OPTIMAL SQUARED-ERROR SIGNAL RECOVERY FROM NONIDEAL SAMPLES

Yonina C. Eldar and Tsvi G. Dvorkind

Department of Electrical Engineering
Technion–Israel Institute of Technology, Haifa, Israel
Email: {yonina@ee,dvorkind@tx}.technion.ac.il.

ABSTRACT

We treat the problem of reconstructing a signal from its non-ideal samples where the sampling and reconstruction spaces as well as the class of input signals can be arbitrary subspaces of a Hilbert space. If the signal is known to lie in an appropriately chosen subspace, then we propose a method that achieves the minimal squared-error approximation. In the general case, we show that the minimal-error reconstruction cannot usually be obtained. Instead, we suggest minimizing the worst-case squared-error between the reconstructed signal, and the best possible (but usually unattainable) approximation of the signal, over all signals that yield the given samples. Interestingly, the optimal method turns out to be linear, and coincides with a recently proposed suboptimal approach for this problem.

1. INTRODUCTION

Digital signal processing entails representing a signal by a set of coefficients and relies on the existence of methods for reconstructing the signal from its samples. A recent approach to sampling and reconstruction is to consider a generalized sampling scheme, in which the samples are represented as the inner products of the input signal x with a set of sampling vectors (associated with the acquisition device), which span the sampling space $\mathcal{S}$ [1, 2, 3, 4, 5]. Reconstruction is obtained by taking linear combinations of a set of reconstruction vectors that span the reconstruction space $\mathcal{W}$. This framework is quite general and includes the conventional Shannon-Whittaker paradigm as a special case.

Since in this setting the reconstructed signal is constrained to lie in $\mathcal{W}$, if x is not in $\mathcal{W}$ to begin with, then perfect reconstruction cannot be obtained, regardless of the sampling and reconstruction method. Our problem then is to process the samples prior to reconstruction such that the reconstructed signal $\hat{x}$ is close to x in some sense. In our setup, the only constraints we impose are that the sampling is linear and bounded, and the reconstruction is constrained to a subspace $\mathcal{W}$ of an arbitrary Hilbert space $\mathcal{H}$. However, we do not require any specific constrains on the spaces involved.

The framework we consider here was first introduced in the context of shift-invariant spaces in [1], in which a consistent approximation method was proposed. In this approach, the reconstructed signal is designed to yield the same samples as the original signal x. This strategy was later extended to a more general setting in [3, 5, 6]. Under a direct-sum condition on the spaces, the consistent reconstruction is given by $\hat{x} = E_{\mathcal{W}\mathcal{S}^{\perp}}x$ where $E_{\mathcal{W}\mathcal{S}^{\perp}}$ is the oblique projection onto $\mathcal{W}$ along the orthogonal compliment of $\mathcal{S}$. Note, however, that the fact that x and $\hat{x}$ yield the same samples does not necessarily imply that $\hat{x}$ is close to x. In fact, for an input x not in $\mathcal{W}$, the norm of the resulting reconstruction error $\hat{x} - x$ can be made arbitrarily large, if $\mathcal{S}$ is close to $\mathcal{W}^{\perp}$.

To ensure that the reconstruction $\hat{x}$ is close to x for all choices of $\mathcal{S}$ and $\mathcal{W}$, we may try to minimize the squared-norm of the reconstruction error $\hat{x} - x$. If the reconstruction space $\mathcal{W}$ is contained in the sampling space $\mathcal{S}$, then by proper pre-processing of the samples the minimal squared-error approximation of x in the space $\mathcal{W}$, given by the orthogonal projection $P_{\mathcal{W}}x$ onto $\mathcal{W}$, can be obtained. However, as we show in Section 3, if $\mathcal{S}$ does not contain $\mathcal{W}$, then the squared-error cannot be minimized over the entire space $\mathcal{H}$ of input signals. In Section 4 we consider the case in which x is known to lie in an appropriately chosen subspace, and show that the minimal squared-error reconstruction can be obtained over all x in the subspace using a linear pre-processing method.

