Persistent Identification and Adaptation: Stabilization of Slowly Varying Systems in $H^\infty$

Le Yi Wang and Lin Lin, Member, IEEE

Abstract—In this paper, the problem of persistent identification and adaptive stabilization of time-varying systems is studied within the framework of deterministic worst case identification and slow $H^\infty$ adaptation. The plants under consideration are unstable and time-varying and cannot be stabilized by a fixed robust controller. Starting from an initial well-designed operating point, the controller must persistently adapt to the time-varying plant to maintain uniform stability over all future time.

A key property which guarantees uniform stability is that the identification-adaptation iteration satisfies a certain invariance principle. We demonstrate that the adaptive design using periodic external inputs, least squares identification, and slow $H^\infty$ adaptation possesses such an invariance property, leading to a successful adaptive stabilization methodology. Generic nature of our findings are discussed.

Index Terms—Adaptation, $H^\infty$ control, identification, stability, time-varying systems.

I. INTRODUCTION

THIS PAPER is concerned with the problem of persistent identification and adaptive stabilization of single-input/single-output (SISO) discrete-time time-varying systems. The plants under consideration are unstable and time-varying and cannot be stabilized by a fixed robust controller. As a result, adaptation is inevitable. Starting from an initial well-designed operating point, the controller must persistently adapt to the time-varying plant to maintain uniform stability over all future time. Some challenging issues must be resolved: How can one design an identification algorithm which can persistently identify time-varying plants with uniformly small error bounds? How should the interaction between identification and adaptation be resolved? In this paper, these issues are investigated for slowly time-varying systems which are, perhaps, the simplest and most tractable time-varying systems. A design methodology is developed which employs the ideas of persistent probing signals, uncertainty-set identification, and slow $H^\infty$ adaptation.

Traditionally, research on adaptive control has been concentrated on parametric systems. Many useful adaptation algorithms have been developed [7]. Rigorous analysis of stability and convergence of these algorithms have been pursued since the 1970’s [14]. Main results obtained in this direction address the issues of boundedness of signals and convergence of parameter estimates when the underlying uncertain plants are time invariant. Extensions to time-varying systems have also been reported [18].

The work reported in this paper differs from the traditional problem formulation and methodologies mainly in three aspects. First, plants are time-varying (versus time-invariant), contain norm-bounded unmodeled dynamics (versus finite-dimensional systems), and cannot be stabilized by a fixed controller (versus the assumption that there exists, albeit unknown, a time-invariant stabilizing controller for the system). Second, we focus on the issue of adaptation to the time-varying aspects of the plant, rather than the initial convergence to a satisfactory controller. This is clearly reflected in our assumption that at the starting time $t_0$ the designed frozen-time controller can stabilize the frozen-time plant. The main challenge becomes: can the controller appropriately adapt itself when the plant varies beyond the robust regions of the initial controller? Finally, we impose a much stronger performance criterion, namely, uniform stability. This requires that controllers be adapted very carefully such that desired performance properties hold all the time after $t_0$. Together, these aspects impose a great challenge on identification and adaptation which must generate plant models sufficiently accurate for control design and maintain performance levels persistently over all future time. We will use the term persistent identification and adaptation for such problems.

While these aspects are relatively unusual and new to the traditional approaches of adaptive control, our approaches employ many ideas and methods of classical adaptive control. For instance, our design can be easily characterized as an indirect adaptive control using certainty equivalence principles, although in the literature of time-varying systems our approaches have been called frozen-time design since the 1960’s. The rank conditions on input signals can be clearly recognized as a type of persistent excitation (PE) conditions, although the classical PE conditions often take very different forms depending on system parameterizations and are used to guarantee asymptotic convergence of parameter estimates. While our estimates may not converge, they must generate sufficiently accurate uncertainty sets containing the plant, persistently over all time, so that robust control can perform. As a result, the conditions we impose take somehow different forms from classical PE conditions.

The main assumptions we make on the time-varying plants are the following.
1) Time-varying plants are expressed in a factorization form $P = ND^{-1}$, in which $N$ and $D$ are frozen-time coprime in the rings of stable systems.

2) $D$ is known and time invariant.

3) $N$ contains (possibly infinite-dimensional) unmodeled dynamics and is slowly time-varying over a large range such that there exists no fixed robust controller to stabilize $P$.

4) The external disturbances $d$ are uniformly bounded by a constant $\varepsilon_d$.

Within these assumptions, the coprimeness is mandatory in our approach since our design relies on robust control to provide a nonzero robustness region for adaptation. It should be noted that coprimeness here is expressed in the rings of stable systems, rather than polynomials. Hence, common stable parts in $N$ and $D$ are allowed. The second assumption is nonessential. The inclusion of this condition is for the simplicity and clarity of our development. To demonstrate its ready extension to the case where $D$ is unknown and slowly time-varying, we cite the recent results on closed-loop persistent identification [32]. The third condition ensures that our problems cannot be reduced trivially to the case of stable system identification by first applying a fixed stabilizing feedback to the plant. The fourth condition follows the paradigm of deterministic control-oriented identification in which a bounded uncertainty set is sought for robust control. The boundedness of the uncertainty set becomes impossible when $d$ is not bounded. A direct consequence is the potential conservativeness of our results: We do not allow occasional bursts of disturbances, no matter how rarely they may occur. The classical stochastic setting may offer advantages in this aspect. The pros and cons of worst case identification versus averaging identification have been vigorously debated and are far from being resolved. Our approach will carry this birthmark until a better framework is developed to incorporate worst case and stochastic methodologies.

It should be emphasized that this paper does not seek weakest conditions on the plant for the given problem. In fact, the assumptions we make on the plant are quite strong. Our main intention is to understand the nature and essential features of persistent identification and adaptation in a robust control framework. As it turns out, these features can be summarized in two invariance principles.

1) **Feedback Invariance Property (FIP)** for probing signals: The feedback system (see Fig. 1) which maps $r$ to $u$ is control-dependent and alters signals significantly. As a result, closed-loop identification becomes very difficult. FIP characterizes the features of signals which are invariant, at least approximately, under the feedback mapping. Periodicity and ranks are such properties. A pleasant surprise arises as these properties are shown to characterize desirable probing signals for persistent identification as well.

2) **Invariance Principle (IP)** for identification-adaptation iteration: To maintain uniform performance, identification errors and performance levels should not increase beyond the uniform limits in any steps of the iteration. This can be viewed as a condition of contraction or invariance on the identification-adaptation iteration.

While this paper addresses these issues in a concrete problem, they seem to be generic in persistent adaptation problems.

Despite a long history of research on adaptive control, some initial (and essential) objectives of adaptation are still not met with rigorous analysis. These objectives are: 1) adaptation is mainly useful for time-varying environment; 2) adaptation is useful only if robust (nonadaptive) control cannot achieve required performance; (3) adaptation is useful only if it can offer better performance. Our work here is a preliminary effort in establishing a framework in which these objectives can be directly and rigorously addressed.

The paper is organized as follows. After introducing necessary mathematics definitions in Section II, we describe the scenario and key issues under study in Section III. The main problems are formulated in Section IV. General approaches are summarized.

Features of probing signals which are desirable for closed-loop persistent identification are sought in Section V. We argue that desirable probing signals should possess features which guarantee accurate identification, on one hand, and are feedback invariant, on the other. We show that full-rank periodic signals have such features and hence are suitable for persistent identification and adaptation of time-varying systems. Section VI studies the problem of system identification of slowly time-varying systems. Explicit error bounds on identification errors are established, which are related to metric complexities of prior uncertainty sets, variation rates of the plants, and probing capability of input signals. The key property of these identification mappings, which is of essential importance for applications in adaptive systems, is expressed in the contraction property (Theorem 2) which establishes a sequence of vanishing invariant neighborhoods under the identification mappings.

The main results on adaptation are presented in Section VII (Theorem 3) in which an adaptive control scheme is developed, employing the results of Sections V and VI as well as the slow $H^\infty$ adaptation introduced in [41]. It is shown that when variation rates of the plants are small, the design procedure produces an adaptive control which achieves persistent bounds on posterior uncertainty sets and stabilizes the time-varying uncertain plants. Since there exists no (nonadaptive) robust control for the time-varying plants, adaptive control is not only a superior choice, but inevitable indeed.
Finally, limitations of our approaches and open research issues along the direction of this paper are summarized in Section VIII.

**Related Early Work:** This paper employs ideas and early findings in deterministic worst case identification, $H^\infty$ adaptation, and classical adaptive control. Hence, it is no surprise that this paper is related to a vast array of early results in several fields.

This paper is a continuation of some early work on identification and frozen-time design of time-varying systems. The measures of persistent identification performance were introduced in [31] for linear time-invariant (LTI) and slowly time-varying stable systems in impulse response models. Further back, these measures are essentially special cases of shift-invariant identification $n$-width introduced in [40]. Extensions to unstable systems were reported recently in [32]. A double algebra framework was introduced in [41] and [35] for slow $H^\infty$ adaptation.

The problem of worst case identification is now a very active research area. The concepts of $\varepsilon$-net and $\varepsilon$-dimension in the Kolmogorov sense [11] were first introduced into the control field by Zames [37] in studies of model complexity and system identification. Complexity issues were investigated by Tse et al. [28], [6], Poolla and Tikku [25], [26], and also [37] and [34]. Milanese is one of the first researchers in recognizing the importance of worst case identification. He and coworkers introduced the problem of set-membership identification and produced many interesting results on the subject [20], [19], [21]. Within the framework of $H^\infty$-frequency domain identification introduced in [10], many efficient algorithms have been developed, including Bai and Raman [1], Chen et al. [3], Gu et al. [8], [9], and Makila et al. [16], [17].

To enhance the feasibility of worst case identification in adaptive applications, major efforts have been made to develop algorithms using time-domain data. Numerous results have been reported, including [43], [22], [12], and [4]. Furthermore, to relate identification to robust control in a direct manner, the problems of closed-loop identification and interaction between identification and control have also been pursued [2], [16].

The adaptation under study follows closely the fundamental philosophy of Zames in developing an information-based theory of identification and adaptation [38], [39], [42]. There is a large treasury of literature on classical identification and adaptive control. The reader is referred to [7] and [14] and the references therein for classical adaptation schemes for deterministic and stochastic systems.

An early version of this paper appeared in [33].

**II. MATHEMATICS PRELIMINARIES**

**A. Basic Notation**

The real numbers, complex numbers, and integers are denoted by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}$, respectively. For $x \in \mathbb{C}$, $x^*$ is its complex conjugate and $|x|$ its absolute value.

Several vector and matrix norms are used in this paper. For a vector $v = [v(1), \cdots, v(n)]^T \in \mathbb{C}^n$, we define

$$
||v||_2 = \left( \sum_{i=1}^{n} |v(i)|^2 \right)^{1/2}, \quad ||v||_\infty = \max_{i=1,\cdots,n} |v(i)|
$$

and for $\sigma > 1$

$$
||v||_\sigma = \sum_{i=1}^{n} \sigma^{-1} |v(i)|.
$$

For a matrix $M = [a_{ij}] \in \mathbb{C}^{k \times k}$, $M^T$ will denote its transpose, $\mathcal{S}(M)$ and $\mathcal{S}_0(M)$ its largest and smallest singular values. The following inequalities are standard. For $v \in \mathbb{C}^n$, $M \in \mathbb{C}^{m \times n}$ and $\sigma > 1$

$$
||v||_\sigma \leq \sqrt{\frac{\sigma^{2n} - 1}{\sigma^2 - 1}} ||v||_2, \quad ||v||_2 \leq ||v||_1 \leq \sqrt{n} ||v||_2
$$

$$
||v||_2 \leq \sqrt{n} ||v||_\infty, \quad ||M||_2 \leq \mathcal{S}(M)||v||_2.
$$

We shall denote throughout this paper

$$
a_\sigma(n) = \sqrt{\frac{\sigma^{2n} - 1}{\sigma^2 - 1}}.
$$

For $\sigma \geq 1$, $l_\sigma^1$ denotes the normed spaces of sequences $\{u(t) \in \mathbb{R}, t = 0, 1, \cdots\}$, for which

$$
||u||_\sigma := \sum_{t=0}^{\infty} \sigma^t |u(t)| < \infty
$$

and $l^\infty$ is the space of sequences $u$ with

$$
||u||_\infty := \sup_{0 \leq t < \infty} |u(t)| < \infty.
$$

Sequences in $l_\sigma^1$ are upper bounded by an exponential decaying envelope. For $\sigma = 1$, $l_1^1$ becomes the Lebesgue space $l^1$, and $||\cdot||_1$ is the usual $l^1$ norm.

