18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

A Method to Sparse Eigen Subspace and Eigenportfolios

Onur Yilmaz
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, New Jersey, USA 07102
Email: onur.yilmaz@NJIT.edu

Ali N. Akansu
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, New Jersey, USA 07102
Email: akansu@NJIT.edu

***Abstract*—A new method to sparse eigen subspaces by using the pdf-optimized zero-zone quantizers is proposed. It is called sparse Karhunen-Loeve Transform (SKLT). The performance of the proposed method is presented for sparse representation of eigenportfolios generated from empirical correlation matrix of stock returns in NASDAQ-100 index. Performance results show that the proposed SKLT outperforms the popular algorithms to sparse eigen subspaces reported earlier in the literature.**

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

I. Introduction

Karhunen-Loeve Transform (KLT), also known as principal component analysis (PCA) and eigenanalysis, has been widely utilized in eigen filtering and dimension reduction applications. In some applications, each element of the principal components (eigenvectors) has special meaning that often requires further interpretation [1], [2], [3], [4]. Eigenvectors are repurposed for the generation and maintenance (rebalancing) of eigenportfolios for investment and trading strategies in finance. Therefore, non-zero loadings (elements) of a principal component (PC, or eigenvector) bring extra cost in such applications [4], [5].

Generating sparse PCs has been proposed in the literature. The straightforward method for sparsity is a simple thresholding [1]. It is easy to implement but it may cause unexpected distortion [1]. Regularization is the most popular method that has been utilized for sparsity. ℓ_0 (norm-0) regularizer leads to a sparse solution. Since it makes the optimization problem non-convex, ℓ_1 regularizer, so called lasso, is also widely used as an approximation [2], [6]. ℓ_1 regularizer based method was proposed in [7] for sparse portfolios. SCoTLASS [2] and SPCA [3] are the two prior methods that utilize the ℓ_1 and ℓ_2 regularizers for sparse approximation to PCs, respectively.

Although the sparse PCA is modeled in [2], [3] as an explained variance maximization problem using ℓ_1 and ℓ_2 regularizers, those frameworks still lead to a non-convex optimization problem due to other constraints. A convex relaxation method called SDP Relaxations for Sparse PCA (DSPCA) using semidefinite programming (SDP) was proposed as an approximation to the original problem [4]. Another lasso based approach, sparse PCA via regularized SVD (sPCA-rSVD), is proposed in [8]. Simulation results for certain cases show that sPCA-rSVD provides competitive results to SPCA. A variation of sPCA-rSVD, sparse principal components (SPC), that utilizes the penalized matrix decomposition (PMD) is proposed in [9]. It utilizes the lasso penalty for sparsity. Unfortunately, none of these methods result in desired sparsity with reasonable distortion regardless of their prohibitive computational cost for high dimensions. Moreover, the lack of mathematical framework to measure and adjust distortion with ease, 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 may cause unexpected distortion levels as called variance loss although it is easy to implement [1]. It performs better than SCoTLASS and slightly worse than SPCA [3]. Soft thresholding (ST) is another technique for sparse representation reported in [3]. Certain experiments show that ST offers slightly better performance than simple thresholding [3]. Therefore, threshold selection plays a central role in sparsity performance as expected.

In this paper, we propose a subspace sparsing framework that may be considered as an extension of the simple and soft thresholding methods [10], [11], [12], [13]. We approach the problem of sparse PCA from the quantization point of view and exploit the mathematical tools used in the source coding field [1], [10], [14], [15], [13]. The method utilizes the rate-distortion theory in order to measure the representation error (distortion) for the desired sparsity. We compare performances of the proposed method with the ST [3], SPCA [3], DSPCA [4] and SPC [9] using the metrics of non-sparsity (NS) and explained variance (EV) as described in [3], [4]. As an example, we sparse eigenportfolios of the stocks in NASDAQ-100 index by using this method and highlight its merit in the following sections of the paper.

The mathematical preliminaries are given in the next section. In Sec. III, the proposed sparse KLT method is summarized. The performance of SKLT is presented for sparse representation of eigenportfolios in Sec. IV. Concluding remarks are presented in Sec. V.

II. MATHEMATICAL PRELIMINARIES

Linearly independent N orthonormal discrete sequences (vectors), $\{\phi_k(n)\}$ $0 \le n \le N-1$, that form an orthonormal subspace satisfy the properties [10]

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

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

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

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.

