# Constrained Spectral Clustering via Exhaustive and Efficient Constraint Propagation

Zhiwu Lu and Horace H.S. Ip

Department of Computer Science, City University of Hong Kong, Hong Kong
AIMtech Centre, City University of Hong Kong, Hong Kong
lzhiwu2@student.cityu.edu.hk, cship@cityu.edu.hk

**Abstract.** This paper presents an exhaustive and efficient constraint propagation approach to exploiting pairwise constraints for spectral clustering. Since traditional label propagation techniques cannot be readily generalized to propagate pairwise constraints, we tackle the constraint propagation problem inversely by decomposing it to a set of independent label propagation subproblems which are further solved in quadratic time using semi-supervised learning based on $k$-nearest neighbors graphs. Since this time complexity is proportional to the number of all possible pairwise constraints, our approach gives a computationally efficient solution for exhaustively propagating pairwise constraint throughout the entire dataset. The resulting exhaustive set of propagated pairwise constraints are then used to adjust the weight (or similarity) matrix for spectral clustering. It is worth noting that this paper first clearly shows how pairwise constraints are propagated independently and then accumulated into a conciliatory closed-form solution. Experimental results on real-life datasets demonstrate that our approach to constrained spectral clustering outperforms the state-of-the-art techniques.

## 1 Introduction

Cluster analysis is largely driven by the quest for more robust clustering algorithms capable of detecting clusters with diverse shapes and densities. It is worth noting that data clustering is an ill-posed problem when the associated objective function is not well defined, which leads to fundamental limitations of generic clustering algorithms. Multiple clustering solutions may seem to be equally plausible due to an inherent arbitrariness in the notion of a cluster. Therefore, any additional supervisory information must be exploited in order to reduce this degeneracy of possible solutions and improve the quality of clustering. The labels of data are potential sources of such supervisory information which has been widely used. In this paper, we consider a commonly adopted and weaker type of supervisory information, called pairwise constraints which specify whether a pair of data belongs to the same cluster or not.

There exist two types of pairwise constraints, known as *must-link* constraints and *cannot-link* constraints, respectively. We can readily derive such pairwise constraints from the labels of data, where a pair of data with the same label

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part VI, LNCS 6316, pp. 1–14, 2010.

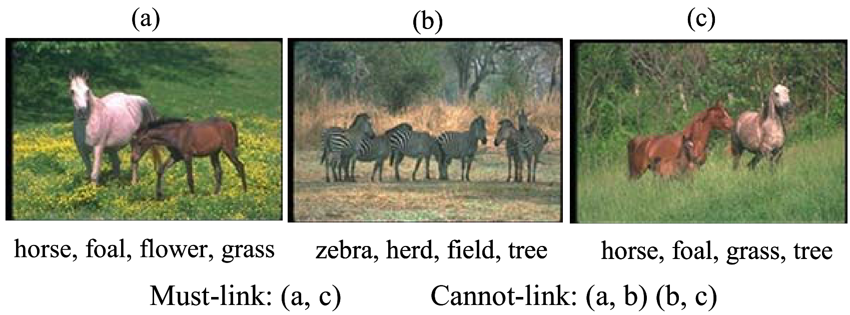



**Fig. 1.** The must-link and cannot-link constraints derived from the annotations of images. Since we focus on recognizing the objects of interests in images, these constraints are formed without considering the backgrounds such as tree, grass, and field.

denotes must-link constraint and cannot-link constraint otherwise. It should be noted, however, that the inverse may not be true, i.e. in general we cannot infer the labels of data from pairwise constraints, particularly for multi-class data. This implies that pairwise constraints are inherently weaker but more general than the labels of data. Moreover, pairwise constraints can also be automatically derived from domain knowledge [1,2] or through machine learning. For example, we can obtain pairwise constraints from the annotations of the images shown in Fig. 1. Since we focus on recognizing the objects of interests (e.g. horse and zebra) in images, the pairwise constraints can be formed without considering the backgrounds such as tree, grass, and field. In practice, the objects of interest can be roughly distinguished from the backgrounds according to the ranking scores of annotations learnt automatically by an image search engine.

Pairwise constraints have been widely used for constrained clustering [1,2,3,4,5], and it has been reported that the use of appropriate pairwise constraints can often lead to the improved quality of clustering. In this paper, we focus on the exploitation of pairwise constraints for spectral clustering [6,7,8,9] which constructs a new low-dimensional data representation for clustering using the leading eigenvectors of the similarity matrix. Since pairwise constraints specify whether a pair of data belongs to the same cluster, they provide a source of information about the data relationships, which can be readily used to adjust the similarities between the data for spectral clustering. In fact, the idea of exploiting pairwise constraints for spectral clustering has been studied previously. For example, [10] trivially adjusted the similarities between the data to 1 and 0 for must-link and cannot-link constraints, respectively. This method only adjusts the similarities between constrained data. In contrast, [11] propagated pairwise constraints to other similarities between unconstrained data using Gaussian process. However, as noted in [11], this method makes certain assumptions for constraint propagation specially with respect to two-class problems, although the heuristic approach for multi-class problems is also discussed. Furthermore, such constraint propagation is formulated as a semi-definite programming (SDP) problem in [12]. Although the method is not limited to two-class problems, it incurs extremely large computational cost for solving the SDP problem. In [13], the constraint propagation is also formulated as a constrained optimization problem, but only must-link constraints can be used for optimization.

