

Binary Partitioning, Perceptual Grouping, and Restoration with Semidefinite Programming¹

Jens Keuchel², Christoph Schnörr, Christian Schellewald, Daniel Cremers

CVGPR-Group

Dept. of Mathematics and Computer Science

University of Mannheim

D-68131 Mannheim, Germany

`{jkeuchel,schnoerr,cschelle,cremers}@uni-mannheim.de`

<http://www.cvgpr.uni-mannheim.de>

Abstract

We introduce a novel optimization method, semidefinite programming, to the field of computer vision and apply it to the combinatorial problem of minimizing quadratic functionals in binary decision variables subject to linear constraints. The approach is (tuning) parameter free and computes high-quality combinatorial solutions using interior-point methods (convex programming) and a randomized hyperplane technique. Apart from a symmetry condition, no assumptions like metric pairwise interactions, for instance, are made with respect to the objective criterion. As a consequence, the approach can be applied to a wide range of problems. Applications to unsupervised partitioning, figure-ground discrimination and binary restoration are presented along with extensive ground-truth experiments. From the viewpoint of relaxation of the underlying combinatorial problem, we show the superiority of our approach to relaxations based on spectral graph theory and prove performance bounds.

Keywords: Image partitioning, segmentation, graph cuts, perceptual grouping, figure-ground discrimination, combinatorial optimization, relaxation, convex optimization, convex programming

¹A preliminary version [57] of this work was presented at the conference “*Energy Minimization Methods in Computer Vision and Pattern Recognition — EMMCVPR 2001*”.

²Corresponding author

1 Introduction

1.1 Motivation and Overview

Optimization problems occur in almost all fields of computer vision and pattern recognition. One of the most important design decisions concerns the compromise between adequacy of the optimization criterion and the difficulty to compute the solution. An inadequate optimization criterion will not solve the application problem, no matter how easy it is to compute the optimum. Conversely, sophisticated criteria which can only be optimized with sufficient *a priori* knowledge like, e.g., a good starting point or after elaborate parameter tuning are useless in practice as well. For this reason optimization approaches are attractive which help make a “good” compromise in this sense.

In this paper we introduce a novel optimization technique, *semidefinite programming*, to the field of computer vision and apply it to minimize quadratic functionals defined over binary decision variables and subject to linear constraints. Numerous problems in computer vision including partitioning and grouping lead to combinatorial optimization problems of this type. In contrast to related work, no specific assumptions are made with respect to the functional form besides a symmetry condition. As a consequence, our approach covers graph-optimization problems, unsupervised and supervised classification tasks, and first-order Markov random field estimates without depending on specific assumptions or problem formulations. Therefore it can be utilized for a wider range of applications.

The combinatorial complexity of the optimization task is dealt with in two steps: Firstly, the decision variables are lifted to a higher-dimensional space where the optimization problem is relaxed to a *convex* optimization problem. Specifically, the resulting semidefinite program comprises a linear objective functional which is defined over a cone in some matrix space, and a number of application-dependent linear constraints. Secondly, the decision variables are recovered from the global optimum of the relaxed problem by using a small set of randomly computed hyperplanes.

Using this optimization technique amounts to a compromise as follows. Advantageous properties are:

- + The original combinatorial problem is transformed to an optimization problem which is *convex*. As a consequence, the *global optimum* of the *transformed* problem can be computed under mild conditions.
- + Using an interior-point algorithm, an ϵ -approximation to this global optimum can be numerically determined in *polynomial time*.
- + *No tuning parameters* are necessary.
- + In contrast to spectral relaxation, *no choice of a suitable threshold value* is necessary. This makes our approach especially suited for unsupervised classification tasks.



Figure 1: A color scene (left) and a gray-value scene comprising some natural textures (right). How to partition such scenes into coherent groups in an unsupervised way based on pairwise (dis)similarities between local measurements?

On the negative side, we have:

- The number of variables is squared.

This limits the application to problems with up to several hundred variables which is, however, sufficient for many problems related to image partitioning and perceptual grouping. Furthermore, the increase of the problem dimension is *necessary* in order to approximate an intricate combinatorial problem by a simpler *convex* optimization problem! Intuitively, nasty combinatorial constraints in the original space are lifted to a higher-dimensional matrix space where these constraints can be better approximated by convex sets which, in turn, are more convenient for numerical optimization. Hence, we add:

- + *High-quality* combinatorial solutions can be computed by solving an appropriate convex optimization problem.

“High-quality” means that the solutions obtained are not far from the unknown global optimum (the computation of which is NP-hard and thus intractable) in terms of the original optimization criterion.

The absence of any specific assumptions about the objective criterion as well as the “+”-properties listed above motivated this investigation.

1.2 Partitioning, Grouping, and Restoration

In this section, we illustrate three problems which lead to different instances of the class of optimization problems considered in this paper. By this we wish to indicate the significance of this problem class for computer vision and to exemplify some non-trivial specific problems. More formal problem definitions will be given in the following sections.

Figure 1 shows two images taken from the VisTex-database at MIT [61]. A common goal of low-level computer vision is to partition such images in an unsupervised way into



Figure 2: Section of an office table shown from the top. The keyboard probably first attracts the attention of the observer. How to compute this figure-ground discrimination (global decision) based on pairwise (dis)similarities between local measurements?

coherent groups based on some locally computed features (color, texture, motion, ...). To this end, the representation of images by graphical structures has recently attracted the interest of researchers [58, 36, 27, 11]. We will show below that when using our approach, some of the assumptions made in the literature concerning admissible objective criteria can be dropped. Moreover, we study in detail the unsupervised bi-partitioning of images by constrained minimal cuts of the underlying graphs and show that, from the optimization point of view, our convex approximation provides a tighter relaxation of the underlying combinatorial optimization problem than recently suggested methods which are based on spectral graph theory.

Figure 2 shows a section of an office table from the top. Probably most human observers focus on the (partially occluded) keyboard first. A typical problem of computer vision is to model such global decisions by solving an optimization problem defined in terms of locally extracted primitives [55, 34, 63]. In this context, the optimization criterion is considered as a saliency measure with respect to decision variables indicating which primitive belongs to the fore- or background, respectively. We show below that quadratic saliency measures which have been considered as difficult [63] due their combinatorial complexity can conveniently be dealt with using our approach.

Figure 3 shows a noisy map of Iceland. The restoration of such images has a long history, in particular in the context of Markov random fields [30, 29, 64, 43]. We will show below that such binary restoration problems can be modeled under less restrictive conditions than with previous approaches.



Figure 3: A noisy binary image (map of Iceland) to be restored.

1.3 Related Work

1.3.1 Optimization Approaches in Computer Vision

Many energy-minimization problems in computer vision like image labeling and partitioning, perceptual grouping, graph matching etc., involve discrete decision variables and therefore are intrinsically combinatorial by nature. Accordingly, optimization approaches to efficiently compute good minimizers have a long history in the literature.

An important class of optimization approaches is based on stochastic sampling which was introduced by Geman and Geman [30] and has been widely applied in the Markov Random Field (MRF) literature [43, 64]. As is well-known, corresponding algorithms are very slow due to the annealing schedules prescribed by theory. Nevertheless, there has been a renewed interest during the last years in conjunction with Bayesian reasoning [40] and complex statistical models (e.g., [71, 68]). For further aspects we refer to [24].

To speed up computations, approaches for computing suboptimal Markov random field estimates like the ICM-algorithm [6], the highest-confidence-first heuristic [12], multi-scale approaches [33], and other approximations [67, 10] were developed. A further important class of approaches comprises continuation methods like Leclerc’s partitioning approach [42], the graduated-non-convexity strategy by Blake and Zisserman [7], and various deterministic (approximate) versions of the annealing approach in applications like surface reconstruction [28], perceptual grouping [34], graph matching [32], or clustering [54, 35].

Apart from simulated annealing (with annealing schedules that are unpractically slow for real-world applications), none of the above-mentioned approaches can guarantee to find the global minimum. And in general, this goal is elusive due to the combinatorial complexity of these minimization problems. Consequently, the important question concerning the *approximation* of the problem arises: How good is a computed minimizer relative to the unknown global optimum? Can a certain quality of solutions in terms of its suboptimality be guaranteed in *each* application? To the best of our knowledge, none of the approaches above (apart from simulated annealing) seems to be immune against getting trapped in some poor local minimum and hence does not meet these criteria.

A further problem relates to the *algorithmic properties* of these approaches. Apart from simple greedy strategies [6, 12], most approaches involve some (sometimes hidden) parameters on which the computed local minimum critically depends. A typical example is given by the artificial temperature parameter in deterministic annealing approaches and the corresponding iterative annealing schedule. It is well known [56] that such approaches exhibit complex bifurcation phenomena, the transitions of which (that is, which branch to follow) cannot be controlled by the user. Furthermore, these approaches involve highly nonlinear numerical fixed-point iterations which tend to oscillate in a parallel (synchronous) update mode (see [34, p. 906] and [50]).

Our approach belongs to the mathematically well-understood class of *convex* optimization problems and contributes to both of the problems discussed above. Firstly, there exists a global optimum under mild assumptions which, in turn, leads to a suboptimal solution of the original problem, along with clear numerical algorithms to compute it. Abstracting from the computational process, we can simply think of a mapping taking the data to this solution. Thus, evidently, no hidden parameter is involved. Secondly, under certain conditions bounds can be derived with respect to the quality of the suboptimal solution. At present, these bounds are not tight with respect to the much better performance measured in practice. Yet it should be noted that for alternative optimization approaches, performance bounds and a corresponding route of research seem to be missing³.

1.3.2 Graph Partitioning, Clustering, Perceptual Grouping

As illustrated in Section 1.2, there is a wide range of problems to which our optimization approach can be applied. While an in-depth discussion of all possible applications is not possible, we next briefly discuss work which relates to the applications we use to illustrate our optimization approach.

Graph partitioning. Approaches to *unsupervised* image segmentation by graph partitioning have been proposed by [46, 58, 27], and references therein. Images are represented by graphs $G(V, E)$ with locally extracted image features as vertices V and pairwise (dis)similarity values as edge weights $w : E \subseteq V \times V \rightarrow \mathbb{R}_0^+$ (Figure 4). A classical approach for the efficient computation of suboptimal cuts is based on the spectral decomposition of the Laplacian matrix [22]. This approach has found applications in many different fields [46]. Accordingly, Shi and Malik [58] propose the “normalized cut” criterion which minimizes the weight of a cut subject to normalizing terms to prevent unbalanced cuts. The resulting combinatorial problem is relaxed using methods from spectral graph theory. For a survey of further work in this direction we refer to [62]. Based on [38], Gdalyahu et al. [27] suggest to compute partitions as “typical average” cuts of the underlying graph using a stochastic sampling method. Although this approach is very interesting, it does not directly relate to an optimization criterion and therefore will not be further discussed.

³A recent notable exception with respect to a more restricted class of optimization problems is [11].

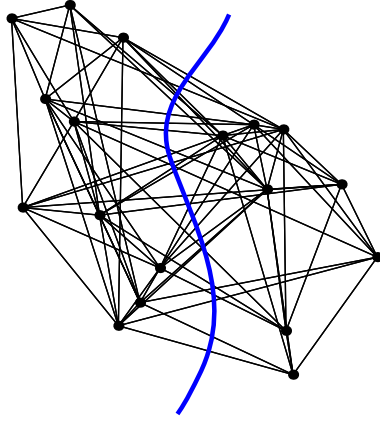


Figure 4: Representing image partitions by graph cuts: The weights of all edges cut provide a measure for the dissimilarity of the resulting groups.

It has been criticized in [27] that methods based on spectral graph theory are not able to partition highly skewed data distributions and non-compact clusters. We will show below, however, that a straightforward remedy is to base the similarity measure on a suitable path metric. Furthermore, our approach yields a tighter relaxation of the underlying combinatorial problem and hence better suboptimal solutions.

Recent approaches to *supervised* graph partitioning (image labeling) include [36, 11] and references therein. These authors consider the case of non-binary labels x_i and the following class of optimization criteria:

$$\sum_{i \in V} D_i(x_i) + \sum_{(i,j) \in E} P_{i,j}(x_i, x_j) \quad (1)$$

While Boykov et al. [11] require $P_{i,j}(\cdot, \cdot)$ to be a semi-metric, Ishikawa [36] makes the stronger assumption $P_{i,j}(x_i, x_j) = P(x_i - x_j)$ with P being convex. In this paper, we consider the case of binary labels and do not make any assumptions with respect to pairwise interaction terms $P_{i,j}$. Unlike a semi-metric, for example, $P_{i,j}$ may not vanish for $x_i = x_j$ or can be negative.

Clustering. The approaches to unsupervised image partitioning discussed above may also be understood as clustering methods, of course. In this paper, we focus in detail on image bi-partitioning by computing constrained minimal cuts of the underlying graph. Consecutive partitions thus lead to a hierarchical clustering method (cf., e.g., [39]).

An important issue in this context is cluster normalization in order to avoid too unbalanced partitions. In general, normalization criteria lead to rational non-quadratic terms of the cost functional [20, 35] to which our optimization approach cannot be directly applied. Below, however, we investigate cluster normalization by imposing various linear constraints, as introduced by Shi and Malik [58] as a relaxation of their specific non-quadratic normalization criterion. Recently, the (natural) use of path-metrics for (dis)similarity-based clustering was

also advocated in [23].

Perceptual grouping. There is a vast literature on perceptual grouping in vision. For a survey, we refer to [55] and, e.g., [37, 2] and references therein. In this paper, we merely focus from an optimization point-of-view on the quadratic saliency measure of Herault and Horaud [34] the application of which has been considered as difficult due to its combinatorial complexity [63]. We show below that this grouping criterion can conveniently be optimized using our approach.

1.3.3 Mathematical Programming

Semidefinite programming is a relatively novel optimization technique which has been successfully applied to optimization problems in such diverse fields as nonconvex and combinatorial optimization, statistics, or control theory. For a survey, we refer to [65].

