

An Asymptotic Scaling Analysis of LQ Performance for an Approximate Adaptive Control Design*

M. French,[†] Cs. Szepesvári,[‡] and E. Rogers[†]

Abstract. We consider the adaptive tracking problem for a chain of integrators, where the uncertainty is static and functional. The uncertainty is specified by L^2/L^∞ or weighted L^2/L^∞ norm bounds. We analyse a standard Lyapunov-based adaptive design which utilises a function approximator to induce a parametric uncertainty, on which the adaptive design is completed. Performance is measured by a modified LQ cost functional, penalising both the tracking error transient and the control effort. With such a cost functional, it is shown that a standard control design has divergent performance when the resolution of a “mono-resolution” approximator is increased. The class of “mono-resolution” approximators includes models popular in applications. A general construction of a class of approximators and their associated controllers which have a uniformly bounded performance independent of the resolution of the approximator is given.

Key words. Adaptive control, Function approximation, LQ performance.

1. Introduction

The use of function approximators within adaptive control designs has been popular since the publication of the papers [SS] and [T]. The idea is to use standard adaptive designs for problems which depend on non-parameterised uncertainties, by utilising function approximators to induce an appropriate approximate parameterisation of unknown system functions. The non-parametric uncertainty has then been converted into an uncertainty in the parameters of the function approximator, to which the standard adaptive designs can be applied, whilst the inherent approximation error is simply treated as a disturbance acting on the system and the standard robust modifications [NA] to the adaptive laws (dead-zones, projections, σ modification, etc.) are made to ensure stability.

It is thus widely appreciated that function approximator-based adaptive

* Date received: 20 April 1999. Date revised: 29 October 2001.

[†] Department of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, England. {mcf,etar}@ecs.soton.ac.uk.

[‡] Mindmaker Ltd., Konkoly Th. M. u. 29-33, Budapest 1112, Hungary. szepes@mindmaker.hu. Also with Research Group on Artificial Intelligence, Hungarian Academy of Sciences—JATE, Szeged, Hungary.

control designs are essentially robust adaptive controllers, with the additional twist that in general the approximation can only be valid over compact domains if a finite-dimensional model is utilised. This gives the theory a semi-global characteristic, whereby the transient must be constrained a priori to lie within the model's compact domain, either by high gain robust terms (often within a sliding mode framework), by high adaption rates or by limiting the uncertainty. By carefully considering the spatial growth of the uncertainty, it is also possible to give global designs based on models with dynamically varying dimension [FSR2]. Designs are available for systems with matched uncertainties and systems in the strict feedback form and output feedback form, see e.g. [F].

Despite the large number of adaptive designs based on function approximators given in the literature, e.g. [SS], [YL] and [CK], little attention has been paid to features (other than the semi-globality) of the approximation theoretic designs which do not appear in the analogous parametric robust adaptive theory. One notable exception to this point is the work in [CS], where a design is given which is in some sense adaptive to the smoothness of the system.

In this paper we demonstrate that there are fundamental questions which arise in the approximate adaptive theory which have no analogue in the parametric robust adaptive theory. These questions centre around the first stage of an approximate adaptive approach: namely the choice of model (i.e. the function approximator). The fundamental question we address is whether increasing the resolution of the model leads to a degradation of the transient performance of the closed-loop system. This question has high relevance since high resolution models are utilised for three reasons:

Scenario 1. To guarantee a small asymptotic L^∞ tracking error: increasing the model resolution decreases the approximation error, and thus permits greater asymptotic tracking accuracy.

Scenario 2. Due to the conservatism of the approximation theoretic bounds which relate approximator resolution to smoothness: when the bounds are conservative, overly high approximator resolutions are utilised to guarantee the required approximation error tolerance.

Scenario 3. Due to the fact that the smoothness of the nonlinearities is hard to estimate in applications: a priori it is difficult to estimate the smoothness of the nonlinearities of a system in the form required for the approximation theoretic bounds.

Hence, it is desirable that scaling the model by increasing the resolution (typically by increasing the dimensionality of the parameter space) does not detrimentally affect the transient performance.

To examine this question we formulate a non-singular linear-quadratic (LQ) type cost, and examine the behaviour of this cost as the resolution of the function approximator is increased. It is important to observe that we are penalising both

the state *and the control effort* in this cost as due to the presence of high gain solutions, the question considered would be ill-posed for singular costs with no penalty on the control. We then require techniques to compute both lower and upper bounds on such costs. A technique for estimating an upper bound was developed in previous work of the authors [FSR2], whilst a technical innovation of this article is a method for estimating lower bounds.

Within the LQ framework we then compare the performances of two classes of function approximator-based controllers as their respective resolutions are increased. We find that two competing designs, one with a mono-resolution and one with a multi-resolution function approximator, have substantially different behaviour. A situation is identified where the cost of increasing the resolution of mono-resolution design increases unboundedly (at the same time we prove that the design yields a stable closed loop so the cost increase cannot be attributed to losing stability); whilst in contrast the cost of increasing the resolution of a multi-resolution function approximator is uniformly bounded.

A statement of greater precision is as follows. We consider systems $\Sigma_{\Delta}(\mathcal{Y}_0)$, where the system's initial condition y_0 lies inside a bounded set \mathcal{Y}_0 , and where the system's nonlinearity f satisfies a matching condition and lies inside the bounded set Δ . The uncertainty Δ is constrained by both spatial L^2 and L^∞ norm bounds. Given a bounded set of reference trajectories \mathcal{Y}_{ref} , the worst case asymptotic tracking requirement is specified by an error set Ω_0^m . The adaptive controllers $\Xi(\Phi^m, \alpha^m, \Omega_0^m)(\mathcal{Y}_{\text{ref}})$ considered in this paper are Lyapunov-based designs dependent on both a function approximator Φ^m , where m indexes the parameter dimension, and on a tunable parameter α^m (an adaption gain). We measure performance of a closed loop $(\Sigma_{\Delta}(\mathcal{Y}_0), \Xi(\Phi^m, \alpha^m, \Omega_0^m)(\mathcal{Y}_{\text{ref}}))$ by a modified worst case LQ cost functional which is denoted by

$$\mathcal{P}_m = \mathcal{P}(\Sigma_{\Delta}(\mathcal{Y}_0), \Xi(\Phi^m, \alpha^m, \Omega_0^m)(\mathcal{Y}_{\text{ref}})). \quad (1)$$

The essence of the modifications to the LQ cost functional is that the cost is not measured inside the error-set Ω_0^m . The cost is worst case with respect to the initial conditions, reference trajectory and system, and is best case with respect to the adaption gain (i.e. we are considering the optimally tuned controller).

In Section 3 we establish a negative result: demonstrating that scaling can be a real issue for these function approximator designs. The class of uncertain systems $\Sigma_{\Delta}(\mathcal{Y}_0)$ considered are MIMO systems of order one. The class of models considered are characterised by uniformly localised basis functions: this class includes B-spline networks Φ^m defined on uniform knot lattices. Such networks are utilised in many applications of approximate adaptive designs [JVL], [OZSP].¹ For this class we show that the designs do not scale when considering Scenario 1 above. In particular, we show that if we satisfy a demand for higher asymptotic tracking accuracy (i.e. $\Omega_0^m \rightarrow \{0\}$ as $m \rightarrow \infty$) by increasing the resolution of the approximator, then the performance diverges although, e.g. stability is still maintained.

¹ The class also includes the Gaussian Radial Basis function (RBF) networks of [SS], subject to an unresolved conjecture (see Conjecture 3.7 below).

Specifically we show that there is a choice of adaptive gains $\{\alpha^m\}_{m \geq 1}$ such that for all sufficiently large m , $\mathcal{P}_m < \infty$, but for any choice of $\{\alpha^m\}_{m \geq 1}$ it follows that

$$\limsup_{m \rightarrow \infty} \mathcal{P}_m = \infty. \quad (2)$$

The poor scaling of performance is solely due to the control design, since under perfect knowledge a feedback linearisation strategy has the property that $\limsup_{m \rightarrow \infty} \mathcal{P}_m < \infty$.

More general control designs for output feedback and strict feedback systems based on backstepping collapse to the controllers we are considering when applied to these first-order systems. Therefore we can conclude that this undesirable performance divergence is a feature of all these adaptive designs when coupled with certain function approximators. We have therefore raised serious questions about the applicability of many designs (and their generalisations) which are popular in the connectionist communities.

In Section 4 we show that this degradation of performance can be avoided in Scenarios 1–3 considered above. The class of systems $\Sigma_{\Delta}(\mathcal{Y}_0)$ considered is a MIMO chain of integrators. This is not as general a class as one would ideally like, but we take this as the first step in comparing the performance of adaptive control designs: given that this class of systems includes the class of first-order MIMO systems considered earlier, we are able to conclude that for a nontrivial class of systems these two designs differ substantially (in Scenario 1). Furthermore, extensions of these positive results have also been made to classes of backstepping controllers [FS].

Specifically, we give a general construction for controllers $\Xi(\Phi^m, \alpha^m, \Omega_0^m)$ based on a wide class of function approximators Φ^m which have the property that there is a choice of $\{\alpha^m\}_{m \geq 1}$ such that

$$\limsup_{m \rightarrow \infty} \mathcal{P}_m < \infty. \quad (3)$$

The results hold in both the cases of $\Omega_0^m \rightarrow \{0\}$ (Scenario 1) and that of a constant $\Omega_0^m = \Omega_0$ (Scenarios 2 and 3). Results are given in both a semi-global setting with finite-dimensional models (in which case the uncertainty level is required knowledge) and in a global setting where the uncertainty level is not required to be known a priori, but the function approximator is of countable dimension. However, by restricting these infinite-dimensional approximator structures to be semi-globally finite-dimensional (SFD), the global controllers can be realised as finite-dimensional controllers since only a finite number of adaptive estimates are updated (the number of updates is dependent on the uncertainty level), see, e.g. [FSR2].

The essential restriction we impose on the approximator class is that higher resolution models are generated by extending the basis function set. Thus function approximators based on series expansions such as polynomial bases, Fourier series, wavelets, etc., are valid; and in a simple manner it is possible to take models of any functional form, and by generating a controller based on the union of all

the models below a particular resolution, also achieve uniformly bounded closed-loop performance.

Whilst we are predominantly concerned with a question which is only applicable to function approximator-based designs, the techniques developed in this paper should be of wider interest: we have to compute both upper and lower bounds on an integral performance measure, techniques for which must be developed if any analytical comparison is to be made between, e.g. robust and adaptive controllers. The upper bound theory extends that of [FSR2], where the resolution scaling issue was first raised, whilst the lower bound estimates are completely new.

2. Problem Formulation

2.1. Notation

We denote by $\|\cdot\|$ the 2-norm of vectors and all norms over vector-valued functions will be defined by this norm. For example, for $f \in C(\Omega, \mathbb{R}^n)$, $\|f\|_{C(\Omega, \mathbb{R}^n)}$ is defined by $\|f\|_{C(\Omega, \mathbb{R}^n)} = \|f\|_{L^\infty(\Omega, \mathbb{R}^n)} = \sup_{x \in \Omega} \|f(x)\|_2$. The matrix norm $\|x\|_Q$ for a positive definite, symmetric matrix Q and vector x is defined by $\|x\|_Q^2 = x^T Q x$. An inner product for an inner product space \mathcal{H} is denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. For a function space \mathcal{F} with domain X and range Y we write $\mathcal{F}(X, Y)$, if $Y = \mathbb{R}$ we write $\mathcal{F}(X)$. The order notation is defined by $f(x) = O(x)$ if and only if $\limsup_{\|x\| \rightarrow \infty} |f(x)|/\|x\| < \infty$. Hence the statement $f(x) \geq O(x)$ means $\exists g$ such that $f(x) \geq g(x)$ and $g(x) = O(x)$, similarly for $f(x) \leq O(x)$.

For a set $\Omega \subset \mathbb{R}^n$, we let Ω° denote the interior, $\text{dist}(x, \Omega) = \inf_{y \in \Omega} \|x - y\|$ and $\text{diam}(\Omega) = \sup\{\|x - y\| : x, y \in \Omega\}$. Elements of Euclidean spaces $\mathcal{W} = \mathbb{R}^m$ are thought of as column vectors. We occasionally let 1 denote the vector $(1, 1, \dots, 1)^T$. $\bar{\lambda}(R), \underline{\lambda}(R)$ denote the maximum and minimum eigenvalues of a matrix R . We define $\mathbb{R}_+ = \{t \in \mathbb{R} : t \geq 0\}$. A closed-loop system is said to be *well-posed* if all solutions of the system are defined on \mathbb{R}_+ , and all closed-loop signals are bounded. The support of a function f is denoted by $\text{supp}(f)$.

2.2. Control Task Formulation

2.2.1. System Specification

We consider systems which are the union of p integrator chains each controlled by a single input. The order of integration for the j th integrator ($1 \leq j \leq p$) will be n_j ($1 \leq n_j$) and its state will be denoted by $Y_j = (y_j, y_j^{(1)}, \dots, y_j^{(n_j-1)})^T \in \mathbb{R}^{n_j}$. The control will be denoted $u = (u_1, \dots, u_p)^T \in \mathbb{R}^p$. The nonlinearities are matched, and hence lie in the span of the control. The nonlinearity for the j th system will be denoted by $f_j: \mathbb{R}^n \rightarrow \mathbb{R}$. These are functions of the *full* state $Y = (Y_1, \dots, Y_p)^T \in \mathbb{R}^n$, $n = \sum_{j=1}^p n_j$, of the system, and $f: \mathbb{R}^n \rightarrow \mathbb{R}^p$ is defined by $f(Y) = (f_1(Y), \dots, f_p(Y))^T$.

