Tariq and Ohno *EURASIP Journal on Advances in Signal Processing* (2016) 2016:29
DOI 10.1186/s13634-016-0326-2

EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Unified LMI-based design of $\Delta\Sigma$ modulators

M. Rizwan Tariq* and Shuichi Ohno

Abstract

Optimal finite impulse response (FIR) error feedback filters for noise shaping in $\Delta\Sigma$ modulators are designed by using weighting functions based on the system norms. We minimize the weighted norms of the quantization error in the output of a $\Delta\Sigma$ modulator, which corresponds to the minimization of the system norm. Three norms, the H_2 system norm, the H_∞ system norm, and the l_1 norm of the impulse response of the system, are adopted. The optimization problem for three types of FIR filters are evaluated by using linear matrix inequalities (LMIs) and then solved numerically via semi-definite programming. Design examples are provided to demonstrate the effectiveness of our proposed methods.

Keywords: Delta-sigma modulation, Noise shaping, Quantization, Semi-definite programming

1 Introduction

Analog-to-digital (A/D) and digital-to-analog (D/A) data converters are some of the most important parts of the electronic systems which act as the interface between the digital signal world and the real analog world. In A/D converters, the continuous-valued signals are discretized and quantized for transmission in wireline or wireless systems [1]. The process of quantization maps the continuous-valued signal to the discrete-valued signal. This usually introduces an undesirable effect, which is known as quantization noise. The important aspect of these converters is their ability to determine whether and how much the conversion can correctly keep the important information of signals, while suppressing undesirable noises.

Currently, the delta-sigma ($\Delta\Sigma$) modulation is a popular technique for making high-resolution A/D and D/A converters [2, 3]. Modern $\Delta\Sigma$ converters offer several benefits including high resolution, low power consumption, and low cost, making them a reasonable choice for the A/D converter for many signal processing applications such as audio devices [4, 5]. These $\Delta\Sigma$ A/D converters are effective for converting analog signals over a wide range of frequencies, from DC to several megahertz.

The $\Delta\Sigma$ modulator mainly consists of a static uniform quantizer and an error feedback filter to shape quantization noise [6], which is called noise shaping filter. The input to the modulator is an oversampled signal which is to be digitized. In oversampling, the signal is sampled at a frequency much higher than the Nyquist frequency (two times the input bandwidth) which reduces the effect of the quantization noise in the frequency band carrying the information signal, while the total noise remains the same.

The high-rate digital output of the modulator has two components, one is the signal which is located in the low-frequency region and the other is the noise which has to be reduced.

In the design of a $\Delta\Sigma$ modulator, the objective is to minimize the in-band quantization noise which as a result improves the signal-to-quantization-noise ratio (SQNR) of a $\Delta\Sigma$ modulator. It has been observed that the technique of oversampling alone may not be enough to improve the SQNR in the band of interest, and we need to exploit the noise shaping properties of the $\Delta\Sigma$ modulator to further reduce the in-band quantization noise. This can be achieved by using a feedback filter which employs the noise shaping to obtain a high SQNR while keeping the oversampling ratio (OSR) not too high. Although the overall quantization noise may not be changed by the noise shaping, the SQNR is increased in the information signal frequency band of the frequency spectrum. Our objective is to design the finite impulse response (FIR) noise shaping

*Correspondence: rizwan-tariq@hiroshima-u.ac.jp
Department of System Cybernetics, Hiroshima University, Saijo, Higashi-Hiroshima, Japan

filter of the $\Delta\Sigma$ modulator so that we can minimize the noise in the frequency region which constitutes our signal bandwidth.

Several designs for feedback filters have been proposed which also use the noise spectrum shaping technique [7, 8]. The FIR error spectrum shaping filters have been proposed for recursive digital filters composed of cascaded second order section in [9]. In [10], the noise transfer function (NTF) is assumed to have an infinite impulse response which is converted to a minimization problem by virtue of generalized Kalman-Yakubovich-Popov (GKYP) lemma. Then, an iterative algorithm is developed to solve this minimization problem subject to quadratic matrix inequalities. The method in [11] is a min-max design to optimize the NTF via GKYP lemma. This approach minimizes the worst case gain of the NTF over the signal frequency band and is shown to be able to improve the overall SNR of $\Delta\Sigma$ modulators as well. However, the method in [11] cannot incorporate the system connected to the quantizer into its design, while we consider a non-ideal output filter to minimize the quantization noise. In [12, 13], the optimization problem based on H_2 norm is formulated as a convex quadratic optimization problem where the weighting function (output filter) impulse response in truncated to finite number of samples.

In this paper, to keep $\Delta\Sigma$ modulators versatile, we utilize the weighting function to design $\Delta\Sigma$ modulators. We minimize the weighted quantization noise in the output of the $\Delta\Sigma$ modulator. Three norms are adopted to measure the quantity of the weighted quantization noise. One is the variance of the weighted quantization noise when the quantization errors at different time are assumed to be independent of each other. The others are the l_2 and the l_∞ norms of the weighted quantization noise. They correspond to the minimization of the H_2 system norm, the H_∞ system norm, and the l_1 norm of the impulse response of a system, respectively, and can be formulated as convex optimization problems with linear matrix inequalities (LMIs), which can be solved efficiently. The three norms considered in this paper are the most commonly used norms for quantifying signals. Our proposed method based on LMIs is termed unified for these three most commonly used norms only. If one imposes a constraint on the filter, then there are nine combinations for the design, three types of objectives, and three types of constraints, which can be handled by LMIs. One of these nine combinations, H_2 norm subjected to the Lee criterion, is similar to the design criteria of the method in [12, 13]. However, for our proposed H_2 norm design, we provide an alternate approach based on expressing the H_2 norm by using LMIs. The similarity lies in the fact that our proposed design and the method in [12, 13] use the idea of incorporating the non-ideal output filter for the minimization of the quantization noise.

The stability condition of $\Delta\Sigma$ modulators is also described by an LMI, which is incorporated into our design. Simulations with our designed noise shaping filters are performed, and comparisons with existing methods are made to demonstrate the effectiveness of our proposed design.

This paper is organized as follows: Section 2 gives the input/output relation of a linearized $\Delta\Sigma$ modulator with a weighting function. Then, we formulate our design problem which minimizes the weighted quantization noise under the stability condition. Section 3 is the main section of this paper, and we propose the design of the FIR feedback filter using LMIs for H_2, H_∞, and l_1 system norms. Section 4 gives design examples to show advantages of our method using simulation results. Section 5 provides us with the conclusion of our study.

Notation: Throughout this paper, $\mathbb{R}$ and $\mathbb{Z}$ denote a set of real numbers and a set of integer respectively. S denotes the set of all stable, proper, and rational transfer functions with real coefficients. The subscript $(\cdot)_+$ is used to indicate a subset restricted to non-negative numbers.

2 $\Delta\Sigma$ modulator and output weighting filter

Let us consider a general linearized model of a $\Delta\Sigma$ modulator for analyzing the noise shaping characteristics and designing the optimal noise shaping filter. We only consider the discretized single-input/single-output system with discrete-time signals. Let us denote the z transform of a sequence $f = \{f_k\}_{k=0}^{\infty}$ as $F[z] = \sum_{k=0}^{\infty} f_k z^{-k}$ and express the output (sequence) b of the linear time invariant (LTI) system $F[z]$ to the input $a = \{a_k\}_{k=0}^{\infty}$ as $b = F[z]\,a$.

