Image Labeling and Grouping by Minimizing Linear Functionals over Cones

Christian Schellewald, Jens Keuchel and Christoph Schnörr

Computer Vision, Graphics, and Pattern Recognition Group
Department of Mathematics and Computer Science;
University of Mannheim, D-68131 Mannheim, Germany
Email: {cschelle,jkeuchel,schnoerr}@ti.uni-mannheim.de;
fax: +49 612 181 2744; http://www.cvgpr.uni-mannheim.de

Abstract. We consider energy minimization problems related to image labeling, partitioning, and grouping, which typically show up at mid-level stages of computer vision systems. A common feature of these problems is their intrinsic combinatorial complexity from an optimization point-of-view. Rather than trying to compute the global minimum – a goal we consider as elusive in these cases – we wish to design optimization approaches which exhibit two relevant properties: First, in *each* application a solution with *guaranteed* degree of suboptimality can be computed. Secondly, the computations are based on clearly defined algorithms which do not comprise *any* (hidden) tuning parameters.

In this paper, we focus on the second property and introduce a novel and general optimization technique to the field of computer vision which amounts to compute a suboptimal solution by just solving a *convex* optimization problem. As representative examples, we consider two binary quadratic energy functionals related to image labeling and perceptual grouping. Both problems can be considered as instances of a general quadratic functional in binary variables, which is embedded into a higher–dimensional space such that suboptimal solutions can be computed as minima of linear functionals over cones in that space (semidefinite programs). Extensive numerical results reveal that, on the average, suboptimal solutions can be computed which yield a gap below 5% with respect to the global optimum in case where this is known.

1 Introduction

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. Important examples include the seminal paper by Geman and Geman [1] on simulated annealing, approaches for suboptimal Markov Random Field (MRF) minimization like the ICM-algorithm [2], the highest-confidence-first heuristic [3], multi-scale approaches [4], and other approximations [5, 6]. A further important class of approaches comprises continuation methods like Leclers partitioning approach [7], the graduated-non-convexity strategy by Blake and Zisserman [8], and various deterministic (approximate) versions of the annealing approach in applications like surface reconstruction [9], perceptual grouping [10], graph matching [11], or clustering [12].

Apart from the simulated annealing approach using annealing schedules which are unpractically slow for real-world applications (but prescribed by theory, see [1]), 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 arises: How good is a minimizer computed 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 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 [2, 3], 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 [13] 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 [10, p. 906] and [15]).

These problems can be avoided by going back to the mathematically well-understood class of *convex* optimization problems. Under mild assumptions there exists a global optimum which, in turn, leads to a suboptimal solution of the original problem, along with clear 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. Concerning global energy-minimization problems in computer vision, this has been exploited for *continuous*-valued functions in [16, 17], for example, to approximate the classical Mumford-Shah functional [18] for image segmentation.

In this paper, however, we focus on more difficult problems by extending this line of research to prototypical energy-minimization problems involving *discrete* decision variables. Our work is based on the seminal paper by Lovász and Schrijver [19] who showed how tight problem relaxations can be obtained by lifting the problem up into some higher-dimensional space and down-projecting to a convex set containing feasible solutions in that space. This idea has been put forward and lead to a remarkable result by Goemans and Williamson [20], who were able to show for a classical combinatorial problem that suboptimal solutions (for the special problem considered) cannot be worse than 14% relative to the unknown global optimum. These two facts – bounds on the suboptimality, and algorithm design based on convex optimization – have motivated our work.

Organisation of the paper. We consider in Section 2 two representatives of the class of quadratic functionals in binary variables. This class of minimization problems is well-known in the context of image labeling, perceptual grouping, MRF-modeling, etc. We derive a problem relaxation leading to a convex optimization problem in Section 3. The corresponding convex programming techniques are sketched in Section 4. In Section 5, we illustrate the properties of our approach by describing ground-truth experiments conducted with one-dimensional signals, for which the global optimum can be easily computed with dynamic programming. Real-world examples are discussed in Section 6, and we conclude our paper by indicating further work in Section 7.

Notation. For a vector $y \in \mathbb{R}^n$, $D(y)$ denotes the diagonal matrix with entries $y_1, \ldots, y_n$. e denotes the vector of one's, $e_i = 1, \forall i$, and $I = D(e)$ the unit matrix. For a matrix X, $D(X)$ denotes the diagonal matrix with the diagonal elements $x_{ii}, \forall i$, of X. $\mathcal{S}^n$ denotes the space of symmetric $n \times n$-matrices $X^t = X$, and $\mathcal{S}^n_+$ denotes the matrices $X \in \mathcal{S}^n$ which are positive semidefinite. For abbreviation, we will also use the symbol $\mathcal{K} = \mathcal{S}^{n+1}_+$. For two matrices $X, Y \in \mathcal{S}^n$, $X \bullet Y = \text{trace}(XY)$ denotes the standard matrix inner product.

