Available online at www.sciencedirect.com
ScienceDirect
Appl. Comput. Harmon. Anal. 23 (2007) 273–283
Applied and Computational Harmonic Analysis
www.elsevier.com/locate/acha

Letter to the Editor

Optimal shift invariant spaces and their Parseval frame generators ☆

Akram Aldroubi [a,*], Carlos Cabrelli [b,c], Douglas Hardin [a], Ursula Molter [b,c]

[a] *Department of Mathematics, Vanderbilt University, 1326 Stevenson Center, Nashville, TN 37240, USA*
[b] *Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón I, 1428 Capital Federal, Argentina*
[c] *CONICET, Argentina*

Available online 23 May 2007

Communicated by Yves F. Meyer on 28 August 2006

Abstract

Given a set of functions $\mathcal{F} = \{f_1, \dots, f_m\} \subset L^2(\mathbb{R}^d)$, we study the problem of finding the shift-invariant space V with n generators $\{\varphi_1, \dots, \varphi_n\}$ that is "closest" to the functions of $\mathcal{F}$ in the sense that

$$V = \underset{V' \in \mathcal{V}_n}{\arg\min} \sum_{i=1}^{m} w_i \| f_i - P_{V'} f_i \|^2,$$

where w_is are positive weights, and $\mathcal{V}_n$ is the set of all shift-invariant spaces that can be generated by n or less generators. The Eckart–Young theorem uses the singular value decomposition to provide a solution to a related problem in finite dimension. We transform the problem under study into an uncountable set of finite dimensional problems each of which can be solved using an extension of the Eckart–Young theorem. We prove that the finite dimensional solutions can be patched together and transformed to obtain the optimal shift-invariant space solution to the original problem, and we produce a Parseval frame for the optimal space. A typical application is the problem of finding a shift-invariant space model that describes a given class of signals or images (e.g., the class of chest X-rays), from the observation of a set of m signals or images $f_1, \dots, f_m$, which may be theoretical samples, or experimental data.

1. Introduction

In many signal and image processing applications, images and signals are assumed to belong to some shift-invariant space of the form:

$$\mathcal{S}(\Phi) := \text{closure}_{L_2} \text{span}\{\varphi_i(x-k)\colon\ i = 1, \dots, n,\ k \in \mathbb{Z}^d\}, \tag{1.1}$$

☆ The research of Akram Aldroubi is supported in part by NSF Grant DMS-0504788. The research of Douglas Hardin is supported in part by NSF Grants DMS-0505756 and DMS-0532154. The research of Carlos Cabrelli and Ursula Molter is partially supported by Grants: PICT 15033, CONICET, PIP 5650, UBACyT X058 and X108.

* Corresponding author.

E-mail addresses: akram.aldroubi@vanderbilt.edu (A. Aldroubi), cabrelli@dm.uba.ar (C. Cabrelli), doug.hardin@vanderbilt.edu (D. Hardin), umolter@dm.uba.ar (U. Molter).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2007.05.001

where $\Phi = \{\varphi_1, \dots, \varphi_n\}$ is a set of functions in $L^2(\mathbb{R}^d)$. The functions $\varphi_1, \varphi_2, \dots, \varphi_n$ are called a *set of generators* for the space $\mathcal{S} = \mathcal{S}(\Phi) = \mathcal{S}(\varphi_1, \dots, \varphi_n)$ and any such space $\mathcal{S}$ is called a *finitely generated shift-invariant space* (FSIS) (see, e.g., [1,6]). For example, if $n = 1$, $d = 1$ and $\phi(x) = \operatorname{sinc}(x)$, then the underlying space is the space of band-limited functions (often used in communications).

Finitely generated shift-invariant spaces, can have different sets of generators. The *length* of an FSIS $\mathcal{S}$ is

$$\mathcal{L}(\mathcal{S}) = \min\bigl\{\ell \in \mathbb{N}\colon \exists \varphi_1, \dots, \varphi_\ell \in \mathcal{S} \text{ with } \mathcal{S} = \mathcal{S}(\varphi_1, \dots, \varphi_\ell)\bigr\}.$$

We will denote by $\mathcal{V}_n$ the set of all shift-invariant spaces with length less than or equal to n. That is, an element in $\mathcal{V}_n$ is a shift-invariant space that has a set of s generators with $s \leqslant n$.

In most applications, the shift-invariant space chosen to describe the underlying class of signals is not derived from experimental data—for example many signal processing applications assume "band-limitedness" of the signal, which has theoretical advantages, but generally does not necessarily reflect the underlying class of signals accurately. Furthermore, in applications, the a priori hypothesis that the class of signals belongs to a shift-invariant space with a known number of generators, may not be satisfied. For example, the class of functions from which the data is drawn may not be a shift-invariant space. Another example is when the shift-invariant space hypothesis is correct but the assumptions about the number of generators is wrong. A third example is when the a priori hypothesis is correct but the data is corrupted by noise. In addition, for computational considerations, a shift-invariant space of length m could be modeled by a shift-invariant model space with length n much smaller than m. For example, in learning theory, the problem of reducing the number of generators for a subspace of a reproducing kernel Hilbert space is also important for improving the efficiency and sparsity of learning algorithms (see [16]). In order to model classes of signals or images by FSIS in realistic cases, or to model a very large data set by a computationally manageable shift-invariant space, we consider the following problem:

Problem 1. *Given a large set of experimental data $\mathcal{F} = \{f_1, f_2, \dots, f_m\} \subset L^2(\mathbb{R}^d)$, we wish to determine a shift-invariant space $V \in \mathcal{V}_n$ (where typically n is chosen to be small compared to m) that models the signals in "some" best way. For this purpose, we consider the following least squares problem*:

$$V = \underset{V' \in \mathcal{V}_n}{\arg\min} \sum_{i=1}^{m} w_i \|f_i - P_{V'} f_i\|^2 \tag{1.2}$$

where w_i are positive weights and where $P_{V'}$ is the orthogonal projection on V'.

A space V satisfying (1.2) will be said to *solve Problem* 1 *for* $(\mathcal{F}, w, n)$.

The weights w_i can be chosen to normalize or to reflect our confidence about the data. For example we can choose $w_i = \|f_i\|^{-2}$ to place the data on a sphere or we can choose a small weight w_i for a given f_i if—due to noise or other factors—our confidence about the accuracy of f_i is low. The goal is to see if we can perform operations on the observed data $\mathcal{F} = \{f_1, f_2, \dots, f_m\}$ to construct (if it exists) a shift-invariant space $\mathcal{S}(\Phi)$ whose length does not exceed a small number n, that minimizes the error with our data $\mathcal{F}$.

