Sublabel-Accurate Discretization of Nonconvex Free-Discontinuity Problems

Thomas Möllenhoff Daniel Cremers
Technical University of Munich
{thomas.moellenhoff,cremers}@tum.de

Abstract

In this work we show how sublabel-accurate multilabeling approaches [15, 18] can be derived by approximating a classical label-continuous convex relaxation of nonconvex free-discontinuity problems. This insight allows to extend these sublabel-accurate approaches from total variation to general convex and nonconvex regularizations. Furthermore, it leads to a systematic approach to the discretization of continuous convex relaxations. We study the relationship to existing discretizations and to discrete-continuous MRFs. Finally, we apply the proposed approach to obtain a sublabel-accurate and convex solution to the vectorial Mumford-Shah functional and show in several experiments that it leads to more precise solutions using fewer labels.

1. Introduction

1.1. A class of continuous optimization problems

Many tasks particularly in low-level computer vision can be formulated as optimization problems over mappings $u : \Omega \to \Gamma$ between sets Ω and Γ. The energy functional is usually designed in such a way that the minimizing argument corresponds to a mapping with the desired solution properties. In classical discrete Markov random field (MRF) approaches, which we refer to as *fully discrete optimization*, Ω is typically a set of nodes (e.g., pixels or superpixels) and Γ a set of labels $\{1, \dots, \ell\}$.

However, in many problems such as image denoising, stereo matching or optical flow where $\Gamma \subset \mathbb{R}^d$ is naturally modeled as a continuum, this discretization into *labels* can entail unreasonably high demands in memory when using a fine sampling, or it leads to a strong label bias when using a coarser sampling, see Figure 1. Furthermore, as jump discontinuities are ubiquitous in low-level vision (e.g., caused by object edges, occlusion boundaries, changes in albedo, shadows, etc.), it is important to model them in a meaningful manner. By restricting either Ω or Γ to a discrete set, one loses the ability to mathematically distinguish between continuous and discontinuous mappings.

(a) (b)

Figure 1: The classical way to discretize continuous convex relaxations such as the vectorial Mumford-Shah functional [26] leads to solutions (**b**), top-left) with a strong bias towards the chosen labels (here an equidistant $5 \times 5 \times 5$ sampling of the RGB space). This can be seen in the bottom left part of the image, where the green color is truncated to the nearest label which is gray. The proposed sublabel-accurate approximation of the continuous relaxation leads to bias-free solutions (**b**), bottom-right).

Motivated by these two points we consider *fully-continuous* optimization approaches, where the idea is to postpone the discretization of $\Omega \subset \mathbb{R}^n$ and $\Gamma \subset \mathbb{R}$ as long as possible. The prototypical class of continuous optimization problems which we consider in this work are nonconvex free-discontinuity problems, inspired by the celebrated Mumford-Shah functional [4, 19]:

$$E(u) = \int_{\Omega \setminus J_u} f\left(x, u(x), \nabla u(x)\right) \mathrm{d}x + \int_{J_u} d\left(x, u^-(x), u^+(x), \nu_u(x)\right) \mathrm{d}\mathcal{H}^{n-1}(x). \tag{1}$$

The first integral is defined on the region $\Omega \setminus J_u$ where u is continuous. The integrand $f : \Omega \times \Gamma \times \mathbb{R}^n \to [0, \infty]$ can be thought of as a combined data term and regularizer, where the regularizer can penalize variations in terms of the (weak) gradient ∇u. The second integral is defined on the $(n-1)$-dimensional discontinuity set $J_u \subset \Omega$ and $d : \Omega \times \Gamma \times \Gamma \times \mathcal{S}^{n-1} \to [0, \infty]$ penalizes jumps from u^- to u^+ in unit direction ν_u. The appropriate function space for (1) are the *special functions of bounded variation*. These are

functions of bounded variation (cf. Section 2 for a defintion) whose distributional derivative Du can be decomposed into a continuous part and a jump part in the spirit of (1):

$$Du = \nabla u \cdot \mathcal{L}^n + \left(u^+ - u^-\right) \nu_u \cdot \mathcal{H}^{n-1} \llcorner J_u, \quad (2)$$

where $\mathcal{L}^n$ denotes the n-dimensional Lebesgue measure and $\mathcal{H}^{n-1} \llcorner J_u$ the $(n-1)$-dimensional Hausdorff measure restricted to the jump set J_u. For an introduction to functions of bounded variation and the study of existence of minimizers to (1) we refer the interested reader to [2].

Note that due to the possible nonconvexity of f in the first two variables a surprisingly large class of low-level vision problems fits the general framework of (1). While (1) is a difficult nonconvex optimization problem, the state-of-the-art are convex relaxations [1, 6, 9]. We give an overview of the idea behind the convex reformulation in Section 3.

Extensions to the vectorial setting, i.e., $\dim(\Gamma) > 1$, have been studied by Strekalovskiy *et al*. in various works [12, 26, 27] and recently using the theory of currents by Windheuser and Cremers [29]. The case when Γ is a manifold has been considered by Lellmann *et al*. [17]. These advances have allowed for a wide range of difficult vectorial and joint optimization problems to be solved within a convex framework.

1.2. Related work

The first practical implementation of (1) was proposed by Pock *et al*. [20], using a simple finite differencing scheme in both Ω and Γ which has remained the standard way to discretize convex relaxations. This leads to a strong label bias (see Figure 1b), top-left) *despite* the initially label-continuous formulation.

In the MRF community, a related approach to overcome this label-bias are *discrete-continuous* models (discrete Ω and continuous Γ), pioneered by Zach *et al*. [30, 31]. Most similar to the present work is the approach of Fix and Agarwal [11]. They derive the discrete-continuous approaches as a discretization of an infinite dimensional dual linear program. Their approach differs from ours, as we start from a different (nonlinear) infinite-dimensional optimization problem and consider a representation of the dual variables which enforces continuity. The recent work of Bach [3] extends the concept of submodularity from discrete to continuous Γ along with complexity estimates.

