Total variation regularization for large-scale X-ray tomography

Keijo Hämäläinen[1], Lauri Harhanen[2], Andreas Hauptmann[3], Aki Kallonen[1], Esa Niemi[3] and Samuli Siltanen[3]

[1]Department of Physics
University of Helsinki
Gustaf Hällströmin katu 2a, Helsinki FI-00014

[2]Department of Mathematics and Systems Analysis
Aalto University
Otakaari 1 M, Espoo FI-00076

[3]Department of Mathematics and Statistics
University of Helsinki
Gustaf Hällströmin katu 2b, Helsinki FI-00014

ABSTRACT

A new large-scale computational total variation regularization algorithm is introduced and tested with examples arising from X-ray tomography with sparsely sampled data. The total variation penalty term is discretized using a basis of discontinuous functions. The approach is motivated by discontinuous Galerkin methods and leads to an additional term of the jump part of total variation. The proposed algorithm combines the usage of the jump part with a subgradient descent scheme. A comparison is provided with the gradient-based projected Barzilai-Borwein method which uses a smoothly approximated total variation penalty. The above two methods are examples of total variation regularization algorithms that can be applied to large-scale tomographic problems in reasonable computation time. A comparison between the methods shows that they use roughly equal computational resources and that the new method produces somewhat blockier reconstructions. Although the test problems are two-dimensional, both methods can be applied to three-dimensional situations as well.

Keywords: X-ray tomography, first-order methods, Barzilai-Borwein, subgradient descent, total variation.

2000 Mathematics Subject Classification: 15A29, 94A08, 92C55.

1 Introduction

In X-ray tomography one has available projection images of a physical body taken from different directions. The goal is to recover the inner structure of the body from the measured data. In this work we concentrate on two-dimensional imaging and model a slice through the body by $\Omega \subset \mathbb{R}^2$. Mathematically, the measurement $\widetilde{g}$ is a collection of line integrals of the non-negative attenuation coefficient function $f : \Omega \to \mathbb{R}$ along the paths of the X-rays. Here $\widetilde{g} = \mathcal{A}f$ with $\mathcal{A}$ a linear operator modeling the measurement geometry. In practice we are given noisy data $g = \widetilde{g} + \varepsilon$. Reconstructing

f from g is an ill-posed inverse problem, so any successful computational inversion method used for tomographic imaging needs to be regularized. See (Kak and Slaney, 1988; Natterer, 1986; Mueller and Siltanen, 2012) for introductions to tomographic imaging.

In most applications of X-ray tomography to medical imaging and non-destructive testing, the attenuation coefficient contains smoothly varying subdomains divided by sharp boundaries. Total variation (TV) regularization is known to be a good reconstruction method for such cases, as it produces noise-robust and edge-preserving reconstructions. The TV-regularized solution is defined as the minimizer of the functional

$$\mathcal{L}_{TV}(f) := \left\|\mathcal{A}f - g\right\|_{L^2}^2 + \alpha TV(f), \tag{1.1}$$

where $\alpha > 0$ is a regularization parameter and $TV(f)$ is the total variation of f defined on a bounded set $\Omega \subset \mathbb{R}^2$ by

$$TV(f) := \sup\left\{\int_\Omega -f \mathrm{div}\varphi \,\mathrm{d}x \ : \ \varphi \in [C_0^1(\Omega)]^2, \|\varphi\|_\infty \leq 1\right\}.$$

The idea of total variation penalty methods was introduced by Rudin, Osher and Fatemi (ROF) (Rudin, Osher and Fatemi, 1992) for removing noise from images. They made the crucial observation that minimizing the total variation (subject to two other conditions regarding the mean and variance of the image) enables better restoration of images containing sharp edges. The total variation regularization is analyzed for example in (Acar and Vogel, 1994). In addition to noise removal, it has been applied to recovering blurred noisy images (Chambolle and Lions, 1995; Vogel and Oman, 1998; Chan and Wong, 1998; Bertalmio, Caselles, Rouge and Solé, 2003; Combettes and Pesquet, 2004). TV regularization can be seen as a regularization method finding solutions whose derivatives are sparse; sparsity requirement compensates under-sampled data in reconstruction as shown in (Daubechies, Defrise and De Mol, 2004; Candès, Romberg and Tao, 2006; Grasmair, Haltmeier and Scherzer., 2008; Hämäläinen, Kallonen, Kolehmainen, Lassas, Niinimäki and Siltanen, 2012). There is considerable recent interest in reducing the X-ray dose to patients going through CT imaging. Several approaches are based on modifying imaging parameters, see (Tsapaki, Aldrich, Sharma, Staniszewska, Krisanachinda, Rehani, Hufton, Triantopoulou, Maniatis, Papailiou and Prokop, September 2006; McCollough, Primak, Braun, Kofler, Yu and Christner, 2009; Yu, Liu, Leng, Kofler, Ramirez-Giraldo, Qu, Christner, Fletcher and McCollough, 2009). However, a more efficient way of reducing dose is to take fewer projection images. The reconstruction problem then becomes more ill-posed, and novel large-scale tomographic algorithms are called for. Compared to the classical filtered back-projection (FBP) algorithm, TV regularization performs better in the case when the number of projection images is small, see for example (Kolehmainen, Siltanen, Järvenpää, Kaipio, Koistinen, Lassas, Pirttilä and Somersalo, 2003, Figure 13) or (Mueller and Siltanen, 2012, Figure 9.5).

Practical total variation regularization is often based on Sobolev functions $f \in W^{1,1}(\Omega)$. In that case the total variation simplifies to

$$TV(f) = \||\nabla f|\|_{L^1} = \int_\Omega |\nabla f| \,\mathrm{d}x = \int_\Omega \sqrt{(\partial_1 f)^2 + (\partial_2 f)^2},$$

where ∇f denotes the weak gradient of f.

The novel computational approach introduced in this paper is based on deriving a representation of the total variation with respect to a basis of piecewise continuous functions on a finite dimensional mesh $\mathcal{T}_h$ and the union of cell boundaries J_h. In this case the total variation is given by

$$TV_d(f) = \int_{J_h} |f^+ - f^-| \, \mathrm{d}s + \sum_{T \in \mathcal{T}_h} \int_T |\nabla f| \, \mathrm{d}x,$$

where, roughly speaking, f^+ and f^- are the values of f on different sides of the pixel boundary. This approach is motivated by discontinuous Galerkin finite element methods, such as in the recent publication (Bartels, 2012). We discuss that only discretizing with respect to a space of piecewise constant functions does not lead to adequate reconstructions in the setting of X-ray tomography. Instead we introduce the jump part of the total variation as an additional term to a subgradient descent scheme.

The attenuation coefficient f is a nonnegative function. This *a priori* information can be taken into account by considering the minimization problem

$$\min_{f \geq 0} \mathcal{L}_{TV}(f). \tag{1.2}$$

Enforcing non-negativity in the reconstruction improves tomographic reconstructions considerably, especially so when the data is sparsely sampled. This can be explained by a simple informal argument: assume that the unknown f is discretized using N pixels. Then the discrete unknown lives in $\mathbb{R}^N$ and must approximately satisfy a finite set of linear equations specified by $\mathcal{A}f = g$. Requiring that the unknown belongs to the part of $\mathbb{R}^N$ where are the coordinates are nonnegative restricts the search space to a tiny fraction 2^{-N} of the full $\mathbb{R}^N$.

Let us provide the reader with an idea of the size of N in practical applications. In dental imaging it is important to see details with size $\frac{1}{10}$mm within a region of interest roughly of size 50×50×50mm, leading to $N = 125\,000\,000$. See (Kolehmainen, Vanne, Siltanen, Järvenpää, Kaipio, Lassas and Kalke, 2006) for a concrete example.

We propose in Section 2 a method for solving the minimization problem (1.2) by employing TV_d and requiring nonnegativity. The basic motivation is to take advantage of the properties of the space BV of functions of bounded variation (piecewise continuous functions belong to BV). In other words, our algorithm is based on a discontinuity assumption on the target f for which we derive the corresponding total variation term and a modified Euler-Lagrange equation using a subgradient representation, similar to (Chambolle, 2004).

Advantages of the proposed method include

- simple enforcement of nonnegativity,
- possibility of region-of-interest imaging by varying the pixel size spatially, and
- applicability to large-scale problems.

The new algorithm is called *discontinuity-based projected subgradient descent* (DB-PSGD). We will compare it to another large-scale approach used in (Barzilai and Borwein, 1988; Niemi, 2010; Jensen, Jørgensen, Hansen and Jensen, 2011; Park, Song, Kim, Park, Kim, Liu, Suh and Song, 2012; Niu and Zhu, 2012) and based on smoothing out the non-differentiability in the objective

functional and applying a projected Barzilai-Borwein (PBB) optimization method to the resulting minimization problem. Our two test cases are two-dimensional X-ray tomography problems based on simulated data from the Shepp-Logan phantom and on X-ray projection data measured from a walnut. We stress that both DB-PSGD and PBB methods can readily be generalized to three-dimensional tomography.

The focus of this work is on fast computational solution of practical large-scale tomographic problems. Established methods such as primal-dual approaches (Chan, Golub and Mulet, 1999; Wu and Tai, 2010) require solving a nonlinear equation either directly or iteratively. Direct solution is out of the question in large-scale situations, so iterative methods such as conjugate gradients must be used. However, this is a major computational effort as each iteration requires a projection and back-projections step. The advantage of both DB-PSGD and PBB methods over primal-dual approaches is the significantly smaller number of such steps.

