# Measurement Feedback Nonlinear $H^\infty$ Control: The Cheap Sensor Case

J.W. Helton* M.R. James† W.M. McEneaney‡

November 27, 2000

LONG DRAFT - will probably be cut in two

**Abstract**

This paper concerns the measurement feedback nonlinear control problem and a practical aspect of extending $H^\infty$ control to it. We consider the *cheap sensor problem*, where some of the state variables are *perfectly* measured. A *reduced order "observer"* describes the evolution of the controller state, resulting in a *huge* reduction of computational complexity. This reduced order observer is implementable with today's technology in parameter ranges which are discussed in this paper. *In terms of computational cost, the cheap sensor measurement feedback problem is closer to the full state feedback problem than to the standard measurement feedback problem.*

**Key words:** Nonlinear $H_\infty$ control, information states, nonlinear estimation, infinite dimensional dynamical systems, reduction of computational complexity, reduced order observers.

# 1 Introduction

Measurement feedback nonlinear $H^\infty$ control theory evolved during the 1990's in response to the need to extend to nonlinear systems the highly successful $H^\infty$ theory for linear

*Department of Mathematics, UCSD, San Diego, CA 92093, USA. E-mail: helton@osiris.ucsd.edu WWW URL: http://www.ucsd.edu, Partially supported by the AFOSR, the NSF, and Ford Motor Co.

†Department of Engineering, Australian National University, Canberra, ACT 0200, Australia. E-mail: Matthew.James@anu.edu.au WWW URL: http://spigot.anu.edu.au/~mat/ Partially supported by the Australian Research Council.

‡Dept. of Mathematics, North Carolina State University, Raleigh, NC 27695–8205, USA, wmm@math.ncsu.edu, http://www4.ncsu.edu/~wmm/. Research partially supported by AFOSR grant F08671–98–0–1098.

systems developed during the 1980's. $H^\infty$ theory is an approach to the analysis and design of controllers that are *robust* with respect to uncertainty.

At the core of nonlinear $H^\infty$ control theory lie two partial differential equations (PDE). One is a first order evolution equation called the *information state* equation which plays the role of a generalized state estimator and in fact constitutes the dynamics of the controller. It must be implemented *online*. The second PDE is defined on the space of possible information states and determines the output law for the controller, namely, it gives a function $f$ producing the control signal $u(t)$ via the formula $u(t) := f(p_t)$ at time $t$ based on the current state $p_t$ of the controller (information state feedback). This can be implemented *offline*. Under special conditions, known as *certainty equivalence conditions*, the second PDE can be replaced by a simpler state feedback PDE on a the state space of the plant. This an Hamilton-Jacobi-Bellman-Isaacs (HJBI) partial differential equation; it is the subject of vast study and for further details we refer to the books [5], [15], [28] and the many references cited therein.

The state feedback HJBI partial differential equation faces the dreaded curse of dimensionality. This stems from the fact it is a PDE on $n$-dimensional space. For example, if we select $G_x$ grid points per dimension and then the corresponding lattice of grid points on $\mathbf{R}^n$ has $G_x^n$ grid points. Taking $G_x = 50$ seems reasonable and if $n = 12$, then it requires $50^{12} \approx 10^{17}$ storage locations and a corresponding number of floating point operations (FLOPS) to perform each calculation. These are of course extremely large numbers.

For the purpose of this paper we will suppose that since the calculations can be done offline the information state feedback function has been obtained, perhaps using clever tricks, special structure, approximation, and long computing time. We will not consider this issue further.

This paper concerns only the information state partial differential equation. It is a dynamical equation on the state space $\mathbf{R}^n$ of the plant to be controlled, in contrast to the state feedback HJBI equation which is not dynamical. In fact the information state PDE constitutes the dynamics of the controller, so it must be solved in real time. The curse of dimensionality makes this seem impossible. However we give a method in this paper, which for cases where all but a few states are measured directly could be used to solve the information state dynamical equation in practice. Our computational savings result from the fact that our method reduces the problem to solving a PDE on a low dimensional space and the fact that updating one time step of a dynamical PDE is much faster than solving a steady state PDE of the same dimension. For example, we show in Section 5.2.3, that for a plant with 6 states of which 4 are perfectly measured and 2 imperfectly that with $G_x = 50$ grid points per dimension, the computational complexity of updating the reduced information state PDE at each time step is $O(6 \times 50^2)$ for memory and $O(10^7)$ Floating Point Operations (FLOPS). This is *significantly less* complex than a direct solution of the full information state PDE in 6 dimensions, which would require $O(6 \times 50^6)$ memory elements and $O(10^{12})$ FLOPS per time step, which is of course infeasible. As another example, if we assume that the input operator to the plant ($\tilde{B}_1(x)$ in equation (3.1) below) is independent of the unknown state, then for state dimension 12 and 10 directly measured states and 2 unmeasured states, with $G_x = 50$ we get that an update costs on the order of $O(12 \times 50^2)$ memory locations and $O(10^7)$ flops.

While the cost of sensors is something which depends on the particular application and changes with time, it is not unreasonable to suppose that in many useful applications sensors are cheap compared to other costs and can be implemented for a significant number of state variables, leaving only a small number of states in need of estimation. Accordingly, we use the terminology *"cheap sensor case"*. The familiar and often used *state feedback case* is a special case of the cheap sensor case; it of course reflects an idealization where *all* states are *perfectly* measured. *In terms of computational cost, the cheap sensor measurement feedback problem is closer to the full state feedback problem than to the standard regular measurement feedback problem.*

To help place the cheap sensor case in perspective, and to prepare for subsequent discussion in the paper, it is useful to recall the commonly used standard form for $H^\infty$ control problems (both linear and nonlinear), where the state space model for a generalized plant $G$ is given by

$$G \; : \; \left[\begin{array}{c|c|c} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ \hline C_2 & D_{21} & D_{22} \end{array}\right]. \tag{1.1}$$

The operator $G$ maps $(w, u) \mapsto (z, y)$, where $w$ is the disturbance input, $u$ is the control input, $z$ is the performance output, and $y$ is the *measurement*; and the internal state is denoted by $\xi$. Thus:

$$G \; : \; \begin{cases} \dot{\xi} = A(\xi) + B_1(\xi) w + B_2(\xi) u \\ z = C_1(\xi) + D_{11} w + D_{12} u \\ y = C_2(\xi) + D_{21} w + D_{22} u \end{cases} \tag{1.2}$$

Throughout, we will use the convention of representing a point in the state space by $x$, and a trajectory of the system (taking values in the state space) by $\xi(\cdot)$. For instance, the initial condition for the above system might be $\xi(0) = x$. As is commonly done, we take $D_{11} = 0$ and $D_{22} = 0$ without loss of generality. Various cases arise depending on the ranks of $D_{12}$ and $D_{21}$. In the *regular case* they both have full rank, otherwise, there are two *singular cases.* * The case where $D_{12}$ is rank deficient has come to be termed the *cheap control case* in the literature (see, e.g. [26] and the references therein), while the case where $D_{21}$ is rank deficient corresponds to our *cheap sensor case.* The cheap sensor case has been studied for linear systems in [27] (without using this terminology), where reduced-order $H^\infty$ compensators were obtained. In the present paper we study a nonlinear analog of [27] in the context of the information state framework [15], and also some special situations not considered in [27]. Specifically, we consider the generalized

*The term *singular* has a different meaning in [15], where it refers to extended real valued information states; these will be used in this paper.

plant structure

$$G \; : \; \left[\begin{array}{c|c|c} A & (\tilde{B}_1, B_1) & B_2 \\ \hline C_1 & (0,0) & D_{12} \\ \hline \begin{pmatrix} I_\kappa \\ C_{2\mu} \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & I_\mu \end{pmatrix} & 0 \end{array}\right], \tag{1.3}$$

with $D_{12}$ full rank, where the state is decomposed as $x = (x_\kappa, x_\mu)$, with $x_\kappa$ measured perfectly and $x_\mu$ imperfectly measured. The state feedback case corresponds to $x = x_\kappa$, while in the full regular case $x = x_\mu$. In standard notation,

$$D_{21} = \begin{bmatrix} 0 & 0 \\ 0 & I_\mu \end{bmatrix} \text{ and } C_2(x) = \begin{bmatrix} x_\kappa \\ C_{2\mu}(x) \end{bmatrix}. \tag{1.4}$$

For references on nonlinear $H_\infty$ control, see for example, [5], [28], [18], [4], [15] and the references therein. The singular $D_{ij}$ problem we study here (and others) has been treated by L. Baramov in [2] [3]. His motivation is the study of plants with special structure such as multiple degrees of freedom. He has an elegant way of embedding singular problems in regular systems which permit factorization. His methods apply to strongly stable systems or to ones which have been strongly stabilized, problems which do not require heavy state estimation. This contrasts strongly with the methods here which focus on estimation and which apply to unstable plants.

## 1.1 Motivation

Exploitation of “cheap sensor”structure may prove to be very important to nonlinear measurement feedback control. There are several reasons for this.

Firstly, almost full state feedback is what actually occurs in many control problems. One example is the currently popular compressor stall problem. Here, in the simplest (Moore-Greitzer) model (cf., [10]), there are three states: the pressure rise in then plenum $\Psi$ , the mass flow $\Phi$ thru the plenum, and a rotating stall parameter $R$. Of these states only $\Phi$ is hard to measure, since this requires putting a sensor in the middle of the jet's flow stream. Thus we can measure two out of three states.

Another example, is that of controlling aircraft control surface actuators. In classical aircraft control problems one has full state information. However, when cost is a major concern as with a missile or with the next generation of expendable unmanned vehicles there are advantages to removing certain problematical sensors such as air probes. Thus all states can measured except one or two.

The second major point is that the online computational burden of state estimation can be vastly less in the cheap sensor versus the corresponding regular case. For linear problems this does not matter much, since computational burdens are usually quite manageable. For nonlinear problems such computational issues are paramount; thus “cheap sensor” algorithms might be essential to making nonlinear measurement feedback “$H^\infty$ control” practical. The body of this paper addresses the theory surrounding this topic.

## 1.2 Contents of this Paper

The layout of the paper is as follows. Section 2 provides a brief review of nonlinear $H_\infty$ control to help set up the framework of the paper. The cheap sensor problem is set up carefully in Section 3, and a RECIPE for the controller construction is given. The RECIPE is applied in Section 4 to the case of linear systems for reasons of illustration and comparison. The computational complexity of the cheap sensor controller is analyzed in Section 5 using a simple finite difference algorithm. Some details of the theory are given in Section 6 which substantiate the RECIPE. An interesting and extreme case is considered in Section 7, and Section 8 verifies the validity of the cheap sensor model by considering an augmented model and a limiting process. Several Appendices contain some detailed calculations and proofs.

# 2 Nonlinear $H_\infty$ Control

Recent advances in nonlinear $H_\infty$ control show that one should use "information state controllers". In this section we state the *regular* nonlinear $H_\infty$ control problem, and briefly summarize some aspects of the information state solution discussed in [18], [15]. In particular, we consider the generalized plant $G$ given by (1.1) in the regular case which is characterized by the first part of the following assumption.

**Assumption 2.1**

$$D_{12}(x)'D_{12}(x) \triangleq E_1(x) > 0 \quad \text{ and } D_{21}(x)\, D_{21}(x)' \triangleq E_2(x) > 0. \tag{2.1}$$

*Assume further for simplicity that $D_{12}$ and $D_{21}$ are independent of $x$.*

Note that these are not the assumptions under which the new results will be obtained, but they are used for the background material of this section. In fact, Assumption 2.1 will be superseded by Assumption 3.1 below.

## 2.1 The Measurement Feedback Case

A measurement feedback controller $K : y \mapsto u$ is said to solve the $H^\infty$ control problem for $G$ provided the closed-loop system (the system $(G, K)$ obtained by interconnecting $G$ and $K$ via $u = K(y)$) is $\gamma$-dissipative and internally stable. The closed-loop system is $\gamma$-*dissipative* if

$$\tfrac{1}{2}\int_0^T |z(s)|^2 ds \le \tfrac{1}{2}\gamma^2 \int_0^T |w(s)|^2 ds + \beta(x) \tag{2.2}$$

where $\xi(0) = x$, for some non-negative function $\beta$ with $\beta(0) = 0$ and every $w \in L_2[0, T]$, for all $T \ge 0$. *Internal stability* means that if $w \in L_2[0, \infty)$ then $u(\cdot), y(\cdot), z(\cdot), \xi(\cdot) \in L_2[0, \infty)$, and consequently $\xi(t) \to 0$ as $t \to \infty$.

We transform the output feedback problem to a new state feedback problem using the information state $p_t(x) = p(x,t)$ defined by the equation

$$\begin{array}{rl} \frac{\partial}{\partial t} p = & -\nabla p(A + B_1 D_{21}'(y - C_2) + B_2 u) + \gamma^{-2}\frac{1}{2}\nabla p B_1 (I - D_{21}' D_{21}) B_1' \nabla p' \\ & -\gamma^2 \frac{1}{2}|y - C_2|^2 + \frac{1}{2}|C_1 + D_{12}u|^2. \end{array} \tag{2.3}$$

In shorthand notation, we write $p_t(\cdot) = p(\cdot, t)$ and regard $p_t$ as the state of a new dynamical system with state equations

$$\dot{p} = F(p, u, y), \tag{2.4}$$

where $F(p,u,y)$ is the nonlinear differential operator defined in (2.3). The state space is an appropriate function space $\mathcal{X}$. It is known ([15, Chapter 3]) that if a controller $K : y \mapsto u$ exists such that the closed-loop is $\gamma$-dissipative, then (2.3) has a solution. If the solution is not smooth, it can be interpreted in the viscosity sense [4], [11].

Solving the control problem for the information state $p_t$ using dynamic programming leads to a nonlinear PDE on the infinite dimensional (Banach) space $\mathcal{X}$, viz.

$$\inf_{u \in \mathbf{R}^m} \sup_{y \in \mathbf{R}^p} \{\nabla W(p)[F(p,u,y)]\} = 0 \tag{2.5}$$

where $\nabla W(p)$ is a linear operator (Frechet derivative). It is known that there exists a value function $W(p)$ solving this equation in a suitable sense, [18], [19], [15]. In general it cannot be expected that $W$ is smooth. However, we assume smoothness to facilitate system-theoretic interpretation, and remark that smoothness issues do not arise in discrete time [18]. For our purposes here it is sufficient to note that a smooth solution of (2.5) defines an output feedback controller $K^*$ via the optimal feedback function obtained by evaluating the infimum in (2.5):

$$u^*(p) = \nabla W(p)[E_1^{-1}(-D_{12}'C_1 + B_2'\nabla p')]. \tag{2.6}$$

This controller is given by

$$K^* \ : \ \begin{cases} \dot{p} = F(p,u,y), \\ u = u^*(p). \end{cases} \tag{2.7}$$

It can be seen that this controller feeds back the information state via the function $u^*(p)$, and the controller dynamics is the PDE (2.3). The *state* of the controller is the information state $p$, a function $p : \mathbf{R}^n \to \mathbf{R}$.

It should be emphasized that this controller construction is *global*; there are no linearization approximations made. Most importantly, when the controller $K^*$ is initialized at a suitable value $p_0$, it produces a $\gamma$-dissipative closed loop, and therefore $K^*$ *is a global solution to the nonlinear $H_\infty$ control problem.* For precise statements and stability results, see [15, Chapter 4]. It is also worth remarking that the information state methodology is quite general, and can be applied to problems with discrete time and state variables, and to stochastic problems.

To implement the controller $K^*$ (given by (2.7)), one must solve the infinite dimensional PDE (2.5) for $W(p)$ to determine the feedback control function $u^*(p)$ via (2.6) (offline), and solve (2.3) for $p_t(x)$ (online). A simplification of this construction occurs when the so-called certainty equivalence assumptions hold, as will be discussed in section 2.3 below. In preparation for this, we describe the state feedback case.

## 2.2 The State Feedback Case

In the idealized case where the controller has perfect state information, we write

$$u = u(x),$$

where $u : \mathbf{R}^n \to \mathbf{R}^m$, and the generalized plant is of the form

$$G \; : \; \left[\begin{array}{c|c|c} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ \hline I_n & 0 & 0 \end{array}\right]. \tag{2.8}$$

If there exists a static state feedback controller $u(x)$ such that the closed-loop system is $\gamma$-dissipative, we know that there exists a storage function $V(x) \geq 0$ which satisfies $V(0) = 0$ and the Hamilton-Jacobi-Bellman-Isaacs (HJBI) equation

$$\nabla V(A - B_2 D'_{12} C_1) + \tfrac{1}{2} \nabla V(\gamma^{-2} B_1 B'_1 - B_2 E_1^{-1} B'_2) \nabla V' + \tfrac{1}{2} C'_1 (I - D_{12} E_1^{-1} D'_{12}) C_1 = 0; \tag{2.9}$$

see, e.g. [28]. The function $V(x)$ need not necessarily be smooth, in which case the PDE (2.9) can be interpreted in the viscosity sense.

Conversely, if (2.9) has a smooth solution $V(\cdot) \geq 0$, $V(0) = 0$, then the state feedback controller

$$u^*_{state}(x) = -E_1^{-1}(D'_{12} C_1(x) + B_2(x)' \nabla V(x)') \tag{2.10}$$

renders the closed loop $\gamma$-dissipative. Since the control law depends on $\nabla V$, we assume that $V(\cdot)$ is differentiable in some sense ($C^1$ with globally Lipschitz derivative is enough). The stability of the closed loop follows from $V(\cdot) \geq 0$ and the zero state detectability of $((A + B_2 u^*_{state})(x), (C_1 + D_{12} u^*_{state})(x))$ (the zero state detectability implies positive definiteness of $V$); see, e.g., [28].

## 2.3 Certainty Equivalence

The optimal information state controller $K^*$ given by (2.7) for the measurement feedback problem simplifies to the certainty equivalence controller of [5] under the following *certainty equivalence conditions.* Assume the state feedback value function $V(x)$ and the information state $p_t(x)$ are smooth, and assume that the *minimum stress state estimate*

$$\bar{\xi}(t) = \arg\max_x \{p(x,t) + V(x)\} \tag{2.11}$$

is unique.

Recall that $u^*_{state}(x)$ denotes the optimal state feedback control, given by (2.10). Then the certainty equivalence controller exists and is given by

$$u^{ce}(t) = u^*_{state}(\bar{\xi}(t)). \tag{2.12}$$

Furthermore, the certainty equivalence controller and the information state controller $K^*$ given by (2.7) coincide:

$$u^*(p_t) = u^{ce}(t).$$

See [5], [19], [15, Section 7.3].

The implication of this is that *under the certainty equivalence assumptions*, the information state controller (which from above we know is a global solution to the $H_\infty$ control problem) simplifies to

$$K^* \ : \ \begin{cases} \dot{p} \ = F(p, u, y), \\ u \ = u^*_{state}(\bar{\xi}(t)), \end{cases} \tag{2.13}$$

a formula which requires the solution of two $n$-dimensional PDEs ((2.9) for $V(x)$ to obtain $u^*_{state}(x)$ via (2.10) (offline) and (2.3) for $p_t(x)$ (online).

In general, it is known that certainty equivalence can fail, in which case one must solve the infinite dimensional PDE (2.5) for $W(p)$ (offline), and implement (2.7) by solving (2.3) for $p_t(x)$ (online).

# 3 The Controller Construction for the Cheap Sensor Problem

We turn now to the *cheap sensor case*, where the state space will have a decomposition as a direct sum of space $X_\kappa$ of *directly measured states* and $X_\mu$ some subspace complementary to $X_\kappa$. Our main interest is the dramatic reduction in computational complexity of solving for the information state $p_t(x)$ which occurs in the cheap sensor case.

We next give precise definitions and formal notation, followed by a RECIPE for a controller which solves the $H^\infty$ problem.

## 3.1 The Formal Setup

Let $P_\kappa$ denote the projection onto $X_\kappa$ along $X_\mu$, and let $P_\mu$ denote the projection onto $X_\mu$ along $X_\kappa$. Think of the state of $G$ decomposed as $x = (x_\kappa, x_\mu)$, with output $y = (x_\kappa, y_\mu)$ where $y_\kappa = x_\kappa = P_\kappa x$ are the perfectly measured states and $y_\mu$ are imperfect measurements of the plant state $x$. The disturbance/command input is denoted $(v, w)$, where the dimension of $w$ equals that of $y_\mu$. The plant $G$ with cheap sensor structure

(1.3) is given by

$$
G: \quad \begin{cases} \dot{\xi} = A(\xi) + B_1(\xi)w + \tilde{B}_1(\xi)v + B_2(\xi)u \\ z = C_1(\xi) + D_{12}(\xi)u \\ y = \begin{bmatrix} y_\kappa \\ y_\mu \end{bmatrix} = \begin{bmatrix} \xi_\kappa \\ C_{2\mu}(\xi) \end{bmatrix} + \begin{bmatrix} 0 \\ w \end{bmatrix} \end{cases} \tag{3.1}
$$

and we often write $y$ as

$$
y = C_2(\xi) + D_{21} \begin{bmatrix} v \\ w \end{bmatrix}. \tag{3.2}
$$

In conventional notation the plant state equation is

$$
\dot{\xi} = A(\xi) + \begin{bmatrix} \tilde{B}_1(\xi) & B_1(\xi) \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} + B_2(\xi)u.
$$

Here, $\xi(t) \in \mathbf{R}^n$, $y(t) \in \mathbf{R}^p$, $z(t) \in \mathbf{R}^r$, $u(t) \in \mathbf{R}^m$, while $v(t) \in \mathbf{R}^q$, $w(t) \in \mathbf{R}^s$ The input $w$ includes sensor noise, while $v$ does not (since $v$ is not directly sensed). Command signals would typically be contained in $w$, and the total number of disturbances plus command signals is $\#Dis = s + q$. The total number of sensors is $\#Sens = p = \dim X_\kappa + s$. Miscellaneous relationships are $\#Sens - \#Dis = \dim X_\kappa - q$ and $n - \#Sens + \#Dis = \dim X_\mu + q$.

