Singular and Principal Subspace of Signal Information System by BROM Algorithm

Władysław Skarbek

Warsaw University of Technology, Faculty of Electronics and Information Technology,
00-665 Warszawa, Nowowiejska 15/19, Poland
W.Skarbek@ire.pw.edu.pl

Abstract. A novel algorithm for finding algebraic base of singular subspace for signal information system is presented. It is based on Best Rank One Matrix (BROM) approximation for matrix representation of information system and on its subsequent matrix residua. From algebraic point of view BROM is a kind of power method for singular value problem. By attribute centering it can be used to determine principal subspace of signal information system and for this goal it is more accurate and faster than Oja's neural algorithm for PCA while preserving its adaptivity to signal change in time and space. The concept is illustrated by an exemplary application from image processing area: adaptive computing of image energy singular trajectory which could be used for image replicas detection.

Keywords: Principal Component Analysis, Singular Value Decomposition, JPEG compression, image replica detection, image energy singular trajectory.

1 Introduction

Signal information system is a special kind of information system in Pawlak's sense [1] in which objects are certain signal (multimedia) objects such as images, audio tracks, video sequences while attributes are determined by certain discrete elements drawn from spatial, temporal, or transform domain of signal objects.

For instance the Discrete Cosine Transform frequency channel represents a DCT coefficient which specifies a share of such frequency in the whole signal object.

Having n signal objects and m DCT frequency channels we get an information system which can be represented by a matrix $A \in \mathbb{R}^{m \times n}$. In case of Discrete Fourier Transform the matrix has complex elements and $A \in \mathbb{C}^{m \times n}$.

The columns of $A \in \mathbb{R}^{m \times n}$ define n attributes $A = [a_1, \ldots, a_n]$ and they can be considered as elements of m dimensional vector space: $a_i \in \mathbb{R}^m$, $i = 1, \ldots, n$.

In order to use algebraic properties of the vectorial space the attributes should have common physical units or they should be made unit-less, for instance by an affine transform, such as attribute centering and scaling.

Having attributes in the vectorial space we can define their dependence by the concept of linear combinations and by the related concepts of linear independence and linear subspace.

J.T. Yao et al. (Eds.): RSKT 2007, LNAI 4481, pp. 157–165, 2007.

In such approach we start from the subspace span(A) which includes all finite linear combinations of columns of A, i.e. attributes $a_1, \ldots, a_n$. If the dimension of span(A) equals to r, i.e. if rank(A) $= r$ then we can find a nested sequence of $r+1$ linear subspaces of increasing dimensionality starting from the null subspace $\mathcal{S}_0 := \{0_m\}$ and ending at $\mathcal{S}_r :=$span(A) :

$$\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_{r-1} \subset \mathcal{S}_r \ .$$

In the infinite number of nested subspace sequences for the given matrix A, there is a specific class of singular subspaces defined for A by the condition of minimum projection error. Namely, let $P_{\mathcal{S}}a$ be the orthogonal projection of a onto the subspace $\mathcal{S}$. Then the *singular subspace* $\mathcal{S}_q$ of dimension $q <$rank(A) minimizes the following squared projection error in norm l_2 :

$$\mathcal{S}_q := \arg \min_{\dim(\mathcal{S})=q} \sum_{i=1}^{n} \|a_i - P_{\mathcal{S}} a_i\|_2^2 \ .$$

The unit vector u spanning $\mathcal{S}_1$ is called the *singular direction.* We say that the attributes $a_1, \ldots, a_n$ are centered if their mean is zero vector, i.e. $\sum_{i=1}^{n} a_i = 0_m$. In case of centered attributes the singular subspace is called the *principal subspace* and the singular direction is called the *principal direction.*

In practice the singular subspace of matrix A is found from Singular Value Decomposition (SVD) of matrix A to orthogonal matrices U, V and diagonal matrix Σ :

$$\begin{aligned} &A = U\Sigma V^t, \ U \in \mathbb{R}^{m\times r}, \ U^tU = I_{r\times r} \\ &\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r), \quad V \in \mathbb{R}^{n\times r}, \ V^tV = I_{r\times r} \ . \end{aligned}$$

Namely, the first q columns of $U = [u_1, \ldots, u_r]$ span the singular subspace $\mathcal{S}_q =$ span$(u_1, \ldots, u_r)$.

