U-Invariant Sampling: Extrapolation and Causal Interpolation From Generalized Samples

Tomer Michaeli, Volker Pohl, *Member, IEEE*, and Yonina C. Eldar, *Senior Member, IEEE*

Abstract—**Causal processing of a signal's samples is crucial in on-line applications such as audio rate conversion, compression, tracking and more. This paper addresses the problems of predicting future samples and causally interpolating deterministic signals. We treat a rich variety of sampling mechanisms encountered in practice, namely in which each sampling function is obtained by applying a unitary operator on its predecessor. Examples include pointwise sampling at the output of an antialiasing filter and magnetic resonance imaging (MRI), which correspond respectively to the translation and modulation operators. From an abstract Hilbert-space viewpoint, such sequences of functions were studied extensively in the context of stationary random processes. We thus utilize powerful tools from this discipline, although our problems are deterministic by nature. In particular, we provide necessary and sufficient conditions on the sampling mechanism such that perfect prediction is possible. For cases where perfect prediction is impossible, we derive the predictor minimizing the prediction error. We also derive a causal interpolation method that best approximates the commonly used noncausal solution. Finally, we study when causal processing of the samples of a signal can be performed in a stable manner.**

Index Terms—**Frames, prediction, Riesz bases, sampling, stationary sequences.**

I. INTRODUCTION

SAMPLING and reconstruction of continuous-time signals play a crucial role in signal processing and communications. During the last several decades, sampling theory has enjoyed rapid development [1], [2] due in part to fruitful fertilizations from other disciplines, such as wavelet theory [3], approximation theory [4], general Hilbert space formulations [5], variational approaches [6], estimation theory [7], [8], and optimization [9]. While these recent developments found widespread use in image processing, their deployment in unidimensional applications, such as audio sampling-rate conversion, is

less common. One of the reasons for this seems to be the relatively few studies treating causality constraints within the above frameworks, which becomes crucial in on-line applications.

Causal recovery of signals from their samples was mainly addressed in the context of spline interpolation and uniform pointwise sampling. Several heuristic methods were developed and analyzed in [10]–[12] for modifying the noncausal prefilter, which is at the heart of cubic spline interpolation, into a causal counterpart. In [13], an H^∞ optimization approach was proposed for approximating the noncausal prefilter by a causal one. This method admits a closed form solution for cubic splines, but has to be solved numerically for higher orders. Causal interpolation was also studied in [14] from an approximation-theory perspective. There, the authors characterized the set of interpolation kernels with a given support for which the prefilter is causal, and whose approximation order is maximal.

In modern sampling theory, sampling of $x(t)$ is often described [1] by an evaluation of inner products $c_n = \langle x, s_n \rangle$ with a set of sampling functions $\{s_n(t)\}$, in contrast to the uniform pointwise framework. This more general formulation can be used to model nonideal sampling devices [15]. It also encompasses a large class of signal representations commonly used in signal processing, including the Gabor and wavelet transforms [16]. The set of sampling functions $\{s_n(t)\}$ typically possesses structure that is either predetermined by the application's hardware (e.g. uniform sampling at the output of an antialiasing filter [15]) or deliberately constructed to obtain a meaningful signal description (e.g. the uniform time-frequency tiling in Gabor representations [16]).

In this paper, we study causal sampling problems with a special type of structure, which we term U-invariance. Specifically, we concentrate on scenarios in which the sampling functions are obtained from a single sampling function s as

$$s_0 = s \quad \text{and} \quad s_{n+1} = \mathrm{U} s_n, \quad n \in \mathbb{Z} \tag{1}$$

where U is some unitary operator. Examples include the translation, modulation, and dilatation operators, which are used, respectively, in classical shift-invariant (SI) sampling problems [1], [2], in magnetic resonance imaging (MRI) and Gabor analysis [17], and in wavelet analysis [16], [18], [19]. The special case in which U is a translation operator has been studied extensively in the sampling literature and corresponds to uniform sampling at the output of an antialiasing filter [15]. For this scenario, a wide variety of non-causal recovery techniques have been developed. Noncausal recovery with an arbitrary U was studied in [20].

Here, we address the problems of *sample extrapolation* and *causal recovery* in U-invariant sampling. These problems are of great significance in on-line applications such as audio sampling-rate conversion, in which it is desired to causally process

the samples of a signal in order to produce samples at a different rate. To develop a general framework, we borrow several mathematical tools from the field of random processes. In this discipline, functions of the form (1) are known as *stationary stochastic sequences* [21]–[23]. As we show, the well established theory of prediction and causal estimation of stochastic processes, can be harnessed to solve the problems of extrapolation and causal interpolation from U-invariant samples. This is despite the fact that the latter problems are completely deterministic by nature.

We begin by addressing the problem of extrapolating future samples. Specifically, given the "past" generalized samples $c_n = \langle x, s_n \rangle$, $n = 0, -1, -2, \ldots$ of a signal x in an arbitrary Hilbert space $\mathcal{H}$, we derive conditions on the U-invariant sequence of sampling functions $\{s_n\}$ such that the "future" samples $c_1, c_2, c_3 \ldots$ can be perfectly predicted. When perfect prediction is impossible, we seek a predictor that minimizes the extrapolation error for the worst-case feasible signal x. To demonstrate our method, we provide formulas for the predictor of samples of a signal $x(t)$ that has passed through an RC circuit and a B-spline filter prior to sampling.

We next turn to causal recovery of a signal x from its U-invariant samples. Here, the purpose is to construct an approximation $\widetilde{x}$ of x by using a linear combination of a set of predefined functions $\{w_n\}$, such that the coefficients of the combination are obtained by causal processing of the samples of x. To obtain a simple generalization of SI interpolation, we assume that the reconstruction sequence is also U-invariant. Using the framework of estimation theory, we derive an explicit formula for the causal prefilter whose output is closest to that of the noncausal solution [20] for the worst-case feasible signal x. As a concrete example, we specialize our result to SI reconstruction and develop causal spline interpolation algorithms. Interestingly, one of the approaches examined in previous works [10]–[12] on causal spline interpolation can be obtained as a special case of our algorithm, when the sampling functions $\{s_n(t)\}$ are orthogonal. This happens, for example, if the support of the sampling filter $s(t)$ is smaller than the sampling period. However, for general sampling filters, the worst-case error of previous methods is larger than that of our solution. Thus, as we demonstrate via simulations, our method often yields a lower reconstruction error than previous approaches.

Finally, we study the concept of *causal stability* in U-invariant sampling, which quantifies whether extrapolation and causal recovery can be done in a stable manner. In particular, we show that perfect extrapolation cannot be carried out in a stable manner.

The paper is organized as follows. In Section II we present the framework of U-invariant sampling in detail, and summarize the main results. Section III provides a short review of basic results from sampling theory and the theory of stationary sequences. Section IV investigates the prediction of future samples based on knowledge of past samples. We provide a formula for the transfer function of the optimal prediction filter and derive an explicit condition such that the prediction error is zero. In Section V we consider causal recovery of signals from their U-invariant samples. We demonstrate the theoretical results in the context of causal spline interpolation. Finally, in Section VI we derive conditions on U-invariant sampling to be stable and causally-stable.

II. PROBLEM FORMULATION AND MAIN RESULTS

A. Notations

Let $L^2(\mathbb{R})$ denote the Hilbert space of complex square integrable functions on the real axis $\mathbb{R}$ with the inner product

$$\langle x, y \rangle = \int_{\mathbb{R}} x(t)\overline{y(t)}\mathrm{d}t$$

and norm $\|x\| = \sqrt{\langle x, x \rangle}$. For every $x \in L^2(\mathbb{R})$

$$\widehat{x}(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t, \quad \omega \in \mathbb{R} \tag{2}$$

denotes the *Fourier transform* of x.

We write $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ for the unit circle in the complex plane $\mathbb{C}$. Let μ be a finite nonnegative measure on the interval $[-\pi, \pi)$. Then $L^2(\mu, \mathbb{T})$ denotes the Hilbert space of all functions on $\mathbb{T}$ with the inner product

$$\langle x, y \rangle_\mu = \frac{1}{2\pi} \int_{-\pi}^{\pi} x(\mathrm{e}^{\mathrm{i}\omega})\overline{y(\mathrm{e}^{\mathrm{i}\omega})}\mathrm{d}\mu(\omega)$$

and norm $\|x\|_\mu = \sqrt{\langle x, x \rangle_\mu}$. When μ is the Lebesgue measure, i.e., $\mathrm{d}\mu(\omega) = \mathrm{d}\omega$, we will simply write $L^2(\mathbb{T})$.

We denote by $\mathcal{P}^+$ the set of all complex polynomials, i.e., the set of all functions of the form

$$p(z) = \sum_{k=0}^{N} \alpha_k z^k, \quad z \in \mathbb{C}$$

for some arbitrary $N \in \mathbb{N}$, and by $\mathcal{P}_0^+ := \{p \in \mathcal{P}^+ : p(0) = 0\}$ the set of all such polynomials with $\alpha_0 = 0$.

Let $\mathcal{H}$ be an arbitrary Hilbert space with inner product $\langle \cdot, \cdot \rangle_\mathcal{H}$ and norm $\| \cdot \|_\mathcal{H}$. If $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, then $\mathrm{P}_\mathcal{M}$ denotes the orthogonal projection of $\mathcal{H}$ onto $\mathcal{M}$, and $\mathcal{M}^\perp$ stands for the orthogonal complement of $\mathcal{M}$ in $\mathcal{H}$. If $\mathcal{I}$ is an arbitrary index set and $\{x_k\}_{k \in \mathcal{I}}$ is a collection of vectors in $\mathcal{H}$ then $\overline{\mathrm{span}}\{x_k : k \in \mathcal{I}\}$ denotes the closed linear span of $\{x_k\}_{k \in \mathcal{I}}$ in $\mathcal{H}$.

Throughout the paper, we omit subscripts denoting the space over which an inner product or a norm is defined, whenever the meaning is clear from the context.

B. From Shift-Invariant to U-Invariant Sampling

Continuous-time signals are often analyzed using discrete-time signal processing methods. This paradigm relies on an initial stage where the analog signal of interest is converted to a discrete-time sequence by means of sampling. A simple model for a practical acquisition device comprises a prefilter $s(-t)$ followed by an ideal sampler [15], as depicted in Fig. 1. In this case, the nth sample is given by

$$c_n = \int_{-\infty}^{\infty} \overline{s(t - na)}x(t)\mathrm{d}t \tag{3}$$

where a is the sampling period. Assuming that both $s(t)$ and $x(t)$ are in $L^2(\mathbb{R})$, (3) can be interpreted as the inner product
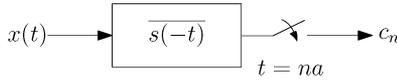
$$c_n = \langle x, s_n \rangle$$

Fig. 1.   SI sampling.

where $s_n(t) = s(t - na)$. The functions $\{s_n\}_{n \in \mathbb{Z}}$ are called the sampling functions, and their closed linear span is termed the sampling space $\mathcal{S}$. One immediate consequence of the fact that each sampling function is a shifted-by-a version of its predecessor, is that if a function $f(t)$ lies in $\mathcal{S}$ then $f(t - na)$ is also in $\mathcal{S}$ for any $n \in \mathbb{Z}$. Spaces possessing this property are called SI spaces.

SI sampling can be considered a special case of a more general acquisition paradigm, where a signal x in an arbitrary Hilbert space $\mathcal{H}$ is measured by a sequence of sampling vectors $\{s_n\}_{n \in \mathbb{Z}}$ in $\mathcal{H}$ that are generated by successive application of some unitary operator U, as in (1). Such sequences of vectors are called stationary sequences. We refer to the space spanned by a stationary sequence as a U-invariant space. Consequently, we term this setting U-*invariant sampling*. Such spaces are a special case of atomic spaces, coined in [20].

Commonly used operators U include compositions in various orders of the *translation*, *modulation*, and *dilatation* operators defined by

$$\mathrm{T}_a \quad : \quad s(t) \mapsto s(t - a) \tag{4}$$
$$\mathrm{M}_a \quad : \quad s(t) \mapsto e^{iat} s(t) \tag{5}$$
$$\mathrm{D}_a \quad : \quad s(t) \mapsto |a|^{-1/2} s(t/|a|) \tag{6}$$

respectively, for an arbitrary $a \neq 0$. SI sampling corresponds to the special case in which $\mathcal{H} = L^2(\mathbb{R})$ and U is the translation operator. Another important U-invariant sampling setting lies at the heart of MRI. There, the continuous-space Fourier transform of a signal x is measured on a lattice of points. Thus, the samples in this imaging modality correspond to inner products $\langle x, \mathrm{M}_a^n s \rangle$, where M_a is the modulation operator (5) and s is some window function. Consequently, this process can be thought of as modulation-invariant sampling [17]. Finally, the Gabor and wavelet transforms of a signal x can be expressed as $c_{m,n} = \langle x, s_{m,n} \rangle$, with $s_{m,n}(t) = \mathrm{M}_{mb} \mathrm{T}_{na} s$ and $s_{m,n}(t) = \mathrm{D}_{mb} \mathrm{T}_{na} s$ respectively [17], for some window $s(t)$ and scalars a and b. Thus, slices of these transforms can be viewed as originating from a U-invariant sampling device.

Note that when U is not the time-shift operator, the index n of the sample sequence c_n does not necessarily correspond to the time variable t of the signal $x(t)$. In fact, our analysis applies also to signal classes, which are not even a function of time. Nevertheless, we are interested in situations where the measurement device produces the samples c_n in a sequential manner, as is the case, e.g., with MRI. Thus, the terms "past" and "future" in this paper correspond to the samples c_n that have already been measured and to those that have not, rather than to the past and future of the signal x itself.

Several fundamental properties of U-invariant sampling and reconstruction were explored in [20]. These include conditions on the function s and operator U in (1) such that the sampling is stable, and a derivation of a consistent recovery technique. Specifically, it was shown how the sequence of

samples $\{c_n\}_{n \in \mathbb{Z}}$ of x, can be processed to produce a recovery $\widetilde{x}$ whose U-invariant samples coincide with those of x. In this paper we study various other aspects of U-invariant sampling by addressing the following problems:

1) *Extrapolation*: Given the past samples $c_0, c_{-1}, c_{-2}, \ldots$ of x, predict the values of the future samples $c_1, c_2, c_3, \ldots$.
2) *Causal Recovery*: Given the sequence of samples $\{c_n\}_{n \in \mathbb{Z}}$ of x, produce a recovery $\widetilde{x}$ by using only causal processing operations.
3) *Causal Stability*: Find conditions on the function s and operator U in (1) such that extrapolation and causal recovery can be carried out in a stable manner.

