5

Explicit Kernel and Sparsity of Eigen Subspace for the AR(1) Process

Mustafa U. Torun, Onur Yilmaz and Ali N. Akansu
New Jersey Institute of Technology, USA

5.1 Introduction

Karhunen–Loeve Transform (KLT), also called Eigen decomposition or principal component analysis (PCA), is the optimal orthogonal subspace method (block transform) that maps wide-sense stationary (WSS) stochastic signals with correlations into nonstationary and pairwise uncorrelated transform coefficients. The coefficient with the highest variance corresponds to the most covariability among the observed signals, hence the most meaningful information (Akansu and Haddad, 1992). Therefore, the coefficients with large variances are kept and the ones with low variances corresponding to noise are discarded in noise-filtering and compression applications (Jolliffe, 2002). KLT basis functions are the eigenvectors of the given signal covariance matrix that define the corresponding unique eigen subspace. Therefore, it is a signal-dependent transform as opposed to some other popular transforms like discrete Fourier transform (DFT) and discrete cosine transform (DCT). DFT and DCT have their kernels that are independent of signal statistics. They are called fixed transforms, and their good performance with efficient implementations make them desirable choices for various applications (Akansu and Haddad, 1992). Fast implementation of KLT is of great interest to several disciplines, and there were attempts to derive closed-form kernel expressions for certain classes of stochastic processes. Such kernels for continuous and discrete stochastic processes with exponential autocorrelation function were reported in the literature (Davenport and Root, 1958; Pugachev, 1959a,b; Ray and Driver, 1970; Wilkinson, 1965). We focus on the discrete autoregressive order one, AR(1), and the process and derivation of its

Financial Signal Processing and Machine Learning, First Edition.
Edited by Ali N. Akansu, Sanjeev R. Kulkarni and Dmitry Malioutov.

explicit eigen kernel, in this chapter. In Section 5.4, we investigate the sparsity of such eigen subspace and present a rate-distortion theory-based sparsing method. Moreover, we highlight the merit of the method for the AR(1) process as well as for the empirical correlation matrix of stock returns in the NASDAQ-100 index.

5.2 Mathematical Definitions

5.2.1 Discrete AR(1) Stochastic Signal Model

Random processes and information sources are often described by a variety of stochastic signal models, including autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) types. Discrete AR source models, also called all-pole models, have been successfully used in applications including speech processing for decades (Atal and Hanauer, 1971). The first-order AR model, AR(1), is a first approximation to many natural signals and has been widely employed in various disciplines. Its continuous analogue is called the Ornstein–Uhlenbeck (OU) process with popular use in physical sciences and mathematical finance (Doob, 1942; Uhlenbeck and Ornstein, 1930). The AR(1) signal is generated through the first-order regression formula written as (Akansu and Haddad, 1992; Kay, 1988):

$$x(n) = \rho x(n-1) + \xi(n), \tag{5.1}$$

where $\xi(n)$ is a zero-mean white noise sequence, that is,

$$\begin{aligned} E\{\xi(n)\} &= 0, \\ E\{\xi(n)\xi(n+k)\} &= \sigma_{\xi}^2 \delta_k. \end{aligned} \tag{5.2}$$

$E\{\cdot\}$ is the expectation operator, and δ_k is the Kronecker delta function. The first-order correlation coefficient, ρ, is real in the range of $-1 < \rho < 1$, and the variance of $x(n)$ is given as follows:

$$\sigma_x^2 = \frac{1}{(1-\rho^2)} \sigma_{\xi}^2. \tag{5.3}$$

The autocorrelation sequence of $x(n)$ is expressed as

$$R_{xx}(k) = E\{x(n)x(n+k)\} = \sigma_x^2 \rho^{|k|}; k = 0, \pm 1, \pm 2, \ldots. \tag{5.4}$$

The resulting $N \times N$ Toeplitz autocorrelation matrix for the AR(1) process is expressed as

$$\mathbf{R}_x = \sigma_x^2 \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{N-1} \\ \rho & 1 & \rho & \ldots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \ldots & 1 \end{bmatrix}. \tag{5.5}$$

5.2.2 Orthogonal Subspace

A family of linearly independent N orthonormal real discrete-time sequences (vectors), $\{\phi_k(n)\}$, on the interval $0 \le n \le N-1$, satisfies the inner product relationship (Akansu and Haddad, 1992)

$$\sum_{n=0}^{N-1} \phi_k(n)\phi_l(n) = \delta_{k-l} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}. \tag{5.6}$$

Equivalently, the orthonormality can also be expressed on the unit circle of the complex plane, $z = e^{j\omega}; -\pi \le \omega \le \pi$, as follows:

$$\sum_{n=0}^{N-1} \phi_k(n)\phi_l(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_k(e^{j\omega})\Phi_l(e^{j\omega})d\omega = \delta_{k-l}, \tag{5.7}$$

where $\Phi_k(e^{j\omega})$ is the discrete time Fourier transform (DTFT) of $\phi_k(n)$. In matrix form, $\{\phi_k(n)\}$ are the rows of the transform matrix, also called basis functions:

$$\Phi = [\phi_k(n)] \; : \; k, n = 0, 1, \ldots, N-1, \tag{5.8}$$

with the matrix orthonormality property stated as

$$\Phi\Phi^{-1} = \Phi\Phi^{\mathrm{T}} = \mathbf{I}, \tag{5.9}$$

where T indicates a transposed version of a matrix or a vector. A random signal vector

$$\mathbf{x} = \begin{bmatrix} x(0) & x(1) & \ldots & x(N-1) \end{bmatrix}^{\mathrm{T}}, \tag{5.10}$$

is mapped onto the orthonormal subspace through a forward transform operator (projection)

$$\boldsymbol{\theta} = \Phi\mathbf{x}, \tag{5.11}$$

where $\boldsymbol{\theta}$ is the transform coefficients vector given as

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \ldots & \theta_{N-1} \end{bmatrix}^{\mathrm{T}}. \tag{5.12}$$

Similarly, the inverse transform reconstructs the signal vector

$$\mathbf{x} = \Phi^{-1}\boldsymbol{\theta} = \Phi^{\mathrm{T}}\boldsymbol{\theta}. \tag{5.13}$$

Hence, one can derive the correlation matrix of transform coefficients as follows:

$$\mathbf{R}_\theta = E\{\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}\} = E\{\Phi\mathbf{x}\mathbf{x}^{\mathrm{T}}\Phi^{\mathrm{T}}\} = \Phi E\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}\Phi^{\mathrm{T}} = \Phi\mathbf{R}_x\Phi^{\mathrm{T}}. \tag{5.14}$$

Furthermore, total energy represented by the transform coefficients is written as:

$$E\{\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}\} = \sum_{k=0}^{N-1} E\{\theta_k^2\} = \sum_{k=0}^{N-1} \sigma_k^2. \tag{5.15}$$

It follows from (5.9) and (5.11) that

$$E\{\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}\} = E\{\mathbf{x}^{\mathrm{T}}\Phi^{\mathrm{T}}\Phi\mathbf{x}\} = E\{\mathbf{x}^{\mathrm{T}}\mathbf{x}\} = \sum_{n=0}^{N-1} \sigma_x^2(n) = N\sigma^2, \tag{5.16}$$

where $\sigma_x^2(n)$ is the variance of the n*th* element of the signal vector given in (5.10) that is equal to σ^2. It follows from (5.15) and (5.16) that

$$\sigma^2 = \frac{1}{N}\sum_{k=0}^{N-1}\sigma_k^2. \tag{5.17}$$

The energy-preserving property of an orthonormal transform is evident in (5.17). It is also noted in (5.11) that the linear transformation of the stationary random vector process $\mathbf{x}$ results in a nonstationary random coefficient vector process $\boldsymbol{\theta}$, that is, $\sigma_k^2 \neq \sigma_l^2$ for $k \neq l$ (Akansu and Haddad, 1992).

5.2.2.1 Karhunen–Loeve Transform (Eigen Subspace)

KLT jointly provides the (i) optimal geometric mean of coefficient variances with a diagonal correlation matrix, $\mathbf{R}_\theta$, in (5.14); and (ii) best possible repacking of signal energy into as few transform coefficients as possible. KLT minimizes the energy of the approximation error due to use of only L basis functions $L \leq N$ in order to approximate covariance subject to the orthonormality constraint given in (5.9). Hence, the cost function is defined as (Akansu and Haddad, 1992):

$$J = \sum_{k=L}^{N-1} J_k = E\{\mathbf{e}_L^{\mathrm{T}}\mathbf{e}_L\} - \sum_{k=L}^{N-1}\lambda_k(\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\phi}_k - 1), \tag{5.18}$$

where λ_k is the k*th* Lagrangian multiplier. (5.18) can be rewritten as

$$J = \sum_{k=L}^{N-1} J_k = \sum_{k=L}^{N-1}\boldsymbol{\phi}_k^{\mathrm{T}}\mathbf{R}_x\boldsymbol{\phi}_k - \sum_{k=L}^{N-1}\lambda_k(\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\phi}_k - 1). \tag{5.19}$$

Taking gradient of one of the components of the error J (i.e., J_k), with respect to $\boldsymbol{\phi}_k$ and setting it to zero as follows (Akansu and Haddad, 1992):

$$\nabla J_k = \frac{\partial J_k}{\partial \boldsymbol{\phi}_k} = 2\mathbf{R}_x\boldsymbol{\phi}_k - 2\lambda_k\boldsymbol{\phi}_k = 0, \tag{5.20}$$

yields

$$\mathbf{R}_x\boldsymbol{\phi}_k = \lambda_k\boldsymbol{\phi}_k, \tag{5.21}$$

which implies that $\boldsymbol{\phi}_k$ is one of the eigenvectors of $\mathbf{R}_x$ and λ_k is the corresponding eigenvalue. It is evident from (5.21) that the basis set for KLT comprises the eigenvectors of the autocorrelation matrix of the input (i.e., $\mathbf{R}_x$), and it needs to be recalculated whenever signal statistics change. It follows from (5.21) that

$$\begin{aligned}\mathbf{R}_x\mathbf{A}_{KLT}^{\mathrm{T}} &= \mathbf{A}_{KLT}^{\mathrm{T}}\boldsymbol{\Lambda},\\ \mathbf{R}_x &= \mathbf{A}_{KLT}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{A}_{KLT} = \sum_{k=0}^{N-1}\lambda_k\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}},\end{aligned} \tag{5.22}$$

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

for the given $\mathbf{R}_x$ where σ_k^2 is the variance of the k*th* transform coefficient, θ_k (Akansu and Haddad, 1992).

5.2.2.2 Performance Metrics for Orthogonal Subspaces

In practice, it is desired that variances of the transform coefficients decrease as the coefficient index k increases, and so the signal energy is consolidated into as small a number of transform coefficients as possible (Akansu and Haddad, 1992). In other words, it is desired to minimize the energy of the approximation error defined as:

$$\mathbf{e}_L = \mathbf{x} - \hat{\mathbf{x}}_L = \sum_{k=0}^{N-1} \theta_k \boldsymbol{\phi}_k - \sum_{k=0}^{L-1} \theta_k \boldsymbol{\phi}_k = \sum_{k=L}^{N-1} \theta_k \boldsymbol{\phi}_k, \tag{5.23}$$

where $0 < L \leq N-1$. There are three commonly used metrics to measure the performance of a given orthonormal transform (subspace) (Akansu and Haddad, 1992). The compaction efficiency of a transform, that is, the ratio of the energy in the first L transform coefficients to the total energy, is defined as

$$\eta_E(L) = 1 - \frac{E\{\mathbf{e}_L^{\mathrm{T}}\mathbf{e}_L\}}{E\{\mathbf{e}_0^{\mathrm{T}}\mathbf{e}_0\}} = \frac{\sum_{k=0}^{L-1}\sigma_k^2}{N\sigma_x^2}. \tag{5.24}$$

This is an important metric to assess the efficiency of a transform for the given signal type.

The gain of transform coding (TC) over pulse code modulation (PCM) performance of an $N \times N$ unitary transform for a given input correlation is particularly significant and widely utilized in transform coding applications as defined:

$$G_{TC}^N = \frac{\frac{1}{N}\sum_{k=0}^{N-1}\sigma_k^2}{\left(\prod_{k=0}^{N-1}\sigma_k^2\right)^{1/N}}. \tag{5.25}$$

Similarly, decorrelation efficiency of a transform is defined as

$$\eta_c = 1 - \frac{\sum_{k=0}^{N-1}\sum_{l=1;l\neq k}^{N-1}|\mathbf{R}_\theta(k,l)|}{\sum_{k=0}^{N-1}\sum_{l=1;l\neq k}^{N-1}|\mathbf{R}_x(k,l)|}. \tag{5.26}$$

It is desired to have high compaction efficiency, $\eta_E(L)$; high gain of TC over PCM, G_{TC}^N; and high decorrelation efficiency, η_c, for a given $N \times N$ orthonormal transform. Detailed discussion on the performance metrics for the orthonormal transforms can be found in Akansu and Haddad, 1992.

