Tight Convex Relaxations for Vector-Valued Labeling Problems

Evgeny Strekalovskiy, Bastian Goldluecke and Daniel Cremers
TU Munich, Germany

Abstract

The multi-label problem is of fundamental importance to computer vision, yet finding global minima of the associated energies is very hard and usually impossible in practice. Recently, progress has been made using continuous formulations of the multi-label problem and solving a convex relaxation globally, thereby getting a solution with optimality bounds. In this work, we develop a novel framework for continuous convex relaxations, where the label space is a continuous product space. In this setting, we can combine the memory efficient product relaxation of [9] with the much tighter relaxation of [5], which leads to solutions closer to the global optimum. Furthermore, the new setting allows us to formulate more general continuous regularizers, which can be freely combined in the different label dimensions. We also improve upon the relaxation of the products in the data term of [9], which removes the need for artificial smoothing and allows the use of exact solvers.

1. Introduction

Labeling problems in computer vision. Recently, there has been a surge of research activity in convex relaxation techniques for energy minimization in computer vision. Particular efforts were directed towards multilabel problems, as they lie at the heart of fundamental problems like segmentation [15], stereo [17], 3D reconstruction [7] and optic flow [9].

In multilabel problems, we are looking for a pointwise labeling $u : \Omega \to \Gamma$ of a domain $\Omega \subset \mathbb{R}^n$ which is optimal in the following sense. Assigning the label $\gamma \in \Gamma$ to a point x incurs the *cost* $c(x, \gamma) \in \mathbb{R}$. The cost usually denotes how well the labeling fits the observed data. It can be an arbitrarily complex function, derived from statistical models or local matching scores. Aside from minimizing the local costs, we want the optimal assignment to exhibit a certain regularity. This requirement is encoded in a prior term $J(u)$. The prior represents our knowledge about which label configurations are more likely, and typically enforces spacial coherence. The solution to the labeling problem is the minimizer of the sum of the local costs and the regularity prior,

$$\operatorname*{argmin}_{u:\Omega\to\Gamma} J(u) + \int_\Omega c(x, u(x))\,\mathrm{d}x. \tag{1}$$

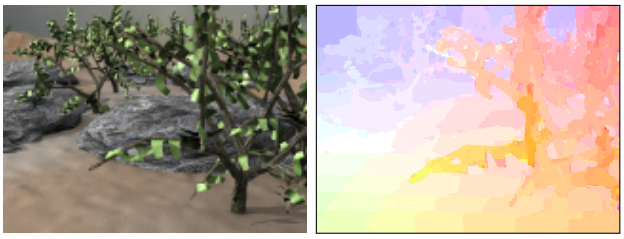


Figure 1: *For multilabel problems like optical flow estimation we propose a novel relaxation which is tighter and more general than previous ones. In contrast to [14, 9], we can use an exact instead of approximated relaxation of the truncated linear regularizer.*

Unfortunately, finding a global minimum of non-convex energies is a very hard problem, and in most cases impossible in practice. In some cases, good results may be obtained by local minimization, starting from a good initialization. Yet, such methods cannot guarantee any form of quality of the result - it could be arbitrarily bad. Therefore, the ultimate goal is to find methods which at least get within provable distance to the global optimum.

Optimal and approximate solutions. True global solutions to (1) can only be found in rare special cases. If the label space is binary and the regularizer submodular, a global minimum can be computed with a minimum cut [10]. In the continuous setting, the two-label segmentation problem with length regularity can be solved by relaxation to a convex problem and subsequent thresholding [15]. A technique called functional lifting was introduced in [17] to solve the special case of the multilabel problem with convex interaction terms and a linearly ordered set of labels. The idea is that the original problem can be reformulated as an equivalent higher-dimensional problem which is convex. It is reminiscent of a similar construction [11] for the discrete case, where a global minimizer for these kinds of problems was computed as a cut in a multi-layered graph.

However, for general non-convex regularizers like the Potts distance, the resulting combinatorial problem is NP-

hard [3]. In the discrete setting, one can approximate a solution for example by solving a sequence of binary problems (α-expansions) [3], linear programming relaxations [18], or quadratic pseudo-boolean optimization [13]. The problem of large label spaces is also tackled in [8], where the authors compute optical flow from an MRF labeling problem using a lower dimensional parametric description for the displacements. In this paper, we work in the fully continuous setting, avoiding typical problems of graph-based discretization like metrication errors [12].

Continuous convex relaxations. Continuous approaches deal with the multi-label problem by *convex relaxation*. The original non-convex energy is replaced with a convex lower bound, which can be minimized globally. We automatically get a bound on the solution and know how far we are from the global optimum. How good the bound is depends on the *tightness* of the relaxation, i.e. how close the new energy is to the old one. While it sometimes can be possible to even achieve global optimality using this class of methods [15, 17], there is no relaxation known which leads to globally optimal solutions of the general problem. Relaxations of different tightness have been proposed in [14, 5, 19]. They all have in common that they are very memory intensive if the number of labels becomes larger, which makes it impossible to use them for scenarios with thousands of labels, like for example optic flow.

Contributions. In [9], the authors reduced memory requirements and made it possible to deal with a very large number of labels, as long as the space of labels carries a product structure. We improve upon their construction in two significant ways. First, their regularizer is based on the relaxation in [14] for multilabel problems with a discrete set of labels, which is known to be less tight than the relaxation introduced in [5] for continuous label spaces. Second, their relaxation of the data term is suboptimal in that it introduces unwanted trivial solutions when relaxing from binary to continuous labels, which have to be avoided by an additional smoothing degrading the quality of solutions.

