2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

An approximate solution to optimal L_p state estimation problems

A. Alessandri, C. Cervellera, A.F. Grassia, M. Sanguineti

Abstract— **We consider optimal estimation problems characterized by a state vector with i) dynamics described via a differential equation with Lipschitz nonlinearities, ii) partial information provided via a Lipschitz nonlinear mapping, and iii) an L_p norm measure of the estimation error to be minimized. An approximate solution of such optimal estimation problem is searched for by restricting the optimization to parameterized nonlinear approximators such as feedforward neural networks. The parameters of a feedforward neural network are the neural weights. This approach entails a constrained nonlinear programming problem, whose constraints are given by the dynamic and measurement equations, and the conditions guaranteeing the stability of the estimation error. To optimize the parameters values of neural networks an algorithm is developed that is based on appropriate sampling of the state and error spaces. Choices of the sample points are devised based on the notion of dispersion, which allow one to obtain an approximate solution of the optimal estimation problem by a small sample complexity.**

I. INTRODUCTION

State estimation problems are usually solved by means of filters and observers, which are treated separately in the literature, depending on the presence or not, respectively, of disturbances acting on the dynamic system.

A possible approach to the solution of the observer problem consists in applying a canonical state-space transformation [1], [2]. This representation of the system dynamics enables one to easily find an observer with linear error dynamics in the transformed state-space. Following different approaches, high-gain and variable-structure observers have been proposed in the literature. The high-gain observers have been considered, for example, in [3], where the convergence of the estimation error is obtained for Lipschitz nonlinear systems by performing a state transformation that allows to dominate the nonlinearities. The design of sliding-mode observers with nonlinear dynamics has been faced in [4].

In filtering, the disturbances that affect the system and/or the measurement equations are regarded either as unknown deterministic inputs or as stochastic random variables within a probabilistic framework. The most popular method to estimate the state of a noisy nonlinear systems is the extended Kalman filter; its convergence properties have been proven in [5]. However, more complex approaches can be devised, related to the conditional probability density function of the Markov process describing the system and measurement equations. Such a density function is conditioned by available on-line measures, and, as it is well known, provides the most complete description of the system state. In general the conditional density function can be determined only approximately: analytical solutions in the form of finite-dimensional filters are very difficult to obtain, except in few cases, e.g., for linear systems with Gaussian noises. For nonlinear systems, conditions for the design of finite-dimensional filters as generators of sufficient statistics can be found, among others, in [6], but a general design methodology to solve such a design problem exactly is not available. This motivates approximation-based approaches as, for instance, the one presented in [7], where the problem of estimation is addressed for nonlinear discrete-time systems by approximating the optimal innovation functions with a filter having a Kalman-like structure. A method is proposed in [8], [9] for particular kinds of discrete-time systems, where the innovation contribution is taken into account by means of polynomial expansions. In [10], a receding-horizon state estimation technique is presented that is based on the idea of minimizing a quadratic estimation cost function defined on a sliding window.

In this paper the estimation problem is addressed for continuous-time, nonlinear dynamic systems, in a framework of L_p signals. We are interested in an optimization-based approach, as the dynamics of the estimator is defined in order to minimize the L_p norm of the estimation error. After restricting the class of nonlinear systems by hypotheses on the system and measurement equations, we define a class of estimators with a certain structure, satisfying a stability requirement and in which an unknown function belonging to a certain smoothness class has to be determined. Then an optimal estimation problem is formulated, in which the unknown function in the estimator has to be determined in such a way to minimize the L_p norm of the estimation error.

An optimal solution to the estimation problem is searched for over families of functions taking on the structure of linear combinations of simple computational units dependent on some parameters [11], [12], [13], [14], [15]. In such a way, a nonlinear (generally) programming problem is obtained that consists in finding the parameters that minimize the L_p norm of the estimation error. This can accomplished under constraints that ensure the existence

Angelo Alessandri, Cristiano Cervellera and Aldo Filippo Grassia are with the Institute of Intelligent Systems for Automation, ISSIA-CNR National Research Council of Italy – Via De Marini 6, 16149 Genova, Italy - E-mail: {angelo, cervellera, filippo}@ge.issia.cnr.it

Marcello Sanguineti is with the Department of Communications, Computer, and System Sciences (DIST), University of Genoa – Via Opera Pia 13, 16145 Genova, Italy – E-mail: marcello@dist.unige.it

0-7803-9098-9/05/$25.00 ©2005 AACC

of a Lyapunov function for the estimation error. If certain smoothness properties of the Lyapunov function are satisfied, a suitable selection of the sampling points that cover in a sufficiently uniform way the state and error space may guarantee the fulfillment of such constraints in a convenient way. More specifically, typical measures of uniformity such as *dispersion* and *discrepancy*, commonly employed in the fields of statistics and number theoretic methods, allow to define favorable asymptotic conditions for the sampling procedure to ensure the convergence of the estimation error.

