TWCCC ⋆ Texas – Wisconsin – California Control Consortium

Technical report number 2013-01

Robust stability of moving horizon estimation under bounded disturbances

Luo Ji*
Dept. Chem. & Biol. Eng.
Univ. of Wisconsin. (USA)

James B. Rawlings†
Dept. Chem. & Biol. Eng.
Univ. of Wisconsin. (USA)

Wuhua Hu‡
INFINITUS
Nanyang Tech. Univ. (Singapore)

Andrew Wynn§
Dept. of Aeron.
Imp. Coll. (United Kingdom)

Moritz Diehl¶
ESAT-SCD & OPTEC (Belgium)

November 14, 2013

[1]

Abstract

The paper proposes a new form of nonlinear state estimator, for which we can establish robust global asymptotic stability (RGAS) in the case of bounded disturbances. In this estimator, a max term is added to the usual sum of stage costs, and one additional assumption is made relating the initial state stage cost to the system's detectability condition. Two simulation examples are presented to illustrate the estimator's performance. Two future problems are discussed: (i) the proof of estimator convergence for convergent disturbances and (ii) changing from full information estimation to moving horizon estimation (MHE), which has a smaller and tractable online computational complexity.

*Email: lji2@wisc.edu
†Email: rawlings@engr.wisc.edu. Author to whom all correspondence should be addressed.
‡Email: hwh@ntu.edu.sg
§Email: a.wynn@imperial.ac.uk
¶Email: moritz.diehl@esat.kuleuven.be

[1]This work was supported by the industrial members of the Texas Wisconsin California Control Consortium (TWCCC). Corresponding author: J.B. Rawlings. E-mail: rawlings@engr.wisc.edu. Telephone: (608) 263-5859. Fax: (608) 265-8794.

Keywords
Moving horizon estimation, nonlinear state estimation, constraints,bounded disturbances, incremental input/output to state stability,robust global asymptotic stability

1 Introduction

Moving horizon estimation (MHE) is an online optimization-based state estimation method that can handle nonlinear systems and satisfy constraints on the estimated states and disturbances. It has been established that MHE can provide a robustly stable estimator in the case of convergent process and measurement disturbances [5, Theorem 12]. A remaining issue is whether the same property can be established for bounded (rather than convergent) disturbance [5, Conjecture 13]. There is some recent research on the issue, which assumes an already-known deterministic estimator [4]. In this paper, we work on a more general case. We show here that [5, Conjecture 13] is true if one adds to the usual full information cost function a max stage cost term, and makes one additional assumption linking the chosen estimator stage cost to the detectability condition of the nonlinear system.

The paper is organized as follows. For the paper to be reasonably self contained, we provide a short summary of $\mathcal{K}$ and $\mathcal{KL}$ functions, define notation and introduce the standard and modified state estimation cost functions. Most of this is taken from [5]. We then state the chosen definition for nonlinear detectability, and define robust global asymptotic stability (RGAS) of a state estimator. Then we prove the main result of the paper: the full information estimator with the modified cost function is RGAS for a detectable nonlinear system subject to bounded disturbances. Two simulation examples are then presented to compare the performances of the state estimators using the conventional and the modified cost functions.

2 Notation and Properties of $\mathcal{K}$ and $\mathcal{KL}$ functions

The symbols $\mathbb{I}_{\geq 0}$ and $\mathbb{R}_{\geq 0}$ denote the sets of nonnegative integers and reals, respectively. The symbol $\mathbb{I}_{0:N-1}$ denotes the set $\{0, 1, \dots, N-1\}$. The symbol $|\cdot|$ denotes the Euclidean norm. The bold symbol $\mathbf{x}$, denote a sequence of a vector-valued variable x, $\{x(0), x(1), \dots\}$. The notation $\|\mathbf{x}\|$ is the sup norm over a sequence, $\sup_{i\geq 0} |x(i)|$, and $\|\mathbf{x}\|_{a:b}$ denotes $\max_{a\leq i\leq b} |x(i)|$. The definition of system detectability and statements and proofs of estimator stability are significantly streamlined using the properties of $\mathcal{K}$ and $\mathcal{KL}$ functions, so we provide a brief summary here. The interested reader may also want to consult [3, pp. 144–147] and [6, Appendix B] for further discussion.

Definition 1 ($\mathcal{K}$, $\mathcal{K}_\infty$, and $\mathcal{KL}$ functions)**.** *A function $\sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ belongs to class $\mathcal{K}$ if it is continuous, zero at zero, and strictly increasing; $\sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ belongs to class $\mathcal{K}_\infty$ if it is a class $\mathcal{K}$ and unbounded ($\sigma(s) \to \infty$ as $s \to \infty$). A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{I}_{\geq 0} \to \mathbb{R}_{\geq 0}$ belongs to class $\mathcal{KL}$ if it is continuous and if, for each $t \in \mathbb{I}_{\geq 0}$, $\beta(\cdot, t)$ is a class $\mathcal{K}$ function and for each $r \geq 0$, $\beta(r, \cdot)$ is nonincreasing and satisfies* $\lim_{t\to\infty} \beta(r, t) = 0$.

The following are useful properties of these functions. Most of these are established in [3, Lemma 4.2]: if $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are $\mathcal{K}$ functions ($\mathcal{K}_\infty$ functions), then $\alpha_1^{-1}(\cdot)$, $(\alpha_1 + \alpha_2)(\cdot)$, and $(\alpha_1 \circ \alpha_2)(\cdot)$[2] are $\mathcal{K}$ functions ($\mathcal{K}_\infty$ functions). Moreover, if $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are $\mathcal{K}$ functions and $\beta(\cdot)$ is a $\mathcal{KL}$ function, then $\sigma(r,t) = \alpha_1(\beta(\alpha_2(r), t))$ is a $\mathcal{KL}$ function.

We require the following basic inequalities to streamline our presentation. Proofs of these properties are given in [5].

1. For $\gamma(\cdot) \in \mathcal{K}$, the following holds for all $a_i \in \mathbb{R}_{\geq 0}$, $i \in \mathbb{I}_{1:n}$

$$\begin{aligned} \gamma\big(a_1 + a_2 + \cdots + a_n\big) &\leq \gamma(na_1) + \gamma(na_2) + \cdots + \gamma(na_n) \\ \gamma\big(a_1 + a_2 + \cdots + a_n\big) &\geq \frac{1}{n}\big(\gamma(a_1) + \gamma(a_2) + \cdots + \gamma(a_n)\big) \end{aligned} \tag{1}$$

2. Similarly, for $\beta(\cdot) \in \mathcal{KL}$ the following holds for all $a_i \in \mathbb{R}_{\geq 0}$, $i \in \mathbb{I}_{1:n}$, and all $t \in \mathbb{R}_{\geq 0}$

$$\begin{aligned} \beta\big((a_1 + a_2 + \cdots + a_n), t\big) &\leq \beta(na_1, t) + \beta(na_2, t) + \cdots + \beta(na_n, t) \\ \beta\big((a_1 + a_2 + \cdots + a_n), t\big) &\geq \frac{1}{n}\big(\beta(a_1, t) + \beta(a_2, t) + \cdots + \beta(a_n, t)\big) \end{aligned} \tag{2}$$

3 Basic definitions and assumptions

We assume that the system generating the measurements is given by the standard discrete time, nonlinear system

$$\begin{aligned} x^+ &= f(x, w) \\ y &= h(x) + v \end{aligned} \tag{3}$$

The state of the systems is $x \in \mathbb{R}^n$, the measurement is $y \in \mathbb{R}^p$, and the notation x^+ means x at the next sample time. A control input u may be included in the model, but it is considered a known variable, and its inclusion is irrelevant to state estimation, so we suppress it in the model under consideration here. We receive a measurement y from the sensor, but the process disturbance, $w \in \mathbb{R}^g$, measurement disturbance, $v \in \mathbb{R}^p$, and system initial state, $x(0)$, are considered unknown variables. These are often modeled as independent, random variables with stationary probability densities in stochastic estimation theory, but we will avoid random variables in this discussion and consider completely deterministic systems. We therefore model $w, v, x(0)$ as unknown, but **bounded** disturbance variables. After this choice, we cannot speak about the statistical properties of the estimator, but we can discuss estimator stability, rate of convergence, and the sensitivity of the estimate error to the disturbances. We assume throughout that functions $f : \mathbb{R}^n \times \mathbb{R}^g \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuous.

