

A convex formulation for Binary Tomography

Ajinkya Kadu and Tristan van Leeuwen

Abstract—Binary tomography is concerned with the recovery of binary images from a few of their projections (i.e., sums of the pixel values along various directions). To reconstruct an image from noisy projection data, one can pose it as a constrained least-squares problem. As the constraints are non-convex, many approaches for solving it rely on either relaxing the constraints or heuristics. In this paper we propose a novel convex formulation, based on the Lagrange dual of the constrained least-squares problem. The resulting problem is a generalized LASSO problem which can be solved efficiently. It is a relaxation in the sense that it can only be guaranteed to give a feasible solution; not necessarily the optimal one. In exhaustive experiments on small images (2×2 , 3×3 , 4×4) we find, however, that if the problem has a unique solution, our dual approach finds it. In the case of multiple solutions, our approach finds the commonalities between the solutions. Further experiments on realistic numerical phantoms and an experiment on X-ray dataset show that our method compares favorably to Total Variation and DART.

Index Terms—Binary tomography, inverse problems, duality, LASSO

I. INTRODUCTION

DISCRETE tomography is concerned with the recovery of discrete images (i.e., images whose pixels take on a small number of prescribed grey values) from a few of their projections (i.e., sums of the pixel values along various directions). Early work on the subject mostly deals with the mathematical analysis, combinatorics, and geometry. Since the 1970s, the development of algorithms for discrete tomography has become an active area of research as well [1]. It has found applications in image processing and computer vision [2], [3], atomic-resolution electron microscopy [4], [5], medicine imaging [6], [7] and material sciences [8]–[11].

A. Mathematical formulation

The discrete tomography problem may be mathematically formulated as follows. We represent an image by a grid of $N = n \times n$ pixels taking values $x_j \in \mathcal{U} = \{u_0, u_1, \dots, u_K\}$. The projections are linear combinations of the pixels along m different (lattice) directions. We denote the linear transformation from image to projection data by

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

where x_j denotes the value of the image in the j^{th} cell, y_i is the (weighted) sum of the image along the i^{th} ray, and a_{ij} is proportional to the length of the i^{th} ray in the j^{th} cell¹.

The goal is to find a solution to this system of equations with the constraint that $x_j \in \mathcal{U}$, i.e.,

$$\text{find } \mathbf{x} \in \mathcal{U}^N \text{ such that } \mathbf{A}\mathbf{x} = \mathbf{y}.$$

A. Kadu and T. van Leeuwen are with the Mathematical Institute, Utrecht University, The Netherlands.

¹We note that other projection models exist and can be similarly represented by a_{ij} .

When the system of equations does not have a unique solution, finding one that only takes values in \mathcal{U}^N has been shown to be an NP-hard problem for more than 3 directions, i.e., $m \geq 3$ [12].

Due to the presence of noise, the system of equations may not have a solution, and the problem is sometimes formulated as a constrained least-squares problem

$$\min_{\mathbf{x} \in \mathcal{U}^N} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2. \quad (1)$$

Obviously, the constraints are non-convex and solving (1) exactly is not trivial. Next, we briefly discuss some existing approaches for solving it.

B. Literature Review

Methods for solving (1) can be roughly divided into four classes: algebraic methods, stochastic sampling methods, (convex) relaxation and (heuristic/greedy) combinatorial approaches.

The *algebraic* methods exploit the algebraic structure of the problem and may give some insight into the (non-) uniqueness and the required number of projections [13], [14]. While theoretically very elegant, these methods are not readily generalized to realistic projection models and noisy data.

The *stochastic sampling methods* typically construct a probability density function on the space of discrete images, allowing one to sample images and use Markov Chain Monte Carlo type methods to find a solution [15]–[17]. These methods are very flexible but may require a prohibitive number of samples when applied to large-scale datasets.

Relaxation methods are based on some form convex or non-convex relaxation of the constraint. This allows for a natural extension of existing variational formulations and iterative algorithms [18]–[22]. While these approaches can often be implemented efficiently and apply to large-scale problems, getting them to converge to the correct discrete solution can be challenging. Another variant of convex relaxation includes the linear-programming based method [23]. This method works well on small-scale images and noise-free data.

The *heuristic* algorithms, finally, combine ideas from combinatorial optimization and iterative methods. Such methods are often efficient and known to perform well in practice [24], [25].

A more extensive overview of various methods for binary tomography and variants thereof (e.g., with more than two grey levels) are discussed in [26].

C. Contributions and outline

We propose a novel, convex, reformulation for discrete tomography with two grey values $\{u_0, u_1\}$ (often referred to

as *binary* tomography). Starting from the constrained least-squares problem (1) we derive a corresponding Lagrange dual problem, which is convex by construction. Solving this problem yields an image with pixel values in $\{u_0, 0, u_1\}$. Setting the remaining zero-valued pixels to u_0 or u_1 generates a *feasible* solution of (1) but not necessarily an *optimal* one. In this sense, our approach is a relaxation. Exhaustive enumeration of small-scale ($n = 2, 3, 4$) images with few directions ($m = 2, 3$) show that if the problem has a unique solution, then solving the dual problem yields the correct solution. When there are multiple solutions, the dual approach finds the common elements of the solutions, leaving the remaining pixels undefined (zero-valued). We conjecture that this holds for larger n and m as well. This implies that we can only expect to usefully solve problem instances that allow a unique solution and characterize the non-uniqueness when there are a few solutions.

For practical applications, the most relevant setting is where the equations alone do not permit a unique solution, but the constrained problem does. Otherwise, more measurements or prior information about the object would be needed in order to usefully image it. With well-chosen numerical experiments on synthetic and real data, we show that our new approach is competitive for practical applications in X-ray tomography as well.

The outline of the paper is as follows. We first give an intuitive derivation of the dual problem for invertible $\mathbf{A}^T \mathbf{A}$ before presenting the main results for general \mathbf{A} . We then discuss two methods for solving the resulting convex problem. Then, we offer the numerical results on small-scale binary problems to support our conjecture. Numerical results on numerical phantoms and real data are presented in Section IV. Finally, we conclude the paper in Section V.

II. DUAL PROBLEM

For the purpose of the derivation, we assume that the problem has pixel values ± 1 . The least-squares *binary* tomography problem can then be formulated as:

$$\begin{aligned} \mathbf{x}^* \triangleq \underset{\mathbf{x}}{\operatorname{argmin}} \left(\inf_{\phi} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 \right), \\ \text{subject to } \mathbf{x} = \operatorname{sign}(\phi), \end{aligned} \quad (2)$$

where $\phi \in \mathbb{R}^N$ is an auxiliary variable, and $\operatorname{sign}(\cdot)$ denotes the elementwise signum function. In our analysis, we consider the signum function such that $\operatorname{sign}(0) = 0$. The Lagrangian for this problem is defined as

$$\mathcal{L}(\mathbf{x}, \phi, \boldsymbol{\nu}) \triangleq \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \boldsymbol{\nu}^T (\mathbf{x} - \operatorname{sign}(\phi)). \quad (3)$$

The variable $\boldsymbol{\nu} \in \mathbb{R}^N$ is the Lagrange multiplier associated with the equality constraint $\mathbf{x} = \operatorname{sign}(\phi)$. We refer to this variable as the *dual* variable in the remainder of the paper. We define the dual function $g(\boldsymbol{\nu})$ corresponding to the Lagrangian (3) as

$$g(\boldsymbol{\nu}) \triangleq \inf_{\mathbf{x}, \phi} \mathcal{L}(\mathbf{x}, \phi, \boldsymbol{\nu})$$

The primal problem (2) has a dual optimization problem expressed as

$$\boldsymbol{\nu}^* \triangleq \underset{\boldsymbol{\nu}}{\operatorname{argmax}} g(\boldsymbol{\nu}).$$

As the dual function is *always* concave [27], this provides a way to define a convex formulation for the original problem. We should note two important aspects of duality theory here: *i)* we are not guaranteed in general that maximizing $g(\boldsymbol{\nu})$ yields a solution to the primal problem; *ii)* the reformulation is only computationally useful if we can efficiently evaluate g . The conditions under which the dual problem yields a solution to the primal problem are known as Slater's conditions [28] and are difficult to check in general unless the primal problem is convex. We will later show, by example, that the dual problem *does not always* solve the primal problem. Classifying under which conditions we *can* solve the primary problem via its dual is beyond the scope of this paper.

It turns out we can obtain a closed-form expression for g . Before presenting the general form of g , we first present a detailed derivation for invertible $\mathbf{A}^T \mathbf{A}$ to provide some insight.