The paper is organized as follows. Section II reports preliminary stability results on the dynamics of the estimation error for the class of state estimators we have considered. The statement of an optimal state estimation problem is given in Section III. A method to find approximate solutions of this problem is presented in Section IV. The issues regarding the practical construction of the estimator are addressed in Section V. The proof of the various results are omitted for the sake of brevity.

Before concluding this section, let us briefly introduce the following notations. For $p \in [1, \infty)$ and a positive n, the space L_p^n consists of all Lebesgue-measurable functions $s : [0, \infty) \to \mathbb{R}^n$ such that $\int_0^\infty \|s(t)\|^p \, dt < \infty$. For every $p \in [1, \infty)$, L_p^n is a Banach space with the norm

$$\|s\|_p \triangleq \left(\int_0^\infty \|s(t)\|^p \, dt \right)^{1/p} .$$

The space L_∞^n is the set of all Lebesgue-measurable functions that are essentially bounded, i.e., such that $\operatorname{ess.sup}_{t \geq 0} \|s(t)\| < \infty$, where "ess. sup" denotes the essential supremum (i.e., supremum except on sets of measure zero). L_∞^n is a Banach space with the norm

$$\|s\|_\infty \triangleq \operatorname*{ess.sup}_{t \geq 0} \|s(t)\| .$$

To deal with possibly unbounded signals, the extension of L_p^n spaces is defined as follows. For $p \in [1, \infty]$, the extended space $L_{p\,e}^n$ is defined as $L_{p\,e}^n \triangleq \left\{ s | s_\tau \in L_p^n, \forall \, \tau \geq 0 \right\}$, where

$$s_\tau(t) \triangleq \begin{cases} s(t) & , \quad \text{if } \; t \leq \tau \, , \\ 0 & , \quad \text{if } \; t > \tau \, . \end{cases}$$

For every $s \in L_{p\,e}^n$ and $p \in [1, \infty]$, let $\|s\|_{p,\,\tau} \triangleq \|s_\tau\|_p$.

II. Assumptions and preliminary results

Let us begin with defining the class of estimation problems we consider. The dynamics of the vector to be estimated and the process of acquisition of information are modelled as:

$$\begin{cases} \dot{x} = f\,(x, w) \\ y = h\,(x, v) \end{cases} \quad , \quad t \geq 0 \tag{1}$$

where $x(t) \in X \subseteq \mathbb{R}^n$ is the vector to be estimated, $y(t) \in Y \subseteq \mathbb{R}^m$ is the vector of available information, and $w(t) \in W \subseteq \mathbb{R}^r$ and $v(t) \in V \subseteq \mathbb{R}^s$ are disturbances. In the following $\| \cdot \|$ denotes the Euclidean norm of its argument.

Assumption 2.1: Let $B_x \triangleq \{x \in X \,|\, \|x\| < \bar{x}, \; \bar{x} > 0\}$, $B_w \triangleq \{w \in W \,|\, \|w\| < \bar{w}\,, \;\; \bar{w} > 0\}$, and $B_v \triangleq \{v \in V \,|\, \|v\| < \bar{v}\,, \, \bar{v} > 0\}$. Then

(i) $f : X \times W \to \mathbb{R}^n$ is locally Lipschitz in $x \in B_x$ uniformly in $w \in B_w$. Moreover, there exist $L_f^x, L_f^w \in \mathbb{R}^+$ such that $\|f(x_1, w) - f(x_2, 0)\| \leq L_f^x \|x_1 - x_2\| + L_f^w \|w\|$, for all $x_1, x_2 \in B_x\,, w \in B_w$.

(ii) $h : X \times V \to \mathbb{R}^m$ is locally Lipschitz in $x \in B_x$, uniformly in $v \in B_v$. Moreover, there exist $L_h^x, L_h^v \in \mathbb{R}^+$ such that $\|h(x_1, v) - h(x_2, 0)\| \leq L_h^x \|x_1 - x_2\| + L_h^v \|v\|$, for all $x_1, x_2 \in B_x, \; v \in B_v$.

□

Assumption 2.1 (i) guarantees the existence and uniqueness of a local solution of the differential equation in (1), describing the dynamics of the vector x (see, for example, [16]).

We consider *full-order state estimators* with the structure

$$\dot{\hat{x}} = f\,(\hat{x}, 0) + g\,(y - h\,(\hat{x}, 0)) \quad , \quad t \geq 0 \, , \tag{2}$$

where $\hat{x}(t) \in \hat{X} \subseteq \mathbb{R}^n$ is the estimate of x and $Z \triangleq \left\{ z \in \mathbb{R}^m : z = y - h\,(\hat{x}, 0) \; , \, y \in Y \, , \, \hat{x} \in \hat{X} \right\} \subseteq \mathbb{R}^m$.

The estimator dynamics is the summation of a prediction term, given by the dynamics, and an innovation term, represented by the function $g : Z \to \mathbb{R}^n\,, Z \subseteq \mathbb{R}^m$; in the following, $y - h(\hat{x}, 0)$ and $g(\cdot)$ will be called *innovation* and *innovation function*, respectively. The innovation function is required to verify the following smoothness assumption.