The method used in this paper for relaxing combinatorial constraints goes back to the seminal work of Lovász and Schrijver [44]. Concerning interior-point methods for convex programming, we refer to numerous textbooks [48, 66, 69] and surveys [59, 65].

To perform step 2 of our approach we adopt the randomized hyperplane technique developed by Goemans and Williamson [31]. For a classical optimization problem from combinatorial graph theory (max-cut problem) these authors were able to show that suboptimal solutions (for the specific problem considered) cannot be worse than 14% relative to the unknown global optimum. Besides the convenient algorithm design based on convex optimization, this fact has motivated our work.

1.4 Organization of the Paper

In Section 2, we formally define three optimization problems related to unsupervised partitioning (Section 2.1), perceptual grouping or figure-ground discrimination (Section 2.2), and binary image restoration (Section 2.3). The mathematical relaxation of combinatorial problems of this class is the subject of Section 3. We explain the derivation of a corresponding semidefinite program (Section 3.1), its feasibility (Section 3.2), related algorithms (Section 3.3), performance bounds (Section 3.4), and the superiority of convex relaxation to spectral relaxation (Section 3.5). We discuss numerical results for real scenes and ground-truth experiments in Section 4. We conclude and indicate further work in Section 5.

1.5 Notation

The following notation will be used throughout the paper. For basic concepts from graph theory, we refer to, e.g., [19].

e	vector of all ones: $e = (1, \dots, 1)^\top$
$D(x)$	diagonal matrix with vector x on its diagonal: $D_{ii} = x_i$
$D(X)$	matrix X with off-diagonal elements set to zero
I	unit matrix $I = D(e)$
$\mathcal{S}^n$	space of symmetric $n \times n$ matrices $X^\top = X$
$\mathcal{S}_+^n$	set of matrices $X \in \mathcal{S}^n$ which are positive semidefinite
$X \bullet Y$	standard matrix scalar product $X \bullet Y = \text{Tr}[X^\top Y]$
$G(V, E)$	undirected graph with vertices $V = \{1, \dots, n\}$ and edges $E \subseteq V \times V$
w	weight function of the graph G : $w : E \rightarrow \mathbb{R}_0^+$
$w(S)$	sum of edge-weights of the subgraph induced by the vertex subset $S \subseteq V$
$\bar{S}$	complement $V \setminus S$ of the vertex subset $S \subset V$
$w(\delta S)$	weight of a cut defined by the partition $S, \bar{S}$
W	weighted adjacency matrix of graph G : $W_{ij} = w(i, j)$, $i, j \in V$
L	Laplacian matrix of graph G : $L = D(We) - W$
$\lambda_k(L)$	eigenvalues $\lambda_1(L) \leq \dots \leq \lambda_n(L)$ of the Laplacian matrix L of graph G
$\ x\ $	Euclidian norm of the vector x : $\ x\ = x^\top x$

2 Problem Statement: Binary Combinatorial Optimization

In this section, we formally define optimization criteria according to the problems introduced in Section 1.2. These criteria will turn out to be special instances of quadratic functionals over binary decision variables subject to linear constraints. Relaxations of these difficult combinatorial problems for computing suboptimal solutions in polynomial time will be studied in Section 3.

2.1 Unsupervised Partitioning

Consider a graph $G(V, E)$ with locally extracted image features as vertices V and pairwise (dis)similarity values as edge-weights $w : E \subseteq V \times V \rightarrow \mathbb{R}_0^+$. We wish to compute a partition of the set V into two coherent groups $V = S \cup \bar{S}$ as depicted in Figure 4. Representing such partitions by an indicator vector $x \in \{-1, +1\}^n$, the weight of a cut is given by (cf. Section 1.5):

$$w(\delta S) = \sum_{i \in S, j \in \bar{S}} w(i, j) = \frac{1}{8} \sum_{i, j \in V} w(i, j) (x_i - x_j)^2 = \frac{1}{4} x^\top L x . \quad (2)$$

If the weight function w encodes a similarity measure between pairs of features, then coherent groups correspond to low values of $w(\delta S)$.

In order to avoid unbalanced partitions which are likely when minimizing $w(\delta S)$, Shi and Malik [58] suggested the following normalized objective function:

$$\frac{w(\delta S)}{w(S) + w(\delta S)} + \frac{w(\delta \bar{S})}{w(\bar{S}) + w(\delta \bar{S})} .$$

Since this optimization problem is intractable, they derived the following relaxation (*normalized cut*):

$$\inf_x \frac{x^\top Lx}{x^\top D(We)x}, \quad e^\top D(We)x = 0, \quad x \in \{-b, 1\}^n, \quad (3)$$

where the number b is *not* known beforehand. Hence, this integer constraint is dropped in practice. Writing D as a shorthand for $D(We)$ and $y = D^{1/2}x$, problem (3) becomes:

$$\inf_{\|y\|=1} y^\top \tilde{L}y, \quad e^\top D^{1/2}y = 0, \quad \tilde{L} = D^{-1/2}LD^{-1/2}.$$

Since e is the eigenvector of the Laplacian matrix L with eigenvalue 0, so is $D^{1/2}e$ with respect to the normalized Laplacian matrix $\tilde{L}$. Consequently, $x = D^{-1/2}y$ solves (3) (without the integer constraint), where y is the eigenvector of $\tilde{L}$ corresponding to the second smallest eigenvalue. Finally, the integer constraint is taken into account by thresholding the eigenvector using some suitable criterion [58].

The approach (3) is close to the following classical partitioning approach from spectral graph theory (see, e.g., [46] for a survey):

$$\inf_x x^\top Lx, \quad e^\top x = 0, \quad x \in \{-1, +1\}^n. \quad (4)$$

Criterion (4) has a clear interpretation: Determine a cut with minimal weight (cf. (2)) subject to the constraint that each group has an equal number of vertices: $e^\top x = 0$. Similarly, dropping the normalization in (3) and setting $x \in \{-1, +1\}^n$, the interpretation of (3) is: Determine a cut with minimal weight such that each partition has an equal number of degrees⁴. In the context of image partitioning, this may be more appropriate than the constraint in (4) and thus explains the success of Shi and Malik’s approach.

The foregoing discussion raises some natural questions: Which other constraints are useful for unsupervised image partitioning? How to take into account the integer constraint with respect to x_i , $i = 1, \dots, n$, for deriving an appropriate relaxation of the combinatorial optimization problem (as opposed to doing that afterwards just by thresholding)? To investigate different constraints, we define the following criterion for unsupervised image partitioning

$$\inf_x x^\top Lx, \quad c^\top x = 0, \quad x \in \{-1, +1\}^n, \quad (5)$$

and focus in Section 3 on an appropriate relaxation of the integer constraint. The vector c in (5) is an application-dependent constraint vector defining what we mean by a “balanced cut”. Examples will be given in Section 4.2.

If the vector c is chosen inappropriately, the problem (5) may have no feasible solution (a simple example is the case $c = e$ with n being an odd number). In this case, an additional variable $x_0 \in \mathbb{R}$ can be incorporated to close the gap in the balancing constraint: $c^\top x + x_0 = 0$.

⁴The degree $(We)_i$ of vertex i is the sum of incident edge-weights.

To minimize x_0 , it is also included into the objective function, hence arriving at the following problem formulation:

$$\inf_{(x, x_0)} \begin{pmatrix} x \\ x_0 \end{pmatrix}^\top \begin{pmatrix} L & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} x \\ x_0 \end{pmatrix}, \quad \begin{pmatrix} c \\ 1 \end{pmatrix}^\top \begin{pmatrix} x \\ x_0 \end{pmatrix} = 0, \quad x \in \{-1, +1\}^n, \quad (6)$$

where a is a sufficiently large number. As an infeasible instance of problem (5) does not necessarily lead to an infeasible relaxation, this extension is not always necessary for our approach to work properly. Exact conditions which the vector c has to satisfy so that the relaxation considered in this paper is feasible will be given in Section 3.2. Since in practice, however, these conditions are satisfied for all the applications studied in Section 4, the extension (6) will not be further considered here.

Remark 1

We note that problem (5) does not conform with optimization problems of the form (1) and corresponding assumptions.

2.2 Perceptual Grouping and Figure-Ground Discrimination

Hérault and Horaud [34] investigated the following combinatorial minimization problem for figure-ground discrimination and perceptual grouping in terms of binary labels $p \in \{0, 1\}^n$ for n primitives:

$$E_{saliency}(p) + \lambda E_{constraint}(p), \quad \lambda \in \mathbb{R}^+, \quad (7)$$

where

$$E_{saliency}(p) = - \sum_{i,j} w(i, j) p_i p_j, \quad E_{constraint}(p) = \left(\sum_i p_i \right)^2. \quad (8)$$

The interaction coefficients $w(i, j)$ encode similarity measures between pairs of primitives like cocircularity, smoothness, proximity, or contrast (see [34]). Accordingly, the first term in (7) measures the mutual reinforcement between pairs of primitives i, j labelled as foreground: $p_i = p_j = 1$. The second term in (7) penalizes primitives which do not receive much “feedback” from other primitives and probably do not belong to some coherent group.

Hérault and Horaud investigated various annealing approaches in order to find good minimizers of (7). The disadvantages of this class of optimization approaches were discussed in Section 1.3.1. Accordingly, in a recent comparison [63], the combinatorial complexity involved has been considered as a decisive disadvantage of using this approach as a saliency measure for perceptual grouping.

We wish to show below that a good minimizer can conveniently be computed with our approach. To this end, we transform the 0/1-variables p to ± 1 -variables $x = 2p - e$ and obtain the following problem formulation (up to constant terms):

$$\inf_x \frac{1}{4} x^\top (\lambda e e^\top - W) x + \frac{1}{2} e^\top (\lambda n I - W) x, \quad x \in \{-1, +1\}^n, \quad (9)$$

with entries of the matrix $W_{ij} = w(i, j)$. The formulation (9) makes more explicit the role of the global parameter λ which acts in a twofold way as a threshold parameter: Primitives i, j reinforce each other if their similarity value W_{ij} is larger than λ (first term), and each primitive i is (additionally) favored if its *average* degree (similarity value) $(We)_i/n$ is larger than λ (second term). Both terms together result in a meaningful global measure of “coherency” based on pairwise comparisons of locally computed primitives.

Remark 2

We note again that problem (9) does not conform with optimization problems of the form (1) and corresponding assumptions.

2.3 Restoration

Consider some scalar-valued feature (gray-value, color feature, texture measure, etc.) $g \in \mathbb{R}^n$ which has been locally computed within the image plane. Suppose that for each pixel i , the feature-value g_i is known to originate from either of two prototypical values u_1, u_2 . In practice, of course, g is real-valued due to measurement errors and noise.

To restore a discrete-valued image function given by the vector $x \in \{-1, +1\}^n$ from the measurements g , we wish to minimize the following functional:

$$z(x) = \frac{1}{4} \sum_i ((u_2 - u_1)x_i + u_2 + u_1 - 2g_i)^2 + \frac{\lambda}{2} \sum_{\langle i, j \rangle} (x_i - x_j)^2. \quad (10)$$

Here, the second term sums over all pairwise adjacent pixels on the regular image grid.

Functional (10) comprises two terms familiar from many regularization approaches [5]: A data-fitting term and a smoothness term modeling spatial context. However, due to the integer constraint $x_i \in \{-1, 1\}$, the optimization problem considered here is much more difficult than standard regularization problems.

Up to constant terms, functional (10) leads to the following optimization problem:

$$\inf_x \frac{1}{4} x^\top Q x + \frac{1}{2} b^\top x, \quad x \in \{-1, +1\}^n, \quad (11)$$

with

$$b_i = (u_2 - u_1)(u_2 + u_1 - 2g_i), \quad Q_{ij} = \begin{cases} -2\lambda & : i, j \text{ adjacent} \\ 0 & : \text{otherwise} \end{cases}.$$

Notice that in this case, in contrast to the problems introduced in the previous sections, the problem matrix Q is very sparse, which is an advantage from the computational point of view.

Remark 3

We are well aware that problem (11) does conform with optimization problems of the form (1) and thus can be solved to optimality using the methods presented in [36, 11]. However, depending on the application considered, it might be useful to modify the terms in (10) to

model properties of the imaging device (data-fitting term) or to take into consideration *a priori* known spatial regularities (smoothness term; see, e.g., [9, 64]). These modifications would lead to other entries for Q and b , which could violate the assumptions for (1) but would not affect the applicability of our approach. Exploring these possibilities, however, is beyond the scope of this paper.

3 Optimization by Mathematical Relaxation

In this section we introduce the optimization approach to solve the different problems presented in the previous section. To this end, notice that the perceptual grouping problem (9) as well as the restoration problem (11) both can be written in the form (5) introduced for the unsupervised partitioning problem. To see that, one has to homogenize the objective functions of (9) and (11), which are both special cases of a general quadratic functional $x^\top Qx + 2b^\top x$, in the following way:

$$x^\top Qx + 2b^\top x = \begin{pmatrix} x \\ 1 \end{pmatrix}^\top L \begin{pmatrix} x \\ 1 \end{pmatrix}, \quad L = \begin{pmatrix} Q & b \\ b^\top & 0 \end{pmatrix}.$$

Hence, both problems are special instances of problem (5) with $c = 0$ and size $n + 1$, if L is generalized to be a symmetric matrix which is subject to no further constraints. Indeed, we do not need L to be the Laplacian matrix of a graph for the following relaxation. Therefore, in this section we will only assume that $L \in \mathcal{S}^n$, and we will always refer to problem (5). The results then apply to all three problems from Section 2, except it is stated otherwise by a special choice of the constraint vector c .

The relaxation approach now consists of two steps: Firstly, the problem variables are lifted into a matrix space, and the combinatorial constraints are *weakly* incorporated in that space (Section 3.1), thus yielding a semidefinite relaxation of (5). This lifting step has been introduced in a more general setting by Lovász and Schrijver [44]. For the resulting *convex* optimization problem a solution is computed using interior point techniques. Secondly, a combinatorial solution must be found based on the relaxed solution computed in the first step. To this end, we apply the randomized hyperplane technique developed by Goemans and Williamson [31] (Section 3.3). We also provide bounds on the quality of the solutions (Section 3.4) and compare these with a spectral relaxation approach (Section 3.5).