2 Problem Statement: Minimizing Binary Quadratic Functionals

In this paper, we consider the problem to minimize functionals of the general form:

$$J(x) = x^t Q x + 2b^t x + const \ , \quad x \in \{-1, 1\}^n, \ Q \in \mathcal{S}^n, \ b \in \mathbb{R}^n \ . \qquad (1)$$

In the field of computer vision, such global optimization problems arise in various contexts. In the following sections, we give two examples related to image labeling and perceptual grouping, respectively.

Note that apart from symmetry, no further constraints are imposed on the matrix Q in (1). Hence, the functional J need not to be convex in general. This property along with the integer constraint $x_i \in \{-1, 1\}$, $i = 1, \ldots, n$, makes the minimization problem (1) intrinsically difficult.

In Section 3 we will relax some of these hard constraints so as to arrive at a convex optimization problem which closely approximates the original one.

2.1 Example 1: Binary Image Restoration and Labeling

Consider some scalar-valued feature (grey-value, color feature, texture measure, etc.) $g : \Omega \to \mathbb{R}$ which has been locally computed within the image plane. Suppose that for each pixel position i, feature g 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. Figure 1 shows an example.

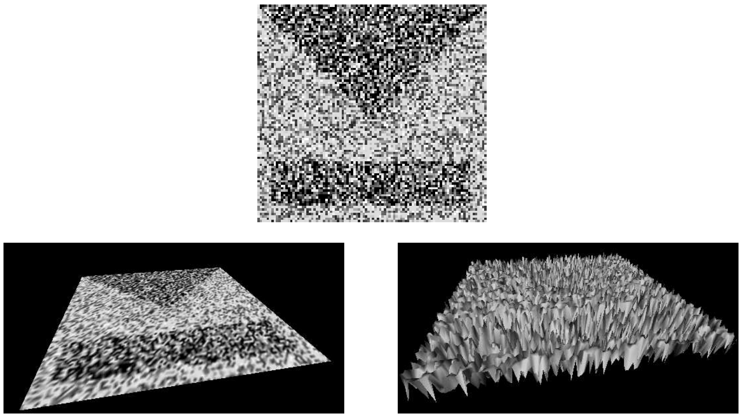


Fig. 1. Top: A binary image, heavily corrupted by (real-valued) noise. **Bottom, left:** The original data textured on a plane. **Bottom, right:** The data as 3D-plot to illustrate the poor signal-to-noise ratio.

To restore a discrete-valued image function x from the measurements g, we wish to compute x as minimizer of a functional which has the form (1):

$$J(x) = \frac{1}{4}\sum_i \big((u_2 - u_1)x_i + u_1 + u_2 - 2g_i\big)^2 + \frac{\lambda}{2}\sum_{\langle i,j\rangle}(x_i - x_j)^2,\ x_i \in \{-1,1\},\ \forall i\ . \tag{2}$$

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

Functional (2) comprises two terms familiar from many regularization approaches [21]: 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.

We note further that, depending on the application considered, it might be useful to modify the terms in (2), either 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., [22]). These modifications, however, would not increase the difficulty of problem (2) from an optimization point-of-view.

2.2 Example 2: Figure-Ground Discrimination and Perceptual Grouping

Let $g_i, i = 1,\dots,n$, denote some feature primitive irregularly distributed over the image plane. Suppose that for each pair of primitives g_i, g_j, we can compute some (dis)similarity measure d_{ij} corresponding to some of the well-known "Gestalt laws", or to some specific object properties learned from examples. For instance, g_i might denote an edge-element computed at location i in the image plane, and d_{ij} might denote some measure corresponding to smooth continuation, co-circularity, etc. For an overview over various features and strategies for perceptual grouping we refer to [23].

According to the spatial context modeled by d_{ij}, we wish to separate familiar configurations from the (unknown) background. To this end, following [10], we label each primitive g_i with a decision variable $x_i \in \{-1, 1\}$ ("1" corresponding to figure, "-1" corresponding to background and noise) and wish to minimize a functional of the form (1):

$$J(x) = \sum_{\langle i,j \rangle} (\lambda - d_{ij}) x_i x_j + 2 \sum_i (\lambda n - \sum_j d_{ij}) x_i, \; x_i \in \{-1, 1\}, \; \forall i \; . \qquad (3)$$

