Sublabel–Accurate Relaxation of Nonconvex Energies

Thomas Möllenhoff[*]
TU München
moellenh@in.tum.de

Emanuel Laude[*]
TU München
laudee@in.tum.de

Michael Moeller
TU München
moellerm@in.tum.de

Jan Lellmann
University of Lübeck
lellmann@mic.uni-luebeck.de

Daniel Cremers
TU München
cremers@tum.de

Abstract

We propose a novel spatially continuous framework for convex relaxations based on functional lifting. Our method can be interpreted as a sublabel–accurate solution to multilabel problems. We show that previously proposed functional lifting methods optimize an energy which is linear between two labels and hence require (often infinitely) many labels for a faithful approximation. In contrast, the proposed formulation is based on a piecewise convex approximation and therefore needs far fewer labels – see Fig. 1. In comparison to recent MRF-based approaches, our method is formulated in a spatially continuous setting and shows less grid bias. Moreover, in a local sense, our formulation is the tightest possible convex relaxation. It is easy to implement and allows an efficient primal-dual optimization on GPUs. We show the effectiveness of our approach on several computer vision problems.

1. Introduction

Energy minimization methods have become the central paradigm for solving practical problems in computer vision. The energy functional can often be written as the sum of a data fidelity and a regularization term. One of the most popular regularizers is the total variation (TV) due to its many favorable properties [4]. Hence, an important class of optimization problems is given as

$$\min_{u:\Omega\to\Gamma} \int_\Omega \rho(x, u(x))\,\mathrm{d}x + \lambda\, TV(u), \qquad (1)$$

---

[*]Those authors contributed equally.
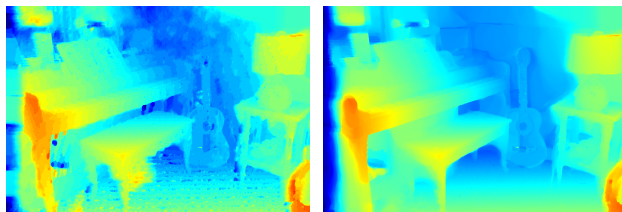
Pock *et al.* [17], 48 labels, 1.49 GB, 52s.   Proposed, 8 labels, 0.49 GB, 30s.

Figure 1. We propose a convex relaxation for the variational model (1), which opposed to existing functional lifting methods [17, 18] allows continuous label spaces *even after* discretization. Our method (here applied to stereo matching) avoids label space discretization artifacts, while saving on memory and runtime.

defined for functions u with finite total variation, arbitrary, possibly nonconvex dataterms $\rho : \Omega \times \Gamma \to \mathbb{R}$, label spaces Γ which are closed intervals in $\mathbb{R}$, $\Omega \subset \mathbb{R}^d$, and $\lambda \in \mathbb{R}^+$. The multilabel interpretation of the dataterm is that $\rho(x, u(x))$ represents the costs of assigning label $u(x)$ to point x. For (weakly) differentiable functions $TV(u)$ equals the integral over the norm of the derivative, and therefore favors a spatially coherent label configuration. The difficultly of minimizing the nonconvex energy (1) has motivated researchers to develop convex reformulations.

Convex representations of (1) and more general related energies have been studied in the context of the calibration method for the Mumford-Shah functional [1]. Based on these works, relaxations for the piecewise constant [15] and piecewise smooth Mumford-Shah functional [16] have been proposed. Inspired by Ishikawa's graph-theoretic globally

optimal solution to discrete variants of (1), continuous analogues have been considered by Pock *et al.* in [17, 18]. Continuous relaxations for multilabeling problems with finite label spaces Γ have also been studied in [11].

Interestingly, the discretization of the aforementioned continuous relaxations is very similar to the linear programming relaxations proposed for MAP inference in the Markov Random Field (MRF) community [10, 22, 24, 26]. Both approaches ultimately discretize the range Γ into a finite set of labels. A closer analysis of these relaxations reveals, however, that they are not well-suited to represent the continuous valued range that we face in most computer vision problems such as stereo matching or optical flow. More specifically, the above relaxations are not designed to assign meaningful cost values to non-integral configurations. As a result, a large number of labels is required to achieve a faithful approximation. Solving real-world vision problems therefore entails large optimization problems with high memory and runtime requirement. To address this problem, Zach and Kohli [27], Zach [25] and Fix and Agarwal [7] introduced MRF-based approaches which retain continuous label spaces after discretization. For manifold-valued labels, this issue was addressed by Lellmann *et al.* [12], however with the sole focus on the regularizer.

1.1. Contributions

We propose the first sublabel–accurate convex relaxation of nonconvex problems in a spatially continuous setting. It exhibits several favorable properties:

- In contrast to existing spatially continuous lifting approaches [17, 18], the proposed method provides substantially better solutions with far fewer labels – see Fig. 1. This provides savings in runtime and memory.

- In Sec. 3 we show that the functional lifting methods [17, 18] are a special case of the proposed framework.

- In Sec. 3 we show that, in a local sense, our formulation is the tightest convex relaxation which takes dataterm and regularizer into account separately. It is unknown whether this "local convex envelope" property also holds for the discrete approach [27].

- Our formulation is compact and requires only half the amount of variables for the dataterm than the formulation in [27]. We prove that the sublabel–accurate total
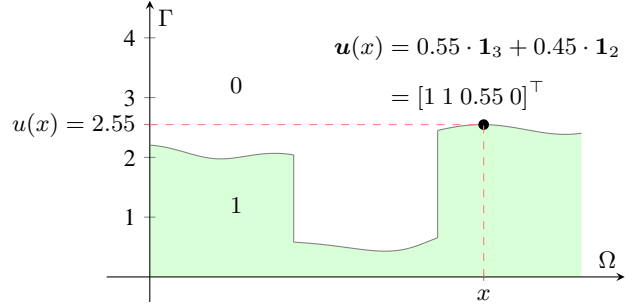

Figure 2. Lifted representation. Instead of optimizing over the function $u : \Omega \to \Gamma$, we optimize over all possible graph functions (here shaded in green) on $\Omega \times \Gamma$. The main idea behind our approach is the finite dimensional representation of the graph at every $x \in \Omega$ by means of $\boldsymbol{u} : \Omega \to \mathbb{R}^k$ (here $k = 4$).

variation can be represented in a very simple way, introducing no overhead compared to [17, 18]. In contrast, the regularizer in [27] is much more involved.