Recently, an alternative method was proposed [7], aimed at reducing the error between the reconstructed signal $\hat{x}$ and the best approximation to x in $\mathcal{W}$. For simplicity, it was suggested to constrain the reconstruction to be linear. Furthermore, the proposed method aimed to minimize the worst-case squared difference between x and the best possible approximation $P_{\mathcal{W}}x$, for all choices of x. This approach does not take the given samples into account, but rather considers the worst-case error over all possible values of the input signal x, although some choices of x are not compatible with the samples. The resulting reconstruction is given by the double orthogonal projection $\hat{x} = P_{\mathcal{W}}P_{\mathcal{S}}x$, where $P_{\mathcal{S}}$ ($P_{\mathcal{W}}$) is the orthogonal projection onto $\mathcal{S}$ ($\mathcal{W}$). It was also shown in [7] that this reconstruction can lead to a smaller squared-norm error than the consistent reconstruction method.

Here, we develop a more general formulation of the problem where we allow for nonlinear reconstruction methods, and we also take the prior information of the given samples into account. Thus, we seek the possibly nonlinear reconstruction that minimizes the worst-case error, where now the worst-case is with respect to the signal values x that are consistent with the given samples. Interestingly, we show that the optimal solution is linear, and coincides with the previous suboptimal approach to this problem: $\hat{x} = P_{\mathcal{W}}P_{\mathcal{S}}x$. Our results provide a stronger optimality property of this choice of reconstruction leading to further justification for its use.

2. PROBLEM FORMULATION

2.1 Sampling Formulation

We denote vectors in an arbitrary Hilbert space $\mathcal{H}$ by lowercase letters, and the elements of a sequence $c \in \ell_2$ by $c[n]$. The operator $P_{\mathcal{A}}$ represents the orthogonal projection onto a closed subspace $\mathcal{A}$ of $\mathcal{H}$, $\mathcal{A}^{\perp}$ is the orthogonal complement of $\mathcal{A}$, and $\mathcal{N}(\cdot)$ and $\mathcal{R}(\cdot)$ are the null space and range space of the corresponding transformation, respectively. The *Moore-Penrose pseudo inverse* and the adjoint of a bounded transformation T are written as $T^{\dagger}$ and T^{*}, respectively. The inner product between vectors $x, y \in \mathcal{H}$ is denoted by $\langle x, y \rangle$, and is linear in the second argument, and $\|x\|^2 = \langle x, x \rangle$ is the squared norm of x. The direct sum between two closed subspaces $\mathcal{W}$ and $\mathcal{S}$ is written as $\mathcal{W} \oplus \mathcal{S}$, and is the sum set $\{w + v;\ w \in \mathcal{W}, v \in \mathcal{S}\}$ with the property that $\mathcal{W} \cap \mathcal{S} = \{0\}$. The oblique projection onto $\mathcal{W}$ along $\mathcal{S}^{\perp}$ is denoted by $E_{\mathcal{W}\mathcal{S}^{\perp}}$, and is defined as the unique projection with $\mathcal{R}(E_{\mathcal{W}\mathcal{S}^{\perp}}) = \mathcal{W}$ and $\mathcal{N}(E_{\mathcal{W}\mathcal{S}^{\perp}}) = \mathcal{S}^{\perp}$. A set transformation $V : \ell_2 \rightarrow \mathcal{H}$ corresponding to frame vectors $\{v_n\} \in \mathcal{H}$ is defined by $Va = \sum_n a[n]v_n$ for all $a \in \ell_2$. From the definition of the adjoint, if $a = V^{*}y$, then $a[n] = \langle v_n, y \rangle$.

We consider a general sampling problem in a Hilbert space $\mathcal{H}$, in which the goal is to reconstruct a signal $x \in \mathcal{H}$ from a sequence of samples $\{c[n]\}$. Our formulation of the problem allows for a broad class of sampling strategies where the basic constraint we

This work was supported by the European Union's Human Potential Programme, under the contract HPRN-CT-2003-00285 (HASSIP).

impose on the sampling process is that it is linear. The samples are modelled as the inner products of the signal x with a set of sampling vectors $\{s_n\}$ that span a space $\mathscr{S}$, so that $c[n] = \langle s_n, x\rangle$. Denoting by S the set transformation corresponding to the vectors $\{s_n\}$, the samples can be written as $c = S^*x$. The problem is to reconstruct x from c, where the reconstruction $\hat{x}$ of x has the form

$$\hat{x} = \sum_n d[n]w_n = Wd \tag{1}$$

for some coefficients $d = H(c)$ that are a (possibly nonlinear) transformation of c. Here W is the set transformation corresponding to a set of vectors $\{w_n\}$ that span the reconstruction space $\mathscr{W}$. The sampling and reconstruction scheme is illustrated in Fig. 1.

