A Theory for Sampling Signals from a Union of Subspaces

Yue M. Lu and Minh N. Do

***Abstract*—One of the fundamental assumptions in traditional sampling theorems is that the signals to be sampled come from a single vector space (e.g. bandlimited functions). However, in many cases of practical interest the sampled signals actually live in a union of subspaces. Examples include piecewise polynomials, sparse representations, nonuniform splines, signals with unknown spectral support, overlapping echoes with unknown delay and amplitude, and so on. For these signals, traditional sampling schemes based on the single subspace assumption can be either inapplicable or highly inefficient. In this paper, we study a general sampling framework where sampled signals come from a known union of subspaces and the sampling operator is linear. Geometrically, the sampling operator can be viewed as projecting sampled signals into a lower dimensional space, while still preserving all the information. We derive necessary and sufficient conditions for invertible and stable sampling operators in this framework and show that these conditions are applicable in many cases. Furthermore, we find the minimum sampling requirements for several classes of signals, which indicates the power of the framework. The results in this paper can serve as a guideline for designing new algorithms for many applications in signal processing and inverse problems.**

***Index Terms*—sampling, signal representations, union of subspaces, linear operators, projections, invertible, stable, shift-invariant spaces.**

I. Introduction

Sampling is a corner stone of signal processing because it allows real-life signals in the continuous-domain to be acquired, represented, and processed in the discrete-domain (e.g. by computers). One of the fundamental assumptions in traditional sampling theorems [1]–[4] is that the signals to be sampled come from a single vector space (e.g. bandlimited functions). For example, the classical Kotelnikov-Shannon-Whittaker sampling theorem can be presented as follow [3]. Denote $\text{sinc}_T(t) = \frac{\sin(\pi t/T)}{\pi t/T}$; then $\{\text{sinc}_T(t - nT)\}_{n\in\mathbb{Z}}$ is an orthogonal basis for the space $\mathcal{S}_T^{(BL)}$ of bandlimited functions whose Fourier transforms are supported within $[-\pi/T, \pi/T]$. Specifically, for all $x \in \mathcal{S}_T^{(BL)}$, we have

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT)\,\text{sinc}_T(t - nT), \tag{1}$$

and

$$x(nT) = \left(x * \frac{\text{sinc}_T}{T}\right)(nT). \tag{2}$$

$$= \langle x, \frac{\text{sinc}_T(\cdot - nT)}{T}\rangle_{L^2(\mathbb{R})} \tag{3}$$

Equation (1) shows that any bandlimited signal $x(t) \in \mathcal{S}_T^{(BL)}$ is uniquely represented by its samples $\{x(nT)\}_{n\in\mathbb{Z}}$ and provides a way to reconstruct $x(t)$ from these samples. Equation (2) corresponds to the practice of passing the signal $x(t)$ through an anti-aliasing filter before taking samples.

From this viewpoint, the Kotelnikov-Shannon-Whittaker sampling theorem has been generalized by considering other signal spaces $\mathcal{S}$ and other sampling functions (see, for example, [3]–[11] and the references therein). In all of these previous studies, the signals to be sampled are assumed to come from a single vector space. However, as we will illustrate with the following examples, in many situations the signals of interest actually live in a *union of subspaces*.

Example 1 (Stream of Diracs): The stream of Diracs is the basic signal model for the recent sampling framework for signals with *finite rate of innovation* [12]–[14]. As illustrated in Figure 1(a), a stream of K Diracs has the form $x(t) = \sum_{k=1}^{K} c_k\,\delta(t - t_k)$, where $\{t_k\}_{k=1}^{K}$ are unknown locations and $\{c_k\}_{k=1}^{K}$ are unknown weights. We see that once the K locations are fixed, the signals live in a K dimensional subspace. Thus, the set of all streams of K Diracs is a union of K-dimensional subspaces.

Example 2 (Piecewise polynomials): Many transient signals in practice can be modeled by piecewise polynomials [see Figure 1(b)]. Let $\mathcal{X}_K^{(PP)}$ denote the set of all signals consisting of K pieces of polynomials supported on $[0, 1]$, where each piece is of degree less than d. We cannot ensure the sum of any two signals in $\mathcal{X}_K^{(PP)}$ still has only K pieces of polynomials, and thus $\mathcal{X}_K^{(PP)}$ is *not* a vector subspace. However, it is easy to verify that we do have a subspace once we fix the locations of the discontinuities. Therefore, $\mathcal{X}_K^{(PP)}$ is the union of the subspaces corresponding to all possible discontinuity locations.

Example 3 (2-D Piecewise Polynomials): Consider 2-D piecewise polynomials of K pieces supported on $[0, 1]^2$, as shown in Figure 1(c). More specifically, each piece is a bivariate polynomial of degree less than d. This kind of signal can be seen as a "cartoon" model for natural images, since natural scenes are often made up from several objects of smooth surfaces with smooth boundaries. Again, it is easy to see that once we fix the boundaries, the signals lie on a subspace of dimension Kd^2. With all possible boundaries, 2-D piecewise polynomials live in a union of subspaces.

Y. M. Lu was with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana IL 61801. He is now with the Audio-Visual Communications Laboratory, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland (e-mail: yue.lu@epfl.ch).

M. N. Do is with the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute, University of Illinois, Urbana IL 61801 (e-mail: minhdo@uiuc.edu).

This work was supported by the U.S. National Science Foundation under Grants CCF-0237633 and CCF-0635234. The material in this paper was presented in part at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Canada, May 2004.

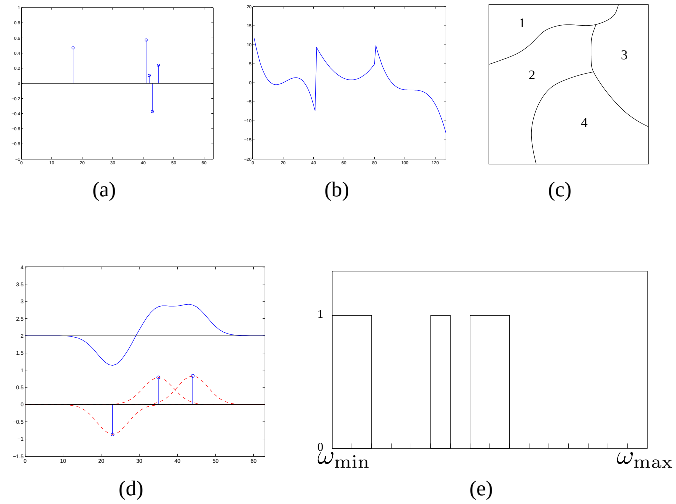

Fig. 1. Several examples in which the signals of interest come from a union of subspaces. (a) A stream of Diracs with unknown locations and weights. (b) A 1-D piecewise polynomial signal with unknown discontinuity locations. (c) A 2-D piecewise polynomial with unknown boundaries. (d) An overlapping echo (shown in solid lines) that is a linear combination of three pulses (shown in dashed lines) with unknown delays and amplitudes. (e) A multiband signal in frequency with unknown spectral support that only occupies a known fraction of the spectral band $[\omega_{\min}, \omega_{\max}]$.

Example 4 (Sparse representation): Sparse representation lies at the heart of modern signal compression and denoising [15], [16]. In these applications, the final output signal is a K-term representation using a fixed basis or dictionary $\{\phi_k\}_{k=1}^{\infty}$ (e.g. a Fourier or wavelets basis), written as

$$\hat{f}_K = \sum_{k \in I} c_k \phi_k, \tag{4}$$

where I is an index set of K selected basis functions or atoms. Clearly, the set of all signals that can be represented by K terms from a given basis or dictionary constitutes a union of subspaces, with each subspace indexed by a set I.

Example 5 (Overlapping echoes): Consider overlapping echoes with unknown delay and amplitude [17], [18]. Illustrated in Figure 1(d), these signals have the form $x(t) = \sum_{k=1}^{K} c_k \, \phi(t - t_k)$, where the pulse shape $\phi(t)$ is known; while the delays $\{t_k\}_{k=1}^{K}$ and amplitudes $\{c_k\}_{k=1}^{K}$ are unknown. Clearly, the set of all possible echoes constitutes a union of subspaces, each of which corresponds to a set of delays $\{t_k\}_{k=1}^{K}$. Signals of this type appear in many applications such as geophysics, radar, sonar, and communications. In these applications, from a limited number of samples of the echo signals, one wishes to find out the delays and amplitudes.

Example 6 (Signals with unknown spectral support): Consider the class of continuous-time signals whose Fourier transforms only occupy a known fraction – but at *unknown* locations – on a spectral band $[\omega_{\min}, \omega_{\max}]$ [see Figure 1(e)]. The sampling problem for this class of signals has been studied in [19]–[22]. Again, for a fixed set of spectral support, these signals live in a subspace. With all possible spectral supports, the signal class can be characterized by a union of subspaces.

For signals given in the above examples, traditional sampling schemes based on the single subspace assumption can be either inapplicable or highly inefficient. In principle, we can always extend the class of signals from a union of subspaces to the smallest linear vector space that contains it, and carry out sampling on that space. However, this strategy is often inefficient since it ignores the additional prior information about the signals.

For instance, the smallest linear space containing the K-term sparse signals in Example 4 is the space spanned by the entire dictionary $\{\phi_k\}_{k=1}^{\infty}$. In contrast, from the definition in (4), we should be able to completely determine these sparse signals by using only $2K$ numbers, with K of them specifying the index set I and the rest recording the coefficients $\{c_k\}$. Similarly, for signals with unknown spectral support in Example 6, the smallest linear space containing them is the space of bandlimited functions supported on the entire spectral band $[\omega_{\min}, \omega_{\max}]$, whose Nyquist rate is based on the whole bandwidth $\omega_{\max} - \omega_{\min}$. However, the work in [20] shows that, by exploiting the additional prior knowledge about the signal spectrum, it is possible to achieve a sampling rate well below the above Nyquist rate.

Thus the examples above motivate us to fundamentally extend the traditional sampling theorems by considering signals from a *union of subspaces* instead of a *single space*. Our proposed sampling framework has close ties to the recent work on sampling signals with *finite rate of innovation* [12]–[14], which demonstrates that several classes of non-bandlimited signals can be uniformly sampled and perfectly reconstructed. In a general sense, signals with finite rate of innovation have a known degree of freedom (i.e. innovation), but the locations of the innovation are unknown (see Examples 1–3). Therefore, these types of signals can often be effectively characterized by unions of subspaces.

Another related work is the recent breakthrough in mathematics under the name *compressed sensing* or *compressive sampling* [23]–[25], which shows that sparse or compressible finite length discrete signals can be recovered from small number of linear, non-adaptive, and random measurements. The number of required measurements has the same order of magnitude as the number of non-zero or “significant” coefficients in the input signal, which is typically much smaller than the length of the signal. The literature on compressed sensing so far only handles finite-dimensional signals.

Our proposed sampling framework with union of subspaces provides a generalized and unified framework for finite rate of innovation sampling, compressed sensing/compressive sampling, and spectrum-blind sampling, in which new results and derivations are discussed. Moreover, the proposed framework provides a geometrical approach to finite rate of innovation sampling and suggest a path for extending the current compressed sensing theory to infinite-dimensional settings and continuous-domain signals.

