

# Conditions for Optimality of Naïve Quantised Finite Horizon Control

Daniel E. Quevedo, Claus Müller and Graham C. Goodwin

(Received 1 March 2006; in revised form 21 November 2006)

This paper presents properties of a control law which quantises the unconstrained solution to a unitary horizon quadratic programme. This *naïve* quantised control law underlies many popular algorithms, such as  $\Sigma\Delta$ -converters and decision feedback equalisers, and is easily shown to be globally optimal for horizon one. However, the question arises as to whether it is also globally optimal for horizons greater than one, i.e., whether it solves a finite horizon quadratic programme, where decision variables are restricted to belong to a quantised set. By using dynamic programming, we develop sufficient conditions for this to hold. The present analysis is restricted to first order plants. However, this case already raises a number of highly non-trivial issues. The results can be applied to arbitrary horizons and quantised sets, which may contain a finite or an infinite (though countable) number of elements.

**Keywords:** finite horizon control; finite-set constraints; quantisation; optimisation

## 1 Introduction

Many interesting engineering applications involve decision variables, which are restricted to belong to a quantised set (Goodwin & Quevedo 2003). Typical examples include: on-off and multi-level control (Bockman 1991, Brockett & Liberzon 2000, Bicchi, Marigo & Piccoli 2002, Ishii & Başar 2004, Picasso & Bicchi 2003, Sznaier & Damborg 1989), power electronics (Rodríguez, Lai & Peng 2002), A/D conversion (Norsworthy, Schreier & Temes 1997, Tewksbury & Hallock 1978, Bölcskei & Hlawatsch 2001), networked control systems (Li & Baillieul 2004, Nair & Evans 2000, Ishii & Francis 2002), design of filters with quantised coefficients (Lim & Parker 1983, Evangelista 2001) and equalisation of bandlimited communication channels (Proakis 1995, Forney 1972). **In addition, quantised variables can also be regarded as an approximation to variables affected by saturation-type constraints.**

One methodology for solving these problems **with quantised decision variables** utilises concepts arising from finite horizon control and estimation (Goodwin, Serón & De Doná 2005, Maciejowski 2002, Rao, Rawlings & Lee 2001, Camacho & Bordons 1999). This approach has been used in audio quantisation (Quevedo & Goodwin 2005c, Goodwin, Quevedo & McGrath 2003), subband coding (Quevedo, Goodwin & Bölcskei 2004), networked control systems (Goodwin, Haimovich, Quevedo & Welsh 2004, Quevedo & Goodwin 2005a, Kiihtelys 2003), channel equalisation (Quevedo, Goodwin & De Doná 2003, Williamson, Kennedy & Pulford 1992), quantised coefficient filter design (Quevedo & Goodwin 2005b) and power conversion (Quevedo & Goodwin 2004a).

Implementation of the schemes described in the above references requires the solution of a quantised finite horizon optimisation problem. As shown in (Quevedo, De Doná & Goodwin 2002), the solution can be implemented by means of a vector quantiser or, equivalently, via a polytopal partition of the state space, see also (Quevedo, Goodwin & De Doná 2004, Bemporad 2003). Unfortunately, for large constraint horizons, which are preferable from a performance perspective, calculating the solution becomes computationally intractable even for relatively modest horizons due to an exponential dependence of complexity on the horizon.

On the other hand, in many applications it has been found that very good performance is achieved using rather simple schemes, such as *decision feedback equalisers* and  $\Sigma\Delta$ -converters, see, e.g., (Paulraj, Nabar & Gore 2003, Norsworthy et al. 1997, Paramesh & von Jouanne 2001, Lipschitz, Vanderkooy & Wannamaker 1991, Bölcskei & Hlawatsch 2001, Nielsen 1989). As documented in our recent work (Quevedo

---

\*School of Electrical Engineering and Computer Science, The University of Newcastle, NSW 2308, Australia; Phone: +61 2 49217378; Fax: +61 2 49601712; Emails: dquevedo@ieee.org, mull@ee.newcastle.edu.au, graham.goodwin@newcastle.edu.au

& Goodwin 2005c, Quevedo et al. 2003), it turns out that these simple methods implicitly amount to utilising a prediction horizon of one. In this case, the optimal solution is given by the *naïve* procedure of simply quantising the unconstrained solution. This is achieved using a standard scalar quantiser and incurs limited computational complexity.

The undeniable success of these reduced-complexity schemes leads to a key question: “Are there situations, in which the solution of the unitary horizon problem is actually optimal for larger horizons?” This question has been addressed for control systems, when affected by saturation in (De Doná & Goodwin 1999, De Doná, Goodwin & Serón 2000), but has not previously been addressed for quantised control, save for initial results in our conference contribution (Quevedo & Goodwin 2004b).

In the present work we develop conditions on the plant and design parameters which ensure that the naïve quantised control law is optimal for non-unitary horizons. We expand and embellish our ideas first expressed in (Quevedo & Goodwin 2004b). We focus our analysis on scalar first order plants. Unlike the unconstrained case, the treatment of quantisation leads to non-trivial issues. Indeed, first order quantised systems exhibit surprisingly rich properties and have been extensively studied in various contexts; see, e.g., (Delchamps 1990, Baillieul 2002, Fagnani & Zampieri 2003, Li & Baillieul 2004).

The remainder of this paper is organised as follows: In Section 2 we will briefly describe  $\Sigma\Delta$ -converters and decision feedback equalisers. Section 3 formulates quantised finite horizon control. In Section 4 we describe the naïve control law (NCL) and state the question posed in this paper in precise terms. Section 5 presents sufficient conditions for finite horizon optimality of the NCL given finite constraint sets. These conditions are further examined in Section 6. Section 7 extends the results to quantised constraint sets with infinite cardinality. Section 8 draws conclusions.

## 2 $\Sigma\Delta$ -Converters and Decision Feedback Equalisers

As foreshadowed in the introduction, the basic motivation for the present work lies in gaining understanding of fundamental properties of simple design solutions for systems where decision variables are quantised. We present here two especially relevant and widespread design solutions, namely  $\Sigma\Delta$ -converters and decision feedback equalisers. Both methodologies are intimately related to the naïve control law to be analysed in later sections.

### 2.1 $\Sigma\Delta$ -Conversion

$\Sigma\Delta$ -converters (and the related *noise-shaping quantisers*) are often deployed in A/D conversion (Candy & Temes 1992, Norsworthy et al. 1997), audio quantisation (Lipschitz et al. 1991), but have also found their way into subband coding (Bölcskei & Hlawatsch 2001), EMI mitigation in Switch-Mode Power Supplies (Paramesh & von Jouanne 2001), and design of FIR filters with quantised coefficients, see, e.g., (Nielsen 1989).

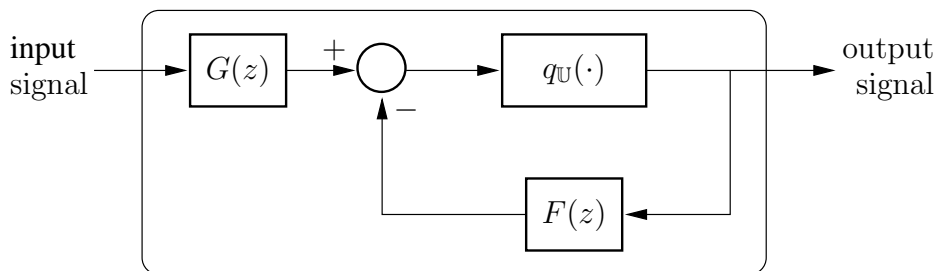


Figure 1. Quantised feedback loop.

A topology which, at least algebraically, encompasses many  $\Sigma\Delta$ -converter structures is shown in Fig. 1. As can be seen from that figure, the circuit consists of a quantiser, denoted here as  $q_U(\cdot)$ , immersed in a feedback loop. The feedback loop contains a feedforward filter,  $G(z)$ , and a feedback filter,  $F(z)$ .

The basic idea underlying  $\Sigma\Delta$ -converters can be understood, at a heuristic level, by modelling  $q_{\mathbb{U}}(\cdot)$  as an additive white noise source. Following this paradigm, the feedback loop will shape the effect of the quantisation noise on the output signal. If the filters  $F(z)$  and  $G(z)$  are designed carefully, then the distortion spectrum can be *pushed out* of critical frequency bands.<sup>1</sup> As documented in various books and articles, see, e.g., (Norsworthy et al. 1997), and references therein, despite their simplicity,  $\Sigma\Delta$ -converters often give good results.

## 2.2 Decision Feedback Equalisation

Decision feedback equalisers (DFE's) underlie many design solutions within digital communication systems. For example, they can be used for channel equalisation (Paulraj et al. 2003) and multiuser detection (Tidestav 1999). DFE's outperform simpler linear approaches by respecting the quantised nature of signals.

Interestingly, as in the case of  $\Sigma\Delta$ -converters described above, DFE's can be characterised by the quantised feedback loop depicted in Fig. 1. Their operation can be visualised by focusing on equalisation of channels which have a digital, i.e., quantised, input and are affected by intersymbol interference. In that case, the input to the loop is given by the received signal (the communication channel output), which is affected by intersymbol interference. The output of the DFE corresponds to the estimated channel input symbols. Thus, the equaliser aims to invert the channel. The feedforward filter  $G(z)$  performs prefiltering of the received signal. The task of the feedback filter  $F(z)$  is to remove the portion of the intersymbol interference, that is a result of earlier symbols, from the current symbol to be detected. The rationale behind the circuit is that, if previous symbols are estimated correctly, then intersymbol interference can be removed and the channel input can be perfectly recovered.<sup>2</sup>

It can be seen that both,  $\Sigma\Delta$ -converters and DFE's, use the power of feedback to give simple design solutions in situations where decision variables are quantised. As we will show in Section 4.1, these schemes are intimately related to quantised finite horizon control.

## 3 Quantised Finite Horizon Control

Consider the first order system

$$x(\ell + 1) = ax(\ell) + bu(\ell), \quad (1)$$

where  $a$  and  $b$  are nonzero scalars and where the input is quantised, i.e., it is required to satisfy

$$u(\ell) \in \mathbb{U}, \quad \forall \ell. \quad (2)$$

In (2),  $\mathbb{U}$  is a given countable, possibly finite, set.

We are interested in evaluating the input for the system (1) via a finite horizon approach. In this methodology, at time instant,  $\ell = k$ , and given the plant state  $x(k)$  the following cost function is minimised:

$$V_N(\vec{u}(k)) \triangleq P(x'(k + N))^2 + \sum_{\ell=k}^{k+N-1} Q(x'(\ell))^2, \quad (3)$$

where:

$$\vec{u}(k) \triangleq [u'(k) \ u'(k + 1) \ \dots \ u'(k + N - 1)]^T \quad (4)$$

<sup>1</sup>We note that there exists a “Bode-type” trade-off, when designing  $F(z)$  and  $G(z)$ , see, e.g., (Gerzon & Craven 1989) and compare to work contained in (Serón, Braslavsky & Goodwin 1997).

<sup>2</sup>The drawback of this idea resides in the fact that channel input estimates are, in general, subject to errors. If wrong decisions are fed back then these errors are propagated and cause performance degradation; see, e.g., (Cantoni & Butler 1976, Quevedo et al. 2003).

contains the decision variables; see, e.g., (Quevedo, Goodwin & De Doná 2004).