In the cheap sensor case, in place of Assumption 2.1 we use the following assumption throughout the paper. (As opposed to other, more particular, assumptions to follow, this assumption will be implicitly assumed throughout without futher reference to it.)

**Assumption 3.1**

$$
D_{12}(x)'D_{12}(x) \triangleq E_1(x) > 0, \quad D_{21} = \begin{bmatrix} 0 & 0 \\ 0 & I_\mu \end{bmatrix}, \text{ and } C_2(x) = \begin{bmatrix} x_\kappa \\ C_{2\mu}(x) \end{bmatrix}. \tag{3.3}
$$

The following signal spaces will be used:

$$
L_2 = L_2([0,\infty), \mathbf{R}^q), \ L_{2,T} = L_2([0,T], \mathbf{R}^q),
$$

$$
\text{and } L_{2,loc} = \{w : [0,\infty) \to \mathbf{R}^q : \ w \in L_2[0,T] \quad \forall T \in [0,\infty)\},
$$

where the dimension of the range space will not be stated explicitly but inferred from the context. It will be assumed throughout that all inputs $(u, v, w)$ will be in $L_{2,loc}$.

## 3.2 Controller Construction

As will be described in Section 6 below, the information state methods can be applied to the cheap sensor case even though the matrix $D_{21}$ is not full rank. Recall that the information state $p$ is in general a function $p : \mathbf{R}^n \to \mathbf{R}$. In the cheap sensor case, this function can be replaced by the pair $(y_\kappa, \tilde{p})$, where

$$\tilde{p} : X_\mu \to \mathbf{R}.$$

This is extremely important because often $\dim X_\mu << n$ (where we recall $n$ is the dimension of the full state space), and so computation of the information state becomes feasible, as will be discussed in Section 5.

### 3.2.1 RECIPE for the Controller

**Controller dynamics.** The states of the information state controller $K^*$ are functions $\tilde{p} : X_\mu \to \mathbf{R}$ with dynamics a *cut down information state PDE* of the form

$$\dot{\tilde{p}} = \tilde{F}(\tilde{p}, u, y), \tag{3.4}$$

with suitable initialization. This is a PDE on the lower dimensional space $X_\mu$. It will be described in Section 3.2.3, and then in more detail in Section 6.1.

**Controller output.**

**General case.** In general the control is a function of the pair $(y_\kappa, \tilde{p})$, i.e.

$$u = u^*(y_\kappa, \tilde{p}). \tag{3.5}$$

**Certainty equivalence case.** Under the certainty equivalence assumptions (Section 2.3), the control is given by

$$\begin{aligned} \mathbf{u}^*(y_\kappa, \tilde{p}) &= u^*_{state}(y_\kappa, \bar{x}_\mu(y_\kappa, \tilde{p})) \\ &= -E_1^{-1}(D'_{12}C_1(\bar{x}(y_\kappa, \tilde{p})) + B_2(\bar{x}(y_\kappa, \tilde{p}))'\nabla_x V(\bar{x}(y_\kappa, \tilde{p}))'), \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} \bar{x}_\mu(y_\kappa, \tilde{p}) &= \operatorname{argmax}_{x_\mu \in X_\mu}\{\tilde{p}(x_\mu) + V(y_\kappa, x_\mu)\} \\ \bar{x}(y_\kappa, \tilde{p}) &= (y_\kappa, \bar{x}_\mu(y_\kappa, \tilde{p})). \end{aligned} \tag{3.7}$$

(Recall that $V(x)$ is the state feedback function solving (2.9), with $[\tilde{B}_1, B_1]$ replacing $B_1$, and $u^*_{state}(x)$ is obtained using (2.10).)

This is discussed in Section 6.2.

Theorem 6.6 in Section 6.2 show that under certainty equivalence Assumption 6.5, the RECIPE produces a $\gamma$-dissipative controller.

To implement this controller RECIPE (under certainty equivalence), one first needs to solve for the state feedback control $u^*_{stae}(x)$, or approximate it. This would involve the solution of a PDE in $\mathbf{R}^n$, or some other approximation or heuristic approach. This occurs *offline*. The cut down information state PDE would be solved *online*. In the cheap sensor context, this online computation is practically feasible with 2000 technology.

### 3.2.2 Special Cases

We illustrate the structure of the cut down controller in the case of certainty equivalence:

(i) $x = x_\kappa$, $y = y_\kappa = x_\kappa$. This is the case of full state feedback

$$u = \mathbf{u}^*(y_\kappa) = u^*_{state}(y_\kappa).$$

No states are estimated ($\bar{x} = y_\kappa$).

(ii) $x = x_\mu$, $y = y_\mu$. This is the case of imperfect measurements, with full information state feedback

$$u = \mathbf{u}^*(\tilde{p}) = u^*_{state}(\bar{x}_\mu).$$

All states must be estimated ($\bar{x} = \bar{x}_\mu$).

(iii) $x = (x_\kappa, x_\mu)$, $y = (x_\kappa, y_\mu)$ as described in section 3, the main case of this paper. The fact that *some* states are perfectly measured means that only the other states need be estimated, and so the controller has the form

$$u = u^*_{state}(y_\kappa, \bar{x}_\mu).$$

Only some states are estimated ($\bar{x}_\mu$).

### 3.2.3 The Cut Down Information State PDE

Our main case of interest is when $\#Sens \leq \#Dis$. In this case the the *cut down information state PDE* (equation (3.4) above) is as follows:

$$\begin{cases} 0 = \frac{\partial}{\partial t}\tilde{p}_t - \sup\limits_{v \in \mathcal{V}_t(x_\mu)} \Big\{ -\nabla_\mu \tilde{p}_t P_\mu [A + B_1(y_\mu(t) - C_{2\mu}) + \widetilde{B}_1 v + B_2 u(t)] \\ \qquad\qquad + \frac{1}{2}|C_1 + D_{12}u(t)|^2 - \gamma^2[|v|^2 + |C_{2\mu} - y_\mu(t)|^2] \Big\} \\ \tilde{p}_0(x_\mu) = \phi_{0\mu}(x_\mu) \end{cases} \tag{3.8}$$

where $\mathcal{V}_t(x_\mu)$ is given by

$$\mathcal{V}_t(x_\mu) \doteq \left\{ v \in (P_\kappa \widetilde{B}_1)^{-1}[\dot{y}_\kappa - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2 u(t))] \right\}. \tag{3.9}$$

Here $(P_\kappa \widetilde{B}_1)^{-1}$ is the set–valued inverse ($(P_\kappa \widetilde{B}_1)^{-1} b = \{a : P_\kappa \widetilde{B}_1 a = b\}$), and we shall assume that it is very well behaved, see Assumption 6.7. This PDE is initialized at $\phi_{0\mu} : X_\mu \to \mathbf{R}$ (see section 6.1). A proof that the (viscosity) solution of this PDE yields the cut down information state appears in Section 6.3 and corresponding appendix.

The supremum in (3.8) looks abstract but it can be computed explicitly. Since $\mathcal{V}_t$ is an affine linear space, the maximizer, $v^*$, exists and it can be shown that it is given by the formula

$$\begin{aligned} v^* = &(I - P)(P_\kappa \widetilde{B}_1)^{-R}[\dot{y}_\kappa - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2 u(t))] \\ &- \tfrac{1}{2\gamma^2} P (P_\mu \widetilde{B}_1)^T \nabla_\mu p_t \end{aligned} \tag{3.10}$$

where $(P_\kappa \widetilde{B}_1)^{-R}$ is a right inverse (i.e. $(P_\kappa \widetilde{B}_1)(P_\kappa \widetilde{B}_1)^{-R} = I$) and $P$ is the projection onto the null space of $P_\kappa \widetilde{B}_1$ (i.e. $(P_\kappa \widetilde{B}_1) P r = 0$ for all $r$). Details are given in Appendix D. Further, in the special case where $P_\kappa \widetilde{B}_1(y_\kappa(t), x_\mu)$ is invertible for all $x_\mu, t$, one obtains

$$v^* = (P_\kappa \widetilde{B}_1)^{-1}[\dot{y}_\kappa - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2 u(t))]$$

(where, abusing notation, we use $(P_\kappa \widetilde{B}_1)^{-1}$ in this last line to indicate the standard, not set–valued, inverse).

In Theorem 6.8 below we show that the cut down information state is a viscosity solution to the cut down PDE (3.8). This interpretation is useful since solutions to such PDE are not smooth in general. However, this *viscosity* solution technicality need not concern us here for three reasons: (i) the information state can be interpreted in an integral form which obviates the need for derivatives, (ii) the viscosity intepretation is not necessary for numerical solution, and (iii) in some cases, such as for linear systems, the cut down information state is smooth, and viscosity solutions are not needed.

The integral form which we take as the *definition* of the *cut down information state* is as follows: for fixed $u, y \in L_{2,loc}$, define

$$\begin{aligned} \tilde{p}_t(x_\mu) \; = \sup_{\tilde{v}(\cdot) \in \mathcal{V}(u, y, (y_\kappa(t), x_\mu), t)} &\{ \phi_{0\mu}(\xi_\mu(0)) + \tfrac{1}{2} \int_0^t \{ |C_1(y_\kappa(s), \xi_\mu(s)) + D_{12} u(s)|^2 \\ &- \gamma^2 (|\tilde{v}(s)|^2 + |C_{2\mu}(y_\kappa(s), \xi_\mu(s)) - y_\mu(s)|^2) \} \, ds \}, \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} \mathcal{V}(u(\cdot), y(\cdot), x, t) = \{ v \in L_{2,loc} : \; &\text{there exists a solution } \xi(\cdot) \text{ of (3.13) consistent with} \\ &u(\cdot), y(\cdot) \text{ on } [0, t] \text{ with } \; y_\kappa(\cdot) = \xi_\kappa(\cdot) \text{ and } \xi(t) = x \}, \end{aligned} \tag{3.12}$$

and $\xi(\cdot) = (y_\kappa(\cdot), \xi_\mu(\cdot))$ has dynamics

$$
\begin{aligned}
\dot{y}_\kappa = \dot{\xi}_\kappa &= P_\kappa \left( A(y_\kappa, \xi_\mu) + B_1(y_\kappa, \xi_\mu)(y_\mu - C_{2\mu}(y_\kappa, \xi_\mu)) + \widetilde{B}_1(y_\kappa, \xi_\mu)\tilde{v} + B_2(y_\kappa, \xi_\mu)u \right) \\
\dot{\xi}_\mu &= P_\mu \left( A(y_\kappa, \xi_\mu) + B_1(y_\kappa, \xi_\mu)(y_\mu - C_{2\mu}(y_\kappa, \xi_\mu)) + \widetilde{B}_1(y_\kappa, \xi_\mu)\tilde{v} + B_2(y_\kappa, \xi_\mu)u \right)
\end{aligned}
\tag{3.13}
$$

with terminal conditions $\xi_\mu(t) = x_\mu$, $\xi_\kappa(t) = x_\kappa$ and constraint $\xi_\kappa(s) = y_\kappa(s)$, $0 \leq s \leq t$. The initial condition is

$$\tilde{p}_0(x_\mu) = \phi_{0\mu}(x_\mu). \tag{3.14}$$

This definition will be supported in section 6.1.

This integral form also makes sense in the case $\#Sens > \#Dis$, which will be discussed in Section 7.

**Remark 3.2** Equations (3.13) consist of a differential equation for the state component $\xi_\mu(t)$ and a *constraint* equation involving the perfectly measured state component $y_\kappa(t) = \xi_\kappa(t)$. This constraint is a reflection of the loss of rank of $D_{21}$ in the cheap sensor case relative to the regular case.

# 4 Linear Systems

As mentioned in the introduction, the paper [27] obtains reduced order $H_\infty$ compensators for cheap sensor linear systems problems. The techniques used there are special to linear systems. In this section we apply our general nonlinear methodology to an example linear systems case, and recover explicit compensator formulas with analogous form to the reduced order compensator in [27]. This shows that the results in this paper are consistent with known linear results. We assume:

**Assumption 4.1** *The plant $G$ given by (1.3) is a linear system, so all the entries are matrices as is standard:*

$$
G: \quad \begin{cases}
\dot{\xi} = A\xi + B_1 w + \tilde{B}_1 v + B_2 u \\
z = C_1 \xi + D_{12} u \\
y = \begin{bmatrix} y_\kappa \\ y_\mu \end{bmatrix} = \begin{bmatrix} \xi_\kappa \\ C_{2\mu}\xi \end{bmatrix} + \begin{bmatrix} 0 \\ w \end{bmatrix}
\end{cases}
\tag{4.1}
$$

*Asssume $\#Sens = \#Dis$, and for simplicity the projection matrix $P = 0$ in equation (3.10), so that $\tilde{B}_{1\kappa}$ (see (4.2) below) is an invertible matrix.*

We partition the matrices relative to $x = (x_\kappa, x_\mu)$, so that

$$A = \begin{bmatrix} A_{\kappa\kappa} & A_{\kappa\mu} \\ A_{\mu\kappa} & A_{\mu\mu} \end{bmatrix}, \quad B_1 = \begin{bmatrix} B_{1\kappa} \\ B_{1\mu} \end{bmatrix}, \quad \tilde{B}_1 = \begin{bmatrix} \tilde{B}_{1\kappa} \\ \tilde{B}_{1\mu} \end{bmatrix}, \quad B_2 = \begin{bmatrix} B_{2\kappa} \\ B_{2\mu} \end{bmatrix}, \quad C_{2\mu} = \begin{bmatrix} C_{2\mu,\kappa} & C_{2\mu,\mu} \end{bmatrix}. \tag{4.2}$$

Write, using the constraint equation (3.13),

$$\begin{aligned} \tilde{y} &= \dot{y}_\kappa - A_{\kappa\kappa} y_\kappa - B_{1\kappa} y_\mu - B_{1\kappa} C_{2\mu,\kappa} y_\kappa - B_{2\kappa} u \\ &= (A_{\kappa\mu} - B_{1\mu} C_{2\mu,\mu}) x_\mu + \tilde{B}_{1\kappa} v^*(x_\mu, t) \end{aligned} \tag{4.3}$$

(c.f. [27, eq. (4.4)]) so that

$$v^*(x_\mu, t) = (B_{1\kappa})^{-1} [\tilde{y}(t) - (A_{\kappa\mu} - B_{1\kappa} C_{2\mu,\mu}) x_\mu] \tag{4.4}$$

an affine function of the unknown state $x_\mu$. The term $(B_{1\kappa})^{-1}$ here denotes matrix inverse which exists under our current assumptions.

With this notation and (3.10), the information state PDE (3.8) can be written

$$\begin{aligned} \tfrac{\partial}{\partial t} p_t^{y_\kappa}(x_\mu) &= -\nabla_\mu p_t^{y_\kappa}(x_\mu) \cdot [A_{\mu\kappa} y_\kappa + A_{\mu\mu} x_\mu + B_{1\mu}(y_\mu - C_{2\mu,\kappa} y_\kappa - C_{2\mu,\mu} x_\mu) \\ &\qquad + B_{2\mu} u + \tilde{B}_{1\mu} v^*(x_\mu)] \\ &\quad + \tfrac{1}{2} |C_{1\kappa} y_\kappa + C_{1\mu} x_\mu + D_{12} u|^2 - \gamma^2 \tfrac{1}{2} |b y_\mu - C_{2\mu,\kappa} y_\kappa - C_{2\mu,\mu} x_\mu|^2 - \gamma^2 \tfrac{1}{2} |v^*(x_\mu)|^2 \end{aligned} \tag{4.5}$$

We take the initial function $\phi_0$ to be of the form

$$\phi_0(y_\kappa, x_\mu) = -\gamma^2 \tfrac{1}{2} y_\kappa' Y_{I\kappa} y_\kappa - \gamma^2 x_\mu' Y_{\mu 0}^{-1} x_\mu, \tag{4.6}$$

with $Y_{I\kappa}$, $Y_{\mu 0}$ symmteric positive definite (see Assumption 6.2 below). Then the PDE (4.5) on $X_\mu$ can be solved as a quadratic form in $x_\mu$:

$$p_t^{y_\kappa}(x_\mu) = -\gamma^2 \tfrac{1}{2} (x_\mu - \hat{x}_\mu(t))' Y_\mu(t)^{-1} (x_\mu - \hat{x}_\mu(t)) + \phi_\mu(t) \tag{4.7}$$

The symmetric non-negative definite matrix $Y_\mu(t)$ solves the Riccati differential equation

$$\dot{Y}_\mu = \bar{A} Y_\mu + Y_\mu \bar{A}' + Y_\mu (\gamma^{-2} C_{1\mu}' C_{1\mu} - \bar{C}_2' \bar{C}_2) Y_\mu \tag{4.8}$$

with initial condition $Y_\mu(0) = Y_{\mu 0}$, and the vector $\hat{x}_\mu(t) \in X_\mu$ solves the differential equation

$$\begin{aligned} \dot{\hat{x}}_\mu &= [\bar{A} + Y_\mu (\gamma^{-2} C_{1\mu}' C_{1\mu} - \bar{C}_2' \bar{C}_2)] \hat{x}_\mu \\ &\quad (\bar{B}_1 + Y_\mu \bar{C}_2' \bar{E}) \bar{y} + (\bar{B}_2 + \gamma^{-2} Y_\mu C_{1\mu} D_{12}) u + \gamma^{-2} Y_\mu C_{1\mu}' C_{1\kappa} y_\kappa \end{aligned} \tag{4.9}$$

with $\hat{x}_\mu(0) = 0$. Here,

$$\bar{A} = A_{\mu\mu} - B_{1\mu} C_{2\mu,\mu} - \tilde{B}_{1\mu} (\tilde{B}_{1\kappa})^{-1} (A_{\kappa\mu} - B_{1\kappa} C_{2\mu,\mu})$$

$$\bar{E} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & -C_{2\mu,\kappa} \end{bmatrix}$$

$$\bar{B}_1 = \left[ \tilde{B}_{1\mu}(\tilde{B}_{1\kappa})^{-1},\ B_{1\mu},\ A_{\mu\kappa} - B_{1\mu}C_{2\mu,\kappa} \right]$$

$$\bar{B}_2 = B_{2\mu}$$

$$\bar{C}_2 = \begin{bmatrix} A_{\kappa\mu} - B_{1\kappa}C_{2\mu,\mu} \\ C_{2\mu,\mu} \end{bmatrix}$$

and

$$\bar{y} = \begin{bmatrix} \tilde{y} \\ y_\mu \\ y_\kappa \end{bmatrix}$$

The state feedback value function $V(x) = V(x_\kappa, x_\mu)$ is a quadratic form with matrix $X$ solving the $n$-dimensional PDE (2.9), so

$$V(y_\kappa, x_\mu) = \tfrac{1}{2}y'_\kappa X_{\kappa\kappa} y_\kappa + y'_\kappa X_{\kappa\mu} x_\mu + \tfrac{1}{2}x'_\mu X_{\mu\mu} x_\mu, \tag{4.10}$$

where

$$X = \begin{bmatrix} X_{\kappa\kappa} & X_{\kappa\mu} \\ X'_{\kappa\mu} & X_{\mu\mu} \end{bmatrix}.$$

Evaluating the certainty equivalence formulas (3.7), (3.6) for the $H_\infty$ controller, we have

$$\bar{x} = (I - \gamma^{-2}Y_\mu X_{\mu\mu})^{-1}(\hat{x}_\mu + \gamma^{-2}Y_\mu X'_{\kappa\mu} y_\kappa), \tag{4.11}$$

and the control is given by

$$\begin{aligned} u^* &= -E_1^{-1}(D'_{12}C_{1\kappa} + B'_{2\kappa}X_{\kappa\kappa} + B'_{2\mu}X'_{\kappa\mu})y_\kappa \\ &\quad -E_1^{-1}(D'_{12}C_{1\mu} + B'_{2\mu}X_{\mu\mu} + B'_{2\kappa}X_{\kappa\mu})\bar{x}_\mu, \end{aligned} \tag{4.12}$$

a formula involving the exactly measured states $y_\kappa$ and the estimate $\hat{x}_\mu$.

The controller formulas as presented depend on the time derivative $\dot{y}_\kappa$ via $\bar{y}$ in equation (4.9). However, as in [27] it is possible to remove this dependence and evaluate the controls using the change of variables

$$\tilde{x}_\mu = \hat{x}_\mu - [\tilde{B}_{1\mu}(\tilde{B}_{1\kappa})^{-1} + Y_\mu(A_{\kappa\mu} - B_{1\kappa}C_{2\mu,\mu})]y_\kappa. \tag{4.13}$$

It can be checked that $\tilde{x}_\mu$ can be obtained by integrating an ODE of the form

$$\dot{\tilde{x}}_\mu = F_\mu \tilde{x}_\mu + G_\mu \begin{bmatrix} u \\ y_\mu \\ y_\kappa \end{bmatrix}, \tag{4.14}$$

for suitable matrices $F_\mu$, $G_\mu$.