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

is the forward transformation of a signal vector $\mathbf{x}$ where $\boldsymbol{\theta}$ is the transform coefficient vector. Similarly,

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

is the inverse transform operator applied to $\boldsymbol{\theta}$ that perfectly reconstructs the signal vector. In transform coding (TC), coefficients are quantized in the transform domain 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 reconstruction error in mean square error (mse) sense due to quantization of coefficients is written as [10]

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

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

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

where $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}$, and $\sigma^2_{q_k,TC} = E\left\{\left|\widetilde{\boldsymbol{\theta}}_k\right|^2\right\}$ is the variance of the quantization error for the kth coefficient. Hence, $\sigma^2_{\epsilon,TC} = \sigma^2_{q,TC}$ for an orthonormal transform that preserves signal energy in the subspace [10].

Sparsity in transform coefficients is desired in TC while the sparsity of transform matrix is desired for sparse representations including sparse KLT. Sparse transform aims to sparse subspace (transform matrix) where basis vector components are interpreted as loading coefficients in applications [16], [17], [18], [19], [20]. Quantization of a given subspace Φ with a quantizer Q, or a set of quantizers $\{Q_k\}$, is defined as

$$\widehat{\Phi} = Q(\Phi) \tag{10}$$

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 (for quantized subspace)

$$\hat{\boldsymbol{\theta}} = \widehat{\Phi}\mathbf{x} \tag{11}$$

In sparse representations, projection of a given signal vector onto quantized subspace is performed to obtain coefficients. Note that the orthogonality of the eigenspace (eigenmatrix) is compromised due to the quantization. Quantization error of the subspace equals to the reconstruction error of the signal vector, in mse, by using non-quantized inverse transform matrix. This error is expressed as

$$\sigma^2_{q,S} = \frac{1}{N^2}\sum_{k=0}^{N-1}\widetilde{\boldsymbol{\phi}_k}^{\mathbf{T}}\widetilde{\boldsymbol{\phi}_k} \tag{12}$$

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

In a sparse representation, pdf-optimized midtread quantizer(s) are specifically designed for eigenmatrix or each eigenvector. pdf-optimized quantizer minimizes quantization error in mse sense for the known function $p(x)$. All bins of a pdf-optimized quantizer have the same level of representation error [11], [12]. The quantization error of an L-bin pdf-optimized quantizer is expressed as follows

$$\sigma^2_q = \sum_{k=1}^{L}\int_{x_k}^{x_{k+1}}(x - y_k)^2\,p(x)dx \tag{13}$$

where quantizer bin intervals, x_k, and quanta values, y_k, are calculated iteratively. The necessary conditions for an mse based pdf-optimized quantizer are given as [11], [12]

$$\begin{aligned} \frac{\partial\sigma^2_q}{\partial x_k} &= 0;\ k = 2, 3, \dots, L \\ \frac{\partial\sigma^2_q}{\partial y_k} &= 0;\ k = 1, 2, 3, \dots, L \end{aligned} \tag{14}$$

leading to the optimal unequal intervals and resulting quanta values as

$$x_{k,opt} = \frac{1}{2}\left(y_{k,opt} + y_{k-1,opt}\right);\ k = 2, 3, \dots, L \tag{15}$$

$$y_{k,opt} = \frac{\int_{x_k}^{x_{k+1,opt}} xp(x)dx}{\int_{x_k}^{x_{k+1,opt}} p(x)dx};\ k = 1, 2, \dots, L \tag{16}$$

where $x_{1,opt} = -\infty$ and $x_{L+1,opt} = \infty$. Sufficient condition to avoid local optimum in (14) is the log-concavity of the pdf function $p(x)$. Log-concave property holds for Uniform, Gaussian and Laplacian pdf types [14]. 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) eigen vectors of an eigen subspace. We will present a practical design example 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) [13]

$$SQNR(dB) = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_q^2}\right) \quad (17)$$

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

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 p(x)dx \quad (18)$$

The first order entropy (rate) of the output for an L-level quantizer with such an input is calculated as [13], [21]

$$H = -\sum_{k=1}^{L} P_k \log_2 P_k \quad (19)$$

where $P_k = \int_{x_k}^{x_{k+1}} p(x)dx$.