In Figure 5.1a, $\eta_E(L)$ of KLT and the popular fixed transform DCT (Akansu and Haddad, 1992) are displayed for the AR(1) process with various correlation coefficients ρ and transform size $N = 31$. Similarly, Figure 5.1b depicts G_{TC}^N performances of KLT and DCT as a function of ρ for $N = 31$. This figure justifies the use of fixed transform DCT as a replacement to signal-dependent KLT in applications where signal samples are highly correlated. Moreover, it is noted that the energy-packing performances of both transforms degrade for lower values of correlation coefficient. Hence, subspace methods bring less value for applications where signal correlation is low.

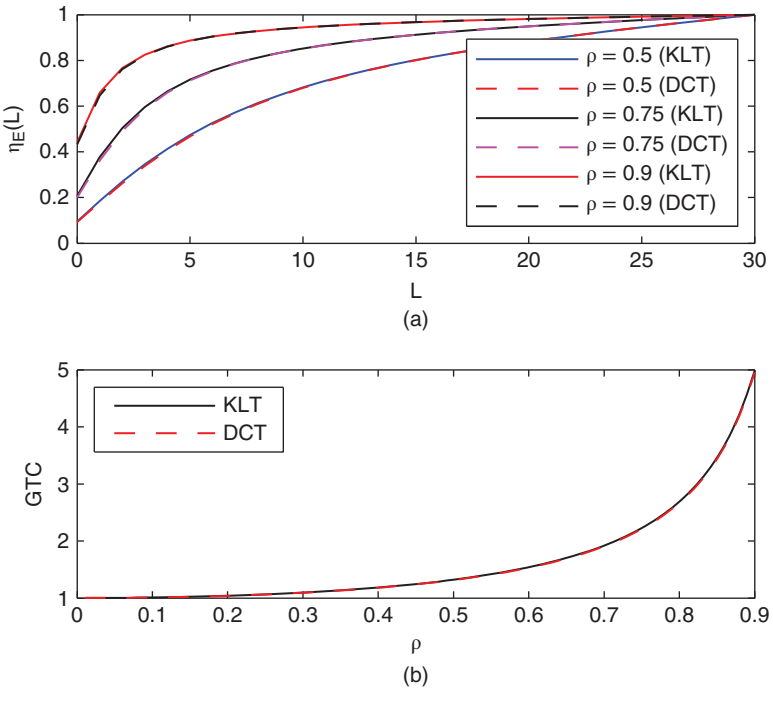


Figure 5.1 (a) $\eta_E(L)$ Performance of KLT and DCT for an AR(1) process with various values of ρ and $N = 31$; (b) G_{TC}^{N} performance of KLT and DCT as a function of ρ for $N = 31$.

5.3 Derivation of Explicit KLT Kernel for a Discrete AR(1) Process

A KLT matrix $\mathbf{A}_{KLT}$ of size $N \times N$ for an AR(1) process is expressed with the closed-form kernel as (Ray and Driver, 1970):

$$\mathbf{A}_{KLT} = [A(k,n)] = \left(\frac{2}{N+\lambda_k}\right)^{1/2} \sin\left[\omega_k\left(n - \frac{N-1}{2}\right) + \frac{(k+1)\pi}{2}\right], \tag{5.27}$$

where $0 \leq k, n \leq N-1$. Corresponding transform coefficient variances (i.e., the eigenvalues of the autocorrelation matrix given in (5.5), λ_k) are derived to be in the closed form (Ray and Driver, 1970):

$$\sigma_k^2 = \lambda_k = \frac{1-\rho^2}{1 - 2\rho\cos(\omega_k) + \rho^2}, \tag{5.28}$$

where $\{\omega_k\}$ are the positive roots of the following transcendental equation (Ray and Driver, 1970):

$$\tan(N\omega) = -\frac{(1-\rho^2)\sin(\omega)}{\cos(\omega) - 2\rho + \rho^2\cos(\omega)}. \tag{5.29}$$

The derivation steps leading to the equations (5.27), (5.28), and (5.29) will be explained throughout this section. Moreover, there is a need to derive an explicit expression for the roots

of the transcendental equation in (5.29) such that the kernel (eigenvectors) and the corresponding variances (eigenvalues) are expressed accordingly.

5.3.1 A Simple Method for Explicit Solution of a Transcendental Equation

A simple method of formulating an explicit solution for the roots of transcendental equations using Cauchy's integral theorem from complex analysis (Strang, 1986) that was introduced by Luck and Stevens (2002) is highlighted in this section. The method determines the roots of a transcendental function by locating the singularities of a reciprocal function. Although derivation steps are detailed in Luck and Stevens (2002), a summary is given here for the completeness of presentation.

Cauchy's theorem states that if a function is analytic in a simple connected region containing the closed curve C, the path integral of the function around the curve C is zero. On the other hand, if a function, $f(z)$, contains a single singularity at z_0 somewhere inside C but analytic elsewhere in the region, then the singularity can be removed by multiplying $f(z)$ with $(z - z_0)$ (i.e., by a pole-zero cancellation). Cauchy's theorem implies that the path integral of the new function $(z - z_0) f(z)$ around C must be zero:

$$\oint_C (z - z_0) f(z) dz = 0. \tag{5.30}$$

Evaluation of the integral given in (5.30) yields a first-order polynomial in z_0 with constant coefficients, and its solution for z_0 provides the location of the singularity as given (Luck and Stevens, 2002)

$$z_0 = \frac{\oint_C z f(z) dz}{\oint_C f(z) dz}. \tag{5.31}$$

This is an explicit expression for the singularity of the function $f(z)$. A root-finding problem is restated as a singularity at the root. It is noted that (5.31) gives the location of the desired root, and it can be evaluated for any closed path by employing either an analytical or a numerical technique. Luck and Stevens (2002) suggested to use a circle in the complex plane with its center h and radius R as the closed curve C, expressed as

$$\begin{aligned} z &= h + Re^{j\theta}, \\ dz &= jRe^{j\theta} d\theta, \end{aligned} \tag{5.32}$$

where $0 \leq \theta \leq 2\pi$; $h \in \mathbb{R}$; and $R \in \mathbb{R}$; Values of h and R do not matter as long as the circle circumscribes the root z_0. Cauchy's argument principle (Brown and Churchill, 2009) or graphical methods may be used to determine the number of roots enclosed by the path C. A function in θ is defined as

$$w(\theta) = f(z)|_{z=h+Re^{j\theta}} = f(h + Re^{j\theta}). \tag{5.33}$$

Then (5.31) becomes (Luck and Stevens, 2002):

$$z_0 = h + R \left[\frac{\int_0^{2\pi} w(\theta) e^{j2\theta} d\theta}{\int_0^{2\pi} w(\theta) e^{j\theta} d\theta} \right]. \tag{5.34}$$

One can easily evaluate (5.34) by employing Fourier analysis since the n*th* Fourier series coefficient for any $x(t)$ is calculated as

$$A_n = \frac{1}{2\pi}\int_0^{2\pi} x(t)e^{jnt}dt. \tag{5.35}$$

It is observed that the term in brackets in (5.34) is equal to the ratio of the second Fourier series coefficient over the first one for the function $w(\theta)$. Fourier series coefficients can be easily calculated numerically by using DFT or by using its fast implementation, Fast Fourier transform (FFT), as is suggested in Luck and Stevens (2002). However, it is observed from (5.34) that one does not need all DFT coefficients to solve the problem since it requires only two Fourier series coefficients. Therefore, it is possible to further improve the computational cost by employing a discrete summation operator to implement (5.34) numerically. Hence, the algorithm would have a computational complexity of $O(N)$ instead of $O(NlogN)$ required for FFT algorithms.

It is also noted that given $f(z)$ is analytic at h, multiplying $f(z)$ by a factor $(z-h) = Re^{j\theta}$ does not change the location of the singularities of $f(z)$. It means that for a given singularity, the term in brackets is also equal to any ratio of the (m + 1)*th* to the m*th* Fourier series coefficients of $w(\theta)$ for $m \geq 1$ (Luck and Stevens, 2002). The MATLAB^{TM} code given in Torun and Akansu (2013) to calculate the roots of (5.43) shows the simplicity of the method to solve such transcendental equations.

5.3.2 Continuous Process with Exponential Autocorrelation

The classic problem of deriving explicit solutions for characteristic values and functions of a continuous random process with an exponential autocorrelation function provides the foundation for the derivation of explicit KLT kernel for a discrete AR(1) process. This problem is discussed in Davenport and Root (1958) and Van Trees (2001). Similar discussions can also be found in Pugachev (1959a,b). After some derivation steps, the characteristic equation of the process is expressed as

$$\phi''(t) + \frac{\alpha(2-\alpha\lambda)}{\lambda}\phi(t) = 0, \tag{5.36}$$

and (5.36) has a solution only in the range of $0 < \lambda < \frac{2}{\alpha}$ and is rewritten as (Davenport and Root, 1958):

$$\phi''(t) + b^2\phi(t) = 0, \tag{5.37}$$

where

$$b^2 = \frac{\alpha(2-\alpha\lambda)}{\lambda}, 0 < b^2 < \infty. \tag{5.38}$$

A general solution to (5.37) is given as

$$\phi(t) = c_1 e^{jbt} + c_2 e^{-jbt}, \tag{5.39}$$

where c_1 and c_2 are arbitrary constants. A solution is possible only when $c_1 = c_2$ or $c_1 = -c_2$. For $c_1 = c_2$, it is shown that one of the unknowns in the general solution is given in (5.39), b, satisfies the equation (Davenport and Root, 1958)

$$b\tan b\frac{T}{2} = \alpha. \tag{5.40}$$

It follows from (5.39) that for every positive b_k that satisfies the transcendental equation (5.40), there is a characteristic function that satisfies the characteristic equation and given as (Davenport and Root, 1958):

$$\phi_k(t) = c_k \cos b_k t, \tag{5.41}$$

where integer $k \geq 0$. Similarly, for $c_1 = -c_2$, b, one can derive the resulting characteristic functions as

$$\phi_k(t) = c_k \sin b_k t. \tag{5.42}$$

The steps required to determine the roots of (5.40) are summarized next. We can rewrite (5.40) for $\alpha = B$ and $T = 2$ as

$$b \tan b = B. \tag{5.43}$$

Positive roots of (5.43), $b_m > 0$, must be calculated in order to determine the even indexed characteristic values and functions. Figure 5.2 displays functions $\tan(b)$ and B/b for various values of B. It is observed from the figure that for the m*th* root, a suitable choice for the closed path C is a circle of radius $R = \pi/4$ centered at $h_m = (m - 3/4)\pi$, as suggested in (Luck and Stevens, 2002). A straightforward way to configure (5.43) to introduce singularities is to use the inverse of (5.43) rearranged as follows (Luck and Stevens, 2002):

$$f(b) = \frac{1}{b \sin(b) - B \cos(b)}. \tag{5.44}$$

Applying (5.34) to (5.44) results in an explicit expression for the m*th* root. This expression can be evaluated by calculating a pair of adjacently indexed DFT coefficients (coefficients of

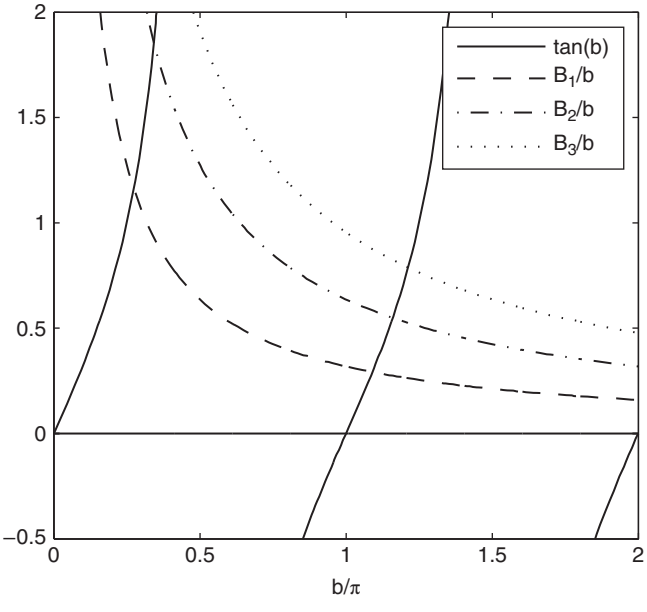


Figure 5.2 Functions $\tan(b)$ and B/b for various values of B where $B_1 = 1$, $B_2 = 2$, and $B_3 = 3$.

