Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization

Emil Y. Sidky and **Xiaochuan Pan**

Department of Radiology MC-2026, University of Chicago, 5841 S. Maryland Ave., Chicago IL, 60637

Abstract

An iterative algorithm, based on recent work in compressive sensing, is developed for volume image reconstruction from a circular cone-beam scan. The algorithm minimizes the total-variation (TV) of the image subject to the constraint that the estimated projection data is within a specified tolerance of the available data and that the values of the volume image are non-negative. The constraints are enforced by use of projection onto convex sets (POCS) and the TV objective is minimized by steepest descent with an adaptive step-size. The algorithm is referred to as adaptive-steepest-descent-POCS (ASD-POCS). It appears to be robust against cone-beam artifacts, and may be particularly useful when the angular range is limited or when the angular sampling rate is low. The ASD-POCS algorithm is tested with the Defrise disk and jaw computerized phantoms. Some comparisons are performed with the POCS and expectation-maximization (EM) algorithms. Although the algorithm is presented in the context of circular cone-beam image reconstruction, it can also be applied to scanning geometries involving other x-ray source trajectories.

1. Introduction

The circular cone-beam scanning configuration is widely used in many computed tomography (CT) applications, because it is a technically simple CT configuration to implement and in many cases the image artifacts caused by the incompleteness of the circular trajectory may not be of practical concern. Circular cone-beam CT is employed for dedicated breast CT (Boone, Nelson, Lindfors & Seibert 2001, Chen & Ning 2002, Boone, Shah & Nelson 2004), micro-CT (Ritman 2002, Lee, Kim, Chun, Cho, Lee & Cho 2003, Ford, Thornton & Holdsworth 2003), image guided radiation therapy (Ford, Chang, Mueller, Sidhu, Todor, Mageras, Yorke, Ling & Amols 2002, Jaffray, Siewerdsen, Wong & Martinez 2002, Letourneau, Wong, Oldham, Gulam, Watt, Jaffray, Siewerdsen & Martinez 2005), and C-arm CT systems employed for surgery (Rafferty, Siewerdsen, Chan, Moseley, Daly, Jaffray & Irish 2005, Daly, Siewerdsen, Moseley, Jaffray & Irish 2006). Although the circular source trajectory is commonly used, it is known that there exists no theoretically exact solution that can stably reconstruct from projection data acquired from this trajectory. As a result, the majority of one-shot image reconstruction algorithms, such as Feldkamp Davis and Kress (FDK) (Feldkamp, Davis & Kress 1984) and its derivatives, for this configuration are approximate. For many applications, particularly for systems with small cone-angles, algorithms such as FDK perform adequately when there are a large number of projection views. But in general it is advantageous to have image reconstruction algorithms that can also yield accurate images for large cone-angles. There can also be situations where the scanning arc has a limited angular range or the angular sampling along the source trajectory is low. For some configurations with limited angular range but dense angular sampling, there is an approximate one-shot algorithm that can

E-mail: sidky@uchicago.edu,xpan@uchicago.edu

provide region of interest images with relatively few artifacts (Yu, Zou, Sidky, Pelizarri, Munro & Pan 2006). But another interesting general approach to counter the effects of limited angular range or angular sampling, or large cone-angles is to develop iterative image reconstruction algorithms.

Many iterative methods view the image reconstruction problem as a discrete linear system, where the projection data are a weighted sum over the image voxels. Though computationally more intensive, iterative methods may have some advantages over one-shot algorithms derived from analytic inversion formulas. When the projection data are sufficient for the analytic inversion formula the corresponding discrete form of the reconstruction algorithm is usually effective. When the projection data, however, are incomplete, such algorithms can implicitly make very unrealistic assumptions about the "missing" data, resulting in prominent artifacts in the reconstructed images. Iterative algorithms, generally make milder assumptions on missing data and allow for the incorporation of image constraints such as positivity, bounds on image roughness and maximum value. Much work has been done on developing iterative methods for CT image reconstruction using both linear and non-linear system models(Lange & Carson 1984, Sauer & Bouman 1993, Manglos, Gagne, Krol, Thomas & Narayanaswamy 1995, Elbakri & Fessler 2003, Kole 2005, Zbijewski, Defrise, Viergever & Beekman 2007). For circular cone-beam CT, in particular, research has focused fast volume reconstruction. The algorithm acceleration is achieved either by reducing the size of the system model or by developing fast cone-beam projection and back-projection algorithms. Methods for reducing the system-model size include seeking expansion functions that are potentially more efficient than voxels (Lewitt 1990, Matej & Lewitt 1996, Carvalho & Herman 2007), or investigating adaptive gridding (Benson & Gregor 2006). Recently, there has been much effort developing accelerated projectors and back-projectors through streamlined algorithms and special purpose hardware (Xu & Mueller 2007, Kachelriess, Knaup & Bockenbach 2007, Kole & Beekman 2006, Sharp, Kandasamy, Singh & Folkert 2007). For the most part the actual iterative algorithms themselves are well-established. In this article, we develop an iterative algorithm for image reconstruction based on constrained, total-variation (TV) minimization (Candes, Romberg & Tao 2006*a*, Candes, Romberg & Tao 2006*b*, Sidky, Kao & Pan 2006*a*). The aim here is not for algorithm efficiency. Instead, we seek to develop an algorithm that can solve the constrained, TV-minimization optimization-problem, and to investigate the algorithm's ability to reconstruct images for circular cone-beam CT with minimal artifacts due to large cone-angles or limited angular sampling.

The idea of constrained TV-minimization originates in the field of compressive sensing from the work by Candes *et al.* on exact recovery of an image from sparse samples of its discrete Fourier transform (DFT) (Candes, Romberg & Tao 2006*a*). The exact recovery depends on the fact that there exists some representation of the image for which the corresponding coefficients are sparse. For example, if the image is known to have only N non-zero voxel values of unknown location, $2N$ measurements of the image's DFT should suffice to completely determine the image. This fact can potentially allow for enormous reduction in the required number of samples: if it is known, for example, that only 1,000 voxels of a 100^3 image are non-zero, only 2,000 DFT samples are needed as opposed to the full 100^3 needed if nothing is known about the image. It turns out that the image can be found by solving the convex optimization problem of minimizing the ℓ_1-norm of the image, subject to the constraint that the image's DFT matches the measured DFT values; although this optimization problem generally requires a little more than $2N$ samples. Other non-convex optimization formulations are being investigated that can potentially improve on the ℓ_1-norm minimization in terms of number of required samples (Chartrand 2007, Sidky, Chartrand & Pan 2007). Sparsity of images in terms of voxels is generally not a widely applicable assumption on the underlying image function; but it is often the case that images have sparse gradient-magnitude images (GMI). For example, the x-ray attenuation coefficient often varies mildly within organs, and

large image variations are usually confined to the borders of internal tissue structures. A sparse GMI may also be a good approximation in industrial or security scanning applications. For such images, minimizing the image TV, which is the ℓ_1-norm of the GMI, subject to the data constraint can yield the accurate images from sparse DFT samples. Though developed for sparse DFT inversion, it was pointed out that constrained TV-minimization could be applied to general linear systems (Candes, Romberg & Tao 2006*b*), but new algorithms need to be developed for the case of non-orthogonal linear systems such as the present case of cone-beam projection. We have conducted a preliminary study of constrained, TV-minimization algorithms to inversion of the x-ray transform from limited numbers and ranges of scanning views (Sidky, Kao & Pan 2006*a*, Sidky, Kao & Pan 2006*b*, Sidky & Pan 2006, Sidky, Chartrand & Pan 2007, Sidky, Reiser, Nishikawa, Pan, Chartrand, Kopans & Moore 2008).

The image TV has been used as a penalty term in iterative, image reconstruction algorithms (Vogel & Oman 1996, Panin, Zeng & Gullberg 1999, Persson, Bone & Elmqvist 2001, Song, Liu, Johnson & Badea 2007, Chen, Tang & Leng 2008). Adding such a term to the data-fidelity-objective function tends to smooth out noise in the image while preserving edges within the image. In the traditional approach of using a regularizer such as TV, there is a trade-off between data fidelity and image regularity; lower image TV implies worse data fidelity. Such a formulation is sensible for fully sampled tomographic systems where there is a unique image (or very limited set of images) that minimize the data-fidelity-objective function. For the present work, we are interested in tomographic image reconstruction where the projection data are incomplete. Because of the incompleteness there will in general be no unique minimizer of the data-fidelity-objective function, and TV is used to select a unique image out of the set of possible images that agree with the available data. For this situation, we find it more natural to use a constrained optimization approach, where image TV is the objective function and data fidelity is a constraint. In this case, image TV may be lowered while data fidelity is maintained. Furthermore, our approach is easily extended to include other constraints such as image positivity or finite upper bound.

The idea of sparse image recovery from sparse projection data has also been reported in the literature. Li *et al*. (Li, Yang & Kudo 2002) investigated an algorithm based on ℓ_1-norm minimization for image reconstruction of blood vessels using contrast agent. In this case, because the underlying image function is sparse in terms of voxels, they were able to achieve accurate reconstructions from as low as 15 projection views. In many practical applications, however, the underlying image functions do not satisfy the sparseness condition assumed by this algorithm. The idea that medical images are often approximately piecewise constant has been recognized before and incorporated into an optimization problem from which new iterative, image reconstruction algorithms were derived (Delaney & Bresler 1998). Although the motivations for that work and the present article are similar, the optimization formulations and reconstruction algorithms are different.

In this article we discuss in detail the sampling issues for circular cone-beam CT and present an image reconstruction algorithm that solves the corresponding constrained, TV-minimization problem in Sec. 2. In Sec. 3 we demonstrate the algorithm with reconstructions of the Defrise disk and FORBILD jaw computerized phantoms, including results with simulated projection data without and with noise. Finally, in Sec. 4 we make remarks about how the present algorithm may affect the design of scanning protocols, and possible generalizations to other imaging applications and modalities.

2. Theory

In this section we develop a constrained, TV-minimization algorithm for image reconstruction in circular cone-beam CT. We first describe the model for the circular cone-beam scan, and

discuss how the corresponding data sampling impacts image reconstruction. An optimization problem is then formulated, the solution of which is the minimum-TV image that satisfies the data and positivity constraints. We derive necessary conditions for the optimality of the solution that can be used as a check on the results from reconstruction algorithm. Finally, we present the algorithm itself, which employs projection onto convex sets (POCS) to enforce the image constraints and steepest descent with an adaptive step-size to reduce the image TV. This adaptive steepest-descent-POCS (ASD-POCS) algorithm yields an approximate solution to the constrained, TV-minimization problem.

When the underlying image function has a sparse GMI, the ASD-POCS algorithm can provide accurate image reconstructions for incomplete sets of projection data. Because the algorithm is based on the standard POCS algorithm, the performance of ASD-POCS in terms of image-reconstruction accuracy should be similar to POCS when the underlying image function does not have a sparse GMI. For the studies presented here, we employ test phantoms that are piece-wise constant (sparse GMI), but we have found that TV minimization yields accurate images from incomplete data for small deviations from having a sparse GMI in the context of 2D fan-beam imaging (Sidky, Kao & Pan 2006*a*).

2.1. Data model and system discretization

Under ideal, noiseless conditions, the CT measurement can be converted to a form that is closely modeled by a line integral through the continuous object function:

$$g\left(\vec{r}_0(\lambda),\hat{\theta}\right)=\int_0^\infty dt f\left(\vec{r}_0(\lambda)+t\hat{\theta}\right),$$

(1)

where $f(\vec{r})$ represents the object function, the x-ray attenuation coefficient; and the data function $g\left(\vec{r}_0,\hat{\theta}\right)$ is the line integral through the object in the direction of the unit vector $\hat{\theta}$ from a source location $\vec{r}_0(\lambda)$. The parametrized x-ray source trajectory considered here is a circle, but other trajectories are also possible:

$$\vec{r}_0(\lambda)=R(\cos\lambda,\sin\lambda,0),$$

(2)

where $\lambda \in [0, 2\pi)$. The parameter R is the radius of the source trajectory, and the source location is specified by the angular parameter λ. The center of rotation is taken to be the origin of the spatial coordinate system, and the z-axis is perpendicular to the plane of the trajectory. For the model considered here, the detector is taken to be a flat panel array with bin locations:

$$\vec{d}_0(\lambda,u,v)=(R-S)(\cos\lambda,\sin\lambda,0)+u(-\sin\lambda,\cos\lambda,0)+v(0,0,1),$$

(3)

where S is the source-detector center distance, and parameters u and v identify a particular bin location. Relating the detector model to the data function in Eq. (1), the direction $\hat{\theta}$ along which a ray passes through the subject is given by

$$\hat{\theta}=\frac{\vec{d}_0(\lambda,u,v)-\vec{r}_0(\lambda)}{|\vec{d}_0(\lambda,u,v)-\vec{r}_0(\lambda)|}.$$

(4)

The goal of image reconstruction is to find $f(\vec{r})$ from knowledge of $g\left(\vec{r}_0(\lambda),\hat{\theta}\right)$. The discussion on the scanning configuration model up until now has been in the limit of continuous

measurement. For any actual CT system, the sampling will naturally be in terms of discrete projection data.

The imaging model, Eq. (1), can be approximated by the following discrete linear system:

$$\tilde{g} = M \vec{f}. \tag{5}$$

The vector $\tilde{g}$ has length N_d, the number of measured projection rays; the vector $\vec{f}$ has length N_{im}, the number of expansion elements used in representing the object function $f(\vec{r})$; and the system matrix M is a discrete model for the integration in Eq. (1). The rays themselves are identified by a particular scheme of discretizing the independent variables λ, u, and v of the projection data space and substituting those values into Eqs. (2) and (3). The particular entries of M depend on the expansion set of the image and the model for the ray integration. For this work, the image expansion functions are voxels, and M is a ray-driven projection operator. Under ideal conditions the image reconstruction problem is tantamount to inverting the linear system Eq. (5), finding $\vec{f}$ from a data set $\tilde{g}$. Even under ideal conditions, solving this linear system can be complicated by ill-conditionedness of the system matrix M.

The system matrix ill-conditionedness can originate from two sources: insufficient coverage in the scanning configuration, e.g. projection data truncation or incomplete x-ray source trajectories; or under-sampling, e.g. too few projection views or low-resolution detectors. For the circular cone-beam scanning configurations considered here, both of these causes for ill-conditionedness are considered. The circular scanning trajectory in the continuous sampling limit does not allow for stable one-shot reconstruction of the volume image, within which most of the points do not satisfy Tuy's condition.

