Lower bounds on the performance of Analog to Digital Converters

Citation: Osqui, Mitra, Alexandre Megretski, and Mardavij Roozbehani. "Lower Bounds on the Performance of Analog to Digital Converters." 50th IEEE Conference on Decision and Control and European Control Conference 2011 (CDC-ECC). 1036–1041.

As Published: http://dx.doi.org/10.1109/CDC.2011.6161525

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/72551

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Lower Bounds on the Performance of Analog to Digital Converters

Mitra Osqui† Alexandre Megretski‡ Mardavij Roozbehani♯

Abstract

This paper deals with the task of finding certified lower bounds for the performance of Analog to Digital Converters (ADCs). A general ADC is modeled as a causal, discrete-time dynamical system with outputs taking values in a finite set. We define the performance of an ADC as the worst-case average intensity of the filtered input matching error, defined as the difference between the input and output of the ADC. The passband of the shaping filter used to filter the error signal determines the frequency region of interest for minimizing the error. The problem of finding a lower bound for the performance of an ADC is formulated as a dynamic game problem in which the input signal to the ADC plays against the output of the ADC. Furthermore, the performance measure must be optimized in the presence of quantized disturbances (output of the ADC) that can exceed the control variable (input of the ADC) in magnitude. We characterize the optimal solution in terms of a Bellman-type inequality. A numerical approach is presented to compute the value function in parallel with the feedback law for generating the worst case input signal. The specific structure of the problem is used to prove certain properties of the value function that simplifies the iterative computation of a certified solution to the Bellman inequality. The solution provides a certified lower bound on the performance of any ADC with respect to the selected performance criteria.

I. INTRODUCTION AND MOTIVATION

Analog to Digital Converters (ADCs) act as the interface between the analog world and digital processors. They are present in almost all digital control and communication systems and modern high-speed data conversion and storage systems. Naturally, the design and analysis of ADCs have, for many years, attracted the attention and interest of researchers from various disciplines across academia and

Project partially supported by: Army Research Office ELASTx program.

†Mitra Osqui is currently a Ph.D. candidate at the department of EECS, Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology, Cambridge, MA. E-mail: mitra@mit.edu

‡ Alexandre Megretski is currently a professor of EECS at LIDS at MIT, Cambridge, MA. E-mail: ameg@mit.edu.

♯ Mardavij Roozbehani is currently a principal research scientist at LIDS at MIT, Cambridge, MA. E-mail: mardavij@mit.edu.

industry. Despite the progress that has been made in this field, the design of optimal ADCs remains an open challenging problem, and the fundamental limitations of their performance are not well understood. This paper is concerned with the latter problem.

A particular class of ADCs primarily used in high resolution applications is the Delta-Sigma Modulator (DSM). Fig. 1, illustrates the classical first-order DSM [1], where Q is a quantizer with uniform step size.

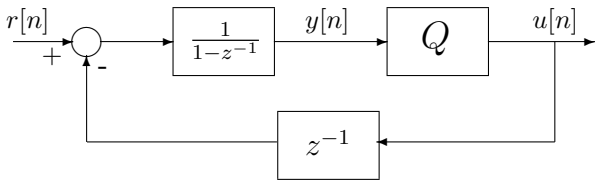


Fig. 1: Classical First-Order Sigma-Delta Modulator

An extensive body of research on DSMs has appeared in the signal processing literature. One well known approach is based on linearized additive noise models and filter design for noise shaping [1]-[5]. The underlying assumption for validity of the linearized additive noise model is availability of a relatively high number of bits. Alternative approaches based on a formalism of the signal transformation performed by the quantizer have been exploited for deterministic analysis in [6]-[8]. Some other works that do not use linearized additive noise models are reported in [9]-[11].

In the control field, [12]-[14] find performance bounds and suboptimal policies for linear stochastic control problems using Bellman inequalities with quadratic value functions. The problem is relaxed and solved using linear matrix inequalities and semidefinite programming. For references on quantized control, please see [15]-[17].

In [18] we provided a characterization of the solution to the optimal ADC design problem and presented a generic methodology for numerical computation of sub-optimal solutions along with computation of a certified upper bound on the performance. The performance of an ADC is evaluated with respect to a cost function which is a measure of the intensity of the error signal (the difference between the input signal and its quantized version) for the worst case input. The error signal is passed through a shaping filter which dictates the frequency region in which the error is to be minimized. Furthermore, we showed that the dynamical system within the optimal ADC is a copy of the shaping filter used to define the performance criteria. In [18] we also presented an exact analytical solution to the optimal ADC for first-order shaping filters, and showed that the classical first-order DSM (Figure 1) is identical to our optimal ADC. This result proved the optimality of the classical first-order DSM with respect to

the adopted performance measure, and was a step towards understanding the limitations of performance.

In this paper, we present a framework for finding *certified lower bounds* for the performance of ADCs with shaping filters of arbitrary order. We use the same ADC model and performance measure adopted in [18]. The objective is to find a lower bound on the infimum of the cost function. The approach is to find a feedback law for generating the input of the ADC such that regardless of its output, the performance is bounded from below by a certain value. Thus, the input of the ADC is viewed as the control, and the problem is posed within a non-linear optimal feedback control/game framework. We show that the optimal control law can be characterized in terms of a *value function* satisfying an analog of the Bellman inequality. The value function in the Bellman inequality and the corresponding control law can be jointly computed via value iteration.

Since searching for the value function involves solving a sequence of infinite dimensional optimization problems, some approximations are needed for numerical computation. First, a finite-dimensional parameterization of the value function is selected. Second, the state space and the input space are discretized. Third, the computations are restricted to a finite subset of the space. The latter step deserves further elaboration. If the dynamical system inside the ADC is strictly stable, then a bounded control invariant set exists, thus it is possible to do the computations over a bounded region. The challenge arises when the filter has poles on the unit circle. In this case, there does not exist a bounded control invariant set, since the disturbances can exceed the control variable in magnitude. Under the condition that there is at most one pole on the unit circle, we present a theorem that states that the value function is zero outside a certain bounded space. Thus, we have an a priori knowledge of an analytic expression for the value function beyond a bounded region. As a result, the computations need to be carried out only over this bounded region. This is in dramatic contrast with the case of upper bound computations [18], something to be discussed in section III.

The organization is as follows. Section II provides a rigorous problem formulation. The main contributions are presented in Section III and IV. Section III describes our methodology for finding certified lower bounds for ADCs. Section IV provides our theoretical results. We provide an example in section V, and section VI concludes the paper.

Notation and Terminology:

- Function $f : \mathbb{R}^m \mapsto \mathbb{R}$ is called BIBO, if the image of every bounded subset $\Omega \subset \mathbb{R}^m$ under f, $f(\Omega)$, is bounded.

- Given a set P, $\ell_+(P)$ is the set of all one-sided sequences x with values in P, i.e. functions $x : \mathbb{Z}_+ \mapsto P$.
- The $\infty-$norm is defined as:

$$\|v\|_\infty = \max |v_i|, \quad \text{for} \quad v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \in \mathbb{R}^m$$

and

$$\|M\|_\infty \stackrel{\text{def}}{=} \sup_{v \neq 0} \frac{\|Mv\|_\infty}{\|v\|_\infty} = \max_{i \in \{1, \cdots, l\}} \sum_{j=1}^{m} |M_{ij}|$$

for a matrix $M = (M_{ij}) \in \mathbb{R}^{l \times m}$.

- Let X be a set and $f : X \mapsto \mathbb{R}$ be a function. For every $\epsilon > 0$,

$$\arg^\epsilon \sup_{x \in X} f(x) \stackrel{\text{def}}{=} \left\{ x \in X : f(x) > -\epsilon + \sup_{x \in X} f(x) \right\}. \tag{1}$$

II. PROBLEM FORMULATION

The problem setup in this section is taken from [18].

A. Analog to Digital Converters

In this paper, a general ADC is viewed as a causal, discrete-time, non-linear system Ψ, accepting arbitrary inputs in the $[-1, 1]$ range, and producing outputs in a fixed finite subset $U \subset \mathbb{R}$, as shown in Fig. 2. We assume that the smallest element in the set U is less than -1 and the largest element is greater than 1.

