymbol{\alpha}_{\text{opt}}$ be the weight-error vector. Let vector $\boldsymbol{c}_v(n)$ be the lexicographic representation of the autocorrelation matrix $\boldsymbol{C}_v(n) = E\{\boldsymbol{v}(n)\, \boldsymbol{v}^\top(n)\}$, i.e., the matrix $\boldsymbol{C}_v(n)$ is stacked column-wise into a vector $\boldsymbol{c}_v(n)$.

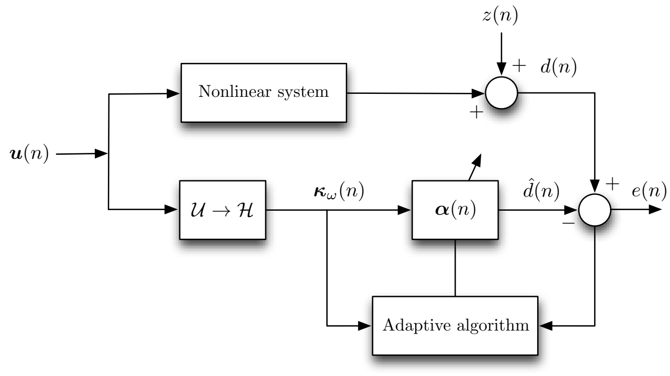

Fig. 1. Kernel-based adaptive system identification.

It was shown in [11] that, under some simplifying statistical assumptions, we have

$$\boldsymbol{c}_v(n+1) = \boldsymbol{G}\,\boldsymbol{c}_v(n) + \eta^2 J_{\min}\,\boldsymbol{r}_{\kappa\kappa} \tag{4}$$

with $\boldsymbol{r}_{\kappa\kappa}$ the lexicographic representation of the correlation matrix $\boldsymbol{R}_{\kappa\kappa} = E\{\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}^\top(n)\}$ of the kernelized input, and $J_{\min}$ the minimum MSE corresponding to the optimum weight vector $\boldsymbol{\alpha}_{\text{opt}} = \boldsymbol{R}_{\kappa\kappa}^{-1}\,\boldsymbol{p}_{\kappa d}$, where $\boldsymbol{p}_{\kappa d} = E\{d(n)\,\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)\}$ is the cross-correlation vector between $\boldsymbol{\kappa}_{\boldsymbol{\omega}}(n)$ and $d(n)$. Matrix $\boldsymbol{G}$, of size $(M^2 \times M^2)$, is defined as

$$\boldsymbol{G} = \Big[\boldsymbol{h}^{11}\,\boldsymbol{h}^{12}\,\ldots\,\boldsymbol{h}^{1M}\,\ldots\,\boldsymbol{h}^{MM}\Big] \tag{5}$$

with $\boldsymbol{h}^{\ell p}$ the $(M^2 \times 1)$ lexicographic representation of the matrix $\boldsymbol{H}^{\ell p}$, given by

$$\begin{array}{ll}
\text{if } (i = j) & \\
\quad [\boldsymbol{H}^{ii}]_{ii} = 1 - 2\eta r_{\text{md}} + \eta^2\mu_1 & \\
\quad [\boldsymbol{H}^{ii}]_{pp} = \eta^2\mu_3 & p \neq i \\
\quad [\boldsymbol{H}^{ii}]_{ip} = \eta^2\mu_2 - \eta r_{\text{od}} = [\boldsymbol{H}^{ii}]_{pi} & p \neq i \\
\quad [\boldsymbol{H}^{ii}]_{p\ell} = \eta^2\mu_4 & \text{otherwise} \\
& \\
\text{if } (i \neq j) & \\
\quad [\boldsymbol{H}^{ij}]_{ij} = [\boldsymbol{H}^{ij}]_{ji} = \frac{1}{2}(1 - 2\eta r_{\text{md}} + 2\eta^2\mu_3) & \\
\quad [\boldsymbol{H}^{ij}]_{pp} = \eta^2\mu_4 & p \neq i, j \\
\quad [\boldsymbol{H}^{ij}]_{ii} = [\boldsymbol{H}^{ij}]_{jj} = \eta^2\mu_2 - \eta r_{\text{od}} & \\
\quad [\boldsymbol{H}^{ij}]_{ip} = [\boldsymbol{H}^{ij}]_{pi} = \frac{1}{2}(2\eta^2\mu_4 - \eta r_{\text{od}}) & p \neq i, j \\
\quad [\boldsymbol{H}^{ij}]_{pj} = [\boldsymbol{H}^{ij}]_{jp} = \frac{1}{2}(2\eta^2\mu_4 - \eta r_{\text{od}}) & p \neq i, j \\
\quad [\boldsymbol{H}^{ii}]_{p\ell} = \eta^2\mu_5 & \text{otherwise}
\end{array}$$

