Quantization of Eigen Subspace for Sparse Representation

Onur Yilmaz and Ali N. Akansu, *Fellow, IEEE*

***Abstract*—We propose sparse Karhunen–Loeve Transform (SKLT) method to sparse eigen subspaces. The sparsity (cardinality reduction) is achieved through the pdf-optimized quantization of basis function (vector) set. It may be considered an extension of the simple and soft thresholding (ST) methods. The merit of the proposed framework for sparse representation is presented for auto-regressive order one, AR(1), discrete process and empirical correlation matrix of stock returns for NASDAQ-100 index. It is shown that SKLT is efficient to implement and outperforms several sparsity algorithms reported in the literature.**

***Index Terms*—Arcsine distribution, cardinality reduction, dimension reduction, eigen decomposition, Karhunen–Loeve Transform (KLT), Lloyd-Max quantizer, midtread (zero-zone) pdf-optimized quantizer, principal component analysis (PCA), sparse matrix, subspace methods, transform coding.**

I. Introduction

ORTHOGONAL transforms (subspace methods) are widely used mathematical tools in many disciplines including signal processing. Eigenanalysis, also known as principal component analysis (PCA) and Karhunen-Loeve Transform (KLT), is the optimal block transform with an orthonormal basis that perfectly decorrelates the given signal in the subspace. However, it is signal dependent and its transform matrix is updated whenever the signal statistics change. Due to high cost of implementation, suboptimal fixed transforms with efficiency, e.g. discrete cosine transform (DCT) and discrete Fourier transform (DFT), are successfully used as good approximations to KLT for various signal types [1].

KLT has been employed in multivariate data processing and dimension reduction although the application specific interpretation of principal components (eigenvectors) is often difficult in some cases [2]–[5]. Moreover, small but non-zero loadings (elements) of each principal component (PC) (or eigenvector) bring implementation cost that is hard to justify in applications such as generation and maintenance (rebalancing) of eigen portfolios in finance [5]–[7]. This and other applications that utilize loading coefficients have motivated researchers to study sparsity of PCs in eigen analysis of matrices. Furthermore, unevenness of signal energy distributed among PCs in eigen subspace is reflected in eigenvalues (coefficient variances) that lead to dimension reduction. The latter is the very foundation of transform coding successfully used in visual signal processing and data compression [1], [8], [9]. Therefore, both dimension reduction and sparsity of basis functions (vectors) are significant attributes of orthogonal transforms widely utilized in many applications. This recent development has paved the way for our study where a rate-distortion based framework to sparse basis functions of subspaces including eigen subspace of a given covariance matrix is proposed. The challenge is to maximize explained variance by minimum number of PCs, also called energy compaction [1], while replacing the less significant samples (loading coefficients) of basis functions with zero to achieve the desired level of sparsity in signal representation.

Regularization methods have been used to make an ill-conditioned matrix invertible or to prevent overfitting [10], [11]. It is achieved by adding an ℓ_1 (norm-1) or ℓ_2 constraint in the optimization. As an example, ridge regression exploits an ℓ_2 penalty for stabilization in the least squares problem [10]. Eigenfiltering is another popular method employed for regularization [10], [11]. More recently, regularization methods have been also utilized for sparsity. ℓ_0 regularizer leads to a sparse solution. On the other hand, it makes the optimization problem non-convex.

ℓ_1 regularizer, so called lasso, is a widely used approximation (convex relaxation) to ℓ_0 case [3], [12]. Another ℓ_1 based method was proposed in [13] for sparse portfolios. SCoTLASS [3] and SPCA [4] utilize the ℓ_1 and ℓ_2 regularizers for sparse approximation to principal components (PCs), respectively. The sparse PCA is modeled in [3], [4] as an explained variance maximization problem where number of non-zero elements in the PCs considered as a basis design constraint. These methods suffer from potentially being stuck in local minima due to the non-convex nature of the optimization. A convex relaxation method called SDP Relaxations for Sparse PCA (DSPCA) using semidefinite programming (SDP) was proposed to deal with a simpler optimization [5]. Empirical performance results for certain cases indicate that DSPCA may generate sparse PCs that preserve slightly more explained variance than SCoTLASS [3] and SPCA [4] for the same sparsity level. A nonnegative variant of the sparse PCA problem that forces the elements of each principal components (PCs) to be nonnegative, is introduced in [14]. Nonnegative sparse PCA (NSPCA) offers competitive performance to SCoTLASS, SPCA and DSPCA in terms of explained variance for a given sparsity. However, sign of the PC elements bear specific information for the applications of interests such as eigenportfolios. Thus,

Manuscript received November 24, 2014; revised February 15, 2015; accepted April 22, 2015. Date of publication May 07, 2015; date of current version June 05, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antonio Napolitano.

The authors are with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, University Heights, Newark, NJ 07102 USA (e-mail: onur.yilmaz@njit.edu; Akansu@NJIT.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2430831

NSPCA is not applicable for all types of applications. Another lasso based approach, so called sparse PCA via regularized SVD (sPCA-rSVD), is proposed in [15]. Simulation results for certain cases show that sPCA-rSVD provides competitive results to SPCA. A variation of sPCA-rSVD, so called sparse principal components (SPC), that utilizes the penalized matrix decomposition (PMD) is proposed in [16]. PMD that computes the rank K approximation of a given matrix is proposed in [16]. It utilizes the lasso penalty for sparsity. Unfortunately, none of these methods result in guaranteed sparsity regardless of their prohibitive computational cost for high dimensions. Moreover, the lack of mathematical framework to measure distortion, or explained variance loss, for a desired sparsity level makes sparse PCA methods of this kind quite ad-hoc and difficult to use. On the other hand, the simple (hard) thresholding technique is easy to implement [2]. It performs better than SCoTLASS and slightly worse than SPCA [4]. Although simple thresholding is easy to implement, it may cause unexpected distortion levels as called variance loss. Soft thresholding (ST) is another technique that is utilized for sparse representation in [4]. Certain experiments show that ST offers slightly better performance than simple thresholding [4]. Therefore, threshold selection plays a central role in sparsity performance.

In this paper, we propose a subspace sparsing framework based on the rate-distortion theory [1], [17]–[19]. It may be considered as an extension of the simple or soft thresholding method to unify sparse representation problem with an optimal quantization method widely used in the source coding field [1], [2], [8], [9], [19]. The method employs a varying size mid-tread (zero-zone) pdf-optimized (Lloyd-Max) quantizer designed for component histogram of each eigenvector (or the entire eigen matrix) to achieve the desired level of distortion (sparsity) in the subspace with reduced cardinality [17], [18], [20]. Although there are studies in the literature that jointly examine compressed sensing (CS) and quantization [21], this is the first attempt to utilize pdf-optimized quantization based methods for sparse PCA problem. We focus on eigen subspace of autoregressive order one, AR(1), discrete process due to the availability of closed form expressions for its eigenvectors and eigenvalues. It is known that AR(1) process approximates well many real world signals [1]. We also sparse eigenportfolios of NASDAQ-100 index by using this method. It is noted that the proposed method to sparse a subspace through quantization of its basis functions is a marked departure from the traditional transform coding where transform coefficients, in the subspace, are quantized for dimension reduction also called zonal sampling in the literature [1], [8], [9]. Therefore, we investigate the trade-off between subspace orthogonality and sparsity from the rate-distortion perspective for the case where original values of transform coefficients are employed. Then, we provide a comparative performance of the proposed method along with the various methods reported in the literature such as ST [4], SPCA [4], DSPCA [5], and SPC [16] with respect to the metrics of non-sparsity (NS) and variance loss (VL).

The mathematical preliminaries are given in Section II. The basic concepts of orthonormal subspaces, rate-distortion theory and transform coding (TC), and the design of pdf-optimized quantizer are revisited in this section. Eigenanalysis of AR(1) process is summarized in Section III. In Section IV, we introduce a pdf model for the components (loading coefficients) of each eigenvector (or the entire eigen matrix) for AR(1) process that leads to design optimal quantizers for cardinality reduction. Then, we detail the proposed method to sparse a subspace, called sparse KLT (SKLT), and elaborate its performance tradeoffs and merit in Section V. Concluding remarks are presented in Section VI of the paper.

