Estimation of Persistently Exciting Subspaces for Robust Parameter Adaptation

Erick Mejia Uzeda*† and Mireille E. Broucke*‡

***Abstract*— Recently we proposed the μ-modification for robust parameter adaptation, premised on the observation that only the parameter dynamics along the subspace with no persistent excitation must be rendered robust. Robustness thereby reduces to a problem of subspace estimation. This paper proposes a new subspace estimator that recovers the non-PE subspace for a large class of regressors. This is achieved through a characterization of persistently exciting subspaces and the use of Principal Component Analysis. Correctness of the design is proved using matrix perturbation theory, while an averaging analysis demonstrates that the design is best employed as a slow process. We develop a general error model capturing those commonly appearing in adaptive control, and we prove that the μ-modification provides a modular robust design, without compromising error regulation.**

***Index Terms*— Adaptive Control, Persistent Excitation, Robustness, Subspace Estimation**

I. Introduction

This paper proposes a new method of robust parameter adaptation. To place ideas in context, consider a static error model $e(t) = w^\intercal(t)\hat{\psi} - y(t) = w^\intercal(t)(\hat{\psi} - \psi)$, where $y(t) = w^\intercal(t)\psi \in \mathbb{R}$ is a measurement, $\psi \in \mathbb{R}^q$ is a vector of unknown parameters, $w(t) \in \mathbb{R}^q$ is a known regressor, and $\hat{\psi}(t)$ is an estimate of ψ. To estimate ψ, the simplest method is to consider the instantaneous cost function

$$J(\hat{\psi}) := \frac{1}{2}e^2(t) = \frac{1}{2}(w^\intercal(t)(\hat{\psi} - \psi))^2 . \tag{1}$$

Computing the gradient, one has $\nabla J(\hat{\psi}) = ew(t)$ so that the *standard gradient algorithm* is

$$\dot{\hat{\psi}} = -\gamma \nabla J(\hat{\psi}) = -\gamma e w(t)$$

where $\gamma > 0$. It has been noted that the instantaneous cost only penalizes errors along w at time t rather than its full range of excitation. For example, if w is persistently exciting (PE) then the full error $\hat{\psi} - \psi$ should be penalized rather than just the component along $w(t)$. To this end, researchers have considered an integral cost function

$$\begin{aligned} J(\hat{\psi}) &:= \frac{\varepsilon}{2}\int_{t_0}^{t} \mathrm{e}^{-\varepsilon(t-\tau)} e^2(\tau)\, d\tau \\ &= \frac{1}{2}(\hat{\psi} - \psi)^\intercal \Sigma(t)(\hat{\psi} - \psi) , \end{aligned} \tag{2}$$

where $\Sigma(t) := \varepsilon \int_{t_0}^{t} \mathrm{e}^{-\varepsilon(t-\tau)} w(\tau) w^\intercal(\tau)\, d\tau$. If it is known that $\Sigma(t)$ is invertible (i.e., $w(t)$ is PE), one obtains the *non-recursive least squares algorithm*: $\hat{\psi}(t) = \Sigma(t)^{-1}Q(t)$, where Σ and Q are generated by

$$\dot{\Sigma} = -\varepsilon\Sigma + \varepsilon w(t)w^\intercal(t) , \quad \Sigma(t_0) = 0 \tag{3a}$$

$$\dot{Q} = -\varepsilon Q + \varepsilon w(t)y(t) , \quad Q(t_0) = 0 . \tag{3b}$$

A recursive solution is provided by the *Kreisselmeier integral algorithm*

$$\dot{\hat{\psi}} = -\gamma\nabla J(\hat{\psi}) = -\gamma(\Sigma\hat{\psi} - Q) ,$$

where Σ and Q are again updated using (3). These algorithms and many more lack robustness and suffer computational problems when w is not PE. In adaptive control, *modifications* of parameter adaptation laws have been suggested to achieve robustness [1]; they generally trade off error regulation.

In [2] we made the observation that only the parameter adaptation dynamics along the subspace with no persistent excitation needs to be rendered robust. We proposed the μ-modification, a modular technique that provides robust adaptation without sacrificing error regulation. To illustrate, suppose it is known that w lies in a subspace $\mathcal{W} \subsetneq \mathbb{R}^q$. Let $W \in \mathbb{R}^{q\times q_{pe}}$ have orthonormal columns such that $\mathcal{W} = \operatorname{Im}(W)$. We can write $w = Ww_{pe}$ for some $w_{pe}(t) \in \mathbb{R}^{q_{pe}}$ (w_{pe} is not known to be PE at this stage). Now the instantaneous cost (1) becomes

$$J(\hat{\psi}) = \frac{1}{2}(w_{pe}^\intercal(t)(\hat{\psi}_{pe} - W^\intercal\psi))^2 = J(\hat{\psi}_{pe}) ,$$

where $\hat{\psi}_{pe} := W^\intercal\hat{\psi}$. The integral cost (2) can similarly be shown to reduce to a function of $\hat{\psi}_{pe}$. It is now clear that the root problem is that only the component $\hat{\psi}_{pe}$ is penalized by the cost function.

The μ-modification regularizes the cost function through a penalty on the unexcited components of ψ. Define $W_\perp$ such that $\begin{bmatrix} W & W_\perp \end{bmatrix} \in \mathbb{R}^{q\times q}$ is an orthogonal matrix. Then $\hat{\psi}$ splits as $\hat{\psi} = WW^\intercal\hat{\psi} + W_\perp W_\perp^\intercal\hat{\psi} =: W\hat{\psi}_{pe} + W_\perp\hat{\psi}_\perp$. Consider the regularized instantaneous cost

$$J_\mu(\hat{\psi}) := J(\hat{\psi}_{pe}) + \frac{\mu}{2\gamma}\|\hat{\psi}_\perp\|^2 = \frac{1}{2}e^2(t) + \frac{\mu}{2\gamma}\|\hat{\psi}_\perp\|^2 .$$

Taking the gradient, we arrive at the μ-modification

$$\dot{\hat{\psi}} = -\gamma\nabla J_\mu(\hat{\psi}) = -\gamma e w(t) - \mu W_\perp W_\perp^\intercal\hat{\psi} .$$

It consists of a leakage term that is only applied along the *non-PE subspace* $\mathcal{W}^\perp = \operatorname{Im}(W_\perp)$, which is appended to

*Supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Electrical and Computer Engineering, University of Toronto, Toronto ON Canada.
†Email: erick.mejiauzeda@mail.utoronto.ca
‡Email: broucke@control.utoronto.ca

the standard gradient algorithm. As a final step, one requires an estimate of the projection map $W_\perp W_\perp^\intercal$ onto the non-PE subspace. One method is to leverage the fact that integral cost functions are meant to capture the full range of excitation of the regressor w. Noting that $\Sigma(t) = W\Lambda_{pe}(t)W^\intercal$ where $\Lambda_{pe}(t) := \varepsilon\int_{t_0}^{t} \mathrm{e}^{-\varepsilon(t-\tau)} w_{pe}(\tau)w_{pe}^\intercal(\tau)\, d\tau$, if $\Lambda_{pe}(t)$ is invertible then $\mathrm{Im}(\Sigma(t)) = \mathrm{Im}(W)$. Therefore, we can use the Moore-Penrose pseudoinverse (denoted $\dagger$) to obtain

$$WW^\intercal = \Sigma(t)\Sigma^\dagger(t)\,.$$

Under the assumption that w_{pe} is PE, this holds eventually and moreover $W_\perp W_\perp^\intercal = I - \Sigma(t)\Sigma^\dagger(t)$. As such, one variant of the μ-modification with a suitable subspace estimator is

$$\dot{\Sigma} = -\varepsilon\Sigma + \varepsilon w(t)w^\intercal(t)\,, \quad \Sigma(t_0) = 0 \tag{4a}$$

$$\Omega = I - \Sigma\Sigma^\dagger \tag{4b}$$

$$\dot{\hat{\psi}} = -\gamma e w(t) - \mu\Omega\Omega^\intercal\hat{\psi}\,, \tag{4c}$$

where we note that $\Omega\Omega^\intercal = W_\perp W_\perp^\intercal W_\perp W_\perp^\intercal = W_\perp W_\perp^\intercal$ eventually. In the remainder of the paper, we will derive an equivalent design for estimation of the non-PE subspace $\mathcal{W}^\perp$ for use in the μ-modification, but starting from first principles concerning exciting subspaces. The final design may be viewed as a recursive algorithm for computing minimum 2-norm parameter estimates.

A. Literature Review

Central to any discussion of parameter adaptation is the notion of regressor excitation, among which the most important is *persistent excitation (PE)* [3]. Arguably the main practical limitation of adaptive control is the PE requirement, since it rarely holds in practice. Additionally, the PE condition is difficult to verify since it must hold over the entire time horizon, leading to a search for alternative characterizations. In [4], the PE condition is related to the degree of sufficient richness of a signal determined by its number of spectral lines. A geometric characterization of PE using an excitation distribution is provided in [5]. They note that almost periodic signals span the entire space they reside in; an insight relevant in our development. When a signal is not PE, one asks how much excitation still remains. In [6] the exosystem state of an LTI exosystem is shown to lie in a subspace whose dimension matches the number of excited modes. An alternative approach is to characterize the subspace that lacks persistent excitation, as in [7]. A more complete characterization of both the excited and unexcited subspaces for discrete-time systems was given earlier in [8].

Beyond characterizing the PE condition, one asks when parameter estimation is possible. The modern line of work in adaptive control is to perform parameter estimation under strictly weaker excitation than PE. A non-exhaustive list of these techniques include: Dynamic Regressor Extension and Mixing (DREM) [9] using regressor filtering, Concurrent Learning (CL) approaches [10] using a separate memory module and an initial excitation (IE) condition, and finite-time convergence (FTC) schemes [11]. Some authors have noticed that sometimes there is no form of excitation that can be leveraged and only partial parameter convergence can be guaranteed. For matrix regressors, a reduced order model guaranteeing parameter convergence has been obtained using the Gram-Schmidt procedure to identify linearly dependent columns [12]. In [13], a semi-IE condition is introduced and partial convergence is guaranteed along a subspace.

This paper adheres to a new design philosophy that parameter adaptation can be improved if one accounts for unexcited dynamics. This idea can be found in our previous work [2] as well as the independent works of [7], [14] and [15]. In [7], [14], the property of *lack of persistency of excitation* is introduced. A subspace estimator based on the Eigenvalue Decomposition (EVD) is used in linear regression and adaptive observers to guarantee parameter convergence by modifying the dynamics along the subspace that lacks persistent excitation. In [15], a DREM scheme for linear regression is considered under various *semi-excitation* conditions. An EVD is used to modify a matrix regressor along directions where it is rank deficient to enhance parameter convergence. Our work demonstrates the generality of this new design philosophy by: (i) characterizing the *inherent* excitation of regressors; (ii) providing a rigorous construction of subspace estimators using the Singular Value Decomposition; and (iii) elaborating the applicability to a general error model.

B. Contributions

Our contributions are multi-faceted and they culminate in the ability to perform robust parameter adaptation without sacrificing asymptotic error regulation. This paper is an extension of ideas in [2]. In [2] we assumed that the exogenous regressor already resides in its *PE subspace*. Instead, here we take a more intrinsic approach by defining the PE decomposition of a regressor arising from its autocovariance. Our new subspace estimator overcomes the following limitations of our prior design: (i) the initial condition can now be arbitrary; (ii) the convergence rate need not be slow; and (iii) we guarantee recovery of the non-PE subspace when using a regressor estimate. Conceptually our previous subspace estimator performed a projection onto the non-PE subspace, while our new subspace estimator starts by estimating the PE subspace. Lastly, while our prior work established robustness by arguing a nominal system is globally exponentially stable (GES), here we derive explicit steady-state robustness bounds.

Our contributions are as follows. In Section II we provide a complete characterization of PE subspaces for a large class of regressors through a PE decomposition. While the idea of exciting subspaces was first introduced in [8] for discrete-time systems, our characterization starting from the autocovariance matrix is not technically equivalent. In Section III, we rigorously construct a non-PE subspace estimator using Kreisselmeier filters and techniques from Principal Component Analysis (PCA). In Section IV, we perform an averaging analysis demonstrating that the design is best run slowly, consistent with biological systems. Lastly, in Section V we define a general error model based on the error models of [16] that can be robustified using the μ-modification, providing a template on how our methods can be applied to new contexts.

Explicit robustness bounds in the presence of various disturbances are presented. First in Theorem 7 we derive explicit bounds on the steady-state error model states in the presence of bounded disturbances, assuming those disturbances appearing in the regressor measurement and subspace estimator are sufficiently small. Theorem 8 removes the requirement of sufficiently small disturbances in the regressor measurement by treating such disturbances as part of the regressor excitation. Applications to linear regression and regulation using state feedback are found in Section VI. Technical proofs are found in Section VII.

C. Notation

Let $\|\cdot\|$ denote the 2-norm and $\sigma_i(\cdot)$ denote the i-th largest singular value. Then $\sigma_{\max}(\cdot) = \sigma_1(\cdot)$ denotes the largest singular value and $\sigma_{\min}(\cdot)$ denotes the smallest singular value. When applied to matrices, note that $\|\cdot\| = \sigma_{\max}(\cdot)$. Define the supremum norm $\|\cdot\|_{\mathcal{L}_\infty} := \sup_{t\geq t_0}\|\cdot\|$. We write $d \sim \mathcal{N}(\mu,\sigma^2)$ if $d(t)$ is white Gaussian noise with mean μ and variance σ^2. Given a subspace $\mathcal{W} \subseteq \mathbb{R}^q$, let $\mathcal{W}^\perp$ denote its orthogonal complement. The inequalities $\succeq, \succ$ denote the ordering of positive (semi-) definite matrices. Lastly, we write $\hat{w} \to w$ if $\hat{w}(t)$ converges asymptotically to $w(t)$.

II. PE Regressors and Subspaces

This section presents inherent properties of regressors and their persistently exciting subspaces. A regressor $w(t) \in \mathbb{R}^q$ is PE if there exist $\beta_0, T > 0$ such that

$$\frac{1}{T}\int_t^{t+T} w(\tau)w^\intercal(\tau)\,d\tau \succeq \beta_0 I\,, \quad \forall\, t\geq 0\,. \tag{5}$$

In [2] we considered a class of regressors that can be expressed as $w = Ww_{pe}$, where $W \in \mathbb{R}^{q\times q_{pe}}$ has orthonormal columns and $w_{pe}(t) \in \mathbb{R}^{q_{pe}}$ is PE. Additionally, we introduced the notion of a PE subspace (therein called PE directions) as the subspace along which a regressor is PE. In [2] the PE subspace associated with w is $\operatorname{Im}(W)$. Since the relationship $w = Ww_{pe}$ may not always hold, here we consider a class of regressors subsuming those previously studied.

Assumption 1: The regressor $w(t) \in \mathbb{R}^q$ is bounded and piecewise continuous, and its *autocovariance matrix*

$$R_w(0) := \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} w(\tau)w^\intercal(\tau)\,d\tau$$

exists, is independent of the initial time t_0, and the convergence of the time average is uniform in $t_0 \geq 0$. ▷

Remark 1: If a regressor w satisfies Assumption 1 and β_0 is the constant from (5), then $R_w(0) \succeq \beta_0 I$. ◁

Definition 1: Suppose Assumption 1 holds. The *PE subspace* $\mathcal{W} \subseteq \mathbb{R}^q$ of the regressor $w(t) \in \mathbb{R}^q$ is the subspace

$$\mathcal{W} := \operatorname{Im}\left(R_w(0)\right)\,.$$

The *non-PE subspace* of w is $\mathcal{W}^\perp$. We denote $q_{pe} := \dim(\mathcal{W})$ as its *degree of persistent excitation*. ▷

When $q_{pe} = q$, w is PE by [17, Proposition 2.7.1]. When $q_{pe} = 0$, we say w has *no persistent excitation*. When $1 \leq q_{pe} < q$, one can split w into a PE part and a part with no persistent excitation.

