Studies of Sensitivity in the Dictionary Learning Approach to Computed Tomography: Simplifying the Reconstruction Problem, Rotation, and Scale

Soltani, Sara

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Studies of Sensitivity in the Dictionary Learning Approach to Computed Tomography: Simplifying the Reconstruction Problem, Rotation, and Scale

Sara Soltani *

Department of Applied Mathematics and Computer Science,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark,
DTU Compute Technical Report-2015-04.

July 2, 2015

Abstract

In this report, we address the problem of low-dose tomographic image reconstruction using dictionary priors learned from training images. In our recent work [22] dictionary learning is used to incorporate priors from training images and construct a dictionary, and then the reconstruction problem is formulated in a convex optimization framework by looking for a solution with a sparse representation in the subspace spanned by the dictionary. The work in [22] has shown that using learned dictionaries in computed tomography can lead to superior image reconstructions comparing to classical methods. Our formulation in [22] enforces that the solution is an exact representation by the dictionary; in this report, we investigate this requirement. Furthermore, the underlying assumption that the scale and orientation of the training images are consistent with the unknown image of interest may not be realistic. We investigate the sensitivity and robustness of the reconstruction to variations of the scale and orientation in the training images and we suggest algorithms to estimate the correct relative scale and orientation of the unknown image to the training images from the data.

*Email: ssol@dtu.dk

This work is part of the project HD-Tomo funded by Advanced Grant No. 291405 from the European Research Council

2010 Mathematics Subject Classification: Primary: 65F22; Secondary: 65K10.

Key words and Phrases: Tomography, Dictionary learning, Inverse problem, Regularization, Sparse representation, Image reconstruction, Scale, Rotation.

1 Introduction

The linear computed tomographic (CT) reconstruction problem is often formulated as $Ax \approx b$ where the vector $x \in \mathbb{R}^n$ represents the unknown image, the vector $b \in \mathbb{R}^m$ is the given (usually inaccurate/noisy) data, and the matrix $A \in \mathbb{R}^{m \times n}$ represents the forward model. To regularize the (often) ill-posed tomographic image reconstruction, we use prior information in the form of "training images" that characterize the geometrical or visual features of the property of interest.

Consider an over-complete *dictionary* $D \in \mathbb{R}^{p \times s}$ with $s > p$ that is composed of s columns – representing a "basis" of s images – which are referred to as *elements* of the dictionary, i.e., the dictionary D contains prototype image-elements. Let $y \in \mathbb{R}^p$ be an arbitrary image. We say that y admits a sparse representation over the dictionary D if one can find a linear combination of a small number of dictionary elements that approximates the image y.

The term dictionary learning was first introduced by Olshausen and Field (1996) [20], referring to discovering basis elements of the dictionary and sparse linear combination of those elements using unlabeled training images. Dictionary learning methods and sparse decomposition are now widely used to model natural signals/images (see e.g., [1, 6, 8, 14]). Sparse representation in terms of learned dictionaries has attracted increased interest in solving imaging problems such as denoising [9], deblurring [18] and restoration [15], in addition to solving tomographic image reconstruction problems [22, 26].

In a recent paper [22] we formulated and implemented a two-stage algorithm for using training images in tomographic reconstruction, in which we first form a dictionary from patches extracted from the training images and then use this dictionary as a prior when computing the reconstruction. Analysis of the image in the patch-based formulation enables the dictionary to find localized features of training images effectively and reduces the computational work. In [22], a nonnegative sparse coding formulation [12] is proposed for learning a single generic dictionary for sparse representation of gray scale training image patches. By means of a sparsity prior on all the non-overlapping patches in the image, the dictionary is used for finding a tomographic solution in the space (cone) defined by the dictionary. Non-overlapping patches are used so that we avoid blurring and over-smoothing the textures in the overlap regions of the reconstruction. Being successful in incorporating the desirable features of the training image in the dictionary prior, leads to a superior solution comparing to classical tomographic reconstruction methods. We also showed that the use of the nonnegative dictionary had a regularizing effect on the solution.

There is no guarantee that the training images have the correct orientation or scale when trying to solve the image reconstruction problem for an unknown object, which is often neglected when using learned dictionary approaches in tomographic image reconstruction, e.g., see [22, 23, 26]. On the other hand in [22] we have been working under the assumption that the solution lies in the cone spanned by the learned dictionary elements. Searching for solutions in the cone spanned by the dictionary elements is a strong assumption in the reconstruction formulation, therefore we are interested to investigate how relaxing this assumption affects the reconstructed solution.

In this report, we continue the work initiated in [22], in order to increase an understanding of the model's limitations and capabilities, we analyze sensitiv-

ity and robustness of our algorithm to scale and rotation variances as well as the assumptions in the problem formulation with various computational tests. Moreover, we propose algorithms to detect rotation and scale of the reconstruction from the sinogram of the tomographic measurement data.

This paper is organized as follows. In Section 2 we first describe the two-stage tomographic reconstruction framework using leaned dictionaries, present a reference example from [22], and describe our numerical test setup. In Section 3 we investigate the assumption that the solution lies strictly in the space spanned by the dictionary elements. Then, in section 4 we study the sensitivity of the reconstruction towards changes in scale and rotation and present algorithms to determine the correct scale and rotation from the measurement tomographic data. The conclusion follows in Section 5.

We use the following notation, where A is an arbitrary matrix:

$$\|A\|_{\mathrm{F}} = \big(\sum_{ij} a_{ij}^2\big)^{1/2}, \quad \|A\|_{\mathrm{sum}} = \sum_{ij} |a_{ij}|.$$

2 The Dictionary Learning Approach in Tomographic Image Reconstruction

For the sake of completeness, in this section we briefly summarize the formulation used in the tomographic reconstruction formulation from [22], and we present our computational scheme and numerical results from a reference problem in [22].

2.1 The Problem Formulation

Let the matrix $Y \in \mathbb{R}_+^{p\times t}$ consist of t training image patches of size $P \times Q$ arranged as vectors of length $p = PQ$. Then the dictionary learning problem can be viewed as the problem of approximating the training matrix as a product of two matrices, $Y \approx DH$, where $D \in \mathbb{R}_+^{p\times s}$ is the dictionary of s basis elements (the columns of D), and $H \in \mathbb{R}_+^{s\times t}$ is the matrix of coefficients. Such a decomposition is not unique, so we need to incorporate further priors on the approximation in order to obtain the dictionary. The matrix D is required to have nonnegative elements, such that its columns represent image patches. Similarly, imposing non-negativity constraints on the representation matrix H corresponds to each training image being represented as a conic combination of dictionary images. Imposing a sparsity-inducing norm penalty on H allows one to control the sparseness of the representation of training patches, as well as alleviate non-uniqueness drawback of the approximation.