[2] $(\alpha_1 \circ \alpha_2)(\cdot)$ is the composition of the two functions $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$, defined by $(\alpha_1 \circ \alpha_2)(s) := \alpha_1(\alpha_2(s))$.

Full information estimation

The most accessible theory that we can present is the theory of full information estimation. Full information estimation also has the best theoretical properties in terms of stability and optimality. Unfortunately, it is also computationally intractable except for the simplest cases, such as a linear system model. Its value therefore lies in clearly defining what is *desirable* in a state estimator. One method for practical estimator design therefore is to come as close as possible to the properties of full information estimation while maintaining a tractable online computation. This design philosophy leads directly to moving horizon estimation (MHE).

First we define some notation necessary to distinguish the system variables from the estimator variables. We have already introduced the system variables (x, w, y, v). In the estimator optimization problem, these have corresponding decision variables, which we denote $(\chi, \omega, \eta, \nu)$. The *optimal* decision variables are denoted $(\hat{x}, \hat{w}, \hat{y}, \hat{v})$ and these optimal decisions are the estimates provided by the state estimator. The relationships between these variables are

$$
\begin{aligned}
x^+ &= f(x, w) \qquad & y &= h(x) + v \\
\chi^+ &= f(\chi, \omega) \qquad & y &= h(\chi) + \nu \\
\hat{x}^+ &= f(\hat{x}, \hat{w}) \qquad & y &= h(\hat{x}) + \hat{v}
\end{aligned}
$$

Notice that it is always the system measurement y that appears in the second column of equations.

Definition 2 (Bounded sequences; set $\mathbb{B}$)**.** *A sequence $w(k)$, $k \geq 0$ is bounded if $\|\mathbf{w}\|$ is finite. The set of bounded sequences is denoted by $\mathbb{B}$.*

We begin with a reasonably general definition of the full information estimator that produces an estimator that is *stable*. The full information objective function is traditionally defined for $T \geq 1$ as

$$
V_T^{\text{sum}}(\chi(0), \boldsymbol{\omega}) = \ell_x(\chi(0) - \overline{x}_0) + \sum_{i=0}^{T-1} \ell_i(\omega(i), \nu(i)) \tag{4}
$$

subject to

$$
\chi^+ = f(\chi, \omega) \qquad y = h(\chi) + \nu
$$

in which T is the current time, $y(i)$ is the measurement at time i, and $\overline{x}_0$ is a user-defined prior value of the initial state. Because $\nu = y - h(\chi)$ is the error in fitting the measurement y, $\ell_i(\omega, \nu)$ costs the model disturbance and the fitting error. These are the two error sources that we reconcile in all state estimation problems.

Next we define a modified objective function, such that we can establish stability properties under bounded disturbances rather than convergent disturbances

$$V_T(\chi(0), \boldsymbol{\omega}) = \frac{1}{T}(1+\delta)\ell_x(\chi(0) - \overline{x}_0) + \frac{1}{T}\sum_{i=0}^{T-1} \ell_i(\omega(i), \nu(i)) + \delta \max_{i \in 0:T-1} \ell_i(\omega(i), \nu(i)) \quad (5)$$

subject to

$$\chi^+ = f(\chi, \omega) \qquad y = h(\chi) + \nu$$

in which δ is a scalar weighting parameter chosen by the user. This modified objective could be viewed as the conventional objective plus a weighted maximum of ℓ_i; note that when $\delta = 0$, the estimator is equivalent to the standard estimation given in (4), because for a specific T, it does not change the optimization result of (4) if we divide V_T^{sum} by T. On the other hand, we could suppress the sum of ℓ_i term by letting $\delta \to \infty$

$$V_T^{\max}(\chi(0), \boldsymbol{\omega}) = \frac{1}{T}\ell_x(\chi(0) - \overline{x}_0) + \max_{i \in 0:T-1} \ell_i(\omega(i), \nu(i))$$

subject to

$$\chi^+ = f(\chi, \omega) \qquad y = h(\chi) + \nu$$

The full information estimator is defined as the solution to

$$\min_{\chi(0), \boldsymbol{\omega}} V_T(\chi(0), \boldsymbol{\omega}) \quad (6)$$

or using V_T^{sum} or $V_T^{\max}$ instead of V_T. For simplicity, we call them the MIX, SUM or MAX estimator according to the objectives (5), (4) or (3). Using the minimization results plus the model we can solve for all $\hat{x}(i|T)$, $i = 0, \cdots, T$, which is the (smoothed) estimated trajectory for a given T. For each T we take the last result $\hat{x}(T|T)$ as the current state estimate, which is usually the estimate passed to a controller.

Assumption 3 (Positive definite stage cost)**.** *The stage costs are continuous functions and satisfy the following inequalities for all* $w \in \mathbb{R}^g$*, and* $v \in \mathbb{R}^p$

$$\underline{\gamma}_x(|x|) \le \ell_x(x) \le \overline{\gamma}_x(|x|) \qquad i \ge 0 \quad (7)$$

$$\underline{\gamma}_w(|w|) + \underline{\gamma}_v(|v|) \le \ell_i(w, v) \le \overline{\gamma}_w(|w|) + \overline{\gamma}_v(|v|) \qquad i \ge 0 \quad (8)$$

in which $\underline{\gamma}_x, \overline{\gamma}_x, \underline{\gamma}_w, \overline{\gamma}_w, \underline{\gamma}_v, \overline{\gamma}_v \in \mathcal{K}_\infty$.

Remark 4. *From Assumption 3 the following holds for all* $\mathbf{w}, \mathbf{v} \in \mathbb{B}$

$$\max_{i \in 0:\infty} \ell_i(w(i), v(i)) \le \overline{\gamma}_w(\|\mathbf{w}\|) + \overline{\gamma}_v(\|\mathbf{v}\|)$$

Remark 5. *Because the stage cost satisfies*

$$\frac{1}{T}\sum_{i=0}^{T-1} \ell_i(\omega(i), \nu(i)) \le \max_{i \in 0:T-1} \ell_i(\omega(i), \nu(i)) \le \sum_{i=0}^{T-1} \ell_i(\omega(i), \nu(i))$$

it is straightforward to show that the two objectives satisfy for all $T \geq 1$,

$$\frac{1}{T}(1+\delta)V_T^{\text{sum}} \leq V_T \leq (1+\delta)V_T^{\text{sum}}$$

Remark 6. *The MIX, SUM, and MAX estimators are well defined because the optimal solution to each of these estimators exists. The solution to the SUM estimator exists for all $T \geq 1$ because (i) the cost $V_T(\cdot)$ is continuous due to the continuity of $f(\cdot)$ and $h(\cdot)$, and (ii) $V_T(\cdot)$ is radially unbounded in the decision variables due to the lower bounds in (7), (8) of Assumption 3. Continuity plus radial unboundedness implies existence of the optimal solution by the Weierstrass theorem. Note that the same existence argument can be made for the MIX and MAX estimators after we transform their objectives into their smoothed versions, as shown in Section 5.2.*

We take incremental input/output-to-state stability (i-IOSS) as the definition of detectability for nonlinear systems [8].