In Section II, we formulate the problem of sampling signals from a union of subspaces and provide a geometrical interpretation. Section III presents general conditions for invertible and stable sampling operators. We then study the sampling problem in two concrete settings: in Section IV, we consider unions of finite-dimensional subspaces; and in Section V, we consider unions of infinite-dimensional shift-invariant subspaces. Section VI concludes the paper with some outlook.

II. Problem Formulation

A. Framework: Linear Sampling from a Union of Subspaces

The examples given in the previous section lead us to consider the following abstract definition for many classes of signals of interest.

First, let $\mathcal{H}$ be an ambient Hilbert space[1] in which our signals live. Some concrete cases of $\mathcal{H}$ include: in Examples 2 and 3 for piecewise polynomials, $\mathcal{H} = L^2(\mathcal{D})$, where $\mathcal{D} = [0,1]$ (or $[0,1]^2$ for 2-D) is the domain of spatial support; for overlapping echoes introduced in Example 5, if the pulse shape $\phi(t)$ is square-integrable, we can choose $\mathcal{H} = L^2(\mathbb{R})$; for signals with unknown spectral support in Example 6, $\mathcal{H}$ can be the space of all functions bandlimited to the largest possible spectral span $[\omega_{\min}, \omega_{\max}]$.

Definition 1 (Union of subspaces): The signals of interest live in a fixed *union of subspaces* that is defined as

$$\mathcal{X} = \bigcup_{\gamma \in \Gamma} \mathcal{S}_\gamma, \tag{5}$$

where $\mathcal{S}_\gamma$ are subspaces of $\mathcal{H}$ and Γ is an index set. In other words, a signal $x \in \mathcal{X}$ if and only if there is some $\gamma_0 \in \Gamma$ such that $x \in \mathcal{S}_{\gamma_0}$.

We consider a general sampling framework in which the input signal $x \in \mathcal{X}$ is sampled via a bounded *linear mapping* A into a sequence of numbers $\{(Ax)_n\}_{n\in\Lambda}$. We refer to $\{(Ax)_n\}_{n\in\Lambda}$ as *samples* of x via the sampling operator A. From the Riesz representation theorem [26], there exists a unique set of vectors $\Psi = \{\psi_n\}_{n\in\Lambda}$ in $\mathcal{H}$ for any such linear mapping A so that

$$(Ax)_n = \langle x, \psi_n \rangle_{\mathcal{H}} \tag{6}$$

and thus

$$A : x \mapsto \{\langle x, \psi_n \rangle\}_{n\in\Lambda}\,. \tag{7}$$

Thus, any bounded linear sampling operator A is uniquely specified by the set of *sampling vectors* $\Psi = \{\psi_n\}_{n\in\Lambda}$. In the form (6), ψ_n resembles the point spreading function of the nth measurement device. A case of particular interest is when the sampling vectors are shifted versions of a common kernel function ψ; for example, $\psi_n(t) = \psi(t - nT)$ and $\mathcal{H} = L^2(\mathbb{R})$. In that case, the sampling procedure given in (7) can be efficiently implemented by filtering followed by uniform pointwise sampling, which is similar to (2) as in classical sampling. In various Fourier imaging systems, including magnetic resonance imaging (MRI), $\{\psi_n\}_{n\in\Lambda}$ are complex exponential signals on a compact support. In computed tomography, inner products with $\{\psi_n\}_{n\in\Lambda}$ represent linear integrals.

[1]We could consider a more general framework where the ambient space is a vector space. However, we will restrict to the Hilbert-space setting as it provides induced norms and is more familiar in the signal processing community.

Given a class of signals defined as a union of subspaces, it is attractive to find a fixed representation as in (7) for them. The natural questions to pursue are the following.

1) When is each object $x \in \mathcal{X}$ *uniquely represented by its sampling data* $\{\langle x, \psi_n \rangle\}_{n\in\Lambda}$?
2) What is the *minimum sampling requirement* for a signal class $\mathcal{X}$?
3) What are the *optimal sampling functions* $\{\psi_n\}_{n\in\Lambda}$?
4) What are *efficient algorithms to reconstruct a signal* $x \in \mathcal{X}$ from its sampling data $\{\langle x, \psi_n \rangle\}_{n\in\Lambda}$?
5) How *stable is the reconstruction* in the presence of noise and model mismatch?

Note that if $\mathcal{X}$ is a single vector space $\mathcal{S}$ then *frame theory* (see, for example, [27, pp. 53-63]) precisely addresses these questions. In particular, one can reconstruct any $x \in \mathcal{S}$ in a numerically stable way from its sampling data $Ax = \{\langle x, \psi_n \rangle\}_{n\in\Lambda}$ whenever $\{\psi_n\}_{n\in\Lambda}$ is a *frame* of $\mathcal{S}$.

In this paper, we study and answer the first two questions outlined above, which involve the feasibility and fundamental performance bounds of the proposed sampling framework. It is our hope that the results from this work, including the geometrical viewpoint, stable sampling bounds, and minimum sampling requirement as discussed below, can provide useful insight and guidelines for the solutions of the remaining questions in future work.

B. Geometrical Viewpoint

In the Hilbert space $\mathcal{H}$, knowing $\{\langle x, \psi_n \rangle\}_{n\in\Lambda}$ is equivalent, up to a linear transformation, to knowing the projection $\mathbf{P}_{\mathcal{S}}\, x$ of x onto the subspace $\mathcal{S} = \overline{\text{span}}\,\{\psi_n\}_{n\in\Lambda}$. We call $\mathcal{S}$ a *representation subspace*. Clearly, $\{\psi_n\}_{n\in\Lambda}$ provides an *invertible* sampling operator for $\mathcal{X}$ if and only if there is a one-to-one mapping between $\mathbf{P}_{\mathcal{S}}\mathcal{X}$ and $\mathcal{X}$.

Figure 2 illustrates a simple case, where the signal space $\mathcal{H}$ is $\mathbb{R}^3$. The set of signals of interest $\mathcal{X} = \bigcup_{i=1}^3 \mathcal{S}_i$ is the union of three one-dimensional subspaces (three lines going through the origin). As shown in Figure 2, we project $\mathcal{X}$ down to a certain subspace (a plane) $\mathcal{S}$ and obtain $\mathbf{P}_{\mathcal{S}}\mathcal{X} = \bigcup_{i=1}^3 \mathbf{P}_{\mathcal{S}}\mathcal{S}_i$. We can see that there is an invertible mapping between $\mathcal{X}$ and $\mathbf{P}_{\mathcal{S}}\mathcal{X}$ as long as no two subspaces in $\{\mathcal{S}_i\}_{i=1}^3$ are projected onto the same line in $\mathcal{S}$ [see Figure 2(a)]. In this case, no information is lost and we have a more compact representation of the original signals. Thus geometrically, we can think of the proposed linear sampling as projecting the set of signals onto a lower dimensional representation space, while still preserving its information.

The first problem is to study the lower bound of the dimension of invertible representation subspaces, which is related to the minimum sampling requirement. In the case of Figure 2, the lower bound is 2 (i.e. a plane), because there would always be information loss if we projected $\mathcal{X}$ onto any single line.

We notice that the representation subspaces that provide invertible or one-to-one mapping are not unique. Although in theory any of them can be used, they are very different in practice. For some representation subspaces, the projected lines are so close to each other [e.g. consider a perturbation

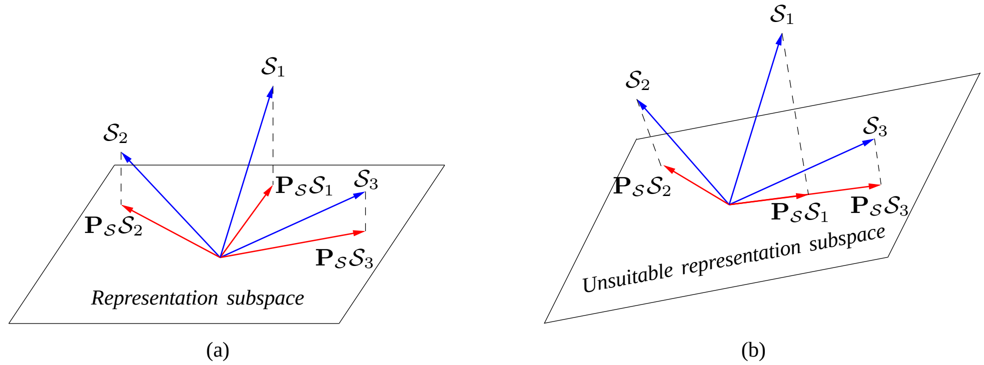

Fig. 2. An example of a union of subspaces $\mathcal{X} = \bigcup_{i=1}^{3} \mathcal{S}_i$ and its projections onto lower dimensional representation subspaces. (a) A case of invertible and stable representation. (b) A case of noninvertible representation. Also, a representation subspace close to this one would lead to unstable representation.

of Figure 2(b)] that sampling becomes very sensitive to noise and numerical error. So there is an issue in how to choose the "optimal" representation subspace, or equivalently the "optimal" sampling vectors.

In the following sections, we will formulate and study the above geometrical intuitions in a rigorous and quantitative way.

III. Conditions for Sampling Operators

A. Definitions

We now go back to the general sampling framework defined in Section II-A, where the set of signals of interest $\mathcal{X}$ is given in (5) and the sampling operator A is given in (7). First, we want to know whether each signal $x \in \mathcal{X}$ is uniquely represented by its sampling data Ax.

Definition 2 (Invertible sampling): We call A an *invertible* sampling operator for $\mathcal{X}$ if each $x \in \mathcal{X}$ is uniquely determined by its sampling data Ax; that means for every x_1 and x_2 in $\mathcal{X}$,

$$Ax_1 = Ax_2 \quad \text{implies} \quad x_1 = x_2. \tag{8}$$

In other words, A is a *one-to-one mapping* between $\mathcal{X}$ and $A\mathcal{X}$.

The invertible (or one-to-one) condition allows us to uniquely identify each $x \in \mathcal{X}$ from Ax. However, in practice, stronger requirements are needed: we want to be able to reconstruct $x \in \mathcal{X}$ in a *numerically stable* way from Ax. To guarantee such an algorithm exists, we need to ensure that if Ax is "close" to Ay then x is "close" to y as well. Furthermore, we want that a small change in the signal x only produces a small change in its sampling data Ax. These requirements motivate the next condition on the sampling operator.

Definition 3 (Stable sampling): We call A a *stable* sampling operator for $\mathcal{X}$ if there exist constants $0 < \alpha \leq \beta < \infty$ such that for every $x_1 \in \mathcal{X}, x_2 \in \mathcal{X}$,

$$\alpha \|x_1 - x_2\|_{\mathcal{H}}^2 \leq \|Ax_1 - Ax_2\|_{l_2(\Lambda)}^2 \leq \beta \|x_1 - x_2\|_{\mathcal{H}}^2. \tag{9}$$

We call α and β *stability bounds* and the tightest ratio $\kappa = \beta/\alpha$ provides a measure of the stability of the sampling operator.

Note that we use the l_2 norm for $Ax_1 - Ax_2$ since it is a sequence of numbers. We can see that stable sampling implies invertible sampling, whereas the reverse is not true.

