

Prior Learning and Convex-Concave Regularization of Binary Tomography

Stefan Weber ^{a,1}, Thomas Schüle ^{a,b,2}, Christoph Schnörr ^{a,3}

^a *University of Mannheim, Dept. M&CS, CVGPR-Group
D-68131 Mannheim, Germany*

^b *Siemens Medical Solutions, D-91301 Forchheim, Germany*

Abstract

In our previous work, we introduced a convex-concave regularization approach to the reconstruction of binary objects from few projections within a limited range of angles. A convex reconstruction functional, comprising the projections equations and a smoothness prior, was complemented with a concave penalty term enforcing binary solutions. In the present work we investigate alternatives to the smoothness prior in terms of probabilistically learnt priors encoding local object structure. We show that the difference-of-convex-functions DC-programming framework is flexible enough to cope with this more general model class. Numerical results show that reconstruction becomes feasible under conditions where our previous approach fails.

Keywords: Discrete Tomography, Markov Random Fields, Prior Learning, Convex-Concave Regularization

¹ Email:wstefan@uni-mannheim.de

² Email:schuele@uni-mannheim.de

³ Email:schnoerr@uni-mannheim.de

1 Introduction

1.1 Overview and Motivation

Discrete Tomography is concerned with the reconstruction of discrete-valued functions from projections. Historically, the field originated from several branches of mathematics like, for example, the combinatorial problem to determine binary matrices from its row and column sums (see the survey [12]). Meanwhile, however, progress is not only driven by challenging theoretical problems [7,10] but also by real-world applications where discrete tomography might play an essential role (cf. [11, chapters 15–21]).

The work presented in this paper is motivated by the reconstruction of volumes from *few* projection directions within a *limited* range of angles. From the viewpoint of established approaches to computational tomography [16], this is a severely ill-posed problem. The motivation for considering this difficult problem relates to the observation that in some specific scenarios [19] it is reasonable to assume that the function f to be reconstructed is *binary-valued*. This poses one of the essential questions of discrete tomography: how can knowledge of the discrete range of f be exploited in order to regularize and solve the reconstruction problem?

In our previous work [18], we introduced a convex-concave regularization approach to the binary reconstruction problem. Minimizing the squared residuals of the projection equations together with a smoothness prior favoring spatially homogeneous reconstructions was shown to considerably alleviate the ill-posedness of the reconstruction problem. Binary solutions were gradually computed in a “reconstruction-sensitive” way by *simultaneously* minimizing a concave penalty term. A primal-dual DC-programming algorithm particularly suited for this class of non-convex optimization problems showed promising performance.

Smoothness priors are convenient from the computational viewpoint because they result in *convex* functionals having relaxed the binary constraint. On the other hand, the signal class modelled thereby is limited to coarse-scale objects with large homogeneous areas or volumes. This motivates to investigate, within the same optimization framework, the use of priors encoding various object structures that have been probabilistically learnt from examples beforehand.

1.2 Related Work

Our objective falls into the area of Markov Random Field modelling which has a long history in image processing [2,8,9,13,20]. In the field of discrete tomography, related work includes [4,14,15]. In a similar way, we learn a standard prior from examples of specific image classes. The novelty of the present work is the incorporation of such priors into the overall convex-concave optimization framework. Rather than computing a-posteriori estimates by MCMC-sampling, we decompose the objective functional into the difference of convex functions and apply the DC-programming technique introduced in our previous work. Numerical results show that this *deterministic* optimization technique that just relies on a sequence of simple convex optimization problems and corresponding fast numerical linear algebra, yields high-quality binary reconstructions.

2 Problem Statement

2.1 Projection

The imaging geometry is represented by a linear system of equations $Ax = b$. Each projection ray corresponds to a row of matrix A , and its projection value is the corresponding component of b . The row entries of A represent the length of the intersection of pixels (voxels in the 3D case) of the (arbitrarily) discretized volume and the corresponding projection ray (see Fig. 1). This corresponds to the assumption that the function to be reconstructed is binary-valued, i.e. x is a binary-valued vector. Each component $x_i \in \{0, 1\}$ indicates whether the corresponding pixel belongs to the reconstructed object, $x_i = 1$, or not, $x_i = 0$ (see Fig. 1). This results in an *under*-determined linear system relating the unknown image (or volume) x and measured projection values b .