Proposition 1: Suppose Assumption 1 holds. If $1 \leq q_{pe} < q$, let $\begin{bmatrix} W & W_\perp \end{bmatrix} \in \mathbb{R}^{q\times q}$ be orthogonal such that

$$\mathcal{W} = \operatorname{Im}(W)\,, \quad \mathcal{W}^\perp = \operatorname{Im}(W_\perp)\,.$$

Then the *PE decomposition of* w is

$$w = WW^\intercal w + W_\perp W_\perp^\intercal w =: Ww_{pe} + W_\perp w_\perp\,, \tag{6}$$

where $w_{pe}(t) \in \mathbb{R}^{q_{pe}}$ is PE and $w_\perp(t) \in \mathbb{R}^{(q-q_{pe})}$ has no persistent excitation. ◇

Proof: Since $\mathbb{R}^q = \operatorname{Im}(W) \oplus \operatorname{Im}(W_\perp)$, $w = Ww_{pe} + W_\perp w_\perp$ for some $w_{pe} \in \mathbb{R}^{q_{pe}}$ and $w_\perp \in \mathbb{R}^{(q-q_{pe})}$. Then by orthogonality of W and $W_\perp$, we have $w_{pe} := W^\intercal w$ and $w_\perp := W_\perp^\intercal w$. Consider the autocovariance matrices

$$R_{w_{pe}}(0) = \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} w_{pe}(\tau)w_{pe}^\intercal(\tau)\,d\tau$$
$$R_{w_\perp}(0) = \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} w_\perp(\tau)w_\perp^\intercal(\tau)\,d\tau\,,$$

yielding the identities

$$R_{w_{pe}}(0) = W^\intercal R_w(0)W\,, \quad R_{w_\perp}(0) = W_\perp^\intercal R_w(0)W_\perp\,.$$

Given that $\operatorname{Im}(W_\perp) = \operatorname{Im}(R_w(0))^\perp = \operatorname{Ker}(R_w(0))$, we have $W_\perp^\intercal R_w(0)W_\perp = 0$ and so $w_\perp$ has no persistent excitation. Next, to show w_{pe} is PE it suffices to show $W^\intercal R_w(0)W \succ 0$ by [17, Proposition 2.7.1]. Suppose not. Then there exists a non-zero $v \in \mathbb{R}^{q_{pe}}$ such that

$$0 = v^\intercal W^\intercal R_w(0)Wv = \|R_w(0)^{1/2}Wv\|^2\,,$$

implying $Wv \in \operatorname{Ker}(R_w(0)) = \operatorname{Im}(W_\perp)$. But $Wv \in \operatorname{Im}(W)$ where $\operatorname{Im}(W)\cap\operatorname{Im}(W_\perp) = \{0\}$. Therefore $Wv = 0$ and so $v \in \operatorname{Ker}(W) = \{0\}$ because W has full column rank. This contradicts the fact that v is non-zero. □

Remark 2: The subspaces $\mathcal{W}$ and $\mathcal{W}^\perp$ are unique, and so are the vectors $Ww_{pe} = WW^\intercal w$ and $W_\perp w_\perp = W_\perp W_\perp^\intercal w$. Note that $WW^\intercal$ and $W_\perp W_\perp^\intercal$ are orthogonal projections. If w has no persistent excitation, then one may consider a PE decomposition with $W = 0$ and $W_\perp = I$; an analogous statement holds when w is PE. When $w = Ww_{pe}$ with w_{pe} PE, it follows that $\mathcal{W} = \operatorname{Im}(R_w(0)) = \operatorname{Im}(W)$, agreeing with the notion of a PE subspace in [2]. Finally, our Definition 1 can be considered a generalization of the independent work [7, Definition 2.1], where the non-PE subspace is the subspace that *lacks persistency of excitation of order* $q - q_{pe}$. ◁

So far we have discussed properties of an exogenous regressor $w(t)$. In practice one rarely has direct access to w; instead one builds an estimate $\hat{w} = w(t) + \tilde{w}$, where $\tilde{w}$ is the estimation error. The following fact from [16, Chapter 6.3.1] ensures that a vanishing transient $\tilde{w}$ does not modify the PE subspace of $\hat{w}$ relative to that of w.

Proposition 2: If $\lim_{t\to\infty}\tilde{w}(t) = 0$, then $\tilde{w}$ has zero average and thus no persistent excitation. ◇

The converse of Proposition 2 is not true in general. In particular, the fact that $w_\perp$ has no persistent excitation does not imply $w_\perp \to 0$. Even if we assume w is bounded and smooth with all its derivatives bounded, it is possible to construct a non-vanishing $w(t) \in \mathbb{R}$ with no persistent excitation. This can be done by partitioning $\mathbb{R}_+$ into a sequence of intervals indexed by k of length 2^k and fitting the same smooth bump function in each of these intervals. There is, nevertheless, an important case when one can conclude $w_\perp \to 0$; namely, the class of almost periodic functions, studied in stability theory [18] and adaptive control [19], among others. Additionally, the output of a linear time-invariant (LTI) exosystem having only simple poles on the imaginary axis, commonly used in regulator theory, is also an almost periodic function. The proof of the following proposition is found in Section VII-A.

Proposition 3: If w is (Bohr) almost periodic, then its PE decomposition satisfies $w_\perp = 0$. ◇

As a consequence of Propositions 2-3, an almost periodic steady-state implies a vanishing non-PE component.

Corollary 1: If there exists w_{ss} almost periodic such that $\lim_{t\to\infty} \|w(t) - w_{ss}(t)\| = 0$, then $\lim_{t\to\infty} w_\perp(t) = 0$. ◇

In light of Corollary 1, in Section III we consider the following regularity assumption. Section IV provides additional results without the assumption.

Assumption 2: The component $w_\perp$ of the regressor w along its non-PE subspace satisfies $\lim_{t\to\infty} w_\perp(t) = 0$. ▷

III. Subspace Estimator

Our approach to achieve robust parameter adaptation involves estimating the non-PE subspace $\mathcal{W}^\perp$ of w (or $\hat{w}$ if $\hat{w} \to w$). We apply the well-established Principal Component Analysis (PCA) used to compute *principal components* along which a collection of data points can be best explained. Principal components are generally computed as the *singular vectors* of the Singular Value Decomposition (SVD) applied to the sample covariance matrix. In our context, we seek singular vectors that span $\mathcal{W} = \mathrm{Im}(R_w(0))$, where $R_w(0)$ plays the analogous role of a sample covariance matrix.

Since $R_w(0)$ is symmetric and positive semi-definite, it has an orthogonal diagonalization which coincides with its SVD. As a result, the principal components are the eigenvectors of $R_w(0)$, which form a basis for $\mathcal{W}$. The singular values of $R_w(0)$ quantify the average excitation the regressor w has along these directions. Because $R_w(0)$ is not directly available, we are naturally lead to the following three step design:

1) generate a proxy of $R_w(0)$, denoted $\hat{\Sigma}$;
2) estimate $\mathcal{W}^\perp$ by applying PCA using $\hat{\Sigma}$;
3) use the estimate of $\mathcal{W}^\perp$, denoted Ω, in the μ-modification to achieve robust parameter adaptation.

A. Generating a Proxy of the Autocovariance Matrix

The first step of the design is to generate a proxy of $R_w(0)$ using a measurement $\hat{w}$. Our inspiration comes from the *Kreisselmeier integral algorithm*. Consider the filter (a minor variation of [20, Eq. (39a)])

$$\dot{\hat{\Sigma}} = -\varepsilon\hat{\Sigma} + \varepsilon\hat{w}\hat{w}^\intercal , \tag{7}$$

with $\varepsilon > 0$ and $\hat{\Sigma}(t) \in \mathbb{R}^{q\times q}$. We make the following observations. When $\hat{w}$ is PE, Kreisselmeier shows in [20, Theorem 3] that the filter matrix [20, Eq. (39a)] becomes positive definite eventually. When $\hat{w}$ is not PE, Marino and Tomei show in [7, Lemma 2.1] that for a particular initial condition, $\hat{\Sigma}$ has the same lack of persistency of excitation as $\hat{w}$. Finally, when $\varepsilon > 0$ is sufficiently small, (7) admits the averaged dynamics

$$\dot{\hat{\Sigma}}_{av} = -\varepsilon\hat{\Sigma}_{av} + \varepsilon R_{\hat{w}}(0) ,$$

which we investigate in Section IV. These observations suggest that $\hat{\Sigma}$ can recover the PE subspace of $\hat{w}$ (and thus w).

Let $\hat{\Sigma} = \Sigma_{pe} + \tilde{\Sigma}$, where Σ_{pe} will correspond to the steady-state component of the solution of (7) and $\tilde{\Sigma}$ corresponds to the transient component. Using the PE decomposition (6) and by linearity, (7) splits as

$$\dot{\Sigma}_{pe} = -\varepsilon\Sigma_{pe} + \varepsilon W w_{pe}(t) w_{pe}^\intercal(t) W^\intercal \tag{8a}$$

$$\dot{\tilde{\Sigma}} = -\varepsilon\tilde{\Sigma} + \varepsilon(\Delta_\perp(t) + \tilde{\Delta}(t)) , \tag{8b}$$

where

$$\Delta_\perp := W w_{pe} w_\perp^\intercal W_\perp^\intercal + W_\perp w_\perp w_{pe}^\intercal W^\intercal + W_\perp w_\perp w_\perp^\intercal W_\perp^\intercal$$
$$\tilde{\Delta} := w(\hat{w} - w)^\intercal + (\hat{w} - w)w^\intercal + (\hat{w} - w)(\hat{w} - w)^\intercal .$$

If $\hat{w} \to w$ and Assumption 2 holds, then we have $\Delta_\perp + \tilde{\Delta} \to 0$, implying that $\tilde{\Sigma} \to 0$ for any $\tilde{\Sigma}(t_0)$; thus confirming the interpretation of Σ_{pe} as the steady-state component of $\hat{\Sigma}$. Indeed, we will assign an initial condition (unknown to the designer) to (8a) so that the excitation properties of Σ_{pe} are invariant over time. The following lemma tells us that the image of Σ_{pe} coincides with the image of $R_w(0)$. Furthermore, we can relate the excitation properties of Σ_{pe} to those of w and the choice of ε. The proof is found in Section VII-B.

Lemma 1: Suppose Assumption 1 holds. Let $q_{pe} \ge 1$, w_{pe}, and $\mathcal{W} = \mathrm{Im}(W)$ result from the PE decomposition of w, and let $\beta_0, T > 0$ be the PE constants in (5) for the regressor w_{pe}. Consider (8a) with $\Sigma_{pe}(t_0) = \varepsilon T R_w(0)$. Then there exists a bounded symmetric $\Lambda_{pe}(t) \in \mathbb{R}^{q_{pe}\times q_{pe}}$ such that

$$\Sigma_{pe}(t) = W\Lambda_{pe}(t)W^\intercal , \quad \Lambda_{pe}(t) \succeq \varepsilon T\beta_0 \mathrm{e}^{-\varepsilon T} I$$

for all $t \ge t_0 \ge 0$. Hence, $\mathrm{Im}(\Sigma_{pe}(t)) = \mathrm{Im}(W)$ for all $t \ge t_0 \ge 0$. ◇

In the case when $q_{pe} = 0$, we have $R_w(0) = 0$ and so there is no PE component w_{pe}. Then the steady-state dynamics (8a) become $\dot{\Sigma}_{pe} = -\varepsilon\Sigma_{pe}$ with $\Sigma_{pe}(t_0) = \varepsilon T R_w(0) = 0$, resulting in $\Sigma_{pe}(t) = 0$ for all $t \ge t_0 \ge 0$.

B. Recovery of the non-PE Subspace using PCA

The second step of the design is to recover the non-PE subspace of w from $\hat{\Sigma}$. We perform an SVD on $\hat{\Sigma}(t)$ (which can be done using numerically stable algorithms), resulting in

$$\begin{aligned}\hat{\Sigma}(t) &= U(t)D(t)V^\intercal(t) \\ &= \begin{bmatrix} U_1(t) & U_2(t) \end{bmatrix} \begin{bmatrix} D_1(t) & 0 \\ 0 & D_2(t) \end{bmatrix} \begin{bmatrix} V_1^\intercal(t) \\ V_2^\intercal(t) \end{bmatrix} ,\end{aligned} \tag{9}$$

where each $D_i(t)$ is diagonal and positive semi-definite, $\sigma_{\min}(D_1(t)) \ge \sigma_{\max}(D_2(t))$, and $D_1(t) \in \mathbb{R}^{q_{pe}\times q_{pe}}$ with all other matrices agreeing in dimension. Also, $U(t), V(t) \in \mathbb{R}^{q\times q}$ are orthogonal matrices; i.e., $UU^\intercal = U^\intercal U = I$, $VV^\intercal =$

$V^\intercal V = I$. When $q_{pe} \in \{1, q\}$ we take the appropriate U_i, D_i, and V_i to be zero to simplify dealing with corner cases. Our goal is to show that $\mathrm{Im}(U_1(t))$ recovers the PE subspace $\mathcal{W}$ and $\mathrm{Im}(U_2(t))$ recovers the non-PE subspace $\mathcal{W}^\perp$. Since $q_{pe} = \dim(\mathcal{W})$ is unknown, the selection of principal components spanning the PE or non-PE subspaces must be based on the singular values. To this end, we apply tools from *matrix perturbation theory* [21].

Consider $\hat{\Sigma} = \Sigma_{pe} + \tilde{\Sigma}$, where $\hat{\Sigma}$ is regarded as a perturbation of Σ_{pe}. Since Σ_{pe} is a steady-state signal with unknown initial condition, it is not available; nevertheless, for theoretical purposes we consider its SVD. By Lemma 1, $\Sigma_{pe}(t)$ has rank q_{pe} and is symmetric positive semi-definite for all $t \geq t_0 \geq 0$. Therefore, its SVD has the form

$$\Sigma_{pe}(t) = \begin{bmatrix} U_3(t) & U_4(t) \end{bmatrix} \begin{bmatrix} D_3(t) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_3^\intercal(t) \\ U_4^\intercal(t) \end{bmatrix}, \qquad (10)$$

with a matrix partition consistent with (9). The following is the main consequence of Lemma 1.

Corollary 2: Let Σ_{pe} be as in Lemma 1 with SVD (10). Then $WW^\intercal = U_3(t)U_3^\intercal(t)$ for all q_{pe} and $t \geq t_0 \geq 0$. If $q_{pe} \geq 1$, then $\sigma_{\min}(D_3(t)) \geq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T}$ for all $t \geq t_0 \geq 0$. ⋄

Proof: When $q_{pe} = 0$, we have $WW^\intercal = 0$ by virtue of the PE decomposition. Additionally, we have $U_3 U_3^\intercal = 0$ given that $\Sigma_{pe} = 0$. Otherwise, the result $WW^\intercal = U_3(t)U_3^\intercal(t)$ follows from the fact that $\mathrm{Im}(W) = \mathrm{Im}(\Sigma_{pe}(t)) = \mathrm{Im}(U_3(t))$ by Lemma 1 and the SVD (10), followed by the fact that orthogonal projections are unique.

Again from Lemma 1 and (10), we have

$$\Sigma_{pe}(t) = W\Lambda_{pe}(t)W^\intercal = U_3(t)D_3(t)U_3^\intercal(t)\,.$$

Given that $\mathrm{rank}(\Sigma_{pe}(t)) = q_{pe}$, its q_{pe}-th singular value is its smallest non-zero singular value for all $t \geq t_0 \geq 0$. Thus

$$\begin{aligned} \sigma_{\min}(D_3(t)) &= \sigma_{q_{pe}}(U_3(t)D_3(t)U_3^\intercal(t)) \\ &= \sigma_{q_{pe}}(W\Lambda_{pe}(t)W^\intercal) \\ &= \sigma_{\min}(\Lambda_{pe}(t)) \geq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T} \end{aligned}$$

for all $t \geq t_0 \geq 0$. The first equality is by definition of the SVD. The second equality is from Lemma 1. The third equality can be seen by writing an SVD for $\Lambda_{pe}(t)$. The last inequality is again from Lemma 1. □

To compare singular values of our computed $\hat{\Sigma}$ and the steady-state Σ_{pe} one applies Weyl's Theorem.

Theorem 1 (Weyl): Let $S, S_{pe} \in \mathbb{R}^{q \times q}$. Then $|\sigma_i(S) - \sigma_i(S_{pe})| \leq \|S - S_{pe}\|$. ⋄

We also want to compare the column spans of $\hat{\Sigma}$ and Σ_{pe}. Wedin's Theorem relates the column spans of the U_i.