2.2. Formulation of the optimization problem and constraints on the solution

Having approximated image reconstruction for circular cone-beam CT to inversion of a finite linear system, we specify the optimization problem that we wish to solve. Constrained TV-minimization yields the discrete image $\vec{f}$ that minimizes its TV

$$\vec{f}^* = \operatorname{argmin} \| \vec{f} \|_{TV} \tag{6}$$

subject to the inequality constraints: (A) data fidelity

$$| M \vec{f} - \tilde{g}_{\text{data}} | \leq \epsilon, \tag{7}$$

and (B) non-negativity

$$\vec{f} \geq 0. \tag{8}$$

The data-consistency constraint is formulated as an inequality constraint, because there will in general be multiple sources of data inconsistency, including a simplified data model, noise, and x-ray scatter, which will make it impossible to always find an image that is perfectly consistent with the data. As a result we only require that the image yields projection data that are within a given ℓ_2 distance ϵ of the actual projection data. We stress that the parameter ϵ is affiected only indirectly by the noise model, and that ϵ accommodates all sources of data inconsistency. Selection of the values of ϵ depends on system configuration parameters and

data quality. The constrained, optimization problem is non-linear because the TV-objective function and the ellipsoidal constraint on the data are non-linear, but it is convex because the objective and constraints are all convex.

Before turning our attention to an algorithm for image reconstruction, we discuss the relationship between the above constrained optimization and the more traditional approach of unconstrained minimization of an objective function that combines data fidelity and regularization terms. Ignoring positivity constraints for the moment, the traditional approach involves solving an unconstrained, convex optimization problem:

$$\vec{f}^* = \operatorname{argmin} \left| M\vec{f} - \tilde{\vec{g}}_{\text{data}} \right|^2 + \tau \| \vec{f} \|_{TV}. \tag{9}$$

This problem can be solved by standard techniques such as Conjugate Gradients or Steepest Descent. This formulation makes sense when the data are complete, where there is a unique image that minimizes the data fidelity term. Adding the regularizing penalty term, sacrifices data fidelity for image regularity. For incomplete data where a number of images have equivalent data fidelity, the role of the TV-norm takes on the additional role of selecting the image with sparsest GMI for the same data fidelity. The two roles of image regularization and sparse GMI selection are more naturally separated in the constrained optimization problem, Eqs. (6) and (7). It is true, however, that there is some equivalence between the two approaches. For any ε strictly greater than zero, there is a τ such that the constrained optimization problem, Eqs. (6) and (7), is equivalent to the unconstrained problem, Eq. (9). However, there is no equivalent value of τ for the case of ε = 0, when Eq. (7) becomes an equality constraint, which is an important case for sparse image recovery. When the projection data are incomplete, a large set of images can be consistent with the available data. In the constrained optimization view, the data fidelity is fixed by the constraint Eq. (7), and the minimum TV-norm is used to select a unique image out of the set of feasible images. In the case of perfectly consistent data, which we have studied in a 2D setting in Sidky, Kao & Pan (2006a), setting ε to zero may not specify a unique image, and the TV-norm is minimized to arrive at a unique solution. The ε = 0 case can be approached for unconstrained optimization by letting τ approach zero, but in practice the algorithms for solving Eq. (9) lose efficiency in this limit. Another advantage of the constrained optimization formulation is that common physical constraints such as image positivity can be incorporated easily.

2.3. Algorithm selection and necessary conditions for the image solution

The next question is how to solve the above constrained, optimization problem. Even though the problem is non-linear, it can be reformulated into a second-order cone program (SOCP) for which there are efficient interior point algorithms for achieving very accurate solutions in "polynomial time" (Boyd & Vanderberghe 2004, Alizadeh & Goldfarb 2003). The SOCP formulation is, however, impractical for cone-beam CT because of the enormous size of optimization problems of practical interest. The known algorithms for efficiently solving SOCPs are not "row-action", requiring simultaneous processing of the whole system matrix for large linear systems at each iteration. For the current case of cone-beam CT image reconstruction the size of the intermediate linear systems are on the order of the size of the cone-beam-projection system-matrix. Even for the smallest image reconstruction problem considered in the Results, the size of system matrix is approximately $10^6 \times 10^6$. Moreover, the SOCP formulation introduces additional auxiliary variables, which can increase the problem size by a factor of two or more.

Another interesting approach is to cast the optimization problem as a variational inequality (VI) (Nagurney & Zhang 1996). Most algorithms for solving VIs, however, require a projection

operator that takes an arbitrary image and yields the closest (in the ℓ_2 sense) image that satisfies the constraint equations. Currently, no such projection operator exists that can project an image onto the nearest image satisfying the constraints Eqs. (7) and (8). An exception, however, is the hybrid steepest descent (HSD) algorithm developed by Yamada for solving VIs (Yamada 2001, Xu & Kim 2003). This algorithm generalizes previous algorithms by allowing the projection to be an iterated operator. POCS is an efficient iterative algorithm for finding images that respect the convex constraints in Eqs. (7) and (8); unfortunately, POCS does not, however, find the *nearest* image satisfying these constraints. And HSD with POCS as the projector may not satisfy the convergence proofs of HSD. As a result, we give the present algorithm, ASD-POCS, a different name so as to not give the impression that there is a proof that it finds the solution of Eqs. (6)-(8).

The ASD-POCS algorithm for solving the constrained, optimization problem uses gradient descent to minimize the objective function combined with POCS to enforce the constraints. Unlike the HSD the proposed ASD-POCS algorithm is adaptive, and therefore convergence proofs for HSD do not apply. The algorithm is described in detail in Sec. 2.4 after the optimality conditions are derived.

The ASD-POCS algorithm is constructed so that the convergence properties of POCS are retained. As a result, images obtained from ASD-POCS obey the constraint Eqs. (7) and (8) for appropriate choices of the algorithm parameters. Within the set of images obeying these constraints one cannot guarantee that ASD-POCS will always find the minimum TV-image. Instead, we can derive conditions for a solution of the constrained, TV-minimization problem. These conditions are used as a check on the images obtained from the present algorithm.

For constrained minimization the necessary conditions for a given image to be the optimal one are called the Karush Kuhn-Tucker (KKT) conditions (see for example Boyd & Vanderberghe (2004) and Nocedal & Wright (2006)). These conditions can be derived through the Lagrangian,

$$L = \|\vec{f}\|_{TV} + \lambda_0 \left(\left| M\vec{f} - \vec{g}_{\text{data}} \right|^2 - \epsilon^2 \right) - \vec{\lambda} \cdot \vec{f},$$

(10)

which combines the objective function with each constraint multiplied by a multiplier λ_i. Because the data constraint is a single constraint, it receives a single multiplier λ_0 in the Lagrangian. The non-negativity constraint is actually a constraint on each voxel, thus there is a multiplier λ_i for each voxel, where i runs from 1 to N_{im}, the number of voxels. The set of multipliers for the non-negativity constraints are abbreviated as $\vec{\lambda}$, a vector of the same dimension as the image size. For inequality constraints two conditions are imposed on the Lagrange multipliers: (a) non-negativity

$$\lambda_i \geq 0,$$

(11)

and (b) complementarity

$$h_i\left(\vec{f}\right)\lambda_i = 0,$$

(12)

where $i = 0, 1, \ldots, N_{im}$, and h_i represents the inequality constraints in the form $h_i(\vec{f}) \leq 0$,

$$h_0\left(\vec{f}\right)=\left|M\vec{f}-\vec{g}_{\text{data}}\right|^2-\epsilon^2 \leq 0$$
$$h_i\left(\vec{f}\right)=-f_i \leq 0 \quad i \in [1, N_{im}].$$

(13)

When an inequality constraint is violated, its corresponding h_i is positive. The complementarity condition Eq. (12) allows the Lagrange multiplier to be non-zero only when its corresponding inequality constraint is active, i.e., when Eq. (13) is satisfied with equality. The conditions for optimality of a potential solution $\vec{f}$ thus are Eqs. (11), (12), and (13) together with

$$\nabla_{\vec{f}}L=0,$$

(14)

where the differential operator $\nabla_{\vec{f}}$ is defined as

$$\nabla_{\vec{f}}Q\left(\vec{f}\right)=\sum_i \frac{\partial}{\partial f_i}Q\left(\vec{f}\right)\vec{\delta}_i,$$

(15)

and $\vec{\delta}_i$ is an image that is 0 everywhere except at the ith voxel where it is 1. Writing out Eq. (14) for the Lagrangian at hand, we obtain

$$\begin{aligned}\nabla_{\vec{f}}L &=\nabla_{\vec{f}}\|\vec{f}\|_{TV}+\lambda_0\nabla_{\vec{f}}h_0\left(\vec{f}\right)+\sum_{i=1}^{N_i}\lambda_i\nabla_{\vec{f}}h_i\left(\vec{f}\right)\\ &=\nabla_{\vec{f}}\|\vec{f}\|_{TV}+2\lambda_0 M^T\left(M\vec{f}-\vec{g}_{\text{data}}\right)-\vec{\lambda}\\ &=0.\end{aligned}$$

(16)

The transpose of the system matrix M^T corresponds to back-projection. Physically, the condition on the gradient of the Lagrangian is like a force balance. The constraint terms can be interpreted as a contact force, because the multipliers are only non-zero when the corresponding constraint is satisfied with equality due to Eq. (12). The strength of the contact force is proportional to the corresponding multiplier, and it always opposes the gradient of the objective function because of Eq. (11).

In general, the optimality conditions can not be tested based on an image estimate alone, because the Lagrange multipliers must be determined also. The structure of the problem at hand, however, does allow us to bypass solving the dual optimization problem directly (i.e., the problem whose solution yields the multipliers). For the present optimization problem, we focus on solving Eq. (16). Because the non-negativity constraints are active only at voxels that are zero, the Lagrange multipliers will be zero for voxels that are strictly positive, due to complementarity. This fact can be utilized to simplify Eq. (16). Using the indicator function

$$\vec{f}_{\text{indicator}}=\begin{cases} 1 & f_i \neq 0 \\ 0 & f_i=0 \end{cases},$$

(17)

we can simplify Eq. (16):

$$diag\left(\vec{f}_{\text{indicator}}\right)\left(\nabla_{\vec{f}}\|\vec{f}\|_{TV}+\lambda_0\nabla_{\vec{f}}h_0\left(\vec{f}\right)\right)=0,$$

(18)

where $diag(\vec{x})$ is a function that yields a diagonal matrix with the elements of $\vec{x}$ placed along the diagonal.

Equation (18) turns out to be the key for assessing the optimality of an image estimate $\vec{f}$. From $\vec{f}$ the following gradients can be computed:

$$\vec{d}_{TV}=diag\left(\vec{f}_{\text{indicator}}\right)\left(\nabla_{\vec{f}}\|\vec{f}\|_{TV}\right) \tag{19}$$

$$\vec{d}_{data}=diag\left(\vec{f}_{\text{indicator}}\right)\left(\nabla_{\vec{f}}h_0\left(\vec{f}\right)\right); \tag{20}$$

these vectors are the TV and data-constraint gradients masked by the indicator function of the image estimate. If $\vec{f}$ is optimal, *the two vectors, $\vec{d}_{TV}$ and $\vec{d}_{data}$, must point in exactly opposite directions* so that Eq.(18) can be satisfied by a positive λ_0. In principle, after determining λ_0 the rest of the KKT conditions can be tested by using Eq. (16) to solve for $\vec{\lambda}$, and by checking that all multipliers are non-negative and that the image estimate respects all the constraints. In practice, we check that the image estimate obeys the constraints and only that the vectors $\vec{d}_{TV}$ and $\vec{d}_{data}$ are back-to-back. We use the cosine of the angle between the two vectors

$$c_\alpha=\frac{\vec{d}_{TV}\cdot\vec{d}_{data}}{|\vec{d}_{TV}||\vec{d}_{data}|} \tag{21}$$

as a test of optimality. Ideally, c_α should be −1.0, but in practice this value is difficult to reach because a large number of iterations are required. Our experience from extensive numerical studies is that c_α is a very sensitive parameter and that there is often only imperceptible changes in the image after c_α goes below −0.5 (e.g., see the image progression of the jaw phantom reconstruction in Fig. 14 of the Results section). This test is crucial to the operation of the algorithm described below. Even though the algorithm is generally effective at finding the minimum-TV image that satisfies the data constraint and non-negativity, it is possible to select algorithm parameters that yield an image far from a true solution of the constrained, TV-minimization problem. The parameter c_α provides a very good test for an image estimate obtained from the algorithm detailed in the next section.

2.4. ASD-POCS: an algorithm for constrained, TV-minimization

The algorithm to find the solution to the constrained, TV-minimization problem is based on our previous algorithm presented in Sidky, Kao & Pan (2006*a*). That algorithm did not explicitly consider data inconsistency, and it was designed to find the solution to the constrained, optimization problem Eqs. (6)-(8) for the case of perfect data consistency, namely $\epsilon = 0$. As the optimization problem is convex, the algorithm of Sidky, Kao & Pan (2006*a*) was simply a combination of gradient descent on the TV objective function, Eq. (6), and POCS to enforce the data and positivity constraints, Eqs. (7) and (8), respectively. The main problem was to find a system to balance TV-gradient descent with POCS, and the balance was achieved by making the TV-gradient-descent step-size proportional to the change in the image due to POCS. Furthermore the proportionality constant for the TV-gradient descent was selected so that the change in the image at every iteration of was decreasing.

The ASD-POCS algorithm presented in this article seeks the solution of Eqs. (6)-(8) for a given ϵ. We emphasize again that is not a measure of the data inconsistency; it is a parameter of the

optimization problem, and it represents a bound on the error of the estimated data with respect to the available data. In developing the present algorithm it is useful to make the following observation: the data constraint, Eq. (7), is always active for practical values of . (Recall from Sec. 2.2, the term active refers to an inequality constraint whose dual variable, in this case λ_0, is strictly positive. Physically, in terms of the Lagrangian, this means that the constraint is exerting a force that prevents the solution from having a lower value of the objective function, in this case the TV norm.) This statement is obvious; if the data constraint is removed, the minimum-TV image that satisfies positivity would be any non-negative flat image, because a flat image has a TV of zero. Because the data-constraint is active, and because the optimization problem is convex, the solution we seek will satisfy Eq. (7) with equality:

$$|M\overrightarrow{f} - \tilde{g}_{\text{data}}| = \epsilon. \tag{22}$$

With this observation, the optimization problem, Eqs. (6)-(8), can be thought of, equivalently, as a variational inequality.

Some of the ideas for the current algorithm are inspired by the understanding of the present optimization problem in terms of the variational inequality (VI) formulation. The VI formulation goes as follows:

$$-\nabla_{\overrightarrow{f}}\|\overrightarrow{f}^*\|_{TV} \cdot \left(\overrightarrow{f}_C - \overrightarrow{f}^*\right) \leq 0, \tag{23}$$

where $\overrightarrow{f}^*$ is the solution to the VI problem; $\overrightarrow{f}_C$ is an image that satisfies the non-negativity and data constraints; and C refers to the convex feasible region of all images satisfying these constraints. In order to give an intuitive picture of the VI problem, a schematic of the problem is provided in Fig. 1. At the solution of the VI problem, $\overrightarrow{f}^*$, the vector $\overrightarrow{v}_2$, which points to any other image $\overrightarrow{f}_C$, must point away from the TV-gradient at $\overrightarrow{f}^*$, namely the vector $\overrightarrow{v}_1$.