II. Mathematical Preliminaries

Historically speaking, transform coding (TC) of image and video signals has been one of the most popular applications of subspace methods where the desired dimension reduction is achieved through quantization of transform coefficients [1], [8], [9]. Original forward and inverse transform matrices are utilized in such a scenario. In contrast, there is a pressing need to sparse subspaces (transform matrices or vectors) rather than transform coefficients for efficiency required by emerging applications involving very large data sets to be processed in real-time. Therefore, researchers have developed several popular methods to answer such a need. The following subsections highlight the analytical framework used in the proposed method to sparse subspaces [3], [5], [12].

A. Transform Coding

A discrete-time orthonormal transform (subspace) is comprised of a set of linearly independent N sequences (vectors), $\{\phi_k(n)\}$ on the interval $0 \leq n \leq N-1$ satisfying the inner product properties [1]

$$\sum_{n=0}^{N-1} \phi_k(n)\phi_l^*(n) = \delta_{k-l} = \begin{cases} 1, & \text{k=l} \\ 0, & otherwise \end{cases} \tag{1}$$

where n is the time variable. In matrix form, the basis sequences $\boldsymbol{\phi}_k = \{\phi_k(n)\}$ are the rows of the transform matrix

$$\Phi = [\phi_k(n)] : \ k, n = 0, 1, \ldots, N-1 \tag{2}$$

with the matrix orthonormality stated as

$$\Phi\Phi^{-1} = \Phi\Phi^{*\mathbf{T}} = \mathbf{I} \tag{3}$$

where $*\mathrm{T}$ indicates the conjugated and transposed version of a matrix, and $\mathbf{I}$ is $N \times N$ identity matrix. Forward transform of a vector $\mathbf{x}$ is defined as

$$\boldsymbol{\theta} = \Phi\mathbf{x} \tag{4}$$

where $\boldsymbol{\theta}$ is the transform coefficient vector. Similarly, inverse transform operator applied to $\boldsymbol{\theta}$ perfectly reconstructs (represents) the signal vector as

$$\mathbf{x} = \Phi^{-1}\boldsymbol{\theta} = \Phi^{*\mathrm{T}}\boldsymbol{\theta} \tag{5}$$

Quantization of coefficients in the transform domain, called transform coding (TC), is defined as

$$\widehat{\boldsymbol{\theta}} = Q\{\boldsymbol{\theta}\} \tag{6}$$

Then, reconstructed signal with quantized coefficient vector $\widehat{\boldsymbol{\theta}}$ is expressed as

$$\widehat{\mathbf{x}} = \Phi^{*\mathrm{T}}\widehat{\boldsymbol{\theta}} \tag{7}$$

The mean square error between the original and reconstructed signal due to quantization of coefficients is written as [1]

$$\sigma^2_{\epsilon,TC} = \frac{1}{N}E\{\widetilde{\mathbf{x}}^{\mathbf{T}}\widetilde{\mathbf{x}}\} \tag{8}$$

for zero mean signal $\mathbf{x}$ where the quantization error $\widetilde{\mathbf{x}} = \mathbf{x} - \widehat{\mathbf{x}}$. Similarly, the mean square error between the original and quantized coefficients in the transform domain is calculated as

$$\sigma^2_{q,TC} = \frac{1}{N}E\{\widetilde{\boldsymbol{\theta}}^{\mathbf{T}}\widetilde{\boldsymbol{\theta}}\} = \frac{1}{N}\sum_{k=0}^{N-1}\sigma^2_{q_k,TC} \tag{9}$$

where $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}$, and $\sigma^2_{q_k,TC} = E\{|\widetilde{\boldsymbol{\theta}}_k|^2\}$ is the variance of the quantization error for the kth coefficient. Hence, $\sigma^2_{\epsilon,TC} = \sigma^2_{q,TC}$ for an orthonormal transform (subspace) [1].

Transform coding (TC) aims to achieve dimension reduction by repacking signal energy unevenly among the minimum possible transform coefficients. The transform coefficients of a signal are quantized for lossy compression (entropy reduction) where most become negligible and replaced by zero in a typical scenario. The benefit of TC over pulse code modulation (PCM) depends on the covariance properties of a given random vector process and has been studied in the literature [1], [8], [9], [19], [22]. It is noted that coefficient sparsity is the main objective in TC, and there is no interest to sparse (quantize) original projection and representation subspaces.

Lloyd-Max [17], [18] quantizer is designed based on the mean square error (mse) criterion for a given probability density function (pdf). In TC, it defines optimal quantizer intervals (bins) and their bin representation (quanta) values according to the pdf of the kth transform coefficient $\boldsymbol{\theta}_k$ in order to minimize $\sigma^2_{q_k,TC}$ with the constraint $\sigma^2_{q_k} = \sigma^2_{q_l}\ \forall k,l$. This quantization process is repeated for all transform coefficients [1]. In contrast for sparse representation, we use midtread pdf-optimized (Lloyd-Max) quantizers with adjustable zero-zone to sparse subspaces with zero valued vector components. We summarize pdf-optimized quantizer in the next section.

1) pdf-Optimized Midtread Quantizer: Quantizers (Q) may be categorized as midrise and midtread [8]. Midtread quantizer is preferred for applications requiring entropy reduction and noise filtering (or sparsity) simultaneously [22]. In this paper, we utilize a midtread quantizer type to quantize each basis function (components of each vector) of a transform to achieve sparse representation.

A celebrated design method to calculate optimum intervals (bins) and representation (quanta) values of a quantizer for the given input signal pdf, so called pdf-optimized quantizer, was independently proposed by Max and Lloyd [17], [18]. It assumes a random information source X with zero-mean and a known pdf function $p(x)$. Then, it minimizes quantization error in the mse sense and also makes sure that all bins of a quantizer have the same level of representation error. The quantization error of an L-bin pdf-optimized quantizer is expressed as follows

$$\sigma^2_q = \sum_{k=1}^{L}\int_{x_k}^{x_{k+1}}(x-y_k)^2p(x)dx \tag{10}$$

where quantizer bin intervals, $[x_k, x_{k+1}]$, and quanta values, y_k, are calculated iteratively. The necessary conditions for an mse based pdf-optimized quantizer are given as [17], [18]

$$\begin{aligned}\frac{\partial\sigma^2_q}{\partial x_k} &= 0; \quad k = 2,3,\ldots,L\\ \frac{\partial\sigma^2_q}{\partial y_k} &= 0; \quad k = 1,2,3,\ldots,L\end{aligned} \tag{11}$$

leading to the optimal unequal intervals and resulting quanta values as

$$x_{k,opt} = \frac{1}{2}(y_{k,opt} + y_{k-1,opt}); \quad k = 2,3,\ldots,L \tag{12}$$

$$y_{k,opt} = \frac{\int_{x_k}^{x_{k+1,opt}} xp(x)dx}{\int_{x_k}^{x_{k+1,opt}} p(x)dx}; \quad k = 1,2,\ldots,L \tag{13}$$

where $x_{1,opt} = -\infty$ and $x_{L+1,opt} = \infty$. Sufficient condition to avoid local optimum in (11) is the log-concavity of the pdf function $p(x)$. Log-concave property holds for Uniform, Gaussian and Laplacian pdf types [8]. The representation point (quantum) of a bin in such a quantizer is its centroid that minimizes the quantization noise for the interval. We are interested in pdf-optimized quantizers with adjustable zero-zone, odd L or midtread quantizer, to sparse (quantize) eigenvectors of an eigensubspace. One can adjust zero-zone(s) of the quantizer(s) to achieve the desired level of sparsity in transform matrix with the trade-off of resulting imperfectness in subspace orthogonality and explained variance. We will present design examples by using the proposed technique to sparse subspaces in the following section.

