Proceedings of the 25th Chinese Control Conference
7–10 August, 2006, Harbin, Heilongjiang

Optimal AD-Conversion via Sampled-Data Receding Horizon Control Theory

Milan S. Derpich, Daniel E. Quevedo and Graham C. Goodwin
School of Electrical Engineering and Computer Science, The University of Newcastle, Australia
E-mails: milan.derpich@studentmail.newcastle.edu.au; dquevedo@ieee.org; graham.goodwin@newcastle.edu.au

Abstract: This paper presents novel results on the joint problem of sampling, reconstruction and quantization of analog signals. Existing literature on this topic deals exclusively with band-limited signals in sampled form. Our key departure from earlier results is that we deal with continuous time reconstruction of not necessarily band-limited signals. Our approach utilizes concepts and tools from optimal sampled-data and receding horizon control theory. The key conclusion from the work presented here is that, in the case under study, the optimal quantizer design problem can be partitioned into two sub-problems, namely (i) the design of an optimal analog pre-filter followed by sampling and (ii) an optimal quantizer, which works directly on the pre-sampled signals. Simulation results are presented which illustrate the performance of the optimal A-D converter designed via these principles.

Key Words: Sampling, quantization, frames, signal processing, sampled-data control.

1 INTRODUCTION

In many applications, one needs to convert analog, continuous time signals into quantized discrete time signals. This leads to an important set of questions regarding the best way to represent a signal by a sequence of sampled and quantized values, such that the information loss inherent in the sampling and quantization process is minimized in some sense. In the present work, we are interested in how to quantize a possible non band-limited signal to obtain the lowest possible reconstruction distortion.

We will show that, for a given sampling rate and reconstruction filters, minimization of reconstruction error, in an L^2 sense, can be converted into a discrete time problem. It turns out that if an appropriate pre-filter is used, then all the information required to find the optimal quantized sequence can *always* be extracted from discrete time samples of its output, even if the continuous time input signal is not band-limited.

Solving the optimal quantization problem amounts to finding the solution of a combinatorial optimization programme, which is in general computationally intractable. Our proposal is to convert the optimal quantization problem into a sampled-data moving horizon optimization problem with quantized decision variables. The proposed method gives excellent results and incurs only limited computational effort. It generalizes our previous work reported in [1][2][3][4] by concentrating on sampled-data signals rather than merely on discrete-time sequences.

Background to the work described here arises from distinct streams. The first of these is associated with the problem of sampling in the absence of quantization [5][6].

The second related field of research is concerned with quantization of signals where the sampling strategy has been pre ordained [7][1][8][9]. The third stream of prior work arises in the area of sampled data control theory. Here, the emphasis has typically been on regulation (zero reference) problems with unconstrained decision variables [10] [11]. In the present work we extend these concepts to account for non zero reference signals and quantized decision variables.

Our approach differs from the work described above by virtue of the fact that we design the joint optimal *sampler and quantizer* using sampled data quantized moving horizon optimization. This leads to significant performance gains, compared with alternative approaches which do not take account of the interaction between sampling and quantization.

The remainder of this work is organized as follows: In Section II we present the continuous time AD-conversion problem and how it can be translated into discrete-time. Section III introduces the continuous time receding horizon quantizer. Simulation studies are included in Section IV. Section V draws conclusions.

2 PROBLEM FORMULATION

The general form of the systems under study is illustrated in Fig. 1.

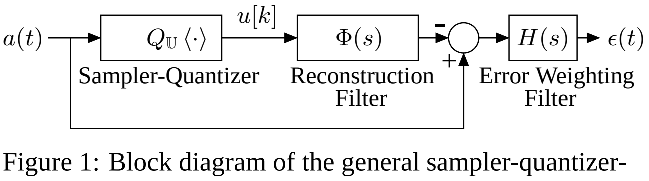

Figure 1: Block diagram of the general sampler-quantizer-reconstruction system.

The *sampler-quantizer* $\mathcal{Q}_{\mathbb{U}}\langle\cdot\rangle$ converts the continuous time signal $a(\cdot)$ into a sequence

$$u = \{u[k]\}_{k\in\mathbb{Z}}\,,\; u[k] \in \mathbb{U},\, \forall k \in \mathbb{Z} \tag{1}$$

where $\mathbb{U}$ is the finite and given set of scalars

$$\mathbb{U} = \{s_1, \ldots, s_{n_{\mathbb{U}}}\} \tag{2}$$

corresponding to the available quantization levels.

The sampling interval is constant and equal to τ seconds. Thus, quantized samples are generated at a rate of $1/\tau$ samples per second.