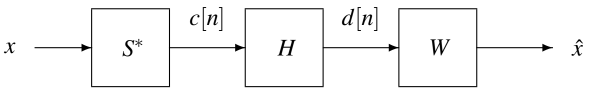

Figure 1: General sampling and reconstruction scheme.

A special case of Fig. 1 is when $\{s_n = s(t-n)\}$ and $\{w_n = w(t-n)\}$ are vectors corresponding to uniform shifts of generators in L_2. In this setting, the sampling and reconstruction scheme of Fig. 1 can be formulated in terms of linear-time invariant (LTI) filters.

If x is in $\mathscr{W}$, and $\mathscr{W}$ and $\mathscr{S}^\perp$ satisfy the direct-sum condition $\mathscr{H} = \mathscr{W} \oplus \mathscr{S}^\perp$, then it was shown in [3, 5, 6] that x can be perfectly reconstructed from the samples $c[n]$ by choosing $H(c) = (S^*W)^\dagger c$. With this choice of H it follows from (4) below that $\hat{x} = E_{\mathscr{W}\mathscr{S}^\perp}x$. If $x \notin \mathscr{W}$, then the reconstruction $\hat{x} = E_{\mathscr{W}\mathscr{S}^\perp}x$ is no longer equal to x, but has the property that it is a *consistent reconstruction* [1], namely, it yields the same samples as x: $S^*x = S^*\hat{x}$. However, the fact that x and $\hat{x}$ have the same samples, does not guarantee that $\hat{x}$ is close to x. In fact, using the relation $x = E_{\mathscr{W}\mathscr{S}^\perp}x + E_{\mathscr{S}^\perp\mathscr{W}}x$ we can express the reconstruction error as $\hat{x} - x = E_{\mathscr{S}^\perp\mathscr{W}}x$, which can have arbitrarily large norm if $\mathscr{S}$ is close to $\mathscr{W}^\perp$. Therefore, our problem is to choose the transformation H in Fig. 1 such that $\hat{x}$ is a good approximation of x.

In [7] it was suggested to seek a linear reconstruction $\hat{x} = WHc$, where H is chosen as the solution to the problem

$$\min_H \max_{\|x\|\leq L} \|WHS^*x - P_{\mathscr{W}}x\|^2. \tag{2}$$

Here, L is an arbitrary constant that does not effect the solution. The objective in (2) measures the difference between the reconstruction with H when the true signal is x, and the best approximation to x in $\mathscr{W}$, over all bounded norm signals. Note, however, that it does not take the given information $c = S^*x$ into account.

In the following sections we propose different strategies for designing H which attempt to control the squared-norm of the reconstruction error $\hat{x} - x$ while directly considering the given samples $c = S^*x$. In the first approach we take advantage of prior information on x in the form of inclusion into a properly chosen subspace. Using this knowledge will allow us to directly minimize the squared-error, as we show in Section 4. The second strategy, considered in Section 5, treats the squared-error criterion over the entire space, and minimizes a worst-case error measure over the set of inputs x that are compatible with the given samples. The solution turns out to be linear, and coincides with the solution to (2).

2.2 Mathematical Preliminaries

To ensure that the sampling is stable we choose the vectors $\{s_n\}$ and $\{w_n\}$ such that they form frames for their closed span, which we denote by $\mathscr{S}$ and $\mathscr{W}$ respectively.

Definition 1 ([8]). A family of vectors $\{h_n\}$ in a Hilbert space $\mathscr{H}$ is called a frame for a subspace $\mathscr{A} \subseteq \mathscr{H}$ if there exist constants $0 < A \leq B < \infty$ such that for all $y \in \mathscr{A}$,

$$A\|y\|^2 \leq \sum_n |\langle y, h_n\rangle|^2 \leq B\|y\|^2. \tag{3}$$

Note that any finite set of vectors that spans $\mathscr{A}$ is a frame for $\mathscr{A}$.