A. Invertible $\mathbf{A}^T \mathbf{A}$

The Lagrangian is *separable* in terms of \mathbf{x} and ϕ . Hence, we can represent the dual function as the sum of two functions, $g_1(\boldsymbol{\nu})$ and $g_2(\boldsymbol{\nu})$.

$$g(\boldsymbol{\nu}) = \underbrace{\inf_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \boldsymbol{\nu}^T \mathbf{x} \right\}}_{g_1(\boldsymbol{\nu})} + \underbrace{\inf_{\phi} \left\{ -\boldsymbol{\nu}^T \operatorname{sign}(\phi) \right\}}_{g_2(\boldsymbol{\nu})} \quad (4)$$

First, we consider $g_1(\boldsymbol{\nu})$. Assuming $\mathbf{A}^T \mathbf{A}$ to be a non-singular matrix, we find the unique minimizer by setting the gradient to zero:

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{y} - \boldsymbol{\nu}). \quad (5)$$

Substituting \mathbf{x}^* back in the expression and re-arranging some terms we arrive at the following expression for g_1 :

$$g_1(\boldsymbol{\nu}) = -\frac{1}{2} \|\mathbf{A}^T \mathbf{y} - \boldsymbol{\nu}\|_{(\mathbf{A}^T \mathbf{A})^{-1}}^2 + \frac{1}{2} \mathbf{y}^T \mathbf{y}, \quad (6)$$

where $\|\mathbf{z}\|_{\mathbf{W}}^2 = \mathbf{z}^T \mathbf{W} \mathbf{z}$ is a weighted ℓ_2 -norm.

Next, we consider $g_2(\boldsymbol{\nu})$. Note that this function is *separable* in terms of ν_i

$$g_2(\boldsymbol{\nu}) = \inf_{\phi} \left\{ -\sum_{i=1}^N (\nu_i \operatorname{sign}(\phi_i)) \right\}.$$

The function $-\nu \operatorname{sign}(\phi)$ achieves its smallest value for $\phi = \nu$ when $\nu \neq 0$. This solution is not unique of course, but that does not matter as we are only interested in the sign of ϕ . When $\nu_i = 0$ the function takes on value 0 regardless of the value of ϕ_i . We thus find

$$g_2(\boldsymbol{\nu}) = -\|\boldsymbol{\nu}\|_1. \quad (7)$$

Hence, the dual function for the Lagrangian in (3) takes the following explicit form:

$$g(\boldsymbol{\nu}) = -\frac{1}{2} \|\mathbf{A}^T \mathbf{y} - \boldsymbol{\nu}\|_{(\mathbf{A}^T \mathbf{A})^{-1}}^2 - \|\boldsymbol{\nu}\|_1 + \frac{1}{2} \mathbf{y}^T \mathbf{y} \quad (8)$$

The maximizer to dual function (8) is found by solving the following minimization problem:

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{\nu} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{\nu} - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^{-1}}^2 + \|\boldsymbol{\nu}\|_1. \quad (9)$$

This optimization problem lies in the class of *least absolute shrinkage and selection operator* (LASSO) [29]. The primal solution can be synthesized from the solution of the dual problem via $\mathbf{x}^* = \text{sign}(\phi^*) = \text{sign}(\boldsymbol{\nu}^*)$.

It is important to note at this point that the solution of the dual problem only determines those elements of the primal problem, x_i , for which $\nu_i \neq 0$. The remaining degrees of freedom in \mathbf{x} need to be determined by alternative means. The resulting solution is a *feasible* solution of the primal problem, but not necessarily the *optimal* one.

To gain some insight into the behaviour of the dual objective, consider a one-dimensional example with $\mathbf{A} = 1$:

$$x^* = \underset{x \in \{-1, 1\}}{\text{argmin}} \frac{1}{2}(x - y)^2. \quad (10)$$

The solution to this problem is given by $x^* = \text{sign}(y)$. The corresponding dual problem is

$$\nu^* = \underset{\nu \in \mathbb{R}}{\text{argmin}} \frac{1}{2}(\nu - y)^2 + |\nu|, \quad (11)$$

the solution of which is given by $\nu^* = \max(|y| - 1, 0) \text{sign}(y)$. Hence, for $|y| > 1$, the solution of the dual problem yields the desired solution. For $|y| \leq 1$, however, the dual problem yields $\nu^* = 0$ in which case the primal solution $x^* = 0$. We will see in section II-C that when using certain iterative methods to solve the dual problem, the iterations will naturally approach the solution $\nu^* = 0$ from the correct side, so that the sign of the approximate solution may still be useful.

B. Main results

We state the main results below. The proofs for these statements are provided in the Appendix section.

Proposition 1. *The dual objective of (2) for general $\mathbf{A} \in \mathbb{R}^{m \times N}$ is given by*

$$g(\boldsymbol{\nu}) = \begin{cases} -\frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 - \|\boldsymbol{\nu}\|_1 + \frac{1}{2} \mathbf{y}^T \mathbf{y} & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}, \\ -\infty & \text{otherwise} \end{cases}$$

where \dagger denotes the pseudo-inverse and $\mathcal{R}_{\mathbf{A}^T}$ is the range of \mathbf{A}^T . This leads to the following optimization problem

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}}{\text{argmin}} \frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 + \|\boldsymbol{\nu}\|_1. \quad (12)$$

Remark 1. *In case $m \geq N$ and \mathbf{A} has full rank (i.e., $\mathcal{R}_{\mathbf{A}^T} = \mathbb{R}^N$), $\mathbf{A}^T \mathbf{A}$ is invertible and the general form (12) simplifies to (9).*

Corollary 1. *The minimization problem (12) can be restated as*

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\mu} \in \mathbb{R}^m}{\text{argmin}} \frac{1}{2}\|\mathbf{A} \mathbf{A}^\dagger (\boldsymbol{\mu} - \mathbf{y})\|^2 + \|\mathbf{A}^T \boldsymbol{\mu}\|_1, \quad (13)$$

and the primal solution is recovered through $\mathbf{x}^* = \text{sign}(\mathbf{A}^T \boldsymbol{\mu}^*)$.

Remark 2. *For $m \leq N$ and \mathbf{A} full row rank, we have $\mathbf{A} \mathbf{A}^\dagger = \mathbf{I}$ and the formulation (13) simplifies to*

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\mu} \in \mathbb{R}^m}{\text{argmin}} \frac{1}{2}\|\boldsymbol{\mu} - \mathbf{y}\|^2 + \|\mathbf{A}^T \boldsymbol{\mu}\|_1. \quad (14)$$

This form implicitly handles the constraints on the search space of $\boldsymbol{\mu}$ in Proposition 1. It allows us to use the functional form for matrix \mathbf{A} thereby reducing the storage and increasing the computational speed to find an optimal dual variable $\boldsymbol{\mu}^$.*

Proposition 2. *The dual problem for a binary tomography problem with grey levels $u_0 < u_1$ is given by:*

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}}{\text{argmin}} \frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 + p(\boldsymbol{\nu}), \quad (15)$$

where $p(\boldsymbol{\nu}) = \sum_i |u_0| \max(-\nu_i, 0) + |u_1| \max(\nu_i, 0)$ is an asymmetric one-norm. The primal solution is obtained using

$$\mathbf{x}^* = u_0 \mathbf{1} + (u_1 - u_0) H(\boldsymbol{\nu}^*),$$

where $H(\cdot)$ denotes the Heaviside function.

We summarize the procedure in algorithm 1 for finding the optimal solution via solving the dual problem. In practical applications, the formulation in step 10 is very useful since the projection matrix \mathbf{A} generally has a low rank, i.e., $\text{rank}(\mathbf{A}) < \min(m, N)$.

