PARAMETER ESTIMATION IN NONLINEAR SYSTEMS WITH SATURATED OUTPUT

ADEL M. ANNABI, JEAN-BAPTISTE POMET, DARIO PRANDI, AND LUDOVIC SACCHELLI *

Abstract. In this work, we propose a method for parameter estimation in highly uncertain nonlinear systems in which the measured output is given by the product of an unknown matrix with a known saturation function evaluated at the (unknown) state. Estimating parameters here is difficult because the dynamics themselves are only partly known. Our approach exploits the structure of the saturation itself. Explicit piecewise-constant control inputs drive the system state into regions where the saturation output takes predetermined values, without requiring knowledge of the exact state or of the underlying dynamics. The resulting signal originates from an unknown function of the state, takes known values at specified times, and varies enough over the considered interval to identify the unknown parameters. We present a canonical form of dynamical systems for which this procedure can be applied. We further extend these results to show output feedback controllability under a slightly stronger condition.

Key words. parameter identification, saturated output, nonlinear control, excitation design, coradiant sets, output feedback

AMS subject classifications. 93B30, 93C10, 93B52, 93B05, 93C41

1. Introduction

Saturations and other activation functions occur in many controlled systems; in neuroscience, they are built into widely used models of brain activity, such as the Wilson–Cowan model [10, 11] and Amari neural fields [1].

All dynamical models depend on parameters whose values influence the behavior of the system and its response to controls. Some of them are known from physical considerations, but most have to be estimated from indirect measurements via *parameter identification* [7]. Identification means using measurements of various outputs of the system, corresponding to some inputs (or excitation), to deduce an estimate of the parameters, the choice of the excitation being an essential part of experiment design. When, moreover, the state is not fully observed and the estimation has to rely on partial measurements, the problem is called *state and parameter estimation* or *adaptive observation*.

These problems have been studied extensively since the foundational work in adaptive control and system identification [7, 3]. The standard framework assumes either contraction or passivity properties ensuring convergence of the estimation errors [7, 3], together with a persistent excitation condition ensuring that the trajectories carry enough information for parameter recovery [3]; persistent excitation of the state has become the standard sufficient condition for identifiability.

These assumptions do not hold generally. Many physical systems fail to satisfy global contraction or passivity, persistent excitation of unmeasured states is difficult to verify, and standard methods require the unknown parameters to multiply either a persistent signal, called the regressor, or a state-dependent function assumed persistently excited through the above-mentioned hypotheses. This is typically the situation of high-uncertainty models, such as those of neuroscience, in which only the structure is known.

*Adel M. Annabi, Jean-Baptiste Pomet, and Ludovic Sacchelli are with Université Côte d'Azur, Inria, CNRS, LJAD, France (email: adel-malik.annabi@inria.fr, jean-baptiste.pomet@inria.fr, ludovic.sacchelli@inria.fr). †‡

In this work, we focus on a class of highly uncertain non-linear systems where the dynamics are unknown and where the output contains an unknown matrix to be retrieved, applied to the saturation of the state. The unknown parameters multiply no known signal, and the classical assumptions for estimation fail [3]. We formalize this class of systems in a canonical form which represents the minimal structural assumptions under which parameter estimation with these techniques is feasible.

We build on this structure. Explicit controls steer the system's state into regions where the saturation takes prescribed values at predetermined times. The estimation procedure consists of three sequential steps. Linear output parameters are first recovered through regression on measurable data under a designed control input; the gain matrix is then identified by activating different saturation regions and exploiting the resulting input-output relationship; state recovery is finally achieved through output feedback control combined with algebraic inversion of the saturation function. Each step comes with explicit control designs and estimation guarantees.

2. Main results

2.1. A general identification result

In this section we consider a system whose measured output is a linear function of the saturated state,

$$\begin{cases} \dot{x} = f(x,u,t), \\ y = WS(x), \end{cases} \tag{2.1}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ the control, $y \in \mathbb{R}^q$ the measured output, $W \in \mathbb{R}^{q\times n}$ an unknown constant matrix and $S : \mathbb{R}^n \to \mathbb{R}^n$ a known continuous function. The vector field f is unknown; we only assume the regularity needed for the solutions to exist and to be absolutely continuous on the time interval under consideration. Nothing else about f is used in this section.

The problem addressed here is the identification of W from the measurement of y on a bounded time interval. The unknown parameters multiply the signal $S(x(\cdot))$, which is neither measured nor known and which is produced by a dynamics that is itself unknown. Classical regression is therefore not directly available, and neither is an observer, since no dissipativity or observability property of f is assumed.

Remark 2.1. All the results of this work can be extended to hold for

$$y = WS(x) + h_0 + G(y,u,t)^\top \theta, \tag{2.2}$$

where $h_0 \in \mathbb{R}^q$ and $\theta \in \mathbb{R}^{n_\theta}$ are unknown and $G : \mathbb{R}^q \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^{n_\theta \times q}$ is known. We chose to present the results in the simpler form (2.1) to avoid overloading the notation and to focus on the main ideas. The generalization is discussed in Remark 2.6.

The structural feature of S that we exploit is that it takes a fixed known value on large parts of the state space. This is what the following definition retains, and it is the only property of S used in this section.

Definition 2.2 (Saturation function). *A continuous function $S : \mathbb{R}^n \to \mathbb{R}^n$ is a* saturation function *if there exist a family of closed sets $\Sigma_i \subset \mathbb{R}^n$ with nonempty interior, i in some index set I, and vectors $v_i \in \mathbb{R}^n$, such that*

$$S(x) = v_i, \qquad \forall x \in \Sigma_i,\ \forall i \in I. \tag{2.3}$$

The set Σ_i is a saturation region *of S and v_i is its* saturation value.

Nothing is required of S outside the saturation regions, and no relation is imposed between the regions themselves. The value v_i is known even though the state producing it is not.

We now state the property on which the identification rests. It involves a trajectory, a finite collection of values and a finite collection of time intervals, and it does not involve the dynamics.

Definition 2.3 (Saturation-excitation condition). *An* excitation design *is a collection $\mathcal{E} = \{(v_i, J_i) \mid i = 1, \dots, N\}$ where the $J_i \subset \mathbb{R}$ are pairwise disjoint closed intervals of positive length and the vectors $v_i \in \mathbb{R}^n$ satisfy*

$$\operatorname{rank}[v_1, \dots, v_N] = n. \tag{2.4}$$

Let $T > 0$. An excitation design is said to be realizable in time T *if $J_i \subset [0, T]$ for all i and $\min J_1 > 0$.*

An absolutely continuous trajectory $x : [0, T] \to \mathbb{R}^n$ is a realization of $\mathcal{E}$ *if $J_i \subset [0, T]$ for every i and*

$$S(x(t)) = v_i \qquad \text{for all } t \in J_i \text{ and all } i \in [\![1, N]\!]. \tag{2.5}$$

We say that x is saturation-exciting *if it realizes an excitation design.*

Condition (2.5) constrains only the value that S takes along the trajectory. In practice, one has $x(t)$ enter the region Σ_i on which S equals v_i and stay there during J_i, the design being fixed beforehand, independently of the solution. The trajectory itself need not be known; what matters is which values are produced and on which intervals. Subsection 2.2 provides sufficient conditions, on the data of the problem and on the control, under which every solution issued from a given ball realizes a prescribed design, uniformly with respect to the unknown part of the system.

One necessary condition can be read on the output. Under (2.5) the output verifies $y \equiv W v_i$ on J_i, and is therefore constant on each J_i even though W is unknown. The converse fails, since a trajectory may be at rest without being saturated.

Given an excitation design $\mathcal{E}$, write $\tau_i := |J_i|$ for the length of the i-th interval and define the excitation Gramian and the estimator

$$\Gamma(\mathcal{E}) := \sum_{i=1}^{N} \tau_i \, v_i v_i^\top \in \mathbb{R}^{n \times n}, \qquad E(y; \mathcal{E}) := \left(\sum_{i=1}^{N} \int_{J_i} y(s) \, v_i^\top \, ds \right) \Gamma(\mathcal{E})^{-1} \in \mathbb{R}^{q \times n}. \tag{2.6}$$

Theorem 2.4. *Let $\mathcal{E}$ be an excitation design and let (x, u, y) be a solution of (2.1) on $[0, T]$ whose state realizes $\mathcal{E}$. Then $\Gamma(\mathcal{E})$ is invertible and*

$$E(y; \mathcal{E}) = W. \tag{2.7}$$

Proof. Let $V := [v_1, \dots, v_N] \in \mathbb{R}^{n \times N}$ and $D := \operatorname{diag}(\tau_1, \dots, \tau_N)$, so that $\Gamma(\mathcal{E}) = V D V^\top$. For $z \in \mathbb{R}^n$ we have $z^\top \Gamma(\mathcal{E}) z = \sum_i \tau_i (v_i^\top z)^2$, which vanishes only if z is orthogonal to every v_i, hence only if $z = 0$ by (2.4). Thus $\Gamma(\mathcal{E})$ is positive definite.

By (2.5) the output verifies $y(t) = W S(x(t)) = W v_i$ for every $t \in J_i$. Therefore

$$\sum_{i=1}^{N} \int_{J_i} y(s) v_i^\top \, ds = \sum_{i=1}^{N} \tau_i \, W v_i v_i^\top = W \, \Gamma(\mathcal{E}), \tag{2.8}$$

and multiplying on the right by $\Gamma(\mathcal{E})^{-1}$ gives the result. ∎

The estimator uses no knowledge of f and no state observer, and it requires no rank assumption on W. Identifiability comes from the saturation itself, not from the output matrix. A recursive least-squares implementation of the same regression [7] is immediate.

Remark 2.5 (Designs of insufficient rank, and output dimension). Two relaxations of the hypotheses of Theorem 2.4 are worth recording. First, suppose the values of the design span only a proper subspace $V := \operatorname{span}\{v_1, \dots, v_N\}$ of $\mathbb{R}^n$, so that (2.4) fails, because S takes fewer than n independent values, or because the values the dynamics can be made to produce do not span. The Gramian $\Gamma(\mathcal{E})$ is then singular with range V, but the data still determine part of W. Writing $\Gamma(\mathcal{E})^+$ for the Moore–Penrose pseudoinverse and P_V for the orthogonal projector onto V, the computation of the proof is unchanged up to its last line and gives

$$\left(\sum_{i=1}^{N} \int_{J_i} y(s)\, v_i^\top \, ds\right) \Gamma(\mathcal{E})^+ \;=\; W\,\Gamma(\mathcal{E})\,\Gamma(\mathcal{E})^+ \;=\; W P_V, \tag{2.9}$$

since $\Gamma(\mathcal{E})$ is symmetric with range V, so that the restriction of W to V is recovered exactly. Second, the equality of the source and target dimensions of S is not used in the proof. If $S : \mathbb{R}^n \to \mathbb{R}^d$ and $W \in \mathbb{R}^{q\times d}$, it suffices that the saturation values of the visited regions span $\mathbb{R}^d$, hence that $N \geq d$. This covers in particular the case where the saturation affects only part of the coordinates.