To overcome these problems, we propose an exhaustive and efficient constraint propagation approach to exploiting pairwise constraints for spectral clustering, which is not limited to two-class problems or using only must-link constraints. Specifically, since traditional label propagation techniques [14,15,16] cannot be readily generalized to propagate pairwise constraints, we tackle the constraint propagation problem inversely by decomposing it to a set of independent label propagation subproblems. Furthermore, we show that through semi-supervised learning based on $k$-nearest neighbors graphs, the set of label propagation subproblems can be solved in quadratic time $O(kN^2)$ with respect to the data size $N$ ($k \ll N$). Since this time complexity is proportional to the total number of all possible pairwise constraints (i.e. $N(N-1)/2$), our constraint propagation approach can be considered computationally efficient. It is worth noting that our approach incurs much less computational cost than [12], given that SDP-based constraint propagation has a time complexity of $O(N^4)$.

The resulting exhaustive set of propagated pairwise constraints can be exploited for spectral clustering through adjusting the similarity matrix with this information. The experimental results on image and UCI datasets demonstrate that our approach outperforms the state-of-the-art techniques. It is worth noting that our approach can be seen as a very general constraint propagation technique, which has the following advantages:

**(1)** This is the first constraint propagation approach that clearly shows how pairwise constraints are propagated independently and then accumulated into a *conciliatory closed-form solution.*

**(2)** Our approach is not limited to two-class problems or using only must-link constraints. More importantly, *our approach allows soft constraints*, i.e., the pairwise constraints can be associated with confidence scores like [17,18].

**(3)** The exhaustive set of pairwise constraints obtained by our approach can also potentially be used to improve the performance of other machine learning techniques by adjusting the similarity matrix.

The remainder of this paper is organized as follows. In Section 2, we propose an exhaustive and efficient constraint propagation approach. In Section 3, we exploit the exhaustive set of propagated pairwise constraints for spectral clustering. In Section 4, our approach is evaluated on image and UCI datasets. Finally, Section 5 gives the conclusions drawn from experimental results.

## 2 Exhaustive and Efficient Constraint Propagation

Given a dataset $\mathcal{X} = \{x_1, ..., x_N\}$, we denote a set of must-link constraints as $\mathcal{M} = \{(x_i, x_j) : z_i = z_j\}$ and a set of cannot-link constraints as $\mathcal{C} = \{(x_i, x_j) : z_i \neq z_j\}$, where $z_i$ is the label of data $x_i$. Our goal is to exploit the two types of

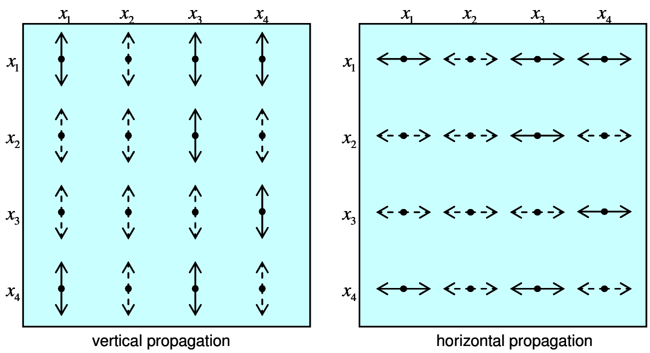



**Fig. 2.** The vertical and horizontal propagation of pairwise constraints. Each arrow denotes the direction of constraint propagation. The solid arrow means that the pairwise constraint is provided initially, while the dashed arrow means that the pairwise constraint is newly generated during constraint propagation.

pairwise constraints for spectral clustering on the dataset $\mathcal{X}$. As we have mentioned, the pairwise constraints can be used to adjust the similarities between data so that spectral clustering can be performed with the adjusted similarity matrix. In previous work [10], only the similarities between the constrained data are adjusted, and thus the pairwise constraints exert very limited effect on the subsequent spectral clustering. In the following, we propose an exhaustive and efficient constraint propagation technique that spreads the effect of pairwise constraints throughout the entire dataset, thereby enabling the pairwise constraints to exert a stronger influence on the subsequent spectral clustering.

A main obstacle of constraint propagation lies in that the cannot-link constraints are not transitive. In this paper, however, we succeed in propagating both must-link and cannot-link constraints. We first represent these two types of pairwise constraints using a single matrix $Z = \{Z_{ij}\}_{N\times N}$:

$$Z_{ij} = \begin{cases} +1, & (x_i, x_j) \in \mathcal{M}; \\ -1, & (x_i, x_j) \in \mathcal{C}; \\ 0, & \text{otherwise}. \end{cases} \tag{1}$$

Here, we have $|Z_{ij}| \leq 1$ for soft constraints [17,18]. Since we can directly obtain the pairwise constraints from the above matrix $Z$, the pairwise constraints have been represented using $Z$ without loss of information. We make further observations on $Z$ column by column. It can be observed that the $j$-th column $Z_{.j}$ actually provides the initial configuration of a *two-class semi-supervised learning problem* with respect to $x_j$, where the "positive class" contains the data that must appear together with $x_j$ and the "negative class" contains the data that cannot appear together with $x_j$. More concretely, $x_i$ can be initially regarded as coming from the positive (or negative) class if $Z_{ij} > 0$ (or $< 0$), but if $x_i$ and $x_j$ are not constrained (i.e. $Z_{ij} = 0$) thus $x_i$ is initially unlabeled. This configuration of a two-class semi-supervised learning is also suitable for soft constraints. The semi-supervised learning problem with respect to $x_j$ can be solved by the label propagation technique [14]. Since the other columns of $Z$ can be handled similarly, we can decompose the constraint propagation problem into $N$ independent label propagation subproblems which can then be solved in parallel. The vertical propagation of pairwise constraints is illustrated in Fig. 2.

