

# Image Reconstruction by Multilabel Propagation

Matthias Zisler<sup>1</sup>(✉), Freddie Åström<sup>1</sup>, Stefania Petra<sup>2</sup>, and Christoph Schnörr<sup>1</sup>

<sup>1</sup> Image and Pattern Analysis Group, Heidelberg University, Heidelberg, Germany  
zisler@math.uni-heidelberg.de

<sup>2</sup> Mathematical Imaging Group, Heidelberg University, Heidelberg, Germany

**Abstract.** This work presents a non-convex variational approach to joint image reconstruction and labeling. Our regularization strategy, based on the KL-divergence, takes into account the smooth geometry on the space of discrete probability distributions. The proposed objective function is efficiently minimized via DC programming which amounts to solving a sequence of convex programs, with guaranteed convergence to a critical point. Each convex program is solved by a generalized primal dual algorithm. This entails the evaluation of a proximal mapping, evaluated efficiently by a fixed point iteration. We illustrate our approach on few key scenarios in discrete tomography and image deblurring.

## 1 Introduction

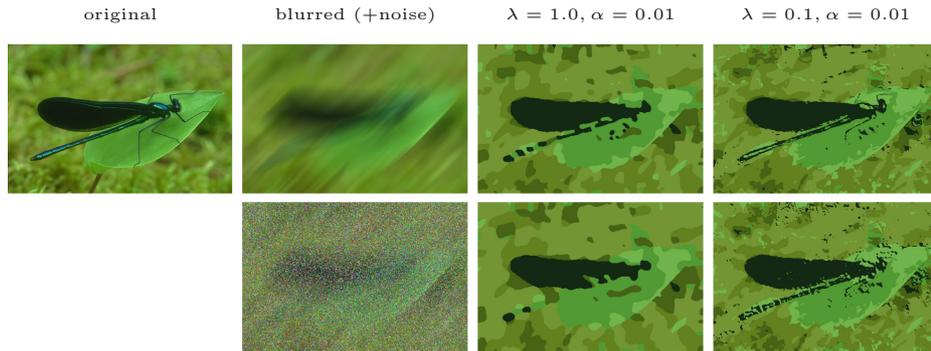
Optimal partition of image data into multiple discrete classes, each representing some semantic information is a relevant problem not only in visual scene understanding but also in, e.g., discrete tomography. A class, or label, in these examples may include sky, road, person and various tissue types such as bone or soft tissue. In addition to be defined from a finite set of image labels, discrete tomography data must first be reconstructed from few discrete projections (data measurements) which constitutes a highly ill-posed problem.

In this work, we propose to jointly solve the labeling problem while enforcing a linear constraint systems, such as the one stemming from discrete tomography. Our smooth variational formulation enables efficient inference of the otherwise NP-hard constrained multilabeling problem. We formulate the objective in a general setting and propose a regularization strategy taking into account the smooth geometry on the space of discrete probability distributions, induced by the Fisher-Rao metric. We focus on the key applications of non-binary discrete tomography (e.g., non-destructive material testing [1]) and deblurring and denoising with joint labeling. As illustrated in Fig. 1 our framework can accurately reconstruct and label severely blurred and noisy data.

**Related Work.** To avoid the combinatorial nature of discrete optimization problems, it is common to use convex relaxations to approximate the integrality constraints [2–5]. However, convex relaxation is loose in connection with a weak

---

**Acknowledgments:** We gratefully acknowledge support by the German Science Foundation, grant GRK 1653.



**Fig. 1.** Reconstruction and labeling with our proposed model (3) of a severely blurred and noisy image of an insect with discrete label set  $\mathcal{L} = \{\square, \square, \square, \square, \square, \square\}$ . In both cases our model can reconstruct fine details present in the original image (see [8]). For the experimental setup we refer to Sect. 4. (Color figure online)

data term, and the performance of the required rounding step (post-processing), projecting the solution of the relaxed problem to the set of feasible solutions, is hard to control. Non-convex approaches perform rounding or discretization already during optimization. On the other hand, these formulations are sensitive to initialization. Recently, a continuous, smooth non-convex approach to image labeling was introduced in [6] which avoids many of the aforementioned drawbacks (see also [7]). In particular, the labeling is initialized with the uninformative barycentric coordinates of the probability simplex but still avoid poor local minima. The underlying mechanism, which governs the inference process, is the evolution of a Riemannian gradient flow defined on the manifold of row-stochastic matrices which terminates at a labeling. Inspired from this work we formulate our constrained labeling problem as an optimization problem enforcing spatial consistency via discrete probability distributions, which entail the joint reconstruction and labeling. The ability of the present approach to simultaneously perform rounding and optimization is a significant conceptual difference to all approaches based on convex relaxations.

Constraining the solution of an inverse problem to piece-wise constant regions motivates the use of sparsifying priors, such as total variation (TV) based energy formulations [9]. However, applying the TV prior component-wise on the simplex variables, e.g., as done in [10] does not respect the geometry of the underlying probability distributions. To overcome this shortcoming, we propose to exploit the Kullback-Leibler (KL) divergence to enforce spatially consistent assignments. Our approach is motivated from the fact that the KL-divergence locally approximates the squared geodesic distance when considering the probability simplex as a Riemannian manifold endowed with the Fisher-Rao metric [11, 12]. Note that we use the same strategy from [10] to obtain a meaningful data term for the constraint labeling problem.

