CCIT Report #582
April 2006

Minimax Approximation of Representation Coefficients From Generalized Samples

Tsvi G. Dvorkind, Hagai Kirshner, Yonina C. Eldar and Moshe Porat

Abstract

Many sources of information are of analogue or continuous-time nature. However, digital signal processing applications rely on discrete data. We consider the problem of approximating L_2 inner products, *i.e.,* representation coefficients of a continuous-time signal, from its generalized samples. Taking a robust approach, we process these generalized samples in a minimax optimal sense. Specifically, for the worst possible signal, we find the best approximation of the desired representation coefficients by proper processing the given sample sequence. We then extend our results to criteria which incorporate smoothness constraints on the unknown function. Finally we compare our methods with the piecewise-constant approximation technique, commonly used for this problem, and discuss the possible improvements by the suggested schemes.

I. Introduction

Signal processing applications are concerned mainly with digital data, although the origin of many sources of information is analogue. This is the case for speech and audio, optics, radar, sonar, biomedical signals and more. In many cases, analysis of a continuous-time signal $\mathbf{x}(t)$ is obtained by evaluating L_2 inner-products $\langle \mathbf{w}_n(t), \mathbf{x}(t) \rangle_{L_2}$ for a set of functions $\{\mathbf{w}_n(t)\}$. For example, one may calculate a Gabor [1] or wavelet [2] representation of a signal. Both are based on finding the signal's representation coefficients; namely performing consecutive L_2 inner products with a set of analysis functions.

Typically, the analysis functions $\{\mathbf{w}_n(t)\}$ are analytically known. On the other hand, in many applications of digital signal processing, there is no knowledge of the continuous-time signal $\mathbf{x}(t)$, but only of its sample sequence. Our problem is to approximate the required L_2 inner-products, by proper processing of the available samples.

In some cases the sampled version of a signal is sufficient to calculate the original function. The classical Whittaker-Shannon sampling theorem is a well known example of the latter; see also [3], [4] for additional shift invariant settings. If the analog input can be determined from the sample sequence, then the required representation coefficients can be calculated as well. Our main focus here is on situations where the knowledge of the continuous-time function is incomplete, such that approximations of the continuous-time inner products must be performed instead.

As an example of facing incomplete knowledge, we mention an initialization problem in wavelet analysis. To initialize the pyramid algorithm [5] one must have the representation coefficients of the continuous time function

This work was supported in part by the HASSIP Research Program HPRN-CT-2002-00285 of the European Commission, and by the Ollendorff Minerva Centre. Minerva is funded through the BMBF.

$\mathbf{x}(t)$, for the initial scale. Unfortunately, those representation coefficients are unavailable, and we only have the samples of $\mathbf{x}(t)$, obtained at the output of some anti-aliasing filter. A common practice in wavelet analysis is to assume that the available samples are the required representation coefficients. This false assumption is also known in the literature as the 'wavelet crime' [6]. In [7] the authors address this problem by suggesting a digital filter to process the available sample sequence, prior to applying the pyramid algorithm. The criteria for designing this correction stage is set to be a minimax approximation of a filtered version of the required representation coefficients. In fact, it can be shown that their result is compatible with a special case of our derivations, presented in Section V-B.

A common approach to cope with incomplete knowledge of the continuous-time function $\mathbf{x}(t)$ is to first interpolate the given sample sequence, thus obtaining an approximation $\hat{\mathbf{x}}(t)$ (see for example [8]). The required calculations can then be performed directly on $\hat{\mathbf{x}}(t)$. The latter method implies that the original signal is first approximated within the span of some synthesis function set, but the best choice of those synthesis functions is not always clear. For rigorous analysis, see [9].

The problem of approximating a continuous-time inner product by properly processing uniform and ideal samples was considered in [10], [11]; in order to approximate a single representation coefficient $\langle \mathbf{w}(t), \mathbf{x}(t) \rangle_{L_2}$, it was suggested to calculate an l_2 inner product $\sum_n b[n]\mathbf{x}(nT)$ instead. Assuming that the input $\mathbf{x}(t)$ is a smooth order one Sobolev [12] function, an upper error bound for the approximation error $|\langle \mathbf{w}(t), \mathbf{x}(t) \rangle_{L_2} - \sum_n b[n]\mathbf{x}(nT)|$ was derived, and the sequence b was determined by minimizing that upper bound.

In practice, however, ideal sampling is impossible to implement. A more practical model may consider generalized samples [4], [13]–[16] instead. Generalized samples of a continuous-time signal are represented as the inner products of this signal with a set of sampling functions $\{\mathbf{s}_n(t)\}$ associated with the acquisition device. Thus, the n'th sample can be written as $c[n] = \langle \mathbf{s}_n(t), \mathbf{x}(t) \rangle_{L_2}$. This sampling model is general enough to describe any linear and bounded acquisition device (Riesz representation theorem [17], [18]). As an example, consider an analog to digital converter which performs pre-filtering prior to sampling, as shown in Figure 1. In such a setting, the sampling vectors $\{\mathbf{s}_n(t) = \mathbf{s}(t - nT)\}$ are shifted and mirrored versions of the impulse response of the pre-filter [13].

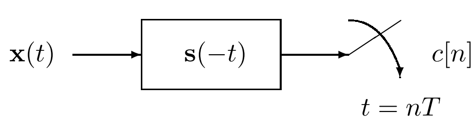


Fig. 1. Filtering with impulse response $\mathbf{s}(-t)$ followed by ideal sampling. The sampling vectors are $\{\mathbf{s}(t - nT)\}$.

Often, the fact that the samples are non-ideal is simply ignored. Assuming that the sample value is close to the mean value of the signal, within some interval of length T, a common practice (referred herein as the 'sum' approach) is to approximate the L_2 inner product by a Riemann-type sum: $\langle \mathbf{w}(t), \mathbf{x}(t) \rangle_{L_2} \approx T \sum_n c[n]\overline{\mathbf{w}(nT)}$, where the over-line stands for complex conjugate. To determine the quality of this approximation, one must assume some regularity conditions on the functions involved. In [19] the author derives convergence rates for such approximations assuming ideal uniform sampling.

In this paper we take a different approach, which is similar in spirit to the works in [7], [16] and a direct generalization of [10]. Given the generalized samples, we approximate the desired representation coefficients $\{\langle \mathbf{w}_n(t), \mathbf{x}(t)\rangle_{L_2}\}$ in a minimax optimal sense. It turns out that the solutions to our robust objectives can also be interpreted as an interpolation of the given samples, followed by applying the analysis functions $\{\mathbf{w}_n(t)\}$ to the interpolation $\hat{\mathbf{x}}(t)$. The nice thing is that the interpolation stage stems naturally from the setup of the problem, rather than being pre-specified arbitrarily. Additionally, the division of the algorithm into interpolation and analysis stages is more of conceptual rather than practical; Both stages can be performed simultaneously, by digital processing of the available samples.

Our results extend [10] in several ways. First, by considering generalized samples (as opposed to the ideal sampling assumption of [10]) our derivations are applicable to practical acquisition devices. Second, if there is prior knowledge that the generalized samples have been obtained from a smooth function, then we show how to incorporate that constraint into the proposed robust solution. Third, our derivations are applicable to a series of representation coefficients. Finally, we analyze the performance of the suggested approach, giving sufficient conditions for it to outperform the naive sum approximation.

The outline of this paper is as follows. In Section II we describe the mathematical preliminaries. Section III discusses situations where one can obtain perfect evaluation of the required L_2 inner products, and establishes a minimax approximation criterion when this is not the case. The latter minimax objective is solved in Section IV. In Section V we consider the problem of incorporating smoothness constraints. Specifically, if there is prior knowledge of the input to be smooth, then we show how to alter the minimax solution by recasting the problem into a proper Sobolev space, presenting [10] as a special setting of our derivations. Section VI discusses the relations between the errors due to the suggested minimax approach and the standard sum approximation method. We show the possible gain in performance by the proposed method and derive sufficient conditions for the minimax solution to dominate the standard approach. Finally, in Section VII, we conclude with several simulations.

II. Mathematical Preliminaries

We denote continuous-time signals by bold lowercase letters, omitting the time dependence, when possible. The elements of a sequence $c \in l_2$ will be written with square brackets, *e.g.* $c[n]$. $\mathbf{X}^F(\omega) = \int \mathbf{x}(t)e^{-j\omega t}dt$ is the continuous time Fourier transform of $\mathbf{x}$ and $C^f(\omega) = \sum_n c[n]e^{-j\omega n}$ is the (2π periodic) *discrete-time Fourier transform* (DTFT) of the sequence c. S_T stands for the ideal sampling operator, such that the n'th element of $S_T\mathbf{w}$ is $\mathbf{w}(nT)$. The operator $P_\mathcal{A}$ represents the orthogonal projection onto a closed subspace $\mathcal{A}$, and $\mathcal{A}^\perp$ is the orthogonal complement of $\mathcal{A}$. The *Moore-Penrose pseudo inverse* [20] and the adjoint of a bounded transformation T are written as $T^\dagger$ and T^*, respectively. $\Re$ stands for the real part. We will denote by $\mathcal{S}$ the sampling space, which is the closure of *span* $\{\mathbf{s}_n\}$. Similarly, $\mathcal{W}$ is the analysis space, obtained by the closure of *span* $\{\mathbf{w}_n\}$.

Inner products and norms are denoted by $\langle a, b\rangle_{\mathcal{H}}$ and $\|a\|_{\mathcal{H}}$, respectively. Here, $\mathcal{H}$ stands for the Hilbert space involved. Usually, we will consider $\mathcal{H}$ to be L_2, l_2 or the order-one Sobolev space W_2^1, which will be discussed in detail in Section V. When the derivations are general enough to describe inner products and norms within L_2 or for that matter any Hilbert space, we will omit the space subscript from the notations, *i.e.,* $\langle f, g\rangle$ or $\|f\|$. All inner products are linear with respect to the second argument. For example, $\langle \mathbf{x}, \mathbf{y}\rangle_{L_2} = \int_{-\infty}^{\infty} \overline{\mathbf{x}(t)}\mathbf{y}(t)dt$.

An easy way to describe linear combinations and inner products is by utilizing set transformations. A set transformation $V : l_2 \rightarrow \mathcal{H}$ corresponding to frame vectors $\{\mathbf{v}_n(t)\}$ is defined by $Va = \sum_n a[n]\mathbf{v}_n(t)$ for all $a \in l_2$ [21]. From the definition of the adjoint, if $a = V^*\mathbf{y}$, then $a[n] = \langle \mathbf{v}_n, \mathbf{y}\rangle$. Denoting by S (W) the set transformation corresponding to the vectors $\{\mathbf{s}_n\}$ ($\{\mathbf{w}_n\}$), the generalized samples $c[n] = \langle \mathbf{s}_n, \mathbf{x}\rangle_{L_2}$ can be written as $c = S^*\mathbf{x}$, and the desired representation coefficients $q[n] = \langle \mathbf{w}_n, \mathbf{x}\rangle_{L_2}$ by $q = W^*\mathbf{x}$. If A is a set transformation with a closed range $\mathcal{A}$, then the orthogonal projection onto $\mathcal{A}$ can be written as $P_{\mathcal{A}} = A(A^*A)^{\dagger}A^*$.

To handle well posed problems, we assume that the sample sequence c and the desired representation coefficients q have finite energy, *i.e.,* $c, q \in l_2$. This will also ensure that if any bounded transformation $G : l_2 \rightarrow l_2$ is applied to the generalized samples c, the error sequence $q - G(c)$ is in l_2 as well. Accordingly, criteria which consider the l_2 norm of the error sequence are well defined. One way to enforce $c, q \in l_2$ is to require that $\{\mathbf{s}_n\}$ and $\{\mathbf{w}_n\}$ form frames [21] for $\mathcal{S}$ and $\mathcal{W}$, respectively, which is precisely what we assume in this work.