Algorithm 1 Dual problem for various cases

Input: $\mathbf{A} \in \mathbb{R}^{m \times N}$, $\mathbf{y} \in \mathbb{R}^m$

Output: $\mathbf{x}^* \in \{-1, 0, 1\}^N$

```

1: if  $\text{rank}(\mathbf{A}) = \min(m, N)$  then
2:   if  $m > N$  then
3:      $\boldsymbol{\nu}^* \triangleq \underset{\boldsymbol{\nu}}{\text{argmin}} \frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^{-1}}^2 + \|\boldsymbol{\nu}\|_1$ 
4:     return  $\mathbf{x}^* = \text{sign}(\boldsymbol{\nu}^*)$ 
5:   else
6:      $\boldsymbol{\nu}^* \triangleq \underset{\boldsymbol{\nu}}{\text{argmin}} \frac{1}{2}\|\boldsymbol{\nu} - \mathbf{y}\|^2 + \|\mathbf{A}^T \boldsymbol{\nu}\|_1$ 
7:     return  $\mathbf{x}^* = \text{sign}(\mathbf{A}^T \boldsymbol{\nu}^*)$ 
8:   end if
9: else
10:   $\boldsymbol{\nu}^* \triangleq \underset{\boldsymbol{\nu}}{\text{argmin}} \frac{1}{2}\|\mathbf{A} \mathbf{A}^\dagger (\boldsymbol{\nu} - \mathbf{y})\|^2 + \|\mathbf{A}^T \boldsymbol{\nu}\|_1$ 
11:  return  $\mathbf{x}^* = \text{sign}(\mathbf{A}^T \boldsymbol{\nu}^*)$ 
12: end if
```

Remark 3. *The realistic tomographic data contains Poisson noise. In such case, the binary tomography problem takes the constrained weighted least-squares form [30], [31]:*

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{argmin}} \left(\inf_{\phi} \frac{1}{2}\|\mathbf{y} - \mathbf{A} \mathbf{x}\|_{\Lambda}^2 \right), \text{ s.t. } \mathbf{x} = \text{sign}(\phi), \quad (16)$$

where $\Lambda \in \mathbb{R}^{m \times m}$ is a diagonal matrix with elements $\Lambda_i > 0$ representing the least-squares weight per projection. The dual objective of (16) is given by

$$g(\boldsymbol{\nu}) = \begin{cases} -\frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T \Lambda \mathbf{y}\|_{\mathbf{B}}^2 - \|\boldsymbol{\nu}\|_1 + \frac{1}{2} \mathbf{y}^T \mathbf{y} & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}, \\ -\infty & \text{otherwise,} \end{cases}$$

where $\mathbf{B} \triangleq (\mathbf{A}^T \Lambda \mathbf{A})^\dagger$. This leads to optimization problem

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}}{\text{argmin}} \frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T \Lambda \mathbf{y}\|_{(\mathbf{A}^T \Lambda \mathbf{A})^\dagger}^2 + \|\boldsymbol{\nu}\|_1. \quad (17)$$

If $\text{rank}(\mathbf{A}) = m$ with $m \leq N$, the problem (17) reduces to

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\mu} \in \mathbb{R}^m}{\text{argmin}} \frac{1}{2}\|\boldsymbol{\mu} - \Lambda^{1/2} \mathbf{y}\|^2 + \|\mathbf{A}^T \Lambda^{1/2} \boldsymbol{\mu}\|_1, \quad (18)$$

and the primal solution is recovered from $\mathbf{x}^* = \text{sign}(\mathbf{A}^T \Lambda^{1/2} \boldsymbol{\mu}^*)$.

C. Solving the dual problem

When $\mathbf{A}^T \mathbf{A}$ is invertible, the dual formulation (12) can be, in principle, solved² using a proximal gradient algorithm ([32], [33]):

$$\boldsymbol{\nu}_{k+1} \triangleq S_{L^{-1}} \left(\boldsymbol{\nu}_k - L^{-1} (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{y} - \boldsymbol{\nu}_k) \right), \quad (19)$$

where $L = \|(\mathbf{A}^T \mathbf{A})^{-1}\|$ and the soft thresholding operator $S_\tau(\cdot) = \max(|\cdot| - \tau, 0) \text{sign}(\cdot)$ is applied component-wise to its input. We can interpret this algorithm as minimizing subsequent approximations of the problem, as illustrated in figure 1.

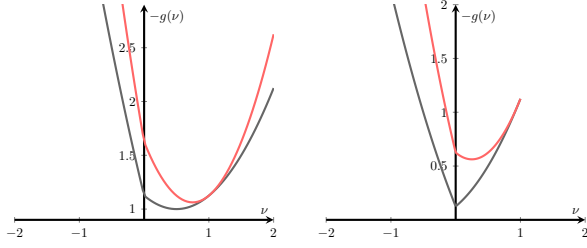


Figure 1. Plot of the dual function g (gray line) corresponding the primal objective $(x - y)^2$ for $y = 1.5$ (left) and $y = 0.5$ (right) and its approximations (red line) at $x = 1$.

An interesting note is that, when starting from $\boldsymbol{\nu}_0 = \mathbf{0}$, the first iteration yields a thresholded version of $\mathbf{A}^\dagger \mathbf{y}$. As such, the proposed formulation is a natural extension of a naive segmentation approach and allows for segmentation in a data-consistent manner.

If $\mathbf{A} \mathbf{A}^T$ is invertible we have $\mathbf{A} \mathbf{A}^\dagger = \mathbf{I}$ and it seems more natural to solve (14) instead. Due to the appearance of \mathbf{A}^T in the one-norm, it is no longer straightforward to apply a proximal gradient method. A possible strategy is to replace the one-norm with a smooth approximation of it, such as $|\cdot| = \sqrt{(\cdot)^2 + \epsilon}$. As illustrated in figure 2, this will slightly shift the minimum of the problem. Since we are ultimately only using the sign of the solution, this may not be a problem. The resulting objective is smooth and can be solved using any gradient-descent algorithm.

We also note that splitting methods can be used to solve (14). For example, the alternating direction method of multipliers (ADMM) [34] and/or split-Bregman method [35]. Another class of method that can solve (14) are the primal-dual methods (e.g., Arrow-Hurwicz primal-dual algorithm [36], Chambolle-Pock algorithm [37]). These methods rely on the proximal operators of functions and iterate towards finding the saddle point of the problem. If the proximal operators are simple, these are computationally faster than the splitting methods. Recently, proximal gradient methods have been developed to solve problems of form (14) [38].

The dual problem (15) for binary tomography problem with grey levels $u_0 < u_1$ is also solved using proximal gradient method. We provide the proximal operator for an asymmetric one-norm in the following proposition.

Proposition 3. *The proximal operator for an asymmetric one-norm function*

$$p(\mathbf{x}) = \sum_{i=1}^N |u_0| \max(-x_i, 0) + |u_1| \max(x_i, 0)$$

with $u_0 < u_1$, is given by

$$\mathcal{P}_{p,\lambda}(\mathbf{z}) \triangleq \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|^2 + \lambda p(\mathbf{x}) \right\} = \mathcal{S}_{\lambda u_0 < \lambda u_1}(\mathbf{z}),$$

where $\lambda > 0$, and $\mathcal{S}_{a < b}(\cdot)$ is an asymmetric soft-thresholding function

$$\mathcal{S}_{a < b}(t) = \begin{cases} t - |b| & t \geq |b| \\ 0 & -|a| < t < |b| \\ t + |a| & t \leq -|a| \end{cases}$$

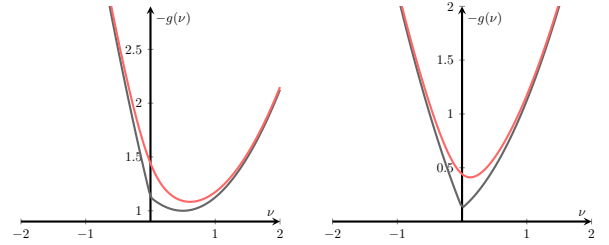


Figure 2. Plot of the dual function g (gray line) corresponding the the primal objective $(x - y)^2$ for $y = 1.5$ (left) and $y = 0.5$ (right) and its smooth approximation (red line) using $|\cdot| \approx \sqrt{(\cdot)^2 + \epsilon}$ with $\epsilon = 0.1$.

III. NUMERICAL EXPERIMENTS - BINARY TOMOGRAPHY

To illustrate the behavior of the dual approach, we consider the simple setting of reconstructing an $n \times n$ image from its sums along m lattice directions (here restricted to the horizontal, vertical and two diagonal directions, so $m \in [2, 4]$). For $m, n > 2$ the problem is known to be NP-hard. For small n we can simply enumerate all possible images, find all solutions in each case by a brute-force search and compare these to the solution obtained by the dual approach. To this end, we solved the resulting dual problem (13) using the CVX package in Matlab [39]. This yields an approximate solution, and we set elements in the numerically computed dual solution smaller than 10^{-9} to zero. We then compare the obtained primal solution, which has values $-1, 0, 1$ to the solution(s) of the binary tomography problem. From performing these computations for $n = 2, 3, 4$ and $m = 2, 3, 4$ we conclude the following:

- If the problem has a unique solution, then the dual approach retrieves it.
- If the problem has multiple solutions, then the dual approach retrieves the intersection of all solutions. The remaining pixels in the dual solution are undetermined (have value zero).

