Available online at www.sciencedirect.com
SCIENCE DIRECT
International Journal of Rock Mechanics & Mining Sciences 43 (2006) 1052–1061
International Journal of Rock Mechanics and Mining Sciences
www.elsevier.com/locate/ijrmms

A spectral method for clustering of rock discontinuity sets

R. Jimenez-Rodriguez[a,b,*], N. Sitar[a]

[a]Department of Civil and Environmental Engineering, University of California at Berkeley, USA
[b]ETS Ing. de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Spain

Accepted 13 February 2006
Available online 3 April 2006

Abstract

We demonstrate the use of a spectral clustering algorithm as a novel approach for the identification of rock discontinuity sets based on discontinuity orientations. We use the spectral clustering approach with a simple measure of similarity between normal unit vectors in spherical space that is specific to the clustering of rock discontinuity orientations. The performance of the algorithm is studied using benchmark test cases with data sets corresponding to real rock masses. The results show that the algorithm provides good clustering results, providing partitions that agree well with the results of several other clustering algorithms that are commonly used in rock engineering. Furthermore, we show an example case with data sets of discontinuity orientation compiled from the literature, in which the spectral clustering algorithm provides more natural partitions than the other algorithms considered. Additional advantages of the algorithm are that convergence is fast, and that it can be easily (and efficiently) implemented using popular software packages for numerical analysis.

Keywords: Joint sets; K-means; Orientation analysis

1. Introduction

The grouping of discontinuities into discontinuity sets and the characterization of their orientation is an important aspect of rock mass characterization for engineering applications (see e.g., [1–8]). The representation of discontinuity orientation data and the identification of discontinuity sets is commonly performed using techniques for hemispherical projection of discontinuity poles (i.e., unit vectors with direction normal to discontinuity planes) [1,9–11]. Methods for visual clustering on density contours computed by counting the number of poles that fall inside a reference circle are commonly used, but they present problems due to sampling bias that need to be corrected [12–14]. Additional difficulties are due to clustering results heavily depending on the size of the counting circle [11], and to subjectivity in the interpretation of the clustering results [1,15,16]. In summary, counting methods have not been found entirely satisfactory in some cases [17,18], leading to the development of alternative techniques for automatic identification of discontinuity sets.

Methods for automatic clustering of discontinuities based on their orientation can be divided into methods that assume an initial probabilistic structure of the discontinuity orientation data [15,19–21], and methods that group discontinuities without considering an a priori probabilistic model [11,16,18,22,23]. Methods for classification of discontinuity sets based on the use of artificial neural networks have also been proposed [24], as well as clustering methods that can incorporate information in addition to the orientation of discontinuities, such as, for instance, planarity, weathering, spacing, or roughness (see e.g., [16,20,23]).

Assuming an a priori probabilistic model of the discontinuity data allows to define confidence intervals and to test statistical hypothesis [21]. In addition, it allows to cluster discontinuities based on their probabilities of membership to each discontinuity set [20,25]. Mahtab and

*Corresponding author. Current address: ETS Ing. de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Spain. Tel.: +34 91 336 6710; fax: +34 91 336 6774.
E-mail address: rjimenez@caminos.upm.es (R. Jimenez-Rodriguez).

1365-1609/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijrmms.2006.02.003

Yegulalp [15] proposed a clustering algorithm for the objective partition of discontinuity data using a rejection scheme based on a randomness test derived from the Poisson distribution; Dershowitz et al. [20] presented an iterative stochastic algorithm in which fractures are assigned in each step to the discontinuity set in which they have the highest probability of membership; and Marcotte and Henry [21] proposed a method for maximum likelihood identification of discontinuity sets assuming that the orientation of each discontinuity set can be modeled as a mixture of truncated bivariate normal distributions.

In some cases, however, making a priori assumptions about the probabilistic structure of discontinuity properties may be a difficult task, and clustering methods that use no a priori probabilistic information have been developed as well. The definition of distance metrics (i.e., measures of the distance between observations in the data set) that are adequate to the structure of the data (i.e., the shape of clusters in the data set) in each particular case has been recognized as a key issue for the successful application of such algorithms (see e.g., [11,22]). In that sense, the ideas for clustering of discontinuity orientations using the fuzzy K-means algorithm originally proposed by Harrison [11] have been extended with the introduction of improved distance metrics and performance measures [16,22,26]; Zhou and Maerz [23] developed a computer program for the characterization of discontinuity sets using several multivariate clustering algorithms; and Klose et al. [18] proposed a clustering approach based on vector quantization and stochastic minimization of a cost function defined in terms of the acute angle between discontinuity poles and the average pole of the discontinuity set to which they are assigned [18,27].

Spectral clustering algorithms group points using eigenvectors of matrices derived from the data, and they have been successfully employed in applications including machine learning, speech processing, and computer vision [28–30]. In this paper we demonstrate the use of a spectral clustering algorithm as a novel approach for the identification of rock discontinuity sets based on discontinuity orientations. We build on the work of Ng et al. [29] and we explore the capabilities of the spectral clustering algorithm for identification of discontinuity sets, using a simple similarity measure in spherical space that is specific to the clustering of rock discontinuity orientations. We further illustrate the performance of the algorithm in a number of test cases, including data sets corresponding to real rock masses, and we compare the clustering results computed using the proposed spectral algorithm with results computed with other clustering algorithms used in rock engineering applications.

2. Representation of directional data

2.1. Introduction

We assume that rock discontinuities can be represented as planar surfaces. Therefore, the orientation of discontinuities can be expressed in terms of the orientation of their unit normal vector; i.e., a vector of unit length with direction normal to the discontinuity plane. We follow the usual convention in rock mechanics that discontinuity unit normal vectors point toward the lower hemisphere of the unit reference sphere.

