Robustness of the Filtered-X LMS Algorithm—Part II: Robustness Enhancement by Minimal Regularization for Norm Bounded Uncertainty

Rufus Fraanje, Stephen J. Elliott, *Senior Member, IEEE*, and Michel Verhaegen

***Abstract*—The relationship between the regularization methods proposed in the literature to increase the robustness of the filtered-x LMS (FXLMS) algorithm is discussed. It is shown that the existing methods are special cases of a more general robust FXLMS algorithm in which particular filters determine the type of regularization. Based on the analysis by Fraanje, Verhaegen, and Elliott [“Robustness of the Filtered-X LMS Algorithm—Part I: Necessary Conditions for Convergence and the Asymptotic Pseudospectrum of Toeplitz Matrices” of this issue], regularization filters are designed that guarantee that the strictly positive real conditions for asymptotic convergence or noncritical behavior are just satisfied for all uncertain systems contained in a particular norm bounded set.**

***Index Terms*—Effort weighting, filtered-x LMS (FXLMS), leaky, model uncertainty, output weighting.**

I. Introduction

It is well known that model errors may result in instability of the filtered-x LMS (FXLMS) update rule. In [1] a strictly positive real (SPR) condition on the model error for which the FXLMS update rule asymptotically converges in the mean was analyzed. But what can be done when the condition is not satisfied and the FXLMS algorithm gets unstable? The solution that lays at hand is to improve the model of the system, e.g., by (online) system identification methods. However, this solution may be computationally demanding in online situations, moreover noise, nonlinearities, and undermodeling still contribute to model errors.

This paper follows the alternative approach of making the stability of the FXLMS update rule less sensitive to model errors. One simple and elegant way to accomplish this is to add *leakage* to the FXLMS update rule. Leakage will bias the asymptotic solution as in ridge regression or Tikhonov regularization, but also relaxes the condition for asymptotic convergence, see, e.g., [2, p. 248] and [3, a.o. see, p. 386]. Another approach, quite similar to leakage, is adding a control effort weighting to the cost function that reduces the power of the control signal, see, e.g., [2, p. 246]. Besides the necessity of tuning a scalar parameter, a drawback of leakage and control effort weighting, is that there is no frequency selectivity in the regularization, resulting in too much conservatism. For this reason, a robust FXLMS algorithm has been proposed in [4] in which the model uncertainty is considered as a frequency dependent stochastic variable with zero mean and known covariance. This robust FXLMS algorithm is equivalent with a frequency dependent control effort weighting. The control effort is reduced especially in the frequency bands where the model uncertainty is large, such that also the condition for asymptotic convergence is relaxed especially in these frequency bands. Another robust variant of the FXLMS algorithm is obtained by output weighting, i.e., the power of the control signal filtered by the system is added to the cost function [5], [6].

Though these four regularization approaches (leakage, control effort, output effort, and model uncertainty weighting) relax the condition for asymptotic convergence of the FXLMS update rule, they do not guarantee that the condition is satisfied. One may increase the regularization by means of a scalar tuning parameter, but this tuning may be cumbersome and resulting in too much conservatism. This paper presents a robust FXLMS algorithm that guarantees the asymptotic convergence condition is satisfied for all model errors contained in a particular norm bounded set. This robust algorithm was suggested by a general robust FXLMS algorithm structure in which the aforementioned regularization methods, except leaky FXLMS, can be embedded. The general robust FXLMS algorithm is obtained by adding to the cost function a term with the power of the control signal filtered by a filter $\mathbf{F}(z)$, that is to be designed. It is equivalent to FXLMS with control effort weighting for $\mathbf{F}(z) = \rho\mathbf{I}$, equivalent to FXLMS with output weighting for $\mathbf{F}(z) = \rho\widehat{\mathbf{G}}(z)$ and to uncertainty weighting for $\mathbf{F}(z) = \widehat{\boldsymbol{\Delta}\mathbf{G}}(z)$, where $\widehat{\boldsymbol{\Delta}\mathbf{G}}(z)$ an estimate of the spectral factor of the model error covariance. Though leaky FXLMS does not fully fit into the general robust FXLMS algorithm structure, the regularization is very similar. In fact, for white noise reference signals with unit covariance matrix, the mean update equation of robust FXLMS with effort weighting is the same as the mean update equation of leaky FXLMS. Following the notation introduced in [1], briefly recapitulated in the following section, the relations between the various robust FXLMS algorithms are summarized for the single-channel case in Table I.

The paper is organized as follows. Section II derives the general (multichannel) robust FXLMS algorithm and the condition for its asymptotic convergence. Section III derives the filter

Manuscript received March 28, 2006; revised December 21, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. John J. Shynk. This work was supported by the Intelligent Materials for Active Noise Reduction (InMAR) project funded under the 6th Framework Programme of the European Commission.

R. Fraanje and M. Verhaegen are with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD, Delft, The Netherlands (e-mail: p.r.fraanje@tudelft.nl; m.verhaegen@moesp.org).

S. J. Elliott is with the Institute of Sound and Vibration Research, Highfield, Southampton S017 1BJ, U.K. (e-mail: sje@isvr.soton.ac.uk).

Digital Object Identifier 10.1109/TSP.2007.896086

TABLE I
SUMMARY OF ROBUST FXLMS ALGORITHMS (SINGLE-CHANNEL CASE)

ALGORITHM	ADDITIONAL UPDATE ON $\hat{\boldsymbol{\theta}}(k)$ $(\mathbf{x}(k) = \phi \otimes x(k))$	COST FUNCTION MINIMIZED $(\widehat{G} = G)$	SPR STABILITY CONDITION (FOR $0 \leq \omega \leq 2\pi$)
LMS	None	$\mathrm{E}[e^2(k)]$	$\mathrm{Re}\left[\widehat{G}^H(e^{\jmath\omega})G(e^{\jmath\omega})\right] > 0$
Leaky	$\rho\hat{\boldsymbol{\theta}}(k)$	$\mathrm{E}[e^2(k)] + \rho\boldsymbol{\theta}^T\boldsymbol{\theta}$	$\mathrm{Re}\left[\widehat{G}^H(e^{\jmath\omega})G(e^{\jmath\omega})\lvert P_{xs}(e^{\jmath\omega})\rvert^2 + \rho\right] > 0$
Effort weighting	$\rho\mathbf{x}(k)u(k)$	$\mathrm{E}[e^2(k)] + \rho\mathrm{E}[u^2(k)]$	$\mathrm{Re}\left[\widehat{G}^H(e^{\jmath\omega})G(e^{\jmath\omega}) + \rho\right] > 0$
Output weighting	$\rho\{\widehat{G}\mathbf{x}(k)\}\{\widehat{G}u(k)\}$	$\mathrm{E}[e^2(k)] + \rho\mathrm{E}\{[Gu(k)]^2\}$	$\mathrm{Re}\left[\widehat{G}^H(e^{\jmath\omega})G(e^{\jmath\omega}) + \rho\lvert\widehat{G}(e^{\jmath\omega})\rvert^2\right] > 0$
Uncertainty weighting	$\{\widehat{\Delta G}\mathbf{x}(k)\}\{\widehat{\Delta G}u(k)\}$	$\mathrm{E}[e^2(k)] + \mathrm{E}\{[\widehat{\Delta G}u(k)]^2\}$	$\mathrm{Re}\left[\widehat{G}^H(e^{\jmath\omega})G(e^{\jmath\omega}) + \lvert\widehat{\Delta G}(e^{\jmath\omega})\rvert^2\right] > 0$
General robust	$\{F\mathbf{x}(k)\}\{Fu(k)\}$	$\mathrm{E}[e^2(k)] + \mathrm{E}\{[Fu(k)]^2\}$	$\mathrm{Re}\left[\widehat{G}^H(e^{\jmath\omega})G(e^{\jmath\omega}) + \lvert F(e^{\jmath\omega})\rvert^2\right] > 0$

$\mathbf{F}(z)$ that guarantees asymptotic convergence for all model errors contained in a particular given norm bounded set. A similar procedure is taken to derive the filter $\mathbf{F}(z)$ that, in addition to asymptotic convergence, also prevents critical behavior. Critical behavior is the initial increase of the error before decaying to zero. It is a transient effect that may happen when the model is not perfect, even though the adaptive algorithm is stable, cf., [1]. The derivations build further on the analysis in [1]. As in [1], the focus is on convergence in mean, the analysis of mean-square convergence, i.e., the power of the error is bounded, involves a lot of algebra for the general multichannel case, and is postponed for future research. Section IV illustrates the robust methods by a simulation example.

II. THE ROBUST FXLMS ALGORITHM AND ITS CONVERGENCE BEHAVIOR

A. Comments on Notation

Let us briefly recapitulate the notation and problem setting from [1]. The disturbance source signal $\mathbf{s}$ is a discrete time m_s-dimensional wide-sense stationary zero mean white noise signal with unit covariance. The primary disturbance path is denoted by $\mathbf{P_{es}}(z) \in \mathbb{R}H_\infty^{m_e \times m_s}$, where $\mathbb{R}H_\infty^{m_y \times m_x}$ is the set of all $m_y \times m_x$ causal and asymptotically stable systems and z the unit forward shift operator. The secondary disturbance path is denoted by $\mathbf{G}(z) \in \mathbb{R}H_\infty^{m_e \times m_u}$ and the reference signal path by $\mathbf{P_{xs}} \in \mathbb{R}H_\infty^{m_x \times m_s}$ (both should have no zeros on the unit-circle). The residual disturbance signal $\mathbf{e}$ and the measured reference signal $\mathbf{x}$ are determined by the relations

$$\mathbf{e}(k) = \mathbf{d}(k) + \mathbf{y}(k) = \mathbf{P_{es}}\mathbf{s}(k) + \mathbf{G}\mathbf{u}(k) \tag{1}$$

$$\mathbf{x}(k) = \mathbf{P_{xs}}\mathbf{s}(k) \tag{2}$$

respectively, where $\mathbf{P_{es}}\mathbf{s}(k)$ denotes the filtering of $\mathbf{s}(k)$ by $\mathbf{P_{es}}$ and $\mathbf{u}(k) \in \mathbb{R}^{m_u}$ is a control signal that is to be determined using measurements $\mathbf{x}(k-i)$, $i \geq 0$. More specifically, $\mathbf{u}(k) = \mathbf{W}\mathbf{x}(k)$ where $\mathbf{W}$ is a FIR operator

$$\mathbf{u}(k) = \sum_{i=0}^{n_w-1} \mathbf{W}_i\mathbf{x}(k-i), \qquad \mathbf{W}_i \in \mathbb{R}^{m_u \times m_x} \tag{3}$$

with $n_w \geq 1$ taps. For ease of notation, we will collect the controller coefficients in a parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_\mathbf{0}^T \ \boldsymbol{\theta}_\mathbf{1}^T \ \cdots \ \boldsymbol{\theta}_{\mathbf{n_w-1}}^T]^T \in \mathbb{R}^{n_w m_u m_x}$, where $\boldsymbol{\theta}_i = \mathrm{vec}(\mathbf{W}_i^T) \in \mathbb{R}^{m_u m_x}$ and $\mathrm{vec}(\mathbf{A})$ denotes the vector stacking of the columns of $\mathbf{A}$. Further, define the regression matrix $\mathbf{R}(k) \in \mathbb{R}^{n_w m_u m_x \times m_e}$ as

$$\mathbf{R}(k) = \boldsymbol{\phi} \otimes \mathbf{G}^T \otimes \mathbf{x}(k) \tag{4}$$

where $\mathbf{G}^T \otimes \mathbf{x}(k)$ is a Kronecker product filtering and

$$\boldsymbol{\phi}(z) = [1 \quad z^{-1} \quad \cdots \quad z^{-n_w+1}]^T. \tag{5}$$

With these definitions, it can be verified, after some algebra, that $\mathbf{e}(k) = \mathbf{d}(k) + \mathbf{R}^T(k)\boldsymbol{\theta}$.