In this cost,  $N \geq 1$  is the constraint horizon and  $P$  and  $Q$  are nonnegative scalar design variables. The cost examines predictions of the future state trajectory. These depend upon the model (1) and the variables contained in (4). More precisely, they are related via:

$$x'(\ell + 1) = ax'(\ell) + bu'(\ell), \quad \ell \in \{k, k + 1, \dots, k + N - 1\}, \quad (5)$$

with initial value  $x'(k) = x(k)$ , the current state, which is assumed to be known. In accordance with (2), all  $u'(\ell) \in \mathbb{U}$ , or, equivalently,

$$\vec{u}(k) \in \mathbb{U}^N, \quad \mathbb{U}^N \triangleq \mathbb{U} \times \dots \times \mathbb{U}.$$

In general choosing long horizons in (3) gives better performance than choosing short horizons, see also (Quevedo & Goodwin 2005c, Quevedo, Goodwin & Bölcskei 2004, Quevedo et al. 2003, Quevedo & Goodwin 2005b). Unfortunately, minimisation of  $V_N$  in (3), requires that one solve a combinatorial programme whose search space has a cardinality which is exponential in the horizon  $N$ . This precludes the use of long horizons in practical applications.

In contrast, using horizon-one, i.e., choosing  $N = 1$  in (3), requires only little computational effort. Thus, a simple alternative to solving (3) resides in computing the optimiser term-wise via  $N$  successive horizon-one optimisations. The main goal of the present work is to examine if there are situations when this, rather naïve, procedure is actually optimal for horizons  $N > 1$ .

As shown in (Quevedo, Goodwin & De Doná 2004), optimal control laws associated with (1)–(4) give rise to a partition of the plant state space. Here, it is interesting to note that whenever the plant state  $x(k)$  lies on a decision boundary, two (or more) candidate sequences, see (4), give identical (and optimal) costs  $V_N$ . Thus, in these situations –which for finite constraint sets are of measure zero– the optimising sequence is not uniquely defined. To avoid ambiguity, throughout this work we define minimisers as those which have minimal index with respect to the orders in  $\mathbb{U}$  and in  $\mathbb{U}^N$  to be introduced in Section 4.

#### 4 The Naïve Control Law

Having presented the main ideas surrounding quantised finite horizon control, in this section we will introduce the associated naïve control law. For that purpose, we will first describe the quantised horizon-one control law.

To keep expressions simple (and without loss of generality) we set  $k = 0$  in the sequel, denote the constrained minimiser to  $V_N(\vec{u}(0))$  as:

$$\vec{u}^* \triangleq \arg \min_{\vec{u}(0) \in \mathbb{U}^N} V_N(\vec{u}(0)) \quad (6)$$

and also write:

$$\vec{u}^* = [u_0^* \ u_1^* \ \dots \ u_{N-1}^*]^T. \quad (7)$$

In the first part of this paper, we will focus our attention on finite constraint sets given by:

$$\mathbb{U} = \{\gamma_0, -\gamma_1, \gamma_1, \dots, -\gamma_M, \gamma_M\}, \quad M \in \mathbb{N}. \quad (8)$$

Later, in Section 7, we will examine the situation of infinite, but countable, constraint sets.

As already mentioned in Section 3, we will first introduce ordering<sup>1</sup> in the constraint sets  $\mathbb{U}$  and  $\mathbb{U}^N$ :

<sup>1</sup>The choice of ordering is not unique. Nevertheless, the definitions used are convenient for our purposes.

Definition 4.1 (*Order in  $\mathbb{U}$* ) We define the order  $\leq$  on  $\mathbb{U}$  given in (8) via the convention:

$$\begin{aligned}\text{Index}(\gamma_0) &= 0 \\ \text{Index}(\gamma_k) &= 2k \\ \text{Index}(-\gamma_k) &= 2k - 1,\end{aligned}$$

where  $k > 0$ .

Definition 4.2 (*Order in  $\mathbb{U}^N$* ) Consider two elements

$$v = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N-1} \end{bmatrix}, \quad w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix} \in \mathbb{U}^N.$$

If  $v = w$ , we define  $\text{Index}(v) = \text{Index}(w)$ . Otherwise, we let  $m \triangleq \min\{0 \leq j < N \mid w_j \neq v_j\}$  and define:

$$\text{Index}(w) < \text{Index}(v) \Leftrightarrow \text{Index}(w_m) < \text{Index}(v_m).$$

For example, if  $N = 2$  and  $\mathbb{U} = \{0, -\delta, \delta\}$  we have

$$\mathbb{U}^2 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -\delta \end{bmatrix}, \begin{bmatrix} 0 \\ \delta \end{bmatrix}, \dots, \begin{bmatrix} \delta \\ -\delta \end{bmatrix}, \begin{bmatrix} \delta \\ \delta \end{bmatrix} \right\}.$$

#### 4.1 The horizon-one solution

If the horizon  $N$  equals 1, then the cost function  $V_N$  in (3) reduces to:

$$V_1(\vec{u}(0)) = Q(x(0))^2 + P(x'(1))^2. \quad (9)$$

Thus, the optimal control action (with minimal index) for initial state  $x(0) \in \mathbb{R}$  is

$$\vec{u}^* = u_0^* = q_{\mathbb{U}}(-ax(0)/b), \quad (10)$$

see (7). In (10),  $q_{\mathbb{U}} : \mathbb{R} \rightarrow \mathbb{U}$  denotes the scalar quantiser:

$$q_{\mathbb{U}}(v) = \begin{cases} -\gamma_M & \text{if } v < -(\gamma_M + \gamma_{M-1})/2 \\ \gamma_M & \text{if } v > (\gamma_M + \gamma_{M-1})/2 \\ \gamma_k & \text{if } (\gamma_k + \gamma_{k-1})/2 < v \leq (\gamma_{k+1} + \gamma_k)/2 \text{ for some } 1 \leq k < M \\ 0 & \text{if } -\gamma_1/2 \leq v \leq \gamma_1/2 \\ -\gamma_k & \text{if } -(\gamma_{k+1} + \gamma_k)/2 \leq v < -(\gamma_k + \gamma_{k-1})/2 \text{ for some } 1 \leq k < M. \end{cases} \quad (11)$$

By bringing (10) together with the system dynamics (1), one obtains that the horizon-one solution can be characterised as in Fig. 2. Comparison of this figure with the quantised loop depicted in Fig. 1 shows that both loops are described by the same dynamics, provided that in the scheme of Fig. 1 the input signal is removed and that:

$$F(z) = \frac{a}{z - a}.$$

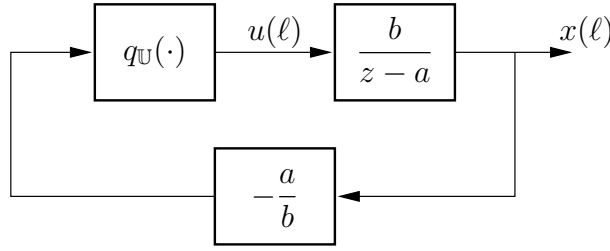


Figure 2. Quantised horizon-one control loop.

*Remark 1* Whilst the above result only applies to a restricted class of systems, in (Quevedo & Goodwin 2005c, Quevedo et al. 2003) we have established that the equivalence between horizon-one solutions and  $\Sigma\Delta$ -converters and DFE's also holds in more general situations.

#### 4.2 The Naïve Control Sequence

Based upon the horizon-one solution we can now define the naïve control sequence as follows:

**Definition 4.3 (Naïve Control Sequence)** The sequence  $(u'_0, \dots, u'_{N-1}) \in \mathbb{U}^N$  given by

$$u'_j = q_{\mathbb{U}}(-ax(j)/b), \quad x(\ell+1) = ax(\ell) + bu'_\ell \quad (12)$$

is called the *naïve control sequence* (with horizon  $N$  and initial state  $x(0)$ ).

We note that the naïve control sequence can be obtained via  $N$  successive  $\Sigma\Delta$ -conversion steps. It therefore requires only minimal computational effort.

In the remainder of this work we will investigate possible multi-step optimality of the naïve control sequence via the following notion:

**Definition 4.4 (Horizon- $N$  optimality of the naïve control law)** We say that the naïve control law (NCL) in (12) is (globally) optimal for horizon  $N$  iff<sup>1</sup>

$$u'_j = u_j^* \quad \forall j \in \{0, 1, \dots, N-1\} \quad \text{and all initial states } x(0) \in \mathbb{R}, \quad (13)$$

where  $u_j^*$  characterises the optimiser as in (7).

To examine horizon- $N$  optimality of the naïve control law, it is convenient to define the scalar quantisation error:

$$\varepsilon(w) \triangleq w - q_{\mathbb{U}}(w), \quad (14)$$

the family of nested nonlinearities:

$$\begin{aligned} g_1(w) &\triangleq w \\ g_{n+1}(w) &\triangleq a\varepsilon(g_n(w)) \end{aligned} \quad (15)$$

and the functions:

$$f_i(v) \triangleq \frac{1}{2} \left( \frac{P}{Q} g_i^2(v) + \sum_{j=1}^{i-1} g_j^2(v) \right), \quad i \geq 2. \quad (16)$$

<sup>1</sup>We write “iff” for “if and only if”.

We note that

$$g_{n+1}(w) = g_n(a\varepsilon(w)), \quad f_{i+1}(v) = f_i(a\varepsilon(v)) + v^2/2.$$

The above definitions allow us to state the main result of this section.

**THEOREM 4.5** *Consider  $N \geq 2$  and let  $\mathbb{U}$  be as in (8). Then, the following are equivalent:*

- 1) *The naïve control law is optimal for all horizons less or equal to  $N$*
- 2)  $\arg \min_{u \in \mathbb{U}} f_i(az + au) = q_{\mathbb{U}}(-z), \quad \forall i \in \{2, 3, \dots, N\}, \quad \forall z \in \mathbb{R}.$

*Proof* The proof is included in Appendix A. □

*Remark 2* The finite horizon optimisation problem (6) underlies receding horizon control with finite-set constraints as examined in (Quevedo et al. 2002, Quevedo, Goodwin & De Doná 2004, Bemporad 2003) and also used in (Goodwin et al. 2004, Quevedo & Goodwin 2005c, Quevedo & Goodwin 2005b). It can be readily verified that, if the NCL is horizon- $N$  optimal (see Definition 4.4), then it is also  $N$ -step optimal in a moving horizon sense. Furthermore, it is easy to show that this condition is necessary and sufficient in the case of horizon  $N = 2$ .

## 5 Sufficient Conditions for Horizon- $N$ Optimality of the NCL

We will next utilise the equivalent form established in Theorem 4.5 to further elucidate possible horizon- $N$  optimality of the NCL. More precisely, we will base the subsequent analysis on the following corollary:

**COROLLARY 5.1** *If all functions  $f_2, f_3, \dots, f_{\hat{N}}$  are strictly monotonic increasing on  $[0, \infty)$ , then the NCL is horizon- $N$  optimal for all horizons  $N \leq \hat{N}$ .*

*Proof* If  $f_2, f_3, \dots, f_{\hat{N}}$  are strictly monotonic increasing on  $[0, \infty)$ , then

$$\arg \min_{u \in \mathbb{U}} f_i(az + au) = q_{\mathbb{U}}(-z), \quad \forall i \in \{2, 3, \dots, \hat{N}\}, \quad \forall z \in \mathbb{R}.$$

The result then follows immediately from Theorem 4.5. □

The above result gives a sufficient condition for the NCL to be horizon- $N$  optimal. Unfortunately, showing that the functions  $f_i$  with  $i \geq 2$  are strictly monotonic is far from trivial. These functions comprise linear combinations of the squares of the functions  $g_i$  defined in (15). Within that family, only  $(g_1)^2$  is convex; all other functions  $(g_i)^2$  are non-convex and non-smooth.

