

Graph Connectivity In Sparse Subspace Clustering

Behrooz Nasihatkon, Richard Hartley
Australian National University
National ICT Australia (NICTA)*

behrooz.nasihatkon@anu.edu.au, richard.hartley@anu.edu.au

Abstract

Sparse Subspace Clustering (SSC) is one of the recent approaches to subspace segmentation. In SSC a graph is constructed whose nodes are the data points and whose edges are inferred from the L^1 -sparse representation of each point by the others. It has been proved that if the points lie on a mixture of independent subspaces, the graphical structure of each subspace is disconnected from the others. However, the problem of connectivity within each subspace is still unanswered. This is important since the subspace segmentation in SSC is based on finding the connected components of the graph. Our analysis is built upon the connection between the sparse representation through L^1 -norm minimization and the geometry of convex polytopes proposed by the compressed sensing community. After introduction of some assumptions to make the problem well-defined, it is proved that the connectivity within each subspace holds for 2- and 3-dimensional subspaces. The claim of connectivity for general d -dimensional case, even for generic configurations, is proved false by giving a counterexample in dimensions greater than 3.

1. Introduction

Subspace clustering is a very important problem with applications in many different areas in computer vision including motion segmentation [11], video shot segmentation [7], illumination invariant clustering [5], image segmentation [12] and image representation and compression [6]. SSC [3] is one of the state-of-the-art methods proposed for subspace segmentation with considerable advantages over the previous methods (see [10]). In SSC the subspace clustering is done based on the neighbourhood graph obtained by the L^1 -norm sparse representation of each point by the other points.

*NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

The basic SSC method works as follows:

Consider a set of points $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ in \mathbb{R}^D , sampled from a mixture of different subspaces such that no point lies on the origin. Considering these points as vectors, each \mathbf{x}_i can be obtained as a linear combination of the others:

$$\mathbf{x}_i = \sum_j a_j \mathbf{x}_j = \mathbf{X} \mathbf{a}, \quad \text{where } a_i = 0, \quad (1)$$

where \mathbf{X} is the matrix $[\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n]$ and \mathbf{a} is the vector of the combination coefficients a_j .

Of course, this combination (if it exists) is not unique in general. In SSC we are interested in a combination with smallest $\|\mathbf{a}\|_1$, i.e. for each \mathbf{x}_i the following is solved:

$$\mathbf{a}^i = \operatorname{argmin}_{\mathbf{a}} \|\mathbf{a}\|_1 \quad \text{s.t. } \mathbf{x}_i = \mathbf{X} \mathbf{a}, \quad a_i = 0 \quad (2)$$

The corresponding points of the nonzero elements of \mathbf{a}^i are set to be the neighbours of \mathbf{x}_i . Doing the same thing for every point forms a directed neighbourhood graph on X .

In [3] it has been proved that if the subspaces are independent, then the neighbours of each point would be in the same subspace. This means that there is no link between the graphs of two different subspaces. Based on this fact, a subspace segmentation method is proposed by finding the connected components of the neighbourhood graph.¹ However, in [3] it is not investigated whether the graphs within each subspace are connected or not. This paper seeks to answer this question.

To investigate the problem, first, an interpretation of the SSC is presented based on the geometry of convex polytopes [4]. The connection between the L^1 -sparse representation and convex polytopes was first noticed in [2], where both strong and weak L^1/L^0 equivalence is investigated using the properties of convex polytopes. In particular, it is shown that the basis vectors \mathbf{x}_i and their negatives $-\mathbf{x}_i$ forming vertices of a so-called k -neighbourly polytope is both necessary and sufficient for L^1/L^0 equivalence under some conditions on the sparse solution. In [8] the polar polytope of the centrosymmetric polytope with vertices

¹In practice this is done by spectral clustering.

$\pm \mathbf{x}_i$ is used to analyze the conditions for unique-optimality of the L^1 -norm minimization for sparse representation.

Based on this geometric representation, a proof of connectivity is given for 2D and 3D subspaces when the points are in general position. Graph connectivity for the general d -dimensional case, even in generic arrangements, is disproved by constructing a counterexample.

2. Basic conventions

In this document the term \mathbf{x}_i may either refer to a vector or the corresponding point in the Euclidean space. For the corresponding graph nodes, we use the indices of the points (i, j , etc.). By \mathcal{N}_i we refer to the set of neighbours of the node i . Therefore, $X_{\mathcal{N}_i}$ shows the set of neighbouring points of \mathbf{x}_i .

Here, when it is said that a point, line, etc. falls inside a region, it means that it can also be on the boundary. When we mean being contained in the (relative) interior of the region we use the term *strictly inside*. Similarly, when we say that a point, line, etc. is on one side of a hyperplane, we mean that it can lie also on the hyperplane, and when we want to exclude the hyperplane itself from each of the half-spaces it bounds, we use the term *strictly on one side*.

Here, \mathbf{O} denotes the origin and $\mathbf{ray}(\mathbf{x})$ refers to the half-line from the origin through the point \mathbf{x} . Also, $\mathbf{hull}(Y)$ refers to the convex hull of the set of points Y and $\mathbf{cone}(Y) = \{a\mathbf{y} \mid a \geq 0, \mathbf{y} \in \mathbf{hull}(Y)\}$ represents the convex cone generated by the set of points Y . If Y is a finite set, $\mathbf{polytope}(Y)$ refers to the polytope structure of $\mathbf{hull}(Y)$. $\mathbf{simplex}(Y)$ shows the m -dimensional simplex whose vertices are the points in the set Y , where $m = |Y|$.

To denote the unique hyperplane passing through a set of points Y , we use $\mathbf{h-plane}(Y)$. For a hyperplane not passing through \mathbf{O} , we use the terms *negative* and *positive sides* to respectively refer to the corresponding half-spaces including and not including \mathbf{O} .