Our contribution is to propose a general framework for convex relaxations of multilabel problems, which is based on a *continuous, multi-dimensional label space*. Regarding the regularization, we are then able to combine the advantages of the efficient multi-dimensional relaxation of [9] with the tight relaxation of [5]. We show that previous relaxations appear as special cases of our framework. In addition, we can formulate more general regularizers on multi-dimensional label spaces and thus solve a more general class of problems efficiently. Furthermore, we propose a *novel convex relaxation of the data term* which is a substantial improvement to [9]. It is not only much tighter, but also avoids the need for an additional approximation as in [9]. Overall, the new framework yields solutions which are provably closer to the global optimum.

2. Multidimensional Labeling Problems

In this work, we propose the setting of a continuous, multi-dimensional label space. In this way, we are able to combine the advantages of the efficient multi-dimensional relaxation of [9] with the tight relaxation of [5].

Let the space of labels be denoted by Γ. We assume that it has the structure of a continuous product space,

$$\Gamma = \Lambda_1 \times ... \times \Lambda_d,$$

where the factors Λ_k are intervals in $\mathbb{R}$. We are looking for an optimal labeling $u : \Omega \to \Gamma$ of the domain $\Omega \subset \mathbb{R}^n$. The motivation for writing the space of labels as a product has been detailed in [9]. In particular, the memory requirements are reduced by orders of magnitude, which makes the method applicable to problems with a very large number of labels.

Regularizer. We write the components of u as $(u^1, \dots, u^d)$, and consider a general separable regularizer

$$J(u) = \sum_{k=1}^{d} J_k(u^k). \tag{2}$$

This way, each J_k acts on the components of u independently.

In order to define the regularizer, we require some technical preliminaries. Recall [2] that for functions u^k in the space $\mathcal{SBV}(\Omega)$ of special functions of bounded variation, the distributional derivative Du^k of u^k can be decomposed as

$$Du^k = \nabla u^k \, \mathrm{d}x + ((u^k)^+ - (u^k)^-)\nu_{u^k} \, dH^{n-1} \llcorner S_{u^k}$$

into a differentiable part and a jump part, see Figure 2. Here, S_{u^k} is the $(n-1)$-dimensional jump set of u^k where the values jump from $(u^k)^-$ to $(u^k)^+$, ν_{u^k} is the normal to S_{u^k} from the $(u^k)^-$ to the $(u^k)^+$ side, and ∇u^k is the approximate gradient of u^k. $\mathcal{H}^{n-1} \llcorner S_{u^k}$ denotes the $(n-1)$-dimensional Hausdorff measure restricted to the set S_{u^k}. We refer to [2] for a comprehensive introduction to functions of bounded variation.

After these preliminary remarks, we are now in a position to define the requirements of a regularizer for the labeling functions u_k.

Definition. In our framework, we consider regularizers for problem (1) of the form (2), with

$$J_k(u^k) = \int_{\Omega \setminus S_{u^k}} h_k(x, u^k(x), \nabla u^k(x)) + \int_{S_{u^k}} d_k\big(s, (u^k)^-(s), (u^k)^+(s)\big), \tag{3}$$

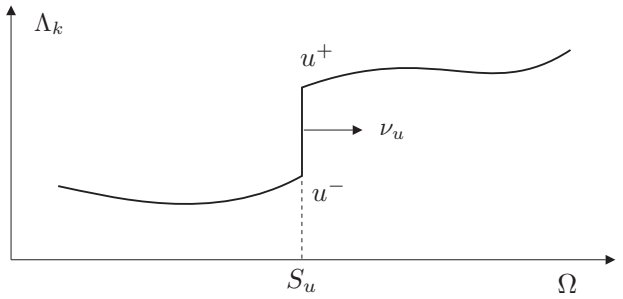


Figure 2: *A special function of bounded variation u has an approximate gradient everywhere except on a nullset S_u, where the values jump from u^- to u^+. The normal ν_u denotes the direction of the jump from small to large values.*

with functions $h_k : \Omega \times \Lambda_k \times \mathbb{R}^n \to \mathbb{R}$ and $d_k : \Omega \times \Lambda_k \times \Lambda_k \to \mathbb{R}$. The functions h_k and d_k have to satisfy the following conditions:

1. $h_k(x, \lambda, p)$ is convex in p for fixed x, λ.
2. $d_k(x, \cdot, \cdot)$ is a metric on Λ_k for fixed x.

The interesting task, of course, is to identify suitable choices of h_k and d_k, and to interpret what the choice means in practice. We will turn to this in section 4. Before we can explore the relationships, however, we need to introduce a convex relaxation of the regularizer in the next section.

