# A Comparison of Two Random-Sampling Approaches to Spectral Clustering in Tissue Classification

William R Crum*

Kings College London, Centre for NeuroImaging Sciences (PO89), Institute of Psychiatry, London SE5 8AF

**Abstract.** There is increasing interest in applying spectral clustering (SC) algorithms to classification problems in medical imaging. These techniques model pair-wise voxel similarity relationships to generate feature eigenvectors which capture complex clustering structure but naïve implementations are computationally impractical for real applications. In previous work we described a more efficient approach to SC using stochastic sampling and sparse matrix methods. An alternative stochastic sampling approach, the Nyström Extension, solves a reduced eigenvector problem and applies a formal interpolation strategy to generate spectral features from the reduced eigenvector set. In this paper we show that these methods are distinct and have different properties. We compare and contrast SC using both approaches applied to classification problems in simulated and real medical images.

## 1 Introduction

Medical image analysis has benefitted from classification techniques for many years [1]. The most common applications are to extract different tissue types using features based on scalar or vector (multi-modal) voxel intensities. In some common applications such as brain-tissue classification, purely data-driven schemes have mostly been replaced by those with prior knowledge, either expressing probabilities of neighbouring tissue types or tissue probability as a function of spatial location (e.g. FAST [2]). However, most techniques still feature a data-driven initialisation and where good models do not yet exist, data-driven classification may be the only technique available.

Recently a new class of data-driven classification technique, Spectral Clustering (SC) [3], has attracted interest. Unlike some of the most common techniques, (e.g. k-means), SC models the interactions between all pairs of data-points in the ensemble. It moves away from the "point-to-cluster" analysis of simpler techniques and instead performs a "point-to-point" analysis. A detailed connectivity map is used to generate new features for each point from an eigenvector analysis; a powerful theoretical framework [3] motivates the use of these features for clustering. As we described in recent work [4] this pair-wise analysis quickly becomes impractical for 3D medical images with number of voxels, n, $\sim 1E6$. Our solution was to adopt a stochastic sampling (SS) approach where only $m << n$ pair-wise interactions are analysed for each voxel leading to a much reduced feature-generation problem. Our approach was motivated from the theory of SC and in particular how the connectivity properties of the data define the resulting spectral features. An alternative approach to SS is the so-called Nyström Extension (NE) [5] where eigenvectors obtained from a reduced data-set are interpolated to approximate those obtained from the full data-set.

Spectral Clustering algorithms using SS or NE have superficial similarities as both use random sampling and solve a reduced problem to generate features for classification. In this paper we consider the operation of both methods to show that they are distinct, and perform an empirical evaluation of their performance in a well-defined clustering problem.