- Since our method is derived in a spatially continuous setting, the proposed approach easily allows different gradient discretizations. In contrast to [25, 27] the regularizer is isotropic leading to noticeably less grid bias.

2. Notation and Mathematical Preliminaries

We make heavy use of the convex conjugate, which is given as $f^*(y) = \sup_{x \in \mathbb{R}^n} \langle y, x \rangle - f(x)$ for functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. The biconjugate f^{**} denotes its *convex envelope*, i.e. the largest lower-semicontinuous convex under-approximation of f. For a set C we denote by δ_C the function which maps any element from C to 0 and is ∞ otherwise. For a comprehensive introduction to convex analysis, we refer the reader to [19]. Vector valued functions $\boldsymbol{u} : \Omega \to \mathbb{R}^k$ are written in bold symbols. If it is clear from the context, we will drop the $x \in \Omega$ inside the functions, e.g., we write $\rho(u)$ for $\rho(x, u(x))$, or α for $\alpha(x)$.

3. Functional Lifting

To derive a convex representation of (1), we rely on the framework of functional lifting. The idea is to reformulate the optimization problem in a higher dimensional space. We numerically show in Sec. 5 that considering the convex envelope of the dataterm and regularizer in this higher dimensional space leads to a better approximation of the original nonconvex energy. We start by sampling the range
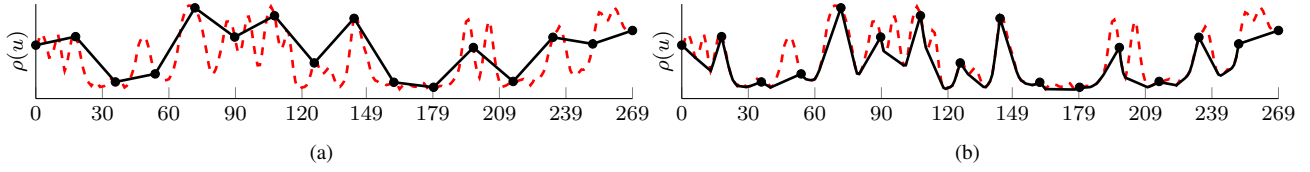
Figure 3. We show the nonconvex energy $\rho(u)$ at a fixed point $x \in \Omega$ (red dashed line in both plots) from the stereo matching experiment in Fig. 9 over the full range of 270 disparities. The black dots indicate the positions of the labels and the black curves show the approximations used by the respective methods. Fig. 3a: The baseline lifting method [17] uses a piecewise linear approximation with labels as nodes. Fig. 3b: The proposed method uses an optimal piecewise convex approximation. As we can see, the piecewise convex approximation is closer to the original nonconvex energy and therefore more accurate.

Γ at $L = k + 1$ labels $\gamma_1 < \ldots < \gamma_L \in \Gamma$. This partitions the range into k intervals $\Gamma_i = [\gamma_i, \gamma_{i+1}]$ so that $\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_k$. For any value in the range of $u : \Omega \to \Gamma$ there exist a label index $1 \le i \le k$ and $\alpha \in [0, 1]$ such that

$$u(x) = \gamma_i^\alpha := \gamma_i + \alpha(\gamma_{i+1} - \gamma_i). \tag{2}$$

We represent a value in the range Γ by a vector in $\mathbb{R}^k$

$$\boldsymbol{u}(x) = \boldsymbol{1}_i^\alpha := \alpha \boldsymbol{1}_i + (1 - \alpha) \boldsymbol{1}_{i-1}, \tag{3}$$

where $\boldsymbol{1}_i$ denotes a vector starting with i ones followed by $k - i$ zeros. We call $\boldsymbol{u} : \Omega \to \mathbb{R}^k$ the *lifted* representation of u, representing the graph of u. This notation is depicted in Fig. 2 for $k = 4$. Back-projecting the lifted $\boldsymbol{u}(x)$ to the range of u using the layer cake formula yields a one-to-one correspondence between $u(x) = \gamma_i^\alpha$ and $\boldsymbol{u}(x) = \boldsymbol{1}_i^\alpha$ via

$$u(x) = \gamma_1 + \sum_{i=1}^k \boldsymbol{u}_i(x)(\gamma_{i+1} - \gamma_i). \tag{4}$$

We write problem (1) in terms of such graph functions, a technique that is used in the theory of Cartesian currents [8].

3.1. Convexification of the Dataterm

For now, we consider a fixed $x \in \Omega$. Then the dataterm from (1) is a possibly nonconvex real-valued function (cf. Fig. 3) that we seek to minimize over a compact interval Γ:

$$\min_{u \in \Gamma} \rho(u). \tag{5}$$

Due to the one-to-one correspondence between γ_i^α and $\boldsymbol{1}_i^\alpha$ it is clear that solving problem (5) is equivalent to finding a minimizer of the lifted energy:

$$\boldsymbol{\rho}(\boldsymbol{u}) = \min_{1 \le i \le k} \boldsymbol{\rho}_i(\boldsymbol{u}), \tag{6}$$

$$\boldsymbol{\rho}_i(\boldsymbol{u}) = \begin{cases} \rho(\gamma_i^\alpha), & \text{if } \boldsymbol{u} = \boldsymbol{1}_i^\alpha, \ \alpha \in [0, 1], \\ \infty, & \text{else.} \end{cases} \tag{7}$$

Note that the constraint in (7) is essentially the nonconvex special ordered set of type 2 (SOS2) constraint [3]. More precisely, we demand that the "derivative" in label direction $(\partial_\gamma \boldsymbol{u})_i := \boldsymbol{u}_{i+1} - \boldsymbol{u}_i$ is zero, except for two neighboring elements, which add up to one. In the following proposition, we derive the tightest convex relaxation of $\boldsymbol{\rho}$.