We use the method proposed by Priest [13] to denote orientations of discontinuity unit normal vectors. That is, the orientation of a discontinuity unit normal vector is expressed as $\Theta \equiv (\alpha, \beta)^{\mathrm{T}}$ (see Fig. 1), where α is the *trend* (i.e., the azimuth angle from the north to the vertical plane containing the discontinuity unit normal vector, measured in clockwise direction), and β is the *plunge* (i.e., the angle measured in a vertical plane between the discontinuity unit normal vector and the horizontal plane) of the downward-pointing discontinuity unit normal vector. As shown in Fig. 1, unit vectors can also be expressed in terms of a three-dimensional Cartesian coordinate system $\mathbf{x} \equiv (x_1, x_2, x_3)^{\mathrm{T}}$. The Cartesian coordinates, $\mathbf{x}$, of a unit vector with orientation $\Theta \equiv (\alpha, \beta)^{\mathrm{T}}$ are given by [13,18]

$$x_1 = \cos\alpha \cos\beta, \tag{1}$$

$$x_2 = \sin\alpha \cos\beta, \tag{2}$$

$$x_3 = \sin\beta. \tag{3}$$

In this paper, we represent discontinuity unit normal vectors by means of their equal-area lower hemisphere projection into the unit reference sphere. Equal-area projections are widely used for representation of rock discontinuities in rock engineering applications (see e.g., [7,15,17–21,24,31]). (Equal-angle projections are commonly used as well (see e.g., [7,11]); for a description of each method of projection and a discussion of its merits, see [10].) Our election of equal-area projections for representation of rock discontinuity data in this work only affects the visualization of the data, but not the results of the proposed spectral clustering algorithm.

2.2. Similarity measures

Clustering algorithms partition data into groups as a function of measures of the distance (or similarity) between

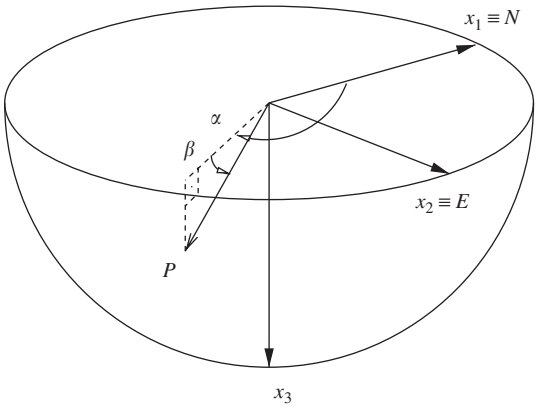

Fig. 1. Three-dimensional Cartesian coordinate system.

observations in the data set. The use of a particular distance metric induces a topology on the pattern space, leading to preferential identification of clusters of certain shape [22]. Therefore, the election of a distance metric that is adequate for each particular case has been identified as a key issue for the successful application of clustering methods [11,22].

Similarity measures based on the sine of the acute angle between discontinuity unit normal vectors have been used in rock engineering applications before [16,22,26]. Distance metrics based on the acute angle itself have also been employed [15,18,19], as well as Euclidean distance metrics [11,23], and Mahalanobis-type distance metrics that improve the capabilities of clustering algorithms to identify discontinuity sets of non-circular shapes [11,22].

We choose to quantify the similarity between the orientations of two discontinuities as the sine of the acute angle between the unit normal vectors of both discontinuities. (Note that more sophisticated distance metrics exist; however, our results show that the spectral clustering algorithm performs well even when such simple similarity measure is employed.) The acute angle between two unit normal vectors expressed in Cartesian coordinates, $\mathbf{x}_1$ and $\mathbf{x}_2$, can be computed as [18]

$$\delta = \arccos|\mathbf{x}_1^{\mathrm{T}} \cdot \mathbf{x}_2|, \tag{4}$$

where the $\cdot$ symbol indicates the dot product operation. Therefore, from Eq. (4), the sine-based similarity measure between discontinuity unit normal vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ becomes [22]

$$d^2(\mathbf{x}_1, \mathbf{x}_2) = 1 - (\mathbf{x}_1^{\mathrm{T}} \cdot \mathbf{x}_2)^2. \tag{5}$$

3. Clustering algorithm

We consider a data set with N measurements of rock discontinuity orientation that we aim to group into K discontinuity sets. The N discontinuities can be partitioned into K clusters (i.e., discontinuity sets) using the following algorithm for spectral clustering [29]:

1. Compute the affinity matrix $A \in \mathbb{R}^{N\times N}$, with elements given by $A_{ij} = \exp(-d^2(\mathbf{x}_i, \mathbf{x}_j)/2\sigma^2)$ if $i \neq j$, and $A_{ii} = 0$, where the squared distance is given by the sine-based distance in Eq. (5). The affinity is a measure of similarity (note that it is a function of the distance metric employed) between the orientation of two discontinuities in the data set. The scaling parameter σ^2 controls how the affinity decays with the distance between two observations. (The higher the value of σ, the slower the decay of affinity as distance increases.)
2. Define D as the diagonal matrix whose (i,i)th element is computed as the sum of the ith row of A, and compute the matrix $L = D^{-1/2} A D^{-1/2}$. (D_{ii} is therefore the sum of affinities of observation i to every other observation in the data set, and L is a normalized affinity matrix.)
3. Compute the K largest eigenvalues of L and their corresponding eigenvectors, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K$ (chosen to be orthogonal to each other in the case of repeated eigenvalues). Stack such eigenvectors by columns, forming the matrix $V = [\mathbf{v}_1\, \mathbf{v}_2 \ldots \mathbf{v}_K]$.
4. Form the matrix U from V by normalizing each row of V so that it has unit length. That is, $U_{ij} = V_{ij}/(\sum_j V_{ij}^2)^{1/2}$.
5. Considering each row of U as a point in $\mathbb{R}^K$, cluster such points (i.e., the rows of U) into K subsets using K-means algorithm.
6. Assign original point $\mathbf{x}_i$ to cluster j *if and only if* row i of matrix U is assigned to cluster j in Step 5.

That is, the spectral clustering algorithm performs a transformation of the N observations of rock discontinuity orientation in the data set, from the original space of Cartesian coordinates of unit normal vectors to a transformed K-dimensional space. The coordinates of points in the transformed space are given by the normalized (i.e., unit length) rows of a matrix obtained by stacking the main eigenvectors of the normalized affinity matrix of observations. For clusters that are connected and separate well (i.e., the affinity is negligible for points within different clusters and non-zero for points within the same cluster), Ng et al. [29] show that the transformed points cluster around K mutually orthogonal points that lie on the surface of the K-dimensional unit sphere; in addition, they show that these clusters correspond to the clusters of the original data set. (Ng et al. [29] also show the conditions under which the algorithm is expected to perform well in other cases.)