We note that the issue of stability has been analyzed only in the context of non-causal recovery [20], wheras Problem 3 refers to the study of stability of causal processing. Problems 1 and 2 were given very little attention even in the SI setting, despite the fact that causal processing is crucial for real-time applications. In this paper we develop extrapolation and causal recovery techniques for U-invariant sampling, and then demonstrate them in the special case of SI sampling.

C. Main Results

We now briefly present the main results of the paper.

1) Extrapolation: Prediction is a well-understood problem in the context of stochastic processes. Here, however, the signal x, and consequently the samples c_n, are not random. In this paper we show that whether perfect extrapolation is possible or not is related to the behavior of the spectral density $\Phi_{\mathbf{s}}(e^{i\omega})$, which is the discrete-time Fourier transform (DTFT) of the autocorrelation sequence $r_{\mathbf{s}}(n) = \langle s_n, s_0 \rangle$. Specifically, we show that if $\int_{-\pi}^{\pi} \log \Phi_{\mathbf{s}}(e^{i\omega}) = -\infty$, then the future samples can be perfectly predicted given the past. Interestingly, this is analogous to the well known Paley-Wiener condition for the setting in which c_n are random and the auto-correlation is defined as $r(n) = \mathbb{E}\{c_n \bar{c}_0\}$, where $\mathbb{E}\{\cdot\}$ is the expectation operator [25]. For cases where this condition is not satisfied, we derive a predictor, whose worst-case error is minimal, and provide an expression for the resulting error.

Our approach can be used, for example, to predict the uniform samples of a signal at the output of an antialiasing filter $\overline{s(-t)}$, as shown in Fig. 1, in which case U is the translation operator T_a. Another example is Gabor analysis, in which prediction of the coefficients $c_{m,n}$ in the time direction corresponds to the shift operator T_a and prediction in the frequency direction corresponds to the modulation operator M_b. These two scenarios are depicted in Fig. 2(a) and (b), respectively. Finally, we can predict the coefficients of the wavelet transform $c_{m,n}$ of a signal in the time direction using the translation operator T_{ab^m}, or in the scale direction using the dilation operator D_b. These last two situations are shown in Fig. 2(c) and (d).

As aforementioned, our predictor minimizes the worst-case extrapolation error, for which we provide a closed form expression as a function of the spectral density associated with s and U. Thus, for each of these settings, our analysis yields a condition on s indicating whether perfect prediction is possible or not. In particular, we obtain as a special case the well known result that a bandlimited signal can be perfectly predicted from its past samples taken above the Nyquist rate [31], [32].
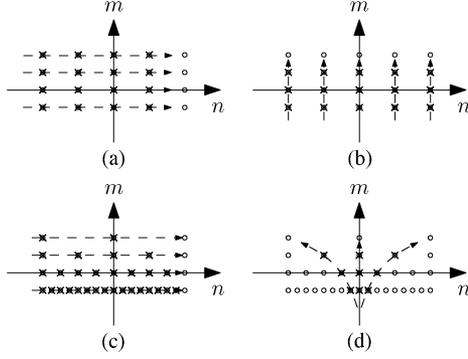
Fig. 2. Scenarios of prediction. (a), (b) Prediction in the time and frequency directions, respectively, of Gabor coefficients. (c), (d) Prediction in the time and scale directions, respectively, of wavelet coefficients.
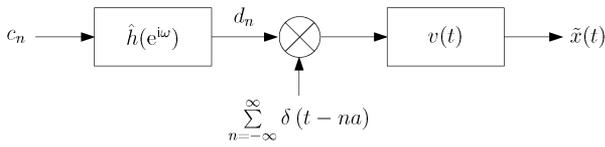

Fig. 3. SI reconstruction.

2) Causal Recovery: Reconstructing a signal $x(t)$ from its SI samples is often approached by employing a SI interpolation formula of the form

$$\tilde{x}(t) = \sum_{n \in \mathbb{Z}} d_n v(t - na) \qquad (7)$$

with some reconstruction kernel $v(t)$ [1]. A widely used method for constructing the sequence d_n relies on the consistency criterion [15]. An attractive feature of this method is that d_n is obtained by feeding the samples c_n into a digital LTI filter. The resulting recovery scheme is shown in Fig. 3.

In [20], a consistent recovery technique was developed for general U-invariant sampling. This was done by replacing the SI reconstruction formula (7) by its U-invariant generalization

$$\tilde{x} = \sum_{n \in \mathbb{Z}} d_n \mathrm{U}^n v. \qquad (8)$$

It was shown that in this case as well d_n is obtained by filtering the samples c_n with an LTI filter. A major drawback of the consistency approach, though, is that it generally leads to a non-causal filter. To make it adequate for use in real-time applications, we address in this paper the problem of designing causal recovery methods. Specifically, in Section V we design a filter whose output $\tilde{d}_n$ minimizes the error $|d_n - \tilde{d}_n|$ for the worst case signal x. We demonstrate our technique in the context of causal SI spline interpolation and show that it is advantageous over the commonly used naive approach of truncating the non-causal solution [10]–[12].

3) Causal Stability: In Section VI we derive conditions such that the past sampling functions $\{s_n\}_{n \le 0}$ corresponding to a U-invariant sampling device form a Riesz basis or a frame. We relate these conditions to the ability to perfectly predict the future samples. In particular, we show that if perfect prediction is possible, then it cannot be done in a stable manner.

III. STATIONARY SEQUENCES

We begin by presenting several basic results from sampling theory and from the theory of stationary sequences, which provide the basis for the derivations in the following sections.

A. Sequences and Sampling

Let $\mathcal{H}$ be a Hilbert space and $\mathbf{s} = \{s_n\}_{n \in \mathbb{Z}}$ be a sequence in $\mathcal{H}$. The set $\mathbf{s}$ is said to be a *frame* for $\mathcal{H}$ if there exist constants $0 < A \le B < \infty$ such that

$$A\|x\|_{\mathcal{H}}^2 \le \sum_{k \in \mathbb{Z}} |\langle x, s_k \rangle_{\mathcal{H}}|^2 \le B\|x\|_{\mathcal{H}}^2 \qquad (9)$$

for all $x \in \mathcal{H}$. The numbers A and B are called *frame bounds*. The sequence $\mathbf{s}$ is called a *Bessel sequence* if it satisfies the right hand side (RHS) inequality of (9), and is a *Riesz basis* for $\mathcal{H}$ if it ceases to be a frame when an arbitrary element is removed.

With every sequence $\mathbf{s}$ in $\mathcal{H}$ we associate the *synthesis operator* $\mathrm{R}_{\mathbf{s}} : \ell^\infty \to \mathcal{H}$ defined by

$$\mathrm{R}_{\mathbf{s}} : \{c_n\}_{n \in \mathbb{Z}} \mapsto \sum_{n \in \mathbb{Z}} c_n s_n. \qquad (10)$$

Its domain is the set of all sequences $\mathbf{c} = \{c_n\}_{n \in \mathbb{Z}}$ of complex numbers for which $\mathrm{R}_{\mathbf{s}}\mathbf{c}$ converges in $\mathcal{H}$. If $\mathbf{s}$ is a Bessel sequence then $\mathrm{R}_{\mathbf{s}}$ is a bounded operator on ℓ^2. The operator $\mathrm{S}_{\mathbf{s}} : \mathcal{H} \to \ell^\infty$ associated with the sequence $\mathbf{s}$ and defined by

$$\mathrm{S}_{\mathbf{s}} : x \mapsto \{\langle x, s_n \rangle_{\mathcal{H}}\}_{n \in \mathbb{Z}}$$

is called the *analysis operator*. If $\mathbf{s}$ is a Bessel sequence then $\mathrm{S}_{\mathbf{s}}$ is a bounded operator which maps $\mathcal{H}$ into ℓ^2 and is the adjoint of $\mathrm{R}_{\mathbf{s}}$. Although from a practical viewpoint it is important that the sampling functions form a Bessel sequence, the results we present in this paper do not require this assumption. Therefore, we refrain from using the adjoint notation $\mathrm{S}_{\mathbf{s}} = \mathrm{R}_{\mathbf{s}}^*$, as commonly done in the sampling literature (e.g., [24]).

Finally, with the sequence $\mathbf{s}$ in $\mathcal{H}$ we associate the subspaces

$$\begin{aligned} \mathcal{S} &= \overline{\operatorname{span}}\{s_n : n \in \mathbb{Z}\} \\ \mathcal{S}_0 &= \overline{\operatorname{span}}\{s_n : n = 0, -1, -2, \ldots\} \end{aligned} \qquad (11)$$

which we call the *sampling space* and the *past sampling space*, respectively.

B. Stationary Sequences

A sequence $\mathbf{s} = \{s_n\}_{n \in \mathbb{Z}}$ of vectors in a separable Hilbert space $\mathcal{H}$ is called a *stationary sequence* if

$$\langle s_{m+k}, s_{n+k} \rangle = \langle s_m, s_n \rangle, \quad \text{for all } m, n, k \in \mathbb{Z}$$

i.e. if the inner product[1] $\langle s_m, s_n \rangle$ depends only on the difference $m - n$. The sequence

$$r_{\mathbf{s}}(n) = \langle s_n, s_0 \rangle, \quad n \in \mathbb{Z} \qquad (12)$$

is called the *covariance function* of $\mathbf{s}$. For every stationary sequence $\mathbf{s}$, there exists a unitary operator U on $\mathcal{H}$ uniquely determined by $\mathbf{s}$, which satisfies

$$s_{n+1} = \mathrm{U}s_n \quad \text{or equivalently } s_n = \mathrm{U}^n s_0$$

[1] In the random signals literature, $\{s_n\}$ is a sequence of random variables and the inner product $\langle s_m, s_n \rangle$ is defined as $\mathbb{E}\{s_m \bar{s}_n\}$.
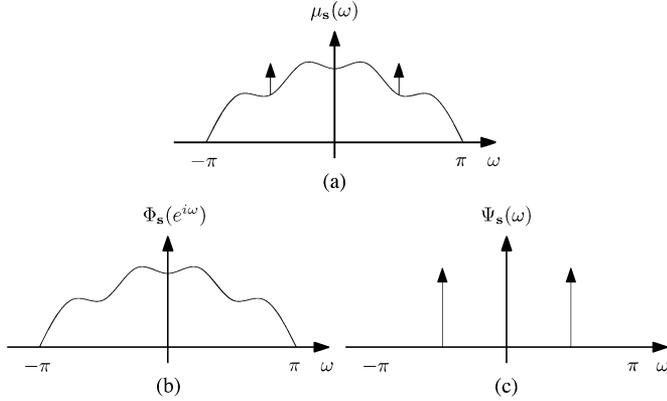
Fig. 4. The Lebesgue decomposition of the spectral measure $\mu_{\mathbf{s}}(\omega)$. (a) corresponds to a combination of the spectral density $\Phi_{\mathbf{s}}(e^{i\omega})$ (b) and a singular part $\Psi_{\mathbf{s}}(\omega)$ (c).

for all $n \in \mathbb{Z}$. Conversely, given an $s \in \mathcal{H}$ and a unitary operator U on $\mathcal{H}$, the sequence $\mathbf{s}$ defined by $s_n = U^n s$, $n \in \mathbb{Z}$ is a stationary sequence. This motivates the following definition.

Definition: Let s be a vector in a Hilbert space $\mathcal{H}$, and let U be a unitary operator on $\mathcal{H}$. The sequence $s_n = U^n s$, $n \in \mathbb{Z}$, is called a stationary sequence generated by (U, s).

By the spectral theorem for unitary operators, the covariance function of a stationary sequence $\mathbf{s}$ generated by (U, s) has a *spectral representation*:

$$r_{\mathbf{s}}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{in\omega} d\mu_{\mathbf{s}}(\omega), \quad n \in \mathbb{Z}$$

where $\mu_{\mathbf{s}}$ is the *spectral measure* of $\mathbf{s}$. In words, the spectral measure $\mu_{\mathbf{s}}(\omega)$ can be thought of as the DTFT of the covariance sequence $r_{\mathbf{s}}(n)$. By Lebesgue decomposition [21]–[23], we have

$$d\mu_{\mathbf{s}}(\omega) = \Phi_{\mathbf{s}}(e^{i\omega})d\omega + d\Psi_{\mathbf{s}}(\omega) \tag{13}$$

where $\Psi_{\mathbf{s}}$ denotes the singular part of $\mu_{\mathbf{s}}$ and $\Phi_{\mathbf{s}} \in L^1(\mathbb{T})$ is the Radon-Nikodym derivative of $\mu_{\mathbf{s}}$ with respect to the Lebesgue measure, which is also known as the *spectral density* of $\mathbf{s}$. Note that in the engineering literature, often no explicit distinction is made between the regular and singular parts, in which case the spectrum may contain delta functions [25]. In our terminology, the singular part $\Psi_{\mathbf{s}}$ comprises the delta functions whereas $\Phi_{\mathbf{s}}$ denotes the rest of the spectrum, as depicted in Fig. 4.

In studying prediction and causal recovery problems, we will encounter expressions of the type $\|\sum_{k \in \mathbb{Z}} c_k s_k\|_{\mathcal{H}}$, where $\mathbf{s} = \{s_k\}_{k \in \mathbb{Z}}$ is a stationary sequence in $\mathcal{H}$ and $\mathbf{c} = \{c_k\}_{k \in \mathbb{Z}}$ is a sequence of scalar coefficients. The spectral representation implies that

$$\langle s_m, s_n \rangle_{\mathcal{H}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)\omega} d\mu_{\mathbf{s}}(\omega)$$

for all $m, n \in \mathbb{Z}$. We thus have the following result.

Lemma 1: Let $R_{\mathbf{s}}\mathbf{c} = \sum_{k \in \mathbb{Z}} c_k s_k$ for some stationary sequence $\mathbf{s} = \{s_k\}_{k \in \mathbb{Z}}$ and scalars $\mathbf{c} = \{c_k\}_{k \in \mathbb{Z}}$, and let $b_{\mathbf{c}}(\omega) =$

$\sum_{k \in \mathbb{Z}} c_k e^{ik\omega}$. Then $b_{\mathbf{c}} \in L^2(\mu_{\mathbf{s}}, \mathbb{T})$ if and only if $R_{\mathbf{s}}\mathbf{c} \in \mathcal{S}$ in which case $\|b_{\mathbf{c}}\|_{\mu_{\mathbf{s}}} = \|R_{\mathbf{s}}\mathbf{c}\|_{\mathcal{H}}$.

