

PAPER

Controlled wavelet domain sparsity for x-ray tomography

To cite this article: Zenith Purisha *et al* 2018 *Meas. Sci. Technol.* **29** 014002

View the [article online](#) for updates and enhancements.

Related content

- [Sparsity-promoting Bayesian inversion](#)
V Kolehmainen, M Lassas, K Niinimäki *et al*.
- [SparseBeads data: benchmarking sparsity-regularized computed tomography](#)
Jakob S Jørgensen, Sophia B Coban, William R B Lionheart *et al*.
- [Statistical image reconstruction from limited projection data with intensity priors](#)
Essam A Rashed and Hiroyuki Kudo

Recent citations

- [Advanced x-ray tomography: experiment, modeling, and algorithms](#)
K Joost Batenburg *et al*



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Controlled wavelet domain sparsity for x-ray tomography

Zenith Purisha^{1,2} , Juho Rimpeläinen¹, Tatiana Bubba¹
and Samuli Siltanen¹

¹ Department of Mathematics and Statistics, University of Helsinki, Finland

² Department of Mathematics, Universitas Gadjah Mada, Indonesia

E-mail: zenith.purisha@helsinki.fi

Received 29 June 2017, revised 26 September 2017

Accepted for publication 10 October 2017

Published 14 December 2017



Abstract

Tomographic reconstruction is an ill-posed inverse problem that calls for regularization. One possibility is to require sparsity of the unknown in an orthonormal wavelet basis. This, in turn, can be achieved by variational regularization, where the penalty term is the sum of the absolute values of the wavelet coefficients. The primal-dual fixed point algorithm showed that the minimizer of the variational regularization functional can be computed iteratively using a soft-thresholding operation. Choosing the soft-thresholding parameter $\mu > 0$ is analogous to the notoriously difficult problem of picking the optimal regularization parameter in Tikhonov regularization. Here, a novel automatic method is introduced for choosing μ , based on a control algorithm driving the sparsity of the reconstruction to an *a priori* known ratio of nonzero versus zero wavelet coefficients in the unknown.

Keywords: sparsity, wavelet, regularization, control, limited data tomography, x-ray tomography

(Some figures may appear in colour only in the online journal)

1. Introduction

Tomographic imaging is based on recording projection images of an object along several directions of view [1–3]. The resulting data can be interpreted as a collection of line integrals of an unknown attenuation coefficient function $f(x)$. In this work, we discretize the problem by approximating f as a vectorized pixel image $\mathbf{f} \in \mathbb{R}^{N^2}$ and using the pencil-beam model for x-rays, so the indirect measurement is modelled by a matrix equation $\mathbf{A}\mathbf{f} = \mathbf{m}$. The inverse problem of reconstructing \mathbf{f} from tomographic data is highly sensitive to noise and modelling errors, or in other words, it is *ill-posed*.

We focus on overcoming ill-posedness by enforcing sparsity of \mathbf{f} with respect to a linear sparsifying transform. Among many possible choices, we consider an orthonormal wavelets basis, which provides a reasonable trade-off between computational cost and quality of the reconstruction, with respect to more recent directional-aware sparsifying transforms.

In practice, the sparse reconstruction $\mathbf{f}_s \in \mathbb{R}^{N^2}$ is defined as the minimizer of this variational regularization functional:

$$\mathbf{f}_s = \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^{N^2}} \left\{ \frac{1}{2} \|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 + \mu \|\mathbf{W}\mathbf{f}\|_1 \right\}, \quad (1)$$

where the matrix \mathbf{W} is a digital implementation of the wavelet transform we consider.

The parameter μ in (1) describes a trade-off between emphasizing more the data fidelity term or the regularizing penalty term. In general, the larger the noise amplitude in the data, the larger μ needs to be.

One popular method to solve the problem (1) is the so-called iterative soft-thresholding algorithm (ISTA). Such an algorithm has been studied already in [4]; the adaptation to sparsity-promoting inversion was introduced in [5] and further developed in [6]. Nevertheless, the convergence rate for a constrained problem, such as non-negativity constraints, is not taken into account in [5, 6]. However, in tomographic

problems, enforcing non-negativity on the attenuation coefficients is highly desirable. This is based on the physical fact that the x-ray radiation can only attenuate inside the target, not strengthen. Thus, the problem we need to solve reads as

$$\mathbf{f}_S = \underset{\mathbf{f} \in \mathbb{R}^{N^2}, \mathbf{f} \geq 0}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 + \mu \|\mathbf{W}\mathbf{f}\|_1 \right\}, \quad (2)$$

where the inequality $\mathbf{f} \geq 0$ is meant component-wise. In their seminal paper [7], Peijun Chen, Jianguo Huang, and Xiaoqun Zhang show that the minimizer of (2) can be computed using the primal-dual fixed point (PDFP) algorithm:

$$\begin{aligned} \mathbf{y}^{(i+1)} &= \mathbb{P}_C \left(\mathbf{f}^{(i)} - \tau \nabla g(\mathbf{f}^{(i)}) - \lambda \mathbf{W}^T \mathbf{v}^{(i)} \right) \\ \mathbf{v}^{(i+1)} &= \left(\mathbf{I} - \mathcal{T}_\mu \right) \left(\mathbf{W} \mathbf{y}^{(i+1)} + \mathbf{v}^{(i)} \right) \\ \mathbf{f}^{(i+1)} &= \mathbb{P}_C \left(\mathbf{f}^{(i)} - \tau \nabla g(\mathbf{f}^{(i)}) - \lambda \mathbf{W}^T \mathbf{v}^{(i+1)} \right), \end{aligned} \quad (3)$$

where τ and λ are positive parameters, $g(\mathbf{f}) = \frac{1}{2} \|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2$, and \mathcal{T} is the soft-thresholding operator defined by

$$\mathcal{T}_\mu(c) = \begin{cases} c + \frac{\mu}{2} & \text{if } c \leq -\frac{\mu}{2} \\ 0 & \text{if } |c| < \frac{\mu}{2} \\ c - \frac{\mu}{2} & \text{if } c \geq \frac{\mu}{2}. \end{cases} \quad (4)$$

Here $\mu > 0$ represents the thresholding parameter, while τ and λ are parameters that need to be suitably chosen to guarantee convergence. In detail, $0 < \lambda < 1/\lambda_{\max}(\mathbf{W}\mathbf{W}^T)$, where λ_{\max} denotes the maximum eigenvalue, and $0 < \tau < 2/\tau_{\text{lip}}$, being τ_{lip} the Lipschitz constant for $g(\mathbf{f})$. Furthermore, in (3) the non-negative ‘quadrant’ is denoted by $C = \mathbb{R}_+^{N^2}$ and \mathbb{P}_C is the Euclidian projection. In other words, \mathbb{P}_C replaces any negative elements in the input vector by zero.