Several computational methods for minimizing the expression (1.1) have been introduced, including quadratic programming (Lassas and Siltanen, 2004; Kolehmainen, Lassas, Niinimäki and Siltanen, 2012) a lagged diffusivity method (Dobson and Vogel, 1997), domain decomposition methods (Fornasier and Schönlieb, 2009; Fornasier, Langer and Schönlieb, 2010), Bregman distance methods (Osher, Burger, Goldfarb, Xu and Yin, 2005; Yin, Osher, Goldfarb and Darbon, 2008; Goldstein and Osher, 2009; Cai, Osher and Shen, 2009; Zhang, Burger and Osher, 2011), primal-dual methods (Chan et al., 1999; Chan and Chen, 2006; Esser, Zhang and Chan, 2010; Nesterov, 2011), finite element methods (Feng and Prohl, 2003; Bartels, 2012), and other methods (Li and Santosa, 1996; Vogel and Oman, 1998; Wang, Yang, Yin and Zhang, 2008; Hale, Yin and Zhang, 2010; Chambolle, Levine and Lucier, 2011). Further treatments of total variation regularization and related methods can be found in the books (Vogel, 2002; Osher and Fedkiw, 2003; Chan and Shen, 2005; Scherzer, Grasmair, Grossauer, Haltmeier and Lenzen, 2009; Hansen, 2010). Total variation regularization has been applied to tomographic problems for instance in (Delaney and Bresler, 1998; Kolehmainen et al., 2003; Kolehmainen et al., 2006; Liao and Sapiro, 2008; Sidky and Pan, 2008; Herman and Davidi, 2008; Tang, Nett and Chen, 2009; Duan, Zhang, Xing, Chen and Cheng, 2009; Bian, Han, Sidky, Cao, Lu, Zhou and Pan, 2010; Jensen et al., 2011; Tian, Jia, Yuan, Pan and Jiang, 2011). The DB-PSGD and PBB methods are convenient for tomographic problems due to effective enforcement of non-negativity.

This paper is organized as follows. In Section 2 we briefly discuss the mathematical model of X-ray tomography and describe the computational method proposed for solving the minimization problem (1.2). Also, the method and equipment used in measuring the X-ray data are described. Section 3 is devoted to numerical results obtained by applying the proposed methods to 2D tomography with both simulated and real X-ray data. Finally, Section 4 presents conclusions.

2 Methods and materials

The main challenges in finding the minimizer (1.2) numerically are

(i) large-scale of the problem,

(ii) non-differentiability of the objective functional, and

(iii) the non-negativity constraint.

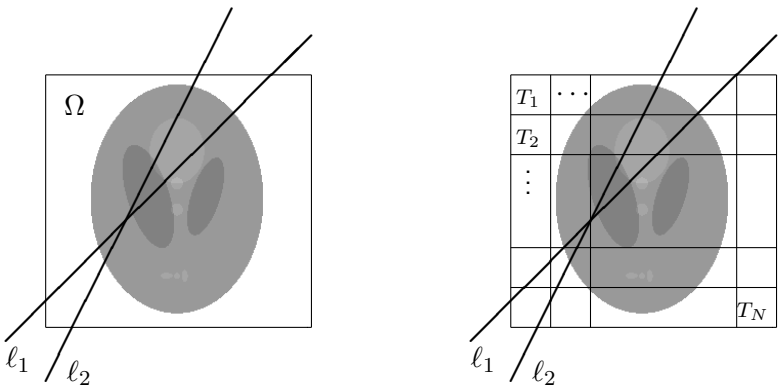

Figure 1: *Left*: A target, i.e. a function $f : \Omega \to \mathbb{R}$, to be reconstructed from a set of line integrals over lines ℓ_i. *Right*: Discretization into N pixels $T_1, T_2, \ldots, T_N$.

In this section we introduce a *discontinuity-based projected subgradient descent* (DB-PSGD) method to handle the above challenges. The idea is to employ the advantage of discontinuous functions, which are included in the space of functions of bounded variation. In addition, we briefly describe a recently introduced method based on smoothing out the non-differentiability of the objective functional and then applying a minimization method known as projected Barzilai-Borwein. These methods are compared by applying them to the tomographic data described in the end of this section. We begin this section by explaining how X-ray tomography can be modeled with a linear equation $\mathcal{A}f = g$ and how it can be approximated by a corresponding matrix equation $A\mathbf{f} = \mathbf{g}$.

2.1 Mathematical model for X-ray tomography

Mathematically the problem of 2D X-ray tomography can be expressed as follows: given the line integrals

$$g = g(\ell) = (\mathcal{A}f)(\ell) = \int_\ell f(x)dx$$

for a set of lines ℓ in the plane $\mathbb{R}^2$, determine the function $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ (if possible). Radon showed in 1917 that if $g(\ell)$ is known for all lines ℓ, then f is uniquely determined, see e.g. (Natterer and Wübbeling, 2001) for details.

In practical computations we need a discretized version $A\mathbf{f} = \mathbf{g}$ for the linear equation $\mathcal{A}f = g$. Thus, we assume here a square shaped domain $\Omega \subset \mathbb{R}^2$ and discretize it into $N = n \cdot n$ pixels such that f is modelled by a vector $\mathbf{f} = [\mathrm{f}_1, \mathrm{f}_2, \ldots, \mathrm{f}_N]^T \in \mathbb{R}^N$ whose jth component f_j approximates the value of f in jth pixel. See Figure 1 for an illustration. Then the approximation for the line integral over the line ℓ_i, $i = 1, 2, \ldots, M$, is given by

$$\mathrm{g}_i = \sum_{j=1}^{N} a_{i,j}\mathrm{f}_j, \tag{2.1}$$

where $a_{i,j}$ is the distance that the line ℓ_i (or X-ray) travels in the jth pixel. Now, interpreting $a_{i,j}$ to be the elements of the matrix $A = (a_{i,j})$ and $\mathbf{g} = [\mathrm{g}_1, \mathrm{g}_2, \ldots, \mathrm{g}_M]^T \in \mathbb{R}^M$ we have a matrix equation model for 2D X-ray tomography. The methods introduced in the following two sections assume that we have computational routines for computing $A\mathbf{z}$ and $A^T\mathbf{w}$ for any vectors $\mathbf{z} \in \mathbb{R}^N$ and $\mathbf{w} \in \mathbb{R}^M$.

We also introduce here the discrete differential operators by two square matrices $D_1, D_2 \in \mathbb{R}^{N\times N}$ for each spatial direction and the discrete gradient $D\mathbf{f}$, by this we have

$$D\mathbf{f} = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} \mathbf{f} = \begin{pmatrix} D_1\mathbf{f} \\ D_2\mathbf{f} \end{pmatrix} \quad \text{and} \quad \operatorname{div} \begin{pmatrix} \mathbf{f} \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} -D_1^T & -D_2^T \end{pmatrix} \begin{pmatrix} \mathbf{f} \\ \mathbf{f} \end{pmatrix} = -D_1^T\mathbf{f} - D_2^T\mathbf{f}.$$

Given the total number of pixels $N = n \cdot n$ with the number of rows and columns n and assume the length of each pixel as $h = 1$, the derivative matrices simply write as

$$(D_1\mathbf{f})_j = \mathrm{f}_{j+n} - \mathrm{f}_j \quad \text{and} \quad (D_2\mathbf{f})_j = \mathrm{f}_{j+1} - \mathrm{f}_j, \tag{2.2}$$

for non-boundary terms, i.e. enforcing Neumann boundary condition gives

$$\begin{aligned} (D_1\mathbf{f})_j &= 0 \text{ if } j \in \{1, 2, \ldots, n, N-n+1, N-n+2, \ldots, N\} \\ (D_2\mathbf{f})_j &= 0 \text{ if } j \in \{1, n, n+1, 2n, \ldots, (n-1)n+1, N\}, \end{aligned}$$

which should in practice be implemented without using any explicit matrices.

2.2 Functions of bounded variation

The aim of our algorithm is to compute a piecewise continuous function by minimizing (1.1) involving its total variation, in fact our implementation will only compute piecewise constant functions. Therefore we introduce the space of functions of bounded variation in order to have a clear definition of the total variation. We will introduce some basic notations and concepts, for a thorough analysis the reader is advised to consult (Ambrosio, Fusco and Pallara, 2000; Giusti, 1984).

A function $f \in L^1(\Omega)$ is called a function of bounded variation on $\Omega \subset \mathbb{R}^2$, if all of its first order partial derivatives (in the distributional sense) are measures with finite total variations in Ω. The derivative of such a function f, is a bounded vector-valued measure, with the finite total variation

$$TV(f) := \sup \left\{ \int_\Omega -f \operatorname{div}\varphi \, \mathrm{d}x \ : \ \varphi \in [C_0^1(\Omega)]^2, \ \|\varphi\|_\infty \leq 1 \right\}. \tag{2.3}$$

The space of functions of bounded variation in Ω is denoted by $BV(\Omega)$, it can be made a Banach space endowed with the norm

$$\|f\|_{BV} = \|f\|_{L^1} + TV(f). \tag{2.4}$$

The main advantage of the space BV is that it includes piecewise smooth functions and their well defined derivatives in the distributional sense.
An often used simplified representation of $TV(f)$ arises if one considers only functions in the Sobolev space $W^{1,1}(\Omega)$, then the total variation simplifies to

$$TV(f) = \||\nabla f|\|_{L^1} = \int_\Omega |\nabla f| \, \mathrm{d}x = \int_\Omega \sqrt{(\partial_1 f)^2 + (\partial_2 f)^2},$$

where ∇f denotes the weak gradient of f.
We want to analyse the behaviour of functions with jump discontinuities, therefore it is important to give a definition for the identification of jump points, that means points where f changes the value by a jump discontinuity, see (Ambrosio et al., 2000) for the following definition. At first we divide

the ball with center x and radius $\rho > 0$, denoted by $B_\rho(x)$, into two halves defined by a direction on the unit sphere $\nu \in S^1$

$$\begin{cases} B_\rho^+(x,\nu) := \{y \in B_\rho(x) : (y-x,\nu)_2 > 0\} \\ B_\rho^-(x,\nu) := \{y \in B_\rho(x) : (y-x,\nu)_2 < 0\} \end{cases}.$$

Definition 2.1. (Ambrosio et al., 2000, Def. 3.67) We say that $f \in BV(\Omega)$ has an approximate jump point at x if there exist $a, b \in \mathbb{R}$, $a \neq b$, and a direction $\nu \in S^1$ such that

$$\lim_{\rho\to 0} \frac{1}{|B_\rho^+(x,\nu)|} \int_{B_\rho^+(x,\nu)} |f(x) - a| \,\mathrm{d}x = 0, \quad \lim_{\rho\to 0} \frac{1}{|B_\rho^-(x,\nu)|} \int_{B_\rho^-(x,\nu)} |f(x) - b| \,\mathrm{d}x = 0. \tag{2.5}$$

We will denote the triplet of the values a and b together with its jump direction ν as $(a, b, \nu) = (f^+(x), f^-(x), \nu_f(x))$. The set of approximate jump points is denoted by J_f.