Assumption 2.2: Let $z \triangleq y - h\,(\hat{x}, 0)$, and $B_z \triangleq \{z \in Z \,|\, \|z\| < \bar{z}\,, \, \bar{z} > 0\}$. Then $g : Z \to \mathbb{R}^n$ is such that $g(0) = 0$ and is locally Lipschitz in B_z. More specifically, there exists $L_g^z \in \mathbb{R}^+$ such that $\|g(z_1) - g(z_2)\| \leq L_g^z \|z_1 - z_2\|$ for all $z_1, z_2 \in B_z$.

□

Assumption 2.2 is a sufficient requirement for having a unique local solution of the differential equation (2) describing the estimator (see, e.g., [16]). The innovation function has been assumed to be of the above-written form for sake of simplicity, although it can be of a more general type [17], e.g., $\tilde{g}\,(\hat{x}, h(x, 0), h(\hat{x}, 0))$, where $\tilde{g}\,(\hat{x}, h(x, 0), h(\hat{x}, 0)) = 0$ if $h(x, 0) = h(\hat{x}, 0)$.

Let $e(t) \triangleq x(t) - \hat{x}(t)$ denote the estimation error; the error dynamics for the estimator is given by

$$\dot{e} = f(x, w) - f(\hat{x}, 0) - g\,(y - h\,(\hat{x}, 0)) \quad , \quad t \geq 0 \, . \tag{3}$$

Note that the condition $g(0) = 0$ guarantees that $e = 0$ is an equilibrium point for (3) in the absence of disturbances.

The existence of a suitable Lyapunov function for the error dynamics, together with Assumptions 2.1 and 2.2, guarantee that (2) is an asymptotic or an exponential estimator for (1) with no disturbances. These properties are sometimes referred to as *weak detectability* (see, e.g., [18]).

Let us now move to the L_p stability issue by assuming that the disturbances belong to an L_{pe} space of signals.

Theorem 2.3: Suppose that Assumptions 2.1 and 2.2 are verified and that there exist a Lyapunov function $V : E \to [0, \infty)$ with $E \supseteq B_e$ and positive constants c_1, c_2, c_3, and c_4 such that

(i) $c_1 \|e\|^2 \le V(e) \le c_2 \|e\|^2$;

(ii) $\dot{V}(x,e) = \dfrac{\partial V}{\partial e}\Big[f(x,0) - f(\hat{x},0) - g\left(h(x,0) - h(\hat{x},0)\right)\Big] \le -c_3 \|e\|^2$;

(iii) $\left\| \dfrac{\partial V}{\partial e} \right\| \le c_4 \|e\|$.

Then, for every $e(0)$ such that $\|e(0)\| < \bar{e}\sqrt{\frac{c_1}{c_2}}$ and for every $w \in L^r_{pe}$ and $v \in L^s_{pe}$ such that $\sup_{0\le\sigma\le\tau} \|w(\sigma)\| < \min\left(\bar{w}, \dfrac{c_1 c_3 \bar{e}}{2 c_2 c_4 L^x_f}\right)$ and $\sup_{0\le\sigma\le\tau} \|v(\sigma)\| < \min\left(\bar{v}, \dfrac{c_1 c_3 \bar{e}}{2 c_2 c_4 L^v_h L^z_g}\right)$, there exist nonnegative constants η, λ, and β such that

$$\|e\|_{p,\tau} \le \eta \|w\|_{p,\tau} + \lambda \|v\|_{p,\tau} + \beta \tag{4}$$

for all $\tau \in [0, \infty)$ with $\eta = \dfrac{c_2\, c_4 L^x_f}{c_1\, c_3}$, $\lambda = \dfrac{c_2\, c_4\, L^v_h\, L^z_g}{c_1\, c_3}$ and $\beta = \sqrt{\dfrac{c_1}{c_2}}\, \|e(0)\|\, \rho$, where

$$\rho = \begin{cases} 1 & , \quad \text{if } p = \infty \\ \left(\dfrac{2\, c_2}{c_3\, p}\right)^{1/p} & , \quad \text{if } p \in [1, \infty). \end{cases}$$

□

According to Theorem 2.3, if suitable Lipschitz conditions are verified by the functions f, h, and g, and if the disturbances belong to L_{pe}, then for every $\tau \ge 0$, in a small-signal context the L_p norm of the estimation error e_τ is bounded by a linear combination of the L_p norms of the disturbances w_τ and v_τ, and one term due to the initial uncertainty in the value of the vector to be estimated.

III. Statement and properties of the Optimal Estimation Problem (OEP)

We exploit the results of Section II to define a cost functional measuring the estimation error and to cast an optimal estimation problem in which such a functional has to be minimized over a certain class of admissible innovation functions. Recall that $B_z \triangleq \{z \in Z \,|\, \|z\| < \bar{z}\,,\ \bar{z} > 0\}$. We define the following set of functions.

