

# Semidefinite Clustering for Image Segmentation with A-priori Knowledge

Matthias Heiler<sup>1</sup>, Jens Keuchel<sup>2</sup>, and Christoph Schnörr<sup>1</sup>

<sup>1</sup> Computer Vision, Graphics, and Pattern Recognition Group,  
Department of Mathematics and Computer Science,  
University of Mannheim, 68131 Mannheim, Germany  
{heiler, schnoerr}@uni-mannheim.de

<sup>2</sup> ETH Zurich, Institute of Computational Science,  
Hirschengraben 84, CH-8092 Zurich  
Jens.Keuchel@inf.ethz.ch

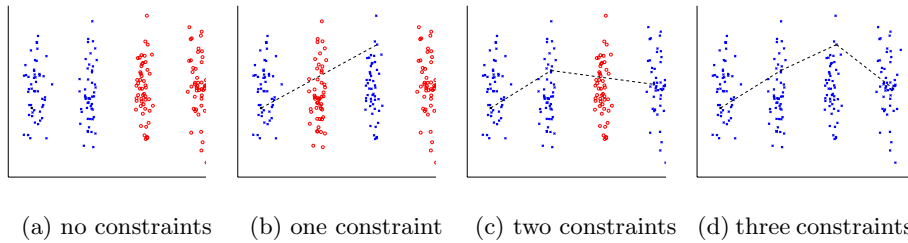
**Abstract.** Graph-based clustering methods are successfully applied to computer vision and machine learning problems. In this paper we demonstrate how to introduce a-priori knowledge on class membership in a systematic and principled way: starting from a convex relaxation of the graph-based clustering problem we integrate information about class membership by adding linear constraints to the resulting semidefinite program. With our method, there is no need to modify the original optimization criterion, ensuring that the algorithm will always converge to a high quality clustering or image segmentation.

## 1 Introduction

When working on clustering problems we often have some a-priori knowledge available: certain samples are known to belong to the same class of objects, or we can make assumptions on the size of the clusters. Occasionally, multiple different clusterings are meaningful and we want to target an algorithm toward one particularly interesting solution. In the extreme case we have a small set of labeled objects and want to generalize their labels to a larger set of new, unseen objects. Instead of training a classifier on the labeled objects only we can employ semi-supervised clustering for this task. This application is usually referred to as *transductive inference*.

Fig. 1 visualizes the idea: given a dataset with a number of “sensible looking” clusterings, find the best (here: binary) clustering consistent with some a-priori information on common class membership. This information is provided in the form of *equivalence constraints* on the class labels of some points. For instance, in Fig. 1(b) a point from the left-most cluster is linked to a point in the middle cluster by a constraint which forces these points to have equal class labels. Interestingly, our results show that such a constraint does not influence these two points only, but the information is propagated through their corresponding clusters.

Our work relates to *graph-based clustering* methods used in machine learning and computer vision [1,2]. Since the corresponding *combinatorial optimization*



**Fig. 1.** Effects of a-priori information: adding constraints (dashed lines) between few points leads to completely different clusterings

problems are NP-hard, a common approach is to compute approximative solutions using eigenvectors or min-flow calculations [1,3]. In this paper, we use an alternative technique that is based on a *semidefinite programming (SDP) relaxation*. Besides its conceptual advantages over spectral relaxation, this method has recently been applied successfully in the context of machine learning [4] and image partitioning [5].

Concerning a-priori information, most spectral and min-flow methods currently require to modify the cost function of the original clustering problem [1,6]. In contrast, the semidefinite relaxation method puts additional constraints on the set of admissible solutions and finds a high quality solution according to the *original* clustering criterion within this restricted set. For the special problem of semi-supervised image segmentation, besides graph-based optimization techniques [7,6] various other approaches were also presented recently [8,9].

