Digital Signal Processing 20 (2010) 1511–1525

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Analysis and design of the receding horizon constrained optimization for class-D amplifier driving signals

Jwu-Sheng Hu [1], Keng-Yuan Chen *

Department of Electrical and Control Engineering, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:
Available online 27 March 2010

Keywords:
Amplifiers
Digital filter
Modulation
Quantization
Receding horizon quadratic optimal control

ABSTRACT

This paper reports the analysis and design method of receding finite horizon constrained optimization approaches to generate driving signals for digital amplifiers. The performance of a digital amplifier that is based on power MOSFETs highly depends on the modulator that modulates the reference signal into binary sequences. The concept of receding finite horizon constrained optimization modulator is to generate the binary output in the sense that the power of filtered error is minimized. The stability analysis of this nonlinear feedback system is derived in detail and the method to estimate the bound of filtered error of the high dimensional quantization is proposed. A second order system with horizon length one and two is designed to produce a 1.5-bit constrained output. Simulation results demonstrate the accuracy of the analysis.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Class D amplifiers are more efficient than class A/B amplifiers and have drawn much attention in recent years [1]. To drive the power stage, the input multiple-bit data must be converted into the driving signal (binary sequence). The binary sequence is typically referred to as the quantized 1-bit signal and has a higher sampling rate than the original source signal which is called over-sampling. The binary sequence can be generated using the digital PWM (DPWM) procedure. To reduce the total harmonic distortion and increase the signal/noise ratio, research based on sigma–delta modulation [2,3], DSP techniques [4,5], feedback control [6], and interpolation methods [7,8] have been adopted. Another design issue in the use of DPWM is the high internal clock frequency. For example, the internal clock frequency of the counter must be 2.89 GHz to yield a 44.1 KHz DPWM with 16-bit resolution [8].

Sigma–delta modulation is an effective means of generating a binary sequence with noise shaping. The output feedback structure can be analyzed by extending control theories of relay feedback [3,9] or sliding mode control [10–12]. The design and related performance analysis are presented [13,14]. Other methods such as genetic algorithm [15], semi-infinite programming [16] and nonlinear circle map [17] are proposed to design a high order stable system in recent years. The generation of the driving signal can be regarded from a broader perspective, and not necessarily following the conventional sigma–delta modulation settings. This broader perspective particularly applies in digital implementation, in which the computing structure is quite flexible.

Feedback quantization using receding horizon linear quadratic control with finite input constraint has recently been reported [18–20]. The approach offers a systematic way of designing the quantizer (or modulator) with various performance measures. However, the stability of this nonlinear feedback system is an important issue in designing the modulator. Though

* Corresponding author. Fax: +886 0305715998.
E-mail addresses: jshu@cn.nctu.edu.tw (J.-S. Hu), bettery.ece94g@nctu.edu.tw (K.-Y. Chen).
[1] Member, IEEE.

1051-2004/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.dsp.2010.03.005

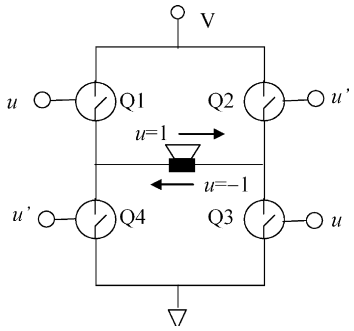


Fig. 1.

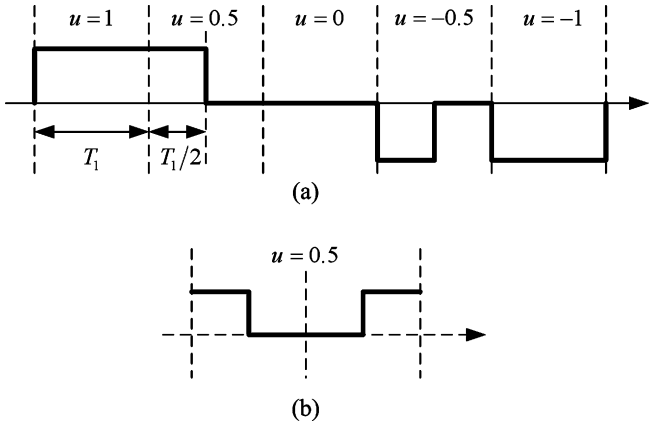


Fig. 2.

proof of the stability in the general case of optimal quantization has been presented elsewhere [20], for class-D amplifier applications, a more detailed procedure must be derived to compute the allowable input level. This work proposes a design method by exploring the structure of the quantizer. The stability of the general horizon system is analyzed without restricting the type of reference input. Simulation results show that the method is quite effective in generating the control sequences.

2. Basic concept

Fig. 1 shows a typical power stage (H-bridge) in a class D amplifier. The control signal u controls the current direction of the load. Suppose that u is updated every T_1 second and that the value of u also contains the net duration of the current. The simplest case is $u = +1$ or -1, meaning that the current flows in either the positive or the negative direction within T_1. If a higher clock rate is available, say T_1/n_H where n_H is an integer, there could be $2^{n_H} + 1$ cases of the duration of the current. For example, Fig. 2(a) depicts five cases for which $n_H = 2$. u can clearly be represented by a 2.5-bit variable. The constrained optimization control [20] can be applied to design a quantization system using over-sampling. This section concisely introduces related work in [20].

2.1. Problem statement

Suppose u belongs to a finite set of scalars U. The purpose of constrained optimization control is to obtain the signal u for a digital input r and a filter W, such that the cost function V is minimized. The cost function is defined as the power of filtered error [20],

$$V \triangleq \frac{1}{2\pi}\int_0^{2\pi} \left|E\left(e^{jw}\right)\right|^2 d\omega \triangleq \frac{1}{2\pi}\int_0^{2\pi} \left|W\left(e^{jw}\right)\left(R\left(e^{jw}\right) - U\left(e^{jw}\right)\right)\right|^2 d\omega \tag{2.1}$$

where $R(e^{jw})$ and $U(e^{jw})$ are the discrete time Fourier-transform of r and u, respectively and $W(e^{jw})$ denotes the frequency response of the filter W. Therefore the modulator quantizes the signal r in a manner that the power of the difference between r and u is minimized in the frequency band that the magnitude response of $W(e^{jw})$ is large. The expression in (2.1) can be translated into the time domain using Parseval's Theorem:

$$V = \sum_{l=0}^{\infty} \big(e(l)\big)^2 \tag{2.2}$$

where $e(l)$ are samples of filtered error. Assume that W is a linear time-invariant filter which can be written as

$$W(z) = D + C(zI - A)^{-1}B$$

The filtered error is expressed as the state-space form:

$$\begin{aligned} x(k+1) &= Ax(k) + B\big(r(k) - u(k)\big) \\ e(k) &= Cx(k) + D\big(r(k) - u(k)\big) \end{aligned} \tag{2.3}$$

where $A \in R^{n\times n}$, $B \in R^{n\times 1}$, $C \in R^{1\times n}$, $D \in R^{1\times 1}$ and $x \in R^n$ are the state vectors of dimension n, i.e. the order of the filter W. Eq. (2.2) is further simplified for practical use with finite decision number N:

$$V_N \triangleq \sum_{m=k}^{k+N-1} \big(e(m)\big)^2 \tag{2.4}$$

From (2.3) and (2.4), it can be seen that V_N is a function of N constrained values u which can be grouped into the vector

$$\mathrm{u}(k) \triangleq \big[u(k)\ u(k+1)\ \cdots\ u(k+N-1)\big]^T \tag{2.5}$$

2.2. Solution and moving horizon

Finding the optimal constrained solution corresponds to finding $\mathbf{u}(k) \in U^{(N)}$, such that V_N is minimized. Assume the constrained optimization control produces a b-bit resolution output u for input r, then the quantization candidate U is a set with K_r elements,

$$U = \{1, 1-\Delta, 1-2\Delta, \ldots, -1+\Delta, -1\} \subset R \tag{2.6}$$

where $\Delta = 2/(2^b - 1)$, $K_r = 2^b$ if $b \in Z$, and $\Delta = 2/2^{b_1}$, $K_r = 2^{b_1} + 1$ if $b = b_1 + 0.5$ and $b_1 \in Z$. Furthermore, $U^{(N)}$ is a set with $(K_r)^N$ vectors,

$$U^{(N)} = \{v_1, v_2, \ldots, v_{(K_r)^N}\} \subset R^N$$

where elements of v_i, $i = 1, 2, \ldots, (K_r)^N$, belong to U. The optimal solution to the cost function V_N can be derived as stated in [20].