# 5 Computational Complexity

In this section, we will indicate two things. The first regards reduction in storage requirements and the second regards the fact that the operation counts required to perform the information state controller computations are *hugely* reduced by the cheap sensor conditions and methods described in this paper. *The chief reason for this huge complexity reduction in the case* $\dim X_\mu << n$ *is that* $G_x^{\dim X_\mu} <<<< G_x^n$. *For example, if there are 2 unmeasured states, then* $G_x^2$ *grid points are required which as we shall see with* $G_x = 50$ *is managable with today's technology.*

The reduction in storage requirements per update is easy to analyze. Clearly, it scales with the total number $G_x^n$ of grid points, so this gives a crude order of magnitude (the storage requirements of the information state). For, say, $n = 12$ and $G_x = 50$ this is of the unmanagable size $50^{12} \approx 10^{17}$. However, for only 2 unmeasured states, this is $50^2 \approx 10^4$ which is not demanding. As we shall sees the CPU demands far exceed this and so henceforth we ignore storage and turn to counting operations per update. Note that we do not consider the storage requirements for the information control feedback function.

The number of operations to update the information state at each time step takes more effort to compute. As a basis for evaluating the computational complexity of solving for the cut down information state, we consider a simple finite difference scheme (Section 5.1). Finite difference and finite element schemes are widely used for numerical solution of PDEs. In the context of control problems, we refer the reader to the publications [8], [9], [21], [11], [4] and the many references therein. The numerical scheme of Section 5.1 is not claimed to be optimal or necessarily ready for production use. Since the purpose of this discussion is the operation count, we do not discuss convergence proofs or assumptions; one may consult the cited references for further information. The detailed operation counting associated with this scheme appears in Appendix A.

The section concludes with some problem whose dimensions sit at the edge for which controller implementing computations can be performed with the computing equipment available today.

## 5.1 Finite Differences

We assume $\#Sens \leq \#Dis$, $\dim(X_\kappa) > 0$, and $\dim(X_\mu) < n$. In particular, we will be mainly interested with the situation where $\dim X_\mu << n$.

The integrated form of information state equation can be written as (see Theorem B.4)

$$\tilde{p}_t(x_\mu) = \sup_{v(\cdot)\in\mathcal{V}_t(x_\mu)} \{\tilde{p}_{t-\Delta}(\xi_\mu(t-\Delta)) + \int_{t-\Delta}^{t} L(y_\kappa(s), \xi_\mu(s), u(s), y_\mu(s), v(s))ds\} \qquad (5.1)$$

where $\Delta > 0$, $\mathcal{V}_t(x_\mu)$ is given by (3.9), and $\xi_\mu(\cdot)$ solves

$$\dot{\xi}_\mu = b_\mu(y_\kappa, \xi_\mu, u, y_\mu, v) \ \text{ on } [t-\Delta, t], \quad \xi_\mu(t) = x_\mu \tag{5.2}$$

Here,

$$\begin{aligned} Q(x, y_\mu) &= A(x) + B_1(x)(y_\mu - C_{2\mu}(x)) + B_2(x)u \\ b(x, u, y_\mu, v) &= Q(x, y_\mu) + \tilde{B}_1(x)v + B_2(x)u \\ b_\mu(y_\kappa, x_\mu, u, y_\mu, v) &= P_\mu b((y_\kappa, x_\mu), u, y_\mu, v) \\ L(x, u, y_\mu, v) &= |C_1(x) + D_{12}u|^2 - \gamma^2(|v|^2 + |y_\mu - C_{2\mu}(x)|^2) \\ L(y_\kappa, x_\mu, u, y_\mu, v) &= L((y_\kappa, x_\mu), u, y_\mu, v) \end{aligned} \tag{5.3}$$

Let $G_x$ denote the number of grid points per dimension of $X_\mu$ used in the spatial discretization. Let

$$\Delta \ = \ \text{time discretization step size}$$

and

$$\delta = \ \text{spatial discretization step size.}$$

We approximate $\tilde{p}_t(x)$ in a region $\mathcal{G}_x \subset X_\mu$ of "size" $\delta G_x$, and denote the set of grid points by $\mathcal{G}_x^\delta$. The number of points in $\mathcal{G}_x^\delta$ is $G_x^{\dim X_\mu}$. The nearest neighbors (in the $X_\mu$ space) to a point $x \in \mathcal{G}_x^\delta$ are given by:

$$\mathcal{N}(x_\mu) = \{x_\mu : \ x_\mu \pm \delta e_i, \ i = 1, \ldots, \dim(X_\mu)\} \subset X_\mu$$

where $e_i$ is the $i$-th unit vector in $\mathbf{R}^n$ (see Figure 1).

We use (5.1) to describe a simple finite difference algorithm. $\tilde{p}_t(x_\mu)$ will be approximated by a piecewise affine function $\breve{p}_t(x_\mu)$ defined by the values it takes on the grid; we describe the procedure for computing the values for $x \in \mathcal{G}_{x_\mu}^\delta$. The approximation will be calculated a time $t = N\Delta$ recursively from its value at time $t - \Delta = (N-1)\Delta$ and including new measurement information.

Now for $\Delta$ small,

$$\begin{aligned} \xi(t-\Delta) &= (y_\kappa(t-\Delta), \xi_\mu(t-\Delta)) \\ &\approx (y_\kappa(t), \xi_\mu) - \Delta b((y_\kappa, x_\mu), u, y_\mu, v), \end{aligned}$$

so

$$\xi_\mu(t-\Delta) \approx x_\mu - \Delta b_\mu(y_\kappa, x_\mu, u, y_\mu, v). \tag{5.4}$$

Note that $\xi(t) \in y_\kappa(t) + X_\mu$, while $\xi(t-\Delta) \in y_\kappa(t-\Delta) + X_\mu$. The value $\xi_\mu(t-\Delta) = P_\mu \xi(t-\Delta) \in X_\mu$ need not necessarily belong to the grid in $X_\mu$; however, we assume it is in the convex hull co$\mathcal{N}(x_\mu)$ of $\mathcal{N}(x_\mu)$.

Define $\rho = \delta/\Delta$ and

$$\lambda(x_\mu, \xi_\mu, u, y, v) = \begin{cases} b^{\mp}_{\mu,i}(P_\kappa y, x_\mu, u, y_\mu, v)/\rho & \text{if } \xi_\mu = x_\mu \pm \delta e_i \ (i = 1, \ldots, \dim(X_\mu)) \\ 1 - (\sum_{i=1}^{\mu} |b_{\mu,i}(P_\kappa y, x_\mu, u, y_\mu, v)|)/\rho & \text{if } \xi_\mu = x_\mu \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

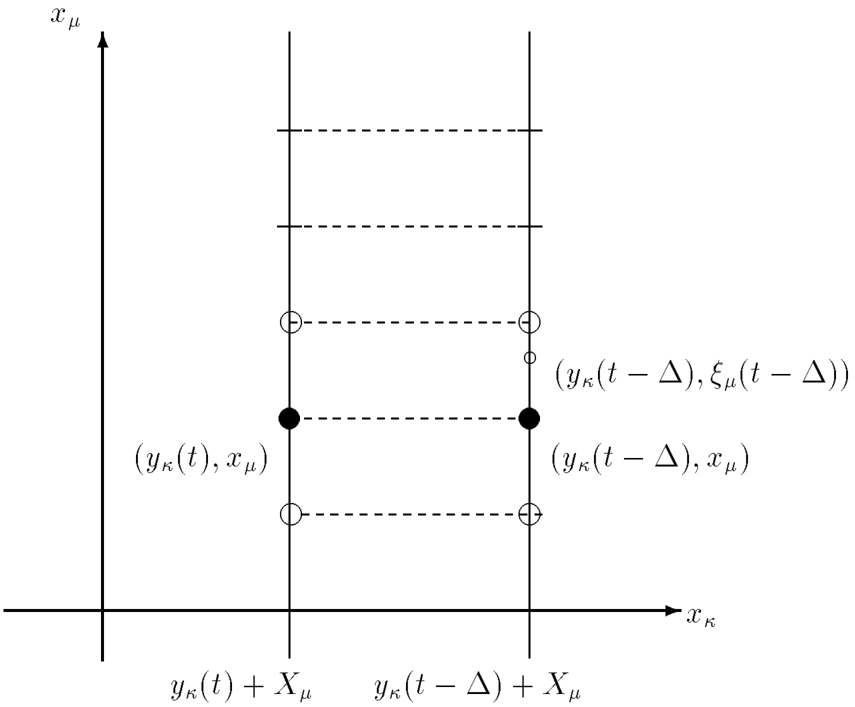


Figure 1: Location of grid $\mathcal{G}^\delta_{x_\mu} \subset X_\mu$ at times $t$ and $t-\Delta$ showing points $(y_\kappa(t), x_\mu)$ and $(y_\kappa(t-\Delta), x_\mu)$ (solid dots), nearest neighbors $\mathcal{N}(x_\mu)$ (empty dots), and $(y_\kappa(t-\Delta), \xi_\mu(t-\Delta))$.

for $\xi_\mu \in \mathcal{N}(x_\mu)$, $x_\mu \in \mathcal{G}_\mu$. Then it can be checked that

$$\sum_{\xi_\mu \in \mathcal{N}(x_\mu)} \xi_\mu \lambda(x_\mu, \xi_\mu, u, y, v) = x_\mu - \Delta b_\mu(P_\kappa y, x_\mu, u, y_\mu, v) \in \mathrm{co}\mathcal{N}(x_\mu) \subset X_\kappa.$$

The algorithm is as follows:

$$\breve{p}_t(x_\mu) = \sup_{v \in \mathcal{V}_t(x_\mu)} \{ \sum_{\xi_\mu \in \mathcal{N}(x_\mu)} \lambda(x_\mu, \xi_\mu, u(t), y_\mu(t), v) \breve{p}_{t-\Delta}(\xi_\mu) + \Delta L(y_\kappa(t), x_\mu, u(t), y(t), v) \} \tag{5.6}$$

for $x_\mu \in \mathcal{G}_{x_\mu}$, $t = N\Delta$, $N \geq 0$ (plus suitable initialization).

*This algorithm is a finite difference scheme on the space $X_\mu$, the dimension of which is relatively small compared to the dimension $n$ of the whole space $\mathbf{R}^n$.* It can be seen that this algorithm involves function evaluations, maximization, and miscellaneous arithmetic operations. Detailed operation counting associated with this algorithm appears in Appendix A, and the outcomes are discussed in the following sections.

## 5.2 Ballpark FLOP Counts

Now we give an idea of the cost of this algorithm a few special cases where there are many sensors. Appendix A derives the costs in detail. Even though we treat a PDE scheme a major cost is that of the evaluating the coefficients of the PDE and we discuss that first.

### 5.2.1 The Cost of Function Evaluations

Costs vary greatly with what type of function is being evaluated. Thus the only general things one can say are very complicated. The vector field

$$Q := A + B_1(y_\mu - C_{2\mu}) + B_2 u \tag{5.7}$$

is important to the algorithm and in order to give an indication of what is involved, consider simply evaluating the $A$ vector field. If, for instance, we have a linear vector field $x \mapsto Ax$, then the cost of evaluating it at $x_0$ is $n^2$. However, in our problem we can be much more efficient. Once we are on an $X_\kappa$ = known slice, we use,

$$Ax = A\begin{pmatrix} 0 \\ x_\mu \end{pmatrix} + A\begin{pmatrix} x_\kappa \\ 0 \end{pmatrix}$$

where $A\begin{pmatrix} x_\kappa \\ 0 \end{pmatrix}$ is evaluated only once for all $x_\mu$ on the slice! Thus evaluating $Ax$ on the $X_\mu$ slice costs

$$2n(\dim X_\mu)G_x^{\dim X_\mu} + 2n(\dim X_\kappa)$$

Thus each function evaluation on $X_\mu$ adds approximately

$$E_A = 2n(\dim X_\mu) \tag{5.8}$$

operations per grid point to the cost. This gives the main idea from which one can analyse more complicated vector fields. For example, piecewise linear vector fields give approximately the same count. Polynomials can lead to powers of $\dim X_\mu$.

### 5.2.2 FLOP Count for Each Information State Update

Performing operation counts for all the terms which need to be computed in each update of the information state algorithm (see Appendix A), one finds that:

The FLOP count per update (in the case $q = n$) is approximately

$$E_Q + E_L + \Big\{ 2n^3 + 3n^2 + n(7D_\mu^2 + 4D_\mu + 1) + 2D_\mu^3 + D_\mu + [D_\mu^3 + 4D_\mu^2 + 4D_\mu + 5]2^{D_\mu} \Big\} M_{D_\mu} \tag{5.9}$$

where $D_\mu = \dim X_\mu$, $E_Q$ and $E_L$ are the FLOPS required for the computation of $Q$ and $\Delta L$ (defined in (5.7) and (5.3) respectively), and $M_{D_\mu}$ is a multiplier (with $M_1 = 1$ and $M_2 = 2$).

A great savings occurs when $P_\kappa \widetilde{B}_1(x)$ does not depend on $x_\mu$, in which case the FLOP count per update (in the case $q = n$) is approximately

$$E_Q + E_L + \Big\{ 3n^2 + n(2D_\mu^2 + 4D_\mu) + 2D_\mu^3 + 5D_\mu + [D_\mu^3 + 4D_\mu^2 + 4D_\mu + 5]2^{D_\mu} \Big\} M_{D_\mu}. \tag{5.10}$$

### 5.2.3 A Few Illustrative Cases

We will indicate some rough FLOP counts for specific problem dimensions. These are based on the detailed counts that appear in Appendix A. The counts are only for the propagation of the information state $\tilde{p}_t$.

To compute the control law from the information state one must have computed the solution $V$ to the HJBI equation offline and have stored it. Since this could not be done by storing the values of $V$ or $\nabla V$ on a rectangular grid on $\mathbf{R}^n$, unless $n$ is relatively small, the designer must produce an algebraic representation for $V$ (such as piecewise linear or piecewise quadratric). When the closed loop system runs and we must produce a controller command at a particular time we must evaluate $V$ and $\nabla V$. We denote the cost of evaluating $V$ and $\nabla V$ at a single point $E_V$. One also needs to find $\operatorname{argmax}\{\tilde{p}_t + V\}$, and then apply the (pre-computed) control at that point. Although a purely brute-force search to compute the argmax could require roughly $50^2$ operations and $50^2$ comparisons to find the argmax, it seems that a more reasonable estimate would be 1000 operations. In any case, these operation count estimates are far below those estimated below for the propagation of the information state, and so we do not include them.

**Example 5.1** Suppose $P_\kappa \widetilde{B}_1$ is independent of $x_\mu$. Then, if $G_x = 50, n = 12$, $q = 12$, and $\dim X_\mu = 2$, we see from formula (5.10) that the total cost of an update of the state estimate is

$$\text{cost: } \approx 50^2(E_Q + E_L + E_V) + 4 \times 10^6 \quad \text{FLOPS}$$

The cost of evaluation of a linear vector field is 50 FLOPS per grid point along $X_\mu$ (see formula 5.8), so we make the (likely over - ) estimate of $3 \times 10^3$ FLOPS to compute $E_Q + E_L + E_V$ per grid point.

$$\text{cost: } \approx 10^7 \text{ FLOPS} = 10 \text{ mega–FLOPS}$$

per second. A modern desktop computer does about 1G FLOPS per second, thus if this were the controller we could update the controller on the scale of at least a hundred times per second. □

**Example 5.2** As another example, consider the case where $n = 6$, $q = 6$, and $\dim X_\mu = 2$, but where $P_\kappa \widetilde{B}_1$ is no longer independent of $x_\mu$. Then one has

$$\text{cost: } \approx 50^2(E_Q + E_L + E_V) + 5 \times 10^6 \quad \text{FLOPS}$$

and again making a (likely over-) estimate of $2 \times 10^3$ FLOPS to compute $E_Q + E_L + E_V$

$$\text{cost: } \approx 10^7 \text{ FLOPS.}$$

The difficulty is the same order of magnitude of as in the previous example. □

We emphasized these two $\dim X_\mu = 2$ cases, because they sit on the borderline of conventional 2000 technology. Obviously, when $\dim X_\mu = 1$ and $G_x$ is still 50, one gets a 50 fold speed up, which means a controller update could be done on the scale of milliseconds.

**Example 5.3** As an example of the degree to which the computations become infeasible without the cheap sensor assumption, consider a case where $n = \dim X_\mu = 6$. With $G_x = 50$, this yields more than $10^{10}$ grid points. Even if a problem had such extremely simple structure that one only performed, say 100 operations per grid point, one would still need to perform more than $10^{12}$ FLOPS per time step. □

### 5.2.4 Future Speed Ups

Notice that our estimates of speed do not take into account the fact that parallelizing methods like those in Section A to some extent is straightforward. Parallelizing would make a big speed improvement. Also, Moore's law implies that by that time the speed up factor will be $2^{10/1.5}$ which is about 100. Thus for $N = 2$ we would have a 10,000 Hertz control law or what is likely more practical, a very cheap 100 Hertz control law or a controller for $\dim X_\mu = 3$.

# 6 Properties of the Controller and Other Theory

We always operate with the following regularity assumptions (not the weakest possible):

**Assumption 6.1** *All of the functions appearing in (3.1) are smooth with bounded first and second order partial derivatives, $B_1$, $\widetilde{B}_1$, $B_2$, $D_{12}$ and $D_{21}$ are bounded, and that zero is an equilibrium: $A(0) = 0$, $C_1(0) = 0$ and $C_{2\mu}(0) = 0$.*

## 6.1 The Integral Equations for the Information State

We now derive the integral form (3.11) given in Section 3.2.3 for the cut down information state from general principles.

Following [15, Chapter 3], for fixed $u, y \in L_{2,loc}$, the information state $p_t$ is defined by

$$p_t(x) = \sup_{\tilde{v}(\cdot),\tilde{w}(\cdot)\in\mathcal{W}(u,y,x,t)} \{\phi_0(\xi(0)) + \tfrac{1}{2}\int_0^t [|\tilde{z}(s)|^2 - \gamma^2(|\tilde{w}(s)|^2 + |\tilde{v}(s)|^2)]\, ds\,\} \tag{6.1}$$

where $\xi(s)$, $0 \le s \le t$, satisfies

$$\dot{\xi} = A(\xi) + \begin{bmatrix} \tilde{B}_1(\xi) & B_1(\xi) \end{bmatrix} \begin{bmatrix} \tilde{v} \\ \tilde{w} \end{bmatrix} + B_2(\xi)u, \tag{6.2}$$

$$\tilde{z}(s) = C_1(\xi(s)) + D_{12}(\xi(s)),$$

and

$$\mathcal{W}(u(\cdot), y(\cdot), x, t) = \{(\tilde{v}, \tilde{w}) \in L_{2,loc} : \ \exists \text{ a solution } \xi(\cdot) \text{ of (6.2) consistent with } u(\cdot), y(\cdot) \text{ on } [0,t] \text{ with } y_\kappa(\cdot) = \xi_\kappa(\cdot) \text{ and } \xi(t) = x\}. \tag{6.3}$$

By convention the supremum over the empty set is $-\infty$ and we should note here that consequently

$$p_t(x_\kappa, x_\mu) = -\infty \quad \text{if } x_\kappa \neq y_\kappa(t).$$

This simplifies once we solve for $\tilde{w}$ via

$$\tilde{w}(s) = y_\mu(s) - C_{2\mu}(\xi(s)) \qquad \forall\, s \in [0,t] \tag{6.4}$$

to give

$$p_t(x) = \sup_{\tilde{v}(\cdot)\in\mathcal{V}(u,y,x,t)} \{\phi_0(\xi(0)) + \tfrac{1}{2}\int_0^t [|\tilde{z}(s)|^2 - \gamma^2(|y_\mu(s) - C_{2\mu}(\xi(s))|^2 + |\tilde{v}(s)|^2)]\, ds\,\} \tag{6.5}$$

where $\mathcal{V}(u(\cdot), y(\cdot), x, t)$ is given by (3.12).

For each time $t$ the information state $p_t$ is *singular* in that it is supported on an affine subspace $x_\kappa = y_\kappa(t)$. For any set $\mathcal{M} \subseteq X$, we define the delta function, $\delta_\mathcal{M}$, by

$$\delta_\mathcal{M}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{M} \\ -\infty & \text{otherwise.} \end{cases}$$

In order to reduce notation, we will write $\delta_{x_\kappa=\alpha}$ in place of $\delta_{\{(x_\kappa,x_\mu):x_\kappa=\alpha\}}$. Then $p_t$ has the form

$$p_t(x_\kappa, x_\mu) = \delta_{x_\kappa=y_\kappa(t)}(x_\kappa, x_\mu) + p_t^{y_\kappa}(x_\mu)$$

where

$$p_t^{y_\kappa}(x_\mu) = p_t(y_\kappa(t), x_\mu)$$

or equivalently (6.5). Thus $p_t^{y_\kappa}(\cdot)$ is a function of $x_\mu \in X_\mu$ and is propagated by

$$p_t^{y_\kappa}(x_\mu) = \sup_{\tilde v(\cdot)\in\mathcal{V}(u,y,(y_\kappa(t),x_\mu),t)} \{\phi_0(y_\kappa(0),\xi_\mu(0)) + \tfrac{1}{2}\int_0^t \{|C_1(y_\kappa(s),\xi_\mu(s)) + D_{12}u(s)|^2$$
$$-\gamma^2(|\tilde v(s)|^2 + |C_{2\mu}(y_\kappa(s),\xi_\mu(s)) - y_\mu(s)|^2)\}\, ds\} \tag{6.6}$$

and (3.13). We assume that $\mathcal{V}(u,y,(y_\kappa(t),x_\mu),t) \neq \emptyset$; otherwise our model would be inconsistent with the data. (However, small observation noise limits are discussed in Section 8.)