Choosing the soft-thresholding parameter μ is analogous to the notoriously difficult problem of picking the optimal regularization parameter in Tikhonov regularization. Many approaches for the regularization parameter selection have been proposed. For a selection of methods designed for total variation (TV) regularization see the following studies: [8–18]. An automatic regularization parameter choice technique using a sparsity-promoting penalty for image denoising is studied in [19]. Iterative hard-thresholding for compressed sensing has been studied in [20].

In this paper, we introduce a novel automatic method for choosing μ based on a control algorithm driving the sparsity of the reconstruction to an *a priori* known ratio $0 \leq \mathcal{C}_{\text{pr}} \leq 1$ of nonzero wavelet coefficients in \mathbf{f} . Our approach is based on the following idea: *in sparsity-promoting regularization, it is natural to assume that the a priori information is given as the percentage of nonzero coefficients in the unknown*. The idea of using the *a priori* known level of sparsity has been used previously [21, 22], however, the idea of using feedback control to achieve this is new.

We think of the iteration (3) as a *plant* which takes the current threshold parameter $\mu^{(i)}$ as an input and returns $\mathcal{C}^{(i)}$, the level of sparsity in the iterate $\mathbf{f}^{(i)}$, as an output. Then, we apply a simple incremental feedback control to $\mu^{(i)}$. The feedback loop we propose is inspired by proportional-integral-derivative

(PID) controllers, which are widely used to control industrial processes [23–25]. If \mathcal{C}_{pr} is the expected degree of sparsity, and $\mathcal{C}^{(i)}$ is the degree of sparsity at the current iterate i , we change $\mu^{(i)}$ adaptively as follows:

$$\mu^{(i+1)} = \mu^{(i)} + \beta(\mathcal{C}^{(i)} - \mathcal{C}_{\text{pr}}), \quad (5)$$

where $\beta > 0$ is a parameter used to tune the controller. We propose a simple method for choosing β based on the wavelet coefficients of the back-projection reconstruction, which is quick and easy to compute. If the β chosen is too large, then the controller results in an oscillating behavior for the sequence $(\mu^{(i)})_i$. However, if the chosen β is too small, reaching the expected sparsity level may take a long time. Therefore we also account for an additional fine-tuning of the controller by exploiting the zero-crossings of the controller error $\mathbf{e}^{(i)} = \mathcal{C}^{(i)} - \mathcal{C}_{\text{pr}}$.

We test our *fully* automatic controlled wavelet domain sparsity (CWDS) method on both simulated and real tomographic data. The results suggest that our method produces robust and accurate reconstructions, when the suitable degree of sparsity is available.

CWDS has a connection to the following studies, which also uses a parameter changing adaptively during the iterations: [26–32]. However, our approach is different from the others as it promotes an *a priori* known level of sparsity. Also, this is not the first study which has used the wavelet transform as a regularization tool in limited data tomography. A non-exhaustive list includes [22, 33–38]. However, the proposed approach is different from the previous works, since it promotes a *fully* automatic choice for the regularization parameter.

2. Materials and methods

2.1. Tomography setup

Consider a physical domain $\Omega \subset \mathbb{R}^2$ and a non-negative attenuation function $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}_+$. As outlined in the introduction, we represent f by a matrix $\mathbf{f} = [\tilde{f}_{ij}] \in \mathbb{R}^{N \times N}$ that is later on intended as a vector belonging to \mathbb{R}^{N^2} , obtained by stacking the entries of the matrix column by column. In x-ray tomography, the detector measures the incoming photons and the measurement data are collected from the intensity losses of x-rays from different directions or angles of view. After calibration, the measurements can be modeled as

$$\int_{L_X} f(x) ds = \sum_{i=1}^N \sum_{j=1}^N a_{ij} \tilde{f}_{ij},$$

where a_{ij} is the distance that an x-ray line L_X travels through the pixel (i, j) . This results in the following matrix model:

$$\mathbf{m} = \mathbf{A}\mathbf{f}, \quad (6)$$

where the measurement matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{P \times N^2}$ contains information about the measurement geometry, and $\mathbf{m} \in \mathbb{R}^P$ is the vector representing the measured data (also called *sino-gram*), P being the number of angles of view multiplied by the number of detector cells.

Notice that in the following, we assume both the measurement matrix \mathbf{A} and the measured data \mathbf{m} to be normalized by the norm $\|\mathbf{A}\|$ of the matrix \mathbf{A} .

2.2. 2D Haar wavelets

For the sake of the readers' convenience, we briefly recall here the main ideas about Haar wavelets.

Consider the two real-valued functions $\varphi(x)$ and $\psi(x)$ defined on the interval $[0, 1]$. Generally, $\varphi(x)$ is referred to as the *scaling function* and $\psi(x)$ as the *mother wavelet*. They are defined as follows:

$$\varphi(x) \equiv 1, \quad \psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1/2, \\ -1 & \text{if } 1/2 \leq x \leq 1. \end{cases}$$

A *discrete Haar wavelet system*, where discrete means that the transform is associated with a discrete parameter set, is built by appropriately scaling and translating the mother wavelet $\psi(x)$:

$$\psi_{jk}(x) := 2^{j/2} \psi(2^j x - k) \quad \text{for } j \leq 0, 0 \leq k \leq 2^j - 1,$$

and the scaling function $\varphi(x)$:

$$\varphi_{jk}(x) := 2^{j/2} \varphi(2^j x - k) \quad \text{for } j \leq 0, 0 \leq k \leq 2^j - 1,$$

where $\varphi(x) = 0$ for $x < 0$ and $x > 1$. Here, $j, k \in \mathbb{Z}$.