The discrepancy between input and output of a quantizer is measured by the signal-to-quantization-noise ratio (SQNR) [19]

$$SQNR(dB) = 10\log_{10}\left(\frac{\sigma^2_x}{\sigma^2_q}\right) \tag{14}$$

where σ^2_x is the variance of an input with zero-mean and known pdf type, and expressed as

$$\sigma^2_x = \int_{-\infty}^{\infty} x^2p(x)dx \tag{15}$$

The first order entropy (rate) of the output for an L-level quantizer with such an input is calculated as [19], [23]

$$\begin{aligned}H &= -\sum_{k=1}^{L}P_k\log_2 P_k\\ P_k &= \int_{x_k}^{x_{k+1}}p(x)dx.\end{aligned} \tag{16}$$

2) Optimum Bit Allocation Among Transform Coefficients: In TC, the method to allocate the allowable total bit rate R

among multiple transform coefficients (information sources) performs an important task. Transform coefficient variances σ_k^2 (or eigenvalues λ_k in KLT) are desired to be maximally uneven in order to achieve dimension reduction in TC. Hence, optimum bit allocation algorithm assigns bit rate R_k for quantization of coefficient θ_k in a way that makes the quantization error for each coefficient to be equal ($\sigma_{q_0}^2 = \sigma_{q_1}^2 = \ldots = \sigma_{q_{N-1}}^2$) [1]. Then, the number of levels for the kth quantizer, for coefficient θ_k, is found as $L_k = 2^{R_k}$.

Rate-distortion theory states that, the quantization error variance is expressed as [19] $\sigma_{q_k}^2 = f(R_k)\sigma_k^2$ where $f(R_k) = \gamma_k 2^{-2R_k}$ and σ_k^2 are the quantizer distortion function for a unit variance input and variance of the kth coefficient, respectively. γ_k depends on the pdf type of information source θ_k and also called *fudge factor*. It is shown with the assumption that all coefficients have the same pdf type, optimum bit rates R_k allocated among multiple information sources for the given total bit budget of *R*are calculated as [1]

$$R_k = R + \frac{1}{2}\log_2 \frac{\sigma_k^2}{\left(\prod_{i=0}^{N-1}\sigma_i^2\right)^{\frac{1}{N}}} \tag{17}$$

where $R = \sum_{k=0}^{N-1} R_k$.

Optimum bit allocation for the coefficient θ_k may yield a negative real number R_k. It implies that representing $\boldsymbol{\theta}_k$ even by zero causes a quantization error less than constrained coefficient distortion $\sigma_k^2 = \sigma_l^2$. Hence, a reduction of one dimension is achieved in the quantized signal representation. Note that $L_k = 2^{R_k}$ needs to be a positive integer number. Therefore, optimum bit allocation is an iterative process in its implementation.

B. Quantization of Transform Matrix (Subspace) for Sparse Representation

In TC, sparsity in transform coefficients is desired. In contrast, any sparse transform including KLT aims to sparse subspace (transform matrix) where values of basis vector components are important and interpreted as loading coefficients in some applications [24]–[29]. Quantization of a given subspace with an optimally designed single quantizer Q, or a set of quantizers $\{Q_k\}$ in the case of quantizing each basis function (vector) independently, is defined as

$$\widehat{\Phi} = Q(\Phi) \tag{18}$$

In this case, Q is a pdf-optimized midtread quantizer designed for the entire transform matrix. Then, transform coefficients are obtained by using the quantized matrix

$$\hat{\boldsymbol{\theta}} = \widehat{\Phi}\mathbf{x} \tag{19}$$

Unlike in TC, coefficients are not quantized in sparse representation methods unless desired for the given application. Instead, coefficients of the projection onto quantized subspace for a given signal vector are obtained. As in TC, quantization error equals to reconstruction error, both in mse, when the signal is reconstructed as (7). Mean squared quantization error due to sparsity of subspace is expressed as

$$\sigma_{q,S}^2 = \frac{1}{N^2}\sum_{k=0}^{N-1} \widetilde{\boldsymbol{\phi}_k}^{\mathbf{T}}\widetilde{\boldsymbol{\phi}_k} \tag{20}$$

where $\widetilde{\boldsymbol{\phi}_k} = \boldsymbol{\phi}_k - \widehat{\boldsymbol{\phi}_k}$.

III. Eigenanalysis of AR(1) Process

An eigenvalue λ and its paired eigenvector $\boldsymbol{\phi}$ of an $N \times N$ matrix $\boldsymbol{\Phi}$ satisfy the matrix equation [1]

$$\begin{aligned} \mathbf{R}_x\boldsymbol{\phi} &= \lambda\boldsymbol{\phi} \\ \mathbf{R}_x\boldsymbol{\phi} - \lambda\mathbf{I}\boldsymbol{\phi} &= (\mathbf{R}_x - \lambda\mathbf{I})\boldsymbol{\phi} = 0 \end{aligned} \tag{21}$$

such that $(\mathbf{R}_x - \lambda\mathbf{I})$ is singular. Namely,

$$\det(\mathbf{R}_x - \lambda\mathbf{I}) = 0 \tag{22}$$

$\mathbf{R}_x$ of AR(1) process, described next in Section III-A, is a real and symmetric matrix, and its eigenvectors are linearly independent. Thus, this determinant is a polynomial in λ of degree N, (22) has N roots and (21) has N solutions for $\boldsymbol{\phi}$ that result in eigenpair set $\{\lambda_k, \boldsymbol{\phi}_k\}$; $0 \le k \le N-1$. Therefore, the eigendecomposition of $\mathbf{R}_x$ is expressed as [1]

$$\mathbf{R}_x = \mathbf{A}_{\mathbf{KLT}}^{\mathbf{T}}\boldsymbol{\Lambda}\mathbf{A}_{\mathbf{KLT}} = \sum_{k=0}^{N-1}\lambda_k\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}} \tag{23}$$

where $\boldsymbol{\Lambda} = diag(\lambda_k)$; $k = 0, 1, \ldots, N-1$, and kth *column* of $\mathbf{A}_{\mathbf{KLT}}^{\mathbf{T}}$ matrix is the kth eigenvector $\boldsymbol{\phi}_k$ of $\mathbf{R}_x$ with the corresponding eigenvalue λ_k. Note that $\{\lambda_k = \sigma_k^2 = \boldsymbol{\phi}_k^{\mathrm{T}}\mathbf{R}_x\boldsymbol{\phi}_k\}\forall k$, for the given $\mathbf{R}_x$ where σ_k^2 is the variance of the kth transform coefficient, θ_k.

A. AR(1) Process

Autoregressive discrete process with order one, AR(1), is a first approximation to many natural signals like images. Therefore, it is widely used for performance analysis and comparison of signal processing methods in the literature. AR(1) signal is modeled as [1]

$$x(n) = \rho x(n-1) + \xi(n) \tag{24}$$

where $\xi(n)$ is a white noise sequence with zero-mean and variance σ_ξ^2, $E\{\xi(n)\xi(n+k)\} = \sigma_\xi^2\delta_{n-k}$. The first order correlation coefficient is in the range of $-1 < \rho < 1$ and defined as follows

$$\begin{aligned} \rho &= \frac{R_{xx}(1)}{R_{xx}(0)} \\ &= \frac{E\{x(n)x(n+1)\}}{E\{x(n)x(n)\}} \end{aligned} \tag{25}$$

The variance of $x(n)$ is calculated as $\sigma_x^2 = (1-\rho^2)^{-1}\sigma_\xi^2$. Hence, autocorrelation sequence of $x(n)$ for AR(1) process is expressed as

$$R_{xx}(k) = E\{x(n)x(n+k)\} = \sigma_x^2\rho^{|k|}; \quad k = 0, \pm1, \pm2, \ldots \tag{26}$$

The resulting Toeplitz correlation matrix of size $N \times N$ is shown to be in the form

$$\mathbf{R}_x = \sigma_x^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix} \quad (27)$$

