Received: 1 September 2023 | Revised: 18 March 2024 | Accepted: 19 April 2024

DOI: 10.1002/acs.3826

RESEARCH ARTICLE

WILEY

Frequency domain stability and relaxed convergence conditions for filtered error adaptive feedforward

Sil T. Spanjer | **Hakan Köroğlu** | **Wouter B. J. Hakvoort**

Precision Engineering, University of Twente, Enschede, The Netherlands

Correspondence
Sil T. Spanjer, Precision Engineering, University of Twente, Enschede, The Netherlands.
Email: s.t.spanjer@utwente.nl

Funding information
University of Twente; MI-partners B.V.; PM B.V.

Summary

The convergence of filtered error and filtered reference adaptive feedforward is limited by three effects: model mismatch, unintended input-disturbance interaction and too fast parameter adaptation. In this article, the first two effects are considered for MIMO systems under the slow parameter adaptation assumption. The convergence with model mismatch is conventionally guaranteed using a strictly positive-real condition. This condition can be easily verified in the frequency domain, but due the high-frequency parasitic dynamics of real systems, it is hardly ever satisfied. Nevertheless, filtered error and filtered reference adaptive feedforward have successfully been implemented in numerous applications without satisfying the strictly positive-real condition. It is shown in this article that the strictly positive-real condition can be relaxed to a power-weighted integral condition, that is less conservative and provides a practical check for the convergence of filtered error adaptive feedforward for real systems in the frequency domain. The effects of input-disturbance interaction are analysed and conditions for the stability are given in the frequency domain. Both conditions give clear indicators for frequency domain filter tuning, and are verified on an experimental active vibration isolation system.

KEYWORDS

adaptive feedforward, convergence, filtered error, frequency domain, stability

1 | INTRODUCTION

Adaptive feedforward (AF) is a method to improve the performance of a feedforward controller under parametric variations and uncertainty. This method is successfully implemented for a multitude of applications, such as a wafer stage,[1] mechanical ventilation,[2] linear motors[3] and wind turbines.[4] A field where adaptive feedforward is especially prevalent is active noise cancellation (ANC) (see Lu et al.[5] for a recent survey) and good results have been obtained for active vibration isolation systems[6] (AVIS).

For systems that use AF, both stability and convergence are important. Stability refers to all internal states of the system being bounded for nonzero initial conditions. Convergence refers to the parameters of the AF adapting to a bounded set. The convergence is consistent if this set is the ideal parameter vector.

The stability of systems that use feedforward is typically not influenced by the feedforward controller if the feedforward controller itself is stable, since no feedback loops are closed by the feedforward controller.[7] For some applications, such as ANC and AVIS, the feedforward is based on measurements of the disturbance, and can be corrupted by the feedforward

action, effectively closing a feedback loop over the feedforward.[5,8] This makes the stability of the system dependent on the feedforward.

For systems with dynamics in the compensation path, such as ANC and AVIS, either filtered reference[9] (Fx) or filtered error[10] (Fe) AF are used. Compared to FxAF, FeAF is computationally more efficient in a MIMO setting. These methods require a model of the system, which is pivotal for the adaptive feedforward parameter convergence. A sufficient condition on the model for convergence is the strictly positive real (SPR) condition by Wang and Ren.[11]

This convergence condition has several downsides. It is derived using a specific choice of the Lyapunov function, which introduces conservatism in the convergence condition. Nonetheless, Fraanje et al.[12] concluded that the specific choice of the Lyapunov function is necessary to prevent critical behavior, which is non-monotonic decrease in error. This makes the specific choice of Lyapunov function actually beneficial for practical implementations due to the actuator and system limits. Furthermore, the condition only holds for slow parameter adaptation. Last is the requirement that the SPR condition should hold at every frequency. This makes the SPR condition impractical for real systems due to their high state dimensional nature and the corresponding significant unmodelled dynamics.

The practical limitation of the SPR condition was overcome by Beijen et al.[6] by using a noise-shaping filter to remove the power content in the frequency regions with modeling error. This allows for a stable implementation of the FeLMS algorithm, however, the convergence condition is still based on the SPR condition and is hence not satisfied. This lack of theoretical underpinning results in a tuning process of the noise-shaping filter without clear objectives, other than the broad notion of convergence, and relies on the experience of the designer. This yields unpredictable behavior. In this article, the SPR condition is relaxed to a power-weighted integral condition. This new condition can be evaluated based on FRF measurements of the dynamics, the disturbance spectrum and an eigenvalue analysis. This has significant value since it provides a less conservative convergence condition, a basis for tuning the noise-shaping filter and a method for evaluating the performance of the system model. Furthermore, it is data-based and does not need an exact model of the plant, other than the one used for adaptation. The analysis for fast parameter adaptation is outside the scope of this article, but is a topic for further research. Furthermore, the effects of the input-disturbance interaction on the stability of the system are analysed. A different problem setup is used here compared to Landau et al.,[13] where the input only influences the measurement of the disturbance. Instead, here the actual disturbance is influenced by the input. The theory will be developed for a FeAF controller for AVIS, but extends straightforwardly to other adaptive feedforward methods such as FxAF and applications such as ANC and reference based AF. Points of attention for other applications and methods will be highlighted in footnotes.

The article is organized as follows. First, the general control structure is introduced. The stability condition with input-disturbance interaction is derived next in Section 3. In Section 4, the filtered error adaptive feedforward method is introduced. The convergence is first analysed in time domain in Section 5 and the resulting condition is converted to the SPR condition in the frequency domain. The SPR condition is relaxed in Section 6. In Section 7, a method for the design of the noise shaping filter based on the relaxed convergence condition is given. The relaxed convergence condition is experimentally validated in Section 8. Sections 9 and 10 are the discussion and conclusion respectively.

1.1 | Notation

In this article, $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ denote respectively the set of real numbers, the set of $n\times 1$ real column vectors and the set of $n\times m$ real matrices; $\mathbb{C}$, $\mathbb{C}^n$ and $\mathbb{C}^{n\times m}$ denote complex numbers, column vectors and matrices respectively. We use $(\bullet)^T$ for the transpose, $(\bullet)^H$ for the Hermitian transpose, $(\bullet)^{-1}$ for the inverse, $\mathrm{Tr}(\bullet)$ for the trace, $\det(\bullet)$ for the determinant, $\inf(\bullet)$ for the infimum, $\sup(\bullet)$ for the supremum, $\mathrm{conv}(\bullet)$ for the convex hull,[14] $Re(\bullet)$ for the real part, $\overline{\bullet}$ for a set frozen in time, and $\mathrm{col}(\bullet)$ for the vectorization of a matrix to a column vector. The set $\mathbb{R}H_\infty^{n\times m}$ contains all $n\times m$ causal stable discrete-time systems. We say that $\boldsymbol{x}\in\mathcal{L}_\bullet$ for $\boldsymbol{x}\in\mathbb{R}^n$ if $\|\boldsymbol{x}\|_\bullet<\infty$, where $\|\boldsymbol{x}\|_\infty=\sup_{k\geq 0}\|\boldsymbol{x}(k)\|_\infty=\sup_{k\geq 0}\max_{1\leq i\leq n}|x_i(k)|$ and $\|\boldsymbol{x}\|_{\mathrm{pow}}=\lim_{N\to\infty}\sqrt{\frac{1}{2N+1}\sum_{i=-N}^{N}\boldsymbol{x}^T(i)\boldsymbol{x}(i)}$ are used. The relevant matrix norms are $\|\boldsymbol{X}\|_{\mathrm{pow}}=\lim_{N\to\infty}\sqrt{\frac{1}{2N+1}\sum_{i=-N}^{N}Tr\left(\boldsymbol{X}^T(i)\boldsymbol{X}(i)\right)}$ and $\|\boldsymbol{X}\|_\infty=\sup_{k\geq 0}\|\boldsymbol{X}(k)\|_\infty=\sup_{k\geq 0}\max_{1\leq i\leq n}\max_{1\leq j\leq m}|x_{ij}(k)|$.[15]

2 | PROBLEM DESCRIPTION

Figure 1 shows the MIMO FeAF control structure, and is based on Beijen et al.[6] The FeAF is described in discrete time. The forward shift operator is defined as q, and $t_k=kt_s$ with the index $k\in\mathbb{N}$ and sampling time t_s. The system consists

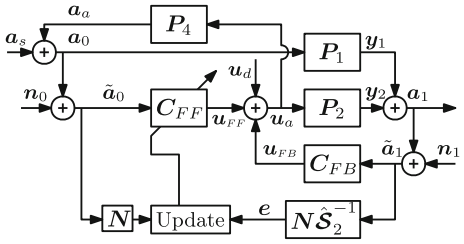

FIGURE 1 MIMO filtered error adaptive feedforward control structure.

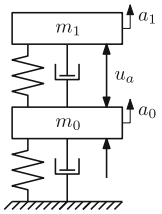

FIGURE 2 Ideal physical model of an AVIS system.

of the primary path $\boldsymbol{P}_1(q) \in \mathbb{R}H_\infty^{n_y \times n_a}$, the secondary path $\boldsymbol{P}_2(q) \in \mathbb{R}H_\infty^{n_y \times n_u}$ and the quaternary path* $\boldsymbol{P}_4(q) \in \mathbb{R}H_\infty^{n_a \times n_u}$. For the AVIS case, the primary path relates the floor acceleration $\boldsymbol{a}_0(k)$ to the payload acceleration $\boldsymbol{a}_1(k)$. The secondary path relates the actuator input $\boldsymbol{u}_a(k)$ to $\boldsymbol{a}_1(k)$ and the quaternary path relates the $\boldsymbol{u}_a(k)$ to $\boldsymbol{a}_0(k)$. In Figure 2 an ideal physical model of the AVIS setup is shown to illustrate these transfer paths. We emphasize that Beijen et al.[6] did not consider the presence of $\boldsymbol{P}_4(q)$. The tertiary path $\boldsymbol{P}_3(q) \in \mathbb{R}H_\infty^{n_a \times n_a}$, that relates $\boldsymbol{a}_0(k)$ to a measurement of $\boldsymbol{a}_0(k)$ is considered unity.[16] The signal $\boldsymbol{a}_s(k) \in \mathbb{R}^{n_a}$ is the exogenous disturbance, whose measurement $\tilde{\boldsymbol{a}}_0(k)$ is corrupted by sensor noise $\boldsymbol{n}_0(k)$ and the input-disturbance interaction signal $\boldsymbol{a}_a(k)$. The measurement $\tilde{\boldsymbol{a}}_0(k)$ is used for the feedforward controller $\boldsymbol{C}_{\text{FF}}(q) \in \mathbb{R}H_\infty^{n_u \times n_a}$. The feedback sensor measurement $\tilde{\boldsymbol{a}}_1(k)$ is corrupted by the sensor noise $\boldsymbol{n}_1(k)$ and is the input for the feedback controller $\boldsymbol{C}_{\text{FB}}(q) \in \mathbb{R}H_\infty^{n_u \times n_y}$. The disturbance input $\boldsymbol{u}_d(k)$ is unmeasured and corrupts the input of $\boldsymbol{P}_2(q)$. The signals $\boldsymbol{a}_s,\ \boldsymbol{n}_0,\ \boldsymbol{u}_d,\ \boldsymbol{n}_1 \in \mathcal{L}_{\text{pow}}$ are assumed wide-sense stationary, uncorrelated and zero-mean.

The output of the system is given by

$$\boldsymbol{a}_1(k) = \boldsymbol{P}_1(q)\boldsymbol{a}_0(k) + \boldsymbol{P}_2(q)\boldsymbol{u}_a(k). \tag{1}$$

The objective of the control structure of Figure 1 is to minimize the variance of the output $\boldsymbol{a}_1(k)$. To this end, both feedforward and the feedback controllers are used. The feedforward is used to attenuate the influence of $\boldsymbol{a}_s(k)$, while the feedback controller is to attenuate all remaining disturbances. The input signal $\boldsymbol{u}_a(k) \in \mathbb{R}^{n_u}$ is given by

$$\begin{aligned}\boldsymbol{u}_a(k) &= \boldsymbol{u}_{\text{FF}}(k) + \boldsymbol{u}_{\text{FB}}(k) + \boldsymbol{u}_d(k),\\ &= \boldsymbol{C}_{\text{FF}}(q)\tilde{\boldsymbol{a}}_0(k) + \boldsymbol{C}_{\text{FB}}(q)\tilde{\boldsymbol{a}}_1(k) + \boldsymbol{u}_d(k),\\ &= \underbrace{\left[\boldsymbol{I} - \boldsymbol{C}_{\text{FF}}(q)\boldsymbol{P}_4(q) - \boldsymbol{C}_{\text{FB}}(q)(\boldsymbol{P}_1(q)\boldsymbol{P}_4(q) + \boldsymbol{P}_2(q))\right]^{-1}}_{\boldsymbol{Q}(q)}\boldsymbol{r}(k),\end{aligned} \tag{2}$$

with

$$\boldsymbol{r}(k) = (\boldsymbol{C}_{\text{FF}}(q) + \boldsymbol{C}_{\text{FB}}(q)\boldsymbol{P}_1(q))\boldsymbol{a}_s(k) + \boldsymbol{C}_{\text{FF}}(q)\boldsymbol{n}_0(k) + \boldsymbol{C}_{\text{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{u}_d(k), \tag{3}$$

$\tilde{\boldsymbol{a}}_0(k) = \boldsymbol{a}_a(k) + \tilde{\boldsymbol{a}}_s(k) = \boldsymbol{a}_s(k) + \boldsymbol{a}_a(k) + \boldsymbol{n}_0(k)$ and $\tilde{\boldsymbol{a}}_1(k) = \boldsymbol{a}_1(k) + \boldsymbol{n}_1(k)$. Note that the occurrence of $\boldsymbol{C}_{\text{FF}}(q)\boldsymbol{P}_4(q)$ in $\boldsymbol{Q}(q)$ implies that it is not quite correct to use the terminology feedforward controller for $\boldsymbol{C}_{\text{FF}}(q)$ since a feedback loop is closed over $\boldsymbol{C}_{\text{FF}}(q)$ and $\boldsymbol{P}_4(q)$. This is nonetheless used in the remainder of the article since it provides a clear distinction between the two controllers. The signal $\boldsymbol{u}_d(k)$ is an unmeasured disturbance. Let

$$\boldsymbol{P}(q) = \boldsymbol{P}_1(q)\boldsymbol{P}_4(q) + \boldsymbol{P}_2(q). \tag{4}$$

Equation (1) can be rewritten as

$$\begin{aligned}\boldsymbol{a}_1(k) &= \boldsymbol{P}_1(q)\boldsymbol{a}_s(k) + \boldsymbol{P}(q)\boldsymbol{u}_a(k),\\ &= \boldsymbol{P}_1(q)\boldsymbol{a}_s(k) + \boldsymbol{P}(q)\boldsymbol{Q}^{-1}(q)\boldsymbol{r}(k),\\ &= \boldsymbol{S}_1(q)\boldsymbol{a}_s(k) + \boldsymbol{S}_2(q)\big[\boldsymbol{C}_{\mathrm{FF}}(q)\boldsymbol{n}_0(k) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{u}_d(k)\big],\end{aligned} \tag{5}$$

with

$$\boldsymbol{S}_1(q) = \boldsymbol{P}_1(q) + \boldsymbol{P}(q)\boldsymbol{Q}^{-1}(q)(\boldsymbol{C}_{\mathrm{FF}}(q) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{P}_1(q)), \tag{6}$$

$$\boldsymbol{S}_2(q) = \boldsymbol{P}(q)\boldsymbol{Q}^{-1}(q). \tag{7}$$

For $\boldsymbol{P}_4(q) = \boldsymbol{0}$, these functions reduce, respectively, to

$$\boldsymbol{\mathcal{T}}(q) = \big[\boldsymbol{I} - \boldsymbol{P}_2(q)\boldsymbol{C}_{\mathrm{FB}}(q)\big]^{-1}\big[\boldsymbol{P}_1(q) + \boldsymbol{P}_2(q)\boldsymbol{C}_{\mathrm{FF}}(q)\big], \tag{8}$$

$$\boldsymbol{C}(q) = \underbrace{\big[\boldsymbol{I} - \boldsymbol{P}_2(q)\boldsymbol{C}_{\mathrm{FB}}(q)\big]^{-1}}_{\boldsymbol{S}_{\mathrm{FB}}}\boldsymbol{P}_2(q), \tag{9}$$

and make

$$\boldsymbol{a}_1(k) = \boldsymbol{\mathcal{T}}(q)\boldsymbol{a}_s(k) + \boldsymbol{C}(q)\big[\boldsymbol{C}_{\mathrm{FF}}(q)\boldsymbol{n}_0(k) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{u}_d(k)\big], \tag{10}$$

which are the well known transmissibility and the (double time derivative of the) compliance functions.[17]

The FeAF controller relies on a feedforward controller that is linear in the parameters, as will be discussed in Section 4. The feedforward controller that minimizes the variance of $\boldsymbol{a}_1(k)$ can be expressed as an infinite series expansion in a suitable pre-chosen basis[18]

$$\boldsymbol{C}^{\circ}_{\mathrm{FF}}(q) = \sum_{i=0}^{\infty}\boldsymbol{W}_i\boldsymbol{F}_i(q), \tag{11}$$

where $\boldsymbol{C}^{\circ}_{\mathrm{FF}}(q) \in \mathbb{R}H_{\infty}^{n_u\times n_a}$, $\boldsymbol{W}_i \in \mathbb{R}^{n_u\times n_a}$ and

$$\boldsymbol{F}_i(q) = \begin{bmatrix} f_i(q) & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & f_i(q)\end{bmatrix} \in \mathbb{R}H_{\infty}^{n_a\times n_a}. \tag{12}$$

This corresponds to the scalar interpretation of (11). The alternative is the multivariable approach, that might be used to reduce the number of required parameters.[19,20] However, the used scalar interpretation offers a more flexible framework.

