Full Article
doi.org/10.1002/ansa.202100018
Chemistry Europe
European Chemical Societies Publishing

Received: 19 March 2021
Revised: 19 May 2021
Accepted: 15 June 2021

Task-adaptive eigenvector-based projection (EBP) transform for compressed sensing: A case study of spectroscopic profiling sensor

Yinsheng Zhang[1,2] | **Haiyan Wang**[1] | **Yongbo Cheng**[3] | **Xiaolin Qin**[4]

[1] School of Management and E-Business, Zhejiang Gongshang University, Hangzhou, China

[2] School of Information Sciences, University of Illinois at Urbana Champaign, Champaign, Illinois, USA

[3] School of Management Science and Engineering, Nanjing University of Finance and Economics, Nanjing, China

[4] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Correspondence
Haiyan Wang and Yinsheng Zhang, School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China.
Email: njue2010@163.com; zhangys@illinois.edu

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 91746202, 61806177, 71433006, 72072163; Zhejiang Provincial Natural Science Foundation of China, Grant/Award Number: LQ20C200004; China Scholarship Council, Grant/Award Number: 201808330609

Abstract

The compressed sensing (CS) theory requires the signal to be sparse under some transform. For most signals (e.g., speech and photos), the non-adaptive transform bases, such as discrete cosine transform (DCT), discrete Fourier transform (DFT), and Walsh-Hadamard transform (WHT), can meet this requirement and perform quite well. However, one limitation of these non-adaptive transforms is that we cannot leverage domain-specific knowledge to improve CS efficiency. This study presents a task-adaptive eigenvector-based projection (EBP) transform. The EBP basis has an equivalent effect of the principal component loading matrix and can generate a sparse representation in the latent space. In a Raman spectroscopic profiling case study, EBP demonstrates better performance than its non-adaptive counterparts. At the 1% CS sampling ratio (k), the reconstruction relative mean square errors of DCT, DFT, WHT and EBP are 0.33, 0.68, 0.32, and 0.00, respectively. At a fixed k, EBP achieves much better reconstruction quality than the non-adaptive counterparts. For specific domain tasks, EBP can significantly lower the CS sampling ratio and reduce the overall measurement cost.

KEYWORDS
compressed sensing, discrete cosine transform, discrete Fourier transform, eigenvector-based projection, sparsity, Walsh-Hadamard transform

1 | INTRODUCTION

In the field of information theory, compressed sensing (CS) breaks through the minimum required sampling rate, as required by the traditional Shannon-Nyquist sampling theorem (the sampling rate be at least twice the interested component's frequency). In engineering, CS has the following advantages[1]: (1) Require fewer data to be sampled and stored. (2) Enhance the analog-to-digital conversion (ADC) efficiency. (3) Reduce measurement requirements and save sensor cost. Because of these benefits, CS has been widely applied in various research and commercial applications, such as medical imaging,[2,3] biochemical analysis,[4] single-pixel imaging,[5] remote sensing,[6] etc.

A key issue in CS is basis selection. A basis Ψ defines a transform that maps the original signal to a latent space. The theory of CS requires this latent representation to be as sparse as possible. The sparsity degree determines the minimum sampling rate and the maximum signal reconstruction quality. Mathematically, Ψ is a unitary matrix, that is, $\Psi^H = \Psi^{-1}$. The most commonly used bases include

Full Article
doi.org/10.1002/ansa.202100018

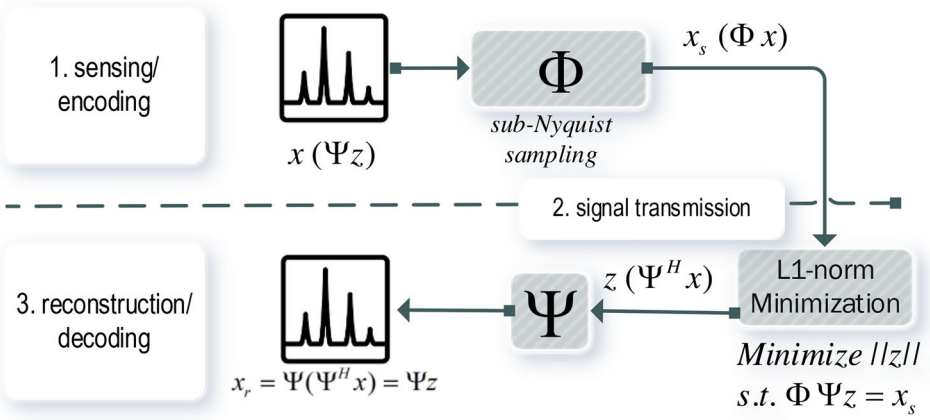

FIGURE 1 A typical compressed sensing workflow. x is the original signal/data. Φ is the sensing matrix. Ψ is a transform basis. z is the latent representation of x under Ψ. Ψ^H is the conjugate transpose of Ψ. x_s is the sampled data from CS. x_s is the actual data transmitted to the receiver side. x_r is the reconstructed data from x_s

discrete cosine transform (DCT), discrete Fourier transform (DFT), and HWT (Hadamard-Walsh transform). These bases are general-purposed and "non-adaptive." We argue that there could exist better "adaptive" counterparts for specific tasks. However, basis design/selection is non-trivial and requires prior knowledge about the domain signal's statistical characteristics. This paper's primary goal is to find and validate such a task-adaptive transform that suits the spectroscopic profiling applications. Spectroscopic profiling is a family of analytical chemistry instruments. It mainly includes vibrational spectrometry (e.g., violet, infra-red, and Raman), NMR (nuclear magnetic resonance), mass spectrometry, and their coupled use, for example, surface-enhanced Raman scattering, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MS), inductively coupled plasma MS. In a previous CS study,[7] we used non-adaptive transforms for Raman spectroscopy. In complement to the previous research, this paper will study adaptive transforms.

