

MANCHESTER 1824
The University of Manchester

The University of Manchester Research

Circular Object Arrangement using Spherical Embeddings

DOI:
10.1016/j.patcog.2019.107192

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
Evangelopoulos, X., Brockmeier, A. J., Mu, T., & Goulermas, J. Y. (2020). Circular Object Arrangement using Spherical Embeddings. *Pattern Recognition*, 107192. https://doi.org/10.1016/j.patcog.2019.107192

Published in:
Pattern Recognition

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

OPEN ACCESS

Download date:29. Mar. 2024

Circular Object Arrangement using Spherical Embeddings

Xenophon Evangelopoulos[a], Austin J. Brockmeier[b],
Tingting Mu[c], John Y. Goulermas[a,*]

[a]*Department of Computer Science, University of Liverpool, Liverpool, UK*
[b]*Department of Electrical and Computer Engineering, University of Delaware, USA*
[c]*School of Computer Science, University of Manchester, Manchester, UK*

Abstract

We consider the problem of recovering a circular arrangement of data instances with respect to some proximity measure, such that nearby instances are more similar. Applications of this problem, also referred to as circular seriation, can be found in various disciplines such as genome sequencing, data visualization and exploratory data analysis. Circular seriation can be expressed as a quadratic assignment problem, which is in general an intractable problem. Spectral-based approaches can be used to find approximate solutions, but are shown to perform well only for a specific class of data matrices. We propose a bilevel optimization framework where we employ a spherical embedding approach together with a spectral method for circular ordering in order to recover circular arrangements of the embedded data. Experiments on real and synthetic datasets demonstrate the competitive performance of the proposed method.

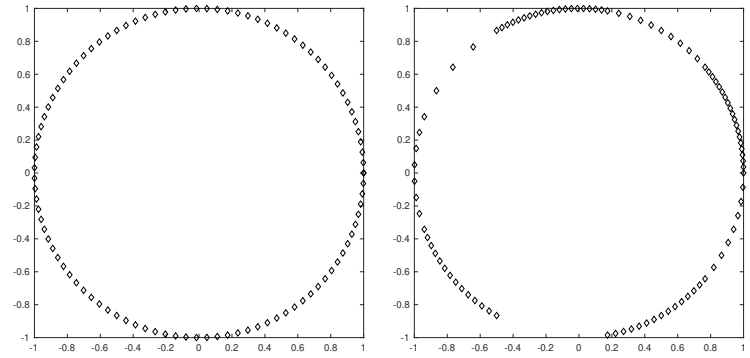
Keywords:
combinatorial data analysis, data sequencing, circular seriation, quadratic assignment problem, spherical embeddings

*Corresponding author:
Email address: j.y.goulermas@liverpool.ac.uk (John Y. Goulermas)

1. Introduction

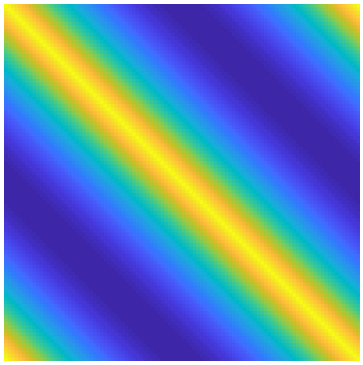
Seriation is an exploratory combinatorial data analysis technique used for the identification of trends of gradually varying similarities between patterns. It operates by reordering a set of arbitrary patterns so that similar ones are placed proximately while dissimilar ones further apart. In essence, seriation aims at inferring an ordering (permutation) consistent with an underlying linear order of the data. Different seriation algorithms work by simultaneously interchanging the rows and columns of a similarity or dissimilarity matrix in order to optimize a specific objective. Originating from the field of archeology where it was used to infer the chronological order of a set of graves based on the artifacts recovered from them [1], seriation has found application in several areas such as sociology and psychology [2], gene sequencing and bioinformatics [3], and exploratory data visualization [4] to identify global patterns (e.g., the number or tendency of clusters). In this context, it has been put to practice to reveal patterns in microarray data [5], and to arrange words or documents in text mining based on their co-occurrence statistics in order to track the flow of conversations [6]. More recently, [7] developed mechanisms for comparing and fusing generated orderings, [8] introduced the notion of robust seriation along with various modeling formulations and solution procedures for that problem, and [9] proposed a set of scalable approximate methods for seriation that can highlight global or local similarity pattern structure. A systematic experimental analysis of seriation methods and measures can be found in [10].

In many cases however, the data objects may be arranged around a closed continuum yielding a rather circular underlying order. This often occurs in disciplines where the data instances are characterized by a cyclic procedure, e.g., in biology where particular cells can have a cyclic evolution procedure [11, 12], or in tomographic reconstruction of a planar object based on its projections taken at various angles in $[0, 2\pi]$ [13]. In such situations, we are interested in recovering object orderings around a closed continuum that maintain the seriation effect, placing similar objects closely. In a matrix representation, this can be visualized as a symmetric matrix of pairwise similarities between objects where elements of each row/column decrease monotonically while

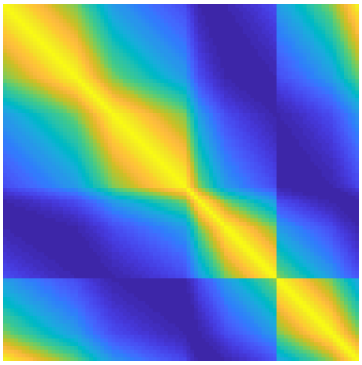


(a) Equispaced points

(b) Randomly allocated points



(c) Circulant-CR map of (1a)



(d) General CR map of (1b)

Figure 1: Illustration of two different types of CR similarity matrices generated by distances between pairs of points residing on a circle.

moving to the right/bottom until some specific element and then increase again monotonically until the end of each row/column and fold back from the left/top of the matrix. Matrices of this form are called circular-Robinsonian (CR) [14]. Two different examples of CR similarity matrices are illustrated in Figure 1 that highlight their structural differences according to different object arrangement on the circle. Higher intensities in the maps correspond to higher similarities between objects. As in the seriation problem, combinatorial optimization procedures or heuristics are sought here to simultaneously interchange rows and columns of such matrices in order to bring them in a CR form.

The aforementioned combinatorial problem, also referred to as *circular seriation* [15, 16] can be modeled as an instance of the quadratic assignment problem (QAP) [17], where a circulant-CR dissimilarity matrix is involved to represent absolute positional differences of n data objects arranged evenly on a circle. As a QAP instance, circular seriation is also an NP-hard combinatorial problem with $(n-1)!$ possible discrete solutions corresponding to different permutations [18]. Solving this problem to optimality can prove impractical when the problem size becomes large. In the ideal and infrequent case where the data can be rearranged to form a circulant-CR similarity matrix, an optimal solution can be identified in polynomial time by sorting the patterns according to the order of the inverse tangent of the angle between the first two non-trivial eigenvectors of a normalized Laplacian [13, 16]. In any other case however, this spectral solution is only guaranteed to approximately minimize the circular seriation problem and therefore alternative approaches are desirable.

Since exact methods such as branch-and-bound [19] and dynamic programming approaches [20] can only be practical for QAP instances of small sizes, suboptimal algorithms and heuristics that maintain good running time performance are often needed. Some of them include improvement methods, such as local search, tabu search [21], simulation approaches such as simulated annealing, and relaxation-based algorithms in the context of graph matching [22, 23, 24]. Nevertheless, such methods are restricted to act explicitly on the given set of similarities by only optimizing the corresponding QAP objective and may thus fail to correctly capture the underlying order of the data. To overcome this limitation, we propose a bilevel optimization framework that employs a spherical embedding approach along with a spectral method for circular ordering to recover a circular arrangement of the embedded representations. This allows for the generation of flexible embedded data points arranged in a circular way that conforms with their original similarities. The resulting ordering yields better circular seriation performance than most compared approaches in terms of various measures.

The rest of the paper is organized as follows. In Section 2, we present recent developments in the field and the current state-of-the-art algorithms. In Section 3, we give a detailed description of the proposed algorithm providing essential information on various optimization-related aspects. Section 4 contains detailed experimental evaluations

and comparisons along with relevant analyses, while Section 5 concludes the present work.

2. Related work

The problem of arranging objects around a circle was first discussed in [25] and [15]. Towards this direction, circular unidimensional scaling was developed in [26] where reconstructed distances between object pairs had to represent the original pattern proximities as closely as possible. In the same context, [14] and [18] introduced the notions of CR and strongly CR matrices, respectively, in an effort to model a more robust fitting of data objects along a closed continuum.