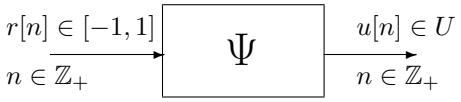


Fig. 2: Analog to Digital Converter as a Dynamical System

Equivalently, an ADC is defined by a sequence of functions $\Upsilon_n : [-1, 1]^{n+1} \mapsto U$ according to

$$\Psi : \; u[n] = \Upsilon_n \left(r[n], r[n-1], \cdots, r[0] \right), \quad n \in \mathbb{Z}_+. \tag{2}$$

The class of ADCs defined above is denoted by $\mathcal{Y}_U$.

B. Asymptotic Weighted Average Intensity (AWAI) of a Signal

The Asymptotic Weighted Average Intensity $\eta_{G,\phi}(w)$ of a signal w with respect to the transfer function $G(z)$ of a strictly causal LTI dynamical system L_G and a non-negative function $\phi : \mathbb{R} \mapsto \mathbb{R}_+$ is given by:

$$\eta_{G,\phi}(w) = \limsup_{N \mapsto \infty} \frac{1}{N+1} \sum_{n=0}^{N} \phi(q[n]), \tag{3}$$

where the sequence q is the response to input w of the dynamical system:

$$L_G : \begin{array}{c} x[n+1] = Ax[n] + Bw[n], \quad x[0] = 0, \\ q[n] = Cx[n] \end{array} \tag{4}$$

and A, B, C are given matrices of appropriate dimensions. Examples of functions ϕ to consider are: $\phi(q) = |q|$ and $\phi(q) = |q|^2$.

C. ADC Performance Measure

The setup that we use to measure the performance of an ADC is illustrated in Fig. 3. The performance measure of $\Psi \in \mathcal{Y}_U$, denoted by $\mathcal{J}_{G,\phi}(\Psi)$, is the worst-case AWAI of the error signal for all input sequences $r \in \ell_+([-1,1])$, that is:

$$\mathcal{J}_{G,\phi}(\Psi) = \sup_{r \in \ell_+([-1,1])} \eta_{G,\phi}(r - \Psi(r)). \tag{5}$$

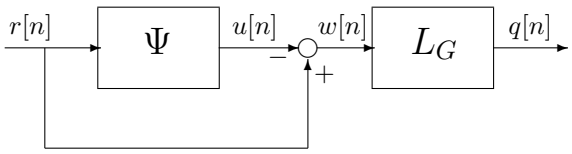


Fig. 3: Setup Used for Measuring the Performance of the ADC

D. ADC Optimization

Given L_G and ϕ, we consider $\Psi_o \in \mathcal{Y}_U$ an optimal ADC if $\mathcal{J}_{G,\phi}(\Psi_o) \leq \mathcal{J}_{G,\phi}(\Psi)$ for all $\Psi \in \mathcal{Y}_U$. The corresponding optimal performance measure $\gamma_{G,\phi}(U)$ is defined as

$$\gamma_{G,\phi}(U) = \inf_{\Psi \in \mathcal{Y}_U} \mathcal{J}_{G,\phi}(\Psi). \tag{6}$$

The objective is to find certified lower bounds for (6).

III. OUR APPROACH

We find the lower bound on the performance of any given ADC belonging to the class $\mathcal{Y}_U$ by associating the problem with a full-information feedback control problem. The objective is to find a feedback law for generating the input of the ADC, r, such that regardless of the output u, the performance is bounded from below by a certain value. Thus, in this setup, r is viewed as the control and u is viewed as the input of a strictly causal system with output r. The setup is depicted in Fig. 4, where the function $K_r : \mathbb{R}^m \mapsto [-1, 1]$ is said to be an admissible controller if there exists $\gamma \in [0, \infty)$ such that every triplet of sequences (x, u, r) satisfying

$$x[n+1] = Ax[n] + Br[n] - Bu[n], \quad x[0] = 0, \tag{7}$$

$$r[n] = K_r\left(x[n]\right), \tag{8}$$

$$q[n] = Cx[n], \tag{9}$$

also satisfies the dissipation inequality

$$\inf_N \sum_{n=0}^{N} \left(\phi\left(q[n]\right) - \gamma\right) > -\infty. \tag{10}$$

Note that if (10) holds subject to (7)-(9), then $\gamma_{G,\phi}\left(U\right) \geq \gamma$. Let γ_o be the minimal upper bound of γ, for which an admissible controller exists. Then K_r is said to be an optimal controller if (10) is satisfied with $\gamma = \gamma_o$.

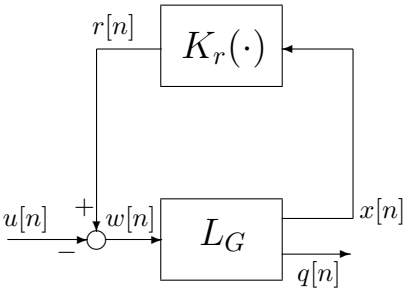


Fig. 4: Full State-Feedback Control Setup

A. The Bellman Inequality

The solution to a well-posed state-feedback optimal control problem can be characterized as the solution to the associated Bellman equation [19]-[22]. Herein, standard techniques are used to show

that a controller K_r satisfying (10) exists if and only if a solution to an analog of the Bellman equation exists. The formulation will be made more precise as follows. Define function $\sigma_\gamma : \mathbb{R}^m \mapsto \mathbb{R}$ by

$$\sigma_\gamma(x) = \gamma - \phi(Cx). \tag{11}$$

It can be shown that a controller K_r in (8) guaranteeing (10) exists if and only if there exists a function $V : \mathbb{R}^m \mapsto \mathbb{R}_+$, such that inequality

$$V(x) \geq \sigma_\gamma(x) + \inf_{r\in[-1,1]} \max_{u\in U} V(Ax + Br - Bu) \tag{12}$$

holds for all $x \in \mathbb{R}^m$ (see Theorem 1). We refer to inequality (12) as the Bellman inequality, and to a function V satisfying (12) as the value function.

B. Numerical Solutions to the Bellman Inequality

In this section, we outline our approach for numerical computation of the value function V and the control function K_r. We can simplify the problem of searching for a solution to inequality (12) by instead finding a solution $V \geq 0$ to the inequality

$$V(x) \geq \sigma_\gamma(x) + \min_{r\in\Gamma_r} \max_{u\in U} V(Ax + Br - Bu), \ \forall x \in \mathbb{R}^m \tag{13}$$

where Γ_r is a finite subset of $[-1, 1]$. Since for every function $g : [-1, 1] \mapsto \mathbb{R}$, we have

$$\inf_{r\in[-1,1]} g(r) \leq \min_{r\in\Gamma_r} g(r), \tag{14}$$

a solution V of (13) is also a solution of (12). In the remainder of this section we focus on finding a solution to (13).

A control invariant set of system (7), with respect to Γ_r, is formally defined as a subset $\mathcal{I} \subset \mathbb{R}^m$ such that:

$$\forall x \in \mathcal{I}, \ \exists r \in \Gamma_r : \ Ax + Br - Bu \in \mathcal{I}, \ \forall u \in U. \tag{15}$$

Furthermore, a strong invariant set of system (7), with respect to Γ_r, is defined as a subset $\mathcal{I} \subset \mathbb{R}^m$ such that:

$$Ax + Br - Bu \in \mathcal{I}, \quad \forall x \in \mathcal{I}, \ r \in \Gamma_r, \ u \in U. \tag{16}$$