If $\{s_n\}$ forms a frame for $\mathscr{S}$, then $c[n] = \langle s_n, x\rangle$ is in ℓ_2 for any signal x that has bounded norm, and therefore the sampling process is stable. Furthermore, S is bounded and $\mathscr{R}(S) = \mathscr{S}$. Similarly, if the vectors $\{w_n\}$ form a frame, then $\mathscr{R}(W) = \mathscr{W}$ and the sum $\sum_n d[n]w_n$ converges for any sequence $d \in \ell_2$ [8].

A useful result on set transformations is given below.

Lemma 1 (Lemma 3.3 [9]). *Let $S : \ell_2 \to \mathscr{H}$ and $W : \ell_2 \to \mathscr{H}$ be bounded transformations on $\mathscr{H}$ with $\mathscr{R}(S) = \mathscr{S}$ and $\mathscr{R}(W) = \mathscr{W}$, where $\mathscr{H} = \mathscr{W} \oplus \mathscr{S}^\perp$. Then*

1. *$\mathscr{N}(S^*W) = \mathscr{N}(W)$;*
2. *$(S^*W)^\dagger$ is a bounded operator from ℓ_2 to ℓ_2;*
3. *$(S^*W)^\dagger S^*W$ is the orthogonal projection onto $\mathscr{N}(W)^\perp$.*

Using part 2 of the lemma, $E_{\mathscr{W}\mathscr{S}^\perp}$ can be expressed as [6],[9]

$$E_{\mathscr{W}\mathscr{S}^\perp} = W\,(S^*W)^\dagger\,S^* \tag{4}$$

where S and W are bounded transformations with $\mathscr{R}(S) = \mathscr{S}$ and $\mathscr{R}(W) = \mathscr{W}$. As a special case,

$$P_{\mathscr{W}} = W\,(W^*W)^\dagger\,W^*. \tag{5}$$

3. MINIMAL SQUARED-ERROR RECONSTRUCTION

A straightforward strategy to designing a reconstruction that is close to x is to minimize the squared-error $\|\hat{x} - x\|^2$. In this approach, the transformation H is the solution to the problem

$$\min_H \|\hat{x} - x\|^2 = \min_H \|WH(S^*x) - x\|^2. \tag{6}$$

For any choice of x,

$$\|WH(S^*x) - x\|^2 = \|WH(S^*x) - P_{\mathscr{W}}x\|^2 + \|P_{\mathscr{W}^\perp}x\|^2 \geq \|P_{\mathscr{W}^\perp}x\|^2. \tag{7}$$

In the special case in which $\mathscr{W} \subseteq \mathscr{S}$, the bound (7) can be achieved by a fixed, linear transformation H defined by

$$H(c) = (W^*W)^\dagger W^*S(S^*S)^\dagger c. \tag{8}$$

Indeed, with this choice of $H(c)$,

$$\hat{x} = WH(S^*x) = W(W^*W)^\dagger W^*S(S^*S)^\dagger S^*x = P_{\mathscr{W}}P_{\mathscr{S}}x = P_{\mathscr{W}}x, \tag{9}$$

where we used the representation (5) of $P_{\mathscr{W}}$ and $P_{\mathscr{S}}$, and the last equality follows from the fact that $\mathscr{W} \subseteq \mathscr{S}$. However, as we now show, when $\mathscr{W}$ is not contained in $\mathscr{S}$, the lower bound cannot be achieved for all x with a transformation $H(c)$ that depends only on the given samples $c = S^*x$ and not directly on x.

Proposition 1. *Let $H : \ell_2 \to \ell_2$ be any solution to*

$$\min_H \|\hat{x} - x\|^2 = \min_H \|WH(S^*x) - x\|^2,$$

where W and S are bounded transformations with $\mathscr{R}(W) = \mathscr{W}$, $\mathscr{R}(S) = \mathscr{S}$, and $\mathscr{W} \nsubseteq \mathscr{S}$. Then for arbitrary choices of x, $H(S^*x)$ cannot achieve the lower bound of (7).

