

# Total Variation and Tomographic Imaging from Projections

Per Christian Hansen\*    Jakob Heide Jørgensen\*  
Technical University of Denmark  
{pch, jakj}@imm.dtu.dk

## Abstract

Total Variation (TV) regularization is a powerful technique for image reconstruction tasks such as denoising, in-painting, and deblurring, because of its ability to produce sharp edges in the images. In this talk we discuss the use of TV regularization for tomographic imaging, where we compute a 2D or 3D reconstruction from noisy projections. We demonstrate that for a small signal-to-noise ratio, this new approach allows us to compute better (i.e., more reliable) reconstructions than those obtained by classical methods. This is possible due to the use of the TV reconstruction model, which incorporates our prior information about the solution and thus compensates for the loss of accuracy in the data. A consequence is that smaller data acquisition times can be used, thus reducing a patient's exposure to X-rays in medical scanning and speeding up non-destructive measurements in materials science.

**Keywords:** Total variation regularization, tomography.

**AMS Subject Classification:** 65K10, 65R32

Tomography is the science of “seeing through objects.” Physical signals — waves, particles, currents — are sent through an object from many different angles, the response of the object to the signal is measured, and an image of the object's interior is reconstructed. Computed tomography (CT) is an indispensable tool in modern science and technology as a non-invasive measurement technique for diagnostics, exploration, analysis, and design, and it has become an independent research field on the border between mathematics, scientific computing, and application sciences [5].

Tomographic imaging is an ill-posed problem, which means that it involves the computation of solutions that are extremely sensitive to data errors, model errors, and rounding errors. Useful reconstructions can only be computed by incorporating prior information in order to define unique, stable, and physically meaningful solutions [4]. *Total variation* (TV) reconstruction, originally proposed for image denoising by Rudin, Osher and Fatemi [11], see also [2], incorporates the prior knowledge that the reconstructions must be piecewise smooth with occasional steep transitions, i.e. sharp edges — the underlying assumption being a Laplacian distribution for the image's gradient magnitude. The TV reconstruction model seeks to do so by explicitly producing an image with a sparse gradient (something that is not achieved by other reconstruction methods such as filtered back projection or Tikhonov regularization), and this fact establishes an interesting connection to compressed sensing [1], [3].

---

\*This work is part of the project CSI: Computational Science in Imaging, supported by grant 274-07-0065 from the Danish Research Council for Technology and Production Sciences.

A variety of TV algorithms have been developed over the years, e.g., time marching algorithms, fixed-point iteration, and various minimization-based methods such as sub-gradient methods, second-order cone programming methods, duality based methods, and graph-cut methods. Many of these algorithms are specifically geared towards 2D problems in image processing, such as denoising, in-painting and deblurring. Other algorithms are more general in nature and therefore also applicable to the large sparse systems of equations that arise in 2D and 3D computed tomography. At any rate, we shall not try to survey all these algorithms here.

The use of TV in MRI tomography was already considered in one of the original papers on compressed sensing [1]; here we focus on conventional CT where the imaging model is not based on random sampling. A basic result regarding the use of TV in tomography is that the TV reconstruction model—due to the way it incorporates prior information about the image—enables us to achieve a good reconstruction quality with less data, or with more noise in the data, than required by a classical reconstruction algorithm. However, one should be careful with such a definitive statement, because several parameters in the model and the algorithm have a non-negligible influence on the TV reconstruction.

Our goal here is thus to illustrate the complex interplay between the choice of these parameters and the quality of the computed TV reconstructions. We consider a number of important tasks: formulate an optimization problem that gives the desired reconstruction and can be solved in realistic time, find an algorithm which is fast enough, find parameter windows that give a useful reconstruction, find adequate stopping criteria, determine the optimal amount of dose pr. views, etc.

Our computations were primarily done with an optimal first-order method developed by us [6]—but our conclusions carry over to other applications and are not specific for our particular TV algorithm.

Below we summarize the steps involved in getting from the measurements to a computed solution, and we introduce important parameters associated with each step.

**Scanner.** In the scanner we can control the dose (the intensity of the source) and the number of views (or positions of the source/detector). The number of bins or pixels of the detector is fixed by the manufacturers of medical scanners, but in other applications we can control this parameter. Associated with the scanner is the *true object* that we want to reconstruct.