III. Problem Formulation

The given data are the generalized samples c of a continuous-time function $\mathbf{x}(t)$, modelled by

$$c[n] = \langle \mathbf{s}_n(t), \mathbf{x}(t)\rangle_{L_2}. \tag{1}$$

We wish to evaluate a set of continuous-time inner products q satisfying

$$q[n] = \langle \mathbf{w}_n, \mathbf{x}\rangle_{L_2}. \tag{2}$$

The analysis functions $\{\mathbf{w}_n\}$ are analytically known. Unfortunately, the input $\mathbf{x}$ is unknown. Our goal is to approximate the required representation coefficients q by proper processing the sample sequence c.

A natural question to be first considered is whether there is an unavoidable error due to our partial knowledge of $\mathbf{x}(t)$, or can we evaluate **exactly** the required L_2 inner products, based on the generalized samples. The following theorem addresses this preliminary question.

Theorem 1: Let $\mathbf{x}$ be an arbitrary function, satisfying $c = S^*\mathbf{x}$. It is possible to obtain the required representation coefficients $q = W^*\mathbf{x}$, by proper processing of the sample sequence c, if and only if $\mathcal{W} \subseteq \mathcal{S}$.

Proof: See Appendix I. □

In some cases, we may have additional prior knowledge on $\mathbf{x}$, such that not all signals in L_2 should be considered. By restricting our attention to a proper subgroup, it is possible to obtain a zero error, even if $\mathcal{W} \nsubseteq \mathcal{S}$. This is true whenever the knowledge of $\mathbf{x}$ allows us to determine a bijection (injective and surjective transformation) between $\mathbf{x}(t)$ and its samples. To illustrate the last point, suppose that $\mathbf{x} \in \mathcal{A}$, where $\mathcal{A}$ is a closed subspace of L_2 satisfying the direct sum condition $L_2 = \mathcal{A} \oplus \mathcal{S}^{\perp}$ (*i.e.,* L_2 can be described by the sum set $\{\mathbf{a}+\mathbf{v};\ \mathbf{a} \in \mathcal{A}, \mathbf{v} \in \mathcal{S}^{\perp}\}$ with the property $\mathcal{A} \cap \mathcal{S}^{\perp} = \{0\}$). Then, we can perfectly reconstruct $\mathbf{x}$ from its generalized samples by

$$\mathbf{x} = A(S^*A)^{\dagger}c, \tag{3}$$

where A is any bounded set transformation with range $\mathcal{A}$ [16]. As a result, we can also perfectly evaluate the coefficients $q = W^*\mathbf{x}$ as

$$q = W^*A(S^*A)^{\dagger}c. \tag{4}$$

Another example in which bijection between the signal and its generalized samples exists is the finite innovation case is considered in [22].

Nevertheless, in the general case the condition $\mathcal{W} \subseteq \mathcal{S}$ may not be satisfied, or there may be no prior knowledge of $\mathbf{x}(t)$. Thus, the coefficients $W^*\mathbf{x}$ cannot be computed exactly and instead must be approximated from the given samples c. A common practice is to perform the l_2 sum approximation:

$$\langle \mathbf{w}(t), \mathbf{x}(t) \rangle_{L_2} \approx T \sum_n c[n]\overline{\mathbf{w}(nT)}, \tag{5}$$

where one implicitly assumes that the generalized samples of $\mathbf{x}$ are close to the mean value of the input signal, within an interval of length T, and similarly the ideal samples of $\mathbf{w}$ are a suitable choice for representing the analysis function. However, in the general case there are no optimality claims for this approach and its analysis must rely on regularity conditions for the functions involved [19].

Alternatively, we may approximate the continuous-time inner products by choosing a sequence d which minimizes the squared norm of the error vector $e = W^*\mathbf{x} - d$. Since $\mathbf{x}$ satisfies $c = S^*\mathbf{x}$, by decomposing $\mathbf{x}$ along $\mathcal{S}$ and $\mathcal{S}^{\perp}$ the error vector can be written as

$$e = W^*S(S^*S)^{\dagger}c + W^*P_{\mathcal{S}^{\perp}}\mathbf{x} - d, \tag{6}$$

where we used $P_{\mathcal{S}}\mathbf{x} = S(S^*S)^{\dagger}c$. This leads to the following objective

$$\min_d \left\| W^*S(S^*S)^{\dagger}c + W^*P_{\mathcal{S}^{\perp}}\mathbf{x} - d \right\|_{l_2}^2. \tag{7}$$

It so happens that the solution of (7) depends on $P_{\mathcal{S}^{\perp}}\mathbf{x}$, which is unknown. To eliminate the dependence on $\mathbf{x}$, one may instead consider a robust approach, where the sequence d is optimized for the worst possible input $\mathbf{x}$. Valid inputs must be consistent with the known samples, *i.e.,* must satisfy $c = S^*\mathbf{x}$. Additionally, if the norm of the input is not bounded, then neither is the norm of the error. Hence, to define a well posed problem, we will assume that $\mathbf{x}$ is norm bounded by some positive constant L. This leads to the minimax objective

$$\min_d \max_{\|x\| \leq L, c = S^*\mathbf{x}} \|W^*\mathbf{x} - d\|_{l_2}^2. \tag{8}$$

In the next sections, we derive a solution for d, and compare its performance with the standard approach, given in (5). As we shown, d does not depend on L, so that exact value of the signal's energy is not required.

IV. Minimax Approximation

The minimax problem of (8) is closely related to the generalized sampling problem considered in [16, Theorem 3]. Relying on results obtained in that context leads to the following theorem.

Theorem 2: Consider the problem

$$\min_{d} \max_{c=S^*\mathbf{x}, \|\mathbf{x}\|\leq L} \|W^*\mathbf{x}-d\|_{l_2}^2 ,$$

where W and S are bounded set transformations with range $\mathcal{W}$ and $\mathcal{S}$, respectively. The (unique) solution is

$$d = W^*S(S^*S)^{\dagger}c. \tag{9}$$

Before going into the details of the proof, note that we have not described the exact Hilbert space in which the bound $\|\mathbf{x}\|_{L_2} \leq L$ and the inner products $S^*\mathbf{x}, W^*\mathbf{x}$ are calculated, since the derivations are general enough to be applicable to any Hilbert space. In Section V we will show how smoothness constraints can be incorporated by applying Theorem 2 in different Hilbert spaces. Additionally, the upper norm bound L is not expressed in the solution (9). Thus, one only has to be sure that the signal has a finite norm, while its exact value is irrelevant for the computation of d. The value of L will be used in Section VI though for analyzing the performance of the proposed algorithm.

Proof: First we note that any $\mathbf{x}$ in the set $\mathcal{D} = \{\mathbf{x} \mid S^*\mathbf{x} = c,\ \|\mathbf{x}\| \leq L\}$ is of the form $\mathbf{x} = S\left(S^*S\right)^{\dagger} c + \mathbf{v}$ for some $\mathbf{v} \in \mathcal{G}$ where

$$\mathcal{G} \triangleq \left\{\mathbf{v} \mid \mathbf{v} \in \mathcal{S}^{\perp}, \|\mathbf{v}\| \leq L'\right\}, \tag{10}$$

and

$$L' = \sqrt{L^2 - \|S(S^*S)^{\dagger}c\|^2}. \tag{11}$$

Thus,

$$\begin{aligned} \max_{\mathbf{x}\in\mathcal{D}} \|W^*\mathbf{x}-d\|_{l_2}^2 &= \\ &= \max_{\mathbf{v}\in\mathcal{G}} \left\|W^*S(S^*S)^{\dagger}c - d + W^*\mathbf{v}\right\|_{l_2}^2 \\ &= \max_{\mathbf{v}\in\mathcal{G}} \|a_d + W^*\mathbf{v}\|_{l_2}^2 \\ &= \max_{\mathbf{v}\in\mathcal{G}} \left(\|a_d\|_{l_2}^2 + 2\Re\left\{\langle a_d, W^*\mathbf{v}\rangle_{l_2}\right\} + \|W^*\mathbf{v}\|_{l_2}^2\right), \end{aligned} \tag{12}$$

where we defined $a_d \triangleq W^*S(S^*S)^{\dagger}c - d$. As a result, the maximum in (12) is achieved when

$$\Re\left\{\langle a_d, W^*\mathbf{v}\rangle_{l_2}\right\} = \left|\langle a_d, W^*\mathbf{v}\rangle_{l_2}\right|. \tag{13}$$

Indeed, let $\mathbf{v} \in \mathcal{G}$ be the vector for which the maximum is achieved. If $\langle a_d, W^*\mathbf{v}\rangle_{l_2} = 0$ then (13) is trivially true. Otherwise, we can define

$$\mathbf{v}_2 \triangleq \frac{\langle W^*\mathbf{v}, a_d\rangle_{l_2}}{|\langle W^*\mathbf{v}, a_d\rangle_{l_2}|}\mathbf{v}. \tag{14}$$

Clearly, $\|\mathbf{v}\| = \|\mathbf{v}_2\|$ and $\mathbf{v}_2 \in \mathcal{G}$. In addition, $\|W^*\mathbf{v}\|_{l_2} = \|W^*\mathbf{v}_2\|_{l_2}$ and $\langle a_d, W^*\mathbf{v}_2\rangle_{l_2} = |\langle a_d, W^*\mathbf{v}\rangle_{l_2}|$ so that the objective in (12) at $\mathbf{v}_2$ is larger than the objective at $\mathbf{v}$ unless (13) is satisfied.

Combining (13) and (12), our problem becomes

$$\min_{d} \max_{\mathbf{v}\in\mathcal{G}} \left(\|a_d\|_{l_2}^2 + 2\left|\langle a_d, W^*\mathbf{v}\rangle_{l_2}\right| + \|W^*\mathbf{v}\|_{l_2}^2 \right). \tag{15}$$

Denoting the optimal objective value by A, and replacing the order of minimization and maximization, we get

$$\begin{aligned} A &\geq \max_{\mathbf{v}\in\mathcal{G}} \min_{d} \left(\|a_d\|_{l_2}^2 + 2\left|\langle a_d, W^*\mathbf{v}\rangle_{l_2}\right| + \|W^*\mathbf{v}\|_{l_2}^2 \right) \\ &= \max_{\mathbf{v}\in\mathcal{G}} \|W^*\mathbf{v}\|_{l_2}^2, \end{aligned} \tag{16}$$

where we used the fact that $\|a_d\|_{l_2}^2 + 2\left|\langle a_d, W^*\mathbf{v}\rangle_{l_2}\right| \geq 0$, with equality for $a_d = 0$, or

$$d = W^*S(S^*S)^{\dagger}c.$$

Thus, for any choice of d,

$$\min_{d} \max_{\mathbf{v}\in\mathcal{G}} \|a_d - W^*\mathbf{v}\|_{l_2}^2 \geq \max_{\mathbf{v}\in\mathcal{G}} \|W^*\mathbf{v}\|_{l_2}^2. \tag{17}$$

The proof then follows from the fact that d given by (9) achieves the lower bound (17). Uniqueness of d follows from (15), as the optimal solution must satisfy $a_d = 0$. □

Note that (9) resembles the solution of the Wiener-Hopf equations, where the Gramian matrix of the autocorrelations is first inverted (pseudo-inverted), and the cross-correlation Gramian matrix is then applied. Another interesting interpretation of (9) is obtained by presenting $P_{\mathcal{S}}\mathbf{x} = S(S^*S)^{\dagger}c$. This leads to the following corollary.

Corollary 1: The solution (9) can be written as

$$d = W^*P_{\mathcal{S}}\mathbf{x}. \tag{18}$$

This means that our robust approach first approximates the signal by its orthogonal projection onto the sampling space, and then applies the analysis functions $\{\mathbf{w}_n\}$. Thus, we can also conclude that the suggested approximation method results in zero error if $\mathcal{W} \subseteq \mathcal{S}$ or if the prior knowledge $\mathbf{x} \in \mathcal{S}$ exists. In fact, by identifying A of (4) with S, the solutions indeed coincide. Interestingly, $P_{\mathcal{S}}\mathbf{x}$ is the minimax approximation of $\mathbf{x}$ over the set $\mathcal{D} = \{\mathbf{x} \mid S^*\mathbf{x} = c,\ \|\mathbf{x}\| \leq L\}$, as incorporated in the following proposition.