3.1 Semidefinite Relaxation

In order to relax the problem (5), we first replace the linear and integer constraint, respectively, by quadratic ones: $(c^\top x)^2 = 0$ and $x_i^2 - 1 = 0$, $i = 1, \dots, n$. Denoting the Lagrangian multiplier variables with y_i , $i = 0, \dots, n$, the Lagrangian of (5) reads:

$$x^\top Lx - y_0(c^\top x)^2 - \sum_{i=1}^n y_i(x_i^2 - 1) = x^\top (L - y_0 cc^\top - D(y))x + e^\top y. \quad (12)$$

The corresponding minimax-problem is:

$$\sup_{y_0, y} \inf_x x^\top (L - y_0 cc^\top - D(y))x + e^\top y .$$

Since x is unconstrained now, the inner minimization is finite-valued only if $L - y_0 cc^\top - D(y)$ is positive semidefinite. Hence we arrive at the relaxed problem:

$$z_d := \sup_{y_0, y} e^\top y , \quad L - y_0 cc^\top - D(y) \in \mathcal{S}_+^n . \quad (13)$$

The important point here is that problem (13) is a *convex* optimization problem! The set $\mathcal{S}_+^n$ is a cone (i.e. a special convex set) which also is self-dual, so that it coincides with its dual cone [48] given by:

$$(\mathcal{S}_+^n)^* = \{Y : X \bullet Y \geq 0, X \in \mathcal{S}_+^n\} .$$

To obtain a connection to our original problem, we derive the Lagrangian dual of (13). Choosing a Lagrangian multiplier $X \in \mathcal{S}_+^n$, similar reasoning as above yields:

$$\begin{aligned} z_d &= \sup_{y_0, y} \inf_{X \in \mathcal{S}_+^n} e^\top y + X \bullet (L - y_0 cc^\top - D(y)) \\ &\leq \inf_{X \in \mathcal{S}_+^n} \sup_{y_0, y} e^\top y + X \bullet (L - y_0 cc^\top - D(y)) \\ &= \inf_{X \in \mathcal{S}_+^n} \sup_{y_0, y} L \bullet X - D(y) \bullet (X - I) - y_0 cc^\top \bullet X . \end{aligned}$$

Here the inner maximization of the last equation is finite only if $D(X) = I$ and $cc^\top \bullet X = 0$. Hence, we obtain the following problem dual to (13):

$$z_p := \inf_{X \in \mathcal{S}_+^n} L \bullet X , \quad cc^\top \bullet X = 0 , \quad D(X) = I , \quad (14)$$

which again is convex.

In order to compare the objective function of this final semidefinite relaxation (14) with that of the original problem (5), we rewrite the latter as follows:

$$\inf_x x^\top Lx = \inf_x L \bullet xx^\top .$$

Note that the matrix $xx^\top$ is positive semidefinite and has rank one. A comparison with the relaxed problem (14) shows that $xx^\top$ is replaced by an arbitrary matrix $X \in \mathcal{S}_+^n$ (i.e. the rank one condition is dropped), and the constraints are lifted to the higher dimensional space accordingly.

In order to illustrate how the *convex* relaxation (14) approximates the *combinatorial, non-convex* problem (5), let us consider the case $n = 3$. In this case the matrix X in (14) has six unknowns due to symmetry (the upper (or lower) triangular part). The intersection of the convex set $\mathcal{S}_+^n$ with the three hyperplanes defined by $D(X) = I$ yields the convex set $\{X \in$

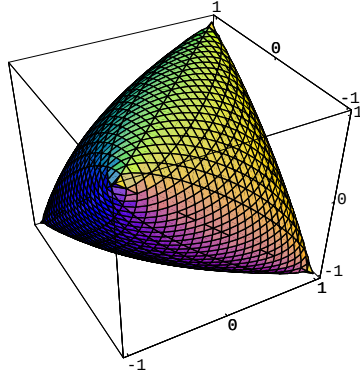


Figure 5: The set of feasible solutions $\{X \in \mathcal{S}_+^n | D(X) = I\}$ for the convex problem relaxation (14), for $n = 3$ and $c = 0$. The subset of feasible combinatorial solutions only contains the four extreme points. For $c \neq 0$, this set additionally has to be intersected with the hyperplane $cc^\top \bullet X = 0$.

$\mathcal{S}_+^n | D(X) = I\}$, which is shown in Figure 5. It looks like a polytope with four vertices (which correspond to the combinatorial solutions of the unrelaxed problem) but with non-linear faces. The set of feasible solutions for (14) is now obtained by additionally intersecting this set with the hyperplane $cc^\top \bullet X = 0$. This shows that the original combinatorial problem (5) only has a feasible solution if at least one of the extreme points of $\{X \in \mathcal{S}_+^n | D(X) = I\}$ lies on this hyperplane. Nevertheless, the *relaxed* solution can always be determined by minimizing numerically the linear functional $L \bullet X$ over the feasible set, provided that this is not empty (see Section 3.2). The nearest extreme point may then be considered as the combinatorial solution of (5), or at least as the combinatorial solution which best approximates the constraint $c^\top x = 0$.

This simple example illustrates a fundamental fact: Intricate constraints can be represented by simpler sets in higher-dimensional spaces. This fact is well-known in other fields like pattern recognition and statistical learning [14, 60, 18].

3.2 Duality and Feasibility

The primal and dual optimization problems (14) and (13) both belong to the class of positive semidefinite programs. The elegant duality theory corresponding to this class of convex optimization problems can be found in [48]. The following duality theorem is also provided there:

Theorem 1 (Strong duality for positive semidefinite programming) *If (14) and (13) both are feasible and there is a strictly interior point for either (14) or (13), then optimal primal and dual solutions $X^*, (y_0^*, y^*)$ exist and the corresponding optimal values are the same, i.e. they yield no duality gap:*

$$z_p - z_d = L \bullet X^* - e^t y^* = 0 .$$

The convex optimization problems considered in this paper are known to be “well-behaved” according to this theorem. A strictly interior point for the dual problem (13) can always be found by setting $y_0 = 0$, $y = -ae$ with a large enough. For the primal problem (14), a feasible solution is given by $X = I$ for $c = 0$ and by $X = \frac{n}{n-1}I - \frac{1}{n-1}ee^\top$ for $c = e$. Otherwise, it may be possible that no feasible solution for the primal problem exists if the vector c is chosen inappropriately. This situation is illustrated by the following example: For the case $n = 2$, the constraints in (14) yield $X = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$ with $a = -\frac{c_1^2 + c_2^2}{2c_1c_2}$. Additionally, $-1 \leq a \leq 1$ has to hold for X to be positive semidefinite. Obviously, this is only valid for $c_1 = c_2$. Thus for all other choices of c the primal problem (14) has no feasible solution in this case.

In general, it is possible to characterize exactly the situations when the primal problem (14) is feasible. The following result is mainly based on a proposition given in [41]:

Theorem 2 *The problem (14) is feasible for a positive constraint vector c ($c_i \geq 0$, $i = 1, \dots, n$) if and only if c is balanced, i. e.*

$$c_i \leq \sum_{j \neq i} c_j \quad \text{for all } i = 1, \dots, n.$$

Proof. As $c \geq 0$, the matrix $cc^\top$ is positive semidefinite. The balancing constraint yields $0 = cc^\top \bullet X = \text{Tr}[cc^\top X] = \sum_i \lambda_i(cc^\top X)$. As X is also positive semidefinite, so is $cc^\top X$, and it follows immediately that $cc^\top X$ has to be the null-matrix. As this is equivalent to $Xc = 0$, c must be contained in the null space $\ker(X)$. Proposition 3.2 in [41] now concludes the proof: The linear space generated by c is contained in $\ker(X)$ for a matrix $X \in \mathcal{S}_+^n$ with $D(X) = I$ if and only if c is balanced. The proof for this proposition can be found in [16]. $\square$

Due to this result, we will only consider examples in Section 4.2 where c is balanced.

3.3 Interior Point Algorithm and Randomized Hyperplanes

To compute the optimal solutions X^* and (y_0^*, y^*) , a wide range of iterative interior-point algorithms can be used. Typically, a sequence of minimizers $\{X_\eta, (y_0)_\eta, y_\eta\}$, parameterized by a parameter η , is computed until the duality gap falls below some threshold ϵ . A remarkable result in [48] asserts that for the family of self-concordant barrier functions, this can always be done in polynomial time, with the complexity depending on the number of variables n and the value of ϵ .

Notice that due to the constraint $cc^\top \bullet X = 0$, the smallest eigenvalue of X has to be equal to 0 (cf. proof of Theorem 2), so that no strictly interior point for the primal problem (14) exists. Because of this observation we decided to use the dual-scaling algorithm from [4] for our experiments. This algorithm has the advantage that it does not need to calculate an interior solution for the primal problem during the iterations, but only for the dual problem. Moreover, it is capable to exploit the sparsity structure of a given problem better than other methods. The

primal solution matrix X^* is not computed until the optimal dual solution (y_0^*, y^*) has been reached.

Based on this solution matrix X^* to the convex optimization problem (14), we have to find a combinatorial solution x to the original problem (5). To achieve this, we used the randomized-hyperplane technique proposed by Goemans and Williamson [31]. To this end, the following interpretation of the relaxation described in Section 3.1 is more convenient: Since $X^* \in \mathcal{S}_+^n$, we can compute $X^* = V^\top V$, $V = (v_1, \dots, v_n)$ using the Cholesky decomposition. From the constraint $D(X) = I$ it follows that $\|v_i\| = 1$, $i = 1, \dots, n$. Hence the relaxation step in Section 3.1 may be interpreted as associating with each primitive x_i a vector v_i on the unit sphere in a high-dimensional space. Accordingly, the matrix $(xx^\top)_{ij} = x_i x_j$ is replaced by the matrix $X_{ij} = v_i^\top v_j$.

Choosing a random vector r from the unit sphere, a combinatorial solution vector x is calculated from $X^* = V^\top V$ by setting $x_i = 1$ if $v_i^\top r \geq 0$ and $x_i = -1$ otherwise. This is done multiple times for different random vectors, letting the final solution x_{SDP} be the one that yields the minimum value for the objective function $x^\top Lx$. This technique may be interpreted as selecting different hyperplanes through the origin, identified by their normal r , that partition the vectors v_i , $i = 1, \dots, n$ in two sets.

Remark 4

Of course for $c \neq 0$, the solution x_{SDP} obtained by this technique does not need to be feasible for (5), as it is not required to satisfy the constraint $c^\top x = 0$. Thus it may yield an objective value $z_{SDP} = x_{SDP}^\top Lx_{SDP}$ that is even smaller than the optimal value of the semidefinite relaxation z_p . Therefore, some modifications of the randomized-hyperplane technique have been proposed in the literature [25, 70], which for the special case $c = e$ guarantee to find a feasible solution to the original problem (5) and even give a performance ratio for the objective value obtained. However, we stuck to the original randomized-hyperplane technique, as for the applications considered in this paper, it is not mandatory to find a feasible solution: We are only interested in solutions that should be quite balanced, but they do not need to be exactly balanced cuts of the associated graph. Hence, the constraint $c^\top x = 0$ may rather be seen as a strong bias to guide the search to a meaningful solution than as a strict requirement, especially in the case when no feasible combinatorial solution exists!

3.4 Performance Bounds

If $c = 0$ as in the examples of Sections 2.2 and 2.3, the combinatorial solution x_{SDP} obtained with the randomized-hyperplane technique is feasible. Based on the results from [31], the following suboptimality bound on the objective value z_{SDP} can be calculated in this case:

Theorem 3 *The expected value $E[z_{SDP}]$ of the objective function $z = x^\top Lx$ calculated with the randomized-hyperplane technique is bounded by*

$$E[z_{SDP}] \leq \alpha z_p + (1 - \alpha) \sum |L_{ij}| ,$$

where

$$\alpha = \min_{0 \leq \gamma \leq \pi} \frac{2}{\pi} \frac{\gamma}{1 - \cos \gamma} \approx 0.878 .$$

Proof. Using the notation of the previous section, we have

$$\begin{aligned} z_{SDP} &= \sum_{i,j} L_{ij} (x_{SDP})_i (x_{SDP})_j \\ &= \sum_{i,j} L_{ij} \operatorname{sgn}(v_i^\top r) \operatorname{sgn}(v_j^\top r) . \end{aligned}$$

Using Lemmas 3.2, 3.4 and 3.2.2 from [31], it follows

$$\begin{aligned} E[z_{SDP}] &= \sum_{i,j} L_{ij} E[\operatorname{sgn}(v_i^\top r) \operatorname{sgn}(v_j^\top r)] \\ &= \sum_{i,j} L_{ij} - 2 \sum_{i,j} L_{ij} \Pr[\operatorname{sgn}(v_i^\top r) \neq \operatorname{sgn}(v_j^\top r)] \\ &= \sum_{i,j} L_{ij} \left(1 - \frac{2}{\pi} \arccos(v_i^\top v_j)\right) \\ &\leq \sum_{i,j} L_{ij} - 2 \sum_{L_{ij} \leq 0} L_{ij} \left(1 - \frac{\alpha}{2} (1 + v_i^\top v_j)\right) - 2 \sum_{L_{ij} > 0} L_{ij} \left(\frac{\alpha}{2} (1 - v_i^\top v_j)\right) \\ &= \alpha \sum_{i,j} L_{ij} v_i^\top v_j + (1 - \alpha) \sum_{i,j} |L_{ij}| \\ &= \alpha z_p + (1 - \alpha) \sum_{i,j} |L_{ij}| . \quad \square \end{aligned}$$

A drawback of this bound is that it contains the problem-dependent constant $\sum_{i,j} |L_{ij}|$. This cannot be omitted, as L may contain negative entries.