We introduce the graph-based clustering framework used in Section 2 and explain how to integrate a-priori knowledge on cluster size and membership. Section 3 presents our semidefinite relaxation approach along with a geometric interpretation. Some experiments in Section 4 show that adding very few constraints already yields appealing results. Section 5 concludes the paper.

## 2 Graph-Based Clustering

In order to cluster  $n$  objects we need to compute a suitable similarity matrix  $W \in \mathbb{R}^{n \times n}$  with  $W_{ij}$  being large when the objects  $i$  and  $j$  are similar. Interpreting the objects as vertices of a fully connected graph  $G(V, E)$  with edge weights  $W_{ij}$ , a classical *binary* partitioning approach from spectral graph theory (see, e.g., [10,11]) is based on the following problem formulation:

$$\max_{x \in \{-1, +1\}^n} x^\top W x \quad \iff \quad \min_{x \in \{-1, +1\}^n} x^\top L x \quad (1)$$

where  $L = \text{diag}(We) - W$  denotes the Laplacian matrix of the graph (with  $e = (1, \dots, 1)^\top \in \mathbb{R}^n$ ). Problem (1) has a clear interpretation: find a binary partitioning with maximum similarity of the objects within each cluster, or, equivalently, determine a cut through  $G$  with minimal weight.

Unfortunately, problem (1) can result in very imbalanced partitions, especially when the similarity matrix  $W$  contains positive entries only: putting every object into one cluster produces the optimal cut of weight 0. As a remedy, different approaches have been proposed in the literature. Several authors (e.g. [3,12]) suggest to scale the objective function in (1) appropriately in order to favor balanced cuts. Another approach uses an additional *balancing constraint*,

$$c^\top x = a, \quad (2)$$

where  $a \geq 0$  specifies the difference between the weighted number of objects in each cluster. For example, setting  $c = e$ ,  $a = 0$  requires that  $G$  is partitioned into clusters of identical size (equipartition problem [10]).

As the resulting problems are NP-hard they are often solved approximately using spectral techniques: dropping the integer constraint, extremal eigenvectors of  $W$  or  $L$  (or of normalized versions of these matrices) are computed and thresholded according to some suitable criterion. In Section 3, we propose a different method to relax and solve constrained problems of type (1), which not only takes the integer constraint on  $x$  into account more accurately than spectral techniques [5], but also permits to include linear and quadratic constraints on  $x$  *without* changing the original objective function.

**Incorporating a-priori knowledge:** Aside from similarities of the objects given by  $W$ , we may often know that some objects belong to the same class. For two objects  $i$  and  $j$ , this is modeled in our framework by the constraint

$$x_i x_j = 1. \quad (3)$$

Conversely, if  $i$  and  $j$  belong to *different* classes, we can use the constraint

$$x_i x_j = -1. \quad (4)$$

In contrast to other approaches [13], adding such *non-equivalent* constraints with our method is not more difficult than adding *is-equivalent* constraints (3): both lead to *quadratic equalities*.

Another example of a-priori information was given above: if the *size* of a cluster is known in advance we can use the linear constraint (2) with an appropriate value for  $a$  to demand a corresponding partitioning of  $G$ .

Note that in contrast to established methods [1], integrating a-priori knowledge into our framework leads to very clear and concise models: entries of the similarity matrix and the corresponding graph remain unchanged. We do not alter the original problem more than absolutely necessary to account for the additional information.

### 3 Semidefinite Programming (SDP) Relaxation

In [5], an approach to approximately solve the combinatorial problem (1) with an additional balancing constraint (2) is presented. This method basically consists

of three steps: first, the decision variables are lifted into a higher-dimensional space where the corresponding problem is relaxed to a *convex* optimization problem [14]. Then, the global optimum of this relaxation is found using interior point techniques. Finally, the decision variables are recovered from the solution using a small number of random hyperplanes [15]. Next, we extend this idea to take a-priori knowledge into account by adding constraints of the form (3) and (4).