3. Preparation

We add the negative points $-\mathbf{x}_i$ to the set of points X to obtain $X_{\pm} = \{\pm \mathbf{x}_i\}$. Indices $n+1, \dots, 2n$ are used to denote the added opposite points. Also, \hat{i} is used for the indices or graph nodes corresponding to $-\mathbf{x}_i$, for $i = 1, 2, \dots, 2n$.

To obtain the neighbourhood graph of X_{\pm} , for each point $\mathbf{x}_i \in X_{\pm}$ we solve:

$$\begin{aligned} \mathbf{a}^i &= \operatorname{argmin}_{\mathbf{a}} \mathbf{1}^T \mathbf{a} \\ \text{s.t. } \mathbf{x}_i &= \mathbf{X}_{\pm} \mathbf{a}, \quad \mathbf{a} \succeq \mathbf{0}, \quad a_i = 0 \end{aligned} \quad (3)$$

where the matrix $\mathbf{X}_{\pm} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{2n}]$ and $\mathbf{a} \succeq \mathbf{0}$ means \mathbf{a} has nonnegative elements. Again, the nodes corresponding to the nonzero elements of \mathbf{a}^i are set to be the neighbours of the node i .

It is known that (2) and (3) are equivalent problems [1]. For any combination $\mathbf{x}_i = \sum_{j \neq i} a_j \mathbf{x}_j$ by replacing $a_j \mathbf{x}_j$ for $-a_j \mathbf{x}_j$ whenever $a_j < 0$ we can find a solution with $\mathbf{a} \succeq \mathbf{0}$. It can be easily shown that for \mathbf{a}^i to be a solution to (3) must we have $a_i^i = 0$ for all i and for any j either a_j^i or a_j^i is zero. Based on this, it is not hard to show that the neighbourhood graph of X obtained by (2), can be achieved by identifying each pair of nodes (i, \hat{i}) in the neighbourhood graph of X_{\pm} obtained by (3) whenever the solution to (2) is unique for all i . When the solution is not unique for any i , there still is a one-to-one correspondence between the graphs obtained by (2) and (3) as long as the set of solutions $\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^{2n}$ to (3) is symmetric in the sense that $a_j^{\hat{i}} = a_j^i$. Then we have:

Proposition 1. *The neighbourhood graph of X is connected if and only if the extended neighbourhood graph of X_{\pm} is connected.*

Here, by the extended neighbourhood graph of X_{\pm} we mean the graph formed by adding edges between every pair of opposite nodes (i, \hat{i}) to the neighbourhood graph.

4. Geometric interpretation of SSC

Using (3) instead of (2) has the advantage of enjoying an interesting geometric interpretation. This property has been noticed by the compressed sensing community for sparse representations (see [2] and [8]). In (3) the search is only done on the conical combination of the points in $X_{-i} \stackrel{\text{def}}{=} X_{\pm} - \{\mathbf{x}_i\}$. In fact we can rewrite the constraints in (3) as:

$$\mathbf{x}_i = \mathbf{X}_{-i} \mathbf{b} = \mathbf{X}_{-i} \frac{\mathbf{b}}{\mathbf{1}^T \mathbf{b}} \cdot \mathbf{1}^T \mathbf{b} = \mathbf{X}_{-i} \mathbf{p} \alpha, \quad (4)$$

where \mathbf{X}_{-i} is the matrix corresponding to X_{-i} and \mathbf{b} can be thought of as \mathbf{a} in (3) with the i -th element removed. Thus, both $\mathbf{b}, \mathbf{p} \in \mathbb{R}^{2n-1}$ have nonnegative elements, $\mathbf{1}^T \mathbf{p} = 1$ and α is a positive scalar. Evidently the set of points $\{\mathbf{X}_{-i} \mathbf{p} \mid \mathbf{p} \succeq \mathbf{0}, \mathbf{1}^T \mathbf{p} = 1\}$ forms $\mathbf{hull}(X_{-i})$, i.e. the convex hull of the points X_{-i} . By taking $\beta = \frac{1}{\alpha}$, equation (3) turns to:

$$\text{maximize } \beta \quad \text{subject to } \beta \mathbf{x}_i \in \mathbf{hull}(X_{-i}) \quad (5)$$

Therefore, we are seeking a vector $\beta_{\max} \mathbf{x}_i$ on $\mathbf{ray}(\mathbf{x}_i)$ and contained in the convex hull of X_{-i} whose length is as large as possible. Thus:

Proposition 2. *If β_{\max} is the solution to (5), then $\mathbf{y}_i \stackrel{\text{def}}{=} \beta_{\max} \mathbf{x}_i$ is the point on the boundary of the convex hull where $\mathbf{ray}(\mathbf{x}_i)$ exits from the convex hull.*

This fact is illustrated in fig. 1.

For the finite set of points X_{-i} , the convex hull is a bounded polytope. Thus, \mathbf{y}_i lies in the relative interior of a unique k -face of this polytope, where $k \in \{0, 1, \dots, d-1\}$.

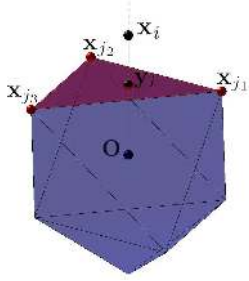


Figure 1. $\text{ray}(\mathbf{x}_i)$ (the dashed line) intersects a face of the polytope at \mathbf{y}_i . The extremal points \mathbf{x}_{j_1} , \mathbf{x}_{j_2} and \mathbf{x}_{j_3} on the intersected face are the neighbours of \mathbf{x}_i

Writing the point \mathbf{y}_i as a convex combination of the points X_{-i} , all the combination coefficients will be zero except those corresponding to the points on the mentioned k -face. This means that $X_{\mathcal{N}_i}$, the set neighbouring points of \mathbf{x}_i , is a subset of the points in X_{-i} lying on the k -face. If there are only $k+1$ points of X_{-i} on the k -face, then the k -face is a k -simplex having these points as its vertices. In this case, these $k+1$ vertices are the unique set of neighbours of \mathbf{x}_i (see [2]).