Theorem 2 (Wedin [22]): Let $S = UDV^\intercal$ and $S_{pe} = U_{pe}D_{pe}V_{pe}^\intercal$ admit SVDs partitioned as in (9) and (10). If there exists a constant $\sigma_{tol} > 0$ such that $\sigma_{\min}(D_1) \geq \sigma_{tol}$, then

$$\|U_1U_1^\intercal - U_3U_3^\intercal\| \leq \frac{\max\{\, \|(S - S_{pe})V_1\|, \|U_1^\intercal(S - S_{pe})\|\,\}}{\sigma_{tol}}\,.$$

⋄

We may now determine the subspaces that our $U_1(t)$ and $U_2(t)$, computed from the SVD (9), recover asymptotically.

Lemma 2: Let $S = UDV^\intercal$ and $\Sigma_{pe}(t)$ admit SVDs partitioned as in (9) and (10), where Σ_{pe} is as in Lemma 1. Let $q_{pe} \geq 1$. Suppose there exists a constant $\sigma_{tol} > 0$ such that $\sigma_{\min}(D_1) \geq \sigma_{tol}$. Then

$$\begin{aligned} \|U_1U_1^\intercal - WW^\intercal\| &= \|U_2U_2^\intercal - W_\perp W_\perp^\intercal\| \\ &\leq \min\{\, \sigma_{tol}^{-1}\|S - \Sigma_{pe}(t)\|, 1\,\} \end{aligned}$$

for all $t \geq t_0 \geq 0$. ⋄

Proof: Using the invariance of the induced 2-norm under multiplication by a unitary matrix, we have

$$\begin{aligned} \|(S - \Sigma_{pe}(t))V_1\| &\leq \|S - \Sigma_{pe}(t)\| \\ \|U_1^\intercal(S - \Sigma_{pe}(t))\| &\leq \|S - \Sigma_{pe}(t)\|\,. \end{aligned}$$

By Corollary 2, $U_3(t)U_3^\intercal(t) = WW^\intercal$ for all $t \geq t_0 \geq 0$. Thus we can apply Wedin's Theorem (Theorem 2) to obtain

$$\|U_1U_1^\intercal - WW^\intercal\| \leq \sigma_{tol}^{-1}\|S - \Sigma_{pe}(t)\|\,.$$

The inclusion of the $\min\{\,\cdot, 1\,\}$ is a consequence of $\|U_1U_1^\intercal - WW^\intercal\| \leq 1$ given that we have a difference of orthogonal projection matrices [23]. The final result then follows by the facts $W_\perp W_\perp^\intercal = I - WW^\intercal$ and $U_2U_2^\intercal = I - U_1U_1^\intercal$. □

We are interested in showing $U_2U_2^\intercal \to W_\perp W_\perp^\intercal$. As such, we need to isolate U_2 in (9).

Lemma 3: Let $S = UDV^\intercal$ and $\Sigma_{pe}(t)$ admit SVDs partitioned as in (9) and (10), where Σ_{pe} is as in Lemma 1. If $q_{pe} \geq 1$, then

$$\begin{aligned} \sigma_{\min}(D_1) &\geq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T} - \|S - \Sigma_{pe}(t)\| \\ \sigma_{\max}(D_2) &\leq \|S - \Sigma_{pe}(t)\| \end{aligned}$$

for all $t \geq t_0 \geq 0$.

Proof: Using Weyl's Theorem (Theorem 1) we have

$$\begin{aligned} |\sigma_{q_{pe}}(S) - \sigma_{q_{pe}}(\Sigma_{pe}(t))| &\leq \|S - \Sigma_{pe}(t)\| \\ |\sigma_{q_{pe}+1}(S) - \sigma_{q_{pe}+1}(\Sigma_{pe}(t))| &\leq \|S - \Sigma_{pe}(t)\|\,. \end{aligned}$$

By our SVDs (9) and (10) we know

$$\begin{aligned} \sigma_{q_{pe}}(S) &= \sigma_{\min}(D_1)\,, & \sigma_{q_{pe}}(\Sigma_{pe}(t)) &= \sigma_{\min}(D_3(t)) \\ \sigma_{q_{pe}+1}(S) &= \sigma_{\max}(D_2)\,, & \sigma_{q_{pe}+1}(\Sigma_{pe}(t)) &= 0\,. \end{aligned}$$

At this point, the second inequality is immediate. Note that when $q_{pe} = q$, we take $D_2 = 0$ and so the second inequality still holds. For the first inequality, we use the reverse triangle inequality to obtain

$$\sigma_{\min}(D_1) \geq \sigma_{\min}(D_3(t)) - \|S - \Sigma_{pe}(t)\|\,.$$

The result then follows by Corollary 2. □

At this stage recall that if $\hat{\Sigma}$ evolves according to (7) and $\hat{w} \to w$, then $\hat{\Sigma} \to \Sigma_{pe}$. Then by Lemma 3, with $S = \hat{\Sigma}(t)$, its SVD satisfies

$$\begin{aligned} \liminf_{t\to\infty} \sigma_{\min}(D_1(t)) &\geq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T} \\ \limsup_{t\to\infty} \sigma_{\max}(D_2(t)) &= 0\,. \end{aligned}$$

Therefore the matrix $U_1(t)$ can be recovered eventually by selecting the principal components of $\hat{\Sigma}$ associated with singular values larger than or equal to $\varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T}$, whereas the remaining principal components must eventually reconstruct $U_2(t)$. As such, we introduce a threshold σ_{tol} to separate $U_1(t)$ from $U_2(t)$ based on their singular values.

To formalize the above, define the binary threshold function

$$\mathrm{bin}(c;\sigma_{tol}) = \begin{cases} 1 & c \geq \sigma_{tol} \\ 0 & c < \sigma_{tol} \end{cases},$$

which is applied component-wise for multivariable inputs. Recall $\hat{\Sigma}(t) = U(t)D(t)V^\intercal(t)$ is its SVD. In practice, the SVD is not computed continuously but rather at an increasing sequence of times $\{ t_i \}_{i=0}^{\infty}$ where $t_i \to \infty$. The output of the non-PE subspace estimator $\Omega(t) \in \mathbb{R}^{q\times q}$ is constructed as

$$\Omega(t) = \begin{cases} U(t)\,(I - \mathrm{bin}(D(t);\sigma_{tol})) & t = t_i \\ \Omega(t_i) & t \in (t_i, t_{i+1}) \end{cases}. \quad (11)$$

Theorem 3: Suppose Assumption 1 holds. Let q_{pe}, w_{pe}, and $\mathcal{W}^\perp = \mathrm{Im}(W_\perp)$ result from the PE decomposition of w, and let $\beta_0, T > 0$ be the PE constants in (5) for the regressor w_{pe}. If $q_{pe} = 0$, set $\beta_0 = \infty$ and $T > 0$ to any finite value. Consider (8a) with $\Sigma_{pe}(t_0) = \varepsilon T R_w(0)$. For any $S \in \mathbb{R}^{q\times q}$ let $S = UDV^\intercal$ denote its SVD and define

$$\Omega := U(I - \mathrm{bin}(D;\sigma_{tol}))\,.$$

Then for every $\sigma_{tol} \in (0, \varepsilon T\beta_0 \mathrm{e}^{-\varepsilon T})$ there exists a constant $c_{tol}(\varepsilon, T, \beta_0, \sigma_{tol}) > 0$ such that

$$\|\Omega\Omega^\intercal - W_\perp W_\perp^\intercal\| \leq \min\{\, c_{tol}\|S - \Sigma_{pe}(t)\|, 1\,\}$$

for all $t \geq t_0 \geq 0$. ⋄

Proof: If $q_{pe} = 0$ then $\Sigma_{pe} = 0$, and so

$$\sigma_{\max}(D) = \sigma_{\max}(S) = \|S\| = \|S - \Sigma_{pe}(t)\|\,,$$

where we use $\sigma_{\max}(\cdot) = \|\cdot\|$. If $\|S - \Sigma_{pe}(t)\| < \sigma_{tol}$ we have $\mathrm{bin}(D;\sigma_{tol}) = 0$, implying that $\Omega = U(I - 0) = U$. Recalling that $UU^\intercal = I$ we obtain

$$\|\Omega\Omega^\intercal - W_\perp W_\perp^\intercal\| = \|UU^\intercal - I\| = 0 \leq \sigma_{tol}^{-1}\|S - \Sigma_{pe}(t)\|\,.$$

Altogether, $\|\Omega\Omega^\intercal - W_\perp W_\perp^\intercal\|$ is upper bounded by

$$\begin{cases} \sigma_{tol}^{-1}\|S - \Sigma_{pe}(t)\| & \|S - \Sigma_{pe}(t)\| < \sigma_{tol} \\ 1 & \text{otherwise} \end{cases}$$
$$= \min\{\, \sigma_{tol}^{-1}\|S - \Sigma_{pe}(t)\|, 1\,\}\,,$$

where the property $\|\Omega\Omega^\intercal - W_\perp W_\perp^\intercal\| \leq 1$ is stated in [23].

Next suppose $q_{pe} \geq 1$. By Lemma 3 and our choice of σ_{tol}, there exists $\delta_{tol}(\varepsilon, T, \beta_0, \sigma_{tol}) > 0$ such that $\|S - \Sigma_{pe}(t)\| \leq \delta_{tol}$ implies

$$\sigma_{\min}(D_1) \geq \sigma_{tol}\,, \quad \sigma_{\max}(D_2) < \sigma_{tol}\,.$$

Therefore, if $\|S - \Sigma_{pe}(t)\| \leq \delta_{tol}$ we have $\Omega = U(I - \mathrm{diag}(I, 0)) = \begin{bmatrix} 0 & U_2 \end{bmatrix}$, implying that $\Omega\Omega^\intercal = U_2U_2^\intercal$. Using Lemma 2 there exists a constant $c_1(\delta_{tol}, \sigma_{tol}) > 0$ such that $\|\Omega\Omega^\intercal - W_\perp W_\perp^\intercal\|$ is upper bounded by

$$\begin{cases} \sigma_{tol}^{-1}\|S - \Sigma_{pe}(t)\| & \|S - \Sigma_{pe}(t)\| \leq \delta_{tol} \\ 1 & \text{otherwise} \end{cases}$$
$$\leq \min\{\, c_1\|S - \Sigma_{pe}(t)\|, 1\,\}\,.$$

As a result, the theorem follows from combining the inequalities derived for $q_{pe} = 0$ and $q_{pe} \geq 1$. □

We conclude in the following that our subspace estimator (7), (11) asymptotically recovers the non-PE subspace and its dimension from a regressor estimate $\hat{w}$.

Theorem 4: Suppose Assumptions 1-2 hold and $\hat{w} \to w$. Let q_{pe}, $W_\perp$, β_0, and T be as defined in Theorem 3. Consider the subspace estimator (7), (11) where $\hat{\Sigma}(t) = U(t)D(t)V^\intercal(t)$ by its SVD and $\{ t_i \}_{i=0}^{\infty}$ is an increasing sequence with $t_i \to \infty$. Then for every $\sigma_{tol} \in (0, \varepsilon T\beta_0\mathrm{e}^{-\varepsilon T})$ we have

$$\lim_{t\to\infty} \|\Omega(t)\Omega^\intercal(t) - W_\perp W_\perp^\intercal\| = 0$$
$$\lim_{t\to\infty} \mathrm{rank}(\Omega(t)\Omega^\intercal(t)) = q - q_{pe}$$

for any $\hat{\Sigma}(t_0) \in \mathbb{R}^{q\times q}$. ⋄

Proof: If Assumption 2 holds and $\hat{w} \to w$ then we know that $\hat{\Sigma} \to \Sigma_{pe}$. Since Assumption 1 holds, we may apply Theorem 3 by letting $S = \hat{\Sigma}(t_i)$ for each t_i, resulting in

$$\|\Omega(t_i)\Omega^\intercal(t_i) - W_\perp W_\perp^\intercal\| \leq \min\{\, c_{tol}\|\hat{\Sigma}(t_i) - \Sigma_{pe}(t_i)\|, 1\,\}\,.$$

The first limit then holds because $\hat{\Sigma} \to \Sigma_{pe}$ and any subsequence of a convergent sequence must converge to the same limit given that $t_i \to \infty$.

To see the second limit we split into two cases. If $q_{pe} = 0$, then $\hat{\Sigma} \to \Sigma_{pe} = 0$ as argued after Lemma 1. Thus there exists a $\Delta t \geq 0$ such that $\Omega(t_i) = U(t_i)$ for all $t_i \geq t_0 + \Delta t$, and the result follows since $U(t_i)U^\intercal(t_i) = I$. Otherwise, by Lemma 3 there exists a $\Delta t \geq 0$ such that

$$\sigma_{\min}(D_1(t_i)) \geq \sigma_{tol}\,, \quad \sigma_{\max}(D_2(t_i)) < \sigma_{tol}$$

for all $t_i \geq t_0 + \Delta t$. Then $\Omega(t_i) = \begin{bmatrix} 0 & U_2(t_i) \end{bmatrix}$ and the result follows because $\mathrm{rank}(U_2(t_i)) = q - q_{pe}$. □

Remark 3: From a design perspective, knowledge of some $\sigma_{tol} \in (0, \varepsilon T\beta_0\mathrm{e}^{-\varepsilon T})$ is not restrictive. Theorem 3 tells us σ_{tol} must merely be selected sufficiently small to achieve a correctly functioning subspace estimator. Further, one could estimate bounds on β_0 and T using past data by probing regressors, similar in spirit to methods in [24]. Finally, we endorse the view that σ_{tol} is a design tolerance: one sets a threshold σ_{tol} to declare a minimum PE level permitted to drive adaptation, thus allowing the designer to distinguish relevant regressor excitation from measurement noise; see also [2]. ◁

IV. Averaging Analysis

This section applies Krylov-Bogoliubov-Mitropolsky averaging [25] to reveal that the proposed design is best run slow. We show that if one relaxes Assumption 2, then it is still possible to approximate $R_w(0)$ (and thus the non-PE subspace) for $\varepsilon > 0$ sufficiently small.

Recall from Lemma 1 that we found a bound $\Lambda_{pe}(t) \succeq \varepsilon T\beta_0\mathrm{e}^{-\varepsilon T}I$. This bound suggests that using a large gain ε in (7) makes recovery of the non-PE subspace numerically difficult, since $\lim_{\varepsilon\to\infty} \varepsilon T\beta_0\mathrm{e}^{-\varepsilon T} = 0$. This observation makes sense because a PE regressor may have long periods of no excitation (as long as sufficient excitation is available periodically), resulting in a $\hat{\Sigma}$ rapidly tracking intervals of no excitation. The same bound $\varepsilon T\beta_0\mathrm{e}^{-\varepsilon T}$ misleads one to think that selecting ε small must result in similar numerical issues. We show using averaging analysis that this is not the case.