The stable sampling condition in (9) is defined in terms of the squared norm (i.e. energy) of the signals and their sample values. However, when we work in $\mathcal{H} = L^2(\mathbb{R})$ and thus all the signals are functions of a (time) variable, it is often desirable to consider a more stringent pointwise stability as discussed in [4]. This additional requirement is due to the fact that two signals $x_1(t)$ and $x_2(t)$ can be close in the L^2 sense, but still differ markedly in their pointwise values within some localized regions.

To bypass this problem, we can adopt the treatment in [4] by restricting the ambient space $\mathcal{H}$ to a (reproducing kernel) subspace of $L^2(\mathbb{R})$ with the following property:

$$\sup_{t \in \mathbb{R}} |x(t)|^2 \leq \alpha' \int_{\mathbb{R}} |x(u)|^2 \, du = \alpha' \, \|x\|_{L^2(\mathbb{R})}^2, \tag{10}$$

for all $x(t) \in \mathcal{H}$, where $0 < \alpha' < \infty$ is some constant. Examples of subspaces having the above property include the space of bandlimited functions, and shift-invariant spaces with the generating function satisfying some mild conditions [4]. By linking (9) and (10), we get

$$\sup_{t \in \mathbb{R}} |x_1(t) - x_2(t)|^2 \leq \alpha'' \, \|Ax_1 - Ax_2\|_{l_2(\Lambda)}^2,$$

for some constant $0 < \alpha'' < \infty$. In this case, the proposed stable sampling condition in (9) implies pointwise stability as well.

B. Key Observation

The main difficulty in dealing with unions of subspaces is that, in the last two definitions, x_1 and x_2 can be from two different subspaces. In other words, the proposed unique and stable sampling conditions are defined on a *nonlinear* set. Consequently, we cannot directly apply various well-known *linear* results in matrix and operator theories to study the proposed sampling conditions. To overcome this problem, we introduce the following subspaces:

$$\begin{aligned} \widetilde{\mathcal{S}}_{\gamma,\theta} &\stackrel{\text{def}}{=} \mathcal{S}_\gamma + \mathcal{S}_\theta \\ &= \{ y : y = x_1 + x_2, \text{ where } x_1 \in \mathcal{S}_\gamma, \; x_2 \in \mathcal{S}_\theta \}. \end{aligned} \tag{11}$$

Typically, $\widetilde{\mathcal{S}}_{\gamma,\theta}$ has simple interpretations. For instance: in Example 1 with streams of K Diracs, $\widetilde{\mathcal{S}}_{\gamma,\theta}$ is a subspace of at most $2K$ Diracs with fixed location; in Example 2 of piecewise polynomials, $\widetilde{\mathcal{S}}_{\gamma,\theta}$ is a subspace of piecewise polynomials with at most $2K-1$ pieces; and so on. It is easy to see that the set

$$\begin{aligned}\widetilde{\mathcal{X}} &\stackrel{\text{def}}{=} \bigcup_{(\gamma,\theta)\in\Gamma\times\Gamma} \widetilde{\mathcal{S}}_{\gamma,\theta} \\ &= \{y : y = x_1 - x_2, \text{ where } x_1 \in \mathcal{X}, x_2 \in \mathcal{X}\}, \end{aligned} \tag{12}$$

consists of all *secant* vectors of the set $\mathcal{X}$, which play a fundamental role in the study of dimensionality reduction [28].

The next two propositions map the invertible and stable conditions on the *union* of subspaces $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_\gamma$ to those for *single* subspaces.

Proposition 1: A linear sampling operator A is *invertible* for $\mathcal{X}$ *if and only if* A is invertible for every $\widetilde{\mathcal{S}}_{\gamma,\theta}$ with $(\gamma,\theta) \in \Gamma\times\Gamma$.

Proof: Consider the "if" part, that is, assume that A is one-to-one on every $\widetilde{\mathcal{S}}_{\gamma,\theta}$, $(\gamma,\theta)\in\Gamma\times\Gamma$. Let x_1, x_2 be vectors in $\mathcal{X}$ such that $Ax_1 = Ax_2$. From the definition of $\mathcal{X}$ there exist $\gamma,\theta\in\Gamma$, such that $x_1 \in \mathcal{S}_\gamma, x_2 \in \mathcal{S}_\theta$. Thus $x_1, x_2 \in \widetilde{\mathcal{S}}_{\gamma,\theta}$, and from the one-to-one assumption for $\widetilde{\mathcal{S}}_{\gamma,\theta}$, it follows that $x_1 = x_2$. Hence A is one-to-one on $\mathcal{X}$.

Now consider the "only if" part, that is, assume that A is one-to-one on $\mathcal{X}$. Let y_1, y_2 be vectors in $\widetilde{\mathcal{S}}_{\gamma,\theta}$, $(\gamma,\theta)\in\Gamma\times\Gamma$ such that $Ay_1 = Ay_2$. Denote $z = y_1 - y_2$. Because $\widetilde{\mathcal{S}}_{\gamma,\theta}$ is a subspace, $z \in \widetilde{\mathcal{S}}_{\gamma,\theta}$. From (12), there exist $x_1 \in \mathcal{X}$ and $x_2 \in \mathcal{X}$ such that $z = x_1 - x_2$. Since A is linear, $Ax_1 - Ax_2 = A\,z = Ay_1 - Ay_2 = 0$. It then follows from the one-to-one assumption for $\mathcal{X}$ that $x_1 = x_2$. This implies $z = 0$, or equivalently, $y_1 = y_2$. Therefore, A is one-to-one on every $\widetilde{\mathcal{S}}_{\gamma,\theta}$, $(\gamma,\theta)\in\Gamma\times\Gamma$. ∎

Proposition 2: A linear sampling operator A is *stable* for $\mathcal{X}$, with stability bounds α and β, *if and only if*

$$\alpha\|y\|^2_{\mathcal{H}} \le \|Ay\|^2_{l_2(\Lambda)} \le \beta\|y\|^2_{\mathcal{H}}, \tag{13}$$

for every $y \in \widetilde{\mathcal{S}}_{\gamma,\theta}$ and $(\gamma,\theta)\in\Gamma\times\Gamma$.

Proof: Starting from the stable sampling condition of A given in (9) and using (12) we have

$$\begin{aligned} &\alpha\|x_1 - x_2\|^2_{\mathcal{H}} \le \|Ax_1 - Ax_2\|^2_{l_2} \le \beta\|x_1 - x_2\|^2_{\mathcal{H}} \\ \Longleftrightarrow\ &\alpha\|y\|^2_{\mathcal{H}} \le \|Ay\|^2_{l_2} \le \beta\|y\|^2_{\mathcal{H}}, \qquad \text{for every } y \in \widetilde{\mathcal{X}} \\ \Longleftrightarrow\ &\alpha\|y\|^2_{\mathcal{H}} \le \|Ay\|^2_{l_2} \le \beta\|y\|^2_{\mathcal{H}}, \end{aligned}$$

for every $y \in \widetilde{\mathcal{S}}_{\gamma,\theta}$ and $(\gamma,\theta)\in\Gamma\times\Gamma$. ∎

IV. Union of Finite-Dimensional Subspaces

In this section we consider the situation where the subspaces $\mathcal{S}_\gamma$ $(\gamma\in\Gamma)$ in $\mathcal{X}$ are finite-dimensional, although the ambient space $\mathcal{H}$ can be infinite-dimensional and the index set Γ can be infinite.

A. Minimum Sampling Requirement

Using Proposition 1, we immediately obtain the following minimum sampling requirement for union of finite-dimensional subspaces.

Proposition 3: Suppose that $A : x \mapsto \{\langle x, \psi_n\rangle\}_{n=1}^N$ is an *invertible* sampling operator for $\mathcal{X}$. Then

$$N \ge N_{\min} \stackrel{\text{def}}{=} \sup_{(\gamma,\theta)\in\Gamma\times\Gamma} \dim(\widetilde{\mathcal{S}}_{\gamma,\theta}). \tag{14}$$

Proof: Suppose that A is one-to-one on $\mathcal{X}$. From Proposition 1, A is one-to-one on every $\widetilde{\mathcal{S}}_{\gamma,\theta}$, $(\gamma,\theta)\in\Gamma\times\Gamma$. It follows that $\dim(\widetilde{\mathcal{S}}_{\gamma,\theta}) = \dim(A(\widetilde{\mathcal{S}}_{\gamma,\theta}))$. Since the range of A is in an N-dimensional vector space, $\dim(A(\widetilde{\mathcal{S}}_{\gamma,\theta})) \le N$. Therefore, $N \ge \dim(\widetilde{\mathcal{S}}_{\gamma,\theta})$ for every (γ,θ), and, hence, $N \ge N_{\min}$. ∎

Proposition 3 provides a minimum sampling requirement (i.e. the minimum number of samples) for *linear sampling*. It states that with a linear sampling scheme, one needs to obtain at least $N_{\min}$ samples to provide an invertible representation for signals from $\mathcal{X}$.

Consider a simple application of Proposition 3 to Example 1 where $\mathcal{X}$ consists of streams of K Diracs.[2] In this case, $\widetilde{\mathcal{S}}_{\gamma,\theta}$ are subspaces of streams of Diracs with impulses at up to $2K$ fixed locations. Thus the minimum sampling requirement is $N_{\min} = 2K$. This is also equal to the number of free parameters for each signal in $\mathcal{X}$.

The situation becomes more interesting when we consider Example 2, where the signal class $\mathcal{X}$ consists of 1-D piecewise polynomial signals supported on $[0,1]$. Every signal in $\mathcal{X}$ contains at most K polynomial pieces, each of degree less than d. We can see that every signal in $\mathcal{X}$ can be fully specified by $Kd + K - 1$ free parameters, with $K - 1$ parameters used to record the locations of the discontinuities and Kd parameters to specify the coefficients of the K polynomial pieces. But is it sufficient to use only $Kd + K - 1$ linear measurements to fully specify signals from $\mathcal{X}$?

The above question can be answered by applying Proposition 3. First, we can check that $\widetilde{\mathcal{S}}_{\gamma,\theta}$ are subspaces of piecewise polynomials with at most $2K-1$ pieces, each of degree less than d. Thus from (14), the minimum sampling requirement for $\mathcal{X}$ is $N_{\min} = (2K-1)d$. Contrary to what one might expect, $N_{\min}$ is *strictly* greater than the number of free parameters $Kd + K - 1$ when $d > 1$. Thus, as a novel application of our minimum sampling bound, we have shown that the sampling algorithm proposed in [12] for piecewise polynomials, which effectively converts the input signal into a stream of Diracs by repeated differentiation, indeed achieves the minimum sampling requirement $N_{\min} = (2K-1)d$.

B. Invertible Conditions on Sampling Vectors

Recall that a linear sampling operator is specified by a set of sampling vectors $\Psi = \{\psi_n\}_{n\in\Lambda}$ as defined in (7). We now study the invertible and stable sampling conditions on Ψ. Let $\Phi = \{\phi_k\}_{k=1}^K$ be a basis for a finite dimensional subspace $\mathcal{S}$.

[2] Technically, streams of Diracs do not belong to a Hilbert space as required in our framework; rather, these generalized functions should be treated as linear functionals on the space of continuous test functions. However, we can verify that Propositions 1, 3, and 4 hold without change under this more general setup. The only difference is that, instead of representing inner products in Hilbert spaces, the notation $\langle x, \psi_n\rangle$ should now be understood as the pairing between the linear functional x (such as Diracs) with its argument ψ_n.