5. Connectivity on a single subspace

5.1. Statement of the problem

The core of the SSC method is the fact that in the SSC graph there is no link between points belonging to different subspaces. Here, we want to study the graph connectivity in a single d -dimensional subspace of \mathbb{R}^D with $n > d$ points in it. Without loss of generality, we can assume that this subspace is \mathbb{R}^d and the set of points on this subspace is again shown by $X = \{\mathbf{x}_i\}$ with $\mathbf{x}_i \in \mathbb{R}^d$.

At this point, we have to make clear what we mean by a single subspace. It is obvious that the points in this single subspace cannot have any arbitrary configuration. At least a single subspace must not be divisible into smaller subspaces.

If the points are in general position, each point has a set of d point as its unique neighbours. Otherwise, there are degenerate cases in which each point has fewer than d neighbours resulting in a disconnected graph (see fig. 2). However, these counterexamples are of no practical significance as they are non-generic cases which are removed by applying any small random perturbation to the set of points. To avoid such cases the following assumption is made:

Assumption 1. *No d points of X lie in a $(d-1)$ -dimensional subspace.*

This assumption implies that no subset of X_{\pm} not containing two opposite points lies on a $(d-1)$ -dimensional subspace. It also requires that no k points of X can lie in a

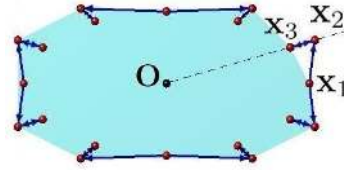


Figure 2. A counterexample to the graph connectivity conjecture for the case where there are d points of X lying on a d -dimensional subspace. Here, \mathbf{x}_2 and \mathbf{x}_3 are on a line passing through \mathbf{O} . $\text{ray}(\mathbf{x}_2)$ (the dashed line) intersects the boundary of $\text{polytope}(X_{-2})$ (the shaded region) on the vertex \mathbf{x}_3 and therefore \mathbf{x}_3 is the only neighbour of \mathbf{x}_2 . Likewise, \mathbf{x}_2 is the only neighbour of \mathbf{x}_3 . Arrows show the neighbourhood relation on the set of points.

$(k-1)$ -dimensional subspace for $k \leq d$. Consequently, no subset of points can form a k -dimensional subspace, $k < d$. It is true as well for 0-dimensional subspaces as it does not allow \mathbf{O} to be among the set of points. Thus, this assumption implies the condition that X cannot be non-trivially decomposed into disjoint subspaces of smaller dimensions.

The main consequence of assumption 1 is that if \mathbf{x}_i is represented as a linear sum, as in (1), then there must be at least d nonzero coefficients a_j . Returning to our geometric representation of the problem, assumption 1 demands that the vector \mathbf{y}_i (introduced in proposition 2) must lie in the interior of a $(d-1)$ -face (a facet) of $\text{polytope}(X_{-i})$. In other words, each point has at least d neighbours.

In a generic configuration of the points, the intersected facet has only d points of X_{\pm} on it which have to be its vertices. In this case (3) has a unique solution and those d points form the unique set of neighbours of \mathbf{x}_i . If there are more than d points on the intersected facet, whether they are vertices or not, the solution to (3) and hence $X_{\mathcal{N}_i}$, the set of neighbours of \mathbf{x}_i , is not unique. To avoid this, we make an extra assumption:

Assumption 2. *No $d+1$ points of X lie on a $d-1$ dimensional affine subspace (a hyperplane),*

which is sufficient for uniqueness of the neighbours.

Assumptions 1 and 2 are not practically restricting because they hold if the points are in general position. In fact, we can avoid assumption 2 by setting the neighbouring points of \mathbf{x}_i to all the points lying on the intersected facet, and all the conclusions in the rest of the paper still hold with slight adjustments. However, for simplicity, we retain this assumption in the sequel. To sum up, we can say that:

Lemma 3. *Under assumptions 1 and 2, the unique set of neighbours of each point \mathbf{x}_i is the set of d vertices of the facet of $\text{polytope}(X_{-i})$ intersected by $\text{ray}(\mathbf{x}_i)$.*

Corollary 4. *$h\text{-plane}(X_{\mathcal{N}_i})$ has all the points in X_{-i} on its negative side. (see sec. 2)*

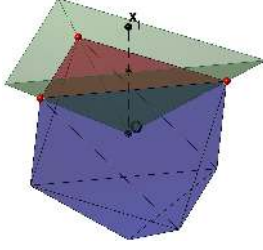


Figure 3. The (boundary of) the neighbourhood cone of \mathbf{x}_i .

Proof. Since the hyperplane $\mathbf{h}\text{-plane}(X_{\mathcal{N}_i})$ corresponds to a facet of $\mathbf{polytope}(X_{-i})$, it has all of the volume of the polytope on one side of it. The origin \mathbf{O} is strictly inside the polytope and therefore on the same side of the hyperplane. This means that this side is the negative side of the hyperplane. \square

Note that \mathbf{x}_i itself can be on either side of $\mathbf{h}\text{-plane}(X_{\mathcal{N}_i})$. From corollary 4 it is immediate that the point $-\mathbf{x}_i$ falls strictly on the negative side of $\mathbf{h}\text{-plane}(X_{\mathcal{N}_i})$ and cannot be a neighbour of \mathbf{x}_i . Also, the points \mathbf{x}_j and $-\mathbf{x}_j$ cannot both be neighbours of \mathbf{x}_i because otherwise it would require \mathbf{O} to be on $\mathbf{h}\text{-plane}(X_{\mathcal{N}_i})$.

5.2. Neighbourhood cones

In this section we introduce the concept of neighbourhood cones and present a theorem which plays a crucial role in proving our main results.

