14

Learning the Right Model from the Data

Akram Aldroubi[1], Carlos Cabrelli[2], and Ursula Molter[2]

[1] Department of Mathematics, Vanderbilt University, Nashville, TN 37240, USA
aldroubi@math.vanderbilt.edu

[2] Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón I, 1428 Capital Federal, Argentina, and CONICET, Argentina
cabrelli@dm.uba.ar, umolter@dm.uba.ar

Summary. In this chapter we discuss the problem of finding the shift-invariant space model that best fits a given class of observed data $\mathcal{F}$. If the data is known to belong to a fixed—but unknown—shift-invariant space $V(\Phi)$ generated by a vector function Φ, then we can probe the data $\mathcal{F}$ to find out whether the data is sufficiently rich for determining the shift-invariant space. If it is determined that the data is not sufficient to find the underlying shift-invariant space V, then we need to acquire more data. If we cannot acquire more data, then instead we can determine a shift-invariant subspace $S \subset V$ whose elements are generated by the data. For the case where the observed data is corrupted by noise, or the data does not belong to a shift-invariant space $V(\Phi)$, then we can determine a space $V(\Phi)$ that fits the data in some optimal way. This latter case is more realistic and can be useful in applications, e.g., finding a shift-invariant space with a small number of generators that describes the class of chest X-rays.

To John, whose mathematics and humanity have inspired us.

14.1 Introduction

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

$$V(\Phi) := \left\{ f = \sum_{i=1}^{n} \sum_{j \in \mathbb{Z}^d} \alpha_i(j)\phi_i(\cdot + j) : \alpha_i \in l^2(\mathbb{Z}^d),\ i = 1, \dots, n \right\}, \quad (14.1)$$

where $\Phi = [\phi_1, \phi_2, \dots, \phi_n]^t$ is a column vector whose elements ϕ_i are functions in $L^2(\mathbb{R}^d)$. These functions are *a set of generators* for the space $V = V(\Phi)$. For example, if $n = 1$, $d = 1$, and $\phi(x) = \text{sinc}(x)$, then the underlying space is the space of band-limited functions (often used in communications) [4], [5], [6].

However, in most applications, the shift-invariant space chosen to describe the underlying class of signals is not derived from experimental data—for example most 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. Thus, in order to derive the appropriate signal model for a class of signals, we consider the following two types of problems:

(I) Given a class of signals belonging to a certain fixed—but unknown—shift-invariant space V, the problem is whether it is possible to determine the space V from a set of m experimental data $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$, where f_i are observed functions (signals) belonging to $V(\Phi)$.

(II) Given a large set of experimental data $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$, where f_i are observed functions (signals) that are not necessarily from a shift-invariant space with a small fixed number of generators, we wish to determine some small space V that models the signals in "some" best way.

For Problem I to be meaningful, we must have some a priori assumption about our signal space V. In particular, we assume that V is a shift-invariant space that can be generated by a set of exactly n generators, $\Phi = [\phi_1, \phi_2, \ldots, \phi_n]^t$, such that $\{\phi_i(\cdot - k) : k \in \mathbb{Z}, i = 1, \ldots, n\}$ forms a Riesz basis for $V(\Phi)$. If a finite set $\mathcal{F}$ of signals is sufficient to determine $V(\Phi)$, then $\mathcal{F}$ is called a *determining set for* $V(\Phi)$. The goal is to see if we can perform operations on the observations $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ to deduce whether they are sufficient to determine the unknown shift-invariant space $V(\Phi)$, and if so, use them to find some set of generators Ψ for $V(\Phi)$, i.e., find some $\Psi = [\psi_1, \psi_2, \ldots, \psi_n]^t$ such that $V(\Psi) = V(\Phi)$. If the observations are not sufficient to determine $V(\Phi)$, then we need to obtain more observations until a determining set is found.

This then becomes a learning problem: If the data is insufficient to determine the model, then the set $S(\mathcal{F}) = \text{closure}_{L^2}\big(\text{span}\{f_i(\cdot - k) : i = 1, \ldots, m, k \in \mathbb{Z}^d\}\big)$ is a proper shift-invariant subspace of V. Thus the data determines some "smaller" shift-invariant space. The acquisition of new data will allow us to "learn" more about the right model, i.e., with the new information we can obtain a more complete description of the space.