There are also *continuous-discrete* models, i.e. the range Γ is discretized into labels but Ω is kept continuous [10, 16]. Recently, these spatially continuous multilabeling models have been extended to allow for so-called *sublabel accurate* solutions [15, 18], i.e., solutions which lie between two labels. These are, however, limited to total variation regularization, due to the separate convexification of data term and regularizer. We show in this work that for general regularizers a joint convex relaxation is crucial.

Finally, while not focus of this work, there are of course also *fully-discrete* approaches, among many [14, 25, 28], which inspired some of the continuous formulations.

1.3. Contribution

In this work, we propose an approximation strategy for *fully-continuous* relaxations which retains continuous Γ even after discretization (see Figure 1b), bottom-right). We summarize our contributions as:

- We generalize the work [18] from total variation to general convex and nonconvex regularization.
- We prove (see Prop. 2 and Prop. 4) that different approximations to a convex relaxation of (1) give rise to existing relaxations [20] and [18]. We investigate the relationship to discrete-continuous MRFs in Prop. 5.
- On the example of the vectorial Mumford-Shah functional [26] we show that our framework yields also sublabel-accurate formulations of extensions to (1).

2. Notation and preliminaries

We denote the Iverson bracket as $[\![\cdot]\!]$. Indicator functions from convex analysis which take on values 0 and ∞ are denoted by $\delta\{\cdot\}$. We denote by f^* the convex conjugate of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set. For a function $u \in L^1(\Omega; \mathbb{R})$ its total variation is defined by

$$TV(u) = \sup \left\{ \int_\Omega u \operatorname{Div} \varphi \, \mathrm{d}x : \varphi \in C_c^1(\Omega; \mathbb{R}^n) \right\}. \quad (3)$$

The space of functions of bounded variation, i.e., for which $TV(u) < \infty$ (or equivalently for which the distributional derivative Du is a finite Radon measure) is denoted by $\mathrm{BV}(\Omega; \mathbb{R})$ [2]. We write $u \in \mathrm{SBV}(\Omega; \mathbb{R})$ for functions $u \in \mathrm{BV}(\Omega; \mathbb{R})$ whose distributional derivative admits the decomposition (2). For the rest of this work, we will make the following simplifying assumptions:

- The Lagrangian f in (1) is separable, i.e.,

  $$f(x, t, g) = \rho(x, t) + \eta(x, g), \quad (4)$$

  for possibly nonconvex $\rho : \Omega \times \Gamma \to \mathbb{R}$ and regularizers $\eta : \Omega \times \mathbb{R}^n \to \mathbb{R}$ which are convex in g.
- The jump regularizer in (1) is isotropic and induced by a concave function $\kappa : \mathbb{R}_{\geq 0} \to \mathbb{R}$:

  $$d(x, u^-, u^+, \nu_u) = \kappa(|u^- - u^+|) \|\nu_u\|_2, \quad (5)$$

  with $\kappa(a) = 0 \Leftrightarrow a = 0$.
- The range $\Gamma = [\gamma_1, \gamma_\ell] \subset \mathbb{R}$ is a compact interval.

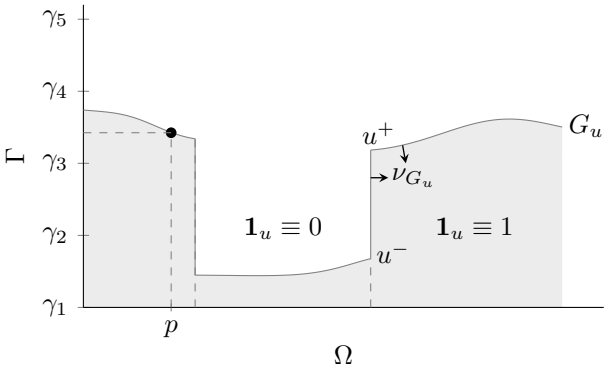


Figure 2: The central idea behind the convex relaxation for problem (1) is to reformulate the functional in terms of the complete graph $G_u \subset \Omega \times \Gamma$ of $u : \Omega \to \Gamma$ in the product space. This procedure is often referred to as "lifting", as one lifts the dimensionality of the problem.

3. The convex relaxation

In [1, 6, 9] the authors propose a convex relaxation for the problem (1). Their basic idea is to reformulate the energy (1) in terms of the *complete graph* of u, i.e. lifting the problem to one dimension higher as illustrated in Figure 2. The complete graph $G_u \subset \Omega \times \Gamma$ is defined as the (measure-theoretic) boundary of the characteristic function of the subgraph $\mathbf{1}_u : \Omega \times \mathbb{R} \to \{0, 1\}$ given by:

$$\mathbf{1}_u(x, t) = [\![t < u(x)]\!]. \tag{6}$$

Furthermore we denote the inner unit normal to $\mathbf{1}_u$ with ν_{G_u}. It is shown in [1] that for $u \in \mathrm{SBV}(\Omega; \mathbb{R})$ one has

$$E(u) = F(\mathbf{1}_u) = \sup_{\varphi \in \mathcal{K}} \int_{G_u} \langle \varphi, \nu_{G_u} \rangle \, \mathrm{d}\mathcal{H}^n, \tag{7}$$

with constraints on the dual variables $\varphi \in \mathcal{K}$ given by

$$\mathcal{K} = \Big\{ (\varphi_x, \varphi_t) \in C_c^1(\Omega \times \mathbb{R}; \mathbb{R}^n \times \mathbb{R}) : \\ \varphi_t(x, t) + \rho(x, t) \geq \eta^*(x, \varphi_x(x, t)), \tag{8}$$

$$\Big\| \int_t^{t'} \varphi_x(x, t) \mathrm{d}t \Big\|_2 \leq \kappa(|t - t'|), \forall t, t', \forall x \Big\}. \tag{9}$$

The functional (7) can be interpreted as the maximum flux of admissible vector fields $\varphi \in \mathcal{K}$ through the cut given by the complete graph G_u. The set $\mathcal{K}$ can be seen as capacity constraints on the flux field φ. This is reminiscent to constructions from the discrete optimization community [14]. The constraints (8) correspond to the first integral in (1) and the non-local constraints (9) to the jump penalization.