two adjacent harmonics), as described in Section 5.3.1, or by using a numerical integration method. Therefore, by setting $b = h + Re^{j\theta}$, $w_m(\theta)$ of (5.33) for this case is defined as:

$$\begin{aligned} w_m(\theta) &= f(h_m + Re^{j\theta}), \\ &= \frac{1}{(h_m + Re^{j\theta})\sin(h_m + Re^{j\theta}) - B\cos(h_m + Re^{j\theta})}, \end{aligned} \tag{5.45}$$

where $0 \leq \theta \leq 2\pi$. Hence, the location of the m*th* root is explicitly defined as

$$b_m = h_m + R\left[\frac{\int_0^{2\pi} w_m(\theta)e^{j2\theta}d\theta}{\int_0^{2\pi} w_m(\theta)e^{j\theta}d\theta}\right]. \tag{5.46}$$

5.3.3 Eigenanalysis of a Discrete AR(1) Process

In this section, derivations of (5.27), (5.28), and (5.29) are given in detail. For a a discrete random signal (process), $x(n)$, a discrete Karhunen–Loeve (K-L) series expansion is given as follows (Akansu and Haddad, 1992):

$$\sum_{m=0}^{N-1} R_{xx}(n,m)\phi_k(m) = \lambda_k\phi_k(n), \tag{5.47}$$

where m and n are the independent discrete variables;

$$R_{xx}(n,m) = E\{x(n)x(m)\}, m, n = 0, 1, \ldots, N-1, \tag{5.48}$$

is the autocorrelation function of the random signal; λ_k is the k*th* eigenvalue; and $\phi_k(n)$ is the corresponding k*th* eigenfunction. The autocorrelation function of the stationary discrete AR(1) process is given as (Papoulis, 1991):

$$R_x(n,m) = R_x(n-m) = \rho^{|n-m|}. \tag{5.49}$$

Hence, the discrete K-L series expansion for an AR(1) process, from (5.47) and (5.49), is stated as follows

$$\sum_{m=0}^{N-1} \rho^{|n-m|}\phi_k(m) = \lambda_k\phi_k(n). \tag{5.50}$$

In order to eliminate the magnitude operator, (5.50) can be rewritten in the form

$$\sum_{m=0}^{n} \rho^{n-m}\phi_k(m) + \sum_{m=n+1}^{N-1} \rho^{m-n}\phi_k(m) = \lambda_k\phi_k(n). \tag{5.51}$$

It follows from the continuous case presented in Section 5.3.2 that the general solution for the k*th* eigenvector is given as (Davenport and Root, 1958; Van Trees, 2001):

$$\phi_k(t) = c_1 e^{j\omega_k t} + c_2 e^{-j\omega_k t}, \tag{5.52}$$

where c_1 and c_2 are arbitrary constants; t is the independent continuous variable; $-T/2 \leq t \leq T/2$; and $\omega_k = b_k$. The eigenfunction given in (5.52) is shifted by $T/2$ and sampled at

$t_n = nT_s$, $0 \le n \le N-1$, where $T_s = T/(N-1)$. Accordingly, the sampled eigenfunction is written as

$$\phi_k(n) = c_1 e^{j\omega_k(n-\frac{N-1}{2})} + c_2 e^{-j\omega_k(n-\frac{N-1}{2})}. \tag{5.53}$$

The solution to (5.50) exists only when $c_1 = \pm c_2$. In the following discussions, the case for $c_1 = c_2$ is considered given the fact that the sister case for $c_1 = -c_2$ is similar. For $c_1 = c_2$, it follows from (5.53) that

$$\phi_k(n) = c_1 \cos\left[\omega_k\left(n - \frac{N-1}{2}\right)\right]. \tag{5.54}$$

By substituting (5.54) in (5.51) and defining a new independent discrete variable $p = m - (N-1)/2$, (5.50) can be rewritten as follows:

$$\sum_{p=-\frac{N-1}{2}}^{n-\frac{N-1}{2}} \rho^{n-p-\frac{N-1}{2}} \cos(\omega_k p) + \sum_{p=n+1-\frac{N-1}{2}}^{\frac{N-1}{2}} \rho^{p+\frac{N-1}{2}-n} \cos(\omega_k p) \tag{5.55}$$

$$= \lambda_k \cos\left[\omega_k\left(n - \frac{N-1}{2}\right)\right]. \tag{5.56}$$

The first summation on the left in (5.55) is rewritten as

$$\frac{1}{2}\rho^{n-\frac{N-1}{2}}\left[\sum_{p=-\frac{N-1}{2}}^{n-\frac{N-1}{2}} \left(\rho^{-1}e^{j\omega_k}\right)^p + \sum_{p=-\frac{N-1}{2}}^{n-\frac{N-1}{2}} (\rho^{-1}e^{-j\omega_k})^p\right]. \tag{5.57}$$

Using the fact that

$$\sum_{n=N_1}^{N_2} \beta^n = \frac{\beta^{N_1} - \beta^{N_2+1}}{1-\beta}, \tag{5.58}$$

and following simple derivation steps, it can be shown that (5.57), hence the first summation on the left in (5.55), is equal to

$$\frac{\rho^{n+2}\cos\omega_1 - \rho\cos\omega_2 - \rho^{n+1}\cos\omega_3 + \cos\omega_4}{1 - 2\rho\cos\omega_k + \rho^2}. \tag{5.59}$$

Similarly, the second summation on the left in (5.55) is equal to

$$\frac{\rho^{N-n+1}\cos\omega_1 + \rho\cos\omega_2 - \rho^{N-n}\cos\omega_3 - \rho^2\cos\omega_4}{1 - 2\rho\cos\omega_k + \rho^2}, \tag{5.60}$$

where

$$\begin{aligned}
\omega_1 &= \omega_k[(N-1)/2], \\
\omega_2 &= \omega_k[n - (N-1)/2 + 1], \\
\omega_3 &= \omega_k[(N-1)/2 + 1], \\
\omega_4 &= \omega_k[n - (N-1)/2],
\end{aligned} \tag{5.61}$$

for both (5.59) and (5.60). It is possible to express λ_k on the right-hand side of (5.55) in terms of ρ and ω_k by taking the discrete K-L expansion given in (5.50) into the frequency domain via DTFT as follows:

$$S_x(e^{j\omega})\Phi_k(e^{j\omega}) = \lambda_k \Phi_k(e^{j\omega}), \tag{5.62}$$

where $S_x(e^{j\omega})$ is the power spectral density (PSD) of the discrete AR(1) process and expressed as

$$S_x(e^{j\omega}) = \mathcal{F}\{\rho^{|n-m|}\} = \frac{1-\rho^2}{1-2\rho\cos\omega+\rho^2}. \tag{5.63}$$

$\mathcal{F}\{\cdot\}$ is the DTFT operator (Akansu and Haddad, 1992). The Fourier transform of the eigenfunction in (5.54) is calculated as

$$\begin{aligned}\Phi_k(e^{j\omega}) &= \mathcal{F}\{\phi_k(n)\},\\ &= c_1 e^{-j\omega_k \frac{N-1}{2}}[\delta(\omega-\omega_k)+\delta(\omega+\omega_k)],\end{aligned} \tag{5.64}$$

where $\delta(\omega-\omega_0)$ is an impulse function of frequency located at ω_0. By substituting (5.63) and (5.64) into (5.62), λ_k is derived as

$$\lambda_k = \frac{1-\rho^2}{1-2\rho\cos\omega_k+\rho^2}. \tag{5.65}$$

Equation (5.65) shows that the eigenvalues are the samples of the PSD given in (5.63). Moreover, (5.28) and (5.65) are identical. By substituting (5.59), (5.60), and (5.65) in (5.51), one can show that

$$\rho = \frac{\cos(\omega_k N/2+\omega_k/2)}{\cos(\omega_k N/2-\omega_k/2)}. \tag{5.66}$$

Using trigonometric identities, the relationship between ω_k and ρ in (5.66) is rewritten as follows

$$\tan\left(\omega_k\frac{N}{2}\right) = \left(\frac{1-\rho}{1+\rho}\right)\cot\left(\frac{\omega_k}{2}\right). \tag{5.67}$$

Similarly, for the case of $c_1 = -c_2$, following the same procedure, the relationship between ω_k and ρ is shown as

$$\tan\left(\omega_k\frac{N}{2}\right) = -\left(\frac{1+\rho}{1-\rho}\right)\tan\left(\frac{\omega_k}{2}\right). \tag{5.68}$$

Finally, from (5.67) and (5.68), it is observed that ω_k are the positive roots of the equation

$$\left[\tan\left(\omega\frac{N}{2}\right)+\frac{1+\rho}{1-\rho}\tan\left(\frac{\omega}{2}\right)\right]\left[\tan\left(\omega\frac{N}{2}\right)-\frac{1-\rho}{1+\rho}\cot\left(\frac{\omega}{2}\right)\right]=0. \tag{5.69}$$

Using trigonometric identities, (5.69) can be rewritten as

$$\tan(N\omega) = -\frac{(1-\rho^2)\sin(\omega)}{\cos(\omega)-2\rho+\rho^2\cos(\omega)}, \tag{5.70}$$

which is the same transcendental equation expressed in (5.29). The roots $\{\omega_k\}$ of the transcendental tangent equation in (5.70) are required in the KLT kernel expressed in (5.27). There are well-known numerical methods like the secant method (Allen and Isaacson, 1997)

to approximate roots of the equation given in (5.70) in order to solve it implicitly rather than explicitly. A method to find explicit solutions to the roots of transcendental equations, including (5.70), is revisited next. That method leads to an explicit definition of KLT kernel given in (5.27) for an AR(1) process.

5.3.4 Fast Derivation of KLT Kernel for an AR(1) Process

In this section, the fast derivation method of KLT kernel for a discrete AR(1) process is explained. Moreover, a step-by-step implementation of the technique is presented.

5.3.4.1 Discrete AR(1) Process

In order to define an explicit expression for the discrete KLT kernel of (5.27), first, one must find $N/2$ positive roots of the following two transcendental equations:

$$\tan\left(\omega\frac{N}{2}\right) = \frac{1}{\gamma}\cot\left(\frac{\omega}{2}\right), \tag{5.71}$$

$$\tan\left(\omega\frac{N}{2}\right) = -\gamma\tan\left(\frac{\omega}{2}\right), \tag{5.72}$$

as discussed in Section 5.3.3. In both equations, N is the transform size and

$$\gamma = (1+\rho)/(1-\rho), \tag{5.73}$$

where ρ is the first-order correlation coefficient of the AR(1) process. Roots of (5.71) and (5.72) correspond to the even and odd indexed eigenvalues and eigenvectors, respectively. Figure 5.3 displays functions $\tan(\omega N/2)$ and $-\gamma\tan(\omega/2)$ for $N = 8$ and various values of ρ. It is apparent from the figure that for the m*th* root of (5.72), a suitable choice for the closed path C in (5.31) is a circle of radius

$$R_m = \begin{cases} \pi/2N & m \le 2 \\ \pi/N & m > 2 \end{cases}, \tag{5.74}$$

centered at $h_m = (m-1/4)(2\pi/N)$, where $1 \le m \le N/2$. It is worth noting that for positively correlated signals, $0 < \rho < 1$, the ratio given in (5.73) is always greater than 1, $\gamma > 1$. However, for negatively correlated signals, $-1 < \rho < 0$, the ratio is between 0 and 1 (i.e., $0 < \gamma < 1$). Therefore, for $\rho < 0$, the last two roots must be smaller than the rest as

$$R_m = \begin{cases} \pi/N & m < N/2-1 \\ \pi/2N & m \ge N/2-1 \end{cases}. \tag{5.75}$$

Similar to the continuous case, (5.72) is reconfigured and the poles of the following inverse function are calculated

$$g(\omega) = \frac{1}{\tan(\omega N/2) + \gamma\tan(\omega/2)}. \tag{5.76}$$

By setting $\omega = h + Re^{j\theta}$, the function $w(\theta)$ of (5.33) for this case is defined as

$$\begin{aligned} w_m(\theta) &= g(h_m + R_m e^{j\theta}), \\ &= \frac{1}{\tan\left[\left(h_m + R_m e^{j\theta}\right)\frac{N}{2}\right] + \gamma\tan\left[\left(h_m + R_m e^{j\theta}\right)\frac{1}{2}\right]}, \end{aligned} \tag{5.77}$$