Then each $x \in \mathcal{S}$ has the basis expansion

$$x = \sum_{k=1}^{K} c_k \phi_k. \tag{15}$$

It follows that

$$(Ax)_n = \langle x, \psi_n \rangle = \sum_{k=1}^{K} c_k \langle \phi_k, \psi_n \rangle.$$

Thus, we can express Ax via a matrix-vector multiplication

$$Ax = \boldsymbol{G}_{\Phi,\Psi} \boldsymbol{c}, \tag{16}$$

where $\boldsymbol{G}_{\Phi,\Psi}$ is the (generalized) Gram matrix between the sets of vectors $\Phi = \{\phi_k\}_{k=1}^K$ and $\Psi = \{\psi_n\}_{n=1}^N$:

$$\boldsymbol{G}_{\Phi,\Psi} \stackrel{\text{def}}{=} \begin{pmatrix} \langle \phi_1, \psi_1 \rangle & \langle \phi_2, \psi_1 \rangle & \cdots & \langle \phi_K, \psi_1 \rangle \\ \langle \phi_1, \psi_2 \rangle & \langle \phi_2, \psi_2 \rangle & \cdots & \langle \phi_K, \psi_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \phi_1, \psi_N \rangle & \langle \phi_2, \psi_N \rangle & \cdots & \langle \phi_K, \psi_N \rangle \end{pmatrix}, \tag{17}$$

and $\boldsymbol{c} = (c_1, \ldots, c_K)^T$ is the column vector of coefficients in the basis expansion of x. Similarly, if $\Phi_{\gamma,\theta} = \left\{ \phi_k^{(\gamma,\theta)} \right\}_{k=1}^{K_{\gamma,\theta}}$ is a basis for $\widetilde{\mathcal{S}}_{\gamma,\theta}$, then for $y \in \widetilde{\mathcal{S}}_{\gamma,\theta}$, we can express Ay via a matrix-vector multiplication as in (16) with the Gram matrix $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}$. Hence, the invertible sampling condition of A in Proposition 1 is translated into the (left) invertible condition on Gram matrices $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}$, as follows.

Proposition 4: Let $\Psi = \{\psi_n\}_{n=1}^N$ be a set of sampling vectors and $\Phi_{\gamma,\theta} = \left\{ \phi_k^{(\gamma,\theta)} \right\}_{k=1}^{K_{\gamma,\theta}}$ be a basis for $\widetilde{\mathcal{S}}_{\gamma,\theta}$. Then Ψ provides an invertible sampling operator for $\mathcal{X}$ *if and only if* $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}$ has full column rank for every $(\gamma, \theta) \in \Gamma \times \Gamma$.

Furthermore, if we suppose $\Phi_{\gamma,\theta} = \left\{ \phi_k^{(\gamma,\theta)} \right\}_{k=1}^{K_{\gamma,\theta}}$ is an *orthonormal* basis for $\widetilde{\mathcal{S}}_{\gamma,\theta}$, then $\|y\|_{\mathcal{H}} = \|\boldsymbol{c}\|_2$. From matrix theory [29], we know that

$$\sigma_{\min}^2(\boldsymbol{G}) \, \|\boldsymbol{c}\|_2^2 \leq \|\boldsymbol{G}\, \boldsymbol{c}\|_2^2 \leq \sigma_{\max}^2(\boldsymbol{G}) \, \|\boldsymbol{c}\|_2^2, \tag{18}$$

for every $\boldsymbol{c}$, where $\sigma_{\min}(\boldsymbol{G})$ and $\sigma_{\max}(\boldsymbol{G})$ are the smallest and largest singular values of $\boldsymbol{G}$, respectively. Moreover, $\sigma_{\min}^2(\boldsymbol{G})$ and $\sigma_{\max}^2(\boldsymbol{G})$ provide the tightest bounds for the inequalities of the type in (18). Hence, the stable sampling condition of A in Proposition 2 is translated into the classical conditioning requirements on Gram matrices $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}$.

Proposition 5: Let $\Psi = \{\psi_n\}_{n=1}^N$ be a set of sampling vectors and $\Phi_{\gamma,\theta} = \left\{ \phi_k^{(\gamma,\theta)} \right\}_{k=1}^{K_{\gamma,\theta}}$ be an orthonormal basis for $\widetilde{\mathcal{S}}_{\gamma,\theta}$. Then Ψ provides a stable sampling operator for $\mathcal{X}$ *if and only if*

$$\begin{aligned} 0 < \alpha &\stackrel{\text{def}}{=} \inf_{(\gamma,\theta)\in\Gamma\times\Gamma} \sigma_{\min}^2(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}) \\ &\leq \beta \stackrel{\text{def}}{=} \sup_{(\gamma,\theta)\in\Gamma\times\Gamma} \sigma_{\max}^2(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}) < \infty. \end{aligned} \tag{19}$$

And α and β are the tightest stability bounds.

C. Application to Finite Rate of Innovation Sampling

To see applications of the results so far, first let's revisit Example 1, where the union of subspaces $\mathcal{X}$ consists of streams of K Diracs and provides the basic signal model for finite rate of innovation sampling [12]–[14]. In this case, each subspace $\widetilde{\mathcal{S}}_{\gamma,\theta}$ has a basis $\{\delta(t-t_k)\}_{k=1}^M$ with $t_1 < t_2 < \ldots < t_M$, and $M \leq 2K$. We have already shown the minimum sampling requirement is $2K$. Since $\langle \delta(t-t_0), \psi(t) \rangle = \psi(t_0)$, it follows from Proposition 4 that a minimum sampling vector set $\Psi = \{\psi_n\}_{n=1}^{2K}$ provides an invertible sampling for streams of K Diracs if and only if

$$\det \begin{pmatrix} \psi_1(t_1) & \psi_1(t_2) & \cdots & \psi_1(t_{2K}) \\ \psi_2(t_1) & \psi_2(t_2) & \cdots & \psi_2(t_{2K}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{2K}(t_1) & \psi_{2K}(t_2) & \cdots & \psi_{2K}(t_{2K}) \end{pmatrix} \neq 0, \tag{20}$$

for every $t_1 < t_2 < \ldots < t_{2K}$.

The set of functions $\Psi = \{\psi_n\}_{n=1}^{2K}$ satisfying the above condition (20) is called a *Tchebycheff system* [30]. The classical example of a Tchebycheff system is the power functions $\psi_n = t^{n-1},\ n = 1, 2, \ldots, 2K$. In this case, the matrix in (20) is the familiar Vandermonde matrix. Tchebycheff systems play a prominent role in several areas of mathematics such as approximation, interpolation, and numerical analysis. Numerous examples of Tchebycheff systems are given in [30], including power functions, Gauss kernels, spline polynomials, $\sin$ and $\cos$ functions, and derived systems from these examples (for instance, if $\{\psi_n(t)\}_{n=1}^N$ is a Tchebycheff system and $w(t)$ is a positive and continuous function, then $\{w(t)\psi_n(t)\}_{n=1}^N$ is also a Tchebycheff system). The particular choices of sampling functions used in the finite rate of innovation sampling literature [12], [14] are of course among these examples.

The above discussion also applies to the signals of overlapping echoes in Example 5. Note that sampling $x(t) = \sum_{k=1}^{K} c_k \, \phi(t-t_k)$ with sampling functions $\{\psi_n(t)\}$ is equivalent to sampling a stream of Diracs $\sum_{k=1}^{K} c_k \, \delta(t-t_k)$ with sampling functions $\left\{ \widetilde{\psi}_n(t) \right\}$, where $\widetilde{\psi}_n(t) = \langle \phi(\cdot - t), \psi_n \rangle$. Thus the invertible sampling condition described in (20) can be used in the case of overlapping echoes as well.

D. Application to Compressed Sensing

In the compressed sensing setup [23]–[25], the signals of interest are supposed to have sparse representation, using up to K terms from an orthonormal basis $\{\phi_k\}$ as in Example 4; i.e.,

$$\mathcal{X} = \left\{ x \; : \; x = \sum_{k \in I} c_k \phi_k, \ |I| \leq K \right\},$$

where I is an index set and $|I|$ denotes its cardinality.

Let $\Psi = \{\psi_n\}_{n=1}^N$ be a set of sampling vectors. For each ϕ_k in the dictionary $\{\phi_k\}_k$, consider the column vector $\boldsymbol{g}_k = (\langle \phi_k, \psi_1 \rangle, \langle \phi_k, \psi_2 \rangle, \ldots, \langle \phi_k, \psi_N \rangle)^T$, and consider the matrix $\boldsymbol{G} = (\boldsymbol{g}_k)_k$ obtained by concatenating all of these columns. Then the problem of reconstructing $x \in \mathcal{X}$ from its sampling data $Ax = \boldsymbol{d}$ is equivalent to solving $\boldsymbol{c}$ from the matrix equation $\boldsymbol{Gc} = \boldsymbol{d}$ under the constraint that $\boldsymbol{c}$ has at most K non-zero entries.

Note that in this case, each subspace $\widetilde{\mathcal{S}}_{\gamma,\theta}$ has an orthonormal basis of the form $\Phi_{\gamma,\theta} = \{\phi_k\}_{k\in I}$ with $|I| \leq 2K$. Therefore, the Gram matrix $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}$ is formed by taking subsets of the columns of $\boldsymbol{G}$ as $\boldsymbol{G}_I = (\boldsymbol{g}_k)_{k\in I}$ with $|I| \leq 2K$. Hence, applying Proposition 5, we can write the stability bounds in this case as

$$\begin{aligned}\alpha &= \inf_{|I|=2K} \lambda_{\min}(G_I^* G_I),\\ \beta &= \sup_{|I|=2K} \lambda_{\max}(G_I^* G_I),\end{aligned} \tag{21}$$

where G_I^* is the conjugate transpose of G_I, and $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues.[3] Note that the stability bounds derived in (21) are closely related to the notion of *restricted isometry* in [31]. By noting that the entries of $G_I^* G_I$ are $\langle \boldsymbol{g}_k, \boldsymbol{g}_l \rangle,\ k,l \in I$, and using the Geršgorin disc theorem [29, pp. 344 - 345] to bound the eigenvalues of these matrices, we obtain

$$\begin{aligned}\alpha &\geq \inf_{|I|=2K-1,\ k\notin I} \left(\langle \boldsymbol{g}_k, \boldsymbol{g}_k \rangle - \sum_{l\in I} |\langle \boldsymbol{g}_k, \boldsymbol{g}_l \rangle| \right)\\ \beta &\leq \sup_{|I|=2K-1,\ k\notin I} \left(\langle \boldsymbol{g}_k, \boldsymbol{g}_k \rangle + \sum_{l\in I} |\langle \boldsymbol{g}_k, \boldsymbol{g}_l \rangle| \right).\end{aligned} \tag{22}$$

Therefore, for stable sampling, the condition $\beta < \infty$ is always satisfied; we only need to ensure $\alpha > 0$. Without loss of generality, we can suppose the columns of $\boldsymbol{G}$ to have unit norm; i.e. $\|\boldsymbol{g}_k\|_2 = 1$. Using the *cumulative coherence* functions that were defined in [32] as

$$\mu_1(m) \stackrel{\text{def}}{=} \sup_{|I|=m,\ k\notin I} \sum_{l\in I} |\langle g_k, g_l \rangle|,$$

we see from (22) that A is a stable sampling operator in this case if

$$\mu_1(2K-1) < 1. \tag{23}$$

It is easy to see that $\mu_1(m) \leq m\mu$, where $\mu \stackrel{\text{def}}{=} \sup_{k\neq l} |\langle \boldsymbol{g}_k, \boldsymbol{g}_l \rangle|$ is called the *coherence* parameter [33]. These coherence measures play a fundamental role in the compressed sensing literature. In comparison with (23), the sharpest available result in [32] shows under the stricter requirement

$$\mu_1(K-1) + \mu_1(K) < 1$$

that two efficient algorithms, Basis Pursuit and Orthogonal Matching Pursuit, can reconstruct K-sparse signals exactly from its sampling data.