The problem of embedding data in lower dimensional spaces while preserving their original characteristics and proximities has been very popular over the last decades for exploratory data analysis and feature generation tasks. Laplacian eigenmaps (LE) [27], Multidimensional scaling (MDS) [28] and Locally Linear Embedding (LLE) [29] are a few typical methods that deal with this problem by exploiting the structure of the original data instances using local neighborhood graphs. An overview of various types of dimensionality reduction methods can be found in [30]. Also, other works operate on pairwise relations of heterogeneous data types to generate low-dimensional representations on a joint common space [31]. Many of the aforementioned approaches have been extended to map proximities on the sphere [32] to suit applications where the data lie on spherical manifolds. In the more recent work of [33], the authors propose to isometrically embed dissimilarities between objects in spherical and hyperbolic spaces.

Recently, several methods have employed embedding-based algorithms for arranging data objects around a close continuum. These include the work in [34] which recovers a circular ordering by sorting the inverse tangent of the angle between the two main eigenvectors of a correlation matrix, and the work in [13] that also relies on the angular position between the first two non-trivial eigenvectors resulting from a Laplacian embedding problem. In a similar context, [35] propose the rank-two ellipse seriation method based on iterative application of correlations of the original dissimilarities to converge directly to a two-dimensional subspace. More recently, [36] tackle

the problem of recovering circular orderings of cell-cycle generated data by also employing MDS and by sorting the angle between the first two embedding coordinates, while [16] propose a unifying framework for capturing the order of general filamentary structures by identifying local data neighborhoods in a lower dimensional space using suitable spectral methods for linear or circular seriation.

In this work we propose a circular seriation method where data instances are embedded in a low-dimensional hypersphere and ordered using a spectral method operating within a bilevel optimization framework.

3. Proposed methodology

3.1. Preliminaries

Let $\boldsymbol{\pi}$ denote a permutation that corresponds to a mapping (treated also for convenience as a vector) of the set $[n] = \{1, 2, \ldots, n\}$ onto itself. We use the convention that a permutation presents a list of objects, and this implies that the object with label π_i is assigned position i. This can also be represented by an $n \times n$ matrix $\mathbf{\Pi}$ from the set of $n!$ permutation matrices $\mathcal{M}^n$, with elements defined by

$$\Pi_{ij} = \begin{cases} 1, & \text{if } \pi_i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The above allows the conversion $\mathbf{\Pi} \boldsymbol{e} = \boldsymbol{\pi}$ between the two structures, where $\boldsymbol{e} = (1, 2, ..., n)^\top$ is the identity permutation.

Combinatorial problems that involve the optimal arrangement of objects can be modeled via objective functions parametrized by permutation vectors or permutation matrices. In particular, the previously mentioned QAP describes models that are quadratic with respect to a permutation matrix, and is expressed as

$$\text{QAP}(\mathbf{A}, \mathbf{B}) \triangleq \text{tr}\Big[\mathbf{\Pi A \Pi}^\top \mathbf{B}^\top\Big] = \sum_{i,j=1}^{n} A_{\pi_i \pi_j} B_{ij}, \tag{2}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the problem parameter matrices.

For circular seriation we are interested in QAP instances with $\mathbf{A}$ being a symmetric data-dependent matrix that encapsulates the pairwise similarities between n objects

and $\mathbf{B}$ a circulant-CR dissimilarity matrix with elements [15]

$$B_{ij} = \begin{cases} |i-j|, & \text{if } |i-j| \leq \lfloor \frac{n}{2} \rfloor, \\ n - |i-j|, & \text{if } |i-j| > \lfloor \frac{n}{2} \rfloor. \end{cases} \tag{3}$$

The above acts as a circular seriation template where the elements of the first row (column) increase monotonically while moving to the right (bottom) until the $\lfloor \frac{n}{2} \rfloor$-th element and then decrease again monotonically until the end of the row (column), and fold back from the left (top) of the matrix. Note that the turning point can also be defined to be any index within $(1, n)$ for each row (column), producing different circular seriation effects [37], but throughout this work for simplicity we use the model of Eq. (3).

In the ideal case where $\mathbf{A}$ can be rearranged to a circulant-CR matrix, it has been shown [16] that the spectral-based approach proposed by Coifman [13] can solve the circular seriation problem and provide approximate solutions for general CR matrices. In detail, the method initially solves the following Laplacian embedding problem

$$\begin{aligned} \underset{\mathbf{Z} \in \mathbb{R}^{n \times k}}{\arg\min} \quad & \operatorname{tr}\left[\mathbf{Z}^\top \mathbf{L}_{\mathbf{A}}^{\text{rw}} \mathbf{Z}\right] \\ \text{s.t.} \quad & \mathbf{Z}^\top \mathbf{Z} = \mathbf{I}_k, \end{aligned} \tag{4}$$

where $\mathbf{Z}$ represents an embedding matrix with each row $\{\mathbf{z}_i\}_{i=1}^n$ corresponding to the coordinates of the i-th embedded data point, and $\mathbf{L}_{\mathbf{A}}^{\text{rw}} = \mathbf{I}_n - \operatorname{diag}\left(\mathbf{A}\mathbf{1}_n\right)^{-1}\mathbf{A}$ with $\mathbf{1}_n$ being the constant vector of all ones; the $\operatorname{diag}(\cdot)$ operator contextually forms either a vector from the diagonal of a matrix or a diagonal matrix from a vector. $\mathbf{L}^{\text{rw}}$ is a normalized Laplacian motivated by its close approximation of the Laplace-Beltrami operator on curves even when the data is distributed nonuniformly [13]. Due to the orthogonality constraint $\mathbf{Z}^\top \mathbf{Z} = \mathbf{I}_k$ Eq. (4) can be solved with an eigen-decomposition. The final output is a circular ordering obtained from sorting the inverse tangent of the angle between the first two non-trivial eigenvectors of $\mathbf{L}_{\mathbf{A}}^{\text{rw}}$. The main steps of this spectral approach are described in Algorithm 1.

Nevertheless, in most cases where the data matrix is far from an ideal circulant-CR form, the previous spectral approach is not guaranteed to solve circular seriation and in practice can have poor performance [16]. Alternatively, QAP solvers, such as

Algorithm 1 Spectral Method for Circular Seriation

Input: data points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^k$

Output: permutation matrix $\mathbf{\Pi} \in \mathcal{M}^n$

1: **function** CircularSeriation($\mathbf{X}$)
2: Construct weight matrix $\mathbf{W}$ from the pairwise distances of $\{\mathbf{x}_i\}_{i=1}^n$
3: Set $\mathbf{G} = \operatorname{diag}\left(\mathbf{W}\mathbf{1}_n\right)^{-1}\mathbf{W}\operatorname{diag}\left(\mathbf{W}\mathbf{1}_n\right)^{-1}$
4: Compute normalized Laplacian $\mathbf{L}_{\mathbf{G}}^{\text{rw}}$ and its first two non-trivial eigenvectors $\mathbf{v}_2, \mathbf{v}_3$
5: Calculate $\boldsymbol{\xi}$: $\xi(i) = \tan^{-1}(\mathbf{v}_3(i)/\mathbf{v}_2(i)) + \mathbb{1}[\mathbf{v}_2(i) < 0]\pi$
6: **return** $\mathbf{\Pi}$ that sorts $\boldsymbol{\xi}$ in ascending order

simulated annealing, tabu search and evolutionary approaches [21], can be employed but these can be sensitive to noise in the data or absence of strong circular trends in the patterns as they are restricted to operate on the given similarities that may not fit perfectly the assumed template. The idea here, is that instead of working solely with the set of original similarities to optimize explicitly the objective of Eq. (2), we use a hybrid model that additionally seeks more flexible low-dimensional pattern representations embedded on a $(k-1)$-dimensional sphere. The motivation for spherical embeddings relies on the observation that when the underlying order of the data is circular these may lie on arcs or circles. In particular, we optimize a bi-objective function under the constraints that the data representations are arranged around a closed continuum such that their latent ordering is preserved. In the following subsection we give a detailed description of the proposed methodology.