In Fig 1, the reconstruction filter Φ converts the discrete time sequence u into a continuous time signal. For example, in case of zero-order hold reconstruction, the impulse response of Φ would be $\phi(t) = \mu(t)+\mu(t-\tau)$. In the classical framework of perfectly band-limited reconstruction, Φ would be an ideal low-pass filter with cutoff frequency $1/2\tau$ [12]. On the other hand, in most practical applications, zero-order hold or some other form of short impulse response filter (sometimes non-causal) is generally used for reconstruction.

The filter H is the *error frequency weighting* filter (see, e.g., [4]). It allows one to represent the different impact of the error at different frequencies for a particular application. For example, if the system is employed for audio signals, then H could be designed to model the psycho-acoustical response of human hearing [13].

We are interested in designing a quantizer which minimizes the L^2 norm of the frequency weighed error ϵ (see Fig. 1), i.e., minimizes the cost function

$$V(\vec{u}) = \|\epsilon(\cdot)\|_{L^2} \tag{3}$$

where $\|\epsilon(\cdot)\|_{L^2}^2$ denotes the standard L^2 norm over the real line, i.e.

$$\|\epsilon(\cdot)\|_{L^2} = \int_{-\infty}^{\infty} \epsilon(\xi)d\xi \tag{4}$$

For the analysis below, it is more convenient to rearrange the system of Fig. 1 to the equivalent form shown in Fig. 2. In this figure, the continuous time filter Ψ is characterized by the transfer function

$$\Psi(s) \triangleq \Phi(s)H(s) \tag{5}$$

and its associated impulse response $\psi(\cdot)$.

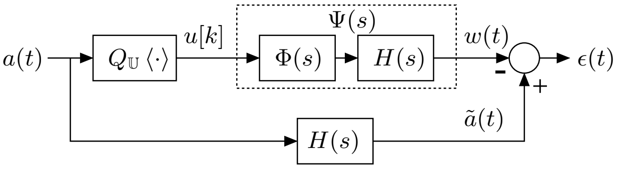

Figure 2: Equivalent block diagram

In order to represent the continuous time filtering performed by Ψ, we define the continuous time version of u as

$$u_c(t) \triangleq \sum_{k\in\mathbb{Z}} \delta(t-k\tau)u[k], \quad \forall t \in \mathbb{R} \tag{6}$$

The output of Ψ, namely $w(\cdot)$, is given by

$$w(t) = \int_{-\infty}^{\infty} \psi(t-\xi)u_c(\xi)d\xi, \quad \forall t \in \mathbb{R} \tag{7}$$

Substituting (6) into (7) one obtains

$$\begin{aligned} w(t) &= \int_{-\infty}^{\infty} \sum_{k\in\mathbb{Z}} \psi(t-\xi)\delta(\xi-k\tau)u[k]d\xi \\ &= \sum_{k\in\mathbb{Z}} \psi(t-k\tau)u[k] \end{aligned} \tag{8}$$

If we denote the impulse response of the frequency weighting filter H by $h(\cdot) \in L^2$, then its output satisfies

$$\tilde{a}(t) \triangleq (h * a)(t) \tag{9}$$

As a consequence, the frequency weighted error ϵ can be written as

$$\epsilon(t) = \tilde{a}(t) - w(t) = \tilde{a}(t) - \sum_{k\in\mathbb{Z}} \psi(t-k\tau)u[k], \; \forall t \in \mathbb{R} \tag{10}$$

Thus, the quest for optimal sampling and quantization can be stated as the optimization problem of finding the sequence $u^{\diamond}$ as in (1) that minimizes the L^2 norm of the reconstruction error, i.e.:

$$u^{\diamond} \triangleq \arg \min_{u[k]\in\mathbb{U},\, \forall k\in\mathbb{Z}} \left\| \tilde{a} - \sum_{k\in\mathbb{Z}} \psi(t-k\tau)u[k] \right\|_{L^2} \tag{11}$$

2.1 Reformulation in Discrete Time

It will now be shown that the L^2 (continuous time) optimization problem in (11) is equivalent to an ℓ^2 (discrete time) optimization problem, where the weighting values depend on the signal inter-sample behaviour. This is established in the following lemma, originally introduced without a formal proof in [2].