Remark 2.6 (Bias and linearly parametrized terms). Consider the more general output

$$y = W S(x) + h_0 + G(y, u, t)^\top \theta, \tag{2.10}$$

where $h_0 \in \mathbb{R}^q$ and $\theta \in \mathbb{R}^{n_\theta}$ are unknown and $G : \mathbb{R}^q \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^{n_\theta \times q}$ is known. Suppose that, besides the design $\mathcal{E}$, the trajectory visits during an interval J_0 preceding J_1 a region on which S vanishes, that is $S(x(t)) = 0$ for $t \in J_0$. On J_0 the output reduces to $y = \Phi\zeta$, with augmented regressor $\Phi := [\, I_q \;\; G(y, u, \cdot)^\top] \in \mathbb{R}^{q\times(q+n_\theta)}$ and $\zeta := (h_0, \theta) \in \mathbb{R}^{q+n_\theta}$, a regression which is linear in the unknowns and which the least-squares estimator

$$E_0(y) := \left(\int_{J_0} \Phi(s)^\top \Phi(s)\, ds\right)^{-1} \int_{J_0} \Phi(s)^\top y(s)\, ds \tag{2.11}$$

inverts exactly as soon as the Gramian $\int_{J_0} \Phi^\top \Phi \, ds \in \mathbb{R}^{(q+n_\theta)\times(q+n_\theta)}$ is invertible. That requirement, unlike the persistent excitation of the classical theory, bears on a regressor which is entirely available, G being known and y and u measured, and can therefore be tested on the recorded data. Applying then Theorem 2.4 to $\tilde{y} := y - h_0 - G(y, u, \cdot)^\top \theta$ on $J_1, \dots, J_N$ returns W. The two stages are sequential and acyclic—the second uses the output of the first, the first uses nothing of W—and their only additional requirement is one more region, the one carrying the value 0, which in the component-wise case means that the set where S vanishes has to be reachable.

Remark 2.7. The design $\mathcal{E}$ plays the role that the persistent excitation of the regressor plays in the classical theory, and the Gramian $\Gamma(\mathcal{E})$ that of the regressor Gramian. The difference is that here the regressor $S(x(\cdot))$ is neither measured nor computable, so no condition bearing on it could be checked; conditions weaker than persistence, such as excitation on a finite interval [6], initial excitation [8], concurrent learning [4] or dynamic regressor extension [2], bear on the regressor as well and are of no help for that reason. What Definition 2.3 requires instead is that the regressor take N prescribed values, and the saturation is what makes those values known in advance. The excitation is designed rather than assumed.

2.2. Structural conditions and input design achieving saturation excitation

Theorem 2.4 is of use only if (2.5) can be produced. We now specify the dynamics and give conditions under which an explicit control forces it. Let us consider the following realization of (2.1).

$$\begin{cases} \dot{x} = Ax + \phi(x,u,t) + Bu, \\ y = WS(x). \end{cases} \tag{2.12}$$

Here, we assume that A is Hurwitz, that ϕ is continuous, and that there exists $\bar{\phi} > 0$ such that $\|\phi\| \leq \bar{\phi}$. The matrix $B \in \mathbb{R}^{n \times m}$ is known.

Three information structures are covered, and they differ only in what is known of the linear part, namely A fully known, a Hurwitz part A_0 known with the remainder small, or A_0 a multiple of the identity.

The regions in which the state has to be brought are required to be unbounded in a controllable direction, which is what the following assumption expresses.

Assumption 2.8 (Coradiant saturation). *The function $S : \mathbb{R}^n \to \mathbb{R}^n$ is continuous. There exist n closed coradiant sets*[1] *$K_1, \dots, K_n \subset \mathbb{R}^n$ with non-empty interior, linearly independent vectors $v_1, \dots, v_n \in \mathbb{R}^n$, and $R > 0$ such that*

$$S(x) = v_i, \quad \forall x \in K_i,\ \|x\| \geq R, \quad i = 1, \dots, n. \tag{2.13}$$

Remark 2.9 (Classical saturations). The component-wise saturation is the model case. Let $S(x) = (S_0(x_1), \dots, S_0(x_n))^\top$ with $S_0 : \mathbb{R} \to \mathbb{R}$ nondecreasing, $S_0 = 0$ on $(-\infty, \underline{R}]$ and $S_0 = 1$ on $[\bar{R}, +\infty)$, put $R := \max\{|\underline{R}|, |\bar{R}|\}$ and, for a sign pattern $\mathbf{s} \in \{0,1\}^n$, let $\mathcal{O}_{\mathbf{s}}$ be the closed orthant carrying that pattern and

$$K_{\mathbf{s}} := \left\{ x \in \mathcal{O}_{\mathbf{s}} :\ |x_j| \geq R,\ j \in [\![1, n]\!] \right\}. \tag{2.14}$$

This set is closed, has nonempty interior, and is coradiant, since multiplying by $\lambda \geq 1$ preserves both the signs and the inequalities $|x_j| \geq R$; and $S = \mathbf{s}$ on the whole of $K_{\mathbf{s}}$, because $s_j = 1$ forces $x_j \geq R \geq \bar{R}$ and $s_j = 0$ forces $x_j \leq -R \leq \underline{R}$. Assumption 2.8 therefore holds with n such sets and with $v_i \in \{0,1\}^n$, provided the corresponding patterns are linearly independent.

For $b \in \operatorname{Im}(B)$ and a Hurwitz A, a constant control $Bu = \mu b$ drives the state towards the steady state $-\mu A^{-1} b$. It is therefore the image of $\operatorname{Im}(B)$ under $-A^{-1}$, and not the image of B itself, which has to meet the regions. We have the following.

Theorem 2.10. *Let Assumption 2.8 hold, let A be Hurwitz, and let $\bar{\phi} > 0$ be such that $\|\phi\| \leq \bar{\phi}$. Assume that*

$$\operatorname{Im}(A^{-1}B) \cap \mathring{K}_i \neq \varnothing \qquad \forall i \in \{1, \dots, n\} \tag{2.15}$$

Then there exists $T^* \geq 0$ such that for any $\rho > 0$ and any $T > T^*$ there exist an excitation design $\mathcal{E}$ realizable in time T, and a control $u : [0,T] \to \mathbb{R}^m$, such that any associated solution x of (2.12) with $x(0) \in \mathcal{B}(0, \rho)$ realizes $\mathcal{E}$.

The above theorem requires the knowledge of A. If this is not completely known we can still say something, as follows.

Theorem 2.11. *Let Assumption 2.8 hold, let $A = A_0 + A_1$ with A_0 Hurwitz, and let $\bar{\phi} > 0$ be such that $\|\phi\| \leq \bar{\phi}$. Assume that $\|A_1\| \leq \varepsilon_0$, where $\varepsilon_0 > 0$ depends only*

[1] Recall that a set $K \subset \mathbb{R}^n$ is coradiant if $\lambda x \in K$ for all $x \in K$ and $\lambda \geq 1$.

on A_0, B and $K_1, \ldots, K_n$ and is given by (3.50)*, and that*

$$\operatorname{Im}(A_0^{-1}B) \cap \mathring{K}_i \neq \varnothing \qquad \forall i \in \{1, \ldots, n\}$$

Then there exists $T^* > 0$ such that for any $\rho > 0$ and any $T > T^*$ there exist an excitation design $\mathcal{E}$ realizable in time T, and a control $u : [0,T] \to \mathbb{R}^m$, computed from A_0, B, $K_1, \ldots, K_n$, $\bar{\phi}$, ρ, T and the bound ε_0 alone, such that any associated solution x of (2.12) *with $x(0) \in \mathcal{B}(0, \rho)$ realizes $\mathcal{E}$.*

Under additional assumptions on the unperturbed part A_0, we can actually do everything in arbitrary time.

Theorem 2.12. *Under the assumptions of Theorem 2.11, assume that either*

i. $A_1 = 0$ and the pair (A_0, B) is controllable;

ii. $A_0 = -\lambda \operatorname{Id}$ for some $\lambda > 0$.

Then $T^* = 0$. For any $\rho > 0$ and any $T > 0$ there exist an excitation design $\mathcal{E}$ realizable in time T and a control $u : [0,T] \to \mathbb{R}^m$, computed from the same data, such that any associated solution x of (2.12) *with $x(0) \in \mathcal{B}(0, \rho)$ realizes $\mathcal{E}$.*

The time T^* of Theorems 2.10 and 2.11 is the total duration of the transients that bring the state into the successive regions; the dwell times spent inside them may be taken arbitrarily short, since they are part of the design and not of the dynamics.

2.3. Recovering the state via output feedback control

Once W has been identified, the output carries information about the state on the part of the space where S has not saturated. This suggests the control problem complementary to the previous one. Instead of steering the state into the saturation regions, keep it out of them. Doing so, and reading the state off the output afterwards, needs more of S than Assumption 2.8 provides, since that assumption says nothing outside the saturation regions. What is needed is that the output be, after a known change of variables, a component-wise saturation of the state.

Assumption 2.13 (Component-wise invertible output). *There exist a matrix $H \in \mathbb{R}^{n \times q}$, levels $\underline{R} < \bar{R}$ and a continuous nondecreasing $S_0 : \mathbb{R} \to [0,1]$, equal to* 0 *on $(-\infty, \underline{R}]$, equal to* 1 *on $[\bar{R}, +\infty)$ and strictly increasing on $[\underline{R}, \bar{R}]$, such that*

$$H\,W\,S(x) \;=\; S_1(x) := \big(S_0(x_1), \ldots, S_0(x_n)\big)^\top, \qquad \forall x \in \mathbb{R}^n. \tag{2.16}$$

The signal $Hy = S_1(x)$ is then computable from the output, and it is a component-wise saturation of the state, on which the scalar feedback used below can act coordinate by coordinate. We write

$$\Omega := [\underline{R}, \bar{R}]^n \tag{2.17}$$

for the region on which S_1 has not saturated; there S_0 is strictly increasing, hence invertible, and that is the only invertibility the recovery uses.

Remark 2.14. The model case is S itself component-wise, as in Remark 2.9, together with $\operatorname{rank} W = n$. Then W admits a left inverse $W^\dagger$ and $H = W^\dagger$ gives (2.16) with $S_1 = S$. More generally, if $S = PS_1$ for a known invertible $P \in \mathbb{R}^{n \times n}$ and a component-wise S_1, one takes $H = P^{-1}W^\dagger$. Conversely (2.16) forces $\operatorname{rank} W = n$, hence $q \geq n$, since the values of S_1 span $\mathbb{R}^n$, so that HW is onto and $\operatorname{rank} W \geq n$. Note that the identification of W by Theorem 2.4 required no rank assumption at all; the rank enters only when the state itself is to be recovered.

Theorem 2.15. *Assume that B is invertible, that W is known, that Assumption 2.13 holds, and that upper bounds on $\|A\|$ and on $\bar{\phi}$ are known. Let $\rho > 0$ and*

$\varepsilon > 0$, let $0 < T_1 < T_2$, and let γ be a C^1 curve with bounded derivative taking its values in a compact subset of $\mathring{\Omega}$. Then there exists $k^\star > 0$ such that, for every $k > k^\star$, every solution of (2.12) under the control $u(t) = \eta(y(t), \gamma(t))$ of (4.3) with $x(0) \in \mathcal{B}(0, \rho)$ is global in time and satisfies

$$x(t) \in \mathring{\Omega} \quad \forall t \geq T_1, \qquad \|x(t) - \gamma(t)\| \leq \varepsilon \quad \forall t \geq T_2. \tag{2.18}$$

The reference γ may be taken constant. The first assertion brings the state out of the saturation regions and keeps it out from a prescribed time on; the second additionally tracks any reference inside the unsaturated region, where the output determines the state.

