

# Multiple Structure Recovery via Robust Preference Analysis<sup>1</sup>

Luca Magri\*

*Dipartimento di Informatica,  
University of Verona, Strada le Grazie, 15 - 37134 Verona, Italy*

Andrea Fusiello\*

*DPIA,  
University of Udine, Via Delle Scienze, 208 - 33100 Udine, Italy*

---

## Abstract

This paper address the extraction of multiple models from outlier-contaminated data by exploiting preference analysis and low rank approximation. First points are represented in the preference space, then Robust PCA (Principal Component Analysis) and Symmetric NMF (Non negative Matrix Factorization) are used to break the multi-model fitting problem into many single-model problems, which in turn are tackled with an approach inspired to MSAC (M-estimator SAmple Consensus) coupled with a model-specific scale estimate. Experimental validation on public, real data-sets demonstrates that our method compares favorably with the state of the art.

*Keywords:* multi-model fitting, model estimation, spectral clustering, matrix factorization

---

## 1. Introduction

Geometric multi-model fitting aims at extracting parametric models from unstructured data in order to organize and aggregate visual content in suitable higher-level geometric structures<sup>2</sup>. This ubiquitous task can be encountered in many Computer Vision applications, for example in 3D reconstruction, where it is employed to estimate multiple rigid moving objects in order to initialize multi-body structure from motion (*e.g.*, [1]), or in the processing of 3D point clouds, where planar patches are fitted to produce intermediate geometric interpretations (*e.g.*, [2]).

Several challenges are afoot. First, segmentation and estimation tasks exhibit a chicken-and-egg pattern, for they are two closely entangled aspects of the same problem: data should be segmented based on their geometric proximity to structures whose unknown parameters must be estimated at the very same time. In other words, in order to estimate models one needs to first segment the data, but conversely, in order to segment the data it is necessary to know the structures associated with each data point.

In addition, the presence of multiple structures hinders robust estimation. Not only visual data are typically affected by arbitrarily large measurement errors –

and require the adoption of robust estimators – but the multi-modality of the data makes the problem even more demanding, as it is necessary to cope also with pseudo-outliers, a concept introduced by Stewart [3] for describing those measurements that do not match a model of interest because they are inliers of a different structure.

Moreover, the problem is inherently ill-posed, since many different interpretations of the same data are possible. Making the problem tractable requires a regularization strategy that constrains the solution using prior information, usually in the form of one or more parameter, such as the number  $\kappa$  of sought structures. Following the Occam’s razor principle – that one should not presume more things than the required minimum –  $\kappa$  should be kept as low as possible, but finding a correct trade-off between data fidelity and model complexity (a.k.a. bias-variance dilemma) is an intricate task, related to the *model selection* problem. Unfortunately estimating this quantity turns to be a thorny problem, and, for this reason, in many scenarios is assumed known.

*Outline.* In this article we present an original method henceforth dubbed RPA (Robust Preference Analysis) which attempts to disentangle the chicken-and-egg nature of multiple structure recovery reducing it to many single robust model estimation problems. In particular, three main steps can be singled out: First, we employ M-estimator to shift the problem into a conceptual space where data points are depicted by the *preference* they have granted to a pool of tentative structures, instantiated via random sampling. Second, a robust version of spectral clus-

---

\*Corresponding author

*Email addresses:* [magri.luca.1@gmail.com](mailto:magri.luca.1@gmail.com) (Luca Magri), [andrea.fusiello@uniud.it](mailto:andrea.fusiello@uniud.it) (Andrea Fusiello)

<sup>2</sup>The “structure” pertains to the arrangement and relations among the data, and it is intrinsic to the data itself, whereas the “model” is the mathematical description of the data that an observer fit onto them. Notwithstanding, we will use the two words interchangeably.

tering is presented: Robust Principal Component Analysis and Symmetric Non negative Matrix Factorization are employed to explicitly model the low rank nature of inlier preferences, in order to produce an accurate soft-segmentation of data. Third, this provisional segmentation is profitably combined with the initial preference representation in a MSAC-like framework to recover the sought structures. A noise scale estimate is computed for each model, with the help of robust statistic.

The next section offers a perspective on the literature that is most relevant to our work. Next (Section 2) our robust preference analysis method is detailed. In Section 2.1 we present the preference representation, devoting Section 2.2 to address the problem of sampling model hypotheses. In Section 2.3 we concentrate on the clustering problem and in Section 2.3.1 we explain how Low Rank matrix factorization techniques can be tailored to our preference embedding, paving the way to Section 2.4 where the robust structure-recovery strategy is described. Finally, in Section 3 we explore the performance of our method on public real datasets. Appendix A reviews some ideas, firstly emerged in the context of subspace clustering, that are extended to our general multi-model fitting problem.

### 1.1. Related work

The analysis of *consensus* and its dual counterpart, the analysis of *preferences*, can be traced as a *fil rouge* linking the vast literature on multi model geometric fitting. The consensus of a model is defined as the set of data points that fit the model within a certain inlier threshold  $\epsilon$ ; likewise, the preference set of a point is the set of models to which that point is closer than  $\epsilon$ .

Most of the multi-model fitting techniques proposed in the literature can be ascribed to one of these two concepts, according to which horn of the chicken-egg-dilemma is addressed first. Consensus-based algorithms put the emphasis on the estimation part and the focus is on models that have to describe as many points as possible. On the other hand, preference approaches concentrate on the segmentation side of the problem, for they are aimed at finding a proper partition of the data, from which estimation follows. In this section we attempt to retrace the path that, starting from consensus throughout preference analysis, have been followed in the literature to address the challenging issues presented by multiple structures recovery.

*Consensus analysis.* Consensus-oriented methods start with a pool of provisional model hypotheses, that are randomly instantiated on minimal sample sets (MSS), *i.e.*, samples composed by the minimum number of data points necessary to define a structure. Then the consensus sets of the models are inspected, and the models that better explain the data are kept. This idea is at the core of the well-known RANSAC (Random Sample Consensus) algorithm and derivations, *e.g.*, MSAC (M-estimator Sample Consensus) and MLESAC (Maximum Likelihood Estimation

Sample Consensus) [4]. Many ameliorations of RANSAC have been proposed in the literature, *e.g.* [5, 6, 7, 8], most of which have been surveyed in [9]. In the case of multiple models, Multi-RANSAC [10] and its modifications [11, 12] rely on the same principle; also the usual Hough transform and its randomized version [13] can be regarded as consensus-based algorithms, where models are detected as consensus maxima in a quantized hypothesis space. The approach presented in [14] combines random sampling, scale estimation and Mean Shift to determine the consensus set when models are multiple linear subspaces. More in general, maximizing the consensus of models is the foundation of many optimization-based geometric fitting algorithms [15].

*Preference analysis.* Preference analysis, introduced by [16], also start with a pool of provisional model hypotheses, but it swaps the role of data and models: rather than considering models and examining which points match them, the preference sets of individual data points are inspected. In particular, [17] build a *conceptual* space in which points are portrayed by the *preferences* they have accorded to provisional models. Within this category, J-Linkage [17] and T-Linkage [18] share the same *first-represent-then-segment* scheme: at first data are represented, respectively, either as characteristic functions or as continuous ones taking values on the hypothesized models, then the preference representations are segmented by greedy bottom-up clustering, exploiting either the Jaccard [19] or the Tanimoto [20] distances in order to measure the agreement between preferences, and using the fact that preferences of inliers from the same structure are correlated. This “preference trick” is a very flexible mechanism that can be applied to a wide varieties of scenarios requiring few assumptions on the desired structures. It is sufficient to have at disposal an error function, aimed at measuring residuals and then the structure recovery problem is shifted in the preference space where it can be addressed using cluster analysis.

Also RCMSA (Random Cluster Model Simulated Annealing) [21] exploits this idea by representing data points as permutations on a set of tentative models constructed iteratively, using subsets larger than minimal. Point preferences are organized in a weighted graph and the multi-model fitting task is stated as a graph cut problem which is solved efficiently in an annealing framework.

Multi-model fitting has been successfully cast as higher order clustering problems [22, 23, 24, 25], which implicitly adopt a preference based approach. In these works higher order similarity tensors are defined between  $n$ -tuple of points as the probability that these points are clustered together. In practice, this measure is approximated exploiting the residual error of the  $n$  points with respect to provisional models; this preference information is encoded in a hypergraph or a multi-way order tensor, which are properly reduced to pairwise similarity and fed to spectral clustering-like segmentation algorithms.

For instance, Sparse Grassmann Clustering (SGC) [24],

approximates the multi-way similarity tensor as the Gramian matrix defined by the inner product of points in the preference space, hence, following the spirit of spectral clustering, projects the Gramian to its best low rank approximation in a least square sense, using Grouse [26]. At the end, the rows of this approximated matrix, are considered as a new representation of the data points in a proper low dimensional space, and are consequently segmented with k-means.

Alternatively, instead of solving a point-point clustering problem, one can formulate a point-model bi-clustering problem directly on the preference hypergraph [27].

It goes without saying that the state-of-the-art on multi-model fitting can be also described along other dimensions. For example multiple structures recovery can be seen by an optimization perspective as the minimization of a global energy functional composed by two terms: a modelling error which can be interpreted as a likelihood term, and a penalty term encoding model complexity mimicking classical MAP-MRF objectives. A survey of multi-model fitting methods from this point of view can be found in [15].

## 2. The proposed method: Robust Preference Analysis (RPA)

In this work we present our attempt to solve the chicken-and-egg dilemma posed by multiple model fitting by tailoring some ideas from sparse representation techniques to the preference analysis, and exploiting the intrinsic interplay between consensus and preferences. As a matter of fact, our approach can be considered as a preference strategy complemented by consensus aspects.

