Discrete Tomography by Continuous Multilabeling Subject to Projection Constraints

Anonymous GCPR 2016 submission

Paper ID ***

Abstract. We present a non-convex variational approach to non-binary discrete tomography which combines non-local projection constraints with a continuous convex relaxation of the multilabeling problem. Minimizing this non-convex energy is is achieved by a fixed point iteration which amounts to solving a sequence of convex problems, with guaranteed convergence to a critical point. A competitive numerical evaluation using standard test-datasets demonstrates a significantly improved reconstruction quality for noisy measurements from a small number of projections.

1 Introduction

Computed tomography [14] deals after spatial discretization in an algebraic set-up with the reconstruction of 2D- or 3D-images $u \in \mathbb{R}^N$ from a small number of 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 this imaging geometry. Applications range from medical imaging [3] to natural sciences and industrial applications, like non-destructive material testing [7]. Many situations require to keep the number of measurements as low as possible, which leads to a small number of projections and hence to a severely ill-posed reconstruction problem.

To cope with such problems, a common assumption in the field of discrete tomography [8] concerns knowledge of a finite range of $u \in \mathcal{L}^N$, $\mathcal{L} := \{c_1, ..., c_K\} \subset$ [0, 1], that is, u represents a piecewise constant function. Our main concern in this paper is to effectively exploit the additional prior knowledge in terms of \mathcal{L} , besides the projection constraints, in order to solve the discrete reconstruction problem

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

which generally is a NP-hard problem.

Related work on discrete tomography considers either *binary* or *non-binary* (multivalued) problems. The latter ones are considerably more involved.

Regarding *binary discrete tomography*, Weber et al. [28, 20] proposed to combine a quadratic program with a non-convex penalty which gradually enforces binary constraints. More recently, Kappes et al. [9] showed how a binary discrete graphical model and a sequence of s-t graph-cuts can be used to take into account the affine projection constraints and to recover high-quality reconstructions.
 $\mathbf{2}$

Regarding *non-binary discrete tomography*, an extension of the latter approach is not straightforward due to the nonlocal projection constraints. Weber [27, Chapter 6] proposed a non-convex term for non-binary discrete tomography that we derive in a natural way in the present work. However, Weber's 41 approach differs with respect to the data term for the projection constraints, 42 regularization and optimization, and additionally requires parameter tuning. 43

Because u is assumed to be piecewise constant, an obvious approach is to consider sparsity promoting priors. The authors of [22] proposed a dynamic pro-gramming approach for minimizing the ℓ_0 -norm of the gradient. The set \mathcal{L} of feasble intensities is not exploited, however. In the convex setting, the integrality constraints are dropped and priors like the ℓ_1 -norm or the total variation (TV) are used [21, 6, 5], with a postprocessing step to round the continuous solution to a piecewise constant one. This approach connects discrete tomography and the fast evolving field of *compressive sensing* with corresponding recovery guar-antees [5]. Again, however, the prior information of the range of the image to be reconstructed is not involved in the optimization process. We focus next on methods that make use of the set \mathcal{L} during the reconstruction process.

Tuysuzoglu et al. [24] casted the non-binary discrete reconstruction problem into a series of submodular binary problems within an α -expansion approach by linearizing the ℓ_2 -fidelity term around an iteratively updated working point. This local approximation discards a lot of information, and a significantly larger number of projections is required to get reasonable reconstructions. Maeda et al. [12] suggested a probabilistic formulation which couples a continuous recon-struction with the Potts model. Alternating optimization is applied to maximize the *a posteriori* probability locally. However, there is no convergence that these alternating continuous and discrete block coordinate steps converge.

Ramlau et al. [10] investigated the theoretical regularization properties of the piecewise constant Mumford-Shah functional [13] applied to linear ill-posed problems. In earlier work [19], they considered discrete tomography reconstruc-tion using this framework. The difficult geometric optimization of the partition is carried out by a level-set approach and additionally the intensities \mathcal{L} were estimated in an alternating fashion. By contrast, our approach is based on a convex relaxation of the perimeter regularization and the set \mathcal{L} is assumed to known beforehand.