3.2. Circular object arrangement using a bilevel model

Our goal is to generate a set of optimal embedded points $\{\mathbf{z}_i\}_{i=1}^n$ lying on the unit sphere of dimension $k-1$ (with $k \ll n$) that are arranged circularly and at the same time maintaining overall their original similarities. Although k can be set to any value, in practice it was observed that smaller values yield better performance. For the optimizing model, we combine a least-squares loss function, together with a QAP objective involving the cosine similarities $\mathbf{Z}\mathbf{Z}^\top$ of the embedded points and the circular template

of Eq. (3) to penalize embedded points deviating from a circular order. A weighted combination of these two objectives allows for the generation of a more flexible circular arrangement of patterns while preserving on average their original similarities. To enforce the circular ordering, we constrain the permutation to capture the circular ordering of the current embedded representations in a secondary optimization step. Additional constraints are also used to avoid trivial cases whereby all embedded data collapse on a single point cluster on the unit sphere. We formulate the above requirements in the following optimization

$$\min_{\substack{\mathbf{Z}\in\mathbb{R}^{n\times k}\\ \mathbf{\Pi}\in\mathcal{M}^n}} \quad \left\{\phi(\mathbf{Z},\mathbf{\Pi}) \triangleq \sum_{i>j}\Big(R_{ij} - \|\mathbf{z}_i-\mathbf{z}_j\|_2\Big)^2 + \frac{1}{n}\mathrm{tr}\Big[\mathbf{\Pi}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Pi}^\top\mathbf{B}^\top\Big]\right\} \tag{5a}$$

$$\text{s.t.} \quad \mathrm{diag}\left(\mathbf{Z}\mathbf{Z}^\top\right) = \mathbf{1}_n, \tag{5b}$$

$$\mathbf{1}_n^\top\mathbf{Z} = \mathbf{0}_k^\top, \tag{5c}$$

$$\mathbf{\Pi} \in \underset{\mathbf{\Pi}'\in\mathcal{M}^n}{\arg\min}\, f\big(\mathbf{Z}\big), \tag{5d}$$

where $\mathbf{0}_k$ is the constant vector of all zeros, and $R_{ij} = \frac{\max(d(\mathbf{A}))}{n}[d(\mathbf{A})]_{ij}$. Here $d(\cdot)$ can be any function that converts similarities to dissimilarities which is then scaled to effectively combine the two objectives for the optimization. To ensure the embedded points lie on the $(k-1)$-dimensional unit sphere, we enforce constraint (5b). However, the objective $\phi(\mathbf{Z},\mathbf{\Pi})$ can admit a minimum energy equal to $\sum_{i>j}R_{ij}^2 + \frac{2}{n}\sum_{i>j}B_{ij} = \sum_{i>j}R_{ij}^2 + \frac{n^2}{4}$ for even n, or $\sum_{i>j}R_{ij}^2 + \frac{(n-1)^2(n+1)}{4n}$ for odd n, due to the trivial solution with all embedded data patterns collapsing on a single point. To control this, we use constraint (5c) which forces the coordinates of the embedded representations $\mathbf{z}_i$ to be centered. Additionally, this constraint encourages better separation among the embedded points (other trivial cases of points collapsing on just few antipodal clustered configurations may arise only when the original data support such configurations).

Problem (5) forms a parametric bilevel optimization problem of mixed discrete and continuous variables and can be solved sequentially as a two-stage decision problem with each level's variable treated as a parameter for the other [38]. Function $f(\cdot)$ in the lower level optimization problem (5d) can be any function of $\mathbf{Z}$ which when optimized recovers a circular ordering of the embedded points currently existing in the

upper level, e.g. $f(\mathbf{Z}) = \mathrm{tr}\Big[\mathbf{\Pi}\mathbf{A}_{\mathbf{Z}}\mathbf{\Pi}^\top\mathbf{B}^\top\Big]$ with $\mathbf{A}_{\mathbf{Z}}$ being any symmetric similarity matrix constructed from $\mathbf{Z}$. In our case, we employ Coifman's spectral method [13] described in Algorithm 1 due to its simplicity and efficiency. An approximate solution to (5) can be found using the augmented Lagrangian multipliers method (ALMM), where the upper level optimizes the Lagrangian formulation accommodating Eqs. (5a), (5b) and (5c), and in each steepest descent iteration the feasible point of the upper level problem is used to find an approximate solution for the lower level problem (using Algorithm 1 to recover a circular ordering). Subsequently, the upper level is updated accordingly and this two-stage process continues iteratively until convergence. Problem (5) effectively relies on two parameters that interact to capture different aspects of the problem, that is $\mathbf{Z}$ and $\mathbf{\Pi}$ to gradually capture locations and order, respectively. The locations are continually adjusted to fit the original similarities and comply with the current ordering, while the ordering is being adjusted to follow the current embedded representations. This fact provides more flexibility in the optimization procedure to result to better seriation results in terms of various measures, without high sensitivity and over-reliance to the data similarity matrix and its conformity to the given seriation template.

The augmented Lagrangian of Eqs. (5a)-(5c) is

$$\begin{aligned}\mathcal{L}(\mathbf{Z},\mathbf{\Pi},\boldsymbol{\lambda},\boldsymbol{\mu},\rho_\lambda,\rho_\mu) = &\sum_{i>j} R_{ij}^2 + \frac{1}{n}\mathrm{tr}\Big[\mathbf{\Pi}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Pi}^\top\mathbf{B}^\top\Big] \\ &+ \mathrm{tr}\Big[\mathbf{Z}^\top\mathbf{V}\mathbf{Z}\Big] - 2\mathrm{tr}\Big[\mathbf{Z}^\top\mathbf{U}\mathbf{Z}\Big] + \big\langle\boldsymbol{\lambda},\mathbf{1}_n^\top\mathbf{Z}\big\rangle + \frac{\rho_\lambda}{2}\left\|\mathbf{1}_n^\top\mathbf{Z}\right\|_2^2 \\ &+ \big\langle\boldsymbol{\mu},\mathrm{diag}\left(\mathbf{Z}\mathbf{Z}^\top\right) - \mathbf{1}_n\big\rangle + \frac{\rho_\mu}{2}\left\|\mathrm{diag}\left(\mathbf{Z}\mathbf{Z}^\top\right) - \mathbf{1}_n\right\|_2^2,\end{aligned} \tag{6}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^k$ and $\boldsymbol{\mu} \in \mathbb{R}^n$ contain the Lagrange multipliers of the $n+k$ constraints and $\rho_\lambda, \rho_\mu > 0$ are the corresponding penalty parameters. Note, that the loss function $\sum_{i>j}\Big(R_{ij} - \|\mathbf{z}_i - \mathbf{z}_j\|_2\Big)^2$ can be equivalently expressed as $\sum_{i>j} R_{ij}^2 + \mathrm{tr}\Big[\mathbf{Z}^\top\mathbf{V}\mathbf{Z}\Big] - 2\mathrm{tr}\Big[\mathbf{Z}^\top\mathbf{U}\mathbf{Z}\Big]$, with

$$V_{ij} = \begin{cases} -1, & i \neq j, \\ n-1, & i = j, \end{cases}$$

and

$$U_{ij} = \begin{cases} -\frac{R_{ij}}{\|\mathbf{z}_i - \mathbf{z}_j\|_2}, & \text{if } i \neq j \wedge \|\mathbf{z}_i - \mathbf{z}_j\|_2 \neq 0, \\ 0, & \text{if } i \neq j \wedge \|\mathbf{z}_i - \mathbf{z}_j\|_2 = 0, \\ -\sum_{j \neq i} U_{ij}, & \text{if } i = j. \end{cases}$$

Consequently, the following bilevel problem can be solved

$$\begin{aligned} \min_{\substack{\mathbf{Z} \in \mathbb{R}^{n \times k} \\ \mathbf{\Pi} \in \mathcal{M}^n}} \quad & \mathcal{L}\big(\mathbf{Z}, \mathbf{\Pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \rho_\lambda, \rho_\mu\big) \\ \text{s.t.} \quad & \mathbf{\Pi} = \text{CircularSeriation}\big(\mathbf{Z}\big), \end{aligned} \tag{7}$$

where the upper-level problem can be solved as an unconstrained problem using any standard steepest descent method. Here we employ a BFGS Quasi-Newton method with a cubic line search procedure, where the gradient of the augmented Lagrangian with respect to $\mathbf{Z}$ is calculated as

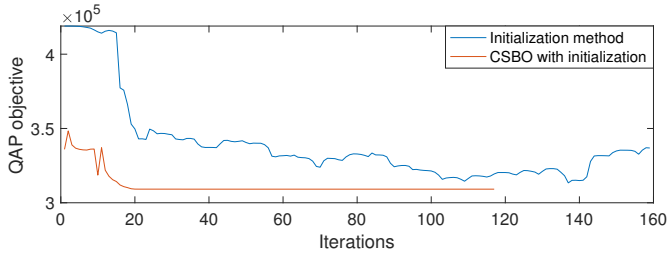
$$\begin{aligned} \nabla_{\mathbf{Z}} \mathcal{L}\big(\mathbf{Z}, \mathbf{\Pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \rho_\lambda, \rho_\mu\big) = 2\mathbf{V}\mathbf{Z} - 2\mathbf{U}\mathbf{Z} + \frac{2}{n}\mathbf{\Pi}^\top \mathbf{B}^\top \mathbf{\Pi}\mathbf{Z} \\ + \mathbf{1}_n \boldsymbol{\lambda}^\top + 2\,\text{diag}\,\big(\boldsymbol{\mu}\big)\mathbf{Z} + 2\rho_\mu\Big(\mathbf{Z}_D\mathbf{Z} - \mathbf{Z}\Big) + \rho_\lambda \mathbf{J}\mathbf{Z}. \end{aligned} \tag{8}$$

with $\mathbf{Z}_D = \text{diag}\,\big(\,\text{diag}\,\big(\mathbf{Z}\mathbf{Z}^\top\big)\big)$ and $\mathbf{J} = \mathbf{1}_n\mathbf{1}_n^\top$. The use of an augmented Lagrangian approach allows for a more flexible embedding procedure on spheres of adjusting radii regulated by penalty values until a better circular configuration is located. We refer to this approach as Circular Seriation using Bilevel Optimization (CSBO).