**Mathematical model.** The mathematical model describes the relation between the rays, the object, and the detectors, and it describes the noise in the data. This step also specifies how we represent the model and the solution on the computer. The model will also (perhaps implicitly) include a deterministic and/or stochastic model of our a priori knowledge of certain properties of object. This model defines the *desired solution*, i.e., the solution we want to compute if there were no errors. In general, this desired solution is only a discrete approximation to the underlying true object.

**Reconstruction model.** The reconstruction model defines an optimization problem which incorporates some kind of regularization in order to handle ill-posedness of the mathematical model, and whose solution is what we want to compute in the face of the above-mentioned errors. The *regularized solution* depends on the regularizing function (used to impose stability) and the regularization parameter (and perhaps other parameters), and it is in general only an approximation to the desired solution.

**Numerical algorithm.** The numerical algorithm defines the particular way we decide to solve the regularization problem. We compute a *numerical solution* which is a (preferably good) approximation to the regularized solution, and whose quality depends on various algorithm parameters, such as the initial guess for the iterations, the stopping criterion, and the choice of algorithm itself.

The *mathematical model* used in this work takes the form of a linear system of equations  $Ax \approx b$  where the sparse system matrix  $A \in \mathbb{R}^{m \times n}$  models the scanning process. The reconstructed  $N \times N$  image is represented by  $x \in \mathbb{R}^n$  (with  $n = N^2$ ), and the right-hand side  $b \in \mathbb{R}^m$  represents the data from the scanner. While the raw data essentially consists of photon counts with Poisson noise, our data  $b$  is obtained by further processing the raw data and we assume that the noise in  $b$  has a Gaussian distribution with standard deviation  $\sigma$ .

Our *TV reconstruction model* has the form

$$\min_{x \geq 0} f(x), \quad f(x) = \frac{1}{2} \|Ax - b\|_2^2 + \alpha \sum_{j=1}^n \phi_\tau(D_j x), \quad (1)$$

where the second term is the TV regularization term:  $\alpha > 0$  is a regularization parameter that controls how much regularization we wish to impose, the matrices  $D_j \in \mathbb{R}^{2 \times n}$  are designed such that  $D_j x \in \mathbb{R}^2$  is a finite difference approximation to the gradient at pixel  $j$ , and  $\phi_\tau(D_j x)$  is our smooth approximation to the gradient magnitude:

$$\phi_\tau(D_j x) = \begin{cases} \|D_j x\|_2 - \frac{\tau}{2} & \text{if } \|D_j x\|_2 \geq \tau, \\ \frac{1}{2\tau} \|D_j x\|_2^2 & \text{else.} \end{cases} \quad (2)$$

This is actually the Huber approximation—other smooth approximations might as well be used, such as  $(\|D_j x\|_2^2 + \tau^2)^{1/2}$ ; both include a smoothing threshold  $\tau$ . We use a smooth approximation because the gradient magnitude  $\|D_j x\|_2$  is not differentiable, and while algorithms for non-smooth optimization do exist, they generally suffer from slow convergence.

Regularization is introduced to prevent solution artifacts, arising from the ill-posedness of the problem that magnifies the noise in the data. One should realize, however, that the regularization also always tends to introduce other artifacts in the solution (compared to the exact and unattainable image). The hope is that the regularization artifacts are different, and that they have a less disturbing influence on the interpretation of the reconstructed image than the original noise artifacts. For example, if we use  $\|x\|_2^2$  as the regularizing function then we know that this leads to smooth reconstructions, and if we wish to reconstruct sharp edges in the image (i.e., pixels with large gradient magnitude) then we obtain severe Gibbs artifacts appearing as “ringing” effects near the edges. The TV function allows better reconstruction of edges, but at the expense of so-called “staircasing” or “cartoon” artifacts [2].

No matter which *numerical algorithm* is used to solve the TV problem, it starts with an initial guess  $x^{(0)}$  and produces a sequence of iterations or approximations  $x^{(k)}$  for  $k = 1, 2, \dots$  until some stopping criterion is satisfied. Standard stopping criteria are based on the change in the objective function  $f(x^{(k-1)}) - f(x^{(k)})$  and the step size  $\|x^{(k-1)} - x^{(k)}\|_2$ , and involve thresholds  $T_{\text{obj}}$  and  $T_{\text{step}}$  for these quantities. Alternatively one can stop when the angle  $\theta$  between the gradients of the two terms in (1) approaches  $\pi$ .