III. Sparse KLT

In this section, we explain the proposed method sparse KLT (SKLT) to sparse eigen subspace. A detailed example of SKLT on a finance application is given in the next section. The steps of SKLT are summarized as follows.

1) For a given signal vector $\mathbf{x}$ with zero-mean and unit variance assumption, correlation matrix $\mathbf{R}$ is calculated. $\mathbf{R}$ is a real, symmetric and positive definite matrix.
2) Eigendecomposition of $\mathbf{R}$ is performed to obtain KLT transform matrix $\mathbf{A}_{KLT}$ that is comprised of N eigenvectors $\{\boldsymbol{\phi}_k\}$; $k = 1, 2, \ldots, N$ with the eigenvalues ordered in descending order as $\lambda_1 > \lambda_2 > \ldots > \lambda_N > 0$.
3) Probability density function (pdf) of eigen subspace (or each eigenvector) components is modeled by inspecting the components histogram in order to design initial pdf-optimized zero-zone quantizers Q_I.
4) L_I-level zero-zone quantizer Q_I optimized for modeled pdf is designed. The level L_I of the initial quantizer is determined based on the computational cost and signal-to-quantization noise (SQNR) of the quantizer. L_I has to be an odd number to create a zero-zone for the quantizer. Note that the rate of the quantizer is be calculated as
$$R = \log_2 L \quad (20)$$
5) Although the initially designed L_I-level pdf-optimized zero-zone quantizer Q_I has zero-zone that offers sparsity, the size of the zero-zone may not be large enough to achieve the desired level of sparsity for certain applications. For the purpose, zero-zone of Q_I is increased by inclusion of its neighboring bins that results in a new L-level zero-zone quantizer Q. L of new zero-zone quantizer that delivers the desired sparsity is calculated using cross-validation. Sparsity is basically tuned by increasing the size of zero-zone of the pre-designed L_I-level pdf-optimized zero-zone quantizers Q_I.
6) L-level quantizer Q is employed to quantize $\mathbf{A}_{KLT}$ shown as
$$\widehat{\mathbf{A}_{KLT}} = Q\left(\mathbf{A}_{KLT}\right) \quad (21)$$
One can also design quantizers specifically for each eigenvector in particular for high dimensions if the application needs different sparsity levels in each eigenvector. Then, quantization of each eigenvector is defined as $\left\{\widehat{\boldsymbol{\phi}_k} = Q_k\left(\boldsymbol{\phi}_k\right)\right\} \forall k$ where Q_k is the quantizer designed for k*th* eigenvector.
7) Then, the transformation is performed using quantized eigen subspace defined as
$$\hat{\boldsymbol{\theta}} = \widehat{\mathbf{A}_{KLT}}\mathbf{x} \quad (22)$$

IV. Sparse Representation of Eigenportfolios

In this section, we will employ SKLT described in Sec. III for a finance application and compare its performance with the ST [3], SPCA [3], DSPCA [4] and SPC [9] methods with respect to the non-sparsity (NS) and explained variance (EV) metrics.

A. Sparse Eigenportfolios for NASDAQ-100 Index

Pairwise correlations among stock returns populate the empirical correlation matrix that reveals important information about portfolio return and its risk. An important application of eigenanalysis for empirical correlation matrix is the creation of eigenportfolios for the given basket of stocks where elements of eigenvectors used as the capital allocation coefficients for each eigenportfolio [5], [19], [22]. Eigenportfolios are widely used in various investment and trading strategies [23]. 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 for the targeted risk levels. Some of the loading coefficients may have relatively small values where their trading cost becomes a practical concern for portfolio managers. Therefore, sparsing eigensubspace of an empirical correlation matrix may offer cost reductions in desired trading activity.