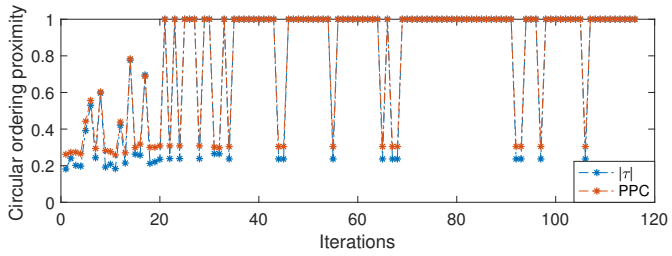
3.3. CSBO initialization

The objective in Eq. (5a) is not necessarily convex in $\mathbf{Z}$ and the minimization process can get trapped in a local minima. In such cases, it is extremely useful to seed the optimization with another method, such as the spectral [13], or classical MDS [39], or any other fast method providing a reasonable starting point. In our case, we initialize problem (5) by solving the following a variant of Eq. (4), given by

$$\begin{aligned} \underset{\mathbf{Z} \in \mathbb{R}^{n \times k}}{\arg\min} \quad & \text{tr}\Big[\mathbf{Z}^\top \mathbf{L}_{\mathbf{A}} \mathbf{Z}\Big] \\ \text{s.t.} \quad & \text{diag}\,\big(\mathbf{Z}\mathbf{Z}^\top\big) = \mathbf{1}_n, \\ & \mathbf{1}_n^\top \mathbf{Z} = \mathbf{0}_k^\top, \end{aligned} \tag{9}$$



(a) Change in QAP objective values during optimization



(b) Proximity between CSBO consecutive solutions

Figure 2: (a) QAP objective values for the initialization method and CSBO logged at each steepest descent iteration. (b) Circular ordering proximity in terms of absolute Kendall's τ [40, 16] and PPC [7], between permutations recovered at consecutive optimization steps in CSBO. Values closer to 1 indicate lesser change between consecutive solutions.

because it showed better performance in practice. Problem (9) embeds points on the sphere while maintaining their similarities using the unnormalized Laplacian $\mathbf{L_A}$ of $\mathbf{A}$. As before, we need to ensure that the embedded points are centered to avoid the trivial solution of a single collapsed cluster. We solve an unconstrained version of problem (9) where the constraints are converted to penalty functions and solve it again with a Quasi-Newton method. The final ordering is determined by the angle of first two embedding coordinates as in step 5 of Algorithm 1. We specifically solve

$$\underset{\mathbf{Z}\in\mathbb{R}^{n\times k}}{\arg\min} \quad \operatorname{tr}\left[\mathbf{Z}^\top \mathbf{L_A} \mathbf{Z}\right] + \sigma \left\|\mathbf{1}_n^\top \mathbf{Z}\right\|_2^2 + \sigma \left\|\operatorname{diag}\left(\mathbf{Z}\mathbf{Z}^\top\right) - \mathbf{1}_n\right\|_2^2, \tag{10}$$

where $\sigma = \max(\mathbf{A}) \times n$. Empirical observations showed that repeating this multiple times, starting from different random $\mathbf{Z}$ on the sphere, and keeping the best to seed CSBO helps improve the final performance.

Algorithm 2 ALMM-based CSBO

Input: dissimilarity matrix $\mathbf{R} \in \mathbb{R}^{n \times n}$, initial points $\mathbf{Z}^{(0)} \in \mathbb{R}^{n \times k}$ and $\mathbf{\Pi}^{(0)} \in \mathcal{M}^n$ returned by (10), embedding dimensionality k

Parameters: $\epsilon = 10^{-1}, \gamma = 5$

Output: $\mathbf{Z} \in \mathbb{R}^{n \times k}, \mathbf{\Pi} \in \mathcal{M}^n$

1: Let $t = 0, \boldsymbol{\lambda}^{(0)} = 0, \boldsymbol{\mu}^{(0)} = 0, \rho_{\lambda}^{(0)} = \rho_{\mu}^{(0)} = 1$

2: **while** *not converged* **do**

3: $\quad \mathbf{g}^{(t)} = \nabla_{\mathbf{Z}} \mathcal{L}\big(\mathbf{Z}^{(t)}, \mathbf{\Pi}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\mu}^{(t)}, \rho_{\lambda}^{(t)}, \rho_{\mu}^{(t)}\big)$

4: $\quad \mathbf{Z}^{(t+1)} = \arg\min \mathcal{L}\big(\mathbf{Z}^{(t)}, \mathbf{\Pi}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\mu}^{(t)}, \rho_{\lambda}^{(t)}, \rho_{\mu}^{(t)}\big)$

5: $\quad \mathbf{\Pi}^{(t+1)} = \text{CircularSeriation}\big(\mathbf{Z}^{(t+1)}\big)$

6: $\quad$ **if** $\big\|\mathbf{g}^{(t)}\big\|_{\infty} \leq \epsilon \wedge \Big\|\text{diag}\left(\mathbf{Z}^{(t+1)}\mathbf{Z}^{(t+1)^{\top}}\right) - \mathbf{1}_n\Big\|_2 < \epsilon$ **then**

7: $\quad\quad$ *converged*

8: $\quad$ **end if**

9: $\quad \boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \rho_{\lambda}^{(t)}\big(\mathbf{1}_n^{\top}\mathbf{Z}^{(t+1)}\big)$

10: $\quad \boldsymbol{\mu}^{(t+1)} = \boldsymbol{\mu}^{(t)} + \rho_{\mu}^{(t)}\Big(\text{diag}\left(\mathbf{Z}^{(t+1)}\mathbf{Z}^{(t+1)^{\top}}\right) - \mathbf{1}_n\Big)$

11: $\quad \rho_{\lambda}^{(t+1)} = \rho_{\mu}^{(t+1)} = \gamma\rho_{\lambda}^{(t)}$

12: **end while**

13: **return** $\mathbf{Z}, \mathbf{\Pi}$

Figure 2a demonstrates the performance difference between problem (10) and the seeded CSBO, in terms of QAP objective for each steepest descent iteration. It is interesting to observe that problem (10) alone can recover orderings that yield good performance, but when used to initialize CSBO, performance and convergence are significantly improved. This further highlights the benefit of the proposed approach. Figure 2b shows how the ordering changes during the optimization for CSBO with initialization and illustrates its fast convergence behavior. The complete steps of the proposed framework are summarized in Algorithm 2, which runs on $\mathcal{O}(n^3)$ time complexity since the most costly operation in each iteration is the eigen-decomposition of step 5.

4. Experimental results

We present a series of experiments in order to compare the proposed methodology with other relevant methods[1] such as heuristics and spectral-based approaches that approximately solve the circular seriation problem. Results are assessed using various seriation measures and permutation proximities as well as the corresponding QAP objective values (2). All methods are also compared in terms of running-time performances. Several datasets ranging from synthetic to real are used in this comparison.

In particular, we experiment with the following methods:

- *CSBO*$_k$: the proposed algorithm for embedding dimensionality k.
- *CSpectral*$_A$: a spectral method [13] that recovers a circular ordering based on the first two non-trivial eigenvectors of the normalized Laplacian.
- *CSpectral*$_B$: a spectral method [16] based on CSpectral$_\mathrm{A}$ that recovers latent orderings using pre-processed similarities based on object proximities in a low dimensional embedding space from a Laplacian embedding.
- *FAQ*: the Fast Approximate QAP method [22] based on the relaxation in the Birkhoff polytope and the Frank-Wolfe method using Eq. (3) as a template.
- *SA*: a simulated annealing-based optimizer applied on problem (2) also based on Eq. (3) as a template.
- *R2E*: a rank-two ellipse seriation method based on iterative application of the correlation function on the original dissimilarities [35].
- *Circ-GnCR*: the GnCR method [9] modified to accommodate circular seriation by using the circulant positional template $B_{ij} = \left[\cos(2\pi\frac{i-1}{n}) - \cos(2\pi\frac{j-1}{n})\right]^2 + \left[\sin(2\pi\frac{i-1}{n}) - \sin(2\pi\frac{j-1}{n})\right]^2,\ i,j \in \{1,\ldots,n\}$, which showed better performance in practice for that method.

[1] The code for the proposed algorithms and other evaluated methods is included in our Matlab toolbox for seriation, available at http://pcwww.liv.ac.uk/~goulerma/software/seriation.zip.