B. Closed-Form Expressions for Eigenvectors and Eigenvalues of AR(1) Process

The eigenvalues of $\mathbf{R}_x$ for an AR(1) process defined in (27) are expressed in the closed-form as [29], [30]

$$\sigma_k^2 = \lambda_k = \frac{1-\rho^2}{1-2\rho\cos(\omega_k)+\rho^2}; 0 \le k \le N-1 \quad (28)$$

where $\{\omega_k\}$ are the positive roots of the transcendental equation

$$\tan(N\omega) = -\frac{(1-\rho^2)\sin(\omega)}{\cos(\omega)-2\rho+\rho^2\cos(\omega)} \quad (29)$$

that is rewritten as

$$\left[\tan\left(\omega\frac{N}{2}\right)+\gamma\tan\left(\frac{\omega}{2}\right)\right]\left[\tan\left(\omega\frac{N}{2}\right)-\frac{1}{\gamma}\cot\left(\frac{\omega}{2}\right)\right]=0$$
$$\gamma = \frac{(1+\rho)}{(1-\rho)}, \quad (30)$$

The resulting KLT kernel for matrix of size $N \times N$ is expressed as [29], [30]

$$\mathbf{A}_{KLT} = [A(k,n)] = c_k \sin\left[\omega_k\left(n-\frac{N-1}{2}\right)+\frac{(k+1)\pi}{2}\right]$$
$$c_k = \left(\frac{2}{N+\lambda_k}\right)^{\frac{1}{2}},\ 0 \le k,n \le N-1 \quad (31)$$

The roots of the transcendental tangent equation in (30), $\{\omega_k\}$, are required in the KLT kernel expressed in (31). An efficient root finding method for explicit solutions of transcendental equations including (30) was proposed in [29]. That method leads to an explicit KLT matrix kernel for an AR(1) process as given in (31).

Note that $\{\lambda_k\}$ are sorted in descending order after the eigenvalues and eigenvectors are calculated. Therefore, first principal component (PC1) is placed in the first column of $\mathbf{A}_{KLT}$ matrix where $k = 0$.

IV. Quantization of Eigen Subspace for AR(1) Process

In Section II-A, we introduced a framework to sparse a subspace by using pdf-optimized zero-zone quantizers. In this section, we investigate and model probability density function (pdf) (or histogram) of eigenvector components (PC loadings) for the Toeplitz correlation matrix of AR(1) source expressed in (31). We design a pdf-optimized zero-zone quantizer for each eigenvector that is being sparsed. One might also use a single quantizer for the entire eigen matrix in order to reduce implementation cost. Rate-distortion performance of such quantizers is evaluated. We present performance comparisons of the proposed sparse KLT (SKLT) method with ST [4], SPCA [4], DSPCA [5], and SPC [16] methods in terms of non-sparsity (NS) and variance loss (VL) metrics in the following section.

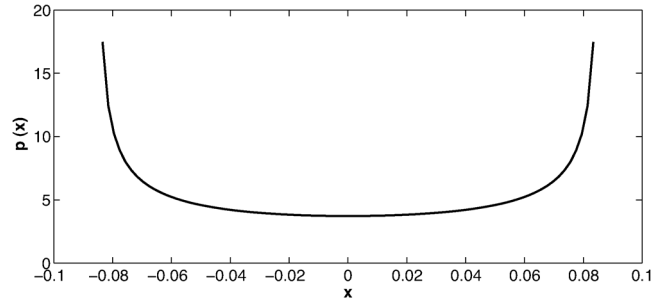


Fig. 1. Probability density function of arcsine distribution for $a = -0.0854$ and $b = 0.0854$. Loadings of second PC for AR(1) signal source with $\rho = 0.9$ and $N = 256$ are fitted to arcsine distribution by finding minimum and maximum values in the PC.

A. Probability Density Functions (pdf) of Eigenvector Components

1) Arcsine Distribution of Continuous Sinusoidal Function: In this section, we model the probability density of eigenvector components in order to design pdf-optimized quantizers to sparse them. Each eigenvector of AR(1) process is generated by a sinusoidal function as expressed in (31). Probability density function (pdf), with arbitrary support, of a continuous sinusoidal function is modeled as [31], [32]

$$p(x) = \frac{1}{\pi\sqrt{(x-a)(b-x)}} \quad (32)$$

where a and b define the support, $a \le x \le b$. Cumulative distribution function (cdf) of such a function type is of arcsine distribution and expressed as

$$P(x) = \frac{2}{\pi}\arcsin\left(\sqrt{\frac{x-a}{b-a}}\right) \quad (33)$$

Mean and variance of the arcsine distribution are calculated as

$$\mu = \frac{a+b}{2} \quad (34)$$
$$\sigma^2 = \frac{(b-a)^2}{8} \quad (35)$$

The pdf of arcsine distribution is symmetric and U-shaped. Fig. 1 shows the pdf of arcsine distribution with parameters $a = -0.0854$ and $b = 0.0854$. Log-concavity of a pdf $p(x)$ is the sufficient condition for the uniqueness of a pdf-optimized quantizer. However, arcsine distribution type has the log-convex property [33]. It is stated in [34] that for exponential sources and the sources with strictly log-convex pdfs, the quantizer intervals (bins) and their bin representation (quanta) values are globally optimum and unique. Therefore, pdf-optimized quantizers can be designed for arcsine distribution [17], [18]. Second principal component, $\boldsymbol{\phi}_1$, of AR(1) source for $\rho = 0.9$ and size of $N = 256$ is shown to be fit by arcsine distribution with $a = \min(\boldsymbol{\phi}_1) = -0.0854$ and $b = \max(\boldsymbol{\phi}_1) = 0.0854$, respectively. Minimum and maximum valued components of the kth eigenvector depend on ρ, ω_k and N as stated in (31). In order to maintain equal distortion levels among quantizers

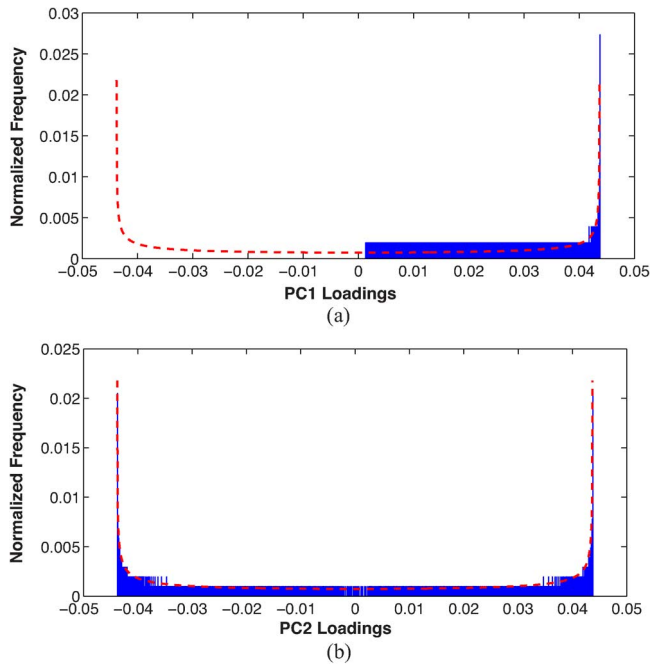


Fig. 2. Normalized histograms of (a) PC1 and (b) PC2 loadings for AR(1) signal source with $\rho = 0.9$ and $N = 1,024$. The dashed lines in each histogram show the probability that is calculated by integrating arcsine pdf for each bin interval.

to sparse eigenvectors, we calculated optimal intervals for zero-zones of pdf-optimized midtread quantizers. Thus, most of the small valued eigenvector components are likely to be quantized as zero.

2) Eigenvector Component Histograms for AR(1) Process: Fig. 2(a) and Fig. 2(b) display the normalized histograms of the first and second eigenvector components (PC1 and PC2 loading coefficients) for AR(1) process with $\rho = 0.9$ and $N = 1,024$. The value of N is selected large enough to generate proper histograms. The intervals of the histograms, Δ_k, are set as $\Delta_k = \frac{\max(\boldsymbol{\phi}_k) - \min(\boldsymbol{\phi}_k)}{N}$ where $\boldsymbol{\phi}_k$ is kth eigenvector. The dashed lines in each normalized histogram show the probability that is calculated by integrating the pdf of arcsine distribution in (32) for each bin interval. The histogram displayed in Fig. 2(a) has only one side of the arcsine pdf as expected from (31). In contrast, Fig. 2(b) displays the histogram with complete arcsine pdf shape. These figures confirm arcsine distribution type for eigenvector components of an AR(1) process.