B. Robust FXLMS Algorithm

The robust FXLMS algorithm is derived similar to the FXLMS algorithm, cf., [1], using the robust cost function

$$J_{rob} = \mathrm{E}\left[\mathbf{e}^T(k)\mathbf{e}(k)\right] + \mathrm{E}\left\{[\mathbf{F}\mathbf{u}(k)]^T[\mathbf{F}\mathbf{u}(k)]\right\} \tag{6}$$

where $\mathbf{F} \in \mathbb{R}H_\infty^{m_f \times m_u}$ to be designed and $m_f \geq 1$ the number of rows of $\mathbf{F}$. By defining $\mathbf{e^a} = [\mathbf{e}^T \ \mathbf{u_F}^T]^T \in \mathbb{R}^{m_e+m_f}$ where $\mathbf{u_F} = \mathbf{F}\mathbf{u}$, we have $J_{rob} = \mathrm{E}[\mathbf{e}^{\mathbf{a}T}(k)\mathbf{e^a}(k)]$. Hence, the robust FXLMS algorithm can be considered as the FXLMS algorithm applied on a system with augmented residual channels. This involves that the secondary plant $\mathbf{G}$ is replaced by $[\mathbf{G}^T \ \mathbf{F}^T]^T$ and the residual signal $\mathbf{e}$ by $\mathbf{e^a}$. The augmented regression matrix $\mathbf{R^a}(k) \in \mathbb{R}^{n_w m_u m_x \times m_e + m_f}$ is defined as

$$\mathbf{R^a}(k) = \boldsymbol{\phi} \otimes \begin{bmatrix}\mathbf{G}\\ \mathbf{F}\end{bmatrix}^T \otimes \mathbf{x}(k) = [\mathbf{R}(k) \quad \mathbf{R_F}(k)] \tag{7}$$

where $\mathbf{R_F}(k) = \boldsymbol{\phi} \otimes \mathbf{F}^T \otimes \mathbf{x}(k)$. Then, the augmented residual signal can be written as

$$\mathbf{e^a}(k) = \begin{bmatrix}\mathbf{d}(k)\\ \mathbf{0}\end{bmatrix} + \mathbf{R^a}(k)^T\boldsymbol{\theta}$$

and the gradient of J_{rob} with respect to $\boldsymbol{\theta}$ as

$$\frac{\mathrm{d}J_{rob}}{\mathrm{d}\boldsymbol{\theta}} = 2\mathrm{E}\left[\mathbf{R^a}(k)\mathbf{e^a}(k)\right]$$

Setting the gradient $\mathrm{d}J_{rob}/\mathrm{d}\boldsymbol{\theta}$ to zero and solving for $\boldsymbol{\theta}$ yields the optimal robust solution $\boldsymbol{\theta}_{\mathrm{rob}}$

$$\boldsymbol{\theta}_{\mathrm{rob}} = -(\mathbf{T_{RR}} + \mathbf{T_{FF}})^{-1}\mathbf{T_{Rd}} \tag{8}$$

where $\mathbf{T_{RR}} = \mathrm{E}[\mathbf{R}(k)\mathbf{R}^T(k)]$, $\mathbf{T_{FF}} = \mathrm{E}[\mathbf{R_F}(k)\mathbf{R_F}^T(k)]$ and $\mathbf{T_{Rd}} = \mathrm{E}[\mathbf{R}(k)\mathbf{d}(k)]$. Note, that $\mathbf{T_{FF}}$ will be nonnegative definite and "pushes" $\boldsymbol{\theta}_{\text{rob}}$ away from the optimal nominal solution $\boldsymbol{\theta}_{\text{opt}} = -\mathbf{T_{RR}}^{-1}\mathbf{T_{Rd}}$ in the direction of zero.

Now, the robust FXLMS algorithm, that updates a current estimate $\hat{\boldsymbol{\theta}}(k)$ of $\boldsymbol{\theta}_{\text{rob}}$ in the negative gradient direction estimated in LMS sense, can be stated as

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) - \mu\widehat{\mathbf{R}}(k)\mathbf{e}(k) - \mu\mathbf{R_F}(k)\mathbf{u_F}(k), \hat{\boldsymbol{\theta}}(0) = 0 \quad (9)$$

where $\widehat{\mathbf{R}}(k) = \boldsymbol{\phi} \otimes \widehat{\mathbf{G}}^T \otimes \mathbf{x}(k)$ is the estimate of $\mathbf{R}(k)$. The additional term $\mu\mathbf{R_F}(k)\mathbf{u_F}(k)$ in (9) determines the difference between the FXLMS and the robust FXLMS algorithm.

Under the assumption that μ is small enough, such that $\mathbf{R}(k)$, $\widehat{\mathbf{R}}(k)$, $\mathbf{R_F}(k)$, $\mathbf{e}(k)$ and $\mathbf{u_F}(k)$ can be supposed as "semi-stationary" and statistically independent of $\hat{\boldsymbol{\theta}}(k)$, the mean update is given by

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}(k+1)\right] = (\mathbf{I} - \mu\mathbf{T_{\hat{R}R}} - \mu\mathbf{T_{FF}})\mathrm{E}\left[\hat{\boldsymbol{\theta}}(k)\right] - \mu\mathbf{T_{\hat{R}d}} \quad (10)$$

with $\mathbf{T_{\hat{R}R}} = \mathrm{E}[\widehat{\mathbf{R}}(k)\mathbf{R}^T(k)]$ and $\mathbf{T_{\hat{R}d}} = \mathrm{E}[\widehat{\mathbf{R}}(k)\mathbf{d}(k)]$. By assuming μ is small enough, also the remaining variation in $\hat{\boldsymbol{\theta}}(k)$ after convergence due to nonzero μ, can be neglected.

Note, that the derivation and the analysis for leaky FXLMS is very similar, when replacing the cost function by $J_{rob} = \mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)] + \rho\boldsymbol{\theta}^T\boldsymbol{\theta}$, using $\mathbf{T_{FF}} = \rho\mathbf{I}$ in the optimal solution and the resulting update equation is given by $\hat{\boldsymbol{\theta}}(k+1) = (1-\mu\rho)\hat{\boldsymbol{\theta}}(k) - \mu\widehat{\mathbf{R}}(k)\mathbf{e}(k)$, c.f., e.g., [7].

It is observed that the dynamics of the mean update (10) of the robust FXLMS algorithm is the same as for the FXLMS algorithm where $\mathbf{T_{\hat{R}R}}$ is replaced by $\mathbf{T_{\hat{R}R}} + \mathbf{T_{FF}}$. Because $\mathbf{T_{FF}}$ has nonnegative real eigenvalues, the eigenvalues of $\mathbf{T_{\hat{R}R}} + \mathbf{T_{FF}}$ will usually have larger real part than the eigenvalues of $\mathbf{T_{\hat{R}R}}$ such that the condition for convergence is relaxed. However, except for $\mathbf{T_{FF}}$ being a scalar times identity, it is difficult or even impossible to relate the eigenvalues of $\mathbf{T_{\hat{R}R}}$ and $\mathbf{T_{FF}}$ to the eigenvalues of $\mathbf{T_{\hat{R}R}} + \mathbf{T_{FF}}$. To say more on the relation between $\mathbf{F}$ (that determines $\mathbf{T_{FF}}$) and asymptotic convergence, the asymptotic analysis of [1], will be applied.

C. Convergence Analysis of Robust FXLMS

Because the robust FXLMS algorithm is equivalent to the FXLMS algorithm applied on the augmented system, in [1, Theorem 2] can be directly applied to derive conditions for asymptotic convergence. This yields the following theorem, which proof follows by transformation of [1, Theorem 2] to the augmented plant.

Theorem 1: Let $\sigma_{\text{ess}}(\mathbf{T_{GF}}) = \{\lambda \in \mathbb{C} : \widehat{\mathbf{G}}(e^{\jmath\omega})^T\mathbf{G}^*(e^{\jmath\omega}) + \mathbf{F}(e^{\jmath\omega})^T\mathbf{F}^*(e^{\jmath\omega}) - \lambda\mathbf{I}$ is singular, $0 \le \omega \le 2\pi\}$ be the essential spectrum of $\mathbf{T_{GF}}$, defined in [1].

i) If $\mathrm{Re}(\lambda) > 0$ for all $\lambda \in \sigma_{\text{ess}}(\mathbf{T_{GF}})$, then for each $n_w \ge 1$ there exists a step-size $\mu > 0$ (sufficiently small) such that the robust FXLMS update equation (9) asymptotically converges in mean.

ii) If $\mathrm{Re}(\lambda) \ge 0$ for $\lambda \in \sigma_{\text{ess}}(\mathbf{T_{GF}})$ and for at least one $\lambda \in \sigma_{\text{ess}}(\mathbf{T_{GF}})$ it holds that $\mathrm{Re}(\lambda) = 0$, then for each $n_w \ge 1$ there exists a step-size $\mu > 0$ (sufficiently small) such that the robust FXLMS update equation (9) does not diverge.

iii) If $\mathrm{Re}(\lambda) < 0$ for at least some $\lambda \in \sigma_{\text{ess}}(\mathbf{T_{GF}})$, then there exists a (sufficiently large) $N_w \ge 1$ such that for each $n_w \ge N_w$ the robust FXLMS update equation (9) diverges for any step-size $\mu > 0$. □

Hence, to determine asymptotic convergence, the real part of the eigenvalues of the matrix $\widehat{\mathbf{G}}(e^{\jmath\omega})^T\mathbf{G}^*(e^{\jmath\omega}) + \mathbf{F}(e^{\jmath\omega})^T\mathbf{F}^*(e^{\jmath\omega})$ have to be examined for $0 \le \omega \le 2\pi$ or equivalently, the real part of the eigenvalues of

$$\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega}) + \mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega}), \qquad 0 \le \omega \le 2\pi. \quad (11)$$

Though the matrix $\mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega})$ is nonnegative definite (usually positive definite) for $0 \le \omega \le 2\pi$ it cannot be said that the minimal real part of the eigenvalues of $\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega}) + \mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega})$ is always greater than the minimal real part of the eigenvalues of $\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega})$ (cf., [8, p. 1]). Though, this would be usually the case. For $\mathbf{F}(e^{\jmath\omega})$ equals a scalar times diagonal, $\mathbf{F} = \rho\mathbf{I}$, (as in control effort weighting) the real part of the eigenvalues are guaranteed to be increased by an amount depending on the value of ρ.