An example is shown in Figure 3.

A summary of these results is presented in Table I. The table shows the number of cases with a unique solution where the dual approach gave the correct solution and, in case of multiple solutions, the number of cases where the

²In practice, computing the inverse of $\mathbf{A}^T \mathbf{A}$ for large-scale problem is difficult.

Table I
SUMMARY OF COMPLETE ENUMERATION EXPERIMENTS.

	n	total	unique	multiple
$m = 2$	2	16	14/14	2/2
	3	512	230/230	282/282
	4	65536	6902/6902	58541/58634*
$m = 3$	2	16	16/16	0/0
	3	512	496/496	16/16
	4	65536	54272/54272	10813/11264*
$m = 4$	2	16	16/16	0/0
	3	512	512/512	0/0
	4	65536	65024/65024	512/512

dual approach correctly determined the intersection of all solutions). In a few instances with multiple solutions, CVX failed to provide an accurate solution (denoted with * in the table).

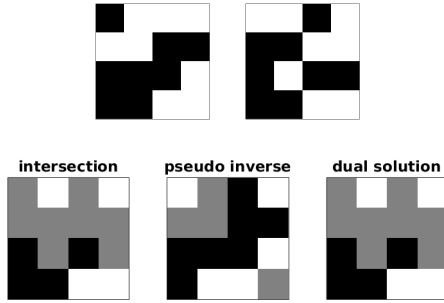


Figure 3. Example for $n = 4$ and $m = 3$ (using directions $(0, 1)$, $(1, 0)$ and $(1, 1)$). Two images with the same projections are shown in the top row while the intersection and the results obtained by the pseudo-inverse and the dual problem are shown in the bottom row.

Based on these experiments, we conjecture that there is a subclass of the described binary tomography problem that is not NP-hard. We should note that, as n grows the number of cases that have a unique solution grows smaller unless m grows accordingly. It has been established that binary images that are h,v,d -convex³ can be reconstructed from their horizontal, vertical and diagonal projections in polynomial time [24], [40]. However, we can construct images that are *not* h,v,d -convex but still permit a unique solution, see figure 4. Such images are also retrieved using our dual approach.

IV. NUMERICAL EXPERIMENTS - X-RAY TOMOGRAPHY

In this section, we present numerical results for limited-angle X-ray tomography on a few numerical phantoms and an experimental X-ray dataset. First, we describe the phantoms and the performance measures used to compare our proposed dual approach (abbreviated as DP) to several state-of-the-art iterative reconstruction techniques. We conclude this section with results on an experimental dataset. All experiments are performed using Matlab in conjunction with the ASTRA toolbox [41].

³For h,v,d -convexity one uses the usual definition of convexity of a set but considers only line segments in the horizontal, vertical and diagonal directions.

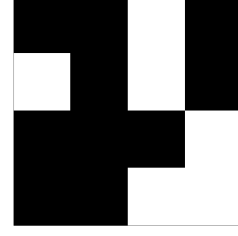


Figure 4. Example of a 4×4 binary image that is not h,v,d -convex but does permit a unique solution.

A. Phantoms

For the synthetic tests, we consider four phantoms shown in figure 5. All the phantoms are binary images of size 128×128 pixels. The grey levels are $u_0 = 0$ and $u_1 = 1$. The detector has 128 pixels, and the distance between the adjacent detectors is the same as the pixel size of the phantoms. We consider a parallel beam geometry for the acquisition of the tomographic data in all the simulation experiments.

B. Tests

We perform three different tests to check the robustness of the proposed method. First, we consider the problem of sparse projection data. For all the phantoms, we first start with 45 projections at angles ranging from 0 to π and subsequently reduce the number of projections. This setup is also known as sparse sampling, where the aim is to reduce the scan time by decreasing the number of angles.

Next, we consider a limited angle scenario. Such a situation usually arises in practice due to the limitations of the setup. For the test, we acquire projections in the range $[0, \theta_{\max}]$ for $\theta_{\max} \in \{5\pi/6, 2\pi/3, 7\pi/12, \pi/2\}$. Reconstruction of limited angle data is known to lead to so-called streak artifacts in the reconstructed image. Strategies to mitigate these streak artifacts include the use of regularization method with some prior information. We will experiment how the discrete tomography can lead to the removal of these artifacts.

Finally, we test the performance of the proposed method in the presence of noise. We consider an additive Gaussian noise in these experiments. We measure the performance of our approach for tomographic data with a signal-to-noise ratio (SNR) of $\{10, 20, 30, 50\}$ dB.

To avoid inverse crime [42] in all the test scenarios, we generate data using strip kernel and use Joseph kernel for modeling.

C. Comparison with other reconstruction methods

There exist a vast amount of reconstruction methods for tomography. Here, consider the following three:

LSQR: Least squares QR method described in [43]. We perform a total of 1000 iterations with a tolerance of 10^{-6} . We segment the resulting reconstruction using Otsu's thresholding algorithm [44].

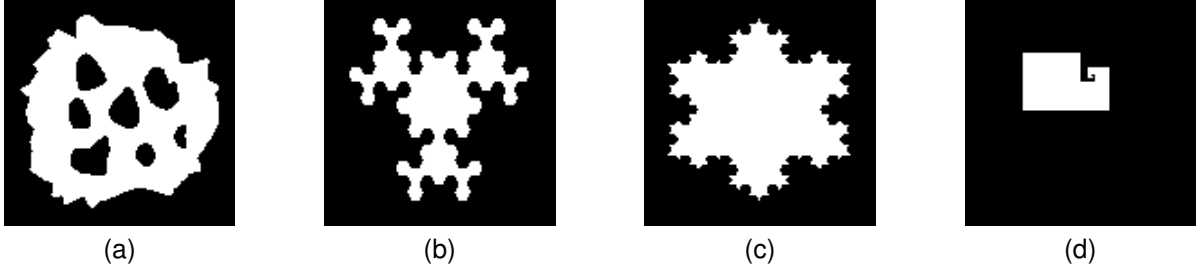


Figure 5. Phantom images used in the simulation experiments. (a) Phantom 1, (b) Phantom 2, (c) Phantom 3, (d) Phantom 4.

TV : The total-variation method leads to an optimization problem described below:

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{Ax} - \mathbf{y}\|^2 + \lambda \|\mathbf{Dx}\|_1,$$

where λ is a corresponding regularization parameter, and \mathbf{D} matrix captures the discrete gradient in both directions. We use the Chambolle-Pock method [37] to solve the above optimization problem with non-negativity constraints on the pixel values. For each case, we perform iterations till the relative duality gap reach a tolerance value of 10^{-4} . To avoid slow convergence, we scale the matrices \mathbf{A} and \mathbf{D} to have unit matrix norm. The regularization parameter λ is selected using Morozov's discrepancy principle using the noise level defined as $\|\mathbf{y} - \mathbf{Ax}_{\text{true}}\|^4$. Finally, We segment the TV-reconstructed image using Otsu's thresholding algorithm.

DART : We use the method described in [25] for DART on the above binary images. The grayvalues are taken to be the same as true grayvalues. We perform 20 Algebraic Reconstruction Method (ARM) iterations initially before performing 40 DART iterations. In each DART iteration, we do 3 algebraic reconstruction iterations. We use the segmented image as a result of the DART iterations to perform the further analysis of the method.

DP : We solve the dual formulation (14) with a smooth approximation of the ℓ_1 -norm (as discussed in section II) using L-BFGS method [45] with a maximum of 500 iterations.

D. Performance Measures

In order to evaluate the performance of reconstruction methods, we use the following two criteria.

RMS The root-mean-square (RMS) error

$$\text{RMS} \triangleq \|\mathbf{Ax}^* - \mathbf{y}\|,$$

measures how well the forward-projected reconstructed image matches the projection data. This measure is useful in practice, as it does not require

knowledge of the ground truth. The RMS value close to the noise level of the data considered as a good reconstruction.

Jl The Jaccard index $\text{JI} \triangleq 1 - \sum_{i=1}^N (\alpha_i + \beta_i) / N$ measures the similarity between the reconstructed image (\mathbf{x}^*) and the ground truth (\mathbf{x}^{true}) in a discrete sense. The parameters α and β represent missing and over-estimated pixels respectively and given by

$$\begin{aligned} \alpha_i &\triangleq (x_i^* = u_0) \times (x_i^{\text{true}} = u_1), \\ \beta_i &\triangleq (x_i^* = u_1) \times (x_i^{\text{true}} = u_0). \end{aligned}$$

The blue and red dots denote the missing and the overestimated pixels in Figure 7, 8, 9, 10, 11. If JI has high value (close to 1), the reconstruction is considered good. Although this measure is not readily applicable to real datasets, it is a handy measure to compare the various reconstruction methods on synthetic examples.