Then the dictionary learning problem is given by the non-negative sparse coding of the non-negative data matrix Y

$$\min_{D,H} \quad \frac{1}{2}\|Y - DH\|_{\mathrm{F}}^2 + \lambda\|H\|_{\mathrm{sum}} \qquad \text{s.t.} \qquad D \in \mathcal{D},\ H \in \mathbb{R}_+^{s\times t}, \tag{1}$$

where $\lambda \geq 0$ is a regularization parameter and we introduce

$$\mathcal{D} \equiv \{D \in \mathbb{R}_+^{p\times s} \mid \|d_i\|_2 \leq \sqrt{p}, i = 1, \ldots, s\}$$

to avoid trivial solutions and to ensure similar bounds on all dictionary elements. The sparsity prior for the representation matrix H is imposed by $\|H\|_{\text{sum}}$. Each training patch and dictionary image is a column vector in Y and D respectively. We use the algorithm based on the Alternating Direction Method of Multipliers (ADMM) [5], for computing a dictionary at a local minima of the non-convex dictionary learning problem (1).

Now let the vector x represent the unknown $M \times N$ image of absorption coefficients with $x \in R_+^n$ and $n = MN$.

Without loss of generality we assume that the size of the image is a multiple of the patch sizes in the dictionary. We partition the image into $q = (M/P)(N/Q)$ non-overlapping blocks of size $(M/P) \times (N/Q)$, i.e., patches are represented by the vectors $x_j \in \mathbb{R}^p$ for $j = 1, \ldots, q$. Then each block of the image x can expressed as a conic combination of dictionary images, i.e., $\Pi x = (I \otimes D)\alpha$. The permutation matrix Π re-orders the vector x such that the correct shuffling of the pixels from the patches is ensured.

The simplicity of this approach is that once the basis elements have been determined, the solution is linear in these new variables, and the data fitting proceeds as usual i.e., from $Ax \approx b$ we obtain $A\Pi^T(I \otimes D)\alpha \approx b$. The tomographic image reconstruction is hence given by:

$$\begin{array}{ll} \text{minimize}_\alpha & \frac{1}{2m}\|A\Pi^T(I \otimes D)\alpha - b\|_2^2 + \frac{\mu}{q}\|\alpha\|_1 + \delta^2\,\psi(\Pi^T(I \otimes D)\alpha) \\ \text{subject to} & \alpha \geq 0 \end{array} \tag{2}$$

where $\alpha \in \mathbb{R}_+^{sq}$. The image prior $\delta^2\psi(\cdot)$ penalize block artifacts that may arise in the reconstruction based on non-overlapping blocks, so the function $\psi(\cdot)$ penalizes jumps across the pixels at the boundary of neighboring patches:

$$\psi(z) = \frac{1/2}{M(M/p-1) + N(N/q-1)}\,\|L\,z\|_2^2, \tag{3}$$

where L is a finite-difference approximation matrix. The denominator is the total number of pixels along the boundaries of the blocks in the image. The 1-norm regularization of α is known to produce sparse coefficients in terms of the dictionary [8]. The problem formulation (2) is normalized by the division of the squared residual norm by the number of measurement m, and division of the 1-norm constraint by the number of blocks q.

We note that (2) is a convex but non-differentiable optimization problem. One can solve the optimization problem (2) by means of sparse approximation methods, see e.g., [24]. We use the software package TFOCS (Templates for First-Order Conic Solvers) [4].

2.2 The Numerical Setup and a Reference Problem

In Sections 3 and 4 we use numerical examples to demonstrate and quantify the behavior of our two-stage algorithm when we encounter uncertainty in the tomographic reconstruction stage such as model assumptions and changes in the scale and orientation of the object. In particular we explore the influence of relaxing the representation in the cone defined by the dictionary, and illustrate the role of scale and orientation of the learned dictionary in the reconstruction stage.

All experiments are run in MATLAB (R2014a) on a 64-bit Linux system. In the dictionary learning stage we use a data set of images which are similar to the ones we wish to reconstruct. The ground-truth or exact image x^{exact} is not contained in the training set, so that we avoid committing an inverse crime. All images are gray-level and scaled in the interval [0, 1]. We use an implementation of the ADMM algorithm presented in [22] to obtain a dictionary and the reconstruction problems are solved using the software package TFOCS version 1.3.1 [4].

All test problems represent a parallel-beam tomographic measurement, and we use the function `paralleltomo` from the MATLAB package AIR Tools [10] to compute the matrix A. The number of rays in each projection is given by $N_{\text{r}} = \lfloor \sqrt{2}N \rfloor$. If the total number of projections is N_{p} then the number of rows in A is $m = N_{\text{r}}N_{\text{p}}$ while the number of columns is $n = MN$. In particular we are interested in scenarios with a small number of projections. The exact data is generated with the forward model after which we add Gaussian white noise, i.e., $b = Ax^{\text{exact}} + e$.

We consider a reference test image from a high-resolution photo of peppers with uneven surfaces, which is an interesting test image for studies of the reconstruction of textures. Figure 1 shows the 1600×1200 high-resolution image and the exact image of dimensions $M \times N = 200 \times 200$.

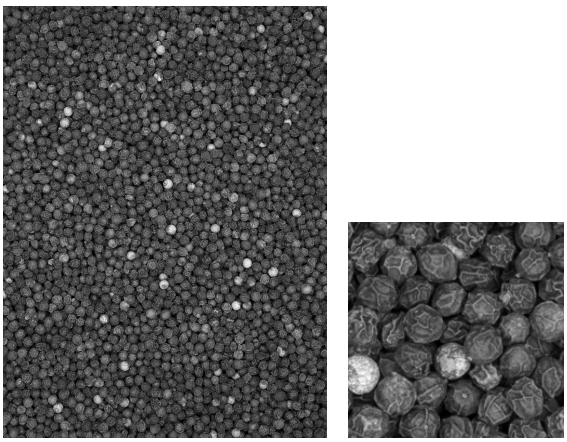


Figure 1: Left: the high-resolution image from which we obtain the training image patches. Right: the 200×200 exact image x^{exact}.

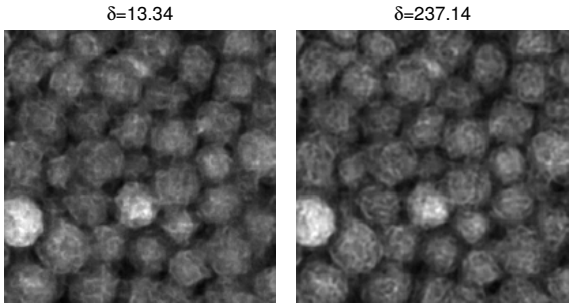
We choose $N_{\text{p}} = 25$ projections corresponding to uniformly distributed angles in $[0°, 180°]$. Hence the matrix A has dimensions $m = \lfloor \sqrt{2} \cdot 200 \rfloor \cdot 25 = 7,075$ and $n = 200^2 = 40,000$, so the problem is highly underdetermined. We use the relative noise level $\|e\|_2 / \|Ax^{\text{exact}}\|_2 = 0.01$. The quality of a solution x is evaluated in terms of the reconstruction error

$$\text{RE} = \frac{\|x - x^{\text{exact}}\|_2}{\|x^{\text{exact}}\|_2}. \tag{4}$$

Patch sizes should be sufficiently large to capture the desired structure in the training images, but the computational cost of the dictionary learning increases as the patch size increases. A study of the patch size $p = P \times Q$ and the number

of dictionary elements s in [22] shows that a reasonably large patch size gives a good trade-off between the computational work and the approximation error by the dictionary, and that the over-representation factor $s/(pr)$ can be smaller for larger patches. For these reasons we use square patch sizes of 10×10 and 20×20 and learn dictionary matrices $D^{(10)}$ and $D^{(20)}$ in $\mathcal{D}$ of size 100×300 and 400×800 respectively. An empirical study of the regularization parameter λ in (1) and (2) suggest that the smallest reconstruction error can be obtained in $D^{(10)}$ and $D^{(20)}$ for $\lambda \approx 3$.