Figure 2 shows a test-problem we use in this paper for illustration. On the left, some "object" is shown which distinguishes itself from background and clutter by the relative angles of edgels. Such edgels typically arise as output of some local edge detector. Accordingly, the difference between the relative angles $\Delta\phi_{ij}$ and the expected ones (due to our knowledge about the object) were chosen as similarity measure d_{ij} with respect to primitives i and j. In addition, we take into consideration inaccuracies of a (fictive) preprocessing stage by virtue of a parameter β (see Figure 2):

$$d_{ij} = \exp(-\beta \prod_k (\Delta\phi_{ij} - \Delta\phi_{k,\text{expected}})^2) \; .$$

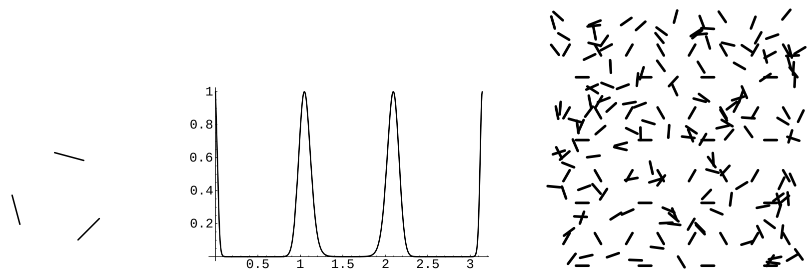


Fig. 2. Left: An "object". **Middle:** The similarity measure for $\Delta\phi_{ij} \in [0, \pi), \beta = 20$, according to the expected relative angles, and allowing for some inaccuracies of a (fictive) preprocessing stage. **Right:** The object was rotated by a fixed arbitrary angle, and translated and scaled copies have been superimposed by noise. Where are these objects?

Clearly, this measure is invariant against translation, rotation and scaling of the object. On the right in Figure 2, an unknown number of translated and scaled copies of the object, which has been rotated in advance by an unknown angle, is shown together with a lot of noisy primitives as "background". Trying to find these objects leads to combinatorial search. By contrast, we are interested in suboptimal minimizers of the functional (3) computed by convex programming.

3 Convex Problem Relaxation

Recall that both problems (2) and (3) (and many others) are special cases of problem (1).

In order to relax problem (1), we first drop the constant and homogenize the objective function as follows:

$$x^t Q x + 2b^t x = \begin{pmatrix} x \\ 1 \end{pmatrix}^t L \begin{pmatrix} x \\ 1 \end{pmatrix} \;, \quad L = \begin{pmatrix} Q & b \\ b^t & 0 \end{pmatrix} \;. \tag{4}$$

With slight abuse of notation, we denote the vector $\begin{pmatrix} x & 1 \end{pmatrix}^t$ again by x.

Next we introduce the Lagrangian with respect to problem (1):

$$x^t L x - \sum_i y_i (x_i^2 - 1) = x^t (L - D(y)) x + e^t y$$

and the corresponding minimax-problem:

$$\sup_y \inf_x x^t (L - D(y)) x + e^t y \;.$$

Since x is unconstrained now, the inner minimization is finite-valued only if $L - D(y) \in \mathcal{S}_+^{n+1} = \mathcal{K}$ (for notation, see section 1). Hence we arrive at the relaxed problem:

$$\sup_y e^t y \;, \quad L - D(y) \in \mathcal{K} \;. \tag{5}$$

The important point here is that problem (5) is a *convex* optimization problem! The set $\mathcal{K}$ is a cone (i.e. a special convex set) and self-dual, that is it coincides with the dual cone [24]

$$\mathcal{K}^* = \{Y : \; X \bullet Y \geq 0, \; X \in \mathcal{K}\} \;.$$

To obtain the connection to our original problem, we derive the dual problem associated with (5). Choosing a Lagrangian multiplier $X \in \mathcal{K}^* = \mathcal{K}$, similar reasoning as above yields:

$$\begin{aligned} \sup_y e^t y &= \sup_y \inf_{X \in \mathcal{K}} e^t y + X \bullet (L - D(y)) \\ &\leq \inf_{X \in \mathcal{K}} \sup_y e^t y + X \bullet (L - D(y)) \\ &= \inf_{X \in \mathcal{K}} \sup_y L \bullet X - D(y) \bullet (X - I) \;. \end{aligned}$$

The inner maximization of the last equation is finite iff $D(X) = I$. Hence, we obtain as the problem dual to (5):

$$\inf_{X \in \mathcal{K}} L \bullet X \ , \quad D(X) = I \ . \tag{6}$$

which again is convex.

In order to compare the relaxation (6) with the problems (1) and (4), respectively, we rewrite the latter as follows:

$$\inf_{x \in \{-1,1\}^{n+1}} x^t L x = \inf_{x \in \{-1,1\}^{n+1}} L \bullet xx^t \ .$$

Note that the matrix $xx^t \in \mathcal{K}$ and has rank one. A comparison with the relaxed problem (6) shows that (i) xx^t is replaced by an arbitrary matrix $X \in \mathcal{K}$ (i.e. the rank one condition is dropped), and (ii) that the integer constraint $x_i \in \{-1, 1\}$ is *weakly* imposed by the constraint $D(X) = I$ in (6).