$$(1) \quad Ax = b, \quad x = (x_1, \dots, x_n)^\top \in \{0, 1\}^n$$

2.2 Reconstruction

In [18], we investigated binary reconstruction in terms of minimizing the functional

$$(2) \quad E(x) = \frac{1}{2} \left\{ |Ax - b|^2 + \alpha \sum_{\langle i,j \rangle} (x_i - x_j)^2 + \mu \langle x, e - x \rangle \right\}, \quad x \in [0, 1]^n.$$

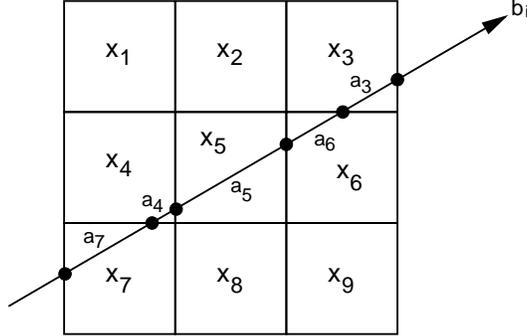


Fig. 1. Discretization model leading to the algebraic representation of the reconstruction problem: $Ax = b$, $x \in \{0, 1\}^n$.

The first term minimizes the residuals of the projection equations (1). The second term is a standard smoothness prior that ranges over all edges $\langle i, j \rangle$ of the underlying grid graph and favors spatial homogeneity of reconstructions x . The third term, e denotes the vector with all components equal to 1, enforces binary solutions x for increasing values of the parameter μ .

Our objective is to replace and to investigate alternatives to the standard smoothness prior, i.e. the second term in (2).

3 Markov Random Field Priors

3.1 Priors

Regarding x as random variables indexed by the pixel sites, and assuming the Markov property that non-adjacent variables x_i and x_j are conditionally independent given the respective neighborhood variables, probability distributions over the x -space may be specified as Gibbs distributions

$$(3) \quad p(x) = \frac{1}{Z} \exp \left(-\frac{1}{T} E_p(x) \right)$$

where the functional E_p is the sum of potentials indexed by the cliques $C \in \mathcal{C}$ of the pixel grid graph, and with the corresponding x -variables as arguments:

$$(4) \quad E_p(x) = \sum_{C \in \mathcal{C}} E_C(x_C).$$

In order to study directly generalizations of the smoothness prior (second term) in (2), we use the very same neighborhood-structure as depicted in

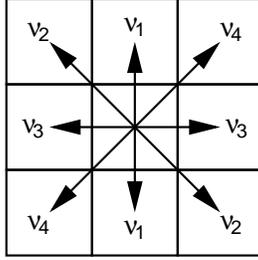


Fig. 2. We consider a second-order neighborhood structure for our MRF approach depending on 5 parameters, $\nu = \{\nu_0, \dots, \nu_4\}$. The first parameter ν_0 depends only on the central pixel itself whereas all other parameters also depend on neighboring pixels, as shown in this figure.

Fig. 2 and replace the smoothness prior by

$$(5) \quad E_p(x) = -\nu_0 \sum_{i=1}^n x_i - \sum_{\langle i,j \rangle} \nu_{\langle i,j \rangle} x_i x_j.$$

Note that this corresponds to (4) with single sites and edges as cliques. Furthermore, as we wish to obtain a stationary (translation-invariant) random field, (5) comprises 5 parameters arranged with respect to pixel site i as illustrated in Fig. 2.

For the purpose of parameter estimation (see next section), we specify the local conditional distribution for the variable x_i given all remaining variables:

$$(6) \quad p(x_i | x_{N(i)}) = \frac{\exp\left(x_i(\nu_0 + \sum_{j \in N(i)} \nu_{\langle i,j \rangle} x_j)\right)}{1 + \exp\left(\nu_0 + \sum_{j \in N(i)} \nu_{\langle i,j \rangle} x_j\right)}.$$

3.2 Parameter estimation

To estimate the parameters or the prior (5) from a sample of given data, we apply the method [3] which has some advantages over more classical estimates. For the readers convenience, we briefly sketch the estimation procedure here. We rewrite (6) as

$$(7) \quad p(x_i | x_{N(i)}) = \frac{\exp(x_i a_h)}{1 + \exp(a_h)}$$

where

$$(8) \quad a_h = w_h^\top \nu, \quad w_h := \begin{pmatrix} 1 \\ x_{i;n} + x_{i;s} \\ x_{i;nw} + x_{i;se} \\ x_{i;w} + x_{i;e} \\ x_{i;ne} + x_{i;sw} \end{pmatrix}, \quad \nu := \begin{pmatrix} \nu_0 \\ \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}.$$

In (8), vector w_h contains values of the neighbour pixels of x_i according to Fig. 2, and $h \in \{1, 2, \dots, 81\}$ indicates which of $3^4 = 81$ possible values this vector takes.