E. Existence of Invertible Minimum Sampling Sets

In the case where $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_\gamma$ is a union of *countable* subspaces, the following proposition shows that the minimum sampling requirement is achieved by a dense set of sampling vectors.

[3] We have used the following equalities: $\lambda_{\min}(G^*G) = \sigma^2_{\min}(G)$ and $\lambda_{\max}(G^*G) = \sigma^2_{\max}(G)$.

Proposition 6 (Existence of Invertible Sampling Operators): Suppose that $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_\gamma$ is a *countable* union of subspaces of $\mathcal{H}$, and suppose that $N_{\min}$ as defined in (14) is finite. Then the collection of sampling vectors $\{\Psi = \{\psi_n\}_{n=1}^{N_{\min}} : \Psi$ provides an invertible sampling operator for $\mathcal{X}\}$ is dense in $\mathcal{H}^{N_{\min}}$.

Proof: Consider the following function that is defined for each $(\gamma,\theta) \in \Gamma\times\Gamma$ as the determinant of the Gram matrix in (17); i.e.

$$f_{\Phi_{\gamma,\theta}}(\psi_1,\ldots,\psi_{N_{\min}}) \stackrel{\text{def}}{=} \det(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}), \tag{24}$$

where $\Phi_{\gamma,\theta} = \{\phi_k^{(\gamma,\theta)}\}_{k=1}^{N_{\min}}$ is some basis for $\widetilde{\mathcal{S}}_{\gamma,\theta}$ (if $\dim(\widetilde{\mathcal{S}}_{\gamma,\theta}) < N_{\min}$ then we augment its basis to a set of $N_{\min}$ linearly independent vectors). From Proposition 4, Ψ provides an invertible sampling operator if $f_{\Phi_{\gamma,\theta}}(\psi_1,\ldots,\psi_{N_{\min}}) \neq 0$ for every $(\gamma,\theta) \in \Gamma\times\Gamma$.

Due to the continuity of the inner products and the continuity of determinant with respect to matrix entries, $f_{\Phi_{\gamma,\theta}}$ is continuous on $\mathcal{H}^{N_{\min}}$. Define the set

$$\begin{aligned}\mathcal{O}_{\Phi_{\gamma,\theta}} &\stackrel{\text{def}}{=} \{(\psi_1,\ldots,\psi_{N_{\min}}) : f_{\Phi_{\gamma,\theta}}(\psi_1,\ldots,\psi_{N_{\min}}) \neq 0\}\\ &= f^{-1}_{\Phi_{\gamma,\theta}}\left((-\infty,0)\cup(0,+\infty)\right).\end{aligned} \tag{25}$$

Since the set $(-\infty,0)\cup(0,+\infty)$ is open and f is continuous, $\mathcal{O}_{\Phi_{\gamma,\theta}}$ is open in $\mathcal{H}^{N_{\min}}$. As shown in the appendix, $\mathcal{O}_{\Phi_{\gamma,\theta}}$ is also a dense set. Now the set of invertible sampling vectors $\mathcal{O} = \bigcap_{(\gamma,\theta)\in\Gamma\times\Gamma} \mathcal{O}_{\Phi_{\gamma,\theta}}$ is a countable intersection of dense open sets in the complete metric space $\mathcal{H}^{N_{\min}}$. Hence, by the Baire theorem [34], $\mathcal{O}$ is dense in $\mathcal{H}^{N_{\min}}$. ■

As a nice application of this result, consider Example 4 of sparse representations. Suppose $\mathcal{H}$ is a separable Hilbert space and let $\{\phi_k\}_{k=1}^{\infty}$ be a countable basis for $\mathcal{H}$. Then the set $\mathcal{X}$ of all possible K-term representations as given in (4) using this basis constitutes a countable union of subspaces of dimension K in $\mathcal{H}$. On the one hand, from Proposition 3, an invertible sampling operator requires at least $2K$ sampling vectors. On the other hand, from Proposition 6, the collection of $2K$ vector sets $\{\psi_n\}_{n=1}^{2K}$ that provide invertible sampling operators is dense.

Existence results on invertible sampling operators of this type were shown in the compressed sensing literature [23]–[25], but only for *finite* unions of finite dimensional subspaces and with extra $\log$ and constant factors on the number of required sampling vectors. However, our result here does not imply stable sampling.

Note that Proposition 6 does not cover the case in Example 1 with streams of K Diracs, in which the index set $\Lambda = \mathbb{R}^K$ is not countable. As discussed in Section IV-C, only Tchebycheff systems lead to invertible sampling operators.

V. UNION OF SHIFT-INVARIANT SUBSPACES

In this section, we consider the case where the ambient space $\mathcal{H} = L^2(\mathbb{R})$ and the set $\mathcal{X}$ of signals of interest is a union of infinite-dimensional shift-invariant subspaces.

A. Shift-Invariant Subspaces

A finitely-generated shift-invariant subspace in $L^2(\mathbb{R})$ is defined as [35]

$$\mathcal{S}_\Phi = \left\{ x(t) = \sum_{k=1}^{K} \sum_{m\in\mathbb{Z}} c_{k,m}\phi_k(t-mT) : \boldsymbol{c} \in l_2 \right\}, \quad (26)$$

where $\Phi = \{\phi_k\}_{k=1}^K$ is called the set of generating functions of $\mathcal{S}_\Phi$, and $\boldsymbol{c} = \{c_{k,m}\}_{1\le k\le K,\, m\in\mathbb{Z}}$ is called the coefficient set of $x(t)$. For expositional simplicity, we will set $T=1$ by rescaling the time axis.

To make the representation in (26) stable and unambiguous, we require that the family of functions $\{\phi_k(t-m)\}_{1\le k\le K,\, m\in\mathbb{Z}}$ form a Riesz basis of $\mathcal{S}_\Phi$ [35], [36], [3]. This means that there must exist positive constants $0 < \alpha_\Phi \le \beta_\Phi < \infty$ such that

$$\alpha_\Phi \|\boldsymbol{c}\|_{l_2}^2 \le \left\| \sum_{k=1}^{K} \sum_{m\in\mathbb{Z}} c_{k,m}\phi_k(t-m) \right\|_{L^2(\mathbb{R})}^2 \le \beta_\Phi \|\boldsymbol{c}\|_{l_2}^2, \quad (27)$$

for all $\boldsymbol{c} \in l_2$, where $\|\boldsymbol{c}\|_{l_2}^2 = \sum_{k=1}^K \sum_{m\in\mathbb{Z}} |c_{k,m}|^2$ is the squared l_2-norm of $\boldsymbol{c}$. Note that this requirement implies any function $x(t) \in \mathcal{S}_\Phi$ has finite energy and is uniquely and stably determined by its coefficients $\boldsymbol{c}$.

Analogous to the *dimension* of a finite-dimensional subspace, the *length* of a shift-invariant subspace $\mathcal{S}$ is defined to be the cardinality of the smallest generating set for $\mathcal{S}$ [35]; i.e.,

$$\begin{aligned}\operatorname{len}(\mathcal{S}) &\stackrel{\text{def}}{=} \\ &\min \left\{ K : \mathcal{S} \text{ can be generated by } \{\phi_k\}_{k=1}^K \right\}.\end{aligned} \quad (28)$$

For example, for the space $\mathcal{S}_\Phi$ given in (26), we have $\operatorname{len}(\mathcal{S}_\Phi) = K$ if the generating functions $\{\phi_k\}_{k=1}^K$ satisfy the Riesz basis condition in (27).

A common approach to studying shift-invariant subspaces is to consider the Fourier domain [35], [36], [10]. Taking the Fourier transform of $x(t)$ in (26) and exchanging the order of integrations, we have

$$\begin{aligned}\widehat{x}(\omega) &= \sum_{k=1}^{K} \sum_{m\in\mathbb{Z}} \int_{-\infty}^{\infty} c_{k,m}\, \phi_k(t-m)\, e^{-j\omega t} dt \\ &= \sum_{k=1}^{K} \widehat{c}_k(\omega)\, \widehat{\phi}_k(\omega),\end{aligned} \quad (29)$$

where $\widehat{\phi}_k(\omega) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \phi_k(t) e^{-j\omega t} dt$ is the Fourier transform of $\phi_k(t)$ and $\widehat{c}_k(\omega) \stackrel{\text{def}}{=} \sum_m c_{k,m} e^{-j\omega m}$ is the discrete-time Fourier transform of the sequence $\boldsymbol{c}_k = \{c_{k,m}\}_{m\in\mathbb{Z}}$. Using (29), one can derive an equivalent Riesz basis requirement in the Fourier domain. We refer to [36], [10] for details.

B. Sampling Signals from a Union of Shift-Invariant Subspaces

Now we consider the class of signals that can be modeled as $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_{\Phi_\gamma}$, where each subspace $\mathcal{S}_{\Phi_\gamma}$ is a shift-invariant subspace generated by a finite set of functions Φ_γ. We want to sample signals from $\mathcal{X}$ by a sampling operator A characterized by a set of sampling vectors $\{\psi_n\}_{n\in\Lambda}$.

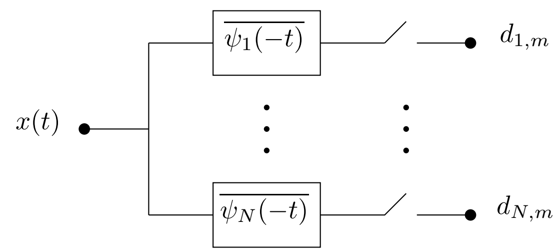

Fig. 3. Multi-channel sampling. The input signal $x(t)$ is first filtered by a bank of N filters, $\left\{\overline{\psi_n(-t)}\right\}_{n=1}^N$, and then the sampling data are taken at time instances $m \in \mathbb{Z}$.

We consider the case where the set of sampling vectors takes the form of $\{\psi_n(t-m)\}_{1\le n\le N,\, m\in\mathbb{Z}}$. In this case, the sampling procedure of computing $\langle x, \psi_n(\cdot - m)\rangle$ can be efficiently implemented by a bank of filtering followed by uniform pointwise sampling, as illustrated in Figure 3.