$\mathcal{G} \triangleq \{\, g : Z \to \mathbb{R}^n$ such that

(i) *bounded in aggregate, i.e.,* $\exists\, L \in \mathbb{R}^+ : \forall g \in \mathcal{G}\ \sup_{z\in Z} \|g(z)\| \le L$;

(ii) Lipschitz in B_z, i.e., $\exists\, L_g \in \mathbb{R}^+ :\ \|g(z) - g(z')\| \le L_g \|z - z'\|$ for all $z, z' \in B_z\}$.

When endowed with the supremum norm, $\mathcal{G}$ is a subset of the normed space $\mathcal{C}(Z, \mathbb{R}^n)$ of continuous, n-valued functions on compact subsets of $\mathbb{R}^n \times \mathbb{R}^m$, equipped with the supremum norm. Conditions (i) and (ii) in the definition of the class $\mathcal{G}$ are related to the compactness of the set $\mathcal{G}$ in $\mathcal{C}(Z, \mathbb{R}^n)$, which plays a basic role in the following. Thus, from now on the state estimate dynamics is given by

$$\dot{\hat{x}} = f(\hat{x}, 0) + g\left(y - h(\hat{x}, 0)\right), \quad t \ge 0, \tag{5}$$

where the innovation function g belongs to $\mathcal{G}$ and is such that $g(0) = 0$.

The unknown innovation function g in an estimator having the structure specified in Assumption 2.2 can be determined by minimizing a performance index. A cost functional well-suited to optimization in an L_p framework is

$$J_{p,T} = \|e\|_{p,T}\,, \tag{6}$$

where $p \in [1, \infty]$ and $T > 0$. Thus, we focus on a cost function of the form (6) and we consider the following Optimal Estimation Problem (OEP, for short):

Problem OEP. Given $p \in [1, \infty]$ and $T > 0$, solve

$$\inf_{g\in\mathcal{G}} J_{p,T}(g)\,, \tag{7}$$

where $J_{p,T}(g) = \|x - \hat{x}\|_{p,T}$, $x,\ \hat{x} \in L^n_{pe}$, $w \in L^r_{pe}$, $v \in L^s_{pe}$, and

$$\begin{cases} \dot{x} = f(x, w) \\ y = h(x, v) \\ \dot{\hat{x}} = f(\hat{x}, 0) + g\left(y - h(\hat{x}, 0)\right). \end{cases} \tag{8}$$

□

The finiteness of the cost functional $J_{p,T}$ and the existence of a solution to OEP can be proven under suitable conditions. Now, basing on the results of Theorem 2.3, we need the following assumption.

Assumption 3.1: The system and measurement disturbances are such that $w \in L^r_{pe}$, $v \in L^s_{pe}$. Moreover, given $\tau > 0$ and the constants defined in Assumptions 2.1 and 2.2, $\sup_{0\le\sigma\le\tau} \|w(\sigma)\| < \min\left(\bar{w}, \dfrac{c_1 c_3 \bar{e}}{2 c_2 c_4 L^x_f}\right)$, $\sup_{0\le\sigma\le\tau} \|v(\sigma)\| < \min\left(\bar{v}, \dfrac{c_1 c_3 \bar{e}}{2 c_2 c_4 L^v_h L^z_g}\right)$, and $\|e(0)\| < \bar{e}\sqrt{\frac{c_1}{c_2}}$.

□

Generally speaking, for a given $g \in \mathcal{G}$, the existence of a unique local solution of the differential equation (8) describing the estimator is guaranteed by the regularity hypotheses on f, g, and h. Assumption 3.1 is quite technical and not particularly restrictive, as it gives conditions on the existence of the solution of the differential equation for small signals (see, for example, [16]).

Due to the very general hypotheses made on the functions f, g, and h, and on the disturbances, finding an analytical solution of Problem OEP is a very difficult task. For this reason, in Section IV we shall propose a methodology of approximate solution that reduces the functional optimization Problem OEP to a sequence of nonlinear (in general) programming problems.

IV. Suboptimal solutions to OEP via nonlinear programming problem

A. Parameterized estimators

OEP, due to its general formulation, cannot be solved analytically. So we search for suboptimal solutions in the following way. We consider the class of *parameterized estimators* defined as:

$$\dot{\hat{x}}= f(\hat{x},0) + \tilde{g}\,(w, y - h\,(\hat{x},0)) \quad , \quad t \geq 0\,, \qquad (9)$$

where $w \in W \subseteq \mathbb{R}^q$ is a parameter vector and $\tilde{g} : W \times Z \to \mathbb{R}^n$ is a parameterized innovation function that is required to verify Assumption 2.2 for every $w \in W$. (9) defines a class of estimators, dependent on the choice of a type of innovation function, which, in turn, depends on a vector of "free" parameters, i.e., parameters that can be chosen according to some optimality criterion.

For any positive integer ν, we define the following family of parameterized functions.

