Sampling from a System-Theoretic Viewpoint: Part II—Noncausal Solutions

Gjerrit Meinsma and Leonid Mirkin, *Member, IEEE*

Abstract

This paper puts to use concepts and tools introduced in Part I to address a wide spectrum of noncausal sampling and reconstruction problems. Particularly, we follow the system-theoretic paradigm by using systems as signal generators to account for available information and system norms (L^2 and L^∞) as performance measures. The proposed optimization-based approach recovers many known solutions, derived hitherto by different methods, as special cases under different assumptions about acquisition or reconstructing devices (e.g., polynomial and exponential cardinal splines for fixed samplers and the Sampling Theorem and its modifications in the case when both sampler and interpolator are design parameters). We also derive new results, such as versions of the Sampling Theorem for downsampling and reconstruction from noisy measurements, the continuous-time invariance of a wide class of optimal sampling-and-reconstruction circuits, etcetera.

Index Terms

Sampling and reconstruction, non-causal filters, least-square optimzation, min-max optimization, Shannon formula, lifting.

I. Introduction and Problem Formulation

In Part I [1] we presented and expanded on the system-theoretic approach to the sampling / reconstruction (SR) problem and related technical tools. The primary goal of this part is to demonstrate how these

This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 1238/08).

G. Meinsma is with the Dept. of Applied Math., University of Twente, 7500 AE Enschede, The Netherlands. E-mail: *g.meinsma@utwente.nl*.

L. Mirkin is with the Faculty of Mechanical Eng., Technion—IIT, Haifa 32000, Israel. E-mail: *mirkin@technion.ac.il*.

ideas can be put to use in various SR problems when no causality constraints are imposed on acquisition / reconstructing devices.

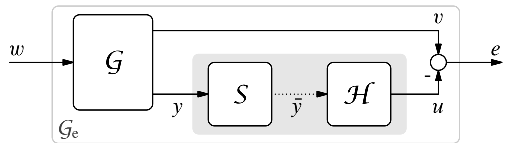


Fig. 1. Sampling / reconstruction (SR) setup

We consider the SR setup depicted in Fig. 1, where v is the analog signal that we want to reconstruct from *sampled* measurements of another, possibly different, signal y. Sampling and reconstruction is carried out by the hybrid signal processor (HSP), highlighted by the gray box in Fig. 1, which comprises a sampler (acquisition or A/D device) $\mathcal{S}$ and a hold (interpolator or D/A device) $\mathcal{H}$ and produces a signal u, which is aimed to be close to v in some sense. In accordance with the system-theoretic approach (see [1] for more details), we express the available information about v, y and their relations via modeling these signals as outputs of a *signal generator*

$$\mathcal{G} = \begin{bmatrix} \mathcal{G}_v \\ \mathcal{G}_y \end{bmatrix}$$

driven by a common (normalized) signal w, so that properties $\mathcal{G}$ reflect those of v and y. For example, the assumption that v and y are realizations of stochastic processes with known power spectral densities $\Phi_v(\omega)$ and $\Phi_y(\omega)$ and a cross-spectral density $\Phi_{yv}(\omega) = [\Phi_{vy}(\omega)]^*$ is equivalent to assuming that w is white noise and $\mathcal{G}$ is a stable system, whose frequency response $G(\mathrm{j}\omega)$ verifies

$$G(\mathrm{j}\omega)[G(\mathrm{j}\omega)]^* = \begin{bmatrix} \Phi_v(\omega) & \Phi_{vy}(\omega) \\ \Phi_{yv}(\omega) & \Phi_y(\omega) \end{bmatrix} \geq 0.$$

As the measure of reconstruction performance we use norms of the *error system*

$$\mathcal{G}_\mathrm{e} := \mathcal{G}_v - \mathcal{H}\mathcal{S}\mathcal{G}_y,$$

which connects w and the reconstruction error e. Specifically, in this paper we are concerned with minimizing the L^2 and L^∞ norms of $\mathcal{G}_\mathrm{e}$. Loosely speaking, the former corresponds to minimizing the energy / variance of e under known v and y (mean-square), while the latter—under the worst-case (min-max) scenario. In each of this cases, both deterministic and stochastic interpretations of the optimization problems exist. Given the norm, the design problems split into three types:

	fixed sampler	free sampler
fixed hold	Type-I	Type-II
free hold	Type-III	Type-IV

where "fixed" might mean both the ideal and generalized sampling / reconstruction. These three types we consider in various settings.

Remark 1.1: There are also Type-I problems, which are problems when both sampler and hold are fixed and only a discrete filter (in between sampler and hold, not shown in the diagram) needs to be designed. It appears that most of the work in the sampling literature are concerned with this scenario. In the system-theoretic formulation, however, Type-I problems, unlike the other three cases, can be reduced to equivalent discrete estimation problems, which, in turn, are solvable by standard methods. Reduction procedures, applicable to non-causal and relaxed-causal setups, are available in [2] for the L^2 norm and in [3] for the L^∞ norm. We therefore do not deal with Type-I problems in this paper. ∇

To the best of our knowledge, noncausal SR problems have not been studied from the system-theoretic viewpoint so far. Nonetheless, we shall demonstrate below that in many cases the approach leads to solutions, already derived in the literature (and used in practice) by rather diverse methods. These are cardinal polynomial and exponential splines [4]–[6], which are produced by solving Type-III problems under certain choices of the signal generator $\mathcal{G}$, and the ubiquitous Sampling Theorem, with its sinc-interpolator, and several of its generalizations [7]–[9], which turn out to be the optimal solutions of various Type-IV problems in both the L^2 and the L^∞ cases. Moreover, we show that the well-known frequency folding phenomenon [10, §6.1] shows up in the determination of singular values of signal generators in the lifted domain (this is a key step in our treatment of Type-IV problems owing to the exhaustive characterisation of hybrid signal processor in the lifted domain via a rank condition in Theorem 4.1). All this deepens the insight into both existing and proposed methods.

At the same time, our machinery goes beyond known results, leading to new solutions and interpretations. We extend the Sampling Theorem to downsampling and SR in the face of noisy measurements, which are new results to the best of our knowledge. We prove an intrinsic continuous-time invariance of both L^2 (mean-square) and L^∞ (worst-case) optimal HSPs. We also present limitations on error-free reconstruction, which expoits an interplay between L^2 and L^∞ systems norms.

The paper is organized as follows. We begin with Type II (Section II) and Type III (Section III) problems. The rest of the paper addresses Type IV problems. Section IV is about a rank characterization of hybrid signal processors and in the following two sections we summarize fixed frequence singular value decompositions in lifted domain and the folding procedure. From Section VII onwards a series of applications is discussed, beginning with a single-channel SR and the ensuing limititations on error free reconstruction. Then, in Sections VIII and IX multi-channel SR and optimal downsampling are discussed. Finally, in Section X we consider SR from noisy measurements. Preliminary conference versions of some of the results presented here can be found in [11]–[13].

A. Notation

In this paper it is convenient to refer to systems that are linear and time invariant with respect to any continuous-time shift as *LCTI* systems, and to systems that are linear and time invariant under discrete times shifts, equal to a multiple of the sampling period h, as *LDTI* systems. For the rest the notion is the same as that of Part I, [1].

II. Type II: Fixed Hold, Optimal Sampler

Type-II (fixed hold) and Type-III (fixed sampler) problems are unconstrained projection problems which makes them easy to solve.

Proposition 2.1: Let $\mathcal{G}_v, \mathcal{G}_y, \mathcal{H} \in L^\infty$ and suppose that $\|\mathcal{G}_v\|_2 < \infty$ and $\mathcal{H}$ is a hold. Then every solution $\mathcal{S}_{\text{opt}} \in L^\infty$ (if any) of the normal equations

$$\mathcal{H}^* \mathcal{G}_v \mathcal{G}_y^* = (\mathcal{H}^* \mathcal{H}) \mathcal{S}_{\text{opt}} (\mathcal{G}_y \mathcal{G}_y^*) \tag{1}$$

is a sampler minimizing $\|\mathcal{G}_{\text{e}}\|_2$ over all $\mathcal{S} \in L^\infty$. The optimal performance level is then $\|\mathcal{G}_{\text{e}}\|_2^2 = \|\mathcal{G}_v\|_2^2 - \|\mathcal{H} \mathcal{S}_{\text{opt}} \mathcal{G}_y\|_2^2$. If in addition

$$(\mathcal{H}^* \mathcal{H})^{-1} \text{ and } (\mathcal{G}_y \mathcal{G}_y^*)^{-1} \text{ exist and are stable,} \tag{2}$$

then

$$\mathcal{S}_{\text{opt}} = (\mathcal{H}^* \mathcal{H})^{-1} \mathcal{H}^* \mathcal{G}_v \mathcal{G}_y^* (\mathcal{G}_y \mathcal{G}_y^*)^{-1} \tag{3}$$

is the unique stable optimal sampler.

Proof: Standard projection combined with the trace-like property [1, Eqn. (34)]. ∎

The optimal sampler (3) can be viewed as the cascade of the LCTI system $\mathcal{G}_v \mathcal{G}_y (\mathcal{G}_y \mathcal{G}_y^*)^{-1}$, the sampler $\mathcal{H}^*$ and the discrete system $(\mathcal{H}^* \mathcal{H})^{-1}$. The first system, $\mathcal{G}_v \mathcal{G}_y (\mathcal{G}_y \mathcal{G}_y^*)^{-1}$, is actually the optimal analog filter, i.e., the filter $\mathcal{F}$ minimizing $\|\mathcal{G}_v - \mathcal{F} \mathcal{G}_y\|_2$ over all stable $\mathcal{F}$.

The above proposition is formulated representation free. To make matters concrete one can employ a specific representation. The lifted frequency response representation is interesting because it shows that the optimal sampler

$$\acute{S}_{\rm opt}(e^{j\theta}) = [(\grave{H}^*\grave{H})^{-1}\grave{H}^*\breve{G}_v\breve{G}_y^*(\breve{G}_y\breve{G}_y^*)^{-1}](e^{j\theta})$$

for each frequency θ satisfies the normal equations associated with the norm $\|\breve{G}_{\rm e}(e^{j\theta})\|_{\rm HS}$. That is, the optimal sampler also frequency-wise minimizes the norm of the frequency response [1, Eqn. (31)]. This is a well known feature in noncausal filter design [14]. If all signals are scalar then the Fourier transform of the optimal sampling function is probably the simplest representation. Indeed in that case $\mathcal{G}_y^*$ cancels in (1) and the optimal sampling function $\phi_{\rm opt}$ then can be shown to have Fourier transform

$$\Psi_{\rm opt}(j\omega) = \frac{h}{\sum_{k\in\mathbb{Z}}|\Phi(j(\omega+2k\omega_{\rm N}))|^2}\Phi(-j\omega)\frac{G_v(j\omega)}{G_y(j\omega)}, \tag{4}$$

where $\Phi(j\omega)$ is the Fourier transform of the hold function $\phi(t)$. This follows for instance from [1, Prop. 5.2].

The L^∞ optimal sampler is more involved but it applies to a larger class of signal generators in that $\|\mathcal{G}_v\|_2$ need not be finite. The following result is proved in Appendix.

Proposition 2.2: Let $\mathcal{G}_v, \mathcal{G}_y, \mathcal{H} \in L^\infty$ and suppose $\mathcal{H}$ is a hold and that (2) is satisfied. Then

$$\|\mathcal{G}_{\rm e}\|_\infty \geq \max(\|(I-\mathcal{H}(\mathcal{H}^*\mathcal{H})^{-1}\mathcal{H}^*)\mathcal{G}_v\|_\infty, \quad \|\mathcal{G}_v(I-\mathcal{G}_y^*(\mathcal{G}_y\mathcal{G}_y^*)^{-1}\mathcal{G}_y)\|_\infty) \tag{5}$$

for any stable sampler, and there exist stable samplers that achieve equality. If $\mathcal{G}_y^{-1}$ exists and is stable then the L^2-optimal (3) is also L^∞-optimal. $\triangledown$

Each term in (5) has a clear interpretation. The first term $\|(I-\mathcal{H}(\mathcal{H}^*\mathcal{H})^{-1}\mathcal{H}^*)\mathcal{G}_v\|_\infty$ is the minimal L^∞-norm for the case that $y = v$, i.e., for the case that all information about the signal v that we want to reconstruct is available for sampling. The second term, $\|\mathcal{G}_v(I-\mathcal{G}_y(\mathcal{G}_y\mathcal{G}_y^*)^{-1}\mathcal{G}_y^*)\|_\infty$, is the L^2-induced norm of the mapping $e = G_{\rm e}w$ for w restricted to $w = (I-\mathcal{G}_y^*(\mathcal{G}_y\mathcal{G}_y^*)^{-1}\mathcal{G}_y)\hat{w}$. These are the signals w for which there is nothing to sample, $y = 0$. Evidently that is a lower bound for $\|\mathcal{G}_{\rm e}\|_\infty$.

A. When $\mathcal{G}_y = \mathcal{G}_v$

Now suppose that the signal available for sampling, y, is exactly the signal to be reconstructed, v. In other words, let $\mathcal{G}_y = \mathcal{G}_v$. For this case the design of L^2 optimal samplers for fixed holds is well documented [15, Section IV] and the optimal sampler is then essentially independent of the signal generator. Including the L^∞ norm we obtain:

Corollary 2.3: Let $\mathcal{G}_y = \mathcal{G}_v$, $\mathcal{H} \in L^\infty$, and suppose that $(\mathcal{H}^*\mathcal{H})^{-1}$ exists and is stable. Then

$$S_{\text{opt}} := (\mathcal{H}^*\mathcal{H})^{-1}\mathcal{H}^* \tag{6}$$

minimizes the L^∞ norm of $\mathcal{G}_\text{e}$ with

$$\|\mathcal{G}_\text{e}\|_\infty = \|(I - \mathcal{H}(\mathcal{H}^*\mathcal{H})^{-1}\mathcal{H}^*)\mathcal{G}_v\|_\infty. \tag{7}$$

If in addition $\|\mathcal{G}_v\|_2 < \infty$, then it minimizes the L^2 norm as well with $\|\mathcal{G}_\text{e}\|_2^2 = \|\mathcal{G}_v\|_2^2 - \|\mathcal{H}(\mathcal{H}^*\mathcal{H})^{-1}\mathcal{H}^*\mathcal{G}_v\|_2^2$. ▽

Indeed $S_{\text{opt}} = (\mathcal{H}^*\mathcal{H})^{-1}\mathcal{H}^*$ solves the normal equation (1) and does not depend on $\mathcal{G}_v$. Another way to think about it is that now there is a single sampler that minimizes the *signal* error norm $\|(I - \mathcal{H}S)\mathcal{G}_v w\|_2$ for every given exogenous input $w \in L^2$. It implies that this sampler is also L^∞-optimal. The Fourier transform (4) of the optimal sampler reduces to

$$\Psi_{\text{opt}}(\mathrm{j}\omega) = \frac{h\Phi(-\mathrm{j}\omega)}{\sum_{k\in\mathbb{Z}} |\Phi(\mathrm{j}(\omega + 2k\omega_\text{N}))|^2}.$$

Example 2.1: The adjoint $\mathcal{H}^*$ is a sampler and according to [1, §V-B], its sampling function is $\psi(t) = \phi(-t)$ with $\phi(t)$ the hold function of $\mathcal{H}$. Thus if the hold is causal then the adjoint hold (a sampler) is anti-causal, and vice-versa. The discrete filter $\bar{\mathcal{K}} := (\mathcal{H}^*\mathcal{H})^{-1}$ because of its symmetry is never causal, unless it is static. For the zero order hold, with hold function $\phi(t) = \mathbb{1}_{[0,h)}(t)$, the discrete filter $\mathcal{H}^*\mathcal{H}$ is the static gain, h. This follows from [1, Eqn. (28a)]. The optimal sampler (6) therefore is $\frac{1}{h}\mathcal{H}^*$. It is the sampler with sampling function $\psi(t) = \frac{1}{h}\phi(-t) = \frac{1}{h}\mathbb{1}_{(-h,0]}(t)$. It is an averaging noncausal sampler, see [1, II-B]. ▽

The optimal sampler (6) makes $\mathcal{H}S_{\text{opt}}$ the classic orthogonal projection (hence self adjoint) onto the image of $\mathcal{H}$, which agrees well with the ideas of [16]. Consequently, we have the trivial identity that $S\mathcal{H} = I$. This implies *consistency*, a term coined by [17]. In the present context consistency means that $S\mathcal{H}S = S$. In other words, in a consistent HSP any reconstructed signal $u := \mathcal{H}Sy$ when reinjected into the sampler recovers the discrete signal Sy that was injected into the hold.

