6th European Conference on Computational Mechanics (ECCM 6)
7th European Conference on Computational Fluid Dynamics (ECFD 7)
1115 June 2018, Glasgow, UK

CHARACTERISING SPHERICAL PACKINGS BY PRINCIPAL COMPONENT ANALYSIS

Tingting Zhao[1] and Y.T. Feng[2]

[1] Zienkiewicz Centre for Computational Engineering, Swansea University
SA1 8EN
842758@swansea.ac.uk

[2]Zienkiewicz Centre for Computational Engineering, Swansea University
SA1 8EN
y.feng@swansea.ac.uk

Key words: Particle packing, Digitalised image, Principal component analysis, Principal variance, Spatial homogeneity and isotropy

Abstract. Particle packings play an important role in the discrete element modelling of particulate systems as different packings can lead to different physical behaviour, and therefore need to be properly characterised and controlled. Apart from a few conventional approaches, there is still a lack of more general, comprehensive and quantitative approaches that can reveal some fundamental features of packings. The current work attempts to develop a novel packing characterising system based on two techniques: digitalised image representation of a packing and subsequent application of Principal Component Analysis to the resulting image. It will prove that the principal components or variances of a packing image can indeed qualify as the signature of the packing, and therefore can be utilised to characterise the packing. Furthermore, a dissimilarity coefficient or a similarity index will be defined which provides a single valued metric to quantitatively compare two packings. Comprehensive investigations for several groups of purposefully generated particle packings are conducted to understand relationships of their principal variances with packing features. The difference between two packings with different features can be revealed by the principal variance (PV) and dissimilarity coefficient (DC).

1 INTRODUCTION

The particle packing plays a fundamental role in leading the physical behaviour of a particle system. Therefore it is of great importance to conduct the spatial-statistical analysis of the geometrical structure of the system. As the topology of the system is highly complex, it is difficult to observe the way particles packed around each other by experiments [1]. With the development of various particle based methods, such as Molecular Dynamics (MD) and the Discrete Element Method [2, 3, 4], more detailed information on the internal structure of particulate systems can be accessed. Currently,

the techniques applied to investigate the features of particle packings mainly focus on the packing density and orientations of the particle contacts.

The informations of a packing can be clarified into two categories: particle related and contact related. The classical characteristics of the particles are particle size distribution and packing density. These informations cannot reflect the spatial feature of the packing. Some researchers attempted to treat this issue using a radial function [5] while it is less effective. The features of the contacts can be checked by coordination number, contact force distribution and the contact orientation distribution[6]. These informations can only be achieved for the dense packing with contacts between particles. Summarily, there is still a lack of more general, comprehensive and quantitative approaches that can reveal some fundamental features of packing.

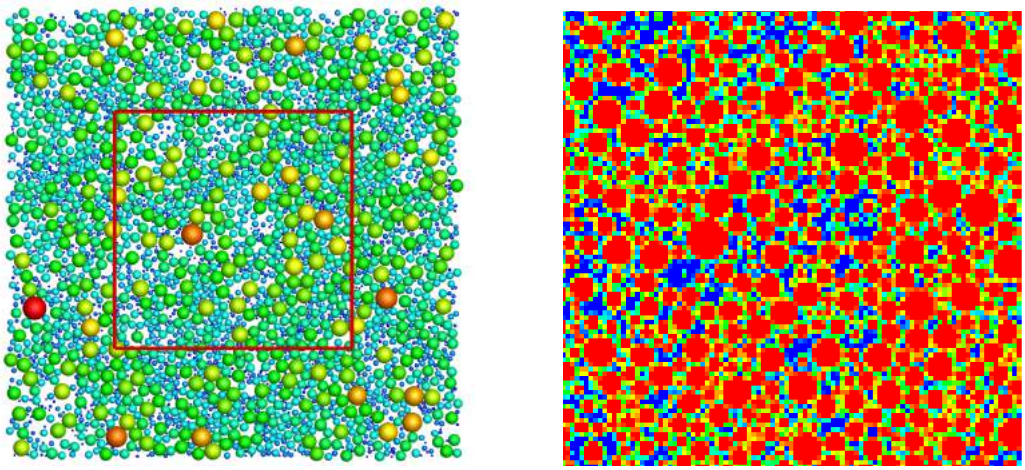
A particle packing can be viewed as a specific spatial variation of solid material (assigned a value 1) and void (a value of 0). The packing can be represented as a digitalised image with grey-scale ranging from 0 to 1 on a regular grid. Then the resulting packing image can be analysed by techniques developed in the field of computer graphics and imaging process [7, 8, 9, 10], and the (dis)similarity of multiple packing images may also be compared in a quantitative manner. Principal component analysis (PCA) [11, 12] is one of the most popular linear transform based statistical techniques, and has been extensively used in a wide variety of applications. It has proved to be a powerful tool that is often employed for data analysis in general, and dimension reduction in multi-variance analysis, and pattern recognition in signal and imaging processing in particular.

