Signal Reconstruction via $H^\infty$ Sampled-Data Control Theory—Beyond the Shannon Paradigm

Yutaka Yamamoto, Fellow, IEEE, Masaaki Nagahara, Member, IEEE, and Pramod P. Khargonekar, Fellow, IEEE

Abstract—This paper presents a new method for signal reconstruction by leveraging sampled-data control theory. We formulate the signal reconstruction problem in terms of an analog performance optimization problem using a stable discrete-time filter. The performance criterion naturally takes intersample behavior into account. We present methods for computing optimal solutions which are guaranteed to be stable and causal. Detailed comparisons to alternative methods are provided. We discuss some applications in sound and image reconstruction.

Index Terms—Sampled-data control theory, digital signal processing, Shannon paradigm, sampling theorem, $H^\infty$ control theory

I. INTRODUCTION

SIGNAL reconstruction from digital data is at the foundations of digital signal processing. For given digital data, stored or transmitted, we need to recover the original signal that generated the data. This procedure is needed and used everywhere: image and sound reconstruction/restoration, moving images, mobile telephones, etc. While discussion of transmission/recovery for digital data only is quite routine, we should note that our ultimate objective is to recover or reconstruct the original analog data from which such digital data are generated.

This signal reconstruction problem dates back to the celebrated paper of Shannon [26]. Using the sampling theorem [42], he showed that we can recover the original analog information from sampled data, provided that the original analog signal is band-limited below the Nyquist frequency. Based on this result, he established a fundamental paradigm for communication and digital processing. We will hereafter refer to this scheme the Shannon paradigm.

This paradigm however leads to various unrealistic constraints. The assumption of perfect band-limitedness, necessary for perfect signal reconstruction, is hardly satisfied in reality. In many applications, the sampling rate is not high enough to allow for this assumption to hold even approximately. To remedy this drawback, one often introduces an anti-aliasing filter to sharply cut high frequency components, but this in turn leads to yet another type of distortion due to the Gibbs phenomenon (see Example 1 in Section V and also Section VII below). Secondly, the sinc function, which is the impulse response of the ideal filter, is not causal and does not decay very fast. This slow decay rate makes it very difficult to implement and various approximations (mostly with respect to the $H^2$-norm criterion) become necessary. This procedure further complicates the total design procedure, making it less transparent.

In view of such drawbacks, there has been revived interest in the extension of the sampling theorem in various forms since the 1990’s. There is by now a stream of papers that aim to study signal reconstruction under the assumption of non-ideal signal acquisition devices; an excellent survey is given in [29]. In this research framework, the incoming signal is acquired through a non-ideal analog filter (acquisition device) and sampled, and then the reconstruction process attempts to recover the original signal. The idea is to place the problem into the framework of the (orthogonal or oblique) projection theorem in a Hilbert space (usually $L^2$), and then project the signal space to the subspace generated by the shifted reconstruction functions. It is often required that the process give a consistent result, i.e., if we subject the reconstructed signal to the whole process again, it should yield the same sampled values from which it was reconstructed [27].

The objective of the present paper is to go beyond the Shannon paradigm, and present an entirely new scheme of signal reconstruction, different from the ones described above. For our scheme, we draw upon modern sampled-data control theory, developed in the control community since the 1990’s. The fundamental accomplishment of modern sampled-data control theory is that it can give us a discrete-time controller (or filter) that optimizes the closed-loop performance with intersample behavior taken into account. In other words, it can optimize an analog (continuous-time) performance. Such a performance is measured according to the $H^\infty$ or $H^2$ performance criteria. This setting gives us an optimal platform to reconstruct the original analog signals from sampled-data under the scenario that the original signal is not band-limited below the Nyquist frequency.

Chen and Francis [7] made a first attempt to apply sampled-data control theory to signal processing (however in a discrete-time domain); see also [12]. Starting in 1995, the present authors and our colleagues have pursued the signal reconstruction problem in the sampled-data context to obtain an optimal analog performance via digital filtering: See [18], [40], [35] for general design frameworks, [13], [23] for sample-rate

The basic approach is as follows: We start with a signal generation model that consists of an analog filter with \( L^2 \) inputs. Then we formulate the signal reconstruction problem as a sampled-data \( H^\infty \) control design problem. The controller to be designed is the digital filter that is desired to reconstruct the original analog signal. The advantage here is that we can formulate an overall error system, and be able to have control over all frequencies including both gain and phase errors, not merely the gain characteristics often observed in many filter designs. Introducing an upsampler, this framework also enables us to optimally interpolate the intersample high frequency components based on the model of the signal generator. We will formulate and discuss this in more detail in Section IV.

The same philosophy of emphasizing the importance of analog performance was proposed and pursued recently by Unser and co-workers [30], [31]. The crucial difference is however that they rely on many filter designs. Introducing an upsampler, this framework can control over all frequencies including both gain and phase errors. Due to this effect, one usually introduces an amplification at another range. Due to this effect, one usually introduces a

\[ \|f\|_2 := \sup_{x > 0} \left\{ \frac{1}{2\pi} \int_{x-j\infty}^{x+j\infty} |f(x + iy)|^2 dy < \infty \right\}^{1/2} \quad (2) \]

It is well known that Laplace transform gives an isometry between \( L^2(0, \infty) \) and \( H^2 \).

The space \( H^\infty \) denotes the Hardy space of functions analytic on \( \mathbb{C}_+ \) and bounded there. It is a Banach space with norm

\[ \|f\|_\infty := \sup_{s \in \mathbb{C}_+} |f(s)|. \quad (3) \]

An element \( f \) of \( H^\infty \) admits nontangential limit to the imaginary axis almost everywhere, which we denote by \( f(j\omega) \), \( \omega \in \mathbb{R} \). Then the \( H^\infty \)-norm of \( f \in H^\infty \) is equal to

\[ \|f\|_\infty = \text{esssup}_{-\infty < \omega < \infty} |f(j\omega)|. \quad (4) \]

Now let \( G \) be the transfer function of a finite-dimensional, asymptotically stable linear time-invariant system. Then \( G \) belongs to \( H^\infty \), and its “size” is measured by the \( H^\infty \)-norm, i.e., the supremum (or maximum) of the Bode magnitude plot as in (4).