We formalise this as follows: Let $A_j \in \mathbb{R}^{n_j \times n_j}$, $B_j \in \mathbb{R}^{n_j \times 1}$, $1 \leq j \leq p$, be defined:

$$A_j = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad B_j = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \quad (4)$$

We consider systems of the form:

$$\Sigma_f(Y_0): \dot{Y}_j = A_j Y_j + B_j(f_j(Y_1, \dots, Y_p) + u_j), \quad 1 \leq j \leq p, \quad Y(0) = Y_0. \quad (5)$$

Typically we are concerned with situations where $f: \mathbb{R}^n \rightarrow \mathbb{R}^p$ is unknown and lies within the set Δ :

$$f = (f_1, \dots, f_p)^T \in \Delta_1 \times \cdots \times \Delta_p = \Delta. \quad (6)$$

We define an uncertain system as

$$\Sigma_\Delta(Y_0) = \{\Sigma_f(Y_0) \mid f \in \Delta\}. \quad (7)$$

Throughout this paper we are concerned with uncertainty sets Δ with the following property:

$$\Delta_j \subset K_j \cap \Delta(\mathcal{H}_j; \delta_{2j}) \cap \Delta(\mathcal{F}_j; \delta_{\infty j}), \quad 1 \leq j \leq p, \quad (8)$$

where

$$\Delta(\mathcal{H}_j; \delta_{2j}) = \{f_j \in \mathcal{H}_j \mid \|f_j\|_{\mathcal{H}_j} \leq \delta_{2j}\}, \quad (9)$$

$$\Delta(\mathcal{F}_j; \delta_{\infty j}) = \{f_j \in \mathcal{F}_j \mid \|f_j\|_{\mathcal{F}_j} \leq \delta_{\infty j}\}. \quad (10)$$

Here \mathcal{H}_j is the space $L^2(\Omega)$ or the weighted space $L^2(\Omega; w_{2j})$, and \mathcal{F}_j is the space $L^\infty(\Omega)$ or the weighted space $L^\infty(\Omega; w_{\infty j})$,² where $\Omega \subset \mathbb{R}^n$. Ω will generally be compact in the unweighted cases, and global in the weighted case. $K = K_1 \times \cdots \times K_p \subset C(\Omega, \mathbb{R}^p)$ is an approximation theoretic smoothness class, which we will discuss subsequently. The motivation for modelling the uncertainty in both an L^2 and an L^∞ sense is discussed in [FSR2], but broadly speaking the L^2 bound suffices to give stability conditions and characterise the state performance, whereas it is necessary to have (pointwise) L^∞ information to bound to the control effort.

As initially our approximation domains Ω will be compact, we first consider semi-global results, and so define an initial condition set as

$$\mathcal{Y}_0 = \{Y_0 \in \mathbb{R}^n \mid \|Y_0\| \leq \gamma_0\}, \quad (11)$$

and let $\Sigma_\Delta(\mathcal{Y}_0) = \{\Sigma_\Delta(Y_0) \mid Y_0 \in \mathcal{Y}_0\}$. Given a reference trajectory, $y_{\text{ref}}: \mathbb{R}_+^p \rightarrow \mathbb{R}_+^p$,

$$y_{\text{ref}} \in C^{n_1}(\mathbb{R}_+, \mathbb{R}) \times C^{n_2}(\mathbb{R}_+, \mathbb{R}) \times \cdots \times C^{n_p}(\mathbb{R}_+, \mathbb{R}), \quad (12)$$

we denote

$$Y_j^{\text{ref}} = (y_{\text{ref}_j}, y_{\text{ref}_j}^{(1)}, \dots, y_{\text{ref}_j}^{(n_j-1)})^T, \quad 1 \leq j \leq p, \quad Y^{\text{ref}} = (Y_1^{\text{ref}}, \dots, Y_p^{\text{ref}}) \quad (13)$$

² The weighted spaces $L^2(\Omega; w_2)$, $L^\infty(\Omega; w_\infty)$ are defined by the inner product $\langle f, g \rangle = \|fgw_2\|_{L^2(\Omega)}^2$ and norm $\|f\| = \|f(\cdot)w_\infty(\|\cdot\|)\|_{L^\infty(\Omega)}$, respectively. We assume throughout that $w_\infty: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is monotonically decreasing and that the weights w_2 lie in L^1 .

(i.e. $Y^{\text{ref}}: \mathbb{R}_+^p \rightarrow \mathbb{R}^n$), and define a reference trajectory set as

$$\mathcal{Y}_{\text{ref}} = \left\{ y_{\text{ref}} \in C^{n_1}(\mathbb{R}_+, \mathbb{R}) \times \cdots \times C^{n_p}(\mathbb{R}_+, \mathbb{R}), \|Y^{\text{ref}}\|_{L^\infty(\mathbb{R}_+)} \leq \gamma_1, \right. \\ \left. \max_{1 \leq j \leq p} \|y_{\text{ref}_j}^{(n_j)}\|_{L^\infty(\mathbb{R}_+)} \leq \gamma_2 \right\}, \quad (14)$$

where $\gamma_1, \gamma_2 \geq 0$ are fixed numbers, known to the control designer. It is also convenient to define

$$\Omega_2 = \{z \in \mathbb{R}^n \mid \|z\| \leq \gamma_1\} \subset \Omega, \quad (15)$$

so (14) implies that for any reference signal $y_{\text{ref}} \in \mathcal{Y}_{\text{ref}}$, $Y^{\text{ref}}(t) \in \Omega_2$ for all $t \geq 0$.

2.2.2. Stability and Performance: Definitions

Let $\Omega_0 \subset \mathbb{R}^n$ be an a priori specified neighbourhood of the origin. The *control task* is to give a controller $\Xi(y_{\text{ref}})$ which drives the worst case tracking error $x = (x_1, \dots, x_p)^T$ to Ω_0 as $t \rightarrow \infty$, where x is defined by

$$x_j(\cdot) = Y_j(\cdot) - Y_j^{\text{ref}}(\cdot), \quad 1 \leq j \leq p, \quad (16)$$

i.e. we want

$$\sup_{f \in \Delta} \sup_{Y_0 \in \mathcal{Y}_0} \sup_{y_{\text{ref}} \in \mathcal{Y}_{\text{ref}}} \sup_{\text{solns}(\Sigma_f(Y_0), \Xi(y_{\text{ref}}))} \lim_{t \rightarrow \infty} \text{dist}\{x(t), \Omega_0\} = 0. \quad (17)$$

Whilst demanding a good asymptotic worst case tracking error, the control designs will be judged by their transient performance. So let $Q_j \in \mathbb{R}^{n_j \times n_j}$ be a symmetric, positive definite matrix and let $r_j > 0$ for $1 \leq j \leq p$. We let $Q \in \mathbb{R}^n \times \mathbb{R}^n$, $R \in \mathbb{R}^p \times \mathbb{R}^p$ be the block diagonal matrices $Q = \text{diag}(Q_1, \dots, Q_p)$, $R = \text{diag}(r_1, \dots, r_p)$. Transient performance is then measured by the following worst case LQ cost:

$$\mathcal{P}(Q, R, \Omega_0, \Omega_1) \\ = \mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi(\mathcal{Y}_{\text{ref}}))}(Q, R, \Omega_0, \Omega_1) \\ = \sup_{f \in \Delta} \sup_{Y_0 \in \mathcal{Y}_0} \sup_{y_{\text{ref}} \in \mathcal{Y}_{\text{ref}}} \sup_{\text{solns}(\Sigma_f(Y_0), \Xi(y_{\text{ref}}))} \left(\int_{T_{\Omega_0}} x^T Q x \, dt + \int_{T_{\Omega_1}} u^T R u \, dt \right), \quad (18)$$

where T_{Ω_i} , $i = 0, 1$, is defined:

$$T_{\Omega_i} = \{t \geq 0 \mid x(t) = Y(t) - Y^{\text{ref}}(t) \notin \Omega_i\}, \quad i = 0, 1, \quad (19)$$

for suitable neighbourhoods Ω_0, Ω_1 of $0 \in \mathbb{R}^n$. Here $x(t)$ and $u(t)$ denote the tracking error and control signals for the (well-posed) system $(\Sigma_f(Y_0), \Xi(y_{\text{ref}}))$.

We motivate this cost as follows. It is reasonable to penalise the state error only

³ For notational simplicity, both here and in what follows, we do not index closed-loop signals x, u , etc., by the closed loop to which they belong: e.g. for x , one should always read $x_{(\Sigma_f(Y_0), \Xi(y_{\text{ref}}))}$, etc.

⁴ The inner supremum is taken over all solutions of the closed loop: it is required as in general the uniqueness of solutions cannot be established, e.g. for controllers with a dead-zone modification [PI].

when it lies above the desired accuracy. It is also reasonable to penalise the control cost with respect to the same threshold, i.e. to take $\Omega_1 = \Omega_0$. However, we are interested in the behaviour as $\Omega_0 \rightarrow \{0\}$. Note that if the nonlinearity of system (5) is known (e.g. $\Delta = \{f^0\}$), then even the feedback linearisation strategy,

$$u_j = -f_j^0 - a_j^T x_j + y_{\text{ref},j}^{(n)}, \quad 1 \leq j \leq p \quad (20)$$

(see Section 2.4 for the definition of a_j), yields

$$\mathcal{P}(Q, R, \Omega_0, \Omega_0) \rightarrow \infty \quad (21)$$

as $\Omega_0 \rightarrow \{0\}$ for $\mathcal{Y}_0 \neq \{0\}$. However, for the feedback linearisation strategy we can also establish:

$$\mathcal{P}(Q, 0, \Omega_0, \Omega_0) \leq \mathcal{P}(Q, 0, \emptyset, \emptyset) < \infty, \quad \forall \Omega_0 \subset \mathbb{R}^n, \quad (22)$$

so it follows that the divergent behaviour in (21) is solely due to the control effort term. There are many different ways of modifying the cost to prevent this divergence whilst still penalising the control effort. Here we elect to consider cost functionals of the form $\mathcal{P}(Q, R, \Omega_0, \Omega_1)$. This is a reasonable basis to judge the behaviour of adaptive designs as $\Omega_0 \rightarrow \{0\}$ (for a fixed Ω_1), as the perfect feedback linearisation strategy gives uniformly bounded behaviour:

$$\mathcal{P}(Q, R, \Omega_0, \Omega_1) \leq \mathcal{P}(Q, R, \{0\}, \Omega_1) < \infty, \quad \forall \Omega_0 \subset \mathbb{R}^n. \quad (23)$$

2.3. Approximation Theory and Notation

To define the control design and to formulate the problem investigated in this paper precisely, we first introduce the notion of a smoothness class and that of a linearly parameterised approximate model class.

For $\Omega \subset \mathbb{R}^n$, a *smoothness class* is a dense nested set of subsets $\{K_\beta\}_{\beta \geq 0}$ of $C(\Omega)$ (i.e. $K_{\beta_1} \subset K_{\beta_2} \subset C(\Omega)$ for all $\beta_1 \leq \beta_2$). Typically a smoothness class is specified by Lipschitz constraints, or by bounds in Sobolev spaces as is typical in approximation theory. A *model class* is a sequence of model bases $\{\phi^m\}_{m \geq 1}$,⁵ $\phi^m: \Omega \rightarrow \mathcal{W}_m$ where \mathcal{W}_m is a Euclidean space called the weight space of the m th model. Typically $\dim \mathcal{W}_m$, the dimension of ϕ^m , will be a divergent function of m .

Generic approximation theory furnishes us with a partially defined function $\rho: \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{N}$, called the *dimension function* that satisfies

$$\sup_{f \in K_\beta} \inf_{\theta \in \mathcal{W}_{\rho(\beta, \varepsilon)}} \|f - \theta^T \phi^{\rho(\beta, \varepsilon)}\|_{C(\Omega)} < \varepsilon, \quad (24)$$

for any smoothness parameter β and approximation error bound ε in the domain of ρ . If $\text{dom } \rho = \mathbb{R}_+ \times \mathbb{R}_+$, then the model class is said to be a $\{K_\beta\}_{\beta \geq 0}$ approximate model class. For the case of compact Ω , there are many different constructions of approximate model classes with finite-dimensional bases. As an

⁵ We use superscripts to denote both powers and index quantities, the usage is determined by context.

example, if ϕ^m is the polynomial basis $\phi^m = [1, x, x^2, \dots, x^m]$, $\Omega = [a, b]$ and K_β is the Lipschitz-class,

$$K_\beta = \{f \in C[a, b] \mid |f(x) - f(y)| \leq \beta|x - y|, \forall x, y \in [a, b]\}, \quad (25)$$

then Jackson's theorem [R] gives the dimension function $\rho(\beta, \varepsilon) = \beta(b - a)/3\varepsilon$. For non-compact Ω , given appropriate regularity, approximate model classes can also be constructed with bases of countable dimension. In the final section in this paper we consider such a model class.

In this paper we are concerned with the approximation of functions in $C(\mathbb{R}^n, \mathbb{R}^p)$, i.e. vector-valued multivariate approximation. For simplicity, we restrict the exposition here to componentwise approximation of these vector-valued functions where the functions corresponding to the individual vector components are approximated independently of each other with a separate set of (multi-variable) function approximators. The corresponding model will be denoted by $\Phi^m: \mathbb{R}^n \rightarrow \mathcal{W}_{m_1} \times \dots \times \mathcal{W}_{m_p}$, where $\Phi^m = (\phi_1^m, \dots, \phi_p^m)^T$, $\phi_j^m: \mathbb{R}^n \rightarrow \mathcal{W}_{m_j}$, where for notational simplicity, we further assume $m_j = m$ for $1 \leq j \leq p$ and take the same model basis for each component $\phi_j^m = \phi_k^m$ for $1 \leq j, k \leq p$. Similarly, we consider the multi-output smoothness class $\{K_\beta\}_{\beta \geq 0}$, $K_\beta \subset C(\Omega, \mathbb{R}^p)$ as defined by

$$K_\beta = K'_\beta \times K'_\beta \times \dots \times K'_\beta = (K'_\beta)^p \quad (26)$$

for some smoothness class $\{K'_\beta\}_{\beta \geq 0}$, $K'_\beta \subset C(\Omega)$.