The value of σ^2 can be chosen by searching over σ until an "adequately small" distortion of clusters is obtained after using K-means on the rows of U in Step 5 [29]. In this research we found that values in the order of $\sigma = 0.1-0.15$ produce adequate clusterings results of discontinuity orientations for the sine-based similarity measure in Eq. (5). (In general, the optimal value of σ depends on the distance measure selected and on the data set; all results presented in this paper are computed using a value of $\sigma = 0.12$.)

As an illustration of the value of the approach, Figs. 2 and 3 show an example in which spectral clustering is an improvement over existing approaches. Discontinuities in the quasi-vertical discontinuity set in Fig. 2 correspond to Site c1904.1 from the MIT fracture-attitude data collection, as reproduced from Herda et al. [17]. Discontinuity data corresponding to the quasi-horizontal discontinuity set are taken from Hammah and Curran [22]. Discontinuity sets with stereological projections of similar shapes to those in Fig. 2 are not uncommon in rock mechanics applications, for instance in the context of folded rock masses (see e.g., [9,31]).

In Fig. 2(a) we show that, in this case, the fuzzy K-means algorithm (with a sine-based distance metric) fails to identify the natural clusters in the data set, as several quasi-vertical discontinuities (with unit normal vectors of

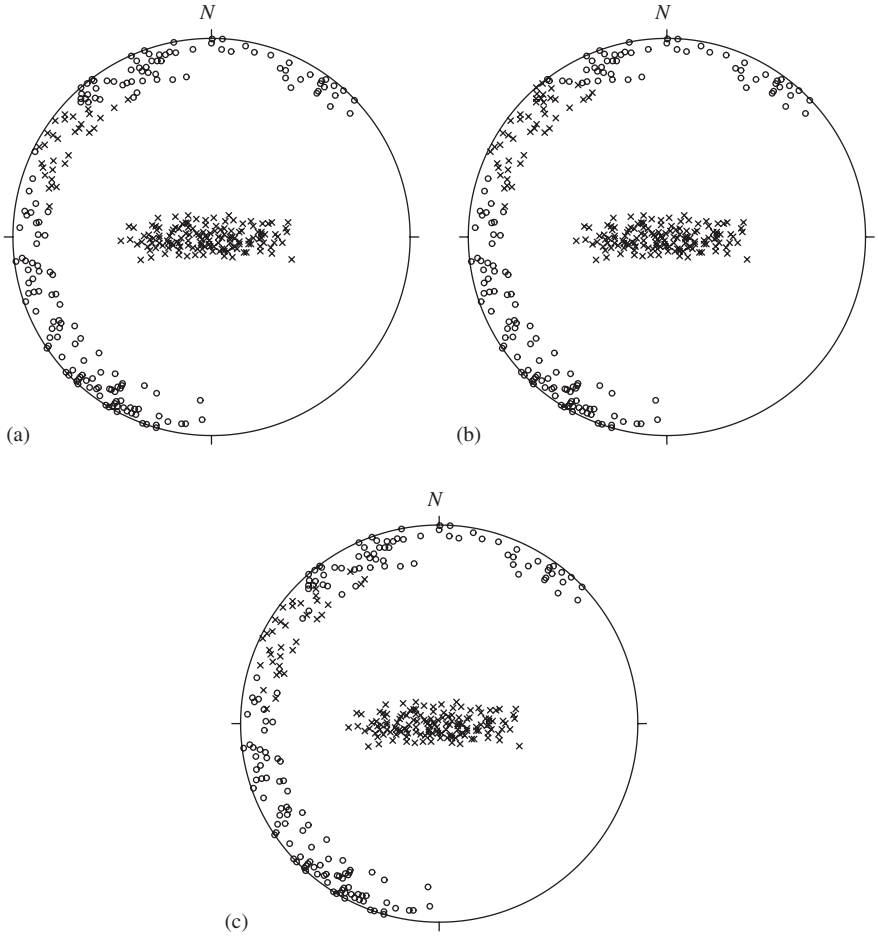

Fig. 2. Examples of difficulties encountered when clustering is performed in the original space of discontinuity orientations: (a) Fuzzy K-means (sine-distance); (b) Fuzzy K-means (Kent-distance); and (c) Vector quantization (acute-angle distance).

trend approximately N300E) are assigned to the joint set of quasi-horizontal discontinuities. Fig. 2(b) shows that using a distance metric based on the Fisher–Bingham distribution (also referred to as Kent distribution) [22,32] does not significantly improve the clustering results in this particular case. (The Kent distribution is an elliptical distribution for spherical data that has been shown to improve the results of the fuzzy K-means algorithm for elliptically shaped clusters in some cases [22].) Fig. 2(c) shows that similar difficulties are encountered when the clustering method based on vector quantization (defining distances as the acute angle between discontinuity unit normal vectors) is employed.

Fig. 3 shows the clustering results computed using the proposed spectral clustering method. These clustering results have been computed for the same data set as in Fig. 2, and we have employed the sine-based similarity measure presented in Eq. (5). Fig. 3(a) shows the transformed space given by the rows of U (as $K = 2$ in this case, it corresponds to a circle with unit radius), where we use K-means to identify partitions in the data set. Note in Fig. 3(a) that the transformed points do form tight clusters. Indeed, the coordinates of points are very similar (it is even difficult to identify individual points) and it is therefore easy to perform clustering using K-means. For ease of visualization, a subset of 10% of data points in Fig. 3(a) have been represented again (imposing small random displacements) in Fig. 3(b). In addition, note that vectors pointing from the origin to each cluster in the transformed space are orthogonal. Fig. 3(c) shows the clustering results obtained once that cluster partitions assigned using K-means in Step 5 are assigned to the corresponding points in the original space of discontinuity orientations in Step 6. The algorithm is shown to perform well (even when such simple sine-based similarity measure is employed), and it successfully separates discontinuities belonging to both discontinuity sets as a human interpreter would do.