However, it is also possible that a column contains no pairwise constraints (for example, see the second column in Fig. 2). That is, the entries of this column may all be zeros, and for such cases, no constraint propagation will occur along this column. We deal with this problem through horizontal constraint propagation (see Fig. 2), which is performed after the vertical constraint propagation. The horizontal propagation can be done similar to the vertical propagation discussed above. The only difference is that we now consider $Z$ row by row, instead of column-wise. More significantly, through combining the vertical and horizontal constraint propagation, we succeed in propagating the pairwise constraints to any pair of data. That is, the semi-supervised learning for constraint propagation could not break down if one type of constraints is missing for some data.

The two sets of constraint propagation subproblems can be solved efficiently through semi-supervised learning based on $k$-nearest neighbors graphs. Let $\mathcal{F} = \{F = \{F_{ij}\}_{N\times N} : |F_{ij}| \leq 1\}$. In fact, each matrix $F \in \mathcal{F}$ denotes a set of pairwise constraints with the associated confidence scores. That is, $F_{ij} > 0$ is equivalent to $(x_i, x_j) \in \mathcal{M}$ while $F_{ij} < 0$ is equivalent to $(x_i, x_j) \in \mathcal{C}$, with $|F_{ij}|$ being the confidence score (i.e. probability) of $(x_i, x_j) \in \mathcal{M}$ or $(x_i, x_j) \in \mathcal{C}$. Particularly, $Z \in \mathcal{F}$, where $Z$ collects the initial pairwise constraints. Given the affinity (or similarity) matrix $A$ for the dataset $\mathcal{X}$, our algorithm for constraint propagation is summarized as follows:

**(1)** Form the weight matrix $W$ of a graph by $W_{ij} = \frac{A(x_i,x_j)}{\sqrt{A(x_i,x_i)}\sqrt{A(x_j,x_j)}}$ if $x_j$ ($j \neq i$) is among the $k$-nearest neighbors ($k$-NN) of $x_i$ and $W_{ij} = 0$ otherwise. Set $W = (W + W^T)/2$ to ensure that $W$ is symmetric.
**(2)** Construct the matrix $\bar{\mathcal{L}} = D^{-1/2}WD^{-1/2}$, where $D$ is a diagonal matrix with its $(i,i)$-element equal to the sum of the $i$-th row of $W$.
**(3)** Iterate $F_v(t+1) = \alpha\bar{\mathcal{L}}F_v(t) + (1-\alpha)Z$ for vertical constraint propagation until convergence, where $F_v(t) \in \mathcal{F}$ and $\alpha$ is a parameter in the range $(0,1)$.
**(4)** Iterate $F_h(t+1) = \alpha F_h(t)\bar{\mathcal{L}} + (1-\alpha)F_v^*$ for horizontal constraint propagation until convergence, where $F_h(t) \in \mathcal{F}$ and $F_v^*$ is the limit of $\{F_v(t)\}$.
**(5)** Output $F^* = F_h^*$ as the final representation of the pairwise constraints, where $F_h^*$ is the limit of $\{F_h(t)\}$.

Below we give a convergence analysis of the above constraint propagation algorithm. Since the vertical constraint propagation in Step (3) can be regarded as label propagation, its convergence has been shown in [14]. More concretely, similar to [14], we can obtain $F_v^* = (1-\alpha)(I-\alpha\bar{\mathcal{L}})^{-1}Z$ as the limit of $\{F_v(t)\}$. As for the horizontal constraint propagation, we have

$$\begin{aligned} F_h^T(t+1) &= \alpha\bar{\mathcal{L}}^T F_h^T(t) + (1-\alpha){F_v^*}^T \\ &= \alpha\bar{\mathcal{L}}F_h^T(t) + (1-\alpha){F_v^*}^T. \end{aligned} \quad (2)$$