The bulk of this paper handles cases in which both sampler and hold are designed simultaneously (Type-IV). Obviously, this generalizes Type-II and hence also in Type-IV problems the hybrid signal processor $\mathcal{H}S$ may be taken (self-adjoint) projections if $\mathcal{G}_y = \mathcal{G}_v$, and they are consistent. If causality requirements are imposed on sampler and / or hold then these properties are lost [18, Part III].

III. Type III: Fixed Sampler, Optimal Hold

Type-III problems are essentially dual to the Type-II problems that we considered in the previous section. This is why in this section we only summarize the results.

Proposition 3.1: Let $\mathcal{G}_v \in L^\infty \cap L^2$ and that a sampler S is given such that $S\mathcal{G}_y \in L^\infty$. Then every $\mathcal{H}_{\text{opt}} \in L^\infty$ (if any) that solves the normal equation

$$\mathcal{G}_v(S\mathcal{G}_y)^* = \mathcal{H}_{\text{opt}} S\mathcal{G}_y(S\mathcal{G}_y)^* \tag{8}$$

minimizes $\|\mathcal{G}_{\text{e}}\|_2$ over all $\mathcal{H} \in L^\infty$ attaining the performance level $\|\mathcal{G}_{\text{e}}\|_2^2 = \|\mathcal{G}_v\|_2^2 - \|\mathcal{H}_{\text{opt}} S\mathcal{G}_y\|_2^2$. In particular if $(S\mathcal{G}_y(S\mathcal{G}_y)^*)^{-1}$ exists and is stable then

$$\mathcal{H}_{\text{opt}} = \mathcal{G}_v(S\mathcal{G}_y)^*(S\mathcal{G}_y(S\mathcal{G}_y)^*)^{-1}. \tag{9}$$

is the unique stable optimal hold. ▽

The optimal hold (9) can be viewed as the cascade of a discrete system $(S\mathcal{G}_y(S\mathcal{G}_y)^*)^{-1}$, a hold $(S\mathcal{G}_y)^*$ and an analog system $\mathcal{G}_v$.

Without loss of generality we can take the sampler to be ideal because its sampling function may always be absorbed into $\mathcal{G}_y$. The required stability of $S_{\text{Idl}}\mathcal{G}_y$ in the above proposition is then ensured if $\mathcal{G}_y$ is LCTI having strictly proper rational transfer function $G(s)$ with no poles on the imaginary axis [1, §VI-A].

The abstract solution (9) for scalar signals and LCTI signal generators $\mathcal{G}_v$ and $\mathcal{G}_y$ is compactly described via the Fourier transform of its hold function

$$\Phi_{\text{opt}}(\mathrm{j}\omega) = \frac{hG_v(\mathrm{j}\omega)G_y(-\mathrm{j}\omega)}{\sum_{k\in\mathbb{Z}}|G_y(\mathrm{j}(\omega+2k\omega_{\text{N}}))|^2}, \tag{10}$$

still under the assumption that $S = S_{\text{Idl}}$. This follows from the jω-axis version of [1, Prop. 5.1].

A. When $\mathcal{G}_y = \mathcal{G}_v$

Let us return to the situation that $\mathcal{G}_y = \mathcal{G}_v$. Then once again the hybrid signal processor becomes consistent because $S\mathcal{H}_{\text{opt}} = I$ for the hold of (9). The normal equation (8) does not simplify much in this case. A crucial difference with Type-II is that now there is no single hold that minimizes the *signal* error norm $\|(\mathcal{G}_v - \mathcal{H}S\mathcal{G}_y)w\|_2$ for all w. Typically, in fact, for almost every given $w \in L^2$ there exists a hold $\mathcal{H}_w$ that makes the reconstruction error $(\mathcal{G}_v - \mathcal{H}_w S\mathcal{G}_y)w$ equal to zero[1], while no single $\mathcal{H}$ exists that does this for all w.

Let us further assume that the sampler is ideal, $S = S_{\text{Idl}}$ (for the L^2 criterion this can be viewed as a noise-free version of the optimal discretization of the Wiener filter solved in [19]). For this case we will establish connections with the cardinal exponential and polynomial spline hold functions of [4]–[6].

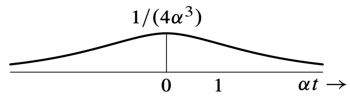
[1] $\grave{H}_w(\mathrm{e}^{\mathrm{j}\theta}) := \breve{v}(\mathrm{e}^{\mathrm{j}\theta})/\bar{y}(\mathrm{e}^{\mathrm{j}\theta})$ is often well defined.

Example 3.1 (Second order signal generator): Let $\mathcal{G}_v$ be the LCTI system with transfer function

$$G_v(s) = \frac{1}{(s+\alpha)^2}, \qquad \alpha > 0.$$

Clearly, $\mathcal{G}_v^* \mathcal{G}_v$ has impulse response $g_v * g_v^{\sim}$ with $g_v(t)$ the impulse response of $\mathcal{G}_v$ and $g_v^{\sim}(t) := g_v(-t)$. In our case

$$(g_v * g_v^{\sim})(t) = \tfrac{1}{4\alpha^3}(1-\alpha t)\mathrm{e}^{\alpha t}\mathbb{1}(-t) + \tfrac{1}{4\alpha^3}(1+\alpha t)\mathrm{e}^{-\alpha t}\mathbb{1}(t)$$

$=$ 

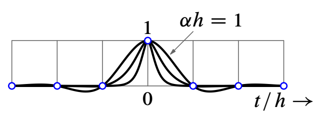
Hence the discrete system $\bar{\mathcal{F}} := S_{\mathrm{Idl}}\mathcal{G}_v(S_{\mathrm{Idl}}\mathcal{G}_v)^*$ has impulse response $f[n] := (g_v * g_v^{\sim})(nh) = \frac{1}{4\alpha^3}(1 + \alpha|n|h)\mathrm{e}^{-\alpha|n|h}$. For the optimal hold (9) we need the inverse of $\bar{\mathcal{F}}$. To this end, we first determine its discrete transfer function (with $r := \mathrm{e}^{-\alpha h}$)

$$\bar{F}(z) = \frac{1}{4\alpha^3}\left(\frac{1-r^2}{(1-rz)(1-r/z)} + \frac{\alpha h r/z}{(1-r/z)^2} + \frac{\alpha h r z}{(1-rz)^2}\right).$$

Its inverse, with $\beta := r(r^2(1+\alpha h) + (ah-1))$, then reads

$$\bar{K}(z) := \bar{F}^{-1}(z) = 4\alpha^3 \frac{(1-r/z)^2(1-rz)^2}{\beta z^{-1} + (1 - r^2(r^2+4\alpha h)) + \beta z}.$$

The hold function of the optimal hold (9) finally can be obtained by filtering $g_v * g_v^{\sim}$ with this $\bar{\mathcal{K}}$. For the three values $\alpha h \in \{1, 5, 10\}$ this results in

$\phi_{\mathrm{opt}}(t) =$ 

For $0 < \alpha h < 1$ the plot is very similar to that for $\alpha h = 1$. Since $g_v * g_v^{\sim}$ is twice continuously differentiable, also $\phi_{\mathrm{opt}}(t)$ has this degree of smoothness. Moreover, since $g_v * g_v^{\sim}$ is piecewise exponential, the optimal hold is a spline that on each sampling interval is a sum of exponential functions. This is an example of the exponential splines of [6]. ∇

Note that the equality $S_{\mathrm{Idl}}\mathcal{H}_{\mathrm{opt}} = I$ for the ideal sampler means that the hold function $\phi_{\mathrm{opt}}(t)$ at the sampling instances, kh, equals the Kronecker delta $\bar{\delta}[k]$. Indeed it does in the above example.

As shown in [16] (see also [20, p. 575]), in many cases, sequences of hold functions $\phi_n(t)$ converge towards $\mathrm{sinc}_h(t)$ as n approaches infinity. For our hold functions that would mean that often sequences of Fourier transforms

$$\Phi_{\mathrm{opt},n}(\mathrm{j}\omega) = \frac{h|G_{v,n}(\mathrm{j}\omega)|^2}{\sum_{k\in\mathbb{Z}}|G_{v,n}(\mathrm{j}(\omega + 2k\omega_{\mathrm{N}}))|^2} \tag{11}$$

converge to $h\mathbb{1}_{[-\omega_{\mathrm{N}},\omega_{\mathrm{N}}]}$ as $n \to \infty$. This convergence occurs iff the corresponding signal generator $\mathcal{G}_{v,n}$ becomes more and more "baseband dominant" as $n \to \infty$. To be more precise, introduce the following definitions:

Definition 3.1: A SISO LCTI system $\mathcal{W}$ is said to be *baseband dominant* if a $c \in [0,1]$ exists such that

$$|W(\mathrm{j}\omega_k)| \le c|W(\mathrm{j}\omega_0)|, \quad \forall k \ne 0,\ \omega_0 \in (-\omega_{\mathrm{N}}, \omega_{\mathrm{N}}).$$

If the inequality above holds for a $c < 1$, then $\mathcal{W}$ is said to be *strict baseband dominant*. ▽

It is easy to see that every real system whose frequency response is monotonically decreasing over positive frequency is strict baseband dominant.

Proposition 3.2: If $\mathcal{W}$ is stable LCTI and strict baseband dominant and if the sampler is ideal, then for $\mathcal{G}_{v,n} := \mathcal{W}^n$ the optimal hold (9) converges to $\mathcal{H}_{\mathrm{sinc}}$ as $n \to \infty$.

Proof: For this $G_{v,n}$ the right-hand side of (11) converges to h if $\omega \in (-\omega_{\mathrm{N}}, \omega_{\mathrm{N}})$ and converges to 0 if $|\omega| > \omega_{\mathrm{N}}$. It converges to the Fourier transform of $\mathrm{sinc}_h(t)$. Stability and strict baseband dominance imply that $\|\mathcal{G}_{v,n}\|_2 < \infty$ and that the denominator in (11) is $< \infty$ for every ω. Moreover, the convergence is in L^2 signal norm, which guarantees that the limit is well defined (in both time and frequency domain). ■

The signal interpretation of this result is intuitive: in the limit $n \to \infty$ the signals $v = \mathcal{G}_{n,v}w$ are effectively bandlimited to $[-\omega_{\mathrm{N}}, \omega_{\mathrm{N}}]$ and indeed, as Shannon dictates, holding with the sinc_h is then the best one can do (irrespective of w).

B. Optimal Hold for Unstable Signal Generators

A popular class of hold functions are the cardinal polynomial spline hold functions [4]. These are polynomial splines of odd degree $2n-1$ ($n = 1, 2, \ldots$) and which are $2n-2$ times continuously differentiable. Further they are in $L^2(\mathbb{R})$ and are required to satisfy the consistency property that $\phi(kh) = \bar{\delta}[k]$. This makes them unique. Fig. 2 shows these hold functions for $n = 1$ and $n = 2$.

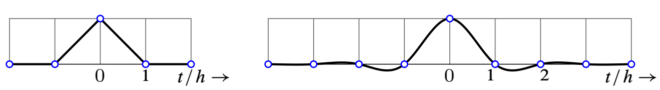


Fig. 2. Cardinal polynomial spline hold functions of degree 1 and 3

A natural question now is: with respect to what class of signals are these hold functions optimal? If in Example 3.1 we let α approach zero then its hold function approaches a cubic spline (this requires some work) and the signal generator approaches the double integrator $1/s^2$. It suggests that cubic cardinal polynomial splines are optimal with respect to doubly integrated white noise (integrated Brownian motion) or doubly integrated L^2 signals, so slowly varying signals. More generally, we claim the following.

Theorem 3.3: If $\mathcal{G}_y = \mathcal{G}_v$ are LCTI integrators of order n, $G_v(s) = G_y(s) = 1/s^n$, then the hold function of (11) is the unique $(2n-2)$-smooth $(2n-1)$-degree polynomial spline in $L^2(\mathbb{R})$ for which $\phi_{\text{opt}}(kh) = \bar{\delta}[k]$. ∇

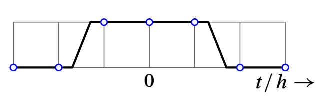
A proper proof is in the Appendix. A dubious derivation, but an insightful one nonetheless, goes as follows. Consider the normal equation $(1-\mathcal{H}S_{\text{Idl}})\mathcal{G}_v\mathcal{G}_v^* S_{\text{Idl}}^* = 0$. Now the adjoint S_{Idl}^* of the ideal sampler is the delta-hold operator and hence $\mathcal{G}_v\mathcal{G}_v^* S_{\text{Idl}}^*\bar{u}$ for any signal $\bar{u}$ is a delta-train integrated $2n$ times, i.e., some $(2n-2)$-smooth $(2n-1)$-degree polynomial spline. So the equality $(1-\mathcal{H}S_{\text{Idl}})\mathcal{G}_v\mathcal{G}_v^* S_{\text{Idl}}^*\bar{u} = 0$ means that the polynomial spline equals $\mathcal{H}$ applied to the sampled polynomial spline. By linearity and discrete-time invariance, it is sufficient to consider the case that $S_{\text{Idl}}\mathcal{G}_v\mathcal{G}_v^* S_{\text{Idl}}^*\bar{u}$ is the unit pulse. There is a unique polynomial spline $\phi \in L^2(\mathbb{R})$ of the given smoothness and degree interpolating the unit pulse [4].

In [18, Remark 20.7] it is shown, as a by product, that the hold (9) is actually the stable hold that makes the error system $(I-\mathcal{H}S)\mathcal{G}_v$ stable and minimizes its L^2-norm. An alternative approach to handle unstable weights (nonstationary processes) can be found in [21]. These formulations circumvent stability of the signal generators and also apply to integrators $G_v(s) = 1/s^n$. Incidentally, since $1/s$ is strict baseband dominant, these cardinal polynomial spline functions converge towards $\text{sinc}_h(t)$ as $n \to \infty$ (Proposition 3.2 and [4, § III.D]).

C. Hybrid Signal Generators: Some Connections

Although we assumed so far that $\mathcal{G}$ is an analog system, Proposition 3.1 (as well as 2.1) is valid for hybrid $\mathcal{G}$, including D/A and A/D components, too. In such cases, the exogenous signal and / or the error to be kept down might also include discrete components. Below, we present two examples of this situation with the aim to highlight connections between our approach and some related results in the literature.

Example 3.2: Following [22], let $\mathcal{G}_y = \mathcal{G}_v = \mathcal{H}_v$ be the hold with

$\phi_v(t) =$ 

and take the ideal sampler S_{Idl}. In fact, the precise shape of $\phi_v(t)$ on $[-2h, h]$ and $[h, 2h]$ is not important, but their symmetry that they add up to 1 for all intersample time,

$$\phi_v(-2h+\tau)+\phi_v(h+\tau)=1, \quad \forall \tau \in [0,h], \tag{12}$$

is. The cascade $\mathcal{S}_{\text{Idl}}\mathcal{G}_v$ is then the discrete FIR system with the transfer function $z+1+z^{-1}$. In [22, p. 1095] it is claimed that then no hold $\mathcal{H}$ exists that reconstructs the input to the sampler error free, because some inverse needed in the process is not defined. That implication is not correct. The normal equation is singular but not unsolvable. To see this, note that (8) in lifted frequency domain reads

$$\grave{H}_v(z)(z+1+z^{-1})=\grave{H}_{\text{opt}}(z)(z+1+z^{-1})^2 \tag{13}$$

and indeed $z=\mathrm{e}^{\mathrm{j}2\pi/3}$ is a zero of the right-most term and so that term has no stable inverse. These zeros, however, cancel against zeros of $\grave{H}_v(z)$, which can be seen via its kernel

$$\begin{aligned}\breve{\phi}_v(z;\tau) &= \diagup\, z^2 + \sqcap\,(z+1) + \diagdown\, z^{-1}\\ &= \Big(\sqcap + \diagup\,(z-1)\Big)(z+1+z^{-1}).\end{aligned} \tag{14}$$

Therefore, the hold with the kernel

$$\breve{\phi}_{\text{opt}}(z;\tau) = \sqcap + \diagup\,(z-1) = \diagup\, z + \diagdown$$

solves (13). This defines an FIR system, reminiscent of the predictive first-order hold [1, Example 3.3]. In hindsight it is easy to see that this hold is optimal, and in fact it is error free (i.e., $\mathcal{G}_{\text{e}}=0$).