where the μ_k's are the fourth-order moments of the kernelized input defined as

$$\begin{aligned}
\mu_1 &:= E\{\kappa_{\boldsymbol{\omega}_i}^4(n)\} \\
\mu_2 &:= E\{\kappa_{\boldsymbol{\omega}_i}^3(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\} \\
\mu_3 &:= E\{\kappa_{\boldsymbol{\omega}_i}^2(n)\,\kappa_{\boldsymbol{\omega}_j}^2(n)\} \\
\mu_4 &:= E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\,\kappa_{\boldsymbol{\omega}_\ell}^2(n)\} \\
\mu_5 &:= E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\,\kappa_{\boldsymbol{\omega}_\ell}(n)\,\kappa_{\boldsymbol{\omega}_p}(n)\}.
\end{aligned} \tag{6}$$

Parameters r_{md} and r_{od} are the main-diagonal and off-diagonal entries of the correlation matrix $\boldsymbol{R}_{\kappa\kappa}$ given by

$$\begin{aligned}
r_{\text{md}} &:= E\{\kappa_{\boldsymbol{\omega}_i}^2(n)\} \\
r_{\text{od}} &:= E\{\kappa_{\boldsymbol{\omega}_i}(n)\,\kappa_{\boldsymbol{\omega}_j}(n)\}.
\end{aligned} \tag{7}$$

As extensively explained in [11], parameters μ_k's, r_{md} and r_{od} can be calculated theoretically in the case of i.i.d. Gaussian inputs $\boldsymbol{u}(n)$. Their value depends on the moments of $\boldsymbol{u}(n)$, the filter length M, and the kernel bandwidth ξ.

Before concluding this section, let us introduce the following inequalities relating the fourth-order moments μ_k and the entries of the correlation matrix $\boldsymbol{R}_{\kappa\kappa}$, which will be used in the sequel. Using Hölder's inequality, it can be shown that

$$\mu_5 \leq \mu_4 \leq \mu_3 \leq \mu_2 \leq \mu_1, \tag{8}$$

and, by virtue of the Chebyshev's sum inequality

$$r_{\text{md}}^2 \leq \mu_3. \tag{9}$$

We refer the reader to the proofs in [11].

We shall now examine the conditions for convergence of the Gaussian KLMS algorithm using model (4). It can be checked that the matrix $\boldsymbol{G}$ is symmetric. This implies that it can be diagonalized, and all its eigenvalues are real-valued. A necessary and sufficient condition for convergence is that all these eigenvalues lie inside $(-1,\,1)$ [9, Section 5.9]. First, we shall consider a sufficient condition based on the Gerschgorin disk theorem. After arguing that these conditions are too restrictive, we provide an easy-to-handle necessary and sufficient condition for convergence, based on a conjecture.