B. Rate-Distortion Performance of Arcsine pdf-Optimized Zero-Zone Quantizer

In this section, we investigate the rate-distortion performance of arcsine pdf-optimized zero-zone quantizer. Rate of quantizer output is calculated by using first order entropy as defined in (16). Distortion caused by the quantizer is calculated in mse and represented in SQNR as defined in (14). Fig. 3 displays rate-distortion performance of such a quantizer with $L = 65$. It is observed that the performance of such a quantizer does not improve significantly for $L > 65$. Therefore, as a design step, we used $L = 65$ for the baseline quantizer where original zero-zone was widened by combining the adjacent bins. Hence,

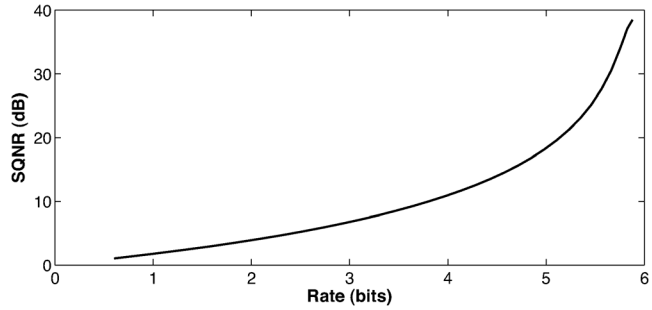


Fig. 3. Rate (bits)-distortion (SQNR) performance of zero mean and unit variance arcsine pdf-optimized quantizer for $L = 65$ bins. Distortion level is increased by combining multiple bins around zero in a larger zero-zone.

distortion level is increased by increasing the zero-zone of the quantizer for more sparsity where rate decreases, accordingly. One may design a midtread quantizer with zero-zone for each eigenvector (PC) or for the entire eigen matrix to achieve the desired level of matrix (subspace) sparsity [17], [18].

C. A Simple Method for Sparse KLT

In this section, we explain the proposed method to sparse eigen subspace of AR(1) process through a design example. The values of relevant parameters for the example are tabulated in Table I. The steps of design are summarized as follows.

1) First order correlation coefficient ρ is calculated from available data set as described in (25). Assume that $\rho = 0.9$ for the given example with $N = 256$.
2) Correlation matrix $\mathbf{R_x}$ for the measured ρ is constructed by using (27).
3) Eigenvalues $\{\lambda_k\}$ and corresponding eigenvectors $\{\boldsymbol{\phi}_k\}$ of $\mathbf{R_x}$ are calculated from (28) and (31), respectively. Then, eigenvalues are sorted in descending order and corresponding eigenvectors are placed in the eigenmatrix. Thus, $\boldsymbol{\phi}_0$ is the first eigenvector (PC1) and $\boldsymbol{\phi}_1$ is the second one (PC2), and so forth. Eigenvalues of first eleven eigenvectors (principal components) are listed in Table I. These eigenvectors explain 57.2% of the total variance. Due to limited space, only the variable values of SKLT for these eigenvectors are tabulated. Values of $\{\omega_k\}$ that are used to calculate each eigenvalue and corresponding eigenvector also shown in Table I. We used the root finding algorithm reported in [29].
4) PC loading coefficients (eigenvector components) are fitted to arcsine distribution by calculating $\{a_k = \min(\boldsymbol{\phi}_k)\}\forall k$ and $\{b_k = \max(\boldsymbol{\phi}_k)\}\forall k$. Then, variances $\left\{\sigma_k^2 = \frac{(b_k - a_k)^2}{8}\right\}\forall k$ are calculated by using (35). Table I also tabulates $\{a_k\}$, $\{b_k\}$ and $\{\sigma_k^2\}$ of eigenvectors.
5) For a given total rate R, $\{R_k\}$ are calculated by plugging $\{\sigma_k^2\}$ in optimum bit allocation equation given in (17). Then, quantizer levels $\{L_k\}$ are calculated as $\{L_k = 2^{R_k}\}\forall k$ and rounded up to the closest odd integer number. R is the sparsity tuning parameter of SKLT. As in all of the sparse PCA methods, R for a given sparsity has to be determined with cross-validation. Table I displays calculated rates and quantizer levels for the total rate of $R = 5.7$.

TABLE I
RELEVANT PARAMETERS OF SKLT METHOD FOR THE FIRST ELEVEN PCS OF AR(1) SOURCE WITH $\rho = 0.9$ AND $N = 256$. FIRST ELEVEN PCS EXPLAIN 57.2% OF THE TOTAL VARIANCE

	ω	λ	a	b	σ^2	R	L_k	S_k
PC1	0.0114	18.77	-0.0853	0.0853	0.0036	5.6546	51	26
PC2	0.0229	18.14	-0.0853	0.0853	0.0036	5.6563	51	28
PC3	0.0344	17.17	-0.0856	0.0856	0.0037	5.6588	51	40
PC4	0.0459	15.97	-0.0857	0.0857	0.0037	5.6620	51	34
PC5	0.0575	14.64	-0.0860	0.0860	0.0037	5.6655	51	36
PC6	0.0691	13.29	-0.0862	0.0862	0.0037	5.6691	51	38
PC7	0.0808	11.97	-0.0864	0.0864	0.0037	5.6725	51	42
PC8	0.0925	10.73	-0.0866	0.0866	0.0037	5.6754	51	42
PC9	0.1043	9.60	-0.0868	0.0868	0.0038	5.6790	51	40
PC10	0.1162	8.58	-0.0869	0.0869	0.0038	5.6819	51	36
PC11	0.1281	7.67	-0.0871	0.0870	0.0038	5.6835	51	44

6) For this design example, $L = 65$ level pdf-optimized zero-zone quantizer of arcsine distribution with zero mean and unit variance is used as the starting point. Then, several adjacent bins around zero are combined to adjust zero-zone for the desired sparsity level. For kth eigenvector, pre-designed $L = 65$ level pdf-optimized zero-zone quantizer is converted to $L_k \leq L$ level zero-zone quantizer.
7) PC loadings (eigenvector components) are normalized to have zero mean and unit variance, $\left\{\boldsymbol{\phi}_k = \frac{(\boldsymbol{\phi}_k - mean(\boldsymbol{\phi}_k))}{std(\boldsymbol{\phi}_k)}\right\} \forall k$ where *mean* and *std* are the mean and standard deviation of eigenvector components, respectively. Quantized (sparsed) eigenvectors are generated by applying quantization on eigenvectors of the original eigensubspace $\{\widehat{\boldsymbol{\phi}_k} = Q_k(\boldsymbol{\phi}_k)\}\forall k$. Number of zero components or sparsity level $\{S_k\}$ of quantized PCs for this example are also given in Table I.

Remark 1: Number of bins for pre-designed pdf-optimized quantizer is selected based on the quantization noise and implementation cost. The increase in signal-to-quantization noise (SQNR) of pdf-optimized zero-zone quanizer optimized for arcsine pdf with $L > 65$ is found not to be that significant.

1) Orthogonality Imperfectness and Subspace Sparsity: Sparsity achieved by quantization of PCs leads to orthogonality imperfectness. We present orthogonality imperfectness ϵ in mse with respect to allowable total rate *R* (desired sparsity level) for various AR(1) sources as defined

$$\epsilon = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{i=0}^{N-1}\left[\mathbf{I}(i,j) - \mathbf{K}(i,j)\right]^2 \tag{36}$$

where $\mathbf{I}$ is $N \times N$ identity matrix and $\mathbf{K} = \mathbf{A}\mathbf{A}^{*\mathbf{T}}$.

Fig. 4 displays the trade-off between subspace sparsity and loss of orthogonality for various AR(1) sources and $N = 256$. It is observed from the figure that the orthogonality imperfectness decreases almost linearly with increasing *R* as expected.