**Contributions and Organization.** Section 2 gives an overview of the constrained multilabeling problem. This section also introduces our *non-convex*

approach which reconstructs an image and simultaneously enforces spatially coherent labeling by our novel regularizer. Our optimization problem is formulated in the framework of difference of convex function (DC) programming in Sect. 3, which comes along with a convergence guarantee. Furthermore we evaluate the generalized proximal mapping of our proposed regularizer by an fixed point iteration rather than solving a large non-linear system of equations. This strategy is numerically efficient even for larger problem instances. In Sect. 4 we compare our approach on few problem instances and Sect. 5 concludes the paper.

**Basic Notation.** Operations and functions are applied component-wise to vectors  $v, w \in \mathbb{R}^n$  and matrices i.e.,  $vw = (\dots, v_i w_i, \dots)$ . The KL-divergence is defined by  $\text{KL}(x, y) = \langle x, \log(x/y) \rangle$  for stochastic vectors as well as row-stochastic matrices where  $\langle \cdot, \cdot \rangle$  denotes the Euclidean scalar product. Moreover we set  $\mathbb{1} = (1, 1, \dots, 1)^T$ .

## 2 Constrained Multilabeling, Model and Relaxation

**Problem Statement.** Consider the linear system of equations

$$Au = b, \quad u_i \in \mathcal{L} \quad \forall i = 1, \dots, N, \quad (1)$$

where the solution is constrained to a discrete set  $\mathcal{L}$  of labels. Note in the general setting this integer constraint formulation leads to NP hard problems. Furthermore, we assume that there are less measurements  $b$  than pixels  $m \ll N$  and hence the inverse problem is ill-posed and requires prior knowledge (regularization). A common choice is the Potts model [13],  $\|\nabla u\|_0 := |\{i \mid (\nabla u)_i \neq 0\}|$  for sparse gradient regularization which favours piecewise constant images. This gives the problem

$$\min_u \lambda \cdot \|\nabla u\|_0 \quad \text{s.t.} \quad Au = b \quad \wedge \quad u_i \in \mathcal{L} \quad \forall i = 1, \dots, N \quad (2)$$

and we refer to (2) as a *constrained multilabeling problem* with Potts regularization. From the viewpoint of graphical models, the system of affine subspace constraints induce (very) high-order potentials. This high-order interaction induced by the non-local constraints results in a non-standard labeling problem that is intractable for large problem sizes. Therefore, we instead adopt the strategy of solving a sequence of convex relaxations in order to minimize a non-convex energy, which properly approximates the original problem.

**Model and Relaxation.** We relax the hard assignment of a label from a given set  $\mathcal{L} := \{c_1, \dots, c_K\}$  of priors to each pixel  $i \in [N] = \{1, 2, \dots, N\}$  to discrete probability distributions  $z \in \mathcal{G} = \{z \in [0, 1]^{N \times K} : \langle z_i, \mathbb{1} \rangle = 1, \forall i \in [N]\}$  and  $(z_{ik})_{k=1}^K$  is the distribution describing the assignment in pixel  $i$ .

*Energy.* We propose the non-convex energy

$$\min_{z \in \mathcal{G}} J(z), \quad J(z) = \lambda R_{\mathcal{G}}(z) - \alpha \langle z, \log(z) \rangle + D(z) \quad (3)$$

which consists of three basic building blocks detailed below: (i) regularization for spatial coherence controlled by parameter  $\lambda \geq 0$ , (ii) an entropy term enforcing an unique decision with weight  $\alpha \geq 0$  and (iii) a dataterm. For convex relaxations [3–5], a rounding scheme is generally required to obtain an integral solution. In our case, however, the concave entropy term promotes an integral solution.

**Dataterm.** We consider two cases of the dataterm: *separable* and *non-separable*. The separable case refers to problem where noisy image data  $u_i^0$  is directly observed. We introduce a distance function  $d_{\mathcal{L}}$  measuring the similarity to the priors  $c_k \in \mathcal{L}, k \in [K]$ , resulting in the dataterm

$$D_{\text{unary}}(z, S) = \langle z, S \rangle \quad \text{where} \quad S_{ik} := d_{\mathcal{L}}(u_i^0, c_k). \quad (4)$$

We introduce the assignment operator  $P_{\mathcal{L}}(z) = zc$  where  $c = (c_1, \dots, c_K)^T$ , which assigns to each pixel  $i$  a convex combination of labels in terms of the distribution  $z_i$ . The non-separable case refers to problems where image data  $u_i^0$  cannot be directly observed since it is the solution of the inverse problem (1). As a consequence, we instead minimize the distance  $d(AP_{\mathcal{L}}(z), b)$ , ( $u = P_{\mathcal{L}}(z)$  in (1)) between the forward projection  $A$  and given measurements  $b$ . In [10] it was shown that if  $D(z) = D(P_{\mathcal{L}}(z))$  is defined over the assigned solution  $P_{\mathcal{L}}(z)$  one is required to introduce, e.g., a concave self-assignment term. Accordingly, we define the dataterm

$$D_{\text{inverse}}(z, A, b) = d(AP_{\mathcal{L}}(z), b) + \langle z, (P_{\mathcal{L}}(z)\mathbb{1}_K^T - \mathbb{1}_N c^T)^2 \rangle, \quad (5)$$

for non-separable inverse labeling problems. Note that when the self-assignment term in (5) is constrained to the simplex, then the vertices of the simplex become its minima. The entropy term (3), which has the same minimizers as the self-assignment term, enforces integral solutions, while the self-assignment term in (5) has a meaningful descent direction w.r.t. to the labels (pushing the assignment  $P_{\mathcal{L}}(z)$  towards the label values  $c_k$ ). Furthermore, the linearization of the self-assignment term in a point  $z^0$  resembles  $D_{\text{unary}}$  with  $u^0 = P_{\mathcal{L}}(z^0)$  and is the squared Euclidean norm.