2.4. Controller Design

The class of controllers considered in this paper are adaptive controllers derived from a simple Lyapunov analysis; as such the designs are well known in the literature, see, e.g. [SS], [KKK], etc. It is important to observe that more complex backstepping designs such as those for the output feedback form and the strict feedback form reduce to the controllers we are considering when applied to MIMO systems of relative degree one, see e.g. [F], [FR] and [FSR1] for dead-zone modified backstepping designs.

The control is taken to be

$$u_j = -\hat{\theta}_j^T \phi_j(Y) - a_j^T x_j + y_{\text{ref}_j^{(n)}}, \quad 1 \leq j \leq p, \quad (27)$$

where $a_j \in \mathbb{R}^{n_j}$ is chosen such that the matrices

$$A_{*j} = A_j - B_j a_j^T, \quad 1 \leq j \leq p, \quad (28)$$

are Hurwitz. The dynamics of the estimator $\hat{\theta}_j \in \mathcal{W}_{m_j}$ is then given by an adaptive law of the form:

$$\dot{\hat{\theta}}_j = \alpha_j x_j^T b_j D(\Omega_0, x) \Gamma_j \phi_j(Y), \quad \hat{\theta}_j(0) = 0 \in \mathcal{W}_{m_j}, \quad 1 \leq j \leq p, \quad (29)$$

where $\Gamma_j \in \mathbb{R}^{\dim \mathcal{W}_{m_j} \times \dim \mathcal{W}_{m_j}}$ is a positive definite matrix, called the adaptive structure matrix, $b_j \in \mathbb{R}^{\dim \mathcal{W}_{m_j} \times 1}$ is a weighting vector (to be defined below), $0 < \alpha_j \in \mathbb{R}$ is the adaption gain and $D(\Omega_0, \cdot)$ is the *dead-zone function*, defined to be the characteristic function of $\mathbb{R}^n \setminus \Omega_0$. We further assume that $\phi_j \in \mathcal{H}_j$, and is locally Lipschitz continuous. We define $\Gamma = \text{diag}(\Gamma_1, \dots, \Gamma_p)$, $b = (b_1, \dots, b_p)^T$

and $\alpha = (\alpha_1, \dots, \alpha_p)^T$. Throughout this paper let P_j be the solution to the Lyapunov equation

$$A_{*j}^T P_j + P_j A_{*j} = -Q_j, \quad 1 \leq j \leq p,$$

where recall that $Q_j = Q_j^T > 0$. Define

$$P = \text{diag}\{P_1, \dots, P_p\}, \quad (30)$$

and let

$$b_j = (P_j^T + P_j)B_j, \quad 1 \leq j \leq p.$$

Equations (27) and (29) then define the controller

$$\Xi(y_{\text{ref}}) = \Xi(\Gamma, \alpha, \Phi, \Omega_0)(y_{\text{ref}}), \quad (31)$$

and together with (5) yield the following closed-loop system $(\Sigma_f(Y_0), \Xi(y_{\text{ref}}))$:

$$\begin{aligned} \dot{x}_j &= A_{*j}x_j + B_j((\theta_j - \hat{\theta}_j)^T \phi_j(Y) + d_{f_j}(Y)), \quad 1 \leq j \leq p, \\ x(0) &= Y(0) - Y^{\text{ref}}(0), \\ \dot{\hat{\theta}}_j &= \alpha_j x_j^T b_j D(\Omega_0, x) \Gamma_j \phi_j(Y), \quad 1 \leq j \leq p, \quad \hat{\theta}_j(0) = 0, \end{aligned} \quad (32)$$

where $d_{f_j}(Y) = f_j(Y) - \theta_j^T \phi_j(Y)$, and θ_j is such that

$$\|d_{f_j}(Y)\|_{C(\Omega)} \leq 2 \inf_{\vartheta \in \mathcal{W}_m} \|f_j - \vartheta^T \phi_j\|_{C(\Omega)}.^6 \quad (33)$$

We let $\theta = (\theta_1, \dots, \theta_p)^T$ and $d_f = (d_{f_1}, \dots, d_{f_p})^T$.

Throughout the paper we take Ω_0, Ω_1 to be of the form

$$\begin{aligned} \Omega_0 &= \{x \in \mathbb{R}^n \mid x^T P x \leq \eta_0^2\}, \\ \Omega_1 &= \{x \in \mathbb{R}^n \mid x^T P x \leq \eta_1^2\}, \end{aligned} \quad (34)$$

for some $\eta_0, \eta_1 \geq 0$. We define the \mathcal{H}_j Gram matrix G_j of the model component $\phi_j: \mathbb{R}^n \rightarrow \mathcal{W}_{m_j}$ by

$$G_j = \{g_{ik}^j\}_{1 \leq i, k \leq \dim \mathcal{W}_m}, \quad g_{ik}^j = \langle (\varphi_j)_i, (\varphi_j)_k \rangle_{\mathcal{H}_j}. \quad (35)$$

2.5. Basic Stability Result

We now give the basic stability/performance result concerning such controllers.

Theorem 2.1. *Let $\Omega \subset \mathbb{R}^n$ be a closed set. Consider the system $\Sigma_\Delta(\mathcal{Y}_0)$ with functional uncertainty:*

$$\Delta_j \subset \Delta(L^2(\Omega; w_{2j}), \delta_{2j}) \cap \Delta(L^\infty(\Omega; w_{\infty j}), \delta_{\infty j}), \quad 1 \leq j \leq p, \quad (36)$$

⁶ The factor of 2 in (33) is purely for mathematical convenience, if best approximations exist, then the factor could be removed, and, in general, the factor could be taken to be any number strictly larger than 1.

where w_{2j} , δ_{2j} and $w_{\infty j}$ are known and the cost function $\mathcal{P}_{(\Sigma_{\Delta}(\mathcal{Y}_0), \Xi(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0, \Omega_1)$ is defined by (18). Let

$$\Xi(\cdot) = \Xi(\Gamma, \alpha, \Phi, \Omega_0)(\cdot), \quad (37)$$

and define

$$W_{\alpha} = \max\{\bar{\lambda}(P)(\gamma_0 + \gamma_1)^2, \eta_0^2\} + \sum_{j=1}^p \frac{1}{2\alpha_j} \frac{(\delta_{2j} + \sup_{f_j \in \Delta} \|d_{f_j}\|_{L^2(\Omega; w_{2j})})^2}{\underline{\lambda}(\Gamma_j)\underline{\lambda}(G_j)}, \quad (38)$$

where G_j is the $L^2(\Omega; w_{2j})$ Gram matrix of the model component ϕ_j , and γ_0, γ_1 define the constraints on \mathcal{Y}_0 and \mathcal{Y}_{ref} , respectively (see (11) and (14)). Further, P is defined by (30). If

1. η_0 satisfies the inequality:

$$\eta_0 \geq \frac{2\|b\|}{\underline{\lambda}(Q)} \sup_{f \in \Delta} \|d_f\|_{C(\Omega, \mathbb{R}^p)}, \quad (39)$$

2. $0 \in \Omega_0 \subset \Omega_1 \subset \Omega$,
3. the adaption gain $\alpha > 0$ is such that

$$\left\{ z \in \mathbb{R}^n \mid z^T P z \leq \left(\sqrt{W_{\alpha}} + \sqrt{\bar{\lambda}(P)\gamma_1} \right)^2 \right\} \subset \Omega^{\circ}, \quad (40)$$

then:

1. $(\Sigma_{\Delta}(\mathcal{Y}_0), \Xi(\mathcal{Y}_{\text{ref}}))$ is well-posed; for any reference signal $y_{\text{ref}} \in \mathcal{Y}_{\text{ref}}$, $x(t) \rightarrow \Omega_0$ as $t \rightarrow \infty$, where $x(t)$ is defined by (16). Also, $Y(t) \in \Omega$ for all $t \geq 0$,
- 2.

$$\mathcal{P}_{(\Sigma_{\Delta}(\mathcal{Y}_0), \Xi(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0, \Omega_1) < \infty. \quad (41)$$

Proof. This result is obtained by a Lyapunov analysis, and is a simple extension of the stabilisation results of [F] and [FSR2]; hence we only sketch the proof. In particular, the well-posedness of the system is not dealt with here.

Let $y_{\text{ref}} \in \mathcal{Y}_{\text{ref}}$, $f \in \Delta$, $y_0 \in \mathcal{Y}_0$, and initially suppose $x_0 \notin \Omega_0$. Consider $(\Sigma_f(y_0), \Xi(y_{\text{ref}}))$ and define $V: \mathbb{R}^n \times \mathcal{W}^{m_1} \times \dots \times \mathcal{W}^{m_p} \rightarrow \mathbb{R}$ by

$$V(x, \hat{\Theta}) = x^T P x + \sum_{j=1}^p \frac{1}{2\alpha_j} (\theta_j - \hat{\theta}_j)^T \Gamma_j^{-1} (\theta_j - \hat{\theta}_j), \quad (42)$$

where θ_j is defined by (33). Let $V_t = V(x(t), \hat{\Theta}(t))$, and T_{Ω_i} , $i = 0, 1$, be defined by (19). Some straightforward inequalities establish that $V_0 \leq W_{\alpha}$. If $Y \in \Omega$ and $x \notin \Omega_0$, then

$$\dot{V}_t = -x^T Q x + x^T b d_f(Y) \leq -(\underline{\lambda}(Q) - g) \|x\|^2, \quad (43)$$

where $g = \|b\| \|d_f\|_{C(\Omega, \mathbb{R}^p)} / \text{diam}(\Omega_0)$. Since by assumption $g \leq \underline{\lambda}(Q)/2$, it follows that the right-hand side of (43) is negative. Hence by condition (40), and a level set argument, we can see that the V_0 level set of V is invariant, hence the closed-loop signals are bounded: in particular $Y(t)$ is bounded by Ω . By definition of the dead-zone, we can establish that V_t is decreasing on T_{Ω_0} and hence by inequality (43) that $x(t) \rightarrow \Omega_0$. The boundedness of the performance follows from the continuity

of x, u , the fact that $x(t), u(t)$ can be bounded in terms of V_t (and hence uniformly by W_α), the measurability of T_{Ω_0} and T_{Ω_1} and the fact that

$$m(T_{\Omega_1}) \leq m(T_{\Omega_0}) \leq \frac{V_0 - \inf_{t \in T_{\Omega_0}} V_t}{\inf_{t \in T_{\Omega_0}} |\dot{V}(t)|} < \infty. \quad (44)$$

The case $x_0 \in \Omega_0$ follows similarly by showing $V_{t^*} \leq W_\alpha$ where $t^* = \inf\{t \geq 0 \mid x(t) \notin \Omega_0\}$. ■

The above theorem is given weight by the fact that by taking Ω to be compact, there are a wide variety of finite-dimensional models satisfying the conditions of the theorem. Any $\{K_\beta\}_{\beta \geq 0}$ approximate model satisfies inequality (39) as $m \rightarrow \infty$, and by choosing α , condition (40) can be satisfied, provided

$\exists \varepsilon > 0$ such that

$$\left\{ z \in \mathbb{R}^n \mid z^T P z \leq \left(\max \left\{ \sqrt{\bar{\lambda}(P)}(\gamma_0 + \gamma_1), \eta_0 \right\} + \sqrt{\bar{\lambda}(P)}\gamma_1 \right)^2 + \varepsilon \right\} \subset \Omega. \quad (45)$$

However, it is important to observe that a large $\alpha > 0$ could lead to excessive control effort. Note that condition (40) restricts the L^2 uncertainty level for a bounded Ω ; condition (40) is a sufficient condition to ensure the state remains in the region where the approximation accuracy is small and clearly a greater uncertainty level will generally lead to a worse output transient for a fixed $\alpha > 0$. Violation of condition (40) can easily lead to complete instability: see [FSR2] for an example.

Note further that this theorem is completely constructive, all the required gains can be computed from the conditions. Explicit bounds for $\mathcal{P}(Q, R, \Omega_0, \Omega_0)$ for the case of stabilisation can be found in [FSR2], these can easily be generalised to the situation considered here.

2.6. Formulation of the Main Results

The problem we now consider is as follows: given a fixed uncertainty Δ , does the performance degrade as the dimensionality of the model increases? In a trivial manner performance can diverge if increasing the dimension of the model causes a loss of stability in the system, by violating condition (40). However, as noted above, if condition (45) is satisfied, then stability can always be maintained by choosing the adaption gain appropriately, so the question we address is whether the performance degrades irrespective of the choice of gains.

Section 3 shows that the performance degrades as the resolution of standard model-based controllers increases. We consider Scenario 1 where we are interested in *decreasing the asymptotic L^∞ error ε by increasing the model resolution (m)*. Increasing the model's resolution reduces the approximation error, and permits a smaller dead-zone (Ω_0^m), hence leading to improved asymptotic L^∞ tracking. Specifically we construct:

1. a class of models Φ^m with localised basis functions,
2. a corresponding set of decreasing L^∞ tracking requirements specified by the sets $\{\Omega_0^m\}_{m \geq 1}$ ($\Omega_0^m \rightarrow \{0\}$ as $m \rightarrow \infty$),
3. a nontrivial choice of Δ , \mathcal{Y}_{ref} satisfying the constraints listed in Section 2.2,

for which the associated controller given by

$$\Xi_m(\cdot) = \Xi(I, \alpha^m, \Phi^m, \Omega_0^m)(\cdot) \quad (46)$$

has the properties that there exists $M \geq 1$ and a positive sequence $\{\alpha^m\}_{m \geq 1}$ such that

$$\mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) < \infty, \quad \forall m \geq M, \quad (47)$$

but *for all* positive sequences $\{\alpha^m\}_{m \geq 1}$ the performance diverges as $m \rightarrow \infty$:

$$\limsup_{m \rightarrow \infty} \mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) = \infty. \quad (48)$$

In Section 4 we give a general construction for a model class whose associated controller performance is uniformly bounded as the model resolution increases, i.e. we construct a class of models $\{\Phi^m\}_{m \geq 1}$ where the associated controller

$$\Xi_m(\cdot) = \Xi(\Gamma^m, \alpha^m, \Phi^m, \Omega_0^m)(\cdot) \quad (49)$$

is such that

$$\limsup_{m \rightarrow \infty} \mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) < \infty. \quad (50)$$

Thus whilst there may be a price associated with increasing the resolution of the model; the worst scenario of divergence can be avoided. In this case Ω_0 can either be fixed independent of m , e.g. $\Omega_0^m = \Omega_1$, or can, e.g. have the property that $\Omega_0^m \rightarrow \{0\}$ as $m \rightarrow \infty$. In the former case the interest in the asymptotic result arises from quantifying whether an overly large model may degrade the performance. As observed previously, overly large models are often utilized as in Scenario 2 or 3 of the Introduction (i.e. due to conservatism in the dimension function ρ or due to the inherent difficulties of determining β). In the latter case we consider Scenario 1, where the resolution is increased to improve the asymptotic tracking accuracy.

