A Fast Leading Eigenvector Approximation for Segmentation and Grouping

Antonio Robles-Kelly[1]*, Sudeep Sarkar[2] and Edwin R. Hancock[3]
[1,3] Computer Science Department, University of York, York Y01 5DD, UK
[2] Computer Science and Engineering, University of South Florida, Tampa, FL, USA
[1]arobkell@cs.york.ac.uk,[2]sarkar@csee.usf.edu,[3]erh@cs.york.ac.uk

Abstract

We present a fast non-iterative method for approximating the leading eigenvector so as to render graph-spectral based grouping algorithms more efficient. The approximation is based on a linear perturbation analysis and applies to matrices that are non-sparse, non-negative and symmetric. For an $N \times N$ matrix, the approximation can be implemented with complexity as low as $\mathcal{O}(4\mathrm{N}^2)$.We provide a performance analysis and demonstrate the usefulness of our method on image segmentation problems.

1 Introduction

Recently, there has been considerable interest in the use of graph-spectral [1] methods for computer vision for segmentation and grouping [2, 3, 4]. These methods all share the feature that they use the eigenvectors of a weighted adjacency matrix to locate salient groupings of objects. Although elegant, one of the criticisms that can be leveled at these methods is that they are computationally demanding because they rely on the numerical determination of eigenvectors of large matrices. The problem of computing the eigenvalues and eigenvectors of a matrix is one of classical linear algebra which arises in many practical problems in science and engineering. However, it is frequently the determination of the leading eigenvector which turns out to be of pivotal importance. The reason for this is that it is intimately related to the Raleigh Quotient [3, 5].

When only the first few eigenvalues and eigenvectors of a matrix are required, then the most commonly used methods are those based on Krylov information [6]. Among these, the most well known are the power method and, for symmetric matrices, the Lanczos method. These methods are iterative, and hence need to be initialised with a starting vector. Moreover, it can be proven that the rate of convergence of the Lanczos method changes in accordance with the error criterion used [7]. Hence, the efficiency of the method is dependant on how distant the extreme eigenpair is from the others.

*Supported by The William Gibbs Plessey Award and CONACYT grant No. 146475/151752.

Other approximation approaches include the Nyström method [8] and matrix perturbation theory [9]. The Nyström method has been applied to a variety of problems, including grouping [10] and solving integral equations with singular kernels [11]. The perturbation theory of eigenvalues and eigenvectors has been researched thoroughly, and the bounds of a single eigenvalue have been set extremely accurately using Gerschgorin's Theorem.

2 Approximation via Matrix Perturbation

Perturbation methods attempt to solve a complex problem approximating it by a simpler problem that is reasonably well known. We will use the perturbation approach to approximate the first eigenvector of a large, real, non-negative symmetric matrix A to a linear combination of perturbation matrices whose spectral properties are well known. In segmentation and grouping this prerequisites are not prohibitive since the matrix A (i.e. the adjacency matrix) is mostly used to capture the pairwise similarity between two image tokens. Hence, it is symmetric in nature and its elements are real and non-negative. It is important to note that the similarity between tokens is not guaranteed to be symmetric in all cases. We refer the interested reader to the work of Lance Williams [12] for counter examples.

For computational purposes, we will constrain the elements of the matrix A to be between 0 and 1. This is a scaling step that will not affect the generality of the method.

We commence expressing the matrix A as a linear combination of perturbation matrices $\hat{A}_i$ as follows

$$A = \frac{1}{N}\sum_{i=1}^{N}\hat{A}_i + E \tag{1}$$

where matrix E is an error matrix.