$T_{[a,b]}$ denotes the truncation operator: $(T_{[a,b]}u)(t) = u(t)$ for $a \leq t \leq b$ and zero otherwise.

In this paper, we often require a system to be exponentially stable. It is well known that exponentially stable systems have transfer functions analytic on the disk of radius larger than one. Hence, we introduce the following notation: $H_{\sigma}^\infty$ is the Hardy space of analytic functions $K$ on the open disk of radius $\sigma$, with norm

$$
||K||_{\sigma} := \sup_{\theta \in [-\pi, \pi]} |K(\sigma e^{i\theta})| < \infty.
$$

For $\sigma = 1$, $H_{\sigma}^\infty$ is reduced to the usual $H^\infty$ space.
B. Systems

In this paper, stable systems \( \mathcal{B} \) will consist of single-input/single-output (SISO) linear time-varying causal bounded discrete-time systems \( K \) on \( F^\infty \) with convolution representations

\[
(Ku)(t) = \sum_{\tau = -\infty}^{t} k(t, \tau)u(\tau), \quad t \in \mathbb{Z}
\]  

(4)

where the kernel \( k(\cdot, \cdot) \) of \( K \) satisfies \( k(t, \tau) = 0, \ t < \tau, \) and \( \sup_{\tau} \|k(t, \cdot)\|_{1} < \infty. \) The norm of \( K \in \mathcal{B} \) is defined by

\[
\|K\|_{1} = \sup_{t} \|k(t, \cdot)\|_{1}.
\]

Unstable systems will belong to \( \mathcal{B}_e \), the extended space of \( \mathcal{B}. \)

The subspace of time invariant systems in \( \mathcal{B} \) will be denoted by \( \mathcal{B}_0. \) If \( K \in \mathcal{B}_0, \) its kernel (impulse response) \( k \in l^2. \)

For \( K \in \mathcal{B}_0 \) with impulse response \( k, \) we shall adopt the mathematics notation (rather than the engineering notation of \( z^{-1} \) form) of \( z \)-transforms

\[
K(z) = \sum_{t = 0}^{\infty} k(t)z^{-t}
\]

for \( z \) in the region of convergence.

\( S \) denotes the right shift operator on \( F^\infty \)

\[
(Sx)(t) = x(t-1).
\]

The frozen-time system of \( K \) at \( \tau \in \mathbb{Z} \) is defined as the (time-invariant) system \( K_{\tau} \) with representation

\[
(K_{\tau}u)(t) := \sum_{\theta = -\infty}^{t} k(\tau, \tau - (t - \theta))u(\theta), \quad t \in \mathbb{Z}.
\]

The kernel of \( K_{\tau} \) will be denoted by \( k_{\tau}, \) and \( k_{\tau}(\theta) = k(\tau, \tau - \theta), \ \tau, \theta \in \mathbb{Z}. \) Obviously, \( k_{\tau}(\tau) \in l^1. \) Hence, \( K_{\tau} \in H^{\infty}. \)

Several norms are frequently used in this paper. For \( K \in \mathcal{B} \) with kernel \( k(\cdot, \cdot), \) we denote, for \( \sigma \geq 1, \) the time-domain norm

\[
\|K\|_{\sigma} := \sup_{t} \|k_{\tau}\|_{\sigma}
\]

and the frequency-domain norm

\[
\|K\|_{(\sigma)} = \sup_{t} \|K_{\tau}\|_{(\sigma)}.
\]

Since for \( \sigma > 1 \)

\[
\frac{\sqrt{\sigma^2 - 1}}{\sigma} \|k_{\tau}\|_{1} \leq \|K_{\tau}\|_{(\sigma)} \leq \|k_{\tau}\|_{\sigma}
\]

we have

\[
\frac{\sqrt{\sigma^2 - 1}}{\sigma} \|K\|_{1} \leq \|K\|_{(\sigma)} \leq \|K\|_{\sigma}.
\]

(5)

A. Feedback Configuration

Consider the feedback system in Fig. 1. \( r, u, y, \) and \( d \) are the external input, plant input, plant output, and disturbance, respectively. The interconnection of a feedback \( F \) and a plant \( P \) in \( \mathcal{B}_e \) is well-posed if all elements of the closed-loop system

\[
K(P, F) := \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} := \begin{pmatrix} (1 + FP)^{-1} & P(1 + FP)^{-1} \\ F(1 + FP)^{-1} & (1 + PF)^{-1} \end{pmatrix}
\]

are in \( \mathcal{B}_e \) and stable if they are in \( \mathcal{B}. \) \( F \in \mathcal{B}_e \) is said to robustly stabilize a subset \( \Omega \subseteq \mathcal{B}_e \) of plants if \( K(P, F) \) is stable for all \( P \in \Omega. \)

B. Main Scenarios and Issues

In adaptive control of time-varying systems, identification must be performed in closed loop. Since both identification and adaptation rely on the input \( u \) to fulfill their individual needs, they interact intimately and often impose competing requirements on the input, leading to nonlinearity and complications in system analysis and design.

Our problems will be formulated around Scenario 1.

Scenario 1: Identification and Adaptation of Time-Varying Systems: Suppose that the plant \( P \) is time-varying. At each time \( t, \) the frozen-time system \( P_{\tau} \) of \( P \) is identified by using input–output observations near \( t. \) The observations result in an uncertainty set \( \Omega \) which contains \( P. \) A frozen-time controller \( F_{\tau} \) is then designed which robustly stabilizes \( \Omega \) and achieves certain desired performances. This identification and controller adaptation procedure is repeated at each discrete time instant.

This scenario, which is nothing but indirect adaptation with certainty equivalence principles, is characterized by several features: 1) The input \( u \) to the plant is generated by the feedback loop and cannot be manipulated directly. Only the external input \( r \in F^\infty \) can be selected at will. 2) Since the plant is time-varying, only a small window of observations near \( t \) can provide useful information about the frozen-time plant \( P_{\tau}, \) even if all past input and output measurements are available. 3) One input \( u \) must provide sufficient probing capability for all possible observation windows. 4) The controller \( F_{\tau} \) must robustly stabilize \( P_{\tau} \) for frozen-time performance, must change smoothly to ensure a tangible relationship between frozen-time performance and that of the time-varying system, and must not produce signals detrimental to identification.
Several fundamental issues arise in this scenario.
1) How can one characterize classes of plant inputs \( u \) which can guarantee accurate identification for all frozen-time systems \( H \)?
2) Can such plant inputs be generated by external inputs \( r \) via the feedback loop?
3) How should controllers be designed?
4) How should the interaction and conflicts between identification and adaptation be resolved?

IV. PROBLEM FORMULATION AND GENERAL APPROACHES

A. Plants
A pair \((N, D) \in \mathcal{B} \times \mathcal{B}\) is coprime in \( \mathcal{B} \) if for some \((X, Y) \in \mathcal{B} \times \mathcal{B}\)

\[
(XN + YD)^{-1} \in \mathcal{B}.
\]

A system \( G \in \mathcal{B} \) has a left (or right) factorization representation in \( \mathcal{B} \) if \( G = D^{-1}N \) (or \( G = ND^{-1} \)), where \((N, D) \in \mathcal{B} \times \mathcal{B}\) and \( D^{-1}\) is well defined. The factorization \( D^{-1}N \) is coprime if the corresponding pair \((N, D)\) is coprime in \( \mathcal{B} \).

Plants and controllers considered in this paper are represented in the feedback system by

\[
y = ND^{-1}u + d
Yu = -Xy + r
\]

where \( ND^{-1} = P \) is the representation of the plant and \( Y^{-1}X = F \) the controller. \( r, d, u, y \) are the external input, disturbance, plant input and output, respectively.

B. Basic Assumptions
Consider the feedback system in (9). The plant \( P \) admits a right factorization representation in \( \mathcal{B} \)

\[
P = ND^{-1}.
\]

Assumption 1: For some \( \sigma > 1 \)
1) \( D \) is time invariant, known, and \( D \in H_\sigma^\infty \). Without loss of generality, assume \( D \) is inner in \( H_\sigma^\infty \), namely,

\[
|D(\sigma e^{i\theta})| = 1, \quad \theta \in [-\pi, \pi].
\]

Denote \( |D|_1 = \kappa_D \).
2) \( N \in \mathcal{B} \) is exponentially stable, \( ||N||_\sigma < \infty \), and is strictly causal. To be explicit, we sometimes write \( N = N_0S \). Let \( ||N||_1 \leq \kappa_N \).
3) A priori information about \( N \) is that the frozen-time systems \( N_t \) are decomposed into

\[
N_t = N_t^m + N_t^u
\]

where \( N_t^m = N_t(0) + N_t(1)z + \cdots + N_t(n-1)z^{n-1} \) is the modeled part of order \( n \), and \( N_t^u = N_t(n)z^n + \cdots \) is the unmodeled dynamics. We do not impose conditions on \( N_t^m \) whose coefficients will be estimated by identification. The unmodeled dynamics is norm-bounded by

\[
||N_t^u||_\sigma \leq \varepsilon_n
\]

but otherwise arbitrary, \( \varepsilon_n \) is assumed to be an exponentially decaying function of \( n \): \( \varepsilon_n \leq d_2\sigma^{-n} \).
4) \( N \) is slowly time-varying with rate \( \gamma \)

\[
d(\sigma)(N) = \sup_t ||N_t - N_{t-1}||_\sigma \leq \gamma.
\]

5) For any \( t, (N_t, D) \) are coprime in \( H_\sigma^\infty \).
6) The disturbance \( d \) is bounded with

\[
||d||_\infty \leq \varepsilon_d.
\]

Examples of systems satisfying Assumption 1 include systems with unstructured additive or multiplicative uncertainties:

\[
P = (N_0 + W\Delta)D_0^{-1} \quad \text{or} \quad P = N_0(I + W\Delta)D_0^{-1}
\]

where \( N_0, D_0, W \) are LTI and known, but the uncertainty part \( \Delta \) is unstructured and slowly varying.

It is noted that when a priori uncertainty on the plant is large, there exists no (nonadaptive) robust control for this class of systems. Consider, for instance, an uncertain time-varying plant with a priori information \( P = (N_0 + W\Delta_t)D_0^{-1} \)

\[
||\Delta_t||_\sigma \leq b.
\]

Let \( X_0, Y_0 \in H_\sigma^\infty \) be solutions to the Bezout equation

\[
X_0N_0 + Y_0D_0 = 1.
\]

Then, it is standard that \( P = (N_0 + W\Delta_t)D_0^{-1} \) is robustly stabilizable in \( H_\sigma^\infty \) if and only if

\[
b < \left( \inf_{Q \in H_\sigma^\infty} ||W(X_0 - DQ)||_\sigma \right)^{-1}.
\]

Consequently, if

\[
b > \left( \inf_{Q \in H_\sigma^\infty} ||W(X_0 - DQ)||_\sigma \right)^{-1}
\]

\( P \) is not robustly stabilizable by a nonadaptive controller. In this case, adaptation becomes inevitable.