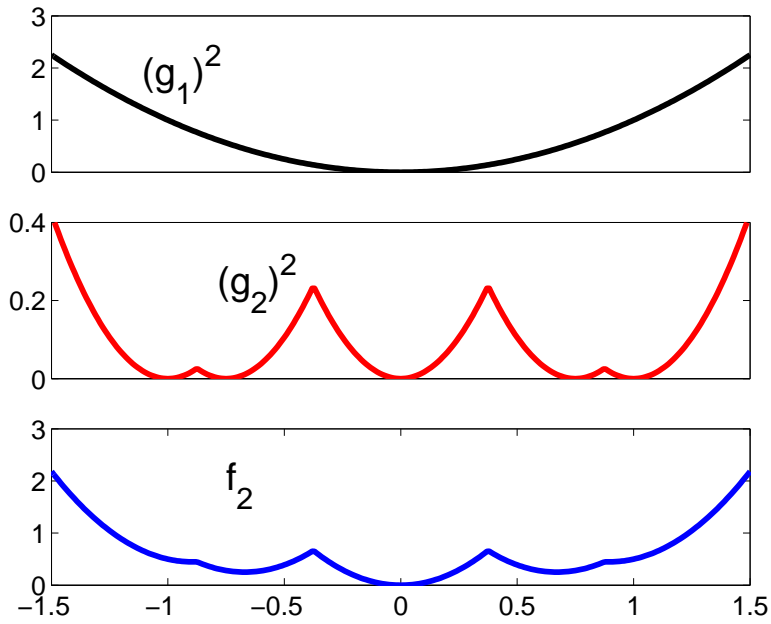
Fig. 3 illustrates this situation. The figure shows plots of functions  $(g_1)^2$ ,  $(g_2)^2$  and of  $f_2$ , for the following parameters:  $a = 1.3$ ,  $(P/Q) = 5$ , and  $\mathbb{U} = \{0, -0.75, 0.75, -1, 1\}$ . As is clear from the figure,  $f_2$  is not monotonic on  $[0, \infty)$ . The main reason for this resides in the fact that  $(g_2)^2$  is weighted heavily in the linear combination which defines  $f_2$ , see (16).

Two key observations follow from Fig. 3. Firstly, (in the case illustrated)  $f_2$  is continuous. Secondly, to apply Corollary 5.1, one may use differential calculus techniques. In the sequel, we will formalise these two observations for general situations.

**LEMMA 5.2** *All functions  $f_n$  defined in (16) are even and continuous.*

*Proof* The proof is included in Appendix B. □

Next, we examine the derivatives of  $(g_n)^2$  and of  $f_n$ . In this context, it is convenient to utilise right-hand

Figure 3. Typical functions  $(g_1)^2$ ,  $(g_2)^2$  and  $f_2$ .

derivatives<sup>1</sup> and to introduce the functions<sup>2</sup>:

$$\tilde{\varepsilon}(v) \triangleq \lim_{h \downarrow 0} \varepsilon(v+h) = \begin{cases} \varepsilon(v), & \text{if } v \notin \left\{ \frac{\gamma_k + \gamma_{k-1}}{2} \mid 1 \leq k \leq M \right\} \\ \frac{\gamma_{k-1} - \gamma_k}{2}, & \text{if } v = \frac{\gamma_k + \gamma_{k-1}}{2} \text{ for some } 1 \leq k \leq M \end{cases}, \quad (17)$$

and

$$\begin{aligned} \tilde{g}_1(v) &\triangleq v, \\ \tilde{g}_{j+1} &\triangleq a\tilde{\varepsilon}(\tilde{g}_j(v)). \end{aligned} \quad (18)$$

LEMMA 5.3 *The functions  $g_n^2$  and  $f_n$  are twice differentiable from the right on  $\mathbb{R}$ , and*

- 1)  $\frac{d^+}{dv} g_n^2(v) \triangleq \lim_{h \downarrow 0} \frac{g_n^2(v+h) - g_n^2(v)}{h} = 2a^{n-1}\tilde{g}_n(v) \quad \text{if } n \geq 1$
- 2)  $\frac{d^+}{dv} \left( \frac{d^+}{dv} g_n^2 \right) (v) = 2a^{2n-2} \quad \text{if } n \geq 1$
- 3)  $\frac{d^+}{dv} f_n(v) = \frac{P}{Q} a^{n-1}\tilde{g}_n(v) + \sum_{j=1}^{n-1} a^{j-1}\tilde{g}_j(v) \quad \text{if } n \geq 2$
- 4)  $\frac{d^+}{dv} \left( \frac{d^+}{dv} f_n \right) (v) = \frac{P}{Q} a^{2n-2} + \sum_{j=1}^{n-1} a^{2j-2} \quad \text{if } n \geq 2$

*Proof* The proof is included in Appendix C. □

COROLLARY 5.4 *The functions  $f_n$  are strictly monotonic increasing on  $[0, \infty)$  iff  $\frac{d^+}{dv} f_n(v) \geq 0$  for all  $v > 0$ .*

*Proof* The proof is included in Appendix D. □

<sup>1</sup>See Remark F1 in Appendix F.

<sup>2</sup>Compare with (14) and (15).

Corollary 5.4 can, in principle, be brought together with Corollary 5.1 to examine horizon- $N$  optimality of the NCL via examination of  $\frac{d^+}{dv}f_n(v)$  for all  $v > 0$ . Fortunately, it turns out that it is sufficient to examine positiveness of  $\frac{d^+}{dv}f_n(v)$  at a finite number of points. To make this statement more precise, we define:

$$\begin{aligned}\mathbb{G}_1 &\triangleq \left\{ \frac{\gamma_k + \gamma_{k-1}}{2} \mid 1 \leq k \leq M \right\} \cup \left\{ -\frac{\gamma_k + \gamma_{k-1}}{2} \mid 1 \leq k \leq M \right\} \\ \mathbb{G}_j &\triangleq \{v \in \mathbb{R} \mid \tilde{g}_j(v) \in \mathbb{G}_1\} \quad \text{if } j \geq 1\end{aligned}\tag{19}$$

and

$$\mathbb{G}_j^+ \triangleq \mathbb{G}_j \cap (0, \infty).$$

LEMMA 5.5 *The sets  $\mathbb{G}_j$  and  $\mathbb{G}_j^+$  satisfy:*

- 1)  $\mathbb{G}_j = -\mathbb{G}_j$  for all  $j \geq 1$
- 2)  $\mathbb{G}_j^+$  is non-empty and finite for all  $j \geq 1$

*Proof* See Appendix E. □

We then obtain the following results:

THEOREM 5.6 *Let  $n \geq 2$ . If*

$$\min_{1 \leq j < n} \min_{v \in \mathbb{G}_j^+} \frac{d^+}{dv}f_n(v) \geq 0,$$

*then  $f_n$  is strictly monotonic increasing on  $[0, \infty)$ .*

*Proof* The proof is included in Appendix F. □

This leads to the main result of this section, which we state as:

COROLLARY 5.7 *Let  $\hat{N} \geq 2$ . If*

$$\min_{1 < n \leq \hat{N}} \min_{1 \leq j < n} \min_{v \in \mathbb{G}_j^+} \left( \frac{P}{Q} a^{n-1} \tilde{g}_n(v) + \sum_{j=1}^{n-1} a^{j-1} \tilde{g}_j(v) \right) \geq 0,\tag{20}$$

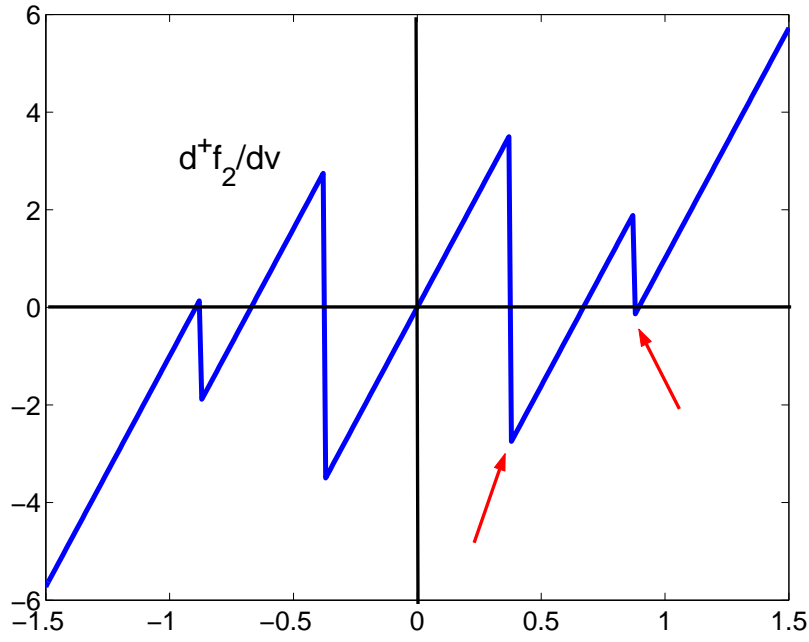
*then the NCL is horizon- $N$  optimal  $\forall N \leq \hat{N}$ .*

*Proof* By Lemma 5.3 and Theorem 5.6 the functions  $f_2, \dots, f_{\hat{N}}$  are strictly monotonic increasing on  $[0, \infty)$ . The assertion follows from Corollary 5.1. □

Note that, as a consequence of Lemma 5.5, only a finite number of evaluations needs to be made to apply this result. This stands in stark contrast to investigating horizon- $N$  optimality of the NCL via Corollaries 5.1 or 5.4, where conditions on  $f_n(v)$  or  $\frac{d^+}{dv}f_n(v)$  need to be examined for all  $v > 0$ .

As an illustration, consider Fig. 4, which contains a plot of the derivative of  $f_2$  as shown in Fig. 3. As can be seen,  $\frac{d^+}{dv}f_2(v)$  is negative on some intervals, where  $v > 0$  and (20) is not satisfied. This explains why, for this case,  $f_2$  is not strictly monotonic on  $[0, \infty)$  and the sufficient condition for horizon-2 optimality in (20) is not satisfied.

Corollary 5.7 establishes conditions on the plant pole  $a$ , the weights  $P$  and  $Q$ , and the constraint set  $\mathbb{U}$  which are sufficient to ensure that the NCL is horizon- $N$  optimal. In the following section we will show how this result can be applied to several specific situations.

Figure 4.  $\frac{d^+}{dv} f_2(v)$  for the case included in Fig. 3.

*Remark 1* Note that the terms  $a^{j-1}\tilde{g}_j(v)$  in (20) do not change their value when  $a$  is replaced by its complement,  $(-a)$ . Thus, the derivative from the right of  $f_i$  is, so to say, *even in*  $a$ . Therefore, it is sufficient to restrict analysis to  $a > 0$ .

## 6 Special Cases

In this section, we will focus our attention on specific cases, namely horizon  $N = 2$ , and stable systems (1). Under these conditions, we will further investigate the results stated in Corollary 5.7.

### 6.1 Horizon $N = 2$

The general methodology described above can be readily applied to horizon-2 optimality. Indeed, we have the following:

**COROLLARY 6.1** *Consider any constraint set  $\mathbb{U}$  of the form (8). If*

$$P \leq a^{-2}Q, \quad (21)$$

*then the NCL is optimal for horizon  $N = 2$ .*

*Proof* We have

$$\begin{aligned} \min_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_2(v) &= \min_{1 \leq k \leq M} \left( \frac{P}{Q} a^2 \tilde{\varepsilon} \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) + \frac{\gamma_k + \gamma_{k-1}}{2} \right) \\ &= \min_{1 \leq k \leq M} \left( \frac{P}{Q} a^2 \left( \frac{\gamma_{k-1} - \gamma_k}{2} \right) + \frac{\gamma_k + \gamma_{k-1}}{2} \right) \\ &= \frac{1}{2} \min_{1 \leq k \leq M} \left( \gamma_{k-1} \left( 1 + \frac{P}{Q} a^2 \right) + \gamma_k \left( 1 - \frac{P}{Q} a^2 \right) \right). \end{aligned}$$

Thus,  $\min_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_2(v) \geq 0 \Leftrightarrow 1 - \frac{P}{Q} a^2 \geq 0$ . □

Condition (21) can be easily interpreted, by noticing that, with  $N = 2$ , the cost function (3) reduces to:

$$V_2(\vec{u}(k)) = Q(x(k))^2 + Q(x'(k+1))^2 + P(x'(k+2))^2. \quad (22)$$

Thus, Corollary 6.1 essentially states that, if the term  $P(x'(k+2))^2$  in  $V_2$  is sufficiently small (for all  $x'(k+2)$ ), then NCL is horizon-2 optimal. Given that for horizon  $N = 1$ , the cost function is as in (9), the result becomes intuitively clear.