Accordingly, we compute 81 histograms using all pixels $x_i = 1$ and $x_i = 0$, respectively, $i = 1, \dots, n$. Using this histograms, the 81 values a_h are estimated using (7), and the following linear system is set up:

$$(9) \quad W\nu = a, \quad W := \begin{pmatrix} w_1^\top \\ w_2^\top \\ w_3^\top \\ \vdots \\ w_{81}^\top \end{pmatrix}, \quad a := \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{81} \end{pmatrix}.$$

Due to the special structure of the matrix W , it is possible to solve this system analytically for the parameters ν .

We point out, that this approach to parameter estimation has been used recently by [14] for discrete tomography as well.

4 Optimization and Reconstruction

4.1 Objective Functional

As discussed and motivated in the introduction, we wish to replace in (2) the second term by the prior (5) with the ν -parameters learnt beforehand as explained in the previous section. The modified objective functional reads:

$$(10) \quad \begin{aligned} E_2(x) &:= \frac{1}{2}|Ax - b|^2 - \tau \left(\nu_0 \sum_{i=1}^n x_i + \sum_{\langle i,j \rangle} \nu_{\langle i,j \rangle} x_i x_j \right) + \frac{1}{2} \mu \langle x, e - x \rangle \\ &= \frac{1}{2}|Ax - b|^2 - \tau(\nu_0 \langle e, x \rangle + \langle x, B, x \rangle) + \frac{1}{2} \mu \langle x, e - x \rangle. \end{aligned}$$

4.2 DC-Programming

In order to minimize our quadratic but non-convex functional (10), we use DC-programming which generally applies to objective functions $f(x) = g(x) - h(x)$ that may be represented as the difference of convex functions g and h .

Specifically, we use following primal-dual subgradient algorithm suggested in [5,6]:

DC-algorithm (DCA):

Choose $x^0 \in \text{dom } g$ arbitrary.

For $k = 0, 1, \dots$ compute (until convergence):

$$\mathbf{y}\text{-step:} \quad y^k \in \partial h(x^k)$$

$$\mathbf{x}\text{-step:} \quad x^{k+1} \in \partial g^*(y^k)$$

Here, g^* denotes the Fenchel conjugate function with respect to g (see, e.g., [17]). The DCA has the following properties:

Proposition 4.1 ([5]) *Assume $g, h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ to be proper, lower-semicontinuous and convex, and $\text{dom } g \subset \text{dom } h$, $\text{dom } h^* \subset \text{dom } g^*$. Then*

- (i) *the sequences $\{x^k\}, \{y^k\}$ according to (11), (11) are well-defined,*
- (ii) *$\{g(x^k) - h(x^k)\}$ is decreasing,*
- (iii) *every limit point x^* of $\{x^k\}$ is a critical point of $g - h$.*

4.3 Reconstruction Algorithm

In order to apply the DCA to the minimization of (10), we have to decompose the functional E_2 into the difference of two convex functions. We point out that this decomposition is not unique. Our choice is motivated by the simplicity and efficiency of the resulting DCA operations which can be easily applied to large-scale problems.

Using the definitions

$$Q := A^\top A, \quad q := -A^\top b, \quad \delta_C(x) = \begin{cases} 0 & \text{if } x \in C = [0, 1]^n \\ +\infty & \text{otherwise} \end{cases}$$

and choosing constants λ_Q, λ_B such that $\lambda_Q I - Q$ and $\lambda_B I + B$ are positive definite, we decompose the objective functional E_2 :

$$(11) \quad E_2(x; \mu) = g(x) - h(x; \mu)$$

where

$$\begin{aligned}
(12) \quad g(x) &= \frac{1}{2} \langle x, (\lambda_Q + 2\lambda_B)Ix \rangle + \delta_C(x) \\
h(x; \mu) &= \frac{1}{2} \langle x, [(\lambda_Q + 2\lambda_B)I + 2\tau B - Q]x \rangle \\
&\quad - \langle q + \tau\nu_0 e, x \rangle - \frac{1}{2} \mu \langle x, (e - x) \rangle \\
&= \frac{1}{2} \langle x, [(\lambda_Q + 2\lambda_B + \mu)I + 2\tau B - Q]x \rangle \\
&\quad - \langle q + \tau\nu_0 e + \frac{1}{2} \mu e, x \rangle.
\end{aligned}$$