In the following sections, we will examine the relaxed problem (6) with respect to the criteria discussed in Section 1.

4 Algorithm

The primal–dual pair of optimization problems (6) and (5), respectively, belongs to the class of conic programs. The elegant duality theory corresponding to this class of convex optimization problems can be found in [24]. For "well-behaved" instances of this problem class, optimal primal and dual solutions X^*, y^*, S^* exist (S denotes a matrix of slack variables) and are complementary to each other: $X^* \bullet S^* = 0$. Moreover, no duality gap exists between the optimal values of the corresponding objective functions:

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

To compute X^*, y^* and S^*, a wide range of iterative interior-point algorithms can be used. Typically, a sequence of minimizers X_η, y_η, S_η, parametrized by a parameter η, is computed until the duality gap falls below some threshold ϵ. A remarkable result in [24] asserts that for the family of self-concordant barrier functions, this can always be done in polynomial time, depending on the number of variables n and the value of ϵ.

For our experiments described in the following two sections, we chose the so-called dual-scaling algorithm using public software from a corresponding website [25]. To get back the solution x to (1) from the solution X to (6), we used the randomized-hyperplane technique described in [20].

A more detailed description of the algorithm, along with useful modifications according to the problem class considered, is beyond the scope of this paper and will be reported elsewhere.

5 Performance: Ground-Truth Numerical Experiments

In this section, we investigate the performance of the relaxed problem (6) experimentally. To this end, we report the statistical results for three different ground-truth experiments using one-dimensional random signals.

We chose one-dimensional signals in this section because ground truth (the global optimum) can be easily computed using dynamic programming. Numerical results concerning two-dimensional signals (images) and grouping experiments are reported in section 6.

In what follows, we denote with x^* the global minimizer of (1), and with x the suboptimal solution reconstructed from the solution X to the convex programming problem (5),(6).

5.1 Ground-Truth Experiments: Partitioning of Random Signals

For the first series of experiments, we generated 1000 random signals, each with 256 pixel values equally distributed in the range $[-1, 1]$. Figure 4, top, shows an example.

To investigate the performance of the relaxed problem we compare the global optimum with the results from the relaxed problem. The optimal objective function is bounded as follows:

$$\inf_{X \in \mathcal{K}, D(X)=I} L \bullet X \;\leq\; J(x^*) \;\leq\; J(x) \,. \tag{7}$$

The left inequality holds true due to the relaxation of problem (1), as described in Section 3. The right inequality is obvious because x^* is the global minimizer.

To evaluate this relationship numerically, we used the following quantities:

- $\overline{J^*}$: the sample mean of the global optimum $J^* = J(x^*)$ of the functional (2) (computed with dynamic programming),
- $\overline{\Delta J}$: the sample mean of the gap $\Delta J = J - J^*$ (measured in % of the optimum) with respect to the objective function values of the suboptimal solution $J(x)$ and the optimal solution J^*, and
- $\sigma_{\Delta J}$: the sample standard deviation of the gap ΔJ.

The resulting values of these quantities are shown in Figure 3, for different values of the global parameter λ (1000 random signals were generated for *each* value of λ). Figure 3 shows that for reasonable values of λ, the gap $\overline{\Delta J}$ is about 5% of the optimal value of the objective function.

Taking into consideration that these suboptimal solutions can be computed by solving a mathematically much simpler convex optimization problem, the quality of these solutions is surprisingly good!

The purely random signals considered in this section exhibit another property: There are many solutions having similar values of the objective function which however differ considerably with respect to the Hamming distance. Figure 4 illustrates this fact for an arbitrary random signal and a solution pair x, x^* leading to a gap of $\Delta J = 6.4\%$, but differing at 58 pixel-positions (=22.7%).