## 6.2 Stable Systems

For open-loop stable systems, i.e., when  $0 < |a| < 1$ , the situation becomes particularly interesting. Perhaps surprisingly, the following expressions hold:

LEMMA 6.2 *Let  $0 < a < 1$  and  $a(\gamma_k - \gamma_{k-1}) < \gamma_1$  for all  $1 \leq k \leq M$ , see (8). Then:*

- 1)  $\tilde{g}_n \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) = a^{n-1}(\gamma_{k-1} - \gamma_k)/2$ , for all  $n \geq 2$ ,  $1 \leq k \leq M$ .
- 2) If  $j \geq 1$  then  $\mathbb{G}_{j+1}^+ \subset (\gamma_M, \infty)$ , and if  $v \in \mathbb{G}_{j+1}^+$ , then  $a\tilde{\varepsilon}(v) \in \mathbb{G}_j^+$ .
- 3) If  $j \geq 1$  and  $n \geq 2$ , then

$$\min_{v \in \mathbb{G}_{j+1}^+} \frac{d^+}{dv} f_{n+1}(v) \geq a \cdot \min_{v \in \mathbb{G}_j^+} \frac{d^+}{dv} f_n(v).$$

- 4) If  $n \geq 2$ , then

$$\min_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_n(v) \geq 0 \Leftrightarrow \frac{P}{Q}(a^2)^{n-1} \leq \frac{1 - 2a^2 + (a^2)^{n-1}}{1 - a^2}.$$

*Proof* The proof is included in Appendix G. □

As a consequence of the above expressions, the analysis simplifies enormously. Indeed, rather than evaluating  $\frac{d^+}{dv} f_n(v)$  in (20) over all  $v \in \mathbb{G}_j^+$ , where  $j \leq \hat{N}$ , optimality of the NCL can be ensured by restricting the analysis to  $v \in \mathbb{G}_1^+$ . This leads to the following simplifications:

COROLLARY 6.3 *Let*

$$0 < a \leq 1/\sqrt{2}, \quad a(\gamma_k - \gamma_{k-1}) < \gamma_1, \quad \forall 1 \leq k \leq M.$$

*If (21) holds, then the NCL is horizon- $N$  optimal for all horizons.*

*Proof* The proof is included in Appendix H. □

COROLLARY 6.4 *Let*

$$1/\sqrt{2} < a < 1, \quad a(\gamma_k - \gamma_{k-1}) < \gamma_1, \quad \forall 1 \leq k \leq M$$

*and consider  $\hat{N} \in \mathbb{N}$  such that*

$$2 \leq \hat{N} < 1 + \frac{\ln(2a^2 - 1)}{2 \ln a}.$$

*If*

$$\frac{P}{Q} \leq \frac{1 - 2a^2 + (a^2)^{\hat{N}-1}}{(a^2)^{\hat{N}-1}(1 - a^2)}, \quad (23)$$

then the NCL is horizon- $N$  optimal  $\forall N \leq \hat{N}$ .

*Proof* The proof is included in Appendix I.  $\square$

As an example of the result stated in Corollary 6.3, consider  $a = 0.7$ ,  $\mathbb{U} = \{0, -1, 1\}$ , and  $\hat{N} = 4$ . In this case, (23) reduces to  $(P/Q) \leq 2.04$ . Fig. 5 contains plots of  $f_2$ ,  $f_3$  and  $f_4$  for the choices  $(P/Q) = 2$  and  $(P/Q) = 4$ . As can be seen in the figure, with  $(P/Q) = 2$ , the functions  $f_2$ ,  $f_3$  and  $f_4$  are all monotonic increasing. However, this does not hold when  $(P/Q) = 4$ . Following Corollary 6.3, we conclude that the NCL with  $(P/Q) = 2$  is optimal for  $N \in \{1, 2, 3, 4\}$ . On the other hand, we cannot ensure that the NCL with  $(P/Q) = 4$  is optimal for horizons larger than one.

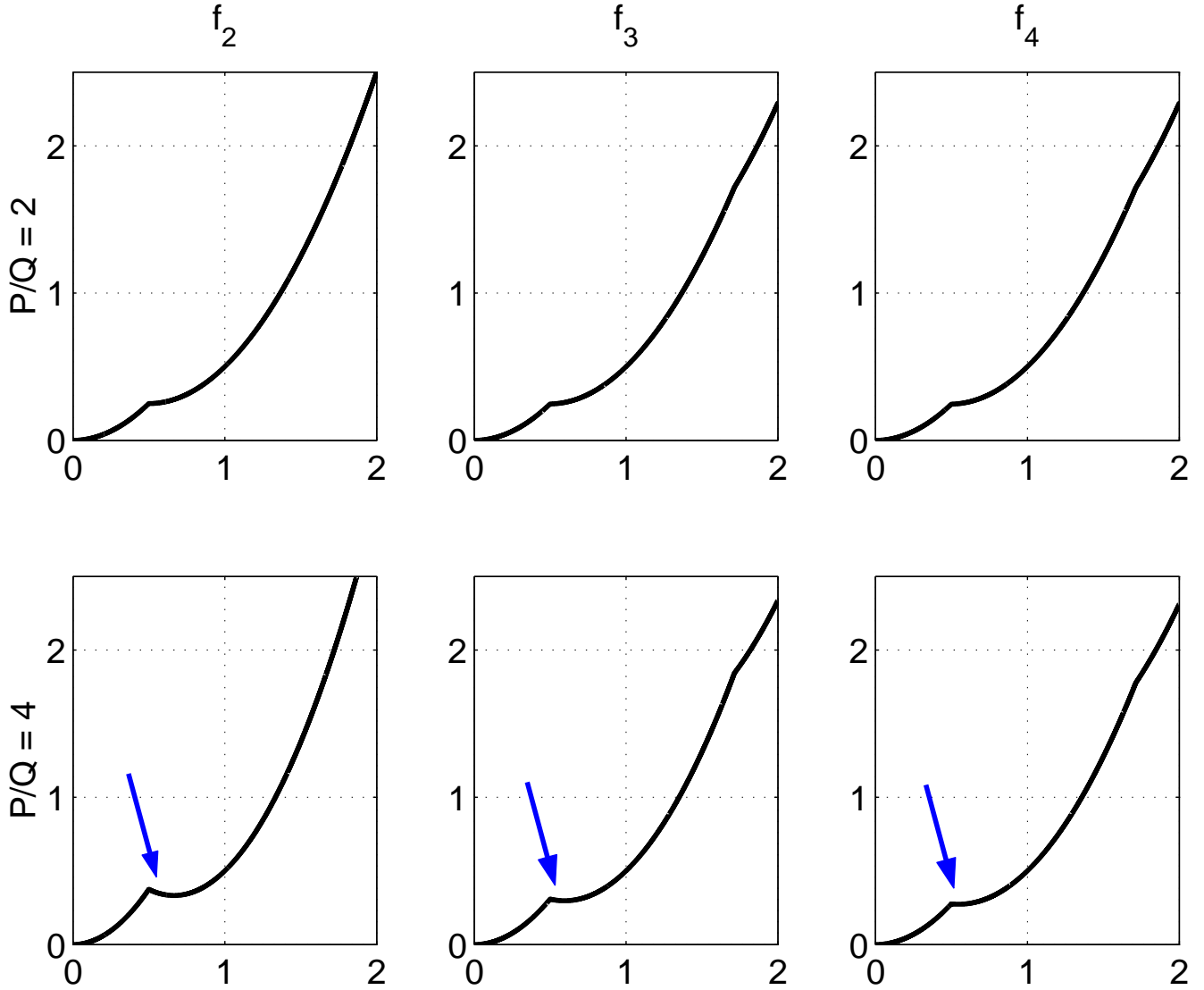


Figure 5. Functions  $f_2$ ,  $f_3$  and  $f_4$  given  $(P/Q) = 2$  (top row) and given  $(P/Q) = 4$  (bottom row).

## 7 Infinite Constraint Sets

So far we have examined finitely quantised control laws, where the constraint set  $\mathbb{U}$  is given as in (8). We will now examine the situation where the constraint set has infinite cardinality. More precisely, we will consider

$$\mathbb{U} = \{\gamma_0, -\gamma_1, \gamma_1, \dots\} \quad (24)$$

where:

$$0 = \gamma_0 < \gamma_1 < \gamma_2 < \dots, \quad \lim_{k \rightarrow \infty} \gamma_k = \infty. \quad (25)$$

Note that we require (25) to be satisfied, so as to ensure the existence of the optimiser (6). As an illustration, this rules out sequences  $(\gamma_k)_{k \in \mathbb{N}}$ , where  $\gamma_k = 1 - 1/k$ , for which the optimiser does not exist in general.

It turns out that the main results developed in the previous sections for finite constraint sets can be applied *mutatis mutandis* to infinite constraint sets as in (24) although some additional technicalities arise in the proofs. To keep the exposition short, we will outline the main aspects and state the key results without proof.

### 7.1 Basic Definitions

In the infinite constraint set case, all the definitions of Sections 3 and 4 hold essentially unaltered. The only exception is that, with  $\mathbb{U}$  as in (24), the scalar quantiser, see (11), has no extremal regions. To be more precise, it is characterised via

$$q_{\mathbb{U}}(v) = \begin{cases} \gamma_k & \text{if } (\gamma_k + \gamma_{k-1})/2 < v \leq (\gamma_{k+1} + \gamma_k)/2 \text{ for some } 1 \leq k < M \\ 0 & \text{if } -\gamma_1/2 \leq v \leq \gamma_1/2 \\ -\gamma_k & \text{if } -(\gamma_{k+1} + \gamma_k)/2 \leq v < -(\gamma_k + \gamma_{k-1})/2 \text{ for some } 1 \leq k < M. \end{cases} \quad (26)$$

### 7.2 Sufficient Conditions for Horizon- $N$ Optimality of the NCL when $\mathbb{U}$ has Infinite Cardinality

The results contained in Section 5, can be directly extended to the present case. Indeed, Lemma 5.3 and Corollary 5.4 still apply. We simply note that, in the infinite constraint set case,  $\tilde{\varepsilon}(v)$  defined in (17) satisfies:

$$\tilde{\varepsilon}(v) = \begin{cases} \varepsilon(v) & \text{if } v \notin \left\{ \frac{\gamma_k + \gamma_{k-1}}{2} \mid k \geq 1 \right\} \\ (\gamma_{k-1} - \gamma_k)/2 & \text{if } v = (\gamma_k + \gamma_{k-1})/2 \text{ for some } k \geq 1 \end{cases}. \quad (27)$$

Similarly, if we re-define

$$\begin{aligned} \mathbb{G}_1 &\triangleq \left\{ \frac{\gamma_k + \gamma_{k-1}}{2} \mid k \geq 1 \right\} \cup \left\{ -\frac{\gamma_k + \gamma_{k-1}}{2} \mid k \geq 1 \right\} \\ \mathbb{G}_j &\triangleq \{v \in \mathbb{R} \mid \tilde{g}_j(v) \in \mathbb{G}_1\}, \\ \mathbb{G}_j^+ &\triangleq \mathbb{G}_j \cap (0, \infty), \quad j \geq 1, \end{aligned} \quad (28)$$

then we obtain the following results:

THEOREM 7.1 *Let  $n \geq 2$ . If*

$$\min_{\substack{1 \leq j < n \\ \mathbb{G}_j^+ \neq \emptyset}} \inf_{v \in \mathbb{G}_j^+} \frac{d^+}{dv} f_n(v) \geq 0,$$

