1. Introduction. In X-ray tomography one collects projection images of an unknown two- or three-dimensional body from different directions. For each direction, an X-ray source is placed on one side of the body, and an X-ray detector is placed on the opposite side. After a calibration step involving a logarithm of the measured projection images, the X-ray tomography experiment is usually modeled by
\[
\tilde{g} = Au + e,
\]
where vector \(\tilde{g} \in \mathbb{R}^M\) is the data and \(u : \Omega \rightarrow \mathbb{R}_+\) is the X-ray attenuation function defined in a bounded domain \(\Omega \subset \mathbb{R}^n\). In this work, we take \(\Omega\) to be the unit square \([0,1]^2\). The operator \(A\) represents integrals of \(u\) over the areas or volumes connecting the X-ray source...
and a pixel in the detector; see Figure 1(a). The model in (1.1) is approximate in the sense that it neglects scattering phenomena [76, 2] and effects of nonmonochromatic radiation, such as beam hardening [83]. Furthermore, \( \mathbf{e} \in \mathbb{R}^{M} \) is a Gaussian random vector modeling measurement noise.

The inverse problem is given by measurement geometry and noisy data \( \tilde{g} \), recover a discrete approximation \( \mathbf{f} \in \mathbb{R}^{N} \) to \( u \). Depending on the number and geometry of projection directions, this inverse problem is either mildly or strongly ill-posed [66, 65]. Ill-posedness means high sensitivity to inaccuracies in modeling and to errors in measurement. Overcoming this sensitivity requires regularization.

In this work, we study regularizing two-dimensional (2D) tomographic problems using total variation (TV) regularization, which was introduced in [74]. We need a finite approximation to the continuous measurement model (1.1). Consider the discrete model

\[
\mathbf{g} = Af,
\]

where \( \mathbf{f} \in \mathbb{R}^{N} \) represents the discretized 2D body under imaging. We divide the square \( \Omega \) into \( N = n \times n \) square pixels with side length \( 1/n \). The function \( u \) is approximated by a function which is constant inside each pixel; those pixel values are listed as elements of the vector \( \mathbf{f} \). Further, \( \mathbf{g} \in \mathbb{R}^{M} \) denotes the measurement data consisting of pixel values in digital X-ray projection images, and \( A \in \mathbb{R}^{M \times N} \) is a matrix approximation to \( A \) (see [65, section 2.3.4]).

The discrete TV-regularized solution is defined by

\[
\hat{\mathbf{f}}_{\alpha,n} := \arg \min_{\mathbf{f} \in \mathbb{R}^{N}} \left\{ \| Af - \tilde{g} \|_{2}^{2} + \frac{\alpha}{n} \sum_{\kappa \sim \kappa'} | f_{\kappa} - f_{\kappa'} | \right\},
\]

where \( \alpha > 0 \) is a so-called regularization parameter. We use the anisotropic definition of TV: in (1.3) the relation \( \kappa \sim \kappa' \) is true whenever elements \( f_{\kappa} \) and \( f_{\kappa'} \) of the vector \( \mathbf{f} \) correspond to values at either horizontally or vertically neighboring pixels.

Figure 1. Schematic illustration about the relationship between continuous and discrete tomographic models. (a) Continuous model for two-dimensional X-ray tomography. The function \( u(x) \) of (1.1) gives the pointwise X-ray attenuation coefficient inside the domain \( \Omega \) shown here as a square. The dots show the five locations of the X-ray source. The detector has 10 pixels, so there is a total of 50 data points. (b)–(d) Discrete models with various resolutions. The attenuation coefficient is modeled as constant inside each pixel. The point of this figure is to demonstrate how the resolution of our computational model can be freely chosen, while the number of data points is fixed.
We remark that the ordinary least-squares formulation in (1.3) is a simplification that does not take into account, e.g., the nonuniform noise variance present in CT; see, for example, [69, 94] for details.

How to choose the regularization parameter $\alpha$ in (1.3)? Many methods have been suggested in the literature. Classical approaches include cross-validation (see [42] for an early example with Tikhonov-regularized tomography), the L-curve method [48], and Stein’s unbiased risk estimates (SURE) [72, 73, 33]. See also the interesting manuscript [75]. One possibility for parameter choice is to use the Bayesian framework by interpreting TV as a MAP estimator [54, 12], provided that the signal has suitable statistical properties. Clason, Jin, and Kunisch proposed a method based on balancing $\ell^1$ and TV norms [25] and tested it on image deblurring problems. Multiresolution-based parameter choices are discussed by Dong, Hintermüller, and Rincon-Camacho [31] and Frick, Marnitz, and Munk [40]. Wen and Chan [93] suggested a discrepancy principle designed for TV and tested it with image restoration. The so-called S-curve method was introduced in [56] and implemented for wavelet-based tomography in [47]. TV parameter choices based on the quasi-optimality principle and Hanke–Raus rules are described by Kindermann, Mutimbu, and Resmerita in [55]. Furthermore, Chen, Loli Piccolomini, and Zama [24] suggest determining the TV parameter in image deblurring by including it in a Karush–Kuhn–Tucker (KKT) system. Toma, Sixou, and Peyrin introduce an iterative choice in [86].

There is so far no parameter choice method that would be the best for all applications. Rather, it seems that for any given application there is a quest for a method that gives consistent, robust, and useful results. For this reason we think that there needs to be a large collection of methods for parameter choice, based on different principles, for scientists and engineers to have a better chance of finding a good option for each practical situation.

We propose a new parameter choice method, based on solving the inverse problem at multiple resolutions. Namely, while the number $M$ of measurement points is fixed (it is simply the number of detector elements times the number of projection images), the number $N = n \times n$ of pixels in the reconstruction can be freely chosen. See Figure 1 for an illustration. Our method is different from all previous methods.

Given a tomographic dataset, we can use definition (1.3) to compute reconstructions in the same square domain but at varying resolutions $n$. We prove in section 3 that if we keep $\alpha > 0$ in (1.3) fixed and let $n \to \infty$, then $\hat{f}_{\alpha,n}$ converges to a limit image. The technique of proof is analogous to the one-dimensional result [61].

We propose the following practical method for choosing the parameter $\alpha > 0$.

**Parameter choice method.** Given a noisy dataset $\tilde{g}$, compute $\hat{f}_{\alpha,n}$ defined by (1.3) for a set of $\alpha$ values and at two or more resolutions $n$, including the resolution intended for showing the final result. Calculate the discrete TV norms of the reconstructions, and define the optimal $\alpha$ to be the smallest value that leads to TV norms that do not depend significantly on $n$.

At this point we do not claim that the above algorithm is a parameter choice rule in the strict sense as defined in [34]. However, the numerical evidence presented in section 5 suggests that the method works well and, as expected, leads to larger $\alpha$ when the noise level is increased.

In light of our new theorem, the TV norms of the reconstructions converge for any choice of $\alpha$ as $n \to \infty$. How exactly can the above approach work at all? It seems that the ill-
posedness of the tomography problem shows up as significant $n$-dependence of the TV norms of reconstructions whenever $\alpha$ is too small to regularize away the instability caused by the term $\|A f - \tilde{g}\|_2^2$. Since the final intended resolution is included in the test resolutions, our choice method adapts to the relevant resolution and the noise level.

The reader may wonder why $\alpha$ is divided by $n$ in (1.3). The TV regularization formulation for smooth functions takes the form

\[
\arg\min_{u \geq 0} \left\{ \|Au - \tilde{g}\|_2^2 + \alpha \int_{\Omega} |\nabla u| dx \right\}
\]

(1.4) to the discrete minimization problem (1.3). The division by $n$ comes from approximations with a finite difference quotient and a midpoint rule integration quadrature:

\[
\int_{\Omega} \left| \frac{\partial u}{\partial x_1} \right| dx \approx \int_{\Omega} \left| u(x_1 + \frac{1}{n}, x_2) - u(x_1, x_2) \right| \frac{1}{n} dx \approx \left( \frac{1}{n} \right)^2 \sum_{\kappa \sim \kappa'} \left| f_{\kappa} - f_{\kappa'} \right| \frac{1}{n}.
\]

(1.5)

In (1.5) the relation $\kappa \sim \kappa'$ is true whenever elements $f_{\kappa}$ and $f_{\kappa'}$ of $f$ correspond to values at horizontally neighboring pixels. A similar computation using vertical differences is needed as well for the analysis regarding (1.4).

Applying TV regularization to tomographic problems dates back to 1998 [29] in the case of simulated data and to 2003 [57] in the case of measured data. Since then, there have been many further studies [58, 63, 80, 50, 84, 32, 8, 52, 85, 53, 78, 22, 23].

There are many computational approaches to minimizing the TV-regularized least-squares functional (1.3). These include quadratic programming [61, 51, 43, 56, 47], lagged diffusivity methods [30], domain decomposition methods [39, 38], Bregman distance methods [70, 95, 44, 14, 96], primal-dual methods [20, 19, 41, 35, 67], finite element methods [37, 5], discontinuous Galerkin methods [46], and other methods [62, 91, 92, 45, 17]. See also [90, 71, 21, 77, 49].

In this paper we use a primal-dual interior-point (PD-IP) quadratic programming method based on dividing the unknown into nonnegative and nonpositive parts in the spirit of [65, section 6.2]. This is not the most efficient TV regularization algorithm, but it works well for us. One advantage of this approach is natural and effective enforcement of nonnegativity of the attenuation coefficient.

Scientific novelties of this paper include

• proving the convergence of TV reconstructions as the resolution is increased;
• testing the S-curve method in the context of TV-regularized tomography; and
• introducing a novel multiresolution parameter choice method and testing it with measured X-ray data.