Proposition 1: $P_{\mathcal{S}}\mathbf{x}$ is the minimax approximation of $\mathbf{x}$ over the set $\mathcal{D} = \{\mathbf{x} \mid S^*\mathbf{x} = c,\ \|\mathbf{x}\| \leq L\}$, *i.e.,* it is the unique solution of

$$\min_{\hat{\mathbf{x}}} \max_{\mathbf{x}\in\mathcal{D}} \|\mathbf{x} - \hat{\mathbf{x}}\|^2.$$

Proof: Projecting $\mathbf{x} - \hat{\mathbf{x}}$ along $\mathcal{S}$ and $\mathcal{S}^{\perp}$ we have

$$\|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \left\|S(S^*S)^{\dagger}c - P_{\mathcal{S}}\hat{\mathbf{x}}\right\|^2 + \|P_{\mathcal{S}^{\perp}}\mathbf{x} - P_{\mathcal{S}^{\perp}}\hat{\mathbf{x}}\|^2. \tag{19}$$

Accordingly, the maximization stage can be rewritten as

$$\begin{aligned} \max_{\mathbf{x}\in\mathcal{D}} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \|e_{\mathcal{S}}(\hat{\mathbf{x}})\|^2 + \max_{\mathbf{x}\in\mathcal{D}} \|P_{\mathcal{S}^{\perp}}\mathbf{x} - P_{\mathcal{S}^{\perp}}\hat{\mathbf{x}}\|^2 = \qquad (20) \\ = \quad \|e_{\mathcal{S}}(\hat{\mathbf{x}})\|^2 + \\ + \max_{\mathbf{x}\in\mathcal{D}} \left(\|P_{\mathcal{S}^{\perp}}\mathbf{x}\|^2 - 2\Re\langle P_{\mathcal{S}^{\perp}}\mathbf{x}, P_{\mathcal{S}^{\perp}}\hat{\mathbf{x}}\rangle + \|P_{\mathcal{S}^{\perp}}\hat{\mathbf{x}}\|^2 \right), \end{aligned}$$

where $e_{\mathcal{S}}(\hat{\mathbf{x}}) = S(S^*S)^\dagger c - P_{\mathcal{S}}\hat{\mathbf{x}}$. Similarly to the proof of Theorem 2, we can replace $-2\Re\langle P_{\mathcal{S}^\perp}\mathbf{x}, P_{\mathcal{S}^\perp}\hat{\mathbf{x}}\rangle$ by its absolute value. Then, the minimax objective can be lower bounded by

$$\min_{\hat{\mathbf{x}}}\Bigg(\|e_{\mathcal{S}}(\hat{\mathbf{x}})\|^2 + \max_{\mathbf{x}\in\mathcal{D}}\big(\|P_{\mathcal{S}^\perp}\mathbf{x}\|^2 + 2\,|\langle P_{\mathcal{S}^\perp}\mathbf{x}, P_{\mathcal{S}^\perp}\hat{\mathbf{x}}\rangle| + \\ + \|P_{\mathcal{S}^\perp}\hat{\mathbf{x}}\|^2\big)\Bigg) \geq \max_{\mathbf{x}\in\mathcal{D}}\|P_{\mathcal{S}^\perp}\mathbf{x}\|^2\,, \tag{21}$$

where we used the fact that for all $\hat{\mathbf{x}}$ we must have $\|e_{\mathcal{S}}(\hat{\mathbf{x}})\|^2 \geq 0$ and $2\,|\langle P_{\mathcal{S}^\perp}\mathbf{x}, P_{\mathcal{S}^\perp}\hat{\mathbf{x}}\rangle| + \|P_{\mathcal{S}^\perp}\hat{\mathbf{x}}\|^2 \geq 0$. The proof then follows since $\hat{\mathbf{x}} = P_{\mathcal{S}}\mathbf{x}$ is the minimizer which achieves this lower bound. Furthermore, it is unique, since from (21) the optimal solution must satisfy $P_{\mathcal{S}^\perp}\hat{\mathbf{x}} = 0$ and $e_{\mathcal{S}}(\hat{\mathbf{x}}) = 0$. □

We conclude that the problem of approximating the representation coefficients in a minimax sense could be split into two parts; first obtaining the minimax approximation of $\mathbf{x}$ itself, then applying the analysis operator W^* to that approximation.

A. *Element-Wise Optimality*

In Theorem 2, we approximate a set of representation coefficients $W^*\mathbf{x}$, by minimizing (in a minimax sense) the squared norm of the error sequence $W^*\mathbf{x} - d$. Instead, one may suggest alternative objectives which combine the entries of the error sequence in an l_1 norm *i.e.,* $\|W^*\mathbf{x} - d\|_{l_1}$ or the l_∞ norm $\|W^*\mathbf{x} - d\|_{l_\infty}$. We now show that the suggested solution (9) is optimal according to all the above criteria, as it is minimax optimal element-wise.

Theorem 3: The sequence $d = W^*S(S^*S)^\dagger c$ is a solution of $\min_d \max_{c=S^*\mathbf{x},\|\mathbf{x}\|\leq L}\|W^*\mathbf{x} - d\|_{l_p}$, for any natural p.

Proof: First assume that a single representation coefficient is to be approximated in a minimax sense

$$\min_{\widetilde{d}}\max_{c=S^*\mathbf{x},\|\mathbf{x}\|\leq L}\left|\langle\mathbf{w},\mathbf{x}\rangle - \widetilde{d}\right|^2, \tag{22}$$

where $\widetilde{d}$ is a scalar. Degenerating the result (9) of Theorem 2 by letting W be the set transformation of the single function $\mathbf{w}(t)$, the solution of (22) is

$$\widetilde{d} = \langle\mathbf{w}, S(S^*S)^\dagger c\rangle.$$

Since (9) satisfies $d[n] = \langle\mathbf{w}_n, S(S^*S)^\dagger c\rangle$ for each entry of the vector d, we conclude that it is element-wise optimal, implying that the solution will not change if we combine the individual errors in l_1, l_∞ or any l_p norm. □

V. Imposing Smoothness by Sobolev Spaces

The objective in Theorem 2 considers functions within the set $\mathcal{D} = \{\mathbf{x}\mid S^*\mathbf{x} = c,\ \|\mathbf{x}\|\leq L\}$. However, sometimes we have prior knowledge of an input signal to be smooth. If we could restrict the set of possible inputs to include only smooth functions, then the performance of the robust objective may be improved.

Sobolev spaces are natural candidates to describe smoothness. For simplicity, our main discussion will concern the Sobolev space of order one.

Definition 1: The Sobolev space of order one W_2^1 is the Hilbert space of functions which have a finite L_2 norm, and whose first derivative also has a finite L_2 norm. The inner product in this space is given by $\langle \mathbf{a}, \mathbf{b} \rangle_{W_2^1} = \langle \mathbf{a}, \mathbf{b} \rangle_{L_2} + \langle \mathbf{a}^{(1)}, \mathbf{b}^{(1)} \rangle_{L_2}$, where $\mathbf{a}^{(1)}$ and $\mathbf{b}^{(1)}$ stand for the first derivative of $\mathbf{a}$ and $\mathbf{b}$, respectively [12].

We note that the definition of the W_2^1 inner product, as stated above, is the common (but definitely not the only) choice. Additionally, note that $\mathbf{x} \in W_2^1$ also implies $\mathbf{x} \in L_2$ (but the opposite is not necessarily true). If we have prior knowledge that the input $\mathbf{x}$ and its first derivative are of finite energy (which in particular implies that $\mathbf{x}$ is continuous), we may consider the set of possible inputs to be $\tilde{\mathcal{D}} = \left\{ \mathbf{x} \mid S^*\mathbf{x} = c,\ \|\mathbf{x}\|_{W_2^1} \leq \tilde{L} \right\}$, where $\tilde{L}$ is an upper bound on the W_2^1 norm of $\mathbf{x}$. This leads to the following minimax objective:

$$\min_{d} \max_{\|\mathbf{x}\|_{W_2^1} \leq \tilde{L}, c = S^*\mathbf{x}} \|W^*\mathbf{x} - d\|_{l_2}^2 . \tag{23}$$

As we will soon show, we can solve this robust objective using Theorem 2. Furthermore, to determine the best sequence d, one does not have to know the upper bound $\tilde{L}$ in advance, however, it has to be finite.

Since the derivations of Theorem 2 are general enough to be applicable to any Hilbert space, the problem can be solved in the order one Sobolev space as well. However, the objective (23) contains mixed inner products and norms; $S^*\mathbf{x}$ and $W^*\mathbf{x}$ describe L_2 inner products, while $\|\mathbf{x}\|_{W_2^1} \leq \tilde{L}$ is a Sobolev norm constraint. Hence, we will first recast the whole problem into the order one Sobolev space, then apply the results of Theorem 2.

To this end, note that W_2^1 inner products can be compactly written in the Fourier domain by

$$\langle \mathbf{s}, \mathbf{x} \rangle_{W_2^1} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{\mathbf{S}^F(\omega)} \mathbf{X}^F(\omega) \left(1 + \omega^2\right) d\omega, \tag{24}$$

where $\mathbf{S}^F(\omega)$ and $\mathbf{X}^F(\omega)$ are the Fourier transforms of $\mathbf{s}(t)$ and $\mathbf{x}(t)$, respectively. As introduced in [10], we can use (24) to rewrite L_2 inner products as W_2^1 inner products. Specifically, for any $\mathbf{a} \in L_2, \mathbf{b} \in W_2^1$

$$\langle \mathbf{a}(t), \mathbf{b}(t) \rangle_{L_2} = \langle \mathbf{a}(t) * \mathbf{u}(t), \mathbf{b}(t) \rangle_{W_2^1}, \tag{25}$$

where $*$ stands for the convolution operation and

$$\mathbf{u}(t) = \frac{1}{2} e^{-|t|}, \tag{26}$$

is the inverse Fourier transform of $1/(1 + \omega^2)$.

Using (25), we can replace the L_2 inner products $W^*\mathbf{x}$ and $S^*\mathbf{x}$, using their Sobolev counterparts, recasting our problem into the order one Sobolev space, which leads to the following theorem.

Theorem 4: Consider the problem

$$\min_{d} \max_{c = S^*\mathbf{x}, \|\mathbf{x}\|_{W_2^1} \leq \tilde{L}} \|W^*\mathbf{x} - d\|_{l_2}^2 , \tag{27}$$

where W and S are bounded set transformations with range $\mathcal{W}$ and $\mathcal{S}$, respectively. The (unique) solution is

$$d = \tilde{W}^* \tilde{S} (\tilde{S}^* \tilde{S})^{\dagger} c, \tag{28}$$

where the inner products described by (28) are computed in the order one Sobolev space, $\tilde{S}, \tilde{W}$ are the set transformations of $\{\tilde{\mathbf{s}}_n = \mathbf{s}_n * \mathbf{u}\}$ and $\{\tilde{\mathbf{w}}_n = \mathbf{w}_n * \mathbf{u}\}$ respectively, and the function $\mathbf{u}(t)$ is given by (26).

Proof: Using (25), we can rewrite (27) as

$$\min_{d} \max_{c=\tilde{S}^*\mathbf{x}, \|\mathbf{x}\| \leq \tilde{L}} \left\| \tilde{W}^*\mathbf{x} - d \right\|_{l_2}^2,$$

where $\tilde{S}$ and $\tilde{W}$ are the set transformation of $\{\tilde{\mathbf{s}}_n = \mathbf{s}_n * \mathbf{u}\}$ and $\{\tilde{\mathbf{w}}_n = \mathbf{w}_n * \mathbf{u}\}$, respectively, and $\mathbf{u}(t)$ is given by (26). Since the derivations of Theorem 2 apply to any Hilbert space, the solution maintains similar form as in (9), resulting in (28). □

We point out that the obtained solution is element-wise optimal. The latter can be shown in a similar way as done in Theorem 3. Also, note that (28) describes W_2^1 inner products. In practice, this means that the i, j'th element of the matrix $\tilde{W}^*\tilde{S}$ is

$$\langle \tilde{\mathbf{w}}_i, \tilde{\mathbf{s}}_j \rangle_{W_2^1} = \langle \mathbf{w}_i, \tilde{\mathbf{s}}_j \rangle_{L_2}, \tag{29}$$

and similarly

$$\langle \tilde{\mathbf{s}}_i, \tilde{\mathbf{s}}_j \rangle_{W_2^1} = \langle \mathbf{s}_i, \tilde{\mathbf{s}}_j \rangle_{L_2}. \tag{30}$$

Rewriting $P_{\widetilde{\mathcal{S}}}\mathbf{x} = \tilde{S}(\tilde{S}^*\tilde{S})^\dagger c$ and using (29), we obtain the following corollary.