Using the fact that the distributional derivative of the subgraph indicator function $\mathbf{1}_u$ can be written as

$$D\mathbf{1}_u = \nu_{G_u} \cdot \mathcal{H}^m \llcorner G_u, \tag{10}$$

one can rewrite the energy (7) as

$$F(\mathbf{1}_u) = \sup_{\varphi \in \mathcal{K}} \int_{\Omega \times \Gamma} \langle \varphi, D\mathbf{1}_u \rangle. \tag{11}$$

A convex formulation is then obtained by relaxing the set of admissible primal variables to a convex set:

$$\mathcal{C} = \Big\{ v \in \mathrm{BV}_{\mathrm{loc}}(\Omega \times \mathbb{R}; [0, 1]) : \\ v(x, t) = 1 \ \ \forall t \leq \gamma_1, v(x, t) = 0 \ \ \forall t > \gamma_\ell, \tag{12}$$

$$v(x, \cdot) \text{ non-increasing} \Big\}.$$

This set can be thought of as the convex hull of the subgraph functions $\mathbf{1}_u$. The final optimization problem is then a convex-concave saddle point problem given by:

$$\inf_{v \in \mathcal{C}} \sup_{\varphi \in \mathcal{K}} \int_{\Omega \times \mathbb{R}} \langle \varphi, Dv \rangle. \tag{13}$$

In dimension one ($n = 1$), this convex relaxation is tight [8, 9]. For $n > 1$ global optimality can be guaranteed by means of a thresholding theorem in case $\kappa \equiv \infty$ [7, 21]. If the primal solution $\widehat{v} \in \mathcal{C}$ to (13) is binary, the global optimum u^* of (1) can be recovered simply by pointwise thresholding $\widehat{u}(x) = \sup\{t : \widehat{v}(x, t) > \frac{1}{2}\}$. If $\widehat{v}$ is not binary, in the general setting it is not clear how to obtain the global optimal solution from the relaxed solution. An a posteriori optimality bound to the global optimum $E(u^*)$ of (1) for the thresholded solution $\widehat{u}$ can be computed by:

$$|E(\widehat{u}) - E(u^*)| \leq |F(\mathbf{1}_{\widehat{u}}) - F(\widehat{v})|. \tag{14}$$

Using that bound, it has been observed that solutions are usually near globally optimal [26]. In the following section, we show how different discretizations of the continuous problem (13) lead to various existing lifting approaches and to generalizations of the recent sublabel-accurate continuous multilabeling approach [18].

4. Sublabel-accurate discretization

4.1. Choice of primal and dual mesh

In order to discretize the relaxation (13), we partition the range $\Gamma = [\gamma_1, \gamma_\ell]$ into $k = \ell - 1$ intervals. The individual intervals $\Gamma_i = [\gamma_i, \gamma_{i+1}]$ form a one dimensional *simplicial complex* (see e.g., [13]), and we have $\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_k$. The points $\gamma_i \in \Gamma$ are also referred to as *labels*. We assume that the labels are equidistantly spaced with label distance $h = \gamma_{i+1} - \gamma_i$. The theory generalizes also to non-uniformly spaced labels, as long as the spacing is homogeneous in Ω. Furthermore, we define $\gamma_0 = \gamma_1 - h$ and $\gamma_{\ell+1} = \gamma_\ell + h$.

The mesh for dual variables is given by *dual complex*, which is formed by the intervals $\Gamma_i^* = [\gamma_{i-1}^*, \gamma_i^*]$ with nodes $\gamma_i^* = \frac{\gamma_i + \gamma_{i+1}}{2}$. An overview of the notation and the considered finite dimensional approximations is given in Figure 3.

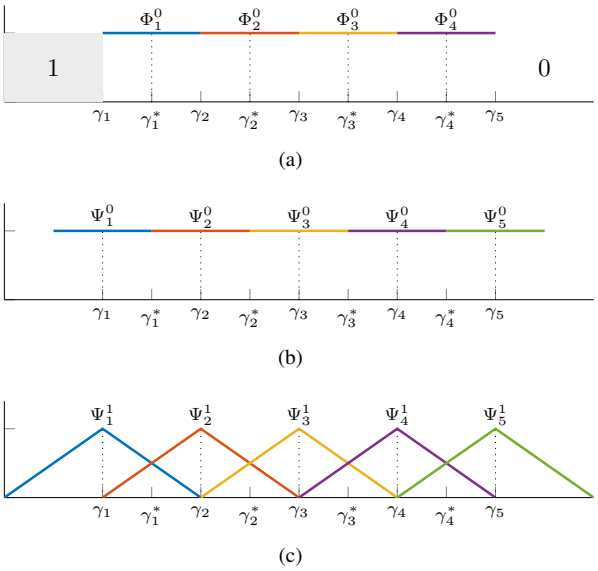


Figure 3: Overview of the notation and proposed finite dimensional approximation spaces.

4.2. Representation of the primal variable

As $\mathbf{1}_u$ is a discontinuous jump function, we consider a piecewise constant approximation for $v \in \mathcal{C}$,

$$\Phi_i^0(t) = [\![t \in \Gamma_i]\!],\ 1 \le i \le k, \tag{15}$$

see Figure 3a). Due to the boundary conditions in Eq. (12), we set v outside of Γ to 1 on the left and 0 on the right. Note that the non-decreasing constraint in $\mathcal{C}$ is implicitly realized as $\varphi_t \in \mathcal{K}$ can be arbitrarily large.

For coefficients $\hat{v} : \Omega \times \{1, \dots, k\} \to \mathbb{R}$ we have

$$v(x,t) = \sum_{i=1}^{k} \hat{v}(x,i)\Phi_i^0(t). \tag{16}$$

As an example of this representation, consider the approximation of $\mathbf{1}_u$ at point p shown in Figure 2:

$$\begin{aligned} \widehat{v}(p,\cdot) &= \sum_{i=1}^{k} e_i \int_\Gamma \Phi_i^0(t)\mathbf{1}_u(p,t)\mathrm{d}t \\ &= h \cdot \begin{bmatrix} 1 & 1 & 0.4 & 0 \end{bmatrix}^\top. \end{aligned} \tag{17}$$

This leads to the sublabel-accurate representation also considered in [18]. In that work, the representation from the above example (17) encodes a convex combination between the labels γ_3 and γ_4 with interpolation factor 0.4. Here it is motivated from a different perspective: we take a finite dimensional subspace approximation of the infinite dimensional optimization problem (13).

