SAMPLING THEORY IN SIGNAL AND IMAGE PROCESSING Vol. 6, No. 2, May 2007, pp. 167-184

ISSN: 1530-6429

Sampling Theorem with Optimum Noise Suppression

Hidemitsu Ogawa
Department of Education
Tokyo University of Social Welfare
2020-1, San'o-cho, Isesaki City, Gunma 372-0831, Japan
hidemitsu-ogawa@kuramae.ne.jp

Akira Hirabayashi
Department of Computer Science and Systems Engineering,
Yamaguchi University,
2-16-1, Tokiwadai, Ube, Yamaguchi 755-8611, Japan
a-hira@yamaguchi-u.ac.jp

Abstract

We propose sampling theorems that reconstruct the optimal approximation under a certain criterion from a *finite* number of *degraded, noisy*, sampled values. In that criterion, we minimize the average difference between a reconstructed function and an individual target function over a noise ensemble subject to the condition that the reconstructed function is an unbiased estimator of the best approximation obtainable from noiseless sampled values. We devise a general form for sampling theorems with a real pulse and with an ideal pulse, thus providing the optimal estimator even for a singular noise covariance matrix. The relationship between the proposed criterion and the Gauss-Markov estimator is also discussed. Finally, we clarify the relationship between the best approximation and the interpolation.

Keywords and phrases : Sampling theorem with noisy sampled values, Sampling theorem with a finite number of sampled values, Best approximation, Optimal noise suppression

The IEEE Transactions on Signal Processing EDICS numbers — 2-SAMP

1 Introduction

The history of sampling theory began with this familiar theorem [1, 2, 3, 4]. If a signal, f, contains no frequencies higher than the frequency, $\Omega/2$, then the signal can be reconstructed from its sampled values via the formula

$$f(x) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{\Omega}\right) \frac{\sin \pi(\Omega x - n)}{\pi(\Omega x - n)}. \tag{1}$$

The theorem was extended in various directions, such as sampling theorems for signals of more than one variable, nonuniform sampling, bandpass signals, and general integral transforms [5, 6, 7, 8].

These theorems were originally proposed for *perfect reconstruction* under certain assumptions, such as band limitation, an infinite number of sampled data, exact sample points, and exact sampled values. For real-world problems, however, these assumptions are generally not satisfied, so the perfect reconstruction cannot be achieved.

To overcome these problems, the *approximation point of view* is useful [9]. In this context, a criterion to evaluate degrees of approximation plays an important role. Let $\hat{f}$ be a function reconstructed by a sampling theorem. A simple measure of the deviation of $\hat{f}$ from f is

$$\|\hat{f} - f\|^2 \tag{2}$$

when exact sampled values are available, or

$$E\|\hat{f} - f\|^2 \tag{3}$$

when sampled values are degraded by noise, where E denotes the expectation over noise ensemble. Since this measure involves the unknown target function, f, it cannot be evaluated. Thus, many alternatives have been proposed so far. They can be classified into two groups.

The first group of the alternatives minimizes, instead of Eq. (2) or (3), the sum of the error term between $\hat{f}$ and f *over sample points* and the term expressing a certain constraint on the target function. These approaches represent a kind of regularization method, and the second term is called the regularization term. Various regularization terms have been proposed [10, 11, 12, 13].

The other group of the alternatives minimizes Eq. (2) or (3) in the sense of the average with respect to target functions f [14, 15], or for the worst target function [14, 15], or for an individual target function even though f is unknown [16, 17, 18].

This article focuses on the minimization of Eq. (3) for an individual target function. We propose sampling theorems that minimize Eq. (3) for a given *finite* number of *degraded noisy* sampled values under the constraint that $\hat{f}$ is the unbiased estimator of the best approximation that can be obtained from noiseless sampled values.

This paper is organized as follows. In Section 2 we formulate the sampling problem[1] from the approximation point of view. In Section 3 a closed form of an

[1]In this paper, the term *sampling problem* refers to a problem of constructing sampling theorems, i.e., reconstruction formulas.

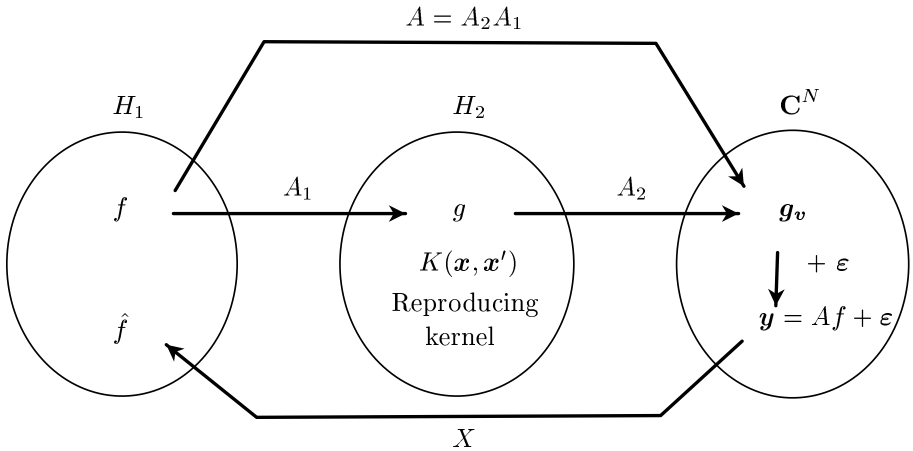

Figure 1: Framework to discuss the sampling problem.

optimal reconstruction operator is devised. Based on the closed form, we develop sampling theorems with optimum noise suppression in Section 4. Finally, the relationship between the best approximation and the interpolation is clarified in Section 5.