C. Design Specifications
Desired performance of the feedback system is characterized by the following specifications.

Design Specifications: The property \( \mathcal{P}(\tau) \) of the closed-loop system (9), defined below, must be satisfied for all \( \tau \).
1) \( ||T_{[-\infty, \tau]}y||_\infty \leq k_y \).
2) \( ||T_{(-\infty, \tau]}u||_\infty \leq k_u \).
3) The estimate \( \hat{N}_\tau \) of \( N_\tau \) must satisfy

\[
\sup_{-\infty < \tau < \tau} ||\hat{N}_\tau - N_\tau||_\sigma \leq \varepsilon.
\]

Here, 1) and 2) are stability requirements (uniform boundedness of all signals). Item 3) demands uniform estimation error bounds. The norm \( ||\cdot||_\sigma \) is used since feedback design must generate exponentially stable systems so that time-varying effects can be compensated.

The problem under study can be stated as follows.
Problem 1: Given a time-varying plant $P$ satisfying Assumption 1, design an identification algorithm $I$, an adaptation mapping $A$, and a probing signal $r \in F^n$ such that “Specification $\mathcal{P}(0)$ holds” implies “$\mathcal{P}(t)$ holds for all $t \geq 0$.”

Departing from traditional convergence analysis, Problem 1 concentrates on the ability of a feedback system in uniformly adapting to a time-varying environment (hence, the requirement “$\mathcal{P}(t)$ holds for all $t \geq 0$”) after arriving at its initial equilibrium points (hence, the condition “Specification $\mathcal{P}(0)$ holds”). Problem 1 will be solved for slowly varying plants, using the approaches of uncertainty-set identification and slow $H^\infty$ adaptation. Although we select stability for concreteness in our development, the framework can certainly be extended to other performance measures.

Here, the three main issues must be resolved: 1) selection of probing signals $r$; 2) construction of identification algorithms; and 3) construction of adaptation mappings. In the sequel, the term “an adaptive design procedure” will mean a specification of these three items.

D. General Approaches

Adaptive control is an iteration process involving the interaction of identification and adaptation at each time step $t$. More precisely, one starts at some time $t_0$ with a certain estimation error

$$e(t_0) = ||\hat{N}_0 - N_0||_\sigma.$$

For $t = t_0 + 1$, $t_0 + 2$, $\cdots$, an identification algorithm $I$ is devised which provides estimates $\hat{N}_t$ with estimation errors

$$||\hat{N}_t - N_t||_\sigma \leq e(t).$$

Then, an adaptation mapping is applied which produces robust controllers $F(t) = Y_t^{-1}X_t$ to achieve performance specifications based on the posterior uncertainty.

This iteration process leads to a sequence of identification errors

$$e(t_0) \to e(t) \to \cdots$$

and signal bounds

$$||T_{\mathcal{I},t_0}^\infty||_\infty \to ||T_{\mathcal{I},t_0}^t||_\infty \to ||T_{\mathcal{I},t}^\infty||_\infty \to \cdots.$$ 

Problem 1 can also be stated in the following invariance principle.

Problem 2 (Invariance Principle): Define an adaptive design procedure such that the intervals

$$\Omega_e = [-\varepsilon, \varepsilon], \quad \Omega_u = [-k_u, k_u], \quad \Omega_y = [-k_y, k_y]$$

are invariant for $e(t)$, $u(t)$, and $y(t)$ under the procedure, namely

$$e(\gamma) \in \Omega_e, \quad \forall \gamma \leq t \quad \Rightarrow \quad e(t+1) \in \Omega_e,$$

$$u(\gamma) \in \Omega_u, \quad \forall \gamma \leq t \quad \Rightarrow \quad u(t+1) \in \Omega_u,$$

$$y(\gamma) \in \Omega_y, \quad \forall \gamma \leq t \quad \Rightarrow \quad y(t+1) \in \Omega_y.$$

In this paper, an adaptive design procedure will be developed which consists of the following identification and adaptation iterations. The details of the procedure will be postponed until later sections.

Design Procedure: For a given real value $\varepsilon > 0$ (for estimation error bounds) and an integer $n > 0$ (for model complexity) we have the following.

1) The identification algorithm is the classical least squares estimation based on the most recent $n$ observations.
2) The adaptation mapping is the suboptimal $H^\infty$ design developed in [41].
3) The probing signal $r$ is $n$-periodic and full rank, namely, the Toeplitz matrix of symbol $r$

$$\Phi_r(0) = \begin{bmatrix} r(n-1) & \cdots & r(0) \\ \vdots & \ddots & \vdots \\ r(2n-1) & \cdots & r(n-1) \end{bmatrix}$$

is full rank. We will denote $\lambda = \text{det}(\Phi_r(0)) > 0$.

While one might quote simplicity and popularity for selecting least squares estimation, the justification here is more fundamental: for persistent identification, it is an optimal identification algorithm, under certain conditions, among all identification algorithms. This finding was first established for stable systems in [31]. In adaptive control of time-varying systems, a frozen-time designed controller must stabilize the frozen-time plant and also vary smoothly. The design in [41] provides an explicit tradeoff between optimal robustness and smoothness of $H^\infty$ controllers. Finally, periodicity and rank conditions are feedback invariant, i.e., they are preserved after the feedback mapping from $r$ to $u$. It turns out that these two properties also characterize desirable probing signals for persistent identification problems.

The main result of the paper is Theorem 3.

Theorem 3: Under Assumption 1, for any required $\varepsilon > 0$, there exist $k_u > 0$, $k_y > 0$, and $\gamma_0 > 0$ such that an integer $n$ (model complexity) can be selected under which the design procedure guarantees $\Omega_e$, $\Omega_u$, and $\Omega_y$ are invariant, provided $\gamma \leq \gamma_0$.

The rest of this paper is devoted to establishing details and properties of the design procedure and this main result.

V. PROBING SIGNALS

We start with a search of probing signals suitable for persistent identification in a closed-loop setting.

Consider a shift-invariant system $G \in B_{k_0}$ with kernel $g$. The input–output relationship of $G$ is expressed as

$$y(t) = Gx + d = \sum_{\tau=-\infty}^{t} g(t - \tau)x(\tau) + d(t)$$

where $||d||_\infty \leq \varepsilon_d$. For a given model order $n \geq 1$, we assume that the unmodeled dynamics

$$G_u(z) = g(n)z^n + \cdots$$

is bounded by $|G_u|_\sigma \leq \varepsilon_n$. Define the truncated regression vectors

$$\phi_T^T(l) = [x(t - ln), x(t - ln - 1), \cdots, x(t - [l + 1]n - 1)],$$

$$l = 0, 1, \cdots, (10)$$
and the vectors of system parameters
\[ \theta_l = \begin{bmatrix} g(ln) \\ \vdots \\ g((l+1)n-1) \end{bmatrix}, \quad l = 0,1,\ldots \]

It follows that
\[ y(t) = \sum_{l=0}^{\infty} \phi_T^l(t) \theta_l + d(t). \]

As shown in [34] and [40], for typical classes of uncertainty sets, \( \theta_0 \) is the optimal choice of models of order \( n \). A standard procedure for estimating \( \theta_0 \) proceeds as follows. First, \( n \) observations on \( y \) in \([t-n+1,\ldots,t]\) are performed and the input–output relationship is rewritten in the form of
\[ Y_n = \sum_{l=0}^{\infty} \Phi(l)\theta_l + D_n = \Phi(0)\theta_0 + \sum_{l=1}^{\infty} \Phi(l)\theta_l + D_n \]
where
\[
Y_n = \begin{bmatrix} y(t-n+1) \\ \vdots \\ y(t) \\ \vdots \\ d(t-n+1) \end{bmatrix}, \quad \Phi(l) = \begin{bmatrix} \phi_T(l) \\ \vdots \\ \phi_T(n+1) \\ \vdots \\ \phi_T(l) \end{bmatrix}, \\
D_n = \begin{bmatrix} d(t-n+1) \\ \vdots \\ d(t) \end{bmatrix}.
\]

Note that \( \Phi(l) \) are Toeplitz matrices. If the input \( x \) is selected such that \( \Phi(0) \) is invertible, \( \theta_0 \) can be represented as
\[ \theta_0 = \Phi(0)^{-1} Y_n - \Phi(0)^{-1} \left( \sum_{l=1}^{\infty} \Phi(l)\theta_l + D_n \right) = \hat{\theta}_0 - E(t) \]
\[ \hat{\theta}_0 = \Phi(0)^{-1} Y_n \]
\[ E(t) = \Phi(0)^{-1} \left( \sum_{l=1}^{\infty} \Phi(l)\theta_l + D_n \right) = \Phi(0)^{-1} \sum_{l=1}^{\infty} \Phi(l)\theta_l + \Phi(0)^{-1} D_n \]
is a term of identification error, where its dependence on the current time of observation windows is expressed explicitly. In persistent identification problems, estimation errors \( E(t) \) must be uniformly small over all \( t \). Hence, an appropriate measure of estimation errors is
\[ \sup_{t \geq 0} \| E(t) \|_\sigma. \]

To obtain bounds on the identification error which depends on the probing signal \( x \), some basic properties of periodic signals must be established, which will be discussed in the next subsection.

### A. Toeplitz Systems with Periodic Symbols

Let \( x \) be an \( n \)-periodic signal (i.e., a periodic signal of period \( n \)) and \( \Phi(l) \) the corresponding Toeplitz matrices defined in (11). Define the discrete Fourier transform of \( x \) as
\[ X(\omega) = \mathcal{F}[x] = \sum_{t=0}^{n-1} x(t) e^{-j\omega t}, \quad \omega \in [-\pi, \pi]. \]

In particular, for \( \omega = \frac{2\pi j}{n}, j = 0,\ldots,n-1 \), it is easy to verify the following shift property: for any \( l = 0,1,\ldots,n-1 \)
\[ \mathcal{F}[S^l x](\omega_j) = e^{j\omega_j l} X(\omega_j), \quad j = 0,1,\ldots,n-1. \]

The following results are fairly standard, especially (1) and (2). The proofs are hence omitted.

**Lemma 1:**
1) \( \Phi(0) = \Phi(l), \quad l = 1,2,\ldots. \)
2) \[ \mathcal{S}(\Phi(0)) = \max_{j=1,\ldots,n-1} |X(\omega_j)| \]
and
\[ \mathcal{S}(\Phi(l)) = \min_{j=1,\ldots,n-1} |X(\omega_j)|. \]
3) Suppose \( x \) is an \( n \)-periodic signal, and \( \Phi(0) \) and \( \Phi_1(0) \) are the Toeplitz matrices with symbols \( x \) and \( S^l x \), respectively. Then \( \Phi(0) \) and \( \Phi_1(0) \) have the same set of singular values.

It follows immediately from Lemma 1 that \( \Phi(0) \) is invertible if and only if its symbol satisfies
\[ \min_{j=1,\ldots,n-1} |X(\omega_j)| > 0. \]

Now, let us revisit the identification errors in (13).

**Proposition 1:** If \( x \) is \( n \)-periodic and \( \lambda = \mathcal{S}(\Phi(0)) > 0 \), then
\[ \sup_{t \geq 0} \| E(t) \|_\sigma \leq \varepsilon_n + \frac{a_\sigma(n)\sqrt{n}\varepsilon_d}{\lambda} \]
where \( a_\sigma(n) \) is defined in (2).