Theorem 1. *For $U^{(N)} = \{v_1, v_2, \ldots, v_{(K_r)^N}\} \subset R^N$, the sequence $\mathbf{u}^*(k)$ of (2.5) which minimize (2.4) under (2.3) is given by*:

$$\mathbf{u}^*(k) = \Psi^{-1} q_{\tilde{U}^{(N)}}\big(\Psi \mathbf{u}_u(k)\big)$$

where $\mathbf{u}_u(k) = \Psi^{-1}\Gamma x(k) + \mathbf{r}(k)$ is the unconstrained optimal control force, and

$$\Psi = \begin{bmatrix} h_\delta & 0 & \cdots & 0 \\ h_{\delta+1} & h_\delta & \ddots & \vdots \\ \vdots & \ddots & h_\delta & 0 \\ h_{\delta+N-1} & \cdots & h_{\delta+1} & h_\delta \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} CA^\delta \\ CA^{\delta+1} \\ \vdots \\ CA^{\delta+N-1} \end{bmatrix}, \qquad \mathbf{r}(k) = \begin{bmatrix} r(k) \\ r(k+1) \\ \vdots \\ r(k+N-1) \end{bmatrix}$$

$$h_i = \begin{cases} D & \text{when } i = 0 \\ CA^{i-1}B & \text{when } i > 0 \end{cases}$$

δ is the relative degree of shaping filter $W(z)$.

The nonlinear function $q_{\tilde{U}^{(N)}}(.)$ is the nearest neighbor vector quantizer, which is described in Definition 1 below. The image of this mapping is the set,

$$\tilde{U}^{(N)} = \{\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{(K_r)^N}\} \subset R^N \quad \text{with } \tilde{v}_i = \Psi v_i,\ v_i \in U^{(N)}$$

Definition 1 *(Nearest neighbor vector quantizer).* (See [18].) For a set of vectors $\tilde{U}^{(N)} = \{\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{(K_r)^N}\} \subset R^N$, the nearest neighbor quantizer is defined as a mapping $q_{\tilde{U}^{(N)}} : R^N \to \tilde{U}^{(N)}$ which assigns to each vector $c \in R^N$ the closest element of $\tilde{U}^{(N)}$ (as measured by the Euclidean norm), i.e., $q_{\tilde{U}^{(N)}}(c) = \tilde{v}_i \in \tilde{U}^{(N)}$ if and only if c satisfies:

$$\|c - \tilde{v}_i\| \leqslant \|c - \tilde{v}_j\|, \quad \forall \tilde{v}_i \neq \tilde{v}_j,\ \tilde{v}_j \in \tilde{U}^{(N)}$$

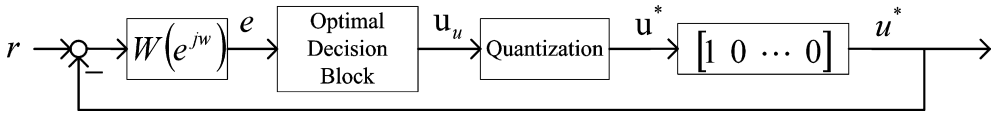


Fig. 3.

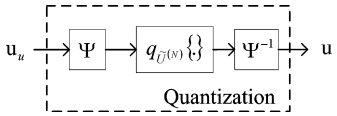


Fig. 4.

Therefore, the nearest neighbor vector quantizer partitions the space R^N into $(K_r)^N$ partitions considering the nearest vector distance to $\tilde{U}^{(N)}$. For future reference, the half space $H(\tilde{v}_i, \tilde{v}_j)$ is defined as the region that the distance to $\tilde{v}_i$ is shorter than or equal to $\tilde{v}_j$,

$$H(\tilde{v}_i, \tilde{v}_j) \triangleq \left\{ \mathbf{y} \in R^N \;\middle|\; \|\mathbf{y} - \tilde{v}_i\| \leqslant \|\mathbf{y} - \tilde{v}_j\| \right\}, \quad j \neq i \tag{2.7}$$

and the partition induced by the constrained vector $\tilde{v}_i$, $P(\tilde{v}_i)$ is the intersection of $(K_r)^N - 1$ half spaces which is written as,

$$P(\tilde{v}_i) = \bigcap_{j \in I \setminus \{i\}} H(\tilde{v}_i, \tilde{v}_j), \quad I = \left\{1, 2, \ldots, (K_r)^N\right\} \tag{2.8}$$

Although, one might think of choosing N as large as the length of the complete input data stream, the implementation complexity increases with the horizon length N. Therefore, using the technique of moving horizon with a small value of N has been proposed in [20]. Accordingly, at time $t = k$, only the first element of the optimizing sequence is used as the output of the converter:

$$u(k) = u^*(k) = [1\ 0\ \cdots\ 0]\Psi^{-1} q_{\tilde{U}^{(N)}}\left(\Psi \mathbf{u}_u(k)\right) \tag{2.9}$$

Also, $u^*(k)$ updates the states according to (2.3), and the new states are used to minimize the cost function V_N, yielding $u(k+1)$. The horizon moves (slides) forward as time increases.

The block diagram of overall system is depicted in Fig. 3 and optimal decision block is responsible for producing unconstrained optimal control force $\mathbf{u}_u$. The operation of quantization block is shown in Fig. 4 according to Theorem 1.

From (2.1), the Fourier transform of the optimized output $u^*(k)$ is expressed as (2.10), to elucidate the effect of noise shaping:

$$U^* = R - W^{-1}E \tag{2.10}$$

where R and E are the Fourier transforms of r and e, respectively. Clearly, the output contains two parts of the signal; the first term is the input signal R, and the second term is the quantization error, which is the residual error E, filtered by the inverse of the system filter, W^{-1}. Consequently, W^{-1} is responsible for spectrally shaping the noise.

3. Bounding the error signal and system states

The stability of (2.3) must be analyzed under the control law (2.9), to determine the design parameters. The techniques to analyze the stability for the Σ–Δ modulator have been reported [19,21]. However, due to the different quantization scheme, these techniques cannot be applied to the optimal quantizer derived in this work (see (2.9)). Further analysis into the quantization scheme is also required to derive the stability boundary. In the following, the analysis is composed of two parts. First, the condition to bound the filtered error and a method to find out the error bound is proposed. Second, conditions for system stability are derived with the input amplitude limitation.

3.1. Bounding the error signal

First, define $\mathbf{d}(k)$ as the input to the nearest neighbor vector quantizer $q_{\tilde{U}^{(N)}}(.)$:

$$\mathbf{d}(k) = \begin{bmatrix} d_1(k) \\ d_2(k) \\ \vdots \\ d_N(k) \end{bmatrix} \triangleq \Psi \mathbf{u}_u(k) \tag{3.1}$$

Proposition 1 states the condition to bound the signal e.

Proposition 1. *For the system* (2.3) *under the control law* (2.9)*, the filtered error e is bounded if the first element of* $\mathbf{d}(k)$ *is bounded, i.e. e is bounded if there exists* $\bar{d}_1 < \infty$ *such that,*

$$\left\|d_1(k)\right\|_\infty \leqslant \bar{d}_1$$

where $\|d_1(k)\|_\infty$ *denotes the peak absolute value of signal* d_1.

Proof. The proof is given in Appendix A. □

Because the half space (see (2.7)) is convex and the intersection of two convex sets is also convex [21], the partitions of the nearest neighbor vector quantizer defined in (2.8) are all convex. Further, the boundaries of the partitions are all linear hyperplanes. Therefore, the problem of finding the upper bound of signal e in N-dimensional quantizer can be formulated as the N-dimensional linear programming problem. For i-th partition $P(\tilde{v}_i)$, the error bound of this region, $\bar{e}_i$, is obtained by solving,

$$\begin{aligned}&\text{Maximize } f(\mathbf{d}) = \left|d_1 - [1\ 0\ \cdots\ 0]\tilde{v}_i\right| \\ &\quad \text{subject to } \|d_1\|_\infty \leqslant \bar{d}_1 \text{ and } \left\{\mathbf{d} \;\middle|\; \mathbf{d} \in P(\tilde{v}_i) \cap S\right\} \\ &\quad \text{where } P(\tilde{v}_i) \text{ is the } i\text{-th partition defined in (2.8) and } S \text{ is the possible input space of } q_{\tilde{U}^{(N)}}(.) \text{ induced by} \\ &\quad \|d_1\|_\infty \leqslant \bar{d}_1\end{aligned}$$

As a result, the upper bound of e, $\bar{e}$, for the system is the maximum error bound selected from $(K_r)^N$ partitions and is written as,

$$\bar{e} = \max \bar{e}_i, \quad i \in \left\{1, 2, \ldots, (K_r)^N\right\} \tag{3.2}$$

Notably, different value of $\bar{d}_1$ will result in different $\bar{e}$. Therefore, we write $\bar{e}$ as a function of $\bar{d}_1$ for subsequent analysis, i.e.

$$\bar{e} = g(\bar{d}_1) \tag{3.3}$$

3.2. Stability and input magnitude limitation

The amplitude of input is limited due to maintaining system stability. The following proposition explains the property.

Proposition 2. *Given the initial conditions,* $x(0)$, $r(0)$ *and* $u(0)$ *where* $u(0)$ *satisfies the control law* (2.9)*, then the system* (2.3) *is stable if there exists a value* $0 < \bar{d}_1 < \infty$ *such that*

$$\begin{cases} h_\delta^{-1}\left(\bar{d}_1 - \left\|P_1(z)\right\|_\infty g(\bar{d}_1)\right) > 0 \\ \left|d_1(0)\right| = \left|CA^\delta x(0) + h_\delta r(0)\right| \leqslant \bar{d}_1 \end{cases} \quad \textit{and}$$

(i) *zeros of the shaping filter* $W(z)$ *are all inside the unit circle of the complex plane*;

(ii) $$\left\|r(k)\right\|_\infty \leqslant h_\delta^{-1}\left(\bar{d}_1 - \left\|P_1(z)\right\|_\infty g(\bar{d}_1)\right) \tag{3.4}$$

where $g(\bar{d}_1)$ *is defined in* (3.3) *and* $P_1(z) = CA^\delta(zI - A + Bh_\delta^{-1}CA^\delta)^{-1}Bh_\delta^{-1}z^\delta$.