The parameters μ and δ in the reconstruction problem (2) both play a role in terms of regularization; to simplify (2) we set $\tau = \mu/q$. As described in [22], the nonnegative constraint in the reconstruction problem plays an extra role of regularization and therefore the reconstruction is not very sensitive to the regularization parameters δ and τ, hence they are chosen from a few numerical experiments such that a solution with the smallest error is obtained. The reconstruction solutions of problem (2) are shown in Fig. 2, which shows that while using $D^{(10)}$ leads to a slightly smaller reconstruction error, the solution obtained with $D^{(20)}$ is also appealing.



(a) 10×10, RE $= 0.220$ (b) 20×20, RE $= 0.226$

Figure 2: Reconstructions for different patch sizes, with $D \in \mathcal{D}$, $\lambda = 3.16$, and $\tau = 0.022$. RE denotes the reconstruction error (4).

3 Importance of the Representation in the Cone Defined by the Dictionary

In this section we perform an empirical study of the reconstruction's robustness to the assumption in the reconstruction step that the solution is a conic combination of dictionary elements and its effect on the success of reconstruction.

Searching for a reconstruction in the space spanned by the dictionary, i.e., $x = (I \otimes D)\alpha$, is a very strong prior. Let us construct our tomographic reconstruction formulation in a different way. We know that a simple/naive tomographic reconstruction problem for Gaussian noise could be formulated as

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 \qquad \text{s.t.} \quad x \in \mathbb{R}^n_+, \tag{5}$$

where the non-negativity of the image is imposed as a prior. Due to the ill-posed nature of the underlying problem, the lack of other priors results in un-

satisfactory results from solving (5). Now to incorporate our dictionary prior, we consider $\Pi x \approx (I \otimes D)\alpha$ rather than assuming that $\Pi x = (I \otimes D)\alpha$, i.e., x does not have an exact representation in the dictionary and instead it is close to a solution that lies in the space spanned by the dictionary elements. Thus we consider the following reconstruction problem:

$$\min_{x,\alpha} \frac{1}{2m}\|Ax - b\|_2^2 + \delta^2\psi(x) + \beta\|x - \Pi^T(I \otimes D)\alpha\|_2^2, \tag{6}$$

$$\text{s.t.} \quad x \geq 0, \ \alpha \geq 0,$$

where the function $\psi(\cdot)$ is defined in (3).

For simplicity of this study, we dropped the sparsity prior $\mu/q\|\alpha\|_1$ from (2) in (6). This is motivated by the results from [22] that for sufficiently large values of δ and patch sizes, the reconstruction error is almost independent of μ as long as it is small.

The problem (6) can equivalently be written as:

$$\min_{x,\alpha} \frac{1}{2}\left\| \begin{pmatrix} \frac{1}{\sqrt{m}}A & 0 \\ \frac{\delta}{\sqrt{\kappa}}L & 0 \\ \sqrt{2\beta}I & -\sqrt{2\beta}\Pi^T(I \otimes D) \end{pmatrix} \begin{pmatrix} x \\ \alpha \end{pmatrix} - \begin{pmatrix} b \\ 0 \\ 0 \end{pmatrix} \right\|_2^2 \tag{7}$$

$$\text{s.t.} \quad \begin{pmatrix} x \\ \alpha \end{pmatrix} \geq 0.$$

Note the similarity of the (7) to the generic nonnegative least squares problem formulation (5).

The regularization parameter β in (6) and (7) balances the fitting term and the regularization induced by the dictionary. The larger the β, the more weight is given to minimization of $\|x - \Pi^T(I \otimes D)\alpha\|_2^2$, while for small β more weight is given to fitting the noisy data, resulting in solutions that are less regular (we obtain the problem (5) and the naive solution when $\beta = 0$). We expect that for sufficiently large β we obtain solutions not far from solutions obtained with the exact dictionary approach (i.e., from problem (2)).

Consider the tomographic problem from section 2.1 with $N_\mathrm{p} = 25$ projections and 1% additive relative noise. Moreover, we use the 20 by 20 patch dictionary $D^{(20)} \in \mathcal{D}$ of size 400×800.

The reconstructions for various values of β are shown in Fig. 3; they are similar across the larger values of β, however pronounced artifacts have appeared for small values of β from over-fitting the noisy data and reducing the weight on the dictionary prior. As can be see in Fig. 3, with larger values of β and less weight given to fit the tomographic data, the solution tends to be smooth.

We define the relative dictionary misfit by $\|\Pi^T(I \otimes D)\alpha - x\|_2/\|x\|_2$. Plots of the reconstruction error and the relative dictionary misfit are given in Fig. 3. As illustrated by these plots the reconstruction error decreases and then levels off for large values of β, e.g., RE= 0.2238 for $\beta = 1000$. The relative dictionary misfit exponentially decreases for large values of β, indicating that the approximation $x \approx \Pi^T(I \otimes D)\alpha$ is almost exact for β sufficiently large.

By considering the problem formulation (7) instead of (2) we are introducing β as a new regularization parameter, which needs further investigations to find a suitable value for it. In general relaxing $\Pi x = (I \otimes D)\alpha$ does not give an