Another bound which allows for L having negative entries was given by Nesterov [47], who extended the results of Goemans and Williamson [31] to maximization problems of the form (5) with $c = 0$. His results can easily be reformulated for the minimization problems considered in this paper, giving:

$$\frac{E[z_{SDP}] - z^*}{z_{\max}^* - z^*} \leq \frac{\pi}{2} - 1 \leq \frac{4}{7} ,$$

with z^* denoting the optimal value of (5) and $z_{\max}^*$ denoting the maximum value of the objective function subject to the integer constraint. However, this relative bound also depends on the problem instance, as it employs the difference between the maximum and minimum values of the objective function, which usually cannot be estimated in advance. Finally observe that the following relations between the different mentioned values of the objective function always hold true for $c = 0$:

$$z_d = z_p \leq z^* \leq z_{SDP} \leq z_{\max}^* .$$

It should be mentioned that the bounds presented above are not tight with respect to the much better performance measured in practice (cf. Section 4.1). However, for most alternative optimization approaches applicable to the general problem class considered here, performance bounds are lacking completely.

3.5 Relation to Spectral Relaxation

In this section we will compare the convex relaxation approach with spectral relaxation approaches. The results will show that convex relaxation always compares favorably with the computation of the so-called “Fiedler vector”, which is often used for the segmentation of graphs [22, 46].

First of all, we reformulate the semidefinite relaxation from Section 3.1 as an eigenvalue optimization problem. This idea dates back to Delorme and Poljak [17]. Starting from the dual problem formulation (13), we parameterize $y = \alpha e - v$, where $e^\top v = 0$. Then the constraint $L - y_0 cc^\top - D(y) = L - y_0 cc^\top + D(v) - \alpha I \in \mathcal{S}_+^n$ is equivalent to $\lambda_{\min}(L - y_0 cc^\top + D(v)) \geq \alpha$, which results in the following representation of (13):

$$z_d = \sup_{y_0, y} e^\top y = \sup_{y_0, \alpha} n\alpha = \sup_{y_0, e^\top v = 0} n\lambda_{\min}(L - y_0 cc^\top + D(v)) . \quad (15)$$

A spectral relaxation approach to (5) is based on the idea to keep the constraint $c^\top x = 0$ out of the Lagrangian formulation of the problem, and instead using it in the minimizing process. Hence we look at the problem:

$$z_{SR} := \min_{x \in \mathbb{R}^n} x^\top Lx , \quad c^\top x = 0, \quad x_i^2 = 1, \quad i = 1, \dots, n . \quad (16)$$

Substituting as above $y = \alpha e - v$, with $e^\top v = 0$, this results in the Lagrangian

$$\begin{aligned} x^\top (L - D(y))x + e^\top y &= x^\top (L - \alpha I + D(v))x + \alpha e^\top e - e^\top v \\ &= x^\top (L + D(v))x , \end{aligned}$$

as $x^\top x = n$ follows from $x_i^2 = 1$, $i = 1, \dots, n$. The corresponding minimax-problem yields the following formulation of the spectral relaxation bound:

$$\begin{aligned} z_{SR} &= \sup_{e^\top v = 0} \inf_{c^\top x = 0} x^\top (L + D(v))x \\ &= \sup_{e^\top v = 0} n \inf_{\|x\|=1, c^\top x = 0} x^\top (L + D(v))x \\ &= \sup_{e^\top v = 0} n\lambda_{\min}(V^\top (L + D(v))V) , \end{aligned} \quad (17)$$

where $V \in \mathbb{R}^{n \times (n-1)}$ contains an orthonormal basis of the complement $c^\perp$, i.e. $V^\top c = 0$, $V^\top V = I$.

For the special case of $c = e$, this bound was first provided by Boppana [8] and Rendl and Wolkowicz [53], independently. Poljak and Rendl [51] showed the equivalence of the spectral

relaxation (17) and the semidefinite relaxation (15) for this case by investigating the broader class of general graph bisection problems. This result can be extended to the general case $c \neq e$ considered here:

Theorem 4 *The semidefinite relaxation (13) yields the same lower bound on the objective function as the spectral relaxation (17):*

$$\begin{aligned} z_d &= z_{SR} \\ &= \sup_{e^\top v=0} n \lambda_{\min}(V^\top (L + D(v))V) . \end{aligned}$$

Notice that if (14) is not feasible, both values (15) and (17) become unbounded!

Proof. For each $y_0 \in \mathbb{R}, v \in \mathbb{R}^n$, the following holds:

$$\begin{aligned} \lambda_{\min}(V^\top (L + D(v))V) &= \inf_{\|x\|=1, c^\top x=0} x^\top (L + D(v))x \\ &= \inf_{\|x\|=1, c^\top x=0} x^\top (L + D(v))x - y_0(c^\top x)^2 \\ &\geq \inf_{\|x\|=1} x^\top (L + D(v))x - y_0(c^\top x)^2 \quad (18) \\ &= \inf_{\|x\|=1} x^\top (L - y_0 c c^\top + D(v))x \\ &= \lambda_{\min}(L - y_0 c c^\top + D(v)) . \end{aligned}$$

Building the supremums immediately yields $z_{SR} \geq z_d$. Now observe that if y_0 approaches negative infinity in (18), the infimum will only be finite if $c^\top x$ becomes 0:

$$\begin{aligned} \lim_{y_0 \rightarrow -\infty} \inf_{\|x\|=1} x^\top (L + D(v))x - y_0(c^\top x)^2 &= \inf_{\|x\|=1, c^\top x=0} x^\top (L + D(v))x \Rightarrow \\ \sup_{y_0} \inf_{\|x\|=1} x^\top (L + D(v))x - y_0(c^\top x)^2 &\geq \inf_{\|x\|=1, c^\top x=0} x^\top (L + D(v))x . \end{aligned}$$

Thus we also have equality for the last equation, and including the supremum over $e^\top v = 0$ gives $z_{SR} = z_d$. $\square$

We now want to derive a comparison of the semidefinite relaxation approach to the computation of the so-called ‘‘Fiedler vector’’. To this end, let's take a closer look at the following *weaker* spectral relaxation of (5):

$$z_{SR2} := \min_{x \in \mathbb{R}^n} x^\top Lx , \quad x^\top x = n, \quad c^\top x = 0 , \quad (19)$$

i.e. the constraint $x \in \{-1, +1\}^n$ is relaxed to $\|x\|^2 = n$ instead of relaxing it to $x_i^2 = 1, i = 1, \dots, n$.

The following Lemma holds:

Lemma 1 Let $V \in \mathbb{R}^{n \times (n-1)}$ denote the matrix defined in (17). Then

$$z_{SR2} = n\lambda_{\min}(V^\top LV) , \quad (20)$$

and the solution of (19) is given by

$$x^* = \sqrt{n}Vw_0 ,$$

where w_0 is the eigenvector corresponding to the smallest eigenvalue of $V^\top LV$ with the norm $\|w_0\| = 1$.

Proof. Let $P = \begin{pmatrix} \frac{1}{\|c\|}c, V \end{pmatrix}$ and $u = P^\top x = (0, w)^\top$, $w = V^\top x$. Then P is orthonormal, and $x^\top Lx = u^\top P^\top LPu = w^\top V^\top LVw$ attains its minimum for $w = \beta w_0$ being proportional to the smallest eigenvector w_0 , $\|w_0\| = 1$ of $V^\top LV$. Hence, $x^* = Pu = \beta Vw_0$, where β follows from the calculation: $n = (x^*)^\top x^* = \beta^2 w_0^\top V^\top Vw_0 = \beta^2$. The value for z_{SR2} follows immediately by inserting x^* into the objective function. $\square$

For the special case of $c = e$, this spectral relaxation corresponds to the computation of the “Fiedler vector”, i.e. the eigenvector x^* to the second smallest eigenvalue $\lambda_2(L)$ of the matrix L . This follows directly from (20) by observing that

$$\lambda_2(L) = \lambda_2(P^\top LP) = \lambda_2 \begin{pmatrix} \frac{1}{\sqrt{n}}e^\top L \frac{1}{\sqrt{n}}e & \frac{1}{\sqrt{n}}e^\top LV \\ V^\top L \frac{1}{\sqrt{n}}e & V^\top LV \end{pmatrix} = \lambda_2 \begin{pmatrix} 0 & 0 \\ 0 & V^\top LV \end{pmatrix} = \lambda_{\min}(V^\top LV)$$

for $c = e$ and P defined as in the proof of Lemma 1. The following fact shows the superiority of the convex relaxation approach; it follows immediately by comparing the result of Lemma 1 with (17):

Corollary 1 For $c = e$, the following inequality on the lower bounds for (5) is valid:

$$z_{SR2} \leq z_{SR} = z_d . \quad (21)$$

Apart from this fact of being less tight concerning the value of the objective function, the spectral relaxation with the Fiedler vector x^* has another disadvantage: To obtain the corresponding combinatorial solution x of (5), a threshold value t is used to set the entries $x_i = 1$ for $x_i^* > t$ and $x_i = -1$ otherwise. This raises the question for an appropriate choice of this threshold value. Two natural approaches seem to be promising: To set $t = 0$ (because of the original $+1/-1$ -constraint on x) or to set t equal to the median of x^* (to meet the balancing constraint $e^\top x = 0$). However, we will show below that an unsupervised choice of the threshold value may fail completely.

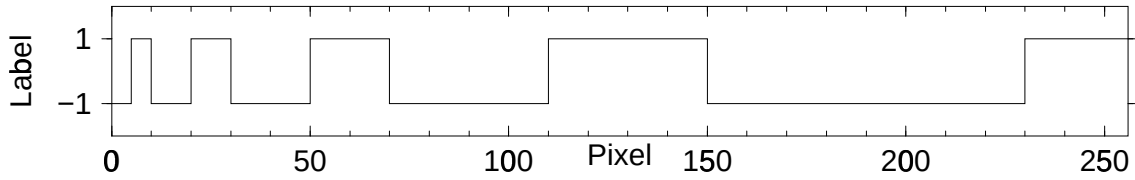


Figure 6: Signal x' comprising multiple spatial scales.

4 Experiments and Discussion

In this section, we investigate the performance of the convex relaxation approach experimentally. In Section 4.1, we start with reporting the statistical results for ground-truth experiments for restoration problems as described in Section 2.3, using noisy one-dimensional signals. The application of the convex relaxation approach to different real scenes from all problem types mentioned in Section 2 will be presented in Section 4.2. Furthermore, a brief discussion of different aspects of the semidefinite relaxation approach will be given in Section 4.3.

4.1 Ground-Truth Experiments

To be able to analyse the performance of the convex relaxation approach described in Section 3 statistically, ground truth data (the global optimum) has to be available for the problem under consideration. Therefore, we decided to investigate the restoration of noisy one-dimensional signals using the functional (10), as in this case the global optimum can be easily computed using dynamic programming.

For our experiments, we took the synthetic signal x' depicted in Figure 6 which involves transitions at multiple spatial scales, and superimposed Gaussian white noise with standard deviation $\sigma = 1.0$. Figure 7, top, shows an example of such a noisy signal. Then both the global optimum x^* of (10) and the solution x of the convex relaxation (14) were computed from this noisy input signal and compared to each other. A representative example of the restoration is given in Figure 7.

To derive some significant statistics, this experiment was repeated 1000 times for varying values of λ and different noisy signals. For each λ -value, we then calculated the following quantities:

$\overline{\Delta z}$: The sample mean of the gap $\Delta z = z - z^*$ (measured in % of the optimum) with respect to the objective function values of the suboptimal solution $z(x)$ and the optimal solution $z^* = z(x^*)$.

$\sigma_{\Delta z}$: The sample standard deviation of the gap Δz .

$\overline{\Delta z'}$: The sample mean of the gap $\Delta z' = z - z'$ (measured in % of the optimum) with respect to the objective function values of the suboptimal solution $z(x)$ and the synthetic signal $z(x')$.

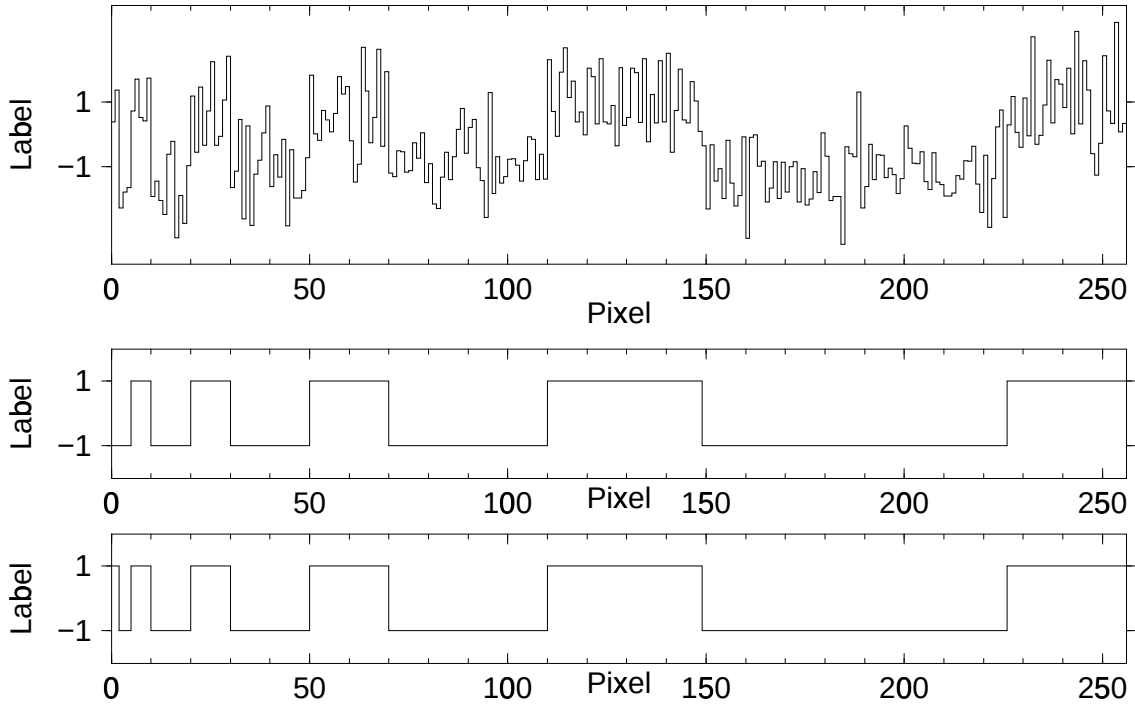


Figure 7: A representative example illustrating the statistics shown in Fig. 8. **Top:** Noisy input signal. **Middle:** Optimal solution x^* . **Bottom:** Suboptimal solution x .