Corollary 2.16 (State recovery). *Under the assumptions of Theorem 2.15, the restriction of S_1 to Ω is a homeomorphism onto $[0, 1]^n$, and for every $t \geq T_1$*

$$x_j(t) = S_0^{-1}\Big(\big(Hy(t)\big)_j\Big), \qquad j \in [\![1, n]\!]. \tag{2.19}$$

The three results compose under the additional assumptions of Theorem 2.15 (B invertible and Assumption 2.13 in force). Theorem 2.11 produces a design $\mathcal{E}$ realizable in time T and a trajectory realizing it, Theorem 2.4 applied to that same $\mathcal{E}$ reads W off the corresponding output, and the feedback of Theorem 2.15 then brings the state into the unsaturated region, where (2.19) returns it, or tracks a prescribed reference inside that region.

3. Technical lemmas and proofs

Theorems 2.10, 2.11 and 2.12 are proved by the same construction. The design $\mathcal{E}$ is ours to choose, and we take for its vectors the $v_1, \ldots, v_n$ of Assumption 2.8, which are linearly independent and therefore satisfy (2.4). By (2.13), realizing such a design amounts to bringing the state into K_1 and holding it there, far enough from the origin, during the interval J_1, then into K_2 during J_2, and so on up to K_n. Two ingredients are needed. The first is a control that brings the state into one prescribed region and keeps it there; the second is a way of chaining n such controls.

Subsections 3.1 and 3.2 prepare the first ingredient, by measuring how much room a coradiant region leaves around a given direction and by writing the trajectory produced by a constant control. Subsection 3.3 then solves the one-region problem for the three information structures of Subsection 2.2, in four cases, corresponding to A known (Lemma 3.6), A known with (A, B) controllable (Lemma 3.7), A_0 known and $\|A_1\|$ small (Lemma 3.8), and A_0 a negative multiple of the identity with $\|A_1\|$ small (Lemma 3.9). The first and third leave a transient to be waited out, the second and fourth work in arbitrary time. Subsection 3.4 chains the steps, and Subsection 3.5 proves the theorems. Theorem 2.15 and Corollary 2.16 are proved in Section 4.

3.1. Geometry of coradiant regions

Throughout, $M \geq 1$ and $\alpha > 0$ denote constants with

$$\|e^{At}\| \leq Me^{-\alpha t}, \qquad t \geq 0, \tag{3.1}$$

which exist because A is Hurwitz. When the decomposition $A = A_0 + A_1$ of Theorems 2.11 and 2.12 is in force, M_0 and α_0 denote the analogous constants for A_0. Integrating (3.1) over $[0, +\infty)$ gives $\|A^{-1}\| \leq M/\alpha$, and likewise $\|A_0^{-1}\| \leq M_0/\alpha_0$.

Solutions of (2.12) exist for the bounded measurable controls used below, are absolutely continuous, and are global as soon as they are bounded. Once a solution x and a control u are fixed we abbreviate $\phi(s) := \phi(x(s), u(s), s)$, bounded by $\bar{\phi}$.

Uniqueness is neither assumed nor used. Every conclusion below holds for every solution issued from the prescribed set of initial states.

Definition 3.1 (Margin). *Let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior, and let $v \in \mathring{K}$, $v \neq 0$. The* margin *of v in K is*

$$\delta_K(v) := \min\left\{1,\ \frac{\operatorname{dist}(v, \mathbb{R}^n \setminus K)}{\|v\|}\right\} \in (0,1]. \tag{3.2}$$

The margin measures the distance from v to the boundary of K, relative to $\|v\|$. On a convex cone it is the sine of the angle from v to the nearest face; on an orthant, and for v in its interior, it is $\min_j |v_j|/\|v\|$.

Proposition 3.2 (Stable scaling of coradiant sets). *Let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior, let $v \in \mathring{K}$ with $v \neq 0$ and set $\delta := \delta_K(v) \in (0,1]$. Then for every $\mu \geq 1$ and every $r \in \mathbb{R}^n$,*

$$\operatorname{dist}\big(\mu v + r,\ \mathbb{R}^n \setminus K\big)\ \geq\ \mu\delta\|v\| - \|r\|. \tag{3.3}$$

In particular, $\mu v + r \in K$ whenever $\|r\| \leq \mu\delta\|v\|$, and $\|\mu v + r\| \geq \mu\|v\| - \|r\|$.

Proof. By definition of δ we have $\mathcal{B}(v, \delta\|v\|) \subset K$; scaling a ball scales its centre and its radius, and coradiancy gives $\mu K \subset K$ for $\mu \geq 1$, so $\mathcal{B}(\mu v,\ \mu\delta\|v\|) = \mu\,\mathcal{B}(v, \delta\|v\|) \subset K$. If $\|r\| \leq \mu\delta\|v\|$ then $\mu v + r$ lies in that ball, hence in K; the distance estimate (3.3) then follows by the triangle inequality applied at μv, and the norm bound is the reverse triangle inequality. ∎

The four steering lemmas of Subsection 3.3 produce trajectories of the same shape, a designed term along a direction $w \in \mathring{K}$ scaled by the amplitude μ, plus a residual. A single comparison, Lemma 3.3, concludes all four cases. In it, c_1 measures the part of the residual that grows with μ, and c_2 the part that does not.

Lemma 3.3 (Margin beats residual). *Let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior, let $w \in \mathring{K}$ with $w \neq 0$ and let $\delta \in (0, \delta_K(w)]$. Let $T \geq 0$ and let x, β and ϱ satisfy*

$$x(t) = \mu\beta(t)\, w + \varrho(t), \qquad \beta(t) \geq \beta_0, \qquad \|\varrho(t)\| \leq \mu\beta(t)\, c_1\|w\| + c_2, \qquad \forall t \geq T, \tag{3.4}$$

for some constants $\beta_0 > 0$, $c_1 \in [0, \delta)$ and $c_2 \geq 0$ and some amplitude $\mu > 0$ with $\mu\beta_0 \geq 1$. If

$$\mu\,\beta_0\,(\delta - c_1)\,\|w\|\ \geq\ R + c_2, \tag{3.5}$$

then

$$x(t) \in K \quad \text{and} \quad \|x(t)\| \geq R, \qquad \forall t \geq T. \tag{3.6}$$

Proof. Fix $t \geq T$. Since $\mu\beta(t) \geq \mu\beta_0 \geq 1$, Proposition 3.2 applies to $\mu\beta(t)w$ and gives $\operatorname{dist}(\mu\beta(t)w, \mathbb{R}^n \setminus K) \geq \mu\beta(t)\delta_K(w)\|w\| \geq \mu\beta(t)\delta\|w\|$. Subtracting the residual and using $\beta(t) \geq \beta_0$ and $\delta > c_1$,

$$\mu\beta(t)\delta\|w\| - \|\varrho(t)\|\ \geq\ \mu\beta(t)(\delta - c_1)\|w\| - c_2\ \geq\ \mu\beta_0(\delta - c_1)\|w\| - c_2\ \geq\ R\ >\ 0. \tag{3.7}$$

In particular $\|\varrho(t)\| \leq \mu\beta(t)\delta\|w\|$, so Proposition 3.2, applied with $v = w$, the amplitude $\mu\beta(t)$ and $r = \varrho(t)$, gives $x(t) \in K$. Its norm bound and (3.7) give $\|x(t)\| \geq \mu\beta(t)\|w\| - \|\varrho(t)\| \geq \mu\beta(t)\delta\|w\| - \|\varrho(t)\| \geq R$, since $\delta \leq 1$. ∎

Figure 3.1 shows the comparison in the plane.

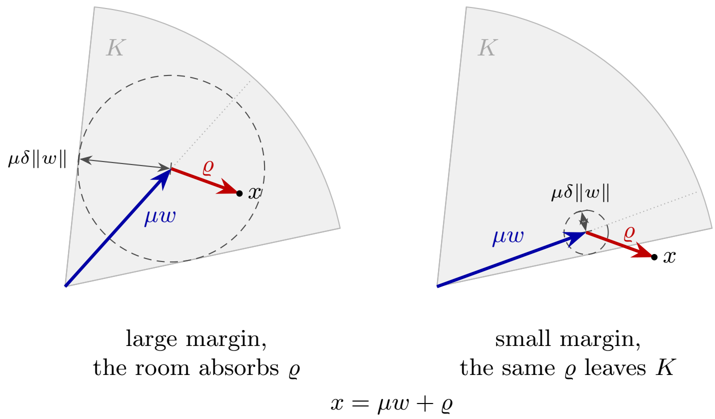

FIGURE 3.1. *The shape* (3.4)*, drawn with* $\beta \equiv 1$*. The set* K *contains the ball of radius* $\mu\delta\|w\|$ *centred at* μw *(dashed). On the left the margin absorbs the residual; on the right, the same* ϱ *leaves* K*.*

3.2. Trajectories under constant controls

The trajectories themselves are obtained from Duhamel's formula. For a solution of (2.12),

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-s)}\big[\phi(s) + Bu(s)\big]ds, \qquad t \geq 0. \tag{3.8}$$

Fix an input direction $b \in \operatorname{Im}(B)$ and take the constant control $Bu \equiv \mu b$ with $\mu > 0$. Since $\int_0^t e^{A(t-s)}ds = A^{-1}(e^{At} - \mathrm{Id})$, that control contributes $\mu(e^{At} - \mathrm{Id})A^{-1}b$ to (3.8), and grouping the constant part of this with the first term gives

$$x(t) = \underbrace{-\mu A^{-1}b}_{\text{equilibrium}} + \underbrace{e^{At}\big(x(0) + \mu A^{-1}b\big)}_{r_1(t)} + \underbrace{\int_0^t e^{A(t-s)}\phi(s)\,ds}_{r_2(t)}, \tag{3.9}$$

where, for $\|x(0)\| \leq \rho$ and by (3.1) and $\|\phi\| \leq \bar{\phi}$,

$$\|r_1(t)\| \;\leq\; Me^{-\alpha t}\big(\rho + \mu\|A^{-1}b\|\big), \qquad \|r_2(t)\| \;\leq\; \frac{M\bar{\phi}}{\alpha}, \qquad t \geq 0. \tag{3.10}$$

The amplitude scales the equilibrium and the transient r_1, but not r_2. Regrouping instead the two terms of (3.9) that carry μ,

$$x(t) = \mu\big(\mathrm{Id} - e^{At}\big)\big(-A^{-1}b\big) + e^{At}x(0) + \int_0^t e^{A(t-s)}\phi(s)\,ds \tag{3.11}$$

puts the whole dependence on the amplitude in the curve $t \mapsto (\mathrm{Id} - e^{At})(-A^{-1}b)$. We use the first form when the transient is waited out and the second when it is not. In either case (3.10) gives the uniform bound

$$\|x(t)\| \;\leq\; R_\mu := \mu(1 + M)\big\|A^{-1}b\big\| + M\rho + \frac{M\bar{\phi}}{\alpha}, \qquad t \geq 0. \tag{3.12}$$

3.3. Steering into one region

Each lemma of this subsection establishes the property below. The radius R is the one of Assumption 2.8.

Definition 3.4 (Steering into a region). *Let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior and let $T > 0$. We say that* (2.12) can be steered into K in time T *if for every $\rho > 0$ and every initial time $t_0 \geq 0$ there exist a bounded measurable control $u : [t_0, +\infty) \to \mathbb{R}^m$ and a radius $\rho' \geq \rho$, neither of them depending on the solution, such that every solution of* (2.12) *with $\|x(t_0)\| \leq \rho$ satisfies*

$$x(t) \in K \quad \textit{and} \quad \|x(t)\| \geq R \quad \forall t \geq t_0 + T, \qquad \|x(t)\| \leq \rho' \quad \forall t \geq t_0. \tag{3.13}$$

The order of the quantifiers matters. The time T is fixed before the radius ρ, so the same T serves every initial ball, and only the control and the radius ρ' depend on ρ.