Therefore, a novel characterising method is developed using the principal component of a packing image as the signature of the packing. Furthermore, a dissimilarity coefficient is defined which can provide a single valued metric to quantitatively compare two packings. This has been made possible by the definition of the principal variance function that maps the ordinal numbers of individual principal variances into a non-dimensional unit interval. The characterisation of the particle packing can be revealed by the subsequent principal component analysis.

The paper is organised as follows. The whole analysis procedure of applying PCA to a spherical particle packing is introduced in Section 2, including the packing digitalisazeion, the formation of packing image, the subsequent numerical treatment of the image matrix and the resulted characterisation of particle packing revealed by the principal variance function and the dissimilarity coefficient. In Section 3, this packing characterising system is applied to several groups of purposefully generated random packings. The PCA analysis is conducted to the random packings. Then comprehensive investigations are carried out to illustrate the relationship between packing features and results obtained by PCA. The result of principal component analysis can evaluate the different effects on packings caused by configuration randomness of the packings. Concluding remarks are made in Section 4.

2 NUMERICAL PROCEDURE OF PCA

This section describe the full numerical procedure of the Principal Component Analysis and the following techniques which are used to characterise the particle packings. As only the main principals of PCA are adopted in the current work, some minor modifications



(a) a packing with a square analysis window

(b) a 100×100 digitalised image

Figure 1: A random particle packing and a digital representation within the analysis window

are made and different terminologies are used.

2.1 Packing digitalisation

Firstly, an analysis window $\mathcal{A}$ of rectangular shape is specified in the domain of the particle system. This window can be divided into $M \times N$ square cells with same spacing h. The grid cell is denoted as $\mathcal{A}_{ij}$ according to the index along the X and Y directions. The average area covered by particles is

$$a_{ij} = \frac{|\Omega_g \cap \mathcal{A}_{ij}|}{|\mathcal{A}_{ij}|} \tag{1}$$

where $|\Omega|$ denotes the measure or area of the particles; and $|\mathcal{A}_{ij}| = h^2$. An empty cell with no overlapping with any particle has $a_{ij} = 0$; while a cell fully covered by a particle has $a_{ij} = 1$. A cell partially covered by particles has $a_{ij} < 1$. So in general $a_{ij} \in [0, 1]$.