Figure 1 shows the error feedback configuration of a $\Delta\Sigma$ modulator. The input to the modulator is y, while the output is u. The filter P[z] acts as a pre-filter to shape the frequency response of the input signal, and $Q(\cdot)$ is our static quantizer. The quantization error w is filtered by R[z] -1 and is fed back to y. We assume that $\lim_{z\to\infty} R[z] = 1$, i.e., the zeroth coefficient of the impulse response of R[z] is 1, which implies R[z] -1 is strictly proper. We also assume that

$$P[z]\,, R[z] \in S. \tag{1}$$

The static uniform quantizer can be described by two parameters, the quantization interval $d \in \mathbb{R}_+$ and the saturation level $L \in \mathbb{Z}_+$. For the continuous-valued input ξ, let the output of the static uniform quantizer be

$$Q(\xi) = \begin{cases} id, & \xi \in \left(\left(i - \frac{1}{2}\right)d, \left(i + \frac{1}{2}\right)d\right) \text{ and } |\xi| \le L \\ L, & \xi > L \\ -L, & \xi < -L \end{cases}, \tag{2}$$

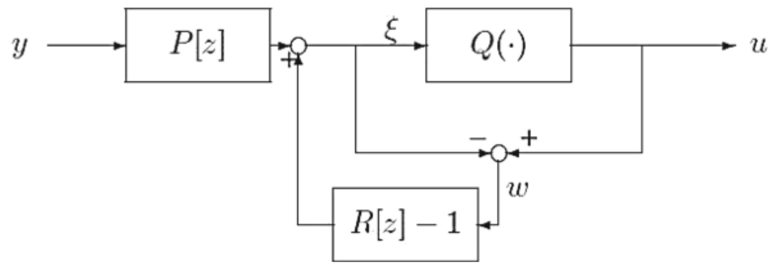

Fig. 1 Error feedback structure of linearized $\Delta\Sigma$ modulator

where d is the quantization interval and i is an integer. We assume that the saturation level is sufficiently large to avoid the saturation.

The difference between the input and the output of the static quantizer Q is known as a quantization error, which is denoted at time k as

$$w_k = u_k - \xi_k. \tag{3}$$

The quantization error is filtered by the noise shaping filter and added to the input to the static quantizer. Then, the input to the static quantizer is expressed as

$$\xi = P[\mathrm{z}]\, y + (R[\mathrm{z}] - 1) w. \tag{4}$$

Then, we have

$$u = w + \xi = P[\mathrm{z}]\, y + R[\mathrm{z}]\, w. \tag{5}$$

The gain from the input y to the output of the modulator u is known as as signal transfer function (STF), while the gain between the quantization error w and the modulator output u is commonly known as noise transfer function (NTF). In our setting, the STF and NTF for the $\Delta\Sigma$ modulator are $P[\mathrm{z}]$ and $R[\mathrm{z}]$, respectively.

The feedback loop acts in such a way that the quantization noise is shifted away from a certain frequency band. If the input to the modulator lies within this certain frequency band, then most of the noise due to quantization lies outside the frequency band of interest.

To design the noise shaping filter, we utilize a weighting function $H_W[\mathrm{z}]$. More specifically, we consider the weighted quantization noise ϵ defined as

$$\epsilon = H_W[\mathrm{z}]\, R[\mathrm{z}]\, w, \tag{6}$$

where $H_W[\mathrm{z}] \in S$. Without loss of generality, we normalize the maximum magnitude of $H_W[\mathrm{z}]$ to be in unity. The weighting function is selected to reduce the effect of the quantization noise in the passband of the y. For example, when the passband of y is $\left[-\omega_p, \omega_p\right]$, we will use the weighting filter that meets $H_W\left[e^{j\omega}\right] \approx 1$ for $\omega \in \left[-\omega_p, \omega_p\right]$ and $|H_W\left[e^{j\omega}\right]|$ is small enough outside the passband to let most of the noise be outside the passband.

Suppose that the output of our $\Delta\Sigma$ modulator u is connected to a system $H_S[\mathrm{z}]$ whose output is denoted by v. Then, we have

$$v = H_S[\mathrm{z}]\, u. \tag{7}$$

Substituting (5) into (7), we get

$$v = H_S[\mathrm{z}]\, P[\mathrm{z}]\, y + H_S[\mathrm{z}]\, R[\mathrm{z}]\, w. \tag{8}$$

In [12], the noise $H_S[\mathrm{z}]\, R[\mathrm{z}]\, w$ is minimized based on the H_2 system norm to reduce the in-band quantization noise. When $H_W[\mathrm{z}] = H_S[\mathrm{z}]$, our minimization based on the H_2 system norm is equivalent to the minimization of [12]. Then, the difference between the proposed method and [12] lies in the usage of different optimization procedures for solving the H_2 norm objective function. As pointed out in [12], if one knows the system $H_S[\mathrm{z}]$ connected to the $\Delta\Sigma$ modulator, one should set $H_W[\mathrm{z}] = H_S[\mathrm{z}]$. If not, we could design more general $\Delta\Sigma$ modulators using weighting functions. Thus, our objective is to obtain the optimal filter $R[\mathrm{z}]$ in (8) for a given $H_W[\mathrm{z}]$ that minimizes $\epsilon = H_W[\mathrm{z}]\, R[\mathrm{z}]\, w$ in a sense.

The signal w which is the difference between the input and the output of the static uniform quantizer satisfies

$$|w_k| \leq \frac{d}{2}. \tag{9}$$

Since the transfer function from w to ϵ is linear, we can put $d = 2$ without the loss of generality so that $|w_k| \leq 1$ and hence $|w_k|^2 \leq 1$.

Let us define the H_2 norm of a system $H[z]=\sum_{k=0}^{\infty} h_k z^{-k}$ as

$$||H[z]||_2 = \left[\sum_{k=0}^{\infty} |h_k|^2\right]^{\frac{1}{2}}. \quad (10)$$

The quantization error may be modeled as a uniform random variable with zero mean and variance σ_w^2 [2]. If the errors at different times are independent of each other, then the variance of the weighted quantization noise ϵ_k is given by

$$\sigma_\epsilon^2 = ||H_W[z]\,R[z]\,||_2^2 \sigma_w^2. \quad (11)$$

As a deterministic sequence, the weighted quantization noise signal ϵ may be measured with its l_p norm defined as

$$||\epsilon||_p = \left[\sum_{k=0}^{\infty} |\epsilon_k|^p\right]^{\frac{1}{p}}. \quad (12)$$

Among l_p, the l_2 and the l_∞ norm are often utilized. The l_∞ norm of a discrete signal $\{\epsilon_k\}_{k=0}^{\infty}$ is defined as

$$||\epsilon||_\infty = \sup_k |\epsilon_k|. \quad (13)$$

The value of $||\epsilon||_\infty$ is the largest absolute value of the error signal and hence can be used to consider the worst case errors. On the other hand, the l_2 norm is defined as

$$||\epsilon||_2 = \left[\sum_{k=0}^{\infty} |\epsilon_k|^2\right]^{\frac{1}{2}}. \quad (14)$$

The stability of $\Delta\Sigma$ modulators should also be considered. Here we consider the l_p bounded stability. Suppose that y is bounded such that $||y||_p < \gamma_y$ for a finite γ_y. Then, it is easy to see that if $||R[z]\,w||_p$ is bounded, then the input ξ to the static quantizer, which is the internal variable of the $\Delta\Sigma$ modulator, is bounded. Thus, to guarantee the l_p stability of the $\Delta\Sigma$ modulator, it is sufficient to assure

$$||R[z]\,w||_p < \gamma, \quad (15)$$

for a finite γ.

If one takes the $p = 2$ norm, then

$$\sup_w ||R[z]\,w||_2 \leq ||R[z]\,||_\infty ||w||_2, \quad (16)$$

where $||R[z]\,||_\infty$ is the H_∞ system norm defined as

$$||R[z]\,||_\infty = \max_\omega |R\left[e^{j\omega}\right]|. \quad (17)$$

The constraint on the H_∞ norm of the NTF is known as Lee criterion [6, 14]. The peak value of the NTF magnitude response must be bounded to some constant value γ, where the value of γ depends on the number of saturation levels. For the case of binary quantizers, the value of γ is usually set as 1.5.

In summary, we would like to minimize σ_ϵ^2, $||\epsilon||_2$, or $||\epsilon||_\infty$ under the stability condition (15), assuming that the variance of w_k, $||w||_2$ or $||w||_\infty$ is finite. It should be noted that without the weighting function, we only have the trivial solution such that $R[z] = 1$, that is, the error feedback is not necessary.

3 Design of FIR noise shaping filters using linear matrix inequalities

Since FIR filters are often preferred, we confine our attention to design of FIR filters of order n, denoting

$$R[z] = \sum_{k=0}^{n} r_k z^{-k}, \;\; r_0 = 1. \quad (18)$$

The coefficient r_0 of the impulse response of the FIR filter $R[z]$ is unitary to ensure $R[z] \in S$, which makes the noise shaping filter strictly proper.