Specifically, by denoting $\xi_n(t) = \overline{\psi_n(-t)}$ we can express the sampling data as

$$d_{n,m} \stackrel{\text{def}}{=} \langle x, \psi_n(\cdot - m)\rangle = (x * \xi_n)(m). \quad (30)$$

In other words, $\{d_{n,m}\}_{m\in\mathbb{Z}}$ is the uniform sampling of the function $(x*\xi_n)(t)$ in the classical sense. Applying the classical sampling formula in the Fourier domain (as obtained from the Poisson summation formula), we can write the discrete-time Fourier transform of the sequence $\boldsymbol{d}_n \stackrel{\text{def}}{=} \{d_{n,m}\}_{m\in\mathbb{Z}}$ as

$$\begin{aligned}\widehat{d}_n(\omega) &\stackrel{\text{def}}{=} \sum_{m\in\mathbb{Z}} d_{n,m} e^{-j\omega m} \\ &= \sum_{m\in\mathbb{Z}} \widehat{x}(\omega + 2\pi m)\, \widehat{\xi}_n(\omega + 2\pi m).\end{aligned} \quad (31)$$

Therefore, if $x(t) \in \mathcal{S}_\Phi$ and is defined as in (26), then substituting (29) into (31), and noting that $\widehat{\xi}_n(\omega) = \overline{\widehat{\psi}_n(\omega)}$ and $\widehat{c}_k(\omega)$ is a 2π-periodic function, we obtain

$$\begin{aligned}\widehat{d}_n(\omega) &= \sum_{k=1}^{K} \sum_{m\in\mathbb{Z}} \widehat{c}_k(\omega+2\pi m)\, \widehat{\phi}_k(\omega+2\pi m)\, \overline{\widehat{\psi}_n(\omega+2\pi m)} \\ &= \sum_{k=1}^{K} \left(\sum_{m\in\mathbb{Z}} \widehat{\phi}_k(\omega+2\pi m)\, \overline{\widehat{\psi}_n(\omega+2\pi m)} \right) \widehat{c}_k(\omega).\end{aligned} \quad (32)$$

This leads to a compact relation between the sampling data $\boldsymbol{d} \stackrel{\text{def}}{=} \{d_{n,m}\}_{1\le n\le N,\, m\in\mathbb{Z}}$ and coefficients $\boldsymbol{c} \stackrel{\text{def}}{=} \{c_{k,m}\}_{1\le k\le K,\, m\in\mathbb{Z}}$ of $x(t) \in \mathcal{S}_\Phi$ via a matrix-vector multiplication in the Fourier domain

$$\widehat{\boldsymbol{d}}(\omega) = \boldsymbol{G}_{\Phi,\Psi}(\omega)\, \widehat{\boldsymbol{c}}(\omega), \quad (33)$$

where $\widehat{\boldsymbol{c}}(\omega) \stackrel{\text{def}}{=} (\widehat{c}_1(\omega), \widehat{c}_2(\omega), \ldots, \widehat{c}_K(\omega))^T$ and $\widehat{\boldsymbol{d}}(\omega) \stackrel{\text{def}}{=} (\widehat{d}_1(\omega), \widehat{d}_2(\omega), \ldots, \widehat{d}_N(\omega))^T$ are column vectors, and $\boldsymbol{G}_{\Phi,\Psi}(\omega)$ is an $N \times K$ matrix with entries

$$\{\boldsymbol{G}_{\Phi,\Psi}(\omega)\}_{n,k} \stackrel{\text{def}}{=} \sum_{m\in\mathbb{Z}} \widehat{\phi}_k(\omega+2\pi m)\, \overline{\widehat{\psi}_n(\omega+2\pi m)}. \quad (34)$$

Note that (33)-(34) closely resemble (16)-(17), and thus we can consider $\boldsymbol{G}_{\Phi,\Psi}(\omega)$ as the Fourier-domain Gram matrix between two sets of generating functions $\Phi = \{\phi_k(t)\}_{k=1}^K$ and $\Psi = \{\psi_n(t)\}_{n=1}^N$.

C. Sampling Conditions for Union of Shift-Invariant Subspaces

Using the results from Section III we can derive the sampling conditions for a union of shift-invariant subspaces $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_{\Phi_\gamma}$ by considering subspaces $\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}} \stackrel{\text{def}}{=} \mathcal{S}_{\Phi_\gamma} + \mathcal{S}_{\Phi_\theta}$. Clearly, $\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}}$ is also a shift-invariant subspace that can be generated by the set of functions $\Phi_\gamma \cup \Phi_\theta$. Denote $\Phi_{\gamma,\theta}$ as a set of generating functions for a Riesz basis for $\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}}$. Thus, it follows from the definition in (28) that $|\Phi_{\gamma,\theta}| = \text{len}(\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}}) \leq \text{len}(\mathcal{S}_{\Phi_\gamma})+\text{len}(\mathcal{S}_{\Phi_\theta}) = |\Phi_\gamma|+|\Phi_\theta|$. Applying the relation given in (33), we can express $A\,y$ for $y \in \widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}}$ via a matrix-vector multiplication in the Fourier domain with the Gram matrix $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)$ of size $N \times \text{len}(\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}})$.

Proposition 7: Suppose that the mapping $A : x \mapsto \{\langle x, \psi_n(\cdot - m)\rangle\}_{1\leq n\leq N,\, m\in\mathbb{Z}}$ is an *invertible* sampling operator for $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_{\Phi_\gamma}$. Then

$$N \geq N_{\min} = \sup_{(\gamma,\theta)\in\Gamma\times\Gamma} \text{len}(\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}}). \tag{35}$$

Proof: From Proposition 1, A is an invertible sampling operator for $\mathcal{X}$ if and only if A is one-to-one on every $\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}}$, $(\gamma,\theta) \in \Gamma\times\Gamma$. Hence, from the matrix-vector representation given in (33), the invertible sampling condition is equivalent to the Gram matrix $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)$ having full column rank, which implies that $N \geq |\Phi_{\gamma,\theta}| = \text{len}(\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}})$ for every $(\gamma,\theta) \in \Gamma\times\Gamma$. ■

Proposition 7 provides an easy-to-compute minimum sampling requirement $N_{\min}$, interpreted as the minimum number of channels in the multi-channel sampling illustrated in Figure 3, or equivalently the minimum number of samples per unit of time, for a union of shift-invariant subspaces. Using the same reasoning leading to Proposition 7, we can obtain the following condition for invertible sampling, whose proof is omitted due to similarity.

Proposition 8: Let $\Psi = \{\psi_n\}_{n=1}^N$ be a set of sampling functions and $\Phi_{\gamma,\theta} = \left\{\phi_k^{(\gamma,\theta)}\right\}_{k=1}^{K_{\gamma,\theta}}$ be a set of generating functions of a Riesz basis for $\widetilde{\mathcal{S}}_{\gamma,\theta}$. Then $\{\psi_n(\cdot - m)\}_{1\leq n\leq N,\, m\in\mathbb{Z}}$ provides an invertible sampling operator for $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_{\Phi_\gamma}$ *if and only if,* for any choice of $(\gamma,\theta) \in \Gamma\times\Gamma$, the corresponding Gram matrix $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)$ has full column rank for almost[4] every ω.

Next we will derive stability conditions for sampling. For simplicity, similar to Proposition 5, we suppose that $\left\{\phi_k^{(\gamma,\theta)}(\cdot - m)\right\}_{1\leq k\leq K_{\gamma,\theta},\, m\in\mathbb{Z}}$ is an orthonormal basis for $\widetilde{\mathcal{S}}_{\Phi_{\gamma,\theta}}$. This assumption is made without loss of generality, since, analogous to [36, Theorem 4.1] on the existence and construction of the dual basis, we can always orthogonalize a set of generating functions for a shift-invariant subspace to obtain an orthonormal basis for it.

Proposition 9: Let $\Psi = \{\psi_n\}_{n=1}^N$ be a set of sampling functions and $\Phi_{\gamma,\theta} = \left\{\phi_k^{(\gamma,\theta)}\right\}_{k=1}^{K_{\gamma,\theta}}$ be a set of generating functions of an orthonormal basis for $\widetilde{\mathcal{S}}_{\gamma,\theta}$. Then $\{\psi_n(\cdot - m)\}_{1\leq n\leq N,\, m\in\mathbb{Z}}$ provides a stable sampling operator for $\mathcal{X} = \bigcup_{\gamma\in\Gamma} \mathcal{S}_{\Phi_\gamma}$ *if and only if*

$$0 < \alpha \stackrel{\text{def}}{=} \operatorname*{ess\,inf}_{(\gamma,\theta)\in\Gamma\times\Gamma,\ \omega\in[-\pi,\pi]} \sigma^2_{\min}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)) \tag{36}$$

$$\leq \beta \stackrel{\text{def}}{=} \operatorname*{ess\,sup}_{(\gamma,\theta)\in\Gamma\times\Gamma,\ \omega\in[-\pi,\pi]} \sigma^2_{\max}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)) < \infty. \tag{37}$$

And α and β are the tightest stability bounds.

Proof: Suppose that $y \in \widetilde{\mathcal{S}}_{\gamma,\theta}$ and $\boldsymbol{c} = \{c_{k,m}\}_{1\leq k\leq K_{\gamma,\theta},\, m\in\mathbb{Z}}$ is the coefficients of y. Since the set of vectors $\left\{\phi_k^{(\gamma,\theta)}(\cdot - m)\right\}_{1\leq k\leq K_{\gamma,\theta},\, m\in\mathbb{Z}}$ is an orthonormal basis of $\widetilde{\mathcal{S}}_{\gamma,\theta}$, it follows that $\|y\|_{L^2} = \|\boldsymbol{c}\|_{l_2}$.

Using the Parseval equality, we have

$$\begin{aligned}\|\boldsymbol{c}\|^2_{l_2} = \sum_{k=1}^{K_{\gamma,\theta}} \|\boldsymbol{c}_k\|^2_{l_2} &= \frac{1}{2\pi}\sum_{k=1}^{K_{\gamma,\theta}} \int_{-\infty}^{\infty} |\widehat{c}_k(\omega)|^2 d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\boldsymbol{c}}^*(\omega)\,\widehat{\boldsymbol{c}}(\omega) d\omega,\end{aligned} \tag{38}$$

where $\widehat{\boldsymbol{c}}(\omega) \stackrel{\text{def}}{=} (\widehat{c}_1(\omega), \widehat{c}_2(\omega), \ldots, \widehat{c}_{K_{\gamma,\theta}}(\omega))^T$ and $\widehat{\boldsymbol{c}}^*(\omega)$ is the conjugate transpose of $\widehat{\boldsymbol{c}}(\omega)$. Similarly, for sampling data $Ay = \boldsymbol{d} = \{d_{n,m}\}_{1\leq n\leq N,\, m\in\mathbb{Z}}$, we have

$$\|\boldsymbol{d}\|^2_{l_2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\boldsymbol{d}}^*(\omega)\,\widehat{\boldsymbol{d}}(\omega) d\omega.$$

Since $\widehat{\boldsymbol{d}}(\omega) = \boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)\,\widehat{\boldsymbol{c}}(\omega)$, we know from matrix theory that for (almost) every ω,

$$\begin{aligned}\sigma^2_{\min}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega))\,\widehat{\boldsymbol{c}}^*(\omega)\,\widehat{\boldsymbol{c}}(\omega) &\leq \widehat{\boldsymbol{d}}^*(\omega)\,\widehat{\boldsymbol{d}}(\omega) \\ &\leq \sigma^2_{\max}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega))\,\widehat{\boldsymbol{c}}^*(\omega)\,\widehat{\boldsymbol{c}}(\omega).\end{aligned}$$

Therefore,

$$\begin{aligned}\left(\operatorname*{ess\,inf}_{\omega\in[-\pi,\pi]} \sigma^2_{\min}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega))\right)\|\boldsymbol{c}\|^2_{l_2} &\leq \|\boldsymbol{d}\|^2_{l_2} \\ &\leq \left(\operatorname*{ess\,sup}_{\omega\in[-\pi,\pi]} \sigma^2_{\max}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega))\right)\|\boldsymbol{c}\|^2_{l_2}.\end{aligned}$$

And the bounds are tight. Combining these bounds for all $(\gamma,\theta) \in \Gamma\times\Gamma$ and using Proposition 2, we obtain the desired result. ■

D. Case Study: Spectrum-Blind Sampling of Multiband Signals

To demonstrate the proposed theory of sampling signals from a union of shift-invariant subspaces, we will revisit here the problem described in Example 6, where the signals of interest are multiband signals with unknown spectral support.