Corollary 2: The solution (28) can be written as

$$d = W^* P_{\widetilde{\mathcal{S}}}\mathbf{x}, \tag{31}$$

where $\widetilde{\mathcal{S}}$ is the closure of *span* $\{\tilde{\mathbf{s}}_n\}$ and $P_{\widetilde{\mathcal{S}}}\mathbf{x}$ stands for the orthogonal projection of $\mathbf{x}$, in the W_2^1 sense, onto $\widetilde{\mathcal{S}}$. The operator W^* describes the usual L_2 inner products.

Another interesting interpretation of Theorem 4 is evident by rewriting all inner products in the L_2 space. Combining (30) with (31) we obtain the following corollary.

Corollary 3: The solution (28) can be written as

$$d = W^*\tilde{S}(S^*\tilde{S})^\dagger S^*\mathbf{x} = W^* E_{\widetilde{\mathcal{S}},\mathcal{S}^\perp}\mathbf{x}, \tag{32}$$

where $E_{\widetilde{\mathcal{S}},\mathcal{S}^\perp}$ stands for the oblique projection operator [16], [23], in the L_2 sense, with a range space $\widetilde{\mathcal{S}}$ and a null space $\mathcal{S}^\perp$.

Note that the oblique projection operator is well defined since there is a bijection between $\mathcal{S}$ and $\widetilde{\mathcal{S}}$.

Similarly as in Proposition IV, it can be shown that $P_{\widetilde{\mathcal{S}}}\mathbf{x} = E_{\widetilde{\mathcal{S}},\mathcal{S}^\perp}\mathbf{x}$ is the unique solution of

$$\arg\min_{\hat{\mathbf{x}}} \max_{c=S^*\mathbf{x}, \|\mathbf{x}\|_{W_2^1} \leq \tilde{L}} \|\mathbf{x} - \hat{\mathbf{x}}\|_{W_2^1}^2.$$

Corollaries 2 and 3 imply that the problem of Theorem 4, could be split into two parts; first obtaining the Sobolev minimax approximation of $\mathbf{x}$ itself, then applying the analysis operator W^* to that approximation.

In this section we have considered the Sobolev space of order one. It is possible to extend the above derivations to higher orders of Sobolev spaces, if sufficient degree of smoothness is known to be present. The order r Sobolev space W_2^r is composed of finite energy functions with r finite energy derivatives; W_2^r inner products can be written as $\langle \mathbf{a}, \mathbf{b} \rangle_{W_2^r} = \sum_{k=0}^{r} \langle \mathbf{a}^{(\mathbf{k})}, \mathbf{b}^{(\mathbf{k})} \rangle_{L_2}$, where the (k) superscript stands for the k'th derivative. Thus, we can obtain similar results, which only require the replacement of the function $\mathbf{u}(t)$ of (26) with the inverse Fourier transform of $1/(1+\omega^2+\omega^4 \ldots + \omega^{2r})$.

As a concluding remark, we note that the obtained solution takes the form of applying the analysis functions to $\hat{\mathbf{x}} = P_{\widetilde{\mathcal{S}}}\mathbf{x} = E_{\widetilde{\mathcal{S}},\mathcal{S}^{\perp}}\mathbf{x}$, which is the minimax approximation of $\mathbf{x}$ within the space $\widetilde{\mathcal{S}}$. Furthermore, this space is determined by the sampling functions $\{\mathbf{s}_n\}$ and the smoothness constraint (manifested by $\mathbf{u}$). Thus, we have obtained a nice counterpart to methods that **arbitrarily** choose the interpolation space.

A. Extension of the Ideal Sampling Case Results

In this section, we show how Theorem 4 extends some recent work by Kirshner and Porat [10], [11].

In [10] [11], it was assumed that a single representation coefficient $\langle \mathbf{w}(t), \mathbf{x}(t) \rangle_{L_2}$ is to be approximated by linearly processing the ideal sample sequence $S_T\mathbf{x} = \{\mathbf{x}(nT)\}$. The processing is performed by calculating the l_2 inner product $\langle b, S_T\mathbf{x} \rangle_{l_2}$ with some sequence b. The ideal sampling case must be treated with caution; restricting $\mathbf{x}(t)$ to be an L_2 signal is not sufficient to guarantee that its ideal samples form an l_2 sequence. For example, take any $\mathbf{x}(t)$ in L_2 and redefine its values on the uniform sampling grid $\{nT\}$ to equal one. Clearly the redefined signal is still in L_2, but its ideal sample sequence has infinite energy. This last example shows that whenever ideal sampling is of concern, we want the signal to be smooth, in addition to having finite energy. Only then, it is guarantied that the l_2 inner product $\langle b, S_T\mathbf{x} \rangle_{l_2}$ is well defined.

To assure finite energy of $S_T\mathbf{x} = \{\mathbf{x}(nT)\}$ it was assumed in [10] that $\mathbf{x}$ lies within an order one Sobolev space. To see that this is a sufficient condition, we refer the reader to [10], [24], or [9, Appendix C] for the general case of a sampling function with a bounded Fourier transform. Then, for any $\mathbf{x} \in W_2^1$ the approximation error was upper bounded by

$$|\langle \mathbf{w}(t), \mathbf{x}(t) \rangle_{L_2} - \langle b, S_T\mathbf{x} \rangle_{l_2}| \leq B \left\|\mathbf{x}\right\|_{W_2^1},$$

where B is a constant that depends on $T, b, \mathbf{w}$ and the function $\mathbf{u}(t)$ of (26). Finally, B was minimized with respect to the processing sequence b.

Reinterpreting the derivations in [10], the approximation problem of [10, Thr. 3] can be restated as a minimax objective:

$$\min_{b} \max_{\|\mathbf{x}\|_{W_2^1} \leq \tilde{L}} |\langle \mathbf{w}, \mathbf{x} \rangle_{L_2} - \langle b, S_T\mathbf{x} \rangle_{l_2}|, \tag{33}$$

where $\tilde{L}$ is some (finite) upper bound on the W_2^1 norm of $\mathbf{x}$. Note that in contrast to our minimax formulation, the prior $c = S_T\mathbf{x}$ is not part of (33). Furthermore, the processing method is restricted to a linear form through $\langle b, S_T\mathbf{x} \rangle_{l_2}$. Yet, it can be shown that the optimal solution of those two minimax problems remains the same (see [16] for a similar setup). In [10] it is found that the optimal sequence b satisfies

$$\sum_n \mathbf{u}(t-nT)b[n] = P_{\mathcal{U}}\left(\mathbf{w} * \mathbf{u}\right), \tag{34}$$

where $P_{\mathcal{U}}$ is the orthogonal projection, in the W_2^1 sense, onto $\mathcal{U}$, which is the closure of $span\left\{\mathbf{u}(t-nT)\right\}$.

We now show this result of [10] as a special case of Theorem 4. First, define U to be the set transformation of the function set $\{\mathbf{u}(t-nT)\}$. It can be shown that on W_2^1, U is the adjoint of the ideal sampling operator S_T [10]. Indeed, for any $\mathbf{x} \in W_2^1$, we have

$$\begin{aligned} \mathbf{x}(nT) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{X}^F(\omega)e^{j\omega nT}\frac{1}{(1+\omega^2)}\left(1+\omega^2\right)d\omega \\ &= \langle \mathbf{u}(t-nT), \mathbf{x}(t)\rangle_{W_2^1}, \end{aligned} \tag{35}$$

or, using operator notations, $c = S_T\mathbf{x} = U^*\mathbf{x}$. We note that U (as well as $U^* = S_T$) is a well defined bounded operator in W_2^1 [9, Appendix C]. Additionally, the single representation coefficient $\langle \mathbf{w}, \mathbf{x}\rangle_{L_2}$ can be written as the order one Sobolev inner product $\tilde{W}^*\mathbf{x}$, with $\tilde{W}$ being the set transform of $\tilde{\mathbf{w}}(t) = \mathbf{w}(t) * \mathbf{u}(t)$. Identifying $\tilde{S}$ with U, we have from (28)

$$\begin{aligned} d &= \tilde{W}^*U(U^*U)^\dagger c \\ &= \langle \tilde{\mathbf{w}}, U(U^*U)^\dagger c\rangle_{W_2^1} \\ &= \langle (U^*U)^\dagger U^*\tilde{\mathbf{w}}, c\rangle_{l_2} = \langle b, c\rangle_{l_2}, \end{aligned} \tag{36}$$

where we denote $b = (U^*U)^\dagger U^*\tilde{\mathbf{w}}$. As a result, Ub is exactly the orthogonal projection of $\tilde{\mathbf{w}} = \mathbf{w} * \mathbf{u}$ onto the space $\mathcal{U}$, which is compatible with (34).

B. The Shift Invariant Case

The approximation (9) was derived for general sampling and analysis subspaces of L_2 or W_2^1. An interesting special case of this setup is when in addition, $\mathcal{S}$ and $\mathcal{W}$ are real *shift invariant* (SI) subspaces, each spanned by shifts of length T of some fixed generating function [13], [16]. In this case, as we will show, the approximation sequence d can be obtained by discrete-time filtering of the sample sequence c. Assume first that $\mathcal{H} = L_2$ and the SI subspaces are

$$\begin{aligned} \mathcal{S} &= \left\{ \mathbf{f}(t) \mid \mathbf{f}(t) = \sum_n a[n]\mathbf{s}(t-nT), a \in \ell_2 \right\}; \\ \mathcal{W} &= \left\{ \mathbf{f}(t) \mid \mathbf{f}(t) = \sum_n a[n]\mathbf{w}(t-nT), a \in \ell_2 \right\}, \end{aligned} \tag{37}$$

where $\mathbf{s}(t)$ and $\mathbf{w}(t)$ are the real generators of $\mathcal{S}$ and $\mathcal{W}$, respectively. In this SI case, the samples $c[n]$, which are given by

$$c[n] = \int \mathbf{s}(t-nT)\mathbf{x}(t)dt = \mathbf{x}(t) * \mathbf{s}(-t)|_{t=nT}, \tag{38}$$

correspond to samples at times $t = nT$ of the output of a filter with impulse response $\mathbf{s}(-t)$, with $\mathbf{x}(t)$ as its input (see Figure 1). Here $\mathbf{g}(t) * \mathbf{z}(t)$ denotes the continuous-time convolution between the signals $\mathbf{g}(t)$ and $\mathbf{z}(t)$, and $\mathbf{y}(t)|_{t=nT} = \mathbf{y}(nT)$.

To ensure that the functions $\{\mathbf{s}_n(t)=\mathbf{s}(t-nT)\}$ and $\{\mathbf{w}_n(t)=\mathbf{w}(t-nT)\}$ form frames for $\mathcal{S}$ and $\mathcal{W}$ respectively, a simple condition can be verified in the frequency domain [21]:

$$\begin{aligned} \alpha \leq R^f_{\mathbf{W},\mathbf{W}}(\omega) \leq \beta, \quad \omega \in \mathcal{I}_{\mathcal{W}}; \\ \gamma \leq R^f_{\mathbf{S},\mathbf{S}}(\omega) \leq \eta, \quad \omega \in \mathcal{I}_{\mathcal{S}}, \end{aligned} \tag{39}$$

for some $0<\alpha\leq\beta<\infty$ and $0<\gamma\leq\eta<\infty$. Here we denote,

$$R^f_{\mathbf{A},\mathbf{B}}(\omega)=\frac{1}{T}\sum_{k=-\infty}^{\infty}\overline{\mathbf{A}^F}\left(\frac{\omega+2\pi k}{T}\right)\mathbf{B}^F\left(\frac{\omega+2\pi k}{T}\right),$$

where $\mathbf{W}^F(\omega),\mathbf{S}^F(\omega)$ are the continuous-time Fourier transforms of the generators $\mathbf{s}(t),\mathbf{w}(t)$, and $\mathcal{I}_{\mathcal{W}},\mathcal{I}_{\mathcal{S}}$ are the set of frequencies ω for which $R^f_{\mathbf{W},\mathbf{W}}(\omega)\neq 0$ and $R^f_{\mathbf{S},\mathbf{S}}(\omega)\neq 0$, respectively.