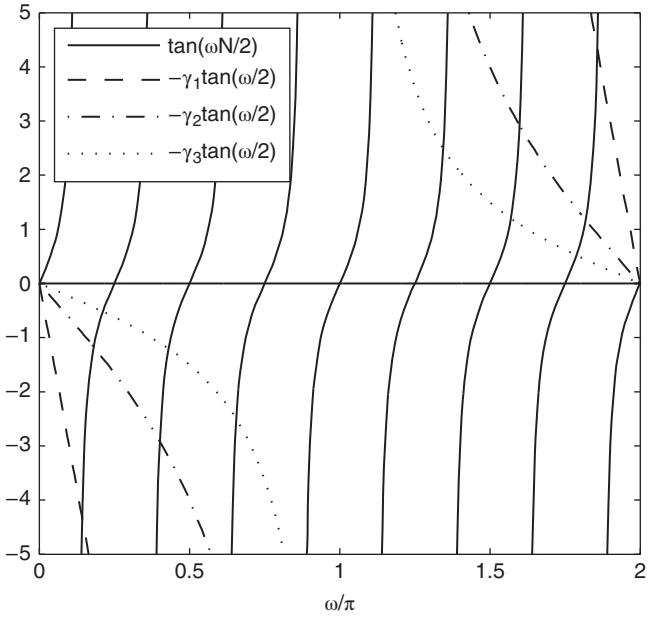


Figure 5.3 Functions $\tan(\omega N/2)$ and $-\gamma\tan(\omega/2)$ for the AR(1) process with $N = 8$ and various values of ρ, where $\rho_1 = 0.9$, $\rho_2 = 0.6$, $\rho_3 = 0.2$, and $\gamma_i = (1+\rho_i)/(1-\rho_i)$, $i = 1, 2, 3$.

where $0 \le \theta \le 2\pi$. Hence, the m*th* root is located at

$$\omega_m = h_m + R_m \left[\frac{\int_0^{2\pi} w_m(\theta) e^{j2\theta} d\theta}{\int_0^{2\pi} w_m(\theta) e^{j\theta} d\theta} \right]. \tag{5.78}$$

The procedure is the same as finding the roots of (5.71) with the exceptions that (5.77) must be modified as follows:

$$w_m(\theta) = \frac{1}{\tan\left[\left(h_m + R_m e^{j\theta}\right)\frac{N}{2}\right] - \frac{1}{\gamma}\cot\left[\left(h_m + R_m e^{j\theta}\right)\frac{1}{2}\right]}, \tag{5.79}$$

and a suitable choice for the closed path C is a circle of radius $R_m = \pi/N$ centered at

$$h_m = \begin{cases} (m-1/2)\ \ (2\pi/N) & m \le 2 \\ (m-1)\ \ (2\pi/N) & m > 2 \end{cases}, \tag{5.80}$$

which can be determined by generating a plot similar to the ones in Figures 5.2 and 5.3.

Finally, the steps to derive an explicit KLT kernel of dimension N for an arbitrary discrete dataset by employing an AR(1) approximation are summarized as follows

1. Estimate the first-order correlation coefficient of an AR(1) model for the given dataset $\{x(n)\}$ as
$$\rho = \frac{R_{xx}(1)}{R_{xx}(0)} = \frac{E\{x(n)x(n+1)\}}{E\{x(n)x(n)\}}, \tag{5.81}$$
where n is the index of random variables (or discrete-time) and $-1 < \rho < 1$.
2. Calculate the positive roots $\{\omega_k\}$ of the polynomial given in (5.29) by substituting (5.77) and (5.79) into (5.78) for odd and even values of k, respectively, and use the following indexing:
$$m = \begin{cases} k/2+1 & k \text{ even} \\ (k+1)/2 & k \text{ odd} \end{cases}. \tag{5.82}$$
3. Plug in the values of ρ and $\{\omega_k\}$ in (5.28) and (5.27) to calculate the eigenvalues λ_k and eigenvectors, respectively.

MATLAB^{TM} code for steps 2 and 3 with DFT (FFT) used in solving (5.78) are provided in (Torun and Akansu, 2013).

The computational cost of deriving KLT matrix for an arbitrary signal source has two distinct components: the calculation of the first-order correlation coefficient ρ for the given signal set, and the calculation of the roots $\{\omega_k\}$ of (5.29) that are plugged in (5.27) to generate the resulting transform matrix $\mathbf{A}_{KLT}$. The roots $\{\omega_k\}$ of the transcendental tangent equation (5.29), calculated by using (5.78), as a function of ρ and for $N = 8$ are displayed in Figure 5.4.

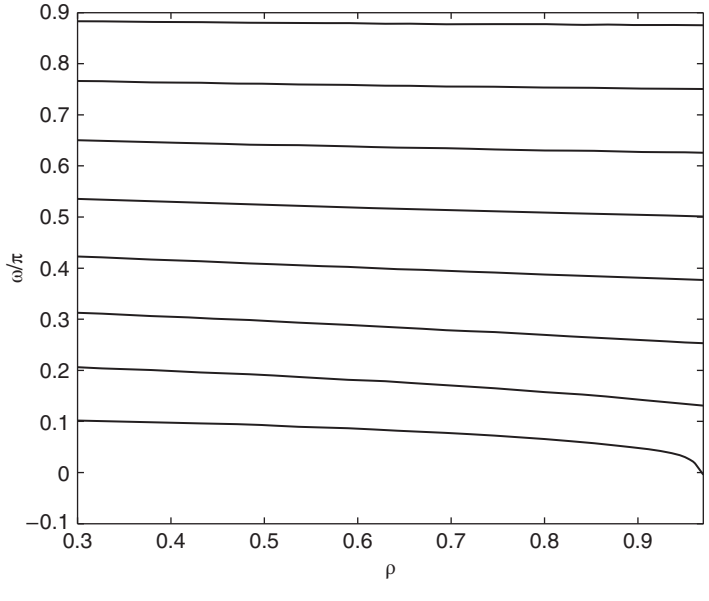


Figure 5.4 The roots of the transcendental tangent equation 5.29, $\{\omega_k\}$, as a function of ρ for $N = 8$.

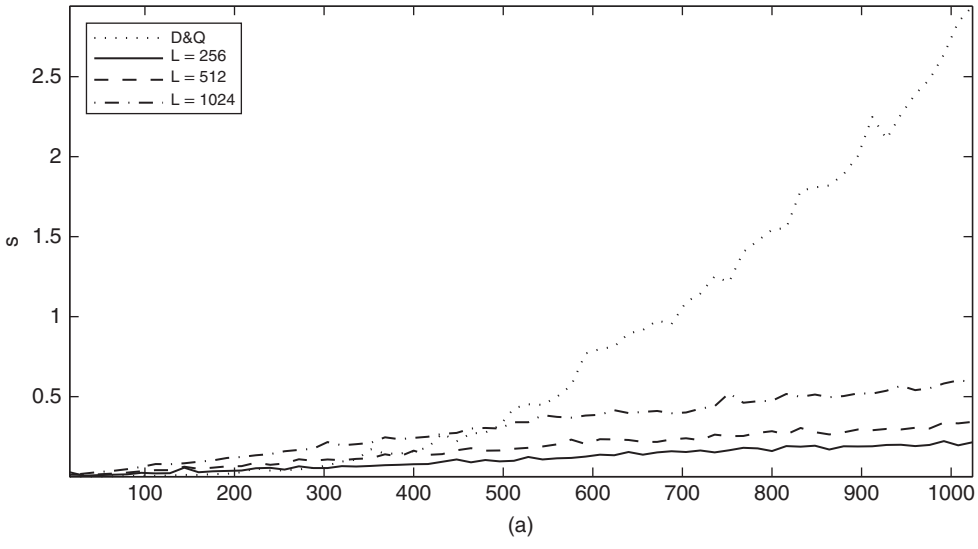


Figure 5.5 Computation time, in seconds, to calculate $\mathbf{A}_{KLT,DQ}$ and $\mathbf{A}_{KLT,E}$ for an AR(1) process with $\rho = 0.95$, and different values of N ($16 \leq N \leq 1024$) and $L = 256, 512, 1024$ (Torun and Akansu, 2013).

The computational cost of generating KLT kernel for the given statistics by using the method presented in this section ($\mathbf{A}_{KLT,E}$) is compared with a widely used numerical algorithm called divide and conquer (D&Q) (Golub and Loan, 1996).

Computation times (in seconds) to generate $\mathbf{A}_{KLT,DQ}$ and $\mathbf{A}_{KLT,E}$ (DFT sizes of $L = 256, 512, 1024$) for the case of $\rho = 0.95$ and $16 \leq N \leq 1024$ are displayed in Figure 5.5. Both computations are performed by using one thread on a single processor. The machine used for the simulations has an Intel® Core™ i5-520M CPU and 8 GB of RAM. It is displayed in the figure that the explicit KLT kernel derivation method significantly outperforms the D&Q algorithm for larger values of N. Furthermore, it has a so-called *embarrassingly parallel* nature. Hence, it can be easily computed on multiple threads and processors for any k. Therefore, by implementing it on a parallel device such as a GPU and FPGA, its speed can be significantly improved.

5.4 Sparsity of Eigen Subspace

The sparse representation of signals has been of great research interest in signal processing and other disciplines. KLT has been one of the most popular mathematical tools employed in multivariate data processing and dimension reduction. The application-specific interpretation of eigenvectors with eigenvalues and their vector components is often a critical task (Cadima and Jolliffe, 1995; d'Aspremont *et al*., 2007; Trendafilov *et al*., 2003; Zou *et al*., 2006). Moreover, small but nonzero 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 eigenportfolios with a large number of financial instruments (d'Aspremont *et al*., 2007; Torun *et al*., 2011). This and other data-intensive applications that utilize loading

coefficients have motivated researchers to study the sparsity of PCs in eigenanalysis of measured covariance 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 (Akansu and Haddad, 1992; Clarke, 1985; Jayant and Noll, 1984). Therefore, both dimension reduction and sparsity of basis functions (vectors) are significant attributes of orthogonal transforms widely utilized in many fields. This recent development has paved the way for the development of several popular methods for sparse eigen subspace. A rate-distortion-based framework to sparse subspaces is presented in this section. The challenge is to maximize explained variance (eigenvalue) by a minimum number of PCs, also called energy compaction (dimension reduction), while replacing the less significant samples (loading coefficients) of basis functions with zero to achieve the desired level of sparsity (cardinality reduction) in representation.

5.4.1 Overview of Sparsity Methods

We provide an overview of the recent literature on sparsity in this subsection for readers with more interest on the subject. Regularization methods have been used to make an ill-conditioned matrix invertible or to prevent overfitting (Bertero and Boccacci, 1998; Engl *et al*., 1996). More recently, regularization methods also have been utilized for sparsity. ℓ_0 regularizer leads to a sparse solution. On the other hand, it makes the optimization problem nonconvex. Eigenfiltering is another popular method employed for regularization (Bertero and Boccacci, 1998; Engl *et al*., 1996). ℓ_1 regularizer, so-called lasso, is widely used as an approximation (convex relaxation) of ℓ_0 regularizer (Tibshirani, 1996; Trendafilov *et al*., 2003). Another ℓ_1-based method was proposed in (Brodie *et al*., 2009) for sparse portfolios. SCoTLASS (Trendafilov *et al*., 2003) and SPCA (Zou *et al*., 2006) utilize the ℓ_1 and ℓ_2 regularizers for sparse approximation to PCs, respectively.

The sparse PCA is modeled in (Trendafilov *et al*., 2003; Zou *et al*., 2006) as an explained variance maximization problem where the number of nonzero elements in the PCs is considered as a basis design constraint. These methods suffer from potentially being stuck in local minima due to the nonconvex 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 (d'Aspremont *et al*., 2007). Empirical performance results for certain cases indicate that DSPCA may generate sparse PCs that preserve slightly more explained variance than SCoTLASS (Trendafilov *et al*., 2003) and SPCA (Zou *et al*., 2006) for the same sparsity level. A nonnegative variant of the sparse PCA problem, which forces the elements of each PC to be nonnegative, is introduced in Zass and Shashua (2006). Nonnegative sparse PCA (NSPCA) offers competitive performance to SCoTLASS, SPCA, and DSPCA in terms of explained variance for a given sparsity. However, signs of the PC elements bear specific information for the applications of interests such as eigenportfolios. Thus, 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 Shen and Huang (2008). Simulation results for certain cases show that sPCA–rSVD provides competitive results to SPCA. A variation of sPCA–rSVD, so-called sparse principal components (SPCs), that utilizes the penalized matrix decomposition (PMD) is proposed in Witten *et al*. (2009). PMD that computes the rank K approximation of a given matrix is proposed in Witten *et al*. (2009). It utilizes

the lasso penalty for sparsity. Unfortunately, none of these methods result in guaranteed sparsity regardless of their prohibitive computational cost. 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 thresholding technique is easy to implement (Cadima and Jolliffe, 1995). It performs better than SCoTLASS and slightly worse than SPCA (Zou *et al*., 2006). Although simple thresholding is easy to implement, it may cause unexpected distortion levels as variance loss. Soft thresholding (ST) is another technique that is utilized for sparse representation in Zou *et al*., (2006). Certain experiments show that ST offers slightly better performance than simple thresholding (Zou *et al*., 2006). Therefore, threshold selection plays a central role in sparsity performance.