4. Example analyses

4.1. San manual copper mine data set

We start with a data set of $N = 286$ discontinuity orientation measurements from San Manual copper mine in Arizona, USA [19]. Klose et al. [18] used this data set as a benchmark of several clustering algorithms in which they considered partitions with $K = 3$ discontinuity sets. In

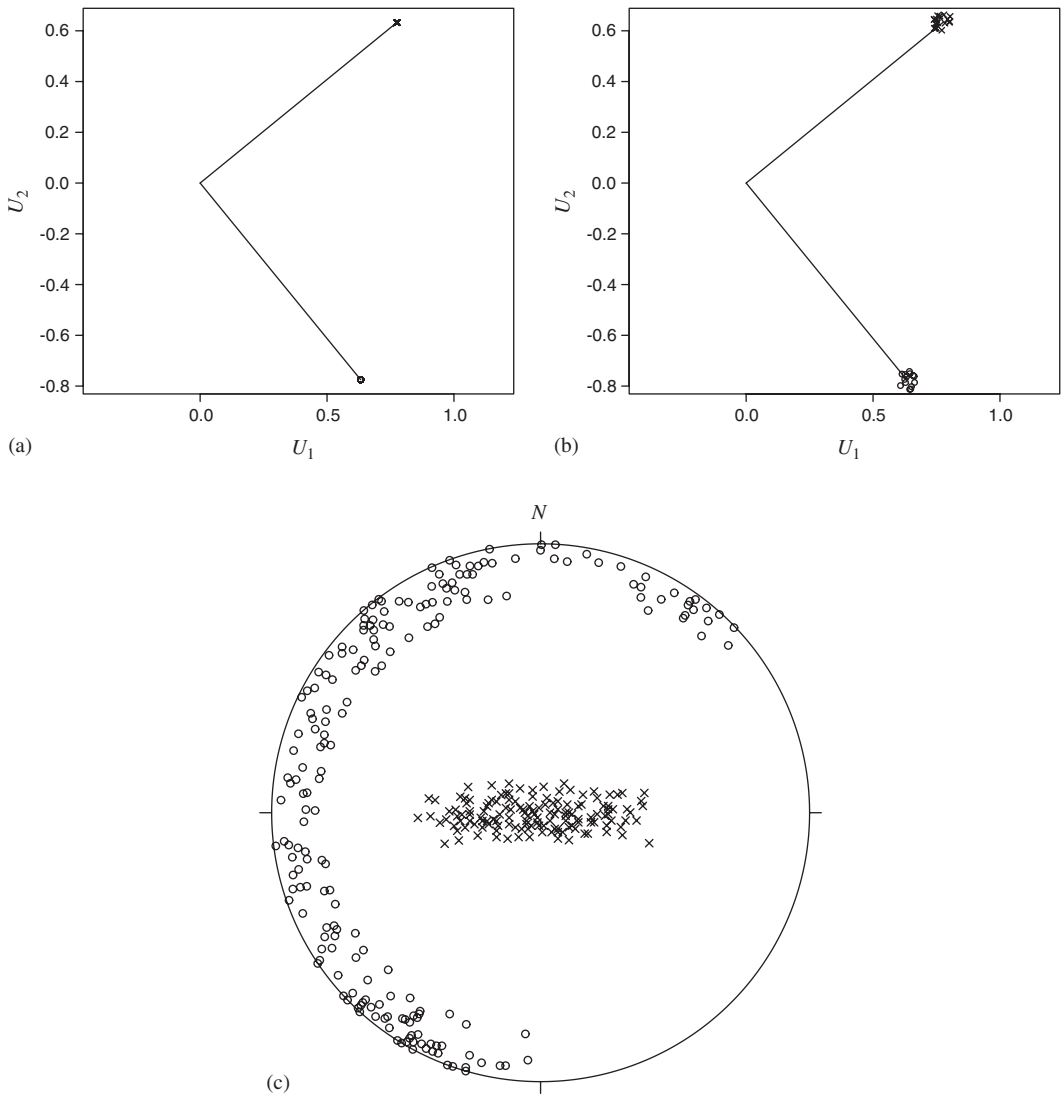

Fig. 3. Results of the proposed spectral clustering algorithm: (a) Results of K-means clustering in the transformed space; (b) Subset of data plotted in (a) (jittered for ease of visualization); and (c) Partitions in the original space.

Fig. 4 we use the same data set as a benchmark, and we compare the results of the spectral clustering algorithm (also with $K = 3$ discontinuity sets; see Fig. 4(d)) with the results of other common algorithms for clustering of discontinuity orientation data. (See Figs. 4(a)–(c).)

We compare our results with the results of the algorithm proposed by Shanley and Mahtab [19] (see Fig. 4(a)), because it is one of the first algorithms for objective clustering of discontinuity orientation data that is still widely used today [18]. We also compare our results with the results of the fuzzy K-means algorithm proposed by Hammah and Curran [16] (see Fig. 4(b)). In this case, we used a distance metric based on the sine of the angle between discontinuity unit normal vectors to compute the fuzzy K-means partitions. (Partitions computed using a distance definition based on the Kent distribution [22] were equivalent to those in Fig. 4(b) and, for the sake of brevity, they are not reproduced herein.) Finally, we compare our results with the results of the vector quantization algorithm (with a distance metric defined in terms of the acute angle between discontinuity unit normal vectors) presented by Klose et al. [18] (see Fig. 4(c)).