The notation of f^+ and f^- dependent on a direction ν will be of further importance for the proposed method.

2.3 A discontinuity-based projected subgradient descent method

We use the following model for the non-discretized (continuum) tomographic measurement: $\mathcal{A}f = g$ with $\mathcal{A} : L^2(\Omega) \to L^2(C)$ a compact linear operator, and Ω, C are suitable bounded sets, e.g. $\Omega, C \subset \mathbb{R}^2$. We note that $BV(\Omega)$ is continuously embedded to $L^2(\Omega)$ for $\Omega \subset \mathbb{R}^2$, see (Ambrosio et al., 2000).

The here proposed method is a discontinuity-based projected subgradient descent scheme, similar to a projected gradient descent algorithm, but without dependence on a smoothing of the total variation or continuity assumptions on the target function.

The idea is based on a space of discontinuous functions on the mesh, or set of pixels, $\mathcal{T}_h := \{T_j : 1 \leq j \leq N\}$ as in Figure 1, defined by

$$S_d^r(\mathcal{T}_h) := \{\varphi \in L^1(\Omega) : \varphi|_T \text{ is a polynomial of degree } r \text{ for each } T \in \mathcal{T}_h\},$$

which is easily seen to be a subspace of $BV(\Omega)$.

For the implementation of X-ray tomography we set $\Omega = [0, n] \times [0, n] \subset \mathbb{R}^2$ and equip the Hilbert space $L^2(\Omega)$ with the L^2-inner product, denoted by $(\cdot,\cdot)_{L^2(\Omega)}$. If we allow the approximating functions f to have discontinuities, that is we assume that $f \in S_d^r(\mathcal{T}_h)$ for $r \geq 0$, some modifications on the mesh need to be specified first. Since the functions may have jumps at the boundary of each T we need to separate those. A natural approach for image processing is to consider a uniform mesh as in Figure 1 that is produced by translation of the half open interval $T_0 = (0, 1]^2$ for each $T \in \mathcal{T}_h$, such that the intersection of two cells is empty.

By this preparation we obtain with integration by parts for $f \in S_d^r(\mathcal{T}_h)$ and $\varphi \in [C_0^1(\Omega)]^2$,

$$-\int_\Omega f \mathrm{div}\varphi \,\mathrm{d}x = -\sum_{T\in\mathcal{T}_h} \int_T f \mathrm{div}\varphi \,\mathrm{d}x = \sum_{T\in\mathcal{T}_h} \left(-\int_{\partial T} \tilde{f}\varphi \cdot \nu_n \,\mathrm{d}s + \int_T (\nabla f, \varphi) \,\mathrm{d}x \right),$$

where $\tilde{f}$ denotes the the C^1 extension of $f|_T$ on the closure $cl(T)$ and is considered separately on each pixel $T \in \mathcal{T}_h$. Introducing the union of the cell boundaries

$$J_h := \bigcup_{T\in\mathcal{T}_h} \partial T$$

and taking the supremum over all $\varphi \in [C_0^1(\Omega)]^2$ with $\|\varphi\|_\infty \leq 1$ we give the total variation on $S_d^r(\mathcal{T}_h)$ as

$$\begin{aligned} TV_d(f) :=& \int_{J_h} |f^+ - f^-| \, \mathrm{d}s + \sum_{T \in \mathcal{T}_h} \int_T |\nabla f| \, \mathrm{d}x \\ :=& \|[f]\|_{L^1(J_h)} + \||\nabla f|\|_{L^1(\Omega)}, \end{aligned} \tag{2.6}$$

with $[f] := f^+ - f^-$ and ∇f is defined element-wise on each $T \in \mathcal{T}_h$.

Remark 2.1. For $f_h \in S_d^r(\mathcal{T}_h)$ we may interpret $TV_d(f_h)$, depending on the mesh $\mathcal{T}_h$, as an approximation of the total variation $TV(f)$ for functions $f \in SBV(\Omega)$, the space of special functions of bounded variation which can be decomposed to their absolute continuous and their jump part, see (Ambrosio et al., 2000). Nevertheless, we can not assure strict convergence of f_h to f in $SBV(\Omega)$, see (Bartels, 2012).

The Euler-Lagrange equation, assuming homogeneous Neumann boundary conditions on $\partial\Omega$, and using a subgradient representation for (1.1) is given by

$$2\mathcal{A}^*(\mathcal{A}f - g) + \alpha p_{TV} = 0 \qquad \text{with } p_{TV} \in \partial TV(f).$$

The subdifferential of the total variation ∂TV with respect to $L^2(\Omega)$ is defined at f by

$$p_{TV} \in \partial TV(f) \Leftrightarrow TV(f) + (p_{TV}, h - f)_{L^2(\Omega)} \leq TV(h) \quad \forall h \in L^2(\Omega).$$

$p_{TV} \in \partial TV(f)$ is called a subgradient of TV in f and coincides with the classical gradient where TV is differentiable, see for instance (Chambolle, 2004; Deimling, 1985). We refer for further discussion and characterization of the subdifferential $\partial TV(f)$ to (Andreu-Vaillo, Caselles and Mazón, 2004). A minima of (1.1) can be obtained by artificial time evolution of the steepest descent equation

$$\begin{cases} f_t = -p_{TV} - 2\mathcal{A}^*(\mathcal{A}f - g) & \text{in } \Omega \times (0,T] \\ \partial_n f = 0 & \text{on } \partial\Omega \times (0,T] \\ f(\cdot,0) = f_0 & \text{in } \Omega \end{cases} \tag{2.7}$$

for a fixed time $T > 0$ and an initial condition f_0. The above gradient flow is also referred as the total variation flow (TV flow) and is analysed for instance in (Andreu, Ballester, Caselles and Mazón, 2001). Many authors constructed algorithms based on the TV flow, including the ROF model (Rudin et al., 1992) and based on that (Li and Santosa, 1996), or such as finite element approximations of the TV flow in (Feng and Prohl, 2003).

To state an Euler-Lagrange like equation, i.e. a minimal condition, for the objective functional involving $TV_d(f)$,

$$\mathcal{L}_{TV_d}(f) := \left\|\mathcal{A}f - g\right\|_{L^2}^2 + \alpha TV_d(f), \tag{2.8}$$

we need to separate the terms acting on Ω and J_h, by which we now have to solve the modified Euler-Lagrange equation for $f \in S_d^r(\Omega)$, given by

$$\begin{cases} 2\mathcal{A}^*(\mathcal{A}f - g) + \alpha p = 0 & \text{in } \Omega \\ \alpha(f^+ - f^-) = 0 & \text{on } J_h \end{cases}, \tag{2.9}$$

with a subgradient $p \in \partial\||\nabla f|\|_{L^1(\Omega)}$.

2.3.1 Discretization in $S_d^0(\Omega)$

As discussed in Section 2.1 the model for X-ray tomography assumes a constant function on each pixel, this suggests to utilize $S_d^0(\Omega)$ for the discretization scheme to obtain a consistency in fidelity and regularization term. The space $S_d^0(\Omega)$ contains only piecewise constants and hence meets the requirement. We will now briefly present a resulting algorithm and discuss its properties, which turned out to be insufficient for our purpose and, hence, a motivation for the proposed DB-PSGD algorithm.

The gradient of a piecewise constant vanishes in every cell and hence the total variation for $f \in S_d^0(\Omega)$ contains only the jump part of the function f, i.e.

$$TV_d(f) = \int_{J_h} |f^+ - f^-| \, \mathrm{d}s, \quad \text{for } f \in S_d^0(\Omega). \tag{2.10}$$

Thus, (2.10) simply sums the jumps across pixels, this can be compared to the introduced discretization in (Bartels, 2012). To cut it short, minimizing (1.1) with the representation of $TV_d(f)$ in a discrete setting results in the descent direction

$$\Delta_0 \mathbf{f} = 2A^T(A\mathbf{f} - \mathbf{g}) + \alpha \left(\operatorname{sign}(D_1\mathbf{f}) + \operatorname{sign}(D_2\mathbf{f}) + \operatorname{sign}(D_1^T\mathbf{f}) + \operatorname{sign}(D_2^T\mathbf{f}) \right) \tag{2.11}$$

for an iterative algorithm

$$\mathbf{f}^{k+1} = \mathbf{f}^k - \lambda \Delta_0 \mathbf{f}^k \quad \text{for } k = 0, 1, \ldots,$$

with an initial iterate $\mathbf{f}^0 \in \mathbb{R}^N$, step size λ, and $\operatorname{sign} : \mathbb{R}^N \to \mathbb{R}^N$ denotes the component-wise sign function. It is important to note that the descent direction $\Delta_0 \mathbf{f}$ only minimizes the anisotropic total variation.

A computed reconstruction using this iterative scheme is shown in Section 3, Figure 4. The results obtained are reasonable but not satisfying in comparison to established methods, i.e. a blocky structure and typical staircasing effects along round edges, this is partly due to the anisotropic total variation. The following approach proposes to combine the idea from the pure $S_d^0(\Omega)$ discretization with a subgradient descent scheme for the total variation starting from the modified Euler-Lagrange equation (2.9).

2.3.2 The discretization scheme for the proposed algorithm

Due to not adequate results using the $S_d^0(\Omega)$ approach we propose the following *discontinuity-based projected subgradient descent* method, which combines the advantage of utilizing the jump part and the subgradient representation. We start by writing (2.9) in a variational form with $f, \varphi \in S_d^r(\mathcal{T}_h)$

$$\mathcal{L}_{TV_d}(f)(\varphi) := 2(\mathcal{A}^*(\mathcal{A}f - g), \varphi)_{L^2(\Omega)} + \alpha(p, \varphi)_{L^2(\Omega)} + \alpha([f], \varphi)_{L^2(J_h)}. \tag{2.12}$$

The condition for a minimizer $f \in S_d^r(\mathcal{T}_h)$ can be stated as

$$\mathcal{L}_{TV_d}(f)(\varphi) = 0, \quad \forall \varphi \in S_d^r(\mathcal{T}_h). \tag{2.13}$$

The main point is to consider each square T_j separately by taking φ_j such that $\operatorname{supp}(\varphi_j) = T_j$, i.e. the test functions are only supported on a specific square. Since (2.13) holds for all $\varphi \in S_d^r(\mathcal{T}_h)$ the condition is still satisfied for every φ_j for $1 \leq j \leq N$.