Proposition 1. *The convex envelope of* (6) *is given as:*

$$\boldsymbol{\rho}^{**}(\boldsymbol{u}) = \sup_{\boldsymbol{v} \in \mathbb{R}^k} \langle \boldsymbol{u}, \boldsymbol{v} \rangle - \max_{1 \le i \le k} \boldsymbol{\rho}_i^*(\boldsymbol{v}), \tag{8}$$

where the conjugate of the individual $\boldsymbol{\rho}_i$ is

$$\boldsymbol{\rho}_i^*(\boldsymbol{v}) = c_i(\boldsymbol{v}) + \rho_i^* \left(\frac{\boldsymbol{v}_i}{\gamma_{i+1} - \gamma_i} \right), \tag{9}$$

with $c_i(\boldsymbol{v}) = \langle \boldsymbol{1}_{i-1}, \boldsymbol{v} \rangle - \frac{\gamma_i}{\gamma_{i+1} - \gamma_i} \boldsymbol{v}_i$ and $\rho_i = \rho + \delta_{\Gamma_i}$.

Proof. See supplementary material. $\square$

The above proposition reveals that the convex relaxation implicitly convexifies the dataterm ρ on each interval Γ_i. The equality $\rho_i^* = \rho_i^{***}$ implies that starting with ρ_i yields exactly the same convex relaxation as starting with ρ_i^{**}.

Corollary 1. *If ρ is linear on each Γ_i, then the convex envelopes of $\boldsymbol{\rho}(\boldsymbol{u})$ and $\boldsymbol{\sigma}(\boldsymbol{u})$ coincide, where the latter is:*

$$\boldsymbol{\sigma}(\boldsymbol{u}) = \begin{cases} \rho(\gamma_i^\alpha), & \text{if } \exists i : \ \boldsymbol{u} = \boldsymbol{1}_i^\alpha, \ \alpha \in \{0, 1\}, \\ \infty, & \text{else.} \end{cases} \tag{10}$$

Proof. Consider an additional constraint $\delta_{\{\gamma_i, \gamma_{i+1}\}}$ for each ρ_i, which corresponds to selecting $\alpha \in \{0, 1\}$ in (7). The fact that our relaxation is independent of whether we choose ρ_i or ρ_i^{**}, along with the fact that the convex hull of two points is a line, yields the assertion. $\square$

For the piecewise linear case, it is possible to find an explicit form of the biconjugate.

Proposition 2. *Let us denote by $\boldsymbol{r} \in \mathbb{R}^k$ the vector with*

$$\boldsymbol{r}_i = \rho(\gamma_{i+1}) - \rho(\gamma_i), \quad 1 \leq i \leq k. \tag{11}$$

Under the assumptions of Prop. 1, one obtains:

$$\boldsymbol{\sigma}^{**}(\boldsymbol{u}) = \begin{cases} \rho(\gamma_1) + \langle \boldsymbol{u}, \boldsymbol{r} \rangle, & \text{if } \boldsymbol{u}_i \geq \boldsymbol{u}_{i+1}, \boldsymbol{u}_i \in [0,1], \\ \infty, & \text{else.} \end{cases} \tag{12}$$

Proof. See supplementary material. □

Up to an offset (which is irrelevant for the optimization), one can see that (12) coincides with the dataterm of [15], the discretizations of [17, 18], and – after a change of variable – with [11]. This not only proves that the latter is optimizing a convex envelope, but also shows that our method naturally generalizes the work from piecewise linear to arbitrary piecewise convex energies. Fig. 3a and Fig. 3b illustrate the difference of $\boldsymbol{\sigma}^{**}$ and $\boldsymbol{\rho}^{**}$ on the example of a nonconvex stereo matching cost.

Because our method allows arbitrary convex functions on each Γ_i, we can prove that, for the two label case, our approach optimizes the convex envelope of the dataterm.

Proposition 3. *In the case of binary labeling, i.e., $L = 2$, the convex envelope of (6) reduces to*

$$\boldsymbol{\rho}^{**}(\boldsymbol{u}) = \rho^{**}\left(\gamma_1 + \boldsymbol{u}(\gamma_2 - \gamma_1)\right), \text{ with } \boldsymbol{u} \in [0,1]. \tag{13}$$

Proof. See supplementary material. □

3.2. A Lifted Representation of the Total Variation

We now want to find a lifted convex formulation that emulates the total variation regularization in (1). We follow [5] and define an appropriate integrand of the functional

$$TV(\boldsymbol{u}) = \int_\Omega \boldsymbol{\Phi}(x, D\boldsymbol{u}), \tag{14}$$

where the distributional derivative $D\boldsymbol{u}$ is a finite $\mathbb{R}^{k \times d}$-valued Radon measure [2]. We define

$$\boldsymbol{\Phi}(\boldsymbol{g}) = \min_{1 \leq i \leq j \leq k} \boldsymbol{\Phi}_{i,j}(\boldsymbol{g}). \tag{15}$$

The individual $\boldsymbol{\Phi}_{i,j} : \mathbb{R}^{k \times d} \to \mathbb{R} \cup \{\infty\}$ are given by:

$$\boldsymbol{\Phi}_{i,j}(\boldsymbol{g}) = \begin{cases} \left|\gamma_i^\alpha - \gamma_j^\beta\right| \cdot |\nu|_2, & \text{if } \boldsymbol{g} = (\mathbf{1}_i^\alpha - \mathbf{1}_j^\beta)\,\nu^\mathsf{T}, \\ \infty, & \text{else,} \end{cases} \tag{16}$$