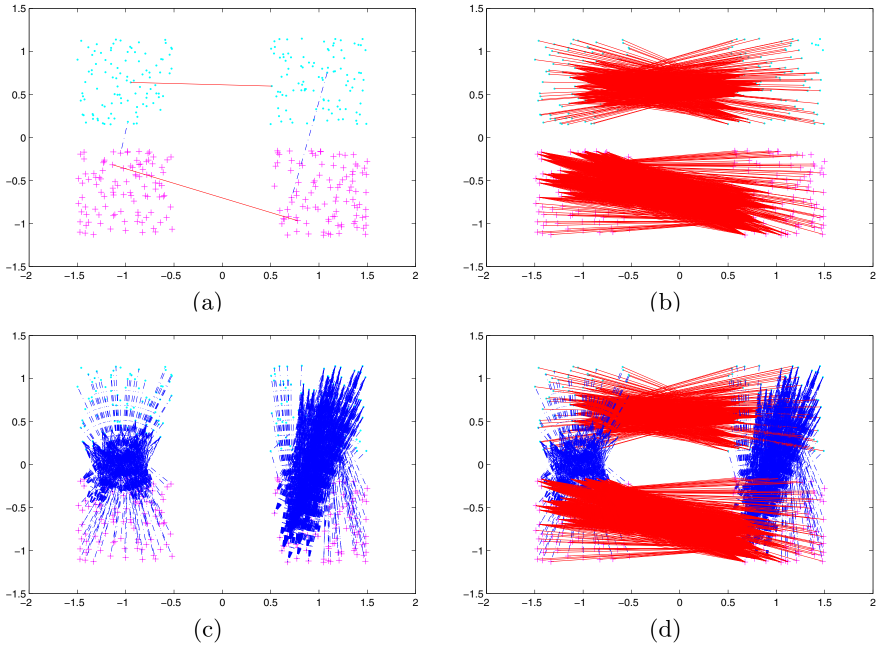



**Fig. 3.** The illustration of our constraint propagation: (a) four pairwise constraints and ideal clustering of the data; (b) final constraints propagated from only two must-link constraints; (c) final constraints propagated from only two cannot-link constraints; (d) final constraints propagated from four pairwise constraints. Here, must-link constraints are denoted by solid red lines, while cannot-link constraints are denoted by dashed blue lines. Moreover, we only show the propagated constraints with predicted confidence scores $> 0.1$ in Figs. 3(b)-3(d).

That is, the horizontal propagation in Step (4) can be transformed to a vertical propagation which converges to ${F_h^*}^T = (1-\alpha)(I-\alpha\bar{\mathcal{L}})^{-1}{F_v^*}^T$. Hence, our constraint propagation algorithm has the following closed-form solution:

$$
\begin{aligned}
F^* = F_h^* &= (1-\alpha)F_v^*(I-\alpha\bar{\mathcal{L}}^T)^{-1} \\
&= (1-\alpha)^2(I-\alpha\bar{\mathcal{L}})^{-1}Z(I-\alpha\bar{\mathcal{L}})^{-1}, \qquad (3)
\end{aligned}
$$

which actually accumulates the evidence to reconcile the contradictory propagated constraints for certain pairs of data. As a toy example, the propagated constraints given by the above equation are explicitly shown in Fig.3. We can find that the propagated constraints obtained by our approach are consistent with the ideal clustering of the data.

Finally, we give a complexity analysis of our constraint propagation algorithm. Through semi-supervised learning based on $k$-nearest neighbors graphs ($k \ll N$), both vertical and horizontal constraint propagation can be performed in quadratic time $O(kN^2)$. Since this time complexity is proportional to the total number of all possible pairwise constraints (i.e. $N(N-1)/2$), our algorithm can be considered computationally efficient. Moreover, our algorithm incurs significantly less computational cost than [12], given that constraint propagation based on semi-definite programming has a time complexity of $O(N^4)$.

## 3 Fully Constrained Spectral Clustering

It should be noted that the output $F^*$ of our constraint propagation algorithm represents an exhaustive set of pairwise constraints with the associated confidence scores $|F^*|$. Our goal is to obtain a data partition that is fully consistent with $F^*$. Here, we exploit $F^*$ for spectral clustering by adjusting the weight matrix $W$ as follows:

$$\tilde{W}_{ij} = \begin{cases} 1-(1-F^*_{ij})(1-W_{ij}), & F^*_{ij} \geq 0; \\ (1+F^*_{ij})W_{ij}, & F^*_{ij} < 0. \end{cases} \tag{4}$$

In the following, $\tilde{W}$ will be used for constrained spectral clustering. Here, we need to first prove that this matrix can be regarded as a weight matrix by showing that $\tilde{W}$ has the following nice properties.

**Proposition 1.** *(i) $\tilde{W}$ is nonnegative and symmetric; (ii) $\tilde{W}_{ij} \geq W_{ij}$ (or $< W_{ij}$) if $F^*_{ij} \geq 0$ (or $< 0$).*

*Proof.* The above proposition is proven as follows:

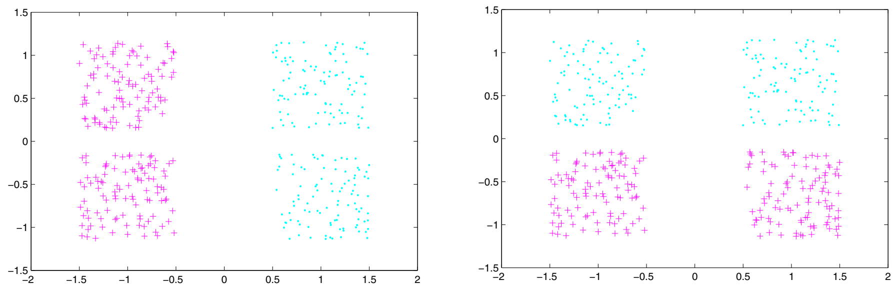
**(i)** The symmetry of both $W$ and $F^*$ ensures that $\tilde{W}$ is symmetric. Since $0 \leq W_{ij} \leq 1$ and $|F^*_{ij}| \leq 1$, we also have: $\tilde{W}_{ij} = 1-(1-F^*_{ij})(1-W_{ij}) \geq 1-(1-W_{ij}) \geq 0$ if $F^*_{ij} \geq 0$ and $\tilde{W}_{ij} = (1+F^*_{ij})W_{ij} \geq 0$ if $F^*_{ij} < 0$. That is, we always have $\tilde{W}_{ij} \geq 0$. Hence, $\tilde{W}$ is nonnegative and symmetric.