In practice however, 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. For these three more realistic cases, we must consider Problem II.

Similarly to Problem I, we must impose some a priori conditions on the space V. In particular, we will search for the optimal space V among those spaces that are generated by exactly n generators. Consider the class $\mathcal{V}$ of all the shift-invariant spaces that are generated by some set of generators $\Phi = [\phi_1, \phi_2, \ldots, \phi_n]^t$, $\phi_i \in L^2(\mathbb{R}^d)$, with the property that $\{\phi_i(\cdot - k) : k \in \mathbb{Z}^d,\ i = 1, \ldots, n\}$ is a Riesz basis for $V(\Phi)$. The problem is then to find a space $V \in \mathcal{V}$ such that

$$V = \operatorname*{argmin}_{V \in \mathcal{V}} \sum_{i=1}^{m} w_i \, \|f_i - P_V f_i\|^2, \tag{14.2}$$

where w_i are positive weights and where P_V is the orthogonal projection on V. 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 use the observations $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ to find some set of generators $\Psi = [\psi_1, \psi_2, \ldots, \psi_n]^t$ that generates the optimal space $V = V(\Psi)$ in (14.2).

14.2 Notation and Preliminaries

Throughout this chapter, we assume that the unknown space V is a Riesz shift-invariant space, i.e., a shift-invariant space that has a set of generators $\Phi = [\phi_1, \ldots, \phi_n]^t$ such that $\{\phi_i(x - k) : i = 1, \ldots, n,\ k \in \mathbb{Z}^d\}$ forms a Riesz basis for V. That is, there exist $0 < A \le B$ such that for all $f \in V(\Phi)$,

$$A\|f\|^2 \le \sum_{i=1}^{n} \sum_{k \in \mathbb{Z}^d} |\langle \phi_i(\cdot - k), f\rangle|^2 \le B\|f\|^2.$$

This Riesz basis assumption can be restated in the Fourier domain using the Grammian matrix of Φ. Specifically, the Grammian G_Θ of a vector function $\Theta = [\theta_1, \ldots, \theta_n]^t$ is defined by

$$G_\Theta(\omega) = \sum_{k \in \mathbb{Z}^d} \widehat{\Theta}(\omega + k)\, \widehat{\Theta}^*(\omega + k)$$

where $\widehat{\Theta}(\omega) := \int_{\mathbb{R}^d} \Theta(x)\, e^{-2\pi i \omega x}\, dx$, and $\widehat{\Theta}^*$ is the adjoint of $\widehat{\Theta}$. With this definition, it is well known that Φ induces a Riesz basis of the space $V = V(\Phi)$ defined by (14.1) if and only if there exist two positive constants $A > 0$ and $B > 0$ such that

$$AI \le G_\Phi(\omega) \le BI, \quad \text{a.e. } \omega, \tag{14.3}$$

where I is the $n \times n$ identity matrix (see, e.g., [1], [8], [9]). The set $B = \{\phi_i(x-k) : i = 1, \ldots, n,\ k \in \mathbb{Z}^d\}$ forms an orthonormal basis if and only if $A = B = 1$ in (14.3). Throughout the chapter we assume that $\Phi = [\phi_1, \ldots, \phi_n]^t$ satisfies (14.3).

We use $\mathcal{F}$ to indicate a set of functions and F to denote the vector-valued function whose components are the elements of $\mathcal{F}$ in some fixed order.

14.3 Problem I

A complete account of the results considered in this section, with proofs, is contained in [3].

Our main goal is to find necessary and sufficient conditions on subsets $\mathcal{F} = \{f_1, \dots, f_m\}$ of $V(\Phi)$ such that any $g \in V$ can be recovered from $\mathcal{F}$ as defined precisely next. A set $\mathcal{F}$ with such a property will be called a *determining set* for $V(\Phi)$. Specifically we have the following definition.