For our analysis, we require the pertubation matrices $\hat{A}_i$ to be real, non-negative symmetric matrices proportional to $\Phi_i{\Phi_i}^T$, where $\Phi_i = [\phi_{i,1}, \ldots, \phi_{i,N}]^T$ is a real, non-negative vector of order N, normalized to satisfy the condition

$$\sum_{k=1}^{N}{\phi_{i,k}}^2 = 1 \tag{2}$$

Providing that these requirements are satisfied, it can be shown that

$$\hat{A}_i = \eta_i \Phi_i {\Phi_i}^T \tag{3}$$

where $\hat{\eta}_i$ is the pertubation parameter for the matrix $\hat{A}_i$. It is relatively easy to prove from the theory of non-negative symmetric matrices [13] that the vector Φ_i is equal to the first eigenvector $\hat{x}$ of A if $A = \hat{A}_1$. Moreover, it can be shown that, in such cases, $\hat{x}$ can be obtained in a closed form at a cost of $\mathcal{O}(2\mathrm{N})$.

In order to obtain the vector Φ_i we re-write equation 3 using the expansion

$$\hat{A}_i = \begin{pmatrix} {(a_{1,1})}_i & \dots & {(a_{1,N})}_i \\ \vdots & \ddots & \vdots \\ {(a_{N,1})}_i & \dots & {(a_{N,N})}_i \end{pmatrix} = \eta_i \begin{pmatrix} {(\phi_{i,1})}^2 & \dots & \phi_{i,1}\phi_{i,N} \\ \vdots & \ddots & \vdots \\ \phi_{i,N}\phi_{i,1} & \dots & {(\phi_{i,N})}^2 \end{pmatrix} \tag{4}$$

After some algebra we find that

$$\eta_i \phi_{i,j} = \sqrt{\sum_{k=1}^{N} {(a_{j,k})_i}^2} \tag{5}$$

So far, we have limited ourselves to the properties of $\hat{A}_i$. We now turn our attention to the task of approximating the leading eigenvalue and eigenvector (i.e. the leading eigenpair) of the matrix A using the vectors Φ_i and the perturbation parameters η_i. Substituting equation 3 into equation 1 yeilds

$$A = \frac{1}{N} \sum_{i=1}^{N} \eta_i \Phi_i {\Phi_i}^T + E \tag{6}$$

The above expression is very similar to the spectral decomposition of the matrix A. The vectors Φ_i are closely related to the first eigenvector of A. Even though the leading eigenvector of matrix A and the vector Φ_i are not equal, the contribution of the former to the latter is proportional to η_i.

When the off-diagonal elements A are all non-zero then E will be a diagonal matrix. As a result, E will be approximately diagonal in structure when the matrix A is dense. For large dense matrices, we can treat E as a diagonal matrix with little loss of accuracy.

Using the spectral expansion of A together with the properties of symmetric non-negative matrices [13, 14], when A is not diagonally dominant then it follows that

$$A \approx \hat{\lambda} \hat{x} \hat{x}^T \tag{7}$$

where $\hat{\lambda}$ and $\hat{x} = [x_1, \dots, x_N]^T$ are the leading eigenvalue and eigenvector of A.

In practice, the requirement that A is not diagonally dominant is not restrictive. The reason being that it can always be made diagonally dominant by adding identity matrices without changing its eigenvectors. The eigenvalues will, of course, be different. Furthermore, it can be proven that

$$[A + I]\hat{x} = (\hat{\lambda} + 1)\hat{x} \tag{8}$$

This property also opens-up the possibility of processing other canonical forms, such as the normalized Laplacian.

From equations 4 and 5, and using the approximation in equation 7, we get

$$\phi_{i,j} = \frac{a_{i,j}}{\sqrt{\sum_{k=1}^{N} {a_{k,j}}^2}}, i \neq j \tag{9}$$

where $a_{i,j}$ is the element indexed i, j of matrix A.