**(ii)** According to (4), we can consider $\tilde{W}_{ij}$ as a *monotonically increasing function* of $F^*_{ij}$. Since $\tilde{W}_{ij} = W_{ij}$ when $F^*_{ij} = 0$, we thus have: $\tilde{W}_{ij} \geq W_{ij}$ (or $< W_{ij}$) if $F^*_{ij} \geq 0$ (or $< 0$).

This proves that $\tilde{W}$ can be used as a weight matrix for spectral clustering. More importantly, according to Proposition 1, the new weight matrix $\tilde{W}$ is derived from the original weight matrix $W$ by increasing $W_{ij}$ for the must-link constraints with $F^*_{ij} > 0$ and decreasing $W_{ij}$ for the cannot-link constraints with $F^*_{ij} < 0$. This is entirely consistent with our original motivation of exploiting pairwise constraints for spectral clustering.

After we have incorporated the exhaustive set of pairwise constraints obtained by our constraint propagation into a new weight matrix $\tilde{W}$, we then perform spectral clustering with this weight matrix. The corresponding algorithm is summarized as follows:

**(1)** Find $K$ largest nontrivial eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_K$ of $\tilde{D}^{-1/2}\tilde{W}\tilde{D}^{-1/2}$, where $\tilde{D}$ is a diagonal matrix with its $(i,i)$-element equal to the sum of the $i$-th row of the weight matrix $\tilde{W}$.




(a) (b)

**Fig. 4.** The results of constrained clustering on the toy data using four pairwise constraints given by Fig. 3(a): (a) spectral learning [10]; (b) our approach. The clustering obtained by our approach is consistent with the ideal clustering of the data.

**(2)** Form $E = [\mathbf{v}_1, ..., \mathbf{v}_K]$, and normalize each row of $E$ to have unit length. Here, the $i$-th row $E_{i.}$ is the low-dimensional feature vector for data $x_i$.

**(3)** Perform $k$-means clustering on the new feature vectors $E_{i.}(i = 1, ..., N)$ to obtain $K$ clusters.

The clustering results on the toy data (see Fig. 3(a)) by the above algorithm are shown in Fig. 4(b). We can find that the clustering obtained by our approach is consistent with the ideal clustering of the data, while this is not true for spectral learning [10] without using constraint propagation (see Fig. 4(a)). In the following, since the pairwise constraints used for constrained spectral clustering (CSC) is obtained by our exhaustive and efficient constraint propagation (E$^2$CP), the above associated clustering algorithm is denoted as E$^2$CSC (or E$^2$CP directly) to distinguish it from other CSC algorithms.

## 4 Experimental Results

In this section, we conduct extensive experiments on real-life data to evaluate the proposed constrained spectral clustering algorithm. We first describe the experimental setup, including the clustering evaluation measure and the parameter selection. Moreover, we compare our algorithm with other closely related methods on two image datasets and four UCI datasets, respectively.

### 4.1 Experimental Setup

For comparison, we present the results of affinity propagation (AP) [11], spectral learning (SL) [10] and semi-supervised kernel k-means (SSKK) [4], which are three closely related constrained clustering algorithms. SL and SSKK adjust only the similarities between the constrained data, while AP and our E$^2$CP propagate the pairwise constraints throughout the entire dataset. Here, it should be noted that AP cannot directly address multi-class problems and we have to take into