For the problem at hand, we do not have a projector that takes an image outside the feasible region and yields the closest image in C; instead we use POCS to move the image estimate toward the feasible set C. Yamada (2001) and Xu & Kim (2003) present an algorithm where the projector is the solution of an iterative process such as POCS, but even so POCS is not strictly a projector for our constraints because the data constraint is a hyper-ellipsoidal, and, as such, the POCS iteration might not yield an image within C that is closest to the starting image. In any case, we have adapted the HSD algorithm to the present problem. Operationally, the change to the HSD algorithm is relatively minor, but it is enough of a difference to void the convergence proof in Xu & Kim (2003). The modified HSD algorithm iterates POCS until the image estimate satisfies the data and non-negativity constraints, followed by a "large" step in the direction opposite to the image TV-gradient. The process is repeated, while reducing the size of the steepest descent step. The algorithm is effective at finding the solution of the constrained, TV-minimization problem, but it is quite inefficient. The proposed ASD-POCS algorithm alternates one iteration of POCS with a steepest-descent step, while adapting the steepest-descent step-size to ensure that the image-constraints are satisfied.

2.4.1. ASD-POCS—In order to explain the ideas of the ASD-POCS algorithm, we provide in Fig. 2 a schematic of the image trajectories during the iteration of the algorithm. The algorithm step that improves data consistency is the algebraic reconstruction technique (ART), and basic projection enforces positivity. Together the ART and positivity projection is POCS. The set of images represented by $I(\epsilon_{\text{min}})$ are the images whose projection have the minimum distance to the given projection data. That ϵmin may be non-zero reflects the fact that the

projection data may not be internally consistent. The set of images represented by $I(\epsilon)$ represent all images that satisfy a given error tolerance. Obviously $\epsilon \geq \epsilon_{min}$. The goal of the ASD-POCS algorithm is to provide the minimum-TV image that is in $I(\epsilon)$. The thick dashed line represents the minimum-TV images as a function of ϵ (The minimum-TV image will necessarily be on the boundary of $I(\epsilon)$ as long as the constant image is not contained in $I(\epsilon)$).

The standard POCS algorithm alone would yield the image in $I(\epsilon)$ that is close to (and not necessarily the closest to) the starting image $\vec{f}_0$. In order to nudge the image toward the minimum-TV solution, we instead alternate POCS steps with TV-steepest descent. The result of this alternation yields a trajectory that is represented by the thin curve. The resulting image will be in $I(\epsilon)$, and its TV will generally be less than the POCS solution. But the resulting image may not be close to the minimum-TV image in $I(\epsilon)$. In order to approach this solution the modified HSD algorithm takes a single large step in the direction of the negative TV-gradient. This same step will in general take the image outside $I(\epsilon)$, but it can be brought back by repeating the POCS and gradient-descent alternation (HSD performs only POCS for this phase). The next time the image hits $I(\epsilon)$ it will in general have a reduced TV. This basic idea is modified in the ASD-POCS algorithm; instead of taking a large gradient-descent step, the gradient-descent step-size is slowly increased relative to the POCS step-size so that the current image estimate gently moves outside $I(\epsilon)$ in the direction of lowered image TV. Once the image is again outside $I(\epsilon)$ the gradient-descent step-size is reduced, so that it is smaller than the POCS step-size and the image returns to $I(\epsilon)$. This step-size adaptation is repeated until the stopping criteria are met. In our tests, the ASD-POCS algorithm appears to converge faster than the modified HSD algorithm. The step-sizes for the POCS and gradient descent are controlled adaptively (described in detail below). The complete iteration trajectory is schematically indicated by the thin curve in Fig. 2, and the proposed algorithm thus yields an image numerically close to the minimum-TV image within projection-data tolerance ϵ.

2.4.2. Pseudo-code for the ASD-POCS algorithm—We present the ASD-POCS algorithm in the form of a pseudo-code and abbreviate the notation where possible. The symbol := means assignment, meaning that the result on the right-hand side gets assigned to the variable on the left-hand side; image-space variables have a vector sign, e.g. $\vec{f}$, and a hat is used if the vector has unit length; data-space variables are denoted by a tilde, e.g. $\tilde{g}$. The vector $\vec{M}_i$ is the row of the system matrix that yields the ith data element.

The pseudo-code for ASD-POCS is:

1. $\beta := 1.0 \; ; \beta_{red} := 0.995 \; ;$

2. $ng := 20; \alpha := 0.2;$

3. $r_{max} := 0.95; \alpha_{red} := 0.95;$

4. $\vec{f} := 0$

5. repeat *main loop (POCS/descent loop)*

6. $\vec{f}_0 := \vec{f}$

7.
   for $i = 1, N_d$ do: if $\quad \vec{f} := \vec{f} + \beta \vec{M}_i \dfrac{g_i - \vec{M}_i \cdot \vec{f}}{\vec{M}_i \cdot \vec{M}_i} \quad ART$

8. for $i = 1, N_i$ do: if $f_i < 0$ then $f_i = 0$ *enforce positivity*

9. $\vec{f}_{res} := \vec{f}$

10. $\tilde{g} := M\vec{f}$

11. $dd := |\tilde{g} - \tilde{g}_0|$

12. $dp := |\vec{f} - \vec{f}_0|$

13. if {first iteration} then $dtvg := \alpha * dp$

14. $\vec{f}_0 := \vec{f}$

15. for $i = 1$, ng do *TV-steepest descent loop*

16. $\vec{df} := \nabla_{\vec{f}} \|\vec{f}\|_{TV}$

17. $\widehat{df} := \widehat{df} / |\widehat{df}|$

18. $\vec{f} := \vec{f} - dtvg * \widehat{df}$

19. end for

20. $dg := |\vec{f} - \vec{f}_0|$

21. if $dg > r_{max} * dp$ and $dd > \epsilon$ then $dtvg := dtvg * \alpha_{red}$

22. $\beta := \beta * \beta_{red}$

23. until {stopping criteria}

24. return $\vec{f}_{res}$

The optimization problem is specified by the projection data $\tilde{g}_0$ and the data-inconsistency-tolerance parameter ϵ. Six parameters control the complete algorithm, which are explained with the steps that make use of them. The values shown in lines 1-3 are typical of the values used in generating the results in the next section. For the studies in this paper the initial image function was initialized to zero, in line 4, but other choices are possible.

The outermost repeat-until loop, lines 5-23, contains two main components: adjustment toward data consistency with the POCS-step loop, and steepest descent toward lower-TV images. The key to the algorithm is how each of the respective step-sizes for POCS and TV-steepest descent are controlled. The image vector $\vec{f}_0$ is used as a place-holder image, in lines 6 and 14, in order to compute changes in the current estimate of the image. The POCS steps, lines 7 and 8, cause the image to respectively move toward data consistency and enforce non-negativity. The *ART* operator depends on the relaxation parameter β, which starts at 1.0 and slowly decreases to 0.0 as the iteration progresses. The current image is stored in $\vec{f}_{res}$ at line 9; on the last iteration it is the image after POCS that is considered to be the "final" one. The data residual is recomputed at lines 10 and 11, and the change in the image due to POCS is computed at line 12. Line 13 is used to convert the steepest-descent step-size from a fraction of a step-size to an absolute image distance on the first iteration. Lines 15-19 implement the TV-steepest descent. In practice, we find that, especially for the early iterations, the algorithm is effective when multiple small descent steps are taken for each POCS step; this point is discussed further at the end of this section. The change in the image due to TV-steepest descent is computed at line 20. The step-size adaptation is performed at line 21 in such a way as to have the image-estimate slide along the boundary of $I(\epsilon)$ toward images of lower TV. If the ratio of the change in the image due to steepest descent to the change in the image due to POCS is greater than r_{max} (usually < 1.0), the gradient-descent step-size is reduced by α_{red}. By controlling the steepest-descent step-size in this way the current image estimate changes toward $I(\epsilon)$. Once the current image satisfies the data-tolerance condition, the gradient-descent step-size is no longer reduced, allowing it to become larger than the POCS step-size, because the POCS step-size is always decreasing. As a result the current image will drift toward lower-TV images. When the data constraint is violated again, the steepest-descent step-size reduction will continue. Finally, in line 22 the ART-relaxation parameter is reduced by a constant fraction

β_{red}. This parameter controls the total number of iterations, because it affects the POCS step-size, which in turn affects the steepest-descent step-size.

The stopping criteria at line 23 are composed of various checks. First, the current image is checked to see if it satisfies the constraints of Eqs (7) and (8). Second, the parameter c_α as defined in Eq. (21) is examined to see if it is close to -1.0. Finally, if β is too small, the iteration is stopped. If the optimality conditions are violated, then the program is rerun with new parameters.

Generally, values of β_{red} closer to 1.0 yield more accurate solutions to the optimization problem, but such values also lengthen the iteration number. The steepest-descent parameters ng and α should be set initially so that ng * α is on the order of, but larger than, one. This is because the steepest-descent step-size dg should be comparable to dp. Note that in the above pseudo-code this product starts at 4.0, but in practice the initial value of dg in the first iteration of the main loop ends up being smaller than dp because the small descent steps are not in the same direction. The initial value of α is not critical, because this parameter is adaptively adjusted; however, one should err to the high side because α only shrinks during the execution of the algorithm. The parameters r_{max} and α_{red} control the evolution of α. The relative size of dg to dp is controlled by r_{max} in line 21. Values of r_{max} should be chosen in the interval(0, 1) to allow the resulting image to respect the data consistency and positivity constraints. And α_{red} should be much less than β_{red} so that α can be reduced fast enough to maintain the march toward data consistency. A progression of images is shown for the jaw phantom reconstruction in Sec. 3.2.2, showing how the ASD-POCS algorithm converges for a particular example.

The algorithm presented above is designed with simplicity in mind, and several steps could be taken to improve the efficiency. The projection at line 10 is computationally expensive, and it is only needed to compute the distance dd at line 11; this could be, for example, computed every 10th loop. The large number of steepest-descent steps at line 15 are needed to make sure that the TV-objective function is decreasing with every step; lines 15-19 could be replaced by a back-tracking search that would take many steps at early iterations when dp is large and few, or one, at later iterations. Also, depending on the application, early truncation may be an option. There is also the possibility of using efficient projectors at line 7 and line 10, or graphics processing unit (GPU) acceleration (Xu & Mueller 2007). The gradient-vector computation at line 16 is particularly well-matched to GPU computation (Sidky & Pan Lindau, Germany 2007).

The run-time for our current implementation of the ASD-POCS algorithm is 280 seconds per loop for the jaw phantom results shown below. The size of the jaw phantom simulation is as follows: the image size is $366 \times 546 \times 216 \approx 43,000,000$ voxels, and the projection data size is 64 views with a 610×93 detector or roughly 3,600,000 measured rays. The POCS step at line 7 and the projection at line 10 take 200 seconds running on one core of a dual-core AMD opteron 285 CPU, and the gradient descent loop with 20 sub-iterations runs in 80 seconds with a GPU implementation on an NVIDIA 7800 graphics card.

The ASD-POCS algorithm is presented here in the context of constrained TV-minimization, but it is clear that other convex objective functions could be employed such as the ℓ_1-norm.

3. Results

To demonstrate the ASD-POCS algorithm for image reconstruction, we perform two sets of studies: the first set of studies are designed in such a way as to acquire some theoretical understanding of image reconstruction with the ASD-POCS algorithm on the circular cone-beam CT configuration, the second set of numerical examples aims at how the ASD-POCS algorithm could be applied to dental scanning with a circular cone-beam CT scanner. We point

out that the studies here are designed to illustrate the ASD-POCS algorithm, and necessarily cannot be "complete" because the algorithm parameter-space is very large, and because 3D iterative image reconstruction is time-consuming. In the following, we use the ASD-POCS algorithm to find the solution to the constrained, TV-minimization problem for various scanning configurations in circular cone-beam CT. As we are using the ASD-POCS algorithm to find an accurate solution to this optimization problem, we refer to the solution of the constrained, TV-minimization problem interchangeably with the result of the ASD-POCS algorithm. Furthermore, for conciseness, we refer to the ASD-POCS algorithm in this application as the TV algorithm.

The goal of the first set of studies in Sec. 3.1 is to acquire some understanding of how the TV algorithm performs on image reconstruction with the circular cone-beam configuration under ideal conditions. For these studies the projection data are generated by applying the projection system-matrix M to a discrete image $\vec{f}_0$, so that the data are perfectly consistent, to within computer accuracy, with the system matrix. Minimal levels of noise are introduced to assess stability of the algorithm. Such studies give an upper bound to the performance of the TV algorithm, and they can also give some insight into the theoretical properties of image reconstruction from the circular cone-beam CT data with constrained TV-minimization.

For the second set of studies in Sec. 3.2, the data model is more realistic in that the projection data are generated by computing line integrals through computer phantoms represented by continuous functions. For these studies there are two sources of data inconsistency: the images are necessarily reconstructed onto voxelized arrays which entails mismatch with data generated from continuous objects, and different levels Gaussian noise is added where the variance is proportional to the line integral values (a reasonable approximation for CT). For typical noise levels and image array sizes these two sources of data inconsistency may be comparable. For these studies there may be no image $\vec{f}$ that satisfies $M\vec{f} = \tilde{g}$. Image reconstruction in this situation is expected to be degraded, but these results give a sense as to how the TV algorithm might perform on actual circular, cone-beam CT projection data. These tests also demonstrate the robustness of the proposed algorithm as neither source of data inconsistency is well-matched with the equally weighted Euclidean norm constraint on the data in Eq. (7).

A potential application to dental CT scanning is studied by simulating data for the FORBILD jaw phantom (http://www.imp.uni-erlangen.de/forbild/english/results/index.htm), shown in Figs. 12 and 13. This phantom is challenging because there are many high-contrast structures along with a low-contrast tumor in the tongue. The problem of imaging the tumor from under-sampled projection data is a particularly difficult task. Again, because we are performing 3D iterative reconstructions, the number of simulations is limited. We selected a few scanning configurations representing different distributions of the projection views, trading off limited angular scanning for angular under-sampling. The particular examples were chosen to illustrate that the TV algorithm is effective at reconstructing circular cone-beam CT data with under-sampled projection data, and furthermore, that the constrained TV-minimization, which is particularly effective with angular under-sampling, may impact the scan design itself.

3.1. Reconstructions from data generated by the matched projection model

Theoretically, it is known that compactly supported continuous functions can be reconstructed for the circular cone-beam scanning configuration (Ramm & Katsevich 1996). It is, however, also known that image reconstruction in this case is highly unstable. The studies in this section explore these theoretical conclusions in the discrete setting, where the true image $\vec{f}_0$ is a voxelized array of gray levels. The problem we address here is the situation where the data vector $\tilde{g}_0$ is given by:

$$\tilde{g}_0 = M\vec{f}_0, \tag{24}$$

as opposed to using Eq. (1) to generate the simulated data, and image reconstruction involves solving this linear system for f_0 from knowledge of $\tilde{g}_0$. Recall that M is the system matrix that yields the discrete set of projection measurements for circular cone-beam scanning from a discrete image array. Based on what is known for the corresponding analytic image reconstruction problem, one might expect to obtain fairly accurate reconstructed images in the situation that the data $\tilde{g}_0$ contain no noise or other inconsistencies. Furthermore, iterative algorithms generally impose additional conditions on the image function that are not done in the analytic case, i.e. positivity or in the present case minimum TV. If the additional assumptions on the image are actually satisfied, it is possible that the instability of image reconstruction in circular cone-beam CT may be reduced.