E. Experimental data setup

We use the experimental X-ray projection data of a carved cheese slice [46]. Figure 6 shows a high-resolution filtered back-projection reconstruction of the data. The cheese contains the letters C and T and the object is (approximately) binary with two grey levels corresponding to calcium-containing organic compounds of cheese and air. The dataset consists of projection data with three different resolutions (128×128 , 256×256 , 512×512) and the corresponding projection matrix modeling the linear operation of X-ray transform.

We perform two sets of experiments: (1) Sparse sampling with 15 angles ranging from 0 to 2π and (2) limited-angle using 15 projections from 0 to $\pi/2$.

F. Sparse projections test

Figure 7 presents the reconstruction results from various methods for phantoms 1. The tomographic data are generated for ten equidistant projection angles from 0 to $\pi/2$. The reconstruction results show the difference between the reconstructed image and the ground truth. It is evident that, compared to the other methods, the proposed method reconstructions are very close to the ground truth. The results from LSQR are the worst as it does not incorporate any prior information about the model. The TV method also leads to artifacts as it includes partial information about the model. The DART and DP are

⁴Since the data is generated from a different modeling kernel, we compute noise level from the mismatch of the projection of true image with the data. The discrepancy principle then selects the parameter which fits data close to such noise level.

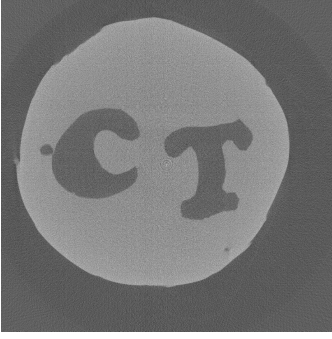


Figure 6. The high resolution (2000×2000 pixels) filtered back-projection reconstruction of the carved cheese from 360 projections from 0 to 2π .

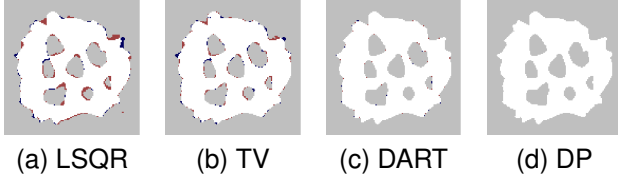


Figure 7. Limited projection test I for Phantom 1. Performance of various reconstruction methods with 10 projection angles from 0 to $\pi/2$.

very close to each other. For all the phantoms, we tabulate the data misfit (RMS) and Jaccard index (JI) in Table II.

We also show reconstructions with the proposed approach for a varying number of projection angles in Figure 8. We note that the problem becomes harder to solve as the number of projections gets smaller. Hence, we may also expect the reconstruction to become poorer. We see that the proposed approach can reconstruct almost correctly with as few as ten projection angles.

G. Limited angle test

Figure 9 shows the results of phantom 2 with various reconstruction methods for limited angle tomography (10 equispaced angles in the range 0 to $\pi/2$). It is visible that the reconstructions from DART and the proposed method are very close to the true images of the phantoms. The values for data misfit and Jaccard index for all the tests with each of the synthetic phantoms are tabulated in Table III.

We also look at how the reconstructions with the proposed method vary with limiting the angle (see Figure 10). As the angle gets limited, the reconstruction problem gets difficult. The proposed method can reconstruct almost correctly with angle limited to $\pi/2$.

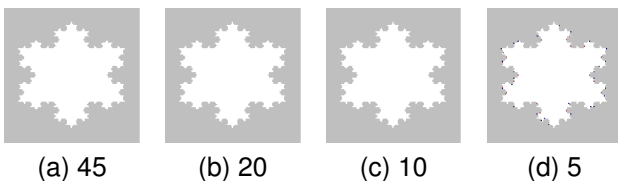


Figure 8. Limited projection test II for Phantom 3. Performance of proposed method vs number of projections.

Table II
LIMITED PROJECTION TEST PERFORMANCE MEASURES

Test	Phantom	LSQR		TV		DART		DP	
		RMS	JI	RMS	JI	RMS	JI	RMS	JI
45	P1	31.4	99.7	19.2	99.9	31.5	99.7	17.6	100
	P2	26.2	99.6	12	100	36.4	99	12	100
	P3	48.1	99	24.5	99.8	52.4	98.8	16.4	100
	P4	11.6	99.8	5	100	13.6	99.6	5	100
20	P1	54.3	97.8	34.5	98.9	24.1	99.5	12	100
	P2	60.5	95.8	12.8	99.9	17	99.6	8.9	100
	P3	54.2	97.6	27	99.4	40.2	98.9	11.1	100
	P4	20.8	99.3	3.4	100	6.8	99.8	3.4	100
10	P1	122.3	93.1	62.4	96.2	22.4	99.1	8.3	99.9
	P2	254.4	76	81	91.1	17	99	11.2	99.7
	P3	73.8	94.9	38.6	98	28.7	99	7.4	100
	P4	42.7	97	5.4	99.9	5.9	99.7	2.4	100
5	P1	269.7	82.4	77.5	87	23.7	97.3	52.8	90.7
	P2	370.8	63.2	158.5	71.3	38.8	76.1	72.8	73.9
	P3	80	90.9	64.9	93.4	22.5	98.5	21.5	97.6
	P4	79.7	86.4	22.9	95.8	4.7	99.6	1.7	100

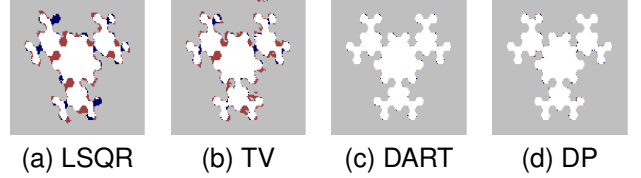


Figure 9. Limited angle test I for Phantom 2. Performance of various reconstruction methods with 10 projection angles from 0 to $\pi/2$.

H. Noisy projection test

This test aims to check the sensitivity of the proposed method to noise in the data. We perform four experiments with varying levels of Poisson noise in the data. In particular, we use expected incident photon counts $I_0 = \{10^6, 10^4, 10^3, 10^2\}$. Note that the lower incident photon counts lead to low SNR. Figure 11 shows the results on phantoms 1 and 2 for increasing noise level. We see that the reconstruction is stable against a moderate amount of noise and degrades gradually as the noise level increases.

I. Real data test

We look at the results of reconstructions from the proposed method for two sets of experiments at various resolutions and compare them with the reconstructions from LSQR and TV.

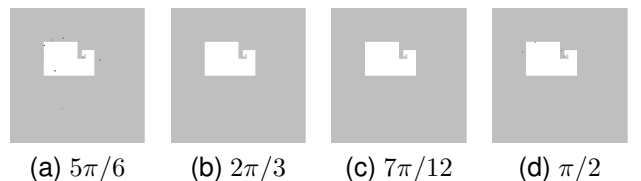


Figure 10. Limited angle test II for Phantom 4. Performance of proposed method vs maximum angle.