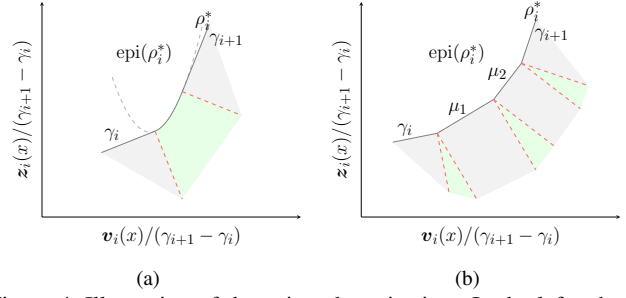Figure 4. Illustration of the epigraph projection. In the left subfigure the projection onto the epigraph of the conjugate of a convex quadratic ρ_i is shown. In the right subfigure the piecewise linear case is illustrated. In the both cases all points that lie in the gray sets are orthogonally projected onto the respective linear parts whereas the points that lie in the green sets are projected onto the parabolic part (in the quadratic case) respectively the kinks (in the piecewise linear case). In the piecewise linear case the green sets are normal cones. The red dashed lines correspond to the boundary cases. γ_i, γ_{i+1}, μ_1, μ_2 are the slopes of the segments of ρ_i^* respectively the (sub-)label positions of ρ_i.

for some $\alpha, \beta \in [0,1]$ and $\nu \in \mathbb{R}^d$. The intuition is that $\Phi_{i,j}$ penalizes a jump from γ_i^α to γ_j^β in the direction of ν. Since $\boldsymbol{\Phi}$ is nonconvex we compute the convex envelope.

Proposition 4. *The convex envelope of (15) is*

$$\boldsymbol{\Phi}^{**}(\boldsymbol{g}) = \sup_{\boldsymbol{p} \in \mathcal{K}} \langle \boldsymbol{p}, \boldsymbol{g} \rangle, \tag{17}$$

where $\mathcal{K} \subset \mathbb{R}^{k \times d}$ is given as:

$$\mathcal{K} = \left\{ \boldsymbol{p} \in \mathbb{R}^{k \times d} \,\middle|\, \left|\boldsymbol{p}^\mathsf{T}(\mathbf{1}_i^\alpha - \mathbf{1}_j^\beta)\right|_2 \leq \left|\gamma_i^\alpha - \gamma_j^\beta\right|, \right.$$
$$\left. \forall\, 1 \leq i \leq j \leq k, \; \forall \alpha, \beta \in [0,1] \right\}. \tag{18}$$

Proof. See supplementary material. □

The set $\mathcal{K}$ from Eq. (18) involves infinitely many constraints which makes numerical optimization difficult. As the following proposition reveals, the infinite number of constraints can be reduced to only linearly many, allowing to enforce the constraint $\boldsymbol{p} \in \mathcal{K}$ exactly.

Proposition 5. *If the labels are ordered ($\gamma_1 < \gamma_2 < \ldots < \gamma_L$) then the constraint set $\mathcal{K}$ from Eq. (18) is equal to*

$$\mathcal{K} = \{\boldsymbol{p} \in \mathbb{R}^{k \times d} \mid |\boldsymbol{p}_i|_2 \leq \gamma_{i+1} - \gamma_i, \; \forall i\}. \tag{19}$$

Proof. See supplementary material. □

Direct Optimization of (25), $t = 0.6s$, 11.78 MB

Baseline ($L = 8$), $t = \infty$, 113 MB

Baseline ($L = 16$), $t = \infty$, 226 MB

Baseline ($L = 256$), $t = \infty$, 3619 MB

Proposed ($L = 2$) $t = 1s$, 27 MB
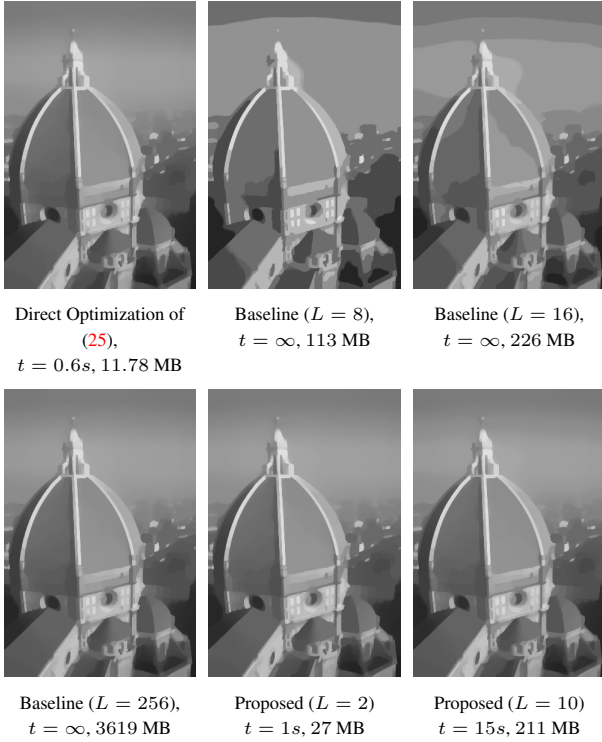
Proposed ($L = 10$) $t = 15s$, 211 MB

Figure 5. Denoising comparison. We compare the proposed method to the baseline method [17] on the convex ROF problem. We show the time in seconds required for each method to produce a solution within a certain energy gap to the optimal solution. As the baseline method optimizes a piecewise linear approximation of the quadratic dataterm, it fails to reach that optimality gap even for $L = 256$ (indicated by $t = \infty$). In contrast, while the proposed lifting method can solve a large class of non-convex problems, it is almost as efficient as direct methods on convex problems.

This shows that the proposed regularizer coincides with the total variation from [5], where it has been derived based on (16) for α and β restricted to $\{0, 1\}$. Prop. 5 together with Prop. 3 show that for $k = 1$ our formulation amounts to unlifted TV optimization with a convexified dataterm.