*then  $f_n$  is strictly monotonic increasing on  $[0, \infty)$ .*

*Proof* The proof follows similar lines to the proof of Theorem 5.6 and is therefore omitted.<sup>1</sup> □

COROLLARY 7.2 *Let  $\hat{N} \geq 2$ . If*

$$\min_{1 < n \leq N} \min_{\substack{1 \leq j < n \\ \mathbb{G}_j^+ \neq \emptyset}} \inf_{v \in \mathbb{G}_j^+} \left( \frac{P}{Q} a^{n-1} \tilde{g}_n(v) + \sum_{j=1}^{n-1} a^{j-1} \tilde{g}_j(v) \right) \geq 0,$$

*then the NCL is horizon- $N$  optimal  $\forall N \leq \hat{N}$ .*

*Proof* The proof is similar to the proof of Corollary 5.7. □

### 7.3 Special Cases

Here we examine some specific situations as was done in Section 6. The extension of Corollary 6.1 to the infinite cardinality case is:

COROLLARY 7.3 *If (21) holds, then the NCL is horizon-2 optimal.*

*Proof* The proof is similar to the proof of Corollary 6.1. □

For stable systems we can derive the following counterpart to Lemma 6.2:

LEMMA 7.4 *Let  $0 < a < 1$ ,  $\mathbb{U}$  as in (24), and  $a(\gamma_k - \gamma_{k-1}) < \gamma_1$  for all  $k \geq 1$ . Then we have:*

- 1)  $\tilde{g}_n \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) = a^{n-1}(\gamma_{k-1} - \gamma_k)/2$  for all  $n \geq 2, k \geq 1$
- 2)  $\mathbb{G}_j = \emptyset$  for all  $j \geq 2$
- 3) If  $n \geq 2$  then

$$\inf_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_n(v) \geq 0 \Leftrightarrow \frac{P}{Q} a^{2n-2} \leq \frac{1 - 2a^2 + a^{2n-2}}{1 - a^2}$$

*Proof* The proof of 1) and 3) is similar to the proof of Lemma 6.2.

2) As in the proof of Lemma 6.2, we see that  $\max_{|v| \leq R} |\tilde{\varepsilon}(v)| < \frac{\gamma_1}{2a}$  for all  $R > 0$ .

If  $v \in \mathbb{G}_2$  then  $a\tilde{\varepsilon}(v) \in \mathbb{G}_1$ , thus  $|a\tilde{\varepsilon}(v)| \geq \gamma_1/2$ , which is a contradiction. Thus  $\mathbb{G}_2 = \emptyset$ .

If  $\mathbb{G}_j = \emptyset$  for some  $j \geq 2$  and  $v \in \mathbb{G}_{j+1}$  then  $a\tilde{\varepsilon}(v) \in \mathbb{G}_j = \emptyset$ , thus  $\mathbb{G}_{j+1} = \emptyset$ . This proves 2). □

As a consequence of this result, only  $\mathbb{G}_1$  need be examined when using Corollary 7.2. It is easy to show – and very interesting to note – that Corollaries 6.3 and 6.4 still apply in the infinite cardinality case, provided  $a(\gamma_k - \gamma_{k-1}) < \gamma_1$  for all  $k \geq 1$ .

<sup>1</sup>Note that  $\mathbb{J}_n$  defined as in (F1) is no longer finite. Thus, we obtain  $\inf_{v \geq 0} \frac{d^+}{dv} f_n(v) = \inf_{v \in \mathbb{J}_n \cup \{0\}} \frac{d^+}{dv} f_n(v)$ .

## 8 Conclusions

This paper has presented conditions which guarantee that the naïve quantised control law is finite horizon optimal for horizons larger than one. We have utilised differential calculus techniques to derive these sufficient conditions. The conditions can be expressed as upper bounds on the ratio of the weighting terms in the quadratic cost function. These bounds depend upon the magnitude of the system pole and the horizon length. The work presented gives further insight into basic properties of  $\Sigma\Delta$ -converters and decision feedback equalisers. The latter structures are used in many application areas. Thus, it is of practical importance to know when using optimisation with horizon length greater than one gives significantly better performance than using horizon one (which turns out to be equivalent to  $\Sigma\Delta$ -conversion). The current paper has taken a first step in that direction by showing that there exist cases where increasing the horizon beyond one makes no difference at all.

Many problems remain open, e.g., how to deal with non-zero reference signals and how to quantify the impact of increasing the horizon in cases where it does affect the result. These topics are currently under study using geometric insights related to those developed in the current paper. Future research may also include the development of both necessary and sufficient conditions for horizon- $N$  optimality of the NCL; closer examination of *moving horizon optimality* (see Remark 2); extensions to higher-order systems; and investigation of *local* conditions for optimality of the NCL. It would also be interesting to weaken the notion of optimality and to study the performance loss which results from using the NCL instead of quantised finite horizon control.

## References

- Baillieul, J. (2002), Feedback coding for information-based control: Operating near the data-rate limit, in 'Proc. IEEE Conf. Decis. Contr.', Las Vegas, Nevada.
- Bemporad, A. (2003), Multiparametric nonlinear integer programming and explicit quantized optimal control, in 'Proc. IEEE Conf. Decis. Contr.', Maui, Hawaii, pp. 3167–3172.
- Bicchi, A., Marigo, A. & Piccoli, B. (2002), 'On the reachability of quantized control systems', *IEEE Trans. Automat. Contr.* **47**(4), 546–562.
- Bockman, S. F. (1991), On-Off, discrete time control of a stable first-order plant, in 'Proc. Amer. Contr. Conf.', Boston, MA, pp. 1258–1262.
- Bölcskei, H. & Hlawatsch, F. (2001), 'Noise reduction in oversampled filter banks using predictive quantization', *IEEE Trans. Inform. Theory* **47**(1), 155–172.
- Brockett, R. W. & Liberzon, D. (2000), 'Quantized feedback stabilization of linear systems', *IEEE Trans. Automat. Contr.* **45**(7), 1279–1289.
- Camacho, E. F. & Bordons, C. (1999), *Model Predictive Control*, Springer-Verlag, New York, N.Y.
- Candy, J. C. & Temes, G. C., eds (1992), *Oversampling Delta-Sigma Data Converters Theory, Design and Simulation*, IEEE Press., New York.
- Cantoni, A. & Butler, P. (1976), 'Stability of decision feedback inverses', *IEEE Trans. Commun.* **24**(9), 970–977.
- De Doná, J. A. & Goodwin, G. C. (1999), 'Disturbance sensitivity issues in predictive control', *Int. J. Adapt. Contr. Signal Proc.* **13**, 507–519.
- De Doná, J. A., Goodwin, G. C. & Serón, M. M. (2000), 'Anti-windup and model predictive control: Reflections and connections', *European J. Contr.* **6**(5), 467–477.
- Delchamps, D. F. (1990), 'Stabilizing a linear system with quantized state feedback', *IEEE Trans. Automat. Contr.* **35**, 916–924.
- Dreyfus, S. E. & Law, A. M. (1977), *The Art and Theory of Dynamic Programming*, Academic Press, New York, NY.
- Evangelista, G. (2001), 'Least mean squared error-design of complex FIR filters with quantized coefficients', *IEEE Trans. Circuits Syst. II* **48**(8), 778–784.
- Fagnani, F. & Zampieri, S. (2003), 'Stability analysis and synthesis for scalar linear systems with a quantized feedback', *IEEE Trans. Automat. Contr.* **48**(9), 1569–1584.
- Forney, G. (1972), 'Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference', *IEEE Trans. Inform. Theory* **18**(3), 363–378.
- Gerzon, M. & Craven, P. G. (1989), Optimal noise shaping and dither of digital signals, in '87th Convention of the AES, New York, NY, preprint 2822'.
- Goodwin, G. C., Haimovich, H., Quevedo, D. E. & Welsh, J. S. (2004), 'A moving horizon approach to networked control system design', *IEEE Trans. Automat. Contr.* **49**(9), 1427–1445.
- Goodwin, G. C. & Quevedo, D. E. (2003), 'Finite alphabet control and estimation', *Int. J. Contr., Automation, and Syst.* **1**(4), 412–430.
- Goodwin, G. C., Quevedo, D. E. & McGrath, D. (2003), 'Moving-horizon optimal quantizer for audio signals', *J. Audio Eng. Soc.* **51**(3), 138–149.
- Goodwin, G. C., Serón, M. M. & De Doná, J. A. (2005), *Constrained Control & Estimation – An Optimization Perspective*, Springer-Verlag, London.
- Ishii, H. & Başar, T. (2004), 'Randomized algorithms for quadratic stability of quantized sampled-data systems', *Automatica* **40**, 839–846.
- Ishii, H. & Francis, B. A. (2002), *Limited Data Rate in Control Systems with Networks*, Springer.
- Kiihtelys, K. (2003), Bandwidth and quantization constraints in feedback control, Master's thesis, Dep. of Systems, Signals and Sensors (S3), Royal Inst. Techn. (KTH), Stockholm, Sweden.
- Li, K. & Baillieul, J. (2004), 'Robust quantization for digital finite communication bandwidth (DFCB) control', *IEEE Trans. Automat. Contr.* **49**(9), 1573–1584.

- Lim, Y. C. & Parker, S. R. (1983), 'Discrete coefficient FIR digital filter design based upon an LMS criteria', *IEEE Trans. Circuits Syst.* **30**(10), 723–739.
- Lipschitz, S. P., Vanderkooy, J. & Wannamaker, R. A. (1991), 'Minimally audible noise shaping', *J. Audio Eng. Soc.* **39**(11), 836–852.
- Maciejowski, J. M. (2002), *Predictive Control with Constraints*, Prentice-Hall.
- Nair, G. N. & Evans, R. J. (2000), 'Stabilization with data-rate-limited feedback: Tightest attainable bounds', *Syst. & Contr. Lett.* **41**, 49–56.
- Nielsen, J. J. (1989), 'Design of linear-phase direct-form FIR digital filters with quantized coefficients using error spectrum shaping', *IEEE Trans. Acoust., Speech, Signal Processing* **37**(7), 1020–1026.
- Norsworthy, S. R., Schreier, R. & Temes, G. C., eds (1997), *Delta-Sigma Data Converters: Theory, Design and Simulation*, IEEE Press, Piscataway, N.J.
- Paramesh, L. & von Jouanne, A. (2001), 'Use of Sigma-Delta modulation to control EMI from switch-mode power supplies', *IEEE Trans. Ind. Electron.* **48**(1), 111–117.
- Paulraj, A., Nabar, R. & Gore, D. (2003), *Introduction to Space-Time Wireless Communications*, Cambridge University Press.
- Picasso, B. & Bicchì, A. (2003), Stabilization of LTI systems with quantized state-quantized input static feedback, in A. Pnueli & O. Maler, eds, 'Hybrid Systems: Computation and Control', Vol. LNCS 2623 of *Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg, Germany, pp. 405–416.
- Proakis, J. G. (1995), *Digital Communications*, McGraw-Hill.
- Quevedo, D. E., De Doná, J. A. & Goodwin, G. C. (2002), Receding horizon linear quadratic control with finite input constraint sets, in 'Proc. 15th IFAC World Congress', Barcelona, Spain.
- Quevedo, D. E. & Goodwin, G. C. (2004a), Control of EMI from switch-mode power supplies via multi-step optimization, in 'Proc. Amer. Contr. Conf.', Boston, MA.
- Quevedo, D. E. & Goodwin, G. C. (2004b), When is the naïve quantized control law globally optimal?, in 'Proc. Asian Control Conference', Melbourne, Australia, pp. 1478–1486.
- Quevedo, D. E. & Goodwin, G. C. (2005a), An improved architecture for networked control systems, in 'Proc. 16th IFAC World Congress', Prague, Czech Republic.
- Quevedo, D. E. & Goodwin, G. C. (2005b), 'Moving horizon design of discrete coefficient FIR filters', *IEEE Trans. Signal Processing* **53**(6), 2262–2267.
- Quevedo, D. E. & Goodwin, G. C. (2005c), 'Multistep optimal analog-to-digital conversion', *IEEE Trans. Circuits Syst. I* **52**(4), 503–515.
- Quevedo, D. E., Goodwin, G. C. & Bölskei, H. (2004), Multi-step optimal quantization in oversampled filter banks, in 'Proc. IEEE Conf. Decis. Contr.', Paradise Island, Bahamas.
- Quevedo, D. E., Goodwin, G. C. & De Doná, J. A. (2003), A multi-step detector for linear ISI-channels incorporating degrees of belief in past estimates, Technical Report EE03013, School of Elect. Eng. and Comput. Sci., The Univ. of Newcastle, NSW 2308, Australia.
- Quevedo, D. E., Goodwin, G. C. & De Doná, J. A. (2004), 'Finite constraint set receding horizon quadratic control', *Int. J. Robust Nonlin. Contr.* **14**(4), 355–377.
- Rao, C. V., Rawlings, J. B. & Lee, J. H. (2001), 'Constrained linear state estimation – A moving horizon approach', *Automatica* **37**, 1619–1628.
- Rodríguez, J., Lai, J.-S. & Peng, F. Z. (2002), 'Multilevel inverters: A survey of topologies, controls, and applications', *IEEE Trans. Ind. Electron.* **49**(4), 724–738.
- Serón, M. M., Braslavsky, J. H. & Goodwin, G. C. (1997), *Fundamental Limitations in Filtering and Control*, Springer-Verlag, London.
- Sznaier, M. & Damborg, M. J. (1989), 'Control of constrained discrete time linear systems using quantized controls', *Automatica* **25**(4), 623–628.
- Tewksbury, S. K. & Hallock, R. W. (1978), 'Oversampled, linear predictive and noise-shaping coders of order  $N > 1$ ', *IEEE Trans. Circuits Syst.* **25**(7), 436–447.
- Tidestav, C. (1999), The Multivariable Decision Feedback Equalizer, PhD thesis, Signals and Systems, Uppsala University.
- Williamson, D., Kennedy, R. A. & Pulford, G. W. (1992), 'Block decision feedback equalization', *IEEE Trans. Commun.* **40**(2), 255–264.