By appropriately choosing $\mathbf{F}$ the condition (11) is strictly positive real (SPR) can be used to derive the conditions for asymptotic convergence for the case of control effort weighting, output weighting and uncertainty weighting.

III. Minimal Regularization for Norm Bounded Uncertainty

The following question naturally arises: what is the best choice of $\mathbf{F}$ to ensure asymptotic convergence for uncertain plants but that does not degrade nominal performance. To answer this question, we will assume $\mathbf{G}$ is such that the model error $\boldsymbol{\Delta}\mathbf{G} = \mathbf{G} - \widehat{\mathbf{G}}$ is contained in a bounded norm set

$$\mathcal{S}_{\boldsymbol{\gamma}} = \left\{\boldsymbol{\Delta}\mathbf{G} \in \mathbb{R}H^{m_e \times m_u} : \|\boldsymbol{\Delta}\mathbf{G}(e^{\jmath\omega})\|_2 \le \gamma(\omega), 0 \le \omega \le 2\pi\right\} \quad (12)$$

where γ is a given function of ω that determines the amount of uncertainty and $\|\mathbf{A}\|_2$ denotes the induced 2-norm of $\mathbf{A}$, i.e., its maximal singular value. Note, that the elements of $\mathcal{S}_\gamma$ are in $\mathbb{R}H^{m_e \times m_u}$, the set of all linear (stable or unstable) systems, rather then $\mathbb{R}H_\infty^{m_e \times m_u}$. Considering this more general class allows us to consider $\boldsymbol{\Delta}\mathbf{G}(e^{\jmath\omega})$ at frequency ω independent of the other frequencies. In practice, $\boldsymbol{\Delta}\mathbf{G}$ will however be stable.

To quantify some optimality criterion, one would be interested for example in the worst case performance, and design a filter $\mathbf{F}$ according to the problem

$$\min_{\mathbf{F}} \max_{\boldsymbol{\Delta}\mathbf{G} \in \mathcal{S}_{\boldsymbol{\gamma}}} \mathrm{E}\left[\mathbf{e}^T(k)\mathbf{e}(k)\right],$$

subject to the robust FXLMS algorithm (9)

and the eigenvalues of (11) being positive real. (13)

However, it is very difficult to determine the worst-case performance $\max_{\boldsymbol{\Delta}\mathbf{G} \in \mathcal{S}_\gamma} \mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)] = \max_{\boldsymbol{\Delta}\mathbf{G} \in \mathcal{S}_\gamma} \hat{\boldsymbol{\theta}}_{\text{rob}}^T\mathbf{T_{RR}}\hat{\boldsymbol{\theta}}_{\text{rob}} + 2\hat{\boldsymbol{\theta}}_{\text{rob}}^T\mathbf{T_{Rd}} + \mathrm{E}[\mathbf{d}^T\mathbf{d}(k)]$ where $\hat{\boldsymbol{\theta}}_{\text{rob}} = -(\mathbf{T_{\hat{R}R}} + \mathbf{T_{FF}})^{-1}\mathbf{T_{\hat{R}d}}$ the solution to which the robust FXLMS algorithm converges. Therefore, we will allow a simplification of the problem. It is known that $\mathbf{T_{FF}}$ will result in a bias on $\hat{\boldsymbol{\theta}}_{\text{rob}}$, such that under the nominal condition

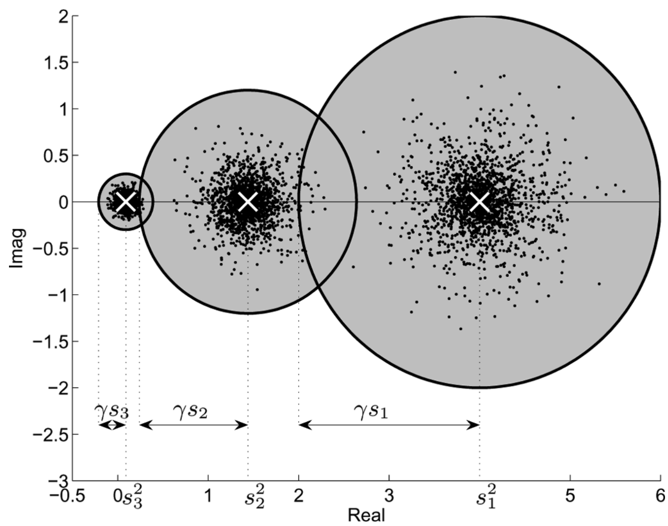

Fig. 1. Eigenvalues of $(\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A}$, $\mathbf{A}=\mathrm{diag}(2,1.2,0.3)$, for 2000 realizations of $\boldsymbol{\Delta}$ such that $\|\boldsymbol{\Delta}\|_2 \leq 1$. The shaded region indicates $\overline{\sigma}_\gamma(\mathbf{A})$.

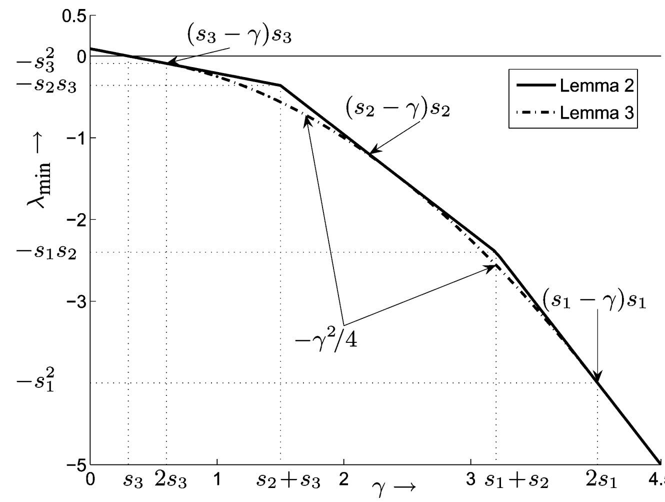

Fig. 2. Values of $\min \mathrm{Re}[\overline{\sigma}_\gamma(\mathbf{A})]$ (Lemma 2) and $\min_{\|\boldsymbol{\Delta}\|_2\leq\gamma} \min \sigma((\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A}+\mathbf{A}^H(\mathbf{A}+\boldsymbol{\Delta}))/2$ (Lemma 3) for $\mathbf{A}=\mathrm{diag}(2,1.2,0.3)$.

the solution will be suboptimal. When the norm of $\mathbf{F}$ is increased, the bias increases resulting in less performance for the nominal, i.e., $\boldsymbol{\Delta}\mathbf{G}=\mathbf{0}$, condition. Therefore, our objective will be to design $\mathbf{F}$ such that its norm (to be specified below) is minimized subject to the condition that the eigenvalues of (11) are positive real. This does not guarantee best performance for the case $\boldsymbol{\Delta}\mathbf{G}\neq\mathbf{0}$, however over-regularization is prevented. The simulation example in Section IV also shows that for this choice of $\mathbf{F}$ better performance is achieved than for leakage/effort weighting and output weighting. To design $\mathbf{F}$ in this minimum norm sense, first some results on the extreme negative real part of the eigenvalues of uncertain matrices need to be derived.

A. Some Matrix Results

In the following, $\mathbf{A}\in\mathbb{C}^{m\times n}$ and its singular value decomposition (SVD) is given by $\mathbf{A}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where $\mathbf{U}\in\mathbb{C}^{m\times m}$ and $\mathbf{V}\in\mathbb{C}^{n\times n}$ both unitary and $\boldsymbol{\Sigma}=[\mathrm{diag}(s_1,s_2,\cdots,s_n)\ 0_{m-n\times n}^T]^T$ for $m\geq n$ and $\boldsymbol{\Sigma}=[\mathrm{diag}(s_1,s_2,\cdots,s_m)\ 0_{m\times n-m}]$ for $n\geq m$ and $s_1\geq s_2\geq\cdots\geq s_p\geq 0$, $p=\min(m,n)$ the singular values of $\mathbf{A}$. Let $\overline{\sigma}_\gamma(\mathbf{A})$ be defined as $\overline{\sigma}_\gamma(\mathbf{A}) := \bigcup_{\|\boldsymbol{\Delta}\|_2\leq\gamma}\sigma((\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A})$, i.e., the union of the sets of eigenvalues of $(\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A}$ for all $\boldsymbol{\Delta}$ such that $\|\boldsymbol{\Delta}\|_2\leq\gamma$. The following two lemmas will be used in Section III-B to derive the minimal regularization filter $\mathbf{F}(z)$ that guarantees asymptotic convergence. Their proofs are given in the Appendix.

Lemma 1: Let $\mathbf{A}\in\mathbb{C}^{m\times n}$ and $\gamma>0$, then $\overline{\sigma}_\gamma(\mathbf{A})=\{\lambda\in\mathbb{C}:|\lambda-s_i^2|\leq\gamma s_i,\text{ for some } i=1,\cdots,p\}$. □

Lemma 2: Let $\mathbf{A}\in\mathbb{C}^{m\times n}$ and $\gamma>0$, then $\min\mathrm{Re}[\overline{\sigma}_\gamma(\mathbf{A})]=(s^o-\gamma)s^o$ where $s^o=\arg\min_i|s_i-\gamma/2|$. □

Fig. 1 illustrates Lemma 1, where $\gamma=1$ and $\mathbf{A}=\mathrm{diag}(2,1.2,0.3)$, such that $s_1=2$, $s_2=1.2$ and $s_3=0.3$. The eigenvalues of $(\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A}$ lay in the disks with centers s_i^2 and radius γs_i for $i=1,2,3$. From this figure, it is clear that the value of $\min\mathrm{Re}[\overline{\sigma}_\gamma(\mathbf{A})]$ is given by $s_3^2-\gamma s_3$, which is in agreement with Lemma 2.