Definition 7 (i-IOSS)**.** *The system $x^+ = f(x,w), y = h(x)$ is incrementally input/output-to-state stable (i-IOSS) if there exist functions $\alpha(\cdot) \in \mathcal{KL}$ and $\gamma_1(\cdot)$, $\gamma_2(\cdot) \in \mathcal{K}$ such that for every two initial states z_1 and z_2, and any two disturbance sequences $\mathbf{w}_1$ and $\mathbf{w}_2$ generating state sequences $\mathbf{x}_1(z_1,\mathbf{w}_1)$ and $\mathbf{x}_2(z_2,\mathbf{w}_2)$, the following holds for all $k \geq 1$*

$$|x(k;z_1,\mathbf{w}_1) - x(k;z_2,\mathbf{w}_2)| \leq \alpha(|z_1 - z_2|, k) + \gamma_1\big(\|\mathbf{w}_1 - \mathbf{w}_2\|_{0:k-1}\big) + \gamma_2\big(\|h(\mathbf{x}_1) - h(\mathbf{x}_2)\|_{0:k-1}\big) \quad (9)$$

The notation $x(k;z,\mathbf{w})$ denotes the solution to $x^+ = f(x,w)$ satisfying initial condition $x(0) = x_0$ with disturbance sequence $\mathbf{w} = \{w(0), w(1), \ldots\}$. The notation $h(\mathbf{x})$ is then defined as $\{h(x(0;x_0,\mathbf{w})), h(x(1;x_0,\mathbf{w})), \ldots\}$.

Definition 8 (Robust global asymptotic stability (RGAS))**.** *The estimate is based on the noisy measurement $\mathbf{y} = h(\mathbf{x}(x_0,\mathbf{w})) + \mathbf{v}$. The estimate is RGAS if for all x_0 and $\overline{x}_0$, and bounded $(\mathbf{w},\mathbf{v})$, there exists functions $\alpha(\cdot) \in \mathcal{KL}$ and $\delta_w(\cdot), \delta_v(\cdot) \in \mathcal{K}$ such that the following holds for all $k \geq 1$*

$$|x(k;x_0,\mathbf{w}) - x(k;\hat{x}(0|k),\widehat{\mathbf{w}})| \leq \phi(|x_0 - \overline{x}_0|, k) + \pi_w(\|\mathbf{w}\|_{0:k-1}) + \pi_v(\|\mathbf{v}\|_{0:k-1}) \quad (10)$$

Remark 9. *The main characteristic of the RGAS definition is that the dynamic system generating the estimate error is input-to-state stable (ISS) [7] considering the disturbances (w,v) as the input.*

Finally we make an additional assumption that enables us to establish the later properties

Assumption 10 (Separability of i-IOSS $\mathcal{KL}$ and $\mathcal{K}$ functions)**.** *The $\mathcal{KL}$ function $\alpha(r,k)$ in Definition 7 satisfies the following separability condition. There exist positive scalars $c_\alpha, p, a > 0$ such that the following holds for all $r \geq 0$, $k \geq 1$*

$$\alpha(r,k) \leq c_\alpha r^p k^{-a}$$

Assumption 11 (Choices of stage costs)**.** *The initial state stage cost $\ell_x(\cdot)$ is chosen so that its lower and upper bounds satisfy*

$$\underline{\gamma}_x(r) = \underline{c}_x r^q$$

for some $\underline{c}_x > 0$ and $q > p/a$.

These assumptions strengthen the detectability condition and link the initial state stage cost to the detectability condition through the condition $q > p/a$. We next show that this requirement is sufficient to establish RGAS of the MIX estimator.

4 Main result

We can now state and prove the main result of the paper.

Theorem 12 (RGAS of full information estimate for bounded disturbances)**.** *Consider an i-IOSS (detectable) system satisfying Assumption 10 and 11 with measurement sequence generated by* (3)*, bounded disturbances satisfying Definition 2, and stage cost satisfying Assumptions 3. Then the full information (MIX) estimator is RGAS.*

Proof. First from the definition of i-IOSS (9), we have the upper bound of the estimation error

$$|x(k;x_0,\mathbf{w}) - x(k;\hat{x}(0|k),\widehat{\mathbf{w}})| \leq \alpha(|x_0 - \hat{x}(0|k)|\,,k) + \gamma_1\big(\,\|\mathbf{w} - \widehat{\mathbf{w}}\|_{0:k-1}\big) + \gamma_2\big(\,\|\mathbf{v} - \widehat{\mathbf{v}}\|_{0:k-1}\big) \quad (11)$$

For each $k \geq 1$, the optimal MIX objective function can be expressed as

$$V_k^0 := V_k(\hat{x}(0|k)),\widehat{\mathbf{w}}_k) = \frac{1}{k}(1+\delta)\ell_x(\hat{x}(0|k) - \overline{x}_0) + \frac{1}{k}\sum_{i=0}^{k-1}\ell_i(\hat{w}(i|k),\hat{v}(i|k)) + \delta \max_{i\in 0:k-1}\ell_i(\hat{w}(i|k),\hat{v}(i|k)) \quad (12)$$

From optimality we know that

$$V_k^0 \leq V_k(x_0,\mathbf{w}) = \frac{1}{k}(1+\delta)\ell_x(x_0 - \overline{x}_0) + \frac{1}{k}\sum_{i=0}^{k-1}\ell_i(w(i),v(i)) + \delta \max_{i\in 0:k-1}\ell_i(w(i),v(i))$$

Then from Remark 5, Assumption 3 and Remark 4 we have the following upper bound $\overline{V}_k$ of V_k^0 valid for $k \geq 1$

$$V_k^0 \leq \overline{V}_k := (1+\delta)(\frac{1}{k}\overline{\gamma}_x\big(\,|x_0 - \overline{x}_0|\,\big) + \overline{\gamma}_w(\|\mathbf{w}\|) + \overline{\gamma}_v(\|\mathbf{v}\|)) \tag{13}$$

Also from Assumption 3 we have the following lower bound of V_k^0

$$V_k^0 \geq \frac{1}{k}(1+\delta)\underline{\gamma}_x\big(\,|\hat{x}(0|k) - \overline{x}_0|\,\big) + \delta\underline{\gamma}_w(|\hat{w}(i|k)|) + \delta\underline{\gamma}_v(|\hat{v}(i|k)|) \tag{14}$$

for any time $i \leq k-1$. Next we proceed to establish an upper bound for $\|\mathbf{w} - \widehat{\mathbf{w}}_k\|_{0:k-1}$. From the triangle inequality and definition of the sup norm we have that

$$\|\mathbf{w} - \widehat{\mathbf{w}}_k\|_{0:k-1} \leq \|\mathbf{w}\|_{0:k-1} + \|\widehat{\mathbf{w}}_k\|_{0:k-1} \tag{15}$$

Next we require a bound for $\|\widehat{\mathbf{w}}_k\|_{0:k-1}$. We have from (14) and (13) that $\underline{\gamma}_w(|\hat{w}(i|k)|) \leq \frac{1}{\delta}\overline{V}_k$ for all $i \leq k-1$. This implies $\|\widehat{\mathbf{w}}_k\|_{0:k-1} \leq \underline{\gamma}_w^{-1}(\frac{1}{\delta}\overline{V}_k)$, and substituting (13) into this result and using (1) gives

$$\begin{aligned}\|\widehat{\mathbf{w}}_k\|_{0:k-1} \leq \underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}\frac{3}{k}\overline{\gamma}_x(|x_0 - \overline{x}_0|)) \\ + \underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}3\overline{\gamma}_w(\|\mathbf{w}\|)) + \underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}3\overline{\gamma}_v(\|\mathbf{v}\|))\end{aligned} \tag{16}$$