Step	Parameters associated with the step
The scanner	$d =$ dose (source’s intensity); $\nu = \#$ views (positions of source); $p =$ number of bins (or pixels) on detector.
Math. model	$m = \#$ data; $n = \#$ pixels; $\sigma =$ noise level.
Reconstr. model	$\alpha =$ reg. parameter; $\tau =$ smoothing threshold.
Numer. algorithm	$x^{(0)} =$ initial guess; $T_{\text{obj}}, T_{\text{step}}, T_\theta =$ thresholds for change in objective function, step length, or angle between gradients.

The table above summarizes the steps and the corresponding parameters that we have introduced here. Below we give examples of the the influence of these parameters on the computed reconstructions.

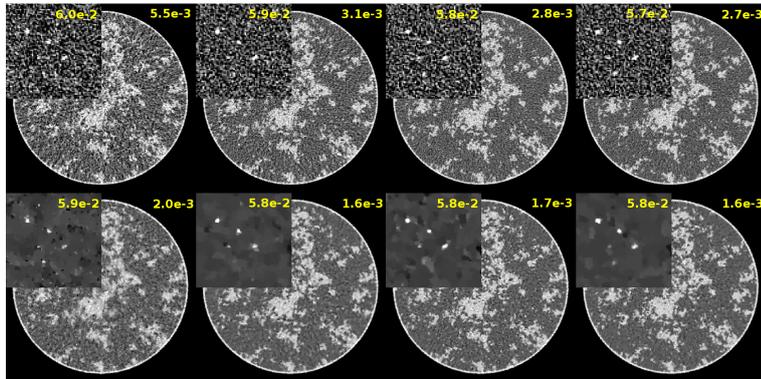


Fig. 1.

### Dose and number of views

In a number of applications the accumulated dose during the measurements must be limited — for example due to safety requirements in a medical scan or due to material limitations in nondestructive testing. This means that the product  $d \cdot \nu$  of the source’s intensity and the number of views is a constant. The signal-to-noise level (SNR) in the data is proportional to the source’s intensity, and therefore we can basically distinguish between two scenarios: many views with dense angular sampling but low SNR in each view, or few views with high SNR in each view but coarse angular sampling. A study of this aspect is given in [7].

The main dilemma in such a study is that when varying the scanner parameters we need to go through all the stages mentioned above to arrive at a reconstruction, making it difficult to make a completely fair comparison. For example, in our study we chose the TV regularization parameter  $\alpha$  by a visual inspection when varying  $d$  and  $\nu$ ; but how should this really be done to make a completely fair comparison?

We use a test image that simulates a cross section of a female breast consisting of four different tissue types: skin, fat, fibroglandular tissue (having a complex structure that is fairly realistic) and micro-calcifications, with different gray level intensities. Of particular interest are the micro-calcifications which are considered an early indicator of a developing cancer. Their tiny size and high contrast make accurate imaging a challenge.

CT screening for breast cancer is being developed as an supplement to conventional mammography, and to make CT feasible in this setting it is necessary to operate at a much lower X-ray dose than conventional CT. In the present study our particular question of interest was therefore: Given a fixed X-ray exposure to the patient (equivalent to mammography levels) what is the best distribution of the dose between the views? We compute noise-free data for  $\nu = 64, 128, 256$  and  $512$  views and manually add noise with increasing intensity to simulate the fixed accumulated dose across all views, i.e., more noise per view in the many view cases.

Figure 1 shows reconstructions computed with two different reconstruction models, filtered back projection (FBP, top) and total variation (TV, bottom), and with four different number of views ranging from  $\nu = 64$  (with high SNR in each view) to  $\nu = 512$  (with low SNR in each view). We also show a zoomed-in version of the region of interest around the micro-calcification structures.