While we discuss here the new multiresolution parameter choice method for 2D tomographic problems only, the computational approach generalizes directly to three-dimensional settings and other inverse problems as well. A one-dimensional analogue of our convergence result is available in [61].

Let us comment on related results in the literature. The interplay between resolution and Tikhonov regularization was studied by Vainikko in [88] in 1992. Chambolle’s study [15] of Mumford–Shah type image segmentation uses geometric arguments very similar to the proof of Lemma 5 below. Estimated TV norm constraint is also considered in [26] in connection.
with an adaptive level set algorithm. The papers [16, 89, 7, 87] study Γ-convergence of TV functionals as the discretization is refined, but without a measurement operator crucial for tomography. We note that the convergence results of the type we prove are quite delicate, as shown by [5, Example 4.1].

Most closely connected to our work is [13], where Cai et al. use wavelet frames to show Γ-convergence of discrete regularization functionals to the infinite-dimensional variational functional. Penalizing \( \ell^1 \) norms of wavelet coefficients in their approach leads to a regularization resembling TV (similarly to [60] in the context of Besov \( B^1_{1,1} \) priors in Bayesian inversion). However, they have the luxury of representing functions using frames, which is not available in our case of actual anisotropic TV regularization.

2. The space \( BV(D) \): Definitions and approximations. Let \( D \) be the square \([0, 1]^2 \subset \mathbb{R}^2\). We use a special \( \ell^1 \)-based anisotropic norm for the space \( BV(D) \) of functions of bounded variation defined on \( D \).

**Definition 1.** The space \( BV(D) \) consists of functions \( u \in L^1(D) \) whose distributional derivative \( \nabla u \) belongs to \( M(D, \mathbb{R}^2) \) and for which
\[
\sup_\varphi \left\{ \langle \nabla u, \varphi \rangle \bigg| \varphi \in C_c(D, \mathbb{R}^2), \|\varphi\|_\infty \leq 1 \right\} < \infty.
\]

Above \( M(D, \mathbb{R}^2) \) denotes \( \mathbb{R}^2 \)-valued Borel measures on \( D \).

**Definition 2.** Define an anisotropic seminorm for \( BV(D) \) in weak form by
\[
V(u) := V(u; D) = \sum_{\ell=1}^2 \sup_\varphi \left\{ \langle \frac{\partial u(x)}{\partial x_\ell}, \varphi_\ell(x) \rangle \bigg| \varphi_j \in C_c(D, \mathbb{R}), \|\varphi_\ell\|_\infty \leq 1 \right\}.
\]

Note that if \( u \in BV(D) \) is smooth, then we have
\[
V(u) = \int_D \left( \left| \frac{\partial u(x)}{\partial x_1} \right| + \left| \frac{\partial u(x)}{\partial x_2} \right| \right) dx.
\]

The anisotropic ABV norm is defined by
\[
\|u\|_{ABV(D)} := \|u\|_{L^1(D)} + V(u; D).
\]

We will use two different concepts of weak convergence in the space \( BV(D) \).

**Definition 3.** A sequence \( u_j \in BV(D) \) converges weakly in \( BV(D) \) to limit \( u \) as \( j \to \infty \), that is, \( u_j \overset{w}{\rightharpoonup} u \), if \( u_j \) and their distributional derivatives \( \nabla u_j \) satisfy
\[
\lim_{j \to \infty} \|u_j - u\|_{L^1(D)} = 0 \quad \text{and} \quad \nabla u_j \rightharpoonup \nabla u \quad \text{weakly in} \quad M(D, \mathbb{R}^2) \quad \text{as} \quad j \to \infty.
\]

Note that \( M(D, \mathbb{R}^2) \) is the dual space of \( C(D; \mathbb{R}^2) \). Recall now from [3, Definition 10.1.3] the definition of another useful type of weak convergence.
Definition 4. Let $u_j$ be a sequence in $BV(D)$ and $u \in BV(D)$. Then $u_j$ converges to $u$ as $j \to \infty$ in the sense of anisotropic intermediate convergence if and only if
\begin{align}
(2.5) & \quad \lim_{j \to \infty} \|u_j - u\|_{L^1(D)} = 0 \quad \text{and} \\
(2.6) & \quad \lim_{j \to \infty} V(u_j) = V(u).
\end{align}

Also, below we use the following distance function leading to a notion of convergence equivalent to (2.5) and (2.6):
\begin{equation}
(2.7) \quad \rho_{1,1}(u,v) = \|u - v\|_{L^1(D)} + |V(u) - V(v)|.
\end{equation}

According to [3, Proposition 10.1.1], anisotropic intermediate convergence of Definition 4 implies weak convergence of Definition 3.

We need a result on approximating elements of $BV(D)$ by functions that are piecewise constant on square pixels of fine enough resolution. For each $j > 0$, consider the closed dyadic squares
\begin{equation}
(2.8) \quad S_{j,\vec{k}} = S_{j,k_1,k_2} = [k_1 2^{-j}, (k_1 + 1)2^{-j}] \times [k_2 2^{-j}, (k_2 + 1)2^{-j}] \subset D,
\end{equation}
where $0 \leq k_1 \leq 2^j - 1$ and $0 \leq k_2 \leq 2^j - 1$. See Figure 2(b) for an illustration of a dyadic grid formed by such squares. Let $T_j : L^2(D) \to L^2(D)$ be the orthogonal projections defined by averaging over squares:
\begin{equation}
(2.9) \quad (T_j u)|_{\text{int}(S_{j,\vec{k}})} = \frac{1}{|S_{j,\vec{k}}|} \int_{S_{j,\vec{k}}} u(x) \, dx.
\end{equation}

For any $j > 0$, the set $Y_j := \text{Range}(T_j)$ consists of $L^2(D)$ functions that are piecewise constant in the open squares belonging to the rectangular $2^j \times 2^j$ grid (2.8).

Lemma 5. For all $u \in BV(D)$ and $\varepsilon > 0$ there is $j_1 > 0$ such that there is a $v_2$ that is piecewise constant in the dyadic $2^{j_1} \times 2^{j_1}$ grid (2.8) and satisfies
\begin{equation}
\rho_{1,1}(u, v_2) < \varepsilon,
\end{equation}
where the $\rho_{1,1}$-distance is defined by (2.7).

Proof. Let $u \in BV$ and $\varepsilon > 0$. We know from [6] that there is a function $v_0$ which is piecewise constant in a finite triangulization of $D$ and satisfies
\begin{equation}
(2.10) \quad \rho_{1,1}(u, v_0) < \varepsilon/3.
\end{equation}

Set $M = \|v_0\|_{L^\infty(D)}$. Let us choose some notation for the concepts related to the fixed triangulization corresponding to $v_0$.

1. Triangles are denoted by $T_\nu \subset D$ with $\nu = 1, 2, \ldots, T$.
2. Edges of the triangles are line segments $\gamma_\ell \subset D$ indexed by $\ell = 1, 2, \ldots, L$. 
3. Vertices are denoted by \( p_r \in D \) with \( r = 1, 2, \ldots, R \).
4. The minimum distance between two separate vertices is denoted by \( \delta_0 > 0 \).
5. The minimum angle between any two edges sharing an endpoint is called \( \alpha_0 > 0 \).
6. The length of the longest edge is denoted by \( G \).

See Figure 2(a) for an illustration of a finite triangulization.

Next we define clusters. Given \( j > 1 \), let \( B'_r = B'_{r,j} \) be the union of all the closed squares \( S_{j,k} \) containing the vertex \( p_r \). See the orange sets in Figure 3. Furthermore, define clusters \( B_r = B_{r,j} \) by

\[
B_r = \bigcup \{ S_{j,k} \mid S_{j,k} \cap B'_r \neq \emptyset \}.
\]

Then \( p_r \in B'_r \subset B_r \). The clusters are shown in gray in Figures 3, 4, and 7.

In the following, consider integers \( j_0 \) large enough for the following to hold:

\[
4\sqrt{2} \cdot 2^{-j_0} < \delta_0.
\]

Now (2.12) implies that the sets \( B_r \) are disjoint and contain at most one vertex each.

Let \( v_1 \) be the function obtained from \( v_0 \) by replacing its value in each cluster \( B_r \) by 0.
Then we have

\[ \|v_0 - v_1\|_{L^1(D)} \leq \sum_{r=1}^{R} \int_{B_r} |v_0 - v_1| \, dm \leq 16RM(2^{-j_0})^2 \]

since each cluster consists of at most 16 squares of area \((2^{-j_0})^2\). Furthermore, the \(V\) expression (2.2) can be very simply computed for functions that are piecewise constant with respect to either a finite triangulation or a square grid. Namely, (2.2) is the sum of (Manhattan distance) lengths of jump curve segments multiplied by the jump in function value over those curve segments. Therefore, we can estimate

\[ |V(v_0) - V(v_1)| \leq 8LMR2^{-j_0} + 32MR2^{-j_0}, \]

where the first term on the right-hand side comes from the triangularization. We used thefacts that the Manhattan length of any edge restricted inside a cluster is at most 4 \cdot 2^{-j_0} andthat \(v_0\) can jump at most 2\(M\) at an edge. The second term on the right-hand side comes fromthe jump of \(v_1\), bounded by 2\(M\), at the cluster boundary whose maximal length is 16 \cdot 2^{-j_0}.

Now we can fix \(j_0\) so large that (2.12) holds and we have

(2.13) \quad \rho_{1,1}(v_0, v_1) < \varepsilon/3.

Next we construct tubular neighborhoods. Denote the union of all clusters by

\[ \mathcal{B} = \bigcup_{r=1}^{R} B_r. \]

Note that \(\mathcal{B} \subset \mathbb{R}^2\) is a closed set. Further, consider the parts of triangle edges that are outside the clusters:

(2.14) \quad \tilde{\gamma}_\ell := \gamma_\ell \setminus \mathcal{B}.