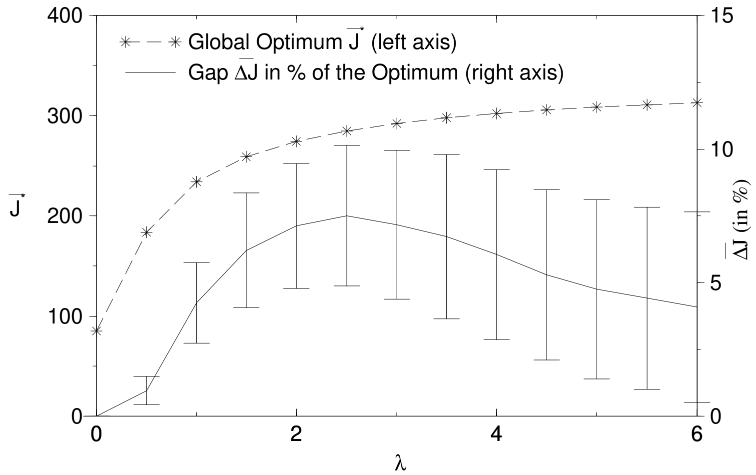


Fig. 3. Sample mean of the optimal value of the objective function $\overline{J^*}$, sample mean of the corresponding gap $\overline{\Delta J}$ with respect to the suboptimal solution, and corresponding standard deviation $\sigma_{\Delta J}$. On the average, the quality of the suboptimal solution is around 5%.

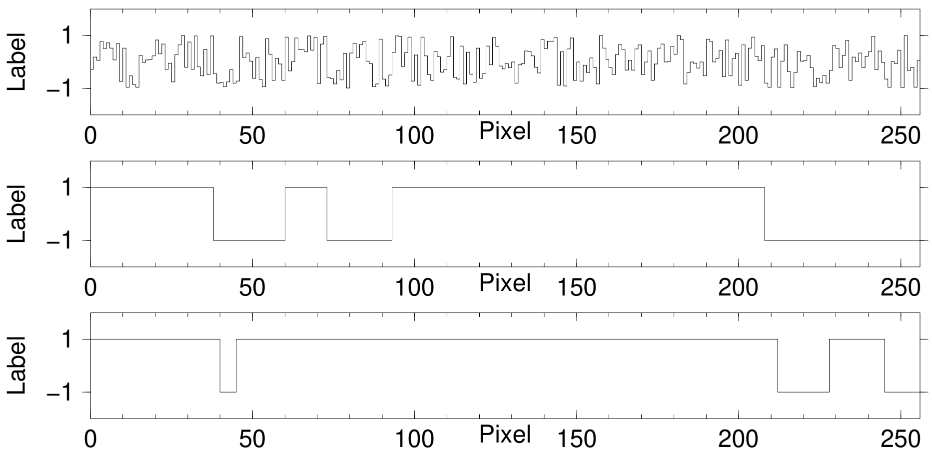


Fig. 4. Top: A purely random input signal. **Middle:** The optimal solution x^*. **Bottom:** The suboptimal solution x. Although the gap $\Delta J = 6.4\%$ only, the Hamming distance between these two solutions is not small. This effect is due to missing structure of the input signal (see text).

On the other hand, no spatial context can be exploited for pure random signals. Accordingly, there is no meaningful parameter value of λ which could give a more accurate solution. Therefore, this negative effect should not be taken too serious because it disappears as soon as the input signal exhibits more structure, as is the case for real signals. This will be confirmed in the following sections.

5.2 Ground-Truth Experiments: Restoration of Noisy Signals Comprising Multiple Scales

In our second series of experiments, we took the synthetic signal x' depicted in Figure 5 which involves transitions at multiple spatial scales, and superimposed Gaussian white noise with standard deviation $\sigma = 1.0$. Figure 7, top, shows an example. The goal was to restore the synthetic signal from the noisy input signals. For each value of λ, we repeated this experiment 1000 times using different noise signals.

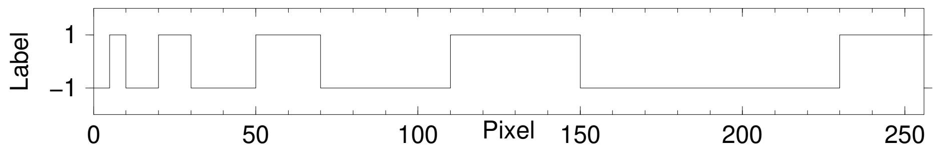


Fig. 5. Signal x' comprising multiple spatial scales.

In addition to the measures introduced in the last section, we computed the following quantities:

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

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

The statistics of our numerical results are shown in Figure 6. Two observations can be made: First, for values of the scale-parameter $\lambda > 1.5$, the restoration is quite accurate: $\overline{\Delta J'} < 3\%$. Secondly, the fact $\overline{\Delta J} < \overline{\Delta J'}$ indicates that more appropriate criteria should exist for the restoration of signals that are structured like x' (see Fig. 5). The derivation of such functionals is not the objective of this paper. However, we point out that such learning problems can probably be solved within the general class (1). In that case, our optimization framework could be applied, too.

In order not to overload Figure 6, we did not include the measures $\sigma_{\Delta J'}$ and $\sigma_{\Delta J}$. The average values are $\sigma_{\Delta J'} = 3.16\%$ and $\sigma_{\Delta J} = 0.80\%$. These values are significantly smaller than those of the previous experiment, and thus they confirm the statements made at the end of the last section.