Under Assumption 2.8 and for $K = K_i$, (2.13) turns (3.13) into $S(x(t)) = v_i$ for every $t \geq t_0 + T$, which is what (2.5) prescribes on J_i. The radius ρ' bounds the state at the end of the step and lets the regions be visited one after the other. The lemmas below are all stated with $t_0 = 0$. The same construction applies verbatim from any initial time, because ϕ enters (2.12) only through the time-uniform bound $\|\phi\| \leq \bar{\phi}$ and the controls produced are constant or built from e^{At}.

All the constants below depend on the equilibrium w only through its margin in the region, so we first record the largest margin available in a region. For a Hurwitz matrix A' set

$$\delta_K^*(A') := \sup\{\delta_K(w) : \ w \in \mathrm{Im}(A'^{-1}B) \cap \mathring{K}\} \ \in \ [0, 1], \tag{3.14}$$

with the convention $\sup \varnothing = 0$, so that $\delta_K^*(A') > 0$ is equivalent to $\mathrm{Im}(A'^{-1}B) \cap \mathring{K} \neq \varnothing$. Every margin below $\delta_K^*(A')$ is achieved by some admissible direction. Indeed, for $\delta < \delta_K^*(A')$ there is $b \in \mathrm{Im}(B)$ with $w = -A'^{-1}b \in \mathring{K}$ and $\delta_K(w) > \delta$.

Each lemma below is stated for a fixed direction and its own margin $\delta_K(w)$, and the theorems of Subsection 3.5 take that margin close to δ_K^*.

Remark 3.5. The supremum in (3.14) is in general not a maximum, and closedness of K does not help. By Proposition 3.2 applied with $r = 0$, the margin δ_K is nondecreasing along each ray, while the competitors are unbounded, so the best of them escape to infinity. For the component-wise saturations of Remark 2.9 the margin increases strictly along each ray, towards the limit $\min_j |w_j|/\|w\|$, which is therefore never attained.

Lemma 3.6 (Steering, known linear part). *Let A be Hurwitz, let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior, and let $b \in \mathrm{Im}(B)$ be such that*

$$w := -A^{-1}b \ \in \ \mathrm{Im}(A^{-1}B) \cap \mathring{K}. \tag{3.15}$$

Set $\delta := \delta_K(w)$. Then (2.12) *can be steered into K in every time*

$$T \ > \ T_{\min} := \frac{1}{\alpha} \ln \frac{M}{\delta}. \tag{3.16}$$

Precisely, for every $\rho > 0$ and every

$$\mu \ \geq \ \max\left\{1, \ \frac{R + M\rho + M\bar{\phi}/\alpha}{\left(\delta - Me^{-\alpha T}\right)\|w\|}\right\}, \tag{3.17}$$

the constant control $Bu \equiv \mu b$ drives every solution of (2.12) *with $x(0) \in \mathcal{B}(0, \rho)$ so that*

$$x(t) \in K \quad \textit{and} \quad \|x(t)\| \geq R \qquad \forall t \geq T, \tag{3.18}$$

and $\|x(t)\| \leq R_\mu$ for every $t \geq 0$.

Proof. Since $b \in \operatorname{Im}(B)$ the constant control $Bu \equiv \mu b$ is admissible, and Duhamel's formula (3.8) together with $\int_0^t e^{A(t-s)}ds = A^{-1}(e^{At} - \mathrm{Id})$ gives, as in (3.9),

$$x(t) = \mu w + \varrho(t), \qquad \varrho(t) := e^{At}\big(x(0) - \mu w\big) + \int_0^t e^{A(t-s)}\phi(s)\,ds, \qquad t \geq 0. \tag{3.19}$$

By (3.10) and $Me^{-\alpha t}\rho \leq M\rho$, for $t \geq T$,

$$\|\varrho(t)\| \;\leq\; \underbrace{\mu M e^{-\alpha T}\|w\|}_{\text{grows with } \mu} + \underbrace{M\rho + \frac{M\bar{\phi}}{\alpha}}_{\text{does not}}. \tag{3.20}$$

This is (3.4) with $\beta \equiv \beta_0 = 1$, $c_1 = Me^{-\alpha T}$ and $c_2 = M\rho + M\bar{\phi}/\alpha$. The condition $T > T_{\min}$ of (3.16) is exactly $c_1 < \delta$, and (3.17) is exactly (3.5), so Lemma 3.3 gives (3.18). The bound $\|x(t)\| \leq R_\mu$ on $[0, +\infty)$ is (3.12), and $R_\mu \geq M\rho \geq \rho$, so (3.13) holds with $\rho' = R_\mu$. ∎

The minimal time reflects the first term of (3.20). Transient and margin grow alike with μ, so no amplitude shortens the wait, whereas against the μ-free second term the amplitude is effective.

That transient can be removed when the pair (A, B) is controllable. Recall that the controllability Gramian

$$\mathcal{G}(T) := \int_0^T e^{As}BB^\top e^{A^\top s}\,ds \tag{3.21}$$

is then positive definite for every $T > 0$, and that for any target $v \in \mathbb{R}^n$ the control $u(t) = B^\top e^{A^\top(T-t)}\mathcal{G}(T)^{-1}v$ satisfies $\int_0^T e^{A(T-s)}Bu(s)\,ds = v$, see [9, 5]. Given an amplitude $\mu > 0$ and a direction b with equilibrium $w = -A^{-1}b$, we steer to μw at time T and hold the state there, using

$$u(t) = \begin{cases} \mu\, B^\top e^{A^\top(T-t)}\mathcal{G}(T)^{-1}w, & t \in [0, T), \\ u_b, & t \geq T, \end{cases} \tag{3.22}$$

where $u_b \in \mathbb{R}^m$ is any vector with $Bu_b = \mu b$.

Lemma 3.7 (Steering in arbitrary time, controllable case). *Let A be Hurwitz with (A, B) controllable, let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior, and let $b \in \operatorname{Im}(B)$ be such that $w := -A^{-1}b \in \mathring{K}$. Set $\delta := \delta_K(w)$. Then (2.12) can be steered into K in* every *time $T > 0$. For every $\rho > 0$ and every*

$$\mu \;\geq\; \max\left\{1,\; \frac{R + (1+M)\big(M\rho + M\bar{\phi}/\alpha\big)}{\delta\,\|w\|}\right\}, \tag{3.23}$$

the control (3.22) drives every solution of (2.12) with $x(0) \in \mathcal{B}(0, \rho)$ so that

$$x(t) \in K \quad \text{and} \quad \|x(t)\| \geq R \qquad \forall t \geq T, \tag{3.24}$$

and $\|x(t)\| \leq \rho'$ for every $t \geq 0$, where

$$\rho' := \mu\big(1 + c_B(T)\big)\|w\| + (1+M)\left(M\rho + \frac{M\bar{\phi}}{\alpha}\right), \qquad c_B(T) := \sup_{0 \leq t \leq T}\left\|\int_0^t e^{A(t-s)}BB^\top e^{A^\top(T-s)}ds \right. \tag{3.25}$$

Proof. The open-loop part of (3.22) places the state at μw at time T, up to the contributions of the initial condition and of ϕ. $\int_0^T e^{A(T-s)}Bu(s)\,ds = \mu w$ gives $\|x(T) - \mu w\| \leq M\rho + M\bar{\phi}/\alpha$ by Duhamel's formula (3.8) at $t = T$. On $[T, +\infty)$

the control is the constant $Bu \equiv \mu b$, so Duhamel's formula run from the initial time T and the initial state $x(T)$, regrouped as in (3.19), gives $x(t) = \mu w + \varrho(t)$ with $\varrho(t) = e^{A(t-T)}(x(T) - \mu w) + \int_T^t e^{A(t-s)}\phi(s)\,ds$, whence

$$\|\varrho(t)\| \;\le\; (1+M)\left(M\rho + \frac{M\bar{\phi}}{\alpha}\right), \qquad t \ge T. \tag{3.26}$$

This is (3.4) with $\beta \equiv \beta_0 = 1$, $c_1 = 0$ and c_2 the right-hand side of (3.26), and (3.23) is then exactly (3.5), so Lemma 3.3 gives (3.24). As for the state bound, on $[0,T]$ Duhamel's formula and (3.22) give $\|x(t)\| \le M\rho + M\bar{\phi}/\alpha + \mu c_B(T)\|w\|$, where $c_B(T)$, defined in (3.25), is finite, while on $[T,+\infty)$ we have $\|x(t)\| \le \mu\|w\| + \|\varrho(t)\|$. Both are at most ρ'. ◻

Because the target is reached exactly at T, the residual (3.26) carries no term growing with μ, and the minimal time disappears; what grows as T decreases is the control effort, through $\mathcal{G}(T)^{-1}$.

We turn to the case where only a part A_0 of the linear dynamics is known, the control being built from A_0 while the equilibrium it reaches, $-A^{-1}b$, involves the unknown part. Write

$$p := \left\|A_0^{-1}A_1\right\| \;\le\; \frac{M_0}{\alpha_0}\|A_1\|, \qquad \vartheta := \frac{p}{1-p} \quad (p<1), \tag{3.27}$$

so that ϑ measures, relative to $\|A_0^{-1}b\|$, how far A_1 can move that equilibrium. One comparison governs the case. The displacement ϑ must be shorter than the margin of the nominal direction. The counterpart of the state bound (3.12), computable from A_0, B and b alone, is

$$R_\mu^0 := \mu(1+M_0)(1+\vartheta)\left\|A_0^{-1}b\right\| + M_0\rho + \frac{2M_0\bar{\phi}}{\alpha_0}. \tag{3.28}$$

Lemma 3.8 (Steering, partially known linear part). *Let A_0 be Hurwitz with constants M_0, α_0 as in (3.1), let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior, and let $b \in \operatorname{Im}(B)$ be such that $w_0 := -A_0^{-1}b \in \mathring{K}$. Set $\delta_0 := \delta_K(w_0)$ and assume*

$$\|A_1\| \;<\; \frac{\alpha_0}{M_0}\cdot\frac{\delta_0}{1+\delta_0}, \qquad \textit{equivalently, by (3.27),} \qquad \vartheta \;<\; \delta_0. \tag{3.29}$$

Then $A = A_0 + A_1$ is Hurwitz and, writing

$$\hat{\delta}_0 := \frac{\delta_0 - \vartheta}{1+\vartheta} \;\in\; (0,1], \qquad T^0_{\min} := \frac{2}{\alpha_0}\ln\frac{M_0}{\hat{\delta}_0}, \tag{3.30}$$

the system (2.12) can be steered into K in every time $T > T^0_{\min}$. Namely, for every $\rho > 0$ and every

$$\mu \;\ge\; \max\left\{1,\; \frac{R + M_0\rho + 2M_0\bar{\phi}/\alpha_0}{\left(\hat{\delta}_0 - M_0e^{-\alpha_0T/2}\right)(1-\vartheta)\|w_0\|}\right\}, \tag{3.31}$$

the constant control $Bu \equiv \mu b$ drives every solution of (2.12) with $x(0) \in \mathcal{B}(0,\rho)$ so that

$$x(t) \in K \quad \textit{and} \quad \|x(t)\| \ge R \qquad \forall t \ge T, \tag{3.32}$$

and $\|x(t)\| \le R_\mu^0$ for every $t \ge 0$.