Varga et al. [25] suggested a heuristic algorithm which is adaptively comb-ing an energy formulation with a non-convex polynomial in order to steer the reconstruction towards the feasible values. Batenburg et al. [2] proposed the Dis-crete Algebraic Reconstruction Technique (DART) algorithm which starts with a continuous reconstruction by a basic algebraic reconstruction method, followed by thresholding to ensure a piecewise constant function. These steps interleaved with smoothing are iteratively repeated to refine the locations where u jumps. This heuristic approach yields good reconstructions in practice but cannot be characterized by an objective function that is optimized.

We regard [2, 25] as state-of-the-art approaches for the experimental comparison.

Contributions. We present a novel variational approach to the discrete tomography reconstruction problem in the general non-binary case. Contrary to existing work, we utilize both the non-local projection constraints and the feasible set of intensities \mathcal{L} in connection with an established convex relaxation of the multilabeling approach with a Potts prior. We show how the resulting non-convex overall energy can be optimized efficiently by a fixed-point iteration which requires to solve a convex problems at each step. In this way, the derivation of our non-convex data and its local updates arise naturally. We also propose a suitable rounding procedure as post-processing step, because the integrality constraints are relaxed. A comprehensive numerical evaluation demonstrates the superior reconstruction performance of our approach compared to related work.

⁹⁴ 2 Reconstruction by Constrained Multilabeling

In this section, we first reformulate the discrete reconstruction problem (1) as
a constraint combinatorial multilabeling problem. Then we derive a tractable
variational approximation and suggest a proper rounding procedure.

98 2.1 Constrained Multilabeling Problem

We assume that there are less measurements than pixels $m \ll N$ and hence that the discrete reconstruction problem (1) is ill-posed and requires regularization. A common choice is the Potts model [18], $R(u) = \|\nabla u\|_0 := |\{i \mid (\nabla u)_i \neq 0\}|$ for sparse gradient regularization which favours piecewise constant images. In presence of noisy measurements b, we use the more general constraints $\underline{b}(\epsilon) \leq$ $Au \leq b(\epsilon)$ instead of Au = b, where ϵ is an upper bound of the noise level. As a result, the discrete reconstruction problem can be rewritten as

$$E(u) = \lambda \cdot \|\nabla u\|_0 \quad \text{s.t.} \quad \underline{b}(\epsilon) \le Au \le \overline{b}(\epsilon) \quad \land \quad u_i \in \mathcal{L} \quad \forall i = 1...N.$$
(2)

We refer to problem (2) as a *constrained multilabeling problem* with Potts regu-larization but point out that, from the viewpoint of graphical models, the system of affine inequalities induces (very) high-order potentials. This high-order inter-action induced by the non-local constraints results in a non-standard labeling problem which becomes intractable for discrete approaches and larger problem sizes. We adopt, therefore, the strategy of solving a sequence of convex relax-ations in order to minimize a non-convex energy, which properly approximates the original problem.

114 2.2 Approximate Variational Problem

Our starting point is the established convex relaxation of the multilabeling prob-

lem [29, 11, 17]. Minimization (3) below over relaxed indicator vectors z_i assigns to each given image pixel u_i^0 a label of the set $\mathcal{L} = \{c_1, \ldots, c_K\}$. The discretized

total variation penalty, weighted by λ , and the simplex constraints G constitute 118 a basic convex relaxation of the integrality constraints with respect to z. 119

$$E(z, u^{0}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left(u_{i}^{0} - c_{k} \right)^{2} + \lambda \sum_{k=1}^{K} \| \nabla z_{k} \|_{1}$$

s.t. $z \in G := \left\{ z \in [0, 1]^{N \times K} : \sum_{k=1}^{K} z_{ik} = 1, \forall i = 1, \dots, N \right\}.$ (3)

Regarding the notation, we denote by $z_k, k \in \{1, ..., K\}$ the column vectors of z and by $z_{ik} = (z_k)_i$ the entries of the matrix z. $Z_{ik} = (z_k)_i$ the entries of the matrix z. $Z_{ik} = (z_k)_i$ the entries of the matrix z.

$$E(z, u^{0}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left(u_{i}^{0} - c_{k} \right)^{2} + \lambda \sum_{k=1}^{K} \|\nabla z_{k}\|_{1} + \delta_{\mathbb{R}^{m}_{+}} (AWz - \underline{b}) + \delta_{\mathbb{R}^{m}_{-}} (AWz - \overline{b}) + \delta_{G}(z).$$
(4)

126 Note that the constraints $\underline{b} \leq AWz \leq \overline{b}$ and $z \in G$ are implemented by indicator 127 functions $\delta_{\mathbb{R}^m_+}$ and $\delta_{\mathbb{R}^m}$. 127 127 127

In tomography, no image u^0 is given, however. Therefore, we cannot drop the unary data term in Eq. (4), however, since the constraints are feasible for all convex combinations of prototypes c_k . In other words, the constraints only constrain the value a pixel but do not indicate how the indicator variables should realize this value (similar to estimating a vector given only its magnitude).