In Fig. 5 we show the K-means partitions computed in the transformed space of the rows of matrix U in Step 5 of the spectral clustering algorithm. In this case $K = 3$, and data points in the transformed space are therefore located on the surface of the unit three-dimensional sphere. Note that vectors pointing from the origin to each cluster average in the transformed space are again orthogonal to each other. Table 1 lists the average directions of the normal unit vectors of each identified discontinuity set (in trend/plunge notation). The spectral clustering algorithm is shown to perform well, providing clustering partitions that are very similar to those obtained with the other clustering methods. The average directions of each discontinuity set computed with the spectral clustering

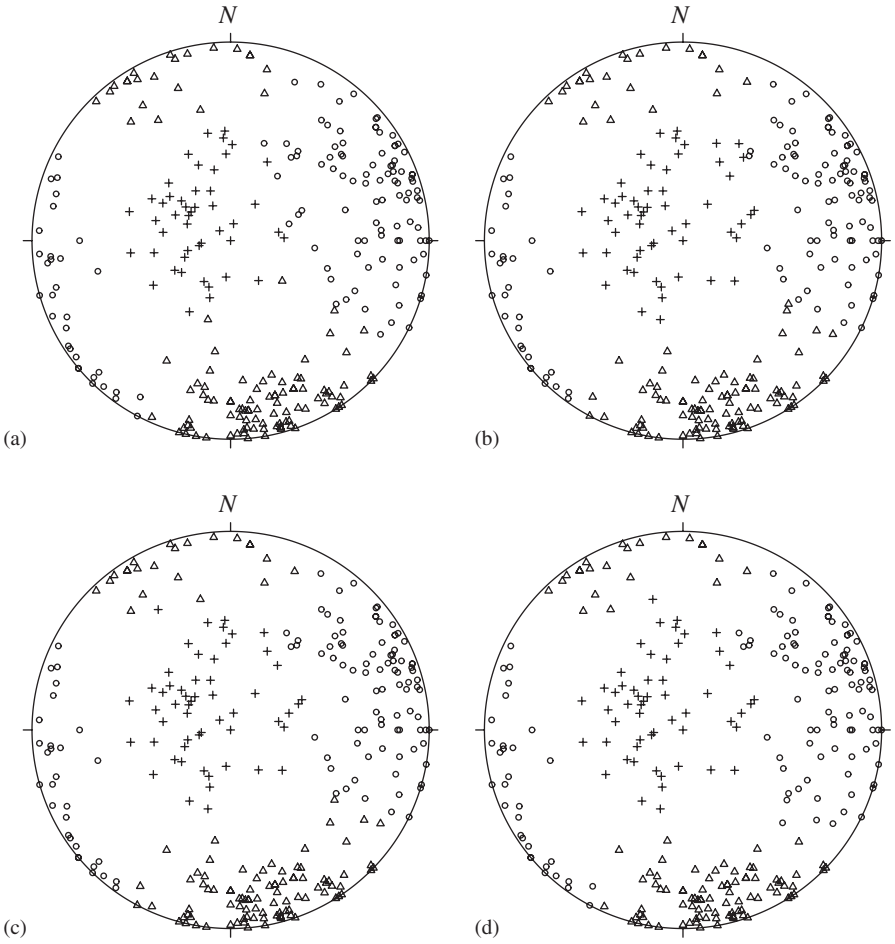

Fig. 4. Comparison of clustering results for different algorithms, considering $K = 3$ discontinuity sets (San Manual mine data set [19]): (a) Shanley and Mahtab; (b) Fuzzy K-means; and (c) Vector quantization; (d) Spectral clustering.

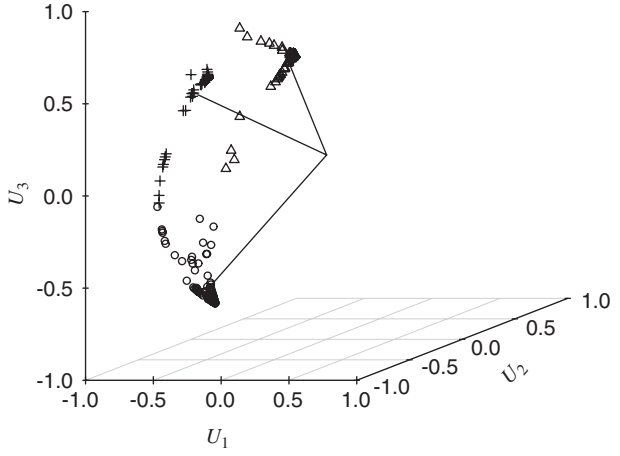

Fig. 5. K-means clustering in the transformed space of rows of U for the San Manual mine data set.

algorithm are further shown to agree well with the average directions of discontinuity sets computed with the other algorithms considered.

Since there is no "ground truth" for this clustering problem [18], it is difficult to know which of the clustering algorithms is better, or what partition should be preferred. In such circumstances, the election of a clustering algorithm may be based on performance, ease of implementation, and issues related to personal preferences and tradition in a particular field. In that sense, in Fig. 2 we showed an example case in which the spectral clustering algorithm provides more natural partitions than other algorithms. Additional advantage of the spectral clustering algorithm is that it can be easily implemented using common software packages for numerical analysis, such as the R environment for statistical computing [33] or MATLAB, which include functions for efficient computation of eigenvalues and eigenvectors, and for K-means clustering. Furthermore, the convergence of K-means clustering in Step 5 of the algorithm is usually fast, as we perform clustering in a transformed space where data are expected to form tight orthogonal clusters (e.g., see Fig. 3(a)). For instance, our implementation in the R language of the spectral clustering algorithm took a mean (after ten test cases) of 487 ms to compute the clustering results presented in Fig. 3.[1] If faster convergence is needed, we could take

[1]We used K-means algorithm using conventional random initialization and 10 restarts to compute the spectral clustering results that we present in

Table 1
Direction of average unit normal vectors for each discontinuity set (San Manual Mine data set)

Set	Shanley and Mahtab[a]	Fuzzy K-means	Vector quantization[a]	Spectral clustering
1 (○)	072/14	070/10	068/15	073/12
2 (△)	163/14	166/09	171/10	167/08
3 (+)	303/81	308/73	310/73	314/76

[a]Values reproduced from Klose et al. [18].