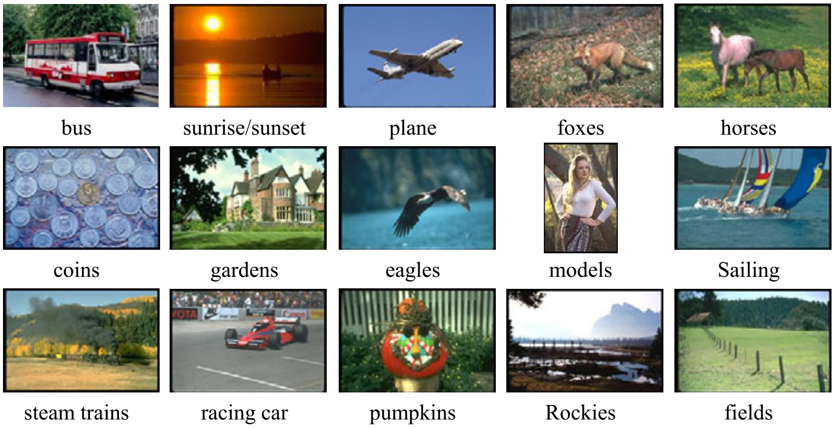



**Fig. 5.** Sample images from 15 categories of the Corel dataset

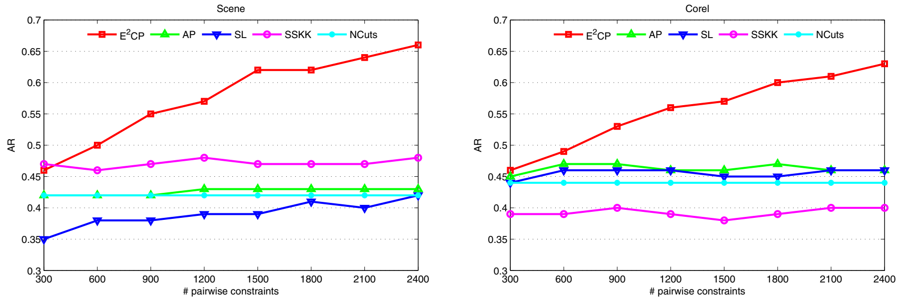
account the heuristic approach discussed in [11]. We also report the baseline results of normalized cuts (NCuts) [8], which is effectively a spectral clustering algorithm but without using pairwise constraints.

We evaluate the clustering results with the adjusted Rand (AR) index [19,20,21], which has been widely used for the evaluation of clustering algorithms. The AR index measures the pairwise agreement between the computed clustering and the ground truth clustering, and takes a value in the range [-1,1]. A higher AR index indicates that a higher percentage of data pairs in the obtained clustering have the same relationship (musk-link or cannot-link) as in the ground truth clustering. In the following, each experiment is randomly run 25 times, and the average AR index is obtained as the final clustering evaluation measure.

We set $\alpha = 0.8$ and $k = 20$ for our $E^2CP$ algorithm. The $k$-NN graph constructed for our constraint propagation is also used for the subsequent spectral clustering. To ensure a fair comparison, we adopt the same $k$-NN graph for the other algorithms. Here, we construct the graph with different kernels for image and UCI datasets. That is, the spatial Markov kernel [15] is defined on the image datasets to exploit the spatial information, while the Gaussian kernel is used for the UCI datasets as in [11]. For each dataset, different numbers of pairwise constraints are randomly generated using the ground-truth cluster labels.

### 4.2 Results on Image Datasets

We select two different image datasets. The first one contains 8 scene categories from MIT [22], including four man-made scenes and four natural scenes. The total number of images is 2,688. The size of each image in this Scene dataset is $256 \times 256$ pixels. The second dataset contains images from a Corel collection. We select 15 categories (see Fig. 5), and each of the categories contains 100 images. In total, this selected set has 1,500 images. The size of each image in this dataset is $384 \times 256$ or $256 \times 384$ pixels.




(a) (b)

**Fig. 6.** The clustering results on the two image datasets by different clustering algorithms with a varying number of pairwise constraints

For these two image datasets, we choose two different feature sets which are introduced in [23] and [15], respectively. That is, as in [23], the SIFT descriptors are used for the Scene dataset, while, similar to [15], the joint color and Gabor features are used for the Corel dataset. These features are chosen to ensure a fair comparison with the state-of-the-art techniques. More concretely, for the Scene dataset, we extract SIFT descriptors of $16 \times 16$ pixel blocks computed over a regular grid with spacing of 8 pixels. As for the Corel dataset, we divide each image into blocks of $16 \times 16$ pixels and then extract a joint color/texture feature vector from each block. Here, the texture features are represented as the means and standard deviations of the coefficients of a bank of Gabor filters (with 3 scales and 4 orientations), and the color features are the mean values of HSV color components. Finally, for each image dataset, we perform $k$-means clustering on the extracted feature vectors to form a vocabulary of 400 visual keywords. Based on this visual vocabulary, we then define a spatial Markov kernel [15] as the weight matrix for graph construction.

In the experiments, we provide the clustering algorithms with a varying number of pairwise constraints. The clustering results are shown in Fig. 6. We can find that our $E^2$CP generally performs the best among the five clustering methods. The effectiveness of our exhaustive constraint propagation approach to exploiting pairwise constraints for spectral clustering is verified by the fact that our $E^2$CP consistently obtains better results. In contrast, SL and SSKK perform unsatisfactorily, and, in some cases, their performance has been degraded to those of NCuts. This may be due to that by merely adjusting the similarities only between the constrained images, these approaches have not fully utilized the additional supervisory or prior information inherent in the constrained images, and hence can not discover the complex manifolds hidden in the challenging image datasets. Although AP can also propagate pairwise constraints throughout the entire dataset like our $E^2$CP, the heuristic approach discussed in [11] may not address multi-class problems for the challenging image datasets, which leads to unsatisfactory results. Moreover, another important observation is that the improvement in the clustering performance by our $E^2$CP with respect to NCuts becomes more obvious when more pairwise constraints are provided, while this