Implementations of the algorithms are in MATLAB ver.9.4, and for timing comparisons we use a 3.4 GHz Intel i7 desktop with 8 GB of memory. We selected a range of synthetic and real datasets, associated with a similarity matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, where we aim to recover a circular ordering of the data that can indicate relations among genes that have cyclic functionality, cortical functional areas of the human brain, semantic contents of images with cyclic gradual variations, etc. These sets include:

- Synthetic datasets:
  - *Points On the Circle (POC)*: a 76×76 similarity matrix $\mathbf{A}$ generated, using Euclidean distances and a Gaussian kernel, with points randomly placed on a circle.
  - *Points On the Ellipse (POE)*: a 94×94 similarity matrix $\mathbf{A}$, generated as above, with points randomly placed on an ellipse.
- Real datasets:
  - *Yeast Cell Cycle Gene Expression (Yeast)*: a 250×18 gene expression data matrix from 250 yeast saccharomyces cerevisiae cell cycle genes [11], converted to dissimilarities using the same preprocessing used in [3], turned to similarities by negation.
  - *Human brain functional coactivation network (Human Brain)*: a 638×638 similarity matrix $\mathbf{A}$ whose elements represent functional associations between cortical areas in the human brain [41].
  - *Rotating Robot (Robot)*: a 72×72 similarity matrix $\mathbf{A}$ from 72 rotating robot images [42] captured at 5 degrees apart, around 360 degrees and organized in 12 classes, each spanning 30 degrees. For the analysis of images the SIFT [43] vector descriptors were employed with patches of 12 pixels long overlapping every 6 pixels. A bag-of-visual-words representation was used and clustering was performed by k-means with a cluster size of 500. To derive the similarity matrix we use the exponentiated χ^2 distance.
  - *Fashion-MNIST (FMNIST)*: a 1,000×1,000 matrix $\mathbf{A}$ from the FMNIST image dataset [44], where images represent different fashion products from

10 categories. To derive the similarity matrix, a similar preprocessing as in the Robot dataset was applied.

– *Drosophila Circandian Gene Expression (Drosophila)*: a 2,837×16 gene expression data matrix $\mathbf{X}$ from 2,837 Drosophila melanogaster circandian genes [12]. For the generation of the similarity matrix $\mathbf{A}$ we assume $A_{ij} = \left|\mathbf{x}_i^\top \mathbf{x}_j\right|$, where $\mathbf{x}_i$ is the i-th row of $\mathbf{X}$.

4.1. Benchmark evaluation

Firstly, we evaluate the utility of the proposed algorithm in terms of the QAP objective and also the Hamiltonian cycle (HC) measure, proposed in [15] to assess circular seriation with emphasis on the local structure of the ordering. We additionally introduce a new measure for circular seriation that counts the number of CR violations of a similarity matrix described by the following loss function

$$\begin{aligned}\mathrm{CRV}(\mathbf{A}) = \min_{\omega\in[n-2]}\Bigg\{ &\sum_{i+\omega<k<j} h(A_{i,k}, A_{i,j}) + \sum_{i+\omega<k<j} h(A_{k-\omega,j}, A_{i,j}) \\ &+ \sum_{j-\omega\le i<k<j} h(A_{i,j}, A_{i,k}) + \sum_{j-\omega\le i<k<j} h(A_{i,j}, A_{k,j})\Bigg\},\end{aligned} \tag{11}$$

where

$$h(x,y) = \begin{cases} 0, & \text{if } x > y, \\ 1, & \text{otherwise.}\end{cases}$$

Specifically, for each diagonal denoting a possible turning point in the matrix, we calculate the number of Robinsonian violations on the left of the diagonal (last two summation terms of Eq. (11)) and the number of anti-Robinsonian violations on the right (first two summation terms of Eq. (11)). We report the minimum over all ω values indicating the original turning point of the matrix and we refer to this measure as CRV (Circular-Robinsonian Violations).

Since the values on different datasets are not comparable, for interpretability we report a normalized value for each measure that quantifies the deviation from the best performer for that dataset. For the i-th dataset the deviation for the j-th algorithm is defined as

$$\Theta_{ij} = \left|\frac{\text{score}_{ij} - \text{best}_i}{\text{best}_i}\right|. \tag{12}$$

Tables 1, 2 and 3 show the normalized deviation from the best QAP, HC and CRV values respectively, for each algorithm and for the 7 datasets (boldfaced table entries denote best performance). For the proposed method, we experimented with various values of embedding dimension k, but we present results for the two best performing cases of $k = 2, 3$. We can see that for $k = 2$ the proposed method outperforms all other algorithms in terms of QAP objective value. For the HC measure CSBO_2 remains competitive against the best performing FAQ in all datasets apart from Human Brain. In terms of CRV, CSBO performs best for $k = 2, 3$, while spectral-based methods remain competitive. In general, the proposed approach maintains a competitive performance also for $k = 3$ for all measures.

	$\text{CSpectral}_\text{A}$	$\text{CSpectral}_\text{B}$	FAQ	SA	Circ-GnCR	R2E	CSBO_2	CSBO_3
POC	0	0	0	0	0	0	0	0
POE	0	0	0	0	0.0860	0	0	0
Yeast	0.0013	0.0025	0.0051	0	0.0159	0.0025	0.0005	0.0006
Human Brain	0.8509	0.1034	0	0.0424	0.1348	0.0230	0.0132	0.0325
Robot	0.0075	0.0018	0.0117	0.0038	0.0034	0.0036	0	0.0011
FMNIST	0.0086	0.0001	0.0086	0.0071	0.0461	0.0085	0	0.0079
Drosophila	0.0119	0.0139	0	0.0139	0.0317	0.0167	0.0048	0.0037
Average	0.1257	0.0174	0.0036	0.0096	0.0454	0.0078	**0.0027**	0.0065

Table 1: Deviation from the best QAP objective value across the 7 datasets.

Empirical evaluations have shown that depending on the problem at hand, different embedding dimensions k can yield different performance, owing to the underlying dimensionality of the data, but in general $k = 2$ appeared to perform best in the present benchmark. It is worth noting the poor performance of the continuation-based Circ-GnCR method, which in this case can be explained due to the nature of the involved optimizing objective that cannot be rendered convex for the initial steps of the continuation scheme; this is an important attribute for the well-behavior of GnCR-based methods as shown by [9].

Figure 3 shows similarity maps of the Yeast gene expression dataset for four dif-

	$\text{CSpectral}_\text{A}$	$\text{CSpectral}_\text{B}$	FAQ	SA	Circ-GnCR	R2E	CSBO_2	CSBO_3
POC	0	0	0	0	0	0	0	0
POE	0	0	0	0	0.1836	0	0	0
Yeast	0.0860	0.1219	0.0300	0	0.1170	0.0929	0.0743	0.1054
Human Brain	0.5455	0.3344	0.0355	0	0.4345	0.4056	0.3775	0.4380
Robot	0.1559	0.0519	0	0.0253	0.0958	0.1354	0.0580	0.1133
FMNIST	0.1994	0	0.1006	0.1259	0.1260	0.1892	0.0049	0.1906
Drosophila	0	0.0017	0.0267	0.0611	0.0602	0.0087	0.0485	0.0503
Average	0.1410	0.0728	**0.0275**	0.0303	0.1453	0.1188	0.0805	0.1282

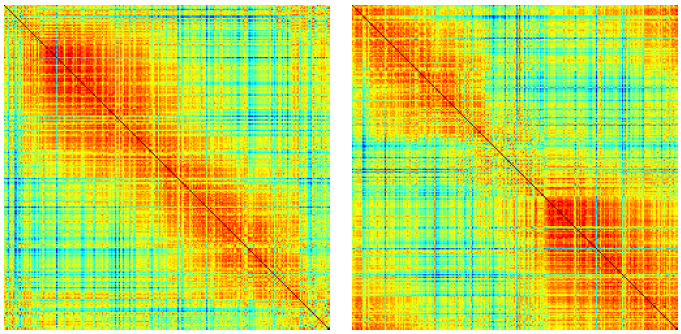
Table 2: Deviation from the best HC score across the 7 datasets.