For all \(\varrho > 0\), define

(2.15) \quad \Gamma_{\ell, \varrho} := \{ x \in \mathbb{R}^2 | \text{dist}(x, \tilde{\gamma}_\ell) < \varrho \} \setminus \mathcal{B},

where “dist” refers to Euclidean distance. The sets \(\Gamma_{\ell, \varrho}\) are open tubular neighborhoods of \(\tilde{\gamma}_\ell\) in \(D \setminus \mathcal{B}\). See Figures 4(a) and 5.

We need to determine an upper bound for \(\varrho > 0\) ensuring that the sets \(\Gamma_{\ell, \varrho}\) are disjoint. It is clear from the construction of the sets \(B_r\) and \(B'_r\) that (Euclidean) \(\text{dist}(p_r, D \setminus \mathcal{B}) > 2^{-j_0}\). As illustrated in Figure 4(a), the minimal angle \(\alpha_0\) between edges determines a lower bound for the distance between the intersection points of two edges with the boundary \(\partial B_r\). A trigonometric consideration shown in Figure 4(b) implies that the tubular neighborhoods \(\Gamma_{\ell, \varrho}\) are disjoint whenever \(\varrho\) satisfies \(0 < \varrho < h_0\) with

(2.16) \quad h_0 := 2^{-j_0} \tan \frac{\alpha_0}{2}. 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Figure 4. (a) Illustration of two disjoint sets $\Gamma_{1,0}$ and $\Gamma_{2,0}$ (pink) as defined in (2.15). The two edges $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, shown in blue, have the minimal angle $\alpha_0$ between them. The construction of $B_r$ and $B'_r$ is done in the square grid of size $2^{-j_0} \times 2^{-j_0}$. (b) The worst-case limit situation when $p_r$ is at the boundary of $B'_r$ is where the definition (2.16) comes from.

See Figure 5 for an illustration of the tubular neighborhoods in a triangulation.

Next we consider refining the square grid outside the clusters as shown in Figure 6. The smaller squares of size $2^{-j_1} \times 2^{-j_1}$ are defined by the formula (2.8) with a large enough $j_1 > j_0$. The idea is to take $j_1$ so large that we can fit a dyadic polygonal chain inside the tubular neighborhoods, as shown in Figure 7(a).

Consider two vertices $p_1$ and $p_2$ connected by an edge $\gamma_1$. (There is no loss of generality since we can always renumber the vertices and edges.) Take $j_1$ so large that

\begin{equation}
2^{-j_1} < \frac{\varrho}{4}.
\end{equation}

Then we can connect the clusters $B_1$ and $B_2$ using a piecewise linear curve

$$c_1 : [0, 1] \to \Gamma_{1,0}$$

with the following properties:

- The edge $\gamma_1$ and the dyadic polygonal chain $c_1$ are disjoint: $\gamma_1 \cap c_1 = \emptyset$.
- The curve starts from the boundary of $B_1$ and ends at the boundary of $B_2$; in other words, $c_1(0) \in \partial B_1$ and $c_1(1) \in \partial B_2$. 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Figure 5. This is an illustration of the tubular neighborhoods $\Gamma_{\ell,\varrho}$ (pink), defined in (2.15), which are disjoint whenever $0 < \varrho < h_0$ with $h_0$ given by (2.16). The vertices $p_{\ell}$ are shown as blue dots, and the “triangle edges outside clusters,” or line segments $\tilde{\gamma}_{\ell}$ defined in (2.14), are shown as blue lines.

- The curve $c_1$ is close to the edge $\gamma_1$:
  \begin{equation}
  \min_{x \in \gamma_1} \| c_1(t) - x \|_2 < 3 \cdot 2^{-j_1} \quad \text{for all } t \in [0,1].
  \end{equation}

- The range $c_1([0,1])$ of the curve consists of boundary segments of squares belonging to the $2^{-j_1} \times 2^{-j_1}$ grid.

- The curve is injective: if $t_1 \neq t_2$, then $c_1(t_1) \neq c_1(t_2)$.

- The dyadic polygonal approximation $c_1$ is optimal in the following sense: if $c_1' : [0,1] \to \Gamma_{1,\varrho}$ is another curve satisfying all of the above conditions for $c_1$, then
  \[ \text{length}(c_1') \geq \text{length}(c_1([0,1])). \]

See Figure 6 for an illustration. Construct $c_\ell$ corresponding to each $\gamma_\ell$ with $\ell = 1, \ldots, L$.

We move on to approximating functions using a square grid. Consider the function $v_1$, which is piecewise constant in the triangulation of $D$, restricted to a tubular neighborhood $\Gamma_{\ell,\varrho}$. The line segment $\tilde{\gamma}_\ell$ divides $\Gamma_{\ell,\varrho}$ into two components, and $v_1$ has a constant value in each component. See Figure 7(b).

Now define another function $v_2$, which is piecewise constant on the square grid of size $2^{-j_1} \times 2^{-j_1}$. Outside the tubular neighborhoods $\Gamma_{\ell,\varrho}$ we set $v_1 \equiv v_2$. Inside $\Gamma_{\ell,\varrho}$ we define the values of $v_2$ uniquely as follows. Consider the dyadic polygonal chains $c_\ell$ that are located close to the edges $\tilde{\gamma}_\ell$. Now $v_2$ is allowed to have discontinuities only along the curves $c_\ell$ or at the
cluster boundaries \( \partial B_r \). See Figure 7(c).

By the construction of the dyadic polygonal chain the area between an edge \( \tilde{\gamma}_\ell \) and the polygonal chain is bounded by \( 3G2^{-j_1} \). Therefore, we can estimate

\[
\|v_1 - v_2\|_{L^1(D)} \leq 6MG2^{-j_1},
\]

since we have \( |v_1(x) - v_2(x)| \leq 2M \) for almost all \( x \in D \).

Now both \( v_1 \) and \( v_2 \) are piecewise constant, and the domains of constant value are polygons. The expressions \( V(v_1) \) and \( V(v_2) \) can thus be calculated by summing the following quantities over the linear boundary segments: the jump in function values over the segment multiplied by the Manhattan length of the segment. Note that in \( |V(v_1) - V(v_2)| \) all of the terms cancel each other except those coming from boundary segments located in the tubular neighborhoods \( \Gamma_{\ell,\rho} \). Due to the construction of \( c_\ell \), the Manhattan and Euclidean lengths of \( c_\ell \) are the same. Further, the Manhattan length of \( \tilde{\gamma}_\ell \) equals the length of \( c_\ell \), apart from the difference between the two red intervals \( I_\ell \) and \( I'_\ell \) shown in Figure 7(a). By (2.18) we have

\[
|I_\ell| - |I'_\ell| \leq |I_\ell| + |I'_\ell| \leq 6 \cdot 2^{-j_1}.
\]

This together with the maximal jump size \( 2M \) allows us to estimate

\[
|V(v_1) - V(v_2)| \leq 12ML2^{-j_1}.
\]

Using (2.19) and (2.20) we see that for large enough \( j_1 \) we have \( \rho_{1,1}(v_1, v_2) \leq \varepsilon/3 \), and the proof is complete.

3. Multiresolution convergence analysis. Define a functional \( S_\infty : BV(D) \to \mathbb{R} \) by

\[
S_\infty(u) = \|Au - m\|_{L^2(\Omega)}^2 + \alpha_1\|u\|_{L^1(D)} + \alpha V(u)
\]

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Figure 7. Square-grid based approximation of functions that are piecewise constant on a triangulation. (a) Detail from a situation similar to that shown in Figure 6. The straight blue line is the edge $\tilde{\gamma}_l$, and the black jagged line is the dyadic polygonal chain $c_l$. The red intervals $I_l$ and $I'_l$ are used for estimating the difference between the $BV$ norms of $v_1$ and $v_2$. (b) A function $v_1$ which is piecewise constant on a triangulation. (c) Approximation of $v_1$ by another function $v_2$ which is piecewise constant on a square grid. The difference of the $BV$ norms of $v_1$ and $v_2$ comes entirely from jumps over the two red vertical intervals shown in (a).

with positive regularization parameters $\alpha_1 > 0$ and $\alpha > 0$, and $V$ given by (2.2). The functional $S_\infty$ defines an infinite-dimensional minimization problem which we will approximate by discrete minimization problems. For this we define functionals $S_j : BV(D) \to \mathbb{R} \cup \{\infty\}$ by

$$ S_j(u) = \begin{cases} S_\infty(u) & \text{for } u \in \text{Range}(T_j), \\ \infty & \text{for } u \notin \text{Range}(T_j). \end{cases} $$

Denote

$$ s_\infty := \inf_{u \in BV} S_\infty(u), \quad s_j := \inf_{u \in BV} S_j(u). $$

We study the existence and properties of minimizers $\tilde{u}_j, \tilde{u}_\infty \in BV(D)$ satisfying $S_j(\tilde{u}_j) = s_j$ and $S_\infty(\tilde{u}_\infty) = s_\infty$, respectively. Such minimizers may not be unique. Therefore, we use the notation

$$ \arg \min(S_j) = \{ u \in BV(D) : S_j(u) = s_j \}, \quad \arg \min(S_\infty) = \{ u \in BV(D) : S_\infty(u) = s_\infty \}. $$

We are now ready to state our main theorem.

**Theorem 6.** Assume either (A) or (B) about the linear operator $A$:

(A) $A : L^2(D) \to L^2(\Omega)$ is compact and $A : L^1(D) \to \mathcal{D}'(\Omega)$ is continuous with some open and bounded set $\Omega \subset \mathbb{R}^2$.