We see that FBP tends to give results that improve slightly with  $\nu$  and with a lot of high-frequency “structure noise” (a well known artifact in FBP) while TV produces reconstructions whose visual appearance varies significantly with  $\nu$ . As expected, the “cartoon” artifacts dominate the TV reconstructions. As the SNR deteriorates we must increase the regularization parameter  $\alpha$  and hence the size of the piecewise

constant regions increases while their number decreases.

While most of the micro-calcifications are visible in each reconstruction, the artifacts and noise texture in the sparse-view images can be distracting and mistaken for additional micro-calcifications. The increased SNR per view impacts the reconstruction less than artifacts due to reduced sampling. Hence, with our choice of  $\alpha$  it appears that the micro-calcifications are better revealed in the reconstructions based on many low-SNR views. This result is interesting and warrants further investigation with more rigorous and quantitative evaluation.

### Number of views and bins

This case study from [9] illustrates the interplay by the scanner, the mathematical model, and the reconstruction model. The ill-posedness of the reconstruction problem can, to a certain extent, be measured by the condition number  $\text{cond}(A)$  of the system matrix, since this number describes the reconstruction’s sensitivity to data errors when regularization is not imposed. For the present simulations we considered fan-beam CT, assuming a circular path for the source. The reconstruction is an  $N \times N$  image with  $N = 32$ , but we fix all pixel values outside a circular region at 0 in order to match the rotational symmetry in the scan geometry, and the number of unknowns is therefore  $n \approx (\pi/4)N^2 = 812$ .

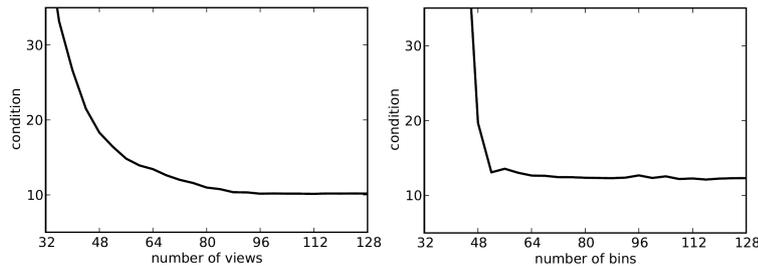


Fig. 2.

Figure 2 shows  $\text{cond}(A)$ , measured in the 2-norm, as a function of the number  $\nu$  of views and the number  $p$  of bins on the detector, for a discrete model with  $n = 812$  pixels in the reconstruction. The largest condition number for the considered sampling range is 825.5 occurring at  $\nu = p = 32$ . The large condition number for the lower number of samples implies that any data inconsistency could be amplified. The condition number decays fast with increasing  $p$  and slower with  $\nu$ . These results seem to suggest the choice  $\nu \approx 2N$  and  $p \approx 2N$  which ensures a small condition number. Increasing  $\nu$  or  $p$  only reduces the condition number marginally.

### TV versus 2-norm regularization

Another case study from [9] illustrates the influence of the regularization function in the reconstruction model; specifically, we compare the TV model (1) with a similar model where the TV term is replaced by the 2-norm  $\|x\|_2$  of the solution. The model problem is the same as above.

Figure 3 shows the root mean square error (RMSE), i.e., the 2-norm of the difference between the exact image and the reconstruction, as a function of the number of views, for the TV and the 2-norm regularizers in the reconstruction model. We already know that the two models give different artifacts in the reconstruction, so the RMSE does not tell the whole story; but the main observation here is that the TV model is able to give much lower RMSE than the 2-norm model as the number of views  $\nu$  decreases. In fact, the RMSE for TV is almost independent of  $\nu$  as long as

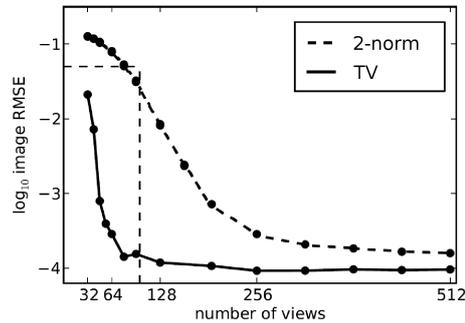


Fig. 3.

this number exceeds 100, while the RMSE increases dramatically for fewer views. The RMSE for the 2-norm regularizer, on the other hand, increases steadily as  $\nu$  decreases. The conclusion is that the TV regularization term represents strong a priori knowledge which is better able to compensate for the reduction in the amount of data than 2-norm regularization.