The structure of the paper proceeds as follows. (1) The first part is a brief review of the CS theory. (2) The second part describes the EBP transform. (3) The third part is a spectroscopic profiling case study. Finally, a comparison between EBP and non-adaptive counterparts is provided.

2 | THEORY

2.1 | A recap of CS

A typical CS workflow contains the following procedures (Figure 1): (1) Sensing/encoding. In this step, a transform basis matrix Ψ must be pre-determined. Ψ defines a latent space in which the original x has a sparse representation z, that is, $x = \Psi z$. A sensing matrix Φ is then applied to x to perform a sub-Nyquist sampling, that is, $x_s = \Phi x$. Φ has a dimensionality of $(kn \times n)$. k is the sensing ratio. (2) Signal transmission. x_s is transmitted to the receiver side. The signal length of x_s is $n_s = kn$. Typically, $k << 1$. A small k means a bigger information loss but a higher signal compression effect. In this sense, CS can significantly reduce the transmission payload and bandwidth requirement. (3) Reconstruction/decoding. Denote $A = \Omega\Psi$ as the measurement matrix. A has $kn \times n$ dimensions. Because $kn << n$, rank$(A) = kn$ (full row rank). Thus $Az = x_s$corresponds to an underdetermined linear system, which has more unknowns than equations. Because of the CS basic assumption (under Ψ, z is sparse), we need to find a sparse candidate z. This can be solved by minimizing the L0 (NP-hard, seldom used) or L1-norm. Finally, the signal can be restored by z. Because Ψ is a unitary matrix, $\Psi^H = \Psi^{-1}$. With$z = \Psi^H x$, the reconstructed data $x_r = \Psi z$.

In the above workflow, the transform basis matrix Ψ plays an essential role. It not only determines the minimum sampling ratio but also decides the signal reconstruction quality. The following section will introduce some of the most commonly used non-adaptive transforms.

2.2 | Commonly used non-adaptive transforms

Table 1 lists three widely used non-adaptive transforms, including DCT, DFT, and HWT (Hadamard-Walsh transform). These transforms can suit most signals (except inherently stochastic ones or white noises) encountered in our daily lives. For example, Joint Photographic Experts Group internally uses DCT, based on the general fact that a photo has most of its information concentrated in only a few low-frequency components (i.e., sparse under DCT).

3 | METHOD

3.1 | Problem statement

Being "non-adaptive" is a double-edged sword. On the one hand, non-adaptive transforms do not need to make any specific assumptions on the domain signal, yet they can perform reasonably well in general

Full Article
doi.org/10.1002/ansa.202100018
Chemistry Europe
European Chemical Societies Publishing

TABLE 1 Commonly used non-adaptive transforms and their bases

Transforms	Basis matrix Ψ
discrete cosine transform (DCT)	$DCT_{N,N} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \sqrt{2}\cos(\frac{\pi}{2N}) & \sqrt{2}\cos(\frac{3\pi}{2N}) & \cdots & \sqrt{2}\cos(\frac{(2N-1)\pi}{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{2}\cos(\frac{(N-1)\pi}{2N}) & \sqrt{2}\cos(\frac{3(N-1)\pi}{2N}) & \cdots & \sqrt{2}\cos(\frac{(2N-1)(N-1N-1)\pi}{2N}) \end{bmatrix}$
discrete Fourier transform (DFT)	$DFT_{N,N} = (\frac{\omega^{jk}}{\sqrt{N}})_{j,k=0,\ldots,N-1} = \frac{1}{\sqrt{N}} \begin{bmatrix} \omega^0 & \omega^0 & \omega^0 & \cdots & \omega^0 \\ \omega^0 & \omega^1 & \omega^2 & \cdots & \omega^{N-1} \\ \omega^0 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \omega^0 & \omega^3 & \omega^6 & \cdots & \omega^{3(N-1)} \\ \omega^0 & \vdots & \vdots & \ddots & \vdots \\ \omega^0 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{bmatrix}$ where $\omega = e^{-2\pi i/N} = \cos(-2\pi/N) + i\sin(-2\pi/N)$
HWT (Hadamard-Walsh transform)*	$HWT_{(2^N)} = \frac{1}{\sqrt{N/2}} \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & -1 & 1 & -1 & \cdots & 1 & -1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & -1 & 1 & -1 & \cdots & 1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & -1 & 1 & -1 & \cdots & 1 & -1 \end{bmatrix}$

*The HWT basis must be padded to 2^N dimensions.