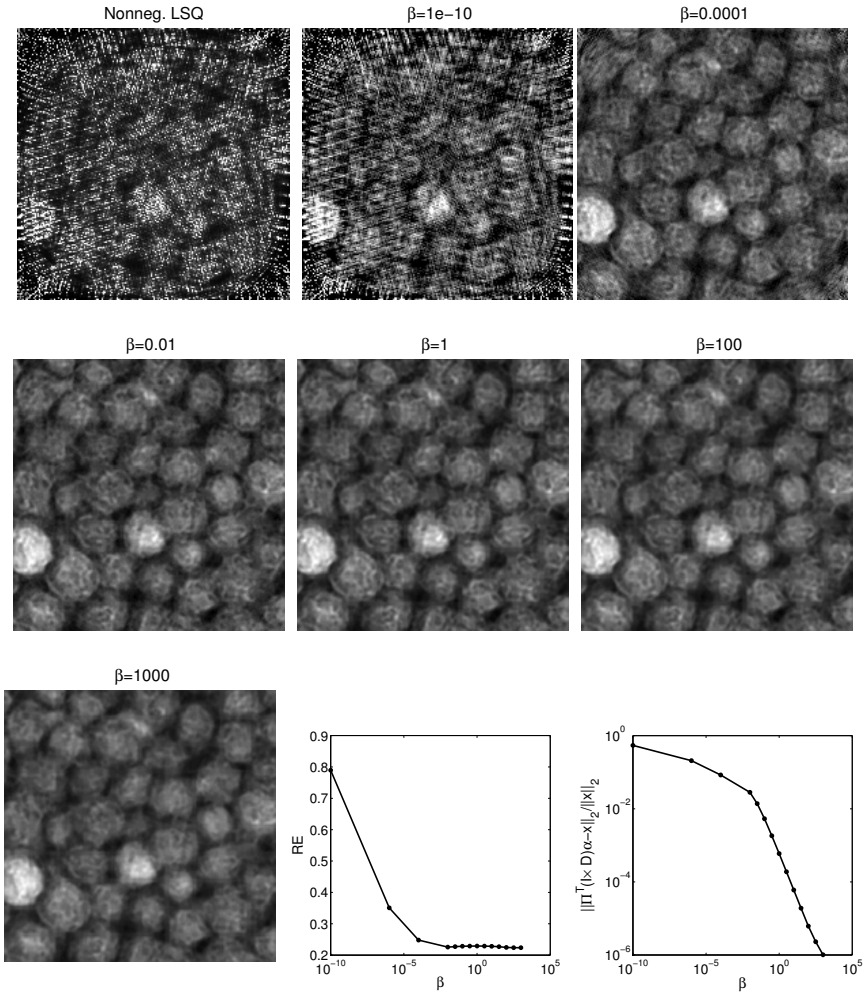


Figure 3: Reconstruction results from solving (6) with $\beta \in [10^{-10}, 1000]$. Bottom: Middle: plot of reconstruction error versus β. Right: plot of the relative dictionary misfit versus β.

advantage, i.e., approximating a solution by $\Pi x \approx (I \otimes D)\alpha$ does not particularly improve the reconstruction quality, and one can safely assume that the solution is a conic combination of the dictionary elements.

4 Experiments with Scale and Rotation

It maybe crucial to include the acts of rotation and geometric scaling of the training images when using the learned dictionaries in the tomographic reconstruction, where there is no guarantee that the training set will have the correct orientation and geometric scaling. Rotation and scaling are two unknown parameters that are needed to be considered in the reconstruction formulation and hence it is advantageous to determine the correct rotation and scaling parameters or obtain a scale and rotation invariant dictionary prior to the reconstruction process.

Invariance to rotation and scale are desirable in many practical applications. For example, in pattern recognition the widely used scale-invariant feature transform (SIFT) algorithm successfully detects the training image under changes in image scale, noise and rotation [19]. The paper [11] presents a face recognition method which uses features that are extracted from the log-polar images which are invariant to scale and rotation. Dictionary learning methods that are independent of orientation and scale, with applications in classification of images or clustering, have also been recently developed. A shift, scale and rotation invariant dictionary learning method for multivariate signals was proposed in [3]. Hierarchical dictionary learning methods for invariant classification have also been proposed in [2]. These methods learn a dictionary in a log-polar domain. In the paper [7] a rotation and scale invariant clustering algorithm using dictionaries is presented where the image features are extracted in the Radon transform domain.

To the best of our knowledge, no study has investigated and explored the role played by scale and rotation in tomographic reconstruction approaches using dictionaries.

4.1 Sensitivity to Scale

It is possible that the scale of the training images differ from the one we would like to achieve in the reconstruction process. While the dictionary learning approaches in image processing problems such as image denoising and image restoration do not directly suffer from scale issues, it has been explored that with the existence of multi-scale features in images, using multi-scale dictionaries would result in superior reconstructions compared to single-scale dictionaries (see, e.g., [15, 16, 21]). Such dictionaries enforce sparsity at multiple scales.

One idea is to train the dictionary on many possible scaling of the training images, this approach is computationally expensive in both the learning and reconstruction stage. Inspired by a multi-scale dictionary, first we investigate if a generic dictionary of smaller patches (with a fixed patch size) or a learned dictionary from different scaling of the training images could result in a "better" reconstruction for an off-scale image.

If the image is represented by a function X then we say $\bar{X}$ is a scaled copy of X with scale factor η if $\bar{X}(u, v) = X(\eta u, \eta v)$.

We look at three test examples that we call "peppers", "matches", and "binary" images. The binary test image – a random image with binary pixel values – is generated by the `phantomgallery` function from the MATLAB package AIR Tools [10]. The exact test images of size 200×200 with the scale factor $\eta = 1.5$ are shown in Fig. 4.

Figure 4: The 200×200 exact images x^{exact} with scale factor $\eta = 1.5$. Left: peppers, middle: matches, and right: binary test images.

To generate different dictionaries for our tests, we consider a large training image for each test case and we denote its scale to be the reference scale (scale 1). Knowing that the scale of the training image is different from the image we want to reconstruct, we can argue that we need a greater over-representation factor to learn a generic dictionary and be able to represent off-scale images. Hence for $\eta = 1$ we learned dictionaries of 5×5 and 10×10 patch sizes with over-representation factors of 10 and 5, respectively, i.e., $D^{(5)} \in \mathbb{R}^{25\times 250}$ and $D^{(10)} \in \mathbb{R}^{100\times 500}$. We also learn a 20×20 patch dictionary of size 400×1200 in which the training patches are chosen randomly from training images that are scaled by a factor of 0.5, 1 and 2. Figure 5 shows examples of 200×200 sub-images of our three training test images with scale factors $\eta = 0.5,\ 1,\ 2$.

The learned multi-scale dictionaries with 20×20 patches and generic dictionaries with 10×10 patches and $\lambda = 1$ are given in Fig. 6. We clearly see the multi-scale features of the dictionary with 20×20 patches.

We solve the reconstruction problem (2) using the exact images given in Fig. 4. We choose $N_{\text{p}} = 25$, projections with uniformly distributed angles in $[0°, 180°]$, $N_{\text{r}} = 283$ and 1% additive noise level. In Figure 7 we compare our reconstructions with those computed by the multi-scale dictionary with 20×20 patches $(\eta = 0.5,\ 1,\ 2)$ and the generic dictionaries of scale factor $\eta = 1$ with 5×5 and 10×10 patch sizes. To be fair, the regularization parameters τ and δ were chosen to yield an optimal reconstruction in terms of the reconstruction error.

The reconstructions shown in the right column of Fig. 7 show no particular advantage in terms of reconstruction errors when using a multi-scale dictionary (learned from patches of various scale) over a sufficiently large generic dictionary of smaller patch sizes, with the reconstructions shown in left and middle columns of Fig. 7.

Now to better understand the role played by the scale parameter η, we solve the peppers tomographic reconstruction problem from the Section 2.1 with the exact image given in the Figure 1 and the matches test problem of size 200×200 where the exact image is given in Figure 8. The scale factor of these test images