**Proof:** See the proof in Appendix 1. \( \square \)

**Remark:** Express \( \Phi(0) = \Phi_x(0) \), with its dependence on the symbol \( x \) explicitly denoted. Let \( x' = \frac{x}{\|x\|_\infty} \) be the normalized input. Obviously, \( \|x'\|_\infty = 1 \). Denote \( \eta = \varepsilon_d/\|x\|_\infty \), the noise/signal ratio. By linearity
\[ \Phi_x(0) = \frac{1}{\|x\|_\infty} \Phi_x(0). \]

As a result, (14) can be expressed as
\[ \sup_{t \geq 0} \| E(t) \|_\sigma \leq \varepsilon_n + \frac{a_\sigma(n)\sqrt{n}\eta}{\lambda}. \]
where \( \lambda' = \mathcal{S}(\Phi_x(0)). \)
B. Feedback Invariance of Periodic Signals

Plant input signals appearing in feedback systems are not directly accessible for experiment design in identification. For example, the external probing signal \( r \) and the plant input \( u \) are related by a feedback system \( M \). \( M \) is LTI (or slowly time-varying) when both \( P \) and \( F \) are LTI (or slowly time-varying). We will show that periodicity and ranks of a signal are invariant, at least approximately, under such feedback mappings.

LTI Systems: Consider signals which are generated by \( n \)-periodic signals passing through an LTI stable system \( M \in \mathbb{B}_0 \)

\[
x(t) = Mu = \sum_{\tau=0}^{\infty} m(\tau)u(t-\tau)
\]

where \( m(\cdot) \) is the kernel of \( M \).

Let \( \omega_j = \frac{2\pi j}{n}, j = 0, \ldots, n-1 \). Denote \( M(\omega) = \sum_{\tau=0}^{\infty} g(\tau)e^{-i\omega\tau} \). Lemma 2 is a standard result [23].

**Lemma 2:** If \( u \) is \( n \)-periodic, then \( x \) is \( n \)-periodic and

\[
X(\omega_j) = M(\omega_j)U(\omega_j), \quad j = 0, 1, \ldots, n-1.
\]

An immediate consequence of Lemmas 1 and 2 is that the Toeplitz matrix \( \Phi_x \) is invertible if

\[
\min_{j=0,\ldots,n-1} |M(\omega_j)| > 0 \quad \text{and} \quad \min_{j=0,\ldots,n-1} |U(\omega_j)| > 0.
\]

It is noted that the condition \( \min_{j=0,\ldots,n-1} |M(\omega_j)| > 0 \) is always satisfied if \( M \) is a stable feedback system in which the plant and controller do not have boundary poles and zeros. As a result, periodicity and ranks of \( u \) are preserved in \( x \).

Slowly Varying Systems: However, when \( M \) is slowly time-varying, in general its output \( x \) is not \( n \)-periodic even when the input \( u \) is \( n \)-periodic. Suppose \( M \in \mathbb{B} \) is slowly varying with rate

\[
d_\ast(M) \leq \gamma \quad \text{and} \quad x = Mu.
\]

First, we will show that if \( u \) is \( n \)-periodic and \( M \) is slowly varying, then \( x \) is approximately \( n \)-periodic.

Observe that \( x \in L^\infty \) is \( n \)-periodic if and only if

\[
x - S^n x = 0.
\]

**Definition:** A signal \( x \in L^\infty \) is said to be nearly \( n \)-periodic of deviation \( \delta \) if

\[
\|x - S^n x\|_\infty \leq \delta.
\]

**Proposition 2:** If \( u \in L^\infty \) is \( n \)-periodic, then \( x \) is nearly \( n \)-periodic of deviation \( \delta \)

\[
\delta = n\gamma\|u\|_\infty.
\]

**Proof:** Since \( M \) is slowly time-varying with rate \( \gamma \), by (6) we have

\[
\|SM - MS\|_\infty \leq \gamma.
\]

It follows that

\[
\|x - S^n x\|_\infty = \|Mu - S^n Mu\|_\infty = \|MS^n u - S^n Mu\|_\infty \quad \text{since } u \text{ is } n \text{-periodic}
\]

\[
= \|(MS - SM)(S^{n-1} u) + \cdots + S^{n-1}(MS - SM)u\|_\infty \leq n\gamma\|u\|_\infty.
\]

Next, we study rank conditions. The frozen-time system of \( M \) at \( t \) is denoted by \( M_t \). Without loss of generality, assume \( t = n - 1 \) and define \( x_n = M_{n-1}u \). \( \Phi_x \) and \( \Phi_{xn} \) are the Toeplitz matrices of symbols \( x \) and \( x_n \), respectively.

\[
\Phi_x = \begin{bmatrix} x(0), & \cdots, & x(-n+1) \\ \vdots & \ddots & \vdots \\ x(n-1), & \cdots, & x(0) \end{bmatrix}, \\
\Phi_{xn} = \begin{bmatrix} x_n(0), & \cdots, & x_n(-n+1) \\ \vdots & \ddots & \vdots \\ x_n(n-1), & \cdots, & x_n(0) \end{bmatrix}.
\]

**Proposition 3:** If \( u \in L^\infty \), then \( \Phi_x \) is approximated by \( \Phi_{xn} \)

\[
\sigma(\Phi_x - \Phi_{xn}) \leq \gamma\alpha
\]

where \( \alpha = (2n-1)(n-1)\|u\|_\infty \).

**Proof:** See the proof in Appendix 1.

Since the frozen-time system \( M_{n-1} \) is shift invariant, the extreme singular values \( \sigma(\Phi_{xn}) \) are characterized by Lemmas 1 and 2. It follows from Proposition 3 that if \( u \) is \( n \)-periodic and full rank, and \( M \) is slowly time-varying, then

\[
\sigma(\Phi_x) \geq \sigma(\Phi_{xn}) - \gamma\alpha, \quad \sigma(\Phi_x) \leq \sigma(\Phi_{xn}) + \gamma\alpha.
\]

Consequently, \( \sigma(\Phi_{xn}) > 0 \) implies \( \sigma(\Phi_x) > 0 \) whenever \( \gamma \) is sufficiently small.

VI. Persistent Identification of Slowly Varying Systems

The problem here is to identify a slowly varying system \( G \in \mathbb{B} \) with rate \( d_\ast(G) \leq \gamma \) via observations from

\[
y = Gz + d
\]

where \( \|z\|_\infty \leq \varepsilon_t \). Initially, \( a \) priori information on \( G \) about its dynamics \( G_t \) at some target time \( t > 0 \) is that \( G_t \in U_\varepsilon \subseteq \mathbb{B}_0 \). The prior uncertainty set \( U_\varepsilon \) will consist of systems \( G \) with exponentially decaying memory, and we denote

\[
\varepsilon_n = \sup_{G \in U_\varepsilon} \|I[\tau,\infty)G\|_\sigma < \infty
\]

and

\[
\zeta(n) = \sup_{G \in U_\varepsilon} \sum_{t=0}^{\infty} |\tau(t)| < \infty.
\]

For systems with exponential decaying memory, \( \zeta(n) \) is an exponentially decaying function of \( n \). Take, for instance, the uncertainty set

\[
U_\varepsilon = \{G \in \mathbb{B}_0 : |g(t)| \leq C\varepsilon_t, \quad t = 0, 1, \cdots\}.
\]
For $\sigma < \sigma_1 < \sigma_0$

$$\zeta(n) \leq C \sum_{t=n}^{\infty} \frac{n}{t} t_0^{-n}$$

$$\leq C \sigma_1^{-n} \sum_{t=n}^{\infty} \frac{t}{t_0} \left( \frac{\sigma_0}{\sigma_1} \right)^{-t}$$

$$= c_1 \sigma_1^{-n}$$

where

$$c_1 = C \sum_{t=n}^{\infty} \left( \frac{\sigma_0}{\sigma_1} \right)^{-t}$$

is a constant. As a result, $\zeta(n)$ approaches zero exponentially as $n \to \infty$.

Note that the identification problem for the systems described in (9) can be easily transformed into a case of (16) with $G = N$ and $x = D^{-1}u$, noting that $D$ is known and time invariant.

The identification of the frozen-time system $G_t$ (with kernel $g_t$) is to be performed on $n$ consecutive observations on $y$ in the time interval $[t - (n - 1), t]$. Without loss of generality, assume $t = n - 1$ (or any fixed $t$). Define

$$\theta_t = \begin{bmatrix} g_{n-1}(t) \\ \vdots \\ g_{n-1}(t + (n - 1) \cdot 1) \\ g_{n-1}(t + (n - 1) \cdot 2) \end{bmatrix}, \quad Y = \begin{bmatrix} y(0) \\ \vdots \\ y(n - 1) \end{bmatrix}$$

and $\Phi_x(I)$ as the Toeplitz matrices of symbol $x$ as in (11).

Also, define $x' = y/n$, the normalized probing signal with $|x'|_{\infty} = 1$, and $\eta = \varepsilon_d/|x'|_{\infty}$ the noise/signal ratio.

**Assumption 2:**

1) $G \in \mathcal{B}$ has variation rate $d_1(G) \leq \gamma$.

2) $\hat{G}_t$ is the estimate of $G_t$

3) $\varepsilon_n = \sup_{G \in \mathcal{U}_n} |\bar{G}_{[n, \infty)}|_{\sigma} < \infty$

$\varepsilon_n$ is an exponentially decaying function of $n$, $\varepsilon_n \leq d_1 \sigma_1^{-n}$ and for some $\sigma_1 > \sigma$

$$\zeta(n) = \sup_{G \in \mathcal{U}_n} \sum_{t=n}^{\infty} |g(t)| \leq C \sigma_1^{-n}.$$

It is observed that for the uncertainty set (17), condition (19) is satisfied.

**A. Open-Loop Identification with Periodic Probing Signals**

**Proposition 4:** Under Assumption 2, if $x$ is $n$-periodic and full rank with $X = g(\Phi_x(0)) > 0$, then

$$||G_t - \hat{G}_t||_{\sigma} \leq c_1$$

where

$$c_1 = 2 \varepsilon_n + \frac{a_\sigma(n)}{\lambda} \sqrt{\frac{(n-1)n}{2} + \sqrt{\eta}}.$$

**Proof:** See the proof in Appendix 2.

**Remark 1:** The first term in $c_1$ is a function of the metric complexity (optimal modeling error) of the a priori information. The second term depends on the variation rate $\gamma$ and probing capability of $x$. Note that the second term is a function of $x'$. As a result, the reduction of the second term cannot be achieved by increasing the magnitude of the probing signal $x$.

The third term is a function of noise/signal ratio, as expected in filtering problems.

2) Asymptotically

$$\lim_{\gamma \to 0} \lim_{\eta \to 0} c_1 = 2 \varepsilon_n$$

which, according to [31], is the optimal radius of the a posteriori uncertainties for some typical classes of LTI systems.

**B. Closed-Loop Identification**

In an adaptive control setting in which both $P$ and $F$ are slowly varying, the probing signal $x$ to the plant $G$ is the output of a slowly varying system $M$, $x = M u$, where $u$ is $n$-periodic and full rank. $M \in \mathcal{B}$ is slowly time-varying with rate $d_2(M) \leq \gamma_m$ and $||M||_{11} = k_m$.

**Theorem 1:** Under Assumption 2

1) $||G_t - \hat{G}_t||_{\sigma} \leq c_2$

where

$$c_2 = 2 \varepsilon_n + \frac{a_\sigma(n)}{\lambda} \frac{\lambda}{2} \gamma_m \gamma_m \sqrt{\eta} + \sqrt{h \eta}$$

provided

$$\gamma_m < \frac{\lambda}{2}$$

where $\lambda_m = g(x_{n-1})$, $x_n = M_{n-1} u$, and $\gamma = (2n-1)(n-1)/||u||_{11}$.