## 2 Methods

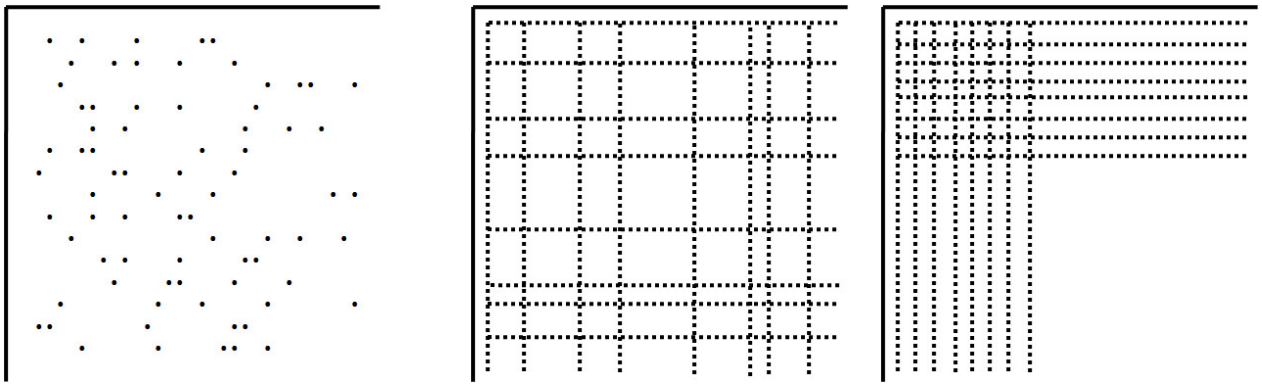
### 2.1 Spectral Clustering

We present a very brief summary of the mechanics of SC for tissue classification and refer the interested reader to [3] for more detail. Spectral Clustering builds two matrices which together describe the similarities between all voxels. The weights matrix, $\mathbf{W}$ is the $n \times n$ symmetric matrix of pair-wise voxel-similarities. The degree matrix $\mathbf{D}$ is the $n$-diagonal degree matrix formed from the row or column sums of $\mathbf{W}$ and summarises the voxel-to-ensemble relationship. These two matrices are combined to form the matrix representation of the Graph Laplacian which is symmetric, positive semi-definite with smallest eigenvalue equal to 0; in this work we use the simplest form of the Laplacian, $\mathbf{L} = \mathbf{D} - \mathbf{W}$. To construct spectral features, the eigenvectors associated with the next $k$ smallest eigenvalues of $\mathbf{L}$, are written into an $n \times k$ matrix $\mathbf{F}$. Spectral features associated with the $i^{\text{th}}$ voxel, are taken from the $i^{\text{th}}$ row of $\mathbf{F}$. The final classification is obtained by clustering the spectral features using standard techniques such as k-means. The number of feature components, $k$, is usually related to the number of expected clusters, $n_c$.

*Email: bill.crum@iop.kcl.ac.uk

### 2.2 Stochastic Sampling

In [4] we described the Stochastic Sampling (SS) strategy for SC. Briefly, for each of the $n$ voxels in the clustering problem, $m$ other voxels are chosen randomly and the $m$ pairwise similarities computed. Eigenvectors are then computed directly from the Laplacian using sparse-matrix techniques [6]. The justification for this approach is that useful spectral features are defined by the global connectivity properties of the underlying clusters in the ensemble rather than the detail of that connectivity. Provided this global connectivity is captured by the sparse-sampling procedure, suitable spectral features will be generated. Random graph theory suggests that a sparsely sampled Laplacian matrix will "almost-always" be connected for $m \sim> \log n$ meaning $m \geq 20$ is a safe choice for $n \sim> 1$ million [3].



(a) Stochastic sampling (b) Original Nyström sampling. (c) Nyström sampling after rearrangement

**Figure 1.** Schematic showing the sampling procedure in the similarity matrix, $\mathbf{W}$, for the Stochastic Sampling and Nyström Extension methods. The sampling density has been exaggerated for illustration. See text for further details

### 2.3 Nystrom Extension

The Nyström Extension (NE) is a well-known method for eigenvector interpolation applied to SC problems in [5]. In the NE method, $m$ voxels are chosen randomly and their pairwise similarities with each other and the other $n - m$ voxels are recorded (Figure 1(b)); the resulting Laplacian is connected by construction since each voxel is either one of the $m$ selected or is connected to one of the $m$ selected. In practice, $\mathbf{W}$ is reordered to place the $m$ fully-sampled rows and columns in a block (Figure 1(c)) to facilitate matrix sub-block manipulation. The justification for this approach is that spectral features in the fully-sampled case are derived from the eigenvectors of the Laplacian. The NE approach formally generates approximations to these eigenvectors by interpolating those obtained from the reduced problem.

### 2.4 Comparison of SS and NE Approaches

It is immediately apparent that SS and NE sampling strategies are not equivalent. In general a SS similarity matrix cannot be reordered to form an NE matrix because the latter has $m$ fully sampled rows and columns; the SS similarity matrix generally has no fully-sampled rows or columns. This is a consequence of choosing a single random set of $m$ voxels in the NE approach but choosing $n$ independent random sets of $m$ voxels in the SS approach. This leads to very different representations of connectivity. Assuming that all sampled voxel pairs have similarities $> 0$, then any voxel in the NE representation is at most two steps away from any other voxel (since all voxels are compared to the $m$ random samples). In the SS representation voxels will generally be separated by more steps (for $1E6$ voxels with $m = 30$ the average separation $\sim 4$ steps).