2 Formulation of the Sampling Problem

We formulate the sampling problem from the approximation point of view. Let f be an original signal to be reconstructed from sampled values associated with f. The signal f is defined on a subset $\mathcal{D}$ in a Euclidean space. Assume that f belongs to a Hilbert space denoted by H_1 (see Fig.1). Note that H_1 is not necessarily a reproducing kernel Hilbert space (RKHS); for example, it is not necessarily band-limited.

When we measure f through a certain device, f is converted into another signal. We denote the signal by g whose domain is also $\mathcal{D}$. Assume that g belongs to an RKHS, denoted by H_2, with the reproducing kernel $K(\boldsymbol{x}, \boldsymbol{x}')$. This guarantees that the value $g(\boldsymbol{x})$ at each point $\boldsymbol{x} \in \mathcal{D}$ is well defined.

Finally, the sampled values of g, denoted by $\{y_n\}_{n=1}^{N}$, at sample points $\{\boldsymbol{x}_n\}_{n=1}^{N}$ are given by

$$y_n = g(\boldsymbol{x}_n) + \varepsilon_n \quad (n = 1, 2, \ldots, N), \tag{4}$$

where ε_n is a random noise. We assume that ε_n has a zero mean and that its covariance matrix Q is given by

$$Q = \sigma^2 Q_1, \tag{5}$$

where Q_1 is a known positive semidefinite matrix and σ is an unknown positive real parameter. By using the Neumann-Schatten product [20], Q is expressed as

$$Q = E(\boldsymbol{\varepsilon} \otimes \overline{\boldsymbol{\varepsilon}}). \tag{6}$$

Let A_1 be the operator that maps f to g:

$$A_1 f = g. \tag{7}$$

Let $\boldsymbol{g_v}$ be the vector whose n-th element is $g(\boldsymbol{x}_n)$. Let A_2 be the operator that maps g to $\boldsymbol{g_v}$, i.e., $A_2 g = \boldsymbol{g_v}$. We denote the totality of A_1 and A_2 by A, i.e., $A = A_2 A_1$.

Let $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ be vectors in $\mathbf{C}^N$ whose n-th elements are y_n and ε_n, respectively. Then,

$$\boldsymbol{y} = Af + \boldsymbol{\varepsilon}. \tag{8}$$

Let X be a linear operator that maps $\boldsymbol{y}$ to an optimal approximation $\hat{f}$ to f under a certain criterion:

$$\hat{f} = X\boldsymbol{y}. \tag{9}$$

Let $\{\boldsymbol{e}_n\}_{n=1}^N$ be the standard basis in $\mathbf{C}^N$. That is, $\boldsymbol{e}_n$ is the N-dimensional vector consisting of zero elements except for the n-th element equal to 1. If we let

$$u_n = X\boldsymbol{e}_n, \tag{10}$$

then we have

$$\hat{f}(\boldsymbol{x}) = \sum_{n=1}^{N} y_n u_n(\boldsymbol{x}). \tag{11}$$

The functions $\{u_n\}_{n=1}^N$ are called reconstruction functions.

Equation (11) is called a sampling theorem with a real pulse. If $H_2 = H_1$ and $A_1 = I$, where I is the identity operator on H_1, then y_n reduces to $y_n = f(\boldsymbol{x}_n) + \varepsilon_n$. In this case Eq. (11) is called a sampling theorem with an ideal pulse [21].

Given a set of sample points, the sampling problem is to obtain an optimal set of reconstruction functions $\{u_n\}_{n=1}^N$ that provides the optimal approximation $\hat{f}$ from the sampled data $\{y_n\}_{n=1}^N$. This is equivalent to obtain X that provides the optimal approximation $\hat{f}$ from the vector $\boldsymbol{y}$.

3 Optimal Reconstruction Operator

In this paper, the following notations are used. Let T^* and $T^\dagger$ be the adjoint operator and the Moore-Penrose generalized inverse of linear operator T, respectively. The range and the null space of T are denoted by $\mathcal{R}(T)$ and $\mathcal{N}(T)$, respectively. Let $S^{\perp}$ and P_S be the orthogonal complement of closed subspace S and the orthogonal projection operator onto S, respectively.

In order to obtain the optimal approximation $\hat{f}$ to f, we introduce the following criterion.

Definition 1. *An operator X is called an 'optimal reconstruction operator' if X minimizes*

$$J[X] = E\|X\boldsymbol{\varepsilon}\|^2 \tag{12}$$

under the constraint of

$$XA = P_{\mathcal{R}(A^*)}. \tag{13}$$

The function $\hat{f}$ in Eq. (9) is called an 'optimal estimator' of f if X is the optimal reconstruction operator.

This criterion is introduced based on the following consideration. We first consider the noiseless case. Then, $\hat{f}$ is given by

$$\hat{f} = XAf.$$

Our goal in this case is to minimize

$$J_1[X] = \| \hat{f} - f \|^2 = \| XAf - f \|^2 .$$

Let S be a closed subspace of H_1 to which XAf belongs. Then, J_1 is minimized by $\hat{f} = P_S f$, where P_S is the orthogonal projection operator onto S. Hence, we require that

$$X\boldsymbol{y} = P_S f \quad \text{for each } f \in H_1. \tag{14}$$

That is, although we do not know the original function f, we want to obtain the best approximation $P_S f$ to each f from sampled values $\{g(\boldsymbol{x}_n)\}_{n=1}^{N}$.

Note that Eq. (14) does not request the interpolation property. The relationship between the best approximation and the interpolation is clarified in Section 5.

Equation (14) is equivalent to the following operator equation:

$$XA = P_S. \tag{15}$$

Here we use the following lemma.

Lemma 1. [19] *For any fixed operators T_1 and T_2 with closed ranges, the following statements are mutually equivalent.*

(i) The equation $XT_1 = T_2$ has a solution.

(ii) $\mathcal{N}(T_1) \subseteq \mathcal{N}(T_2)$.