Problem 1 can be viewed as non-linear infinite dimensional constrained minimization problem. It may also be viewed in light of the recent learning theory developed in [2,5,15], and estimates of model fit in terms of noise and approximation space may be derived. Beside the fundamental question of existence of an optimal space, it will be important for applications to have a way to construct the generators of the optimal space if it exists, and to estimate the error $\mathcal{E}(\mathcal{F}, w, n) = \sum_{i=1}^{m} w_i \|f_i - P_V f_i\|^2$, where $V \in \mathcal{V}_n$ is an optimal space for $\mathcal{F}$, w and n.

Typical applications involve large data sets (for example consider the problem of finding a shift-invariant space model for the collection of chest X-rays using data collected by a hospital during the last 10 years). The space $\mathcal{S}(\mathcal{F})$ generated by a set of experimental data contains all the data as possible signals, but it is too large to be an appropriate model for use in applications. A space with a "small" number of generators is more suitable, since if the space is chosen correctly, it would reduce noise, and would give a computationally manageable model for a given application.

Least squares problems of the form above in *finite dimensional* spaces can be solved using the singular value decomposition (SVD). Shift-invariant spaces are infinite dimensional and the SVD cannot be applied directly. However, due to the special structure of shift-invariant spaces, the Fourier transform converts Problem 1 into finite dimensional least square problems at each frequency as will be discussed in Section 4.1.

2. Main theorems

In this paper we will sometimes deal with the standard Hilbert space $\mathbb{C}^N$. Elements of this vector space are column vectors with N coordinates. We will use the notation A^t and A^* to denote the transpose and the conjugate transpose respectively of a complex matrix A. We will say that a vector $y \in \mathbb{C}^N$ is a left eigenvector of the matrix A associated to the eigenvalue λ, if $y^t A = \lambda y^t$.

For clarity in the exposition, we will consider the unweighted case ($w_i = 1$, $i = 1, \dots, m$). The general case can be derived by simply applying the results of the unweighted case to the set of normalized observations $\mathcal{F} = \{f_1/w_1^2, \dots, f_m/w_m^2\}$.

The first theorem establishes the existence of an optimal space V. It also establishes that V can always be chosen to be a subspace of the shift-invariant space $\mathcal{S}(\mathcal{F})$ generated by the totality of the data. This optimal space V may not be unique. However, under additional assumptions that are often satisfied in practice, there is only one optimal space V, as stated in Theorem 2.4.

Theorem 2.1. *Let $\mathcal{F} = \{f_1, \dots, f_m\}$ be a set of functions in $L^2(\mathbb{R}^d)$. Then*

(1) *There exists $V \in \mathcal{V}_n$ such that*

$$\sum_{i=1}^{m} \|f_i - P_V f_i\|^2 \leqslant \sum_{i=1}^{m} \|f_i - P_{V'} f_i\|^2, \quad \forall V' \in \mathcal{V}_n. \tag{2.1}$$

(2) *The optimal shift-invariant space V in* (2.1) *can be chosen such that $V \subset \mathcal{S}(\mathcal{F})$.*

Remarks. (i) Although we do not make the assumption that $n \leqslant m$, if $n > m$, then $\mathcal{S}(\mathcal{F})$ is an optimal space that belongs to $\mathcal{V}_n$. Thus, we will always assume that $n \leqslant m$ for the remainder of this paper.

(ii) In practice it will often be the case that n is chosen (or found) to be much smaller than m.

We still need to explicitly find an optimal space V and estimate the error

$$\mathcal{E}(\mathcal{F}, n) = \min_{V' \in \mathcal{V}_n} \sum_{i=1}^{m} \|f_i - P_{V'} f_i\|^2. \tag{2.2}$$

To compute the error $\mathcal{E}(\mathcal{F}, n)$ we need to consider the Gramian matrix $G_{\mathcal{F}}$ of $\mathcal{F} = \{f_1, \dots, f_m\}$. Specifically, the *Gramian* G_Φ of a set of functions $\Phi = \{\varphi_1, \dots, \varphi_n\}$ with elements in $L^2(\mathbb{R}^d)$ is defined to be the $n \times n$ matrix of $\mathbb{Z}^d$-periodic functions

$$\big[G_\Phi(\omega)\big]_{i,j} = \sum_{k \in \mathbb{Z}^d} \hat{\varphi}_i(\omega + k)\overline{\hat{\varphi}_j(\omega + k)}, \quad \omega \in \mathbb{R}^d, \tag{2.3}$$

where $\hat{\varphi}_i$ denotes the Fourier transform of φ_i, and where $\bar{\hat{\varphi}}_i$ denotes the complex conjugate of $\hat{\varphi}_i$. It is known that G_Φ is $\mathbb{Z}^d$-periodic non-negative and self-adjoint for almost every ω. In this paper, we use the following definition for the Fourier transform of a function $\phi \in L^2(\mathbb{R}^d)$:

$$\hat{\phi}(\omega) := \int_{\mathbb{R}^d} \phi(x) e^{-i2\pi x^t \omega}\, \mathrm{d}x, \quad \omega \in \mathbb{R}^d, \tag{2.4}$$

where $\mathrm{d}x$ denotes Lebesgue measure on $\mathbb{R}^d$.

If $V = \{v_1, \dots, v_n\}$ is a set of vectors in a Hilbert space $\mathcal{H}$, we will denote by $\mathfrak{G}(V) = \mathfrak{G}(v_1, \dots, v_n)$ the matrix

$$\big[\mathfrak{G}(V)\big]_{i,j} = \langle v_i, v_j\rangle_{\mathcal{H}}, \quad i, j = 1, \dots, n. \tag{2.5}$$

Our next theorem produces a generator for an optimal space V and provides a formula for the exact value of the error, but we first recall the definition and some properties of frames used in its statement (see for example [4,9,11]).

Definition 2.2. Let $\mathcal{H}$ be a Hilbert space and $\{u_i\}_{i\in I}$ a countable subset of $\mathcal{H}$. The set $\{u_i\}_{i\in I}$ is said to form a *frame* for $\mathcal{H}$ if there exist $q, Q > 0$ such that

$$q\|f\|^2 \leqslant \sum_{i\in I} \left|\langle f, u_i\rangle\right|^2 \leqslant Q\|f\|^2, \quad \forall f \in \mathcal{H}.$$

If $q = Q$, then $\{u_i\}_{i\in I}$ is called a *tight frame*, and it is called a *Parseval frame* if $q = Q = 1$.

If $\{u_i\}_{i\in I}$ is a Parseval frame for a subspace W of a Hilbert space $\mathcal{H}$, and if $a \in \mathcal{H}$, then the orthogonal projection of a onto W is given by

$$P_W(a) = \sum_{i\in I} \langle a, u_i\rangle u_i. \tag{2.6}$$

Thus, a Parseval frames acts as if it were an orthonormal basis of W, even though it may not be one.