Traditionally principal subspaces are obtained from Eigenvector Decomposition (EVD) of the outer product of centered matrix A :

$$AA^t = U\Lambda U^t$$

This procedure is known as Principal Component Analysis (PCA) – one of the most famous transformations in signal theory [2]. The traditional approach, though very efficient, is not adaptive to change of matrix A. In case of PCA there is also well known Oja neural scheme [3] which stochastically approximates the principal direction. However, it can be used to centered data only and therefore is not applicable to the general case of singular direction. In this paper another adaptive scheme is presented. It is based on analysis of rank one matrix approximations of the information system.

2 BROM Algorithm for Singular Subspace

Let us consider a signal information system with m objects and n real valued attributes. Then it is represented by a matrix $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m\times n}$, $a_i \in \mathbb{R}^m$.

If the linear subspace spanned by attributes a_i has the dimension r, i.e. if $\mathrm{rank}(A) = r$, then for any $q < r$ we consider the following problem: *Find a matrix* $X \in \mathbb{R}^{m\times n}$ *of rank* q *which is the best approximation of* A *in Frobenius norm, i.e. it minimizes* $\|A - X\|_F$. We call this problem as *best rank* q *matrix* and in particular for $q = 1$ we have BROM problem, i.e. *best rank one matrix* problem.

It appears that it is enough to know an algorithm for BROM problem, in order to get incrementally the solution X_q for any $q < r$:

$$X_0 = 0_{m\times n};\ \text{for } q = 1, \dots, \mathrm{rank}(A) - 1 :\ X_q := X_{q-1} + \mathrm{BROM}(A - X_{q-1}); \quad (1)$$

On the other hand the matrix X of rank one has a shorter nonlinear parametrization with $m + n$ variables. Namely the following property is true: *The matrix* $X \in \mathbb{R}^{m\times n}$ *is of rank one if and only if there exist vectors* $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$, $u \neq 0, v \neq 0$, *such that* $X = uv^t$. Therefore we can state the following optimization goal function e of two vectorial parameters u, v :

$$e(u, v) := \|A - uv^t\|_F^2,\ u \in \mathbb{R}^m,\ v \in \mathbb{R}^n \quad (2)$$

It is easy to find a necessary and sufficient condition for the stationary points of e, i.e. the zero gradient points of $e(u, v)$:

$$v = \frac{A^t u}{\|u\|^2},\ u = \frac{Av}{\|v\|^2} \quad (3)$$

Moreover, at fixed u (v) the actual minimum of e can be explicitly found:

- At fixed $u \in \mathbb{R}^m, u \neq 0$ the optimal $v_{opt} \in \mathbb{R}^n$ minimizing $e(u, v)$ has the form:
$$\begin{aligned} v_{opt} &= \tfrac{A^t u}{\|u\|^2},\ e(u, v_{opt}) \leq e(u, v),\ \forall v \in \mathbb{R}^n \\ e(u, v_{opt}) &= \|A\|_F^2 - \|u\|_2^2 \|v_{opt}\|_2^2 \end{aligned} \quad (4)$$
- At fixed $v \in \mathbb{R}^n, v \neq 0$ the optimal $u_{opt} \in \mathbb{R}^m$ minimizing $e(u, v)$ is of the form:
$$\begin{aligned} u_{opt} &= \tfrac{Av}{\|v\|^2},\ e(u_{opt}, v) \leq e(u, v),\ \forall u \in \mathbb{R}^m \\ e(u_{opt}, v) &= \|A\|_F^2 - \|u_{opt}\|_2^2 \|v\|_2^2 \end{aligned} \quad (5)$$

The algorithm BROM looks for the best rank one matrix uv^t of the matrix A by the following locally optimal steps for $i = 0, 1, \cdots$:

1. for u_i determine the optimal v_{i+1};
2. for v_{i+1} determine the optimal u_{i+1}.