Letting S and W be the set transformations of $\{\mathbf{s}(t-nT)\}$ and $\{\mathbf{w}(t-nT)\}$, respectively, it is easy to see that W^*Sa is equivalent to filtering the sequence $a[n]$ by a discrete-time LTI filter with DTFT $R^f_{\mathbf{W},\mathbf{S}}(\omega)$. Similarly, the pseudo-inverse operator $(S^*S)^\dagger$ can be represented by application of a filter with DTFT $1/R^f_{\mathbf{S},\mathbf{S}}(\omega)$ for $\omega\in\mathcal{I}_{\mathcal{S}}$ and zero otherwise. Combining the above, we have that for $\mathcal{H}=L_2$ the sequence $d=W^*S(S^*S)^\dagger c$ can be obtained by filtering the sample sequence c with a digital filter having the frequency response

$$G^f(\omega)=\begin{cases}\frac{R^f_{\mathbf{W},\mathbf{S}}(\omega)}{R^f_{\mathbf{S},\mathbf{S}}(\omega)}, & \omega\in\mathcal{I}_{\mathcal{S}};\\ 0, & \omega\notin\mathcal{I}_{\mathcal{S}}.\end{cases} \tag{40}$$

We point out that by a proper choice of the sampling and analysis functions, the filter (40) is compatible with the solution for the 'wavelet crime' problem, as obtained in [7].

We now find the digital filter, when a smoothness constraint is incorporated, by considering the results of Theorem 4. Let $\mathbf{u}(t)$, as given by (26), be the inverse Fourier transform of $1/(1+\omega^2)$ and define $\tilde{\mathbf{s}}=\mathbf{s}*\mathbf{u}$. In the SI case, the solution (32) then takes the form

$$G^f(\omega)=\begin{cases}\frac{R^f_{\mathbf{W},\tilde{\mathbf{S}}}(\omega)}{R^f_{\mathbf{S},\tilde{\mathbf{S}}}(\omega)}, & \omega\in\mathcal{I}_{\mathcal{S}};\\ 0, & \omega\notin\mathcal{I}_{\mathcal{S}},\end{cases} \tag{41}$$

where $\tilde{\mathbf{S}}^F(\omega)=\mathbf{S}^F(\omega)/(1+\omega^2)$ is the Fourier transform of $\tilde{\mathbf{s}}(t)$.

VI. Error Analysis

In this section we investigate the error resulting from the minimax method. We then derive sufficient conditions for our method to outperform the sum approximation (5). Although we use the S and W operators (as opposed to their Sobolev counterparts $\tilde{S},\tilde{W}$), all derivations are applicable to Sobolev spaces by considering the appropriate inner products.

Let

$$e_{mx}=W^*\mathbf{x}-d \tag{42}$$

be the error sequence due to the minimax approach, where d is given by (9). Using (18) we can express the error as

$$e_{mx} = W^* P_{\mathcal{S}^\perp} \mathbf{x}. \tag{43}$$

Define e_{sum} to be the error sequence due to the sum approximation method (5). The n'th element of e_{sum} satisfies

$$e_{sum}[n] = \langle \mathbf{w}_n, \mathbf{x} \rangle_{L_2} - T \langle S_T \mathbf{w}_n, c \rangle_{l_2}. \tag{44}$$

It should be noted that in many engineering applications, pre-filtering followed by ideal sampling is performed. Then, the natural choice of T is the shift interval for the acquisition device impulse response (Figure 1).

We now examine the conditions which will assure that $\|e_{mx}\|_{l_2}^2 \leq \|e_{sum}\|_{l_2}^2$. In the following lemma we first introduce tight bounds for the difference $\|e_{sum}\|_{l_2}^2 - \|e_{mx}\|_{l_2}^2$. Clearly, if the difference is positive, then the minimax method is preferable to the sum approximation method, and vice versa.

Lemma 1: For any $\mathbf{x}$ in the set $\left\{\mathbf{x} |\ c = S^* \mathbf{x}, \|\mathbf{x}\|_{L_2} \leq L \right\}$ the squared norm difference $\|e_{sum}\|_{l_2}^2 - \|e_{mx}\|_{l_2}^2$ lies within the tight bounds

$$B_L \leq \|e_{sum}\|_{l_2}^2 - \|e_{mx}\|_{l_2}^2 \leq B_H, \tag{45}$$

where

$$\begin{aligned} B_L &= \|a\|_{l_2}^2 - 2 \left| \langle a, e_{mx} \rangle_{l_2} \right| \\ B_H &= \|a\|_{l_2}^2 + 2 \left| \langle a, e_{mx} \rangle_{l_2} \right|, \end{aligned} \tag{46}$$

and

$$a[n] = \langle \mathbf{w}_n, S(S^*S)^\dagger c \rangle_{L_2} - T \langle S_T \mathbf{w}_n, c \rangle_{l_2}. \tag{47}$$

Proof: Using (42) and (44) we can relate the two error sequences by

$$e_{sum} = a + e_{mx}, \tag{48}$$

with a given by (47). Note that since the sample sequence c is available, and so are T, S and W, the sequence a is known as well. Furthermore, $a \in l_2$. The latter is evident by rewriting $a = W^* \left(S(S^*S)^\dagger - T S_T^* \right) c$. Since $c \in l_2$, $S_T^* = U$ is a bounded transformation [9, Appendix C] and so are $S(S^*S)^\dagger$ and W^* (due to the frame assumptions) the sequence a has finite energy. Taking the squared norm of both sides of (48) and rearranging terms, we get

$$\|e_{sum}\|_{l_2}^2 - \|e_{mx}\|_{l_2}^2 = \|a\|_{l_2}^2 + 2\Re \left\{ \langle a, e_{mx} \rangle_{l_2} \right\}.$$

The bounds (46) then follow from

$$- \left| \langle a, e_{mx} \rangle_{l_2} \right| \leq \Re \left\{ \langle a, e_{mx} \rangle_{l_2} \right\} \leq \left| \langle a, e_{mx} \rangle_{l_2} \right|.$$

We now show that the bounds are tight. Assume to the contrary that for all $\mathbf{x}$ in the set $\left\{\mathbf{x} |\ c = S^* \mathbf{x}, \|\mathbf{x}\|_{L_2} \leq L \right\}$,

$$\Re \left\{ \langle a, e_{mx} \rangle_{l_2} \right\} < \left| \langle a, e_{mx} \rangle_{l_2} \right|.$$

Define $\mathbf{x}_2 = S(S^*S)^\dagger c + \frac{\langle W^* P_{\mathcal{S}^\perp}\mathbf{x}, a\rangle_{l_2}}{|\langle W^* P_{\mathcal{S}^\perp}\mathbf{x}, a\rangle_{l_2}|} P_{\mathcal{S}^\perp}\mathbf{x}$. We point out that $\mathbf{x}_2$ is a valid input since it satisfies the norm constraint $\|\mathbf{x}_2\| \leq L$ and is consistent with the known samples (*i.e.,* $c = S^*\mathbf{x}_2$). However, by examining the minimax error at $\mathbf{x}_2$, we have

$$\Re\left\{\langle a, W^* P_{\mathcal{S}^\perp}\mathbf{x}_2\rangle_{l_2}\right\} = \left|\langle a, W^* P_{\mathcal{S}^\perp}\mathbf{x}_2\rangle_{l_2}\right|,$$

thus contradicting our initial assumption. The proof of tightness for the lower bound is similar. □

Since the tight upper bound B_H is nonnegative for all choices of e_{mx}, we conclude that the sum approximation method **cannot** outperform the proposed minimax approach, for **all** possible inputs. On the other hand, in some cases, it is possible to have better performance by the minimax approach, for all possible inputs. To assure this, the lower bound B_L must be positive. In the following lemma, we provide a tight upper bound on $\|e_{mx}\|_{l_2}$ assuming that the set $\{\mathbf{w}_n(t)\}$ is orthonormal. Using this bound, we then state a sufficient condition for the minimax method to outperform the standard sum approximation approach for all possible $\mathbf{x}$.

Lemma 2: Let $\{\mathbf{w}_n(t)\}$ be an orthonormal set, and let $\mathbf{x}$ satisfy $\|\mathbf{x}\| \leq L, c = S^*\mathbf{x}$. Then

$$\|e_{mx}\|_{l_2} \leq B_{mx} = \sqrt{1 - \cos^2(\mathcal{W}, \mathcal{S})}L', \tag{49}$$

where

$$L' = \sqrt{L^2 - \|S^*(S^*S)^\dagger c\|^2} \tag{50}$$

is the norm of $P_{\mathcal{S}^\perp}\mathbf{x}$ and $\cos(\mathcal{W}, \mathcal{S}) = \inf_{\mathbf{y}\in\mathcal{W}, \|\mathbf{y}\|=1} \|P_{\mathcal{S}}\mathbf{y}\|$.

Before giving the proof, we mention that $\cos(\mathcal{W}, \mathcal{S})$ is related to the largest angle [13], [16] between the spaces. An explicit expression for $\cos(\mathcal{W}, \mathcal{S})$ in the case of SI spaces is given in [13].

Proof: Since

$$\|e_{mx}\|_{l_2}^2 = \|W^* P_{\mathcal{S}^\perp}\mathbf{x}\|_{l_2}^2 = \langle W^* P_{\mathcal{S}^\perp}\mathbf{x}, W^* P_{\mathcal{S}^\perp}\mathbf{x}\rangle_{l_2},$$

we also have

$$\begin{aligned}
\|e_{mx}\|_{l_2}^2 &= \langle P_{\mathcal{S}^\perp}\mathbf{x}, WW^* P_{\mathcal{S}^\perp}\mathbf{x}\rangle \\
&= \langle P_{\mathcal{S}^\perp}\mathbf{x}, P_{\mathcal{W}} P_{\mathcal{S}^\perp}\mathbf{x}\rangle \\
&= \langle P_{\mathcal{W}} P_{\mathcal{S}^\perp}\mathbf{x}, P_{\mathcal{W}} P_{\mathcal{S}^\perp}\mathbf{x}\rangle \\
&= \|P_{\mathcal{W}} P_{\mathcal{S}^\perp}\mathbf{x}\|^2,
\end{aligned} \tag{51}$$

where we utilized the orthonormality of the analysis set $\{\mathbf{w}_n(t)\}$ to rewrite $WW^* = P_{\mathcal{W}}$. For any $\mathbf{x}$ satisfying $c = S^*\mathbf{x}$, $\|x\| \leq L$ we have $\|P_{\mathcal{S}^\perp}\mathbf{x}\| \leq L'$, where L' is given by (50). Thus we can bound

$$\|P_{\mathcal{W}} P_{\mathcal{S}^\perp}\mathbf{x}\| \leq \sin\left(\mathcal{S}^\perp, \mathcal{W}^\perp\right) L',$$

where we denote $\sin\left(\mathcal{S}^\perp, \mathcal{W}^\perp\right) = \sup_{\mathbf{y}\in\mathcal{S}^\perp, \|\mathbf{y}\|=1} \|P_{\mathcal{W}}\mathbf{y}\|$. From [13], [25] $\sin\left(\mathcal{S}^\perp, \mathcal{W}^\perp\right) = \sqrt{1 - \cos^2(\mathcal{W}, \mathcal{S})}$ and the proof follows. □

Corollary 4: Let $\{\mathbf{w}_n(t)\}$ be an orthonormal set. A sufficient condition for the minimax method to outperform the sum approach for all $\mathbf{x}$ in the set $\{\mathbf{x}|\ \|\mathbf{x}\| \leq L, c = S^*\mathbf{x}\}$ is $\|a\|_{l_2} \geq 2B_{mx}$, where a and B_{mx} are given by (47) and (49), respectively.