Theorem 2.3. *Under the same assumptions as in Theorem* 2.1, *let* $\lambda_1(\omega) \geqslant \lambda_2(\omega) \geqslant \cdots \geqslant \lambda_m(\omega)$ *be the eigenvalues of the Gramian* $G_{\mathcal{F}}(\omega)$. *Then*

(1) *The eigenvalues* $\lambda_i(\omega)$, $1 \leqslant i \leqslant m$ *are* $\mathbb{Z}^d$*-periodic, measurable functions in* $L^2([0,1]^d)$ *and*

$$\mathcal{E}(\mathcal{F}, n) = \sum_{i=n+1}^{m} \int_{[0,1]^d} \lambda_i(\omega)\,\mathrm{d}\omega. \tag{2.7}$$

(2) *Let* $E_i := \{\omega\colon \lambda_i(\omega) \neq 0\}$, *and define* $\tilde{\sigma}_i(\omega) = \lambda_i^{-1/2}(\omega)$ *on* E_i *and* $\tilde{\sigma}_i(\omega) = 0$ *on* E_i^c. *Then, there exists a choice of measurable left eigenvectors* $y_1(\omega), \ldots, y_n(\omega)$ *with* $y_i = (y_{i1}, \ldots, y_{im})^t$, $i = 1, \ldots, n$, *associated with the first* n *largest eigenvalues of* $G_{\mathcal{F}}(\omega)$ *such that the functions defined by*

$$\hat{\varphi}_i(\omega) = \tilde{\sigma}_i(\omega) \sum_{j=1}^{m} y_{ij}(\omega) \hat{f}_j(\omega), \quad i = 1, \ldots, n,\ \omega \in \mathbb{R}^d \tag{2.8}$$

are in $L^2(\mathbb{R}^d)$. *Furthermore, the corresponding set of functions* $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ *is a generator for an optimal space* V *and the set* $\{\varphi_i(\cdot - k),\ k \in \mathbb{Z}^d,\ i = 1, \ldots, n\}$ *is a Parseval frame for* V.

The following example shows that the optimal space V does not need to be unique. Let $m = 2$, $n = 1$, and let f_1, f_2 be two orthonormal functions. For this situation, $G_{\mathcal{F}}(\omega)$ is the 2×2 identity matrix for almost all $\omega \in \mathbb{R}^d$. It follows that any function $\varphi = c_1 f_1 + c_2 f_2$ with $c = (c_1, c_2)$ a unit vector in $\mathbb{R}^2$ generates an optimal space and $\mathcal{E}(\mathcal{F}, 1) = 1$. Obviously, in this particular case there are infinitely many optimal spaces. However, under some mild assumptions, there exists a unique optimal space V as described in the following theorem:

Theorem 2.4. *Let* $\mathcal{F} = \{f_1, \ldots, f_m\}$ *functions in* $L^2(\mathbb{R}^d)$ *be given. If* $\lambda_n(\omega) > \lambda_{n+1}(\omega)$ *for almost all* ω, *then the optimal space* V *in* (2.1) *is unique. In this case,* $n \leqslant r_{\min} = \min_{\omega\in[0,1]^d} \operatorname{rank} G_{\mathcal{F}}(\omega)$ *and the set* $\{\varphi(\cdot - k),\ k \in \mathbb{Z}^d,\ i = 1, \ldots, n\}$ *in part* (2) *of Theorem* 2.3 *is an orthonormal basis for* V.

Remarks.

(i) In case that $n = \mathcal{L}(\mathcal{S}(f_1, \ldots, f_m))$, Theorem 2.3 gives a proof of the known result that every FSIS has a set of generators forming a Parseval frame.

(ii) It will be clear from the proofs of Theorems 2.1 and 2.3 that the optimal space V can be decomposed as $V = S(\varphi_1) \bigoplus \cdots \bigoplus S(\varphi_\ell)$ where $\ell = \mathcal{L}(V)$, the direct sum is orthogonal and each φ_i is a Parseval frame generator of $S(\varphi_i)$.

(iii) Theorem 2.4 can only be used when $n < m$. When $n = m$ then $\mathcal{S}(\mathcal{F})$ is an optimal space and it is the unique optimal space if and only if $\mathcal{L}(\mathcal{S}(\mathcal{F})) = m$.

(iv) Obviously, if $n = m$ then the error between the model and the observation is null. However, by plotting the error in (2.7) in terms of the number of generators, an optimal number n may be heuristically derived. Alternatively, one may choose n so that a cost functional (depending on the error and on n) is optimized as in other dimension reduction schemes.

2.1. From theory to practice

In this section we briefly discuss some of the implementation issues that may arise in applications.

In order to find the generators of an optimal space described by Theorem 2.3, we need to find the eigenvalues of $G_{\mathcal{F}}$ and corresponding left eigenvectors for all $\omega \in [0,1]^d$ which, in practice, is not always possible. However, under some restriction on the data set $\mathcal{F}$, often assumed in practice, the optimal space can be approximated as closely as we wish.

For example, if $\mathcal{F}$ is a subset of the Wiener amalgam space W^1 of bounded functions with sufficient decay (specifically, $f \in W^1$ if $\|f\| = \sum_k \operatorname{ess\,sup}\{|f(x+k)|\colon x \in [0,1]^d\} < \infty$), then it can be shown that $G_{\mathcal{F}}$ is a continuous matrix function of ω (see [1]). For this case, we can approximate $G_{\mathcal{F}}$ by the piecewise constant matrix function

$$G^{\ell}_{\mathcal{F}}(\omega) = \sum_k G_{\mathcal{F}}(\omega_k)\chi_{I_k}(\omega),$$

where $\omega_k = \frac{1}{\ell}k$ for $k \in \mathbb{Z}^d$ and $I_k = \omega_k + [0, 1/\ell]^d$. Continuity of $G_{\mathcal{F}}$ implies that $G^{\ell}_{\mathcal{F}}$ converges to $G_{\mathcal{F}}$ uniformly as $\ell \to \infty$. We can compute the eigenvalues $\lambda_i^{\ell} = \sum_k \lambda_i(\omega_k)\chi_{I_k}$ and eigenvectors $y_i^{\ell} = \sum_k y_i(\omega_k)\chi_{I_k}$. (Note that because $G_{\mathcal{F}}$ is a $\mathbb{Z}^d$-periodic function, only a finite number of eigenvalues and eigenvectors need to be calculated in the approximation.)

For the case of interest when the number of generators n is less than the number of samples m, it is generically the case that the n eigenvalues λ_i are distinct and uniformly positive. Thus, for the generic case, the eigenvalues are continuous functions of ω and for each ω the eigenspaces are one dimensional and the eigenvectors can be chosen to be continuous in ω [13, p. 110]. Therefore, λ_i^{ℓ} and y_i^{ℓ} can be used in formula (2.8) of Theorem 2.3 to obtain a basis Φ^{ℓ} that approaches Φ (in the L^2 sense, using that y^{ℓ} converges uniformly to y) as ℓ approaches infinity.