Ideally we would like to have a bounded invariant set, so that the search for V satisfying the Bellman inequality is restricted to a bounded region of the state space. If $\max|\operatorname{eig}(A)| < 1$, then a bounded set $\mathcal{I}$ satisfying (16) is guaranteed to exist. However, if $\max|\operatorname{eig}(A)| = 1$, then there is no bounded set $\mathcal{I}$ satisfying (15), due to the assumption that the smallest element in the set U is less than -1 and

the largest element is greater than 1. Hence, the case when $\max|\operatorname{eig}(A)| = 1$ presents the challenge of searching for a numerical solution to (13) over an unbounded state space. However, for the case that there is only one pole on the unit circle, we will establish in Theorem 2 that the value function is zero for all x outside a certain bounded region. Hence, the numerical search for V satisfying (13) needs to be carried out only over a bounded subset of the state space. Next, uniform grids are created for the state space. In this paper, these are uniformly-spaced, discrete subsets of the Euclidean space, and are defined as follows. The set

$$\mathbb{G} = \{i\Delta \mid i \in \mathbb{Z}\}$$

is a grid on $\mathbb{R}$, where $D = 1/\Delta$ is a positive integer. The corresponding grid on $\mathcal{I}$ is

$$\Gamma = \mathbb{G}^m \cap \mathcal{I}.$$

Furthermore, we define $\Gamma_r = \{r_1,\ r_2, \cdots, r_L\}$ as

$$\Gamma_r = \mathbb{G} \cap [-1, 1]\,.$$

The next step is to create a finite-dimensional parameterization of V. In this paper, the search is performed over the class of *piecewise constant* (PWC) functions assuming a constant value over a *tile*. A *tile* in $\mathbb{G}^n$, $n \in \mathbb{N}$ is defined as the smallest hypercube formed by 2^n points on the grid, and thus, has $2n$ faces (the faces are hypercubes of dimension $n-1$). By convention, we assume that the n faces that contain the lexicographically smallest vertex are closed, and the rest are open. The union of all such tiles covers $\mathbb{R}^n$ and their intersection is empty. Let T_i denote the i^{th} tile over the grid $\mathbb{G}^m$, and $\mathcal{T}$ the set of all tiles that lie within $\mathcal{I}$, and N_T the number of all such tiles:

$$\mathcal{T} = \{T_i \mid i \in \{1, 2, \cdots, N_T\}\}\,.$$

The PWC parameterization of V is as follows

$$V(x) = V_i,\ \forall x \in T_i,\ i \in \{1, 2, \cdots, N_T\} \tag{17}$$

where $V_i \in \mathbb{R}_+$. We then search for a solution $V : \mathcal{I} \mapsto \mathbb{R}_+$ of (13) for all $x \in \mathcal{I}$ within the class of PWC functions defined in (17). The corresponding PWC control function $K_r : \mathcal{I} \mapsto \Gamma_r$ is given by

$$K_r(x) = \arg\min_{r \in \Gamma_r} \max_{u \in U, \bar{x} \in T(x)} V\left(A\bar{x} + Br - Bu\right),\ \forall x \in \mathcal{I}. \tag{18}$$

where $T(x) = T_i$ for $x \in T_i$. In the next subsection we show how to search and certify functions V and K_r satisfying (13) and (18).

C. Searching for Numerical Solutions

The Bellman inequality (13) is solved via value iteration. The algorithm is initialized at $\Lambda_0(x) = 0$, for all $x \in \mathcal{T}$, and at stage $k+1$ it computes a PWC function $\Lambda_{k+1} : \mathcal{T} \mapsto \mathbb{R}_+$ satisfying

$$\Lambda_{k+1}(x) = \max\left\{0,\ \sigma_\gamma(x) + \min_{r\in\Gamma_r} \max_{u\in U, \bar{x}\in T(x)} \Lambda_k(A\bar{x} + Br - Bu)\right\}. \tag{19}$$

At each stage of the iteration, Λ_{k+1} is computed and certified to satisfy (19) for all $x \in \mathcal{T}$ as follows:

1) For every $i \in \{1, 2, \cdots, N_T\}$ and $j \in \{1, 2, \cdots, L\}$, define

$$\sigma_i = \sup_{x\in T_i} \sigma_\gamma(x),$$
$$Y_{ij} = \{Ax + Br_j - Bu \mid x \in T_i,\ r_j \in \Gamma_r,\ u \in U\},$$

and find all the tiles that intersect with Y_{ij}

$$\Theta_{ij} = \{p \mid T_p \cap Y_{ij} \neq \{\emptyset\},\ p \in \{1, 2, \cdots, N_T\}\}.$$

2) Let

$$v_s = \Lambda_k(x),\ x \in T_s,\ s \in \{1, 2, \cdots, N_T\}.$$

Compute

$$v_{ij} = \max_{s\in\Theta_{ij}} v_s.$$

3) For every tile $x \in T_i$ compute PWC functions:

$$\Lambda_{k+1}(x) = \max\left\{0,\ \sigma_i + \min_j v_{ij}\right\}.$$

When the iteration converges, it converges pointwise to a limit $\Lambda : \mathcal{T} \mapsto \mathbb{R}_+$, where the limit satisfies, for all $x \in \mathcal{T}$, the equality

$$\Lambda(x) = \max\left\{0,\ \sigma_\gamma(x) + \min_{r\in\Gamma_r} \max_{u\in U, \bar{x}\in T(x)} \Lambda(A\bar{x} + Br - Bu)\right\}. \tag{20}$$

The largest γ for which (19) converges is found through line search. We take $V(x) = \Lambda(x)$, for all $x \in \mathcal{T}$. The associated suboptimal control law is a PWC function defined over all tiles T_i in the control invariant set $\mathcal{I}$ that satisfies (18).

IV. THEORETICAL STATEMENTS

In this section, we show that under some technical assumptions, the value function in (12) is zero beyond a bounded region. However, we first present a theorem that establishes the link between the full information feedback control problem and the Bellman inequality (12). Note that in this section we use subscript notation for values of sequences at specific time instances instead of the bracket notion used elsewhere in the paper, that is x_n is used in place of $x[n]$.

Theorem 1: Let X be a metric space, $\Omega \subset \mathbb{R}$ be a compact metric space, $U \subset \mathbb{R}$ be a finite set, and $f : X \times \Omega \times U \mapsto X$ and $\sigma : X \mapsto \mathbb{R}$ be continuous functions. Then the following statements are equivalent:

(i)

$$V_\infty(\bar{x}) \stackrel{\text{def}}{=} \sup_{\tau \in \mathbb{Z}_+} V_\tau(\bar{x}) < \infty, \quad \forall \bar{x} \in X, \tag{21}$$

where $V_\tau : X \mapsto \mathbb{R}_+$ is defined by

$$V_\tau(\bar{x}) = \max_{\theta_0} \min_{r_0} \max_{u_0,\theta_1} \cdots \min_{r_{\tau-2}} \max_{u_{\tau-2},\theta_{\tau-1}} \sum_{n=0}^{\tau-1} h_{n+1}\sigma(x_n), \tag{22}$$

with $r_n,\ u_n,\ \theta_n$ restricted by $r_n \in \Omega,\ u_n \in U,\ \theta_n \in \{0,1\}$ and $x_n,\ h_n$ defined by

$$x_{n+1} = f(x_n, r_n, u_n), \quad x_0 = \bar{x}, \quad \forall n \in \mathbb{Z}_+ \tag{23}$$

$$h_{n+1} = \theta_n h_n, \quad h_0 = 1, \quad \forall n \in \mathbb{Z}_+. \tag{24}$$

(ii) The sequence of functions $\Lambda_k : X \mapsto \mathbb{R}_+$ defined by

$$\begin{aligned} \Lambda_0(x) &\equiv 0 \\ \Lambda_{k+1}(x) &= \max\left\{0,\ \sigma(x) + \min_{r\in\Omega} \max_{u\in U} \Lambda_k(f(x,r,u))\right\} \end{aligned} \tag{25}$$

converges pointwise to a limit $\Lambda_\infty : X \mapsto \mathbb{R}_+$.

(iii) There exists a function $V : X \mapsto \mathbb{R}_+$ such that

$$V(x) = \max\left\{0,\ \sigma(x) + \min_{r\in\Omega} \max_{u\in U} V(f(x,r,u))\right\} \tag{26}$$

for every $x \in X$.

(iv) There exists a function $V : X \mapsto \mathbb{R}_+$ such that

$$V(x) \geq \sigma(x) + \min_{r\in\Omega} \max_{u\in U} V(f(x,r,u)), \quad \forall x \in X. \tag{27}$$

Moreover, if conditions (i)$-$(iv) hold, then V_∞ is a solution of (26) and

$$V_\infty = \Lambda_\infty \geq V_k = \Lambda_k, \quad \forall k \in \mathbb{Z}_+ \tag{28}$$

$$V \geq V_\infty. \tag{29}$$

for V satisfying (iii). Furthermore, for every x_n satisfying (23),

$$\sup_\tau \sum_{n=0}^{\tau-1} \sigma(x_n) < \infty. \tag{30}$$

Proof: Please see the Appendix. ■

Definition 1: For $v \in \mathbb{R}^m \backslash \{0\}$, a *cylinder with axis* v is a set of the form:

$$\mathcal{C}_{Q,\beta}(v) = \left\{ p \in \mathbb{R}^m : \inf_{t \in \mathbb{R}} (p - tv)^T Q (p - tv) \leq \beta \right\} \tag{31}$$

where $Q \in \mathbb{R}^{m \times m}$, $Q = Q' > 0$, and $\beta > 0$.