We define the *neighbourhood cone* of each point \mathbf{x}_i to be $C_{\text{ng}}(\mathbf{x}_i) = \mathbf{cone}(X_{\mathcal{N}_i})$, i.e. the convex cone generated by the neighbouring points of \mathbf{x}_i . From the discussion in sec. 4 we can conclude that assumption 1 requires $C_{\text{ng}}(\mathbf{x}_i)$ to have the entire $\mathbf{ray}(\mathbf{x}_i)$ (except \mathbf{O}) strictly inside it. Fig. 3 shows the neighbourhood cone of a point \mathbf{x}_i .

Neighbourhood cones are worth considering since the intersection between each pair of them can give useful information about connectivity of their corresponding graph nodes. This is described in the next theorem:

Theorem 5. *Two points \mathbf{x}_i and \mathbf{x}_j in X_{\pm} are neighbours in the neighbourhood graph of X_{\pm} if and only if their neighbourhood cones strictly intersect.*

By two cones strictly intersecting we mean that their intersection has nonempty interior.

Proof. First, assume that \mathbf{x}_j is a neighbouring point of \mathbf{x}_i . Then the line segment from \mathbf{x}_i to \mathbf{x}_j except at \mathbf{x}_j lies in the interior of $C_{\text{ng}}(\mathbf{x}_i)$. Now, \mathbf{x}_j is strictly inside $C_{\text{ng}}(\mathbf{x}_j)$. Thus, there is an open ball centred at \mathbf{x}_j being contained in the interior of $C_{\text{ng}}(\mathbf{x}_j)$. Evidently, the line segment from \mathbf{x}_j to \mathbf{x}_i intersects this ball in points other than \mathbf{x}_j and these points are in the interior of both cones. Taking one of these

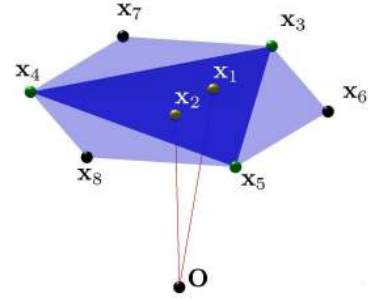


Figure 4. In this figure all the points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_8$ lie on a single facet of $\mathbf{polytope}(X_{\pm})$. Both $\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$ and $\{\mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8\}$ are candidate neighbours for \mathbf{x}_1 and also for \mathbf{x}_2 . If $\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$ and $\{\mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8\}$ are chosen to be neighbours of \mathbf{x}_1 and \mathbf{x}_2 respectively, then $\mathbf{cone}(\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\})$ and $\mathbf{cone}(\{\mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8\})$ strictly intersect, while \mathbf{x}_1 and \mathbf{x}_2 are not neighbours. This shows the reliance of theorem 5 on assumption 2.

points and considering the open ball around it being in the interior of both points shows that the intersection of the two cones must have an interior.

Now, assume there is a point \mathbf{x} strictly inside the intersection of the two cones. In this case, there exists an open ball contained in both cones centred at \mathbf{x} . It means that a set of d points $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_d$ not lying on any $(d-1)$ -dimensional subspace can be found inside the ball. Since the intersection of two cones is again a cone, for each \mathbf{z}_k , $\mathbf{ray}(\mathbf{z}_k)$ is entirely inside the intersection cone. For each \mathbf{z}_k , $\mathbf{ray}(\mathbf{z}_k)$ must intersect both $\mathbf{simplex}(X_{\mathcal{N}_i})$ and $\mathbf{simplex}(X_{\mathcal{N}_j})$. According to assumption 2, $\mathbf{simplex}(X_{\mathcal{N}_i})$ and $\mathbf{simplex}(X_{\mathcal{N}_j})$ do not lie on a common hyperplane, hence at least for one k , $\mathbf{ray}(\mathbf{z}_k)$ intersect the two simplices at different points. Without loss of generality, we say that $\mathbf{ray}(\mathbf{z}_k)$ has intersected $\mathbf{simplex}(X_{\mathcal{N}_j})$ at a less distance from \mathbf{O} than $\mathbf{simplex}(X_{\mathcal{N}_i})$. This means that the intersected point on $\mathbf{simplex}(X_{\mathcal{N}_i})$ is strictly on the positive side of $\mathbf{h}\text{-plane}(X_{\mathcal{N}_j})$. It follows that at least one of the extremal points of $\mathbf{simplex}(X_{\mathcal{N}_i})$, i.e. one of the points in $X_{\mathcal{N}_i}$, must be strictly on the positive side of $\mathbf{h}\text{-plane}(X_{\mathcal{N}_j})$. But, according to corollary 4 this point can only be \mathbf{x}_j , and hence, \mathbf{x}_j has to be a neighbour of \mathbf{x}_i . \square

Theorem 5 is very useful as it establishes a link between the connectivity of the neighbourhood graph and the connectivity of the neighbourhood cones of the set of points. For example, it tells us that for a single connected component C_1 of the neighbourhood graph, the interior of $\cup_{i \in C_1} C_{\text{ng}}(\mathbf{x}_i)$ forms a connected region. For two distinct connected components C_1 and C_2 the interiors of $\cup_{i \in C_1} C_{\text{ng}}(\mathbf{x}_i)$ and $\cup_{i \in C_2} C_{\text{ng}}(\mathbf{x}_i)$ are disjoint.

Notice that the results obtained in theorem 5 is not true without assumption 2 (see Fig. 4.)

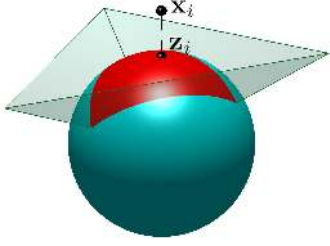


Figure 5. Projection on the unit hypersphere. The shaded region shows $S_{\text{ng}}(\mathbf{z}_i)$, the neighbourhood simplex of \mathbf{z}_i , which is the spherical triangle formed by projecting $C_{\text{ng}}(\mathbf{x}_i)$ on S^2 .

5.3. Projection on the unit hypersphere

The geometric structure made by a set of cones can be further compressed by projecting them on a hypersphere centred at the origin. Here, for simplicity we consider the unit hypersphere S^{d-1} .