It is well known that the above 1D construction leads to an orthonormal system. In 2D, we consider the standard tensor-product extension of the 1D Haar wavelet transform. In detail, a 2D Haar system is spanned by four types of functions. Three of these types have the following form:

$$\varphi_{jk}(x)\psi_{jk}(y), \quad \psi_{jk}(x)\varphi_{jk}(y), \quad \psi_{jk}(x)\psi_{jk}(y), \quad (7)$$

and the fourth type is given by $\varphi_{j_0 k}(x)\varphi_{j_0 k}(y)$. Notice that the fourth type describes the coarsest scale j_0 . The associated matrix underlying the *discrete wavelet transform* of a function \mathbf{f} is given by

$$\mathbf{W} = \begin{bmatrix} W_\varphi & W_\psi^V \\ W_\psi^H & W_\psi^D \end{bmatrix} \in \mathbb{R}^{N^2 \times N^2}$$

where

$$W_\varphi = \frac{1}{\sqrt{RS}} \sum_{r=1}^R \sum_{s=1}^S \mathbf{f}\{q\} \varphi_{j_0 k}\{r\} \varphi_{j_0 k}\{s\} \quad (8)$$

$$W_\psi^H = \frac{1}{\sqrt{RS}} \sum_{r=1}^R \sum_{s=1}^S \mathbf{f}\{q\} \psi_{jk}\{r\} \varphi_{jk}\{s\} \quad (9)$$

$$W_\psi^V = \frac{1}{\sqrt{RS}} \sum_{r=1}^R \sum_{s=1}^S \mathbf{f}\{q\} \varphi_{jk}\{r\} \psi_{jk}\{s\} \quad (10)$$

$$W_\psi^D = \frac{1}{\sqrt{RS}} \sum_{r=1}^R \sum_{s=1}^S \mathbf{f}\{q\} \psi_{jk}\{r\} \psi_{jk}\{s\} \quad (11)$$

with $q = r + R * (s - 1)$, and the brackets $\{\cdot\}$ indicating that we now intend also ϕ and ψ to be discrete functions defined in the intervals $[1, R]$ or $[1, S]$, being $R, S \in \mathbb{Z}$.

Thus, the vector collecting all the wavelet coefficients is given by

$$\mathbf{W}\mathbf{f} \in \mathbb{R}^{N^2}. \quad (12)$$

With the above notation, the minimization problem (2) reads

$$\mathbf{f}_S = \underset{\mathbf{f} \in \mathbb{R}_+^{N^2}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 + \mu \|\mathbf{W}\mathbf{f}\|_1 \right\}. \quad (13)$$

One of the main benefits of wavelets is that the transform coefficients are easy to compute and many fast algorithmic implementations are available.

For more information about the Haar wavelet transform and its implementation, we refer to the classic text [39].

2.3. Sparsity promoting-regularization

We consider the functional in (2) with $\{\psi_\gamma\}_{\gamma \in \Gamma}$ being the Haar wavelet basis as described in section 2.2. To solve the minimization problem (2), we implement the PDFP algorithm (3).

2.4. Sparsity selection

We assume that we have an object available \mathbf{f}_{pr} similar to the one we are imaging.

Given $\kappa \geq 0$, for a vector $\mathbf{w} \in \mathbb{R}^{N^2}$ we define the number of elements larger than κ in absolute value as follows:

$$\#_\kappa \mathbf{w} := \#\{i \mid 1 \leq i \leq N^2, |w_i| > \kappa\}.$$

Now, the prior sparsity level is defined by

$$\mathcal{C}_{\text{pr}} = \frac{\#_\kappa \{\mathbf{W}\mathbf{f}_{\text{pr}}\}}{N^2},$$

where N^2 is the total number of coefficients. In practical computations, the value of κ is set to be small but positive.

2.5. Automatic selection of the soft-thresholding parameter μ

Assume that we know *a priori* the expected degree of sparsity $0 < \mathcal{C}_{\text{pr}} \leq 1$ in the reconstruction. We introduce a simple feedback loop to drive the soft-thresholding parameter μ to the desired ratio \mathcal{C}_{pr} of nonzero wavelet coefficients.

The core idea is to allow $\mu = \mu^{(i)}$ to vary during the iterations by adaptively tuning it at each iteration by the following updating rule:

$$\mu^{(i+1)} := \mu^{(i)} + \beta(\mathcal{C}^{(i)} - \mathcal{C}_{\text{pr}}),$$

where $0 \leq \mathcal{C}^{(i)} \leq 1$ is the sparsity level of the reconstruction $\mathbf{f}^{(i)}$ at the i th iteration. The above controller is a special case of an incremental PID-controller, where only integral control is performed.

2.6. The tuning parameter β

Selecting the tuning parameter β is easier than selecting the soft-thresholding parameter μ . Indeed, β has to be small enough to avoid oscillations in the sparsity $\mathcal{C}^{(i)}$ of the iterates

as a function of i . If the chosen β is too small, this only results in a slower convergence of the algorithm.

To this purpose, we choose β by making a suitable estimate for the initial $\mu^{(0)}$. First, we compute the back-projection of the measured data to get a rough reconstruction. Back-projection is quick to compute and shows the dominant features of the target, but noise and artefacts are still predominant, especially when only a few projection views are available. As a result, the back-projection reconstruction is only good enough for estimating an initial estimate for $\mu^{(0)}$, which is done by computing its wavelet coefficients. The initial value of the thresholding parameter $\mu^{(0)}$ is set equal to the mean of the absolute values of the M smallest wavelet coefficients. In our case, we choose $M = n(1 - \mathcal{C}_{pr})$, where n is the total number of wavelet coefficients. Lastly, the tuning parameter is set to be $\beta = \omega\mu^{(0)}$, where ω is a positive parameter. To start with a small value of β , ω is required to be small, and vice versa.

In addition, the controller is fine-tuned by detecting when the sign of difference $\mathbf{e}^{(i)} = \mathcal{C}^{(i)} - \mathcal{C}_{pr}$ changes. When this happens, β is updated by $\beta|\mathbf{e}^{(i)} - \mathbf{e}^{(i-1)}|$. The underlying idea is that, if the desired sparsity level is crossed, that is, e changes sign, either β is far too large and oscillations have emerged, or we are already reasonably close to the optimal μ , and β can be decreased without affecting the performance too much.

2.7 Pseudo-algorithm

A step-by-step description of the proposed CWDS algorithm is summarized in algorithm 1.

Algorithm 1. Controlled wavelet domain sparsity algorithm.