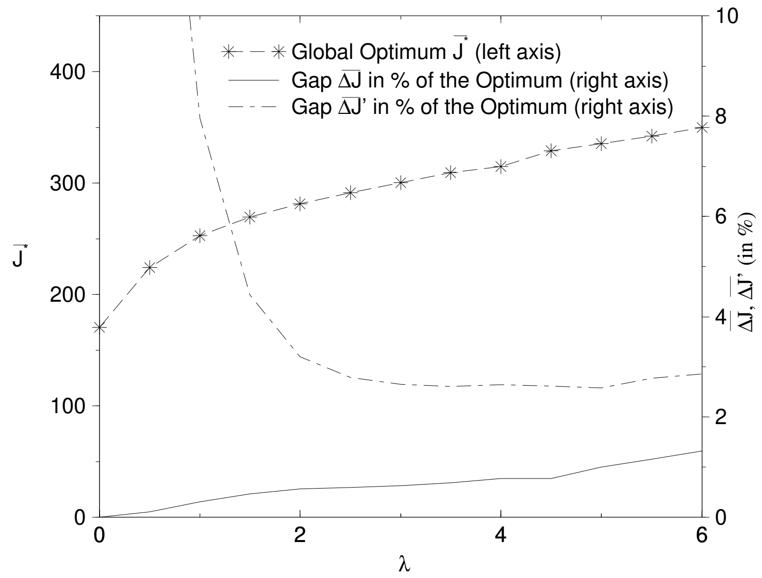


Fig. 6. Average gaps $\overline{\Delta J}$ and $\overline{\Delta J'}$ for noisy versions ($\sigma = 1$) of the signal x' shown in Fig. 5, and for different values of the scale parameter λ. According to the dominating spatial scales in signal x', for $\lambda > 1.5$ the quality of the restoration is remarkably good (below 4%).

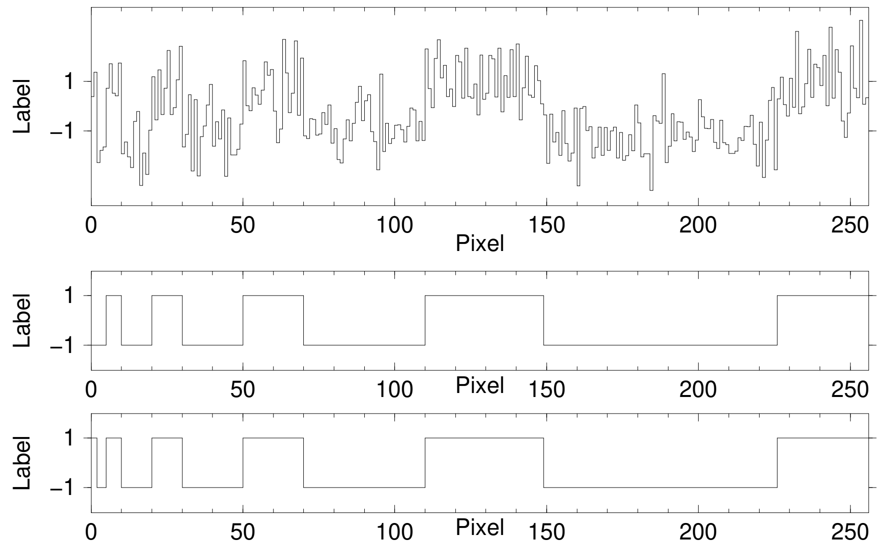


Fig. 7. A representative example illustrating the statistics shown in Fig. 6. **Top:** Noisy input signal. **Middle:** Optimal Solution x^*. **Bottom:** Suboptimal solution x.

5.3 Ground-Truth Experiment: Real 1D-signal

Before turning to two-dimensional signals in the next section, it is quite illustrative to look at numerical results for a real one-dimensional signal, namely a column of the noisy image depicted in Figure 1. In Figure 8, top, the noisy column of this image is shown.

The following two plots in Figure 1 show the global minimizer x^* computed with dynamic programming, and the suboptimal solution x computed with convex programming, respectively, for an appropriate value of the scale-parameter $\lambda = 2$.

This result demonstrates once more the "tightness" of the convex approximation of the combinatorial optimization problem (1).

