EFFECT OF NORMALIZATION OF EIGENVECTORS ON THE PAST AND RP ALGORITHMS FOR PCA

R. Landqvist, A. Mohammed

School of Engineering, Blekinge Institute of Technology
PO BOX 520, 372 25, Ronneby, Sweden
phone: + (46) 457 385713, fax: + (46) 457 27914, email: ronnie.landqvist@bth.se
web: http://www.its.bth.se/staff/rgs/

ABSTRACT

This paper investigates the effect of incorporating normalization of the eigenvectors between iterations in the PAST and RP algorithms for principal component analysis (PCA). In addition, an algorithm denoted as exact eigendecomposition (EE) is proposed for PCA. The algorithms are compared for different configurations using Monte Carlo simulations. Simulation results show that EE has the best performance and that normalization may be used for improving PAST and RP.

1. INTRODUCTION

Principal component analysis (PCA) is an important concept in statistical signal processing. PCA is used in various applications such as feature extraction (compression), signal estimation and detection. In blind signal separation, PCA may be used to prewhite (i.e. decorrelate) signals before independent component analysis (ICA) is applied for finding the independent signal components. ICA algorithms are generally more efficient if the input data is white since the number of possible solutions to the problem decreases significantly.

There are two different approaches to PCA. If the whole data set is available, analytical methods can be used to calculate the principal components (PC). On the other hand, if PCA is to be used in real time applications, then the PCs have to be estimated on-line for each new sample. Algorithms that operate on the whole date set are generally denoted off-line algorithms, while the latter type is denoted on-line algorithms.

Two on-line algorithms for PCA are the projection approximation subspace tracking (PAST) algorithm and the Rao-Principe (RP) algorithm. The first algorithm was proposed by Yang in [1]. Essentially, PAST finds the PCs by minimizing the linear PCA criterion [2,3,4] using a gradient-descent technique or any recursive least squares variant. In this paper, the well known recursive least squares (RLS) algorithm [5] is used. The second algorithm, RP, was proposed by Rao and Principe in [6,7]. RP estimates the PCs by using update equations derived from the Rayleigh quotient, without the use of any external parameter such as a step-size or forgetting factor. PAST, on the other hand, relies on a forgetting factor parameter to be tuned before it can be used.

This paper investigates the effect of incorporating normalization of the eigenvectors between iterations in the PAST and RP algorithms. In addition, a new on-line algorithm for PCA, denoted the exact eigendecomposition (EE) algorithm, is proposed. EE is based on direct estimation of the correlation matrix followed by calculation of the PCs using the EIG-command in Matlab. The performance of these algorithms will be assessed for different configurations using Monte Carlo computer simulations in Matlab.

2. PRINCIPAL COMPONENT ANALYSIS

On-line PCA can be viewed as a functional block where the input is a complex m-by-1 data vector $\mathbf{x}(n)$ at the nth time instant. At each time instant the outputs are estimated eigenvectors and eigenvalues of the correlation matrix $\mathbf{R}$ of the data. These eigenvectors and eigenvalues are denoted PCs of the data. Note that in this paper it is assumed that the correlation matrix $\mathbf{R}$ of the data is time-invariant.

2.1 The PAST algorithm

The PAST algorithm [1] minimizes an approximation of the linear PCA cost criterion,

$$J(\mathbf{w}(n)) = \sum_{i=1}^{n} \beta^{n-i} \left\| \mathbf{x}(i) - \mathbf{w}(n)x'(i) \right\|^2 ,$$

where $x'(n) = \mathbf{w}^H(n-1)\mathbf{x}(n)$, β is a scalar forgetting factor, $\mathbf{w}(n)$ is an m-by-1 coefficient vector and $\|\cdot\|$ denotes the vector norm. This approximate version of the cost criterion is quadratic and has the same form as the cost criterion of the RLS algorithm, with exception of the error signal (inside the vector norm) that is a vector instead of a scalar. Thus, the minimization may be performed by incorporating the new signal $x'(n)$ in the RLS. The RLS will update the coefficient vector $\mathbf{w}(n)$ using $x'(n)$ as the input signal and $\mathbf{x}(n)$ as the desired signal. When the algorithm has converged, $\mathbf{w}(n)$ will contain the eigenvector corresponding to the largest eigenvalue of the correlation matrix $\mathbf{R}$, i.e. the largest PC. The eigenvalue can be found directly by the RLS via a reformulation of the update equation for the inverse correlation matrix. The deflation technique is used for sequential estimation of the remaining PCs.