If there is no dominant perturbation parameter η_i and if the matrix A is large, then from equations 7 and 9 we have

$$x_i \approx \frac{1}{N-1} \sum_{j=1; j \neq i}^{N} \phi_{i,j} \tag{10}$$

Expanding the right-hand side of equation 10 including the elements $e_{i,j}$ of matrix E and assuming E to be diagonal, we find that

$$\frac{1}{N-1} \sum_{j=1; j \neq i}^{N} \phi_{i,j} = \frac{1}{N-1} \left\{ \frac{a_{i,1}}{\sqrt{\sum_{k=1}^{N} {a_{k,1}}^2 + {(a_{1,1} - e_{1,1})}^2}} + \dots + \frac{1}{N-1} \frac{a_{i,N}}{\sqrt{\sum_{k=1}^{N} {a_{k,N}}^2 + {(a_{N,N} - e_{N,N})}^2}} \right\} \tag{11}$$

Hence, the accuracy of the approximation will be closely related to the diagonal elements of A. More precisely, it is governed by terms of the form ${(a_{i,i} - e_{i,i})}^2$.

A less computationally expensive estimate of the approximation accuracy is given by the norm condition of the eigenvector. Hence, we define the accuracy variable σ to be

$$\sigma = \mid 1 - \sum_{i=1}^{N} {x_i}^2 \mid \tag{12}$$

where x_i is the ith element of the approximated first eigenvector.

3 Pseudocode

Having presented the theoretical prerequisites, we can develop a practical algorithm for computing an approximation to the first eigenvector of the matrix A. The pseudocode is as follows

For $i = 1 \dots N$
  $\hat{x}_i = 0$;
  For $j = 1 \dots N$
    If $i \neq$ **j** $\hat{x}_i = {a_{i,j}}^2 + \hat{x}_i$;
For $i = 1 \dots N$
  $\hat{x}_i = \sqrt{\hat{x}_i}$;
For $i = 1 \dots N$
  For $j = 1 \dots N$
    If $j \neq i$ $a_{i,j} = \frac{a_{i,j}}{\hat{x}_i}$;
  If $j = i$ $a_{i,j} = \hat{x}_i$;
For $i = 1 \dots N$
  $\sigma = 0$;

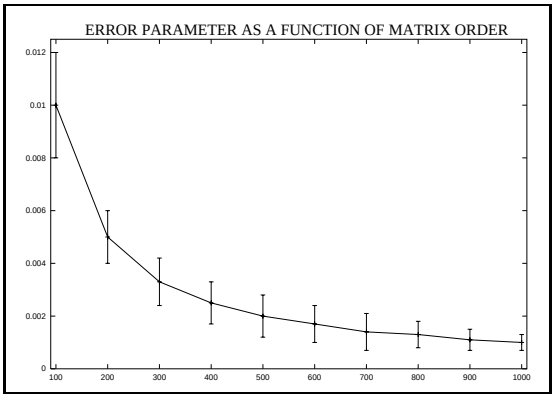

Figure 1. Effect of N on σ

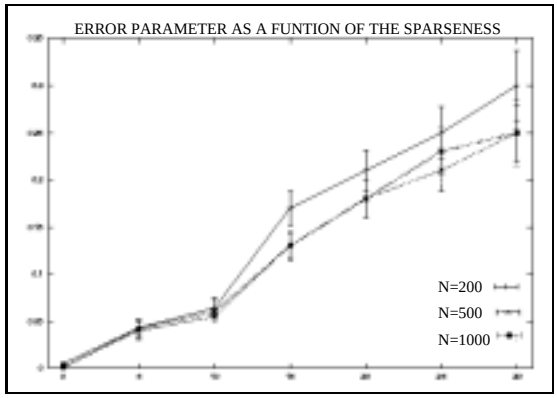

Figure 2. Effect of the sparseness on σ

For $j = 1 \ldots N$
  If $i \neq j$ $\sigma = \sigma + a_{i,j}$**;**
$x_i = \sigma$**;**
For $i = 1 \ldots N$
$\hat{x}_i = \frac{\hat{x}_i}{N-1}$**;**
$\sigma = 0$**;**
For $i = 1 \ldots N$
$\sigma = \sigma + {x_i}^2$**;**

There are two important practical considerations underpinning this algorithm. First, in order to reuse memory, the matrix A is overwritten at output and σ has been used more than once as a variable. Second, the sequence of operations has been slightly altered in order to reduce the number of floating point operations. The complexity of the algorithm is $4N^2$. The memory requirement, assuming matrix A has already been allocated, is $N + 3$ memory locations.