Proof: Using Lemma 1, the Cauchy-Schwartz inequality and Lemma 2, we have

$$B_L \geq \|a\|_{l_2}^2 - 2\|a\|_{l_2}\|e_{mx}\|_{l_2} \geq \|a\|_{l_2}^2 - 2\|a\|_{l_2} B_{mx},$$

from which the proof follows. □

The error analysis is summarized in Figure 2.

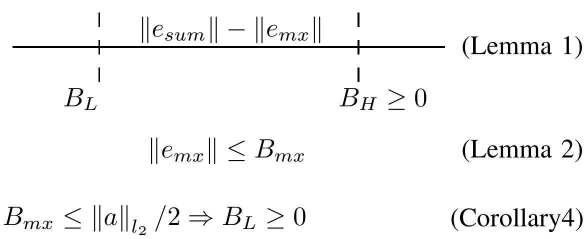


Fig. 2. Regions of $\|e_{sum}\| - \|e_{mx}\|$ for the case where $\{\mathbf{w}_n(t)\}$ is an orthonormal set. If the maximal norm of the minimax error (49) is smaller than $\|a\|_{l_2}/2$, then the suggested robust approach is superior to the sum method, for all possible inputs.

Another interesting case, which is easy to evaluate, is when a single representation coefficient $\langle \mathbf{w}, \mathbf{x}\rangle$ is to be approximated. In this setting, e_{mx}, e_{sum} and a are all scalars. It can then be shown that the minimax method and the sum approximation approach are tightly upper bounded by

$$\begin{aligned} |e_{mx}| &\leq B_{mx} = L'\|P_{\mathcal{S}^\perp}\mathbf{w}\| \\ |e_{sum}| &\leq B_{sum} = |a| + B_{mx}, \end{aligned} \tag{52}$$

where the worst possible input, which achieves those upper bounds is

$$\mathbf{x}_{worst} = S(S^*S)^\dagger c + \frac{aL'}{|a|\,\|P_{\mathcal{S}^\perp}\mathbf{w}\|} P_{\mathcal{S}^\perp}\mathbf{w}. \tag{53}$$

Additionally, a sufficient condition for the minimax method to outperform the sum approach becomes

$$|a| \geq 2L'\|P_{\mathcal{S}^\perp}\mathbf{w}\|. \tag{54}$$

For proof, we refer the reader to Appendix II.

As a conclusion from the above analysis, we get that when the spaces $\mathcal{W}$ and $\mathcal{S}$ are *close* (such that $\cos(\mathcal{W}, \mathcal{S})$ is close to one), or when most of the signal's energy lies within the sampling space $\mathcal{S}$ (such that L' is small), then the minimax method will outperform the standard approach. Similarly, for large sampling intervals T, we can make $\|a\|_{l_2}$ large enough, again guaranteeing better performance by the minimax method.

VII. SIMULATIONS

Suppose that we wish to approximate a single representation coefficient $\langle \mathbf{w}, \mathbf{x} \rangle_{L_2}$, where $\mathbf{w}(t)$ is a modulated and normalized Gaussian

$$\mathbf{w}(t) = \alpha e^{-t^2/2} \cos(4\pi t), \tag{55}$$

with α chosen such that $\|\mathbf{w}\|_{L_2} = 1$. Consider an input

$$\mathbf{x}(t) = e^{-50t^2} - e^{-50(t-0.75)^2}, \tag{56}$$

composed of two Gaussians, synchronized with the analysis function $\mathbf{w}(t)$ (Figure 3). For this example $\langle \mathbf{w}, \mathbf{x} \rangle_{L_2} \approx 0.2$.

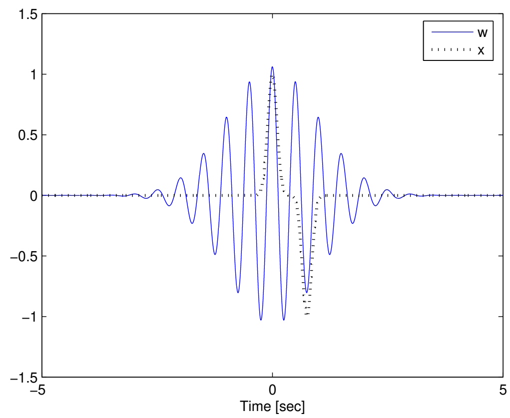


Fig. 3. The analysis function $\mathbf{w}(t)$ and the input signal $\mathbf{x}(t)$.

A. ZOH Sampling

Assume that the generalized samples of $\mathbf{x}(t)$ were obtained by averaging the value of $\mathbf{x}(t)$ within a small interval of length Δ, *i.e.,*

$$c[n] = \frac{1}{\Delta} \int_{nT}^{nT+\Delta} \mathbf{x}(t) dt. \tag{57}$$

In this setting, the n'th sampling vector $\mathbf{s}_n(t)$ of (1) is

$$\mathbf{s}_n(t) = \begin{cases} 1/\Delta, & t \in [nT, nT+\Delta]; \\ 0, & \text{otherwise.} \end{cases} \tag{58}$$

By processing the generalized samples $\{c[n] = \langle \mathbf{s}_n, \mathbf{x} \rangle_{L_2}\}$ using the transformation (9), we obtain the minimax approximation of $\langle \mathbf{w}, \mathbf{x} \rangle_{L_2}$. The latter can be obtained in the L_2 space, or transformed into a proper Sobolev space, *i.e.,* using (28), when smoothness is of concern. Note that the input signal of the example (56) indeed satisfies $\mathbf{x} \in W_2^1$. Accordingly, as we will show, the minimax solution with the smoothness constraints outperforms the minimax method which just assumes an L_2 input signal.

Interpreting the minimax solutions as the application of the analysis operator W^* to the approximates $P_{\mathcal{S}}\mathbf{x}$ and $E_{\widetilde{\mathcal{S}},\mathcal{S}^\perp}\mathbf{x}$ (*i.e.,* Equations (18) and (32) respectively), it is of interest to show the building blocks of those approximations. Figure 4 depicts the sampling functions $s_0(t) \in \mathcal{S}$ and $\tilde{s}_0(t) \in \widetilde{\mathcal{S}}$, for $\Delta = 0.05$. In Figure 5 we

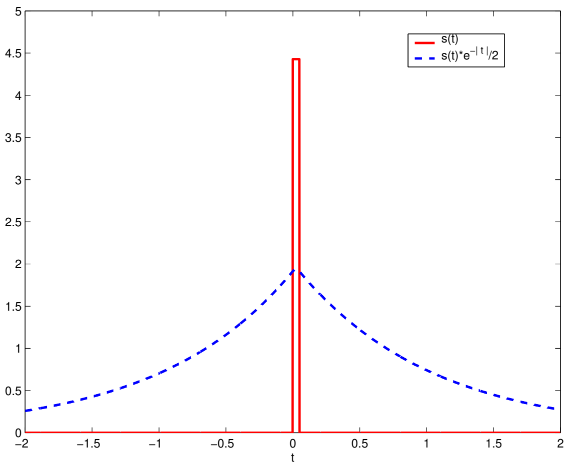


Fig. 4. *ZOH sampling. Shown are typical sampling functions $s_0(t)$ (58) and $\tilde{s}_0(t) = s_0(t) * e^{-|t|}/2$, which are the generators for the shift invariant spaces $\mathcal{S}$ and $\widetilde{\mathcal{S}}$, respectively. Here, $\Delta = 0.05$ and both functions are normalized for presentation purposes.*

plot a section of $\mathbf{x}$ with its projections onto the appropriate sampling spaces. The parameters T and Δ were set to 0.1[sec] and 0.05[sec], respectively. For this example, the space $\widetilde{\mathcal{S}}$ captures most of the signal's energy. Indeed, as can be seen from Figure 5, the approximation $E_{\widetilde{\mathcal{S}},\mathcal{S}^\perp}\mathbf{x}$ is very close to the original input and one may expect good results while evaluating the minimax objective with the smoothness constraints.

In addition to performing the minimax approximations, we have also processed the samples using the standard sum approach (5). In Figure 6 we present the specific errors for the current input, using several choices of T. The minimax solution is optimized for the worst possible inputs within the considered set. However, the assumed test signal (56) is by no means the worst possible input. It was merely used to produce the generalized samples. As a result, for some sampling intervals, the suggested robust solutions are better, while for others they are outperformed by the sum approximation.

It is also of interest to examine the signals that cause the highest value of the cost function. In Figure 7 we plot those worst inputs. In both the presented cases, the worst possible input is calculated according to (53), and is given by a projection of $\mathbf{x}$ onto the sampling space, and a vector in $\mathcal{S}^\perp$, which has the smallest angle with the analysis function $\mathbf{w}(t)$. As can be seen from part (a) of Figure 7, the worst possible input in the set $\mathcal{D} = \left\{\mathbf{x} \mid \|\mathbf{x}\|_{L_2} \leq L, c = S^*\mathbf{x}\right\}$ is a highly non smooth function. This input is indeed possible in the L_2 space, but it is not likely to appear if we know the signal to be smooth. If we consider only order one Sobolev functions, the worst input is a smooth function, and is much closer to the original input (part (b) of Figure 7). The exceptionally good results of Figure 7 (b) are due to the fact that for this example, most of the signals energy lies within the space $\widetilde{\mathcal{S}}$ (alternatively, L' is small). As a result, the approximation $P_{\widetilde{\mathcal{S}}}\mathbf{x} = E_{\widetilde{\mathcal{S}},\mathcal{S}^\perp}\mathbf{x}$ describes well the original input.

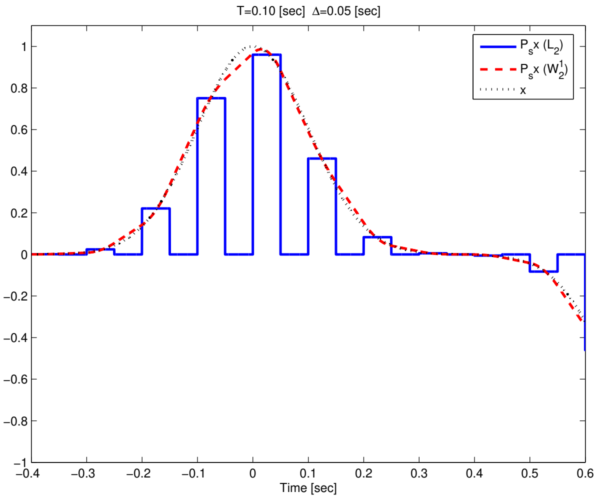


Fig. 5. A section of $\mathbf{x}$ and its approximation in the sampling space; The L_2 orthogonal projection onto $\mathcal{S}$ yields rectangular pulses. The oblique projection onto $\widetilde{\mathcal{S}}$ yields a smooth function, which is very close to the original input.

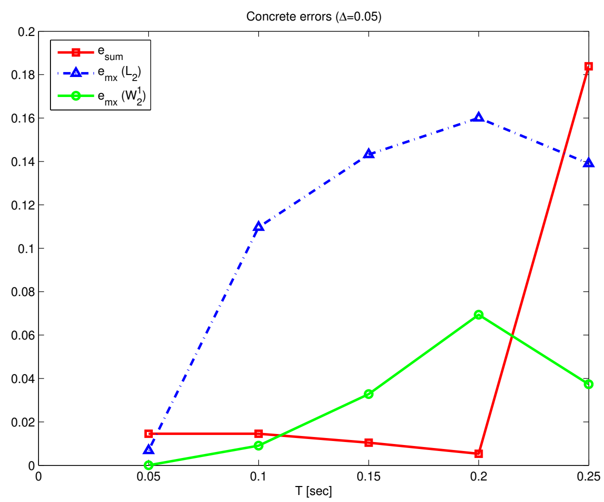


Fig. 6. The errors as a function of T.