First, we drop Assumption 2, implying that $\tilde{\Sigma}$ of (8b) does not necessarily converge to 0. As before we split $\hat{\Sigma} = \Sigma_{pe} + \Sigma_{\perp} + \tilde{\Sigma}$, where Σ_{pe} corresponds to the steady-state PE component, $\Sigma_{\perp}$ corresponds to the steady-state non-PE component, and $\tilde{\Sigma}$ is the transient component. With some overloaded notation and by linearity, (7) can be split as

$$\dot{\Sigma}_{pe} = -\varepsilon\Sigma_{pe} + \varepsilon W w_{pe}(t) w_{pe}^{\mathsf{T}}(t) W^{\mathsf{T}} \tag{12a}$$

$$\dot{\Sigma}_{\perp} = -\varepsilon\Sigma_{\perp} + \varepsilon\Delta_{\perp}(t) \tag{12b}$$

$$\dot{\tilde{\Sigma}} = -\varepsilon\tilde{\Sigma} + \varepsilon\tilde{\Delta}(t)\,, \tag{12c}$$

where $\hat{\Sigma}(t_0) = \Sigma_{pe}(t_0) + \Sigma_{\perp}(t_0) + \tilde{\Sigma}(t_0)$ and $\Delta_{\perp}$, $\tilde{\Delta}$ are defined in Section III-A. If $\hat{w} \to w$, then $\tilde{\Delta}, \tilde{\Sigma} \to 0$ as before. Instead $\Delta_{\perp}$ does not necessarily vanish. Also notice $(\Sigma_{pe}, \Sigma_{\perp}) = (R_w(0), 0)$ need not be an equilibrium of (12a)-(12b). Define $\tilde{\Sigma}_{pe} := \Sigma_{pe} - R_w(0)$. Using (12a), we have

$$\dot{\tilde{\Sigma}}_{pe} = -\varepsilon\tilde{\Sigma}_{pe} + \varepsilon\left(W w_{pe}(t) w_{pe}^{\mathsf{T}}(t) W^{\mathsf{T}} - R_w(0)\right)\,.$$

We are interested in computing the average dynamics, which exist because $R_w(0)$ exists with convergence uniform in $t_0 \geq 0$. Using the PE decomposition $w = W w_{pe} + W_{\perp} w_{\perp}$, we have

$$w(t) w^{\mathsf{T}}(t) = W w_{pe}(t) w_{pe}^{\mathsf{T}}(t) W^{\mathsf{T}} + \Delta_{\perp}(t)\,. \tag{13}$$

Recall $w_{pe} = W^{\mathsf{T}} w$ and $\mathcal{W}^{\perp} = \operatorname{Ker}(R_w(0))$. Then

$$\begin{aligned} R_{W w_{pe}}(0) &= W W^{\mathsf{T}} R_w(0) W W^{\mathsf{T}} \\ &= (I - W_{\perp} W_{\perp}^{\mathsf{T}}) R_w(0) (I - W_{\perp} W_{\perp}^{\mathsf{T}}) = R_w(0)\,. \end{aligned}$$

Thus, $W w_{pe} w_{pe}^{\mathsf{T}} W^{\mathsf{T}} - R_w(0)$ has zero average. Now take the time average of (13):

$$R_w(0) = R_{W w_{pe}}(0) + \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} \Delta_{\perp}(\tau)\, d\tau\,.$$

Combining the previous two calculations, we find $\Delta_{\perp}$ has zero average. We conclude that the averaged dynamics of (12a)-(12b) are

$$(\dot{\tilde{\Sigma}}_{pe})_{av} = -\varepsilon(\tilde{\Sigma}_{pe})_{av} \tag{14a}$$

$$(\dot{\Sigma}_{\perp})_{av} = -\varepsilon(\Sigma_{\perp})_{av}\,. \tag{14b}$$

The next result shows that Σ_{pe} approximates $R_w(0)$ to an arbitrary precision provided $\varepsilon > 0$ is sufficiently small.

Lemma 4: Suppose Assumption 1 holds. Consider the systems (12a)-(12b) with $(\Sigma_{pe}, \Sigma_{\perp})(t_0) = (R_w(0), 0)$. Then there exist class-$\mathcal{K}$ functions $\delta_{pe}^{w}(\cdot)$, $\delta_{\perp}^{w}(\cdot)$ and a constant $\varepsilon_1(w) > 0$ such that

$$\|\Sigma_{pe}(t) - R_w(0)\| \leq \delta_{pe}^{w}(\varepsilon)\,, \quad \|\Sigma_{\perp}(t)\| \leq \delta_{\perp}^{w}(\varepsilon)\,,$$

for all $\varepsilon \in (0, \varepsilon_1]$ and $t \geq t_0 \geq 0$. ◇

Proof: Clearly the equilibrium $((\tilde{\Sigma}_{pe})_{av}, (\Sigma_{\perp})_{av}) = (0, 0)$ is GES for (14). Due to our choice of initial conditions $(R_w(0), 0)$, we have $((\tilde{\Sigma}_{pe})_{av}, (\Sigma_{\perp})_{av})(t) = (0, 0)$ for all $t \geq t_0 \geq 0$. Then by the Hovering Theorem [25, Theorem 5.5.1] there exist class-$\mathcal{K}$ functions $\delta_{pe}^{w}(\cdot)$, $\delta_{\perp}^{w}(\cdot)$ and a constant $\varepsilon_1(w) > 0$ such that

$$\begin{aligned} \delta_{pe}^{w}(\varepsilon) &\geq \|\tilde{\Sigma}_{pe}(t) - (\tilde{\Sigma}_{pe})_{av}(t)\| = \|\Sigma_{pe}(t) - R_w(0)\| \\ \delta_{\perp}^{w}(\varepsilon) &\geq \|\Sigma_{\perp}(t) - (\Sigma_{\perp})_{av}(t)\| = \|\Sigma_{\perp}(t)\| \end{aligned}$$

for all $\varepsilon \in (0, \varepsilon_1]$ and $t \geq t_0 \geq 0$. □

With $\Sigma_{pe}(t_0) = R_w(0)$ one may re-apply the argument at the beginning of the proof of Lemma 1 to deduce that $\Sigma_{pe}(t) = W \Lambda_{pe}(t) W^{\mathsf{T}}$. By unitary invariance of the induced 2-norm and Lemma 4, we have

$$\|W^{\mathsf{T}}(\Sigma_{pe}(t) - R_w(0)) W\| \leq \|\Sigma_{pe}(t) - R_w(0)\| \leq \delta_{pe}^{w}(\varepsilon)\,.$$

Then we obtain

$$\begin{aligned} \Lambda_{pe}(t) &= W^{\mathsf{T}} \Sigma_{pe}(t) W \\ &= R_{w_{pe}}(0) + W^{\mathsf{T}}\left(\Sigma_{pe}(t) - R_w(0)\right) W \\ &\succeq \left(\beta_0 - \delta_{pe}^{w}(\varepsilon)\right) I \end{aligned} \tag{15}$$

for all $\varepsilon \in (0, \varepsilon_1]$ and $q_{pe} \geq 1$. This lower bound provides a significantly better approximation of the regressor excitation level compared to our previous bound $\varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T}$ in Lemma 1, if $\varepsilon > 0$ is sufficiently small. We can see this because

$$\beta_0 = \lim_{\varepsilon\to 0^+} \left(\beta_0 - \delta_{pe}^{w}(\varepsilon)\right) > \lim_{\varepsilon\to 0^+} \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T} = 0\,.$$

The implication is that rather than first selecting ε and then setting σ_{tol} sufficiently close to 0, instead we can select σ_{tol} to be representative of the excitation level β_0 and then set ε to sufficiently slow down the subspace estimator. Effectively, we have increased the allowable range for σ_{tol} up to the PE lower bound β_0. As stated below, with proof in Section VII-C, this is possible when system (7) is run sufficiently slowly.

Theorem 5: Suppose Assumption 1 holds. Let q_{pe}, w_{pe}, and $\mathcal{W}^{\perp} = \operatorname{Im}(W_{\perp})$ result from the PE decomposition of w, and let $\beta_0 > 0$ be the PE constant in (5) for the regressor w_{pe}. If $q_{pe} = 0$, set $\beta_0 = \infty$. Consider (8a) with $\Sigma_{pe}(t_0) = R_w(0)$. For any $S \in \mathbb{R}^{q\times q}$ let $S = U D V^{\mathsf{T}}$ denote its SVD and define

$$\Omega := U(I - \operatorname{bin}(D; \sigma_{tol}))\,.$$

Then for every $\sigma_{tol} \in (0, \beta_0)$ there exists constants $\varepsilon_{\star}(w, \sigma_{tol})$, $c_{tol}(w, \sigma_{tol}) > 0$ such that

$$\|\Omega\Omega^{\mathsf{T}} - W_{\perp} W_{\perp}^{\mathsf{T}}\| \leq \min\{\, c_{tol}\|S - \Sigma_{pe}(t)\|, 1\,\}$$

for all $\varepsilon \in (0, \varepsilon_{\star}]$ and $t \geq t_0 \geq 0$. ◇

Similar to Theorem 4, the following result establishes convergence results when employing the subspace estimator (7), (11). We note that the dimension of the non-PE subspace $\mathcal{W}^{\perp}$ can still be recovered even if $\mathcal{W}^{\perp}$ itself cannot be. Its proof is found in Section VII-D.

Theorem 6: Suppose Assumption 1 holds and $\hat{w} \to w$. Let q_{pe}, $W_{\perp}$, and β_0 be as defined in Theorem 5. Also, let $\delta_{\perp}^{w}(\cdot)$ be given by Lemma 4. Consider the subspace estimator (7), (11) where $\hat{\Sigma}(t) = U(t) D(t) V^{\mathsf{T}}(t)$ by its SVD and $\{\, t_i \,\}_{i=0}^{\infty}$ is an increasing sequence with $t_i \to \infty$. Then for every $\sigma_{tol} \in (0, \beta_0)$ there exists constants $\varepsilon_{\star}(w, \sigma_{tol})$, $\varepsilon_{\star\star}(w, \sigma_{tol})$, $c_{tol}(w, \sigma_{tol}) > 0$ such that

$$\limsup_{t\to\infty} \|\Omega(t)\Omega^{\mathsf{T}}(t) - W_{\perp} W_{\perp}^{\mathsf{T}}\| \leq c_{tol}\delta_{\perp}^{w}(\varepsilon)$$

for all $\varepsilon \in (0, \varepsilon_\star]$ and $\lim_{t\to\infty} \operatorname{rank}(\Omega(t)\Omega^\intercal(t)) = q - q_{pe}$ for all $\varepsilon \in (0, \varepsilon_{\star\star}]$, both for any $\hat{\Sigma}(t_0) \in \mathbb{R}^{q\times q}$. If Assumption 2 also holds then we have

$$\lim_{t\to\infty} \|\Omega(t)\Omega^\intercal(t) - W_\perp W_\perp^\intercal\| = 0$$

for all $\varepsilon \in (0, \varepsilon_\star]$ and any $\hat{\Sigma}(t_0) \in \mathbb{R}^{q\times q}$. ◇

Remark 4: The $\limsup$ is used because $\Omega\Omega^\intercal$ need not have a limit as $t \to \infty$. We do not pursue obtaining an estimate of $\varepsilon_\star(w, \sigma_{tol})$ nor of $\delta^w_{pe}(\cdot)$, $\delta^w_\perp(\cdot)$ as this would involve a lengthy discussion about averaging theory. ◁

The primary forte of Theorems 5-6 is that one has an effective design even when Assumption 2 does not hold. In comparing Theorem 4 and Theorem 6, we observe that the Hovering Theorem [25, Theorem 5.5.1] allows us to incorporate a meaningful bound on the mismatch between the subspace estimate $\Omega(t)\Omega^\intercal(t)$ and the projection onto the non-PE subspace $W_\perp W_\perp^\intercal$. One cannot obtain this desirable behaviour with a large gain ε. Additionally, in comparing Theorem 3 and Theorem 5, we may now threshold with the significantly larger constant $\sigma_{tol} \in (0, \beta_0)$ rather than $\sigma_{tol} \in (0, \varepsilon T\beta_0 \mathrm{e}^{-\varepsilon T})$ as $\varepsilon \to 0^+$, thus improving numerical stability.

From our analyses in Section III and Section IV, we have designed a subspace estimator whose steady-state behaviour is agnostic to the initial condition $\hat{\Sigma}(t_0)$, is valid for all gains $\varepsilon > 0$, and is agnostic to the presence of transients $\hat{w} \to w$.

V. Robust Parameter Adaptation

Consider a parameter adaptation law of the form

$$\dot{\hat{\psi}} = -\gamma e \hat{w} \tag{16}$$

with adaptation gain $\gamma > 0$, scalar error signal $e(t) \in \mathbb{R}$, and regressor estimate $\hat{w}(t) \in \mathbb{R}^q$. It is well known that (16) is not in general robust [1]. We are interested to render the parameter adaptation dynamics (16) robust without sacrificing error regulation. This problem has remained largely unsolved in the adaptive control literature without introducing further assumptions such as having a priori information about ψ [1, Section 8.5] or assuming weaker notions of excitation [26]. We forgo such assumptions by combining our newly developed subspace estimator with our μ-modification presented in [2].

A. General Error Model

Our development of a robust parameter adaptation scheme begins by presenting a general error model encompassing all the relevant features of the error models appearing in [16, Ch. 7]. We consider an error model $e = \mathscr{E}[\hat{\psi}, w, \nu]$, where $\mathscr{E}[\cdot]$ is a time-varying system of the form

$$\dot{\xi} = f(t, \xi, \hat{w}_\circ^\intercal \hat{\psi} - w_\circ^\intercal(t)\psi) \tag{17a}$$

$$e = g(t, \xi, \hat{w}^\intercal \hat{\psi} - w^\intercal(t)\psi)\,, \tag{17b}$$

where $\xi(t) \in \mathbb{R}^n$ is the error state; $\hat{\psi}(t) \in \mathbb{R}^q$ is the parameter estimate; $\psi \in \mathbb{R}^q$ is the unknown parameter; $w_\circ(t) \in \mathbb{R}^q$ is a regressor related to w; and $\nu(t) \in \mathbb{R}^v$ is a transient state. The transient state ν arises due to transients that vanish independently of all other dynamics when using the estimates $\hat{w}_\circ$ and $\hat{w}$. These transients are assumed to be generated by an asymptotically stable system

$$\dot{\nu} = \Delta(t, \nu) \tag{18a}$$

$$\hat{w}_\circ = w_\circ(t) + \tilde{w}_\circ(t, \nu) \tag{18b}$$

$$\hat{w} = w(t) + \tilde{w}(t, \nu)\,. \tag{18c}$$

The motivation for a new regressor $w_\circ$ is to capture the fact that the regressor present in the error dynamics need not be the same regressor used to drive parameter adaptation. An example is *error model 4* [16, Ch. 7.5], where the regressor appearing in the plant dynamics must be filtered through an LTI system before it can be used for parameter adaptation.

Next we state technical assumptions on the error model. In brief, Assumption 3 restricts attention to well-behaved regressors via (E0); states relevant structural properties of the dynamics in (E1)-(E3); and imposes nominal stability requirements in (E4)-(E7).

Assumption 3: The regressor w and the closed-loop error model (16)-(18) satisfy:

(E0) regressors $w_\circ$ and w satisfy Assumptions 1-2;

(E1) the PE subspace of $w_\circ$ and w coincide;

(E2) functions $f(\cdot)$ and $g(\cdot)$ are piecewise continuous in t and globally Lipschitz uniformly in t, for $t \geq t_0 \geq 0$. Moreover, they satisfy $f(t, 0, 0) = 0$ and $g(t, 0, 0) = 0$;

(E3) the functions $\tilde{w}_\circ(\cdot)$ and $\tilde{w}(\cdot)$ are piecewise continuous in t and continuous uniformly in t, for $t \geq t_0 \geq 0$. Moreover, they satisfy $\tilde{w}_\circ(t, 0) = \tilde{w}(t, 0) = 0$;

(E4) the equilibrium $\nu = 0$ of (18a) is globally uniformly asymptotically stable (GUAS);

(E5) the equilibrium $\xi = 0$ of $\dot{\xi} = f(t, \xi, 0)$ is GES;

(E6) given any $q \in \mathbb{N}$, $\psi \in \mathbb{R}^q$, and appropriate $w_\circ(t) \in \mathbb{R}^q$ satisfying (E0)-(E1), if $\nu = 0$ and $w(t) \in \mathbb{R}^q$ is PE then the equilibrium $(\xi, \hat{\psi}) = (0, \psi)$ is GES for

$$\begin{aligned} \dot{\xi} &= f(t, \xi, w_\circ^\intercal(t)(\hat{\psi} - \psi)) \\ e &= g(t, \xi, w^\intercal(t)(\hat{\psi} - \psi)) \\ \dot{\hat{\psi}} &= -\gamma e w(t)\,. \end{aligned}$$

(E7) given $q_{pe} \geq 1$, $(w_\circ)_{pe}$, and w_{pe} from the PE decomposition of $w_\circ$ and w, let $\hat{\psi}(t), \psi \in \mathbb{R}^{q_{pe}}$. Then the equilibrium $(\xi, \hat{\psi}) = (0, \psi)$ is GES for

$$\begin{aligned} \dot{\xi} &= f(t, \xi, (w_\circ)_{pe}^\intercal(t)(\hat{\psi} - \psi)) \\ e &= g(t, \xi, w_{pe}^\intercal(t)(\hat{\psi} - \psi)) \\ \dot{\hat{\psi}} &= -\gamma e w_{pe}(t)\,. \end{aligned}$$

▷

Remark 5: The choice of $w_\circ$ for (E6) depends on context. As mentioned, in *error model 4* [16, Ch. 7.5] $w_\circ$ is filtered component-wise through a stable LTI system to obtain the regressor w appearing in the augmented error e. By the Swapping Lemma [17, Lemma 3.6.5], one can show that the PE subspace of $w_\circ$ and w coincide provided (e.g.) the LTI filter is minimum phase. In contrast, $w_\circ = w$ for *error model 3* [16, Ch. 7.4]. Special care is needed to verify that a regressor $w_\circ$ is suitable with respect to (E6). This includes the derived PE regressors $(w_\circ)_{pe}$ and w_{pe} used in Theorems 7-8. ◁

Remark 6: The reader may wonder why (E6) and (E7) are stated separately. First, (E6) is a nominal stability property concerning PE regressors not tied to any PE decomposition. Instead, (E7) says that this stability property should be retained in the reduced system after performing a PE decomposition. Presenting both (E6) and (E7) suggests the possibility to bootstrap from known results concerning PE regressors; indeed, one could remove (E6) and only ask for (E7). ◁

Through an appropriate identification of states, every error model considered in [16, Ch. 7] can be represented as the system (16)-(18) satisfying Assumption 3. An important feature of Assumption 3 is that the only stability properties of the error dynamics that we ask for are (E5)-(E7), both dealing only with the two extreme cases ($w = 0$ and w PE) that are most amenable to analysis. As we show in the proof of Theorem 7, the fact that the PE subspace of $w_\circ$ and w coincide in (E1) allows us to deduce stability properties when w has some non-trivial PE subspace.