The expansion in (11) is truncated for practical reasons to order n_e as

$$\boldsymbol{C}^{\circ}_{\mathrm{FF}}(q) \approx \boldsymbol{C}_{\mathrm{FF}}(q) = \sum_{i=0}^{n_e}\boldsymbol{W}_i\boldsymbol{F}_i(q). \tag{13}$$

The choice of basis functions depends on the dynamics of the system. The most well known choice of basis functions are finite impulse response (FIR) filters

$$f_i(q) = q^{-i}. \tag{14}$$

FIR filters are commonly used in ANC systems due to their simplicity. For systems with long impulse response times, FIR filters are however inefficient because a large n_e is required for an accurate system description. The system description can be made more efficient by using other bases that allow incorporating knowledge of the system dynamics in the basis. A suitable choice for systems with possibly multiple complex poles are Takenaka-Malmquist functions.[18] Both the FIR filter basis and the Takenaka-Malmquist function basis are orthonormal. This gives favorable numerical properties,

but for the content of this article only linear independence is required. A suitable basis for AVIS are weak integrators,[6] possibly in combination with Takenaka-Malmquist functions.[21] The matrices $\boldsymbol{W}_i$ are uniquely defined if the basis is linearly-independent.

For later use, the output of the feedforward controller for a general input $\boldsymbol{a}(k)$ can be expressed as

$$\begin{aligned}\boldsymbol{u}_{\mathrm{FF}}(k) &= \boldsymbol{C}_{\mathrm{FF}}(q)\boldsymbol{a}(k) = \sum_{i=0}^{n_e} \boldsymbol{W}_i \boldsymbol{F}_i(q)\boldsymbol{a}(k) = \boldsymbol{W}\boldsymbol{F}(q)\boldsymbol{a}(k) \\ &= \underbrace{\begin{bmatrix} \tilde{\boldsymbol{\psi}}^T(k) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \tilde{\boldsymbol{\psi}}^T(k) \end{bmatrix}}_{\tilde{\boldsymbol{\Psi}}(k)} \boldsymbol{w} = \boldsymbol{w}_r \tilde{\boldsymbol{\psi}}(k) = \underbrace{\begin{bmatrix} \boldsymbol{f}_t^T(q) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{f}_t^T(q) \end{bmatrix}}_{\boldsymbol{\mathcal{F}}(q)} \boldsymbol{w}_d \begin{bmatrix} \boldsymbol{a}(k) \\ \vdots \\ \boldsymbol{a}(k) \end{bmatrix},\end{aligned} \tag{15}$$

with

$$\tilde{\boldsymbol{\psi}}(k) = \underbrace{\begin{bmatrix} \boldsymbol{F}_1(q) \\ \vdots \\ \boldsymbol{F}_{n_e}(q) \end{bmatrix}}_{\boldsymbol{F}(q)} \boldsymbol{a}(k), \tag{16}$$

and $\boldsymbol{a}(k) \in \mathbb{R}^{n_a}$, $\tilde{\boldsymbol{\Psi}}(k) \in \mathbb{R}^{n_u \times n_a n_e n_u}$. The parameter vector $\boldsymbol{w} = col(\boldsymbol{W}) \in \mathbb{R}^{n_a n_u n_e \times 1}$, and $\boldsymbol{w}_d \in \mathbb{R}^{n_a n_u n_e \times n_a n_u n_e}$ is the zero matrix with $\boldsymbol{w}$ on the diagonal. The matrix $\boldsymbol{w}_r \in \mathbb{R}^{n_u \times n_a n_e}$ contains the elements of $\boldsymbol{W}$. One order of the diagonal generating transfer functions is $\boldsymbol{F}_i(q) \in \mathbb{R}H_\infty^{n_a \times n_a}$, and the total set of block diagonal generating transfer functions $\boldsymbol{\mathcal{F}}(q) \in \mathbb{R}H_\infty^{n_u \times n_a n_e n_u}$ that contain repetitions of the elements of $\boldsymbol{F} \in \mathbb{R}H_\infty^{n_a n_e \times n_a}$. The nonzero entries of $\boldsymbol{F}(q)$ can be vectorized as $\boldsymbol{f}_t = col(\boldsymbol{F}(q)) \in \mathbb{R}^{n_a n_e}$. Equations (15) and (16) are various forms of the tensor Equation (13). Equation (15) implies the same expansion for all $\boldsymbol{u}_{\mathrm{FF}}(k)$. From an implementation point of view this is often not desired.[20] Using a different expansion for the different $u_{\mathrm{FF}}(k)$ translates to variation between the different blocks on the diagonal of $\boldsymbol{\mathcal{F}}(q)$. The same notation can be kept by taking the unique functions in $\boldsymbol{\mathcal{F}}(q)$, substituting these for $\boldsymbol{f}_t(q)$ and setting the correspondingly added parameters in $\boldsymbol{w}$ to zero.

The basis of the feedforward controller is to be designed such that it minimizes $\boldsymbol{a}_1(k)$, and will be detailed Section 4 for the AVIS case. It is made adaptive to make the performance robust against parameter variations. Making the feedforward adaptive creates a nonlinear feedback loop in the system. The convergence of this nonlinear feedback loop is conventionally guaranteed by the SPR lemma, and is discussed in Section 5. This lemma is too restrictive for practical applications, since it requires an accurate model up to the Nyquist frequency. Therefore, this condition is relaxed in Section 6. Synthesis guidelines are derived based on the relaxed convergence condition and are presented in Section 7. The inclusion of $\boldsymbol{P}_4(q)$ in the system creates additional feedback loops in the system. The stability of this interconnection is discussed in Section 3.

The stability and convergence conditions are given in the frequency domain, and require the following measurements of the system

1. Power spectral density (PSD) of $\boldsymbol{a}_s(k) + \boldsymbol{n}_0(k)$.
2. Frequency response function (FRF) of $\boldsymbol{P}(q)^\dagger$ and $\boldsymbol{P}_4(q)$.

The stability and convergence conditions can be used to design $\boldsymbol{C}_{\mathrm{FF}}(q)$, $\boldsymbol{C}_{\mathrm{FB}}(q)$, $\boldsymbol{N}(q)$, $\hat{\boldsymbol{S}}_2^{-1}(q)$. The latter two filters are used for the adaptation process, and will be introduced in Section 4.

3 | STABILITY WITH INPUT-DISTURBANCE INTERACTION

The additional feedback loops that are closed by $\boldsymbol{P}_4(q) \neq \boldsymbol{0}$ can cause instability over $\boldsymbol{P}_4(q)$ and the feedforward controller, where the feedforward controller is described by (15). The stability for this feedback loop is considered in this section, and

the range of $\boldsymbol{w}$ for which the closed loop is stable is quantified. This stability condition is derived using the generalized Nyquist criterion.

Theorem 1 (Skogestad and Postlethwaite[22]). *Let n_{ol} be the number of unstable open loop poles in $\boldsymbol{L}(q)$. The closed-loop $\boldsymbol{S}(q) = (\boldsymbol{I} + \boldsymbol{L}(q))^{-1}$ is stable if and only if the image of* $\det(\boldsymbol{I} + \boldsymbol{L}(q))$

1. *makes n_{ol} anti-clockwise encirclements of the origin, and*
2. *does not pass through the origin*

for all q in the discrete Nyquist contour[23].

Lemma 1. *Let the conditions of Theorem 1 be satisfied for*

$$\boldsymbol{L}(q) = -\boldsymbol{C}_{FB}(q)\boldsymbol{P}(q). \tag{17}$$

Let $\mathcal{W}_g$ represent the set of finitely many vertices of a polytopic region $\mathcal{W}_l = conv(\mathcal{W}_g)$[14] *and $\overline{\boldsymbol{w}}$ a vector of parameters for $\boldsymbol{C}_{FF}(q, \overline{\boldsymbol{w}})$. Then $\boldsymbol{S}_1(q, \overline{\boldsymbol{w}})$ and $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}})$ will be stable for all $\overline{\boldsymbol{w}} \in \mathcal{W}_l$, if the following condition is satisfied*:

$$\sigma_{max}\left(\boldsymbol{C}_{FF}\left(e^{j\omega}, \overline{\boldsymbol{w}}\right)\right)\sigma_{max}\left(\boldsymbol{P}_4\left(e^{j\omega}\right)\right) < \sigma_{min}\left(\boldsymbol{I} - \boldsymbol{C}_{FB}\left(e^{j\omega}\right)\left(\boldsymbol{P}_1\left(e^{j\omega}\right)\boldsymbol{P}_4\left(e^{j\omega}\right) + \boldsymbol{P}_2\left(e^{j\omega}\right)\right)\right) \quad \forall \overline{\boldsymbol{w}} \in \mathcal{W}_g. \tag{18}$$

Proof. From (5), it can be concluded that $\boldsymbol{S}_1(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_a}$ and $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_u}$ if $\boldsymbol{Q}^{-1}(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_u}$. First $\boldsymbol{C}_{\mathrm{FB}}$ is assumed to satisfy the generalized Nyquist criterion of Theorem 1 with $\boldsymbol{C}_{\mathrm{FF}}(q, \overline{\boldsymbol{w}})\boldsymbol{P}_4(q) = \boldsymbol{0}$. The second term of $\boldsymbol{Q}^{-1}(q, \overline{\boldsymbol{w}})$ is considered a perturbation. The total stability of the system is determined by $\det(\boldsymbol{I} + \boldsymbol{L}(q) - \boldsymbol{C}_{\mathrm{FF}}(q, \overline{\boldsymbol{w}})\boldsymbol{P}_4(q))$. Following the argument of Doyle,[24] the conditions of Theorem 1 can be violated if

$$\det(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4) = \boldsymbol{0}, \tag{19}$$

for some $0 \leq \eta \leq 1$ and $-\pi \leq \omega \leq \pi$. The argument $e^{j\omega}$ is left out for notational simplicity. The following identities and inequalities can be found in Skogestad and Postlethwaite.[22] Since $\det(\boldsymbol{A}) = \prod_i(\lambda_i(\boldsymbol{A}))$, if $\det(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4) = \boldsymbol{0}$ for $0 \leq \eta \leq 1$ and $-\pi \leq \omega \leq \pi$, then $\lambda_{\min}(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4) = 0$, with $\lambda_{\min}(.)$ the eigenvalue with the smallest absolute value. Furthermore, if $\lambda_{\min}\left(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) = 0$, then $\sigma_{\min}\left(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) = 0$ since $0 \leq \sigma_{\min} \leq |\lambda_i| \leq \sigma_{\max}$. This can be used to infer

$$\sigma_{\min}\left(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) \geq \sigma_{\min}(\boldsymbol{I} + \boldsymbol{L}) - \eta\sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) \geq \sigma_{\min}(\boldsymbol{I} + \boldsymbol{L}) - \sigma_{\max}(\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4). \tag{20}$$

Thus, to ensure $\lambda_{\min}\left(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) > 0$, it would be sufficient to have $\sigma_{\min}\left(\boldsymbol{I} + \boldsymbol{L} - \eta\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) > 0$, which would be implied by $\sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) < \sigma_{\min}(\boldsymbol{I} + \boldsymbol{L})$. Since

$$\sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\boldsymbol{P}_4\right) \leq \sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}(\overline{\boldsymbol{w}})\right)\sigma_{\max}(\boldsymbol{P}_4), \tag{21}$$

a sufficient condition for the stability of $\boldsymbol{Q}^{-1}$ is (18). Therefore, (18) implies $\boldsymbol{S}_1(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_a}$, $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_u} \forall\ \overline{\boldsymbol{w}} \in \mathcal{W}_g$. Equation (18) can be written as a linear matrix inequality (LMI) since $\boldsymbol{C}_{\mathrm{FF}}\left(e^{j\omega}, \overline{\boldsymbol{w}}\right)$ depends linearly on $\overline{\boldsymbol{w}}$, according to (15). The set $\mathcal{W}_l$ is therefore described by the convex hull $\mathcal{W}_l = conv\left(\mathcal{W}_g\right)$,[14] and $\boldsymbol{S}_1(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_a}$, $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_u} \forall\ \overline{\boldsymbol{w}} \in \mathcal{W}_g$ implies that $\boldsymbol{S}_1(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_a}$, $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}})$ are stable for all $\overline{\boldsymbol{w}} \in \mathcal{W}_l$. ∎

Remark 1. The convex dependence of $\sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}\left(e^{j\omega}, \overline{\boldsymbol{w}}\right)\right)$ on $\overline{\boldsymbol{w}}$ implies that it is sufficient to check (18) on the generators $\mathcal{W}_g$ of $\mathcal{W}_l$.[14] The generators $\mathcal{W}_g$ can be estimated in the frequency domain. This is shown in Section 8.

For sufficiently slowly varying $\boldsymbol{w}(k)$, $\boldsymbol{S}_1(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_a}$, $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_u} \forall\ \overline{\boldsymbol{w}} \in \mathcal{W}_g$ implies that $\boldsymbol{S}_1(q, \boldsymbol{w}(k))$ and $\boldsymbol{S}_2(q, \boldsymbol{w}(k))$ are stable for all $\boldsymbol{w}(k) \in \mathcal{W}_l$.[25] Therefore, this condition can also be used for the adaptive feedforward presented in the next section.