(iii) $T_2 T_1^{\dagger} T_1 = T_2$.

When these conditions hold, a general solution of $XT_1 = T_2$ is given by

$$X = T_2 T_1^\dagger + Y(I - T_1 T_1^\dagger),$$

where Y is an arbitrary linear operator.

By Lemma 1 Eq. (15) has a solution if and only if $S \subseteq \mathcal{R}(A^*)$. Then we have the following.

Lemma 2. *For each $f \in H_1$ the best approximation to f in S can be obtained from sampled values of f if and only if*

$$S \subseteq \mathcal{R}(A^*). \tag{16}$$

The range of A^* is the largest subspace in which we can obtain the best approximation to f from $\boldsymbol{y}$.

The larger the subspace S is, the better the approximation $P_S f$ becomes. Hence, we concentrate our attention on this maximal subspace $\mathcal{R}(A^*)$ hereafter, i.e., $S = \mathcal{R}(A^*)$. Then, the best approximation of f becomes the orthogonal projection of f onto $\mathcal{R}(A^*)$, and Eq. (15) reduces to Eq. (13).

Now, we consider the noisy case with the measure in Eq. (3):

$$J_1 = E \parallel \hat{f} - f \parallel^2 . \tag{17}$$

Equations (8) and (9) yield

$$\hat{f} = XAf + X\boldsymbol{\varepsilon}. \tag{18}$$

In order to be consistent with the noiseless case, we preserve the condition (13) for the noisy case. Let us denote the mean of $\hat{f}$ by $\overline{\hat{f}}$. Since $\boldsymbol{\varepsilon}$ has a mean value 0, it follows from Eq. (18) that $\overline{\hat{f}} = P_{\mathcal{R}(A^*)} f$ if and only if $XA = P_{\mathcal{R}(A^*)}$. Thus, we have the following Lemma.

Lemma 3. *The estimator $\hat{f}$ in Eq. (18) is an unbiased estimator of $P_{\mathcal{R}(A^*)} f$ if and only if $XA = P_{\mathcal{R}(A^*)}$.*

The measure J_1 in Eq. (17) has the following bias-variance decomposition:

$$J_1 = \parallel \overline{\hat{f}} - f \parallel^2 + E \parallel \hat{f} - \overline{\hat{f}} \parallel^2 . \tag{19}$$

Since $\overline{\hat{f}} = P_{\mathcal{R}(A^*)} f$, Eq. (19) becomes

$$J_1 = \parallel P_{\mathcal{R}(A^*)} f - f \parallel^2 + E \parallel X\boldsymbol{\varepsilon} \parallel^2 . \tag{20}$$

The first term of this equation is independent of X. Hence, the minimization of J_1 with respect to X is equivalent to the minimization of Eq. (12) under the constraint of $XA = P_{\mathcal{R}(A^*)}$. In light of these considerations, we established Definition 1.

The optimal reconstruction operator in the sense of Definition 1 is given as follows.

Theorem 1. (Optimal reconstruction operator) *The optimal reconstruction operator always exists. Its general form is given by*

$$X = V^{\dagger}A^{*}U^{\dagger} + Y(I - UU^{\dagger}), \tag{21}$$

where U and V are operators defined by

$$U = AA^{*} + Q_1, \tag{22}$$
$$V = A^{*}U^{\dagger}A, \tag{23}$$

and Y is an arbitrary linear operator from $\mathbf{C}^N$ to H_1. The minimum value, say J_0, of $J[X]$ is given by

$$J_0 = \sigma^2(\mathrm{tr}(V^{\dagger}) - M_0), \tag{24}$$

where M_0 is the dimension of $\mathcal{R}(A^{*})$ and $\mathrm{tr}(V^{\dagger})$ is the trace of the operator $V^{\dagger}$ [20].

The operator U depends only on Q_1 in Eq. (5), and independent of the unknown parameter σ^2. Hence, so is X in Eq. (21).

Theorem 1 is proved along the following line. (i) Properties of operators U and V are summerized in Lemma 4. (ii) The constrained optimization problem is reduced to a pair of linear equations in Lemma 5. (iii) The pair of linear equations is reduced to a single equation in Lemma 6. (iv) By using these lemmas, Theorem 1 is proved.

The operators U and V have the following properties.

Lemma 4. *The operators U and V are positive semidefinite. Furthermore, it holds that*

$$UU^{\dagger}A = A, \tag{25}$$
$$VV^{\dagger} = V^{\dagger}V = P_{\mathcal{R}(A^{*})}, \tag{26}$$
$$\mathcal{R}(U) = \mathcal{R}(A)\dot{+}Q_1\mathcal{R}(A)^{\perp}. \tag{27}$$

The proof of this lemma is reserved for Section 6.1. Equation (27) is a (generally non-orthogonal) direct sum decomposition of $\mathcal{R}(U)$.

The constrained optimization problem is reduced to a pair of linear equations as follows.

Lemma 5. *An operator X is an optimal reconstruction operator if and only if X together with an operator C satisfies*

$$XA = P_{\mathcal{R}(A^{*})}, \tag{28}$$
$$XQ_1 = CA^{*}. \tag{29}$$

In this case the minimum value J_0 of $J[X]$ is given by

$$J_0 = \sigma^2\mathrm{tr}(CP_{\mathcal{R}(A^{*})}). \tag{30}$$

The rigorous proof of this lemma is reserved for Section 6.2. It is, however, worthy to show an intuitive derivation of Eqs. (28) and (29), which provides the meaning of the operator C in Lemma 5. By using the notion of the Schmidt inner product $\langle T_1, T_2\rangle$ of operators T_1 and T_2 [20], Eq. (12) can be expressed as

$$J[X] = \sigma^2 \mathrm{tr}\,(XQ_1X^*) = \sigma^2\langle XQ_1, X\rangle, \tag{31}$$

which is proved in Section 6.2. If the Lagrange multiplier operator is denoted by C, then the minimization of Eq. (12) subject to $XA = P_{\mathcal{R}(A^*)}$ is reduced to the unconditional problem of variation that minimizes

$$J[X, C] = \langle XQ_1, X\rangle - 2\Re\langle XA - P_{\mathcal{R}(A)}, C\rangle,$$

where $\Re$ stands for the real part of a complex number. Equating the partial derivative of $J[X, C]$ with respect to C and X to zero yields Eqs. (28) and (29), respectively. That is, the operator C in Lemma 5 is nothing but the Lagrange multiplier operator.