Lemma 1 Let the sequence of functions $\{\psi(\cdot - k\tau)\}_{k\in\mathbb{Z}}$ be a frame for $\overline{\text{span}}\{\psi(\cdot - k\tau)\}_{k\in\mathbb{Z}} = \mathcal{W} \subset L^2$, with frame bounds $0 \leq R \leq P < \infty$, i.e.

$$R\|w\|_{L^2}^2 \leq \sum_{k\in\mathbb{Z}} |\langle\psi(\cdot - k\tau), w\rangle|^2 \leq P\|w\|_{L^2}^2 \tag{12}$$

for all $w \in \mathcal{W}$. Let $w(\cdot) \in L^2$ be defined as in (8) for the sequence of scalars $\{u[k]\}_{k\in\mathbb{Z}} \in \ell^2$. Let the signal $\tilde{a} \in L^2$, and define $F[j,k]$ and $Y[k]$, $k, n \in \mathbb{Z}$ via

$$\begin{aligned} Y[j] &\triangleq \langle\tilde{a}, \psi(\cdot - j\tau)\rangle_{L^2} && (13)\\ F[j,k] &\triangleq \langle\psi(\cdot - j\tau), \psi(\cdot - k\tau)\rangle_{L^2} && (14)\\ & j,k \in \mathbb{Z} && (15) \end{aligned}$$

where $\langle\cdot,\cdot\rangle_{L^2}$ denotes the standard inner product in L^2. Then:

$$\|\epsilon(\cdot)\|_{L^2}^2 = \|\tilde{a}\|_{L^2}^2 - 2\sum_{j\in\mathbb{Z}} u[j]Y[j] + \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} u[j]u[k]F[j,k] \tag{16}$$

Proof 1 Substituting (10) in (3) we can write

$$V = \int_{-\infty}^{\infty} (\tilde{a}(t) - w(t))^2 \, dt = \langle \tilde{a} - w, \overline{\tilde{a} - w} \rangle_{L^2}$$
$$= \|\tilde{a}\|_{L^2}^2 - 2\langle \tilde{a}, \overline{w} \rangle_{L^2} + \langle w, \overline{w} \rangle_{L^2} \quad (17)$$

We note that the first term in the last line of (17) *is well defined since $\tilde{a} \in L^2$ as required by the Lemma. The second term is finite by virtue of the Cauchy-Schwartz inequality $\langle \tilde{a}, w \rangle \leq \|\tilde{a}\|_{L^2} \|w\|_{L^2}$ and the fact that $w \in \mathcal{W} \subset L^2$, which in turn implies the third integral is also bounded.*

Since inner products are by definition linear and $\tilde{a}$ and w are real signals, one obtains by substituting (8) *into* (17) *that*

$$V = \|\tilde{a}\|_{L^2}^2 - 2\sum_{k\in\mathbb{Z}} u[k] \langle \tilde{a}, \psi(t - k\tau) \rangle_{L^2} + \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} u[j]u[k] \langle \psi(t - j\tau), \psi(t - k\tau) \rangle_{L^2} \quad (18)$$

which is equivalent to (16).

Remark 1 *If we make the change of variables*

$$e[j] \triangleq u[j] - u^\star[j], \forall j \in \mathbb{Z} \quad (19)$$

where $u^\star$ is an un-quantized sequence that yields the global minimum of V in (16)*, namely $V^\star$, then*

$$V = V^\star + \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} e[j]e[k]F[j,k] \quad (20)$$

Remark 2 *From lemma 1 and remark 1, it is clear that to minimize* (3) *the only information needed is an optimal un-quantized sequence $u^\star$ (or, alternatively, the sequence $\{Y[k]\}_{k\in\mathbb{Z}}$) and the coefficients $F[j,k]$. The latter corresponds to samples of the autocorrelation function of $\psi(\cdot)$, which can be determined off-line after choosing reconstruction and error weighting filters and then incorporated to the quantization algorithm. On the other hand, $\{Y[k]\}_{k\in\mathbb{Z}}$ can be obtained from $u^\star$ by differentiating* (16) *with respect to u and equating to zero, which leads to*

$$Y[j] = \sum_{j\in\mathbb{Z}} u^\star[k]F[j,k], \quad \forall j \in \mathbb{Z} \quad (21)$$

2.2 Pre-filter Requirements for Optimal Quantization

In practice, any quantization algorithm has to work with discrete-time values. Remarks 1 and 2 arise the need to determine whether a quantization algorithm can elaborate or obtain $u^\star$ or $\{Y[k]\}_{k\in\mathbb{Z}}$ from samples of the input signals. Consider, first, the determination of the series of coefficients $\{Y[k]\}_{k\in\mathbb{Z}}$. From definition (13), we have

$$\begin{aligned} Y[k] &= \langle \tilde{a}(\cdot), \psi(t - k\tau) \rangle (k\tau) = (\tilde{a} * \psi^\vee)(k\tau) \\ &= (\tilde{a} * h * [h * \phi]^\vee)(k\tau) \\ &= (\tilde{a} * h * h^\vee * \phi^\vee)(k\tau), \forall k \in \mathbb{Z} \end{aligned} \quad (22)$$