4.3. Representation of the dual variables

4.3.1 Piecewise constant φ_t

The simplest discretization of the dual variable φ_t is to pick a piecewise constant approximation on the dual intervals Γ_i^* as shown in Figure 3b): The functions are given by

$$\Psi_i^0(t) = [\![t \in \Gamma_i^*]\!],\ 1 \le i \le \ell, \tag{18}$$

As φ is a vector field in C_c^1, the functions Ψ vanish outside of Γ. For coefficient functions $\hat{\varphi}_t : \Omega \times \{1, \dots, \ell\} \to \mathbb{R}$ and $\hat{\varphi}_x : \Omega \times \{1, \dots, k\} \to \mathbb{R}^n$ we have:

$$\varphi_t(t) = \sum_{i=1}^{\ell} \hat{\varphi}_t(i)\Psi_i^0(t),\ \varphi_x(t) = \sum_{i=1}^{k} \hat{\varphi}_x(i)\Phi_i^0(t). \tag{19}$$

To avoid notational clutter, we dropped $x \in \Omega$ in (19) and will do so also in the following derivations. Note that for φ_x we chose the same piecewise constant approximation as for v, as we keep the model continuous in Ω, and ultimately discretize it using finite differences in x.

Discretization of the constraints In the following, we will plug in the finite dimensional approximations into the constraints from the set $\mathcal{K}$. We start by reformulating the constraints in (8). Taking the infimum over $t \in \Gamma_i$ they can be equivalently written as:

$$\inf_{t \in \Gamma_i} \varphi_t(t) + \rho(t) - \eta^*\left(\varphi_x(t)\right) \ge 0,\ 1 \le i \le \ell. \tag{20}$$

Plugging in the approximation (19) into the above leads to the following constraints for $1 \le i \le k$:

$$\begin{aligned} \hat{\varphi}_t(i) + \inf_{t \in [\gamma_i, \gamma_i^*]} \rho(t) &\ge \eta^*(\hat{\varphi}_x(i)), \\ \hat{\varphi}_t(i+1) + \underbrace{\inf_{t \in [\gamma_i^*, \gamma_{i+1}]} \rho(t)}_{\text{min-pooling}} &\ge \eta^*(\hat{\varphi}_x(i)). \end{aligned} \tag{21}$$

These constraints can be seen as min-pooling of the continuous unary potentials in a symmetric region centered on the label γ_i. To see that more easily, assume one-homogeneous regularization so that $\eta^* \equiv 0$ on its domain. Then two consecutive constraints from (21) can be combined into one where the infimum of ρ is taken over $\Gamma_i^* = [\gamma_i^*, \gamma_{i+1}^*]$ centered the label γ_i. This leads to capacity constraints for the flow in vertical direction $-\hat{\varphi}_t(i)$ of the form

$$-\hat{\varphi}_t(i) \le \inf_{t \in \Gamma_i^*} \rho(t),\ 2 \le i \le \ell - 1, \tag{22}$$

as well as similar constraints on $\hat{\varphi}_t(1)$ and $\hat{\varphi}_t(\ell)$. The effect of this on a nonconvex energy is shown in Figure 4 on the left. The constraints (21) are convex inequality constraints,

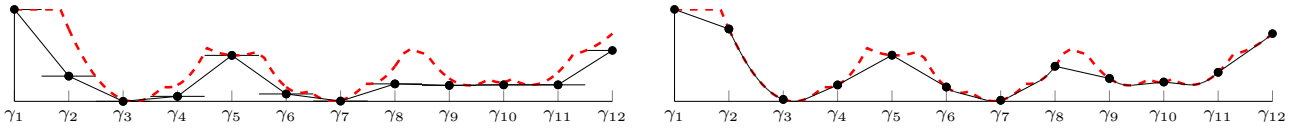


Figure 4: **Left:** piecewise constant dual variables φ_t lead to a linear approximation (shown in black) to the original cost function (shown in red). The unaries are determined through min-pooling of the continuous cost in the Voronoi cells around the labels. **Right:** continuous piecewise linear dual variables φ_t convexify the costs on each interval.

which can be implemented using standard proximal optimization methods and orthogonal projections onto the epigraph $\operatorname{epi}(\eta^*)$ as described in [21, Section 5.3].

For the second part of the constraint set (9) we insert again the finite-dimensional representation (19) to arrive at:

$$\begin{aligned}&\left\|(1-\alpha)\hat{\varphi}_x(i) + \sum_{l=i+1}^{j-1} \hat{\varphi}_x(l) + \beta\hat{\varphi}_x(j)\right\| \\ &\quad\leq \frac{\kappa(\gamma_j^\beta - \gamma_i^\alpha)}{h},\ \forall 1 \leq i \leq j \leq k, \alpha, \beta \in [0,1],\end{aligned} \tag{23}$$

where $\gamma_i^\alpha := (1-\alpha)\gamma_i + \alpha\gamma_{i+1}$. These are infinitely many constraints, but similar to [18] these can be implemented with finitely many constraints.

Proposition 1. *For concave $\kappa : \mathbb{R}_0^+ \to \mathbb{R}$ with $\kappa(a) = 0 \Leftrightarrow a = 0$, the constraints (23) are equivalent to*

$$\left\|\sum_{l=i}^{j} \hat{\varphi}_x(l)\right\| \leq \frac{\kappa(\gamma_{j+1} - \gamma_i)}{h},\ \forall 1 \leq i \leq j \leq k. \tag{24}$$

Proof. Proofs are given in the supplementary material. □

This proposition reveals that only information from the labels γ_i enters into the jump regularizer κ. For $\ell = 2$ we expect all regularizers to behave like the total variation.