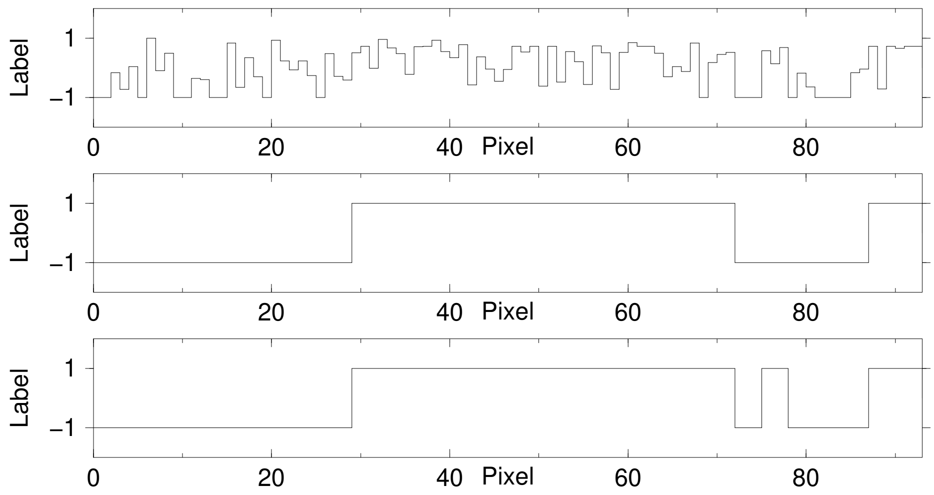


Fig. 8. Top: Column 28 of the very noisy image shown in Fig. 1. **Middle:** Optimal solution x^* computed by dynamic programming. **Bottom:** Suboptimal solution x computed by convex programming.

6 Numerical Experiments: 2D-images and Grouping

In the previous section, we showed the performance of the algorithm in the context of one-dimensional signals. We will next discuss the results of applying the algorithm to two-dimensional images. Computing the global optimum for real 2D-signals (images) is no longer possible. To demonstrate the wide range of problems that can in principle be tackled by the general approach (1), we also include results with respect to a grouping problem (see Figure 2).

The results concerning the restoration of the real image shown in Figure 1 are shown in Figure 9. Taking into consideration the quite poor signal-to-noise ratio, the quality of the restoration is encouraging.

Figure 10 shows the same experiment with respect to another image. Note that the desired object to be restored comprises structures at both large and small spatial scales. Again the restoration result using convex programming is surprisingly good.

Next, Figure 11 shows the well known checkerboard experiment. As can be expected, small errors only occur at corners, that is at local structures with a very small spatial structure close to noise.

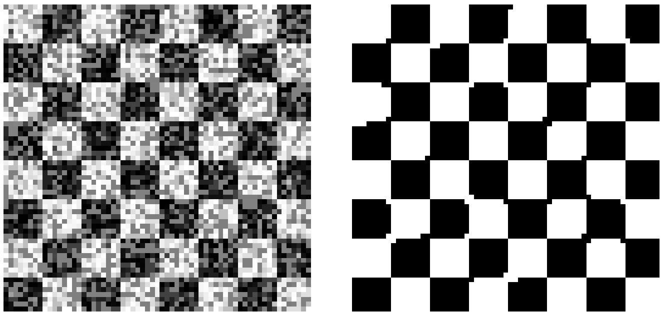
Finally, the results of the grouping problem (see section 2.2) are depicted in Figure 12. The suboptimal solution computed by convex programming clearly separates structure from background, apart from a small number of edgels. The presence of these extra edgels however is not caused by our optimization approach but is consistent with the chosen similarity measure which fails to label them as dissimilar.

(a) (b) (c) (d)

Fig. 9. Arrow and bar real image. **(a)** Noisy original. **(b), (c), (d):** Suboptimal solutions computed by convex programming for $\lambda = 0.8, 1.5, 3.0$.

(a) (b) (c)

Fig. 10. Iceland image. **(a)** Binary noisy original. **(b)** Suboptimal solution computed by convex programming with $\lambda = 2.0$. **(c)** Original before adding noise.



(a) (b)

Fig. 11. Checkerboard image. **(a)** Noisy original. **(b)** Suboptimal solution computed by convex programming with $\lambda = 1.5$.