	$\text{CSpectral}_\text{A}$	$\text{CSpectral}_\text{B}$	FAQ	SA	Circ-GnCR	R2E	CSBO_2	CSBO_3
POC	0.1277	0.1771	1.7852	3.1821	0	0	0.0631	0.0631
POE	0.4443	0.5295	0	0.4891	1.0011	0.4443	0.4419	0.4419
Yeast	0.1054	0.1162	0.1539	0	0.1736	0.0885	0.1026	0.1045
Human Brain	1.3291	0.2509	0	0.1124	0.3600	0.1268	0.1275	0.1532
Robot	0.2883	0.1725	0.8489	0.2661	0.2184	0.1211	0	0.0130
FMNIST	0.1033	0.0363	0.0778	0.0618	0.6043	0.2229	0	0.0826
Drosophila	0.2216	0.2287	0.0131	0.2393	0.2558	0.2348	0.0004	0
Average	0.4366	0.2519	0.4798	0.7251	0.4355	0.2064	**0.1226**	0.1431

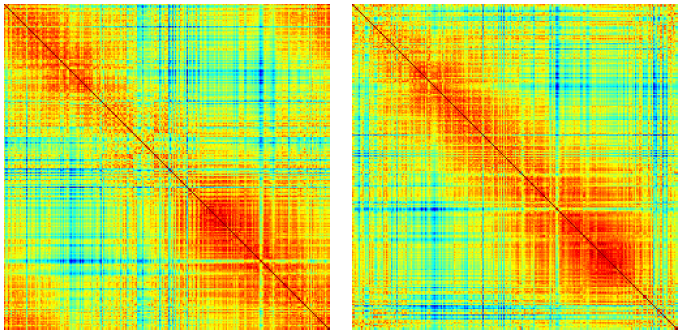
Table 3: Deviation from the best CRV value across the 7 datasets.

ferent methods highlighting various circular seriation effects. SA and the proposed method reveal slightly more distinct concentrations of high and low similarities. This is also reflected in Tables 1, 2 and 3. In a similar fashion, Figure 4 demonstrates circularly seriated similarity maps of the Human Brain dataset along with the corresponding functional association connectivity graphs. We can observe that the proposed method can reveal functional associations between more cortical areas of the human brain compared to $\text{CSpectral}_\text{A}$, showing similar results to the original meta-analysis of [41], and additionally suggesting a circular functional association among the different cortical

areas.



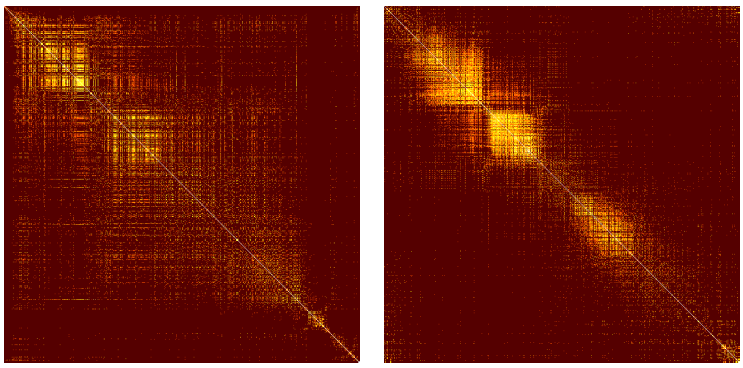
(a) $\text{CSpectral}_{\text{A}}$ map (b) R2E map



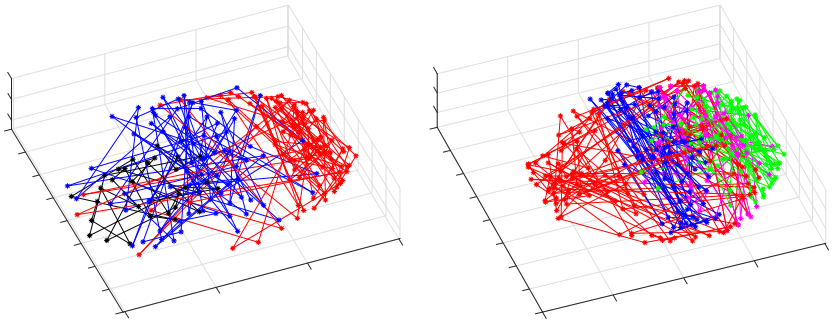
(c) SA map (d) CSBO_2 map

Figure 3: Reconstructed similarity maps of the Yeast dataset for four different methods. High similarities (red) appear as block formations along the main diagonal and the anti-diagonal corners, while low similarities (blue) concentrate in off-diagonal regions.

As far as running-time performance is concerned, Figure 5 shows that when the dataset size increases significantly, the proposed method becomes slower compared to spectral methods. It has been empirically observed however, that in cases where the data is near circulant, CSBO converges faster to a local solution. This can be explained as the corresponding QAP is easier to solve in the special case where the similarity matrix is a circulant-CR one (as shown later in the Appendix). Figure 6 illustrates the convergence behavior of CSBO for two different cases of similarity matrices. Overall, the proposed method has a reasonable trade-off between running time and seriation



(a) CSpectral$_{\mathrm{A}}$ Human Brain map (b) CSBO$_2$ Human Brain map



(c) CSpectral$_{\mathrm{A}}$ connectivity graph (d) CSBO$_2$ connectivity graph

Figure 4: Reconstructed similarity maps of the Human Brain dataset and corresponding functional association connectivity graphs of the brain, for CSpectral$_{\mathrm{A}}$ and CSBO$_2$. Straight line colors in the bottom plots correspond to groups of regions with similar functional association.

300 quality.

4.2. Evaluating circularly ordered images using class information

We further assess the effectiveness of the proposed model using class label information for two image datasets. We define the binary class association matrix as $C_{ij} = 1$ iff objects i and j belong to the same class, and we use a suitably modified measure for circular seriation, similar to [9], that quantifies the occurences a seriation algorithm

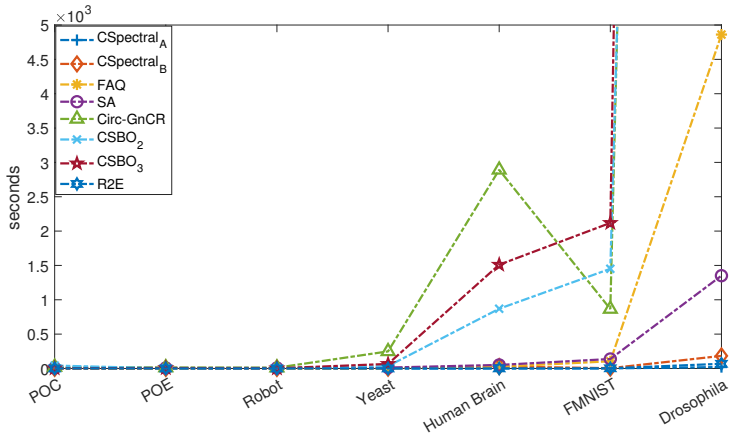


Figure 5: Running time performance of the compared methods for the 7 different datasets. Slower running times of CSBO_2, CSBO_3 and Circ-GnCR for the Drosophila dataset (correspondingly measured at 58×10^3, 66×10^3 and 93×10^3 seconds) are clipped.

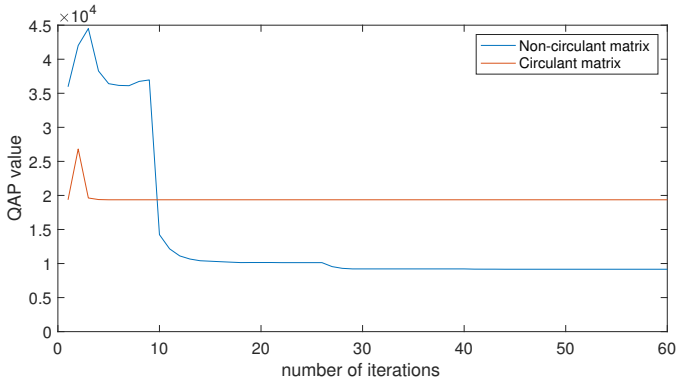


Figure 6: Convergence behavior of the proposed method for circulant and non-circulant CR matrices.

places objects from different classes adjacently, as

$$\delta_{\text{count}}(\boldsymbol{\pi}, \mathbf{C}) \triangleq \sum_{i=1}^{n-1} \left(1 - C_{\pi(i)\pi(i+1)}\right) + \left(1 - C_{\pi(1)\pi(n)}\right). \tag{13}$$

Additionally, we employ another adapted measure for assessing circular orderings that penalizes objects from the same class that are placed apart [9]. It can be formulated

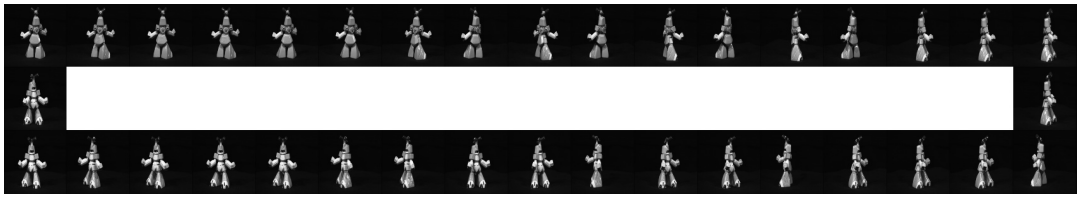
as (2), but the similarity matrix $\mathbf{A}$ is replaced with $\mathbf{C}$, according to

$$\text{cQAP}_{\text{sup}}(\boldsymbol{\pi}, \mathbf{C}, \mathbf{B}) \triangleq \sum_{i,j=1}^{n} C_{\pi_i \pi_j} B_{ij}. \tag{14}$$