3. Divergent Performance of Mono-Resolution Models

In this section we develop a class of examples which have divergent closed-loop performance as the model dimension is increased, irrespective of the choice of the adaptive gain.

3.1. Model Structure

We first construct a multivariate model on the domain

$$\mathcal{V} = [-\eta, \eta]^{n_1} \times \cdots \times [-\eta, \eta]^{n_p}. \quad (51)$$

We assume the following properties on the model structure:

1. (*Basis function*). Let $F: \mathbb{R} \rightarrow \mathbb{R}$ be such that F is continuous, F has a maximum at 0, $F(0) > 0$, and $F(x) \geq 0, \forall x \in \mathbb{R}$.
2. (*Tensor product*). Define $s^m: \mathbb{R} \rightarrow \mathbb{R}^{2m+1}$ by $s^m = (\varphi_{-m}^m, \varphi_{-m+1}^m, \dots, \varphi_{m-1}^m, \varphi_m^m)$ for $m \geq 1$ where

$$\varphi_i^m(x) = F(mx - \eta i), \quad -m \leq i \leq m. \quad (52)$$

We define a multivariate model $\Phi^m: \mathbb{R}^n \rightarrow (\mathbb{R}^{(2m+1)^n})^p$, $\Phi^m = (\phi_1^m, \dots, \phi_p^m)^T$, $\phi_j^m: \mathbb{R}^n \rightarrow \mathbb{R}^{(2m+1)^n}$, $1 \leq j \leq p$, by the tensor product construction:

$$\phi_j^m(x) = \phi_j^m(x_1, \dots, x_n) = s^m(x_1) \otimes s^m(x_2) \otimes \dots \otimes s^m(x_n) \quad (53)$$

(note that $\phi_j^m = \phi_k^m$ for $1 \leq j, k \leq p$), so that for some suitable bijection

$$\xi: \{1, 2, \dots, (2m+1)^n\} \rightarrow \{-m, -m+1, \dots, m\}^n \quad (54)$$

we can write

$$(\phi_j^m)_k(x_1, \dots, x_n) = \varphi_{\xi(k)_1}^m(x_1) \varphi_{\xi(k)_2}^m(x_2) \dots \varphi_{\xi(k)_n}^m(x_n). \quad (55)$$

The corresponding weights are denoted by $\hat{\theta}_j^m \in \mathcal{W}_{m_j} = \mathbb{R}^{(2m+1)^n}$. We let $\xi: \mathbb{N} \rightarrow \mathbb{N}^n$ denote the map $\xi(k) = (i_1, \dots, i_n)$.

3. (*Linear Independence*). The functions $\{(\phi_j^m)_k|_{\Omega_2}: \Omega_2 \rightarrow \mathbb{R}\}_{-m \leq k \leq m}$ are linearly independent for all $1 \leq j \leq p$ (recall that Ω_2 is specified by (15)).
4. (*Uniform strength*). There exist constants Υ_1, Υ_2 such that for all $m \geq 1$,

$$0 < \Upsilon_1 \leq \inf_{x \in \Omega_2} |1^T \phi_j^m(x)| \leq \sup_{x \in \Omega_2} |1^T \phi_j^m(x)| \leq \Upsilon_2, \quad 1 \leq j \leq p. \quad (56)$$

Note that condition 4 implies an exponential spatial decay rate for the function F .

3.1.1. Examples: B-Spline and Gaussian RBF Networks

Many common models utilised in approximate adaptive designs satisfy the above assumptions. In particular consider the Gaussian radial basis function networks defined on the regular grid:

$$\mathcal{L}_j(m) = \{-\eta, \dots, -\eta + k\eta/m, \dots, \eta - \eta/m, \eta\}^n, \quad (57)$$

$$\mathcal{L}(m) = \mathcal{L}_1(m) \times \mathcal{L}_2(m) \times \dots \times \mathcal{L}_p(m) \quad (58)$$

and obtained by taking

$$F(x) = \exp(-\rho x^2), \quad (59)$$

and then following the tensor-product construction of the previous section (see conditions 1 and 2). This yields the familiar basis functions with the scaling of, e.g. [SS]:

$$\phi_{jk}^m = \exp\left(-m^2 \rho \left\|x - \frac{\eta}{m} \xi(k)\right\|^2\right). \quad (60)$$

Here $\xi(k) \in \{-m, -m+1, \dots, m\}^n$ is viewed as an element of \mathbb{N}^n .

Condition 3 follows from the linear independence of $\{(\phi_j^m)_k: \mathbb{R}^n \rightarrow \mathbb{R}\}_{-m \leq k \leq m}$ and the analyticity of the Gaussian function. Condition 4 follows from the inequalities

$$0 < \exp(-n\rho\eta) \leq |1^T \phi_j^m(x)| \leq |1^T \phi_j^m(0)| \leq 1 + \sum_{j=1}^{\infty} \exp(-\rho\eta^2 j^2), \quad \forall x \in \mathbb{R}^n, \quad (61)$$

since the right-hand side of the final inequality clearly does not depend on m and is easily shown to be convergent by, e.g. the ratio test. Thus the standard lattice-based Gaussian RBFs satisfy conditions 1–4.

For the k th-order B-spline networks we take F to be the k -fold convolution of the characteristic function on $[-\frac{1}{2}, \frac{1}{2}]$ with itself. The multivariate tensor product B-splines are then given by following the tensor-product construction of conditions 1 and 2 of the previous section. The linear independence of the B-spline basis functions over their domain of definition is a standard fact, hence condition 3 follows if $\Omega_2 = \mathfrak{U}$. Condition 4 follows from the fact that B-splines form a partition of unity.

3.2. System Structure

We take $p \geq 2$, $n_j = 1$ for $1 \leq j \leq p$. The initial condition set, \mathcal{Y}_0 , and reference signal set, \mathcal{Y}_{ref} , are defined by $\gamma_0, \gamma_1, \gamma_2 > 0$. Let $\{\varepsilon_m\}_{m \geq 1}$ be a sequence of approximation error bounds where $\varepsilon_m > 0$, $\forall m \geq 1$, and where ε_m monotonically decreases to zero. The trick is to define the smoothness class $\{K_\beta\}_{\beta \geq 0}$ so that the ε_m is an approximation error bound for the model ϕ_j^m , $1 \leq j \leq p$. In particular, we define the smoothness class $\{K_\beta\}_{\beta \geq 0}$ as follows:

$$K_\beta = K'_\beta \times \cdots \times K'_\beta,$$

where

$$K'_\beta = \bigcap_{m \geq \beta} \{f \in C(\Omega) \mid \text{supp } f \subset \mathfrak{U}, f = \theta^T \phi_j^m + d, \theta \in \mathcal{W}_m, \|d\|_{C(\Omega)} \leq \varepsilon_m\}. \quad (62)$$

The uncertainty set is taken to be

$$\Delta_j = K'_\beta \cap \Delta(L^2(\Omega); \delta_{2j}) \cap \Delta(L^\infty(\Omega); \delta_{\infty j}), \quad \delta_{2j}, \delta_{\infty j} > 0, \quad \beta \geq 0, \quad 1 \leq j \leq p, \quad (63)$$

where $\Omega = \mathbb{R}^n$. This is equivalent to the knowledge of a dimension function ρ with the property that

$$\rho(\beta, \varepsilon_m) = \dim \mathcal{W}_m, \quad \forall m \geq 1. \quad (64)$$

If $\text{span } \Phi^m \subset \text{span } \Phi^{m+1}$ (such as with B-splines defined on refined lattices) it is straightforward to observe that the smoothness class K_β is non-trivial for any sequence $\{\varepsilon_m\}_{m \geq 1}$ and $\text{span } \Phi^m \subset K_\beta$. On the other hand, if $\text{span } \Phi^m \not\subset \text{span } \Phi^{m+1}$ as for the Gaussian RBF model, then the rate of decay is critical in establishing that $\text{span } \Phi^m \subset K_\beta$. For the remainder of this section we assume that $\text{span } \Phi^m \subset K_\beta$ for all $m \geq 1$.

We have introduced a requirement that the supports of the nonlinearities are contained in \mathfrak{U} . This is simply for convenience, so that the stability at large can be established independently of the choice of α . In particular it allows us to take $\Omega_2 = \mathfrak{U}$ (hence that condition 3 can be verified for, e.g. the compactly supported B-spline basis), without introducing instability (due to the violation of condition (40)).

3.3. Controller and Cost Functional

The performance is measured by the cost functional $\mathcal{P}(Q, R, \Omega_0^m, \Omega_1)$. Ω_1 is defined by η_1 (see (34)) and is fixed independently of m , whereas Ω_0^m is defined by η_0^m :

$$\eta_0^m = \frac{2\|b\|\varepsilon_m}{\lambda(Q)}, \quad (65)$$

where ε_m is as in Section 3.2. Given a model structure $\{\Phi^m\}_{m \geq 1}$ as defined in Section 3.1, the controllers $\{\Xi_m(\cdot)\}_{m \geq 1}$ are defined by

$$\Xi_m(\cdot) = \Xi(I, \alpha^m, \Phi^m, \Omega_0^m)(\cdot). \quad (66)$$

3.4. A Theorem Proving the Lack of Scalability

We now demonstrate that the class of examples defined above suffer from divergent closed-loop behaviour as $m \rightarrow \infty$. By applying Theorem 2.1 to the closed loop $(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))$ defined by Sections 3.2 and 3.3, we can show that there exists a positive sequence $\{\alpha^m\}_{m \geq 1}$ such that

$$\mathcal{P}(Q, R, \Omega_0^m, \Omega_1) < \infty, \quad \forall m \geq \beta. \quad (67)$$

Therefore there is a choice of adaption gain which gives a finite performance at each model resolution; we now establish that there can be no uniform bound.

Firstly we give a lower bound for the control cost:

Proposition 3.1. *Consider the closed-loop system $(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))$ defined by Sections 3.2 and 3.3. Then*

$$\mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(0, R, \Omega_0^m, \Omega_1) \geq O\left(\sqrt{\frac{\alpha_j^m}{m}}\right), \quad 1 \leq j \leq p. \quad (68)$$

Proof. See the Appendix. ■

The important feature of this bound is that if we can force α_j^m to have greater than linear growth in m , then the control effort diverges. Therefore we next address the scaling of the state performance. In the following results we are able to compute the state cost accurately, and show that α_j^m must be selected to have greater than linear growth to stop the state performance from diverging. This essentially shows that divergence of the full cost is inevitable irrespective of the choice of α_j^m .

To compute the state performance accurately we exploit the fact that persistently exciting reference signals can cause parameter convergence. An extension of a well-known calculation [KKK] then computes the state performance accurately (see the proof of Theorem 3.6). Recall the following definition and theorem:

Definition 3.2. A function $\xi: \mathbb{R}_+ \rightarrow \mathbb{R}^m$ is said to be (τ, w_1, w_2) *persistently exciting* if there exists $T > 0$ such that for all $t \geq T$ and for all $c \in \mathbb{R}^m$,

$$w_2 \|c\|^2 \geq \int_t^{t+\tau} (c^T \xi(t))^2 dt \geq w_1 \|c\|^2. \quad (69)$$

Theorem 3.3. *Consider the (unperturbed) system:*

$$\begin{aligned} \dot{x}_j &= A_{*j} x_j + B_j ((\theta_j - \hat{\theta}_j)^T \phi_j(Y^{\text{ref}})), & 1 \leq j \leq p, \\ \dot{\hat{\theta}}_j &= \alpha_j x_j^T b_j \phi_j(Y^{\text{ref}}), & 1 \leq j \leq p. \end{aligned} \quad (70)$$

If $\phi_j(Y^{\text{ref}})$ is persistently exciting, then $(x_j(t), (\theta_j - \hat{\theta}_j)(t))$ is globally exponentially stable.

Proof. This is the standard result on the exponential stability of systems with persistently excited regressors, see, e.g. Theorem 2.6.5 of [SB]. ■

To be able to establish parameter convergence for our controllers, we need a similar result in the presence of a small dead-zone. The following result establishes the practical convergence of the parameter estimators for sufficiently small dead-zones.

Proposition 3.4. *Let $m \geq 1$, $\alpha^m > 0$, $\kappa > 0$, $\rho_m > 0$ be arbitrary. Let $(Y^m, \hat{\theta}^m)$ ($1 \leq j \leq p$) denote the solution of the closed-loop system $(\Sigma_{f^m}, \Xi_m(Y_m^{\text{ref}}))$, where $f^m = (f_1^m, \dots, f_p^m)^T$, $f_j^m = (\theta_j^m)^T \phi_j^m$, $\theta_j^m = \kappa 1$. Assume further that the reference signal Y_m^{ref} is such that $\phi_j^m(Y_m^{\text{ref}})$ is persistently exciting. Then there exists $\bar{\epsilon}_m > 0$ such that if $\eta_0^m \leq \bar{\epsilon}_m$, then*

$$\limsup_{t \rightarrow \infty} \|\hat{\theta}_j^m(t) - \theta_j^m\| \leq \rho_m, \quad 1 \leq j \leq p, \quad \forall m \geq \beta. \quad (71)$$

Proof. See the Appendix. ■

The critical step is to establish the existence of a persistently exciting reference signal $Y^{\text{ref}} \in \mathcal{Y}_{\text{ref}}$. The proof of the following result constructs (for fixed m) a signal $y_{\text{ref}} \in \mathcal{Y}_{\text{ref}}$ which guarantees that the regressors $\phi_j^m(Y)$, $1 \leq j \leq p$, are persistently exciting. The construction is an extension of the ideas in [KNW].