For the present studies we employ a relatively small image array of $100 \times 100 \times 100$ voxels. To obtain the true image, the Defrise disk phantom is embedded into this array. This phantom is challenging for studying image reconstruction in circular cone-beam CT because of the rapid variations in the z-direction. A slice of the true image is shown in Fig. 3. In order to save time on the image reconstruction, we model a half-cone geometry. The source to iso-center distance is 50 cm and the source to detector distance is 100 cm. The detector height is 20.7 cm with the bottom edge of the detector at the plane of the circular trajectory. This configuration is equivalent to a full cone-angle of $23°$. The source trajectory covers the full $360°$; in one case the source trajectory is covered by 128 view angles, and for a few-view simulation the circular trajectory is executed with only 25 views. The former study with relatively high angular sampling isolates the effect of the cone-angle, while the latter study demonstrates the effectiveness of the TV algorithm on few-view projection data in the setting of circular cone-beam scanning. In both cases, the focus is on reconstruction from noiseless, consistent data, but we also introduce perturbations in the form of Gaussian noise to the projection data to assess stability of the image reconstruction in each case. For comparison, the TV-algorithm results are shown together with POCS, which is essentially the TV algorithm without the TV-steepest-descent steps.

For the first study, projection data are generated for the discretized disk phantom at 128 views. Because there is no inconsistency in the projection data, theoretically it is possible to drive the data residual down to zero. Hence the data-tolerance parameter is set to zero. Slices of the volumes reconstructed by use of the TV algorithm and POCS are shown in Fig. 4. It appears that under the conditions of this idealized simulation, both the TV and POCS algorithms yield accurate reconstructions. The number of measured rays here is 2,880,000, which is greater than the number of voxels 1,000,000. So it is possible that the discrete linear system describing this imaging problem has a unique solution. But it is also known that image reconstruction in circular cone-beam CT is unstable for continuous image functions. We investigate the instability of image reconstruction in the present discrete case by adding noise to the projection data.

In the next set of reconstructions, we add Gaussian distributed noise to the simulated projection data, where the variance is set to 0.1% of the corresponding detector bin value. In Fig. 5, slices of the reconstructed volumes are shown for the TV and POCS algorithms in a narrow contrast window. Visually, the POCS image and the TV image for $\epsilon = 1.775$ appear very similar, because this value of ϵ is near ϵ_{min} and with dense ray sampling there is likely not much room in the set of feasible images. As a result, the TV algorithm has little effect. Loosening the data constraint by setting $\epsilon = 2.0$ allows the feasible set of images to expand, and the TV algorithm clearly has a greater effect. The speckle noise is substantially reduced, while the true structures

still are well-defined. Further increasing ϵ to 2.5 yields an image that appears to lose resolution at high cone-angles near the top of the image.

The effect of ϵ can be seen quantitatively by plotting the root-square error (RSE) of the reconstructed image as a function of ϵ in Fig. 6. The trend of this plot is interesting in that the minimum RSE error occurs at an $\epsilon > \epsilon_{min}$, mirroring the visual trend of Fig. 5. The reason why the TV reconstruction can become more accurate with increasing ϵ is that the underlying image function is known to have a sparse GMI. Another interesting point about this result is that the minimum image-RSE occurs near the noise level introduced in the simulated data. The RSE of the projection data from the noiseless data is 2.14. The proximity of the minimum image-RSE to $\epsilon = 2.14$ is likely due to the fact that the noise introduced is Gaussian which matches the Euclidean distance used in the data constraint Eq. (7). At first glance the fact that there exist images with $\epsilon < 2.14$ seems impossible, but on further reflection it is possible to have ϵ lower than the data RSE because a component of the data noise will be consistent with the cone-beam projection matrix. When ϵ is near ϵ_{min}, the TV algorithm for this scanning configuration tends toward the POCS result, because the present linear system appears to be over-determined. In other words, the constraints Eq. (13) specify a very limited set of images, and minimizing the image TV has little effect. As the data tolerance ϵ is loosened, the set of feasible images expands, and there is more freedom to find feasible images with lower TV. As ϵ is increased further, the image becomes more blurry and less accurate in the RSE sense. In practice, the true image is not known and such a plot cannot be made, but the operating ϵ can be found by simulation of the scanning configuration and imaging subject.

In the previous example the sampling of disk phantom projections over-determines the values of the voxels in image array. For the next set of studies we investigate a decidedly under-determined system where the scan simulates only 25 projection views with a 100×100 bin detector. The number of samples in this case is a quarter of the number of voxels in the image; moreover, the voxel sampling is non-uniform. Figure 7 shows slices from the volumes reconstructed by TV and POCS algorithms. For this case, the difference between the two results is more dramatic. The TV reconstruction appears to be very accurate, while the POCS image has significant artifacts. What allows the TV algorithm to recover the image for this case is the extra condition imposed on the image that it has a sparse GMI. The disk phantom has 100,327 non-zero voxels in its GMI. In order for the TV algorithm to reconstruct accurately the number of measurements should be at least twice the number of non-zero GMI values, and the data set with 25 views contains 250,000 measurements. Indeed, under ideal conditions, the image recovery is very accurate for the TV algorithm.

With 25 projection view scan, we repeat the study on the effect of the data tolerance on image reconstruction by constrained TV-minimization. The standard deviation for the Gaussian noise model is again set to 0.1% of the corresponding data values. Using the POCS algorithm, the data tolerance minimum RSE, ϵ_{min} is estimated to be 0.033, which is much lower than the RSE of 0.63 introduced into the noiseless projection data. The reason why the data RSE can be reduced so much is that the projection data are under-sampled in view angle, and as a result only a small component of the noise is inconsistent with the cone-beam transform. In Fig. 9 the dependence of the reconstructed image RSE is shown as a function of ϵ. Again, the RSE error in the reconstructed image first drops then increases with increasing ϵ. Interestingly, the RSE values for TV algorithm at $\epsilon = \epsilon_{min}$ are lower than that for the POCS algorithm. Furthermore, looking at the reconstructed images in Fig. 8, the TV algorithm for $\epsilon = \epsilon_{min}$ appears to give a more accurate reconstructed image than that of POCS. The reason for the difference in the results between POCS and TV with $\epsilon = \epsilon_{min}$ is that the image sampling is highly under-determined in this case. The constraint on the image for $\epsilon = \epsilon_{min}$ no longer specifies a unique image, and in general the images that do satisfy the data constraint do not all have the same TV. The TV algorithm selects the one with the minimum TV, which will not

in general be the same as the image arrived at by POCS. Again the success of the TV algorithm in terms of accuracy of the image reconstruction depends on the fact that the underlying image function has a sparse GMI.

The system configurations investigated in this section are idealized in the sense that the simulated projection data are generated from the same discrete cone-beam projection operator used in the reconstruction algorithms. Furthermore, in the examples with data noise, the inconsistency in the data is well characterized and understood. These studies do, however, help in comparing algorithms for inversion of the imaging linear system, and they give an upper bound to the performance of the reconstruction algorithms. In the case of circular cone-beam scanning with full sampling in the angular direction, both POCS and TV algorithms yield accurate reconstructions when there is no data inconsistency. Thus it suggests that the circular cone-beam configuration can be highly accurate even though this configuration does not satisfy Tuy's data sufficiency condition. When data noise is introduced, the reconstructed images are degraded, but the TV algorithm appears to provide improved image reconstruction over POCS. For the case of under-sampled projection data, the gap between the images reconstructed by POCS and constrained TV-minimization is much larger. For the discretized disk phantom, the TV algorithm yields accurate image reconstruction while the POCS results contain severe artifacts.

An important point to understand about the data inconsistent case, for both the sufficiently- and under-sampled projection data, is the role of the data tolerance ϵ and how to determine it. In the studies above, we have studied the dependence of the image-RSE as a function of ϵ, and it is clear that this dependence is not monotonic. As a result there is an ϵ that gives an optimal reconstructed image in the sense of minimal RSE. This optimal ϵ, however, depends on numerous factors. It depends on the underlying image function, the form and level of the data inconsistency, the system matrix, and also the metric by which the quality of the image is being judged (in this case RSE). For the curves computed in Figs. 6 and 9, the difference is the system matrix which represents sufficient sampling in the former and under sampling in the latter. As can be seen the ϵ that yields the optimal reconstructed image, in terms of RSE, is very different in both cases. Additionally, the importance of finding the optimal ϵ also depends on the above factors and the desired goal of the imaging task. In the under-sampled case studied above, a wide range of ϵ for constrained TV-minimization leads to substantially better images than POCS, and the RSE appears to be flat, while for the densely sampled case, the image quality metric may depend more strongly on ϵ.

3.2. Reconstructions from data generated by a continuous object data model

The examples in this section aim at investigating the performance of the TV algorithm using a continuous object data model which more realistically models cone-beam CT imaging. Previously, the disk phantom used was a discretized 3D volume array, and data were generated by discrete projection. In this section, the phantoms are defined as combinations of analytic shapes such as ellipsoids and cylinders, and the projection data are modeled by analytic line integrals through these shapes. As a result, even in the case of no noise, there will be a mismatch between the projection operator that generates the data and the discrete projector used in the reconstruction algorithms. Theoretically, the inconsistency between the actual data model and the projector used in the algorithm can be reduced by increasing image resolution (Zbijewski & Beekman 2006), but in practice there will always be a limitation due to the correspondingly larger computational effort. Noise is also considered in the data model. The first set of reconstructions are performed again on the half disk phantom, and the set of studies employ the FORBILD jaw phantom, which has both high contrast and low contrast objects.

3.2.1. Disk phantom—For the image reconstruction from projections of the continuous disk phantom, the size of the projection data set and reconstruction volume are increased. The simulated detector now has 400×400 bins. The projection data are taken at only 32 views. The volume is reconstructed into a 400^3 array. We show reconstructed volume slices of the disk phantom for the case where no noise is added to the projection data and for Gaussian noise added at a level of 1% and 5%. Figure 10 shows a vertical slice through the reconstructed disk phantom for the 32 view data containing no noise. The TV result clearly shows fewer artifacts than that of POCS. This difference is not surprising considering that there are only 32 views in the data set. The TV reconstruction for the present continuous disk phantom and unmatched system matrix appears to be less accurate than the TV reconstructions of the previous section with the discrete disk phantom and matched system matrix. Again, this is the expected result as the present noiseless projection data contain inconsistencies as well as being insufficient for exact reconstruction. The TV algorithm here is run with $\epsilon = 10.0$. Some results for TV reconstruction with noisy data are shown in Fig. 11 along with the previous noiseless reconstruction. The data tolerances for the 1% and 5% noisy data sets are respectively $\epsilon = 20.0$ and $\epsilon = 64.0$. Clearly, the higher noise level results in a noisier image, but the TV algorithm appears to be robust against data inconsistencies. It is particularly interesting in the 5% noise case that the few-view streak artifacts seen in the POCS image of Fig. 10 do not appear here. Although TV images are shown here for single values of , it may be important to explore other values of ϵ as is done in Sec. 3.1. Such a study, which is beyond the scope of the current work, is, in any case, usually highly task dependent; the optimal ϵ depends on what the image is being used for.

3.2.2. Jaw phantom—In order to provide a more stringent and realistic test for the TV algorithm, we employ a more detailed phantom; namely, we generate data from the analytic FORBILD jaw phantom. This phantom has more structures with varying contrast levels than does the Defrise disk phantom, as seen in Figs. 12 and 13. The investigations here explore different configurations with a limited number of projection views on a circular scanning orbit. The visual test for the various configurations and the TV algorithm is the apparent contrast of the low-contrast tumor in the tongue. Such a test is challenging for any reconstruction algorithm, because streak artifacts from the surrounding high contrast objects such as the teeth can swamp low-contrast structures. In each of the scans simulated for the jaw phantom, the number of views is fixed at 64; however, the angular range of the x-ray source is varied. The 64 views are distributed uniformly over a 180° arc in front of the jaw (front scan), in back of the jaw (back scan), and over the full 360° circle around the jaw (2π scan). The trade-offs for these scans are as follows: for the back and front scans the angular resolution will be better than that of the 2π scan while the 2π scan obviously has better angular coverage; and between the back and front scans the part of the volume opposite the x-ray source will have better cone-beam coverage due to beam divergence while the angular coverage will be diminished for the same region. For all of the projection data Gaussian noise is added at a level of 0.1%. The simulated detector has 610×93 detector bins. As the simulated number of detector rows is only 93, the projection data are truncated in the vertical direction. The reconstructed volume is composed of 366×546×216 cubic voxels.

Before comparing the different scanning configurations, we look in depth at the progress of the TV algorithm for the 2π scan. Frames for a sagittal slice through the jaw are shown in Fig. 14 at increasing iteration number. The data tolerance for this reconstruction is set to $\epsilon = 156.1$, and this value is reached at iteration $i = 500$. The second image in the series at $i = 600$ already satisfies the data constraint, but the value for c_α is substantially larger than −1.0. Thus the $i = 600$ image is not the minimum-TV image satisfying the data constraint. As the various step-sizes in the TV algorithm are adjusted, the data distance converges to the set value 156.1 and c_α nears −1.0. With the current version of the TV algorithm the image will not in general reach the solution of the optimization problem Eqs. (6)-(8), but the series of images shown in Fig.

14 indicate that in practice the present algorithm is close enough and that c_α provides a good test for the constrained, TV-minimization solution. It is clear from the figure that there is little change in the current image estimate from iteration 1200 to 1800 even though the images are displayed in a narrow contrast window. There is, however, still quite some change in c_α. We have found empirically that there is little significant difference in the image once c_α goes below -0.5, but we aim to select algorithm parameters so that we can attain c_α in the neighborhood of -0.9. It is possible to achieve images with c_α arbitrarily close to -1.0 by replacing the POCS step with gradient descent, but such an algorithm is slow and appears to be unnecessary.

The large number of iterations to achieve near convergence may lead to questions about the algorithm practicality. But we point out that no other algorithm that we know of is capable of recovering the low contrast tumor no matter how many iterations are performed. The number of iterations necessary to achieve near convergence can vary substantially depending on scanning configuration and properties of the object being scanned. The jaw phantom example considered here is particularly challenging, requiring a large number of iterations. This number may be substantially reduced if the structures of interest have higher contrast. Finally, mathematical convergence may be unecessary in practice (Sidky et al. 2008).