Proof. To prove the proposition, suppose to the contrary that there exits a solution $H(c)$ that depends only on the available samples $c = S^*x$. Consider the signal x defined by $x = x_{\mathscr{S}^\perp} + x_{\mathscr{W}}$ where $x_{\mathscr{S}^\perp}$ is in $\mathscr{S}^\perp$ but not in $\mathscr{W}^\perp$ (such a vector always exists since $\mathscr{W} \nsubseteq \mathscr{S}$) and $x_{\mathscr{W}} \in \mathscr{W}$. For this choice, $S^*x = S^*x_{\mathscr{W}} = c$ so that

$$WH(S^*x) = WH(S^*x_{\mathscr{W}}). \tag{10}$$

On the other hand, since H achieves the lower bound in (7), $WH(S^*x) = P_{\mathscr{W}}x$ and $WH(S^*x_{\mathscr{W}}) = P_{\mathscr{W}}x_{\mathscr{W}} = x_{\mathscr{W}}$ which implies that $P_{\mathscr{W}}x_{\mathscr{S}^\perp} = 0$, or $x_{\mathscr{S}^\perp} \in \mathscr{W}^\perp$, contradicting our assumption. $\square$

To circumvent the problem associated with minimizing the squared-error we develop two strategies which differ in their assumptions on x. In the first approach x is assumed to lie in a subspace, which will allow us to directly minimize the squared-error. In the second approach, we eliminate the dependency of the squared-error on x by considering a worst-case error measure.

4. RECONSTRUCTION ON A SUBSPACE

We have seen in Proposition 1 that if $\mathscr{W} \nsubseteq \mathscr{S}$, then the lower bound in (7) cannot be achieved *for all* $x \in \mathscr{H}$ with a transformation that depends only on the given samples. However, this does not preclude the possibility that a fixed transformation achieves this bound when restricting attention to a subset of input signals. Indeed, if we consider only signals $x \in \mathscr{W}$, then we showed that under the direct-sum condition $\mathscr{H} = \mathscr{W} \oplus \mathscr{S}^\perp$, perfect reconstruction (which is therefore the minimal-error reconstruction) is possible with $H(c) = (S^*W)^\dagger c$. We now generalize this result to a broader class of input signals.

Suppose that x lies in a subspace $\mathscr{A}$ satisfying $\mathscr{H} = \mathscr{A} \oplus \mathscr{S}^\perp$. Theorem 1 below shows that in this case the minimal error reconstruction can be achieved using a fixed linear transformation.

Theorem 1. *Consider the problem*

$$\min_H \|\hat{x} - x\|^2 = \min_H \|WH(S^*x) - x\|^2, \quad x \in \mathscr{A}$$

where $\mathscr{A} \subseteq \mathscr{H}$ is an arbitrary subspace such that $\mathscr{H} = \mathscr{A} \oplus \mathscr{S}^\perp$ and W, S are bounded set transformations with $\mathscr{R}(W) = \mathscr{W}, \mathscr{R}(S) = \mathscr{S}$. A possible solution is $H(c) = H_{\mathscr{A}}c$ where

$$H_{\mathscr{A}} = (W^*W)^\dagger W^*A(S^*A)^\dagger, \tag{11}$$

and A is any bounded transformation with $\mathscr{R}(A) = \mathscr{A}$. The resulting reconstruction $\hat{x}$ is the minimal-error solution $\hat{x} = P_{\mathscr{W}}x$.

Before proving the theorem, we note that from Lemma 1 $(S^*A)^\dagger$ is a well defined bounded operator. Furthermore, it is shown in [9] that $A(S^*A)^\dagger$ in (11) is independent of the choice of the bounded transformation $A : \ell^2 \to \mathscr{H}$, as long as $\mathscr{R}(A) = \mathscr{A}$.

Proof. We begin by noting that since $x \in \mathscr{A}$, it can be expressed as $x = Ay$ for some vector $y \in \mathscr{N}(A)^\perp$. In addition, we know that

$$c = S^*x = S^*Ay. \tag{12}$$

Multiplying both sides of (12) by $(S^*A)^\dagger$ and using Lemma 1,

$$(S^*A)^\dagger c = (S^*A)^\dagger(S^*A)y = P_{\mathscr{N}(A)^\perp}y = y. \tag{13}$$