The steady-state response of \( G \) against a sinusoid \( e^{j\omega t} \) is given by \( G(j\omega)e^{j\omega t} \), and its magnitude is bounded as

\[ |G(j\omega)e^{j\omega t}| \leq \sup_{-\infty < \omega < \infty} |G(j\omega)| \cdot |e^{j\omega t}| = \|G\|_\infty. \]

In general, for \( u \in H^2 \), it is known that

\[ \|Gu\|_2 \leq \|G\|_\infty \cdot \|u\|_2, \quad (5) \]

and this bound is tight. Hence the \( H^\infty \) norm gives the \( L^2 \) energy induced-gain, and minimizing it yields a system that works uniformly well for the whole frequency range.

For this reason, it is recognized that the \( H^\infty \)-norm criterion is often superior to the \( H^2 \)-norm criterion, where the \( H^2 \)-norm for a stable matrix transfer function is defined as

\[ \|G\|_2 := \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace} \{G^*(j\omega)G(j\omega)\} dt \right)^{1/2}. \]

The \( H^\infty \) norm has been used successfully in the control literature [8], [9], [11].

The \( H^\infty \) norm criterion is naturally extended to sampled-data systems. The problem here is that such systems have two time sets: continuous and discrete. Hence the overall system is not time-invariant in the classical sense. This difficulty can be remedied by the now-standard technique called lifting, which converts a linear time-invariant continuous-time system to an infinite-dimensional discrete-time system. It is then possible to naturally extend the notion of the \( H^\infty \)-norm to sampled-data systems. To be more precise, let \( \mathcal{G} \) denote the input/output operator of such a system. Then its \( H^\infty \)-norm is defined to be the induced norm against all \( L^2 \) inputs:

\[ \|\mathcal{G}\|_\infty := \sup_{u \in L^2, u \neq 0} \frac{\|\mathcal{G}u\|_2}{\|u\|_2}. \quad (6) \]

However, it is to be noted that it is not possible to uniformly attenuate \( |G(j\omega)| \). If we attenuate \( G(j\omega) \) for a certain frequency range, it will yield an amplification at another range. Due to this effect, one usually introduce a frequency weighting \( W(s) \), and minimize \( |W(j\omega)G(j\omega)| \).

II. PRELIMINARIES

Let us introduce some basic function spaces and performance measures. Let \( L^2(a, b) \) or \( L^2(a, b), L^2[a, b], \) etc. be the space of Lebesgue square integrable functions on the interval \( (a, b) \), \( a < b \). For a function \( f \) valued in \( \mathbb{R}^n \) or \( \mathbb{C}^n \), its \( L^2 \)-norm is denoted by

\[ \|f\|_2 = \left\{ \int_a^b |f(t)|^2 dt \right\}^{1/2}, \quad (1) \]

where \(| \cdot | \) denotes the Euclidean norm in \( \mathbb{C}^n \). Let \( H^2 \) denote the space of \( \mathbb{C}^n \)-valued functions \( f \) that are analytic on the open right half plane \( \mathbb{C}_+ := \{ s : \text{Re } s > 0 \} \) and satisfy

\[ \sup_{x > 0} \int_{x-j\infty}^{x+j\infty} |f(x + iy)|^2 dy < \infty. \]
Via lifting, this norm is shown to be equivalent to the maximum gain of the frequency response operator of $G$ as in (3).

For details, see Appendix I.

III. SIGNAL RECONSTRUCTION AND SAMPLING THEOREM

Consider the block diagram depicted in Fig. 1:

\[\begin{array}{c}
\text{w}_c \\
\xrightarrow{F} \\
\text{y}_c \\
\xrightarrow{\text{K}(z)} \\
\text{y}_{K,d} \\
\xrightarrow{\Phi} \\
y
\end{array}\]

Fig. 1. Signal Reconstruction System

In this diagram, the signal $w_c \in L^2$ denotes the external analog signal to be reconstructed. It is filtered by an analog filter (acquisition device) $F$, and then sampled by the sampler with sampling period $h$. If $f(t)$ denotes the impulse response of the analog filter $F$, then the discrete-time signal $y_d[k]$ is easily seen to be given by

$$y_d[k] = (f * w_c)[k] = \langle f(-kh), w_c \rangle$$

(7)

where $\tilde{f}(t) := f(-t)$ is the mirror image (with respect to time) of $f$. The obtained signal $y_d$ is then processed by a discrete-time filter $K$ and then the filtered discrete-time signal $y_{K,d}$ is converted back to an analog signal $y$ via a reconstruction device $\Phi$. Denoting by $\phi$ the impulse response of $\Phi$, the reconstructed $y$ is given by

$$y(t) = \sum_{k=-\infty}^{\infty} y_{K,d}[k] \phi(t - kh).$$

(8)

In the Shannon paradigm, the analog filter $F$ is taken to be the ideal filter, and $\phi$ above is the sinc function [42], [29]. As mentioned in the Introduction, this has several limitations. To take care of this, one often employs an approximation of the ideal filter with respect to $H^2$ norm [10], and this unfortunately yields a sharp ringing effect in the frequency domain.

Unser and co-workers published series of papers of generalized sampling theorems where the acquisition device $F$ is not the ideal filter [27], [28], [29]. First define the subspace

$$V_f := \{ \sum_{k=-\infty}^{\infty} \alpha[k] f(t - kh) : \{\alpha[k]\} \in \ell^2 \}$$

(9)
generated by the translates of the impulse response of the acquisition filter, and the reconstruction space

$$V_\phi := \{ \sum_{k=-\infty}^{\infty} \beta[k] \phi(t - kh) : \{\beta[k]\} \in \ell^2 \}$$

(10)
generated by the translates of the reconstruction function $\phi$. From the consistency requirement [27], a key step in their procedure is the oblique projection of $L^2$ onto $V_\phi$ perpendicular to $V_f$. A precise comparison of this approach with our work is given in Section VI.