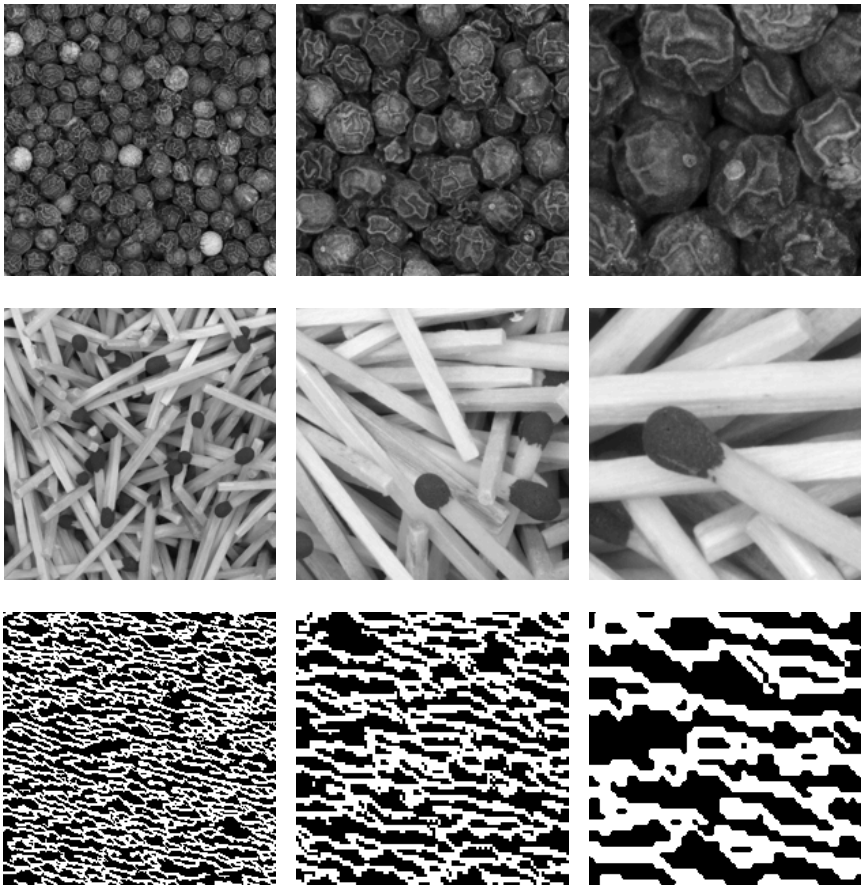


Figure 5: Examples of 200×200 sub-images of the training test images with scale factors $\eta = 0.5$, 1, 2. Top: peppers, middle: matches, and bottom: binary test images.

is assumed to be $\eta = 1$. We use $N_{\mathrm{p}} = 25$ projections with angles in $[0°, 180°]$ and relative noise level 0.01. We keep the size of the patches 10×10, and the dictionary size $s = 500$, and we learn 11 new dictionaries of size 100×500 where the scale factor of the training images η is varied in the interval $[0.4, 4]$. Plots of the reconstruction error versus the scale factor of the training patches, which we learned our dictionaries from, are given in Figure 9. We also plot the structural similarity index measure (SSIM) [25] for measuring the similarity between the reconstructed solution and the exact image in Figures 1 and 9. Recall that a larger SSIM means a better reconstruction.

Figure 9 shows that unless we are looking for a solution with a higher resolution than the training images, i.e., if the scale of the training images are smaller than the desired image that we want to reconstruct, the reconstruction is not very sensitive to the scaling factor, choosing a generic dictionary and sufficiently large number of elements. This is no surprise, one cannot expect to perfectly reconstruct a high resolution image from a dictionary learned from

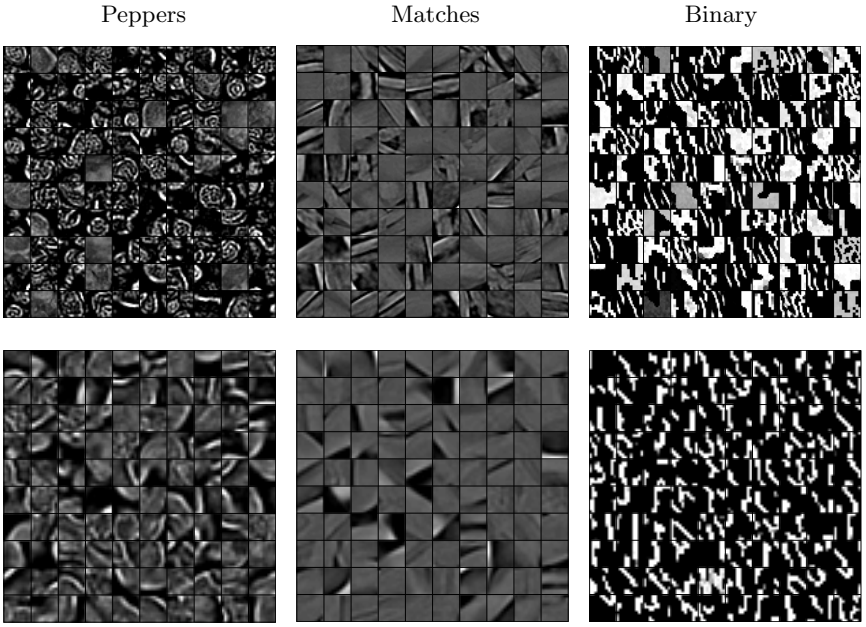


Figure 6: Top: Examples of the multi-scale dictionary elements (images) with 20×20 patches and $\lambda = 1$. Bottom: Examples of the generic dictionary elements (images) with scale factor 1, 10×10 patches and $\lambda = 1$.

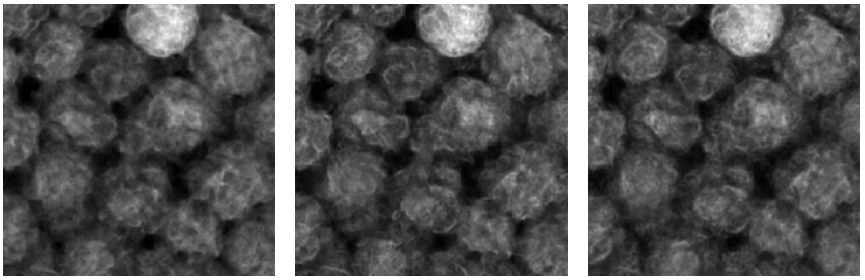
lower resolution training images since some important details of textures and structure are missing in those images.

4.2 An Algorithm to Determine Scale

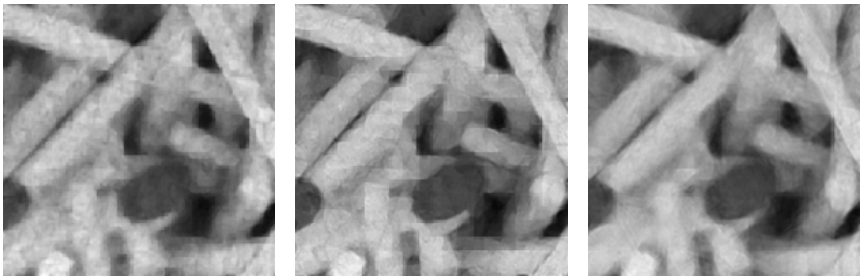
One may think of a preprocessing step to find the appropriate scale of the image before training the dictionary. The simplest case is downsizing the training images and learn the dictionary in the right scale or downsizing/shrinking the dictionary images in the right way. One simple way is to reconstruct a naive filter back projection (FBP) solution – see, e.g., [17, §2.3]) – and compare the solution with the training images to find the correct scale. The scale can be detected by comparing similar single objects in both images; however the limited tomographic data and presence of noise often result in obtaining unreliable naive solutions where most textures and image structures have disappeared, which makes such an estimation difficult.