Applying $\gamma_1(\cdot)$ to (15) and using (16) gives

$$\begin{aligned}&\gamma_1(\|\mathbf{w} - \widehat{\mathbf{w}}_k\|_{0:k-1}) \leq \gamma_1\big(\,\|\mathbf{w}\| + \|\widehat{\mathbf{w}}_k\|\,\big) \\ &\leq \gamma_1\big(\,\|\mathbf{w}\| + \underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}\frac{3}{k}\overline{\gamma}_x(|x_0 - \overline{x}_0|)) \\ &\quad + \underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}3\overline{\gamma}_w(\|\mathbf{w}\|)) + \underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}3\overline{\gamma}_v(\|\mathbf{v}\|))\big) \\ &\leq \gamma_1\big(3\underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}\frac{3}{k}\overline{\gamma}_x(|x_0 - \overline{x}_0|))\big) \\ &\quad + \gamma_1\big(3\,\|\mathbf{w}\| + 3\underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}3\overline{\gamma}_w(\|\mathbf{w}\|))\big) + \gamma_1\big(3\underline{\gamma}_w^{-1}(\frac{1+\delta}{\delta}3\overline{\gamma}_v(\|\mathbf{v}\|))\big)\end{aligned}$$

Noting that $\beta(r,k) := (1/k)\overline{\gamma}_x(r)$ is $\mathcal{KL}$, and using the properties of $\mathcal{K}$ and $\mathcal{KL}$ functions, this equation can be expressed as

$$\gamma_1\big(\,\|\mathbf{w} - \widehat{\mathbf{w}}_k\|_{0:k-1}\,\big) \leq \phi_x^w(|x_0 - \overline{x}_0|\,, k) + \pi_w^w(\|\mathbf{w}\|) + \pi_v^w(\|\mathbf{v}\|) \tag{17}$$

with $\phi_x^w \in \mathcal{KL}$ and $\pi_w^w, \pi_v^w \in \mathcal{K}$. Furthermore, notice that the same reasoning applies to $\|\mathbf{v} - \widehat{\mathbf{v}}_k\|_{0:k-1}$ yielding

$$\gamma_2\big(\,\|\mathbf{v} - \widehat{\mathbf{v}}_k\|_{0:k-1}\,\big) \leq \phi_x^v(|x_0 - \overline{x}_0|\,, k) + \pi_w^v(\|\mathbf{w}\|) + \pi_v^v(\|\mathbf{v}\|) \tag{18}$$

for $\phi_x^v \in \mathcal{KL}$ and $\pi_w^v, \pi_v^v \in \mathcal{K}$.

From (14) and (13) we also have that

$$\begin{aligned}\underline{\gamma}_x\big(\,|\hat{x}(0|k) - \overline{x}_0|\,\big) &\leq \frac{k}{1+\delta}V_k^0 \leq \frac{k}{1+\delta}\overline{V}_k \\ &\leq \overline{\gamma}_x\big(\,|x_0 - \overline{x}_0|\,\big) + k\overline{\gamma}_w(\|\mathbf{w}\|) + k\overline{\gamma}_v(\|\mathbf{v}\|)\end{aligned}$$

Then taking the inverse of the $\mathcal{K}$ function and using (1) generates

$$\begin{aligned}&|\hat{x}(0|k) - \overline{x}_0| \\ &\leq \underline{\gamma}_x^{-1}\big(\overline{\gamma}_x\big(\,|x_0 - \overline{x}_0|\,\big) + k\overline{\gamma}_w(\|\mathbf{w}\|) + k\overline{\gamma}_v(\|\mathbf{v}\|)\big) \\ &\leq \underline{\gamma}_x^{-1}\big(3\overline{\gamma}_x\big(\,|x_0 - \overline{x}_0|\,\big)\big) + \underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big) + \underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_v(\|\mathbf{v}\|)\big)\end{aligned}$$

Again from the triangle inequality we have that

$$\begin{aligned}|\hat{x}(0|k) - x_0| &= |(\hat{x}(0|k) - \overline{x}_0) - (x_0 - \overline{x}_0)| \leq |\hat{x}(0|k) - \overline{x}_0| + |x_0 - \overline{x}_0| \\ &\leq |x_0 - \overline{x}_0| + \underline{\gamma}_x^{-1}\big(3\overline{\gamma}_x\big(\,|x_0 - \overline{x}_0|\,\big)\big) + \underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big) + \underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_v(\|\mathbf{v}\|)\big)\end{aligned}$$

Then we have the upper bound of the first part of (11) as

$$\begin{aligned}&\alpha(|\hat{x}(0|k) - x_0|\,, k) \\ &\leq \alpha\big(\,|x_0 - \overline{x}_0| + \underline{\gamma}_x^{-1}\big(3\overline{\gamma}_x\big(\,|x_0 - \overline{x}_0|\,\big)\big) \\ &\quad + \underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big) + \underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_v(\|\mathbf{v}\|)\big), k\big) \\ &\leq \alpha\big(3\,|x_0 - \overline{x}_0| + 3\underline{\gamma}_x^{-1}\big(3\overline{\gamma}_x\big(\,|x_0 - \overline{x}_0|\,\big)\big), k\big) \\ &\quad + \alpha\big(3\underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big), k\big) + \alpha\big(3\underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_v(\|\mathbf{v}\|)\big), k\big)\end{aligned}$$

The first term on the right-hand side of the inequality is a $\mathcal{KL}$ function. Using Assumption 10, the second term satisfies

$$\alpha\big(3\underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big), k\big) \leq \left(\frac{c_\alpha 3^{(p+p/q)}}{\underline{c}_x^{p/q}}\right) k^{p/q-a}\overline{\gamma}_w(\|\mathbf{w}\|)^{p/q} := \phi_w^x(\|\mathbf{w}\|\,, k)$$

Note that due to Assumption 10, $p/q - a < 0$ and ϕ_w^x is therefore a $\mathcal{KL}$ function. Similar analysis applies to the third term giving

$$\alpha(|\hat{x}(0|k) - x_0|\,, k) \leq \phi_x^x(|x_0 - \overline{x}_0|\,, k) + \phi_w^x(\|\mathbf{w}\|\,, k) + \phi_v^x(\|\mathbf{v}\|\,, k)$$

for $\phi_x^x, \phi_w^x, \phi_v^x \in \mathcal{KL}$. Note that $\phi_w^x(\|\mathbf{w}\|\,, k) \leq \phi_w^x(\|\mathbf{w}\|\,, 1)$ and $\phi_v^x(\|\mathbf{v}\|\,, k) \leq \phi_v^x(\|\mathbf{v}\|\,, 1)$ giving

$$\alpha(|\hat{x}(0|k) - x_0|\,, k) \leq \phi_x^x(|x_0 - \overline{x}_0|\,, k) + \pi_w^x(\|\mathbf{w}\|) + \pi_v^x(\|\mathbf{v}\|) \tag{19}$$

for $\pi_w^x, \pi_v^x \in \mathcal{K}$. We substitute (19), (18), and (17) into (11) to obtain for all $k \geq 1$,

$$|x(k; x_0, \mathbf{w}) - x(k; \hat{x}(0|k), \widehat{\mathbf{w}}_k)| \leq \phi(|x_0 - \overline{x}_0|\,, k) + \pi_w(\|\mathbf{w}\|) + \pi_v(\|\mathbf{v}\|)$$

in which $\phi := \phi_x^x + \phi_x^w + \phi_x^v \in \mathcal{KL}$, and $\pi_w := \pi_w^x + \pi_w^w + \pi_w^v \in \mathcal{K}$, and $\pi_v := \pi_v^x + \pi_x^w + \pi_v^v \in \mathcal{K}$. Since $w(j)$, $v(j)$ for $j \geq k$ affect neither $x(k)$ nor $\hat{x}(k|k)$, this result also implies that

$$|x(k; x_0, \mathbf{w}) - x(k; \hat{x}(0|k), \widehat{\mathbf{w}}_k)| \leq \phi(|x_0 - \overline{x}_0|\,, k) + \pi_w(\|\mathbf{w}\|_{0:k-1}) + \pi_v(\|\mathbf{v}\|_{0:k-1})$$

The estimate error therefore satisfies (10) and RGAS has been established. □

Note that RGAS of the MAX estimator can be established similarly.