Proof. The proof is given in Appendix B. □

Notably, there may be infinite many selections of $\bar{d}_1$ for one system and different $\bar{d}_1$ will result in different input amplitude limitation. Therefore, we prefer choosing $\bar{d}_1$ that results in the maximum value of the bound in (3.4) so as to maximize the input amplitude. Furthermore, employing more levels in quantization will decrease the bound of error and then increase the stable input amplitude. It has been shown that the sigma–delta modulator is a special case of the constrained optimization control when $N = 1$ [20]. Therefore, Propositions 1–2 could be also applied to the analysis of the sigma–delta modulator.

4. Design example and simulation results

This section describes an example of a second order constrained optimization control system for the application of audio-band full digital amplifier. The bandwidth of the input audio signal r is 24 KHz and the system converts it into a 1.5-bit signal u at a sampling rate of 6.144 MHz (OSR = 128).

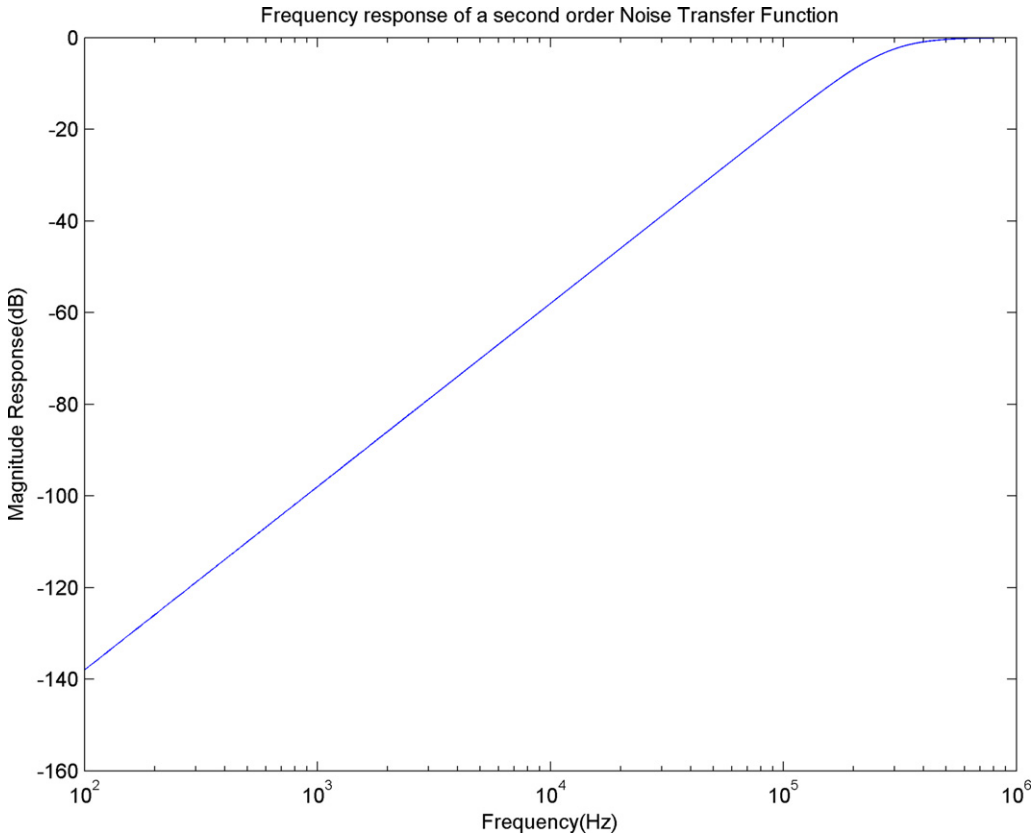


Fig. 5.

4.1. Overall system

The shaping filter $W(z)$ is designed that all zeros are stable and all poles are on the unit circle $|z| = 1$ to have better noise attenuation within the bandwidth of input signal:

$$W(z) = \frac{1.22 - 1.96z^{-1} + 0.82z^{-2}}{1 - 2z^{-1} + z^{-2}}$$

Therefore, the frequency response of the noise transfer function W^{-1} is shown in Fig. 5 implying that the attenuation of signal e is -40 dB or lower within the bandwidth 0–20 kHz.

Once the shaping filter is selected, we obtained that

$$\left\| P_1(z) \right\|_\infty = 1.36$$

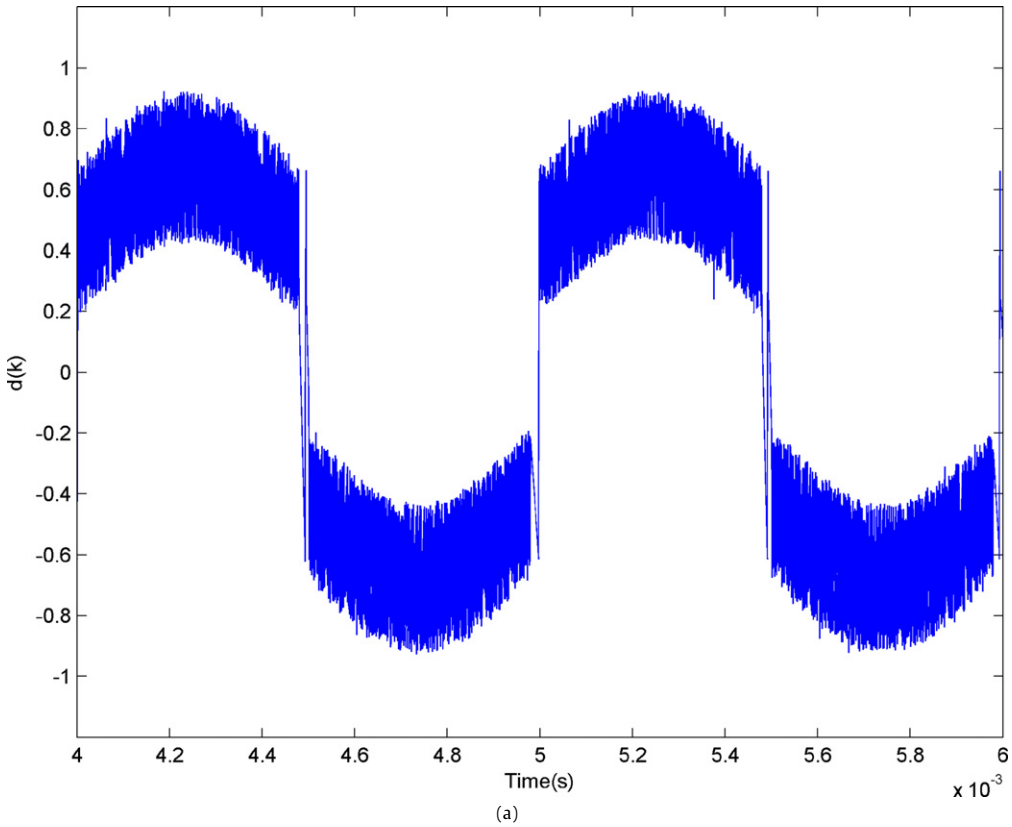
The bound of filtered error are calculated using linear programming by limiting $\|d_1\|_\infty \leqslant \bar{d}_1 = 2.41$. Therefore, $\|e\|_\infty \leqslant 1.18$ for $N = 1$ and 2. The stable input level can be estimated by (3.3),

$$\|r\|_\infty \leqslant h_\delta^{-1}\left(\bar{d}_1 - \left\| P_1(z) \right\|_\infty \left\| e(k) \right\|_\infty\right) = (1.22)^{-1}(2.41 - 1.36 \times 1.18) = 0.66$$