### The TV regularization parameter

The TV reconstruction model includes the parameter  $\alpha$  that controls the weight given to the regularization term, and studies in [13] demonstrate that  $\alpha$  acts as a “scale parameter” or “resolution limit” that controls the size of the smallest features that can be reconstructed by the TV model. This parameter depends on the noise level in the data and a too small value will result in a useless reconstruction that is contaminated by influence from the noise, while a too large value will result in a very “cartoony” reconstruction with too few details.

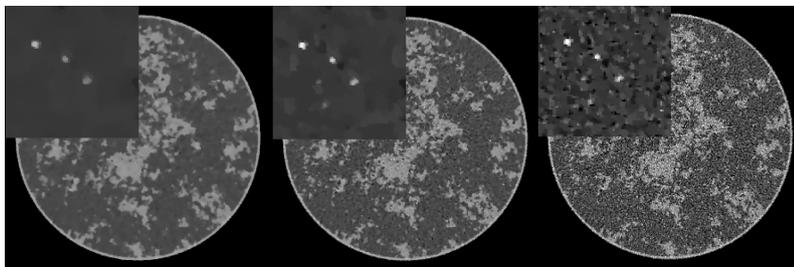
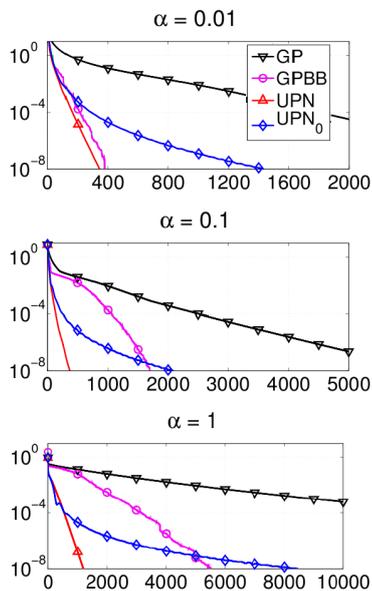


Fig. 4.

Figure 4 illustrates this aspect. In the left image,  $\alpha$  is too large—the inverted noise is suppressed but the regions of constant intensity are too large giving a “cartoony” reconstruction. In the right image,  $\alpha$  is too small such that  $f(x)$  is dominated by the residual term, and hence the solution is dominated by inverted noise due to the ill conditioned  $A$  matrix. The middle image contains more details without being influenced by the noise.

At the same time, the size of  $\alpha$  influences how “difficult” it is to solve the optimization problem (1) numerically: A small  $\alpha$  means the the objective function  $f(x)$  is dominated by the residual term which is smooth, while a large  $\alpha$  puts more emphasis on the less smooth TV term. An additional issue—which we will not consider here—is the choice of the smoothing threshold  $\tau$  that controls how much we smooth the TV term in (2).



**Figure 5** (from [6]) shows the convergence of four numerical algorithms for solving the TV regularization problem (1) for three choices of  $\alpha$ . Note the different iteration ranges on the abscissa axis! The four algorithms are:

**GP** The standard gradient projection algorithm.

**GPBB** GP with Barzilai–Borwein acceleration.

**UPN<sub>0</sub>** An optimal first-order method from [6] *not* exploiting strong convexity.

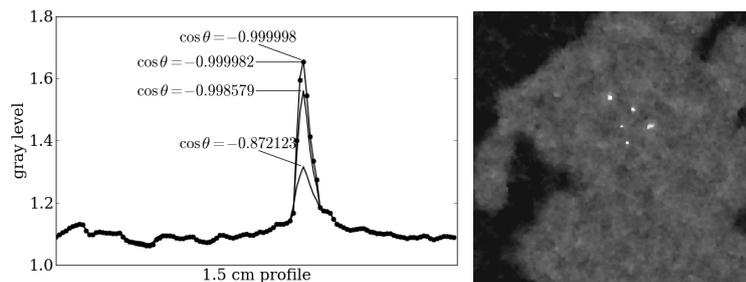
**UPN** An optimal first-order method from [6] that exploits strong convexity.