133A straightforward approach would be to start with some initial guess u^0 , e.g.133134computed using some another reconstruction method, followed by iteratively134135applying this approach above. This gives the fixed point iteration135

$$z^{n+1} = \operatorname*{arg\,min}_{z} E(z, Wz^n). \tag{5}$$

At every iteration a convex problem has to solved whose solution updates the unary data term. This raises the question whether the iteration convergences and which overall energy is actually optimized?

To address these questions, we first eliminate u^0 in a principled way. Noting 139 139 that the term of $E(z, u^0)$ in Eq. (4) depending from u^0 is differentiable with 140 140 respect to u^0 , we invoke Fermat's (first order) optimality condition $u^* = Wz$ 141 141 which says that the optimal u^* must be equal to the weighted average of the 142 142 labels c_k . Substituting this optimality condition back into the energy (4) results 143 143 in the final version of the proposed energy which only depends on z, 144 144

$$E(z) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left((Wz)_i - c_k \right)^2 + \lambda \sum_{k=1}^{K} \|\nabla z_k\|_1 + \delta_{\mathbb{R}^m_+} (AWz - \underline{b}) + \delta_{\mathbb{R}^m_-} (AWz - \overline{b}) + \delta_G(z).$$
(6)

145 146 This energy, Eq. (6), is non-convex because of the products in the first term 145 which is measuring the discreteness of z. We call this term *phase data term* and 146 denote it by

$$D(z) := \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left((Wz)_i - c_k \right)^2.$$
(7)

Fig. 1. Visualization of the phase data term term D(z) for N = 1 over the probability simplex G, the vertices correspond to the values $\mathcal{L} = \{0.0, 0.4, 1.0\}$. Note that the minimum is attained at the vertices of the simplex which correspond to unit vectors.



Figure 1 shows a plot of D(z), Eq. (7). Weber [27, Chapter 6] proposed this term for discrete tomography which arises here in a natural way, whereas his overall approach differs with respect to the data term for the projection constraints, regularization and optimization. We show in the Appendix A, as Lemma 2, that D(z), Eq. (7), is concave over the domain G.

153 **3** Optimization

In this section, we reformulate the objective function (6) as DC program and work out a corresponding optimization algorithm.

156 **3.1** DC Programming

A large subclass of non-convex optimization problems are DC functions (difference of convex functions) which can be solved by DC Programming [15], this generalizes subgradient optimization of convex functions to local optimization of DC functions. Accordingly, basic concepts of convex optimization like duality and KKT conditions were extended to DC functions [23]. The basic form of a DC Program is given by

$$z^* = \operatorname*{arg\,min}_{z} g(z) - h(z), \tag{8}$$

where g(z) and h(x) are proper, lower semicontinuous, convex functions. There exists a simplified version of the DC algorithm [16] for minimizing (8) which guarantees convergence to a critical point by starting with $z^0 \in \text{dom}(g)$ and then alternatingly applying the updates

$$v^n \in \partial h(z^n)$$
 and $z^{n+1} \in \partial g^*(v^n)$ (9)

until a termination criterion is reached, where g^* denotes the Legendre-Fenchel 167 conjugate of g. To apply the DC algorithm to our non-convex energy E(z), Eq. 168

153

154

155

156

157

158

159

160

161

162

147

148

149

150

151

169 (6), we rewrite E(z) = g(z) - h(z) as a DC function where h(z) = -D(z) since 169 the phase data term (7) is concave (see Appendix A: Lemma 2) and g(z) is 170 denoting the remaining convex terms from Eq. (6). 171 In order to make explicit the step $z^{n+1} \in \partial q^*(v^n)$, we apply the subgradient 172 In order to make explicit the step $z^{n+1} \in \partial q^*(v^n)$, we apply the subgradient