(B) $A : L^1(D) \to \mathbb{R}^M$ is bounded.

Let the functionals $S_j$ and $S_\infty$ be given by (3.2) and (3.1), respectively, and define $s_j$ and $s_\infty$ as in (3.3). Then the following hold:

(i) $\lim_{j \to \infty} s_j = s_\infty$.
(ii) There exists a minimizer $\tilde{u}_j \in \arg \min(S_j)$ for all $j = 1, 2, 3, \ldots$.
(iii) There exists a minimizer $\tilde{u}_\infty \in \arg \min(S_\infty)$.
Any sequence $\tilde{u}_j \in \text{arg min}(S_j)$ of minimizers has a subsequence $\tilde{u}_{j(\ell)}$ that converges weakly in $BV(D)$ to some limit $w \in BV(D)$, or, more briefly, $\tilde{u}_{j(\ell)} \Rightarrow w$ as $\ell \to \infty$. Furthermore, $\lim_{\ell \to \infty} V(\tilde{u}_{j(\ell)}) = V(w)$; that is, the subsequence $\tilde{u}_{j(\ell)}$ converges to $w$ in the anisotropic intermediate topolgy of $BV(D)$.

The limit $w =: \tilde{u}_\infty$ in (iv) is a minimizer: $\tilde{u}_\infty \in \text{arg min}(S_\infty)$.

The set $\Omega$ in (3.1) can be an open rectangle with $L^2(\Omega)$, modeling sinogram space, or $\Omega = \{1, 2, \ldots, M\}$, equipped with the counting measure, leading to $L^2(\Omega) = \mathbb{R}^M$. These two options correspond to assumptions (A) and (B) in Theorem 6.

The connection between Theorem 6 and the computational explanation in the introduction is this. In the computational part of this paper we take $\alpha_1 = 0$, which roughly corresponds to having a really small $\alpha_1 > 0$ in the theory. The minimizing vector $\tilde{f}_{n,n} \in \mathbb{R}^{n \times n}$ defined in (1.3) can be seen as an $n \times n$ matrix with $n = 2^j$. The matrix $\tilde{f}_{n,n}$ is a 2D array of the values of a piecewise constant minimizer function $\tilde{u}_j \in \text{arg min}(S_j)$ in the pixels (2.8). Thus we see that for any choice of regularization parameter $\alpha$ in (1.3), the TV norms of $\tilde{f}_{n,n}$ with $n = 2^j$ converge to a limit as $j \to \infty$. The measurement operator discussed in the introduction is as in assumption (B).

Remark. Our convergence result is quite delicate. Even a slight change of norms will destroy the argumentation. See Appendix A for more details about this.

Before proving the main theorem we prove a couple of lemmas.

Lemma 7. Let the assumptions of Theorem 6 hold. Assume that $u_j \to u$ as $j \to \infty$ in the weak topology of $BV(D)$ given in Definition 3. Then

$$\lim_{j \to \infty} Au_j = Au$$

with convergence in $L^2(\Omega)$ or in $\mathbb{R}^M$ in the cases of assumption (A) or (B) in Theorem 6, respectively. Moreover, it follows that

$$\|Au - m\|_{L^2(\Omega)}^2 + \alpha_1\|u\|_{L^1(D)} + \alpha V(u) \leq \liminf_{j \to \infty} \left(\|Au_j - m\|_{L^2(\Omega)}^2 + \alpha_1\|u_j\|_{L^1(D)} + \alpha V(u_j)\right).$$

Proof. Assume (A) of Theorem 6. By [3, Theorem 10.1.3] we have

$$\|u\|_{L^2(D)} \leq C_1\|u\|_{BV(\Omega)}.$$ Moreover, there is $c_1$ such that $\|u_j\|_{BV(\Omega)} \leq c_1$ for all $n$. Hence, $\|u_j\|_{L^2(D)} \leq C$ for all $j$. To show that $Au_j$ converges in $L^2(\Omega)$ to $Au$ as $j \to \infty$, assume the opposite. Then there is $\varepsilon > 0$ and a subsequence $j(\ell) \to \infty$ as $\ell \to \infty$ such that

$$\|Au_{j(\ell)} - Au\|_{L^2(\Omega)} \geq \varepsilon.$$ Now $\|u_{j(\ell)}\|_{L^2(D)} \leq C$, and as $A : L^2(D) \to L^2(\Omega)$ is compact, there is a subsequence $u_{j(\ell_p)}$ and $w \in L^2(\Omega)$ such that

$$\lim_{p \to \infty} \|Au_{j(\ell_p)} - w\|_{L^2(\Omega)} = 0.$$
On the other hand, as \( u_j \xrightarrow{\text{weak}} u \), we have that \( u_j \to u \) in the strong topology of \( L^1(D) \) by definition. This implies that
\[
(3.8) \quad \lim_{p \to \infty} A u_j(t_p) = A u \quad \text{in } D'(\Omega).
\]
The above two limits (3.7) in \( L^2(\Omega) \) and (3.8) in \( D'(\Omega) \) imply that \( A u = w \). This, (3.6), and (3.7) are in contradiction. Hence we have (3.4) in the case of assumption (A) in Theorem 6.

Inequality (3.5) follows from the above and [28, Proposition 2.4.12(c)].

Assumption (B) can be handled by adding a projection operator similar to (2.9).

**Lemma 8.** Let the assumptions of Theorem 6 hold. Consider the functionals \( S_j : (BV(D), T'') \to [0, +\infty] \) defined in (3.2) and \( S_\infty : (BV(D), T'') \to [0, +\infty] \) defined in (3.1). Here \( T'' \) denotes the anisotropic intermediate topology of Definition 4. Then the sequence \( S_j \) of functionals \( \Gamma \)-converges to \( S_\infty \) as \( j \to \infty \).

**Proof.** Note that one can use the distance function (2.7) to show that \( T'' \) is metrizable. Because of the metrizability, showing \( \Gamma \)-convergence can be done by proving two things [28, Proposition 8.1]. The lower-bound inequality
\[
(3.9) \quad S_\infty(u) \leq \liminf_{j \to \infty} S_j(u_j)
\]
should hold whenever there is anisotropic intermediate convergence of \( u_j \in BV(D) \) to the limit \( u \in BV(D) \). Further, we must show that for every \( u \in BV(D) \) there is a sequence \( u_j \in BV(D) \) that converges to \( u \) in the anisotropic intermediate topology and satisfies the upper-bound inequality
\[
(3.10) \quad S_\infty(u) \geq \limsup_{j \to \infty} S_j(u_j).
\]
This proves the lower-bound inequality (3.9). Assume that \( u_j \) converges to \( u \) in the anisotropic intermediate topology of \( BV \) as \( j \to \infty \). Then by definition we have \( \lim_{j \to \infty} \| u_j \|_{L^1(D)} = \| u \|_{L^1(D)} \). Furthermore, since anisotropic intermediate convergence implies weak convergence by [3, Proposition 10.1.1], Lemma 7 shows that \( \lim_{j \to \infty} A u_j = A u \) in \( L^2(\Omega) \), and so
\[
\lim_{j \to \infty} \| A u_j - m \|_{L^2(\Omega)} = \| A u - m \|_{L^2(\Omega)}.
\]
Moreover, anisotropic intermediate convergence implies \( V(u) = \lim_{j \to \infty} V(u_j) \). Now
\[
S_\infty(u) = \| A u - m \|_{L^2(\Omega)}^2 + \alpha_1 \| u \|_{L^1(D)} + \alpha V(u)
\]
\[
= \lim_{j \to \infty} \left( \| A u_j - m \|_{L^2(\Omega)}^2 + \alpha_1 \| u_j \|_{L^1(D)} + \alpha V(u_j) \right)
\]
\[
= \limsup_{j \to \infty} S_\infty(u_j)
\]
\[
\leq \liminf_{j \to \infty} S_j(u_j).
\]

Proving the upper-bound inequality (3.10). Take any \( u \in BV(D) \). Recall the orthogonal projection operators \( T_j : L^2(D) \to L^2(D) \) defined in (2.9), and the spaces \( Y_j = \text{Range}(T_j) \) of piecewise constant functions. As \( Y_j \subset Y_{j+1} \), by Lemma 5 there are \( u_j \in Y_j \) such that \( \lim_{j \to \infty} p_{1,1}(u_j, u) = 0 \), so there is anisotropic intermediate convergence of \( u_j \) to \( u \) as \( j \to \infty \). Now \( u_j \in Y_j \) and formula (3.2) together imply \( S_j(u_j) = S_\infty(u_j) \). The upper-bound inequality (3.10) now follows from (3.11).
Proof of Theorem 6. Let us restrict $S_j$ and $S_\infty$ to the set

$$B_R = \{ u \in BV(D) : (\alpha_1\|u\|_{L^1(D)} + \alpha V(u)) \leq R \}$$

and call the resulting functionals $S_j^{(R)}$ and $S_\infty^{(R)}$, respectively. Denote by $T'$ the topology in $B_R \subset BV(D)$ induced by the weak topology of $BV(D)$. Since $C(D)$ is separable, the space $(B_R, T')$ is metrizable and, in particular, satisfies the first axiom of countability as a topological space.