4 | FILTERED ERROR ADAPTIVE FEEDFORWARD

To maintain desirable performance in the face of parameter variations, the feedforward control is made adaptive using the Fe method.[9] To this end, $\boldsymbol{C}_{\mathrm{FF}}(q)$ is replaced by $\boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))$, and the output of the adaptive feedforward controller is given by

$$\begin{aligned}\boldsymbol{u}_{\mathrm{FF}}(k) &= \boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))\tilde{\boldsymbol{a}}_0(k)\\ &= \boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))\tilde{\boldsymbol{a}}_s(k) + \boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))\boldsymbol{a}_a(k) = \tilde{\boldsymbol{\Psi}}(k)\boldsymbol{w}(k) + \boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))\boldsymbol{a}_a(k).\end{aligned} \tag{22}$$

Here $\tilde{\boldsymbol{\Psi}}(k)$ is defined as in (15), with $\boldsymbol{\psi}(k) = \boldsymbol{f}(q)(\boldsymbol{a}_s(k) + \boldsymbol{n}_0(k)) \in \mathbb{R}^{n_i n_a}$. Note that the implemented feedforward controller uses $\boldsymbol{\psi}(k) = \boldsymbol{f}(q)(\boldsymbol{a}_a(k) + \boldsymbol{a}_s(k) + \boldsymbol{n}_0(k)) \in \mathbb{R}^{n_i n_a}$. The split in (22) is for analysis purposes. The structure of the Fe method can be seen in Figure 1. The filtered error minimized by adaptation is the signal $\tilde{\boldsymbol{a}}_1(k)$ filtered by $\boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q) \in \mathbb{R}H_\infty^{n_u \times n_y}$. Here $\boldsymbol{N}(q)$ is the noise shaping filter with

$$\boldsymbol{N}(q) = N(q)\boldsymbol{I} \in \mathbb{R}H_\infty^{n_u \times n_u}, \tag{23}$$

and $\hat{\boldsymbol{S}}_2^{-1}(q)$ is the model of the inverse of the process sensitivity. The filtered error is thereby given as

$$\begin{aligned}\boldsymbol{e}(k,\boldsymbol{w}(k)) &= \boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q)\tilde{\boldsymbol{a}}_1(k),\\ &= \boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q)(\boldsymbol{y}_1(k) + \boldsymbol{y}_2(k) + \boldsymbol{n}_1(k)).\end{aligned} \tag{24}$$

With some abuse of notation, we indicate the dependence of the error on $\boldsymbol{w}(k)$ to motivate the derivations in the sequel. Note that $\hat{\boldsymbol{S}}_2^{-1}(q)$ only exists if $n_u = n_y$ while for systems with $n_u \neq n_y$ either a squaring down approach or an inner/outer factorization approach can be used; see van Zundert[26] and Berkhoff and Nijsse[27] respectively. $\boldsymbol{N}(q)$ can be used to make $\boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q)$ causal and stable. The noise shaping filter on $\tilde{\boldsymbol{a}}_0(k)$ is $\boldsymbol{N}(q) = N(q)\boldsymbol{I} \in \mathbb{R}H_\infty^{n_a \times n_a}$. The same notation is used since they only differ in size.

The parameter vector is estimated using the gradient-based update law[28]

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \frac{\boldsymbol{\Gamma}(k)}{2}\left(\frac{\partial J}{\partial \boldsymbol{w}}\right)^T, \tag{25}$$

with

$$J(\boldsymbol{w}) = \mathbb{E}\left[\boldsymbol{e}^T(k,\boldsymbol{w}(k))\boldsymbol{e}(k,\boldsymbol{w}(k))\right], \tag{26}$$

and gain matrix $\boldsymbol{\Gamma}(k) = \boldsymbol{\Gamma}(k)^T$. The gradient in (25) is written as

$$\left(\frac{\partial J(\boldsymbol{w})}{\partial \boldsymbol{w}}\right)^T = \left(\frac{\partial J(\boldsymbol{w})}{\partial \boldsymbol{e}(k,\boldsymbol{w}(k))}\frac{\partial \boldsymbol{e}(k,\boldsymbol{w}(k))}{\partial \boldsymbol{w}}\right)^T. \tag{27}$$

The filtered error $\boldsymbol{e}(k,\boldsymbol{w}(k))$ can be rewritten by substituting (5) and $\tilde{\boldsymbol{a}}_1(k) = \boldsymbol{a}_1(k) + \boldsymbol{n}_1(k)$ into (24). This yields

$$\boldsymbol{e}(k,\boldsymbol{w}(k)) = \boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q)\left[\boldsymbol{P}_1(q)\boldsymbol{a}_s(k) + \boldsymbol{n}_1(k) + \boldsymbol{S}_2(q,\boldsymbol{w}(k))\left[\boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))\tilde{\boldsymbol{a}}_s(k) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{P}_1\boldsymbol{a}_s(k) + \boldsymbol{u}_d(k)\right]\right]. \tag{28}$$

Substituting (22) gives

$$\boldsymbol{e}(k,\boldsymbol{w}(k)) = \boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q)\left[\boldsymbol{P}_1(q)\boldsymbol{a}_s(k) + \boldsymbol{n}_1(k) + \boldsymbol{S}_2(q,\boldsymbol{w}(k))\left(\tilde{\boldsymbol{\Psi}}(k)\boldsymbol{w}(k) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{P}_1\boldsymbol{a}_s(k) + \boldsymbol{u}_d(k)\right)\right]. \tag{29}$$

The gradient of (29) w.r.t. the parameter vector $\boldsymbol{w}$ is required for the Fe method. To derive an implementable AF, two assumptions are made:

Assumption 1. Slow parameter adaptation.[6]

Assumption 2. $\boldsymbol{S}_2(q,\boldsymbol{w}(k)) \approx \hat{\boldsymbol{S}}_2(q)\ \forall \boldsymbol{w} \in \mathcal{W}_l$.

Assumption 1 is an inherent limitation of the Fe method and will be used at several points of the derivation. The first point is in the following derivation of the gradient. Later, it shows up in the transformation to the frequency domain in Section 5. It is also already used in Section 3 following Lemma 1. The model $\hat{\boldsymbol{S}}_2(q)$ is a design element for the Fe method that is used to render the error in (29) (approximately) affine in $\boldsymbol{w}$. This is the basis for the derivation of the Fe method. The influence of the second item of Assumption 2 on the convergence of the Fe method is discussed in Sections 5 and 6. It implies that the change in the dynamics of $\boldsymbol{S}_2(q, \boldsymbol{w})$ as a function of $\boldsymbol{w}$ is relatively small. The bounds of the mismatch of this assumption is the main topic of Lemmas 3 and 4. If the Assumption 2 holds to equality, then the cost function is guaranteed to be convex.

Using Assumptions 1 and 2, the gradient of (29) with respect to $\boldsymbol{w}$ can be approximated as[6]

$$\frac{\partial \boldsymbol{e}(k, \boldsymbol{w}(k))}{\partial \boldsymbol{w}} \approx \tilde{\boldsymbol{\Psi}}_N(k) = \boldsymbol{N}(q)\tilde{\boldsymbol{\Psi}}(k). \tag{30}$$

This approximation holds to equality if the parameter variation is arbitrarily slow, and the Assumption 2 is an equality. Using this result, the gradient of (26) can be evaluated as

$$\left(\frac{\partial J}{\partial \boldsymbol{w}}\right)^T \approx 2\mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k))\right]. \tag{31}$$

This yields the update law

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \boldsymbol{\Gamma}(k)\mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k))\right]. \tag{32}$$

The stochastic information of $\mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k))\right]$ is rarely available in practice, hence the instantaneous approximation $\mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k))\right] \to \tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k))$ is typically used.[29]

The matrix $\boldsymbol{\Gamma}(k)$ is of major importance for the transient behavior of the parameter adaption and there are several commonly used methods to define $\boldsymbol{\Gamma}(k)$. The matrix $\boldsymbol{\Gamma}(k)$ is split as

$$\boldsymbol{\Gamma}(k) = \gamma(k)\overline{\boldsymbol{\Gamma}}(k), \tag{33}$$

with $\gamma(k) > 0$ being a scalar gain. The realizations of this equation for two commonly used methods are given in Appendix A.

Assumption 3. The following assumptions are made for $\boldsymbol{\Gamma}(k)$:

- $\boldsymbol{\Gamma}(k) = \boldsymbol{\Gamma}(k)^T \succeq 0$.
- $\boldsymbol{\Gamma}(k) \prec c_h \boldsymbol{I}$ with $c_h > 0$.
- $-c_v(k)\boldsymbol{I} \prec \mathbb{E}\left[\overline{\boldsymbol{\Gamma}}(k)\right] - \overline{\boldsymbol{\Gamma}}(k-1) \prec c_v(k)\boldsymbol{I}$ with $c_v(k) \geq 0$, and some small $c_v(k)$.
- $c_v(k) = \gamma^3(k)\delta_v$ for small $\gamma(k)$, see Appendix A.

The third item of Assumption 3 implies that $\boldsymbol{\Gamma}(k)$ is slowly varying.

The update law of (32) with the instantaneous gradient approximation and the parametrization of (33) yields

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \gamma(k)\overline{\boldsymbol{\Gamma}}(k)\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k)). \tag{34}$$

For the convergence analysis, it is convenient to express the update law of (34) in terms of the parameter error with respect to a stationary point. Therefore, the parameter error is defined as

$$\tilde{\boldsymbol{w}}(k) = \boldsymbol{w}(k) - \boldsymbol{w}^*, \tag{35}$$

where $\boldsymbol{w}^*$ are the stationary points of (34) and are defined as the solutions of the noise free orthogonality condition

$$\mathbb{E}\left[\boldsymbol{\Psi}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}^*)\right] = \boldsymbol{0}, \tag{36}$$

with the noise-free regression matrix $\boldsymbol{\Psi}(k)$ defined equivalently to (15) but with $\boldsymbol{\psi}(k) = \boldsymbol{f}(q)\boldsymbol{a}_s(k) \in \mathbb{R}^{n_i n_a}$. The update of the parameter error is

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1) &= \boldsymbol{w}(k+1) - \boldsymbol{w}^*, \\ &= \tilde{\boldsymbol{w}}(k) - \gamma(k)\overline{\boldsymbol{\Gamma}}(k)\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k)).\end{aligned} \tag{37}$$

The error vector $\boldsymbol{e}(k, \boldsymbol{w}(k))$ depends nonlinearly on $\boldsymbol{w}(k)$ and thereby on $\tilde{\boldsymbol{w}}(k)$. For the convergence analysis, the affine terms of $\tilde{\boldsymbol{w}}(k)$ and $\boldsymbol{w}^*$ are extracted. The fixed-basis, linear-in-the-parameters feedforward controller can be written as

$$\boldsymbol{\Psi}(k)\boldsymbol{w}^* = -\boldsymbol{\mathcal{S}}_2^{-1}(q, \boldsymbol{w}^*)(\boldsymbol{\mathcal{S}}_1(q, \boldsymbol{w}^*)\boldsymbol{a}_s(k) - \boldsymbol{\varepsilon}_0(k)), \tag{38}$$

where $\boldsymbol{\varepsilon}_0(k)$ contains the controller mismatch due to the truncation error in (13).

This can be used to write $\boldsymbol{y}_1(k)$ from (24) as

$$\boldsymbol{y}_1(k) = \boldsymbol{\mathcal{S}}_1(q, \boldsymbol{w}(k))\boldsymbol{a}_0(k) = -\boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))\boldsymbol{\Psi}(k)\boldsymbol{w}^* + \boldsymbol{\varepsilon}_0(k). \tag{39}$$

Similarly, (22) can be used for

$$\boldsymbol{y}_2(k) = \boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))\left(\tilde{\boldsymbol{\Psi}}(k)\boldsymbol{w}(k) + \boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{u}_d(k)\right). \tag{40}$$

Substituting (39) and (40) into (24) yields

$$\begin{aligned}\boldsymbol{e}(k, \boldsymbol{w}(k)) = &\boldsymbol{N}(q)\hat{\boldsymbol{\mathcal{S}}}_2^{-1}(q)\boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))\left[\tilde{\boldsymbol{\Psi}}(k)\tilde{\boldsymbol{w}}(k)\right] \\ &+ \boldsymbol{N}(q)\hat{\boldsymbol{\mathcal{S}}}_2^{-1}(q)\boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))\left[\tilde{\boldsymbol{\Psi}}(k) - \boldsymbol{\Psi}(k)\right]\boldsymbol{w}^* \\ &+ \boldsymbol{N}(q)\hat{\boldsymbol{\mathcal{S}}}_2^{-1}(q)\left[\boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))(\boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{u}_d(k)) + \boldsymbol{\varepsilon}_0(k) + \boldsymbol{n}_1(k)\right].\end{aligned} \tag{41}$$

Clearly, the generating functions $\boldsymbol{f}(q)$ should be chosen such that $\boldsymbol{\varepsilon}_0(k)$ is small for a given n_e. Substitution of (41) into (37) yields

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I} - \gamma(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\right]\tilde{\boldsymbol{w}}(k) + \gamma(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{g}(k) \tag{42}$$

with

$$\begin{aligned}\boldsymbol{G}_1(k) &= \tilde{\boldsymbol{\Psi}}_N^T(k)\left[\boldsymbol{N}(q)\hat{\boldsymbol{\mathcal{S}}}_2^{-1}(q)\boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))\tilde{\boldsymbol{\Psi}}(k)\right], \\ \boldsymbol{g}_2(k) &= \tilde{\boldsymbol{\Psi}}_N^T(k)\left[\boldsymbol{N}(q)\hat{\boldsymbol{\mathcal{S}}}_2^{-1}(q)\boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))\left(\boldsymbol{\Psi}(k) - \tilde{\boldsymbol{\Psi}}(k)\right)\right]\boldsymbol{w}^*, \\ \boldsymbol{g}_3(k) &= -\tilde{\boldsymbol{\Psi}}_N^T(k)\left[\boldsymbol{N}(q)\hat{\boldsymbol{\mathcal{S}}}_2^{-1}(q)\left(\boldsymbol{\mathcal{S}}_2(q, \boldsymbol{w}(k))(\boldsymbol{C}_{\mathrm{FB}}(q)\boldsymbol{n}_1(k) + \boldsymbol{u}_d(k)) + \boldsymbol{\varepsilon}_0(k) + \boldsymbol{n}_1(k)\right)\right],\end{aligned} \tag{43}$$

and

$$\boldsymbol{g}(k) = \boldsymbol{g}_2(k) + \boldsymbol{g}_3(k). \tag{44}$$

Equation (42) is the basis for the following convergence analysis, and can be used to analyse the bias and minimum mean-square error.[6,29] The analysis of the bias and mean-square error is outside the scope of this article. The evolution of the feedforward parameters is determined by (42), and is driven by the stochastic signals $\boldsymbol{a}_s(k)$, $\boldsymbol{n}_0(k)$, $\boldsymbol{n}_1(k)$ and $\boldsymbol{u}_d(k)$. Since the specific realization of these stochastic signals is a-priori unknown, the evolution of the feedforward parameters is also unknown. Therefore, a projection operator can be used to enforce $\boldsymbol{w}(k) \in \mathcal{W}_l$ such that the input-disturbance feedback loop does not become unstable during adaptation.[30] This is not pursued here since it diverges from the main objective.

5 | CONVERGENCE ANALYSIS

Assumption 2 is a suitable starting point to develop the Fe method since it renders the error in (29) approximately affine in $\boldsymbol{w}(k)$. This assumption does however hardly ever hold to equality for real systems. In this section limits for violation of this assumption will be derived.

The convergence properties of the FeAF algorithm are determined by (42). This is a discrete stochastic system due to the stochastic nature of $\boldsymbol{a}_s$, $\boldsymbol{n}_0$, $\boldsymbol{u}_d$ and $\boldsymbol{n}_1$. There exist several general tools for the analysis of discrete stochastic systems, such as the associated ordinary differential equation (ODE) method,[31] and martingale methods.[32] The analysis presented here is based on the martingale methods due to the clear analogy to deterministic stability, and shows uniform boundedness in probability of the system. This is defined as follows:

Definition 1 (Kushner[32‡]). For a given initial condition $\boldsymbol{x}(0)$, the system is uniformly bounded in probability, with probability ρ if and only if

$$\mathbb{P}\left[\sup_{0\leq k\leq\infty}\|\boldsymbol{x}(k)\|_\infty \geq \epsilon\right] \leq 1-\rho \tag{45}$$

with $\mathbb{P}(.)$ being the probability operator.