Lemma 5 reduces a variational problem for the optimal reconstruction operator to an algebraic problem, in which the operator is characterized by using a pair of linear equations. Furthermore, it is reduced to a single equation as follows.

Lemma 6. *An operator X is an optimal reconstruction operator if and only if X satisfies*

$$XU = V^\dagger A^*. \tag{32}$$

For the solution of Eq. (32) it holds that

$$XQ_1 = (V^\dagger - P_{\mathcal{R}(A^*)})A^*. \tag{33}$$

The proof of this lemma is reserved for Section 6.3. By using these lemmas, we shall prove Theorem 1.

Proof of Theorem 1

As shown in the proof of Lemma 6, Eq. (32) has a solution. Its general form is given by Eq. (21) because of Lemma 1. We shall show Eq. (24). It follows from Eqs. (29) and (33) that

$$CA^* = XQ_1 = (V^\dagger - P_{\mathcal{R}(A^*)})A^*. \tag{34}$$

Since $A^*(A^*)^\dagger = P_{\mathcal{R}(A^*)}$, Eqs. (34) and (26) yield

$$\begin{aligned} CP_{\mathcal{R}(A^*)} &= CA^*(A^*)^\dagger = (V^\dagger - P_{\mathcal{R}(A^*)})A^*(A^*)^\dagger \\ &= (V^\dagger - P_{\mathcal{R}(A^*)})P_{\mathcal{R}(A^*)} = V^\dagger P_{\mathcal{R}(A^*)} - P_{\mathcal{R}(A^*)} \\ &= V^\dagger - P_{\mathcal{R}(A^*)}, \end{aligned}$$

and, hence, Eq. (24) holds because of Eq. (30). □

The optimal estimator of f is directly obtained by $\hat{f} = X\boldsymbol{y}$ with X in Eq. (21):

$$\hat{f} = V^{\dagger}A^{*}U^{\dagger}\boldsymbol{y} + Y(I - UU^{\dagger})\boldsymbol{y}. \tag{35}$$

Equations (21) and (35) are expressions associated with the orthogonal direct sum decomposition $\mathbf{C}^{N} = \mathcal{R}(U) \oplus \mathcal{R}(U)^{\perp}$. Further, $\mathcal{R}(U)$ has the (generally non-orthogonal) direct sum decomposition as shown in Eq. (27). For the decomposition, we have the following.

Corollary 1. *An operator X is the optimal reconstruction operator if and only if*

$$Xu = \begin{cases} A^{\dagger}u & : u \in \mathcal{R}(A), \\ 0 & : u \in Q_1\mathcal{R}(A)^{\perp}. \end{cases} \tag{36}$$

Proof. By Lemma 5, it is enough to show that two equations in Eq. (36) are equivalent to Eqs. (28) and (29), respectively. The first equation in Eq. (36) is equivalent to $XA = A^{\dagger}A$, which is also equivalent to Eq. (28). On the other hand, the second equation in Eq. (36) is equivalent to $XQ_1\mathcal{R}(A)^{\perp} = \{0\}$, which is equivalent to $\mathcal{N}(A^{*}) \subseteq \mathcal{N}(XQ_1)$. This is also equivalent to the fact that for any fixed operator X, Eq. (29) has a solution C because of Lemma 1. Hence, the second equation in Eq. (36) is equivalent to Eq. (29). □

Corollary 1 shows that the optimal reconstruction operator works as $A^{\dagger}$ for elements in $\mathcal{R}(A)$. However, although $A^{\dagger}$ reduces to zero all elements in $\mathcal{R}(A)^{\perp}$, the optimal reconstruction operator reduces to zero all elements in $Q_1\mathcal{R}(A)^{\perp}$. This guarantees the optimal noise suppression ability of the reconstruction operator.

For special A and Q_1, the expression of the optimal reconstruction operator in Eq. (21) becomes much simpler. For example, if $\mathcal{N}(A) = \{0\}$ and Q_1 is nonsingular, then

$$X = (A^{*}Q_1^{-1}A)^{-1}A^{*}Q_1^{-1}, \tag{37}$$

which is already described in [16] in the context of Gauss-Markov estimator.

Eq. (37) can be extended to the positive semidefinite Q_1 as follows. If $\mathcal{R}(Q_1) \supseteq \mathcal{R}(A)$, then the optimal reconstruction operator is expressed as

$$X = (A^{*}Q_1^{\dagger}A)^{\dagger}A^{*}Q_1^{\dagger} + Y(I - Q_1Q_1^{\dagger}). \tag{38}$$

The proof of this equation is reserved for Section 6.4. This equation means that if $\mathcal{R}(Q_1) \supseteq \mathcal{R}(A)$, then we can replace U in Eq. (21) with Q_1.

Remark The optimal estimator in the sense of Definition 1 is a special case of the Gauss-Markov estimator [16]. It is also referred to as the best linear unbiased estimator (BLUE) [22]. The Gauss-Markov estimator includes several parameters. However, since the Gauss-Markov theory is a general theory of estimation, it cannot specify those parameters. They should be determined for specific applications. Definition 1 has specified those parameters from the point of view of the sampling theorem.