IV. $H^\infty$ SIGNAL RECONSTRUCTION PROBLEM

We are now ready to precisely state our signal reconstruction problem. The basic features are the following:

- We allow a finite step preview for reconstruction.
- The acquisition device, sampling and hold elements are fixed.

Consider the block diagram Fig. 2. The external continuous-time signal $w_c \in L^2$ is first band-limited by going through the analog low-pass filter $F(s)$, which is linear and time-invariant, and finite-dimensional. This $F(s)$ is a rational function of $s$ which is strictly proper (i.e., the degree of the numerator polynomial is less than that of the denominator). As is well known, it is represented by a linear, time-invariant system

$$\frac{dx}{dt}(t) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t)$$

where $A, B, C$ are constant matrices of appropriate sizes, and $F(s) = C(sI - A)^{-1}B$. Hence $F$ is not an ideal filter unlike the case of the Shannon paradigm, and is physically realizable through the above state space model. The signal $w_c$ is the external signal that drives $F$ and produces the actual signal $y_c$ to be processed. The produced signal $y_c$ is then sampled by the ideal sampler $S_h$ and becomes a discrete-time signal $y_d$ with sampling period $h$. This signal is then upsampled by $\uparrow L$ to allow for processing (interpolation) between the original sampling period $h$. The digital filter $K(z)$ processes this upsampled signal to produce $y_{K,d}$. The signal $y_{K,d}$ then goes through the zero-order hold $H_{h/L}$ and becomes a continuous-time signal. It is then further processed by an analog low-pass filter $P(s)$ to become the final analog output $y_p$.

In the upper part of the diagram, we allow $m$ steps of delay for the analog signal $y_c$ and obtain $y_c(t - mh)$. This is a setup for allowing a preview in a causal framework, and it is very effective compared to reconstruction without a preview. This also takes care of certain processing delays caused by the processing filter. The integer $m$ is a design parameter that can be chosen by the designer. This is in marked contrast to the conventional design methodologies: These methods usually allow a non-causal impulse response for reconstruction, e.g., [31], [20], [21]. But in real implementation, one has to truncate it, and it is often unclear how many steps one would need to obtain a desired accuracy. In the present setup, one can prespecify an allowable step of delays (preview), and obtain an optimal design under such a constraint.

Finally, the processed signal $y_p$ is compared with this delayed $y_c(t - mh)$ and subtracted from it to obtain the error signal $e_c$. The design objective is to make the error as small as possible. Observe also that this design framework is...
formulated in the continuous-time domain in contrast to the usual discrete-time setups.

We must specify a performance index to give a precise meaning to this problem. The following $L^2$ induced norm from $w_c$ to $e_c$ (or the sampled-data $H^\infty$ norm) is the one we take:

$$ J := \sup_{w_c \in L^2, w_c \neq 0} \frac{\|e_c\|_2}{\|w_c\|_2}. $$

(11)

We thus arrive at the following design problem:

**Problem 1:** Let $T_{ew}$ denote the input/output operator from $w_c$ to $e_c$ which is defined as $y_c := y_c(\cdot) - w_c(\cdot - mhs)$ in Fig. 2. Given an attenuation level $\gamma > 0$, find, if one exists, a stable digital (discrete-time) filter $K(z)$ such that

$$ \|T_{ew}\|_\infty := J = \sup_{w_c \in L^2(0,\infty)} \frac{\|T_{ew}w_c\|_2}{\|w_c\|_2} < \gamma. $$

(12)

The performance index (12) intends to minimize the maximum error induced by an (unknown) input $w_c$ that gives rise to the largest norm of $e_c$ among all inputs. This is made possible by the $H^\infty$ design methodology. Note that the actual error is not known, but due to the min-max nature of the problem, we can minimize the worst transmission error. Observe also that this setup allows for a capability of minimizing continuous-time phase errors due to the continuous-time nature of the performance index, as opposed to the conventional gain-phase design principles.

This min-max problem differs sharply from the orthogonal projection based methods. Also, due to sampling, $T_{ew}$ is not even time-invariant (in continuous-time).

It is now known however that this problem is reducible to a linear time-invariant problem via lifting; see Appendix I; the problem is now solvable via now-standard $H^\infty$ control theory, see, e.g., [6], [4] (see also [9] for standard treatments of $H^\infty$ control in the continuous-time setting).

The existence of $e^{-mhs}$ makes this an infinite-dimensional $H^\infty$ problem; see [18], [34], [38], [40], etc. The simplest one is to employ the so-called fast-sampling/fast-hold approximation, which we will outline in the next section.

**V. Solution via Fast Sample/Hold**

While Problem 1 is known to be reducible to a finite-dimensional problem [18], [22], it is not necessarily appealing computationally. It is often more convenient to resort to an approximation method. We employ the fast sample/hold approximation [17], [6], [34], [38]. This method approximates continuous-time inputs and outputs via a sampler and hold that operate in the period $h/N$ for some positive integer $N$. The method usually works fairly well for $N \sim 5L$, where $L$ is the upsampling ratio given in Section IV, and the convergence of such an approximation is shown in [34], [38]. We show here the design procedure of $K(z)$ by the fast sampling/hold approximation.

