IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 May 2022, accepted 21 July 2022, date of publication 29 July 2022, date of current version 4 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3194926

RESEARCH ARTICLE

Continuous Compressed Sensing Hilbert-Schmidt Integral Operator

MOHAMMADREZA ROBAEI, (Member, IEEE), AND ROBERT AKL, (Senior Member, IEEE)
Department of Computer Science and Engineering, University of North Texas, Denton, TX 76210, USA
Corresponding author: Mohammadreza Robaei (mohammadrezarobaei@my.unt.edu)

ABSTRACT Continuous Compressed-Sensing-Karhunen-Loéve Expansion (CS-KLE) has been proposed. Compressed sensing has been proposed as a highly efficient computational method to represent compressible signals using a few numbers of linear functional. On the other hand, KLE is known to be the optimum orthogonal decomposition. While both methodologies have been addressed comprehensively and independently in the literature, their relationship has not been studied. In this work, we study the relation between random sampling and KLE. In particular, we examine how doubly orthogonal property is affected by the mutual coherency and RIP of the compressed sensing. A detailed theoretical study of random sampling and KLE is conducted. We prove the Compressed Sensing Hilbert-Schmidt integral operator as double integral acting on the signal space and its dual space. The proof of the proposed integral operator follows from the Kolmogorov Conditional Expectation theorem. Then, two formulations are proposed to compute CS-KLE relation, (1) through Mercer's theorem and (2) through Green's theorem. Also, the convergence of CS-KLE with respect to RIP is proved. It has been shown that there is a transition point in the spectral overlap between the estimated and actual signal spaces. The transition point occurs for the optimum subspace of the given compressible signal. Numerical simulation is presented by applying CS-KLE to semi-infinite and infinite-dimensional signals and also Magnetic Resonance Images (MRI).

INDEX TERMS Karhunen-Loéve expansion, continuous compressed sensing, separable Hilbert space, Kernel, random process, restricted isometry property, infinite-dimensional signals.

I. INTRODUCTION

Compressed sensing aims to represent and reconstruct sparse vectors using infimum possible subspace such that the signal space can be decomposed into orthogonal and orthogonal complement subspace [1]. For this purpose, greedy-pursuit [2]–[5] and ℓ_1-convex optimization solutions [6], [7] have been recommended. While ℓ_1-solutions are good at the accurate reconstruction of sparse vectors, the greed-pursuit algorithms are known for a fast convergence rate.

On the other hand, Karhunen-Loéve Expansion (KLE) [8, Chapter 6] is known for the optimum orthogonal decomposition of the signal. However, KLE suffers from numerous numerical problems, including computational complexity decomposing signal space using an infinite number of eigenpairs. To reduce computational complexity, truncated KLE has also been recommended to solve the partial differential equation using the stochastic Galerkin method [9].

The relation between CS and KLE has not been studied in the literature, even though both address a random process hidden in data. In this paper, we investigate the relation between compressed sensing and Karhunen-Loéve Expansion to propose compressed sensing Hilbert-Schmidt operator in Theorem 6. Accordingly, we propose Compressed sensing-KLE (CS-KLE) based on compressed sensing Hilbert-Schmidt operator. Two distinct formulations for CS-KLE have been proved in Theorems 7 and 11.

II. MOTIVATION

A. TRANSITION POINT, AND ℓ_1 THEORETICAL RESTRICTION

It has been shown by Donoho [10] that theoretically, regions of success and failure of the sparse recovery algorithms

The associate editor coordinating the review of this manuscript and approving it for publication was Qingchun Chen.

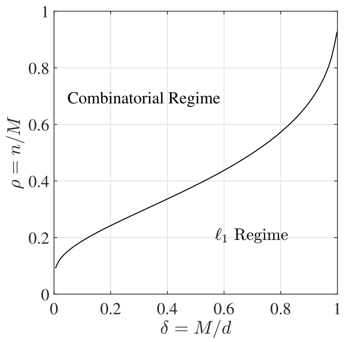

FIGURE 1. **Theoretical ℓ_1 transition phase [11, SparsLab].**

are well separated by a *sharp threshold* with tight upper and lower bounds. As shown in Fig. 1, the transition phase indicates that the performance of the sparse recovery algorithms is optimum for a certain set of problems that satisfy the range requirements determined by a normalized ratio $\rho = \frac{n}{M}$ and undersampling ratio $\delta = \frac{M}{d}$. Beyond the transition phase, one needs to consider the combinatorial search to estimate the signal space.

In [11], the authors have examined several algorithms, including LASSO [6] and LARS [12], and compared them with the theoretical ℓ_1 performance. From their works, one notices that LASSO shows the best performance in addressing different ranges of ρ and δ. Nevertheless, the significant range of problems regarding ρ and δ lies in the combinatorial search region.

Due to the transition point, one should expect an optimum subspace for which the performance metric is optimum for a given number of measurements, M. We define the transition point in the following.

Definition 1 (Transition Point): For a given basis totally ordered in non-increasing rearrangement-invariant form with respect to their spectral power, the transition point is defined as the basis index in a directed support set for which performance metric is optimum.

In many applications such as directional communication channel estimation, image processing, and continuous spectrum approximation, the optimum sparsity level n is unknown. More important, n can be significantly larger than the expected sparsity level. The example signals with continuous spectrum extension are studied in section VI. Nevertheless, the prediction of the optimum subspace dimension n is out of our scope in this paper and has been studied in [13].

B. CONTINUOUS SPECTRUM APPROXIMATION USING COMPRESSED SENSING

Compressed sensing, recommended by Donoho [1], is widely adopted in science and engineering to estimate finite-dimensional signals. An n-dimensional signal can be described as a signal in $\mathbb{K}^n$ that lies in a space $\mathbb{K}^d$, $d \gg n$. Nevertheless, in practical application, not only is n usually unknown priori, the sparsity level can extend toward d (d can be infinite), as discussed in section VI-D for the image signals. The problem discussed here is a part of a larger class of problems studied as continuous spectrum approximation in the linear operator theories [14], [15]. Obviously, compressed sensing and operator theory perfectly coincide if we notice that the random sampling matrix is, in fact, a linear operator.

The diagonalization of infinite-dimensional signals, also referred to as signals with continuous spectrum or continuous spectral extension during the paper, has practical importance in some applications such as image processing. In [16], finite-dimensional compressed sensing has been applied to the MRI. To decrease the point spread of the image (i.e., blurring), the authors have recommended leveraging the wavelet basis with finite support in the space-frequency joint domain and sampling the image in the Fourier domain. However, if the wavelet base is Haar basis, then its Fourier representation consists of the infinite number of basis (iteration).

The approximation of infinite-dimensional signals has practical importance in engineering (such as MRI, Computerized Tomography, and tomography in general) and physics (such as Helium Atom Scattering and Quantum mechanics in general). We will see that the proposed CS-KLE formulations through Mercer's theorem (in Theorem 7) and Green's function (in Theorem 11) are capable of approximating signals with continuous spectral extension while leveraging the undersampling operator.

III. CONTRIBUTION

Our contributions is as in the following:

1) Examining the stochastic orthogonality of KLE and random sampling operator, CS-KLE is proposed in Theorem 7. CS-KLE estimates the optimum subspace that maximizes spectral overlap between estimated and actual basis.
2) Using separable Hilbert space, the theoretical aspects of the CS-KLE, including compactness, trace-class positive semidefinite, and continuity of the kernel, are discussed. Accordingly, uniform convergence of CS-KLE relation considering Restricted Isometry Property (RIP) is proved.
3) We study CS and KLE in separable Hilbert space to derive the Compressed Sensing Hilbert-Schmidt integral operator. The condition for the Compressed Sensing Hilbert-Schmidt Operator to be an integral operator is studied. It is proved that the RIP condition guarantees the existence of an integral operator.
4) It is shown that the outermost integral of CS-KLE through Mercer's theorem enables combinatorial search. In fact, for a certain undersampling ratio $\delta \geq 0.6$, CS-KLE can estimate the whole range of $\rho = \frac{n}{M}$. Also, the outermost integral of CS-KLE enables continuous spectrum approximation.
5) The relation between the Compressed Sensing Hilbert-Schmidt operator and Green's function is examined. Considering the fact that the kernel of

the integral operator is the Green's function of the CS-KLE, CS-KLE through Green's function is formulated in the Theorem 11. Similar to its Mercer's counterpart, CS-KLE formulation via Green's function extends compressed sensing solution to the combinatorial search region in the theoretical ℓ_1 transition phase.

6) Continuous spectrum approximation is evaluated by applying CS-KLE to semi-infinite and infinite-dimensional signals. Accordingly, n-pseudospectrum and optimum subspace are measured by optimizing spectral overlap between estimated and actual signals.
7) Last but not least important, it is shown that the optimum subspace coincides with the transition point. For subspaces larger or smaller than the transition point, the accuracy of local and global models is sub-optimal.

IV. THEORETICAL DEVELOPMENT OF COMPRESSED SENSING KARHUNEN-LOÉVE EXPANSION

In this section, we derive CS-KLE. Theorems 7 and 6 represent our proposed approach to robustly estimate the underlying compressible signal $X(\omega, t)$ from a noisy measurement $Y(\omega, t)$.