Crucial in the derivation is the symmetry (12). If this symmetry is absent then the unit circle zeros of $z+1+z^{-1}$ reappear in the (unique) solution of (13) as poles, rendering it unstable. Yet even in this case one can approach the perfect reconstruction arbitrarily close by a stable $\mathcal{H}$. ▽

For LCTI signal generators $\mathcal{G}_y=\mathcal{G}_v$ and the ideal sampler, the conclusions are very similar and this, once again, is best seen from its classic Fourier transform: while the explicit formula (9) requires $\mathcal{S}_{\text{Idl}}\mathcal{G}_v(\mathcal{S}_{\text{Idl}}\mathcal{G}_v)^*$ to be stably invertible, for the normal equations to hold for some stable hold we merely need that its Fourier transform

$$\Phi_{\text{opt}}(\mathrm{j}\omega)=\frac{h|G_v(\mathrm{j}\omega)|^2}{\sum_{k\in\mathbb{Z}}|G_v(\mathrm{j}(\omega+2k\omega_{\text{N}}))|^2}$$

determines a stable system. Evidently, we have $|\Phi_{\text{opt}}(\mathrm{j}\omega)|\le h$ for every ω and so stability of the hold is, for instance, ensured if $|\omega|^{\gamma}G_v(\mathrm{j}\omega)$ is bounded for some $\gamma>1/2$ [1, §VI-A]. Note that $\mathcal{S}_{\text{Idl}}\mathcal{G}_v(\mathcal{S}_{\text{Idl}}\mathcal{G}_v)^*$ is stable and stably invertible iff $\frac{1}{\epsilon}>\sum_{k\in\mathbb{Z}}|G_v(\mathrm{j}(\omega+2k\omega_{\text{N}}))|^2>\epsilon$ for some $\epsilon\in(0,1)$ and all $\omega\in\mathbb{R}$.

Example 3.3: The configuration in Fig. 3 corresponds to a particular case of the optimization-based approach proposed in [23] (see also [19, Eqn. (12)] for the noise-free case). The goal is to reconstruct

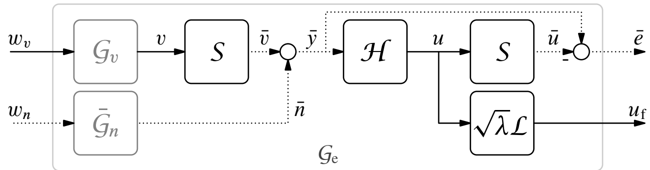


Fig. 3. Reconstruction setup for Example 3.3

an analog signal v from measurements of its sampled version $\bar v$ corrupted by a discrete noise $\bar n$. This is done via choosing the hold $\mathcal{H}$ that minimizes $\|\bar e\|_2^2 + \lambda\|\mathcal{L}u\|_2^2$. This cost penalizes both the deviation of the sampled version of u from the measurement $\bar y$ (data term) and the weighted reconstruction itself (regularization term), where the "amount of regularization imposed on the reconstruction" is controlled by the parameter $\lambda > 0$ and the dynamic weight $\mathcal{L}$ is designated to determine the level of smoothness of u. To minimize the L^2-norm of the error system,

$$\mathcal{G}_e = \begin{bmatrix} S\mathcal{G}_v & \bar{\mathcal{G}}_n \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} S \\ \sqrt{\lambda}\mathcal{L} \end{bmatrix} \mathcal{H} \begin{bmatrix} S\mathcal{G}_v & \bar{\mathcal{G}}_n \end{bmatrix},$$

we may use Proposition 2.1, which yields the following normal equation (providing $S\mathcal{G}_v(S\mathcal{G}_v)^* + \bar{\mathcal{G}}_n\bar{\mathcal{G}}_n^*$ is nonsingular):

$$S^* = \bigl(S^*S + \lambda\mathcal{L}^*\mathcal{L}\bigr)\mathcal{H}_{\text{opt}}. \tag{15}$$

Remarkably, (15) is independent of $\mathcal{G}_v$ and $\bar{\mathcal{G}}_n$, i.e., of properties of v and $\bar n$. This appears to be the rationale behind the choice $\bar e = \bar y - \bar u$ rather than the seemingly more natural $\bar e = \bar v - \bar u$. Assuming that $\mathcal{L}$ is stably invertible[2], (15) rewrites as $(S\mathcal{W})^* = \bigl((S\mathcal{W})^*S\mathcal{W} + \lambda I\bigr)\mathcal{W}^{-1}\mathcal{H}_{\text{opt}}$, where $\mathcal{W} := \mathcal{L}^{-1}$. This leads to

$$\begin{aligned}\mathcal{H}_{\text{opt}} &= \mathcal{W}\bigl(\lambda I + (S\mathcal{W})^*S\mathcal{W}\bigr)^{-1}(S\mathcal{W})^* \\ &= \mathcal{W}(S\mathcal{W})^*\bigl(\lambda I + S\mathcal{W}(S\mathcal{W})^*\bigr)^{-1}.\end{aligned}$$

As $\lambda \to 0$ (see [19]), this recovers (9) with $\mathcal{G}_y = \mathcal{G}_v = \mathcal{L}^{-1}$. For nonzero λ we can end up with this reconstructor with our setup in Fig. 1 by adding a discrete white noise of variance λ between S and $\mathcal{H}$, see [12, Thm. 1].

We thus see that there might be a number of approaches, viz. optimization criteria, enabling us to end up with the very same solution. An interesting question is now to compare these approaches. This may

[2]Otherwise, the arguments of the proof of Theorem 3.3 may be used.

be a right place to reemphasize our statement from [1, Remark 2.2] that we consider the optimization as merely a design tool. We therefore believe that a fair comparison between our formulation and those of [19], [23] should be drawn from the transparency of the design steps and tuning the weighting functions and from the extensibility of the method. These issues deserve an in-depth analysis, which goes beyond the scope of this paper. ∇

IV. Rank Theorem

Samplers, by their very nature, reduce continuous-time signals to discrete-time signals. Clearly then sampling normally brings about a loss of information. Dually, the output of a hold is continuous time, but as the hold is shift-invariant and driven by a discrete signal, the richness of the set of its continuous-time outputs is limited. Typically this set is nevertheless infinite dimensional and it is difficult to get a handle on the richness of the set in time domain. In lifted frequency domain matters are transparent and in fact one can fully characterize what it means for an LDTI system to be a series interconnection of a sampler and a hold.

First, recall that the series interconnection $u = \mathcal{H}Sy$ in lifted frequency domain is an integral operator

$$\breve{u}(\mathrm{e}^{\mathrm{j}\theta};\tau) = \int_0^h \breve{f}_{\mathrm{HSP}}(\mathrm{e}^{\mathrm{j}\theta};\tau,\sigma)\breve{y}(\mathrm{e}^{\mathrm{j}\theta},\sigma)\mathrm{d}\sigma \tag{16}$$

whose kernel can be expressed in terms of its sampling and hold functions as

$$\breve{f}_{\mathrm{HSP}}(\mathrm{e}^{\mathrm{j}\theta};\tau,\sigma) = \breve{\phi}(\mathrm{e}^{\mathrm{j}\theta};\tau)\,\breve{\psi}(\mathrm{e}^{\mathrm{j}\theta};-\sigma), \tag{17}$$

see Appendix for a derivation. At each θ the range of the integral operator (16) is contained in the subspace spanned by $\breve{\phi}(\mathrm{e}^{\mathrm{j}\theta};\tau)$. If the input of the hold is a channel with $n_{\bar{u}}$ elements then the dimension of this subspace is $n_{\bar{u}}$ (at most). The ramification of this observation is:

Theorem 4.1 (Rank Theorem): Let $\mathcal{F} \in L^\infty$ and suppose that its frequency response kernel $\breve{f}(\mathrm{e}^{\mathrm{j}\theta};\tau,\sigma)$ is piecewise continuous. Then $\mathcal{F}$ is an HSP iff there is $r \in \mathbb{N}$ such that $\operatorname{rank} \breve{F}(\mathrm{e}^{\mathrm{j}\theta}) \le r \;\; \forall\theta \in [-\pi,\pi]$. In this case $r \le \min(n_{\bar{y}}, n_{\bar{u}})$ for any HSP implementation of $\mathcal{F}$, and HSP-implementations of $\mathcal{F}$ exist for which $r = n_{\bar{y}} = n_{\bar{u}}$.

Proof: See Appendix. ■

The assumption on piecewise continuity of the kernel avoids issues with Lebesgue measure but other than that it is not essential to the result. It is because of this Rank Theorem that of all representations of systems, the lifted frequency response is the most useful one, at least for the design problems considered in the remainder of this paper.

V. Singular Values and Optimal HSP

Having characterized HSPs as having a uniform finite rank frequency response at each θ, the design of HSPs amounts to frequency-wise approximation of given operators by finite rank operators. In the (finite-dimensional) matrix case this could be done via the singular vale decomposition (SVD) machinery [24, §2.5.5]. An extension of this to infinite-dimensional Hilbert space operators is called the *Schmidt decomposition*. Omitting some technical details, for which the reader may refer to [25, Ch. VI] or [26, §A.4.2], a compact Hilbert space operator $O : \mathbb{H}_1 \to \mathbb{H}_2$ admits a representation of the form

$$O = \sum_{k\in\mathbb{N}} \sigma_k \langle \cdot, e_k\rangle_{\mathbb{H}_1} v_k, \tag{18}$$

where real $\sigma_1 \geq \sigma_2 \geq \cdots$ are nonnegative and $\{e_1, e_2, \ldots\}$ and $\{v_1, v_2, \ldots\}$ are orthonormal bases in $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively. The notation $\langle \cdot, e_k\rangle_{\mathbb{H}_1}$ means $\langle \eta, e_k\rangle_{\mathbb{H}_1}$ for any input $\eta \in \mathbb{H}_1$ of O. The numbers σ_k and pairs (e_k, v_k) are called the *singular values* (s-numbers) and the *Schmidt pairs* of O, respectively (the Schmidt pairs may be thought of as counterparts of the singular vectors in the matrix case).

We then have:

Theorem 5.1: Let $\mathcal{G} \in L^\infty$ and suppose that at almost every $\theta \in [-\pi, \pi]$ the operator $\breve{G}(\mathrm{e}^{\mathrm{j}\theta}) : \mathbb{L} \to \mathbb{L}$ has SVD of the form (18) for θ-dependent singular values and Schmidt pairs. Then if the HSP $\mathcal{G}_{\mathrm{HSP}}$ satisfying

$$\breve{G}_{\mathrm{HSP}}(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{k=1}^{r} \sigma_k \langle \cdot, e_k\rangle_{\mathbb{L}} v_k, \tag{19}$$

is well defined, it minimizes $\|\mathcal{G} - \mathcal{G}_{\mathrm{HSP}}\|_\infty$ over all HSPs of rank $\leq r$, attaining $\|\mathcal{G} - \mathcal{G}_{\mathrm{HSP}}\|_\infty = \operatorname{ess\,sup}_{\theta\in(-\pi,\pi)} \sigma_{r+1}(\theta)$. If $\mathcal{G}$ has finite L^2-norm, then the HSP (19) minimizes $\|\mathcal{G} - \mathcal{G}_{\mathrm{HSP}}\|_2$ as well, attaining $\|\mathcal{G} - \mathcal{G}_{\mathrm{HSP}}\|_2^2 = \frac{1}{2\pi h} \int_{-\pi}^{\pi} \sum_{k=r+1}^{\infty} \sigma_k^2(\theta) \mathrm{d}\theta$.

Proof: The L^2 and L^∞ norms involve nonnegative integrals over frequency θ, see [1, Eqns. (31) and (30)]. Therefore, if $\breve{G}_{\mathrm{HSP}}(\mathrm{e}^{\mathrm{j}\theta})$ minimizes the norms for every fixed frequency then it is optimal. The rest is standard. ■

This theorem does not settle the potentially complicated matter of existence of such SVDs and whether or not the frequency-wise defined HSP (19) can be implemented. For the applications that we have in mind, however, the SVD of $\breve{G}(\mathrm{e}^{\mathrm{j}\theta})$ exists and is explicit and the pointwise HSP can be implemented as convolutions.

Typically HSPs are not LCTI and it is not hard to formalize that the subset of HSPs that are LCTI form a set of measure zero. However if $\mathcal{G}$ is LCTI then often the optimal finite-rank approximation $\mathcal{G}_{\mathrm{HSP}}$ of $\mathcal{G}$ is LCTI as well. This follows from explicit representations in the next section but it can also be understood from the fact that the L^2 and L^∞ system norms are invariant under continuous time shift:

Proposition 5.2: Given LCTI system $\mathcal{G}$, the minimizer $\mathcal{G}_{\text{HSP}}$ of $\|\mathcal{G}-\mathcal{G}_{\text{HSP}}\|_2$ or $\|\mathcal{G}-\mathcal{G}_{\text{HSP}}\|_\infty$ over noncausal LDTI HSPs of given rank is LCTI if it is unique.

Proof: The L^2 and L^∞ norms do not depend on shifts of input and output: $\|\Delta^\tau(\mathcal{G}-\mathcal{G}_{\text{HSP}})\Delta^{-\tau}\| = \|\mathcal{G}-\mathcal{G}_{\text{HSP}}\|$ where Δ^τ is delay operator ($\tau \in \mathbb{R}$). By continuous time-invariance of $\mathcal{G}$ the $\mathcal{G}_{\text{HSP}}$ hence is optimal iff $\Delta^{-\tau}\mathcal{G}_{\text{HSP}}\Delta^\tau$ is optimal for all $\tau \in \mathbb{R}$. ∎

Subsequently, we shall also need the following result:

Corollary 5.3: Let $\mathcal{G}$ be as in Theorem 5.1. Then the rank-r $\mathcal{F}_{\text{HSP}}$ with the frequency response satisfying

$$\breve{F}_{\text{HSP}}(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{k=1}^{r} \langle \cdot, v_k\rangle_{\mathbb{L}} v_k, \quad \forall \breve{y} \in \mathbb{L},$$

minimizes both $\|(I-\mathcal{F}_{\text{HSP}})\mathcal{G}\|_\infty$ and $\|(I-\mathcal{F}_{\text{HSP}})\mathcal{G}\|_2$ (provided $\|\mathcal{G}\|_2 < \infty$) with respect to stable rank-r HSPs, attaining the same norms as in Theorem 5.1.

Proof: Then $\breve{G}_{\text{HSP}} := \breve{F}_{\text{HSP}}\breve{G}$ equals (19). ∎

VI. SVD of LCTI Systems—Frequency Folding

LCTI systems have an explicit fixed frequency SVD. This is very similar to what [27, p. 1770] derived in discrete time and for spectral densities. We need it for signal generators:

Proposition 6.1: Let $\mathcal{G} \in L^\infty \cap L^2$. Then $\breve{G}(\mathrm{e}^{\mathrm{j}\theta})$ exists for almost every $\theta \in [-\pi, \pi]$ and has SVD

$$\breve{G}(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{k\in\mathbb{Z}} |G(\mathrm{j}\omega_k)|\, \langle \cdot, e_k\rangle_{\mathbb{L}}\, v_k, \tag{20}$$

in which

$$e_k(\tau) := \frac{1}{\sqrt{h}}\mathrm{e}^{\mathrm{j}\omega_k\tau}, \quad k \in \mathbb{Z}, \tag{21}$$

is the standard orthonormal basis of $\mathbb{L}$ and $v_k := e_k \mathrm{e}^{\mathrm{j}\arg G(\mathrm{j}\omega_k)}$. The singular values in this case are well defined at almost every θ and equal $\sigma_k(\theta) = |G(\mathrm{j}\omega_k)|$, $k \in \mathbb{Z}$, modulo ordering.