Pivotal in the derivation is the following theorem:

Theorem 2 (Kushner[32§]). *Let $\mathcal{V}(\boldsymbol{x}(k))$ be nonnegative, scalar-valued and continuous in the set $\mathcal{Z}_m \triangleq \{\boldsymbol{x}(k) : \varepsilon(\boldsymbol{x}(k)) < \mathcal{V}(\boldsymbol{x}(k)) < m\}$ and $\mathcal{Z}_0 \triangleq \{\boldsymbol{x}(k) : 0 < \mathcal{V}(\boldsymbol{x}(k)) \leq \varepsilon(\boldsymbol{x}(k))\}$. $\boldsymbol{x}(k)$ is a discrete Markov process defined until at least some $\tau > \tau_z = \inf\{k : \boldsymbol{x}(k) \notin \boldsymbol{Z}_m\}$. Let*

$$\Delta\mathcal{V}(\boldsymbol{x}(k+1)) = \mathbb{E}\left[\boldsymbol{\mathcal{V}}(\boldsymbol{x}(k+1))\right] - \boldsymbol{\mathcal{V}}(\boldsymbol{x}(k)) \leq c(\boldsymbol{x}(k+1)), \tag{46}$$

with

$$c(\boldsymbol{x}(k+1))\begin{cases}\leq 0\ \forall\ \boldsymbol{x}(k+1)\in\mathcal{Z}_m,\\ >0\ \forall\ \boldsymbol{x}(k+1)\in\mathcal{Z}_0,\end{cases} \tag{47}$$

and $c(\boldsymbol{x}(k))$ a uniformly continuous function. Then there is some $t \leq \infty$ such that

$$\boldsymbol{x}(k) \to \{\boldsymbol{x}(k) : c(\boldsymbol{x}) = 0\}, \tag{48}$$

as $k \to t$, with a probability no less than $1 - \frac{\mathcal{V}(\boldsymbol{x}_0)}{m}$. Also

$$\mathbb{P}\left[\sup_{0\leq k\leq\tau}\mathcal{V}(\boldsymbol{x}(k+1)) \geq m\right] \leq \frac{\mathcal{V}(\boldsymbol{x}(0))}{m}. \tag{49}$$

Applying Theorem 2 to (42) yields the following result:

Lemma 2. *Let the conditions of Theorem 1 be satisfied for (17), $\boldsymbol{w}(k) \in \mathcal{W}$ with $\mathcal{W} = \mathcal{W}_a \bigcap \mathcal{W}_l$, where $\mathcal{W}_l$ is defined according to equation (18) and $\mathcal{W}_a$ is defined by*

$$\boldsymbol{A} + \boldsymbol{A}^T \succ 0 \quad \forall\ \boldsymbol{w}(k) \in\ \mathcal{W}_a, \tag{50}$$

where

$$\boldsymbol{A} = \mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\boldsymbol{w}(k))\tilde{\boldsymbol{\Psi}}_N(k)\right]\right]. \tag{51}$$

Then there exists a sufficiently small $\gamma(k)$ such that the parameters, that are updated with (42), are uniformly bounded in probability.

Proof. Consider the Lyapunov candidate function

$$\mathcal{V}(\tilde{\boldsymbol{w}}(k)) = \tilde{\boldsymbol{w}}^T(k)\overline{\boldsymbol{\Gamma}}^{-1}(k-1)\tilde{\boldsymbol{w}}(k) \geq 0, \tag{52}$$

with

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I} - \gamma(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\right]\tilde{\boldsymbol{w}}(k) + \gamma(k)\overline{\boldsymbol{\Gamma}}(k)\mathbf{g}(k). \tag{53}$$

This system is uniformly bounded in probability according to Theorem 2 if

$$\Delta\mathcal{V}(\tilde{\boldsymbol{w}}(k+1)) = \mathbb{E}\left[\mathcal{V}(\tilde{\boldsymbol{w}}(k+1))\right] - \mathcal{V}(\tilde{\boldsymbol{w}}(k)) \leq c(\tilde{\boldsymbol{w}}(k+1)). \tag{54}$$

This is evaluated as

$$\begin{aligned}
\Delta\mathcal{V}(\tilde{\boldsymbol{w}}(k+1)) &= \mathbb{E}\left[\mathcal{V}(\tilde{\boldsymbol{w}}(k+1))\right] - \mathcal{V}(\tilde{\boldsymbol{w}}(k)) \\
&= \mathbb{E}\left[\tilde{\boldsymbol{w}}^T(k+1)\overline{\boldsymbol{\Gamma}}^{-1}(k)\tilde{\boldsymbol{w}}(k+1)\right] - \tilde{\boldsymbol{w}}^T(k)\overline{\boldsymbol{\Gamma}}^{-1}(k-1)\tilde{\boldsymbol{w}}(k) \\
&= \mathbb{E}\left[\gamma^2(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{G}_1^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\tilde{\boldsymbol{w}}(k)\right] - \mathbb{E}\left[\gamma(k)\tilde{\boldsymbol{w}}^T(k)\left(\boldsymbol{G}_1(k)+\boldsymbol{G}_1^T(k)\right)\tilde{\boldsymbol{w}}(k)\right] \\
&\quad - \mathbb{E}\left[\gamma^2(k)\left(\tilde{\boldsymbol{w}}^T(k)\boldsymbol{G}_1^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{g}(k) + \boldsymbol{g}^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\tilde{\boldsymbol{w}}(k) - \boldsymbol{g}^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{g}(k)\right)\right] \\
&\quad + \mathbb{E}\left[\gamma(k)\left(\tilde{\boldsymbol{w}}^T(k)\boldsymbol{g}(k) + \boldsymbol{g}^T(k)\tilde{\boldsymbol{w}}(k)\right)\right] + \mathbb{E}\left[\tilde{\boldsymbol{w}}^T(k)\overline{\boldsymbol{\Gamma}}(k)^{-1}\tilde{\boldsymbol{w}}(k)\right] - \tilde{\boldsymbol{w}}^T(k)\overline{\boldsymbol{\Gamma}}(k-1)^{-1}\tilde{\boldsymbol{w}}(k).
\end{aligned} \tag{55}$$

The stochasticity of the system originates from the different noise sources. These are present in $\boldsymbol{G}_1(k)$, $\boldsymbol{g}(k)$, $\gamma(k)$ and $\overline{\boldsymbol{\Gamma}}(k)$ and thereby in $\tilde{\boldsymbol{w}}(k+1)$. The realization of $\tilde{\boldsymbol{w}}(k)$ is however deterministic, since it is determined by the previous and fixed measurements on the system. Equation (55) can now be rewritten as

$$\begin{aligned}
\Delta\mathcal{V}(\tilde{\boldsymbol{w}}(k+1)) &= \tilde{\boldsymbol{w}}^T(k)\mathbb{E}\left[\gamma^2(k)\boldsymbol{G}_1^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\right]\tilde{\boldsymbol{w}}(k) - \tilde{\boldsymbol{w}}^T(k)\left(\mathbb{E}\left[\gamma(k)\boldsymbol{G}_1(k)\right] + \mathbb{E}\left[\gamma(k)\boldsymbol{G}_1^T(k)\right]\right)\tilde{\boldsymbol{w}}(k) \\
&\quad - \tilde{\boldsymbol{w}}^T(k)\mathbb{E}\left[\gamma^2(k)\boldsymbol{G}_1^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{g}(k)\right] - \mathbb{E}\left[\gamma^2(k)\boldsymbol{g}^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\right]\tilde{\boldsymbol{w}}(k) \\
&\quad + \mathbb{E}\left[\gamma^2(k)\boldsymbol{g}^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{g}(k)\right] + \tilde{\boldsymbol{w}}^T(k)\mathbb{E}\left[\gamma(k)\boldsymbol{g}(k)\right] + \mathbb{E}\left[\gamma(k)\boldsymbol{g}^T(k)\right]\tilde{\boldsymbol{w}}(k) \\
&\quad + \tilde{\boldsymbol{w}}^T(k)\left(\mathbb{E}\left[\overline{\boldsymbol{\Gamma}}(k)^{-1}\right] - \overline{\boldsymbol{\Gamma}}(k-1)^{-1}\right)\tilde{\boldsymbol{w}}(k) \\
&\leq \tilde{\boldsymbol{w}}^T(k)\mathbb{E}\left[\gamma^2(k)\boldsymbol{G}_1^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\right]\tilde{\boldsymbol{w}}(k) - \tilde{\boldsymbol{w}}^T(k)\left(\mathbb{E}\left[\gamma(k)\boldsymbol{G}_1(k)\right] + \mathbb{E}\left[\gamma(k)\boldsymbol{G}_1^T(k)\right]\right)\tilde{\boldsymbol{w}}(k) \\
&\quad - \tilde{\boldsymbol{w}}^T(k)\mathbb{E}\left[\gamma^2(k)\boldsymbol{G}_1^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{g}(k)\right] - \mathbb{E}\left[\gamma^2(k)\boldsymbol{g}^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\right]\tilde{\boldsymbol{w}}(k) \\
&\quad + \mathbb{E}\left[\gamma^2(k)\boldsymbol{g}^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{g}(k)\right] + \tilde{\boldsymbol{w}}^T(k)\mathbb{E}\left[\gamma(k)\boldsymbol{g}(k)\right] + \mathbb{E}\left[\gamma(k)\boldsymbol{g}^T(k)\right]\tilde{\boldsymbol{w}}(k) + \gamma^3(k)\delta_v\tilde{\boldsymbol{w}}^T(k)\tilde{\boldsymbol{w}}(k),
\end{aligned} \tag{56}$$

using Assumption 3 for the inequality. Note that only the first, second and last term of (56) depend quadratically on $\tilde{\boldsymbol{w}}(k)$. Let

$$\boldsymbol{A} = \mathbb{E}\left[\boldsymbol{G}_1(k)\right], \tag{57}$$

which is evaluated as (51). Here (23) is used to chance the order of $\boldsymbol{N}(q)$ and $\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\boldsymbol{w}(k))$. Lets define the set $\mathcal{W}_n$ as

$$\Delta\mathcal{V}(\tilde{\boldsymbol{w}}(k+1)) \geq 0 \quad \forall\ \tilde{\boldsymbol{w}}(k) \in \mathcal{W}_n, \tag{58}$$

and where $\Delta\mathcal{V}(\tilde{\boldsymbol{w}}(k+1)) = 0$ on the boundary of $\mathcal{W}_n$. Since $\gamma(k) > 0$,

$$\gamma^3(k)\delta_v\boldsymbol{I} + \mathbb{E}\left[\gamma^2(k)\boldsymbol{G}_1^T(k)\overline{\boldsymbol{\Gamma}}(k)\boldsymbol{G}_1(k)\right] - \gamma(k)\left(\boldsymbol{A}+\boldsymbol{A}^T\right) \prec 0 \quad \forall\ \boldsymbol{w}(k) \in \mathcal{W}_a \tag{59}$$

implies that $\Delta\mathcal{V}(\tilde{\boldsymbol{w}}(k+1)) \leq 0$ if $\tilde{\boldsymbol{w}}(k)$ is outside $\mathcal{W}_n$, and inside $\mathcal{W}_a$ since (59) is dominant for large $\tilde{\boldsymbol{w}}(k)$. Note that this implies that $\mathcal{W}_n \subset \mathcal{W}_a$. If $\boldsymbol{A}+\boldsymbol{A}^T \succ \boldsymbol{0}\ \forall\ \boldsymbol{w}(k) \in \mathcal{W}_a$, there always exists a sufficiently small $\gamma(k)$ if $\boldsymbol{w}(k) \in \mathcal{W}_a$ such that (59) is satisfied. The set $\mathcal{W}_a$ is compact if all signals are bounded.

According to the relations of Appendix B, the boundedness of the signals in (56) reduces to:

1. $\|\boldsymbol{G}_1(k)\|_{\text{pow}} \leq \infty$,
2. $\|\boldsymbol{g}(k)\|_{\text{pow}} \leq \infty$,
3. $\|\overline{\boldsymbol{\Gamma}}(k)\|_{\text{pow}} \leq \infty$.

The first and second condition hold if $\boldsymbol{S}_2(q,\overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_u}$, $\boldsymbol{f}_i(q) \in \mathbb{R}H_\infty^{n_i n_a}$, $\boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q) \in \mathbb{R}H_\infty^{n_y \times n_u}$, and $\boldsymbol{a}_0(k),\ \boldsymbol{n}_0(k),\ \boldsymbol{n}_1(k),\ \boldsymbol{u}_d(k),\ \boldsymbol{\varepsilon}_0(k) \in \mathcal{L}_{\text{pow}}$, and if $\boldsymbol{w}(k)$ is sufficiently slowly varying. The generating transfer functions $\boldsymbol{f}_i(q)$ and filter $\boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q)$ can be chosen to be in $\mathbb{R}H_\infty$ by design. From the proof of Lemma 1 it follows that $\boldsymbol{S}_2(q,\overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_u}\ \forall \overline{\boldsymbol{w}} \in \mathcal{W}_l$. The conditions $\boldsymbol{a}_0(k),\ \boldsymbol{n}_0(k),\ \boldsymbol{n}_1(k),\ \boldsymbol{u}_d(k) \in \mathcal{L}_{\text{pow}}$ hold by definition. Rewrite (39) as

$$\boldsymbol{\varepsilon}_0(k) = \boldsymbol{S}_1(q,\boldsymbol{w}(k))\boldsymbol{a}_0(k) + \boldsymbol{S}_2(q,\boldsymbol{w}(k))\boldsymbol{\Psi}(k)\boldsymbol{w}^*. \tag{60}$$

Following the proof of Lemma 1 it follows that $\boldsymbol{S}_1(q,\overline{\boldsymbol{w}}) \in \mathbb{R}H_\infty^{n_y \times n_a}\ \forall\ \overline{\boldsymbol{w}} \in \mathcal{W}_l$, hence $\boldsymbol{\varepsilon}_0(k) \in \mathcal{L}_{\text{pow}}$. The third condition is satisfied under Assumption 3.

Equations (52) and (56) imply that $\tilde{\boldsymbol{w}}(k) \in \mathcal{L}_\infty$ if $\boldsymbol{A} + \boldsymbol{A}^T \succ \boldsymbol{0}\ \forall \boldsymbol{w}(k) \in \mathcal{W}_a$ for some sufficiently small $\gamma(k)$ and if $\boldsymbol{w}(k) \in \mathcal{W}_l$, then (52) is a Lyapunov function and the update law in (42) is uniformly bounded in probability. ∎

Note that the probability of the boundedness depends on $\mathcal{W}_a$. For this, let w_a be the boundary of $\mathcal{W}_a$, such that $w_a = \mathcal{Z}_m$ with the set $\mathcal{Z}_m$ of (47). Therefore, boundedness w.p. 1. can only be achieved when $\inf(w_a) = \infty$, or when a projection operator is used.

Equation (59) is used to prove that the elements of (56) that are quadratic in $\tilde{\boldsymbol{w}}(k)$ are smaller than 0. Therefore, $\Delta\mathcal{V}(\tilde{\boldsymbol{w}}(k+1)) < 0$ for some sufficiently large $\tilde{\boldsymbol{w}}(k)$. The remainder of the elements of (56) might have varying polarity over time. Hence, $\mathcal{W}_n$ is in the neighborhood of the origin. According to Theorem 2, the algorithm will converge to the boundary of $\mathcal{W}_n$.

In the convergence analysis, the sets $\mathcal{W}_l$, $\mathcal{W}_a$ and $\mathcal{W}_n$ are used. It was concluded that the system is stable if $\tilde{\boldsymbol{w}}(k) \in \mathcal{W}$ with $\mathcal{W} = \mathcal{W}_a \bigcap \mathcal{W}_l$, and that the algorithm converges to the boundary of $\mathcal{W}_n$. Loosely speaking, this means that the system converges if Assumption 2 holds, the current parameters do not destabilize the system over the input-disturbance interaction, and that the change in dynamics incurred by the parameters stays within some bound. This bound will be quantified in the frequency domain in the sequel. This also means that the stationary point of the algorithm with noisy data should be within $\mathcal{W}$. This poses limits on the input-disturbance interaction. For AVIS, it is straightforward to use this limit to derive minimal mass-ratios between m_1 and m_0 (Figure 2), as well as minimal stiffness ratios and limits on the relative damping of the floor.

The set $\mathcal{W}_a$ is a sublevel set that is in general non-convex and that determines partly the convergence of the update Equation (42). The quantification of $\mathcal{W}_a$ requires an exact description of $\boldsymbol{S}_2(q,\boldsymbol{w}(k))$, and is usually not available for real systems. To relax this requirement, the nowadays classical result of Wang and Ren[11] is used to rewrite the condition $\boldsymbol{A} + \boldsymbol{A}^T \succ \boldsymbol{0}$ into the frequency domain. Originally Wang and Ren[11] considered the FxAF framework. This is adapted to the FeAF framework by Beijen et al.[6] Here, it is modified to accommodate for $\boldsymbol{P}_4(q)$. Originally Wang and Ren[11] used the associated ODE method to derive their results. Beijen et al.[6] switched to the independence assumption, which is well-known in the adaptive filtering literature.[29] Here, the martingale method is used. It is interesting to note that the same result is found here.