2.2 The RP algorithm

The RP [6,7] algorithm is derived from the Rayleigh quotient and uses the following rule for extraction of the first PC:

$$\mathbf{w}(n) = \frac{\mathbf{w}(n-1) + \mathbf{R}\mathbf{w}(n-1)}{1 + \mathbf{w}^H(n-1)\mathbf{R}\mathbf{w}(n-1)}.$$

In the implementation the terms $\mathbf{R}\mathbf{w}(n-1)$ and $\mathbf{w}^H(n-1)\mathbf{R}\mathbf{w}(n-1)$ are redefined as $\mathbf{P}(n-1)$ and $Q(n-1)$ and updated independently before calculating the new coefficient vector $\mathbf{w}(n)$. This is convenient since $Q(n)$ will be an estimate of the largest eigenvalue. The update rules for $\mathbf{P}(n-1)$ and $Q(n-1)$ may be rewritten so that the need for an explicit estimate of the correlation matrix $\mathbf{R}$ vanishes. No forgetting factor is incorporated in the estimation of $\mathbf{P}(n-1)$ and $Q(n-1)$, nevertheless this could be useful in a time-varying environment. The remaining PCs are estimated by using the same deflation technique as in the PAST algorithm.

2.3 The EE Algorithm

In this paper the EE algorithm is proposed for PCA. EE uses the following update rule

$$\mathbf{R}(n) = \beta\mathbf{R}(n-1) + \mathbf{x}(n)\mathbf{x}^H(n),$$

where β is a scalar forgetting factor. The correlation matrix is estimated by $(1-\beta)\mathbf{R}(n)$ and is used for calculating the eigenvectors and eigenvalues (i.e. the PCs) by calling the EIG-function in Matlab. A key difference between EE and the other algorithms is that EE uses direct estimation of the correlation matrix as the basis for calculating the PCs, while PAST and RP directly estimates the PCs without making a detour via the correlation matrix.

3. CHOICE OF COEFFICIENTS AND NORMALIZATION

The PAST and RP algorithms estimate the different PCs in an iterative way, where the largest PC is estimated first followed by a deflation step. After the deflation, the second largest PC is estimated followed by a new deflation step and so on. This procedure continues until all desired PCs are estimated, then the sample instant n is advanced by one and the procedure is repeated for the next data vector $\mathbf{x}(n)$. The deflation step is essentially a simple algebraic rule that removes the contribution of the latest estimated PC from the data vector. The most straight forward approach is to use the most current estimate of the eigenvector in this rule, but one could also choose to use the previous estimate. Which choice that is the best is not clear and is therefore investigated by Monte Carlo simulations in this paper. Both the PAST and RP algorithm are simulated in two configurations, using the old or the new estimate.

Eigenvectors are by definition normalized so that $\mathbf{w}^H\mathbf{w}$ equals 1. However, the formulation of the PAST and RP algorithms does not guarantee this condition. Depending on the application, the eigenvectors can be normalized after each iteration, every l iteration or when the coefficients have converged [7]. In this paper, it is proposed that both algorithms are adjusted so that the eigenvectors are normalized between iterations. Normalization is not required for EE, since the EIG-function assures that the eigenvectors are normalized at all time.

According to the previous discussion, there are now a total of four possible configurations of PAST and RP. First, the old or the new estimate may be used in the deflation rule. Second, normalization of the coefficient vectors is *on* or *off*. The different configurations are denoted as A, B, C or D: old weights and normalization off (A), old weights and normalization on (B), new weights and normalization off (C), new weights and normalization on (D). The configuration letter (A, B, C or D) is added to the algorithm name in order to differentiate between them. Examples are PAST-A, PAST-B, …, RP-A, RP-B, etc. EE does not have these configurations and hence it is simply denoted as EE. The performance of PAST and RP is dependent of the configuration and is investigated in the next section.

4. PERFORMANCE EVALUATION

In this section we evaluate the performance of the PAST-x, RP-x and EE algorithms by means of computer simulations that employ a Monte Carlo approach. Here, x denotes configuration A, B, C or D, respectively. In the simulations, a white Gaussian random noise signal with variance 1 is filtered by an first order AR filter $H(z) = 1/(1-0.9495z^{-1})$. The resulting filtered signal is then embedded by a 15-tap delay line resulting in input data vectors $\mathbf{x}(n)$ of size 15-by-1. The five largest true eigenvalues of the correlation matrix $\mathbf{R} = E\{\mathbf{x}(n)\mathbf{x}^H(n)\}$ can be found to be 119, 17.8, 5.58, 2.65 and 1.55, respectively. The condition number or the eigenvalue spread of $\mathbf{R}$ is 451. The choice of the first order AR filter was made so that it results in approximately the same eigenvalues used in [6].