Proof: By [1, (17b)] we have that the kernel of $\breve{G}(\mathrm{e}^{\mathrm{j}\theta})$ equals

$$\breve{g}(\mathrm{e}^{\mathrm{j}\theta}; \tau, \sigma) = \frac{1}{h}\sum_{k\in\mathbb{Z}} G(\mathrm{j}\omega_k)\mathrm{e}^{\mathrm{j}\omega_k(\tau-\sigma)} \tag{22}$$

so its frequency response, mapping $\breve{w}(\mathrm{e}^{\mathrm{j}\theta})$ to $\breve{v}(\mathrm{e}^{\mathrm{j}\theta})$, reads

$$\begin{aligned}\breve{v}(\mathrm{e}^{\mathrm{j}\theta}; \tau) &= \frac{1}{h}\sum_{k\in\mathbb{Z}}\int_0^h G(\mathrm{j}\omega_k)\mathrm{e}^{\mathrm{j}\omega_k(\tau-\sigma)}\breve{w}(\mathrm{e}^{\mathrm{j}\theta}; \sigma)\,\mathrm{d}\sigma \\ &= \sum_{k\in\mathbb{Z}} G(\mathrm{j}\omega_k)\,\langle \breve{w}(\mathrm{e}^{\mathrm{j}\theta}), e_k\rangle_{\mathbb{L}}\, e_k(\tau).\end{aligned}$$

Since the functions e_k are orthonormal in $\mathbb{L}$, the absolute values $|G(\mathrm{j}\omega_k)|$ are the singular values (modulo order). The fact that $G \in L^2 \cap L^\infty$ implies the existence of singular values and, by Plancherel, that $\breve{G}(\mathrm{e}^{\mathrm{j}\theta})$ has finite Hilbert-Schmidt norm almost everywhere. ■

This establishes that the singular values of $\breve{G}(\mathrm{e}^{\mathrm{j}\theta})$ are actually the magnitudes of the continuous-time frequency response $G(\mathrm{j}\omega)$ at all its *aliased frequencies* ω_k. This can be visualized by folding the magnitude plot of $G(\mathrm{j}\omega)$, see Fig. 4. Folding reduces the infinite frequency bands to the finite baseband $[0, \omega_\mathrm{N}]$ and we end up with a zig-zag plot that at each $\theta/h = \omega_0 \in [0, \omega_\mathrm{N}]$ captures its countably many singular values $\sigma_1, \sigma_2, \ldots$ Frequency folding is well known in the literature as a way to explain aliasing or to visualize the sampled spectrum [10, §6.1]. In the lifting approach we do not add up the $G(\mathrm{j}\omega_k)$—which would result in the sampled spectrum and thus loose intersample information—but keep them as separate entities.

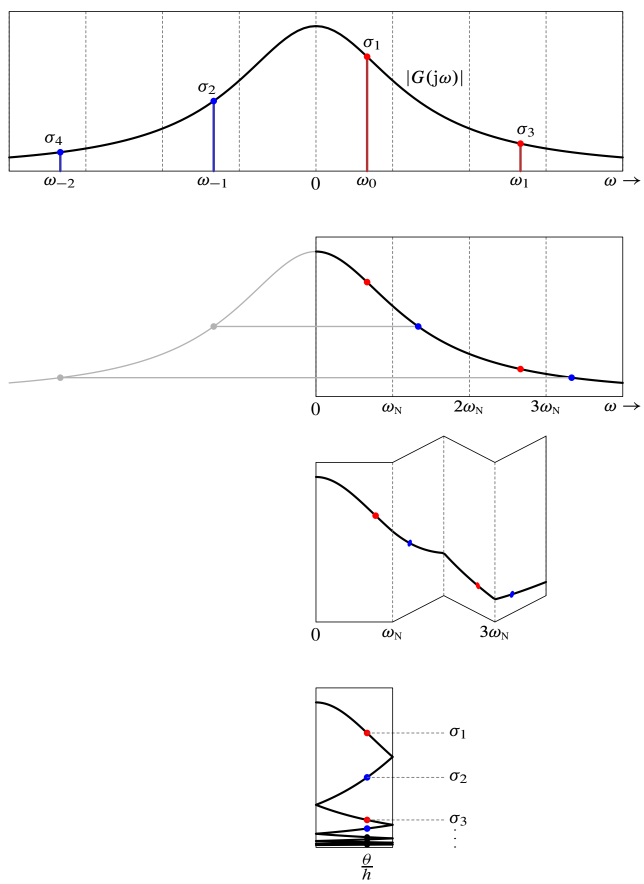


Fig. 4. Frequency folding for LCTI systems: at each $\theta/h \in [0, \omega_\mathrm{N}]$ the $\breve{G}(\mathrm{e}^{\mathrm{j}\theta})$ has countably many singular values $\sigma_k = |\breve{G}(\mathrm{e}^{\mathrm{j}\omega_k})|$, modulo ordering

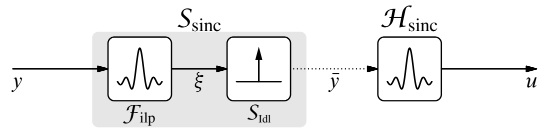


Fig. 5. WKS hybrid signal processor

Example 6.1 (WKS-block): Consider the HSP of Fig. 5. It comprises the sinc sampler

$$\bar{y} = S_{\rm sinc}(y): \quad \bar{y}[k] = \int_{-\infty}^{\infty} \tfrac{1}{h}\,\mathrm{sinc}_h(kh - s)y(s)\mathrm{d}s \tag{23}$$

(presented in the figure as the cascade of the ideal lowpass filter $\mathcal{F}_{\rm ilp}$ and $S_{\rm Idl}$) and the sinc-hold

$$u = \mathcal{H}_{\rm sinc}(\bar{u}): \quad u(t) = \sum_{i\in\mathbb{Z}} \mathrm{sinc}_h(t - ih)\bar{u}[i]. \tag{24}$$

Here $\xi := \mathcal{F}_{\rm ilp} y$ is the projection of y into the space of $\omega_{\rm N}$-bandlimited signals. It follows from the Sampling Theorem that $u = \xi$ and, moreover, if y itself is $\omega_{\rm N}$-bandlimited, that we have perfect reconstruction, $u = y$. We call this system the *Whittaker-Kotel'nikov-Shannon* (WKS) block, and denote it as $\mathcal{F}_{\rm WKS}$. According to (17) and [1, Examples 4.5 and 4.6], the frequency response kernel of $\breve{F}_{\rm WKS}(\mathrm{e}^{\mathrm{j}\theta})$ is $\breve{f}_{\rm WKS}(\mathrm{e}^{\mathrm{j}\theta}; \tau, \sigma) = \breve{\phi}_{\rm sinc}(\mathrm{e}^{\mathrm{j}\theta}; \tau)\breve{\psi}_{\rm sinc}(\mathrm{e}^{\mathrm{j}\theta}; -\sigma) = \frac{1}{h}\mathrm{e}^{\mathrm{j}\theta(\tau-\sigma)/h}$. Note that this kernel has a Toeplitz structure. Together with the discrete-time invariance of $\mathcal{F}_{\rm WKS}$, this implies that $\mathcal{F}_{\rm WKS}$ is actually LCTI. This may appear remarkable, taking into account that generically HSPs are LDTI and typically not LCTI.

Alternatively, since the WKS-block is LCTI with the real frequency response $F(\mathrm{j}\omega) = \mathbb{1}_{[-\omega_{\rm N},\omega_{\rm N}]}(\omega)$ we have, according to Proposition 6.1, that $\breve{F}(\mathrm{e}^{\mathrm{j}\theta}) = \langle\cdot, e_0\rangle_{\mathbb{L}} e_0$ and that its frequency response kernel is $\breve{f}(\mathrm{e}^{\mathrm{j}\theta}; \tau, \sigma) = e_0(\tau)e_0^*(\sigma) = \frac{1}{h}\mathrm{e}^{\mathrm{j}\theta(\tau-\sigma)/h}$. Indeed. ▽

VII. Single-Channel Optimal SR

We are now in a position to formulate and solve a number of Type-IV signal reconstruction problems, i.e., problems where both sampler and hold are available for design.

In this section we return to the case that $\mathcal{G}_y = \mathcal{G}_v$. The error system we write as $\mathcal{G}_{\rm e} = (I - \mathcal{F}_{\rm HSP})\mathcal{G}_v = \mathcal{G}_v - \mathcal{F}_{\rm HSP}\mathcal{G}_v$, where $\mathcal{F}_{\rm HSP} := \mathcal{H}\mathcal{S}$. We further restrict attention to single channel HSPs. Single-channel refers to the case that the sampled signal $\bar{y}$ is scalar, i.e., that we have only one sensor. The rank theorem thus states that rank $\breve{F}_{\rm HSP}(\mathrm{e}^{\mathrm{j}\theta}) \le 1, \forall\theta \in [-\pi, \pi]$ for any such HSP. This clearly implies that the best we can do with our HSP is to match the directions and norm (Schmidt pair) corresponding to the largest singular value of $\breve{G}_v(\mathrm{e}^{\mathrm{j}\theta})$ at each frequency θ and have a unit gain there. To simplify the outline, we assume that

$\boldsymbol{\mathcal{A}_1}$: $\mathcal{G}_v$ is baseband dominant

(see Definition 3.1). $\boldsymbol{\mathcal{A}_1}$ says that at each $\theta \in [-\pi, \pi]$ the largest singular value of $\breve{G}_v(\mathrm{e}^{\mathrm{j}\theta})$ is attained in the baseband. By Corollary 5.3 and Proposition 6.1, the optimal rank-one $\breve{F}(\mathrm{e}^{\mathrm{j}\theta})$ has the kernel

$$\breve{f}_{\mathrm{HSP}}(\mathrm{e}^{\mathrm{j}\theta}; \tau, \sigma) = v_0(\tau) v_0^*(\sigma) = \tfrac{1}{h} \mathrm{e}^{\mathrm{j}\theta(\tau-\sigma)/h},$$

meaning that the optimal HSP is actually $\mathcal{F}_{\mathrm{WKS}}$. Thus, we just proved the following result:

Theorem 7.1: Suppose $\mathcal{G}_v \in L^\infty \cap L^2$ is LCTI and that it satisfies $\boldsymbol{\mathcal{A}_1}$. Then the WKS block $\mathcal{F}_{\mathrm{WKS}}$ considered in Example 6.1 is the HSP that minimizes both L^2 and L^∞ norms of $\mathcal{G}_{\mathrm{e}}$, and the optimal performance indices are

$$\|(I - \mathcal{F}_{\mathrm{WKS}})\mathcal{G}_v\|_2^2 = \frac{1}{\pi} \int_{\omega_{\mathrm{N}}}^{\infty} |G_v(\mathrm{j}\omega)|^2 \, \mathrm{d}\omega \tag{25}$$

in the L^2 case, and

$$\|(I - \mathcal{F}_{\mathrm{WKS}})\mathcal{G}_v\|_\infty = \sup\nolimits_{\omega > \omega_{\mathrm{N}}} |G_v(\mathrm{j}\omega)| \tag{26}$$

in the L^∞ case.

This result is not new for the L^2-norm. It was derived earlier in [27] using similar methods, but then for the discrete time case. An elegant and entirely different derivation can be found in [15, p. 3593], again for the L^2 norm. Computation of the L^2 norm (25) can be done without gridding [28].

If $\mathcal{G}_v$ is *strict* baseband dominant then the optimal HSP is unique. Theorem 7.1 establishes that sinc-sampler (23) and sinc-hold (24) are optimal from both L^2 and L^∞ points of view. Interestingly, neither the optimal sampler nor the optimal hold depends on $\mathcal{G}_v$ as long as $\mathcal{G}_v$ is baseband dominant. Clearly under the baseband dominance assumption the norm of the reconstruction error is zero iff $G_v(\mathrm{j}\omega) = 0$ almost everywhere outside the baseband $[-\omega_{\mathrm{N}}, \omega_{\mathrm{N}}]$. This is the classic Sampling Theorem.

If $\mathcal{G}_v$ is not baseband-dominant, then the optimal $\mathcal{F}_{\mathrm{HSP}}$ should account for frequency band(s) in which the frequency response gain of $\mathcal{G}_v$ is dominant. In this case, the optimal sampler comprises the ideal sampler and an ideal passband filter. The frequency pattern of the latter might be rather complicated. Also, the perfect reconstruction conditions will be different in this case. The sampled signal need no longer have zero frequency content outside the baseband. Rather, we should require that $G_v(\mathrm{j}\omega_k) \neq 0$ for at most one k (which is not necessarily $k = 0$). The optimal $\mathcal{F}_{\mathrm{HSP}}$ is nonetheless selfadjoint, consistent and LCTI and its classic Fourier transform is piecewise constant having value 0 or 1, a so called brickwall filter [27].

Remark 7.1: It is straightforward to extend these ideas to multi-input-multi-output (MIMO) systems $\mathcal{G}_v$. In such cases, $\breve{G}_v(\mathrm{j}\omega_k)$ is a matrix and, for every k, has a finite number of singular values $\sigma_{k,n}(\theta)$,

$n \in \mathbb{N}$, with respect to the standard Euclidean norm. Thus, at each $\theta \in [0, \pi]$ we end up with doubly indexed singular values, but the task of the HSP remains the same: to delete the largest singular value. The optimal HSP is again a (modulated) WKS-block, but then pre- and post processed by MIMO LCTI systems that select, so the say, the direction of the largest singular value of G_v. ∇

A. Fundamental Limit for Error-Free Reconstruction

The optimal mapping $\mathcal{F}_{\mathrm{HSP}}$ selects frequency bands where $|G_v(\mathrm{j}\omega)|$ is maximal and with that in mind one can obtain the upper bound $\|\mathcal{F}_{\mathrm{HSP}}\mathcal{G}_v\|_2^2 \leq \|\mathcal{G}_v\|_\infty^2/h$ and that the upper bound is tight (in a ratio sense) if $h \to \infty$ [28]. By orthogonality we also have the upper bound $\|\mathcal{F}_{\mathrm{HSP}}\mathcal{G}_v\|_2^2 \leq \|\mathcal{G}_v\|_2^2$. The two upper bounds meet at

$$h_G := \|\mathcal{G}\|_\infty^2/\|\mathcal{G}_v\|_2^2,$$

which has an interesting property:

Proposition 7.2: Whatever $\mathcal{G}_v$ is, error free reconstruction is impossible for $h > h_G$. ∇

This follows from the lower bound on the error reconstruction, $\|\mathcal{G}_\mathrm{e}\|_2^2 = \|\mathcal{G}_v\|_2^2 - \|\mathcal{F}_{\mathrm{HSP}}\mathcal{G}_v\|_2^2 \geq \|\mathcal{G}_v\|_2^2 - \|\mathcal{G}_v\|_\infty^2/h = \|\mathcal{G}_v\|_2^2(1 - h_G/h)$. Stated differently, the "signal-to-error ratio" (SER) is bounded from above by

$$\mathrm{SER} := \frac{\|\mathcal{G}_v\|_2^2}{\|\mathcal{G}_\mathrm{e}\|_2^2} \leq \frac{1}{1 - h_G/h}, \quad \forall h > h_G.$$

Also the L^∞ norm gives rise to limitations on perfect reconstruction. In fact, for certain values of h the L^∞ norm may not be reducible at all if $|G_v(\mathrm{j}\omega)|$ is not monotonically decaying. Indeed, suppose that the peak value of $|G_v(\mathrm{j}\omega)|$ is attained at some frequency, called resonance frequency,

$$\omega_{\mathrm{res}} := \arg\max_{\omega > 0} |G_v(\mathrm{j}\omega)|.$$

Suppose further that we sample at an integer fraction of the resonance frequency, i.e., at $\omega_\mathrm{N} = \omega_{\mathrm{res}}/k$, for some $k \in \mathbb{N}$ Then folding of $|G_v(\mathrm{j}\omega)|$ shows that there are at least two singular values σ_k equal to $\|\mathcal{G}_v\|_\infty$ at either $\omega = 0$ or $\omega = \omega_\mathrm{N}$, see Fig. 6. Since a single channel hybrid signal processor can cancel