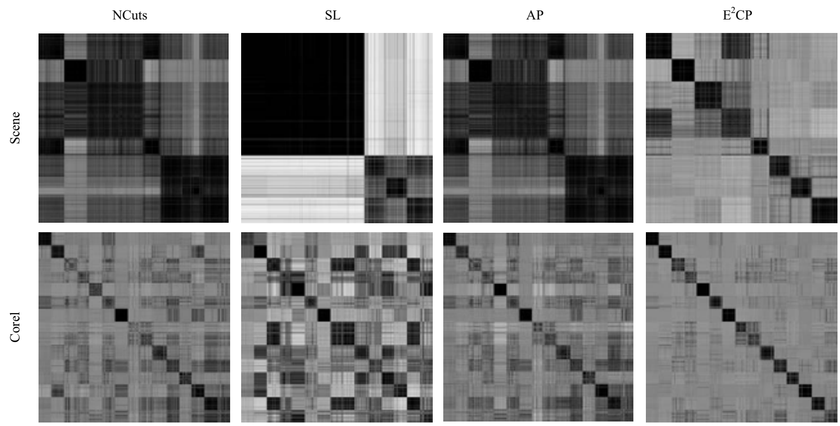



**Fig. 7.** Distance matrices of the low-dimensional data representations for the two image datasets obtained by NCuts, SL, AP, and E$^2$CP, respectively. For illustration purpose, the data are arranged such that images within a cluster appear consecutively. The darker is a pixel, the smaller is the distance.

is not the case for AP, SL or SSKK. In other words, the pairwise constraints has been exploited more exhaustively and effectively by our E$^2$CP.

To make it clearer how our E$^2$CP exploits the pairwise constraints for spectral clustering, we show the distance matrices of the low-dimensional data representations obtained by NCuts, SL, AP, and E$^2$CP in Fig. 7. We can find that the block structure of the distance matrices of the data representations obtained by our E$^2$CP on the two image datasets is significantly more obvious, as compared to those of the data representations obtained by NCuts, SL, and AP. This means that after being adjusted by our E$^2$CP, each cluster associated with the new data representation becomes more compact and different clusters become more separated. Hence, we can conclude that our E$^2$CP does lead to better spectral clustering through our exhaustive constraint propagation.

The pairwise constraints used here are actually very sparse. For example, the largest number of pairwise constraints (i.e. 2,400) used for constrained clustering are generated with only 2.6% of the images in the Scene dataset. Here, images from the same cluster form the must-link constraints while images from different clusters form the cannot-link constraints. Through our exhaustive constraint propagation, we obtain 3,611,328 pairwise constraints with nonzero confidence scores from this sparse set of pairwise constraints. That is, we have successfully propagated 2,400 pairwise constraints throughout the entire dataset.

It is noteworthy that the running time of our E$^2$CP is comparable to that of the constrained clustering algorithms without using constraint propagation (e.g. SL and NCuts). Moreover, as for the two constraint propagation approaches, our E$^2$CP runs faster than AP, particularly for multi-class problems. For example, the time taken by E$^2$CP, AP, SL, SSKK, and NCuts on the Scene dataset is 20, 42, 15, 17, and 12 seconds, respectively. We run all the five algorithms (Matlab code) on a PC with 2.33 GHz CPU and 2GB RAM.

### 4.3 Results on UCI Datasets

We further conduct experiments on four UCI datasets, which are described in Table 1. The UCI data are widely used to evaluate clustering and classification algorithms in machine learning. Here, as in [11], the Gaussian kernel is defined on each UCI dataset for computing the weight matrix during graph construction. The experimental setup on the UCI datasets is similar to that for the image datasets. The clustering results are shown in Fig. 8.

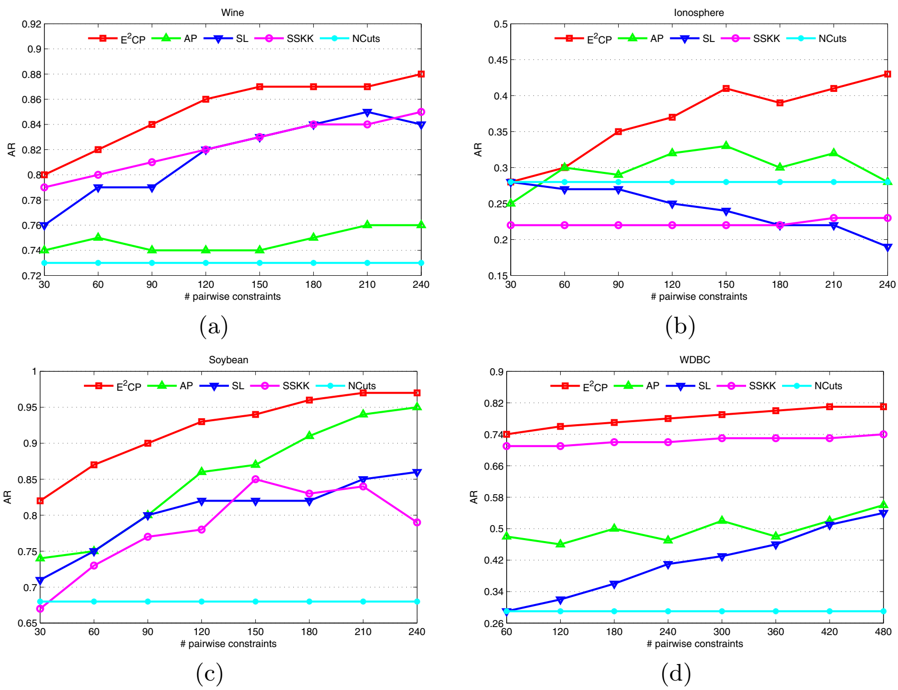



**Fig. 8.** The clustering results on the four UCI datasets by different clustering algorithms with a varying number of pairwise constraints