From Lemma 1 and 2 it follows that $\overline{\sigma}_\gamma(\mathbf{A})$ lies in the open right half plane if $0<\gamma<s_p$ and $\overline{\sigma}_\gamma(\mathbf{A})$ contains elements with negative real part if $s_p<\gamma$. The following lemma will be used in Section III-C to derive the minimal regularization filter $\mathbf{F}(z)$ that also guarantees noncritical behavior. Its proof is given in the Appendix.

Lemma 3: Let $\mathbf{A}\in\mathbb{C}^{m\times n}$ and $\gamma>0$. If $m<n$ let $s_{m+1}=\cdots=s_n=0$ be zero-valued singular values. Let

$$\lambda_{\min,\gamma}(\mathbf{A})=\min_{\|\boldsymbol{\Delta}\|_2\leq\gamma}\min\sigma\left[(\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A}+\mathbf{A}^H(\mathbf{A}+\boldsymbol{\Delta})\right]$$

Then

$$\lambda_{\min,\gamma}(\mathbf{A})=\begin{cases}2(s_n-\gamma)s_n & 0<\gamma/2\leq s_n\\ -\gamma^2/2 & s_n\leq\gamma/2\leq s_1\\ 2(s_1-\gamma)s_1 & s_1\leq\gamma/2.\end{cases}$$

□

From Lemma 3 it follows that $\lambda_{\min,\gamma}(\mathbf{A})>0$ if $0<\gamma<s_n$ and $\lambda_{\min,\gamma}(\mathbf{A})<0$ if $s_n<\gamma$.

Combining the results of Lemma 2 and 3 it follows that $(1/2)\min_{\|\boldsymbol{\Delta}\|_2\leq\gamma}\min\sigma[(\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A}+\mathbf{A}^H(\mathbf{A}+\boldsymbol{\Delta})]\leq\min\mathrm{Re}[\overline{\sigma}_\gamma(\mathbf{A})]$, which is intuitive since the condition for noncritical behavior is stronger than the condition for asymptotic convergence (equality holds for $\gamma\leq 2s_n$ and $\gamma\geq 2s_1$). Fig. 2 illustrates the results of Lemma 2 and 3, and shows that the results of both lemmas deviate especially for γ in the regions around s_i+s_{i+1} ($i=1,2$, etc.).

B. Minimal Regularization for Guaranteed Convergence

Let $s_1(\omega)\geq s_2(\omega)\geq\cdots\geq s_{m_u}(\omega)$ be singular values of $\widehat{\mathbf{G}}(e^{\jmath\omega})$. Then with the identifications $\widehat{\mathbf{G}}(e^{\jmath\omega})\rightarrow\mathbf{A}$,

$\Delta\mathbf{G}(e^{\jmath\omega}) \rightarrow \mathbf{\Delta}$ and $\gamma(\omega) \rightarrow \gamma$ Lemma 2 says that for $\mathbf{G} = \widehat{\mathbf{G}} + \Delta\mathbf{G}$ and for each $\omega \in [0, 2\pi]$

$$\min_{\Delta\mathbf{G}\in\mathcal{S}_\gamma} \operatorname{Re}\left\{\sigma\left[\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega})\right]\right\} = [s^o(\omega) - \gamma(\omega)]\, s^o(\omega) \tag{14}$$

where $s^o(\omega) = \arg\min_i |s_i(\omega) - \gamma/2|$. If the value (14) is negative for some $\omega \in [0, 2\pi]$, then according to Theorem 1 the nominal (i.e., with $\mathbf{F} = \mathbf{0}$) FXLMS algorithm will diverge for sufficiently large number of adaptive filter coefficients. When $\mathbf{F}$ is chosen such that $\mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega}) = \{\epsilon - [s^o(\omega) - \gamma(\omega)]s^o(\omega)\}\mathbf{I}$, where $\epsilon > 0$ an arbitrarily small user chosen number, then the value of $\min \operatorname{Re}[\sigma(\mathbf{T_{GF}})]$ will be no smaller then ϵ. Hence, for this choice of $\mathbf{F}$ the robust FXLMS will be asymptotically converging according to Theorem 1. The result is summarized in the following theorem.

Theorem 2: Let $s_1(\omega) \geq s_2(\omega) \geq \cdots \geq s_{m_u}(\omega)$ be singular values of $\widehat{\mathbf{G}}(e^{\jmath\omega})$. For any $\epsilon > 0$, let $\mathbf{F}(e^{\jmath\omega}) \in \mathbb{C}^{m_u\times m_u}$ be such that

$$\mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega}) = \begin{cases}\{\epsilon - [s^o(\omega) - \gamma(\omega)]\, s^o(\omega)\}\,\mathbf{I}, & \text{if } \gamma(\omega) > s^o(\omega)\\ \epsilon\mathbf{I}, & \text{otherwise}\end{cases}$$

then $\min \operatorname{Re}[\sigma_{\mathrm{ess}}(\mathbf{T_{GF}})] \geq \epsilon$ such that the robust FXLMS algorithm (9) asymptotically converges in mean for sufficiently small step-size $\mu > 0$. □

Note, that $\mathbf{F}(e^{\jmath\omega}) = \sqrt{\epsilon}\mathbf{I}$ (in the case where $\gamma(\omega) = 0$) induces an effort weighting, and thus ϵ can also be used to combine effort weighting together with minimal norm regularization, cf., Table I.

So, given a particular bound $\gamma(\omega)$ Theorem 2 can be applied to compute the spectrum $\mathbf{F}^H(^{\jmath\omega})\mathbf{F}(e^{\jmath\omega})$ such that $\|\mathbf{F}(e^{\jmath\omega})\|_2$ is minimal. To obtain an explicit expression for $\mathbf{F}(e^{\jmath\omega})$ a parametric spectral factorization has to be made. The good news is that only a scalar (single-channel) spectral factorization has to be computed, even in the matrix $(m_u > 1)$ case. However, the spectrum $\mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega})$ obtained by Theorem 2 is not guaranteed to be *rational*, i.e., there may not exist a *finite* dimensional spectral factor with real coefficients $\widehat{\mathbf{F}}(z) \in \mathbb{R}H_\infty^{m_f\times m_u}$ such that $\widehat{\mathbf{F}}^H(e^{\jmath\omega})\widehat{\mathbf{F}}(e^{\jmath\omega}) = \mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega})$ for all $\omega \in [0, 2\pi]$. In this case the spectrum $\mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega})$ has to be approximated. There are various techniques for computing the (approximate) spectral factor $\widehat{\mathbf{F}}(e^{\jmath\omega})$ (such as based on the cepstrum). In our simulations, we obtained good results with the Power Spectrum SubSpace Identification (PSSSID) algorithm [9] that estimates a state-space realization of the (minimum-phase) spectral factor. The algorithm is based upon subspace estimation and solving a conic programming problem and is numerically well conditioned even for spectra that are close to singular.

C. Minimal Regularization for Guaranteed Noncritical Convergence Behavior

A similar approach can be taken to prevent critical behavior. To this end, the stronger condition

$$\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega}) + \mathbf{G}^H(e^{\jmath\omega})\widehat{\mathbf{G}}(e^{\jmath\omega}) + 2\mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega}) > 0 \tag{15}$$

for $\omega \in [0, 2\pi]$ need to be satisfied for all $\Delta\mathbf{G} \in \mathcal{S}_\gamma$, that is obtained by application of [1, Theorem 3] on the augmented plant. By Lemma 3, it is inferred that for each $\omega \in [0, 2\pi]$

$$\begin{aligned}&\min\nolimits_{\Delta\mathbf{G}\in\mathcal{S}_\gamma} \min \sigma\left[\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega}) + \mathbf{G}^H(e^{\jmath\omega})\widehat{\mathbf{G}}(e^{\jmath\omega})\right]\\ &\quad= \begin{cases}2\,[s_{m_u}(\omega) - \gamma(\omega)]\, s_{m_u}(\omega), & 0 < \gamma(\omega)/2 \leq s_{m_u}(\omega)\\ -\gamma(\omega)^2/2, & s_{m_u}(\omega) \leq \gamma(\omega)/2 \leq s_1(\omega)\\ 2\,[s_1(\omega) - \gamma(\omega)]\, s_1(\omega), & s_1(\omega) \leq \gamma(\omega)/2\end{cases}\end{aligned} \tag{16}$$

which is negative if $\gamma(\omega) > s_{m_u}(\omega)$. Hence, because $\min_{\Delta\mathbf{G}\in\mathcal{S}_\gamma} \min \sigma[\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega}) + \mathbf{G}^H(e^{\jmath\omega})\widehat{\mathbf{G}}(e^{\jmath\omega})]$ is known, we can choose $\mathbf{F}$ such that (15) is guaranteed to be satisfied and the following theorem can be stated.

Theorem 3: Let $s_1(\omega) \geq s_2(\omega) \geq \cdots \geq s_{m_u}(\omega)$ be singular values of $\widehat{\mathbf{G}}(e^{\jmath\omega})$ and $\epsilon > 0$. If $\gamma(\omega) \leq s_{m_u}(\omega)$ let $\mathbf{F}(e^{\jmath\omega}) = \sqrt{\epsilon}\mathbf{I}$, else let $\mathbf{F}(e^{\jmath\omega})$ be such that for all $\omega \in [0, 2\pi]$

$$\mathbf{F}^H(e^{\jmath\omega})\mathbf{F}(e^{\jmath\omega}) = \begin{cases}\{\epsilon/2 - [s_{m_u}(\omega) - \gamma(\omega)]\, s_{m_u}(\omega)\}\,\mathbf{I}, & s_{m_u}(\omega) \leq \gamma(\omega) \leq 2s_{m_u}(\omega)\\ \{\epsilon/2 + \gamma(\omega)^2/2\}\,\mathbf{I}, & 2s_{m_u}(\omega) \leq \gamma(\omega) \leq 2s_1(\omega)\\ \{\epsilon/2 - [s_1(\omega) - \gamma(\omega)]\, s_1(\omega)\}\,\mathbf{I}, & 2s_1(\omega) \leq \gamma(\omega)\end{cases}$$

then (15) is satisfied for all $\Delta\mathbf{G} \in \mathcal{S}_\gamma$ such that the robust FXLMS algorithm (9) asymptotically converges in mean for sufficiently small step-size $\mu > 0$ and does not show critical behavior. □

As aforementioned, it may be impossible to find a rational spectral factor $\mathbf{F}(z)$ such that Theorem 3 holds, and a rational approximation $\widehat{\mathbf{F}}(z)$ has to be made.