## Appendix A: Proof of Theorem 4.5

We use the principle of optimality; see, e.g., (Dreyfus & Law 1977). We first show that the NCL achieves the optimal cost for all horizons less than or equal to  $N$  iff  $u = q_{\mathbb{U}}(-z)$  is a minimiser of the function  $f_i(az + au)$  on  $\mathbb{U}$  for all  $2 \leq i \leq N$  and all  $z \in \mathbb{R}$ .

Towards this goal we denote by  $S_i(x)$  the optimal cost-to-go with initial state  $x \in \mathbb{R}$  and horizon  $N - i$ ,  $0 \leq i < N$ , i.e.

$$S_i(x) \triangleq \min_{u_i, \dots, u_{N-1} \in \mathbb{U}} Px(N)^2 + \sum_{\ell=i}^{N-1} Qx(\ell)^2,$$

where  $x(i) = x$  and  $x(\ell + 1) = ax(\ell) + bu_{\ell}$  if  $i \leq \ell < N$ .

We also define  $S_N(x) \triangleq Px^2$ . Then it is easy to see that

$$S_i(x) = \min_{u \in \mathbb{U}} (Qx^2 + S_{i+1}(ax + bu)) \quad \text{for } 0 \leq i < N, \quad (\text{A1})$$

and  $u \in \mathbb{U}$  minimises (A1) iff  $u$  is the first component of some optimal control sequence ( $u_i = u, \dots, u_{N-1}$ ) with initial state  $x$ . We note that if  $x$  is fixed then, in general,  $u$  is not uniquely determined.

We next show by induction that, if  $2 \leq i \leq N$  and the NCL gives the optimal cost for all horizons less or equal to  $i - 1$ , then we have

$$S_{N-i}(x) = Qx^2 + 2Q(b/a)^2 \cdot f_i \left( \frac{a}{b}(ax + bu^*(x)) \right) \quad \text{for all } x \in \mathbb{R}, \quad (\text{A2})$$

where  $u^*(x)$  is the first component of some optimal control sequence  $(u_{N-i} = u^*(x), \dots, u_{N-1})$ .

Next, let  $i = 2$ . Then  $S_{N-2}(x) = Qx^2 + S_{N-1}(ax + bu^*)$ , where  $(u^* = u_{N-2}, u_{N-1})$  is some optimal control sequence with initial state  $x$ .

Define  $\hat{u}(y) \triangleq q_{\mathbb{U}}(-ay/b)$  and  $v \triangleq ax + bu^*$ . Then,

$$S_{N-1}(y) = \min_{u \in \mathbb{U}} Qy^2 + P(ay + bu)^2 = Qy^2 + P(ay + b\hat{u}(y))^2.$$

Thus,

$$\begin{aligned} S_{N-2}(x) &= Qx^2 + Qv^2 + P(av + b\hat{u}(v))^2 = Qx^2 + Q(b/a)^2 \left( (av/b)^2 + \frac{P}{Q}a^2(av/b + \hat{u}(v))^2 \right) \\ &= Qx^2 + Q(b/a)^2 \left( \frac{P}{Q}a^2\varepsilon^2(av/b) + (av/b)^2 \right) = Qx^2 + 2Q(b/a)^2 f_2(av/b). \end{aligned}$$

Let us assume the assertion is true for some  $2 \leq i < N$ , and that the NCL gives the optimal cost for all horizons less or equal to  $i$ . Then  $S_{N-(i+1)}(x) = Qx^2 + S_{N-i}(ax + bu')$ , where  $(u' = u_{N-(i+1)}, \dots, u_{N-1})$  is some optimal control sequence with initial state  $x$ . Let  $w \triangleq ax + bu'$ . By assumption we have

$$S_{N-(i+1)}(x) = Qx^2 + Qw^2 + 2Q(b/a)^2 f_i \left( \frac{a}{b}(aw + bu^*(w)) \right),$$

where  $u^*(w)$  is the first component of some optimal control sequence  $(u^*(w) = s_i, \dots, s_{N-1})$  with initial state  $w$ .

Again, by assumption, we may choose  $u^*(w) = q_{\mathbb{U}}(-aw/b)$ . Thus,

$$\begin{aligned} S_{N-(i+1)}(x) &= Qx^2 + Q(b/a)^2 \left( (aw/b)^2 + 2f_i \left( \frac{a}{b}(aw - bq_{\mathbb{U}}(aw/b)) \right) \right) \\ &= Qx^2 + 2Q(b/a)^2 f_{i+1}(aw/b), \quad \text{since } 2f_i(a\varepsilon(aw/b)) = 2f_{i+1}(aw/b) - (aw/b)^2 \\ &= Qx^2 + 2Q(b/a)^2 f_{i+1} \left( \frac{a}{b}(ax + bu') \right). \end{aligned}$$

This establishes (A2).

We have previously noted that, if  $1 \leq i \leq N$  and  $x \in \mathbb{R}$ , then the two sets

$$M_i \triangleq \left\{ u \in \mathbb{U} \mid \begin{array}{l} u \text{ is the first component of some optimal control sequence} \\ (u_{N-i} = u, \dots, u_{N-1}) \text{ with initial state } x \end{array} \right\}$$

and  $N_i \triangleq \{u \in \mathbb{U} \mid u \text{ minimises } S_{N-i}(x)\}$  coincide.

Now, we have:

The NCL achieves the optimal cost for all horizons less or equal to  $N$

$\Leftrightarrow$  For all  $x \in \mathbb{R}$  and all  $2 \leq i \leq N$  the number  $q_{\mathbb{U}}(-ax/b)$  is the first component

of some optimal control sequence  $(u_{N-i} = q_{\mathbb{U}}(-ax/b), \dots, u_{N-1})$  with initial state  $x$

$\Leftrightarrow q_{\mathbb{U}}(-ax/b) \in M_i$  for all  $x \in \mathbb{R}, 2 \leq i \leq N$

$\Leftrightarrow q_{\mathbb{U}}(-ax/b) \in N_i$  for all  $x \in \mathbb{R}, 2 \leq i \leq N$

$\Leftrightarrow$  For all  $x \in \mathbb{R}$  and all  $2 \leq i \leq N$  the number  $u = q_{\mathbb{U}}(-ax/b)$  is a minimiser of  $f_i(a/b(ax + bu))$  on  $\mathbb{U}$

$\Leftrightarrow$  For all  $z \in \mathbb{R}$  and all  $2 \leq i \leq N$  the number  $q_{\mathbb{U}}(-z)$  is a minimiser of  $f_i(az + au)$  on  $\mathbb{U}$ .

Next, we show  $1) \Rightarrow 2)$ .

Let  $z \in \mathbb{R}, 2 \leq i \leq N$ , and let  $(v_{N-i}, \dots, v_{N-1}) \in \mathbb{U}^i$  be the optimal control sequence with initial state  $bz/a$  and with minimal index in  $\mathbb{U}^i$ . By assumption, we have  $v_{N-i} = q_{\mathbb{U}}(-z)$ , and since the NCL gives the optimal cost for all horizons less than or equal to  $N$ , the number  $q_{\mathbb{U}}(-z)$  minimises  $u \mapsto f_i(az + au)$  on  $\mathbb{U}$ .

Assume that there is some  $u' \in \mathbb{U}$  with  $\text{Index}(u') < \text{Index}(q_{\mathbb{U}}(-z))$  that minimises  $u \mapsto f_i(az + au)$  on  $\mathbb{U}$ . Then, by (A2),  $u'$  minimises  $S_{N-i}(bz/a)$ . Thus  $u'$  is the first component of some optimal control sequence  $(w_{N-i} = u', \dots, w_{N-1})$  with initial state  $bz/a$ . Also, by assumption, we have  $\text{Index}(v_{N-i}, \dots, v_{N-1}) \leq \text{Index}(w_{N-i}, \dots, w_{N-1})$ .

However, this cannot be true, since

$$\text{Index}(v_{N-i}) = \text{Index}(q_{\mathbb{U}}(-z)) > \text{Index}(u') = \text{Index}(w_{N-i}),$$

by virtue of the order in  $\mathbb{U}^i$ .

Finally, we will prove  $2) \Rightarrow 1)$ .

Let  $x \in \mathbb{R}, 2 \leq i \leq N$  and let  $[v_{N-i} \dots v_{N-1}]^T \in \mathbb{U}^i$  be the optimal control sequence with initial state  $x$  and with minimal Index in  $\mathbb{U}^i$ . Let  $[u_{N-i} \dots u_{N-1}]^T$  be the naïve control sequence with initial state  $x$ . From the analysis presented above, it follows that

$$V_i([u_{N-i} \dots u_{N-1}]^T) = V_i([v_{N-i} \dots v_{N-1}]^T),$$

i.e.,  $u_{N-i}$  and  $v_{N-i}$  both minimise  $S_{N-i}(x)$ . Therefore,  $u \mapsto f_i(az + au)$ , where  $z = ax/b$ . Thus  $\text{Index}(u_{N-i}) = \text{Index}(q_{\mathbb{U}}(-z)) \leq \text{Index}(v_{N-i})$  and  $u_{N-i} = v_{N-i}$ .

Now assume there exists some  $\ell, 2 \leq \ell \leq i$ , such that  $u_{N-k} = v_{N-k}$  for all  $k = \ell, \dots, i$ . Then  $(v_{N-(\ell-1)}, \dots, v_{N-1})$  is the optimal control sequence in  $\mathbb{U}^{\ell-1}$  with the same initial state, say  $y$ , as the naïve control sequence  $(u_{N-(\ell-1)}, \dots, u_{N-1})$ , where we have again invoked the order in  $\mathbb{U}^{\ell}$ .