If the functions f_i have compact support, then the entries of $G_{\mathcal{F}}$ are trigonometric polynomials. The construction formula (2.8) of Theorem 2.3 shows that the elements of the optimal basis $\{\varphi_1, \dots, \varphi_n\}$, generating the function space, are linear combinations of the functions $f_j \in \mathcal{F}$. We also know that the coefficients of these linear combination are ℓ^2 sequences and hence decay to zero. Thus we can truncate these coefficients and obtain compactly supported generators $\{\varphi_1^c, \dots, \varphi_n^c\}$.

3. Preliminaries on finitely generated shift-invariant spaces

In this section we state some known results about finitely generated shift-invariant spaces that we will need later. See for example [3,6,7,10,14].

We need first to introduce some definitions.

Given $f \in L^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ the *fiber* of f at x is the sequence $\Gamma_x f = \{f(x+k)\colon k \in \mathbb{Z}^d\}$.

If V is a FSIS (recall Definition (1.1)) and $\omega \in [0,1]^d$ we set $V_\omega = \{\Gamma_\omega \hat{f};\ f \in V\}$ the *fiber space* associated to V and ω.

If $\mathcal{M}$ is a closed subspace of a Hilbert space $\mathcal{H}$, throughout this article we will denote by $P_{\mathcal{M}}$ the orthogonal projection operator in $\mathcal{H}$ onto $\mathcal{M}$.

With this notation we have:

Lemma 3.1. *If $f \in L^2(\mathbb{R}^d)$, then*

(1) *The sequence $(\Gamma_\omega \hat{f})_k = (\hat{f}(\omega + k))$ is a well-defined sequence in $\ell_2(\mathbb{Z}^d)$ a.e. $\omega \in \mathbb{R}^d$; and*
(2) *$\|\Gamma_\omega \hat{f}\|_{\ell_2}$ is a measurable function of ω and $\|f\|^2 = \|\hat{f}\|^2 = \int_{[0,1]^d} \|\Gamma_\omega \hat{f}\|^2_{\ell_2}\, d\omega$.*

Lemma 3.2. *Let V be a FSIS in $L^2(\mathbb{R}^d)$. Then we have*:

(i) *V_ω is a closed subspace of $\ell_2(\mathbb{Z}^d)$ for almost all $\omega \in [0,1]^d$.*
(ii) *$V = \{f \in L^2(\mathbb{R}^d)\colon\ \Gamma_\omega \hat f \in V_\omega$ for almost all $\omega \in [0,1]^d\}$.*
(iii) *For each $f \in L^2(\mathbb{R}^d)$ we have that $\|\Gamma_\omega(\widehat{P_V f})\|_{\ell_2}$ is a measurable function of the variable ω and $\Gamma_\omega(\widehat{P_V f}) = \Gamma_\omega P_{\hat V}\hat f = P_{V_\omega}(\Gamma_\omega \hat f)$.*
(iv) *Let $\varphi_1, \dots, \varphi_r \in L^2(\mathbb{R}^d)$. We have that*
  (a) *$\{\varphi_1, \dots, \varphi_r\}$ is a set of generators of V, if and only if the fibers $\Gamma_\omega\hat\varphi_1, \dots, \Gamma_\omega\hat\varphi_r$ span V_ω for almost all $\omega \in [0,1]^d$,*
  (b) *the integer translates of $\varphi_1, \dots, \varphi_r$ are a frame of V, if and only if $\Gamma_\omega\hat\varphi_1, \dots, \Gamma_\omega\hat\varphi_r$ are a frame of V_ω with the same frame bounds, for almost all $\omega \in [0,1]^d$.*

Lemma 3.3. *Let $\mathcal{F} = \{f_1, \dots, f_m\}$ be functions in $L^2(\mathbb{R}^d)$ and let $A(\omega)$ be the infinite matrix $A_{kj}(\omega) = (\Gamma_\omega \hat f_j)(k) = \hat f_j(\omega + k)$, $j = 1, \dots, m$, $k \in \mathbb{Z}^d$, and $\omega \in \mathbb{R}^d$. Then $G_{\mathcal{F}}(\omega) = A^t(\omega)\overline{A(\omega)}$, and $\operatorname{rank} G_{\mathcal{F}}(\omega) = \operatorname{rank} A(\omega) = \operatorname{rank} A^*(\omega)$, a.e. $\omega \in \mathbb{R}^d$. In particular, $G_{\mathcal{F}}(\omega) = \mathfrak{G}(\Gamma_\omega \hat f_1, \dots, \Gamma_\omega \hat f_m)$.*

4. Proofs

To prove the theorems in Section 2, we proceed in several steps. First we reduce the optimization problem into an uncountable set of finite dimensional problems in the Hilbert space $\mathcal{H} = \ell_2(\mathbb{Z}^d)$. We then apply the Eckart–Young theorem to prove that the reduced problems have solutions. Finally, we construct the generators of the optimal space patching together the solutions of the reduced problems to obtain the solution to the original problem.

4.1. Reduction

In this section, we reduce Problem 1 to a set of finite dimensional problems. To see this let us first consider the following:

Problem 2. *Let H be a Hilbert space, n, m positive integers and $A = \{a_1, \dots, a_m\}$ a set of vectors in $\mathcal{H}$. We want to find a closed subspace S of $\mathcal{H}$ with* $\dim(S) \leqslant n$ *that satisfies*

$$\sum_{i=1}^m \|a_i - P_S a_i\|^2 \leqslant \sum_{i=1}^m \|a_i - P_{S'} a_i\|^2 \tag{4.1}$$

for every subspace $S' \subset \mathcal{H}$ with $\dim(S') \leqslant n$.

If such an S exists, we say that S solves Problem 2 for the data (A, n).

If $B = \{b_1, \dots, b_r\}$ is a set of vectors from $\mathcal{H}$ with $S = \operatorname{span}(B)$ we will say that the vectors in B solve Problem 2 for the data (A, n). *The error for Problem 2 is*

$$\mathfrak{E}(A, n) = \min_{\dim(S') \leqslant n} \sum_{i=1}^m \|a_i - P_{S'} a_i\|^2.$$

Note that in Problem 2 we take the minimum over all subspaces of dimension less than n, while in Problem 1 the minimization is taken over a particular class of infinite dimensional subspaces, so the two problems are essentially different.

In the next section we state and prove an extension of the Eckart–Young theorem. We conclude from this extension that Problem 2 always has a solution for any set of data (A, n) in an arbitrary Hilbert space. That is, given A and n there always exists a subspace S with $\dim(S) \leqslant n$ satisfying (4.1). We will also see that a solution S can be chosen in such a way that $S \subset \operatorname{span}(A)$ when $n \leqslant \dim(\operatorname{span}(A))$.

Before proving these results let us see how Problem 2 helps our original question.