4. Numerical Optimization

Discretizing $\Omega \subset \mathbb{R}^d$ as a d-dimensional Cartesian grid, the relaxed energy minimization problem becomes

$$\min_{\boldsymbol{u}:\Omega\to\mathbb{R}^k} \sum_{x\in\Omega} \boldsymbol{\rho}^{**}(x, \boldsymbol{u}(x)) + \boldsymbol{\Phi}^{**}(x, \nabla\boldsymbol{u}(x)), \quad (20)$$

where ∇ denotes a forward-difference operator with $\nabla\boldsymbol{u} : \Omega \to \mathbb{R}^{k\times d}$. We rewrite the dataterm given in equation (8) by replacing the pointwise maximum over the conjugates $\boldsymbol{\rho}_i^*$ with a maximum over a real number $q \in \mathbb{R}$ and obtain

Input image f

Proposed ($L = 5$), $E = 20494$, $t = 14.6s$

Proposed ($L = 10$), $E = 18844$, $t = 30.5s$

Proposed ($L = 20$), $E = 18699$, $t = 123.9s$

Baseline ($L = 256$), $E = 18660$, $t = 1001s$

Baseline ($L = 5$), $E = 23864$, $t = 4.7s$

Baseline ($L = 10$), $E = 19802$, $t = 6.3s$
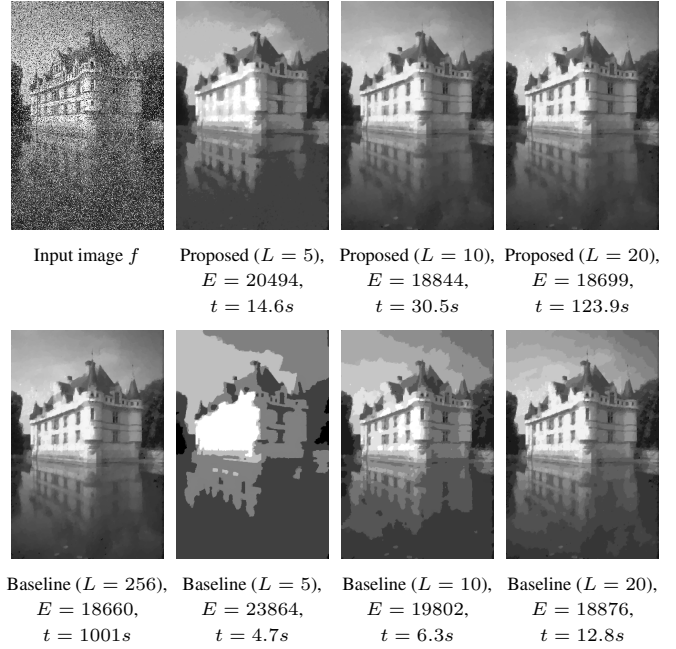
Baseline ($L = 20$), $E = 18876$, $t = 12.8s$

Figure 6. Denoising using a robust truncated quadratic dataterm. The top row shows the input image along with the result obtained by our approach for a varying number of labels L. The bottom row illustrates the results obtained by the baseline method [17]. The energy of the final solution as well as the total runtime are given below each image.

the following saddle point formulation of problem (20):

$$\min_{\substack{\boldsymbol{u}:\Omega\to\mathbb{R}^k \\ \boldsymbol{p}:\Omega\to\mathcal{K}}} \max_{\substack{(\boldsymbol{v},q)\in\mathcal{C}}} \langle\boldsymbol{u},\boldsymbol{v}\rangle - \sum_{x\in\Omega} q(x) + \langle\boldsymbol{p}, \nabla\boldsymbol{u}\rangle, \quad (21)$$

$$\mathcal{C} = \{(\boldsymbol{v}, q) : \Omega \to \mathbb{R}^k \times \mathbb{R} \mid q(x) \geq \boldsymbol{\rho}_i^*(\boldsymbol{v}(x)), \forall x, \forall i\}. \quad (22)$$

We numerically compute a minimizer of problem (21) using a first-order primal-dual method [6, 16] with diagonal preconditioning [14] and adaptive steps [9]. It alternates between a gradient descent step in the primal variable and a gradient ascent step in the dual variable. Subsequently the dual variables are orthogonally projected onto the sets $\mathcal{C}$ respectively $\mathcal{K}$. In the following we give some hints on the implementation of the individual steps. For a detailed discussion we refer to [9]. The projection onto the set $\mathcal{K}$ is a simple ℓ_2-ball projection. To simplify the projection onto $\mathcal{C}$, we transform the k-dimensional epigraph constraints in (22) into 1-dimensional scaled epigraph constraints by introducing an additional variable $z : \Omega \to \mathbb{R}^k$ with:

$$\boldsymbol{z}_i(x) = [q(x) - c_i(\boldsymbol{v}(x))](\gamma_{i+1} - \gamma_i). \quad (23)$$

| $E = 279394$ | $E = 208432$ | $E = \mathbf{196803}$ | $E = 194855$ |

| $E = 278108$ | $E = \mathbf{208112}$ | $E = 196810$ | $E = 194845$ |

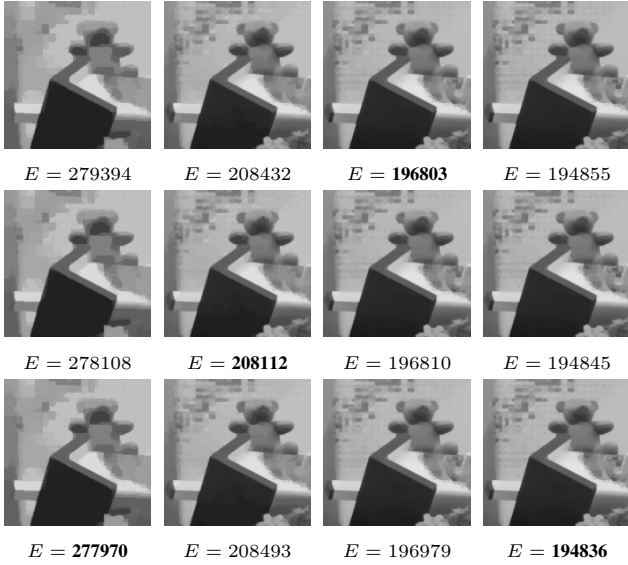| $E = \mathbf{277970}$ | $E = 208493$ | $E = 196979$ | $E = \mathbf{194836}$ |