3. Convex Relaxation and Discretization

Convex representation of the regularizer. In order to arrive at a convex representation of the regularizer, we formulate it in terms of new unknowns v^k defined as

$$v^k(x, \lambda) := \delta(u^k(x) - \lambda). \tag{4}$$

Note that the new unknowns are distributions on the higher dimensional space $\Omega \times \Lambda_k$, which however will reduce to regular functions after discretization. They are a generalization of the label indicator functions in [9, 14] to the continuous label space, in particular they satisfy the relations

$$\int_{\Lambda_k} v^k(x, \lambda)\, \mathrm{d}\lambda = 1, \quad \int_{\Lambda_k} \lambda\, v^k(x, \lambda)\, \mathrm{d}\lambda = u^k(x). \tag{5}$$

Intuitively, this means that for each fixed $x \in \Omega$, $v^k(x, \cdot)$ has a total mass of 1 and is concentrated on the point $u^k(x)$. The functional J_k can then be represented in the following form, which is an application of the calibration method discussed in [1]. We refer to their work for an in-depth introduction to this technique.

Theorem 3.1. *Let J_k be of the form (3), and the indicator distributions v_k defined as in (4). Then*

$$J_k(u^k) = \sup_{(p,b) \in C_k} \int_{\Omega \times \Lambda_k} (-\mathrm{div}(p) - b)\, v^k\, \mathrm{d}(x, \lambda), \tag{6}$$

with the convex set

$$\begin{aligned} C_k = \Big\{ (p, b) : \Omega \times \Lambda_k \to \mathbb{R}^n \times \mathbb{R} \\ \text{such that for all } x \in \Omega \text{ and } \lambda, \gamma \in \Lambda_k, \\ b(x, \lambda) \geq h_k^*\big(x, \lambda, \partial_\lambda p(x, \lambda)\big), \\ |p(x, \lambda) - p(x, \gamma)| \leq d_k(x, \lambda, \gamma) \Big\}. \end{aligned} \tag{7}$$

Above, $h_k^*(x, \lambda, q)$ denotes the convex conjugate of $h_k(x, \lambda, p)$ with respect to p.

Proof. See appendix. □

Discretization and convex representation of the data term. In practice, the label space Λ_k needs to be discretized at $N_k = |\Lambda_k|$ levels for each k. This means that instead of the indicator distributions $v^k(x, \lambda)$, we end up with a finite number of indicator functions v_λ^k for each discrete label. Optimization in v takes thus place over the set

$$D = \left\{ (v^k)_{1..d} \,\middle|\, v_\lambda^k \in \mathcal{SBV}(\Omega, \{0, 1\}) \;:\; \sum_{\lambda \in \Lambda_k} v_\lambda^k = 1 \right\}. \tag{8}$$

which is non-convex, since each v_λ^k must be binary. Ultimately, we wish to replace D with its convex hull $\mathrm{co}(D)$, which consists of functions taking values in the full range $[0, 1]$. Similarly to D, the discrete version of the set C_k in (7) consists of tuples $(p_\lambda, b_\lambda)_{\lambda \in \Lambda_k}$ of functions. Using the relationships (5) to transform the data term of the multilabel problem (1), we see that it is pointwise of the form

$$E_{\text{data}}(v) = \sum_{\gamma \in \Gamma} c_\gamma\, v_{\gamma_1}^1 \cdot \ldots \cdot v_{\gamma_d}^d \tag{9}$$

with $c_\gamma(x) := c(x, \gamma)$, which is non-convex due to the multiplication terms. In [9], the authors convexified the data term by replacing each of the monomials by its convex envelope. In the two-label case for example, this leads to the relaxation

$$R_{\text{previous}}(v) = \sum_{\gamma \in \Gamma} c_\gamma \max(0, v_{\gamma_1}^1 + v_{\gamma_2}^2 - 1). \tag{10}$$

However, this approach has the drawback that it introduces the trivial solution $v_\gamma^k = 1/N_k$ as the point-wise minimizer of the data term with value $R_{\text{previous}}(v) = 0$ when moving from D to $\mathrm{co}(D)$, i.e. from binary to continuous v. In [9], this problem was avoided by an additional smoothing, which increases the energy of the trivial solution and thus rules it out. Unfortunately, the smoothing leads to non-exact solutions and is thus not ideal. In this work we therefore propose a different relaxation of the data term, which does not suffer from this problem. We propose the relaxation

$$R_{\text{data}}(v) := \sup_{q \in \mathcal{Q}} \sum_{\lambda \in \Lambda_1} q_\lambda^1 v_\lambda^1 + \ldots + \sum_{\lambda \in \Lambda_d} q_\lambda^d v_\lambda^d \tag{11}$$

of $E_{\text{data}}(v)$, which is to be understood pointwise for each $x \in \Omega$. The additional dual variables $q = (q^k)_{k=1..d}$ range over the convex set

$$\mathcal{Q} := \big\{(q^k)_{k=1..d} \,|\, q^k : \Lambda_k \to \mathbb{R} \text{ s.t. for all } \gamma \in \Gamma, \; q^1_{\gamma_1} + \ldots + q^d_{\gamma_d} \leq c_\gamma\big\}. \tag{12}$$

One can easily see that the relaxation coincides with the original energy for binary functions. In addition, one can prove the following theorem, which shows that the relaxation of the data term has the correct pointwise minimizers, in contrast to the one proposed in [9]. This means that no smoothing is necessary and an exact minimization algorithm can be employed to obtain solutions. It is one of the key theoretical contributions of this paper.

Theorem 3.2. *Let $\hat{v} \in D$ be a binary (pointwise) minimizer of $E_{data}(v)$. Then $\hat{v}$ is also a minimizer of*

$$\min_{v \in co(D)} R_{data}(v). \tag{13}$$

In particular, $E_{data}(\hat{v}) = R_{data}(\hat{v}) = \inf_{\gamma\in\Gamma} c_\gamma$.