In order to make explicit the step $z^{n+1} \in \partial g^*(v^n)$, we apply the subgradient inversion rule of convex analysis to obtain

$$z^{n+1} \in \partial g^*(v^n) \quad \Leftrightarrow \quad v^n \in \partial g(z^{n+1}) \quad \Leftrightarrow \quad 0 \in \partial g(z^{n+1}) - v^n$$
(10)

174 which is equivalent to the convex optimization problem

$$z^{n+1} = \arg\min_{z} g(z) - \langle v^n, z \rangle.$$
(11)

175 Because h is differentiable, the other step of (9) reads

$$v^n \in \partial h(z^n) \quad \Leftrightarrow \quad v^n = -\nabla D(z^n),$$
 (12)

where the gradient of D at z for pixel i and label c_k is given by (see Appendix A: Lemma 1) 177

$$(\nabla D(z))_{ik} = \frac{\partial D(z)}{\partial z_{ik}} = ((Wz)_i - c_k)^2, \qquad i = 1, \dots, N, \quad k = 1, \dots, K.$$
 (13)

178 Combining equations (12) and (13) and inserting into equation (11) yields

$$z^{n+1} = \arg\min_{z} E(z, Wz^n) = \arg\min_{z} g(z) + \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left((Wz^n)_i - c_k \right)^2. \quad (14)$$

We notice that the DC algorithm, summarized as Algorithm 1 below, agrees with the iteration (5), and hence proves its convergence. We apply the primal

Algorithm 1: DC Fixed Point Algorithm

1. Initialization: choose any $z^0 \in \mathbb{R}^{n \times k}$

2. Generate a sequence $(z^n)_{n \in \mathbb{N}}$ by solving the convex problems

$$z^{n+1} = \arg\min_{z} E(z, Wz^n) \tag{15}$$

until a termination criterion is met.

180		180
181	dual (PD) algorithm proposed by $[4]$ to solve each convex subproblem (15) .	181

182 3.2 Rounding Step

183	Recall that the data term $D(z)$ of (6) only steers the solution to the finite set	183
184	of feasible values \mathcal{L} . As a consequence, for vanishing regularization parameter	184

175

173

174

182

(weight) λ , the minimizer z will correspond to indicator vectors z_i that assign a unique label to each pixel i. For larger values of λ which are more common in practice, however, the minimizing vectors z_i will not be integral in general. Therefore, a post-processing step for rounding the solution is required. Given the minimizer z^* of (6), we propose to select a label for each pixel i189

Given the minimizer z^* of (6), we propose to select a label for each pixel ias post-processing step by solving the local problems

$$\hat{u}_i^* = \operatorname*{arg\,min}_{c \in \mathcal{L}} |(Wz^*)_i - c|, \qquad i = 1, \dots, N.$$
 (16)

191 Note that this method differs from the common rounding procedure of multi-192 labeling approaches which select the label c_k if $z_{ik} = \max\{z_{i1}, \ldots, z_{iK}\}$. 192 193

193 4 Numerical Experiments

193

197

198

199

200

201

202

203

204

205

206

207

190

Set-up. In this section, we compare our approach to state-of-the-art approaches
for non-binary discrete tomography in limited angles scenarios. Specifically, we
considered the Discrete Algebraic Reconstruction Technique (DART) [2] and
the energy minimization method from Varga et al. [25] (Varga). As multivalued



Fig. 2. The 5 different phantoms used for the numerical evaluation.

197

test-datasets we adapted the binary phantoms from Weber et al. [26] to more 198 labels, shown as phantom 1,2 and 3 by Figure 2. Phantom 4 in Figure 2 was 199 taken from [2] and phantom 5 is the well-known Shepp-Logan phantom. We 200 created noisy scenarios by applying Poisson noise to the measurements b with a 201 signal-to-noise ratio of SNR = 20 db. The geometrical setup was created by the 202 ASTRA-toolbox [1], where we used parallel projections along equidistant angles 203 between 0 and 180 degrees. Each entry a_{ij} of the matrix A corresponds to the 204 length of the line segment of the *i*-th projection ray passing through the j-th 205 pixel in the image domain. The width of the sensor-array was set 1.5 times the 206 image size, so that every pixel intersects with a least a single projection ray. 207