So far, the Gauss-Markov theory has given solutions for special cases such as Eq. (37) [16, 22]. Furthermore, they are not all the solutions, but rather special solutions. On the other hand, Theorem 1 has devised a general form of the optimal estimator that is valid even for the singular noise covariance matrix.

4 Sampling Theorem with Optimum Noise Suppression

If we apply the optimal reconstruction operator given in Theorem 1 to Eqs. (10) and (11), then we arrive at the following theorem.

Theorem 2 (Sampling theorem with real pulse). *Let* $\{u_n\}_{n=1}^{N}$ *be functions given by*

$$u_n = \hat{u}_n + \tilde{u}_n \quad (n = 1, 2, \ldots, N), \tag{39}$$

where

$$\hat{u}_n = V^{\dagger} A^{*} U^{\dagger} \boldsymbol{e}_n, \tag{40}$$

$$\tilde{u}_n = Y(I - UU^{\dagger})\boldsymbol{e}_n. \tag{41}$$

Then the optimal estimator $\hat{f}$ *is obtained by*

$$\hat{f}(\boldsymbol{x}) = \sum_{n=1}^{N} y_n u_n(\boldsymbol{x}). \tag{42}$$

Since Eq. (21) is a general form of the optimal reconstruction operator X, Eq. (42) is a general form of the sampling theorem with a real pulse that provides the optimal estimator in the sense of Definition 1. That is, by changing Y in Eq. (41) we can construct all sets of optimal reconstruction functions $\{u_n\}_{n=1}^{N}$.

Corollary 2. *The optimal estimator* $\hat{f}$ *in Eq. (42) agrees with* $\sum_{n=1}^{N} y_n \hat{u}_n$ *with probability* 1. *That is, it holds that*

$$E \parallel \hat{f} - \sum_{n=1}^{N} y_n \hat{u}_n \parallel^2 = 0.$$

Proof. Since $\mathcal{R}(U) = \mathcal{R}(A) + \mathcal{R}(Q_1)$, it holds that $\mathcal{R}(A) \subseteq \mathcal{R}(U)$ and $\mathcal{R}(Q_1) \subseteq \mathcal{R}(U)$. Thus, it follows from Eqs. (40), (35), (6), and (5) that

$$\begin{aligned} E \parallel \hat{f} - \sum_{n=1}^{N} y_n \hat{u}_n \parallel^2 &= E \parallel \hat{f} - V^\dagger A^* U^\dagger \boldsymbol{y} \parallel^2 \\ &= E \parallel Y P_{\mathcal{R}(U)^\perp} \boldsymbol{y} \parallel^2 \\ &= E \parallel Y P_{\mathcal{R}(U)^\perp} (Af + \boldsymbol{\varepsilon}) \parallel^2 \\ &= E \parallel Y P_{\mathcal{R}(U)^\perp} \boldsymbol{\varepsilon} \parallel^2 \\ &= E \mathrm{tr}(Y P_{\mathcal{R}(U)^\perp} \boldsymbol{\varepsilon} \otimes \overline{Y P_{\mathcal{R}(U)^\perp} \boldsymbol{\varepsilon}}) \\ &= \sigma^2 \mathrm{tr}(Y P_{\mathcal{R}(U)^\perp} Q_1 P_{\mathcal{R}(U)^\perp} Y^*) = 0. \end{aligned}$$

This establishes the corollary. □

Corollary 2 means that $\hat{f}$ is independent of $\{\tilde{u}_n\}_{n=1}^N$ in Eq. (41) with probability 1. Hence, from the scientific point of view, the set $\{\hat{u}_n\}_{n=1}^N$ plays an essential role in reconstruction. From an engineering point of view, however, the redundancy of $\{\tilde{u}_n\}_{n=1}^N$ provides the potential ability to design the reconstruction functions $\{u_n\}_{n=1}^N$. That will be discussed in a separate paper.

The matrix U in Theorem 2 is defined by using the operator A as shown in Eq. (22). It can be calculated by using only matrices as follows. Let $\{v_n^*\}_{n=1}^N$ and $\{u_n^*\}_{n=1}^N$ be functions defined by

$$v_n^*(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}_n) \quad \textit{and} \quad u_n^* = A_1^* v_n^*, \tag{43}$$

respectively. Let G be the Gram matrix of $\{u_n^*\}_{n=1}^N$. Then, it holds that

$$U = G + Q_1. \tag{44}$$

Indeed, it follows from Eqs. (43) and (7) that

$$\langle f, u_n^* \rangle = \langle g, v_n^* \rangle = g(\boldsymbol{x}_n). \tag{45}$$

Hence, by using $\{u_n^*\}_{n=1}^N$, the operator A is expressed as

$$A = \sum_{n=1}^{N} \boldsymbol{e}_n \otimes \overline{u_n^*}.$$

This implies $AA^* = G$, and Eq. (44) holds.

Because of Eq. (45), $\{v_n^*\}_{n=1}^N$ and $\{u_n^*\}_{n=1}^N$ are said to be the sampling functions with an ideal pulse and a real pulse, respectively. They are unique for any fixed set of sample points $\{\boldsymbol{x}_n\}_{n=1}^N$.

Consider the case that $H_2 = H_1$ and $A_1 = I$, where I is the identity operator on H_1. Then, Theorem 2 reduces to the sampling theorem with an ideal pulse.

In this case, the Gram matrix G of $\{u_n^*\}_{n=1}^N$ reduces to the Gram matrix of $\{v_n^*\}_{n=1}^N$. Since

$$\langle v_j^*, v_i^* \rangle = K(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{46}$$

the Gram matrix of $\{v_n^*\}_{n=1}^N$ can be obtained without calculating inner product of functions.