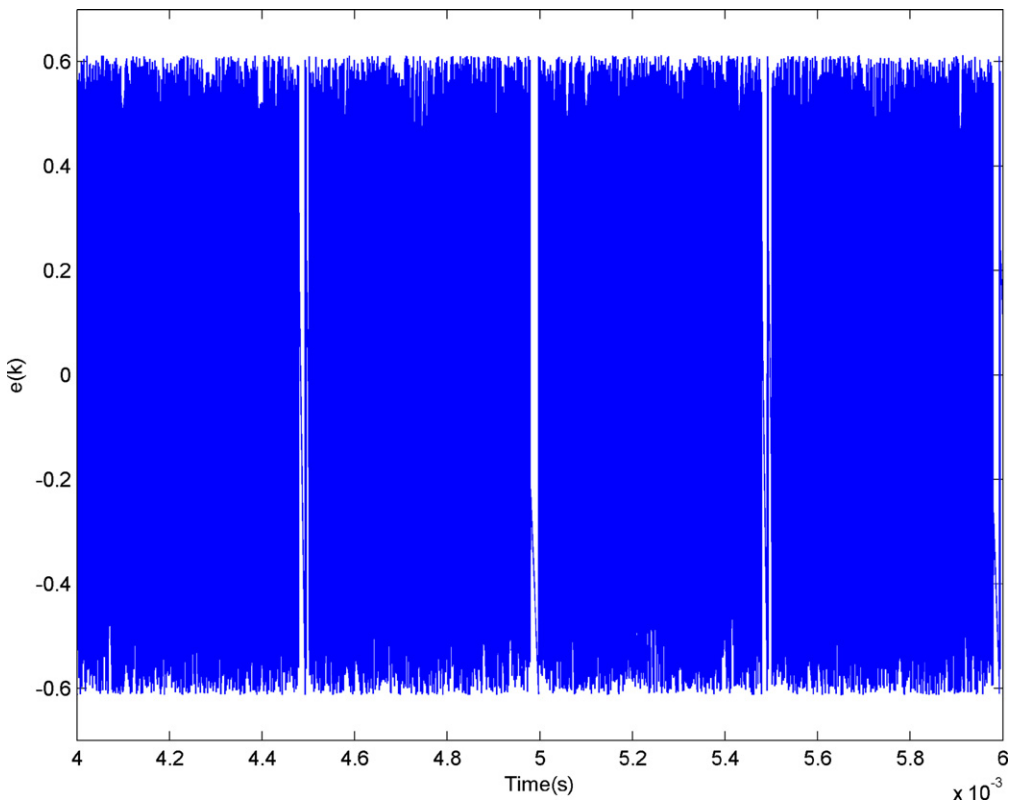
Notably, for the multi-level power stage, an input range larger than 0.66 is obtained. For example, stable input level is 0.83 for 2.5-bit quantization. Nevertheless, in the application of 2-level power stage, the designed system converts the reference signal into a 1.5-bit sequence to reduce the total harmonic distortion and increase the signal-to-noise ratio.

4.2. Simulation results

A 1 kHz sine wave with a normalized amplitude of 0.66, sampled at 48 kHz, is applied to the system as an input. Fig. 6 plots the waveforms of $d_1(k)$ and $e(k)$, respectively, in $N = 1$ quantization. Boundaries are $\|d_1(k)\|_\infty \leqslant 0.93$ and $\|e(k)\|_\infty \leqslant 0.61$ which are within the theoretical upper bound: 2.41 and 1.18. The power of the error is calculated, $\frac{1}{T_f}\sum_{k=0}^{T_f} e^2(k) = 0.23$ where T_f is the number of samples during the time $t = 0$–100 ms. Fig. 7 displays output spectrum obtained by taking the DFT (Discrete Fourier Transform) of the binary sequence in the period $t = 0$–100 ms. Total Harmonic Distortion (THD), Total


d(k)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
4
4.2
4.4
4.6
4.8
5
5.2
5.4
5.6
5.8
6
Time(s)
x 10^-3

(a)


e(k)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
4
4.2
4.4
4.6
4.8
5
5.2
5.4
5.6
5.8
6
Time(s)
x 10^-3

(b)

Fig. 6.

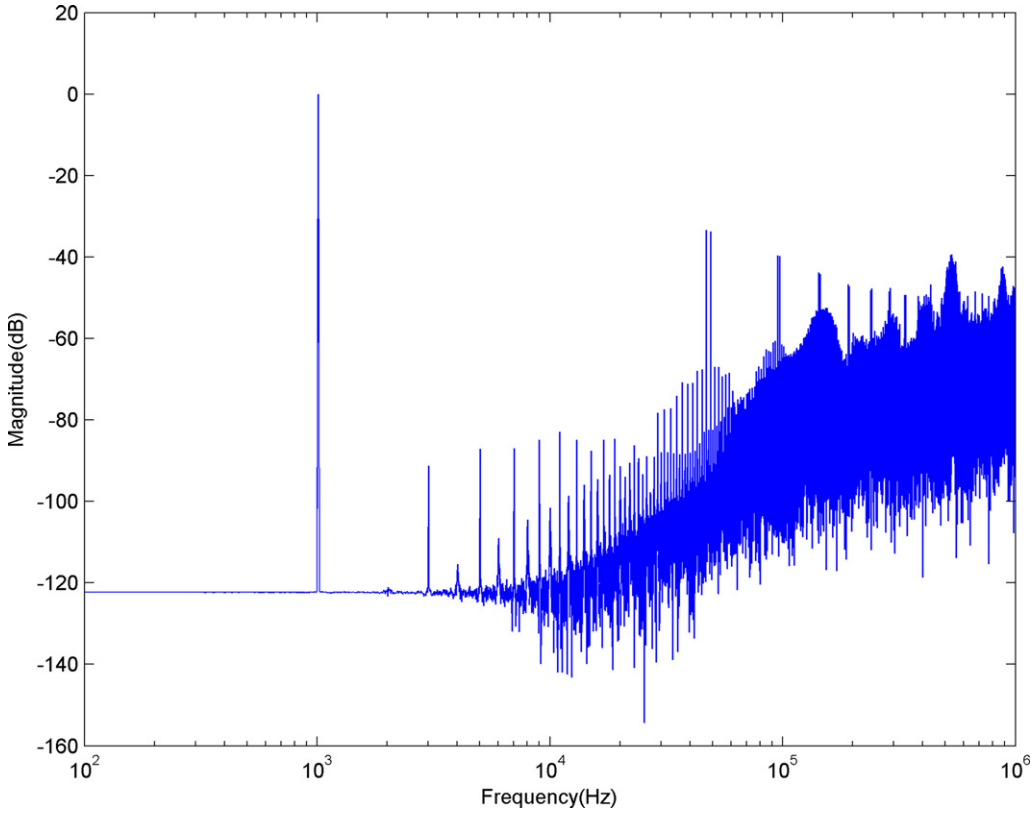


Fig. 7.

Harmonic Distortion Plus Noise (THD + N) and Signal-to-Noise Ratio (SNR) are calculated based on [22]. The results are THD = 0.032%, THD + N = 0.037% and SNR = 101 dB.

For horizon length two ($N = 2$), signals $d_1(k)$, $d_2(k)$ and $e(k)$ are as plotted in Figs. 8 and 9. The figures reveal that $\|d_1(k)\|_\infty \leqslant 1.04$, $\|e(k)\|_\infty \leqslant 0.76$ and the power of the error is 0.21. Fig. 9(b) presents the spectrum of the switching output during the period $t = 0$–100 ms; the results are THD = 0.022%, THD + N = 0.026% and SNR = 104 dB implying the better results comparing to the system with $N = 1$ (which is essentially a sigma–delta modulator).

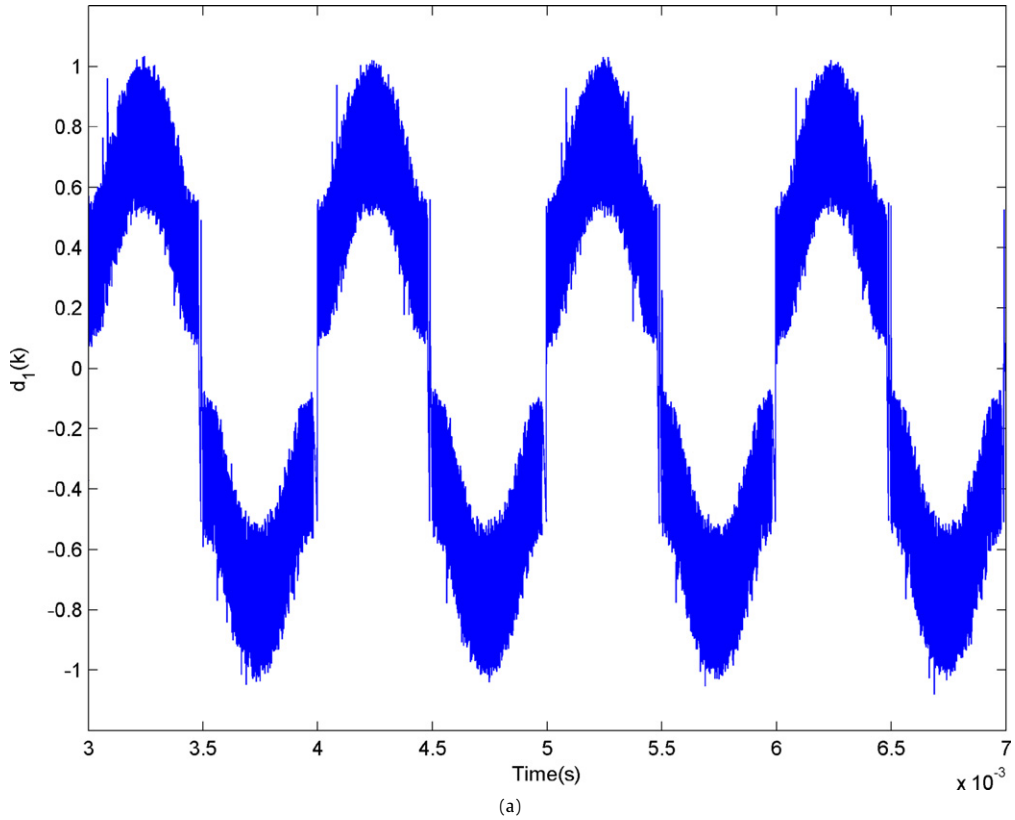
Figs. 10–12 depict the results of power of the error, THD + N and SNR for different input frequency. It is shown that for $N = 2$ ($N = 1$) system, power of the modulation error is below 0.21 (0.23); THD + N is below 0.065% (0.152%) and SNR is above 101 dB (98 dB), yielding a good performance in the audio band. Further, the constrained optimization system with $N = 2$ has better performance than $N = 1$, indicating that by looking one step ahead, the constrained optimization system performs better than sigma–delta modulator at the filter order of 2.