Discretization of the energy For the discretization of the saddle point energy (13) we apply the divergence theorem

$$\int_{\Omega\times\mathbb{R}} \langle \varphi, Dv\rangle = \int_{\Omega\times\mathbb{R}} -\operatorname{Div}\varphi \cdot v \,\mathrm{d}t\,\mathrm{d}x, \tag{25}$$

and then discretize the divergence by inserting the piecewise constant representations of φ_t and v:

$$\begin{aligned}&\int_{\mathbb{R}} -\partial_t\varphi_t(t)v(t)\,\mathrm{d}t = \\ &-\hat{\varphi}_t(1) - \sum_{i=1}^{k} \hat{v}(i)\left[\hat{\varphi}_t(i+1) - \hat{\varphi}_t(i)\right].\end{aligned} \tag{26}$$

The discretization of the other parts of the divergence are given as the following:

$$\int_{\mathbb{R}} -\partial_{x_j}\varphi_x(t)v(t)\,\mathrm{d}t = -h\sum_{i=1}^{k} \partial_{x_j}\hat{\varphi}_x(i)\hat{v}(i), \tag{27}$$

where the spatial derivatives ∂_{x_j} are ultimately discretized using standard finite differences. It turns out that the above discretization can be related to the one from [20]:

Proposition 2. *For convex one-homogeneous η the discretization with piecewise constant φ_t and φ_x leads to the traditional discretization as proposed in [20], except with min-pooled instead of sampled unaries.*

4.3.2 Piecewise linear φ_t

As the dual variables in $\mathcal{K}$ are continuous vector fields, a more faithful approximation is given by a continuous piecewise linear approximation, given for $1 \leq i \leq \ell$ as:

$$\Psi_i^1(t) = \begin{cases} \frac{t-\gamma_{i-1}}{h}, & \text{if } t \in [\gamma_{i-1}, \gamma_i], \\ \frac{\gamma_{i+1}-t}{h}, & \text{if } t \in [\gamma_i, \gamma_{i+1}], \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

They are shown in Figure 3c), and we set:

$$\varphi_t(t) = \sum_{i=1}^{\ell} \hat{\varphi}_t(i)\Psi_i^1(t). \tag{29}$$

Note that the piecewise linear dual representation considered by Fix *et al.* in [11] differs in this point, as they do not ensure a continuous representation. Unlike the proposed approach their approximation does not take a true subspace of the original infinite dimensional function space.

Discretization of the constraints We start from the reformulation (20) of the original constraints (8). With (29) for φ_t and (19) for φ_x, we have for $1 \leq i \leq k$:

$$\begin{aligned}\inf_{t\in\Gamma_i}\ &\hat{\varphi}_t(i)\frac{\gamma_{i+1}-t}{h} + \hat{\varphi}_t(i+1)\frac{t-\gamma_i}{h} \\ &+ \rho(t) \geq \eta^*(\hat{\varphi}_x(i)).\end{aligned} \tag{30}$$

While the constraints (30) seem difficult to implement, they can be reformulated in a simpler way involving ρ^*.

Proposition 3. *The constraints (30) can be equivalently reformulated by introducing additional variables $a \in \mathbb{R}^k$, $b \in \mathbb{R}^k$, where $\forall i \in \{1, \dots, k\}$:*

$$\begin{aligned} &r(i) = (\hat{\varphi}_t(i) - \hat{\varphi}_t(i+1))/h, \\ &a(i) + b(i) - (\hat{\varphi}_t(i)\gamma_{i+1} - \hat{\varphi}_t(x, i+1)\gamma_i)/h = 0, \\ &r(i) \geq \rho_i^*\left(a(i)\right), \hat{\varphi}_x(i) \geq \eta^*\left(b(i)\right), \end{aligned} \tag{31}$$

with $\rho_i(x,t) = \rho(x,t) + \delta\{t \in \Gamma_i\}$.

The constraints (31) are implemented by projections onto the epigraphs of η^* and ρ_i^*, as they can be written as:

$$(r(i), a(i)) \in \mathsf{epi}(\rho_i^*),\ (\hat{\varphi}_x(i), b(i)) \in \mathsf{epi}(\eta^*). \tag{32}$$

Epigraphical projections for quadratic and piecewise linear ρ_i are described in [18]. In Section 5.1 we describe how to implement piecewise quadratic ρ_i. As the convex conjugate of ρ_i enters into the constraints, it becomes clear that this discretization only sees the *convexified* unaries on each interval, see also the right part of Figure 4.

Discretization of the energy It turns out that the piecewise linear representation of φ_t leads to the same discrete bilinear saddle point term as (26). The other term remains unchanged, as we pick the same representation of φ_x.

Relation to existing approaches In the following we point out the relationship of the approximation with piecewise linear φ_t to the sublabel-accurate multilabeling approaches [18] and the discrete-continuous MRFs [31].

Proposition 4. *The discretization with piecewise linear φ_t and piecewise constant φ_x, together with the choice $\eta(g) = \|g\|$ and $\kappa(a) = a$ is equivalent to the relaxation [18].*

Thus we extend the relaxation proposed in [18] to more general regularizations. The relaxation [18] was derived starting from a discrete label space and involved a separate relaxation of data term and regularizer. To see this, first note that the convex conjugate of a convex one-homogeneous function is the indicator function of a convex set [23, Corollary 13.2.1]. Then the constraints (8) can be written as

$$-\varphi_t(x,t) \leq \rho(x,t), \tag{33}$$

$$\varphi_x(x,t) \in \mathsf{dom}\{\eta^*\}, \tag{34}$$

where (33) is the data term and (34) the regularizer. This provides an intuition why the separate convex relaxation of data term and regularizer in [18] worked well. However, for general choices of η a joint relaxation of data term and regularizer as in (30) is crucial. The next proposition establishes the relationship between the data term from [31] and the one from [18].