The main idea of RPA, pictorially represented in Figure 1, is to build on the preference analysis and integrating decomposition techniques, such as Robust Principal Component Analysis (Robust PCA) and symmetric Non-negative Matrix Factorization (symNMF). Roughly speaking, this method can be regarded as a sort of “robust spectral clustering”, for it mimics the customary two steps of spectral clustering, which first project data on the space spanned by the first eigenvectors of the Laplacian matrix and then apply k-means. The shortcoming of this approach is that it is not robust to outliers. We propose to follow the same scheme, while enforcing robustness in every step: the eigen-decomposition stage is replaced by Robust PCA on a pairwise affinity matrix, and Symmetric NMF [28] takes the place of k-means. In this way we decompose the multi-model fitting problem into many single-fitting problems which are solved individually by inspecting – with the help of robust statistics – the product between the preference matrix and the output of Symmetric NMF.

The stages of RPA will be described in the next sections.

### 2.1. Preference embedding

The specific preference representation proposed by J-Linkage and T-Linkage can be seen as a particular instance

of a more general framework in which points preferences are depicted by exploiting M-estimators. In particular, the preference granted by a point  $x$  to  $h_j \in H$ , where  $H$  is a set of  $m$  sampled tentative structures, are expressed as  $w(r(x, h_j))$  where  $r(\cdot, \cdot)$  measures the distance between a datum  $x$  and a provisional model  $h_j$ , and  $w$  can be any M-estimator weighting function. If a step function  $w$  that assigns 1 to residual smaller than the inlier threshold  $\epsilon$  and zero otherwise is adopted, the binary preference set of J-Linkage is obtained. Or else, hard descenders can be employed as in T-Linkage or MSAC. In our approach we advocate the use of soft-descenders – such as the Cauchy weighting function – that do not have a finite rejection point, but force the influence of points to decrease to zero as the residuals tends to infinity, for they are less sensitive to the choice of the inlier threshold. The preference trick, that shift the data points  $X = \{x_1, \dots, x_n\}$ , from their ambient space to the preference one, consequently becomes:

$$\phi = (\phi_1, \dots, \phi_m)^\top: X \rightarrow [0, 1]^m, \quad (1)$$

where  $\phi$  is defined component-wise as the Cauchy weighting function:

$$\phi_j(x) = w(r(x, h_j)) = \frac{1}{1 + (r(x, h_j)/\theta\sigma_n)^2}, \quad j = 1, \dots, m, \quad (2)$$

where  $\sigma_n$  is an estimate of the standard deviation of the residuals and  $\theta$  is a tuning constant (set to 5.0 in all our experiments). In practice,  $\phi_j(x)$  assess the fitness to  $x$  with respect to the structure  $h_j$ .

Interestingly, this construction can be ascribed to the framework of *conceptual representation* settled by [29], for it can be interpreted in statistical sense as the posterior probabilities of the point  $x$  with respect to the  $m$  classes determined by the consensus set of the putative structures:

$$[\text{Prob}(x|h_1), \dots, \text{Prob}(x|h_m)] \in \mathbb{R}^m \quad (3)$$

Seen in this way, this representation is linked to the stream of research on higher-order clustering where probabilities are used to defined higher-order affinity between points.

In summary, the map  $\phi$  collects in a vector  $\Phi^i = \phi(x_i) \in \mathbb{R}^m$  the preferences granted by  $x_i$  to the models  $h_j$  ( $j = 1, \dots, m$ ) with a vote in  $[0, 1]$  that depends the residuals via a weighting function. We indicate with  $\Phi = [\Phi^1, \dots, \Phi^n] \in \mathbb{R}^{m \times n}$  the matrix whose columns are the point coordinates in the preference space. Rows corresponds to tentative models. The agreement between the preferences of two points reveals the multiple structures hidden in the data: points sharing the same preferences are likely to belong to the same structures. This notion is captured by defining a positive semi-definite kernel matrix  $K \in [0, 1]^{n \times n}$  on  $\Phi$ :

$$K(i, j) = \exp(-d_\tau(\Phi^i, \Phi^j)^2), \quad (4)$$

where

$$d_\tau(\Phi^i, \Phi^j) = 1 - \frac{\langle \Phi^i, \Phi^j \rangle}{\|\Phi^i\|^2 + \|\Phi^j\|^2 - \langle \Phi^i, \Phi^j \rangle}. \quad (5)$$

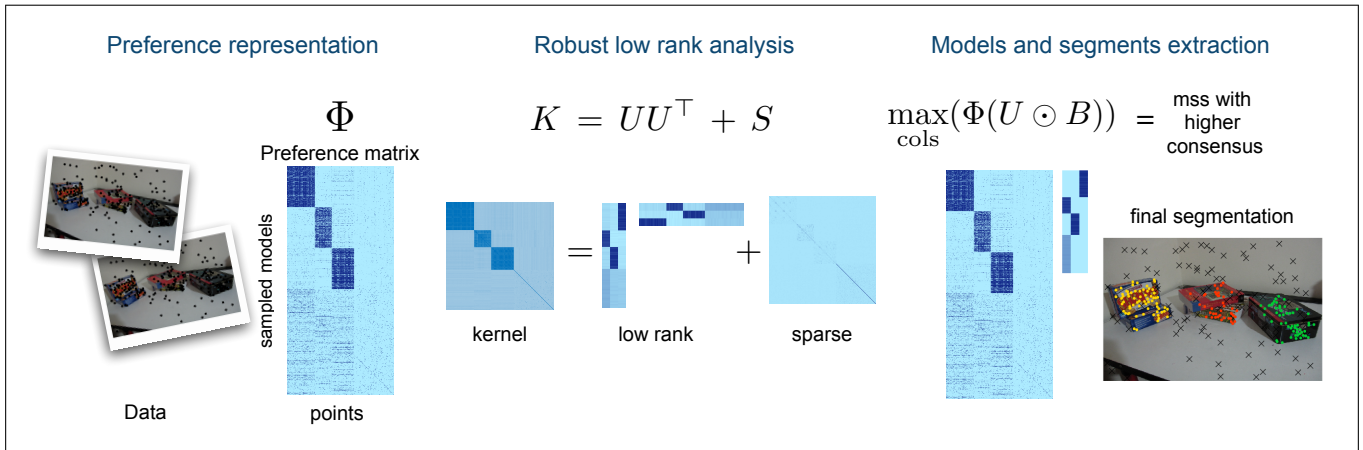


Figure 1: RPA in a nutshell: data points are shifted in a conceptual space where they are framed as a preference matrix  $\Phi$ . A similarity matrix  $K$  is defined exploiting agreement between preferences. Robust PCA and Symmetric NMF are used to robustly decompose  $K = UU^T + S$ , where  $S$  is a sparse matrix modelling outliers, and  $U$  is a low rank matrix representing the segmentation. Finally, models are extracted inspecting the product of the preference matrix with thresholded  $U$ , mimicking the MSAC strategy (points are ordered by cluster for visualization purposes).

Observe that  $\Phi$  endowed with this kernel bear a resemblance to the representation proposed in [18], where the notions of consensus and preference sets are relaxed to continuous functions and the *Tanimoto distance* [20]  $d_\tau$  is used to measure preference agreement. However some differences can be pointed out. In fact  $\phi$  is a continuous robust weighting function, whereas the embedding proposed in T-Linkage uses the inlier threshold  $\epsilon$  to cut off the preferences of points having distance greater than  $\epsilon$ ; moreover the kernel  $K$  replaces the direct use of the metric  $d_\tau$ .

Matrix  $\Phi$  can also be thought as a weighted hypergraph: hyper-edges represent sampled models and weights correspond to point preferences. In this interpretation the role of the kernel matrix  $K$  is to reduce the multi-way similarity to a pairwise affinity. Along, this line, it is worth noting that the Tanimoto distance is defined in terms of inner product of columns of the preference matrix  $\Phi$ , as the Gramian used, for example, in SGC [30] to encapsulate in a strict graph the flattened similarity tensor.

## 2.2. Biased Random Sampling in Tanimoto space

The exploration of the parameter space of all the possible models, hereinafter denoted by  $\Pi$ , straddles all the methods based on either consensus or preferences. Indeed, the designing of the pool of tentative models  $H$  has a pivotal role, for the quality of the embedding  $\phi_H$  is strictly linked to the ability of the sampled space  $H$  to adequately represent  $\Pi$ . In this section we propose a straightforward method to enhance the generation of tentative hypotheses leveraging on the geometric information embodied in the Tanimoto space.

In principle, also a simple uniform sampling strategy can capture the hidden multi-modality of a multi-model fitting problem. However this comes at the cost of extensive sampling and increased computational burden, since a large number of trials is required for reaching a reasonable

probability of hitting at least a *pure* (*i.e.*, outlier free) MSS per model. The number of required minimal sample sets can be significantly reduced when information about the distribution of the data points is available. This information can be either provided by the user, or can be extracted from the data itself through an auxiliary estimation process. Many strategies have been proposed along this line in order to guide sampling towards promising models both in the case of single-model [6, 31, 11], and in the multiple models scenario [32]. GROUPSAC [11], for example, relies on the assumption that inliers tends to be similar to each other, therefore data points are separated into groups that are similar according to some criterion, and intra-group MSS are favoured. A popular choice is the use of a localized sampling [33], where neighbouring points in the data space are selected with higher probability, thereby reducing the number of hypotheses that have to be generated. However, depending on the application, introducing a locality bias in the ambient space of data can be tricky, as different structures may obey different spatial distributions of data in the ambient space. Think for example to motion segmentation where different moving objects could have very different shapes or very different sizes due to perspective effects. Moreover one of the shortcomings of these strategies is that enforcing the spatial proximity requirement can make the estimation susceptible to degeneracies. As a matter of fact, some methods in the literature [34] enforce the opposite condition, *i.e.*, that MSS should not contain data points that are too close to each other.