Next we show that the restricted functionals $S_j^{(R)}$ $\Gamma$-converge to the restricted limit functional $S_\infty^{(R)}$. When $u_j \in B_R$ converges in the weak topology to the limit $u \in B_R$, the lower-bound inequality (3.9) follows from Lemma 7. Furthermore, Lemma 8 yields that for all $u \in B_R$ there is some sequence $u_j \in BV(D)$ converging to $u$ in the anisotropic intermediate topology (and therefore in the weak topology) of $BV$ for which the the upper-bound inequality (3.10) holds. Then, we see that the sequence

$$(3.13) \quad v_j(x) := \frac{V(u) + \|u\|_{L^1(D)}}{V(u_j) + \|u_j\|_{L^1(D)}} + \frac{1}{j} u_j(x)$$

satisfies $v_j \in B_R$, $v_j$ converges to $u$ in the anisotropic intermediate topology of $BV$, and the upper-bound inequality (3.10) is valid also for the sequence $v_j$ and $u$. As the anisotropic intermediate convergence in $BV$ implies the weak convergence in $BV$, these facts show that $S_j^{(R)}$ $\Gamma$-converge to $S_\infty^{(R)}$, in $(B_R, T')$, as $j \to \infty$ for any fixed $R > 0$; see, e.g., [4, 3].

Let us prove the existence of minimizers for $S_\infty^{(R)}$ and $S_j^{(R)}$. Note first that

- by Lemma 7 the functionals $S_\infty$ and $S_j$ are lower semicontinuous with respect to the weak topology of $BV(D)$, and
- by the Banach–Alaoglu theorem and [3, Theorem 10.1.4] the space $(B_R, T')$ is compact.

Therefore, the functionals $S_\infty$ and $S_j$ have minimizers for any $R > 0$.

Set $R = s_\infty + 1$ with $s_\infty$ defined in (3.3). Then there exists a minimizer $\tilde{u}_\infty \in \arg\min(S_\infty^{(R)})$ belonging to the open interior of $B_R$.

Using (3.13), we see that there are $u_j \in Y_j \cap B_R$ for which $u_j \to \tilde{u}_\infty$ in the anisotropic intermediate topology and (3.12) holds. For all large enough $j$ we have

$$S_\infty(u_j) \leq S_\infty(\tilde{u}_\infty) + 1 = s_\infty + 1.$$ 

As $S_j(u_j) = S_\infty(u_j)$ and $S_m(u) \leq S_j(u)$ for $m \geq j$, we see that for all large enough $j$ the minimizers of the functionals $S_j^{(R)}$ coincide with the minimizers of $S_j$. Therefore, we do not lose any generality in our minimization analysis by restricting to $B_R$.

Let us now study the convergence properties of minimizers. As we saw above, for large enough $j$ all minimizers $\tilde{u}_j \in \arg\min(S_j) = \arg\min(S_j^{(R)})$ and $\tilde{u}_\infty \in \arg\min(S_\infty) = \arg\min(S_\infty^{(R)})$ are in the ball $B_R$. The compactness of $(B_R, T')$ implies that any sequence $(\tilde{u}_j)_{j=1}^{\infty}$ of minimizers has a subsequence $(\tilde{u}_j(\ell))_{\ell=1}^{\infty}$ converging to some $w \in B_R$ in the weak topology of $BV$.

By [28, Corollary 7.20], the limit $w$ is then a minimizer of $S_\infty^{(R)}$, and we have

$$(3.14) \quad \lim_{\ell \to \infty} S_j(\ell)(\tilde{u}_j(\ell)) = \lim_{\ell \to \infty} S_j^{(R)}(\tilde{u}_j(\ell)) = S_\infty^{(R)}(w) = S_\infty(w) = s_\infty.$$
Moreover, by [28, Corollary 7.20], (3.14) holds for an arbitrary sequence $u_{j(\ell)} \in \text{argmin}(S_{j(\ell)})$, $\ell = 1, 2, \ldots$, that converges in the weak topology of $BV$ to some limit $w$, and $w$ is a minimizer of $S_\infty$.

Let us next consider a subsequence $\tilde{u}_{j(\ell)} \in \text{argmin}(S_{j(\ell)})$ converging to $\tilde{u}_\infty \in \text{argmin}(S_\infty)$ in the weak topology of $BV$. Then we have

$$V(\tilde{u}_\infty) \leq \liminf_{\ell \to \infty} V(\tilde{u}_{j(\ell)}).$$

(3.15)

Assume that we would have here a strict inequality.

By definition we have for any subsequence $\tilde{u}_{j(\ell)}$ converging to $\tilde{u}_\infty$ in the weak topology of $BV$

$$\|\tilde{u}_\infty\|_{L^1(D)} = \lim_{\ell \to \infty} \|\tilde{u}_{j(\ell)}\|_{L^1(D)}.$$

By Lemma 7, we have

$$\|A\tilde{u}_\infty - m\|_{L^2(\Omega)} = \lim_{\ell \to \infty} \|A\tilde{u}_{j(\ell)} - m\|_{L^2(\Omega)}.$$

(3.16)

These and a strict inequality in (3.15) would imply that $\liminf_{\ell \to \infty} S_{j(\ell)}(u_{j(\ell)}) < s_\infty$, but that is not possible in view of (3.14). Therefore,

$$V(\tilde{u}_\infty) = \liminf_{\ell \to \infty} V(\tilde{u}_{j(\ell)}).$$

(3.17)

Similar arguments with $\liminf$ replaced by $\limsup$ yield the desired equality. 

Remark 9. In the numerical computations below we include a nonnegativity constraint in the minimization problem. This models the physical fact that X-rays can only attenuate inside matter—not intensify. The nonnegativity constraint can be added also to the above convergence proofs. Let us explain.

The set $L^1_+(D) = \{u \in L^1(D); u(x) \geq 0 \text{ a.e. } x \in D\}$ is closed in $L^1(D)$, and thus the set $BV_+(D) = \{u \in BV(D); u(x) \geq 0 \text{ a.e. } x \in D\}$ is closed both in the weak topology of $BV(D)$ and in the anisotropic intermediate topology of $BV(D)$. Also, let $u^+(x) = \max(u(x), 0)$ denote the positive part of $u$, let $\chi_+(u)$ be an indicator function that is zero for $u \in BV_+(D)$ and equal to the infinity for $u \in BV(D) \setminus BV_+(D)$, and let

$$S_j^{(R,+)}(u) = S_j^{(R)}(u) + \chi_+(u), \quad S_\infty^{(R,+)}(u) = S_\infty^{(R)}(u) + \chi_+(u).$$

Moreover, let $B_R^+ = B_R \cap BV_+(D)$.

By [36, section 5.5, Theorem 1], the map $v \mapsto v^+$ satisfies $\|v\|_{BV} \leq \|v^+\|_{BV}$. Using this and [3, Proposition 10.1.1], we observe that if $v_n \in Y_n \cap B_R$ converge in the weak topology of $BV$ to $u \in BV(D)$ as $n \to \infty$, then also the positive parts of $v_n$ satisfy $v_n^+ \in Y_n \cap B_R$ and the functions $v_n^+$ converge in the weak topology of $BV$ to $u^+ \in BV_+(D)$ as $n \to \infty$.

Using these facts, we can see, similarly to the proof of Theorem 6, that $S_j^{(R,+)} \Gamma$-converges, with respect to the weak topology of $BV$, to $S_\infty^{(R,+)}$ as $n \to \infty$. This shows that the result of Theorem 6 holds also with the positivity constraint.
4. Computational models and algorithms.

4.1. X-ray attenuation model. In X-ray imaging the radiation source is placed on one side of an object, and the detector is placed at the opposite side. The rays pass through the object, and the attenuated signal is detected by a digital sensor (an array of almost pointlike detectors); see Figure 8 for an illustration. We model the (2D slice of the) object by a rectangle $\Omega \subset \mathbb{R}^2$ and by a nonnegative attenuation coefficient $u : \Omega \to [0, \infty)$. The attenuation coefficient $u$ gives the relative intensity loss of an X-ray traveling a small distance $ds$ at $s \in \Omega$. This leads to the following linear model:

$$g_j = -\log \left( \frac{I_j}{I_0} \right) = \int_{L_j} u(s) ds,$$

where $L_j$ is the $j$th line of the X-ray, $g_j$ is the value of the projection measurement of the $j$th source to detector line $L_j$, $I_j$ is the measured X-ray intensity, and $I_0$ is the initial intensity of the X-ray beam before entering $\Omega$.

We discretize this model by dividing the domain $\Omega$ into a lattice of pixels $\Omega_i$ and by computing the length of the path $L_j$ inside each pixel $\Omega_i$; see Figure 8. Assuming that the attenuation function $u(s)$ is constant inside each pixel $\Omega_i$, the projection measurements $g_j$ can be approximated by

$$g_j = \int_{L_j} u(s) ds \approx \sum_i f_i |\Omega_i \cap L_j|,$$

where $f_i$ is the constant value of the function $u$ in the $i$th pixel and $|\Omega_i \cap L_j|$ denotes the length of the ray $L_j$ through pixel $\Omega_i$. Arranging the set of $M$ measurements into a vector $\mathbf{g} = (g_1, g_2, \ldots, g_M) \in \mathbb{R}^M$, we obtain the model

$$\mathbf{g} = A\mathbf{f},$$
where \( f \in \mathbb{R}^N = \mathbb{R}^{n \times n} \) is the vector containing the attenuation values in the pixels and 
\( A \in \mathbb{R}^{M \times N} \) is the matrix that implements the transform from the pixel values to the projection data; see (4.2).