Empirical correlation matrix of the end of day (EOD) stock returns for NASDAQ-100 index with $W = 30$ day time window ending on April 9, 2014 is measured [5]. The vector of 100 stock returns at time n is created as [19]

$$\mathbf{r}(n) = \left[r_k(n)\right]; k = 1, 2, \ldots, 100 \quad (23)$$

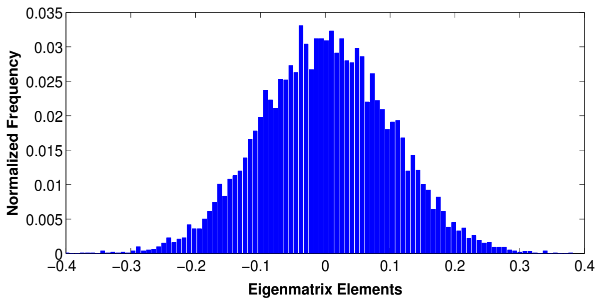


Fig. 1: Normalized histogram of eigenmatrix elements for empirical correlation matrix of end of day (EOD) returns for 100 stocks in NASDAQ-100 index with $W = 30$-day measurement window ending on April 9, 2014.

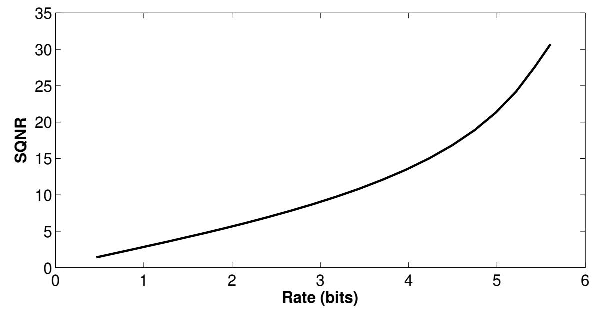


Fig. 2: Rate (bits)-distortion (SQNR) performance of zero mean and unit variance gaussian pdf-optimized quantizer for $L = 65$ bins. Distortion level is increased by combining multiple bins around zero in a larger zero-zone.

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

$$\begin{aligned} \mathbf{R}_E(n) &\triangleq \left[E\left\{\mathbf{r}(n)\mathbf{r}^T(n)\right\}\right] = [R_{k,l}(n)] \quad (24) \\ &= \begin{bmatrix} R_{1,1}(n) & R_{1,2}(n) & \cdots & R_{1,100}(n) \\ R_{2,1}(n) & R_{2,2}(n) & \cdots & R_{2,100}(n) \\ \vdots & \vdots & \ddots & \vdots \\ R_{100,1}(n) & R_{100,2}(n) & \cdots & R_{100,100}(n) \end{bmatrix} \end{aligned}$$

where the matrix elements

$$R_{k,l}(n) = E\left\{r_k(n)r_l(n)\right\} = \frac{1}{W}\sum_{m=0}^{W-1} r_k(n-m)r_l(n-m) \quad (25)$$

represent measured pairwise correlations for an observation window of $W = 30$ 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. Eigendecomposition of $\mathbf{R}_E$, is defined as follows

$$\mathbf{R}_E(n) = \mathbf{A}_{KLT}^T \mathbf{\Lambda} \mathbf{A}_{KLT} = \sum_{k=1}^{N} \lambda_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k^T \quad (26)$$

where $\{\lambda_k, \boldsymbol{\phi}_k\}$ are eigenvalue-eigenvector pairs. 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 row of $\mathbf{A}_{KLT}$ matrix.

The histogram of $\mathbf{A}_{KLT}$ matrix elements is inspected to model with a proper probability density function (pdf). It is observed to be a Gaussian pdf as displayed in Fig. 1. Now, we define the level L_I of the initial zero-zone quantizer optimized for Gaussian pdf. The level is set as $L_I = 65$ in this paper since the SQNR of the quantizer does not improve significantly for $L_I > 65$. In addition, it has a reasonable computational cost.

Fig. 2 displays the rate-distortion performance of Gaussian pdf-optimized zero-zone quantizer with respect to the size of zero-zone. Rate of quantizer output is calculated by using first order entropy defined as (19). One can calculate the level of the quantizer for the given rate by using (20). Distortion caused by the quantizer is calculated in mse and represented in SQNR as expressed in (17). In this figure, distortion level is increased when the zero-zone of the quantizer increases with more sparsity, and the rate decreases, accordingly [11], [12]. Thus, one can analytically calculate the resulting distortion for the required sparsity. Therefore, the proposed SKLT is a theoretically trackable method.