In order to overcome this difficulty we propose to sample the hypotheses directly in the preference space. This can be easily done in three steps: at first a preliminary uniform sampling of hypotheses is performed, then data are represented in the Tanimoto space according to these putative models, finally a biased sampling in this space is performed. In particular, if a point  $y$  has already been



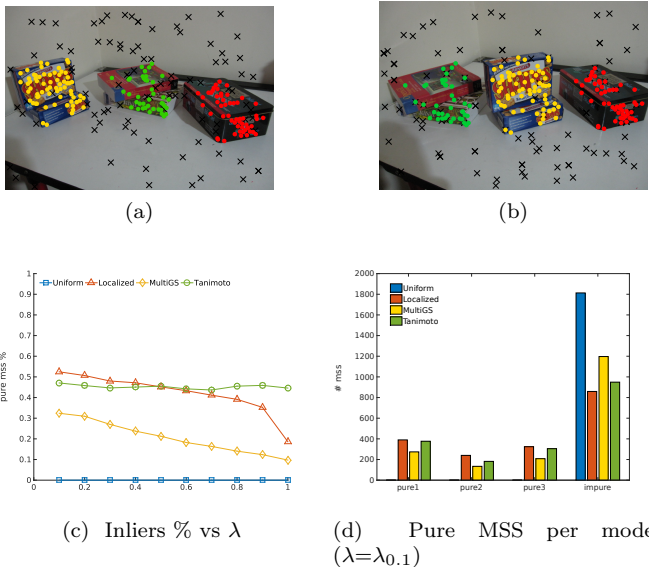


Figure 2: Comparison of guided sampling methods on the **biscuitbookbox** pair (a-b), from the AdelaideRMF dataset described in Section 4. Points are described by three fundamental matrices, outliers are marked as black crosses. (c) reports the percentage of pure MSS with respect to the local bias parameter  $\lambda$ . (d) the number of pure MSS per structures in the data. It is worth observing that uniform sampling struggles to find genuine MSS.

selected, then a point  $x$  such that  $x \neq y$  has the following probability of being drawn:

$$\text{Prob}(x|y) = \frac{1}{Z} \exp \frac{-d_\tau(\phi(x), \phi(y))^2}{\lambda^2}, \quad (6)$$

where  $Z$  is a normalization constant and  $\lambda$  controls the locality bias. Tanimoto distances can be then updated on the fly based on the hypotheses already sampled.

We illustrate the effectiveness of this sampling strategy on the image pair depicted in Figure 2a-b, where the model is a fundamental matrix. Our biased sampling in Tanimoto space is compared with respect to uniform sampling, localized sampling, and Multi-GS a method proposed in [32], which exploits a kernel based on preferences intersection. All these methods can be led back to the conditional sampling scheme presented here, substituting  $d_\tau$  in (6) with an appropriate distance function:  $d_U = 1$  for uniform sampling,  $d_L = \|\cdot\|$  for localized sampling and the intersection kernel  $d_{GS}$ . We run these methods with different values of  $\lambda$ ; in particular we set  $\lambda = \lambda_q$  as the  $q$ -th quantile of all these distances, varying  $q \in [0.1, 1]$ . The experiments demonstrate that our biased sampling provides results comparable with localized sampling for more values of  $\lambda$  (Figure 2c) and produces many pure MSS per model (Figure 2d).

This behaviour can be motivated in a probabilistic setting considering the lower density of outliers in the Tanimoto space with respect to the inlier distribution. The number  $m$  of MSS to be drawn is related to the percentage of outlier and must be large enough so that a certain number (at least) of outlier-free MSS are obtained with a

given probability for all the models. As explained in [17], if  $n_i$  is the number of inliers of the smaller structure contained in the data, the probability  $p$  of drawing a MSS of cardinality  $\zeta$  composed only of inliers is given by the product

$$p = \text{Prob}(E_1) \text{Prob}(E_2|E_1) \cdots \text{Prob}(E_\zeta|E_1, E_2 \dots E_{\zeta-1}), \quad (7)$$

where  $E_j$  is the event extract an inlier at the  $j$ -th drawing. It is worth noting that this probability exponentially decrease as  $\zeta$  increases, therefore, even if in principle the space  $\Pi$  can be explored by instantiating structures on subset with cardinality larger than the minimum as proposed in [35], in practice it is better to keep  $\zeta$  as low as possible and, if a consensus set can be defined, re-estimate the structure via least square or robust technique on its supporting points.

In the case of uniform sampling we have

$$\text{Prob}(E_j|E_1, E_2 \dots E_{j-1}) = \frac{n_i - j + 1}{n - j + 1}. \quad (8)$$

In our case, we can assume that the first point is sampled with uniform probability, hence  $\text{Prob}(E_1) = n_i/n$ , while the others are sampled with the probability function (6), therefore, after expanding the normalization constant  $Z$ , for every  $j = 2, \dots, \zeta$ , the conditional probability  $\text{Prob}(E_j|E_1, \dots, E_{j-1})$  can be approximated as

$$\frac{(n_i - j + 1)e^{-\alpha^2/\lambda^2}}{(n - n_i - j + 1)e^{-\omega^2/\lambda^2} + (n_i - j + 1)e^{-\alpha^2/\lambda^2}}, \quad (9)$$

where  $\alpha$  represents the average inlier-inlier distance, and  $\omega$  is the average inlier-outlier distance. Assuming that the cardinality of MSS is smaller with respect to the number of inliers,  $n_i \gg \zeta$ , we have

$$p \approx \delta \left( \frac{\delta e^{-\alpha^2/\lambda^2}}{(1 - \delta)e^{-\omega^2/\lambda^2} + \delta e^{-\alpha^2/\lambda^2}} \right)^{\zeta-1}. \quad (10)$$

where  $\delta = n_i/n$  is the inlier fraction for a given model. Since inliers determine compact cluster with respect to outliers ( $\omega > \alpha$ ), we have shown that this sampling strategy increases the probability of extracting a pure outlier-free MSS. In order to complete the picture on hypothesis generation, we can observe, following [16], that the probability of drawing at least  $k$  outlier-free MSS out of  $m$  with a given level of confidence  $\rho$  is obtained as:

$$\rho = 1 - \sum_{j=0}^{k-1} \binom{m}{j} p^j (1-p)^{m-j}. \quad (11)$$

The increased effectiveness of the sampling can be converted into either less samples or into an increase of the quality of the sampled structures (while preserving the same number of MSS).

Said that, drawing of MSS should be regarded as a computational procedure for approximating the parameter space  $\Pi$ : any information that can be profitably introduced to make the approximation more accurate can be

easily integrated in this step. For example various model verification tests can be adopted in order to enforce desired properties on the sampled structures. A notable example consists in ensuring geometrical non-degeneracy of MSS. A configurations of points is termed *degenerate* with respect to a model if it does not admit a unique solution with respect to that model [36], *e.g.*, collinear triplet of point in case of plane estimation. For instance, in the case of fundamental matrix estimation, a set of correspondence is deemed as degenerate if five or more points lie on the same plane in the scene [37], therefore a specific test aimed to identify MSS where five or more correspondences are related by a homography can be used to prune  $H$  from ambiguous structure estimate. Other problem-depending constraints can be imposed, *e.g.*, chirality constraints [38].

### 2.3. Clustering

To a certain extent, the Tanimoto distance promotes the aggregations of points lying in the low rank component of the preference space, since it takes into account the linear dependence of the preference representations, and characterizes outliers as the most separate points in that space. In the outlier-free scenario of SGC, this space is simply recovered via least-square regression: the eigen-decomposition of the (flattened) preference-tensor translates the multi-modality of the data in the language of the linear algebra, where the Occam principle is converted in a bound on the rank of the similarity matrix.

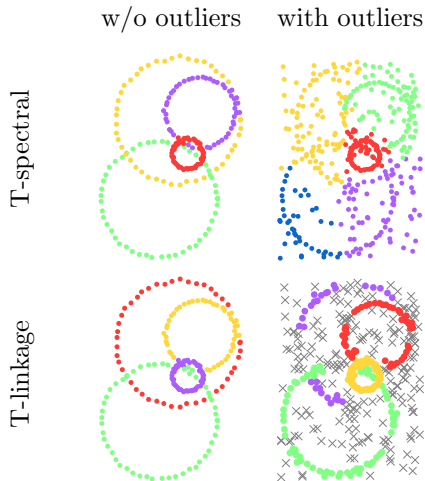


Figure 3: T-Spectral *vs.* T-linkage. The results are obtained with the knowledge of the correct number of models. Both methods are effective in dealing with pseudo outliers (left column), however T-spectral breaks down when data are contaminated. Indeed, the output of T-Spectral (top right) can hardly be interpreted for model fitting purposes. On the contrary, the four dominant models produced by T-Linkage (bottom right) are intelligible, albeit not fully accurate.

As a matter of fact, spectral clustering has several advantages over other methods. It is able to deal with clusters of arbitrary shape, (as opposed to other partitioning methods, as k-means, where the clusters are assumed to

lie in disjoint Voronoi cells). Moreover, it is less sensitive than T-linkage (and J-linkage) to the choice of the inlier threshold, because the segmentation is affected to less extent by  $\epsilon$ , being mainly controlled by the number of desired cluster specified as input. This is not surprising since the parameter  $\epsilon$  in T-Linkage controls the inlier threshold but it also implicitly governs the orthogonality between points represented in the Tanimoto space, and in practice, decides the number of attained models.

Unfortunately, spectral clustering suffers the presence of outliers. In this case, the naive strategy of increasing the number of desired segments in order to collect together outliers is of no help. In order to better grasp these aspects, it might be instructive to contrast T-linkage with a method (henceforth dubbed T-spectral) that combines the embedding in Tanimoto space (which is part of T-Linkage) with spectral clustering, as illustrated in Figure 3. If gross outliers contaminate the data, the clustering produced by T-spectral is useless, as the outliers are attached to some of the segments composed by inliers. On the contrary, even though the presence of outlying data hinders T-Linkage, the result of hierarchical clustering can be easily paired with some a-posteriori outlier rejection strategy (in this case consisting in selecting the four largest segments).