The basic lifting step of the SDP relaxation is based on the observation that the objective function in (1) can be rewritten in the form of a standard matrix inner product as  $x^\top W x = \text{tr}(W x x^\top) =: W \bullet x x^\top$ . Interpreting this as an optimization problem in a higher dimensional matrix space, the relaxation consists of replacing the positive semidefinite *rank one* matrix  $x x^\top \in \mathbb{R}^{n \times n}$  by a positive semidefinite matrix  $X \succeq 0$  of *arbitrary rank*. Since the combinatorial constraints on the entries of  $x$  in (1) can be lifted easily into this matrix space by requiring  $X_{ii} = 1$ , we obtain the following basic relaxation of (1):

$$\begin{aligned} & \max_{X \succeq 0} W \bullet X \\ & \text{subject to } X_{ii} = 1 \quad \forall i = 1, \dots, n \end{aligned} \tag{5}$$

While solving this relaxation is trivial in case of a positive matrix  $W$  (cf. Section 2), it is also applicable if  $W$  contains negative entries. Moreover, note that the integer constraint on  $x$  is still accounted for in (5), which contrasts spectral relaxation techniques which usually drop it completely [3].

Problem (5) belongs to the class of *semidefinite programs*, for which the global optimum can be computed to arbitrary precision in polynomial time (see, e.g., [16]). For this problem class, the additional constraints on  $x$  which describe the a-priori knowledge can easily be incorporated by lifting them into the matrix space: the balancing constraint (2) is squared to become  $c^\top x x^\top c = a^2$ , which results in the linear constraint  $cc^\top \bullet X = a^2$  after relaxation. Each equivalence constraint of the form (3) can be transformed directly to  $X_{ij} = X_{ji} = 1$ . To represent this as a linear constraint based on a symmetric matrix, we use the alternative formulation  $X_{ij} + X_{ji} = 2$ . Due to the constraints  $X_{ii} = 1$  and the fact that  $X$  is positive semidefinite, this imposes no additional relaxation as no entry of  $X$  can become larger than 1. Equivalently, the non-equivalence constraints (4) are represented by  $X_{ij} + X_{ji} = -2$ .

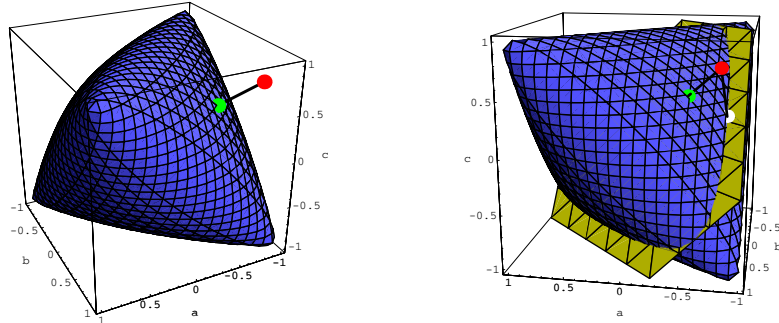
Let  $P_1$  ( $P_2$ ) denote the set containing the pairs  $(i, j)$  of objects that are known to belong to the same class (different classes). Representing all constraints in linear form, we finally obtain the following semidefinite program:

$$\begin{aligned} & \max_{X \succeq 0} W \bullet X \\ & \text{subject to } e_i e_i^\top \bullet X = 1 \quad \forall i = 1, \dots, n \end{aligned} \tag{6a}$$

$$(e_i e_j^\top + e_j e_i^\top) \bullet X = 2 \quad \forall (i, j) \in P_1 \tag{6b}$$

$$(e_i e_j^\top + e_j e_i^\top) \bullet X = -2 \quad \forall (i, j) \in P_2 \tag{6c}$$

$$cc^\top \bullet X = a^2 \tag{6d}$$



**Fig. 2.** SDP clustering as matrix approximation. Depicted is the set  $M_{\succeq 0}$  of all positive semidefinite symmetric matrices of dimension  $3 \times 3$  with diagonal fixed to unity (blue object). For a matrix  $W$  with negative eigenvalues (red point) the solution  $X$  of problem (5) (green point) is given by the projection of  $W$  onto  $M_{\succeq 0}$  (left). Incorporating a-priori information about relative cluster sizes (2) leads to an additional linear constraint (right, green plane). The SDP relaxation finds a solution  $X$  (white point) satisfying this linear constraint and approximating  $W$  (red point) optimally.