5 Discussion

5.1 Discussion of Assumption 10

The new restriction in this paper is Assumption 10, so we elaborate further on this assumption. If we step back, remove some details, and look at the general issue that is being addressed in Assumption 10, it is basically this. When can the linear growth in the first term of a $\mathcal{KL}$ function be overcome by the decrease in its second argument, i.e., for what $\alpha(\cdot) \in \mathcal{KL}$, does there exist $\gamma(\cdot) \in \mathcal{K}$ such that for some $\beta(\cdot)$ in $\mathcal{KL}$

$$\alpha(\gamma(rk), k) \leq \beta(r, k)$$

for all $r \geq 0$, and $k \geq 1$? This general question appears to be rather new and unexplored, especially here because it arises in the context of the system's detectability, which is already a complex issue for nonlinear systems.

There are, however, some important cases in which Assumption 10 clearly is satisfied, and Assumption 11 is not required. The first case is nonlinear observability. Nonlinear observability rather than detectability corresponds to the case of $\alpha(\cdot) = 0$ in Definition 7.

Remark 13. *Assumption 10 is satisfied for a nonlinear observable system, in which $\alpha(\cdot) = 0$. Any $\ell_x(\cdot)$ satisfying (7) may be used in this case.*

The next case of interest is (constrained) linear systems

$$\begin{aligned} x^+ &= Ax + Gw \\ y &= Cx + v \end{aligned} \tag{20}$$

with quadratic penalties

$$\ell_x(x) = (1/2)\,|x|^2_{P_0} \qquad \ell_i(w, v) = (1/2)(|w|^2_Q + |v|^2_R)$$

for penalty matrices $P_0, Q, R > 0$. We have the following result.

Proposition 14 (Constrained linear systems with quadratic costs)**.** *Assumption 10 is satisfied for LQ estimation of a detectable system. In this case there exists $c > 0$ and $\lambda < 1$ such that for all $k \geq 1$*

$$\alpha\big(3\underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big), k\big) \leq c\,\|\mathbf{w}\|\,\lambda^k$$

and the right-hand side is an exponential $\mathcal{KL}$ function.

Proof. For a detectable linear system it can be shown that the $\mathcal{KL}$ function $\alpha(\cdot)$ satisfies an exponential decay rate

$$\alpha(r, k) \leq c_\alpha r \overline{\lambda}^k, \quad 0 < \overline{\lambda} < 1 \tag{21}$$

The stage cost bounds satisfy

$$\begin{aligned} \underline{\gamma}_x(|x|) &= \underline{a}_x\,|x|^2 & \overline{\gamma}_x(|x|) &= \overline{a}_x\,|x|^2 \\ \underline{\gamma}_w(|w|) &= \underline{a}_w\,|w|^2 & \overline{\gamma}_w(|w|) &= \overline{a}_w\,|w|^2 \\ \underline{\gamma}_v(|v|) &= \underline{a}_v\,|v|^2 & \overline{\gamma}_v(|v|) &= \overline{a}_v\,|v|^2 \end{aligned}$$

in which $\underline{a}_x$, $\underline{a}_w$, and $\underline{a}_v$ ($\overline{a}_x$, $\overline{a}_w$, and $\overline{a}_v$) denote (1/2) times the smallest (largest) singular values of P_0, Q, and R, respectively. Using these results, $\gamma_x^{-1}(\cdot) = \sqrt{(1/\underline{a}_x)(\cdot)}$ and we have that

$$3\underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big) \leq c_\gamma \sqrt{k}\,\|\mathbf{w}\|$$

with $c_\gamma = 3\sqrt{3k\overline{a}_w/\underline{a}_x}$. Applying $\alpha(\cdot)$ gives

$$\alpha\big(3\underline{\gamma}_x^{-1}\big(3k\overline{\gamma}_w(\|\mathbf{w}\|)\big), k\big) \leq c_\alpha c_\gamma \|\mathbf{w}\| \sqrt{k}\, \overline{\lambda}^k$$

We can then increase $\overline{\lambda}$ slightly and obtain an exponential bound, i.e., there exits λ such that

$$c_\alpha c_\gamma \|\mathbf{w}\| \sqrt{k}\, \overline{\lambda}^k \leq c \|\mathbf{w}\| \lambda^k, \quad \overline{\lambda} < \lambda < 1$$

and the result is established. □

Note that for the LQ case, both Assumption 10 and 11 are satisfied. Assumption 10 holds for any $a > 0$ since λ^k decays faster than k^{-a}. The LQ values of $p = 1$ and $q = 2$ satisfies $q > p/a$ in Assumption 11 for $a > 1/2$.

5.2 Computation of the MIX and MAX estimator

In order to computationally solve the MIX estimator (5), we can redefine it as

$$\min_{\chi(0),\boldsymbol{\omega},\ell_{\max}} V_T(\chi(0), \boldsymbol{\omega}) = (1+\delta)\frac{1}{T}\ell_x(\chi(0) - \overline{x}_0) + \frac{1}{T}\sum_{j=0}^{T-1} \ell_i(\omega(i), \nu(i)) + \delta \ell_{\max} \tag{22}$$

subject to

$$\chi^+ = f(\chi, \omega) \qquad y = h(\chi) + \nu \qquad \ell_{\max} - \ell_i(\omega(i), \nu(i)) \geq 0, \quad i \in 0 : T-1$$

Then it has been transformed into a smooth optimization problem that can be solved by standard nonlinear optimization tools. A similar transformation can be applied on the MAX estimator. From this form of the definition, we can see the MIX and MAX estimators increase the dimension of the constraints by T, so they are usually slower and more difficult to solve than the SUM estimator, especially when T is large. An MHE version of these estimators would significantly reduce the computational requirements.

5.3 Uniqueness and convergence of the solution

Besides RGAS, there are important properties that should be studied. One is the uniqueness of the optimal solution. For nonlinear systems, it is difficult to ensure convexity of the SUM estimator since $\nu(i)$ is a nonlinear function of $\chi(0)$ and $\omega(0), \cdots, \omega(i-1)$. However, it is easy to show the objective function is convex when the system is linear, which is the main reason that we anticipate uniqueness of the SUM estimator in mildly nonlinear applications. When considering the MIX estimator, based on (22), the objective function contains an additional linear term of $\ell_{\max}$, which should not affect the objective convexity (or the constraints). So the same argument can be applied to the MIX estimator for linear systems; and we have the same reason to anticipate that the MIX estimator also may have a unique solution for mildly nonlinear applications.

Another important property is the convergence of the state estimate when the disturbances are not only bounded but also converge to zero.