Remark 1: A cylinder that is an invariant set for system (7) is called an *invariant cylinder.*

The following theorem establishes that the value function is zero for all x outside a certain bounded region.

Theorem 2: Let $U \subset \mathbb{R}$ be a fixed finite set. Consider the system defined by equation (7), where $x \in \ell_+(\mathbb{R}^m),\ r \in \ell_+([-1,1]),\ u \in \ell_+(U)$, and the pair (A, B) is controllable. Suppose that A has exactly one eigenvalue on the unit circle. Let e_1 denote the eigenvector corresponding to the eigenvalue of A that is on the unit circle. Let $\beta > 0$ and $Q \in \mathbb{R}^{m \times m}$, $Q = Q' > 0$ be such that $\mathcal{C}_{Q,\beta}(e_1)$ is an invariant cylinder for system (7). Let V be defined by (21) and σ be BIBO. If the set

$$S_0 = \{ x \in \mathcal{C}_{Q,\beta}(e_1) : \sigma(x) > -\epsilon_0 \} \tag{32}$$

is bounded for some $\epsilon_0 > 0$, then the set

$$M = \{ x \in \mathcal{C}_{Q,\beta}(e_1) : V(x) \neq 0 \} \tag{33}$$

is also bounded.

Proof: Please see the Appendix. ■

V. NUMERICAL EXAMPLE

Consider the example in [18], were the dynamical system L_G (4) has transfer function

$$H(z) = \frac{z+1}{z(z-1)}.$$

Let $U = \{-1.5, 0, 1.5\}$, $\phi(x) = |Cx|$, and $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$. From [18], the strong invariant set $\mathcal{I}$ is given by

$$\mathcal{I} = \{x \in \mathbb{R}^2 \ : \ |x_1 - x_2| \leq 2.5\}. \tag{34}$$

Due to the pole at $z = 1$, the strong invariant set $\mathcal{I}$ given by (34) is unbounded and defines an infinite strip in $\mathbb{R}^2$. However, according to Theorem 2 we need to search for $V(x)$ only inside a bounded region within this infinite strip, since $V(x) = 0$ for all x outside a certain bounded region. The bounded region is found via trial and error. We select a grid spacing of $\Delta = 1/64$. Following the procedures outlined in subsections III-B and III-C, the largest γ for which the iteration in (19) converges to the limit Λ in (20), is $\gamma = 0.925$, which is a certified lower bound on the performance of any arbitrary ADC with respect to the specific performance measure selected. Figures 5, 6a, and 6b show the value function V, the cross section of V, and the zero level set of V, respectively. Figures 7a and 7b show the control function and its cross section, respectively. The certified upper bound for the performance of the ADC designed in [18] with respect to the same performance criteria is 1.1875.

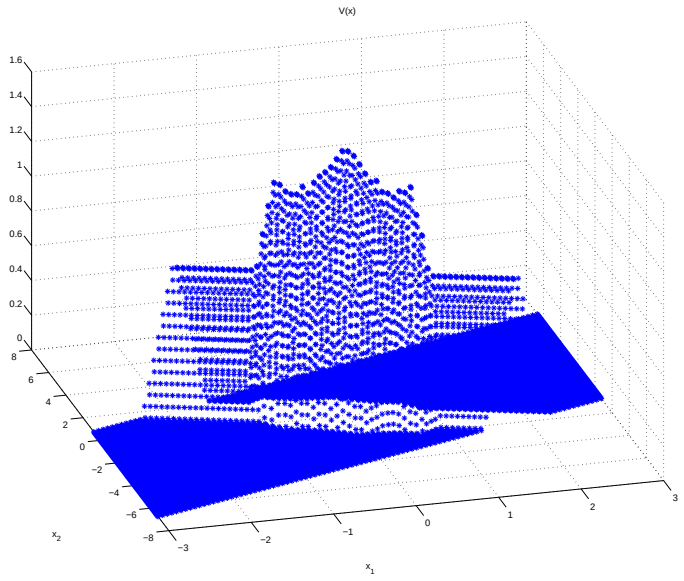
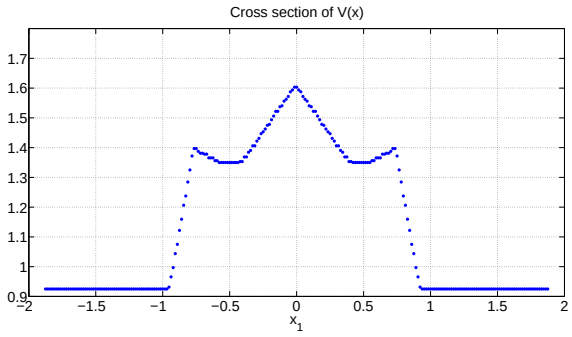
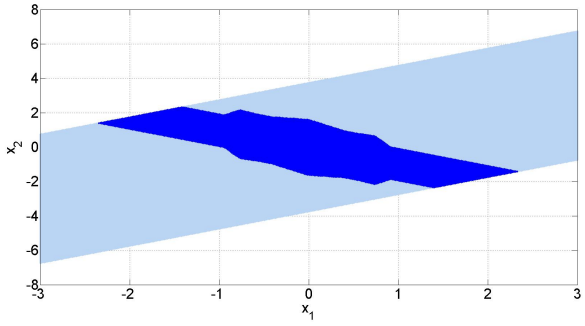


Fig. 5: Value Function $V(x)$ for Lower Bound

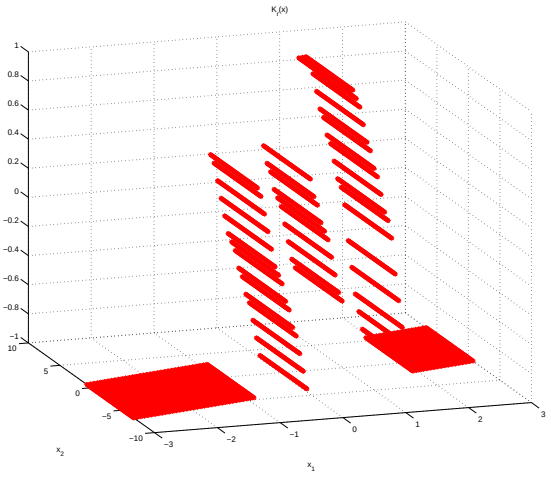


(a) Cross Section of $V(x)$ along $x_1 = -x_2$

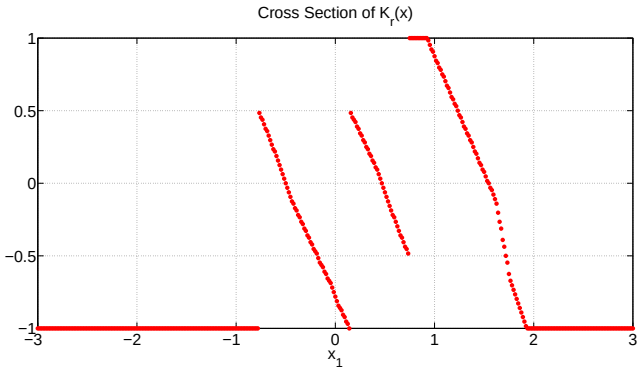


(b) Zero Level Set of $V(x)$

Fig. 6: Value Function $V(x)$



(a) Control $K_r(x)$ for Lower Bound



(b) Cross Section of Control $K_r(x)$ along $x_1 = x_2$

Fig. 7: Control Function $K_r(x)$

VI. CONCLUSION

In this paper, we studied performance limitations of Analog to Digital Converters (ADCs). The performance of an ADC was defined in terms of a measure that represents the worst case average intensity of the filtered input matching error. The passband of the shaping filter defines the frequency region in which the error is to be minimized. The problem of finding a lower bound for the performance of an ADC was associated with a full information feedback optimal control problem and formulated as a dynamic game in which the input of the ADC (control variable) played against the output of the ADC (quantized disturbance). Since the disturbances can exceed the control variable in magnitude, if

the shaping filter has a pole on the unit circle, then there does not exist a bounded control invariant set, which presents a challenge for numerical computations. This challenge is overcome with theoretical results that show that the value function is zero beyond a bounded region, thus computations need to be done only over this bounded region. A numerical algorithm was presented that provided certified solutions to the underlying Bellman inequality in parallel with the control law; hence, certified lower bounds on the performance of arbitrary ADCs with respect to the adopted performance criteria.