2) Suppose, in addition, $\gamma = \inf_{\omega} |M_{n-1}(\omega)| > 0$. Denote

$$u' = \frac{u}{||u||_{11}}, \quad \eta = \frac{\varepsilon_d}{||u||_{11}}$$

Then

a) $||G_t - \hat{G}_t||_{\sigma} \leq c_3$

where

$$c_3 = 2 \varepsilon_n + \frac{a_\sigma(n)}{\lambda} \left( 2n-1 \right) ||u||_{11} \left( \gamma_m (2n-1)(n-1) \right)$$

$$\times \left( 2n-1 \right) \sqrt{\eta} + \frac{\gamma (n-1)n}{2} k_m + \sqrt{h \eta}. \right)$$. (21)

3 To make the notation consistent for this section, we use $x$ and $u$ in place of $u$ and $r$ in Fig. 1.
b) For any given \( \varepsilon_0 > 0 \), there exist an integer \( n \) (model order) and \( \gamma_0 > 0 \) such that one can construct an \( n \)-periodic full rank probing signal \( u_k \) for which

\[
\gamma_3 \leq \varepsilon_0
\]

provided \( \gamma \leq \gamma_0, \gamma_m \leq \gamma_0 \).

Proof: See the proof in Appendix 2.

Conceptually, it is quite straightforward to see that to reduce \( \gamma_3 \) below \( \varepsilon_0 \), one can reduce \( \varepsilon_n \) by increasing \( n \) (Assumption 2), reduce the second and third terms by allowing smaller \( \gamma \) and \( \gamma_m \), and the fourth term by increasing \( \|u\|_\infty \). However, when \( M \) is actually a feedback system, its variation rate \( \gamma_m \) is a function of the plant and controller variation rates, as well as estimation error \( \varepsilon_3 \). To avoid circular arguments, we must demonstrate that \( \varepsilon_3 \) is uniformly bounded. This will be sought in the next subsection.

C. Contraction Properties

When one applies the identification mapping (18) in an adaptation scheme, the feedback mapping \( M \) is a time-varying system. Conceptually, it is obvious that due to identification-adaptation interaction, the variation rates of \( M \) will depend on the variation rates of the plant and controller, as well as estimation errors. It will be shown in Section VII that for the adaptation schemes employed in this paper, the variation rates of \( M \) are bounded by

\[
\gamma_m \leq a_1 \varepsilon_3(t) + a_2 \gamma
\]

(22)

where \( a_1, a_2 \) are constants, and \( \varepsilon_3(t) \) is the identification error bound at \( t \). As a result, according to (21) and (22) the error bound \( \varepsilon_3(t) \) in (21) evolves dynamically in an iteration inequality

\[
\varepsilon_3(t + 1) \leq f(\varepsilon_3(t)).
\]

(23)

To establish uniform boundedness of identification errors in adaptation, we must show that there exists a neighborhood of \( \varepsilon_3 \) at zero which is invariant under the iteration mapping (23).

Theorem 2 (Contraction Properties): There exists a sequence of neighborhoods

\[
\Omega_k = [-\xi_k, \xi_k], \quad k = 0, 1, \cdots
\]

(24)

with \( \xi_k \to 0 \) as \( k \to \infty \), such that for each \( k \), there exist \( \gamma(k) \), \( \eta(k) \), \( n(k) \) for which \( \Omega_k \) is invariant under the iteration mapping (23), provided

\[
\gamma \leq \gamma(k), \quad \frac{\varepsilon_d}{\|u\|_\infty} \leq \eta(k), \quad n = n(k).
\]

Proof: First, let \( \gamma = 0, \eta = \varepsilon_d/\|u\|_\infty = 0 \). Then by (22)

\[
\gamma_m \leq a_1 \varepsilon_3. \quad \text{By (21)}
\]

\[
f(\varepsilon_3(t)) = 2\varepsilon_n + \frac{\varepsilon_3(t)\eta(n)}{b - \varepsilon_3p(n)}
\]

where

\[
\eta(n) = a_1a_\sigma(n)n(2n - 1)(n - 1)\zeta(n) / b = a_1n \eta(n),
\]

\[
p(n) = a_1(2n - 1)(n - 1).
\]

Define

\[
\xi_k = \frac{b}{(k + 2)p(n)}, \quad k = 0, 1, \cdots
\]

Then for \( \varepsilon_3(t) \leq \xi_k \), we have

\[
f(\varepsilon_3(t)) \leq 2\varepsilon_n + \frac{\eta(n)}{(k + 1)p(n)}.
\]

(25)

To have \( \varepsilon_3(t + 1) \leq \xi_k \), we only need to show that there exists \( n \) such that

\[
f(\varepsilon_3(t)) \leq 2\varepsilon_n + \frac{\eta(n)}{(k + 1)p(n)} < \xi_k = \frac{b}{(k + 2)p(n)}.
\]

(26)

A sufficient condition for (25) is

\[
2(k + 2)\varepsilon_n p(n) + 2q(n) < b.
\]

(27)

However, by Assumption 2, since \( \varepsilon_n \leq d_0 \sigma^{-n} \) and \( \zeta(n) \leq \alpha_1^{-n} \) with \( \alpha_1 > \sigma, 2(k + 2)\varepsilon_n p(n) + 2q(n) \) is bounded by

\[
2(k + 2)d_1 \sigma^{-1} p(n) + 2a_1 a_\sigma(n)n(2n - 1)(n - 1)\alpha_1^{-n}
\]

which, for sufficiently large \( n \), is monotone decreasing in \( n \) and

\[
2(k + 2)d_1 \sigma^{-1} p(n) + a_1 a_\sigma(n)n(2n - 1)(n - 1)\alpha_1^{-n} \to 0,
\]

as \( n \to \infty \).

As a result, there exists \( n_0 \) for which the strict inequality (25) is valid for \( n = n_0 \). Define

\[
n(k) = n_0 \quad \text{and} \quad \xi_k = \frac{b}{(k + 2)p(n_0)},
\]

(26)

Now, by the continuity of \( f(\cdot) \) with respect to \( \gamma \) and \( \varepsilon_d/\|u\|_\infty \) as well as the strict inequality of (25), there exist \( \gamma(k) \) and \( \eta(k) \) such that for any \( \varepsilon_3(t) \leq \xi_k \)

\[
f(\varepsilon_3(t)) < \xi_k
\]

provided \( \gamma \leq \gamma(k) \) and \( \varepsilon_d/\|u\|_\infty \leq \eta(k) \). Therefore, \( \Omega_k = [-\xi_k, \xi_k] \) is invariant under the iteration mapping (23).

Finally, since \( \xi_k \) defined in (27) satisfies \( \xi_k \to 0 \) as \( k \to \infty \), the invariant neighborhoods \( \Omega_k \) are vanishing indeed. This completes the proof.

VII. PERSISTENT ADAPTATION

A. Adaptation Mapping

The feedback controller \( F = Y^{-1}X \) is updated at every \( t \) via the frozen-time approach. At time \( t \), the posterior information \( \hat{\Omega}_t \) of the plant is obtained via identification and \( F(t) = Y_t^{-1}X_t \) is designed which robustly stabilizes \( \hat{N}_t \).

For the uncertainty sets considered in this paper, \( \hat{N}_t \) (to be constructed later) takes the form of

\[
\hat{N}_t = \{ \hat{A}_t = (\hat{N}_t + \Delta)D^{-1}, \|\Delta\|_{\sigma} \leq \varepsilon \}
\]

where \( \hat{N}_t \) is an estimate of \( N_t \) and \( \varepsilon \) the identification error in \( H_\varepsilon^2 \). By the \( H_\varepsilon^2 \) theory of robust stabilization, \( F(t) = \)

4It should be pointed out here that in general, \( Y_t^{-1}X_t \neq Y^{-1}X \). To avoid confusion, the symbol \( F(t) \) is used to represent \( Y_t^{-1}X_t \).
robustly stabilizes \( \Omega_t \) in \( H^\infty_\sigma \) if one solves the Bezout equation

\[
X_t \hat{N}_t + Y_t D = 1, \quad X_t, Y_t \in H^\infty_\sigma \tag{28}
\]

and

\[
\|X_t\|_\sigma \leq \frac{1}{\varepsilon}. \tag{29}
\]

Let

\[
\mu_t = \inf_{Q \in H^\infty_\sigma} \|X_0^0 - DQ\|_\sigma \tag{30}
\]

where the infimum is taken over all \( X_t \in H^\infty_\sigma \) satisfying (28) (with some \( Y_t \in H^\infty_\sigma \)). More precisely, if \( X_0^0, Y_0^0 \in H^\infty_\sigma \) solve (28), then by Youla’s parameterization, all stabilizing controllers for \( \hat{N}_t D^{-1} \) are parameterized by

\[
X_t = X_0^0 - DQ, \quad Y_t = Y_0^0 + \hat{N}_t Q
\]

with \( Q \in H^\infty_\sigma \) and \( (Y_0^0 + \hat{N}_t Q)^{-1} \) well-posed.

It follows that

\[
\|X_t\|_\sigma \leq \mu_t + \delta.
\]

Consequently, we can compute

\[
Q = D^{-1}(X_0^0 - X_t) \quad \text{and} \quad Y_t = Y_0^0 + \hat{N}_t Q.
\]

\( Y_t^{-1}X_t \) robustly stabilizes \( \Omega_t \) in \( H^\infty_\sigma \) if

\[
(\mu_t + \delta)\varepsilon < 1.
\]

Define the adaptation mapping as

\[
A[\hat{N}_t] = \begin{bmatrix} X_t \\ Y_t \end{bmatrix} \tag{33}
\]

It has been shown [35] that the mapping is Lipschitz continuous \( (H^\infty_\sigma \rightarrow H^\infty_\sigma) \), namely

\[
\left\| \begin{bmatrix} X_{t+1} \\ Y_{t+1} \end{bmatrix} - \begin{bmatrix} X_t \\ Y_t \end{bmatrix} \right\|_{H^\infty_\sigma} \leq c_1 \|\hat{N}_{t+1} - \hat{N}_t\|_\sigma \tag{34}
\]

where the Lipschitz constant \( c_1 = 2\sqrt{2} (\mu_t + \delta)^2 \).

The Lipschitz continuity (34) is required to guarantee that \( X \) and \( Y \) are slowly time-varying when \( \hat{N} \) is slowly varying.

### B. Identification Precision

Robust stabilizability requirements mandate that the identification error \( \varepsilon \) must satisfy

\[
\varepsilon < \frac{1}{\mu_t + \delta}. \tag{35}
\]

For (35) to hold for all \( t \), \( \mu_t \) must be uniformly bounded. Observe that \( \mu_t \) is a function of \( \hat{N}_t \), which will be explicitly denoted by \( \mu_t(\hat{N}_t) \).

Since \( \mu_t \) is a function of \( \hat{N}_t \) and \( D \), it depends on the \( a \) priori information \( U_a \) as well as identification. A uniform upper bound on \( \mu_t \) can be computed. Observe that \( \hat{N}_t \in U_a \) and

\[
\|\hat{N}_t - \bar{N}_t\|_\sigma \leq \varepsilon
\]

which implies

\[
\hat{N}_t \in \bar{U} = U_a + \text{Ball}_\sigma(\varepsilon)
\]

where

\[
\text{Ball}_\sigma(\varepsilon) := \{ \Delta \in H^\infty_\sigma : \|\Delta\|_\sigma \leq \varepsilon \}
\]

Define

\[
\mu(\varepsilon) = \sup_{\hat{N}_t \in \bar{U}} \mu_t(\hat{N}_t). \tag{36}
\]

\( \mu(\varepsilon) \) is an upper bound on \( \mu_t \) when the underlying identification error is bounded by \( \varepsilon \). Note that \( \mu(\varepsilon) \) is a monotone increasing function of \( \varepsilon \).