$\sigma_{\Delta z'}$: The sample standard deviation of the gap $\Delta z'$.

Moreover, we also calculated the sample mean of the number of misclassified pixels in comparison with the optimal solution.

The results are shown in Figure 8. They reveal the accuracy of the suboptimal solutions obtained with the semidefinite relaxation: The average relative error for both the objective function value and the number of correctly classified pixels is below 2%, compared to the optimum values of the objective function. The corresponding measures for $\sigma_{\Delta z}$ lie between 0.12% and 1.33%. This shows that in practice, the performance of the semidefinite relaxation approach is much better than the bounds presented in Section 3.4.

Concerning the restoration of the original signal x' , it should be mentioned that the signal x' is not the best solution of the functional (10), which results in $\overline{\Delta z} < \overline{\Delta z'}$. This indicates that more appropriate criteria should be used for the restoration of signals, e.g. by incorporating suitable priors with respect to x' (cf. [9]). The derivation of such criteria is not the objective of this paper. However, the performance is still remarkably good (cf. Figure 8): For values of the scale parameter $\lambda > 1.5$, the average relative error for the objective function value $\overline{\Delta z'}$ is below 3%, with the corresponding measures for $\sigma_{\Delta z'}$ lying between 1.95% and 2.69%. The high error rates for $\lambda < 1.5$ can be explained with the dominating larger spatial scales in the signal x' .

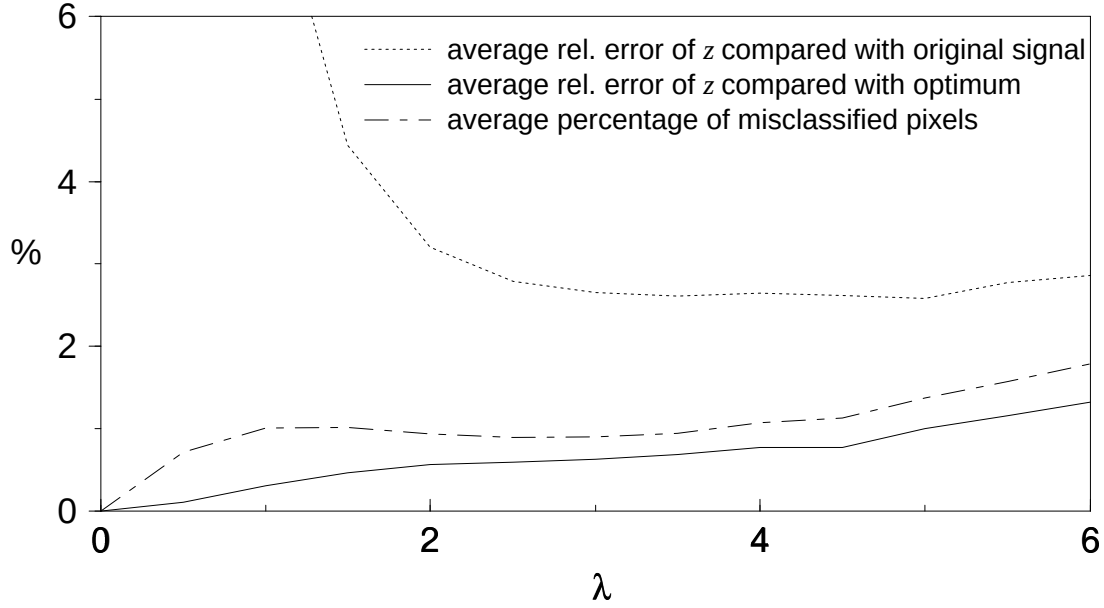


Figure 8: Average relative errors $\overline{\Delta z}$ and $\overline{\Delta z'}$ of the objective function for the suboptimal solution x in comparison to the optimal signal x^* and the synthetic signal x' , respectively, for different values of the scale parameter λ . Also shown is the average percentage of misclassified pixels for the suboptimal solution x compared to the optimal solution x^* .

4.2 Real Scenes

In this section we present the results of the semidefinite relaxation approach by applying it to problems from the different fields presented in Section 2. For the unsupervised partitioning examples, we also compare the results with the segmentation obtained by thresholding the Fiedler vector.

4.2.1 Preliminary Remark (Similarity Measures)

Recall that the objective in unsupervised partitioning is to split a graph with some extracted image features as vertices into two coherent groups. To this end, the edge weights $w(i, j)$ building the similarity matrix are computed from distances $d(i, j)$ between the extracted image features i and j as

$$w(i, j) = e^{-\frac{(d(i, j))^2}{\sigma}}, \quad (22)$$

where $d(i, j)$ and σ are chosen application dependent. We studied two different methods to calculate the similarity measures:

- (i) Compute $w(i, j)$ for all feature pairs (i, j) directly, thus not including any spatial information.

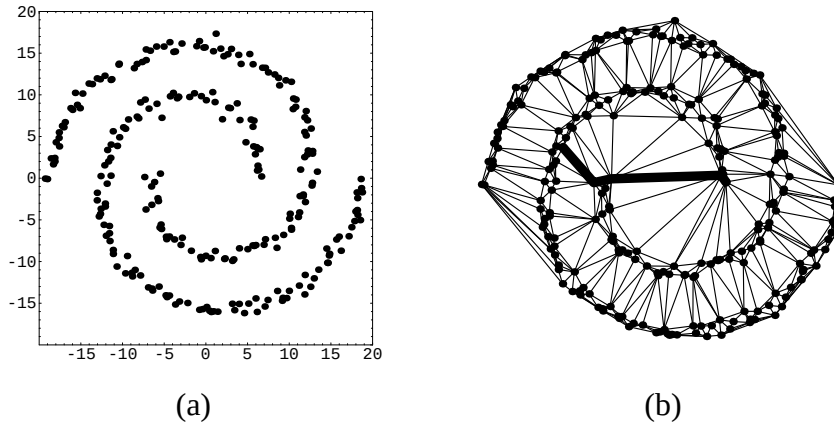


Figure 9: **(a)** A skewed data distribution with two spiral-shaped groups. **(b)** The shortest Euclidean path within the Delaunay-graph between two points of the same group.

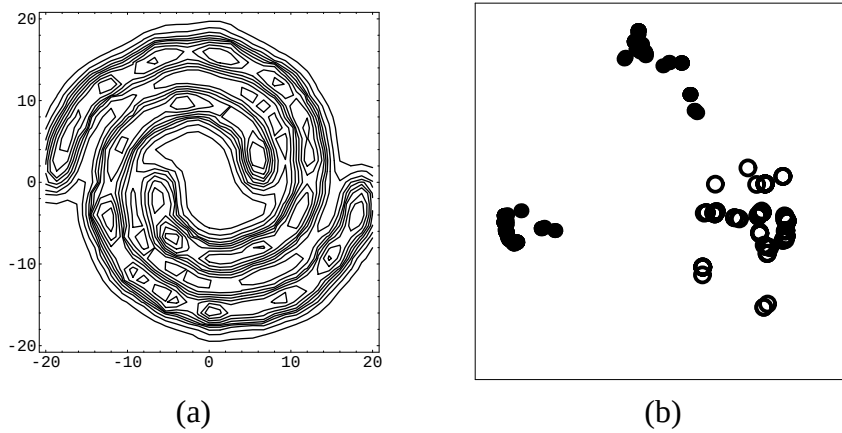


Figure 10: **(a)** Level lines of a Parzen estimate of the data distribution. **(b)** Application of a metric scaling technique [15] yields a Euclidean approximation in 2D-space of the weighted-path distances (many points project to almost equal locations): As compared to Figure 9(a), the two spiral-shaped groups form more compact clusters.

- (ii) Compute $w(i, j)$ only for neighboring features, and derive the other edge weights by calculation of a path connecting them. This is done by computing the shortest paths for the graph derived from G by changing the similarity weights to dissimilarities and transforming them back afterwards. This results in a similarity measure which favors spatially coherent structures.

To motivate the latter method (ii), consider Figure 9(a), which shows a set of points with the shape of two spirals. It was critically observed in [27] that spectral methods fail to partition such “skewed” coherent groups. Indeed, Figure 9(b) shows that in the corresponding Delaunay-graph, the shortest *Euclidean* path between two points of the same group traverses the other group. As a consequence, a direct pairwise comparison of Euclidean distances and, more general, method (i) above which ignores spatial context, are not appropriate.

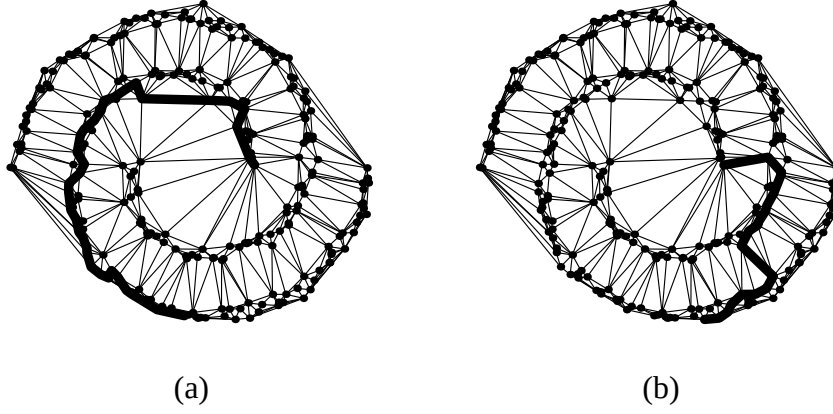


Figure 11: Weighted shortest paths using the density from Figure 10(a). **(a)** As compared to the Euclidean distance, within-group paths have become shorter. **(b)** The partition problem is still not trivial since shortest paths between points of the same group do not always lie within this group.

Method (ii) provides a simple remedy in this situation (cf. also [23]). Apart from “location”, additional attributes like color, texture, etc. are typically used to define pairwise distances/similarities as edge-weights. As a result, by using weighted paths as a distance measure, spatial coherency is exploited in the appropriate way and results in shorter paths *within* a group. For the example shown in Figure 9(a), we simulated such an additional attribute by a Parzen estimate [26] of the spatial data distribution (Figure 10(a)). Figure 10(b) visualizes the resulting distances by approximating them with Euclidean distances within 2D-space using a classical metric scaling technique [15]. This result shows that points of a coherent group have become more similar to each other. Accordingly, the partition task has become more well-defined but not trivial, of course: Weighted paths within groups have been shortened (Figure 11(a)), but the shortest paths between two points of the same group do still not always lie within this group (Figure 11(b)).

As we are mainly interested here in the results of the semidefinite relaxation approach from an optimization point of view, we did not work on more elaborate computations of the $w(i, j)$ -values. For a survey of numerous other (dis)similarity measures, see [52].

4.2.2 Unsupervised Partitioning

Point sets. Figure 12 shows the partitions computed with convex (Figure 12(a)) and spectral (Figure 12(b)) relaxation, respectively, for the example given in the previous section (Figure 9(a)). As both spirals have the same number of points, we used $c = e$ for this experiment, with each point defining a vertex in the corresponding graph. Although the similarity weights $w(i, j)$ are calculated using the path-metric from method (ii), spectral relaxation with the Fiedler vector still fails to compute the correct cut whereas convex relaxation does. This reflects the theoretical results of Section 3.5, showing the superiority of the convex relaxation

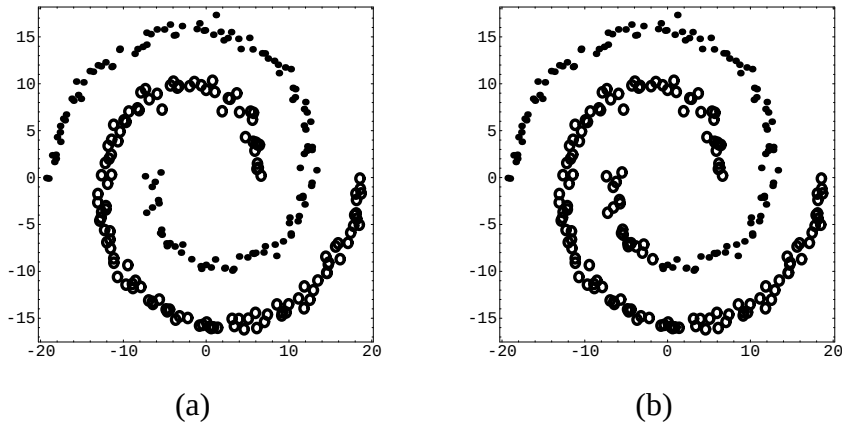


Figure 12: Point set clustering for Figure 9(a). **(a)** Using distances based on weighted paths, convex relaxation computed the correct partition. **(b)** Partition computed with the Fiedler vector thresholded at 0. Correct separation is still not possible due to the less tight relaxation of the underlying combinatorial problem.