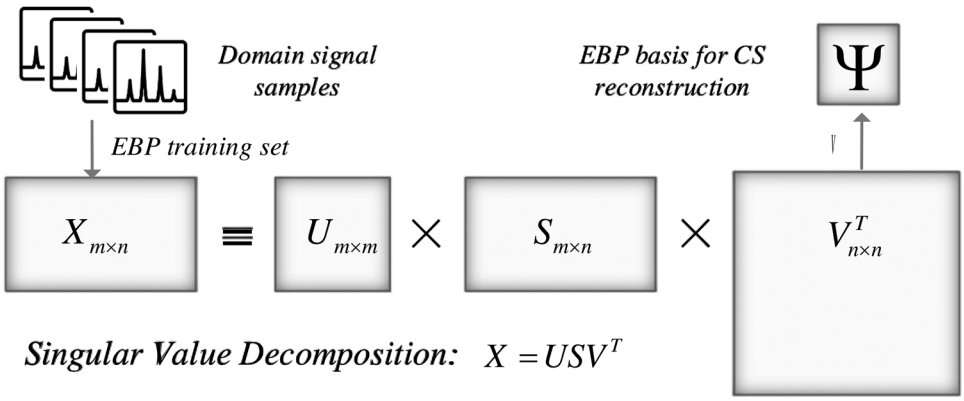

FIGURE 2 Use SVD to generate the EBP basis. The image shows the following steps: (1) Collect domain signal samples to construct the training set *X*. (2) Perform SVD on *X*. (3) Use the last rotational factor *V* as the EBP basis for CS reconstruction. Any new signal sample from the same domain is expected to be sparse under EBP

cases. On the other hand, they cannot leverage any domain-specific prior knowledge, putting them in a disadvantageous status for concrete tasks. Therefore, this study's primary goal is to present and validate a task-adaptive transform that better suits spectroscopic profiling signals.

3.2 | Eigenvector-based projection transform

3.2.1 | Basis generation

This section presents an eigenvector-based projection (EBP) transform based on singular value decomposition (SVD). The SVD is a matrix factorization technique that decomposes a matrix into three parts: $X = USV^T$. *U* and *V* are orthogonal (also unitary). Columns in *U* are left eigenvectors. Columns in *V* are right eigenvectors. *S* is diagonal. The diagonal elements in *S* are called singular values. *U* and *V* represent the "rotational" factors in *X*, while *S* represents the "stretching" factor in *X*.

Figure 2 shows the EBP basis training process. First, we need to collect sufficient domain signal samples. "Being sufficient" means these samples should capture enough distribution information (variance in the sense of SVD). With *m* samples of the *n*-dimensional signal, we will construct an *m*×*n* matrix *X*. Then, perform SVD on *X*. The last rotational factor *V* is essentially a group of eigenvectors. *V* can be used as an EBP transform basis in the CS scenario.

The following section will describe EBP's relationship with PCA. We will see that the EBP basis *V* is also the left rotational factor of the covariance matrix in PCA. It equals the component loading matrix and has an "information concentration" effect. Applying *V* to a new signal from the same domain is expected to get a sparse projection.

Full Article
doi.org/10.1002/ansa.202100018
Chemistry Europe
European Chemical Societies Publishing

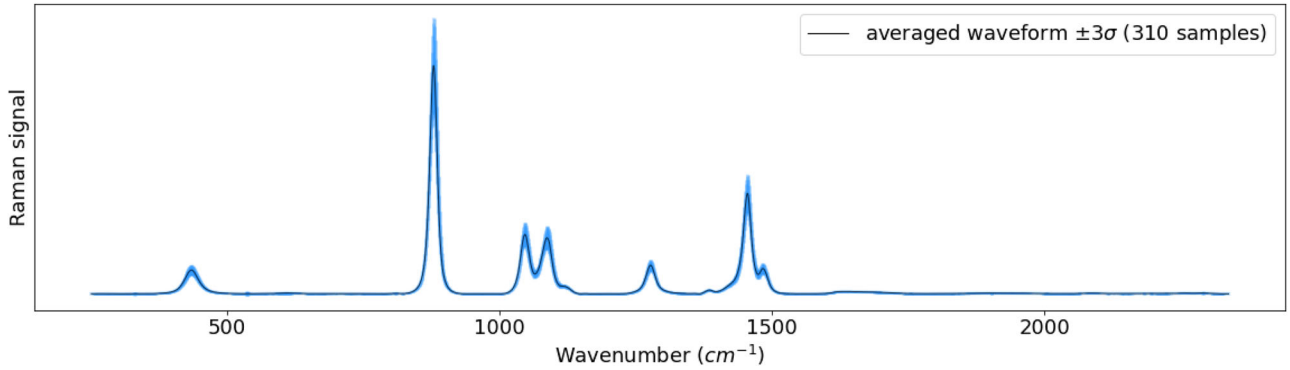

FIGURE 3 The averaged Raman signal of 310 vintage liquor samples with ± 3 standard deviations (99.74% samples will fall in this region). The shaded region is the deviation region

3.2.2 | Relationship with PCA

EBP is also closely related to principal component analysis (PCA). PCA involves the SVD of the covariance matrix ($\frac{X^TX}{m}$, X is demeaned), as follows.

$$Cov = \frac{X^TX}{m} = (VS^TU^T)USV^T/m = V(\frac{S^TS}{m})V^T \triangleq V\Lambda V^T$$

Λ is a diagonal matrix of eigenvalues. Because $\Lambda = \frac{S^TS}{m}$, so each eigenvalue equals the square of singular value divided by the number of observations, that is, $\lambda_i = \frac{s_i^2}{m}$. The above uses the biased version of the covariance matrix. The unbiased version is $Cov = \frac{X^TX}{m-1}$.