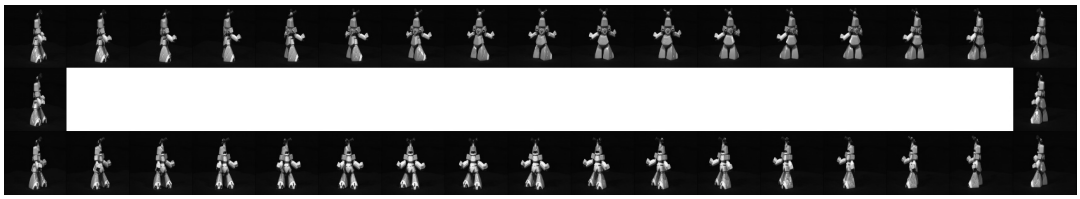
Table 4 shows results for the above measures and the two image datasets. For the Robot dataset and for both measures, we can see that $\text{CSpectral}_{\text{B}}$ performs best, while CSBO_2 scores very closely. Figure 7 demonstrates two different circular seriation effects on the robot images for $\text{CSpectral}_{\text{A}}$ and CSBO_2, with the latter recovering an order that reflects the smooth variation across the patterns as opposed to $\text{CSpectral}_{\text{A}}$ which fails to do so. Similarly, for the FMNIST dataset, the proposed CSBO_2 scores very close to the best (FAQ) in terms of δ_{count}, while both CSBO_2 and CSBO_3 perform very well in terms of cQAP_{sup}, with the latter scoring best. Figure 8 illustrates the underlying cyclic order of the 10 different categories of FMNIST identified by CSBO_2, which reveals a smooth variation according to the semantic content of the fashion images. In general, we can see that despite the fact that image seriation is challenging due to the quality of the extracted image descriptors, the proposed approach appears to perform well in terms of recovering circular orderings that place images of the same category closely.

	δ_{count}		cQAP_{sup}	
	Robot	FMNIST	Robot	FMNIST
$\text{CSpectral}_{\text{A}}$	53	518	1,924	1,068 $(\times 10^4)$
$\text{CSpectral}_{\text{B}}$	**14**	414	**876**	851 $(\times 10^4)$
FAQ	44	**364**	1,896	983 $(\times 10^4)$
SA	35	392	1,800	813 $(\times 10^4)$
Circ-GnCR	40	435	3,044	902 $(\times 10^4)$
R2E	51	473	1,916	848 $(\times 10^4)$
CSBO_2	16	400	896	**812** $(\times 10^4)$
CSBO_3	49	456	1,872	820 $(\times 10^4)$

Table 4: δ_{count} and cQAP_{sup} measures for each algorithm for the labeled Robot and FMNIST image datasets.



(a) $\text{CSpectral}_{\text{A}}$



(b) CSBO_2

Figure 7: Image sequence of a robot captured throughout different angles, ordered circularly using $\text{CSpectral}_{\text{A}}$ and CSBO_2. For presentation reasons, a subset of the 72 robot images is displayed here.

Finally we evaluate the ability of CSBO to find a solution that is close to the true underlying ordering of the Robot dataset. Table 5 employs a modified absolute Kendall's τ score to calculate the rank correlation between two circular orderings [40, 16], and also the agreement in terms of positional proximities using the PPC [7] measure. We can see that $\text{CSpectral}_{\text{B}}$ and CSBO_2 maintain scores close to the optimum, with the former scoring slightly better.

5. Conclusion

In this work we proposed a new algorithm for the circular seriation problem. Our approach operates within a bilevel optimization framework by iteratively embedding data instances circularly on a spherical space such that their original similarities are preserved. The flexibility of this approach is that it is not restricted to rely exclusively on the original similarities, but it also maintains a set of low-dimensional representa-

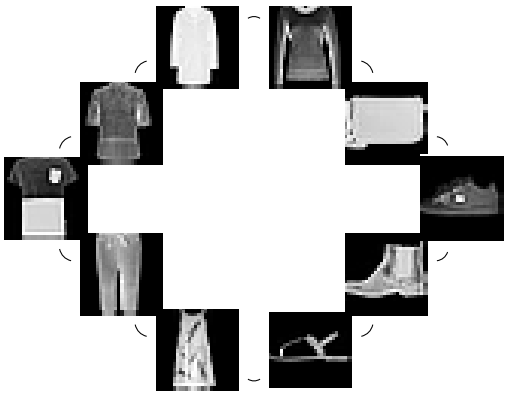


Figure 8: Circular ordering of the distinct image categories from the FMNIST dataset as recovered by CSBO_2.

	$\lvert\tau\rvert$	PPC
$\text{CSpectral}_\text{A}$	0.218	0.424
$\text{CSpectral}_\text{B}$	**0.995**	**1**
FAQ	0.261	0.440
SA	0.241	0.428
Circ-GnCR	0.451	0.566
R2E	0.225	0.429
CSBO_2	0.991	0.999
CSBO_3	0.225	0.429

Table 5: Modified Kendall's $\lvert\tau\rvert$ and PPC scores between final solution and true underlying circular ordering, for the Robot dataset. Values closer to 1 indicate better ordering agreement.

tions to recover the underlying ordering in a more robust manner. This makes it a useful alternative for exploratory data analysis as a means to recover a consistent arrangement of gradually varying data in a wide range of applications. Its practical utility is demonstrated through experiments which show a very competitive performance against other circular seriation methods in terms of various measures. As running time performance is concerned, CSBO appears to be less scalable when the problem size rapidly increases

and therefore a future improvement of this work would be to address this particular issue.

6. Appendix

In this appendix we show that given a circulant-CR similarity matrix (decreasing circulant) $\mathbf{S} \in \mathbf{R}_{\geq 0}^{n \times n}$ and a circulant-CR dissimilarity matrix (increasing circulant) $\mathbf{Q} \in \mathbf{R}_{\geq 0}^{n \times n}$, then all cyclic shifts of the identity permutation are optimal solutions to the problem QAP($\mathbf{S}, \mathbf{Q}$).

Let $\mathbf{N}^{\Delta}$ be a binary increasing symmetric circulant matrix, sufficiently defined by its first row

$$N_{ij}^{\Delta} = \begin{cases} 0 & \text{if } |i-j| \bmod (n-\Delta+1) < \Delta \\ 1 & \text{otherwise,} \end{cases} \tag{15}$$

for $1 \leq j \leq i \leq n$ and with $1 \leq \Delta \leq \lfloor n/2 \rfloor + 1$, denoting the first element that increases in the first row. Accordingly let the decreasing symmetric circulant matrix be defined as

$$M_{ij}^{\Delta} = \begin{cases} 1 & \text{if } |i-j| \bmod (n-\Delta+1) < \Delta \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

We can now define both $\mathbf{S}$ and $\mathbf{Q}$ as a conic combination of the binary circulant matrices $\mathbf{M}^{\Delta}$ and $\mathbf{N}^{\Delta}$ [45], respectively, as

$$\mathbf{S} = \alpha_0 \mathbf{J} + \sum_{\Delta=1}^{\lfloor n/2 \rfloor + 1} \alpha_{\Delta} \mathbf{M}^{\Delta}, \tag{17}$$

$$\mathbf{Q} = \beta_0 \mathbf{J} + \sum_{\Delta=1}^{\lfloor n/2 \rfloor + 1} \beta_{\Delta} \mathbf{N}^{\Delta}, \tag{18}$$

with $1 \leq \alpha_{\Delta}, \beta_{\Delta} \leq \lfloor n/2 \rfloor + 1$. In our case $\mathbf{Q}$ is a dissimilarity matrix and thus $\beta_0 = 0$. By definition, $\langle \mathbf{M}^{\Delta}, \mathbf{N}^{\Delta} \rangle = \langle \mathbf{M}^{\Delta}, \mathbf{J} \rangle - \langle \mathbf{M}^{\Delta}, \mathbf{M}^{\Delta} \rangle = 0$, and hence there exists no permutation $\boldsymbol{\pi}$ that can result to a smaller value for QAP($\mathbf{M}^{\Delta}$, $\mathbf{N}^{\Delta}$).