The function $\phi_0$ used in the definitions above represents the initial information concerning the state $x_0$. Suppose that $\phi_0$ is a finite function. Then the value of the RHS of (6.5) at time $t = 0$ is

$$\begin{aligned} p_0(x_\kappa, x_\mu) &= \sup_{v(\cdot)\in\mathcal{V}(u,y,(y_\kappa(t),x_\mu),0)} \{\phi_0(\xi_\kappa(0),\xi_\mu(0)) + 0\} \\ &= \delta_{x_\kappa = y_\kappa(0)}(x_\kappa, x_\mu) + \phi_0(y_\kappa(0), x_\mu). \end{aligned}$$

In this case,

$$p_0^{y_\kappa}(x_\mu) = \phi_0(y_\kappa(0), x_\mu).$$

We make the following assumption:

**Assumption 6.2** *We assume that $\phi_0(x_\kappa, x_\mu)$ is of the form*

$$\phi_0(x_\kappa, x_\mu) = \phi_{0\kappa}(x_\kappa) + \phi_{0\mu}(x_\mu), \tag{6.7}$$

*where each term is of class $C^1$ and of at most quadratic growth.*

**Remark 6.3** The form (6.7) means that the cut down information state can be initialized without knowledge of $y_\kappa(0)$. However, in general one could set $\phi_{0\mu}(x_\mu) = \phi_0(y_\kappa(0), x_\mu)$, $\phi_{0\kappa} = 0$.

Now define the *cut down information state* $\tilde p_t(x_\mu)$ by (3.11). Then it is clear that

$$p_t^{y_\kappa} = \tilde p_t + \phi_{0\kappa}(y_\kappa(0)). \tag{6.8}$$

The cut down information state characterizes the dissipation property (2.2):

**Lemma 6.4** *Let $\tilde p_0$ be given by (3.14), and define $\tilde p_t$ by (3.11). Then if $K$ is an admissible controller such that for the closed loop system $(G,K)$ the cut down information state satisfies*

$$\tilde p_t(x_\mu) \le -\phi_{0\kappa}(y_\kappa(0)) \quad \text{for all } t \ge 0, x_\mu \in X_\mu, \tag{6.9}$$

*where $u = K(y)$ and $y$ are the closed loop control and measurement signals, then the closed loop system $(G,K)$ is dissipative; i.e. (2.2) holds.*

PROOF. Consider the closed loop system $(G, K)$ with initial state $x_0$ and inputs $w_0(\cdot), v_0(\cdot)$. The controller $K$ determines the closed loop signals $u(\cdot) = K(y(\cdot))$ and $y(\cdot)$, together with state trajectory denoted $\xi_0(\cdot)$. In view of (6.9) and (3.11) (setting $\tilde{v}(\cdot) = v_0(\cdot)$, $\xi(\cdot) = \xi_0(\cdot)$, $\tilde{z}(\cdot) = z_0(\cdot)$) we see that

$$\phi_{0\mu}(x_{\mu 0}(0)) + \int_0^t \{|z_0(s)|^2 - \gamma^2(|v_0(s)|^2 + |w_0(s)|^2)\}\, ds \leq -\phi_{0\kappa}(y_\kappa(0)) \tag{6.10}$$

for all $t \geq 0$, since $w_0(s) = y_\mu(s) - C_{2\mu}(\xi_0(s))$, $0 \leq s \leq t$. This implies,

$$\int_0^t |z_0(s)|^2 ds \leq \gamma^2 \int_0^t (|v_0(s)|^2 + |w_0(s)|^2)\, ds - \phi_{0\kappa}(y_\kappa(0)) - \phi_{0\mu}(\xi_{\mu 0}(0)) \tag{6.11}$$

This is (2.2) with $\beta(y_\kappa(0), x_\mu) = -\phi_0(y_\kappa(0), \xi_{\mu 0}(0))$, by Assumption 6.2. □

## 6.2 Effectiveness of the Controller

It was explained in section 2 that the information state theory of nonlinear $H_\infty$ control is global, and is both necessary and sufficient. It is beyond the scope of this paper to go into all these details here, and we refer the reader to the monograph [15]. In this section we give conditions under which the RECIPE of Section 3.2.1 works. These conditions are stated for the certainty equivalence case; the general case can also be considered and will be dealt with elsewhere.

We treat the case $\#Sens \leq \#Dis$. Consider the RECIPE (Section 3.2.1) under the certainty equivalence assumptions (Section 2.3), which produces the certainty equivalence controller $K^*$ (see (2.13)). To be clear, the certainty equivalence assumptions in the present context are as follows:

**Assumption 6.5** *(i) The cut down information state $\tilde{p}_t(x_\mu)$ defined in section 3.2.3 is smooth and solves the cut down PDE (3.8).*

*(ii) The state feedback value function $V(x_\kappa, x_\mu)$ (section 2.2) is smooth and solves the state feedback PDE (2.9) (with $B_1$ replaced by $[\tilde{B}_1, B_1]$).*

*(iii) The minimum stress estimate $\bar{x}_\mu(t)$ exists and is uniquely defined by*

$$\bar{\xi}_\mu(t) = \arg \max_{x_\mu \in X_\mu} \{\tilde{p}_t(x_\mu) + V(y_\kappa(t), x_\mu)\}. \tag{6.12}$$

*(iv) The control is given by*

$$u(t) = u^*_{state}(y_\kappa(t), \bar{x}_\mu(t)) \tag{6.13}$$

*(c.f (3.6)).*

The certainty equivalence controller $K^*$ is given by (3.8), (6.12), and (6.13).

We must show that under this assumption the closed loop $(G, K^*)$ is dissipative. By Lemma 6.4, this will follow if we can show that the cut down information state is bounded above as in (6.9). In order to show this, we define a function

$$\bar{W}(y_\kappa, \tilde{p}) = \max_{x_\mu \in X_\mu} \{\tilde{p}(x_\mu) + V(y_\kappa, x_\mu)\}.$$

Since $V \geq 0$ it is clear that $\bar{W}$ satisfies the important inequality:

$$\bar{W}(x_\kappa, \tilde{p}) \geq \sup_{x_\mu \in X_\mu} \{\tilde{p}(x)\} \qquad \forall\, x_\kappa \in X_\kappa. \tag{6.14}$$

**Theorem 6.6** *Assume $\#Sens \leq \#Dis$ and Assumption 6.5. Assume that the initial information state is chosen so that*

$$\bar{W}(y_\kappa(0), \phi_{0\mu}) \leq -\phi_{0\kappa}(y_\kappa(0)). \tag{6.15}$$

*Consider the controller $K^*$ produced by the RECIPE and the closed loop $(G, K^*)$. Then:*

*(i) The function (of $t$) $\bar{W}(y_\kappa(t), \tilde{p}_t)$ is decreasing, and in particular,*

$$\bar{W}(y_\kappa(t), \tilde{p}_t) \leq \bar{W}(y_\kappa(0), \tilde{p}_0). \tag{6.16}$$

*(ii) The closed loop $(G, K^*)$ is dissipative.*

Proof. Consider the closed loop system $(G, K^*)$ with initial plant state $x_0$ and inputs $w_0(\cdot), v_0(\cdot)$. The controller $K^*$ determines the closed loop signals $u(\cdot) = K^*(y(\cdot))$ and $y(\cdot)$, together with state trajectory denoted $\xi_0(\cdot)$ and cut down information state $\tilde{p}$. Then for $t > 0$ we have

$$\bar{W}(y_\kappa(t), \tilde{p}_t) = \tilde{p}_t(\bar{x}_\mu(t)) + V(y_\kappa(t), \bar{x}_\mu(t)).$$

Differentiating with respect to $t$ we have

$$\begin{aligned} \tfrac{d}{dt}\bar{W}(y_\kappa(t), \tilde{p}_t) &= \nabla_\kappa V(y_\kappa(t), \bar{x}_\mu(t))\dot{y}_\kappa(t) + \nabla_\mu V(y_\kappa(t), \bar{x}_\mu(t))\dot{\bar{x}}_\mu(t) \\ &\quad + \tfrac{\partial}{\partial t}\tilde{p}_t(\bar{x}_\mu(t)) + \nabla_\mu \tilde{p}_t(\bar{x}_\mu(t))\dot{\bar{x}}_\mu(t) \\ &= \nabla_\kappa V(y_\kappa(t), \bar{x}_\mu(t))\dot{y}_\kappa(t) + \tfrac{\partial}{\partial t}\tilde{p}_t(\bar{x}_\mu(t)) \end{aligned} \tag{6.17}$$

since

$$\nabla_\mu V(y_\kappa(t), \bar{x}_\mu(t)) + \nabla_\mu \tilde{p}_t(\bar{x}_\mu(t)) = 0. \tag{6.18}$$

(The differentiation is justified by Danskin's Theorem, [5, Appendix B].)

The cut down information state PDE (3.8) gives

$$\begin{aligned} \tfrac{\partial}{\partial t}\tilde{p}_t(\bar{x}_\mu(t)) = -\nabla_\mu \tilde{p}_t P_\mu[A + B_1(y_\mu(t) - C_{2\mu}) + \widetilde{B}_1 v^* + B_2 u(t)] \\ + \tfrac{1}{2}|C_1 + D_{12}u(t)|^2 - \tfrac{1}{2}\gamma^2[|v^*|^2 + |C_{2\mu} - y_\mu(t)|^2] \end{aligned}$$

where $v^* = v^*(x_\mu, t)$ is given by (3.10) and the RHS is evaluated at $(y_\kappa(t), \bar{x}_\mu(t))$ and $t$ appropriately. The constraint (3.13) yields

$$\begin{aligned}\dot{y}_\kappa(t) = P_\kappa[&A(y_\kappa(t), \bar{x}_\mu(t)) + B_1(y_\kappa(t), \bar{x}_\mu(t))(y_\mu(t) - C_{2\mu}(y_\kappa(t), \bar{x}_\mu(t)) \\ &+\widetilde{B}_1(y_\kappa(t), \bar{x}_\mu(t))\tilde{v} + B_2(y_\kappa(t), \bar{x}_\mu(t))u(t))]\end{aligned}$$

where $\tilde{v}(t) = v^*(\bar{x}_\mu(t), t) \in \mathcal{V}_t(\bar{x}_\mu(t))$. We also have

$$y_\mu(t) = C_{2\mu}(y_\kappa(t), \bar{x}_\mu(t)) + w_0(t).$$

These facts together with (6.17), (6.18) give

$$\begin{aligned}\tfrac{d}{dt}\bar{W}(y_\kappa(t), \tilde{p}_t) &= \nabla_x V(y_\kappa(t), \bar{x}_\mu(t))[A + \tilde{B}_1 v^* + B_1 w_0(t) + B_2 u(t)] \\ &\quad +\tfrac{1}{2}|C_1 + D_{12}u(t)|^2 - \tfrac{1}{2}\gamma^2[|v^*|^2 + |C_{2\mu} - y_\mu(t)|^2 \\ &\leq 0\end{aligned} \tag{6.19}$$

with the RHS is evaluated at $(y_\kappa(t), \dot{\bar{x}}_\mu(t))$ and $t$ appropriately. The inequality in (6.19) follows from the PDE (2.9) for $V$ (with $B_1$ replaced by $[\tilde{B}_1, B_1]$), and the certainty equivalence control law (6.13) (Assumption 6.5). This proves the first part of the theorem upon integration.

By (6.14), (6.15) and (6.16) we have

$$\tilde{p}_t(x_\mu) \leq \bar{W}(y_\kappa(t), \tilde{p}_t) \leq \bar{W}(y_\kappa(0), \tilde{p}_0) \leq -\phi_{0\kappa}(y_\kappa(0)), \tag{6.20}$$

and so applying Lemma 6.4 we see that the closed loop system is dissipative. $\square$

## 6.3 Viscosity Solutions

As promised in Section 3.2.3, we see here when $\#Sens \leq \#Dis$ that the cut down information state is a viscosity solution of the cut down PDE. This theorem is slightly unusual in that it asserts that the cut down information state, as a function of the $x_\mu$ components of the state, is a solution of (3.8) on this moving affine subspace of the state space. We do not fully analyse the $\#Sens \leq \#Dis$ case, but we require the following assumption:

**Assumption 6.7** *We assume that $P_\kappa\widetilde{B}_1(x_\kappa, x_\mu)$ has range $X_\kappa$ for all $(x_\kappa, x_\mu)$, so that for any $x_\mu, t$, there exists some $v \in \mathcal{V}$ such that (3.13) holds with $\xi_\kappa(t) \equiv y_\kappa(t)$. Specifically, the set*

$$\mathcal{V}_t(x_\mu) \doteq \left\{v \in (P_\kappa\widetilde{B}_1)^{-1}[\dot{y}_\kappa - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2 u(t))]\right\}$$

*where $(P_\kappa\widetilde{B}_1)^{-1}$ is the set–valued inverse ($(P_\kappa\widetilde{B}_1)^{-1}b = \{a : P_\kappa\widetilde{B}_1 a = b\}$) is nonempty. We also assume uniform non–degeneracy, by which we mean that we assume that there exists $C_{PB} < \infty$ such that*

$$\min\{|v| : v \in (P_\kappa\widetilde{B}_1)^{-1}\xi\} = |(P_\kappa\widetilde{B}_1)^{\#}\xi| \leq C_{PB}|\xi| \qquad \forall\, x \in \mathbf{R}^n,\ \xi \in \mathbf{R}^p \tag{6.21}$$

*where superscript # indicates the Moore–Penrose generalized inverse. Since the Moore–Penrose generalized inverse is a bounded linear operator (see Appendix D, or more generally [16]), it is automatically Lipschitz. However, noting that $\widetilde{B}_1$ (and consequently $(P_\kappa \widetilde{B}_1)^\#$) may depend on $x_\mu$, we explicitly assume that*

$$(P_\kappa \widetilde{B}_1)^\# [\dot{y}_\kappa(t) - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2 u(t))] \tag{6.22}$$

*be uniformly Lipschitz in $x_\mu$.*

It might be helpful to note that, loosely speaking, one would typically expect the range condition of the assumption to hold when $\#Sens \leq \#Dis$ for "most" $P_\kappa \widetilde{B}_1$, although this can obviously fail. Also, of course $\#Sens \leq \#Dis$ is a necessary condition for the assumption to hold. We make these remarks only to provide the casual reader a feel for the situation being considered.

**Theorem 6.8** *Fix $\tilde{p}_0$, and let $u, y \in L_{2,loc}$ be given. Assume:*

*(i) The cut down information state $\tilde{p}_t$ is defined by the integral definition (3.11).*

*(ii) For some $\gamma_0$, $\tilde{p}_t$ satisfies the bound (6.9).*

*(iii) Assumption 6.2 on $\phi_0$ holds.*

*(iv) The uniform non-degeneracy and Lipschitz conditions Assumption 6.7 hold.*

*Then, for any $\gamma > \gamma_0$, $\tilde{p}_t(x)$ is finite and a continuous viscosity solution of (3.8).*

The proof is long, and is provided in Appendix B.

# 7 More Measurements than Disturbances: $\#Sens > \#Dis$

In this section we treat the situation where dim $X_\kappa > q$, as opposed to the dim $X_\kappa \leq q$ case we have been studying. As we shall see in this ideal situation the support (region of finiteness) of the information state collapses to a few points. Later in Section 8 we show that small amounts of noise broaden this support in predictable ways.

## 7.1 The Basic Idea

Let's first illustrate the basic idea by the case where $v$ does not occur ($q = 0$). This gives a very informative picture, because when $v$ does occur we solve for it or parts of it; thereby producing equations like the ones where $v$ does not occur.

Look on (3.13) as the dynamics of a time varying system (determined by the fixed signal $u$, $y$, with no controlling inputs). Here we ignore the partition $P_\kappa$, $P_\mu$. We can look on $\xi_\kappa(\cdot) = y_\kappa(\cdot)$ as an observation of this system. If the system is observable (which is true for almost all systems), then by definition this means that knowing $y_\kappa(\cdot)$ up to a time $T > 0$, determines the trajectory $\xi(t)$ for $t > T$. In this case $p_t$ may start at $t = 0$ as any function on the set $\{\xi : \ \xi_\kappa(0) = y_\kappa(0)\}$, but then $p_t$ will be supported on a single vector $\xi(t)$ for $t > T$ (the constraint set $\mathcal{V}(u, y, x, t)$ will correspond to a unique state trajectory by observability).

When $v$ does occur ($q > 0$), then there are various cases, but what we see as the generic situation:

> *When* $\dim\ X_\kappa > q$*, the information state* $p_t$ *behaves as it does above, starting as "anything" at* $t = 0$ *and then for* $t > T$ *developing support on a curve* $\{x : \ x = \xi(t), \ t > 0\}$ *or on a set of isolated curves.*

Naturally this is a noiseless idealization and is just a convenience. The introduction of observation noise in $y_\kappa$ will produce $p_t$ with support not equal to $\{x : \ x = \xi(t), \ t > 0\}$, but rather a $p_t$ which is approaching $-\infty$ off of that set as the added noise goes to zero. Some of the effects of such noise will be treated in section 8.

## 7.2 Details of the Situation

To justify this description, we start with the assumption:

**Assumption 7.1**

$$\mathcal{R} := \ \text{range}\ P_\kappa \widetilde{B}_1$$

*is a fixed subspace of* $X_\kappa$ *and does not depend on* $t$ *and* $x$*, and also that* $\text{rank}\ \widetilde{B}_1 = q$.

Let $P_1$ denote a projection whose null space equals the null space of $P_\kappa \widetilde{B}_1$. Let $\rho_1$ denote any matrix such that $\rho_1 P_\kappa \widetilde{B}_1 = P_1$, and such that the range of $P_1$ equals the range of $\rho_1$.

Let $P_\mathcal{R}$ be the orthogonal projection of $X_\kappa$ onto $\mathcal{R}$. Then multiply equations (3.13) by $\rho_1 P_\mathcal{R}$ and by $P_\kappa - P_\mathcal{R}$ to obtain

$$\begin{aligned} \rho_1 P_\mathcal{R} \dot{y}_\kappa &= \rho_1 P_\mathcal{R} \dot{\xi}_\kappa = \rho_1 P_\mathcal{R} P_\kappa \Big( A(y_\kappa, \xi_\mu) + B_1(y_\kappa, \xi_\mu)(y_\mu - C_{2\mu}(y_\kappa, \xi_\mu)) \\ &\qquad + B_2(y_\kappa, \xi_\mu) u \Big) + P_1 v \end{aligned} \tag{7.1}$$

$$
\begin{aligned}
(I_\kappa - P_{\mathcal{R}})\dot{y}_\kappa &= (I_\kappa - P_{\mathcal{R}})\dot{\xi}_\kappa = (I_\kappa - P_{\mathcal{R}})\big(A(y_\kappa,\xi_\mu) + B_1(y_\kappa,\xi_\mu)(y_\mu - C_{2\mu}(y_\kappa,\xi_\mu)) \\
&\qquad +B_2(y_\kappa,\xi_\mu)u\big)
\end{aligned}
\tag{7.2}
$$

and

$$
\begin{aligned}
\dot{\xi}_\mu = {} & P_\mu\Big(A(y_\kappa,\xi_\mu) + B_1(y_\kappa,\xi_\mu)(y_\mu - C_{2\mu}(y_\kappa,\xi_\mu)) \\
& +B_2(y_\kappa,\xi_\mu)u + \widetilde{B}_1(y_\kappa,\xi_\mu)P_1 v + \widetilde{B}_1(y_\kappa,\xi_\mu)(1-P_1)v\Big)
\end{aligned}
\tag{7.3}
$$

where $I_\kappa$ is the identity on $X_\kappa$. Solve the first algebraic equation (7.1) for $P_1 v$ and plug into the next to last term of the differential equation (7.3).

The resulting version of (7.3) becomes the state dynamics for a smaller system on the $X_\mu$ subspace, with (7.2) being the observation. To make this more explicit, note first that the input $u$ and the observation $y_\kappa$ are known functions of time. Consequently, $(I_\kappa - P_{\mathcal{R}})\dot{y_\kappa}$ is also a known function of time. Let

$$
\begin{aligned}
\widehat{y} &\doteq (I_\kappa - P_{\mathcal{R}})\dot{y_\kappa}, \\
h(t,\xi_\mu) &\doteq A(y_\kappa,\xi_\mu) + B_1(y_\kappa,\xi_\mu)(y_\mu - C_{2\mu}(y_\kappa,\xi_\mu)) + B_2(y_\kappa,\xi_\mu)u \\
\mathcal{A}(t,\xi_\mu) &\doteq P_\mu\left\{h(t,\xi_\mu) + \widetilde{B}_1(y_\kappa,\xi_\mu)\left[\rho_1 P_{\mathcal{R}}(\dot{y}_\kappa - h(t,\xi_\mu))\right]\right\},
\end{aligned}
$$

and

$$
\tilde{\mathcal{B}}(t,\xi_\mu) \doteq P_\mu \widetilde{B}_1(y_\kappa,\xi_\mu).
$$

Then, the reduced system has its dynamics and observations given by

$$
\begin{aligned}
\dot{\xi_\mu} &= \mathcal{A}(t,\xi_\mu) + \tilde{\mathcal{B}}(t,\xi_\mu)(1-P_1)v \\
\widehat{y}(t) &= \mathcal{C}(t,\xi_\mu) := (I_\kappa - P_{\mathcal{R}})h(t,\xi_\mu).
\end{aligned}
$$

Alternatively, one might choose to define $\widehat{v} \doteq (1-P_1)v$ to obtain the form

$$
\begin{aligned}
\dot{\xi_\mu} &= \mathcal{A}(t,\xi_\mu) + \mathcal{B}(t,\xi_\mu)\widehat{v} \\
\widehat{y}(t) &= \mathcal{C}(t,\xi_\mu).
\end{aligned}
\tag{7.4}
$$

We call a system of the form (7.4) *observable despite noise* provided that the trajectory $\xi$ of the system is determined by any brief interval $[0,\nu]$ of the observations, for any unknown input function $v$. For example, when $\mathcal{B}$ is 0 this is the usual notion of observability for the system $(\mathcal{C},\mathcal{A})$.