B. Robust Design using the μ-modification

The μ-modification takes the form

$$\dot{\hat{\psi}} = -\gamma e\hat{w} - \mu\Omega\Omega^\intercal\hat{\psi}\,, \quad \mu > 0\,, \tag{19}$$

where Ω is the output of any general non-PE subspace estimator, including (7), (11) or the one presented in [2]. Our goal is to show that this modification renders the closed-loop system robust when considering the general error model in Section V-A. Additionally, we want to recover error regulation when the closed-loop system using the μ-modification is unperturbed.

The next two results (with proofs in Section VII) establish robustness to bounded disturbances in every state and measurement of our design. Looking ahead at (20)-(21), there are three key observations. (i) Arbitrarily small perturbations (d_w, d_Σ) will not cause unbounded growth of states regardless of the PE properties of w. This is not the case using (16) without w being PE [1]. (ii) Error regulation is retained when there are no disturbances. This can be seen in the steady-state bound $\|d_2(t)\|_{\mathcal{L}_\infty}\|(d_2, d_e)(t)\|_{\mathcal{L}_\infty} + \|d(t)\|_{\mathcal{L}_\infty}$ implying that if $d = 0$ then all states converge to their equilibrium and $e \to 0$. (iii) While our results require d_Σ to be sufficiently small, we note that (d_ψ, d_Σ) typically are negligible because (20c)-(20d) are generally implemented on a computer. We also show in Theorem 8 that smallness of d_w can be removed provided it introduces sufficient excitation. The following results pertains to the design in Section III; a similar statement could be made for the design in Section IV.

Theorem 7: Consider the system (16)-(18) satisfying Assumption 3. Let q_{pe}, w_{pe}, and $\mathcal{W} = \operatorname{Im}(W)$ result from the PE decomposition of w, and let β_0, $T > 0$ be the PE constants in (5) for the regressor w_{pe}. If $q_{pe} = 0$, set $\beta_0 = \infty$ and $T > 0$ to any finite value. Fix $\varepsilon > 0$, $\sigma_{tol} \in (0, \varepsilon T\beta_0 \mathrm{e}^{-\varepsilon T})$, and consider the subspace estimator (7), (11). Define the perturbed closed-loop error model

$$\dot{\xi} = f(t, \xi, \hat{w}_\circ^\intercal\hat{\psi} - w_\circ^\intercal(t)\psi) + d_\xi \tag{20a}$$
$$e = g(t, \xi, \hat{w}^\intercal\hat{\psi} - w^\intercal(t)\psi) + d_e \tag{20b}$$
$$\dot{\hat{\psi}} = -\gamma e\hat{w} - \mu\Omega\Omega^\intercal\hat{\psi} \qquad + d_\psi \tag{20c}$$
$$\dot{\hat{\Sigma}} = -\varepsilon\hat{\Sigma} + \varepsilon\hat{w}\hat{w}^\intercal \qquad + d_\Sigma \tag{20d}$$

with transients

$$\dot{\nu} = \Delta(t, \nu) \tag{21a}$$
$$\hat{w}_\circ = w_\circ(t) + \tilde{w}_\circ(t, \nu) + d_1 \tag{21b}$$
$$\hat{w} = w(t) + \tilde{w}(t, \nu) \quad + d_2\,, \tag{21c}$$

which replaces (16) with the μ-modification (19) and where $d(t) := (d_\xi, d_e, d_\psi, d_w, d_\Sigma)(t)$ is a bounded disturbance perturbing the system with $d_w := (d_1, d_2)$. Then all states $(\xi, \hat{\psi}, \hat{\Sigma})$ are uniformly bounded and there exist constants $\gamma_\star$, $d_\star > 0$ such that

$$\limsup_{t\to\infty} \|(\xi, \hat{\psi} - WW^\intercal\psi, \hat{\Sigma} - \Sigma_{pe})(t)\|$$
$$\le \gamma_\star\left(\|d_2(t)\|_{\mathcal{L}_\infty}\|(d_2, d_e)(t)\|_{\mathcal{L}_\infty} + \|d(t)\|_{\mathcal{L}_\infty}\right)$$

for all $\|(d_w, d_\Sigma)(t)\|_{\mathcal{L}_\infty} \le d_\star$ sufficiently small, where Σ_{pe} evolves according to (8a) with $\Sigma_{pe}(t_0) = \varepsilon TR_w(0)$. ◇

Remark 7: We do not perturb the dynamics of ν directly since GUAS in (E4) of (18a) and [27, Lemma 9.3] imply that it suffices to place the perturbations in (21b)-(21c). Also, the proof of Theorem 7 assumes continuous differentiability by relying on [27, Theorem 4.14]. This can be relaxed by considering upper Dini derivatives. Finally, notice that $\gamma_\star$ and $d_\star$ depend on various constants not explicitly mentioned and that each $\|\cdot\|_{\mathcal{L}_\infty}$ could be replaced with $\limsup_{t\to\infty}\|\cdot\|$, meaning only steady-state bounds are relevant. ◁

In the next result we redefine the regressors $(w_\circ, w)$ to absorb the disturbances $d_w = (d_1, d_2)$, under the rationale that d_w introduces excitation that can be leveraged for parameter adaptation. The theorem shows that d_w aids in recovering $WW^\intercal\psi$ (rather than ψ). Thus, by viewing large disturbances as a source of excitation, we can derive an analogous result to Theorem 7. Interestingly, the theorem shows that if d_1 and d_2 live in the non-PE subspace $\mathcal{W}^\perp$ and all other disturbances are zero, then there is no steady-state error. This can be seen from the fact that (d_1, d_2) is pre-multiplied by $W^\intercal$ in appropriate locations in the estimates below.

Theorem 8: Consider the perturbed closed-loop error model of Theorem 7. Define the perturbed regressors $\bar{w}_\circ := w_\circ + d_1$ and $\bar{w} := w + d_2$, and suppose that these regressors substituted into the error model satisfy Assumption 3. Let $\bar{q}_{pe} \ge 1$, $\bar{w}_{pe}$, and $\bar{\mathcal{W}} = \operatorname{Im}(\bar{W})$ result from the PE decomposition of $\bar{w}$. If the perturbed regressor $\bar{w}$ satisfies

$$\frac{1}{T}\int_t^{t+T} \bar{w}_{pe}(\tau)\bar{w}_{pe}^\intercal(\tau)\,d\tau \succeq \beta_0 I\,, \quad \forall\, t \ge 0\,,$$

then all states $(\xi, \hat{\psi}, \hat{\Sigma})$ are uniformly bounded and there exist constants $\gamma_\star(d_w)$, $d_\star > 0$ such that

$$\limsup_{t\to\infty} \|(\xi, \hat{\psi} - \bar{W}\bar{W}^\intercal WW^\intercal\psi, \hat{\Sigma} - \bar{\Sigma}_{pe})(t)\|$$
$$\le \gamma_\star\left(\|d_2(t)\|_{\mathcal{L}_\infty}\|(W^\intercal d_2, d_e)(t)\|_{\mathcal{L}_\infty} + \|d_{pe}(t)\|_{\mathcal{L}_\infty}\right)$$

for all $\|d_\Sigma(t)\|_{\mathcal{L}_\infty} \le d_\star$ sufficiently small, where

$$d_{pe} := (d_\xi, d_e, d_\psi, W^\intercal d_1, W^\intercal d_2, d_\Sigma)$$

and $\bar{\Sigma}_{pe}$ evolves according to

$$\dot{\bar{\Sigma}}_{pe} = -\varepsilon\bar{\Sigma}_{pe} + \varepsilon\bar{W}\bar{w}_{pe}(t)\bar{w}_{pe}^\intercal(t)\bar{W}^\intercal$$

with $\bar{\Sigma}_{pe}(t_0) = \varepsilon T R_{\bar{w}}(0)$. ◇

VI. Applications

To apply the μ-modification with our new subspace estimator, the designer must select three parameters:

1) μ, determining the leakage rate of $\hat{\psi}_\perp$;
2) ε, affecting how well $\mathcal{W}^\perp$ can be recovered;
3) σ_{tol}, setting a soft excitation threshold for w.

To showcase the effectiveness of our design, we first consider the simplest case of linear regression, known as *error model 1* [16, Ch. 7.2]. Afterwards, we solve the regulator problem for a known plant and unknown exosystem using state feedback. This second application is related to *error model 2* [16, Ch. 7.3] and focuses on the effects of transients arising from the construction of an internal model to estimate the exogenous regressor. In a companion paper [28], we solve the regulator problem for a known plant and unknown exosystem with output feedback and using the subspace estimator in [2]; this context highlights the relevance of (E1).

A. Error Model 1 - Linear Regression

The linear regression model corresponds to an error

$$e = \mathscr{E}[\hat{\psi}, w, \nu] = \hat{w}^\intercal\hat{\psi} - w^\intercal(t)\psi\,. \tag{22}$$

Suppose $\psi = \begin{bmatrix}1 & 1\end{bmatrix}^\intercal$ and $w(t) = \begin{bmatrix}\sin(t) & 0.8(t^2+1)^{-1}\end{bmatrix}^\intercal$. Its PE decomposition is

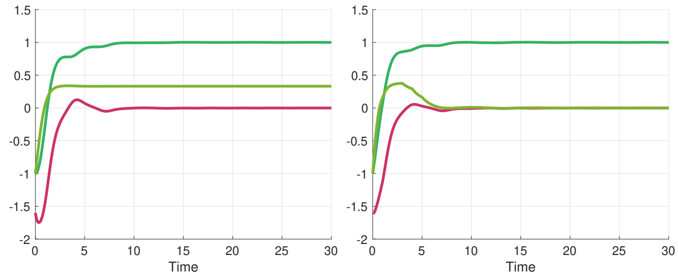
$$w(t) = Ww_{pe}(t) + W_\perp w_\perp(t) = \begin{bmatrix}1\\0\end{bmatrix}\sin(t) + \begin{bmatrix}0\\1\end{bmatrix}\frac{0.8}{t^2+1}\,.$$

One can verify this error model satisfies Assumption 3, where the ξ dynamics are vacuous, and there is no transient state ν if we take $\hat{w} = w$. Note that

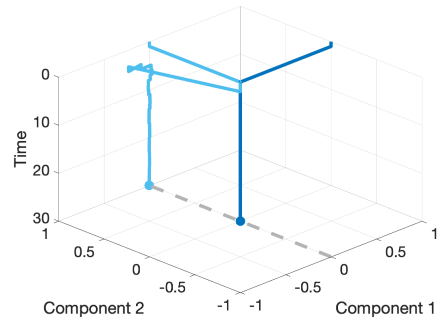
$$\frac{1}{\pi}\int_t^{t+\pi} w_{pe}(\tau)w_{pe}^\intercal(\tau)\,d\tau = \frac{1}{\pi}\int_t^{t+\pi}\sin^2(\tau)\,d\tau = 0.5$$

so $\beta_0 = 0.5$ and $T = \pi$. Figure 1 simulates (22) and (16) with and without the μ-modification (19). For the μ-modification, we employ the subspace estimator (7), (11) using the threshold $\sigma_{tol} = 0.3\varepsilon T\mathrm{e}^{-\varepsilon T}$ and update time $t_{i+1} - t_i = 1$. We set $\gamma = \mu = \varepsilon = 1$, $\hat{\psi}(t_0) = -\psi$, and $\hat{\Sigma}(t_0) = 0$.

From Figure 1c, one observes that $\mathcal{W}^\perp = \mathrm{Im}(W_\perp) = \mathrm{Im}(\begin{bmatrix}0 & 1\end{bmatrix}^\intercal)$, as expected from the PE decomposition of w. The PE dynamics consist of $\hat{\psi}_{pe} = \hat{\psi}_1 = W^\intercal\hat{\psi}$ (green) and the non-PE dynamics consist of $\hat{\psi}_\perp = \hat{\psi}_2 = W_\perp^\intercal\hat{\psi}$ (olive green). Using the standard gradient algorithm without the μ-modification, Figure 1a shows that only $\hat{\psi}_1$ converges to $\psi_1 = 1$, whereas $\hat{\psi}_2$ is temporarily driven by $w_\perp$ without converging to its true value $\psi_2 = 1$. Regardless, the error e is regulated to zero. With the μ-modification on in Figure 1b, the leakage term drives the non-PE dynamics $\hat{\psi}_2$ to zero, whereas

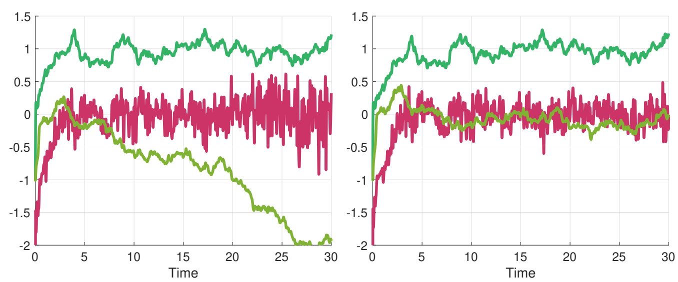


(a) μ-modification off (b) μ-modification on

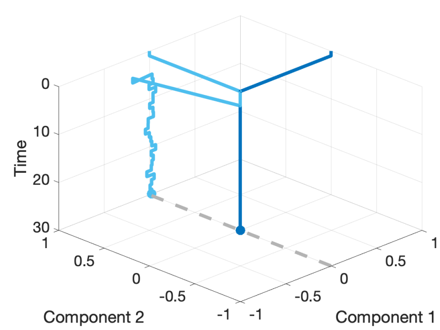


(c) Subspace estimator

Fig. 1: Comparison of the error (red) and parameter adaptation dynamics (green, olive green) for error model 1 as well as a plot of the columns of the subspace estimator (blue, light blue) and the non-PE subspace (grey dashed line).



(a) μ-modification off (b) μ-modification on

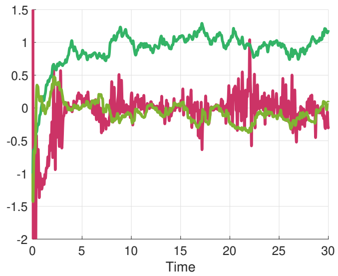


(c) Subspace estimator

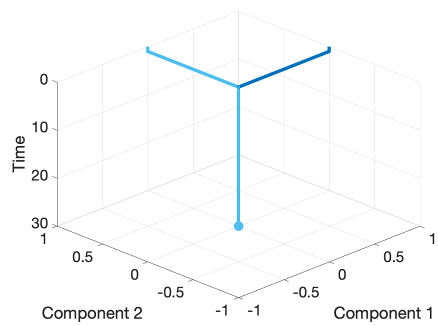
Fig. 2: Comparison of the noisy error (red) and parameter adaptation dynamics (green, olive green) for error model 1 as well as a plot of the columns of the subspace estimator (blue, light blue) and the non-PE subspace (grey dashed line).

the other dynamics remain relatively unchanged. This agrees with Theorem 7 in the case when $d = 0$.

(a) μ-modification off



(b) μ-modification on



(c) Subspace estimator

Fig. 3: Comparison of the noisy error (red) and parameter adaptation dynamics (green, olive green) for error model 1 in the presence of large noise in the regressor as well as a plot of the columns of the subspace estimator (blue, light blue). Notice the non-PE subspace is $\{0\}$.