Implementation details. Each subproblem of Algorithm 1 was approxi-208 208 mately solved using the primal dual (PD) algorithm [4] limited to 1000 iterations 209 209 or till the primal dual gap dropps below 0.1. The outer iteration was terminated 210 210 if change of the energy between two iterations, normalized by the number of 211 211 pixels, was smaller than 10^{-5} in the noiseless case and 10^{-4} in the noisy case. 212 212 Additionally, we limited the number of outer iterations to 20. For DART we 213 213



Fig. 3. Visual results of experiment phantom 1.

used the publicly available implementation included in the ASTRA-toolbox [1] and for the method of Varga [25] we used our own implementation in MATLAB because no public code was available. We tried to use the default parameters of the competing approaches as proposed by their authors. However, since the test-datasets differ in size, we slightly adjusted the parameters in order to get best results for every algorithm and problem instance.

Performance measure. For the evaluation we measured the relative pixel error, that is the relative number of erroneously reconstructed pixels as compared to the groundtruth.

Results. Figure 4 shows all results of the numerical evaluation. For each test-datasets (phantoms 1 - 5), the left plot displays the relative pixel error for increasing numbers of projection angles. On the right, the corresponding runtime is shown as log-scaled plot. For each algorithm two curves are drawn: filled makers correspond to the noiseless case and non-filled markers correspond to the noisy case. The results show that the proposed approach returns a perfect reconstruction with the least number of projection angles in the noiseless case among all approaches. In the scenarios, the proposed algorithm is performing better, too. In the noiseless case phantom 1 can be almost perfectly reconstructed from only 3 projection angles and fully from 4 by the proposed approach whereas DART needs 7 projection to get an almost perfect reconstruction and the method of Varga needs at least 7 projections to get a reasonable reconstruction. These visual differences can be seen in Figure 3. Thhis ranking of the performance of the approaches is similar for phantoms 2,3 and 5, except for the phantom 5 in



Fig. 4. Numerical evaluation of the approaches for the different test-datasets and increasing (but small) numbers of projections, in the noiseless case (filled markers) and in the noisy case (non-filled markers), with noise level SNR = 20 db. The relative pixel error is shown. The proposed approach gives perfect reconstructions with the

least number of projection angles in the noiseless case and also returns high-quality

reconstructions in the presence of noise, compared to the other approaches.

with noise

 \rightarrow Varga \rightarrow with noise

DART

Proposed \rightarrow with noise



Fig. 5. Visual results of experiment phantom 4.

the noiseless case where the approach of Varga performs better than DART. Figure 5 shows the results for phantom 4, where our approach is able to fully reconstruct from merely 7 projections in the noiseless case and returns a good piecewise-constant result in the noisy case. Due to a lack of space, we refer to Appendix B of the supplementary material for the visualization of the results for all test-datasets.

Regarding the runtime (right plots from figure 4), DART is the fastest approach, Varga is in between DART, and the proposed approach is clearly consuming more runtime to return more accurate solutions. Obviously, if computational performance is important, the proposed approach could be easily parallelized and implemented e.g. in CUDA to run on modern graphics cards.

²⁴⁸ 5 Conclusion and Future Work

We presented a novel non-convex variational approach for solving the discrete tomography reconstruction in the general non-binary case. The approach com-bines a convex relaxation of the multilabeling problem with Potts prior and the non-local tomographic projection constraints. The feasible set of labels is taken into account by a non-convex data term which naturally emerges when the func-tion to be reconstructed is represented as convex combination of these values. A DC algorithm reliable minimizes the overall objective function and provably converges. The reconstruction performance turned out to be superior to the state of the art.

In future work, we plan to improve the running time and focus on the theoretical aspects of this approach. The proposed data term (7), in particular, fits nicely into spatially continuous variational formulations and thus may indicate ways for further improvement.