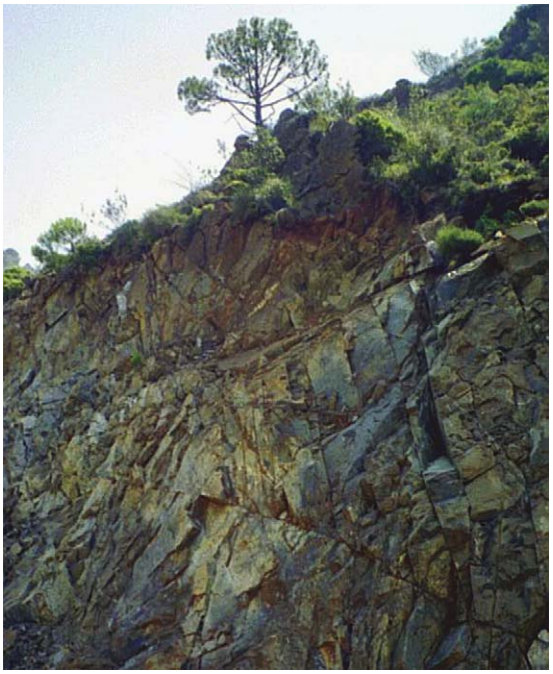

Fig. 6. Observations of discontinuities at exposed rock face.

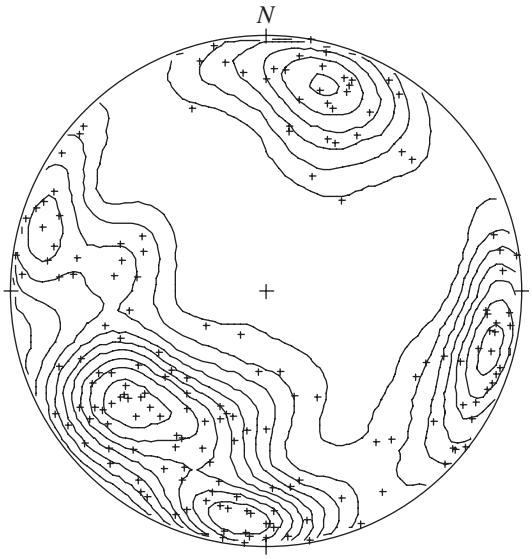

Fig. 7. Data set of discontinuity orientations in a serpentine rock mass in Ronda, southern Spain.

advantage of the orthogonality of cluster centers in the transformed space to improve the rate of convergence of the K-means algorithm by providing "good" initial cluster center estimates that are approximately orthogonal to each other [29].

4.2. Data set of discontinuities in a serpentine rock mass

We further test the spectral clustering algorithm using a data set of 185 measures of discontinuity orientations in a serpentine rock mass in southern Spain. Discontinuity orientation data are obtained by means of compass measurements of discontinuities selected using the scanline sampling method at exposed rock faces [13,34], and several scanlines are employed to reduce the effects of sampling bias. Fig. 6 shows typical discontinuity traces observed at exposed rock faces at the site. Fig. 7 (plotted with program GEOPLOT [35]) shows an equal-area, lower hemisphere projection of the normal unit vectors of each discontinuity mapped, together with contours that represent the estimated frequency of discontinuity normals in each direction in space. (Contour maps of unit normal vectors are commonly used to identify discontinuity sets in rock engineering projects; within that context, Fig. 7 can help a trained analyst to decide the number of discontinuity sets that should be used in a particular application.)

As an example, in Fig. 8 we present the computed spectral clustering partitions when $K = 2$ and 3 discontinuity sets are considered. For comparison, the results of the fuzzy K-means clustering with sine-based distance of Hammah and Curran [16] are presented as well. (The results using the distance metric based on the Kent distribution were found to be identical and they are not reproduced herein.) The clustering results in Fig. 8 show that the spectral clustering algorithm provides very similar partitions to those obtained with fuzzy K-means. As in the San Manual Mine data set, there is no "ground truth" to this clustering problem [18], and it is difficult to know which partition should be preferred. Based on visual observation of Fig. 8, however, we consider that partitions computed with the spectral clustering results are at least as natural as partitions computed with the fuzzy K-means algorithm. Similarly, Table 2 shows that the average unit

(*footnote continued*)
this work. Less than 10 iterations were needed for K-means convergence in all cases. Computations were performed in a Pentium IV computer (3.4 GHz processor; 1024 MB RAM) running the GNU/Linux operating system.

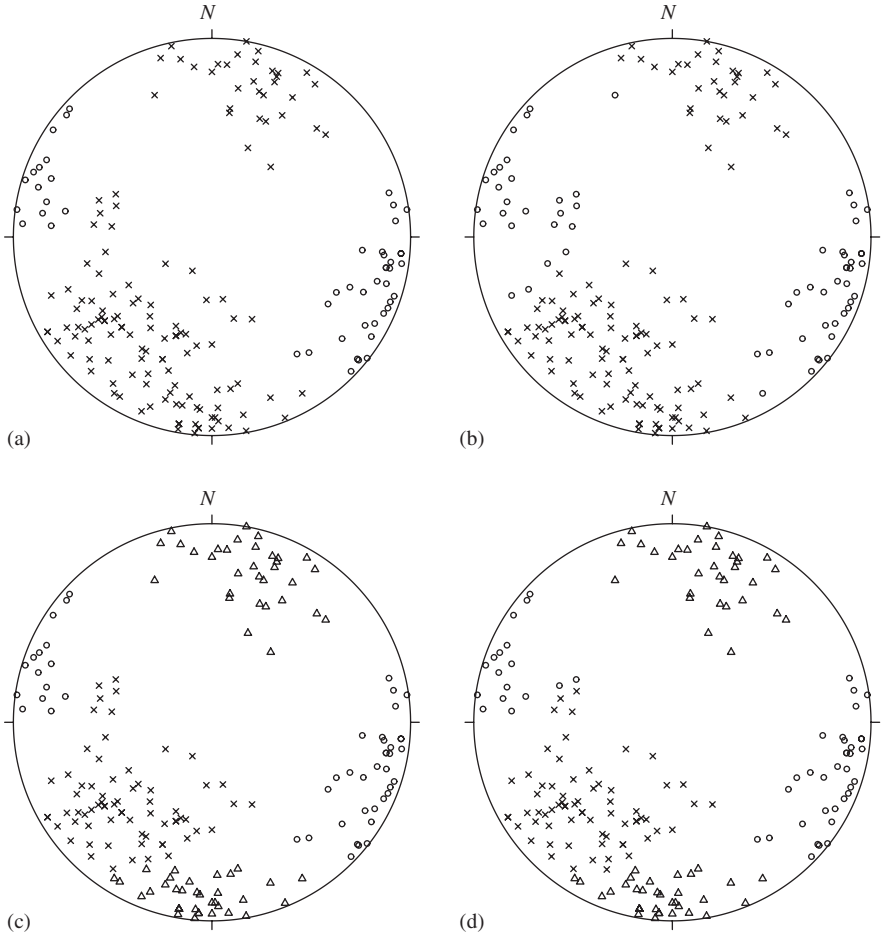

Fig. 8. Identified discontinuity sets for the serpentine rock mass data set. (a) Spectral clustering; $K = 2$; (b) Fuzzy K-means clustering; K = 2; (c) Spectral clustering; $K = 3$; and (d) Fuzzy K-means clustering; K = 3.