Since

\[
\frac{1}{\mu(\varepsilon) + \delta} < \frac{1}{\mu_t + \delta}
\]

a sufficient condition for (35) is

\[
\varepsilon < \frac{1}{\mu(\varepsilon) + \delta}. \tag{37}
\]

Inequality (37) is always satisfied for sufficiently small \( \varepsilon \), since

\[
\frac{1}{\mu(\varepsilon) + \delta} > 0.
\]

### C. Design Specifications and Procedures

The closed-loop (time-varying) system is given by

\[
y = ND^{-1}u + d
\]

\[
Y u = -X y + r. \tag{38}
\]

By defining \( v = D^{-1}u \), we get

\[
u = Dv
\]

\[
y = N v + d
\]

\[
(XN + YD)v = r +Xd.
\]

The design objective is expressed by design specifications \( \mathcal{P} \) given in Section IV. Namely, we would like to establish \( \mathcal{P}(t) \) holds for all \( t \)."

Now, we specify our design parameters and procedures. Let \( \Theta_k = [-\xi_k, \xi_k] \), \( k = 0, 1, \ldots \) be the monotone decreasing invariant neighborhoods defined in Theorem 2, and \( n(k), \gamma(k), \rho(k) \) the corresponding parameters. Also, define \( \mu(\xi_k) \) by (36).
Design Procedure $\mathcal{D}_k$: For the selected $\xi_k$ and $n = n(k)$, we have the following.

1) The adaptation mapping is defined by the central $\delta$-suboptimal design, given by (33).

2) The estimate $\hat{N}_t$ of $N_n$ is constructed from $n$ observations on $y$ in $[t-n+1, t]$ via the least squares estimation

$$\hat{\theta}^n = \Phi_e^{-1}(0)Y_n$$

where $\Phi_e(0)$ is the $n \times n$ Toeplitz matrix of symbol $v$, $Y_n = [y(t-n+1), \ldots, y(t)]^T$, and

$$\hat{N}_t(z) = [1, z, z^2, \ldots, z^{n-1}]\hat{\theta}^n.$$ 

3) The probing signal $r$ is $n$-periodic and full rank.

D. Main Results

Theorem 3: Under Assumption 1, there exists $k_0 \geq 0$ such that for any $k \geq k_0$ a corresponding $\gamma(k) > 0$ can be found for which the design procedure $\mathcal{D}_k$ guarantees that $P(t)$ holds for all $t \geq 0$ with $\varepsilon = \xi_k$ in (39), provided that $P(0)$ holds and the variation rate $\gamma$ of the plant satisfies $\gamma \leq \gamma(k)$.

Remark: Theorem 3 claims that for any disturbance bounds $\varepsilon_d$ and any a priori information $U_a$ satisfying Assumption 1, an adaptive design procedure $\mathcal{D}_k$ can be devised to achieve uniform stabilization and uniform identification error bounds $\xi_k$, provided that the variation rate $\gamma$ of the plant is sufficiently small. The allowable variation rate $\gamma$ as well as the signal bounds $k_u$ and $k_y$, however, are functions of the metric complexity $\xi_n$ of $U_a$, disturbance bound $\varepsilon_d$, and the required identification precision $\xi_k$.

The remaining part of this section is devoted to the proof of Theorem 3. First, Property $\mathcal{P}$ will be expanded.

Design Specifications $\mathcal{P}$: $\mathcal{P}(t)$, consisting of the following components, must hold for all $t$.

a) Uniform Identification Error Bounds:

$$\sup_{-\infty \leq \tau \leq t} ||\hat{N}_t - N_t||_{(\sigma)} \leq \varepsilon. \tag{39}$$

b) Slowly Varying Estimates:

$$\sup_{-\infty \leq \tau \leq t} ||\hat{N}_t - \hat{N}_{t-1}||_{(\sigma)} \leq \dot{\varepsilon}, \tag{40}$$

c) $||I_{(-\infty,T]}v||_{\infty} \leq \kappa_r ||v||_{\infty} + \kappa_e \epsilon d$. \tag{41}$

It is easy to show that

$$\mathcal{P}(t) \Rightarrow \mathcal{P}(t),$$

Indeed, since

$$u = Dv, \quad y = Nv + d$$

we have, by the causality of $N$ and $D$

$$||I_{(-\infty,T]}u||_{\infty} \leq ||D||_1 ||I_{(-\infty,T]}v||_{\infty} \leq \kappa D(\kappa_r ||v||_{\infty} + \kappa_e \epsilon d)$$

$$||I_{(-\infty,T]}y||_{\infty} \leq ||N||_1 ||I_{(-\infty,T]}v||_{\infty} + \epsilon d \leq \kappa_N \kappa_r ||v||_{\infty} + (\kappa_N \kappa_d + 1) \epsilon d.$$ Therefore $\mathcal{P}(t)$ implies $\mathcal{P}(t)$ with $k_u = \kappa_D(\kappa_r ||v||_{\infty} + \kappa_e \epsilon d)$ and $k_y = \kappa_N \kappa_r ||v||_{\infty} + (\kappa_N \kappa_d + 1) \epsilon d$. Hence, we only need to demonstrate $\mathcal{P}$ here.

Theorem 3 will be proved by induction. Namely, we will establish the implication

$$\mathcal{P}(t) \Rightarrow \mathcal{P}(t+1),$$

Then the conclusion $\mathcal{P}(0) \Rightarrow \mathcal{P}(t)$ follows by induction.

Recall that the closed-loop system is expressed by (38) with $N$ and $D$ satisfying Assumption 1 as well as $X$ and $Y$ designed by the design procedure $\mathcal{D}_k$. To facilitate the analysis, we define the following artificial time-varying systems: At time $t$, define

$$\tilde{P}(t) = \tilde{N} \mathcal{D}^{-1}, \quad \tilde{P}(t) = \tilde{Y}^{-1} \tilde{X}$$

where for $\tilde{K} \in \{\tilde{N}, \tilde{X}, \tilde{Y}\}$

$$\tilde{K}_t = \begin{cases} K_{\tau}, & \tau \leq t \\ K_t, & \tau > t. \end{cases}$$

In other words, $\tilde{N}$ assumes the same frozen-time systems as $N$ before $t$ and is time invariant after $t$ with dynamics $\tilde{N}_t$, similarly for $\tilde{X}$ and $\tilde{Y}$. Then, the following artificial system can be constructed:

$$\tilde{y} = \tilde{N} \mathcal{D}^{-1} \tilde{u} + d$$

$$\tilde{Y}_{\tilde{u}} = -\tilde{X} \tilde{y} + r.$$ By defining $\tilde{u} = D \tilde{u}$

$$\tilde{y} = \tilde{N} \tilde{u} + d$$

$$(\tilde{X}^{\dagger} \tilde{N}^{\dagger} + \tilde{Y}^{\dagger} D) \tilde{v} = r + \tilde{X}^{\dagger} d.$$ By causality

$$\tilde{y}(\tau) = y(\tau), \quad \tilde{u}(\tau) = u(\tau), \quad \tilde{v}(\tau) = v(\tau), \quad \tau \leq t.$$ Define the local and global resolvents of the artificial system by

$$\tilde{R}_{L} = \tilde{X} \odot \tilde{N} + \tilde{Y} \odot D$$

$$\tilde{R}_{g} = \tilde{X} \tilde{N} + \tilde{Y} \mathcal{D}$$

where the local product of two time-varying systems $\tilde{X}$ and $\tilde{N}$ is defined by

$$(\tilde{X} \odot \tilde{N})_{\tilde{K}} = \tilde{X}_{\tilde{K}} \tilde{N}_{\tilde{K}}, \quad \forall \tilde{t}.$$ Two lemmas will be established first.

Lemma 3: Under Assumption 1, there exist $\gamma_2 > 0$ and a design procedure $\mathcal{D}_k$ (with the corresponding $\xi_k$), $k$ being independent of $t$, such that $\tilde{R}_{g}^{-1}$ is the closed-loop system $\tilde{R}_{g}$ is given by

$$d_2(\tilde{R}_{g}) \leq c_1 \xi_k + c_2 \gamma$$

where $c_1, c_2$ are constants, provided $\mathcal{P}(t)$ holds with $\varepsilon = \xi_k$ and $\gamma \leq \gamma_2$.

Proof: See the proof in Appendix 3. \qed
**Lemma 4:** Under Assumption 1, there exist $\gamma_3 > 0$ and a design procedure $\mathcal{D}_k$ (with the corresponding $\xi_k$, $k$ being independent of $t$, such that 1) and 2) of $\hat{p}(t+1)$ are satisfied whenever $\hat{p}(t)$ holds with $\epsilon = \xi_k$ and $\gamma \leq \gamma_3$.

**Proof:** See the proof in Appendix 3. \qed

Now, we are ready to prove Theorem 3.

**Proof of Theorem 3:** Suppose $\hat{p}(t)$ holds. Let $\gamma \leq \gamma_k = \min\{\gamma_2, \gamma_3\}$. By Lemma 4, 1) and 2) of $\hat{p}(t+1)$ are satisfied, when $\gamma \leq \gamma_3$ and the design procedure $\mathcal{D}_k$ is employed.

To prove 3) of $\hat{p}(t+1)$, we define the artificial systems $(\hat{p}(t+1), \hat{f}(t+1))$:

$$\hat{p}(t+1) = \hat{N}^1 D^{-1}, \quad \hat{f}(t+1) = (Y^1)^{-1} \hat{X}^1$$

where $\hat{K}_4 \in \{\hat{N}^1, \hat{X}^1, \hat{Y}^1\}$

$$\hat{K}_4^t = \begin{cases} \hat{K}^t_4, & \tau \leq t + 1 \\ \hat{K}^t_4, & \tau > t + 1. \end{cases}$$

More precisely, the artificial system $(\hat{p}(t+1), \hat{f}(t+1))$ is given by

$$\hat{\eta} = \hat{N}^1 D^{-1} \hat{u} + d$$

$$\hat{Y}^1 \hat{u} = -\hat{X}^1 \hat{y} + r.$$ (42)

By defining $\hat{v} = D^{-1} \hat{u}$, we get

$$\hat{\eta} = \hat{N}^1 \hat{v} + d$$

$$(XN + YD)v = r + Xd.$$ (43)

Denote the corresponding global resolvent as

$$\hat{R}_g^1 = \hat{X}^1 \hat{N}^1 + \hat{Y}^1 D.$$ (44)

As a result

$$\hat{R}_g^1 \hat{\eta} = r + \hat{X}^1 d.$$ (45)

Following the same arguments leading to (53), we conclude

$$||((\hat{R}_g^1)^{-1})||_1 \leq \frac{1}{1 - \zeta},$$

where $\zeta := \frac{\sigma}{\sqrt{\sigma^2 - 1}}(\mu(\hat{\xi}) + \delta)\epsilon_k + \frac{\mu(\hat{\xi}) + \delta}{\sigma^2 - 1}\gamma < 1$. Therefore

$$||\hat{\eta}||_\infty \leq \frac{1}{1 - \zeta}||r||_\infty + \frac{1}{1 - \zeta}||X^1||_1 \epsilon_d$$

$$\leq \frac{1}{1 - \zeta}||r||_\infty + \frac{\sigma}{\sqrt{\sigma^2 - 1}}(\mu(\epsilon) + \delta)\epsilon_d$$

where the last inequality follows from

$$||X^1||_1 \leq \frac{\sigma}{\sqrt{\sigma^2 - 1}}||X^1||_\sigma$$

$$= \frac{\sigma}{\sqrt{\sigma^2 - 1}} \sup_{\tau \leq t} ||X^1||_\tau$$

$$\leq \mu(\epsilon) + \delta$$

by the suboptimal design. Since

$$v(\tau) = \hat{v}(\tau), \quad \tau \leq t + 1$$

we get

$$||T_{(\infty, \infty)} v||_\infty \leq \kappa_r ||r||_\infty + \kappa_d \epsilon_d$$

where $\kappa_r = \frac{1}{1 - \zeta}$ and $\kappa_d = \frac{\sigma}{\sqrt{\sigma^2 - 1}}\mu(\epsilon) + \delta$. This verifies 3) of $\hat{p}(t+1)$.