262 References

263	1.	van Aarle, W., Palenstijn, W.J., Beenhouwer, J.D., Altantzis, T., Bals, S., Baten-	263
264		burg, K.J., Sijbers, J.: The {ASTRA} Toolbox: A platform for advanced algo-	264
265		rithm development in electron tomography. Ultramicroscopy 157, 35 – 47 (2015).	265
266		http://www.sciencedirect.com/science/article/pii/S0304399115001060	266
267	2.	Batenburg, K., Sijbers, J.: DART: A Practical Reconstruction Algorithm for Dis-	267
268		crete Tomography. Image Processing, IEEE Transactions on 20(9), 2542–2553 (Sept	268
269		2011)	269
270	3.	Bushberg, J., Seibert, J., Leidholdt, E., Boone, J.: The Essential Physics of Medical	270
271		Imaging, Wolters Kluwer, 3rd edn. (2011)	271
272	4.	Chambolle, A., Pock, T.: A First-Order Primal-Dual Algorithm for Convex Prob-	272
273		lems with Applications to Imaging, J. Math. Imaging Vis. 40(1), 120–145 (May	273
274		2011)	274
275	5	Denitiu A Petra S Schnörr C Schnörr C Phase Transitions and Cosparse	275
276	0.	Tomographic Recovery of Compound Solid Bodies from Few Projections Funda-	275
277		Tomographic receivery of compound solid bodies from Lew Trojections. Fundational Informatical 135, $73-102$ (2014)	270
278	6	Coris B den Broek WV Batenburg K Mezerii HH Bals	211
270	0.	S: Electron tomography based on a total variation minimization	270
279		s. Electron tomography based on a total variation minimization	219
280		reconstruction technique. Ultrannicroscopy 115, 120 – 150 (2012),	280
281	7	Huple, P. Fucha, T. Uhlmann, N. Y. raw based methods for non-destructive test	281
282	1.	inanke, R., Fuchs, T., Olimiani, N.: A-ray based methods for hom-destructive test-	282
283		ing and material characterization. Nuclear Instruments and Methods in Physics	283
284		Research Section A: Accelerators, Spectrometers, Detectors and Associated Equip-	284
285	0	ment $591(1), 14 - 18 (2008)$	285
286	8.	Herman, G., Kuba, A.: Discrete Tomography: Foundations, Algorithms and Ap-	286
287	0	plications. Birkhauser (1999)	287
288	9.	Kappes, J.H., Petra, S., Schnorr, C., Zisler, M.: TomoGC: Binary Tomography by	288
289	10	Constrained GraphCuts. Proc. GCPR 30	289
290	10.	Klann, E., Ramlau, R.: Regularization properties of mumford-shah-type function-	290
291		als with perimeter and norm constraints for linear ill-posed problems. SIAM Jour-	291
292		nal on Imaging Sciences $6(1)$, $413-436$ (2013)	292
293	11.	Lellmann, J., Becker, F., Schnorr, C.: Convex optimization for multi-class image	293
294		labeling with a novel family of total variation based regularizers. In: Proceedings of	294
295		the IEEE Conference on Computer Vision (ICCV 09) Kyoto, Japan. pp. 646–653	295
296		(2009), 1	296
297	12.	Maeda, S., Fukuda, W., Kanemura, A., Ishii, S.: Maximum a posteriori X-ray	297
298		computed tomography using graph cuts. Journal of Physics: Conference Series 233	298
299		(2010)	299
300	13.	Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions	300
301		and associated variational problems. Communications on Pure and Applied Math-	301
302		ematics $42(5), 577-685$ (1989)	302
303	14.	Natterer, F., Wübbeling, F.: Mathematical Methods in Image Reconstruction.	303
304		SIAM (2001)	304
305	15.	Pham Dinh, T., El Bernoussi, S.: Algorithms for Solving a Class of Nonconvex Opti-	305
306		mization Problems. Methods of Subgradients. In: Hiriart-Urruty, J.B. (ed.) Fermat	306
307		Days 85: Mathematics for Optimization, North-Holland Mathematics Studies, vol.	307
308		129, pp. 249 – 271. North-Holland (1986)	308
309	16.	Pham Dinh, T., Hoai An, L.: Convex Analysis Approach to D.C. Programming:	309
310		Theory, Algorithms and Applications. Acta Math. Vietnamica 22(1), 289–355	310
311		(1997)	311