Note that in all cases, the worst inputs look the same for the acquisition device, as they both produce the same generalized samples. To illustrate the last point, in Figure 8 we plot a section of $\mathbf{x}$ and the worst possible inputs (for the L_2 and the W_2^1 sets). In addition, we present the orthogonal projection $P_{\mathcal{S}}\mathbf{x}$, in the L_2 sense, which is composed of rectangular pulses describing the integration zone due to the sampling functions (58). As can be seen, all signals yield the same generalized samples, as they all have the same area within the rectangular pulses.

In Figure 9 we plot the upper bounds of the performance for the different approximation methods. The upper two curves are due to Equation (52). If in addition the input is known to be smooth, we can perform all the inner products and norms in the order one Sobolev space. As a result, the value of the upper bound B_{mx} changes, and

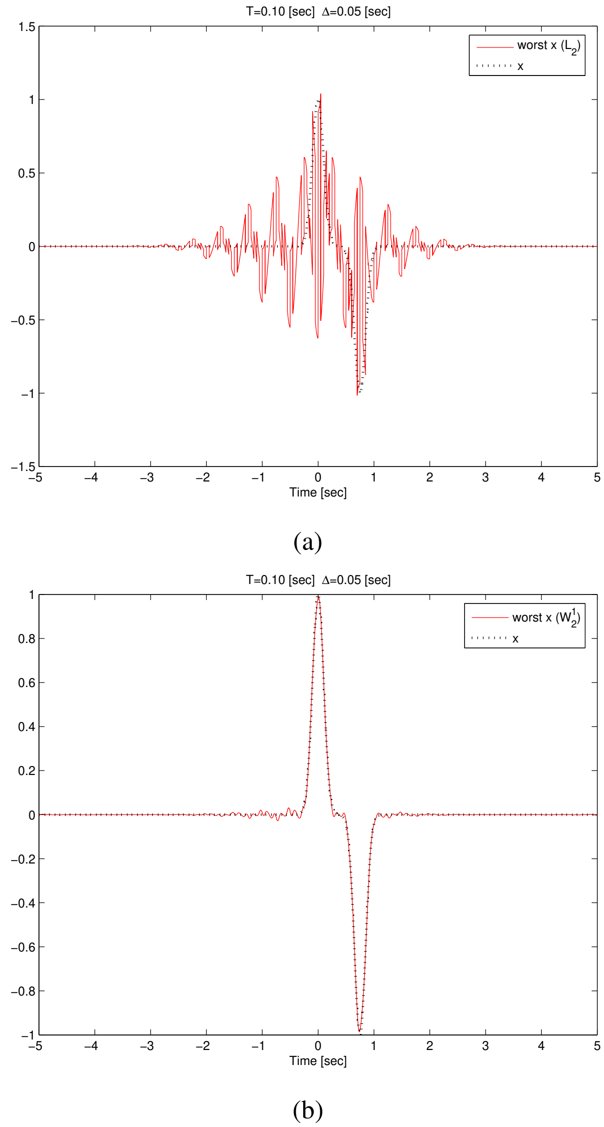


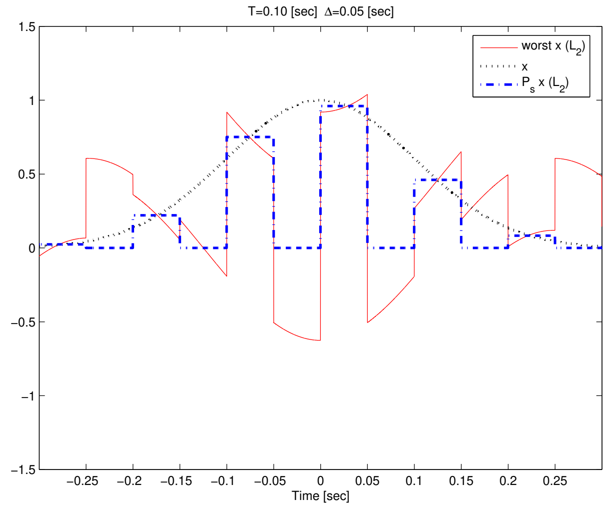
Fig. 7. The original input and the worst possible counterpart in (a) L_2, (b) Sobolev space of order one.

so is B_{sum} (the lower two curves of Figure 9).

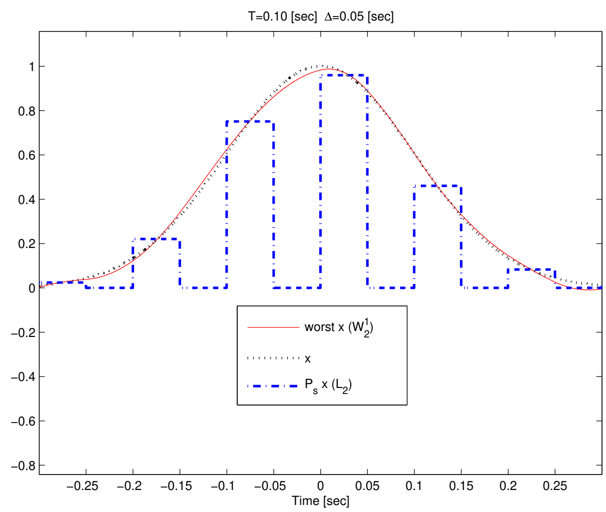
Note that those upper bounds are obtained by the worst possible inputs plotted in Figure 7. Specifically, the signal at the bottom part of Figure 7 achieves the lower two curves of Figure 9 (with the lowest curve for the minimax method with the smoothness constraints, and the one above it for the sum approach). Similarly, when smoothness is not of concern, the signal at the top part of Figure 7 achieves the top two error bound curves of Figure 9 (with the higher curve for the sum approximation).

B. RC Sampling

As an additional example, suppose that the acquisition device is a low pass RC circuit, followed by an ideal sampler with interval T (Figure 10). Here, the frequency response of the acquisition filter is given by $1/(1+j\omega RC)$,



(a)



(b)

Fig. 8. A section of $\mathbf{x}$ and the worst possible counterpart in (a) L_2, (b) Sobolev space of order one. Both are plotted against $P_{\mathcal{S}}\mathbf{x}$ to describe the integration zones.

and the n'th sampling vector is shifted and mirrored version of the impulse response

$$\mathbf{s}_n(t) = \begin{cases} (RC)^{-1}e^{\frac{t-nT}{RC}}, & t \leq nT; \\ 0, & \text{otherwise.} \end{cases} \tag{59}$$

Figure 11 is similar to Figure 7, while considering the RC circuit sampling function (59), with the time constant $RC = 0.5$. Here as well, the sampling functions posses discontinuities, giving rise to a non-smooth worst-case function, as shown in the top part of Figure 11. When we expect the input to be smooth, the minimax objective with the smoothness constraint can be used. For such a criterion, the worst-case input function behaves accordingly (part (b) of Figure 11).

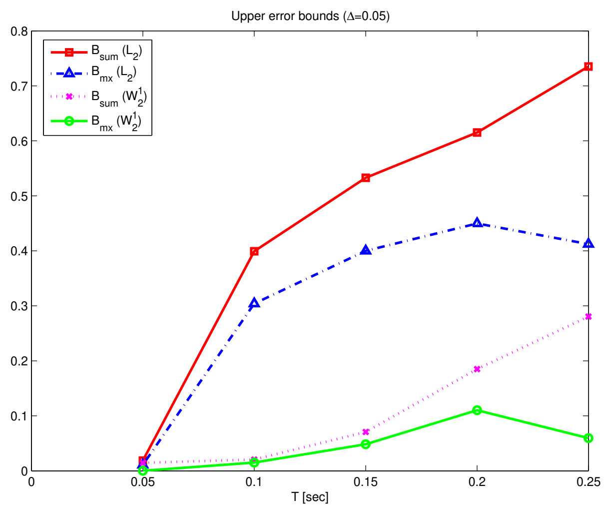


Fig. 9. Upper error bounds according to equation (52). The sampling functions are given by (58).

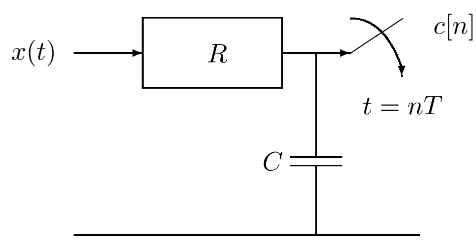
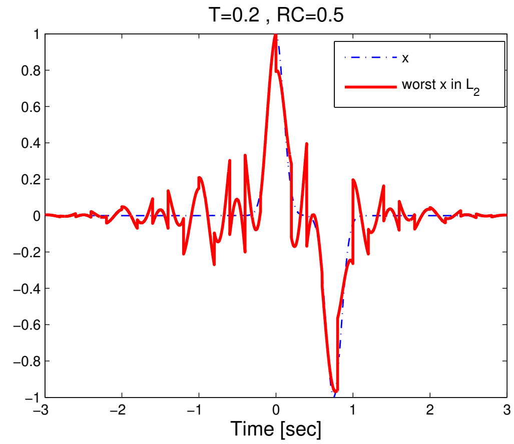


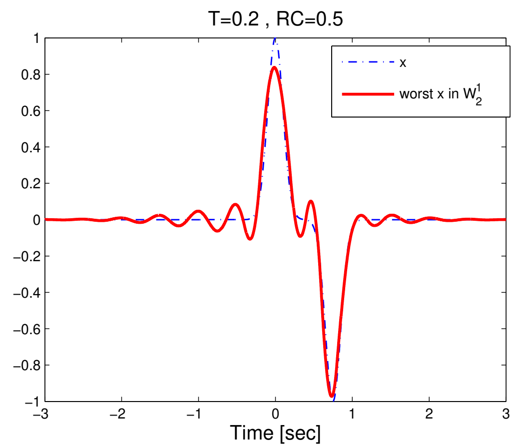
Fig. 10. An RC circuit, followed by ideal sampler, serves as the acquisition device.

Figure 12 shows the approximation error for the input $\mathbf{x}(t)$ of (56). Since input (56) is a smooth function, imposing the smoothness constraint indeed improves the performance of the minimax methods. Here as well, the proposed robust criterions do not always outperform the Rieman sum approximation (part (b) of Figure 12). However, by considering the worst possible input, the superiority of the minimax methods is guaranteed. In Figure 13 we show the upper error bounds for several values of T and RC. As expected, the robust approaches outperform the sum counterpart. Additionally, when we restrict the set of possible inputs to order one Sobolev functions, the worst case errors are smaller. As with the previous simulation, the presented error bounds are tight. For example, the worst case inputs of Figure 11 achieve the error bounds in the top part of Figure 13.

As a final remark, note that the worst-case signal is dependent on the sampling and analysis functions (53). Therefore, when either of them is non-smooth, the worst-case function might be non-smooth as well, being the sum of functions with discontinuities. As a result, if we have prior knowledge that the input $\mathbf{x}(t)$ is smooth, it is recommended to implement the minimax solution with the smoothness constraint.



(a)

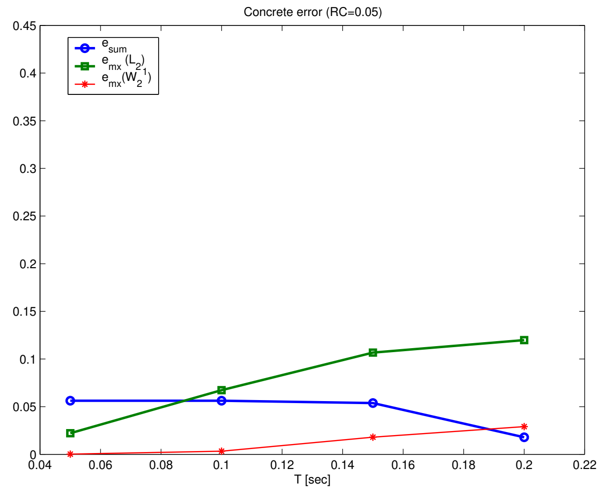


(b)

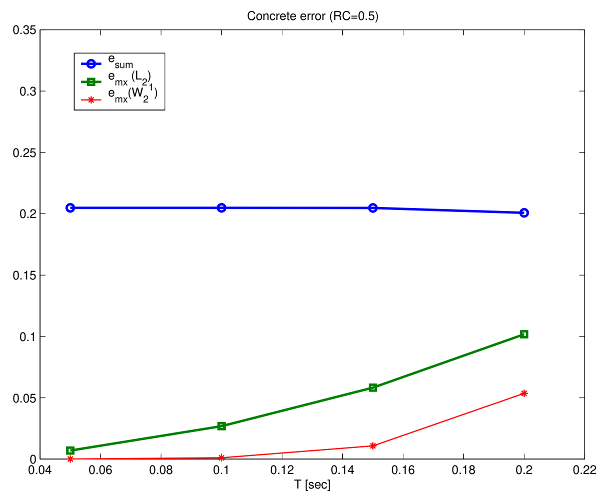
Fig. 11. The original input and the worst possible counterpart in (a) L_2, (b) Sobolev space of order one. The sampling functions are given by (59).