Definition 4.1:
$A_\nu \stackrel{\triangle}{=} \{\ \gamma_\nu : K \times \mathbb{R}^l \to \mathbb{R}^n\ ,\ K \text{ compact}\,,\ \text{such that}$

(i) $\gamma_{\nu\,j}(\xi, \omega_{\nu_j}) = \sum_{i=1}^{\nu} c_{ij}\,\varphi_i(\xi,\kappa_i)$, $\varphi_i : K \times \mathbb{R}^l \to \mathbb{R}$, $|c_{ij}| \leq C$, $C \in \mathbb{R}^+$, $\kappa_i \in \mathbb{R}^l$, $i = 1,\ldots,\nu$, $j = 1,\ldots,n$, $\omega_{\nu\,j} \stackrel{\triangle}{=} \text{col}(c_{ij},\kappa_i : i = 1,\ldots,\nu)$;

(ii) the functions $\varphi_i(\cdot,\kappa_i)$ are bounded in aggregate, i.e., $\exists\, M \in \mathbb{R}^+$ such that $\forall i = 1,\ldots,\nu, \forall \kappa_i \in \mathbb{R}^l$, $\sup_{\xi \in K} |\varphi_i(\xi,\kappa_i)| \leq M$;

(iii) the functions $\varphi_i(\cdot,\kappa_i)$ are equicontinuous, i.e.,$\forall\,\epsilon > 0\ \exists\,\delta_\epsilon > 0$ such that $\forall\, i = 1,\ldots,\nu$, $\forall\,\kappa_i,\kappa_i' \in \mathbb{R}^l$, if $\|\xi - \xi'\| < \delta_\epsilon$ then $|\varphi_i(\xi,\kappa_i) - \varphi_i(\xi',\kappa_i')| \leq \epsilon$;

(iv) the functions $\varphi_i(\cdot,\kappa_i)$ are Lipschitz, i.e., $\forall\, i = 1,\ldots,\nu\ \exists\, L_i \in \mathbb{R}^+$ such that $\forall \kappa_i \in \mathbb{R}^l$, $|\varphi_i(\xi,\kappa_i) - \varphi_i(\xi',\kappa_i)| \leq L_i|\xi - \xi'|$;

(v) $\bigcup_{\nu=1}^{\infty} A_\nu$ is dense in $\mathcal{G}$ with respect to the supremum $\}$.

□

Due to requirements (ii), (iii), and (iv) in the above definition, for every positive integer ν the set A_ν is compact in $\mathcal{C}(K,\mathbb{R}^n)$ (see the proof of Theorem 4.3). The compactness property of A_ν plays a basic role in the following.

Functions in A_ν are linear combinations with coefficients of ν basis functions, and they are bounded in aggregate and Lipschitz; thus, the following proposition holds.

Proposition 4.2: For every integer ν, the elements of A_ν are admissible innovation functions.

□

Feedforward neural networks of the perceptron type, with at most ν hidden units and bounded parameters, and Radial-basis-functions with at most ν hidden units and bounded parameters are examples of widely used sets A_ν.

B. A nonlinear programming problem approximating OEP

To avoid burdening the notation, in the following we omit the dependence of J on p and T and we write merely J instead of $J_{p,T}$. Let ν be a positive integer. We introduce the following problem.

Problem OEP$_\nu$. Given $p \in [1,\infty]$ and $T > 0$, solve

$$\inf_{\gamma_\nu \in A_\nu} J(\gamma_\nu), \qquad (10)$$

where $J(\gamma_\nu) = \|x - \hat{x}\|_{p,T},\ x,\ \hat{x} \in L^n_{p\,e}, w \in L^r_{p\,e},\ v \in L^s_{p\,e}$, and

$$\begin{cases} \dot{x}= f\,(x,w) \\ y = h\,(x,v) \\ \dot{\hat{x}}_\nu= f(\hat{x}_\nu,0) + \gamma_\nu\,(\omega_\nu, y - h\,(\hat{x}_\nu,0))\,. \end{cases} \qquad (11)$$

□

Theorem 4.3: If Assumption 3.1 is satisfied, and the hypotheses of Theorem 2.3 hold, then for every $p \in [1,\infty]$, $T > 0$, and every positive integer ν there exists $\gamma_\nu^\circ \in A_\nu$ such that

$$J(\gamma_\nu^\circ) \stackrel{\triangle}{=} J_\nu^\circ = \min_{\gamma_\nu \in A_\nu} J(\gamma_\nu).$$

□

As each A_ν is a set of parameterized functions with a fixed structure, the minimization has to be performed with respect to the finite-dimensional vector of parameters $\omega_\nu \in \mathbb{R}^{\mathcal{N}(\nu)}$, whereas OEP entails an infinite-dimensional minimization.
This turns out to be evident by substituting γ_ν into the differential equation of the estimator and then into J. The cost functional is a function of the parameter vector ω_ν. With a little notational abuse, we denote such a function by $J_\nu(\omega_\nu)$. Thus, for each positive integer ν the minimization (7) with respect to the infinite-dimensional set $\mathcal{G}$ is replaced by the minimization with respect to the finite-dimensional vector $\omega_\nu \in \mathbb{R}^{\mathcal{N}(\nu)}$. Hence we can define the following.