Using the explicit formulas for u_{opt} and v_{opt} we get the following iterative scheme:

$$\begin{aligned} u_0 &:= \text{ nonzero column of } A \\ v_{i+1} &:= \tfrac{A^t u_i}{\|u_i\|^2},\ u_{i+1} := \tfrac{Av_{i+1}}{\|v_{i+1}\|^2},\ i = 0, 1, 2, \dots \end{aligned} \quad (6)$$

For practical use the following form of BROM has been elaborated which returns the singular direction in vector u.

algorithm $[u, v] := brom(A)$
$u :=$ the first nonzero column of A
if $u = 0$ **then** *return* **endif**
do
$v := \frac{A^t u}{\|u\|^2}; \ u := \frac{Av}{\|v\|^2}$
until $(\|u\| \cdot \|v\|$ stabilizes$)$
$v := v * \|u\|; \ u := u/\|u\|$
endalgorithm

Since at the exit of BROM we have $v = A^t u$, the coordinate of column a_i w.r.t. to the vector u is v_i, $i = 1, \ldots, n$. It is interesting that when a symmetric matrix A is input of the BROM, the algorithm produces as u the eigenvector corresponding to the eigenvalue λ_{max} of maximum absolute value and $\lambda_{max} = u^t A u$. This follows from the observation that modulo a scaling factor, the BROM algorithm performs iterations of *the power method* for the matrix A^2 which has the same eigenvectors as A, but for squared eigenvalues. Since the power method computes the maximal eigenvalue for A^2 then for A it corresponds to the maximum absolute eigenvalue.

3 Outline of BROM's Convergence Analysis

The strict proof of convergence for BROM has been recently developed by the author, but the limit of pages for this paper allows only for an outline of BROM's convergence analysis.

We analyze the sequences defined by (6). The first observation concerns the behavior of norms for the sequences. Namely, the norms of vectorial sequences u_i and v_i satisfy the following inequalities for $i = 1, 2, \cdots$:

$$\begin{aligned} &\|u_i\|\|v_i\| \leq \|A\|_F \\ &\|u_i\|\|v_i\| \leq \|u_i\|\|v_{i+1}\| \leq \|u_{i+1}\|\|v_{i+1}\| \\ &\|u_i\| \leq \|u_{i+1}\|, \ 1 \leq \|v_i\| \leq \|v_{i+1}\| \end{aligned}$$

Hence, the norms of the sequences are bounded and monotonic. Thus the sequences of norms $\|u_i\|, \|v_i\|$ are convergent.

Let λ be an eigenvalue of the matrix B. Then we denote by $W(B, \lambda)$ the subspace of all eigenvectors defined by eigenvalue λ :

$$W(B, \lambda) := \{u : \ Bu = \lambda u\}$$

The remaining convergence analysis can be summarized in six properties which are stated in the following theorem.

Theorem 1 (on convergence of BROM).

1. *The vectorial sequence u_i is convergent in l_2 to an eigenvector u_* of the matrix AA^t corresponding to the largest eigenvalue λ for which the initial vector u_0 is not perpendicular to the eigenvector subspace $W(AA^t, \lambda)$.*

2. The vectorial sequence v_i *is convergent in* l_2 *to the vector* $v_* = A^t u_*/\|u_*\|^2$.
3. The vector $u_*/\|u_*\|$ *is the singular direction with the singular value* $\sqrt{\lambda}$ *and the singular coordinates* $\sqrt{\lambda} v_*/\|v_*\|$.
4. The matrix sequence $u_i v_i^t$ *is convergent in Frobenius norm to the matrix* $u_* v_*^t$ *which is the stationary point of the objective function* $e(u, v)$.
5. If u_0 *is not perpendicular to the eigenvector subspace* $W(AA^t, \lambda_{max})$ *for the maximal eigenvalue of the matrix* AA^t *then the matrix sequence* $u_i v_i^t$ *is convergent w.r.t. Frobenius norm to the matrix* $u_* v_*^t$ *which is the global minimum of the objective function* $e(u, v)$ *and* $e(u_*, v_*) = \|A\|_F^2 - \lambda_{max}$.
6. The stop condition for BROM algorithm selected in the form:

$$\|u_{i+1}\|\|v_{i+1}\| - \|u_i\|\|v_i\| < \epsilon$$

implies the stabilization of the objective function: $e(u_i, v_i) - e(u_{i+1}, v_{i+1}) < \epsilon$.