Table III
LIMITED ANGLE TEST PERFORMANCE MEASURES

Test	Phantom	LSQR		TV		DART		DP	
		RMS	JI	RMS	JI	RMS	JI	RMS	JI
$5\pi/6$	P1	158.3	91.6	101.9	93.9	23.6	99.1	5.4	100
	P2	215.9	79.5	79.6	91.3	16.4	99.2	5	100
	P3	82.4	94.5	53.3	96.7	28.5	99	5.3	100
	P4	54.9	95.6	4.7	99.9	5.3	99.7	9.5	99.2
$2\pi/3$	P1	199.7	88.4	141.7	91.5	18.7	99.3	6.7	100
	P2	189.3	77.8	146.5	83	22	98	13.3	99.3
	P3	96.2	92.9	70.5	95	27.4	98.6	3.4	100
	P4	68.2	91.7	9.1	99.3	7.1	99.5	0.8	100
$7\pi/12$	P1	213.2	87	164.6	89.7	20.1	99.2	7.7	99.9
	P2	231.8	76.4	182.1	81.2	21.7	98	17	98.5
	P3	105.5	92.3	79.9	94.3	28.3	98.5	3.7	100
	P4	84.5	89.7	8.1	99.4	7.6	99.5	0.9	100
$\pi/2$	P1	258.6	84.9	293	85.2	19.1	99.3	18	99.2
	P2	205.3	75.3	192.5	80.9	22.7	98.2	18.2	98.5
	P3	127.1	90.5	99.2	93.5	31	98.3	4	100
	P4	128.2	82.9	27.1	96.7	7.8	0.99.5	7.2	99.5

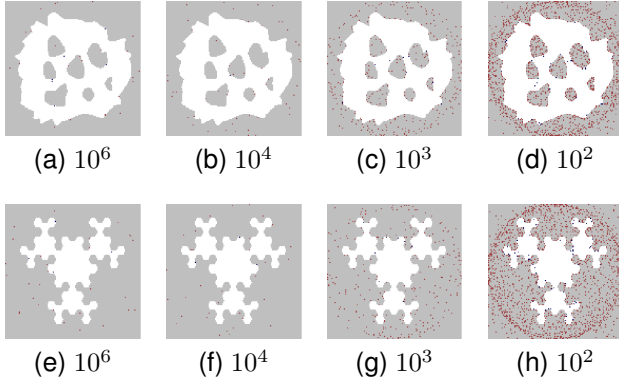


Figure 11. Noisy Projection test on phantoms 1 and 2. Performance of proposed method vs expected total incident photon counts.

Since the ground truth image is not available, we compare these reconstructions visually.

In order to apply DP, we first need to estimate the grey values of the object. The object, a thin slice of cheese, consists of two materials; the organic compound of the cheese, which is we assume to be homogeneous, and air. For air, the grey value is zero. We estimate the grey value of the organic compound of cheese from the histogram of an FBP reconstruction provided with the data. Figure 12 represents the histogram. We obtain a value of 0.00696 for this compound.

We first consider the reconstructions from sparse angular sampling (15 projections spanning from 0 to 2π). The tests are performed on two different resolutions: 128×128 , and 512×512 . Figure 13 presents the results of the reconstructions with LSQR, TV, and DP for these resolutions. The DP reconstruction is discrete and correctly identifies the letters C and T with also a little hole at the left side of C. Although LSQR reconstruction is poor for 128×128 , it improves with the resolution. We still see the mild streak artifacts in these reconstructions. The TV reconstruction removes these streak

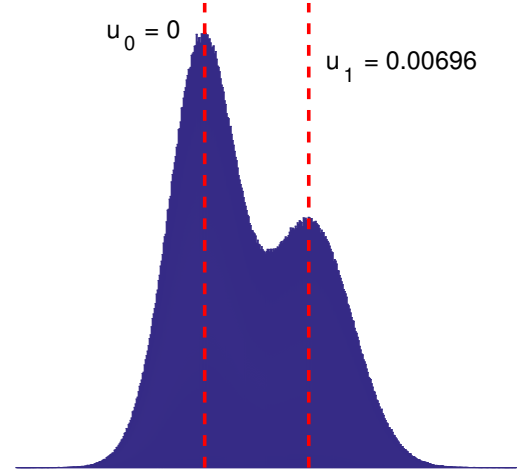


Figure 12. Histogram of filtered backprojection image of the carved cheese.

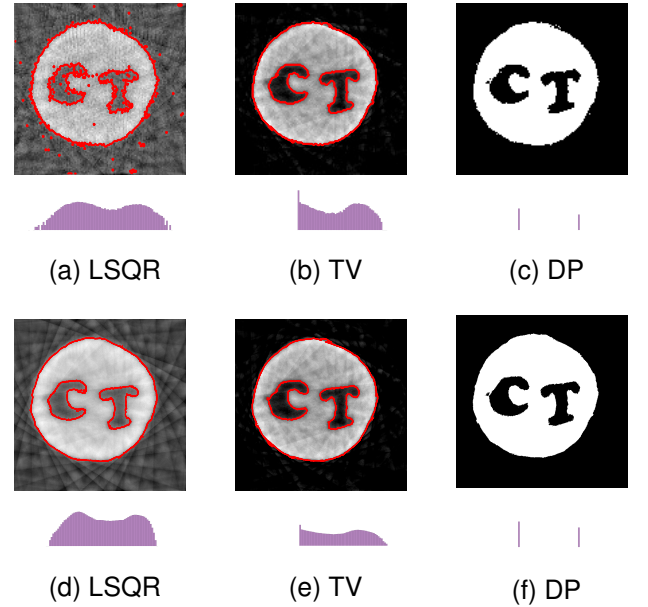


Figure 13. Real Data Test I - Sparse projection tomography. Performance of various methods with different resolutions. Top row corresponds to 128×128 pixels. Bottom row corresponds to 512×512 pixels. Figure below each image denote the histogram. The red contours represent the thresholded image.

artifacts but fails to identify the homogeneous cheese slice correctly.

In the second test, we limit the projection angles to $0 - \pi/2$. Figure 14 shows the results of the reconstructions from LSQR, TV, and DP for two different resolutions. We see that the reconstructions improve with increment in the resolution. LSQR reconstructions have severe streak artifacts, which are the characteristics of the limited data tomography. TV and DP reconstructions do not possess these artifacts. TV reconstruction can capture the shape of the cheese, but it blurs out the carved parts C and T. DP reconstructs the shape of cheese quite accurately and has C and T are also identified.

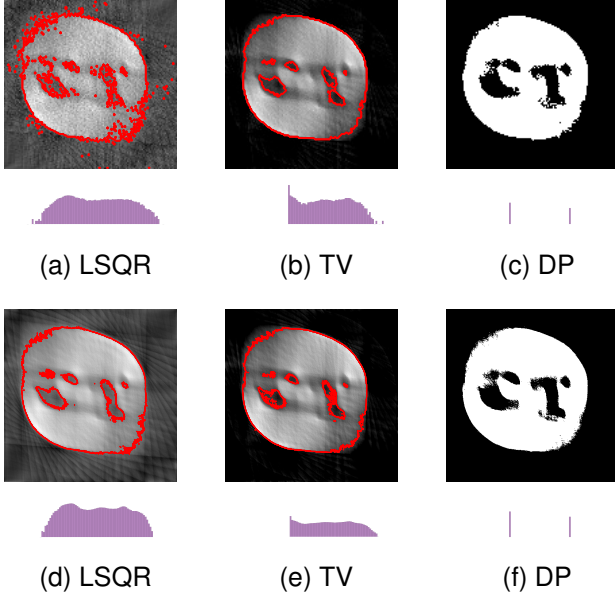


Figure 14. Real Data Test II - Limited angle tomography. Performance of various methods with different resolutions. Top row corresponds to 128×128 pixels. Bottom row corresponds to 512×512 pixels. Figure below each image denote the histogram. The red contours represent the thresholded image.

V. CONCLUSION

We presented a novel convex formulation for binary tomography. The problem is primarily a generalized LASSO problem that can be solved efficiently when the system matrix has full row rank or full column rank. Solving the dual problem is not guaranteed to give the optimal solution, but can at least be used to construct a feasible solution. In a complete enumeration of small binary test cases (images of $n \times n$ pixels for $n = 2, 3, 4$) we observed that if the problem has a unique solution, then the proposed dual approach finds it. In case the problem has multiple solutions, the dual approach finds the part that is common in all solutions. Based on these experiments we conjecture that this holds in the general case (beyond the small test images). Of course, verifying beforehand if the problem has a unique solution may not be possible.

We test the proposed method on numerical phantoms and real data, showing that the method compares favorably to some of the state-of-the-art reconstruction techniques (Total Variation, DART). The proposed method is also reasonably stable against a moderate amount of noise.

We currently assume the grey levels are known apriori. Extension to multiple (i.e., more than 2) unknown grey levels is possible in the same framework but will be left for future work. To make the method more robust against noise additional regularization may be added.