## 3 Experiments and Results

### 3.1 Experimental Data and Evaluation Methods

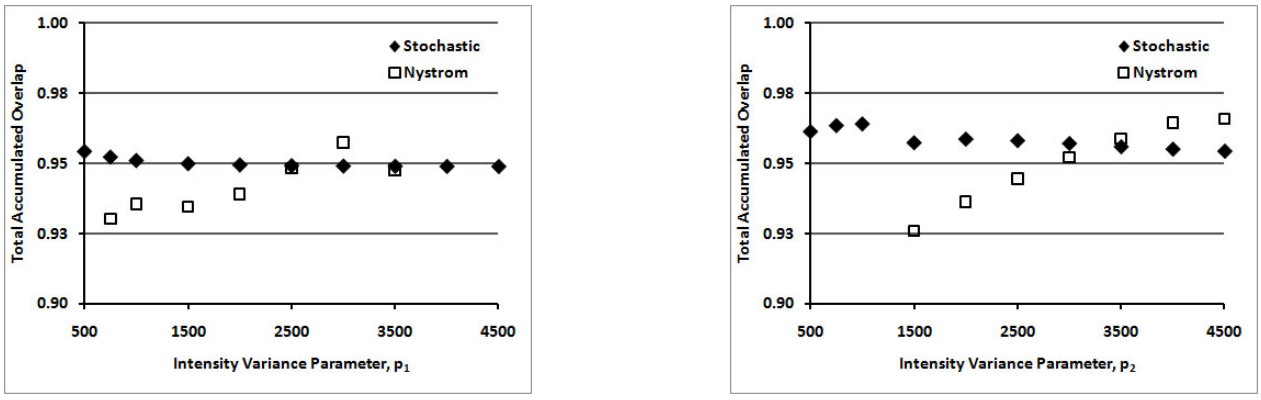
For the comparison of methods we return to the well-understood and well-characterised example of brain-tissue classification used in [4]. For intial parameter selection and sample-number studies, we used the MNI BrainWeb [1] digital brain phantom T1-weighted (T1w) and T2-weighted (T2w) 1mm isotropic voxel images [7]. For a broader accuracy test in simulated data we used 10 subjects chosen at random from the 20 available BrainWeb simulations of normal

[1] http://www.bic.mni.mcgill.ca/brainweb

brains [8]. In both cases the brain and cerebral-spinal-fluid (CSF) were extracted from background to leave $n_c = 3$ underlying clusters (CSF, grey-matter (GM), white-matter (WM)) and Rician noise of $2\%$ maximum intensity was added. For a semi-quantitative comparison we used multi-spectral 3D MRI from $10$ normal volunteers acquired as part of an imaging protocol evaluation on a General Electric 1.5T Signa HDx scanner. The images were axial T1-weighted (T1w) MPRAGE $(256 \times 256 \times 180)$ and T2-weighted (T2w) $(512 \times 512 \times 36)$ corrected for intensity inhomogeneity using N3 [9] and preprocessed using BET to remove non-brain/CSF voxels. The T2w scans were rigidly registered with the T1w scans and resampled resulting in all scans having voxel dimensions of $0.9375 \times 0.9375 \times 1.2$mm. To measure overall classification accuracy and compare classifications we computed the Total Accumulated Overlap (TAO) [10] over GM, WM and CSF. The TAO is in the range $[0, 1]$ which reduces to the standard Dice overlap for a single label comparison but allows the overall agreement of multiple labels to be summarised by a single index.