Proof. See appendix. □

Final saddle point problem. Summarizing, in order to transform the multilabel problem into the final form which we are going to solve, we formulate it in terms of the indicator functions v^k_λ and plug in the representation (6) for the regularizer and use the relaxation (11) of the data term. Using the relationships (5) to transform the data term, we end up with the following saddle point problem:

$$\min_{(v^k_\lambda)\in \text{co}(D)} \; \max_{(p^k,b^k)\in C_k} \int_\Omega \left[R_{\text{data}}(v) + \sum_{k,\lambda\in\Lambda_k} \big(-\operatorname{div}(p^k_\lambda) - b^k_\lambda\big)\, v^k_\lambda \right] \mathrm{d}x. \tag{14}$$

Note that we performed the relaxation to the convex hull $\text{co}(D)$, which means we optimize over non-binary functions taking values between 0 and 1. Thus, the problem is now fully convex, and can be solved with a first order primal-dual algorithm, as detailed in section 5 later. If the minimizer already lies in D, we have found the global optimum, otherwise we have to project the result from $\text{co}(D)$ back to the smaller set D and usually find a non-optimal solution. However, we can compute an optimality bound to estimate how far we are from the global optimum by comparing the global optimum of the convex functional to the projected solution in D.

4. Multilabel Regularizers

In this section, we will explore suitable choices of the regularizer, and how they fit within the proposed framework. In particular, we will see how our model can be specialized to the previous work in [14, 9], but we will also discuss additional regularizers which become possible.

Previous approaches as special cases. We first consider the special case of a *piecewise constant* labeling. This means that the approximate gradient ∇u^k is constant zero for all k. Then the regularizer J_k penalizes only the jumps of u and reduces to

$$J_k(u^k) = \int_{S_{u^k}} d_k\big((u^k)^-, (u^k)^+\big). \tag{15}$$

Applying theorem 3.1 leads to a convex representation of J_k, which we formulate in the following proposition in its discretized form.

Proposition 4.1. *The convex relaxation of (15) is given by*

$$J_k(u^k) = \sup_{p\in C_k} \sum_{\lambda\in\Lambda_k} \int_\Omega p_\lambda(x) \nabla u^k_\lambda(x) \tag{16}$$

with

$$C_k = \Big\{p : \Omega \times \Lambda_k \to \mathbb{R}^n \,|\, |p_\lambda(x) - p_\gamma(x)| \leq d_k(\lambda,\gamma)\Big\}.$$

In the following we show that the label metrications proposed in [14] for one-dimensional and in [9] for multi-dimensional label spaces appear as special cases of our framework. Assume that the metric d_k has an *Euclidean representation*, which means that each label $\lambda \in \Lambda_k$ is represented by a vector $a_\lambda \in \mathbb{R}^{N_r}$ with some $N_r \geq 1$ and that with these vectors $d_k(\lambda,\gamma) = |a_\lambda - a_\gamma|$.

Proposition 4.2. *Let the matrix A consist of the columns a_λ. Then*

$$J_k(u^k) \geq TV_v(Au^k), \tag{17}$$

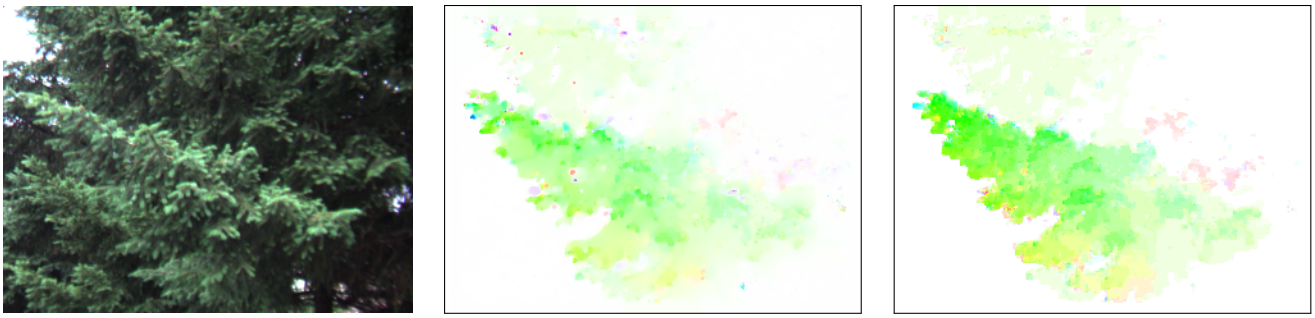
where

$$TV_v(\hat{u}) := \int_\Omega \sqrt{\sum_{i=1}^{N_r} |\nabla \hat{u}_i|^2}$$

denotes the vectorial total variation for functions taking values in $\mathbb{R}^{N_r}$. Equality holds if and only if u^k is binary.