In Figs. 15-18 we compare image reconstruction of the jaw phantom by the TV algorithm and by expectation maximization (EM) for data acquired with the front, back, and 2π scans. We compare with the EM algorithm because it is commonly used for iterative image reconstruction. We do not make, here, any conclusions about the performance of the TV algorithm versus EM because there are many parameters in the algorithm implementations. We only seek to illustrate that the additional assumption of a sparse GMI in the underlying object function for constrained TV-minimization may yield reconstructed images of greater accuracy. (The reconstructed images using POCS without TV gradient descent are qualitatively similar to the EM images). For the present implementation of EM, the current image estimate is regularized by convolving it with Gaussian with the width of half the voxel length on every iteration. In Fig. 15 transaxial slices at a z-plane through the tumor are shown, while in Fig. 16 transaxial slices in the plane of the x-ray source trajectory are shown. Figures 17 and 18 show sagittal and coronal slices respectively. The EM reconstructions all show prominent streak artifacts due to the fact that the projection data contain only 64 views. The artifacts in the EM results completely swamp the low contrast tumor in the tongue. The TV images, however, have no significant streaking due to the limited number of views, and the resulting images can be displayed in a gray scale window narrow enough to pick up the tumor.

The tumor in the TV images seems to be most easily seen in the case of the back scan. The tumor is also clearly visible in the 2π scan, but it is more blurred in comparison with the back scan. One might expect that the front scan would by the most capable of resolving the tumor, because, from the point of view of the tumor, the angular coverage in front scan should be more complete. The angular coverage turns out not to be the deciding factor. The relevant property of the scan that explains this result is the cone-beam coverage. The tumor is off the mid-plane of the circular scan; in fact, it is near the edge of the cone-beam. Looking at the TV images in Fig. 17, the effect of the cone-beam geometry becomes clear. For the front scan, the modeled x-ray source is closer to the tumor on average, and as a result the cone-beam is too narrow to see the tumor for some of the views. For the back scan, the cone-beam is larger at the tumor and it is visible from all the views. Visually, the TV algorithm appears to have relatively uniform image quality going from the back of the jaw to the front of the jaw, so the back scan appears to be advantageous due to the increased cone-beam coverage. Note that the EM sagittal images in Fig. 17 show the intuitive image quality behavior. The front scan for the EM resolves the front teeth better than the back scan. It is interesting to note the effect of increasing the angular spacing between views, and thereby the angular coverage, for the TV algorithm. The

resulting image appears to lose resolution without introducing streak artifacts typical of angular under-sampling.

4. Discussion

In this article, we have developed and investigated a compressive-sensing image reconstruction algorithm which aims at solving constrained minimization of the estimated image TV norm. Although the application presented here is circular cone-beam CT, it is clear that this algorithm can be applied to other tomographic imaging modalities with linear system models. The algorithm is closely related to the well-known POCS algorithm, taking advantage of an additional property of many underlying object functions: namely, that the object being scanned often times has a sparse GMI. When this assumption is valid, the ASD-POCS algorithm may yield accurate image reconstruction even when the projection data are highly under-sampled. Additionally, for the case of circular cone-beam CT, the minimum-TV restriction on the object function appears to stabilize image reconstruction at parts of the imaging volume at large cone-angles for circular cone-beam scanning. When the assumption of sparse GMI does not apply, the resulting reconstructed images are generally no worse than the POCS result due to the design of the algorithm. The presented TV algorithm used for TV-minimization is simple with few algorithm parameters so that analyzing its performance on image reconstruction is possible. Future work will focus on improving algorithm efficiency, as was discussed at the end of Sec. 2.4.2, and studying the impact of early truncation, so that the TV algorithm can be practical for actual scanning systems.

As demonstrated by the results using the FORBILD jaw phantom, the TV algorithm appears to be able to resolve low contrast structures in the presence of high-contrast objects even with the projection data sets are limited in angular range or view number, and when the low contrast object is not at the mid-plane of the circular x-ray trajectory. The fact that the scan with the x-ray source sweeping over the back of the head appears to yield the best results with constrained TV-minimization, is counter-intuitive since the tumor being imaged is more toward the front of the head. This example illustrates a property of the TV algorithm that may play a role in system scanning design when using this algorithm. In general, the constrained TV-minimization performs well with respect to angular under-sampling, so it may be advantageous to sacrifice angular view sampling density for improving cone-beam or angular range coverage. We emphasis, however, that it is difficult to make general conclusions about the performance of the TV algorithm because its performance depends on the structure of the object being scanned.

Although this article presents the TV algorithm in the context of circular cone-beam CT scanning, it is clear that this algorithm is directly applicable to other cone-beam CT configurations. More generally, the TV algorithm applies with minor modifications to a host of other tomographic imaging applications as long as the data model depends linearly on the underlying object function. In future work, we will investigate the impact of other physical factors that complicate the data model, and the present algorithm will be applied to actual data sets from x-ray based tomographic scanning systems.

Acknowledgments

References

Alizadeh F, Goldfarb D. Second-order cone programming. Math. Program., Ser. B 2003;95:3–51.

Benson TM, Gregor J. Three-dimensional focus of attention for iterative cone-beam micro-CT reconstruction. Phys. Med. Bio 2006;51(18):4533–4546. [PubMed: 16953041]

Boone JM, Shah N, Nelson TR. A comprehensive analysis of DgN(CT) coefficients for pendant-geometry cone-beam breast computed tomography. Med. Phys 2004;31(2):226–235. [PubMed: 15000608]

Boone JM, Nelson TR, Lindfors KK, Seibert JA. Dedicated breast CT: Radiation dose and image quality evaluation. Radiology 2001;221(3):657–667. [PubMed: 11719660]

Boyd, S.; Vanderberghe, L. Convex Optimization. Cambridge University Press; 2004.

Candes E, Romberg J, Tao T. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. IEEE Trans. Inf. Theory 2006a;52:489–509.

Candes E, Romberg J, Tao T. Stable Signal Recovery from Incomplete and Inaccurate Measurements. Comm. Pure Appl. Math 2006b:1207–1223.

Carvalho, Bruno A.; Herman, Gabor T. Low-dose, large-angled cone-beam helical CT data reconstruction using algebraic reconstruction techniques. Im. Vis. Comp 2007;25(1):78–94.

Chartrand, Rick. Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process. Lett 2007;14:707–710.

Chen B, Ning R. Cone-beam volume CT breast imaging: Feasibility study. Med. Phys 2002;29(5):755–770. [PubMed: 12033572]

Chen G-H, Tang J, Leng S. Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. Med. Phys 2008;35:660–663. [PubMed: 18383687]

Daly MJ, Siewerdsen JH, Moseley DJ, Jaffray DA, Irish JC. Intraoperative cone-beam CT for guidance of head and neck surgery: Assessment of dose and image quality using a C-arm prototype. Med. Phys 2006;33(10):3767–3780. [PubMed: 17089842]

Delaney AH, Bresler Y. Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography. IEEE Trans. Image Proc 1998;7:204–221.

Elbakri IA, Fessler JA. Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation. Phys. Med. Bio 2003;48(15):2453–2477. [PubMed: 12953909]

Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. JOSA A - Im. Sci. Vis 1984;1(6):612–619.

Ford EC, Chang J, Mueller K, Sidhu K, Todor D, Mageras G, Yorke E, Ling CC, Amols H. Cone-beam CT with megavoltage beams and an amorphous silicon electronic portal imaging device: Potential for verification of radiotherapy of lung cancer. Med. Phys 2002;29(12):2913–2924. [PubMed: 12512728]

Ford NL, Thornton MM, Holdsworth DW. Fundamental image quality limits for microcomputed tomography in small animals. Med. Phys 2003;30(11):2869–2877. [PubMed: 14655933]

Jaffray DA, Siewerdsen JH, Wong JW, Martinez AA. Flat-panel cone-beam computed tomography for image-guided radiation therapy. Int. J. Rad. Onc. Bio. Phys 2002;53(5):1337–1349.

Kachelriess, Marc; Knaup, Michael; Bockenbach, Olivier. Hyperfast parallel-beam and cone-beam backprojection using the cell general purpose hardware. Med. Phys 2007;34(4):1474–1486. [PubMed: 17500478]

Kole JS. Statistical image reconstruction for transmission tomography using relaxed ordered subset algorithms. Phys. Med. Bio 2005;50(7):1533–1545. [PubMed: 15798342]

Kole JS, Beekman FJ. Evaluation of accelerated iterative x-ray CT image reconstruction using floating point graphics hardware. Phys. Med. Bio 2006;51:875–889. [PubMed: 16467584]

Lange K, Carson R. EM reconstruction algorithms for emission and transmission tomography. J. Comp. Asst. Tomo 1984;8(2):306–316.

Lee SC, Kim HK, Chun IK, Cho MH, Lee SY, Cho MH. A flat-panel detector based micro-CT system: performance evaluation for small-animal imaging. Phys. Med. Bio 2003;48(24):4173–4185. [PubMed: 14727760]

Letourneau D, Wong JW, Oldham M, Gulam M, Watt L, Jaffray DA, Siewerdsen JH, Martinez AA. Cone-beam-CT guided radiation therapy: technical implementation. Rad. Onc 2005;75(3):279–286.

Lewitt RM. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. JOSA A - Im. Sci. Vis 1990;7(10):1834–1846.

Li MH, Yang HQ, Kudo H. An accurate iterative reconstruction algorithm for sparse objects: Application to 3D blood vessel reconstruction from a limited number of projections. Phys. Med. Biol 2002;47:2599–2609. [PubMed: 12200927]

Manglos SH, Gagne GM, Krol A, Thomas FD, Narayanaswamy R. Transmission maximum-likelihood reconstruction with ordered subsets for cone-beam CT. Phys. Med. Bio 1995;40(7):1225–1241. [PubMed: 7568379]

Matej S, Lewitt RM. Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. IEEE Trans. Med. Im 1996;15(1):68–78.

Nagurney, A.; Zhang, D. Projected dynamical systems and variational inequalities with applications. Kluwer; Boston, Massachussets: 1996.

Nocedal, J.; Wright, S. Numerical Optimization. 2nd ed.. Springer; 2006.

Panin VY, Zeng GL, Gullberg GT. Total Variation Regulated EM Algorithm. IEEE Trans. Nuc. Sci 1999;46:2202–2210.

Persson M, Bone D, Elmqvist H. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. Phys. Med. Biol 2001;46:853–866. [PubMed: 11277230]

Rafferty MA, Siewerdsen JH, Chan Y, Moseley DJ, Daly MJ, Jaffray DA, Irish JC. Investigation of C-arm cone-beam CT-guided surgery of the frontal recess. Laryn 2005;115(12):2138–2143.

Ramm, AG.; Katsevich, AI. The Radon Transform and Local Tomography. CRC Press; Boca Raton, FL: 1996.

Ritman EL. Molecular imaging in small animals - Roles for micro-CT. *J. Cell. Bio.* Suppl. 2002;39:116–124.

Sauer K, Bouman C. A local update strategy for iterative reconstruction from projections. IEEE Trans. Sig. Proc 1993;41(2):534–548.

Sharp GC, Kandasamy N, Singh H, Folkert M. GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration. Phys. Med. Bio 2007;52(19):5771–5783. [PubMed: 17881799]

Sidky EY, Kao C-M, Pan X. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. J. X-ray Sci. Tech 2006a;14:119–139.

Sidky EY, Kao C-M, Pan X. Effect of the data constraint on few-view, fan-beam CT image reconstruction by TV minimization. IEEE Nuc. Sci. Conf. Rec. 2006 2006b;4:2296–2298.

Sidky, EY.; Reiser, I.; Nishikawa, RM.; Pan, X.; Chartrand, R.; Kopans, DB.; Moore, RH. Practical iterative image reconstruction in digital breast tomosynthesis by non-convex TpV optimization. In: Hsieh, J.; Samei, E., editors. Proc. SPIE Med. Imag. 2008: Phys. Med. Imag. 6913. 2008. p. 691328

Sidky EY, Chartrand R, Pan X. Image reconstruction from few views by non-convex optimization. IEEE Nuc. Sci. Conf. Rec. 2007 2007;5:3526–3530.

Sidky EY, Pan X. Accurate image reconstruction in circular cone-beam computed tomography by total variation minimization: a preliminary investigation. IEEE Nuc. Sci. Conf. Rec. 2006 2006;5:2904–2907.

Sidky, EY.; Pan, X.; Germany, Lindau. Few-view, cone-beam CT image reconstruction by GPU-accelerated total variation minimization. In: Kacheriess, M.; Beekman, F.; Müller, K., editors. 9th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine; 2007. p. 60-63.

Song J, Liu QH, Johnson GA, Badea CT. Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT. IEEE Trans. Nuc. Sci 2007;34:4476–4483.

Vogel CR, Oman ME. Iterative methods for total variation denoising. SIAM J. Sci Comp 1996;17:227–238.

Xu, Fang; Mueller, Klaus. Real-time 3D computed tomographic reconstruction using commodity graphics hardware. Phys. Med. Bio 2007;52(12):3405–3419. [PubMed: 17664551]

Xu HK, Kim TH. Convergence of Hybrid Steepest-Descent Methods for Variational Inequalities. J. Opt. Th. App 2003;119:185–201.

Yamada, I. The hybrid steepest-descent method for variational inequality problems over the intersection of the fixed-point sets of nonexpansive mappings. In: Butnariu, D.; Censor, Y.; Reich, S., editors. Inherently parallel algorithms in feasibility and optimization and their applications. North-Holland, Amsterdam, Holland: 2001. p. 473-504.

Yu L, Zou Y, Sidky EY, Pelizarri CA, Munro P, Pan X. Region of Interest Reconstruction from Truncated Data in Circular Cone-beam CT. IEEE Trans. Med Imag 2006;25(7):869–881.