From the last line of (22), it is clear that the series of coefficients $\{Y[k]\}_{k\in\mathbb{Z}}$ can be obtained by passing the input signal through a filter with frequency response $G_Y(j\omega)$ given by

$$G_Y(j\omega) \triangleq |H(j\omega)|^2 \Phi^*(j\omega) = H(j\omega)\Psi^*(j\omega) \quad (23)$$

and then taking the samples every τ seconds. i.e., if we denote the impulse response of G_Y by $g_Y(\cdot)$, then $Y[j] = (a * g_Y)(j\tau)$, $\forall j \in \mathbb{Z}$,

Let us next consider the determination of $u^\star$, the sequence of samples which minimizes reconstruction error in the absence of quantization. It is known from sampling theory [14][15] that, for *any* input signal, $u^\star$ can be obtained by sampling the output of a pre-filter $G_S(j\omega)$ matched to the reconstruction filter. From these results, for the system depicted in Fig. 2, the ideal matched pre-filter for a given reconstruction filter $\Psi(j\omega)$ is given by

$$G_S(j\omega) = \begin{cases} \frac{|H(j\omega)|^2 \Phi^*(j\omega)}{\mathcal{A}_\Psi(e^{j\omega\tau})} & \text{, if } \mathcal{A}_\Psi(e^{j\omega\tau}) \neq 0 \\ 0 & \text{, if } \mathcal{A}_\Psi(e^{j\omega\tau}) = 0 \end{cases} \quad (24)$$

where

$$\mathcal{A}_\Psi(e^{j\omega\tau}) \triangleq \tfrac{1}{\tau} \sum_{k\in\mathbb{Z}} \left| \Psi(j[w + \tfrac{2\pi}{\tau}k]) \right|^2 \quad (25)$$

is the discrete time Fourier transform of the sampled autocorrelation function of $\psi(\cdot)$ [5]. Notice that $H(j\omega) = 0 \, \forall \omega \in \{w : \mathcal{A}_\Psi(e^{j\omega\tau}) = 0\}$.

The above results suggest that all the necessary information about the input signal for optimal quantization to be feasible can be obtained from samples of the filtered input signal, and that the required pre-filter is not unique. We will provide next necessary and sufficient conditions for a pre-filter to yield samples that allow for optimal quantization.

Consider the discrete Fourier transform of $u^\star$, and let $\hat{f}(e^{j\omega\tau})$ and $\hat{g}(j\omega)$ denote the discrete and continuous Fourier transforms of any $f \in \ell^2$ and $g \in L^2$, respectively. Since $u^\star$ is a sequence of samples of a filtered by $G_S(j\omega)$, we have

$$\widehat{u^\star}(e^{j\omega\tau}) = \tfrac{1}{\tau} \sum_{k\in\mathbb{Z}} G_S(j[\omega - \tfrac{2\pi}{\tau}k]) \widehat{a}(j[\omega - \tfrac{2\pi}{\tau}k]) \quad (26)$$

Suppose the quantizer gets samples v of the input signal pre-filtered by another filter $G_X(j\omega)$. The discrete Fourier transform of such sequence of samples would be

$$\widehat{v}(e^{j\omega\tau}) \triangleq \tfrac{1}{\tau} \sum_{k\in\mathbb{Z}} G_X(j[\omega - \tfrac{2\pi}{\tau}k]) \widehat{a}(j[\omega - \tfrac{2\pi}{\tau}k]) \quad (27)$$

Recovery of $\widehat{u^\star}(e^{j\omega\tau})$ can be achieved in the discrete-time domain by applying a discrete-time filter $\Gamma(e^{j\omega\tau})$ to the

There might exist more than one optimal sequence if the reconstruction stage is redundant.

sequence v, such that

$$\begin{aligned}\widehat{u^{\star}}(e^{j\omega\tau}) &= \Gamma(e^{j\omega\tau})\widehat{v}(e^{j\omega\tau}) \\ &= \tfrac{1}{\tau}\Gamma(e^{j\omega\tau})\sum_{k\in\mathbb{Z}} G_X(j[\omega - \tfrac{2\pi}{\tau}k])\widehat{a}(j[\omega - \tfrac{2\pi}{\tau}k]) \\ &= \tfrac{1}{\tau}\sum_{k\in\mathbb{Z}} \Gamma(e^{j\omega\tau})G_X(j[\omega - \tfrac{2\pi}{\tau}k])\widehat{a}(j[\omega - \tfrac{2\pi}{\tau}k])\end{aligned} \tag{28}$$