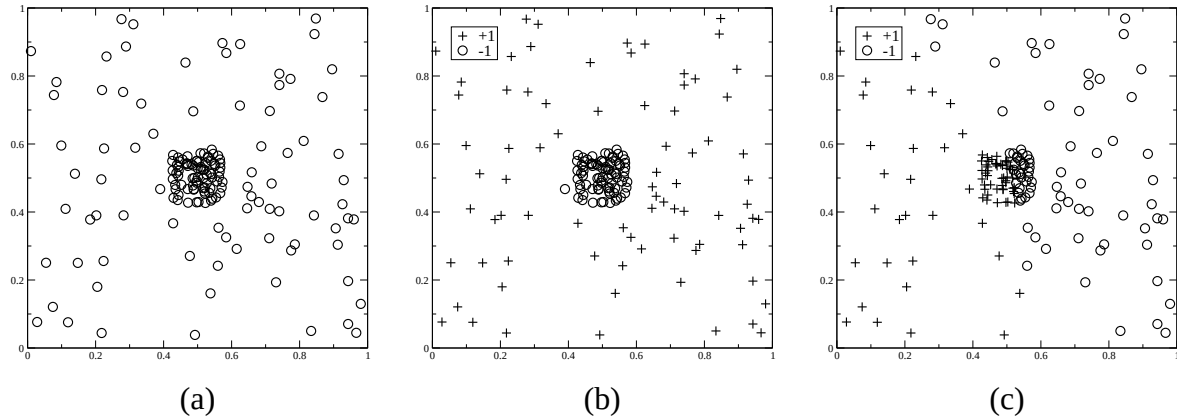


Figure 13: Point set clustering. **(a)** Input data with 160 points, weights calculated with $\sigma = 0.0025$ in (22). **(b)** Solution computed with convex optimization, using $c = e$. **(c)** Solution computed with the Fiedler vector, thresholded at the median value: Spectral relaxation fails!

approach from an optimization point of view. This disadvantage of spectral relaxation is particularly relevant in the *unsupervised* case, as it is not known beforehand which heuristic for thresholding the eigenvector might yield the desired result. On the other hand, note that convex relaxation works *without any* threshold.

Another situation is depicted in Figure 13(a). This point set comprises a dense cluster in the middle and equally distributed background clutter, both containing 80 points. The similarity weights $w(i, j)$ are now computed directly from the Euclidian distances $d(i, j)$ between all points (method (i)). The results for this example (Figure 13) again show the superiority of the convex optimization approach: While it successfully separates the dense cluster from the background, the spectral relaxation only achieves an unsatisfactory partition. This is due to the fact that the Fiedler vector does not give a clear cut value (cf. Figure 14). Also notice that

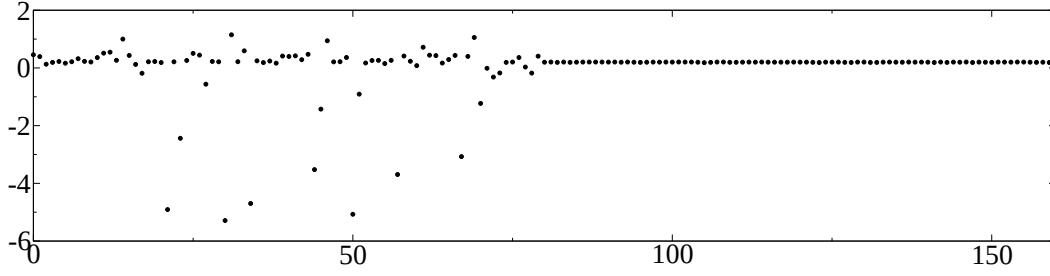


Figure 14: The Fiedler vector for the example in Figure 13.

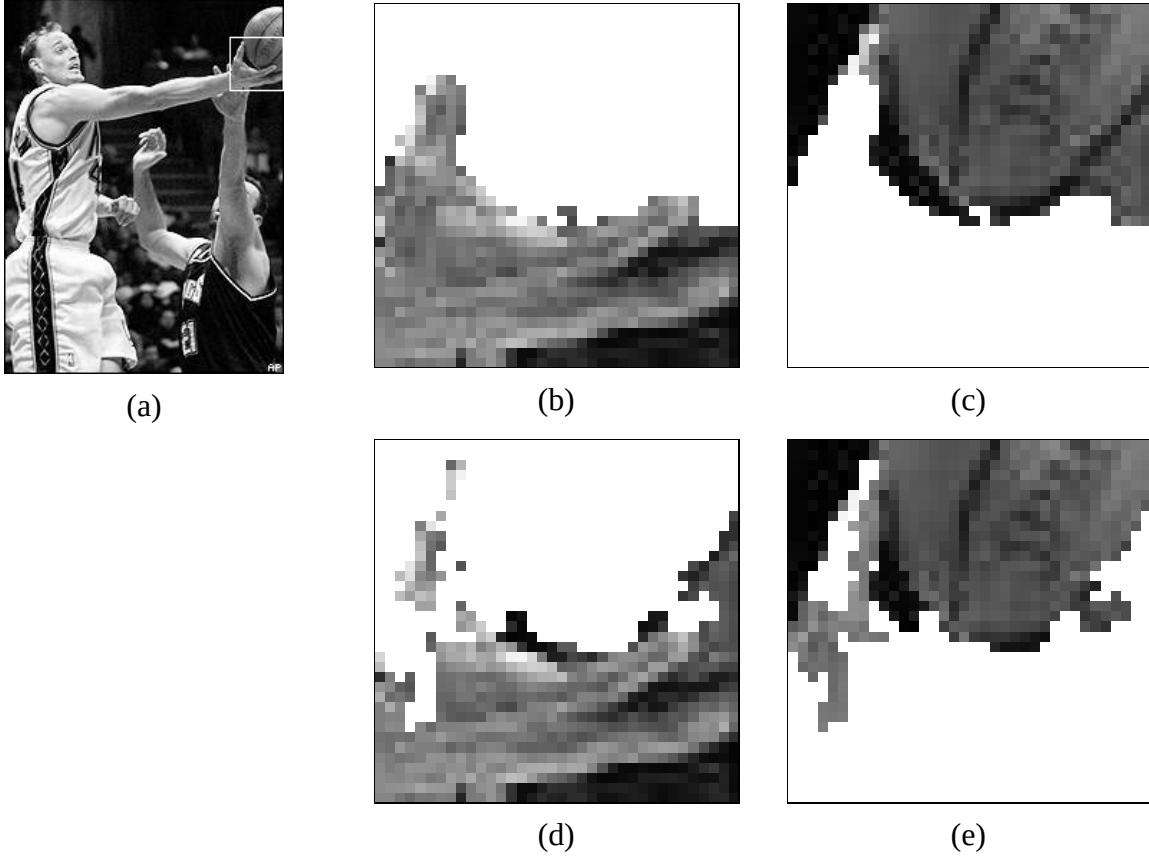


Figure 15: Grayscale image partitioning. **(a)** Input image (36×36 pixels) as part of a larger image ($\sigma = 8 \cdot 255^2$). **(b),(c)** Segmentation computed with convex optimization: The hand is clearly separated from the ball. **(d),(e)** Solution computed with the Fiedler vector: No clear separation is obtained by median thresholding. Thresholding at 0 just separates one pixel from the rest of the image.

the randomized hyperplane technique does not give an exactly balanced cut: The two parts contain 78 and 82 points, respectively. But as this corresponds to the visual impression, this is no drawback.

Grayscale images. To study the partitioning of grayscale images, we created the neigh-

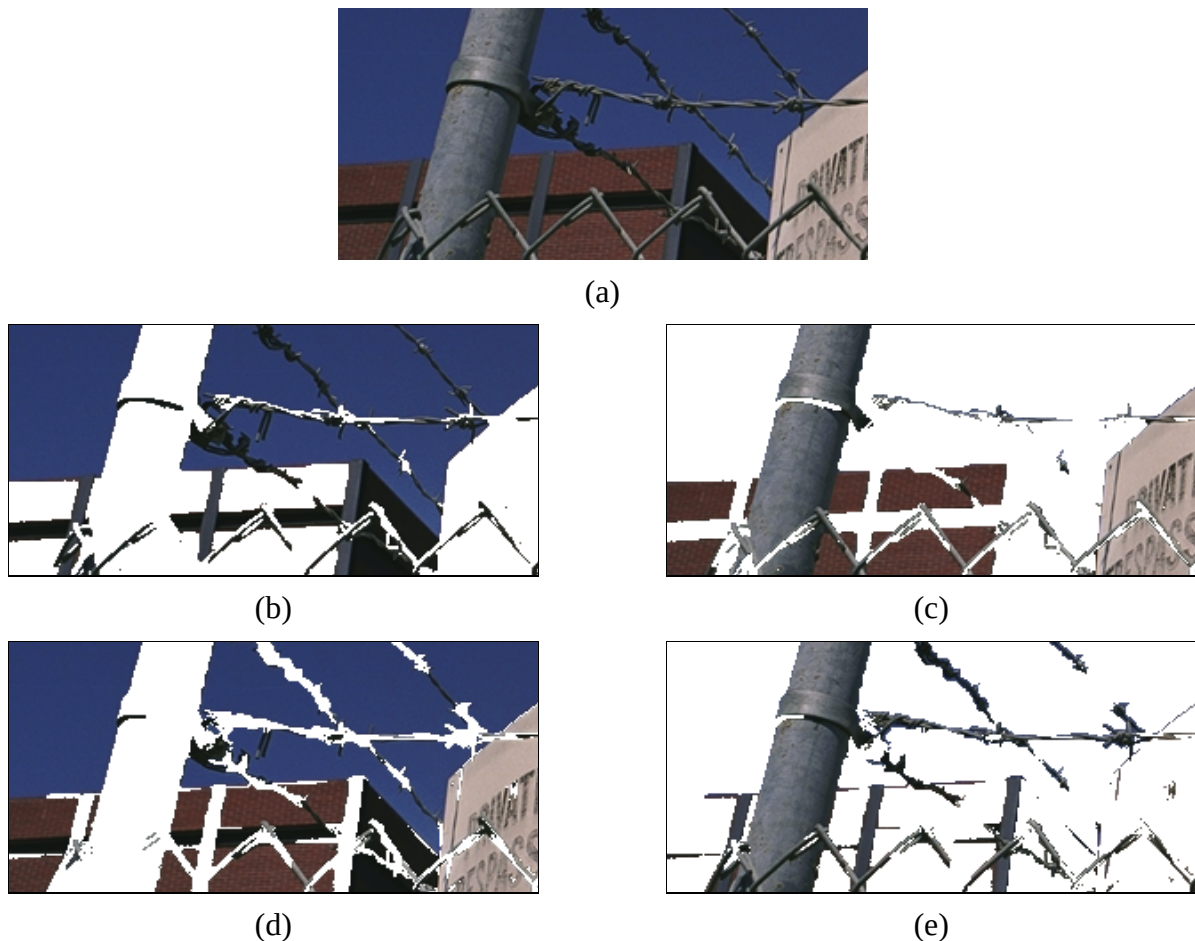


Figure 16: Color image partitioning, using method (i). **(a)** Input image (298×141 pixels), yielding 209 clusters ($\sigma = 8 \cdot 255^2$). **(b),(c)** Segmentation computed with convex optimization: Similar colors are grouped together. **(d),(e)** Segmentation computed with the Fiedler vector thresholded at the median. Thresholding at 0 just separates one pixel from the rest of the image.

neighborhood graph G which contains a vertex for each pixel, and calculated the similarity weights $w(i, j)$ for adjacent pixels as in (22), with $d(i, j)$ denoting the gray-value difference of pixels i and j . The remaining similarity weights were computed using method (ii) to favor spatially coherent structures. Applying the convex relaxation for $c = e$ then yields a segmentation of the image into two parts of nearly the same size. The result for one sample image is shown in Figure 15. Here, the Fiedler vector only yields an unsatisfactory segmentation, whereas the convex relaxation approach clearly separates the hand from the ball, giving two groups containing 647 and 649 pixels, respectively.

Color images. For larger images, we first computed an oversegmentation by applying the mean shift technique [13] at a fine spatial scale in order not to destroy any perceptually significant structure. Instead of thousands of pixels, the graph vertices are then formed by the obtained clusters, and $d(i, j)$ is computed as the color difference of two clusters in the

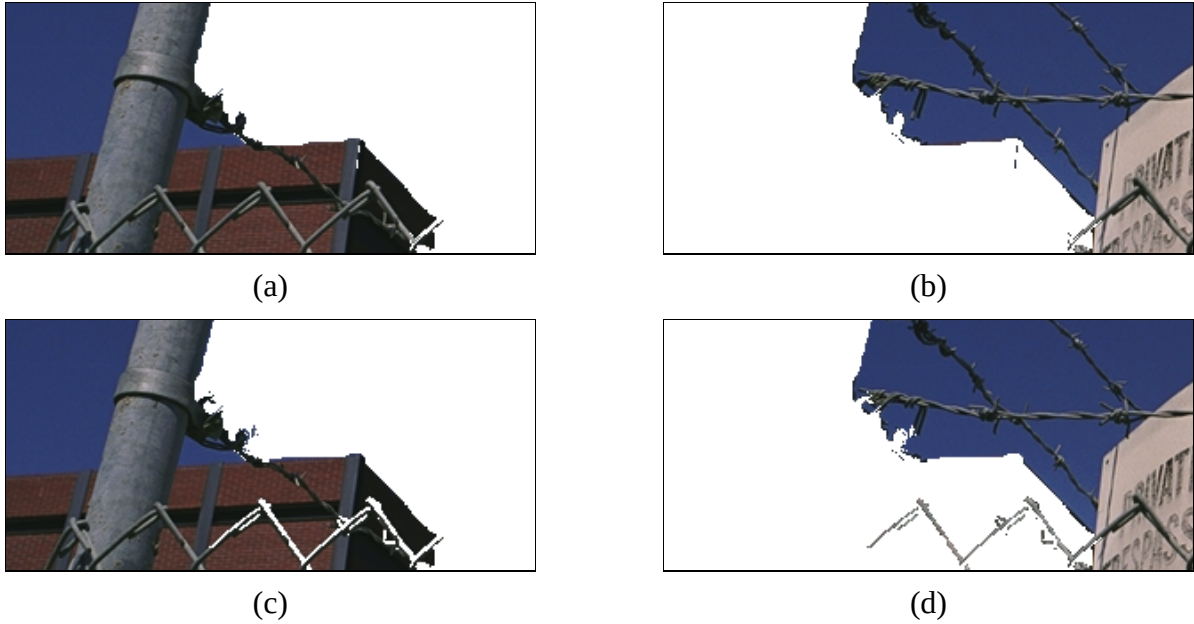


Figure 17: Color image partitioning for the image from Figure 16(a), using method (ii). **(a),(b)** Segmentation computed with convex optimization: Spatially coherent structures are favored. **(c),(d)** Segmentation computed with the Fiedler vector thresholded at the median: The requirement that both parts have the same size influences the result negatively.

perceptually uniform LUV space. We applied both methods (i) and (ii) for the color image shown in Figure 16(a). The results approve the wide range of applicability and the success of the semidefinite relaxation approach: Whereas the similarity measure from method (i) groups together pixels of similar colors (see Figure 16(b),(c)), method (ii) yields a segmentation into two reasonable, spatially coherent parts (see Figure 17(a),(b)). For this example, the results obtained by thresholding the Fiedler vector at its median value are also quite reasonable (see Figures 16(d),(e) and 16(c),(d)), but a crucial point is that the threshold value does not emerge naturally: Looking at the Fiedler vector in more detail shows that basically one cluster is separated from the rest of the image, as only one entry has a large positive value while most of the others contain small negative values. Notice once more that on the other hand, no choice of a threshold value is necessary for the semidefinite relaxation approach!