**Regularizer.** To enforce spatial coherence over pixel-wise probability distributions Zach et al. [2] regularize each individual layer  $z_k$  to get a convex relaxation of the Potts model. A tighter relaxation is obtained by regularizing across the layers [3, 4]. However, these works employ Euclidean norms that disregard the underlying geometry of the discrete probability distributions and hence necessitate an additional re-projection step onto the simplex.

Instead, we propose a regularizer  $R_{\mathcal{G}}(z)$  (see (3)) which respects the underlying geometry of the probability simplex by coupling probability distributions across layers via the KL divergence. Our regularizer is defined as

$$R_{\mathcal{G}}(z) := \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} \frac{1}{N_s} \text{KL}(z_i, z_j) \quad \text{where} \quad N_s := |\mathcal{N}(i)|. \quad (6)$$

which enforces spatial coherence by pairwise interactions in neighborhoods  $\mathcal{N}(i)$  induced by the underlying grid-graph of the image.

It is well-known that the KL divergence locally approximates the *squared* geodesic distance on the probability simplex equipped with the Fisher-Rao metric [11]. In this sense, (6) naturally respects the information geometric properties of the underlying manifold. Furthermore, for this particular manifold, our formulation *without* approximation of the quadratic geodesic distance would correspond to a *non-local* extension of a quadratic regularizer in the framework of [14]. We have

**Lemma 1 (Basic properties).** *Let  $z \in \mathcal{G}$  and define  $R_{\mathcal{G}}(z)$  by (6). Then*

1.  $R_{\mathcal{G}}(z)$  is a convex function
2.  $R_{\mathcal{G}}(z)$  is the KL-divergence between  $z_i$  and the geometric mean of the vectors  $z_j$  indexed by  $j \in \mathcal{N}(i)$ ,

$$R_{\mathcal{G}}(z) = \sum_{i=1}^N \text{KL}(z_i, \text{gm}(\{z_j\}_{j \in \mathcal{N}(i)})), \quad \text{gm}(\{z_j\}_{j \in \mathcal{N}(i)}) := \prod_{j \in \mathcal{N}(i)} z_j^{\frac{1}{N_s}}. \quad (7)$$

*Proof.* Assertion 1 follows from the joint convexity of the KL-divergence [15]. The second claim can be seen by

$$R_{\mathcal{G}}(z) = \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} \frac{1}{N_s} \langle z_i, \log\left(\frac{z_i}{z_j}\right) \rangle = \sum_{i=1}^N \langle z_i, \log\left(\prod_{j \in \mathcal{N}(i)} \left(\frac{z_i}{z_j}\right)^{\frac{1}{N_s}}\right) \rangle \quad (8a)$$

$$= \sum_{i=1}^N \langle z_i, \log\left(z_i^{N_s \frac{1}{N_s}} \prod_{j \in \mathcal{N}(i)} (z_j^{\frac{1}{N_s}})^{-1}\right) \rangle = \sum_{i=1}^N \text{KL}\left(z_i, \prod_{j \in \mathcal{N}(i)} z_j^{\frac{1}{N_s}}\right) \quad (8b)$$

□

Next we reformulate the objective function (3) as a difference of convex (DC) program [16] and work out a corresponding optimization algorithm.

### 3 Optimization

**DC Programming.** A large subclass of non-convex objective functions are DC functions which can be (locally) minimized by DC Programming [16]. The basic form of a DC program is given by

$$z^* = \arg \min_z g(z) - h(z), \quad (9)$$

where  $g(z)$  and  $h(z)$  are proper, lower semicontinuous, convex functions. There exists a simplified version of the DC algorithm [17] for minimizing (9) which guarantees convergence to a critical point by starting with  $z^0 \in \text{dom}(g)$  and then alternately applying the updates  $v^n \in \partial h(z^n)$  and  $z^{n+1} \in \partial g^*(v^n)$  until a termination criterion is reached, where  $g^*$  denotes the Legendre-Fenchel conjugate [18] of  $g$ .

**Algorithm 1.** Iterated Primal Dual Algorithm**Init:** choose the barycenter for  $z^0 \in \mathcal{G}$ ,  $q^0 \in \text{dom}(D^*)$  and  $\tau, \sigma > 0$ **while not converged do**  Set  $\hat{z} = z^l$   **while not converged do**

$$z^{n+1} = \arg \min_{z \in \mathcal{G}} \lambda R_{\mathcal{G}}(z) + \langle z, A^* q^n - \nabla h(\hat{z}) \rangle + \frac{1}{\tau} \text{KL}(z, z^n) \quad (11)$$

$$q^{n+1} = \arg \min_q D^*(q) - \langle q, A(2z^{n+1} - z^n) \rangle + \frac{1}{2\sigma} \|q - q^n\|_2^2 \quad (12)$$