The error system in Fig. 2 is a multirate system due to the upsampler $\uparrow L$. We first reduce this system to a single-rate one. Introduce the discrete-time lifting, also known as the polyphase decomposition [41], $L_L$ and its inverse $L_L^{-1}$ as

$$ L_L := (\uparrow L) \begin{bmatrix} 1 & z & \cdots & z^{L-1} \end{bmatrix}^T, $$

$$ L_L^{-1} := \begin{bmatrix} 1 & z^{-1} & \cdots & z^{-L+1} \end{bmatrix} (\uparrow L). $$

(13)

Then $K(z)(\uparrow L)$ can be rewritten by a lifted system as

$$ K(z)(\uparrow L) = L_L^{-1} \tilde{K}(z), $$

$$ \tilde{K}(z) := L_L K(z) L_L^{-1} \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T. $$

The filter $\tilde{K}(z)$ is an LTI (linear and time-invariant), single-input/L-output system that satisfies

$$ K(z) = \begin{bmatrix} 1 & z^{-1} & \cdots & z^{-L+1} \end{bmatrix} \tilde{K}(z^L). $$

(14)

Define $\tilde{H}_h := \mathcal{H}_h/L L_L^{-1}$, and we obtain the following equality

$$ \mathcal{H}_h/L K(z)(\uparrow L) \mathcal{S}_h = \tilde{H}_h \tilde{K}(z) \mathcal{S}_h. $$

Hence the multirate system in Fig. 2 is reduced to the single-rate system shown in Fig. 3.

We then employ the fast sample/hold approximation for the error system $T_{ew}$ in Fig. 3. We connect fast sample and hold devices $\mathcal{S}_{h/N}$, $\mathcal{H}_{h/N}$ with the error system $T_{ew}$ as shown in Fig. 4. For brevity of notation, let us adopt the following shorthand notation for the transfer function $D + C(zI - A)^{-1}B$:

$$ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A P_c & B P_c \\ C P & D P \end{bmatrix}, \quad P(s) = \begin{bmatrix} A P_c & B P_c \\ C P & D P \end{bmatrix}. $$

(15)

The sampled-data error system $T_{ew}$ can be approximated by a discrete-time LTI system as in the following theorem:

**Theorem 1:** Let state-space realizations of $F(s)$ and $P(s)$ be given by

$$ F(s) = \begin{bmatrix} A_F & B_F \\ C_F & 0 \end{bmatrix}, \quad P(s) = \begin{bmatrix} A_P & B_P \\ C_P & D_P \end{bmatrix}. $$

Let $N = Ll$ where $l$ is a positive integer, and define the discrete-time LTI system $T_N$ as follows:

$$ T_N(z) = z^{-m} F_N(z) - P_N(z) H \tilde{K}(z) S F_N(z), $$

(16)
Fig. 5. Discrete-time system for $H^\infty$ optimization

$$
\begin{bmatrix}
    A_F^N & A_F^{N-1} & A_F^{N-2} & \cdots & B_F \\
    C_F & 0 & 0 & \cdots & 0 \\
    C_F A_F & C_F B_F & 0 & \cdots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    C_F A_F^{N-2} & C_F A_F^{N-3} & \cdots & C_F B_F & 0 \\
    A_F & A_F^{N-1} & A_F^{N-2} & \cdots & B_F \\
    C_F & D_F & 0 & \cdots & 0 \\
    C_F A_F & C_F B_F & D_F & \ddots & \vdots \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    C_F A_F^{N-2} & C_F A_F^{N-3} & \cdots & C_F B_F & D_F \\
\end{bmatrix}
$$

$$
F_N = 
\begin{bmatrix}
    A_F^N \\
    C_F A_F \\
    \vdots \\
    C_F A_F^{N-2} \\
    A_F \\
    C_F \\
    \vdots \\
    C_F A_F^{N-2} \\
\end{bmatrix}, 
\begin{bmatrix}
    B_F \\
    0 \\
    \vdots \\
    0 \\
    B_F \\
    D_F \\
    \vdots \\
    D_F \\
\end{bmatrix}
$$

Then, for each fixed $\tilde{K}$ and for each $\omega \in [0,2\pi/h)$, the frequency response

$$
\|T_N(e^{j\omega h})\|_\infty \to \|T_{ew}(e^{j\omega h})\|_\infty.
$$

(17)

as $N \to \infty$, and this convergence is uniform with respect to $\omega \in [0,2\pi/h)$. Furthermore, this convergence is also uniform in $\tilde{K}$ if $\tilde{K}$ ranges over a compact set of filters.

**Proof.** See Appendix II.

In view of the uniformity of convergence $\|T_N\|_\infty$ in $\tilde{K}$, our design problem (12) can be approximated by

$$
\|T_N\|_\infty < \gamma.
$$

This is a discrete-time $H^\infty$ optimization problem. To obtain a filter $\tilde{K}(z)$ satisfying the above inequality, we can adopt numerical softwares as MATLAB with robust control toolbox [19], by the generalized plant representation depicted in Fig. 5, where $\tilde{w}_d = L_N w_d$ and $\tilde{e}_d = L_N e_d$. Once the optimal filter $\tilde{K}(z)$ is obtained, one can obtain the interpolation filter $\tilde{K}(z)$ by formula (14).

**Example 1:** Let us make a comparison with a usual linear phase filter—the Johnston filter [14]. We design the proposed filter $K(z)$ with interpolation ratio $L = 2$, sampling period $h = 1$, and delay step $m = 2$. The analog filters $F(s)$ and $P(s)$ are given by

$$
F(s) = \frac{1}{(Ts + 1)(0.1Ts + 1)}, \quad T = 7.0187, \quad P(s) = 1.
$$

The time constant $T = 7.0187$ is taken to be equivalent to 1 kHz with sampling frequency 44.1 kHz. It therefore corresponds to an energy distribution that decays by $-20$ dB per decade from 1 kHz and $-40$ dB per decade from 10 kHz.

Fig. 6 shows the Bode gain plots of the proposed filter and the Johnston FIR filter with 32 taps. We can see that the Johnston filter has a sharp decay around the cutoff frequency $\omega = \pi/2$, while the filter obtained by the proposed method shows a rather mild decay.

Fig. 7 shows the response of the Johnston filter against a rectangular wave. It exhibits a very sharp ringing effect. This is because the filter has a sharp cut-off characteristic, and inevitably introduces the well-known Gibbs phenomenon due to the fact that the frequency components beyond the pass-band are sharply truncated. In contrast, Fig. 8 shows the response of the filter designed by the present method. It shows virtually no ringing.