Since h is smooth, the y -step of the DCA amounts to evaluate the gradient:

$$\begin{aligned}
(13) \quad y^k &= \nabla h(x^k; \mu) \\
&= [(\lambda_Q + 2\lambda_B + \mu)I - 2\tau B - Q]x^k - [q - (\tau\nu_0 - \frac{1}{2}\mu)e].
\end{aligned}$$

On the other hand, function g is non-smooth due to the constraint $x \in C$, and we have to solve:

$$\begin{aligned}
(14) \quad x^{k+1} &\in \partial g^*(y^k) \\
&= \operatorname{argmin}_x \{g(x) - \langle y^k, x \rangle\} \\
&= \operatorname{argmin}_x \left\{ \frac{\lambda_Q + 2\lambda_B}{2} \|x\|^2 - \langle y^k, x \rangle + \delta_C(x) \right\} \\
&= \operatorname{argmin}_x \{\tilde{g}(x) + \delta_C(x)\}.
\end{aligned}$$

The solution is easily found to be:

$$(15) \quad (x^{k+1})_i = \begin{cases} 0, & y_i^k \leq 0 \\ 1, & y_i^k \geq (\lambda_Q + 2\lambda_B) \quad , \quad i = 1, \dots, n. \\ \frac{1}{\lambda_Q + 2\lambda_B} y_i^k & \text{otherwise} \end{cases}$$

Remark: For the specific decomposition (12), the reconstruction algorithm turns out to be a special instance of the Goldstein-Levitin-Polyak projection method [1]. However, our approach proves convergence for the damping parameter $\frac{1}{\lambda_Q + 2\lambda_B}$, too.

5 Evaluation

In order to evaluate the performance of the MRF prior within the convex-concave regularization framework we created different texture-like images and estimated the MRF parameters. Throughout all experiments we compared the reconstruction based on the criterion (10) with the reconstruction using

the standard smoothness prior, i.e. criterion (2).

If not mentioned otherwise we initialized parameter μ with $\mu_0 = 0.0$ and used increments $\mu_\Delta = 0.1$ after each termination of the DCA. We recall that this gradually enforces the binary constraints $x_i \in \{0, 1\}$.

5.1 Experiment I

Figure 3(a) shows a chessboard texture from which we took just the horizontal and the vertical projections as measurements b . The reconstruction using (2) is shown in Fig. 3(b). Obviously, this image structure does not at all conform to the smoothness prior. Consequently, the reconstructions are poor, no matter how the parameter α is chosen.

The remaining figures depict the reconstruction results using the MRF prior: 3(c) for $\tau = 3.0$, and 3(d) for $\tau = 4.0$. The notable fact here is that the DCA computes a very good local minimum of the non-convex objective functional (10).

5.2 Experiment II

Figure 4(a) shows the original image from which we again took the horizontal and the vertical projection. Reconstruction of this problem with the standard smoothness prior, $\alpha = 0.25$, fails again, as shown in Fig. 4(b).

The texture in Fig. 4(c) was used for estimating the parameters ν of the MRF prior. The reconstructions for the following parameter values are shown: Figure 4(d) $\tau = 2.0$, 4(e) $\tau = 3.0$, and 4(f) $\tau = 10.0$. For $\tau = 2.0$, Fig. 4(d), the reconstruction fits the projection constraints best. By increasing the parameter τ , Figs. 4(e) and 4(f), the prior influences more and more the reconstruction so as to get closer to the texture image, Fig. 4(c), from its parameters were learnt.

5.3 Experiment III

For this experiment we created a reconstruction problem by taking the vertical and both diagonal projections from the image shown in Fig. 5(a). The original image was also used for the parameter estimation. We sampled the corresponding Gibbs distribution (3) with a Gibbs sampler and artificial temperature parameter $T^{-1} = 10$, Fig. 5(b). Figure 5(c) shows the poor reconstruction with the standard smoothness prior, $\alpha = 0.25$. The MRF prior, on the other hand, enables a very good reconstruction shown in Fig. 5(d) $\tau = 10$, based on the prior knowledge illustrated in Fig. 5(b)