Table 2
Direction of average unit normal vectors for each discontinuity set (serpentine rock mass data set)

Set	Fuzzy K-means	Spectral clustering
(a) $K = 2$		
1 (○)	287/01	108/05
2 (×)	203/17	205/18
(b) $K = 3$		
1 (○)	108/04	108/05
2 (△)	225/37	227/37
3 (×)	008/03	009/03

normal vectors for each discontinuity set computed with both algorithms are very similar as well.

Fig. 9 shows the results of K-means clustering in the transformed space of the rows of U. For the case of two discontinuity sets ($K = 2$; see Fig. 9(a)) we show that observations in the cluster indicated with the circle (○) symbol are grouped close to the cluster average, suggesting a high degree of membership to that cluster. Similarly, note that memberships of several discontinuities assigned to the set labeled with the cross (×) symbol are not so certain. A similar discussion applies to the case of three discontinuity sets ($K = 3$; see Fig. 9(b)), where the separation between discontinuity sets labeled with the triangle (△) and cross (×) symbols is less crisp than their separation with respect to the discontinuity set labeled with circle (○) symbols.

Techniques for automatic estimation of the "optimal" number of partitions could be used to provide an estimate of the most adequate number of discontinuity sets in the rock mass in this case (see e.g., [11,16,26]). The automatic identification of the optimal number of partitions is, however, beyond the scope of this paper, and we do not discuss validity measures herein. The selection of the number of clusters should be done by an expert [3,11,18], depending on the purpose of the rock mass characterization study, on the geometry of the excavation, and on the engineering performance of rock mass models obtained after considering the different clustering alternatives.

5. Conclusions

We present a new approach based on spectral clustering for grouping of discontinuities and for identification of discontinuity sets based on their orientation. We demonstrate the use of the spectral clustering algorithm with a

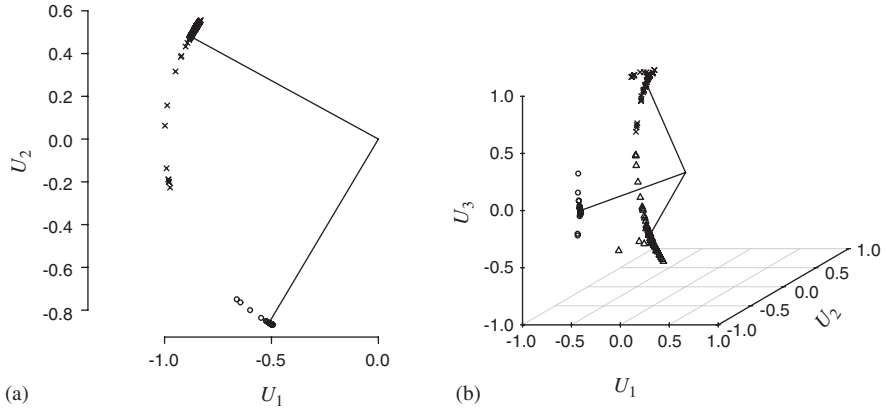

Fig. 9. K-means clustering in the transformed space (serpentine rock mass data set).

simple measure of similarity that is specific to the problem of clustering of rock discontinuity orientations, defined as the sine of the acute angle between the directions of discontinuity unit normal vectors.

The spectral clustering algorithm makes a transformation of the original discontinuity orientation data into a transformed space where clustering is performed. The advantage of such transformation is that it is easier to perform clustering in the transformed space as, for K discontinuity sets, points in the transformed space form tight clusters around K mutually orthogonal points that lie on the surface of the K-dimensional unit sphere. The coordinates of points in the transformed space are given by the normalized rows of a matrix constructed using the main eigenvectors of the normalized affinity matrix. The affinity matrix is defined in terms of the distance between observations in the data set and in terms of a scaling parameter σ^2. We show that values of the scaling parameter in the order of $\sigma = 0.12$ produce adequate clustering results in a number of test for the sine-based similarity measure considered.

The performance of the spectral clustering algorithm is studied using benchmark test cases based on real rock discontinuity data sets. The results show that the algorithm exhibits good clustering capabilities, even when a simple sine-based similarity measure is considered. In particular, the discontinuity set partitions computed with the spectral clustering algorithm are shown to be very similar to those computed with a number of clustering algorithms (and their related distance measures) that are commonly used in rock engineering applications. We also show that the average directions of discontinuity sets identified with the spectral clustering algorithm agree well with the average directions of discontinuity sets identified with such commonly used algorithms. Furthermore, we show an example case with data sets of discontinuity orientations compiled from the literature, in which the spectral clustering algorithm provides more natural partitions than the other algorithms considered. Additional advantages of the algorithm are that convergence is fast, and that it can be easily (and efficiently) implemented using popular software packages for numerical analysis such as R or MATLAB.

Acknowledgements

Christian Klose kindly provided a computer implementation of their vector quantization algorithm that was used during development of our own implementation. Financial support for this research was provided by the National Science Foundation (Grant CMS-0355079). The support of "Grupo de Investigaciones Medioambientales: Riesgos Geológicos e Ingeniería del Terreno" (RNM 221. PAI. Junta de Andalucía) of the University of Granada is gratefully acknowledged as well.