**VI. COMPARISON WITH OTHER METHODS**

A. **Remark on the consistency requirement**

As noted in the Introduction, the notion of consistent reconstruction is quite widely accepted in the literature, e.g., [27], [29]. We start with a discussion of this property.
Consider the spaces \( V_f \) and \( V_\phi \) in (9) and (10). When \( V_f \) and \( V_\phi \) are equal, we are in the situation of orthogonal projection. The Shannon paradigm is a particular case.

When \( V_f \not= V_\phi \), there is freedom in choosing an oblique projection, but consistency requirement [27] makes it unique: that is, one takes the oblique projection of \( L^2 \) onto \( V_\phi \) perpendicular to \( V_f \). This gives a perfect reconstruction for elements in \( V_\phi \) and also consistency. That is, when one injects any reconstructed signal \( \sum_k c[k] \phi(t - kh) \) into the acquisition device \( F \) and sampling, one should obtain the same sampled data \( c[k] \) [27], [29]. Unlike the least square error case, however, this process may yield a large error when the two spaces \( V_f \) and \( V_\phi \) are apart. This can be seen from an error analysis given in [27], [29] that depends on the angle of two spaces \( V_f \) and \( V_\phi \).

To see this more clearly, take a pure sinusoid \( \sin t \) (over some bounded interval if we strictly require it to belong to \( L^2 \)), and suppose that we sample it with sampling period \( h = \pi/2 \), and define the acquisition device to be the ideal filter and the reconstruction device to be the one given by

\[
\phi(t) = \begin{cases} 
1, & 0 \leq t \leq h \\
0, & \text{otherwise}.
\end{cases}
\]

If we sample \( \sin t \) at times \( 2n\pi/h \), \( n = 0, 1, 2, \ldots \), and if we adopt the consistency requirement (i.e., sampled \( y \) gives the same values that we started with), the reconstructed signal \( \tilde{y} \) becomes

\[
y(t) = \begin{cases} 
0, & 2kh \leq t < (2k + 1)h, \\
1, & (2k + 1)h \leq t < 2(2k + 1)h, \\
-1, & (2k + 3)h \leq t < 2(2k + 2)h.
\end{cases}
\]

On the other hand, it is easily seen that the function

\[
\tilde{y}(t) = \begin{cases} 
1/2, & 2kh \leq t2(k + 1)h, \\
-1/2, & 2(k + 1)h \leq t \leq 2(2k + 2)h,
\end{cases}
\]

gives a better approximation for \( \sin t \) with respect to the \( L^2 \) norm, and hence the consistency requirement does not necessarily lead to a good approximation result. See Figs. 10 and 11.

Hence the consistent reconstruction does not necessarily yield a desirable result when analog performance is taken into account.

### B. Comparison with cardinal exponential splines

Unser and Blu [30], [31] proposed to use cardinal exponential splines to recover analog information from sampled data. They used the philosophy of placing emphasis on analog performance exactly the same as ours here. Their method is however very different from the present one, and in some cases it does not necessarily lead to a desirable result. Even a stability issue may arise. We here give a detailed analysis of their method, and make some comparisons.

Consider the block diagram Fig. 12. In this figure, \( F(s) \) and \( P(s) \) are analog filters, and \( S \) and \( H \) are, respectively, an ideal sampler and the zero-order hold with sampling period \( h = 1 \). The filter \( F(s) \) represents the acquisition filter for sampling, and \( P(s) \) is a postfilter which smooths out the output of the
zero-order hold \( \mathcal{H} \). We assume that these analog filters are causal and LTI systems defined by

\[
F(s) = \prod_{m=1}^{m_1} \left( s - \gamma_{1m} \right) \prod_{m=1}^{m_2} \left( s - \gamma_{2m} \right), \quad P(s) = \prod_{n=1}^{n_1} \left( s - \alpha_{1n} \right) \prod_{n=1}^{n_2} \left( s - \alpha_{2n} \right),
\]

with impulse responses \( f(t) \) and \( p(t) \), respectively. Under these assumptions, Unser [31] derived the optimal reconstruction filter which achieves the consistency constraint

\[
\langle y_c - y_p, \phi(-\hat{k}) \rangle = 0, \quad \hat{k} = 0, \pm 1, \pm 2, \ldots,
\]

where \( \phi(t) = f(-t) \). This idea states that there remains no extra component in the error \( y_c - y_p \) that can be expanded with elements in \( V_n \) (see (10)). The optimal filter is obtained by the oblique projection technique [27], [29], [30], [31] as follows:

\[
K_{\text{op}}(z) = \frac{\Delta_1(z) \Delta_2(z)}{\sum_{k=0}^{n_1} \beta(k) z^{-k}},
\]

(20)

where

\[
\Delta_i(z) := \prod_{n=1}^{n_i} \left( 1 - e^{\alpha_i n} z^{-1} \right), \quad i = 1, 2,
\]

and \( \beta(0), \beta(1), \ldots, \beta(n_1 + n_2 + 1) \) are the sampled values of \( \beta(t) \) which is defined by its Laplace transform

\[
\hat{\beta}(s) = \frac{1 - e^{-s}}{s} F(s) P(s) 2 \prod_{i=1}^{n_1} \prod_{n=1}^{n_i} \left( 1 - e^{\alpha_i n} e^{-s} \right).
\]

(22)

The filter (20) is generally an IIR filter. We can prove that this filter is given via system inversion as follows:

**Theorem 2:** The optimal filter \( K_{\text{op}}(z) \) in (20) can be equivalently realized as

\[
K_{\text{op}}(z) = 1 \frac{H_d(z)}{H_d(z)},
\]

(23)

where \( H_d(z) \) is the step-invariant equivalent discretization of \( F(s) P(s) \), that is, if a state-space realization of \( F(s) P(s) \) is given by \( \{ A, B, C, 0 \} \), then

\[
H_d(z) = S F(s) P(s) \mathcal{H} = \left[ e^A \right] C \left[ \int_0^t e^{A\tau} B d\tau \right].
\]