On the other hand, the greediness of linkage clustering is a weak spot of T-linkage that affects the segmentation outcomes. This can be recognized, for example, in the tendency of losing inliers during the segmentation step, which results in lower statistical efficiency in the estimation step. During the hierarchical clustering short-sighted decisions can possibly be made with relatively high probability that lead to suboptimal results.

In the light of these considerations, it appears sensible to combine the preference representation of T-linkage with a “robust” spectral clustering approach, that caters for outliers.

Hereinafter we will see how the ideas of exploiting a low rank representation to segment the data, which is stylized in spectral clustering, can be “robustified”, by leveraging on the low-rank and sparse decomposition literature.

#### 2.3.1. Robust Spectral Clustering

The kernel matrix  $K$  defined in Equation (4) encodes agreement between points on the preferred models.

Consider an ideal, binary  $n \times n$  affinity matrix  $F$  which encodes point membership to the same model/segment:

$$F_{i,j} = \begin{cases} 1 & \text{if } x_i \text{ and } x_j \text{ are clustered together} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

If points belonging to the same segment are arranged consecutively, the matrix  $F$  exhibits a block structure and therefore has rank  $\kappa$  equal to the number of clusters in the data.

As described in [25, 39] the problem of partitioning a set of data points in  $\kappa$  segments starting from a positive

semi-definite kernel matrix  $K$  is equivalent to approximating  $K$  in a least square sense by means of an ideal affinity matrix  $F$ . In formulae, denoting by  $\|\cdot\|_F$  the Frobenius norm of a matrix, we are interested in:

$$\min_F \|K - F\|_F^2, \quad (13)$$

under conditions on  $F$  to be further specified. This problem is usually formulated by introducing a matrix  $U \in \mathbb{R}^{n \times k}$  such that  $F = UU^T$ , which represents a soft segmentation of the data: the element  $U_{ij}$  measures the probability that the  $i$ -th point belongs to the  $j$ -th segment.

According to the constraints imposed on  $U$ , the solution of (13) corresponds to different classical clustering algorithms, such as spectral clustering or k-means. More precisely Equation (13) can be expanded as:

$$\begin{aligned} \min \|K - UU^T\|_F^2 &\iff \\ \min \text{trace}[(K - UU^T)^T(K - UU^T)] &\iff \\ \min \text{trace}(K^T K) - 2 \text{trace}(U^T K U) + \text{trace}(I) &\iff \\ \max \text{trace}(U^T K U). & \end{aligned} \quad (14)$$

The last equation becomes the objective of spectral clustering when  $K$  represents the opposite of the Laplacian of a graph and the columns of  $U$  are orthogonal. Since, it has been demonstrated [39] that the Laplacian of a similarity matrix can be viewed as the closest double stochastic approximation in relative entropy of the similarity matrix  $K$ , balanced partition are implicitly promoted. If  $K$  is chosen as the Gramian matrix of the data and if orthogonality, double stochasticity and non negativity of  $U$  are enforced, the considered trace maximization problem corresponds to k-means. Finally, for sake of completeness, we recall that in the case in which  $K = XX^T$  is the covariance matrix of the data, under orthogonality constraint, solving Equation (13) is tantamount of doing Principal Component Analysis.

In summary the constraints that are usually imposed on  $U$  are:  $U \geq 0$ ,  $\text{rank}(U) = \kappa$ ,  $U^T U = I$  and  $UU^T$  is doubly stochastic. Hard-clustering assignment implies orthogonality; being doubly stochastic represents a balancing condition on the sizes of the clusters; the non negativity of  $U$  ensures physical meaning of the entry of  $U$  which can be interpreted as the probability of points to belong to a given segment. The last constraint is the most important according to [25, 39], where it is highlighted as the key ingredients for solving Problem (14) are the low-rank nature of both the affinity matrix and  $U$  (since since  $k \ll n$ ), together with the non-negativeness of  $U$ .

Symmetric NMF (SymNMF) [28], that recently stood out in the literature of cluster analysis, enforces exactly these two proprieties. The idea at the basis of SymNMF is to rephrase (14) in the equivalent formulation

$$\min_{U \in \mathbb{R}_+^{n \times \kappa}} \|K - UU^T\|_F^2 \quad (15)$$

and hence to find  $U$  minimizing (15) using an improved Newton-like algorithm that exploits the second-order information efficiently.

When data are contaminated by gross outliers  $K$  has no longer low rank. For this reason, before applying SymNMF, we search *robustly* for the lowest-rank matrix  $L$  and the column-sparsiest matrix  $S$  such that the data matrix can be decomposed as

$$K = L + S. \quad (16)$$

This Robust PCA [40] step mimics in an outlier-resilient way the projection of data on the space of  $\kappa$  eigenvectors of the similarity matrix performed in spectral clustering. Moreover it is easy to recognize the formulation of LRR presented in Equation A.6 when the dictionary  $A$  is chosen as the identity matrix. The decomposition (16) can be computed with the Augmented Lagrangian Method (ALM) [41], which solves the problem

$$\arg \min \|L\|_* + \mu \|S\|_1 \quad \text{s.t.} \quad K = L + S. \quad (17)$$

The parameter  $\mu$  has a provable optimal value [40] at  $\mu = \frac{1}{\sqrt{n}}$ , where  $n$  is the dimension of the square matrix  $K$ . In other words, following the ideas in LRR (Equation (A.6)) we are retaining the low rank part of the similarity matrix, rejecting the sparse part of  $K$  that corresponds to micro-clusters of outlying preferences. Please note that this approximation differs from the one adopted by SGC [24] where a low rank space is fit to a Gramian matrix in a least square sense. We turned away from this approach because a least squares fit is reliable as long as the MSS are pure, but this property can not be ensured in presence of outliers.

We can now apply the SymNMF machinery to  $L$  (instead of  $K$ ) in order to find a completely positive rank- $\kappa$  factorization  $L = UU^T$ . A segmentation is readily obtained from  $U$  by considering the matrix  $B$  with the same dimension of  $U$  that has a one in each row where  $U$  achieves its row-maximum, and zero otherwise.  $B$  represents a segmentation, namely  $B_{i,j} = 1$  means that point  $i$  belongs to segment  $j$ . This last step is similar to the customary k-means that comes at the end of spectral clustering.

Interestingly we can interpret the clustering produced by SymNMF as a dimensionality reduction result in which the data are projected on the directions corresponding to the structures that explains better the data: the columns of  $U$  form a basis of a latent space of the data, whereas the rows collect the coefficients that express the data as linear combinations of the basis vectors. Since in the preference space the basis over which the points are represented is determined by the sampled structures, the columns of  $U$  can be thought as the  $\kappa$  ideal structures that well describe the data. Once these models have been recovered, the chicken-and-egg dilemma is disentangled and the multi-model fitting problem is reduced to many single-model fitting problems that can be solved, with the help of robust statistics, by maximizing consensus.

At this point, the matrix  $B$  is regarded as a provisional segmentation of the points into  $\kappa$  segments containing outliers. The goal of the next section is to refine this segmentation and prune outliers, by solving, within each segment, a robust *single model* fitting.

#### 2.4. Structure recovery and outlier pruning

Our robust spectral clustering produces an accurate soft-segmentation of data, represented by matrix  $B$  which allows us to reduce the multi-model fitting problem to many single-fitting problems, which are solved by consensus maximization followed by a refinement of the segmentation guided by the principles of robust statistics.

##### 2.4.1. Consensus maximization

Let us first observe that  $\Phi\mathbf{1}$  (where  $\mathbf{1}$  is a vector of ones) is the sum of the preference vectors of all the points in  $\Phi$ , so its entries are the votes obtained by each model. Hence finding the maximal entry of  $\Phi\mathbf{1}$  is equivalent to doing a sort of MSAC (M-estimator SAmples and Consensus) with the Cauchy weighting function (Equation 2).

We have seen that columns of  $B = [B^1, \dots, B^k]$  can be regarded as indicators of the segments. Hence  $\Phi B^i$  is the sum of the preference vectors of the points in the segment  $i$ , and its maximal entry represents the most preferred model in that segment. Therefore, the maxima over the columns of  $\Phi B$  are the indices of the models in  $\Phi$  that achieve maximum consensus in each segment. According to the observation above, this is equivalent to running a MSAC within each segment  $i$  with preference matrix  $(\Phi \text{diag}(B^i))$ . The above reasoning can be extended to the matrix  $U \circ B$  with entries in  $[0, 1]$ , that corresponds to a soft segmentation in which outliers are under-weighted ( $\circ$  denotes the component-wise or Hadamard product).

Before this step we found beneficial to remove "spurious" models, according to the segmentation represented by  $B$ , and to replace them with (tentatively) pure models. In particular, we deem as "spurious" those models that have less than 50% of their points belonging to a single segment; in practice, we label the points in  $\Phi$  according to the segmentation given by  $B$  and we remove the rows where no label occurs at least than 50% of the times. New models are obtained by drawing random MSS within each segment  $i$  with probabilities given by the non-zero entries of  $(U \circ B)^i$ .

In summary, the maximal entry in each column of  $\Phi(U \circ B)$  corresponds to the index of the most preferred model by the points of the segment, hence we choose it as the model that represents the segment. This could have been a final result if the goal was to find the correct models. However, having recognized the entangled nature of model fitting and segmentation problems, we will update the segmentation after refining the model.

##### 2.4.2. Robust Segmentation

The models computed in the previous step implicitly define a new segmentation, by assigning points to the near-

est model. Within this segmentation, outliers are singled-out as points with a residual higher than a threshold  $\epsilon = \theta \hat{\sigma}$  where  $\hat{\sigma}$  is an estimate of the scale of the noise affecting a particular model and  $\theta$  is the same tuning constant as in Equation (2) (set to 5.0 in our experiments).