We conclude that the only vector in $\mathscr{A}$ with samples given by c is the vector

$$x = Ay = A(S^*A)^\dagger c, \tag{14}$$

so that given c we can reconstruct the vector x exactly. Once we know x, the approximation in $\mathscr{W}$ minimizing the squared-error is

$$\hat{x} = P_{\mathscr{W}}x = P_{\mathscr{W}}A(S^*A)^\dagger c = WH_{\mathscr{A}}c, \tag{15}$$

where we used the representation (5) of $P_{\mathscr{W}}$. Finally, since $c = S^*x$,

$$\hat{x} = P_{\mathscr{W}}A(S^*A)^\dagger S^*x = P_{\mathscr{W}}E_{\mathscr{A}\mathscr{S}^\perp}x = P_{\mathscr{W}}x, \tag{16}$$

where we used the fact that from (4), $A(S^*A)^\dagger S^* = E_{\mathscr{A}\mathscr{S}^\perp}$, and since $x \in \mathscr{A}$, $E_{\mathscr{A}\mathscr{S}^\perp}x = x$. $\square$

A special case of Theorem 1 is when $\mathscr{A} = \mathscr{S}$, for which

$$H_{\mathscr{A}} = (W^*W)^\dagger W^*S(S^*S)^\dagger. \tag{17}$$

In Theorem 2 below we will see that this solution is equivalent to the minimax transformation. This implies that the minimax approach minimizes the squared-error over all $x \in \mathscr{S}$.

As another example, suppose that $\mathscr{H} = \mathscr{W} \oplus \mathscr{S}^\perp$, and let $\mathscr{A} = \mathscr{W}$. With this choice,

$$H_{\mathscr{A}} = (W^*W)^\dagger W^*W(S^*W)^\dagger = P_{\mathscr{N}(\mathscr{W})^\perp}(S^*W)^\dagger = (S^*W)^\dagger, \tag{18}$$

where we used the fact that $\mathscr{R}\left((S^*W)^\dagger\right) = \mathscr{N}(S^*W)^\perp = \mathscr{N}(W)^\perp$; the last equality follows from Lemma 1. Thus, $H_{\mathscr{A}}$ is equal to the consistent reconstruction transformation, which agrees with the fact that the consistent strategy minimizes the error over all $x \in \mathscr{W}$.

4.1 Geometric Interpretation

We now consider a geometric interpretation of our results. We first note that sampling x with sampling vectors in $\mathscr{S}$, is equivalent to sampling the orthogonal projection of x onto $\mathscr{S}$, denoted by $x_{\mathscr{S}} = P_{\mathscr{S}}x$. This follows from the relation

$$\langle s_n, x\rangle = \langle P_{\mathscr{S}}s_n, x\rangle = \langle s_n, P_{\mathscr{S}}x\rangle. \tag{19}$$

Since $x_{\mathscr{S}} \in \mathscr{S}$ and the vectors $\{s_n\}$ span $\mathscr{S}$, $x_{\mathscr{S}}$ is uniquely determined by the samples $c[n]$. Therefore, knowing $c[n]$ is equivalent to knowing $x_{\mathscr{S}}$. The reconstruction problem then becomes that of reconstructing a signal in $\mathscr{A}$ from its orthogonal projection $x_{\mathscr{S}}$ onto a subspace $\mathscr{S}$ of $\mathscr{H}$. In Fig. 2 we illustrate the fact that there is only one vector in $\mathscr{A}$ whose orthogonal projection onto $\mathscr{S}$ is $x_{\mathscr{S}}$. Thus, given $x_{\mathscr{S}}$, we can immediately determine the original vector x, if we know that x is in $\mathscr{A}$. In our setup we are constrained to obtain a reconstruction in $\mathscr{W}$. But, since we can determine x from $x_{\mathscr{S}}$, we can also determine its orthogonal projection onto $\mathscr{W}$, which is the minimal error reconstruction.

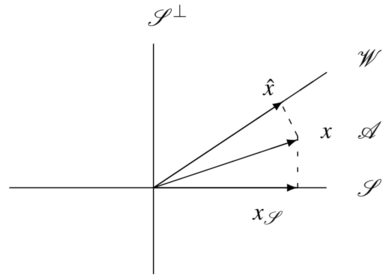

Figure 2: Illustration of minimal-error reconstruction $\hat{x} = P_{\mathscr{W}}x$ of $x \in \mathscr{A}$ from $x_{\mathscr{S}} = P_{\mathscr{S}}x$, with $\mathscr{H} = \mathscr{A} \oplus \mathscr{S}^\perp$.

5. MINIMAX SQUARED-ERROR

The previous section treated the case in which x is restricted to a subspace of $\mathscr{H}$. We now consider the general formulation of the sampling problem in which no such information on x is available.