Proposition 5. *The data term from [18] (which is in turn a special case of the discretization with piecewise linear φ_t) can be (pointwise) brought into the primal form*

$$\mathcal{D}(\widehat{v}) = \inf_{\substack{x_i \geq 0, \sum_i x_i = 1 \\ \widehat{v} = y/h + I^\top x}} \sum_{i=1}^{k} x_i \rho_i^{**}\left(\frac{y_i}{x_i}\right), \tag{35}$$

where $I \in \mathbb{R}^{k \times k}$ is a discretized integration operator.

The data term of Zach and Kohli [31] is precisely given by (35) except that the optimization is directly performed on $x, y \in \mathbb{R}^k$. The variable x can be interpreted as 1-sparse indicator of the interval Γ_i and $y \in \mathbb{R}^k$ as a sublabel offset. The constraint $\widehat{v} = y/h + I^\top x$ connects this representation to the subgraph representation $\widehat{v}$ via the operator $I \in \mathbb{R}^{k \times k}$ (see supplementary material for the definition). For general regularizers η, the discretization with piecewise linear φ_t differs from [18] as we perform a *joint convexification* of data term and regularizer and from [31] as we consider the spatially continuous setting. Another important question to ask is which primal formulation is actually optimized after discretization with piecewise linear φ_t. In particular the distinction between jump and smooth regularization only makes sense for continuous label spaces, so it is interesting to see what is optimized after discretizing the label space.

Proposition 6. *Let $\gamma = \kappa(\gamma_2 - \gamma_1)$ and $\ell = 2$. The approximation with piecewise linear φ_t and piecewise constant φ_x of the continuous optimization problem (13) is equivalent to*

$$\inf_{u:\Omega\to\Gamma} \int_\Omega \rho^{**}(x, u(x)) + (\eta^{**} \,\square\, \gamma\|\cdot\|)(\nabla u(x))\,\mathrm{d}x, \tag{36}$$

where $(\eta \,\square\, \gamma\|\cdot\|)(x) = \inf_y\ \eta(x-y) + \gamma\|y\|$ denotes the infimal convolution (cf. [23, Section 5]).

From Proposition 6 we see that the minimal discretization with $\ell = 2$ amounts to approximating problem (1) by globally convexifying the data term. Furthermore, we can see that Mumford-Shah (truncated quadratic) regularization ($\eta(g) = \alpha\|g\|^2$, $\kappa(a) \equiv \lambda[\![a > 0]\!]$) is approximated by a convex Huber regularizer in case $\ell = 2$. This is because the infimal convolution between x^2 and $|x|$ corresponds to the Huber function. While even for $\ell = 2$ this is a reasonable approximation to the original model (1), we can gradually increase the number of labels to get an increasingly faithful approximation of the original nonconvex problem.

4.3.3 Piecewise quadratic φ_t

For piecewise quadratic φ_t the main difficulty are the constraints in (20). For piecewise linear φ_t the infimum over a linear function plus ρ_i lead to (minus) the convex conjugate of ρ_i. Quadratic dual variables lead to so called generalized Φ-conjugates [24, Chapter 11L*, Example 11.66].

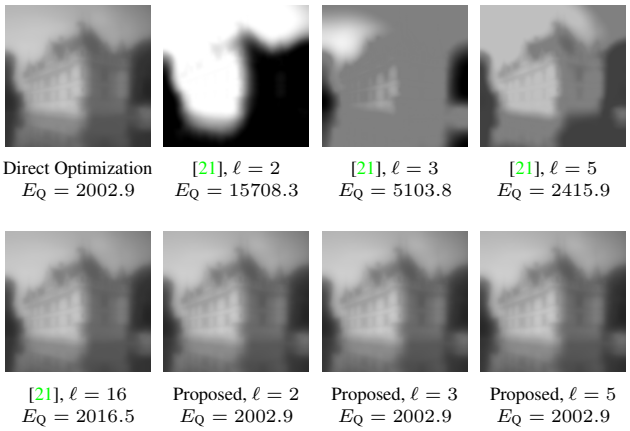


Figure 5: To verify the tightness of the approximation, we optimize a convex problem (quadratic data term with quadratic regularization). The discretization with piecewise linear φ_t recovers the exact solution with 2 labels and remains tight (numerically) for all $\ell > 2$, while the traditional discretization from [21] leads to a strong label bias.

Such conjugates were also theoretically considered in the recent work [11] for discrete-continuous MRFs, however an efficient implementation seems challenging. The advantage of this representation would be that one can avoid convexification of the unaries on each interval Γ_i and thus obtain a tighter approximation. While in principle the resulting constraints could be implemented using techniques from convex algebraic geometry and semi-definite programming [5] we leave this direction open to future work.

5. Implementation and extensions

5.1. Piecewise quadratic unaries ρ_i

In some applications such as robust fusion of depth maps, the data term ρ has a piecewise quadratic form:

$$\rho(u) = \sum_{m=1}^{M} \min\left\{\nu_m, \alpha_m (u - f_m)^2\right\}. \tag{37}$$

The intervals on which the above function is a quadratic are formed by the breakpoints $f_m \pm \sqrt{\nu_m/\alpha_m}$. In order to optimize this within our framework, we need to compute the convex conjugate of ρ on the intervals Γ_i, see Eq. (31). We can write the data term (37) on each Γ_i as

$$\min_{1\le j\le n_i} \underbrace{a_{i,j}u^2 + b_{i,j}u + c_{i,j} + \delta\{u \in I_{i,j}\}}_{=:\rho_{i,j}(u)}, \tag{38}$$

where n_i denotes the number of pieces and the intervals $I_{i,j}$ are given by the breakpoints and Γ_i. The convex conjugate is then given by $\rho_i^*(v) = \max_{1\le j\le n_i} \rho_{i,j}^*(v)$. As the epigraph of the maximum is the intersection of the epigraphs, $\mathsf{epi}(\rho_i^*) = \bigcap_{j=1}^{n_j} \mathsf{epi}\left(\rho_{i,j}^*\right)$, the constraints for the data term $(r^i, a^i) \in \mathsf{epi}(\rho_i^*)$, can be broken down:

$$(r^{i,j}, a^{i,j}) \in \mathsf{epi}\left(\rho_{i,j}^*\right), r^i = r^{i,j}, a^i = a^{i,j}, \forall j. \tag{39}$$

The projection onto the epigraphs of the $\rho_{i,j}^*$ are carried out as described in [18]. Such a convexified piecewise quadratic function is shown on the right in Figure 4.