The value of  $\hat{\sigma}$  can be obtained in several ways: it can be user provided ( $\hat{\sigma} = \sigma_n$ ) or can be estimated robustly as the standard deviation of the residuals of the *inliers* of a specific model.

The second solution is to be preferred, as it leaves the choice of the global noise scale  $\sigma_n$  a non-critical step and makes the threshold  $\epsilon$  data-adaptive, *i.e.* different for each model. As a robust scale estimator we selected [42]:

$$\hat{\sigma} = S_n \doteq c \text{ med}_i(\text{med}_j(|r_i - r_j|)), \quad (18)$$

(where  $r_i$ ,  $i = 1, \dots, n$  denotes the residual between the data  $x_i$  and the considered model) as a valid alternative to the more common median absolute deviation (MAD), which is aimed at symmetric distributions, and has a low (37%) Gaussian efficiency.  $S_n$  instead copes with skewed distributions, has the same breakdown as MAD but a higher Gaussian efficiency (58%). The *efficiency* of a robust estimator is defined as the ratio between the lowest achievable variance in an estimate to the actual variance of a (robust) estimate, with the minimum possible variance being determined by a target distribution such as the normal distribution. Asymptotic efficiency is the limit in efficiency as the number of data points tends to infinity.

The factor  $c$  can be set to 1.1926 for consistency with a normal distribution, but other distributions require different values (see [42] for details). In our experiments it has been tuned heuristically by analysing the distribution of the residuals of inliers given by the ground-truth

For each sequence (or pair) the true standard deviation of inliers and the estimator  $S_n$  had been computed, their ratio being a measure of  $c$ . The value of  $c$  for each experiment is obtained as the solution of a scalar least-squares system. Values are reported in Table 1.

Experiment	$\sigma_n$	$c$
Vanishing point detection	0.397	1.19
Motion segmentation (two-views)	0.005	1.53
Planar segmentation (two-views)	0.013	2.11
Motion segmentation (video)	0.005	1.51

Table 1: Parameters used in the experiments.  $\sigma_n$  is the overall standard deviation of the residuals of the inliers, as computed from ground-truth (units refer to normalized image coordinates).  $c$  is the value in Equation (18) tuned on the ground-truth data.

We noticed that in some cases most of the outliers are assigned to a single segment, resulting in a contamination greater than 50% that inevitably skews  $S_n$ . As a safeguard against this,  $S_n$  is computed only on the residuals smaller than  $\theta\sigma_n$  (with  $\theta = 5.0$ ).

---

**Algorithm 1** Robust Preference Analysis (RPA)

---

**Input:** data points  $X$ , an estimation of the noise variance  $\sigma_n$ , the number  $\kappa$  of sought structures

**Output:** segmentation

```
Generate tentative structures  $H = \{h_1, \dots, h_m\}$ ; ▷ biased sampling in Tanimoto space;
Build the matrix  $\Phi$ , using the Cauchy weighting function in Equation (2);
Define the matrix  $K$  as in Eq.(4);
 $K = L + S$ ; ▷ low-rank decomposition
 $L = UU^\top$ ; ▷ SymNMF
 $B = \max_{\text{rows}} U$ ; ▷ segmentation matrix
Replace spurious models with pure ones in  $H$ , according to the segmentation  $B$ ;
 $\{\iota_1, \dots, \iota_\kappa\} = \max_{\text{cols}} \Phi(U \odot B)$  ▷ consensus maximization
for  $j=1, \dots, \kappa$  do ▷ robust segmentation
  for  $\text{iter}=1, \dots, 2$  do
    estimate  $S_n$  for the model  $h_{\iota_j}$  using Eq. (18);
    outliers are points with residual  $> 5.0S_n$ ;
    recompute  $h_{\iota_j}$  with least-squares fit on the inliers;
  end for
end for
```

---

The model is then refined with a least-squares fit on the inliers, and the threshold  $\epsilon$  is computed again (for each model) to determine the final segmentation.

The whole procedure is summarized in Algorithm 1.

### 3. Experimental evaluation

In this section we assess experimentally the effectiveness of our algorithm. All the code is written in Matlab and is available for download<sup>3</sup>. We used the inexact ALM code [43], whereas the SymNMF implementation is taken from [28].

As customary in the literature, the results report the misclassification errors (ME), *i.e.* the percentage of misclassified points.

The input parameters of RPA are reported in Table 1 and have been kept fixed in each experiment.

*Vanishing point detection.* In this experiment we compare the performances of RPA with MFIPG on vanishing point detection using the York Urban Line Segment Database [44], a collection of 102 images of architectural Manhattan-like environments (*i.e.* scenes dominated by two or three mutually orthogonal vanishing directions). Annotated line-segments that match with the 3-d orthogonal frame of the urban scene are provided with the ground-truth; no outliers are present in the data.

The aim is to group the supplied segments in order to recover two or three orthogonal vanishing points. Specifically, in this scenario, the sought models are two or three points on the projective plane (possibly lying on the line at infinity) which identify vanishing direction, and a segment

is said “to belong” to a vanishing point if its supporting line, embedded in the projective plane, is incident to it.

MFIPG (Model Fitting with-Interacting Geometric Priors) is a recently proposed method [45] that builds on the PEARL [46] algorithm adding high-level geometric priors. In particular, in this application, an additional term expressing interaction between vanishing points is included to promote the extraction of orthogonal vanishing points. We used the code available from [43] with parameters tuned manually for each single image, starting from those recommended as default in the original paper. We take into comparison also J-Linkage and T-Linkage. In this experiment all the competing method were tuned with the knowledge that the data were outlier-free.

The differences among the methods can be appreciated by analyzing the average and median ME, collated in Table 2: these quantitative results confirms that RPA achieves the lowest ME, followed by T-linkage. The attained segmentations are highly accurate and the estimated vanishing points are orthogonal without enforcing this constraints as a prior information on segments distribution, as happen in MFIPG.

Figure 4 shows three images where RPA achieved the worst ME. The sub-optimal performance of RPA on these three images can all be ascribed to SymNMF that, similarly to spectral clustering, struggles in dealing with clus-

	J-linkage	T-linkage	MFIPG	RPA
Mean	2.85	1.44	3.51	<b>1.08</b>
Median	1.80	<b>0.00</b>	0.16	<b>0.00</b>

Table 2: Misclassification error (ME %) on YorkUrbanDB.

<sup>3</sup><http://www.diegm.uniud.it/fusiello/demo/rpa/>





Figure 4: A sample of the worst RPA results on **vanishing point detection**. Lines incident to the same vanishing points are represented with the same color.

ters of heavily unbalanced sizes.

*Two-views segmentation.* In this experiment we exploit the Adelaide Robust Model Fitting Data Set [47], which consists of 38 image pairs, 19 related by multiple fundamental matrices and 19 by multiple homographies, with gross outliers. The task consist in grouping points exhibiting a consistent motion between the two images by fitting fundamental matrices to subsets of point matches or segmenting them according to plane membership by fitting homographies.

We compared RPA with T-Linkage, which uses preference analysis and agglomerative clustering, and RCMSA (available from [48]), a robust method which relies on an efficient graph cut clustering based on a label optimization framework.

RPA and T-Linkage shared the same biased sampling in preference space: we drew  $3n$  hypotheses by uniform sampling and we used them to instantiate other  $3n$  MSSs according to Equation (6). In all the experiments  $\lambda$  was set to the median of the Tanimoto distances between data points.

We provided T-Linkage with the inlier thresholds computed from the ground-truth segmentation for each single image pair, and we retained as inliers the largest  $\kappa$  clusters,  $\kappa$  being the correct number of models according to ground-truth. Parameters of RCSA are those provided by the authors in their implementation [21].

Results are reported in Table 4 and Table 3, and demonstrate that our method outperforms its competitors, obtaining the lowest ME in most cases and the best mean and median results overall. Some of the worst instances, where RPA achieves the highest ME, are reported in Figure 5 and Figure 6.

While for fundamental matrices (Figure 5) the subjective quality of the worst result appears to be acceptable, in the case of homographies the three worst results represent defective segmentations, and the ME is indeed higher. By scrutinizing the intermediate results, it turns out that for `johnsonb` the blame should be put on Symmetric NMF, which fails to find a correct segmentation of the data, whereas for `library` and `oldclassicswing` it is the value

of  $\sigma_n$  that is respectively too low (over-segmentation) and too high (under-segmentation). While there are no clear remedies for the first case, the last two can be cured by a better choice of  $\sigma_n$ : for example, the ME drops to 24.53% for `library` and to 0.55% for `oldclassicswing` after assigning to  $\sigma_n$  the standard deviation of the residuals of the inliers for that specific image pair.

	$\kappa$	%out	T-linkage	RCMSA	RPA
<code>biscuitbookbox</code>	3	37.21	<b>3.10</b>	16.92	3.88
<code>breadcartoychips</code>	4	35.20	14.29	25.69	<b>7.50</b>
<code>breadcubechips</code>	3	35.22	<b>3.48</b>	8.12	5.07
<code>breadtoyocar</code>	3	34.15	9.15	18.29	<b>7.52</b>
<code>carchipscube</code>	3	36.59	<b>4.27</b>	18.90	6.50
<code>cubebreadtoychips</code>	4	28.03	9.24	13.27	<b>4.99</b>
<code>dinobooks</code>	3	44.54	20.94	23.50	<b>15.14</b>
<code>toycubecar</code>	3	36.36	15.66	13.81	<b>9.43</b>
<code>biscuit</code>	1	57.68	16.93	14.00	<b>1.15</b>
<code>biscuitbook</code>	2	47.51	<b>3.23</b>	8.41	<b>3.23</b>
<code>boardgame</code>	1	42.48	21.43	19.80	<b>11.65</b>
<code>book</code>	1	44.32	3.24	4.32	<b>2.88</b>
<code>breadcube</code>	2	32.19	19.31	9.87	<b>4.58</b>
<code>breadtoy</code>	2	37.41	5.40	3.96	<b>2.76</b>
<code>cube</code>	1	69.49	7.80	8.14	<b>3.28</b>
<code>cubetoy</code>	2	41.42	<b>3.77</b>	5.86	4.04
<code>game</code>	1	73.48	<b>1.30</b>	5.07	3.62
<code>gamebiscuit</code>	2	51.54	9.26	9.37	<b>2.57</b>
<code>cubechips</code>	2	51.62	6.14	7.70	<b>4.57</b>
mean			9.36	12.37	<b>5.49</b>
median			7.80	9.87	<b>4.57</b>

Table 3: Misclassification error (ME %) in **motion segmentation**.  $\kappa$  is the number of ground truth structures and % out is the percentage of outliers. All figures are the average of the middle 3 out of 5 runs.