Zbijewski, Wojciech; Beekman, Freek J. Comparison of methods for suppressing edge and aliasing artefacts in iterative x-ray CT reconstruction. Phys. Med. Bio 2006;51:1877–1889. [PubMed: 16552111]

Zbijewski, Wojciech; Defrise, Michel; Viergever, Max A.; Beekman, Freek J. Statistical reconstruction for x-ray CT systems with non-continuous detectors. Phys. Med. Bio 2007;52(2):403–418. [PubMed: 17202623]
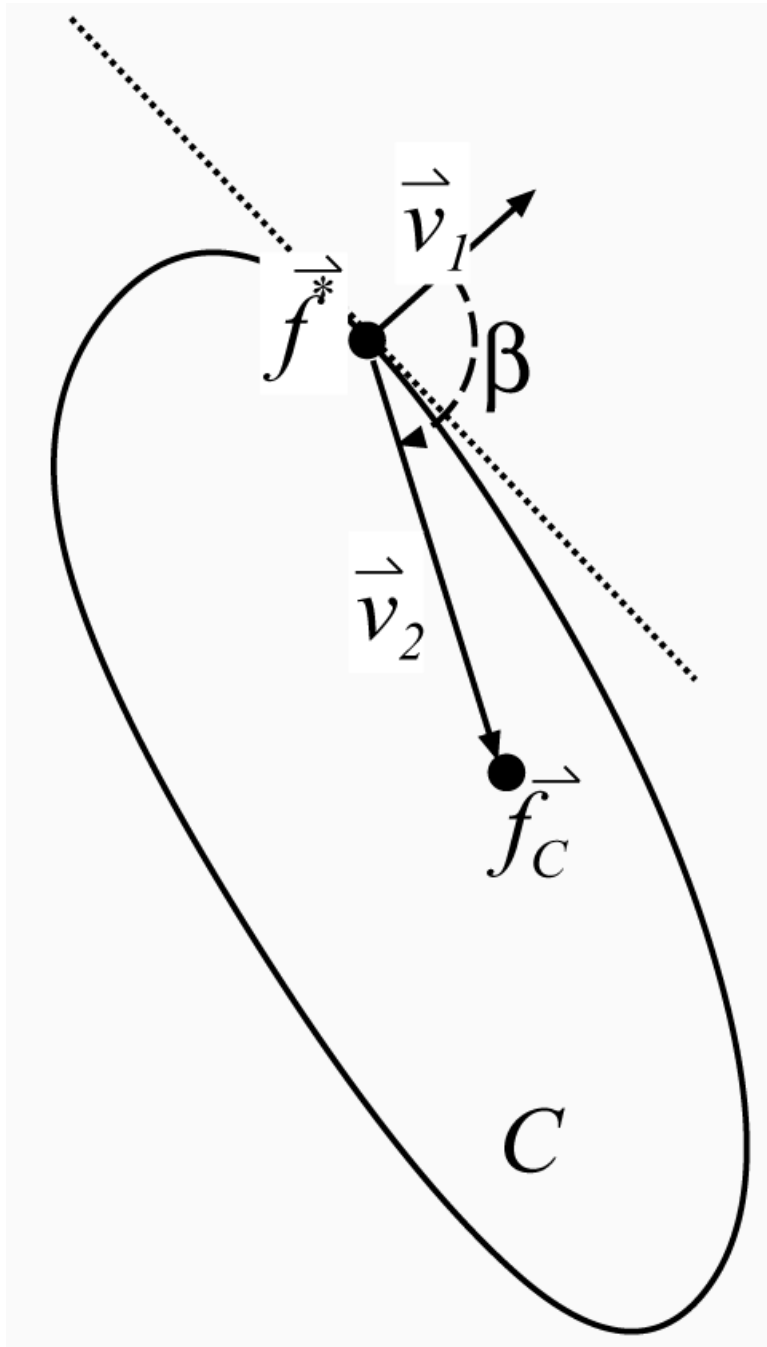
Figure 1.
Illustration of the structure of the variational inequality (VI) formulation of the TV optimization problem. The region C represents all feasible images, in this case images that satisfy the data and positivity constraints. $\vec{v} = -\vec{\nabla}\|f^*\|_{TV}$ and $\vec{v_2} = \vec{f_C} - \vec{f^*}$. Solving the VI problem amounts to finding $\vec{f^*}$ such that $\vec{v_1} \cdot \vec{v_2} \leq 0$ (or equivalently $\pi/2 \leq \beta \leq 3\pi/2$) for any feasible image $\vec{f_C}$.
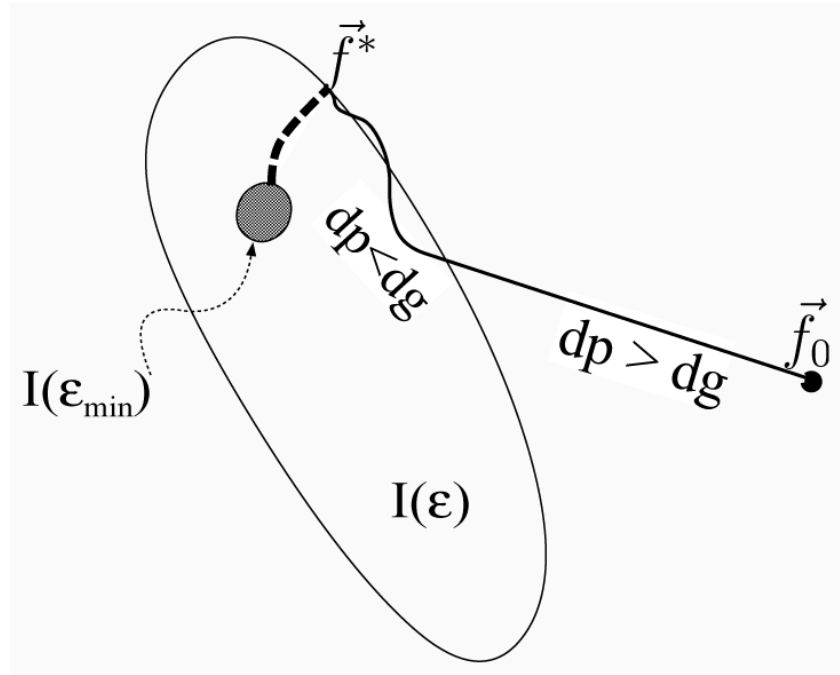
Figure 2.
Schematic of the ASD-POCS algorithm in the space of possible images. The dark circle labeled $I(\epsilon_{min})$ represents the image or set of images with minimum error ϵmin in the estimated projection data. The region labeled $I(\epsilon)$ represents all images that agree with the data to within tolerance ϵ. The image within $I(\epsilon)$ with minimum TV will in general be on the boundary of $I(\epsilon)$, and the thick dashed line represents the minimum-TV images for the regions $I(\epsilon')$ where $\epsilon_{min} < \epsilon' < \epsilon$. The dot outside of $I(\epsilon)$ and the thin curve are respectively the starting image $\vec{f}_0$ and image trajectory for the ASD-POCS algorithm. In the ASD-POCS algorithm specified by the pseudo-code in this section, two image distances dp and dg control the image trajectory. The distances dp and dg are respectively the magnitudes of the change in the image due to POCS and TV-steepest descent. When the current image estimate is outside of $I(\epsilon)$, dg is adaptively controlled to be less than dp. When the current image estimate is inside of $I(\epsilon)$, dg is adaptively controlled to be greater than dp. The resulting image trajectory that is schematically denoted by the thin curve terminates near the VI solution f^*.
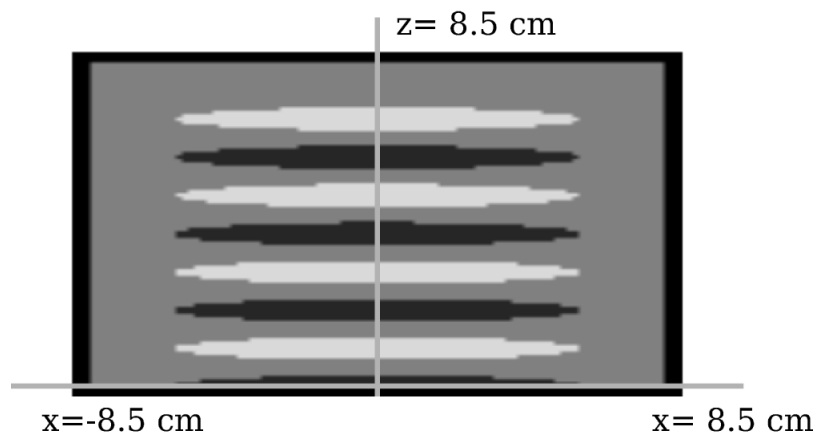
Figure 3.
Vertical slice at $y = 0$ through the discrete Defrise disk phantom. The phantom is cut in half at the $z = 0$ plane in order to reduce computational time. The plane of the circular x-ray source trajectory is at $z = 0$.