From (26) and (28) it can be seen that a sufficient and necessary condition is the existence of a periodic transfer function $\Gamma(e^{j\omega\tau})$, such that

$$G_X(j\omega) = \frac{G_S(j\omega)}{\Gamma(e^{j\omega\tau})} \quad , \quad \forall w \in \{w : \widehat{a}(j\omega) \neq 0\} \tag{29}$$

$$K_1 \leq \left|\Gamma(e^{j\omega\tau})\right| \leq K_2 \quad , \quad \forall w \in \{w : \widehat{a}(j\omega) \neq 0\} \tag{30}$$

for the quantizer stage to be able to determine $u^{\star}$ from the samples v (and allow for optimal quantization). As a particular case, if $\mathcal{A}_{\Psi} \equiv 1$, then, from (23) and (24), $G_S(j\omega) = G_Y(j\omega), \forall w \in \mathbb{R}$ and $Y[j] = u^{\star}[j]$, $\forall j \in \mathbb{Z}$. The latter equality can also be obtained from (21) by noting that $\mathcal{A}_{\Psi} \equiv 1$ if and only if $F[j,k] = \delta_{j,k}, \forall j,k \in \mathbb{Z}$.

Of course the quantizer stage would need to implement the correction filter $\Gamma(e^{j\omega\tau})$ based upon knowledge of $G_S(j\omega)$ and $G_X(j\omega)$, according to (29). Two important special cases are to be highlighted:

- If a has no frequency components beyond π/τ [rad/s], then any pre-filter G_X satisfying

  $$\begin{aligned} &C_1 \leq |G_X(j\omega)| \leq C_2, \\ &\qquad \forall w \in \{w : \widehat{a}(j\omega) \neq 0\,,\, G_S(j\omega) \neq 0\}\end{aligned} \tag{31}$$

  for some constants $0 < C_1 \leq C_2 < \infty$ would make optimal quantization possible from the samples v. This result is not surprising, since by Shannon's sampling theorem, the samples of a band-limited signal contain all the information about the complete signal [12].

- If Ψ is band-limited to a frequency $\alpha = \pi/\tau$ but a has energy at frequencies greater than α, then any pre-filter G_X band-limited exactly to α satisfying

  $$\begin{aligned} K_1 \leq |G_X(j\omega)| \leq K_2,\ \forall w \in \{w : G_S(j\omega) \neq 0\} \\ G_X(j\omega) = 0,\ \forall w \in \{w : G_S(j\omega) = 0\}\end{aligned} \tag{32}$$

  for some constants $0 < K_1 \leq K_2 < \infty$ would have a feasible correction filter that makes optimal quantization possible from the samples v.

The conclusion from the above cases is that quantization for optimal reconstruction of an input signal not band-limited to π/τ is possible, but demands either the use of an appropriate pre-filter satisfying (29) to get the samples from, or, alternatively, the quantizer needs to "know" the signal between sampling instants.

3 THE SAMPLED-DATA RECEDING HORIZON QUANTIZER

For the general case, minimization of (16) would require the evaluation of (16) for every possible sequence $\{u[k]\}_{k\in\mathbb{Z}}$, $u[k] \in \mathbb{U}$. For sufficiently long sequences, the optimization programme becomes computationally intractable. To overcome this problem, we propose to use concepts from the receding horizon control framework [16] and optimize over a short receding horizon of samples. A quantizer based on this idea has been recently proposed by the current authors in [3][4] for an all discrete-time system, achieving near optimal performance with rather short horizon lengths [17]. In what follows we will extend this idea to the sampled data case, and show, via simulations, that significant distortion reduction is obtained when converting non band-limited signals.

3.1 Optimal Quantization Over a Finite Horizon

Consider the system at $t_\ell \triangleq \ell\tau, \ell \in \mathbb{Z}$. Instead of attempting to optimize the cost over $t \in \mathbb{R}$, we will aim to minimize the cost within a finite time interval $[(\ell - M)\tau, (\ell + N)\tau)$, where $M, N \in \mathbb{Z}^{+}$ are design parameters. We will only concentrate on the optimal sequence of quantized coefficients to be generated for the interval $[\ell\tau, (\ell + N)\tau)$, defined as

$$\vec{u}_\ell \triangleq (u[\ell]\ u[\ell+1]\ \ldots\ u[\ell+N-1])^T \tag{33}$$

The number M, therefore, accounts for the non-causality of Ψ.