Figure 7. Comparison to the MRF approach presented in [27]. The first row shows DC-Linear, second row DC-MRF and third row our results for 4, 8, 16 and 32 convex pieces on the truncated quadratic energy (26). Below the figures we show the final nonconvex energy. We achieve competitive results while using a more compact representation and generalizing to isotropic regularizers.

Using equation (9) we can write the constraints in (22) as

$$\frac{\boldsymbol{z}_i(x)}{\gamma_{i+1} - \gamma_i} \geq \rho_i^* \left(\frac{\boldsymbol{v}_i(x)}{\gamma_{i+1} - \gamma_i} \right). \qquad (24)$$

We implement the newly introduced equality constraints (23) introducing a Lagrange multiplier $\boldsymbol{s} : \Omega \to \mathbb{R}^k$. It remains to discuss the orthogonal projections onto the epigraphs of the conjugates ρ_i^*. Currently we support quadratic and piecewise linear convex pieces ρ_i. For the piecewise linear case, the conjugate ρ_i^* is a piecewise linear function with domain $\mathbb{R}$. The slopes correspond to the x-positions of the sublabels and the intercepts correspond to the function values at the sublabel positions.

The conjugates as well as the epigraph projections of both, a quadratic and a piecewise linear piece are depicted in Fig. 4. For the quadratic case, the projection onto the epigraph of a parabola is computed using [23, Appendix B.2].

5. Experiments

We implemented the primal-dual algorithm in CUDA to run on GPUs. [1] For $d = 2$, our implementation of the func-

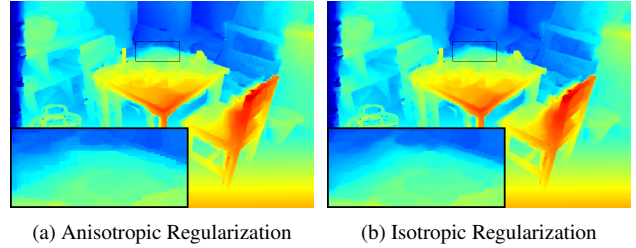(a) Anisotropic Regularization   (b) Isotropic Regularization

Figure 8. We compare the proposed relaxation with anistropic regularizer to isotropic regularization on the stereo matching example. Using an anisotropic formulation as in [27] leads to grid bias.

tional lifting framework [17], which will serve as a baseline method, requires $4N(L - 1)$ optimization variables, while the proposed method requires $6N(L - 1) + N$ variables, where N is the number of points used to discretize the domain $\Omega \subset \mathbb{R}^d$. As we will show, our method requires much fewer labels to yield comparable results, thus, leading to an improvement in accuracy, memory usage, and speed.

5.1. Rudin-Osher-Fatemi Model

As a proof of concept, we first evaluate the novel relaxation on the well-known Rudin-Osher-Fatemi (ROF) model [20]. It corresponds to (1) with the following dataterm:

$$\rho(x, u(x)) = (u(x) - f(x))^2, \qquad (25)$$

where $f : \Omega \to \mathbb{R}$ denotes the input data. While there is no practical use in applying convex relaxation methods to an already convex problem such as the ROF model, the purpose of this is two-fold. Firstly, it allows us to measure the overhead introduced by our method by comparing it to standard convex optimization methods which do not rely on functional lifting. Secondly, we can experimentally verify that the relaxation is tight for a convex dataterm.

In Fig. 5 we solve (25) directly using the primal-dual algorithm [9], using the baseline functional lifting method [17] and using our proposed algorithm. First, the globally optimal energy was computed using the direct method with a very high number of iterations. Then we measure how long each method took to reach this global optimum to a fixed tolerance.

The baseline method fails to reach the global optimum even for 256 labels. While the lifting framework introduces a certain overhead, the proposed method finds the same globally optimal energy as the direct unlifted optimization approach and generalizes to nonconvex energies.

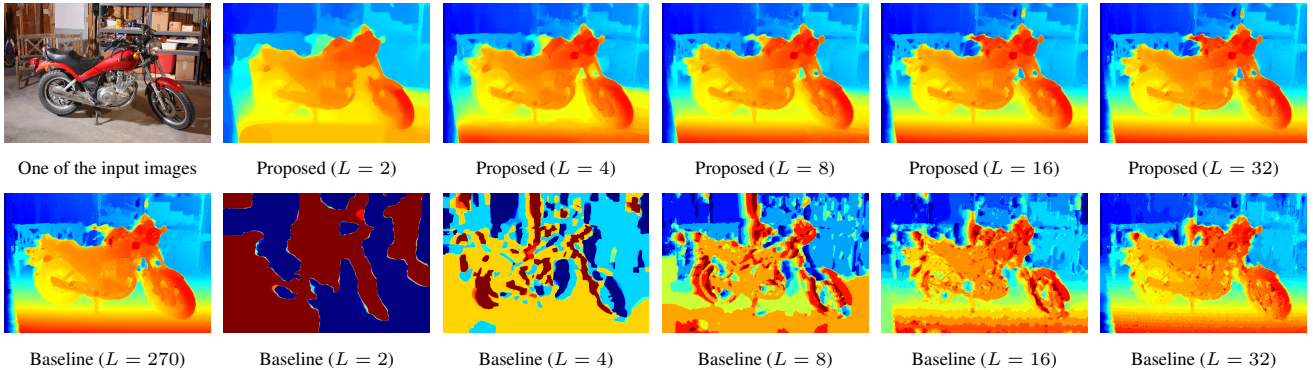|					
One of the input images	Proposed ($L = 2$)	Proposed ($L = 4$)	Proposed ($L = 8$)	Proposed ($L = 16$)	Proposed ($L = 32$)
Baseline ($L = 270$)	Baseline ($L = 2$)	Baseline ($L = 4$)	Baseline ($L = 8$)	Baseline ($L = 16$)	Baseline ($L = 32$)