In order to develop CS-KLE, we need to show the relationship between the KLE and compressed sensing. Sections IV-A and IV-B explain the basics of KLE in separable Hilbert space. In sections IV-D to IV-F, we formulate CS-KLE and will prove two main theorems of CS-KLE in Theorems 6 and 7. The existence of compressed sensing Kernel will be discussed in section IV-E. In section IV-F, we prove that it is impossible to formulate CS-KLE without violation of the stochastical orthogonality of KLE. Properties of the CS-KLE and proof of convergence with respect to Restricted Isometry Property (RIP) are presented in sections IV-F.

A. SEPARABLE HILBERT SPACE

Separable Hilbert space is essential to formulate KLE as a weighted combination of linear functional. Hilbert spaces that have a countable number of basis are said to be separable. Note that the separable Hilbert space can be infinitely countable or finitely countable. Both finitely and infinitely countable Hilbert spaces are of interest in this work.

Definition 2 (Separable Hilbert Space): Let $\mathscr{B}$ be the subset of Hilbert space H defined over scalar field $\mathbb{K}$. Then,

(a) $\mathscr{B} \subset H$ are the Hilbert space basis, if and only if, for every pair of basis $b_i, b_j \in \mathscr{B}$, then $\langle b_i, b_j \rangle = \delta_{i,j}$.
(b) $span\{\mathscr{B}\} = H$ is a closed linear hull of $\mathscr{B}$ that spans H.
(c) If $\mathscr{B} = \{i \in \mathbb{N} \mid b_i \in H\}$, then H is said to be separable Hilbert space.
(d) The system of orthogonal basis $\mathscr{B} = \{i \in \mathbb{N} \mid b_i \in H\}$ is called orthonormal basis of H if $\|b_i\|_2 = 1$. According to the Gram-Schmidt theorem, every separable Hilbert space has orthonormal basis and can be subsequently represented as a weighted linear combination of orthonormal basis.

In order to apply separable Hilbert space to compressed sensing, the following property has to be considered in addition to Definition 2.

Properties 1 (Separable Space for Compressed Sensing):

- *Let X be a finite-dimensional compressible random process in Hilbert space. Given n-dimensional subspace X_n of X such that X_n uniformly converges to X, the set of the orthonormal basis of X_n is a minimal orthonormal subset of a basis of X.*
- *Let X be N infinite-dimensional compressible random process in Hilbert space with a continuous spectrum extension. An n-pseudospectrum of X is defined as the discrete subset of a continuous spectrum of X that uniformly converges to the continuous spectrum of X.*

Theorem 7 and its convergence analysis require additional structures in $\|\cdot\|_{1,\infty}$ and $\|\cdot\|_1$. The appearance of these additional structures is justified by their necessity to formulate the compressed sensing problem in Hilbert Space using the compressed sensing integral operators in Theorem 6 and Theorem 7. The reader should note that these additional structures are inherited directly from the Banach space. Mainly, these additional structures are needed to clearly define the boundedness for the kernel function in Theorem 4 and 3.

Theorem 1 (Multiplication of ℓ_2 by ℓ_2): [17, Problem/Solution 29] For a given sequence α_n bounded in ℓ_2, i.e., $\sum_n |\alpha_n|^2 < \infty$, then $\sum_n |\alpha_n^* \beta_n| < \infty$ if

$$\sum_n |\beta_n|^2 < \infty \tag{1}$$

Theorem 1 indicates that $C(X_1, X_2)$, for $X_1, X_2 \in \ell_2$ exists, and it is bounded in L^1-space. The product space in L^1 theoretically is enough for the Carleman kernel. However, we assume that the $X^1 \otimes X^2 \in L^2$, a stronger condition than Theorem 1, is a valid assumption for many applications. The kernel in L^2-space is called the Hilbert-Schmidt kernel.

Corollary 1 (Extension of Theorem 1 to $\|\cdot\|_{1,\infty}$): If $\sum_n |\alpha_n|^2 < \infty$ and $\sum_n |\beta_n|^2 < \infty$, then

$$\text{ess sup}_n \sum_n |\alpha_n \beta_n| < \infty \tag{2}$$

According to Corollary 1, the product space of $X_1, X_2 \in L^2$ has bounded essential supremum in L^1-space. In other words, the kernel $C(X_1, X_2)$ is bounded by the essential supremum of the product of two ℓ_2 sequences.

B. BASIC DEFINITIONS

Definition 3 (Compact Operator): Let X and Y be separable Hilbert spaces. A bounded linear functional $\Lambda : X \to Y$ is compact if it maps bounded subsets $U \subset X$ onto precompact subsets in Y, that is, $\overline{\Lambda(U)} \subset Y$.

Definition 4 (Operator Norm): Let $\Lambda : X \to Y$ be a linear functional from Hilbert space $X \in H_1$ to Hilbert space

$Y \in H_2$. The operator norm is defined as

$$\|\Lambda\|_Y = \left\{\|\Lambda x\|_\infty | x \in X,\ \|x\|_2 \le 1\right\} \tag{3}$$

for all $x \in X$.

Theorem 2 (Direct Sum of Subspaces): [18, Theorem 18.1] Let $\{H^i\}_{i\in I}$ be a set of separable vector spaces in Hilbert spaces with a well-defined inner product. Then, H can be expressed as

$$H = \oplus_{i=1}^{n} H_i \tag{4}$$

where $\oplus$ denotes the internal direct sum.

Theorem 2 indicates that the internal and external direct sums are equal provided that H_1 and H_2 are disjoint, that is, $H_1 \bigcap H_2 = \emptyset$, and conditions of vector spaces are satisfied.

Theorem 3 (Tensor Product and Quotient Spaces): [18, Theorem 18.2] Let H_1 and H_2 be Hilbert spaces. Then, for every bilinear map $B : H_1 \times H_2 \to H_3$, there is a tensor product defined as linear mapping $l : (H_1 \otimes H_2) \to H_3$ such that

$$\begin{aligned} B\left(x_i, x_j\right) &= l\left(x_i \otimes x_j\right) \\ &= \Lambda\left(H_1, H_2\right) / \Lambda_0 \\ &= \Lambda\left(H_1, H_2\right) + \Lambda_0 \end{aligned} \tag{5}$$

where Λ_0 is the null space of linear functional Λ.

If the Λ_0 is nontrivial, the quotient space is not unique. However, discussed in [13], given an optimum $\Lambda\left(H_1, H_2\right)$, $B\left(x_i, x_j\right)$ is almost unique. Theorem 3 can be formulated coordinate-wise in product space as represented in the following theorem.

Theorem 4 Tensor Product of Subspaces and Compact Kernel: Let $\mathscr{B} \in L^2$-space be a set of orthonormal basis spanning Hilbert space H. Also, let H_1, H_2 be separable restricted representations of H such that $H_1 := X_1 \otimes \mathscr{B}_1$ and $H_2 := X_2 \otimes \mathscr{B}_2$ where $X_1 = \{x_i\}_{i\in I_1}$ and $X_2 = \{x_j\}_{j\in I_2}$ are sequences of compressible random processes in L^2-space such that $I_1 \cap I_2 = \emptyset$. The set of basis $\mathscr{B}_1$ and $\mathscr{B}_2$ are the subspace of $\mathscr{B}$ spanning H_1 and H_2. For $\{b^i\} \in \mathscr{B}_1$ and $\{b^j\} \in \mathscr{B}_2$, $i, j \in \mathbb{I}$, and for every column vectors $\{x_i\} \in X_1$, $\{x_j\} \in X_2$, kernel $C\left(X_1, X_2\right)$ is independent of orthonormal basis $\mathscr{B}_1$ and $\mathscr{B}_2$.

Proof:

$$\begin{aligned} C\left(X_1, X_2\right) &= \sum_{i=1}^{m}\sum_{j=1}^{m} \left(x_i \otimes b^i\right) \otimes \left(x_j \otimes b^j\right) \\ &\overset{a}{=} \sum_{i=1}^{m}\sum_{j=1}^{m} x_i b^{i^*} \otimes x_j b^{j^*} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} x_i b^{i^*} \left(x_j b^{j^*}\right)^* \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} x_i b^{i^*} b^j x_j^* \\ &\overset{b}{=} \sum_{i=1}^{m} x_i x_j^* = X_1 \otimes X_2 \end{aligned} \tag{6}$$

where (a) is due to $A \otimes B = \sum_i \sum_j a^i b^{j^*}$ for typical vector space A and B with vectors $a^i \in A$, and $b^j \in B$. And, (b) is because the inner product of the basis vector b^i and b^j defined as $\langle b^i, b^j\rangle = \delta_{ij}$ where δ_{ij} is the Kronecker delta. ■

Applying Theorems 2 and 3 to 4, $C\left(X_1, X_2\right)$ can be rephrased as a quotient space

$$C\left(X_1, X_2\right) = \Lambda\left(X_1, X_2\right) / \Lambda_0 \tag{7}$$

For compressible signals with continuous spectral extension where $j \lessapprox d$, the diagonalization through deterministic orthogonality is not accurate. In this case, the kernel can be expressed as

$$C\left(X_1, X_2\right) = \left(\Lambda\left(X_1, X_2\right) + \Xi\right) + \Lambda_0 \tag{8}$$

where Ξ represents the continuous spectral extension of the compressible random process. Obviously, failure to approximate Ξ contributes to an estimation error of the signal. The proper approximation of the Ξ provides continuous spectrum approximation, which is important in many applications, including MRI image processing and quantum operators.