 $n \leftarrow n + 1$    $l \leftarrow l + 1$ **Output:**  $z^* = z^l$ 

To apply the DC algorithm to our non-convex energy  $J(z)$  in (3), we rewrite  $J(z) = g(z) - h(z)$  as a DC function. We set  $h_{\text{unary}}(z) = \alpha \langle z, \log(z) \rangle$  for the case of a separable dataterm and  $h_{\text{inverse}}(z) = \alpha \langle z, \log(z) \rangle - \langle z, (\text{P}_{\mathcal{L}}(z) \mathbb{1}_K^T - \mathbb{1}_N c^T)^2 \rangle$  for the non-separable case since the entropy and the self assignment term are concave. We denote by  $g(z) = \lambda R_{\mathcal{G}}(z) + D(z)$  the remaining convex terms from (3), where  $D(z)$  corresponds to the convex part of (5). The DC algorithm results in the fixed point iteration

$$z^{n+1} = \arg \min_{z \in \mathcal{G}} \lambda R_{\mathcal{G}}(z) + D(z) - \langle z, \nabla h(z^n) \rangle, \quad (10)$$

where the gradient  $\nabla h(z)$  for the separable case is given by  $\nabla h_{\text{unary}}(z) = \alpha(\log(z) + \mathbb{1}_N \mathbb{1}_K^T)$  and in the non-separable case by  $\nabla h_{\text{inverse}}(z) = \alpha(\log(z) + \mathbb{1}_N \mathbb{1}_K^T) - (\text{P}_{\mathcal{L}}(z) \mathbb{1}_K^T - \mathbb{1}_N c^T)^2$ . We refer to [10] for the gradient of the second term of  $\nabla h_{\text{inverse}}(z)$ .

**Solving the Fixed Point Iteration.** Algorithm 1 solves the fixed point iteration (10) iteratively using the generalized primal dual algorithm [19].

*Primal Update.* The primal update step (11) requires to evaluate the generalized proximal operator of the regularizer (6). We rewrite (11) as

$$z^{n+1} = \arg \min_{z \in \mathcal{G}} R_{\mathcal{G}}(z) + \frac{1}{\lambda\tau} \text{KL}(z, p), \quad (13)$$

where the argument  $p \in \mathcal{G}$  is given by the non-linear gradient descent step

$$p = \arg \min_{z \in \mathcal{G}} \langle z, A^* q^n - \nabla h(\hat{z}) \rangle + \frac{1}{\tau} \text{KL}(z, z^n) \quad (14a)$$

$$= \frac{z^n \exp(-\tau(A^* q^n - \nabla h(\hat{z})))}{\langle z^n, \exp(-\tau(A^* q^n - \nabla h(\hat{z}))) \rangle}. \quad (14b)$$

Note that the argmin induces normalization of  $p$ , thus  $p \in \mathcal{G}$ .

Theorem 1 below states that evaluating the proximal mapping (13) can be done approximately by an efficient fixed point iteration rather than solving a large non-linear equation system (optimality conditions). Even for larger problem instances, this fixed point iteration evaluates the proximal mapping very efficiently. Specifically, in our numerical experiments we observed convergence within few iterations and we initialize with  $p$  for warm start. Due to the fact that the variation in (7) with respect to geometric averaging is significantly smaller than in the first argument of the KL-divergence (see [6]), we have

**Theorem 1 (Evaluation of the proximal mapping).** *Let  $p \in \mathcal{G}$  be fixed and define  $R_{\mathcal{G}}(z)$  by (6), then the fixed point iteration converges for every  $z^0 \in \mathcal{G}$*

$$z_i^{m+1} = \arg \min_{z \in \mathcal{G}} \text{KL}(z_i, \text{gm}(\{z_j^m\}_{j \in \mathcal{N}(i)})) + \frac{1}{\tau\lambda} \text{KL}(z_i, p_i), \quad \forall i \in [N]. \quad (15)$$

*Proof.* We evaluate the fixed point iteration (15) and obtain

$$z_i^{m+1} = \frac{(p_i)^{\frac{1}{1+\tau\lambda}} \text{gm}(\{z_j^m\}_{j \in \mathcal{N}(i)})^{\frac{\tau\lambda}{1+\tau\lambda}}}{\langle (p_i)^{\frac{1}{1+\tau\lambda}}, \text{gm}(\{z_j^m\}_{j \in \mathcal{N}(i)})^{\frac{\tau\lambda}{1+\tau\lambda}} \rangle} \quad \forall i \in [N], \quad (16)$$

Without loss of generality we skip the intermediate normalizations and normalize only the *last* iterate since the normalization of the intermediate steps cancel out. This yields the fixed point iteration with  $p \in \mathcal{G}$  fixed and initial point  $z^0 \in \mathcal{G}$

$$z_i^{m+1} = (p_i)^{\frac{1}{1+\tau\lambda}} \text{gm}(\{z_j^m\}_{j \in \mathcal{N}(i)})^{\frac{\tau\lambda}{1+\tau\lambda}}. \quad (17)$$

Taking the logarithm of (17), substituting  $u^m = \log(z^m)$  and  $r = \log(p)$ , gives

$$u_i^{m+1} = \frac{1}{1+\tau\lambda} r_i + \frac{\tau\lambda}{1+\tau\lambda} \sum_{j \in \mathcal{N}(i)} \frac{1}{N_s} u_j^m. \quad (18)$$