In this section, we present in detail a subspace sparsing framework based on the rate-distortion theory (Akansu and Haddad, 1992; Berger, 2003; Lloyd, 1982; Max, 1960). It may be considered as an extension of the simple or soft thresholding method to combine a sparse representation problem with an optimal quantization method used in the source coding field (Akansu and Haddad, 1992; Berger, 2003; Cadima and Jolliffe, 1995; Clarke, 1985; Jayant and Noll, 1984). The method employs a varying-size midtread (zero-zone) probability density function (pdf)-optimized (Lloyd–Max) quantizer designed for a component histogram of each eigenvector (or the entire eigenmatrix) to achieve the desired level of distortion (sparsity) in the subspace with reduced cardinality (Hajnal, 1983; Lloyd, 1982; Max, 1960; Sriperumbudur *et al*., 2009). Herein, we focus specifically on eigen subspace of a discrete AR(1) process with closed–form expressions for its eigenvectors and eigenvalues as derived earlier in the chapter. It is known that the AR(1) process approximates well many real-world signals (Akansu and Haddad, 1992). We also sparse eigenportfolios of the NASDAQ-100 index by using this method. Note that the 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 (Akansu and Haddad, 1992; Clarke, 1985; Jayant and Noll, 1984). Therefore, we investigate the trade-off between subspace orthogonality and sparsity from a rate-distortion perspective. Then, we provide a comparative performance of the presented method along with the various methods reported in the literature, such as ST (Zou *et al*., 2006), SPCA (Zou *et al*., 2006), DSPCA (d'Aspremont *et al*., 2007), and SPC (Witten *et al*., 2009), with respect to the metrics of nonsparsity (NS) and variance loss (VL).

5.4.2 pdf-Optimized Midtread Quantizer

Quantizers may be categorized as midrise and midtread (Jayant and Noll, 1984). A midtread quantizer is preferred for applications requiring entropy reduction and noise filtering (or sparsity) simultaneously (Gonzales and Akansu, 1997). In this section, we utilize a midtread type to quantize basis function (vector) samples (components) of a transform (subspace) to achieve sparse representation in the signal domain.

A celebrated design method to calculate optimum intervals (bins) and representation (quanta) values of a quantizer for the given input signal pdf, a so-called pdf-optimized quantizer, was independently proposed by Max and Lloyd (Lloyd, 1982; Max, 1960). It assumes a random information source X with zero-mean and a known pdf function $p(x)$. Then, it minimizes quantization error in the mean squared error (mse) sense and also makes sure

that all bins of a quantizer have the same level of error. The quantization error of an L-bin pdf-optimized quantizer is expressed as follows:

$$\sigma_q^2 = \sum_{k=1}^{L} \int_{x_k}^{x_{k+1}} (x - y_k)^2 p(x) dx, \tag{5.83}$$

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 (Lloyd, 1982; Max, 1960):

$$\begin{aligned} \frac{\partial \sigma_q^2}{\partial x_k} &= 0;\ k = 2, 3, \ldots, L, \\ \frac{\partial \sigma_q^2}{\partial y_k} &= 0;\ k = 1, 2, 3, \ldots, L, \end{aligned} \tag{5.84}$$

leading to the optimal unequal intervals and resulting quanta values as

$$x_{k,opt} = \frac{1}{2}(y_{k,opt} + y_{k-1,opt});\ k = 2, 3, \ldots, L, \tag{5.85}$$

$$y_{k,opt} = \frac{\int_{x_k}^{x_{k+1,opt}} xp(x)dx}{\int_{x_k}^{x_{k+1,opt}} p(x)dx};\ k = 1, 2, \ldots, L, \tag{5.86}$$

where $x_{1,opt} = -\infty$ and $x_{L+1,opt} = \infty$. A sufficient condition to avoid local optimum in (5.84) is the log-concavity of the pdf function $p(x)$. The log-concave property holds for uniform, Gaussian, and Laplacian pdf types (Jayant and Noll, 1984). 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 an adjustable zero-zone, odd L, or midtread quantizer, to sparse (quantize) eigenvectors of an eigen subspace. One can adjust zero-zone(s) of the quantizer(s) to achieve the desired level of sparsity in a transform matrix with the trade-off of resulting imperfectness in orthogonality and explained variance. We will present design examples by using the presented technique to sparse subspaces in Section 5.4.3.

The discrepancy between input and output of a quantizer is measured by the signal-to-quantization-noise ratio (SQNR) (Berger, 2003)

$$SQNR(dB) = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_q^2}\right), \tag{5.87}$$

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

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 p(x) dx. \tag{5.88}$$

The first-order entropy (rate) of the output for an L-level quantizer with such an input is calculated as (Berger, 2003; Brusewitz, 1986)

$$H = -\sum_{k=1}^{L} P_k \log_2 P_k, \tag{5.89}$$

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

Rate-distortion theory states that the quantization error variance is expressed as (Berger, 2003):

$$\sigma_q^2 = f(R)\sigma_x^2, \tag{5.90}$$

where $f(R) = \gamma 2^{-2R}$ and the number of quantizer levels found as $L = 2^R$. The parameter γ, also called the *fudge factor*, depends on the pdf type of the information source.

The optimum allocation of the total bits Ramong multiple information sources (transform coefficients in transform coding (TC)) is an important task in lossy compression. Transform coefficient variances σ_k^2 (or eigenvalues λ_k) are quite uneven to achieve dimension reduction in TC. Therefore, an optimum bit allocation algorithm assigns bit rate R_k for quantization of coefficient θ_k in such a way that the resulting quantization errors for all coefficients are forced to be equal, $\sigma_{q_0}^2 = \sigma_{q_1}^2 = \ldots = \sigma_{q_{N-1}}^2$ (Akansu and Haddad, 1992). The number of levels for the k^{th} quantizer, for coefficient θ_k, is found as $L_k = 2^{R_k}$. Optimally allocated bits R_k among multiple sources for the total bit budget R,with the assumption that all sources have the same pdf type, are calculated as (Akansu and Haddad, 1992)

$$R_k = R + \frac{1}{2}\log_2 \frac{\sigma_k^2}{\left(\prod_{i=0}^{N-1} \sigma_i^2\right)^{\frac{1}{N}}}, \tag{5.91}$$

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

5.4.3 Quantization of Eigen Subspace

In TC, sparsity in transform coefficients is desired. In contrast, any sparse transform including KLT aims to sparse subspace (transform matrix) where basis vector components are interpreted as loading coefficients in some applications (Akansu and Torun, 2012; Bollen and Whaley, 2009; Choi and Varian, 2012; Mamaysky *et al*., 2008; Ohlson and Rosenberg, 1982; Torun and Akansu, 2013). Quantization of a given subspace with an optimally designed single quantizer, Q, or a set of quantizers $\{Q_k\}$, is defined as

$$\hat{\Phi} = Q(\Phi). \tag{5.92}$$

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

$$\hat{\boldsymbol{\theta}} = \hat{\Phi}\boldsymbol{x}. \tag{5.93}$$

Unlike in TC, coefficients are not quantized in sparse representation methods. Instead, transform coefficients of the quantized subspace for a given signal vector are obtained. As in TC, quantization error equals to reconstruction error, both in mse. Mean-squared quantization error is expressed as

$$\sigma_{q,S}^2 = \frac{1}{N^2} \sum_{k=0}^{N-1} \widetilde{\phi_k}^{\mathbf{T}} \widetilde{\phi_k}, \tag{5.94}$$

where $\widetilde{\phi_k} = \phi_k - \widehat{\phi_k}$.

5.4.4 *pdf of Eigenvector*

We model the probability density of eigenvector components in order to design pdf-optimized quantizers to sparse them. Each eigenvector of an AR(1) process is generated by a sinusoidal function as expressed in (5.27). The pdf, with arbitrary support, of a continuous sinusoidal function is modeled as (Balakrishnan and Nevzorov, 2004; Hejn *et al*., 1998):

$$p(x) = \frac{1}{\pi\sqrt{(x-a)\ (b-x)}}, \tag{5.95}$$

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

$$P(x) = \frac{2}{\pi}\arcsin\left(\sqrt{\frac{x-a}{b-a}}\right). \tag{5.96}$$

Mean and variance of the arcsine distribution are calculated as

$$\mu = \frac{a+b}{2}, \tag{5.97}$$

$$\sigma^2 = \frac{(b-a)^2}{8}. \tag{5.98}$$

The pdf of arcsine distribution is symmetric and U-shaped. The arcsine distribution with $a = 0$ and $b = 1$, namely standard arcsine distribution, is a special case of the beta distribution with the parameters $\alpha = \beta = 0.5$. Figure 5.6 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 (Bagnoli and Bergstrom, 2005). It is stated in Yee (2010) 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

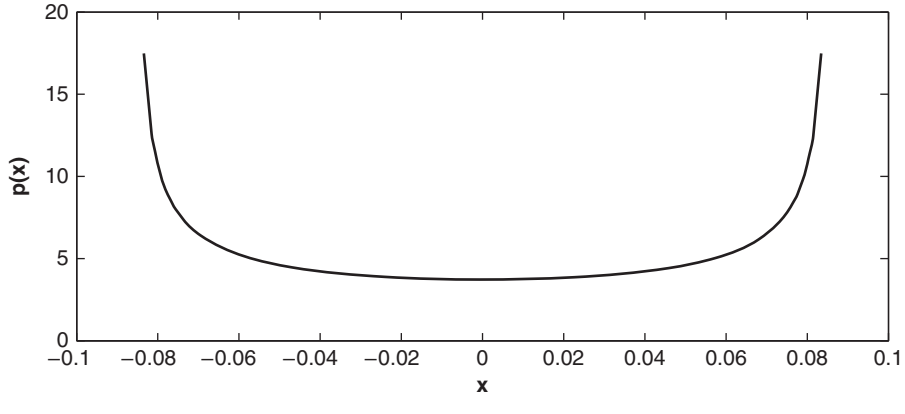


Figure 5.6 Probability density function of arcsine distribution for $a = -0.0854$ and $b = 0.0854$. Loadings of a second PC for an AR(1) signal source with $\rho = 0.9$ and $N = 256$ are fitted to arcsine distribution by finding minimum and maximum values in the PC.

quantizers can be designed for arcsine distribution (Lloyd, 1982; Max, 1960). The second principal component, ϕ_1, of an AR(1) source for $\rho = 0.9$ and size of $N = 256$ is shown to be fit by arcsine distribution with $a = \min(\phi_1)$ and $b = \max(\phi_1)$, respectively. Minimum and maximum valued components of the k*th* eigenvector depend on ρ, ω_k, and N as stated in (5.27). In order to maintain equal distortion levels among quantizers 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.

Figure 5.7a and Figure 5.7b display the normalized histograms of the first and second eigenvector components (PC1 and PC2 loading coefficients) for an AR(1) process with $\rho = 0.9$ and $N = 1024$. The value of N is selected large enough to generate proper histograms. The intervals of the histograms, Δ_k, are set as

$$\Delta_k = \frac{\max(\phi_k) - \min(\phi_k)}{N}, \tag{5.99}$$

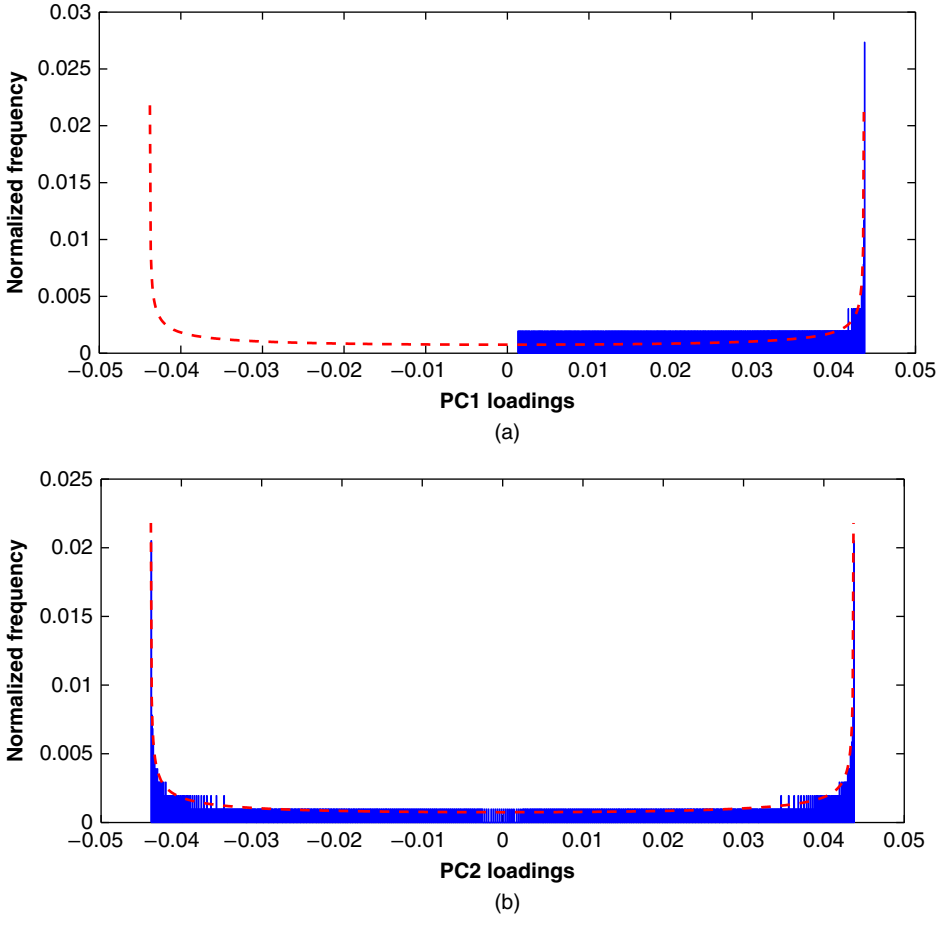


Figure 5.7 Normalized histograms of (a) PC1 and (b) PC2 loadings for an AR(1) signal source with $\rho = 0.9$ and $N = 1024$. The dashed lines in each histogram show the probability that is calculated by integrating an arcsine pdf for each bin interval.