4 Error and Performance Analysis

Our aim in this section is to explore the effects of varying the size N and sparseness of the matrix A on our approximate eigenvector calculation.

For our sensitivity study we have generated matrices whose elements are normally distributed random variables with mean 0.5 and variance 0.1667. The variance has been chosen in order to ensure $a_{i,j} \in [0, 1]$. It can be proven that, as a consequence of the normal distribution of the matrix elements, a modification of the variance would not alter the parameter σ.

First, we investigate the effect of varying the matrix order N. To do this we have generated a set of 5 matrices for each value N. None of the matrices generated has null off-diagonal elements. In Figure 1 we show the plot of the error variable σ, averaged over the 5 matrices, as a function of the matrix order N. The error variable falls rapidly with increasing matrix size, and becomes insignificant when $N > 500$.

Next we consider the effect of increasing the sparseness of the matrices. To do this we have generated random element matrices of size $N = 200, 500, 1000$. We have controlled the degree of sparseness of these matrices by setting to zero a specified fraction of randomly selected off-diagonal elements. For each fraction of zero elements, we have generated 5 different matrices and computed the average value of σ. Figure 2 shows the result of this study. Here the different curves are for different values of N.

5 Experiments

To evaluate the practical utility of our eigenvector approximation method, we have used it in conjunction with the grouping method described in [15]. This eigenclustering algorithm makes heavy use of non-sparse, non-negative matrices. We have compared the results obtained when the leading eigenvector is computed using the approximation described in this and the the well-known power method. Since the order of the matrix A decreases with each iteration of the grouping process, provisions have been taken when using the approximation method in order to avoid processing matrices with $N < 100$.

We commence by showing results obtained using simulated images. For this study, we use scenes composed of three rectangular regions, as illustrated in Figure 3. To the synthetic images we have added Gaussian noise of increasing standard deviation. In the figure, from left to right the different columns show the original images, the segmentaion result obtained using the power method and the result obtained using our eigenvector approximation method. The rows corespond to different levels of Gaussian noises, i.e., 35% and 50% of the grey scale difference between the regions. There are no significant differences between the results delivered by the two methods. The result of a more systematic quantitative evaluation is shown in Figure 4. Here we plot the percentage of mis-assigned pixels as a function of the noise standard deviation quoted as a fraction of the grey-scale difference between regions. Again, there is little to distinguish the performance of the two methods.

Finally, we study some real-world imagery. In the left hand column of Figure 5 we show the original images, the middle column shows the segmentations obtained when the leading eigenvector is computed using the power method and the right hand column shows the result obtained when we use our approximation method. There is no apparent segmentation difference between the two methods. However, the new approximation method is on average 2 times

faster when applied to the synthetic images and 2.5 times faster when processing the real images. In our experiments the error parameter σ never exceeded 0.005.

6 Conclusions

In this paper, we have presented a non-iterative method for approximating the leading eigenvector of a large, dense, non-negative symmetric matrix with the aim of developing more efficient graph-spectral image segmentation algorithms. The method decomposes the original matrix using a linear combination of perturbation matrices whose spectral properties are well known. This results in a significant improvement in computational efficiency. Moreover, in our sensitivity experiments, the approximation proven to be a reliable one which results in only a small loss of accuracy. Moreover, in a practical setting, the method reduces the required computation time by a factor of at least 2.

References

[1] Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[2] J. Shi and J. Malik. Normalized cuts and image segmentations. In *Proc. IEEE CVPR*, pages 731–737, 1997.

[3] S. Sarkar and K. L. Boyer. Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. *Computer Vision and Image Understanding*, 71(1):110–136, 1998.