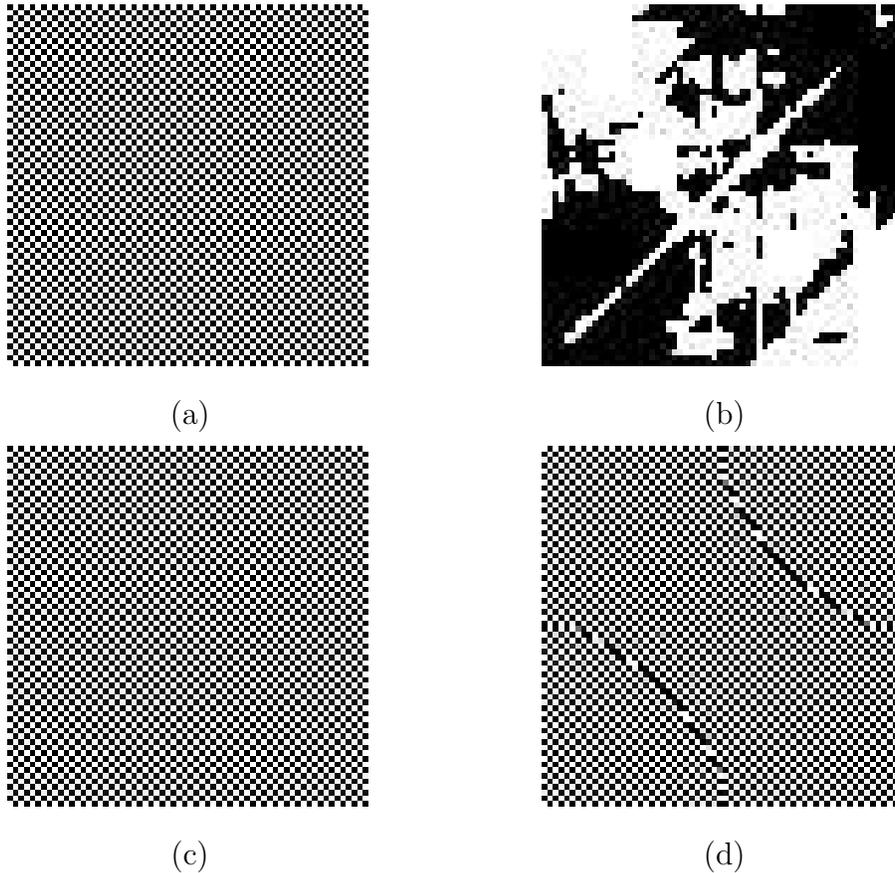


Fig. 3. (a) Original image, 64×64 , from which we took the vertical and the horizontal projection for the reconstruction problem. (b) DCA with smoothness prior, $\alpha = 0.25$. (c) DCA with MRF prior, $\tau = 3.0$. (d) DCA with MRF prior, $\tau = 4.0$.

5.4 Experiment IV

In the final experiment, we show that the MRF prior is also able to cope with situations where the standard smoothness prior is *not* inferior. To this end, we took two projections, horizontal and one diagonal (north-west to south-east), from the image shown in Fig. 6(a). The reconstruction with the standard smoothness prior is shown in Fig. 6(b) for $\alpha = 0.25$.

In order to apply the MRF prior, we estimated corresponding the parameters again from the original image, Fig. 6(a). Figures 6(c) and 6(d) show the reconstruction based on the MRF prior for different values of τ .

These results demonstrate that the MRF prior turns into a standard smoothness prior if the sample data used for learning provide corresponding evidence.