### 3.2 Parameter Sensitivity and Stochastic Consistency in the BrainWeb Digital Phantom

There are two key parameters common to both schemes. The similarity function used to compare voxel-pairs is $s_{ij} = \exp(-x_{ij}^2/p^2)$ where $x_{ij}$ is the Euclidean intensity distance between voxels $i$ and $j$ and $s_{ij}$ is parameterised by $\mathbf{p}$, a scalar/vector for single-/multi- images respectively. The second shared parameter is the number of random samples, $m$.




(a) T1w Image

(b) T1w and T2w Image

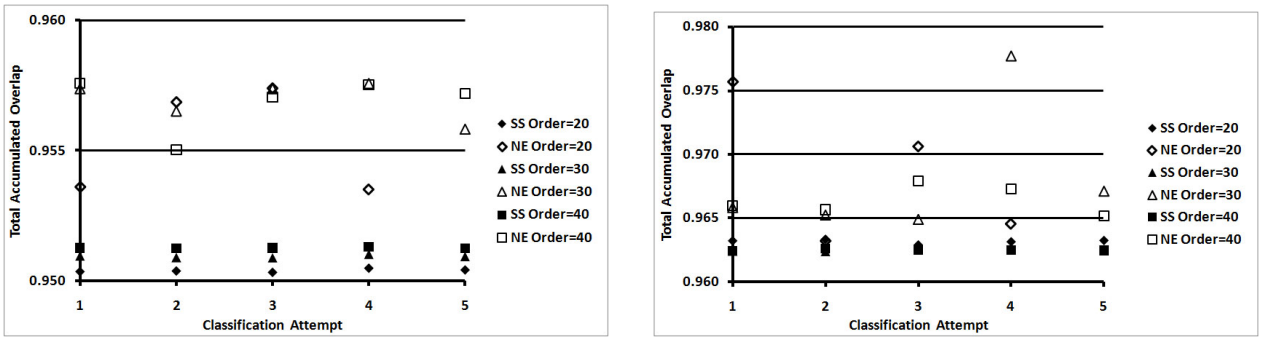
**Figure 2.** Classification accuracy of SS and NE as a function of intensity similarity parameter in the BrainWeb digital phantom

From previous experience with SS, we set the number of samples high at $m = 40$ and compared the sensitivity of the classification accuracy to the intensity parameters $p_1^2$ and $p_2^2$ associated with the T1w and T2w images respectively. We chose suitable values for $p_1$ from experiments using the T1w image (intensity range $[0, 4500]$) alone and then evaluated the performance in multi-channel classification using T1w and T2w images with fixed optimal $p_1$ and varying $p_2$. Figure 2(a) shows how the classification accuracy is a stronger function of $p_1$ for NE than SS but that peak accuracy is similar. Figure 2(b) shows similar behaviour for multi-channel data. We chose $\mathbf{p} = (1000, 1000)$ (SS) and $\mathbf{p} = (3000, 4500)$ (NE) for the next experiment.

We then varied the number of random samples $m = \{20, 30, 40\}$ and ran the classification repeatedly using the previously selected values for $\mathbf{p}$. It can be seen in figure 3 that the NE approach consistently results in higher classification accuracy than the SS approach but also has higher variability too. For the two-channel classification experiment, the variability is more pronounced in NE, particularly for low numbers of samples.

### 3.3 Classification Accuracy in $10$ Simulated Normal Brains

We applied SS and NE Spectral Clustering and the standard method, FAST [2], to the $10$ T1w BrainWeb normal subjects and assessed the classification accuracy as before (figure 4). This is a comparatively straight-forward classification problem and the results are comparable across all techniques with the exception of case $45$. We further examined the SS result for this case and found that the k-means clustering of the spectral features had not converged to $3$ distinct clusters. We repeated the classification of this subject with all parameters the same as before (including the random-number initialisation), but increasing the number of k-means random initialisations from $25$ to $50$. The TAO for this case then increased to $0.90$ showing that the original spectral features had captured the cluster-structure of the data but the subsequent k-means clustering of spectral features had reached a local optima.