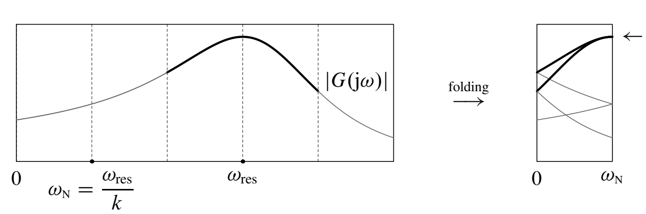


Fig. 6. Samplint at $\omega_\mathrm{N} = \omega_{\mathrm{res}}/k$

only one singular value, the largest singular value can not be reduced at all in this case and therefore we have:

Proposition 7.3: If $|G_v(\mathrm{j}\omega)|$ is continuous and $\omega_{\mathrm{res}} > 0$ then sampling with $\omega_{\mathrm{N}} = \omega_{\mathrm{res}}/k$ is futile: $\|\mathcal{G}_{\mathrm{e}}\|_\infty = \|\mathcal{G}_v\|_\infty$ is the best we can do and $\mathcal{F}_{\mathrm{HSP}} = 0$ is an L^∞-optimal solution. ∇

Example 7.1 (Resonance peaks): Consider the second order LCTI system $\mathcal{G}_v$ with resonance peak near $\omega = 1$,

$$G_v(\mathrm{j}\omega) = \frac{1}{(\mathrm{j}\omega + .2)^2 + 1}$$

Because of the peak, the reconstruction errors norms $\|\mathcal{G}_{\mathrm{e}}\|_2$ and $\|\mathcal{G}_{\mathrm{e}}\|_\infty$ need not be monotonous in the sampling period h, and indeed they are not: Fig. 7 shows the numerically computed $\|\mathcal{G}_{\mathrm{e}}\|_2^2$ and $\|\mathcal{G}_v\|_\infty$ as a function of h. The reconstruction error norms converges to zero as $h \to 0$ and converge to $\|\mathcal{G}_v\|_2$ and $\|\mathcal{G}_v\|_\infty$ respectively as $h \to \infty$. In this example the fundamental time limit is $h_G = \|\mathcal{G}_v\|_\infty^2/\|\mathcal{G}_v\|_2^2 = \frac{2.5^2}{125/104} = 5.2$ exactly. As predicted, the L^∞ norm can not be reduced if $\omega_{\mathrm{N}} = \omega_{\mathrm{res}}/k \approx 1/k$, that is, if $h = k\pi/\omega_{\mathrm{res}} \approx k\pi$. As Fig. 7 suggests also the L^2 norm is close to a local maximum at these values.

Fig. 7. Optimal $\|\mathcal{G}_{\mathrm{e}}\|_2^2$ (left) and $\|\mathcal{G}_{\mathrm{e}}\|_\infty$ (right) as a function of h

This can be interpreted as being close to pathological sampling (see next subsection). ∇

B. Unstable Signal Generators and Pathological Sampling

To avoid technicalities it was assumed so far that the signal generator $\mathcal{G}_v$ is stable. But it is tempting to consider unstable signal generators as well. Bypassing the mathematical difficulties (this will be fixed later), suppose that $G_v(s)$ has several imaginary poles. Clearly after folding we end up with a two or more infinite singular values (poles) at some θ iff

$$\omega_a - \omega_b = 2k\omega_{\mathrm{N}} \quad \text{for some poles } \mathrm{j}\omega_a \neq \mathrm{j}\omega_b \text{ of } G_v(s)$$

and certain $k \in \mathbb{Z}$. This situation is known as *pathological sampling* and it is the case when controllability and/or observability may be lost after standard discretization of a system in state space [29]. Since an HSP can delete only one singular value per discrete frequency, one expects that no HSP can achieve a finite norm if we have pathological sampling. If, on the other hand, no such ω_a, ω_b, and k exist then no two poles overlap after folding, and then an HSP can be found that deletes all infinite singular values (poles), rendering the error system stable. This is indeed the case. For technical reasons we formulate the result for rational $G_v(s)$ only:

Proposition 7.4: Suppose $G_v(s)$ is rational and strictly proper, but possibly with imaginary poles. Then a single channel HSP exists that renders $(I - \mathcal{F}_{\text{HSP}})\mathcal{G}_v$ stable iff h is not pathological with respect to $G_v(s)$. In that case any brick-wall filter $\breve{F}_{\text{HSP}}(\mathrm{e}^{\mathrm{j}\theta})$ that at each θ cancels the largest singular value $\sigma_{\max}(\theta)$ of $\breve{G}_v(\mathrm{e}^{\mathrm{j}\theta})$ (and leaves the other singular values unaffected) is an L^2 optimal rank-1 HSP.

Proof: See Appendix. ■

In particular, for the integrators $G_v(s) = 1/s^n$ the WKS-block once again is optimal under all $h > 0$ (no pathological sampling in this case).

VIII. Multichannel SR, Shannon Extension

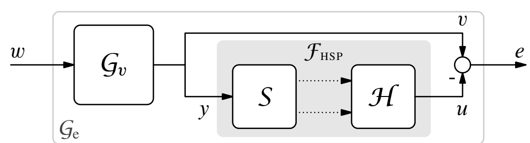


Fig. 8. Two-channel SR setup (Section VIII)

Next we consider the setup depicted in Fig. 8. It is the case where we have two channels, i.e., *two* samplers and two holds. The HSP in this case has the form $\mathcal{F}_{\text{HSP}} = \mathcal{F}_{\text{HSP1}} + \mathcal{F}_{\text{HSP2}}$ for some *scalar* HSPs $\mathcal{F}_{\text{HSP1}}$ and $\mathcal{F}_{\text{HSP2}}$. This leads to the following rank constraint: rank $\breve{F}_{\text{HSP}}(\mathrm{e}^{\mathrm{j}\theta}) \leq 2 \quad \forall\theta \in [-\pi, \pi]$. Assuming $2\omega_{\text{N}}$-baseband dominance of $\mathcal{G}_v$ and following the arguments of the previous section, we obtain the optimal HSP in terms of its lifted frequency response kernel as

$$\breve{f}_{\text{HSP}}(\mathrm{e}^{\mathrm{j}\theta}; \tau, \sigma) = \frac{1}{h}\big(\mathrm{e}^{\mathrm{j}\omega_0(\tau-\sigma)} + \mathrm{e}^{\mathrm{j}\omega_{-1}(\tau-\sigma)}\big), \tag{27}$$

for $\theta \geq 0$ (the negative part follows by symmetry using the assumption that the system is real) and that the optimal L^2 and L^∞ performance indices are as in (25) and (26) with ω_{N} replaced by $2\omega_{\text{N}}$. The optimal HSP is again LCTI and its frequency response is $F_{\text{HSP}}(\mathrm{j}\omega) = \mathbb{1}_{[-2\omega_{\text{N}}, 2\omega_{\text{N}}]}(\omega)$.

Expression (27) does not determine optimal $\mathcal{F}_{\text{HSP1}}$ and $\mathcal{F}_{\text{HSP2}}$ unambiguously. In fact, there is an infinite number of possible combinations in this case. Yet it is clear that we have perfect reconstruction iff we sample at half the Nyquist rate or faster, i.e., iff $G_v(\mathrm{j}\omega)$ is zero outside $[-2\omega_{\text{N}}, 2\omega_{\text{N}}]$ (given the assumed $2\omega_\mathrm{N}$-baseband dominance). In other words there are two scalar HSPs that, combined, can perfectly reconstruct any ω_b-bandlimited signal if *and only if* $\omega_\mathrm{b} < 2\omega_\mathrm{N}$.

The optimal kernel (27) naturally splits into two channels by decomposing it as

$$\begin{aligned}\breve{f}_{\mathrm{HSP}}(\mathrm{e}^{\mathrm{j}\theta};\tau,\sigma) &= \frac{1}{h}\left(\mathrm{e}^{\mathrm{j}\omega_0(\tau-\sigma)} + \mathrm{e}^{\mathrm{j}\omega_{-1}(\tau-\sigma)}\right)\\ &= \begin{bmatrix} \mathrm{e}^{\mathrm{j}\omega_0\tau} & \mathrm{e}^{\mathrm{j}\omega_{-1}\tau}\end{bmatrix}\begin{bmatrix}\frac{1}{h}\mathrm{e}^{-\mathrm{j}\omega_0\sigma}\\ \frac{1}{h}\mathrm{e}^{-\mathrm{j}\omega_{-1}\sigma}\end{bmatrix}\\ &= \begin{bmatrix}\phi_1(\mathrm{e}^{\mathrm{j}\theta}) & \phi_2(\mathrm{e}^{\mathrm{j}\theta})\end{bmatrix}\begin{bmatrix}\psi_1(\mathrm{e}^{\mathrm{j}\theta})\\ \psi_2(\mathrm{e}^{\mathrm{j}\theta})\end{bmatrix} \qquad (28)\end{aligned}$$

with hold and sampling functions defined as

$$\begin{bmatrix}\phi_1(\mathrm{e}^{\mathrm{j}\theta}) & \phi_2(\mathrm{e}^{\mathrm{j}\theta})\end{bmatrix} = \begin{bmatrix}\mathrm{e}^{\mathrm{j}\omega_0\tau} & \mathrm{e}^{\mathrm{j}\omega_{-1}\tau}\end{bmatrix}$$
$$\begin{bmatrix}\psi_1(\mathrm{e}^{\mathrm{j}\theta})\\ \psi_2(\mathrm{e}^{\mathrm{j}\theta})\end{bmatrix} = \frac{1}{h}\begin{bmatrix}\mathrm{e}^{-\mathrm{j}\omega_0\sigma}\\ \mathrm{e}^{-\mathrm{j}\omega_{-1}\sigma}\end{bmatrix}.$$

This corresponds to one channel $\mathcal{H}_1 S_1$ being the standard WKS-block and the other channel $\mathcal{H}_2 S_2$—its modulated version. Many other splittings exist. In fact, (28) holds true for

$$\begin{bmatrix}\phi_1(\mathrm{e}^{\mathrm{j}\theta}) & \phi_2(\mathrm{e}^{\mathrm{j}\theta})\end{bmatrix} = \begin{bmatrix}\mathrm{e}^{\mathrm{j}\omega_0\tau} & \mathrm{e}^{\mathrm{j}\omega_{-1}\tau}\end{bmatrix}\bar{A}^{-1}(\theta) \qquad (29)$$
$$\begin{bmatrix}\psi_1(\mathrm{e}^{\mathrm{j}\theta})\\ \psi_2(\mathrm{e}^{\mathrm{j}\theta})\end{bmatrix} = \bar{A}(\theta)\frac{1}{h}\begin{bmatrix}\mathrm{e}^{-\mathrm{j}\omega_0\sigma}\\ \mathrm{e}^{-\mathrm{j}\omega_{-1}\sigma}\end{bmatrix} \qquad (30)$$

for any 2×2 discrete system $\bar{A}(\theta)$ that is bistable (stable and having stable inverse). This way the two channels could be time varying (as continuous time systems) while we know that their sum is LCTI. An interesting and still rather general splitting is depicted in Fig. 9. Here the signal y is first given to the

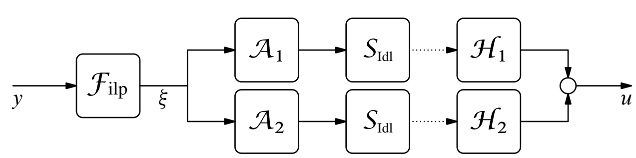


Fig. 9. Alternative implementation of a two-channel HSP

ideal lowpass filter $\mathcal{F}_{\mathrm{ilp}}$ with the cut-off frequency $2\omega_\mathrm{N}$. With this choice, we do not need to prefilter measurements if they are already $2\omega_\mathrm{N}$-bandlimited. The outcome is then fed to two different LCTI filters

$\mathcal{A}_1$ and $\mathcal{A}_2$ followed by ideal samplers and then two holds. This corresponds to the case that

$$\bar{A}(\theta) = \begin{bmatrix} A_1(\mathrm{j}\omega_0) & A_1(\mathrm{j}\omega_{-1}) \\ A_2(\mathrm{j}\omega_0) & A_2(\mathrm{j}\omega_{-1}) \end{bmatrix} \tag{31}$$

if $\theta > 0$ (see Appendix for a derivation).

Example 8.1 (Samples with derivatives): If $\mathcal{A}_1$ is the identity and $\mathcal{A}_2$ the differentiator we get a mixing matrix

$$\bar{A}(\theta) = \begin{bmatrix} 1 & 1 \\ \mathrm{j}\omega_0 & \mathrm{j}\omega_{-1} \end{bmatrix}.$$

This matrix has constant nonzero determinant $-\mathrm{j}2\pi/h$. The hold functions (29) now become (for $\theta \in [0, \pi]$)

$$\begin{aligned} &\begin{bmatrix} \phi_1(\mathrm{e}^{\mathrm{j}\theta}) & \phi_2(\mathrm{e}^{\mathrm{j}\theta}) \end{bmatrix} \\ &\quad = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\omega_0\tau} & \mathrm{e}^{\mathrm{j}\omega_{-1}\tau} \end{bmatrix} A^{-1}(\theta) \\ &\quad = \begin{bmatrix} \dfrac{\mathrm{e}^{\mathrm{j}\omega_0\tau}\mathrm{j}\omega_{-1} - \mathrm{e}^{\mathrm{j}\omega_{-1}\tau}\mathrm{j}\omega_0}{-\mathrm{j}2\pi/h} & \dfrac{-\mathrm{e}^{\mathrm{j}\omega_0\tau} + \mathrm{e}^{\mathrm{j}\omega_{-1}\tau}}{-\mathrm{j}2\pi/h} \end{bmatrix}. \end{aligned}$$

The inverse Fourier transformation subsequently yields (see [1, Example 4.2] for ϕ_1) the two hold functions

$$\phi_1(t) = \mathrm{sinc}_h^2(t), \qquad \phi_2(t) = t\,\mathrm{sinc}_h^2(t)$$

and we get the well known reconstruction formula

$$f(t) = \sum_{k\in\mathbb{Z}} \phi_1(t - kh) f(kh) + \phi_2(t - kh) f'(kh)$$

provided $f(t)$ is $2\omega_{\mathrm{N}}$-bandlimited. ∇

For two channels the mixing matrix $\bar{A}(\theta)$ is 2×2. It is straightforward to extend the ideas to more than two channels. For instance when M derivative samples, $y^{(i)}(kh)$ for $i = 0, \ldots, M-1$, are available etcetera. The formulae are unwieldy though.

For recurring non-uniform sampling the method recovers Yen's original work [8]. In this case the formulae are manageable for any M:

Example 8.2 (Recurring non-uniform sampling): If $\mathcal{A}_1$ is the identity and $\mathcal{A}_2$ the T-delay operator $A_2(\mathrm{j}\omega) = \mathrm{e}^{-\mathrm{j}T\omega}$ then the mixing matrix (31) becomes the Vandermonde matrix

$$\bar{A}(\theta) = \begin{bmatrix} 1 & 1 \\ \mathrm{e}^{-\mathrm{j}T\omega_0} & \mathrm{e}^{-\mathrm{j}T\omega_{-1}} \end{bmatrix} \quad \text{for } \theta \in [0, \pi].$$

It is invertible iff the delay T is not a multiple of the sampling period h, in which case

$$\bar{A}^{-1}(\theta) = \frac{1}{e^{-jT\omega_{-1}} - e^{-jT\omega_0}} \begin{bmatrix} e^{-jT\omega_{-1}} & -1 \\ -e^{-jT\omega_0} & 1 \end{bmatrix}.$$

Direct inverse Fourier transformation of (29) now yields the optimal hold functions

$$\phi_1(t) = \operatorname{sinc}_h(t)\frac{\sin(\omega_N(t+T))}{\sin(\omega_N T)}, \quad \phi_2(t) = \phi_1(-t-T)$$

see Fig. 10. This $\phi_1(t)$ is the unique[3] $2\omega_N$-bandlimited signal that is 1 at $t = 0$ and is 0 at both all other sampling instances, kh, $k \neq 0$, and delayed sampling instances $kh - T$, $k \in \mathbb{Z}$. By symmetry $\phi_2(t) = \phi_1(-t-T)$ has comparable interpolation properties, see Fig. 10.