Let us denote the matrices of a state-space realization of $R[z]$ by $(A_R, B_R, C_R, 1)$, where

$$A_R = \begin{bmatrix} 0 & 1 & & 0 \\ \vdots & \ddots & \ddots & \\ \vdots & & \ddots & 1 \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \; B_R = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (19)$$

$$C_R(r) = \left[\, r_n, \; r_{n-1}, \; \cdots \; r_1 \,\right]. \quad (20)$$

It is noted that A_R and B_R are constant. Our design parameter is

$$r = [r_1, \ldots, r_n] \quad (21)$$

which defines $C_R(r)$ above.

The weighted quantization noise ϵ in (6) to be minimized is characterized by the the composite system $H_W[z]\,R[z]$, which has to be internally stable.

Let $H_W[z]$ be a proper function, whose (A, B, C, D) matrices of a state-space realization is (A_H, B_H, C_H, D_H). Let the order of $R[z]$ be n and let $(A_R, B_R, C_R, 1)$ be (A, B, C, D) matrices of a state-space realization of $R[z]$. Then, one can express the state-space realization of $H_W[z]\,R[z]$ as

$$x_{k+1} = Ax_k + Bw_k \quad (22)$$

$$\epsilon_k = Cx_k + Dw_k \quad (23)$$

where

$$A = \begin{bmatrix} A_R & B_R C_H \\ \mathbf{0} & A_H \end{bmatrix}, \; B = \begin{bmatrix} B_R D_H \\ B_H \end{bmatrix},$$

$$C = \left[\, C_R \;\; C_H \,\right], \; D = D_H. \quad (24)$$

First of all, let us consider the minimization of the variance σ_ϵ^2 of the weighted quantization error under the white noise assumption. It is sufficient to minimize the H_2 norm of $H_W[z]\,R[z]$ to minimize σ_ϵ^2 given by (11).

For FIR $R[z]$, $||H_W[z]\,R[z]\,||_2^2$ can be expressed as a quadratic function of $r = [r_1, \ldots, r_n]$ by using inverse Fourier transform of $|H_W\left[e^{j\omega}\right]|^2$ [8], which requires numerical integrations. On the other hand, a truncated impulse response of $H_W[\mathrm{z}]\,R[\mathrm{z}]$ is utilized in [12], where the order of some parameters is scaled by the length of the truncated impulse response. Here, we adopt LMIs to numerically evaluate the H_2 norm based on the next lemma.

Lemma 1. *([15]) Let $G[\mathrm{z}]$ be a proper stable rational function, whose state-space realization is (A, B, C, D). Then, A is Schur and*

$$||G[\mathrm{z}]\,||_2^2 < \mu_2 \tag{25}$$

if and only if there exist positive definite matrices P and Z which satisfy

$$APA^T - P + BB^T \prec 0 \tag{26}$$

$$Z - DD^T - CPC^T \succ 0 \tag{27}$$

$$\mathrm{trace}(Z) < \mu_2. \tag{28}$$

Using the Schur complement, one can show that (26) holds true if and only if

$$\begin{bmatrix} P & PA & PB \\ A^TP & P & 0 \\ B^TP & \mathbf{0} & 1 \end{bmatrix} \succ 0. \tag{29}$$

Similarly, since our system has a single input and a single output, Eq. (27) for (A, B, C, D) can be expressed as

$$\begin{bmatrix} \mu_2 & C & D \\ C^T & P & \mathbf{0} \\ D^T & \mathbf{0} & 1 \end{bmatrix} \succ 0. \tag{30}$$

On the other hand, the l_2 norm of the weighted quantization noise is bounded as

$$||\epsilon||_2 = ||H_W[\mathrm{z}]\,R[\mathrm{z}]\,w||_2 \le ||H_W[\mathrm{z}]\,R[\mathrm{z}]\,||_\infty ||w||_2. \tag{31}$$

We can utilize the bounded real lemma that provides us an LMI to evaluate the gain.

Lemma 2. *([16]) Let $G[\mathrm{z}]$ be a proper stable rational function, whose state-space realization is (A, B, C, D). Then, A is Schur and*

$$||G[\mathrm{z}]\,||_\infty^2 < \mu_\infty \tag{32}$$

if and only if there exists a positive definite matrix P which satisfies

$$\begin{bmatrix} A^TPA - P + C^TC & A^TPB + C^TD \\ B^TPA + D^TC & B^TPB + D^TD - \mu_\infty I \end{bmatrix} \prec 0. \tag{33}$$

By using the Schur complement, (33) can be converted into an LMI given by

$$\begin{bmatrix} -P & PA & PB & \mathbf{0} \\ A^TP & -P & \mathbf{0} & C^T \\ B^TP & 0 & -\mu_\infty & D^T \\ \mathbf{0} & C & D & -1 \end{bmatrix} \prec 0. \tag{34}$$

Let us define the l_∞ norm of the impulse response of a system $H[\mathrm{z}] = \sum_{k=0}^{\infty} h_k \mathrm{z}^{-k}$ as

$$||H[\mathrm{z}]\,||_{\mathrm{imp}} = \sum_{k=0}^{\infty} |h_k|. \tag{35}$$

We call $||H[\mathrm{z}]\,||_{\mathrm{imp}}$ as H_{imp} norm for convenience. Then, the l_∞ norm $||\epsilon_k||_\infty$ is bounded as

$$||\epsilon||_\infty = ||H_W[\mathrm{z}]\,R[\mathrm{z}]\,w||_\infty \le ||H_W[\mathrm{z}]\,R[\mathrm{z}]\,||_{\mathrm{imp}} ||w||_\infty$$

$$= ||H_W[\mathrm{z}]\,R[\mathrm{z}]\,||_{\mathrm{imp}}. \tag{36}$$

We can reduce $||\epsilon||_\infty$ by minimizing $||H_W[\mathrm{z}]\,R[\mathrm{z}]\,||_{\mathrm{imp}}$.

Unlike the H_2 norm and the H_∞ norm, only upper bounds of the H_{imp} norm are available. In [17, 18], an upper bound based on the invariant set of a discrete-time system has been utilized to design infinite impulse response (IIR) error feedback filters for dynamic quantizers. The invariant set of a discrete-time system is defined as follows [19]:

Definition 1. Let $x_k \in \mathbb{R}^n$ be the state vector of the LTI system given by

$$x_{k+1} = Ax_k + Bw_k \tag{37}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $w_k \in \mathbb{R}^m$. A set $\mathcal{X}$ that satisfies $x_{k+1} \in \mathcal{X}$ if $x_k \in \mathcal{X}$ and $w_k^T w_k \le 1$ is called an invariant set of the system given by (37).

The following lemma describes how to obtain an ellipsoid which is an invariant set of the system (37).

Lemma 3. *([19]) Let $\mathcal{E}(P)$ be the ellipsoid defined by an $n \times n$ real symmetric matrix $P \succ 0$ as $\mathcal{E}(P) = \{x \in \mathbb{R}^n : x^TPx \le 1\}$.*

The ellipsoid $\mathcal{E}(P)$ is an invariant set of the system (37) if and only if there exists a scalar $\alpha \in \left[0, 1 - \rho^2(A)\right]$ which satisfies

$$\begin{bmatrix} A^TPA - (1-\alpha)P & A^TPB \\ B^TPA & B^TPB - \alpha I \end{bmatrix} \preceq 0 \tag{38}$$

where $\rho(A)$ is the spectrum radius of A.

It should be noted that unlike H_2 and H_∞, H_{imp} depends on parameter α.

If $x_k \in \mathcal{E}(P)$, then

$$\sup_{x_k \in \mathcal{E}(P)} |Cx_k|^2 = CP^{-1}C^T. \tag{39}$$

It follows from $|\epsilon_k| = |Cx_k + Dw_k| \le |Cx_k| + |Dw_k|$ that

$$||H_W[\mathrm{z}]\,R[\mathrm{z}]\,||_{\mathrm{imp}} \le |CP^{-1}C^T|^{\frac{1}{2}} + |D|. \tag{40}$$

Thus, we can conclude that $|CP^{-1}C^T|^{\frac{1}{2}} + |D|$ is an upper bound of the norm.