5 Best Approximation and Interpolation

Remember that $\boldsymbol{g_v}$ is the vector whose n-th element is $g(\boldsymbol{x}_n)$, i.e., $\boldsymbol{g_v} = Af$. If we observe the reconstructed function $\hat{f}$ by the operator A again, then

$$E(A\hat{f}) = \boldsymbol{g_v}, \tag{47}$$

because of $XA = P_{\mathcal{R}(A^*)}$. In the noiseless case, Eq. (47) reduces to

$$A\hat{f} = \boldsymbol{g_v}. \tag{48}$$

When $H_2 = H_1$ and $A_1 = I$, the vector with n-th element $f(\boldsymbol{x}_n)$ is well defined, and is denoted by $\boldsymbol{f_v}$, i.e., $\boldsymbol{f_v} = Af$. In this case, Eqs. (47) and (48) reduce to

$$E(A\hat{f}) = \boldsymbol{f_v} \quad \textit{and} \quad A\hat{f} = \boldsymbol{f_v}, \tag{49}$$

respectively. These equations imply that $\hat{f}$ is an interpolation function of $\boldsymbol{f_v}$. Hence, we call Eqs. (47) and (48) the *generalized interpolation property* or merely the *interpolation property.*

In Eq. (14), we required only the fact that the best approximation $P_S f$ to each f can be obtained from sampled values $\{g(\boldsymbol{x}_n)\}_{n=1}^N$, although we do not know the original function f. Where does the interpolation property come from? Is it a desirable property? In this section, we discuss the relationship between the obtainability of the best approximation and the interpolation property. Since the noise $\boldsymbol{\varepsilon}$ has zero mean, Eqs. (47) and (48) are essentially equivalent. Thus, we concentrate our attention on the noiseless case.

The prime concern in this paper is how to obtain the best approximation $P_S f$ to each f, as mentioned above. Thus, if $P_S f$ has the interpolation property, then Eq. (48) is a desired property. Otherwise it is not, because in that case Eq. (48) means that $\hat{f}$ is not the best approximation. In this context, the following lemma is important.

Lemma 7. *Let S be a closed subspace in H_1 . For each $f \in H_1$, the best approximation to f in S has the generalized interpolation property if and only if*

$$S \supseteq \mathcal{R}(A^*). \tag{50}$$

The range of A^* is the smallest subspace that has the generalized interpolation property.

Proof. The generalized interpolation property is expressed as $AP_S f = Af$ for any f in H_1 . This is equivalent to $AP_S = A$, which is also equivalent to $\mathcal{N}(P_S) \subseteq \mathcal{N}(A)$ because of Lemma 1. Taking the adjoint of this equation yields Eq. (50). Thus, the lemma is established. □

Note that Eq. (50) is the opposite result of Eq. (16). Lemma 2 and Lemma 7 have the following consequences:

(1) If $S \subset \mathcal{R}(A^*)$, the best approximation $P_S f$ to f can be obtained from the observed vector $\boldsymbol{g_v}$. However, it has no interpolation property.

(2) Conversely, if $S \supset \mathcal{R}(A^*)$, $P_S f$ has the interpolation property. However, it cannot be obtained from $\boldsymbol{g_v}$ anymore.

(3) Finally, if and only if $S = \mathcal{R}(A^*)$, $P_S f$ can be obtained from $\boldsymbol{g_v}$ and has the interpolation property at the same time.

6 Proofs of auxiliary results

6.1 Proof of Lemma 4

Positive semidefiniteness of U and V are clear from their definitions. Since

$$\mathcal{R}(U) = \mathcal{R}(A) + \mathcal{R}(Q_1), \tag{51}$$

it holds that $\mathcal{R}(U) \supseteq \mathcal{R}(A)$, which yields Eq. (25). Since V is self-adjoint and $\mathcal{R}(V) = \mathcal{R}(A^*)$, Eq. (26) is clear. Finally, we shall prove Eq. (27) . Since it follows from Proposition 4.3 in [16] that for a subspace S and a positive semidefinite operator T

$$S \cap TS^{\perp} = \{0\}, \tag{52}$$

we have

$$\mathcal{R}(A) \cap Q_1\mathcal{R}(A)^{\perp} = 0.$$

Hence, it is enough to show that $\mathcal{R}(A) + Q_1\mathcal{R}(A)^{\perp} = \mathcal{R}(U)$. It follows from Eq. (51) that

$$\mathcal{R}(A) + Q_1\mathcal{R}(A)^{\perp} \subseteq \mathcal{R}(U).$$

We shall show the converse. For any $u \in \mathcal{R}(A)^{\perp} \cap (Q_1\mathcal{R}(A)^{\perp})^{\perp}$, it follows from Eq. (52) that

$$Q_1 u \in (Q_1\mathcal{R}(A)^{\perp}) \cap Q_1(Q_1\mathcal{R}(A)^{\perp})^{\perp} = \{0\}.$$

Then, $Q_1 u = 0$ and

$$\mathcal{R}(A)^{\perp} \cap (Q_1\mathcal{R}(A)^{\perp})^{\perp} \subseteq \mathcal{N}(Q_1).$$

Taking the orthogonal complement of this equation yields

$$\mathcal{R}(A) + Q_1\mathcal{R}(A)^{\perp} \supseteq \mathcal{R}(Q_1).$$

Then, Eq. (51) yields

$$\mathcal{R}(A) + Q_1\mathcal{R}(A)^{\perp} \supseteq \mathcal{R}(A) + \mathcal{R}(Q_1) = \mathcal{R}(U).$$

□

6.2 Proof of Lemma 5

We first show that the system of Eqs. (28) and (29) has solutions. Let

$$X = V^{\dagger}A^{*}U^{\dagger}. \tag{53}$$

It follows from Eqs. (23) and (26) that

$$XA = V^{\dagger}A^{*}U^{\dagger}A = V^{\dagger}V = P_{\mathcal{R}(A^{*})}.$$

That is, X in Eq. (53) satisfies Eq. (28). Since Lemma 4 yields $A^{*}U^{\dagger}U = A^{*}$, it follows from Eqs. (22), (53), and (28) that

$$XQ_1 = XU - XAA^{*} = V^{\dagger}A^{*}U^{\dagger}U - P_{\mathcal{R}(A^{*})}A^{*} = V^{\dagger}A^{*} - P_{\mathcal{R}(A^{*})}A^{*},$$

and, hence, $XQ_1 = (V^{\dagger} - P_{\mathcal{R}(A^{*})})A^{*}$. This implies that the system of Eqs. (28) and (29) has solutions with $C = V^{\dagger} - P_{\mathcal{R}(A^{*})}$.