Thus, we have almost the same equation as (2.12) but on separate squares

$$\mathcal{L}_{TV_d}(f)(\varphi_j) = 2(\mathcal{A}^*(\mathcal{A}f - g), \varphi_j)_{L^2(T_j)} + \alpha(p, \varphi_j)_{L^2(T_j)} + \alpha([f], \varphi_j)_{L^2(\partial T_j)}. \tag{2.14}$$

From this point we switch to the discrete matrix-vector equations. We use only constant functions on each square, that means the test functions φ_j are represented as a vector indicating each square T_j by the unit vector $\mathbf{e_j} \in \mathbb{R}^N$. This allows a fast and simple representation of the resulting equations. Thus, (2.14) writes in the discrete case as

$$L(\mathbf{f})_{TV_d}(\mathbf{e_j}) = 2(A^T(A\mathbf{f} - \mathbf{g}), \mathbf{e_j})_{l^2(T_j)} + \alpha(\mathbf{p}, \mathbf{e_j})_{l^2(T_j)} + \alpha([\mathbf{f}], \mathbf{e_j})_{l^2(\partial T_j)}, \tag{2.15}$$

which will be the descent direction for the algorithm, compare to (2.9) and (2.7). In the following we derive an explicit representation for the computations.

We denote the indices of the neighbourhood of a square T_j by

$$\mathcal{N}(T_j) := \{i : T_i \text{ has an edge togther with } T_j,\ 1 \leq i \leq N\}.$$

The inner product along the boundary ∂T_j with an edge to a neighbouring square is given by

$$(\mathbf{f}^+ - \mathbf{f}^-, \mathbf{e_j})_{l^2(\partial T_j)} = \sum_{i \in \mathcal{N}(T_j)} (\mathrm{f}_i - \mathrm{f}_j) = -4\mathrm{f}_j + \sum_{i \in \mathcal{N}(T_j)} \mathrm{f}_i, \tag{2.16}$$

which can be represented by using the derivative matrices as

$$(D_1 + D_2)\mathbf{f} + (D_1^T + D_2^T)\mathbf{f}.$$

Similarly for the inner products on T_j, we obtain

$$(2A^T(A\mathbf{f} - \mathbf{g}) + \alpha\mathbf{p}, \mathbf{e_j})_{l^2(T_j)} = (2A^T(A\mathbf{f} - \mathbf{g}) + \alpha\mathbf{p})_j,$$

where the finite dimensional subgradient $\mathbf{p} \in \mathbb{R}^N$ is given component-wise by

$$\mathbf{p} = \begin{cases} -\mathrm{div}\left(\dfrac{D\mathbf{f}}{|D\mathbf{f}|}\right) & \text{where } |D\mathbf{f}| \neq 0, \\ 0 & \text{else,} \end{cases}$$

with $\left(\frac{D\mathbf{f}}{|D\mathbf{f}|}\right)_j := \frac{(D\mathbf{f})_j}{|(D\mathbf{f})_j|}$, that is, the absolute values and divisions are taken element-wise.
This leads to the subgradient descent scheme

$$\mathbf{f}^{k+1} = \mathbf{f}^k - \lambda_k \Delta\mathbf{f}^k \qquad \text{for } k = 0, 1, \ldots,$$

for an initial iterate $\mathbf{f}^0 \in \mathbb{R}^N$. The step $\Delta\mathbf{f}$ is the explicit representation of (2.15) given by

$$\begin{aligned} \Delta\mathbf{f} = 2A^T(A\mathbf{f} - \mathbf{g}) - \alpha &\left((D_1^T \frac{D_1\mathbf{f}}{|D\mathbf{f}|}) + (D_2^T \frac{D_2\mathbf{f}}{|D\mathbf{f}|}) \right) \\ &+ \alpha \left((D_1 + D_2)\mathbf{f} + (D_1^T + D_2^T)\mathbf{f} \right). \end{aligned} \tag{2.17}$$

It is left to give a step size rule and a stopping condition, for this purpose we define the discrete functional

$$L(\mathbf{f}) := \|A\mathbf{f} - \mathbf{g}\|_2^2 + \alpha\|\tilde{D}\mathbf{f}\|_1 \tag{2.18}$$

with $\tilde{D}\mathbf{f}$ a discrete approximation of the Euclidean norm of the gradient of f with its jth component given by

$$(\tilde{D}\mathbf{f})_j = \sqrt{(\mathrm{f}_{j+n} - \mathrm{f}_j)^2 + (\mathrm{f}_{j+1} - \mathrm{f}_j)^2}. \tag{2.19}$$

By this the step size is calculated under the feasibility constraint

$$\lambda_k \in \mathrm{FR}_\lambda := \left\{\lambda : L(\mathbf{f}^{k+1}) < L(\mathbf{f}^k),\ \mathbf{f}^{k+1} = \mathbf{f}^k - \lambda \mathbf{f}^k\right\}$$

as

$$\lambda_k = \min\left\{\lambda_{\max},\ \max\left\{\lambda_{\min},\ \left\{\lambda \in \mathrm{FR}_\lambda :\ \lambda = \frac{\lambda_{k-1}}{2^j},\ j \geq -1\right\}\right\}\right\} \tag{2.20}$$

and since $j \geq -1$ it is possible that λ increases during the iteration.

DB-PSGD is an iterative algorithm and hence a stopping condition is needed. In our implementation the algorithm will stop when a maximum number of iterations $J_{\max}$ was performed, due to a better comparison with the PBB method. If one wants to choose an automatic stopping condition, a suitable choice would be to require that the condition

$$L(\mathbf{f}^{k-1}) - L(\mathbf{f}^k) < \varepsilon L(\mathbf{f}^0) \tag{2.21}$$

is satisfied for a preassigned small number $\varepsilon > 0$.

To extend the method for solving the constrained minimization problem (1.2), we apply a projection strategy. More precisely, having computed the iterate f^k, we project it to the feasible region $\mathrm{FR} := \{\mathbf{z} \in \mathbb{R}^N \,|\, \mathbf{z} \geq 0\}$ using the projection operator $P : \mathbb{R}^N \to \mathrm{FR}$ defined component-wise by

$$\big(P(\mathbf{z})\big)_j := \begin{cases} \mathrm{z}_j & \text{if } \mathrm{z}_j \geq 0 \\ 0 & \text{if } \mathrm{z}_j < 0 \end{cases} \tag{2.22}$$

for $j = 1, \ldots, N$ and $\mathbf{z} = (\mathrm{z}_1, ..., \mathrm{z}_N) \in \mathbb{R}^N$.

At last we can formulate Algorithm 1 the discontinuity-based projected subgradient descent implementation for 2D X-ray tomography, which needs only vectors of size N to compute the reconstruction and no solving of any nonlinear equation.

Algorithm 1 DB-PSGD implementation for 2D X-ray tomography

```
Set k = 0 and choose λ_0, f^k, λ_min
while k ≤ J_max do
    Compute Δf^k by (2.17)
    f^{k+1} = P(f^k − λ_k Δf^k)
    if λ_k ∈ FR_λ then
        λ_{k+1} = 2λ_k
    else
        Determine λ_k by (2.20)
        f^{k+1} = P(f^k − λ_k Δf^k)
    end if
    k = k + 1
end while
```

2.4 Projected Barzilai-Borwein

The Barzilai-Borwein method (Barzilai and Borwein, 1988) is especially suitable for large-scale optimization problems due to its relatively inexpensive matrix-free computations and low memory requirements. It has proven to outperform the classical steepest descent method and be competitive also with conjugate-gradient methods in many optimization problems. In addition, the method is simple to implement.

To solve the minimization problem (1.2) numerically using Barzilai-Borwein, we discretize the problem similarly to (2.18) as

$$\min_{\mathbf{f}\geq 0} L_\beta(\mathbf{f}) := \|A\mathbf{f}-\mathbf{g}\|_2^2 + \alpha\|\tilde{D}\mathbf{f}\|_{1,\beta}, \tag{2.23}$$

where A, $\mathbf{f}$ and $\mathbf{g}$ are as described in section 2.1, $\tilde{D}\mathbf{f}$ is a discrete approximation of the Euclidean norm of the gradient of f defined in (2.19), $\beta > 0$ is a small parameter, and $\|\cdot\|_{1,\beta}$ denotes an approximation of the 1-norm with smoothed absolute value function given by

$$\|\mathbf{f}\|_{1,\beta} := \sum_{j=1}^{N} \sqrt{\mathrm{f}_j^2+\beta}, \quad \mathbf{f} = (\mathrm{f}_1,\ldots,\mathrm{f}_N) \in \mathbb{R}^N,$$

similarly to, for example, (Acar and Vogel, 1994; Vogel and Oman, 1996; Dobson and Vogel, 1997). On the boundary of the image Ω we assume Neumann boundary condition as explained in the previous subsection.

We need to compute the gradient

$$\nabla L_\beta(\mathbf{f}) = \nabla\left(\|A\mathbf{f}-\mathbf{g}\|_2^2\right) + \alpha\,\nabla\left(\|\tilde{D}\mathbf{f}\|_{1,\beta}\right) \tag{2.24}$$

in order to apply Barzilai-Borwein method to (2.23). For the first term we note that

$$\nabla\left(\|A\mathbf{f}-\mathbf{g}\|_2^2\right) = 2A^T(A\mathbf{f}-\mathbf{g}), \tag{2.25}$$

while the second term requires a simple differentiation of

$$\|\tilde{D}\mathbf{f}\|_{1,\beta} = \sum_{i=1}^{N} \sqrt{(\tilde{D}\mathbf{f})_i^2+\beta}$$

with respect to f_j, i.e.

$$\begin{aligned}\frac{\partial}{\partial \mathrm{f}_j}\|\tilde{D}\mathbf{f}\|_{1,\beta} =& \frac{2\mathrm{f}_j-\mathrm{f}_{j+n}-\mathrm{f}_{j+1}}{\sqrt{(\mathrm{f}_{j+n}-\mathrm{f}_j)^2+(\mathrm{f}_{j+1}-\mathrm{f}_j)^2+\beta}} \\ &+ \frac{\mathrm{f}_j-\mathrm{f}_{j-n}}{\sqrt{(\mathrm{f}_j-\mathrm{f}_{j-n})^2+(\mathrm{f}_{j-n+1}-\mathrm{f}_{j-n})^2+\beta}} \\ &+ \frac{\mathrm{f}_j-\mathrm{f}_{j-1}}{\sqrt{(\mathrm{f}_j-\mathrm{f}_{j-1})^2+(\mathrm{f}_{j+n-1}-\mathrm{f}_{j-1})^2+\beta}}\end{aligned} \tag{2.26}$$

for non-boundary components f_j (and zero otherwise).