**Table 1.** Four UCI datasets used in the experiment. The features are first normalized to the range [-1, 1] for all the datasets.

| Datasets | Wine | Ionosphere | Soybean | WDBC |
|---|---|---|---|---|
| # samples | 178 | 351 | 47 | 569 |
| # features | 13 | 34 | 35 | 30 |
| # clusters | 3 | 2 | 4 | 2 |

Again, we can find that our $E^2CP$ performs the best in most cases. Moreover, the other three constrained clustering approaches (i.e. AP, SL, and SSKK) are shown to have generally benefited from the pairwise constraints as compared to NCuts. This observation is different from that on the image datasets. As we have mentioned, this may be due to that, considering the complexity of the image datasets, a more exhaustive propagation (like our $E^2CP$) of the pairwise constraints is needed in order to fully utilize the inherent supervisory information provided by the constraints. Our experimental results also demonstrated that an exhaustive propagation of the pairwise constraints in the UCI data through our $E^2CP$ leads to improved clustering performance over the other three constrained clustering approaches (i.e. AP, SL, and SSKK).

## 5 Conclusions

We have proposed an exhaustive and efficient constraint propagation approach to exploiting pairwise constraints for spectral clustering. The challenging constraint propagation problem for both the must-link and cannot-link constraints is decomposed into a set of independent label propagation subproblems, which can then be solved efficiently and in parallel through semi-supervised learning based on $k$-nearest neighbors graphs. The resulting exhaustive set of propagated pairwise constraints with associated confidence scores are further used to adjust the weight matrix for spectral clustering. It is worth noting that this paper first clearly shows how pairwise constraints are propagated independently and then accumulated into a conciliatory closed-form solution. Experimental results on image and UCI datasets demonstrate clearly that by exhaustively propagating the pairwise constraints throughout the entire dataset, our approach is able to fully utilize the additional supervisory or prior information inherent in the constrained data for spectral clustering and then achieve superior performance compared to the state-of-the-art techniques. For future work, our approach will also be used to improve the performance of other graph-based methods by exhaustively exploiting the pairwise constraints.

## Acknowledgements

The work described in this paper was supported by a grant from the Research Council of Hong Kong SAR, China (Project No. CityU 114007) and a grant from City University of Hong Kong (Project No. 7008040).

## References

1. Wagstaff, K., Cardie, C., Rogers, S., Schroedl, S.: Constrained k-means clustering with background knowledge. In: ICML, pp. 577–584 (2001)
2. Klein, D., Kamvar, S., Manning, C.: From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In: ICML, pp. 307–314 (2002)

3. Basu, S., Bilenko, M., Mooney, R.: A probabilistic framework for semi-supervised clustering. In: KDD, pp. 59–68 (2004)
4. Kulis, B., Basu, S., Dhillon, I., Mooney, R.: Semi-supervised graph clustering: A kernel approach. In: ICML, pp. 457–464 (2005)
5. Lu, Z., Peng, Y.: A semi-supervised learning algorithm on Gaussian mixture with automatic model selection. Neural Processing Letters 27, 57–66 (2008)
6. Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems, vol. 14, pp. 849–856 (2002)
7. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17, 395–416 (2007)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
9. Veksler, O.: Star shape prior for graph-cut image segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 454–467. Springer, Heidelberg (2008)
10. Kamvar, S., Klein, D., Manning, C.: Spectral learning. In: IJCAI, pp. 561–566 (2003)
11. Lu, Z., Carreira-Perpinan, M.: Constrained spectral clustering through affinity propagation. In: CVPR (2008)
12. Li, Z., Liu, J., Tang, X.: Pairwise constraint propagation by semidefinite programming for semi-supervised classification. In: ICML, pp. 576–583 (2008)
13. Yu, S., Shi, J.: Segmentation given partial grouping constraints. IEEE Trans. on Pattern Analysis and Machine Intelligence 26, 173–183 (2004)
14. Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: Advances in Neural Information Processing Systems, vol. 16, pp. 321–328 (2004)
15. Lu, Z., Ip, H.: Image categorization by learning with context and consistency. In: CVPR, pp. 2719–2726 (2009)
16. Lu, Z., Ip, H.: Combining context, consistency, and diversity cues for interactive image categorization. IEEE Transactions on Multimedia 12, 194–203 (2010)
17. Law, M., Topchy, A., Jain, A.: Clustering with soft and group constraints. In: Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, pp. 662–670 (2004)
18. Law, M., Topchy, A., Jain, A.: Model-based clustering with probabilistic constraints. In: Proceedings of SIAM Data Mining, pp. 641–645 (2005)
19. Hubert, L., Arabie, P.: Comparing partitions. Journal of Classification 2, 193–218 (1985)
20. Lu, Z., Peng, Y., Xiao, J.: From comparing clusterings to combining clusterings. In: AAAI, pp. 665–670 (2008)
21. Lu, Z., Peng, Y., Ip, H.: Gaussian mixture learning via robust competitive agglomeration. Pattern Recognition Letters 31, 539–547 (2010)
22. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV 42, 145–175 (2001)
23. Bosch, A., Zisserman, A., Muñoz, X.: Scene classification via pLSA. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 517–530. Springer, Heidelberg (2006)