Rewriting the neighborhood interactions by the associated stochastic matrix  $Q$ , with  $Q_{ij} := 1/N_s$  for  $j \in \mathcal{N}(i)$  and 0 otherwise, we get the explicit expression

$$u^{m+1} = \frac{1}{1+\tau\lambda} r + \underbrace{\frac{\tau\lambda}{1+\tau\lambda} Q}_{:= \tilde{P}} u^m = \frac{1}{1+\tau\lambda} \sum_{l=0}^m \tilde{P}^l r + \tilde{P}^{m+1} u_0. \quad (19)$$

Since  $Q$ , per definition, is a stochastic matrix and  $\tau\lambda(1+\tau\lambda)^{-1} < 1$  it follows that  $\lim_{m \rightarrow \infty} \tilde{P}^m = 0$  thus  $|\lambda_i| < 1$  holds for all eigenvalues from  $\tilde{P}$  and  $(I - \tilde{P})$  is invertible. This implies that the geometric series of the matrix  $\tilde{P}$  converges to

$$u = \lim_{m \rightarrow \infty} u^m = \frac{1}{1+\tau\lambda} (I - \tilde{P})^{-1} r. \quad (20)$$

Resubstitute the continuous functions  $z = \exp(u)$ ,  $r = \log(p)$  into (20) and normalization finally gives

$$z^* = \frac{\exp\left(\frac{1}{1+\tau\lambda} (I - \tilde{P})^{-1} \log(p)\right)}{\langle \exp\left(\frac{1}{1+\tau\lambda} (I - \tilde{P})^{-1} \log(p)\right), \mathbb{1} \rangle}, \quad (21)$$

which yields the limit point  $z^*$  independent from the starting point  $z^0$ .  $\square$

*Dual Update.* Due to the convexity of  $D$  and the standard Euclidean proximal mapping, the dual step can be evaluated in a straightforward manner.

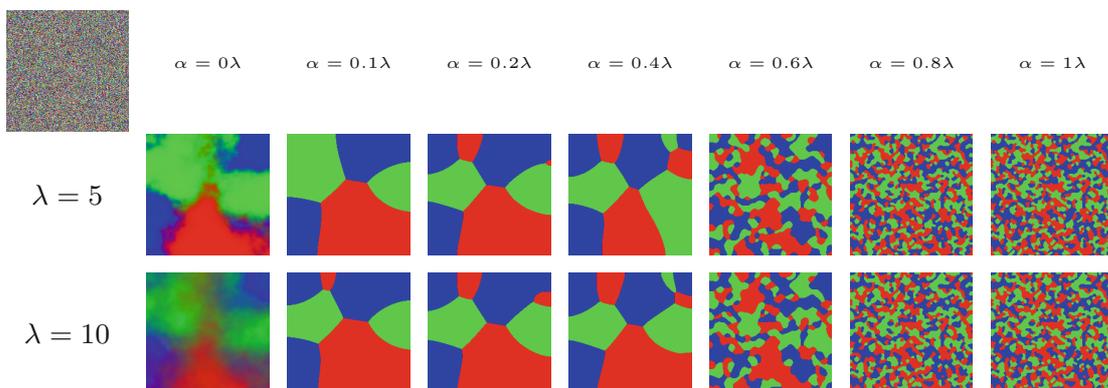
*Parameter Selection.* Following the parameter selection of [19, Example 7.2] we set  $\tau = \sqrt{K}/L_{12}^2$  in the primal update and  $\sigma = 1/\sqrt{K}$  in the dual update. Note that this parameter configuration implies that the condition  $\sigma\tau \leq \|A\|^2$  holds, where operator norm of  $A$  is given by  $L_{12} = \|A\| = \sup_{\|x\|_1 \leq 1} \|Ax\|_2 = \max_j \|A_j\|_2$  with respect to the mixed  $L_1 - L_2$ -norm. This stems from the fact that, in the primal, the negative entropy is 1-strongly convex with respect to the  $L_1$ -norm when restricted to the simplex, which induce our KL-divergence.

## 4 Experiments

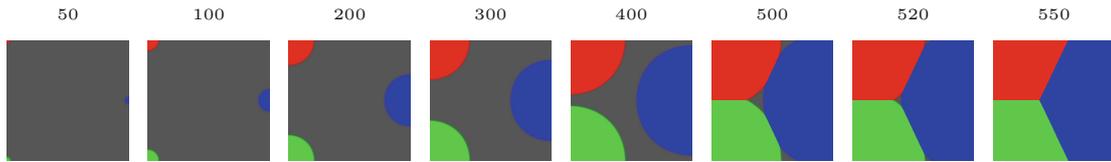
In this section we evaluate our proposed model (3) for separable and non-separable dataterms. We used a  $3 \times 3$  neighborhood system in all experiments. To guarantee fully discrete solutions we use only a simple pixelwise maximum likelihood (argmax) rounding scheme. We avoid numerical issues when evaluating the KL-divergences by adopting the renormalization strategy from [6].