A. Gerschgorin disk conditions

The eigenvalues of matrix $\boldsymbol{G}$ lie inside the union of Gerschgorin disks [4], each disk being centered at a diagonal element of $\boldsymbol{G}$, with radius given by the sum of the absolute values of the remaining elements of the same row. A sufficient condition for stability of (4) is thus given by

$$|[\boldsymbol{G}]_{ii}| + \sum_{\substack{\ell=1 \\ \ell \neq i}}^{M^2} |[\boldsymbol{G}]_{i\ell}| < 1 \tag{10}$$

for $i = 1, \ldots, M^2$. The definition of $\boldsymbol{G}$ shows that the rows of this matrix 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 Gerschgorin disks can be distinguished. Using (8), it can be shown that all the entries of $\boldsymbol{G}$ are positive except possibly for $[\boldsymbol{G}]_{i\ell} = \eta^2\,\mu_2 - \eta\,r_{\text{od}}$ and $[\boldsymbol{G}]_{i\ell} = \frac{1}{2}(2\,\eta^2\,\mu_4 - \eta\,r_{\text{od}})$. Expression (10) thus leads to only two sufficient conditions, defined as follows for $M \geq 3$,

$$\begin{aligned}
\lambda_{\text{ger}}^{(1)} := (1 - 2\eta r_{\text{md}} + \eta^2\mu_1) &+ (M-1)\eta^2\mu_3 \\
&+ 2(M-1)|\eta^2\mu_2 - \eta r_{\text{od}}| \\
&+ (M-1)(M-2)\eta^2\mu_4 < 1,
\end{aligned} \tag{11a}$$

TABLE I
STABILITY RESULTS FOR EXAMPLE 1

ξ	M	$\eta_{\max}, \eta_{\text{conj}}$	η_{ger}
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 II
STABILITY RESULTS FOR EXAMPLE 2

ξ	M	$\eta_{\max}, \eta_{\text{conj}}$	η_{ger}
0.05	7	2.33	–
0.065	4	2.49	0.68
0.075	3	2.60	1.92
0.125	2	2.39	2.32

TABLE III
STABILITY RESULTS FOR EXAMPLE 3

ξ	M	$\eta_{\max}, \eta_{\text{conj}}$	η_{ger}
0.15	11	1.17	–
0.20	7	1.19	–
0.25	5	1.24	–
0.30	3	1.59	–

$$\begin{aligned}\lambda_{\text{ger}}^{(2)} := (1 - 2\eta r_{\text{md}} + 2\eta^2\mu_3) + 2|\eta^2\mu_2 - \eta r_{\text{od}}| \\ + (M-2)\eta^2\mu_4 \\ + 2(M-2)|2\eta^2\mu_4 - \eta r_{\text{od}}| \\ + (M-2)(M-3)\eta^2\mu_5 < 1.\end{aligned} \tag{11b}$$

The intersection of these two conditions provides the following sufficient condition for stability

$$\lambda_{\text{ger}}(\eta, M, \xi) := \max\{\lambda_{\text{ger}}^{(1)}, \lambda_{\text{ger}}^{(2)}\} < 1. \tag{12}$$

which avoids multiple time consuming diagonalizations of the matrix $\boldsymbol{G}$. Solving $\lambda_{\text{ger}} = 1$ to derive upper-bounds with respect to η, M or ξ requires (basic) numerical methods. We observe that $\lambda_{\text{ger}}^{(1)}$ and $\lambda_{\text{ger}}^{(2)}$ are piecewise polynomial functions in η. Because they are both equal to 1 for $\eta = 0$, their derivative at the origin must be strictly negative for the conditions (11a)–(11b) to be meaningful. This leads to the condition $(M-1)r_{\text{od}} < r_{\text{md}}$, which is very restrictive. Application examples in Section III show situations where the Gerschgorin disk test is ineffective.

B. Conjectured necessary and sufficient condition

It can be shown that there exist θ_1, $\theta_2 \in \mathbb{R}$ not simultaneously equal to zero, so that the $(M^2 \times 1)$ vector $\boldsymbol{w}$ with i-th entry defined by

$$\begin{cases} w_i = \theta_1, & \text{if } (i-1) \in M\mathbb{Z} \\ w_i = \theta_2, & \text{otherwise,} \end{cases} \tag{13}$$

is an eigenvector of $\boldsymbol{G}$. The conjecture is that the largest eigenvalue of $\boldsymbol{G}$ in absolute value is associated to an eigenvector of the form (13). While we currently have no proof, we have not numerically experienced any contradiction. The difficulty in proving this result is that it does not only rely on the specific structure of the matrix, but also on the expression and/or some order relations of its entries such as (8).