Let $\mathcal{F} = \{f_1, \dots, f_m\} \subset L^2(\mathbb{R}^d)$. We want to find out if there exists $V \in \mathcal{V}_n$ such that V minimizes $\sum_{i=1}^m \|f_i - P_V f_i\|^2$. Using Lemma 3.1 we obtain that for any $V \in \mathcal{V}_n$,

$$\sum_{i=1}^m \|f_i - P_V f_i\|^2 = \sum_{i=1}^m \int_{[0,1]^d} \|\Gamma_\omega \hat f_i - \Gamma_\omega \widehat{P_V f_i}\|_{\ell_2}^2 \, \mathrm{d}\omega = \int_{[0,1]^d} \sum_{i=1}^m \|\Gamma_\omega \hat f_i - \Gamma_\omega \widehat{P_V f_i}\|_{\ell_2}^2 \, \mathrm{d}\omega. \tag{4.2}$$

By Lemma 3.2(iii), $\Gamma_\omega \widehat{P_V f_i} = P_{V_\omega} \Gamma_\omega \hat{f}_i$. So from (4.2) we conclude that,

$$\sum_{i=1}^{m} \|f_i - P_V f_i\|^2 = \int_{[0,1]^d} \sum_{i=1}^{m} \|\Gamma_\omega \hat{f}_i - P_{V_\omega} \Gamma_\omega \hat{f}_i\|_{\ell_2}^2 \, d\omega. \tag{4.3}$$

The sum inside the integral on the right-hand side of (4.3) is of the same type than the sum that is involved in Problem 2 in the case that $\mathcal{H} = \ell_2(\mathbb{Z}^d)$ and $S = V_\omega$. Since we are assuming that Problem 2 always has a solution, we know that for almost each $\omega \in [0,1]^d$ there exists a subspace $S_\omega \subset \ell_2(\mathbb{Z}^d)$ that solves Problem 2 for the data $(\mathcal{F}_\omega, n)$ where $\mathcal{F}_\omega = \{\Gamma_\omega \hat{f}_1, \ldots, \Gamma_\omega \hat{f}_m\}$. Note that the subspace S_ω does not need to be related with the fiber space of any FSIS. If the function $\omega \mapsto \sum_{i=1}^{m} \|\Gamma_\omega \hat{f}_i - P_{S_\omega} \Gamma_\omega \hat{f}_i\|_{\ell_2}^2$, were a measurable function of ω then we would have

$$\int_{[0,1]^d} \sum_{i=1}^{m} \|\Gamma_\omega \hat{f}_i - P_{S_\omega} \Gamma_\omega \hat{f}_i\|_{\ell_2}^2 \, d\omega \leqslant \sum_{i=1}^{m} \|f_i - P_{V'} f_i\|^2 \tag{4.4}$$

for every $V' \in \mathcal{V}_n$.

Therefore, in case that there exists a FSIS $V \in \mathcal{V}_n$ such that $V_\omega = S_\omega$ a.e. $\omega \in [0,1]^d$, then by Lemmas 3.1 and 3.2 the above function would be measurable and V necessarily will be a solution to Problem 1, since

$$\begin{aligned}\sum_{i=1}^{m} \|f_i - P_V f_i\|^2 = \int_{[0,1]^d} \sum_{i=1}^{m} \|\Gamma_\omega \hat{f}_i - P_{S_\omega} \Gamma_\omega \hat{f}_i\|_{\ell_2}^2 \, d\omega &\leqslant \int_{[0,1]^d} \sum_{i=1}^{m} \|\Gamma_\omega \hat{f}_i - P_{V'_\omega} \Gamma_\omega \hat{f}_i\|_{\ell_2}^2 \, d\omega \\ &= \sum_{i=1}^{m} \|f_i - P_{V'} f_i\|^2 \end{aligned}\tag{4.5}$$

for every $V' \in \mathcal{V}_n$.

We will see later that such a FSIS indeed exists. More precisely we will construct a set of generators such that its integer translates form a frame of the optimal FSIS. We will do that by patching together the fibers of the generators of each of the optimal subspaces S_ω.

4.2. Solution to Problem 2

We now prove that Problem 2 always has a solution.

Theorem 4.1. *Let $\mathcal{H}$ be an infinite dimensional Hilbert space, $\mathcal{F} = \{f_1, \ldots, f_m\} \subset \mathcal{H}$, $\mathcal{X} = \operatorname{span}\{f_1, \ldots, f_m\}$, $\lambda_1 \geqslant \cdots \geqslant \lambda_m$ the eigenvalues of the matrix $\mathfrak{G}(\mathcal{F})$ defined as in (2.5) and $y_1, \ldots, y_m \in \mathbb{C}^m$, with $y_i = (y_{i1}, \ldots, y_{im})^t$ orthonormal left eigenvectors associated to the eigenvalues $\lambda_1, \ldots, \lambda_m$. Let $r = \dim \mathcal{X} = \operatorname{rank} \mathfrak{G}(\mathcal{F})$.*

Define the vectors $q_1, \ldots, q_n \in \mathcal{H}$ by

$$q_i = \tilde{\sigma}_i \sum_{j=1}^{m} y_{ij} f_j, \quad i = 1, \ldots, n, \tag{4.6}$$

where $\tilde{\sigma}_i = \lambda_i^{-1/2}$ if $\lambda_i \neq 0$, and $\tilde{\sigma}_i = 0$ otherwise. Then $\{q_1, \ldots, q_n\}$ is a Parseval frame of $W = \operatorname{span}\{q_1, \ldots, q_n\}$ and the subspace W is optimal in the sense that

$$\mathfrak{E}(\mathcal{F}, n) = \sum_{i=1}^{m} \|f_i - P_W f_i\|^2 \leqslant \sum_{i=1}^{m} \|f_i - P_{W'} f_i\|^2, \quad \forall \text{ subspace } W',\ \dim W' \leqslant n.$$

Furthermore we have the following formula for the error

$$\mathfrak{E}(\mathcal{F}, n) = \sum_{i=n+1}^{m} \lambda_i. \tag{4.7}$$

Remark. If r is small (i.e. $r \leqslant n$) then all the vectors $q_{r+1}, \ldots, q_n$ are null and $\{q_1, \ldots, q_r\}$ is an orthonormal set.

One could also choose $q_{r+1},\dots,q_n$ to be any orthonormal set in the orthogonal complement of $\mathcal{X}$ and so obtain an orthonormal set of n elements and the formula for the error would still hold.

If $\mathcal{H}$ is finite dimensional and $n \leqslant r$, then Theorem 4.1 is a consequence of the Eckart–Young theorem (see Appendix A). To prove Theorem 4.1 we will reduce it to the finite dimensional case and then use the Eckart–Young result.