For the purpose of including in the horizon the effect of past errors, it is convenient to describe the continuous time filter Ψ by its state space representation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u_c(t) \tag{34}$$

$$w(t) = \mathbf{C}\mathbf{x}(t + \lambda) \tag{35}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times 1}$, $\mathbf{C} \in \mathbb{R}^{1\times n}$ and $u_c(\cdot)$ is as defined in (6). In (35), $\lambda \geq 0$ accounts for a possible non-causal Ψ. If Ψ is causal, then $\lambda = 0$.

The cost function to minimize is

$$V_N(\vec{u}_\ell) = \int_{(\ell-M)\tau}^{(\ell+N)\tau} (\tilde{a}(t) - w(t))^2\, dt \tag{36}$$

If $\mathbf{x}(t_\ell)$ is known, then, from (34) and (35), $w(t)$ can be determined by

$$w(t) \triangleq \sum_{k=0}^{N-1} \psi(t\text{–}[k{+}\ell]\tau)u_\ell[k] + \mathbf{C}\mathrm{e}^{\mathbf{A}(t+\lambda-\ell\tau)}\mathbf{x}(t_\ell) \tag{37}$$

for $t \in [t_\ell - M\tau,\, t_\ell + N\tau)$.

Notice that the first term on the right hand side of (37) captures the effect of the choice of $\vec{u}_\ell$ within the horizon. This term corresponds to the forced response of Ψ to the input $\vec{u}_\ell$, which can be conveniently represented as

$$w_\ell(r) \triangleq \sum_{k=0}^{N-1} \psi(r - k\tau)u_\ell[k],\ r \in [0, N\tau) \tag{38}$$

On the other hand, the second term on the right hand side of (37) represents the natural response of Ψ when the initial state is $\mathbf{x}(t_\ell)$. We define the difference between this initial state response and the filtered input signal $\tilde{a}(t)$ within the horizon as the *target function* $y_\ell(\cdot)$, for the horizon at $t = \ell\tau$:

$$y_\ell(r) \triangleq \tilde{a}(r+t_\ell) - \mathbf{C}e^{\mathbf{A}(r+\lambda)}\mathbf{x}(t_\ell),\ r \in [0, N\tau) \quad (39)$$

By using (38) and (39) , the cost (36) can be expressed as the L^2 norm of the difference between the zero-initial-state response $w_\ell(\cdot)$ and the target function $y_\ell(\cdot)$:

$$V_N(\vec{u}_\ell) = \int_{-M\tau}^{N\tau} (y_\ell(x) - w_\ell(x))^2\, dx \quad (40)$$

Substituting (38) into (40) one obtains

$$V_N(\vec{u}_\ell) = \int_{-M\tau}^{N\tau} \left(y_\ell(x) - \sum_{k=0}^{N-1} \psi(x-k\tau)u_\ell[k] \right)^2 dx \quad (41)$$

Since $\psi(\cdot) \in L^2$, one can exchange the order of sum and integration in (41) and rewrite it in matrix form as

$$V_N(\vec{u}_\ell) = \vec{u}_\ell^T \mathbf{F}_N \vec{u}_\ell - 2\mathbf{Y}_\ell^T \vec{u}_\ell + \int_{-M\tau}^{N\tau} y_\ell^2(x)dx \quad (42)$$

where the vector $\mathbf{Y}_\ell \in \mathbb{R}^N$ and the symmetric, positive definite matrix $\mathbf{F}_N \in \mathbb{R}^{N\times N}$ are defined element-wise as

$$F_N[j,k] \triangleq \int_{-M\tau}^{N\tau} \psi(x-j\tau)\psi(x-k\tau)dx \quad (43)$$

$$Y_\ell[j] \triangleq \int_{-M\tau}^{N\tau} y(x)\psi(x-j\tau)dx \quad (44)$$

$$j,k = 0,1,\dots N-1 \quad (45)$$

3.2 The Sampled-Data Receding Horizon Quantizer

The expressions derived above for the cost function over a finite horizon allow us to introduce the sampled-data receding horizon quantizer. The algorithm finds, at a given instant $\ell\tau$, the vector of quantized coefficients $\vec{u}_\ell$ that minimizes the total filtered reconstruction error from $(\ell - M)\tau$ to $(\ell + N)\tau$ defined in (36). Then, the first element of $\vec{u}_\ell$ is sent to the output of the quantizer. The horizon is then shifted forward by τ, and iteration $\ell + 1$ begins.