Due to symmetries in matrix $\boldsymbol{G}$, the eigensystem $\boldsymbol{G}\boldsymbol{w} = \lambda\boldsymbol{w}$ of M^2 linear equations in unknowns θ_1 and θ_2 reduces to the equation $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$, where $\boldsymbol{A}$ is the (2×2) matrix whose entries a_{ij} are given by

$$\begin{aligned} a_{11} &= \eta^2\,(\mu_1 + (M-1)\mu_3) - 2\eta r_{\text{md}} + 1 \\ a_{12} &= (M-1)\left(\eta^2(2\mu_2 + (M-2)\mu_4) - 2\eta r_{\text{od}}\right) \\ a_{21} &= \eta^2\,(2\mu_2 + (M-2)\mu_4) - 2\eta r_{\text{od}} \\ a_{22} &= \eta^2\,(2\mu_3 + 4(M-2)\mu_4 + (M-2)(M-3)\mu_5) \\ &\quad - 2\eta(r_{\text{md}} + (M-2)r_{\text{od}}) + 1 \end{aligned} \tag{14}$$

Solving the above-mentioned equation yields the following two real-valued eigenvalues

$$\begin{aligned} \lambda &= \frac{1}{2}(a_{11} + a_{22} - \sqrt{\Delta}) \\ \lambda' &= \frac{1}{2}(a_{11} + a_{22} + \sqrt{\Delta}) \end{aligned} \tag{15}$$

with $\Delta = (a_{11} - a_{22})^2 + 4(M-1)\,a_{21}^2$. This finally implies the conjectured necessary and sufficient condition for convergence

$$\lambda_{\text{conj}}(\eta, M, \xi) := \frac{1}{2}(|a_{11} + a_{22}| + \sqrt{\Delta}) < 1. \tag{16}$$

Obviously, exploiting this condition is much less computationally demanding than diagonalizing the $(M^2 \times M^2)$ matrix $\boldsymbol{G}$, and checking if its eigenvalues lie inside $(-1, 1)$. In addition, it provides an upper-bound that can be easily studied even if solving $\lambda_{\text{conj}} = 1$ with respect to η, M or ξ requires (basic) numerical methods.

III. EXPERIMENTATIONS

We shall now consider the experiments described in [11], and compare the upper bounds λ_{ger} and λ_{conj} provided by the Gerschgorin disk conditions (11a)–(11b), and the (conjectured) necessary and sufficient condition (16), respectively. We shall also check that λ_{conj} matches the estimated largest eigenvalue $\lambda_{\max}$ in absolute value of the matrix $\boldsymbol{G}$. Let η_{ger}, η_{conj} and $\eta_{\max}$ be the maximum step sizes provided by these three approaches, for fixed parameters M and ξ.

All the Matlab codes used in this paper are available on the personal website of the first author: www.cedric-richard.fr

A. Experiment 1

We consider the problem studied in [10], for which

$$y(n) = \frac{y(n-1)}{1 + y^2(n-1)} + u^3(n-1) \tag{17}$$

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

Table I reports the maximum step sizes η_{ger}, η_{conj} and $\eta_{\max}$, for several values of M and ξ. It can be observed that the condition imposed by the Gerschgorin disks is very restrictive compared to the two others. Figure 2 (left) represents λ_{ger}, λ_{conj} and $\lambda_{\max}$ as a function of η, with parameters M and ξ defined as in the first row of Table I. It can be noticed that the two latter superimpose perfectly. Figure 3 represents the conjectured largest eigenvalue λ_{conj} of $\boldsymbol{G}$ as a function of parameters η (left), M (middle), and ξ (right), in the vicinity