5. Conclusion

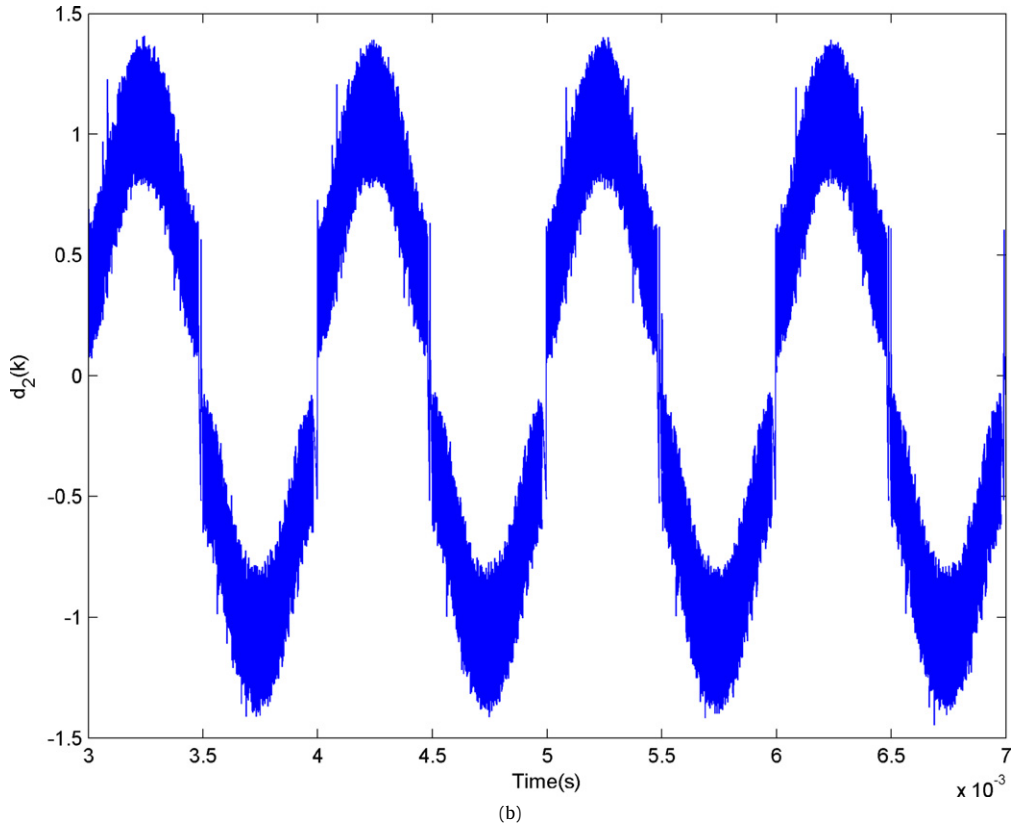
The design of finite horizon constrained optimization system for a full digital amplifier is presented. The stability analysis for general horizon length N is conducted to determine the bound of the input reference signal, based on a detailed derivation of the optimal quantizer. The bound of filtered error is obtained by solving the linear programming problem. Simulation results demonstrate the theoretical bound on the error signal. The function of optimal quantization with horizon length one ($N = 1$) is exactly the sigma–delta modulator. Therefore, the performance can be further improved without raising the operating frequency (i.e. without increasing the OSR) when the optimal quantization is further modified to horizon two ($N = 2$). In real practice, it is necessary to add dead-time to the switching signals and this will influence the performance. This topic will be investigated in the future research.

Acknowledgments

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC94-2218-E009064 and MOE ATU Program under the account number 95W803E.

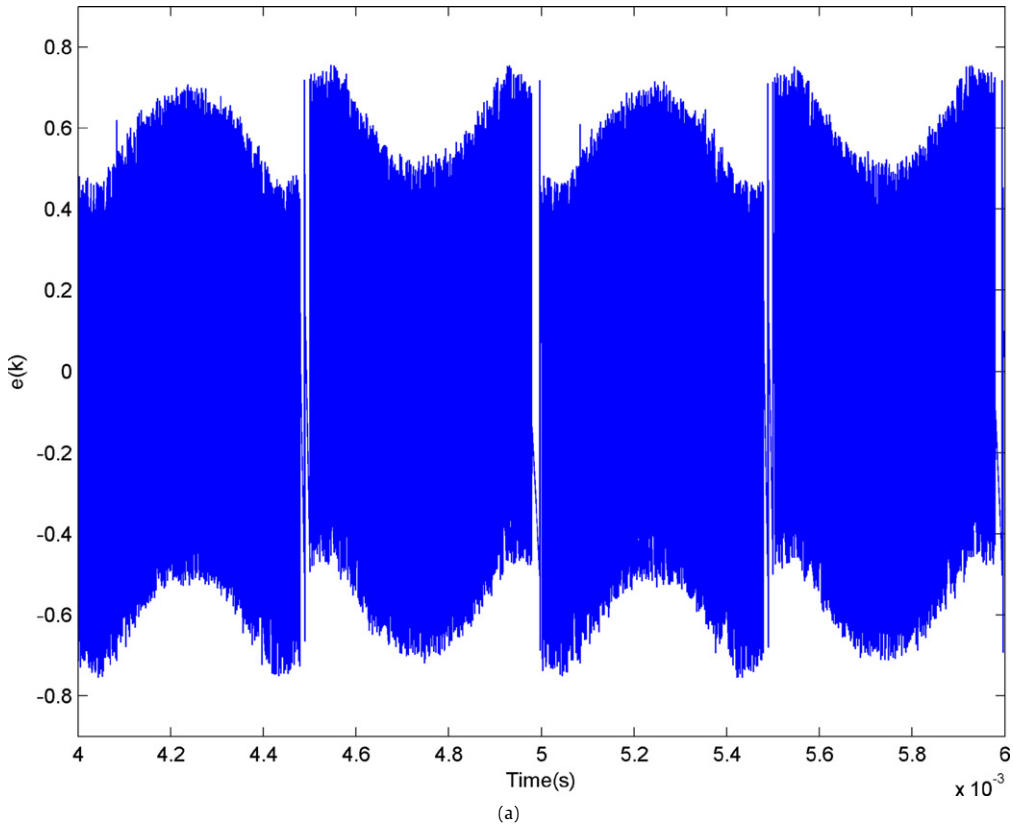


(a)

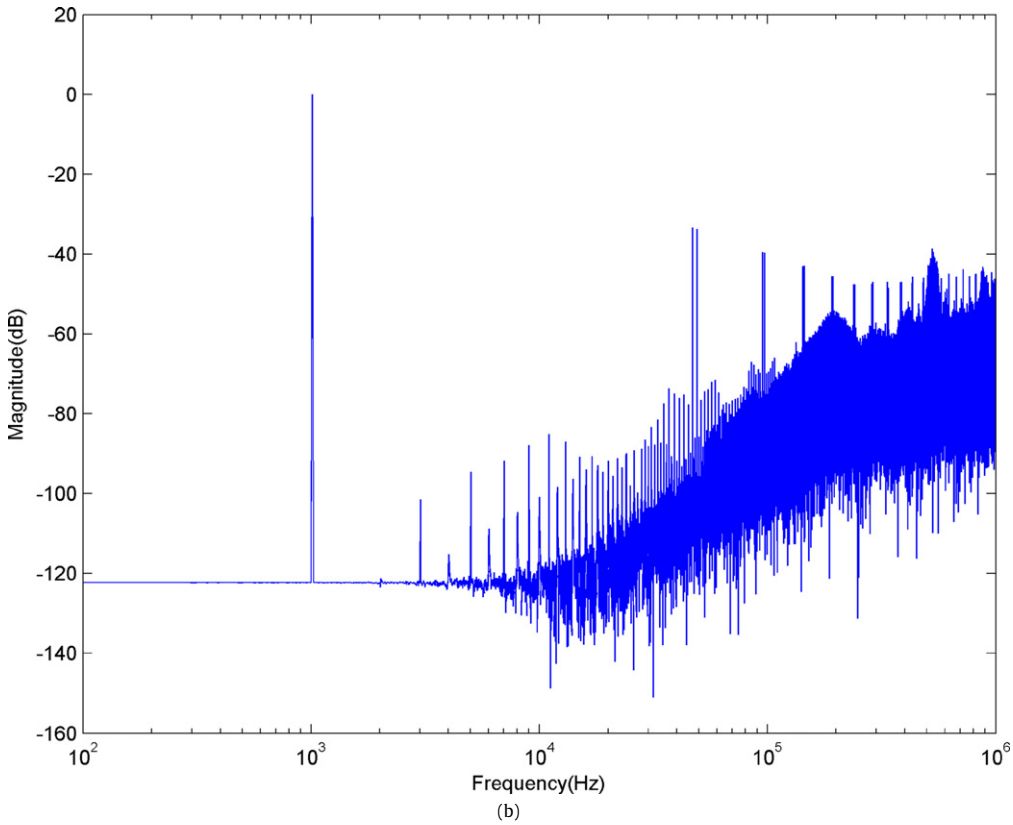


(b)

Fig. 8.