**Proof** See Appendix III.

While our sampled-data method always yields a stable filter, the above optimal filter \( K_{\text{op}} \) derived by Unser [31], according to this theorem, can have a pole in \( \mathbb{D}_+ := \{ z \in \mathbb{C} : |z| \geq 1 \} \), when the relative degree of \( F(s) P(s) \) is strictly greater than 2. This follows from the following well-known result on limiting zeros [3]:

**Fact 2:** Let \( \Sigma \) be a continuous-time, linear time-invariant single-input, single-output system with relative degree strictly greater than 2, and let \( \Sigma_h \) be its step-invariant discretized system with sampling period \( h \). Then there always exists an \( h \) such that \( \Sigma_h \) possesses a zero in \( \mathbb{D}_+ \).

Even if the sampling time \( h = 1 \) in our present normalization is not small relative to the time-constants of \( F(s) \) and \( P(s) \), the discretized system \( H_d(z) \) can still have an unstable zero, thereby yielding an unstable pole of \( K_{\text{op}}(z) \). In such a case, Unser and Blu [30] propose to use a non-causal filter, folding back the anti-causal part associated with the pole in \( \mathbb{D}_+ \) to the negative time axis. This can, in principle, lead to a very long delay for reconstruction.

Let us see this by an example. Consider

\[
F(s) = \frac{1}{s + 1}, \quad P(s) = \frac{1}{(s + d)(s + 2)}, \quad d = 1.5.
\]

The zeros of the discretized system \( H_d(z) \) are

\[
\{ -1.28549, -0.0816767 \},
\]

and hence the optimal filter

\[
K_{\text{op}}(z) = \frac{1}{H_d(z)} = \frac{z^3 - 0.7263z^2 + 0.1621z - 0.01111}{0.05725z^2 + 0.27372z + 0.006011}
\]

has an unstable pole at \( z = -1.28549 \). This \( K_{\text{op}} \) agrees exactly with the one obtained via oblique projections; see, e.g., [5]. To implement this filter, we first shift \( K_{\text{op}}(z) \) to obtain a proper transfer function, that is, we use \( K_{\text{op}}(z) = \frac{1}{H_d(z)} \), as suggested in [31, Subsection V-C]. Then, we split it into two parts: the causal and anti-causal part. Fig. 13 and Fig. 14 show the impulse responses of the causal and anti-causal part.

The anti-causal impulse response exhibits much oscillation because the pole \(-1.28549\) is close to the point \( z = -1 \). Using this filter, we reconstruct a response against a rectangular wave. Fig. 15 shows the result. The reconstructed signal exhibits much oscillation around the edges. This is due to the oscillatory impulse response shown in Fig. 14. Moreover, the result shows a rather long delay, about 31 steps.
Let us design the (sub)optimal filter $K_{sd}(z)$ by our sampled-data $H^\infty$ optimization method. Assume the analog characteristic of the input signals to be

$$P(s) = \frac{1}{s + 0.05}.$$  

We also assume the reconstruction delay $m = 5$. Fig. 16 shows the reconstructed signal against the same rectangular wave. It is clearly seen that the proposed method provides a much better result.

Let us further discuss the stability issue. It is claimed in [31] that there will be no zeros on the unit circle in the optimal filter, but this does not hold. Fig. 17 shows the locus of a pole of $K_{op}(z)$ for $d \in [2, 3.5]$. There exists a real number $d$ (approximately 2.72778) such that the filter $K_{op}(z)$ has a pole at $z = -1$. This clearly shows that the filter cannot be implemented as it is, and the optimal filter is not stable.

Fig. 18 shows the frequency response of the filters $K_{op}(z)$ and $K_{sd}(z)$.

The optimal filter $K_{op}(z)$ shows a higher gain than the sampled-data filter $K_{sd}(z)$ in high frequency, and this explains the ringing around the edges in Fig. 15.
C. Comparison with the results of Meinsma and Mirkin

Recently, Meinsma and Mirkin [20], [21] have also studied the signal reconstruction problem in a framework that is similar to ours. Some key features of their results are the following:

- Solutions are given to signal reconstruction with free sampler or free hold, or both free sampler and hold.
- Instead of a fixed preview length, they allow non-causal filters.
- An $L^\infty$ bound for the error performance as well as $L^2$ (or $H^2$) type closed-form solutions are given.

Our work which started in 1995 has the following features:

- fixed sampler and hold devices,
- fixed preview length, and
- causal and stable $H^\infty$-optimal (suboptimal) filter construction.

Our motivation is that the fixed sampler and hold is a more commonly encountered and practical situation. We have also chosen a fixed preview length ($e^{-mhs}$ in Fig. 2) as a design parameter, rather than considering optimal filter design with non-causal filters. In the latter, when we have to implement it in practice, we need to truncate the impulse response at a certain length. A priori estimate of the resulting performance after truncation is difficult to obtain. When one imposes certain filter characteristics (e.g., fast decay beyond the pass-band), the amount of delays necessary to implement such filters can be very large, often reaching thousands of steps, as is the case of FIR filters having a sharp cut-off characteristic. On the other hand, once we fix a delay length, which corresponds to the preview length, our design Problem 1 in Section IV always gives rise to an optimal performance bound $\gamma$ (cf. (12)). Moreover, due to the very nature of sampled-data $H^\infty$ control, this will always yield a stable filter.

VII. APPLICATIONS TO SOUND AND IMAGE PROCESSING

We here present some applications of the proposed method to sound and image processing.

A. Application to sounds

As seen in Example 1 in Section V, the sharp cut-off characteristic based on the Shannon paradigm generally induces a high distortion due to the Gibbs phenomenon.