In the context of PCA, V is the component loading matrix. By only keeping the first K row eigenvectors in V, X (an $m \times n$ matrix) can be projected into an $m \times K$ matrix in the PCA latent space. Typically, $K < n$, so PCA has a dimensionality reduction effect. It also comes with a lossy compression effect, with information loss = $\sum_{K+1}^{n} \lambda_i / \sum_{1}^{n} \lambda_i$.

For a specific domain dataset X, the first few eigenvectors from SVD are projection directions that capture most signal information/variance. After EBP, the transformed representation z in the latent space will be very sparse. In the following part, we will carry out a case study to verify EBP and compare it with non-adaptive ones.

3.3 | Mathematical symbols and glossary

The mathematical symbols used in this manuscript are provided in Table 2.

4 | CASE STUDY

4.1 | Dataset

Background: The purpose of this case study is to profile vintage liquors with Raman spectroscopy. Raman spectroscopy is a vibrational spectroscopic technology.[8] Researchers have used Raman spectroscopy to profile various materials and products, such as wines,[9] herbs,[10] oil,[11] milk.[12] At the core of the Raman spectroscopy is a charge-coupled device sensor, which can be enhanced with CS to speed up the ADC.

Dataset: A Raman spectroscopic dataset of 310 vintage liquor samples. These samples are from five batches of 8-year *Gujing Tribute* (古井贡, a national geographical iconic brand of China) vintage liquor. Figure 3 plots the averaged signal.

Instrument: Prott-ezRaman-D3 Raman spectrometer, manufactured by Enwave Optronics, US. Spectral resolution: 1 cm^{-1}. Spectral range: 251 ~ 2338 cm^{-1}.

4.2 | Signal sparsity analysis in the latent space

CS requires the signal to be "sparse" under a specific transform. The sparsity of the transformed signal z determines the minimum sampling ratio k. The basis that generates the sparsest representation in the latent space is considered the best. Therefore, we first perform a preliminary sparsity analysis on different CS transforms. In Figure 4, the second column shows each transform's z in the latent space. (1) For identify matrix (IDM), $z = x$, that is, the transformed signal z equals the original signal x. (2) DCT and DFT have sparse z in the latent space. Most of the energy concentrates in the low-frequency components. (3) For HWT, the non-zero elements in z are evenly distributed (not concentrated). (4) EBP has the sparsest z representation. Considering z's sparsity, EBP is expected to perform better than its non-adaptive counterparts.

The last four columns in Figure 4 are the reconstructed results under different transforms and sampling ratios. By comparing the reconstructed signals, we have the following observations: (1) When $k = 0.01$, $n_s = 20$. It means using only 20 points to reconstruct the original 2088 points. With such limited information, the reconstruction of the non-adaptive transforms is quite messy. In contrast, EBP has a near-perfect reconstruction. (2) With k increases, more information is reserved, and the reconstruction quality gradually improves. (3) Regarding the reconstruction quality, EBP > DCT > DFT > HWT. At the same k, the reconstructed signals of DFT and HWT are more "spiky/noisy" than DCT.

Full Article
doi.org/10.1002/ansa.202100018
Chemistry Europe
European Chemical Societies Publishing