Therefore, to evaluate the expression $\|\sum_{k \in \mathbb{Z}} c_k s_k\|_{\mathcal{H}}$, we can replace each s_k by an exponent $e^{ik\omega}$ and the norm on $\mathcal{H}$ by the norm $L^2(\mu_{\mathbf{s}}, \mathbb{T})$. This implies that the sampling space $\mathcal{S}$ (11) associated with $\mathbf{s}$ is isometric isomorphic to $L^2(\mu_{\mathbf{s}}, \mathbb{T})$.

Proof: By definition

$$\begin{aligned} \|R_{\mathbf{s}}\mathbf{c}\|_{\mathcal{H}}^2 &= \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} c_m \bar{c}_n \langle s_m, s_n \rangle_{\mathcal{H}} \\ &= \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} c_m \bar{c}_n \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)\omega} d\mu_{\mathbf{s}}(\omega) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k \in \mathbb{Z}} c_k e^{ik\omega} \right|^2 d\mu_{\mathbf{s}}(\omega) \\ &= \|b_{\mathbf{c}}\|_{\mu_{\mathbf{s}}}^2 \end{aligned} \tag{14}$$

which completes the proof. ∎

Two stationary sequences $\mathbf{s}$ and $\mathbf{w}$ are said to be *stationary correlated* if

$$\langle s_{m+k}, w_{n+k} \rangle = \langle s_m, w_n \rangle, \quad \text{for all } m, n, k \in \mathbb{Z}.$$

In this case, the sequence

$$r_{\mathbf{s},\mathbf{w}}(n) = \langle s_n, w_0 \rangle, \quad n \in \mathbb{Z}$$

is called the *cross-covariance* function. It has a spectral representation

$$r_{\mathbf{s},\mathbf{w}}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{in\omega} d\mu_{\mathbf{s},\mathbf{w}}(\omega), \quad n \in \mathbb{Z}.$$

and *cross-spectral measure* $\mu_{\mathbf{s},\mathbf{w}}$. As before, the cross-spectral measure $\mu_{\mathbf{s},\mathbf{w}}(\omega)$ can be thought of as the DTFT of $r_{\mathbf{s},\mathbf{w}}(n)$. By Lebesgue decomposition we have again $d\mu_{\mathbf{s},\mathbf{w}}(\omega) = \Phi_{\mathbf{s},\mathbf{w}}(e^{i\omega})d\omega + d\Psi_{\mathbf{s},\mathbf{w}}(\omega)$ with the *cross spectral density* $\Phi_{\mathbf{s},\mathbf{w}} \in L^1(\mathbb{T})$. Two stationary sequences $\mathbf{s}$ and $\mathbf{w}$ are stationary correlated if and only if they are generated by the same unitary operator U.

C. Examples

We close this section with some examples of stationary sequences often used in sampling applications.

Example 1 (Translation Operator): We begin by examining the widely studied SI case. Specifically, assume that $\mathcal{H} = L^2(\mathbb{R})$, $s \in L^2(\mathbb{R})$ is an arbitrary function, and $a \neq 0$ is a real number. Let

$$s_n(t) = (T_a^n s)(t) = s(t - na), \quad n \in \mathbb{Z}. \tag{15}$$

Computing the DTFT of the covariance function $r_{\mathbf{s}}(n) = \langle s_n, s_0 \rangle$, the spectral density of $\mathbf{s}$ is given by

$$\Phi_{\mathbf{s}}(e^{i\omega}) = \frac{1}{|a|} \sum_{k \in \mathbb{Z}} \left| \hat{s}\left(\frac{-\omega + 2\pi k}{a} \right) \right|^2, \quad \omega \in [-\pi, \pi)$$

where $\hat{s}$ denotes the Fourier transform (2) of s.

Example 2 (Modulation Operator): Next, we examine the modulation-invariant setting. Specifically, assume that $\mathcal{H} = L^2(\mathbb{R})$, s is an arbitrary function in $L^2(\mathbb{R})$, and $a \neq 0$ is a real number. Let

$$s_n(t) = (\mathrm{M}_a^n s)(t) = \mathrm{e}^{\mathrm{i}ant} s(t), \quad n \in \mathbb{Z}.$$

The spectral density in this case is given by

$$\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = \frac{2\pi}{|a|} \sum_{k \in \mathbb{Z}} \left| s\left(\frac{\omega + 2\pi k}{a}\right) \right|^2, \quad \omega \in [-\pi, \pi).$$

Example 3 (Dilatation Operator): For an arbitrary $a > 0$, consider the sequence

$$s_n(t) = (\mathrm{D}_a^n s)(t) = a^{-n/2} s(a^{-n} t), \quad n \in \mathbb{Z}$$

generated by applying the dilation operator on $s \in L^2(\mathbb{R})$. Its spectral density is given by the Fourier series

$$\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = \sum_{n \in \mathbb{Z}} \alpha_n \mathrm{e}^{\mathrm{i}n\omega}, \quad \omega \in [-\pi, \pi)$$

with the Fourier coefficients

$$\alpha_n = \int_{-\infty}^{\infty} s(a^{n/2} t) \overline{s(a^{-n/2} t)} \mathrm{d}t.$$

Note that in all three examples the singular part $\Psi_{\mathbf{s}}$ of the spectral measure is identically zero. The next example demonstrates that this is not always the case.

Example 4 (Finite Dimensional Spaces): Consider the space $\mathcal{H} = \mathbb{C}^N$ with the standard inner product and let

$$s_n = \mathrm{U}^n s, \quad n \in \mathbb{Z}$$

where $s \in \mathbb{C}^N$ and U is a unitary $N \times N$ matrix. Since U is unitary, it can be factored as $\mathrm{U} = \mathrm{VDV}^*$, where V is unitary and D is diagonal with $\mathrm{D}_{k,k} = \mathrm{e}^{\mathrm{i}\omega_k}$ for some $\omega_k \in [-\pi, \pi)$. Consequently

$$r_{\mathbf{s}}(n) = \langle \mathrm{U}^n s, s \rangle = \sum_{k=1}^{N} |a_k|^2 \mathrm{e}^{\mathrm{i}\omega_k n}$$

where a_k is the kth element of the vector $\mathrm{V}^* s$. This implies that the spectral measure comprises only a singular part, which satisfies[2] $\Psi_{\mathbf{s}}(\{\omega_k\}) = 2\pi |a_k|^2$ and $\Psi_{\mathbf{s}}(\{\Omega\}) = 0$ for every set Ω that does not contain $\{\omega_k\}_{k=1}^{N}$.

Note that the spectral measure $\mu_{\mathbf{s}}$ depends not only on U and s but also on the Hilbert space $\mathcal{H}$ on which the sequence $\mathbf{s}$ is defined, as illustrated next.

Example 5 (Translation Operator in Sobolev Spaces): Assume that $\mathcal{H} = W^{p,2}$ is the Sobolev space of order p

$$W^{p,2} = \left\{ x \in L^2(\mathbb{R}) : x^{(r)} \in L^2(\mathbb{R}) \text{ for all } r \leq p \right\}$$

---

[2] In engineering terms, the spectrum comprises only delta functions, namely $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = 2\pi \sum_{k=1}^{N} |a_k|^2 \delta(\omega - \omega_k)$.

endowed with the inner product

$$\langle x, y \rangle_{W^{p,2}} = \sum_{r=0}^{p} \left\langle x^{(r)}, y^{(r)} \right\rangle$$

where $x^{(r)}$ and $y^{(r)}$ are the rth-order derivatives of x and y, respectively. Then for every $s \in W^{p,2}$ the translation operator T_a considered in Example 1 defines by (15) a stationary sequence $\mathbf{s}$ in $W^{p,2}$. Its spectral density can be calculated by taking the DTFT of $r_{\mathbf{s}}(n) = \langle s_n, s_0 \rangle_{W^{p,2}}$, which yields

$$\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = \frac{1}{|a|} \sum_{k \in \mathbb{Z}} \left(\left| \widehat{s}\left(\frac{-\omega + 2\pi k}{a}\right) \right|^2 \sum_{r=0}^{p} \left| \frac{-\omega + 2\pi k}{a} \right|^{2r} \right)$$

for every $\omega \in [-\pi, \pi)$.

Example 6 (Stationary Correlated Sequences): Consider next the translation operator T_a of Example 1 on $\mathcal{H} = L^2(\mathbb{R})$ with $a \neq 0$. If $s \in L^2(\mathbb{R})$ and $w \in L^2(\mathbb{R})$ are two different functions, then $\mathbf{s} = \{s_n = \mathrm{T}_a^n s\}_{n \in \mathbb{Z}}$ and $\mathbf{w} = \{w_n = \mathrm{T}_a^n w\}_{n \in \mathbb{Z}}$ are two stationary sequences in $L^2(\mathbb{R})$. Moreover, both sequences are stationary correlated since

$$\langle s_{m+k}, w_{n+k} \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}(n-m)a\omega} \widehat{s}(\omega) \overline{\widehat{w}(\omega)} \mathrm{d}\omega = \langle s_m, w_n \rangle.$$

The cross-spectral density is

$$\Phi_{\mathbf{s},\mathbf{w}}(\mathrm{e}^{\mathrm{i}\omega}) = \frac{1}{|a|} \sum_{k \in \mathbb{Z}} \widehat{s}\left(\frac{-\omega + 2\pi k}{a}\right) \overline{\widehat{w}\left(\frac{-\omega + 2\pi k}{a}\right)} \quad (16)$$

for every $\omega \in [-\pi, \pi)$.

However, the two sequences $\mathbf{s} = \{s_n = \mathrm{T}_a^n s\}_{n \in \mathbb{Z}}$ and $\mathbf{w} = \{w_n = \mathrm{T}_b^n w\}_{n \in \mathbb{Z}}$ generated by two different unitary operators T_a and T_b, are not stationary correlated since

$$\langle s_{m+k}, w_{n+k} \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{q\mathrm{i}(nb-ma)\omega} \mathrm{e}^{\mathrm{i}k(b-a)\omega} \widehat{s}(\omega) \overline{\widehat{w}(\omega)} \mathrm{d}\omega$$

depends on the absolute position k and not only on the difference $m - n$.

IV. EXTRAPOLATION

A. General Approach

We consider the following problem: Based on the knowledge of past samples $\mathbf{c}_0 = \{c_n\}_{n \leq 0}$ of a vector $x \in \mathcal{H}$, we want to predict the future samples c_n for $n > 0$, where $c_n = \langle x, s_n \rangle$ for every $n \in \mathbb{Z}$. We begin our analysis with general sampling vectors $\mathbf{s} = \{s_n\}_{n \in \mathbb{Z}}$, and then specialize the discussion to stationary sequences in Sections IV-B and -C.

Our goal is to minimize the prediction error

$$|c_n - \widetilde{c}_n| = |\langle x, s_n \rangle - \widetilde{c}_n| \quad (17)$$

where $\widetilde{c}_n$ is an estimate of c_n that is a function of $\mathbf{c}_0$. Unfortunately, (17) depends on x, which is unknown. Let $\mathbf{s}_0 = \{s_n\}_{n \leq 0}$

denote the subsequence of sampling vectors with a nonpositive index, and let S_{s_0} be the corresponding analysis operator. To eliminate the dependency on x, we instead consider the following min-max problem

$$\min_{\widetilde{c}_n} \max_{x \in \mathcal{B}} |\langle x, s_n \rangle - \widetilde{c}_n| \tag{18}$$

where $\mathcal{B} = \{\xi \in \mathcal{H} : S_{s_0}\xi = \mathbf{c}_0, \|\xi\| \leq L\}$ is the set of signals that could have generated the observed sequence of past samples $\mathbf{c}_0$, and whose norm is bounded by some scalar $L > 0$. Thus, we look for the predictor $\widetilde{c}_n$ that minimizes the prediction error for the worst-case feasible signal x. The following proposition shows that the solution of (18) is equal to the inner product between x and the best approximation of s_n in the past sampling space $\mathcal{S}_0$.

Proposition 2: Let $\mathbf{s}$ be a sequence in a Hilbert space $\mathcal{H}$. Then the unique solution of

$$\arg\min_{\widetilde{c}_n} \max_{x \in \mathcal{B}} |\langle x, s_n \rangle - \widetilde{c}_n| \tag{19}$$

is given by

$$\widetilde{c}_n = \langle x, \mathrm{P}_{\mathcal{S}_0} s_n \rangle \tag{20}$$

where $\mathcal{S}_0$ is the past sampling space (11).

Proof: See Appendix A. ∎

Proposition 2 was proved in [26] for the case in which the sequence $\mathbf{s}_0$ forms a frame for the past sampling space $\mathcal{S}_0$. The proof we provide in Appendix A does not require this assumption and is therefore more general.

Note that (20) does not depend on the bound L. However, it seems to depend on x, which is unknown. Nevertheless, $\widetilde{c}_n$ of (20) can be written explicitly as a linear combination of the past samples, which are given. Indeed, we can write[3]

$$\mathrm{P}_{\mathcal{S}_0} s_n = \sum_{k=0}^{\infty} \gamma_k s_{-k} \tag{21}$$

for some sequence of coefficients $\boldsymbol{\gamma} = \{\gamma_k\}_{k=0}^{\infty}$ (which depends on n). Consequently,

$$\widetilde{c}_n = \left\langle x, \sum_{k=0}^{\infty} \gamma_k s_{-k} \right\rangle = \sum_{k=0}^{\infty} \overline{\gamma}_k c_{-k}. \tag{22}$$

Interestingly, the min-max predictor is linear in the past samples, although we did not restrict ourselves to linear schemes in (19). In Sections IV-B and -C we discuss how to compute the coefficients $\{\gamma_k\}_{k=0}^{\infty}$.

When $\mathbf{s}$ is a stationary sequence, the U-invariance structure implies that for any $m \in \mathbb{Z}$, the min-max predictor of c_{m+n} based on the samples $\{c_k\}_{k \leq m}$ is given by $\widetilde{c}_{m+n} = \sum_{k=0}^{\infty} \overline{\gamma}_k c_{m-k}$. Therefore, in the stationary setting, the n-step-ahead predicted sequence $\{\widetilde{c}_{k+n}\}_{k \in \mathbb{Z}}$ is computed by feeding the sample sequence $\{c_k\}_{k \in \mathbb{Z}}$ into the causal filter whose impulse response is $h(k) = \gamma_k$ for $k \geq 0$ and $h(k) = 0$ for $k < 0$. The transfer function of this filter is $\Gamma(e^{i\omega}) = \sum_{k=0}^{\infty} \gamma_k e^{-ik\omega}$.

---

[3]More precisely, we can approximate $\mathrm{P}_{\mathcal{S}_0} s_n$ as closely as desired by linear combinations of the form (21).