**Parameter Influence.** This experiment shows the influence of the regularization parameter  $\lambda$  and discretization parameter  $\alpha$ . We generated random color noise  $u^0$  and used the dataterm  $D_{\text{unary}}(z, S) = \langle z, S \rangle$  (4) with  $S_{ik} = d(u_i^0, c_k) = \|u_i^0 - c_k\|_2^2$  where the labels  $c_k \in \mathcal{L} = \{\text{red}, \text{green}, \text{blue}\}$ . Figure 2 shows that larger  $\lambda$  acts as a smoothing parameter enforcing larger constant regions, whereas  $\alpha$  favors consistency over the discrete label space, i.e., the data range. The presence of the entropy term is promoting an integral solution as illustrated in the left most-column. In this experiment no rounding was applied.

**Interface Propagation.** In this example we illustrate the information propagation in the case when the dataterm is uninformative. We use the same model configuration and  $\mathcal{L}$  as the in previous experiment with  $\lambda = 10$  and  $\alpha = 1$ .



**Fig. 2.** Random color noise, see top left corner, is labeled with  $\mathcal{L} = \{\text{red}, \text{green}, \text{blue}\}$  with varying regularization parameter  $\lambda$  and discretization parameter  $\alpha$ . (Color figure online)



**Fig. 3.** The evolution of the interfaces are shown for increasing iteration number with the label set  $c_k \in \mathcal{L} = \{\text{red}, \text{green}, \text{blue}\}$ . The interfaces are propagated uniformly from the three seed-pixels into the uninformative image until they meet in a triple junction. (Color figure online)

We set the input data  $u^0$  to a constant gray image with three seed pixels: one (red)-pixel in the top left corner, one (green)-pixel in the bottom left corner and one (blue)-pixel in the middle of the right edge. Figure 3 displays the evolution of the interfaces for increasing iterations. We see that the information given by the three seed-pixels is uniformly propagated into the image until the interfaces meet in a triple junction which demonstrates uniform propagation speeds. In this experiment no rounding was applied.

**Joint Deblurring and Labeling.** We used the non-separable data term  $D_{\text{inverse}}$  (5) implemented by the  $L_1$ -norm and the self-assignment term was extended component-wise to each color channel. The label set  $\mathcal{L} = \{\text{red}, \text{green}, \text{blue}, \text{yellow}, \text{cyan}, \text{magenta}\}$  was generated by K-means clustering with 6 cluster of the original image seen in Fig. 1. The same figure shows the reconstruction of a severe blurred picture of an insect (motion blur of 65 pixel length) and joint labeling with two different parameter configurations: high regularization and low regularization. In a more challenging setting we additionally corrupt 50% pixels of the blurred image with random colors drawn from a uniform distribution. In both cases we reconstruct fine details of the original image.

**Discrete Tomography Reconstruction.** The reconstruction problem in discrete tomography aims to recover an image  $u \in \mathbb{R}^N$  from a small number of possibly noisy measurements  $b = Au + \nu \in \mathbb{R}^m$ . The latter correspond to line integrals that sum up all absorptions over each ray transmitted through the object. A given projection matrix  $A \in \mathbb{R}^{m \times N}$  encodes the imaging geometry, here we used the parallel beam setup. The width of the sensor-array was set 1.5 times the image size, so that every pixel intersects with a least a single projection ray. We used the non-separable data term  $D_{\text{inverse}}$  (5) implemented by the indicator function to enforce the constraints (see [10] for details). We compare our model to state-of-the-art approaches for non-binary discrete tomography in limited angles scenarios. Specifically, we considered DART [20], the energy minimization method from Varga et al. [21] (Varga) and [10] (LayerTV) with a layer-wise total variation regularizer.

*Setup.* For the evaluation we measured the relative pixel error, that is the relative number of erroneously reconstructed pixels as compared to the groundtruth. We tried to use the default parameters of the competing approaches as proposed by

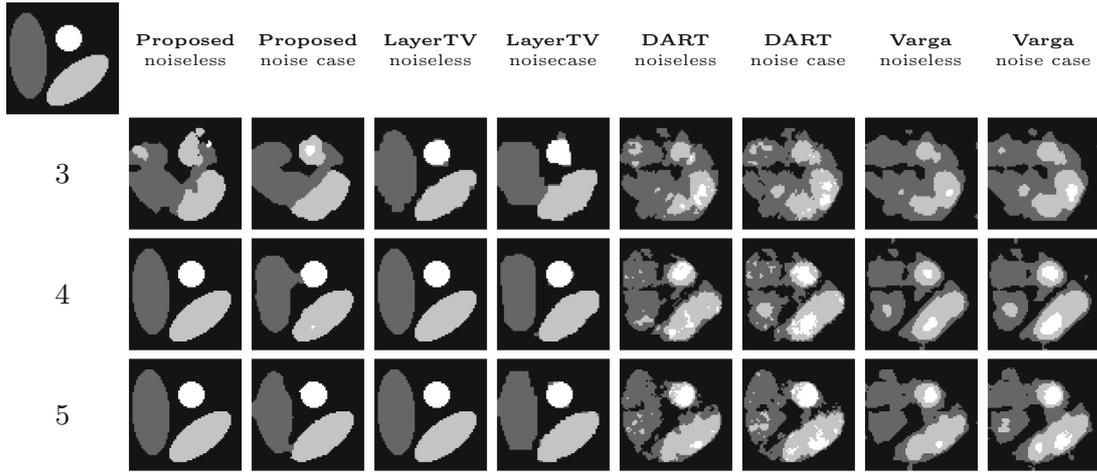


Fig. 4. Visual results of experiment with ellipses phantom.