VIII. SUMMARY

A minimax approach has been introduced for approximating inner-product calculations within the continuous-time domain, while having the generalized samples of the signal as the only available data. We have shown that if the input signal is known to be a smooth function, then a smoothness constraint can be incorporated into the minimax criterion. The latter was achieved by recasting the problem into a proper Sobolev space. A comparison of our proposed robust methods with the standard Rieman sum approximation has been presented. Error bounds for the different methods were derived, showing the possible improvement by the proposed methods. The derivations presented herein extend recent results concerning the ideal sampling case, allowing for practical acquisition devices to be incorporated.



(a)



(b)

Fig. 12. Concrete approximation errors for the input (56) processed by the RC circuit. (a) $RC = 0.05$, (b) $RC = 0.5$.

APPENDIX I

PROOF OF THEOREM 1

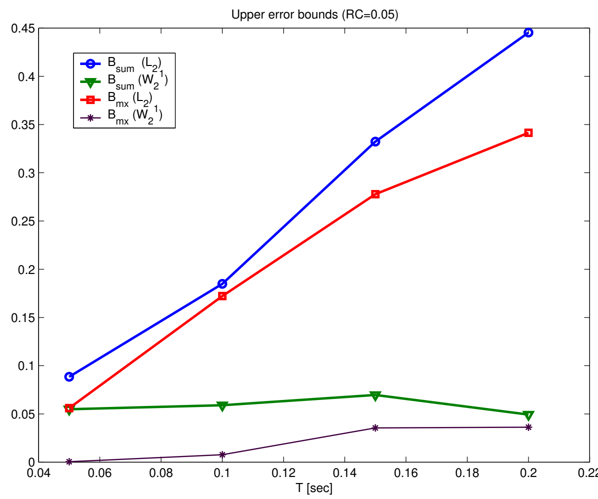
In this appendix we show that for a general $\mathbf{x} \in \mathcal{H}$, satisfying $c = S^*\mathbf{x}$, it is possible to obtain the required inner products $W^*\mathbf{x}$ if and only if $\mathcal{W} \subseteq \mathcal{S}$. The proof of this claim is similar to the proof of a sampling problem, considered in [16, Sec. 3]. For completeness, we bring the derivations below.

Assume $\mathcal{W} \subseteq \mathcal{S}$ and let $d = Gc$ where

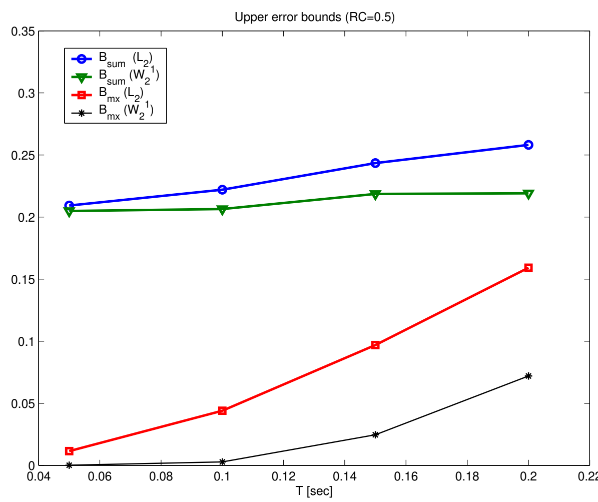
$$G = W^*S\,(S^*S)^\dagger\,. \tag{60}$$

We now show that $d = W^*\mathbf{x}$. Indeed, since for any function $\mathbf{f}$, $W^*\mathbf{f} = W^*P_{\mathcal{W}}\mathbf{f}$ we have $Gc = W^*P_{\mathcal{W}}S\,(S^*S)^\dagger\,c$. Substituting $c = S^*\mathbf{x}$,

$$d = Gc = W^*P_{\mathcal{W}}P_{\mathcal{S}}\mathbf{x} = W^*P_{\mathcal{W}}\mathbf{x} = W^*\mathbf{x}, \tag{61}$$



(a)



(b)

Fig. 13. Upper error bounds. The sampling functions are given by (59) with (a) $RC = 0.05$, (b) $RC = 0.5$.

where we used the fact that $P_{\mathcal{W}} P_{\mathcal{S}} = P_{\mathcal{W}}$ since $\mathcal{W} \subseteq \mathcal{S}$.

Now, assume that $\mathcal{W} \nsubseteq \mathcal{S}$ and suppose that there exits a transformation $d = G(c)$ achieving $d = W^* \mathbf{x}$. Consider the signal $\mathbf{x}$ defined by $\mathbf{x} = \mathbf{x}_{\mathcal{S}^\perp} + \mathbf{x}_{\mathcal{W}}$ where $\mathbf{x}_{\mathcal{S}^\perp}$ is in $\mathcal{S}^\perp$ but not in $\mathcal{W}^\perp$ (such a function always exists since $\mathcal{W} \nsubseteq \mathcal{S}$) and $\mathbf{x}_{\mathcal{W}} \in \mathcal{W}$. For this choice, $c = S^* \mathbf{x} = S^* \mathbf{x}_{\mathcal{W}}$ but $W^* \mathbf{x} - W^* \mathbf{x}_{\mathcal{W}} = W^* \mathbf{x}_{\mathcal{S}^\perp} \neq 0$. Since we assumed $W^* \mathbf{x} = G(S^* \mathbf{x})$ and $W^* \mathbf{x}_{\mathcal{W}} = G(S^* \mathbf{x}_{\mathcal{W}})$ we also have

$$W^* \mathbf{x}_{\mathcal{S}^\perp} = G(S^* \mathbf{x}) - G(S^* \mathbf{x}_{\mathcal{W}}) = 0, \tag{62}$$

which implies that $\mathbf{x}_{\mathcal{S}^\perp} \in \mathcal{W}^\perp$, contradicting our assumption. $\square$

APPENDIX II

ERROR BOUNDS FOR THE SCALAR CASE

In this appendix we prove (52), (53) and (54).

To show (52) note that for a single representation coefficient we have

$$\begin{aligned} |e_{mx}| &= |\langle \mathbf{w}, P_{\mathcal{S}^\perp} x\rangle| \\ &= |\langle P_{\mathcal{S}^\perp}\mathbf{w}, P_{\mathcal{S}^\perp} x\rangle| \\ &\leq \|P_{\mathcal{S}^\perp}\mathbf{w}\|\, L', \end{aligned} \tag{63}$$

where we used the Cauchy-Schwartz inequality and the norm constraint $\|P_{\mathcal{S}^\perp}x\| \leq L'$, with L' given by (50). The bound is tight, since

$$\mathbf{x} = S(S^*S)^\dagger c + \frac{P_{\mathcal{S}^\perp}\mathbf{w}}{\|P_{\mathcal{S}^\perp}\mathbf{w}\|} L' \tag{64}$$

is a valid input which achieves (63) with equality. Similarly, we can bound the error due to the sum method. Using (48),

$$|e_{sum}| \leq |a| + |e_{mx}| \leq |a| + \|P_{\mathcal{S}^\perp}\mathbf{w}\|\, L'. \tag{65}$$

This upper bound is obtained by setting $\mathbf{x} = \mathbf{x}_{worst}$ as in (53). In fact, the signal $\mathbf{x}_{worst}$ of (53) also achieves the maximal bound in (63). Thus, there is a valid input which makes both the sum and the minimax methods to operate as worst as possible.

To prove (54), we must find a sufficient condition that ensures that the lower bound B_L of (46) is positive. Using (46) and (63) we have that $B_L \geq |a|^2 - 2\,|a|\,\|P_{\mathcal{S}^\perp}\mathbf{w}\|\, L'$ from which (54) follows.

REFERENCES

[1] H. G. Feichtinger and T. S. (Ed.), *Advances in Gabor Analysis*. Birkhäuser, Boston, 2003.

[2] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

[3] G. G. Walter, *Wavelets and Other Orthogonal Systems With Applications*. CRC Press, Boca Raton, FL., 1994.

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

[5] S. G. Mallat, “A theory of multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, 1989.

[6] G. Strang and T. Nguyen, *Wavelet and Filter Banks*. Cambridge, U.K.: Wellesley-Cambridge, 1997.

[7] P. Qian and B. A. Francis, “Optimal initialization of the discrete wavelet transform,” *in Proc. Workshop Recent Advances in Control*, 1998.

[8] N. Kaiblinger, “Approximation of the Fourier transform and the dual Gabor window,” *J. Fourier Anal. Appl.*, vol. 11, no. 1, pp. 25–42, 2005.

[9] T. Blu and M. Unser, “Approximation error for quasi-interpolators and (multi-)wavelet expansions,” *Applied and Computational Harmonic Analysis*, vol. 6, p. 219251, 1999.

[10] H. Kirshner and M. Porat, “On sampling invariant signal representation,” Technion, CCIT report 520, Feb. 2005, submitted to *IEEE Trans. Signal Processing*.

[11] ——, “A new approach to sampling finite energy functions and sobolev signal representation,” *SampTA05 International Conference*, Jul. 2005.

[12] R. A. Adams, *Sobolev spaces*. Academic Press, New York, NY, 1975.

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

[14] Y. C. Eldar, “Sampling and reconstruction in arbitrary spaces and oblique dual frame vectors,” *J. Fourier Analys. Appl.*, vol. 1, no. 9, pp. 77–96, Jan. 2003.

[15] M. Unser and J. Zerubia, “Generalized sampling: Stability and performance analysis,” *IEEE Trans. Signal Processing*, vol. 45, no. 12, pp. 2941–2950, Dec. 1997.

[16] Y. C. Eldar and T. Dvorkind, “A minimum squared-error framework for sampling and reconstruction in arbitrary spaces,” to appear in *IEEE Trans. Signal Processing*.

[17] I. Gohberg and S. Goldberg, *Basic operator theory*. Birkhäuser, 1981.

[18] D. Luenberger, *Optimization by vector space methods*. John Wiley and Sons, Inc., 1969.

[19] M. Fornasier, “Function spaces inclusion and rate of convergence of Riemann-type sums in numerical integration,” *Numerical Functional Analysis and Opt.*, vol. 24, no. 1,2, pp. 45–57, 2003.

[20] A. Ben-Israel and T. N.E.Greville, *Generalized Inverses: Theory and Applications*. John Wiley and Sons, Inc., 1974.

[21] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA: Birkhäuser, 2002.

[22] M. V. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 50, pp. 1417–1428, June 2002.

[23] S. Kayalar and H. L. Weinert, “Oblique projections: Formulas, algorithms, and error bounds,” *Math. Contr. Signals Syst.*, vol. 2, no. 1, pp. 33–45, 1989.

[24] H. Feichtinger and T. Werther, *Robustness of minimal norm interpolation in Sobolev algebras*. Birkhäuser, 2004, in: Sampling, Wavelets, and Tomography, eds. J.J Benedetto and A. Zayed, , 83-113.

[25] W. S. Tang, “Oblique projections, biorthogonal Riesz bases and multiwavelets in Hilbert space,” *Proc. Amer. Math. Soc.*, vol. 128, no. 2, pp. 463–473, 2000.