Problem OEP$'_\nu$. Given $T > 0$, find

$$\inf_{\omega_\nu \in \mathbb{R}^{\mathcal{N}(\nu)}} J_\nu(\omega_\nu)\,. \qquad (12)$$

where $J_\nu(\omega_\nu) = \|x - \hat{x}\|_{p,T},\ x,\ \hat{x} \in L^n_{p\,e}, w \in L^r_{p\,e},\ v \in L^s_{p\,e}$, and

$$\begin{cases} \dot{x}= f\,(x,w) \\ y = h\,(x,v) \\ \dot{\hat{x}}_\nu= f(\hat{x}_\nu,0) + \gamma_\nu\,(\omega_\nu, y - h\,(\hat{x}_\nu,0))\,. \end{cases} \qquad (13)$$

□

It is worth noting that the solution of this last problem may not be unique, even if there exists a unique minimum of OEP$_\nu$, as it might happen that there is no one-to-one correspondence between a vector $\omega_\nu \in \mathbb{R}^{\mathcal{N}(\nu)}$ and an element $\gamma_\nu \in A_\nu$.

As OEP$'_\nu$ is a nonlinear programming problem, it can be solved determining the optimal parameters vector w^o_ν by a suitable descent algorithm. Typically used algorithms are those based on gradient descent with stochastic perturbations [19, pp. 38-40, 103-104], genetic algorithms [20], simulated annealing [21], global stochastic optimization based on Monte Carlo [22] or quasi-Monte Carlo [23, Chapter 4] methods, etc. When the basis functions in A_ν are functions computable by neural-network computational units, various standard learning algorithms can be applied (see, e.g., [24], [25], [26], [27] and the references therein).

C. On the convergence of suboptimal solutions

As concerns the convergence of J°_ν to J°, the following proposition holds.

Proposition 4.4: The sequence $\{\gamma^\circ_\nu\}_{\nu=1}^\infty$ of optimal solutions to the sequence of nonlinear programmming problems $\{\text{OEP}_\nu\}_{\nu=1}^\infty$ is a minimizing sequence for OEP, i.e., $\lim\limits_{\nu\to\infty} J^\circ_\nu = J^\circ$.

□

As regards the convergence of a sequence $\{\gamma^\circ_\nu\}_{\nu=1}^\infty$ of optimal solutions of Problem OEP$_\nu$ to an optimal solution γ° of OEP, that depends heavily on the properties of the functional to be minimized, and is an open problem in most applications of this kind of methods (see, for example, [28] and [29]).

V. Solution of the approximating problems

A. Stability issues

When optimizing the parameters vector $\omega_\nu \in \mathbb{R}^{\mathcal{N}(\nu)}$, we must take into account that the corresponding estimator has to be stable.

The dynamics of the estimation error of estimator (9) for the system (1) without disturbances (i.e., with $w = 0$ and $v = 0$) is given by

$$\dot{e} = f(x,0) - f(\hat{x},0) - \tilde{g}\left(\omega, y - h\left(t,\hat{x},0\right)\right) , \quad t \geq 0 \, . \tag{14}$$

In the following, we shall assume that X is a compact set, and consider a compact set $\bar{E}$ that belongs to E; if $E = \mathbb{R}^n$, then $\bar{E}$ can be an arbitrarily large compact set.

Given the compact set $S = X \times \bar{E}$, let us denote by S_M a set of M sample points $s_i = \text{col}(x_i, e_i) \stackrel{\triangle}{=} (x_i^{\mathcal{T}}, e_i^{\mathcal{T}})^{\mathcal{T}}$, $i = 1, 2, \ldots, M$, that belong to S. Let us define the dispersion of S_M (see [30]) as

$$\theta(S_M) \stackrel{\triangle}{=} \sup_{s\in S} \min_{1\leq i\leq M} \|s - s_i\| \, .$$

Consider a Lyapunov function $V : \bar{E} \to [0, \infty)$, which is assumed to be continuously differentiable with respect to time. For estimator (9), if we consider the equality $\hat{x} = x - e$, the time derivative of the Lyapunov function depends on both s (i.e., x and e) and the parameters vector ω.

We have the following stability result (see [31]).