2.2 Formation of packing image

The collection of all the cell average areas a_{ij} forms an $M \times N$ matrix $\mathbf{A}_h = \{a_{ij}\}$, which can be viewed as a digitalised grey-scale representation of the original packing Ω_p in the region $\mathcal{A}$, thus is termed as the *packing matrix* or *image*. Figure 1(b) illustrates such a digitalisation of a disc particle packing shown in Figure 1(a). Clearly, $\mathbf{A}_h$ is accurate within particles or void space of the packing, but may introduce approximation around particle boundaries. The accuracy of this representation for the original packing depends on the grid spacing h, and will be accurate in the limit case:

$$\lim_{h \to 0} \mathbf{A}_h = \Omega_g \cap \mathcal{A} \tag{2}$$

2.3 Numerical procedures and formulations

Let m_j be the mean value of the j-th column of the packing matrix $\mathbf{A}_h$

$$m_j = \frac{1}{M}\sum_{i=1}^{M} a_{ij} \tag{3}$$

By subtraction of its mean from each column vector of $\mathbf{A}_h$, the column centralised matrix $\bar{\mathbf{A}}_N$ of $\mathbf{A}_h$ is obtained as:

$$\bar{\mathbf{A}}_N = \mathbf{A}_h - \mathbf{e}_{_M}\mathbf{m}_N \tag{4}$$

where $\mathbf{e}_{_M}$ is an $M \times 1$ column vector with all its elements being 1's; and $\mathbf{m}_N$ is the $1 \times N$ mean value vector $\mathbf{m}_N = \{m_j\}$.

Define the covariance matrix of $\bar{\mathbf{A}}_N$ as

$$\mathbf{S}_N = \frac{1}{M}\bar{\mathbf{A}}_N^{^T}\bar{\mathbf{A}}_N \tag{5}$$

where $\mathbf{S}_N$ is a $N \times N$ square matrix. Notice in the above that M instead of $M-1$ is used. Further define the column-wise total variance as

$$\sigma_{_N}^c = \frac{1}{N}\,\mathrm{Tr}(\mathbf{S}_N) = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{S}_N)_{ii} \tag{6}$$

which may be (slightly) different from the total variance σ_h in general.

By solving the eigenvalue problem of $\mathbf{S}_N$, it yields the following matrix decomposition

$$\mathbf{S}_N\mathbf{V}_N = \mathbf{V}_N\mathbf{D}_N \tag{7}$$

with

$$\mathbf{D}_N = \mathbf{V}_N^T\mathbf{S}_N\mathbf{V}_N, \qquad \mathbf{V}_N^T\mathbf{V}_N = \mathbf{I}_N$$

where the diagonal matrix $\mathbf{D}_N = \mathrm{diag}\{d_i\}$ contains all the eigenvalues d_i in descending order, which are termed the *principal variances* (PVs); and $\mathbf{V}_N = \{\mathbf{v}_i\}$ are the orthonormal vectors, termed the *principal modes*. As $\bar{\mathbf{A}}_N$ is column centralised, $\mathbf{S}_N$ is a semi-positive definite matrix with at least one zero principal variance. It is also well known that the sum of the PVs and the total column-wise variance is related by

$$\frac{1}{N}\sum_{i=1}^{N} d_i = \sigma_N^c \tag{8}$$

$\mathbf{S}_N$ can be recovered from the principal variances and modes as

$$\mathbf{S}_N = \mathbf{V}_N\mathbf{D}_N\mathbf{V}_N^T = \sum_{i=1}^{N-1} d_i\mathbf{v}_i\mathbf{v}_i^T \tag{9}$$

In many applications, only the first few principal variances are needed to approximate $\mathbf{S}_N$ to a reasonable degree, thereby significantly reducing the dimension of the problem concerned. This is often the main objection of PCA, but not an issue for the current problem.

Further define the projection $\mathbf{U}_N$ of $\bar{\mathbf{A}}_N$ onto the space spanned by $\mathbf{V}_N$ as

$$\mathbf{U}_N = \bar{\mathbf{A}}_N \mathbf{V}_N \tag{10}$$

Then it has

$$\mathbf{D}_N = \frac{1}{N} \mathbf{U}_N^T \mathbf{U}_N \tag{11}$$

and $\bar{\mathbf{A}}_N$ and $\mathbf{A}_h$ can be recovered by

$$\bar{\mathbf{A}}_N = \mathbf{U}_N \mathbf{V}_N^T; \quad \mathbf{A}_h = \bar{\mathbf{A}}_N + \mathbf{e}_M \mathbf{m}_N \tag{12}$$

Similarly to $\mathbf{S}_N$, $\bar{\mathbf{A}}_N$ or $\mathbf{A}_h$ can be optimally approximated by the leading principal variances.