Proof. Step 1: where the equilibrium goes. Write $P := A_0^{-1}A_1$, so that $A = A_0(\mathrm{Id} + P)$ with $\|P\| = p$. By (3.29) and $\delta_0 \le 1$ we have $p < \delta_0/(1+\delta_0) \le 1/2$, so the Neumann series inverts $\mathrm{Id} + P$, hence A, and the equilibrium $w := -A^{-1}b$ is $(\mathrm{Id} + P)^{-1}w_0$, with

$$\|w - w_0\| = \big\|(\mathrm{Id} + P)^{-1}Pw_0\big\| \;\le\; \frac{p}{1-p}\|w_0\| \;=\; \vartheta\|w_0\|, \tag{3.33}$$

whence $(1-\vartheta)\|w_0\| \le \|w\| \le (1+\vartheta)\|w_0\|$ by the triangle inequality. Since $p < \delta_0/(1+\delta_0)$ is exactly $\vartheta < \delta_0$, the displacement (3.33) is shorter than the radius $\delta_0\|w_0\|$ of the ball that K contains around w_0, so

$$\mathrm{dist}\big(w, \mathbb{R}^n \setminus K\big) \;\ge\; (\delta_0 - \vartheta)\|w_0\| \;>\; 0,$$

which places w in $\mathring{K}$. Dividing by $\|w\| \le (1+\vartheta)\|w_0\|$ and using $\hat{\delta}_0 \le 1$ gives

$$\delta_K(w) \;\ge\; \hat{\delta}_0 \;>\; 0. \tag{3.34}$$

Step 2: the decay of e^{At}. From $e^{At} = e^{A_0t} + \int_0^t e^{A_0(t-s)}A_1e^{As}ds$ and (3.1) for A_0, the function $\psi(t) := e^{\alpha_0 t}\|e^{At}\|$ satisfies $\psi(t) \le M_0 + M_0\|A_1\|\int_0^t \psi(s)ds$, so Grönwall's lemma gives $\|e^{At}\| \le M_0e^{-(\alpha_0 - M_0\|A_1\|)t}$. Here $M_0\|A_1\| < \alpha_0\delta_0/(1+\delta_0) \le \alpha_0/2$, so (3.1) holds with $M = M_0$ and $\alpha > \alpha_0/2$. In particular A is Hurwitz and $M\bar{\phi}/\alpha < 2M_0\bar{\phi}/\alpha_0$.

Step 3: Duhamel along the perturbed equilibrium. The control $Bu \equiv \mu b$ is admissible and constant, so (3.19) holds verbatim with the equilibrium w of Step 1, and Step 2 bounds its residual, for $t \ge T$, by

$$\|\varrho(t)\| \;\le\; \mu M_0 e^{-\alpha_0 T/2}\|w\| + M_0\rho + \frac{2M_0\bar{\phi}}{\alpha_0}. \tag{3.35}$$

This is (3.4) with $\beta \equiv \beta_0 = 1$, $c_1 = M_0e^{-\alpha_0T/2}$ and $c_2 = M_0\rho + 2M_0\bar{\phi}/\alpha_0$, and we apply Lemma 3.3 with the computable lower bound $\delta = \hat{\delta}_0$ of (3.34). The condition $T > T^0_{\min}$ of (3.30) is exactly $c_1 < \hat{\delta}_0$. Since $\|w\| \ge (1-\vartheta)\|w_0\|$ by Step 1, (3.31) implies (3.5). Lemma 3.3 therefore gives (3.32). Finally $\|x(t)\| \le \mu\|w\| + \|\varrho(t)\| \le \mu(1+M_0)\|w\| + M_0\rho + 2M_0\bar{\phi}/\alpha_0 \le R^0_\mu$ on $[0, +\infty)$, again by $\|w\| \le (1+\vartheta)\|w_0\|$. ☐

The factor 2 in R^0_μ and in (3.31) absorbs the decay rate of e^{At}; Step 2 of the proof gives (3.1) with $M = M_0$ and some $\alpha > \alpha_0/2$.

Every quantity in (3.29)–(3.31) is computed from A_0, B, K, ρ, $\bar{\phi}$ and an upper bound on $\|A_1\|$. Neither A_1 itself nor the true margin $\delta_K(w)$ of the realized equilibrium $w = -A^{-1}b$ enters, $\hat{\delta}_0$ being a computable lower bound for the latter. All the constants are monotone in ϑ, so an upper bound may be substituted for it throughout.

Condition (3.29) compares a displacement with a margin, both measured relative to $\|w_0\|$. The unknown part may move the equilibrium anywhere inside the ball that K already contains around the nominal one, and no further; this is the comparison of Figure 3.1, with the displacement $\vartheta\|w_0\|$ in place of $\|\varrho\|$. Raising the amplitude does not repair a violation. By (3.19) the state is μw up to an error that grows strictly more slowly than μ, so once w has left K, a larger amplitude only drives the state further from the region. The hypothesis is monotone in δ_0, so a direction realizing it exists as soon as

$$\|A_1\| \;<\; \frac{\alpha_0}{M_0} \cdot \frac{\delta^*_K(A_0)}{1 + \delta^*_K(A_0)}, \tag{3.36}$$

a condition on the data alone. It suffices to take δ_0 close enough to $\delta^*_K(A_0)$ in (3.14).

When the known part is a multiple of the identity, the transient no longer has to be waited out. What must stay inside K is then the curve $t \mapsto (\mathrm{Id} - e^{At})(-A^{-1}b)$ of (3.11), which stays within a controlled error of its limit at every time. The counterpart is a stronger condition on A_1.

Lemma 3.9 (Steering in arbitrary time, scalar known part). *Let $A_0 = -\lambda\,\mathrm{Id}$ with $\lambda > 0$, so that $M_0 = 1$, $\alpha_0 = \lambda$ and $p = \|A_1\|/\lambda$ in (3.27). Let $K \subset \mathbb{R}^n$ be closed and coradiant with nonempty interior and let $b \in \mathrm{Im}(B)$ be such that $w_0 := b/\lambda \in \mathring{K}$. Set $\delta_0 := \delta_K(w_0)$ and assume*

$$2\vartheta \; < \; \delta_0. \tag{3.37}$$

Then (2.12) can be steered into K in every *time $T > 0$. For every $\rho > 0$ and every*

$$\mu \;\geq\; \frac{1}{1 - e^{-\lambda T}} \max\left\{1,\;\; \frac{\lambda(1+\vartheta)\big(R + \rho + 3\bar{\phi}/(2\lambda)\big)}{(1-\vartheta)(\delta_0 - 2\vartheta)\,\|b\|}\right\}, \tag{3.38}$$

the constant control $Bu \equiv \mu b$ drives every solution of (2.12) with $x(0) \in \mathcal{B}(0,\rho)$ so that

$$x(t) \in K \quad \text{and} \quad \|x(t)\| \geq R \qquad \forall t \geq T, \tag{3.39}$$

and $\|x(t)\| \leq R^0_\mu$ for every $t \geq 0$.

Proof. Since $2\vartheta < \delta_0 \leq 1$ we have $\vartheta < \delta_0/2$, hence $p = \vartheta/(1+\vartheta) < \delta_0/(2+\delta_0) < \delta_0/(1+\delta_0)$, and $p = \|A_1\|/\lambda$ makes this the hypothesis (3.29) of Lemma 3.8. Step 1 of its proof places $w := -A^{-1}b$ in $\mathring{K}$, with

$$\delta_K(w) \;\geq\; \hat{\delta}_0 = \frac{\delta_0 - \vartheta}{1+\vartheta}, \qquad \|w\| \;\geq\; (1-\vartheta)\frac{\|b\|}{\lambda}. \tag{3.40}$$

Moreover $\vartheta < 1/2$ gives $p < 1/3$, hence, in the estimates below, $\alpha = \lambda(1-p) > 2\lambda/3$.

Since $-\lambda\,\mathrm{Id}$ commutes with A_1 we have $e^{At} = e^{-\lambda t}e^{A_1 t}$, so $\|e^{At}\| \leq e^{-(\lambda - \|A_1\|)t}$ and $M = 1$ may be taken in (3.1). Under the constant control $Bu \equiv \mu b$, Duhamel's formula (3.8) written in the form (3.11), which keeps the whole amplitude in a single term, reads

$$x(t) = \mu\big(\mathrm{Id} - e^{At}\big)w + \varrho(t), \qquad \varrho(t) := e^{At}x(0) + \int_0^t e^{A(t-s)}\phi(s)\,ds, \qquad \|\varrho(t)\| \leq \rho + \frac{\bar{\phi}}{\alpha}, \tag{3.41}$$

the residual ϱ being bounded uniformly in t and in μ. It therefore suffices to keep the curve $t \mapsto (\mathrm{Id} - e^{At})w$ far enough inside K for every $t \geq T$. Now

$$\big(\mathrm{Id} - e^{At}\big)w = \big(1 - e^{-\lambda t}\big)w + z(t), \qquad z(t) := e^{-\lambda t}\big(\mathrm{Id} - e^{A_1 t}\big)w. \tag{3.42}$$

The first term lies on the ray $\mathbb{R}_+ w$ for every $t > 0$; in the general case the transient rotates the state away from the ray, whereas here the entire deviation is carried by z. From $\|\mathrm{Id} - e^{A_1 t}\| \leq e^{\|A_1\|t} - 1$,

$$\|z(t)\| \;\leq\; e^{-\lambda t}\big(e^{\|A_1\|t} - 1\big)\|w\| \;=\; \big(1 - e^{-\lambda t}\big)\frac{e^{-\lambda t}\big(e^{\|A_1\|t} - 1\big)}{1 - e^{-\lambda t}}\,\|w\|, \tag{3.43}$$

and we claim that the fraction is bounded by $p = \|A_1\|/\lambda$ for every $t > 0$. Setting $s := e^{-\lambda t} \in (0,1)$, it reads $(s^{1-p} - s)/(1-s)$. Since $p \in (0,1)$, the function $s \mapsto s^{1-p}$ is concave, hence lies below its tangent at $s = 1$, which is $s^{1-p} \leq (1-p)s + p$; rearranging gives $(s^{1-p} - s)/(1-s) \leq p$. Hence

$$\|z(t)\| \;\leq\; \big(1 - e^{-\lambda t}\big)\frac{\|A_1\|}{\lambda}\,\|w\|, \qquad t > 0. \tag{3.44}$$

Both terms of (3.42) carry the same factor $1-e^{-\lambda t}$, so their comparison is independent of t and of μ, and no minimal time is needed.