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
e(k)
4
4.2
4.4
4.6
4.8
5
5.2
5.4
5.6
5.8
6
Time(s)
x 10^-3

(a)


20
0
-20
-40
-60
-80
-100
-120
-140
-160
Magnitude(dB)
10^2
10^3
10^4
10^5
10^6
Frequency(Hz)

(b)

Fig. 9.

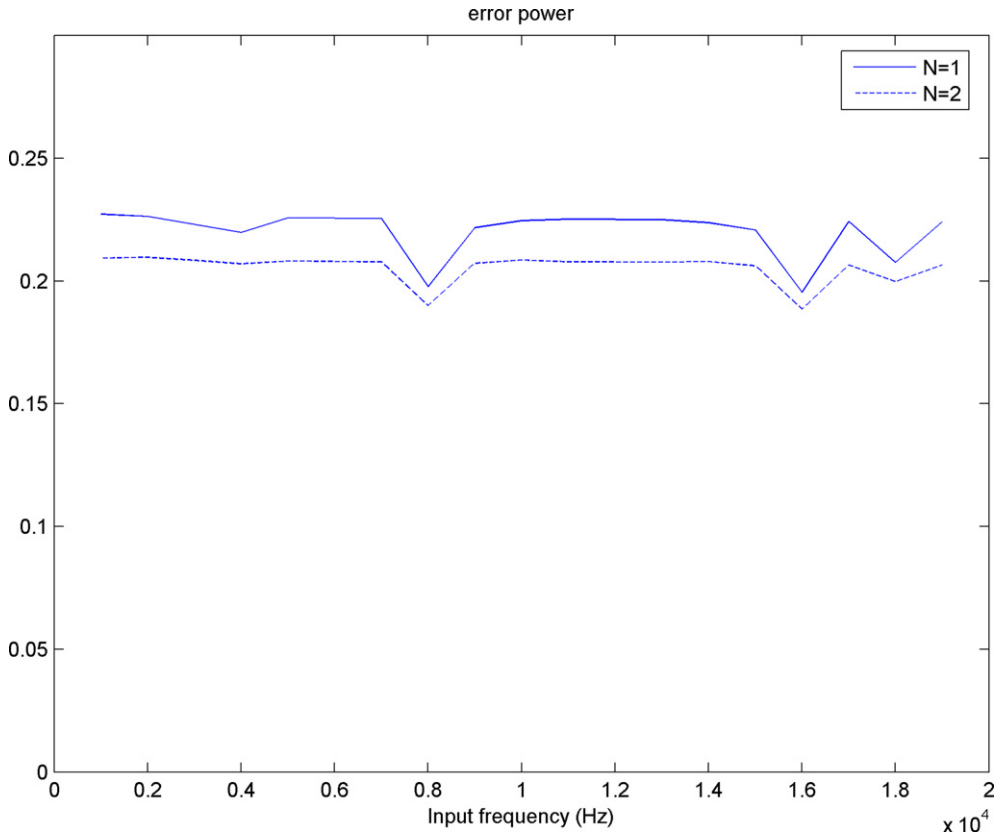

error power
N=1
N=2
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Input frequency (Hz)
x 10^4

Fig. 10.

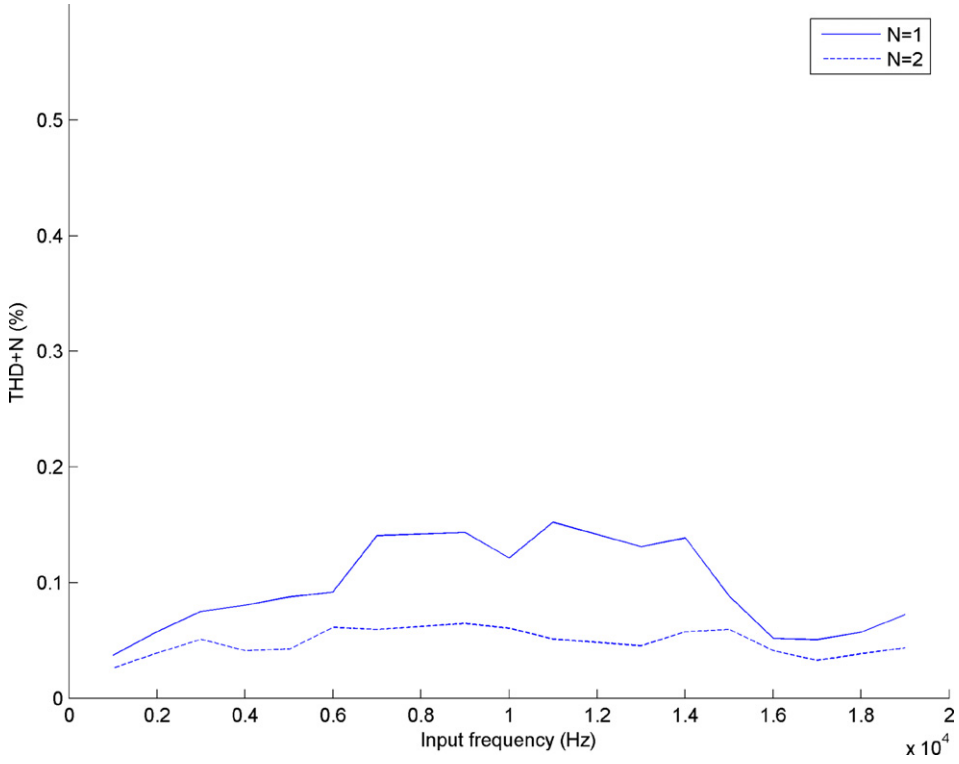

N=1
N=2
THD+N (%)
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Input frequency (Hz)
x 10^4

Fig. 11.

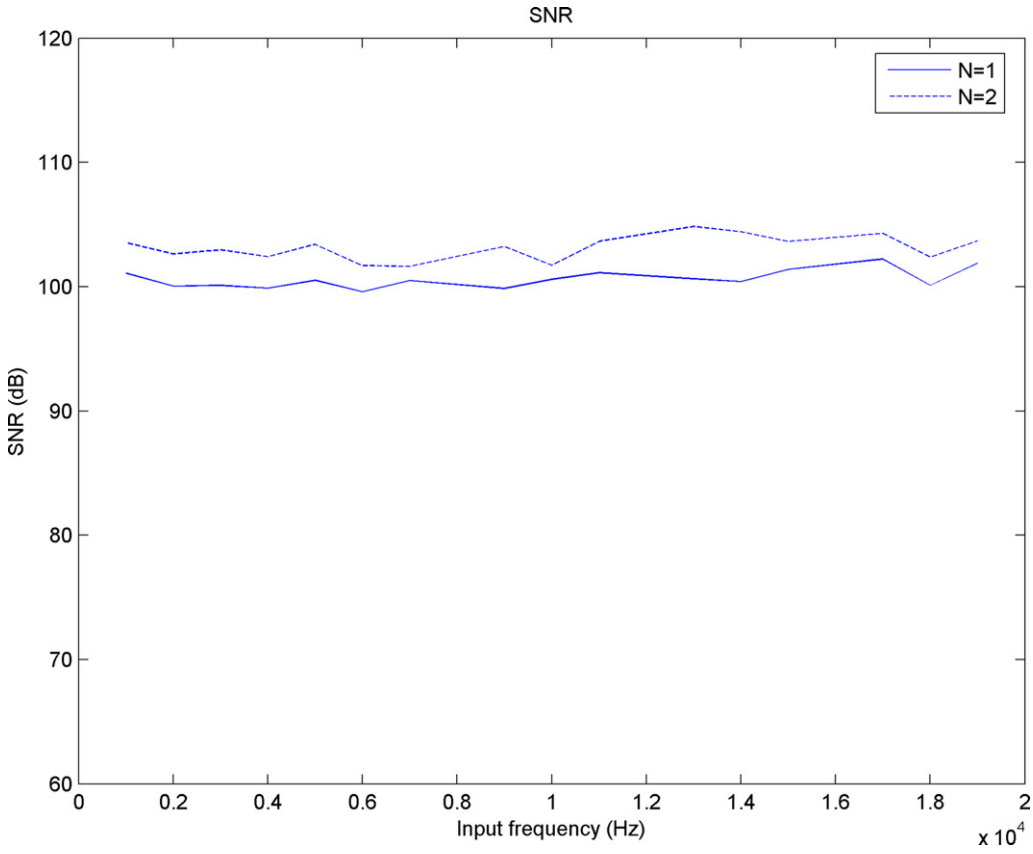


Fig. 12.