We here show a brief example of sound recovery in high frequency. We consider a sound sample whose Fourier transform is shown in Fig. 19. This signal has the frequency components up to 22.05 kHz. We apply downsampler $\downarrow 4$ to this signal to obtain a signal whose bandwidth is limited to 5.51 kHz, and attempt to recover the original high frequency components.

We upsample it by the factor of 4, and then apply two filters: a conventional equi-ripple filter and the proposed filter. Fig. 20 shows the Fourier transform of the reconstructed signal by the equi-ripple filter, while Fig. 21 shows the Fourier transform of the recovered sounds by the proposed method, with a suitable $P(s)$ as in Example 1. We can see that high frequency components beyond 6 kHz are well recovered by the sampled-data method. On the other hand, if we apply the equi-ripple filter with cut-off frequency 5.51 kHz, it does not give any frequency components beyond 6 kHz, naturally, since there is no guiding principle for reconstructing such components beyond 5.51 kHz. The advantage of the present method is that it can evaluate the overall performance of the error signal $e_c$ in Fig. 2 in terms of the $H^\infty$-norm of $T_{M,m}$ of the transfer operator.

The present method has been applied to sound processing, particularly in supplementing lost high-frequency components in compression audio, and has been quite successful. In sound compression, the bandwidth is often limited to a rather narrow range (e.g., only up to 12 or 16 kHz, as in the MP3 or AAC format). The digital filter using sampled-data theory here, along with upsampling, can recover the lost intersample information optimally in the $H^\infty$ sense, thereby expanding the effective bandwidth to the original range. This has been patented [36], [37] and already marketed as sound processing LSI chips by Sanyo Semiconductors, and successfully used in mobile phones, digital voice recorders and MP3 players; their cumulative production has exceeded 25 million units as of 2011.

B. Application to Images

The same idea is applicable to images. However, since images are two-dimensional, we should be careful about how our (essentially) one-dimensional method can be applied. There is no universal recipe for this, and the simplest is to apply this in two steps: first process the data in the horizontal direction, save the temporary data in buffer memories, and then process them in the vertical direction.
We can interpolate lost intersample data by the present framework. For example, take the well-known sample picture of Lena, Fig. 22. The next picture Fig. 23 shows its twice downsampled image. To make the image more visible, Fig. 23 is enlarged by the zero-order hold.

To recover the original resolution, we first upsample the downsampled image by the factor of $2$. We then apply two interpolation filters: one is the conventional equi-ripple FIR filter with 128 taps with cut-off frequency at the downsampled frequency, and the other is the proposed sampled-data $H^\infty$ filter designed with $h = 1$, $m = 1$, $P(s) = 1$, with

$$F(s) = \frac{1}{20s + 1}.$$

Fig. 24 shows the reconstructed image by the equi-ripple filter. We see many artifacts on the face. This is due to the Gibbs phenomenon shown in Fig. 7 in the case of 1-dimensional processing, since the equi-ripple filter has a very sharp cut-off characteristic as the Johnston filter does.

Fig. 25 shows the result of the proposed method, which looks more natural. The edges in the original image are well recovered and the face does not necessarily have such artifacts as Fig. 24.

Observe again that the ideal lowpass filter which is optimal in the Shannon paradigm does not give an appropriate result because of the sharp cut-off characteristic and the associated Gibbs phenomenon.

VIII. CONCLUDING REMARKS

We have presented a new framework for digital signal processing. The fundamental philosophy is the emphasis on analog (continuous-time) performance with discrete-time signal processing. This naturally leads to a technical difficulty because two different time-sets are involved: continuous and discrete. Leveraging the sampled-data $H^\infty$ control theory, we have presented computable procedures for designing optimal, stable, causal filters. These filters are optimal with respect to a uniform analog performance measure. Our methodology is applicable to a wide variety of theoretical and application problems in digital signal processing. We believe that it has many merits and we hope it will be more widely used in the future.
Fig. 26. Lifting: a continuous-time signal $w(t)$ (left) is converted to a
function-valued discrete time signal (right).

ACKNOWLEDGMENTS

This research is supported in part by the JSPS Grant-in-Aid for Scientific Research (B) No. 2136020318360203, and also by Grant-in-Aid for Exploratory Research No. 22656095.

APPENDIX I

LIFTING, TRANSFER FUNCTIONS, AND FREQUENCY RESPONSES

As mentioned in the main text, the major difficulty in sampled-data systems lies in the mixture of two different time sets: continuous and discrete. Lifting [4], [6], [32], [33] is a method that makes it possible to describe continuous-time systems in a discrete-time setting, without introducing any approximation, thereby merging the two time sets into one.

We start by placing a continuous-time signal in a discrete-time framework. Take a continuous-time signal $w(t)$, and consider the following mapping $L$ (with a suitable domain and codomain) that maps $w$ into a sequence of functions as

$$(Lw)[k] := w[k] := \{w(kh + \theta)\}_{\theta \in [0,h]}, \ k = 0, 1, 2, \ldots$$

See Fig. 26. The operator $L$ is called lifting.

This idea makes it possible to view time-invariant, or even periodically time-varying continuous-time systems as linear, time-invariant discrete-time systems.

Using this operator, one can describe a linear, time-invariant continuous-time system with a linear, time-invariant discrete-time system. Consider the following linear system:

$$\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t), \\
y(t) &= Cx(t),
\end{align*}$$

(25)

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$ are the state, input and output of this system, respectively. Let us assume, e.g., $u \in L^2_{loc}[0,\infty)$, the set of locally square-integrable functions on $[0,\infty)$. The idea is that we view the continuous-time system (25) as one with discrete-timing $t = kh$, $k = 0, 1, 2, \ldots$ such that it receives function-valued inputs at these instants and produces function-valued outputs at these times also. Suppose that (25) is at state $x(kh)$ at time $t = kh$. Then

$$\begin{align*}
x((k+1)h) &= e^{Ah}x(kh) + \int_0^h e^{A(h-\tau)}Bu(kh + \tau)d\tau, \\
y(kh + \theta) &= Ce^{A\theta}x(kh) + \int_0^\theta e^{A(\theta-\tau)}Bu(kh + \tau)d\tau.
\end{align*}$$

where $0 \leq \theta < h$ denotes the intersample parameter. Lifting the input $u(t)$ and the output $y(t)$ as per (24), we can rewrite these formulas as a lifted discrete-time system [6], [33]:

$$\begin{align*}
x[k+1] &= Ax[k] + Bu[k], \\
y[k] &= Cx[k] + Du[k], \ k = 0, 1, 2, \ldots
\end{align*}$$

where $x[k] = x(kh)$, $u[k] = (Cu)[k]$, $y[k] = (Ly)[k]$, and $A : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto e^{Ah}x$.