We first need the following lemma:

Lemma 4.2. *Let $\mathcal{H}$ be a Hilbert space, $\{f_1,\dots,f_m\} \subset \mathcal{H}$, $\mathcal{X} = \operatorname{span}\{f_1,\dots,f_m\}$. Assume that there exists $M \subset \mathcal{H}$ with $\dim M \leqslant n$ such that*

$$\sum_{i=1}^{m} \|f_i - P_M f_i\|^2 \leqslant \sum_{i=1}^{m} \|f_i - P_{M'} f_i\|^2$$

for any subspace $M' \subset \mathcal{H}$ with $\dim M' \leqslant n$, then there exists $W \subset \mathcal{X}$, with $\dim W \leqslant n$, such that

$$\sum_{i=1}^{m} \|f_i - P_W f_i\|^2 = \sum_{i=1}^{m} \|f_i - P_M f_i\|^2.$$

Proof. Define the subspace $W = P_{\mathcal{X}} M$ as the orthogonal projection of M onto $\mathcal{X}$. By construction, $W \subset \mathcal{X}$, and $\dim W \leqslant n$.

Let $f \in \mathcal{X}$, then we have

$$\|f - P_W f\|^2 = \inf\bigl\{\|f-g\|^2\colon g \in W\bigr\} \leqslant \|f - P_{\mathcal{X}} P_M f\|^2 = \|P_{\mathcal{X}} f - P_{\mathcal{X}} P_M f\|^2 \leqslant \|f - P_M f\|^2. \qquad \square$$

This lemma shows that, in a possibly infinite dimensional Hilbert space $\mathcal{H}$, the problem of finding a finite dimensional subspace $M \subset \mathcal{H}$ with $\dim M \leqslant n$ that "best approximates" m vectors $\{f_1,\dots,f_m\}$, can always be reduced to a search in the finite dimensional space $\mathcal{X} = \operatorname{span}\{f_1,\dots,f_m\}$.

We now prove Theorem 4.1.

Proof of Theorem 4.1. Let $\tau : \mathcal{X} \to \mathbb{C}^m$ be an isometric isomorphism with its image. Set $b_i = \tau(f_i)$, and let B be the matrix having the vectors b_i as columns. So, $r = \dim \mathcal{X} = \operatorname{rank}(B)$ and $B^t \overline{B}$ coincides with $\mathfrak{G}(\mathcal{F}) = \{\langle f_i, f_j\rangle_{\mathcal{H}}\}_{i,j}$.

Choose orthonormal left eigenvectors $y_1,\dots,y_m \in \mathbb{C}^m$, with $y_i = (y_{i1},\dots,y_{im})^t$ associated to the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_m$ of $B^t\overline{B}$, and define the vectors

$$u_i = \tilde{\sigma}_i \sum_{j=1}^{m} y_{ij} b_j, \quad i = 1,\dots,n, \tag{4.8}$$

where as before $\tilde{\sigma}_i = \lambda_i^{-1/2}$ if $\lambda_i \neq 0$, and $\tilde{\sigma}_i = 0$ otherwise.

Then, if $n \leqslant r$ by Theorem 4.6 in Appendix A, the subspace $M \subset \mathbb{C}^m$, $M = \operatorname{span}\{u_1,\dots,u_n\}$ satisfies:

$$\sum_{i=1}^{m} \|b_i - P_M b_i\|^2 \leqslant \sum_{i=1}^{m} \|b_i - P_{M'} b_i\|^2 \tag{4.9}$$

for every subspace $M' \subset \mathbb{C}^m$ with $\dim M' \leqslant n$.

If however, $n \geqslant r$ then the left side of (4.9) is 0 and therefore the inequality is also satisfied.

Setting $W = \tau^{-1}(M)$ and noting that $P_M \tau = \tau P_W$ we have from (4.9)

$$\mathfrak{E}(B,n) = \mathfrak{E}(\mathcal{F},n) = \sum_{i=1}^{m} \|f_i - P_W f_i\|^2 \leqslant \sum_{i=1}^{m} \|f_i - P_{W'} f_i\|^2 \tag{4.10}$$

for every subspace $W' \subset \mathcal{H}$ with $\dim W' \leqslant n$. So, $W \subset \mathcal{H}$ is optimal for $(\mathcal{F}, n)$ and $q_i = \tau^{-1}(u_i)$, $i = 1,\dots,n$ is a Parseval frame for W. Furthermore, the formula (4.7) also holds. $\square$

Remarks. (i) If $n > m$ the optimal space W is not unique since any space W' of dimension n containing $\operatorname{span}\{a_1,\dots,a_m\}$ will be optimal. The same argument also shows that the space W is not unique if $n > r = \dim \mathcal{X}$.

(ii) If $n \leqslant r$, the vectors u_i and y_i are related by $\sqrt{\lambda_i}\, u_i = A y_i$ as described in Appendix A.

4.3. Solution to Problem 1

In order to solve Problem 1, we need the following technical proposition concerning the measurability of the eigenvalues and the existence of measurable eigenvectors of a non-negative matrix with measurable entries (cf. [14, Lemma 2.3.5]).

Lemma 4.3. *Let $G = G(\omega)$ be an $m \times m$ self-adjoint matrix of measurable functions defined on a measurable subset $E \subset \mathbb{R}^d$ with eigenvalues $\lambda_1(\omega) \geqslant \lambda_2(\omega) \geqslant \cdots \geqslant \lambda_m(\omega)$. Then the eigenvalues λ_i, $i = 1, \ldots, m$, are measurable on E and there exists an $m \times m$ matrix of measurable functions $U = U(\omega)$ on E such that $U(\omega)U^*(\omega) = I$ a.e. $\omega \in E$ and such that*

$$G(\omega) = U(\omega)\Lambda(\omega)U^*(\omega) \quad \text{a.e. } \omega \in E, \tag{4.11}$$

where $\Lambda(\omega) := \operatorname{diag}(\lambda_1(\omega), \ldots, \lambda_m(\omega))$.

Proof of Theorems 2.1 and 2.3. In what follows we will apply Theorem 4.1 to find the solution to Problem 1. As before, let $\mathcal{F} = \{f_1, \ldots, f_m\} \subset L^2(\mathbb{R}^d)$ and for $\omega \in [0,1]^d$ let $G_{\mathcal{F}}(\omega)$ be the associated Gramian matrix with eigenvalues $\lambda_1(\omega) \geqslant \cdots \geqslant \lambda_m(\omega) \geqslant 0$. Let $U(\omega)$ be a measurable $m \times m$ matrix as in Lemma 4.3. Since $G_{\mathcal{F}}(\omega)$ is $\mathbb{Z}^d$-periodic on $\mathbb{R}^d$, we can choose $U(\omega)$ to be $\mathbb{Z}^d$-periodic as well. Let $U_i(\omega)$ denote the ith row of $U(\omega)$. Multiplying (4.11) on the left by $U^*(\omega)$ shows that $y_i(\omega) := U_i(\omega)^*$ is a left-eigenvector of $G(\omega)$ with eigenvalue $\lambda_i(\omega)$ for $i = 1, \ldots, m$. Furthermore, the left eigenvectors $y_i(\omega) = (y_{i1}(\omega), \ldots, y_{im}(\omega))^t$, $i = 1, \ldots, m$, form an orthonormal basis of $\mathbb{C}^m$.