**Example 7.2** (dim $X_\kappa = 1 =$ dim $X_\mu$ and $\tilde{B}_1 = 0$, that is, there is no $v$ ). The equation (7.2) has the form $r(t,\xi_\mu) = 0$ where $r$ is a real valued function. It dominates in that typically this equation has solutions which are isolated points (by Sard's theorem). Thus it reduces our uncertainty about $\xi_\mu$ from a continuum of possibilities to a single point or a set of single points. The other equations in (7.2) (7.3) must be consistent with this, since the equations come from running a system.

We indulge for the moment in a trivial example intended as a tutorial. Suppose we have a time varying polynomial.

$$r(t,x) := 1 + a_1(t)x + a_2(t)x^2$$

The equation $0 = r(t,x)$ has 0, 1 or 2 real solutions $\xi(t)$. In each case the equation forces $\xi(t)$ to lie in a 0 dimensional or empty manifold. Globally, there are 0 such manifolds, 1 such, or 2 such manifolds. If there are 2, then as we move $t$ they sweep out 2 curves which might touch and then disappear. Also there might be no solutions and then two solutions simultaneously appear. The upshot of all this is that the support of the information state $p_t$ is very heavily determined if we have a polynomial constraint attached to its underlying dynamics. If no solutions at all exist to the polynomial we are in an unphysical situation or our model is incorrect. □

## 7.3 The Linear Case

In the case of linear systems, equation (7.4) takes the form

$$\begin{aligned} \dot{\xi}_\mu &= \mathcal{A}\xi_\mu + \mathcal{B}v + K(t) \\ \bar{y} &= \mathcal{C}\xi_\mu \end{aligned} \tag{7.5}$$

for a suitable function of time $K(t)$ and observation $\bar{y}(t)$. We assume $\dim \bar{y} \geq 2$ and break up $\bar{y}$ into two linearly independent components:

$$\begin{aligned} \bar{y}_1 &= \mathcal{C}_1\xi_\mu \\ \bar{y}_2 &= \mathcal{C}_2\xi_\mu \end{aligned} \tag{7.6}$$

Now

$$\xi_\mu(t) = e^{\mathcal{A}t}\xi_\mu(0) + \int_0^t e^{\mathcal{A}(t-s)}[\mathcal{B}v(s) + K(s)]ds$$

and so

$$\begin{aligned} \bar{y}_1(t) &= \mathcal{C}_1 e^{\mathcal{A}t}\xi_\mu(0) + \textstyle\int_0^t \mathcal{C}_1 e^{\mathcal{A}(t-s)}[\mathcal{B}v(s) + K(s)]ds \\ \bar{y}_2(t) &= \mathcal{C}_2 e^{\mathcal{A}t}\xi_\mu(0) + \textstyle\int_0^t \mathcal{C}_2 e^{\mathcal{A}(t-s)}[\mathcal{B}v(s) + K(s)]ds \end{aligned} \tag{7.7}$$

We shall use the first equation in (7.7) to solve for the noise term $v$. We re-write it as

$$\tilde{y}_1(t) = \bar{y}_1(t) - \tilde{K}_1(t) = \mathcal{C}_1 e^{\mathcal{A}t}\xi_\mu(0) + \int_0^t \mathcal{C}_1 e^{\mathcal{A}(t-s)}\mathcal{B}v(s)ds \tag{7.8}$$

where $\tilde{K}_1(t) = \int_0^t \mathcal{C}_1 e^{\mathcal{A}(t-s)}K(s)ds$. The equation

$$\Gamma_1(v)(t) = \int_0^t \mathcal{C}_1 e^{\mathcal{A}(t-s)}\mathcal{B}v(s)ds$$

defines a linear operator $\Gamma_1 : L_{2,loc} + \mathbf{R}\,\delta_0 \to L_{2,loc}$, so that

$$\tilde{y}_1(t) = \mathcal{C}_1 e^{\mathcal{A}t}\xi_\mu(0) + \Gamma_1(v)(t). \tag{7.9}$$

Here $\delta_0$ is the delta function at 0 and $\mathbf{R}$ denotes the real numbers. We assume $\Gamma_1$ is invertible, with inverse $\Gamma_1^{-1} : L_{2,loc} \to L_{2,loc} + \delta_0\mathbf{R}$. This certainly implies that

$$\operatorname{rank}\mathcal{B} \leq \operatorname{rank}\mathcal{C}_1. \tag{7.10}$$

Indeed for generic choice of $\mathcal{A}, \mathcal{B}, \mathcal{C}_1$ satisfying (7.10), the operator $\Gamma_1$ is invertible. Consider equation (7.9) written as

$$\tilde{y}_1 - \mathcal{C}_1 e^{\mathcal{A}t}\xi_\mu(0) = \Gamma_1(v) \tag{7.11}$$

Solving for $v$ gives

$$\begin{aligned} v(t) &= \Gamma_1^{-1}[\tilde{y}_1(t) - \mathcal{C}_1 e^{\mathcal{A}t}\xi_\mu(0)] \\ &= \Lambda_{11}(\tilde{y}_1)(t) + \Lambda_{12}(t)\xi_\mu(0) + \phi\,\delta_0 \end{aligned} \tag{7.12}$$

for suitable $\Lambda_{11} : L_{2,loc} \to L_{2,loc}$, $\Lambda_{12} : X_\mu \to L_{2,loc}$, $\phi \in \mathbf{R}$.

Next we use the second component in equation (7.7) to solve for the unknown initial condition $\xi_\mu(0)$:

$$\check{y}(t) := \bar{y}_2(t) - \tilde{K}_2(t) = \mathcal{C}_2 e^{\mathcal{A}t}\xi_\mu(0) + \int_0^t \mathcal{C}_2 e^{\mathcal{A}(t-s)}\mathcal{B}v(s)ds \tag{7.13}$$

where $\tilde{K}_2(t) = \int_0^t \mathcal{C}_2 e^{\mathcal{A}(t-s)}K(s)ds$. Substituting (7.12) into (7.13) gives

$$\begin{aligned} \check{y}(t) \;=\; & \int_0^t \mathcal{C}_2 e^{\mathcal{A}(t-s)}\mathcal{B}\Lambda_{11}(\tilde{y}_1)(s)ds \\ & + \mathcal{C}_2 e^{\mathcal{A}t}[I + \int_0^t \mathcal{C}_2 e^{-\mathcal{A}s}\mathcal{B}\Lambda_{12}(s)ds]\xi_\mu(0) + \phi\,\mathcal{C}_2 e^{\mathcal{A}t}\mathcal{B} \end{aligned} \tag{7.14}$$

In principle, the initial condition $\xi_\mu(0)$ and $\phi \in \mathbf{R}$ can be determined from a system of equations obtained by differentiating equation (7.14) $\dim X_\mu$ times with respect to $t$ and setting $t = 0$; this gives

$$\begin{bmatrix} \check{y}(0) \\ \check{y}'(0) \\ \check{y}^{(2)}(0) \\ \cdots \\ \check{y}^{(\mu)}(0) \end{bmatrix} = \begin{bmatrix} \mathcal{C}_2 \\ \mathcal{C}_2\mathcal{A} + \mathcal{C}_2\mathcal{B}\Lambda_{12}(0) \\ \mathcal{C}_2\mathcal{A}^2 + \mathcal{C}_2\mathcal{A}\mathcal{B}\Lambda_{12}(0) + \mathcal{C}_2\mathcal{B}\Lambda'_{12}(0) \\ \cdots \\ \mathcal{C}_2\mathcal{A}^\mu + \cdots \end{bmatrix} \xi_\mu(0) + \begin{bmatrix} \mathcal{C}\mathcal{B} \\ \mathcal{C}\mathcal{A}\mathcal{B} \\ \vdots \\ \mathcal{C}\mathcal{A} \end{bmatrix} \phi \tag{7.15}$$

*This requires that the matrix in (7.15) has full rank.* Note that ordinary observability requires that the rank of the observability matrix

$$\begin{bmatrix} \mathcal{C}_2 \\ \mathcal{C}_2\mathcal{A} \\ \mathcal{C}_2\mathcal{A}^2 \\ \\ \cdots \\ \mathcal{C}_2\mathcal{A}^\mu \end{bmatrix} \tag{7.16}$$

be full. Conditions such as (7.15) and (7.16) are true for generic systems. Provided the dimension of the $y$ measurement space is greater than the dimension of the space of disturbances $v$.

With regard to our original problem the key condition for a generic system to be observable despite noise is that $\mathcal{C}_2$ not be 0. From this (and inequality (7.10)) we conclude that a generic linear system is observable despite noise provided the number of accurate sensors exceeds the number of channels of $v$ noise.

## 7.4 Sensor Counts

As illustrated in the linear case the expected behavior even of a system (7.4) is that it is observable despite noise provided $\dim \operatorname{range} \mathcal{C} > \dim \operatorname{domain} \mathcal{B} = \dim \operatorname{range}(1 - P_1)$. The next lemma provides the equivalent and more appealing condition that observable despite noise is the expected behavior when the total number of sensors strictly exceeds the number of disturbances.

**Lemma 7.3** *Generically dim range $\mathcal{C} >$ dim domain $\mathcal{B}$ is equivalent to* $\dim X_\kappa > q$.

PROOF. We use a few facts, generically:

- $\operatorname{rank} P_1 = \operatorname{rank} P_{\mathcal{R}}$
- $\dim \operatorname{domain} \mathcal{B} = \dim \operatorname{range}(1 - P_1) = q - \operatorname{rank} P_1 = q - \operatorname{rank} P_{\mathcal{R}}$
- $\dim \operatorname{range} \mathcal{C} \leq \operatorname{rank} P_\kappa - \operatorname{rank} P_{\mathcal{R}}$.
- $q = \operatorname{rank} \tilde{B}_1$

These combined with the assumption of the lemma give

$$\begin{aligned}
\dim X_\kappa - \operatorname{rank} P_{\mathcal{R}} &= \operatorname{rank} P_\kappa - \operatorname{rank} P_{\mathcal{R}} \\
&\geq \dim \operatorname{range} \mathcal{C} \\
&> \dim \operatorname{domain} \mathcal{B} \\
&= \operatorname{rank}(1 - P_1) \\
&\geq q - \operatorname{rank} P_{\mathcal{R}}
\end{aligned}$$

in which cancellation yields the assertion of the lemma. □

# 8 The Effect of Noise - Perturbation Estimates

So far we have considered the case where the measurements of the $x_\kappa$ state components are exact (i.e. $y_\kappa(t) = \xi_\kappa(t)$ for all $t$). Let us call this the *perfect measurement* case. However, in practice there is always some measurement error, and so this is just a convenient approximation. Indeed, the often-studied state feedback problem is such an idealization. Thus we analyze small perturbations of the exact measurement situation. In order to validate the model considered in the previous sections, one would like to show that the information state for the system with perturbed measurements converges to the information state of the previous sections. In this section we prove a key step towards this goal.

We consider the system (3.1) with an additional "small" input $\sqrt{\varepsilon}\overline{w}$ whose effect on the output is

$$y^\varepsilon_\kappa = \xi_\kappa + \sqrt{\varepsilon}\overline{w} \tag{8.1}$$

and call it the *imperfect measurement* system. This imperfect measurement system leads to an information state $p^\varepsilon_t$ which we shall study as $\varepsilon$ goes to 0. To deal with it efficiently we set the following notation. For fixed $u, y \in L_{2,loc}$, define the $\varepsilon, \gamma$-cost function for the imperfect measurement system by

$$J^\varepsilon(t, x_\kappa, x_\mu, v; u, y) =$$

$$\phi_0(\xi(0)) + \tfrac{1}{2}\int_0^t [|C_1(\xi(s)) + D_{12}(\xi(s))u(s)|^2 - \gamma^2(|w(s)|^2 + |v(s)|^2 + \frac{1}{\varepsilon}|\overline{w}(s)|^2)]\, ds, \tag{8.2}$$

where $\xi_\kappa(\cdot), \xi_\mu(\cdot)$ satisfy (3.1) with $\xi_\kappa(t) = x_\kappa, \xi_\mu(t) = x_\mu$. As before, $w, \overline{w}$ do not appear on the left hand side since

$$\begin{aligned} w &= y_\mu - C_{2\mu}(\xi) \\ \overline{w} &= \tfrac{1}{\sqrt{\varepsilon}}[y^\varepsilon_\kappa - \xi_\kappa]. \end{aligned} \tag{8.3}$$

The information state for the imperfect measurement system is then given by

$$p^\varepsilon_t(x_\kappa, x_\mu) = \sup_{v \in L_{2,loc}} J^\varepsilon(t, x_\kappa, x_\mu, v; u, y). \tag{8.4}$$

In order to analyze the behavior of $p^\varepsilon_t(x_\kappa, x_\mu)$ as $\varepsilon \to 0$, we will need some setup. It is natural to expect that if a controller $K_0$ solves the $H^\infty$ problem for the imperfect measurement system, then $K_0$ solves the $H^\infty$ problem for the case of $\kappa$ perfectly measured states; however, the converse is not generally true.

For definiteness, fix a controller $K_0$, an initial state $x_0 \in \mathbf{R}^n$, and $\tilde{w}, \tilde{\overline{w}}, \tilde{v} \in L_{2,loc}$. The imperfect measurement system will have a state trajectory $\tilde{\xi}^\varepsilon(\cdot)$ determined by (3.1) and $\tilde{u}^\varepsilon = K_0(\tilde{y}^\varepsilon)$, where of course the measurement equation of (3.1) is replaced by

$$\tilde{y}^\varepsilon_\kappa(t) = \tilde{\xi}^\varepsilon_\kappa(t) + \sqrt{\varepsilon}\tilde{\overline{w}}(t), \quad \tilde{y}^\varepsilon_\mu(t) = C_{2\mu}(\tilde{\xi}^\varepsilon_\kappa(t), \tilde{\xi}^\varepsilon_\mu(t)) + \tilde{w}(t). \tag{8.5}$$

For the perfect measurement system, let the state trajectory, $\tilde{\xi}$ be determined by (3.1) with the same $\tilde{w}, \tilde{\overline{w}}, \tilde{v}$ (although now $\tilde{\overline{w}}$ is irrelevant), the same $x_0 \in \mathbf{R}^n$, and the same $K_0$ yielding control $\tilde{u} = K_0(\tilde{y})$ where

$$\tilde{y}_\kappa(t) = \tilde{\xi}_\kappa(t), \quad \tilde{y}_\mu(t) = C_{2\mu}(\tilde{\xi}_\kappa(t), \tilde{\xi}_\mu(t)) + \tilde{w}(t).$$

The measurements $\tilde{y}^\varepsilon$ from the imperfect measurement system will be input to the information state $p_t^\varepsilon$, while the perfect measurements $\tilde{y}$ will be input into the perfect information state $p_t$. The theorems to follow in this section will be concerned with the rates at which such a sequence (indexed by $\varepsilon$ for any fixed $\tilde{\overline{w}}$) of $p_t^\varepsilon$ converge to $-\infty$ on the set of $x$ for which $p_t(x) = -\infty$. We will make the following assumptions concerning the convergence of the imperfect measurement system state to the perfect measurement system state when they have the same inputs as discussed just above.

**Assumption 8.1** *Let $\tilde{w}, \tilde{\overline{w}}, \tilde{v} \in L_{2,loc}$. Fix a controller $K_0$, $x_0 \in \mathbf{R}^n$. We assume there exists $C_1 < \infty$ such that*

$$\begin{aligned} &\limsup_{\varepsilon \to 0} \varepsilon^{-1/2} \parallel \tilde{\xi}^\varepsilon - \tilde{\xi} \parallel_{L_\infty[0,t]} \leq C_1 \\ &\limsup_{\varepsilon \to 0} \varepsilon^{-1/2} \parallel \tilde{u}^\varepsilon - \tilde{u} \parallel_{L_2[0,t]} \leq C_1. \end{aligned} \tag{8.6}$$

**Remark 8.2** We include a trivial example indicating the reasonableness of Assumption 8.1. Consider the simple, one–dimensional system where

$$\begin{aligned} \dot{\tilde{\xi}} &= \tilde{\xi} + \tilde{u} + \tilde{v} \\ \tilde{y} &= \tilde{\xi} \\ \tilde{u}(t) &= K_0[\tilde{y}](t) = -2\tilde{y}(t). \end{aligned}$$

Then

$$\dot{\tilde{\xi}} = -\tilde{\xi} + \tilde{v}.$$

Also,

$$\begin{aligned} \dot{\tilde{\xi}}^\varepsilon &= \tilde{\xi}^\varepsilon + \tilde{u}^\varepsilon + \tilde{v} \\ \tilde{y}^\varepsilon &= \tilde{\xi}^\varepsilon + \sqrt{\varepsilon}\tilde{\overline{w}} \\ \tilde{u}^\varepsilon &= K_0[\tilde{y}^\varepsilon](t) = -2\tilde{y}^\varepsilon(t) \end{aligned}$$

so that

$$\dot{\tilde{\xi}}^\varepsilon = -\tilde{\xi}^\varepsilon + \tilde{v} - 2\sqrt{\varepsilon}\tilde{\overline{w}}.$$

Then

$$\dot{\tilde{\xi}}^\varepsilon - \dot{\tilde{\xi}} = -(\tilde{\xi}^\varepsilon - \tilde{\xi}) - 2\sqrt{\varepsilon}\tilde{\overline{w}},$$

and from this it is easy to show that

$$\begin{aligned} \|\tilde{\xi}^\varepsilon - \tilde{\xi}\|_{L_\infty} &\leq \sqrt{\varepsilon}\|\tilde{\overline{w}}\|_{L_2} \\ \|\tilde{u}^\varepsilon - \tilde{u}\|_{L_2} &\leq \sqrt{(4t+1)\varepsilon}\|\tilde{\overline{w}}\|_{L_2} \end{aligned}$$

which completes the remark. □

**Assumption 8.3** *It is assumed that the controller $K_0$ makes the imperfect measurement system dissipative, which in terms of the information state means: there exists $\varepsilon_0 > 0$ and $\gamma_0 > 0$ such that for all $0 < \varepsilon \leq \varepsilon_0$*

$$\begin{aligned}\phi_0(\xi^\varepsilon(0)) + \tfrac{1}{2}\int_0^t [|C_1(\xi^\varepsilon(s)) + D_{12}(\xi^\varepsilon(s))\tilde{u}^\varepsilon(s)|^2 \\ -\gamma_0^2(|\tilde{y}^\varepsilon_\mu(s) - C_{2\mu}(\xi^\varepsilon(s))|^2 + |v(s)|^2 + \tfrac{1}{\varepsilon}|\tilde{y}^\varepsilon_\kappa(s) - \xi^\varepsilon_\kappa(s)|^2)]\, ds \leq 0\end{aligned} \tag{8.7}$$

*for all $v \in L_{2,loc}$ and all final states $\xi^\varepsilon(t) = x$, where $\xi^\varepsilon(\cdot)$ is the solution of the state equation in (3.13) driven by $\tilde{y}^\varepsilon, \tilde{u}^\varepsilon, v$.*

Note that this implies that the controller $K_0$ makes the perfect system dissipative: there exists $\gamma_0 > 0$ such that

$$\begin{aligned}\phi_0(\xi(0)) + \tfrac{1}{2}\int_0^t [|C_1(\xi(s)) + D_{12}(\xi(s))\tilde{u}(s)|^2 \\ -\gamma_0^2(|\tilde{y}_\mu(s) - C_{2\mu}(\xi(s))|^2 + |v(s)|^2)]\, ds \leq 0\end{aligned} \tag{8.8}$$

for all $v \in \mathcal{V}$ and all final states $\xi(t) = x$, where $\xi(\cdot)$ is the solution of the state equation in (3.13) driven by $\tilde{y}, \tilde{u}, v$.

Lastly, we will also use the following.

**Assumption 8.4** *We assume that $C_{2\mu}$ is bounded.*

## 8.1 General Case

In this section, we wish to show that $p^\varepsilon$ decays to $-\infty$ off the time–varying affine space $\{x = (x_\kappa, x_\mu) : x_\kappa = y_\kappa(t)\}$.

**Theorem 8.5** *Fix $t > 0$. Let $\tilde{y}$ be generated by the perfect measurement system with some inputs $\tilde{w}, \tilde{v}$, initial condition $\tilde{\xi}(0) = x_0$, and controller $K_0$. For each $\varepsilon > 0$, let $\tilde{y}^\varepsilon$ be generated by the imperfect measurement system with the same inputs, initial condition and controller as the perfect measurement systems, but now also with input $\tilde{\overline{w}}$. We suppose that Assumptions 8.1, 8.3 and 8.4 hold as well as 6.1. Let $\gamma > \gamma_0$ (where $\gamma_0$ is given in Assumption 8.3). Then there exist constants $\alpha_1 > 0$, $\varepsilon_1 > 0$ and $k_1 > 0$ such that for $0 < \varepsilon \leq \varepsilon_1$*

$$p^\varepsilon_t(x) \leq -\frac{k_1}{\varepsilon^{1/2}}|x_\kappa - \tilde{\xi}_\kappa(t)|^2 + \alpha_1 + (\gamma^2 - \gamma_0^2)\|\tilde{\overline{w}}\|^2_{L_2} \tag{8.9}$$

*where $p^\varepsilon_t$ is generated using observations $\tilde{y}^\varepsilon$.*

The proof is given in Appendix C.