Substituting (3.42) into (3.41) gives $x(t) = \mu\big(1-e^{-\lambda t}\big)w + \big(\mu z(t) + \varrho(t)\big)$, which by (3.44) is (3.4) with

$$\beta(t) = 1-e^{-\lambda t} \;\geq\; \beta_0 := 1-e^{-\lambda T}, \qquad c_1 = \frac{\|A_1\|}{\lambda} = \frac{\vartheta}{1+\vartheta}, \qquad c_2 = \rho + \frac{\bar\phi}{\alpha} \;<\; \rho + \frac{3\bar\phi}{2\lambda},$$

for $t \geq T$, since $t \mapsto 1-e^{-\lambda t}$ is nondecreasing. We apply Lemma 3.3 with $\delta = \hat\delta_0$, legitimate by (3.40). Then

$$\delta - c_1 \;\geq\; \frac{\delta_0 - \vartheta}{1+\vartheta} - \frac{\vartheta}{1+\vartheta} \;=\; \frac{\delta_0 - 2\vartheta}{1+\vartheta} \;>\; 0 \tag{3.45}$$

by (3.37). With $\|w\| \geq (1-\vartheta)\|b\|/\lambda$, the condition (3.5) is implied by

$$\mu\big(1-e^{-\lambda T}\big)\frac{(\delta_0 - 2\vartheta)(1-\vartheta)}{1+\vartheta}\cdot\frac{\|b\|}{\lambda} \;\geq\; R + \rho + \frac{3\bar\phi}{2\lambda},$$

which is the second entry of the maximum in (3.38), the requirement $\mu\beta_0 \geq 1$ being the first. Lemma 3.3 therefore gives (3.39). Finally, $\|e^{At}\| \leq 1$ gives $\|x(t)\| \leq \mu\big(1 + \|e^{At}\|\big)\|w\| + \rho + 3\bar\phi/(2\lambda) \leq 2\mu(1+\vartheta)\|w_0\| + \rho + 2\bar\phi/\lambda = R^0_\mu$ on $[0,+\infty)$, since $M_0 = 1$ and $\alpha_0 = \lambda$ here. ∎

As in Lemma 3.8, a suitable direction exists as soon as A_1 obeys a single bound on the data, here

$$\|A_1\| \;<\; \lambda \cdot \frac{\delta^*_K(A_0)}{2 + \delta^*_K(A_0)}, \tag{3.46}$$

which is equivalent to $2\vartheta < \delta^*_K(A_0)$. A direction of margin $\delta_0 > 2\vartheta$ then exists by (3.14).

Condition (3.46) does not involve T; it acts on the amplitude instead, which by (3.38) grows like $1/(\lambda T)$ as the prescribed time shrinks. It is also of a different nature from (3.29); it bounds the rotation of the curve $t \mapsto (\mathrm{Id} - e^{At})w$ at small times, whereas (3.29) bounds the displacement of the equilibrium. Both conditions enter here, and their sum is the factor two of (3.37). Sharpening the estimate would not improve the threshold, since $\|A_1\|/\lambda$ is the exact supremum of the fraction bounded in the proof, approached as $t \to 0$.

3.4. Concatenating the steps

Lemma 3.10 (Sequential steering). *Let Assumption 2.8 hold and, for each $i \in [\![1,n]\!]$, let (2.12) be steerable into K_i in some time $T_i > 0$ in the sense of Definition 3.4. Then for every $\rho > 0$ and every $T > \sum_{i=1}^n T_i$ there exist an excitation design $\mathcal{E} = \{(v_i, J_i)\}_{i\in[\![1,n]\!]}$ realizable in time T and a control $u : [0,T] \to \mathbb{R}^m$ such that every solution of (2.12) with $x(0) \in \mathcal{B}(0,\rho)$ satisfies*

$$x(t) \in K_i \quad \text{and} \quad S(x(t)) = v_i \qquad \forall t \in J_i, \quad \forall i \in [\![1,n]\!]. \tag{3.47}$$

Proof. Write $s_0 := 0$ and choose dwell times $\tau_i > 0$ with $\sum_{i=1}^n (T_i + \tau_i) < T$, which is possible since $T > \sum_i T_i$. Set

$$t_1 := T_1, \qquad t_i := s_{i-1} + T_i, \qquad s_i := t_i + \tau_i, \quad i \in [\![1,n]\!], \tag{3.48}$$

and $J_i := [t_i, s_i]$, so that $t_1 \geq T_1$, $t_i \geq s_{i-1} + T_i$ and $s_n < T$; in particular $\mathcal{E} := \{(v_i, J_i)\}_{i\in[\![1,n]\!]}$ is realizable in time T. The construction of the control is by induction on i, the induction hypothesis being a bound ρ_i on the norm of the state at time s_{i-1}, with $\rho_1 := \rho$. Given ρ_i, Definition 3.4 applied with the initial time s_{i-1} furnishes for K_i and the radius ρ_i a control u_i and a radius ρ_i' such that any solution issued at time s_{i-1} from a state of norm at most ρ_i satisfies $x(t) \in K_i$ and $\|x(t)\| \geq R$ for every $t \geq s_{i-1} + T_i$, hence for every $t \in J_i$ by (3.48), and $\|x(t)\| \leq \rho_i'$ throughout. By (2.13), $S(x(t)) = v_i$ on J_i, which is (3.47). Setting $\rho_{i+1} := \rho_i'$ closes the induction, and the concatenation $u := \sum_{i=1}^{n} u_i \mathbf{1}_{[s_{i-1}, s_i)}$, extended to $[0, T]$ by u_n on $[s_n, T]$, is the desired control. The v_i are linearly independent by Assumption 2.8, so $\mathcal{E}$ is a design of full rank. □

Only the amplitudes and the radii ρ_i' carried from one step to the next depend on the initial radius. The times T_i do not, so the induction does not lengthen the horizon as it progresses, and the T^* of Theorems 2.10 and 2.11 is the same for every initial ball.

3.5. Proofs of the theorems

Proof of Theorem 2.10. For each i the hypothesis is (3.15) with $K = K_i$. Let $\delta_i := \delta^*_{K_i}(A) > 0$. By Lemma 3.6, a direction $w \in \operatorname{Im}(A^{-1}B) \cap \mathring{K}_i$ makes (2.12) steerable into K_i in every time $T_i > \alpha^{-1} \ln\big(M/\delta_{K_i}(w)\big)$, and (3.14) makes every margin below δ_i available in K_i. Set

$$T^* := \sum_{i=1}^{n} \frac{1}{\alpha} \ln \frac{M}{\delta_i} \;\geq\; 0, \tag{3.49}$$

which is nonnegative because $M \geq 1 \geq \delta_i$. Given $T > T^*$, the margins $\delta_{K_i}(w)$ can be taken close enough to the δ_i that the sum of the corresponding times T_i is below T, and Lemma 3.10 then returns an excitation design realizable in time T and a control realizing it. □

The larger the margins available in the K_i, the shorter T^*.

Proof of Theorem 2.11. For each i let $\delta_i := \delta^*_{K_i}(A_0) > 0$, a quantity determined by A_0, B and K_i alone, and set

$$\varepsilon_0 := \frac{\alpha_0}{M_0} \min_{1 \leq i \leq n} \frac{\delta_i}{3 + \delta_i}. \tag{3.50}$$

Let $\|A_1\| \leq \varepsilon_0$, so that $(M_0/\alpha_0)\|A_1\| \leq \delta_i/(3 + \delta_i)$ for every i. By (3.14) each K_i contains a direction of margin δ_0 as close to δ_i as desired, and since $\delta_i/(3 + \delta_i) < \delta_0/(1 + \delta_0)$ once δ_0 is close enough to δ_i, such a direction satisfies the hypothesis (3.29) of Lemma 3.8. That lemma therefore applies to each K_i, with $\vartheta \leq \delta_i/3$, and as $\delta_0 \to \delta_i$ the time in (3.30) decreases to at most

$$T^* := \sum_{i=1}^{n} \frac{2}{\alpha_0} \ln \frac{M_0(3 + \delta_i)}{2\delta_i} \;>\; 0, \tag{3.51}$$

because $\vartheta \leq \delta_i/3$ gives $(\delta_i - \vartheta)/(1 + \vartheta) \geq 2\delta_i/(3 + \delta_i)$. Given $T > T^*$, choosing the margins close enough to the δ_i gives times T_i with $\sum_i T_i < T$, as in the previous proof, and Lemma 3.10 concludes. The directions, the margins δ_i, the times T_i and the amplitudes are all computed from A_0, B, the K_i, $\bar{\phi}$, ρ and the bound ε_0; A_1 is nowhere used. □

Remark 3.11 (Choice of ε_0). The constant 3 in (3.50), in place of the 1 of (3.36), leaves room for the stronger condition (3.37) of Lemma 3.9, which case *ii.* of Theorem 2.12 meets exactly. The same ε_0 then serves both theorems, with a correspondingly larger T^* here.

Proof of Theorem 2.12, case i.. Here $A_1 = 0$, so $A = A_0$ is known and Hurwitz and the hypothesis of Theorem 2.11 is (3.15) for each K_i. Since (A, B) is controllable, Lemma 3.7 makes (2.12) steerable into K_i in *every* time $T_i > 0$. Given $T > 0$, taking the T_i small enough that $\sum_i T_i < T$ and applying Lemma 3.10 concludes. Hence $T^* = 0$. ∎

Proof of Theorem 2.12, case ii.. Here $A_0 = -\lambda\,\mathrm{Id}$ with $\lambda > 0$, so $M_0 = 1$, $\alpha_0 = \lambda$ and $A = -\lambda\,\mathrm{Id} + A_1$ has the structure required by Lemma 3.9. By (3.50), $\|A_1\| \le \varepsilon_0$ gives $\vartheta \le \delta_i/3$ for every i, so $2\vartheta \le 2\delta_i/3 < \delta_i$ and (3.14) supplies in each K_i a direction of margin $\delta_0 > 2\vartheta$, which is (3.37). Lemma 3.9 then makes (2.12) steerable into K_i in *every* time $T_i > 0$, and one concludes as in case *i*. Hence $T^* = 0$, the control on the i-th step being computed from λ, B, K_i, $\bar{\phi}$, ρ and the bound ε_0. ∎

4. Output feedback and state recovery

4.1. The output feedback law

The feedback constructed here reads the measurement coordinate by coordinate. Where a coordinate indicates saturation it pushes the state back with a constant effort; where the measurement lies in the invertible range of the saturation it switches to proportional tracking. No direction enters the law, and a scalar gain k replaces the amplitude of Section 3. The construction uses H, S_0 and W, all taken known; A and ϕ remain unknown and enter only through a known linear-growth bound on the drift, stated in Lemma 4.1. For $k > 0$, $\omega \in (0, 1/2)$ and a scalar reference value $\gamma \in \mathbb{R}$, define the scalar feedback

$$\eta_0(\sigma; k, \gamma, \omega) = \begin{cases} k, & \sigma \le 0, \\ \ell_1(\sigma; k, \gamma), & 0 < \sigma < \omega, \\ -k\big(S_0^{-1}(\sigma) - \gamma\big), & \omega \le \sigma \le 1 - \omega, \\ \ell_2(\sigma; k, \gamma), & 1 - \omega < \sigma < 1, \\ -k, & \sigma \ge 1, \end{cases} \tag{4.1}$$

where ℓ_1 and ℓ_2 are the functions of σ, affine on $[0, \omega]$ and on $[1 - \omega, 1]$ respectively, making η_0 continuous. Figure 4.1 shows the five branches.

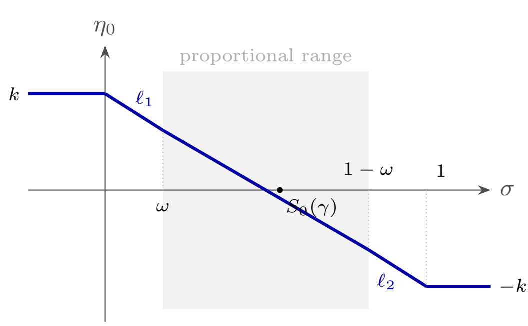

FIGURE 4.1. *The scalar feedback* (4.1) *as a function of the measured coordinate* σ. *Outside* $[0, 1]$ *the measurement carries no information about the state and the law applies the constant effort* $\pm k$. *On* $[\omega, 1 - \omega]$ *it is the proportional term* $-k(S_0^{-1}(\sigma) - \gamma)$, *which vanishes at* $\sigma = S_0(\gamma)$. *The affine pieces* ℓ_1 *and* ℓ_2 *make the law continuous across the two deadbands.*

The feedback of Theorem 2.15 is obtained by applying (4.1) to each coordinate of $Hy = S_1(x)$, with the corresponding coordinate of the reference, and inverting B.