Since the orthogonal projection is self-adjoint, one obtains from (20) that $\widetilde{c}_n = \langle \mathrm{P}_{\mathcal{S}_0} x, s_n \rangle$. This allows the interpretation that the optimal prediction $\widetilde{c}_n$ is obtained by first approximating the signal x by its orthogonal projection onto the past sampling space $\mathcal{S}_0$, and then sampling this approximation with the sampling function s_n.

From (20), it follows that the prediction error attained by the min-max predictor for any $x \in \mathcal{H}$ is

$$|c_n - \widetilde{c}_n| = |\langle x, s_n - \mathrm{P}_{\mathcal{S}_0} s_n \rangle| = \left| \left\langle x, \mathrm{P}_{\mathcal{S}_0^\perp} s_n \right\rangle \right|. \tag{23}$$

This error will usually be nonzero. However, there always exist signals $x \in \mathcal{H}$ with zero prediction error. This happens, for example, for all $x \in \mathcal{S}_0$ because then $\widetilde{c}_n = \langle \mathrm{P}_{\mathcal{S}_0} x, s_n \rangle = \langle x, s_n \rangle = c_n$. An interesting question is under what conditions on the sampling functions $\mathbf{s}$ the prediction error becomes zero for *every* signal $x \in \mathcal{H}$. If this happens, we will say that the sample c_n can be *perfectly predicted* from the past samples $\mathbf{c}_0$. From (23) we immediately have the following observation.

Proposition 3: Let $\mathbf{s}$ be a sequence in a Hilbert space $\mathcal{H}$. Then the sample c_n can be perfectly predicted from the past samples $\mathbf{c}_0$ if and only if $s_n \in \mathcal{S}_0$. In particular, perfect prediction is possible if the spaces $\mathcal{S}_0$ and $\mathcal{S}$ of (11) are equal.

B. Stationary Perfect Prediction

We now show that when $\mathbf{s}$ is stationary, Proposition 3 translates into a simple condition on the spectral density $\Phi_{\mathbf{s}}$. This result is obtained by using Szegö's famous theorem [27].

Theorem 4: Let $\mathbf{s} = \{s_n\}_{n \in \mathbb{Z}}$ be a sequence of sampling functions in a Hilbert space $\mathcal{H}$, let $x \in \mathcal{H}$ be arbitrary, and let $\mathbf{c}_0 = \{\langle x, s_n \rangle\}_{n \leq 0}$ be the past generalized samples of x. Assume that $\mathbf{s}$ is a stationary sequence with spectral measure $\mathrm{d}\mu_{\mathbf{s}}(\omega) = \Phi_{\mathbf{s}}(e^{i\omega})\mathrm{d}\omega + \mathrm{d}\Psi_{\mathbf{s}}(\omega)$. Then the future samples $c_1, c_2, c_3, \ldots$ can be perfectly predicted from $\mathbf{c}_0$ if and only if

$$\int_{-\pi}^{\pi} \log \Phi_{\mathbf{s}}(e^{i\omega})\mathrm{d}\omega = -\infty. \tag{24}$$

Before proving the theorem, we note that condition (24) also characterizes perfect predictability in stochastic scenarios. Specifically, the well known Paley-Wiener theorem [25] states that a sequence c_n of random variables with autocorrelation $r(n) = \mathbb{E}\{c_n \overline{c}_0\}$ can be perfectly predicted if and only if the DTFT $\Phi(e^{i\omega})$ of $r(n)$ satisfies (24). Thus, the spectral density $\Phi_{\mathbf{s}}(e^{i\omega})$ in our case can be thought of as the analog of the spectrum of the samples in the stochastic setting.

Proof: First, we consider the case of one step prediction $n = 1$. From Proposition 2, the prediction error obtained by the min-max predictor can be written using (23) as

$$|c_1 - \widetilde{c}_1| = |\langle x, r \rangle| \leq \|x\| \|r\| \tag{25}$$

where we denoted $r = s_1 - \mathrm{P}_{\mathcal{S}_0} s_n$ and used the Cauchy–Schwarz inequality. Moreover, there always exists an $x \in \mathcal{H}$ for which equality holds in (25). Using (21), r can be written as $r = s_1 - \sum_{k=0}^{\infty} \gamma_k s_{-k}$. Thus, perfect prediction is possible if and only if $\|r\| = \|s_1 - \sum_{k=0}^{\infty} \gamma_k s_{-k}\|_{\mathcal{H}}$ can be made arbitrarily small by a proper choice of the sequence of coefficients $\boldsymbol{\gamma}$.
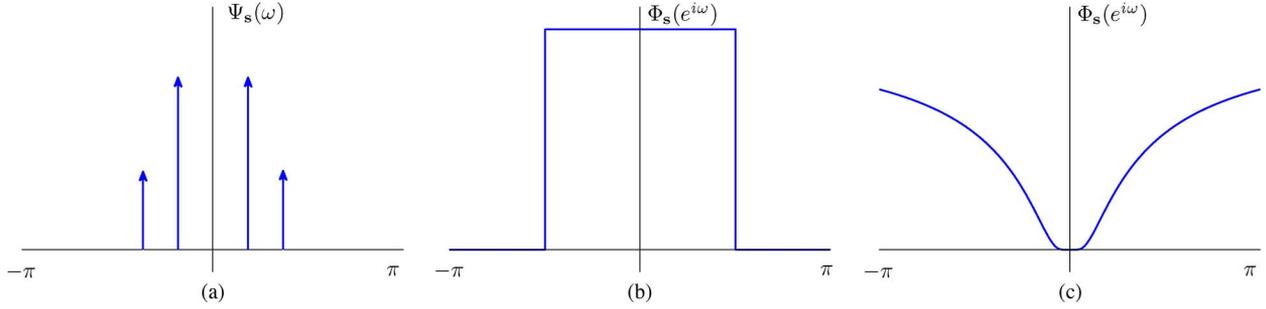
Fig. 5. Scenarios of perfect prediction. (a) A completely singular measure. (b) A bandlimited spectral density. (c) A spectral density with a locally flat zero: $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = \exp\{-1/|\omega|\}$.

Employing Lemma 1

$$\left\| s_1 - \sum_{k=0}^{N-1} \gamma_k s_{-k} \right\|_{\mathcal{H}}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \mathrm{e}^{\mathrm{i}\omega} - \sum_{k=0}^{N-1} \gamma_k \mathrm{e}^{-\mathrm{i}\omega k} \right|^2 \mathrm{d}\mu_{\mathbf{s}}(\omega)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| 1 - \sum_{k=1}^{N} \gamma_{k-1} \mathrm{e}^{-\mathrm{i}\omega k} \right|^2 \mathrm{d}\mu_{\mathbf{s}}(\omega)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| 1 - \widetilde{p}_0(\mathrm{e}^{-\mathrm{i}\omega}) \right|^2 \mathrm{d}\mu_{\mathbf{s}}(\omega)$$

where $\widetilde{p}_0 \in \mathcal{P}_0^+$ denotes the causal complex polynomial $\widetilde{p}_0(\mathrm{e}^{\mathrm{i}\omega}) = \sum_{k=1}^{N} \gamma_{k-1} \mathrm{e}^{\mathrm{i}\omega k}$ of a certain finite degree N. Now, Szegö's theorem states that

$$\inf_{\widetilde{p}_0 \in \mathcal{P}_0^+} \int_{-\pi}^{\pi} \left| 1 - \widetilde{p}_0(\mathrm{e}^{-\mathrm{i}\omega}) \right|^2 \mathrm{d}\mu_{\mathbf{s}}(\omega) = 0$$

if and only if (24) is satisfied (see, e.g., [28, Ch. IV.3]). In this case there exists a function $p_0(\mathrm{e}^{\mathrm{i}\omega}) = \sum_{k=0}^{\infty} \gamma_{k-1} \mathrm{e}^{\mathrm{i}\omega k}$ in the closure of $\mathcal{P}_0^+$ in $L^2(\mu_{\mathbf{s}}, \mathbb{T})$ such that the RHS of (25) vanishes, i.e., such that the estimation error is zero.

By induction, it follows that if we can perfectly predict c_1, then we can also perfectly predict c_n for all $n > 1$. ∎

There are three different cases in which condition (24) is satisfied, as demonstrated in Fig. 5:

a) *Singular case*: The spectral measure $\mu_{\mathbf{s}} = \Psi_{\mathbf{s}}$ is completely singular, as shown in Fig. 5(a).

b) *Bandlimited case*: The spectral density $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$ vanishes on a set of positive Lebesgue measure, as shown in Fig. 5(b).

c) *Locally-flat zero case*: $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) \neq 0$ almost everywhere on $[-\pi, \pi)$ but (24) holds, as shown in Fig. 5(c). For example, $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = \exp\{-|\omega|^{-\alpha}\}$ with $\alpha \geq 1$.

To determine the predictor, recall from the proof of Theorem 4 that the prediction error is upper bounded by

$$|c_1 - \widetilde{c}_1|^2 \leq \|x\|^2 \inf_{\widetilde{p}_0 \in \mathcal{P}_0^+} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| 1 - \widetilde{p}_0(\mathrm{e}^{-\mathrm{i}\omega}) \right|^2 \mathrm{d}\mu_{\mathbf{s}}(\omega).$$

Let ϵ be an arbitrary small admissible prediction error. Then it suffices to find a finite-degree polynomial $\widetilde{p}_0 \in \mathcal{P}_0^+$ such that

$$\int_{-\pi}^{\pi} \left| 1 - \widetilde{p}_0(\mathrm{e}^{-\mathrm{i}\omega}) \right|^2 \mathrm{d}\mu_{\mathbf{s}}(\omega) < \frac{2\pi\epsilon^2}{L^2} \tag{26}$$

in order that the prediction error $|c_0 - \widetilde{c}_0|$ be smaller than ϵ for every signal $x \in \mathcal{H}$ whose norm is bounded by L. If $\widetilde{p}_0$ is such a polynomial then it can be written as

$$\widetilde{p}_0(\mathrm{e}^{\mathrm{i}\omega}) = \sum_{k=1}^{N} \gamma_{k-1} \mathrm{e}^{\mathrm{i}k\omega} \tag{27}$$

and the coefficients of the corresponding linear predictor (22) are given by the coefficients $\{\gamma_k\}_{k=0}^{N-1}$ of $\widetilde{p}_0$.

a) Singular case: In this setting, the measure $\mu_{\mathbf{s}}$ is concentrated on a set $\Omega \subset [-\pi, \pi)$ of Lebesgue measure zero. Assume first that Ω is composed of a finite number of frequencies: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. Then the polynomial

$$\widetilde{p}_0(\mathrm{e}^{\mathrm{i}\omega}) = 1 - \left(1 - \mathrm{e}^{\mathrm{i}\omega_1} \mathrm{e}^{\mathrm{i}\omega}\right) \times \cdots \times \left(1 - \mathrm{e}^{\mathrm{i}\omega_N} \mathrm{e}^{\mathrm{i}\omega}\right)$$

clearly satisfies $|1 - \widetilde{p}_0(\mathrm{e}^{-\mathrm{i}\omega})| = 0$ for every $\omega \in \Omega$ so that $\int_{-\pi}^{\pi} |1 - p_0(\mathrm{e}^{-\mathrm{i}\omega})|^2 \mathrm{d}\Psi_{\mathbf{s}}(\omega) = 0$. By writing this polynomial in the form (27), we extract the coefficients $\{\gamma_k\}_{k=0}^{N-1}$ of the predictor, which is a finite-impulse response (FIR) filter.

The more general case, in which $\Omega \subset [-\pi, \pi)$ is an arbitrary set of Lebesgue measure zero, can be treated in a similar manner. Specifically, let $\mathcal{A}$ be the disk algebra, i.e., the closure of the polynomials $\mathcal{P}^+$ in the infinity norm $\|p\|_\infty = \inf_{|z|<1} |p(z)|$. Then a theorem of Fatou [29] states that there exists a function $f \in \mathcal{A}$ such that for every z with $|z| \leq 1$ we have $f(z) = 0$ if and only if $z = \mathrm{e}^{-\mathrm{i}\omega}$ with $\omega \in \Omega$ (see, e.g., [30, Ch. 2.2] where also an explicit construction of f is given). In particular $f(0) \neq 0$ and we can define $p(z) := f(z)/f(0)$ which can be written as

$$p(z) = 1 - \sum_{k=1}^{\infty} \gamma_{k-1} z^k = 1 - \widetilde{p}_0(z).$$

By this construction $\int_{-\pi}^{\pi} |1 - \widetilde{p}_0(\mathrm{e}^{-\mathrm{i}\omega})|^2 \mathrm{d}\Psi_{\mathbf{s}}(\omega) = 0$ and the desired polynomial, which satisfies (26), is obtained from $\widetilde{p}_0$ by writing $\widetilde{p}_0(z) = \sum_{k=1}^{N} \gamma_{k-1} z^k$ for a sufficiently large degree N.
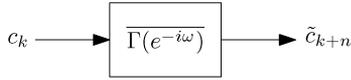
Fig. 6. The min-max optimal n-step-ahead predictor corresponds to the causal filter $\Gamma(\mathrm{e}^{\mathrm{i}\omega})$ of (28).

Fig. 7. Sampling at the output of an *RC* circuit.

b) Bandlimited and locally-flat zero cases: The bandlimited case is well studied in the engineering literature in the context of stochastic processes (see, e.g., the discussion and references in [31]). In this case, the spectral density $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$ is concentrated on a set $\Omega \subset [-\pi, \pi)$ of Lebesgue measure strictly smaller than 2π. Here, the existence of a polynomial $\widetilde{p}_0 \in \mathcal{P}_0^+$ which satisfies (26) follows from the fact that the set $\{\mathrm{e}^{\mathrm{i}n\omega}\}_{n=1}^{\infty}$ is dense in $L^2(\Omega)$, the set of square integrable functions on Ω [30], [31].

In terms of constructing a polynomial $\widetilde{p}_0$ complying with a given (arbitrarily small) prediction error, both the locally-flat zero scenario and the bandlimited setting can be treated in the spirit of the approach presented in [32]. Specifically, the spectral density $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$ can be approximated arbitrarily well by a different spectral density $\widetilde{\Phi}_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$, which does not satisfy (24). In Section IV-C, we present a closed form formula for the predictor minimizing the prediction error associated with such a $\widetilde{\Phi}_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$. Therefore, by an appropriate choice of $\widetilde{\Phi}_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$, and using the result we present in Section IV-C, a predictor meeting the desired admissible error, can be designed.