While both Theorem 3 and 4 are equivalent, they provide different capabilities. Especially, Theorem 3 can be extended to n-dimensional Hilbert spaces as an n-fold tensor product

$$H = H_1 \otimes H_2 \cdots \otimes H_n \tag{9}$$

On the other hand, Theorem 4 implies that given orthonormal basis $\mathscr{B}$ and separable Hilbert spaces H_1 and H_2, the kernel only depends on X_1 and X_2. As a result, the decomposition of $C\left(X_1, X_2\right)$ provides information about the structure of the X_1 and X_2. The most important information that eigenfunctions of $C\left(X_1, X_2\right)$ reveals is about the distribution of X_2 given X_1, e.g., the coordinates of non-zero components. The decomposition of $C\left(X_1, X_2\right)$ can be obtained using Mercer's Theorem given in 5. This has great implications since $C\left(X_1, X_2\right)$ can be obtained as the covariance of X_1 and X_2 for two Hilbert spaces H_1 and H_2, and as an n-fold Tensor product for multi-dimensional signal $H_1, H_2, \cdots, H_n$.

Definition 5 (Continuity of Compact Kernel): A compact kernel is continuous on a given vector space. That is, compressible vector X is closed under addition as $x + x' \in X$, for every column vector $x, x' \in X$. X is also closed under scalar multiplication since $\alpha x \in X$, for $\alpha \in \mathbb{C}$. Note that compressed sensing is only isomorphic for the well-defined subspace of $\Gamma_S \subset \Gamma$. However, this is not a problem since CS-KLE in Theorem 7 is continuous inherently due to the continuity of the kernel.

Definition 6 (Trace Class Positive Definite Operator): Kernel function $C := X_1 \otimes X_2$ can be defined as linear functional $C : L^2\left(X_1 \times X_2\right)$ for $H_1 \in X_1 \otimes \mathscr{B}_{X_1}$, $H_2 \in X_2 \otimes \mathscr{B}_{X_2}$, and $\mathscr{B}_{X_1}, \mathscr{B}_{X_2} \subset \mathscr{B}$. The Hilbert-Schmidt kernel $C\left(X_1, X_2\right)$ satisfies the trace class positive semidefinite property with respect to the inner product and $b_i \in \mathscr{B}_{X_1}$ and $b'_j \in \mathscr{B}_{X_2}$

$$\mathrm{Tr}\{C\} = \sum_{i,j} \langle C b_i, b'_j\rangle < \infty \tag{10}$$

C. HILBERT-SCHMIDT OPERATOR AND PROPERTIES

Definition 7 (Hilbert-Schmidt Operator): Let X and X' be compact metric spaces with measurable spaces $(X, \mathcal{B}_X, \mu_X)$, and $\left(X', \mathcal{B}_{\mathcal{X}'}, \mu_{X'}\right)$. Provided continuous kernel $C : L^2\left(X \times X'\right)$, Hilbert-Schmidt operator $\mathbf{G} : L^2\left(\mathbb{C}^d\right) \rightarrow L^2\left(\mathbb{C}^d\right)$ is defined as

$$(\mathbf{G}\psi)\left(x'\right) = \int C\left(x, x'\right)\psi(x)\,d\mu(x) \tag{11}$$

where $\psi(x)$ is the eigenfunction of the kernel operator, and $d\mu(x)$ is increment measurement corresponding to Lebesgue measurement μ.

Linear functional $C\left(x, x'\right)$ is analogous to covariance function, and the Hilbert-Schmidt operator extends the covariance matrix to integral form using integral operator G. Th implication is that if one knows the integral operator $\mathbf{G}$ and the covariance matrix $C\left(x, x'\right)$, then the unknown state of a system $\psi\left(x'\right)$ can be estimated using known $\psi(x)$. The eigenfunction $\psi(x)$ plays a crucial role in finding KLE and Mercer's formulation of CS-KLE. The easiest method to obtain $\psi(x)$ is through Mercer's theorem.

Properties 2: Hilbert-Schmidt operator satisfies the following properties:

(a) $\mathbf{G}$ is linear
(b) $\mathbf{G}$ is positive semidefinite
(c) $\mathbf{G}$ is compact

Theorem 5 (Mercer's Theorem): Let $C : L^2\left(X \times X'\right)$ be a continuous, compact, and positive-semidefinite covariance function. Then, there is an infinite number of eigenpairs $\{\lambda_i, \psi_i(t)\}_{i\in\mathbb{N}}$ such that

$$C\left(x, x'\right) = \sum_{i=1}^{\infty}\lambda_i\psi_i(x)\,\psi_i^*\left(x'\right) \tag{12}$$

At the heart of the compressed sensing lies the random sampling, where the randomness is according to the few dominant random macrostates that a system takes from a universe space. As a result, $Y(\omega, t)$ is a separable random function represented as a product of deterministic eigenfunction $\psi(t)$ and stochastic random process $\gamma_i(\omega) \in \Gamma(\omega)$ for $i \in S$, where S is the support subset of $X(t, \omega)$. For the sake of simplicity, we drop the ω in the remaining of this letter.

$$Y(\omega, t) = \sum_{i\in S}\sqrt{\lambda}\psi(t)\,\gamma_i(\omega) \tag{13}$$

In the remaining of the paper, the Hilbert-Schmidt operator and Mercer's Theorem are rephrased according to $Y(\omega, t)$ and $Y\left(\omega', t'\right)$. Note that t and x represent some longitudinal variables hereafter. A measure space is also updated to $(Y, \mathcal{B}_Y, [0, 1])$, where $\mathcal{B}$ is a σ-Borel set, and $[0, 1]$ is the segment of the real line corresponding to weights of $\gamma_i(\omega)$, $i \in [1, d]$, also in S.

D. COMPRESSED SENSING HILBERT-SCHMIDT OPERATOR AND CS-KLE FORMULATION

Let $f = g + h$ be a compressible signal of interest with main compact signal space g and a continuous spectrum h. Then, the Hilbert-Schmidt operator needs not only to measure the subset of the main signal space g but also it has to measure the extension of the signal in the subset of Hilbert space basis that spans h. In other words, a potential candidate for compressed sensing Hilbert-Schmidt operator has to integrate over the orthonormal basis $\mathscr{B}$. The following theorem extends the Hilbert-Schmidt Operator to the *Compressed Sensing Hilbert-Schmidt Operator*.

Theorem 6 (Compressed Sensing Hilbert-Schmidt Operator): Let X and X' be compact metric spaces with measurable spaces $(X, \mathcal{B}_X, \mu_X)$, and $\left(X', \mathcal{B}_{X'}, \mu_{X'}\right)$. Also, let $\Gamma \in \mathbb{C}^{M\times d}$ be an underdetermined sampling matrix. Provided continuous kernel $C : L^2\left(X \times X'\right)$, compressed sensing Hilbert-Schmidt operator $\mathbf{G_c} : L^2\left(\mathbb{C}^d\right) \rightarrow L^2\left(\mathbb{C}^d\right)$ is obtained as

$$(\mathbf{G}_c\psi(\Gamma))\left(x'\right) = \int_{X^*}\int_X C\left(x, x'\right)\psi(x)\,d\mu(\gamma x) \tag{14}$$

for all $x \in X$ and $\gamma \in X^*$.

Proof: From averaging property of Kolmogorov conditional expectation

$$\mathbb{E}\left[\mathbb{E}^{\mathcal{F}}\zeta; A\right] = \mathbb{E}[\zeta; A], \zeta \in L^1, A \in F \tag{15}$$

where

$$\mathbb{E}[\zeta; A] = \mathbb{E}(\zeta\mathbf{1}_A) = \int_A \zeta\,d\mu(a) \tag{16}$$

Substituting (16) into the LHS of (15)

$$\begin{aligned}\mathbb{E}[\zeta; A] &= \mathbb{E}\left(\mathbb{E}^{\mathcal{F}}\zeta\mathbf{1}_A\right)\\ &= \int_A \mathbb{E}\left(\mathbb{E}^{\mathcal{F}}\zeta\mathbf{1}_A\right)d\mu(a)\end{aligned} \tag{17}$$

By defining evaluation map $\mathbb{E}^{\mathcal{F}} : L^1 \rightarrow L^1(\mathcal{F})$ acting on random process ζ as

$$\mathbb{E}^{\mathcal{F}}\zeta = \int_X C\left(x, x'\right)\psi(x)\,d\mu(x) \tag{18}$$

such that $\zeta := \psi(x)$, $\mathcal{F} := C\left(x, x'\right)$, and $\mathcal{F} \subset X^*$ is a measurement on σ-algebra, where X^* is a dual space of X,[1].[2] By substituting (18) into (17) and applying Hilbert-Schmidt operator

$$\begin{aligned}\mathbb{E}\left[\psi\left(x'\right); \Gamma\right] &= \mathbb{E}\left[\mathbb{E}^{\mathcal{F}}\psi\left(x'\right); \Gamma\right]\\ &= \int_\Gamma\int_X C\left(x, x'\right)\psi(x)\,d\mu(x)\,d\mu(\gamma)\\ &\overset{a}{=} \int_\Gamma\int_X C\left(x, x'\right)\psi(x)\,d\mu(\gamma x)\end{aligned} \tag{19}$$

where a is due to Fubini-Tonelli's theorem. And, this ends the proof. ■

If sampling matrix Γ qualifies RIP condition, then compressed sensing Hilbert-Schmidt operator exists with high probability. Also, Theorem 6 satisfies the Fubini-Tonelli

[1] X^* is a vector space with continuous linear functional elements acting on vector space X.
[2] $\mathcal{F} : \Gamma X \times \Gamma X' \rightarrow C_Y \in L(\Gamma)$.

theorem [19, Theorem 1.27], if and only if, the sampling operator Γ satisfies completeness and absolute integrability conditions [20, Remark 1.7.20].