- 1: Inputs: measurement data vector \mathbf{m} , system matrix \mathbf{A} , parameters τ , $\lambda > 0$ to ensure convergence, *a priori* degree of sparsity \mathcal{C}_{pr} , initial thresholding parameter $\mu^{(0)}$, maximum number of iterations $I_{\max} > 0$, tolerances $\epsilon_1, \epsilon_2 > 0$ for the stopping rule and control stepsize $\beta > 0$.
- 2: $\mathbf{f}^{(0)} = \mathbf{0}$, $i = 0$, $e = 1$, and $\mathcal{C}^{(0)} = 1$
- 3: **while** $i < I_{\max}$ and $|e| \geq \epsilon_1$ or $d \geq \epsilon_2$ **do**
- 4: $e = \mathcal{C}^{(i)} - \mathcal{C}_{pr}$
- 5: **if** $\text{sign}(\mathbf{e}^{(i+1)}) \neq \text{sign}(\mathbf{e}^{(i)})$ **then**
- 6: $\beta = \beta(1 - |\mathbf{e}^{(i+1)} - \mathbf{e}^{(i)}|)$
- 7: $\mu^{(i+1)} = \max\{0, \mu^{(i)} + \beta e\}$
- 8: $\mathbf{y}^{(i+1)} = \max\{0, \mathbf{f}^{(i)} - \gamma \nabla g(\mathbf{f}^{(i)}) - \lambda \mathbf{W}^T \mathbf{v}^{(i)}\}$
- 9: $\mathbf{v}^{(i+1)} = (\mathbf{I} - \mathcal{T}_{\mu^{(i)}})(\mathbf{W}\mathbf{y}^{(i+1)} + \mathbf{v}^{(i)})$
- 10: $\mathbf{f}^{(i+1)} = \max\{0, \mathbf{f}^{(i)} - \gamma \nabla g(\mathbf{f}^{(i)}) - \lambda \mathbf{W}^T \mathbf{v}^{(i+1)}\}$
- 11: $\mathcal{C}^{(i+1)} = N^{-2} \#_{\kappa}(\mathbf{W}\mathbf{f}^{(i+1)})$
- 12: $d = \|\mathbf{f}^{(i+1)} - \mathbf{f}^{(i)}\|_2 / \|\mathbf{f}^{(i+1)}\|_2$
- 13: $i := i + 1$

3. Data acquisition

In this paper, we consider both simulated data (see section 3.1) and real data (see section 3.2).

Table 1. Initial values $\mu^{(0)}$ of the soft-thresholding parameter.

Target	120 data	30 data
Shepp–Logan	0.0202	0.0195
Walnut	0.0019	0.0021



Figure 1. The Shepp–Logan phantom, sized 328×328 , generated with Matlab.

3.1. Simulated data

We use the Shepp–Logan phantom, available, for instance, in the Matlab image processing toolbox (see figure 1). The phantom is sized $N \times N$, with $N = 328$. The projection data (i.e. sinogram) of the simulated phantom is corrupted by a white Gaussian process with zero mean and 0.1% variance.

3.2. Real data

We use the tomographic x-ray real data of a walnut, consisting of a 2D cross-section of a real 3D walnut measured with a custom-built CT device available at the University of Helsinki (Finland). The dataset is available and freely downloadable at <http://fips.fi/dataset.php>. For detailed documentation of the acquiring setup, including the specifications of the x-ray systems, see [40]. Here, we only mention that the sinogram is sized 328×120 . Sinograms with different resolutions for the angle of view can be obtained by further downsampling.

4. Numerical experiments

In this section, we present numerical results in the framework of 2D fan-beam geometry.

4.1. Algorithm parameters

In all the experiments, we set $\lambda = 0.99$ (being $\lambda_{\max}(\mathbf{W}\mathbf{W}^T) = 1$) and $\tau = 1$ to ensure convergence. Also, we choose $\epsilon_1 = 5 \times 10^{-4}$ and $\epsilon_2 = 5 \times 10^{-4}$ for the stopping rule, and $I_{\max} = 1500$ as a safeguard maximum number of iterations (which is never attained in the results reported in section 4.3), $\beta = \omega\mu^{(0)}$, where $\omega = 1$ and the values of $\mu^{(0)}$ for each experiments are shown in table 1.

All the algorithms were implemented in Matlab 8.5 (R2015a) and performed on Intel Core i5 at 2.9 GHz and CPU 8GB 1867MHz DDR3 memory. The Haar matrix \mathbf{W} is generated by using Spot–A Linear-Operator Toolbox [41]. The number of scales for the wavelet transform is set equal to 3 (see figure 2).

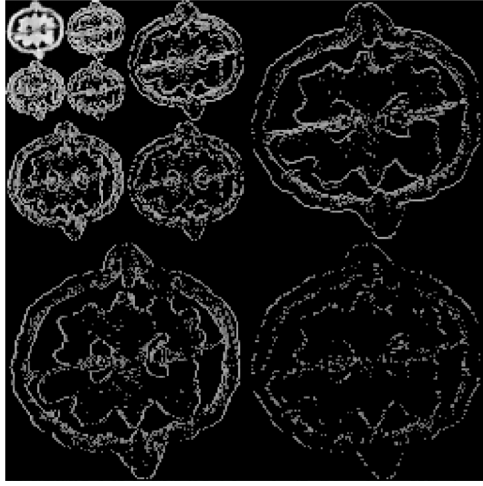


Figure 2. Wavelet transform of the left photograph in figure 3. The original image is high-pass filtered, yielding the three large images. It is then low-pass filtered and downsampled, yielding an approximation image; this image is high-pass filtered to produce three smaller detailed images, and low-pass filtered to produce the final approximation image in the upper-left corner.

4.2. A priori sparsity level

To compute the desired sparsity level, we choose $\kappa = 10^{-6}$ for both the Shepp–Logan phantom and the walnut, and we apply the strategy outlined in section 2.5. In particular, for the walnut case, since we do not have at our disposal the ‘original’ target, we compute the sparsity level from the photographs of two walnuts cut in half (see figure 3). The *a priori* sparsity level C_{pr} for the walnut is the average of those two sparsity levels.

For the Shepp–Logan phantom, the percentage of nonzero coefficients was estimated to be 12%. The percentage of the nonzero coefficients for the walnut case was estimated to be 32%.