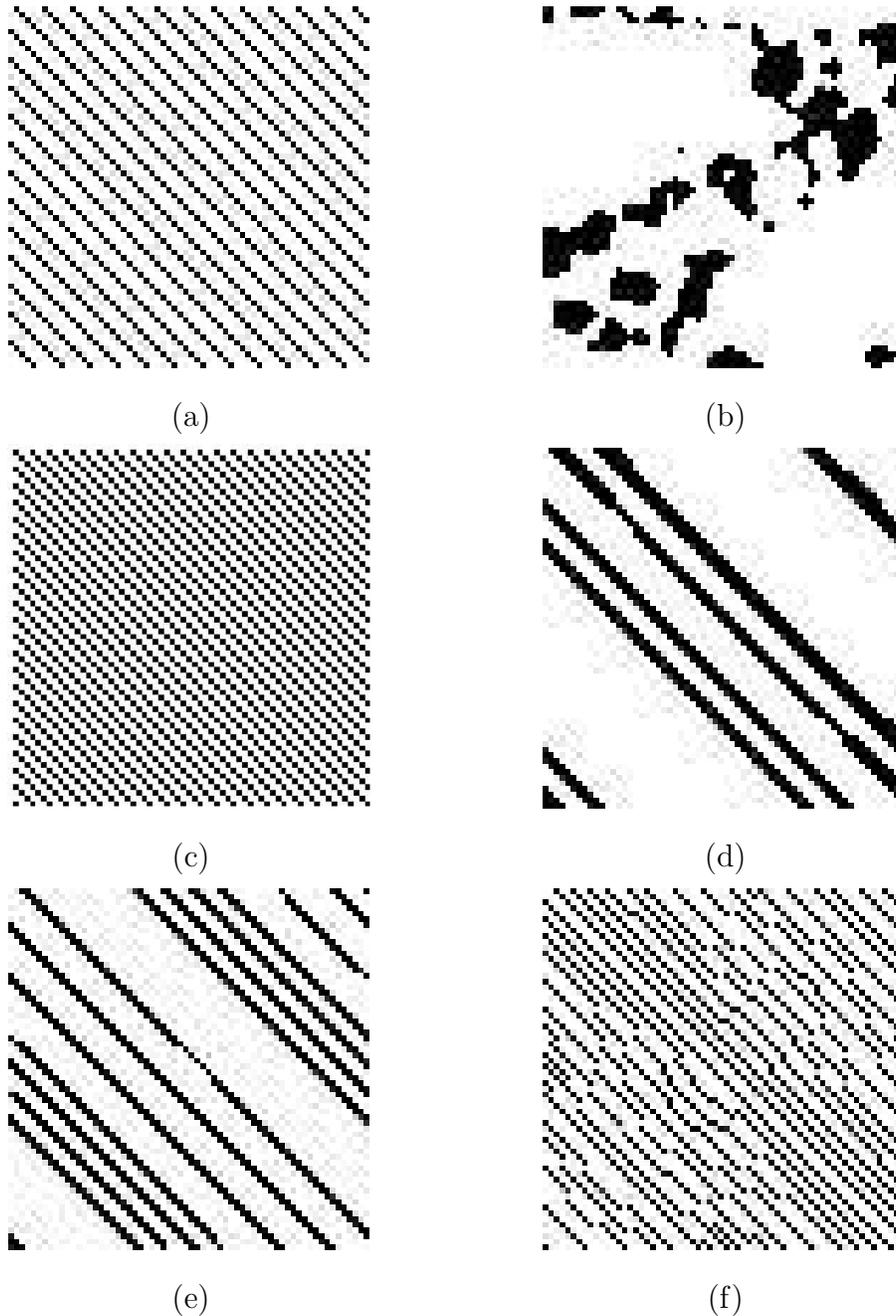


Fig. 4. (a) Original image, 64×64 , from which the vertical and the horizontal projections were taken as measurements b for the reconstruction problem. (b) DCA with smoothness prior, $\alpha = 0.25$. (c) Texture image used for estimating the MRF parameters, $\nu_0 = 0.206522$, $\nu_1 = -0.22789$, $\nu_2 = 0.468601$, $\nu_3 = -0.22789$, and $\nu_4 = -0.22789$. (d) DCA with MRF prior, $\tau = 2.0$. (e) DCA with MRF prior, $\tau = 3.0$. (f) DCA with MRF prior, $\tau = 10.0$.