If  $\ell = 2$  then  $v_{N-1} = u_{N-1}$  and thus  $(u_{N-i}, \dots, u_{N-1}) = (v_{N-i}, \dots, v_{N-1})$ .

If  $\ell > 2$  then  $u_{N-(\ell-1)}$  and  $v_{N-(\ell-1)}$  both minimise  $S_{N-(\ell-1)}(y)$  and therefore  $u \mapsto f_{\ell-1}(az + au)$ , where  $z = ay/b$ .

Thus,  $\text{Index}(u_{N-(\ell-1)}) = \text{Index}(q_{\mathbb{U}}(-z)) \leq \text{Index}(v_{N-(\ell-1)})$  and  $u_{N-(\ell-1)} = v_{N-(\ell-1)}$ .

This shows that  $(v_{N-i}, \dots, v_{N-1}) = (u_{N-i}, \dots, u_{N-1})$ . As a consequence, the NCL is optimal for all horizons less or equal to  $N$ .

## Appendix B: Proof of Lemma 5.2

It is easy to see that  $g_n$  is odd by induction, thus  $f_n$  is even.

To show that  $f_n$  is continuous it is sufficient to show that  $\varepsilon^2$  is continuous. The points of discontinuity of  $q_{\mathbb{U}}$  are exactly  $\pm(\gamma_k + \gamma_{k+1})/2$  for  $0 \leq k < M$ . Since  $\varepsilon^2$  is even it suffices to show that  $\varepsilon^2$  is continuous in  $w_k \triangleq (\gamma_k + \gamma_{k+1})/2$  for  $0 \leq k < M$ .

We have

$$\varepsilon^2(w_k) = (w_k - q_{\mathbb{U}}(w_k))^2 = (w_k - \gamma_k)^2 = (\gamma_{k+1} - \gamma_k)^2/4.$$

If  $\delta > 0$  is small enough then

$$\varepsilon^2(w_k + \delta) = (w_k + \delta - \gamma_{k+1})^2.$$

Moreover, this expression converges to  $(\gamma_{k+1} - \gamma_k)^2/4$  as  $\delta$  goes to 0.

Similarly we have  $\varepsilon^2(w_k - \delta) = (w_k - \delta - \gamma_k)^2$ , and this expression converges to  $(\gamma_{k+1} - \gamma_k)^2/4$  as  $\delta$  goes to 0.

### Appendix C: Proof of Lemma 5.3

First note that  $\tilde{g}_{j+1}(v) = \tilde{g}_j(a\tilde{\varepsilon}(v)) = \lim_{h \downarrow 0} g_{j+1}(v+h)$ , and

$$|\tilde{\varepsilon}(v)| = |\tilde{\varepsilon}(-v)|, \quad |\tilde{g}_j(v)| = |\tilde{g}_j(-v)|, \quad \forall j \geq 1, v \in \mathbb{R}.$$

We only prove the lemma for  $a > 0$ . The proof for  $a < 0$  is similar.

1) The assertion is true if  $n = 1$ . Assume it is true for some  $n \geq 1$ . Then

$$\lim_{h \downarrow 0} \frac{g_{n+1}^2(v+h) - g_n^2(v)}{h} = \lim_{h \downarrow 0} \frac{g_n^2(a\varepsilon(v+h)) - g_n^2(a\varepsilon(v))}{h} \quad (C1)$$

First let  $(\gamma_k + \gamma_{k-1})/2 < v < (\gamma_{k+1} + \gamma_k)/2$  for some  $1 \leq k < M$ . Then, for small  $h > 0$  it follows that:

$$\begin{aligned} \varepsilon(v) &= v - \gamma_k = \tilde{\varepsilon}(v) \\ \varepsilon(v+h) &= v + h - \gamma_k. \end{aligned}$$

Thus, the expression (C1) becomes

$$\begin{aligned} \lim_{h \downarrow 0} \frac{g_n^2(a(v - \gamma_k) + ah) - g_n^2(a(v - \gamma_k))}{ah} \cdot a &= a \frac{d^+}{dv} g_n^2(a(v - \gamma_k)) = a \frac{d^+}{dv} g_n^2(a\tilde{\varepsilon}(v)) \\ &= 2a^n \tilde{g}_n(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_{n+1}(v). \end{aligned}$$

Now let  $-(\gamma_{k+1} + \gamma_k)/2 \leq v < -(\gamma_k + \gamma_{k-1})/2$  for some  $1 \leq k < M$ . Then, for small  $h > 0$  it follows that:

$$\begin{aligned} \varepsilon(v) &= v + \gamma_k = \tilde{\varepsilon}(v) \\ \varepsilon(v+h) &= v + h + \gamma_k, \end{aligned}$$

so that now the expression (C1) becomes:

$$\lim_{h \downarrow 0} \frac{g_n^2(a(v + \gamma_k) + ah) - g_n^2(a(v + \gamma_k))}{ah} \cdot a = a \frac{d^+}{dv} g_n^2(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_n(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_{n+1}(v).$$

Now let  $-\gamma_1/2 \leq v < \gamma_1/2$ . Then, for small  $h > 0$ ,  $\varepsilon(v) = v = \tilde{\varepsilon}(v)$  and  $\varepsilon(v+h) = v+h$ . As a

consequence, (C1) becomes:

$$\lim_{h \downarrow 0} \frac{g_n^2(av + ah) - g_n^2(av)}{ah} \cdot a = a \frac{d^+}{dv} g_n^2(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_n(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_{n+1}(v).$$

Now let  $v = (\gamma_{k+1} + \gamma_k)/2$  for some  $0 \leq k < M$ . Then  $\varepsilon(v) = v - \gamma_k = (\gamma_{k+1} - \gamma_k)/2$ ,  $\tilde{\varepsilon}(v) = (\gamma_k - \gamma_{k+1})/2$  and  $\varepsilon(v + h) = v + h - \gamma_{k+1} = (\gamma_k - \gamma_{k+1})/2 + h$ . Thus, expression (C1) becomes

$$\begin{aligned} \lim_{h \downarrow 0} \frac{g_n^2\left(a \frac{\gamma_k - \gamma_{k+1}}{2} + ah\right) - g_n^2\left(a \frac{\gamma_{k+1} - \gamma_k}{2}\right)}{ah} \cdot a &= a \frac{d^+}{dv} g_n^2\left(a \frac{\gamma_k - \gamma_{k+1}}{2}\right) \quad (\text{recall that } g_n^2 \text{ is even}) \\ &= a \frac{d^+}{dv} g_n^2(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_n(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_{n+1}(v). \end{aligned}$$

Now let  $v > (\gamma_M + \gamma_{M-1})/2$ . Then  $\varepsilon(v) = v - \gamma_M = \tilde{\varepsilon}(v)$  and  $\varepsilon(v + h) = v + h - \gamma_M$  for small  $h > 0$ . As a consequence, (C1) becomes

$$\lim_{h \downarrow 0} \frac{g_n^2(a(v - \gamma_M) + ah) - g_n^2(a(v - \gamma_M))}{ah} \cdot a = a \frac{d^+}{dv} g_n^2(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_n(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_{n+1}(v).$$

Now let  $v < -(\gamma_M + \gamma_{M-1})/2$ . Then  $\varepsilon(v) = v + \gamma_M = \tilde{\varepsilon}(v)$  and  $\varepsilon(v + h) = v + h + \gamma_M$  for small  $h > 0$ . Thus, expression (C1) becomes

$$\lim_{h \downarrow 0} \frac{g_n^2(a(v + \gamma_M) + ah) - g_n^2(a(v + \gamma_M))}{ah} \cdot a = a \frac{d^+}{dv} g_n^2(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_n(a\tilde{\varepsilon}(v)) = 2a^n \tilde{g}_{n+1}(v).$$

This proves 1).

2) can be proved in a similar way, and 3) and 4) are direct consequences of 1) and 2).

#### Appendix D: Proof of Corollary 5.4

It is only necessary to show that, if  $\frac{d^+}{dv} f_n \geq 0$  on  $(0, \infty)$ , then  $f_n$  is strictly monotonic increasing on  $[0, \infty)$ .

Let  $0 \leq v < w$ . We show that  $f_n(v) < f_n(w)$ . Since  $f_n$  is continuous at 0, we may assume  $v > 0$ . If we assume  $f_n(v) \geq f_n(w)$  then  $f_n$  is constant on  $[v, w]$  since  $f_n$  is monotonic increasing on  $[v, \infty)$ . Thus, by Lemma 5.3 and since  $a \neq 0$ , it follows that

$$0 = \frac{d^+}{dv} \left( \frac{d^+}{dv} f_n \right) (s) > 0 \quad \text{for all } s \in [v, w].$$

This is a contradiction, thus establishing the result.

#### Appendix E: Proof of Lemma 5.5

1) Since  $|\tilde{g}_j(v)| = |\tilde{g}_j(-v)|$  for all  $j \geq 1$  and all  $v \in \mathbb{R}$ , we have

$$v \in \mathbb{G}_j \Rightarrow \tilde{g}_j(v) \in \mathbb{G}_1 \Rightarrow |\tilde{g}_j(v)| \in \mathbb{G}_1^+ \Rightarrow |\tilde{g}_j(-v)| \in \mathbb{G}_1^+ \Rightarrow \tilde{g}_j(-v) \in \mathbb{G}_1 \Rightarrow -v \in \mathbb{G}_j.$$

2) We first show that  $\mathbb{G}_j^+ \neq \emptyset$ . By 1) it suffices to show that  $\mathbb{G}_j \neq \emptyset$ .

This is true if  $j = 1$ . Assume  $\mathbb{G}_j \neq \emptyset$  for some  $j \geq 1$ . Since  $\mathbb{U}$  is finite the function  $\tilde{\varepsilon} : \mathbb{R} \rightarrow \mathbb{R}$  is onto, and since  $\mathbb{G}_j \neq \emptyset$  there is some  $v' \in \mathbb{R}$  such that  $a\tilde{\varepsilon}(v') \in \mathbb{G}_j$ . Thus  $\tilde{g}_{j+1}(v') = \tilde{g}_j(a\tilde{\varepsilon}(v')) \in \mathbb{G}_1$ , and  $v' \in \mathbb{G}_{j+1}$ .

Finally, we show that  $\mathbb{G}_j$  is finite. Since  $\mathbb{G}_1$  is finite it suffices to show that the set  $\tilde{g}_j^{-1}(\{w\})$  is finite for all  $w \in \mathbb{R}$  and all  $j \geq 1$ . (Recall the definition of  $\mathbb{G}_j$ .) To this end we show, by induction, that  $\lim_{|w| \rightarrow \infty} |\tilde{g}_j(w)| = \infty$  for all  $j \geq 1$ .

The claim is true if  $j = 1$ . Assume it is true for some  $j \geq 1$ . Then

$$\lim_{|w| \rightarrow \infty} |\tilde{g}_{j+1}(w)| = \lim_{|w| \rightarrow \infty} |a\tilde{g}_j(a\tilde{\varepsilon}(w))| = \infty$$

by assumption and by the fact that  $\lim_{|w| \rightarrow \infty} |a\tilde{\varepsilon}(w)| = \infty$ .

## Appendix F: Proof of Theorem 5.6

If  $n \geq 2$  we let

$$\mathbb{J}_n \triangleq \left\{ v > 0 \mid \frac{d^+}{dv} f_n \text{ is not continuous in } v \right\} \quad (\text{F1})$$

be the set of all positive points of discontinuity of  $\frac{d^+}{dv} f_n$ . It is easy to see that  $\mathbb{J}_n$  is finite.