V. SPARSITY PERFORMANCE

Now, we compare performance of the proposed SKLT method with the ST [4], SPCA [4], DSPCA [5], and SPC [16] methods for AR(1) process, and also for empirical correlation matrix of stock returns in NASDAQ-100 index in the following subsections. In order to provide a fair comparison, sparsity levels of all methods considered here are tuned in a way that compared PCs have almost same number of non-zero components. In most cases, number of non-zero components of each PC in SKLT method are kept slightly lower than the others in order to show its merit under mildly disadvantageous test conditions.

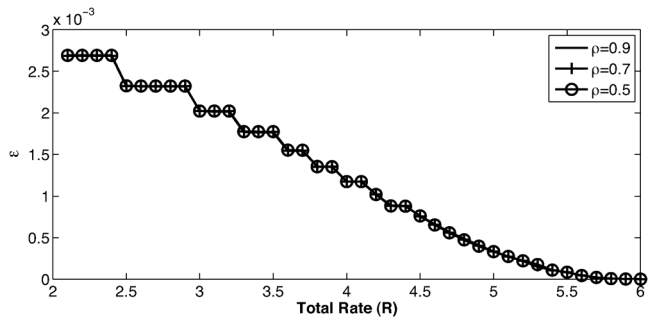


Fig. 4. Orthogonality imperfectness-rate (sparsity) trade-off of sparse eigen subspaces of three AR(1) sources with $N = 256$.

A. Sparsity of Eigen Subspace for AR(1) Process

The sparsity imposed on PCs may degrade the explained variance described in [5]. The explained variances (eigenvalues) of the PCs are calculated as $\{\lambda_k = \sigma_k^2 = \boldsymbol{\phi}_k^{\mathrm{T}}\mathbf{R}_x\boldsymbol{\phi}_k\}\forall k$ where $\boldsymbol{\phi}_k$ is the kth eigenvector for a given $\mathbf{R}_x$. For the sparsed PCs, new explained variances (eigenvalue) are calculated as $\{\widehat{\lambda}_k = \widehat{\sigma_k^2} = \widehat{\boldsymbol{\phi}_k}^{\mathbf{T}}\mathbf{R}_x\widehat{\boldsymbol{\phi}_k}\}\forall k$ where $\widehat{\boldsymbol{\phi}_k}$ is the kth sparse eigenvector. Then, the percentage of explained variance loss (VL) as a performance metric is defined as $\left\{V_k = \frac{(\lambda_k - \widehat{\lambda}_k)}{\lambda_k} \times 100\right\}\forall k$. Cumulative explained variance loss of first L number of PCs is also defined as $C_L = \sum_{k=1}^{N}\lambda_k - \sum_{k=1}^{L}\widehat{\lambda}_k$. In addition, we also used non-sparsity (NS) performance metric for comparison. It is defined as the percentage of non-zero components in a given sparsed eigenvector. Thus, the performance is measured as the variance loss for the given non-sparsity level [4], [5], [35]. We are unable to provide their comparative rate-distortion performance due to the lack of models to generate sparse PCs for all methods reported here.

Fig. 5 displays the variance loss (VL) measurements of sparsed first PC generated by SKLT, SPCA, SPC, ST and DSPCA methods with respect to non-sparsity (NS) for AR(1) source with $\rho = 0.9$ and $N = 256$. For SKLT, $L = 65$ level quantizer optimized for arcsine pdf with zero-mean and unit

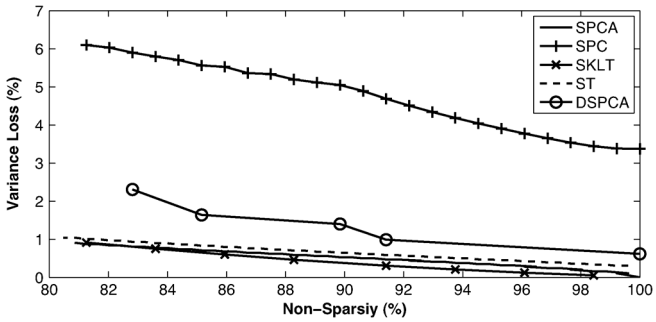


Fig. 5. Variance loss (VL) measurements of sparsed first PC generated by SKLT, SPCA, SPC, ST and DSPCA methods with respect to non-sparsity (NS) for AR(1) source with $\rho = 0.9$ and $N = 256$.

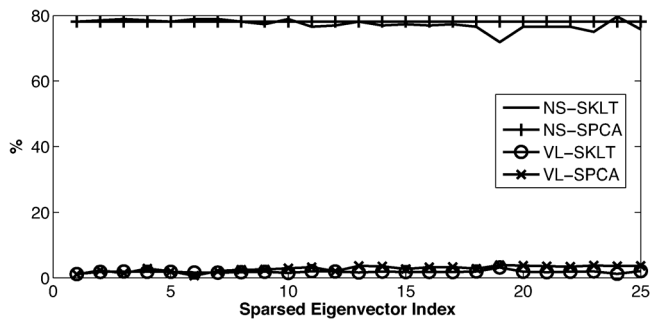


Fig. 6. Non-sparsity (NS) and variance loss (VL) measurements of sparsed eigenvectors generated by SKLT method and SPCA algorithm for AR(1) source with $\rho = 0.9$ and $N = 256$.

variance is used as the initial quantizer. The zero-zone width of the initial quantizer is adjusted for required sparsity as explained earlier. Then, the generated quantizer is employed. Fig. 5 shows that SKLT offers less variance loss than the other methods. SPCA provides competitive performance to SKLT. Fig. 6 displays non-sparsity (NS) and variance loss (VL) performance comparisons of sparse PCs generated by SKLT and by SPCA for the same AR(1) process. The original eigenvectors that explain 90% of the total variance are selected for sparsity comparison. Fig. 6 shows that the VL performance of SKLT is slightly better than SPCA. Note that NS of SKLT is slightly lower than SPCA in this comparison.

B. Sparsity of Eigenportfolios for NASDAQ-100 Index

Empirical correlation matrix of stock returns in an investment portfolio statistically measures their relative return performance and price inefficiencies. Eigendecomposition of empirical correlation matrix is a popular mathematical tool in finance employed for various tasks including eigenfiltering of measurement noise and creation of eigenportfolios for baskets of stocks [6], [7], [27], [36]. In this section, we present a finance application of the proposed method to sparse a subspace that may lead to trading cost reduction. Empirical correlation matrix for the end of day (EOD) stock returns for NASDAQ-100 index with $W = 30$ day time window ending on April 9, 2014 is measured [6], [7]. The vector of 100 stock returns in NASDAQ-100 index at time n is created as [7], [27]

$$\mathbf{r}(n) = [r_k(n)]; \quad k = 1, 2, \ldots, 100 \tag{37}$$

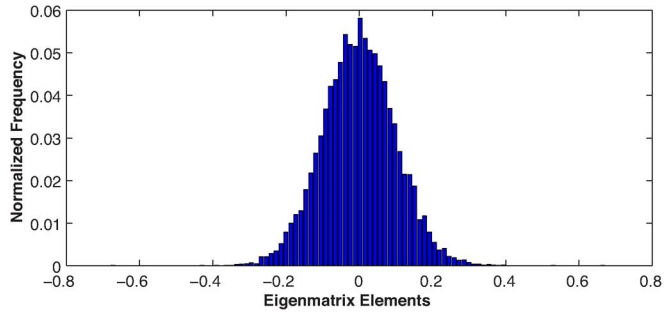


Fig. 7. Normalized histogram of eigenmatrix elements for empirical correlation matrix of end of day (EOD) returns for 100 stocks in NASDAQ-100 index with $W = 30$-day measurement window ending on April 9, 2014.