TABLE 2 Symbols and terms

Symbol/Term	Explanation
CS	Compressed sensing
Shannon-Nyquist theorem	To prevent aliasing, the minimum sampling rate is twice the interested component's frequency.
n	The original signal dimensionality. The length of the source signal.
m	The number of samples/signals/observations.
k	CS sampling rate, i.e., the percentage to be sampled from the original signal. Value range is *(0,1)*. A lower k means higher sampling efficiency but more info loss.
Φ	Sensing matrix. A *kn-by-n* square matrix.
Ψ	A transform basis. An *n-by-n* square matrix. Ψ is a unitary matrix, i.e.,$\Psi^H = \Psi^{-1}$.
A	The measurement matrix. $A = \Omega\Psi$.
x	The original sample/observation/signal.
X	The training dataset. An *m-by-n* matrix.
z	Signal representation in the latent space. $z = \Psi^H x$
x_s	Sampled signal. $x_s = \Phi x = \Phi\Psi z = Az$
n_s	The dimensionality of the sampled signal x_s. $n_s = kn$. In CS, n_s means how many points to be randomly picked from the original signal x.
x_r	Reconstructed signal.
rank	The number of linearly-independent column/row vectors of a matrix. For the measurement matrix A, $rank(A) = kn$.
column space	The column space of matrix A is the set of all possible outputs of $A\vec{v}$ ($\vec{v}$ is any arbitrary vector).
null space	The set of vectors that become$\vec{0}$ (lose their identities) after transform A, i.e., all solutions to$A\vec{v} = \vec{0}$. Column space is also known as the kernel of a matrix. Because $A\vec{v} = \lambda\vec{v}$, the eigenspaces of A are the null spaces of $A - \lambda I$. The general solution for the underdetermined linear system $Az = x_s$ can be represented as $z_{sp} + z_0$, where z_{sp} is a specific solution and $z_0 \in null(A)$. This is due to$A(z_{sp} + z_0) = Az_{sp} + Az_0 = x_s + \vec{0} = x_s$
nullity	The dimension of the null space/kernel.
Underdetermined linear system	In CS, $Az = x_s$ defines an underdetermined linear system (with more unknowns than equations), as the measurement matrix A is a full row rank matrix, i.e., $kn < n$.
rank-nullity theorem	*rank(A) + nullity(A) = column number(A)*, i.e., the nullity of A is the complement to its rank.
I, or IDM	An identity matrix
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
HWT	Hadamard-Walsh transform
EBP	Eigenvector-based projection
SVD	Singular value decomposition
U	The first orthogonal matrix (left "rotational" factor) of SVD. Columns in U are left eigen vectors.
S	The middle diagonal matrix from SVD. S represents the "stretch" factor in X. The diagonal elements are singular values.
s_i	The *i-th* singular value in S
V	The second orthogonal matrix (right "rotational" factor) of SVD. Columns in V are right eigenvectors. In the context of PCA, V is the component loadings.
PCA	Principal component analysis
K	In PCA, K (capital) is the number of principal components to be kept.
Cov	The covariance matrix. Cov of a de-meaned matrix X is $\frac{X^TX}{m}$ (biased estimator) or $\frac{X^TX}{m-1}$(unbiased estimator).
Λ	Eigen-value matrix. Λcomes from the SVD of Cov. Each diagonal element is an eigenvalue. $Cov = \frac{X^TX}{m} = (VS^TU^T)USV^T/m = V(\frac{S^TS}{m})V^T \triangleq V\Lambda V^T$, *so* $\Lambda = \frac{S^TS}{m}$
λ_i	The *i-th* eigenvalue in Λ. It is related to s_i by $\lambda_i = \frac{s_i^2}{m}$.
MSE	Mean square error. $MSE = \frac{\|\|x-x_r\|\|^2}{n}$. MSE is used to measure the reconstruction error of x_r.
RMSE	Relative mean square error. $RMSE = \frac{\|\|x-x_r\|\|^2}{\|\|x\|\|^2}$
SNR	Signal-to-noise ratio. $SNR = \frac{\|\|x\|\|^2}{\|\|x-x_r\|\|^2} = \frac{1}{RMSE}$. It can also be measured in dB as $10\log_{10}(SNR)$.

Full Article
doi.org/10.1002/ansa.202100018
Chemistry Europe
European Chemical Societies Publishing

