Digital Signal Processing 32 (2014) 79–84

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Atmospheric radar signal processing using principal component analysis

D. Uma Maheswara Rao, T. Sreenivasulu Reddy *, G. Ramachandra Reddy

Sri Venkateswara University College of Engineering, Department of Electronics and Communication Engineering, Tirupati, Andhra Pradesh, 517 502, India

ARTICLE INFO

Article history:
Available online 27 May 2014

Keywords:
Principal component analysis
MST radar
GPS sonde
Wavelet-based denoising
Cepstral thresholding

ABSTRACT

Principal Component Analysis (PCA) is a simple non-parametric method for extracting relevant information from high-dimensional data sets. In this paper, we analyze the data collected from the Indian MST (Mesosphere, Stratosphere, Troposphere) radar at Gadanki (13.5°N, 79.2°E) using PCA. We tested the PCA for various simulated signals like narrowband, wideband and exponential signals which may contain more than one frequency both in absence and presence of noise. For the simulated data, it is observed that PCA works for low SNR, i.e. it successfully detects the frequency in the highly noise-corrupted signal also. Finally, we applied PCA to the radar data for estimating the power spectrum and thus in turn estimating the Doppler frequency components. We estimate the zonal (U), meridional (V), wind speed (W) etc. from the Doppler frequencies. Compared with existing algorithms, PCA works well at higher altitudes and results have been validated using the GPS sonde data.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Indian MST radar provides information on wind data in the mesosphere, stratosphere and troposphere with a resolution of 150 m starting above 3.5 km. The radar uses Doppler Beam Swinging (DBS) method to determine the three wind components U, V and W [1]. The spectral data are collected by the radar using multiple beam positions (east, west, zenith-X, zenith-Y, north and south) with 16 μs coded pulse and 1000 μs Inter Pulse Period (IPP). The complex time series of the decoded and integrated signal samples are subjected to the process of Fast Fourier Transform (FFT) for on-line computation of the Doppler power spectra for each bin of the selected range window. The off-line data processing involves the following steps: the removal of dc, estimation of average noise level, the removal of interference, incoherent integration and computation of low-order (0th, 1st and 2nd) moments. The three moments are signal strength, weighted mean Doppler shift and half-width parameters of the spectrum respectively. Up to a certain height ($\leq$18 km), this technique estimates the Doppler frequencies of the returned echoes accurately. At heights greater than 18 km, estimation fails. It is also observed that in many cases that when noise interferes with data at lower altitudes (3.5 km to 12 km), we get incorrect results. Several authors proposed various algorithms for denoising the spectrum, finding the Doppler frequencies from the estimated spectrum and thus the U, V and W components. Bispectral-based estimation algorithm eliminates the noise [2]. But, this algorithm involves a complex mathematical computation. Multitaper spectral estimation algorithm produces broadened spectral peak [3]. Wavelet-based denoising method has been applied for spectrum cleaning and thus estimating the Doppler frequencies and wind components [4]. The Cepstrum thresholding approach also has been applied to the radar data to estimate the frequencies [5]. These methods have an advantage of total variance reduction of the estimated spectrum. But, they fail at higher heights where SNR will be very low. Hence, there should be an algorithm which yields correct results at medium as well as higher altitudes and take an advantage of reduced variance. PCA is such method which reduces the complexity and gives us the good results.

The remainder of this paper is organized as follows. In Section 2, the PCA concept and its properties for exponentials are explained. In Section 3, we showed the reduction in variance and mean square error (MSE) when PCA is applied to the simulated data and illustrate the detection of frequencies of one or more sinusoids in presence of high noise. We applied our approach to the MST radar data to estimate the Doppler shift in Section 4. Wind velocities are computed from the estimated Doppler shifts. In Section 5, the concluding remarks are given.

2. Principal component analysis

PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated

* Corresponding author.
E-mail address: mettu86@yahoo.co.in (T. Sreenivasulu Reddy).

http://dx.doi.org/10.1016/j.dsp.2014.05.009
1051-2004/© 2014 Elsevier Inc. All rights reserved.

variables into a set of values of uncorrelated variables called principal components (PCs). The main advantage of PCA is dimensionality reduction. There are various spectrum estimation algorithms to estimate the frequency. But, there will be certain cases where we need to estimate only the frequencies and amplitudes of the spectral components. There is no need to estimate the entire spectrum. These are known as frequency estimation techniques and are applicable to a harmonic process that consists of a sum of sinusoids or complex signals. These methods may use the vectors that lie in noise subspace or signal subspace. The signal subspace methods form a low-rank approximation to the autocorrelation matrix which is then incorporated to a spectrum estimation algorithm. PCA is one of such signal subspace methods.

The autocorrelation matrix (ACM) for the given set of data $x(n)$ consisting of p exponentials plus noise is a sum of autocorrelation matrices due to signal s and noise n [6]. Let the size of ACM be $M \times M$. Thus,

$$\mathbf{R}_x = \mathbf{R}_s + \mathbf{R}_n \tag{1}$$

The Eigen decomposition of the ACM can be expressed as below assuming that the eigenvalues (λ_i) are arranged in the descending order ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$),

$$\mathbf{R}_x = \sum_{i=1}^{M} \lambda_i \mathbf{v}_i \mathbf{v}_i^H = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^H + \sum_{i=p+1}^{M} \lambda_i \mathbf{v}_i \mathbf{v}_i^H \tag{2}$$

where $\mathbf{v}_i$ is eigenvector corresponding to eigenvalue λ_i. The first term of (2) is due to signal alone and second term of (2) is due to noise alone. If we retain only the principal eigenvectors of (2), a reduced rank approximation is formed to the signal ACM. Thus,

$$\hat{\mathbf{R}}_s = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^H \tag{3}$$

Now, any spectral estimator can be used for the above approximated ACM of (3). The noise part of (2) is eliminated and only signal part is retained. Thus, the estimation of spectral component due to signal is enhanced. The principal components representation of (2) imposes a rank-p constraint on $\mathbf{R}_x$ since it is assumed that the signal has p exponentials and also the rank of the ACM due to the signal is p.