By induction, “$\hat{p}(0)$ holds” implies “$\hat{p}(t)$ holds for all $\tau \geq 0$.” \qed

**VIII. CONCLUDING REMARKS**

Historically, adaptation was introduced as a means of achieving superior performance for time-varying plants and environment. While some practical adaptive systems often demonstrate these expected benefits, rigorous analysis of them remains a daunting problem. This paper is a preliminary effort in establishing a framework in which the ability of a controller to adapt to a time-varying environment and to achieve performance beyond robust control can be rigorously studied. Conceptually, the notion of persistent identification and adaptation is introduced to capture these aspects of adaptive systems. It is shown that the fundamental features of adaptive schemes in such problems are feedback invariance properties for probing signals and invariance principles for identification-adaptation iteration. A stabilization problem is solved in detail.

There remain numerous unresolved issues in persistent adaptation problems. It is highly desirable that system performance, beyond stability, can be analyzed. The main obstacle is the lack of tangible robust performance synthesis methods (available design methods like $H^\infty$ robust performance, $\mu$ synthesis, etc., still encounter difficulties in characterization and numerical solutions). Also, in trading for a concise introduction of the framework, we used some academic arguments in the design procedure. For instance, we assume that noise/signal ratios can be freely changed by increasing the signal magnitudes. When this is not the case (as always in practice), signals must be more carefully selected to minimize the effect of disturbances on identification errors. Furthermore, to fit into the robust control framework, identification is performed in a worst-case paradigm. Inevitably, results may become very conservative. The conservativeness becomes especially severe in persistent adaptation problems since allowable variation rates of the plant are functions of time complexity of the prior uncertainty sets. It is now well understood that time complexity in worst case identification can be extremely high.

**APPENDIX 1**

**Proofs for Section V**

**Proof of Proposition 1:** It follows from Lemma 1 that

$$\Phi^{-1}(0) \sum_{t=1}^\infty \Phi(t) \Phi_t = \sum_{t=1}^\infty \theta_t.$$
As a result
\[ \sup_{\|g\| \leq \varepsilon_n} \left\| \sum_{i=1}^{\infty} \theta_i g \right\|_\sigma \leq \sup_{\|g\| \leq \varepsilon_n} \|T_{[n,\infty)} g\|_\sigma = \varepsilon_n. \]

Furthermore, by (1)
\[ \left\| \Phi^{-1}(0) D_n \right\|_\sigma \leq a_r(n) \| \Phi^{-1}(0) D_n \|_2 \]
\[ \leq a_r(n) \gamma(n) \| \Phi(0) \|_\sigma \]
\[ \leq \frac{a_r(n) \sqrt{n} \varepsilon_d}{\lambda}. \]

Therefore
\[ \sup_{t \geq 0} \| E(t) \|_\sigma \leq \varepsilon_n + \frac{a_r(n) \sqrt{n} \varepsilon_d}{\lambda}. \]

\[ \square \]

Proof of Proposition 3:
\[ \left\| T_{[n+1,n+1]}(x - x_n) \right\|_1 = \left\| T_{[n+1,n+1]}(M - M_{n+1}) x \right\|_1 \]
\[ \leq \sum_{t=n+1}^{n+1} \left\| (M_t - M_{n+1}) x \right\|_\infty \]
\[ \leq \sum_{t=n+1}^{n+1} (n - 1) \gamma \| x \|_\infty \]
\[ = \gamma(n - 1) (2n - 1) \| x \|_\infty \]
\[ = \gamma n. \]

Since \( \Phi_x - \Phi_{x_n} = \Phi_{x_n} x_n \) is a Toeplitz matrix of symbol \( x - x_n \)
\[ \gamma(\Phi_x - \Phi_{x_n}) \leq \left\| T_{[n+1,n+1]}(x - x_n) \right\|_1 \]
\[ \leq \gamma n. \]

\[ \square \]

APPENDIX 2
PROOFS FOR SECTION VI

Proof of Proposition 4: By the definition of frozen-time systems
\[ y(t) = (G_t x)(t) + d(t), \quad t = 0, \ldots, n - 1. \]

Observe that \( y \) is approximated by \( y_{n+1} = G_{n+1} x + d \) in the interval of observation
\[ \left\| T_{[0,n+1]}(y - y_{n+1}) \right\|_1 = \left\| T_{[0,n+1]}(G - G_{n+1}) x \right\|_1 \]
\[ \leq \sum_{t=0}^{n+1} \left\| (G_t - G_{n+1}) x \right\|_\infty \]
\[ \leq \sum_{t=0}^{n+1} (n - 1) \gamma \| x \|_\infty \]
\[ = \gamma n \frac{n(n - 1)}{2} \| x \|_\infty. \]

Define
\[ Y_n = \begin{bmatrix} y_{n+1}(0) \\ \vdots \\ y_{n+1}(n-1) \end{bmatrix}. \]

We have
\[ Y_n = \Phi_x(0) \theta_0 + \sum_{l=1}^{\infty} \Phi_x(l) \theta_l + D_n. \]

Since the observation is performed on \( Y \), not \( Y_n \), we express \( \theta_0 \) as
\[ \theta_0 = \Phi_x^{-1}(0) Y + \Phi_x^{-1}(0) (Y_n - Y) \]
\[ - \Phi_x^{-1}(0) \sum_{l=1}^{\infty} \Phi_x(l) \theta_l - \Phi_x^{-1}(0) D_n. \]

When \( x \) is \( n \)-periodic
\[ \Phi_x(0) = \Phi_x(l), \quad l = 1, 2, \ldots \]

by Lemma 1. Consequently
\[ \theta_0 = \Phi_x^{-1}(0) Y + \Phi_x^{-1}(0) (Y_n - Y) - \sum_{l=1}^{\infty} \theta_l - \Phi_x^{-1}(0) D_n \]
\[ = \hat{\theta}_0 + E \]
where \( \hat{\theta}_0 = \Phi_x^{-1}(0) Y \) is an estimate of \( \theta_0 \). The error term
\[ E = \Phi_x^{-1}(0) (Y_n - Y) - \sum_{l=1}^{\infty} \theta_l - \Phi_x^{-1}(0) D_n \]
is bounded in the \( \| \cdot \|_\sigma \) norm by
\[ \| E \|_\sigma = \left\| \Phi_x^{-1}(0) (Y_n - Y) - \sum_{l=1}^{\infty} \theta_l - \Phi_x^{-1}(0) D_n \right\|_\sigma \]
\[ \leq \left\| \Phi_x^{-1}(0) (Y_n - Y) \right\|_\sigma + \left\| T_{[n+1,\infty)} g \right\|_\sigma \]
\[ + \left\| \Phi_x^{-1}(0) D_n \right\|_\sigma \]
\[ \leq \left\| \Phi_x^{-1}(0) (Y_n - Y) \right\|_\sigma + \varepsilon_n + \frac{a_r(n) \sqrt{n} \varepsilon_d}{\lambda} \]
by Assumption 2 and (43). Moreover, by (1)
\[ \left\| \Phi_x^{-1}(0) (Y_n - Y) \right\|_\sigma \leq a_r(n) \gamma(n) \left\| (Y_n - Y) \right\|_\sigma \]
\[ \leq a_r(n) \gamma(n) \frac{(n-1)n}{2} \| x \|_\infty \]
by (44). Therefore
\[ \left\| E \right\|_\sigma \leq \varepsilon_n + \frac{a_r(n)}{\lambda} \left( \gamma(n-1)n \| x \|_\infty + \sqrt{n} \varepsilon_d \right) \]
\[ = \varepsilon_n + \frac{a_r(n)}{\lambda} \left( \gamma(n-1)n + \sqrt{n} \varepsilon_d \right). \]

Finally
\[ \left\| G_{n+1} - \Phi_{n+1} \right\|_\sigma \leq \left\| \theta_0 - \hat{\theta}_0 \right\|_\sigma + \left\| \sum_{l=1}^{\infty} \theta_l \right\|_\sigma \]
\[ \leq \left\| E \right\|_\sigma + \varepsilon_n \]
\[ \leq 2 \varepsilon_n + \frac{a_r(n)}{\lambda} \left( \gamma(n-1)n + \sqrt{n} \varepsilon_d \right). \]

\[ \square \]
Proof of Theorem 1—1): By Proposition 3

\[ \mathcal{S}(\Phi_x(0) - \Phi_{x_t}(0)) \leq \gamma_m \alpha \]

where \( \alpha = (2n-1)(n-1)\|u\|_\infty \). Suppose \( \lambda_n > 0 \). Then for \( \gamma_m \alpha < \lambda_n \), \( I + (\Phi_x(0) - \Phi_{x_t}(0))(\Phi_{x_t}(0))^{-1} \) is invertible and

\[ (\Phi_x(0))^{-1} = (\Phi_{x_t}(0))^{-1}[I + (\Phi_x(0) - \Phi_{x_t}(0))(\Phi_{x_t}(0))^{-1}]^{-1} \]

and

\[ \mathcal{S}(\Phi_x(0))^{-1} \leq \frac{\mathcal{S}(\Phi_{x_t}(0))^{-1}}{1 - \mathcal{S}(\Phi_{x_t}(0))^{-1}} \gamma_m \alpha \]

\[ = \frac{\lambda_n^{-1}}{1 - \lambda_n^{-1} \gamma_m \alpha} = \frac{1}{\lambda_n - \gamma_m \alpha}, \quad (46) \]

It follows that the error in (45) can be expressed as

\[ E = -\Phi_x^{-1}(0) \left( \sum_{l=1}^\infty \Phi_x(l) \theta_l \right) \]

\[ + \Phi_x^{-1}(0)[Y_n - Y] - \Phi_x^{-1}(0)D_n \]

\[ = -\Phi_x^{-1}(0) \left( \sum_{l=1}^\infty [\Phi_x(l) - \Phi_x(0)] \theta_l \right) \]

\[ + \Phi_x^{-1}(0)[Y_n - Y] - \Phi_x^{-1}(0)D_n \]

\[ = -\sum_{l=1}^\infty (\Phi_x^{-1}(0)[\Phi_x(l) - \Phi_x(0)] + I) \theta_l \]

\[ + \Phi_x^{-1}(0)[Y_n - Y] - \Phi_x^{-1}(0)D_n \]

\[ = -\sum_{l=1}^\infty \theta_l - \sum_{l=1}^\infty \Phi_x^{-1}(0)[\Phi_x(l) - \Phi_x(0)] \theta_l \]

\[ + \Phi_x^{-1}(0)[Y_n - Y] - \Phi_x^{-1}(0)D_n \]

It is easy to verify that

\[ \mathcal{S}(\Phi_x(l) - \Phi_x(0)) \leq \gamma_m \alpha n \alpha \]