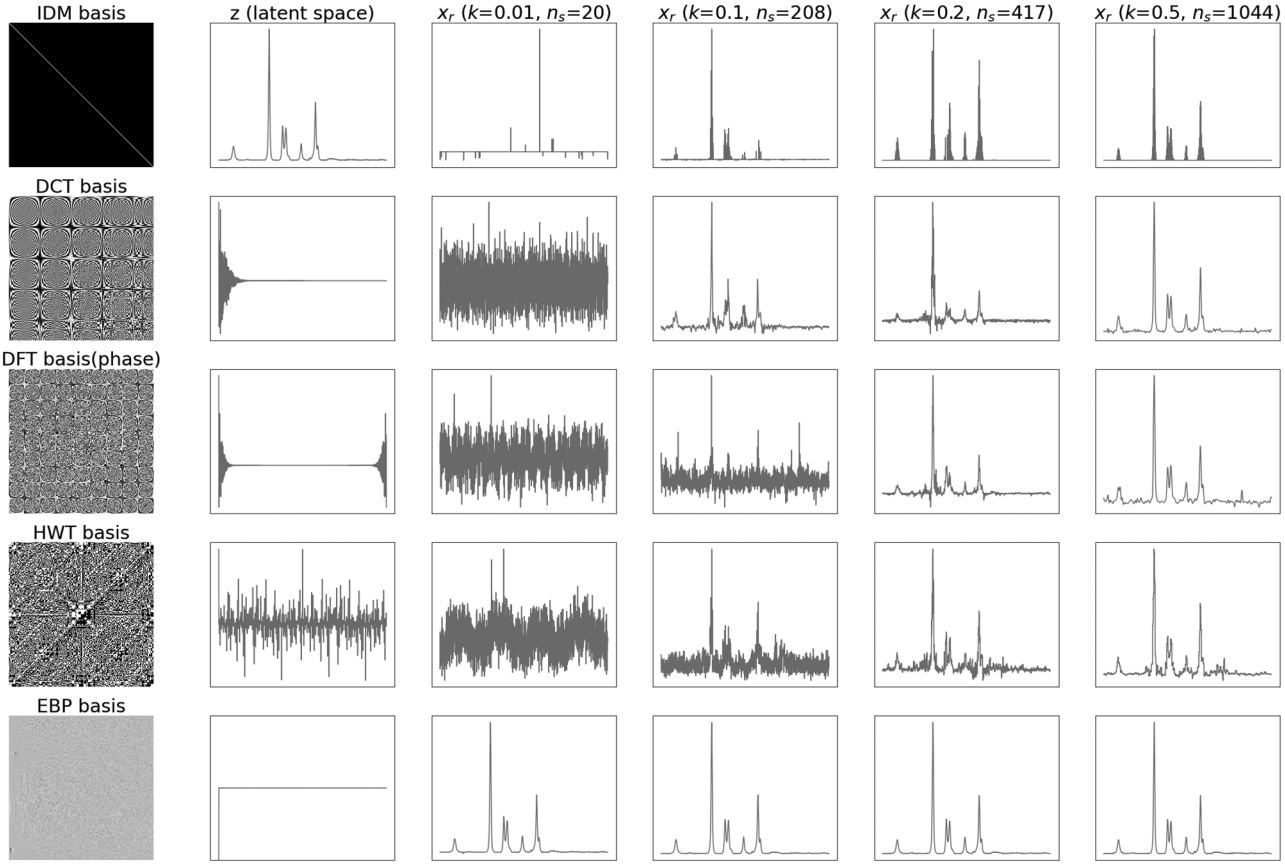

FIGURE 4 CS reconstruction results of different transform bases at varied sampling ratios. The last four columns display the reconstructed signal (x_r) at different k levels. Each column corresponds to a specific k value, while each row is for one transform. IDM's z (in the first row and second column) equals the original signal x and is the ground truth for reconstruction

In the next section, we will use cross-validation (CV) to compare their signal reconstruction qualities.

4.3 | Evaluation strategy

To evaluate the different CS transforms, we may consider existing CV or bootstrapping strategies. Although both are resampling methods, CV is more dedicated to model evaluation, while bootstrapping is more used in statistical parameter estimation and ensemble learning. In actual datasets, the k-fold cross validation arguably performs better than others.[13] In this study, we choose repeated double k-fold CV (rdk-CV).[14,15] The outer loop of rdk-CV is a repeated 10-fold CV, and the inner loop is another five-fold CV for optimizing the CS reconstruction hyper-parameters.

Each CV iteration includes the following steps.

4.3.1 | Basis training

The training set is used to construct the EBP basis by the SVD decomposition (Figure 2). Three non-adaptive transform bases, that is, DCT, DFT, and HWT (Hadamard-Walsh transform), are generated by their definitions (Table 1). They don't need an explicit training process. The identity transform is also included as a theoretical baseline. Its basis is simply an IDM.

The bases of IDM, EBP, DCT, and DFT are all 2088×2088 square matrices, whereas the HWT basis is 4096×4096, as HWT requires the basis's dimension to be 2^N. The generated bases are plotted in the first column of Figure 4.

4.3.2 | CS

Perform CS on the test set under different transform bases and k values. Because the source signal x is an n-dimensional ($n = 2088$) vector, performing CS on the source signal equals randomly drawing kn (k is the sampling ratio. $0 < k \leq 1$) points from x. Therefore, the length of x_s is $n_s = kn$. For good ADC efficiency, k should be small. However, small k also means big information loss and low reconstruction quality. In real applications, users should weigh the pros and cons when choosing k value. This study will try a wide range of k values, that is, 0.01, 0.1, 0.2, and 0.5.

4.3.3 | Reconstruction

In CS reconstruction, we use Least Absolute Shrinkage and Selection Operator[16] (LASSO) to solve L1-norm minimization. In the inner loop

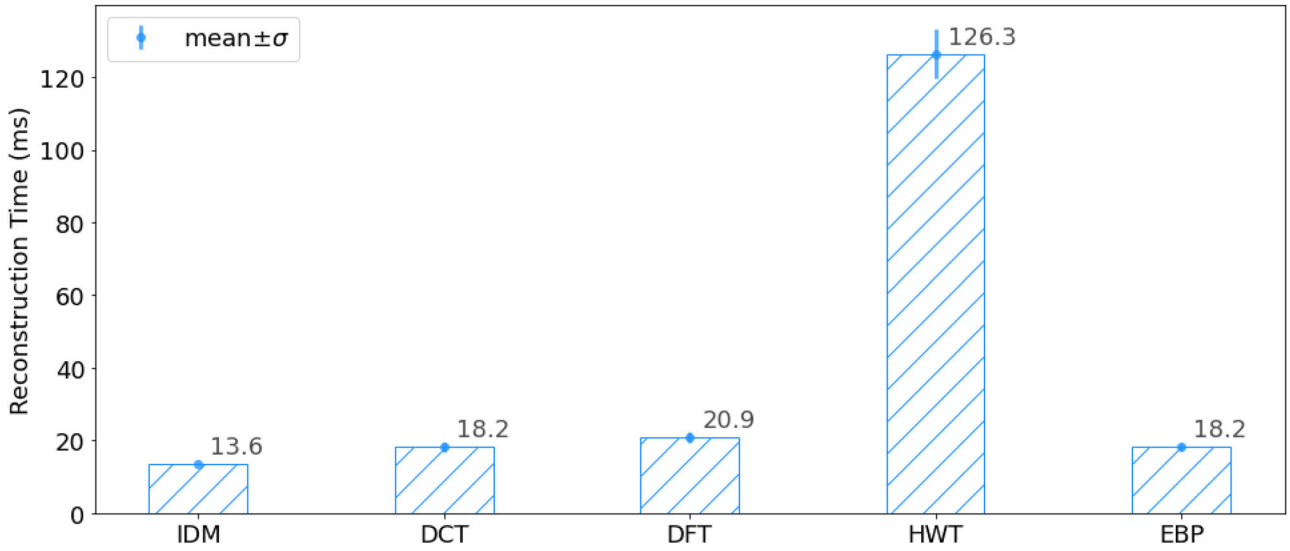

FIGURE 5 Averaged CS reconstruction time of different transform bases

TABLE 3 Signal reconstruction qualities. k = CS sample ratio, ranging from 0 to 1.0. n_s = how many points are samples from the original signal, that is, the length of x_s. $n_s = nk$

k		0.01	0.1	0.2	0.5
n_s		20	208	417	1044
RMSE	IDM	0.319	0.289	0.258	0.161
	DCT	0.326	0.114	0.062	0.033
	DFT	0.681	0.316	0.184	0.151
	HWT	0.322	0.276	0.149	0.058
	EBP	0.0	0.0	0.0	0.0
SNR	IDM	3.134	3.461	3.876	6.206
	DCT	3.063	8.756	16.172	30.058
	DFT	1.468	3.168	5.438	6.608
	HWT	3.102	3.617	6.706	17.195
	EBP	4.8×10^3	2.3×10^5	4.6×10^5	1.5×10^6