The number of principal components is less than or equal to the number of original variables. The first principal component has high variance as it accounts for as much of the variability in the data as possible. The succeeding component in turn has the highest variance possible under the constraint that it has to be orthogonal to (uncorrelated with) the preceding components. Its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. Fig. 1 shows the steps involved in PCA spectral estimate.

The PCA of the ACM may be used in conjunction with any of the spectrum estimation techniques and thus forming principal components spectrum estimate. The following are some of the methods used for this PC spectrum estimate.

1. Blackman–Tukey Frequency Estimation method (PCA-BT),
2. Minimum variance Frequency Estimation method (PCA-MV) and
3. AR Frequency Estimation method (PCA-AR).

In this paper, we implemented the first two methods. The equations of the PC version of the above first two methods are given by,

$$\hat{P}_{\text{PCA-BT}}(e^{jw}) = \frac{1}{M} \mathbf{e}^H \hat{\mathbf{R}}_s \mathbf{e} = \frac{1}{M} \sum_{i=1}^{p} \lambda_i \left| \mathbf{e}^H \mathbf{v}_i \right|^2 \tag{4}$$

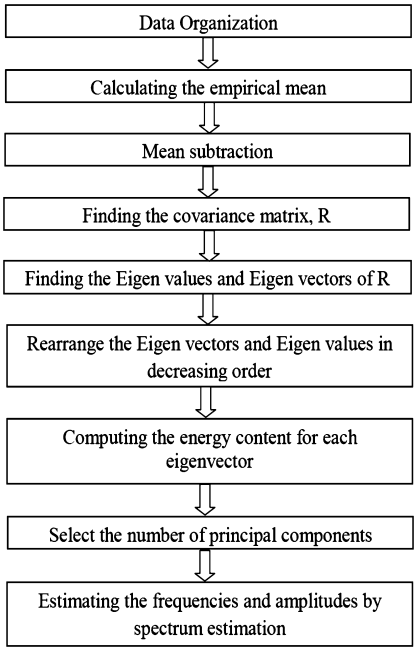

Fig. 1. Flowchart of the steps involved in PCA.

$$\hat{P}_{\text{PCA-MV}}(e^{jw}) = \frac{M}{\sum_{i=1}^{p} \frac{1}{\lambda_i} |\mathbf{e}^H \mathbf{v}_i|^2} \tag{5}$$

where $\mathbf{e}$ is the vector of complex exponentials orthogonal to $\mathbf{v}_i$, $i = 1, 2, \ldots, p$.

Selecting the number of PCs

Let $\mathbf{R}$ be the $M \times M$ covariance matrix obtained from the mean subtracted data vector, $x(n)$. $\mathbf{R}$ can be expressed as,

$$\mathbf{V}^{-1} \mathbf{R} \mathbf{V} = \mathbf{D} \tag{6}$$

In (6), $\mathbf{V}$ is the matrix of eigenvectors that diagonalizes the covariance matrix $\mathbf{R}$ and $\mathbf{D}$ is the diagonal matrix of eigenvalues of $\mathbf{R}$. $\mathbf{D}$ is an $M \times M$ diagonal matrix, where

$$\mathbf{D}[p, q] = \begin{cases} \lambda_m, & p = q = m \\ 0, & p \neq q \end{cases} \tag{7}$$

with λ_m being the mth eigen value and the elements of the diagonal matrix $\mathbf{D}$ are in descending order. The eigenvalues represent the distribution of source data's energy among each of the eigenvectors. The cumulative energy content $\mathbf{E}$ of the mth eigenvector is the sum of the energy content across all of the eigenvalues from 1 through m.

$$\mathbf{E}[m] = \sum_{q=1}^{m} \mathbf{D}[q, q], \quad m = 1, \ldots, M \tag{8}$$

Save the first L columns of $\mathbf{V}$ as the $M \times L$ matrix $\mathbf{W}$.

$$\mathbf{W}[p, q] = \mathbf{V}[p, q], \quad p = 1, \ldots, M, \ q = 1, \ldots, L \tag{9}$$

where $1 \leq L \leq M$. The $\mathbf{E}[m]$ can be used as a guide in choosing an appropriate value of L. The value of L is as small as possible while achieving a reasonably high value of $\mathbf{E}$ on a percentage basis. For example, if we want to choose L so that the cumulative energy $\mathbf{E}[m]$ is above a certain threshold, like 90 percent. The smallest value of L is chosen such that,

$$\frac{\mathbf{E}[m = L]}{\sum_{q=1}^{M} \mathbf{D}[q, q]} \geq 90\% \tag{10}$$

Thus, the number of PCs is selected.

Download English Version:

https://daneshyari.com/en/article/558750

Download Persian Version:

https://daneshyari.com/article/558750

Daneshyari.com