The column-wise total variance $\sigma_{_N}^c$, the mean value vector $\mathbf{m}_N$, the principal variance matrix $\mathbf{D}_N$ and the corresponding modes $\mathbf{V}_N$ form a unique set $\mathcal{S}_\mathcal{N}$, termed the column-wise *characteristic set*, that fully determines the packing in the vertical direction

$$\mathcal{C}_\mathcal{N} = \{\sigma_{_N}^c, \mathbf{m}_N, \mathbf{D}_N, \mathbf{V}_N\} \tag{13}$$

As the PVs and the column-wise total variance is related by (8), and the total variance (and also the column-wise variance) is related to the density, the PVs play a dominant role to characterise a packing image and therefore can be viewed as the (column-wise) *signature* of the packing.

2.4 Principal variance function

To facilitate the comparison between different sets of principal variances, particularly when they are obtained from different resolutions P, the ordinal number i of a principal variance d_i is mapped from 1 to P to a non-dimensional "position" variable $x \in [0, 1]$ by

$$x(i) = \frac{1}{P}\Big(i - \frac{1}{2}\Big) \tag{14}$$

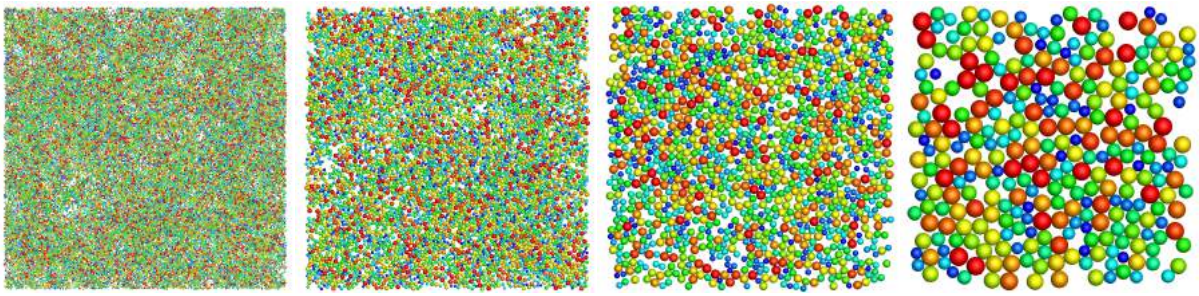
Then a continuous function $\mathsf{d}(x), x \in [0, 1]$, termed the *principal variance function*, can be constructed to interpolate the discrete variances d_i using piecewise linear or higher order interpolation functions such that

$$\mathsf{d}(x_i) = d_i, \quad x_i = x(i), \quad i = 1, \cdots, P \tag{15}$$

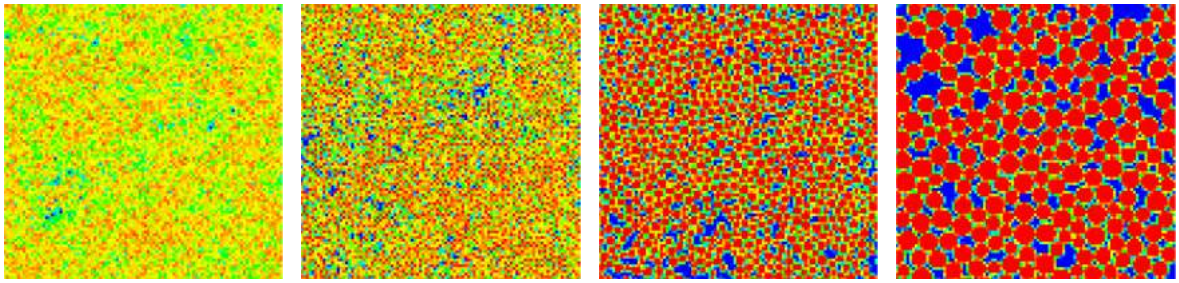
2.5 Packing image similarity

Consider two packing images with their principal variance functions $\mathsf{d}_1(x)$ and $\mathsf{d}_2(x)$ obtained, define a so-called *dissimilarity coefficient* (DC) between these two images as