Let X_0 be a solution of Eqs. (28) and (29). We shall show that for any X which satisfies Eq. (28) it holds that $J[X] \geq J[X_0]$. Equations (12), (6), and (5) yield

$$\begin{aligned} J[X] &= E\|X\boldsymbol{\varepsilon}\|^2 = E\mathrm{tr}\left((X\boldsymbol{\varepsilon}) \otimes \overline{(X\boldsymbol{\varepsilon})}\right) = \mathrm{tr}\left(XE(\boldsymbol{\varepsilon} \otimes \overline{\boldsymbol{\varepsilon}})X^{*}\right) \\ &= \mathrm{tr}\left(XQX^{*}\right) = \sigma^2\mathrm{tr}\left(XQ_1X^{*}\right). \end{aligned} \tag{54}$$

From the definition of X_0 it holds that

$$X_0A = P_{\mathcal{R}(A^{*})} \quad \text{and} \quad X_0Q_1 = CA^{*}. \tag{55}$$

Then, Eq. (28) yields

$$\begin{aligned} XQ_1X_0^{*} &= X(X_0Q_1)^{*} = X(CA^{*})^{*} = (XA)C^{*} = P_{\mathcal{R}(A^{*})}C^{*} \\ &= (X_0A)C^{*} = X_0(CA^{*})^{*} = X_0(X_0Q_1)^{*} = X_0Q_1X_0^{*} \end{aligned}$$

and, hence, $XQ_1X_0^{*} = X_0Q_1X_0^{*}$. Since $X_0Q_1X_0^{*}$ is self-adjoint, $XQ_1X_0^{*}$ is also self-adjoint and it holds that

$$X_0Q_1X_0^{*} = XQ_1X_0^{*} = X_0Q_1X^{*}.$$

Thus, Eq. (54) yields

$$J[X] - J[X_0] = \sigma^2 \mathrm{tr}[(X - X_0)Q_1(X - X_0)^*] \geq 0. \tag{56}$$

Then, $J[X] \geq J[X_0]$ because $Q_1 \geq 0$. That is, X_0 is an optimal reconstruction operator.

Conversely, assume that $J[X] = J[X_0]$ for X in Eq. (28). Equation (56) implies $(X - X_0)Q_1 = 0$. Hence, $XQ_1 = X_0Q_1 = CA^*$. That is, X satisfies Eq. (29). This means that all optimal reconstruction operators are given by solutions of Eqs. (28) and (29).

Finally we show Eq. (30). It follows from Eq. (55) that

$$X_0Q_1X_0^* = CA^*X_0^* = C(X_0A)^* = CP_{\mathcal{R}(A^*)}$$

and, hence, $X_0Q_1X_0^* = CP_{\mathcal{R}(A^*)}$. This implies Eq. (30) because of Eq. (54) □

6.3 Proof of Lemma 6

From Lemma 5 it is enough to show that the system of Eqs. (28) and (29) is equivalent to Eq. (32). Let X be a solution of Eqs. (28) and (29) with an operator C. It follows from Eqs. (22), (28), and (29) that

$$XU = XAA^* + XQ_1 = P_{\mathcal{R}(A^*)}A^* + CA^*$$

and, hence,

$$XU = (P_{\mathcal{R}(A^*)} + C)A^*. \tag{57}$$

It follows from Eqs. (57), (26), (23), (25), and (28) that

$$\begin{aligned} XU &= (P_{\mathcal{R}(A^*)} + C)VV^\dagger A^* = (P_{\mathcal{R}(A^*)} + C)(A^*U^\dagger A)V^\dagger A^* \\ &= ((P_{\mathcal{R}(A^*)} + C)A^*)U^\dagger AV^\dagger A^* = XUU^\dagger AV^\dagger A^* \\ &= XAV^\dagger A^* = P_{\mathcal{R}(A^*)}V^\dagger A^* = V^\dagger A^*, \end{aligned}$$

which implies Eq. (32). This proof also means that Eq. (32) has a solution because the system of Eqs. (28) and (29) has a solution.

Conversely, let X be a solution of Eq. (32). It follows from Eqs. (25), (32), (23), and (26) that

$$XA = XUU^\dagger A = V^\dagger A^*U^\dagger A = V^\dagger V = P_{\mathcal{R}(A^*)},$$

which implies Eq. (28). It follows from Eqs. (22), (32), and (28) that

$$XQ_1 = X(U - AA^*) = XU - XAA^* = V^\dagger A^* - P_{\mathcal{R}(A^*)}A^*$$

and, hence, Eq. (33) holds. Therefore, if we let $C = V^\dagger - P_{\mathcal{R}(A^*)}$, then $XQ_1 = CA^*$. This implies Eq. (29). □

6.4 Proof of Eq. (38)

In order to prove Eq. (38), the following lemma is used.

Lemma 8. [23] *Let T_1 be an operator from a Hilbert space H_2 to a Hilbert space H_1. Let T_2 be a positive semidefinite operator on H_1. If and only if $\mathcal{R}(T_1) \subseteq \mathcal{R}(T_2)$, it holds*

$$(T_1T_1^* + T_2)^\dagger = T_2^\dagger - T_2^\dagger T_1(I_2 + T_1^*T_2^\dagger T_1)^{-1}T_1^*T_2^\dagger,$$

where I_2 is the identity operator on H_2.