Writing, for $\sigma \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}^n$,

$$\boldsymbol{\eta}_0(\sigma; k, \gamma, \omega) := \big(\eta_0(\sigma_1; k, \gamma_1, \omega), \dots, \eta_0(\sigma_n; k, \gamma_n, \omega)\big)^\top, \tag{4.2}$$

the feedback is

$$\eta(y, \gamma) := B^{-1}\boldsymbol{\eta}_0\big(Hy; k, \gamma, \omega\big). \tag{4.3}$$

Since B^{-1} is applied last, the control enters the dynamics as

$$Bu = \boldsymbol{\eta}_0\big(S_1(x); k, \gamma, \omega\big), \tag{4.4}$$

so that the loop is closed coordinate by coordinate and the analysis is scalar.

For $\omega \in (0, 1/2)$ write

$$\Omega_\omega := \big[S_0^{-1}(\omega),\ S_0^{-1}(1-\omega)\big]^n \subset \mathring{\Omega} \tag{4.5}$$

for the box obtained from Ω by staying at distance ω from the saturation levels, measured on the output side, and, for $L > 0$ and $X \subset \mathbb{R}^n$, let $C^{1,L}(\mathbb{R}, X)$ denote the set of C^1 curves with values in X whose derivative is bounded by L.

Lemma 4.1 (Tracking under a saturated measurement). *Let B be invertible, let $g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n$ be continuous with*

$$\|g(x, u, t)\| \le a\|x\| + c_g, \qquad (x, u, t) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}, \tag{4.6}$$

for some known *constants $a, c_g > 0$, and consider*

$$\dot{x} = g(x, u, t) + Bu, \qquad y = WS(x), \tag{4.7}$$

with W known and H, S_0, S_1 as in Assumption 2.13. Let $\varepsilon, \rho, L > 0$, let $0 < T_1 < T_2$, let $\omega \in (0, 1/4)$ and let $\gamma \in C^{1,L}(\mathbb{R}, \Omega_{2\omega})$. Then there exists $k^\star > 0$, given by (4.12) below, such that for every gain $k > k^\star$ the solutions of (4.7) under the control $u(t) = \eta(y(t), \gamma(t))$ of (4.3), issued from $\mathcal{B}(0, \rho)$, are global in time and satisfy

$$x(t) \in \Omega_\omega \quad \forall t \ge T_1, \qquad \|x(t) - \gamma(t)\| \le \varepsilon \quad \forall t \ge T_2. \tag{4.8}$$

Proof. By (2.16), $Hy(t) = HWS(x(t)) = S_1(x(t))$, so (4.3) is indeed an output feedback. Continuity of S, g and η_0 gives local existence and absolute continuity of the solutions, and as noted after (4.3) the closed loop reads

$$\dot{x}_j = g_j(x, u, t) + \eta_0\big(S_0(x_j); k, \gamma_j, \omega\big), \qquad j \in [\![1, n]\!]. \tag{4.9}$$

Set

$$\Lambda := \max\big\{\rho, |\underline{R}|, |\bar{R}|\big\}, \qquad \kappa := a\sqrt{n}\,\Lambda + c_g + L, \tag{4.10}$$

$$c_\omega := \min\Big\{1,\ S_0^{-1}(1-\omega) - S_0^{-1}(1-2\omega),\ S_0^{-1}(2\omega) - S_0^{-1}(\omega)\Big\}\ >\ 0, \tag{4.11}$$

the positivity of c_ω following from $\omega < 1/4$ and the strict monotonicity of S_0 on $[\underline{R}, \bar{R}]$, and

$$k^\star := \max\left\{\frac{1}{c_\omega}\left(\frac{2\Lambda}{T_1} + a\sqrt{n}\,(\Lambda + 1) + c_g\right),\ \ \frac{2\sqrt{n}\,\kappa}{\varepsilon},\ \ \frac{1}{T_2 - T_1}\ln\frac{2\sqrt{n}\,(\bar{R} - \underline{R})}{\varepsilon}\right\}. \tag{4.12}$$

Throughout, $k > k^\star$, whose first term in particular gives

$$k\ >\ \frac{a\sqrt{n}\,(\Lambda + 1) + c_g}{c_\omega}. \tag{4.13}$$

Step 1: the feedback is uniformly negative on the upper saturated range. Since $\gamma_j(t) \in [S_0^{-1}(2\omega), S_0^{-1}(1-2\omega)]$, the value of η_0 at $\sigma = 1-\omega$ is $-k(S_0^{-1}(1-\omega)-\gamma_j) \leq -kc_\omega$ by (4.11), and its value at $\sigma = 1$ is $-k \leq -kc_\omega$. Being affine in σ on $[1-\omega, 1]$ and constant beyond, η_0 is therefore at most $-kc_\omega$ on the whole of $[1-\omega, +\infty)$. Symmetrically $\eta_0 \geq kc_\omega$ on $(-\infty, \omega]$. In terms of the state,

$$\eta_0\big(S_0(\xi); k, \gamma_j, \omega\big) \begin{cases} \leq -kc_\omega, & \xi \geq S_0^{-1}(1-\omega), \\ \geq kc_\omega, & \xi \leq S_0^{-1}(\omega). \end{cases} \tag{4.14}$$

Step 2: an invariant box, and globality. We claim that $\{\|x\|_\infty \leq \Lambda\}$ is positively invariant; we argue on the larger box $\{\|x\|_\infty \leq \Lambda+1\}$, where (4.6) remains usable. Suppose the solution reaches $\|x(t_1)\|_\infty = \Lambda+1$ at some first time t_1, and let j be an index with $x_j(t_1) = \Lambda+1$, say, the other case being symmetric. Since $x_j(0) \leq \rho \leq \Lambda$, the set of times before t_1 with $x_j = \Lambda$ is nonempty and closed; let t_0 be its maximum, so that $x_j > \Lambda$ on $(t_0, t_1]$. On $[t_0, t_1]$ we have $\|x\|_\infty \leq \Lambda+1$, so (4.6) gives $|g_j| \leq a\sqrt{n}(\Lambda+1)+c_g$, and $x_j \geq \Lambda \geq \bar{R} \geq S_0^{-1}(1-\omega)$ there, so (4.9) and (4.14) give $\dot{x}_j \;\leq\; a\sqrt{n}\,(\Lambda+1)+c_g-kc_\omega \;<\; 0$ almost everywhere on $[t_0, t_1]$ by (4.13), contradicting $x_j(t_1) > x_j(t_0)$. The solution therefore stays in the box, hence is global, and

$$\|g(x(t), u(t), t)\| \;\leq\; a\sqrt{n}\,\Lambda + c_g, \qquad \forall t \geq 0. \tag{4.15}$$

Step 3: Ω_ω is reached before T_1 and is invariant. By (4.14) and (4.15), whenever $x_j(t) \geq S_0^{-1}(1-\omega)$ one has $\dot{x}_j \leq -(kc_\omega - a\sqrt{n}\Lambda - c_g) < 0$, and symmetrically $\dot{x}_j > 0$ whenever $x_j(t) \leq S_0^{-1}(\omega)$. The interval $[S_0^{-1}(\omega), S_0^{-1}(1-\omega)]$ is therefore positively invariant for x_j, and since $|x_j(0)| \leq \Lambda$ and $|S_0^{-1}(\sigma)| \leq \Lambda$ for $\sigma \in [0,1]$, it is reached after a time at most $\frac{2\Lambda}{kc_\omega - a\sqrt{n}\,\Lambda - c_g}$, which tends to 0 as $k \to \infty$ and is at most T_1 by the first term of (4.12). Hence $x(t) \in \Omega_\omega$ for every $t \geq T_1$, the first assertion of (4.8).

Step 4: tracking. For $t \geq T_1$ we have $S_0(x_j(t)) \in [\omega, 1-\omega]$, so $S_0^{-1}(S_0(x_j)) = x_j$ and (4.9) reads $\dot{x}_j = g_j - k(x_j - \gamma_j)$. With $e_j := x_j - \gamma_j$ and $\|\dot{\gamma}\| \leq L$, (4.15) and (4.10) give

$$D^+|e_j| \;\leq\; \kappa - k|e_j| \qquad \text{almost everywhere on } [T_1, +\infty), \tag{4.16}$$

where D^+ is the upper right Dini derivative, $|e_j|$ being absolutely continuous but not differentiable at its zeros. Hence, by Grönwall's inequality and $|e_j(T_1)| \leq \bar{R} - \underline{R}$,

$$|e_j(t)| \;\leq\; \big(\bar{R} - \underline{R}\big)e^{-k(t-T_1)} + \frac{\kappa}{k}, \qquad t \geq T_1. \tag{4.17}$$

The constant κ does not depend on k, so both terms are at most $\varepsilon/(2\sqrt{n})$ for every $t \geq T_2$ as soon as

$$k \;>\; \max\left\{\frac{2\sqrt{n}\,\kappa}{\varepsilon},\; \frac{1}{T_2 - T_1}\ln\frac{2\sqrt{n}\,(\bar{R} - \underline{R})}{\varepsilon}\right\}, \tag{4.18}$$

whose right-hand side collects the two remaining terms of (4.12). Then $\|x(t) - \gamma(t)\| \leq \sqrt{n}\max_j |e_j(t)| \leq \varepsilon$ for $t \geq T_2$. ◻

The three terms of (4.12), all computed from the data declared known, correspond respectively to attraction and invariance of Ω_ω before T_1, to the offset bound, and to the decay of the initial error before T_2.

The reference is required to stay in the deeper box $\Omega_{2\omega}$, while the feedback is built with the deadband ω. The gap makes the proportional term nonzero at the edge of Ω_ω, hence makes that box invariant. With a single deadband the term vanishes when γ sits exactly on the edge, and nothing pushes the state back.

Proof of Theorem 2.15. Let $\mathcal{K} \subset \mathring{\Omega}$ be a compact set containing the values of γ and let $L > 0$ bound $\|\dot{\gamma}\|$. Since $\mathcal{K}$ is compact and contained in the open box $\mathring{\Omega}$, there is $\zeta > 0$ with $\mathcal{K} \subset [\underline{R} + \zeta, \bar{R} - \zeta]^n$, and then, S_0^{-1} being continuous on $[0,1]$ with $S_0^{-1}(0) = \underline{R}$ and $S_0^{-1}(1) = \bar{R}$, there is some $\omega \in (0, 1/4)$ with $\mathcal{K} \subset \Omega_{2\omega}$, so that $\gamma \in C^{1,L}(\mathbb{R}, \Omega_{2\omega})$. The system (2.12) is of the form (4.7) with $g(x,u,t) = Ax + \phi(x,u,t)$, which is continuous and satisfies (4.6) with $a = \|A\| > 0$ and $c_g = \bar{\phi} > 0$, uniformly in u and t. Lemma 4.1 then gives (4.8), and $\Omega_\omega \subset \mathring{\Omega}$ turns its first assertion into that of (2.18). The deadband ω and the gain threshold $k^\star$ of (4.12) are the explicit output of this construction, computed from upper bounds on $\|A\|$ and on $\bar{\phi}$. ◻

The construction used one property of the output, namely that Hy be a component-wise saturation of the state. Outside the class (2.16) the scalar law (4.1) has no component-wise meaning, the unsaturated region is no longer a box, and its faces are not aligned with the coordinates, so the recovery of Section 4.2 does not carry over. The steering results of Section 3 hold for general coradiant regions; the orthant geometry enters at this last stage only.