Remark 15. *Notice if the disturbances satisfy* $|w(k)|, |v(k)| \to 0$ *as* $k \to \infty$*, Definition 8 does not imply* $|x(k; x_0, \mathbf{w}) - x(k; \hat{x}(0|k), \widehat{\mathbf{w}})| \to 0$*, because* $\|\mathbf{w}\|_{0:k-1}$*,* $\|\mathbf{v}\|_{0:k-1}$ *(and* $\|\widehat{\mathbf{w}}\|_{0:k-1}$*,* $\|\widehat{\mathbf{v}}\|_{0:k-1}$*) can remain large for large* k *and corresponding small* $|w(k)|$ *and* $|v(k)|$*.*

Nevertheless, in [5, Proposition 11] the convergence of the SUM estimator has been established, but the argument used in that case does not extend to the MIX and MAX estimators. Therefore a proof of the convergence of MIX and MAX estimators is a valuable topic for future research. When δ is taken small, the MIX estimator can be viewed as the SUM estimator with a small perturbation. Therefore, it is reasonable to conjecture that the MIX estimator is convergent for δ small enough. This conjecture will be shown to be the case with the two simulation examples in Section 6. Obviously when δ is large this assumption does not hold and the convergence is not expected to hold. The linear example in 6.1 is a counterexample of the MAX estimator's convergence (with $\delta \to \infty$). On the other hand, if δ is chosen too small, according to (17) and (18), a large RGAS upper bound could result. Therefore there is a trade off in the value of δ in order to ensure both properties and good performance.

6 Example

Here we design two examples to illustrate the behaviors of the presented estimators. The first is a trial linear example with an occasional bad measurement outlier, and the second is a physical nonlinear example with random bounded noises for process and measurement disturbances. All simulations were performed with the free software package GNU Octave [1], and the software is available upon request.

6.1 Linear example with measurement noise outlier

This example aims to illustrate why we do not expect the MAX estimator to be useful in practice even though it ensures RGAS. A simple detectable (and therefore also i-IOSS) linear system is defined as

$$\begin{aligned} x^+ &= x + w \\ y &= x + v \end{aligned} \tag{23}$$

where $x, y \in \mathbb{R}^1$. We set $\ell_x(x) = x^2$ and $\ell_i(w, v) = w^2 + v^2$. Assume we already have a precise initial guess $\bar{x}_0 = x(0) = 0$. The disturbances $w(k), v(k)$ are generally all zero except for the nonzero $v(2) = 1$. In other words, all disturbances are both bounded and convergent, all states x and outputs y are zero except for one erroneous measurement $y(2) = 1$. The system is easy to estimate, and it is reasonable to expect a good estimator

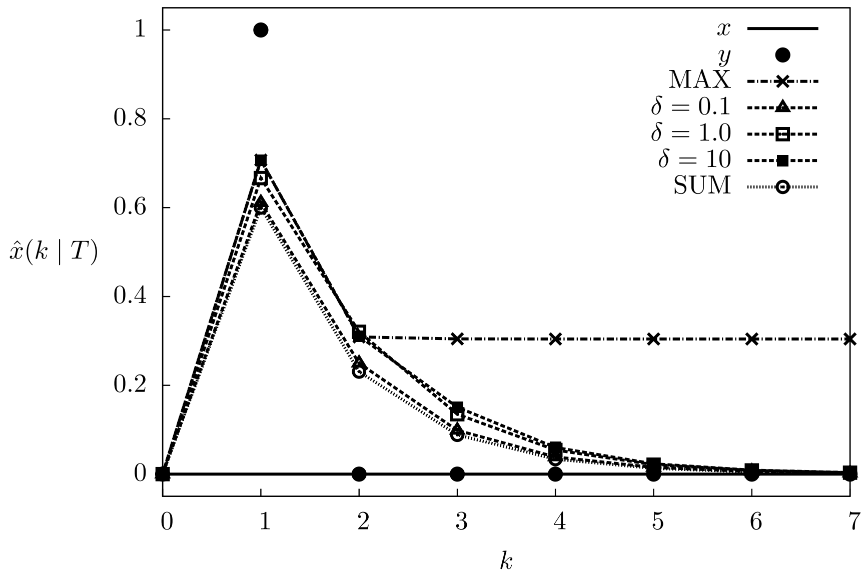

Figure 1: Data and estimates of Example 6.1

should reject the single poor measurement and obtain a converged result given that the disturbances converge. Here we conduct the simulation with maximum time $T = 7$. The MIX estimators are implemented for different δ values, as well as the MAX and SUM estimator. Estimate results are shown in Figure 1. From Figure 1, the state estimate of the SUM estimator converges to zero, which is consistent with its stated properties. The MIX estimators also converge for all values of parameter δ. Finally, the MAX estimator fails to converge to zero and its estimate error remains large regardless of how many measurements become available. Yet the MAX estimator is RGAS since $\|\mathbf{v}\| = |v(2)| = 1$ and the RGAS upper bound in (10) is a fixed constant.

6.2 Nonlinear example

Here we use the gas-phase irreversible reaction example proposed in [2]

$$2\text{A} \longrightarrow \text{B}$$

with the reaction rate $r = kc_A^2$, $k = 0.16$. We define the two states x_1, x_2 as the partial pressures of species A and B, and the measurement as the total pressure. Assuming the ideal gas law holds and the batch reactor is well-mixed and isothermal, the model of system is

$$f(x) = \begin{bmatrix} -2kx_1^2 \\ kx_1^2 \end{bmatrix}, \qquad x^+ = \int_0^\Delta f(x)dt + w, \qquad y = \begin{bmatrix} 1 & 1 \end{bmatrix} x + v$$

with the sample time $\Delta = 0.1$, and the initial state $x(0) = [3 \quad 1]^T$. In the simulation we assume the plant suffers random noises $w \sim \mathcal{N}(0, Q_w)$ and $v \sim \mathcal{N}(0, R_v)$ where $Q_w =$

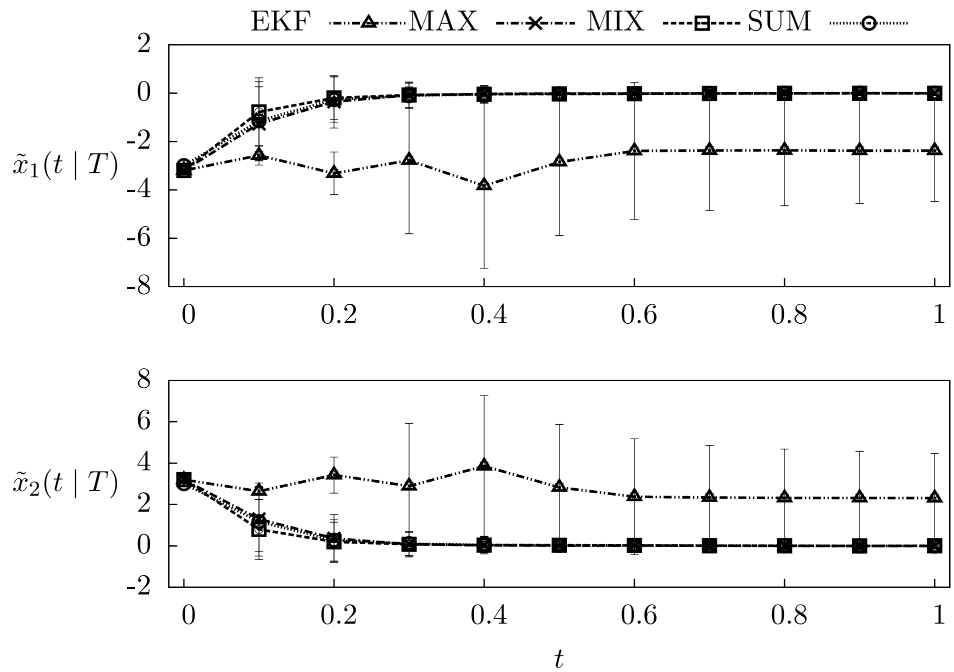

Figure 2: Mean and variation of $\tilde{x}(t \mid T)$ at $T = 1$.