## 8.2 Noise Insensitive Observability Case

In this subsection, we consider the situation similar to that of Section 7. As we saw in there, in the case where $\dim X_\kappa > q$, the information state $p_t(x)$ for the perfect measurement system system is supported on a single point $x = \tilde{\xi}(t)$ under a natural and generic observability condition. In this section we show that for the imperfect measurement system, the information state $p_t^\varepsilon(x)$ converges to $-\infty$ for $x \neq \tilde{\xi}(t)$. Here, $\tilde{\xi}(\cdot)$ is the unique trajectory satisfying the constraints in (3.13).

We will need an assumption related to the *observable despite noise* assumption of Section 7, called *noise insensitive observability*, which will specify the behavior as noise is added to the previously perfect measurements. The assumption is phrased in a somewhat more general context.

**Assumption 8.6** *(noise insensitive observability) Fix $t > 0$. Let $\xi^1$ and $\xi^2$ be generated by the first equation of (3.1) with inputs $y_\mu^1, v^1, u^1$ and $y_\mu^2, v^2, u^2$, respectively, and terminal conditions $\xi^1(t) = x^1$ and $\xi^2(t) = x^2$. (To be specific, the $w$ input for each is given by $w^1(r) = y_\mu^1(r) - C_{2\mu}(\xi^1(r))$ and $w^2(r) = y_\mu^2(r) - C_{2\mu}(\xi^2(r))$ for all $r$, respectively.) Suppose there exist $\beta > 0$, $\zeta > 0$, $\nu_0 > 0$, $\delta_0 > 0$ and $g : [0,\infty) \to [0,\infty)$ such that $g(\eta) \geq \beta|\eta|^\zeta$ for all $\eta \in [0,\infty)$, such that the following holds. If there exist $\eta > 0$ such that*

$$|\xi_\mu^1(r) - \xi_\mu^2(r)| \geq \eta \qquad \forall\, r \in [t-\nu_0, t], \tag{8.10}$$

*then for all $\delta \in (0, \delta_0]$ and all $\nu \in (0, \nu_0]$*

$$\begin{aligned} &\|y_\mu^1 - y_\mu^2\|_{L_2(t-\nu,t)} + \|u^1 - u^2\|_{L_2(t-\nu,t)} + \delta\nu\|v^1 - v^2\|_{L_2(t-\nu,t)} \\ &+ \frac{1}{\nu\delta}\|\xi_\kappa^1 - \xi_\kappa^2\|_{L_2(t-\nu,t)} \geq g(\eta)\sqrt{\nu}. \end{aligned} \tag{8.11}$$

**Remark 8.7** We include a trivial example indicating that Assumption 8.6 defines a reasonable class of systems. In this example, we see that Noise Insensitive Observability can occur even when one only has $\#Sens = \#Dis$ as opposed to the strict $\#Sens > \#Dis$ case where one would typically expect it to occur. In particular, consider the simple system given by

$$\begin{aligned} \dot{\xi}_\kappa &= \xi_\mu \\ \dot{\xi}_\mu &= v + u. \end{aligned}$$

In this case, the assumption follows from a standard Sobolev interpolation inequality (see for instance [1]). In particular, let $z \in W^{2,2}[0,1]$. Then, given any $\delta_1 > 0$, there exists $k < \infty$ (dependent on $\delta_1$) such that for any $\delta \in (0, \delta_1)$,

$$\|\dot{z}\|_{L_2} \leq k\delta\|\ddot{z}\|_{L_2} + \frac{k}{\delta}\|z\|_{L_2}.$$

In order to apply this inequality in the example at hand, simply map $[t-\nu,t]$ onto $[0,1]$ to obtain

$$\|\xi_\mu\|_{L_2(t-\nu,t)} \leq k\nu\delta\|u+v\|_{L_2(t-\nu,t)} + \frac{k}{\nu\delta}\|\xi_\kappa\|_{L_2(t-\nu,t)}.$$

With a little work (and taking $\delta\nu$ sufficiently small), one sees that this implies that Assumption 8.6 holds for this example.

**Remark 8.8** *Noise insensitive observability* implies the property *observable despite noise* of Section 7, since essentially the former is a quantification of the latter expressed in terms of signal norms. Now we give some details of this correspondence. To compare *noise insensitive observablity* with the *observable despite noise* assumption of Section 7, consider two output measurements which are the same (i.e. $y^1_\kappa(s) = y^2_\kappa(s)$ ) for all $s \in [t-\nu_0, t]$. This implies $\xi^1_\kappa(s) = \xi^2_\kappa(s)$ for all $s \in [t-\nu_0,t]$. If $u^1(s) = u^2(s)$ and $y^1(s) = y^2(s)$ for almost all $s \in [t-\nu_0,t]$, then Assumption 8.6 (by taking $\delta \to 0$) implies $g(\eta) = 0$. Thus $\eta = 0$ which by (8.10) implies there exists an $s_0$ in $[t-\nu_0,t]$ for which $\xi^1_\mu(s_0) = \xi^2_\mu(s_0)$, This happens for all $\nu_0$ and $t$, so $\xi^1_\mu(s) = \xi^2_\mu(s)$ for all $s$ (using continuity of $\xi_m u$ - which we have). Thus, *noise insensitive observablity* is a stronger condition than *observable despite noise.*

**Theorem 8.9** *Fix $t > 0$. Let $\tilde{y}$ be generated by the perfect measurement system with some inputs $\widetilde{w}, \widetilde{v}$, initial condition $\tilde{\xi}(0) = x_0$, and controller $K_0$. For each $\varepsilon > 0$, let $\tilde{y}^\varepsilon$ be generated by the imperfect measurement system with the same inputs, initial condition and controller as the perfect measurement system, but now also with additional input $\widetilde{\widetilde{w}}$ (see (8.5)). We suppose that Assumptions 8.1, 8.3, 8.4 and 8.6 hold as well as 6.1. Let $\gamma > \gamma_0$ (where $\gamma_0$ is given in Assumption 8.3). Then there exist constants $\alpha_2 > 0$, $\varepsilon_2 > 0$ and $k_2 > 0$ such that for $0 < \varepsilon \leq \varepsilon_2$*

$$\begin{aligned} p^\varepsilon_t(x) \leq -k_2 \max\left\{ \frac{1}{\varepsilon^{1/2}}|x_\kappa - \tilde{\xi}_\kappa(t)|^2,\ \frac{1}{\varepsilon^{1/5}}|x_\mu - \widetilde{\xi}_\mu(t)|^{\min\{2,4\zeta\}} \right\} \\ +\alpha_2 + (\gamma^2-\gamma_0^2)\|\widetilde{\widetilde{w}}\|^2_{L_2} \qquad (8.12) \end{aligned}$$

*where $p^\varepsilon_t$ is generated using observations $\tilde{y}^\varepsilon$.*

The proof is given in Appendix C.

**Remark 8.10** A close examination of the proof of Theorem 8.9 will show that one can get arbitrarily close to a growth rate of $1/\varepsilon^{1/4}$ in the last term inside the max in (8.12), but that this is at the expense of $\varepsilon_2$. Also, the authors do not claim that these growth rate estimates are sharp.

**Remark 8.11** One can replace the $g(\eta) \geq \beta|\eta|^\zeta$ bound in Assumption 8.6 with a different functional form yielding a diffrenent form for the last term in (8.12).

# A Computational Cost of the Finite Difference Numerical Scheme

## A.1 Algebraic Reductions

In order to simplify the calculations, we define the following notation. Let

$$\widetilde{b}_{\mu,i} \doteq [P_\mu[A + B_1(y - C_2) + B_2 u]]_i$$

where we are suppressing arguments. Let $\widetilde{B}_\mu \doteq P_\mu \widetilde{B}_1$ and $d_{\mu,i}$ be the transpose of the $i^{th}$ row of $\widetilde{B}_\mu$. Also, let $\widetilde{L} = L + \gamma^2|v|^2$, and note that $\widetilde{L}$ is independent of $v$. Finally, let $D_\mu$ be the dimension of $X_\mu$. Then (5.6) can be written as

$$\begin{aligned}\breve{p}_t(x_\mu) = \breve{p}_{t-\Delta}(x_\mu) + \Delta\widetilde{L} + \sup_{v\in\mathcal{V}_t(x_\mu)} \Bigg\{ \sum_{i=1}^{D_\mu} \frac{1}{\rho}\Big[ &(\widetilde{b}_{\mu,i} + d_{\mu,i}^T v)^- (\breve{p}_{t-\Delta}(x_\mu + \delta e_i) - \breve{p}_{t-\Delta}(x_\mu)) \\ &+ (\widetilde{b}_{\mu,i} + d_{\mu,i}^T v)^+ (\breve{p}_{t-\Delta}(x_\mu - \delta e_i) - \breve{p}_{t-\Delta}(x_\mu))\Big] \\ &- \Delta\gamma^2|v|^2 \Bigg\}\end{aligned}$$

Letting $\alpha_{i,\pm} \doteq \breve{p}_{t-\Delta}(x_\mu \pm \delta e_i) - \breve{p}_{t-\Delta}(x_\mu)$, this becomes

$$\breve{p}_t(x_\mu) = \breve{p}_{t-\Delta}(x_\mu) + \Delta\widetilde{L} + \sup_{v\in\mathcal{V}_t(x_\mu)} \left\{ \sum_{i=1}^{D_\mu} \frac{1}{\rho}\Big[(\widetilde{b}_{\mu,i} + d_{\mu,i}^T v)^- \alpha_{i,+} + (\widetilde{b}_{\mu,i} + d_{\mu,i}^T v)^+ \alpha_{i,-}\Big] - \Delta\gamma^2|v|^2 \right\}. \tag{A.1}$$

## A.2 Explication of $\mathcal{V}_t$

Recall that we are in the case $\#Sens \le \#Dis$ which implies $\dim(X_\kappa) \le q$ where we recall that $q$ is the dimension of the $v$-space. The case $\dim(X_\kappa) = q$ is rather trivial, so we consider only $\dim(X_\kappa) < q$. One may write $\mathcal{V}_t(x_\mu)$ as

$$\mathcal{V}_t(x_\mu) = v_0(x_\mu) + \widetilde{\mathcal{V}}_t(x_\mu)$$

where $\widetilde{\mathcal{V}}_t(x_\mu)$ is a subspace of dimension $D_V \doteq q - \dim(X_\kappa)$ (containing the origin), and

$$v_0 \perp \widetilde{\mathcal{V}}_t(x_\mu). \tag{A.2}$$

We drop the argument, shortening this to $\mathcal{V}_t = v_0 + \widetilde{\mathcal{V}}_t$. Let the projection of each $d_{\mu,i}$ onto $\widetilde{\mathcal{V}}$ be denoted by $\widetilde{d}_{\mu,i}$. Also let $\widehat{b}_{\mu,i} \doteq \widetilde{b}_{\mu,i} + d_{\mu,i}^T v_0$ for all $i$. Then, by (A.1) and (A.2), one has

$$\begin{aligned}\breve{p}_t(x_\mu) &= \breve{p}_{t-\Delta}(x_\mu) + \Delta\widetilde{L} - \Delta\gamma^2|v_0|^2 \\ &\quad + \sup_{\widetilde{v}\in\widetilde{\mathcal{V}}_t(x_\mu)} \left\{ \sum_{i=1}^{D_\mu} \frac{1}{\rho}\Big[(\widehat{b}_{\mu,i} + \widetilde{d}_{\mu,i}^T \widetilde{v})^- \alpha_{i,+} + (\widehat{b}_{\mu,i} + \widetilde{d}_{\mu,i}^T \widetilde{v})^+ \alpha_{i,-}\Big] - \Delta\gamma^2|\widetilde{v}|^2 \right\}.\end{aligned} \tag{A.3}$$

## A.3 Computing the Supremum

Note that the supremum in (A.3) is a maximum of a sum of piecewise-linear functions, each with a single discontinuity in the gradient along the hyperplane $\widehat{b}_{\mu,i} + \widetilde{d}_{\mu,i}^T \widetilde{v} = 0$, and the convex quadratic $-\Delta\gamma^2|\widetilde{v}|^2$ over the space $\widetilde{\mathcal{V}}_t$. Suppose that the maximum occurs in the interior of one of the smooth regions. (Checking the hyperplanes and lower dimensional subspaces is of similar difficulty.) Then the maximizing $\widetilde{v}$ lies in the span of the $\widetilde{d}_{\mu,i}$. In other words, the maximizer takes the form

$$\widetilde{v}_m = \sum_{i=1}^{D_\mu} \beta_i \widetilde{d}_{\mu,i}$$

for some choice of $\beta_j \in \mathbf{R}$. Let $\beta = (\beta_1, \beta_2, \ldots \beta_{D_\mu})$. (Note that the number of elements in the sum on the right–hand side may sometimes be smaller than this. For instance, if $D_V < D_\mu$, then letting $e_j$ span $\widetilde{\mathcal{V}}$, the sum could be replaced by $\sum_{j=1}^{D_V} \beta_j e_j$.) With this, (A.3) becomes

$$\begin{aligned} \breve{p}_t(x_\mu) &= \breve{p}_{t-\Delta}(x_\mu) + \Delta\widetilde{L} - \Delta\gamma^2|v_0|^2 \\ &\quad + \sup_{\beta\in\mathbf{R}^{D_\mu}} \left\{ \sum_{i=1}^{D_\mu} \frac{1}{\rho}\left[ \left(\widehat{b}_{\mu,i} + \sum_{j=1}^{D_\mu} \lambda_{i,j}\beta_j\right)^- \alpha_{i,+} + \left(\widehat{b}_{\mu,i} + \sum_{j=1}^{D_\mu} \lambda_{i,j}\beta_j\right)^+ \alpha_{i,-}\right] \right. \\ &\qquad\qquad \left. -\Delta\gamma^2 \sum_{i,j} \beta_i\beta_j\lambda_{i,j} \right\} \end{aligned} \tag{A.4}$$

where $\lambda_{i,j} \doteq \widetilde{d}_{\mu,i} \cdot \widetilde{d}_{\mu,j}$. Letting $\widetilde{\lambda}_i \doteq (\lambda_{i,1}, \ldots \lambda_{i,D_\mu})^T$ and $\Lambda$ be the positive semidefinite matrix of the $\lambda_{i,j}$, this becomes

$$\begin{aligned} \breve{p}_t(x_\mu) &= \breve{p}_{t-\Delta}(x_\mu) + \Delta\widetilde{L} - \Delta\gamma^2|v_0|^2 \\ &\quad + \sup_{\beta\in\mathbf{R}^{D_\mu}} \left\{ \sum_{i=1}^{D_\mu} \frac{1}{\rho}\left[ \left(\widehat{b}_{\mu,i} + \widetilde{\lambda}_i^T\beta\right)^- \alpha_{i,+} + \left(\widehat{b}_{\mu,i} + \widetilde{\lambda}_i^T\beta\right)^+ \alpha_{i,-}\right] - \Delta\gamma^2\beta^T\Lambda\beta \right\}. \end{aligned} \tag{A.5}$$

One must check each possible smooth region separately. It is not difficult to show that there are at most $2^{D_\mu}$ such regions. For instance, consider the region where all the $\widehat{b}_{\mu,i} + \widetilde{\lambda}_i^T\beta$ are positive. Then one has

$$\sup_{\beta\in\mathbf{R}^{D_\mu}} \left\{ \sum_{i=1}^{D_\mu} \frac{1}{\rho}\left[ \left(\widehat{b}_{\mu,i} + \widetilde{\lambda}_i^T\beta\right)\alpha_{i,-}\right] - \Delta\gamma^2\beta^T\Lambda\beta \right\},$$

and the maximizer, $\beta_m$, is given by

$$\frac{1}{\rho}\left[\sum_{i=1}^{D_\mu} \tilde{\lambda}_i\right] - 2\Delta\gamma^2\Lambda\beta_m = 0.$$

If $\beta_m$ satisfies $\widehat{b}_{\mu,i} + \widetilde{\lambda}_i^T\beta_m > 0$ for all $i$, then it is a candidate for the global maximizer. One computes all possible local maxima on interiors of the smooth regions similarly. One must then check those hyperplanes corresponding to *concave* discontinuities in the gradient and so on with the lower dimensional subspaces.

## A.4 Computational Complexity

Recalling that the number of grid points per dimension is $G_x$, one sees that in the situation at hand, the total number of grid points is $G_x^{D_\mu}$ as opposed to $G_x^n$. For instance, in a case with $n = 6$, $D_\mu = 2$ (i.e. 4 of 6 states measured "perfectly") and $G_x = 50$, the number of grid points is reduced by a factor of $50^4$. If the operations necessary to compute $\breve{p}_t(x_\mu)$ are $c_1(n, q, D_\mu)$ for each grid point, then the total operation count per time-step would be roughly $50^2 c_1(n, q, D_\mu)$. This compares quite favorably with the operation estimate for $D_\mu = n$ of $50^6 c_1(n, q, n) \simeq 1.5 \times 10^{10} c_1(n, q, n)$ (an improvement by a factor of more than $10^6$).

Now, in order to justify the feasibility of even this reduced level of computation we will perform a rough estimate of the operations required per grid point, $c_1(n, q, D_\mu)$. Note that the first two items in the list below are function evaluations, and so the exact number of operations will depend on the particular example. The remainder are straight operation counts for matrix/vector operations.

*Operation counts:*

- Function evaluations:
  - $Q \doteq A + B_1(y - C_2) + B_2 u \rightarrow$ assume $E_Q$ operations
  - $\Delta\widetilde{L} \rightarrow$ assume $E_L$ operations
- Simple operation counts:
  - all $\widetilde{b}_{\mu,i} \rightarrow 2nD_\mu$ operations (not including the operations for $Q$ above)
  - all $d_{\mu,i} \rightarrow nD_\mu$ operations
  - all $\alpha_{i,\pm} \rightarrow 4D_\mu$ operations
  - $\dot{y}_\kappa - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2 u(t)) = \dot{y}_\kappa - P_\kappa Q \rightarrow D_\kappa(1 + n)$ operations
  - $v_0$ and all $\widetilde{d}_{\mu,i} \rightarrow D_\kappa^2 q + D_\kappa^3 + 2D_\kappa q + 7qD_\mu D_V$ operations (where $D_V$ is given in Subsection A.2)
  - all $\widehat{b}_{\mu,i} \rightarrow D_\mu(1 + q)$ operations
  - $\Lambda \rightarrow 2D_V D_\mu^2$ operations
  - Each $\beta_m \rightarrow D_\mu^3$ operations
  - $\breve{p}_t(x_\mu)$ given a $\beta_m \rightarrow 5 + 4D_\mu + 4D_\mu^2$ operations

Recall that each $\beta_m$ is for only one of the smooth regions. Consequently the operations for $\beta_m$ and $\breve{p}_t^{y_\kappa}(x_\mu)$ need to be performed up to once per smooth region ($2^{D_\mu}$ times in the worst case). Recall also that one needs to check all hyperplanes corresponding to concave discontinuities in the gradient. In the case where $D_\mu = 1$, it is easy to see that this is essentially trivial. In the case of $D_\mu = 2$, the two hyperplanes are lines, and so one needs to check the four rays emanating from the intersection of the two lines, *assuming* both of the slope discontinuities are concave. (In the case where a slope discontinuity is convex, one cannot have a maximum there of course, so this is a worst case analysis.) A conservative figure is to assume this doubles the cost of the four smooth regions.

Summing up the above operation counts, and taking the last paragraph into consideration, one finds that the total operation count per time-step from

$$
\begin{aligned}
c_1 = {} & E_Q + E_L \\
& + \Big\{ 3nD_\mu + 4D_\mu + (n+1)D_\kappa + qD_\kappa^2 + D_\kappa^3 + 2qD_\kappa + 7qD_\mu D_V + (q+1)D_\mu + 2D_\mu^2 D_V \\
& + [D_\mu^3 + 4D_\mu^2 + 4D_\mu + 5]2^{D_\mu} \Big\} M_{D_\mu}
\end{aligned}
$$

where $M_{D_\mu}$ is the multiplier which approximates the additional cost due to checking the lower dimensional hyperplanes. As noted above, we take $M_1 = 1$ and $M_2 = 2$; we do not consider higher values of $D_\mu$ here.

To simplify this, use a conservative bound of replacing $D_\kappa$ by $n$ above (keeping in mind that we typically expect $D_\kappa$ nearly equal to $n$ (or equivalently, $D_\mu$ small). Further, let us specialize to the case $q = n$ which implies $D_V = D_\mu$. (Note that the case $q > n$ can be reduced to the case $q = n$ by a change of variables for the disturbances, albeit with a change of cost criterion from $-\gamma^2|v|^2$ to a different concave quadratic, however, we do not consider those details here.) With these reductions, the above total count becomes

$$
\begin{aligned}
c_1 = {} & E_Q + E_L \\
& + \Big\{ 2n^3 + 3n^2 + n(7D_\mu^2 + 4D_\mu + 1) + 2D_\mu^3 + D_\mu \\
& + [D_\mu^3 + 4D_\mu^2 + 4D_\mu + 5]2^{D_\mu} \Big\} M_{D_\mu}.
\end{aligned}
$$

Continuing with the example of $n = 6$, $D_\mu = 2$, $D_\kappa = n - 2 = 4$, we also suppose say, $q = 6$, so that $D_V = q - D_\kappa = q - (n - D_\mu) = 2$. In this case, one would have at worst roughly $\tilde{5}0^2 c_1(6,6,2) \simeq 50^2 \cdot (E_Q + E_\Delta + 2000) = 2500(E_Q + E_\Delta) + 5 \times 10^6$ FLOPS per time-step.