Remark 1 (Extension to Beyond Combinatorial Search): The operator $\mathbf{G}_c$ formulation in Definition 6 predicts important property. The compressed sensing Hilbert Schmidt operator $\mathbf{G}_c$ expands to the combinatorial search region beyond the theoretical ℓ_1 transition phase.

Remark 2 Universality of Solution: Universal compressed sensing aims to reconstruct the sparse vector from under-sampled measurement $\bar{y}$ without a priori knowledge about the structure of the underlying sparse vector. Here, the structure of the compressible signal is determined mainly by the probability distribution law that governs the compressibility. The most important consequence of such a structure is the compressibility level of the signal. However, the compressibility level is vague since it follows a power law pattern in the spectral domain and does not follow the non-zero support definition commonly considered in the compressed sensing literature [21, Chapter 2]. The support subset of a compressible signal is obtained through ad-hoc hard thresholding. In general, hard thresholding can change the spectral structure of a signal by enforcing significant entries to zero. In other words, the ad-hoc threshold level may not reveal the optimum subspace. Considering averaging property, compressed sensing Hilbert Schmidt operator is capable of estimating the optimum subspace provided that the kernel exists. The existence of a kernel, in general, depends on the random mask used for sampling. Considering that the kernel is almost zero everywhere, the concentration of measurement needs to be considered somehow in the construction of a mask. This is proved in the next section while studying the existence of a kernel for compressed sensing Hilbert-Schmidt. Considering the fact that the outermost integral of the Compressed Sensing Hilbert-Schmidt operator enables to search the whole dual space X^* and reconstruct signal space X holistically.

Considering KLE signal reconstruction, we propose CS-KLE as a compressed sensing reconstruction algorithm in the following theorem. Note that unlike KLE, which estimates the measurement vector $\bar{y}$, CS-KLE formulation in the following estimates the non-decreasing rearrangement-invariant form of the underlying sparse signal.

Theorem 7 (CS-KLE Relation): Let $Y(\omega, t) = \Gamma(\omega) X(t) + e$ be a noisy measurement of compressible random process $X \in L^2(\mathbb{C}^d)$ filtered by the sampling process $\Gamma(\omega) \in L^2(\mathbb{C}^{m\times d})$, where $\omega \in \Omega$ is a random process path $\omega \to \Gamma(\omega)$. Also, let $C_Y(t, t') \in L^2(Y \times Y')$ be a continuous, compact, and positive-semidefinite covariance matrix of compressed sensing measurement vector Y. Then, CS-KLE relation is defined as

$$\operatorname{ess\,sup}_{\gamma_i(\omega)} \sum_{i\in[1,d]} \left|\sqrt{\lambda}\psi(t)\gamma_i(\omega)\right| \longmapsto \left\|X(t)\right\|_{1,\infty} \tag{20}$$

for all $\gamma_i \in \Gamma$. $\|X(t)\|_{1,\infty}$ denotes the non-increasing rearrangement-invariant of X.

Proof: Considering RIP and imperfect stochastical orthogonal property, i.e., mutual coherency, we prove that the convergence of Theorem 7 to ℓ_2-error of best s-term approximation, $\sigma_s(X)_2$.

$$\left\|X(t)\right\|_{1,\infty} \underbrace{\left\|\mathbf{I} - \Gamma_S^*(\omega)\Gamma_S(\omega)\right\|_{1,\infty}}_{\geq \left\|\mathbf{I}\right\|_{1,\infty} - \left\|\Gamma_S^*(\omega)\Gamma_S(\omega)\right\|_{1,\infty}} \leq \left\|\sum_{i\in S}\sqrt{\lambda}\psi(t)\gamma_i(\omega)\right\|_{1,\infty} \tag{21}$$

Inequality (21) can be rewritten as

$$\begin{aligned}
&\left|\left\|X(t)\right\|_{1,\infty} - \left\|\sum_{i\in S}\sqrt{\lambda}\psi(t)\gamma_i(\omega)\right\|_{1,\infty}\right| \\
&\quad\leq \left\|\Gamma_S^*(\omega)\Gamma_S(\omega)\right\|_{\infty}\left\|X(t)\right\|_{1,\infty} + \eta_Y \\
&\quad\leq C_2\mu\left\|X(t)\right\|_{1,\infty} + \eta_Y \\
&\quad\overset{a}{\leq} C_2\mu + \eta_Y
\end{aligned} \tag{22}$$

for some constant $C_2 \geq 0$, η_Y is least-square estimation error, and (a) is because of compactness or $\left\|X(t)\right\|_{\infty} \leq 1$. And this proves Theorem 7. ■

Theorem 7 integrates random sampling of compressed sensing with KLE random weights, which describes the stochastic behavior of the random process, to find the support subset of X.Intuitively, Theorem 7 can be described using Theorem (2), where the left-hand side can be written in the form of the direct sum of subspaces. On the left-hand side, the coordinates of pivots in eigenfunctions coincide with the dominant components of X^*. The random sampling matrix evaluates the eigenfunction coordinate-wise to uncover the location of pivots. As it has been proved in [13], this observation is an example of the Schwartz class test functions.

E. EXISTENCE OF KERNEL FOR COMPRESSED SENSING HILBERT-SCHMIDT INTEGRAL OPERATOR

An important question is if the proper orthogonal decomposition formulation in the previous section can be used for compressed sensing. More accurate questions are (1) if there is a kernel for compressed sensing Hilbert-Schmidt operator? And (2) if the answer is YES, what are the conditions for the compressed sensing operator to have a kernel? These questions are answered in this section. As mentioned in [17, Section 173], it is not easy to compute integral operator $\mathbf{G}_c$ from a given kernel matrix in general. Then, related to the questions above, the following statements are equivalent

1) The compressed sensing sampling operator has no kernel.
2) The compressed sensing sampling operator is not an integral operator.
3) The compressed sensing sampling operator does not solve the problem.

The answer to the first question is YES with some probability. The condition for the compressed sensing sampling operator to be an integral operator is the same as RIP. As it is described in the following, the existence of a kernel for the compressed sensing sampling operator is probabilistic and is a function of the undersampling ratio δ and random mask concentration (for continuous spectrum scenarios). In compressed sensing, one aims to find an optimum pseudospectrum subset $\mathcal{F}$ in $\mathfrak{X}$ using undersampled operator Γ. However, it is known that the identity operator obtained for an interval is not an integral operator in general. For convenience, we prove the idea in the following theorem.

Theorem 8 (Identity and Integral Operators): [22, Theorem 8.5] The identity operator on interval $L^2(t, 1)$ does not admit integral operator.

Theorem 8 indicates that the compressed sensing sampling operator measured that collects samples from a *certain interval* with a high probability does not have a kernel. However, for a properly randomized operator that covers all the $L^2(X)$ and also a sufficient number of measurements, the compressed-sensing integral operator tends to have a kernel. The randomness is the crucial condition to avoid the condition $f \in L^2(1, t)$. The condition is well-qualified for $f(x) = \psi(x)$, where $\psi(x)$ is the eigenfunction obtained from Mercer's Theorem 5. For a given $\psi(x)$, the random location of the pivots along $\psi(x)$ prevents the concentration of $\psi(x)$ in a certain region $[t, 1]$ in general. However, if the signal of interest has a continuous extension in the spectral domain, i.e., in the dual space X^*, then one should properly select a mask to collect enough samples from continuous extensions. As a result, *a sufficient number of measurements* M_S, either uniformly at random or using a random mask, generates random entries at custom from the *whole* $L^2(X')$ domain. Then, the integral operator has a kernel with respect to the coordinates of the pivots and the Borel σ-algebra. The existence of the integral operator entangled with the pivots to be in the correct coordinates in $\psi(x')$. As a result, for a sufficient number of measurements, the integral operator can recover the underlying pivots almost surely.

The sufficient number of measurements is not a quantitative criterion and can be understood twofold: (1) for a given sparsity level ρ, there is a sufficient number of measurements that provides optimum subspace, equivalent to minimal linear functional [13], and (2) the optimum subspace is the function of the number of measurement, i.e., we are only able to recover certain subsets of an actual support subset given the undersampling ration δ. Obviously, this is equivalent to the transition phase formulation proposed by Donoho *et al.* [23].