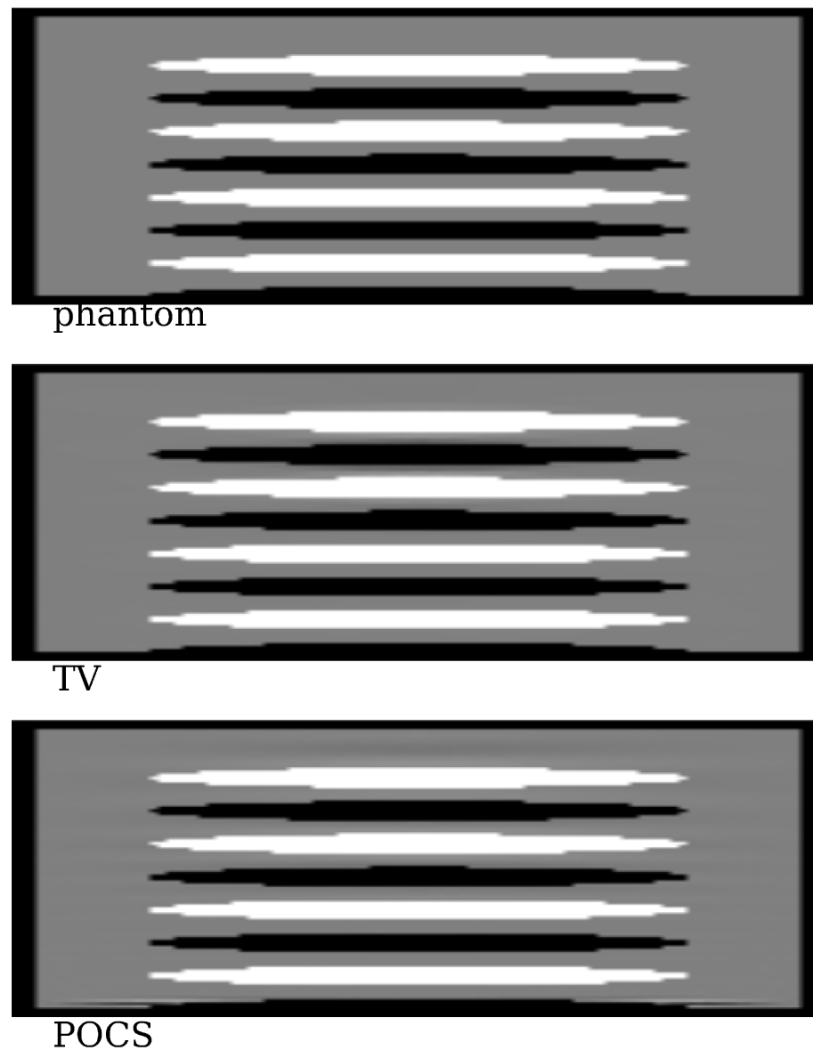
Figure 4.
Vertical slice at $y = 0$ through the half Defrise disk phantom; top: phantom, middle: reconstructed image by the TV algorithm, and bottom: reconstructed image by POCS. The simulated projection data contains 128 views covering 2π radians. The gray scale of the images is [0.9,1.1].
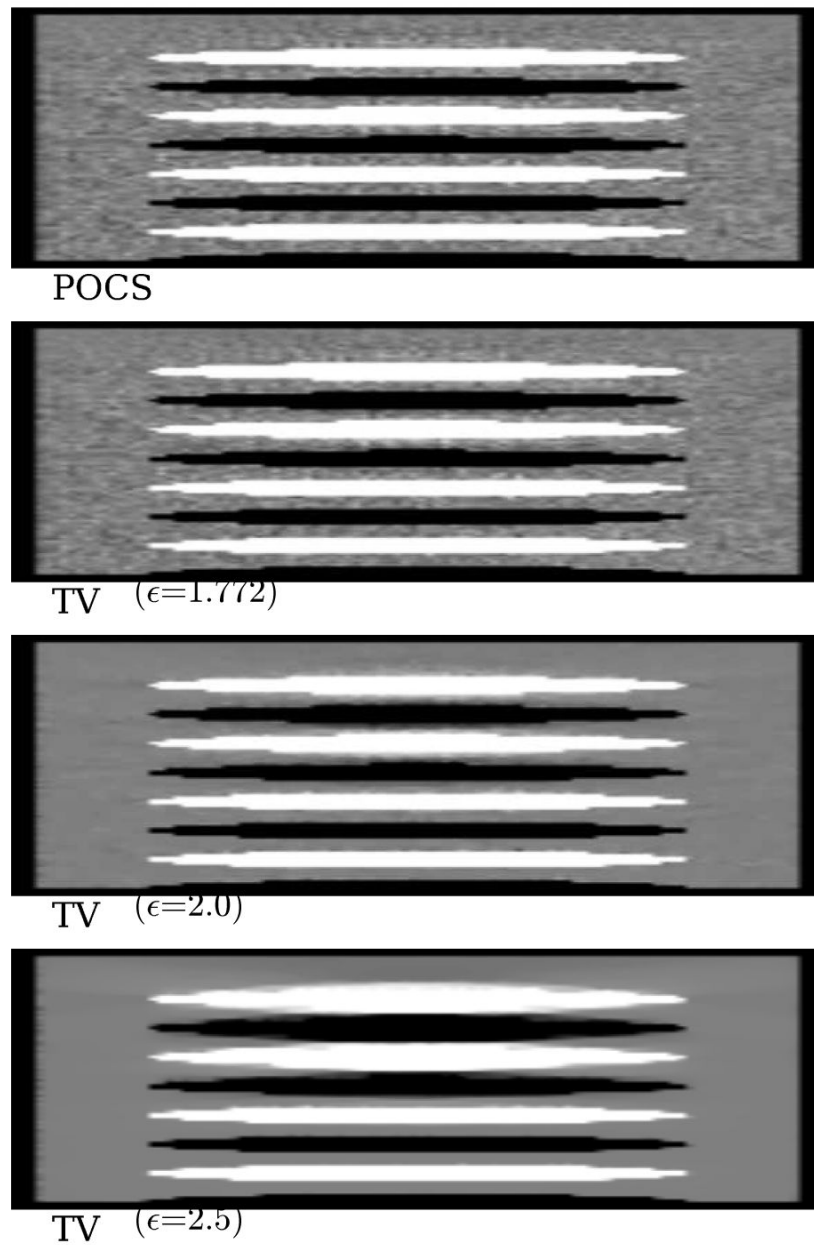
Figure 5.
Reconstructed images at the $y = 0$ slice of the half Defrise disk phantom with simulated 0.1% Gaussian noise added to the projection data. The top image is the POCS reconstruction, and the other three are TV reconstructions for three different values of ϵ. The simulated projection data contains 128 views covering 2π radians. The gray scale is [0.95,1.05].
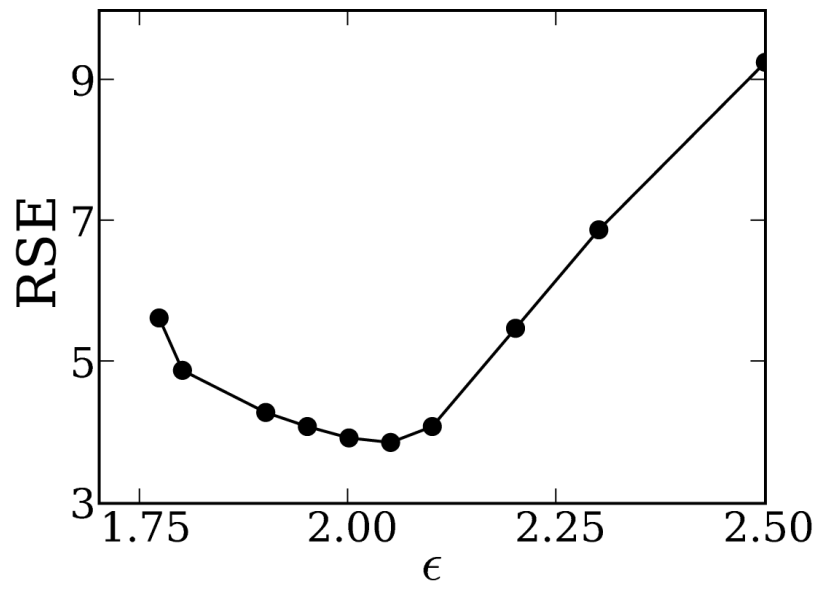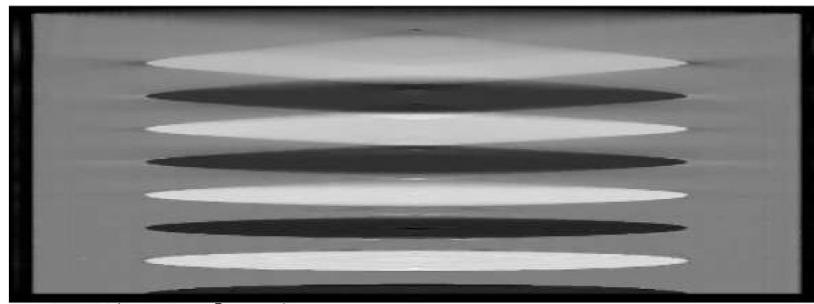
Figure 6.
Image RSE as a function of data tolerance ϵ. For reference, the RSE deviation from the noiseless data is 2.14, and the POCS estimate for ϵ_{min} is 1.772.
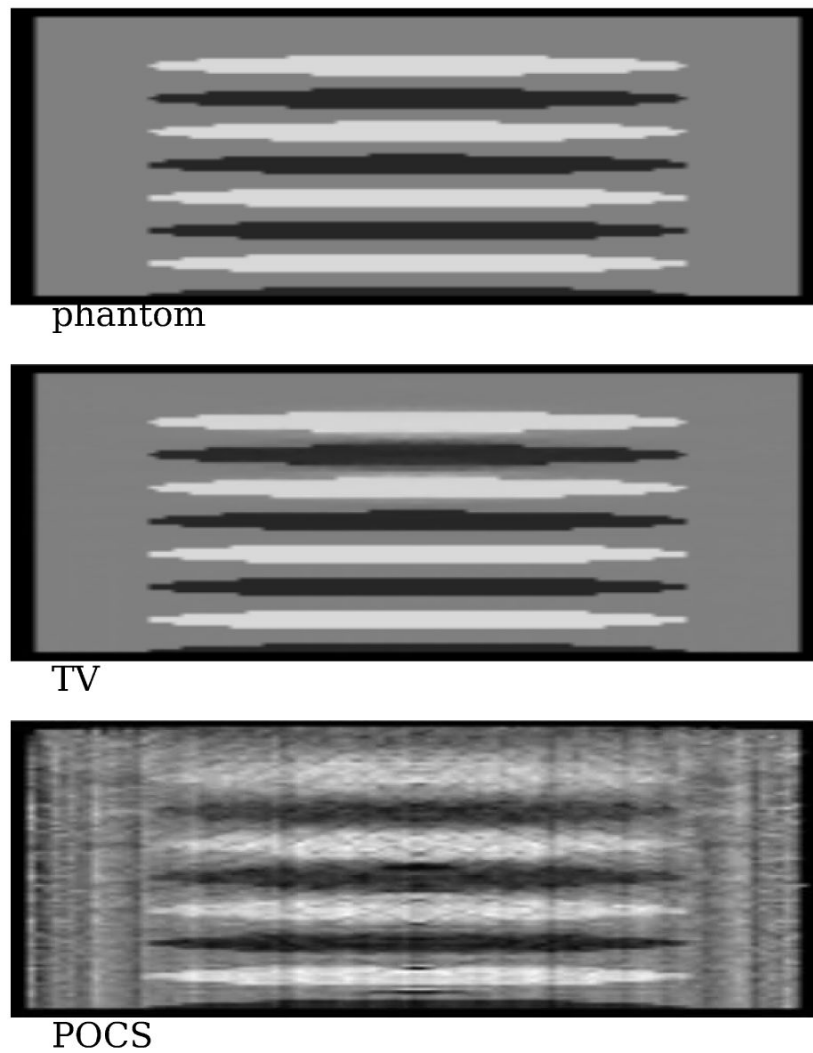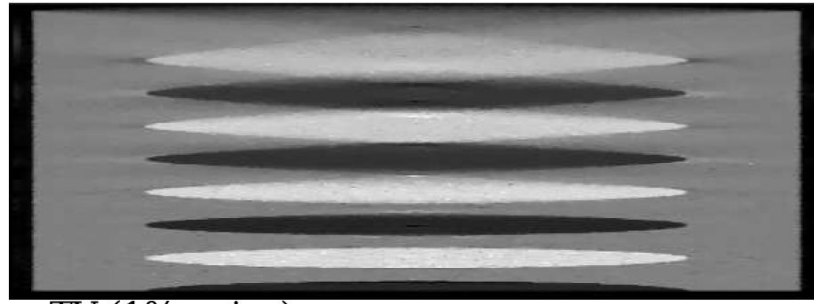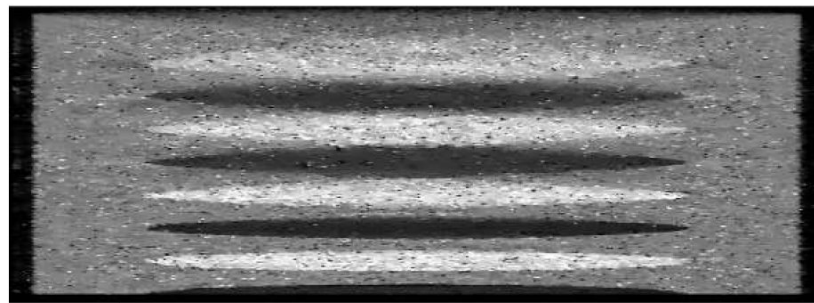
Figure 7.
Vertical slice at $y = 0$ through the half Defrise disk phantom; top: phantom, middle: reconstructed image by the TV algorithm, and bottom: reconstructed image by POCS. The simulated projection data contains only 25 projection views covering 2π radians. The gray scale of the images is [0.0,2.0].
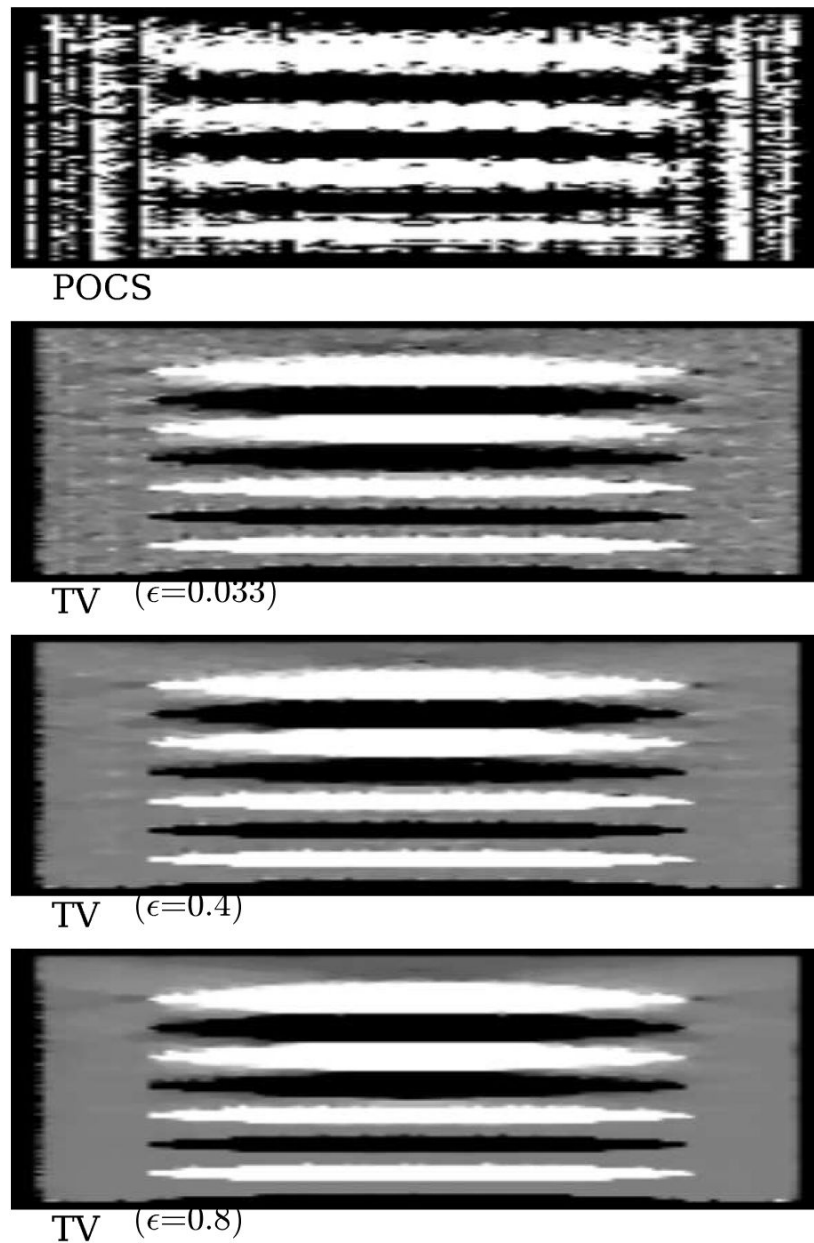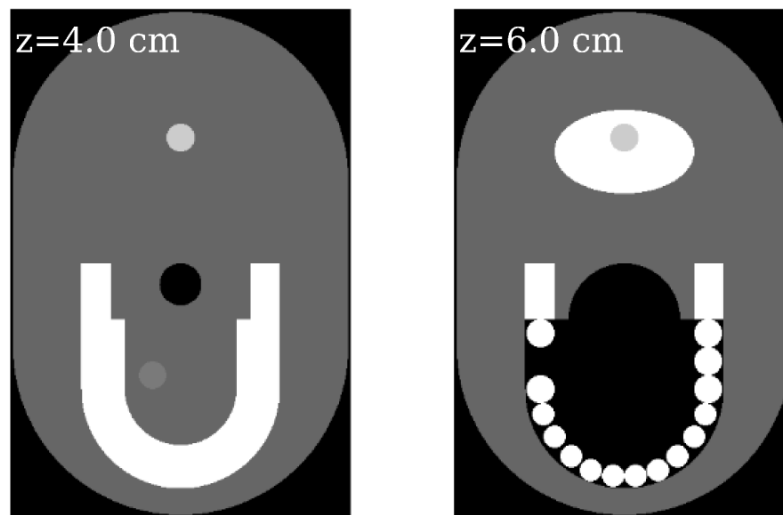
Figure 8.
Reconstructed images at the $y = 0$ slice of the half Defrise disk phantom with simulated 0.1% Gaussian noise added to the projection data. The top image is the POCS reconstruction, and the other three are TV reconstructions for three different values of ϵ. The simulated projection data contains 25 views covering 2π radians. The gray scale is [0.95,1.05].
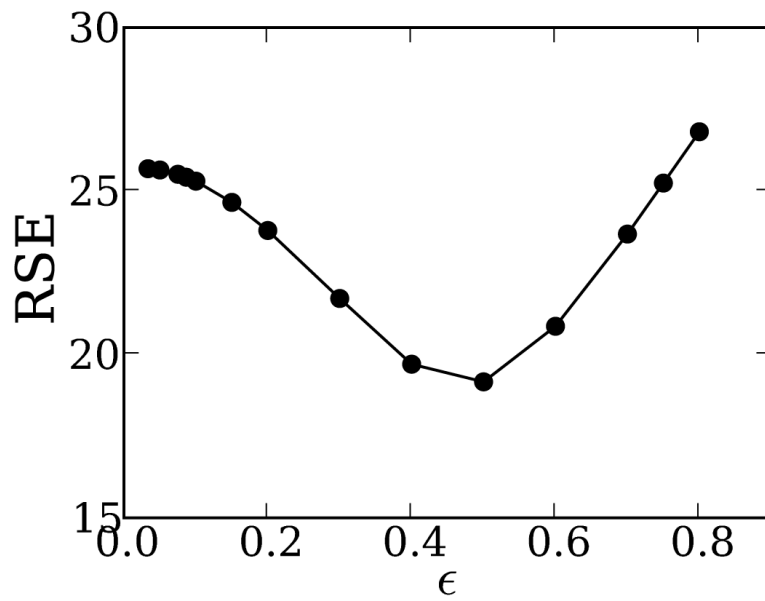
Figure 9.
Image RSE, for images reconstructed from the 25-view data set, as a function of data tolerance ϵ. For reference, the RSE deviation from the noiseless data is 0.63, and the POCS estimate for $\epsilon_{\min}$ is 0.033.
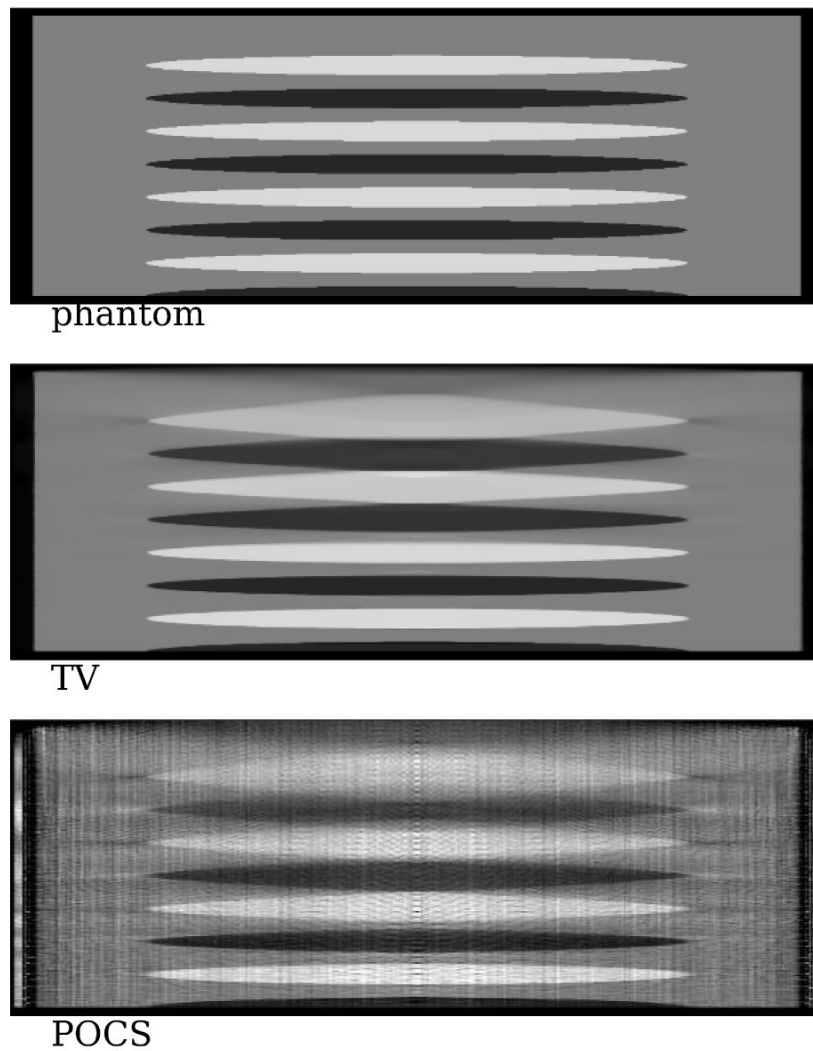
Figure 10.
Vertical slice at $y = 0$ through the half Defrise disk phantom; top: phantom, middle: reconstructed image by the TV algorithm, and bottom: reconstructed image by POCS. The simulated, analytic projection data contains only 32 projection views covering 2π radians. The gray scale of the images is [0.0,2.0].