*Video motion segmentation.* In these experiments we considered Sparse Subspace Clustering (SSC) [30] a state-of-

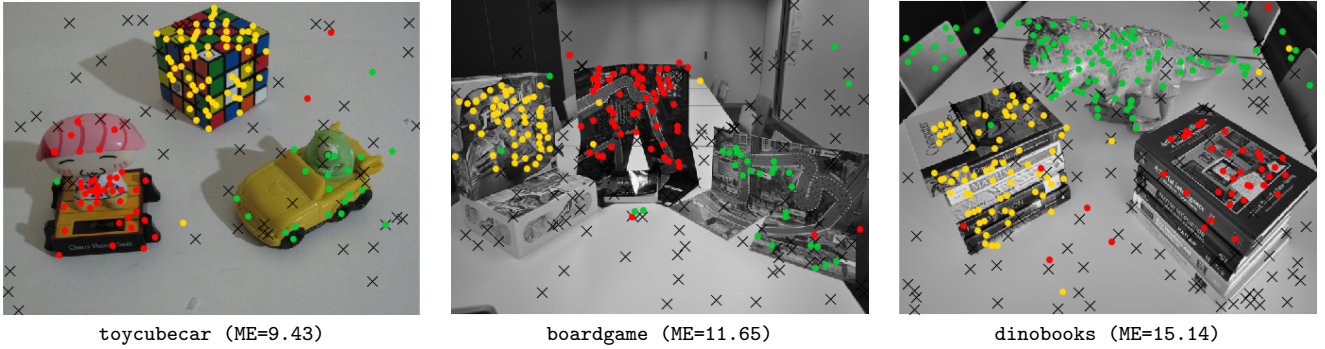


Figure 5: Instances of **motion segmentation** where RPA achieved the worst ME. Model membership is colour coded, black crosses ( $\times$ ) are outliers.

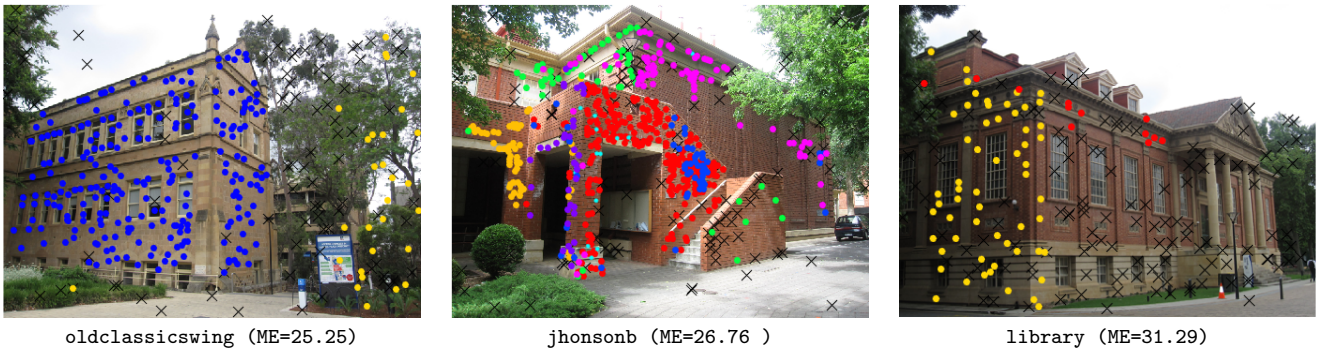


Figure 6: Instances of **plane segmentation** where RPA achieved the worst ME. Model membership is colour coded, black crosses ( $\times$ ) are outliers.

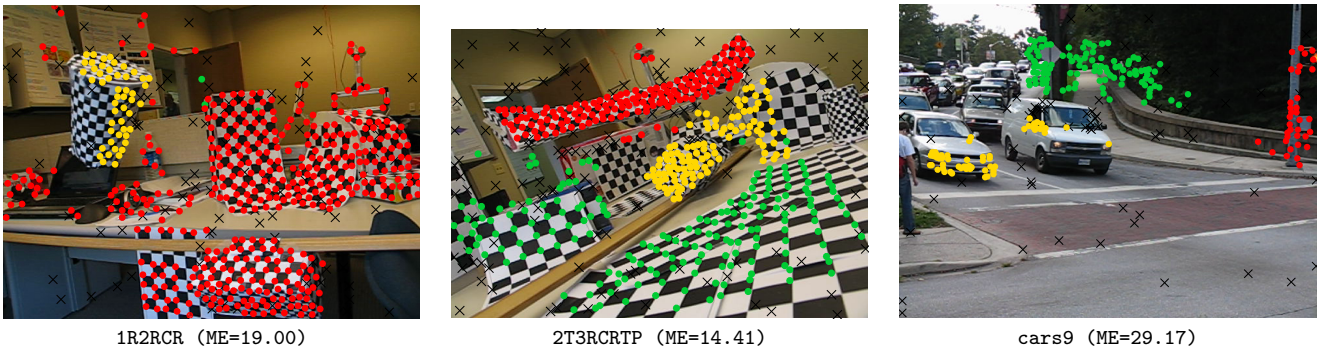


Figure 7: Some of the worst results obtained by RPA on **video motion segmentation**. Model membership is colour coded, black crosses ( $\times$ ) are outliers.

the-art technique that exploits a sparse representation to build an affinity matrix, which in turns is segmented by spectral clustering. The input data is a set of features trajectories across a video taken by a moving camera, and the aim is to recover the different rigid bodies moving independently.

Motion segmentation can be seen as a subspace segmentation problem under the assumption of affine cameras. In this case all feature trajectories associated with a single moving object lie in a linear subspace of dimension at most 4 in  $\mathbb{R}^{2f}$  (where  $F$  is the number of video frames). For this reason feature trajectories of a dynamic scene containing  $n$  rigid motion lie in the union of  $n$  low

dimensional subspace of  $\mathbb{R}^{2f}$  and segmentation can be reduced to clustering data points in a union of subspaces. Please note that SSC is tailored specifically to subspace segmentation, hence it cannot be applied to fit fundamental or homography models, for they are not linear nor affine subspaces.

We use the 155 video sequences from the Hopkins 155 dataset [49], each containing two or three moving objects. In order to evaluate the robustness of the method, we add to each sequence of the dataset 20% of outlying trajectories which are generated by starting a random walk at a random point in the image and adding increments taken from a trajectory (picked at random in the sequence) be-

	$\kappa$	%out	T-linkage	RCMSA	RPA
unionhouse	5	18.78	48.99	<b>2.64</b>	10.87
bonython	1	75.13	<b>11.92</b>	17.79	15.89
physics	1	46.60	29.13	48.87	<b>0.00</b>
elderhalla	2	60.75	10.75	29.28	<b>0.93</b>
ladysymon	2	33.48	<b>24.67</b>	39.50	<b>24.67</b>
library	2	56.13	<b>24.53</b>	40.72	31.29
nese	2	30.29	7.05	46.34	<b>0.83</b>
sene	2	44.49	7.63	20.20	<b>0.42</b>
napiera	2	64.73	28.08	31.16	<b>9.25</b>
hartley	2	62.22	21.90	37.78	<b>17.78</b>
oldclassicswing	2	32.23	<b>20.66</b>	21.30	25.25
barrsmith	2	69.79	49.79	<b>20.14</b>	36.31
neem	3	37.83	25.65	41.45	<b>19.86</b>
elderhallb	3	49.80	31.02	35.78	<b>17.82</b>
napierb	3	37.13	<b>13.50</b>	29.40	31.22
johnsona	4	21.25	34.28	36.73	<b>10.76</b>
johnsonb	7	12.02	24.04	<b>16.46</b>	26.76
unihouse	5	18.78	33.13	<b>2.56</b>	5.21
bonhall	6	6.43	21.84	<b>19.69</b>	41.67
mean			24.66	28.30	<b>17.20</b>
median			23.38	29.40	<b>17.53</b>

Table 4: Misclassification error (ME %) in **plane segmentation**.  $\kappa$  is the number of ground truth structures and % out is the percentage of outliers. All figures are the average of the middle 3 out of 5 runs.

		$\kappa = 2$		$\kappa = 3$	
		SSC	RPA	SSC	RPA
Checker	mean	8.19	<b>4.53</b>	9.58	<b>6.09</b>
	median	<b>0.32</b>	2.72	<b>2.91</b>	3.77
Traffic	mean	9.89	<b>7.16</b>	12.21	<b>7.88</b>
	median	<b>1.93</b>	4.80	5.87	<b>4.77</b>
Others	mean	17.97	<b>13.68</b>	22.84	<b>19.15</b>
	median	<b>1.07</b>	6.93	22.84	<b>19.15</b>
All	mean	<b>6.33</b>	9.84	10.94	<b>7.24</b>
	median	3.65	<b>0.82</b>	<b>3.68</b>	4.38

Table 5: Misclassification error (ME %) in **video motion segmentation** with 20% of outliers. Every row correspond to a subset of sequences.  $\kappa$  is the number of ground truth structures.

tween consecutive frames (again picked at random).