(a) T1w Image

(b) T1w and T2w Image

**Figure 3.** Classification consistency for repeated independent applications of Spectral Clustering with Stochastic Sampling and the Nyström Extension in the BrainWeb digital phantom

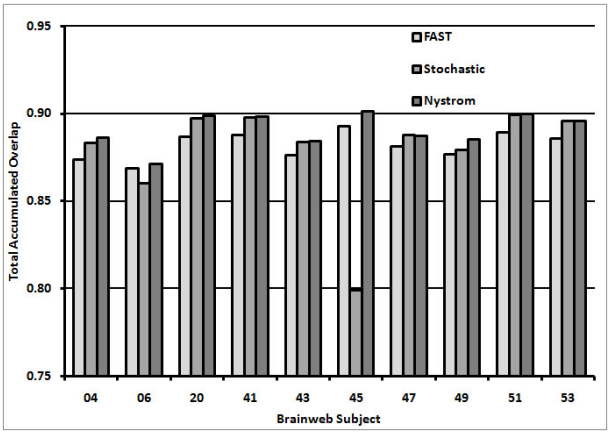


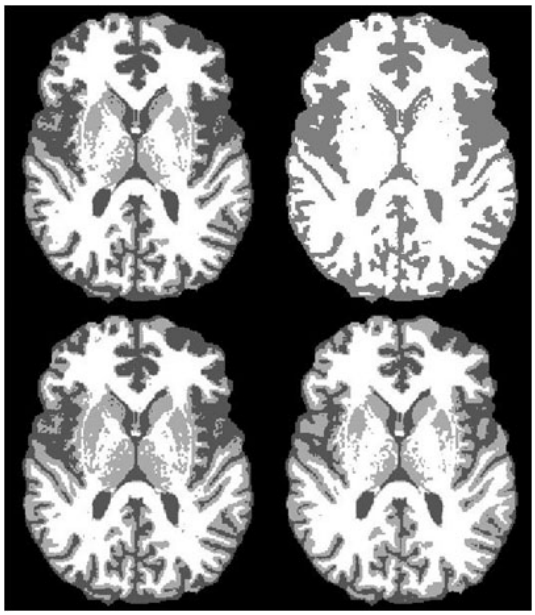

**Figure 4.** Classification accuracy in 10 simulated normal subjects using FAST, SS and NE

### 3.4 Comparison in Real Data

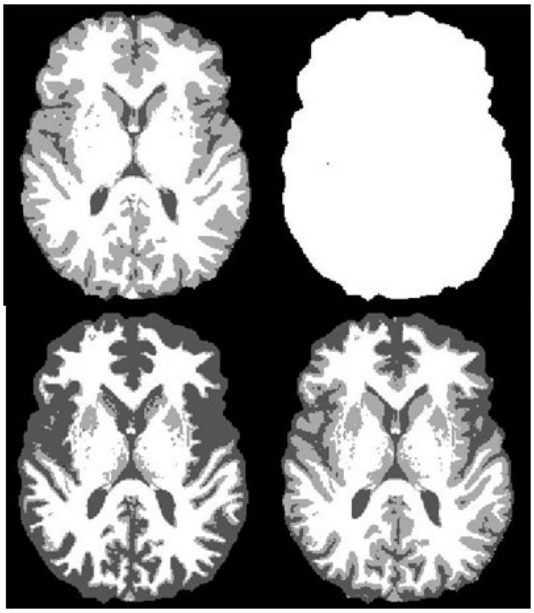
We applied SS and NE to the T1w images of 10 normal subjects. The similarity parameters, $p_1$ and $p_2$ were expressed as fractions, $q = \{1, 2, 3\}$, of the mean intensity of foreground voxels. We assessed consistency within-technique by computing the TAO over all 10 subjects and all 3 q-values. We assessed consistency across-technique by computing the TAO over all 10 subjects for each of the 6 combinations of q-values. The within-technique TAO for SS/NE in T1w images was $0.95/0.58$. The lower consistency for NE is can be partially attributed to outright classification failures (Figure 5). The highest across-technique TAO was $0.83$ for q-values for SS/NE of $3/1$. The results for T1w+T2w images were similar: within-technique TAO for SS/NE was $0.93/0.33$ with highest across-technique TAO $0.79$, again for q-values for SS/NE of $3/1$. These results are consistent with section 3.2