C. Stationary Estimation of Future Samples

If (24) is not met, then perfect prediction of future samples is impossible. However, (20) still provides a predictor which minimizes the worst-case error. To explicitly compute this predictor, we have to find the orthogonal projection of s_n onto the past sampling space $\mathcal{S}_0$. For stationary sequences $\mathbf{s}$ whose spectral measure contains no singular part (i.e. $\mathrm{d}\mu_{\mathbf{s}}(\omega) = \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega$) and whose spectral density $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$ does not vanish anywhere, the solution is well known in the field of stochastic-processes and is given by (see, e.g., [23])

$$\mathrm{P}_{\mathcal{S}_0} s_n = \sum_{k=0}^{\infty} \gamma_k s_{-k}$$

where $\{\gamma_k\}_{k=0}^{\infty}$ is the inverse DTFT of the Wiener predictor

$$\Gamma(\mathrm{e}^{\mathrm{i}\omega}) = \frac{\left\{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})\mathrm{e}^{\mathrm{i}\omega n}\right\}_+}{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})}. \tag{28}$$

Here, $\Phi_{\mathbf{s}}^+$ denotes a spectral factor of the spectral density $\Phi_{\mathbf{s}}$, which is a function in $L^2(\mathbb{T})$ satisfying $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = |\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})|^2$, and whose inverse DTFT is a causal stable sequence.[4] The operator $\{\cdot\}_+ : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ sets the past coefficients of its argument to zero in the time domain: $\{\sum_{k=-\infty}^{\infty} \alpha_k \mathrm{e}^{-\mathrm{i}\omega k}\}_+ = \sum_{k=0}^{\infty} \alpha_k \mathrm{e}^{-\mathrm{i}\omega k}$. To summarize, according to (22), the optimal n-step-ahead prediction $\{\widetilde{c}_{k+n}\}_{k\in\mathbb{Z}}$, is obtained by feeding the samples c_k into the causal filter with transfer function $\overline{\Gamma(\mathrm{e}^{-\mathrm{i}\omega})}$ corresponding to (28). This is demonstrated schematically in Fig. 6.
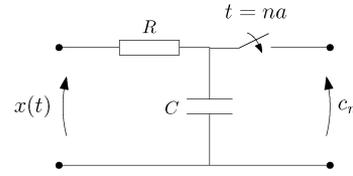
[4]It exists since $\int_{-\pi}^{\pi} \log \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega > -\infty$.

For the case of one-step-ahead prediction $(n = 1)$, Szegö's theorem also provides an upper bound for the resulting error in terms of the spectral density of $\mathbf{s}$.

Theorem 5: Assume that $\mathbf{s}$ is a stationary sequence in $\mathcal{H}$ whose spectral density $\Phi_{\mathbf{s}}$ does not satisfy (24). Then the minimal one step prediction error is upper bounded by

$$|c_1 - \widetilde{c}_1|^2 \leq \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega\right\} \|x\|^2$$

with equality if and only if $x = \alpha \mathrm{P}_{\mathcal{S}_0^{\perp}} s_1$ for some $\alpha \in \mathbb{C}$.

Proof: It was already shown in the proof of Theorem 4 that the minimal prediction error is upper bounded by

$$|c_1 - \widetilde{c}_1|^2 \leq \|x\|^2 \|s_1 - \mathrm{P}_{\mathcal{S}_0} s_1\|^2$$

$$= \|x\|^2 \inf_{\widetilde{p}_0 \in \mathcal{P}_0^+} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|1 - \widetilde{p}_0(\mathrm{e}^{-\mathrm{i}\omega})\right|^2 \mathrm{d}\mu_{\mathbf{s}}(\omega).$$

Szegö's theorem (see, e.g., [28, Ch. IV.3]) states that the infimum on the RHS is equal to $\exp\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega\}$. The first inequality in the above equation follows from the Cauchy-Schwarz inequality (cf. proof of Theorem 4). Consequently, equality holds if and only if x is a multiple of $s_1 - \mathrm{P}_{\mathcal{S}_0} s_1 = \mathrm{P}_{\mathcal{S}_0^{\perp}} s_1$. ∎

Note that the upper bound is independent of the singular part of the spectral measure $\mu_{\mathbf{s}}$ since the sample component associated with this part can be perfectly predicted. Furthermore, the bound is attained by signals whose past samples are identically zero. Indeed, we have seen that the bound is attained by all signals of the form $x = \alpha \mathrm{P}_{\mathcal{S}_0^{\perp}} s_1$, whose samples are given by $c_n = \langle x, s_n \rangle = \alpha \langle \mathrm{P}_{\mathcal{S}_0^{\perp}} s_1, \mathrm{P}_{\mathcal{S}_0} s_n \rangle = 0$. For these signals, the optimal predictor of the future samples will give $\widetilde{c}_n = 0$ for all $n = 1, 2, 3, \ldots$.

D. RC-Circuit Sample Prediction

To demonstrate prediction of future samples, we consider the SI setting of Fig. 1 with an *RC* sampling filter, as shown in Fig. 7. The impulse response in this case is given by

$$s(-t) = \frac{1}{\tau}\mathrm{e}^{-\frac{t}{\tau}}u(t)$$

where $\tau = RC$ is the time constant of the circuit and $u(t)$ is the unit step function.

The autocorrelation function associated with $s(t)$ is

$$r_{\mathbf{s}}(n) = \int_{-\infty}^{\infty} s(t - na)s(t)\mathrm{d}t = \frac{1}{2\tau}\mathrm{e}^{-\frac{|n|a}{\tau}}.$$

TABLE I
FREQUENCY RESPONSE OF ONE-STEP-AHEAD B-SPLINE SAMPLE PREDICTORS FOR VARIOUS DEGREES N

	$N = 0$	$N = 1$	$N = 2$
$\Phi^+(z)$	a_0	$a_1 \left(1 + \left(2 - \sqrt{3}\right) z^{-1}\right)$	$a_2 \left(1 + 0.375 z^{-1} + 0.005 z^{-2}\right)$
$\Gamma(z)$	0	$\dfrac{2 - \sqrt{3}}{1 + (2 - \sqrt{3}) z^{-1}}$	$\dfrac{0.375 + 0.005 z^{-1}}{1 + 0.375 z^{-1} + 0.005 z^{-2}}$

This sequence can be factored as

$$r_{\mathbf{s}}(n) = \left(r_{\mathbf{s}}^+(\cdot) * r_{\mathbf{s}}^+(-\cdot)\right)(n)$$

where

$$r_{\mathbf{s}}^+(k) = \sqrt{\frac{1 - \mathrm{e}^{-\frac{2a}{\tau}}}{2\tau}} \mathrm{e}^{-\frac{ak}{\tau}} u(k) \qquad (29)$$

is a stable sequence. We therefore conclude that $r_{\mathbf{s}}^+(k)$ is the inverse DTFT of the spectral factor $\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})$.

Using (29), for any $n \geq 0$ the numerator $\{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})\mathrm{e}^{\mathrm{i}\omega n}\}_+$ in (28) corresponds to the sequence

$$r_{\mathbf{s}}^+(k + n) u(k) = \sqrt{\frac{1 - \mathrm{e}^{-\frac{2a}{\tau}}}{2\tau}} \mathrm{e}^{-\frac{a(k+n)}{\tau}} u(k + n) u(k)$$

$$= \mathrm{e}^{-\frac{an}{\tau}} r_{\mathbf{s}}^+(k)$$

implying that $\{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})\mathrm{e}^{\mathrm{i}\omega n}\}_+ = \mathrm{e}^{-(an/\tau)} \Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})$. Consequently, the prediction filter (28) is given by

$$\gamma_k = \begin{cases} \mathrm{e}^{-\frac{an}{\tau}} & k = 0 \\ 0 & k \neq 0. \end{cases}$$

This implies that the min-max optimal n-step ahead prediction is obtained by shrinking the last available sample, namely

$$\widetilde{c}_{k+n} = \mathrm{e}^{-\frac{an}{\tau}} c_k.$$

E. B-Spline Sample Prediction

As another example, consider the SI setting of Fig. 1 with a B-spline sampling filter $s(t)$ and a sampling period of $a = 1$. A B-spline of degree N, denoted $\beta^N(t)$, is recursively defined by

$$\beta^N(t) = (\beta^{N-1} * \beta^0)(t) \qquad (30)$$

where $\beta^0(t)$ is the unit square

$$\beta^0(t) = \begin{cases} 1 & -\frac{1}{2} < t < \frac{1}{2}; \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $\beta^N(t)$ is obtained by the $(N+1)$-fold convolution of $\beta^0(t)$. The filter $\beta^0(t)$ is a good model for optical systems in which the effect of the point spread function (PSF) of the lens is negligible with respect to pixel size. B-splines of higher degrees have a bell-like shape, which constitutes a good model for imaging devices with nonnegligible PSF.

The spectral density of an Nth-degree B-spline sampler can be computed using the convolution property (30) and the fact that $\beta^N(t)$ is symmetric

$$r_{\mathbf{s}}(n) = \langle s_n, s_0 \rangle = \int_{-\infty}^{\infty} \beta^N(t - n)\beta^N(t)\mathrm{d}t = \beta^{2N+1}(n).$$

Expressions for the Z-transform $\Phi_{\mathbf{s}}(z)$ of $r_{\mathbf{s}}(n)$ and for the corresponding spectral factorization $\Phi_{\mathbf{s}}^+(z)$ for B-splines of various

degrees can be found in [33]. In Table I, we provide the formula for $\Phi_{\mathbf{s}}^+(z)$ for B-splines of degrees 0, 1, and 2, as well as the resulting frequency response of the one-step-ahead ($n = 1$) Wiener predictor $\Gamma(z)$. The values a_0, a_1, and a_2, appearing in the table, are constants which do not affect $\Gamma(z)$.

As can be seen in the table, the predictor corresponding to a B-spline of degree 0, is identically zero. This follows from the fact that the functions $\{\beta^0(t - n)\}_{n \in \mathbb{Z}}$ do not overlap so that the past samples carry no information about the future. For B-splines of degree 1 and 2, the predictor is a simple infinite impulse response (IIR) filter, which can be implemented efficiently. For example, for $N = 2$, we have the recursive implementation $\widetilde{c}_{n+1} = 0.375(c_n - \widetilde{c}_n) + 0.005(c_{n-1} - \widetilde{c}_{n-1})$.

V. CAUSAL RECOVERY

A common task in signal processing is that of recovering a continuous-time signal $x(t)$ from a sequence of uniformly spaced samples $\mathbf{c} = \{c_n\}_{n \in \mathbb{Z}}$ taken at the output of a filter $\overline{s(-t)}$, as depicted in Fig. 1. As discussed in Section I, the samples in this case correspond to the inner products $\langle x, s_n \rangle$, with the functions $s_n(t) = s(t - na) = (\mathrm{T}_a^n s)(t)$. This task is often approached by employing a recovery formula of the form (7) with some predefined reconstruction kernel $v(t)$ [1]. The overall recovery scheme, therefore, comprises a digital correction filter followed by a digital-to-analog reconstruction stage, as shown in Fig. 3.

In our setting, $s_n = \mathrm{U}^n s$ for some unitary operator U, and thus it is reasonable to replace the SI reconstruction formula (7) by its U-invariant generalization (8). A common method for constructing the sequence of coefficients $\mathbf{d} = \{d_n\}_{n \in \mathbb{Z}}$ results from requiring that $\langle \widetilde{x}, s_n \rangle = c_n$ for all $n \in \mathbb{Z}$ [15]. In [20] it was shown that such a consistent recovery is unique if the cross-spectral density $\Phi_{\mathbf{s}, \mathbf{v}}(\mathrm{e}^{\mathrm{i}\omega})$ associated with the sampling kernel s and reconstruction kernel v, satisfies $|\Phi_{\mathbf{s}, \mathbf{v}}(\mathrm{e}^{\mathrm{i}\omega})| \geq A$ for some $A > 0$ and for all $\omega \in [-\pi, \pi)$. In this case, d_n is obtained by filtering the samples c_n with

$$\widehat{h}_{\mathrm{nc}}(\mathrm{e}^{\mathrm{i}\omega}) = \frac{1}{\Phi_{\mathbf{s}, \mathbf{v}}(\mathrm{e}^{-\mathrm{i}\omega})}. \qquad (31)$$

A major drawback of the consistency approach is that the correction filter (31) is generally noncausal, i.e., $h_{\mathrm{nc}}(n)$ does not vanish for $n < 0$. Thus, our goal is to approximate d_n by relying only on the set of samples $\{c_k\}_{k \leq 0}$. To this end, we observe that d_n can be written as

$$d_n = \sum_{k \in \mathbb{Z}} c_k h_{\mathrm{nc}}(n - k)$$

$$= \sum_{k \in \mathbb{Z}} \langle x, s_k \rangle h_{\mathrm{nc}}(n - k)$$

$$= \left\langle x, \sum_{k \in \mathbb{Z}} \overline{h_{\mathrm{nc}}(n - k)} s_k \right\rangle$$

$$= \langle x, w_n \rangle,$$

where $w_n = \mathrm{U}^n w$ with w given by

$$w = \sum_{k \in \mathbb{Z}} \overline{h_{\mathrm{nc}}(-k)} s_k. \tag{32}$$

This representation allows a simple interpretation of our task. Given the past samples $c_k = \langle x, s_k \rangle$, $k \leq 0$, corresponding to a stationary sequence generated by (U, s), we would like to produce an estimate $\widetilde{d}_n$ of the generalized sample $d_n = \langle x, w_n \rangle$ for some $n \in \mathbb{Z}$, corresponding to the stationary sequence $\mathbf{w} = \{w_k\}_{k \in \mathbb{Z}}$ generated by (U, w). Note that the discussion in Section IV on extrapolation of future samples is a special case of this problem, corresponding to $w = s$. We now address this problem in detail.

A. Causal Estimation of a Different Representation

The problem of approximating a set of generalized samples of a signal based on a second set of generalized samples corresponding to a frame sequence was addressed in [26]. Here, we do not require that $\mathbf{s}_0$ and $\mathbf{w}$ constitute frames. Furthermore, we would like to exploit the knowledge that $\mathbf{s}$ and $\mathbf{w}$ are stationary correlated, to obtain simple closed form expressions for the optimal predictor.

As in Section IV, we look for an estimate $\widetilde{d}_n$ for the generalized sample d_n, which minimizes the error for the worst-case feasible signal

$$\min_{\widetilde{d}_n} \max_{x \in \mathcal{B}} \left| \langle x, w_n \rangle - \widetilde{d}_n \right|$$

where $\mathcal{B} = \{\xi \in \mathcal{H} : \mathrm{S}_{\mathbf{s}_0} \xi = \mathbf{c}_0, \|\xi\| \leq L\}$ is the set of signals that could have generated the observed sequence of past samples $\mathbf{c}_0$, and whose norm is bounded by some scalar $L > 0$. The solution to this min-max problem (the proof is as in Lemma 2) is given by

$$\widetilde{d}_n = \langle x, \mathrm{P}_{\mathcal{S}_0} w_n \rangle.$$

The resulting estimation error is given by

$$|d_n - \widetilde{d}_n| = |\langle x, w_n - \mathrm{P}_{\mathcal{S}_0} w_n \rangle| = \left| \left\langle x, \mathrm{P}_{\mathcal{S}_0^\perp} w_n \right\rangle \right|.$$

This implies that the estimation error becomes zero for every signal $x \in \mathcal{H}$ (perfect estimation) if and only if $w_n \in \mathcal{S}_0$.