Theorem 5.1: Suppose that Assumptions 2.1 and 2.2 are verified and that there exists a Lyapunov function $V : \bar{E} \to [0, \infty)$ with time derivative that is Lipschitz with respect to $s = \text{col}(x, e) \in S$ uniformly in $\omega \in \mathbb{R}^{\mathcal{N}(\nu)}$ such that Assumption 3.1 is verified and

(i) $c_1 \|e\|^2 \leq V(e) \leq c_2 \|e\|^2, \quad$ for every $e \in \bar{E}$,

(ii) $\left\| \dfrac{\partial V}{\partial e} \right\| \leq c_3 \|e\| \quad$ for every $e \in \bar{E}$,

where $c_1, c_2, c_3 > 0$. Let $\bar{\omega} \in \mathbb{R}^{\mathcal{N}(\nu)}$ be a parameter vector and $S_M \subseteq S$ be a set of M sample points such that

(iii) $\dot{V}\,(s_i, \bar{\omega}) \leq -c_4 \|e_i\|^2 \quad$ for $s_i \in S_M$, $i = 1, 2, \ldots, M$,

where $c_4 > 0$, and let L_F be the Lipschitz constant of the function $F(s, \bar{\omega}) \stackrel{\triangle}{=} \dot{V}\,(s, \bar{w}) + c_1 \|e\|^l$ with respect to s. Then there exists $\varepsilon_M > 0$ such that, if

(iv) $\theta(S_M) < \dfrac{\varepsilon_M}{L_F}$,

for every $e(0) \in \bar{E}$, the estimator (9) provides an asymptotically stable estimation error for the system (1).

□

B. Choice of the discretization points

Condition (iv) of Theorem 5.1 requires that the M points of the discretization of S are "close enough" to each other, and spread in the most uniform way, without leaving regions of the space "undersampled."

Various methods have been developed for the generation of well uniformly scattered *deterministic* sequences, such as the *Good Lattice Points sequences*, the *Niederreiter sequence*, the *Halton sequence*, the *Hammersley sequence* and the *Sobol' sequence* [30], [32]. Their construction varies from mehod to method. In [32] a common framework for the construction of (t, n)-*sequences*, which generalize many of the aforementioned techniques, sometimes called *low-discrepancy sequences*, is presented, together with the most relevant theoretical properties.

In particular, it can be shown that it is possible to construct (t, n)-sequences that satisfy deterministically

$$\theta(S_M) \leq O(\sqrt{2n} M^{-1/2n}),$$

and thus ensure the convergence of the estimation error according to Theorem 5.1.

The use of low-discrepancy sequences for function learning by neural networks is described in [33].

C. A nonlinear programming algorithm for the design of the the optimal estimator

On the basis of Theorem 5.1, the design of the optimal estimator can be designed via the following algorithm.

1) Choose a time horizon T, an L_p measure for the estimation error, and compact sets $X, \bar{E} \subset \mathbb{R}^n$.

2) Choose a composite model A_ν with ν basis functions φ; the admissible innovation functions have to belong to A_ν.

3) Choose a set $S_M \subset X \times \bar{E}$ of M sample points $s_i \triangleq \mathrm{col}(x_i, e_i)$, $i = 1, \ldots, M$ belonging to a low-discrepancy sequence.

4) Given a Lyapunov function $V : \bar{E} \mapsto [0, \infty)$, find $\bar{\omega}_\nu \in \mathbb{R}^{\mathcal{N}(\nu)}$ such that

4.1) $J_\nu(\bar{\omega}_\nu) \triangleq J_\nu(\bar{\gamma}_\nu) = \min_{\omega_\nu \in \mathbb{R}^{\mathcal{N}(\nu)}} \|e(\omega_\nu)\|_{p,T}$;

4.2) $c_1 \|e\|^2 \le V(e) \le c_2 \|e\|^2$ for every $e \in \bar{E}$;

4.3) $\left\| \frac{\partial V}{\partial e} \right\| \le c_3 \|e\|$ for every $e \in \bar{E}$, where $c_1, c_2, c_3 > 0$;

4.4) $\dot{V}\ (s_i, \bar{\omega}) \le -c_4 \|e_i\|^2$ for $s_i \in S_M$, $i = 1, 2, \ldots, M$, where $c_4 > 0$;

4.5) $\theta(S_M) < \frac{\varepsilon_M}{L_F}$, where L_F be the Lipschitz constant of $F(s, \bar{\omega}) \triangleq \dot{V}\ (s, \bar{\omega}) + c_1 \|e\|^l$ with respect to s.

When the composite model A_ν corresponds to commonly used neural networks, Step 4), in which the parameters of the innovation function γ_ν have to be optimized, can be viewed as a particular kind of neural-network training, in order to satisfy the constraints (i)–(iv) of Theorem 5.1. This is a supervised learning that differs from standard techniques used to optimize the parameters of approximating networks corresponding to commonly used neural networks; such algorithms minimize the distance from given target values. In contrast to them, to minimize J_ν and satisfy the constraints such as those corresponding to (i)–(iv) in Theorem 5.1 one has to employ ad-hoc techniques, sometimes referred to as *distal training* [34], which often are modified versions of standard minimization algorithms. For example, in [34] the satisfaction of constraints to design a neural controller is obtained by minimizing a suitable quadratic penalty function using a specialized version of the Levenberg-Marquardt algorithm. A possible approach for the solution of the problem at step 4) is proposed in [31] and will be the subject of future researches.

References

[1] A. J. Krener and A. Isidori, "Linearization by output injection and nonlinear observers", *Systems & Control Letters*, vol. 3, pp. 47–52, 1983.