VII. APPENDIX

Observation 1: The sequence of functions $\Lambda_k : X \mapsto \mathbb{R}_+$ defined by (25) is monotonically increasing.

Proof: The proof is done by induction. Since $\Lambda_0(x) = 0$ for all $x \in X$, it follows that

$$\Lambda_1(x) = \max\{0,\ \sigma(x)\} \geq \Lambda_0(x), \qquad \forall x \in X. \tag{35}$$

Assume, $\Lambda_k(x) \geq \Lambda_{k-1}(x)$ for all $x \in X$. This assumption in conjunction with equation (25), results in the following inequality

$$\begin{aligned}\Lambda_{k+1}(x) &\geq \max\left\{0,\ \sigma(x) + \min_{r\in\Omega}\max_{u\in U}\Lambda_{k-1}(f(x,r,u))\right\}\\ &= \Lambda_k(x).\end{aligned}$$

Therefore,

$$\Lambda_0(x) \leq \Lambda_1(x) \leq \cdots \leq \Lambda_k(x), \quad \forall x \in X,\ k \in \mathbb{Z}_+$$

■

Proof of Theorem 1: (i) $\Longrightarrow$ (ii) For $\tau = 0$, equation (22) simplifies to:

$$V_0(x_0) = 0, \quad \forall x_0 \in X. \tag{36}$$

For $\tau = 1$, we have:

$$V_1(x_0) = \max\{0, \sigma(x_0)\}.$$

The rest of the proof is done by induction over τ. For $\tau = 2$, we have:

$$\begin{aligned}V_2(x_0) &= \max\{0, \sigma(x_0) + \min_{r_0}\max_{u_0,\theta_1}\theta_1\sigma(x_1)\}\\ &= \max\{0, \sigma(x_0) + \min_{r_0}\max_{u_0} V_1(f(x_0, r_0, u_0))\}.\end{aligned}$$

Assume that,

$$V_k(x_0) = \max\{0, \sigma(x_0) + \min_{r_0}\max_{u_0} V_{k-1}(f(x_0, r_0, u_0))\}.$$

Define $\widetilde{h}_{n+1} = \theta_n \widetilde{h}_n$, $\widetilde{h}_1 = 1$, for $n = 1, 2, 3 \cdots$. Equation (22) can be equivalently written for $\tau = k+1$ as follows:

$$V_{k+1}(x_0) = \max_{\theta_0} \left[\theta_0 \Big(\sigma(x_0) + \min_{r_0} \max_{u_0} \underbrace{\max_{\theta_1} \cdots \min_{r_{k-1}} \max_{u_{k-1}, \theta_k} \sum_{n=1}^{k} \widetilde{h}_{n+1} \sigma(x_n)}_{V_k(x_1)} \Big) \right].$$

Therefore,

$$V_{k+1}(x_0) = \max\{0, \sigma(x_0) + \min_{r_0} \max_{u_0} V_k(f(x_0, r_0, u_0))\}. \tag{37}$$

From Observation 1, we know that the sequence of functions V_k is monotonically increasing. Since a monotonic sequence of functions converge if and only if it is bounded, we have convergence of the sequence.

(ii) $\Longrightarrow$ (i) Again from Observation 1, the sequence of functions V_k is monotonically increasing; thus, in conjunction with convergence of the sequence, we have $\Lambda_\infty < \infty$. Furthermore, since $\Lambda_k(x) \geq 0$ for all $k \in \mathbb{Z}_+$, we also have $\Lambda_\infty(x) \geq 0$. It only remains to show that equation (25) is equivalent to (22). Equation (25) for $k = 1$ is trivially equivalent to:

$$\Lambda_1(x_0) = \max_{\theta_0 \in \{0,1\}} \theta_0 \sigma(x_0), \quad \forall x_0 \in X$$

For $k = 2$, equation (25) is equivalent to:

$$\begin{aligned}
\Lambda_2(x_0) &= \max \left\{ 0, \ \sigma_0(x_0) + \min_{r_0} \max_{u_0} \Lambda_1(x_1) \right\} \\
&= \max_{\theta_0} \left[\theta_0 \left(\sigma_0(x_0) + \min_{r_0} \max_{u_0, \theta_1} \{\theta_1 \sigma(x_1)\} \right) \right] \\
&= \max_{\theta_0} \min_{r_0} \max_{u_0, \theta_1} \sum_{n=0}^{1} h_{n+1} \sigma(x_n).
\end{aligned}$$

Assume,

$$\Lambda_k(x_0) = \max_{\theta_0} \min_{r_0} \max_{u_0, \theta_1} \cdots \min_{r_{k-2}} \max_{u_{k-2}, \theta_{k-1}} \sum_{n=0}^{k-1} h_{n+1} \sigma(x_n).$$

Substituting the above equation into (25) we have:

$$\Lambda_{k+1}(x_0) = \max_{\theta_0} \left[\theta_0 \Big(\sigma(x_0) + \min_{r_0} \max_{u_0} \underbrace{\max_{\theta_1} \cdots \min_{r_{k-1}} \max_{u_{k-1}, \theta_k} \sum_{n=1}^{k} \widetilde{h}_{n+1} \sigma(x_n)}_{\Lambda_k(x_1)} \Big) \right],$$

which is equivalent to (22). Finally, since the sequence of functions Λ_k is monotonically increasing, the limit as $k \to \infty$ of Λ_k is equivalent to its supremum over k.

(i) $\Longrightarrow$ (iii) Substituting (37) into (21) and interchanging the order of the supremum over τ with the maximum, we get

$$V_\infty(\bar{x}) = \max\{0, \sigma(x_0) + \sup_{\tau\in\mathbb{Z}_+} \min_{r_0} \max_{u_0} V_{\tau-1}(f(x_0, r_0, u_0))\}.$$

As discussed in the proof of (ii) $\Longrightarrow$ (i), the supremum over τ, in the expression above, is equal to the limit as $\tau \to \infty$. Moreover, a well-known theorem from Analysis states that: given a metric space X, a compact metric space Ω, and a continuous function $g : X \times \Omega \mapsto \mathbb{R}$, the function $\hat{g} : X \mapsto \mathbb{R}$ defined by

$$\hat{g}(x) = \max_{r\in\Omega} g(x, r), \quad \text{or} \quad \hat{g}(x) = \min_{r\in\Omega} g(x, r)$$

is continuous. Furthermore, given a compact metric space Ω, and a monotonically increasing sequence of continuous functions $g_k : \Omega \mapsto \mathbb{R}$ such that $\lim_{k\to\infty} g_k(r)$ is finite for every $r \in \Omega$, the following equality holds:

$$\lim_{k\to\infty} \min_{r\in\Omega} g_k(r) = \min_{r\in\Omega} \lim_{k\to\infty} g_k(r)$$

Therefore, we have (26).

(iii) $\Longrightarrow$ (iv) Trivially true.

(iv) $\Longrightarrow$ (i) Since $V(x) \geq 0$ for all x, we can rewrite (27) as

$$V(x) \geq \max\left\{0,\ \sigma(x) + \min_{r} \max_{u} V(f(x, r, u))\right\}. \tag{38}$$

Inequality (38) holds for all $x \in X$, thus it holds for the sequence $\{x_0, x_1, \cdots, x_{k-1}\}$, where x_{k-1} satisfies (23). Now take inequality (38) with x replaced by x_0 and substitute for $V(f(x_0, r_0, u_0))$ the corresponding inequality for $V(x_1)$:

$$V(x_0) \geq \max\left\{0,\ \sigma(x_0) + \min_{r_0} \max_{u_0} \max\left[0,\ \sigma(x_1) + \min_{r_1} \max_{u_1} V(f(x_1, r_1, u_1))\right]\right\}.$$

Equivalently,

$$V(x_0) \geq \max_{\theta_0} \theta_0 \left(\sigma(x_0) + \min_{r_0} \max_{u_0,\theta_1} \theta_1 \left[\sigma(x_1) + \min_{r_1} \max_{u_1} V(f(x_1, r_1, u_1))\right]\right).$$

Repeating this process, we have:

$$V(x_0) \geq \overbrace{\max_{\theta_0} \min_{r_0} \max_{u_0,\theta_1} \cdots \min_{r_{k-2}} \max_{u_{k-2},\theta_{k-1}}}^{S(x_0)} \left[\sum_{n=0}^{k-1} h_{n+1}\sigma(x_n) + \min_{r_{k-1}} \max_{u_{k-1}} h_n V(f(x_{k-1}, r_{k-1}, u_{k-1}))\right].$$

After rearranging terms we have:

$$S(x_0) \leq V(x_0) - \max_{\theta_0} \min_{r_0} \max_{u_0,\theta_1} \cdots \min_{r_{k-1}} \max_{u_{k-1}} h_n V(x_k).$$

Non-negativity and existence of V guarantees:

$$S(x_0) \leq V(x_0) < \infty. \tag{39}$$

Since (39) holds for all k, we have (21).