[4] P. Perona and W. T. Freeman. Factorization approach to grouping. In *Proc. ECCV*, pages 655–670, 1998.

[5] Y. Weiss. Segmentation using eigenvectors: A unifying view. In *IEEE International Conference on Computer Vision*, pages 975–982, 1999.

[6] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, 1965.

[7] G. M. Del Corso. Randomized error estimation for eigenvalue approximation. *Calcolo*, 37:21–46, 2000.

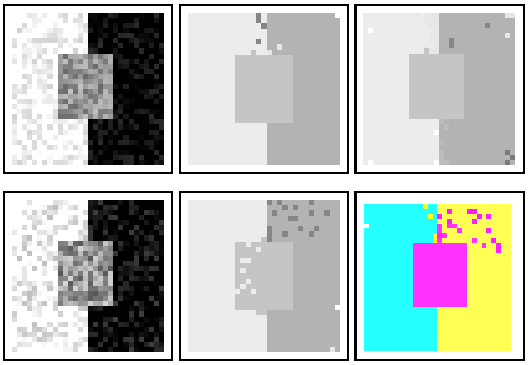

Figure 3. Left: Orignal images with different noise levels, Middle: Output of the power method, Right: Our approximation

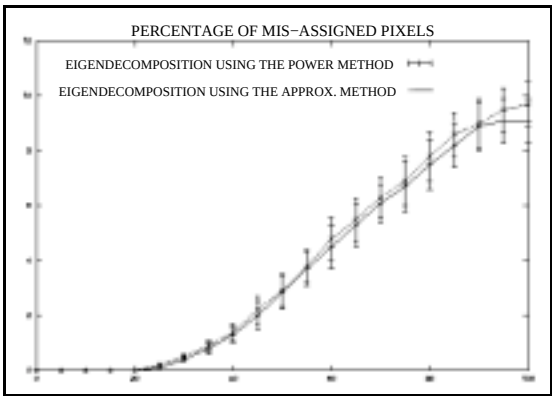

Figure 4. Percentage of mis-assigned pixels

[8] E. J. Nyström. Über die Praktische Auflösung von Linearen Integralgleichungen mit Anwndungen auf Randwertaufgaben der Potentialtheorie. *Commentationes Physico-Mathematica*, 4(15):1–52, 1928.

[9] G. W. Stewart and Ji guang Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

[10] C. Fowlkes, S. Belongie, and J. Malik. Efficient spatiotemporal grouping using the Nyström method. In *CVPR01*, 2001.

[11] W. T. Vetterlin W. H. Press, S. A. Teukolsky and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.

[12] L. R. Williams. Perceptual organization of occluding contours. In *IEEE International Conference on Computer Vision*, pages 133–137, 1990.

[13] R. S. Varga. *Matrix Iterative Analysis*. Springer, second edition, 2000.

[14] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*, volume 9 of *Classics in Applied Mathematics*. SIAM, 1994.

[15] A. Robles-Kelly and E. R. Hancock. A maximum likelihood framework for iterative eigendecomposition. In *Proc. of the IEEE International Conference on Conputer Vision*, 2001.

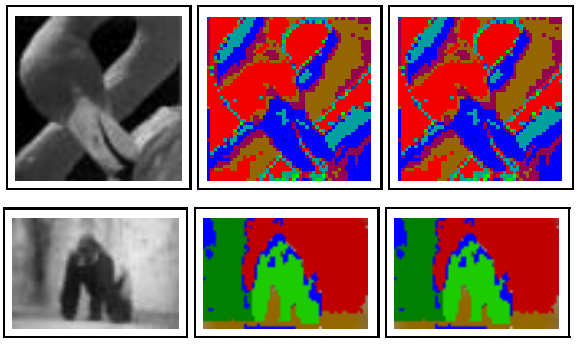

Figure 5. Left: Orignal images, Middle: Output of the power method, Right: Our approximation