Lemma 3 (Wang and Ren[11]). *If, for sufficiently small* $\gamma(k)$,

$$\left(\hat{\boldsymbol{S}}_2^{-1}\left(e^{j\omega}\right)\boldsymbol{S}_2\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\right)^H + \hat{\boldsymbol{S}}_2^{-1}\left(e^{j\omega}\right)\boldsymbol{S}_2\left(e^{j\omega},\overline{\boldsymbol{w}}\right) \succ \boldsymbol{0}\ \forall\ \overline{\boldsymbol{w}} \in \mathcal{W}_b, \quad -\pi \le \omega \le \pi, \tag{61}$$

then $\boldsymbol{A} + \boldsymbol{A}^T \succ \boldsymbol{0}\ \forall\ \boldsymbol{w}(k) \in \mathcal{W}_b$.

Proof. The property $\boldsymbol{A} + \boldsymbol{A}^T \succ \boldsymbol{0}$ is ensured by $\overline{\boldsymbol{A}} + \overline{\boldsymbol{A}}^T \succ \boldsymbol{0}$ under Assumption 1, similar to the proof of Lemma 2, with

$$\overline{\boldsymbol{A}} = \mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\overline{\boldsymbol{w}})\tilde{\boldsymbol{\Psi}}_N(k)\right]\right]. \tag{62}$$

The order of $\boldsymbol{N}$ and $\hat{\boldsymbol{S}}_2^{-1}\boldsymbol{S}_2$ in (57) is exchanged as a result of (23). Let $\boldsymbol{l} \in \mathbb{R}^{n_i n_a n_u \times 1}$ be an arbitrary nonzero vector, and

$$\boldsymbol{l}^T\left(\overline{\boldsymbol{A}} + \overline{\boldsymbol{A}}^T\right)\boldsymbol{l} = \mathbb{E}\left[\xi(k)^T\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\overline{\boldsymbol{w}})\xi(k)\right]\right] + \mathbb{E}\left[\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\overline{\boldsymbol{w}})\xi(k)\right]^T\xi(k)\right], \tag{63}$$

where $\boldsymbol{\xi}(k) = \tilde{\boldsymbol{\Psi}}_N(k)\boldsymbol{l}$. Equation (63) has a scalar output, therefore

$$\boldsymbol{l}^T\left(\overline{\boldsymbol{A}} + \overline{\boldsymbol{A}}^T\right)\boldsymbol{l} = Tr\left\{\mathbb{E}\left[\boldsymbol{\xi}(k)^T\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\overline{\boldsymbol{w}})\boldsymbol{\xi}(k)\right]\right] + \mathbb{E}\left[\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\overline{\boldsymbol{w}})\boldsymbol{\xi}(k)\right]^T\boldsymbol{\xi}(k)\right]\right\}. \tag{64}$$

The cyclic property $Tr\{\boldsymbol{u}_1\boldsymbol{u}_2\} = Tr\{\boldsymbol{u}_2\boldsymbol{u}_1\}$ is used to express (64) as

$$\boldsymbol{l}^T\left(\overline{\boldsymbol{A}} + \overline{\boldsymbol{A}}^T\right)\boldsymbol{l} = Tr\left\{\mathbb{E}\left[\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\overline{\boldsymbol{w}})\boldsymbol{\xi}(k)\right]\boldsymbol{\xi}(k)^T\right] + \mathbb{E}\left[\boldsymbol{\xi}(k)\left[\hat{\boldsymbol{S}}_2^{-1}(q)\boldsymbol{S}_2(q,\overline{\boldsymbol{w}})\boldsymbol{\xi}(k)\right]^T\right]\right\}. \tag{65}$$

Parseval's theorem can be used to convert this expression into the frequency domain[33]

$$\boldsymbol{l}^T\left(\overline{\boldsymbol{A}} + \overline{\boldsymbol{A}}^T\right)\boldsymbol{l} = \frac{1}{2\pi}\int_{-\pi}^{\pi} Tr\left\{\boldsymbol{S}_{\xi}\left(e^{j\omega}\right)\left[\left(\hat{\boldsymbol{S}}_2^{-1}\left(e^{j\omega}\right)\boldsymbol{S}_2\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\right)^H + \hat{\boldsymbol{S}}_2^{-1}\left(e^{j\omega}\right)\boldsymbol{S}_2\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\right]\right\}\mathrm{d}\omega, \tag{66}$$

where $\boldsymbol{S}_{\xi}\left(e^{j\omega}\right)$ is the power spectrum of $\boldsymbol{\xi}(k)$. If there is persistent excitation that is sufficiently rich[34] it holds that $\boldsymbol{S}_{\xi}\left(e^{j\omega}\right) \succ \boldsymbol{0}$. Since the trace of the product of positive-definite matrices is positive, (61) implies that $\boldsymbol{l}^T\left(\overline{\boldsymbol{A}} + \overline{\boldsymbol{A}}^T\right)\boldsymbol{l} > 0$. For sufficiently small $\gamma(k)$, Equation (61) is a sufficient condition for $\boldsymbol{A} + \boldsymbol{A} \succ \boldsymbol{0}$. ■

For a SISO plant, the condition of (61) reduces to[11]

$$Re\left[\hat{S}_2^{-1}\left(e^{j\omega}\right)S_2\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\right] > 0. \tag{67}$$

Hence the name SPR. Lemma 3 provides a sufficient, but very restrictive condition that can be used in Lemma 2, for which holds that $\mathcal{W}_b \subset \mathcal{W}_a$. This condition is almost never satisfied in practice, since it requires a good phase approximation up to the Nyquist frequency. This typically requires high dimensional models, which are difficult to obtain, have a high computational complexity and are sensitive for changes in the system. Therefore, in the next section, the SPR condition will be relaxed to a less restrictive and thus more practically useful condition.

6 | RELAXED STABILITY CONDITION

The sufficient condition of Lemma 3 is restrictive for practical applications. This condition can be relaxed with the following power weighted strictly positive real (PWSPR) lemma. For this lemma, the index notation is used, see Appendix C for an overview of the relevant properties of this notation.

Lemma 4. *Let*

$$\boldsymbol{Q}(\overline{\boldsymbol{w}}) \succ \boldsymbol{0}\ \forall\ \overline{\boldsymbol{w}} \in \mathcal{W}_a, \tag{68}$$

with

$$\boldsymbol{Q}(\overline{\boldsymbol{w}}) = \int_{-\pi}^{\pi} e_{mpv}S_{\tilde{\psi},N,pq}\left(e^{j\omega}\right)\mathcal{P}_{mr}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)e_{rqw}\mathrm{d}\omega, \tag{69}$$

and

$$\boldsymbol{\mathcal{P}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right) = \left(\hat{\boldsymbol{S}}_2^{-1}\left(e^{j\omega}\right)\boldsymbol{S}_2\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\right)^H + \hat{\boldsymbol{S}}_2^{-1}\left(e^{j\omega}\right)\boldsymbol{S}_2\left(e^{j\omega},\overline{\boldsymbol{w}}\right), \tag{70}$$

and mpqrtvw indices in the index notation. Then, for sufficiently small $\gamma(k)$,

$$\boldsymbol{A} + \boldsymbol{A}^T \succ 0\ \forall\ \boldsymbol{w}(k) \in \mathcal{W}_a. \tag{71}$$

Equation (71) ensures that Lemma 2 is satisfied.

Proof. The property $\boldsymbol{A}+\boldsymbol{A}^T \succ \boldsymbol{0}$ is ensured by $\overline{\boldsymbol{A}}+\overline{\boldsymbol{A}}^T \succ \boldsymbol{0}$ for sufficiently slowly varying $\boldsymbol{w}(k)$, similarly to the proof of Lemmas 2 and 3. Let $\boldsymbol{l}_i \in \mathbb{R}^{n_i n_a \times 1}$ be an arbitrary nonzero vector and $\boldsymbol{l} = \left[\boldsymbol{l}_1^T, \ \dots \ , \boldsymbol{l}_{n_u}^T\right]^T \in \mathbb{R}^{n_i n_a n_u}$. The block diagonal definition $\boldsymbol{\Psi}(k)$ of (15) can be used to rewrite

$$\boldsymbol{\zeta}(k) = \tilde{\boldsymbol{\Psi}}(k)\boldsymbol{l} = \underbrace{\begin{bmatrix} \boldsymbol{l}_1^T \\ \vdots \\ \boldsymbol{l}_{n_u}^T \end{bmatrix}}_{\boldsymbol{L}} \tilde{\boldsymbol{\psi}}(k), \tag{72}$$

with $\boldsymbol{L} \in \mathbb{R}^{n_u \times n_i n_a}$. The PSD of $\boldsymbol{\zeta}(k)$ is given by

$$\boldsymbol{S}_{\zeta}\left(e^{j\omega}\right) = \boldsymbol{L}\boldsymbol{S}_{\tilde{\psi}}\left(e^{j\omega}\right)\boldsymbol{L}^T, \tag{73}$$

with $\boldsymbol{S}_{\tilde{\psi}}\left(e^{j\omega}\right)$ the PSD of $\tilde{\boldsymbol{\psi}}(k)$. The PSD of $\boldsymbol{\xi}(k) = \tilde{\boldsymbol{\Psi}}_N\boldsymbol{l}$ can now be expressed as

$$\boldsymbol{S}_{\xi}\left(e^{j\omega}\right) = \boldsymbol{N}\left(e^{j\omega}\right)\boldsymbol{L}\boldsymbol{S}_{\tilde{\psi}}\left(e^{j\omega}\right)\boldsymbol{L}^T\boldsymbol{N}^H\left(e^{j\omega}\right). \tag{74}$$

Using (23) yields

$$\boldsymbol{S}_{\xi}\left(e^{j\omega}\right) = \boldsymbol{L}\boldsymbol{S}_{\tilde{\psi}}\left(e^{j\omega}\right)S_N\left(e^{j\omega}\right)\boldsymbol{L}^T, \tag{75}$$

where $S_N\left(e^{j\omega}\right) = N\left(e^{j\omega}\right)N^H\left(e^{j\omega}\right)$. Thereby, (66) can be rewritten by substituting (70) as

$$\boldsymbol{l}^T\left(\overline{\boldsymbol{A}}+\overline{\boldsymbol{A}}^T\right)\boldsymbol{l} = \frac{1}{2\pi}\int_{-\pi}^{\pi} Tr\left\{\boldsymbol{L}\boldsymbol{S}_{\tilde{\psi},N}\left(e^{j\omega}\right)\boldsymbol{L}^T\boldsymbol{\mathcal{P}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\right\}\mathrm{d}\omega, \tag{76}$$

with $\boldsymbol{S}_{\tilde{\psi},N}\left(e^{j\omega}\right) = S_N\left(e^{j\omega}\right)\boldsymbol{S}_{\tilde{\psi}}\left(e^{j\omega}\right)$. Now reorder this equation as

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} Tr\left\{\boldsymbol{L}\boldsymbol{S}^H_{\tilde{\psi},N}\left(e^{j\omega}\right)\boldsymbol{L}^T\boldsymbol{\mathcal{P}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\right\}\mathrm{d}\omega = \frac{1}{2\pi}Tr\left\{\boldsymbol{L}\int_{-\pi}^{\pi}\boldsymbol{S}^H_{\tilde{\psi},N}\left(e^{j\omega}\right)\boldsymbol{L}^T\boldsymbol{\mathcal{P}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\mathrm{d}\omega\right\}. \tag{77}$$

The part inside the trace operator can be rewritten as

$$\begin{aligned}\boldsymbol{L}\int_{-\pi}^{\pi}\boldsymbol{S}_{\tilde{\psi},N}\left(e^{j\omega}\right)\boldsymbol{L}^T\boldsymbol{\mathcal{P}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\mathrm{d}\omega &= \sum_p\sum_q\sum_r L_{\mathrm{mp}}\int_{-\pi}^{\pi}S_{\tilde{\psi},N,\mathrm{pq}}\left(e^{j\omega}\right)L_{\mathrm{rq}}\mathcal{P}_{\mathrm{rt}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\mathrm{d}\omega \\ &= L_{\mathrm{mp}}\int_{-\pi}^{\pi}S_{\tilde{\psi},N,\mathrm{pq}}\left(e^{j\omega}\right)L_{\mathrm{rq}}\mathcal{P}_{\mathrm{rt}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\mathrm{d}\omega,\end{aligned} \tag{78}$$

where first index notation is used, and in the second equality the Einstein summation convention. A short introduction in the index notation is given in Appendix C. This can be reordered to

$$\begin{aligned}\frac{1}{2\pi}Tr\left\{L_{\mathrm{mp}}\int_{-\pi}^{\pi}S_{\tilde{\psi},N,\mathrm{pq}}\left(e^{j\omega}\right)L_{\mathrm{rq}}\mathcal{P}_{\mathrm{rt}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\mathrm{d}\omega\right\} &= \frac{1}{2\pi}Tr\left\{L_{\mathrm{mp}}\int_{-\pi}^{\pi}S_{\tilde{\psi},N,\mathrm{pq}}\left(e^{j\omega}\right)\mathcal{P}_{\mathrm{rt}}\left(e^{j\omega}\right)\mathrm{d}\omega\ L_{\mathrm{rq}}\right\}, \\ &= \frac{1}{2\pi}L_{\mathrm{mp}}\int_{-\pi}^{\pi}\delta_{\mathrm{mt}}S_{\tilde{\psi},N,\mathrm{pq}}\left(e^{j\omega}\right)\mathcal{P}_{\mathrm{rt}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\mathrm{d}\omega\ L_{\mathrm{rq}}, \\ &= \frac{1}{2\pi}L_{\mathrm{mp}}\int_{-\pi}^{\pi}S_{\tilde{\psi},N,\mathrm{pq}}\left(e^{j\omega}\right)\mathcal{P}_{\mathrm{mr}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\mathrm{d}\omega\ L_{\mathrm{rq}},\end{aligned} \tag{79}$$

where δ_{mt} is the Kronecker delta. Now introduce the contraction tensor e such that

$$L_{\mathrm{mp}} = e_{\mathrm{mpv}}l_v, \quad L_{\mathrm{rq}} = e_{\mathrm{rqw}}l_w, \tag{80}$$

see Appendix C. Substitution in (79) yields

$$\frac{1}{2\pi}L_{\text{mp}}\int_{-\pi}^{\pi}S_{\tilde{\psi},N,\text{pq}}\left(e^{j\omega}\right)\mathcal{P}_{\text{mr}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\text{d}\omega\; L_{\text{rq}}=\frac{1}{2\pi}l_v\underbrace{\int_{-\pi}^{\pi}e_{\text{mpv}}S_{\tilde{\psi},N,\text{pq}}\left(e^{j\omega}\right)\mathcal{P}_{\text{mr}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)e_{\text{rqw}}\text{d}\omega}_{Q_{vw}(\overline{\boldsymbol{w}})}\;l_w,$$

$$=\frac{1}{2\pi}l_vQ_{vw}(\overline{\boldsymbol{w}})\;l_w, \tag{81}$$

■

hence for sufficiently small $\gamma(k)$, Equation (68) implies (71).

Equation (69) is an exact frequency domain version of (50), hence in both (69) and (50) the set $\mathcal{W}_a$ is used. The quantification of $\mathcal{W}_a$ is a computationally virtually intractable problem since (81) is nonlinear in $\boldsymbol{w}(k)$. This nonlinearity originates from (7). An approximate quantification of $\mathcal{W}_a$ can be found by validating (68) for a grid of $\boldsymbol{w}(k)\in\mathcal{W}_l$.

Satisfying the SPR condition of (61) implies satisfying (68). Thereby, Equation (68) relaxes the convergence condition for filtered error adaptive feedforward.

Lemma 2 can be combined with Lemma 1, and Lemma 4 to yield the main result of this article.

Theorem 3. *Let the conditions of Theorem 1 be satisfied for (17), and let $\boldsymbol{w}(k)\in\mathcal{W}$ with $\mathcal{W}=\mathcal{W}_l\bigcap\mathcal{W}_a$, with $\mathcal{W}_l$ defined in (18), and $\mathcal{W}_a$ defined in (50). Then the update law in Equation (42) is uniformly bounded in probability, for sufficiently small $\gamma(k)$ and arbitrary small c_v.*

Proof. The proof of Theorem 3 follows directly from the proofs of Lemmas 1,2, and 4 and is therefore omitted. ■

For systems with $\boldsymbol{P}_4(q)=\boldsymbol{0}$, we have the following corollary. The proof directly follows from (18), and Lemma 2.