If the uncertainty bound $\gamma(\omega)$ is such that $\gamma(\omega) > s_{m_u}(\omega)$ for some ω, then observe that for $\mathbf{F} = 0$ there always exists a $\Delta\mathbf{G} \in \mathcal{S}_\gamma$ such that the condition for asymptotic convergence is not satisfied as well as the condition for noncritical behavior. For both conditions there exists a "worst-case" model error $\Delta\mathbf{G}$. It is illustrative to determine this worst-case $\Delta\mathbf{G}$ for the single-channel case. Note, that in the single-channel case the conditions for asymptotic convergence and noncritical behavior are equivalent. The SVD of scalar systems $\widehat{\mathbf{G}}$ is given by $\widehat{\mathbf{G}}(e^{\jmath\omega}) = u(\omega)s_1(\omega)v^*(\omega)$ where $u(\omega) = 1$, $s_1(\omega) = |\widehat{\mathbf{G}}(e^{\jmath\omega})|$ and $v^*(\omega) = \widehat{\mathbf{G}}(e^{\jmath\omega})/|\widehat{\mathbf{G}}(e^{\jmath\omega})|$. Then it can be easily verified that the $\Delta\mathbf{G} \in \mathcal{S}_\gamma$ that minimizes (14) as well as (16) satisfies $\Delta\mathbf{G}(e^{\jmath\omega}) = -\gamma(\omega)\widehat{\mathbf{G}}(e^{\jmath\omega})/|\widehat{\mathbf{G}}(e^{\jmath\omega})|$. This model error has maximal magnitude and also maximal phase-difference with $\widehat{\mathbf{G}}(e^{\jmath\omega})$.

IV. Simulation Results

To illustrate and support the results of this paper let us consider the following (academic) example, where

$$\mathbf{P_{es}}(z) = \begin{bmatrix}\frac{z+0.85}{1.86z^2} & 0\\ 0 & \frac{z-0.85}{1.86z^2}\end{bmatrix} \quad \mathbf{P_{xs}}(z) = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

$$\widehat{\mathbf{G}}(z) = \begin{bmatrix}\frac{z^2+z+0.1}{2z^2} & 0\\ 0 & \frac{z^2-z+0.1}{2z^2}\end{bmatrix} \quad \mathbf{G}(z) = \widehat{\mathbf{G}}(z) + \Delta\mathbf{G}(z)$$

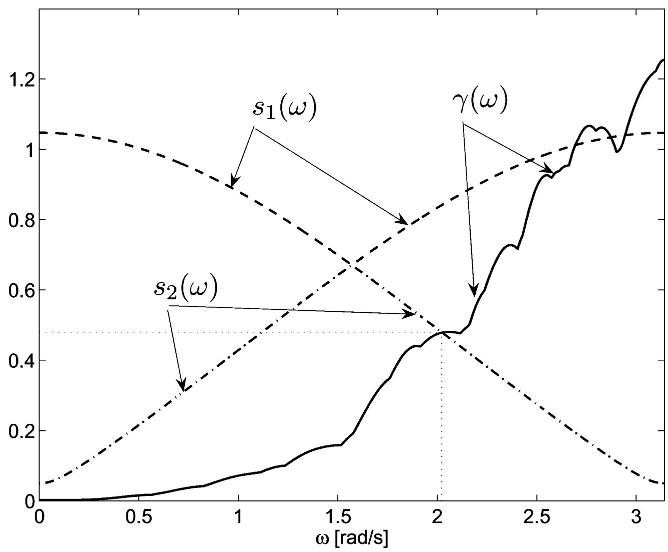

Fig. 3. Worst case uncertainty bound $\gamma(\omega)$ on $\Delta\mathbf{G}(e^{\jmath\omega})$ together with the largest and smallest singular value of $\widehat{\mathbf{G}}(e^{\jmath\omega})$, denoted by $s_1(\omega)$ and $s_2(\omega)$, respectively.

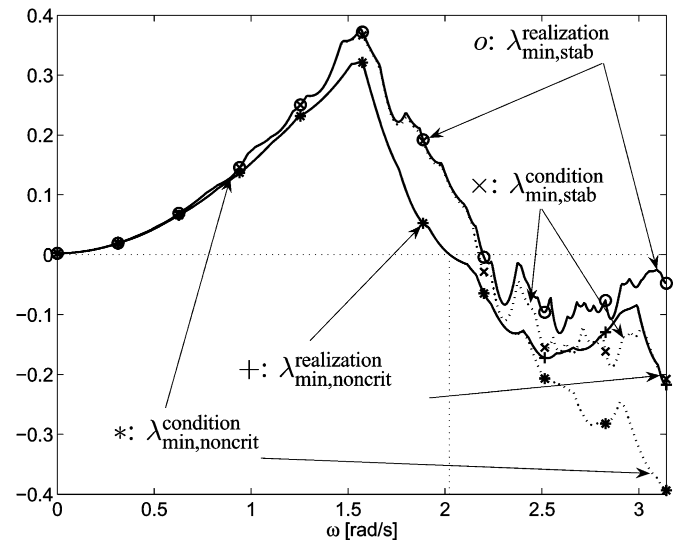

Fig. 4. $\lambda^{\text{realization}}_{\text{min,noncrit}}$ and $\lambda^{\text{realization}}_{\text{min,stab}}$, the minimal real part of the eigenvalues of $\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega})$ and $(1/2)(\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega}) + \mathbf{G}^H(e^{\jmath\omega})\widehat{\mathbf{G}}(e^{\jmath\omega}))$ for 100 realizations of $\Delta\mathbf{G}$ respectively, together with $\lambda^{\text{condition}}_{\text{min,noncrit}}$ and $\lambda^{\text{condition}}_{\text{min,stab}}$ that are obtained by minimizing over the whole uncertainty set $\mathcal{S}_\gamma$.

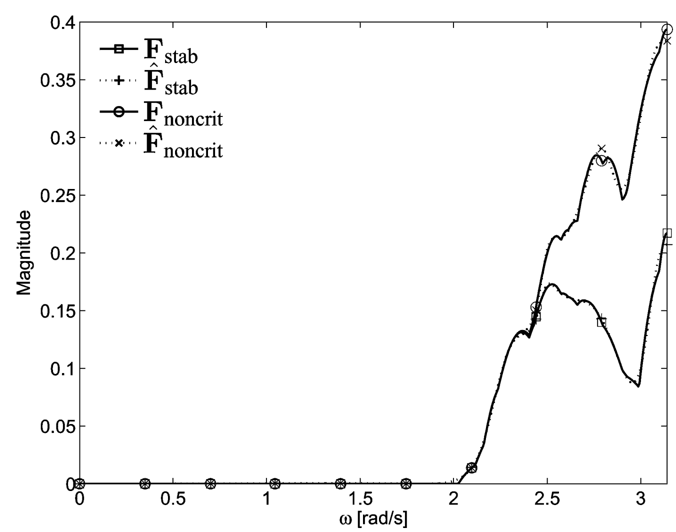

Fig. 5. Magnitude of the diagonals of the regularization filters $\mathbf{F}_{\text{stab}}$ and $\mathbf{F}_{\text{noncrit}}$ together with their rational approximations.

and $\Delta\mathbf{G}(z)$ is a full 2×2 transfer-function matrix, which magnitude is dominant at high frequencies. Each i, jth element $\Delta\mathbf{G}_{i,j}(z)$ of $\Delta\mathbf{G}(z)$ is determined as follows. Let $p(z) = (0.06746z^2 - 0.1349z + 0.06746)/(z^2 + 1.143z + 0.4128)$ be a second order high-pass Butterworth filter with cut-on frequency at 0.8π rad/s. Let $\delta(z) = \sum_{i=0}^{19} \delta_i z^{-i}$ where the coefficients $\delta_i \in \mathbb{R}$ are realizations of a Gaussian distributed random variable with zero mean and variance 0.04. Then the entries $\Delta\mathbf{G}_{i,j}(z)$ are obtained by convolution of $\delta(z)$ and the 20-taps FIR approximation of $p(z)$, resulting in a realization of a (stable) random $\Delta\mathbf{G}(z)$ that is dominant at high frequencies. The off-diagonal elements of $\Delta\mathbf{G}(z)$ cause a cross-coupling, that distorts the decentralization of the filtering problem.

The bound $\gamma(\omega)$ is approximated on the basis of the worst-case realization of each $\Delta\mathbf{G}(e^{\jmath\omega})$, $0 \leq \omega \leq 2\pi$, using 100 realizations of $\Delta\mathbf{G}(z)$. The resulting $\gamma(\omega)$ is depicted by the solid curve in Fig. 3. The figure also shows the largest and smallest singular value of $\widehat{\mathbf{G}}(e^{\jmath\omega})$, denoted by $s_1(\omega)$ and $s_2(\omega)$ respectively. Observe that for $\omega > 2$ rad/s it holds that $\gamma(\omega) > s_2(\omega)$, such that there exists $\Delta\mathbf{G}(e^{\jmath\omega})$ bounded by $\gamma(\omega)$ for which the conditions for asymptotic convergence and noncritical behavior are *not* satisfied. For each of the 100 realizations of $\Delta\mathbf{G}$ the minimal real part of the eigenvalues of $\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega})$ and the minimal eigenvalue of $[\widehat{\mathbf{G}}^H(e^{\jmath\omega})\mathbf{G}(e^{\jmath\omega})+\mathbf{G}^H(e^{\jmath\omega})\widehat{\mathbf{G}}(e^{\jmath\omega})]/2$ have been computed to evaluate the conditions for asymptotic and noncritical behavior respectively. Taking the minimal values over all 100 realizations, for each $0 \leq \omega \leq \pi$, yields the values $\lambda^{\text{realization}}_{\text{min,stab}}(\omega)$ and $\lambda^{\text{realization}}_{\text{min,noncrit}}(\omega)$, which are depicted in Fig. 4. If the 100 realizations sufficiently cover the model uncertainty set $\mathcal{S}_\gamma$ defined in (12) $\lambda^{\text{realization}}_{\text{min,stab}}(\omega)$ will be equal to the optimal value in (14), denoted by $\lambda^{\text{condition}}_{\text{min,stab}}(\omega)$. Similar, $\lambda^{\text{realization}}_{\text{min,noncrit}}(\omega)$ would be equal to the optimal value in (15) divided by 2, denoted by $\lambda^{\text{condition}}_{\text{min,noncrit}}(\omega)$. From Fig. 4, that also shows $\lambda^{\text{condition}}_{\text{min,stab}}$ and $\lambda^{\text{condition}}_{\text{min,noncrit}}$ (dotted curves), we observe that $\lambda^{\text{condition}}_{\text{min,stab}}$ is below $\lambda^{\text{realization}}_{\text{min,stab}}$ and $\lambda^{\text{condition}}_{\text{min,noncrit}}$ is below $\lambda^{\text{realization}}_{\text{min,noncrit}}$, because the 100 realizations of $\Delta\mathbf{G}$ do not cover the whole set $\mathcal{S}_\gamma$. Hence the regularizations based on $\lambda^{\text{condition}}_{\text{min,stab}}$ and $\lambda^{\text{condition}}_{\text{min,noncrit}}$ may be (a bit) too conservative. Moreover, at higher frequencies $\lambda^{\text{condition/realization}}_{\text{min,noncrit}}$ and $\lambda^{\text{condition/realization}}_{\text{min,stab}}$ deviates, such that the regularization to prevent noncritical behavior need to be stronger in this frequency range than for achieving guaranteed asymptotic convergence. Using Theorem 2 and Theorem 3 the power spectra of the filters $\mathbf{F}_{\text{stab}}$ and $\mathbf{F}_{\text{noncrit}}$ are computed that guarantee asymptotic convergence and noncritical behavior of the general robust FXLMS algorithm respectively ($\epsilon = 10^{-8}$ to guarantee a positive definite spectrum). Fig. 5 shows the magnitude of the diagonal elements of $\mathbf{F}_{\text{stab}}$ and $\mathbf{F}_{\text{noncrit}}$ together with their finite dimensional (8 order) approximations $\widehat{\mathbf{F}}_{\text{stab}}$ and $\widehat{\mathbf{F}}_{\text{noncrit}}$, determined by the PSSSID algorithm [9].