Proposition 3.5. *Let $m \geq \beta$. Then there exists a reference trajectory $Y_m^{\text{ref}} \in \mathcal{Y}_{\text{ref}}$ and $\tilde{\epsilon}_m > 0$ such that if $Y: \mathbb{R}_+ \rightarrow \mathbb{R}^n$ is continuous and*

$$\limsup_{t \rightarrow \infty} \|Y(t) - Y_m^{\text{ref}}(t)\| \leq \tilde{\epsilon}_m,$$

then $\phi_j^m(Y(\cdot))$ is persistently exciting for $1 \leq j \leq p$.

Proof. See the Appendix. ■

We now give the main result of this section of the paper.

Theorem 3.6. *There exists a sequence $\{\epsilon_m\}_{m \geq 1}$ defining both the smoothness class K_β and the dead-zone by (62) and (65), respectively, such that if*

$$\text{span } \Phi^m \subset K_\beta, \quad \forall m \geq \beta, \quad (72)$$

and Δ, Ξ^m are defined by

$$\Delta_j = K_\beta' \cap \Delta(L^2(\Omega); \delta_{2j}) \cap \Delta(L^\infty(\Omega); \delta_{\infty j}), \quad 1 \leq j \leq p, \quad (73)$$

$$\Xi_m(\cdot) = \Xi(I, \alpha^m, \Phi^m, \Omega_0^m)(\cdot), \quad (74)$$

then

1. there exists a positive sequence $\{\alpha^m\}_{m \geq 1}$ such that

$$\mathcal{P}_{(\Sigma_\Delta(\mathcal{U}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) < \infty, \quad \forall m \geq \beta, \quad (75)$$

but,

2. for all sequences $\{\alpha^m\}_{m \geq 1}$,

$$\mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) \geq O(m^{(p-1)/3}). \quad (76)$$

Proof. Define the sequence $\{\varepsilon_m\}_{m \geq 1}$ by

$$\varepsilon_m = \min\left(\frac{\underline{\lambda}(\mathcal{Q})}{2\|b\|} \tilde{\varepsilon}_m, \bar{\varepsilon}_m\right), \quad (77)$$

where $\tilde{\varepsilon}_m$ and $\bar{\varepsilon}_m$ are defined by Propositions 3.5 and 3.4, respectively.

Inequality (75) follows from Theorem 2.1. To establish inequality (76) we first consider the state performance. We apply the construction of Proposition 3.4 together with Proposition 3.5. For each $m \geq 1$ let $f_j^m = \kappa 1^T \phi_j^m$, where $\kappa > 0$ is chosen such that $f_j^m \in \Delta$ (e.g. one may choose $\kappa = 1/\Upsilon_2 \min(\delta_\infty, \delta_2/m(\Omega))$). Let Y_m^{ref} be the reference signal whose existence is proved in Proposition 3.5 and let $(Y, \hat{\theta})$ be the solution of the closed-loop system $(\Sigma_{f^m}(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))$. Then $\limsup_{t \rightarrow \infty} \|Y_m^{\text{ref}}(t) - Y(t)\| \leq \eta_0^m \leq \tilde{\varepsilon}_m$ and thus $\phi_j^m(Y(t))$ is persistently exciting, by Proposition 3.5.

Therefore taking $\rho_m = \rho > 0$ and applying Proposition 3.4, we have

$$\begin{aligned} \int_{T_{\Omega_0^m}} x^T Q x \, dt &= \int_{T_{\Omega_0^m}} -\dot{V} \, dt \\ &= V(0) - \limsup_{t \rightarrow \infty} V(t) \\ &= x_0^T P x_0 - \frac{2\|b\|\varepsilon_m}{\underline{\lambda}(\mathcal{Q})} \\ &\quad + \limsup_{t \rightarrow \infty} \sum_{j=1}^p \frac{1}{2\alpha_j^m} (\theta_j^T \theta_j - (\theta_j - \hat{\theta}_j(t))^T (\theta_j - \hat{\theta}_j(t))) \\ &= x_0^T P x_0 - \frac{2\|b\|\varepsilon_m}{\underline{\lambda}(\mathcal{Q})} + \limsup_{t \rightarrow \infty} \sum_{j=1}^p \frac{1}{2\alpha_j^m} (\theta_j^T \theta_j - \rho) \\ &= O\left(m^p \sum_{j=1}^p \frac{1}{2\alpha_j^m}\right) \\ &= O\left(\frac{m^p}{\alpha_j^m}\right), \quad 1 \leq j \leq p. \end{aligned} \quad (78)$$

However, by Proposition 3.1,

$$\mathcal{P}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) \geq O\left(\sqrt{\frac{\alpha_j^m}{m}}\right), \quad 1 \leq j \leq p. \quad (79)$$

Combining this with (78) we have

$$\begin{aligned} \mathcal{P}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) &\geq O\left(\max\left(\sqrt{\frac{\alpha_j^m}{m}}, \frac{m^p}{\alpha_j^m}\right)\right) \\ &\geq O(m^{(p-1)/3}), \quad 1 \leq j \leq p, \end{aligned} \quad (80)$$

as required. ■

We have therefore established that a wide class of models lead to divergent closed-loop performance when the resolution of the approximator is increased.

It is fair to criticise the construction due to the excessively fast rate of convergence required for ε_m and hence the required rate of shrinking of the dead-zone. Indeed, it can be argued that unless η_0^m and hence ε_m can be taken to have a polynomial decay, then the smoothness class is not prototypical of standard smoothness classes widely considered in the approximation literature, where we can expect the approximation errors ε_m to be of the order of $(1/m^p)^\beta$, where $\beta \geq 0$ is the measure of the smoothness. For example, if the decay is superpolynomial, then the Gaussian RBF considered previously yields the smoothness class $K'_\beta = \{0\}$ which clearly does not satisfy the requirement that $\text{span } \Phi^m \subset K_\beta, \forall m \geq \beta$. We formalise this as a conjecture:

Conjecture 3.7. *The decay rate of ε in the definition of the smoothness class K in Theorem 3.6 can be taken to be polynomial.*

We consider the resolution of this conjecture to be an important open issue in this field.

Of course, as observed above, model classes such as the B-splines satisfy the requirements of the theorem regardless of the decay rate, and thus we have indicated that the scaling problem is a real issue in these Lyapunov-based designs.

4. A General Design for Resolution Scalability

In this section we give a general construction for resolution scalability based on any approximate model class derived from basis extensions. Since the results will hold for any sequence of dead-zones $\Omega_0^m \subset \Omega$ of the form of (34), it follows that the results are applicable in Scenarios 1–3.

Let $\{\tilde{\Phi}^k: \Omega \rightarrow \mathcal{W}_k\}_{k \geq 1}$ be a sequence of models, and define

$$\Phi^m = [\tilde{\Phi}^1 | \tilde{\Phi}^2 | \dots | \tilde{\Phi}^m]. \quad (81)$$

We assume that $\{\Phi^m\}_{m \geq 1}$ is a $\{K_\beta\}_{\beta \geq 0}$ approximate model class. Note that we are imposing a considerable structure on the model class, for example we have ruled out the mono-resolution model class of the previous section. We are essentially requiring that the higher resolution models are simply basis extensions of the lower resolution models, hence standard approximation bases such as Taylor series, Fourier series, wavelets, etc., are all of the appropriate form. Basis

functions from any $\{K_\beta\}_{\beta \geq 0}$ approximation class can be utilized, for if $\{\tilde{\Phi}^k\}_{k \geq 1}$ is a $\{K_\beta\}_{\beta \geq 0}$ approximation class, then so is $\{\Phi^m\}_{m \geq 1}$, but note that Φ^m is of much higher dimension than $\tilde{\Phi}^m$ as it is the union of all the lower resolution models. The important difference between the requirement we are making in this section and the conditions imposed in the previous section is that the models contain basis functions corresponding to all resolutions, whereas the models of the previous section are “mono-resolution”. For simplicity, we assume that the components of the model are all equal, i.e. $\phi_j^m = \phi_i^m$ for $1 \leq i, j \leq p$.

Firstly we treat the case of compact Ω . We define

$$\Gamma_j^m = \text{diag}\{\omega_1^2 I_{\dim \mathcal{W}_1 \times \dim \mathcal{W}_1}, \omega_2^2 I_{\dim \mathcal{W}_2 \times \dim \mathcal{W}_2}, \dots, \omega_m^2 I_{\dim \mathcal{W}_m \times \dim \mathcal{W}_m}\}, \quad (82)$$

$$\Gamma^m = \text{diag}\{\Gamma_1^m, \dots, \Gamma_p^m\}, \quad (83)$$

where $\{\omega_i\}_{i \geq 1}$ is the positive sequence defined by

$$\omega_i = \frac{|c_i|}{\sup_{x \in \Omega} \|\tilde{\phi}^i(x)\|}. \quad (84)$$

Here $\{c_i\}_{i \geq 1}$ is any sequence for which $\sum_{i=1}^{\infty} |c_i| < +\infty$. Without loss of generality we may assume that $\sum_{i=1}^{\infty} |c_i| = 1$.

Theorem 4.1. *Let $\Omega \subset \mathbb{R}^n$ be a fixed compact set, and let $\{K_\beta\}_{\beta \geq 0}$ be a smoothness class. Suppose $\{\Phi^m\}_{m \geq 1}$ is a $\{K_\beta\}_{\beta \geq 0}$ approximate model class. Consider the system $\Sigma_\Delta(\mathcal{Y}_0)$ with functional uncertainty:*

$$\Delta_j \subset K_\beta \cap \Delta(L^2(\Omega; w_{2j}), \delta_{2j}) \cap \Delta(L^\infty(\Omega; w_{\infty j}), \delta_{\infty j}), \quad 1 \leq j \leq p, \quad (85)$$

where δ_{2j} , ω_{2j} and $\omega_{\infty j}$ are known. Let

$$\Xi_m(\cdot) = \Xi_m(\Gamma^m, \alpha, \Phi^m, \Omega_0^m)(\cdot), \quad (86)$$

where Γ^m and Φ^m are as in (81)–(83). If

1. $0 \in \Omega_0^m \subset \Omega_1 \subset \Omega$, $\forall m \geq 1$,
2. $\exists \varepsilon > 0$ such that

$$\left\{ z \in \mathbb{R}^n \mid z^T P z \leq \left(\max \left\{ \sqrt{\bar{\lambda}(P)}(\gamma_0 + \gamma_1), \eta_0 \right\} + \sqrt{\bar{\lambda}(P)}\gamma_1 \right)^2 + \varepsilon \right\} \subset \Omega, \quad (87)$$

then there exists $\alpha > 0$ and $M \geq 1$ such that, for all $m \geq M$,

1. $(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))$ is well-posed, $x(t) \rightarrow \Omega_0$ as $t \rightarrow \infty$; $Y(t) \in \Omega$ for all $t \geq 0$,
- 2.

$$\limsup_{m \rightarrow \infty} \mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) < \infty. \quad (88)$$

Proof. Choose any M satisfying

$$M \geq \rho \left(\beta, \eta_0 \frac{\lambda(\mathcal{Q})}{2\|b\|} \right) \quad (89)$$

and define

$$W_\alpha^* = \max\{\bar{\lambda}(P)(\gamma_0 + \gamma_1)^2, \eta_0^2\} + \sum_{j=1}^p \frac{1}{2\alpha_j} \frac{(\delta_{2j} + q_M)^2}{\underline{\lambda}(\Gamma^M)\underline{\lambda}(G^M)}, \quad (90)$$

where

$$q_M = \max_{1 \leq j \leq p} \sup_{f_j \in \Delta_j} \|d_{f_j}\|_{L^2(\Omega; w_{2j})} \leq s_M \max_{1 \leq j \leq p} \|w_{2j}\|_{L^1(\Omega)} < \infty, \quad (91)$$

and $s_M > 0$ is such that $M \geq \rho(\beta, s_M)$. It therefore follows by condition (87) that there exists an adaption gain $\alpha > 0$ such that

$$\{z \in \mathbb{R}^n \mid z^T P z \leq W_\alpha^* + \bar{\lambda}(P)\gamma_1^2\} \subset \Omega. \quad (92)$$

Let $f \in \Delta$, $Y^{\text{ref}} \in \mathcal{Y}_{\text{ref}}$, $y_0 \in \mathcal{Y}_0$ and initially assume that $x_0 \notin \Omega_0$. First we consider well-posedness and convergence to Ω_0 . Let $m \geq M$. Let V^m be the Lyapunov function of Theorem 2.1 corresponding to the m th model. Following the proof of Theorem 2.1, it suffices to show that the level set defined by V_0^m is contained in Ω . From (92) it suffices to show that

$$V_0^m \leq W_\alpha^*, \quad (93)$$

where V_0^m is the value of V^m at time zero.

Since the model Φ^m for $m > M$ is simply a basis extension of the model Φ^M , we can define $\theta_f^m \in \mathcal{W}_1 \times \cdots \times \mathcal{W}_m$ by $\theta_f^m = [\theta_f^M | 0]$, where θ_f^M is a parameter-vector that satisfies

$$\|d_f^M\|_{C(\Omega)} = \|f - (\theta_f^M)^T \Phi^M\|_{C(\Omega)} \leq \eta_0 \frac{\underline{\lambda}(Q)}{2\|b\|}. \quad (94)$$

Since $(\theta_j^m)^T (\Gamma_j^m)^{-1} \theta_j^m = (\theta_j^M)^T (\Gamma_j^M)^{-1} \theta_j^M$, $(\theta_j^M)^T G^M \theta_j^M \leq (\delta_{2j} + q_M)^2$, we have

$$(\theta_j^m - \hat{\theta}_j^m(0))^T (\Gamma_j^m)^{-1} (\theta_j^m - \hat{\theta}_j^m(0)) \leq \frac{(\delta_{2j} + q_M)^2}{\underline{\lambda}(\Gamma_j^M)\underline{\lambda}(G^M)}, \quad (95)$$

and so $V_0^m \leq W_\alpha^*$ as required.