The Barzilai-Borwein (BB) is a gradient based method given by the iteration

$$\mathbf{f}^{k+1} = \mathbf{f}^k - \lambda_k \nabla L_\beta(\mathbf{f}^k)$$

with a special choice of the step size

$$\lambda_k = \frac{(\mathbf{f}^k-\mathbf{f}^{k-1})^T(\mathbf{f}^k-\mathbf{f}^{k-1})}{(\mathbf{f}^k-\mathbf{f}^{k-1})^T(\nabla L_\beta(\mathbf{f}^k)-\nabla L_\beta(\mathbf{f}^{k-1}))}. \tag{2.27}$$

This choice of step size is based on a two-point approximation for the secant equation underlying quasi-Newton methods. We remark that (2.27) is not only easier to compute than the line-search-based classical steepest descent step size but it has also proven to yield significantly better convergence speed than the classical steepest descent method, see e.g. (Fletcher, 2005). An interesting property of the BB choice of step size is that it leads to a non-monotone method, i.e. the objective function value is not guaranteed to decrease at every iteration step. The BB method has been proven to be globally convergent for the case of strictly convex quadratic objective functions (Raydan, 1993). However, for general objective functionals some type of globalization strategy may be needed to guarantee convergence, see (Raydan, 1997).

Despite the fact that our objective functional in (2.23) is not quadratic, we have chosen to use no globalization strategy in this work. Our numerical experience on several different X-ray data sets with different measurement settings suggests that such a strategy is not necessary in the present problem of TV regularization for X-ray tomography. However, we emphasize that if a guarantee for convergence is desired then a globalization similar to that for example in (Raydan, 1997) might be necessary.

In order to solve the *constrained* optimization problem (2.23) using BB we apply a strategy of projected gradient methods similarly as explained in the previous subsection. More precisely, having computed the iterate f^k with the BB method, we project it to the feasible region $\mathrm{FR} := \{\mathbf{z} \in \mathbb{R}^N \,|\, \mathbf{z} \geq 0\}$ by the projection operator $P : \mathbb{R}^N \to \mathrm{FR}$ defined in (2.22). With this modification the BB method is known as the projected Barzilai-Borwein (PBB) method. The PBB method is considered in detail for example in (Dai and Fletcher, 2005). Since PBB is an iterative method we need to specify a stopping condition. A useful condition could be for example a gradient condition like

$$\|\nabla_{\mathrm{test}} L_\beta(\mathbf{f}^k)\|_2 \leq \varepsilon \|\nabla L_\beta(\mathbf{f}^0)\|_2, \tag{2.28}$$

where $\varepsilon > 0$ is a preassigned number and

$$\left(\nabla_{\mathrm{test}} L_\beta(\mathbf{f}^k)\right)_j := \begin{cases} \left(\nabla L_\beta(\mathbf{f}^k)\right)_j & \text{if } (\mathbf{f}^k)_j > 0 \\ \min\{\left(\nabla L_\beta(\mathbf{f}^k)\right)_j, 0\} & \text{if } (\mathbf{f}^k)_j = 0 \end{cases} .$$

In this study, however, we simply stop the iteration after some preassigned maximum number $J_{\max}$ of iterations. Let us finally summarize the resulting algorithm for solving (2.23) with PBB, see Algorithm 2.

2.5 Tomographic data

To illustrate the computational methods described in the previous subsections, we test them in a setting of 2D X-ray tomography. This is done by cases of both simulated and real data; these data sets are described in the following two subsections.

2.5.1 Simulated data

Our test case with simulated data employs the widely-used Shepp-Logan phantom as an image to be reconstructed from 20 (parallel-beam) projection images. To avoid the most obvious "inverse crime" making the inversion "too easy", we create the projection data first on a slightly finer mesh

Algorithm 2 PBB algorithm for solving (2.23)

Set $k = 0$, and choose $\mathbf{f}^0 = 0$, $\lambda_0 = 10^{-5}$, $\beta > 0$
Compute $\nabla L_\beta(\mathbf{f}^0)$ using (2.24), (2.25) and (2.26)
while $k \leq J_{\max}$ **do**
  $\mathbf{f}^{k+1} = P(\mathbf{f}^k - \lambda_k \nabla L_\beta(\mathbf{f}^k))$
  Compute $\nabla L_\beta(\mathbf{f}^{k+1})$ using (2.24), (2.25) and (2.26)
  if $k < J_{\max}$ **then**
    $$\lambda_{k+1} = \frac{(\mathbf{f}^{k+1} - \mathbf{f}^k)^T(\mathbf{f}^{k+1} - \mathbf{f}^k)}{(\mathbf{f}^{k+1} - \mathbf{f}^k)^T(\nabla L_\beta(\mathbf{f}^{k+1}) - \nabla L_\beta(\mathbf{f}^k))}$$
  end if
  Set $k = k + 1$
end while

and then interpolate the data to the final resolution with which the reconstructions are computed. In addition, to simulate random measurement errors, we add 2% Gaussian random noise to the data.

2.5.2 Data measured from a walnut

The X-ray microtomography measurements of the walnut were performed with the custom-built μCT device nanotom supplied by Phoenix|Xray Systems + Services GmbH (Wunstorf, Germany). The sample was mounted on an acrylic rod (diameter 10 mm) with beeswax. The X-ray detector used was a CMOS flat panel detector with 2304 $\times$ 2284 pixels of 50 μm size (Hamamatsu Photonics, Japan). A set of 1200 projection images were acquired over a full 360 degree rotation with an angular step of 0.3 degrees between projections. Each projection image was composed of an average of six 750 ms exposures. The X-ray tube acceleration voltage was 80 kV and tube current 200 μA.

For the purposes of this work only the projection images corresponding to the central cross-section of the walnut are of interest; these form a set of (fan-beam) projection images of the 2D cross-section. From these data, we choose (sub)sets of 1200, 120, 60 and 30 projection images with angular steps of 0.3, 3, 6 and 12 degrees, respectively.

3 Numerical results

In this section we present a demonstration of the DB-PSGD and PBB methods applied to the 2D X-ray tomography test cases described previously. In the numerical computations we run the methods for some preassigned number of iterations, i.e. no specific stopping condition is employed. In the last chapter we only mentioned appropriate choices of stopping conditions for convenience of the reader. The idea is to illustrate the convergence speeds of the methods by the example with simulated data, while the real data example only demonstrates how the methods work with real and sparse X-ray data.

In subsection 3.1 the simulated test case is considered, while in subsection 3.2 we present results computed from real X-ray data of the walnut. The computations were performed with a machine equipped with 2.2 GHz Intel Core i7 CPU and 8 GB memory.

3.1 Simulated data

The PBB and DB-PSGD reconstructions after 200, 50 and 10 iterations and the original image are shown in Figure 2. A demonstration of convergence speeds of the algorithms can be found in Figure 3. The reported relative L^2 errors of the reconstructions are computed as

$$\frac{\|\text{original image} - \text{reconstruction}\|_2}{\|\text{original image}\|_2}.$$

The regularization parameter $\alpha > 0$ was chosen such that the relative L^2 errors were (approximately) minimized.

A comparison of DB-PSGD with a reconstruction using the pure $S_d^0(\Omega)$ discretization is shown in Figure 4. The reconstruction is computed by substituting the descent direction in step 3 of Algorithm 1 by $\Delta_0 \mathbf{f}$ given in (2.11). That means, both results are computed with the same algorithm by only changing the descent direction.

3.2 Real X-ray data of walnut

Let us then turn to an illustration how the methods work with real X-ray data of the walnut. The reconstructions from 120, 60 and 30 projection images with two different regularization parameters α are shown in Figures 5 – 10, respectively. In addition to the PBB and DB-PSGD reconstructions we present a filtered back-projection (FBP) reconstruction of the same target but computed from the full set of 1200 projection images. This reconstruction serves as a "ground truth" to which the PBB and DB-PSGD reconstructions may be compared.

The regularization parameter α was chosen by visual inspection such that there is a balance between two properties: the reconstructions should be as "blocky" as possible and as close to the FBP reconstruction as possible. The smoothing parameter for PBB is $\beta = 10^{-5}$ in all computations.