$\text{diag}(0.001^2, 0.001^2)$ and $R_v = 0.1^2$. However, to ensure boundedness, hard constraints are included such that $|w_1|, |w_2| \leq 0.01$, $|v| \leq 1.0$. As in [2], the extended Kalman filter (EKF) given a poor initial guess $\bar{x}_0 = [0.1 \quad 4.5]^T$, $\Pi_0 = \text{diag}(6^2, 6^2)$ and accurate Q_w, R_v covariance values is not stable. For comparison we also simulate the full information estimators with the same $\bar{x}_0$ and stage costs defined by

$$\ell_x(\chi) := \chi^T P_0 \chi, \qquad P_0 = \Pi_0^{-1}$$
$$\ell_i(\omega, \nu) := \omega^T Q \omega + \nu^T R \nu, \qquad Q = Q_w^{-1}, \qquad R = R_v^{-1}$$

The optimization is conducted over $T = 11$ steps. The MIX estimator uses $\delta = 1$. To provide sufficient statistical samples, a total of $s = 300$ simulations are performed for each estimator.

Defining the state estimate error as $\tilde{x}(i) := x(i) - \hat{x}(i|i)$, Figure 2 shows the sample averages (over all the runs) and variations of $\tilde{x}_1(T)$ and $\tilde{x}_2(T)$ of all four estimators. The EKF's estimate error does not converge to zero, and the estimator errors of MAX, MIX and SUM do converge to zero. To better compare their performances, we define the benchmark as $\left|\tilde{x}^j(T)\right|^2$ where j denotes the jth simulation run. The histogram at the final time is shown in Figure 3, which clearly indicates the performance differences. To make the comparison more straightforward, we can also look at the statistical expectation

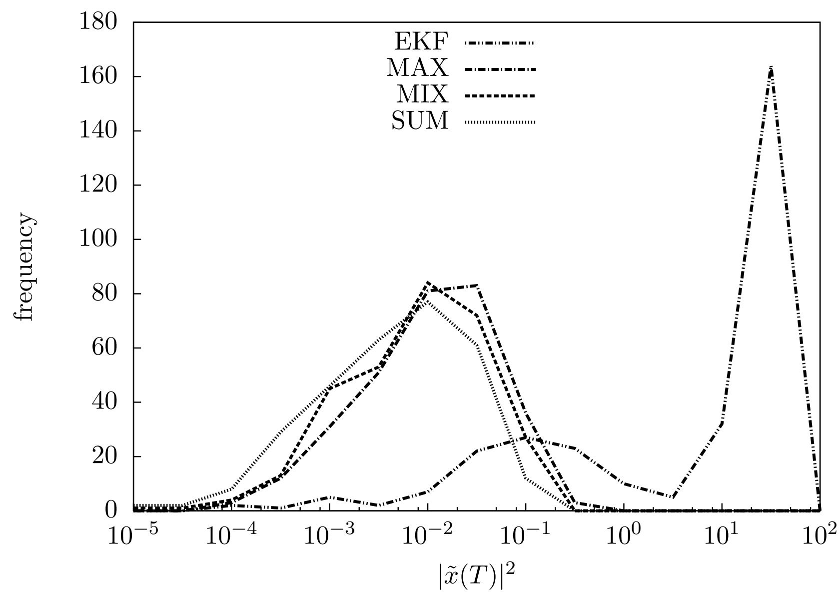

Figure 3: Histogram of $|\tilde{x}(T)|^2$; EKF: $\langle|\tilde{x}(T)|^2\rangle = 20.101$, MAX: $\langle|\tilde{x}(T)|^2\rangle = 0.029$, MIX: $\langle|\tilde{x}(T)|^2\rangle = 0.023$, SUM: $\langle|\tilde{x}(T)|^2\rangle = 0.015$

$\mathrm{E}(|\tilde{x}(T)|^2)$. In practice, the sample averages can be used to approximate the expectation

$$\langle|\tilde{x}(T)|^2\rangle := \frac{1}{s}\sum_{j=1}^{s}\left|\tilde{x}^j(T)\right|^2$$

These values are given in the caption of Figure 3.

To better understand how the full information estimators perform, we show the entire trajectory of optimal stage costs ($\hat{l}_i := \ell(\hat{w}(i|T), \hat{w}(i|T))$) at the final time in one typical run (Figure 4). The MAX and MIX estimators have smaller maximum $\hat{l}_i$ than the SUM estimator, and the result of the MIX estimator is close to the MAX estimator. Similarly, we can use V_T^0, $V_T^{0,\max}$ and $V_T^{0,\mathrm{sum}}$ values as another performance benchmark. The histograms and corresponding sample averages are shown in Figure 5, 6, and 7. Not too surprisingly, each of the three estimators performs best on the benchmark corresponding to its own objective function. To provide a better indication that how the optimization works, the 'Actual' plots in these figures show the values of the benchmarks using the actual disturbances(w, v) in the plant ($\ell(w(i), v(i))$, $V_T(x(0), \mathbf{w})$, $V_T^{\mathrm{sum}}(x(0), \mathbf{w})$ and $V_T^{\max}(x(0), \mathbf{w})$).

7 Conclusion

Establishing robust global asymptotic stability (RGAS) ensures that the estimate error of a nonlinear estimator has an upper-bound depending on the sizes of the initial estimate error and the process and measurement disturbances. In previous work, RGAS of the full

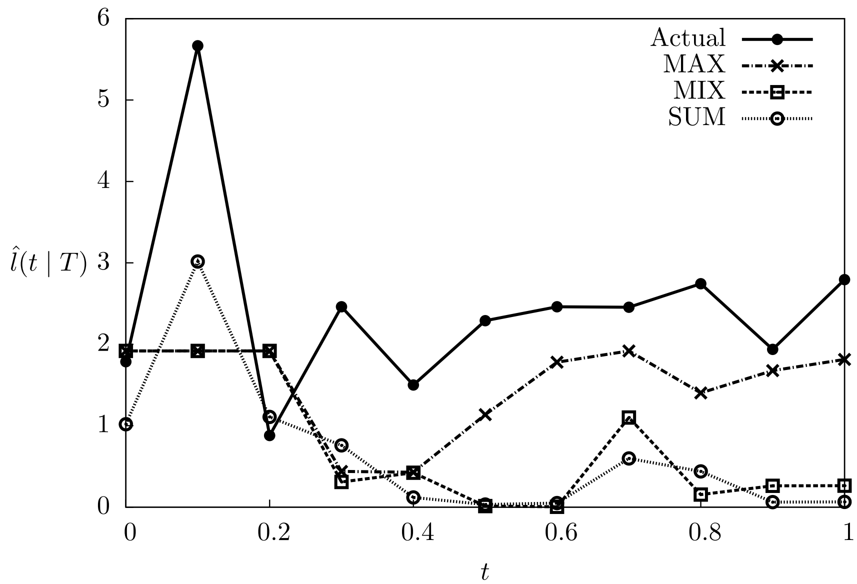

Figure 4: Open-loop estimated stage costs at $T = 1$.