5.2. The vectorial Mumford-Shah functional

Recently, the free-discontinuity problem (1) has been generalized to vectorial functions $u : \Omega \to \mathbb{R}^{n_c}$ by Strekalovskiy *et al.* [26]. The model they propose is

$$\sum_{c=1}^{n_c} \int_{\Omega\setminus J_u} f_c(x, u_c(x), \nabla_x u_c(x))\,\mathrm{d}x + \lambda\mathcal{H}^{n-1}(J_u), \tag{40}$$

which consists of a separable data term and separable regularization on the continuous part. The individual channels are coupled through the jump part regularizer $\mathcal{H}^{n-1}(J_u)$ of the joint jump set across all channels. Using the same strategy as in Section 4, applied to the relaxation described in [26, Section 3], a sublabel-accurate representation of the vectorial Mumford-Shah functional can be obtained.

5.3. Numerical solution

We solve the final finite dimensional optimization problem after finite-difference discretization in spatial direction using the primal-dual algorithm [20] implemented in the convex optimization framework `prost` [1].

6. Experiments

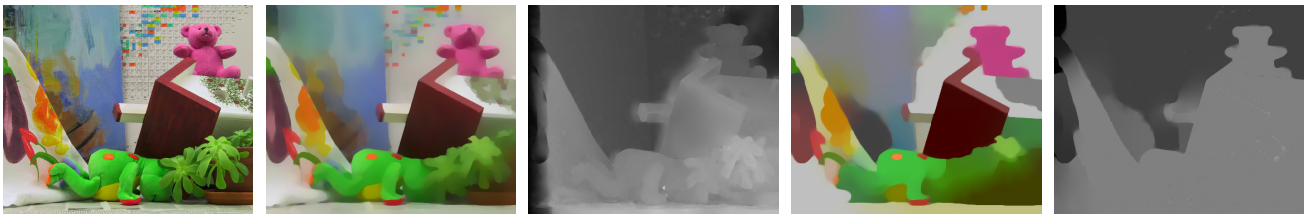
6.1. Exactness in the convex case

We validate our discretization in Figure 5 on the convex problem $\rho(u) = (u - f)^2$, $\eta(\nabla u) = \lambda|\nabla u|^2$. The global minimizer of the problem is obtained by solving $(I - \lambda\Delta)u = f$. For piecewise linear φ_t we recover the exact solution using only 2 labels, and remain (experimentally) exact as we increase the number of labels. The discretization from [21] shows a strong label bias due to the piecewise constant dual variable φ_t. Even with 16 labels their solution is different from the ground truth energy.

6.2. The vectorial Mumford-Shah functional

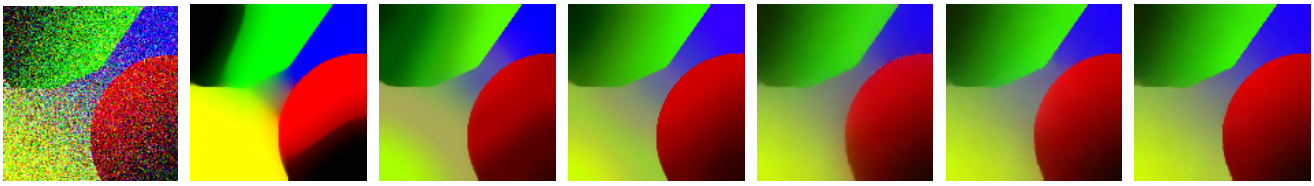
Joint depth fusion and segmentation We consider the problem of joint image segmentation and robust depth fusion from [22] using the vectorial Mumford-Shah functional from Section 5.2. The data term for the depth channel is given by (37), where f_m are the input depth hypotheses, α_m is a depth confidence and ν_m is a truncation parameter to be robust towards outliers. For the segmentation, we use

[1] https://github.com/tum-vision/prost



(a) Left input image (b) Proposed, (Segmentation) (c) Proposed, (Depth map) (d) [26], (Segmentation) (e) [26], (Depth map)

Figure 6: Joint segmentation and stereo matching. **b), c)** Using the proposed discretization we can arrive at smooth solutions using a moderate ($5 \times 5 \times 5 \times 5$) discretization of the 4-dimensional RGB-D label space. **d), e)** When using such a coarse sampling of the label space, the classical discretization used in [26] leads to a strong label bias. Note that with the proposed approach, a piecewise constant segmentation as in **d)** could also be obtained by increasing the smoothness parameter.



Noisy Input, (PSNR=10.4) | [26], $\ell = 2 \times 2 \times 2$ (PSNR=14.7) | [26], $\ell = 4 \times 4 \times 4$ (PSNR=25.0) | [26], $\ell = 6 \times 6 \times 6$ (PSNR=29.3) | Ours, $\ell = 2 \times 2 \times 2$, (PSNR=24.8) | Ours, $\ell = 4 \times 4 \times 4$, (PSNR=28.0) | Ours, $\ell = 6 \times 6 \times 6$, (PSNR=**30.0**)

Figure 7: Denoising of a synthetic piecewise smooth image degraded with 30% Gaussian noise. The standard discretization of the vectorial Mumford-Shah functional shows a strong bias towards the chosen labels (see also Figure 8), while the proposed discretization has no bias and leads to the highest overall peak signal to noise ratio (PSNR).

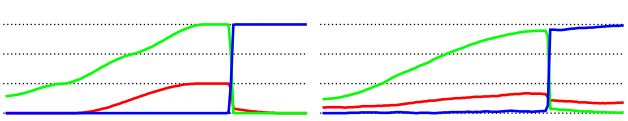


Figure 8: We show a 1D-slice through the resulting image in Figure 7 (with $\ell = 4 \times 4 \times 4$). The discretization [26] (left) shows a strong bias towards the labels, while the proposed discretization (right) yields a sublabel-accurate solution.

a quadratic difference dataterm in RGB space. For Figure 6 we computed multiple depth hypotheses f_m on a stereo pair using different matching costs (sum of absolute (gradient) differences, and normalized cross correlation) with varying patch radii (0 to 2). Even for a moderate label space of $5 \times 5 \times 5 \times 5$ we have no label discretization artifacts.