Next, we investigate robustness in the presence of noise. Consider the perturbed closed-loop error model (20)-(21), where we ignore d_ξ and set $d_w = d_1 = d_2$. Suppose we inject the following white Gaussian noise:

$$d_e,\, (d_\psi)_1 \sim \mathcal{N}(0, 0.5^2)\,, \qquad (d_\psi)_2 \sim \mathcal{N}(-0.05, 0.5^2)\,,$$
$$(d_w)_i \sim \mathcal{N}(0, 0.15^2)\,, \quad (d_\Sigma)_{ij} \sim \mathcal{N}(0, 0.05^2)\,,$$

where $i,\, j \in \{1, 2\}$. Simulation results are found in Figure 2. Note that we plot the error e without the disturbance d_e, since from the perspective of error regulation we are not interested in the measured error with $d_e \neq 0$. Without the μ-modification, we observe in Figure 2a that the parameter $\hat{\psi}_2$ exhibits a negative linear growth; that is, unstable behaviour. In contrast, Figure 2b shows that $\hat{\psi}_2$ converges near zero, illustrating how parameter adaptation dynamics are rendered robust using the μ-modification in the presence of noise. Furthermore, looking at the subspace estimator in Figure 2c we note that it too is robust to noise as it still closely recovers the non-PE subspace. These observations also agree with Theorem 7.

Lastly, we investigate the effect of a large disturbance d_w. Leaving the other noise profiles the same, we now assume:

$$(d_w)_1 \sim \mathcal{N}(0, 0^2)\,, \quad (d_w)_2 \sim \mathcal{N}(0, 1^2)\,.$$

By Figure 3c it is clear that d_w introduces sufficient excitation so that the conditions of Theorem 8 are satisfied. In particular, given that our subspace estimator vanishes, we deduce that $w + d_w$ is now PE, despite w not being PE. As observed in Figure 3 and expected from Theorem 8, we have robustness of all states in the presence of noise, including large d_w. Moreover, we notice that we have robustness both with and without the μ-modification. This fact is not a coincidence but a consequence of Theorem 8, which can be explored as future work.

B. Error Model 2 - Regulation using State Feedback

Consider the single-input LTI system

$$\dot{x} = Ax + B\,(u - \Gamma\zeta) \tag{23a}$$
$$\dot{\zeta} = S\zeta\,, \tag{23b}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}$ is the control input, and $\zeta(t) \in \mathbb{R}^q$ is the exosystem state. Define the disturbance to reject $d := \Gamma\zeta$. Our goal is to regulate the state x to 0 while rendering the closed-loop dynamics exponentially stable. In other words, we consider a disturbance rejection problem. As is standard in regulator theory, we assume the following.

Assumption 4: The system (23) satisfies:

(A1) the pair (A, B) is known and controllable;
(A2) the matrix S only has simple eigenvalues on the $j\omega$-axis;
(A3) wlog the pair (Γ, S) is observable;
(A4) the dimension q is interpreted as a known upper bound on the exosystem order;
(A5) the measurement is x. ▷

Following the regulator design in [29, Ch. 4.1.2], we construct a regulator of the form

$$u = u_s + u_{im}\,, \tag{24}$$

where u_s is for closed-loop stability and u_{im} is for disturbance rejection. The stabilization piece is a pole placement controller $u_s = K^\intercal x$, where $K \in \mathbb{R}^n$ is selected such that $A_{cl} := A + BK^\intercal$ is Hurwitz. To perform disturbance rejection, select a controllable pair $(F, G) \in \mathbb{R}^{q\times q} \times \mathbb{R}^q$ with F Hurwitz. Then there exists a new exosystem state $w := M\zeta$ and an unknown parameter $\psi^\intercal := \Gamma M^{-1}$ yielding the exosystem

$$\dot{w} = Fw + Gd$$
$$d = \psi^\intercal w\,,$$

for some $M \in \mathbb{R}^{q\times q}$. Then the disturbance rejection piece $u_{im} = \hat{\psi}^\intercal \hat{w}$ consists of two parts. First, an internal model

$$\dot{\eta} = F\eta + (FN - NA)x - NBu \tag{25a}$$
$$\hat{w} = \eta + Nx\,, \tag{25b}$$

where N satisfies $NB = -G$. Second, an adaptive law

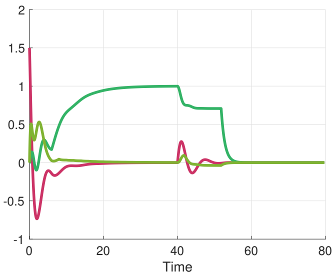
$$\dot{\hat{\psi}} = -\gamma(B^\intercal Px)\hat{w}\,, \tag{26}$$

where we select $\gamma > 0$ and $P \succ 0$ solves the Lyapunov equation $A_{cl}^\intercal P + PA_{cl} = -I$.

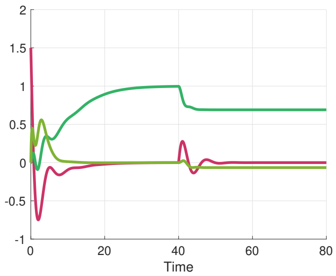
Define $\tilde{w} := \hat{w} - w$. The resulting closed-loop dynamics are

$$\dot{x} = A_{cl}x + B(\hat{w}^\intercal\hat{\psi} - w^\intercal(t)\psi) \tag{27a}$$
$$\dot{\hat{\psi}} = -\gamma(B^\intercal Px)\hat{w} \tag{27b}$$
$$\dot{\tilde{w}} = F\tilde{w}\,. \tag{27c}$$

To map (27) back to the error model in Section V-A, identify the error state as $\xi = x$, the error signal $e = B^\intercal Px$, and the transient state $\nu = \tilde{w}$. Since w is the state of an LTI exosystem,

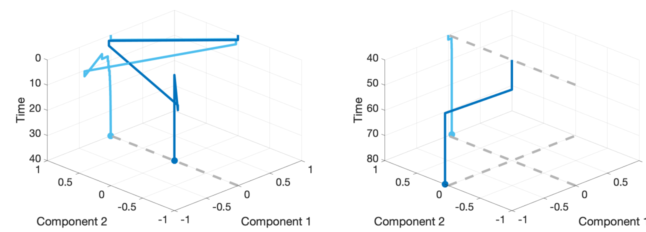


(a) New subspace estimator

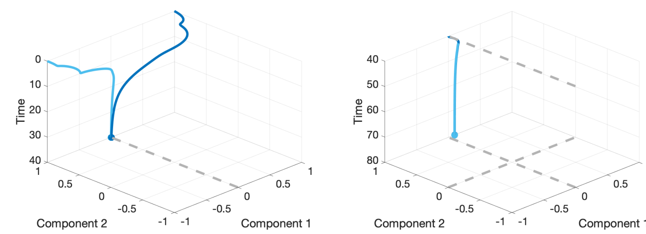


(b) Old subspace estimator

Fig. 4: Comparison of the error (red) and parameter adaptation dynamics (green, olive green) for error model 2.



(a) New subspace estimator



(b) Old subspace estimator

Fig. 5: Columns of our subspace estimators (blue, light blue) and a plot of non-PE subspaces (grey dashed lines).

it is an almost periodic signal and, by Proposition 3, the PE decomposition of w is $w = Ww_{pe}$. Note that $w_\circ = w$. Also, the nominal stability requirements (E5)-(E7) are well known facts of adaptive control.

We simulate system (23) with matrices

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \Gamma = 1, \quad S = 0,$$

and initial conditions $x(t_0) = \begin{bmatrix} 1 & 1 \end{bmatrix}^\intercal$ and $\zeta(t_0) = 0.5$. Moreover, at $t = 40$ we turn off the exosystem by setting $\Gamma = 0$. The regulator (24)-(26) is built with values $N = -GB^\intercal$,

$$F = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad K = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \gamma = 2,$$

and initial conditions $\eta(t_0) = 0$ and $\hat{\psi}(t_0) = 0$. Note that our internal model is of order $q = 2$, whereas the exosystem is of order strictly less than q. Therefore, parameter adaptation will not be robust and we must employ the μ-modification. Here we compare, without introducing noise, our proposed subspace estimator with our previous subspace estimator [2].

When using our new subspace estimator (7), (11) we set parameters $\mu = 1$ and $\varepsilon = \sigma_{tol} = 0.1$, update time $t_{i+1} - t_i = 2$, and initial condition $\hat{\Sigma}(t_0) = 0$. On the other hand, recall our old subspace estimator

$$\dot{v} = -\varepsilon \hat{w}\hat{w}^\intercal v + \varepsilon \sigma_{tol}(1 - \|v\|^2)v$$

from [2], where $v(t) \in \mathbb{R}^q$ denotes any column of $\Omega(t) \in \mathbb{R}^{q \times q}$. For the latter subspace estimator, we let $\varepsilon = 0.75$ and select the (full rank) initial condition

$$\Omega(t_0) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

while other parameters remain the same as above. We do not verify that σ_{tol} is sufficiently small but instead treat it as a design specification for the expected minimum excitation threshold for w.

In Figure 4 we compare the closed-loop dynamics resulting from employing both subspace estimators. First of all, we observe that asymptotic error regulation is always achieved. In contrast, the parameter adaptation dynamics differ during the window $t \in [40, 80]$. Notice that during $t \in [0, 40]$ the disturbance d is a constant, meaning that the PE subspace of w is 1-dimensional. One can show $\mathcal{W} = \mathrm{Im}(\begin{bmatrix} 1 & 0 \end{bmatrix}^\intercal)$ and $\mathcal{W}^\perp = \mathrm{Im}(\begin{bmatrix} 0 & 1 \end{bmatrix}^\intercal)$. We see in Figure 5 that by $t = 40$ the columns of both subspace estimators align, and each spans the non-PE subspace. Returning to Figure 4, the component $\hat{\psi}_1$ is driven to its true value of $\psi_1 = 1$ while the μ-modification drives $\hat{\psi}_2$ to 0. When the exosystem is suddenly turned off for $t \in [40, 80]$, there is no excitation to drive parameter adaptation, so the non-PE subspace is all of $\mathbb{R}^2$, and we expect $\hat{\psi} \to 0$. In Figure 5a we see that our new subspace estimator recovers all of $\mathbb{R}^2$, but in Figure 5b our old one fails to do so. This follows from the fact that $\Omega(40)$ in Figure 5b no longer appears to be full rank given that both of its columns align, which is needed for correct functioning of the subspace estimator in [2]. Consequently, the μ-modification with the old subspace estimator fails to forget unexcited parameters to provide robustness during $t \in [40, 80]$ as seen in Figure 4b. Altogether, we have illustrated that our new subspace estimator provides the advantage of being live in suddenly changing environments given that its convergence properties are independent of its initial condition.

VII. Proofs

A. Proof of Proposition 3

Since w is almost periodic, so is $ww^\intercal$. By [30, Appendix, Theorem 6] w satisfies Assumption 1. Then by Proposition 1, w has a PE decomposition given by $w = Ww_{pe} + W_\perp w_\perp$. We want to show $w_\perp = W_\perp^\intercal w = 0$. For the sake of contradiction, suppose $W_\perp^\intercal w \neq 0$. Then there exists a non-zero $v \in \mathrm{Im}(W_\perp)$ such that $f(t) := (v^\intercal w(t))^2 \geq 0$ is not identically 0. As a result, there are some constants ϵ, $t_\star > 0$ such that $f(t_\star) \geq 3\epsilon$. Again since $w(\cdot)$ is almost periodic, so is $f(\cdot)$. Recalling that an almost periodic function is continuous, we have that for all $\delta > 0$ sufficiently small, $f(t) \geq 2\epsilon$ for $t \in (t_\star - \delta, t_\star + \delta)$. By definition of almost periodicity in [30, Appendix, Definition 5], the set $T(f, \epsilon) := \{\tau : |f(t + \tau) - f(t)| < \epsilon \ \forall \, t \in \mathbb{R}\}$ is relatively dense. Thus there exists $L > 0$ such that $[\tau, \tau + L] \cap T(f, \epsilon) \neq \emptyset$ for all $\tau \geq 0$.

Now partition $\mathbb{R}$ into intervals of length $3L$ and position them such that $t_\star$ is in the middle third of any one of the intervals. Wlog we assume that $\delta < L$. By relative density of $T(f,\epsilon)$, we can construct a strictly increasing sequence $\{\tau_i\}_{i=0}^{\infty}$ with $\tau_0 = t_\star$, where each τ_{i+1} lies in the middle third of each successive interval, and with $\tau_i - \tau_0 \in T(f,\epsilon)$. By construction of the τ_i and the definition of $T(f,\epsilon)$, we have that $f(t) \geq \epsilon$ for all $t \in (\tau_i - \delta, \tau_i + \delta)$ and for all i. Thus the integral of $f(\cdot) \geq 0$ over every interval of length $3L$ must be at least $\epsilon \cdot 2\delta$. As a result, one has

$$\begin{aligned} v^\intercal R_w(0)v &= \lim_{T\to\infty} \frac{1}{T}\int_{t_0}^{t_0+T} v^\intercal w(\tau)w^\intercal(t)v\, d\tau \\ &= \lim_{T\to\infty} \frac{1}{T}\int_{t_0}^{t_0+T} f(\tau)\, d\tau \geq \frac{2\epsilon\delta}{3L} > 0\,. \end{aligned}$$

But $v \in \mathrm{Im}(W_\perp) = \mathrm{Ker}(R_w(0))$ so that $v^\intercal R_w(0)v = 0$, which is a contradiction. □

B. Proof of Lemma 1

Let $W_\perp$ also result from the PE decomposition of w. By pre- and post-multiplying (8a) by $W_\perp$, we have

$$W_\perp^\intercal \dot{\Sigma}_{pe} = -\varepsilon W_\perp^\intercal \Sigma_{pe} + 0\,, \quad W_\perp^\intercal \Sigma_{pe}(t_0) = 0 \tag{28a}$$

$$\dot{\Sigma}_{pe} W_\perp = -\varepsilon \Sigma_{pe} W_\perp + 0\,, \quad \Sigma_{pe}(t_0) W_\perp = 0 \tag{28b}$$

since $\mathrm{Im}(W_\perp) = \mathrm{Ker}(R_w(0)) = \mathrm{Im}(W)^\perp$. Thus $W_\perp^\intercal \Sigma_{pe}$ and $\Sigma_{pe} W_\perp$ are identically zero. Recalling that $I = WW^\intercal + W_\perp W_\perp^\intercal$ by orthogonality, one has that $\Sigma_{pe}(t) = WW^\intercal \Sigma_{pe}(t) WW^\intercal$. Therefore, it suffices to show that $\Lambda_{pe}(t) := W^\intercal \Sigma_{pe}(t) W$ satisfies the conclusions of Lemma 1. Taking the time derivative, the Λ_{pe} dynamics are

$$\begin{aligned} \dot{\Lambda}_{pe} &= -\varepsilon \Lambda_{pe} + \varepsilon w_{pe}(t) w_{pe}^\intercal(t) \\ \Lambda_{pe}(t_0) &= \varepsilon T R_{w_{pe}}(0)\,. \end{aligned} \tag{29}$$

Given that $\varepsilon > 0$, w_{pe} is bounded by Assumption 1, and $\Lambda_{pe}(t_0)$ is symmetric, it is clear that Λ_{pe} is bounded and symmetric. We will show that $\Lambda_{pe}(t) \succeq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T} I$ for all $t \geq t_0 \geq 0$ inductively, effectively employing the proof of [7, Lemma 2.1]. Let $v \in \mathbb{R}^{q_{pe}}$ be any unit vector and partition

$$[t_0, \infty) = \bigcup_{k=1}^{\infty} [t_0 + (k-1)T, t_0 + kT)\,.$$

Base Case: First, we show the bound holds for $t \in [t_0, t_0 + T)$ (i.e., $k = 1$). Recall that $R_{w_{pe}}(0) \succeq \beta_0 I$. Since the system (29) is linear and all terms involved are positive, we have that the time solution satisfies

$$\begin{aligned} v^\intercal \Lambda_{pe}(t) v &= v^\intercal \left(\mathrm{e}^{-\varepsilon(t-t_0)} \Lambda_{pe}(t_0) \right) v \\ &\quad + v^\intercal \left(\varepsilon \int_{t_0}^{t} \mathrm{e}^{-\varepsilon(t-\tau)} w_{pe}(\tau) w_{pe}^\intercal(\tau)\, d\tau \right) v \\ &\geq \mathrm{e}^{-\varepsilon T} v^\intercal \Lambda_{pe}(t_0) v \geq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T}\,. \end{aligned}$$