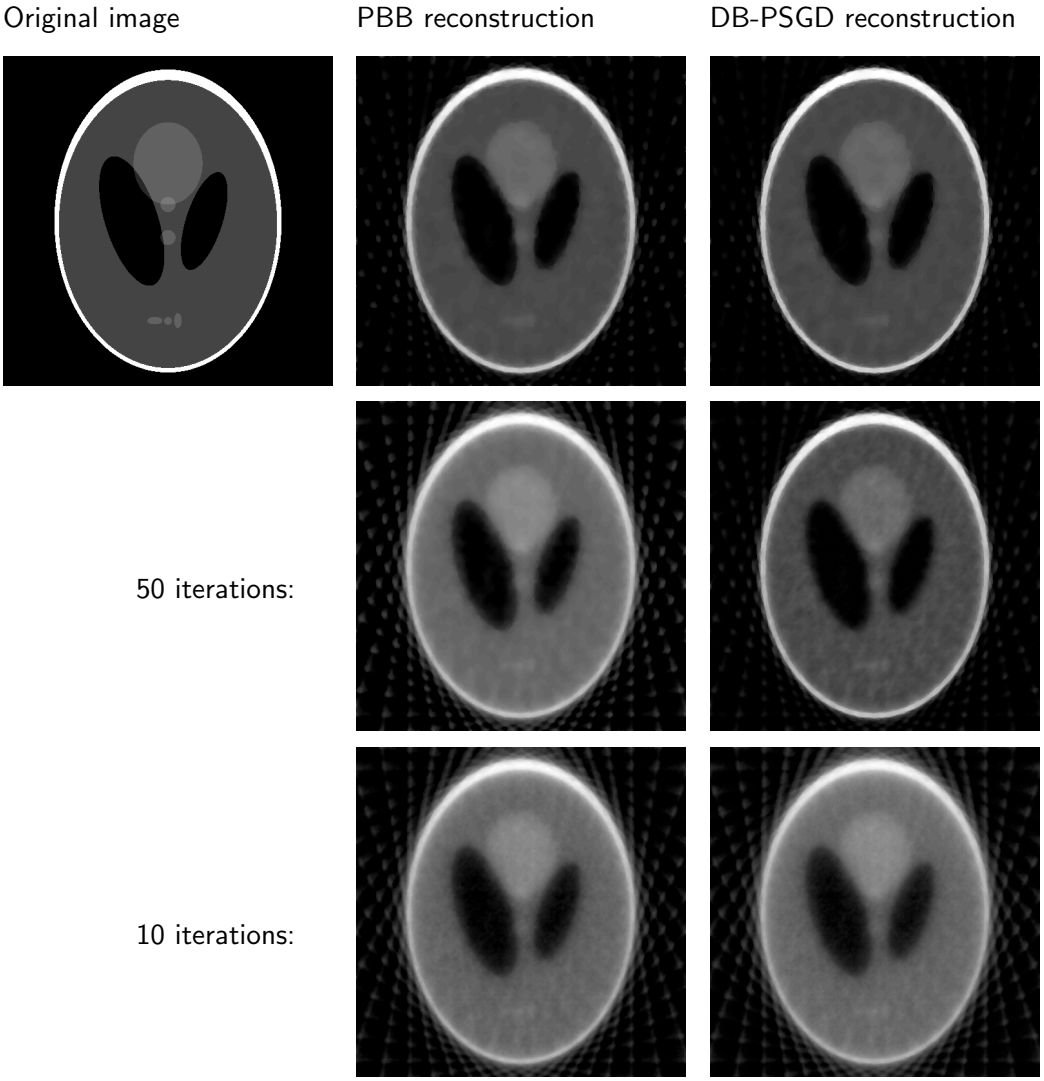

Figure 2: Reconstructions from simulated parallel beam data with 20 projection images. The regularization parameter is $\alpha = 100$, and the smoothing parameter for PBB $\beta = 10^{-5}$. Noiselevel in data is 2%. On the upper row, the number of iterations is 200 for both methods, and the computation time 70 seconds for PBB and 130 seconds for DB-PSGD. Relative L^2 error is 45.5% for PBB and 45.2% for DB-PSGD. For demonstration of convergence speed, also the reconstructions after 50 and 10 iterations are shown on two lower rows. The resolution of the images is 512x512.

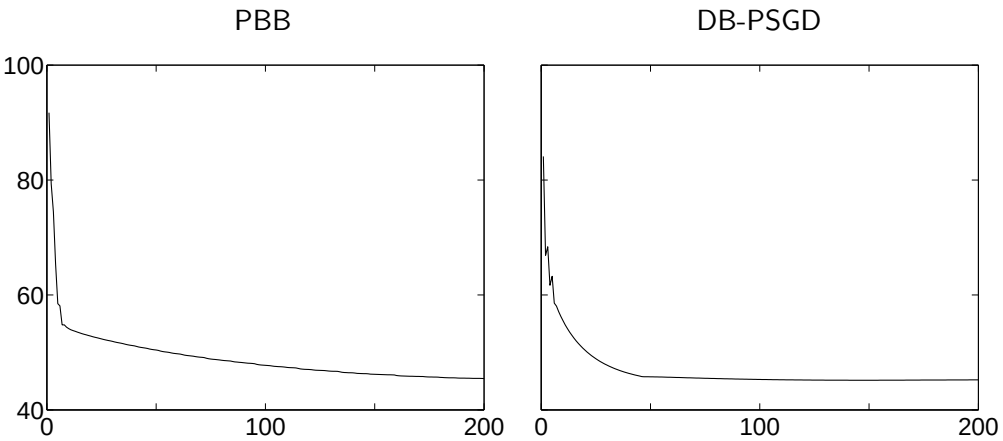

Figure 3: Relative L^2 errors of PBB and DB-PSGD reconstructions (in percents) as functions of iterations.

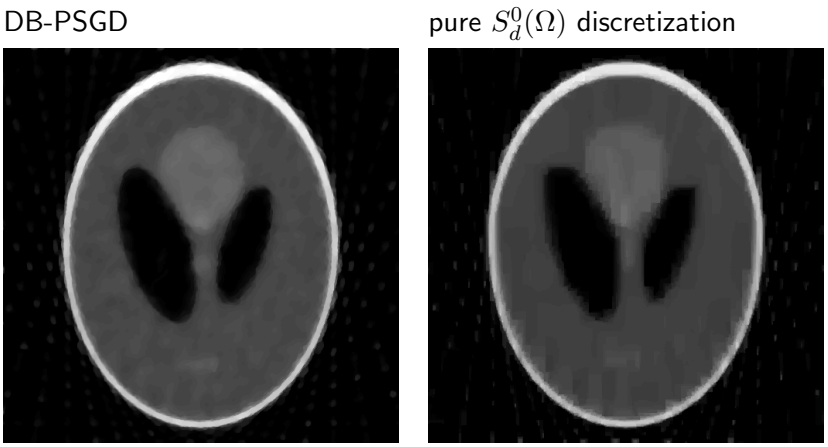

Figure 4: Comparison of DB-PSGD and pure $S_d^0(\Omega)$ discretization. The Reconstructions are from simulated parallel beam data with 20 projection images. The regularization parameter is $\alpha = 100$ and a noiselevel in data of 2%. The results after 200 iterations are shown. Relative L^2 error of DB-PSGD is 45.2% and 51.2% for pure $S_d^0(\Omega)$ discretization. Computation times are 130 sec. for DB-PSGD and 118 for the $S_d^0(\Omega)$ discretization. The resolution of the images is 512x512.

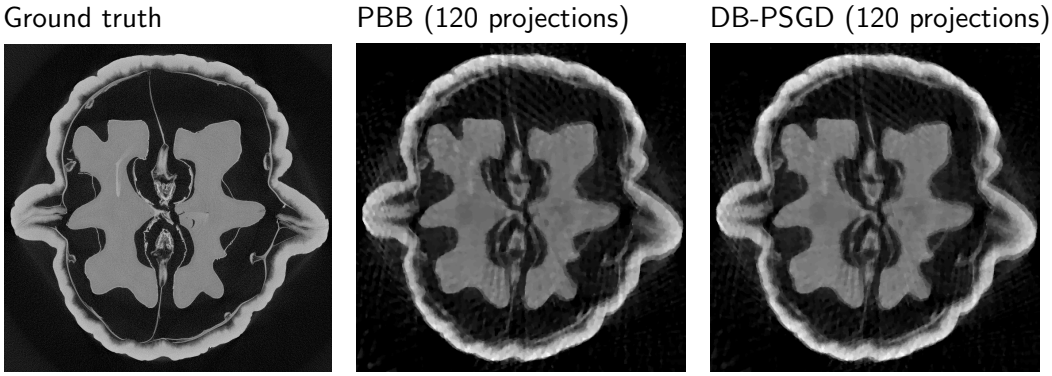

Figure 5: Reconstructions from 120 fan-beam projection images. The regularization parameter is $\alpha = 10^{-3}$. The number of iterations is 200 for both methods. Computation times of PBB and DB-PSGD were 740 and 1230 seconds, respectively. The resolution of the PBB and DB-PSGD reconstructions is 512x512.

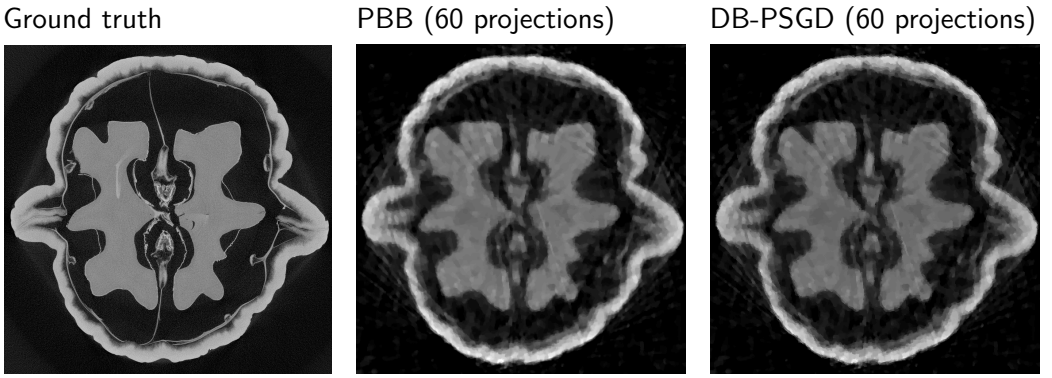

Figure 6: Reconstructions from 60 fan-beam projection images. The regularization parameter is $\alpha = 10^{-3}$. The number of iterations is 200 for both methods. Computation times of PBB and DB-PSGD were 360 and 600 seconds, respectively. The resolution of the PBB and DB-PSGD reconstructions is 512x512.

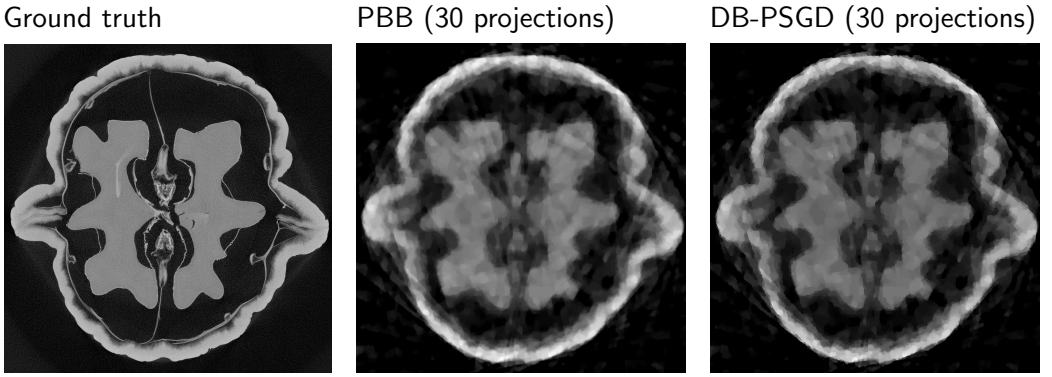

Figure 7: Reconstructions from 30 fan-beam projection images. The regularization parameter is $\alpha = 10^{-3}$. The number of iterations is 200 for both methods. Computation times of PBB and DB-PSGD were 180 and 300 seconds, respectively. The resolution of the PBB and DB-PSGD reconstructions is 512x512.