APPENDIX A PROOFS

A. Proposition 1

Proof. In (4), the $g_1(\nu)$ has a closed-form expression for general \mathbf{A} . To see this, let us first denote

$$f(\mathbf{x}, \nu) \triangleq \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \nu^T \mathbf{x}. \quad (\text{A.1})$$

We are interested in the infimum value of this function with respect to \mathbf{x} . To obtain this, we set the gradient of f with respect to \mathbf{x} to zero

$$\nabla_{\mathbf{x}} f = \mathbf{A}^T (\mathbf{A}\mathbf{x} - \mathbf{y}) + \nu = 0.$$

Since \mathbf{A} is a general matrix, it may be rank-deficient. Hence, the optimal value \mathbf{x}^* only exists if ν is in the range of \mathbf{A}^T and it is given by

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^\dagger (\mathbf{A}^T \mathbf{y} - \nu).$$

Substituting this value in (A.1), we get the following:

$$\begin{aligned} g_1(\nu) &= \inf_{\mathbf{x}} f(\mathbf{x}, \nu) \\ &= \begin{cases} f(\mathbf{x}^*, \nu) & \nu \in \mathcal{R}_{\mathbf{A}^T} \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} -\frac{1}{2} \|\nu - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 + \frac{1}{2} \mathbf{y}^T \mathbf{y} & \nu \in \mathcal{R}_{\mathbf{A}^T} \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

Now we return to the dual objective in equation (4). Substituting the explicit forms for $g_1(\nu)$ from above and $g_2(\nu)$ from equation (7), we get the expression for the dual objective:

$$g(\nu) = \begin{cases} -\frac{1}{2} \|\nu - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 - \|\nu\|_1 + \frac{1}{2} \mathbf{y}^T \mathbf{y} & \nu \in \mathcal{R}_{\mathbf{A}^T} \\ -\infty & \text{otherwise.} \end{cases}$$

The above dual objective leads to the following maximization problem with respect to the dual variable ν

$$\nu^* = \operatorname{argmax}_{\nu \in \mathbb{R}^N} g(\nu).$$

As we are only interested in the maximum value of the dual objective, the space of ν can be constrained to the range of \mathbf{A}^T . This is valid as the dual objective is $-\infty$ for the ν outside the range of \mathbf{A}^T . Hence, the maximization problem reduced to the following minimization problem:

$$\nu^* = \operatorname{argmin}_{\nu \in \mathcal{R}_{\mathbf{A}^T}} \frac{1}{2} \|\nu - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 + \|\nu\|_1.$$

□

B. Corollary 1

Proof. Since the search space for the dual variable ν is constrained to the range of \mathbf{A}^T , we can express this variable as $\mathbf{A}^T \mu$, where $\mu \in \mathbb{R}^m$. Substituting $\nu = \mathbf{A}^T \mu$ in (12), we get

$$\mu^* = \operatorname{argmin}_{\mu \in \mathbb{R}^m} \frac{1}{2} \|\mathbf{A}^T (\mu - \mathbf{y})\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 + \|\mathbf{A}^T \mu\|_1. \quad (\text{A.2})$$

Using the identities $(\mathbf{A}^T \mathbf{A})^\dagger \mathbf{A}^T = \mathbf{A}^\dagger$ and $\mathbf{A} = \mathbf{A} \mathbf{A}^\dagger \mathbf{A}$ [47], we can re-write the weighted norm $\|\mathbf{A}^T \mathbf{r}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2$ as $\|\mathbf{A} \mathbf{A}^\dagger \mathbf{r}\|^2$. The dual problem (A.2) now reads

$$\mu^* = \operatorname{argmin}_{\mu \in \mathbb{R}^m} \frac{1}{2} \|\mathbf{A} \mathbf{A}^\dagger (\mu - \mathbf{y})\|^2 + \|\mathbf{A}^T \mu\|_1.$$

Correspondingly, the primal solution \mathbf{x}^* related to the dual optimal μ^* is

$$\mathbf{x}^* = \operatorname{sign}(\phi^*) = \operatorname{sign}(\nu^*) = \operatorname{sign}(\mathbf{A}^T \nu^*).$$

□

C. Proposition 2

Proof. The primal problem for binary tomography problem with grey levels $u_0 < u_1$ can be stated as:

$$\begin{aligned} \mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \left(\inf_{\phi} \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|^2 \right), \\ \text{subject to } \mathbf{x} = u_0 \mathbf{1} + (u_1 - u_0) H(\phi), \end{aligned}$$

where $H(\cdot)$ denotes the Heaviside function and ϕ is an auxiliary variable. Such problem admits a Lagrangian

$$\mathcal{L}(\mathbf{x}, \phi, \boldsymbol{\nu}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|^2 + \boldsymbol{\nu}^T (\mathbf{x} - u_0 \mathbf{1} - (u_1 - u_0) H(\phi)),$$

where $\boldsymbol{\nu} \in \mathbb{R}^N$ is a Lagrangian multiplier (also known as dual variable) corresponding to the equality constraint. This gives rise to a dual function

$$\begin{aligned} g(\boldsymbol{\nu}) = \underbrace{\inf_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|^2 + \boldsymbol{\nu}^T \mathbf{x} \right\}}_{g_1(\boldsymbol{\nu})} \\ + \underbrace{\inf_{\phi} \left\{ -(u_1 - u_0) \boldsymbol{\nu}^T H(\phi) \right\} - u_0 \boldsymbol{\nu}^T \mathbf{1}}_{g_2(\boldsymbol{\nu})}. \end{aligned}$$

Since we already know $g_1(\boldsymbol{\nu})$ (refer to equation (6)), we require the explicit form for $g_2(\boldsymbol{\nu})$. For its computation, we use the componentwise property of the Heaviside function to separate the infimum.

$$\begin{aligned} g_2(\boldsymbol{\nu}) &= \sum_{i=1}^N \inf_{\phi_i} \left\{ -(u_1 - u_0) \nu_i H(\phi_i) \right\} \\ &= - \sum_{i=1}^N \sup_{\phi_i} \left\{ (u_1 - u_0) \nu_i H(\phi_i) \right\} \end{aligned}$$

Since the range of Heaviside function is only two values, namely $\{0, 1\}$, we get the simple form for $g_2(\boldsymbol{\nu})$:

$$\begin{aligned} g_2(\boldsymbol{\nu}) &= - \sum_{i=1}^N q(\nu_i) \\ \text{where } q(\nu_i) &= \begin{cases} (u_1 - u_0) \nu_i & \text{if } \nu_i > 0 \\ 0 & \text{otherwise} \end{cases} \\ &= (u_1 - u_0) \max(\nu_i, 0). \end{aligned}$$

This infimal value is attained at $\phi^* = H(\boldsymbol{\nu})$. Now the dual problem reads

$$\begin{aligned} \boldsymbol{\nu}^* = \underset{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\boldsymbol{\nu} - \mathbf{A}^T \mathbf{y}\|_{(\mathbf{A}^T \mathbf{A})^\dagger}^2 + \right. \\ \left. \sum_i (u_1 - u_0) \max(\nu_i, 0) + u_0 \boldsymbol{\nu}^T \mathbf{1} \right\}. \end{aligned} \quad (\text{A.3})$$

We note that the last two terms in the dual objective can be compactly represented by

$$p(\boldsymbol{\nu}) = \sum_i |u_0| \max(-\nu_i, 0) + |u_1| \max(\nu_i, 0),$$

where $p(\cdot)$ is known as an asymmetric one-norm. The optimal point of the problem (A.3) is denoted by $\boldsymbol{\nu}^*$ and the corresponding primal optimal is retrieved using

$$\mathbf{x}^* = u_0 \mathbf{1} + (u_1 - u_0) H(\phi^*) = u_0 \mathbf{1} + (u_1 - u_0) H(\boldsymbol{\nu}^*).$$

□

D. Proposition 3

Proof. The minimization problem for the proximal operator of an asymmetric one-norm function $p(\cdot)$ reads

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|^2 + \lambda p(\mathbf{x}), \quad (\text{A.4})$$

where $\lambda > 0$ is a parameter. Since the function is convex, we get the following from the first-order optimality condition [27]:

$$\begin{aligned} \mathbf{0} &\in \partial f(\mathbf{x}^*), \\ &\in \mathbf{x}^* - \mathbf{z} + \lambda \partial p(\mathbf{x}^*), \end{aligned} \quad (\text{A.5})$$

where \mathbf{x}^* is an optimal point of (A.4), and $\partial p(\mathbf{x})$ is a sub-differential of function $p(\cdot)$ at \mathbf{x} . This sub-differential is

$$\partial p(x_i) = \begin{cases} |u_1| & x_i > 0 \\ [-|u_0|, |u_1|] & x_i = 0 \\ -|u_0| & x_i < 0 \end{cases}.$$

Now coming back to the first-order optimality condition in (A.5), we get the explicit form for optimal solution \mathbf{x}^* :

$$x_i^* = \begin{cases} z_i - \lambda |u_1| & z_i \geq \lambda |u_1| \\ 0 & -\lambda |u_0| \leq z_i \leq \lambda |u_1| \\ z_i + \lambda |u_0| & z_i \leq -\lambda |u_0| \end{cases}.$$

We recognize this function as an asymmetric soft-thresholding function. □

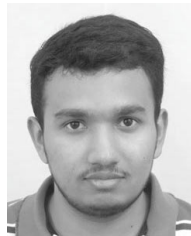
ACKNOWLEDGMENT

This work is part of the Industrial Partnership Programme (IPP) ‘Computational sciences for energy research’ of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). This research programme is co-financed by Shell Global Solutions International B.V. The second author is financially supported by the Netherlands Organisation for Scientific Research (NWO) as part of research programme 613.009.032. We thank four anonymous reviewers and associate editor for their comments and sufficiently improving the manuscript.