### 4.2. Discrete anisotropic TV.

We consider the anisotropic TV

\[
\|f\|_{TV} = \|D_H f\|_1 + \|D_V f\|_1,
\]

where the \( N \times N \) matrices \( D_H \) and \( D_V \) implement horizontal and vertical differences in the \( n \times n \) pixel image represented by vector \( f \). We use periodic boundary conditions. For example, in the case of \( 3 \times 3 \) images

\[
f = [f_{11}, f_{21}, f_{31}, f_{12}, f_{22}, f_{32}, f_{13}, f_{23}, f_{33}]^T \in \mathbb{R}^9,
\]

the matrix \( D_H \) takes the form

\[
D_H = \frac{1}{3} \begin{bmatrix}
-1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1
\end{bmatrix}.
\]

The TV-regularized X-ray imaging problem (1.3) can now be rewritten as follows:

\[
(4.3) \quad \arg \min_{f \in \mathbb{R}^N} \left\{ \frac{1}{2} \| Af - \tilde{g} \|_2^2 + \alpha (\| D_H f \|_1 + \| D_V f \|_1) \right\}.
\]

### 4.3. Quadratic programming and primal-dual interior-point (PD-IP) method.

The minimization problem of (4.3) can be reformulated as a quadratic programming problem. Denoting \( D_H f = h^+ - h^- \) and \( D_V f = v^+ - v^- \) and following [68], we get

\[
\min_{z} \frac{1}{2} z^T Q z + c^T z + d \\
\text{s.t.} \quad B z = b, \\
\quad z \geq 0,
\]

where \( Q \) is a symmetric and semi–positive definite matrix and

\[
B = \begin{bmatrix} D_H & -I & I & 0 & O \\
D_V & O & O & -I & I \end{bmatrix},
\]

where \( I \) denotes an identity matrix and \( O \) denotes a matrix of all zeros.
We seek to solve (4.4) by a PD-IP method similar to that proposed in [68]. Thus by introducing the Lagrangian function

\[ L(z, y; \mu) = \frac{1}{2} z^T Q z + c^T z + d - \mu \sum \log(z) - y^T (B z - b), \]

with \( y \) being the Lagrangian multiplier and \( \mu > 0 \), we can derive the first order necessary optimality conditions, i.e., the Karush–Kuhn–Tucker (KKT) conditions, of the problem in the following reduced form:

\[ \begin{bmatrix} -(Q + Z^{-1} \tilde{X}) & B^T & \Delta z \\ B & 0 & \Delta y \end{bmatrix} = \begin{bmatrix} p_1 - Z^{-1} \tilde{X} p_3 \\ p_2 \end{bmatrix}, \]

where \( Z = \text{diag}(z_1, z_2, \ldots, z_n), \tilde{X} = \text{diag}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n), p_1 = c + Qz - B^T y - \tilde{x}, p_2 = b - Bz, p_3 = \frac{\mu}{\tilde{x}} - z - \Delta \tilde{x} \Delta z / \tilde{x}, \) and \( \Delta \tilde{x} = Z^{-1} \tilde{X} (p_3 - \Delta z). \) For further details see [68].

The PD-IP algorithm that we use to solve our quadratic programming problem is an iterative method based on Mehrotra’s predictor-corrector approach [64]. The algorithm is explained in detail in section 4 of [68].

4.4. Projected Barzilai–Borwein method. As another method for solving (4.3) approximately, we apply a projected Barzilai–Borwein (PBB) minimization algorithm, which is a gradient-based method suitable for large-scale problems due to its relatively cheap matrix-free computations. Because of the differentiability assumption of the algorithm, we use the well-known differentiable approximation

\[ ||f||_{\beta} := \sqrt{\sum_{i=1}^{N} f_i^2 + \beta}, \quad f \in \mathbb{R}^N, \]

for the 1-norm. Here \( \beta > 0 \) is a small parameter; we use \( \beta = 10^{-2} \) in computations. The resulting functional to be minimized is given by

\[ L_{\beta}(f) := \frac{1}{2} \| Af - \tilde{g} \|_2^2 + \alpha (\| D_H f \|_\beta + \| D_V f \|_\beta), \]

whose nonnegativity constrained iterative PBB minimization is given by

\[ f^{k+1} = P \left( f^k - \lambda_k \nabla L_{\beta}(f^k) \right), \quad k = 0, \ldots, I_{\text{max}} - 1, \]

where the step size is

\[ \lambda_k = \frac{(f^k - f^{k-1})^T (f^k - f^{k-1})}{(f^k - f^{k-1})^T (\nabla L_{\beta}(f^k) - \nabla L_{\beta}(f^{k-1}))} \]

and \( P : \mathbb{R}^N \to \mathbb{R}^N \) is a projection operator to the nonnegative quadrant of \( \mathbb{R}^N \), i.e.,

\[ (P(f))_i = \begin{cases} f_i & \text{if } f_i \geq 0, \\ 0 & \text{if } f_i < 0, \end{cases} \quad i = 1, \ldots, N. \]

We choose the first step size to be \( \lambda_0 = 10^{-4} \) and the number of iterations to be \( I_{\text{max}} = 200 \).
5. Results. This section is organized as follows. Section 5.1 presents the datasets we used to test our method. Section 5.2 is devoted to testing the proposed method, whereas section 5.3 adapts the S-curve method to TV-regularized tomography and presents the associated results. In section 5.4 we use the L-curve results as a reference and present a comparison of all three parameter selection methods.

5.1. X-ray measurement data. The proposed parameter selection method was tested using three different data sets: simulated projection data from the Shepp–Logan phantom, experimental sparse angle projection data from a walnut specimen, and experimental projection data from sugar cubes.

5.1.1. Simulated data. The simulated measurement data was computed using the modified Shepp–Logan phantom. Gaussian measurement noise $N(0, \sigma^2 I)$ with standard deviation of 5% of the maximum of the computed noise-free data was added to the projection data. A more dense grid was used in the generation of the noise-free data than was used in any of the reconstructions, thus avoiding inverse crime. The simulated dataset consists of 180 projections with uniform angular sampling from angular spanning of 180°.

5.1.2. Experimental data. The projection data of the walnut and of the sugar cubes were acquired with a custom-built μCT device nanotom 180 supplied by Phoenix Xray Systems + Services GmbH (Wunstorf, Germany). The measurement set-up is presented in Figure 9. The X-ray detector is a 12-bit CMOS flat panel detector with $2304 \times 2284$ pixels of 50 μm size (Hamamatsu Photonics, Japan).

From the walnut dataset, a set of 90 projection images was acquired over a 180 degree rotation with uniform angular step of two 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, the tube current was 200 μA, and the full polychromatic beam was used for image acquisition. For this work we chose only the projections corresponding to the middle cross-section of the walnut, thus resulting in a 2D reconstruction task.

For the sugar cubes, a full-cycle 120-projection set with three degree angular step was measured. Each projection was composed of eight 500 ms exposures. The X-ray tube voltage was 80 kV, and the tube current was 225 μA. Similarly to the walnut data, only the sinogram corresponding to the middle 2D cross-section of the sugar cubes is considered in this work, and the measured sinogram was down-sampled to resolution $512 \times 120$ for computations.

We consider the following five test cases:

(i) Reconstruction using simulated noisy projection data from Shepp–Logan phantom with discretization levels $n = 64, 128, 192, 256$.

(ii) Reconstruction using the measured walnut data with $n = 128, 192, 256$.

(iii) Reconstruction using the measured walnut data with 5% additional additive random noise, i.e., using measurement $\tilde{g} + e$, where $e$ is the realization from multivariate Gaussian with standard deviation equivalent to 5% of the maximum value of the actual measurement $\tilde{g}$. The discretization levels were set to $n = 128, 192, 256$.

(iv) Reconstruction using the measured sugar cube data with discretization levels $n = 128, 256, 512$.

1The data is openly available at http://www.fips.fi/dataset.php.
Figure 9. Left: Experimental setup for collecting tomographic X-ray data of a walnut. The detector plane is on the left and the X-ray source is on the right in the picture. The walnut is attached to a computer-controlled rotator platform. Right: Two examples of the resulting projection images.

(v) Reconstruction using the measured sugar cube data with 5% additional additive random noise (similar to (iii) above) with $n = 128, 256, 512$.

5.2. Selection of $\alpha$ using the parameter choice method. In this work we consider the following levels of discretization: $n = 64, 128, 192, 256, 512$ within the five different test cases, each resulting in a 2D reconstruction on an $n \times n$ grid. TV-regularized reconstructions $\hat{f}_{\alpha,n}$ were computed with several values of regularization parameter $\alpha$. The results of cases (i), (ii), and (iii) were computed by solving the minimization problem of (4.3) using the PD-IP algorithm of [68]. In cases (iv) and (v), the PBB algorithm described in section 4.4 was used for reconstruction.

For $f \in \mathbb{R}^{n \times n}$, consider the discrete anisotropic TV norm

$$\|f\|_{TV} = \frac{1}{n} \sum_{\kappa \sim \kappa'} |f_\kappa - f_{\kappa'}|,$$

where $\kappa$ and $\kappa'$ are as in (1.3). The norms $\|\hat{f}_{\alpha,n}\|_{TV}$ of the reconstructed images were computed as a function of $\alpha$ at each discretization level $n$. Results from the simulated data (case (i)) are shown in Table 1. Table 1 also shows the relative reconstruction errors

$$\delta_{\hat{f}_{\alpha,n}} = \frac{\|f_{\text{true}} - \hat{f}_{\alpha,n}\|}{\|f_{\text{true}}\|}$$

of $\hat{f}_{\alpha,n}$ for all discretization levels as a function of $\alpha$. A few samples of the corresponding reconstructions in case (i) are shown in Figure 10.