The above theorem explains why in the stop condition we can replace the original requirement for $u_i v_i^t$ stabilization by less costly condition of stabilization for the norm product $\|u_i\|\|v_i\|$. Namely, the convergence of the matrix sequence $u_i v_i^t$ enables observing of this convergence indirectly in the range of the error function $e(u_i, v_i)$. But from the last property of the theorem we have seen that the convergence of $e(u_i, v_i)$ can be detected from the convergence of the sequence $\|u_i\|\|v_i\|$.

4 Application: Image Energy Singular Trajectory

In many applications images are decomposed into small size blocks in order to make local analysis which is more efficient or more problem relevant. The blocks of the decomposition can be disjoint or overlapping. For instance in JPEG compression [4] the blocks are of size 8×8 and they are not overlapping. In some applications the order of blocks is irrelevant while in some their relative locations in the sequence is important.

Having a fixed ordering of blocks, for instance according the raster scan or along the Hilbert curve, we can introduce a concept of signal energy trajectory for the given image. Namely, each image block is characterized by the signal energy measured by the sum of squared pixel intensities.

The signal energy as a block feature is not invariant to most image processing operations. However, if we consider the fractional distribution of the energy in singular channels defined by singular directions of image blocks, the situation is much better and such features can be used for instance to image replica detection [5] even if replica have been *invisibly processed* to cheat web robots.

Let us define the *image energy singular trajectory* of rank r more formally. Let $f_1, \ldots, f_L$ be the sequence of pixel blocks drawn from the image f. It means that $f_i := f|_{D_i}$ for a rectangular sub-domain D_i of the image domain D. Performing the singular decomposition of the matrix f_i, we consider only r dominant singular values $\sigma_i(1), \ldots, \sigma_i(r)$.

It is well known that the signal energy of block f_i is decomposed into the sum of all squared singular values of f_i :

$$\|f_i\|_F^2 = \sum_k \sigma_i^2(k), \ \sum_k \frac{\sigma_i^2(k)}{\|f_i\|_F^2} = 1$$

The image energy singular trajectory of rank r is defined as the sequence of points in r dimensional unit cube $[0,1]^r$:

$$\left(\frac{\sigma_i^2(1)}{\|f_i\|_F^2}, \dots \frac{\sigma_i^2(r)}{\|f_i\|_F^2} \right), \ i = 1, \dots, L$$

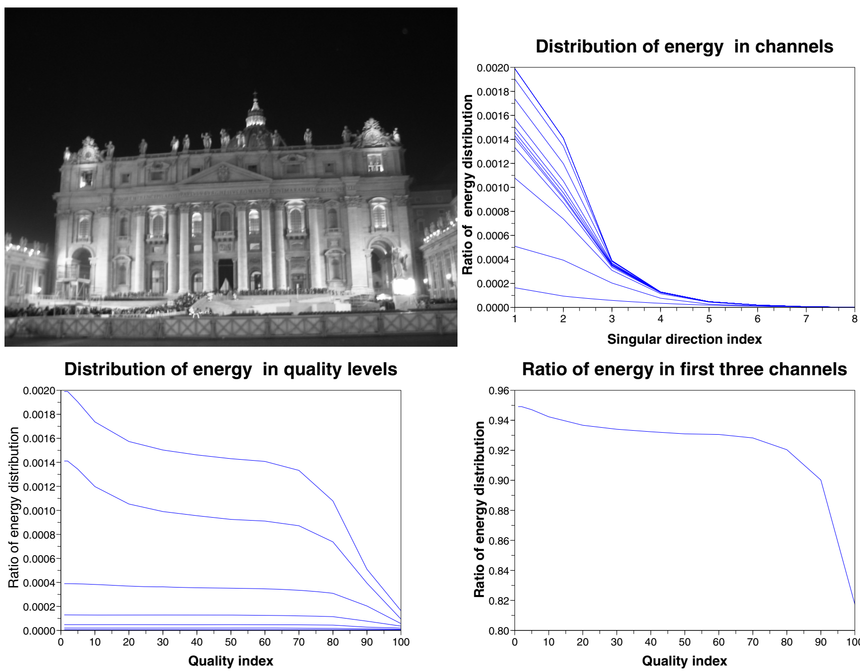

Fig. 1. Dependence of energy distribution in singular channels for various JPEG compression quality. For bottom graphs lower curves correspond to higher index of energy channel while for the right upper graph to the higher quality index.