One may also consider the case where $P_\kappa \widetilde{B}_1$ is independent of $x_\mu$. In this case, the operation count is greatly reduced. Specifically, one need only do most of the computations for $v_0$ and all $\widetilde{d}_{\mu,i}$ once per time-step rather than once per grid point as well as time-step. Since this step is one of the most time-consuming, this is a great reduction. The operation count for those computations which still need to be done at each grid point a for $v_0$ and all $\widetilde{d}_{\mu,i}$ is only $2qD_\mu D_V + 2qD_\kappa$. Again taking $D_\kappa$ to be roughly $n$ and $q = n$, one obtains

$$
c_1 = E_Q + E_L
$$

$$
\begin{aligned}
&+\Big\{3n^2 + n(2D_\mu^2 + 4D_\mu) + 2D_\mu^3 + 5D_\mu \\
&+[D_\mu^3 + 4D_\mu^2 + 4D_\mu + 5]2^{D_\mu}\Big\} M_{D_\mu}.
\end{aligned}
$$

In this case, consider the situation where $n = 12$ and $D_\mu = 2$. The computations per time-step are only roughly $50^2\tilde{c}_1 \simeq 50^2(E_Q + E_\Delta + 1600) = 2500(E_Q + E_\Delta) + 4 \times 10^6$ FLOPS. This is the same order of magnitude as the $n = 6$, $D_\mu = 2$ case when $P_\kappa \widetilde{B}_1$ was not independent of $x_\mu$.

# B Proof of Theorem 6.8

The first step is to obtain a bound on the size of optimal disturbances, that is, we obtain a bound on the size of $\varepsilon$–optimal $v$ in (6.6). This will yield the asserted finiteness as well as a uniform continuity with respect to time off the corresponding state trajectories.

For notational simplicity, define

$$
\begin{aligned}
L_\gamma(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; s) \; &= |C_1(y_\kappa(s), \xi_\mu(s)) + D_{12}u(s)|^2 \\
&\quad -\gamma^2(|v(s)|^2 + |C_{2\mu}(y_\kappa(s), \xi_\mu(s)) - y_\mu(s)|^2) \qquad \text{(B.1)}
\end{aligned}
$$

where $\xi_\mu(\cdot)$ is given by (3.13) for given $y, u, v$ and $\xi_\mu(t) = x_\mu$ so that (6.6) becomes

$$
\tilde{p}_t(x_\mu) = \sup_{v(\cdot)\in\mathcal{V}} \{\phi_{0\mu}(\xi(0)_\mu) + \tfrac{1}{2}\int_0^t L_\gamma(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; s)\, ds\} \qquad \text{(B.2)}
$$

For clarity we reiterate that the input functions $u, y, v$ determine $\xi_\kappa(u, y, v)$ via the dynamical equations (3.1) for $G$. Note the $t$ dependence in (B.1) is 0, but we allow $L_\gamma$ to depend on $t$ since our theorems using $L_\gamma$ go through without added difficulty.

Throughout, we assume $\gamma > \gamma_0$ and $u, y$ are as given in the statement of Theorem 6.8.

**Lemma B.1** *Let $\varepsilon \in (0, 1]$. Then $\varepsilon$–optimal $v$ in (6.6) satisfy*

$$
\|v\|^2_{l_2(0,t)} \le M_3^t(\gamma^2 - \gamma_0^2)^{-1}(1 + R^2) \doteq M_{t,R}
$$

*for all $|x_\mu| \le R$ and all $R < \infty$.*

Proof. This proof is only a slight variant of similar proofs which appear in [22], [23], [24]. The first part differs from the references in that the choice of $u, y, \gamma_0$ automatically yields an upper bound by the assumptions.

Let $v \in \mathcal{V}(K_0[y], y, (y_\kappa(t), x_\mu), t)$ be $\varepsilon$–optimal . Let

$$\begin{aligned} J_t^{y_\kappa}(x_\mu, v(\cdot)) &\doteq \phi_{0\mu}(\xi_\mu(0)) + \tfrac{1}{2}\int_0^t L_\gamma(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; s)\, ds \\ &= \phi_{0\mu}(\xi_\mu(0)) + \tfrac{1}{2}\int_0^t L_{\gamma_0}(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; s)\, ds - (\gamma^2 - \gamma_0^2)\int_0^t |v(s)|^2\, ds \end{aligned}$$

which by the assumptions of Theorem 6.8 yields the upper bound

$$\leq -(\gamma^2 - \gamma_0^2)\int_0^t |v(s)|^2\, ds. \tag{B.3}$$

In the references mentioned at the top of the proof, the second part of the proof is to compare this bound on the cost with the cost for $v \equiv 0$, and show that the above cost is less optimal than that for $v \equiv 0$ for large enough $\|v\|^2$. This would yield the desired result. The variation here is that $v \equiv 0$ might not be in $\mathcal{V}(u, y, (y_\kappa(t), x_\mu), t)$. In order to remedy this, take

$$\widehat{v}(t, x_\mu) \doteq \mathrm{argmin}\Big\{|v| : v \in (P_\kappa \widetilde{B}_1)^{-1}[\dot{y}_\kappa - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2 u(t))]\Big\}$$

Then $\widehat{v}$ is measurable in $t$, and by (6.22) there exists $K_t < \infty$ such that

$$|\widehat{v}(s, x_\mu) - \widehat{v}(s, \overline{x}_\mu)| \leq K_t |x_\mu - \overline{x}_\mu| \qquad \forall\, s \in (0, t),\ x_\mu, \overline{x}_\mu \in X_\mu.$$

Consider the second equation of (3.13) with disturbance $\widehat{v}$

$$\dot{\xi}_\mu = P_\mu\left(A(y_\kappa, \xi_\mu) + B_1(y_\kappa, \xi_\mu)(y_\mu - C_{2\mu}(y_\kappa, \xi_\mu)) + \widetilde{B}_1(y_\kappa, \xi_\mu)\widehat{v} + B_2(y_\kappa, \xi_\mu)u\right).$$

Note that the right hand side of this ODE is globally Lipschitz in $\xi_\mu$ with a constant $C_t < \infty$ for any $t \in (0, \infty)$. By standard results, there exists a solution $\xi_\mu(\cdot)$ and further,

$$|\xi_\mu(s)|^2 \leq \overline{C}_t(1 + |x_\mu|^2) \qquad \forall\, s \in [0, t]$$

for proper choice of $\overline{C}_t$. Consequently, defining $v^m(\cdot) \doteq \widehat{v}(\cdot, \xi_\mu(\cdot))$, we see that $v^m \in L_2$, and $\mathcal{V} \neq \emptyset$. Further, by Assumption 6.7 there exists $M_1^t < \infty$ such that

$$|v^m(s)| \leq M_1^t(1 + |\xi_\mu(s)|) \qquad \forall\, s \in [0, t].$$

Then

$$J_t^{y_\kappa}(x_\mu, v^m) \geq -\gamma^2 \overline{M}_1^t \left[1 + \|\xi_\mu^m(\cdot)\|^2_{L_2[0,t]}\right] \tag{B.4}$$

where $\xi_\mu^m$ is given by (3.13) with $v = v^m$. Also note that there exists $M_2^t$ such that

$$\|\xi_\mu^m(\cdot)\|^2 \leq M_2^t(1 + |x_\mu|^2). \tag{B.5}$$

By (B.4), (B.5)

$$J_t^{y_\kappa}(x_\mu, v^m) \geq -\gamma^2 M_3^t(1 + |x_\mu|^2) \tag{B.6}$$

for proper $M_3^t$. Comparing (B.3) and (B.6) yields the result. □

**Theorem B.2** *$\tilde{p}_s(x_\mu)$ is finite for all $(s, x_\mu) \in (0,t) \times X_\mu$.*

PROOF. The bound from above is inherent in the hypotheses of Theorem 6.8. So one only needs to obtain $p_s^{y_\kappa}(x_\mu) > -\infty$, but this follows directly from the existence of $v^m$ and (B.6) in the proof of the previous lemma. □

We will now show that $\tilde{p}_t(x_\mu)$ is a continuous viscosity solution of (3.8). The results are essentially standard with the slight exceptions due to quadratic growth and the dependence of $\mathcal{V}_t(x_\mu)$ on $t, x_\mu$. Consequently, we will not include redundant details.

The next lemma follows directly from Lemma B.1 by some simple computations using Gronwalls bound.

**Lemma B.3** *Let $x_\mu \in X_\mu$, $R < |x_\mu|$, and let $v$ be $\varepsilon$–optimal . Then there exists $C_1$ depending on $t, R$ such that for any $s \in [0,t]$,*

$$|\xi_\mu(s) - x_\mu| \le C_1[(t-s) + \sqrt{t-s}]. \tag{B.7}$$

The next result is a dynamic programming principle particularly suited to our needs.

**Theorem B.4** *Let $0 \le s \le t$. Let $x_\mu \in X_\mu$. For any $R \ge |x_\mu|$ and any $M \ge M_{t,R}$ (where $M_{t,R}$ is given in Lemma B.1), one has*

$$\tilde{p}_t(x_\mu) = \sup_{||v||^2 \le M, v \in \mathcal{V}} \left\{ \tilde{p}_s(\xi_\mu(s)) + \int_s^t L_\gamma(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; r)\, dr \right\} \tag{B.8}$$

PROOF. Let $R \ge |x_\mu|$, $M \ge M_{t,R}$. Define

$$R_t(x_\mu) = \sup_{||v||^2 \le M, v \in \mathcal{V}} \left\{ \tilde{p}_s(\xi_\mu(s)) + \int_s^t L_\gamma(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; r)\, dr \right\}.$$

Then

$$\begin{aligned} R_t(x_\mu) &\le \sup_{v \in \mathcal{V}} \Big\{ \sup_{v \in \mathcal{V}} [\phi_{0\mu}(\xi_\mu(0)) + \textstyle\int_0^s L_\gamma(y(\cdot), u(\cdot), v(\cdot), \xi_\mu(s), s; r)\, dr] \\ &\qquad\qquad + \textstyle\int_s^t L_\gamma(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; r)\, dr \Big\} \\ &= \tilde{p}_t(x_\mu) \end{aligned} \tag{B.9}$$

On the other hand, for any $\overline{x}_\mu$

$$\tilde{p}_s(\overline{x}_\mu) = \sup_{v \in \mathcal{V}} \left[ \phi_{0\mu}(\xi_\mu(0)) + \int_0^s L_\gamma(y(\cdot), u(\cdot), v(\cdot), \overline{x}_\mu, s; r)\, dr \right].$$

Therefore,

$$
\begin{aligned}
R_t(x_\mu) &= \sup_{\|v_1\|^2 \le M, v_1 \in \mathcal{V}} \Big\{ \sup_{v_2 \in \mathcal{V}} \left[ \phi_{0\mu}(\xi_\mu(0)) + \textstyle\int_0^s L_\gamma(y(\cdot), u(\cdot), v_2(\cdot), \xi_\mu(s), s; r)\, dr \right] \\
&\qquad\qquad + \textstyle\int_s^t L_\gamma(y(\cdot), u(\cdot), v_1(\cdot), x_\mu, t; r)\, dr \Big\} \\
&\ge \sup_{\|v\|^2 \le M, v \in \mathcal{V}} \left\{ \phi_{0\mu}(\xi_\mu(0)) + \textstyle\int_0^t L_\gamma(y(\cdot), u(\cdot), v(\cdot), x_\mu, t; r)\, dr \right\}.
\end{aligned} \tag{B.10}
$$

Consequently, by Lemma B.1, $R_t(x_\mu) \ge \tilde{p}_t(x_\mu)$ which completes the proof. □

**Theorem B.5** *$\tilde{p}.(\cdot)$ is a continuous viscosity solution of (3.8).*

The proof of Theorem 6.8 follows directly from the following two lemmas which are easily proved. See [McR, Lemmas 2.8, 2.9, Theorem 2.10] for proofs of similar (but slightly more complex) results.

**Lemma B.6** *Let $(\bar{t}, \overline{x}_\mu) \in (0,t) \times X_\mu$. Let $\theta > 0$, $g \in C^1(\mathbf{R} \times X_\mu)$ satisfy*

$$
-\theta \ge g_t(\bar{t}, \overline{x}_\mu) + H(\bar{t}, \overline{x}_\mu, \nabla_\mu g)
$$

*where*

$$
\begin{aligned}
H(t,x,q) \doteq -\sup_{v \in \mathcal{V}_t(x_\mu)} \Big\{ & -[A + B_1(y_\mu(t) - C_{2\mu}) + \widetilde{B}_1 v + B_2 u(t)]^T P_\mu^T q \\
& + \tfrac{1}{2}|C_1 + D_{12}u(t)|^2 - \gamma^2[|v|^2 + |C_{2\mu} - y_\mu(t)|^2] \Big\}.
\end{aligned} \tag{B.11}
$$

*Then there exists $M' < \infty$ such that for all $M \ge M'$, there exists $\rho_0 = \rho_0(\bar{t}, \overline{x}_\mu, M) < \bar{t}$ and $v \in \mathcal{V}$ such that $\|v\|^2_{L_2[0,t]} \le M$ and*

$$
\begin{aligned}
\int_{\bar{t}-\rho_0}^{\bar{t}} \Big\{ & L_\gamma(y(\cdot), u(\cdot), v(\cdot), \overline{x}_\mu, \bar{t}; r) \\
& - [A + B_1(y_\mu(r) - C_{2\mu}) + \widetilde{B}_1 v + B_2 u(r)]^T P_\mu^T \nabla_\mu g + g_t \Big\}\, dr \le \frac{-\rho\theta}{2}
\end{aligned}
$$

*for all $\rho \in [0, \rho_0]$.*

**Lemma B.7** *Let $(\bar{t}, \overline{x}_\mu) \in (0,t) \times X_\mu$ and $R \ge |x_\mu|$. Let $\theta > 0$, $g \in C^1(\mathbf{R} \times X_\mu)$ satisfy*

$$
\theta \le g_t(\bar{t}, \overline{x}_\mu) + H(\bar{t}, \overline{x}_\mu, \nabla_\mu g)
$$

*where $H$ is given in (B.11). Then there exists $\rho_0 = \rho_0(\bar{t}, \overline{x}_\mu, M) < \bar{t}$ such that for all $v \in \mathcal{V}$ such that $\|v\|^2_{L_2[0,t]} \le M_{t,R}$, one has*

$$
\begin{aligned}
\int_{\bar{t}-\rho_0}^{\bar{t}} \Big\{ & L_\gamma(y(\cdot), u(\cdot), v(\cdot), \overline{x}_\mu, \bar{t}; r) \\
& - [A + B_1(y_\mu(r) - C_{2\mu}) + \widetilde{B}_1 v + B_2 u(r)]^T P_\mu^T \nabla_\mu g + g_t \Big\}\, dr \ge \frac{\rho\theta}{2}
\end{aligned}
$$

*for all $\rho \in [0, \rho_0]$.*

# C Proofs of Perturbation Theorems in Section 8

PROOF OF THEOREM 8.5.

Note that the observation process, $\tilde{y}^\varepsilon$ (for each $\varepsilon > 0$) will be fixed, but that the corresponding $p_t^\varepsilon(x)$ is represented as a supremum of the cost over possible system trajectories $\xi^\varepsilon$ which may have generated $\tilde{y}^\varepsilon$, but are not necessarily the same as $\tilde{\xi}^\varepsilon$.

By definition (8.4) and Assumption 8.3

$$
\begin{aligned}
p_t^\varepsilon(x) \le \sup_{v\in L_2}\Big\{&\phi_0(\xi^\varepsilon(0)) + \tfrac{1}{2}\int_0^t [|C_1(\xi^\varepsilon(s)) + D_{12}\tilde{u}^\varepsilon(s)|^2 \\
&-\gamma_0^2(|\tilde{y}_\mu^\varepsilon(s) - C_{2\mu}(\xi^\varepsilon(s))|^2 + |v(s)|^2 + \tfrac{1}{\varepsilon}|\tilde{y}_\kappa^\varepsilon(s) - \xi_\kappa^\varepsilon(s)|^2)]\,ds \\
&-\tfrac{1}{2}\int_0^t(\gamma^2-\gamma_0^2)[|\tilde{y}_\mu^\varepsilon(s) - C_{2\mu}(\xi^\varepsilon(s))|^2 + |v(s)|^2]\,ds \\
&-\frac{1}{2\varepsilon}\int_0^t\Big[\gamma^2-\gamma_0^2\Big]|\tilde{y}_\kappa^\varepsilon - \xi_\kappa^\varepsilon(s)|^2\,ds\Big\} \\
\le -\tfrac{1}{2}&\int_0^t(\gamma^2-\gamma_0^2)[|\tilde{y}_\mu^\varepsilon(s) - C_{2\mu}(\xi^\varepsilon(s))|^2 + |v(s)|^2]\,ds \\
&-\frac{1}{2\varepsilon}\int_0^t\Big[\gamma^2-\gamma_0^2\Big]|\tilde{y}_\kappa^\varepsilon - \xi_\kappa^\varepsilon(s)|^2\,ds\Big\}
\end{aligned}
$$

where $\xi^\varepsilon$ satisfies (3.1) driven by some $v \in L_2(0,t)$, $w = \tilde{y}_\mu^\varepsilon - C_{2\mu}(\xi^\varepsilon)$, and known controller $\tilde{u}^\varepsilon$ (the control being known since it is in the past) with terminal condition $\xi^\varepsilon(t) = x$. This implies,

$$
p_t^\varepsilon(x) \le \sup_{v\in L_2}\Big\{-\int_0^t[\eta|v|^2 + \eta_\varepsilon|\tilde{y}_\kappa^\varepsilon(s) - \xi_\kappa^\varepsilon(s)|^2]\,ds\Big\} \tag{C.1}
$$

where $\eta_\varepsilon := \frac{1}{2}(\gamma_0^2 - \gamma^2)/\varepsilon$ and $\eta := \frac{1}{2}(\gamma^2 - \gamma_0^2)$.

Now we suppose $\Delta_1 \doteq |\tilde{y}_\kappa(t) - x_\kappa| > 0$, and we use this to obtain the asserted bound (where we recall from the assumptions that $\tilde{y}$ corresponds to the perfect measurement system).

By the triangle inequality and a simple bound on the square of a sum, we have

$$
-|\tilde{y}_\kappa^\varepsilon(r) - \xi_\kappa^\varepsilon(r)|^2 \le |\tilde{y}_\kappa^\varepsilon(r) - \tilde{y}_\kappa(r)|^2 - \tfrac{1}{2}|\tilde{y}_\kappa(r) - \xi_\kappa^\varepsilon(r)|^2.
$$

Also,

$$
\begin{aligned}
|\tilde{y}_\kappa^\varepsilon(r) - \tilde{y}_\kappa(r)|^2 &\le 2|\tilde{\xi}_\kappa^\varepsilon(r) - \tilde{y}_\kappa(r)|^2 + 2\varepsilon|\tilde{\overline{w}}|^2 \\
&\le 2\alpha(\varepsilon) + 2\varepsilon|\tilde{\overline{w}}|^2
\end{aligned}
$$

where there exists $C_1 < \infty$ such that $\limsup_{\varepsilon\downarrow 0}[\alpha(\varepsilon)/\varepsilon] \le C_1$ by Assumption 8.1.

Substituting these inequalities into (C.1) we get

$$
p_t^\varepsilon(x) \le \sup_{v\in L_2}\Big\{-\int_s^t[\eta|v|^2 + \eta_\varepsilon|\tilde{y}_\kappa(r) - \xi_\kappa^\varepsilon(r)|^2]\,dr\Big\} + 2\eta t\alpha(\varepsilon)/\varepsilon + 2\eta \parallel \tilde{\overline{w}} \parallel_{L^2}^2 \tag{C.2}
$$

for any $0 \le s \le t$.