4.3. Reconstruction results

In this section, we present numerical results for the CWDS method, using both simulated and real data. As a benchmark comparison, we computed the reconstructions also with filtered back-projection (FBP) and with the anisotropic total variation (TV) approach. For both the simulated and real data, we computed the reconstructions for two different resolutions of the angle of view, namely 120 and 30 projection directions, respectively.

The FBP and CWDS reconstructions of the Shepp–Logan phantom are shown in figure 4. Plots of the sparsity levels, as the iteration progresses, are reported in figure 8. For the 120 projections case, the proposed approach converges in 885 iterations, while in the 30 projections case, it converges in 301 iterations. As a figure of merit, we use the relative error: the obtained values are summarized in table 2, where we also report the values of the relative error obtained for the FBP reconstructions.

Concerning the TV approach, we consider smoothed TV (sTV) [42], where a small quantity δ is considered to remove the singularity of the discrete gradient:

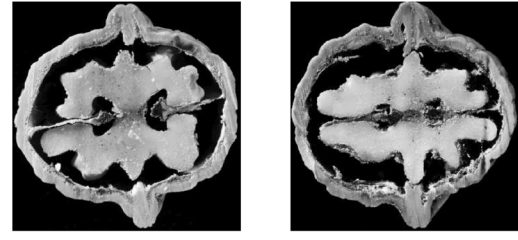


Figure 3. Photographs of the walnuts split in half. The sparsity level of each image was calculated to provide the *a priori* information of the sparsity level for the measured walnut. The above photographs do not include the measured walnut.

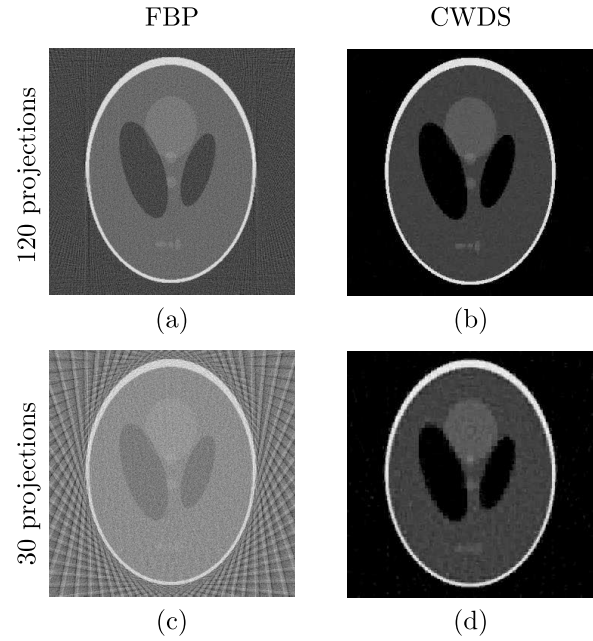


Figure 4. Reconstructions of the Shepp–Logan phantom using FBP with (a) 120 projections, and (c) 30 projections. Reconstructions using CWDS with (b) 120 projections, and (d) 30 projections.

Table 2. The relative error of the Shepp–Logan phantom reconstructions for FBP and CWDS.

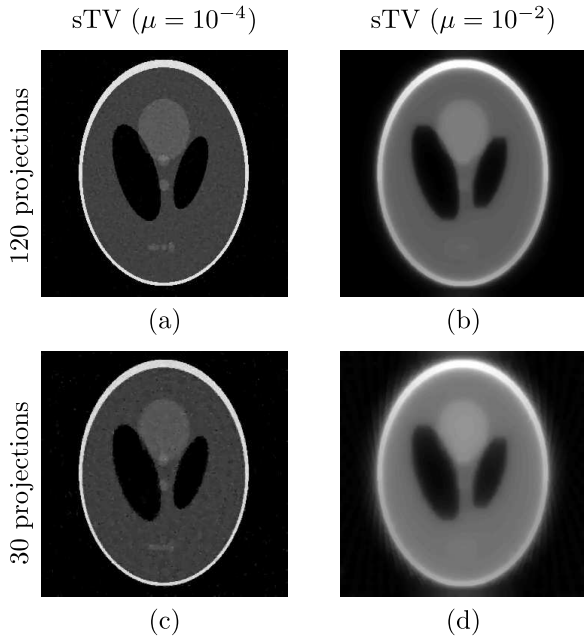
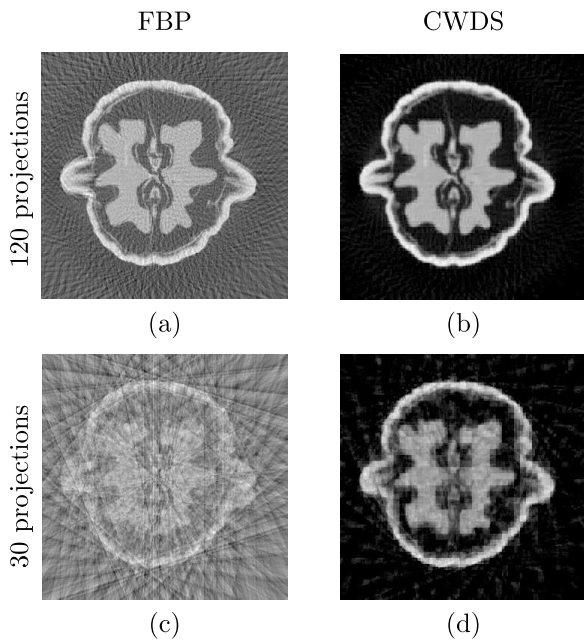
Method	120 data	30 data
FBP	0.15	0.27
CWDS	0.04	0.08

$$\text{TV}_\delta(\mathbf{f}) = \sum_{i=1}^{N^2} \left\| \begin{pmatrix} \nabla_i \mathbf{f} \\ \delta \end{pmatrix} \right\|_2$$

where $\nabla_i \in \mathbb{R}^{2 \times N^2}$ is the i th submatrix of the discrete gradient operator ∇ , namely $(\nabla \mathbf{f})_i = \nabla_i \mathbf{f}$. Similar to wavelets, the TV operator can be understood as a sparsifying transformation, which transforms the original image into an edge map. The reconstructions are computed using the gradient descent method, with the Barzilai–Borwein [43] strategy for the step length selection. For the regularization parameter, we sampled ten different values $\mu \in [10^{-8}, 10^1]$ and the smoothing parameter δ is set equal to 10^{-6} . For the simulated Shepp–Logan phantom, we computed the relative error: the values for each reconstruction are reported in table 3. One of the drawbacks

Table 3. The relative error for the Shepp–Logan reconstructions using sTV.