The performance evaluation is set up so that each algorithm (PAST-x, RP-x and EE) is simulated for 10000 Monte Carlo runs consisting of 2000 iterations. After the last iteration in each Monte Carlo run, the five largest estimated eigenvalues and the corresponding eigenvectors are examined. An estimation error (the difference between the true eigenvalue and the estimated eigenvalue) is calculated and saved. Also, the absolute value of the directional cosines (DC) between the true and the estimated eigenvectors are calculated and saved. A value of +1 indicates perfect alignment of the estimated and true eigenvectors, while 0 indicate that the estimated and true eigenvector are orthogonal. The forgetting factor is equal to 1 for all simulations of PAST and EE.

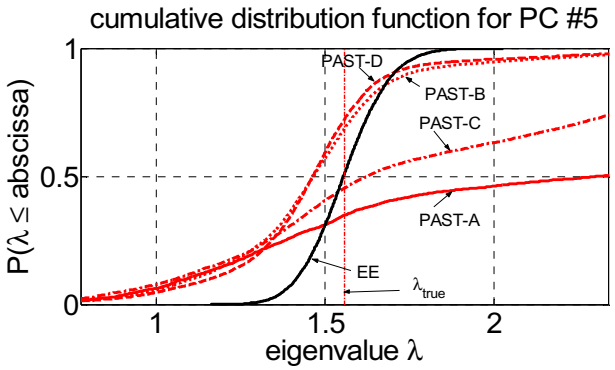

Figure 1: Cumulative distribution functions for the 5th principal component for PAST and EE. The vertical line corresponds to the true eigenvalue.

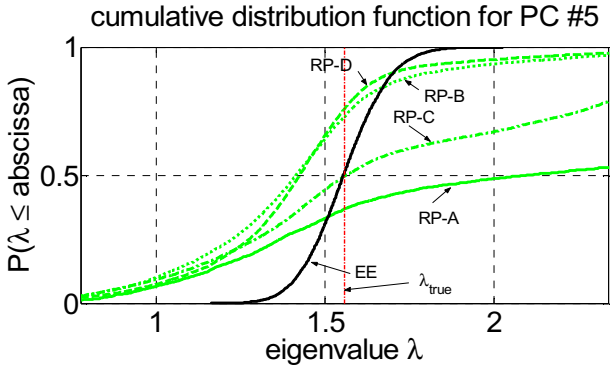

Figure 2: Cumulative distribution functions for the 5th principal component for RP and EE. The vertical line corresponds to the true eigenvalue.

The saved eigenvalue estimates and directional cosines may be evaluated by plotting cumulative distribution functions (CDF). The CDF for the eigenvalue estimates for the 5th principal component is plotted in figures 1 and 2. Figure 1 shows the results for PAST and Figure 2 for RP, respectively. The CDF for EE is also plotted in both figures as well as the true eigenvalue denoted by λ_{true}. The figures show that PAST and RP has similar performance, that EE has the best performance and that the B and D modes perform better than the A and C modes of the algorithms.

A compilation of all results are shown in tables 1 to 4. All tables show performance results for all nine algorithms for the five largest PCs. Figures in **bold** highlights the best result for each PC, not taking the reference algorithm EE into account. Table 1 shows the averaged estimation error for the eigenvalues, Table 2 the variance of the eigenvalue estimates, Table 3 the average DC for the eigenvectors and Table 4 the variance of the DCs. From these results, the following observations can be made:

- The EE algorithm has the best overall performance considering the performance measures listed in tables 1 to 4.
- PAST-B has the lowest eigenvalue estimation error.
- PAST-D has the lowest eigenvalue estimation variance.
- PAST-D has the lowest eigenvector DC error (closest to 1).
- PAST-D and RP-D have the lowest eigenvector DC variance.
- The performance of PAST-B and RP-B is similar.
- The performance of PAST-D and RP-D is similar.
- The performance of PAST-B and PAST-D is similar.
- The performance of PAST-A, PAST-C, RP-A and RP-C is worse than for the B and D versions.

From these observations, it can be concluded that the B and D configurations of PAST and RP have similarly good performance. The A and C configurations have the worst performance. Thus, it is concluded that the choice of the old or the new weights in the deflation step is not crucial. On the other hand, normalization is highly recommended. Finally, the EE algorithm has the most excellent overall performance and may therefore be used as a benchmark for PCA algorithms.