Naturally, the question that arises here is the probability of compressed sensing sampling operator to have a kernel for given undersampling ratio δ. In the compressed sensing literature, including [21, Chapters 7, and 8] [24], the probability of success is derived from the Bernstein inequality

$$Pr\left(\left|\sum_{l=0}^{M-1} \epsilon_l a_l\right| > u\|a\|_2\right) < \frac{1}{1-\delta} e^{-u^2\delta} \tag{23}$$

where $a \in L^2\left(\mathbb{K}^d\right)$ is an estimated sequence, $\{\epsilon_l\}$ for all $l \in [0, M-1]$ is Steinhaus sequence, $0 < \delta < 1$, $u > 0$. Here, we should be careful about two pitfalls using (25). First, the left-hand side of the Bernstein inequality expresses the probability, while the right-hand side is a more analytic expression since it is a function of the fixed δ and u. The value of u depends on the underlying problem and should be chosen such that it leads to probabilistic results rather than a certain outcome. For this, δ and u should be determined asymptotically for a sufficiently large number of trials. However, after an even sufficiently large number of trials, it may be difficult to determine the proper value of u. Second, we should carefully define the quality for which the probability is measured. The probability is assigned for the compressed sensing sampling operator to be an integral operator. Due to ambiguity in the definition of compressibility, such a quality is not a binary decision, but it is best characterized by a probability of the reconstruction error. In the following, δ denotes undersampling ratio, and u is the mean square error measured for orthogonal decomposition as $\left|X_n^\perp\right|^2 = |X - X_n|^2$, where $X_n^\perp \subset X$ is the orthogonal complement subspace of X_n. The next two theorems prove the existence of a sufficient number of measurements M_S.

Theorem 9 (Integral Operator for $\delta \to 0$): Let $X = X_n + X_n^\perp$ be an optimum orthogonal decomposition. The probability of the Compressed Sensing Hilbert-Schmidt operator to have a kernel approaches zeros as $\delta \to 0$. Equivalently

$$Pr\left(\|X_n^\perp\| \le u \middle| M, \delta \to 0\right) \approx 0 \tag{24}$$

Proof: Let rephrase (24) as

$$Pr\left(\|X_n^\perp\| \le u \middle| M, \delta \to 0\right) = 1 - Pr\left(\|X_n^\perp\| > u \middle| M, \delta \to 0\right) \tag{25}$$

The second statement on the right-hand side can be obtained using conditional Benstein inequality

$$Pr\left(\|X_n^\perp\| > u \middle| M, \delta \to 0\right) < \frac{1}{1-\delta} e^{-u^2\delta}, u > 0 \tag{26}$$

Since $\lim_{\delta \to 0} \frac{1}{1-\delta} e^{-u^2\delta} = 1$,

$$Pr\left(\|X_n^\perp\| \le u \middle| M, \delta \to 0\right) = 1 - 1 = 0 \tag{27}$$

■

Theorem 10 Integral Operator for $\delta \to 1$: As δ approaches 1, the probability that the Compressed Sensing Hilbert-Schmidt operator to have a kernel approaches one.

Proof: To prove that the compressed sensing integral operator has a kernel with probability one for $\delta \to 1$, we need to show that the following inequality is satisfied for some u

$$Pr\left(\|X_n^\perp\| < 0.1 \middle| M_S, \delta \to 1\right) \to 1 \tag{28}$$

For $\delta \rightarrow 1$, the limit of the right-hand side in (25) can be found using L'Hôpital's theorem as

$$\lim_{\delta\to 1} \frac{1}{1-\delta} e^{-u^2\delta} = \lim_{\delta\to 1} u^2 e^{-u^2\delta} = u^2 e^{-u^2} \tag{29}$$

and substituting $u = 0.1$, the probability that the compressed sensing integral operator to have kernel is obtained as

$$Pr\left(\|\Xi\| < 0.1 \middle| M_S, \delta \to 1 \right) = 1 - 0.1^2 e^{-0.1^2} \approx 1 \tag{30}$$

And this proves the theorem. ■

F. CS-KLE PROPERTIES AND CONVERGENCE WITH RESPECT TO RIP

Corollary 2 (CS-KLE and Doubly Orthogonal Property): Let X be a compressible random process with continuous spectral extension. Considering CS-KLE, deterministic orthogonality is guaranteed via mutually orthogonal eigenfunctions. However, doubly orthogonality is disrupted slightly by the violation of stochastic orthogonality as

$$C\left(Y, Y'\right) = X\Gamma\left(\omega\right)^* \Gamma\left(\omega\right) X' + X_\Xi + \mathcal{E} \tag{31}$$

Proof:

$$\begin{aligned} C\left(Y, Y'\right) &= \sum_{i=1}^{d}\sum_{j=1}^{d} \left(x_i \otimes \gamma_i\right) \otimes \left(x_j' \otimes \gamma_j\right) \\ &= \sum_{i=1}^{d}\sum_{j=1}^{d} x_i \gamma_i^* \gamma_j x_j'^* \\ &= X\Gamma^*\left(\omega\right)\Gamma\left(\omega\right) X' + X_\Xi + \mathcal{E} \end{aligned} \tag{32}$$

■

From Theorem 3 and 2, one can conclude that Theorem 7 represents a Hilbert space H as a normal and completely continuous operator, $\Gamma = D + \Xi + \mathcal{N}(\Gamma)$ (analogous to $X = X_D + X_\Xi + \mathcal{E}$). Here, D is a diagonal operator, Ξ is a completely continuous spectral extension operator, and $\mathcal{N}(\Gamma)$ is the null space corresponding to $\|\mathcal{E}\| < \epsilon$ for sufficiently small $\epsilon \in \mathbb{R}^+ \cup \{0\}$. This is equivalent to the optimum orthogonal decomposition where the left-hand side of (7) converges in norm (also called quadratic norm [8, Chapter X]) to $X(t)$ such that

$$\begin{aligned} \|\mathcal{E}\| &= \left\| \sup_{\gamma_i(\omega)} \sum_{i\in[d]} \left|\sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right)\right| - \left\|X\left(t\right)\right\|_{1,\infty} \right\| \\ &= \mathbb{E}\left\{ \left| \sup_{\gamma_i(\omega)} \sum_{i\in[d]} \left|\sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right)\right| - \left\|X\left(t\right)\right\|_{1,\infty} \right|^2 \right\} < \epsilon \end{aligned} \tag{33}$$

The direct consequence of Corollary 2 is the violation of the separable Hilbert space assumption. It is already known that as the dictionary gets larger, it becomes more difficult for the sparse recovery algorithm to distinguish nontrivial supports from the surrounding trivial basis in the given neighborhood of the bases. In addition, (32) illustrates the direct impact of the mutual coherency on the covariance matrix as the off-diagonal components are generated because of the violation of stochastical orthogonality. *Note how the Compressed Sensing Hilbert-Schmidt kernel depends on the basis in comparison to ideal incoherency in Theorem 4.*

Properties 3: (Convergence Properties) Considering the imperfect doubly orthogonal property, the convergence of CS-KLE is characterized by Restricted Isometry Property (RIP) and quadratic convergence. For the estimated support set S

(a) RIP

$$\left\|\Gamma_S^*\Gamma_S - I\right\|_{2\to 2} \le \mu_1\left(s-1\right), \text{for} \mu_1\left(s-1\right) < 1 \tag{34}$$

(b) Quadratic convergence

$$\begin{aligned} &\mathbb{E}\left\{ \left| Y\left(\omega\prime, t\right) - \sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \right|^2 \right\} \\ &\qquad \le C_2\left(1 - \frac{\delta_s}{s-1}\right)\lambda \end{aligned} \tag{35}$$

Proof: (a) characterizes separability of (20). It can be proved with respect to RIP and mutual coherency as $\|\Gamma_S^*\Gamma_S - \mathbf{I}\|_{2\to 2} \le \delta_s$ and $\delta_s \le \mu_1(s-1)$ [21, Chapter 6].