References

[1] Priest SD. Discontinuity analysis for rock engineering, 1st ed. London, New York: Chapman & Hall; 1993.
[2] Kulatilake P, Fiedler R, Panda BB. Box fractal dimension as a measure of statistical homogeneity of jointed rock masses. Eng Geol 1997;48(3–4):217–29.
[3] Kulatilake PHSW, Um JG, Wang M, Escandon RF, Narvaiz J. Stochastic fracture geometry modeling in 3-D including validations for a part of Arrowhead east tunnel, California, USA. Eng Geol 2003;70(1–2):131–55.
[4] Kulatilake P, Wathugala DN, Poulton M, Stephansson O. Analysis of structural homogeneity of rock masses. Eng Geol 1990;29(3): 195–211.
[5] Goodman RE. Block theory and its application. Geotechnique 1995;45(3):383–422.
[6] Starzec P, Andersson J. Probabilistic predictions regarding key blocks using stochastic discrete fracture networks—example from a rock cavern in south-east Sweden. Bull Eng Geol Environ 2002;61(4): 363–78.
[7] Martin MW, Tannant DD. A technique for identifying structural domain boundaries at the EKATI diamond mine. Eng Geol 2004; 74(3–4):247–64.
[8] Park HJ, West TR, Woo I. Probabilistic analysis of rock slope stability and random properties of discontinuity parameters, Interstate Highway 40, Western North Carolina, USA. Eng Geol 2005; 79(3–4):230–50.

[9] Billings MP. Structural geology. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall; 1972.
[10] Priest SD. Hemispherical projection methods in rock mechanics. London; Boston: Allen & Unwin; 1985.
[11] Harrison JP. Fuzzy objective functions applied to the analysis of discontinuity orientation data. In: Hudson JA, editor. Rock characterization: Proceedings of ISRM symposium, Eurock'92. British Geotechnical Society, London: Thomas Telford; 1992.
[12] Terzaghi RD. Sources of errors in joint surveys. Geotechnique 1965;15:287–304.
[13] Priest SD. The collection and analysis of discontinuity orientation data for engineering design with examples. In: Hudson JA, editor. Comprehensive rock engineering; principles, practice & projects: Rock testing and site characterization. Oxford: Pergamon Press; 1993. p. 167–92.
[14] Zhang L, Einstein HH. Estimating the mean trace length of rock discontinuities. Rock Mech Rock Eng 1998;31(4):217–35.
[15] Mahtab MA, Yegulalp TM. A rejection criterion for definition of clusters in orientation data. In: Goodman RE, Heuze FE, editors. Proceedings 22nd US symposium rock mechanics. Soc. Min. Eng. Am. Inst. Min. Metall. Petrol. Eng.; 1982. p. 116–23.
[16] Hammah RE, Curran JH. Fuzzy cluster algorithm for the automatic identification of joint sets. Int J Rock Mech Mining Sci 1998;35(7): 889–905.
[17] Herda HHW, Einstein HH, Dershowitz WS. Problems with representation of rock fracture clusters. J Geotech Eng 1991;117(11):1754–71.
[18] Klose C, Seo S, Obermayer K. A new clustering approach for partitioning directional data. Int J Rock Mech Mining Sci Geomech Abstracts 2005;42(2):315–21.
[19] Shanley RJ, Mahtab MA. Delineation and analysis of clusters in orientation data. J Math Geol 1976;8(1):9–23.
[20] Dershowitz W, Busse R, Geier J, Uchida M. A stochastic approach for fracture set definition. In: Aubertin M, Hassani F, Mitri S, editors. Proceedings of the second North American rock mechanics symposium; NARMS'96, a regional conference of ISRM. Rock mechanics tools and techniques, vol. 2. Rotterdam, Netherlands (NLD): A. A. Balkema; 1996. p. 1809–13.
[21] Marcotte D, Henry E. Automatic joint set clustering using a mixture of bivariate normal distributions. Int J Rock Mech Mining Sci 2002;39(3):323–34.
[22] Hammah RE, Curran JH. On distance measures for the fuzzy K-means algorithm for joint data. Rock Mech Rock Eng 1999;32(1):1–27.
[23] Zhou W, Maerz NH. Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes. Comput Geosci 2002;28(7):827–39.
[24] Sirat M, Talbot C. Application of artificial neural networks to fracture analysis at the Äspö HRL, Sweden: fracture sets classification. Intl J Rock Mech Mining Sci 2001;38:621–39.
[25] Henry E, Marcotte D, Kavanagh P. Classification of non-oriented fractures along boreholes to joint sets and its success degree. Rock Mech Rock Eng 2001;34(4):257–73.
[26] Hammah RE, Curran JH. Validity measures for the fuzzy cluster analysis of orientations. IEEE Trans Pattern Anal Mach Intell 2000; 22(12):1467–72.
[27] Klose C. A new objective approach for fracture grouping—theory and advantages. In: Schubert W, editor. Rock engineering theory and practice, ISRM regional symposium Eurock 2004 & 53rd geomechanics colloquium. VGE, 2004. Also available at ⟨http://www.easypublic.com/pages/publications/Klose2004b.pdf⟩.
[28] Weiss Y. Segmentation using eigenvectors: a unifying view. In: Proceedings of the seventh IEEE international conference on computer vision, vol. 2. 1999. p. 975–82.
[29] Ng AY, Jordan M, Weiss Y. On spectral clustering: analysis and an algorithm. In: Dietterich TG, Becker S, Ghahramani Z, editors. Advances in neural information processing systems, vol. 14. Cambridge, MA: MIT Press; 2002. p. 849–56.
[30] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 2000;22(8):888–905.
[31] Ebblin C, Zille A, Rossi G. The running-cone method for the interpretation of conical fold geometries: an example from the Badia Valley, Northern Dolomites (NE Italy). J Struct Geol 2005;27(1): 139–44.
[32] Kent JT. The Fisher–Bingham distribution on the sphere. J R Statist Soc Ser B (Methodological) 1982;44(1):71–80.
[33] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2004. ⟨http://www.R-project.org⟩.
[34] Priest SD, Hudson JA. Estimation of discontinuity spacing and trace length using scanline surveys. Int J Rock Mech Mining Sci Geomech Abstracts 1981;18(3):183–97.
[35] Ahlgren S. GeoPlot: plotting and analysis software for microsoft windows. ⟨http://www.geo.arizona.edu/coulomb/pages/geoplot/index.html⟩; 2000.