Since D is constant, we minimize $CP^{-1}C^T$ with respect to α and $C_R(r)$. It should be also remarked that we can

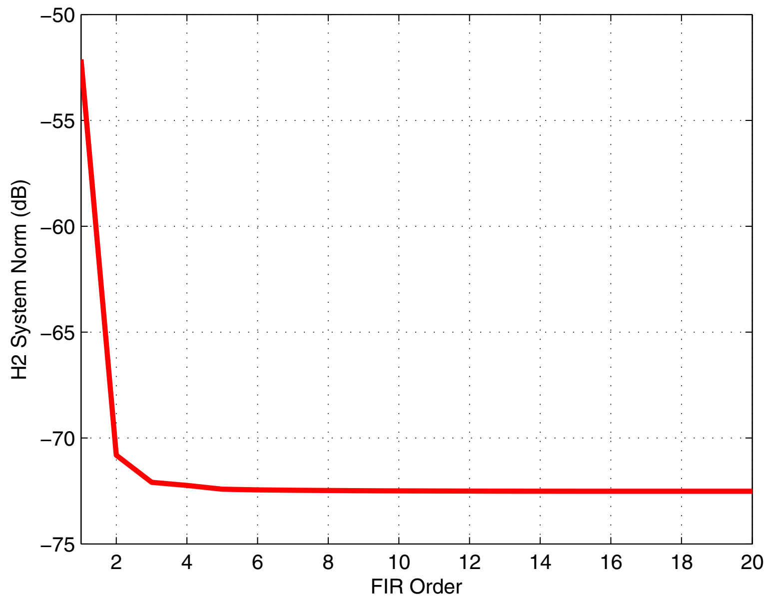

Fig. 2 The H_2 norm of $H_W[z]\ R[z]$ as a function of order of $R[z]$ for the first-order lowpass weighting function, where $R[z]$ is designed based on the H_2 norm

assume that $\alpha \neq 0$ since our B matrix is not zero. Similarly, we can express (38) with (A, B, C, D) as

$$\begin{bmatrix} (1-\alpha)P & 0 & A^T P \\ \mathbf{0} & \alpha & B^T P \\ PA & PB & P \end{bmatrix} \succeq 0. \tag{41}$$

Moreover, using the Schur complement, we can express $CP^{-1}C^T \leq \mu$ as an LMI given by

$$\begin{bmatrix} P & C^T \\ C & \mu \end{bmatrix} \succeq 0. \tag{42}$$

For a fixed α, the minimization of μ is a semidefinite program, which can be numerically solved by existing optimization packages, e.g., CVX [20]. Then, all we have to do is to find α which gives the minimum. Since A is our

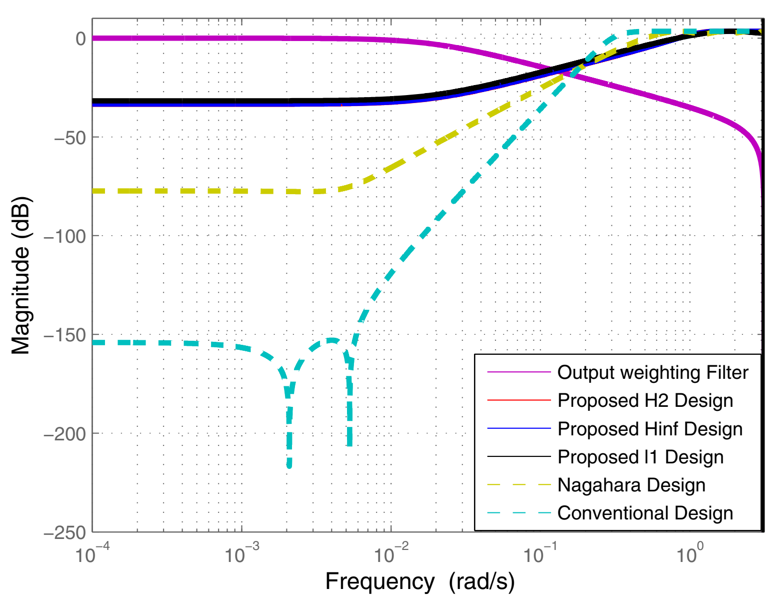

Fig. 3 Frequency responses of filters designed by the proposed method and the referenced methods. The weighting function is of order unity

design parameter, a line search for $\alpha \in (0, 1)$ is required to obtain the minimum. The optimal (A, B, C, D) is given by the arguments corresponding to the optimal α.

Not only the objective function but also the condition (15) on the stability can be described by LMIs. For example, as shown in [11], it follows from Lemma 2 that the Lee criterion

$$||R[z]||_\infty < \gamma \tag{43}$$

is satisfied if and only if there exists a positive definite matrix P_R which meets

$$\begin{bmatrix} -P_R & P_R A_R & P_R B_R & \mathbf{0} \\ A_R^T P_R & -P_R & \mathbf{0} & C_R^T \\ B_R^T P_R & 0 & -\gamma^2 & 1 \\ \mathbf{0} & C_R & 1 & -1 \end{bmatrix} \prec 0. \tag{44}$$

Thus, if one would like to design the FIR noise shaping filter that minimizes σ_ϵ^2 under the Lee criterion, it suffices to solve the following convex optimization problem

$$\min_{r_1,\ldots,r_n} \mu_2 \tag{45}$$

subject to (29), (30), and (44).

The LMIs for other stability conditions $||R[z]||_2 < \gamma$ and $||R[z]||_{\mathrm{imp}} < \gamma$ can be obtained similarly, which are omitted to avoid the duplication.

In summary, our unified approach enables the design of the FIR noise shaping filter to minimize the H_2, the H_∞, or the l_1 system norm under the H_2, the H_∞, or the l_1 norm constraint. Moreover, since norms are described by LMIs, different types of problems can be solved numerically. For example, some signal processing applications may require us to design an error feedback filter for a $\Delta\Sigma$ modulator by adding a constraint that limits the magnitude of the weighted quantization noise to a certain value. Then, our design objective is to design the noise shaping filter that attains the optimal value of the stability threshold γ under the maximum weighted quantization noise constraint. If we adopt the Lee criterion, we can obtain the most stable error feedback filter by minimizing (43) subject to $||\epsilon||_\infty \leq c$, where c is the maximum bound on the weighted quantization noise ϵ, by using LMIs in (41), (42), and (44).

4 Design examples

In this section, simulations for lowpass and bandpass $\Delta\Sigma$ modulators have been shown using the proposed design method based on H_2, H_∞, and l_1 system norms. For the design of a conventional $\Delta\Sigma$ modulator by NTF zero optimization method [6], the DELSIG toolbox [21] is utilized to obtain the frequency response of an IIR noise shaping filter with *synthesizeNTF* MATLAB function. The frequency response and the noise shaping characteristics of the FIR feedback filter proposed in [11] are also compared with our designed filters. As our proposed H_2 norm minimization is mathematically equivalent to the method proposed in [12], the numerical results which we have performed also show that there is no significant difference between our proposed H_2 norm method and [12]. To avoid redundancy in our simulations, we omit the comparison with [12].