The TV norms of the reconstructed images computed from the experimental walnut data (case (ii)) and from the walnut data corrupted with 5% of random additional noise (case (iii)) are shown in Table 2. A few samples with the walnut data are shown in Figures 11 and 12 for test cases (ii) and (iii), respectively.
The reconstructions are computed with the PD-IP algorithm from the simulated X-ray data on four different discretization levels with several values of $\alpha$ ranging in the interval $[10^{-4}, 10^6]$. The discretization levels were set to $n = 64, 128, 192, 256$. Roughly stable TV norm regimes are indicated by yellow. See Figure 10 for reconstructions.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$64^2$</th>
<th>$128^2$</th>
<th>$192^2$</th>
<th>$256^2$</th>
<th>$64^2$</th>
<th>$128^2$</th>
<th>$192^2$</th>
<th>$256^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-4}$</td>
<td>7.62</td>
<td>17.35</td>
<td>35.34</td>
<td>59.19</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>7.62</td>
<td>17.35</td>
<td>35.33</td>
<td>59.19</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>7.62</td>
<td>17.34</td>
<td>35.32</td>
<td>59.20</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>7.61</td>
<td>17.33</td>
<td>35.32</td>
<td>59.18</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^0$</td>
<td>7.61</td>
<td>17.33</td>
<td>35.22</td>
<td>58.98</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^1$</td>
<td>7.58</td>
<td>17.02</td>
<td>34.22</td>
<td>59.18</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^{1.33}$</td>
<td>7.54</td>
<td>16.67</td>
<td>33.01</td>
<td>54.66</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^{1.66}$</td>
<td>7.45</td>
<td>15.94</td>
<td>30.61</td>
<td>49.84</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^2$</td>
<td>7.28</td>
<td>14.51</td>
<td>26.04</td>
<td>40.52</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$10^{2.33}$</td>
<td>6.96</td>
<td>12.14</td>
<td>19.09</td>
<td>27.16</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>$10^{2.66}$</td>
<td>6.45</td>
<td>9.10</td>
<td>11.26</td>
<td>13.55</td>
<td>0.19</td>
<td>0.09</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>$10^3$</td>
<td>5.82</td>
<td>6.67</td>
<td>6.83</td>
<td>7.00</td>
<td>0.19</td>
<td>0.09</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>$10^{3.33}$</td>
<td>5.35</td>
<td>5.78</td>
<td>5.81</td>
<td>5.84</td>
<td>0.19</td>
<td>0.10</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>$10^{3.66}$</td>
<td>4.91</td>
<td>5.17</td>
<td>5.25</td>
<td>5.28</td>
<td>0.19</td>
<td>0.11</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>$10^4$</td>
<td>4.16</td>
<td>4.40</td>
<td>4.46</td>
<td>4.48</td>
<td>0.19</td>
<td>0.14</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>$10^{4.33}$</td>
<td>3.23</td>
<td>3.37</td>
<td>3.41</td>
<td>3.42</td>
<td>0.24</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>$10^{4.66}$</td>
<td>2.22</td>
<td>2.27</td>
<td>2.28</td>
<td>2.28</td>
<td>0.31</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>$10^5$</td>
<td>1.30</td>
<td>1.31</td>
<td>1.32</td>
<td>1.32</td>
<td>0.37</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>$10^6$</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
</tr>
</tbody>
</table>

For the experimental sugar cube data (cases (iv) and (v)), the TV norms of the reconstructions are presented in Table 3, and the reconstructions computed with the parameters $\alpha$ chosen by the proposed parameter selection strategy are shown in Figure 13.

The reconstructions in Figures 10, 11, and 12 are selected as follows: On the top row the value of the regularization parameter is too small, producing a “noisy” reconstruction, and on the bottom row the regularization parameter is too big, producing an “overregularized” reconstruction. In the middle row the value of the regularization parameter is selected as the smallest, keeping the TV norms of reconstructions approximately independent of resolution. See Tables 1 and 2.

5.3. Selection of $\alpha$ using the S-curve method. The S-curve method was originally introduced in [56, 47] within a Bayesian inversion framework. In this work we apply the S-curve method for TV-regularized tomography. The S-curve results are presented here only for test cases (ii) and (iii). For further studies of the S-curve method applied to simulated and measured datasets, see [56, 47].

The S-curve method was applied as follows. For test case (ii), we computed TV-regularized
Table 2

TV norms of 2D reconstructions \( \hat{f}_{\alpha,n} \) computed from the walnut data on three different discretization levels with several values of \( \alpha \) ranging in the interval \([10^{-6}, 10^6]\). The discretization levels were set to \( n = 128, 192, \) and 256. The results are computed from projection data with 90 projection angles. Data is corrupted with additional additive random noise (5\%). The choice of \( \alpha \) is the smallest possible leading to stable TV norms: \( \alpha = 1 \) in the case of low noise, and \( \alpha = 10 \) in the case of 5\% added noise. Roughly stable TV norm regimes are indicated by yellow. See Figures 11 and 12.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>Low noise</th>
<th>5% noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128(^2)</td>
<td>192(^2)</td>
</tr>
<tr>
<td>( 10^{-4} )</td>
<td>1.51 2.29 3.64</td>
<td>2.42 5.05 8.71</td>
</tr>
<tr>
<td>( 10^{-3} )</td>
<td>1.51 2.29 3.46</td>
<td>2.43 5.05 8.59</td>
</tr>
<tr>
<td>( 10^{-2} )</td>
<td>1.50 2.23 2.97</td>
<td>2.42 5.01 8.59</td>
</tr>
<tr>
<td>( 10^{-1} )</td>
<td>1.43 1.85 1.93</td>
<td>2.37 4.83 8.16</td>
</tr>
<tr>
<td>( 10^0 )</td>
<td>1.08 1.11 1.11</td>
<td>1.99 3.50 5.12</td>
</tr>
<tr>
<td>( 10^1 )</td>
<td>0.78 0.78 0.77</td>
<td>0.86 0.86 0.88</td>
</tr>
<tr>
<td>( 10^2 )</td>
<td>0.48 0.48 0.48</td>
<td>0.48 0.48 0.48</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>0.12 0.12 0.12</td>
<td>0.12 0.12 0.12</td>
</tr>
<tr>
<td>( 10^4 )</td>
<td>0.04 0.04 0.04</td>
<td>0.04 0.04 0.04</td>
</tr>
<tr>
<td>( 10^5 )</td>
<td>0 0 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>( 10^6 )</td>
<td>0 0 0</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>

Table 3

TV norms of 2D reconstructions \( \hat{f}_{\alpha,n} \) computed with projected Barzilai–Borwein from the sugar cube X-ray data on three different discretization levels with several values of \( \alpha \) ranging in the interval \([10^{-4}, 10^3]\). The discretization levels were set to \( n = 128, 256, \) and 512. The results were computed from projection data with 120 equiangular fan-beam projections with and without additional Gaussian noise (5\%). The choice of \( \alpha \) is the smallest possible leading to stable TV norms: \( \alpha = 10 \) in the case of low noise, and \( \alpha = 100 \) in the case of 5\% added noise. Roughly stable TV norm regimes are indicated by yellow. See Figure 13.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>Low noise</th>
<th>5% noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128(^2)</td>
<td>256(^2)</td>
</tr>
<tr>
<td>( 10^{-4} )</td>
<td>1.36 5.41 18.3</td>
<td>2.96 12.0 27.8</td>
</tr>
<tr>
<td>( 10^{-3} )</td>
<td>1.38 5.96 17.7</td>
<td>2.95 10.7 24.3</td>
</tr>
<tr>
<td>( 10^{-2} )</td>
<td>1.36 4.84 13.5</td>
<td>2.92 10.9 25.4</td>
</tr>
<tr>
<td>( 10^{-1} )</td>
<td>1.26 2.48 3.95</td>
<td>2.64 7.26 9.98</td>
</tr>
<tr>
<td>( 10^0 )</td>
<td>0.95 1.11 1.28</td>
<td>1.66 2.41 2.90</td>
</tr>
<tr>
<td>( 10^1 )</td>
<td>0.63 0.64 0.65</td>
<td>0.82 0.91 0.98</td>
</tr>
<tr>
<td>( 10^2 )</td>
<td>0.45 0.45 0.45</td>
<td>0.48 0.49 0.50</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>0.34 0.34 0.36</td>
<td>0.34 0.34 0.34</td>
</tr>
</tbody>
</table>

reconstructions \( \hat{f}_{\alpha,n} \) with fixed \( n \) for 14 different values of \( \alpha \) ranging in the interval \([10^{-6}, 10^6]\) and computed the TV norms

\[
S(\alpha) := \| \hat{f}_{\alpha,n} \|_{TV}
\]
for the reconstructions. Then the data \( \{ \alpha, S(\alpha) \} \) were interpolated to get the S-curve. The S-curve was computed using the same three different discretization levels \( n = 128, 192, 256 \) that were used in section 5.2 with the walnut data.

In the test case (iii) we computed reconstructions with 14 different values of \( \alpha \) ranging in the interval \([10^{-4}, 10^8]\) and applied the S-curve interpolation procedure similarly as in case (ii).

As in [47], the a priori value for the sparsity level \( \hat{S} \) was estimated using three digital photographs of split walnuts. The photographs of the three walnuts are shown in Figure 14. The walnut used to measure the X-ray data is not included in the photos. In reality, as is the case also in this work, the a priori information, which we use to estimate the value of \( \hat{S} \), comes from a different modality (e.g., anatomical atlases) than the one we are considering.
Figure 11. TV-regularized reconstructions $\hat{f}_{\alpha,n}$ from the projection data of the walnut computed with different values of regularization parameter $\alpha$. The values of $\alpha$ are indicated on the right. Discretization levels $n = 128, 192, 256$ are indicated on the top. See Table 2.