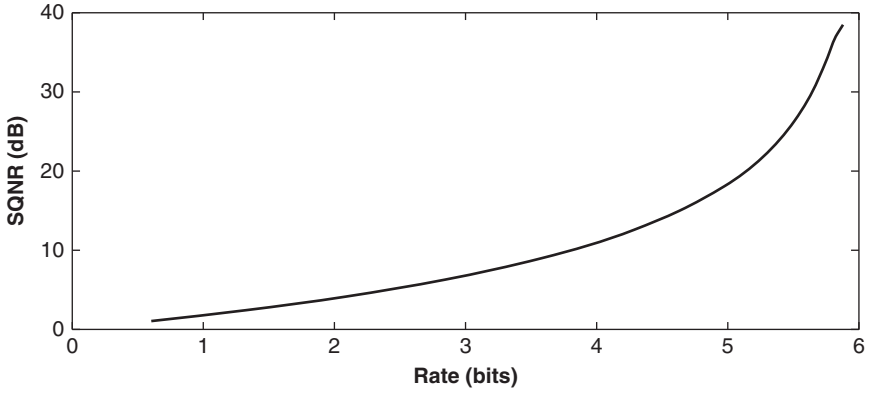


Figure 5.8 Rate (bits)-distortion (SQNR) performance of zero mean and unit variance arcsine pdf-optimized quantizer for $L = 65$ bins. The distortion level is increased by combining multiple bins around zero in a larger zero-zone.

where ϕ_k is the k*th* eigenvector. The dashed lines in each normalized histogram show the probability that is calculated by integrating the pdf of arcsine distribution in (5.95) for each bin interval. The histogram displayed in Figure 5.7a has only one side of the arcsine pdf, as expected from (5.27). In contrast, Figure 5.7b displays the histogram with a complete arcsine pdf shape. These figures confirm the arcsine distribution type for eigenvector components of an AR(1) process.

Now, we investigate the rate-distortion performance of an arcsine pdf-optimized zero-zone quantizer. Rate of quantizer output is calculated by using first-order entropy as defined in (5.89). Distortion caused by the quantizer is calculated in mse and represented in SQNR as defined in (5.87). Figure 5.8 displays rate-distortion performance of such a quantizer with $L = 65$. In this figure, distortion level is increased by increasing the zero-zone of the quantizer for more sparsity where rate decreases accordingly. One can design a quantizer with zero-zone for each eigenvector (PC) or for the entire eigenmatrix to achieve the desired level of matrix sparsity (Lloyd, 1982; Max, 1960).

5.4.5 Sparse KLT Method

In this subsection, we present in detail a simple method to sparse eigen subspace for an AR(1) process through a design example. The relevant parameter values for the sparse KLT (SKLT) example considered are tabulated in Table 5.1. The steps of design are summarized as follows.

1. First-order correlation coefficient ρ is calculated from the available dataset as described in (5.81), and $\mathbf{R}_x$ is constructed. Assume that $\rho = 0.9$ for the given example with $N = 256$.
2. Eigenvalues $\{\lambda_k\}$ and corresponding eigenvectors $\{\phi_k\}$ of $\mathbf{R}_x$ are calculated from (5.28) and (5.27), respectively. Eigenvalues of the first 16 eigenvectors (PCs) are listed in Table 5.1. These eigenvectors explain 68.28% of the total variance. Values of $\{\omega_k\}$ used to calculate each eigenvalue and corresponding eigenvector are also shown in Table 5.1. Note that $\eta_E(L)$ of (5.24) and the explained variance used in Table 5.1 measure energy compaction efficiency of a subspace.

3. PC loading coefficients (eigenvector components) are fitted to arcsine distribution by calculating $\{a_k = \min(\phi_k)\}\forall k$ and $\{b_k = \max(\phi_k)\}\forall k$. Then, variances $\left\{\sigma_k^2 = \frac{(b_k - a_k)^2}{8}\right\}\forall k$, are calculated by using (5.98). Table 5.1 also tabulates $\{a_k\}$, $\{b_k\}$, and $\{\sigma_k^2\}$ of eigenvectors.
4. For a given total rate R (desired level of sparsity), $\{R_k\}$ are calculated by plugging $\{\sigma_k^2\}$ in the optimum bit allocation equation given in (5.91). 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 5.1 displays calculated rates and quantizer levels for the total rate of $R = 5.7$.
5. For this design example, a $L = 65$ 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 the zero-zone for the desired sparsity level. For the k*th* eigenvector, a predesigned $L = 65$ level pdf-optimized the zero-zone quantizer is converted to an $L_k \leq L$ level zero-zone quantizer.
6. PC loadings (eigenvector components) are normalized to have zero mean and unit variance, $\left\{\phi_k = \frac{(\phi_k - mean(\phi_k))}{std(\phi_k)}\right\}\forall k$, 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 eigen subspace $\{\widehat{\phi_k} = Q_k(\phi_k)\}\forall k$. The number of zero components or sparsity level (S) of quantized PCs for this example are also given in Table 5.1.

Remark 5.1 *The number of bins for a predesigned pdf-optimized quantizer is selected based on the quantization noise and implementation cost. The increase in signal-to-quantization noise (SQNR) of a pdf-optimized zero-zone quantizer optimized for arcsine pdf with $L > 65$ is found not to be that significant.*

Table 5.1 Relevant parameter values of SKLT example for the first 16 PCs of an AR(1) source with $\rho = 0.9$ and $N = 256$. They explain 68.28% of the total variance.

	ω	λ	a	b	σ^2	R	L	S
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
PC12	0.1400	6.88	−0.0872	0.0872	0.0038	5.6866	53	36
PC13	0.1520	6.17	−0.0873	0.0873	0.0038	5.6881	53	36
PC14	0.1640	5.56	−0.0874	0.0874	0.0038	5.6902	53	36
PC15	0.1760	5.02	−0.0875	0.0875	0.0038	5.6915	53	36
PC16	0.1881	4.55	−0.0876	0.0876	0.0038	5.6930	53	36

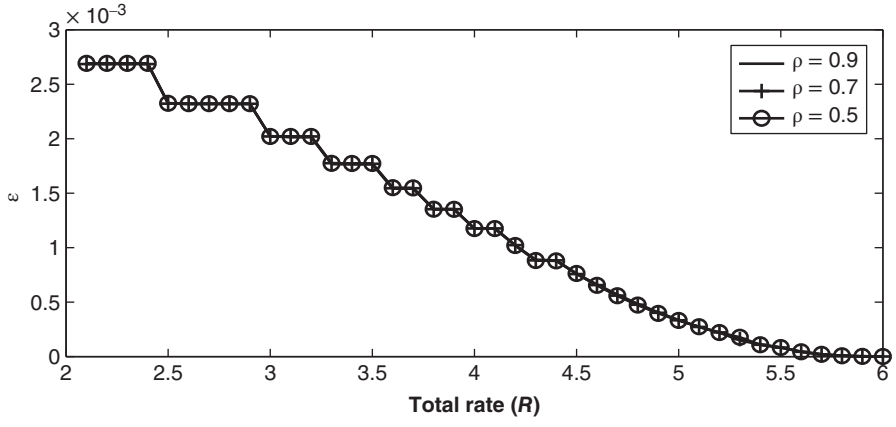


Figure 5.9 Orthogonality imperfectness-rate (sparsity) trade-off for sparse eigen subspaces of three AR(1) sources with $N = 256$.

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 defined as

$$\epsilon = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{i=0}^{N-1} [\mathbf{I}(i,j) - \mathbf{K}(i,j)]^2, \tag{5.100}$$

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

Figure 5.9 displays the trade-off between subspace sparsity and loss of orthogonality for various AR(1) sources and $N = 256$. It is observed from Figure 5.9 that the orthogonality imperfectness decreases almost linearly with increasing R, as expected.

5.4.6 Sparsity Performance

Let us compare performances of the presented SKLT method with the ST (Zou *et al*., 2006), SPCA (Zou *et al*., 2006), DSPCA (d'Aspremont *et al*., 2007), and SPC (d'Aspremont *et al*., 2007) methods for the AR(1) process, and also for an empirical correlation matrix of stock returns in the NASDAQ-100 index in this subsection. In order to provide a fair comparison, sparsity levels of all methods considered here are tuned in such a way that compared PCs have almost the same number of nonzero components. In most cases, the number of nonzero components of each PC in the SKLT method are kept slightly lower than the others in order to show the method's merit under mildly disadvantageous test conditions.

5.4.6.1 Eigen Subspace Sparsity for the AR(1) Process

The sparsity imposed on PCs may degrade the explained variance described in d'Aspremont *et al*. (2007). The explained variances of the PCs are calculated as $\{\lambda_k = \sigma_k^2 = \phi_k^{\mathrm{T}} \mathbf{R}_x \phi_k\} \forall k$, where ϕ_k is the k*th* eigenvector for a given $\mathbf{R}_x$. For the sparsed PCs, new explained

variances (eigenvalues) are calculated as $\left\{\widehat{\lambda}_k = \widehat{\sigma_k^2} = \widehat{\phi_k}^{\mathbf{T}} \mathbf{R}_x \widehat{\phi_k}\right\} \forall k$, where $\widehat{\phi_k}$ is the k*th* sparse eigenvector. Then, the percentage of explained variance loss (VL) as a performance metric is defined as $\left\{V_k = \frac{(\lambda_k - \widehat{\lambda}_k)}{\lambda_k} \times 100\right\} \forall k$. Similarly, the cumulative explained variance loss of first L PCs is defined as $C_L = \sum_{k=1}^{N} \lambda_k - \sum_{k=1}^{L} \widehat{\lambda}_k$. In addition, we also used a nonsparsity (NS) performance metric for comparison. It is defined as the percentage of nonzero components in a given sparsed eigenvector. Thus, the performance is measured as the variance loss for the given NS level (d'Aspremont *et al*., 2007; Zou and Hastie, 2005; Zou *et al*., 2006). 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.