To obtain a closed form expression for the min-max optimal estimator, we note that $\widetilde{d}_n$ is given by a linear combination of the past samples. Indeed, we can write

$$\mathrm{P}_{\mathcal{S}_0} w_n = \sum_{k=0}^{\infty} \gamma_k s_{-k}$$

with a sequence $\boldsymbol{\gamma} = \{\gamma_k\}_{k=0}^{\infty}$, which depends on n. Consequently, the min-max optimal estimator is of the form

$$\widetilde{d}_n = \left\langle x, \sum_{k=0}^{\infty} \gamma_k s_{-k} \right\rangle = \sum_{k=0}^{\infty} \overline{\gamma}_k c_{-k}. \tag{33}$$

Similarly, it can be easily verified that when $\mathbf{s}$ and $\mathbf{w}$ are stationary correlated, the min-max estimator of d_{m+n} based on the samples $\{c_k\}_{k \leq m}$ is given by $\widetilde{d}_{m+n} = \sum_{k=0}^{\infty} \overline{\gamma}_k c_{m-k}$. Therefore, the n-step-ahead estimated sequence $\{\widetilde{d}_{k+n}\}_{k \in \mathbb{Z}}$ is computed by feeding the sample sequence $\{c_k\}_{k \in \mathbb{Z}}$ into the causal filter whose impulse response is $h(k) = \overline{\gamma}_k$ for $k \geq 0$ and

$h(k) = 0$ for $k < 0$. The transfer function of this filter is $\Gamma(\mathrm{e}^{\mathrm{i}\omega}) = \sum_{k=0}^{\infty} \gamma_k \mathrm{e}^{-\mathrm{i}k\omega}$.

To obtain an explicit expression for the coefficient sequence $\boldsymbol{\gamma}$, we next assume that $\mathbf{s}$ and $\mathbf{w}$ are stationary correlated with cross-spectral measure $\mathrm{d}\mu_{\mathbf{s},\mathbf{w}}(\omega) = \Phi_{\mathbf{s},\mathbf{w}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega$. For simplicity, we also assume that the spectral measure of the sequence $\mathbf{s}$ is given by $\mathrm{d}\mu_{\mathbf{s}}(\omega) = \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega$ with a spectral density $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$ that does not vanish anywhere. It is well known in the stochastic literature that the sequence $\boldsymbol{\gamma}$ is given in this case by the inverse DTFT of the causal Wiener filter (see, e.g., [23])

$$\Gamma(\mathrm{e}^{\mathrm{i}\omega}) = \frac{1}{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})} \left\{ \frac{\Phi_{\mathbf{w},\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) \mathrm{e}^{\mathrm{i}\omega n}}{\overline{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})}} \right\}_+. \tag{34}$$

Thus, according to (33), the filter $\overline{\Gamma(\mathrm{e}^{-\mathrm{i}\omega})}$ corresponding to (34) causally converts between the two signal representations associated with the sampling sequences $\mathbf{s}$ and $\mathbf{w}$.

B. Application to Causal Recovery

We now utilize (34) to approximate the consistent recovery approach using a causal digital filter. To do that, we first need to express $\Phi_{\mathbf{w},\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})$ in terms of s and v. Using (32)

$$r_{\mathbf{w},\mathbf{s}}(n) = \langle w_n, s_0 \rangle$$

$$= \left\langle \sum_{k \in \mathbb{Z}} \overline{h_{\mathrm{nc}}(n-k)} s_k, s_0 \right\rangle$$

$$= \sum_{k \in \mathbb{Z}} \overline{h_{\mathrm{nc}}(n-k)} r_{\mathbf{s}}(k).$$

In the Fourier domain, this relation becomes

$$\Phi_{\mathbf{w},\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = \overline{\widehat{h}_{\mathrm{nc}}(\mathrm{e}^{-\mathrm{i}\omega})}\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = \frac{\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})}{\Phi_{\mathbf{s},\mathbf{v}}(\mathrm{e}^{\mathrm{i}\omega})} \tag{35}$$

where we used (31). Substituting (35) into (34), and using the fact that $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) = |\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})|^2$, we conclude that the min-max optimal causal filter is given by $\widehat{h}_c(\mathrm{e}^{\mathrm{i}\omega}) = \overline{\Gamma(\mathrm{e}^{-\mathrm{i}\omega})}$, where

$$\Gamma(\mathrm{e}^{\mathrm{i}\omega}) = \frac{1}{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})} \left\{ \frac{\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega}) \mathrm{e}^{\mathrm{i}\omega n}}{\Phi_{\mathbf{s},\mathbf{v}}(\mathrm{e}^{\mathrm{i}\omega})} \right\}_+. \tag{36}$$

This filter, in general, does not lead to a consistent recovery. However, among all causal filters, the expansion coefficients it produces are closest to those of the consistency approach for the worst-case signal x.

An interesting phenomenon occurs when the sampling functions $\{s_n\}$ are orthogonal. This happens, for example, in the SI setting $s_n(t) = s(t - na)$, when $s(t)$ is the rectangular window $\beta^0(t/a)$ or when $s(t)$ is the ideal low-pass filter $\mathrm{sinc}(t/a)$. In this situation, $\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega}) = A$ for some constant $A \neq 0$ and thus $\Gamma(\mathrm{e}^{\mathrm{i}\omega})$ reduces to $\{\mathrm{e}^{\mathrm{i}\omega n}/\Phi_{\mathbf{s},\mathbf{v}}(\mathrm{e}^{\mathrm{i}\omega})\}_+$, which corresponds to a simple truncation of the impulse response of the non-causal filter $\widehat{h}_{\mathrm{nc}}(\mathrm{e}^{\mathrm{i}\omega})$ of (31). The important thing to note, however, is that when the functions $\{s_n\}$ are not orthogonal, $\Phi_{\mathbf{s}}^+(\mathrm{e}^{\mathrm{i}\omega})$ is not constant and simple truncation is no longer optimal.

C. Causal Spline Interpolation

To demonstrate the causal recovery technique discussed above, we next apply it to causal spline interpolation. The class of splines is a popular choice of SI spaces in many image processing applications. A spline $x(t)$ of degree N is a piecewise
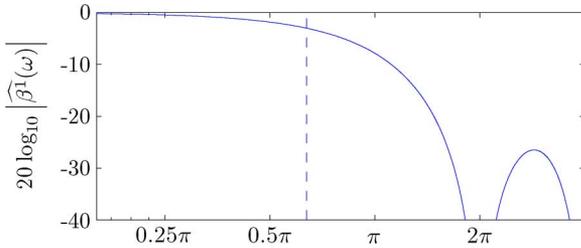
Fig. 8. Frequency response of the filter $\beta^1(t)$.

Fig. 9. Reconstruction of $x(t)$ from its samples c_n using the noncausal filter $\widehat{h}_{\mathrm{nc}}(z)$ of (37), the truncated filter $\widehat{h}_{\mathrm{tr}}(z)$ of (38) and the min-max optimal causal filter $\widehat{h}_{\mathrm{c}}(z)$ of (39).

polynomial with the pieces combined at knots, such that the function is continuously differentiable $N-1$ times. It can be shown that any spline of degree N with knots at the integers can be generated using (7) where $v(t) = \beta^N(t)$ is the B-spline function of degree N.

As a simple example, consider again the SI setting of Fig. 1 where the sampling period is $a = 1$ and the sampling filter $s(t)$ is a B-spline of degree 1. The frequency response of $\beta^1(t)$ is given by $\mathrm{sinc}^2(\omega/(2\pi))$. Therefore, in this case $s(t)$ can be considered a nonideal antialiasing low-pass filter whose cutoff frequency is slightly smaller than the sampling rate $\omega = \pi$, as shown in Fig. 8. We would like to approximate the signal $x(t)$ using (7) with[5] $v(t) = \beta^0(t)$, where the coefficients d_n are obtained by causal processing of the samples c_n.

Using the convolution property (30) of B-splines, we have

$$r_{\mathbf{s},\mathbf{v}}(n) = \langle s_n, v_0 \rangle = \int_{-\infty}^{\infty} \beta^1(t-n)\beta^0(t)\mathrm{d}t = \beta^2(n).$$

As shown in [33], the Z-transform of $\beta^2(n)$ is given by $(1 - z_2 z^{-1})(1 - z_2 z)/(-8 z_2)$, where $z_2 = \sqrt{8} - 3$. Therefore, the noncausal consistent filter (31) is given by

$$\widehat{h}_{\mathrm{nc}}(z) = \frac{1}{\Phi_{\mathbf{s},\mathbf{v}}(z)} = \frac{-8z_2}{(1 - z_2 z^{-1})(1 - z_2 z)} \qquad (37)$$

and corresponds to the impulse response

$$h_{\mathrm{nc}}(n) = \frac{-8z_2}{1 - z_2^2} z_2^{|n|}.$$

A naive solution for obtaining a causal correction filter follows from truncating the impulse response $h_{\mathrm{nc}}(n)$ [10]–[12]. This approach results in the filter $h_{\mathrm{tr}}(n) = h_{\mathrm{nc}}(n)u(n)$, whose Z-transform is

$$\widehat{h}_{\mathrm{tr}}(z) = \frac{-8z_2}{1 - z_2^2} \frac{1}{1 - z_2 z^{-1}}. \qquad (38)$$

Our framework, however, dictates a different strategy. Specifically, recall from Table I that $\Phi_{\mathbf{s}}^+(z) = a_1(1 - z_1 z^{-1})$, where $z_1 = \sqrt{3} - 2$. Substituting this expression into (36) leads to

$$\widehat{h}_{\mathrm{c}}(z) = \frac{-8z_2}{1 - z_1 z^{-1}} \left\{ \frac{1 - z_1 z^{-1}}{(1 - z_2 z^{-1})(1 - z_2 z)} \right\}_+.$$

This expression can be shown to equal

$$\widehat{h}_{\mathrm{c}}(z) = \frac{-8z_2}{1 - z_2^2} \frac{(1 - z_1 z_2) - z_1 \left(1 - z_2^2\right) z^{-1}}{1 - (z_1 + z_2)z^{-1} + z_1 z_2 z^{-2}}. \qquad (39)$$

Fig. 9 compares the above methods in the task of recovering a randomly generated spline $x(t)$ of degree 2. In this example, the noncausal solution (37) attains a signal-to-noise ratio (SNR) of[6] $20\log_{10}(\|x\|/\|x - \widetilde{x}_{\mathrm{nc}}\|) = 8.67$ dB. The truncated filter (38) suffers from a significant degradation attaining an SNR of 7.59 dB. By contrast, our min-max solution (39) results in an SNR of 8.21 dB, which is only slightly worse than the noncausal approach.

VI. CAUSAL STABILITY

In sampling theory and signal representations, it is often desired that the sequence $\{s_n\}_{n\in\mathbb{Z}}$ of sampling functions be a frame or a Riesz basis. This guarantees that small changes in the signal x result in small changes in the sample sequence $\mathbf{c} = \{c_n\}_{n\in\mathbb{Z}}$. Moreover, it ensures that slight perturbations in the samples $\mathbf{c}$ can only be a result of minor perturbations in the signal x. Consequently, recovery of x from $\mathbf{c}$ can be conducted in a stable manner. However, in causal processing tasks, such as extrapolation and causal recovery, one only has access to the past samples $\mathbf{c}_0 = \{c_n\}_{n\le 0}$. Therefore, stability in these applications, is ensured if the past sampling functions $\{s_n\}_{n\le 0}$ form a frame or a Riesz basis for their span. In this section, we derive conditions for the past of a stationary sequence to form a Riesz basis or a frame. We then study the relation between stability and extrapolation.

We begin with a condition on an entire stationary sequence $\mathbf{s} = \{s_n\}_{n\in\mathbb{Z}}$ to be a Bessel-, Riesz-, or frame sequence.

Theorem 6: Let $\mathbf{s}$ be a stationary sequence in a Hilbert space $\mathcal{H}$ with spectral measure $\mu_{\mathbf{s}}$ with Lebesgue decomposition (13). Then

1) $\mathbf{s}$ is a Bessel sequence with bound B if and only if

$$\Psi_{\mathbf{s}} \equiv 0 \text{ and } \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) \le B \text{ a.e. } \omega \in [-\pi, \pi).$$

---

[5]Reconstruction with $\beta^0(t)$ is a shifted-by-1/2 version of zero-order hold recovery, also known as nearest neighbor interpolation in image processing.
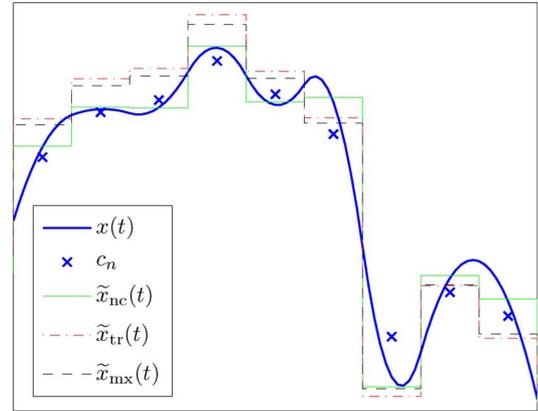
[6]The SNR was computed for a long time segment, only a small portion of which is shown in Fig. 9.
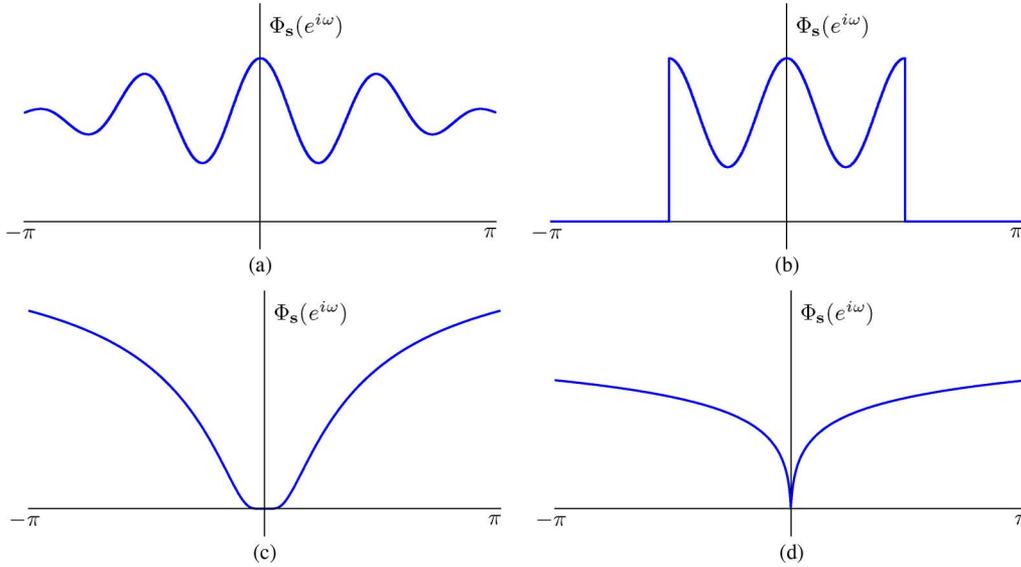
Fig. 10. The spectral density $\Phi_{\mathbf{s}}(e^{i\omega})$ corresponding to various scenarios of stability and perfect extrapolation. (a) Riesz basis. (b) Frame. (c) Bessel sequence with a locally flat zero: $\Phi_{\mathbf{s}}(e^{i\omega}) = \exp\{-1/|\omega|\}$. (d) Bessel sequence with a locally sharp zero: $\Phi_{\mathbf{s}}(e^{i\omega}) = \exp\{-1/|\omega|^{1/4}\}$.