REFERENCES

- [1] G. T. Herman and A. Kuba, *Advances in discrete tomography and its applications*. Springer Science & Business Media, 2008.
- [2] J. L. Sanz, E. B. Hinkle, and A. K. Jain, “Radon and projection transform-based computer vision: algorithms, a pipeline architecture, and industrial applications,” 1988.
- [3] B. Sharif and B. Sharif, “Discrete tomography in discrete deconvolution: Deconvolution of binary images using ryser’s algorithm,” *Electronic Notes in Discrete Mathematics*, vol. 20, pp. 555–571, 2005.
- [4] M. C. San Martin, N. P. J. Stamford, N. Dammerova, N. E. Dixon, and J. M. Carazo, “A structural model for the escherichia coli dnab helicase based on electron microscopy data,” *Journal of structural biology*, vol. 114, no. 3, pp. 167–176, 1995.
- [5] J.-M. Carazo, C. Sorzano, E. Rietzel, R. Schröder, and R. Marabini, “Discrete tomography in electron microscopy,” in *Discrete Tomography*. Springer, 1999, pp. 405–416.
- [6] M. Servieres, J. Guédon, and N. Normand, “A discrete tomography approach to pet reconstruction,” in *Fully 3D Reconstruction In Radiology and Nuclear Medicine*, no. 1, 2003, pp. 1–4.

- [7] B. M. Carvalho, G. T. Herman, S. Matej, C. Salzberg, and E. Vardi, "Binary tomography for triplane cardiography," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 1999, pp. 29–41.
- [8] P. A. Midgley and R. E. Dunin-Borkowski, "Electron tomography and holography in materials science," *Nature materials*, vol. 8, no. 4, p. 271, 2009.
- [9] M. Balaskó, A. Kuba, A. Nagy, Z. Kiss, L. Rodek, and L. Ruskó, "Neutron-, gamma- and x-ray three-dimensional computed tomography at the budapest research reactor site," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 542, no. 1-3, pp. 22–27, 2005.
- [10] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo, "Three-dimensional atomic imaging of crystalline nanoparticles," *Nature*, vol. 470, no. 7334, p. 374, 2011.
- [11] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. Midgley, J. Hernandez, U. Kaiser, E. Encina, E. Coronado, and G. Van Tendeloo, "3d imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.
- [12] R. J. Gardner, P. Gritzmann, and D. Prangenberg, "On the computational complexity of reconstructing lattice sets from their X-rays," *Discrete Mathematics*, vol. 202, no. 1–3, pp. 45–71, 1999. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0012365X98003471>
- [13] A. E. Yagle, "An algebraic solution for discrete tomography," in *Discrete Tomography*. Springer, 1999, pp. 265–284.
- [14] L. Hajdu and R. Tijdeman, "Algebraic aspects of discrete tomography," *Journal für die Reine und Angewandte Mathematik*, vol. 534, pp. 119–128, 2001.
- [15] S. Matej, A. Vardi, G. T. Herman, and E. Vardi, "Binary tomography using gibbs priors," in *Discrete Tomography*. Springer, 1999, pp. 191–212.
- [16] M. T. Chan, G. T. Herman, and E. Levitan, "Probabilistic modeling of discrete images," in *Discrete Tomography*. Springer, 1999, pp. 213–235.
- [17] T. Frese, C. A. Bouman, and K. Sauer, "Multiscale bayesian methods for discrete tomography," in *Discrete Tomography*. Springer, 1999, pp. 237–264.
- [18] Y. Censor and S. Matej, "Binary steering of nonbinary iterative algorithms," in *Discrete tomography*. Springer, 1999, pp. 285–296.
- [19] Y. Vardi and C.-H. Zhang, "Reconstruction of binary images via the em algorithm," in *Discrete Tomography*. Springer, 1999, pp. 297–316.
- [20] T. Capricelli and P. Combettes, "A convex programming algorithm for noisy discrete tomography," in *Advances in Discrete Tomography and Its Applications*. Springer, 2007, pp. 207–226.
- [21] T. Schüle, C. Schnörr, S. Weber, and J. Hornegger, "Discrete tomography by convex-concave regularization and dc programming," *Discrete Applied Mathematics*, vol. 151, no. 1-3, pp. 229–243, 2005.
- [22] A. Tuysuzoglu, Y. Khoo, and W. C. Karl, "Fast and robust discrete computational imaging," *Electronic Imaging*, vol. 2017, no. 17, pp. 49–54, 2017.
- [23] J. Kuske, P. Swoboda, and S. Petra, "A novel convex relaxation for non-binary discrete tomography," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2017, pp. 235–246.
- [24] E. Barucci, S. Brunetti, A. Del Lungo, and M. Nivat, "Reconstruction of lattice sets from their horizontal, vertical and diagonal X-rays," *Discrete Mathematics*, vol. 241, no. 1-3, pp. 65–78, 2001.
- [25] K. J. Batenburg and J. Sijbers, "Dart: a practical reconstruction algorithm for discrete tomography," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2542–2553, 2011.
- [26] G. T. Herman and A. Kuba, *Discrete Tomography: Foundations, Algorithms, and Applications*. Springer Science & Business Media, 1999.
- [27] R. Rockafellar, "Convex analysis," 1970.
- [28] M. Slater, "Lagrange multipliers revisited," Cowles Foundation for Research in Economics, Yale University, Tech. Rep., 1959.
- [29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [30] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE transactions on medical imaging*, vol. 13, no. 2, pp. 290–300, 1994.
- [31] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 534–548, 1993.
- [32] I. Daubechies, M. Debrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [33] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [34] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [35] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM journal on imaging sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [36] K. J. Arrow, L. Hurwicz, and H. Uzawa, "Studies in linear and non-linear programming," 1958.
- [37] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [38] R. Gu and A. Dogandžić, "Projected nesterov's proximal-gradient algorithm for sparse signal recovery," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3510–3525, 2017.
- [39] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2008.
- [40] E. Barucci, A. Del Lungo, M. Nivat, and R. Pinzani, "Reconstructing convex polyominoes from horizontal and vertical projections," *Theoretical Computer Science*, vol. 155, no. 2, pp. 321–347, 1996.
- [41] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The astra toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.
- [42] J. Kaipio and E. Somersalo, "Statistical inverse problems: discretization, model reduction and inverse crimes," *Journal of computational and applied mathematics*, vol. 198, no. 2, pp. 493–504, 2007.
- [43] C. C. Paige and M. A. Saunders, "Lsq: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software (TOMS)*, vol. 8, no. 1, pp. 43–71, 1982.
- [44] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [45] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical programming*, vol. 45, no. 1-3, pp. 503–528, 1989.
- [46] T. A. Bubba, M. Juvonen, J. Lehtonen, M. März, A. Meaney, Z. Purisha, and S. Siltanen, "Tomographic x-ray data of carved cheese," *arXiv preprint arXiv:1705.05732*, 2017.
- [47] S. Barnett, *Matrices: methods and applications*. Clarendon Press, 1990.



Ajinkya Kadu received BSc. and MSc. degree in Aerospace Engineering from the Indian Institute of Technology, Bombay, India, in 2015. He is working towards the PhD. degree at Mathematical Institute of Utrecht University, The Netherlands. His PhD. is part of 'Computational Sciences for Energy Research', a research programme of NWO in partnership with Shell. His research interests include computational imaging, full-waveform inversion and distributed optimization.



Tristan van Leeuwen received his BSc. and MSc. in Computational Science from Utrecht University. He obtained his PhD. in geophysics at Delft University in 2010. After spending some time as a postdoctoral researcher at the University of British Columbia in Vancouver, Canada and the Centrum Wiskunde & Informatica in Amsterdam, the Netherlands, he returned to Utrecht University in 2014 as an assistant professor at the mathematical institute. His research interests include: inverse problems, computational imaging, tomography and numerical optimization.