Definition 14.1. *The set $\mathcal{F} = \{f_1, f_2, \dots, f_m\} \subset V(\Phi)$ is said to be a* determining set *for $V(\Phi)$ if any $g \in V(\Phi)$ can be written as*

$$\widehat{g}(\omega) = \widehat{\alpha}_1(\omega)\,\widehat{f}_1(\omega) + \widehat{\alpha}_2(\omega)\,\widehat{f}_2(\omega) + \cdots + \widehat{\alpha}_m(\omega)\,\widehat{f}_m(\omega),$$

where $\widehat{\alpha}_1, \dots, \widehat{\alpha}_m$ are some 1-periodic measurable functions. In addition, if $\mathcal{F}$ is a determining set of $V(\Phi)$, then we will say that $V(\Phi)$ is determined by *$\mathcal{F}$.*

Remark 14.2. (i) The integer translates of the functions in the set $\mathcal{F} = \{f_1, \dots, f_m\}$ need not form a Riesz basis for V. In fact, series of the form

$$\sum_{i=1}^{m} \sum_{k \in \mathbb{Z}^d} c_i(k) f_i(x-k)$$

need not even be convergent for all $c_i \in l^2$.

(ii) An equivalent definition of a determining set is the following (e.g., see [8, Thm. 1.7]): a set $\mathcal{F}$ is a determining set for $V(\Phi)$ if and only if $V(\Phi) \subset \mathrm{closure}_{L_2}\big(\mathrm{span}\{f_i(x-k) : f_i \in \mathcal{F}\}\big)$.

It is not surprising that if V has a Riesz basis of n generators, then the cardinality m of a determining set $\mathcal{F}$ must be larger or equal to n. This result is stated in the following proposition.

Proposition 14.3. *Let V be a shift-invariant space generated by some Riesz basis $\{\phi_i(x-k) : i = 1, \dots, n,\ k \in \mathbb{Z}^d\}$, where $\Phi = [\phi_1, \dots, \phi_n]^t$ is a vector of functions in V. If $\mathcal{F}$ is a determining set for V, then $card(\mathcal{F}) \geq n$.*

Because of the proposition above, we will only consider sets $\mathcal{F}$ of cardinality m larger than or equal to the number n of the generators for V. Given such a set $\mathcal{F}$ there are $L = \binom{m}{n}$ subsets $\mathcal{F}_\ell \subset \mathcal{F}$ of size n. For each such subset $\mathcal{F}_\ell$ of size n, we define the set

$$A_\ell = \{\omega : \det G_{F_\ell}(\omega) \neq 0\}, \qquad 1 \leq \ell \leq L, \tag{14.4}$$

where G_{F_ℓ} is the $n \times n$ Grammian matrix for the vector F_ℓ, and we "disjointize" the sets A_ℓ by introducing the sets $\{B_\ell\}_{\ell=1}^{L}$ defined by

$$B_1 := A_1, \quad B_\ell := A_\ell - \bigcup_{j=1}^{\ell-1} A_j, \quad \ell = 2, \ldots, L.$$

Below, we state (and give a reduced version of the proof of) a theorem from [3] that solves Problem I. The result characterizes determining sets, and produces an orthonormal basis for a shift-invariant space V when it is determined by the data.

Theorem 14.4 ([3]). *A set $\mathcal{F} = \{f_1, \ldots, f_m\} \subset V(\Phi)$ is a determining set for $V(\Phi)$ if and only if the set $\bigcup_{\ell=1}^{L} A_\ell$ has Lebesgue measure one.*

Moreover, if $\mathcal{F}$ is a determining set for $V(\Phi)$, then the vector function

$$\widehat{\Psi}(\omega) := G_{F_1}^{-\frac{1}{2}}(\omega)\, \widehat{F_1}(\omega)\, \chi_{B_1}(\omega) + \cdots + G_{F_L}^{-\frac{1}{2}}(\omega)\, \widehat{F_L}(\omega)\, \chi_{B_L}(\omega) \tag{14.5}$$

generates an orthonormal basis $\{\psi_i(x-k) : i = 1, \ldots, n,\ k \in \mathbb{Z}^d\}$ of $V(\Phi)$.