We now consider the state performance bound (to avoid an explosion of indices, we omit the index denoting the dependency on the model-size unless the dependency is crucial). We follow the derivation of [FSR2], which we repeat here for completeness. By (43),

$$\mathcal{P}' \stackrel{\text{def}}{=} \int_{T_{\Omega_0}} x^T(t) Q x(t) dt = \int_{T_{\Omega_0}} -\dot{V}_t dt + \int_{T_{\Omega_0}} x^T b d(x) dt. \quad (96)$$

We consider the first term of the right-hand side. Since $x(t)$ is continuous, we can write $T_{\Omega_0} = \bigcup_{b \in B} E_b$ where $E_b = (t_b^-, t_b^+)$ are maximal disjointed connected subsets of \mathbb{R} , and define $B_n = \{b \in B \mid m(E_b) \geq 1/n\}$. By the definition of the dead-zone:

$$\int_{\bigcup_{b \in B_n} E_b} -\dot{V}_t dt = \sum_{b \in B_n} V_{t_b^-} - V_{t_b^+} \leq V_0 - \eta_0^2, \quad (97)$$

since the summation telescopes and $V_t \geq \eta_1^2$ for all $t \geq 0$. Then applying the monotone convergence theorem we obtain

$$\int_{T_{\Omega_0}} \dot{V}_t dt = \lim_{n \rightarrow \infty} \int_{\bigcup_{b \in B_n} E_b} -\dot{V}_t dt \leq V_0 - \eta_0^2. \quad (98)$$

Then since

$$|x^T b d(x)| \leq g \|x\|^2 \leq \frac{g}{\underline{\lambda}(\mathcal{Q})} x^T \mathcal{Q} x, \quad (99)$$

we have

$$\begin{aligned} \mathcal{P}' &\leq (V_0 - \eta^2) + \int_{T_{\Omega_0}} x^T b d(x) dt \leq (V_0 - \eta_0^2) + \frac{g}{\underline{\lambda}(\mathcal{Q})} \int_{T_{\Omega_0}} \|x\|_{\mathcal{Q}}^2 dt \\ &\leq (V_0 - \eta_0^2) + \frac{g}{\underline{\lambda}(\mathcal{Q})} \mathcal{P}'. \end{aligned} \quad (100)$$

Rearranging the terms yields

$$\mathcal{P}' \leq \frac{\underline{\lambda}(\mathcal{Q})}{\underline{\lambda}(\mathcal{Q}) - g} (V_0 - \eta_0^2) \leq \frac{\underline{\lambda}(\mathcal{Q})}{\underline{\lambda}(\mathcal{Q}) - g} (W_\alpha^* - \eta_0^2), \quad (101)$$

so we are left with estimating the control effort integral. The idea is to change the integration over T_{Ω_1} to an integration over $[\eta_1^2, V_0]$ as a function of V_t . We first establish an inequality for $u_j^2(t)$, $1 \leq j \leq p$, in terms of V_t :

$$\begin{aligned} u_j^2(t) &= (-\hat{\theta}_j^T(t) \phi_j^m(Y(t)) - a_j^T x_j + y_{\text{ref}_j}^{(n)})^2 \\ &\leq (-a_j^T x_j - \theta_j^T \phi_j^m(Y(t)) + (\theta_j - \hat{\theta}_j)^T \phi_j^m(Y(t)) + y_{\text{ref}_j}^{(n)})^2 \\ &\leq (-a_j^T x_j - f_j(Y(t)) + d_j(Y(t)) + (\theta_j - \hat{\theta}_j)^T \phi_j^m(Y(t)) + y_{\text{ref}_j}^{(n)})^2 \\ &\leq \left(\|a_j\| \sqrt{\frac{x_j^T P_j x_j}{\underline{\lambda}(P_j)}} + \delta_{\infty j} w_{\infty j}^{-1} (\|Y(t)\|) + s_M \right. \\ &\quad \left. + \sqrt{(\theta_j - \hat{\theta}_j)^T (\Gamma_j^m)^{-1} (\theta_j - \hat{\theta}_j)} \|(\Gamma_j^m)^{1/2} \phi_j^m(Y(t))\| + \gamma_2 \right)^2 \\ &\leq \left(\|a_j\| \sqrt{\frac{x_j^T P_j x_j}{\underline{\lambda}(P_j)}} + \delta_{\infty j} w_{\infty j}^{-1} \left(\sqrt{\frac{Y^T P_j Y}{\underline{\lambda}(P_j)}} \right) + s_M \right. \\ &\quad \left. + \sqrt{(\theta_j - \hat{\theta}_j)^T (\Gamma_j^m)^{-1} (\theta_j - \hat{\theta}_j)} + \gamma_2 \right)^2 \\ &\leq \tilde{u}_j^2(V(t)), \end{aligned} \quad (102)$$

where the function \tilde{u}_j is defined by

$$\tilde{u}_j(v) = \frac{\|a_j\|}{\sqrt{\underline{\lambda}(P_j)}} \sqrt{v} + \delta_{\infty j} w_{\infty j}^{-1} \left(\frac{\sqrt{v} + \bar{\lambda}(P_j) \gamma_1}{\sqrt{\underline{\lambda}(P_j)}} \right) + s_M + \sqrt{2\alpha_j v} + \gamma_2, \quad (103)$$

and where we have used the crucial inequality

$$\|(\Gamma_j^m)^{1/2} \phi_j^m(x)\| = \sum_{k=1}^m \omega_k \|\tilde{\phi}^k(x)\| \leq \sum_{k=1}^m |c_k| \leq 1, \quad (104)$$

which follows from the construction of Γ^m . The bound on the control effort term is now obtained by the substitution $v = V_t$, which is valid, since by inequality (43) and the dead-zone definition, V is decreasing on T_{Ω_1} , hence

$$\begin{aligned} \int_{T_{\Omega_1}} u^T R u \, dt &\leq \bar{\lambda}(R) \sum_{j=1}^p \int_{T_{\Omega_1}} \tilde{u}_j^2(V_t) \, dt \\ &= \bar{\lambda}(R) \sum_{j=1}^p \int_{v \in v(T_{\Omega_1})} \frac{\tilde{u}_j^2(v)}{\dot{v}} \, dv \\ &\leq \frac{\bar{\lambda}(R)}{\eta_1^2(\underline{\lambda}(Q) - g)} \sum_{j=1}^p \int_{\eta_1^2}^{V_0} \tilde{u}_j^2(v) \, dv \\ &\leq \frac{\bar{\lambda}(R)}{\eta_1^2(\underline{\lambda}(Q) - g)} \sum_{j=1}^p \int_{\eta_1^2}^{W^*} \tilde{u}_j^2(v) \, dv, \end{aligned} \quad (105)$$

where we have used the fundamental inequality (43). The result now follows by observing that the case $x_0 \in \Omega_0$ follows by shifting time so that at time $t = 0$, $x_0 \in \partial\Omega_0$. \blacksquare

To apply the above result, it is necessary to have a priori knowledge of the uncertainty level δ_2 in order to satisfy condition (87). We now relax the need for this knowledge, by considering global results, i.e. by taking $\Omega = \mathbb{R}^n$. We measure the uncertainties in global weighted L^2 , L^∞ spaces.

As discussed in [FSR2], it is clear that in general we will require a countably infinite-dimensional model to satisfy the approximation constraints. However, in general infinite-dimensional models will not yield physically realisable controllers, as we cannot update even a countable number of parameters at a time step. Thus we restrict ourselves to a particular class of locally finite-dimensional models, the so-called semi-global finite-dimensional (SFD) models. These are models whose basis functions have compact support, and moreover for any compact set Ω , there are only a finite number of basis functions whose supports intersect Ω [FSR2].⁷ Examples of such models include splines and compactly supported wavelets. Such models yield controllers whose (finite) dimension varies according to the uncertainty level, in the sense that only a finite number of adaptive estimates are non-constant: hence the controller can be implemented.

Let $\{\tilde{\Phi}^m\}_{m \geq 1}$ be a sequence of a (countably infinite) SFD model. Defining the

⁷ The proof of *global* stability given in [FSR2] is however only valid if the model basis is orthonormal (orthonormal SFD models exist, e.g. Daubechies wavelets [D]). Note however that the construction given next also establishes the required *global* result for general SFD models.

basis function extension Φ^m as in (81), it is clear that Φ^m is also an SFD model. We assume that $\{\Phi^m\}_{m \geq 1}$ is a $\{K_\beta\}_{\beta \geq 0}$ approximate model class. A suitable construction of a suitable Γ^m for the SFD model Φ^m is as follows.

As in the previous section, for simplicity, we assume the model components are equal, i.e. $\phi_j^m = \phi_i^m = \phi^m$ for $1 \leq i, j \leq p$. Then we partition the infinite-dimensional models $\tilde{\phi}^m$ into a countable number of finite-dimensional parts:

$$\tilde{\phi}^m = [(\bar{\phi}_1^m)^T | (\bar{\phi}_2^m)^T | \dots]^T, \quad \dim \bar{\phi}_l^m < \infty, \quad l \geq 1, \quad (106)$$

where the basis functions are ordered such that if $\bar{\Omega}_l^m = \bigcup_{1 \leq i \leq l} \text{supp } \bar{\phi}_i^m$, then

$$\bar{\Omega}_l \subset \bar{\Omega}_{l+1}, \quad \bar{\Omega}_l \cap (\bar{\Omega}_k \setminus \bar{\Omega}_{l+1}) = \emptyset, \quad \forall k \geq l+2, \quad \forall l \geq 1. \quad (107)$$

Such a partition into finite-dimensional parts can always be achieved by a recursive construction, for if

$$[(\bar{\phi}_1^m)^T | (\bar{\phi}_2^m)^T | \dots | (\bar{\phi}_k^m)^T]^T \quad (108)$$

satisfies (107), then there are at most a finite number of remaining basis functions which: (a) are contained in $\tilde{\phi}^m$, (b) are not contained in $[(\bar{\phi}_1^m)^T | (\bar{\phi}_2^m)^T | \dots | (\bar{\phi}_k^m)^T]^T$ and (c) have supports which intersect $\bar{\Omega}_k^m$. Choose a finite-dimensional $\bar{\phi}_k^m$ to have at least one of the basis functions satisfying both (a) and (b) and to contain all the basis functions satisfying (c) (a finite number, since the model is SFD). Then

$$[(\bar{\phi}_1^m)^T | (\bar{\phi}_2^m)^T | \dots | (\bar{\phi}_{k+1}^m)^T]^T \quad (109)$$

also satisfies (107), and such a recursive construction defines a suitable partition of the whole of ϕ^m .

We let the corresponding parameter vectors be denoted by $\bar{\theta}_j^m, \bar{\theta}_l^m$, etc., and define Gram matrices \bar{G}_{jl}^m corresponding to the basis functions $[(\bar{\phi}_1^m)^T | \dots | (\bar{\phi}_l^m)^T]^T$ and with respect to the space $L^2(\bar{\Omega}_l^m; w_{2j})$ for $m \geq 1, l \geq 1, 1 \leq j \leq p$.

Let $H_j^m, 1 \leq j \leq p$, be the operators defined by

$$H_j^m = \text{diag} \left\{ \frac{1}{c'_1 \lambda(\bar{G}_{j2}^m)} I_{\dim \bar{\phi}_1^m \times \dim \bar{\phi}_1^m}, \frac{1}{c'_2 \lambda(\bar{G}_{j3}^m)} I_{\dim \bar{\phi}_2^m \times \dim \bar{\phi}_2^m}, \dots \right\}, \quad (110)$$

where (c'_1, c'_2, \dots) is a sequence of positive numbers satisfying $\sum_{i=1}^{\infty} c'_i = 1$.

Now, let

$$\Gamma_j^m = \text{diag} \{ \omega_{j1}^2 H_j^1, \omega_{j2}^2 H_j^2, \dots, \omega_{jm}^2 H_j^m \}, \quad \Gamma^m = (\Gamma_1^m, \dots, \Gamma_p^m)^T, \quad (111)$$

where ω_i is chosen as follows. Let (c_1, c_2, \dots) be a positive sequence satisfying $\sum_{m=1}^{\infty} c_m = 1$, and let

$$r_{jm} = \sup_x \| (H_j^m)^{1/2} \tilde{\phi}^m(x) \|. \quad (112)$$

Note that by the K_β -approximation property of the model, we must have $r_{ji} > 0$, and since by the compact support and continuity of the basis functions we have $\sup_x \|\tilde{\phi}^i(x)\| < +\infty$, hence we also have $r_{ji} < +\infty$. Now let

$$\omega_{jm} = \frac{c_m}{r_{jm}}. \quad (113)$$

The reason for defining H^m in the definition of Γ_j^m above instead of taking $H^m = I$ (as was the case in Theorem 4.1) is that we need an appropriate scaling to deal with the problem of the countable number of terms in the expressions:

$$(\tilde{\theta}_j^m)^T (\Gamma_j^m)^{-1} \tilde{\theta}_j^m, \quad 1 \leq j \leq p. \quad (114)$$

Theorem 4.2. *Let $\{K_\beta\}_{\beta \geq 0}$ be a smoothness class, and suppose $\{\Phi^m\}_{m \geq 1}$ is defined by (81), where $\{\tilde{\Phi}^m\}_{m \geq 1}$ is a $\{K_\beta\}_{\beta \geq 0}$ SFD approximate model class. Consider the system $\Sigma_\Delta(\mathcal{Y}_0)$ with functional uncertainty:*

$$\Delta_j \subset K_\beta \cap \Delta(L^2(\mathbb{R}^n; w_{2j}), \delta_{2j}) \cap \Delta(L^\infty(\mathbb{R}^n; w_{\infty j}), \delta_{\infty j}), \quad 1 \leq j \leq p, \quad (115)$$

where ω_{2j} and $\omega_{\infty j}$ are known, but the uncertainty levels $\delta_{2j}, \delta_{\infty j}$ are unknown. Let

$$\Xi_m(\cdot) = \Xi(\Gamma^m, \alpha, \Phi^m, \Omega_0^m)(\cdot), \quad (116)$$

where Γ^m and Φ^m are as in (81) and (111). If $\Omega_0^m \subset \Omega_1 \subset \Omega$, $\forall m \geq 1$, then there exists $M \geq 1$ such that for all $m \geq M$:

1. $(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))$ is well-posed and $x(t) \rightarrow \Omega_0$ as $t \rightarrow \infty$,
2. only a finite number of parameters are adapted,
- 3.

$$\limsup_{m \rightarrow \infty} \mathcal{P}_{(\Sigma_\Delta(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(\mathcal{Q}, R, \Omega_0^m, \Omega_1) < \infty. \quad (117)$$

Proof. The theorem follows analogously to that of Theorem 4.1, and the proof technique of the global result in [FSR2], given the inequalities (119) and (120) below. Inequality (119) is required to obtain a uniform bound on $(\theta_m)^T (\Gamma_j^m)^{-1} \theta_m$ (see inequality (95)) which leads to the uniform bound on V_0^m (by W_α^*), whilst inequality (120) is required to obtain a uniform bound on $\|(\Gamma_j^m)^{1/2} \phi_j^m(x)\|$ (which forms part of the control effort bound).