## 4 Discussion

We have investigated the differences between two methods of classification using Spectral Clustering with reduced sampling and have obtained good results with very low numbers of samples compared to solving the full problem. There is clearly an interaction between the number of samples used and the expected number and size of clusters. Small clusters are more likely to be poorly represented when few samples are used and the relationship of sample to cluster size will be investigated in future work. The techniques summarise the connectivity of the data in markedly different ways. The Nyström Extension method produces higher accuracy classifications in simulated data but is less stable with respect to parameter choice. In practice we also found that NE occasionally resulted in outlier values in some spectral features and/or spectral features which resulted in spurious clustering. Here, we deleted outliers and excluded clusters of size less than $1/n_c\%$ of the input volume during the k-means phase. For our implementation, SC with NE is approximately $4$ times faster than the SS equivalent. However, while promising in terms of computational efficiency, the stability and consistency of SS with NE warrants further investigation.



(a) T1w Classification: left/right = q = $1/3$, top/bottom = NE/SS



(b) T1w and T2w Classification: left/right = q = $1/3$, top/bottom = NE/SS

**Figure 5.** Classification examples in a single subject using Stochastic Sampling and the Nyström Extension. Note that in both cases NE with $q = 3$ is a classification failure. See text for further details.

## Acknowledgements

The author acknowledges financial support from the Department of Health via the National Institute for Health Research (NIHR) Specialist Biomedical Research Centre (BRC) for Mental Health award to South London and Maudsley NHS Foundation Trust (SLaM) and the Institute of Psychiatry at King's College London. The author also thanks Andrew Simmons and the BRC for the volunteer data.

## References

1. A. Liew & H. Yan. "Current methods in the automatic tissue segmentation of 3D magnetic resonance brain images." *Current Medical Imaging Reviews* **2(1)**, pp. 91–103, 2006.
2. Y. Zhang, M. Brady & S. Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm." *IEEE Transactions on Medical Imaging* **20(1)**, pp. 45–47, 2001.
3. U. von Luxburg. "A tutorial on spectral clustering." *Statistics and Computing* **17**, pp. 395–416, 2007.
4. W. R. Crum. "Spectral clustering and label fusion for 3D tissue classification: Sensitivity and consistency analysis." In *Medical Image Understanding and Analysis*, pp. 149–153. 2008.
5. C. Fowlkes, S. Belongie, F. Chung et al. "Spectral grouping using the Nyström method." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28(2)**, pp. 214–225, 2004.
6. R. Geus. *The Jacobi-Davidson algorithm for solving large sparse symmetric eigenvalue problems with application to the design of accelerator cavities*. Doctor of technical sciences, Swiss Federal Institite of Technology, Zurich, 2002.
7. C. Cocosco, V. Kollokian, R.-S. Kwan et al. "BrainWeb: Online interface to a 3D MRI simulated brain database." *NeuroImage* **5(4)**, 1997.
8. B. Aubert-Broche, A. Evans & L. Collins. "A new improved version of the realistic digital brain phantom." *NeuroImage* **32(1)**, pp. 138–145, 2006.
9. J. G. Sled, A. P. Zijdenbos & A. C. Evans. "A non-parametric method for automatic correction of intensity non-uniformity in MRI data." *IEEE Transactions on Medical Imaging* **17(1)**, pp. 87–97, 1998.
10. W. R. Crum, O. Camara & D. L. Hill. "Generalized overlap measures for evaluation and validation in medical image analysis." *IEEE Transactions on Medical Imaging* **25(11)**, pp. 1451–1461, 2006.