Since $\mathcal{R}(Q_1) \supseteq \mathcal{R}(A)$, Eq. (51) yields $\mathcal{R}(U) = \mathcal{R}(Q_1)$. Then,

$$UU^\dagger = P_{\mathcal{R}(U)} = P_{\mathcal{R}(Q_1)} = Q_1Q_1^\dagger.$$

This implies that the second terms on the right-hand sides of Eqs. (21) and (38) agree with each other.

In order to prove that the first terms on the right-hand sides of these equations agree with each other, let us temporarily define an operator T by

$$T = A^*Q_1^\dagger A.$$

Since $Q_1 \geq 0$, when $\mathcal{R}(A) \subseteq \mathcal{R}(Q_1)$, it follows from Lemma 8 that

$$(AA^* + Q_1)^\dagger = Q_1^\dagger - Q_1^\dagger A(I + A^*Q_1^\dagger A)^{-1}A^*Q_1^\dagger. \tag{58}$$

It follows from Eqs. (22) and (58) that

$$\begin{aligned} A^*U^\dagger &= A^*(AA^* + Q_1)^\dagger \\ &= A^*(Q_1^\dagger - Q_1^\dagger A(I+T)^{-1}A^*Q_1^\dagger) \\ &= A^*Q_1^\dagger - (A^*Q_1^\dagger A)(I+T)^{-1}A^*Q_1^\dagger \\ &= (I - T(I+T)^{-1})A^*Q_1^\dagger \\ &= (I+T)^{-1}A^*Q_1^\dagger \end{aligned}$$

and, hence,

$$A^*U^\dagger = (I+T)^{-1}A^*Q_1^\dagger. \tag{59}$$

Then, Eq. (23) yields

$$V = A^*U^\dagger A = (I+T)^{-1}A^*Q_1^\dagger A = (I+T)^{-1}T$$

and, hence, $V = (I+T)^{-1}T$. Since $T^* = T$, it holds that $V^\dagger = T^\dagger(I+T)$. Hence, Eq. (59) yields

$$V^\dagger A^*U^\dagger = T^\dagger(I+T)(I+T)^{-1}A^*Q_1^\dagger = T^\dagger A^*Q_1^\dagger,$$

which implies that the first terms on the right-hand sides of Eqs. (21) and (38) agree with each other. □

ACKNOWLEDGEMENT

The authors are very thankful to the anonymous reviewers for their valuable comments.

References

[1] E. T. Whittaker, On the functions which are represented by the expansions of interpolation-theory, *Proc. Royal Society of Edinburgh*, **35**, 181–194, 1915.

[2] K. Ogura, On a certain transcendental integral function in the theory of interpolation, *Tohoku Mathematical Journal*, **17**, 64–72, 1920.

[3] I. Someya, *Waveform Transmission*, Shukyo Ltd., Tokyo, 1949.

[4] C. E. Shannon, Communications in the presence of noise, *Proc. IRE*, **37**, 10-21, 1949.

[5] A. J. Jerri, The Shannon sampling theorem – Its various extensions and applications: A tutorial review, *Proc. IEEE*, **65**(11), 1565–1596, 1977.

[6] R. Marks II, *Introduction to Shannon Sampling and Interpolation Theory*, Springer-Verlag, New York, 1991.

[7] A. Zayed, *Advances in Shannon's Sampling Theory*, CRC Press, New York, 1993.

[8] J. R. Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundations*, Clarendon Press, Oxford, 1996.

[9] M. Unser, Sampling–50 years after Shannon, *Proc. IEEE*, **88**(4), 569–587, 2000.

[10] A. N. Tikhonov and V. Y. Asenin (Translated by F. John), *Solutions of Ill-Posed Problems*, V. H. Winston and Sons, 1977.

[11] M. Z. Nashed and G. Wahba, Convergence rates of approximate least squares solutions of linear integral and operators equations of the first kind, *Math. Comp*, **28**, 69-80, 1974.

[12] G. Wahba, *Spline Models for Observational Data*, BMS-NSF Regional Conference Series in Applied Mathematics 59, Philadelphia, 1990.

[13] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers Group, Dordrecht, 1996.

[14] J. F. Traub, G. W. Wasilkowski and H. Wozniakowski, *Information-based Complexity*, Academic Press, New York, 1988.

[15] L. Plaskota, *Noisy Information and Computational Complexity*, Cambridge Univ. Press, Cambridge, 1996.

[16] M. L. Eaton, *Multivariate Statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York 1983.

[17] H. Ogawa, A unified approach to generalized sampling theorems, *Proc. ICASSP'86, IEEE-IECEJ-ASJ Int. Conf. on Acoustics, Speech, and Signal Processing*, **3**, 1657–1660, Tokyo, 1986.

[18] H. Ogawa, A generalized sampling theorem, *Transactions of IEICE*[2] *A*, **J71-A**(2), 163–170, 1988 (in Japanese). Its English translation appeared in *Electronics and Communication in Japan, Part 3*, Scripta Technica, USA, **72**(3), 97–105, 1989.

[19] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York and London, 1972.

[20] R. Schatten, *Norm Ideals of Completely Continuous Operators*, Springer-Verlag, Berlin, 1960.

[21] A. Papoulis, Error analysis in sampling theory, *Proc. IEEE*, **54**(7), 947–955, 1966.

[22] M. A. Golberg and H. A. Cho, *Introduction to Regression Analysis*, WIT Press, Southampton, 2004.

[23] H. Ogawa, An operator pseudo-inversion lemma, *SIAM J. Appl. Math.*, **48**(6), 1527-1531, 1988.

[2] *IEICE = The Institute of Electronics, Information and Communication Engineers, Japan*