312	17.	Pock, T., Chambolle, A., Cremers, D., Bischof, H.: A convex relaxation approach	312
313		for computing minimal partitions. In: Computer Vision and Pattern Recognition,	313
314		2009. CVPR 2009. IEEE Conference on. pp. 810–817 (June 2009)	314
315	18.	Potts, R.B.: Some generalized order-disorder transformations. Mathematical Pro-	315
316		ceedings of the Cambridge Philosophical Society 48, 106–109 (1 1952)	316
317	19.	Ramlau, R., Ring, W.: A mumford-shah level-set approach for the inversion and	317
318		segmentation of x-ray tomography data. Journal of Computational Physics 221(2),	318
319		539-557 (2007)	319
320	20.	Schüle, T., Schnörr, C., Weber, S., Hornegger, J.: Discrete tomography by con-	320
321		vexconcave regularization and D.C. programming. Discrete Applied Mathematics	321
322		$151(13), 229 - 243 \ (2005)$	322
323	21.	Sidky, E.Y., Pan, X.: Image reconstruction in circular cone-beam computed to-	323
324		mography by constrained, total-variation minimization. Physics in Medicine and	324
325		Biology 53(17), 4777 (2008), http://stacks.iop.org/0031-9155/53/i=17/a=021	325
326	22.	Storath, M., Weinmann, A., Frikel, J., Unser, M.: Joint image reconstruction and	326
327		segmentation using the Potts model. Inverse Problems 31(2), 025003 (2015)	327
328	23.	Toland, J.: Duality in nonconvex optimization. Journal of Mathematical Analysis	328
329		and Applications 66(2), 399–415 (1978)	329
330	24.	Tuysuzoglu, A., Karl, W., Stojanovic, I., Castanon, D., Unlu, M.: Graph-Cut Based	330
331		Discrete-Valued Image Reconstruction. Image Processing, IEEE Transactions on	331
332		24(5), 1614-1627 (May 2015)	332
333	25.	Varga, L., Balázs, P., Nagy, A.: An Energy Minimization Reconstruction Algorithm	333
334		for Multivalued Discrete Tomography. In: 3rd International Symposium on Com-	334
335		putational Modeling of Objects Represented in Images, Rome, Italy, Proceedings	335
336		(Taylor & Francis). pp. 179–185 (2012)	336
337	26.	Weber, S., Nagy, A., Schüle, T., Schnörr, C., Kuba, A.: A Benchmark Evaluation of	337
338		Large-Scale Optimization Approaches to Binary Tomography. In: Discrete Geom-	338
339		etry for Computer Imagery (DGCI 2006). LNCS, vol. 4245, pp. 146–156. Springer	339
340		(2006)	340
341	27.	Weber, S.: Discrete tomography by convex-concave regularization using linear and	341
342		quadratic optimization. PhD thesis, Ruprecht-Karls-Universität, Heidelberg, Ger-	342
343		many (2009)	343
344	28.	Weber, S., Schnörr, C., Hornegger, J.: A linear programming relaxation for binary	344
345		tomography with smoothness priors. Electronic Notes in Discrete Mathematics 12,	345
346		243-254 (2003)	346
347	29.	Zach, C., Gallup, D., Frahm, J., Niethammer, M.: Fast global labeling for real-	347
348		time stereo using multiple plane sweeps. In: Proceedings of the Vision, Modeling,	348
349		and Visualization Conference 2008, VMV 2008, Konstanz, Germany, October 8-10,	349
350		2008. pp. 243–252 (2008)	350

351 6 Appendix A

Lemma 1. Let D(z) defined as

$$D(z) := \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left((Wz)_i - c_k \right)^2.$$
(17)

Then the component $(\nabla D(z))_{jl}$ of the gradient of D at z with respect to z for pixel $j = \{1, ..., N\}$ and label $c_l \in \mathcal{L} = \{c_1, ..., c_K\}$ is given by 354

$$(\nabla D(z))_{jl} = \frac{\partial D(z)}{\partial z_{jl}} = ((Wz)_j - c_l)^2.$$
(18)

Proof. Using $Wz = \sum_{k=1}^{K} c_k z_k$, we compute

$$\frac{\partial D(z)}{\partial z_{jl}} = ((Wz)_j - c_l)^2 + 2\sum_{i=1}^N \sum_{k=1}^K z_{ik} ((Wz)_i - c_k) \frac{\partial}{\partial z_{jl}} \Big(\sum_m z_{im} c_m - c_k\Big)$$
(19)

$$= ((Wz)_j - c_l)^2 + 2c_l \sum_{k=1}^{K} z_{jk} ((Wz)_j - c_k)$$
(20)

$$= ((Wz)_j - c_l)^2 + 2c_l \left((Wz)_j \sum_{k=1}^K z_{jk} - \sum_{k=1}^K z_{jk} c_k \right)$$
(21)

$$= ((Wz)_j - c_l)^2 + 2c_l ((Wz)_j \cdot 1 - (Wz)_j)$$
(22)

$$= ((Wz)_j - c_l)^2$$
(23)