It is easy to see that

$$\begin{aligned}
&\langle \mathbf{M}_{\pi}^{\Delta}, \mathbf{N}^{\Delta} \rangle \geq \langle \mathbf{M}^{\Delta}, \mathbf{N}^{\Delta} \rangle \\
&\Rightarrow \sum_{\Delta=1}^{\lfloor n/2 \rfloor+1} \beta_{\Delta} \langle \mathbf{M}_{\pi}^{\Delta}, \mathbf{N}^{\Delta} \rangle \geq \sum_{\Delta=1}^{\lfloor n/2 \rfloor+1} \beta_{\Delta} \langle \mathbf{M}^{\Delta}, \mathbf{N}^{\Delta} \rangle \\
&\Rightarrow \alpha_0 \mathbf{J} + \sum_{\Delta=1}^{\lfloor n/2 \rfloor+1} \alpha_{\Delta} \langle \mathbf{M}_{\pi}^{\Delta}, \mathbf{Q} \rangle \geq \alpha_0 \mathbf{J} + \sum_{\Delta=1}^{\lfloor n/2 \rfloor+1} \alpha_{\Delta} \langle \mathbf{M}^{\Delta}, \mathbf{Q} \rangle \\
&\Rightarrow \langle \mathbf{S}_{\pi}, \mathbf{Q} \rangle \geq \langle \mathbf{S}, \mathbf{Q} \rangle. \qquad (19)
\end{aligned}$$

This concludes the proof that the identity permutation (including its $n-1$ cyclic shifts) is an optimal solution to the problem QAP($\mathbf{S}, \mathbf{Q}$).

Acknowledgements

The work was supported by a studentship from the School of Electrical Engineering, Electronics and Computer Science, at the University of Liverpool, U.K. We thank the anonymous reviewers for the very constructive comments and suggestions.

References

[1] F. R. Hodson, The La Tène cemetery at Münsingen-Rain: catalogue and relative chronology, Vol. 5, Stämpfli, 1968.

[2] I. Liiv, Seriation and matrix reordering methods: An historical overview, Statistical Analysis and Data Mining 3 (2) (2010) 70–91.

[3] D. Tsafrir, I. Tsafrir, L. Ein-Dor, O. Zuk, D. A. Notterman, E. Domany, Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices, Bioinformatics 21 (10) (2005) 2301–2308.

[4] T. C. Havens, J. C. Bezdek, An efficient formulation of the improved visual assessment of cluster tendency (iVAT) algorithm, IEEE Transactions on Knowledge and Data Engineering 24 (5) (2012) 813–822.

[5] Y.-J. Tien, Y.-S. Lee, H.-M. Wu, C.-H. Chen, Methods for simultaneously identifying coherent local clusters with smooth global patterns in gene expression profiles, BMC Bioinformatics 9 (1) (2008) 155.

[6] D. Mavroeidis, E. Bingham, Enhancing the stability and efficiency of spectral ordering with partial supervision and feature selection, Knowledge and Information Systems 23 (2) (2010) 243–265.

[7] J. Y. Goulermas, A. Kostopoulos, T. Mu, A new measure for analyzing and fusing sequences of objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (5) (2016) 833–848.

[8] A. Recanati, N. Servant, J.-P. Vert, A. d'Aspremont, Robust Seriation and Applications to Cancer Genomics, ArXiv e-prints arXiv:1806.00664.

[9] X. Evangelopoulos, A. J. Brockmeier, T. Mu, J. Y. Goulermas, Continuation methods for approximate large scale object sequencing, Machine Learning 108 (4) (2019) 595–626.

[10] M. Hahsler, An experimental comparison of seriation methods for one-mode two-way data, European Journal of Operational Research 257 (1) (2017) 133–143.

[11] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, B. Futcher, Comprehensive identification of cell cycle–regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization, Molecular Biology of the Cell 9 (12) (1998) 3273–3297.

[12] M. J. McDonald, M. Rosbash, Microarray analysis and organization of circadian gene expression in drosophila, Cell 107 (5) (2001) 567 – 578.

[13] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, A. Singer, Graph laplacian tomography from unknown random projections, IEEE Transactions on Image Processing 17 (10) (2008) 1891–1899.

[14] L. Hubert, P. Arabie, The analysis of proximity matrices through sums of matrices having (anti-)robinson forms, British Journal of Mathematical and Statistical Psychology 47 (1) (1994) 1–40.

[15] L. Hubert, J. Schultz, Quadratic assignment as a general data analysis strategy, British Journal of Mathematical and Statistical Psychology 29 (2) (1976) 190–241.

[16] A. Recanati, T. Kerdreux, A. d'Aspremont, Reconstructing Latent Orderings by Spectral Clustering, ArXiv e-prints arXiv:1807.07122.

[17] E. Çela, The quadratic assignment problem: theory and algorithms, Vol. 1, Springer Science & Business Media, 2013.

[18] L. Hubert, P. Arabie, J. Meulman, Graph-theoretic representations for proximity matrices through strongly-anti-robinson or circular strongly-anti-robinson matrices, Psychometrika 63 (4) (1998) 341–358.

[19] M. J. Brusco, S. Stahl, Compact integer-programming models for extracting subsets of stimuli from confusion matrices, Psychometrika 66 (3) (2001) 405–419.

[20] N. Christofides, E. Benavent, An exact algorithm for the quadratic assignment problem on a tree, Operations Research 37 (5) (1989) 760–768.

[21] E. M. Loiola, N. M. M. de Abreu, P. O. Boaventura-Netto, P. Hahn, T. Querido, A survey for the quadratic assignment problem, European Journal of Operational Research 176 (2) (2007) 657–690.

[22] J. T. Vogelstein, J. M. Conroy, V. Lyzinski, L. J. Podrazik, S. G. Kratzer, E. T. Harley, D. E. Fishkind, R. J. Vogelstein, C. E. Priebe, Fast approximate quadratic programming for graph matching, PLOS ONE 10 (4) (2015) 1–17.

[23] Y. Lu, K. Huang, C.-L. Liu, A fast projected fixed-point algorithm for large graph matching, Pattern Recognition 60 (2016) 971 – 982.

[24] D. E. Fishkind, S. Adali, H. G. Patsolic, L. Meng, D. Singh, V. Lyzinski, C. E. Priebe, Seeded graph matching, Pattern Recognition 87 (2019) 203 – 215.

[25] L. Guttman, A new approach to factor analysis: The radex., Mathematical thinking in the social sciences (1954) 258–348.

[26] L. Hubert, P. Arabie, J. Meulman, Linear and circular unidimensional scaling for symmetric proximity matrices, British Journal of Mathematical and Statistical Psychology 50 (2) (1997) 253–284.

[27] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373–1396.

[28] J. B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika 29 (1) (1964) 1–27.

[29] S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[30] T. Mu, J. Y. Goulermas, J. Tsujii, S. Ananiadou, Proximity-based frameworks for generating embeddings from multi-output data, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (11) (2012) 2216–2232.

[31] Y. Wu, T. Mu, P. Liatsis, J. Y. Goulermas, Computation of heterogeneous object co-embeddings from relational measurements, Pattern Recognition 65 (2017) 146 – 163.

[32] T. F. Cox, M. A. Cox, Multidimensional scaling on a sphere, Communications in Statistics - Theory and Methods 20 (9) (1991) 2943–2953.

[33] R. C. Wilson, E. R. Hancock, E. Pekalska, R. P. W. Duin, Spherical and hyperbolic embeddings of data, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (11) (2014) 2255–2269.

[34] M. Friendly, Corrgrams: Exploratory displays for correlation matrices, The American Statistician 56 (4) (2002) 316–324.

[35] C.-H. Chen, Generalized association plots: Information visualization via iteratively generated correlation matrices, Statistica Sinica (2002) 7–29.

[36] J. Liu, D. Lin, G. G. Yardımcı, W. S. Noble, Unsupervised embedding of single-cell Hi-C data, Bioinformatics 34 (13) (2018) i96–i104.

[37] D. Vicari, M. Vichi, Non-hierarchical classification structures, in: W. Gaul, O. Opitz, M. Schader (Eds.), Data Analysis: Scientific Modeling and Practical Application, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 51–65.

[38] M. Köppe, M. Queyranne, C. T. Ryan, Parametric integer programming algorithm for bilevel mixed integer programs, Journal of Optimization Theory and Applications 146 (1) (2010) 137–150.

[39] T. F. Cox, M. A. Cox, Multidimensional scaling, Chapman & Hall/CRC, 2000.

[40] D. E. Critchlow, Metric methods for analyzing partially ranked data, Vol. 34, Springer Science & Business Media, 2012.

[41] N. A. Crossley, A. Mechelli, P. E. Vértes, T. T. Winton-Brown, A. X. Patel, C. E. Ginestet, P. McGuire, E. T. Bullmore, Cognitive relevance of the community structure of the human brain functional coactivation network, Proceedings of the National Academy of Sciences 110 (28) (2013) 11583–11588.

[42] J.-M. Geusebroek, G. J. Burghouts, A. W. Smeulders, The amsterdam library of object images, International Journal of Computer Vision 61 (1) (2005) 103–112.

[43] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[44] H. Xiao, K. Rasul, R. Vollgraf, Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms, ArXiv e-prints arXiv:1708.07747.

[45] V. G. Deineko, G. J. Woeginger, A solvable case of the quadratic assignment problem., Operations Research Letters 22 (1) (1998) 13 – 17.