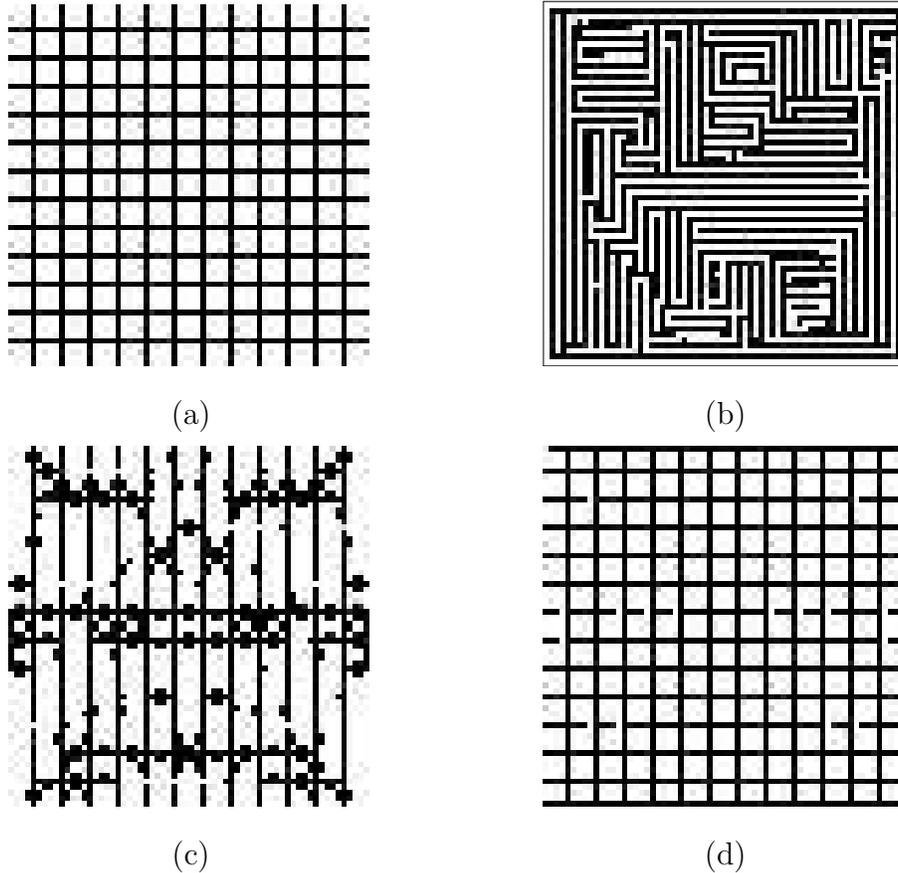


Fig. 5. (a) Original image, 64×64 , from which vertical and both diagonal projections were taken as measurements b for the reconstruction problem. (b) Sample of the Gibbs distribution with the estimated Markov random field parameters, $\nu_0 = -0.116639$, $\nu_1 = 0.361875$, $\nu_2 = -0.321148$, $\nu_3 = 0.361875$, $\nu_4 = -0.321148$, and artificial temperature $T^{-1} = 10$. (c) Result of the DCA with smoothness prior and $\alpha = 0.25$, (d) Result of the DCA with MRF prior, $\tau = 10$.

6 Conclusion

We investigated reconstructions in discrete tomography from few projections using MRF-based priors. Incorporation of this prior into a convex-concave regularization framework allows to efficiently minimize a highly non-convex reconstruction functional with a sequence of simple convex optimization problems and related deterministic algorithms, as opposed to MCMC sampling commonly used in the MRF literature.