μ	120 data	30 data
10^{-4}	0.014	0.044
10^{-2}	0.236	0.289

**Figure 5.** sTV reconstructions of the Shepp–Logan phantom using $\mu = 10^{-4}$ with (a) 120 projections and (c) 30 projection. sTV reconstructions of the Shepp–Logan phantom using $\mu = 10^{-2}$ with (b) 120 projections and (d) 30 projection.**Figure 6.** Reconstructions of the walnut using FBP with (a) 120 projections, and (c) 30 projections. Reconstructions of the walnut using CWDS with (b) 120 projections, and (d) 30 projections.**Table 4.** Computation times for FBP reconstructions and CWDS reconstructions, in seconds.

Target	Method	120 data	30 data
Walnut	FBP	0.45	0.09
	CWDS	17.40	16.30
Shepp–Logan	FBP	0.02	0.01
	CWDS	98.90	29.50

of the standard sTV approach, compared to the CWDS, is that to select the ‘best’ reconstruction, quite often in practice, one computes the reconstructions for different values of the regularization parameter μ and then the ‘best’ one is manually chosen, by selecting the image that yields the smallest relative error or, when such a measure is not available, the one which seems most appealing through visual inspection. To show how sensible the quality of the reconstruction is to the choice of the regularization parameter, we report sTV reconstructions for the Shepp–Logan phantom using two different values of μ : 10^{-4} , corresponding to the smallest relative errors and 10^{-2} to bigger relative errors, as presented in table 3. The reconstructions are collected in figure 5.

Similarly for the walnut, we report in figure 7 sTV reconstructions for two different values of the regularization parameter μ . Once again, we show the best and a rather poor result, this time choosing through visual inspection. The FBP and CWDS reconstructions for the walnut dataset, for both 120 and 30 projections, are collected in figure 6. The corresponding sparsity plots are shown in figure 10. Concerning the number of iterations to convergence, the 120 projections case required 180 iterations, while in the 30 projections case convergence was reached in 206 iterations.

Lastly, the computation times for all the reconstructions are reported in table 4.

5. Discussion

We presented results for both simulated and real x-ray data, also in the limited data case of only 30 projection views, with the *fully* automatic CWDS method. As can be seen in figures 4 and 6, the reconstructions for both the Shepp–Logan phantom and the walnut data outperform the FBP reconstructions. For the Shepp–Logan case, this is confirmed by the relative errors reported in table 2. In detail, the reconstructions using CWDS produce sharper images, with less artefacts. Overall, the quality of the reconstruction remains good even when the number of projections is reduced to 30, while, for the FBP reconstructions, streak artefacts overwhelm the reconstructions. Finally, the presence of ℓ_1 -norm term combined with a sparsity transform, that produce denoising, and the non-negativity constraint (which is not enforced in the classical FBP scheme) definitively improves the reconstructions.

The image quality of the CWDS reconstructions are comparable to those obtained using sTV. However, for

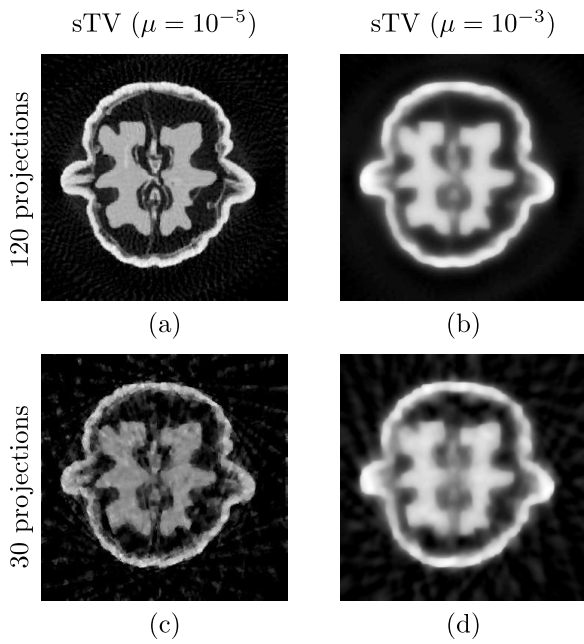


Figure 7. sTV reconstructions of the walnut using $\mu = 10^{-5}$ with (a) 120 projections and (c) 30 projection. sTV reconstructions of the walnut using $\mu = 10^{-3}$ with (a) 120 projections and (c) 30 projections.

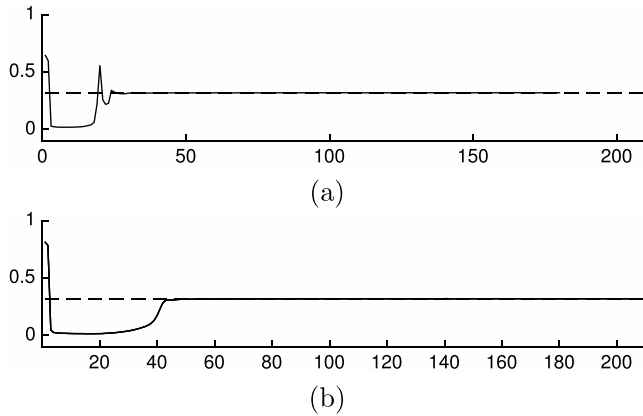


Figure 8. The ratio of nonzero wavelet coefficients as the iteration progresses for the walnut case. (a) 120 projections. (b) 30 projections. The dashed line shows the sparsity prior C_{pr} .

the sTV reconstructions, we can obtain satisfying results only if we use a proper regularization parameter μ . In these experiments, we compute sTV reconstructions for ten different values of the regularization parameter. For the Shepp–Logan phantom, the smallest relative errors for 120 and 30 projection data are attained by using $\mu = 10^{-4}$. Using different μ may yield poor reconstructions, as can be seen in figures 5 and 7 for the Shepp–Logan phantom and the walnut, respectively. We stress that one of the main advantages of the proposed CWDS method is that it does not require any tuning for the regularization parameter. Also, compared to some other approaches that appear simpler, like iterative hard thresholding by keeping a constant number of largest wavelet coefficients [20], the CWDS seemingly can be extended to different forms of