4.2. Recovering the state from the output

It remains to invert the output map on the unsaturated region. Two inversions are involved. That of W is linear and valid at all times; that of the saturation is nonlinear and valid only where the state is unsaturated.

Proof of Corollary 2.16. By (2.16) the signal $Hy(t) = S_1(x(t))$ is available at every time and whatever the state. Next, S_0 is continuous and strictly increasing on $[\underline{R}, \bar{R}]$ with $S_0(\underline{R}) = 0$ and $S_0(\bar{R}) = 1$, hence an increasing bijection from $[\underline{R}, \bar{R}]$ onto $[0,1]$, whose inverse is continuous by compactness. Applying this coordinate by coordinate, the restriction of S_1 to Ω is a homeomorphism onto $[0,1]^n$, with inverse $(\sigma_1, \ldots, \sigma_n) \mapsto (S_0^{-1}(\sigma_1), \ldots, S_0^{-1}(\sigma_n))$. By Theorem 2.15 the state lies in $\mathring{\Omega}$ for $t \geq T_1$, and (2.19) follows by applying that inverse to $Hy(t) = S_1(x(t))$. ◻

Remark 4.2 (Conditioning of the recovery). Assume S_0 is C^1 with $S_0' > 0$ on $(\underline{R}, \bar{R})$ and set $m_\omega := \min\{S_0'(\xi) : S_0(\xi) \in [\omega/2, 1 - \omega/2]\} > 0$, where ω is the deadband produced by the proof of Theorem 2.15. Suppose the output is measured as $y + \nu$ with $\|H\nu\|_\infty \leq \min\{\bar{\nu}, \omega/2\}$. For $t \geq T_1$ the state lies in Ω_ω, so the coordinates of $H(y+\nu)$ still lie in $[\omega/2, 1-\omega/2]$, on which S_0^{-1} is m_ω^{-1}-Lipschitz, and the recovered state $\hat{x}_j := S_0^{-1}\big((H(y+\nu))_j\big)$ obeys

$$\|\hat{x}(t) - x(t)\|_\infty \ \leq \ \frac{\bar{\nu}}{m_\omega}, \qquad t \geq T_1. \tag{4.19}$$

For a saturation which is C^1 across its levels, S_0' vanishes there and $m_\omega \to 0$ as $\omega \to 0$, so the recovery is ill-conditioned near saturation. This is why the feedback is designed to keep the state at a positive distance from the levels, invertibility of S_0 alone being of no use there. The noise is accounted for here in the recovery step alone, the trajectory being the one produced by the noiseless loop of Theorem 2.15. If the feedback is itself driven by $y + \nu$, the sign estimate (4.14) survives with c_ω replaced by $c_\omega - \|H\nu\|_\infty / m_\omega$. The same conclusions then hold for a noise small in that ratio, with a correspondingly larger gain.

5. Numerical simulation

We illustrate the identification of W on a two-dimensional example, the smallest one in which the mechanism is visible. Two regions must be visited, and the excitation design is read directly on the phase plane. We take $n = q = 2$, $m = 1$ and consider

$$\begin{cases} \dot{x} = -\lambda x + CS(x) + w_k(t) + Bu(t), \\ y = WS(x) + \nu_k(t), \end{cases} \tag{5.1}$$

with $B = \begin{bmatrix} 1 & -1 \end{bmatrix}^\top$, $\lambda = 1.2$, and C, W unknown matrices whose entries are drawn independently and uniformly in $[-1, 1]$. The saturation is piecewise linear with levels $\underline{R} = 0$ and $\bar{R} = 1.5$,

$$S_0(\xi) = \operatorname{clamp}\left(\frac{\xi}{1.5}, 0, 1\right) = \begin{cases} 0 & \xi \leq 0, \\ \dfrac{\xi}{1.5} & 0 < \xi < 1.5, \\ 1 & \xi \geq 1.5, \end{cases} \tag{5.2}$$

applied component-wise. The system is integrated by an Euler scheme with step $T_s = 10^{-3}$, and w_k and ν_k are independent zero-mean Gaussian variables drawn afresh at each step, with standard deviations $\sigma_w = 0.5$ and $\sigma_\nu = 0.3$. Averaging the output over a window of length τ suppresses the output noise by a factor of order $(T_s/\tau)^{1/2}$, which puts the errors reported below at the scale $\sigma_\nu (T_s/\tau)^{1/2} \simeq 3 \times 10^{-3}$. In the notation of Subsection 2.2 this is (2.12) with $A_0 = -\lambda\,\mathrm{Id}$, $A_1 = 0$ and $\phi = CS(x) + w_k$. The coupling term is bounded, since $\|S\|_\infty = 1$, but the Gaussian noise w_k is not, so the simulation probes the method slightly outside the class covered by the theorems, which assume a uniform bound $\bar{\phi}$ on the whole of ϕ.

The image of B is the line spanned by $(1, -1)^\top$, which meets the two orthants of pattern $(+, -)$ and $(-, +)$. The corresponding saturation values $v_1 = (1, 0)^\top$ and $v_2 = (0, 1)^\top$ are linearly independent, so Assumption 2.8 is satisfied with the two sets (2.14) $K_1 = [1.5, +\infty) \times (-\infty, -1.5]$ and $K_2 = (-\infty, -1.5] \times [1.5, +\infty)$, which correspond to $R = \max\{|\underline{R}|, |\bar{R}|\} = 1.5$. Since $A_1 = 0$, case *ii.* of Theorem 2.12 applies and the design may be realized on an arbitrarily short horizon. We take long windows nonetheless, $\tau = 10$, so as to display the averaging effect of (2.6) on the output noise. With $t_1 = \tau$ and $t_2 = t_1 + 2\tau$, the piecewise constant control is $u = \mu_1 \mathbf{1}_{[0, t_1 + \tau]} - \mu_2 \mathbf{1}_{[t_1 + \tau, t_2 + \tau]}$ with $\mu_1 = 10$ and $\mu_2 = 12$, which is the control of Lemma 3.10 for this design. The associated equilibria are $\mu_1 (1, -1)^\top / \lambda \simeq (8.3, -8.3)$ and $\mu_2 (-1, 1)^\top / \lambda \simeq (-10, 10)$, both well inside the respective regions. Figure 5.1 shows the resulting trajectory in the phase plane, with the two segments on which $S(x(t)) = v_i$ highlighted. These are the segments the estimator integrates over.

The matrix W is then estimated by the estimator $E(y; \mathcal{E})$ of (2.6) with $\mathcal{E} = \{(v_1, [t_1, t_1 + \tau]), (v_2, [t_2, t_2 + \tau])\}$, the integrals being evaluated by the trapezoidal rule on the sampled output. Over 100 runs, with C, W and the noises redrawn each time, the error $\|E(y; \mathcal{E}) - W\|$ had mean 9×10^{-3}, within a factor of three of the predicted scale 3×10^{-3}. Three effects explain the gap. During the saturated phases $S(x(t)) = v_i$ regardless of small perturbations of x, so the estimator is insensitive to the state noise w_k; the output noise ν_k has zero mean, so the time averaging in (2.6) acts as a low-pass filter and reduces its contribution by a factor of order $\tau^{-1/2}$; and the inversion by $\Gamma(\mathcal{E})^{-1}$ in (2.6) is well conditioned, the two values v_1 and v_2 being orthogonal.

The simulation covers the identification stage only, in the setting of case *ii.* of Theorem 2.12; the output feedback of Section 4 and the recovery of Corollary 2.16 are not simulated here.

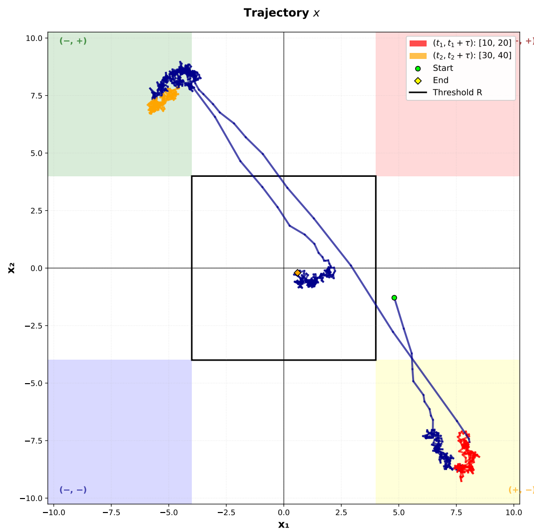

FIGURE 5.1. *Trajectory of x in the phase plane, colour-coded by saturation regime. The red segment is the interval $[t_1, t_1 + \tau]$, on which $S(x(t)) = v_1 = (1, 0)^\top$, and the yellow segment is $[t_2, t_2 + \tau]$, on which $S(x(t)) = v_2 = (0, 1)^\top$. The grey portions are the transients between them.*

6. Conclusion

We have shown that the saturation of an unmeasured state can be turned into a regressor whose values are known in advance, so that an unknown output matrix is recovered exactly by a closed-form linear regression, with no observer and no persistence of excitation. The excitation is designed, not assumed. Indeed, explicit piecewise-constant or Gramian-based controls steer every trajectory issued from a ball of initial conditions through the saturation regions in a prescribed order, under a geometric condition on the known part of the dynamics, given at three levels of knowledge of the linear part. Once the matrix is known, an output feedback keeps the state away from saturation, and the state is recovered by component-wise inversion.

The identification stage assumes the linear part of the dynamics to be known, or known up to a perturbation small enough to satisfy an explicit bound. The recovery stage is more demanding, requiring a square invertible input matrix and at least as many independent outputs as states. The robustness of the identification to measurement noise is established only empirically, through the simulation of Section 5; the only noise bound proved is the conditioning estimate of the recovery step. Finally, the feedback gain is computed from upper bounds on the norm of the linear part and on the perturbation term, which must therefore be available.

REFERENCES

[1] S.-I. AMARI, *Dynamics of pattern formation in lateral-inhibition type neural fields*, Biological cybernetics, 27 (1977), pp. 77–87.

[2] S. ARANOVSKIY, A. BOBTSOV, R. ORTEGA, AND A. PYRKIN, *Performance enhancement of parameter estimators via dynamic regressor extension and mixing procedure*, IEEE Transactions on Automatic Control, 62 (2017), pp. 3546–3550.

[3] G. BESANÇON, *Nonlinear observers and applications*, vol. 363, Springer, 2007.

[4] G. CHOWDHARY AND E. JOHNSON, *Concurrent learning for convergence in adaptive control without persistency of excitation*, in 49th IEEE Conference on Decision and Control (CDC),

IEEE, 2010, pp. 3674–3679.

[5] J.-M. Coron, *Control and nonlinearity*, no. 136, American Mathematical Soc., 2007.

[6] G. Kreisselmeier and G. Rietze-Augst, *Richness and excitation on an interval—with application to continuous-time adaptive control*, IEEE Transactions on Automatic Control, 35 (1990), pp. 165–171.

[7] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, 2nd ed., 1999.

[8] S. B. Roy, S. Bhasin, and I. N. Kar, *Combined mrac for unknown mimo lti systems with parameter convergence*, IEEE Transactions on Automatic Control, 63 (2018), pp. 283–290.

[9] E. Trélat, *Contrôle optimal: théorie & applications*, vol. 36, Vuibert Paris, 2005.

[10] H. Wilson and J. Cowan, *A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue*, Kybernetik, 13 (1973), pp. 55–80.

[11] H. R. Wilson and J. D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophysical Journal, 12 (1972), pp. 1–24, https://doi.org/10.1016/S0006-3495(72)86068-5, https://www.cell.com/biophysj/fulltext/S0006-3495(72)86068-5.