Corollary 1. *If $\boldsymbol{P}_4(q)=\boldsymbol{0}$, the set $\mathcal{W}$ is unbounded since $\boldsymbol{S}_2(q)$ becomes independent of $\boldsymbol{w}(k)$, and the additional feedback loop over the feedforward controller disappears.*

The implication of this corollary is that if there is no input-disturbance interaction, then the PWSPR lemma does not depend on $\boldsymbol{w}(k)$, and only one evaluation is needed for a convergence guarantee.

The upperbound for $\gamma(k)$ is typically determined by the slow parameter variation assumption, and is hence related to the system dynamics. An exact quantification of this upperbound is difficult to obtain and to the best of the authors knowledge no general procedure is available for the upperbound. An approximate guideline for the user is to choose the time-constant for the adaptation such that it is an order of magnitude longer than the slowest time-constant of the closed loop system dynamics.

6.1 | Interpretation

From experimental results, it is known that the SPR condition for real mechanical systems is almost never satisfied, but the adaptive feedforward controller parameters can still converge.[6] The engineering intuition of this phenomena is that if there is no excitation power on the frequencies with the model mismatch, this does not result in divergence either. This intuition is made explicit by Equations (68) and (69), where $\boldsymbol{S}_{\tilde{\psi},N}$ can be seen as a frequency dependent weighting on the integral over $\boldsymbol{\mathcal{P}}$. This weighting is dependent on $\boldsymbol{f}$, see (15), the input spectrum of $\tilde{\boldsymbol{a}}_s(k)$ and $N(q)$. The spectrum of $\tilde{\boldsymbol{a}}_s(k)$ is easily measured in the system by turning off all controllers. $\boldsymbol{f}$ is chosen to minimize $\boldsymbol{\epsilon}_0(k)$ for a given n_e and $\tilde{\boldsymbol{a}}_s(k)$ is determined by the experimental setup, $N(q)$ is relevant for the convergence condition, while it is not in the feedforward and can therefore specifically be designed to satisfy the convergence condition in Equation (68).

7 | FILTER DESIGN

The condition of Equation (68) provides a convergence check of the system. However, it does not directly yield insight in the design considerations for $N(q)$. To this end, rewrite Equation (81) as

$$\frac{1}{2\pi}l_v\int_{-\pi}^{\pi}e_{\text{mpv}}S_{\tilde{\psi},N,\text{pq}}\left(e^{j\omega}\right)\mathcal{P}_{\text{mr}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)e_{\text{rqw}}\;l_w=\frac{1}{2\pi}l_v\int_{-\pi}^{\pi}\mathcal{M}_{\text{wv}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\text{d}\omega\;l_w, \tag{82}$$

and take the eigenvalue decomposition of (68) as

$$\boldsymbol{Q}(\overline{\boldsymbol{w}})\boldsymbol{V}(\overline{\boldsymbol{w}}) = \boldsymbol{V}(\overline{\boldsymbol{w}})\boldsymbol{D}(\overline{\boldsymbol{w}}), \tag{83}$$

with $\boldsymbol{D}(\overline{\boldsymbol{w}})$ a diagonal matrix. Since $\boldsymbol{Q}(\overline{\boldsymbol{w}})$ is a symmetric matrix, the right eigenvectors are orthogonal. The matrix of right eigenvectors $\boldsymbol{V}(\overline{\boldsymbol{w}})$ can be used to decompose the symmetric $\mathcal{M}_{\text{wv}}$ into

$$\boldsymbol{\mathcal{D}}(e^{j\omega}, \overline{\boldsymbol{w}}) = \boldsymbol{V}^T(\overline{\boldsymbol{w}})\boldsymbol{\mathcal{M}}(e^{j\omega}, \overline{\boldsymbol{w}})\boldsymbol{V}(\overline{\boldsymbol{w}}). \tag{84}$$

The off-diagonal elements of $\boldsymbol{\mathcal{D}}(e^{j\omega}, \overline{\boldsymbol{w}})$ are non-zero, but integrate to zero due to the diagonalization in (83). The off-diagonal elements do therefore not contribute to the eigenvalues of $\boldsymbol{Q}(\overline{\boldsymbol{w}})$ and it is sufficient to check the diagonal elements $\mathcal{D}_{ii}(e^{j\omega}, \overline{\boldsymbol{w}})$ of $\boldsymbol{\mathcal{D}}(e^{j\omega}, \overline{\boldsymbol{w}})$. This diagonal contains the magnitude and the sign of the contribution of $\boldsymbol{\mathcal{M}}(e^{j\omega}, \overline{\boldsymbol{w}})$ per frequency to the eigenvalues of $\boldsymbol{Q}(\overline{\boldsymbol{w}})$. A positive definite matrix has only positive eigenvalues, hence $\boldsymbol{\mathcal{D}}(e^{j\omega}, \overline{\boldsymbol{w}})$ can be used to identify frequencies where the improvement of the model $\hat{\boldsymbol{S}}_2(q)$ improves the convergence behavior of the AF.

The filter $\boldsymbol{N}(q)$ serves three purposes. First, it is used to shape the spectrum of the filtered error $\boldsymbol{e}(k)$ to be minimized. Next is to make $\boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q)$ stable and proper, that is, $\boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q) \in \mathbb{R}H_\infty^{n_y \times n_u}$, and third it can be used to attenuate $\boldsymbol{S}_{\tilde{\boldsymbol{\psi}}}(e^{j\omega})$ on frequencies where $\boldsymbol{\mathcal{M}}(e^{j\omega}, \overline{\boldsymbol{w}})$ contributes negatively to $\boldsymbol{Q}(\overline{\boldsymbol{w}})$. For these purposes, it is proposed to use the filter structure

$$N(q) = N_s(q)N_f(q), \tag{85}$$

where $N_s(q)$ is used to shape the spectrum, and $N_f(q)$ for power attenuation in regions with model mismatch. To design the filter, the following steps can be taken:

1. Choose $N_s(q)$ to shape the spectrum of $\boldsymbol{e}(k)$ for the desired error reduction in $\boldsymbol{a}_1(k)$.
   a. For example, $N_s(q)$ can contain notch filters to remove the 50 Hz mains frequency and its harmonics.
   b. This filter can in be used to compensate partly for the filtering introduced by $\hat{\boldsymbol{S}}_2^{-1}(q)$.[35]
2. Choose $N_f(s)$ as a bandpass filter.
   a. The orders of the high- and low-pass filter are lower bounded by $\boldsymbol{N}(q)\hat{\boldsymbol{S}}_2^{-1}(q) \in \mathbb{R}H_\infty^{n_y \times n_u}$. The upper bound on the order is determined by the computational complexity of the filter.
   b. Choose the corner frequencies of the high- and low-pass filters such that $N_f(e^{j\omega}) \approx 1 \; \forall \; \omega \in \Omega_p$, with $\left\{\omega \in \Omega_p \; : \; \mathcal{D}_{ii}(e^{j\omega}) \geq 0\right\}$, and $N_f(e^{j\omega}) \approx 0 \; \forall \; \omega \notin \Omega_p$.

Note that the design for $N_s(q)$ is case dependent and can be unity.

8 | EXPERIMENTAL VALIDATION

Theorem 3 is evaluated for the experimental AVIS setup that is used by Beijen et al.[6] and Hakvoort and Beijen,[35] and with the filters described by Hakvoort and Beijen.[35] The reader is referred to Tjepkema,[36] Beijen et al.[6] and Hakvoort and Beijen[35] for more details about the setup. The system has $n_a = 6$, $n_u = 6$, $n_y = 6$.

One element of the transfer function matrices $\boldsymbol{P}_1(s)$ and $\boldsymbol{P}_2(s)$, with s the Laplace variable, is modeled by

$$P_1(s) = \frac{ds + k}{ms^2 + ds + k}\mathcal{U}(s), \quad P_2(s) = \frac{s^2}{ms^2 + ds + k}\frac{\omega_i}{s + \omega_i}\mathcal{U}(s) \tag{86}$$

with the mass m, damping constant d, stiffness k, actuator pole ω_i and $\mathcal{U}(s)$ the unmodelled high frequency dynamics. $\boldsymbol{P}_2(s)$ is diagonally dominant with similar dynamics on the diagonal. The relevant filters, which are designed using engineering insight, are repeated here for convenience. They are defined in the continuous domain, and their discrete counterparts are obtained by Tustin discretization. The feedback controller is given as

$$\boldsymbol{C}_{\text{FB}}(s) = \frac{\omega_i}{s + \omega_i}k_v\frac{s^2 + 0.01\omega_n s + \omega_n^2}{s^2 + \omega_n s + \omega_n^2}\boldsymbol{I}, \tag{87}$$

with $k_v = 225$, $\omega_i = 1$ and $\omega_n = 962 \cdot 2\pi$. The generating transfer functions

$$\boldsymbol{f}(s) = \begin{bmatrix} 1 & \beta H(s) & \beta^2 H^2(s) \end{bmatrix}, \tag{88}$$

are formed with the weak integrator

$$H(s) = \frac{1 - L(s)}{s}, \tag{89}$$

and

$$L(s) = \left(\frac{\alpha}{s+\alpha}\right)^5, \tag{90}$$

where $\alpha = 4\pi$ and $\beta = \|H(q)\|_2^{-1}$. To make the implementation internally stable, a minimal realization of $H(q)$ should be used. Furthermore,

$$\hat{\boldsymbol{S}}_2(s) = \left[\boldsymbol{I} - \boldsymbol{P}_2(q)\boldsymbol{C}_{\mathrm{FB}}(q)\right]^{-1}\boldsymbol{P}_2(q) = \hat{S}_2(s)\boldsymbol{I}. \tag{91}$$

This description can be used since $\boldsymbol{S}(q, \boldsymbol{w}(k))$ is a diagonally dominant system, with similar entries on the diagonal. The noise shaping filter is defined as

$$N(s) = \hat{S}_2(s)W_n(s)\left(\frac{1000}{s+1000}\right)^3\left(\frac{s}{s+3\alpha}\right)^3, \tag{92}$$

with $W_n(s)$ a series of notch filters on 50, 100, 200, 948 Hz. The notch filters are discretized using pole-zero matching. The set $\mathcal{W}_l = conv(\mathcal{W}_g)$ depends on $\sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}(e^{j\omega}, \overline{\boldsymbol{w}})\right)$, see Equation (18), with

$$\sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}\left(e^{j\omega}, \overline{\boldsymbol{w}}\right)\right) = \sigma_{\max}\left(\boldsymbol{\mathcal{F}}\left(e^{j\omega}\right)\boldsymbol{w}_d\right). \tag{93}$$

$\boldsymbol{\mathcal{F}}(q)$ contains repeated entries, as can be seen in (15), therefore it is sufficient to evaluate (18) for a reduced set of parameters related to each of the the basis functions

$$\sigma_{\max}\left(\boldsymbol{C}_{\mathrm{FF}}\left(e^{j\omega}, \overline{\boldsymbol{w}}\right)\right) = \sigma_{\max}\left(\boldsymbol{f}_r\left(e^{j\omega}\right)\boldsymbol{w}_r\right), \tag{94}$$

with

$$\boldsymbol{f}_r(q) = \sqrt{n_u}\boldsymbol{f}(q). \tag{95}$$

The singular values are sign-invariant, hence the set $\mathcal{W}_g \in \mathbb{R}^{2^{n_a n_u n_e}}$ is the hypercube constructed from all permutations of $\boldsymbol{w}_{r,\max}$ and $-\boldsymbol{w}_{r,\max}$ where $\boldsymbol{w}_{r,\max}$ are the maximum values of $\boldsymbol{w}_r(k)$

$$\mathcal{W}_g = \left\{ col(w_{g,1}, \ \dots \ , w_{g,n_a n_u n_e}) \ \middle| \ w_{g,1} = \left\{-w_{r,\max,1}, w_{r,\max,1}\right\}, \ \dots \ , w_{g,n_a n_u} = \left\{-w_{r,\max,n_a n_u n_e}, w_{r,\max,n_a n_u n_e}\right\}\right\}. \tag{96}$$

The construction of $\mathcal{W}_g$ reduces to finding a $\boldsymbol{w}_{r,\max}$ that satisfies

$$\sigma_{\max}\left(\boldsymbol{f}_r\boldsymbol{w}_r\right) \leq \sigma_{\min}(\boldsymbol{I} - \boldsymbol{C}_{\mathrm{FB}}(\boldsymbol{P}_1\boldsymbol{P}_4 + \boldsymbol{P}_2))\sigma_{\max}(\boldsymbol{P}_4)^{-1}, \tag{97}$$

where the argument $e^{j\omega}$ is left out, and maximizes $\|\boldsymbol{V}\boldsymbol{w}_{r,\max}\|_2$ with a suitable diagonal scaling matrix $\boldsymbol{V}$. This scaling matrix is used to shape $\mathcal{W}_g$. This can be written as an LMI with a constraint based on (97) for every frequency of the FRF. For the considered system only a reduced set $\boldsymbol{w}_r \in \mathbb{R}^3$ is required. The transfer function $\boldsymbol{P}_4(q)$ is measured to validate (97), and the singular values of the FRF can be seen in Figure 3. The vector $\boldsymbol{w}_r$ is found by maximizing $\|\boldsymbol{V}\boldsymbol{w}_{r,\max}\|_2$ for

$$\boldsymbol{V} = \begin{bmatrix} (w_1^\circ)^{-1} & 0 & 0 \\ 0 & (w_2^\circ)^{-1} & 0 \\ 0 & 0 & (w_3^\circ)^{-1} \end{bmatrix}, \tag{98}$$

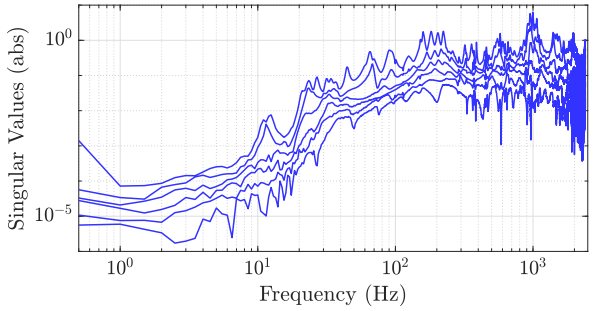

FIGURE 3 Singular values of the FRF of $\boldsymbol{P}_4(q)$.

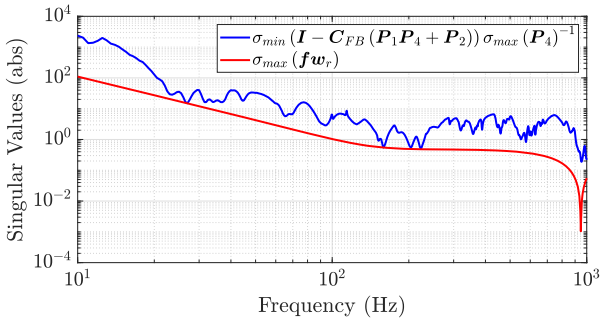

FIGURE 4 Evaluation of the linear stability condition in Equation (97).

with w_i° the maximal value that satisfies

$$\sigma_{\max}\left(f_{r,i}w_i^\circ\right) \leq \sigma_{\min}(\boldsymbol{I}-\boldsymbol{C}_{\mathrm{FB}}(\boldsymbol{P}_1\boldsymbol{P}_4+\boldsymbol{P}_2))\sigma_{\max}(\boldsymbol{P}_4)^{-1}, \tag{99}$$

and where $f_{r,i}$ is the ith element of $\boldsymbol{f}_r$. Combined with the measurement of $\boldsymbol{P}_4(q)$, this motivates the extension of (88) with a notch filter as

$$\boldsymbol{f}(s) = \begin{bmatrix} 1 & \beta H(s) & \beta^2 H^2(s) \end{bmatrix} \frac{s^2+0.01\omega_n s+\omega_n^2}{s^2+\omega_n s+\omega_n^2}, \tag{100}$$

to expand $\mathcal{W}_l$. This yields a factor 2 improvement in the element of $\boldsymbol{w}_r$ that relates to the proportional generator. This parametrization can be seen in of Figure 4. This comes however with a performance penalty.

The SPR condition of (61) can be checked by the eigenvalue decomposition of $\mathcal{P}_{\mathrm{mr}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)$ in (70) as

$$\boldsymbol{\mathcal{P}}\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\boldsymbol{V}_p\left(e^{j\omega},\overline{\boldsymbol{w}}\right) = \boldsymbol{V}_p\left(e^{j\omega},\overline{\boldsymbol{w}}\right)\boldsymbol{D}_p\left(e^{j\omega},\overline{\boldsymbol{w}}\right), \tag{101}$$

and is shown in Figure 5, where the red dots indicate frequencies and directions where the SPR condition is not satisfied. It can be concluded that the SPR condition is not satisfied.