All simulation results are obtained by using MATLAB programming, while semi-definite programming (SDP)

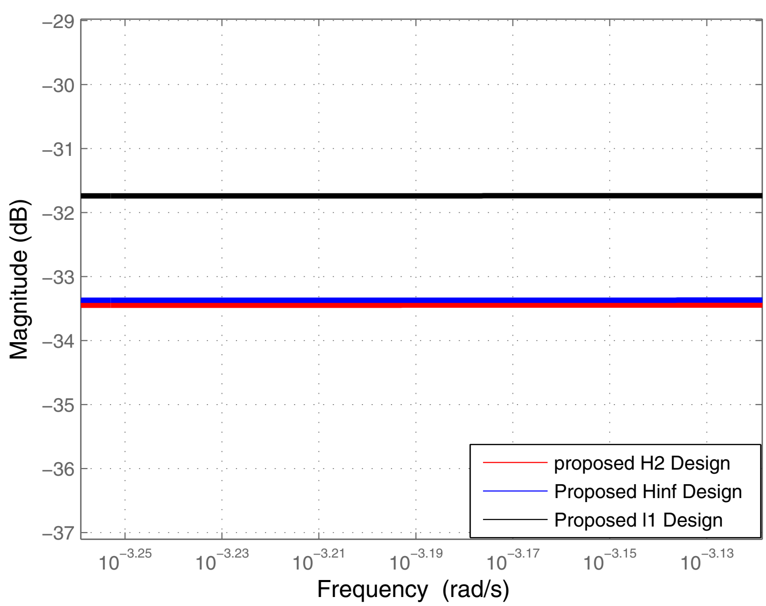

Fig. 4 Enlarged frequency response of our proposed filters in Fig. 3

problems are solved by using CVX tool [20], which is an effective solver for convex optimization problems.

4.1 Lowpass $\Delta\Sigma$ modulator with the first-order weighting function

Now, let us design a lowpass $\Delta\Sigma$ modulator by using a first-order lowpass Butterworth filter as our weighting function $H_W[z]$. The first-order Butterworth filter provides us the maximum flat response in the passband at the expense of a wide transition band as the filter changes from the passband to the stopband. The input signal y to the lowpass $\Delta\Sigma$ modulator is assumed to be oversampled with an oversampling ratio (OSR) of 512. Then, the cut-off frequency of the first-order Butterworth filter is set at $\pi/\mathrm{OSR} \approx 0.0061$ in the normalized angular frequency interval $[0, \pi]$.

For the stability of the $\Delta\Sigma$ modulator, we assume the value of the Lee coefficient γ to be 1.5 which is equivalent to 3.52 in decibels ; however, the value of γ can be increased further as long as the $\Delta\Sigma$ modulator remains stable.

The order of the FIR feedback filter $R[z]$ is chosen based on the convergence behavior of the objective function. Figure 2 shows that the H_2 norm of $H_W[z]\,R[z]$ reaches a value as we keep on increasing the order of FIR filter.

Table 1 $||H_W[z]\,R[z]\,||_2$, $||H_W[z]\,R[z]\,||_\infty$, and l_1 norms of the impulse response of $H_W[z]\,R[z]$ for the first-order lowpass weighting function

	H_2 norm	H_∞ norm	l_1 norm
H_2 norm design	1.54×10^{-2}	2.19×10^{-2}	2.62×10^{-2}
H_∞ norm design	1.54×10^{-2}	2.16×10^{-2}	2.59×10^{-2}
l_1 norm design	1.63×10^{-2}	2.59×10^{-2}	2.59×10^{-2}
Nagahara design [11]	1.92×10^{-2}	3.82×10^{-2}	4.89×10^{-2}
Conventional design [6]	2.61×10^{-2}	6.92×10^{-2}	11×10^{-2}

Above the FIR order 8, the norm of the weighted quantization noise remains almost constant in terms of the H_2 norm, resulting in a high convergence rate. In this example, the FIR feedback filter $R[z]$ for noise shaping is set to be 8.

Figure 3 depicts the frequency responses of H_2, H_∞, and l_1 norm-based filters compared with the referenced methods in [6] and [11]. The order of FIR feedback filter in [11] is also chosen to be 8, while the order of IIR feedback filter for the conventional design [6] is set to be 4. Our designed FIR filters have almost the same frequency response. It can be observed that the frequency responses of our designed FIR filters have

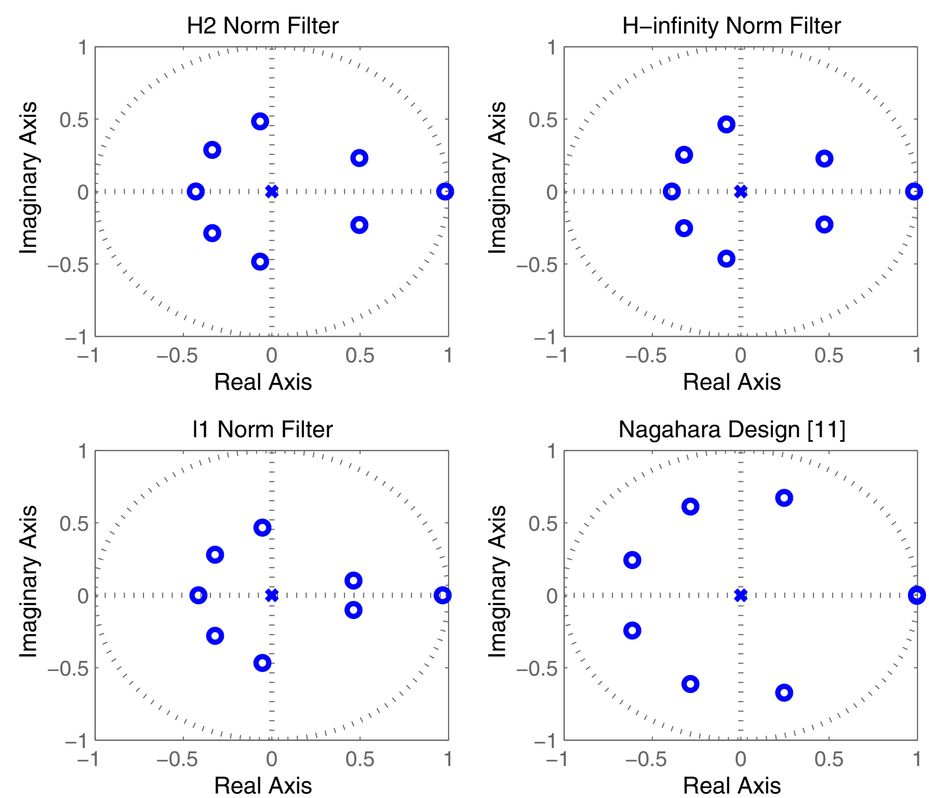

Fig. 5 Pole-zero placements of proposed designs and Nagahara design [11]

uniform attenuation in the low-frequency region of the frequency spectrum, while the conventional design shows a peak in the magnitude response near the cut-off frequency.

To precisely see the difference between the magnitude responses of our designed filters in low-frequency region, the enlarged view of Fig. 3 is shown in Fig. 4.

The method in [11] designs the FIR noise shaping filter based on the weighted H_∞ norm of $R[z]$. Near the cut-off frequency, the magnitude response of the FIR filter in [11] increases rapidly showing the high steepness in the transition band, while all of our proposed filters exhibit good performance, matching the steepness of the weighting function. Note that the maximum magnitude value of all filters are bounded to 3.52 dB approximately due to stability constraint which utilizes the Lee coefficient $\gamma = 1.5$.

Table 1 lists the H_2 norm $||H_W[z]\,R[z]\,||_2$, the H_∞ norm $||H_W[z]\,R[z]\,||_\infty$, and the l_1 norm of the impulse response of $H_W[z]\,R[z]$ for our designed FIR filters compared with the referenced designs in [6] and [11]. All three designed filters have less H_2, H_∞, and l_1 norms as compared with optimal feedback filters in [6] and [11]. Although the referenced designs have lower gains in the passband as observed in Fig. 3, our designed filters have better performance in the weighted norms. This is because the referenced designs only take into account the passband, while our design does the whole band by incorporating a lowpass output filter which is assumed to be non-ideal in practice. Indeed, if an ideal lowpass filter can be used as our weighting function, our H_∞ norm-based filter is equivalent to the weighted H_∞ norm-based filter in [11]. Since any ideal lowpass filter is not available in practice, it is important to consider the noise in the stopband. Our method can trade off the properties of the noise shaping filter in the passband and the stopband using an appropriate weighting function.

The H_∞ and l_1 norm designs exhibit an equivalent l_1 norm, while the H_2 and H_∞ norm designs have an equivalent H_2 norm. This may be partially due to the implementation and the numerical errors in our numerical optimization. It should be noted that we minimize the upper bounds, which implies that we cannot guarantee that the quantizer designed based on a norm is optimal in the sense of the norm.