Although the initial value $L_I = 65$ offers some sparsity, the size of the zero-zone may not be large enough to provide the desired level of sparsity. For such a case, a new L-level quantizer, $L < L_I$, is created by combining multiple bins around the zero zone in a *new larger zero-zone*. The levels of new quantizers are calculated with cross-validation. One can also design different quantizers for each eigenvector if the application justifies it. Rate-distortion performance of the new L-level quantizer is calculated and displayed in Fig. 2.

A single quantizer is applied to the entire eigen subspace as $\widehat{\mathbf{A}_{KLT}} = Q\left(\mathbf{A}_{KLT}\right)$ or to one eigenvector (PC) $\left\{\widehat{\boldsymbol{\phi}_k} = Q_k\left(\boldsymbol{\phi}_k\right)\right\} \forall k$. The component values of sparsed eigenvectors $\left\{\widehat{\boldsymbol{\phi}_k}\right\}$ are repurposed as the capital allocation coefficients to create the k*th* sparse eigenportfolio. Therefore, transaction cost is reduced since some of the capital allocation coefficients are quantized as zero.

B. Performance Comparisons

In this section, we compare performance of SKLT with the ST [3], SPCA [3], DSPCA [4] and SPC [9] methods in terms of explained variance (EV) metric for the given nonsparsity (NS) level that is defined as the percentage of nonzero components in a given sparsed PC (eigenvector). The explained variance (eigenvalue) of the PCs are calculated as $\left\{\lambda_k = \sigma_k^2 = \boldsymbol{\phi}_k^{\mathrm{T}} \mathbf{R}_x \boldsymbol{\phi}_k\right\} \forall k$ where $\boldsymbol{\phi}_k$ is the k*th* PC (eigenvector) for a given correlation matrix $\mathbf{R}$. For the sparsed PCs, new explained variances (eigenvalue) are calculated as $\left\{\widehat{\lambda}_k = \widehat{\sigma}_k^2 = \widehat{\boldsymbol{\phi}_k}^{\mathbf{T}} \mathbf{R}_x \widehat{\boldsymbol{\phi}_k}\right\} \forall k$ where $\widehat{\boldsymbol{\phi}_k}$ is the k*th* sparse PC (eigenvector). We are unable to provide their comparative rate-