5. CONCLUSIONS

The PAST and RP algorithms are both robust and efficient algorithms for PCA. This paper has investigated the performance of the algorithms and the effect when using normalization of the eigenvectors between iterations. Simulation results clearly demonstrate that the algorithms operate more reliably when normalization is adopted. It is also shown that the EE algorithm provides the best performance and can be used as a benchmark for comparing PCA algorithms.

ACKNOWLEDGEMENT

The authors thank Dr. Jan-Mark de Haan for reviewing the paper and giving valuable feedback.

REFERENCES

[1] Yang, B.: ”Projection Approximation Subspace Tracking”, IEEE Transactions on Signal Processing, 1995, vol. 43, (1), pp. 95-107.

[2] Xu, L.: ”Least Mean Square Error Reconstruction Principle for Self-Organizing Neural-Nets”, Neural Networks, 1993, vol. 6, pp. 627-648.

[3] Pajunen, P. and Karhunen, J.: ”Least-Squares Methods for Blind Source Separation Based on Nonlinear PCA”, International Journal of Neural Systems, 1997, vol. 8, pp. 601-612.

[4] Karhunen, J. and Joutsensalo, J.: ”Representation and Separation of Signals using Nonlinear PCA Type Learning”, Neural Networks, 1994, vol. 7, issue 1, pp. 113-127.

[5] Haykin, S.: ”Adaptive Filter Theory”, 2002, Prentice Hall, 4th edition.

[6] Rao, Y.N. and Principe, J.C.: ”Robust On-line Principal Component Analysis Based on a Fixed-point Approach”, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002, vol. 1, pp. I-981-I-984.
[7] Rao, Y.N. and Principe, J.C.: "A Fast, On-line Algorithm for PCA and its Convergence Characteristics", IEEE Signal Processing Society Workshop, 2000, vol. 1, pp. 299-307.

PC	PAST				RP				EE
	A	B	C	D	A	B	C	D	
1	2.07	0.82	2.07	0.95	1.07	**0.59**	1.05	1.06	0.35
2	-10.88	**0.18**	1.51	0.26	-10.13	0.20	0.99	0.28	-0.02
3	-17.55	**0.15**	-0.33	0.19	-18.44	0.20	0.52	0.22	0.00
4	-18.73	**0.12**	-0.47	0.13	-21.00	0.19	-0.14	0.16	0.00
5	-20.33	**0.09**	-0.41	**0.09**	-18.27	0.13	-0.17	0.13	0.00

Table 1: Average estimation errors for the eigenvalues

PC	PAST				RP				EE
	A	B	C	D	A	B	C	D	
1	430.02	401.92	430.97	**400.04**	412.01	424.12	410.60	415.38	414.45
2	1111.2	2.92	24.50	**2.78**	1009.2	2.95	6.29	2.81	2.75
3	1505.6	0.35	11.13	**0.31**	1632.3	0.37	3.50	0.32	0.19
4	1533.5	0.18	3.75	**0.14**	1853.4	0.22	1.36	0.17	0.04
5	1666.3	0.10	1.11	**0.08**	1548.7	0.13	0.38	0.10	0.01

Table 2: Variance of the eigenvalue estimates

PC	PAST				RP				EE
	A	B	C	D	A	B	C	D	
1	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	1.00
2	0.74	**1.00**	0.93	**1.00**	0.80	**1.00**	0.99	**1.00**	1.00
3	0.46	**0.99**	0.75	**0.99**	0.49	**0.99**	0.86	**0.99**	1.00
4	0.31	0.94	0.57	**0.95**	0.36	0.92	0.66	0.94	1.00
5	0.23	0.84	0.45	**0.88**	0.25	0.79	0.50	0.85	1.00

Table 3: Average direct cosines

PC	PAST				RP				EE
	A	B	C	D	A	B	C	D	
1	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.00
2	0.11	**0.00**	0.05	**0.00**	0.09	**0.00**	0.01	**0.00**	0.00
3	0.13	**0.00**	0.15	**0.00**	0.13	**0.00**	0.08	**0.00**	0.00
4	0.09	**0.02**	0.18	**0.02**	0.09	0.03	0.15	**0.02**	0.00
5	**0.06**	0.07	0.16	**0.06**	**0.06**	0.07	0.16	**0.06**	0.00

Table 4: Variance of direct cosines