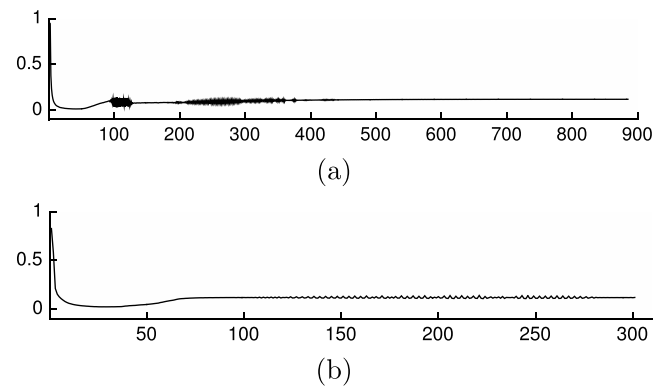


Figure 9. The ratio of nonzero wavelet coefficients as the iteration progresses, for the Shepp–Logan phantom. (a) 120 projections. (b) 30 projections.

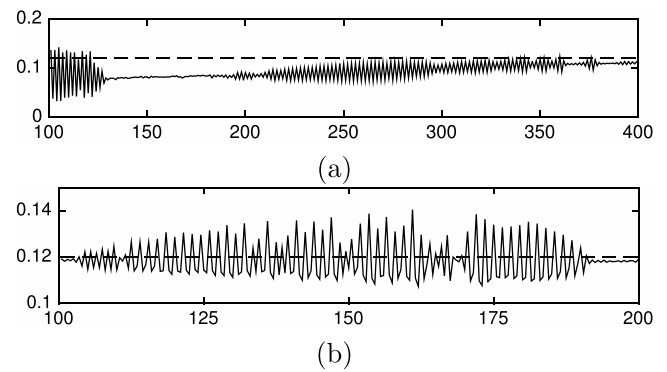


Figure 10. A closer look at the sparsity level shows oscillations for the Shepp–Logan phantom. (a) 120 projections (iterations 100–400). (b) 30 projections (iterations 100–200). The dashed line shows the sparsity prior C_{pr} .

thresholding with slight changes. For instance, future work could focus in coupling the tuning-free idea introduced with CWDS with the anisotropic TV formulation. The CWDS could potentially be applied as well to other techniques than thresholding-based iterative reconstruction algorithms.

Concerning the behavior of the sparsity level for the walnut case, it can be seen in the first row of figure 8 that the initial rapid oscillations decays fast. This is due to the role of the additional controller tuning β , as presented in section 2.6.

For the Shepp–Logan case, it can be seen in figure 9, and with a closer look for some iterations in figure 10, that the ratio of nonzero wavelet coefficients produces long lasting high frequency oscillations. In the case of 30 projections, the oscillations are fixated around the desired level of sparsity, and thus they are likely controller induced: the tuning parameter β is too large and thus the ratio of nonzero coefficients keeps overshooting, and, at the same time, since the absolute difference of the controller error e of two consecutive iterations is very small, β decreases slowly. In the case of 120 projections, however, the oscillations are not centered around the desired level of sparsity. This unexpected behavior is likely to be an indication of the nonlinearities in the system. Fortunately, the problem seems relatively minor in this case: the ratio of

nonzero coefficients of the reconstruction eventually gets sufficiently close to the desired level.

The proposed feedback controller is relatively easy to implement and some related-experiments have been done successfully using the method. However, if, for the suggested simple controller, the nonlinearity and the uncertainty of the system prove to be too much in some cases, alternative controller designs are required. Future research might therefore involve the use of nonlinear control techniques to avoid problems caused by the nonlinearity. Also, future research could delve into alternative adaptive self-tuning controllers, such as the adaptive integral controller introduced in [44]. Such controllers might improve the system response to unexpected disturbances and help with the oscillations caused by the slow decay of β and the nonlinearities demonstrated in figure 9. Additionally, careful analysis of the dynamics of the algorithm (3) is required to see if convergence of CWDS can always be guaranteed with the methods presented in this paper, or if a more advanced control is required.

Anyhow, what is remarkable is that, for all numerical experiments, the sparsity level eventually converges to the desired sparsity level C_{pr} .

6. Conclusion

In this paper, we proposed a new approach in tuning the regularization parameter, in this case the sparsity level of the reconstruction in the wavelet domain. The CWDS seems to be a promising strategy, especially in real-life applications where the end-users could avoid manually tuning the parameters.

In the case of sparsely collected projection data, the fully automatic CWDS outperforms the conventional FBP algorithm in terms of image quality (measured as relative RMS error).

Acknowledgments

This work was supported by the Academy of Finland through the Finnish Centre of Excellence in Inverse Problems Research 2012–2017 (Academy of Finland CoE-project 284715).