Proof (Sketch). Since $\mathcal{F}_\ell \subset V(\Phi)$ and $\operatorname{card}(\mathcal{F}_\ell) = n$, we can write $\widehat{F_\ell} = \widehat{C_{F_\ell}} \widehat{\Phi}$ for some $n \times n$ square matrix $\widehat{C_{F_\ell}}$ with $L^2([0,1]^d)$ entries, and we have

$$\begin{aligned} G_{F_\ell}(\omega) &= \sum_k (\widehat{C_{F_\ell}}(\omega+k)\, \widehat{\Phi}(\omega+k))\, (\widehat{C_{F_\ell}}(\omega+k)\, \widehat{\Phi}(\omega+k))^* \\ &= \sum_k \widehat{C_{F_\ell}}(\omega+k)\, \widehat{\Phi}(\omega+k)\, \widehat{\Phi}^*(\omega+k)\, \widehat{C_{F_\ell}}^*(\omega+k) \\ &= \widehat{C_{F_\ell}}(\omega)\, G_\Phi(\omega)\, \widehat{C_{F_\ell}}^*(\omega), \end{aligned}$$

since $\widehat{C_{F_\ell}}(\omega)$ is 1-periodic.

Moreover, since Φ induces a Riesz basis, it follows that G_Φ is positive definite. It is also true that $\widehat{C_{F_\ell}}(\omega)$ is non-singular for a.e. $\omega \in B_\ell$. Thus, G_{F_ℓ} is self-adjoint and positive definite on B_ℓ.

Therefore, if we define Ψ as in (14.5), then it can be seen that the set $\{\psi_i(x-k) : i = 1, \ldots, n,\ k \in \mathbb{Z}^d\}$ forms an orthonormal basis for $V(\Phi)$.

For the converse see [3]. □

Remark 14.5. (i) Theorem 14.4 provides a method for checking whether and when a set of functions generates a fixed (yet unknown) shift-invariant space generated by some unknown Φ of known size n. Since other than the value n, the only requirement is that the set of functions must belong to the same (unknown) shift-invariant space, we can apply the theorem to a set of *observed* functions (the data) if we know that they are all from some shift-invariant space V. We can either determine the space, or conclude that we do not have enough data to do so and need to acquire more data. If we cannot acquire more data, we can still determine the space $S(\mathcal{F}) = \operatorname{closure}_{L^2}(\operatorname{span}\{f_i(\cdot - k) : i = 1, \ldots, m, k \in \mathbb{Z}\})$, which is a subspace of the unknown space V. However, the subspace $S(\mathcal{F})$ is not necessarily generated by a Riesz basis.

(ii) The functions of the orthonormal basis constructed in Theorem 14.4 are in L^2 but not in $L^1 \cap L^2$ in general. Further investigation is needed for the construction of better-localized bases.

14.4 Problem II

The intuition—or idea—behind Problem II is that one has a large amount of data (for example the data base of all chest X-rays during the last 10 years). The space

$$\mathcal{S}(\mathcal{F}) = \text{closure}_{L_2}\left(\text{span}\{f_i(x-k) : f_i \in \mathcal{F}\}\right)$$

generated by our 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. Since in general the data does not belong to a shift-invariant space with n generators (n small), the goal is to find—among all possible shift-invariant spaces with n generators—the one that fits the data optimally.

Accordingly, in this section we do not assume that $\mathcal{F} = \{f_1, \ldots, f_m\}$ belongs to a space V with exactly n generators.

Let us consider the function

$$r(\omega) = \text{rank } G_{\mathcal{F}}(\omega),$$

where $G_{\mathcal{F}}(\omega)$ is the Grammian matrix at ω. Let $r_{\min}$ and $r_{\max}$ denote the minimum and the maximum value that $r(\omega)$ can attain in $[0,1]^d$, i.e.,

$$r_{\min} = \min_{\omega \in [0,1]^d} r(\omega) \quad \text{and} \quad r_{\max} = \max_{\omega \in [0,1]^d} r(\omega).$$

Clearly, if $r_{\max}$ is already small, the problem is not interesting. So we will assume that $r_{\min} \geq n$, where n is the number of generators for the space V that we are seeking to model the observed data $\mathcal{F}$. This hypothesis is not strictly necessary for our results, but we will impose it for simplicity.