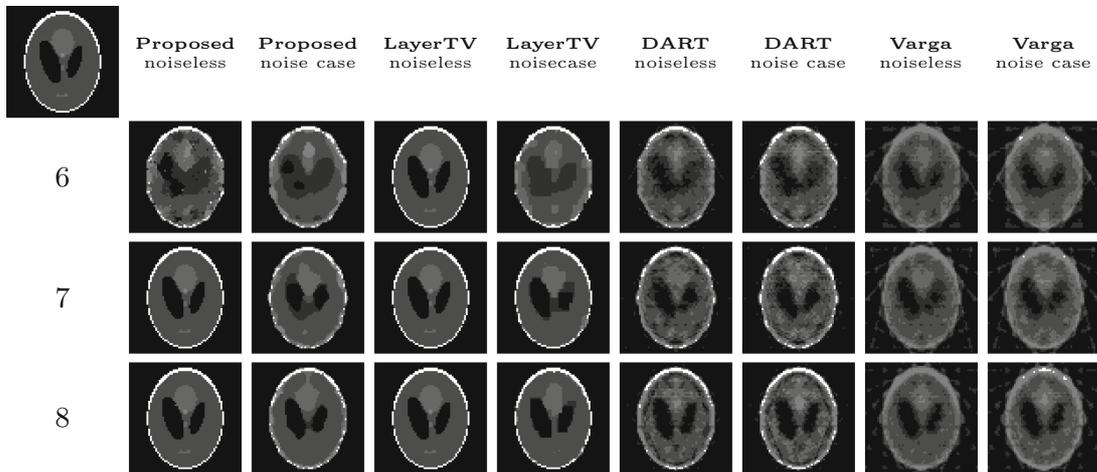
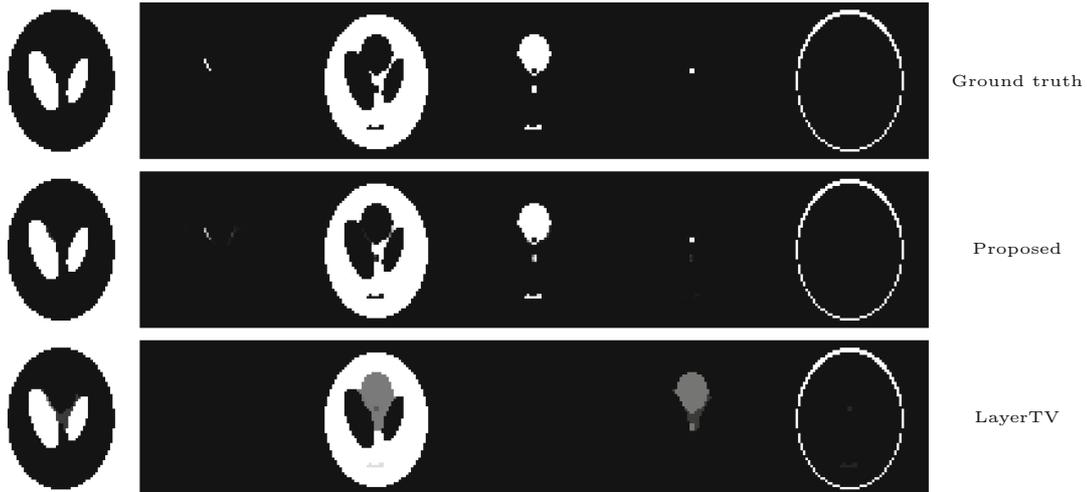


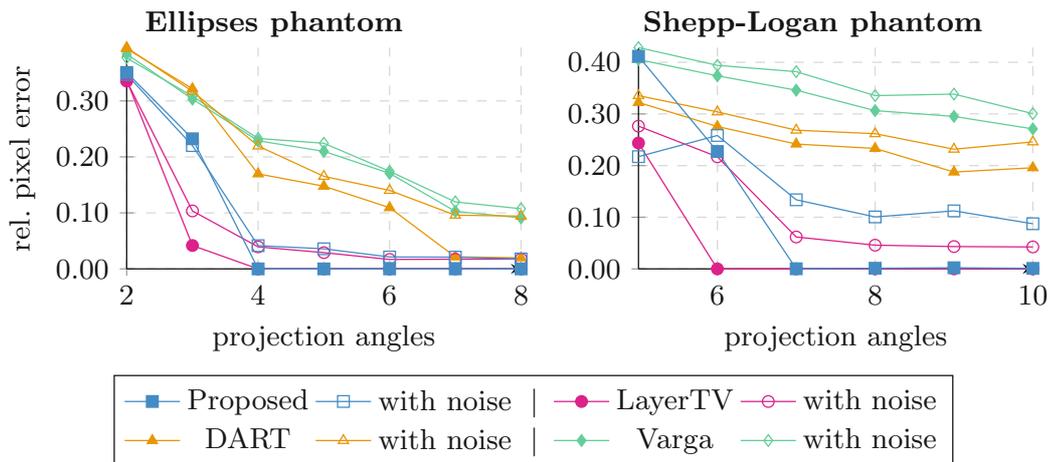
Fig. 5. Visual results of experiment with Shepp-Logan phantom.

their authors. However, the test-datasets differ in size, we slightly adjusted the parameters to get best results for every algorithm and problem instance.