ORCID iDs

Zenith Purisha  <https://orcid.org/0000-0003-0801-3087>

References

- [1] Kak A C and Slaney M 2001 *Principles of Computerized Tomographic Imaging* (Philadelphia: SIAM) (<https://doi.org/10.1137/1.9780898719277>)
- [2] Natterer F 2001 *The Mathematics of Computerized Tomography* (Philadelphia: SIAM) (<https://doi.org/10.1137/1.9780898719284>)
- [3] Shepp L A and Kruskal J 1978 Computerized tomography: the new medical x-ray technology *Am. Math. Mon.* **85** 420–39
- [4] Lions P-L and Mercier B 1979 Splitting algorithms for the sum of two nonlinear operators *SIAM J. Numer. Anal.* **16** 964–79
- [5] Daubechies I, Defrise M and DeMol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint *Commun. Pure Appl. Math.* **57** 1413–57
- [6] Loris I and Verhoeven C 2011 On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty *Inverse Problems* **27** 125007
- [7] Chen P, Huang J and Zhang X 2016 A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions *Fixed Point Theory Appl.* **2016** 54
- [8] Rullgård H 2008 A new principle for choosing regularization parameter in certain inverse problems (arXiv:0803.3713v2 [math.NA])
- [9] Clason C, Jin B and Kunisch K 2010 A duality-based splitting method for ℓ^1 -tv image restoration with automatic regularization parameter choice *SIAM J. Sci. Comput.* **32** 1484–505
- [10] Dong Y, Hintermüller M and Rincon-Camacho M M 2011 Automated regularization parameter selection in multi-scale total variation models for image restoration *J. Math. Imaging Vis.* **40** 82104
- [11] Frick K, Marnitz P and Munk A 2012 Statistical multiresolution dantzig estimation in imaging: fundamental concepts and algorithmic framework *Electron. J. Stat.* **6** 231–68
- [12] Wen Y-W and Chan R H 2011 Parameter selection for total-variation-based image restoration using discrepancy principle *IEEE Trans. Image Process.* **21** 1770–81
- [13] Chen K, Loli Piccolomini E and Zama F 2014 An automatic regularization parameter selection algorithm in the total variation model for image deblurring *Numer. Algorithms* **67** 7392
- [14] Toma A, Sixou B and Peyrin F 2015 Iterative choice of the optimal regularization parameter in TV image restoration *Inverse Problems Imaging* **9** 1171–91
- [15] Niinimäki K, Lassas M, Hämmäläinen K, Kallonen A, Kolehmainen V, Niemi E and Siltanen S 2014 Multi-resolution parameter choice method for total variation regularized tomography *SIAM Journal on Imaging Sciences* **9.3** 938–974
- [16] Liu J, Huang T-Z, Lv X-G and Wang S 2017 High-order total variation-based poissonian image deconvolution with spatially adapted regularization parameter *Appl. Math. Modelling* **45** 516–29
- [17] Becker H, Albera L, Comon P, Nunes J-C, Gribonval R, Fleureau J, Guillotel P and Merlet I 2017 Sissy: an efficient and automatic algorithm for the analysis of EEG sources based on structured sparsity *NeuroImage* **157** 157–72
- [18] Hao B and Zhu J 2017 Fast l1 regularized iterative forward backward splitting with adaptive parameter selection for image restoration *J. Vis. Commun. Image Represent.* **44** 139–47
- [19] Pfister L and Bresler Y 2017 Automatic parameter tuning for image denoising with learned sparsifying transforms *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* <https://doi.org/10.1109/ICASSP.2017.7953316>
- [20] Blumensath T and Davies M E 2009 Iterative hard thresholding for compressed sensing *Applied and Computational Harmonic Analysis* **27** 265–74
- [21] Kolehmainen V, Lassas M, Niinimäki K and Siltanen S 2012 Sparsity-promoting bayesian inversion *Inverse Problems* **28** 025005
- [22] Hämmäläinen K, Kallonen A, Kolehmainen V, Lassas M, Niinimäki K and Siltanen S 2013 Sparse tomography *SIAM J. Sci. Comput.* **35** B644–65
- [23] Åström K J and Häggglund T 1995 *PID Controllers: Theory, Design, and Tuning* vol 2 (Research Park, NC: ISA)

- [24] Araki M 2009 PID control *Control Systems, Robotics and Automation: System Analysis and Control: Classical Approaches II* ed H Unbehauen (Oxford: EOLSS Publishers Co. Ltd.) pp 58–79
- [25] Bennett S 1993 *A History of Control Engineering, 1930–1955* vol 47 (London: IET)
- [26] Bahraoui M and Lemaire B 1994 Convergence of diagonally stationary sequences in convex optimization *Set-Valued Analysis* **2** 49–61
- [27] Attouch H 1996 Viscosity solutions of minimization problems *SIAM J. Optim.* **6** 769–806
- [28] Attouch H and Cominetti R 1996 A dynamical approach to convex minimization coupling approximation with the steepest descent method *J. Differ. Equ.* **128** 519–40
- [29] Cabot A 2005 Proximal point algorithm controlled by a slowly vanishing term: applications to hierarchical minimization *SIAM J. Optim.* **15** 555–72
- [30] Rosasco L, Tacchetti A and Villa S 2014 Regularization by early stopping for online learning algorithms (arXiv:1405.0042v1)
- [31] Rosasco L, Villa S and Vũ B C 2016 A stochastic inertial forward–backward splitting algorithm for multivariate monotone inclusions *Optimization* **65** 1293–314
- [32] Hale E T, Yin W and Zhang Y 2008 Fixed-point continuation for ℓ_1 -minimization: methodology and convergence *SIAM J. Optim.* **19** 1107–30
- [33] Rantala M, Vanska S, Jarvenpää S, Kalke M, Lassas M, Moberg J and Siltanen S 2006 Wavelet-based reconstruction for limited-angle x-ray tomography *IEEE Trans. Med. Imaging* **25** 210–7
- [34] Niinimäki K, Siltanen S and Kolehmainen V 2007 Bayesian multiresolution method for local tomography in dental x-ray imaging *Phys. Med. Biol.* **52** 6663
- [35] Soussen C and Idier J 2008 Reconstruction of three-dimensional localized objects from limited angle x-ray projections: an approach based on sparsity and multigrid image representation *J. Electron. Imaging* **17** 033011–1
- [36] Klann E, Ramlau R and Reichel L 2011 Wavelet-based multilevel methods for linear ill-posed problems *BIT Numer. Math.* **51** 669–94
- [37] Klann E, Quinto E T and Ramlau R 2015 Wavelet methods for a weighted sparsity penalty for region of interest tomography *Inverse Problems* **31** 025001
- [38] Helin T and Yuditskiy M 2013 Wavelet methods in multi-conjugate adaptive optics *Inverse Problems* **29** 085003
- [39] Daubechies I 1992 *Ten Lectures on Wavelets* (Philadelphia, PA: SIAM) (<https://doi.org/10.1137/1.9781611970104>)
- [40] Hämäläinen K, Harhanen L, Kallonen A, Kujanpää A, Niemi E and Siltanen S 2015 Tomographic x-ray data of a walnut (arXiv:1502.04064)
- [41] Vanden Berg E and Friedlander M 2013 Spot a linear-operator toolbox www.cs.ubc.ca/labs/scl/spot/ (accessed: 02 August 2013)
- [42] Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia, PA: SIAM) (<https://doi.org/10.1137/1.9780898717570>)
- [43] Barzilai J and Borwein J M 1988 Two point step size gradient methods *IMA J. Numer. Anal.* **8** 141–8
- [44] Logemann H and Townley S 1997 Adaptive integral control of time-delay systems *IEE Proc. Control Theory Appl.* **144** 531–6