Abbreviations: RMSE, relative mean square error; SNR, signal-to-noise ratio.

of rdk-CV, we use another five-fold CV to decide the best L1 regularization hyper-parameter for LASSO.

4.3.4 | Measurement

We measure the signal reconstruction quality of the test set with the relative mean square error (RMSE). $RMSE = \frac{||x-x_r||^2}{||x||^2}$) and signal-to-noise ratio (SNR) ($SNR = \frac{||x||^2}{||x-x_r||^2} = \frac{1}{RMSE}$) metrics. SNR can also be calculated in dB as $10\log_{10}(SNR)$. RMSE measures the ratio of the residual/noise and the source. An SNR greater than 1.0 (0 dB) means more useful information than noise. The reconstruction time T is also recorded as a performance metric.

4.4 | Result

After the rdk-CV, the *RMSE* and *SNR* of all iterations are averaged (Table 3). The result is consistent with Figure 4. At k = 0.01, 0.1, 0.2, and 0.5, EBP has nearly zero RMSE. The SNRs are 4.8×10^3(37 dB), 2.3×10^5 (44 dB), 4.6×10^5 (47 dB), and 1.5×10^6 (52 dB), respectively. EBP has much better RMSE and SNR than the non-adaptive transforms.

During the CV evaluation, we also recorded the reconstruction time at each iteration. The averaged time for the five transforms is plotted in Figure 5. EBP has a comparable reconstruction time with DCT (18.2 ms) and is slightly better than DFT (20.9 ms). HWT uses significantly more time (126.3 ms) than others. That is because HWT requires to pad the original signal to 2^N dimensions (Ψ is also enlarged from 2088 × 2088 to 4096×4096), so it takes more computation time.

5 | DISCUSSIONS AND FUTURE WORK

5.1 | Identity matrix as a transform basis

In practice, we seldom use IDM. In the case study, we use IDM as a theoretical baseline. IDM has the following properties: (1) Under IDM, the latent space is the original signal space, so $z = x$. (2) In the reconstructed phase, $A = \Phi\Psi = \Phi I = \Phi$ and $x_r = z_r$, so $Az = x_s$ becomes $\Phi x_r = x_s$. Because each row in Φ is one-hot encoding, that is, only one element is 1, and all others are zero, the resulting x_r is just restoring each point in x_s to its original position. (3) When $k = 1$, Φ becomes a square matrix ($kn = n$), and x_s is an n-dimensional vector. x_s becomes a shuffled version of x. In the reconstruction phase, the linear system $\Phi x_r = x_s$ has equal numbers of unknowns and equations. Therefore, it has a unique solution, and x_r is an exact restoration of x.

5.2 | Determination of the required sample size

The EBP basis is generated by the SVD decomposition on the training set. One fundamental question worth further researching is how to determine the minimum sample size. More samples will usually provide more reliable and accurate information on the domain signal, while a small sample size will produce unstable results due to randomness. To determine the minimum sample size, we need to measure such "unstableness." One approach is to find/design a statistic and calculate its variance, for example, samples' mean or signal reconstruction quality. Another is to use bootstrapping to build surrogate models and measure the deviation between surrogate models. How to align these different approaches and find the most suitable one remains to explore.

5.3 | CS under-sampling schemes

In CS, the sampling pattern/scheme determines the sensing matrix Φ. This study uses the uniform random sampling pattern (denote as Φ_U), as it has been well tested and widely used. According to the CS theory, Φ and the transform basis Ψ must be incoherent. It can be proved that Φ_U has very low coherence (denote as μ) with common transform bases, for example, DCT ($\mu = \sqrt{2}$), DFT ($\mu = 1$), and HWT ($\mu = 1$). However, Φ_U may not be the only choice. In other disciplines, such as magnetic resonance imaging,[17,18] multiple two-dimensional under-sampling schemes have been studied. For spectroscopic data, designing and finding other one-dimensional sampling schemes is another future research topic.

5.4 | Basis selection from the dictionary learning perspective

The core problem in CS signal reconstruction is to solve the linear system $Az = x_s$, where $A = \Phi\Psi$. Because Φ is a ($kn \times n$) matrix, and Ψ is an ($n \times n$) matrix, A is a ($kn \times n$) matrix. In most CS cases, $k < 1$, so A is "overcomplete," that is, has more columns than rows. A is also "full row rank," with rank(A) = nk (its row number). If we rewrite A as $A = [\, a_1 \; a_2 \; \ldots \; a_n \,]$, A will be a dictionary/frame with excessive entries (each column vector is a dictionary entry). Therefore, we can find a sparse representation z under dictionary A. In this sense, the CS basis selection equals a dictionary learning problem. We will continue to research existing dictionary learning techniques to find other excessive/redundant dictionaries for x_s.