For each fixed $\omega \in [0,1]^d$, we consider Problem 2 in the space $\ell_2(\mathbb{Z}^d)$ for the data $(\mathcal{F}_\omega, n)$ with $\mathcal{F}_\omega = \{\Gamma_\omega \hat{f}_1, \ldots, \Gamma_\omega \hat{f}_m\}$. Define $q_1(\omega), \ldots, q_n(\omega) \in \ell_2(\mathbb{Z}^d)$ by

$$q_i(\omega) = \tilde{\sigma}_i(\omega) \sum_{j=1}^{m} y_{ij}(\omega) \Gamma_\omega \hat{f}_j, \quad i = 1, \ldots, n, \tag{4.12}$$

where $\tilde{\sigma}_i(\omega) = \lambda_i^{-1/2}(\omega)$ if $\lambda_i(\omega) \neq 0$, and $\tilde{\sigma}_i(\omega) = 0$ otherwise. Since $G_{\mathcal{F}}(\omega) = \mathfrak{G}(\mathcal{F}_\omega)$ (see Lemma 3.3), Theorem 4.1 shows that the space $S_\omega := \operatorname{span}\{q_1(\omega), \ldots, q_n(\omega)\}$ optimizes Problem 2. Moreover, the vectors $\{q_1(\omega), \ldots, q_n(\omega)\}$ form a Parseval frame for S_ω and we have the following formula for the error:

$$\mathfrak{E}(\mathcal{F}_\omega, n) = \sum_{i=n+1}^{m} \lambda_i(\omega). \tag{4.13}$$

Define now the functions $h_i : \mathbb{R}^d \to \mathbb{C}$, $i = 1, \ldots, m$,

$$h_i(\omega) = \tilde{\sigma}_i(\omega) \sum_{j=1}^{m} y_{ij}(\omega) \hat{f}_j(\omega). \tag{4.14}$$

Since $\tilde{\sigma}_i$ and y_i are measurable functions of ω, then h_i is also measurable. Moreover, h_i is in $L^2(\mathbb{R}^d)$ as the following simple argument shows. Since

$$\left|h_i(\omega)\right|^2 = h_i(\omega)\bar{h}_i(\omega) = \tilde{\sigma}_i(\omega)^2 \sum_{j,s=1}^{m} y_{ij}(\omega) \hat{f}_j(\omega) \bar{\hat{f}}_s(\omega) \bar{y}_{is}(\omega)$$

we have (using that if y_i is a left eigenvector of the self-adjoint matrix $G_{\mathcal{F}}$, then $\bar{y}_i$ is a right eigenvector for that matrix associated to the same eigenvalue),

$$\begin{aligned}\sum_{k \in \mathbb{Z}^d} \left|h_i(\omega + k)\right|^2 &= \tilde{\sigma}_i(\omega)^2 \sum_{j=1}^{m} y_{ij}(\omega) \sum_{s=1}^{m} \left[G_{\mathcal{F}}(\omega)\right]_{js} \bar{y}_{is}(\omega) \\ &= \tilde{\sigma}_i(\omega)^2 \lambda_i(\omega) \sum_{j=1}^{m} y_{ij}(\omega) \bar{y}_{ij}(\omega) = \tilde{\sigma}_i(\omega)^2 \lambda_i(\omega).\end{aligned} \tag{4.15}$$

If $\lambda_i(\omega) \neq 0$ then the product in (4.15) is one, otherwise it is zero. That is $\sum_{k\in\mathbb{Z}^d} |h_i(\omega+k)|^2 = \mathbf{1}_{\{\omega:\ \lambda_i(\omega)>0\}}$ and by Lemma 3.1, $\|h_i\| \leqslant 1$.

Now define functions $\varphi_1, \ldots, \varphi_n$ in $L^2(\mathbb{R}^d)$ by

$$\hat{\varphi}_i(\omega) = h_i(\omega), \quad i = 1, \ldots, n,$$

and let $V = S(\varphi_1, \ldots, \varphi_n)$. The space V is a shift-invariant space of length no bigger than n. So $V \in \mathcal{V}_n$. Furthermore, by Lemma 3.2 (iv-a), the space V_ω is spanned by $\Gamma_\omega \hat{\varphi}_i$, $i = 1, \ldots, n$.

Since $(\Gamma_\omega \hat{\varphi}_i)(k) = h_i(\omega+k) = q_i(\omega)(k)$, $k \in \mathbb{Z}^d$, $i = 1, \ldots, n$ a.e., then $V_\omega = S_\omega$ (the optimal space for the data $(\mathcal{F}_\omega, n)$) in $\ell_2(\mathbb{Z}^d)$.

By Eq. (4.5) and the comment before, V is optimal, that is V solves Problem 1 for the data $(\mathcal{F}, n)$.

Now, since $\{\Gamma_\omega \hat{\varphi}_1, \ldots, \Gamma_\omega \hat{\varphi}_n\}$ is a Parseval frame of S_ω for a.e. $\omega \in [0,1]^d$ then by Lemma 3.2 (iv-b) the integer translates of $\varphi_1, \ldots, \varphi_n$ form a Parseval frame of V. On the other hand, formula (4.3) says that

$$\mathcal{E}(\mathcal{F}, n) = \int\limits_{[0,1]^d} \mathfrak{E}(\mathcal{F}_\omega, n)\, d\omega. \tag{4.16}$$

Thus using (4.13) we have that $\mathcal{E}(\mathcal{F}, n) = \sum_{i=n+1}^{m} \int_{[0,1]^d} \lambda_i(\omega)\, d\omega$. □

Proof of Theorem 2.4. Under the hypothesis of Theorem 2.4, Theorem 4.1 and the Remark after it guarantee the uniqueness of the optimal spaces S_ω associated to the data $(\mathcal{F}_\omega, n)$ for almost all ω. Since these fiber spaces characterize the optimal space V, then the theorem follows. □

Acknowledgments

We thank Yves Meyer for the encouragement to write this paper. We also thank the anonymous referee for his helpful comments and suggestions.

Appendix A

A.1. Best linear approximation and the SVD

Here we review the singular value decomposition (SVD) of a matrix and its relation to finite dimensional least-squares problems. For an overview see [18], and for a very detailed treatment see for example [12].

We start with the following proposition.