Figure 5 shows the pole-zero placement for the lowpass $\Delta\Sigma$ modulator with proposed error feedback filters compared with the FIR filter in [11].

4.2 Lowpass $\Delta\Sigma$ modulator with the fourth-order weighting function

Now let us introduce a higher order lowpass Butterworth filter of order 4 as our weighting function, where the OSR is 32. The maximum magnitude of NTF is limited to 3.52 dB by using the Lee coefficient $\gamma = 1.5$. The fourth-order Butterworth filter with a cut-off frequency of $\pi/\text{OSR} \approx 0.0098$ has a better stopband attenuation than the first-order Butterworth filter by increasing the steepness of the passband to the stopband transition at the cost of reduced passband flatness.

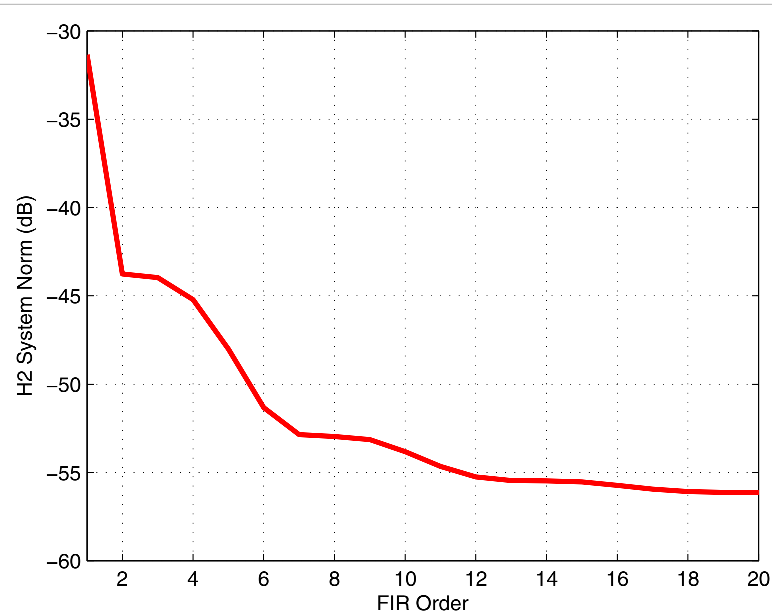

Fig. 6 The H_2 norm of $H_W[z]\,R[z]$ as a function of order of $R[z]$ for the fourth-order lowpass weighting function, where $R[z]$ is designed based on the H_2 norm

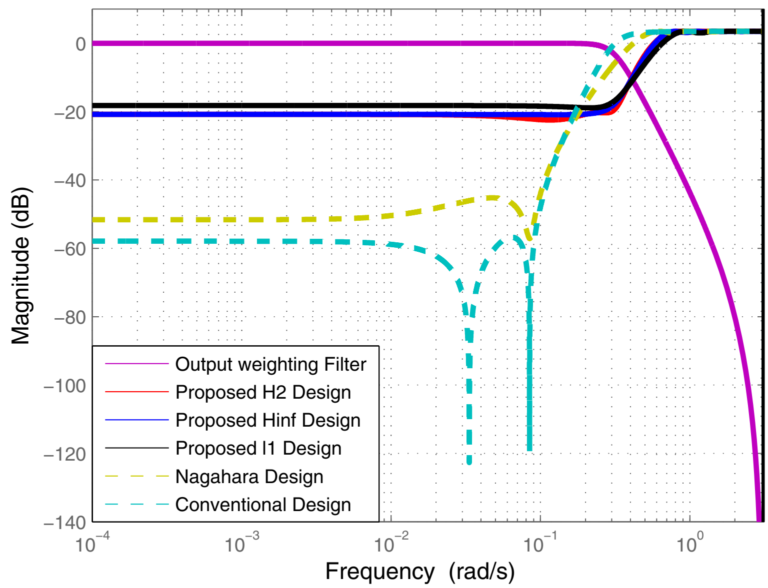

Fig. 7 Frequency responses of the filters designed by the proposed method and the referenced methods. The weighting function is of order 4

For this lowpass $\Delta\Sigma$ modulator, Fig. 6 shows the convergence behavior of the H_2 norm of $H_W[z]R[z]$ for the H_2 norm-based design. From this, the FIR feedback filter of order 20 is chosen for th epro-posed designs and referenced design in [11], while the IIR feedback filter for conventional design [6] is of order 4.

In Fig. 7, we give the frequency responses of proposed H_2, H_∞, and l_1 norm-based filters compared with the referenced methods. Our proposed designs show better performance by providing uniform attenuation in the low-frequency region and exhibiting better magnitude responses near the cut-off frequency as compared to the referenced methods in [11] and [6], for the three designed noise shaping filters.

Table 2 shows the H_2 norm, the H_∞ norm, and the l_1 norm of the impulse response of $H_W[z]R[z]$ for our designed FIR filters compared with the referenced designs in [6] and [11]. It can be observed that all three designed filters have less H_2, H_∞ and, l_1 norms than the optimal feedback filters in [6] and [11]. The H_2, H_∞, and l_1 norm designs have the least H_2, H_∞, and l_1 norms, respectively.

Table 2 $||H_W[z]R[z]||_2$, $||H_W[z]R[z]||_\infty$, and l_1 norms of the impulse response of $H_W[z]R[z]$ for the fourth-order lowpass weighting function

	H_2 norm	H_∞ norm	l_1 norm
H_2 norm design	3.95×10^{-2}	9.71×10^{-2}	1.40×10^{-1}
H_∞ norm design	4.07×10^{-2}	9.09×10^{-2}	1.24×10^{-1}
l_1 norm design	4.43×10^{-2}	1.22×10^{-1}	1.23×10^{-1}
Nagahara design [11]	9.18×10^{-2}	3.53×10^{-1}	4.74×10^{-1}
Conventional design [6]	1.49×10^{-1}	6.69×10^{-1}	9.01×10^{-1}

To assess the performance of the lowpass $\Delta\Sigma$ modulator with an error feedback filter obtained by our proposed H_2 norm-based design, the MATLAB function *simulateDSM* in DELSIG toolbox [21] is used to simulate the $\Delta\Sigma$ modulator for obtaining the digital output. The input to the $\Delta\Sigma$ modulator is a sinusoidal wave with a frequency of 100 Hz and and amplitude of 0.5. We assume a uniform quantizer with saturation levels $L = 2$ and quantization interval $d = 2$.

The output of this uniform quantizer is a digital signal which is represented by using +1 and −1 volts for binary 0 and 1, respectively, which is shown in the upper part of Fig. 8. The lower part of Fig. 8 is the frequency spectrum of the digital output, which gives the performance of our lowpass $\Delta\Sigma$ modulator. Our lowpass $\Delta\Sigma$ modulator attenuates the quantization noise in the frequency region which contains the information signal. The frequency notch for the input signal appears at 100 Hz, which is the same with the sinusoidal wave, and the magnitude of quantization noise is low in the passband. Our proposed H_2 filter efficiently shifts the quantization noise towards the high-frequency region which does not carry much information. Similar results can be found for H_∞ and l_1 norm-based designs, which are omitted.