(b) Considering convergence in quadratic mean

$$\begin{aligned} &\mathbb{E}\left\{ \left| Y\left(\omega, t\right) - \sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \right|^2 \right\} \\ &= \mathbb{E}\{|Y\left(\omega, t\right)|^2\} - 2\mathbb{E}\left\{ Y\left(t, \omega\right) \sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right)^* \right\} \\ &\quad + \mathbb{E}\left\{ \left| \sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \right|^2 \right\} \\ &= tr\{C\left(t, t'\right)\} \\ &\quad -2\mathbb{E}\left\{ \sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \left(\sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \right)^* \right\} \\ &\quad +\mathbb{E}\left\{ \sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \left(\sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \right)^* \right\} \\ &= tr\{C\left(t, t'\right)\} \\ &\quad -\mathbb{E}\left\{ \sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \left(\sum_{i\in S} \sqrt{\lambda}\psi\left(t\right)\gamma_i\left(\omega\right) \right)^* \right\} \\ &= \lambda - \mathbb{E}\left\{ \sum_{i\in S} \lambda\psi\left(t\right)\gamma_i\left(\omega\right)\gamma_i\left(\omega\right)^*\psi\left(t\right)^* \right\} \\ &= \lambda - \mathbb{E}\left\{ \lambda\psi\left(t\right)\psi\left(t\right)^* \right\} \sum_{i\in S} \gamma_i\left(\omega\right)\gamma_i\left(\omega\right)^* \\ &= \lambda - \lambda\mathbb{E}\left\{ \psi\left(t\right)\psi\left(t\right)^* \right\} \Gamma_S^*\Gamma_S \\ &= \left(\mathbf{I} - \Gamma_S^*\Gamma_S\right)\lambda \\ &\le C_1\left(1-\mu\right)\lambda, \quad C_1 \ge 1 \end{aligned} \tag{36}$$

■

V. CS-KLE AND GREEN's THEOREM

Theorem 11 (Continuous Comrpessed Sensing and Green's Function): The kernel $C(t,t')$ is the Green's function of the problem formulated in Theorem 7. Then, the following properties have been satisfied.

$$\int_t C_Y(t,t')\, X(\omega,t)\, dt \longmapsto X(\omega,t') \tag{37}$$

$$\int_t C_Y(t,t')\, X(\omega,t')\, dt' \longmapsto X(\omega,t) \tag{38}$$

Definition 8 (Green's Theorem): Let $Lu\psi(t) = f$ be a linear algebra problem. By solving this equation, one aims to find an unknown vector $\psi(t)$. Let L^* be the adjoint operator of L. There is a function $G(t,t')$ such that $L^*G(t-t') = \delta(t-t')$. The Green's function $G(t,t')$ satisfies

$$\psi(t') = \int \psi(\omega,t)\, G(t,t')\, dt \tag{39}$$

Now we can prove Theorem 11.

Proof: By Theorem 7

$$\int \Gamma(\omega)\, \psi(\omega,t)\, dw \longmapsto X(\omega,t') \tag{40}$$

Substituting $\psi(t',\omega)$ by Hilbert-Schmidt operator

$$\int \Gamma(\omega)\left(C_Y(t,t')\, \psi(\omega,t')\, dt\right) d\omega \longmapsto X(\omega,t') \tag{41}$$

By applying Fubini's Theorem

$$\int C_Y(t,t')\left(\Gamma(\omega)\, \psi(\omega,t')\, d\omega\right) dt \longmapsto X(\omega,t') \tag{42}$$

Substituting (20) into (42)

$$\int C_Y(t,t')\, X(\omega,t')\, dt' \longmapsto X(\omega,t) \tag{43}$$

By comparing (43) and (39), one notices that $C_Y(t,t')$ is the Green's function of the (43). The eigenfunction $\psi(\omega,t')$ is derived from the covariance function of $Y(\omega,t)$. $\psi(\omega,t)$ can be generalized using the chain rule as a function of a random sampling process $\Xi(\omega) : \omega \rightarrow X(\omega,t)$ as $\psi(t,\Xi(\omega))$. Then, we obtain

$$\int \Gamma(\omega)\, \psi(t,\Xi(\omega))\, d\omega \longmapsto X(\omega,t') \tag{44}$$

$$\int \Gamma(\omega)\, (\psi(t)\circ\Xi)(\omega)\, d\omega \longmapsto X(\omega,t') \tag{45}$$

$$\int (\Gamma(\omega)\, \psi(t))\, \Xi(\omega)\, d\omega \longmapsto X(\omega,t') \tag{46}$$

where $\circ$ is for function composition. (46) can be written as

$$\int X(t,\omega)\left(\int C_Y(t,t')\, e^{-j2\pi\omega' t'} dt'\right) d\omega \longmapsto X(\omega',t) \tag{47}$$

■

Theorem 11 has two main advantages: (1) CS-KLE formulation in Theorem 6 relies on Mercer's theorem. The computation of eigenfunctions using Mercer's theorem requires eigendecomposition of the kernel that may cause a memory bottleneck for large data. Also, from the computational point of view, the outermost integral in (47) *breaks the integration into finitely many Fourier transforms with fast and parallel implementation possibilities. And (2) the outermost integral in* (47) *guarantees continuous spectrum approximation using a smaller undersampling ratio δ.*

VI. NUMERICAL RESULTS

A. TRANSITION PHASE – CS-KLE VS. LASSO

Fig. 2 shows the transition phase for CS-KLE for noiseless and noisy measurements with $SNR = \{\infty, 15, 3\}$dB. Fig. 2a shows that the CS-KLE divides the error regions almost vertically as a function of undersampling ratio δ. In other words, the performance of CS-KLE becomes almost independent of the ρ, and only it depends on the undersampling rate δ. This behavior is also observed for different ranges of *SNR*. For comparison, the theoretical ℓ_1 transition phase for LASSO has shown in Fig. 2d with a median error between 0.3 and 0.4 at $SNR = 15$dB. Obviously, the median error of LASSO is a function of both δ and ρ, since there is no a certain δ for which the LASSO error estimation is constant for the whole range of sparsity ρ.

From Fig. 2b at $SNR = 15$dB, it is obvious that as $\delta \rightarrow 0.7$, the transition phase has an error performance approximately the same as LASSO with the advantage that CS-KLE acts independently of sparsity level ρ. For $\delta \gtrapprox 0.8$, the median error drops to approximately 0.3. We conclude that the optimum undersampling ratio with respect to the Nyquist rate lies about $0.6 \lessapprox \delta$, for which CS-KLE converges for all possible values of ρ with an error less than 0.5. As the *SNR* decreases further to 3dB in Fig. 2c, the median error of 0.4 can be achieved if $\delta \gtrapprox 0.8$.

We end this section with the following conclusion. *For a large range of SNR, there is a reasonable range of undersampling ratio δ where CS-KLE can solve compressed sensing problems for the almost whole range of sparsity levels ρ up to the median error of* 0.4.

B. CONTINUOUS SPECTRUM APPROXIMATION OF INFINITE-DIMENSIONAL SIGNAL

This section evaluates the reconstruction of the infinite-dimensional signal in (48) with main signal space g and continuous spectral extensions h_1 and h_2.

$$\begin{aligned} f(t) &= g + h_1 + h_2 \\ &= \sin(2\pi f t) + h_1 + h_2 \end{aligned} \tag{48}$$

In (48), $d = +\infty$, $f = 2$ is the fundamental frequency. h_1, h_2 are the first and second infinite-dimensional local fluctuations with amplitudes of $A_1 = 0.15$ and $A_2 = 0.25$, respectively.

Fig. 3a shows the signal in (48) with a finite-dimensional spectrum and continuous spectral extensions of h_1 and h_2 in Fig. 3c-3d, respectively. Obviously, $d \rightarrow \infty$. Fig. 4 shows the approximated spectrum and reconstructed signal

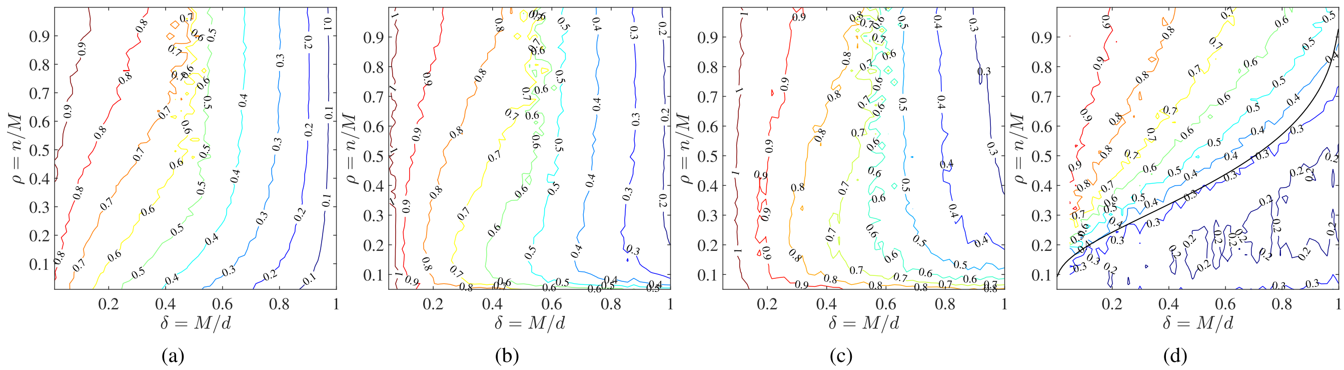

FIGURE 2. Median NMSE, (a) noiseless measurements, infinite *SNR*, (b) noisy data, *SNR* = 15dB, (c) noisy data, *SNR* = 3dB, (d) LASSO, *SNR* = 15dB.