Proof. The claim follows from our general formulation (16) by choosing a special form of the dual variables p together with additional relaxations of the equations in C_k, see appendix. □

The right hand side of inequality (17) is exactly the regularizer used in [9, 14]. This implies that for binary functions, the regularizers coincide, which can already be seen from representation (15), see [14]. However, if we perform the relaxation to functions taking values between 0 and 1,



One of two input images | Previous relaxation (25 minutes) | Proposed relaxation (1 minute)

Figure 3: *Optical flow fields with* 32×32 *labels computed on an image with resolution* 320×240*. With the new relaxation of the regularizers, we achieve optimality bounds which are on average three times lower than with previous relaxations from [9, 14], Here, we used total variation and in both cases the proposed data term relaxation* (11)*. Since the scaling of the regularity term is not directly comparable, we chose optimal parameters for both algorithms manually. The large time difference results from a narrow constraint on the time step for [14].*

inequality (17) implies that our relaxation is more tight, leading to solutions closer to the global optimum.

In addition, our method allows for much more complex regularizers allowing also a *smooth variation* of the labels which is not possible with the piecewise constant approach of [14, 9] which uses vectorial total variation. For instance, our formulation is capable of representing more sophisticated regularizers such as Huber-TV and the piecewise smooth Mumford-Shah functional, as we will show in the following paragraphs. For the regularizers presented in the remainder of this section, relaxations have previously been proposed for the case of a one-dimensional label space in [5, 16, 17]. However, our framework is more general and allows to combine them freely in the different label dimensions.

Huber-TV. The TV regularization is known to produce staircasing effects in the reconstruction, i.e. the solution will be piecewise constant. While this is natural in case of a discrete label space, for continuous label spaces it impedes smooth variations of the solution. A remedy for this is replacing the norm $|\nabla u|$ of the gradient by the Huber function

$$|\nabla u|_\alpha := \begin{cases} \frac{1}{2\alpha}|\nabla u|^2, & \text{if } |\nabla u| \leq \alpha \\ |\nabla u| - \frac{\alpha}{2}, & \text{else.} \end{cases} \tag{18}$$

which smoothes out the kink at the origin [17]. The Huber-TV regularizer is then defined by

$$J_k(u^k) = \int_\Omega |\nabla u^k(x)|_\alpha.$$

Theorem (3.1) gives a convex representation for J_k. The constraint set in (7) is found to be

$$C_k = \Big\{ (p,b) : \Omega \times \Lambda_k \to \mathbb{R}^n \times \mathbb{R} \,\Big|\, b_\lambda \geq \frac{\alpha}{2}|\partial_\lambda p_\lambda|^2, \quad |\partial_\lambda p_\lambda| \leq 1 \Big\}.$$

Piecewise smooth Mumford-Shah. The celebrated Mumford-Shah regularizer [1, 16]

$$J_k(u^k) = \int_{\Omega \setminus S_{u^k}} \frac{1}{2\alpha}|\nabla u|^2 \; + \; \nu |S_{u^k}|$$

allows to estimate a denoised image u which is *piecewise smooth*. Parameter ν can be used to easily control the total length of the jump set S_{u^k}. Bigger values of ν lead to a smaller jump set, i.e. the solution is smooth on wider subregions of Ω. The constraint set in the convex representation of theorem 3.1 becomes

$$C_k = \Big\{ (p,b) : \Omega \times \Lambda_k \to \mathbb{R}^n \times \mathbb{R} \,\Big|\, b_\lambda \geq \frac{\alpha}{2}|\partial_\lambda p_\lambda|^2, \quad |p_\lambda - p_\gamma| \leq \nu \Big\}.$$

The limiting case $\alpha = 0$ corresponds to the piecewise constant Mumford-Shah regularizer.

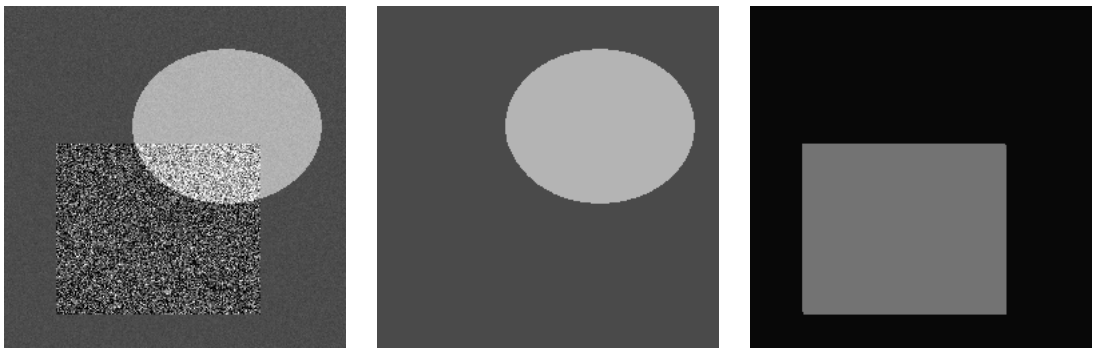
Truncated linear. For many applications, it is useful to consider two function values a and b just as "different" if their difference $|a - b|$ is big enough, i.e. jumps from a to b are penalized by a constant t regardless of how big $|b - a|$ is, if $|b - a|$ is greater than a certain threshold. Using linear penalization for small values this leads to the robust *truncated linear* regularizer [5]

$$J_k(u^k) = \int_{\Omega \setminus S_{u^k}} |\nabla u| + \int_{S_{u^k}} \min\left(t, |(u^k)^+ - (u^k)^-|\right).$$

The constraint set for this case is

$$C_k = \Big\{ (p,b) : \Omega \times \Lambda_k \to \mathbb{R}^n \times \mathbb{R} \,\Big|\, |\partial_\lambda p_\lambda| \leq 1, \quad |p_\lambda - p_\gamma| \leq t, \quad b = 0 \Big\}.$$

The second constraint must be imposed only for $\lambda, \gamma \in \Lambda_k$ with $|\lambda - \gamma| \geq t$, since otherwise it is already satisfied by the first constraint.