The input parameters of SSC are those provided by the authors in the code, available from [50]. Feature coordinates are normalized to the interval  $[-1, 1]$ .

Some failure cases of RPA are reported in Figure 7. In particular, in **1R2RCR** and **cars9** two independent motions are incorrectly merged because the intermediate segmen-

tation produced by SymNMF is inaccurate. As a consequence RPA hallucinates a small moving object among the outliers, whereas in **cars9** the background is over-segmented in two motions. The suboptimal segmentation of **2T3RCRTP** is due to the looseness of  $\sigma_n$ , as the inlier threshold of the motion defined by the translating box is not tightly estimated and some background points are incorrectly assigned to it. This problem is common in presence of nearly-degenerate motions, since, when the dimension of some subspace drops, the scale of the residuals of its points is accordingly affected and determining a global inlier threshold becomes difficult.

The figures reported in Table 5 show that RPA achieves results comparable to SSC, which anyway is specifically designed for subspace recovery. RPA, instead, is able to handle a variety of structures, as it combines the main ideas of SSC – *i.e.* modeling outliers as sparse components – with the versatility of the preference representation.

*Sampling.* In order to complement Figure 2, where a comparison of biased sampling methods is reported on a two-view motion segmentation example, we also investigate the performance of the guided sampling in the preference space on instances of vanishing points estimation, homography fitting, and video motion segmentations problems. Results are collected in Figure 8 where it can be appreciated that in all the applications is beneficial to exploit the information provided by Tanimoto in terms of pure MSS extracted. Moreover the specification of the locality bias  $\lambda$  seems to be less critical.

## 4. Conclusions

We argued that preference analysis coupled with robust matrix decompositions yield a versatile tool for geometric fitting, which exploits profitably the interplay between consensus and preference and is robust to outliers.

Our strategy was to reduce the multi-model fitting task to many single robust model estimation problems attempting to solve the chicken-and-egg dilemma. In particular, we conceived three levels of protection against outliers. The first one is the adoption of the Cauchy function to model points preferences. The second level appears in the robust low rank approximation, where Robust PCA and SymNMF are used to give rise to a soft segmentation where outliers are under-weighted. Robust extraction of models in a MSAC-like framework, together with outlier rejection based on a robust, model-specific scale estimate, is our third safeguard against outliers. The value of  $\sigma_n$  and the number of models  $\kappa$  are the only inputs required from the user.

Experiments have provided evidence that our method compares favorably with state-of-the-art competing algorithms, being more accurate than T-linkage, MFIGP and RCMSA and as accurate as SSC, although the latter only works with subspaces.



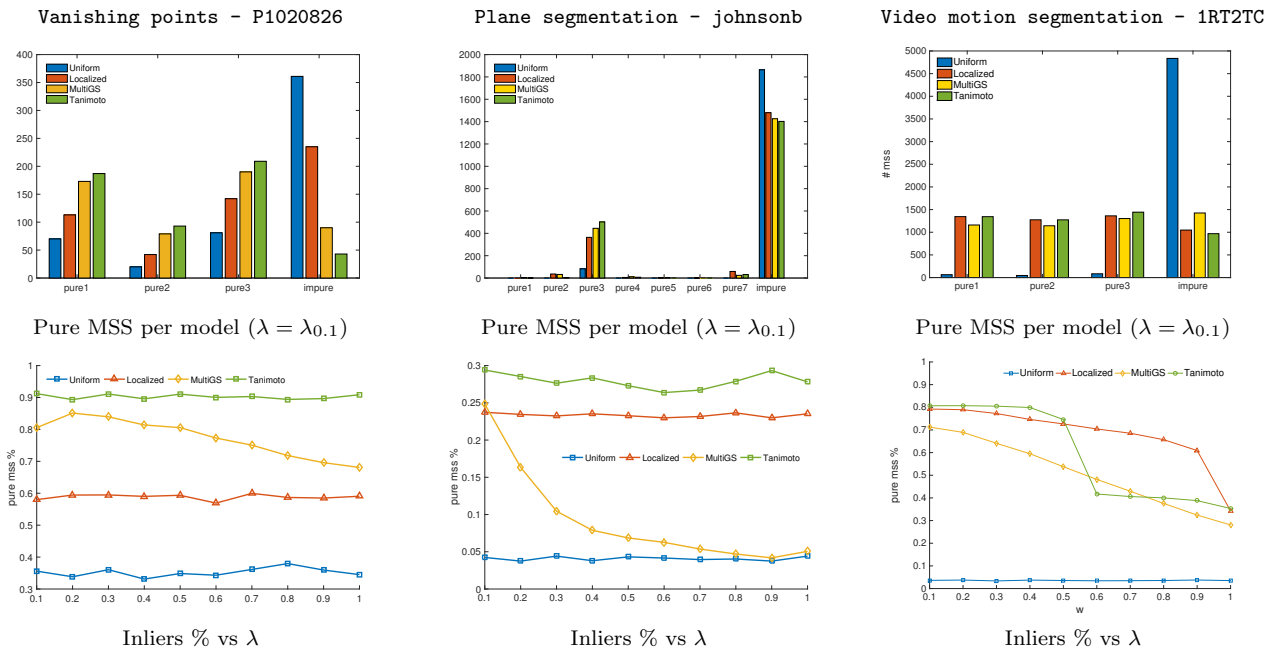


Figure 8: Comparison of different sampling strategies on sample vanishing point (first column), plane segmentation (second column) and video motion segmentation (third column) experiments.

## References

- [1] K. E. Ozden, K. Schindler, L. Van Gool, Multibody structure-from-motion in practice, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (6) (2010) 1134–1141.
- [2] R. Toldo, A. Fusiello, Image-consistent patches from unstructured points with j-linkage, *Image and Vision Computing* 31 (10) (2013) 756–770.
- [3] C. V. Stewart, Bias in robust estimation caused by discontinuities and multiple structures, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (8) (1997) 818–833.
- [4] P. H. S. Torr, A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry, *Computer Vision and Image Understanding* 78 (2000) 2000.
- [5] K. Lebeda, J. Matas, O. Chum, Fixing the locally optimized RANSAC—full experimental evaluation, in: *British Machine Vision Conference, 2012*, pp. 1–11.
- [6] O. Chum, J. Matas, Matching with PROSAC - progressive sample consensus, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005*, pp. 220–226.
- [7] V. Fragoso, P. Sen, S. Rodriguez, M. Turk, Evsac: accelerating hypotheses generation by modeling matching scores with extreme value theory, in: *Proceedings of the IEEE International Conference on Computer Vision, 2013*, pp. 2472–2479.
- [8] R. Litman, S. Korman, A. Bronstein, S. Avidan, Inverting ransac: Global model detection via inlier rate estimation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015*, pp. 5243–5251.
- [9] S. Choi, T. Kim, W. Yu, Performance evaluation of RANSAC family, in: *British Machine Vision Conference, 2009*.
- [10] M. Zuliani, C. S. Kenney, B. S. Manjunath, The multiRANSAC algorithm and its application to detect planar homographies, in: *Proceedings of the IEEE International Conference on Image Processing, 2005*.
- [11] K. Ni, H. Jin, F. Dellaert, GroupSAC: Efficient consensus in the presence of groupings, in: *Proceedings of the International Conference on Computer Vision, 2009*, pp. 2193–2200.
- [12] O. Gallo, R. Manduchi, A. Rafii, CC-RANSAC: Fitting planes in the presence of multiple surfaces in range data, *Pattern Recognition Letters* 32 (3) (2011) 403–410.
- [13] L. Xu, E. Oja, P. Kultanen, A new curve detection method: randomized Hough transform (RHT), *Pattern Recognition Letters* 11 (5) (1990) 331–338. doi:http://dx.doi.org/10.1016/0167-8655(90)90042-Z.
- [14] S. Mittal, S. Anand, P. Meer, Generalized projection based estimator: Theory and applications., in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011*.
- [15] H. Isack, Y. Boykov, Energy-based geometric multi-model fitting, *International Journal of Computer Vision* 97 (2) (2012) 123–147.
- [16] W. Zhang, J. Kosecká, Nonparametric estimation of multiple structures with outliers, in: *Workshop on Dynamic Vision, European Conference on Computer Vision 2006, Vol. 4358 of Lecture Notes in Computer Science, Springer, 2006*, pp. 60–74.
- [17] R. Toldo, A. Fusiello, Robust multiple structures estimation with j-linkage, in: *Proceedings of the European Conference on Computer Vision, 2008*, pp. 537–547.
- [18] L. Magri, A. Fusiello, T-linkage: A continuous relaxation of J-linkage for multi-model fitting, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014*.
- [19] P. Jaccard, Étude comparative de la distribution florale dans une portion des Alpes et des Jura, *Bulletin del la Société Vaudoise des Sciences Naturelles* 37 (1901) 547–579.
- [20] T. Tanimoto, An elementary mathematical theory of classification and prediction, Internal technical report, IBM (1957).
- [21] T.-T. Pham, T.-J. Chin, J. Yu, D. Suter, The random cluster model for robust geometric fitting, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (8) (2014) 1658–1671.
- [22] S. Agarwal, J. Lim, L. Zelnik-manor, P. Perona, D. Kriegman, S. Belongie, Beyond pairwise clustering, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005*, pp. 838–845.
- [23] V. M. Govindu, A Tensor Decomposition for Geometric Grouping and Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1 (2005)* 1150–1157.
- [24] S. Jain, V. M. Govindu, Efficient higher-order clustering on the grassmann manifold, in: *Proceedings of the International Conference on Computer Vision, 2013*, pp. 3511–3518.
- [25] R. Zass, A. Shashua, A unifying approach to hard and probabilistic clustering, in: *Proceedings of the International Confer-*