Induction Step: Suppose that the bound holds for $t \in [t_0, t_0 + kT)$, we want to show it holds for $t \in [t_0 + kT, t_0 + (k+1)T)$. Let $t = \Delta t + T$ such that $\Delta t \in [t_0, t_0 + kT)$, then one can show

$$\begin{aligned} &\Lambda_{pe}(\Delta t + T) = \\ &\mathrm{e}^{-\varepsilon T} \left(\Lambda_{pe}(\Delta t) + \varepsilon \int_{\Delta t}^{\Delta t + T} \mathrm{e}^{\varepsilon(\tau - \Delta t)} w_{pe}(\tau) w_{pe}^\intercal(\tau)\, d\tau \right)\,. \end{aligned}$$

By the fact that $\Lambda_{pe}(\Delta t) \succeq 0$ from the induction hypothesis, and using both the PE constants for w_{pe} as well as the fact that $\mathrm{e}^{\varepsilon(\tau - \Delta t)} \geq 1$ for $\tau \in [\Delta t, \Delta t + T)$ we get

$$\begin{aligned} &v^\intercal \Lambda_{pe}(\Delta t + T) v \\ &\geq \varepsilon T \mathrm{e}^{-\varepsilon T} \frac{1}{T} \int_{\Delta t}^{\Delta t + T} \mathrm{e}^{\varepsilon(\tau - \Delta t)} (v^\intercal w_{pe}(\tau))^2\, d\tau \geq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T} \end{aligned}$$

for all $\Delta t \in [t_0, t_0 + kT)$. This implies $\Lambda_{pe}(t) \succeq \varepsilon T \beta_0 \mathrm{e}^{-\varepsilon T} I$ for $t \in [t_0 + T, t_0 + (k+1)T)$, proving the result. □

C. Proof of Theorem 5

The proof follows the same procedure as that in Theorem 3, with a few changes. If $q_{pe} = 0$, then once again $\Sigma_{pe} = 0$ and the proof follows verbatim Theorem 3. Now suppose $q_{pe} \geq 1$. Using Assumption 1, let $\delta_{pe}^w(\cdot)$ and $\delta_\perp^w(\cdot)$ be class-$\mathcal{K}$ functions and $\varepsilon_1(w) > 0$ be a constant provided by Lemma 4. Then, as we have shown in (15), we have results similar to Lemma 1 and Corollary 2 but with

$$\Lambda_{pe}(t) \succeq (\beta_0 - \delta_{pe}^w(\varepsilon)) I\,, \quad \sigma_{\min}(D_3(t)) \geq \beta_0 - \delta_{pe}^w(\varepsilon)$$

for all $\varepsilon \in (0, \varepsilon_1]$ and $t \geq t_0 \geq 0$. Then Lemma 3 follows but with the changes

$$\begin{aligned} \sigma_{\min}(D_1) &\geq \beta_0 - \delta_{pe}^w(\varepsilon) - \|S - \Sigma_{pe}(t)\| \\ \sigma_{\max}(D_2) &\leq \|S - \Sigma_{pe}(t)\| \end{aligned}$$

for all $\varepsilon \in (0, \varepsilon_1]$ and $t \geq t_0 \geq 0$. Using the facts $\sigma_{tol} < \beta_0$ and $\delta_{pe}^w(\cdot)$ is a class-$\mathcal{K}$ function, we can select $\varepsilon_\star(w, \sigma_{tol}) > 0$ no larger than ε_1 such that

$$\beta_0 - \delta_{pe}^w(\varepsilon_\star) > \sigma_{tol}\,.$$

Moreover, notice that $\beta_0 - \delta_{pe}^w(\varepsilon) \geq \beta_0 - \delta_{pe}^w(\varepsilon_\star) > \sigma_{tol}$ for all $\varepsilon \in (0, \varepsilon_\star]$. Thus there exists $\delta_{tol}(w, \sigma_{tol}) > 0$ such that for all $\|S - \Sigma_{pe}(t)\| \leq \delta_{tol}$ and $\varepsilon \in (0, \varepsilon_\star]$ we have

$$\sigma_{\min}(D_1) \geq \sigma_{tol}\,, \quad \sigma_{\max}(D_2) < \sigma_{tol}\,;$$

that is, $\Omega = U(I - \mathrm{diag}(I, 0)) = \begin{bmatrix} 0 & U_2 \end{bmatrix}$. At this point, one can directly follow the same procedure as the proof of Theorem 3 involving Wedin's Theorem. □

D. Proof of Theorem 6

The proof follows the same procedure as that in Theorem 4, with a few changes. To prove the first limit, apply Theorem 5 with $S = \hat{\Sigma}(t_i)$ for each t_i. In particular, recalling that $\hat{\Sigma} = \Sigma_{pe} + \Sigma_\perp + \tilde{\Sigma}$, we use the fact that

$$\|\hat{\Sigma}(t) - \Sigma_{pe}(t)\| \leq \|\Sigma_\perp(t)\| + \|\tilde{\Sigma}(t)\| \leq \delta_\perp^w(\varepsilon) + \|\tilde{\Sigma}(t)\|$$

for all $\varepsilon \in (0, \varepsilon_\star]$ and $t \geq t_0 \geq 0$, with $\tilde{\Sigma} \to 0$ because $\hat{w} \to w$. To prove the third limit, note that if Assumption 2 holds then it is wlog to assume $\delta_\perp^w(\varepsilon) = 0$ by absorbing $\Delta_\perp$

into $\tilde{\Delta}$; that is, consider (8) rather than (12). Finally, to prove the second limit, let $\varepsilon_{\star\star}(w,\sigma_{tol}) > 0$ be such that

$$\beta_0 - \delta^w_{pe}(\varepsilon_{\star\star}) - \delta^w_{\perp}(\varepsilon_{\star\star}) > \sigma_{tol}\,, \quad \delta^w_{\perp}(\varepsilon_{\star\star}) < \sigma_{tol}\,.$$

Then, for $q_{pe} \geq 1$, substituting $\hat{\Sigma} - \Sigma_{pe} = \Sigma_\perp + \tilde{\Sigma}$ we get

$$\begin{aligned}\sigma_{\min}(D_1(t)) &\geq \beta_0 - \delta^w_{pe}(\varepsilon) - \delta^w_{\perp}(\varepsilon) - \|\tilde{\Sigma}(t)\|\\ \sigma_{\max}(D_2(t)) &\leq \delta^w_{\perp}(\varepsilon) + \|\tilde{\Sigma}(t)\|\end{aligned}$$

for all $\varepsilon \in (0, \varepsilon_{\star\star}]$ and $t \geq t_0 \geq 0$, where $\tilde{\Sigma} \to 0$ when $\hat{w} \to w$. At this point, one can directly follow the same procedure as the proof of Theorem 4. □

E. Proof of Theorem 7

We only prove the case when $1 \leq q_{pe} < q$ as the other cases follow by specialization of the proof. There are four main steps of the proof. First, by (E0)-(E1) there exists PE decompositions

$$w_\circ = W(w_\circ)_{pe} + W_\perp(w_\circ)_\perp\,, \quad w = Ww_{pe} + W_\perp w_\perp$$

where $(w_\circ)_\perp$, $w_\perp \to 0$. Define $\tilde{\Omega} := \Omega\Omega^\intercal - W_\perp W_\perp^\intercal$ and the coordinate transformation

$$\begin{bmatrix}\tilde{\psi}_{pe}\\ \hat{\psi}_\perp\end{bmatrix} = \begin{bmatrix}W^\intercal\\ W_\perp^\intercal\end{bmatrix}(\hat{\psi} - WW^\intercal\psi)\,.$$

Second, the perturbed closed-loop error model becomes

$$\begin{aligned}\dot{\xi} &= f(t,\xi,(w_\circ)^\intercal_{pe}\tilde{\psi}_{pe}) + p_1(\cdot) + d_\xi\\ \dot{\tilde{\psi}}_{pe} &= -\gamma e_\circ w_{pe} \qquad\qquad + p_2(\cdot) + W^\intercal d_\psi\\ \dot{\hat{\psi}}_\perp &= -\mu\hat{\psi}_\perp \qquad\qquad\quad + p_3(\cdot) + W_\perp^\intercal d_\psi\end{aligned}$$

where $e_\circ := g(t,\xi,w_{pe}^\intercal\tilde{\psi}_{pe})$ and

$$\begin{aligned}p_1(\cdot) &= f(t,\xi,\hat{w}_\circ^\intercal\hat{\psi} - w_\circ^\intercal\psi) - f(t,\xi,(w_\circ)^\intercal_{pe}\tilde{\psi}_{pe})\\ p_2(\cdot) &= -\gamma(e-e_\circ)w_{pe} - \gamma eW^\intercal(\hat{w}-w)\\ &\quad - \mu W^\intercal\tilde{\Omega}(W\tilde{\psi}_{pe} + W_\perp\hat{\psi}_\perp + WW^\intercal\psi)\\ p_3(\cdot) &= -\gamma e(W_\perp^\intercal(\hat{w}-w) + w_\perp)\\ &\quad - \mu W_\perp^\intercal\tilde{\Omega}(W\tilde{\psi}_{pe} + W_\perp\hat{\psi}_\perp + WW^\intercal\psi)\,.\end{aligned}$$

Noting that we can write $\hat{w}^\intercal\hat{\psi} - w^\intercal\psi$ as

$$\begin{aligned}&w_{pe}^\intercal\tilde{\psi}_{pe} + (\hat{w}-w)^\intercal W\tilde{\psi}_{pe} + (w_\perp + W_\perp^\intercal(\hat{w}-w))^\intercal\hat{\psi}_\perp\\ &\quad + (-W_\perp w_\perp + WW^\intercal(\hat{w}-w))^\intercal\psi\end{aligned}$$

and synonymously for $\hat{w}_\circ^\intercal\hat{\psi} - w_\circ^\intercal\psi$, we may use the global Lipschitz property from (E2) to obtain the bounds

$$\begin{aligned}\|p_1(\cdot)\| &\leq \ell_f(\|(w_\circ)_\perp\| + \|\hat{w}_\circ - w_\circ\|)(\|(\tilde{\psi}_{pe},\hat{\psi}_\perp)\| + \|\psi\|)\\ \|e-e_\circ\| &\leq \ell_g(\|w_\perp\| + \|\hat{w}-w\|)(\|(\tilde{\psi}_{pe},\hat{\psi}_\perp)\| + \|\psi\|)\\ &\quad + \|d_e\|\\ \|e\| &\leq \ell_g(\|w_\perp\| + \|\hat{w}-w\|)(\|(\tilde{\psi}_{pe},\hat{\psi}_\perp)\| + \|\psi\|)\\ &\quad + \|d_e\| + \ell_g(1 + \|w(t)\|_{\mathcal{L}_\infty})\|(\xi,\tilde{\psi}_{pe})\|\end{aligned}$$

where ℓ_f, $\ell_g \geq 0$ are global Lipschitz constants for $f(\cdot)$, $g(\cdot)$. Defining $\eta = (\eta_1,\eta_2)$ where

$$\eta_1 := \|(w_\circ)_\perp\| + \|\hat{w}_\circ - w_\circ\|\,, \quad \eta_2 := \|w_\perp\| + \|\hat{w}-w\|$$

we can derive the bounds

$$\begin{aligned}\|p_1(\cdot)\| &\leq \ell_f\eta_1(\|(\tilde{\psi}_{pe},\hat{\psi}_\perp)\| + \|\psi\|)\\ \|(p_2,p_3)(\cdot)\| &\leq \gamma\ell_g(1 + \|w(t)\|_{\mathcal{L}_\infty})\eta_2\|(\xi,\tilde{\psi}_{pe})\|\\ &\quad + \gamma(\|w(t)\|_{\mathcal{L}_\infty} + \eta_2)(\ell_g\eta_2(\|(\tilde{\psi}_{pe},\hat{\psi}_\perp)\| + \|\psi\|) + \|d_e\|)\\ &\quad + \mu\|\tilde{\Omega}\|(\|(\tilde{\psi}_{pe},\hat{\psi}_\perp)\| + \|\psi\|)\,.\end{aligned}$$

By defining the aggregate state $\chi := (\xi,\tilde{\psi}_{pe},\hat{\psi}_\perp)$ we can compactly write the perturbed closed-loop system as

$$\dot{\chi} = F(t,\chi) + P(\cdot)$$

where

$$F(t,\chi) = \begin{bmatrix}f(t,\xi,(w_\circ)^\intercal_{pe}(t)\tilde{\psi}_{pe})\\ -\gamma e_\circ w_{pe}(t)\\ -\mu\hat{\psi}_\perp\end{bmatrix}$$

and

$$\begin{aligned}\|P(\cdot)\| &\leq \gamma_1(\|\eta\| + \eta_2^2 + \|\tilde{\Omega}\|)\|\chi\|\\ &\quad + \gamma_1(\|\eta\| + \eta_2^2 + \|\tilde{\Omega}\| + \eta_2\|d_e\| + \|(d_\xi,d_e,d_\psi)\|)\\ &=: P_1(\eta,\tilde{\Omega})\|\chi\| + P_2(\eta,\tilde{\Omega},d)\end{aligned}$$

for some constant $\gamma_1 > 0$ (whose dependence on other constants we do not track). We introduce further constants $\gamma_i > 0$ below, as needed.

Third, to carry out a Lyapunov analysis to prove the desired result, we need to establish steady-state bounds on η and $\tilde{\Omega}$. Using $(w_\circ)_\perp$, $w_\perp \to 0$ by (E0), $\tilde{w}_\circ(\cdot)$, $\tilde{w}(\cdot)$ are continuous uniformly in t with $\tilde{w}_\circ(t,0) = \tilde{w}(t,0) = 0$ by (E3), and $\nu \to 0$ by (E4), we have

$$\limsup_{t\to\infty}\eta_i(t) \leq \|d_i(t)\|_{\mathcal{L}_\infty}\,, \quad \limsup_{t\to\infty}\|\eta(t)\| \leq \|d_w(t)\|_{\mathcal{L}_\infty}$$

by the fact that $\limsup_{t\to\infty}\|\cdot\| \leq \|\cdot\|_{\mathcal{L}_\infty}$. Next, letting $\Delta_\perp$ and $\tilde{\Delta}$ be as defined in Section III-A and using the fact that w is bounded (in conjunction with the aforementioned properties), we have

$$\begin{aligned}&\limsup_{t\to\infty}\|\Delta_\perp(t)\| = 0\\ &\quad\limsup_{t\to\infty}\|\tilde{\Delta}(t)\| \leq \gamma_2(1 + \|d_2(t)\|_{\mathcal{L}_\infty})\|d_2(t)\|_{\mathcal{L}_\infty}\,.\end{aligned}$$

Letting $\tilde{\Sigma} := \hat{\Sigma} - \Sigma_{pe}$ we have

$$\dot{\tilde{\Sigma}} = -\varepsilon\tilde{\Sigma} + \varepsilon(\Delta_\perp(t) + \tilde{\Delta}(t)) + d_\Sigma\,.$$

Using [31, Lemma 2] and the Comparison Lemma we have

$$\begin{aligned}&\limsup_{t\to\infty}\|\hat{\Sigma}(t) - \Sigma_{pe}(t)\|\\ &\qquad\leq \limsup_{t\to\infty}\|\tilde{\Delta}(t)\| + \limsup_{t\to\infty}\varepsilon^{-1}\|d_\Sigma(t)\|\\ &\qquad\leq \gamma_2(1 + \|d_2(t)\|_{\mathcal{L}_\infty})\|d_2(t)\|_{\mathcal{L}_\infty} + \varepsilon^{-1}\|d_\Sigma(t)\|_{\mathcal{L}_\infty}\,.\end{aligned}$$

Then by Theorem 3 we have

$$\begin{aligned}\limsup_{t\to\infty}\|\tilde{\Omega}(t)\| &\leq \limsup_{t\to\infty}\Big(\min\{\,c_{tol}\|\hat{\Sigma}(t) - \Sigma_{pe}(t)\|, 1\,\}\Big)\\ &\leq \gamma_3\|(d_2,d_\Sigma)(t)\|_{\mathcal{L}_\infty}\end{aligned}$$

where the $\|d_2(t)\|^2_{\mathcal{L}_\infty}$ term can be ignored using the $\min\{\,\cdot\,\}$ provided $\gamma_3 > 0$ is selected sufficiently large. Altogether, $\limsup_{t\to\infty} P_1(\eta(t), \tilde{\Omega}(t))$ is upper bounded by

$$\gamma_4(\|d_2(t)\|^2_{\mathcal{L}_\infty} + \|(d_w, d_\Sigma)(t)\|_{\mathcal{L}_\infty})$$

and $\limsup_{t\to\infty} P_2(\eta(t), \tilde{\Omega}(t), d(t))$ is upper bounded by

$$\gamma_5(\|d_2(t)\|_{\mathcal{L}_\infty}\|(d_2, d_e)(t)\|_{\mathcal{L}_\infty} + \|d(t)\|_{\mathcal{L}_\infty})\,.$$

Furthermore, each $P_i(\cdot)$ is uniformly bounded as a function of t because $\tilde{w}_\circ(\cdot)$ and $\tilde{w}(\cdot)$ are continuous uniformly in t and $\nu(t)$ is uniformly bounded provided $\nu(t_0)$ is restricted to a compact set.