The proposed algorithm, beginning at instant $\ell\tau$, can be formalized as follows:

Step 1.- Calculate the matrix $\mathbf{F}_N$ in (43)
Step 2.- Calculate $\mathbf{Y}_\ell$
Step 3.- Find the optimizer $u_\ell^\diamond$ by minimizing (42)
Step 4.- Output $u_\ell^\diamond[0]$, the first element of $u_\ell^\diamond$ (see (33))
Step 5.- Increment ℓ by 1 and go to Step 2.

The sequence $\dots, u_{\ell-1}^\diamond[0], u_\ell^\diamond[0], u_{\ell+1}^\diamond[0], \dots$ of step 4 forms the output of the sampled-data receding horizon sampler quantizer. If the input signal is not band-limited to π/τ, the algorithm reduces filtered aliasing and frequency weighted quantization noise. For that purpose, it does simultaneous adaptive filtering on the input signal and adaptive noise shaping of the quantization noise, thus respecting the interaction between both phenomena.

It is interesting to note that, as the horizon is made larger, the output of the sampler-quantizer defined above approaches the optimal feasible output sequence possible defined in (11).

4 SIMULATION STUDY

We will first show an example comparing the performance of the proposed sampled-data receding horizon quantizer (SDRHQ) against the so called *all-discrete-time* (DTRHQ) receding horizon quantizer introduced in [3] in the following situation:

- The input signal, a, is an audio signal that has frequency components up to 22 [kHz]. Its frequency energy content is shown in Fig. 3.

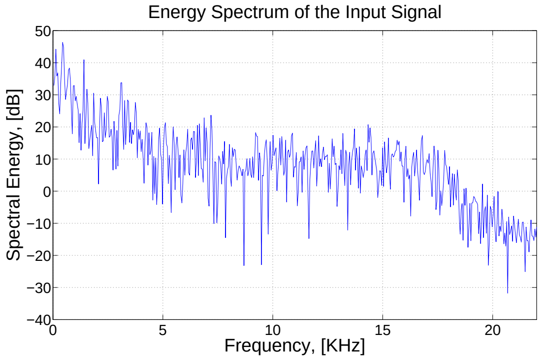

Figure 3: Spectral composition of the input signal $a(\cdot)$.

- The sampling frequency is half the required to avoid aliasing i.e., $1/\tau = 11$ [kHz].
- The filter H performs zero-order hold reconstruction, i.e., it has impulse response $\phi(t) = \mu(t) + \mu(t-\tau)$.
- The frequency error weighting filter corresponds to the third order model for the psycho-acoustical response of the human ear [13]. Its frequency response is shown in Fig.4
- No pre-filtering is used, i.e., $G_X(j\omega) = 1$.

In the simulation, the DTRHQ has full knowledge of the filters H and Φ, and utilizes the matrix $\mathbf{F}_N$ defined in (43). However, it operates based on direct samples of the input signal. Thus, since the implicit pre-filter is a unity gain, (29) predicts that the all-discrete-time quantizer will be unable to determine the target function to be approximated by the reconstruction stage. On the other hand, the SDRHQ

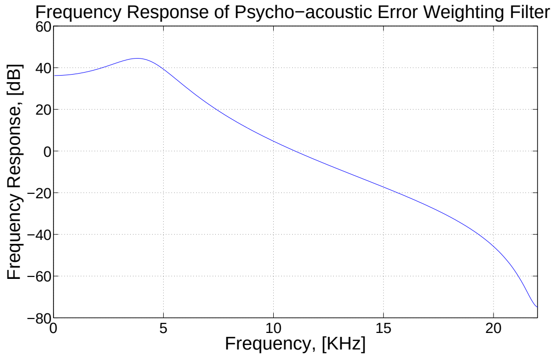

Figure 4: Frequency response of filter H.

utilizes the same matrix $\mathbf{F}_N$ but has access to the inter-sample behaviour of the input signal.

Fig. 5 shows the normalized reconstruction error from the outputs of both quantizers, for several horizon lengths. It can clearly be appreciated how, in this case, the sampled-data converter proposed in the present work outperforms the all-discrete-time converter of [3]. It can also be seen that the distortion exhibits a small decrease with the increase of the horizon length N. This suggests that, in this example, the main contributor to the distortion is aliasing noise, for both converters. Notice that the SDRHQ operated without an anti-aliasing filter. This, together with its much lower distortion in comparison with that of the DTRHQ, suggests that the sampled-data receding horizon optimization algorithm accomplishes a form of pre-filtering of the input signal that effectively reduces aliasing.