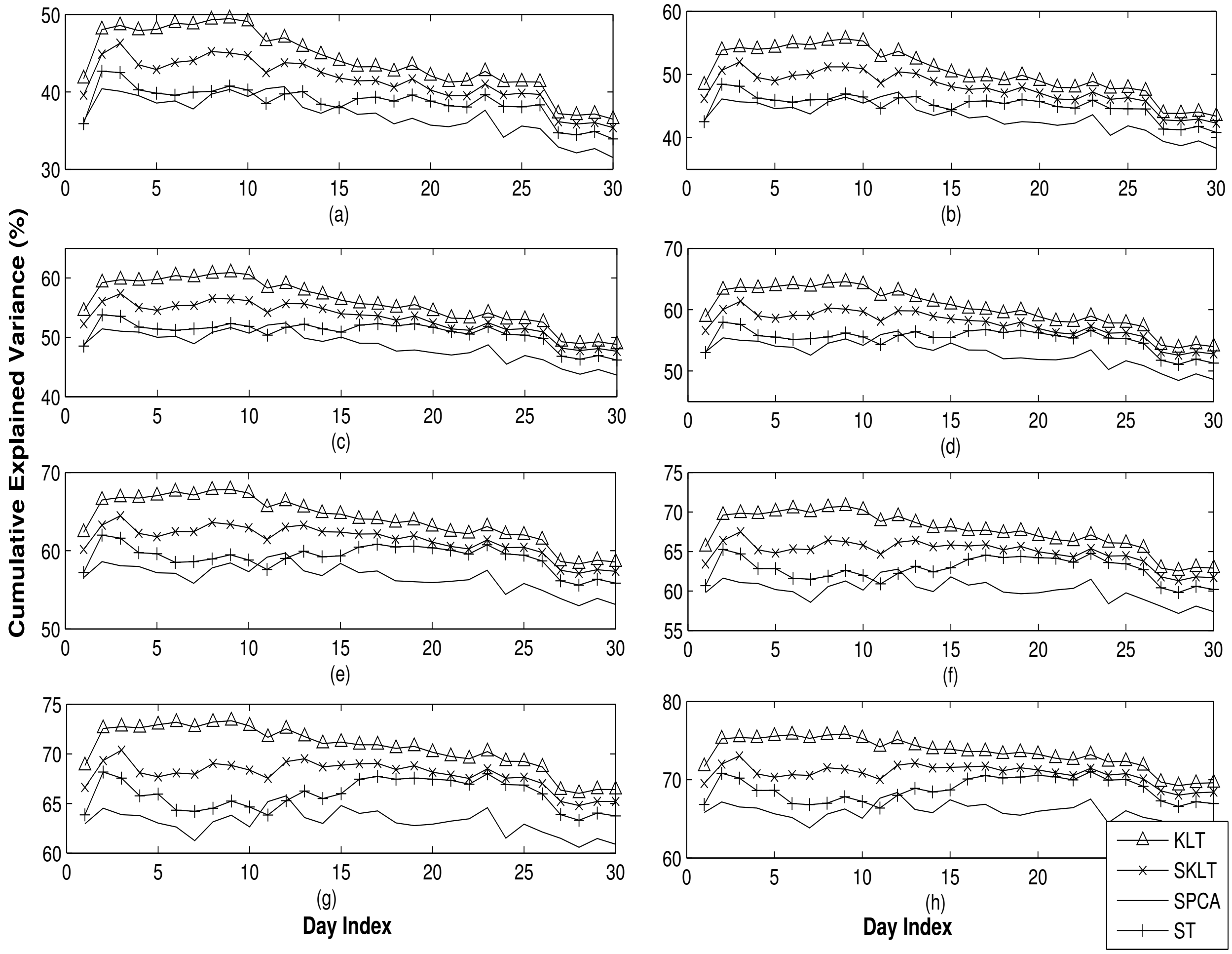


Fig. 4: (a) Explained variance of sparsed first PC and cumulative explained variance of sparsed (b) two PCs, (c) three PCs, (d) four PCs, (e) five PCs, (f) six PCs, (g) seven PCs, and (h) eight PCs generated daily from empirical correlation matrix of EOD returns between April 9, 2014 and May 22, 2014 for 100 stocks in NASDAQ-100 index by using KLT, SKLT, SPCA and ST methods. Non-sparsity level is set to 85% for each PCs.

distortion performance due to the lack of model to generate sparse PCs with other popular methods.

In this section, we assume that the required non-sparsity level is known for the desired cost reduction. In fact, the calculation of required non-sparsity level depends on the application. It is not a trivial task and beyond the scope of this paper. For the given non-sparsity level, we tuned the zero-zone size of 65-level Gaussian pdf-optimized midtread quantizer by combining multiple bins around zero in a larger zero-zone to sparse the eigenvectors of eigen subspace represented by $\mathbf{A}_{\mathbf{KLT}}$. Thus, lower level quantizers for the required non-sparsity level are created out of the pre-designed 65-level Gaussian pdf-optimized midtread quantizer. Note that the level of the quantizer is the sparsity tuning parameter for SKLT. For other methods, sparsity tuning parameters are set to the best possible values that provide the desired non-sparsity level. It is noted that the trade-off between the savings due to the forced sparsity of eigenmatrix and the profit-and-loss (PNL) performances of such sparsed eigenportfolios need to be quantified in such a real world investment scenario.

Fig. 3 displays the explained variance of sparsed first PC (PC1) generated by SKLT, DSPCA, SPCA, ST and SPC with respect to increasing non-sparsity level 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. It is clear that SKLT provides better explained variance than other popular methods for the different non-sparsity levels. Note that SKLT also provides better performance with respect to increased non-sparsity level for daily generated empirical correlation matrix of EOD returns between April 9, 2014 and May 22, 2014.