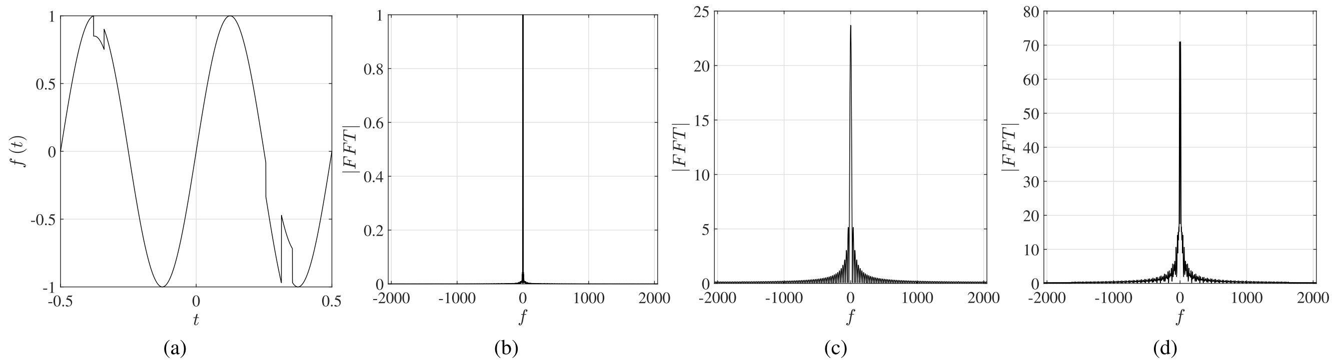

FIGURE 3. Infinite-dimensional signal, (a) Original signal $f(t)$, (b) infinite-dimensional continuous spectrum of $f(t)$, (c) infinite-dimensional continuous spectrum of $h_1(t)$, (d) infinite-dimensional continuous spectrum of $h_2(t)$.

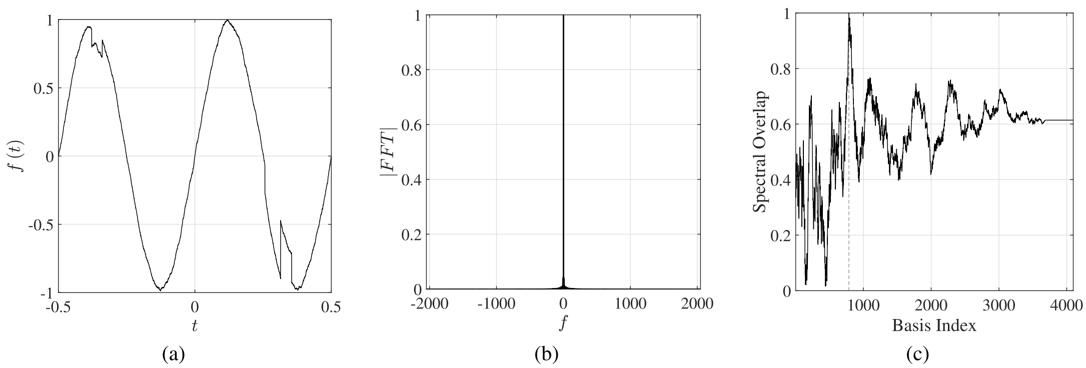

FIGURE 4. CS-KLE reconstruction using Mercer's theorem (a) reconstructed semi-infinite-dimensional signal, (b) reconstructed semi-infinite-dimensional continuous spectrum, (c) index-wise spectral overlap with maximum occurs at $n = 786$.

for $d = 4096$ and undersampling rate of $\delta = 0.9$. As shown in Fig. 4c, the reconstruction process is slow convergence due to the continuous spectrum. The reconstructed signal using CS-KLE via Mercer's theorem is shown in Fig. 4a.

In [13], the authors have proved that in order to approximate continuous spectrum, it is sufficient to approximate the set of n pseudospectrum. As shown in Fig. 4c, the maximum spectral overlap of 0.999 has been achieved for the n-pseudospectrum with $n = 786$ components. Accordingly, a mean square error of 0.0127 is obtained.

C. COMPARISON WITH GENERALIZED SAMPLING COMPRESSED SENSING

This section compares CS-KLE with the continuous compressed sensing formulation recommended using the Generalized Sampling theorem [25, Section 7]. The GS-CS approach relies on the perfect knowledge about the distribution of the original signal space g and the continuous spectral extensions. This is reflected in the set of measurements $f \in \overline{span}\{\varphi_j\}_{j\in\mathbb{B}}$ [25, Section 4.3 - Equation 4.9 and Equation 4.8], where f contains the exact local property

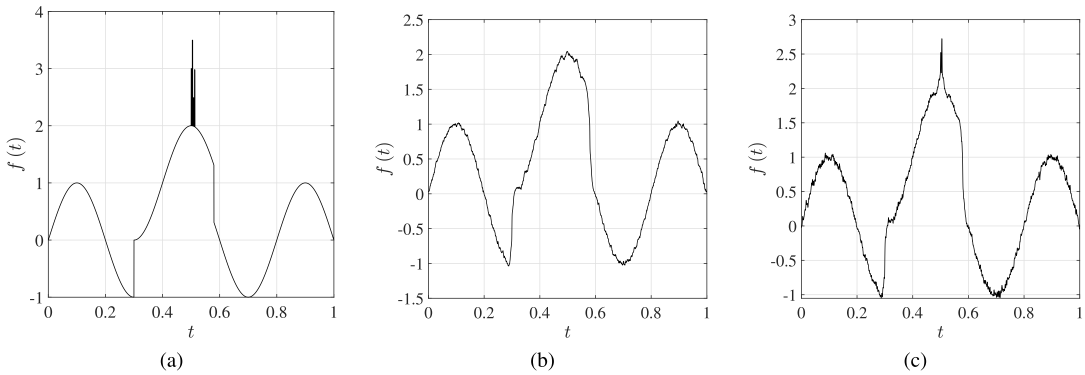

FIGURE 5. **GS-CS failure due to lack of information about the distribution of data. (a) Original signal space, reproduced from [26] (b) reconstructed signal using compressed sensing measurement vector, which lacks information about local perturbation,, reproduced from [26] (c) reconstructed signal using compressed sensing measurement vector contains information about two out of four perturbations.**

of the continuous spectral extension. This is equivalent to saying that the compressed sensing measurement vectors are constructed with a perfect priori knowledge about the local distribution of the signal, i.e., it contains all the information required for the perfect recovery of the signal. We examine two scenarios in that GS-CS collapses due to partial priori knowledge about the distribution of the signal.

The first scenario shown in Fig. 5 occurs when the compressed sensing measurement vector does not contain perfect information about the continuous spectral extension. Two examples are given. First, when the knowledge about the local distribution of the continuous spectral extension is absent from the measurement vector, as shown in Fig. 5b. Obviously, the original signal space g has been recovered properly; however, the local perturbations h are completely ignored. This is due to the fact that the approximated spectrum lacks the spectrum of the continuous extension portion corresponding to the h. In the second example, it is assumed that the compressed sensing measurement vector contains only partial information about the local perturbations, e.g., two out of four local spikes. Then, as it is observed in Fig. 5c, the recovered signal recovers two of the local perturbations.

The second scenario occurs when the compressed sensing measurement vector contains information about the fake local perturbation. As shown in Fig. 6b, the GS-CS predicts local perturbation in recovered signals, which do not exist in the actual signal space given in Fig. 6a.

The two scenarios discussed above show that GS-CS only is feasible if one has perfect knowledge of the global and local distribution of the signal. Compared to GS-CS, CS-KLE approximates the continuous spectrum of an infinite-dimensional signal with proper local perturbation without a priori knowledge about the distribution of the spectrum of a signal. Fig. 7 compares GC-SC with CS-KLE. GS-CS leverages the perfect knowledge about the distribution of the signal, but CS-KLE samples DFT basis to generate sampling matrix. Also, it is obvious that CS-KLE reconstructs the infinite-dimensional signal more accurately. *We conclude this section with a claim that the CS-KLE approach provides a universal solution for compressed sensing in the sense that it does not require priori knowledge about the local and global distributions of the signal.*

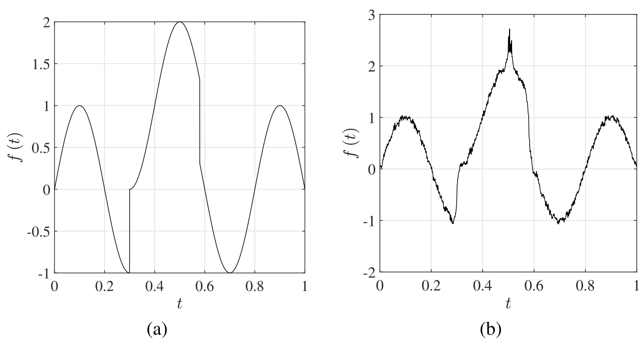

FIGURE 6. **GS-CS failure due to non-existence distribution, reproduced from [26]. (a) Original signal space, reproduced from [26], (b) reconstructed signal using compressed sensing measurement vector with non-existence perturbation.**

D. IMAGE RECONSTRUCTION

As discussed in section II, images are the examples of the signals that are not sparse in the Fourier domain. In image processing, sampling is performed not uniformly at random but randomly with respect to the weights that are determined by a certain law. While the mask used to generate the sampling operator itself is not unique, the requirements to design an effective mask are unique. In order to reconstruct MRI images with sufficient resolution, one needs to generate a mask that samples lower frequency components heavier than the higher frequency components. Obviously, the deviation from the uniform random sampling increases the cost of measurement and reconstruction. Since high-frequency components lie outward in the K-space, one can generate a mask that puts a larger weight sampling the inward frequency components.