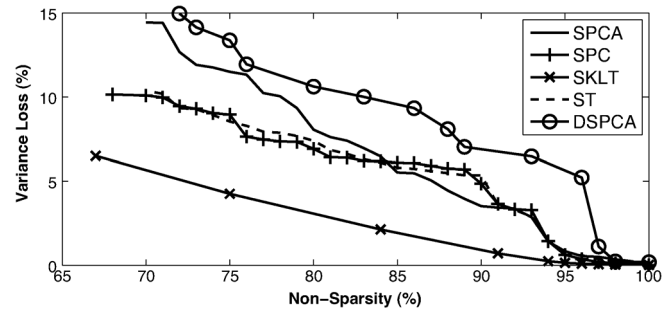


Fig. 8. Variance loss (VL) measurements of sparsed first PC generated by SKLT, SPCA, SPC, ST and DSPCA methods with respect to non-sparsity (NS) for empirical correlation matrix of end of day (EOD) returns for 100 stocks in NASDAQ-100 index with $W = 30$-day measurement window ending on April 9, 2014.

The empirical correlation matrix of returns at time n is expressed as

$$\mathbf{R}_E(n) \triangleq \left[E\left\{\mathbf{r}(n)\mathbf{r}^T(n)\right\}\right] = [R_{k,l}(n)] = \begin{bmatrix} R_{1,1}(n) & R_{1,2}(n) & \cdots & R_{1,100}(n) \\ R_{2,1}(n) & R_{2,2}(n) & \cdots & R_{2,100}(n) \\ \vdots & \vdots & \ddots & \vdots \\ R_{100,1}(n) & R_{100,2}(n) & \cdots & R_{100,100}(n) \end{bmatrix} \tag{38}$$

where the matrix elements

$$R_{k,l}(n) = E\left\{r_k(n)r_l(n)\right\} = \frac{1}{W}\sum_{m=0}^{W-1} r_k(n-m)r_l(n-m)$$

represent measured pairwise correlations for an observation window of W samples. The returns are normalized to be zero mean and unit variance, and $\mathbf{R}_E(n)$ is a real, symmetric and positive definite matrix. Now, we introduce eigendecomposition of $\mathbf{R}_E$ as follows

$$\mathbf{R}_E(n) = \mathbf{A}_{KLT}^T \mathbf{\Lambda} \mathbf{A}_{KLT} = \sum_{k=1}^{N} \lambda_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k^T \tag{39}$$

where $\{\lambda_k, \boldsymbol{\phi}_k\}$ are eigenvalue-eigenvector pairs.

The original eigenvectors that explain almost 90% of the total variance are selected for sparsity comparison. Due to simplicity, we employed a single quantizer for the SKLT method to sparse the entire eigenmatrix $\mathbf{A_{KLT}}$. It is optimized for the histogram of its elements as displayed in Fig. 7. It is observed to be a Gaussian pdf. Fig. 8 displays the variance loss (VL)

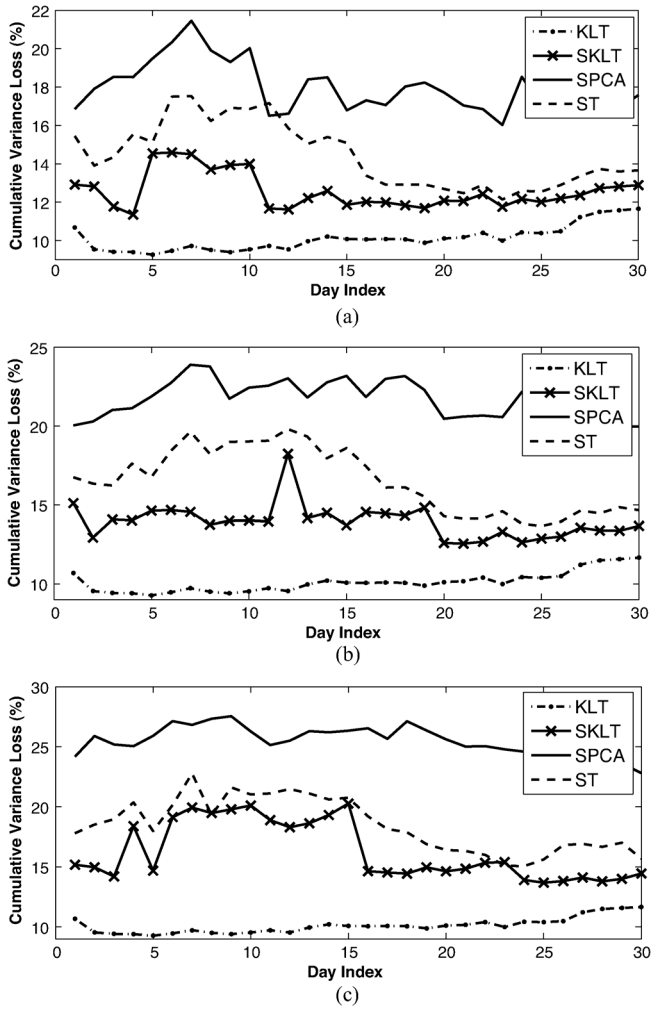


Fig. 9. Cumulative explained variance loss of first sixteen sparsed PCs generated from daily empirical correlation matrix of EOD returns during the time interval between April 9, 2014 and May 22, 2014 for 100 stocks in NASDAQ-100 index by using KLT, SKLT, SPCA and ST methods. Non-sparsity levels of 85%, 80% and 75% for each PC are forced in (a), (b) and (c), respectively, with $W = 30$-days.

measurements of sparsed first PC generated by SKLT, SPCA, SPC, ST and DSPCA methods with respect to non-sparsity (NS). Figure shows that SKLT offers less variance loss than compared methods. Similarly, Fig. 9 displays the cumulative explained variance loss of first sixteen sparsed PCs generated from daily empirical correlation matrix of EOD returns during the time interval between April 9, 2014 and May 22, 2014 for 100 stocks in NASDAQ-100 index by using KLT, SKLT, SPCA and ST methods. The measurement window of the last 30 days, $W = 30$, is used to calculate empirical correlation matrix for each day. Non-sparsity levels of 85%, 80% and 75% for each PC are forced in experiments displayed in Fig. 9(a), Fig. 9(b), and Fig. 9(c), respectively. The superior performance of the SKLT method is observed for this scenario as well where empirical correlation matrix of EOD returns changes every day.

The difference between the original $\mathbf{R}_E(n)$ and the modified correlation matrix $\widehat{\mathbf{R}_E(n)}$ due to sparsed eigenvectors is defined as $d^{\mathbf{R}} = \|\mathbf{R}_E(n) - \widehat{\mathbf{R}_E(n)}\|_2$ where $\|.\|_2$ is the norm-2 of a matrix. $d^{\mathbf{R}}_{SKLT} = 10.35$, $d^{\mathbf{R}}_{SPCA} = 17.15$, and $d^{\mathbf{R}}_{ST} = 17.38$ are measured for empirical correlation matrix of EOD returns for 100 stocks in NASDAQ-100 index with $W = 30$-days ending on April 9, 2014 with 85% non-sparsity level. Similarly, the distance between the original and the sparsed eigenmatrices is expressed as $d^{\mathbf{A}} = \|\mathbf{A}_{KLT} - \widehat{\mathbf{A}_{KLT}}\|_2$. The measured distances for the same experiment are $d^{\mathbf{A}}_{SKLT} = 0.23$, $d^{\mathbf{A}}_{SPCA} = 1.99$, and $d^{\mathbf{A}}_{ST} = 2.00$ for SKLT and ST methods, respectively. These objective measures also show that the proposed SKLT sparses eigen subspace of NASDAQ-100 index better than the ST and SPCA methods for the experiments presented here.

The component values of eigenvector $\{\phi_k\}$ are repurposed as the capital allocation coefficients to create the kth eigenportfolio for a group of stocks where the resulting coefficients $\{\theta_k\}$ are pairwise uncorrelated. These coefficients represent eigenportfolio returns in this application. Eigenportfolios are used in various investment and trading strategies [7], [37]. It is required to buy and sell certain stocks in amounts defined by the loading (capital allocation) coefficients in order to build and rebalance eigenportfolios in time. Some of the loading coefficients may have relatively small values where their trading cost becomes a practical concern for portfolio managers. Therefore, sparsing eigen subspace of an empirical correlation matrix $\mathbf{R}_E(n)$ may offer cost reductions in desired portfolio creation, maintenance and trading activity. In contrast, although theoretically appealing, the optimization algorithms like SPCA, DSPCA and SPC with constraints for forced sparsity (cardinality reduction of a set) may substantially alter intrinsic structures of the original eigenportfolios and their assets. Therefore, such a forced sparse representation might cause to significantly deviate from the measured empirical correlation matrix. Hence, financial performance degradations may happen in eigenportfolios generated by sparsity constrained optimization.