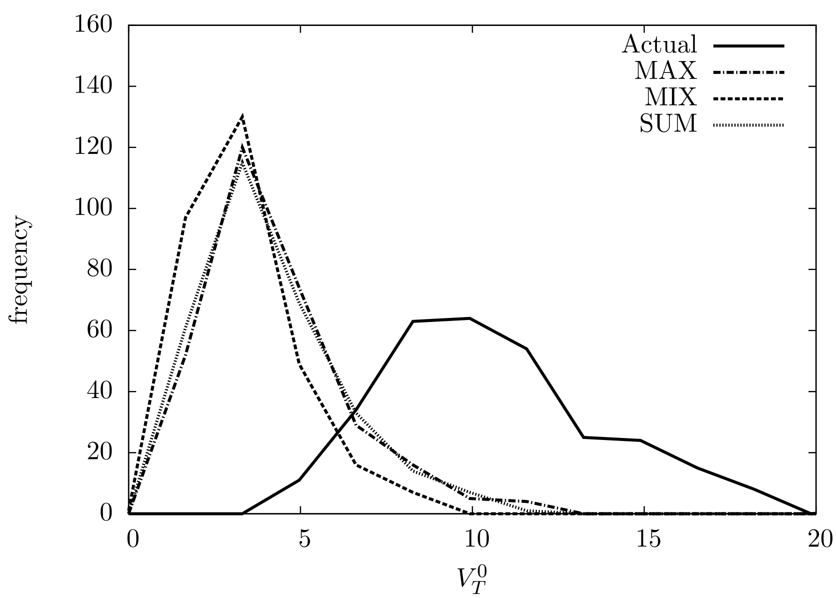

Figure 5: Histogram of V_T^0; Data: $\langle V_T^0 \rangle = 10.627$, MAX: $\langle V_T^0 \rangle = 4.192$, MIX: $\langle V_T^0 \rangle = 3.275$, SUM: $\langle V_T^0 \rangle = 4.129$

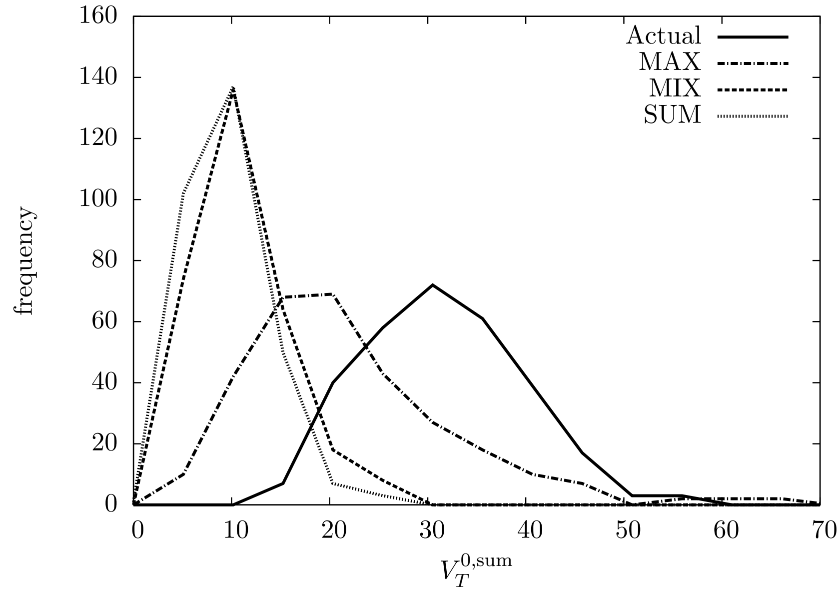

Figure 6: Histogram of $V_T^{0,\text{sum}}$; Data: $\langle V_T^{0,\text{sum}} \rangle = 31.58$, MAX: $\langle V_T^{0,\text{sum}} \rangle = 21.938$, MIX: $\langle V_T^{0,\text{sum}} \rangle = 11.077$, SUM: $\langle V_T^{0,\text{sum}} \rangle = 9.733$

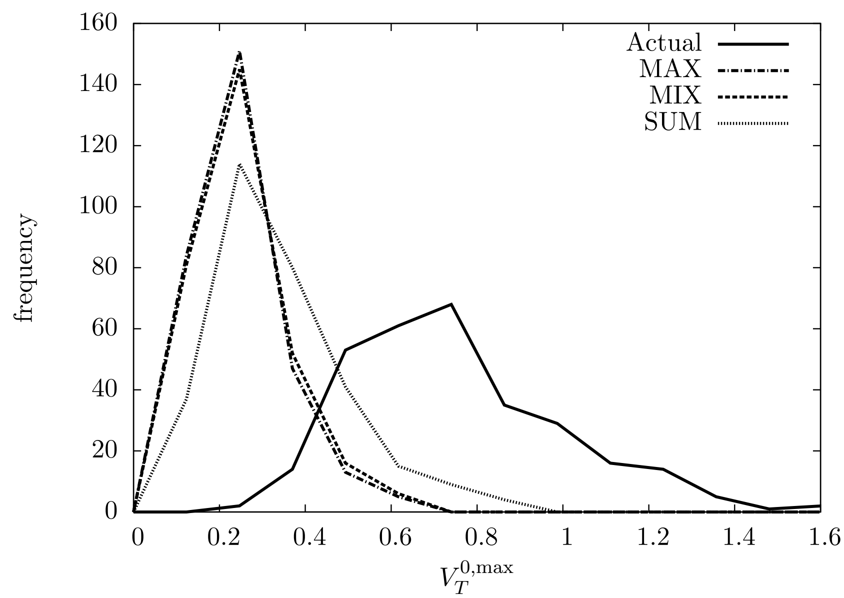

Figure 7: Histogram of $V_T^{0,\text{max}}$; Data: $\langle V_T^{0,\text{max}} \rangle = 0.753$, MAX: $\langle V_T^{0,\text{max}} \rangle = 0.248$, MIX: $\langle V_T^{0,\text{max}} \rangle = 0.254$, SUM: $\langle V_T^{0,\text{max}} \rangle = 0.343$

information estimator has been established for i-IOSS (detectable) nonlinear systems for convergent disturbances, i.e., disturbances that converge to zero as time increases to infinity. In applications, however, it is more reasonable to assume that the system disturbances are only bounded and not convergent. In this paper we defined a new form of the full information estimator that provides RGAS for detectable nonlinear systems for bounded disturbances. The new objective function includes the maximum over all stage costs as well as the standard sum of stage costs. The estimator is still optimization based and can incorporate constraints. To establish RGAS, we made one additional assumption relating the stage cost functions to the i-IOSS property. For observably nonlinear systems and constrained linear quadratic estimation of detectable systems, this assumption is always satisfied. An alternative form of this estimator is provided in order to implement the state estimator with standard nonlinear programming solvers.

We should emphasize that this RGAS definition does not automatically imply that the estimate error converges to zero for convergent disturbances due to the addition of the maximal stage cost term. It remains an open problem to establish that a single full information estimator has both RGAS and convergence properties. In future studies a moving horizon version of this new full information estimator will be valuable since it offers the same stability properties with a significantly smaller and tractable computational complexity.

Acknowledgments

The authors would like to thank Professors D.Q. Mayne and A.R. Teel for helpful discussion of this paper. The authors would like to thank the industrial members of the Texas Wisconsin California Control Consortium (TWCCC) for financial support of this research.

References

[1] John W. Eaton, David Bateman, and Sren Hauberg. *GNU Octave Manual Version 3*. Network Theory Limited, 2008.

[2] Eric L. Haseltine and James B. Rawlings. A critical evaluation of extended Kalman filtering and moving horizon estimation. Technical Report 2002–03, TWMCC, Department of Chemical Engineering, University of Wisconsin–Madison, August 2002.

[3] Hassan K. Khalil. *Nonlinear Systems*. Prentice-Hall, Upper Saddle River, NJ, third edition, 2002.

[4] J. Liu. Moving horizon state estimation for nonlinear systems with bounded uncertainties. *Chem. Eng. Sci.*, 93:376–386, February 2013.

[5] James B. Rawlings and Luo Ji. Optimization-based state estimation: Current status and some new results. *J. Proc. Cont.*, 22:1439–1444, 2012.

[6] James B. Rawlings and David Q. Mayne. *Model Predictive Control: Theory and Design.* Nob Hill Publishing, Madison, WI, 2009. 576 pages, ISBN 978-0-9759377-0-9.

[7] E. D. Sontag and Y. Wang. On the characterization of the input to state stability property. *Sys. Cont. Let.*, 24:351–359, 1995.

[8] Eduardo D. Sontag and Yuan Wang. Output-to-state stability and detectability of nonlinear systems. *Sys. Cont. Let.*, 29:279–290, 1997.