$$\mathcal{D}_c = \Big[\frac{1}{\Sigma_{_1} + \Sigma_{_2}} \int_0^1 [\mathsf{d}_1(x) - \mathsf{d}_2(x)]^2 d\,x\Big]^{1/2} \in [0, 1] \tag{16}$$



(a) Packing U1 (b) Packing U2 (c) Packing U4 (d) Packing U8



(e) Image U1(N) (f) Image U2(N) (g) Image U4(N) (h) Image U8(N)

Figure 2: Four uniform packing groups in region [-0.1, 1.1] × [-0.1, 1.1] (a-d); and their digital images (with N=100) within region [0, 1] × [0, 1] (e-h)

where Σ_1 and Σ_2 are defined as

$$\Sigma_i = \int_0^1 \mathsf{d}_i^2(x)\, dx \quad (i = 1, 2)$$

Consequently, the degree of similarity of these two packing images can be quantified by the *similarity index* $\in [0, 100]$ defined as

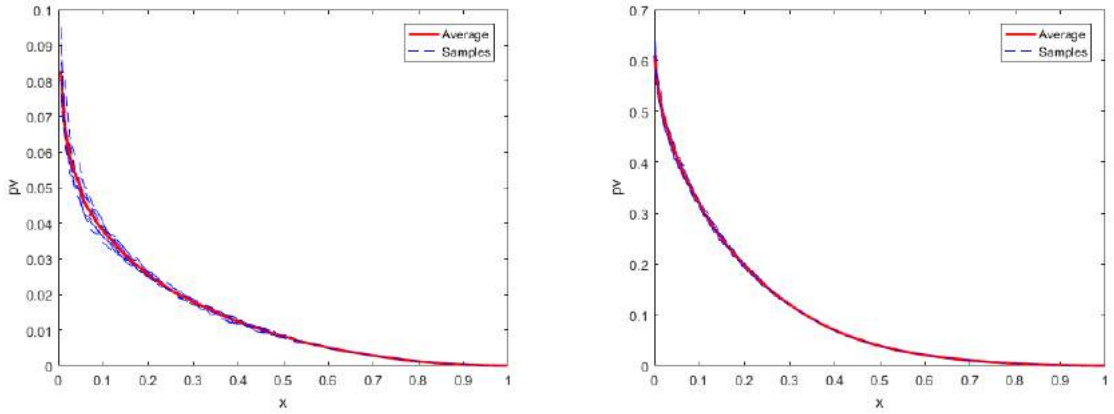
$$\mathcal{S}_{\mathcal{I}} = (1 - \mathcal{D}_c) \times 100 \tag{17}$$

3 PACKING CHARACTERISATION

This section is devoted to illustrating how the principal variance function and dissimilarity coefficient defined in the previous section can be applied to quantitatively characterise the features of different packings.

Four groups of random particle packings are generated within the domain $[-0.1, 1.1] \times [-0.1, 1.1]$ with the periodic condition applied to both directions. All groups have particle sizes obeying uniform distribution within a range. Each group has the particle size range doubled from the previous group. The size distribution of groups U2, U4, U8 are respectively 2, 4 and 8 times of the base group U1. 10 random packing samples with the same size distribution are generated in each group. The packings and their images at $N = 100$ are displayed in Figure 2.