To obtain a reconstruction that is close to the optimal approximation $P_{\mathscr{W}}x$ of x in $\mathscr{W}$, we would like to minimize the error $\|\hat{x} - P_{\mathscr{W}}x\| = \|Wd - P_{\mathscr{W}}x\|$. Since this error depends on x, which is unknown, we seek to minimize the worst-case error over all norm-bounded values of x that are consistent with our prior information $S^*x = c$. This results in the problem

$$\min_{d} \max_{c=S^*x, \|x\|\leq L} \|Wd - P_{\mathscr{W}}x\|^2, \tag{20}$$

where L is some constant; as we show, the solution does not depend on the choice of L.

Theorem 2. *Consider the problem*

$$\min_{d} \max_{c=S^*x, \|x\|\leq L} \|Wd - P_{\mathscr{W}}x\|^2,$$

where W and S are bounded set transformations with $\mathscr{R}(W) = \mathscr{W}$ and $\mathscr{R}(S) = \mathscr{S}$. A possible solution is

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

The resulting reconstruction is $\hat{x} = P_{\mathscr{W}} P_{\mathscr{S}} x$.

Proof. First we note that any x satisfying $S^*x = c$ and $\|x\| \leq L$ is of the form $x = S(S^*S)^{\dagger}c + v$ for some $v \in \mathscr{G}$ where

$$\mathscr{G} \triangleq \left\{ v \mid v \in \mathscr{S}^{\perp}, \|v\| \leq L' \right\},$$

and $L'^2 = L^2 - \left\|S(S^*S)^{\dagger}c\right\|^2$. Thus,

$$\begin{aligned} \max_{c=S^*x, \|x\|\leq L} \|Wd - P_{\mathscr{W}}x\|^2 &= \max_{v\in\mathscr{G}} \left\|Wd - P_{\mathscr{W}}S(S^*S)^{\dagger}c - P_{\mathscr{W}}v\right\|^2 \\ &= \max_{v\in\mathscr{G}} \|a_d - P_{\mathscr{W}}v\|^2 \\ &= \max_{v\in\mathscr{G}} \left\{ \|a_d\|^2 - 2\Re\{\langle a_d, P_{\mathscr{W}}v\rangle\} + \|P_{\mathscr{W}}v\|^2 \right\}, \end{aligned} \tag{21}$$

where we defined $a_d = W(d - (W^*W)^{\dagger}W^*S(S^*S)^{\dagger}c)$. Now, the maximum in (21) is achieved when

$$\Re\{\langle a_d, P_{\mathscr{W}}v\rangle\} = -|\langle a_d, P_{\mathscr{W}}v\rangle|. \tag{22}$$

Indeed, let $v \in \mathscr{G}$ be the vector for which the maximum is achieved. If $\langle a_d, P_{\mathscr{W}}v\rangle = 0$ than (22) is trivially true. Otherwise, we can define

$$v_2 \triangleq -\frac{\langle P_{\mathscr{W}}v, a_d\rangle}{|\langle a_d, P_{\mathscr{W}}v\rangle|} v. \tag{23}$$

Clearly, $\|v\| = \|v_2\|$ and $v_2 \in \mathscr{G}$. In addition, $\|P_{\mathscr{W}}v\| = \|P_{\mathscr{W}}v_2\|$ and $\langle a_d, P_{\mathscr{W}}v_2\rangle = -|\langle a_d, P_{\mathscr{W}}v\rangle|$ so that the objective in (21) at v_2 is larger than the objective at v unless (22) is satisfied.

Combining (22) and (21) our problem becomes

$$\min_{d} \max_{v\in\mathscr{G}} \left\{ \|a_d\|^2 + 2|\langle a_d, P_{\mathscr{W}}v\rangle| + \|P_{\mathscr{W}}v\|^2 \right\}. \tag{24}$$

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

$$\begin{aligned} A &\geq \max_{v\in\mathscr{G}} \min_{d} \left\{ \|a_d\|^2 + 2|\langle a_d, P_{\mathscr{W}}v\rangle| + \|P_{\mathscr{W}}v\|^2 \right\} \\ &= \max_{v\in\mathscr{G}} \|P_{\mathscr{W}}v\|^2, \end{aligned} \tag{25}$$

where we used the fact that $\|a_d\|^2 + 2\,|\langle a_d, P_{\mathscr{W}}v\rangle| \geq 0$ with equality for $a_d = 0$, or

$$d = (W^*W)^{\dagger}W^*S(S^*S)^{\dagger}c. \tag{26}$$

Thus, for any choice of d,

$$\min_{d} \max_{v\in\mathscr{G}} \|a_d - P_{\mathscr{W}}v\|^2 \geq \max_{v\in\mathscr{G}} \|P_{\mathscr{W}}v\|^2. \tag{27}$$

The proof then follows from the fact that d given by (26) achieves the lower bound (27). □

Theorem 2 establishes that the solution obtained in [7] is in fact minimax optimal over all nonlinear transformations that are compatible with the samples $c = S^*x$.