Another option is to find scales from the sinogram of the 2D unknown image. Recall that the tomographic data can be represented – for some 2D applications – as a matrix called the sinogram. We denote the sinogram by the matrix S. The 2D Radon transform is graphically represented as the sinogram, which means by the intensity values in the coordinate system of variables (t, θ). The Radon transform of a two variable function X is defined as

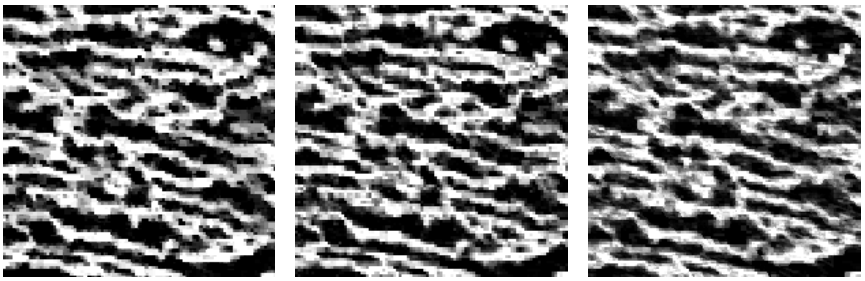
$$\mathrm{R}_\theta X(t) = \int_{-\infty}^{+\infty} X(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta)ds,$$
$$(t, \theta) \in (-\infty, \infty) \times [0, \pi).$$



(a) 5×5, RE=0.1973 (b) 10×10, RE=0.2025 (c) 20×20, RE=0.2035



(d) 5×5, RE=0.0782 (e) 10×10, RE=0.0712 (f) 20×20, RE=0.0717



(g) 5×5, RE=0.4236 (h) 10×10, RE=0.4577 (i) 20×20, RE=0.4624

Figure 7: Reconstructions for the generic and multi-scale dictionaries with different patch sizes (Figure 6), using the exact images given in Figure 4. RE denotes the reconstruction error.

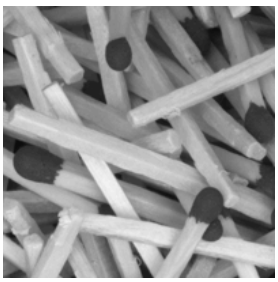


Figure 8: The 200×200 matches exact image x^{exact} with scale factor $\eta = 1$.

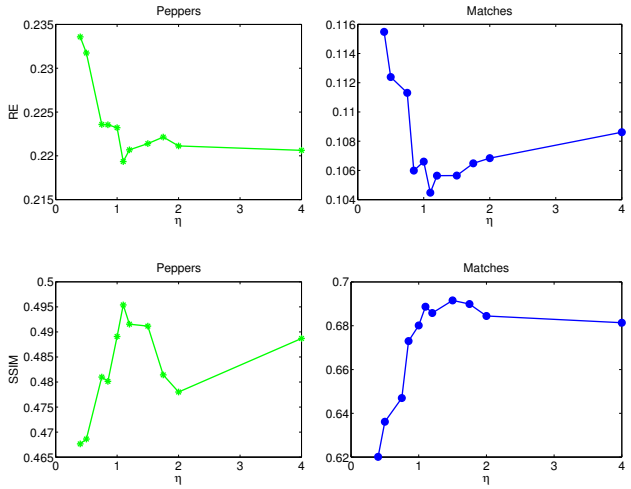


Figure 9: Top: left: peppers, and right: matches reconstruction errors versus the scaling factor of dictionaries. Bottom: left: peppers, and right: matches SSIM measures versus the scaling factor of dictionaries.

Let $\bar{X}$ be a scaled copy of X with the scaling factor η. Then the Radon transforms of $\bar{X}$ and X are related as follows:

$$\mathrm{R}_\theta \bar{X}(t) = \int_{-\infty}^{+\infty} \bar{X}(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta)ds \tag{8}$$

$$= \int_{-\infty}^{+\infty} X(\eta t\cos\theta - \eta s\sin\theta, \eta t\sin\theta + \eta s\cos\theta)ds \tag{9}$$

$$= \frac{1}{\eta}\mathrm{R}_\theta x(\eta t). \tag{10}$$

Let us define:

$$M_X = \max_{t,\theta} |\mathrm{R}_\theta X(t)|$$

Then for any pair $\bar{X}$ and X related by $\bar{X}(u,v) = X(\eta u, \eta v)$ with $\eta > 0$ the following holds:

$$M_{\bar{X}} = \frac{1}{\eta} M_X.$$

Since from (8):

$$\begin{aligned} M_{\bar{X}} &= \max_{t,\theta} |\mathrm{R}_\theta \bar{X}(t)| \\ &= \max_{t,\theta} |\frac{1}{\eta}\mathrm{R}_\theta X(\eta t)| \\ &= \frac{1}{\eta}\max_{t,\theta} |\mathrm{R}_\theta X(\eta t)| \\ &= \frac{1}{\eta}\max_{t,\theta} |\mathrm{R}_\theta X(\psi)| \quad (\text{if } \psi = \eta t) \\ &= \frac{1}{\eta} M_X. \end{aligned}$$

This proof is adopted from [7]. In the sinogram matrix S given by the discretized Radon transform, column indices correspond to discrete values of θ, while row indices correspond to discrete values of t. Hence M_X is the element-wise maximum of the values in the sinogram matrix.

Consider an unknown image X, where a noisy sinogram of X is available. We can make an artificial sinogram of a training image with the same tomographic setting/scenario. We can claim that if the training image $\bar{Z}$ with a similar dimension as X is given, then we can compute the relative scale factor η by

$$\eta \approx \frac{M_X}{M_{\bar{Z}}}.$$

We emphasize that the practical use of this approach relies on a careful implementation, and use of the Radon transform such that the integrals are correctly evaluated. Matlab's `radon` satisfies this requirement.

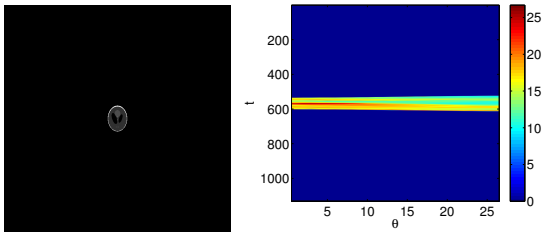
For a test problem we use the 200×200 resolution Shepp-Logan phantom in a 800×800 image grid given in Fig. 10 with $\eta = 1$. We compute the matrix A and the measurement data b with $N_{\mathrm{p}} = 25$ projections, $N_{\mathrm{r}} = 1131$ rays per projection and 1% additive noise. We construct $\bar{Z}$ as reference training images with scale factors 0.5, 2, 3, and 4 (see Fig. 10). We should here mention that it is important that all of these images have the same number of pixels, to avoid scaling issues with the numerical computations. We create an artificial noise-free sinogram of this training images. The images X, $\bar{Z}$ and the corresponding sinograms of our tomographic data are shown in Fig. 10. The number of pixels in the images given in Fig. 10 is 800^2.

We compute M_X and $M_{\bar{Z}}$ from the given sinograms in Fig. 10. We obtain $\eta = [0.51, 2.05, 3.11, 4.17]$, which are an approximation to the correct scale factors $[0.5, 2, 3, 4]$.