where  $e_i \in \mathbb{R}^n$  denotes the  $i$ th standard unit vector. Note that the (non-)equivalence constraints (3),(4) can also be combined into a single constraint: adding the matrices from (6b),(6c) as  $E_{P_k} = \sum_{(i,j) \in P_k} (e_i e_j^\top + e_j e_i^\top)$  gives the equivalent constraints  $E_{P_1} \bullet X = 2|P_1|$  and  $E_{P_2} \bullet X = -2|P_2|$ , respectively. As already mentioned above, this results in no further relaxation.

After a solution  $X$  of (6) is found we apply the *randomized hyperplane technique* [15] to recover a binary solution  $x$ . In this step, no adaption is necessary to enforce the additional (non-)equivalence constraints (3),(4) as the corresponding constraints (6b),(6c) already do this efficiently. Depending on the application we also may not enforce the balancing constraint (2): since the a-priori knowledge on the size of the clusters usually is given only approximately, it merely serves as a bias to guide the search for convenient clusters than as a strict requirement. For more details on the SDP relaxation approach, we refer to [5].

Fig. 2 visualizes the geometry of our method: solving (6) corresponds to projecting the problem matrix  $W$  (which is not necessarily positive semidefinite) onto the set  $M_{\succeq 0}$  of all positive semidefinite matrices with diagonal fixed to unity, which is equivalent to finding the closest approximation of  $W$  within  $M_{\succeq 0}$ .

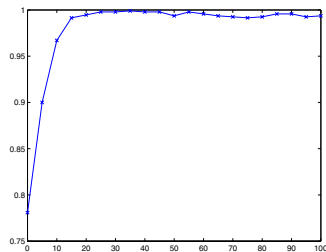
The resulting solution matrix  $X \in M_{\succeq 0}$  is, by construction, positive semidefinite and therefore can be interpreted as a matrix whose entries are inner products of points located on the unit sphere in some Euclidean space. The randomized hyperplane algorithm then places a cut through this sphere and retrieves a binary clustering which maximizes the original objective function in (1). Geometrically, this is a projection of  $X$  onto the closest vertex of the set  $M_{\succeq 0}$  (Fig. 2, left).

The linear constraints (6b)–(6d) further limit the set of admissible solutions: whereas constraint (6d) represents a plane cutting through  $M_{\succeq 0}$  (Fig. 2,

right), the constraints (6b),(6c) correspond to tangential planes. Together, they move the solution  $X$  toward vertices obeying the a-priori knowledge *and* representing a good clustering measured in terms of the original objective function  $x^\top Wx$ .

## 4 Experiments

As a proof-of-concept we created a very simple dataset consisting of four clearly separated clusters of 50 points each distributed according to a Gaussian distribution (Fig. 1(a)). Using a centered Gaussian kernel as similarity measure we compute multiple clusterings by subsequently adding constraints on the class labels. Fig. 1 shows that the constraints were met in each case and led to completely different clusterings. Although we did not provide a-priori information about the size of the clusters the centered kernel favored balanced solutions and flipped the unconstrained clusters accordingly.