The projection is given by $\mathbf{x} \mapsto \frac{\mathbf{x}}{\|\mathbf{x}\|}$ for each point \mathbf{x} in the space. Let \mathbf{z}_i denote the projected point of \mathbf{x}_i . The projection of the neighbourhood cone of each point \mathbf{x}_i on the hypersphere, which is the intersection of the surface of the hypersphere with that cone, forms a *hyperspherical simplex* called the neighbourhood simplex of \mathbf{z}_i and is denoted by $S_{\text{ng}}(\mathbf{z}_i)$. This is shown in fig. 5 for S^2 .

It has to be noticed that the neighbourhood structure is built before applying the projection, i.e. the optimization problem (3) is run upon $\{\mathbf{x}_i\}$ and not $\{\mathbf{z}_i\}$.

As the neighbourhood cones strictly intersecting is equivalent to their corresponding neighbourhood simplices strictly intersecting, restricting ourselves to S^{d-1} , most of the interesting properties mentioned about the neighbourhood cones hold. The point \mathbf{z}_i must lie strictly inside $S_{\text{ng}}(\mathbf{z}_i)$ and theorem 5 obtains by substituting neighbourhood simplices for neighbourhood cones. Again, for a connected component C of the neighbourhood graph, the interior of $\cup_{i \in C} S_{\text{ng}}(\mathbf{z}_i)$ forms a connected region, disjoint from that of the other connected components.

This projection is useful as it decreases the dimensionality by one and eliminates the intrinsic redundancy of the convex cones. As far as the connectivity of the cones is concerned, we can forget about the cones $C_{\text{ng}}(\mathbf{x}_i)$ and work with the simpler structures $S_{\text{ng}}(\mathbf{x}_i)$.

5.4. Proof of connectivity for 2D

The proof of connectivity of the neighbourhood graph for 2D subspaces is quite simple and the concept of projection of the neighbourhood cones on the unit circle S^1 makes it even easier. Here the projected neighbourhood simplices are in the form of arcs on the circumference of the circle.

Notice that being strictly inside its neighbourhood arc, each projected point \mathbf{z}_i must have its two neighbours on left

and right sides of it on the unit circle. By left and right here we mean clockwise and anticlockwise directions respectively.

It is easily shown that having more than one connected component is impossible. A connected component means the set of points in X_{\pm} or Z_{\pm} corresponding to the nodes of a single connected component of the neighbourhood graph.

Consider a single connected component C_1 . Since each point of C_1 has a left neighbour and the points are finite in number, there must be a loop $\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \dots, \mathbf{z}_{i_m}, \mathbf{z}_{i_{m+1}}$, where $i_{m+1} = i_1$ and $\mathbf{z}_{i_{k+1}}$ is the left neighbour of \mathbf{z}_{i_k} . It is clear that the circumference of the circle from \mathbf{z}_{i_k} to its left neighbour $\mathbf{z}_{i_{k+1}}$ (not including $\mathbf{z}_{i_{k+1}}$) is strictly inside the neighbouring arc of \mathbf{z}_{i_k} , and thus, the neighbourhood arcs of the points in this loop occupy all the circumference of the circle. Therefore, if another connected component coexists with C_1 , the neighbourhood arc of each of its points strictly intersect with one of the arcs in C_1 and this is impossible due to theorem 5.

5.5. A proof for 3D

In 3D the neighbourhood simplices (triangles) of one connected component may not occupy the whole surface of the unit sphere and the proof of connectivity is harder than in 2D.

As suggested in sec. 5.3, the points in X_{\pm} and the cone structures are projected on S^2 . The neighbourhood cone of each point \mathbf{x}_i , is projected on S^2 as a spherical triangle called the neighbourhood triangle of \mathbf{z}_i . It follows from theorem 5 that if two nodes belong to different connected components, their neighbourhood triangles can never intersect.

The point \mathbf{z}_i strictly lies inside its neighbourhood triangle and the points \mathbf{z}_{N_i} lie on the corners of the triangle. In the next two lemmas, a single connected component and its corresponding set of neighbourhood triangles are considered.

Lemma 6. *Consider a single connected component C_1 of the neighbourhood graph. Then $S^2 - \cup_{i \in C_1} S_{\text{ng}}(\mathbf{z}_i)$ consists of a (possibly empty) set of regions each of which is topologically an open disk.*

Proof. Since the spherical triangles $S_{\text{ng}}(\mathbf{z}_i)$ are finite and all closed then $S^2 - \cup_{i \in C_1} S_{\text{ng}}(\mathbf{z}_i)$ is open and consequently each of its connected subsets is open. Consider one connected area on the surface of the sphere not containing any spherical triangles. This open region must be bounded by the boundary of spherical triangles. If this area is not topologically a disk, it has to have more than one boundary. But, this means that the spherical triangles related to one boundary are disconnected from those related to the other boundary, which is impossible due to theorem 5 as the neigh-

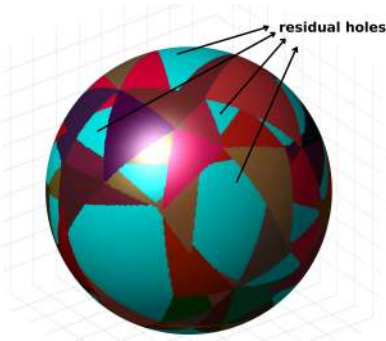


Figure 6. The neighbourhood triangles of a single connected component and the residual holes left on the surface of the sphere.

neighbourhood triangles correspond to a single connected component. \square

Here, we refer to these remaining regions as *residual holes* of a connected component. Fig. 6 shows the residual holes left on the sphere after projecting the neighbourhood triangles.

Lemma 7. *Each of the residual holes of a connected component is a spherical polygon whose internal angles are each less than π radians.*

Proof. Each hole is bounded by parts of the boundaries of neighbourhood triangles. This means that the boundary is piecewise geodesic, and each hole is a spherical polygon.