Appendix A. Proof of Proposition 1

It is composed of two parts. First we show that the input space of $q_{\tilde{U}^{(N)}}(.)$ is bounded if $d_1(k)$ is bounded. Then it is shown that the filtered error e is bounded when the input space of $q_{\tilde{U}^{(N)}}(.)$ is bounded.

From (2.3) and Theorem 1,

$$\begin{aligned} d_1(k+1) &= CA^{\delta}x(k+1) + h_{\delta}r(k+1) \\ &= \left[CA^{\delta+1}x(k) + h_{\delta+1}r(k) + h_{\delta}r(k+1)\right] - h_{\delta+1}u(k) \\ &= d_2(k) - [h_{\delta+1}\ 0\ \cdots\ 0]\,\mathbf{u}(k) \end{aligned}$$

Therefore,

$$d_2(k) = d_1(k+1) + [h_{\delta+1}\ 0\ \cdots\ 0]\mathbf{u}(k) \tag{A.1}$$

Because $\|d_1(k)\|_{\infty} \leqslant \bar{d}_1$, the upper (lower) bound of $d_1(k+1)$ is $\bar{d}_1$ $(-\bar{d}_1)$. Further, from (A.1), the bound of $d_2(k)$ is

$$\bar{d}_1 + [h_{\delta+1}\ 0\ \cdots\ 0]\mathbf{u}(k) \leqslant d_2(k) \leqslant -\bar{d}_1 + [h_{\delta+1}\ 0\ \cdots\ 0]\mathbf{u}(k) \tag{A.2}$$

If input of $q_{\tilde{U}^{(N)}}(.)$ locates at the partition that

$$[1\ 0\ \cdots\ 0]\,\mathbf{u}(k) = u(k) = 1$$

then the bound of $d_2(k)$ is

$$-\bar{d}_1 + h_{\delta+1} \leqslant d_2(k) \leqslant \bar{d}_1 + h_{\delta+1}$$

at that partition. An example of 1.5-bit constrained optimal control for $N = 2$ is shown in Fig. 13. The marks '$\times$' in Fig. 13(a) are the constrained set $\tilde{U}^{(2)}$. According to nearest vector distance, nine partitions are obtained (see Fig. 13(b)). The signal $u(k)$ for each region is shown below basing on Theorem 1:

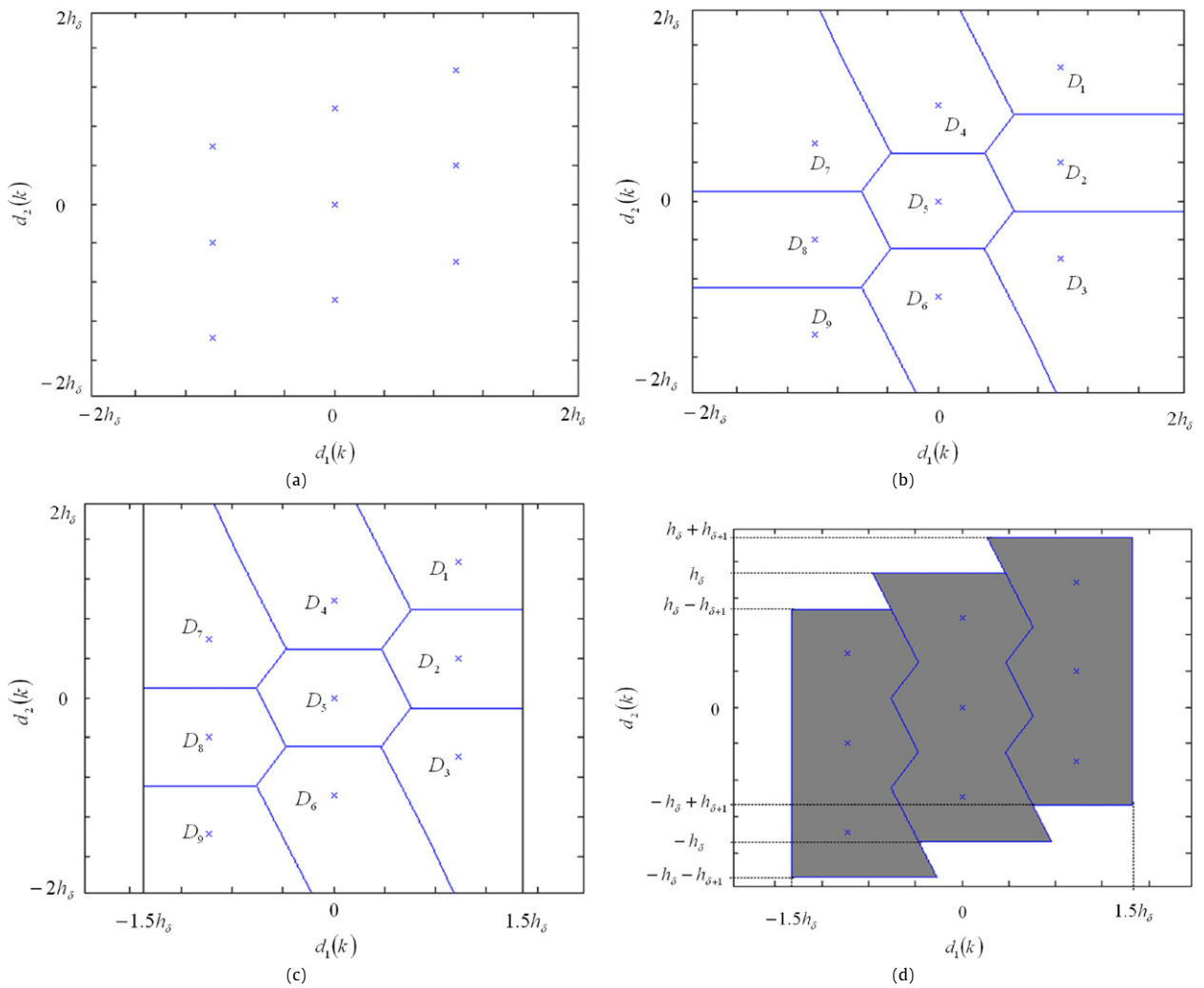


Fig. 13.

$$u(k)=\begin{cases}1, & \text{if } \mathbf{d}(k)=\begin{bmatrix} d_1(k) & d_2(k)\end{bmatrix}^T \in D_1, D_2, D_3\\ 0, & \text{if } \mathbf{d}(k)=\begin{bmatrix} d_1(k) & d_2(k)\end{bmatrix}^T \in D_4, D_5, D_6\\ -1, & \text{if } \mathbf{d}(k)=\begin{bmatrix} d_1(k) & d_2(k)\end{bmatrix}^T \in D_7, D_8, D_9\end{cases} \tag{A.3}$$

If $\|d_1(k)\|_\infty \leqslant \bar{d}_1 = 1.5h_\delta$ is satisfied (Fig. 13(c)), the bound of $d_2(k)$ is obtained according to (A.2) and (A.3):

$$\begin{cases} -1.5h_\delta + h_{\delta+1} \leqslant d_2(k) \leqslant 1.5h_\delta + h_{\delta+1}, & \text{for region } D_1, D_2, D_3\\ -1.5h_\delta \leqslant d_2(k) \leqslant 1.5h_\delta, & \text{for region } D_4, D_5, D_6\\ -1.5h_\delta - h_{\delta+1} \leqslant d_2(k) \leqslant 1.5h_\delta - h_{\delta+1}, & \text{for region } D_7, D_8, D_9\end{cases}$$

Therefore, the shaded region in Fig. 13(d) is the possible input region of $q_{\tilde{U}^{(2)}}(.)$ for a 1.5-bit constrained optimal control system.

To consider the general case N, the following equations are obtained from (2.3) and Theorem 1,

$$\begin{aligned} d_1(k+2) &= CA^{\delta}x(k+2) + h_\delta r(k+2)\\ &= CA^{\delta+1}x(k+1) + h_{\delta+1}r(k+1) + h_\delta r(k+2) - h_{\delta+2}u(k+1)\\ &= \left[CA^{\delta+2}x(k) + h_{\delta+2}r(k) + h_{\delta+1}r(k+1) + h_\delta r(k+2)\right] - h_{\delta+2}u(k) - h_{\delta+1}u(k+1)\\ &= d_3(k) - [\, h_{\delta+2}\ h_{\delta+1}\ 0\ \cdots\ 0\,]\mathbf{u}(k)\end{aligned}$$

$$\vdots$$

$$d_1(k+N-1) = d_N(k) - [\, h_{\delta+N-1}\ h_{\delta+N-2}\ \cdots\ h_{\delta+1}\ 0\,]\mathbf{u}(k)$$

Therefore, the vector $\mathbf{d}(k)$ is expressed as,

$$\mathbf{d}(k)=\begin{bmatrix} d_1(k)\\ d_2(k)\\ \vdots\\ d_N(k)\end{bmatrix}=\begin{bmatrix} d_1(k)\\ d_1(k+1)\\ \vdots\\ d_1(k+N-1)\end{bmatrix}+\begin{bmatrix} 0 & 0 & \cdots & 0\\ h_{\delta+1} & 0 & & 0\\ \vdots & \ddots & \ddots & \vdots\\ h_{\delta+N-1} & \cdots & h_{\delta+1} & 0\end{bmatrix}\mathbf{u}(k) \tag{A.4}$$

Because the elements of $\mathbf{u}$ belong to constrained set U (see (2.6)), (A.4) shows that if d_1 is bounded, elements of $\mathbf{d}$ are all bounded indicating that possible input space of $q_{\tilde{U}^{(N)}}(.)$ is a bounded subspace of R^N.