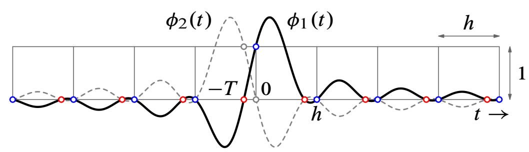


Fig. 10. Optimal hold functions for $M = 2$ (Example 8.2)

If instead of 2 we have M samples every $[hk, hk+h)$ at $t = hk + T_1$, $t = hk + T_2, \ldots, t = hk + T_M$, then the M optimal hold functions $\phi_1, \ldots, \phi_M$ are [8]

$$\phi_n(t) = \operatorname{sinc}_h(t + T_n) \prod_{k \neq n} \frac{\sin(\omega_N(t + T_k))}{\sin(\omega_N(-T_n + T_k))}.$$

Indeed, they satisfy the interpolation conditions and are $M\omega_N$-bandlimited by the fact that they are M products of ω_N-bandlimited signals, and thus they are the solutions we seek (provided $\mathcal{G}_v$ is $M\omega_N$-band dominant). ∇

Besides [8], the results in this section bears close resemblance with the generalized sampling theorems of [9], with the difference that [9] assumes from the outset that the signal is sufficiently bandlimited. Paper [30] treats the same problem but then aims at consistent rather than norm-optimal HSPs. This, however, is closely related to norm-optimality because consistency is an interpolation condition and in Footnote 3 we saw that norm-optimality under certain assumptions is equivalent to an interpolation condition.

[3]Since $\mathcal{G}_{e,opt} = 0$ for any $\mathcal{G}_v$ that is $2\omega_N$-bandlimited, we have that $\mathcal{H}S = I$ when restricted to $2\omega_N$-bandlimited signals. Suppose η and ζ are two $2\omega_N$-bandlimited signals with the same samples, then $\zeta = \mathcal{H}S\zeta = \mathcal{H}S\eta = \eta$ i.e., then they are the same.

IX. Downsampling

Consider again the case $\mathcal{G}_y = \mathcal{G}_v$, but now assume that the generator of v is itself an HSP,

$$\mathcal{G}_v = \mathcal{H}_{h'} \mathcal{S}_{h'} \tag{32}$$

with a sampling period h' different from h. To maintain h-periodicity we assume that this sampling period is an integer fraction of h,

$$h' = h/m, \qquad \text{for some } m \in \mathbb{N}.$$

The problem is to find a single channel $\mathcal{F}_{\text{HSP}}$ with sampling period h that minimizes the L^2 or L^∞ norm of the error system $\mathcal{G}_\text{e}$. In the present context this is an example of downsampling by a factor m. System (32) has kernel $g(t,s) = \sum_{k\in\mathbb{Z}} \phi_{h'}(t-kh')\psi_{h'}(kh'-s)$ and it can be seen as the superposition of m advanced-delayed h-periodic systems, $g(t,s) = \sum_{n=0}^{m-1}\sum_{k\in\mathbb{Z}} \phi_{h'}(t-nh'-kh)\psi_{h'}(nh'+kh-s)$. It has frequency response kernel

$$\breve{g}(\text{e}^{\text{j}\theta};\tau,\sigma) = \sum_{n=0}^{m-1} \breve{\phi}_{h'}(\text{e}^{\text{j}\theta};\tau-nh')\breve{\psi}_{h'}(\text{e}^{\text{j}\theta};-(\sigma-nh')).$$

Using the Key Lifting Formula for the sampling function ψ shows that

$$\breve{g}(\text{e}^{\text{j}\theta};\tau,\sigma) = \sum_{k\in\mathbb{Z}}\left(\sum_{n=0}^{m-1} \phi_{h'}(\text{e}^{\text{j}\theta};\tau-nh')\text{e}^{-n\text{j}\omega_k h'}\right) \times \Psi(\text{j}\omega_k)\frac{1}{h}\text{e}^{-\text{j}\omega_k\sigma}.$$

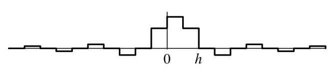
Since $\mathcal{G}_v$ is not LCTI it is not immediate what the fixed-frequency SVD (Proposition 6.1) is, but for certain examples of $\mathcal{G}_v$ it can be done:

Example 9.1 (Downsampling by factor 2): Let $m=2$ and $\mathcal{G}_v = \mathcal{H}_{\text{ZOH}}\mathcal{S}_{\text{Idl}}\mathcal{G}_{\text{ilp}}$, where the ideal sampler $\mathcal{S}_{\text{Idl}}$ and the zero-order hold $\mathcal{H}_{\text{ZOH}}$ have the sampling period $h/2$ and $\mathcal{G}_{\text{ilp}}$ is the ideal lowpass filter with bandwidth $2\omega_\text{N}$. By the bandlimitness of the prefilter we have, for $\theta \in [0,\omega_\text{N}]$,

$$\begin{aligned}\breve{g}(\text{e}^{\text{j}\theta};\tau,\sigma) &= \sum_{k=1,2}\left(\sum_{n=0,1} \phi_{h'}(\text{e}^{\text{j}\theta};\tau-nh')\text{e}^{-n\text{j}\omega_k h'}\right) \times \Psi(\text{j}\omega_k)\frac{1}{h}\text{e}^{-\text{j}\omega_k\sigma} \\ &= \underbrace{\begin{bmatrix} \mathbb{1}_{[0,h/2]}(\tau) & \mathbb{1}_{[h/2,h]}(\tau)\end{bmatrix}\begin{bmatrix} 1 & 1 \\ \text{e}^{\text{j}\theta/2} & -\text{e}^{\text{j}\theta/2}\end{bmatrix}}_{V(\theta)} \\ &\quad \times \begin{bmatrix}\Psi(\text{j}\omega_0) & 0 \\ 0 & \Psi(\text{j}\omega_1)\end{bmatrix}\begin{bmatrix}\frac{1}{h}\text{e}^{-\text{j}\omega_0\sigma} \\ \frac{1}{h}\text{e}^{-\text{j}\omega_1\sigma}\end{bmatrix}.\end{aligned} \tag{33}$$

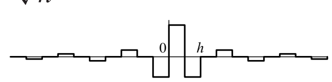
The two shifted hold functions $\mathbb{1}_{[0,h/2]}(\tau)$ and $\mathbb{1}_{[h/2,h]}(\tau)$ have non-overlapping support and therefore are orthogonal (and with the same $\mathbb{L}$-norm of $\sqrt{h/2}$), making the $V(\theta)$ defined above orthogonal at each θ and $V'(\theta)V(\theta) = hI_2$. Equation (33) at each θ is therefore an SVD with singular values $\{h|\Psi(\mathrm{j}\omega_0)|, h|\Psi(\mathrm{j}\omega_1)|\}$. By Corollary 5.3, the optimal HSP should cancel the largest singular value. If Ψ is baseband dominant then according to this corollary $\breve{F}_{\mathrm{HSP}}(\mathrm{e}^{\mathrm{j}\theta}) = \langle \cdot, v_1\rangle_{\mathbb{L}} v_1$ with v_1 the θ dependent first column of $V(\theta)$ normalized to have $\mathbb{L}$-norm 1. That is, its kernel is $f_{\mathrm{HSP}}(t,s) = \phi(t)\phi(s)$ with optimal hold and sampler equal to the inverse Fourier transform of the first column of V (scaled by $\sqrt{h}$ for orthonormality),

$$\begin{aligned}\phi(t) = \psi(t) &= \frac{1}{\sqrt{h}}\mathfrak{F}^{-1}\{V_1\}\\ &= \frac{1}{\sqrt{h}}\mathfrak{F}^{-1}\left\{\mathbb{1}_{[0,h/2]}(\tau) + \mathbb{1}_{[h/2,h]}(\tau)\mathrm{e}^{\mathrm{j}\theta/2}\right\}\\ &= \frac{1}{\sqrt{h}}\Big(\mathbb{1}_{[0,h/2]}(t) + \sum_{k\in\mathbb{Z}} \mathrm{sinc}_1(k+\tfrac{1}{2})\mathbb{1}_{[h/2,h]}(t-kh)\Big)\\ &=\end{aligned}$$



The optimal HSP is $S^*S = \mathcal{H}\mathcal{H}^*$. In the somewhat special case that Ψ is passband dominant in the sense that the second band is dominant, that is, $|\Psi(\mathrm{j}\omega_1)| \geq |\Psi(\mathrm{j}\omega_{k\neq 1})|$, $\forall\theta \in [0,\pi)$, then we should select the second column of V, rendering the optimal hold/sampler equal to

$$\begin{aligned}\phi(t) = \psi(t) &= \frac{1}{\sqrt{h}}\mathfrak{F}^{-1}\{V_2\}\\ &= \frac{1}{\sqrt{h}}\mathfrak{F}^{-1}\left\{\mathbb{1}_{[0,h/2]}(\tau) - \mathbb{1}_{[h/2,h]}(\tau)\mathrm{e}^{\mathrm{j}\theta/2}\right\}\\ &=\end{aligned}$$



The hold function is unique (modulo frequency dependent scaling that could be absorbed into the sampler or discrete filter) but the sampler is not unique in this case because the signal generator is singular. Neither $\mathcal{G}_v$ nor the optimal HSP is LCTI. ∇

X. SR with Noisy Measurements

In this final section we consider the case where the signal y available for sampling is corrupted by colored noise. This very common situation can be modeled as in Fig. 11 where n is the colored noise which is seen as the output of a system $\mathcal{G}_n$ driven by white noise w_n, assumed to be independent of w_v which drives the system $\mathcal{G}_v$ that generates the signal v that we aim to reconstruct. This is a special case of the setup in Fig. 1 for

$$\mathcal{G} = \begin{bmatrix} \mathcal{G}_v & 0 \\ \mathcal{G}_v & \mathcal{G}_n \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} w_v \\ w_n \end{bmatrix}.$$

The signal generators $\mathcal{G}_v$ and $\mathcal{G}_n$ are assumed to be real LCTI systems satisfying

$\mathcal{A}_2$: $|G_v(\mathrm{j}\omega)|^2 + |G_n(\mathrm{j}\omega)|^2 > 0$ for all ω.

This assumption guarantees that the optimization problems are non-singular.

The requirement that $\mathcal{F}_{\text{HSP}}$ is an HSP can be viewed as a structural constraint imposed on the reconstructor (estimator). This suggests that the problem can be addressed via the solution of the unconstrained problems, where the L^2 or L^∞ norms of the error system $\mathcal{G}_\mathrm{e}$ are minimized by an analog filter $\mathcal{F}$ (not necessarily an HSP). We thus start with the latter problem, following the ideas of [14].

First, recall that the L^2-norm of $\mathcal{G}_\mathrm{e}$, $\|\mathcal{G}_\mathrm{e}\|_2$, is the square root of the (operator) trace of $\mathcal{G}_\mathrm{e}\mathcal{G}_\mathrm{e}^*$ and the L^∞-norm of the error system $\|\mathcal{G}_\mathrm{e}\|_\infty \le \gamma$ iff $\mathcal{G}_\mathrm{e}\mathcal{G}_\mathrm{e}^* \le \gamma^2 I$, [14]. This is to say that the system $\mathcal{G}_\mathrm{e}\mathcal{G}_\mathrm{e}^*$ plays a central role in both optimization problems. Now,

$$\begin{aligned} \mathcal{G}_\mathrm{e}\mathcal{G}_\mathrm{e}^* &= (I - \mathcal{F})\mathcal{G}_v\mathcal{G}_v^*(I - \mathcal{F})^* + \mathcal{F}\mathcal{G}_n\mathcal{G}_n^*\mathcal{F}^* \\ &= \mathcal{G}_v\mathcal{G}_v^* - \mathcal{F}\mathcal{G}_v\mathcal{G}_v^* - \mathcal{G}_v\mathcal{G}_v^*\mathcal{F}^* + \mathcal{F}(\mathcal{G}_v\mathcal{G}_v^* + \mathcal{G}_n\mathcal{G}_n^*)\mathcal{F}^* \\ &= \mathcal{Q} + (\mathcal{G}_v\mathcal{G}_v^*\mathcal{R}^{-1} - \mathcal{F})\mathcal{R}(\mathcal{G}_v\mathcal{G}_v^*\mathcal{R}^{-1} - \mathcal{F})^*, \end{aligned} \tag{34}$$

where $\mathcal{R} := \mathcal{G}_v\mathcal{G}_v^* + \mathcal{G}_n\mathcal{G}_n^*$ is invertible by $\mathcal{A}_2$ and, in fact, $\mathcal{G}_v\mathcal{G}_v^*\mathcal{R}^{-1}$ is then well defined and stable. Also,

$$\mathcal{Q} := \mathcal{G}_v(I - \mathcal{G}_v^*\mathcal{R}^{-1}\mathcal{G}_v)\mathcal{G}_v^* = \mathcal{G}_v\mathcal{G}_v^*\mathcal{R}^{-1}\mathcal{G}_n\mathcal{G}_n^*.$$

As no causality constraints are imposed, it is readily seen [14] that the optimal solution in both L^2 and L^∞ cases is

$$\mathcal{F} = \mathcal{F}_{\text{wiener}} := \mathcal{G}_v\mathcal{G}_v^*\mathcal{R}^{-1} = \mathcal{G}_v\mathcal{G}_v^*(\mathcal{G}_v\mathcal{G}_v^* + \mathcal{G}_n\mathcal{G}_n^*)^{-1}$$

(in the L^∞ case it might be non-unique). This is the classical LCTI Wiener filter. It is not necessarily an HSP and in fact it generally is not an HSP, and as such $\mathcal{F}_{\text{wiener}}$ is not the solution we seek.

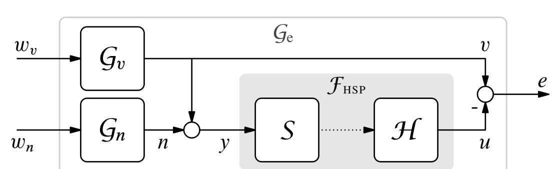


Fig. 11. Setup for SR with noisy measurements (Section X)

Important is that (34) can be used to reduce the original signal reconstruction problem to a simpler problem, similar to the noise-free problem studied in Section VII. This reduction, however, is different in the L^2 and L^∞ cases.

A. L^2 Optimization

Because of the linearity of the operator trace, (34) gives that

$$\|\mathcal{G}_\mathrm{e}\|_2^2 = \|\mathcal{Q}^{1/2}\|_2^2 + \|(\mathcal{F}_\mathrm{wiener} - \mathcal{F})\mathcal{R}^{1/2}\|_2^2. \tag{35}$$

Hence, the L^2 signal reconstruction problem is equivalent to the problem of

$$\min_{\mathcal{F}_\mathrm{HSP}} \|\underbrace{\mathcal{F}_\mathrm{wiener}\mathcal{R}^{1/2}}_{\mathcal{G}_2} - \underbrace{\mathcal{F}_\mathrm{HSP}\mathcal{R}^{1/2}}_{\mathcal{F}_2}\|_2, \tag{36}$$

which is a one-block problem. In the noise-free setting, the systems $\mathcal{R}^{1/2}$ and $\mathcal{F}_\mathrm{wiener}$ should be replaced with $\mathcal{G}$ and I, respectively. The presence of $\mathcal{R}^{1/2}$ and $\mathcal{F}_\mathrm{wiener}$ does not lead to any conceptual difference though. By the invertibility of $\mathcal{R}^{1/2}$, the series interconnection $\mathcal{F}_2$ is a rank-1 HSP iff $\mathcal{F}_\mathrm{HSP}$ is. Now, the optimal rank-1 approximation $\mathcal{F}_2$ of an LCTI system $\mathcal{G}_2$ is itself LCTI and therefore the optimal rank-1 $\mathcal{F}_\mathrm{HSP} = \mathcal{F}_2\mathcal{R}^{-1/2}$ is LCTI as well. To circumvent exotic HSPs we again assume baseband dominance:

$\boldsymbol{\mathcal{A}}_\mathbf{3}$: $\mathcal{G}_2 = \mathcal{G}_v\mathcal{G}_v^*(\mathcal{G}_v\mathcal{G}_v^* + \mathcal{G}_n\mathcal{G}_n^*)^{-1/2}$ is baseband dominant.