Recall from sec. 5.2 that the neighbourhood cone of \mathbf{x}_i has the whole $\text{ray}(\mathbf{x}_i)$, except \mathbf{O} , strictly inside it, and thus each corner of any triangle is strictly inside another triangle and cannot exist on the boundary of the holes. Therefore, all the angles on the boundary have to be created by the intersection of two geodesics each being a side of some triangle. The intersection creates four angles each smaller than or equal to π radians. Since the intersection is not on the corner of any triangle, in a sufficiently small neighbourhood around the intersection point, the regions corresponding to three of these four angles are occupied by some neighbourhood triangles. The internal angle of the boundary of the residual hole is thus one of the four angles at the intersection and hence is $\leq \pi$ radians.

Not all of the angles can be equal to π radians, since otherwise, some triangle corners must lie on the boundary. We can disregard possible π radian angles and say that all the angles of the polygon are less than π radians. \square

Lemma 8. *The area of each residual hole of one connected component is less than the area of a half-sphere, i.e. 2π .*

Proof. Being topologically a disk on the surface of a sphere, the closure of each residual hole is a compact two-dimensional Riemannian manifold and thus, the Gauss-

Bonnet theorem [9] can be applied. Applying the Gauss-Bonnet theorem to a residual hole which is a spherical polygon gives:

$$A = 2\pi - \sum_i \alpha_i \quad (6)$$

where A is the total area of the hole and α_i -s are the *jump angles* on the boundary, which are here equal to the external angles of the spherical polygon. Since the internal angles of the hole are less than π radians, the external angles α_i are more than zero and thus the area is less than a half-sphere (i.e. 2π). \square

Theorem 9. *Under assumptions 1 and 2, the neighbourhood graph of every set of points in \mathbb{R}^3 made by the SSC method is connected.*

Proof. Assume there is more than one connected component, call two of them A and B ; Since, neighbourhood triangles of each of the connected components cannot intersect, all neighbourhood triangles of the connected component B must fall inside one of the residual holes of the connected component A . Let us show this hole by \mathcal{H} . Since the area of \mathcal{H} is less than a half-sphere, the area of $S^2 - \mathcal{H}$ has to be more than a half-sphere. But, since the connected component B also leaves holes on the sphere, and the entire neighbourhood triangles of B lie inside \mathcal{H} , $S^2 - \mathcal{H}$ has to be contained in one of the holes of B . But, this requires the area of that hole of B to be bigger than a half-sphere, which is impossible due to lemma 8. \square

5.6. A counterexample for dimensions ≥ 4

We observed that the restrictions imposed by the neighbourhood simplices of a single connected component in 3D were less severe compared to 2D, as residual holes were not allowed in 2D. One can expect that in higher dimensions the restrictions would be more relaxed and at some point, they are weak enough to allow more than one connected component. In fact, this break point is 4D. Here a simple counterexample is illustrated and proved for 4D. For greater than 4 dimensions the construction of the counterexample is given without a proof.

We consider two sets of points aligned around (not on) two non-intersecting great circles $[\cos \alpha, \sin \alpha, 0, 0]$ and $[0, 0, \cos \alpha, \sin \alpha]$ of the hypersphere $\|\mathbf{x}\|^2 = 1$. The points are $X_{\pm} = X_{C_1} \cup X_{C_2}$ where X_{C_1} consists of the points $[\cos \theta_k, \sin \theta_k, s\delta, s'\delta]^T$, with $\theta_k = k\pi/m$ for the positive integer m , for all $k=0, 1, \dots, 2m-1$ and all $s, s' \in \{-1, 1\}$. Here, δ is a positive constant. Thus, there are a total of $8m$ points in X_{C_1} . Similarly, points in X_{C_2} are $[s\delta, s'\delta, \cos \theta_k, \sin \theta_k]^T$. Notice that the data points are aligned on the hypersphere $\|\mathbf{x}\|^2 = 1 + 2\delta^2$ and the negative of each point of each X_{C_j} is also in X_{C_j} for $j = 1, 2$.

Here we show that for some values of m and δ the four neighbours of each point $[\cos \theta_k, \sin \theta_k, s\delta, s'\delta]^T$ are:

$$\begin{aligned} & [\cos \theta_{k+1}, \sin \theta_{k+1}, s\delta, s'\delta]^T \\ & [\cos \theta_{k-1}, \sin \theta_{k-1}, s\delta, s'\delta]^T \\ & [\cos \theta_k, \sin \theta_k, -s\delta, s'\delta]^T \\ & [\cos \theta_k, \sin \theta_k, s\delta, -s'\delta]^T. \end{aligned}$$

It means that the neighbours of each node in C_1 remain in C_1 . Giving the similar arguments for C_2 shows that C_1 and C_2 are disconnected.

Since the algorithm is not sensitive to rotation and reflection, for a typical point $[\cos \theta_k, \sin \theta_k, s\delta, s'\delta]^T$ in X_{C_1} , we make the last two coordinate positive by possible reflections across their axes and then rotate the first two coordinates by $-\theta_k$ to obtain $\mathbf{x}_i = [1, 0, \delta, \delta]^T$. The whole set of points in X_{C_1} and X_{C_2} is reflected and rotated accordingly. Note that X_{C_1} is closed under the mentioned rotation and reflection while X_{C_2} is only closed under the reflection. Hence, we have to show that the neighbours of \mathbf{x}_i are:

$$\begin{aligned} \mathbf{x}_1 &= [\cos \theta_1, \sin \theta_1, \delta, \delta]^T \\ \mathbf{x}_2 &= [\cos \theta_1, -\sin \theta_1, \delta, \delta]^T \\ \mathbf{x}_3 &= [1, 0, -\delta, \delta]^T \\ \mathbf{x}_4 &= [1, 0, \delta, -\delta]^T \end{aligned}$$