[2] A. J. Krener and W. Respondek, "Nonlinear observer with linearizable error dynamics", *SIAM J. on Control and Optimization*, vol. 23, no. 2, pp. 197–216, 1985.

[3] J. P. Gauthier, H. Hammouri, and S. Othman, "A simple observer for nonlinear systems applications to bioreactors", *IEEE Trans. on Automatic Control*, vol. 37, no. 6, pp. 875–880, 1992.

[4] B. L. Walcott and S. H. Zak, "State observation of nonlinear uncertain dynamical systems", *IEEE Trans. on Automatic Control*, vol. 32, no. 2, pp. 166–170, 1987.

[5] K. Reif, S. Günter, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter", *IEEE Trans. on Automatic Control*, vol. 44, no. 4, pp. 714–728, 1999.

[6] M. Kouritzin, "On exact filters for continuous signals with discrete observations", *IEEE Trans. on Automatic Control*, vol. 43, no. 5, pp. 709–715, 1998.

[7] T. Parisini and R. Zoppoli, "Neural networks for nonlinear state estimation", *Int. J. of Robust and Nonlinear Control*, vol. 4, pp. 231–248, 1994.

[8] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering for linear discrete time non-gaussian systems", *SIAM J. on Control and Optimization*, vol. 34, no. 5, pp. 1666–1690, 1996.

[9] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering of discrete-time linear systems with multiplicative state noise", *IEEE Trans. on Automatic Control*, vol. 42, no. 8, pp. 1106–1126, 1997.

[10] A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli, "A neural state estimator with bounded errors for nonlinear systems", *IEEE Trans. on Automatic Control*, vol. 44, no. 11, pp. 2028–2042, 1999.

[11] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function", *IEEE Trans. on Inf. Theory*, vol. 39, pp. 930–945, 1993.

[12] L. Breiman, "Hinging hyperplanes for regression, classification and function approximation", *IEEE Trans. Inform. Theory*, vol. 39:3, pp. 999–1013, 1993.

[13] V. Kůrková and M. Sanguineti, "Comparison of worst-case errors in linear and neural network approximation", *IEEE Trans. on Inform. Theory*, vol. 48, no. 1, pp. 265–275, 2002.

[14] P. C. Kainen, V. Kůrková, and M. Sanguineti, "Minimization of error functionals over variable-basis functions", *SIAM J. on Optimization*, vol. 14, no. 3, pp. 732742, 2004.

[15] V. Kůrková and M. Sanguineti, "Error estimates for approximate optimization by the extended Ritz method", *SIAM J. on Optimization*, to appear.

[16] H. K. Khalil, *Nonlinear Systems*, Prentice Hall, Upper Saddle River, NJ, 1996.

[17] X. Xia and W. Gao, "On exponential observers for nonlinear systems", *Systems & Control Letters*, vol. 11, no. 4, pp. 319–325, 1988.

[18] M. Vidyasagar, "On the stabilization of nonlinear systems using state detection", *IEEE Trans. on Automatic Control*, vol. 25, no. 3, pp. 504–509, 1980.

[19] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.

[20] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[21] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*, John Wiley & Sons, 1989.

[22] G. Yin, "Rates of convergence for a class of global stochastic optimization algorithms", *SIAM J. on Optimization*, vol. 10, pp. 99–120, 1999.

[23] J. F. Traub and A. G. Werschulz, *Complexity and Information*, Cambridge University Press, Cambridge, 2000.

[24] A. Alessandri, M. Sanguineti, and M. Maggiore, "Optimization-based learning with bounded error for feedforward neural networks", *IEEE Trans. on Neural Networks*, vol. 13, pp. 261–273, 2002.

[25] D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems", *SIAM J. on Optimization*, vol. 7, pp. 913–926, 1997.

[26] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.

[27] P. Tseng, "Incremental gradient(-projection) method with momentum term and adaptive stepsize rule", *SIAM J. on Optimization*, vol. 8, pp. 506–531, 1998.

[28] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Prentice Hall, Englewood Cliffs, N. J., 1963.

[29] J. W. Daniel, *The Approximate Minimization of Functionals*, Prentice-Hall, Englewood Cliffs, N. J., 1971.

[30] K.-T. Fang and Y. Wang, *Number-theoretic Methods in Statistics*, Chapman & Hall, London, 1994.

[31] A. Alessandri, C. Cervellera, A. F. Grassia, and M. Sanguineti, "Design of observers for continuous-time nonlinear systems using neural networks", in *Proc. American Control Conference*, Boston, MA, 2004, pp. 2433–2438.

[32] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

[33] C. Cervellera and M. Muselli, "Deterministic design for neural network learning: an approach based on discrepancy", *IEEE Trans. on Neural Networks*, vol. 15, no. 3, pp. 533–544, 2004.

[34] S. Yu and A. M. Annaswamy, "Stable neural controllers for nonlinear dynamic systems", *Automatica*, vol. 34, no. 5, pp. 641–650, 1998.