Figure 5.10a displays the VL measurements of sparsed first PCs generated by SKLT, SPCA, SPC, ST, and DSPCA methods with respect to NS for an AR(1) source with $\rho = 0.9$ and $N = 256$. For SKLT, an $L = 65$ level quantizer optimized for arcsine pdf with zero mean and

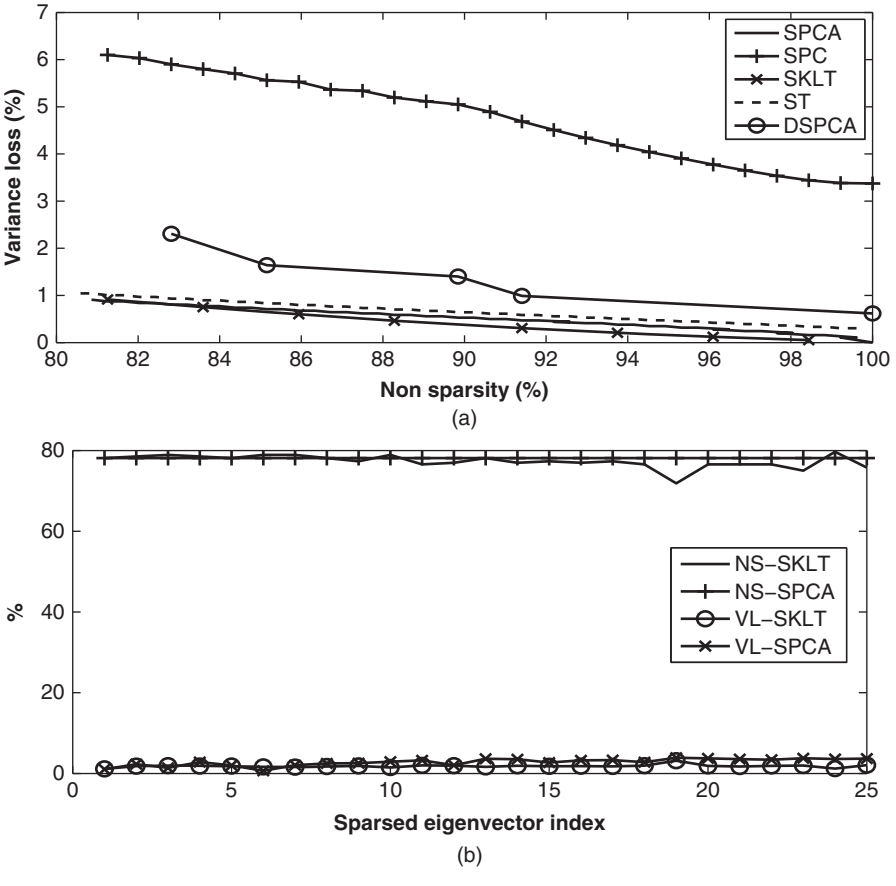


Figure 5.10 (a) Variance loss (VL) measurements of sparsed first PCs generated by SKLT, SPCA, SPC, ST, and DSPCA methods with respect to nonsparsity (NS) for an AR(1) source with $\rho = 0.9$ and $N = 256$; (b) NS and VL measurements of sparsed eigenvectors for an AR(1) source with $\rho = 0.9$ and $N = 256$ generated by the SKLT method and SPCA algorithm.

unit variance is used as the initial quantizer. The zero-zone width of the initial quantizer is adjusted for required sparsity, as explained in this chapter. Then, the generated quantizer is employed. Figure 5.10a shows that SKLT offers less variance loss than the other methods. SPCA provides competitive performance to SKLT. Figure 5.10b displays NS and 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. Figure 5.10b shows that the VL performance of SKLT is slightly better than that of SPCA. Note that the NS of SKLT is slightly lower than that of SPCA in this comparison.

5.4.6.2 Eigen Subspace Sparsity for the NASDAQ-100 Index

We present an example to sparse an eigen subspace that leads to the creation of the corresponding sparse eigenportfolios (Akansu and Torun, 2015). Eigendecomposition of empirical correlation matrices is a popular mathematical tool in finance employed for various tasks including eigenfiltering of measurement noise and creation of eigenportfolios for baskets of stocks (Akansu and Torun, 2012; Markowitz, 1959; Torun *et al*., 2011). An empirical correlation matrix for the end-of-day (EOD) stock returns for the NASDAQ-100 index with $W = 30$-day time window ending on April 9, 2014, is measured (Torun *et al*., 2011). The vector of 100 stock returns in the NASDAQ-100 index at time n is created as (Akansu and Torun, 2012):

$$\mathbf{r}(n) = [r_k(n)]; k = 1, 2, \ldots, 100. \tag{5.101}$$

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

$$\begin{aligned} \mathbf{R}_E(n) &\triangleq [E\{\mathbf{r}(n)\mathbf{r}^T(n)\}] = [R_{k,l}(n)], \\ &= \begin{bmatrix} R_{1,1}(n) & R_{1,2}(n) & \ldots & R_{1,100}(n) \\ R_{2,1}(n) & R_{2,2}(n) & \ldots & R_{2,100}(n) \\ \vdots & \vdots & \ddots & \vdots \\ R_{100,1}(n) & R_{100,2}(n) & \ldots & R_{100,100}(n) \end{bmatrix}, \end{aligned} \tag{5.102}$$

where the matrix elements

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

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:

$$\mathbf{R}_E(n) = \mathbf{A}_{KLT}^T \mathbf{\Lambda} \mathbf{A}_{KLT} = \sum_{k=1}^{N} \lambda_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k^T, \tag{5.103}$$

where $\{\lambda_k, \boldsymbol{\phi}_k\}$ are eigenvalue–eigenvector pairs (Akansu and Torun, 2012).

The component values of eigenvector $\{\boldsymbol{\phi}_k\}$ are repurposed as the capital allocation coefficients to create the k*th* eigenportfolio for a group of stocks where the resulting coefficients $\{\theta_k\}$ are pairwise uncorrelated. These coefficients represent eigenportfolio returns. Eigenportfolios are used in various investment and trading strategies (Chamberlain and Rothschild, 1983). It is

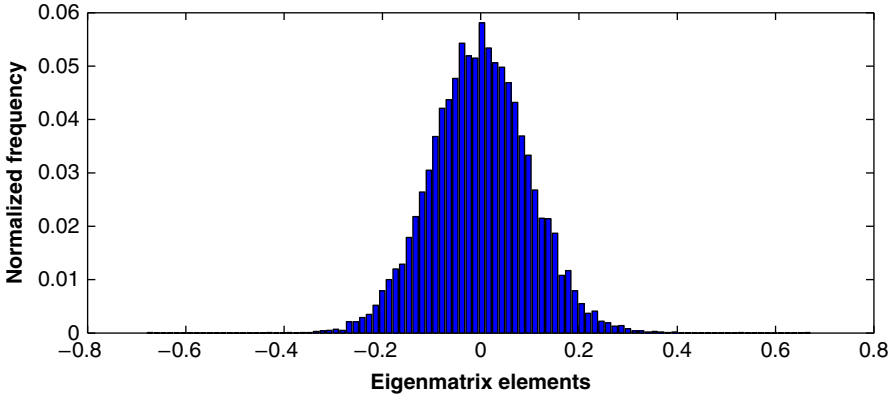


Figure 5.11 Normalized histogram of eigenmatrix elements for an empirical correlation matrix of end-of-day (EOD) returns for 100 stocks in the NASDAQ-100 index. $W = 30$-day measurement window ending on April 9, 2014.

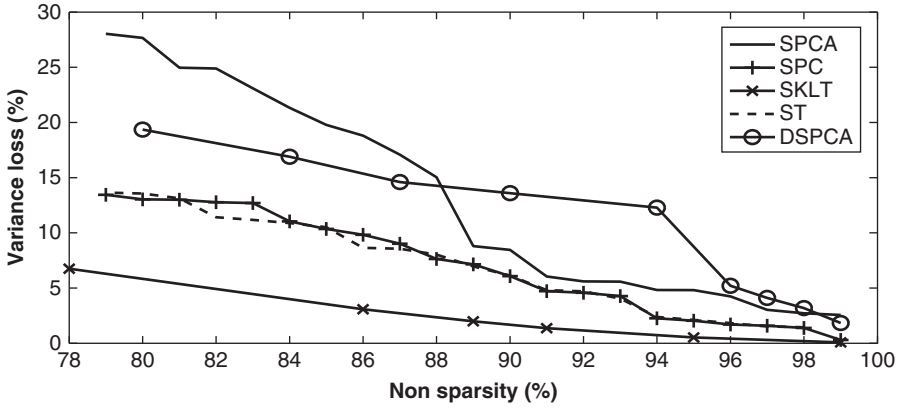
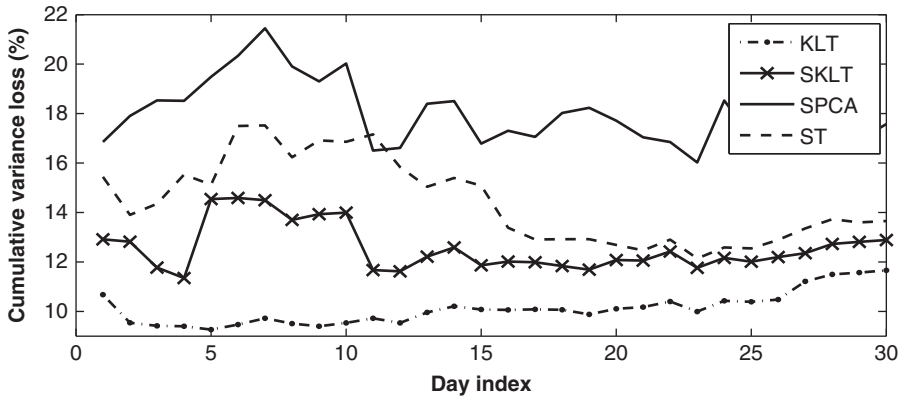
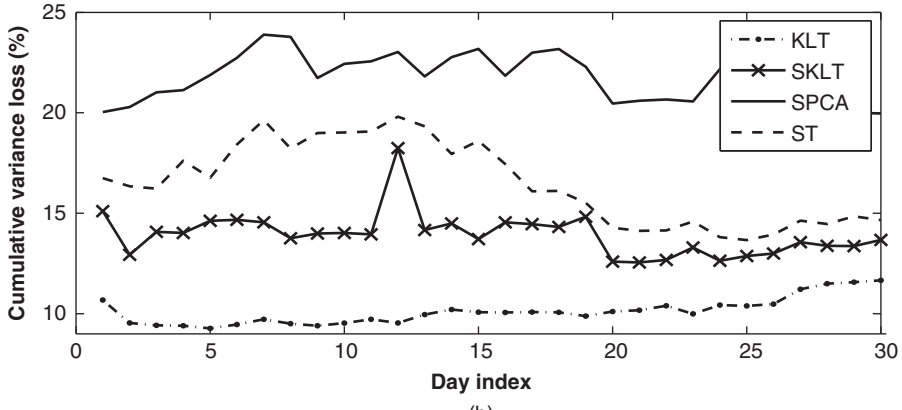


Figure 5.12 VL measurements of sparsed first PCs generated by SKLT, SPCA, SPC, ST, and DSPCA methods with respect to NS for an empirical correlation matrix of EOD returns for 100 stocks in the NASDAQ-100 index with $W = 30$-day measurement window ending on April 9, 2014.

required to buy and sell certain stocks in the 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 eigen subspace of an empirical correlation matrix $\mathbf{R}_E(n)$ may offer cost reductions in desired 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 original eigenportfolios. Therefore, such a sparse representation might cause a significant deviation from the measured empirical correlation matrix. Hence, the use of eigenportfolios generated by a sparsity-constrained optimization in a trading strategy may lead to poor performance.





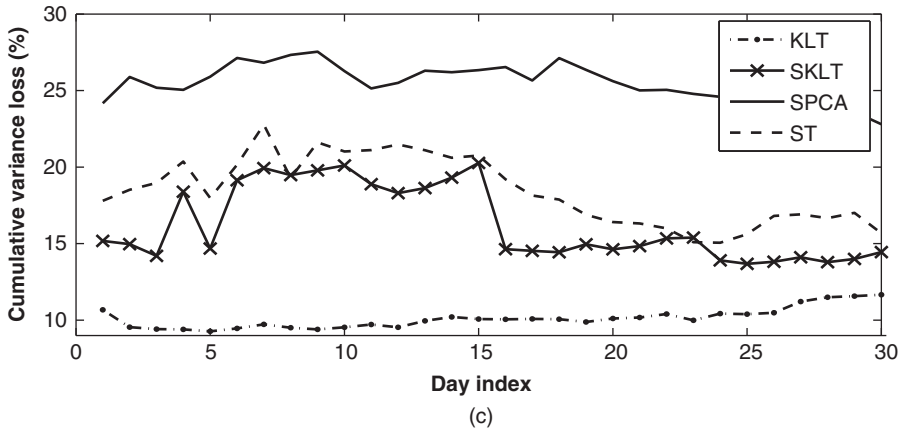


Figure 5.13 Cumulative explained variance loss with $L = 16$ generated daily from an empirical correlation matrix of EOD returns between April 9, 2014, and May 22, 2014, for 100 stocks in the NASDAQ-100 index by using KLT, SKLT, SPCA, and ST methods. NS levels of 85%, 80%, and 75% for all PCs are forced in (a), (b), and (c), respectively, using $W = 30$ days.