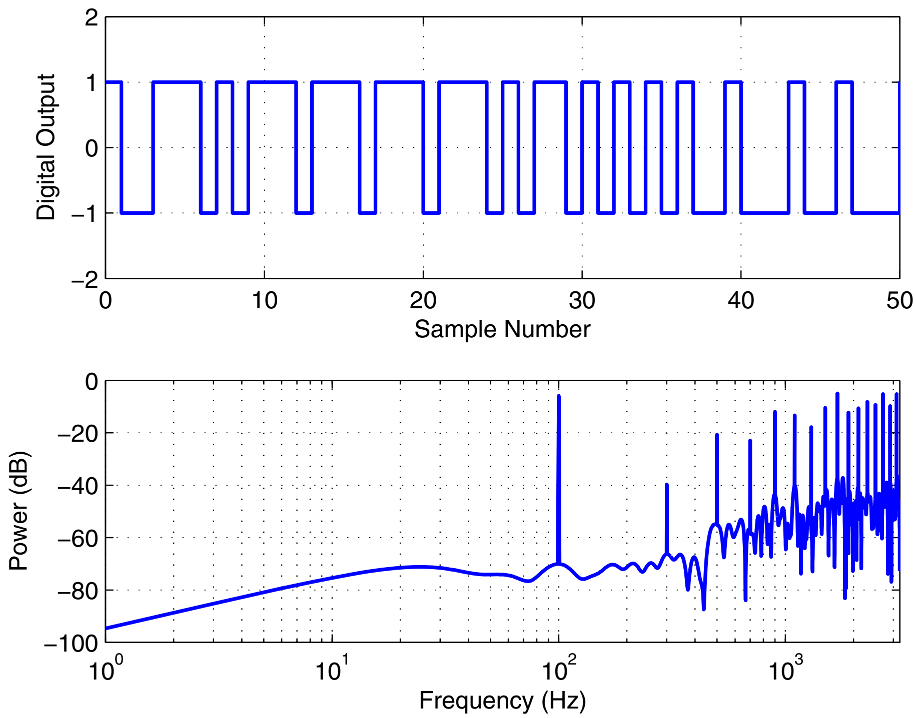

Fig. 8 Output and frequency spectrum plot of the lowpass $\Delta\Sigma$ modulator obtained by the proposed H_2 norm-based design

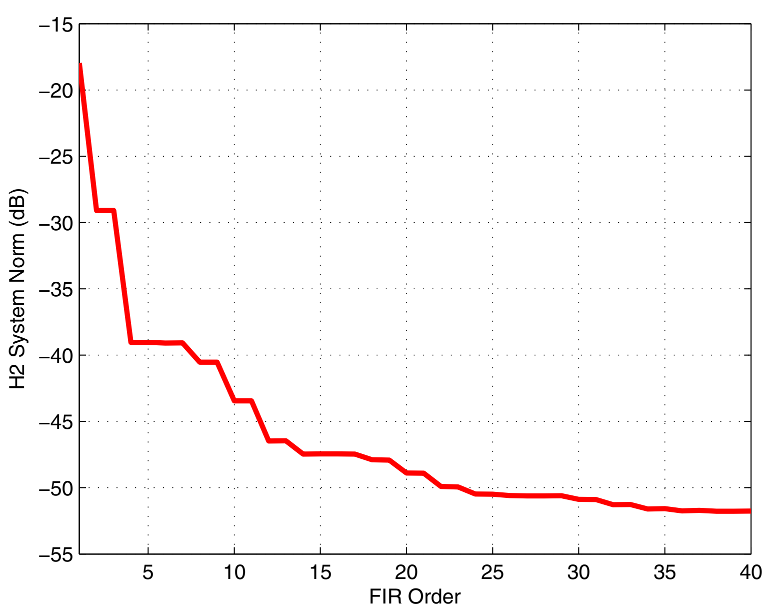

Fig. 9 The H_2 norm of $H_W[z]\,R[z]$ as a function of order of $R[z]$ for the sixth-order bandpass weighting function, where $R[z]$ is designed based on the H_2 norm

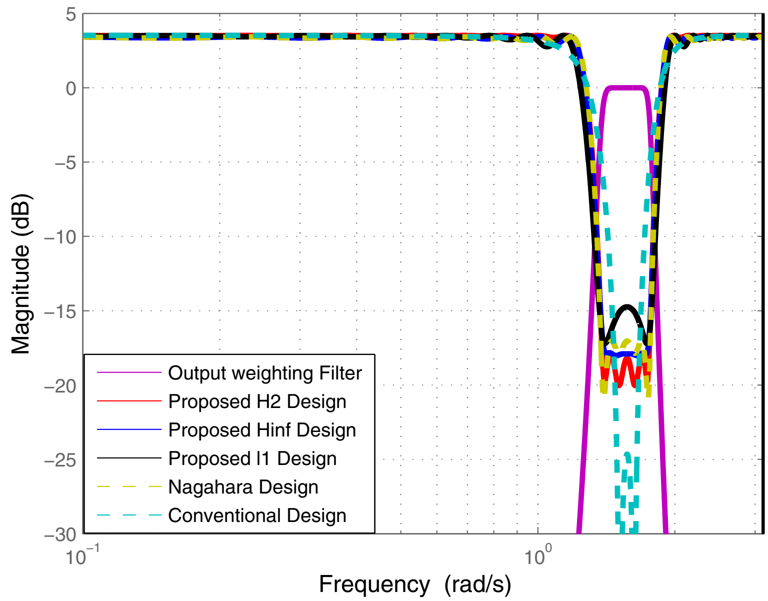

Fig. 10 Frequency responses of filters designed by the proposed method and the conventional method. The weighting function is of order 6

4.3 Bandpass $\Delta\Sigma$ modulator with the sixth order weighting function

Finally, we adopt a sixth-order bandpass Butterworth filter as our weighting function, whose frequency response is found in Fig. 10.

The input to the modulator is assumed to have the center frequency $\omega_\circ = \pi/2$ and bandwidth parameter $\Omega = \pi/16$. For the passband $\omega \in [\pi/2 - \pi/16, \pi/2 + \pi/16]$, we use the bandpass Butterworth filter that meets $H_W[e^{j\omega}] \approx 1$ for $\omega \in [\omega_\circ - \Omega, \omega_\circ + \Omega]$, and $|H_W[e^{j\omega}]|$ is small enough outside the passband to let most of the noise be outside the passband. For the conventional design [6], OSR is set to be 16.

As illustrated in Fig. 9 the H_2 norm of $H_W[z]R[z]$ for H_2 norm-based design converges slowly compared to the previous examples. A longer order is required to adjust to the sixth-order bandpass Butterworth filter. Thus, the order of proposed FIR feedback filters $R[z]$ is chosen to be 40. The order of FIR feedback filter in [11] is also set to be 40. For the conventional bandpass $\Delta\Sigma$ modulator [6],

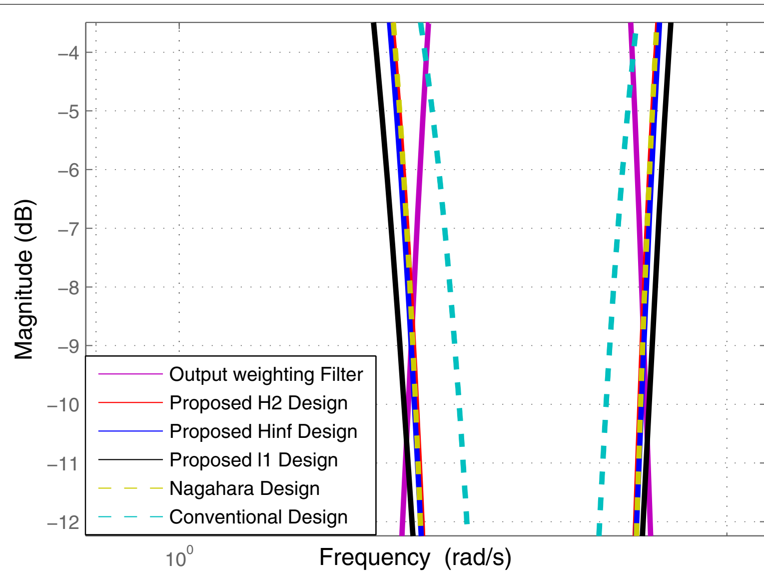

Fig. 11 Enlarged frequency response of our proposed filters in Fig. 10

Table 3 $||H_W[z]\,R[z]\,||_2$, $||H_W[z]\,R[z]\,||_\infty$, and l_1 norms of the impulse response of $H_W[z]\,R[z]$ for the sixth-order bandpass weighting function

	H_2 norm	H_∞ norm	l_1 norm
H_2 norm design	5.08×10^{-2}	1.385×10^{-1}	2.094×10^{-1}
H_∞ norm design	5.38×10^{-2}	1.277×10^{-1}	1.916×10^{-1}
l_1 norm design	6.08×10^{-2}	1.833×10^{-1}	1.858×10^{-1}
Nagahara design [11]	5.45×10^{-2}	1.408×10^{-1}	2.222×10^{-1}
Conventional design [6]	10.19×10^{-2}	4.253×10^{-1}	5.461×10^{-1}

the order of the IIR feedback filter is 4, whereas the center frequency is $f_\circ = 1/4$.