Now consider the signal e. From Theorem 1 and system (2.3),

$$\begin{aligned} e(k+\delta)&=CA^{\delta}x(k)+h_{\delta}\big(r(k)-u(k)\big)\\ &=\big[CA^{\delta}x(k)+h_{\delta}r(k)\big]-h_{\delta}[\,1\;\;0\;\cdots\;0\,]\Psi^{-1}q_{\tilde{U}^{(N)}}\big(\Gamma x(k)+\Psi\vec{r}(k)\big)\\ &=[\,1\;\;0\;\cdots\;0\,]\big[\mathbf{d}(k)-q_{\tilde{U}^{(N)}}\big(\mathbf{d}(k)\big)\big] \end{aligned} \tag{A.5}$$

As a result, the filtered error e is the first element of the difference between input and output of $q_{\tilde{U}^{(N)}}(.)$ implying that e is bounded under the condition $\|d_1(k)\|_{\infty}\leqslant\bar{d}_1$ because the input space and output of $q_{\tilde{U}^{(N)}}(.)$ are all bounded.

Appendix B. Proof of Proposition 2

First, it is shown that error signal e is bounded under the input constraint (ii). From the second equation of (2.3), the constrained signal $u(k)$ is written in the form,

$$u(k)=r(k)-h_{\delta}^{-1}e(k+\delta)+h_{\delta}^{-1}CA^{\delta}x(k) \tag{B.1}$$

Substituting (B.1) into the first equation of (2.3), yields the expression for the system,

$$x(k+1)=\big(A-Bh_{\delta}^{-1}CA^{\delta}\big)x(k)+Bh_{\delta}^{-1}e(k+\delta) \tag{B.2}$$

From the equation above, it can be shown that z-transform of signal $CA^{\delta}x$ is,

$$Z\big\{CA^{\delta}x\big\}=P_1(z)E(z) \tag{B.3}$$

Further, using (B.3), $\|d_1(k)\|$ is written as in [12]

$$\big\|d_1(k)\big\|=\big\|CA^{\delta}x(k)+h_{\delta}r(k)\big\|\leqslant\big\|P_1(z)\big\|_{\infty}g(\bar{d}_1)+h_{\delta}\big\|r(k)\big\|_{\infty} \tag{B.4}$$

To bound the signal e, the following inequality is imposed to satisfy $\|d_1(k)\|_{\infty}\leqslant\bar{d}_1$ (see Proposition 1),

$$\big\|P_1(z)\big\|_{\infty}g(\bar{d}_1)+h_{\delta}\big\|r(k)\big\|_{\infty}\leqslant\bar{d}_1$$

or alternatively,

$$\big\|r(k)\big\|_{\infty}\leqslant h_{\delta}^{-1}\big(\bar{d}_1-\big\|P_1(z)\big\|_{\infty}g(\bar{d}_1)\big) \tag{B.5}$$

Since signals $e(0)\leqslant g(\bar{d}_1)$ and $d_1(0)\leqslant\bar{d}_1$ are bounded initially, system is stable at the beginning. The signal d_1 will be bounded at all times if (B.5) is satisfied [12] indicating that the bound of e will be maintained under the input constraint stated in (B.5).

Further, since e is bounded under the condition (ii), the states are bounded if the system (B.2) is stable, indicating that the eigenvalues of the matrix $(A-Bh_{\delta}^{-1}CA^{\delta})$ must lie inside the unit circle. If the shaping filter $W(z)$ is designed without pole zero cancellation, then the eigenvalues of $(A-Bh_{\delta}^{-1}CA^{\delta})$ are exactly the zeros of $W(z)$. Therefore the system states are all bounded under the conditions (i) and (ii).

References

[1] B. Putzeys, Digital audio's final frontier, IEEE Spectrum 40 (2003) 34–41.
[2] E. Gaalaas, B.Y. Liu, N. Nishimura, R. Adams, K. Sweetland, Integrated stereo $\Delta\Sigma$ class D amplifier, IEEE J. Solid-State Circuits 40 (2005) 2388–2397.
[3] S.H. Yu, J.S. Hu, Stability and performance of single-bit sigma–delta modulators operated in quasi-sliding mode, Circuits Systems Signal Process. 25 (2006) 571–590.
[4] R.E. Hiorns, M.B. Sandler, Power digital to analogue conversion using pulse width modulation and digital signal processing, IEE Circuits Devices Syst. Proc. G 140 (1993) 329–338.
[5] S. Logan, M.O.J. Hawksford, Linearization of class D output stages for high-performance audio power amplifiers, in: 2nd Int. Conf. on Advanced A–D and D–A Conversion Techniques and Their Applications, 1994, pp. 136–141.
[6] J. Park, C.G. Kim, J. Jeong, B.H. Cho, A novel controller for switching audio power amplifier with digital input, in: IEEE 33rd Annual Power Electronics Specialists Conference, vol. 1, 2002, pp. 39–44.

[7] H. Li, B.H. Gwee, J.S. Chang, A digital class D amplifier design embodying a novel sampling process and pulse generator, in: IEEE Int. Symp. on Circuits and Systems (ISCAS'01), vol. 4, 2001, pp. 826–829.
[8] K.M. Smedley, Digital-PWM audio power amplifiers with noise and ripple shaping, in: 25th Annual IEEE Power Electronics Specialists Conference, vol. 1, 1994, pp. 566–570.
[9] A. Ucar, Bounding integrator output of sigma–delta modulator by time delay feedback control, IEE Proc. Circuits Dev. Syst. 150 (2003) 31–37.
[10] J.S. Hu, S.H. Yu, Analysis and design of 1-bit noise-shaping quantizer using variable structure control approach, in: American Control Conference, vol. 2, 2004, pp. 1271–1276.
[11] J.F. Silva, PWM audio power amplifiers: sigma delta versus sliding mode control, in: IEEE Int. Conf. on Electronics, Circuits and Systems, vol. 1, 1998, pp. 359–362.
[12] S.H. Yu, Analysis and design of single-bit sigma–delta modulators using the theory of sliding modes, IEEE Trans. Control Syst. Technol. 14 (2006).
[13] J.S. Hu, K.Y. Chen, Implementation of a full digital amplifier using feedback quantization, in: American Control Conference, 2006, pp. 2759–2764.
[14] S.H. Yu, Noise-shaping coding through bounding the frequency-weighted reconstruction error, IEEE Trans. Circuits Syst. II, Exp. Briefs 53 (2006) 67–71.
[15] Yongtao Wang, K. Muhammad, K. Roy, Design of sigma–delta modulators with arbitrary transfer functions, IEEE Trans. Signal Process. 55 (2007) 677–683.
[16] C.Y.-F. Ho, B.W.-K. Ling, J.D. Reiss, Y.-Q. Liu, K.-L. Teo, Design of interpolative sigma–delta modulators via semi-infinite programming, IEEE Trans. Signal Process. 54 (2006) 4047–4051.
[17] Amin Z. Sadik, Zahir M. Hussain, Xinghuo Yu, Peter O'Shea, An approach for stability analysis of a single-bit high-order digital sigma–delta modulator, Digital Signal Process. 17 (2007) 1040–1054.
[18] D.E. Quevedo, J.A.D. Doná, G.C. Goodwin, Receding horizon linear quadratic control with finite input constraint sets, in: IFAC 15th Triennial World Congress, 2002.
[19] D.E. Quevedo, G.C. Goodwin, Audio quantization from a receding horizon control perspective, in: Proc. American Control Conference, vol. 5, 2003, pp. 4131–4136.
[20] D.E. Quevedo, G.C. Goodwin, Multi-step optimal analog-to-digital conversion, IEEE Trans. Circuits Syst. I, Reg. Pap. 52 (2005) 503–505.
[21] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, Spatial tessellations, pp. 23–24.
[22] Testing DDX ®Digital Amplifier, Apogee Technology, Inc., Norwood, USA [online]. Available at: http://www.metatech.com.hk/appnote/apogee/pdf/TestingNote.pdf.

Jwu-Sheng Hu received the B.S. degree from the Department of Mechanical Engineering, National Taiwan University, Taiwan, in 1984, and the M.S. and Ph.D. degrees from the Department of Mechanical Engineering, University of California at Berkeley, in 1988 and 1990, respectively.

He is currently a Professor in the Department of Electrical and Control Engineering, National Chiao-Tung University, Taiwan, ROC. His current research interests include microphone array signal processing, active noise control, intelligent mobile robots, embedded systems and applications.

Keng-Yuan Chen received the B.S. degree from the Department of Electrical and Control Engineering, National Chiao-Tung University, Taiwan, in 2003, and the M.S. degree from the Department of Electrical and Control Engineering, National Chiao-Tung University, Taiwan, in 2005. She is currently a Ph.D. candidate in the Department of Electrical and Control Engineering, National Chiao-Tung University, Taiwan, ROC. Her main research interests cover digital signal processing and class-d amplification.