[4]This technicality is due to the fact that, for some generating functions $\Phi_{\gamma,\theta}$ and sampling functions Ψ, the corresponding Gram matrix $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)$ may not have well-defined pointwise values on a set of measure zero.

Our discussions differ in style as well as in technical details from some of the original results of Bresler *et al.* [20]–[22], who first proposed and studied the spectrum-blind sampling and reconstruction of these multiband signals.

As shown in Figure 1(e), we partition the spectral span $\mathcal{F} = [\omega_{\min}, \omega_{\max}]$ into L equally spaced spectral cells $\{\mathcal{C}_i\}_{i=0}^{L-1}$. For simplicity of exposition, we set $\omega_{\max} - \omega_{\min} = 2\pi L$ (after rescaling the time axis); each cell can then be specified as $\mathcal{C}_i = [\omega_{\min}+2\pi i,\ \omega_{\min}+2\pi(i+1)]$. The signals to be sampled have nonzero frequency values in at most K spectral cells (with $K \ll L$), though we do not know the exact locations of these cells.

Clearly, the signals of interest form a union of subspaces, and can be written as

$$\mathcal{X}^{(MB)} = \bigcup_{\gamma \in \Gamma} \mathcal{S}^{(BL)}_{\mathcal{F}_\gamma},$$

where $\gamma \stackrel{\text{def}}{=} \{q_1, q_2, \ldots, q_K : 0 \le q_k < L\}$ represents a set of indices, specifying a possible choice of K, out of L, spectral cells; $\mathcal{F}_\gamma \stackrel{\text{def}}{=} \bigcup_{q_k \in \gamma} \mathcal{C}_{q_k}$ is the finite union of these K cells; and $\mathcal{S}^{(BL)}_{\mathcal{F}_\gamma}$ is the subspace of all continuous functions bandlimited to $\mathcal{F}_\gamma$.

To apply the results in Section V-C, we consider the subspace $\widetilde{\mathcal{S}}_{\mathcal{F}_{\gamma,\theta}} = \mathcal{S}^{(BL)}_{\mathcal{F}_\gamma} + \mathcal{S}^{(BL)}_{\mathcal{F}_\theta}$, which consists of all continuous functions bandlimited to $\mathcal{F}_{\gamma,\theta} = \mathcal{F}_\gamma \cup \mathcal{F}_\theta$. Let $\phi_i(t)$ represent the function whose Fourier transform is the indicator function $\mathbb{1}_{\mathcal{C}_i}(\omega)$ of the ith cell, i.e.,

$$\widehat{\phi}_i(\omega) = \mathbb{1}_{\mathcal{C}_i}(\omega), \quad 0 \le i < L. \tag{39}$$

We can then verify that the shift-invariant subspace $\widetilde{\mathcal{S}}_{\mathcal{F}_{\gamma,\theta}}$ has an orthonormal basis $\{\phi_{q_k}(\cdot - m)\}_{q_k \in \gamma \cup \theta,\, m \in \mathbb{Z}}$, where $\gamma \cup \theta = \{q_1, q_2, \ldots, q_M\}$ are the indices of the M different cells in $\mathcal{F}_{\gamma \cup \theta}$. Since $\text{len}(\widetilde{\mathcal{S}}_{\mathcal{F}_{\gamma,\theta}}) = M = |\gamma \cup \theta| \le |\gamma| + |\theta| = 2K$ with equality when γ and θ are disjoint, it follows from the minimum sampling requirement in Proposition 7 that we need at least $N_{\min} = 2K$ samples per unit of time to determine uniquely all signals from $\mathcal{X}^{(MB)}$ from their sampling data. This is twice the rate we would need if we possessed prior knowledge about the frequency support. However, this minimum sampling rate can still be much more efficient than the Nyquist rate, which is based on the entire bandwidth $\omega_{\max} - \omega_{\min} = 2\pi L$ and, therefore, requires $L(\gg 2K)$ samples per unit of time.

Next, we will show that the above minimum sampling rate can be achieved, i.e., there exist suitable choices of $2K$ sampling functions $\{\psi_n\}_{n=1}^{2K}$ providing stable sampling for $\mathcal{X}^{(MB)}$.

Proposition 10: The sampling process $A : x \mapsto \{\langle x, \psi_n(\cdot - m)\rangle\}_{1 \le n \le 2K,\, m \in \mathbb{Z}}$ is a *stable* sampling operator for the multiband signals $\mathcal{X}^{(MB)}$ if the Fourier transforms of the sampling functions $\widehat{\Psi} = \left\{\widehat{\psi}_n(\omega)\right\}_{n=1}^{2K}$ are continuous and form a Tchebycheff system on the interval $[\omega_{\min}, \omega_{\max}]$, i.e.,

$$\det \begin{pmatrix} \widehat{\psi}_1(\omega_1) & \widehat{\psi}_1(\omega_2) & \cdots & \widehat{\psi}_1(\omega_{2K}) \\ \widehat{\psi}_2(\omega_1) & \widehat{\psi}_2(\omega_2) & \cdots & \widehat{\psi}_2(\omega_{2K}) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\psi}_{2K}(\omega_1) & \widehat{\psi}_{2K}(\omega_2) & \cdots & \widehat{\psi}_{2K}(\omega_{2K}) \end{pmatrix} \neq 0, \tag{40}$$

for all choices of $\omega_{\min} \le \omega_1 < \omega_2 < \ldots < \omega_{2K} < \omega_{\max}$.

Proof: First, we substitute (39) into (34), and write the entries of the Gram matrix as

$$\begin{aligned}\left\{\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)\right\}_{n,k} \\ &= \sum_{m \in Z} \mathbb{1}_{\mathcal{C}_{q_k}}(\omega + 2\pi m)\, \overline{\widehat{\psi}_n(\omega + 2\pi m)},\end{aligned} \tag{41}$$

where q_k is the index of the kth cell in $\mathcal{F}_{\gamma \cup \theta}$, and $1 \le k \le 2K$. Since (41) represents a 2π-periodic function, we only need to evaluate its values in one period. On the interval $\omega \in [\omega_{\min}, \omega_{\min} + 2\pi]$, we have

$$\mathbb{1}_{\mathcal{C}_{q_k}}(\omega + 2\pi m) = \begin{cases} 1 & \text{when } m = q_k, \\ 0 & \text{when } m \neq q_k, \end{cases} \tag{42}$$

and thus

$$\left\{\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)\right\}_{n,k} = \overline{\widehat{\psi}_n(\omega + 2\pi q_k)},$$

for $1 \le n, k \le 2K$. Consequently, the condition in (40) simply implies that the Gram matrix $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)$ always has full column rank.

Next, denote $\alpha_{\gamma,\theta}(\omega) = \sigma^2_{\min}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega))$ and $\beta_{\gamma,\theta}(\omega) = \sigma^2_{\max}(\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega))$. For any fixed (γ, θ), we know from the full rank property of $\boldsymbol{G}_{\Phi_{\gamma,\theta},\Psi}(\omega)$ that $0 < \alpha_{\gamma,\theta}(\omega) \le \beta_{\gamma,\theta}(\omega) < \infty$ for all $\omega \in [\omega_{\min}, \omega_{\min} + 2\pi]$. Since both $\alpha_{\gamma,\theta}(\omega)$ and $\beta_{\gamma,\theta}(\omega)$ are continuous functions of ω (due to the assumption that $\left\{\widehat{\psi}_n(\omega)\right\}$ are continuous functions on $[\omega_{\min}, \omega_{\max}]$), we can further conclude that there exist $\alpha_{\gamma,\theta}$ and $\beta_{\gamma,\theta}$ (independent of ω) such that $0 < \alpha_{\gamma,\theta} \le \alpha_{\gamma,\theta}(\omega) \le \beta_{\gamma,\theta}(\omega) \le \beta_{\gamma,\theta} < \infty$ for all ω on the finite and closed interval $[\omega_{\min}, \omega_{\min}+2\pi]$. Moreover, since there is only a *finite* number of choices for (γ, θ) (corresponding to all possible configurations of choosing up to $2K$ cells out of L cells), we can find constants α, β such that $0 < \alpha \le \alpha_{\gamma,\theta} \le \beta_{\gamma,\theta} \le \beta < \infty$ for all (γ, θ), which implies the conditions in (36) and (37) for stable sampling. ■

In the following, we give two concrete examples of the sampling functions $\{\psi_n\}_{n=1}^{2K}$ that satisfy the conditions required in Proposition 10. The first is to consider

$$\widehat{\psi}_n(\omega) = e^{-j\frac{n-1}{L}\omega}\, \mathbb{1}_{\mathcal{F}}(\omega), \quad 1 \le n \le 2K, \tag{43}$$

where $\mathbb{1}_{\mathcal{F}}(\omega)$ is the indicator function of the spectral span $\mathcal{F}$. It is easy to verify (from the property of the Vandermonde matrix) that the determinant of the matrix in (40) is always different from zero. Since any function $x(t)$ from $\mathcal{X}^{(MB)}$ is bandlimited within $\mathcal{F}$, we can obtain from (30) that the resulting sampling data can be written as

$$d_{n,m} = x\left(m + \frac{n-1}{L}\right), \tag{44}$$

for $1 \leq n \leq N$ and $m \in \mathbb{Z}$. This becomes exactly the original sampling scheme proposed in [20], where the sampling data are obtained by directly taking the pointwise values of the input signal on a periodic nonuniform pattern $\bigcup_{n=1}^{N} \bigcup_{m\in\mathbb{Z}} (m + \frac{n-1}{L})$.

In the second example, we propose a new sampling scheme that has not been considered in the previous work [20]–[22]. Let the sampling functions be Gaussian kernels defined (in the Fourier domain) as

$$\widehat{\psi}_n(\omega) = e^{-\omega^2 n/\sigma^2}, \quad 1 \leq n \leq 2K, \tag{45}$$

for some constant $\sigma > 0$. In this case, the matrix in (40) becomes a (generalized) Vandermonde matrix, whose elements $\left\{e^{-\omega_k^2/\sigma^2}\right\}_{k=1}^{2K}$ are always distinct for arbitrary choice of $\omega_{\min} \leq \omega_1 < \omega_2 < \ldots < \omega_{2K} < \omega_{\max}$, as long as we assume[5] $\omega_{\min} \geq 0$. It then follows from Proposition 10 that the proposed sampling vectors given in (45) can also provide stable sampling for the multiband signals.