*Results.* In Figs. 5 and 4 the proposed approach gives perfect reconstructions with a low number of projection angles in the noiseless case and also returns high-quality reconstructions in the presence of noise, only LayerTV needs one projection less however a non-trivial rounding strategy is used. This is depicted in Fig. 6 for the Shepp-Logan phantom from 7 projections where LayerTV clearly gives a non-integral solution and requires a special rounding strategy to obtain a meaningful reconstruction, further details are given in the caption. Figure 7 shows the numerical evaluation of the approaches for increasing (but small) numbers of projections, in the noiseless case (filled markers) and in the noisy case (non-filled markers), with Poisson noise  $SNR = 20$  dB.



**Fig. 6.** Reconstruction of the Shepp-Logan phantom from 7 projections, where the indicator variables  $z^k$  are shown for each layer  $k \in [K]$  from left to right. White denotes the selected label. The LayerTV produces a non-integral solution with a convex combination of the labels - illustrating the need for rounding, whereas our proposed model directly gives an unique labeling.



**Fig. 7.** Relative pixel error compared to the number of projections. In the Shepp-Logan phantom LayerTV can reconstruct the phantom with one less projection, however a special rounding strategy is performed to obtain a meaningful solution.

## 5 Conclusion

In this work we presented a novel variational approach to joint image reconstruction and labeling. Opposed to state of the art reconstruction algorithms which use intra layer coupling strategies, or basic Euclidean inter layer coupling, we have instead derived the first inter layer coupling which preserves the information geometric properties of the underlying statistical manifold. Additionally, we have shown that the evaluation of a generalized proximal mapping, relying on the

KL-divergence, can be efficiently evaluated. The numerical evaluation illustrate the competitiveness of our approach compared to state of the art discrete tomography reconstruction and deblurring and denoising with joint labeling.

## References

1. Hanke, R., Fuchs, T., Uhlmann, N.: X-ray based methods for non-destructive testing and material characterization. *Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrom. Detect. Assoc. Equip.* **591**(1), 14–18 (2008)
2. Zach, C., Gallup, D., Frahm, J., Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: *VMV*, pp. 243–252 (2008)
3. Chambolle, A., Cremers, D., Pock, T.: A convex approach to minimal partitions. *SIAM J. Imaging Sci.* **5**(4), 1113–1158 (2012)
4. Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) *SSVM 2009. LNCS*, vol. 5567, pp. 150–162. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-02256-2\\_13](https://doi.org/10.1007/978-3-642-02256-2_13)
5. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sci.* **4**(4), 1049–1096 (2011)
6. Åström, F., Petra, S., Schmitzer, B., Schnörr, C.: Image labeling by assignment. *J. Math. Imaging Vis.* **58**, 1–28 (2017)
7. Bergmann, R., Fitschen, J.H., Persch, J., Steidl, G.: Iterative multiplicative filters for data labeling. *Int. J. Comput. Vis.*, 1–19 (2017)
8. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings of the ICCV*, pp. 416–423 (2001)
9. Zisler, M., Kappes, J.H., Schnörr, C., Petra, S., Schnörr, C.: Non-binary discrete tomography by continuous non-convex optimization. *IEEE Trans. Comput. Imaging* **2**(3), 335–347 (2016)
10. Zisler, M., Petra, S., Schnörr, C., Schnörr, C.: Discrete tomography by continuous multilabeling subject to projection constraints. In: Rosenhahn, B., Andres, B. (eds.) *GCPR 2016. LNCS*, vol. 9796, pp. 261–272. Springer, Cham (2016). doi:[10.1007/978-3-319-45886-1\\_21](https://doi.org/10.1007/978-3-319-45886-1_21)
11. Kass, R.E.: The geometry of asymptotic inference. *Stat. Sci.* **4**(3), 188–234 (1989)
12. Amari, S.I., Cichocki, A.: Information geometry of divergence functions. *Bull. Pol. Acad. Sci.: Tech. Sci.* **58**(1), 183–195 (2010)
13. Potts, R.B.: Some generalized order-disorder transformations. *Math. Proc. Camb. Philos. Soc.* **48**, 106–109 (1952)
14. Weinmann, A., Demaret, L., Storath, M.: Total variation regularization for manifold-valued data. *SIAM J. Imaging Sci.* **7**(4), 2226–2257 (2014)
15. Cover, T., Thomas, J.: *Elements of Information Theory*, 2nd edn. Wiley, Hoboken (2006)
16. Pham Dinh, T., El Bernoussi, S.: Algorithms for solving a class of nonconvex optimization problems. *Methods of subgradients*. In: Hiriart-Urruty, J.B. (ed.) *Fermat Days 85: Mathematics for Optimization*. North-Holland Mathematics Studies, vol. 129, pp. 249–271. North-Holland, Amsterdam (1986)
17. Pham-Dinh, T., Hoai An, L.: Convex analysis approach to D.C. programming: theory, algorithms and applications. *Acta Math. Vietnam.* **22**(1), 289–355 (1997)
18. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*. Springer, Heidelberg (2009)

19. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. *Math. Program.* **159**(1), 253–287 (2016)
20. Batenburg, K., Sijbers, J.: DART: a practical reconstruction algorithm for discrete tomography. *IEEE Trans. Image Process.* **20**(9), 2542–2553 (2011)
21. Varga, L., Balázs, P., Nagy, A.: An energy minimization reconstruction algorithm for multivalued discrete tomography. In: *3rd International Symposium on Computational Modeling of Objects Represented in Images, Italy*, pp. 179–185 (2012)