The diagonal of $\boldsymbol{D}(e^{j\omega})$ of (83) is evaluated on a discrete frequency grid with measurement data and is shown in Figure 6 with $N(q)=1$. The fit of the model $\hat{\boldsymbol{S}}_2(q)$ is good on the interval [2, 800] Hz. This can be seen by the positive real parts in $\boldsymbol{D}(e^{j\omega})$ inside this interval. This suggests that the bandpass of $N(q)$ should not exceed this frequency range. This is indeed the case for (92), and explains in retrospect the band-pass filter that was obtained by an engineering approach.

The integral of Equation (69) is discretized for numerical evaluation using (92). The eigenvalues of Equation (68) are shown in Figure 7 for $\boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))=\boldsymbol{0}$. It shows that there are no negative eigenvalues. Since $\boldsymbol{Q}(\boldsymbol{w}(k))$ depends on $\boldsymbol{w}(k)$ for systems with $\boldsymbol{P}_4(q)\neq\boldsymbol{0}$, it needs to be evaluated for a grid of $\boldsymbol{w}(k)\in\mathcal{W}_l$. This shows that some additional dynamics are introduced due to $\boldsymbol{C}_{\mathrm{FF}}(q,\boldsymbol{w}(k))\neq\boldsymbol{0}$. This is most prominent at 220 Hz. This is the resonance frequency associated with the rigid body motion of the floor. This gridding can be done numerically, since $\boldsymbol{Q}(\boldsymbol{w}(k))$ can be constructed according to

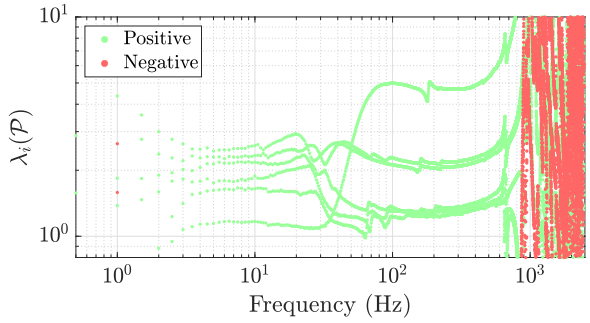

FIGURE 5 SPR condition of (61) for the experimental FRF data. Positive points correspond with eigenvalues with positive real parts, negative points are defined similar but with negative real parts.

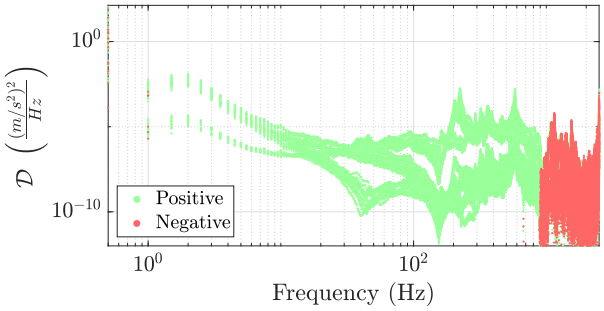

FIGURE 6 Evaluation of (84) for the experimental data. Positive points correspond with elements of the diagonal of $\boldsymbol{D}(e^{j\omega})$ with positive real parts, negative points are defined similar but with negative real parts. The negative points are displayed in front of the positive points.

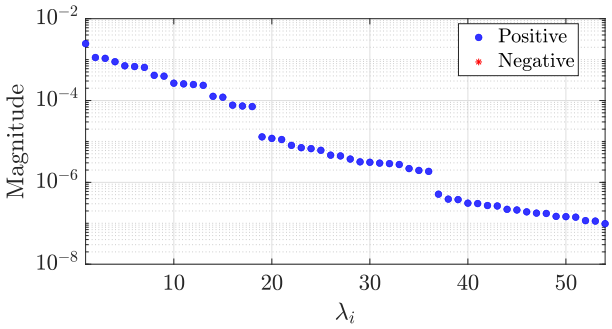

FIGURE 7 Experimentally determined eigenvalues of $\boldsymbol{Q}(\boldsymbol{0})$.

(69) using the measurements of $\boldsymbol{P}(q)$, the filter $\boldsymbol{N}(q)$, the model $\hat{\boldsymbol{S}}_2(q)$, the feedback controller $\boldsymbol{C}_{\text{FB}}(q)$, the basis $\boldsymbol{f}(q)$, the relation (7) and the measured spectrum of $\tilde{\boldsymbol{a}}_s(k)$. This last spectrum can be measured if $\boldsymbol{C}_{\text{FB}}(q) = \boldsymbol{C}_{\text{FF}}(q) = \boldsymbol{0}$. The elements of $\boldsymbol{w}_{r,\max}$ are reduced w.r.t. (97) to adhere to the actuator power and peak amplitude limits. A linear grid is chosen between (the reduced) $\boldsymbol{w}_{r,\max}$. This additional dynamics decreases the eigenvalues, but for the evaluated grid points, the PWSPR condition remains satisfied. This grid evaluation of $\boldsymbol{Q}(\boldsymbol{w}(k))$ gives a good indication of the convergence properties, but is no conclusive proof.

Based on the results presented above, the AF should be stable. This is validated by implementing the RLS algorithm of (A3) to the AVIS system with $\lambda = e^{-1/10t_s}$. The adaptation is started at $t = 0$ and parameter convergence is recorded for 100 s. The parameters converge to a stationary point. This is shown in Figures 8 and 9. These figures contain the same data, but in Figure 8 the parameter trajectories are normalized on the vertices of $\mathcal{W}_l$, and in Figure 9 the trajectories of the parameter errors $\tilde{\boldsymbol{w}}_n(t)$ normalized to the final value are displayed. The parameters are normalized to the vertices of $\mathcal{W}_l$ to indicate the distance to linear instability due to the input-disturbance interaction, and since no complete quantification of

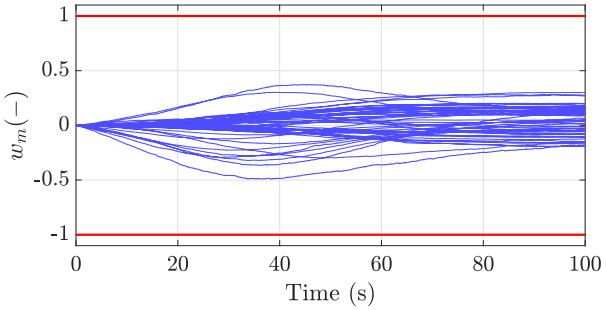

FIGURE 8 $\tilde{w}_{m,i}(t) = \frac{w_i(t)}{w_{r,i}}$ are the normalized trajectories of $\boldsymbol{w}(t)$, where $w_{r,i}$ is the element of w_r which corresponds to w_i, such that the stable region is between $[-1,\ 1]$.

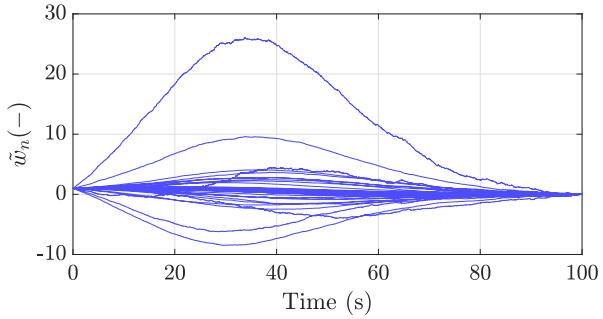

FIGURE 9 Normalized trajectories of the $\tilde{\boldsymbol{w}}(t)$, $\tilde{w}_{n,i}(t) = \frac{w_i(t) - w_i(100)}{w_i(0) - w_i(100)}$.

$\mathcal{W}_a$ is available. It can be seen that the convergence is not monotonic, and that the overshoot of the weights appears to be significant. This highlights the relevance of checking the condition for a set of weights larger than the hypercube spanned by the nominal weights. Here only the parameter adaptation is considered, the performance of the AVIS is similar to Beijen et al.[6] and Hakvoort and Beijen.[35]

9 | DISCUSSION

It is shown in the previous section that the PWSPR condition significantly reduces the conservatism in the convergence condition for FeAF. This is validated for an experimental setup. The evaluation of the PWSPR condition requires a frequency domain representation of $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}})$, a PSD of $\tilde{\boldsymbol{a}}_s(k)$ and the filters required for the control. The filters are up to the designer. For the frequency domain representation of $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}})$ a FRF measurement can be used and obviates the need for accurately fitting a parametric model. The sensors and actuators required for the FRF measurement are available as part of the FeAF structure. The PSD of $\tilde{\boldsymbol{a}}_s$ is typically available for the designer, since either measurements or models of the PSD of $\tilde{\boldsymbol{a}}_s(k)$ can be used. For example, the PSD of $\tilde{\boldsymbol{a}}_s(k)$ corresponds with the PSD of the noise corrupted floor acceleration. This PSD is easily measured by disabling the feedback and feedforward controllers. In the absence of measurements, standards can be used.[37] Therefore the PWSPR provides a practically applicable condition for the convergence of FeAF.

It has to be noted that since the PWSPR condition depends on the PSD of $\tilde{\boldsymbol{a}}_s$, the convergence can be lost over time if $\boldsymbol{a}_s + \boldsymbol{n}_0$ is not wide-sense stationary. It is therefore recommended to design $N(q)$ with sufficient margin, that is,

$$\boldsymbol{Q}(\overline{\boldsymbol{w}}) \succ \varepsilon \boldsymbol{I} \ \forall\ \overline{\boldsymbol{w}} \in \mathcal{W}_b, \tag{102}$$

with some suitable $\varepsilon > 0$. Furthermore, it is up to the designer to guarantee sufficiently accurate FRF measurements of $\boldsymbol{S}_2(q, \overline{\boldsymbol{w}})$. This is also the case for the SPR condition.

The Lyapunov function of (52) provides only information about convergence of the expected update step, therefore, peaking of the parameters can still occur. This effect becomes more prominent when the eigenvalues of $\boldsymbol{Q}(\overline{\boldsymbol{w}})$ in (68) tend

to zero, and can cause $\boldsymbol{w}(k) \notin \mathcal{W}$. This can be avoided if a projection operator is used. Even if the projection operator is used, this will degrade the performance of the feedforward controller until the parameters converge to a stationary point again. This further motivates a sufficiently large ε in (102).

There is some conservatism left in this convergence condition. This originates from the specific choice of Lyapunov function in (52). It is noted by Fraanje et al.[12] that the choice of this Lyapunov function does not capture convergence with an initial increase in error. This is called critical behavior and can saturate the sensors and actuators and destabilize the system due to the associated unmodelled non-linearities. It can furthermore negatively affect the lifetime of the mechanism due to large excitations. This conservatism might therefore actually be beneficial from a design perspective.

The evaluation of the small gain condition in (97) with the measurement data from the experimental setup motivates the addition of a notch filter in the feedback and feedforward path in (87) and (100). This improves the stable region with respect to the input-disturbance interaction. It is observed in experiments that without this filter, the convergence of the system varies between runs, which can be explained by the large overshoot of the parameters related to the proportional generating transfer function in (88) and the reduced stability range. This overshoot is depending on the specific realization of the floor acceleration and can therefore vary between runs.

The obtained results are derived under the condition that $\gamma(k)$ is sufficiently small, such that the slow parameter variation assumption is satisfied. At the moment, only heuristic upperbounds for $\gamma(k)$ are available, and it is not clear how to quantify the exact upperbound. Also, the exact quantification of the set $\mathcal{W}_a$ is not clear. These are topics for further research.

10 | CONCLUSION

In this article, the stability and convergence properties of FeAF are considered under the influence of actuator reference interaction and model inaccuracies. It is shown that this input-disturbance interaction closes a nonlinear feedback loop, which makes the stability dependent on the values of the feedforward parameters. Using the slow parameter variation assumption, this reduces to a linear stability condition for a frozen set of parameters, and can be evaluated in the frequency domain using a small gain stability condition.

Second, it is shown that the conventional convergence conditions for FeAF can be relaxed by taking into account the PSD of the input. This yields a condition that is usable for systems with significant unmodelled dynamics. The convergence condition is derived under the slow parameter variation assumption, which is also used by Wang and Ren.[11] It can be verified in the frequency domain and is applicable for MIMO systems. The condition has been verified on measurement data from an experimental AVIS. This experimental setup is also used in previous research and the presented conditions explain the choice of filtering and the obtained results of Beijen et al.[6] and Hakvoort and Beijen[35] in retrospect. Guidelines are given for the design of the noise shaping filter using the relaxed convergence condition.

ACKNOWLEDGMENTS

The authors would like to thank MI partners, Veldhoven, the Netherlands and PM, Dedemsvaart, the Netherlands, for their financial support.

FUNDING INFORMATION

This work was supported in part by MI-partners B.V. and PM B.V.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

ENDNOTES

*For AVIS, $\boldsymbol{P}_4(q)$ corresponds to the actuator floor interaction. In contrast, for ANC the actuators typically only influence the reference sensors. This can be easily adapted in this formulation by setting $\boldsymbol{P}_1(q)\boldsymbol{P}_4(q) = \boldsymbol{0}$. When $\boldsymbol{P}_1(q)\boldsymbol{P}_4(q) = \boldsymbol{0}$, echo cancellation can be used to prevent instability.[5] For reference based adaptive feedforward $\boldsymbol{P}_4 = \boldsymbol{0}$.

†Note that the measurement of the FRF of $\boldsymbol{P}(q)$ is typically easier than measuring $\boldsymbol{P}_1(q)$ and $\boldsymbol{P}_2(q)$ separately.

‡Definition 4 in Kushner.[32]

§This theorem is a combination of Theorem 12 and Corollary 3-1 of Kushner.[32]

ORCID

Sil T. Spanjer https://orcid.org/0009-0004-7787-6745
Hakan Köroğlu https://orcid.org/0000-0001-5285-7784
Wouter B. J. Hakvoort https://orcid.org/0000-0001-5919-4442