Proposition 4.4 *(SVD). Let* $A = [a_1, a_2, \ldots, a_m]$ *be the matrix with columns* $a_i \in \mathbb{C}^N$, $m \leqslant N$. *Let* $r := \dim \operatorname{span}\{a_1, \ldots, a_m\}$. *Then there are* m *numbers* $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0$, *an orthonormal collection of* m *(column) vectors* $y_1, \ldots, y_m \in \mathbb{C}^m$, *and an orthonormal collection of* m *(column) vectors* $u_1, \ldots, u_m \in \mathbb{C}^N$ *such that*

$$A = \sum_{k=1}^{m} \sqrt{\lambda_k} u_k y_k^* = U \Lambda^{1/2} Y^*, \tag{4.17}$$

where $U \in \mathbb{C}^{N\times m}$ *is the matrix* $U = [u_1, \ldots, u_m]$, $\Lambda^{1/2} = \operatorname{diag}(\lambda_1^{1/2}, \ldots, \lambda_m^{1/2})$, *and* $Y = [y_1, \ldots, y_m] \in C^{m\times m}$ *with* $U^*U = I_m = Y^*Y = YY^*$.

The representation of A *given in* (4.17) *is called* the singular value decomposition (*SVD*) *of* A.

The SVD of a matrix A can be obtained as follows. Consider the matrix $A^*A \in \mathbb{C}^{m\times m}$. Since A^*A is self-adjoint and positive semi-definite, its eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$ are non-negative and the associated eigenvectors $y_1, \ldots, y_m$ can be chosen to form an orthonormal basis of $\mathbb{C}^m$. Note that the rank r of A corresponds to the largest index i such that $\lambda_i > 0$. The left singular vectors $u_1, \ldots, u_r$ can then be obtained from

$$\sqrt{\lambda_i} u_i = A y_i, \quad \text{that is} \quad u_i = \lambda_i^{-1/2} \sum_{j=1}^{m} y_{ij} a_j \quad (1 \leqslant i \leqslant r).$$

Here $y_i = (y_{i1}, \dots, y_{im})^t$. The remaining left singular vectors $u_{r+1}, \dots, u_m$ can be chosen to be any orthonormal collection of $m - r$ vectors in $\mathbb{C}^N$ that are perpendicular to $\operatorname{span}\{a_1, \dots, a_m\}$. One may then readily verify that (4.17) holds.

The *Frobenius norm* of a matrix $X = [x_1, \dots, x_m] \in \mathbb{C}^{N\times m}$ is $\|X\|_F = \operatorname{tr}(X^*X)$, where tr denotes the trace of a matrix.

Now, the following approximation theorem of Schmidt (cf. [17]) and later rediscovered by Eckart and Young [8] shows that the SVD can be used to find the subspace of dimension n that is closest to a given finite numbers of vectors.

Theorem 4.5. *Let $\{a_1, \dots, a_m\}$, be a set of vectors in $\mathbb{C}^N$, such that $r = \dim(\operatorname{span}\{a_1, \dots, a_m\})$, and suppose $A = [a_1, \dots, a_m]$, has SVD $A = U\Lambda^{1/2}Y^*$ with $0 < n \leqslant r$. Then $A_n := \sum_{j=1}^{n} \sqrt{\lambda_j} u_j y_j^*$ satisfies*

$$\|A - A_n\|_F = \min_{\operatorname{rank} B \leqslant n} \|A - B\|_F = \left(\sum_{j=n+1}^{r} \lambda_j\right)^{1/2}.$$

If $\lambda_{n+1} \neq \lambda_n$, then A_n is the unique such matrix of rank at most n.

Equivalently,

Theorem 4.6. *Let $\{a_1, \dots, a_m\}$, be a set of vectors in $\mathbb{C}^N$ such that $r = \dim(\operatorname{span}\{a_1, \dots, a_m\})$, and suppose $A = [a_1, \dots, a_m]$, has SVD $A = U\Lambda^{1/2}Y^*$ and that $0 < n \leqslant r$. If $W = \operatorname{span}\{u_1, \dots, u_n\}$, then*

$$\{P_W a_1, \dots, P_W a_m\} = \sum_{i=1}^{n} \sqrt{\lambda_i} u_i y_i^* = A_n$$

and

$$\sum_{i=1}^{m} \|a_i - P_W a_i\|_2^2 \leqslant \sum_{i=1}^{m} \|a_i - P_M a_i\|_2^2, \quad \forall M,\ \dim M \leqslant n, \tag{4.18}$$

and the space W is unique if $\lambda_{n+1} \neq \lambda_n$.

References

[1] A. Aldroubi, K.-H. Gröchenig, Non-uniform sampling in shift-invariant space, SIAM Rev. 43 (4) (2001) 585–620.
[2] P. Binev, A. Cohen, W. Dahmen, R. DeVore, Universal algorithms for learning theory part I: Piecewise constant functions, J. Mach. Learn. Res. 6 (2005) 1297–1321.
[3] M. Bownik, The structure of shift-invariant subspaces of $L^2(\mathbb{R}^n)$, J. Funct. Anal. 177 (2000) 282–309.
[4] P. Cazassa, O. Christensen, Perturbation of operators and applications to frame theory, J. Fourier Anal. Appl. 3 (1997) 543–557.
[5] F. Cucker, S. Smale, On the mathematical foundations of learning, Bull. Amer. Math. Soc. (N.S.) 39 (1) (2002) 1–49 (electronic).
[6] C. de Boor, R. DeVore, A. Ron, Approximation from shift-invariant subspaces of $L_2(R^d)$, Trans. Amer. Math. Soc. 341 (1994) 787–806.
[7] C. de Boor, R. DeVore, A. Ron, The structure of finitely generated shift-invariant subspaces of $L_2(R^d)$, J. Funct. Anal. 119 (1994) 37–78.
[8] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, Psychometrica 1 (1936) 211–218.
[9] D. Han, D. Larson, Frames, bases and group representations, Mem. Amer. Math. Soc. 147 (697) (2000), x+94 pp.
[10] H. Helson, Lectures on Invariant Subspaces, Academic Press, London, 1964.
[11] E. Hernández, D. Labate, G. Weiss, A unified characterization of reproducing systems generated by a finite family, II, J. Geom. Anal. 12 (2002) 615–662.
[12] R. Horn, C. Johnson, Matrix Analysis, Cambridge Univ. Press, Cambridge, 1985.
[13] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1984.
[14] A. Ron, Z. Shen, Frames and stable bases for shift invariant subspaces of $L2(Rd)$, Canad. J. Math. 47 (1995) 1051–1094.
[15] S. Smale, D.-X. Zhou, Estimating the approximation error in learning theory, Anal. Appl. (Singap.) 1 (2003) 17–41.
[16] S. Smale, D.-X. Zhou, Shannon sampling and function reconstruction from point values, Bull. Amer. Math. Soc. 41 (2004) 279–305.
[17] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener, Math. Ann. 63 (1907) 433–476.
[18] G.W. Stewart, On the early history of the singular value decomposition, SIAM Rev. 35 (1993) 551–566.