Several robust FXLMS algorithms have been evaluated: 1) leakage and effort weighting ($\mathbf{F} = \sqrt{\rho}\mathbf{I}$), which provide the

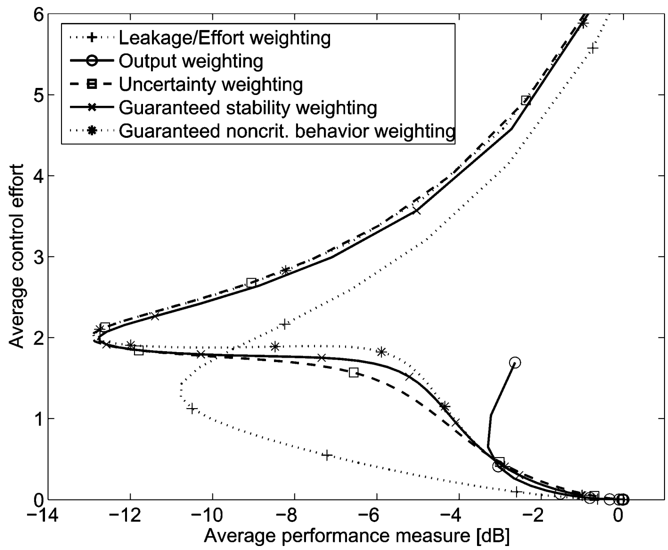

Fig. 6. Control effort versus optimal value of the performance measure achieved by various robust FXLMS algorithms averaged over 100 realizations of $\mathbf{\Delta G}$ (only depicted cases for which the adaptive algorithms are stable).

same result since $\mathbf{P_{xs}} = \mathbf{I}$; 2) output weighting ($\mathbf{F} = \sqrt{\rho}\mathbf{G}$); 3) uncertainty weighting ($\mathbf{F} = \sqrt{\rho}\widehat{\mathbf{\Delta G}}$), where $\widehat{\mathbf{\Delta G}}$ the estimated uncertainty spectral factor as in [4]; 4) guaranteed stability weighting ($\mathbf{F} = \sqrt{\rho}\hat{\mathbf{F}}_{\rm stab}$); and 5) guaranteed noncritical behavior weighting ($\mathbf{F} = \sqrt{\rho}\hat{\mathbf{F}}_{\rm noncrit}$). The weighting parameter ρ, that is used as a tuning parameter in leakage, effort, output and uncertainty weighting, has been used in the other methods as well for comparison. Though, usually for guaranteed stability and guaranteed noncritical behavior weighting ρ is set to $\rho = 1$. To get insight in the tradeoff between control effort and the value of the performance measure as well as to determine the minimal value of ρ for which the update algorithm is stable, ρ has been varied over 81 logarithmically spaced values between 10^{-4} and 10^4. The length of the adaptive filter has been set to $n_w = 30$ taps, for which the nominal disturbance suppression is -66 dB (note that $\widehat{\mathbf{G}}$ is minimum phase and $\mathbf{P_{es}}$ and $\widehat{\mathbf{G}}$ have the same poles and approximately the same zeros).

For each value of ρ the asymptotic solution of the various robust algorithms for vanishing stepsize have been computed for 100 realizations of the uncertainty $\mathbf{\Delta G}$. Note, that Fig. 6 shows the resulting tradeoffs between control effort and the mean-square error (mse) $\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$ to which the FXLMS algorithm converges for vanishing stepsize ($\mu \rightarrow 0$). In this figure, both the control effort $\mathrm{E}[\mathbf{u}^T(k)\mathbf{u}(k)]$ and the mse $\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$ are averaged over the 100 realizations of $\mathbf{\Delta G}$, provided the update algorithm is stable (i.e., the eigenvalues of $\mathbf{T_{\hat{R}R}} + \mathbf{T_{FF}}$ are in the right half plane).

We observe that the results for uncertainty, stability and noncritical behavior weighting do not differ that much, whereas the output weighting yields a rather bad tradeoff and leakage/effort weighting has lower cost but is not able to give best performance. The behavior of output weighting can be explained by the fact that the 1,1 element of $\widehat{\mathbf{G}}$ consists of a low pass filter. To be able to stabilize the update rule for the model uncertainty in the high frequency band a large value for ρ needs to be selected, which significantly suppresses the performance at the lower frequency band. Similarly, for leakage/effort weighting the regularization, though less severe as in output weighting, is not according to the model uncertainty. Note, that for very small ρ the control effort becomes significant but due to the uncertainty in the system the performance is getting worse.

Table II shows the values of ρ, and the asymptotic values of $\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$ (in dB) and $\mathrm{E}[\mathbf{u}^T(k)\mathbf{u}(k)]$ (in dB) for vanishing stepsize averaged over the 100 uncertainty realizations of $\mathbf{\Delta G}$ for several situations. The first rows and columns give the values for which the update algorithm is just stable, i.e., the eigenvalues of $\mathbf{T_{\hat{R}R}} + \mathbf{T_{FF}}$ are in the right half plane. The values on the first rows and last columns are obtained for ρ being such that the SPR condition of Theorem 1 is just satisfied. Comparing both results, we infer that to assure stability of the update rule, ρ can be significantly smaller than is necessary to satisfy the SPR condition. This is because the number of adaptive filter coefficients is not too large, i.e., $n_w = 30$. When n_w is increased the values of ρ to achieve stability will approach the values of ρ that satisfy the SPR condition. Note, that for $\mathbf{F} = \sqrt{\rho}\hat{\mathbf{F}}_{\rm stab}$ the SPR condition is, indeed, satisfied for $\rho = 1$.

Also observe, that the minimal ρ to achieve stability, does not yield best performance. This is because the uncertainty in the secondary path $\mathbf{\Delta G}$ distorts the performance significantly. The influence of $\mathbf{\Delta G}$ on the performance is reduced by increasing the regularization parameter ρ, such that there exists a particular value ρ for which the performance, on average over all realizations $\mathbf{\Delta G}$, is optimal. These values are given on the last rows of Table II, that show the optimal performance averaged over all the 100 realizations of $\mathbf{\Delta G}$ obtained by the adaptive algorithms and obtained by the model-based offline computed filter $-\{\mathrm{E}[\widehat{\mathbf{R}}(k)\widehat{\mathbf{R}}^T(k)] + \mathbf{T_{FF}}\}^{-1}\mathrm{E}[\widehat{\mathbf{R}}(k)\mathbf{d}(k)]$. We observe, that, except for output weighting, the optimal performance obtained by the adaptive robust FXLMS algorithms is better than the performance achieved by the model-based offline computed filters. Observe, that in this experiment the robust FXLMS algorithm with $\mathbf{F} = \sqrt{\rho}\hat{\mathbf{F}}_{\rm stab}$ and the offline computed filter with $\mathbf{F} = \sqrt{\rho}\hat{\mathbf{F}}_{\rm noncrit}$ give optimal performance for $\rho = 1$. However, this need not to be the case in general. Note, that though the regularization by $\hat{\mathbf{F}}_{\rm noncrit}$ to achieve noncritical behavior is stronger than by $\hat{\mathbf{F}}_{\rm stab}$ to achieve asymptotic convergence, this difference does not influence the performance significantly. For the case of output weighting, the nonadaptive offline computed filter provides better performance than the adaptive algorithm, because ρ can be much smaller ($\rho = 0.200$) in the offline case than is necessary to guarantee stability in the adaptive case.

Finally, Fig. 7 shows the mse, $\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$, averaged over the 100 realizations of $\mathbf{\Delta G}$, obtained by the nominal robust FXLMS algorithms with effort, output and stability weighting. For the robust algorithms, ρ is selected such that optimal average performance is achieved. The stepsize is chosen as $\mu = 0.001$, which is relatively small, such that the misadjustment can be neglected. The results for uncertainty and noncritical behavior weighting are comparable with stability weighting. We observe, that indeed there exists realizations $\mathbf{\Delta G}$ for which the nominal FXLMS algorithm diverges. The converged values of $\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$ are in agreement with Table II (last rows, first columns).