The proof for (28)-(30) is as follows: Substituting (21) into the right hand side of (26) and using the reasoning in (i) $\implies$ (iii) it is easy to see that (21) is a solution of (26). Furthermore, (28) was proved within the proof of (i) $\implies$ (ii). Inequality (29) states that (21) is the minimal solution of (26), this is proven by induction. Let F be a function that maps function V on X into function FV on X, defined according to:

$$(FV)(x) = \max\left\{0,\ \sigma(x) + \min_{r\in\Omega}\max_{u\in U} V(f(x,r,u))\right\},$$

then $V = FV$. Since $V \geq 0$, we have $V \geq V_0$. Assume $V \geq V_k$, and apply mapping F to both sides to get

$$FV \geq FV_k = V_{k+1}.$$

Therefore, $V \geq V_k$ for all k, and thus (21) is the minimal solution of (26). Finally, (30) is obtained by substituting into (21) the argument of minimums and maximums for the sequences r, u, and θ respectively.

■

The proof of Theorem 2 relies on Lemmas 1 and 2 below.

Lemma 1: Let (A, B) be a controllable pair. Then, for every bounded set $\Xi \subset \mathbb{R}^m$, there exists a finite set $\widetilde{X} \subset \mathbb{R}^m$ and a function $\rho : \Xi \mapsto [-1, 1]^m$, such that $x_m \in \widetilde{X}$ whenever $x_0 \in \Xi$ for every solution (x, r) of

$$x_{n+1} = Ax_n + Br_n - Bu_n, \quad n \leq m \tag{40}$$

$$r_n = \rho(x_0)_n, \quad n \leq m \tag{41}$$

for every $u \in \ell_+(U)$, where $\rho(x_0)_n$ denotes the n-th element of $\rho(x_0)$.

Proof: The solution to (40) is given by

$$x_m = A^m x_0 + \sum_{i=0}^{m-1} A^i B r_{m-i-1} - \sum_{i=0}^{m-1} A^i B u_{m-i-1}. \tag{42}$$

Since Ξ is bounded, $A^m\Xi$ is also bounded, thus the first term on the right hand side of (42) is bounded. Let L_c denote the controllability matrix:

$$L_c = [A^{m-1}B \cdots AB \quad B].$$

Construct a finite set $\widetilde{\Xi}_F \subset \widetilde{\Xi}$ as follows: let $\widetilde{\Xi}_F$ be the intersection of $\widetilde{\Xi}$ and the set consisting of uniformly spaced Cartesian grid points with spacing Δ, where

$$\Delta \leq 2/||L_c^{-1}||_\infty.$$

Then for every $y_0 \in \widetilde{\Xi}$, there exists $\tilde{\xi} \in \widetilde{\Xi}_F$ such that:

$$||y_0 - \tilde{\xi}||_\infty \leq \Delta/2$$

Thus,

$$||L_c^{-1}(y_0 - \tilde{\xi})||_\infty \leq 1,$$

which implies

$$L_c^{-1}(A^m x_0 - \tilde{\xi}) \in [-1, 1]^m, \quad \forall x_0 \in \Xi \tag{43}$$

Then for

$$\rho(x_0) = -L_c^{-1}(A^m x_0 - \tilde{\xi}), \tag{44}$$

we have

$$x_m = \tilde{\xi} - \sum_{i=0}^{m-1} A^i B u_{m-i-1}.$$

Since $\widetilde{\Xi}_F$ and U are finite sets, x_m takes only a finite number of values. ■

Lemma 2: Assume (A, B) is controllable and the function $\sigma : \mathbb{R}^m \mapsto \mathbb{R}$ is BIBO. If V_∞ in (21)-(22) satisfies

$$V_\infty(x) < \infty, \quad \forall x \in \mathbb{R}^m \tag{45}$$

then V_∞ is BIBO.

Proof: Let $\alpha : \mathbb{R}^m \times \ell_+([-1, 1]) \times \ell_+(U) \times \mathbb{Z}_+ \mapsto \mathbb{R}^m$ be a function that maps the initial state x_0 and sequences r and u to the state x at time k, where the evolution of the state is given by $x[n + 1] = Ax[n] + Br[n] - Bu[n]$:

$$\alpha(x_0, r, u, k) = x_k. \tag{46}$$

Equation (22) can be equivalently written as:

$$V_\tau(x_0) = \max_{\theta_0} \min_{r_0} \max_{u_0,\theta_1} \cdots \min_{r_{m-1}} \max_{u_{m-1}} \left\{ \sum_{n=0}^{m-1} h_{n+1}\sigma(x_n) + \max_{\theta_m} \min_{r_m} \max_{u_m,\theta_{m+1}} \cdots \min_{r_{\tau-2}} \max_{u_{\tau-2},\theta_{\tau-1}} \sum_{n=m}^{\tau-1} h_{n+1}\sigma(x_n) \right\} \tag{47}$$

Denote,

$$\widehat{r} = \{r_i\}_{i=0}^{m-1}, \quad \widehat{u} = \{u_i\}_{i=0}^{m-1}, \quad \widehat{\theta} = \{\theta_i\}_{i=0}^{m-1}.$$

We can rewrite (47) as:

$$V_\tau(x_0) = \max_{\theta_0} \min_{r_0} \max_{u_0,\theta_1} \cdots \min_{r_{m-1}} \max_{u_{m-1}} \left\{ \sum_{n=0}^{m-1} h_{n+1}\sigma(x_n) + V_\tau(\alpha(x_0, \widehat{r}, \widehat{u}, m)) \right\} \tag{48}$$

Let Ξ be a bounded subset of $\mathbb{R}^m$ and $x_0 \in \Xi$. Furthermore, let $\rho : \Xi \mapsto [-1,1]^m$ be the function defined in (44) and $\widetilde{X} \subset \mathbb{R}^m$ denote the set of all states that can be reached in exactly m steps for some input sequence $u \in \ell_+(U)$. According to Lemma 1, the set $\widetilde{X}$ is finite. Let $\widehat{r} = \rho(x_0)$, consequently $\alpha(x_0, \rho(x_0), \widehat{u}, m) = x_m \in \widetilde{X}$ for every sequence $\widehat{u}$. Denote $\bar{r} = \{r_i\}_{i=0}^{m-2}$, $\bar{u} = \{u_i\}_{i=0}^{m-2}$, and

$$\bar{\nu}\left(x_0, \bar{r}, \bar{u}, \widehat{\theta}\right) \stackrel{\text{def}}{=} \sum_{n=0}^{m-1} h_{n+1}\sigma(x_n).$$

Let $\bar{r} = \{\rho(x_0)_i\}_{i=0}^{m-2}$ and denote,

$$\nu\left(x_0, \bar{u}, \widehat{\theta}\right) \stackrel{\text{def}}{=} \bar{\nu}\left(x_0, \bar{r}, \bar{u}, \widehat{\theta}\right)\Big|_{\bar{r}=\{\rho(x_0)_i\}_{i=0}^{m-2}}.$$

Thus,

$$V_\tau(x_0) \leq \max_{\widehat{u},\widehat{\theta}} \left\{ \nu\left(x_0, \bar{u}, \widehat{\theta}\right) + V_\tau(x_m) \right\},$$

where $x_m \in \widetilde{X}$ for every sequence $\widehat{u}$. Taking supremum over τ from both sides of the above inequality, we have:

$$\begin{aligned} V_\infty(x_0) &\leq \sup_\tau \max_{\widehat{u},\widehat{\theta}} \left\{ \nu\left(x_0, \bar{u}, \widehat{\theta}\right) + V_\tau(x_m) \right\}, \\ &= \max_{\widehat{u},\widehat{\theta}} \left\{ \nu\left(x_0, \bar{u}, \widehat{\theta}\right) + V_\infty(x_m) \right\}. \end{aligned}$$