Since  $\frac{d^+}{dv} f_n$  is piecewise monotonic increasing and continuous from the right by Lemma 5.3 we see that

$$\inf_{v \geq 0} \frac{d^+}{dv} f_n(v) = \min_{v \geq 0} \frac{d^+}{dv} f_n(v) = \min_{v \in \mathbb{J}_n \cup \{0\}} \frac{d^+}{dv} f_n(v). \quad (\text{F2})$$

By Corollary 5.4 and Equation (F2), we need to show that  $\min_{v \in \mathbb{J}_n \cup \{0\}} \frac{d^+}{dv} f_n(v) \geq 0$ .

If  $n \geq 1$  let

$$\mathbb{V}_n \triangleq \{v \in \mathbb{R} \mid \tilde{g}_n \text{ is not continuous in } v\}.$$

First, we show that

$$\mathbb{V}_{n+1} \subset \bigcup_{j=1}^n \mathbb{G}_j \quad \text{for all } n \geq 1. \quad (\text{F3})$$

This is true if  $n = 1$ , since  $\mathbb{V}_2 = \{v \in \mathbb{R} \mid \tilde{\varepsilon} \text{ is not continuous in } v\} = \mathbb{G}_1$ .

If it is true for some  $n \geq 1$ , then we have

$$\begin{aligned} \mathbb{V}_{n+2} &= \{v \in \mathbb{R} \mid \tilde{g}_{n+2}(v) = \tilde{g}_{n+1}(a\tilde{\varepsilon}(v)) \text{ is not continuous in } v\} \\ &\subset (\mathbb{V}_2 \cup \{v \in \mathbb{R} \mid a\tilde{\varepsilon}(v) \in \mathbb{V}_{n+1}\}) \subset \left( \mathbb{V}_2 \cup \bigcup_{j=1}^n \{v \in \mathbb{R} \mid a\tilde{\varepsilon}(v) \in \mathbb{G}_j\} \right) \quad \text{by assumption} \\ &= \mathbb{V}_2 \cup \bigcup_{j=1}^n \{v \in \mathbb{R} \mid \tilde{g}_j(a\tilde{\varepsilon}(v)) \in \mathbb{V}_2\} = \mathbb{V}_2 \cup \bigcup_{j=2}^{n+1} \{v \in \mathbb{R} \mid \tilde{g}_j(v) \in \mathbb{V}_2\} = \mathbb{G}_1 \cup \bigcup_{j=2}^{n+1} \mathbb{G}_j = \bigcup_{j=1}^{n+1} \mathbb{G}_j. \end{aligned}$$

This establishes (F3).

Thus, if  $n \geq 2$ , then it follows that

$$\mathbb{J}_n \subset \left( (0, \infty) \cap \bigcup_{k=2}^n \mathbb{V}_k \right) \subset \bigcup_{j=1}^{n-1} \mathbb{G}_j^+.$$

Thus,

$$\min_{v \in \mathbb{J}_n} \frac{d^+}{dv} f_n(v) \geq \min_{1 \leq j < n} \min_{v \in \mathbb{G}_j^+} \frac{d^+}{dv} f_n(v) \geq 0,$$

so that

$$\min_{v \in \mathbb{J}_n \cup \{0\}} \frac{d^+}{dv} f_n(v) \geq 0,$$

since  $\frac{d^+}{dv} f_n(0) = 0$ .

*Remark F1* Actually, from (F2), it is clear why we did not consider left-hand derivatives  $\frac{d^-}{dv}$ , since then we would end up with the more complicated expression

$$\inf_{v \geq 0} \frac{d^-}{dv} f_n(v) = \min_{v \in \mathbb{J}_n \cup \{0\}} \lim_{h \downarrow 0} \frac{d^-}{dv} f_n(v+h),$$

where  $\mathbb{J}_n^- \triangleq \left\{ v > 0 \mid \frac{d^-}{dv} f_n \text{ is not continuous in } v \right\}$ . This is why we use  $\tilde{g}_j$  and  $\tilde{\varepsilon}$  instead of the original functions  $g_j$  and  $\varepsilon$ , defined in (14) and (15).

## Appendix G: Proof of Lemma 6.2

1) We have  $\tilde{g}_2 \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) = a\tilde{\varepsilon} \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) = a(\gamma_{k-1} - \gamma_k)/2$ .

Now assume the assertion is true for some  $n \geq 2$ . Then

$$\tilde{g}_{n+1} \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) = a\tilde{\varepsilon} \left( \tilde{g}_n \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) \right) = a\tilde{\varepsilon} \left( a^{n-1} \frac{\gamma_{k-1} - \gamma_k}{2} \right) = a^n \frac{\gamma_{k-1} - \gamma_k}{2}.$$

2) First we note that since  $\tilde{\varepsilon}$  is continuous from the right, piecewise monotonic increasing with non-negative points of discontinuity  $(\gamma_k + \gamma_{k-1})/2$  ( $1 \leq k \leq M$ ). Since  $\tilde{\varepsilon}(0) = \tilde{\varepsilon}(\gamma_M) = 0$ , we have that, if  $0 \leq v \leq \gamma_M$ , then:

$$\tilde{\varepsilon}(v) \geq \min \left( \{0\} \cup \left\{ \tilde{\varepsilon} \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) \mid 1 \leq k \leq M \right\} \right) = \min \left( \{0\} \cup \left\{ \frac{\gamma_{k-1} - \gamma_k}{2} \mid 1 \leq k \leq M \right\} \right) > -\frac{\gamma_1}{2a},$$

and

$$\tilde{\varepsilon}(v) \leq \max \left( \{0\} \cup \left\{ \varepsilon \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) \mid 1 \leq k \leq M \right\} \right) = \max \left( \{0\} \cup \left\{ \frac{\gamma_k - \gamma_{k-1}}{2} \mid 1 \leq k \leq M \right\} \right) < \frac{\gamma_1}{2a}.$$

Since  $|\tilde{\varepsilon}(v)| = |\tilde{\varepsilon}(-v)|$ , we obtain

$$\max_{|v| \leq \gamma_M} |\tilde{\varepsilon}(v)| < \frac{\gamma_1}{2a}.$$

Next we utilise induction to show  $\mathbb{G}_{j+1} \subset ((-\infty, -\gamma_M) \cup (\gamma_M, \infty))$ . If  $v \in \mathbb{G}_2$ , then  $a\tilde{\varepsilon}(v) \in \mathbb{G}_1$ , thus  $|a\tilde{\varepsilon}(v)| \geq \gamma_1/2$ , and therefore  $|v| > \gamma_M$ .

Assume the assertion is true for some  $j \geq 1$ . If  $v \in \mathbb{G}_{j+2}$  then  $a\tilde{\varepsilon}(v) \in \mathbb{G}_{j+1}$ , thus  $|a\tilde{\varepsilon}(v)| > \gamma_M > \gamma_1/2$ , and therefore  $|v| > \gamma_M$ .

Finally if  $v \in \mathbb{G}_{j+1}^+$  then  $\tilde{g}_j(a\tilde{\varepsilon}(v)) \in \mathbb{G}_1$  and  $v > \gamma_M$ , thus  $a\tilde{\varepsilon}(v) \in \mathbb{G}_j$  and  $\tilde{\varepsilon}(v) > 0$ . As a consequence,  $a\tilde{\varepsilon}(v) \in \mathbb{G}_j^+$ .

3) We have

$$\begin{aligned} \min_{v \in \mathbb{G}_{j+1}^+} \frac{d^+}{dv} f_{n+1}(v) &= \min_{v \in \mathbb{G}_{j+1}^+} \left( \frac{P}{Q} a^n \tilde{g}_n(a\tilde{\varepsilon}(v)) + \sum_{k=2}^n a^{k-1} \tilde{g}_{k-1}(a\tilde{\varepsilon}(v)) + v \right) \\ &\geq \min_{v \in \mathbb{G}_j^+} \left( \frac{P}{Q} a^n \tilde{g}_n(v) + \sum_{k=2}^n a^{k-1} \tilde{g}_{k-1}(v) \right) \quad \text{by 2)} \\ &= a \cdot \min_{v \in \mathbb{G}_j^+} \left( \frac{P}{Q} a^{n-1} \tilde{g}_n(v) + \sum_{k=1}^{n-1} a^{k-1} \tilde{g}_k(v) \right) = a \cdot \min_{v \in \mathbb{G}_j^+} \frac{d^+}{dv} f_n(v). \end{aligned}$$

4) The assertion follows from

$$\begin{aligned} \min_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_n(v) &= \min_{1 \leq k \leq M} \left( \frac{P}{Q} a^{n-1} \tilde{g}_n \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) + \frac{\gamma_k + \gamma_{k-1}}{2} + \sum_{j=2}^{n-1} a^{j-1} \tilde{g}_j \left( \frac{\gamma_k + \gamma_{k-1}}{2} \right) \right) \\ &= \min_{1 \leq k \leq M} \left( \frac{P}{Q} a^{2n-2} \frac{\gamma_{k-1} - \gamma_k}{2} + \frac{\gamma_k + \gamma_{k-1}}{2} + \sum_{j=2}^{n-1} a^{2j-2} \frac{\gamma_{k-1} - \gamma_k}{2} \right) \\ &= \min_{1 \leq k \leq M} \left( \frac{\gamma_k + \gamma_{k-1}}{2} + \frac{\gamma_{k-1} - \gamma_k}{2} \left( \frac{P}{Q} a^{2n-2} + \frac{a^2 - a^{2n-2}}{1 - a^2} \right) \right) \\ &= \frac{1}{2} \min_{1 \leq k \leq M} \left( \gamma_{k-1} \left( 1 + \frac{P}{Q} a^{2n-2} + \frac{a^2 - a^{2n-2}}{1 - a^2} \right) + \gamma_k \left( 1 - \frac{P}{Q} a^{2n-2} - \frac{a^2 - a^{2n-2}}{1 - a^2} \right) \right). \end{aligned}$$

## Appendix H: Proof of Corollary 6.3

Let  $N \geq 2$ . Since  $0 < a \leq 1/\sqrt{2}$ , the sequence  $\left( \frac{1-2a^2+a^{2n-2}}{a^{2n-2}(1-a^2)} \right)_{n=2}^\infty$  is monotonic increasing. Thus,

$$\frac{P}{Q} \leq \frac{1}{a^2} \leq \frac{1-2a^2+a^{2n-2}}{a^{2n-2}(1-a^2)}, \quad \forall 2 \leq n \leq N.$$

By Lemma 6.2 we have:

$$\begin{aligned} \min_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_n(v) &\geq 0, \\ \min_{v \in \mathbb{G}_j^+} \frac{d^+}{dv} f_n(v) &\geq a^{j-1} \min_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_{n-j+1}(v), \end{aligned} \quad \forall 2 \leq n \leq N, \quad 1 \leq j < n$$

Thus,

$$\min_{1 < n \leq N} \min_{1 \leq j < n} \min_{v \in \mathbb{G}_j^+} \frac{d^+}{dv} f_n(v) \geq \min_{1 < n \leq N} \min_{1 \leq j < n} a^{j-1} \min_{v \in \mathbb{G}_1^+} \frac{d^+}{dv} f_{n-j+1}(v) \geq 0.$$

The assertion follows from Corollary 5.7.

**Appendix I: Proof of Corollary 6.4**

Since  $1/\sqrt{2} < a < 1$  the sequence  $\left(\frac{1-2a^2+a^{2n-2}}{a^{2n-2}(1-a^2)}\right)_{n=2}^{\infty}$  is monotonic decreasing, and since  $\hat{N} < 1 + \frac{\ln(2a^2-1)}{2\ln a}$  we have  $1 - 2a^2 + a^{2\hat{N}-2} > 0$ .

From

$$\frac{P}{Q} \leq \frac{1 - 2a^2 + a^{2\hat{N}-2}}{a^{2\hat{N}-2}(1-a^2)}$$

we obtain

$$\frac{P}{Q} \leq \frac{1 - 2a^2 + a^{2n-2}}{a^{2n-2}(1-a^2)}, \quad \forall 2 \leq n \leq \hat{N}.$$

The rest of the proof parallels that of Corollary 6.3.