The singular values of $\breve{G}_2(\mathrm{e}^{\mathrm{j}\theta})$ at each θ can be expressed as

$$\sigma_k = \frac{|G_v(\mathrm{j}\omega_k)|^2}{\sqrt{|G_v(\mathrm{j}\omega_k)|^2 + |G_n(\mathrm{j}\omega_k)|^2}} = |G_v(\mathrm{j}\omega_k)|\sqrt{\frac{\rho(\omega_k)}{1+\rho(\omega_k)}},$$

where

$$\rho(\omega) := \frac{|G_v(\mathrm{j}\omega)|^2}{|G_n(\mathrm{j}\omega)|^2} \tag{37}$$

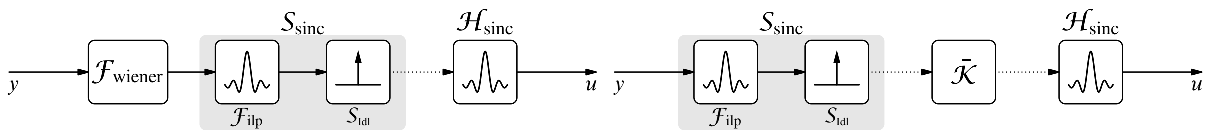
can be interpreted as the signal-to-noise ratio spectrum.

Given $\boldsymbol{\mathcal{A}}_\mathbf{3}$, the $F_2(\mathrm{j}\omega)$ that minimizes (36) equals $G_2(\mathrm{j}\omega)$ in the baseband $\omega \in [-\omega_\mathrm{N}, \omega_\mathrm{N}]$ and is zero elsewhere. The optimal $\mathcal{F}_\mathrm{HSP} = \mathcal{F}_2\mathcal{R}^{-1/2}$ therefore is the LCTI system that is zero outside the baseband, and in the baseband equals $F_\mathrm{HSP}(\mathrm{j}\omega) = G_2(\mathrm{j}\omega)R(\mathrm{j}\omega)^{-1/2} = F_\mathrm{wiener}(\mathrm{j}\omega)$. In the baseband the optimal $\mathcal{F}_\mathrm{HSP}$ acts as the classic Wiener filter making the error $G_\mathrm{e}(\mathrm{j}\omega)G_\mathrm{e}(\mathrm{j}\omega)^*$ equal to $Q(\mathrm{e}^{\mathrm{j}\theta})$, and outside the baseband it does nothing. Therefore:

Theorem 10.1: Let $\mathcal{G}_v$ and $\mathcal{G}_n$ be real stable LCTI systems and suppose assumptions $\boldsymbol{\mathcal{A}}_\mathbf{2\text{-}3}$ hold. Then the HSP depicted in Fig. 12(a) minimizes the L^2 norm of $\mathcal{G}_\mathrm{e}$ and attains the optimal performance

$$\|\mathcal{G}_\mathrm{e}\|_2^2 = \frac{1}{\pi}\int_0^{\omega_\mathrm{N}} \frac{|G_v(\mathrm{j}\omega)|^2}{1+\rho(\omega)}\mathrm{d}\omega + \frac{1}{\pi}\int_{\omega_\mathrm{N}}^\infty |G_v(\mathrm{j}\omega)|^2\mathrm{d}\omega,$$

where $F_{\text{wiener}}(j\omega) = \frac{\rho(\omega)}{1+\rho(\omega)}$ and $\rho(\omega)$ is defined by (37). All components are stable and the overall HSP is LCTI. ▽



(a) Configuration with analog Wiener filter $\mathcal{F}_{\text{wiener}}$ (b) Configuration with discrete filter, $\bar{K}(e^{j\theta}) = F_{\text{wiener}}(\theta/h)$

Fig. 12. The optimal HSP for SR with noisy measurements

The optimal reconstructor is very similar to the WKS-block with the sole difference that the analog Wiener filter preprocesses the measurement. The frequency response of $\mathcal{F}_{\text{wiener}}$ is real valued for all frequencies, so it is noncausal (unless it is static, which happens if $\mathcal{G}_n$ is scalar multiple of $\mathcal{G}_v$). An alternative form of the optimal HSP is presented in Fig. 12(b), in which the Wiener filter is, in a sense, converted to the discrete filter $\bar{\mathcal{K}}$ with the frequency response $\bar{K}(e^{j\theta}) = F_{\text{wiener}}(\theta/h)$. This filter is also generically noncausal. Moreover, it is normally not a rational function of $e^{j\theta}$ even if the analog Wiener filter is rational. Hence, unless $\bar{\mathcal{K}}$ is static, it is infinite dimensional.

B. *L^∞ Optimization*

The situation here is more complicated than in the L^2 case. Clearly from (34) we have that $\mathcal{G}_e\mathcal{G}_e^* \le \gamma^2 I$ iff

$$(\mathcal{F}_{\text{wiener}} - \mathcal{F})\mathcal{R}(\mathcal{F}_{\text{wiener}} - \mathcal{F})^* \le \gamma^2 I - \mathcal{Q}. \tag{38}$$

This requires that $\gamma \ge \gamma_{\text{wiener}}$, where

$$\gamma_{\text{wiener}} := \sqrt{\|\mathcal{Q}\|_\infty}$$

is the optimal L^∞ performance achievable with $\mathcal{F} = \mathcal{F}_{\text{wiener}}$.

If $\gamma > \gamma_{\text{wiener}}$, the system $I - \gamma^{-2}\mathcal{Q}$ is stably invertible and then there is an HSP guaranteeing that $\|\mathcal{G}_e\|_\infty \le \gamma$ iff

$$\|(I - \gamma^{-2}\mathcal{Q})^{-1/2}(\mathcal{F}_{\text{wiener}} - \mathcal{F}_{\text{HSP}})\mathcal{R}^{1/2}\|_\infty \le \gamma \tag{39}$$

for some $\mathcal{F}_{\text{HSP}}$. The system in (39) is of the one-block type

$$\underbrace{(I - \gamma^{-2}\mathcal{Q})^{-1/2}\mathcal{F}_{\text{wiener}}\mathcal{R}^{1/2}}_{\mathcal{G}_\infty} - \underbrace{(I - \gamma^{-2}\mathcal{Q})^{-1/2}\mathcal{F}_{\text{HSP}}\mathcal{R}^{1/2}}_{\mathcal{F}_\infty}$$

and, similarly to the L^2 case, $\mathcal{F}_\infty$ is a rank-1 HSP iff $\mathcal{F}_{\text{HSP}}$ is and by the fact that optimal rank-1 $\mathcal{F}_\infty$ can be taken LCTI also $\mathcal{F}_{\text{HSP}} = (I - \gamma^2 \mathcal{Q})^{1/2} \mathcal{F}_\infty \mathcal{R}^{-1/2}$ can be taken LCTI. Now if we were to cancel the singular value $|G_\infty(e^{j\omega_0})|$ in the baseband then this would result in $F_{\text{HSP}}(j\omega) = F_{\text{wiener}}(j\omega)$ in $\omega \in [-\omega_N, \omega_N]$ and zero elsewhere. This is exactly the same HSP as in the L^2 case. This choice of $\mathcal{F}_{\text{HSP}}$ achieves $\|\mathcal{G}_e\|_\infty \le \gamma$ if and only if $\sup_{\omega > \omega} |G_\infty(j\omega)| \le \gamma$. At first sight, this condition appears hard to check. There, however, holds:

Lemma 10.2: Let $\gamma > \gamma_{\text{wiener}}$. Then at each ω we have $|G_\infty(j\omega)| \le \gamma \iff |G_v(j\omega)| \le \gamma$.

Proof: $|G_\infty(j\omega)| \le \gamma$ iff (38) holds for $F(j\omega) = 0$ at the given frequency, which in turn is equivalent to $|G_e(j\omega)| \le \gamma$, but $G_e(j\omega) = G_v(j\omega)$ for $F(j\omega) = 0$. ■

This property allows to bypass baseband dominance of $\mathcal{G}_\infty$ (which is rather involved as $\mathcal{G}_\infty$ depends on γ). Sufficient is to assume baseband dominance of $\mathcal{G}_v$. Thus, we have:

Theorem 10.3: Suppose assumptions $\mathcal{A}_{1,2}$ are satisfied. Then the optimal HSP is the same as that of Theorem 10.1 and

$$\|\mathcal{G}_e\|_\infty = \max\left(\sup_{\omega \in [0, \omega_N]} \frac{|G_v(j\omega)|}{\sqrt{1 + \rho(\omega)}}, \sup_{\omega \in (\omega_N, \infty)} |G_v(j\omega)| \right)$$

is the optimal L^∞ performance level.

Proof: Let γ_∞ be the minimal achievable norm of $\|\mathcal{G}_e\|_\infty$ by rank-1 $\mathcal{F}_{\text{HSP}}$. Assume first that $\gamma_\infty > \gamma_{\text{wiener}}$. Then $|G_\infty(j\omega_k)| > \gamma_\infty$ for at most one of the aliased frequencies ω_k, which by Lemma 10.2 is equivalent to $|G_v(j\omega_k)| > \gamma_\infty$ (for the same one k). By the baseband dominance of $\mathcal{G}_v$, this must be $k = 0$. I.e., the baseband has to be removed, leaving $|G_e(j\omega)| = Q^{1/2}(j\omega)$ in the baseband and $G_e(j\omega) = G_v(j\omega)$ elsewhere. The formula for γ_∞ follows on noting that $Q(j\omega) = |G_v(j\omega)|^2/(1 + \rho(\omega))$.

If $\gamma_\infty = \gamma_{\text{wiener}}$ then for any $\gamma > \gamma_{\text{wiener}} = \gamma_\infty$ by the above argument the given $\mathcal{F}_{\text{HSP}}$ achieves $\|\mathcal{G}_e\|_\infty \le \gamma$. I.e., then for any $\gamma > \gamma_{\text{wiener}}$ inequality (38) is satisfied for $\mathcal{F} = \mathcal{F}_{\text{HSP}}$. Since $\mathcal{F}_{\text{HSP}}$ is independent of γ, the inequality (38) then holds for $\gamma = \gamma_\infty$ as well. ■

Both L^2 and L^∞ equivalent one-block problems (36) and (39), respectively, can be interpreted as (weighted) approximations of the analog optimal reconstructor $\mathcal{F}_{\text{wiener}}$ by $\mathcal{F}_{\text{HSP}}$. In other words, the choice of "good" HSPs can be viewed as an attempt to imitate their analog counterparts. This interpretation repeats the main point of [31, Sec. 6] made in the context of the sampled-data feedback control with causal controllers.

Remark 10.1: The optimal performance indices in Theorems 10.1 and 10.3 have two components representing two extreme situations. The first of these components reflects the contribution of the baseband, $[0, \omega_N]$, and is a size of $\mathcal{Q}$ in this frequency range. The frequency response of $\mathcal{Q}$ is actually the spectrum of the estimation error under the optimal analog reconstruction. Thus, the baseband contributes, in a sense, by the optimal analog performance. The second component of the optimal indices reflects the contribution of the high-frequency range, $(\omega_{\mathrm{N}}, \infty)$, and is a size of $\mathcal{G}_v$. Thus, high frequency components contribute by the estimator-free performance. Thus, in $[0, \omega_{\mathrm{N}}]$ the sampled-data reconstructor recovers the analog performance, whereas in $(\omega_{\mathrm{N}}, \infty)$ it does nothing. ∇

Remark 10.2: In the L^2 case, Theorem 10.1 requires that the function $|G_v(\mathrm{j}\omega)|^2 \frac{\rho(\omega)}{1+\rho(\omega)}$ is baseband-dominant. If $|G_v(\mathrm{j}\omega)|$ is baseband-dominant, this requirement is clearly guaranteed if the signal-to-noise ratio $\rho(\omega)$ is a non-increasing function of ω, which is a reasonable assumption in many applications. The dominance requirement might fail if $\frac{\rho(\omega)}{1+\rho(\omega)}$ increases faster than $|G_v(\mathrm{j}\omega)|^2$ decreases. This, in turn, is possible if the signal-to-noise ratio increases considerably faster then the spectrum of v decays. Spectral properties of $\mathcal{G}_n$ do not affect the baseband-dominance in the L^∞ case. ∇

XI. Concluding Remarks

The main message of this part is that the system-theoretic approach—the use of systems as signal generators to account for available information and system norms as performance measures—facilitates a unified treatment of a wide spectrum of sampling and reconstruction problems. We have considered the design of L^2 and L^∞ optimal acquisition and / or interpolation devices when no causality constraints are imposed on them. Remarkably, this single approach recovers many known HSPs derived hitherto by different methods. For example, when sampling circuits are fixed (Type III problems), certain choices of signal generators produce conventional cardinal polynomial or exponential splines as the optimal reconstructors. Another example is the recovery of the classical Sampling Theorem and its modifications (samples with derivatives, recurring non-uniform sampling) when both sampling and reconstruction devices are design parameters (Type IV problems) under different assumptions about the sampling process. We believe that the capability to reproduce known results as special cases of a general framework is an important property, offering an additional insight into both existing and the proposed approaches. The presented proofs of the continuous-time invariance of certain optimal HSPs and the necessity of a bandllimited assumption in multi-channel sampling attest to it. At the same time, we have shown that the approach can produce new solutions and interpretations, like the interplay between L^2 and L^∞ norms, leading to limitations on error free reconstruction, and optimal downsampling and a version of the Sampling Theorem for reconstructing signals from noisy measurements. Many more extensions can be added to this list. One of them—imposing causality constraints on the design of L^2-optimal reconstructors—is reported in [18, Part III].

APPENDIX

Proof of Proposition 2.2: This is a known result, often called the Parrott lower bound, see [32], [33]. The idea is to transform the operator whose L^∞ norm we want to minimize into one of the form

$$\begin{bmatrix} \mathcal{R}_{11} - S & \mathcal{R}_{12} \\ \mathcal{R}_{21} & \mathcal{R}_{22} \end{bmatrix}$$

with $\mathcal{R}_{ij}$ fixed operators and S our free parameter (sampler). Parrott [32] showed that then its L^∞ norm is bounded from below by

$$\max\left(\left\| \begin{bmatrix} \mathcal{R}_{12} \\ \mathcal{R}_{22} \end{bmatrix} \right\|_\infty, \left\| \begin{bmatrix} \mathcal{R}_{21} & \mathcal{R}_{22} \end{bmatrix} \right\|_\infty \right)$$

and that equality can be achieved [33]. To simplify the exposition, we assume that $\mathcal{H}^*\mathcal{H} = I$ and $\mathcal{G}_y\mathcal{G}_y^* = I$. Then $\begin{bmatrix} \mathcal{G}_y^* & I - \mathcal{G}_y^*\mathcal{G}_y \end{bmatrix}$ is co-inner, meaning that

$$\begin{bmatrix} \mathcal{G}_y^* & I - \mathcal{G}_y^*\mathcal{G}_y \end{bmatrix} \begin{bmatrix} \mathcal{G}_y \\ I - \mathcal{G}_y^*\mathcal{G}_y \end{bmatrix} = I.$$

Therefore $\mathcal{G}_v - \mathcal{H}S\mathcal{G}_y$ and

$$(\mathcal{G}_v - \mathcal{H}S\mathcal{G}_y)\begin{bmatrix} \mathcal{G}_y^* & I - \mathcal{G}_y^*\mathcal{G}_y \end{bmatrix} = \begin{bmatrix} \mathcal{G}_v\mathcal{G}_y^* - \mathcal{H}S & \mathcal{G}_v(I - \mathcal{G}_y^*\mathcal{G}_y) \end{bmatrix} \quad (40)$$

have the same L^∞ norm. Notice that the second block here does not depend on S. Similarly $\begin{bmatrix} \mathcal{H}^* \\ I - \mathcal{H}\mathcal{H}^* \end{bmatrix}$ is inner and therefore (40) in turn has the same L^∞ norm as

$$\begin{aligned} &\begin{bmatrix} \mathcal{H}^* \\ I - \mathcal{H}\mathcal{H}^* \end{bmatrix} \begin{bmatrix} \mathcal{G}_v\mathcal{G}_y^* - \mathcal{H}S & \mathcal{G}_v(I - \mathcal{G}_y^*\mathcal{G}_y) \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{H}^*\mathcal{G}_v\mathcal{G}_y^* - S & \mathcal{H}^*\mathcal{G}_v(I - \mathcal{G}_y^*\mathcal{G}_y) \\ (I - \mathcal{H}\mathcal{H}^*)\mathcal{G}_v\mathcal{G}_y^* & (I - \mathcal{H}\mathcal{H}^*)\mathcal{G}_v(I - \mathcal{G}_y^*\mathcal{G}_y) \end{bmatrix} \\ &=: \begin{bmatrix} \mathcal{R}_{11} - S & \mathcal{R}_{12} \\ \mathcal{R}_{21} & \mathcal{R}_{22} \end{bmatrix}. \end{aligned}$$