Hence,

$$\sup_{x_0\in\Xi} V_\infty(x_0) \leq \sup_{x_0\in\Xi} \max_{\bar{u},\widehat{\theta}} \nu\left(x_0, \bar{u}, \widehat{\theta}\right) + \sup_{x_0\in\Xi} \max_{\widehat{u},\widehat{\theta}} V_\infty(x_m). \tag{49}$$

Since Ξ and U are bounded, the set

$$\left\{ A^n x + \sum_{k=-1}^{n-2} A^{k+1} B(r_{n-k} - u_{n-k}) : x \in \Xi, r_k \in [-1,1], u_k \in U \right\}$$

is also bounded for every finite n. Furthermore, σ is BIBO, which immediately implies that the first supremum on the right side of inequality (49) is bounded. Moreover, x_m can take only a finite number of values for every $x_0 \in \Xi$ and every sequence $\widehat{u}$. Since V_∞ is finite and the supremum over a finite set is finite, the second term on the right side of inequality (49) is also bounded. Hence V_∞ is BIBO. ■

Proof of Theorem 2: For system (7) with exactly one pole z_1 on the unit circle and all other poles strictly inside the unit circle, with $e_1 \in \mathbb{R}^m \backslash \{0\}$ such that $Ae_1 = z_1 e_1$, and $Q \in \mathbb{R}^{m\times m}$, $Q = Q' > 0$ such that $Q \geq A'QA$, there exists an invariant cylinder with axis e_1 for some $\beta > 0$. Furthermore, the intersection of $\mathcal{C}_{Q,\beta}(e_1)$ with the set $\{|e_1 x| < \zeta : x \in \mathbb{R}^m, \zeta > 0\}$ is bounded whenever $Ce_1 \neq 0$. Define

$$M_0 = \sup_{x\in S_0} V_\infty(x).$$

Lemma 2 guarantees finiteness of the supremum. Let $\alpha : \mathbb{R}^m \times \ell_+([-1,1]) \times \ell_+(U) \times \mathbb{Z}_+ \mapsto \mathbb{R}^m$ be a function defined in the proof of Lemma 2. Let L denote the smallest integer strictly larger than M_0/ϵ_0. Define

$$S_L = \{x \in \mathcal{C}_{Q,\beta}(e_1) : \alpha(x, r, u, k) \notin S_0,\ \forall k \leq L,\ \forall r \in \ell_+([-1,1]),\ \forall u \in \ell_+(U)\}$$

That is, S_L is the set of all states within the invariant cylinder from which S_0 cannot be reached in L steps or less. The complement of the set S_L is the region of the cylinder for which there exist sequences r and u such that the state gets to S_0 in L steps or less. Since, both r and u are uniformly bounded, the set S_L^c is bounded.

For $j = \{0, 1, 2, \cdots, \tau - 2\}$, define functions $g_j : \mathbb{R}^m \times \ell_+(\{0,1\}) \times \ell_+([-1,1]) \times \ell_+(U) \mapsto \mathbb{R}$,

$$g_j(\bar{x}, \{\theta_i\}_{i=0}^{j}, \{r_i\}_{i=0}^{j-1}, \{u_i\}_{i=0}^{j-1}) = \min_{r_j} \max_{u_j,\theta_{j+1}} \cdots \min_{r_{\tau-2}} \max_{u_{\tau-2},\theta_{\tau-1}} \sum_{n=0}^{\tau-1} h_{n+1}\sigma(x_n) \tag{50}$$

$$g_{\tau-1}(\bar{x}, \{\theta_i\}_{i=0}^{\tau-1}, \{r_i\}_{i=0}^{\tau-2}, \{u_i\}_{i=0}^{\tau-2}) = \sum_{n=0}^{\tau-1} h_{n+1}\sigma(x_n) \tag{51}$$

For $\epsilon > 0$ and $\bar{x} \in \mathbb{R}^m$, let $\widetilde{\tau}(\bar{x}) \in \mathbb{Z}_+$, $\widetilde{\Theta}_0(\bar{x}) \in \{0,1\}$, $\widetilde{R}_0(\bar{x}) \in [-1,1]$, $\widetilde{\mathcal{U}}_0(\bar{x}) \in U$, $\widetilde{\Theta}_1(\bar{x}) \in \{0,1\}$ be functions such that:

$$\widetilde{\tau}(\bar{x}) \in \arg^{\epsilon} \sup_{\tau\in\mathbb{Z}_+} V_\tau(\bar{x}), \tag{52}$$

$$\widetilde{\Theta}_0(\bar{x}) = \theta_0 \in \arg\max_{\theta_0} g_0(\bar{x}, \theta_0), \tag{53}$$

$$\widetilde{R}_0(\bar{x}) = r_0 \in \arg\min_{r_0} \max_{u_0,\theta_1} g_1\left(\bar{x}, \left(\widetilde{\Theta}_0(\bar{x}), \theta_1\right), r_0, u_0\right), \tag{54}$$

$$\left(\widetilde{\mathcal{U}}_0(\bar{x}), \widetilde{\Theta}_1(\bar{x})\right) = (u_0, \theta_1) \in \arg\max_{u_0,\theta_1} g_1\left(\bar{x}, \left(\widetilde{\Theta}_0(\bar{x}), \theta_1\right), \widetilde{R}_0(\bar{x}), u_0\right). \tag{55}$$

For $\bar{x} \in \mathbb{R}^m$ and $j = \{0, 1, 2, \cdots, \widetilde{\tau}(\bar{x}) - 2\}$, let $\widetilde{\Theta} \in \{0,1\}^{j+2}$, $\widetilde{R} \in [-1,1]^{j+1}$, $\widetilde{\mathcal{U}} \in U^{j+1}$, $\widetilde{R}_j(\bar{x}) \in [-1,1]$, $\widetilde{\mathcal{U}}_j(\bar{x}) \in U$, and $\widetilde{\Theta}_{j+1}(\bar{x}) \in \{0,1\}$ be functions such that:

$$\widetilde{\Theta} = \left(\{\widetilde{\Theta}_i(\bar{x})\}_{i=0}^{j}, \theta_{j+1}\right), \quad \widetilde{R} = \left(\{\widetilde{R}_i(\bar{x})\}_{i=0}^{j-1}, r_j\right), \quad \widetilde{\mathcal{U}} = \left(\{\widetilde{\mathcal{U}}_i(\bar{x})\}_{i=0}^{j-1}, u_j\right). \tag{56}$$

$$\widetilde{R}_j(\bar{x}) = r_j \in \arg\min_{r_j} \max_{u_j, \theta_{j+1}} g_{j+1}\left(\bar{x}, \widetilde{\Theta}, \widetilde{R}, \widetilde{U}\right), \tag{57}$$

$$\left(\widetilde{\mathcal{U}}_j(\bar{x}), \widetilde{\Theta}_{j+1}(\bar{x})\right) = (u_j, \theta_{j+1}) \in \arg\max_{u_j, \theta_{j+1}} g_{j+1}\left(\bar{x}, \widetilde{\Theta}, \{\widetilde{R}_i(\bar{x})\}_{i=0}^{j}, \widetilde{\mathcal{U}}\right). \tag{58}$$

Assuming that $\bar{x} \in S_L$, there are two cases to consider, either:

1) $\alpha(\bar{x}, \{\widetilde{R}_i(\bar{x})\}_{i=0}^{J-1}, \{\widetilde{\mathcal{U}}_i(\bar{x})\}_{i=0}^{J-1}, J) \notin S_0$, for all $J \in \{0, 1, 2, \cdots \widetilde{\tau} - 1\}$.
2) There exists an integer $J > L$ such that $\alpha(\bar{x}, \{\widetilde{R}_i(\bar{x})\}_{i=0}^{J-1}, \{\widetilde{\mathcal{U}}_i(\bar{x})\}_{i=0}^{J-1}, J) \in S_0$.