Fig. 4a displays the explained variance of sparsed first PC

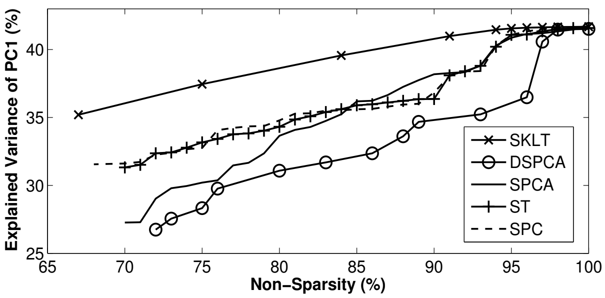
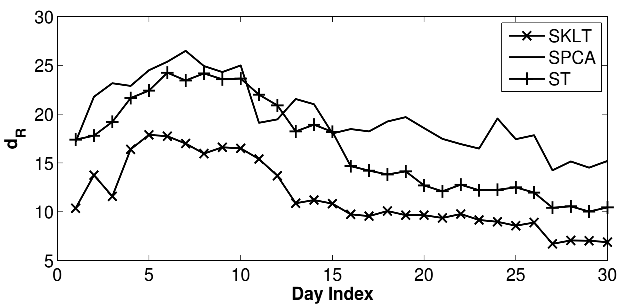
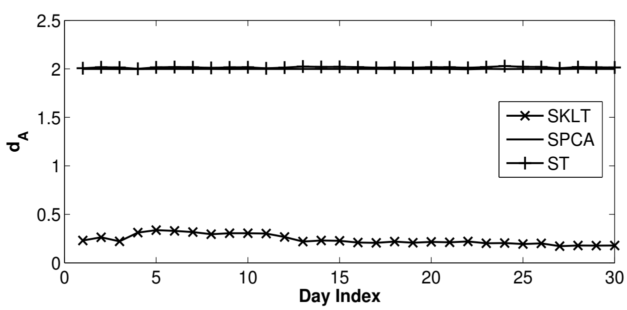


Fig. 3: Explained variance of sparsed first PC (PC1) generated by SKLT, DSPCA, SPCA, ST and SPC with respect to increasing non-sparsity level for empirical correlation matrix of end of day (EOD) returns for 100 stocks in NASDAQ-100 index with $W = 30$-day measurement window ending on April 9, 2014.



(a)



(b)

Fig. 5: (a) $d_{\mathbf{R}}$ and (b) $d_{\mathbf{A}}$ of sparse eigen subspaces generated daily from empirical correlation matrix of EOD returns between April 9, 2014 and May 22, 2014 for 100 stocks in NASDAQ-100 index by using SKLT, SPCA and ST methods, respectively. Non-sparsity level of 85% for all PCs is forced with $W = 30$-days.

generated by KLT, SKLT, SPCA and ST for 30 different empirical correlation matrices of end of day (EOD) returns of 100 stocks in NASDAQ-100 index with $W = 30$-days measurement window for the same data. Non-sparsity level is set to 85% for each PC. Since SPCA and ST perform better than DSPCA and SPC for the 85% non-sparsity level as displayed in Fig. 3, SKLT is only compared with KLT, SPCA and ST for convenience. The original (non-quantized) KLT gives us the upper limit for the explained variance. Fig. 4b, Fig. 4c, Fig. 4d, Fig. 4e, Fig. 4f, Fig. 4g, and Fig. 4h display the cumulative explained variance of sparsed two, three, four, five, six, seven and eight PCs for the same experiment, respectively. Note that the cumulative explained variance of all sparsed PCs for the same experiment has similar results with the ones displayed in Fig. 4. SKLT provides better explained variance performance than other methods even if the non-sparsity level is decreased to 80% and 75%. These figures highlight the superior sparsity performance of the SKLT method over the popular algorithms proposed in the literature. In addition, the SKLT is much simpler to implement than the others and trackable.

The distance between the original $\mathbf{R}_E(n)$ and the modified correlation matrix $\widehat{\mathbf{R}_E(n)}$ due to sparsed eigenvectors is defined as

$$d_{\mathbf{R}} = \left\| \mathbf{R}_E(n) - \widehat{\mathbf{R}_E(n)} \right\|_2 \tag{27}$$

where $\|.\|_2$ is the norm-2 of a matrix. Similarly, the distance between the original and the sparsed eigenmatrices is expressed as

$$d_{\mathbf{A}} = \left\| \mathbf{A}_{KLT} - \widehat{\mathbf{A}_{KLT}} \right\|_2 \tag{28}$$

Fig. 5a and Fig. 5b display the $d_{\mathbf{R}}$ and $d_{\mathbf{A}}$ of sparse eigen subspaces generated daily from empirical correlation matrix of EOD returns between April 9, 2014 and May 22, 2014 for 100 stocks in NASDAQ-100 index by using SKLT, SPCA and ST methods, respectively. Non-sparsity level of 85% for all PCs is forced with $W = 30$-days. These measures show that the proposed SKLT sparses eigen subspace of NASDAQ-100 index better than the SPCA and ST methods. Moreover, the SKLT does not alter the actual covariance structure like other methods due to forced sparsity as their methodology.