The piecewise linear approximation of the unaries in [26] leads to an almost piecewise constant segmentation of the image. To highlight the sublabel-accuracy of the proposed approach we chose a small smoothness parameter which leads to a piecewise smooth segmentation, but with a higher smoothness term or different choice of unaries a piecewise constant segmentation could also be obtained.

Piecewise-smooth approximations In Figure 7 we compare the discretizations for the vectorial Mumford-Shah functional. We see that the approach [26] shows strong label bias (see also Figure 8 and 1) while the discretiziation with piecewise linear duals leads to a sublabel-accurate result.

7. Conclusion

We proposed a framework to numerically solve *fully-continuous* convex relaxations in a sublabel-accurate fashion. The key idea is to implement the dual variables using a piecewise linear approximation. We prove that different choices of approximations for the dual variables give rise to various existing relaxations: in particular piecewise constant duals lead to the traditional lifting [20] (with min-pooling of the unary costs), whereas piecewise linear duals lead to the sublabel lifting that was recently proposed for total variation regularized problems [18]. While the latter method is not easily generalized to other regularizers due to the separate convexification of data term and regularizer, the proposed representation generalizes to arbitrary convex and non-convex regularizers such as the scalar and the vectorial Mumford-Shah problem. The proposed approach provides a systematic technique to derive sublabel-accurate discretizations for continuous convex relaxation approaches, thereby boosting their memory and runtime efficiency for challenging large-scale applications.

References

[1] G. Alberti, G. Bouchitté, and G. Dal Maso. The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Partial Differential Equations*, 16(3):299–333, 2003. 2, 3

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, USA, 2000. 2

[3] F. Bach. Submodular functions: from discrete to continous domains. *arXiv:1511.00394*, 2015. 2

[4] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987. 1

[5] G. Blekherman, P. A. Parrilo, and R. R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry*. SIAM, 2012. 7

[6] G. Bouchitté. Recent convexity arguments in the calculus of variations. *Lecture notes from the 3rd Int. Summer School on the Calculus of Variations, Pisa*, 1998. 2, 3

[7] G. Bouchitté and I. Fragalà. Duality for non-convex variational problems. *Comptes Rendus Mathematique*, 353(4):375–379, 2015. 3

[8] M. Carioni. A discrete coarea-type formula for the Mumford-Shah functional in dimension one. *arXiv preprint arXiv:1610.01846*, 2016. 3

[9] A. Chambolle. Convex representation for lower semicontinuous envelopes of functionals in L^1. *J. Convex Anal.*, 8(1):149–170, 2001. 2, 3

[10] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *SIAM J. Imaging Sciences*, 5(4):1113–1158, 2012. 2

[11] A. Fix and S. Agarwal. Duality and the continuous graphical model. In *Proceedings of the European Conference on Computer Vision, ECCV*, 2014. 2, 5, 7

[12] B. Goldluecke, E. Strekalovskiy, and D. Cremers. Tight convex relaxations for vector-valued labeling. *SIAM J. Imaging Sciences*, 6(3):1626–1664, 2013. 2

[13] A. N. Hirani. *Discrete exterior calculus*. PhD thesis, California Institute of Technology, 2003. 3

[14] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, 2003. 2, 3

[15] E. Laude, T. Möllenhoff, M. Moeller, J. Lellmann, and D. Cremers. Sublabel-accurate convex relaxation of vectorial multilabel energies. In *Proceedings of the European Conference on Computer Vision, ECCV*, 2016. 1, 2

[16] J. Lellmann and C. Schnörr. Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sciences*, 4(4):1049–1096, 2011. 2

[17] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. In *Proceedings of the IEEE International Conference on Computer Vision, ICCV*, 2013. 2

[18] T. Möllenhoff, E. Laude, M. Moeller, J. Lellmann, and D. Cremers. Sublabel-accurate relaxation of nonconvex energies. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2016. 1, 2, 3, 4, 5, 6, 7, 8

[19] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42(5):577–685, 1989. 1

[20] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In *Proceedings of the IEEE International Conference on Computer Vision, ICCV*, 2009. 2, 5, 7, 8

[21] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM J. Imaging Sci.*, 3(4):1122–1145, 2010. 3, 5, 7

[22] T. Pock, C. Zach, and H. Bischof. Mumford-Shah meets stereo: Integration of weak depth hypotheses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2007. 7

[23] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1996. 6

[24] R. T. Rockafellar, R. J.-B. Wets, and M. Wets. *Variational analysis*. Springer, 1998. 6

[25] M. Schlesinger. Sintaksicheskiy analiz dvumernykh zritelnikh signalov v usloviyakh pomekh (Syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika*, 4:113–130, 1976. 2

[26] E. Strekalovskiy, A. Chambolle, and D. Cremers. A convex representation for the vectorial Mumford-Shah functional. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2012. 1, 2, 3, 7, 8

[27] E. Strekalovskiy, A. Chambolle, and D. Cremers. Convex relaxation of vectorial problems with coupled regularization. *SIAM J. Imaging Sciences*, 7(1):294–336, 2014. 2

[28] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(7):1165–1179, 2007. 2

[29] T. Windheuser and D. Cremers. A convex solution to spatially-regularized correspondence problems. In *Proceedings of the European Conference on Computer Vision, ECCV*, 2016. 2

[30] C. Zach. Dual decomposition for joint discrete-continuous optimization. In *Proceedings of the International Conference on Artificial Intelligence and Statistics, AISTATS*, 2013. 2

[31] C. Zach and P. Kohli. A convex discrete-continuous approach for Markov random fields. In *Proceedings of the European Conference on Computer Vision, ECCV*, 2014. 2, 6