(Here digital photographs versus X-ray attenuation function). Therefore, in order to compute the $\hat{S}$ for the TV-regularized X-ray tomography, each of the digital photographs $f_p$ was scaled such that the norm of the computed X-ray projection data of the photograph is the same as the norm of the measured projection data $\tilde{g}$. This was obtained by

\[
\tilde{f}_p = \frac{\| \tilde{g} \|}{\| A f_p \|} f_p.
\]

This scaling of the photographs is essential since the TVs of the photographs are not directly comparable to the TV of the X-ray attenuation function.
5.4. Comparison of the results. As a reference, results with the L-curve method were computed with both of the walnut data test cases (ii) and (iii). The L-curve plots and corresponding reconstructions are shown here only for the discretization level $n = 128$, but the results for other discretization levels were similar. Figures 17 and 18 present the L-curve plots and the corresponding reconstructions for the test cases (ii) and (iii), respectively.

A comparison of the results obtained with all of the three methods, i.e., TV parameter choice $\alpha$, number of discretization levels $n = 128, 192, 256$, and the measured projection data corrupted with 5% additional noise, are presented in Figure 12. The S-curve plots and the resulting reconstructions for each discretization levels are presented in Figure 15 for the walnut data. Corresponding results with 5% additional noise are presented in Figure 16.

Figure 12. TV-regularized reconstructions $\hat{f}_{\alpha,n}$ from the measured projection data corrupted with 5% additional noise computed with different values of regularization parameter $\alpha$. The values of $\alpha$ are denoted on the right. Discretization levels $n = 128, 192, 256$ are indicated on the top. See Table 2.
Figure 13. TV-regularized reconstructions computed from the measured projection data of 10 sugar cubes with projected Barzilai–Borwein and smoothed TV. Top: Filtered backprojection reconstruction (for reference). Middle: TV reconstructions computed from the X-ray data with no additional noise and with three different discretization levels, $n = 128$, 256, and 512. The regularization parameter is $\alpha = 10$. Bottom: TV reconstructions computed from X-ray data with 5% additional Gaussian noise. Here $\alpha = 100$. See Table 3.

6. Discussion and conclusions. The results listed in Tables 1, 2, and 3 provide numerical evidence supporting the proposed parameter choice method. It seems that the new multiresolution-based parameter choice method is useful and robust. The fact that we also use measured X-ray projection data lends credibility to these initial findings. From a practical point of view, a very nice feature of the approach is that the parameter selection does not
Figure 14. Photographs of three walnuts split in half. These images were used to provide a priori information about the expected TV of the target. Of course, these are optical photographs and thus physically very different objects from the reconstructions (tomographic slices representing X-ray attenuation coefficient). However, we assume that the TVs of the photographs, when scaled by (5.1), are comparable to the TV of the X-ray attenuation coefficient.

require (quantitative) prior information about the measurement noise (e.g., expected norm of the noise) or the unknowns (e.g., expected value of the TV norm) as input variables for the parameter selection.

The results of Table 1 show that including a coarser discretization level in the computations of \( \alpha \), i.e., level \( n = 64 \), leads to the choice \( \alpha = 10^{3.66} \). While this choice of \( \alpha \) produces the smallest reconstruction error for estimate \( \hat{f}_{\alpha,n} \), the corresponding errors for \( n = 128, 192, 256 \) are not the smallest. A smaller error for \( n = 128, 192, 256 \) is obtained with \( \alpha = 10^3 \). However, reconstructions with \( \alpha = 10^3 \) and \( \alpha = 10^{3.66} \) both yield reasonable estimates for \( \hat{f}_{\alpha,n} \) for all resolutions \( n \) considered. Thus it seems that the proposed method works acceptably also at low resolutions.

The precise mechanism behind the proposed method is still unclear. We presented a mathematical proof that the reconstructions converge for any choice of \( \alpha \) as the resolution is refined. Therefore, it remains an open question whether the unstable norms in Tables 1, 2, and 3 would converge with higher resolutions, or whether those numbers are inaccurate, reflecting the numerical instability arising from the ill-posedness of the underlying tomographic problem. There may be a connection to [82], where the scale of image features is defined by measuring their tendency to change under TV regularization. In any case, in these examples at least, the proposed method automatically adapts to the relevant resolution and actual noise level.

The S-curve method, introduced in [56, 47], was found to work robustly and reliably for the present datasets. Curiously, the S-curve method produces parameter values of the same order of magnitude as the new resolution-based method. However, a priori information about the TV of the unknown function is needed for applying the S-curve method, while the proposed multiresolution-based method does not need anything other than the data.

The proposed method has the drawback that many reconstructions need to be computed with various choices of resolution \( n \) and regularization parameter \( \alpha \). This can be computationally demanding. In this work we solve (1.3) using a PD-IP method and a PBB method. The Barzilai–Borwein method is a gradient-based method which is relatively cheap due to matrix-free computations. For the Barzilai–Borwein method, the absolute values of the 1-
norms are approximated due to the differentiability requirement of the method. By dividing the unknown into nonnegative and nonpositive parts, we can solve (1.3) without approximations. However, this division has the drawback of increasing the number of unknowns in the problem. The new, bigger problem is, however, sparse and can be solved efficiently using PD-IP methods as demonstrated in [41], for example. With the PD-IP method that we use, the most demanding part is the computation of (4.6). However, (4.6) can be solved itera-

Figure 15. S-curve method applied to the measured projection data with no additional noise (case (i)). Left column: Plots of the S-curves used to determine the values of $\alpha$. Right column: Reconstruction computed with the selected $\alpha$. 
Figure 16. S-curve method applied to the measured projection data corrupted with 5% additional noise (case (ii)). Left column: Plots of the S-curves which were used to determine the values of $\alpha$. Right column: Reconstruction computed using the selected $\alpha$.

- For $n = 128$, $\alpha = 14$, $\hat{S} = 0.74$.
- For $n = 192$, $\alpha = 28$, $\hat{S} = 0.75$.
- For $n = 256$, $\alpha = 37$, $\hat{S} = 0.76$.

The $S$-curve method is one of the techniques used in the parameter choice problem. Effectively using a suitable preconditioned iterative scheme, such as the preconditioned conjugate gradient method, a generalized minimal residual method, or a preconditioned quasi-minimal residual method, for example. An efficient implementation of a suitable preconditioner will enable the acceleration of the convergence of a selected iterative solver. All computations can also be made matrix-free, making the method suitable also for large-scale problems. One of the advantages of PD-IP methods is the natural and effective enforcement of nonnegativity.
Figure 17. L-curve method applied to walnut data; no additional noise. Left: L-curve plot. Right: Reconstruction computed using the selected $\alpha = 1358$. The discretization level is $n = 128$.

Figure 18. L-curve method applied to walnut data with 5% additional noise. Left: L-curve plot. Right: Reconstruction computed using the selected $\alpha = 10.4$. The discretization level is $n = 128$.

Figure 19. The final results from the walnut data with low noise. Reconstruction from left to right: TV parameter choice $\alpha_{TV} = 1$, S-curve method $\alpha_{S\text{-curve}} = 3.00$, and L-curve method $\alpha_{L\text{-curve}} = 1358$, respectively. The discretization level is $n = 128$. 
Figure 20. The final results from the walnut data with 5\% additional noise. Reconstruction from left to right: TV parameter choice $\alpha_{TV} = 10$, S-curve method $\alpha_{S\text{curve}} = 14.31$, and L-curve method $\alpha_{L\text{curve}} = 10.4$, respectively. The discretization level is $n = 128$.

Also, with primal-dual methods the evaluation of the duality gap provides a nonheuristic check of convergence. In this work we used a preconditioned quasi-minimal residual method for solving (4.6). All the PD-IP computations of this work were performed using MATLAB version R2014.

We note that, in addition to the methods discussed in the introduction, (1.3) could also be solved using proximal splitting methods [27], for example. However, these methods introduce additional variables into the optimization problem, and without acceleration the number of iterations is considerable. With proper preconditioning and/or acceleration schemes, proximal splitting methods provide feasible algorithms for solving convex, nonsmooth imaging problems; see [18, 1, 79, 11, 10, 59], for example.

We stress that this work is just an initial feasibility study for the new parameter choice method. More comprehensive testing and comparisons to other methods are still needed to assess the properties of the new approach. Also, it would be interesting to see how the method works with generalizations of TV regularization, such as [9, 81].

Appendix A. A note on $BV$ norms. What would happen if one used the $BV$ norm

$$
\|u\|_{L^1(D)} + V_p(u) := \|u\|_{L^1(D)} + \int_D \left( \left| \frac{\partial u(x)}{\partial x_1} \right|^p + \left| \frac{\partial u(x)}{\partial x_2} \right|^p \right)^{1/p} dm(x)
$$

with $p > 1$ in the analysis of section 3? The norm (2.2) used there corresponds to the choice $p = 1$. First, taking $p > 1$ would only give an inequality in (2.6),

$$
\lim_{j \to \infty} V_p(v_j) \geq V_p(u),
$$

and the proof would not work.

Let us point out a specific example using $p = 2$. Consider the characteristic function $\chi = \chi_{\{|x|<1\}}$ of the unit disc, whose BV norm is $2\pi$: the length of the boundary (which is $2\pi$) times the jump (which is 1). Now approximate $\chi$ using functions equal to one inside pixel-based set approximations to the unit disc; see Figure 21. The total length of the boundary of each such set is 8. Thus the BV norms of the approximations converge to 8 instead of $2\pi$. 
Figure 21. Pixel-based piecewise linear approximations to the characteristic function of the unit disc. Here white denotes zero value, and gray denotes value one. The TV norm of all of these functions is exactly 8 (equal to the length of the boundary).

Now using the BV norm with \( p = 1 \) leads to the (Manhattan) length of the unit circle to be 8, ensuring the appropriate convergence.

REFERENCES


TV PARAMETER CHOICE