The sampling data in this case are $d_{n,m} = (x(t) * \psi_n(t))(m)$, where the spatial domain sampling functions are $\psi_n(t) = e^{-t^2\sigma^2/(4n)}\sqrt{\sigma^2/(4n\pi)}$. Compared with the previous sampling scheme shown in (44), the proposed new scheme differs in two ways: first, instead of directly taking pointwise values, the sampling data are now obtained by averaging the input signals with Gaussian kernels; second, all the sampling data in the new scheme are taken at the same sampling instances (i.e. $m \in \mathbb{Z}$) without a timing shift. The latter property can be desirable in practical implementations, since it eliminates the need to carefully control the timing offsets between different sampling cosets, which was required in the periodic nonuniform sampling procedure in (44).

Finally, we would like to point out that the class of stable sampling vectors for the multiband signals are not limited to the two choices given in (43) and (45). As we have shown in Proposition 10, a set of $2K$ sampling functions provide stable sampling for $\mathcal{X}^{(MB)}$ if their Fourier transforms are continuous and form a Tchebycheff system on the interval $\omega \in [\omega_{\min}, \omega_{\max}]$. The two particular choices in (43) and (45) are just special cases of the Tchebycheff systems, which contain many other possibilities as mentioned in Section IV-C. This generalization about suitable sampling functions opens door to greater flexibilities in the design of the sampling systems.

VI. Conclusions

We proposed a new sampling problem where the signals of interest live on a union of subspaces. The first two questions outlined in Section II-A were addressed in this work, which involve the feasibility and performance bounds of the proposed sampling framework. The key geometrical viewpoint was to find a suitable sampling operator which projects the signals of interest into a lower dimensional representation space while still preserves all the information. Starting from the case of unions of finite-dimensional subspaces, we derived necessary and sufficient conditions for such sampling operators to exist, and found the minimum sampling requirement. Next, we extended all the results to the case of unions of infinite-dimensional shift-invariant subspaces.

The proposed sampling framework for unions of subspaces has close ties to the prior work of finite rate of innovation sampling, compressed sensing/compressive sampling, and spectrum-blind sampling, in which new results and derivations were discovered. It is our hope that the proposed framework can serve as a common ground and facilitate the interplay between the above three lines of thinking. Moreover, the idea of modeling signals as coming from unions of subspaces provides a useful geometrical viewpoint for finite rate of innovation sampling and suggests a path for extending the current compressed sensing/compressive sampling work from discrete and finite dimensional cases to continuous and infinite-dimensional cases (e.g. by considering unions of shift-invariant subspaces).

[5] This assumption is made without loss of generality, since we can always apply a frequency modulation to the signals before sampling, to make the assumed condition hold.

Appendix

We will show that for a linearly independent set of vectors $\Phi \stackrel{\text{def}}{=} \{\phi_1, \phi_2, \ldots, \phi_N\}$, the set $\mathcal{O}_\Phi$ defined in (25) is dense in $\mathcal{H}^N$. Geometrically, this means that, given an N-dimensional subspace, the set of N-dimensional subspaces onto which the former subspace can be projected without losing dimensions is dense.

Suppose that $\Psi \stackrel{\text{def}}{=} \{\psi_1, \psi_2, \ldots, \psi_N\} \notin \mathcal{O}_\Phi$. We will show that there exists a $\widetilde{\Psi} \in \mathcal{O}_\Phi$ that is arbitrarily close to Ψ. For the Gram matrix $\boldsymbol{G}_{\Phi,\Psi}$ as defined in (17), its singular value decomposition has the form

$$\boldsymbol{G}_{\Phi,\Psi} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^*,$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are two unitary matrices, and $\boldsymbol{\Lambda}$ is a diagonal matrix with real and *non-negative* entries. We can always find another diagonal matrix $\boldsymbol{\Lambda}_2$ such that for all $\alpha > 0$, $\boldsymbol{\Lambda} + \alpha\boldsymbol{\Lambda_2}$ is a diagonal matrix with real and *positive* entries.

Since $\{\phi_n\}_{n=1}^N$ is a linearly independent set, it is easy to verify that $\boldsymbol{G}_{\Phi,\Phi}$ is invertible. Let $\widetilde{\Psi} = \Psi + \alpha\,\boldsymbol{U}\boldsymbol{\Lambda_2}\boldsymbol{V}^*\boldsymbol{G}_{\Phi,\Phi}^{-1}\,\Phi$. Because the Gram matrix is linear with respect to its constituent sets of vectors, we have

$$\boldsymbol{G}_{\Phi,\widetilde{\Psi}} = \boldsymbol{G}_{\Phi,\Psi} + \alpha\,\boldsymbol{U}\boldsymbol{\Lambda_2}\boldsymbol{V}^*\boldsymbol{G}_{\Phi,\Phi}^{-1}\boldsymbol{G}_{\Phi,\Phi} = \boldsymbol{U}(\boldsymbol{\Lambda} + \alpha\boldsymbol{\Lambda_2})\boldsymbol{V}^*.$$

Thus, by construction $\det(\boldsymbol{G}_{\Phi,\widetilde{\Psi}}) \neq 0$, which means $\widetilde{\Psi} \in \mathcal{O}_\Phi$. Since α can be arbitrarily small, we are done.

Acknowledgment

The authors would like to thank R. Laugesen, L. Jacques, and anonymous reviewers for their helpful comments and suggestions.

References

[1] C. E. Shannon, “Communications in the presence of noise,” *Proc. IRE*, vol. 37, pp. 10–21, 1949.

[2] A. Jerri, “The Shannon sampling theorem - its various extensions and applications: A tutorial review,” *Proc. IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.

[3] M. Unser, "Sampling - 50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

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

[5] G. G. Walter, "A sampling theorem for wavelet subspaces," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 881–884, Mar. 1992.

[6] M. Unser and A. Aldroubi, "A general sampling theory for non-ideal acquisition devices," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2915–2925, Nov. 1994.

[7] R. A. Gopinath, J. E. Odegard, and C. S. Burrus, "Optimal wavelet representation of signals and the wavelet sampling theorem," *IEEE Trans. Circ. and Syst. II*, vol. CAS-41, pp. 262–277, Apr. 1994.

[8] I. Djokovic and P. P. Vaidyanathan, "Generalized sampling theorems in multiresolution subspaces," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 583–599, Mar. 1997.

[9] A. Aldroubi and K. Gröchenig, "Nonuniform sampling and reconstruction in shift-invariant spaces," *SIAM Review*, vol. 43, no. 4, pp. 585–620, 2001.

[10] C. Zhao and P. Zhao, "Sampling theorem and irregular sampling theorem for multiwavelet subspaces," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 705–713, Feb. 2005.

[11] P. Zhao, C. Zhao, and P. G. Casazza, "Pertubation of regular sampling in shift-invariant spaces for frames," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4643–4648, Oct. 2006.

[12] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[13] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[14] P. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[15] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.

[16] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego: Academic Press, 1999.

[17] A. M. Bruchstein, T. J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 33, no. 6, pp. 1357–1367, Dec. 1985.

[18] S. F. Yau and Y. Bresler, "Maximum likelihood parameter estimation of superimposed signals by dynamic programming," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 804–820, Feb. 1993.

[19] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, 1967.

[20] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Atlanta, USA, 1996.

[21] Y. Bresler and P. Feng, "Spectrum-blind minimum-rate sampling and reconstruction of 2-D multiband signals," in *Proc. IEEE Int. Conf. on Image Proc.*, Lausanne, Switzerland, Sep. 1996.

[22] R. Venkataramani and Y. Bresler, "Further results on spectrum blind sampling of 2D signals," in *Proc. IEEE Int. Conf. on Image Proc.*, Chicago, IL, Oct. 1998.

[23] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[24] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[25] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[26] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York, NY: Springer-Verlag, 1982.

[27] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[28] D. S. Broomhead and M. Kirby, "A new approach to dimensionality reduction: Theory and algorithms," *SIAM J. Appl. Math.*, vol. 60, no. 6, pp. 2114–2142, 2000.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[30] S. Karlin and W. J. Studden, *Tchebycheff Systems: with Applications in Analysis and Statistics*. New York, NY: John Wiley & Sons, 1966.

[31] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[32] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, Oct. 2004.

[33] D. L. Donoho and M. Elad, "Maximal sparsity representation via l_1 minimization," *Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, Mar. 2003.

[34] H. L. Royden, *Real Analysis*. Upper Saddle River, NJ: Prentice Hall, 1988.

[35] C. de Boor, R. A. DeVore, and A. Ron, "The structure of finitely generated shift-invariant spaces in $L_2(\mathbb{R}^d)$," *J. Funct. Anal.*, vol. 119, pp. 37 – 78, 1994.

[36] J. S. Geronimo, D. P. Hardin, and P. R. Massopust, "Fractal functions and wavelet expansions based on several scaling functions," *Journal of Approximation Theory*, vol. 78, no. 3, pp. 373–401, 1994.

Yue M. Lu received the B.Eng and M.Eng degrees in electrical engineering from Shanghai Jiao Tong University, China, in 1999 and 2002, respectively. He received the M.Sc degree in mathematics and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 2007.

He was a Research Assistant at the University of Illinois at Urbana-Champaign, and has worked for Microsoft Research Asia, Beijing, China and Siemens Corporate Research, Princeton, NJ. He is now with the Audio-Visual Communications Laboratory at the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland. His research interests include signal processing for sensor networks; the theory, constructions, and applications of multiscale geometric representations for multidimensional signals; image and video processing; and sampling theory.

He received the Most Innovative Paper Award of IEEE International Conference on Image Processing (ICIP) in 2006 for his paper (with Minh N. Do) on the construction of directional multiresolution image representations, and the Student Paper Award of IEEE ICIP in 2007.

Minh N. Do was born in Thanh Hoa, Vietnam, in 1974. He received the B.Eng. degree in computer engineering from the University of Canberra, Australia, in 1997, and the Dr.Sci. degree in communication systems from the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, in 2001.

Since 2002, he has been an Assistant Professor with the Department of Electrical and Computer Engineering and a Research Assistant Professor with the Coordinated Science Laboratory and the Beckman Institute, University of Illinois at Urbana-Champaign (UIUC). His research interests include image and multi-dimensional signal processing, wavelets and multiscale geometric analysis, computational imaging, and visual information representation.

He received a Silver Medal from the 32nd International Mathematical Olympiad in 1991, a University Medal from the University of Canberra in 1997, the best doctoral thesis award from EPFL in 2001, and a CAREER award from the National Science Foundation in 2003. He was named a Beckman Fellow at the Center for Advanced Study, UIUC, in 2006, and received of a Xerox Award for Faculty Research from the College of Engineering, UIUC, in 2007. He is a member of the IEEE Signal Processing Society Signal Processing Theory and Methods and Image and MultiDimensional Signal Processing Technical Committees, and an Associate Editor of the IEEE Transactions on Image Processing.