Figure 9. Stereo comparison. We compare the proposed method to the baseline method on the example of stereo matching. The first column shows one of the two input images and below the baseline method with the full number of labels. The proposed relaxation requires much fewer labels to reach a smooth depth map. Even for $L = 32$, the label space discretization of the baseline method is strongly visible, while the proposed method yields a smooth result already for $L = 8$.

5.2. Robust Truncated Quadratic Dataterm

The quadratic dataterm in (25) is often not well suited for real-world data as it comes from a pure Gaussian noise assumption and does not model outliers. We now consider a robust truncated quadratic dataterm:

$$\rho(x, u(x)) = \frac{\alpha}{2} \min \left\{ (u(x) - f(x))^2, \nu \right\}. \quad (26)$$

To implement (26), we use a piecewise polynomial approximation of the dataterm. In Fig. 6 we degraded the input image with additive Gaussian and salt and pepper noise. The parameters in (26) were chosen as $\alpha = 25$, $\nu = 0.025$ and $\lambda = 1$. It can be seen that the proposed method requires fewer labels to find lower energies than the baseline.

5.3. Comparison to the Method of Zach and Kohli

We remark that Prop. 4 and Prop. 5 hold for arbitrary convex one-homogeneous functionals $\phi(\nu)$ instead of $|\nu|_2$ in equation (16). In particular, they hold for the anisotropic total variation $\phi(\nu) = |\nu|_1$. This generalization allows us to directly compare our convex relaxation to the MRF approach of Zach and Kohli [27].

In Fig. 7 we show the results of optimizing the two models entitled "DC-Linear" and "DC-MRF" proposed in [27], and of our proposed method with anisotropic regularization on the robust truncated denoising energy (26). We picked the parameters as $\alpha = 0.2$, $\nu = 500$, and $\lambda = 1$. The label space is also chosen as $\Gamma = [0, 256]$ as described in [27]. Note that overall, all the energies are better than the ones reported in [27]. It can be seen from Fig. 7 that the proposed relaxation is competitive to the one pro-

posed by Zach and Kohli. In addition, the proposed relaxation uses a more compact representation and extends to isotropic and convex one-homogeneous regularizers. To illustrate the advantages of isotropic regularizations, Fig. 8a and Fig. 8b show a comparison of our proposed method for isotropic and anisotropic regularization for the example of stereo matching discussed in the next section.

5.4. Stereo Matching

Given a pair of rectified images, the task of finding a correspondence between the two images can be formulated as an optimization problem over a scalar field $u : \Omega \to \Gamma$ where each point $u(x) \in \Gamma$ denotes the displacement along the epipolar line associated with each $x \in \Omega$. The overall cost functional fits Eq. (1). In our experiments, we computed $\rho(x, u(x))$ for 270 disparities on the Middlebury stereo benchmark [21] in a 4×4 patch using a truncated sum of absolute gradient differences. We convexify the matching cost ρ in each range Γ_i by numerically computing the convex envelope using the gift wrapping algorithm.

The first row in Fig. 9 shows the result of the proposed relaxation using the convexified energy between two labels. The second row shows the baseline approach using the same amount of labels. Even for $L = 2$, the proposed method produces a reasonable depth map while the baseline approach basically corresponds to a two region segmentation.

5.5. Phase Unwrapping

Many sensors such as time-of-flight cameras or interferometric synthetic aperture radar (SAR) yield cyclic data ly-

| One of the input images | Proposed ($L = 2$) | Proposed ($L = 4$) | Proposed ($L = 8$) | Proposed ($L = 16$) | Proposed ($L = 32$) |

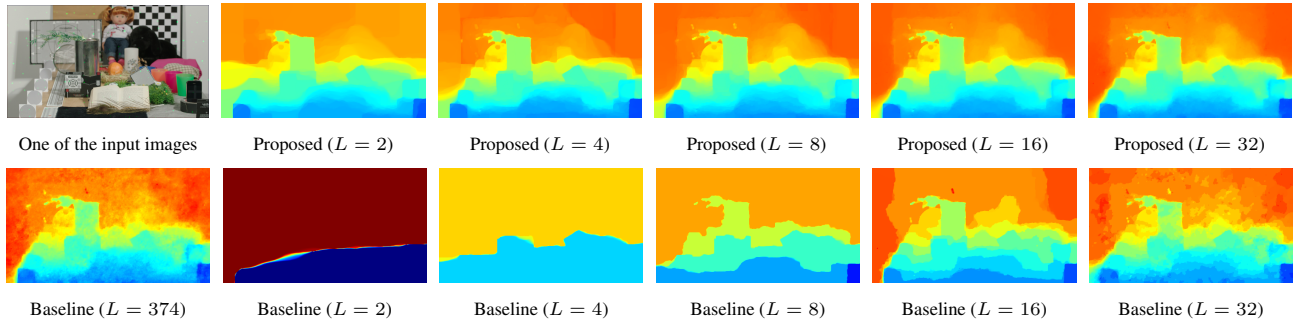| Baseline ($L = 374$) | Baseline ($L = 2$) | Baseline ($L = 4$) | Baseline ($L = 8$) | Baseline ($L = 16$) | Baseline ($L = 32$) |