Although theoretically appealing, the optimization algorithms like DSPCA, SPCA and SPC with constraints for forced sparsity (cardinality reduction of a set) may substantially alter intrinsic structures of original eigenportfolios. Therefore, such a sparse representation might cause to deviate significantly from the measured empirical correlation matrix. Hence, financial performance degradations may happen in eigenportfolios generated by sparsity constrained optimization methods.

V. Conclusions

In this paper, we propose a simple method called sparse KLT (SKLT) to sparse eigen subspaces. The proposed SKLT method utilizes the mathematical framework developed for transform coding and rate-distortion theory. Due to the non-convex nature of the problem, regularization based optimization methods proposed in the literature are unable to guarantee good performance for an arbitrary covariance matrix. The sparsity performance comparisons of eigenportfolios for NASDAQ-100 index demonstrate the superiority of SKLT over the popular algorithms reported earlier in the literature.

REFERENCES

[1] J. Cadima and I. T. Jolliffe, “Loading and correlations in the interpretation of principle compenents,” *Journal of Applied Statistics*, vol. 22, no. 2, pp. 203–214, 1995.

[2] N. Trendafilov, I. T. Jolliffe, and M. Uddin, “A modified principal component technique based on the LASSO,” *Journal of Computational and Graphical Statistics*, vol. 12, pp. 531–547, 2003.

[3] H. Zou, T. Hastie, and R. Tibshirani, “Sparse principal component analysis,” *Journal of Computational and Graphical Statistics*, vol. 15, pp. 262–286, 2006.

[4] 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.

[5] M. U. Torun, A. N. Akansu, and M. Avellaneda, “Portfolio risk in multiple frequencies,” *IEEE Signal Processing Magazine, Special Issue on Signal Processing for Financial Applications*, vol. 28, no. 5, pp. 61–71, Sep. 2011.

[6] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1996.

[7] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, “Sparse and stable markowitz portfolios,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 30, pp. 12 267–12 272, 2009.

[8] 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.

[9] 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, p. 515, 2009.

[10] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets.* Academic Press, Inc., 1992.

[11] J. Max, “Quantizing for minimum distortion,” *Information Theory, IRE Transactions on*, vol. 6, no. 1, pp. 7–12, 1960.

[12] S. Lloyd, “Least squares quantization in PCM,” *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129–137, 1982.

[13] T. Berger, *Rate-Distortion Theory*. John Wiley and Sons, Inc., 2003.

[14] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall Professional Technical Reference, 1984.

[15] R. J. Clarke, *Transform Coding of Images.* Academic Press, Inc., 1985.

[16] J. Ohlson and B. Rosenberg, “Systematic risk of the CRSP equal-weighted common stock index: A history estimated by stochastic-parameter regression,” *The Journal of Business*, vol. 55, no. 1, pp. 121–145, Jan. 1982.

[17] H. Mamaysky, M. Spiegel, and H. Zhang, “Estimating the dynamics of mutual fund alphas and betas,” *Rev. Financ. Stud.*, vol. 21, pp. 233–264, 2008.

[18] N. P. B. Bollen and R. E. Whaley, “Hedge fund risk dynamics: Implications for performance appraisal,” *The Journal of Finance*, vol. 64, no. 2, pp. 985–1035, 2009.

[19] A. N. Akansu and M. U. Torun, “Toeplitz approximation to empirical correlation matrix of asset returns: A signal processing perspective,” *Journal of Selected Topics in Signal Processing*, vol. 6, no. 4, pp. 319–326, Aug. 2012.

[20] M. U. Torun and A. N. Akansu, “An efficient method to derive explicit KLT kernel for first-order autoregressive discrete process,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 15, pp. 3944–3953, Aug 2013.

[21] H. Brusewitz, *Quantization with Entropy Constraint and Bounds to the Rate Distortion Function*, ser. Trita-TTT. Telecommunication Theory, Electrical Engineering, Royal Inst. of Technology, 1986.

[22] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York, 1959.

[23] 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.