Now, only the upper left block depends on S and it can be assigned any operator that we like and therefore Parrott's theorem applies. It is readily seen that

$$\left\| \begin{bmatrix} \mathcal{R}_{12} \\ \mathcal{R}_{22} \end{bmatrix} \right\|_\infty = \|\mathcal{G}_v(I - \mathcal{G}_y^*\mathcal{G}_y)\|_\infty$$

and

$$\left\| \begin{bmatrix} \mathcal{R}_{21} & \mathcal{R}_{22} \end{bmatrix} \right\|_\infty = \|(I - \mathcal{H}\mathcal{H}^*)\mathcal{G}_v\|_\infty.$$

The formula for the optimal S is very involved [33]. Yet if $\mathcal{G}_y$ is stably invertible, then (3) achieves the lower bound (5). ■

Proof of Theorem 3.3: For $G_v(s) = 1/s^n$ the Fourier transform (11) becomes

$$\Phi_{\text{opt}}(\mathrm{j}\omega) = \frac{1/\omega^{2n}}{\frac{1}{h}\sum_{k\in\mathbb{Z}} 1/(\omega + 2k\omega_{\mathrm{N}})^{2n}}.$$

Since $\mathrm{e}^{\mathrm{j}2k\omega_{\mathrm{N}}h} = 1$ this Fourier transform equals

$$\Phi_{\text{opt}}(\mathrm{j}\omega) = \frac{W(\mathrm{j}\omega)^{2n}}{\frac{1}{h}\sum_{k\in\mathbb{Z}}(W(\mathrm{j}(\omega + 2k\omega_{\mathrm{N}}))^{2n}} \tag{41}$$

for $W(\mathrm{j}\omega) := (1 - \mathrm{e}^{-\mathrm{j}\omega h})/(\mathrm{j}\omega)$. Now, W is the Fourier transform of the zero degree B-spline (not centered around zero) and so W^{2n} corresponds to the degree $2n-1$ B-spline. The numerator in (41) is the result of passing W^{2n} through a stable discrete filter that makes $\phi(kh) = \bar{\delta}[k]$, see [17, §V.B]. So $\phi(t)$ is the cardinal polynomial spline of degree $2n-1$. ■

Proof of Equation (17)*:* According to [1, Eqn. (7)], the kernel $g(t,s)$ of the continuous-time mapping $u = \mathcal{H}Sy$ is $g(t,s) = \sum_{i\in\mathbb{Z}} \phi(t - ih)\psi(ih - s)$. Therefore the kernel $\breve{g}(z;\tau,\sigma)$ of the transfer function is

$$\begin{aligned} \breve{g}(z;\tau,\sigma) &= \sum_{k\in\mathbb{Z}}\sum_{i\in\mathbb{Z}} \phi(\tau + kh - ih)\psi(ih - \sigma)z^{-k} \\ &= \sum_{k\in\mathbb{Z}}\sum_{i\in\mathbb{Z}} \phi(\tau + (k-i)h)z^{-(k-i)}\psi(ih - \sigma)z^{-i} \\ &= \breve{\phi}(z;\tau)\breve{\psi}(z;-\sigma). \end{aligned}$$

This completes the proof. ■

Proof of Theorem 4.1 (Rank Theorem): The if part is trivial. Now the only-if part. If $g \in L^2(\mathbb{R})$, then by Parseval we have that $\int_{-\pi}^{\pi} \|\breve{G}(\mathrm{e}^{\mathrm{j}\theta})\|_{\mathrm{HS}}^2 \mathrm{d}\theta < \infty$. Hence $\|\breve{G}(\mathrm{e}^{\mathrm{j}\theta})\|_{\mathrm{HS}} < \infty$ for almost all θ (for all θ except possibly on a set of zero measure). By the definition of the Hilbert-Schmidt norm then,

$$\int_0^h \int_0^h |\breve{g}(\mathrm{e}^{\mathrm{j}\theta};\tau,\sigma)|^2 \mathrm{d}\tau\mathrm{d}\sigma < \infty \tag{42}$$

for almost all $\theta \in [-\pi,\pi]$. For any of those θ the mapping $\int_0^h \int_0^h \breve{g}(\mathrm{e}^{\mathrm{j}\theta};\tau,\sigma)\breve{u}(\sigma)\mathrm{d}\tau\mathrm{d}\sigma$ is readily seen to be a bounded mapping from $\mathbb{L}$ to $\mathbb{L}$ and therefore is a compact operator and so has an SVD with countably many singular values (at most r in fact) [26, A.3.24 and A.4.23], that is, has a representation of the form $\sum_{k=1}^{r} \alpha_k(\tau)\langle \breve{u}, \beta_k \rangle$ where the inner product is that of $\mathbb{L}$ (all α_ks and β_ks still depend on θ). The kernel of this mapping hence is

$$\breve{g}(\mathrm{e}^{\mathrm{j}\theta}; \tau, \sigma) = \breve{\phi}(\mathrm{e}^{\mathrm{j}\theta}; \tau)\breve{\psi}(\mathrm{e}^{\mathrm{j}\theta}; \sigma)$$
$$:= \begin{bmatrix} \alpha_1(\tau) & \cdots & \alpha_r(\tau) \end{bmatrix} \begin{bmatrix} \beta_1'(\sigma) \\ \vdots \\ \beta_r'(\sigma) \end{bmatrix}.$$

Having finite norm (42) both parts $\breve{\psi}(\mathrm{e}^{\mathrm{j}\theta})$ and $\breve{\phi}(\mathrm{e}^{\mathrm{j}\theta})$ have finite $\mathbb{L}$ norm—which by scaling may be taken to be the same—almost everywhere and then have well defined inverse Fourier transforms in $L^2(\mathbb{R})$. The assumption of continuity on some finite partition is sufficient to guarantee that the factors are Lebesgue integrable. ■

Proof of Proposition 7.4 (Pathological sampling): Define $G_\epsilon(\mathrm{j}\omega)$ as the magnitude of $G_v(\mathrm{j}\omega)$ upto at most $1/\epsilon$, i.e., $G_\epsilon(\mathrm{j}\omega) := \min(1/\epsilon, |G_v(\mathrm{j}\omega)|)$. This G_ϵ is stable and, for every frequency $s = \mathrm{j}\omega$ that is not a pole of $G_v(s)$, it converges pointwise to $G(\mathrm{j}\omega)$ as $\epsilon \to 0$. Therefore in the case of pathological sampling two or more singular values $\sigma_k(\theta)$ of $G_\epsilon(\mathrm{e}^{\mathrm{j}\theta})$ converge to ∞ for some θ. So then (given the rationality of G_v) the error norm for the stabilized generator $\mathcal{G}_{\mathrm{e},\epsilon} := (I - \mathcal{F}_\epsilon)\mathcal{G}_\epsilon$ converges to ∞ as $\epsilon \to 0$. Now, since

$$\|(I - \mathcal{F})\mathcal{G}_v\| \geq \|(I - \mathcal{F})\mathcal{G}_\epsilon\| \geq \|(I - \mathcal{F}_\epsilon)\mathcal{G}_\epsilon\|,$$

we necessarily have that $\|(I - \mathcal{F})\mathcal{G}_v\| = \infty$ for any $\mathcal{F}$ (LCTI or LDTI), which is what we had to prove.

If we have no pathological sampling, then $F_0 := \lim_{\epsilon\to 0} F_\epsilon$ is well defined (frequency-wise, and by rationality). We claim that then $\|(I - \mathcal{F})\mathcal{G}_v\|_2 \geq \|(I - \mathcal{F}_0)\mathcal{G}_v\|_2$, so that $\mathcal{F}_0$ is optimal for $\mathcal{G}_v$. Indeed, if $\|(I - \mathcal{F})\mathcal{G}_v\|_2 < \|(I - \mathcal{F}_0)\mathcal{G}_v\|_2$, then by continuity in ϵ also $\|(I - \mathcal{F})\mathcal{G}_\epsilon\|_2 < \|(I - \mathcal{F}_\epsilon)\mathcal{G}_\epsilon\|_2$ for some small enough ϵ. This contradicts the optimality of $\mathcal{F}_\epsilon$. ■

Mixing matrices (Eqn. (31)*):* We prove that (31) is the mixing matrix for the scheme of Fig. 9. The mapping from y to $\bar{u}_1$ is a sampler $\mathcal{S}_{\mathrm{Idl}}\mathcal{A}_1\mathcal{F}_{\mathrm{idl}}$ where the ideal low pass filter has cut off frequency $2\omega_{\mathrm{N}}$. The sampling function of this sampler is the impulse response of $\mathcal{A}_1\mathcal{F}_{\mathrm{idl}}$. Its frequency response according to the Key Lifting Formula [1, (17b)] is $\frac{1}{h}\sum_{k\in\mathbb{Z}} A_1(\mathrm{j}\omega_k)F_{\mathrm{idl}}(\mathrm{j}\omega_k)\mathrm{e}^{\mathrm{j}\omega_k\tau}$, which for $\theta \in [0, \pi)$ and by the bandlimitness of the ideal low-pass filter becomes

$$\frac{1}{h}[A_1(\mathrm{j}\omega_0)\mathrm{e}^{\mathrm{j}\omega_0\tau} + A_1(\mathrm{j}\omega_{-1})\mathrm{e}^{\mathrm{j}\omega_{-1}\tau}]$$
$$= \begin{bmatrix} A_1(\mathrm{j}\omega_0) & A_1(\mathrm{j}\omega_{-1}) \end{bmatrix} \begin{bmatrix} \mathrm{e}^{\mathrm{j}\omega_0\tau}/h \\ \mathrm{e}^{\mathrm{j}\omega_{-1}\tau}/h \end{bmatrix}.$$

For the lower loop, the A_1 has to be replaced with A_2. ∎

Acknowledgment

The authors would like to thank anonymous reviewers for their valuable comments and suggestions that were helpful in improving the paper.

References

[1] G. Meinsma and L. Mirkin, "Sampling from a system-theoretic viewpoint: Part I—Concepts and tools," *IEEE Trans. Signal Processing*, vol. 58, 2010 (to appear).

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

[3] L. Mirkin and G. Tadmor, "Yet another H^∞ discretization," *IEEE Trans. Automat. Control*, vol. 48, no. 5, pp. 891–894, 2003.

[4] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part I—Theory," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 821–833, 1993.

[5] ——, "B-spline signal processing: Part II—Efficiency design and applications," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 834–848, 1993.

[6] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1425–1438, 2005.

[7] A. J. Jerri, "The Shannon sampling theorem—its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, no. 11, pp. 1565–1596, 1977.

[8] J. L. Yen, "On nonuniform sampling of bandwidth-limited signals," *IRE Trans. Circuit Theory*, vol. CT-3, pp. 251–257, 1956.

[9] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. 24, no. 5, pp. 652–654, 1977.

[10] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2007.

[11] G. Meinsma and L. Mirkin, "System theoretic perspectives of the Sampling Theorem," in *Proc. 17th MTNS Symposium*, Kyoto, Japan, 2006, pp. 730–741.

[12] ——, "Sampled signal reconstruction via H^2 optimization," in *Proc. 2006 IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. III, Toulouse, France, 2006, pp. 365–368.

[13] ——, "Noncausal sampled signal reconstruction from noisy measurements: a system-theoretic approach," in *Proc. 46th IEEE Conf. Decision and Control*, New Orleans, LA, 2007, pp. 444–449.

[14] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories*. Philadelphia: SIAM, 1999.

[15] M. Unser, "On the optimality of ideal filters for pyramid and wavelet signal approximation," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3591–3596, 1993.

[16] A. Aldroubi and M. Unser, "Sampling procedures in function spaces and asymptotic equivalence with Shannon's sampling theory," *Numer. Func. Anal. Opt*, vol. 15, no. 1, pp. 1–21, 1994.

[17] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 2915–2925, 1994.

[18] G. Meinsma and L. Mirkin, "Sampling from a system-theoretic viewpoint," Department of Applied Mathematics, University of Twente, Enschede, Memorandum 1907, November 2009. [Online]. Available: http://eprints.eemcs.utwente.nl/16463/

[19] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Trans. Signal Processing*, vol. 53, no. 6, pp. 2146–2159, 2005.

[20] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[21] T. Blu and M. Unser, "Self-similarity: Part II—Optimal estimation of fractal processes," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1364–1378, 2007.

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

[23] S. Ramani, D. Van De Ville, T. Blu, and M. Unser, "Nonideal sampling and regularization theory," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 2915–2925, 2008.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: The Johns Hopkins University Press, 1996.

[25] I. Gohberg, S. Goldberg, and M. A. Kaashoek, *Classes of Linear Operators*. Basel: Birkhäuser, 1990, vol. I.

[26] R. F. Curtain and H. Zwart, *An Introduction to Infinite-Dimensional Linear Systems Theory*. New York: Springer-Verlag, 1995.

[27] M. K. Tsatsans and G. B. Giannakis, "Principal component filter banks for optimal multiresolution analysis," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 1766–1777, 1995.

[28] G. Meinsma and H. S. Shekhawat, "Truncated norms and limitations on signal reconstruction," in *Proc. 19th MTNS Symposium*, Budapest, Hungary, 2010 (to appear).

[29] R. E. Kalman, Y. C. Ho, and K. S. Narendra, "Controllability of linear dynamical systems," *Contributions to Differential Equations*, vol. 1, no. 2, pp. 189–213, 1963.

[30] M. Unser and J. Zerubia, "A generalized sampling theorem without band-limiting constraints," *IEEE Trans. Circuits Syst. I*, vol. 45, no. 8, pp. 959–969, 1998.

[31] L. Mirkin, H. Rotstein, and Z. J. Palmor, "H^2 and H^∞ design of sampled-data systems using lifting – Part I: General framework and solutions," *SIAM J. Control Optim.*, vol. 38, no. 1, pp. 175–196, 1999.

[32] S. Parrott, "On a quotient norm and the Sz-Nagy-Foias lifting theorem," *J. Funct. Anal.*, vol. 30, pp. 311–328, 1978.

[33] K. Glover, D. J. N. Limebeer, J. C. Doyle, E. M. Kasenally, and M. G. Safonov, "A characterization of all solutions to the four block general distance problem," *SIAM J. Control Optim.*, vol. 92, no. 2, pp. 283–324, 1991.

Gjerrit Meinsma was born in Opeinde, The Netherlands in 1965. He received his Ph.D. degree at the University of Twente in 1993. During the following three years he held a postdoc position at the University of Newcastle, Australia. Since 1997 he is with the Department of Applied Mathematics at the University of Twente, The Netherlands. His research interests are in mathematical systems and control theory, in particular robust control theory.

Leonid Mirkin (M'99) is a native of Frunze, Kirghiz SSR, USSR (now Bishkek, Kyrgyz Republic). He received the electrical engineer degree from Frunze Polytechnic Institute and the PhD (candidate of sciences) degree in automatic control from the Institute of Automation, Academy of Sciences of Kyrgyz Republic, in 1989 and 1992, respectively.

From 1989 to 1993 he was with the Institute of Automation, Academy of Sciences of Kyrgyz Republic. In 1994 he joined the Faculty of Mechanical Engineering at the Technion—Israel Institute of Technology, first as a postdoc and then as a faculty member. His research interests include systems theory, control and estimation of sampled-data systems, dead-time compensation, systems with preview, the application of control to electro-mechanical and optical devices, and robustifying properties of corruption.