As  $\alpha$  increases, the TV regularization term in  $f(x)$  becomes increasingly important and the problem becomes more difficult to solve, resulting in an increasing number of iterations for all four methods to reach a solution of the desired accuracy

In all three cases GPBB is superior to GP, and for large values of  $\alpha$  the two first-order methods are even faster. The four methods differ by the amount of “information” about the optimization problem that they exploit. Ranging from GP—a basic steepest-descent type method—to UPN, which adaptively estimates the Lipschitz constant and the strong convexity parameter of the objective function.

### The stopping criterion

Our last example shows the influence of the stopping criterion on the TV reconstruction. The stopping criterion used here is based on the optimality criterion  $\cos \theta = -1$ , where  $\theta$  is the angle between the gradients of the squared residual and the TV regularization term [12], and we stop when  $\cos \theta$  is sufficiently close to  $-1$ . The algorithm used here is GPBB from the previous section.



**Fig. 6.**

Figure 6 (from [10]) shows a particular profile through a single micro-calcification, see the inserted image, for different iterations that are increasingly close to satisfying  $\cos \theta = -1$ . As the number of iterations increases the profile’s sharp peak gets better resolved. Interestingly, the low-frequency components of the profile are captured already after a small number of iterations, while many more iterations are needed to capture the correct shape and magnitude of the peak. The TV reconstruction model focuses on providing an accurate representation of the image’s gradient, and our exam-

ple shows that it is important to be close to the minimum of  $f(x)$  in order to achieve this.

An important point is that it is unclear precisely *how* close  $\cos\theta$  should be to  $-1$ , and whether a different stopping criterion, e.g., exploiting local information around the micro-calcification, could be more reliable. On the other hand, we could simply take an extremely large number of iterations to ensure an accurate solution, but in practice this is of course not feasible. Accepting an inadequate reconstruction can have clinical impact, as it might fail to provide enough contrast for detecting the micro-calcification.

## Acknowledgement

The research presented here was carried out in collaboration with Dr. E. Y. Sidky and Prof. X. Pan from University of Chicago and T. L. Jensen from Aalborg University. We also thank Dr. S. Schmidt from Risø DTU for his insight.

## References

- [1] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory, 52 (2006), pp. 489–509.
- [2] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational PDE, Wavelet, and Stochastic Methods*, SIAM, PA, 2005.
- [3] D. L. Donoho, *Compressed Sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.
- [4] H. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 2000.
- [5] G. T. Herman, *Fundamentals of Computerized Tomography*, Springer, Berlin, 2009.
- [6] T. L. Jensen, J. H. Jørgensen, P. C. Hansen, and S. H. Jensen, *Implementation of an optimal first-order method for strongly convex total variation regularization*, BIT, to appear. Available from [www.imm.dtu.dk/~pch/TVReg](http://www.imm.dtu.dk/~pch/TVReg).
- [7] J. H. Jørgensen, P. C. Hansen, E. Y. Sidky, I. S. Reiser, and X. Pan, *Toward optimal X-ray flux utilization in breast CT*; in proceedings of 11th Fully 3D Meeting, Potsdam, Germany, July 2011.
- [8] J. H. Jørgensen, T. L. Jensen, P. C. Hansen, S. H. Jensen, E. Y. Sidky, and X. Pan, *Accelerated gradient methods for total-variation-based CT image reconstruction*; in proceedings of 11th Fully 3D Meeting, Potsdam, Germany, July 2011.
- [9] J. H. Jørgensen, E. Y. Sidky, and X. Pan, *Analysis of discrete-to-discrete imaging models for iterative tomographic image reconstruction and compressive sensing*; submitted to IEEE Transactions on Medical Imaging. Available from <http://arxiv.org/abs/1109.0629>.
- [10] J. H. Jørgensen, E. Y. Sidky, and X. Pan, *Ensuring convergence in total-variation-based reconstruction for accurate microcalcification imaging in breast X-ray CT*, Proceedings of the 2011 IEEE Nuclear Science Symposium and Medical Imaging Conference, Valencia, Spain.
- [11] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
- [12] E. Y. Sidky and X. Pan, *Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization*, Phys. Med. Biol., 53 (2008), pp. 4777–4807.
- [13] D. Strong, and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Prob., 19 (2003), pp. S165–S187.