The rationale for this novel concept is getting a fine characterization of signal energy distribution with its spatial coherency to be represented by trajectory concept. Since trajectories can be normalized by its re-sampling to a standard discrete interval, changes of image resolutions, cutting of windows, and local affine image transformations could be detected in trajectory segments by point proximity analysis for trajectories in time-energy space.

If small rank trajectories are enough for a particular application then we expect that BROM algorithm can be recommended in place of standard SVD algorithm. In the remaining part of this section few such practical cases are analyzed.

JPEG quality measure. In image compression the quality is usually measured by error image global analysis. The error image is the difference between decoder's output image and encoder's input image computed pixel-wise.

The most popular image fidelity measures are based on mean squared error (MSE) which is a scaled version of squared Frobenius norm for the error image. More subtle fidelity measures have vectorial character such as SVD based measures [6].

We analyze the distribution of compression error energy for an image of architectural scenes. Let JPEG quality index be from the set:

$$Q = \{1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100\} .$$

Then for the image of Fig. 1 the average distribution of energy in error images for those quality indexes w.r.t. all eight singular directions is presented. We see that most of energy is included in the first three singular directions (so called the energy channels) – from 95% for low quality images down to 80% for high quality image. We observe that the proposed measure is uncorrelated with human eye sensitivity: the highest drop of energy corresponds to best visual quality interval $[70, 100]$. This is very desirable property for image replica detection.

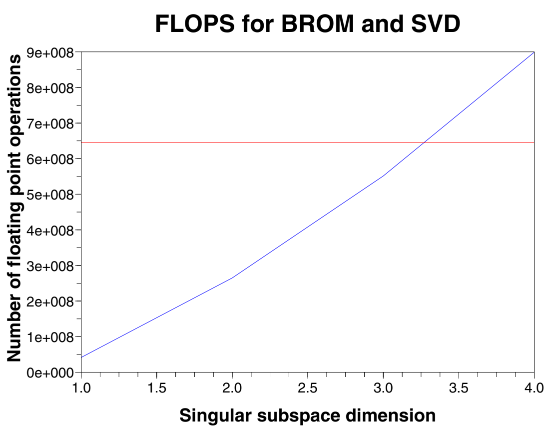

Fig. 2. Number of floating point operation at computing image energy singular trajectory by BROM and SVD (horizontal line) as function of singular subspace dimension

Trajectory complexity for BROM and SVD. It is interesting to compare the computational efficiency of using BROM to find the image energy singular trajectory w.r.t. the classical approach using Singular Value Decomposition for image blocks. Since SVD returns all singular values of image blocks it seems that

BROM should be much faster. The expectation is confirmed up to rank $r = 3$ (cf. Fig. 2) at the arithmetic precision 10^{-10} for singular values.

For this kind of applications when *signal information systems* are relatively small the adaptivity of BROM results in less than 5% reducing of complexity.

At comparison, the SVD complexity has been evaluated on the basis of the formulas given in [7]. Other experiments show that BROM algorithm is more accurate and faster than Oja algorithm [3] when applied for centered data to find the principal direction.