We compare the frequency responses of our proposed FIR feedback filters for the bandpass $\Delta\Sigma$ modulator with the referenced designs in [6] and [11]. Figure 10 shows that the magnitude responses of proposed H_2 and H_∞ design FIR filters have higher attenuation levels as compared to the method proposed in [11]. Again, the magnitude responses of our proposed design filters are uniformly attenuated over the passband, while the conventional design shows a peak near the edges of the band which can be observed in Fig. 11.

Table 3 gives the H_2 norm, the H_∞ norm, and the l_1 norm of the impulse response of $H_W[z]\,R[z]$ for our designed FIR filters compared with the referenced designs in [6] and [11]. Again, our proposed H_2, H_∞, and l_1 norm designs have the least H_2, H_∞, and l_1 norms, respectively.

4.4 Stability under the l_∞ norm constraint on the weighted quantization noise

Here, to obtain the most stable error feedback filter for a lowpass $\Delta\Sigma$ modulator, we minimize (43) under the l_∞ norm constraint on the weighted quantization noise such that $||\epsilon||_\infty = 1.96 \times 10^{-2}$. We use the same first-order Butterworth filter in Section 4.1.

The minimum magnitude value of the in-band quantization noise is -34.2 dB. The obtained upper bound of the Lee criterion is $\gamma = 1.92$, which is equivalent to 5.7dB. It is larger than 1.5 used in the l_1 norm design in Table 1, since we impose a slight tighter constraint on the $||\epsilon||_\infty = 1.96 \times 10^{-2}$ than 2.59×10^{-2} in Table 1. The frequency response of the designed feedback filter is illustrated in Fig. 12.

5 Conclusions

We have proposed a design method of the FIR noise shaping filters of $\Delta\Sigma$ modulators based on H_2, H_∞, and l_1 norms. The minimization of the norm of the weighted quantization error is cast into a convex optimization problem by using LMIs, which can be efficiently and numerically solved. To ensure the stability of a $\Delta\Sigma$ modulator, we have also included LMI constraints which subsumes the Lee criterion. Our results show that the frequency response of our filters exhibits good performance throughout the low-frequency region providing uniform attenuation and matching the weighting function. Also, our proposed H_2, H_∞, and l_1 norm designed error feedback filters are shown to provide us with minimum H_2,

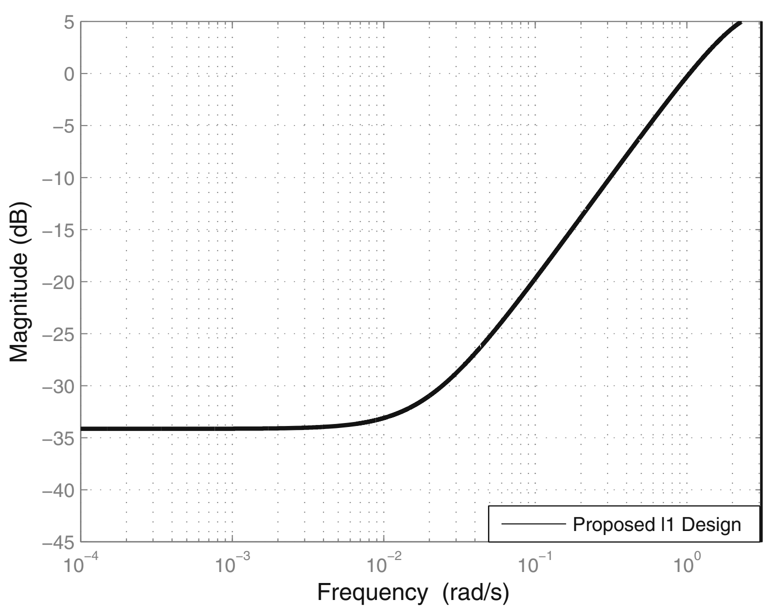

Fig. 12 Frequency response of the error feedback filter designed by minimizing the upper bound of the Lee coefficient under the constraint on the l_∞ norm of the weighted quantization noise

H_∞, and l_1 norms of the weighted quantization error, respectively, which shows the effectiveness of our proposed design method.

Competing interests

The authors declare that they have no competing interests.

Received: 16 June 2015 Accepted: 18 February 2016
Published online: 29 February 2016

References

1. J Proakis, *Digital Communications*, 4th. (McGraw-Hill, New York, 2001)
2. JC Candy, GC Temes, *Oversampling Delta-Sigma Data Converters*. (IEEE Press, New York, 1991)
3. KCH Chao, S Nadeem, WL Lee, CG Sodini, A higher order topology for interpolative modulator for oversampling A/D conversion. IEEE Trans. Circuits Syst. **37**, 309–318 (1990)
4. E Janssen, D Reefman, Super-audio CD: an introduction. IEEE Signal Process. Mag. **20**(4), 83–90 (2003)
5. U Zoler, *Digital Audio Signal Processing*, 2nd. (Wiley, New York, 2008)
6. R Schreier, GC Temes, *Understanding Delta-Sigma Data Converters*, 1st. (Wiley-IEEE Press, US, 2004)
7. Tran-Thong, B Liu, Error spectrum shaping in narrow-band recursive filters. IEEE Trans. Acoustics, Speech Signal Process. **25**(2), 200–203 (1977)
8. T Laakso, I Hartimo, Noise reduction in recursive digital filters using high-order error feedback. IEEE Trans. Signal Process. **40**(5), 1096–1107 (1992)
9. W Higgins, DC J Munson, Noise reduction strategies for digital filters: error spectrum shaping versus the optimal linear state-space formulation. IEEE Trans. Acoustics Speech Signal Process. **30**(6), 963–973 (1982)
10. X Li, C Yu, H Gao, Design of delta-sigma modulators via generalized Kalman-Yakubovich-Popov lemma. Automatica. **50**, 2700–2708 (2014)
11. M Nagahara, Y Yamamoto, Frequency domain min-max optimization of noise-shaping delta-sigma modulators. IEEE Trans. Signal Process. **60**(6), 2828–2839 (2012)
12. S Callegari, F Bizzarri, Output filter aware optimization of the noise shaping properties of delta-sigma modulators via semi-definite programming. IEEE Trans. Circuits Syst. I. **60**(9), 2352–2365 (2013)
13. S Callegari, F Bizzarri, Noise weighting in the design of delta-sigma modulators (with a psychoacoustic coder as an example). IEEE Trans. Circuits Syst. II: Express Briefs. **60**(11), 756–760 (2013)
14. KCH Chao, WLL S. Nadeem, CG Sodini, A higher order topology for interpolative modulators for oversampling a/d converters. IEEE Trans. Circuits Syst. **37**(3), 309–318 (1997)
15. S Boyd, LE Ghaoul, E Feron, *Linear Matrix Inequalities in System and Control Theory*. (Society for Industrial and Applied Mathematics, USA, 1997)
16. P Gahinet, P Apkarian, A linear matrix inequality approach to h_∞ control. Int. J. Robust Nonlinear Control. **4**, 421–448 (1994)
17. K Sawada, S Shin, Synthesis of dynamic quantizers for quantized feedback systems within invariant set analysis framework. Am. Control Conf. (ACC), 1662–1667 (2011)
18. S Ohno, Y Wakasa, M Nagata, Optimal error feedback filters for uniform quantizers at remote sensors. IEEE Int. Conf. Acoustics, Speech Signal Process, 3866–3870 (2015)
19. H Shingin, Y Ohta, in *10th IFAC/IFORS/IMACS/IFIP Symposium on Large Scale systems: Theory and Applications (LSS)*. Optimal invariant sets for discrete-time systems: approximation of reachable sets for bounded inputs, (2004), pp. 401–406
20. M Grant, S Boyd, Cvx: Matlab software for disciplined convex programming, version 2.0 beta (2012). http://cvxr.com/cvx/. Accessed date Feb 2015
21. R Schreier, Delta sigma toolbox [online]. http://jp.mathworks.com/matlabcentral/fileexchange/19-delta-sigma-toolbox. Accessed date Feb 2015