As a result

\[ \|E\|_\sigma \leq \left\| \sum_{l=1}^\infty \theta_l \right\|_\sigma + \left\| \sum_{l=1}^\infty \Phi_x^{-1}(0)[\Phi_x(l) - \Phi_x(0)] \theta_l \right\|_\sigma \]

\[ + \Phi_x^{-1}(0)[Y_n - Y] - \Phi_x^{-1}(0)D_n \]

\[ \leq \varepsilon_n + \alpha(n) \mathcal{S}(\Phi_x^{-1}(0)) \left( \sum_{l=1}^\infty \mathcal{S}(\Phi_x(l) - \Phi_x(0))\|\theta_l\|_2 \right) \]

\[ \leq \varepsilon_n + \frac{\alpha(n)}{\lambda_n - \gamma_m \alpha} \left( n \gamma_m \alpha \sum_{l=1}^\infty \|\theta_l\|_1 \right) \]

\[ + \|Y_n - Y\|_1 + \sqrt{\mathbb{d}} \|D_n\|_\infty \]

by (46)

\[ \leq \varepsilon_n + \frac{\alpha(n)}{\lambda_n - \gamma_m \alpha} \left( n \gamma_m \alpha (n) + \gamma (n - 1) n k_m ||u||_\infty \right) \]

\[ + \sqrt{n}\varepsilon_\mathbb{d} \]

by (43) and (19). Therefore

\[ \|G_t - \hat{G}_t\|_\sigma \]

\[ \leq \|E\|_\sigma + \varepsilon_n \leq 2\varepsilon_n + \frac{\alpha(n)}{\lambda_n - \gamma_m \alpha} \]

\[ \times \left( \gamma_m n (n - 1) k_m ||D_n||_\infty + \sqrt{n}\varepsilon_\mathbb{d} \right). \]

2)-a): It follows from Lemmas 1 and 2 that

\[ \lambda_n \geq a \lambda. \]

Let \( u' = u/||u||_\infty \) be the normalized probing signal. By linearity, \( \lambda = ||u'|| \). Now, by (20)

\[ \varepsilon_\mathbb{c} \leq 2\varepsilon_n + \frac{\alpha(n)}{\lambda \gamma_m (2n - 1)(n - 1)} \]

\[ \times \left( \gamma_m n (2n - 1)(n - 1) \zeta(n) \right) \]

\[ + \gamma \frac{n - 1}{2} k_m + \sqrt{n}\varepsilon_\mathbb{d} \cdot \]

b): Let \( \gamma_0 = \frac{1}{2}(\frac{2a\lambda'}{\gamma_m (2n - 1)(n - 1)} \gamma_m n (2n - 1)(n - 1) \zeta(n) \]

\[ \leq \alpha(n) n \zeta(n). \]

Since by Assumption 2, \( \varepsilon_n \leq d_t \sigma^{-n} \) and \( \zeta(n) \leq c_0 t^{-n} \) with \( \sigma_1 > \sigma \), for any \( c_0 > 0 \), there exists \( n \) for which

\[ 2\varepsilon_n \leq \frac{c_0}{4}; \quad \alpha(n) n \zeta(n) \leq \frac{c_0}{4}; \]

Moreover, by selecting \( \gamma_0 \) and \( ||u'||_\infty \) to satisfy

\[ \frac{2a\alpha(n)(n - 1)n}{\alpha \lambda'} k_m \leq \frac{c_0}{4} \]

and

\[ \frac{2a\alpha(n)}{\alpha \lambda'} \sqrt{n}\varepsilon_\mathbb{d} \leq \frac{c_0}{4} \]

we have

\[ c_3 \leq \frac{c_0}{4} \leq \frac{c_0}{4} + \frac{c_0}{4} + \frac{c_0}{4} = c_0. \]

The proof is completed by defining

\[ \gamma_0 = \min \{ \gamma_0, \gamma_3 \}. \]

\[ \square \]
Appendix 3
Proofs for Section VII

Proof of Lemma 3: By the Lipschitz continuity of the adaptation mapping
\[
\| \hat{X}_t - \hat{X}_{t-1} \|_{\| F \|_1} \leq c \| \hat{N}_t - \hat{N}_{t-1} \|_\sigma, \quad t \leq t.
\]
The right-hand side of (47) is bounded by
\[
\| \hat{N}_t - \hat{N}_{t-1} \|_\sigma \leq \| \hat{N}_t - N_t \|_\sigma + \| N_t - N_{t-1} \|_\sigma \\
+ \| \hat{N}_{t-1} - N_{t-1} \|_\sigma \\
\leq 2\epsilon + \gamma.
\]
Furthermore, by $\delta$-suboptimality
\[
\| \hat{X} \|_\sigma \leq \mu(\epsilon) + \delta.
\]
By (5), we have
\[
d_4(\hat{X}) \leq \frac{\sigma c}{\sqrt{\sigma^2 - 1}} (2\epsilon + \gamma), \\
d_4(\hat{Y}) \leq \frac{\sigma c}{\sqrt{\sigma^2 - 1}} (2\epsilon + \gamma)
\]
\[
\| \hat{X} \|_1 \leq \frac{\sigma}{\sqrt{\sigma^2 - 1}} (\mu(\epsilon) + \delta).
\]
Also
\[
d_4(N) \leq \frac{\sigma}{\sqrt{\sigma^2 - 1}} (2\epsilon + \gamma).
\]
Note that by Assumption 1 and (a) of $\bar{P}(t)$
\[
\| \hat{N} \|_1 \leq \| \hat{N} \|_1 + \| N - \hat{N} \|_1 \\
\leq \| N \|_1 + \frac{\sigma c}{\sqrt{\sigma^2 - 1}} \epsilon \\
\leq \kappa_N + \frac{\sigma c}{\sqrt{\sigma^2 - 1}} \epsilon \\
\| D \|_1 \leq \kappa_D.
\]
It follows that
\[
d_4(\hat{R}_g) \leq \| \hat{R}_g T - T\hat{R}_g \|_1 \\
= \| \hat{X} \|_1 d_4(\hat{N}) + \| \hat{N} \|_1 d_4(\hat{X}) + \| D \|_1 d_4(\hat{Y}) \\
\leq \frac{\sigma^2}{\sqrt{\sigma^2 - 1}} (\mu(\epsilon) + \delta) (2\epsilon + \gamma) \\
+ \left( \kappa_N + \frac{\sigma c}{\sqrt{\sigma^2 - 1}} \epsilon \right) \frac{\sigma c}{\sqrt{\sigma^2 - 1}} (2\epsilon + \gamma) \\
+ \kappa_D \frac{\sigma c}{\sqrt{\sigma^2 - 1}} (2\epsilon + \gamma) \\
\leq a_1 \epsilon + a_2 \gamma
\]
where $a_1, a_2$ are constants.
Now the local resolvent can be expressed as
\[
\hat{R}_g = \hat{X} \hat{N} + \hat{Y} \hat{D}
\]
where $\hat{D}$ is time-invariant, by (41) we have
\[
\hat{Y} \hat{D} = 0.
\]
As a result, $\hat{R}_g - \hat{R}_d = \hat{X} \hat{N} \hat{N} - \hat{X} \hat{N} = 0.$

Proof of Lemma 4: Let us consider the identification problem at $t + 1$. Here, $\hat{N}_{t+1}$ is to be estimated based on input–output observations
\[
y(t) = N v + d.
\]
Since $\hat{N}$ is strictly causal, $\hat{N} = N_0 S$, the identification of $\hat{N}_{t+1}$ requires output observations $y(t - n + 2), \ldots, y(t + 1)$ and input $v(t - 2n), \ldots, v(t)$. Denote
\[
\Phi_v(0) = \begin{bmatrix} v(t), & \cdots, & v(t - n + 1) \\
                    \vdots, & \cdots, & \vdots \\
v(t - n + 1), & \cdots, & v(t - 2n) \end{bmatrix}
\]
and $\Phi_v(0) = \Phi_v(0).$
Observe that when \( \varepsilon \leq \varepsilon_k \) in (52), \( \tilde{R}^1 \in \mathcal{B} \) and
\[
\tilde{v} = \tilde{R}^1 y + \tilde{R}^0 \tilde{X} d = \tilde{v}_y + \tilde{v}_d.
\]

Let
\[
r' = \frac{r}{\|r\|_{\infty}}, \quad d = \frac{d}{\|d\|_{\infty}}, \quad \text{and} \quad \eta = \frac{\varepsilon_d}{\|d\|_{\infty}}.
\]

By linearity
\[
\tilde{v} = \|v\|_{\infty} (\tilde{R}^1 r' + \tilde{R}^0 \tilde{X} d') = \|v\|_{\infty} (\tilde{v}_y' + \tilde{v}_d').
\]

Now
\[
\Phi_{\tilde{v}}(0) = \|v\|_{\infty} (\Phi_{\tilde{v}_y}(0) + \Phi_{\tilde{v}_d}(0)).
\]

If \( \Phi(\Phi_{\tilde{v}}(0)) > 0 \) and \( \eta \) sufficiently small (which is always true for sufficiently large \( \|v\|_{\infty} \)), we have \( \Phi(\Phi_{\tilde{v}}(0)) > 0 \). By Lemma 3 and Proposition 4, the identification error is bounded by
\[
e_3(t + 1) = \| \hat{N} t + 1 - \hat{N} t + 1 \|_{\sigma} \leq f(e_3(t)).
\]

Select \( \varepsilon_k \) to satisfy
\[
\varepsilon_k < \frac{\gamma}{2},
\]

where \( \gamma \) is the required rate in the design specifications \( \tilde{P} \). Then for \( \xi = \min(\varepsilon_k, \varepsilon_0) \), by Theorem 2 there exists \( \gamma > 0 \) for which the neighborhood \( [-\xi, \xi] \) is invariant under the iteration mapping (54) when \( \gamma \leq \gamma_3 \). As a result
\[
e_3(t + 1) + \xi_k
\]

provided \( \gamma \leq \gamma_3 \). This implies 1) of \( \tilde{P}(t + 1) \).

Finally, by (48)
\[
\| \hat{N} t + 1 - \hat{N} t + 1 \|_{\gamma} \leq 2 \xi + \gamma.
\]

Since \( \xi \leq \varepsilon_k < \frac{\gamma}{2} \), there exists \( \gamma_0 > 0 \) such that whenever \( \gamma \leq \gamma_0 \), we have
\[
2 \xi + \gamma \leq 2 \xi + \gamma_0 < \frac{\gamma}{2}.
\]

Therefore, 2) of \( \tilde{P}(t + 1) \) is satisfied. Now, Lemma 4 follows by choosing
\[
\gamma_3 = \min \left\{ \gamma_3 + \gamma_3 \right\}.
\]

REFERENCES

Le Yi Wang received the M.E. degree in computer control from the Shanghai Institute of Mechanical Engineering, China, in 1982, and the Ph.D. degree in electrical engineering from McGill University, Montreal, Canada, in 1990. Since 1990 he has been with Wayne State University, Detroit, MI, where he is currently an Associate Professor in the Department of Electrical and Computer Engineering. His research interests include $H_\infty$ optimization, robust control, time-varying systems, system identification, adaptive systems, as well as hybrid and nonlinear systems with automotive applications.

Dr. Wang was awarded the Research Initiation Award in 1992 from the National Science Foundation. He also received the Faculty Research Award from Wayne State University in 1992 and the College Outstanding Teaching Award from the College of Engineering, Wayne State University, in 1995.

Lin Lin (S’91–M’94) received the B.E. and M.E. degrees from Tsinghua University, Beijing, China, in 1984 and 1987, respectively, and the Ph.D. degree from McGill University, Montreal, Canada, in 1993, all in electrical engineering.

From 1985 to 1987, he participated in a research project on power system stabilization at Je Jiang Hydro, China. Since 1993, he has worked on various control and signal processing problems at several places including Nortel and CAE Electronics in Montreal, Canada, and MPD Technologies, New York, NY, where he is currently an Engineering Manager. He has been an Adjunct Professor in the Department of Electrical Engineering, McGill University since 1994. His research interests include wireless communications and industrial process control.