$B : L^2(0, h) \to \mathbb{R}^n : u \mapsto \int_0^h e^{A(h-\tau)}Bu(\tau)d\tau$

$C : \mathbb{R}^n \to L^2[0, h] : x \mapsto Ce^{A\theta}x$,

$D : L^2(0, h) \to L^2[0, h] : u \mapsto \int_0^\theta e^{A(\theta-\tau)}Bu(\tau)d\tau$,

where $\theta \in [0, h)$ describes the intersample parameter. Observe that the operators $A, B, C, D$ above do not depend on time $k$, and hence system (26) is a time-invariant discrete-time system, albeit with infinite-dimensional input and output spaces. Hence it is straightforward to connect this system with a discrete-time controller (or a filter), and the obtained sampled-data system is again a linear, time-invariant discrete-time system without sacrificing any intersampling information. The resulting system can also be described by a 4-tuple of operators $A, B, C, D$, and its transfer function (operator) of the lifted system is defined as

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

with such $A, B, C, D$. Note that for each fixed $z \in \mathbb{C} \setminus \sigma(e^{Ah})$, $(\sigma(e^{Ah})$ is the set of eigenvalues of $e^{Ah})$, $G(z)$ is a linear operator acting on $L^2(0, h)$ into itself. The frequency response operator is then defined as $G(e^{j\omega h})$, and the gain at frequency $\omega$ is defined as

$$\|G(e^{j\omega h})\| = \sup_{v \in L^2(0, h)} \frac{\|G(e^{j\omega h}) v\|}{\|v\|}.$$

The $H^\infty$ norm of $G$ then becomes

$$\|G\|_{\infty} = \sup_{\omega \in [0, 2\pi/h]} \|G(e^{j\omega h})\|,$$

which is known to be identical to the $L^2$-induced norm given by (6) in Section II [6].

APPENDIX II

PROOF OF THEOREM 1

Let $T_N$ be the fast discretization shown in Fig. 4, namely,

$$T_N = S_h/N T \circ H_{h/N} N - S_h/N P \tilde{H}_h \tilde{K} S_h F H_{h/N}.$$

By using the identities

$$\tilde{H}_h \circ H_{h/N} L_N^{-1} H, \quad H_{h/N} S_h/N H_{h/N} = H_{h/N}, \quad S_h = S L_N S_h/N,$$

where

$$H := \text{diag} \{I_l\} \in \mathbb{R}^{N \times L}, \quad I_l := [1, 1, \ldots, 1]^T \in \mathbb{R}^l,$$

$$S := [1, 0, \ldots, 0] \in \mathbb{R}^{1 \times N},$$

we have

$$T_N = z^{-mN} F_N - P_N L_N^{-1} H \tilde{K} S L_N F_N,$$
where \( F_N := S_{h/N} F H_{h/N} \) and \( P_N := S_{h/N} P H_{h/N} \). Applying the discrete-time lifting \( L_N \) and its inverse \( L_N^{-1} \) gives

\[
T_N = L_N T_N L_N^{-1} = z^{-m} L_N F_N L_N^{-1} - L_N P_N L_N^{-1} H \tilde{K} S L_N F_N L_N^{-1} = z^{-m} F_N - P_N H \tilde{K} S F_N.
\]

The state space matrices for \( P_N \) and \( F_N \) are given by the formulas in Theorem 8.2.1 [6, Chap. 8]. The convergence in (17) is proved in [34] and in [38]. It is uniform in frequency when the filter is confined to a compact set [38].

**APPENDIX III**

**PROOF OF THEOREM 2**

First, we consider the denominator of \( K_{op}(z) \). The coefficients \( \beta(k), k = 0, 1, 2, \ldots \) are obtained by sampling the inverse Laplace transform of \( \hat{\beta} \) given in (22). Since \( (1-e^{-s})/s \) is the Laplace transform of the zero-order hold \( H \), the denominator of \( K_{op}(z) \) is the \( Z \)-transform of the step-invariant transformation of

\[
\hat{\beta}(s) := \prod_{i=1}^{2} \prod_{n=1}^{n_i} (1-e^{a_n e^{-s}}) F(s) P(s).
\]

That is, the denominator is given by

\[
\mathcal{Z}[S\hat{\beta}(s)H] = \mathcal{Z} \left[ \prod_{i=1}^{2} \prod_{n=1}^{n_i} (1-e^{a_n e^{-s}}) F(s) P(s) H(s) \right].
\]

By using the relation \( Se^{-s} = z^{-1} S \) and the definition of \( \Delta_i(z) \) in (21), we have

\[
S \prod_{i=1}^{2} \prod_{n=1}^{n_i} (1-e^{a_n e^{-s}}) = \Delta_1(z) \Delta_2(z) S.
\]

It follows that

\[
S\hat{\beta}(s)H = \Delta_1(z) \Delta_2(z) SF(s) P(s) H = \Delta_1(z) \Delta_2(z) H_d(z).
\]

Then, since the numerator of \( K_{op}(z) \) is \( \Delta_1(z) \Delta_2(z) \), we conclude that

\[
K_{op}(z) = \frac{\Delta_1(z) \Delta_2(z)}{S\hat{\beta}(s)H} = \frac{1}{H_d(z)}.
\]

**REFERENCES**