Consider as before the class $\mathcal{V}$ of all the shift-invariant spaces that are generated by some set of n generators Φ with the property that $\{\phi_i(\cdot - k) : k \in \mathbb{Z}^d,\ i = 1, \ldots, n\}$ is an orthogonal basis for $V(\Phi)$. Note that the assumption of orthogonality does not change the class $\mathcal{V}$ considered in the introduction. Let $w = (w_1, \ldots, w_m)$ be a vector of weights, (i.e., $w_i \in \mathbb{R}$, $w_i > 0$).

Our goal is, given $\mathcal{F}$, to find a space $V \in \mathcal{V}$ such that V minimizes the least square error

$$E(\mathcal{F}, w, n) = \sum_{i=1}^{m} w_i \, \|f_i - P_V f_i\|^2,$$

where P_V is the orthogonal projection onto V. This problem can be viewed as a nonlinear infinite-dimensional constrained minimization problem. It is remarkable that it has a constructive solution, as shown in [2]. This problem may also be viewed in the framework of the recent learning theory developed in [7] and estimates of "model fit" in terms of noise and approximation space may be derived (see the next section).

The first question that arises is if such a space exists at all. In the case that it exists, in order to be useful for applications, it will be important to have a way to construct the generators of the space and to estimate the error $E(\mathcal{F}, w, n)$.

Surprisingly, in [2] the following theorem is proved.

Theorem 14.6. *With the previous notation, let n be given, assume that $n \le r_{\min}$, and let w be a vector of weights, $w = (w_1, \dots, w_m)$. Then there exists a space $V \in \mathcal{V}$ such that*

$$\sum_{i=1}^{m} w_i \, \|f_i - P_V f_i\|^2 \le \sum_{i=1}^{m} w_i \, \|f_i - P_{V'} f_i\|^2, \quad \forall V' \in \mathcal{V}. \tag{14.6}$$

Proof (Sketch). The proof is quite technical and therefore not suitable for this chapter (see [2]); however, it is constructive. We will skip the details and try to give an idea of the construction of the space V.

We consider the space $\mathcal{S}(\mathcal{F})$ and look at the Grammian matrix $G_{\mathcal{F}}$. Since $r_{\min} \ge n$, we always have at least n non-zero eigenvalues. For $i = 1, \dots, n$, consider $\hat{g}_i(\omega) = v_1^i \hat{f}_1 + \cdots + v_m^i \hat{f}_m(\omega)$ where $v^i \in \mathbb{C}^m$ are some choice of eigenvectors associated to the n largest eigenvalues of $G_{\mathcal{F}}(\omega)$. If this choice can be made in such a way that the resulting functions are linearly independent functions in $\mathcal{S}(\mathcal{F})$, then the space generated by these n functions will be the space V we are looking for.

Note that it is not immediate to see that the functions obtained in this way belong to L^2 (or are even measurable functions!). However, after solving this technical part ([2]), one sees that if $r_{\min}$ is greater than or equal to n, we can *always* solve Problem II. □

We will call a space $V \in \mathcal{V}$ satisfying (14.6) an *optimal space* (for the data $\mathcal{F}$). Moreover, it can be seen that the space V is (under minor assumptions) unique.

In view of the preceding construction, we can now state two consequences of the previous theorem that are relevant for this chapter.

Theorem 14.7. *Let $V \in \mathcal{V}$ be an optimal space. Then $V \subset \mathcal{S}(\mathcal{F})$.*

This shows that every optimal space should be contained in the space $\mathcal{S}(\mathcal{F})$ spanned by the data.

Further, we have the following estimate for the error.

Theorem 14.8. *Let again $V \in \mathcal{V}$ be an optimal space, $w = (w_1, \dots, w_m)$ a vector of weights, and $\lambda_1(\omega) \geq \lambda_2(\omega) \geq \cdots \geq \lambda_{r_{\max}}(\omega)$ the eigenvalues of $G_{\mathcal{F}}$ at ω. Then*

$$E(\mathcal{F}, w, n) = \sum_{i=1}^{m} w_i \, \|f_i - P_V f_i\|^2 = \sum_{i=n+1}^{r_{\max}} w_i \int_{[0,1]^d} \lambda_i(\omega) d\omega.$$

Remark 14.9. Obviously, if $n = m$ then the error between the model and the observation is null. However, by plotting the error in Theorem 14.8 in terms of the number of generators, an optimal number n may be derived if the behavior of the error in terms of n shows a horizontal asymptote.