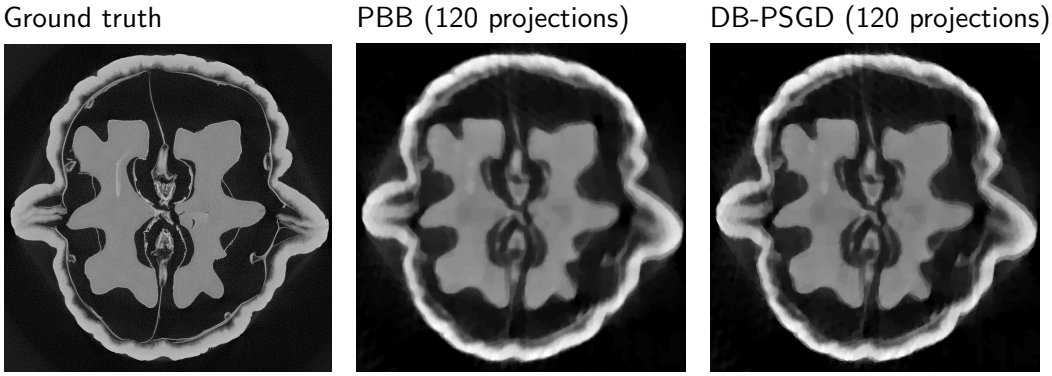

Figure 8: Reconstructions from 120 fan-beam projection images. The regularization parameter is $\alpha = 5 \cdot 10^{-3}$. The number of iterations is 200 for both methods. Computation times of PBB and DB-PSGD were 740 and 1230 seconds, respectively. The resolution of the PBB and DB-PSGD reconstructions is 512x512.

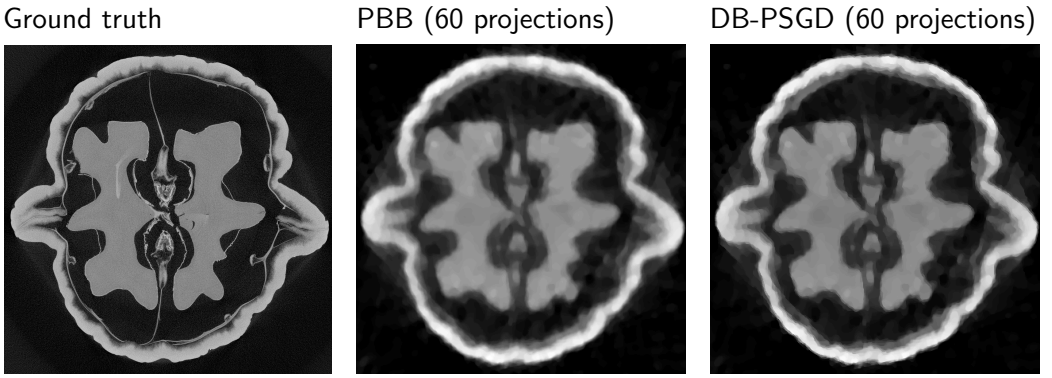

Figure 9: Reconstructions from 60 fan-beam projection images. The regularization parameter is $\alpha = 5 \cdot 10^{-3}$. The number of iterations is 200 for both methods. Computation times of PBB and DB-PSGD were 360 and 600 seconds, respectively. The resolution of the PBB and DB-PSGD reconstructions is 512x512.

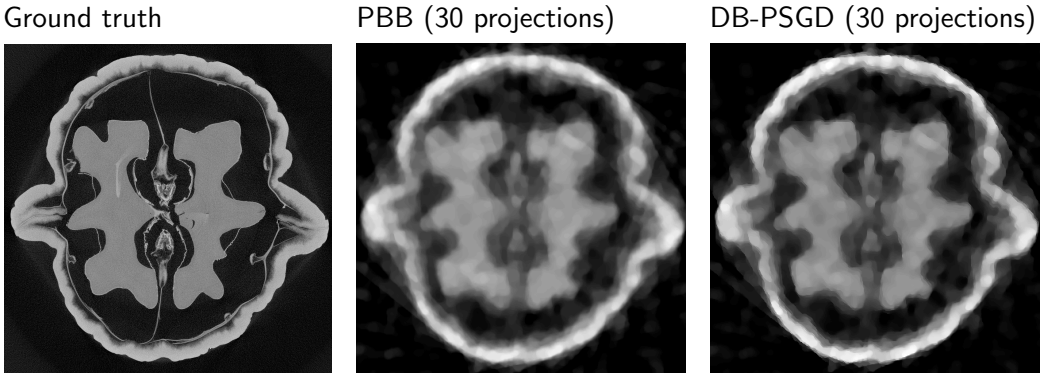

Figure 10: Reconstructions from 30 fan-beam projection images. The regularization parameter is $\alpha = 5 \cdot 10^{-3}$. The number of iterations is 200 for both methods. Computation times of PBB and DB-PSGD were 180 and 300 seconds, respectively. The resolution of the PBB and DB-PSGD reconstructions is 512x512.

4 Discussion

Tomographic imaging from sparsely collected projection data allows new low-dose medical imaging possibilities. However, the mathematical reconstruction problem is very ill-posed and calls for efficient regularization methods. Also, practical applications require fine discretization of the unknown attenuation coefficient, leading to computationally demanding large-scale problems.

We introduce a novel large-scale algorithm called *discontinuity-based projected subgradient descent* method (DB-PSGD) for computing total variation (TV) regularized reconstructions. The method is inspired by a discontinuous Galerkin discretization. We test DB-PSGD on realistically large-scale tomographic problems with both simulated and measured data.

How to choose a benchmark algorithm providing a comparison for the new method? The chosen algorithm should be applicable to large-scale problems, be relatively established, have computational complexity comparable to DB-PSGD, allow enforcement of non-negativity of the unknown attenuation coefficient, and work with a matrix-free implementation of the measurement operator and its adjoint. The very popular primal-dual approaches are not optimal for this as they require much more operator evaluations than DB-PSGD. We decided to choose the projected Barzilai-Borwein (PBB) optimization method applied to total variation penalty with smoothed-out absolute value function.

PBB and DB-PSGD turn out to have rather similar properties in terms of computational expense and image quality. Both methods are good choices for computing tomographic reconstructions from sparsely sampled X-ray projection images, reconstructing the main features and edges of the unknown even if the data sets are noisy and strongly limited. However, DB-PSGD has a couple of advantages over PBB. First, DB-PSGD requires no smoothing of the TV functional and thus yields reconstructions with sharper edges compared to algorithms with smoothed TV. This is illustrated for example in the walnut reconstructions in Figure 8 computed from real X-ray data. Second, DB-PSGD offers the possibility to focus on a region-of-interest by varying the pixel size spatially; this is an interesting topic for further research.

The choice of the regularization parameter α has a significant effect on the reconstructions with both methods. The smoothing parameter β for PBB plays a crucial role as well; too large β leads to reconstructions with smooth (blurred) edges while too small β causes numerical difficulties. Automatic choice of these parameters would be useful for practical applications, but creating such choice rules is a deep problem and outside the scope of this feasibility study. In conclusion, DB-PSGD is a promising new large-scale method for computing crisp TV regularized tomographic reconstructions.

Acknowledgment

This work was supported by the Academy of Finland (project 141094 and 1259526) and by the Finnish Centre of Excellence in Inverse Problems Research 2012-2017 (Academy of Finland CoE-project 250215). Lauri Harhanen was supported by the Finnish Doctoral Programme in Computational Sciences (FICS), the Finnish Cultural Foundation and the Magnus Ehrnrooth foundation.

References

Acar, R. and Vogel, C. R. 1994. Analysis of total variation penalty methods, *Inverse Problems* **10**: 1217–1229.

Ambrosio, L., Fusco, N. and Pallara, D. 2000. *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, Clarendon Press, Oxford.

Andreu, F., Ballester, C., Caselles, V. and Mazón, J. 2001. Minimizing total variation flow, *Differential and Integral Equations* **14**(3): 321–360.

Andreu-Vaillo, F., Caselles, V. and Mazón, J. 2004. *Parabolic quasilinear equations minimizing linear growth functionals*, Vol. 223 of *Progress in Mathematics*, Birkhäuser Verlag, Basel.

Bartels, S. 2012. Total variation minimization with finite elements: Convergence and iterative solution, *SIAM Journal on Numerical Analysis* **50**(3): 1162–1180.

Barzilai, J. and Borwein, J. 1988. Two-point step size gradient methods, *IMA Journal of Numerical Analysis* **8**(1): 141–148.

Bertalmio, M., Caselles, V., Rouge, B. and Solé, A. 2003. TV based image restoration with local constraints, *Journal of Scientific Computing* **19**(1-3): 95–122.

Bian, J., Han, X., Sidky, E. Y., Cao, G., Lu, J., Zhou, O. and Pan, X. 2010. Investigation of sparse data mouse imaging using micro-CT with a carbon-nanotube-based X-ray source, *Tsinghua Sci Technol* **15**(1): 74–78.

Cai, J.-F., Osher, S. and Shen, Z. 2009. Linearized Bregman iterations for frame-based image deblurring, *SIAM J. Imaging Sci.* **2**(1): 226–252.

Candès, E., Romberg, K. J. and Tao, T. 2006. Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics* **59**(8): 1207–1223.

Chambolle, A. 2004. An algorithm for total variation minimization and applications, *Journal of Mathematical imaging and vision* **20**(1): 89–97.

Chambolle, A., Levine, S. E. and Lucier, B. J. 2011. An upwind finite-difference method for total variation-based image smoothing, *SIAM J. Imaging Sci.* **4**(1): 277–299.

Chambolle, A. and Lions, P. 1995. Image recovery via total variation minimization and related problems, *Numerische Mathematik* **76**: 167–188.