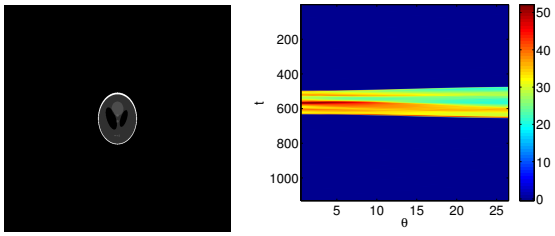
Now let us consider our textural 200×200 peppers test image with $\eta = 1$ given in Fig. 1. We identically consider $\bar{Z}$ as training images of size 200×200, similar to our test image with scale factors 0.5, 2, 3, 4 and compute the sinogram matrix S with analogous tomographic scenario, i.e., $N_{\mathrm{p}} = 25$ projections in $[0°, 180°]$, $N_{\mathrm{r}} = 283$ rays per projection and 1% additive noise (see Fig. 11). Computing M_X and $M_{\bar{Z}}$ from the given sinograms in Fig. 11 results in approximating the scale factors to be $[0.9647, 1.0000, 1.0738, 0.9649, 1.3451]$, showing that this method is not suited for images with textures.

We can conclude that this method only works well if the unknown image is a single object with an unknown scale, and a training image includes a similar object with a different scale.

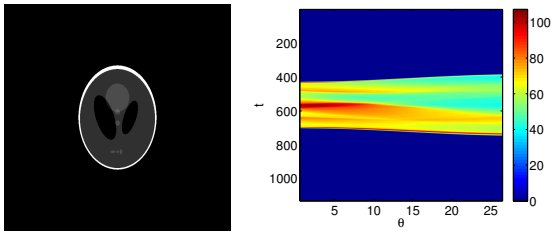
Finding the scale factor in 3D tomographic reconstruction where the tomographic data is available in form of projection images in which a multitude of details of the shapes and features are already visible, is a fairly straightforward process. Because the shapes in 2D slices of training images can be compared with similar shapes in the 2D projection data and the scale factor can be found with simple mathematical functions from geometry, e.g., we need to find a corresponding side in each similar shape in two images where we can measure the length of both. The ratio between the length of these sides is the scale factor.



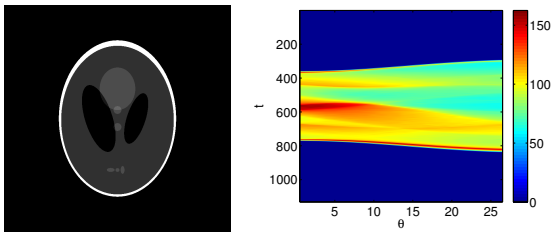
(a) $\bar{Z}, \eta = 0.5$ (b) Clean S, $M_{\bar{Z}} = 26.70$



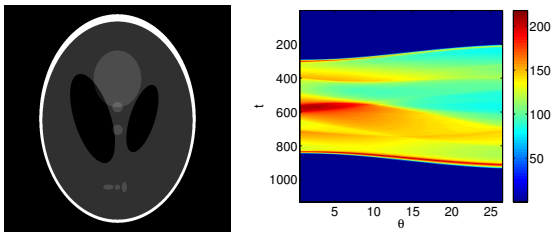
(c) $X, \eta = 1$ (d) Noisy S, $M_X = 52.11$



(e) $\bar{Z}, \eta = 2$ (f) Clean S, $M_{\bar{Z}} = 107.22$

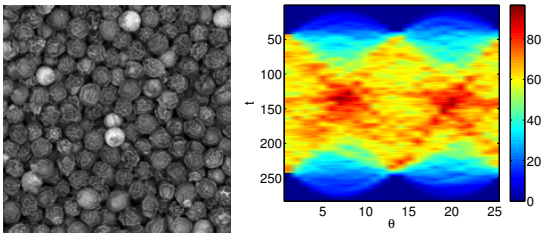


(g) $\bar{Z}, \eta = 3$ (h) Clean S, $M_{\bar{Z}} = 162.56$

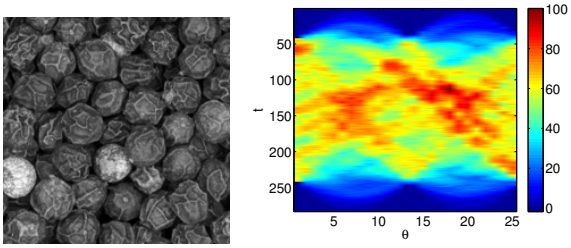


(i) $\bar{Z}, \eta = 4$ (j) Clean S, $M_{\bar{Z}} = 217.86$

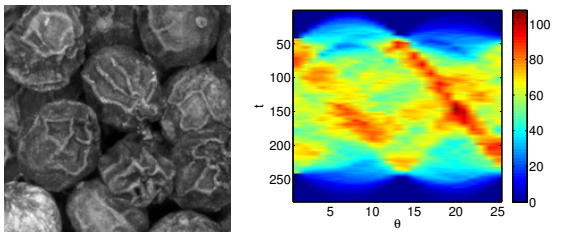
Figure 10: Right: The reference Shepp-logan phantom image X, $\eta = 1$ and training images $\bar{Z}$ with scale factor $\eta = 0.5$, 2, 3, 4. Left: The clean sinograms ($S \in \mathbb{R}^{N_\mathrm{r} \times N_\mathrm{p}}$) of $\bar{Z}$ and noisy sinogram of X with $N_\mathrm{p} = 25$ projections and $N_\mathrm{r} = 1131$ rays.



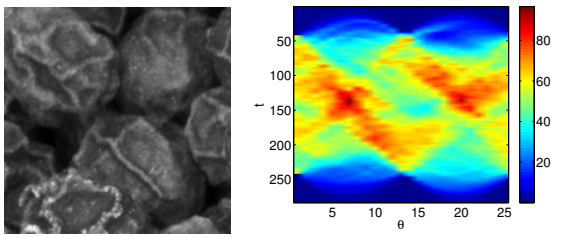
(a) $\bar{Z}, \eta = 0.5$ (b) Clean S, $M_{\bar{Z}} = 99.03$



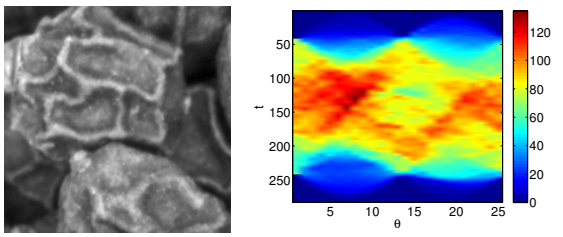
(c) $X, \eta = 1$ (d) Noisy S, $M_X = 102.27$



(e) $\bar{Z}, \eta = 2$ (f) Clean S, $M_{\bar{Z}} = 107.98$



(g) $\bar{Z}, \eta = 3$ (h) Clean S, $M_{\bar{Z}} = 97.84$



(i) $\bar{Z}, \eta = 4$ (j) Clean S, $M_{\bar{Z}} = 137.33$

Figure 11: Right: The reference peppers image X, $\eta = 1$ and training images $\bar{Z}$ with scale factor $\eta = 0.5$, 2, 3, 4. Left: The clean sinograms ($S \in \mathbb{R}^{N_\mathrm{r} \times N_\mathrm{p}}$) of $\bar{Z}$ and noisy sinogram of X with $N_\mathrm{p} = 25$ projections and $N_\mathrm{r} = 283$ rays.