Choice of c . So far, the size of the clusters had no influence on the similarity weights. This may yield unsatisfactory separation results. Figure 18 gives an example: Here the sky accounts for nearly half of the image (approx. 44%), but the oversegmentation puts all of its pixels into one cluster. As for $c = e$, all clusters are of the same importance no matter how large they are, the semidefinite relaxation approach segments the image into two parts by cutting the city, which contains many small clusters (see Figure 18(b),(c)). To derive a segmentation which takes into account the different sizes of the clusters, the balancing constraint can be changed in the following way: Calculate the number of pixels m_i contained in each cluster i and set $c = m$ instead of $c = e$. Thus we now search for a segmentation which parti-



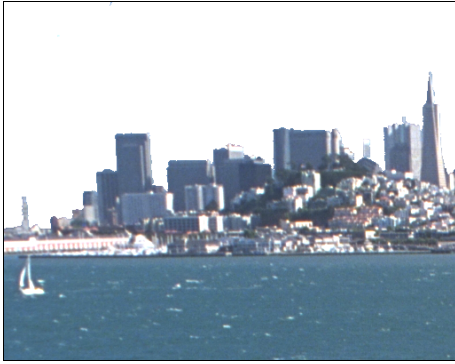
(a)



(b)



(c)



(d)



(e)

Figure 18: Color image partitioning. **(a)** Input image (512×404 pixels), yielding 408 clusters ($\sigma = 8 \cdot 255^2$). **(b),(c)** Segmentation computed with convex optimization, with $c = e$: The image is cut in two coherent parts. **(d),(e)** Segmentation computed with convex optimization, with c_i equal to the number of pixels in cluster i : The largest cluster is separated from the rest of the image.

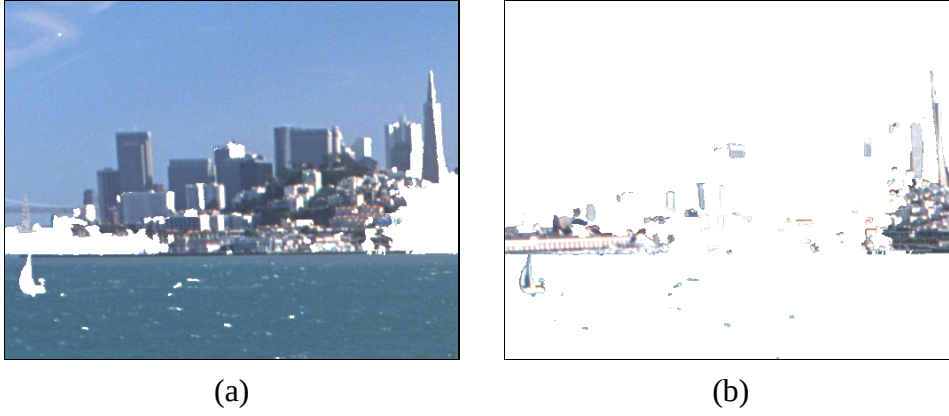


Figure 19: Segmentation computed with the Fiedler vector thresholded at the median for the image from Figure 18(a): Separation into spatially coherent parts fails completely.

tions the image in two coherent parts with each containing approximately the same number of pixels instead of the same number of clusters. The result shown in Figure 18(d),(e) approves the validity of this approach: Now the sky is separated from the rest of the image, giving a segmentation in accordance to our balancing constraint. Notice that for this example, the Fiedler vector fails completely to give a meaningful separation (see Figure 19): Thresholding at the median value yields no coherent segmentation, whereas thresholding at 0 only separated 3 clusters from the city from the rest of the image.

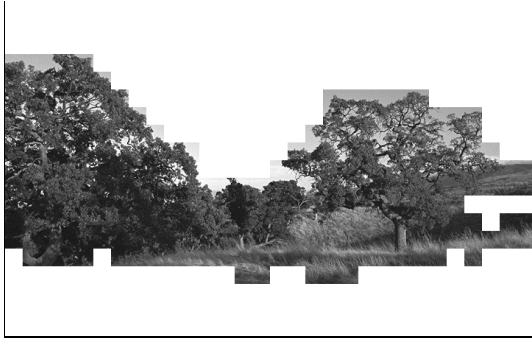
Texture. The final experiment for binary partitioning deals with grayscale images comprising some natural textures. An example is shown in Figure 20(a). To derive a texture measure for this image, we subdivided it into 24×24 pixel windows, and calculated local histograms for two texture features within these windows. Each window then corresponds to a graph vertex, and $d(i, j)$ is computed as the χ^2 -distance of the histograms for all window pairs (i, j) , thus using method (i). Considering the simplicity of this texture measure, the segmentation result obtained in this way is excellent (see Figure 20(b),(c)). In order to yield a satisfactory result, the Fiedler vector has to be thresholded at 0 for this example. The median threshold does not make sense here, as the image does not contain two parts of the same size. Again we note that, in the unsupervised case, this is *not* known beforehand.

4.2.3 Perceptual Grouping and Figure-Ground Discrimination

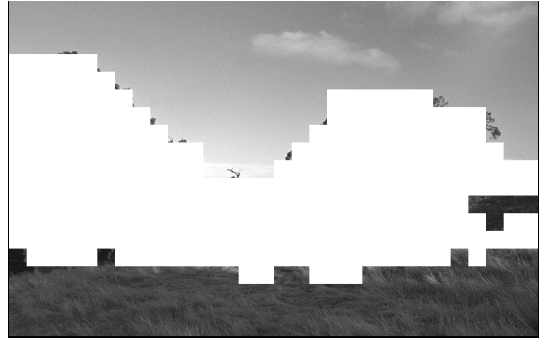
Figure 21 depicts the result computed by minimizing (9) using the convex relaxation approach. As input data we used a line-finder developed by the group of Prof. Förstner [21]. Figure 21(b) shows a few hundred line-fragments computed for the scene depicted in (a). As similarity measure $w(i, j)$ between two primitives (line-fragments) i and j we used the relative angle between these primitives. Correspondingly, as the graph of w shows in Figure 21(c), two lines are similar if their relative angle is close to a multiple of $\pi/2$. We refer to [34] for more elaborate similarity measures which, however, are not essential for testing our approach from the optimization point of view.



(a)



(b)



(c)



(d)



(e)

Figure 20: Grayscale-texture partitioning. **(a)** Input image (720×456 pixels), yielding 570 vertices of 24×24 -pixel windows ($\sigma = 1$), similarity weights computed with method (i). **(b),(c)** Segmentation computed with convex optimization. **(d),(e)** Segmentation computed with the Fiedler vector, using the threshold value 0.

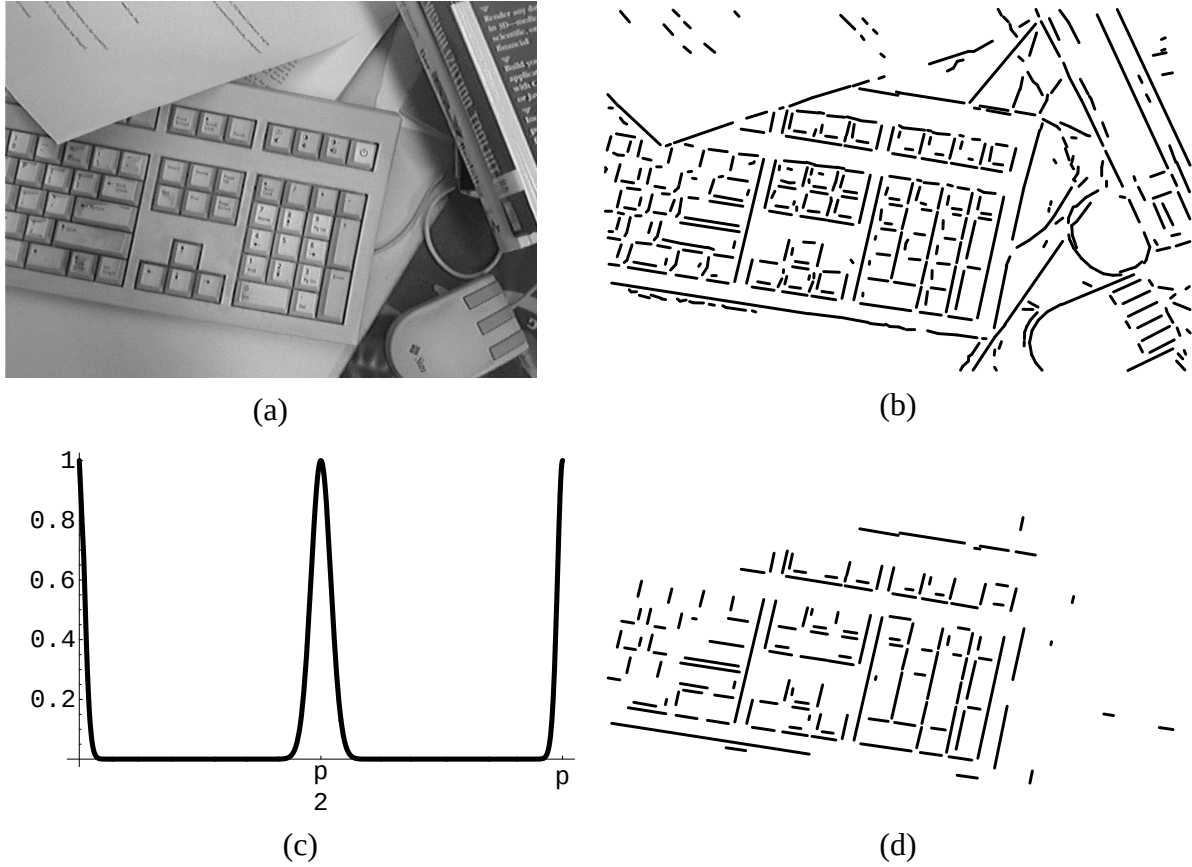


Figure 21: Perceptual grouping. **(a)** Section of an office table shown from the top. **(b)** Output of a line finder. **(c)** The similarity measure $w(i, j)$ as a function of the angle between two line fragments. Two fragments are most similar if they are (nearly) orthogonal to each other. Several coherent groups of different cardinality exist in (b). **(d)** The minimizer of (9) determines the “most coherent” group of line fragments and suppresses the other groups.

We note that *several* groups exist in Figure 21(b) which are coherent according to the similarity measure w . The minimizer of (9) shown in Figure 21(d) determines the keyboard as the “most coherent” group, as expected from a visual inspection of the scene.

4.2.4 Restoration

In Section 4.1, we already presented the results of the convex relaxation approach with respect to the restoration of noisy one-dimensional signals. The result concerning the restoration of a two-dimensional real image is shown in Figure 22. Considering that the desired object to be restored comprises structures at both large and small spatial scales, the restoration result is fairly good.

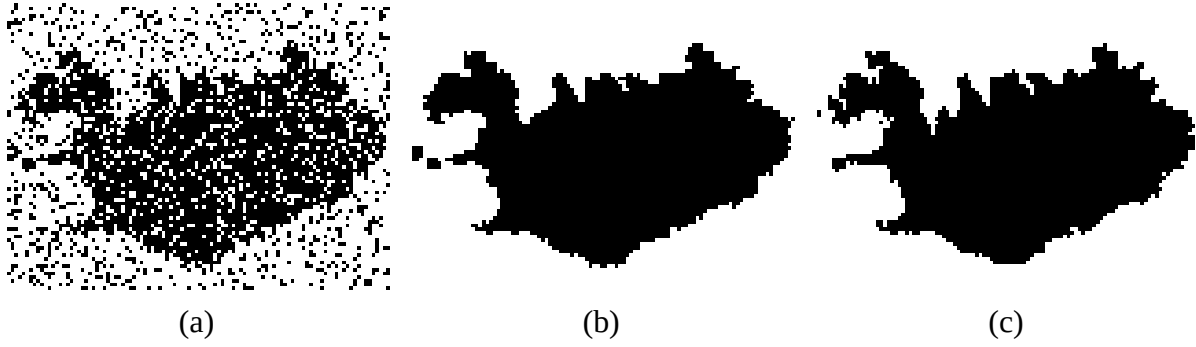


Figure 22: Restoration. **(a)** Binary noisy original image (map of Iceland). **(b)** Suboptimal solution computed by convex optimization with $\lambda = 2.0$. **(c)** Original before adding noise.