Noisy input | Denoised image | Reconstructed standard deviation

Figure 4: *The algorithm allows to both recover the unknown standard deviation σ of the noise as well as the intensity of a denoised image by solving a single optimization problem. Ground truth: Within rectangle Gaussian noise with standard deviation $\sigma = 0.25$, outside $\sigma = 0.02$; image intensity within ellipsoid $u = 0.7$, outside $u = 0.3$. Image resolution is 256×256 using 32×32 labels. Computation time is 10 minutes.*

5. Implementation

We minimize the energy (14) with a recent general fast primal-dual algorithm in [6]. It is essentially a gradient descent in v and gradient ascent in p and b, with a subsequent application of proximation operators which act as generalized reprojections. The proximation for the primal variable v consists of solving

$$\min_{v} \sum_{k,\lambda\in\Lambda_k} \frac{(v_\lambda^k - f_\lambda^k)^2}{2\theta} + R_{\text{data}}(v) \tag{19}$$

for each $x \in \Omega$ with some fixed $f : \Lambda_k \to \mathbb{R}$ and $\theta > 0$. Here, $R_{\text{data}}(v)$ is the data term relaxation in (11). This is a *pointwise* convex problem which can be solved in parallel for each $x \in \Omega$, again using the algorithm [6]. To make sure that the dual variables q of R_{data} lie in the constraint set (12), we add the Lagrange multiplier terms

$$\sup_{\mu_\gamma \geq 0} \sum_{\gamma\in\Gamma} \mu_\gamma (q_{\gamma_1}^1 + \ldots + q_{\gamma_d}^d - c_\gamma)$$

to the energy and optimize (19) over v and μ. Note that for each $x \in \Omega$ we thus need $\mathcal{O}(N_1 \cdots N_d)$ memory to solve (19), which at first *seems* to contradict our statement to substantially reduce the overall memory requirements to $\mathcal{O}((N_1 + \ldots + N_d)|\Omega|)$. However, the problems (19) are *independend* of each other for different $x \in \Omega$ and can be solved in chunks of $\mathcal{O}\big((N_1 + \ldots + N_d)|\Omega|/N_1 \cdots N_d\big)$ points x in parallel.

The proximation for the dual variables is the projection onto the constraint set C_k. We implement it by introducing new dual variables $d_\lambda = \partial_\lambda p_\lambda$ or $d_{\lambda,\gamma} = p_\lambda - p_\gamma$, respectively, depending on the kind of constraints in C_k, and adding the corresponding Lagrange multiplier terms $\inf_\eta \eta \cdot (\partial_\lambda p_\lambda - d_\lambda)$ respectively $\inf_\eta \eta \cdot (p_\lambda - p_\gamma - d_{\lambda,\gamma})$ to the energy to enforce these equalities. The optimization (14) is then performed over v, p, b and $\eta : \Omega \to \mathbb{R}^n$.

6. Experiments

We demonstrate the correctness and usability of our method on two examples. The first example shows computation of optic flow fields, the second example a denoising scheme where regularity is adapted optimally over the image. Several different regularizers are used in the examples. In the cases where the regularizer can be simulated with the previous relaxation [9], we compared the resulting optimality bounds. On average, our bounds were approximately three times better ($3 - 5\%$ with the proposed framework compared to $10 - 15\%$ with the previous relaxation). We used a parallel CUDA implementation on an nVidia GTX 480 GPU.

Optical flow. We compute optic flow between two color input images $I_0, I_1 : \Omega \to \mathbb{R}^3$ taken at two different time instants. The space of labels is two-dimensional, with $\Lambda_1 = \Lambda_2$ denoting the possible components of flow vectors in x and y-direction. Details on the data term can be found in [9]. Both directions are regularized either with total variation in figure 3, or a truncated linear penalizer in figure 1. Note that we can provide a tight relaxation of the exact penalizer, which was only coarsely approximated in the previous approaches [9, 14], while we can provide a tight relaxation of the exact penalizer.

Adaptive denoising. As a novel application of a multi-dimensional label space, we present adaptive denoising, where we *jointly* estimate a noise level and a denoised image by solving a single minimization problem. Note that here we require the continuous label space to represent the image intensity range. The Mumford-Shah energy can be interpreted as a denoising model which yields the maximum a posteriori estimate for the original image under the assumption that the input image f was distorted with Gaussian noise of standard deviation σ. If this standard deviation is itself viewed as an unknown which varies over the

image, the label space becomes two-dimensional, with one dimension representing the unknown intensity u of the original image, the second dimension representing the unknown standard deviation σ of the noise. The data term of the energy can then be written as [4]

$$\int_\Omega \frac{(u-f)^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2)\,\mathrm{d}x. \tag{20}$$

Results of the optimization can be observed in figure 4. For the regularizer, we used piecewise constant Mumford-Shah for both σ and u.