First we let q_M be defined by (91) with $\Omega = \mathbb{R}^n$. Choose M large, and then let $\theta^m = [0 | (\tilde{\theta}^M)^T | 0 \cdots]^T$ be defined so that $(\theta^m)^T \phi^m = (\tilde{\theta}^M)^T \tilde{\phi}^M$. Then

$$(\theta_j^m)^T (\Gamma_j^m)^{-1} \theta_j^m = (\omega_{jM}^2)^{-1} (\tilde{\theta}_j^M)^T (H_j^M)^{-1} \tilde{\theta}_j^M \quad (118)$$

and

$$\begin{aligned} (\tilde{\theta}_j^M)^T (H_j^M)^{-1} \tilde{\theta}_j^M &\leq \sum_{k=1}^{\infty} c'_k \underline{\lambda}(\bar{G}_{j(k+1)}^M) (\bar{\theta}_{jk}^M)^T \bar{\theta}_{jk}^M \\ &\leq \sum_{k=1}^{\infty} c'_k \underline{\lambda}(\bar{G}_{j(k+1)}^M) \left(\sum_{l=1}^{k+1} (\bar{\theta}_{jl}^M)^T \bar{\theta}_{jl}^M \right) \\ &\leq \sum_{k=1}^{\infty} c'_k [(\bar{\theta}_{j1}^M)^T | \cdots | (\bar{\theta}_{j(k+1)}^M)^T] \bar{G}_{j(k+1)}^M [(\bar{\theta}_{j1}^M)^T | \cdots | (\bar{\theta}_{j(k+1)}^M)^T]^T \\ &= \sum_{k=1}^{\infty} c'_k \| [(\bar{\theta}_{j1}^M)^T | \cdots | (\bar{\theta}_{j(k+1)}^M)^T] [(\bar{\phi}_1^M)^T | \cdots | (\bar{\phi}_{k+1}^M)^T]^T \|_{L^2(\bar{\Omega}_k^M; w_{2j})}^2 \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{k=1}^{\infty} c'_k \|(\tilde{\theta}_j^M)^T \tilde{\phi}^M\|_{L^2(\bar{\Omega}_k^M; w_{2j})}^2 \\
&\leq \sum_{k=1}^{\infty} c'_k \|(\theta_j)^T \phi^M\|_{L^2(\bar{\Omega}_k^M; w_{2j})}^2 \\
&\leq (\delta_{2j} + q_M)^2.
\end{aligned} \tag{119}$$

Now we derive the bound analogous to (104) that was required in bounding the control effort term of the performance measure. By the definition of Γ_j^m , it follows that

$$\|(\Gamma_j^m)^{1/2} \phi_j^m(x)\| \leq \sum_{k=1}^m \omega_{jk} \|(H_j^k)^{1/2} \tilde{\phi}^k(x)\| \leq \sum_{k=1}^m c_k \leq 1, \tag{120}$$

hence completing the proof. \blacksquare

5. Conclusions

We consider the closed-loop performance scaling question to be the fundamental issue in the design and implementation of controllers based on function approximators. We have proved in Section 3 that standard approximate adaptive control designs lead to divergent closed-loop performance when the resolution of the approximator is increased in Scenario 1, i.e. the case where $\Omega_0^m \rightarrow \{0\}$. It remains open to establish whether bad scaling can be proved to occur for a fixed Ω_0 , and whether poor scaling can occur in the SISO case.

For a simple class of systems, namely a tracking problem for a chain of integrators, we have exhibited both global and semi-global control designs (Section 4) which do not lead to divergent performance in Scenarios 1–3, i.e. the performance is uniformly bounded in cases where $\Omega_0^m \rightarrow \{0\}$ and where $\Omega_0^m = \Omega_0$ is fixed. Furthermore, the proofs of the results give an explicit uniform performance bound. These positive results can be generalised to wider classes of systems [KKK] with backstepping controllers [FS].

The dynamics of the two designs are interesting. In the former case the divergent performance can be roughly understood as follows. Because the model is mono-resolution, a high adaption gain must be used to ensure a good transient tracking performance. On the other hand, a uniformly high adaption gain leads to excessive control effort via increased overshoots. In the latter designs the effective adaption rates can be chosen to be lower for the basis functions corresponding to higher resolutions. In this manner we can ensure good transient tracking, whilst not leading to high adaption gains which force excessive control effort.

The results in this paper give a strong mathematical incentive for studying function approximator designs as opposed to their parametric counterparts. In a problem specified by a functional uncertainty, the nonlinearities of the system are not highly structured: they are specified solely by a norm bound. In contrast, a problem specified by a parametric uncertainty has a highly structured nonlinearity which is specified by both the parameters *and the basis functions which they multiply*. On the other hand, in the approximate adaptive case, the regularity of the basis func-

tions themselves can be exploited. Hence, in the approximate adaptive scenario we have a much better handle on the nonlinearity and, e.g. the “worst case” system, and can give meaningful performance results for whole classes of systems. In the parametric case we can only expect to give good performance results for classes of systems which impose particular sets of assumptions on the allowable nonlinearities: more probably by proceeding on a case by case basis. This simply reflects the fact that less structured uncertainties are simpler to handle than more structured uncertainties. Furthermore, the approximate adaptive theory has a rich asymptotic theory (as $m \rightarrow \infty$), which by definition has no counterpart in the parametric case.

Acknowledgements. The authors thank the reviewers for their careful reading of the manuscript and for their many constructive comments.

Appendix. Proofs of the Results of Section 3

Proposition 3.1. Consider the closed-loop system $(\Sigma_f(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))$ defined by Sections 3.2 and 3.3. Then

$$\mathcal{P}_{(\Sigma_f(\mathcal{Y}_0), \Xi_m(\mathcal{Y}_{\text{ref}}))}(0, R, \Omega_0^m, \Omega_1) \geq O\left(\sqrt{\frac{\alpha_j^m}{m}}\right), \quad 1 \leq j \leq p. \quad (121)$$

Proof. Let $f = (f_1, f_2, \dots, f_p)^T = (0, 0, \dots, 0)^T \in \Delta$, $n_j = 1$, $1 \leq j \leq p$ (i.e. $\dot{x}_j = u_j$, $1 \leq j \leq p$), and consider the zero reference signal $y_{\text{ref}} = 0 \in \mathcal{Y}_{\text{ref}}$. Fix m and $1 \leq j \leq p$ and consider the initial condition $x(0) = x_0$ defined by $x_{0i} = 0$, $i \neq j$, $x_{0j} = \eta/2$. Applying the controller Ξ_m , it is straightforward to observe that $x_i = 0$, $\hat{\theta}_i = 0$, $u_i = 0$ for all $t \geq 0$, and for $i \neq j$. Then

$$\begin{aligned} \hat{\theta}_j^T \phi_j^m(x) &= \hat{\theta}_j^T \phi_j^m(0, \dots, 0, x_j, 0, \dots, 0) \\ &= \sum_k \hat{\theta}_{jk} \varphi_{\xi(k)_j}^m(x_j) \prod_{l \neq j} \varphi_{\xi(k)_l}^m(0) \\ &= \sum_{-m \leq i \leq m} \varphi_i^m(x_j) \sum_{k: \xi(k)_j=i} \hat{\theta}_{jk} \prod_{l \neq j} \varphi_{\xi(k)_l}^m(0) \\ &= \tilde{\theta}_j^T s_j^m(x_j), \end{aligned}$$

where we define

$$\tilde{\theta}_{ji} = \sum_{k: \xi(k)_j=i} \hat{\theta}_{jk} \prod_{l \neq j} \varphi_{\xi(k)_l}^m(0). \quad (122)$$

Rewriting the closed-loop system in terms of $\tilde{\theta}_j$, we only need consider the $(1 + 2m + 1)$ -dimensional system:

$$\begin{aligned} \dot{x}_j &= -\tilde{\theta}_j^T s^m(x_j) - a_j x_j, & x_j(0) &= \eta/2, \\ \dot{\tilde{\theta}}_j &= \alpha_j^m x_j b_j D([- \eta_0(m), \eta_0(m)], x_j) s^m(x_j), & \tilde{\theta}_j(0) &= 0, \\ u_j &= -\tilde{\theta}_j^T s^m(x_j) - a_j x_j. \end{aligned} \quad (123)$$

Since $R = \text{diag}(r_1, \dots, r_p)$, we clearly have

$$\int_{\{t \geq 0 \mid |x_j(t)| > \eta_1\}} r_j u_j^2 dt \leq \mathcal{P}_{(\Sigma_\Delta(\mathcal{Q}_0), \Xi_m(\mathcal{Q}_{\text{ref}}))}(\mathbf{0}, R, \Omega_0^m, \Omega_1), \quad (124)$$

where the left-hand side is computed for the system (123).

Let $t = \inf\{s \geq 0 \mid \text{dist}\{x(s), \Omega_1\} = 0\}$. As $t < \infty$ by Theorem 2.1, we can establish the following inequality for $u_j = \dot{x}_j$ ($x_{0j} \geq \eta_1$):

$$\|x_j\|_{L^2[0,t]} \|u_j\|_{L^2[0,t]} \geq \left| \int_0^t x_j u_j dt \right| = \left| \int_0^t x_j \dot{x}_j dt \right| = x_j^2(0) - \eta_1^2, \quad (125)$$

by an application of the Cauchy-Schwartz inequality. Hence

$$\int_{T_{\Omega_1}} u_j^2 dt \geq \left(\frac{x_0^2 - \eta_1^2}{\|x_j\|_{L^2[0,t]}} \right)^2. \quad (126)$$

Now we consider upper bounds on $\|x_j\|_{L^2[0,t]}$.

By the first part of condition 4, Section 3.1, given any $m \geq 1$ we know that $\bar{\theta}_m = (\eta \sqrt{\alpha_j^m} / \Upsilon_1 \sqrt{m}) \mathbf{1}$ satisfies

$$\bar{\theta}_m^T s^m(x_j) \geq \sqrt{\frac{\alpha_j^m}{m}} x_j, \quad \forall x_j \in [0, \eta_1]. \quad (127)$$

Now consider the Lyapunov function:

$$W_m = \frac{Q_j x_j^2}{2a_j} + \frac{(\bar{\theta}_m - \tilde{\theta})^T (\bar{\theta}_m - \tilde{\theta})}{2\alpha_j^m}. \quad (128)$$

Using the differential inequality

$$\dot{x}_j \leq (\bar{\theta}_m - \tilde{\theta})^T s^m(x_j) - \left(a_j + \sqrt{\frac{\alpha_j^m}{m}} \right) x_j \quad (129)$$

we can establish on the interval $[0, t)$ that

$$\begin{aligned} \dot{W}_m &\leq \frac{Q_j x_j}{a_j} (\bar{\theta}_m - \tilde{\theta})^T s^m(x_j) - \frac{a_j + \sqrt{\alpha_j^m/m}}{a_j} Q_j x_j^2 - (\bar{\theta}_m - \tilde{\theta})^T x_j b_j s^m(x_j) \\ &\leq - \frac{a_j + \sqrt{\alpha_j^m/m}}{a_j} Q_j x_j^2, \end{aligned} \quad (130)$$

since $b_j = Q_j/a_j$ (because of the Lyapunov equation $A_{*j}^T P_j + P_j A_{*j} = -Q_j$, $P_j \in \mathbb{R}$, $A_{*j} = -a_j$, $b_j = 2p_j$ and $b_j a_j = Q_j$). Hence

$$\begin{aligned} \int_0^t x_j^2 dt &\leq \frac{a_j}{Q_j(a_j + \sqrt{\alpha_j^m/m})} \int_0^t -\dot{W}_m dt \\ &= \frac{a_j}{Q_j(a_j + \sqrt{\alpha_j^m/m})} (W_m(0) - W_m(t)) \\ &\leq \frac{a_j}{Q_j(a_j + \sqrt{\alpha_j^m/m})} \left(\frac{Q_j(x_j^2(0) - \eta_1^2)}{2a_j} + \frac{\bar{\theta}_m^T \bar{\theta}_m}{2\alpha_j^m} \right), \end{aligned} \quad (131)$$

and so we can establish

$$\begin{aligned}
\int_{T_{\Omega_1}} u_j^2 dt &\geq \frac{(x_j^2(0) - \eta_1^2)^2}{\|x\|_{L^2[0,t]}^2} \\
&\geq \frac{Q_j(x_j^2(0) - \eta_1^2)^2}{a_j} \left(a_j + \sqrt{\frac{\alpha_j^m}{m}} \right) \left(\frac{Q_j(x_j^2(0) - \eta_1^2)}{2a_j} + \frac{\bar{\theta}_m^T \bar{\theta}_m}{2\alpha_j^m} \right)^{-1} \\
&= O\left(\sqrt{\frac{\alpha_j^m}{m}} \right). \quad \blacksquare
\end{aligned} \tag{132}$$