TABLE II
STABILITY, CAUSAL STABILITY, AND PERFECT PREDICTION IN
U-INVARIANT SAMPLING

	Riesz	Frame	Bessel	
			Flat Zero	Sharp Zero
Stability	✓	✓	✗	✗
Causal stability	✓	✗	✗	?
Perfect extrapolation	✗	✓	✓	✗

2) $\mathbf{s}$ is a Riesz basis with bounds A, B if and only if

$$\Psi_{\mathbf{s}} \equiv 0 \text{ and } A \leq \Phi_{\mathbf{s}}(e^{i\omega}) \leq B \text{ a.e. } \omega \in [-\pi, \pi). \quad (40)$$

3) $\mathbf{s}$ is a frame with bounds A, B if and only if

$$\Psi_{\mathbf{s}} \equiv 0 \text{ and } A \leq \Phi_{\mathbf{s}}(e^{i\omega}) \leq B \text{ a.e. } \omega \in [-\pi, \pi) \setminus \Omega \quad (41)$$

where $\Omega = \{\omega \in [-\pi, \pi) : \Phi_{\mathbf{s}}(e^{i\omega}) = 0\}$.

Note that if $\mathbf{s}$ satisfies (40) then it also satisfies (41), reflecting the fact that every Riesz basis is also a frame. Conversely, if the Lebesgue measure of the set Ω, defined under point 3), is positive, then (40) is not satisfied and $\mathbf{s}$ is a frame which is not a Riesz basis.

Proof: See Appendix B. ∎

Remark 1: The above theorem is well known for the case in which the sequence $\mathbf{s}$ is generated by the translation operator given in Example 1 (see, e.g., [16], [19], [24], [34]). Part 2 of the theorem was also derived in [20] for arbitrary stationary sequences. For completeness, we present in Appendix B a unified proof for all three cases.

We now proceed to study causal stability by discussing several cases, as demonstrated in Fig. 10 and summarized in Table II. We assume in the sequel that $\Psi_{\mathbf{s}} \equiv 0$.

a) $\mathbf{s}$ is a Riesz sequence: Assume that $\mathbf{s}$ is a Riesz basis for $\mathcal{S}$, as demonstrated in Fig. 10(a). If one element is removed from a Riesz basis, then it ceases being a Riesz basis. Consequently, the subset $\mathbf{s}_0 = \{s_k\}_{k \leq 0}$ of past sampling functions cannot be a Riesz basis for the whole sampling space $\mathcal{S}$. However, it is still a

Riesz basis for the past sampling space $\mathcal{S}_0$ (see, e.g., [16]) so that causal processing can be pursued in a stable manner. Therefore, if $\mathbf{s}$ is a Riesz sequence then U-invariant sampling is both stable and causally stable. Nevertheless, since $\Phi_{\mathbf{s}}$ does not satisfy (24), perfect extrapolation is impossible in this setting, as can be seen in the first column of Table II.

b) $\mathbf{s}$ is a frame sequence: Assume next that $\mathbf{s}$ is a frame for $\mathcal{S}$, as depicted in Fig. 10(b). In this case, $\Phi_{\mathbf{s}}$ satisfies condition (24) so that the past functions $\mathbf{s}_0$ also span $\mathcal{S}$. However, the next result shows that $\mathbf{s}_0$ *cannot be a frame* for $\mathcal{S}$.

Corollary 7: Let $\mathbf{s} = \{s_n\}_{n \in \mathbb{Z}}$ be a stationary frame sequence for $\mathcal{S}$ with bounds A and B and assume that its spectral density $\Phi_{\mathbf{s}}$ satisfies (24). Then

$$0 < \sum_{k \leq 0} |\langle x, s_k \rangle|^2 \leq B\|x\|^2 \quad (42)$$

for all nonzero $x \in \mathcal{S}$, but there exists no constant $A > 0$ such that $A\|x\|^2 \leq \sum_{k \leq 0} |\langle x, s_k \rangle|^2$ for all $x \in \mathcal{S}$. Thus, $\mathbf{s}_0 = \{s_n\}_{n \leq 0}$ is not a frame for $\mathcal{S}$.

Proof: See Appendix C. ∎

According to (42), the analysis operator $S_{\mathbf{s}_0}$, corresponding to the past sampling functions $\mathbf{s}_0$, is one-to-one. Therefore, it is invertible on its range in the pure algebraic sense. However, since $S_{\mathbf{s}_0}$ is not bounded from below, the corresponding inverse is not bounded, i.e. $S_{\mathbf{s}_0}$ is not invertible if it is viewed as a bounded linear operator $\mathcal{S} \to \ell^2$. This implies that although perfect prediction is theoretically possible, it cannot be achieved by any stable procedure, as summarized in the second column of Table II.

c) $\mathbf{s}$ is a Bessel sequence but not a Riesz basis or a frame: When $\mathbf{s}$ is a Bessel sequence that is not a Riesz basis or a frame, its spectral density $\Phi_{\mathbf{s}}$ is not bounded from below by any positive constant outside the set $\{\omega : \Phi_{\mathbf{s}}(e^{i\omega}) = 0\}$. This scenario can be divided into two subcategories, which differ by the behavior of $\Phi_{\mathbf{s}}$ around its zeros.

1) *Locally flat zero*: $\Phi_{\mathbf{s}}(e^{i\omega})$ satisfies (24). For example, $\Phi_{\mathbf{s}}(e^{i\omega}) = \exp\{-|\omega|^{-\alpha}\}$ with $\alpha \geq 1$, as shown in Fig. 10(c), or $\Phi_{\mathbf{s}}(e^{i\omega})$ being a continuous function, which

vanishes for all $\omega \in \Omega$ where $\Omega \subset [-\pi, \pi)$ is of positive Lebesgue measure.

2) Locally sharp zero: $\Phi_{\mathbf{s}}(e^{i\omega})$ does not satisfy (24). For example, $\Phi_{\mathbf{s}}(e^{i\omega}) = \exp\{-|\omega|^{-\alpha}\}$ with $\alpha < 1$, as demonstrated in Fig. 10(d).

In both cases $\mathbf{s}$ does not constitute a frame for $\mathcal{S}$ since the lower bound in (41) is not satisfied. As we have seen, in the flat zero case condition (24) is satisfied and, therefore, $\mathcal{S}_0 = \mathcal{S}$. Since $\mathbf{s}$ is not a frame, there exist unit-norm functions $x \in \mathcal{S}$ such that $\sum_{k \in \mathbb{Z}} |\langle x, s_k \rangle|^2$ is arbitrarily small. For these functions, of course, $\sum_{k \le 0} |\langle x, s_k \rangle|^2$ is also arbitrarily small. This, in turn, implies that $\bar{\mathbf{s}}_0$ is also not a frame for $\mathcal{S}$, despite the fact that it spans $\mathcal{S}$. Therefore, in this case perfect prediction is possible but it is not a stable operation, as shown in the third column of Table II.

The same argumentation shows that $\mathbf{s}_0$ does not constitute a frame for $\mathcal{S}$ in the sharp-zero case as well. However, in this setting condition (24) is not satisfied so that $\mathcal{S}_0 \neq \mathcal{S}$. Therefore, this does not yet imply that causal stability is lacking. Whether $\mathbf{s}_0$ forms a frame for $\mathcal{S}_0$ or not, remains an open question. All we know is that in this setting, perfect prediction is impossible.

VII. CONCLUSION

In this paper we explored the use of the theory of stationary stochastic sequences for solving sampling problems, which are of a deterministic nature. The theory of stationary sequences seems to have been primarily used in the probability and statistics communities, while barely fertilizing the sampling literature. In this work, we showed that several well-developed tools from this theory can be used to predict future samples of a signal based on its past and causally interpolate a signal from its samples. We also introduced and studied the notion of causal stability, which quantifies the ability to causally predict and interpolate a signal in a stable fashion. Our results indicate that in cases where perfect prediction is theoretically possible, it cannot be done stably.

APPENDIX A
PROOF OF PROPOSITION 2

Any signal x can be written as $x = u + v$ where $u = \mathrm{P}_{\mathcal{S}_0} x$ and $v = \mathrm{P}_{\mathcal{S}_0^{\perp}} x$. The signals $x \in \mathcal{B}$ all satisfy $\mathrm{S}_{\mathbf{s}_0} x = \mathbf{c}_0$ and thus differ from one another only in the second component v. Using this observation, the inner maximization in (19) can be written as

$$|\langle u, s_n \rangle - \widetilde{c}_n|^2 + \max_{v \in \mathcal{C}} \left\{ |\langle v, s_n \rangle|^2 + 2\mathrm{Re}\left\{ \overline{\langle v, s_n \rangle} (\langle u, s_n \rangle - \widetilde{c}_n) \right\} \right\} \quad (43)$$

where $\mathcal{C} = \{ v \in \mathcal{S}_0^{\perp} : \|v\|^2 \le L^2 - \|u\|^2 \}$.

The vector v attaining the maximum must satisfy

$$\mathrm{Re}\left\{ \overline{\langle v, s_n \rangle} (\langle u, s_n \rangle - \widetilde{c}_n) \right\} \ge 0$$

since we can change the sign of v without effecting the constraint. Consequently

$$\max_{v \in \mathcal{C}} \left\{ |\langle v, s_n \rangle|^2 + 2\mathrm{Re}\left\{ \overline{\langle v, s_n \rangle} (\langle u, s_n \rangle - \widetilde{c}_n) \right\} \right\}$$
$$\ge \max_{v \in \mathcal{C}} \left\{ |\langle v, s_n \rangle|^2 \right\} \quad (44)$$

combining (43) and (44), we have that

$$\min_{\widetilde{c}_n} \max_{x \in \mathcal{B}} |\langle x, s_n \rangle - \widetilde{c}_n|^2$$
$$\ge \min_{\widetilde{c}_n} \left\{ |\langle u, s_n \rangle - \widetilde{c}_n|^2 + \max_{v \in \mathcal{C}} \left\{ |\langle v, s_n \rangle|^2 \right\} \right\}$$
$$= \max_{v \in \mathcal{C}} \left\{ |\langle v, s_n \rangle|^2 \right\}$$

where the equality is a result of solving the minimization, which is obtained at

$$\widetilde{c} = \langle u, s_n \rangle = \langle \mathrm{P}_{\mathcal{S}_0} x, s_n \rangle = \langle x, \mathrm{P}_{\mathcal{S}_0} s_n \rangle. \quad (45)$$

We now show that the inequality becomes an equality with $\widetilde{c}$ of (45). Indeed, substituting (45) into (43), we have that

$$\max_{x \in \mathcal{B}} |\langle x, s_n \rangle - \widetilde{c}_n|^2$$
$$= \max_{v \in \mathcal{C}} \left\{ |\langle v, s_n \rangle|^2 + 2\mathrm{Re}\left\{ \overline{\langle v, s_n \rangle} (\langle u, s_n \rangle - \widetilde{c}_n) \right\} \right\}$$
$$= \max_{v \in \mathcal{C}} \left\{ |\langle v, s_n \rangle|^2 \right\}$$

from which the proof follows.

APPENDIX B
PROOF OF THEOREM 6

First, we show that the synthesis operator $\mathrm{R}_{\mathbf{s}}$ given in (10) is unbounded on ℓ^2 if the singular part $\Psi_{\mathbf{s}}$ of the spectral measure is not identically zero. The singular measure $\Psi_{\mathbf{s}}$ is concentrated on a set $\Omega \subset [-\pi, \pi)$ of Lebesgue measure zero. Without loss of generality, we assume that $0 \in \Omega$ and that $\Psi_{\mathbf{s}}(\{0\}) > 0$. For an arbitrary $N \in \mathbb{N}$, we consider the scalar sequence $\mathbf{c}_N = \{c_k\}_{k=-N}^N$ given by $c_0 = 0$ and $c_k = 1/(2|k|)$ for all $1 \le |k| \le N$. One easily verifies that

$$\|\mathbf{c}_N\|_{\ell^2}^2 \le 1 - 1/N < 1.$$

Next, we use the isometric isomorphism between $\mathcal{S}$ and $L^2(\mu_{\mathbf{s}}, \mathbb{T})$ (cf. Lemma 1). Therewith, we get

$$\|\mathrm{R}_{\mathbf{s}} \mathbf{c}_N\|_{\mathcal{H}}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |b_{\mathbf{c}_N}(e^{i\omega})|^2 \left(\Phi_{\mathbf{s}}(e^{i\omega}) d\omega + d\Psi_{\mathbf{s}}(\omega) \right)$$
$$\ge \frac{1}{2\pi} \int_{-\pi}^{\pi} |b_{\mathbf{c}_N}(e^{i\omega})|^2 d\Psi_{\mathbf{s}}(\omega)$$
$$\ge \frac{1}{2\pi} |b_{\mathbf{c}_N}(e^{i0})|^2 \Psi_{\mathbf{s}}(\{0\})$$

where $b_{\mathbf{c}_N}(e^{i\omega}) = \sum_{k=-N}^N c_k e^{ik\omega} = \sum_{k=1}^N \cos(k\omega)/k$. At $\omega = 0$, we obviously have

$$|b_{\mathbf{c}_N}(e^{i0})| = \sum_{k=1}^N \frac{1}{k} \ge \log(N+1)$$

such that

$$\|\mathrm{R}_{\mathbf{s}} \mathbf{c}_N\|_{\mathcal{H}}^2 \ge \frac{1}{2\pi} |\log(N+1)|^2 \Psi_{\mathbf{s}}(\{0\}).$$

This shows that to every $C_0 > 0$ there exists a sequence $\mathbf{c}_N \in \ell^2$ with $\|\mathbf{c}_N\|_{\ell^2} \le 1$ such that $\|\mathrm{R}_{\mathbf{s}} \mathbf{c}_N\|_{\mathcal{H}} > C_0$. Thus $\mathrm{R}_{\mathbf{s}}$ is unbounded on ℓ^2 and it follows that $\mathbf{s}$ cannot be a Bessel sequence

and consequently it also cannot be a Riesz basis or a frame for $\mathcal{S}$ (see, e.g., [16]).