Figure 3 illustrates the minimax reconstruction geometrically. In Section 4.1 we showed that knowing the samples c is equivalent to knowledge of $P_{\mathscr{S}}x$. The reconstruction problem then becomes that of approximating an arbitrary signal in $\mathscr{H}$ from its orthogonal projection $x_{\mathscr{S}} = P_{\mathscr{S}}x$, where the reconstruction is constrained to lie in a subspace $\mathscr{W}$ of $\mathscr{H}$. Fig. 3(a) depicts the orthogonal projection of $x \in \mathscr{H}$ onto $\mathscr{S}$. The minimax reconstruction chooses the orthogonal projection of $x_{\mathscr{S}}$ onto $\mathscr{W}$, as illustrated in Fig. 3(b).

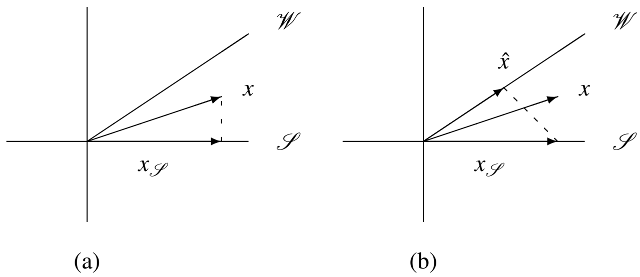

Figure 3: Illustration of minimax reconstruction of x from $x_{\mathscr{S}} = P_{\mathscr{S}}x$ (a) orthogonal projection of x onto $\mathscr{S}$ (b) minimax reconstruction $\hat{x} = P_{\mathscr{W}}P_{\mathscr{S}}x$.

In [7], the error resulting from the reconstruction $\hat{x} = P_{\mathscr{W}}P_{\mathscr{S}}x$ is analyzed and compared with the error of the consistent reconstruction method $\hat{x} = E_{\mathscr{W}\mathscr{S}^{\perp}}x$ when x can be an arbitrary signal in $\mathscr{H}$. It can be shown that if $\|P_{\mathscr{W}^{\perp}}x\|^2 \geq \gamma_1 \|P_{\mathscr{S}^{\perp}}x\|^2$ for a constant γ_1, in which case most of the energy of the signal is in the sampling space, then the minimax approach will lead to a smaller error than the consistent method. On the other hand, when $\|P_{\mathscr{W}^{\perp}}x\|^2 \leq \gamma_2 \|P_{\mathscr{S}^{\perp}}x\|^2$ for a constant γ_2, so that most of the energy of the signal is in the reconstruction space, then the consistent reconstruction method leads to a smaller squared-error.

REFERENCES

[1] 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.

[2] M. Unser, "Sampling—50 years after Shannon," *IEEE Proc.*, vol. 88, pp. 569–587, Apr. 2000.

[3] 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.

[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] Y. C. Eldar, "Sampling without input constraints: Consistent reconstruction in arbitrary spaces," in *Sampling, Wavelets and Tomography*, A. I. Zayed and J. J. Benedetto, Eds., pp. 33–60. Boston, MA: Birkhauser, 2004.

[6] T. Werther and Y. C. Eldar, "General framework for consistent sampling in Hilbert spaces," to appear in *International Journal of Wavelets, Multiresolution and Information Processing*, 2005.

[7] Y. C. Eldar and T. G. Dvorkind, "Minimax sampling with arbitrary spaces," *The 11th IEEE Int. Conf. on Electronics, Circuits and Systems (ICECS-2004)*, pp. 559–562, Dec. 2004.

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

[9] Y. C. Eldar and O. Christansen, "Characterization of frame duals," to appear in *J. Applied Signal Processing*, 2005.