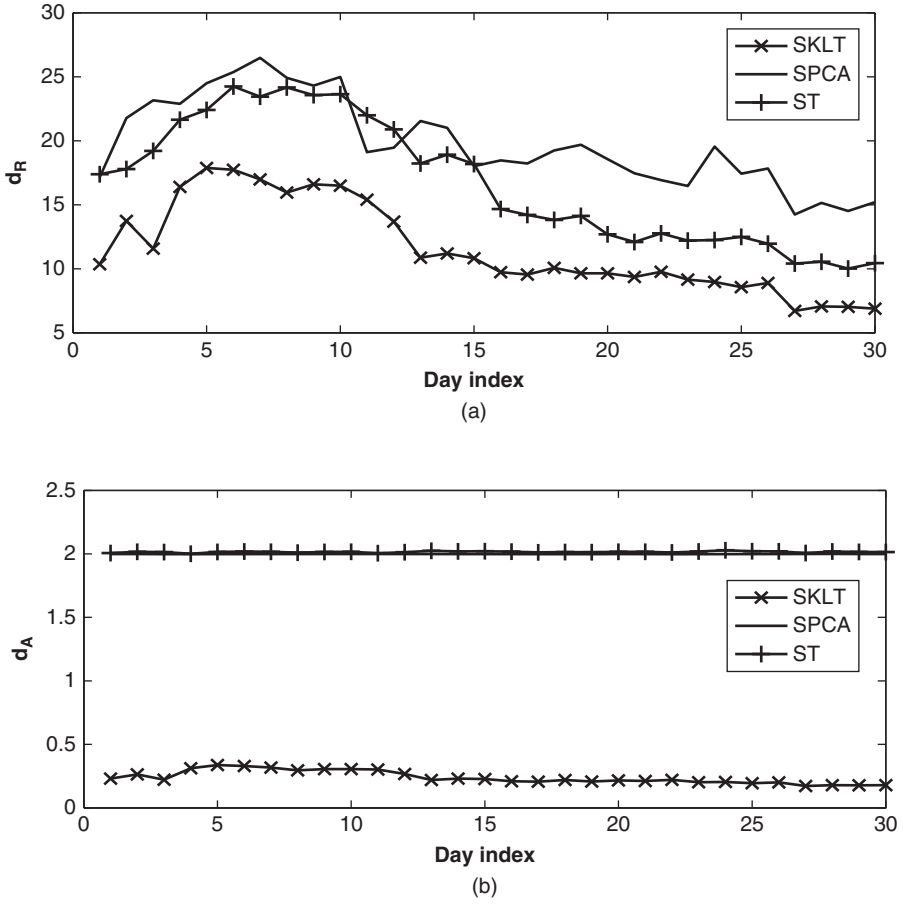


Figure 5.14 (a) $d_{\mathbf{R}}$ and (b) $d_{\mathbf{A}}$ of sparse eigen subspaces generated daily from an empirical correlation matrix of EOD returns between April 9, 2014, and May 22, 2014, for 100 stocks in the NASDAQ-100 index by using SKLT, SPCA, and ST methods, respectively. NS level of 85% for all PCs is forced with $W = 30$ days.

For simplicity, a single quantizer in the SKLT method is utilized to sparse the entire eigenmatrix $\mathbf{A}_{\text{KLT}}$. It is optimized for the histogram of its elements as displayed in Figure 5.11. It is observed to be a Gaussian pdf. Figure 5.12 displays the VL measurements of sparsed first PCs generated by SKLT, SPCA, SPC, ST, and DSPCA methods with respect to NS. It is shown that SKLT offers less variance loss than other methods considered in this chapter. Figure 5.13 displays the cumulative explained variance loss with $L = 16$ generated daily from an empirical correlation matrix of EOD returns between April 9, 2014, and May 22, 2014, for 100 stocks in the NASDAQ-100 index by using KLT, SKLT, SPCA, and ST methods. The measurement window of the last 30 days, $W = 30$, is used. NS levels of 85%, 80%, and 75% for each PC are forced in experiments as displayed in Figure 5.13a, Figure 5.13b, and Figure 5.13c, respectively. The SKLT method consistently outperforms the others for this time-varying scenario.

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

$$d_{\mathbf{R}} = \left\| \mathbf{R}_E(n) - \widehat{\mathbf{R}_E(n)} \right\|_2, \tag{5.104}$$

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

$$d_{\mathbf{A}} = \left\| \mathbf{A}_{KLT} - \widehat{\mathbf{A}_{KLT}} \right\|_2. \tag{5.105}$$

Figure 5.14a and Figure 5.14b display the $d_{\mathbf{R}}$ and $d_{\mathbf{A}}$ of sparse eigen subspaces generated daily from an empirical correlation matrix of EOD returns between April 9, 2014, and May 22, 2014, for 100 stocks in the NASDAQ-100 index by using SKLT, SPCA, and ST methods, respectively. The NS level of 85% for all PCs is forced with $W = 30$ days. These performance measures highlight that the SKLT method sparses eigen subspace of the NASDAQ-100 index better than the SPCA and ST methods. Moreover, the SKLT does not force an alteration of the actual covariance structure like other methods.

5.5 Conclusions

Closed-form expressions for KLT kernel (eigenvectors) and corresponding transform coefficient variances (eigenvalues) of the AR(1) process were reported in the literature (Ray and Driver, 1970). However, they require solving a transcendental tangent equation (5.29). Mathematical steps leading to equations (5.27), (5.28), and (5.29) are discussed in detail, following the methodology used for a continuous stochastic process with exponential autocorrelation function (Davenport and Root, 1958; Pugachev, 1959a,b; Wilkinson, 1965). Then, a simple and fast method to find the roots of a transcendental equation is employed to derive the roots of (5.29) explicitly. That derivation made it possible to express the $N \times N$ KLT kernel and corresponding coefficient variances in explicit form, leading to extremely fast KLT implementations for processes that can be modeled with AR(1) process. The technique is shown to be more efficient than the D&Q algorithm (Golub and Loan, 1996).

The constrained optimization algorithms to generate sparse PCs do not guarantee a good performance for an arbitrary covariance matrix due to the nonconvex nature of the problem. The SKLT method to sparse subspaces is presented in this chapter. It utilizes the mathematical framework developed for transform coding in rate-distortion theory. The sparsity performance comparisons demonstrate the superiority of SKLT over the popular algorithms known in the literature.

References

Akansu, A.N. and Haddad, R.A. (1992). *Multiresolution signal decomposition: transforms, subbands, and wavelets*. New York: Academic Press.

Akansu, A.N. and Torun, M.U. (2012). Toeplitz approximation to empirical correlation matrix of asset returns: a signal processing perspective. *Journal of Selected Topics in Signal Processing*, 6 (4), 319–326.

Akansu, A.N. and Torun, M.U. (2015). *A primer for financial engineering: financial signal processing and electronic trading*. New York: Academic Press-Elsevier.

Allen, M. and Isaacson, E. (1997). *Numerical analysis for applied science*. New York: John Wiley & Sons.

Atal, B. and Hanauer, S. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, 50, 637.

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, 26 (2), 445–469.

Balakrishnan, N. and Nevzorov, V.B. (2004). *A primer on statistical distributions*. Hoboken, NJ: John Wiley & Sons.

Berger, T. (2003). *Rate-distortion theory*. Hoboken, NJ: John Wiley and Sons.

Bertero, M. and Boccacci, P. (1998). *Introduction to inverse problems in imaging*. London: Institute of Physics Publishing.

Bollen, N.P.B. and Whaley, R.E. (2009). Hedge fund risk dynamics: implications for performance appraisal. *The Journal of Finance*, 64 (2), 985–1035.

Brodie, J., Daubechies, I., De Mol, C., Giannone D. and Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106 (30), 12267–12272.

Brown, J. and Churchill, R. (2009). *Complex variables and applications*. New York: McGraw-Hill.

Brusewitz, H. (1986). Quantization with entropy constraint and bounds to the rate distortion function, ser. Trita-TTT. *Telecommunication Theory, Electrical Engineering*, 8605, 28–29.

Cadima, J. and Jolliffe, I.T. (1995). Loading and correlations in the interpretation of principle components. *Journal of Applied Statistics*, 22 (2), 203–214.

Chamberlain, G. and Rothschild, M. (1983) Arbitrage, factor structure and mean-variance analysis on large asset markets. *Econometrica*, 51 (5), 1281–1304.

Choi, H. and Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88, 2–9.

Clarke, R.J. (1985). *Transform coding of images*. New York: Academic Press.

d'Aspremont, A. El Ghaoui, L. Jordan, M.I. and Lanckriet, G.R.G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Review*, 49 (3), 434–448.

Davenport, W.B. and Root, W.L. (1958). *An introduction to the theory of random signals and noise*. New York: McGraw-Hill.

Doob, J. (1942). The Brownian movement and stochastic equations. *Annals of Mathematics*, 43 (2), 351–369.

Engl, H.W. Hanke, M. and Neubauer, A. (1996). *Regularization of inverse problems*. New York: Kluwer Academic.

Golub, G.H. and Loan, C.F.V. (1996). *Matrix computations*. Baltimore: Johns Hopkins University Press.

Gonzales, C. and Akansu, A.N. (1997). A very efficient low-bit-rate subband image/video codec using shift-only PR-QMF and zero-zone linear quantizers. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 4, 2993–2996.

Hajnal, A. and Juhasz, I. (1983). Remarks on the cardinality of compact spaces and their Lindelof subspaces. *Proceedings of the American Mathematical Society*, 51 (5), 146–148.

Hejn, K., Pacut, A. and Kramarski, L. (1998). The effective resolution measurements in scope of sine-fit test. *IEEE Transactions on Instrumentation and Measurement*, 47 (1), 45–50.

Jayant, N.S. and Noll, P. (1984). *Digital coding of waveforms: principles and applications to speech and video*. Englewood Cliffs, NJ: Prentice Hall Professional Technical Reference.

Jolliffe, I.T. (2002). *Principal component analysis*. New York: Springer-Verlag.

Kay, S. (1988). *Modern spectral estimation: theory and application*. Upper Saddle River, NJ: Prentice Hall.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2), 129–137.

Luck, R. and Stevens, J. (2002). Explicit solutions for transcendental equations. *SIAM Review*, 44, 227–233.

Mamaysky, H. Spiegel, M. and Zhang, H. (2008). Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies*, 21, 233–264.

Markowitz, H.M. (1959). *Portfolio selection: efficient diversification of investments*. New York: John Wiley & Sons.

Max, J. (1960). Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6 (1), 7–12.

Ohlson, J. and Rosenberg, B. (1982). Systematic risk of the CRSP equal-weighted common stock index: a history estimated by stochastic-parameter regression. *The Journal of Business*, 55 (1), 121–145.

Papoulis, A. (1991). *Probability, random variables, and stochastic processes*. New York: McGraw-Hill.

Pugachev, V. (1959a). A method for the determination of the eigenvalues and eigenfunctions of a certain class of linear integral equations. *Journal of Applied Mathematics and Mechanics* (translation of the Russian journal *Prikladnaya Matematika i Mekhanika*), 23 (3), 527–533.

Pugachev, V. (1959b). A method of solving the basic integral equation of statistical theory of optimum systems in finite form. *Journal of Applied Mathematics and Mechanics (translation of the Russian journal Prikladnaya Matematika i Mekhanika)*, 23 (1), 3–14.

Ray, W. and Driver, R. (1970). Further decomposition of the Karhunen-Loeve series representation of a stationary random process. *IEEE Transactions on Information Theory*, 16 (6), 663–668.

Shen, H. and Huang, J.Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, 99 (6), 1015–1034.

Sriperumbudur, B.K. Torres, D.A. and Lanckriet, G.R.G. The sparse eigenvalue problem. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.243.8515&rep=rep1&type=pdf

Strang, G. (1986). *Introduction to applied mathematics*. Wellesley, MA: Wellesley-Cambridge Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Torun, M.U. and Akansu, A.N. (2013). An efficient method to derive explicit KLT kernel for first-order autoregressive discrete process. *IEEE Transactions on Signal Processing*, 61 (15), 3944–3953.

Torun, M.U., Akansu, A.N. and Avellaneda, M. (2011). Portfolio risk in multiple frequencies. *IEEE Signal Processing Magazine, Special Issue on Signal Processing for Financial Applications*, 28 (5), 61–71.

Trendafilov, N. Jolliffe, I.T. and Uddin, M. (2003). A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, 12, 531–547.

Uhlenbeck, G. and Ornstein, L. (1930). On the theory of the Brownian motion. *Physical Review*, 36, 823–841.

Van Trees, H.L. (2001). *Detection, estimation, and modulation theory*. New York: John Wiley & Sons.

Wilkinson, J. (1965). *The algebraic eigenvalue problem*. Oxford: Oxford University Press.

Witten, D.M. Tibshirani, R. and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10 (3), 515.

Yee, V.B. (2010). Studies on the asymptotic behavior of parameters in optimal scalar quantization. Ph.D. dissertation, The University of Michigan.

Zass, R. and Shashua, A. (2006). Nonnegative sparse PCA. In *Advances in* neural information processing systems (ed. Scholkopf, B., Platt, J. and Hoffman, T.). Cambridge, MA: MIT Press, 1561–1568.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67 (2), 301–320.

Zou, H. Hastie, T. and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15, 262–286.