First, observe that \mathbf{x}_i is strictly inside $\mathbf{cone}(X_N)$, where $X_N = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$. It means that \mathbf{x}_i is on the same side of $\mathbf{h-plane}(\{\mathbf{O}\} \cup (X_N - \{\mathbf{x}_j\}))$ as \mathbf{x}_j for $j = 1, 2, 3, 4$, i.e. $\mathbf{a}_j^T \mathbf{x}_i > 0$ where \mathbf{a}_j is the solution to $\mathbf{a}_j^T \mathbf{X}_N = \mathbf{e}_j^T$, with \mathbf{e}_j being the j -th column of the 4×4 identity matrix and $\mathbf{X}_N = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4]$. In other words, all elements of $\mathbf{X}_N^{-1} \mathbf{x}_i$ are positive. We have:

$$\mathbf{X}_N^{-1} = \frac{1}{4d} \begin{bmatrix} 0 & 2d/\sin \theta_1 & 1 & 1 \\ 0 & -2d/\sin \theta_1 & 1 & 1 \\ 2\delta & 0 & -1 - \cos \theta_1 & 1 - \cos \theta_1 \\ 2\delta & 0 & 1 - \cos \theta_1 & -1 - \cos \theta_1 \end{bmatrix}$$

We see that $\mathbf{X}_N^{-1} \mathbf{x}_i = \frac{1}{2}[1, 1, 1 - \cos \theta_1, 1 - \cos \theta_1]$ has positive elements for every positive value of m , and hence, \mathbf{x}_i is strictly inside $\mathbf{cone}(X_N)$.

Now, consider the hyperplane $\mathbf{h-plane}(X_N)$. We show that for large enough values of m and a range of values for δ , this hyperplane has all the points except $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ and \mathbf{x}_i strictly on its negative side and hence $\mathbf{x}_1, \dots, \mathbf{x}_4$ are vertices of a facet of $\mathbf{polytope}(X_{-i})$. The hyperplane $\mathbf{h-plane}(X_N)$ has the equation $\mathbf{a}^T \mathbf{y} = \mathbf{1}^T \mathbf{X}_N^{-1} \mathbf{y} = 1$ and for points strictly on its negative side we have $\mathbf{a}^T \mathbf{y} < 1$. Now, $\mathbf{a}^T = \frac{1}{2\delta}[2\delta, 0, 1 - \cos \theta_1, 1 - \cos \theta_1]$.

First we check for points in X_{C_1} . Since X_{C_1} is closed under reflections across the axes of the last two coordinates

and also rotations by $-\theta_k$ on the first two coordinates, we only check it for a typical point $[\cos \theta_k, \sin \theta_k, s\delta, s'\delta]^T$ in X_{C_1} :

$$\begin{aligned} \mathbf{a}^T [\cos \theta_k, \sin \theta_k, \delta, \delta]^T &= \cos \theta_k - \cos \theta_1 + 1 \\ \mathbf{a}^T [\cos \theta_k, \sin \theta_k, \pm\delta, \mp\delta]^T &= \cos \theta_k \\ \mathbf{a}^T [\cos \theta_k, \sin \theta_k, -\delta, -\delta]^T &= \cos \theta_k + \cos \theta_1 - 1 \end{aligned}$$

which is equal to 0 for $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and \mathbf{x}_4 , > 0 for \mathbf{x}_i and < 0 for all the other points in X_{\pm} .

The set X_{C_2} is closed under reflection across its last two coordinates but not rotation by $-\theta_k$ on the first two coordinates. Therefore, we first apply this rotation to a point $[s\delta, s'\delta, \cos \theta_l, \sin \theta_l]^T$ to obtain $\mathbf{y} = [(s \cos \theta_k + s' \sin \theta_k)\delta, (s' \cos \theta_k - s \sin \theta_k)\delta, \cos \theta_l, \sin \theta_l]^T$ and then check for this new rotated point:

$$\mathbf{a}^T \mathbf{y} = (s \cos \theta_k + s' \sin \theta_k)\delta + \frac{(\cos \theta_l + \sin \theta_l)(1 - \cos \theta_1)}{2\delta}.$$

For $(s \cos \theta_k + s' \sin \theta_k)$ and $(\cos \theta_l + \sin \theta_l)$, $\sqrt{2}$ is a tight upper bound, and hence, the above is less than 1 if $\sqrt{2}\delta^2 - \delta + \frac{\sqrt{2}}{2}(1 - \cos \theta_1) < 0$, i.e. $\cos \theta_1 > \frac{3}{4}$ and $|\delta - \frac{\sqrt{2}}{4}| < (\frac{1}{8} - \frac{1}{2}(1 - \cos \theta_1))^{\frac{1}{2}}$. It means that the condition holds for $\delta \in (\Delta, \frac{\sqrt{2}}{2} - \Delta)$, where $\Delta = \frac{\sqrt{2}}{4} - (\frac{1}{8} - \frac{1}{2}(1 - \cos \theta_1))^{\frac{1}{2}}$. Notice that $\cos \theta_1 > \frac{3}{4}$ can be obtained and also Δ can be made small enough by choosing a sufficiently large m .

The above argument tells us that the points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and \mathbf{x}_4 must be vertices of a facet of $\mathbf{polytope}(X_{-i})$. As \mathbf{x}_i is strictly inside $\mathbf{cone}(X_N)$, the interior of this facet is intersected by $\mathbf{ray}(\mathbf{x}_i)$ and hence X_N is the set of neighbours of \mathbf{x}_i . It is the unique set of neighbours because it has only $d = 4$ points on it.

Fig. 7 shows an illustration of the described example. Fig. 7(c) is worth considering as the distance between the two connected components is small. As δ can be made close enough to $\sqrt{2}/2$, the distance between the two sets of circles $[\cos \alpha, \sin \alpha, s\delta, s'\delta]^T$ and $[t\delta, t'\delta, \cos \beta, \sin \beta]^T$ on which our data points lie can be arbitrarily close to zero, e.g. at the two point where $\alpha = \beta = \frac{\pi}{4}$ and $s = s' = t = t' = 1$. This shows that the data points being close together (i.e. being on a trajectory) does not necessarily imply connectedness.