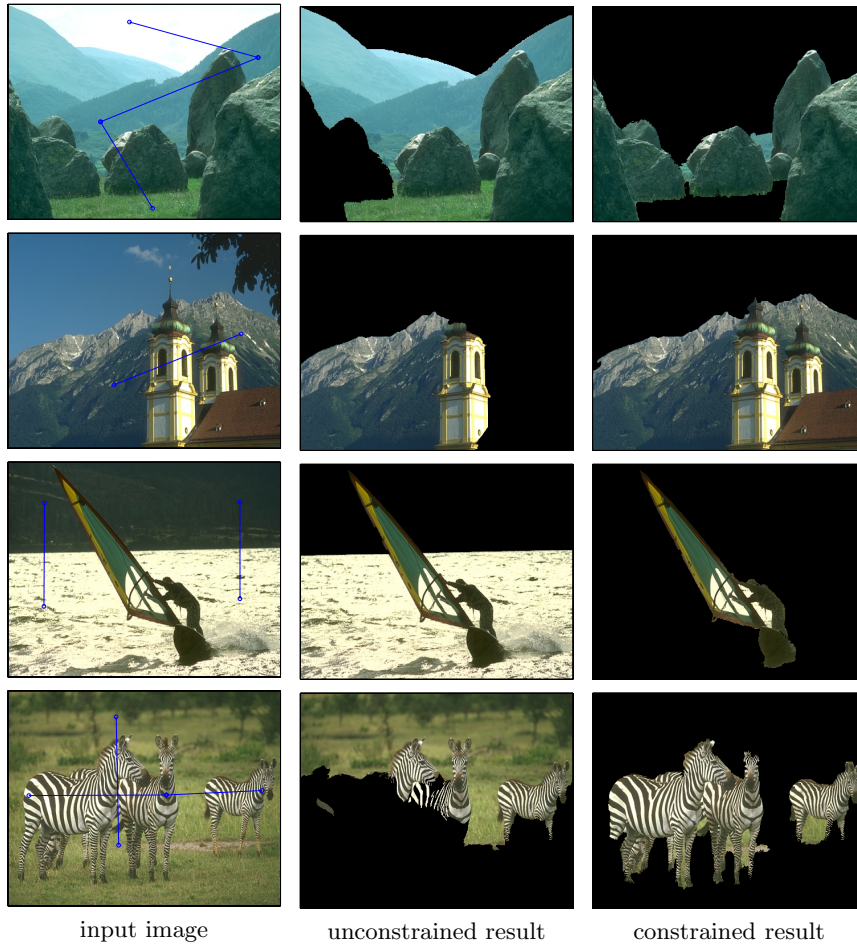


**Fig. 3.** The soybean experiment

truth we generated random constraints on the training sets, clustered, and measured accuracies on the corresponding test sets. The mean performance over 10 repeated experiments is visualized in Fig. 3: without any constraints 78% of the instances are correctly clustered. This is worse than the 87% reported for kmeans [17]. However, with only 5 constraints this improves to 90% correctly clustered points ( $\approx 88\%$  for kmeans), and we need only 15 constraints to achieve an accuracy of 99% (kmeans needs 100 constraints). Thus, for this dataset adding a-priori constraints is highly effective, leading to dramatic improvements in accuracy.

In Fig. 4 we show segmentations obtained for images from the Berkeley segmentation dataset [19] using a similarity measure based on color and spatial proximity (cf. [21]). In order to reduce the problem size appropriately, the images were over-segmented in a preprocessing step by applying the mean shift algorithm [20], which results in less than 1000 image patches [21]. These are clustered by our SDP relaxation with and without additional equivalence constraints. It is clearly visible that adding very few constraints can lead to dramatically different and visually more appealing segmentations.

Following [17] we tested our approach on the soybean dataset which comprises 35 attributes for 47 objects from four classes. To apply our method to this *multiclass problem* we assigned each class a *two bit binary code* and clustered on the corresponding binary digits using a centered exponential kernel as similarity measure. As in [17] consistency w.r.t. the known correct clustering was determined using the *Rand index* [18] and 10-fold cross-validation. With the class labels as ground



**Fig. 4.** Effects of prior information on image segmentation. An unsupervised segmentation based on color and spatial proximity may not partition the image in a visually meaningful way (2<sup>nd</sup> column). Adding only 1–3 equivalence constraints (blue lines in 1<sup>st</sup> column) can dramatically improve the segmentation (3<sup>rd</sup> column).