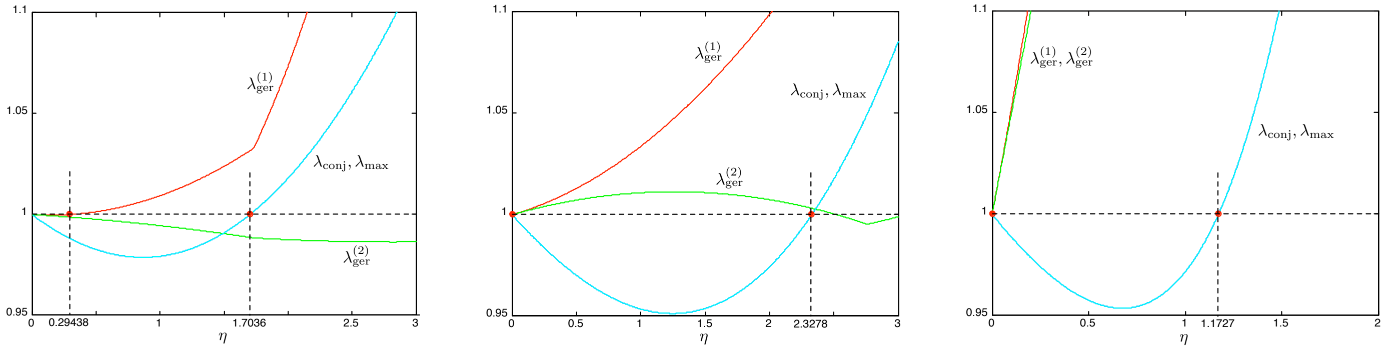

Fig. 2. Comparison of the upper bounds ($\lambda_{\text{ger}}^{(1)}, \lambda_{\text{ger}}^{(2)}$) and λ_{conj} provided by the Gerschgorin disk conditions (11a)–(11b), and the (conjectured) necessary and sufficient condition (16), respectively, with the largest eigenvalue $\lambda_{\max}$ of $\boldsymbol{G}$ in absolute value. The three experimental setups are described in the first row of Table I (left), Table II (middle) and Table III (right).

of the stability limit defined by the first row of Table I. Finally, Figure 4 (left) illustrates the convergence of the mean-square error estimated by averaging over 500 runs. The step size was arbitrarily chosen to be $1/3$ of the maximum step size η_{conj}. We encourage the reader to refer to [11] for an analysis of the stochastic behavior of the KLMS algorithm.

B. Experiment 2

We now consider the nonlinear dynamic system identification problem studied in [14]. The input signal is a sequence of statistically independent vectors $\boldsymbol{u}(n) = [u_1(n)\, u_2(n)]$ with correlated samples satisfying $u_1(n) = 0.5\, u_2(n) + \eta_u(n)$. The second component of $\boldsymbol{u}(n)$ 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 $\sigma_{u_1}^2 = 0.0156$. The nonlinear system under study consists of the linear system with memory defined by

$$y(n) = u_1(n) + 0.5 u_2(n) - 0.2\, y(n-1) + 0.35\, y(n-2), \quad (18)$$

and the nonlinear Wiener function

$$\varphi_y(n) = \begin{cases} \dfrac{y(n)}{3(0.1 + 0.9\, y^2(n))^{1/2}} & \text{if } y(n) \geq 0 \\ \dfrac{-y^2(n)(1-\exp(0.7y(n)))}{3} & \text{otherwise.} \end{cases} \quad (19)$$

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

Table II reports the maximum step sizes η_{ger}, η_{conj} and $\eta_{\max}$, for several values of M and ξ. Observe in Figure 2 (middle) that, with the experimental setup described in the first row of Table II, no bound on η was provided by the Gerschgorin disk condition (12). The reason is that $(M-1)r_{\text{od}} < r_{\text{md}}$ is not satisfied in this case, because $r_{\text{md}} = 0.0439$ and $r_{\text{od}} = 0.0088$. Finally, Figure 4 (middle) illustrates the convergence of the mean-square error estimated by averaging over 500 runs. The step size was arbitrarily chosen to be $1/3$ of η_{conj}.