Fourth, by (E7) and because $\mu > 0$ there exists a converse Lyapunov function $V(t, \chi)$ for the dynamics $\dot{\chi} = F(t, \chi)$ satisfying the conclusions of [27, Theorem 4.14] globally with constants $c_i > 0$. Note that if $q_{pe} = 0$ then we use (E5) instead. Taking its time derivative with respect to trajectories of the perturbed closed-loop error model and applying Young's Inequality we have

$$\begin{aligned}\dot{V}(t,\chi) &\le -c_3\|\chi\|^2 + c_4\|\chi\|\|P(\cdot)\| \\ &\le -\left(c_3 - c_5 - c_4 P_1(\cdot)\right)\|\chi\|^2 + \left(\frac{c_4}{2\sqrt{c_5}}P_2(\cdot)\right)^2\end{aligned}$$

for some $c_5 > 0$ selected so that $c_3 - c_5 > 0$. From here, a standard Lyapunov argument using the Comparison Lemma, the same two step approach as [2, Proposition 2], and dealing with the $\limsup$ using [31, Lemma 2], which we omit for brevity, proves the result. In particular, one is to use the fact that $P_1(\cdot)$ can be made arbitrarily small in steady-state by selecting $\|(d_w, d_\Sigma)(t)\|_{\mathcal{L}_\infty}$ sufficiently small. □

F. Proof of Theorem 8

The proof follows the same procedure as the proof of Theorem 7, so here we only present the required modifications. First, there exists PE decompositions

$$\bar{w}_\circ = \bar{W}(\bar{w}_\circ)_{pe} + \bar{W}_\perp(\bar{w}_\circ)_\perp\,, \quad \bar{w} = \bar{W}\bar{w}_{pe} + \bar{W}_\perp\bar{w}_\perp$$

where $(\bar{w}_\circ)_\perp$, $\bar{w}_\perp \to 0$. Define $\tilde{\Omega} := \Omega\Omega^\intercal - \bar{W}_\perp\bar{W}_\perp^\intercal$ and the coordinate transformation

$$\begin{bmatrix}\tilde{\psi}_{pe}\\ \hat{\psi}_\perp\end{bmatrix} = \begin{bmatrix}\bar{W}^\intercal\\ \bar{W}_\perp^\intercal\end{bmatrix}(\hat{\psi} - \bar{W}\bar{W}^\intercal W W^\intercal\psi)\,.$$

Second, noting that

$$w^\intercal\psi = \bar{w}^\intercal(WW^\intercal\psi) + (W_\perp w_\perp - W(W^\intercal d_2))^\intercal\psi$$

we can write $\hat{w}^\intercal\hat{\psi} - w^\intercal\psi$ as

$$\begin{aligned}&\bar{w}_{pe}^\intercal\tilde{\psi}_{pe} + (\hat{w} - \bar{w})^\intercal\bar{W}\tilde{\psi}_{pe} + (\bar{w}_\perp + \bar{W}_\perp^\intercal(\hat{w} - \bar{w}))^\intercal\hat{\psi}_\perp\\ &\quad + (-\bar{W}_\perp\bar{w}_\perp + \bar{W}\bar{W}^\intercal(\hat{w} - \bar{w}))^\intercal(WW^\intercal\psi)\\ &\quad + (-W_\perp w_\perp + W(W^\intercal d_2))^\intercal\psi\,,\end{aligned}$$

and synonymously for $\hat{w}_\circ^\intercal\hat{\psi} - w_\circ^\intercal\psi$. Following through similar algebra as in the proof of Theorem 7, using the fact that $\|\bar{w}_{pe}\| \le \|w\| + \|d_2\|$, and defining $\chi := (\xi, \tilde{\psi}_{pe}, \hat{\psi}_\perp)$ as well as $\eta = (\eta_1, \eta_2)$ where

$$\eta_1 := \|(\bar{w}_\circ)_\perp\| + \|\hat{w}_\circ - \bar{w}_\circ\|\,, \quad \eta_2 := \|\bar{w}_\perp\| + \|\hat{w} - \bar{w}\|$$

we can compactly write the perturbed closed-loop system as

$$\dot{\chi} = F(t, \chi, d_w) + P(\cdot)$$

where $e_\circ := g(t, \xi, \bar{w}_{pe}^\intercal(t)\tilde{\psi}_{pe})$,

$$F(t,\chi,d_w) = \begin{bmatrix} f(t,\xi,(\bar{w}_\circ)_{pe}^\intercal(t)\tilde{\psi}_{pe})\\ -\gamma e_\circ\bar{w}_{pe}(t)\\ -\mu\hat{\psi}_\perp\end{bmatrix},$$

and

$$\begin{aligned}\|P(\cdot)\| \le\ &\gamma_1(\|\eta\| + \eta_2^2 + \|\tilde{\Omega}\| + \eta_2\|d_2\|)\|\chi\|\\ &+ \gamma_1(\|\eta\| + \eta_2^2 + \|\tilde{\Omega}\| + \eta_2\|d_2\|)\\ &+ \gamma_1(\|d_2\| + \eta_2)(\|w_\perp\| + \|W^\intercal d_2\| + \|d_e\|)\\ &+ \gamma_1(\|(d_\xi, d_e, d_\psi, W^\intercal d_1, W^\intercal d_2)\|)\\ &+ \gamma_1(\|(w_\circ)_\perp\| + \|w_\perp\|)\\ =:\ & P_1(\eta, \tilde{\Omega}, d_2)\|\chi\| + P_2(t, \eta, \tilde{\Omega}, d)\end{aligned}$$

for some constant $\gamma_1 > 0$.

Third, observe that $\hat{w}_\circ - \bar{w}_\circ = \tilde{w}_\circ(t,\nu)$ and $\hat{w} - \bar{w} = \tilde{w}(t,\nu)$ vanish because $\nu \to 0$. Hence $\eta \to 0$. Using

$$\frac{1}{T}\int_t^{t+T}\bar{w}_{pe}(\tau)\bar{w}_{pe}^\intercal(\tau)\,d\tau \succeq \beta_0 I\,, \quad \forall\, t \ge 0$$

we have that $\bar{\Sigma}_{pe}$ satisfies the conclusions of Lemma 1 with respect to the perturbed regressor $\bar{w}$ for $\bar{q}_{pe} \ge 1$. Comparing the difference between $\hat{\Sigma}$ and $\bar{\Sigma}_{pe}$, one has that

$$\limsup_{t\to\infty}\|\tilde{\Omega}(t)\| \le c_{tol}\varepsilon^{-1}\|d_\Sigma(t)\|_{\mathcal{L}_\infty}\,.$$

Thus $\limsup_{t\to\infty} P_1(\eta(t), \tilde{\Omega}(t), d_2(t))$ is upper bounded by

$$\gamma_2\|d_\Sigma(t)\|_{\mathcal{L}_\infty}$$

and $\limsup_{t\to\infty} P_2(t, \eta(t), \tilde{\Omega}(t), d(t))$ is upper bounded by

$$\gamma_3(\|d_2(t)\|_{\mathcal{L}_\infty}\|(W^\intercal d_2, d_e)(t)\|_{\mathcal{L}_\infty} + \|d_{pe}(t)\|_{\mathcal{L}_\infty})\,.$$

Fourth, the converse Lyapuonv function $V(t, \chi; d_w)$ for the dynamics $\dot{\chi} = F(t, \chi, d_w)$ now depends on d_w, meaning that the constants $c_i(d_w) > 0$ also depend on d_w. In turn, we obtain a $\gamma_\star(d_w) > 0$ for the scaling factor describing the steady-state behaviour of all the states. Also, notice that $P_1(\cdot)$ can be made arbitrarily small in steady-state by now only selecting $\|d_\Sigma(t)\|_{\mathcal{L}_\infty}$ sufficiently small. Altogether, the result follows by the same Lyapunov argument used in Theorem 7. □

VIII. Conclusion

This paper is premised on the philosophy that robust parameter adaptation can be achieved by forgetting the unexcited dynamics. In systems neuroscience, this approach is referred to as the *Use it or Lose it Principle*. To formalize this principle, we developed a framework based on PE subspaces yielding the PE decomposition of a general class of regressors. Then we presented a new subspace estimator that can be employed to provide robustness to parameter adaptation laws without compromising on asymptotic error regulation. We showed that our new design can be operated on a fast time-scale, while arguing that it is most effective as a slow process.

An important next step is to expand the class of error models for which we can apply the μ-modification. We made the assumption that we have a regressor estimate $\hat{w} = w(t) + \tilde{w}$ with $\tilde{w}$ vanishing independently. Some examples where this assumption holds were considered in Section VI. In other control problems such as model reference adaptive control (MRAC) or adaptive output regulation of unknown plants and unknown exosystems, the evolution of $\tilde{w}$ may be coupled with the plant and parameter dynamics. In such cases our design using a subspace estimator must be revisited.

Our developments revolved around the classical notion of persistent excitation. The μ-modification can be adapted for different characterizations of regressor excitation, such as semi-initial excitation [13] or the conditions established for Dynamic Regressor Extension and Mixing (DREM) [26]. We believe such extensions will be relatively straightforward; the key question is which notion of excitation best characterizes the properties of any regressor.

REFERENCES

[1] P. Ioannou and J. Sun, *Robust Adaptive Control*. Dover, 2012.

[2] E. Mejia Uzeda and M. Broucke, "Robust parameter adaptation and the μ-modification," *Systems and Control Letters*, vol. 171, p. 105416, 2023.

[3] K. Astrom and B. Wittenmark, *Adaptive Control*. Dover Publications, 2008.

[4] S. Boyd and S. Sastry, "Necessary and sufficient conditions for parameter convergence in adaptive control," *Automatica*, vol. 22, 1986.

[5] A. Padoan, G. Scarciotti, and A. Astolfi, "A geometric characterization of the persistence of excitation condition for the solutions of autonomous systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5666–5677, 2017.

[6] A. Serrani, "Rejection of harmonic disturbances at the controller input via hybrid adaptive external models," *Automatica*, vol. 42, no. 11, pp. 1977–1985, 2006.

[7] R. Marino and P. Tomei, "On exponentially convergent parameter estimation with lack of persistency of excitation," *Systems & Control Letters*, vol. 159, no. 105080, 2022.

[8] S. Bittanti, P. Bolzern, and M. Campi, "Recursive least-squares identification algorithms with incomplete excitation: convergence analysis and application to adaptive control," *IEEE Transactions on Automatic Control*, vol. 35, no. 12, pp. 1371–1373, 1990.

[9] M. Korotina, J. Romero, S. Aranovskiy, A. Bobtsov, and R. Ortega, "A new on-line exponential parameter estimator without persistent excitation," *Systems & Control Letters*, vol. 159, no. 105079, 2022.

[10] A. Parikh, R. Kamalapurkar, and W. E. Dixon, "Integral concurrent learning: Adaptive control with parameter convergence using finite excitation," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1775–1787, 2019.

[11] J. Wang, D. Efimov, and A. A. Bobtsov, "On robust parameter estimation in finite-time without persistence of excitation," *IEEE Transactions on Automatic Control*, vol. 65, no. 4, pp. 1731–1738, 2020.

[12] A. Ovcharov, A. Vedyakov, S. Kazak, and A. Pyrkin, "Overparameterized model parameter recovering with finite-time convergence," *International Journal of Adaptive Control and Signal Processing*, vol. 36, no. 6, pp. 1305–1325, 2022.

[13] S. Basu Roy and S. Bhasin, "Novel model reference adaptive control architecture using semi-initial excitation-based switched parameter estimator," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1759–1774, 2019.

[14] P. Tomei and R. Marino, "An enhanced feedback adaptive observer for nonlinear systems with lack of persistency of excitation," *IEEE Transactions on Automatic Control*, vol. 68, no. 8, pp. 5067–5072, 2023.

[15] A. I. Gluschenko and K. A. Lastochkin, "Relaxation of conditions for convergence of dynamic regressor extension and mixing procedure," *Automation and Remote Control*, vol. 84, pp. 14–41, 2023.

[16] K. Narendra and A. Annaswamy, *Stable Adaptive Systems*. Dover Publications, 1989.

[17] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. Prentice-Hall, 1989.

[18] T. Yoshizawa, *Stability Theory and the Existence of Periodic Solutions and Almost Periodic Solutions*, vol. 14. Springer New York, 1975.

[19] B. D. O. Anderson, R. Bitmead, *et al.*, *Stability of Adaptive Systems*. The MIT Press, 1986.

[20] G. Kreisselmeier, "Adaptive observers with exponential rate of convergence," *IEEE Transactions on Automatic Control*, vol. 22, no. 1, pp. 2–8, 1977.

[21] G. W. Stewart, "Perturbation theory for the singular value decomposition," 1990.

[22] P. Å. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT Numerical Mathematics*, vol. 12, pp. 99–111, 1972.

[23] G. W. Stewart, "On the perturbation of pseudo-inverses, projections and linear least squares problems," *SIAM Review*, vol. 19, no. 4, pp. 634–662, 1977.

[24] G. Chowdhary, T. Yucelen, M. Muhlegg, and E. Johnson, "Concurrent learning adaptive control of linear systems with exponentially convergent bounds," *Int. J. Adaptive Control Signal Processing*, vol. 27, pp. 280–301, 2013.

[25] J. Sanders, F. Verhulst, and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems, 2nd Ed.* Springer, 2007.

[26] R. Ortega, V. Nikiforov, and D. Gerasimov, "On modified parameter estimators for identification and adaptive control. A unified framework and some new schemes," *Annual Reviews in Control*, vol. 50, 2020.

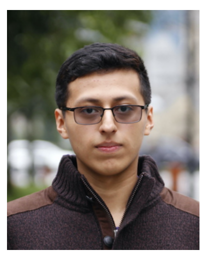
[27] H. Khalil, *Nonlinear Systems, 3rd Ed.* Prentice-Hall, 2002.

[28] E. Mejia Uzeda and M. E. Broucke, "Adaptive output regulation and the use it or lose it principle," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 3547–3554, 2023.

[29] V. Nikiforov and D. Gerasimov, *Adaptive Regulation*, vol. 491. Springer International Publishing, 2022.

[30] J. K. Hale, *Ordinary Differential Equations, 2nd Ed.* Kreiger Publishing Company, 1980.

[31] E. Mejia Uzeda and M. E. Broucke, "Training reflexes using adaptive feedforward control," *IEEE Open Journal of Control Systems*, vol. 2, pp. 396–409, 2023.



Erick Mejia Uzeda received the B.A.Sc. degree in electrical engineering and M.A.Sc. degree in systems control from the University of Toronto, Toronto, Canada, in 2021 and 2023, respectively. He is currently pursuing the Ph.D degree in systems control at the University of Toronto. His research interests include adaptive control, regulator theory, and applications of control theory to systems neuroscience.

Mireille E. Broucke obtained the BSEE degree in Electrical Engineering from the University of Texas at Austin in 1984 and the MSEE and PhD degrees from the University of California, Berkeley in 1987 and 2000, respectively. She has six years industry experience at Integrated Systems, Inc. and several aerospace companies. From 1993 to 1996 she acted as program manager and researcher at Partners for Advanced Transportation and Highways (PATH) at UC Berkeley. She is a professor in Electrical and Computer Engineering at the University of Toronto. Her research interests are in the application of control theory to systems neuroscience.