It remains to show that $\mathbf{s}$ is a Bessel sequence, Riesz basis, or frame if the spectral density $\Phi_{\mathbf{s}}$ satisfies the conditions of Theorem 6. It is known (see, e.g., [16]) that $\mathbf{s}$ is a Riesz basis for $\mathcal{S}$ with Riesz bounds A, B if and only if

$$A\|\mathbf{c}\|_{\ell^2}^2 \leq \|\mathrm{R}_{\mathbf{s}}\mathbf{c}\|_{\mathcal{H}}^2 \leq B\|\mathbf{c}\|_{\ell^2}^2 \quad \text{for all } \mathbf{c} \in \ell^2. \tag{46}$$

Assume that $\Phi_{\mathbf{s}}$ satisfies (40). Then it follows from Lemma 1 and from Parseval's theorem that

$$\|\mathrm{R}_{\mathbf{c}}\mathbf{c}\|_{\mathcal{H}}^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|b_{\mathbf{c}}(\mathrm{e}^{\mathrm{i}\omega})\right|^2 \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega$$

$$\leq \frac{B}{2\pi}\int_{-\pi}^{\pi}\left|b_{\mathbf{c}}(\mathrm{e}^{\mathrm{i}\omega})\right|^2 \mathrm{d}\omega = B\|\mathbf{c}\|_{\ell^2}^2$$

for all $\mathbf{c} \in \ell^2$, where $b_{\mathbf{c}}(\mathrm{e}^{\mathrm{i}\omega}) = \sum_{k\in\mathbb{Z}}c_k\mathrm{e}^{\mathrm{i}k\omega}$. In the same way, one gets $\|\mathrm{R}_{\mathbf{c}}\mathbf{c}\|_{\mathcal{H}}^2 \geq A\|\mathbf{c}\|_{\ell^2}^2$ which shows that $\mathbf{s}$ is a Riesz basis for $\mathcal{S}$. Conversely, assume that (46) holds. Then Parseval's theorem and Lemma 1 give

$$\frac{A}{2\pi}\int_{-\pi}^{\pi}\left|b_{\mathbf{c}}(\mathrm{e}^{\mathrm{i}\omega})\right|^2 \mathrm{d}\omega \leq \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|b_{\mathbf{c}}(\mathrm{e}^{\mathrm{i}\omega})\right|^2 \Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega$$

$$\leq \frac{B}{2\pi}\int_{-\pi}^{\pi}\left|b_{\mathbf{c}}(\mathrm{e}^{\mathrm{i}\omega})\right|^2 \mathrm{d}\omega.$$

Since this holds for all $b_{\mathbf{c}} \in L^2(\mathbb{T})$, (40) follows.

The proof of (41) is similar to the proof of [16, Th. 7.2.3] for the special case of the translation operator. Here, only the isometric relation of Lemma 1 has to be used at the appropriate points. Therefore, the details are omitted.

APPENDIX C
PROOF OF COROLLARY 7

The upper bound in (42) follows from

$$\sum_{k\leq 0}\left|\langle x, s_k\rangle\right|^2 \leq \sum_{k\in\mathbb{Z}}\left|\langle x, s_k\rangle\right|^2 \leq B\|x\|^2.$$

To prove the lower bound in (42) we show first that to every nonzero $x \in \mathcal{S}$ there always exists at least one index $k_0 \leq 0$ such that $c_{k_0} = \langle x, s_{k_0}\rangle > 0$. Indeed, assume to the contrary that $c_k = \langle x, s_k\rangle = 0$ for all $k = 0, -1, -2, \ldots$. Then the min-max optimal predictor (20) of the future sample gives $\widetilde{c}_1 = 0$. Since $\Phi_{\mathbf{s}}$ satisfies (24), Theorem 4 implies that this prediction is perfect, i.e., that $c_1 = \widetilde{c}_1 = 0$. By induction, it follows that $c_k = 0$ for all $k = 1, 2, 3, \ldots$ and consequently $\sum_{k\in\mathbb{Z}}\left|\langle x, s_k\rangle\right|^2 = 0$. However, this contradicts the lower bound in (9) which has to be satisfied because $\mathbf{s}$ is assumed to be a frame for $\mathcal{S}$. Consequently

$$\sum_{k\leq 0}\left|\langle x, s_k\rangle\right|^2 \geq \max_{k\leq 0}\left|\langle x, s_k\rangle\right|^2 \geq \left|c_{k_0}\right|^2 > 0$$

which proves the lower bound in (42).

To prove the last statement, it suffices to show that to every $\epsilon > 0$ there exists an $x \in \mathcal{S}$ with $\|x\|_{\mathcal{H}} \geq 1$ such that $\sum_{k\leq 0}\left|\langle x, s_k\rangle_{\mathcal{H}}\right|^2 < \epsilon$. Let $\Omega \subset [-\pi, \pi)$ be an open interval on which $\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega}) > 0$ for all $\omega \in \Omega$. Without loss of generality we can assume that $\Omega = (-\Omega_0, \Omega_0)$ for some $0 < \Omega_0 < \pi$. For an arbitrary positive integer $N \in \mathbb{N}$ define the function

$$\widetilde{x}_N(\mathrm{e}^{\mathrm{i}\omega}) = \begin{cases} \sqrt{\frac{\pi B}{\Omega_0}}\dfrac{\mathrm{e}^{\mathrm{i}N\omega}}{\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})} & \text{if } \omega \in \Omega \\ 0 & \text{if } \omega \notin \Omega. \end{cases}$$

Using (41), one easily verifies that $1 \leq \|\widetilde{x}_N\|_{\mu_{\mathbf{s}}} \leq \sqrt{B/A}$. Consequently $\widetilde{x}_N \in L^2(\mu_{\mathbf{s}}, \mathbb{T})$ and therefore there exists an $x_N \in \mathcal{S}$ with $\|x_N\|_{\mathcal{H}} = \|\widetilde{x}_N\|_{\mu_{\mathbf{s}}}$ (cf. Section III-B). For the generalized samples of x_N, one obtains

$$\langle x_N, s_k\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi}\widetilde{x}_N(\mathrm{e}^{\mathrm{i}\omega})\mathrm{e}^{-\mathrm{i}k\omega}\Phi_{\mathbf{s}}(\mathrm{e}^{\mathrm{i}\omega})\mathrm{d}\omega$$

$$= \sqrt{\frac{B}{\pi\Omega_0}}\frac{\sin\left([N-k]\Omega_0\right)}{N-k}$$

and therewith, one obtains

$$\sum_{k\leq 0}\left|\langle x_N, s_k\rangle\right|^2 \leq \frac{B}{\pi\Omega_0}\sum_{k=N}^{\infty}\frac{1}{k^2} \leq \frac{B}{\pi\Omega_0}\frac{1}{N-1}.$$

Consequently, for a sufficiently large N one gets $\sum_{k\leq 0}\left|\langle x_N, s_k\rangle\right|^2 < \epsilon$.

REFERENCES

[1] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.

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

[3] A. Aldroubi and M. Unser, "Families of wavelet transforms in connection with Shannon sampling theory and the Gabor transform," in *Wavelets: A Tutorial in Theory and Applications*, C. K. Chui, Ed. San Diego, CA: Academic, 1992.

[4] T. Blu and M. Unser, "Quantitative Fourier analysis of approximation techniques-part I: Interpolators and projectors," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2783–2795, Oct. 1999.

[5] Y. C. Eldar, "Sampling without input constraints: Consistent reconstruction in arbitrary spaces," in *Sampling, Wavelets and Tomography*, J. J. Benedetto and A. I. Zayed, Eds. Boston, MA: Birkhäuser, 2004.

[6] J. Kybic, T. Blu, and M. Unser, "Generalized sampling: A variational approach-part I: Theory," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1965–1976, Aug. 2002.

[7] Y. C. Eldar and M. Unser, "Nonideal sampling and interpolation from noisy observations in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2636–2651, Jul. 2006.

[8] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2146–2159, Jun. 2006.

[9] T. Michaeli and Y. C. Eldar, "Optimization techniques in modern sampling theory," in *Convex Optimization in Signal Processing and Communications*, Y. C. Eldar and D. Palomar, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[10] T. J. Lim and M. D. Macleod, "On-line interpolation using spline functions," *IEEE Signal Process. Lett.*, vol. 3, no. 5, pp. 144–146, May 1996.

[11] D. Petrinovic, "Causal cubic splines: Formulations, interpolation properties and implementations," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5442–5453, Nov. 2008.

[12] D. Petrinovic, "Continuous time domain properties of causal cubic splines," *Signal Process.*, vol. 89, no. 10, pp. 1941–1958, Oct. 2009.

[13] M. Nagahara and Y. Yamamoto, "Causal spline interpolation by H^{∞} optimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, Apr. 2007, vol. 3, pp. 1469–1472.

[14] T. Blu, P. Thévenaz, and M. Unser, "High-quality causal interpolation for online unidimensional signal processing," in *Proc. 11th Eur. Signal Process. Conf. (EUSIPCO'04)*, Vienna, Austria, Sep. 2004, pp. 1417–1420.

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

[16] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA: Birkhäuser, 2003.

[17] E. Matusiak, T. Michaeli, and Y. C. Eldar, "Noninvertible Gabor transforms," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2597–2612, May 2010.

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

[19] P. Wojtaszczyk, *A Mathematical Introduction to Wavelets*. Cambridge, MA: Cambridge Univ. Press, 1999.

[20] A. Aldroubi, "Oblique projections in atomic spaces," *Proc. Amer. Math. Soc.*, vol. 124, no. 7, pp. 2051–2060, 1996.

[21] Y. A. Rozanov, *Stationary Random Processes*. San Francisco, CA: Holden-Day, 1967.

[22] J. Lamperti, *Stochastic Processes*. New York: Springer-Verlag, 1977.

[23] A. N. Shiryayev, *Probability*, 2nd ed. New York: Springer-Verlag, 1996.

[24] O. Christensen and Y. C. Eldar, "Oblique dual frames and shift-invariant spaces," *Appl. Computat. Harmon. Anal.*, vol. 17, no. 1, pp. 48–68, Jul. 2004.

[25] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 2002.

[26] T. G. Dvorkind, H. Kirshner, Y. C. Eldar, and M. Porat, "Minimax approximation of representation coefficients from generalized samples," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4430–4443, Sep. 2007.

[27] G. Szegö, "Beiträge zur Theorie the Toeplitzen Formen (Erste Mitteilung)," *Math. Zeit.*, vol. 6, pp. 167–202, 1920.

[28] J. B. Garnett, *Bounded Analytic Functions*. New York: Academic, 1981.

[29] P. Fatou, "Séries trigonométriques et séries de Taylor," *Acta Math.*, vol. 30, pp. 335–400, 1906.

[30] J. R. Partington, *Interpolation, Identification, and Sampling*. Oxford, U.K.: Clarendon, 1997.

[31] J. L. Brown, "On the prediction of a band-limited signal from past samples," *Proc. IEEE*, vol. 74, no. 11, pp. 1596–1698, Nov. 1986.

[32] A. Papoulis, "A note on the predictiability of band-limited processes," *Proc. IEEE*, vol. 73, no. 8, pp. 1332–1333, Aug. 1985.

[33] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part II-Efficiency design and applications," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 834–848, Feb. 1993.

[34] J. Benedetto and S. Li, "The theory of multiresolution analysis frames and applications to filter banks," *Appl. Comp. Harm. Anal.*, vol. 5, no. 4, pp. 389–427, 1998.

Tomer Michaeli received the B.Sc. degree in electrical engineering in 2004 (*summa cum laude*) from the Technion-Israel Institute of Technology, Haifa, Israel.

He is currently pursuing the Ph.D. degree in electrical engineering at the Technion. From 2000 to 2008, he was a Research Engineer at RAFAEL Research Laboratories, Israel Ministry of Defense, Haifa. In 2008, he was awarded the Andrew and Erna Finci Viterbi Fellowship and in 2010 he was a Jacobs-QUALCOMM Fellow. His research interests include statistical signal processing, estimation theory, sampling theory, image processing, and computer vision.

Volker Pohl (M'10) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the Technische Universität Berlin, Germany, in 2000 and 2006, respectively.

From 2000 to 2007, he was a Research Associate with the Department of Broadband Mobile Communications Networks of the Heinrich-Hertz-Institut für Nachrichtentechnik Berlin, Germany, and at the Institute for Communications Systems, Technische Universität Berlin. Between 2007 to 2009, he was a Postdoctoral Fellow with the Department of Electrical Engineering, Technion-Israel Institute of Technology. He is currently with the Institute for Theoretical Information Technology, Technische Universität München, Germany.

Yonina C. Eldar (S'98–M'02–SM'07) received the B.Sc. degree in physics in 1995 and the B.Sc. degree in electrical engineering in 1996 both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in electrical engineering and computer science in 2001 from the Massachusetts Institute of Technology (MIT), Cambridge.

From January 2002 to July 2002, she was a Postdoctoral Fellow with the Digital Signal Processing Group, MIT. She is currently a Professor with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa. She is also a Research Affiliate with the Research Laboratory of Electronics, MIT, and a visiting Professor with Stanford University, Stanford, CA. Her research interests are in the general areas of sampling theory, statistical signal processing, optimization methods, and their applications to biology and optics.

Dr. Eldar was in the Program for Outstanding Students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel and David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009 the Technion's Award for Excellence in Teaching, and in 2010 the Michael Bruno Memorial Award from the Rothschild Foundation. She is a member of the IEEE Signal Processing Theory and Methods technical committee and the Bio Imaging Signal Processing Technical Committee, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the EURASIP *Journal of Signal Processing*, the SIAM *Journal on Matrix Analysis and Applications*, and the SIAM *Journal on Imaging Sciences*, and on the Editorial Board of *Foundations and Trends in Signal Processing*.