355

Lemma 2. The phase data term D(z) is concave with respect to z over the domain G.

$$D(z) := \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left((Wz)_i - c_k \right)^2.$$
(24)

$$G := \left\{ z \in [0,1]^{N \times K} \colon \sum_{k=1}^{K} z_{ik} = 1, \ \forall i = 1, \dots, N \right\}.$$
 (25)

Proof. The domain G is a convex set by definition. Consider the function

$$H: G \times G \to \mathbb{R}_+, \quad (x, y) \mapsto \sum_{i=1}^N \sum_{k=1}^K y_{ik} \left((Wx)_i - c_k \right)^2 \tag{26}$$

then the components $(\nabla_x H(x, y))_{jl}, (\nabla_y H(x, y))_{jl}$ of the gradient of H at (x, y) with respect to x and y for pixel $j = \{1, ..., N\}$ and label $l = \{1, ..., K\}$ are given

352

355

by

$$(\nabla_x H(x,y))_{jl} = \frac{\partial H(x,y)}{\partial x_{jl}} = 2c_l (W(x-y))_j$$
(27)

$$(\nabla_y H(x,y))_{jl} = \frac{\partial H(x,y)}{\partial y_{jl}} = ((Wx)_j - c_l)^2.$$
(28)

357 Using

$$D(z) = H(x, y) \circ (z, z) = H(z, z)$$
(29)

and applying the chain rule, we get

$$(\nabla D(z))_{jl} = \left((\nabla_x H(z,z))_{jl}, (\nabla_y H(z,z))_{jl} \right) \begin{pmatrix} 1\\ 1 \end{pmatrix} = ((Wz)_j - c_l)^2, \qquad (30)$$

$$D(z) = \langle \nabla D(z), z \rangle, \tag{31}$$

$$H(x,y) = \langle \nabla D(x), y \rangle.$$
(32)

Now, for any $x, y \in G$, we obtain the inequality

$$D(x) + \langle \nabla D(x), y - x \rangle = D(x) + \langle \nabla D(x), y \rangle - \langle \nabla D(x), x \rangle$$
(33)

$$= D(x) + \langle \nabla D(x), y \rangle - D(x) \tag{34}$$

$$= \langle \nabla D(x), y \rangle \tag{35}$$

$$=H(x,y) \tag{36}$$

$$\geq \inf_{x \in G} H(x, y) = H(y, y) = D(y), \tag{37}$$

where the infimum of H(x, y) with respect to $x \in G$ follows from the necessary condition:

$$(\nabla_x H(x,y))_{jl} = 2c_l(W(x-y))_j = 0, \qquad \forall j, \ \forall l, \tag{38}$$

$$\implies (Wx)_j = (Wy)_j, \tag{39}$$

$$\implies \quad x = y. \tag{40}$$

Since the function $H(\cdot, y)$ is invariant under pixelwise weighted averaging, and because $H(\cdot, y)$ is a sum of positive parabolas, the condition x = y is sufficient for attaining the infimum. As a result, we have shown the first-order condition for concavity of the phase for attaining the phase

As a result, we have shown the first-order condition for concavity of the phase data term D(z)

$$D(x) + \langle \nabla D(x), y - x \rangle \ge D(y), \quad \forall x, y \in G.$$
(41)

363

362

357

364 7 Appendix B



Fig. 6. Visual results of experiment phantom 1.

364



Fig. 7. Visual results of experiment phantom 2.



Fig. 8. Visual results of experiment phantom 3.



Fig. 9. Visual results of experiment phantom 4.

	Proposed noiseless	Proposed noise case	DART[2] noiseless	DART[2] noise case	Varga[25] noiseless	Varga[25] noise case
2						
3						
4						
5						
6		X				
7	A.		1			
8		N.	the second s			
9					S.	
10			S.	N	S.	

Fig. 10. Visual results of experiment phantom 5.