Figure 10. Depth from focus comparison. We compare our method to the baseline approach on the problem of depth from focus. First column: one of the 374 differently focused input images and the baseline method for full number of labels. Following columns: proposed relaxation (top row) vs. baseline (bottom row) for 2, 4, 8, 16 and 32 labels each.

ing on the circle $\mathcal{S}^1$. Here we consider the task of total variation regularized unwrapping. As is shown on the left in Fig. 11, the dataterm is a nonconvex function where each minimum corresponds to a phase shift by 2π:

$$\rho\left(x, u(x)\right) = d_{\mathcal{S}^1}\left(u(x), f(x)\right)^2. \tag{27}$$

For the experiments, we approximated the nonconvex energy by quadratic pieces as depicted in Fig. 11. The label space is chosen as $\Gamma = [0, 4\pi]$ and the regularization parameter was set to $\lambda = 0.005$. Again, it is visible in Fig. 11 that the baseline method shows label space discretization and fails to unwrap the depth map correctly if the number of labels is chosen too low. The proposed method yields a smooth unwrapped result using only 8 labels.

5.6. Depth From Focus

In depth from focus the task is to recover the depth of a scene, given a stack of images each taken from a constant position but in a different focal setting, so that in each image only the objects of a certain depth are sharp. images. We compute the dataterm cost ρ by using the modified Laplacian function [13] as a contrast measure.

Similar to the stereo experiments, we convexify the cost on each label range by computing the convex hull. The results are shown in Fig. 10. While the baseline method clearly shows the label space discretization, the proposed approach yields a smooth depth map. Since the proposed method uses a convex lower bound of the lifted energy, the regularizer has slightly more influence on the final result. This explains why the resulting depth maps in Fig. 10 and Fig. 9 look overall less noisy.

| Piecewise convex energy | Input image | Ground truth |

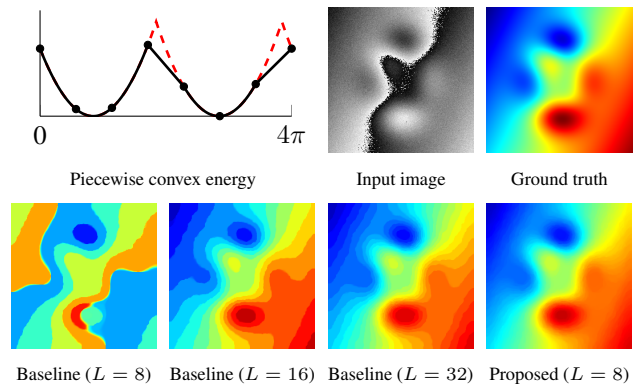| Baseline ($L = 8$) | Baseline ($L = 16$) | Baseline ($L = 32$) | Proposed ($L = 8$) |

Figure 11. We show the piecewise convex approximation of the phase unwrapping energy, followed by the cyclic input image and the unwrapped ground truth. With only 8 labels, the proposed method already yields a smooth reconstruction. The baseline method fails to unwrap the heightmap correctly using 8 labels, and for 16 and 32 labels, the discretization is still noticable.

6. Conclusion

In this work we proposed a tight convex relaxation that can be interpreted as a sublabel–accurate formulation of classical multilabel problems. The final formulation is a simple saddle-point problem that admits fast primal-dual optimization. Our method maintains sublabel accuracy even after discretization and for that reason outperforms existing spatially continuous methods. Interesting directions for future work include higher dimensional label spaces, manifold valued data and more general regularizers.

References

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

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, 2000. 4

[3] E. Beale and J. Tomlin. Special facilities in a general mathematical programming system for nonconvex problems using ordered sets of variables. *Proceedings of the fifth international conference on operational research*, pages 447–454, 1970. 3

[4] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9:263–340, 2010. 1

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

[6] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010. 5

[7] A. Fix and S. Agarwal. Duality and the continuous graphical model. In *Computer Vision ECCV 2014*, volume 8691 of *Lecture Notes in Computer Science*, pages 266–281. Springer International Publishing, 2014. 2

[8] M. Giaquinta, G. Modica, and J. Souček. *Cartesian currents in the calculus of variations I, II.*, volume 37-38 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3.* Springer-Verlag, Berlin, 1998. 3

[9] T. Goldstein, E. Esser, and R. Baraniuk. Adaptive Primal-Dual Hybrid Gradient Methods for Saddle-Point Problems. *arXiv Preprint*, 2013. 5, 6

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

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

[12] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. In *ICCV*, December 2013. 2

[13] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(8):824–831, Aug. 1994. 8

[14] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *ICCV*, pages 1762–1769, 2011. 5

[15] T. Pock, A. Chambolle, H. Bischof, and D. Cremers. A convex relaxation approach for computing minimal partitions. In *CVPR*, pages 810–817, 2009. 1, 4

[16] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In *ICCV*, 2009. 1, 5

[17] 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. 1, 2, 3, 4, 5, 6

[18] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *European Conference on Computer Vision (ECCV)*, Marseille, France, October 2008. 1, 2, 4

[19] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1996. 2

[20] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. 6

[21] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nei, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, volume 8753, pages 31–42. Springer, 2014. 7

[22] 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

[23] E. Strekalovskiy, A. Chambolle, and D. Cremers. Convex relaxation of vectorial problems with coupled regularization. *SIAM Journal on Imaging Sciences*, 7(1):294–336, 2014. 6

[24] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(7):1165–1179, July 2007. 2

[25] C. Zach. Dual decomposition for joint discrete-continuous optimization. In *AISTATS*, pages 632–640, 2013. 2

[26] C. Zach, C. Hane, and M. Pollefeys. What is optimized in tight convex relaxations for multi-label problems? In *CVPR*, pages 1664–1671, 2012. 2

[27] C. Zach and P. Kohli. A convex discrete-continuous approach for markov random fields. In *ECCV*, volume 7577, pages 386–399. Springer Berlin Heidelberg, 2012. 2, 6, 7