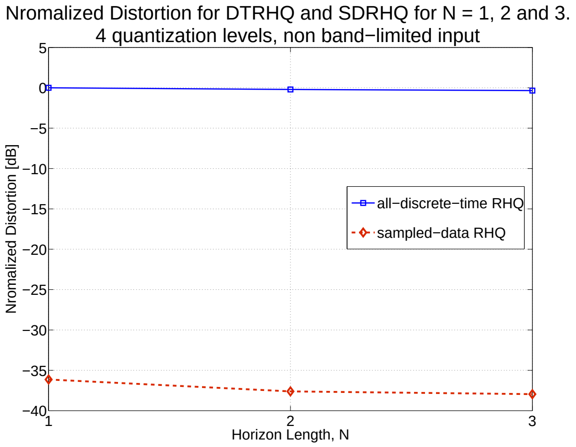

Figure 5: Reconstruction error (normalized to DTRHQ distortion for $N = 1$) from the outputs of all-discrete-time quantizer (DTRHQ) and sampled-data quantizer (SDRHQ), for N = 1 to 3, for a non band-limited input.

5 CONCLUSIONS

This paper has shown how receding horizon sampled-data control methods can be utilized to design optimal AD Converters. A key departure from earlier results in this area is that we optimize a filtered version of the continuous time reconstruction error for not necessarily band limited input signals. Inter alia, we show that the optimal design problem can be decomposed into two subproblems, namely the design of an optimal analogue pre-filter together with an optimal quantizer . We showed that the latter is feasible if only samples of the signals are available. The efficacy of the proposed method has been illustrated by an example using an audio signal sampled below its Nyquist rate.

References

[1] G. C. Goodwin, D. E. Quevedo, and D. McGrath, “Moving-horizon optimal quantizer for audio signals,” *J. Audio Eng. Soc.*, vol. 51, no. 3, pp. 138–149, Mar. 2003.

[2] M. S. Derpich, D. E. Quevedo, G. C. Goodwin, and A. Feuer, “Quantization and sampling of not necessarily band-limited signals,” to appear in Proc. of the International Conf. on Audio, Speech and Signal Proc. ICASSP-2006.

[3] D. E. Quevedo and G. C. Goodwin, “Audio quantization from a receding horizon control perspective,” in *Proc. Amer. Contr. Conf.*, 2003, pp. 4131–4136.

[4] ——, “Multi-step optimal analog-to-digital conversion,” School of Elect. Eng. and Comput. Sci., The Univ. of Newcastle, NSW 2308, Australia, Tech. Rep. EE03032, 2003.

[5] M. Unser, “Sampling – 50 years after Shannon,” in *Proc. IEEE*, vol. 88, no. 4, April 2000.

[6] P. Vaidyanathan, “Generalizations of the sampling theorem: Seven decades after Nyquist,” *IEEE Trans. Circuits Syst. I*, vol. 48, no. 9, pp. 1094–1109, September 2001.

[7] S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta–Sigma Data Converters: Theory, Design and Simulation*. Piscataway, N.J.: IEEE Press, 1997.

[8] D. E. Quevedo and G. C. Goodwin, “An improved architecture for networked control systems,” in *Proc. 16th IFAC World Congress, Prague, Czech Republic*, 2005.

[9] ——, “Finite alphabet constraints in engineering,” 2004, in preparation.

[10] T. Chen and B. A. Francis, *Optimal Sampled-Data Control Systems*. London: Springer-Verlag, 1995.

[11] A. Feuer and G. C. Goodwin, *Sampling in digital signal processing and control*. Cambridge, Mass.: Birkäusser Boston, 1996.

[12] C. E. Shannon, “Communication in presence of noise,” in *Proc. IRE*, vol. 37, 1949, pp. 10–21.

[13] R. A. Wannamaker, “Psycho-acoustically optimal noise shaping,” *J. Audio Eng. Soc.*, vol. 40, no. 7/8, pp. 611–620, July/Aug. 1992.

[14] A. Aldroubi and M. Unser, “Sampling procedures in function spaces and asymptotic equivalence with Shannon’s sampling theory,” *Numer. Funct. Anal. Optimizat.*, vol. 15,, no. 1-2, pp. 1–21, Feb 1994.

[15] O. Christensen and Y. C. Eldar, “Oblique dual frames and shift-invariant spaces,” *Appl. Comput. Anal.*, vol. 17, no. 1, pp. 46–48, Jul. 2004.

[16] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, “Constrained model predictive control: Optimality and stability,” *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[17] D. E. Quevedo and G. C. Goodwin, “Multistep optimal analog-to-digital conversion,” *IEEE Trans. Circuits Syst. I*, vol. 52, Issue 3, pp. 503–515, March 2005.