Figure 11.
Reconstructed images at the $y = 0$ slice of the half Defrise disk phantom with simulated Gaussian noise, of varying strength, added to the analytic, 32-view projection data. The gray scale is [0.0,2.0].

Figure 12.
Slices at $x = 1.0$ cm(left)andat $y = 7.0$ cm (right) through the FORBILD jaw phantom. The slices were chosen to intersect the tumor located at (−1.0 cm, 7.0 cm, 4.0 cm). The displayed gray scale is [0.9,1.15].
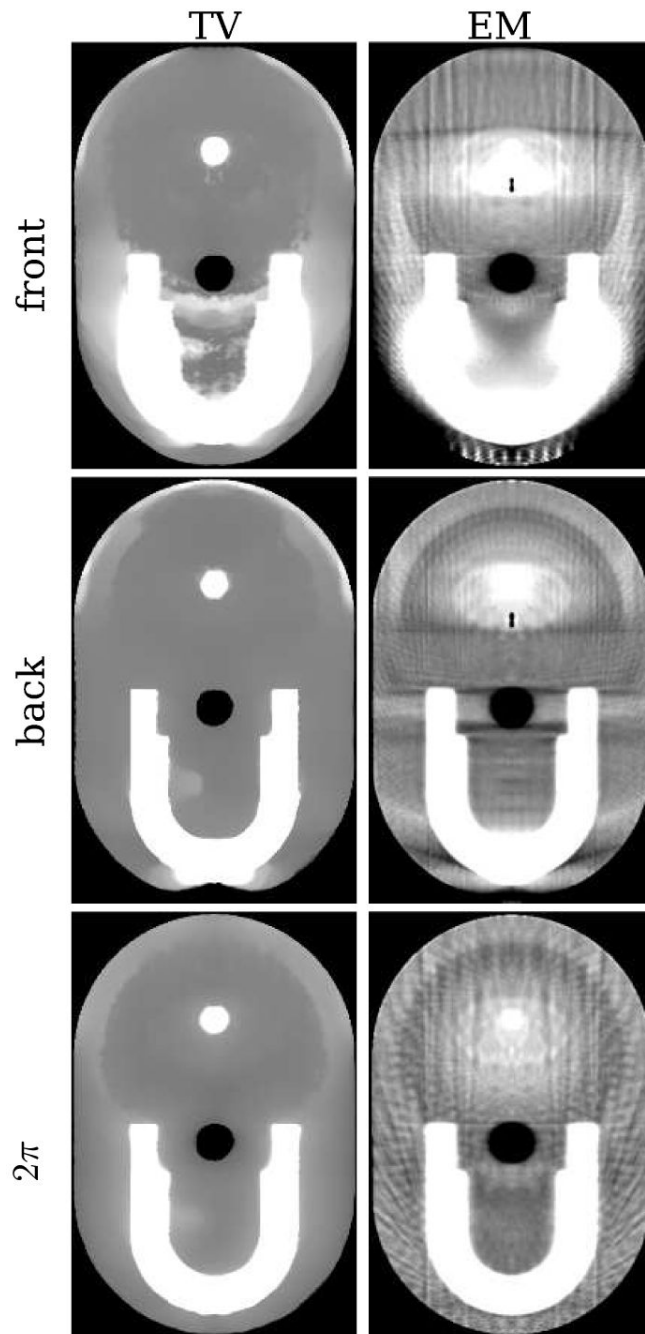
Figure 13.
Two trans-axial slices at $z = 4.0$cmand $z = 6.0$ cm through the FORBILD jaw phantom. The slice at $z = 4.0$ cm intersects the tumor located at (−1.0 cm, 7.0 cm, 4.0 cm). The displayed gray scale is [0.9,1.15].
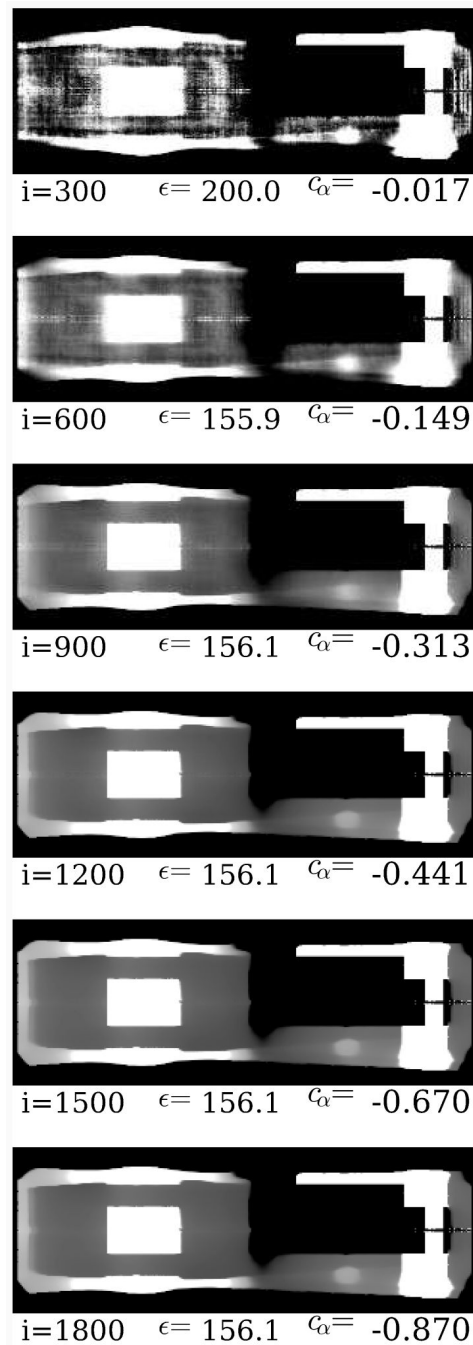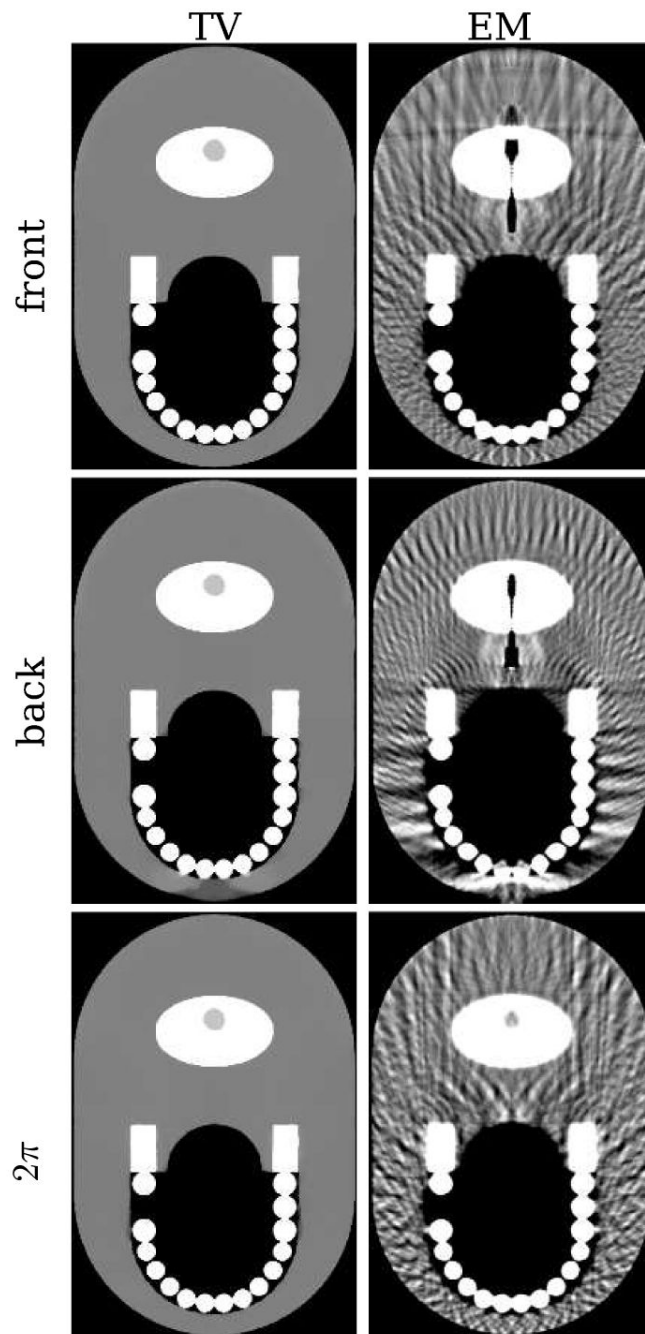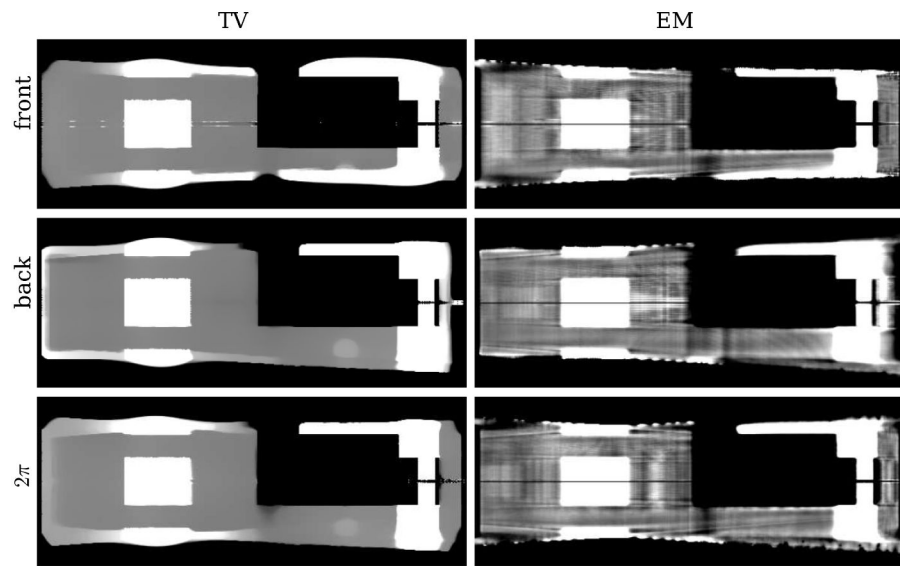
Figure 14.
Intermediate images showing the progress of the iterative ASD-POCS algorithm for constrained TV-minimization. Shown is a sagittal slice at $x = -1.0$ cm with a gray scale of [0.995,1.01]. The value of ϵ is set to 156.1 which was reached at iteration number 500.

Figure 15.
Transaxial slice at z = 4.0 cm for images reconstructed by the TV and EM algorithms for 64-view projection data distributed in various configurations. The TV and EM images are shown at a gray scale of [0.98,1.02] and [0.8,1.2], respectively. The larger gray scale window for the EM results is necessary due to the prominent streaking artifacts.
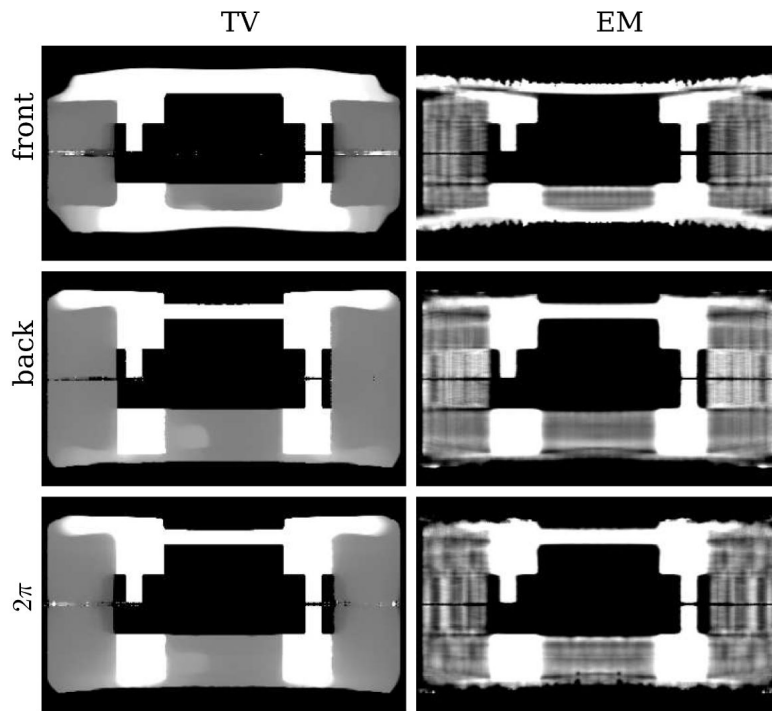
Figure 16.
Transaxial slice at $z = 6.0$ cm for images reconstructed by the TV and EM algorithms for 64-view projection data distributed in various configurations. The TV and EM images are both shown at a gray scale of [0.8,1.2].

Figure 17.
Sagittal slice at $x = -1.0$ cm for images reconstructed by the TV and EM algorithms for 64-view projection data distributed in various configurations. The TV and EM images are shown at a gray scale of [0.98,1.02] and [0.8,1.2], respectively. The larger gray scale window for the EM results is necessary due to the prominent streaking artifacts.

Figure 18.
Coronal slice at $y = 7.0$ cm for images reconstructed by the TV and EM algorithms for 64-view projection data distributed in various configurations. The TV and EM images are shown at a gray scale of [0.98,1.02] and [0.8,1.2], respectively. The larger gray scale window for the EM results is necessary due to the prominent streaking artifacts.