C. Experiment 3

Finally, as a third example, we considered the fluid-flow control problem studied in [1], [15]. The input signal was a sequence $\boldsymbol{u}(n) = [u_1(n)\, u_2(n)]$ of statistically independent vectors with samples satisfying $u_1(n) = 0.5\, u_2(n) + \eta_u(n)$. The second component $u_2(n)$ is a i.i.d. Gaussian sequence with variance $\sigma_{u_2}^2 = 0.0625$, and $\eta_u(n)$ is a i.i.d. Gaussian noise so that $u_1(n)$ has variance $\sigma_{u_1}^2 = 0.0625$. The nonlinear system under study consists of the linear system

$$\begin{aligned} y(n) =& 0.1044\, u_1(n) + 0.0883\, u_2(n) \\ &+ 1.4138\, y(n-1) - 0.6065\, y(n-2) \end{aligned} \quad (20)$$

and the nonlinear Wiener function

$$\varphi_y(n) = \frac{0.3163\, y(n)}{\sqrt{0.10 + 0.90\, y^2(n)}}. \quad (21)$$

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

It can be noticed in Table III that no upper bound for the step size η was provided by the Gerschgorin disk condition. As previously, condition $(M-1)r_{\text{od}} < r_{\text{md}}$ was not satisfied in these cases. Figure 2 (right) represents λ_{ger}, λ_{conj} and $\lambda_{\max}$ as a function of η, with parameters M and ξ defined as in the first row of Table III. It can be noticed that λ_{conj} and $\lambda_{\max}$ superimpose perfectly. Finally, Figure 4 (middle) illustrates the convergence of the mean-square error estimated by averaging over 500 runs. The step size was arbitrarily chosen to be $1/3$ of the maximum step size η_{conj}.

IV. Conclusion

The kernel least-mean-square filter has become a popular algorithm in nonlinear adaptive filtering due to its simplicity and robustness. One of our recent works has brought a new contribution to the analysis of this approach by providing the first analytical models of convergence of the Gaussian kernel least-mean-square algorithm. Checking its stability can be computationally expensive as it needs to calculate the extreme eigenvalues of large matrix, for each candidate parameter setting. To circumvent this drawback, in this paper, we presented two easy-to-handle conditions. The first one is a sufficient condition based on the Gerschgorin disk theorem. The second one is a conjectured necessary and sufficient condition for convergence that allows to greatly simplify calculations.

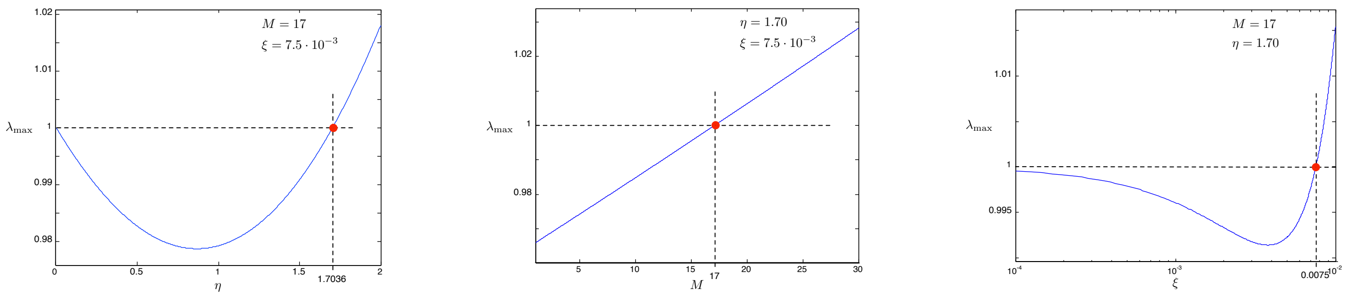

Fig. 3. Largest eigenvalue of $\boldsymbol{G}$ in absolute value provided by the (conjectured) expression λ_{conj} as a function of η (left), M (middle), and ξ (right), in the vicinity of the stability limit defined by the first row of Table I.