TABLE II
VALUES OF ρ, $\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$ (IN dB) AND $\mathrm{E}[\mathbf{u}^T(k)\mathbf{u}(k)]$ (IN dB) AVERAGED OVER THE 100 UNCERTAINTY REALIZATIONS OF $\boldsymbol{\Delta}\mathbf{G}$ FOR WHICH THE ADAPTIVE ALGORITHMS ARE JUST STABLE (FIRST ROWS AND COLUMNS), FOR WHICH THE SPR STABILITY CONDITION OF THEOREM 1 IS JUST SATISFIED (FIRST ROWS, LAST COLUMNS), FOR WHICH THE LOWEST AVERAGED VALUE FOR $\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$ IS ACHIEVED BY THE ADAPTIVE ALGORITHMS (LAST ROWS, FIRST COLUMNS) AND BY THE OFFLINE COMPUTED FILTER (LAST ROWS AND COLUMNS)

Method:	Minimal ρ for stability			Minimal ρ for SPR condition		
	ρ	$\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$	$\mathrm{E}[\mathbf{u}^T(k)\mathbf{u}(k)]$	ρ	$\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$	$\mathrm{E}[\mathbf{u}^T(k)\mathbf{u}(k)]$
Leaky/Effort ($\mathbf{F}=\sqrt{\rho}\mathbf{I}$)	0.0079	5.25	12.1	0.158	−10.0	−0.0297
Output ($\mathbf{F}=\sqrt{\rho}\widehat{\mathbf{G}}$)	1.00	−2.64	2.28	15.8	−0.527	−20.7
Uncertainty ($\mathbf{F}=\sqrt{\rho}\widehat{\boldsymbol{\Delta}\mathbf{G}}$)	0.00025	5.11	12.1	15.8	−1.39	−9.05
Stability ($\mathbf{F}=\sqrt{\rho}\mathbf{F}_{\text{stab}}$)	0.100	4.04	11.4	1.00	−12.8	3.03
Noncritical ($\mathbf{F}=\sqrt{\rho}\widehat{\mathbf{F}}_{\text{noncrit}}$)	0.0316	5.13	12.1	0.794	−12.9	3.01
Method:	Opt. performance adaptive algorithms			Opt. performance offline least squares		
	ρ	$\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$	$\mathrm{E}[\mathbf{u}^T(k)\mathbf{u}(k)]$	ρ	$\mathrm{E}[\mathbf{e}^T(k)\mathbf{e}(k)]$	$\mathrm{E}[\mathbf{u}^T(k)\mathbf{u}(k)]$
Leaky/Effort ($\mathbf{F}=\sqrt{\rho}\mathbf{I}$)	0.100	−10.8	1.04	0.0794	−9.22	0.932
Output ($\mathbf{F}=\sqrt{\rho}\widehat{\mathbf{G}}$)	1.58	−3.30	−1.88	0.200	−7.66	2.22
Uncertainty ($\mathbf{F}=\sqrt{\rho}\widehat{\boldsymbol{\Delta}\mathbf{G}}$)	0.0063	−13.0	3.01	0.0100	−10.5	2.66
Stability ($\mathbf{F}=\sqrt{\rho}\widehat{\mathbf{F}}_{\text{stab}}$)	1.00	−12.8	3.03	1.58	−10.4	2.67
Noncritical ($\mathbf{F}=\sqrt{\rho}\widehat{\mathbf{F}}_{\text{noncrit}}$)	0.631	−12.9	3.11	1.00	−10.4	2.76

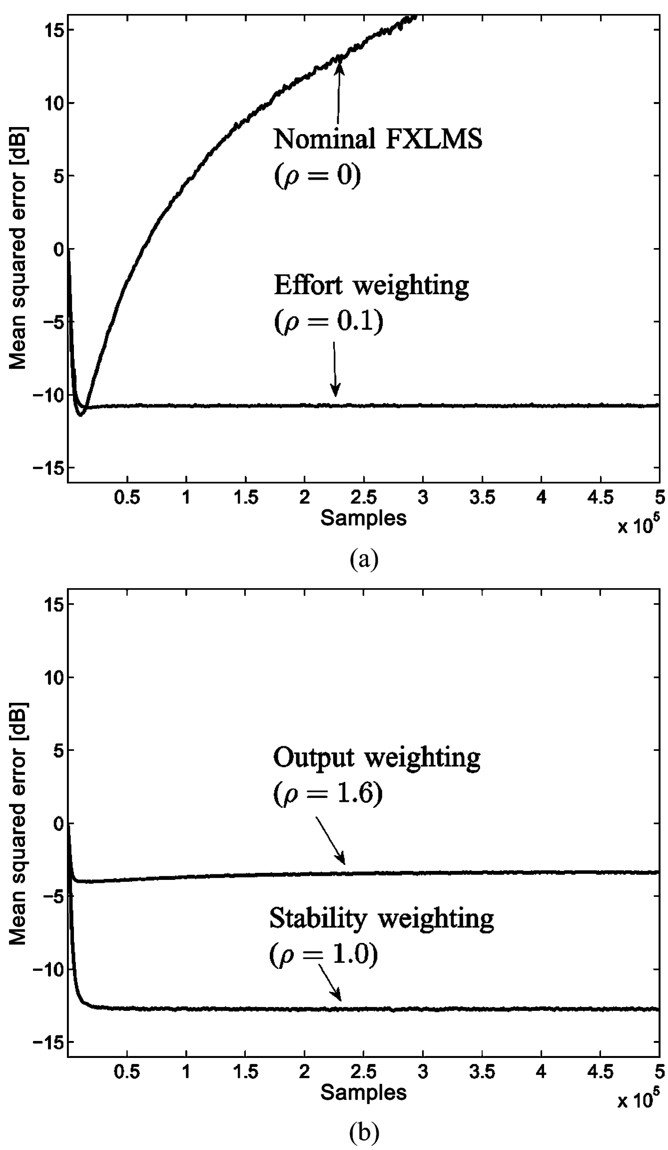

Fig. 7. Mean-square error averaged over 100 realizations of $\boldsymbol{\Delta}\mathbf{G}$ obtained by the nominal Filtered-X LMS algorithm and various robust variants, the stepsize is $\mu = 0.001$, and the vertical intervals indicate the upper and lower bounds of the mse for all 100 realizations of $\boldsymbol{\Delta}\mathbf{G}$. (a) Nominal filtered-X LMS and effort weighting; (b) output filtered-X LMS and stability weighting.

For this experiment uncertainty, stability and noncritical behavior weighting show similar results that are better than leakage/effort and output weighting. This does not mean that in practice one always need to choose for these weightings. A drawback is an increase of computational complexity because of additional filter actions and an increase of the number of columns in the regression vector. For example, for diagonal $\mathbf{F}$ with m_u equivalent FIR filters of n_f taps, the computational complexity is increased by approximately $2m_x(n_f + m_u n_w)$ floating point operations (flops). In comparison, the increase of computational complexity for online secondary path modeling is approximately $4m_u m_e n_s$ flops where n_s the number of taps of the FIR secondary path model.

The worth of the analysis of this paper together with the experimental results is an increase of insight in the behavior of (robust) FXLMS algorithms and stresses the need to evaluate in which frequency range the model uncertainty is dominant.

V. CONCLUSION

Model uncertainty not only may degrade the performance of the FXLMS algorithm, it may also yield an unstable update rule. Various regularization methods that may recover stability are discussed. When the uncertainty is dominant in a particular frequency region, regularization methods like leaky FXLMS, effort weighting and output weighting FXLMS, may yield too conservative results. Minimum norm regularization filters have been proposed that stabilize the update algorithm for a norm bounded uncertainty set and prevent critical behavior (i.e., the initial increase of the error before decaying to zero). The minimum norm regularization filters provide good performance, but the price to be paid is an increase of computations.

APPENDIX

A. Proof of Lemma 1

Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ be the SVD of $\mathbf{A}$. Assume $\lambda \in \mathbb{C}$ such that $|\lambda - s_i^2| \leq \gamma s_i$ for some singular value s_i. If $s_i = 0$ then $\lambda = 0$ which is an eigenvalue of $(\mathbf{A}+\boldsymbol{\Delta})^H\mathbf{A}$ for $\boldsymbol{\Delta} =$

0. Assume $s_i \neq 0$, then choose $\mathbf{\Delta} = \mathbf{U}\mathbf{\Phi}\mathbf{V}^H$ where $\mathbf{\Phi} = [((\lambda - s_i^2)/s_i)I_n \; 0_{m-n\times n}^T]^T$ for $m \geq n$ and $\mathbf{\Phi} = [((\lambda - s_i^2)/s_i)I_m \; 0_{m\times n-m}^T]^T$ for $n \geq m$. Because $\|\mathbf{\Delta}\|_2 = \|\mathbf{\Phi}\|_2 = (\lambda - s_i^2)/s_i$ and the assumption $|\lambda - s_i^2| \leq \gamma s_i$, it follows that $\|\mathbf{\Delta}\|_2 \leq \gamma$ such that this choice of $\mathbf{\Delta}$ is valid. Then, $(\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} = \mathbf{V}(\mathbf{\Sigma}^H\mathbf{\Sigma} + \mathbf{\Phi}^H\mathbf{\Sigma})\mathbf{V}^H$ is an eigenvalue decomposition, where the ith element on the diagonal of $\mathbf{\Sigma}^H\mathbf{\Sigma} + \mathbf{\Phi}^H\mathbf{\Sigma}$ equals λ such that $\lambda \in \overline{\sigma}_\gamma(\mathbf{A})$.

On the other hand, assume that there exists a $\mathbf{\Delta}$ for which $\|\mathbf{\Delta}\|_2 \leq \gamma$ such that λ is an eigenvalue of $(\mathbf{A}+\mathbf{\Delta})^H\mathbf{A}$. But assume that $\lambda \notin \{\lambda \in \mathbb{C} : |\lambda - s_i^2| \leq \gamma s_i, i = 1, \cdots, \min(m,n)\}$ such that $|\lambda - s_i^2| > \gamma s_i$ for $i = 1, \cdots, \min(m,n)$. Let $\mathbf{v}$ be an eigenvector of $(\mathbf{A} + \mathbf{\Delta})^H\mathbf{A}$ corresponding with λ, such that $((\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$ and $\|(\mathbf{A}^H\mathbf{A} - \lambda\mathbf{I})\mathbf{v}\|_2 = \|\mathbf{\Delta}^H\mathbf{A}\mathbf{v}\|_2$. Now, define $\mathbf{w} = \mathbf{V}^H\mathbf{v}$ such that $\mathbf{\Delta}^H\mathbf{A}\mathbf{v} = \mathbf{\Delta}^H\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{v} = \mathbf{\Delta}^H\mathbf{U}\mathbf{\Sigma}\mathbf{w}$. Hence, $\|\mathbf{\Delta}^H\mathbf{A}\mathbf{v}\|_2^2 = \|\mathbf{\Delta}^H\mathbf{U}\mathbf{\Sigma}\mathbf{w}\|_2^2 \leq \gamma^2\|\mathbf{\Sigma}\mathbf{w}\|_2^2 = \sum_{i=1}^{\min(m,n)}(\gamma s_i)^2\mathbf{w}_i < \sum_{i=1}^{\min(m,n)} |\lambda - s_i^2|^2\mathbf{w}_i^2 = \|(\mathbf{A}^H\mathbf{A} - \lambda\mathbf{I})\mathbf{v}\|_2^2$, where we made use of the assumption $|\lambda - s_i^2| > \gamma s_i$. This result contradicts with $\|(\mathbf{A}^H\mathbf{A} - \lambda\mathbf{I})\mathbf{v}\|_2 = \|\mathbf{\Delta}^H\mathbf{A}\mathbf{v}\|_2$ and thus λ should be contained in $\overline{\sigma}_\gamma(\mathbf{A})$. □