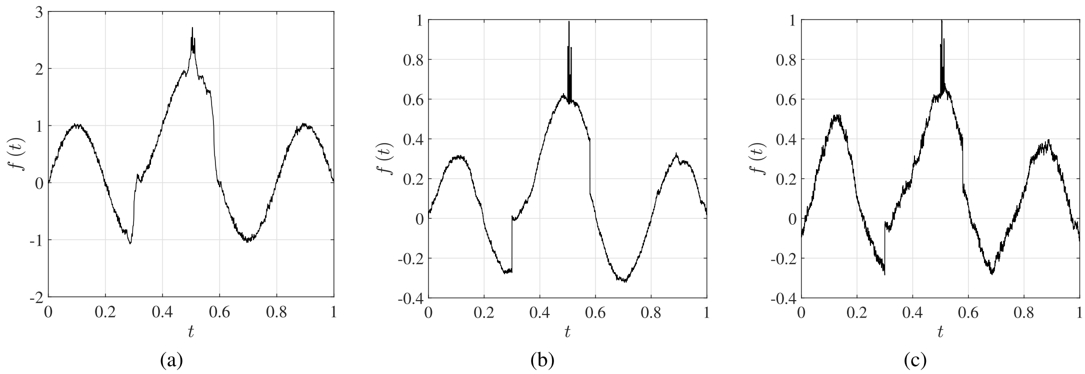

FIGURE 7. Comparing the performance of GS-CS and CS-KLE reconstructing signal with continuous spectral extension. (a) GS-CS, reproduced from [26], (b) CS-KLE through Mercer's theorem, (c) CS-KLE through Green's function.

FIGURE 8. MRI reconstruction using CS-KLE. (a) original 160 × 160 image, (b) reconstructed image, (c) PSF and transition point detect at $n = 21482$.

This section studies the degree of continuous spectral extension of the images in the Fourier domain. In particular, we are interested to see what is the range of n in n-pseudospectral set. Fig. 8c shows the Point Spread Function, $PSF = \phi^*\psi^*\phi\psi$, where ϕ is the Fourier transform of the original image in 8a and ψ is the Fourier transform of the CS-KLE reconstruction in Fig. 8b using CS-KLE. The original image has 160×160 pixels. The sampling matrix is generated using a dictionary matrix with a Fourier basis of size $160 * 160$ and an undersampling ratio of $\delta = 0.5$. Obviously, PSF, which measures the spectral overlap between the original image and reconstructed image, saturates at $PSF = 0.8$ for a certain index $n = 21482$ in the directed support set. However, the continuous spectral extension continues till the $n = 25594$ where the maximum is obtained. This indicates that the image in Fig. 8a has an infinite-dimensional continuous spectrum that occupies 0.9998% of its full spectrum in the Fourier domain.

VII. DISCUSSION AND CONCLUSION

This work proposed CS-KLE algorithm as a universal continuous compressed sensing scheme for a wide range of applications. Independent of the application, the numerical results indicate that CS-KLE algorithm guarantees to reconstruct the compressible signal with a high probability for the $0.5 \lessapprox \delta \lessapprox 0.9$. For the signals with both Gaussian and non-Gaussian distributions, the $0.5 \lessapprox \delta \lessapprox 0.9$ provides the optimum subspace that optimizes spectral overlap between the reconstructed and the actual signals. Finally, the proposed algorithm can be applied when the sparsity level of the signal is unknown, since for almost every possible sparsity level ρ by setting measurement numbers to satisfy $0.5 \lessapprox \delta \lessapprox 0.8$.

We observed that both formulations of the CS-KLE through Mercer's theorem and Green's function can approximate signals with a continuous spectrum. In particular, Green's formulation of the CS-KLE could estimate the n-pseudospectrum with an integral operator generated with an undersampling ratio of $\delta = 0.6$. However, Mercer's formulation of the CS-KLE needs the undersampling ratio to be as high as 0.9. We believe that the difference between the required undersampling ratios of the two formulations is due to the fact that eigenfunction estimation using Mercer's theorem requires a larger number of measurements. However, in Greens' function approach, the test function $C(t, \omega)$ obtained using inner integral already provides all the required spectral components for continuous spectrum approximation. Another difference between the Mercer's and Green's function formulation of CS-KLE is their complexity. Obviously, the eigendecomposition of the covariance matrix leads to a memory bottleneck for large data sets. CS-KLE with

Green's function formulation breaks the problem into simple vector multiplication, which can be handled using parallel processing. In particular, the CS-KLE through Green's application can be potentially solved using Kilo-core GPGPU, which not only prevents memory bottleneck bit it also decreases convergence time significantly.

Compared to GS-CS, CS-KLE does not require knowledge about the distribution of data. It has been shown that by adopting a proper philosophy, n-pseudospectrum can approximate the continuous spectrum of the continuous operators. The optimum subspace n (resp. the optimum n-pseudospectrum for semi-infinite and infinite-dimensional signal) has been computed by measuring spectral overlap between the estimated spectrum and actual signal spectrum. It is obvious that such a measurement requires knowledge about the actual signal, which does not sound feasible. Measuring optimum subspace n (resp. the optimum n-pseudospectrum) from a compressed sensing measurement vector $\bar{y}$ is a subject have not been addressed in this work. Readers can refer to [13], where compressible signals are studied as locally convex spaces.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2016.

[2] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 40–44.

[3] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Apr. 2010.

[4] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[5] T. Blumensath and M. E. Davies, "Stagewise weak gradient pursuits," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4333–4346, Nov. 2009.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 2009.

[8] M. Loève, *Probability Theory I, II*. New York, NY, USA: Springer, 1979.

[9] C. Schwab and R. A. Todor, "Karhunen–Loève approximation of random fields by generalized fast multipole methods," *J. Comput. Phys.*, vol. 217, no. 1, pp. 100–122, Sep. 2006.

[10] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete Comput. Geometry*, vol. 35, no. 4, pp. 617–652, May 2006.

[11] D. Donoho and V. Stodden, "Breakdown point of model selection when the number of variables exceeds the number of observations," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2006, pp. 1916–1921.

[12] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[13] M. Robaei and R. Akl, "Compressible topological vector spaces," 2022, *arXiv:4427780*.

[14] A. Hansen, "On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators," *J. Amer. Math. Soc.*, vol. 24, no. 1, pp. 81–124, Jan. 2011.

[15] E. B. Davies, "A defence of mathematical pluralism," *Philosophia Math.*, vol. 13, no. 3, pp. 252–276, Oct. 2005.

[16] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[17] P. R. Halmos, *A Hilbert Space Problem Book*. New York, NY, USA: Springer, 1982.

[18] P. Blanchard and E. Brüning, *Mathematical Methods in Physics: Distributions, Hilbert Space Operators, Variational Methods, and Applications in Quantum Physics*, vol. 69. Basel, Switzerland: Birkhäuser, 2015.

[19] O. Kallenberg and O. Kallenberg, *Foundations of Modern Probability*, vol. 2. New York, NY, USA: Springer, 2002.

[20] T. Tao, *An Introduction to Measure Theory*, vol. 126. Providence, RI, USA: American Mathematical Society Providence, 2011.

[21] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," *Bull. Amer. Math.*, vol. 54, no. 2017, pp. 151–165, 2017.

[22] P. R. Halmos and V. S. Sunder, *Bounded Integral Operators on L^2 Spaces*, vol. 96. Berlin, Germany: Springer, 1978.

[23] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proc. IEEE*, vol. 98, no. 6, pp. 913–924, Jun. 2010.

[24] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Dec. 2009.

[25] B. Adcock and A. C. Hansen, "Generalized sampling and infinite-dimensional compressed sensing," *Found. Comput. Math.*, vol. 16, no. 5, pp. 1263–1323, Oct. 2016.

[26] M. Gataric and C. Poon, "A practical guide to the recovery of wavelet coefficients from Fourier measurements," *SIAM J. Sci. Comput.*, vol. 38, no. 2, pp. A1075–A1099, Jan. 2016.

MOHAMMADREZA ROBAEI (Member, IEEE) received the B.Sc. degree in electrical and electronics from Tabriz Azad University, Iran, in 2006, the M.Sc. degree in electrical engineering from Middle East Technical University, Ankara, Turkey, in 2015, and the M.Sc. degree in information systems and technology from the University of Michigan, Dearborn, MI, USA, in 2017. He is currently pursuing the Ph.D. degree with the University of North Texas, Denton, TX, USA, working on advanced stochastic signal processing.

His research interests include a wide variety of topics in wireless communications, signal processing, and spectral and time-frequency analysis. His recent works focus on millimeter-wave physical layer signal processing, uniform convergence of compressible vector spaces, topological analysis of sparsity, optimization, joint-distribution analysis, and signal processing in general.

ROBERT AKL (Senior Member, IEEE) received the B.S. degree in computer science and the B.S., M.S., and D.Sc. degrees in electrical engineering from Washington University in St. Louis, in 1994, 1996, and 2000, respectively.

He is currently a Tenured Associate Professor at the University of North Texas. He has designed, implemented, and optimized both hardware and software aspects of several wireless communication systems for cellular, Wi-Fi, and sensor networks. He has broad expertise in wireless communications, bluetooth, cellular, Wi-Fi, VoIP, telephony, computer architecture, and computer networks. He has been awarded many research grants by leading companies in the industry and the National Science Foundation. He has developed and taught over 100 courses in his field. He has received several awards and commendation for his work, including the 2008 IEEE Professionalism Award and was the Winner of the 2010 Tech Titan of the Future Award.