From equations (50)−(58), we have:

$$g_0(\bar{x}, 0) = 0, \tag{59}$$

$$g_0(\bar{x}, 1) = \min_{\{r_i\}_{i=0}^{\widetilde{\tau}(\bar{x})-2}} \sum_{n=0}^{\widetilde{\tau}(\bar{x})-1} h_{n+1}\sigma(x_n) \tag{60}$$

$$\leq \sum_{n=0}^{\widetilde{\tau}(\bar{x})-1} h_{n+1}\sigma(x_n)\Big|_{\{r_i\}_{i=0}^{\widetilde{\tau}(\bar{x})-2} = \{\widetilde{R}_i(\bar{x})\}_{i=0}^{\widetilde{\tau}(\bar{x})-2}} \tag{61}$$

where $\{u_i\}_{i=0}^{\widetilde{\tau}(\bar{x})-2} = \{\widetilde{\mathcal{U}}_i(\bar{x})\}_{i=0}^{\widetilde{\tau}(\bar{x})-2}$ and $\{\theta_i\}_{i=1}^{\widetilde{\tau}(\bar{x})-1} = \{\widetilde{\Theta}_i(\bar{x})\}_{i=1}^{\widetilde{\tau}(\bar{x})-1}$. Furthermore, by (21) and (52), we have:

$$V_\infty(\bar{x}) < \epsilon + V_{\widetilde{\tau}(\bar{x})}(\bar{x}) = \epsilon + \max_{\theta_0} g_0(\bar{x}, \theta_0) \tag{62}$$

$$= \epsilon + \max\left\{g_0(\bar{x}, 0), g_0(\bar{x}, 1)\right\}. \tag{63}$$

Consider case (1): since $\sigma(x) \leq -\epsilon_0$ for all $x \notin S_0$, the sum over $h_{n+1}\sigma(x_n)$ will be negative for all $\widetilde{\tau}(\bar{x})$. Thus,

$$V_\infty(\bar{x}) < \epsilon, \quad \forall \bar{x} \in S_L. \tag{64}$$

For case (2), we can write $g_0(\bar{x}, 1)$ equivalently as:

$$g_0(\bar{x}, 1) = \min_{r_0} \max_{u_0, \theta_1} \cdots \min_{r_{J-2}} \max_{u_{J-2}, \theta_{J-1}} \left\{ \sum_{n=0}^{J-1} h_{n+1}\sigma(x_n) + \min_{r_{J-1}} \max_{u_{J-1}, \theta_J} \cdots \min_{r_{\widetilde{\tau}(\bar{x})-2}} \max_{u_{\widetilde{\tau}(\bar{x})-2}, \theta_{\widetilde{\tau}(\bar{x})-1}} \sum_{n=J}^{\widetilde{\tau}(\bar{x})-1} h_{n+1}\sigma(x_n) \right\} \tag{65}$$

Since the first summation term in (65) is bounded above by $-J\epsilon_0$ for $\{\widetilde{\Theta}_i(\bar{x})\}_{i=0}^{J-1}$ and all sequences $\{r_i\}_{i=0}^{J-2}$ and $\{u_i\}_{i=0}^{J-2}$, and the second summation term is equal to $g_0(\alpha(\bar{x}, \{r_i\}_{i=0}^{J-1}, \{u_i\}_{i=0}^{J-1}, J), \theta_J)$, we have:

$$g_0(\bar{x}, 1) \leq \min_{r_0} \max_{u_0, \theta_1} \cdots \min_{r_{J-2}} \max_{u_{J-2}, \theta_{J-1}} \left\{ -J\epsilon_0 + g_0(\alpha(\bar{x}, \{r_i\}_{i=0}^{J-1}, \{u_i\}_{i=0}^{J-1}, J), \theta_J) \right\}. \tag{66}$$

Since

$$g_0(\alpha(\bar{x}, \{\widetilde{R}_i(\bar{x})\}_{i=0}^{J-1}, \{\widetilde{\mathcal{U}}_i(\bar{x})\}_{i=0}^{J-1}, J), \widetilde{\Theta}_J(\bar{x})) \leq M_0, \tag{67}$$

we have:

$$g_0(\bar{x}, 1) \leq M_0 - J\epsilon_0. \tag{68}$$

Furthermore, $J > M_0/\epsilon_0$, thus,

$$V_\infty(\bar{x}) < \epsilon + \max\{0,\ M_0 - J\epsilon_0\} = \epsilon. \tag{69}$$

Since, $V_\infty(\bar{x}) < \epsilon$ for every $\epsilon > 0$,

$$V_\infty(\bar{x}) = 0, \quad \forall \bar{x} \in S_L. \tag{70}$$

Finally, the complement of the set S_L is bounded; therefore, (33) is bounded.

REFERENCES

[1] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Prentice-Hall, 1999.

[2] M. Derpich, E. Silva, D. Quevedo, and G. Goodwin, "On optimal perfect reconstruction feedback quantizers," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3871 –3890, Aug 2008.

[3] S. Ardalan and J. Paulos, "An analysis of nonlinear behavior in delta - sigma modulators," *IEEE Transactions on Circuits and Systems*, vol. 34, no. 6, pp. 593 – 603, jun 1987.

[4] A. Marques, V. Peluso, M. S. Steyaert, and . W. M. Sansen, "Optimal parameters for $\Delta\Sigma$ modulator topologies," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 9, pp. 1232–1241, Sep. 1998.

[5] S. Norsworthy, R. Schreier, and G. C. Temes, *Delta-Sigma Data Converters: Theory, Design, and Simulation*. IEEE Press, John Wiley and Sons, Inc, 1997.

[6] N. T. Thao and M. Vetterli, "A deterministic analysis of oversampled A/D conversion and $\Sigma\Delta$ modulation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 468–471, Apr. 1993.

[7] ——, "Deterministic Analysis of Oversampled A/D Conversion and Decoding Improvement Based on Consistent Estimates," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 519–531, Mar. 1994.

[8] N. T. Thao, "The Tiling Phenomenon in $\Sigma\Delta$ Modulation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 7, pp. 1365 – 1378, Jul. 2004.

[9] D. Quevedo and G. Goodwin, "Multistep optimal analog-to-digital conversion," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 3, pp. 503 – 515, march 2005.

[10] P. Steiner and W. Yang, "A framework for analysis of high-order sigma-delta modulators," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 44, no. 1, pp. 1 –10, jan 1997.

[11] H. Wang, "A geometric view of sigma; delta; modulations," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 39, no. 6, pp. 402 –405, jun 1992.

[12] Y. Wang and S. Boyd, "Performance bounds and suboptimal policies for linear stochastic control via LMIs," *International Journal of Robust and Nonlinear Control*, vol. 21, no. 14, pp. 1710–1728, 2011, available: http://dx.doi.org/10.1002/rnc.1665. [Online]. Available: http://www.stanford.edu/~boyd/papers/gen_ctrl_bnds.html

[13] ——, "Performance bounds for linear stochastic control," *Systems and Control Letters*, vol. 58, no. 3, pp. 178 – 182, 2009.

[14] ——, "Approximate dynamic programming via iterated bellman inequalities," April 2010. [Online]. Available: http://www.stanford.edu/~boyd/papers/adp_iter_bellman.html

[15] F. Bullo and D. Liberzon, "Quantized control via locational optimization," *IEEE Transactions on Automatic Control*, vol. 51, no. 1, pp. 2 – 13, jan. 2006.

[16] R. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Transactions on Automatic Control*, vol. 45, no. 7, pp. 1279 –1289, Jul. 2000.

[17] N. Elia and S. Mitter, "Stabilization of linear systems with limited information," *IEEE Transactions on Automatic Control*, vol. 46, no. 9, pp. 1384 –1400, sep 2001.

[18] M. Osqui, A. Megretski, and M. Roozbehani, "Optimality and Performance Limitations of Analog to Digital Converters," *Conference on Decision and Control*, pp. 7527–7532, Dec. 2010.

[19] J. W. Helton and M. R. James, *Extending H-infinity Control to Nonlinear Systems: Control of Nonlinear Systems to Achieve Performance Objectives*. Society for Industrial Mathematics (SIAM), 1999.

[20] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I*. Athena Scientific, 2005.

[21] ——, *Dynamic Programming and Optimal Control, Vol. II*. Athena Scientific, 2007.

[22] R. Bellman, *Dynamic Programming*. Dover Publications, 2003.

[23] A. Megretski, "Robustness of finite state automata," in *Multidisciplinary Research in Control: The Mohammed Dahleh Symposium 2002*, ser. Lecture Notes in Control and Information Sciences, L. Giarre and B. Bamieh, Eds. Springer, 2003, vol. 289, pp. 147–160.

[24] M. Osqui, M. Roozbehani, and A. Megretski, "Semidefinite Programming in Analysis and Optimization of Performance of Sigma-Delta Modulators for Low Frequencies," *American Control Conference*, pp. 3582–3587, Jul. 2007.

[25] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.

[26] T. Basar and P. Bernhard, H^∞ - *Optimal Control and related Minimax Design Problems: A Dynamic Game Approach*. Birkhauser, 1995.