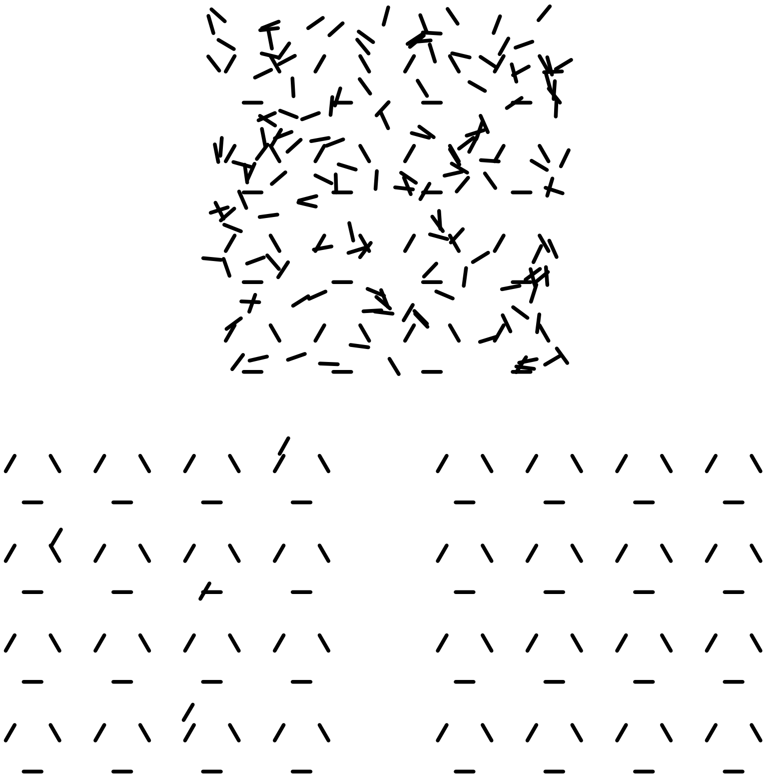


Fig. 12. Top: Input data (see Section 2.2). **Bottom, left:** The suboptimal solution computed with convex optimization ($\lambda = 0.9$). Four false primitives are included which – according to the relative angle measure – cannot be distinguished from object primitives. **Bottom, right:** The true solution.

7 Conclusion and Further Work

In this paper, we introduced a novel optimization technique to the field of image processing and computer vision. This technique applies to various energy minimization problems of mid-level vision, the objective function of which typically belongs to the large class of binary quadratic functionals.

The most important property which distinguishes our approach from related work is its mathematical simplicity: Suboptimal solutions can be computed by just solving a *convex* optimization problem. As a consequence, no additional tuning parameters related to search heuristics, etc. are needed, apart from the parameters of the original model itself, of course.

For two representative functionals related to image labeling and grouping, extensive numerical experiments revealed a surprising quality of suboptimal solutions with an error below 5% on the average. Due to this fact as well as the clear algorithmic properties of our approach, we consider it as an attractive candidate in the context of computational vision.

We will continue our work as follows: First, we will try to prove bounds with respect to the suboptimality of solutions (see Eqn. (7)). Furthermore, we will focus on the algorithmic properties in order to exploit sparsity of specific problems. Finally, other problems in the general class (1) like matching of relational object representations, for example, will be investigated.

Acknowledgements. The authors enjoyed discussing this paper with Daniel Cremers. Programming support by Gregor Schlosser concerning the laborious experiments in Section 5 is gratefully acknowledged.
Jens Keuchel has been supported by the German National Science Foundation (DFG) under the grant Sch457/3-1.

References

1. 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.
2. J.E. Besag. On the analysis of dirty pictures (with discussion). *J. R. Statist. Soc. B*, 48:259–302, 1986.
3. 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.
4. 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.
5. 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.
6. 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.

7. Y.G. Leclerc. Constructing simple stable descriptions for image partitioning. *Int. J. of Comp. Vision*, 3(1):73–102, 1989.
8. A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
9. 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.
10. L. Herault and R. Horaud. Figure-ground discrimination: A combinatorial optimization approach. *IEEE Trans. Patt. Anal. Mach. Intell.*, 15(9):899–914, 1993.
11. S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18(4):377–388, 1996.
12. T. Hofmann and J. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(1):1–14, 1997.
13. M. Sato and S. Ishii. Bifurcations in mean-field-theory annealing. *Physical Review E*, 53(5):5153–5168, 1996.
14. E. Aarts and J.K. Lenstra, editors. *Local Search in Combinatorial Optimization*, Chichester, 1997. Wiley & Sons.
15. C. Peterson and B. Söderberg. Artificial neural networks. In Aarts and Lenstra [14], chapter 7.
16. C. Schnörr. Unique reconstruction of piecewise smooth images by minimizing strictly convex non-quadratic functionals. *J. of Math. Imag. Vision*, 4:189–198, 1994.
17. C. Schnörr. A study of a convex variational diffusion approach for image segmentation and feature extraction. *J. of Math. Imag. and Vision*, 8(3):271–292, 1998.
18. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
19. L. Lovász and A. Schrijver. Cones of matrices and set–functions and 0–1 optimization. *SIAM J. Optimization*, 1(2):166–190, 1991.
20. 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.
21. M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proc. IEEE*, 76:869–889, 1988.
22. G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, volume 27 of *Appl. of Mathematics*. Springer-Verlag, Heidelberg, 1995.
23. S. Sarkar and K.L. Boyer. Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Tr. Systems, Man, and Cyb.*, 23(2):382–399, 1993.
24. Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming*. SIAM, 1994.
25. 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.