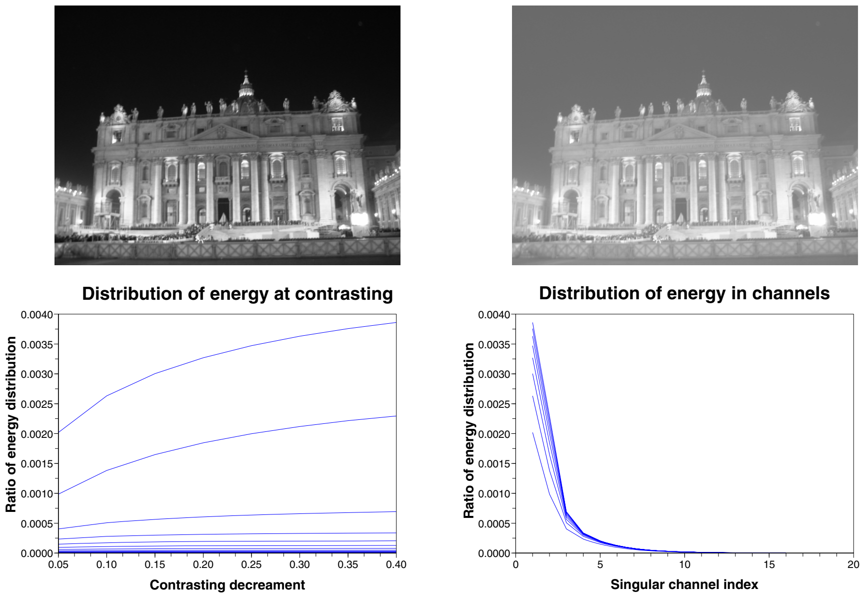

Fig. 3. Dependence of energy distribution in singular channels for various image contrasting and brightening operations

Energy distribution at image contrasting and brightening. In this experiment we change the image contrast and its brightness using scaling parameter $s \in (0, 1)$:

$$f_s := (1 - s) \cdot f + s \cdot f_{max}$$

where f_{max} is the maximum value of image f. This kind of image processing operations in the same time reduces contrast by factor $(1 - s)$ and in a sense compensates this by increase of brightness to preserve the maximum value of image intensity. For small s the processing effect is invisible, but the change in the signal energy is significant.

Figure 3 shows two processed versions of image from Fig. 1. The left image is processed with $s = 0.05$ and the right one with $s = 0.4$.

Experiments show that the error in image energy distribution slightly depends on s and for singular channels with indices higher than three (lower curves on the left graph of Fig. 3) is marginal (less than 0.04%). The maximum change is observed for the dominant singular subspace (top curve) but even for maximal $s = 0.4$ the change of energy share for this channel is less than 0.4%. It means that the energy distribution is a good invariant for image contrasting and brightening operations and it in this context is useful for image replica detection.

The right graph on Fig. 3 confirms *the rule of three* observed in context of JPEG image quality: *the most of change in image energy distribution is observed in the first three singular channels.* The curves in this graph are indexed by the parameter s with $s = 0.05$ for the bottom curve and $s = 0.4$ for the top one.

5 Conclusions

BROM algorithm is a practical alternative for SVD in finding algebraic base of singular subspace for signal information systems.

From algebraic point of view BROM is a kind of power method applied for singular value problem and it can be specialized to be a novel power method for eigenvalue problem.

By attribute centering it can be used to determine principal subspace of signal information system and for this goal it is more accurate and faster than Oja's neural algorithm for PCA while preserving its adaptivity to signal change in time and space.

In exemplary application computing image energy singular trajectory it appears faster than SVD for rank less than four. The trajectory approach is the useful tool applicable to image replica detection.

Acknowledgment. The work presented was developed within VISNET 2, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 Programme.

References

1. Pawlak, Z.: Information Systems Theoretical Foundations. Information Systems **3** (1981) 205–218
2. Jolliffe, I.T.: Principal Component Analysis. Springer-Verlag (2002)
3. Skarbek, W., Pietrowcew, A., Sikora, R.: The modified Oja-RLS algorithm, stochastic convergence analysis and application for image compression. Fundamenta Informaticae **36** (1998) 345–365
4. Pennebaker, W.B., Mitchell, J.L.: JPEG: Still Image Data Compression Standard. Springer (2006)
5. Maret, Y., DuFaux, F., Ebrahimi, T.: Adaptive image replica detection based on support vector classifiers. Signal Processing: Image Communication **21** (2006) 688–703
6. Shnayderman, A., Gusev, A., Eskicioglu, A.M.: An SVD-Based Grayscale Image Quality Measure for Local and Global Assessment. IEEE Trans. Image Process **15** (2006) 422–429
7. Golub, G., Loan, C.: Matrix Computations. The Johns Hopkins University Press (1989)