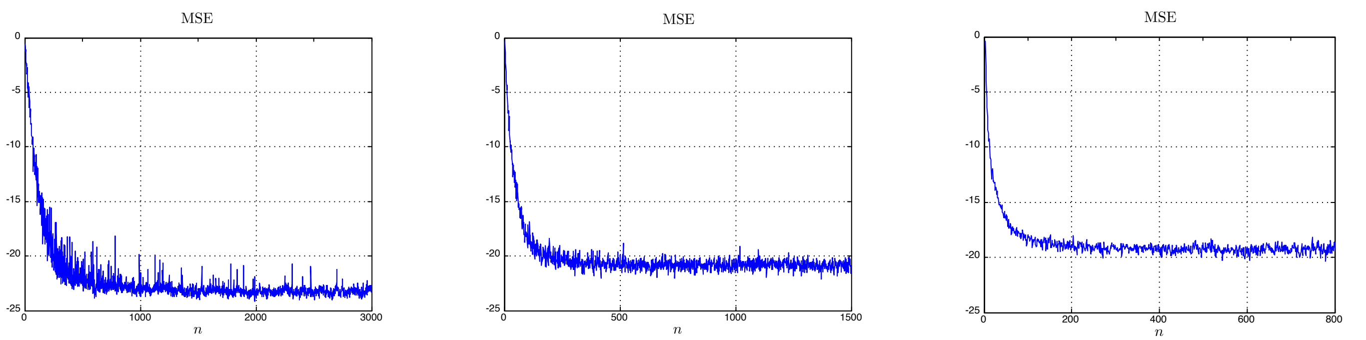

Fig. 4. Monte-Carlo simulation of KLMS algorithm with Gaussian kernel. The three experimental setups are described in the first row of Table I (left), Table II (middle) and Table III (right). In each case, the step size η was arbitrarily chosen to be $1/3$ of the maximum step size.

REFERENCES

[1] H. Al-Duwaish, M. N. Karim, and V. Chandrasekar. Use of multilayer feedfoorward neural networks in identification and control of wiener model. In *Proc. Control Theory Appl.*, volume 143, May 1996.

[2] D. L. Duttweiler and T. Kailath. An RKHS approach to detection and estimation theory: Some parameter estimation problems (Part V). *IEEE Trans. Inf. Theory*, 19(1):29–37, 1973.

[3] Y. Engel, S. Mannor, and R. Meir. Kernel recursive least squares. *IEEE Trans. Signal Process.*, 52(8):2275–2285, 2004.

[4] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, 1996.

[5] P. Honeine, C. Richard, and J. C. M. Bermudez. On-line nonlinear sparse approximation of functions. In *Proc. IEEE ISIT'07*, pages 956–960, Nice, France, June 2007.

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

[7] W. Liu, P. P. Pokharel, and J. C. Principe. The kernel least-mean-squares algorithm. *IEEE Trans. Signal Process.*, 56(2):543–554, February 2008.

[8] W. Liu, J. C. Principe, and S. Haykin. *Kernel Adaptive Filtering*. Jonh Wiley & Sons, Inc., 2010.

[9] D. G. Luenberger. *Introduction Dynamic Systems: Theory, Models & Applications*. Jonh Wiley & Sons, Inc., Stanford, CA, 1979.

[10] D. P. Mandic. A generalized normalized gradient descent algorithm. *IEEE Signal Processing Letters*, 2:115–118, February 2004.

[11] 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.*, 60(5):2208–2222, 2012.

[12] C. Richard, J. C. M. Bermudez, and P. Honeine. Online prediction of time series data with kernels. *IEEE Trans. Signal Process.*, 57(3):1058 –1067, March 2009.

[13] K. Slavakis and S. Theodoridis. Sliding window generalized kernel affine projection algorithm using projection mappings. *EURASIP Journal on Advances in Signal Processing*, 2008:ID 735351, 2008.

[14] J. Vörös. Modeling and identification of Wiener systems with two-segment nonlinearities. *IEEE Transactions on Control Systems Technology*, 11(2):253 – 257, March 2003.

[15] J.-S. Wang and Y.-L. Hsu. Dynamic nonlinear system identification using a Wiener-type recurrent network with okid algorithm. *Journal of Information Science and Engineering*, 24:891–905, 2008.