Our further work will focus on larger image classes relevant for, e.g. med-

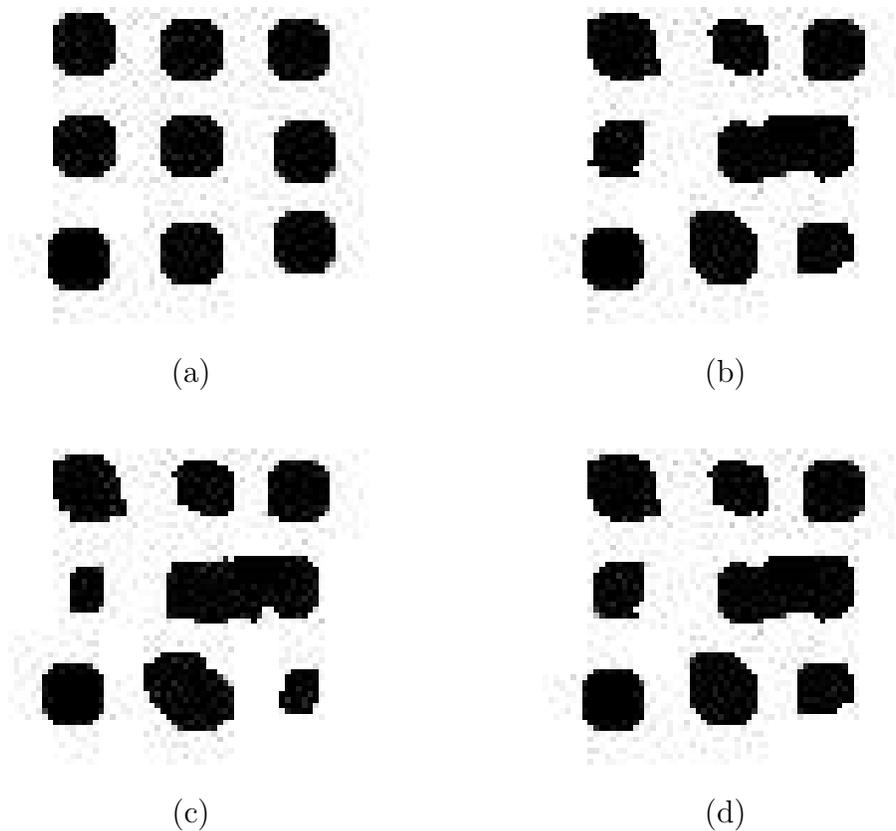


Fig. 6. (a) Original image, 64×64 , from the horizontal and one diagonal projection (north-west to south-east) were taken as measurements b . (b) Result of the DCA with smoothness prior, $\alpha = 0.25$. (c) Result of the DCA with MRF prior, $\tau = 0.001$. (d) Result of the DCA with MRF prior, $\tau = 0.005$.

ical imaging, and correspondingly extended MRF priors.

References

- [1] Bertsekas, D., *On the goldstein-levitin-polyak gradient projection method*, IEEE Transactions on Automatic Control **21** (1976), 174–184.
- [2] Besag, J., *On the analysis of dirty pictures (with discussion)*, J. R. Statist. Soc., Series B **48** (1986), 259–302.
- [3] Borges, C., *On the estimation of markov random field parameters*, IEEE Trans. on Pattern Analysis and Machine Intelligence **21** (1999).

- [4] Chan, M., G. Herman and E. Levitan, *Bayesian image reconstruction using image-modeling gibbs priors*, Int. J. Imag. Syst. Technol. **9** (1998), 85–98.
- [5] Dinh, T. P. and L. H. An, *A d.c. optimization algorithm for solving the trust-region subproblem*, SIAM J. Optim. **8** (1998), 476–505.
- [6] Dinh, T. P. and S. Elberoussi, *Duality in d.c. (difference of convex functions) optimization subgradient methods*, in: *Trends in Mathematical Optimization*, Int. Series of Numer. Math. **84**, Birkhäuser Verlag, Basel, 1988 pp. 277–293.
- [7] Gardner, R. and P. Gritzmann, *Discrete tomography: Determination of finite sets by x-rays*, Trans. Amer. Math. Soc. **349** (1997), 2271–2295.
- [8] Geman, D., *Random fields and inverse problems in imaging*, in: P. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XVIII – 1988*, Lect. Notes in Math. **1427**, Springer Verlag, Berlin, 1990 pp. 113–193.
- [9] Geman, S. and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, IEEE Trans. Patt. Anal. Mach. Intell. **6** (1984), 721–741.
- [10] Gritzmann, P., D. Prangenberg, S. de Vries and M. Wiegelmann, *Success and failure of certain reconstruction and uniqueness algorithms in discrete tomography*, Int. J. Imag. Syst. Technol. **9** (1998), 101–109.
- [11] Herman, G. and A. Kuba, editors, *Discrete Tomography: Foundations, Algorithms, and Applications*, Birkhäuser, Boston, 1999.
- [12] Kuba, A. and G. Herman, *Discrete tomography: A historical overview*, in: G. T. Herman and A. Kuba, editors, *Discrete Tomography: Foundations, Algorithms, and Applications*, Birkhäuser, Boston, 1999 pp. 3–34.
- [13] Li, S., *Markov Random Field Modeling in Image Analysis*, Springer Verlag, Tokyo, 2001.
- [14] Liao, H. Y. and G. T. Herman, *Automated estimation of the parameters of gibbs priors to be used in binary tomography*, Discrete Appl. Math. **139** (2004), 149–170.
- [15] Matej, S., G. Herman and A. Vardi, *Binary tomography on the hexagonal grid using gibbs priors*, Int. J. Imag. Syst. Technol. **9** (1998), 126–131.
- [16] Natterer, F. and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia, 2001.
- [17] Rockafellar, R., *Convex analysis*, Princeton Univ. Press, Princeton, NJ, 1972, 2 edition.

- [18] Schüle, T., C. Schnörr, S. Weber and J. Hornegger, *Discrete tomography by convex-concave regularization and d.c. programming*, Technical report 15/2003, University of Mannheim (2003), to appear in *Discrete Applied Mathematics*.
- [19] Weber, S., T. Schüle, C. Schnörr and J. Hornegger, *A linear programming approach to limited angle 3d reconstruction from dsa projections*, Special Issue of *Methods of Information in Medicine* **4** (2004), pp. 320–326.
- [20] Winkler, G., *Image analysis, random fields and dynamic monte carlo methods*, *Appl. of Mathematics*, Springer-Verlag, Heidelberg **27** (1995).