7. Conclusion

We proposed a general framework for convex relaxations of multilabel problems based on a continuous, multi-dimensional label space. The first main contribution is to improve upon the regularization by combining the advantages of the efficient multi-dimensional relaxation presented in [9] with the tight relaxation of [5]. The second main contribution is to introduce a much tighter relaxation for the products in the data term, which avoids the problem of a trivial pointwise solution without the need for additional smoothing as in [9]. In this way, we arrive at solutions which are provably closer to the global optimum. In addition, the framework allows to formulate more general regularizers on multi-dimensional label spaces and thus solve a more general class of problems efficiently. For example, we can explicitly encourage the solution to be smooth in certain regions, and can represent Huber-TV and truncated linear regularization by an exact and tight relaxation. In contrast to previous work, the regularizers can be arbitrarily mixed, in the sense that each dimension of the label space can have its own type of regularity.

Appendix

Proof of Theorem 3.1. We use the graph function $1_{u^k} : \Omega \times \Lambda_k \to \mathbb{R}$ of u^k, defined as

$$1_{u^k}(x,\lambda) = \begin{cases} 1, & \text{if } \lambda < u^k(x) \\ 0, & \text{else.} \end{cases}$$

Under the assumptions on h_k and d_k in the definition of J_k in section 2, the functional J_k can be represented in the following way [1]:

$$J_k(u^k) = \sup_{\phi \in K} \int_{\Omega\times\Lambda_k} \phi^x \cdot \nabla_x(1_{u^k}) + \phi^\lambda\, \partial_\lambda(1_{u^k}) \tag{21}$$

with the convex set

$$\begin{aligned} K = \Big\{ \phi = (\phi^x, \phi^\lambda) : \Omega\times\Lambda_k \to \mathbb{R}^n \times \mathbb{R} \,\big|\, & \\ \phi^\lambda(x,\lambda) \geq h_k^*(x,\lambda,\phi^x(x,\lambda)), & \\ \Big| \int_\lambda^\gamma \phi^x(x,s)\,\mathrm{d}s \Big| \leq d_k(x,\lambda,\gamma) \Big\}. & \end{aligned} \tag{22}$$

Representation (6) follows from (21) by noting that

$$1_{u^k}(x,\lambda) = \int_\lambda^{\lambda_{\max}} v_k(x,s)\,\mathrm{d}s \tag{23}$$

and defining $p : \Omega \times \Lambda_k \to \mathbb{R}^n$ and $b : \Omega \times \Lambda_k \to \mathbb{R}$ by

$$p(x,\lambda) := \int_{\lambda_{\min}}^\lambda \phi^x(x,s)\,\mathrm{d}s, \quad b(x,\lambda) := \phi^\lambda(x,\lambda).$$

Specifically, the second addend in (21) converts to

$$\phi^\lambda\, \partial_\lambda(1_{u^k}) = -b\, v_k. \tag{24}$$

Furthermore, from the definition of p it follows $\phi^x = \partial_\lambda p$ and we obtain using the relation (23):

$$\begin{aligned} &\int_{\Omega\times\Lambda_k} \phi^x \cdot \nabla_x(1_{u^k}) \\ &\quad = \int_\Omega \int_{\Lambda_k} \partial_\lambda p(x,\lambda) \cdot \Big(\int_\lambda^{\lambda_{\max}} \nabla v_k(x,s)\,\mathrm{d}s \Big)\,\mathrm{d}\lambda\,\mathrm{d}x \\ &\quad = \int_\Omega \int_{\Lambda_k} p(x,\lambda) \cdot \nabla v_k(x,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}x. \end{aligned} \tag{25}$$

Here, we applied integration by parts together with $p(x,\lambda_{\min}) = 0$. The claim in the proposition now follows directly from (21) using (24) and (25). □

Proof of Theorem 3.2. Let v be arbitrary, $\hat{c} = \inf_\gamma c_\gamma$ and set $q_\lambda^i := \hat{c}/d$. Then

$$\sum_{i=1}^d \sum_{\lambda\in\Lambda^i} q_\lambda^i v_\lambda^i = \sum_{i=1}^d \frac{\hat{c}}{d} \sum_{\lambda\in\Lambda^i} v_\lambda^i = d\frac{\hat{c}}{d} = \hat{c}, \tag{26}$$

and $\sum_i q_{\gamma_i}^i = \hat{c} \leq c_\gamma$ for all γ, so $q \in \mathcal{Q}$. This shows that $R(v) \geq \hat{c}$. Conversely, let $\hat{\gamma}$ be the label indicated by $\hat{v}$, i.e. where all $\hat{v}_{\hat{\gamma}_i} = 1$. Then $E(\hat{v}) = \hat{c}$ and for all $q \in \hat{C}$,

$$\sum_{i=1}^d \sum_{\lambda\in\Lambda^i} q_\lambda^i v_\lambda^i = \sum_{i=1}^d q_{\hat{\gamma}^i}^i v_{\hat{\gamma}^i}^i = \sum_{i=1}^d q_{\hat{\gamma}^i}^i \leq \hat{c}. \tag{27}$$

This shows $R(\hat{v}) = \hat{c}$ and completes the proof.