To show that this counterexample is not non-generic (nowhere dense) note that for each point \mathbf{x}_i , $\mathbf{ray}(\mathbf{x}_i)$ intersects the interior of a facet of $\mathbf{polytope}(X_{-i})$. This means that the neighbourhood structure is preserved if all the points have an arbitrary perturbation within a small enough ball. Therefore, the graph is disconnected at least within one open set of the space of all the points and thus cannot be nowhere dense. In other words, the set of point arrangements with a connected SSC graph in 4D cannot be dense and hence is not generic.

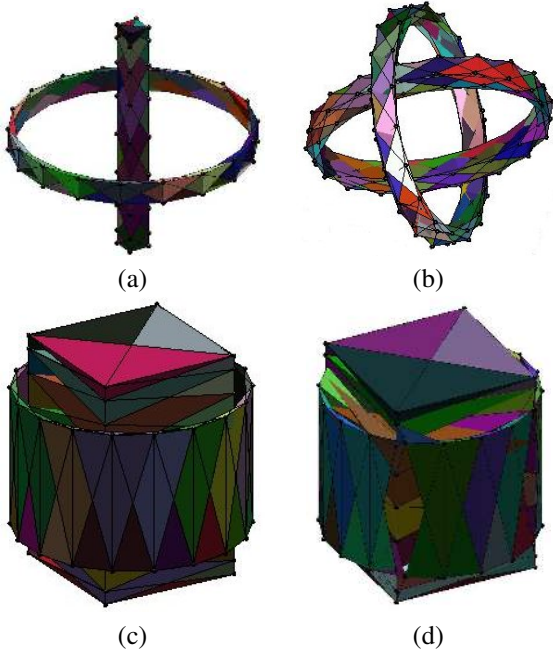


Figure 7. The orthographic projection of $\cup_i S_{ng}(\mathbf{z}_i)$ on the 4th coordinate axis, for $m = 9$ ($\Delta \approx 0.3491$). For simplicity, the 1-skeleton of the hyperspherical tetrahedra $S_{ng}(\mathbf{z}_i)$ after projection is shown as straight lines. (a) $\delta = \Delta + \epsilon$, (b) $\delta = \Delta + \epsilon$ and rotated before projection, (c) $\delta = \frac{\sqrt{2}}{2} - \Delta - \epsilon$, the two connected components are quite close together, and (d) $\delta = \frac{\sqrt{2}}{2} - \Delta + \epsilon$, there is just one connected component.

Here, we did not mention assumptions 1 and 2 because their purpose was showing the connectivity in 2- and 3-dimensional cases for generic arrangements. In 4D the connectivity is not a generic property, however, as these assumptions are true for an open and dense set of point configurations and our counterexample holds in an open ball, there must exist an open ball in which the counterexample exists and assumptions 1 and 2 are satisfied.

A similar counterexample can be made for higher dimensional spaces by arranging the data around non-intersecting great circles of a hypersphere. For example two sets of points $[\cos \theta_k, \sin \theta_k, s_1 \delta, s_2 \delta, \dots, s_{d-2} \delta]^T$, and $[s_1 \delta, s_2 \delta, \cos \theta_k, \sin \theta_k, s_3 \delta, \dots, s_{d-2} \delta]^T$. The structure of the proof is similar to that of 4D, however it is slightly more intricate.

6. Conclusion

The paper investigated the problem of connectivity in the SSC method when the points lie on a single subspace and cannot be segmented into smaller subspaces. The treatment was based on a geometric interpretation of the sparse representation problem. The connectivity of a set of nodes in SSC is then converted to the connectivity of the region formed by the union of the neighbouring cones. It was

shown that, if the points are in general position, the restriction imposed by the neighbouring cones of one connected component prevents the existence of extra connected components for 2- and 3-dimensional subspaces and hence, the graph is connected in these cases. However, the conjecture of connectivity was rejected in the general d -dimensional case by providing counterexamples.

As a conclusion, we can say that it is possible for the SSC algorithm to over-segment subspaces for dimensions higher than 3. It has to be studied whether some simple treatments such as adding a post-processing stage can resolve this intrinsic drawback, especially in the presence of noise and outliers. The next step is to seek for a variant of the SSC not suffering from this disadvantage.

References

- [1] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.
- [2] D. L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report, Department of Statistics, Stanford University, 2005.
- [3] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2797. IEEE, June 2009.
- [4] B. Grunbaum. *Convex polytopes (Graduate Texts in Mathematics)*. Springer, second edition, 2003.
- [5] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–11 – I–18 vol.1, 2003.
- [6] W. Hong, J. Wright, K. Huang, and Y. Ma. A multiscale hybrid linear model for lossy image representation. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 764 – 771 Vol. 1, 2005.
- [7] L. Lu and R. Vidal. Combined central and subspace clustering for computer vision applications. In *Proceedings of the 23rd international conference on Machine learning, ICML '06*, pages 593–600, New York, NY, USA, 2006. ACM.
- [8] M. Plumbley. On polar polytopes and the recovery of sparse representations. *Information Theory, IEEE Transactions on*, 53(9):3188–3195, 2007.
- [9] M. Spivak. *A comprehensive introduction to differential geometry. Vol. III*. Publish or Perish Inc., Wilmington, Del., second edition, 1979.
- [10] R. Vidal. A tutorial on subspace clustering. *IEEE Signal Processing Magazine*, to appear.
- [11] R. Vidal, R. Tron, and R. Hartley. Multiframe motion segmentation with missing data using powerfactorization and gpca. *International Journal of Computer Vision*, 79:85–105, 2008. 10.1007/s11263-007-0099-z.
- [12] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry. Unsupervised segmentation of natural images via lossy data compression. *Comput. Vis. Image Underst.*, 110:212–225, May 2008.