By the triangle inequality

$$\begin{aligned} |\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)| &\ge |\tilde{y}_\kappa(t) - \xi^\varepsilon_\kappa(t)| - [|\tilde{y}_\kappa(t) - \tilde{y}_\kappa(r)| + |\xi^\varepsilon_\kappa(t) - \xi^\varepsilon_\kappa(r)|] \\ &= \Delta_1 - [|\tilde{y}_\kappa(t) - \tilde{y}_\kappa(r)| + |\xi^\varepsilon_\kappa(t) - \xi^\varepsilon_\kappa(r)|]. \end{aligned}$$

By the continuity of $\tilde{y}$, there exists $\nu > 0$ such that for all $r \in [t-\nu, t]$

$$|\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)| \ge \frac{\Delta_1}{2} - |\xi^\varepsilon_\kappa(t) - \xi^\varepsilon_\kappa(r)|. \tag{C.3}$$

Note by (3.1) with the above inputs, one has

$$\dot{\xi^\varepsilon} = A(\xi^\varepsilon) + B_1(\xi^\varepsilon)(\tilde{y}^\varepsilon_\mu - C_{2\mu}(\xi^\varepsilon)) + \widetilde{B}_1(\xi^\varepsilon)v + B_2(\xi^\varepsilon)\tilde{u}^\varepsilon.$$

Using this and Assumptions 6.1 and 8.4 in (C.3) yields

$$|\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)| \ \ \ge \frac{\Delta_1}{2} - c\left[(t-r) + \|\tilde{y}^\varepsilon\|_{L_2}\sqrt{t-r} + \|\tilde{u}^\varepsilon\|_{L_2}\sqrt{t-r} + \|v\|_{L_2}\sqrt{t-r}\right]$$

and so for some smaller $\nu$ (using Assumption 8.1 for uniform bounds on $\|\tilde{y}^\varepsilon\|, \|\tilde{u}^\varepsilon\|$)

$$\ge \frac{\Delta_1}{4} - c\|v\|_{L_2}\sqrt{t-r} \qquad \forall\, r \in [t-\nu, t],$$

or, squaring

$$2|\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)|^2 + 2c^2\left[\int_r^t |v|^2\,dr\right](t-r) \ge \frac{\Delta_1^2}{16}.$$

Integrating, one has

$$\int_s^t |\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)|^2\,dr + c^2\left[\int_s^t |v(r)|^2\,dr\right](t-s)^2 \ge \frac{\Delta_1^2}{32}(t-s)$$

which implies

$$\int_s^t |\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)|^2\,dr + \left[\int_s^t |v(r)|^2\,dr\right](t-s)^2 \ge \frac{\Delta_1^2}{32(1+c^2)}(t-s).$$

Let $\tau = t-s$, and substitute this last inequality into (C.2) to obtain

$$p^\varepsilon_t(x) \le \sup_{v\in L_2}\left\{-(1-\tau^2/\varepsilon)\int_s^t \eta|v|^2\,ds\right\} - \eta^\varepsilon\frac{\Delta_1^2}{32(1+c^2)}\tau + 2\eta t\alpha(\varepsilon)/\varepsilon + 2\eta \parallel \widetilde{\widetilde{w}} \parallel^2_{L^2}$$

Choosing $\tau = \varepsilon^{1/2}$, and taking $\varepsilon$ sufficiently small we obtain

$$p^\varepsilon_t(x) \le -\frac{\eta\frac{\Delta_1^2}{32(1+c^2)}}{\varepsilon^{1/2}} + \alpha_1 + (\gamma^2 - \gamma_0^2) \parallel \widetilde{\widetilde{w}} \parallel^2_{L^2} \tag{C.4}$$

for suitable $\alpha_1 > 0$. This proves (8.12) for $\Delta_1 > 0$. □

Proof of Theorem 8.9.

The proof that the left hand side is bounded by the first term in the maximum (note minus sign multiplying maximum) in (8.12) is identical to that in the proof of Theorem 8.5 above, and so is not repeated here. We now turn to proving that the left hand side is bounded by the second term as well. This will complete the proof of Theorem 8.9. Suppose $\Delta_2 \doteq |\tilde{\xi}_\mu(t) - x_\mu| > 0$.

Let $\xi^\varepsilon$ be generated by (3.1) driven by some $v \in L_2(0,t)$, $w = \widetilde{y}^\varepsilon_\mu - C_{2\mu}(\xi^\varepsilon)$, and known controller $\tilde{u}^\varepsilon$ (the control being known since it is in the past) with terminal condition $\xi^\varepsilon(t) = x$. As in the proof of Theorem 8.9, these $\xi^\varepsilon$ trajectories (for different inputs $v$) are the "test trajectories" over which one takes the supremum in the definition of $p^\varepsilon_t(x)$. Applying the triangle inequality,

$$|\tilde{\xi}_\mu(s) - \xi^\varepsilon_\mu(s)| \geq |\tilde{\xi}_\mu(t) - \xi^\varepsilon_\mu(t)| - \left[|\tilde{\xi}_\mu(t) - \tilde{\xi}_\mu(s)| + |\xi^\varepsilon_\mu(t) - \xi^\varepsilon_\mu(s)|\right]. \tag{C.5}$$

If we take $\nu > 0$ sufficiently small, then by continuity

$$|\widetilde{\xi}_\mu(t) - \widetilde{\xi}_\mu(s)| \leq \frac{\Delta_2}{4} \qquad \forall\, s \in [t-\nu, t].$$

Combining this with (C.5) yields

$$|\widetilde{\xi}_\mu(s) - \xi^\varepsilon_\mu(s)| \geq \frac{3\Delta_2}{4} - |\xi^\varepsilon_\mu(t) - \xi^\varepsilon_\mu(s)|. \tag{C.6}$$

for all $s \in [t-\nu, t]$ for $\nu$ sufficiently small.

Note by (3.1) with the above inputs, one has

$$\dot{\xi^\varepsilon} = A(\xi^\varepsilon) + B_1(\xi^\varepsilon)(\widetilde{y}^\varepsilon_\mu - C_{2\mu}(\xi^\varepsilon)) + \widetilde{B}_1(\xi^\varepsilon)v + B_2(\xi^\varepsilon)\tilde{u}^\varepsilon.$$

Using this and Assumptions 6.1 and 8.4 yields (for proper choice of $c < \infty$)

$$|\xi^\varepsilon_\mu(t) - \xi^\varepsilon_\mu(s)| \quad \leq c\left[(t-s) + \|\tilde{y}^\varepsilon\|_{L_2}\sqrt{t-s} + \|\tilde{u}^\varepsilon\|_{L_2}\sqrt{t-s} + \|v\|_{L_2}\sqrt{t-s}\right]$$

and so for some smaller $\nu$ (using Assumption 8.1 for uniform bounds on $\|\tilde{y}^\varepsilon\|, \|\tilde{u}^\varepsilon\|$)

$$\leq \frac{\Delta_2}{4} + c\|v\|_{L_2}\sqrt{t-s} \qquad \forall\, s \in [t-\nu, t],$$

Combine this with (C.6) to find that

$$|\tilde{\xi}_\mu(s) - \xi^\varepsilon_\mu(s)| \geq \frac{\Delta_2}{2} - c\|v\|_{L_2(t-\nu,t)}\sqrt{\nu} \qquad \forall\, s \in [t-\nu, t].$$

Squaring,

$$|\tilde{\xi}_\mu(s) - \xi^\varepsilon_\mu(s)|^2 \geq \frac{\Delta_2^2}{8} - c^2\|v\|^2_{L_2(t-\nu,t)}\nu \qquad \forall\, s \in [t-\nu, t]. \tag{C.7}$$

Now suppose that $v$ is such that

$$\|v\|^2_{L_2(t-\nu,t)} \le \frac{\Delta_2^2}{\varepsilon^{1/5}}. \tag{C.8}$$

Combining (C.7) and (C.8) we get

$$|\tilde{\xi}_\mu(s) - \xi^\varepsilon_\mu(s)|^2 \ge \Delta_2^2(\frac{1}{8} - \frac{c^2\nu}{\varepsilon^{1/5}}),$$

and selecting

$$\nu = \nu^\varepsilon = \varepsilon^{1/5}/(16c^2) \tag{C.9}$$

(where we restrict $\varepsilon$ to be sufficiently small so that $\nu^\varepsilon$ is no larger than our above bounds on $\nu$) gives

$$|\tilde{\xi}_\mu(s) - \xi^\varepsilon_\mu(s)|^2 \ge \frac{\Delta_2^2}{16} \qquad \forall\, s \in [t-\nu^\varepsilon, t]. \tag{C.10}$$

Now, applying *noise insensitive observability* (i.e. Assumption 8.6) with trajectories $\tilde{\xi}$ and $\xi^\varepsilon$, one has (for $\delta, \varepsilon$ sufficiently small with $\nu = \nu^\varepsilon = \varepsilon^{1/5}/(16c^2)$ of course)

$$\begin{aligned}
&\|\widetilde{y}_\mu - \widetilde{y}^\varepsilon_\mu\|_{L_2(t-\nu,t)} + \|\tilde{v} - v\|_{L_2(t-\nu,t)} + \nu\delta\|\tilde{u} - \tilde{u}^\varepsilon\|_{L_2(t-\nu,t)} \\
&+\frac{1}{\nu\delta}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|_{L_2(t-\nu,t)} \ge g\left(\Delta_2^2/16\right)\sqrt{\nu}.
\end{aligned}$$

Squaring, using a simple inequality and rearranging yields

$$\begin{aligned}
&\nu^2\delta^2\|\tilde{v} - v\|^2_{L_2(t-\nu,t)} + \frac{1}{\nu^2\delta^2}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|^2_{L_2(t-\nu,t)} \\
&\ge \frac{\nu}{4}g^2\left(\Delta_2^2/16\right) - \left[\|\widetilde{y}_\mu - \widetilde{y}^\varepsilon_\mu\|^2_{L_2(t-\nu,t)} + \|\tilde{u} - \tilde{u}^\varepsilon\|^2_{L_2(t-\nu,t)}\right].
\end{aligned}$$

Using the triangle inequality and rearranging, this yields

$$\begin{aligned}
&2\nu\delta^2\|v\|^2_{L_2(t-\nu,t)} + \frac{1}{\nu^3\delta^2}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|^2_{L_2(t-\nu,t)} \\
&\ge \frac{1}{4}g^2\left(\Delta_2^2/16\right) - \frac{1}{\nu}\left[\|\widetilde{y}_\mu - \widetilde{y}^\varepsilon_\mu\|^2_{L_2(t-\nu,t)} + \|\tilde{u} - \tilde{u}^\varepsilon\|^2_{L_2(t-\nu,t)} + 2\nu^2\delta^2\|\tilde{v}\|^2_{L_2(t-\nu,t)}\right].
\end{aligned}$$

Using (C.9), and taking $\delta^2 = \varepsilon^{1/5}$ yields

$$\begin{aligned}
&2\varepsilon^{1/5}\nu\|v\|^2_{L_2(t-\nu,t)} + \frac{(16c^2)^3}{\varepsilon^{4/5}}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|^2_{L_2(t-\nu,t)} \\
&\ge \left\{\frac{1}{4}g^2\left(\Delta_2^2/16\right) - \frac{16c^2}{\varepsilon^{1/5}}\left[\|\widetilde{y}_\mu - \widetilde{y}^\varepsilon_\mu\|^2_{L_2(t-\nu,t)} + \|\tilde{u} - \tilde{u}^\varepsilon\|^2_{L_2(t-\nu,t)}\right] - \frac{\varepsilon^{2/5}}{8c^2}\|\tilde{v}\|^2_{L_2(t-\nu,t)}\right\}.
\end{aligned}$$

Note also, that by Assumption 6.1, there exists $\overline{K} < \infty$ such that $\|\widetilde{y}_\mu - \widetilde{y}^\varepsilon_\mu\|^2_{L_2(t-\nu,t)} = \|C_{2\mu}(\tilde{\xi}) - C_{2\mu}(\tilde{\xi}^\varepsilon)\|^2_{L_2(t-\nu,t)} \le \overline{K}\|\tilde{\xi} - \tilde{\xi}^\varepsilon\|^2_{L_2(t-\nu,t)}$. Consequently, one has

$$\begin{aligned}
&2\varepsilon^{1/5}\nu\|v\|^2_{L_2(t-\nu,t)} + \frac{(16c^2)^3}{\varepsilon^{4/5}}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|^2_{L_2(t-\nu,t)} \\
&\ge \frac{1}{4}g^2\left(\Delta_2^2/16\right) - \frac{16c^2}{\varepsilon^{1/5}}\left[\overline{K}\|\tilde{\xi} - \tilde{\xi}^\varepsilon\|^2_{L_2(t-\nu,t)} + \|\tilde{u} - \tilde{u}^\varepsilon\|^2_{L_2(t-\nu,t)}\right] - \frac{\varepsilon^{2/5}}{8c^2}\|\tilde{v}\|^2_{L_2(t-\nu,t)}.
\end{aligned}$$

Now, using Assumption 8.1, we see that for possibly smaller $\varepsilon > 0$ and corresponding $\nu = \nu^\varepsilon$,

$$\frac{(16c^2)^3}{\varepsilon^{4/5}}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|^2_{L_2(t-\nu,t)} \geq \left\{\frac{1}{5}g^2\left(\Delta_2^2/16\right)\right\} - 2\varepsilon^{1/5}\nu\|v\|^2_{L_2(t-\nu,t)}$$

which using (C.8) and (C.9)

$$\geq \left\{\frac{1}{5}g^2\left(\Delta_2^2/16\right)\right\} - 2\varepsilon^{1/5}\frac{\Delta_2^2}{16c^2}.$$

Consequently, for possibly smaller $\varepsilon > 0$ and corresponding $\nu = \nu^\varepsilon$,

$$\frac{(16c^2)^3}{\varepsilon^{4/5}}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|^2_{L_2(t-\nu,t)} \geq \frac{1}{6}g^2\left(\Delta_2^2/16\right).$$

Then, recalling the details of Assumption 8.6, one has

$$\frac{(16c^2)^3}{\varepsilon}\|\widetilde{\xi}_\kappa - \xi^\varepsilon_\kappa\|^2_{L_2(t-\nu,t)} \geq \frac{1}{6\varepsilon^{1/5}}g^2\left(\Delta_2^2/16\right) \geq \frac{1}{\varepsilon^{1/5}}\left(\frac{\beta^2}{6}\right)\left|\Delta_2^2/16\right|^{2\zeta}. \tag{C.11}$$

Recall that this was obtained under (C.8); alternatively, one has

$$\|v\|^2_{L_2(t-\nu,t)} > \frac{\Delta_2^2}{\varepsilon^{1/5}}. \tag{C.12}$$

Combining (C.11) and (C.12), one sees that in either case

$$\begin{aligned}\int_s^t [\eta|v|^2 + \eta_\varepsilon|\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)|^2]\,dr &\geq \int_{t-\nu^\varepsilon}^t [\eta|v|^2 + \eta_\varepsilon|\tilde{y}_\kappa(r) - \xi^\varepsilon_\kappa(r)|^2]\,dr \\ &\geq \frac{c_1}{\varepsilon^{1/5}}\min\left\{\Delta_2^2,\, \Delta_2^{4\zeta}\right\}\end{aligned}$$

for proper choice of $c_1 < \infty$. Combining this with (C.2) yields the remaining bound asserted in the theorem. □

# D Expression for Maximum in PDE

In this appendix, we obtain equation (3.10).

Let

$$\begin{aligned} c &\doteq [A + B_1(y_\mu(t) - C_{2\mu}) + B_2u(t)]^T P_\mu^T \nabla_\mu p_t - \tfrac{1}{2}|C_1 + D_{12}u(t)|^2 + \gamma^2|C_{2\mu} - y_\mu(t)|^2 \\ b^T &\doteq \nabla_\mu p_t P_\mu \widetilde{B}_1 \\ Q &\doteq \dot{y}_\kappa - P_\kappa(A + B_1(y_\mu(t) - C_{2\mu}) + B_2u(t)) \\ F &\doteq P_\kappa \widetilde{B}_1. \end{aligned} \tag{D.1}$$

Then, $v^*$ is a minimizer of the problem

$$\begin{array}{l} \text{Minimize } c + b^T v + \gamma^2 |v|^2 \\ \text{subject to } Fv = Q. \end{array} \tag{D.2}$$

Let $F^{-R}$ be a right inverse of $F$ (i.e. $FF^{-R} = I$). Then by the side condition, we must have

$$v^* = F^{-R}Q + \eta$$

where $F\eta = 0$. Let $P$ (a $q \times q$ matrix) be a projection onto the null space of $F$ (i.e. $FPr = 0$ for all $r \in \mathbf{R}^q$). Then

$$v^* = F^{-R}Q + Pr^* \tag{D.3}$$

for some $r^*$ achieving the minimum.

Define

$$\overline{g}(r) \doteq c + b^T(F^{-R}Q + Pr) + \gamma^2 |F^{-R}Q + Pr|^2.$$

Then, for any $h \in \mathbf{R}^q$, one must have

$$0 = \frac{d}{dt}[\overline{g}(r^* + th)]|_{t=0} = h^T P^T [b + 2\gamma^2 F^{-R}Q + 2\gamma^2 Pr^*]$$

which implies

$$0 = P^T [b + 2\gamma^2 F^{-R}Q + 2\gamma^2 Pr^*].$$

Now, using $P = P^T$ and $PP = P$, this yields

$$Pr^* = -PF^{-R}Q - \frac{1}{2\gamma^2}Pb$$

Substituting this into (D.3) yields

$$v^* = (I - P)F^{-R}Q - \frac{1}{2\gamma^2}Pb$$

which is (3.10).

# Thanks

Thanks to Andy Packard for pointing out the references on linear problems with measurements of $P_\kappa x$. Thanks are due to Carl Nett, Meir Pachter, and Kevin Wise for valuable conversations.

# References


[1] R.A. Adams, "Sobolev Spaces ", Academic Press, 1975.

[2] L. Baramov, "$H^\infty$ control of nonlinear nonstandard systems based on factorization", MTNS 98.

[3] L. Baramov, "Solutions to a class of nonstandard nonlinear $H^\infty$ control problems", to appear Intern. Jour. of Control

[4] M. Bardi and I. Capuzzo-Dolcetta, "Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations", Birkhauser, Boston, 1997.

[5] T. Basar and P. Bernhard, *$H_\infty$ Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, Birkhuaser, Boston, 1995.

[6] G. Didinsky, Z. Pan and T. Basar, *Parameter identification for Uncertain Plants using $H^\infty$ Methods*, Automatica, 31(9), 1995, 1227-1250.

[7] J. Doyle, K. Glover, P. P. Kargonekar, B. A. Francis *State space solutions to standard $H^2$ and $H^\infty$ control problems.* TAC vol. 34 no. 8 p831 - 847, 1989

[8] P. Dupuis and M. Boué, *Markov chain approximations for deterministic control problems with affine dynamics and quadratic cost in the control*, SIAM J. on Numerical Analysis, **36** (1999), pp. 667–695.

[9] P. Dupuis and A. Szpiro, *Convergence of the optimal feedback policies in a numerical method for a class of deterministic optimal control problems*, to appear in SIAM J. on Control and Opt.

[10] K.M. Eveker, D.L. Gysling, C.N. Nett, and O.P Sharma "Integrated Control of Rotating Stall and Surge in Aeroengines," *1995 SPIE Conf. on Sensing, Actuation, and Control in Aeropropulsion.*

[11] W.H. Fleming and H.M. Soner, "Controlled Markov Process and Viscosity Solutions", Springer-Verlag, New York, 1993.

[12] J.W. Helton and M.R. James, *Reduction of Controller Complexity in Nonlinear $H_\infty$ Control*, Proc. 34th IEEE CDC, New Orleans, Dec. 1995.

[13] J.W. Helton and M.R. James, *On the Stability of the Information State System,* Sys Control Letters vol 29 (1996) p61-72

[14] J.W. Helton and M.R. James, *On Validating the Certainty Equivalence Assumptions in Nonlinear $H_\infty$ Control*, Proc. 37th IEEE CDC, Dec. 1998.

[15] J.W. Helton and M.R. James, "Extending $H_\infty$ Control to Nonlinear Systems: Control of Nonlinear Systems to Achieve Performance Objectives", SIAM 1999.

[16] R.A. Horn and C.R. Johnson, "Matrix Analysis", Cambridge Univ. Press 1985.

[17] M.R. James, *On the Certainty Equivalence Principle and the Optimal Control of Partially Observed Dynamic Games*, IEEE Trans. Aut. Control, 39(11) 1994, 2321-2324.

[18] M.R. James and J.S. Baras, *Robust $H_\infty$ Output Feedback Control for Nonlinear Systems*, IEEE TAC **40**, 1995, 1007-1017.

[19] M.R. James and J.S. Baras, *Partially Observed Differential Games, Infinite Dimensional HJB Equations, and Nonlinear $H_\infty$ Control*, SIAM J. Control Optim., 34, 1996, 1342-1364.

[20] M. Joannides and F. Le Gland, *Nonlinear Filtering with Continuous Time Perfect Observations and Noninformative Quadratic Variation*, Proc. 36th IEEE CDC, Dec. 1997, 1645–1650.

[21] H.J. Kushner and P. Dupuis,

[22] W.M. McEneaney, "Robust control and differential games on a finite time horizon", Math. of Controls Signals and Systems, **8** (1995), 138–166.

[23] W. M. McEneaney, "A Uniqueness Result for the Isaacs equation corresponding to Nonlinear $\mathrm{H}_\infty$ control", Math. Controls, Signals and Systems, **11** (1998), 303–334.

[24] W. M. McEneaney, "Elimination of troublesome disturbances with application to representation results for $\mathrm{H}_\infty$ control DPEs", Proc. Seventh International Symposium on Dynamic Games and Applications, Shonan Village (1996), 662–671.

[25] W.M. McEneaney, "Robust/$\mathrm{H}_\infty$ filtering for nonlinear systems", Systems and Control Letters, **33** (1998), 315–325.

[26] M.M. Seron, J.H. Braslavsky, P.V. Kokotovic, D.Q. Mayne, *Feedback Limitations in Nonlinear Systems: From Bode Diagrams to Cheap Control*, IEEE Trans. Aut. Control, **44**(4), 829-833, 1999.

[27] A.A. Stoorvogel, A. Saberi and B.M. Chen, IEEE TAC,vol 39, Feb 1994, pp 355-360

[28] A. van der Schaft, "$L_2$-Gain and Passivity Techniques in Nonlinear Control", Springer Verlag, New York, 1996.

[29] Xin Guo, Feng, TITLE ??XX Automatica, 1996, pp 1581-1585