## 5 Conclusion

We presented a method for clustering and segmentation based on a semidefinite relaxation of the well-known minimal cut problem on graphs. The advantage over alternative approaches is that it allows incorporating a-priori knowledge in the clustering process without changing the target function. Instead, available equivalence information is modeled by additional constraints on the optimization problem. This simplifies interpretation of the results and ensures that different constraints can be combined arbitrarily.

We gave two examples for a-priori information which lead to linear constraints on the set of admissible solutions of the semidefinite relaxation and explained their geometric meaning. In an experimental section we showed that the method works in practice and can lead to improved image segmentation results.

In the future, we will investigate how to integrate further types of a-priori information and evaluate the method for constrained multiclass-clustering. Besides using binary codes, binary clustering can be applied hierarchically [21] or the SDP relaxation can be extended to multiclass settings.

## References

1. A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *ICML*, pp. 19–26, 2001.
2. Y. Weiss, "Segmentation using eigenvectors: A unifying view," in *ICCV*, pp. 975–982, 1999.
3. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE PAMI*, vol. 22, no. 8, pp. 888–905, 2000.
4. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. Jordan, "Learning the kernel matrix with semi-definite programming," in *ICML*, pp. 323–330, 2002.
5. J. Keuchel, C. Schnörr, C. Schellewald, and D. Cremers, "Binary partitioning, perceptual grouping, and restoration with semidefinite programming," *IEEE PAMI*, vol. 25, no. 11, pp. 1364–1379, 2003.
6. S. X. Yu and J. Shi, "Segmentation given partial grouping constraints," *IEEE PAMI*, vol. 26, no. 2, pp. 173–183, 2004.
7. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in *ICCV*, vol. 1, pp. 105–112, 2001.
8. L. Hermes and J. M. Buhmann, "Semi-supervised image segmentation by parametric distributional clustering," in *Energy Min. Meth. in Comp. Vis. a. Patt. Recog. (EMMCVPR)*, no. 2683 in LNCS, pp. 229–245, Springer, 2003.
9. R. Nock and F. Nielsen, "Grouping with bias revisited," in *CVPR*, 2004.
10. B. Mohar and S. Poljak, "Eigenvalues in combinatorial optimization," in *Combinatorial and Graph-Theoretical Problems in Linear Algebra* (R. Brualdi, S. Friedland, and V. Klee, eds.), vol. 50 of *IMA Vol. Math. Appl.*, pp. 107–151, Springer, 1993.
11. P. Perona and W. Freeman, "A factorization approach to grouping," in *ECCV'98* (H. Burkhardt and B. Neumann, eds.), LNCS, pp. 655–670, Springer, 1998.
12. S. Sarkar and P. Soundararajan, "Supervised learning of large perceptual organization: Graph spectral partitioning and learning automata," *IEEE PAMI*, vol. 22, no. 5, pp. 504–525, 2000.
13. T. Hertz, N. Shental, A. Bar-Hillel, and D. Weinshall, "Enhancing image and video retrieval: Learning via equivalence constraints," in *CVPR*, 2003.
14. L. Lovász and A. Schrijver, "Cones of matrices and set-functions and 0-1 optimization," *SIAM J. Optimization*, vol. 1, no. 2, pp. 166–190, 1991.
15. M. Goemans and D. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. of the ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.
16. Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming*. SIAM, 1994.
17. K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, "Constrained k-means clustering with background knowledge," in *ICML*, pp. 577–584, 2001.



18. W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. of the Am. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
19. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *ICCV*, pp. 416–423, 2001.
20. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE PAMI*, vol. 24, no. 5, pp. 603–619, 2002.
21. J. Keuchel, C. Schnörr, and M. Heiler, "Hierarchical image segmentation based on semidefinite programming," in *Proc. DAGM*, 2004.