Proof of Proposition 4.1. We can enforce a piecewise constant labeling u, if we enforce the approximate gradient ∇u^k to be constant zero for all k. In (3), this can be achieved by setting $h(x,u^k(x),\nabla u^k(x)) = c\,|\nabla u^k|$ with a constant $c > 0$ and then letting $c \to \infty$ to enforce $\nabla u^k \equiv 0$ in $\Omega \setminus S_{u^k}$. Inserting the convex conjugate $h_k^*(x,\lambda,q) = \delta_{\{|q|\leq c\}}$, we find that the conditions in (7) now reduce to

$$\begin{aligned} b_\lambda(x) &\geq 0, \\ |\partial_\lambda p_\lambda(x)| &\leq c, \\ |p_\lambda(x) - p_\gamma(x)| &\leq d_k(\lambda,\gamma). \end{aligned} \tag{28}$$

The supremum over $b_\lambda(x) \geq 0$ is easily eliminated from (6) since $v_\lambda^k(x) \geq 0$, i.e. $-b_\lambda(x)v_\lambda^k(x) \leq 0$ with 0 being the maximum possible value. The second constraint in (28) follows from the third if we choose $c \geq \max_{\lambda>\gamma} \frac{d_k(\lambda,\gamma)}{|\lambda-\gamma|}$. Thus we arrive at (16) with the set C_k as claimed in the proposition. □

Proof of Proposition 4.2. We choose a special form for p_λ as

$$p_\lambda = \sum_{i=1}^{N_r} A_{i\lambda} q_i, \tag{29}$$

with $q : \Omega \times \{1, \dots, M_k\} \to \mathbb{R}^n$ such that $|q| \leq 1$ and the matrix $A : \mathbb{R}^{N_r \times N_k}$ of the euclidean representation of d_k. The constraint for p in (16) is then satisfied, since by the Cauchy-Schwarz inequality,

$$\begin{aligned} |p_\lambda - p_\gamma| &= \left| \sum_{i=1}^{N_r} (A_{i\lambda} - A_{i\gamma}) q_i \right| \\ &\leq \sqrt{\sum_{i=1}^{N_r} (A_{i\lambda} - A_{i\gamma})^2} \cdot \sqrt{\sum_{i=1}^{N_r} |q_i|^2} \\ &= |Ae_\lambda - Ae_\gamma| \, |q| \leq d_k(\lambda, \gamma). \end{aligned}$$

For the second inequality, we made use of the definition of A, which implies $|Ae_\lambda - Ae_\gamma| = d_k(\lambda, \gamma)$. Plugging (29) into (16) we obtain the desired result

$$\begin{aligned} J_k(u^k) &\geq \sup_{|q|\leq 1} \sum_{\lambda \in \Lambda_k} \int_\Omega \left(\sum_{i=1}^{M_k} A_{i\lambda} q_i \right) \cdot \nabla u_\lambda^k \, \mathrm{d}x \\ &= \sup_{|q|\leq 1} \int_\Omega \sum_{i=1}^{M_k} q_i \cdot \nabla \left(\sum_{\lambda \in \Lambda_k} A_{i\lambda} u_\lambda^k \right) \mathrm{d}x \\ &= \sup_{|q|\leq 1} \int_\Omega \sum_{i=1}^{M_k} q_i \cdot \nabla (Au^k)_i \, \mathrm{d}x = TV_{\mathrm{v}}(Au^k). \end{aligned}$$

The inequality in the first step is a consequence of choosing a the special form of p's, thus reducing the set over which the supremum is taken. This completes the proof. □

References

[1] G. Albert, G. Bouchitt, and G. D. Maso. The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calculus of Variations and Partial Differential Equations*, 16(3):299–333, 2002. 3, 5, 7

[2] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV Spaces*. MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2006. 2

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Intelligence*, 23(11):1222–1239, 2001. 2

[4] T. Brox and D. Cremers. On local region models and a statistical interpretation of the piecewise smooth Mumford-Shah functional. *International Journal of Computer Vision*, 84:184–193, 2009. 7

[5] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical Report TR-2008-05, Dept. of Computer Science, University of Bonn, 2008. 1, 2, 5, 7

[6] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *preprint*, 2010. 6

[7] D. Cremers and K. Kolev. Multiview stereo and silhouette consistency via convex functionals over convex domains. *IEEE Transactions on Pattern Analysis and Intelligence*, 2010. 1

[8] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab. Optical flow estimation with uncertainties through dynamic mrfs. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2008. 2

[9] B. Goldluecke and D. Cremers. Convex relaxation for multilabel problems with product label spaces. In *Proc. European Conference on Computer Vision*, 2010. 1, 2, 3, 4, 5, 6, 7

[10] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *J. Royal Statistics Soc.*, 51(Series B):271–279, 1989. 1

[11] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Intelligence*, 25(10):1333–1336, 2003. 1

[12] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *Proc. European Conference on Computer Vision*, pages 332–345, 2008. 2

[13] V. Kolmogorov and C. Rother. Minimizing non-submodular functions with graph cuts - a review. *IEEE Transactions on Pattern Analysis and Intelligence*, 29(7), 2007. 2

[14] J. Lellmann, , F. Becker, and C. Schnörr. Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In *Proc. International Conference on Computer Vision*, 2009. 1, 2, 3, 4, 5, 6

[15] M. Nikolova, S. Esedoglu, and T.Chan. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal of Applied Mathematics*, 66(5):1632–1648, 2006. 1, 2

[16] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In *Proc. International Conference on Computer Vision*, 2009. 5

[17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*, 2010. 1, 2, 5

[18] M. Wainwright, T. Jaakkola, and A. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Trans. Inf. Theory*, 51(11):3697–3717, 2005. 2

[19] C. Zach, D. Gallup, J. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization*, pages 243–252, 2009. 2