B. Proof of Lemma 2

By Lemma 1 $\overline{\sigma}_\gamma(\mathbf{A}) = \{\lambda \in \mathbb{C} : |\lambda - s_i^2| \leq \gamma s_i, i = 1, \cdots, \min(m,n)\}$, such that $\min \mathrm{Re}[\overline{\sigma}_\gamma(\mathbf{A})] = \min_i(s_i - \gamma)s_i = (s^o - \gamma)s^o$, where s^o is the singular value nearest to the minimum of the function $f(s) = (s - \gamma)s$ at $s = \gamma/2$. □

C. Proof of Lemma 3

Because $(\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} + \mathbf{A}^H(\mathbf{A} + \mathbf{\Delta})$ is Hermitian it holds that $\min \sigma((\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} + \mathbf{A}^H(\mathbf{A} + \mathbf{\Delta})) = \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^H((\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} + \mathbf{A}^H(\mathbf{A} + \mathbf{\Delta}))\mathbf{x}$ [10, Th. .2.2, p. 176]. Because of continuity the order of the minimization can be changed such that $\min_{\|\mathbf{\Delta}\|_2\leq\gamma} \min \sigma((\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} + \mathbf{A}^H(\mathbf{A} + \mathbf{\Delta})) = \min_{\|\mathbf{x}\|_2=1} \min_{\|\mathbf{\Delta}\|_2\leq\gamma} \mathbf{x}^H(\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} + \mathbf{A}^H(\mathbf{A} + \mathbf{\Delta}))\mathbf{x}$. Let $\mathbf{y} = \mathbf{A}\mathbf{x}$ and $\mathbf{z} = \mathbf{\Delta}\mathbf{x}$, then $\mathbf{x}^H(\mathbf{A}+\mathbf{\Delta})^H\mathbf{A}+\mathbf{A}^H(\mathbf{A}+\mathbf{\Delta}))\mathbf{x} = (\mathbf{y}+\mathbf{z})^H\mathbf{y}+\mathbf{y}^H(\mathbf{y}+\mathbf{z})$. For each $\mathbf{x}$, for which $\|\mathbf{x}\|_2 = 1$, and for each $\mathbf{z}$, for which $0 < \mathbf{z}^H\mathbf{z} \leq \gamma^2$, there exists a $\mathbf{\Delta}$ for which $\|\mathbf{\Delta}\|_2 \leq \gamma$ such that $\mathbf{z} = \mathbf{\Delta}\mathbf{x}$, e.g., choose $\mathbf{\Delta} = \mathbf{z}\mathbf{x}^H$. Hence $\mathbf{z}$ can point in every direction, only its magnitude is limited from above by γ. Hence, $\min_{\|\mathbf{\Delta}\|_2\leq\gamma} \mathbf{x}^H(\mathbf{A} + \mathbf{\Delta})^H\mathbf{A} + \mathbf{A}^H(\mathbf{A} + \mathbf{\Delta}))\mathbf{x} = \min_{\mathbf{z}^H\mathbf{z}\leq\gamma^2}(\mathbf{y} + \mathbf{z})^H\mathbf{y} + \mathbf{y}^H(\mathbf{y} + \mathbf{z})$. The minimizing value $\mathbf{z}$ is obtained by perpendicular projection of $\mathbf{z}$ onto $\mathbf{y}$ and given by $\mathbf{z} = -(\gamma/\|\mathbf{y}\|_2)\mathbf{y}$ for $\mathbf{y} \neq 0$, and $\min_{\mathbf{z}^H\mathbf{z}\leq\gamma^2}(\mathbf{y} + \mathbf{z})^H\mathbf{y} + \mathbf{y}^H(\mathbf{y} + \mathbf{z}) = 2(\|\mathbf{y}\|_2 - \gamma)\|\mathbf{y}\|_2$. In the trivial case $\mathbf{y} = 0$, $\mathbf{z}$ can point in any direction and $\min_{\mathbf{z}^H\mathbf{z}\leq\gamma^2}(\mathbf{y}+\mathbf{z})^H\mathbf{y}+\mathbf{y}^H(\mathbf{y}+\mathbf{z}) = 0$. Remains to minimize $2(\|\mathbf{y}\|_2 - \gamma)\|\mathbf{y}\|_2$ subject to $\mathbf{y} = \mathbf{A}\mathbf{x}$ over all $\mathbf{x}$ for which $\|\mathbf{x}\|_2 = 1$. The unconstrained minimum of $2(\|\mathbf{y}\|_2 - \gamma)\|\mathbf{y}\|_2$ is obtained at $\|\mathbf{y}\|_2 = \gamma/2$, which may or may not be achieved because $s_n \leq \|\mathbf{y}\|_2 \leq s_1$. Therefore, distinguish the three cases (as in the lemma):

- $0 < \gamma/2 \leq s_n$: then $2(\|\mathbf{y}\|_2 - \gamma)\|\mathbf{y}\|_2$ is minimized for $\|\mathbf{y}\|_2 = s_n$ such that $\lambda_{\min,\gamma}(\mathbf{A}) = 2(s_n - \gamma)s_n$, which proves the first case.
- $s_n \leq \gamma/2 \leq s_1$: then $2(\|\mathbf{y}\|_2 - \gamma)\|\mathbf{y}\|_2$ is minimized for $\|\mathbf{y}\|_2 = \gamma/2$ such that $\lambda_{\min,\gamma}(\mathbf{A}) = -\gamma^2/2$ which proves the second case.
- $s_1 \leq \gamma/2$: then $2(\|\mathbf{y}\|_2 - \gamma)\|\mathbf{y}\|_2$ is minimized for $\|\mathbf{y}\|_2 = s_1$ such that $\lambda_{\min,\gamma}(\mathbf{A}) = 2(s_1 - \gamma)s_1$ which proves the third case. □

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] R. Fraanje, M. Verhaegen, and S. J. Elliott, "Robustness of the filtered-X LMS algorithm—Part I: Necessary conditions for convergence and the asymptotic pseudospectrum of Toeplitz matrices," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4029–4037, Aug. 2007.

[2] S. J. Elliott, *Signal Processing for Active Control*. London, U.K.: Academic, 2001.

[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[4] R. Fraanje, M. Verhaegen, and N. J. Doelman, "Increasing the robustness of a preconditioned filtered-x LMS algorithm," *IEEE Signal Process. Lett.*, vol. SPL-11, no. 2, pp. 285–288, Feb. 2004.

[5] P. Darlington, "Performance surfaces of minimum effort estimators and controllers," *IEEE Trans. Signal Process.*, vol. SP-43, no. 2, pp. 536–539, Feb. 1994.

[6] M. H. Costa and J. C. M. Bermudez, "A new adaptive algorithm for reducing non-linear effects from saturation in active noise control systems," *Int. J. Adapt. Control Signal Process.*, vol. 19, no. 2–3, pp. 177–196, 2005.

[7] O. J. Tobias and R. Sera, "Leaky-FXLMS algorithm: Stochastic analysis for Gaussian data and secondary path modeling error," *IEEE Trans. Speech Audio Process.*, vol. SAP-13, no. 6, pp. 1217–1230, Nov. 2005.

[8] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge Univ. Press, 1991.

[9] K. Hinnen, M. Verhaegen, and N. Doelman, "Robust spectral factor approximation of discrete-time frequency domain power spectra," *Automatica*, vol. 41, no. 10, pp. 1791–1798, Oct. 2005.

[10] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.

Rufus Fraanje was born in The Netherlands in 1976. He received the M.Sc. degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 1999, and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 2004.

Since 2004, he has been a Postdoctoral Researcher with the Delft Center for Systems and Control from the Delft University of Technology. He was a visiting researcher with INRIA, Rennes, France, the Adaptive Systems Lab, University of California, Los Angeles, and the Institute of Sound and Vibration Research (ISVR), Southampton, U. K. His research interests are in adaptive filtering, system identification, and distributed control for application to active noise and vibration suppression, structural acoustics, and adaptive optics.

Stephen J. Elliott (M'83–SM'92) received the B.Sc. degree in physics and electronics from the University of London, U.K., in 1976, and the Ph.D. degree from the University of Surrey, U.K., in 1979 for a dissertation on musical acoustics.

After a short period as a Research Fellow at the ISVR working on acoustic intensity measurement and as a temporary Lecturer at the University of Surrey, he was appointed Lecturer at the Institute of Sound and Vibration Research (ISVR), University of Southampton, in 1982. He became a Senior Lecturer in 1988, Professor in 1994, and Director of the ISVR in 2005. His research interests have been mainly concerned with the connections between the physical world, signal processing and control, mainly in relation the active control of sound using adaptive filters and the active feedback control of vibration. This work has resulted in the practical demonstration of active control in propeller aircraft, cars, and helicopters. His current research interests include active isolation, modular systems for active feedback control, and modeling the active processes within the cochlear. He has published more than 150 papers in refereed journals and 350 conference papers. He is a coauthor of *Active Control of Sound* (with P.A. Nelson, London, U.K.: Harcourt Brace, 1992), *Active Control of Vibration* (with C. R. Fuller and P. A. Nelson, London, U.K.: Academic, 1996) and the author of *Signal Processing for Active Control* (San Diego, CA, 2001).

Prof. Elliott is a member of the Acoustical Society of America and a Fellow of the Institute of Electrical Engineering and the IoA. He was jointly awarded the Tyndall Medal from the Institute of Acoustics in 1992 and the Kenneth Harris James Prize from the Institution of Mechanical Engineers in 2000.

Michel Verhaegen received an engineering degree in aeronautics from the Delft University of Technology, The Netherlands, in August 1982, and the Ph.D. degree in applied sciences from the Catholic University Leuven, Belgium, in November 1985. During his graduate study, he held a research assistantship sponsored by the Flemish Institute for Scientific Research (IWT).

From 1985 to 1994, he has been a Research Fellow of the U.S. National Research Council (NRC), affiliated with the NASA Ames Research Center in California, and of the Dutch Academy of Arts and Sciences, affiliated with the Network Theory Group of the Delft University of Technology. During 1994–1999, he was an Associate Professor of the Control Laboratory of the Delft University of Technology and became appointed Full Professor with the faculty of Applied Physics of the University of Twente, The Netherlands, in 1999. In 2001, he returned to the University of Delft and is now directing, together with Prof. van den Hof, the new Delft Center for Systems and Control. This center has brought together three research groups in Delft that focus on practically motivated fundamental research in the area of systems and control. The center, established in 2004, currently has five full-time professors, nine permanent staff members, 14 support staff members, approximately 40 Ph.D. degree students, and 40 M.Sc. degree students. His main research interest is the interdisciplinary domain of numerical algebra and system theory. In this field, he has published more than 100 papers. Current activities focus on the transfer of knowledge about new identification and controller design methodologies to industry. Application areas include control of large scale wind turbines, adaptive optics, and X-by-wire global chassis control.