Proposition 3.4. *Let $m \geq 1$, $\alpha^m > 0$, $\kappa > 0$, $\rho_m > 0$ be arbitrary. Let $(Y^m, \hat{\theta}^m)$ ($1 \leq j \leq p$) denote the solution of the closed-loop system $(\Sigma_{f^m}, \Xi_m(Y_m^{\text{ref}}))$, where $f^m = (f_1^m, \dots, f_p^m)^T$, $f_j^m = (\theta_j^m)^T \phi_j^m$, $\theta_j^m = \kappa \mathbf{1}$. Assume further that the reference signal Y_m^{ref} is such that $\phi_j^m(Y_m^{\text{ref}})$ is persistently exciting. Then there exists $\bar{\varepsilon}_m > 0$ such that if $\eta_0^m \leq \bar{\varepsilon}_m$, then*

$$\limsup_{t \rightarrow \infty} \|\hat{\theta}_j^m(t) - \theta_j^m\| \leq \rho_m, \quad 1 \leq j \leq p, \quad \forall m \geq \beta. \tag{133}$$

Proof. Let $m \geq \beta$. Write the closed-loop system (Σ_{f^m}, Ξ_m) in the form

$$\begin{aligned}
\dot{x}_j &= A_{*j} x_j + B_j((\theta_j - \hat{\theta}_j)^T \phi_j(Y^{\text{ref}})) + d_{xj}(t), \quad 1 \leq j \leq p, \quad x(0) = x_0, \\
\dot{\hat{\theta}}_j &= \alpha_j x_j^T b_j \phi_j(Y^{\text{ref}}) + d_{\hat{\theta}_j}(t), \quad \hat{\theta}_j(0) = 0, \quad 1 \leq j \leq p,
\end{aligned} \tag{134}$$

where

$$d_{xj}(t) = B_j(\theta_j - \hat{\theta}_j(t))^T (\phi_j(Y(t)) - \phi_j(Y^{\text{ref}}(t))), \tag{135}$$

$$d_{\hat{\theta}_j} = \alpha x_j^T b_j (D(\Omega_0, x) \phi_j(Y(t)) - \phi_j(Y^{\text{ref}}(t))). \tag{136}$$

We denote solutions of this system by $(x(t, \eta_0), \hat{\theta}(t, \eta_0))$. We know from Theorem 2.1 that

$$\sup_{0 < \eta_0 \leq \varepsilon} \limsup_{t \rightarrow \infty} \|x(t, \eta_0)\| \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0, \tag{137}$$

and since $V(x(t, \eta_0), \hat{\theta}(t, \eta_0)) \leq W_\alpha$ we also have

$$\sup_{0 < \eta_0 \leq \varepsilon} \sup_{t \rightarrow \infty} \|x(t, \eta_0)\| < \infty \tag{138}$$

and

$$\sup_{0 < \eta_0 \leq \varepsilon} \sup_{t \rightarrow \infty} \|\theta - \hat{\theta}(t, \eta_0)\| < \infty. \tag{139}$$

Hence it follows that

$$\sup_{0 < \eta_0 \leq \varepsilon} \limsup_{t \rightarrow \infty} \|(d_{x_j}^m, d_{\hat{\theta}_j}^m)\| \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0. \tag{140}$$

The right-hand side of (134) with $(d_{x_j}^m, d_{\hat{\theta}_j}^m) = 0$ is locally Lipschitz, hence uniformly Lipschitz along the trajectories by (138), and also exponentially stable by

(70). Hence by a simple modification of, e.g. Theorem 5.3.1 of [SB], we have $\exists K_m > 0$ such that

$$\limsup_{t \rightarrow \infty} \|(x(t, \eta_0), \theta - \hat{\theta}(t, \eta_0))\| \leq K_m \limsup_{t \rightarrow \infty} \|d(t, \eta_0)\|. \quad (141)$$

Then it follows that $\exists \tilde{\epsilon}_m > 0$ such that

$$\limsup_{t \rightarrow \infty} \|(x(t, \eta_0), \theta - \hat{\theta}(t, \eta_0))\| \leq \rho \quad (142)$$

from which the result follows. \blacksquare

Proposition 3.5. *Let $m \geq \beta$. Then there exists a reference trajectory $Y_m^{\text{ref}} \in \mathcal{Y}_{\text{ref}}$ and $\tilde{\epsilon}_m > 0$ such that if $Y: \mathbb{R}_+ \rightarrow \mathbb{R}^n$ is continuous and*

$$\limsup_{t \rightarrow \infty} \|Y(t) - Y_m^{\text{ref}}(t)\| \leq \tilde{\epsilon}_m,$$

then $\phi_j^m(Y(\cdot))$ is persistently exciting for $1 \leq j \leq p$.

Proof. Fix $j, t > 0$. The upper bound in the persistently excitation condition is easily established for any continuous signal Y and $\tau > 0$:

$$\int_t^{t+\tau} (c^T \phi_j^m(Y(t)))^2 dt \leq \tau \Upsilon_2 \|c\|^2. \quad (143)$$

To establish the lower bound in the persistently excitation condition we argue as follows.

By condition 3, Section 3.1, there are $m' = (2m + 1)^p$ linearly independent basis functions of each model component ϕ_j^m over the domain Ω_2 . It easily follows that there exist m' disjoint points $Z_1, \dots, Z_{m'} \in \Omega_2$ such that the matrix $H_m = (\phi_{jk}^m(Z_i))_{1 \leq k, i \leq m'}$ is invertible. For $x_i \in \mathbb{R}^n$, $1 \leq i \leq m'$, define

$$E_m(x_1, \dots, x_{m'}) = (\phi_{jk}^m(Z_i + x_i))_{1 \leq i, k \leq m'} \quad (144)$$

and observe that $E_m(0, 0, \dots, 0) = H_m$ and thus $\underline{\lambda}(E_m(0, \dots, 0))^T E_m(0, \dots, 0) > 0$.

We define

$$v(\phi_j^m, \varepsilon) = \inf_{\|x_1\|, \dots, \|x_{m'}\| \leq \varepsilon} \underline{\lambda}(E_m(x_1, \dots, x_{m'}))^T E_m(x_1, \dots, x_{m'}). \quad (145)$$

Then there exists $\tilde{\varepsilon} > 0$ such that if $0 \leq \varepsilon < \tilde{\varepsilon}/2$, then $0 < v(\phi_j^m, \varepsilon) < \infty$. This is a simple consequence of the continuity of the basis functions, the continuous dependence of the eigenvalues of a matrix with respect to its entries and the previous observation that $\underline{\lambda}(E_m(0))^T E_m(0) > 0$.

Let $\tau > 0$ be large enough such that there exist a sufficiently smooth periodic signal $y_{\text{ref}} \in \mathcal{Y}_{\text{ref}}$ (i.e. such that $\|Y^{\text{ref}}\| \leq \gamma_1$ and $\|y_{\text{ref}}^{(n_j)}\| \leq \gamma_2$) with period τ for which the sets

$$\bar{I}_i(t) = \{s \in [t, t + \tau): \|Z_i - Y^{\text{ref}}(s)\| < \tilde{\varepsilon}/2\} \quad (146)$$

have positive measure (i.e. $m(\bar{I}_i(t)) > 0$) for $1 \leq i \leq m'$.

Choose any continuous signal Y such that $\limsup_{s \rightarrow \infty} \|Y(s) - Y^{\text{ref}}(s)\| \leq \varepsilon$. We show that for t large enough

$$\int_t^{t+\tau} |c^T \phi_j^m(Y(t))|^2 dt \geq w_1 \|c\|^2 \quad (147)$$

for some $w_1 > 0$. Since $\limsup_{s \rightarrow \infty} \|Y(s) - Y^{\text{ref}}(s)\| \leq \varepsilon$, there exists $T > 0$ such that for $s > T$ $\|Y(s) - Y^{\text{ref}}(s)\| \leq \tilde{\varepsilon}/2$. Let $t > T$. Therefore

$$0 < m(\bar{I}_i(t)) \leq m(\{s \in [t, t + \tau]: \|Z_i - Y(s)\| \leq \tilde{\varepsilon}\}), \quad 1 \leq i \leq m'. \quad (148)$$

Since $Z_1, \dots, Z_{m'}$ are disjoint there exists a system of pairwise disjoint (measurable) sets $I_i(t) \subseteq \{s \in [t, t + \tau]: \|Z_i - Y(s)\| \leq \tilde{\varepsilon}\}$ such that $m(I_i(t)) > 0$, $1 \leq i \leq m'$. Then

$$\int_t^{t+\tau} |c^T \phi_j^m(Y(t))|^2 dt \geq \sum_{i=1}^{m'} \int_{I_i(t)} |c^T \phi_j^m(Y(t))|^2 dt, \quad 1 \leq j \leq p. \quad (149)$$

Let $1 \leq i \leq m'$. Applying the intermediate value theorem, we have that there exist $x_i \in \mathbb{R}^n$, $\|x_i - Z_i\| \leq \tilde{\varepsilon}$ ($Y(t) = x_i$ for some $t \in I_i(t)$) such that

$$\int_{I_i(t)} |c^T \phi_j^m(Y(t))|^2 dt = |c^T \phi_j^m(x_i)|^2 m(I_i(t)) \geq \tau_0 |c^T \phi_j^m(x_i)|^2, \quad (150)$$

where $0 < \tau_0 = \inf_{t \geq T} \min_{1 \leq i \leq m'} m(I_i(t))$.

By the definition of v ,

$$\begin{aligned} & |c^T \phi_j^m(x_i)|^2 \\ &= \sum_{k,l=1}^{m'} c_k c_l \phi_{jk}^m(x_i) \phi_{jl}^m(x_i) \\ &= c^T E_m(x_1 - Z_1, \dots, x_{m'} - Z_{m'})^T E_m(x_1 - Z_1, \dots, x_{m'} - Z_{m'}) c \\ &\geq \underline{\lambda}(E_m(x_1 - Z_1, \dots, x_{m'} - Z_{m'})^T E_m(x_1 - Z_1, \dots, x_{m'} - Z_{m'})) \|c\|^2. \end{aligned} \quad (151)$$

Since $\|x_i - Z_i\| \leq \tilde{\varepsilon}$ it follows that $|c^T \phi_j^m(x_i)|^2 \geq \|c\|^2 v(\phi_j^m, \tilde{\varepsilon}_m)$. Therefore

$$\int_t^{t+\tau} |c^T \phi_j^m(Y(t))|^2 dt \geq \tau_0 m' v(\phi_j^m, \tilde{\varepsilon}_m) \|c\|^2. \quad (152)$$

This completes the proof. ■

References

- [CS] M. Cannon and J. J. E. Slotine. Space-frequency localized basis functions for nonlinear system estimation and control. *Neurocomputing*, 9(3):293–342, 1995.
- [CK] F.-C. Chen and H. K. Khalil. Adaptive control of nonlinear systems using neural networks. *Internat. J. Control*, 55(6):1299–1317, January 1992.
- [D] I. Daubechies. *Ten Lectures on Wavelets*. Regional Conference Series on Applied Mathematics. CBMS–NSF, Philadelphia, PA, 1992.

- [F] M. French. Adaptive Control of Functionally Uncertain Systems. Ph.D. thesis, University of Southampton, 1998.
- [FR] M. French and E. Rogers. Control and state performance measures in unmatched approximate adaptive nonlinear control. In *Proceedings IFAC Nonlinear Control Systems Design Symposium '98*, volume 3, pages 679–684, Enschede, July 1998.
- [FS] M. French and Cs. Szepesvári. Function approximator based control designs for strict feedback systems: LQ performance and scaling. Submitted for publication.
- [FSR1] M. French, Cs. Szepesvári, and E. Rogers. LQ performance bounds for adaptive output feedback controllers for functionally uncertain systems. *Automatica*, 38: 683–693, 2002.
- [FSR2] M. French, Cs. Szepesvári, and E. Rogers. Uncertainty, performance and model dependency in approximate adaptive nonlinear control. *IEEE Trans. Automat. Control*, 45(2):353–358, 2000.
- [JVL] S. Jagannathan, M. W. Vandegrift, and F. L. Lewis. Adaptive fuzzy logic control of discrete-time dynamical systems. *Automatica*, 36:229–241, 2000.
- [KKK] M. Krstić, I. Kanellakopoulos, and P. V. Kokotović. *Nonlinear and Adaptive Control Design*, 1st edition. Wiley, New York, 1995.
- [KNW] A. J. Kurdila, F. J. Narcowich, and J. D. Ward. Persistency of excitation in identification using radial basis functions. *SIAM J. Control Optim.*, 33(2):625–642, March 1995.
- [NA] K. S. Narendra and A. M. Annaswamy. *Stable Adaptive Control*. Information and System Sciences Series. Prentice-Hall, Englewood Cliffs, NJ, 1990.
- [OZSP] R. Ordonez, J. Zumberge, J. T. Spooner, and K. M. Passino. Adaptive fuzzy control: experiments and comparative analyses. *IEEE Trans. Fuzzy Systems*, 5(2):339–359, 1997.
- [PI] M. M. Polycarpou and P. A. Ioannou. On the existence and uniqueness of solutions in adaptive control systems. *IEEE Trans. Automat. Control*, 38(3):474–479, 1993.
- [R] T. J. Rivlin. *An Introduction to the Approximation of Functions*, 1st edition. Blaisdell, Walton, MA, 1969.
- [SS] R. M. Sanner and J. J. E. Slotine. Gaussian networks for direct adaptive control. *IEEE Trans. Neural Networks*, 3(6):837–863, 1992.
- [SB] S. Sastry and M. Bodson. *Adaptive Control—Stability, Convergence and Robustness*, 1st edition. Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [T] E. Tzirkel-Hancock. Stable Control of Nonlinear Systems Using Neural Networks. Ph.D. thesis, Department of Engineering, University of Cambridge, 1992.
- [YL] A. Yeslidirek and F. L. Lewis. Feedback linearization using neural networks. *Automatica*, 31(11):1659–1664, November 1995.