5.5 | Other adaptive transforms

Under EBP, each component in z is a linear combination of the original signal's n features. Besides this linear EBP transform, there are also non-linear transforms. For example, recent work in the deep-learning field, such as the auto-encoder, may be used to design non-linear transforms. In the future, we will continue to research other possibilities for CS basis design.

6 | CONCLUSION

This study presents an EBP transform and its application in spectroscopic profiling signals. Like PCA's component loading matrix, EBP has an information concentration effect. In the case study, EBP has demonstrated a remarkable performance. EBP's z in the latent space is much sparser than the non-adaptive transforms. At the same sampling ratio k, EBP has the best signal reconstruction quality. In conclusion, EBP is a competitive alternative to traditional non-adaptive transforms for specific spectroscopic profiling tasks. It can further lower the sampling ratio and reduce the overall measurement cost.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grant numbers: 91746202, 61806177, 71433006, and 72072163), Zhejiang Provincial Natural Science Foundation of China (grant number: LQ20C200004) and China Scholarship Council (grant number: 201808330609).

CONFLICT OF INTEREST

The authors do not declare any conflict of interest.

DATA AVAILABILITY STATEMENT

The dataset and codes have been uploaded to the public repository.

DOI: 10.17632/8cg4sctwxm
URL: https://doi.org/10.17632/8cg4sctwxm
License: Creative Commons Attribution (CC BY NC 3.0)

ORCID

Yinsheng Zhang https://orcid.org/0000-0003-4971-8809

REFERENCES

1. Gamez G. Compressed sensing in spectroscopy for chemical analysis. *J Anal At Spectrom*. 2016;31(11):2165–2174.
2. Jaspan ON, Fleysher R, Lipton ML. Compressed sensing MRI: a review of the clinical literature. *Br J Radiol*. 2015;88(1056):1–12.
3. Bao P, Sun H, Wang Z, et al. Convolutional sparse coding for compressed sensing CT reconstruction. *IEEE Trans Med Imaging*. 2019;38(11):2607–2619.
4. Conrad TOF, Genzel M, Cvetkovic N, et al. Sparse proteomics analysis – a compressed sensing-based approach for feature selection and classification of high-dimensional proteomics mass spectrometry data. *BMC Bioinformatics*. 2017;18(1):160–180.
5. Duarte MF, Davenport MA, Takhar D, et al. Single-pixel imaging via compressive sampling. *IEEE Signal Process Mag*. 2008;25(2):83–91.
6. Ghahremani M, Ghassemian H. Remote sensing image fusion using ripplet transform and compressed sensing. *IEEE Geosci Remote Sens Lett*. 2015;12(3):502–506.
7. Zhang Y, Zhang Z, Zhao Y, et al. Adaptive compressed sensing of Raman spectroscopic profiling data for discriminative tasks. *Talanta*. 2020;211(5):120681.
8. Smith E, Dent G. *Modern Raman Spectroscopy: A Practical Approach*. 2nd ed. Chichester, UK: Wiley; 2019.
9. Zanuttin F, Gurian E, Ignat I, et al. Characterization of white wines from north-eastern Italy with surface-enhanced Raman spectroscopy. *Talanta*. 2019;203:99–105.

10. Wang H-Y, Song C, Sha M, Liu J, Li L-P, Zhang Z-Y. Discrimination of medicine radix astragali from different geographic origins using multiple spectroscopies combined with data fusion methods. *J Appl Spectrosc*. 2018;85(2):313–319.
11. Jimenez-Carvelo AM, Osorio MT, Koidis A, Gonzalez-Casado A, Cuadros-Rodriguez L. Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy. *Lwt Food Sci Technol*. 2017;86:174–184.
12. Ahmad N, Saleem M. Raman spectroscopy based characterization of desi ghee obtained from buffalo and cow milk. *Int Dairy J*. 2019;89:119–128.
13. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int Jt Conf Artif Intell*. 1995;14:1137–1145.
14. Filzmoser P, Liebmann B, Varmuza K. Repeated double cross validation. *J Chemometrics*. 2009;23(4):160–171.
15. Ishibuchi H, Nojima Y. Repeated double cross-validation for choosing a single solution in evolutionary multi-objective fuzzy classifier design. *Knowl-Based Syst*. 2013;54:22–31.
16. Ranstam J, Cook JA. LASSO regression. *Br J Surg*. 2018;105(10):1348–1348.
17. Zhan Z, Cai J-F, Guo D, Liu Y, Chen Z, Qu X. Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. *IEEE Trans Biomed Eng*. 2016;63(9):1850–1861.
18. Deka B, Datta S. A practical under-sampling pattern for compressed sensing MRI. In: Bora PK, Prasanna SRM, Sarma KK, Saikia N, eds. *Advances in Communication and Computing*. Springer India: Lecture Notes in Electrical Engineering; 2015:115–125.

How to cite this article: Zhang Y, Wang H, Cheng Y, Qin X. Task-adaptive eigenvector-based projection (EBP) transform for compressed sensing: A case study of spectroscopic profiling sensor. *Anal Sci Adv*. 2022;3:29–37. https://doi.org/10.1002/ansa.202100018