VI. Conclusions

The constrained optimization algorithms to generate sparse PCs are unable to guarantee good performance for an arbitrary covariance matrix due to the non-convex nature of the problem. In this paper, we propose a procedure to sparse subspaces. The proposed SKLT method utilizes the mathematical framework developed in rate-distortion theory for transform coding using pdf-optimized quantizers. The sparsity performance comparisons demonstrate the superiority of SKLT over the popular algorithms including ST, SPCA, DSPCA and SPC. SKLT is theoretically tractable, simple to implement and serves to sparse any subspace of interest.

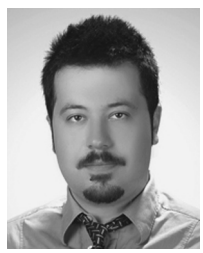
References

[1] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*. New York, NY, USA: Academic, 1992.

[2] J. Cadima and I. T. Jolliffe, "Loading and correlations in the interpretation of principle components," *J. App. Statist.*, vol. 22, no. 2, pp. 203–214, 1995.

[3] N. Trendafilov, I. T. Jolliffe, and M. Uddin, "A modified principal component technique based on the LASSO," *J. Comput. Graphic. Statist.*, vol. 12, no. 3, pp. 531–547, Sep. 2003.

[4] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graphic. Statist.*, vol. 15, no. 2, pp. 262–286, 2006.
[5] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM Rev.*, vol. 49, no. 3, pp. 434–448, Jul. 2007.
[6] M. U. Torun, A. N. Akansu, and M. Avellaneda, "Portfolio risk in multiple frequencies," *IEEE Signal Process. Mag., Special Issue on Signal Processing for Financial Applications*, vol. 28, no. 5, pp. 61–71, Sep. 2011.
[7] A. N. Akansu and M. U. Torun, *A Primer for Financial Engineering: Financial Signal Processing and Electronic Trading*. New York, NY, USA: Academic, 2015.
[8] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs, NJ, USA: Prentice-Hall Professional Technical Reference, 1984.
[9] R. J. Clarke, *Transform Coding of Images*. New York, NY, USA: Academic, 1985.
[10] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. London, U.K.: Instit. of Physics Publishing, 1998.
[11] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Norwell, MA, USA: Kluwer Academic, 1996.
[12] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc., Series B*, vol. 58, no. 1, pp. 267–288, 1996.
[13] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, "Sparse and stable Markowitz portfolios," *Proc. Nat. Acad. Sci.*, vol. 106, no. 30, pp. 12 267–12 272, Apr. 2009.
[14] R. Zass and A. Shashua, "Nonnegative sparse PCA," in *Proc. NIPS*, 2007, pp. 1561–1568, MIT Press.
[15] H. Shen and J. Z. Huang, "Sparse principal component analysis via regularized low rank matrix approximation," *J. Multivar. Anal.*, vol. 99, no. 6, pp. 1015–1034, Jul. 2008.
[16] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, Jul. 2009.
[17] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.
[18] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
[19] T. Berger, *Rate-Distortion Theory*. New York, NY, USA: Wiley, 2003.
[20] A. Hajnal and I. Juhasz, "Remarks on the cardinality of compact spaces and their Lindelof subspaces," *Proc. Amer. Math. Soc.*, vol. 51, no. 5, pp. 146–148, 1983.
[21] P. Boufounos and R. Baraniuk, "Quantization of sparse representations," in *Proc. Data Compress. Conf.*, Mar. 2007, pp. 378–378.
[22] C. A. Gonzales and A. N. Akansu, "A very efficient low-bit-rate subband image/video codec using shift-only PR-QMF and zero-zone linear quantizers," in *Proc. IEEE ICASSP*, Apr. 1997, vol. 4, pp. 2993–2996.
[23] H. Brusewitz, *Quantization With Entropy Constraint and Bounds to the Rate Distortion Function, Telecommunication Theory*. Stockholm, Sweden: Electrical Engineering, Royal Inst. of Technol., 1986.
[24] J. Ohlson and B. Rosenberg, "Systematic risk of the CRSP equal-weighted common stock index: A history estimated by stochastic-parameter regression," *J. Business*, vol. 55, no. 1, pp. 121–145, Jan. 1982.
[25] H. Mamaysky, M. Spiegel, and H. Zhang, "Estimating the dynamics of mutual fund alphas and betas," *Rev. Financ. Stud.*, vol. 21, no. 1, pp. 233–264, 2008.
[26] N. P. B. Bollen and R. E. Whaley, "Hedge fund risk dynamics: Implications for performance appraisal," *J. Finance*, vol. 64, no. 2, pp. 985–1035, 2009.
[27] M. U. Torun and A. N. Akansu, "Toeplitz approximation to empirical correlation matrix of asset returns: A signal processing perspective," *J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 319–326, Aug. 2012.
[28] H. Choi and H. Varian, "Predicting the present with Google Trends," *Econom. Rec.*, vol. 88, pp. 2–9, 2012.
[29] M. U. Torun and A. N. Akansu, "An efficient method to derive explicit KLT kernel for first-order autoregressive discrete process," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3944–3953, Aug. 2013.
[30] W. Ray and R. Driver, "Further decomposition of the Karhunen-Loeve series representation of a stationary random process," *IEEE Trans. Inf. Theory*, vol. 16, no. 6, pp. 663–668, Sep. 1970.
[31] K. Hejn, A. Pacut, and L. Kramarski, "The effective resolution measurements in scope of sine-fit test," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 1, pp. 45–50, Feb. 1998.
[32] N. Balakrishnan and V. B. Nevzorov, *A Primer on Statistical Distributions*. New York, NY, USA: Wiley, 2004.
[33] M. Bagnoli and T. Bergstrom, "Log-concave probability and its applications," *Econom. Theory*, vol. 26, no. 2, pp. 445–469, Aug. 2005.
[34] V. B. Yee, "Studies on the asymptotic behavior of parameters in optimal scalar quantization," Ph.D. dissertation, Univ. of Michigan, Ann Arbor, MI, USA, 2010.
[35] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc. B (Statist. Methodol.)*, vol. 67, no. 2, pp. 301–320, 2005.
[36] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. New York, NY, USA: Wiley, 1959.
[37] G. Chamberlain and M. Rothschild, "Arbitrage, factor structure and mean-variance analysis on large asset markets," *Econometrica*, vol. 51, no. 5, pp. 1281–1304, 1983.



Onur Yilmaz received his B.S. degree in Computer and Educational Sciences and M.S. degree in Computer Science, both from Ege University, Izmir, Turkey, in 2007 and 2011 respectively. Since 2011, he has been a Ph.D. candidate in the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology, Newark, NJ. His research interests include high performance digital signal processing and distributed systems.

Ali N. Akansu (F'08) received his B.S. degree from the Technical University of Istanbul, Turkey, M.S. and Ph.D. degrees from the Polytechnic University, Brooklyn, New York, all in Electrical Engineering. He has been with the Department of Electrical & Computer Engineering at the New Jersey Institute of Technology where he is Professor of Electrical & Computer Engineering. He was a Founding Director of the New Jersey Center for Multimedia Research and NSF Industry-University Cooperative Research Center for Digital Video. Dr. Akansu was the Vice President for Research and Development of IDT Corporation. He was the founding President and CEO of PixWave, Inc., and Senior VP for Technology Development of TV.TV (IDT subsidiaries). He had been on the boards of several companies and an investment fund. He visited David Sarnoff Research Center, IBM T.J. Watson Research Center, GEC-Marconi Electronic Systems Corp., and Courant Institute of Mathematical Sciences at the New York University. Dr. Akansu has published numerous articles and several books on his research work. His current professional and research interests include theories of signals and transforms, quantitative finance and high frequency trading, high performance DSP and computing.