4.3 Discussion

Combinatorial optimization. The experimental results demonstrate that our approach is a versatile tool for solving a broad range of difficult combinatorial problems in a convenient way. For example, the user can focus on how to choose a constraint vector appropriate for the application he is interested in, rather than to worry about heuristics and technical details in order to avoid bad local minima. Across several different application fields we obtained meaningful solutions according to the criterion which was optimized. The optimization problem (9) suggested by Herault and Horaud [34], for instance, is a useful saliency measure for perceptual grouping but has gained only a mixed reputation in the literature [63] due to the combinatorial complexity involved. Our approach considerably mitigates this latter problem.

Convex vs. spectral relaxation. We have shown in Section 3.5 that convex relaxation always provides a tighter relaxation of the underlying combinatorial problem in comparison to spectral relaxation. In fact, this improvement often can be made explicit by representing the convex optimum as solution to an eigenvalue optimization problem in terms of the multiplier variables.⁵ In many cases this theoretical result also turned out to be significant in practice: Spectral relaxation may yield a solution which doesn't make sense in respect of the chosen optimization criterion. Furthermore, spectral relaxation has a decisive disadvantage in the case of *unsupervised* classification. The suitability of the criterion for thresholding the eigenvector often depends on the particular data being processed.

Again we note (cf. Section 2.1) that alternative improvements of spectral relaxation techniques exist [58, 62]. Concerning the normalized cut criterion (3), we suppose that the choice of a different constraint vector improves the classical spectral approach of Fiedler. In this respect, we have considerably generalized the approach by taking into account arbitrary constraint vectors. The effect of normalizing the Laplacian matrix in (3) on the tightness of the relaxation and the thresholding problem are difficult to analyze and left as an open problem. For very recent work based on the normalized cut criterion we refer to [45].

⁵From the computational viewpoint, this representation is less convenient than the underlying convex optimization problem, however.

Image	n	time (in sec.)
1D signal (Fig. 7)	256	3
spiral (Fig. 12)	256	30
point set (Fig. 13)	160	10
hand/ball (Fig. 15)	1296	7745
color image (Fig. 16), method (i)	209	12
color image (Fig. 17), method (ii)	209	12
color image (Fig. 18(b),(c)), $c = e$	408	142
color image (Fig. 18(d),(e)), $c = m$	408	165
texture image (Fig. 20)	570	569
keyboard (Fig. 21)	466	184
iceland (Fig. 22)	8113	64885

Table 1: Sizes and computation times for the different problems considered in this paper.

Computational complexity. The price for the convenient properties of our optimization approach listed in Section 1.1 is the squared number of variables of the semidefinite relaxation. Although the approach of Benson et al. [4] is able to exploit a sparse problem structure very well, the computation time quickly grows with the number of variables such that problems with ten thousands of variables cannot be solved. While this is not a problem for the perceptual grouping of a couple of hundred primitives, it prevents at present the application to large-scale problems like, for instance, combinatorial image restoration (cf. Table 1).

An interesting theoretical result in this context concerns bounds which have been derived for the maximal rank r of a matrix X solving a semidefinite program [3, 49]. Applying this result to the program (14), we obtain the bound

$$\frac{1}{2}r(r+1) \leq n+1,$$

hence $r < \sqrt{2n}$ for large n . This means that *in principle* the large number of n^2 problem variables can be reduced by setting $n-r$ rows of the matrix V in the decomposition $X = V^\top V$ to zero (cf. Section 3.3). The future will show whether algorithms will come up which exploit this property along with a considerable saving of memory and speed-up of computation.

5 Conclusion and Further Work

We worked out a semidefinite programming framework applicable to a broad class of binary combinatorial optimization problems in computer vision. In our opinion, the major contribution of this work is to put into perspective a fairly general and novel optimization technique as an attractive alternative to established techniques, due to sound underlying mathematical principles and the absence of tuning parameters.

So far, the focus of our work was primarily on mathematical optimization: Convex relaxation, existence of feasible constraints/solutions, and performance bounds. Although we demonstrated the applicability of the approach for three different non-trivial problems, we did not spend much time on working out tailor-made similarity measures for specific applications. Accordingly, in future work, our focus will shift to the related and general problem of learning suitable metrics for classification. Furthermore, we will investigate the case of non-binary classification and more intricate constraints such as those of relational graph matching.

Acknowledgement

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG; grant Schn457/3-1).

References

- [1] E. Aarts and J.K. Lenstra, editors. *Local Search in Combinatorial Optimization*, Chichester, 1997. Wiley & Sons.
- [2] A. Amir and M. Lindenbaum. A generic grouping algorithm and its quantitative analysis. *IEEE Trans. Patt. Anal. Mach. Intell.*, 20(2):168–185, 1998.
- [3] A. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discr. Comput. Geom.*, 13(2):189–202, 1995.
- [4] S.J. Benson, Y. Ye, and X. Zhang. Mixed linear and semidefinite programming for combinatorial and quadratic optimization. *Optimiz. Methods and Software*, 11&12:515–544, 1999.
- [5] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proc. IEEE*, 76:869–889, 1988.
- [6] J.E. Besag. On the analysis of dirty pictures (with discussion). *J. R. Statist. Soc. B*, 48:259–302, 1986.
- [7] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
- [8] R. B. Boppana. Eigenvalues and graph bisection: an average-case analysis. In *Proceedings of the 28th Annual IEEE Symposium on Foundations of Computer Science*, pages 280–285. IEEE Computer Society Press, 1987.
- [9] C.F. Borges. On the estimation of markov random field parameters. *IEEE Trans. Patt. Anal. Mach. Intell.*, 21(3):216–224, 1999.

- [10] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog. (CVPR'98)*, pages 648–655, Santa Barbara, California, 1998.
- [11] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Patt. Anal. Mach. Intell.*, 23(11):1222–1239, 2001.
- [12] P.B. Chou and C.M. Brown. Multimodal reconstruction and segmentation with markov random fields and hcf optimization. In *Proc. DARPA Image Underst. Workshop*, pages 214–221, Cambridge, Massachussetts, April 6-8 1988.
- [13] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *7th Int. Conf. on Comp. Vision (ICCV'99)*, pages 1197–1203, 1999.
- [14] T.M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electronic Computers*, 14:326–334, 1965.
- [15] T.F. Cox and M.A. Cox. *Multidimensional scaling*. Chapman & Hall, London, 1994.
- [16] C. Delorme and S. Poljak. Combinatorial properties and the complexity of a max-cut approximation. *European Journal of Combinatorics*, 14:313–333, 1993.
- [17] C. Delorme and S. Poljak. Laplacian eigenvalues and the maximum cut problem. *Mathematical Programming*, 62:557–574, 1993.
- [18] L. Devroye. Automatic pattern recognition: A study of the probability of error. *IEEE Trans. Patt. Anal. Mach. Intell.*, 10(4):530–543, 1988.
- [19] R. Diestel. *Graph Theory*. Springer, 1996.
- [20] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 1973.
- [21] Fex-homepage. <http://www.ipb.uni-bonn.de/ipb/projects/fex/fex.html>, Inst. of Photogrammetry, University of Bonn, Germany.
- [22] M. Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25(2):619–633, 1975.
- [23] B. Fischer, T. Zöllner, and J.M. Buhmann. Path based pairwise data clustering with application to texture segmentation. In M. Figueiredo, J. Zerubia, and A.K. Jain, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2134 of *Lect. Not. Comp. Sci.*, pages 235–250, Berlin, 2001. Springer.

- [24] D.A. Forsythe, J. Haddon, and S. Ioffe. The joy of sampling. *Int. J. of Comp. Vision*, 41(1/2):109–134, 2001.
- [25] A. Frieze and M. Jerrum. Improved approximation algorithms for max k-cut and max bisection. *Algorithmica*, 18:67–81, 1997.
- [26] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.
- [27] Y. Gdalyahu, D. Weinshall, and M. Werman. Self-organization in vision: Stochastic clustering for image segmentation, perceptual grouping, and image database organization. *IEEE Trans. Patt. Anal. Mach. Intell.*, 23(10):1053–1074, 2001.
- [28] D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrf’s: Surface reconstruction. *IEEE Trans. Patt. Anal. Mach. Intell.*, 13(5):401–412, 1991.
- [29] D. Geman. Random fields and inverse problems in imaging. In P.L. Hennequin, editor, *École d’Été de Probabilités de Saint-Flour XVIII – 1988*, volume 1427 of *Lect. Notes in Math.*, pages 113–193, Berlin, 1990. Springer Verlag.
- [30] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 6(6):721–741, 1984.
- [31] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42:1115–1145, 1995.
- [32] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18(4):377–388, 1996.
- [33] F. Heitz, P. Perez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *Comp. Vis. Graph. Image Proc.: IU*, 59(1):125–134, 1994.
- [34] L. Herault and R. Horaud. Figure-ground discrimination: A combinatorial optimization approach. *IEEE Trans. Patt. Anal. Mach. Intell.*, 15(9):899–914, 1993.
- [35] T. Hofmann and J. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(1):1–14, 1997.
- [36] H. Ishikawa. *Global Optimization Using Embedded Graphs*. PhD thesis, Dept. Comp. Science, Courant Inst. Math. Sciences, New York University, 2000.
- [37] D.W. Jacobs. Robust and efficient detection of salient convex groups. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18(1):23–37, 1996.

- [38] D.R. Karger and C. Stein. A new approach to the minimum cut problem. *J. of the ACM*, 43(4):601–640, 1996.
- [39] L. Kaufman and P.J. Rousseeuw. *Finding Groups in Data*. John Wiley & Sons, New York, 1990.
- [40] D.C. Knill and W. Richards, editors. *Perception as Bayesian Inference*. Cambridge Univ. Press, 1996.
- [41] M. Laurent and S. Poljak. On a positive semidefinite relaxation of the cut polytope. *Linear Algebra and its Applications*, 223/224(1–3):439–461, 1995.
- [42] Y.G. Leclerc. Constructing simple stable descriptions for image partitioning. *Int. J. of Comp. Vision*, 3(1):73–102, 1989.
- [43] S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, Tokyo, 1995.
- [44] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0–1 optimization. *SIAM J. Optimization*, 1(2):166–190, 1991.
- [45] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *Int. J. of Comp. Vision*, 43(1):7–27, 2001.
- [46] B. Mohar and S. Poljak. Eigenvalues in combinatorial optimization. In R.A. Brualdi, S. Friedland, and V. Klee, editors, *Combinatorial and Graph-Theoretical Problems in Linear Algebra*, volume 50 of *IMA Vol. Math. Appl.*, pages 107–151. Springer, 1993.
- [47] Y. Nesterov. Quality of semidefinite relaxation for nonconvex quadratic optimization. Technical report, CORE, Université Catholique de Louvain, Belgium, 1997.
- [48] Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming*. SIAM, 1994.
- [49] G. Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.*, 23(2):339–358, 1998.
- [50] C. Peterson and B. Söderberg. Artificial neural networks. In Aarts and Lenstra [1], chapter 7.
- [51] S. Poljak and F. Rendl. Nonpolyhedral relaxations of graph-bisection problems. *SIAM Journal on Optimization*, 5:467–487, 1995.
- [52] J. Puzicha, J.M. Buhmann, Y. Rubner, and C. Tomasi. Empirical evaluation of dissimilarity measures for color and texture. In *7th Int. Conf. on Comp. Vision (ICCV’99)*, pages 1165–1172, Los Alamitos, CA, 1999. IEEE Comp. Soc.

- [53] F. Rendl and H. Wolkowicz. A projection technique for partitioning the nodes of a graph. *Annals of Operations Research*, 58:155–180, 1995.
- [54] K. Rose, E. Gurewitz, and G.C. Fox. Constrained clustering as an optimization method. *IEEE Trans. Patt. Anal. Mach. Intell.*, 15(8):785–794, 1993.
- [55] S. Sarkar and K.L. Boyer. Integration, inference, and management of spatial information using bayesian networks: Perceptual organization. *IEEE Trans. Patt. Anal. Mach. Intell.*, 15(3):256–274, 1993.
- [56] M. Sato and S. Ishii. Bifurcations in mean-field-theory annealing. *Physical Review E*, 53(5):5153–5168, 1996.
- [57] C. Schellewald, J. Keuchel, and C. Schnörr. Image labeling and grouping by minimizing linear functionals over cones. In M. Figueiredo, J. Zerubia, and A.K. Jain, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2134 of *Lect. Not. Comp. Sci.*, pages 235–250, Berlin, 2001. Springer.
- [58] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 22(8):888–905, 2000.
- [59] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.
- [60] V. Vapnik. *Estimation of Dependencies Based on Empirical Data*. Springer, 1982.
- [61] Vision texture database. <http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>.
- [62] Y. Weiss. Segmentation using eigenvectors: a unifying view. In *Proc. Int. Conf. Comp. Vision (ICCV'99)*, pages 975–982, 1999.
- [63] L.R. Williams and K.K. Thornber. A comparison of measures for detecting natural shapes in cluttered backgrounds. *Int. J. of Comp. Vision*, 34(2/3):81–96, 2000.
- [64] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, volume 27 of *Appl. of Mathematics*. Springer-Verlag, Heidelberg, 1995.
- [65] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming*, Boston, 2000. Kluwer Acad. Publ.
- [66] S.J. Wright. *Primal–Dual Interior–Point Methods*. SIAM, 1996.
- [67] C.-h. Wu and P.C. Doerschuk. Cluster expansions for the deterministic computation of bayesian estimators based on markov random fields. *IEEE Trans. Patt. Anal. Mach. Intell.*, 17(3):275–293, 1995.

- [68] Y.N. Wu, S.C. Zhu, and X. Liu. Equivalence of julesz ensembles and frame models. *Int. J. of Comp. Vision*, 38(3):247–265, 2000.
- [69] Y. Ye. *Interior Point Algorithms: Theory and Analysis*. Wiley, 1997.
- [70] Y. Ye. A .699-approximation algorithm for max-bisection. *Mathematical Programming*, 90:101–111, 2001.
- [71] D. Zhu, S.C. and Mumford. Prior learning and gibbs reaction–diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(11):1236–1250, 1997.