- ence on Computer Vision, Vol. 1, 2005, pp. 294–301.
- [26] L. Balzano, R. Nowak, B. Recht, Online identification and tracking of subspaces from highly incomplete information, in: Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing, 2010.
- [27] M. Tepper, G. Sapiro, A biclustering framework for consensus problems, *SIAM Journal on Imaging Sciences* 7 (4) (2014) 2488–2525.
- [28] D. Kuang, S. Yun, H. Park, SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering, *Journal of Global Optimization* (2014) 1–30.
- [29] E. Pekalska, R. P. Duin, The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence), World Scientific Publishing, 2005.
- [30] M. Soltanolkotabi, E. Elhamifar, E. J. Candès, Robust subspace clustering, *Annals of Statistics* 42 (2) (2014) 669–699.
- [31] O. Chum, J. Matas, Randomized RANSAC with  $T_{d,d}$  test, in: *Image and Vision Computing*, Vol. 22, 2002, pp. 837–842.
- [32] T.-J. Chin, J. Yu, D. Suter, Accelerated hypothesis generation for multistructure data via preference analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2012) 533–546.
- [33] Y. Kanazawa, H. Kawakami, Detection of planar regions with uncalibrated stereo using distributions of feature points, in: *British Machine Vision Conference*, 2004, pp. 247 – 256.
- [34] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence* 78 (1-2) (1995) 87–119.
- [35] P. Purkait, T.-J. Chin, H. Ackermann, D. Suter, Clustering with hypergraphs: the case for large hyperedges, in: *Proceedings of the European Conference on Computer Vision*, 2014, pp. 672–687.
- [36] P. H. S. Torr, A. Zisserman, S. Maybank, Robust detection of degeneracy, in: *Proceedings of the International Conference on Computer Vision*, Springer-Verlag, 1995, pp. 1037–1044.
- [37] O. Chum, T. Werner, J. Matas, Two-view geometry estimation unaffected by a dominant plane, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, 2005, pp. 772–779.
- [38] O. Chum, T. Werner, J. Matas, Epipolar geometry estimation via RANSAC benefits from the oriented epipolar constraint, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
- [39] R. Zass, A. Shashua, Doubly stochastic normalization for spectral clustering, in: *Neural Information Processing Systems*, 2006, pp. 1569–1576.
- [40] E. J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, *Journal of the ACM* 58 (3) (2011) 11:1–11:37.
- [41] Z. Lin, M. Chen, Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted Low-Rank matrices, eprint arXiv:1009.5055 (2010).
- [42] P. J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, *Journal of the American Statistical Association* 88 (424) (1993) 1273–1283.
- [43] [http://perception.csl.illinois.edu/matrix-rank/sample\\_code.html](http://perception.csl.illinois.edu/matrix-rank/sample_code.html).
- [44] P. Denis, J. H. Elder, F. J. Estrada, Efficient edge-based methods for estimating manhattan frames in urban imagery, in: *Proceedings of the European Conference on Computer Vision*, 2008, pp. 197–210.
- [45] T. T. Pham, T.-J. Chin, K. Schindler, D. Suter, Interacting geometric priors for robust multimodel fitting, *IEEE Transactions on Image Processing* 23 (10) (2014) 4601–4610.
- [46] A. Delong, O. Veksler, Y. Boykov, Fast fusion moves for multimodel estimation, in: *Proceedings of the European Conference on Computer Vision*, 2012, pp. 370–384.
- [47] H. S. Wong, T.-J. Chin, J. Yu, D. Suter, Dynamic and hierarchical multi-structure geometric model fitting, in: *Proceedings of the International Conference on Computer Vision*, 2011.
- [48] [http://cs.adelaide.edu.au/~trung/lib/exe/fetch.php?media=rCMSA\\_robust\\_fitting.zip](http://cs.adelaide.edu.au/~trung/lib/exe/fetch.php?media=rCMSA_robust_fitting.zip).
- [49] R. Tron, R. Vidal, A benchmark for the comparison of 3-d motion segmentation algorithms, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2007, pp. 1–8.
- [50] <http://www.vision.jhu.edu/code/>.
- [51] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2013) 171–184.
- [52] A. M. Bruckstein, D. L. Donoho, E. Michael, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM review* 51 (1) (2009) 34–81.

## Appendix A. Background: subspace recovery

Multiple subspace estimation is a particular instance of multi-model fitting that can be encountered in many Computer Vision applications – from image segmentation to motion segmentation, or temporal video segmentation – where a mixture of subspaces has to be fitted to high dimensional data.

In this scenario, a recent trend [30, 51] has been concentrating on sparse representation and low rank constraints in order to derive a similarity matrix robust to outliers that, in turn, is fed to spectral clustering.

The notion of sparsity [52] is straightforward: a vector is sparse if it can be exactly or approximately represented as a linear combination of only a few vectors selected from a predetermined dictionary.

This property is encoded by the  $\ell_0$  “norm”  $\|v\|_0 = |\{k: (v)_k \neq 0\}|$ : a vector admits a  $k$ -sparse representation with respect to a dictionary  $D$  if it can be written as  $Dc$  and  $\|c\|_0 = k$ .

While the reconstruction of a signal from its sparse representation is a simple linear transform, the inverse problem

$$\arg \min_c \|c\|_0 \text{ such that } Dc = v, \quad (\text{A.1})$$

is a non-linear optimization that, in general, is intractable. This fact has motivated the flourishing of many methods in the compressed sensing literature based on the convex relaxation of the  $\ell_0$ -norm. The  $\ell_1$ -norm, defined as the sum of the absolute values of the entries  $\|v\|_1 = \sum_k |(v)_k|$ , serves to replace the problem in Equation (A.1) with the following tractable optimization objective:

$$\arg \min_c \|c\|_1 \text{ such that } Dc = v. \quad (\text{A.2})$$

At a high level, the effectiveness of sparsity-oriented approaches can be explained viewing this property as a useful way of constraining the complexity of a vector representation, which can be very generally justified by Occam’s razor. Sparse Subspace Clustering (SSC [30]) exploits this principle to derive a segmentation of high dimensional data. The main idea of SCC is to take advantage of the “self-expressiveness” of the input: every point can be expressed as a linear combination of few other points lying in

the same subspace. A sparse  $\ell_1$  optimization program captures this property by defining a collection of vectors of coefficients  $c_i$  using as a dictionary the data matrix  $D = X$ :

$$\arg \min_{c_i \in \mathbb{R}^N} \|c_i\|_1 \quad \text{subject to} \quad x_i = Xc_i \quad \text{and} \quad (c_i)_i = 0. \quad (\text{A.3})$$

The constraint  $(c_i)_i = 0$  removes the trivial solution that decomposes a point  $x_i$  as a linear combination of itself. In this way the sparsest representation of  $x_i$  would only select vectors from the subspace in which  $x_i$  happens to lie. In matrix notation the problem in Equation (A.3) can be rewritten as

$$\min_C \|C\|_1 \quad \text{such that} \quad X = XC^T, \text{diag}(C) = 0. \quad (\text{A.4})$$

In case of data contaminated by noise and outliers, instead of expressing a data point as an exact linear combination of other points, it is convenient to introduce a penalty term:

$$\min_C \|C\|_1 + \mu \|E\|_{2,1} \quad \text{such that} \quad X = XC^T + E, \text{diag}(C) = 0. \quad (\text{A.5})$$

The  $\ell_{1,2}$ -norm is defined as  $\|E\|_{2,1} = \sum_j \sqrt{\sum_i |E_{i,j}|^2}$ . The underlying assumption is that a data point can be written as  $x_i = Xc_i + e_i$  where  $e_i$  is a sparse vector that models gross outlying entries of  $x_i$ .

A related approach, termed Low Rank Representation (LRR) [51], derives a similar convex optimization problem

$$\min_C \|C\|_* + \mu \|E\|_1 \quad \text{such that} \quad X = DC^T + E \quad (\text{A.6})$$

where  $D$  is the dictionary matrix, either constructed in advance or equal to  $X$ , and  $\|C\|_* = \sum \sigma_i(C)$  is the nuclear norm that equals to the sum of all the singular values  $\sigma_i(C)$  of  $C$ .

Both SSC and LRR then use the optimal  $C$  to define an affinity matrix. It is quite natural to define a similarity measure between points as  $S = |C| + |C|^T$ , because non-zero elements of  $c_i$  correspond to points from the same subspace of  $x_i$ . This similarity matrix is finally used to feed spectral clustering and a partition of the data is obtained.

The main difference between the two methods is that SSC minimizes the  $\ell_1$ -norm of the representation matrix to induce sparsity while LRR tries to minimize nuclear norm to promote a low-rank structure. Both methods however rely on the same idea: taking advantage of the intrinsic redundancy of multi-model data to embed the data point in a discrete metric space to facilitate the segmentation task that is solved thanks to spectral clustering.

Interestingly, this *first-represent-then-clusterize* approach is evocative of the preference approach philosophy.

If T-Linkage is taken in comparison, the representation matrix  $C$  corresponds to the preference embedding, whereas the similarity  $S$  plays the same role of the Tanimoto distances. Moreover, as sparsity is concerned, outliers can be recognized in practice as sparse vectors in the preference space, since the number of sampled structures supporting

outliers is considerable smaller than the number of structures supporting an inlier (typically an outlier is explained only by the structures it has happen to generate by the MSS it belongs to).

There are also some deep differences, though. The preference representation proposed in T-Linkage is not limited to vectorial (and affine) subspace. This comes at the cost of choosing a correct inlier threshold – a parameter that has geometrical meaning but is highly data depended – and of the effectiveness of the sampling scheme adopted to generate hypotheses. On the other hand, SSC and LRR depends on the regularization coefficient  $\mu$  to handle outliers and noisy data. As the segmentation is concerned, spectral clustering requires to know in advance the number of models, whereas the greedy linkage strategy, adopted by T-Linkage, can work without this parameter.