REFERENCES

1. Butler H. Adaptive feedforward for a wafer stage in a lithographic tool. *IEEE Trans Control Syst Technol*. 2012;21(3):875-881.
2. Reinders J, Hunnekens B, Heck F, Oomen T, van de Wouw N. Adaptive control for mechanical ventilation for improved pressure support. *IEEE Trans Control Syst Technol*. 2020;29(1):180-193.
3. Zhao S, Tan K. Adaptive feedforward compensation of force ripples in linear motors. *Control Eng Pract*. 2005;13(9):1081-1092.
4. Wang N, Johnson KE, Wright AD. FX-RLS-based feedforward control for LIDAR-enabled wind turbine load mitigation. *IEEE Trans Control Syst Technol*. 2011;20(5):1212-1222.
5. Lu L, Yin KL, de Lamare RC, et al. A survey on active noise control in the past decade—part I: linear systems. *Signal Process*. 2021;183:108039.
6. Beijen M, Heertjes M, Van Dijk J, Hakvoort W. Self-tuning MIMO disturbance feedforward control for active hard-mounted vibration isolators. *Control Eng Pract*. 2018;72:90-103.
7. Åström KJ, Murray RM. *Feedback Systems: an Introduction for Scientists and Engineers*. Princeton University Press; 2021.
8. Landau ID, Alma M, Airimitoaie TB. Adaptive feedforward compensation algorithms for active vibration control with mechanical coupling. *Automatica*. 2011;47(10):2185-2196.
9. Morgan D. An analysis of multiple correlation cancellation loops with a filter in the auxiliary path. *IEEE Trans Acoust Speech Signal Process*. 1980;28(4):454-467.
10. Wan EA. Adjoint LMS: an efficient alternative to the filtered-X LMS and multiple error LMS algorithms. *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. IEEE; 1996:1842-1845.
11. Wang AK, Ren W. Convergence analysis of the multi-variable filtered-X LMS algorithm with application to active noise control. *IEEE Trans Signal Process*. 1999;47(4):1166-1169.
12. Fraanje R, Verhaegen M, Elliott SJ. Robustness of the filtered-X LMS algorithm—part I: necessary conditions for convergence and the asymptotic pseudospectrum of Toeplitz matrices. *IEEE Trans Signal Process*. 2007;55(8):4029-4037.
13. Landau ID, Airimiţoaie TB, Alma M. IIR Youla–Kucera parameterized adaptive feedforward compensators for active vibration control with mechanical coupling. *IEEE Trans Control Syst Technol*. 2012;21(3):765-779.
14. Scherer C, Weiland S. Linear matrix inequalities in control. Lecture notes. Dutch Institute for Systems and Control; 2021: 3(2).
15. Tao G. *Adaptive Control Design and Analysis*. Vol 37. John Wiley & Sons; 2003.
16. van der Poel GW. *An Exploration of Active Hard Mount Vibration Isolation for Precision Equipment*. Ph.D. Thesis. University of Twente; 2010.
17. Collette C, Matichard F. Sensor fusion methods for high performance active vibration isolation systems. *J Sound Vib*. 2015;342:1-21.
18. Heuberger PSC, van den Hof PMJ, Wahlberg B. *Modelling and Identification with Ratioal Orthogonal Basis Functions*. Springer; 2005.
19. van den Hof P, Wahlberg B, Heuberger P, Ninness B, Bokor J, Silva TO. Modelling and identification with rational orthogonal basis functions. *IFAC Proc Vol*. 2000;33(15):445-455.
20. van Donkelaar ET. *Improvement of Efficiency in Identification and Model Predictive Control of Industrial Processes - A Flexible Linear Parametrization Approach*. Ph.D. Thesis. Delft University of Technology; 2000.
21. Beijen MA, Heertjes MF, Butler H, Steinbuch M. Disturbance feedforward control for active vibration isolation systems with internal isolator dynamics. *J Sound Vib*. 2018;436:220-235.
22. Skogestad S, Postlethwaite I. *Multivariable Feedback Control: Analysis and Design*. john Wiley & Sons; 2005.
23. Åström KJ, Wittenmark B. *Computer-Controlled Systems: Theory and Design*. Courier Corporation; 2013.
24. Doyle J. Robustness of multiloop linear feedback systems. *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*. IEEE; 1979:12-18.
25. Shamma JS. An overview of LPV systems. In: Mohammadpour J, Scherer C, eds. *Control of Linear Parameter Varying Systems with Applications*. Springer; 2012:3-26.
26. van Zundert J, Oomen T. On inversion-based approaches for feedforward and ILC. *Mechatronics*. 2018;50:282-291.
27. Berkhoff AP, Nijsse G. A rapidly converging filtered-error algorithm for multichannel active noise control. *Int J Adapt Control Signal Process*. 2007;21(7):556-569.
28. Elliott S. *Signal Processing for Active Control*. Elsevier; 2000.
29. Sayed AH. *Fundamentals of Adaptive Filtering*. John Wiley & Sons; 2003.
30. Lavretsky E, Wise KA. *Robust Adaptive Control*. Springer; 2012.
31. Ljung L. Analysis of recursive stochastic algorithms. *IEEE Trans Automat Contr*. 1977;22(4):551-575.
32. Kushner HJ. *Stochastic Stability and Control*. Academic Press; 1967.
33. Hayes MH. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons; 1996.
34. Bai EW, Sastry SS. Persistency of excitation, sufficient richness and parameter convergence in discrete time adaptive control. *Syst Control Lett*. 1985;6(3):153-163.

35. Hakvoort WB, Beijen MA. Filtered-error RLS for self-tuning disturbance feedforward control with application to a multi-axis vibration isolator. *Mechatronics*. 2023;89:102934.
36. Tjepkema D. *Active Hard Mount Vibration Isolation for Precision Equipment*. Ph.D. Thesis. University of Twente; 2012.
37. Gordon CG. Generic criteria for vibration-sensitive equipment. *Vibration Control in Microelectronics, Optics, and Metrology*. SPIE; 1992:71-85.
38. van Ophem S, Berkhoff AP. Multi-channel Kalman filters for active noise control. *J Acoust Soc Am*. 2013;133(4):2105-2115.
39. Ljung L. *System Identification: Theory for the User*. 2nd ed. Prentice-Hall; 1999.
40. Cao L, Schwartz HM. Analysis of the Kalman filter based estimation algorithm: an orthogonal decomposition approach. *Automatica*. 2004;40(1):5-19.
41. Lavergne P. A Cauchy-Schwarz inequality for expectation of matrices. Technical report. Simon Fraser University; 2008.
42. Irgens F. *Tensor Analysis*. Springer; 2019.
43. Einstein A. Die Grundlage der allgemeinen Relativitätstheorie. *Ann Phys*. 1916;354(7):769-822.

How to cite this article: Spanjer ST, Köroğlu H, Hakvoort WBJ. Frequency domain stability and relaxed convergence conditions for filtered error adaptive feedforward. *Int J Adapt Control Signal Process*. 2024;1-25. doi: 10.1002/acs.3826

APPENDIX A. DEFINITIONS OF Γ

Equation (32) can be interpreted as a regularized Newton recursion if[29]

$$\boldsymbol{\Gamma}(k) = \gamma(k)\overline{\boldsymbol{\Gamma}}(k) = \gamma(k)(\boldsymbol{E}(k) + \boldsymbol{R})^{-1}, \tag{A1}$$

where $\boldsymbol{R} = \mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\tilde{\boldsymbol{\Psi}}_N(k)\right] \succeq \mathbf{0}$. $\boldsymbol{E}(k) = \boldsymbol{E}^T(k) \succ \mathbf{0}$ is a regularization matrix such that $\overline{\boldsymbol{\Gamma}} = \overline{\boldsymbol{\Gamma}}^T \succ \mathbf{0}$ and $\gamma(k) \geq 0$ is a scalar gain. In general, the statistical information of $\boldsymbol{R}$ is a priori not available and hence are usually implemented as approximations. Let $\hat{\boldsymbol{R}}_r(k) = \hat{\boldsymbol{R}}_r(k)^T = \left(\boldsymbol{E}(k) + \hat{\boldsymbol{R}}(k)\right) \succ \mathbf{0}$. For the approximation $\boldsymbol{\Gamma}(k) = \gamma(k)\hat{\boldsymbol{R}}_r^{-1}(k)$ there exist several popular choices such as normalized least mean squares (ϵ-LMS), recursive least squares (RLS) and Kalman filters, see Beijen et al.,[6] Hakvoort and Beijen[35] and van Ophem and Berkhoff[38] respectively. Here only the ϵ-LMS and RLS are discussed. These methods all use the instantaneous approximation $\mathbb{E}\left[\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k))\right] \rightarrow \tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{e}(k, \boldsymbol{w}(k))$ in (32).[29] The ϵ-LMS algorithm uses

$$\begin{aligned}
\overline{\boldsymbol{\Gamma}}(k) &= \left(\epsilon + \hat{\boldsymbol{R}}(k)\right)^{-1}\boldsymbol{I}, \\
\hat{\boldsymbol{R}}(k) &= \|\tilde{\boldsymbol{\Psi}}_N(k)\|_2^2, \\
\gamma(k) &= \overline{\gamma},
\end{aligned} \tag{A2}$$

with a static $\epsilon \in \mathbb{R}^1$. The RLS algorithm uses

$$\begin{aligned}
\overline{\boldsymbol{\Gamma}}(k) = \hat{\boldsymbol{R}}_r^{-1}(k) &= \lambda^{-1}\left[\overline{\boldsymbol{\Gamma}}(k-1) - \overline{\boldsymbol{\Gamma}}(k-1)\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{L}(k)^{-1}\tilde{\boldsymbol{\Psi}}_N(k)\overline{\boldsymbol{\Gamma}}(k-1)\right], \\
\boldsymbol{L}(k) &= \left[\lambda\boldsymbol{I} + \tilde{\boldsymbol{\Psi}}_N(k)\overline{\boldsymbol{\Gamma}}(k-1)\tilde{\boldsymbol{\Psi}}_N^T(k)\right], \\
\gamma &= 1,
\end{aligned} \tag{A3}$$

with $\lambda \in (0, 1]$ being the exponential forgetting factor and $\hat{\boldsymbol{R}}(0)$ chosen appropriately.[35] This yields an exponentially decaying $\boldsymbol{E}(k) = \frac{\lambda^{k+1}\hat{\boldsymbol{R}}(0)}{k+1}$.

The ϵ-LMS and RLS are often found in the literature in the previous form. To comply to Assumption 3, they are rewritten to a more convenient form. Equation (A2) is rewritten as

$$\begin{aligned}
\overline{\boldsymbol{\Gamma}} &= \hat{\boldsymbol{R}} = \boldsymbol{I}, \\
\gamma(k) &= \frac{\overline{\gamma}}{\epsilon + \|\tilde{\boldsymbol{\Psi}}_N(k)\|_2^2},
\end{aligned} \tag{A4}$$

and the terms of (A3) are rewritten to the normalized gain variant[39]

$$\begin{aligned}\overline{\boldsymbol{\Gamma}}(k) &= \hat{\boldsymbol{R}}_r^{-1}(k) = \frac{1}{1-\gamma}\left[\overline{\boldsymbol{\Gamma}}(k-1) - \overline{\boldsymbol{\Gamma}}(k-1)\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{L}(k)^{-1}\tilde{\boldsymbol{\Psi}}_N(k)\overline{\boldsymbol{\Gamma}}(k-1)\right],\\ \boldsymbol{L}(k) &= \left[\frac{1-\gamma}{\gamma}\boldsymbol{I} + \tilde{\boldsymbol{\Psi}}_N(k)\overline{\boldsymbol{\Gamma}}(k-1)\tilde{\boldsymbol{\Psi}}_N^T(k)\right],\\ \gamma &= 1-\lambda.\end{aligned} \tag{A5}$$

In this form, $\hat{\boldsymbol{R}}_r^{-1}(k)$ is the exponentially weighted sample sum, scaled by γ. This means that in this form, $\overline{\boldsymbol{\Gamma}}(k)$ varies slowly over time if λ approaches 1 since it is assumed that the input signal of $\tilde{\boldsymbol{\Psi}}_N(k)$ is wide sense stationary. It is noted that (32) and (A1) are the basis for a larger class of adaptive algorithms, where the implementations of ϵ-LMS and RLS serve as examples. The multivariable versions of the ϵ-LMS and RLS are given, however, due to the block diagonal nature of $\tilde{\boldsymbol{\Psi}}$ it is possible to reduce the size of the estimation problem.[35]

The first item of Assumption 3 holds for both ϵ-LMS and RLS due to the regularization that is applied. The second item of Assumption 3 is satisfied if there is persistent excitation.[40] For AVIS this is the case due to the nature of $\boldsymbol{a}_s(k)$ and $\boldsymbol{n}_0(k)$. For reference based AF, this condition needs to be guaranteed by the designer. The third item holds for ϵ-LMS by definition, and for RLS if $\lambda \to 1$. The fourth item holds for ϵ-LMS with $\delta_v = 0$. Item four can be proven for RLS by

$$\begin{aligned}\lim_{\gamma(k)\to 0}\overline{\boldsymbol{\Gamma}}(k) &= \lim_{\gamma(k)\to 0}\frac{1}{1-\gamma}\left[\overline{\boldsymbol{\Gamma}}(k-1) - \overline{\boldsymbol{\Gamma}}(k-1)\tilde{\boldsymbol{\Psi}}_N^T(k)\boldsymbol{L}(k)^{-1}\tilde{\boldsymbol{\Psi}}_N(k)\overline{\boldsymbol{\Gamma}}(k-1)\right],\\ &= \lim_{\gamma(k)\to 0}\frac{1}{1-\gamma}\left[\overline{\boldsymbol{\Gamma}}(k-1) - \underbrace{\overline{\boldsymbol{\Gamma}}(k-1)\tilde{\boldsymbol{\Psi}}_N^T(k)\gamma(k)\tilde{\boldsymbol{\Psi}}_N(k)\overline{\boldsymbol{\Gamma}}(k-1)}_{\boldsymbol{B}(k)}\right].\end{aligned} \tag{A6}$$

The value of c_v is determined by $\boldsymbol{B}(k)$, where $\boldsymbol{B}(k)$ contains $\overline{\boldsymbol{\Gamma}}(k-1)$, $\tilde{\boldsymbol{\Psi}}_N(k)$ and $\gamma(k)$. The matrix $\tilde{\boldsymbol{\Psi}}_N(k)$ does not scale with $\gamma(k)$. The $\overline{\boldsymbol{\Gamma}}(k-1)$ is the inverse of the exponentially weighted sample average[29,39]

$$\overline{\boldsymbol{\Gamma}}^{-1}(k-1) = \frac{1}{\gamma(N+1)}\sum_{i=0}^{N}\lambda^{N-j}\tilde{\boldsymbol{\Psi}}_N^T(k-1)\tilde{\boldsymbol{\Psi}}_N(k-1). \tag{A7}$$

Thereby $\boldsymbol{\Gamma}(k-1)$ scales with γ. The matrix $\boldsymbol{B}(k)$ therefore scales with γ^3 and justifies the fourth item of Assumption 3.

APPENDIX B. BOUNDEDNESS OF THE EXPECTATION OF PRODUCTS

Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{Y} \in \mathbb{R}^{p\times n}$ be two possibly dependent jointly distributed random variables. The extension of the Cauchy-Schwarz inequality for the expectation of random matrices is[41]

$$\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right] - \mathbb{E}\left[\boldsymbol{X}\boldsymbol{Y}^T\right]\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^T\right]^{-1}\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{X}^T\right] \succcurlyeq 0. \tag{B1}$$

Therefore, if $\|\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]\|_\infty \leq \infty$ and $\|\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^T\right]\|_\infty \leq \infty$, then $\|\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{X}^T\right]\|_\infty = \|\mathbb{E}\left[\boldsymbol{X}\boldsymbol{Y}^T\right]^T\|_\infty \leq \infty$. If $\|\boldsymbol{X}\|_{\text{pow}} \leq \infty$ and $\|\boldsymbol{Y}\|_{\text{pow}} \leq \infty$, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ have finite second order moments and it holds that $\|\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]\|_\infty \leq \infty$ and $\|\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^T\right]\|_\infty \leq \infty$. This can be extended for the expectation of the product of an arbitrary number of random matrices.

APPENDIX C. INDEX NOTATION

The index notation is used to write tensors and tensor operations. The content of this appendix is based on Irgens.[42] Let $\boldsymbol{b} \in \mathbb{C}^3$ be a first order tensor, and $\boldsymbol{A} \in \mathbb{C}^{2\times 3}$ a second order tensor

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1\\ b_2\\ b_3\end{bmatrix}. \tag{C1}$$

This is in index notation

$$\boldsymbol{A} = A_{\mathrm{mp}}, \quad \boldsymbol{b} = b_p, \tag{C2}$$

and can be generalized to higher order tensors as $\boldsymbol{C} \in \mathbb{C}^{n_1 \times n_2 \times \cdots \times n_m}$. The inner product is expressed as

$$\boldsymbol{d} = \boldsymbol{A}\boldsymbol{b} = \sum_{p=1}^{3} A_{\mathrm{mp}} b_p = A_{\mathrm{mp}} b_p, \tag{C3}$$

where the last equality is by the Einstein summation convention. This convention implies summation over repeated indices.[43] The transpose of a first order tensors is

$$\boldsymbol{A}^T = A_{\mathrm{pm}}. \tag{C4}$$

The Kronecker delta is

$$\delta_{\mathrm{mp}} = \begin{cases} 1 & m = p, \\ 0 & m \neq p. \end{cases} \tag{C5}$$

The trace operation can be expressed in terms of a Kronecker delta as

$$Tr(A) = A_{\mathrm{pp}} = \delta_{\mathrm{pm}} A_{\mathrm{mp}}. \tag{C6}$$

The contraction tensor $e_{\mathrm{mpq}} \in \mathbb{R}^{n_1 \times n_2 \times n_1 n_2}$ that maps

$$L_{\mathrm{mp}} = e_{\mathrm{mpq}} l_q, \tag{C7}$$

with $L_{\mathrm{mp}} \in \mathbb{C}^{n_1 \times n_2}$ and $l_q \in \mathbb{C}^{n_1 n_2}$ is defined as

$$e_{\mathrm{mpq}} = \begin{cases} 1 & m + n_1(p-1) = q, \\ 0 & \text{else}. \end{cases} \tag{C8}$$