14.5 Problem II as a Learning Problem

Problem II has an interpretation as a learning problem as defined in [7].

Consider a class of signals or images (e.g., electroencephalograms or MRI images). This class of signals belongs to some unknown space that we can assume to be a shift-invariant space $\mathcal{T} \subset L^2(\mathbb{R}^d)$. The space $\mathcal{T}$ (*the target space*) is often very large. For processing, analysis, and manipulation of the data it is necessary to restrict the model to a smaller class of spaces with enough structure. For example, shift-invariant spaces that can be generated by a Riesz basis are appropriate, and are often used in many signal processing applications.

Therefore, we fix a positive integer n, and consider the class $\mathcal{V}$ (*the hypothesis class*) as before. We want to learn about the space $\mathcal{T}$ from some sample elements. Assume that we have m sample signals, say $\mathcal{F} = \{f_1, \dots, f_m\}$ (*the training set*). Using Theorem 14.6, we see that from our data set $\mathcal{F}$ we can obtain some space $V_{\mathcal{F}} \in \mathcal{V}$ that best fits our data.

However, a realistic assumption should consider that our samples are noisy. Therefore, they may not belong to the space $\mathcal{T}$. This means that the space $V_{\mathcal{F}}$ will, in general, be different from the space $V_{\widetilde{\mathcal{F}}}$ that we would have found from signals that are not corrupted by noise.

The noisy data introduces an error. This error can be quantified using some distance between the subspaces $V_{\mathcal{F}}$ and $V_{\widetilde{\mathcal{F}}}$. (We can, for example, consider the distance between the orthogonal projections in some operator norm.) This error is usually called the *sample error* in learning theory.

There is another error (*the approximation error*) due to the fact that our family of spaces is constrained to have only n generators.

Estimation of these errors in terms of the number of samples and the number of generators is an ongoing research by the authors.

Acknowledgments

The authors were partially supported by grants from the NSF (DMS-0103104, DMS-0139740), CIES (CIES-87426189), the University of Buenos Aires (UBA-CyT X058 and 108), and the ANPCyT (BID 1201/OC-AR PICT 15033).

References

1. A. Aldroubi, Oblique projections in atomic spaces, *Proc. Amer. Math. Soc.*, **124** (1996), pp. 2051–2060.
2. A. Aldroubi, C. A. Cabrelli, D. Hardin, and U. M. Molter, Optimal shift-invariant spaces and their Parseval frame generators, preprint (2006).
3. A. Aldroubi, C. A. Cabrelli, D. Hardin, U. Molter, and A. Rodado, Determining sets of shift invariant spaces, in: *Wavelets and their Applications* (Chennai, January 2002), M. Krishna, R. Radha, and S. Thangavelu, eds., Allied Publishers, New Delhi (2003), pp. 1–8.
4. A. Aldroubi and K. Gröchenig, Non-uniform sampling and reconstruction in shift-invariant spaces, *SIAM Review*, **43** (2001), pp. 585–620.
5. J. J. Benedetto and P. J. S. G. Ferreira, eds., *Modern Sampling Theory: Mathematics and Applications*, Birkhäuser, Boston, 2001.
6. J. J. Benedetto and A. I. Zayed, eds., *Sampling, Wavelets, and Tomography*, Birkhäuser, Boston, 2004.
7. F. Cucker and S. Smale, On the mathematical foundations of learning, *Bull. Amer. Math. Soc. (N.S.)*, **39** (2002), pp. 1–49.
8. C. de Boor, R. De Vore, and A. Ron, The structure of finitely generated shift-invariant subspaces of $L_2(R^d)$, *J. Funct. Anal.*, **119** (1994), pp. 37–78.
9. T. N. T. Goodman, S. L. Lee, and W. S. Tang, Wavelets in wandering subspaces, *Trans. Amer. Math. Soc.*, **338** (1993), pp. 639–654.