For each packing group, the PVs of all 10 samples are computed and their averages are taken to be the PVs of the group. For illustrative purpose, the principal variance



(a) U1(100) (b) U1(400)

Figure 3: Principal variance of group U1

Table 1: Average dissimilarity coefficient of each group of random packings

N	U1	U2	U4	U8
400	0.0126	0.0215	0.0372	0.0602
100	0.0378	0.0379	0.0465	0.0662

functions of the 10 samples for group U1 at two resolutions N=100 and 400 are displayed in Figure 3. Clearly the PV functions of the 10 samples at each set are located within a narrow band around the group mean value. The maximum difference appears at the leading variances but is much reduced for smaller PVs. This indicates that these samples randomly generated from the same distribution indeed have very similar statistical features.

To quantify the difference, the dissimilarity coefficient of 10 samples in each group are calculated based on the formula (16) against their group average for two resolutions: N=100 and 400. The average dissimilarity coefficients of 10 samples in each group for the three resolutions are provided in Table 1. As Figure 3 visually shows a minor difference among PVs functions of samples in each group, which is confirmed by very small dissimilarity coefficients in Table 1, it can be concluded that the effects of the particle distribution randomness is indeed insignificant. It has also been found that the difference decreases when the number of particles in the packing increases.

Besides reflecting the influence of the configuration randomness of the particle system, Principal Component Analysis can also indicate the different level of effects on packing caused by particle distribution, packing density and particle size. The uniformity and isotropy of a packing can also been investigated by this PCA based approach.

4 CONCLUSIONS

- The current has developed a novel packing characterising system based on two techniques: digitalised image representation of a packing and subsequent application of Principal Component Analysis to the resulting image. It has proved that the principal components or variances of a packing image can indeed qualify as the signature of the packing, and therefore can be utilised to characterise the packing. Furthermore, a dissimilarity coefficient or a similarity index can be defined which provides a single valued metric to quantitatively compare two packings.

- The values of PVs and DC can indicate different levels of effects on packing caused by configuration randomness, particle distribution, packing density and particle size. The uniformity and isotropy of a packing can also been investigated by this PCA based approach.

- The methodology developed can be extended to both 3D cases and non-spherical particle packings, and can also be applicable to some other problems in particle systems

REFERENCES

[1] Majmudar, Trushant S and Behringer, Robert P. Contact force measurements and stress-induced anisotropy in granular materials. *Nature.* 435(7045): 1079-1082, 2005.

[2] P. A. Cundall, and O. D. L. Strack. A discrete numerical model for granular assemblies. *Geotechnique*, 29(1):47-65, 1979.

[3] Y. T. Feng, K. Han, D. R. J. Owen. Filling domains with disks: An advancing front approach. *International Journal For Numerical Methods in Engineering.* 56: 699-713, 2003.

[4] K. Han, Y. T. Feng, D. R. J. Owen. Sphere packing with a geometric based compression algorithm. *Powder Technology*, 155(1):33-41, 2005.

[5] Lochmann, Kristin and Oger, Luc and Stoyan, Dietrich. Statistical analysis of random sphere packings with variable radius distribution. *Solid State Sciences*, 8(12):1397-1413, 2006.

[6] Cambou, B and Dubujet, Ph and Nouguier-Lehon, C. Anisotropy in granular materials at different scales. *Mechanics of materials*, 36(12):1185-1194, 2004.

[7] H. H. Barret. Foundations of Image Science. 3rd Edition, John Wiley & Sons, New Jersey, UK, 2004.

[8] R. C. Gonzales and R. E. Woods. Digital Image Processing. 2nd Edition, Prentice Hall, 2002.

[9] S. Haykin. Neural Networks, A Comprehensive Foundation. IEEE Press, USA, 1984.

[10] M. Petrou and P. Bosddogianni. Image Processing: The Fundamentals. John Wiley & Sons, UK, 2000.

[11] I. T. Jolliffe. Principal Component Analysis, Springer Series in Statistics, 2nd ed., Springer, NY, 2002

[12] H. Abdi and L. J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics.* 2 (4): 433459, 2010.