Chan, T. F. and Chen, K. 2006. On a nonlinear multigrid algorithm with primal relaxation for the image total variation minimisation, *Numer. Algorithms* **41**(4): 387–411.

Chan, T. F., Golub, G. H. and Mulet, P. 1999. A nonlinear primal-dual method for total variation-based image restoration, *SIAM Journal on Scientific Computing* **20**(6): 1964–1977.

Chan, T. F. and Shen, J. 2005. *Image processing and analysis*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Variational, PDE, wavelet, and stochastic methods.

Chan, T. F. and Wong, C. 1998. Total variation blind deconvolution, *IEEE Transactions on Image Processing* **7**(6): 370–375.

Combettes, P. L. and Pesquet, J. 2004. Image restoration subject to a total variation constraint, *IEEE Transactions on Image Processing* **13**(9): 1213–1222.

Dai, Y.-H. and Fletcher, R. 2005. Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming, *Numerische Mathematik* **100**(1): 21–47.

Daubechies, I., Defrise, M. and De Mol, C. 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Communications on pure and applied mathematics* **57**(11): 1413–1457.

Deimling, K. 1985. *Nonlinear Functional Analysis*, Springer-Verlag.

Delaney, A. H. and Bresler, Y. 1998. Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography, *IEEE Transactions on Image Processing* **7**(2): 204–221.

Dobson, D. and Vogel, C. 1997. Convergence of an iterative method for total variation denoising, *SIAM Journal on Numerical Analysis* **34**(5): 1779–1791.

Duan, X., Zhang, L., Xing, Y., Chen, Z. and Cheng, J. 2009. Few-view projection reconstruction with an iterative reconstruction-reprojection algorithm and TV constraint, *Nuclear Science, IEEE Transactions on* **56**(3): 1377 –1382.

Esser, E., Zhang, X. and Chan, T. F. 2010. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science, *SIAM J. Imaging Sci.* **3**(4): 1015–1046.

Feng, X. and Prohl, A. 2003. Analysis of total variation flow and its finite element approximations, *M2AN Math. Model. Numer. Anal.* **37**(3): 533–556.

Fletcher, R. 2005. On the Barzilai-Borwein method, *Optimization and Control with Applications*, Vol. 96 of *Applied Optimization*, Springer US, pp. 235–256.

Fornasier, M., Langer, A. and Schönlieb, C.-B. 2010. A convergent overlapping domain decomposition method for total variation minimization, *Numer. Math.* **116**(4): 645–685.

Fornasier, M. and Schönlieb, C.-B. 2009. Subspace correction methods for total variation and l_1-minimization, *SIAM J. Numer. Anal.* **47**(5): 3397–3428.

Giusti, E. 1984. *Minimal surfaces and functions of bounded variation*, Vol. 80 of *Monographs in Mathematics*, Birkhäuser.

Goldstein, T. and Osher, S. 2009. The split Bregman method for $L1$-regularized problems, *SIAM J. Imaging Sci.* **2**(2): 323–343.

Grasmair, M., Haltmeier, M. and Scherzer., O. 2008. Sparse regularization with l^q penalty term, *Inverse Problems* **24**(5): 055020 (13pp).

Hale, E. T., Yin, W. and Zhang, Y. 2010. Fixed-point continuation applied to compressed sensing: implementation and numerical experiments, *J. Comput. Math.* **28**(2): 170–194.

Hämäläinen, K., Kallonen, A., Kolehmainen, V., Lassas, M., Niinimäki, K. and Siltanen, S. 2012. Sparse tomography, *Submitted manuscript* .

Hansen, P. C. 2010. *Discrete inverse problems*, Vol. 7 of *Fundamentals of Algorithms*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Insight and algorithms.

Herman, G. T. and Davidi, R. 2008. Image reconstruction from a small number of projections, *Inverse Problems* **24**(4): 045011.

Jensen, T. L., Jørgensen, J. H., Hansen, P. C. and Jensen, S. H. 2011. Implementation of an optimal first-order method for strongly convex total variation regularization, *BIT Numerical Mathematics* **DOI: 10.1007/s10543-011-0359-8**.

Kak, A. and Slaney, M. 1988. Principles of computerized tomographic imaging, *IEEE Service Center, Piscataway, NJ* .

Kolehmainen, V., Lassas, M., Niinimäki, K. and Siltanen, S. 2012. Sparsity-promoting Bayesian inversion, *Inverse Problems* **28**(2): 025005.

Kolehmainen, V., Siltanen, S., Järvenpää, S., Kaipio, J., Koistinen, P., Lassas, M., Pirttilä, J. and Somersalo, E. 2003. Statistical inversion for medical X-ray tomography with few radiographs: II. application to dental radiology, *Physics in Medicine and Biology* **48**: 1465–1490.

Kolehmainen, V., Vanne, A., Siltanen, S., Järvenpää, S., Kaipio, J., Lassas, M. and Kalke, M. 2006. Parallelized Bayesian inversion for three-dimensional dental X-ray imaging, *IEEE Transactions on Medical Imaging* **25**(2): 218–228.

Lassas, M. and Siltanen, S. 2004. Can one use total variation prior for edge-preserving Bayesian inversion?, *Inverse Problems* **20**: 1537–1564.

Li, Y. and Santosa, F. 1996. A computational algorithm for minimizing total variation in image restoration, *IEEE Transactions on Image Processing* **5**(6): 987 –995.

Liao, H. and Sapiro, G. 2008. Sparse representations for limited data tomography, *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 1375 –1378.

McCollough, C. H., Primak, A. N., Braun, N., Kofler, J., Yu, L. and Christner, J. 2009. Strategies for reducing radiation dose in ct, *Radiol Clin North Am* **47**(1): 27–40.

Mueller, J. and Siltanen, S. 2012. *Linear and Nonlinear Inverse Problems with Practical Applications*, SIAM.

Natterer, F. 1986. *The mathematics of computerized tomography*, Vol. 32, John Wiley & Sons, Chichester, USA, and B. G. Teubner, Stuttgart, Germany.

Natterer, F. and Wübbeling, F. 2001. *Mathematical Methods in Image Reconstruction*, Monographs on Mathematical Modeling and Computation, SIAM.

Nesterov, Y. 2011. Barrier subgradient method, *Math. Program.* **127**(1, Ser. B): 31–56.

Niemi, E. 2010. *Projected Barzilai–Borwein method for total variation regularization with nonnegativity constraints*, Bachelor of science thesis, Tampere University of Technology.

Niu, T. and Zhu, L. 2012. Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: Phantom studies, *Medical Physics* **39**(7): 4588–4598.

Osher, S., Burger, M., Goldfarb, D., Xu, J. and Yin, W. 2005. An iterative regularization method for total variation-based image restoration, *Multiscale Model. Simul.* **4**(2): 460–489 (electronic).

Osher, S. and Fedkiw, R. 2003. *Level set methods and dynamic implicit surfaces*, Vol. 153 of *Applied Mathematical Sciences*, Springer-Verlag, New York.

Park, J. C., Song, B., Kim, J. S., Park, S. H., Kim, H. K., Liu, Z., Suh, T. S. and Song, W. Y. 2012. Fast compressed sensing-based CBCT reconstruction using barzilai-borwein formulation for application to on-line IGRT, *Medical Physics* **39**(3): 1207–1217.

Raydan, M. 1993. On the Barzilai-Borwein choice of steplength for the gradient method, *IMA Journal of Numerical Analysis* **13**: 321–326.

Raydan, M. 1997. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM Journal on Optimization* **7**(1): 26–33.

Rudin, L., Osher, S. and Fatemi, E. 1992. Nonlinear total variation based noise removal algorithms, *Physica D: Nonlinear Phenomena* **60**(1-4): 259–268.

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M. and Lenzen, F. 2009. *Variational methods in imaging*, Vol. 167 of *Applied Mathematical Sciences*, Springer Verlag.

Sidky, E. Y. and Pan, X. 2008. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization, *Physics in Medicine and Biology* **53**(17): 4777.

Tang, J., Nett, B. E. and Chen, G.-H. 2009. Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms, *Physics in Medicine and Biology* **54**(19): 5781.

Tian, Z., Jia, X., Yuan, K., Pan, T. and Jiang, S. B. 2011. Low-dose CT reconstruction via edge-preserving total variation regularization, *Physics in Medicine and Biology* **56**(18): 5949.

Tsapaki, V., Aldrich, J. E., Sharma, R., Staniszewska, M. A., Krisanachinda, A., Rehani, M., Hufton, A., Triantopoulou, C., Maniatis, P. N., Papailiou, J. and Prokop, M. September 2006. Dose reduction in CT while maintaining diagnostic confidence: Diagnostic reference levels at routine head, chest, and abdominal CT—IAEA-coordinated research project, *Radiology* **240**(3): 828–834.

Vogel, C. 2002. *Computational methods for inverse problems*, number 23, SIAM, Frontiers in Applied Mathematics.

Vogel, C. and Oman, M. 1996. Iterative methods for total variation denoising, *SIAM Journal on Scientific Computing* **17**: 227–238.

Vogel, C. and Oman, M. 1998. Fast, robust total variation-based reconstruction of noisy, blurred images, *IEEE Transactions on Image Processing* **7**(6): 813–824.

Wang, Y., Yang, J., Yin, W. and Zhang, Y. 2008. A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imaging Sci.* **1**(3): 248–272.

Wu, C. and Tai, X.-C. 2010. Augmented lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models, *SIAM Journal on Imaging Sciences* **3**(3): 300–339.

Yin, W., Osher, S., Goldfarb, D. and Darbon, J. 2008. Bregman iterative algorithms for l_1-minimization with applications to compressed sensing, *SIAM J. Imaging Sci.* **1**(1): 143–168.

Yu, L., Liu, X., Leng, S., Kofler, J. M., Ramirez-Giraldo, J. C., Qu, M., Christner, J., Fletcher, J. G. and McCollough, C. H. 2009. Radiation dose reduction in computed tomography: techniques and future perspective, *Imaging in Medicine* **1**(1): 65–84.

Zhang, X., Burger, M. and Osher, S. 2011. A unified primal-dual algorithm framework based on Bregman iteration, *J. Sci. Comput.* **46**(1): 20–46.