4.3 Sensitivity to Rotation

In this section we analyze the sensitivity of the reconstruction results to a rotation parameter. We use three test images of size 200×200 which we call "peppers", "binary" and "D53". The D53 test image is chosen from the normalized brodatz texture database [27]. For the peppers test image we use the exact image given in Fig. 1. The binary and D53 test images are given in Figure 12. We expect that the peppers test image is invariant to rotation while the binary and D53 test images, as can be seen in Figure 12, are highly directional and sensitive to rotation.

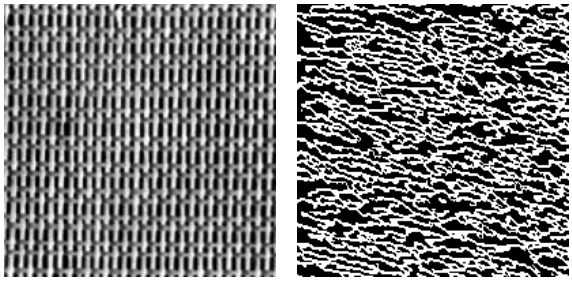


Figure 12: The 200×200 test images for the rotation sensitivity analysis. Left: the D53 and right: the binary test images.

We choose rotation angles of $[5°, 10°, 30°, 45°, 60°, 90°]$ and we rotate the test images with the chosen angles. Since the rotated images are not exactly equivalent to the original test images, for the comparison of the reconstruction qualities to be fair, we extracted 4 smaller test images of size 50×50 from each rotated image. We use a reconstruction scenario with 12 projections and 70 rays in $[0°, 180°]$ and 1% noise. We obtain a reconstruction for each 50×50 image in every rotation and average over the reconstruction errors and SSIM measures. Figure 13 shows the plots for the average reconstruction errors (RE) and SSIM measures versus the rotation angles for our three test images.

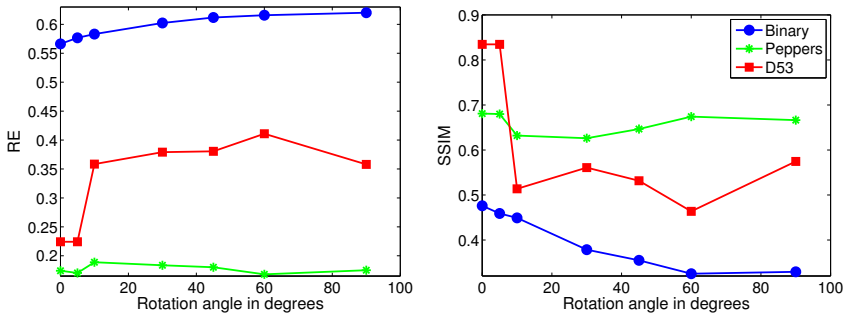


Figure 13: Left: The plots for the average reconstruction errors (RE) and, Right: SSIM measures versus the rotation degrees for our three test images, where 4 smaller test images of size 50×50 are extracted from each rotated test image.

The plots in Figure 13 show that while, as expected, no particular sensitivity

trends for the peppers test image can be detected by changing the rotation angles, the binary test image and D53 are highly sensitive to rotation and the worst reconstructions in term of RE and SSIM measures are obtained with the 60° and 90° rotation.

4.4 An Algorithm to Determine Rotation Angle

If the angle of rotation is known in advance to the reconstruction step, one can learn the dictionaries with larger patch sizes and then rotate the dictionary images by the known angle. The pixels that fall outside the boundaries of the original dictionary image are, in MATLAB, set to 0 and appear as a black background in the rotated image. However, we can specify a smaller patch size and exclude the boundary pixels with zero values in the rotated dictionary elements and extract rotated dictionary images of smaller size than the original one, to include in the reconstruction step.

If an image is given, the principal direction of the image can be estimated from the Radon transform of the image [7]. The Radon transform can be used to detect linear trends in images. For general images, the principal orientation may be taken as the direction along which the Radon transform has the maximum variability.

Let v_j denote the variance of the sinogram data for the jth projection, i.e., the jth column of the sinogram matrix S:

$$v_j = \frac{1}{N_\mathrm{r} - 1} \sum_{k=1}^{N_\mathrm{r}} \left(S_{k,j} - \mathscr{M}_j \right)^2, \quad \forall\ j = 1, \ldots, N_\mathrm{p},$$

where $\mathscr{M}_j$ is the mean of each column vector in S,

$$\mathscr{M}_j = \frac{1}{N_\mathrm{r}} \sum_{k=1}^{N_\mathrm{r}} S_{k,j}, \quad \forall\ j = 1, \ldots, N_\mathrm{p}.$$

An important observation in [13] was that the sinogram $R_\theta X(t)$ along θ has larger variations with respect to t for the principal angle with most directional lines. Hence in our case with angles θ_j, $j = 1, \ldots, N_\mathrm{p}$:

$$\Theta = \theta_{j^\star}, \quad j^\star = \arg\max_j v_j$$

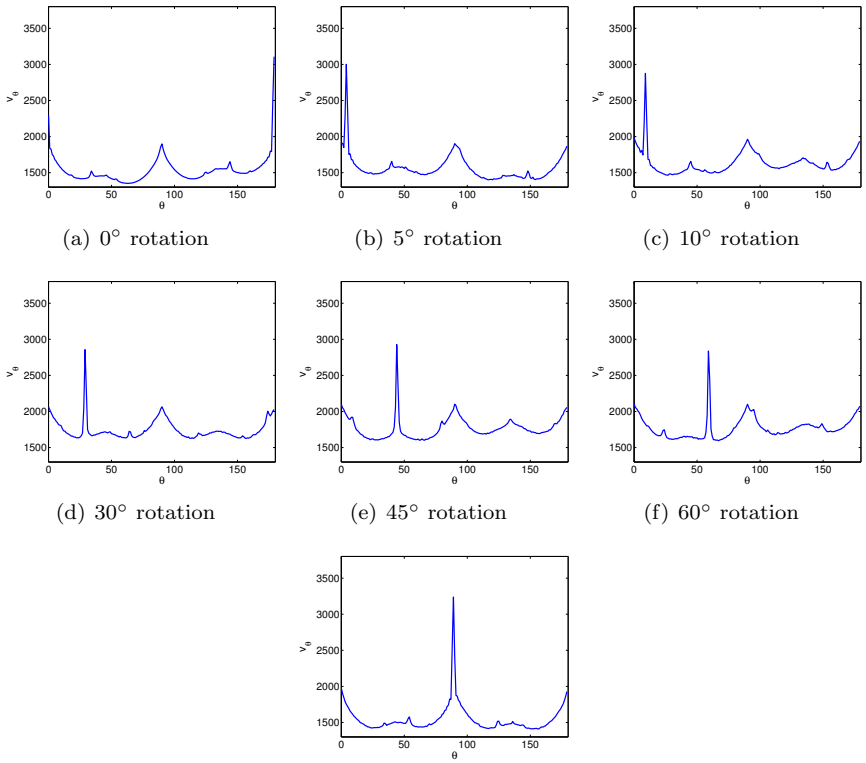
is the direction with most linear trends along it. Such an estimate is useful for estimating the presence of rotation in the images.

We can assume that $\tilde{z} \in \mathbb{R}^n$ is a sub-image from the training image of a similar size as the unknown image x. We compute the sinogram of $\tilde{z}$ by generating the tomographic data by $A\tilde{z}$ and representing it as a matrix. We compute $\max_\theta \tilde{v}_\theta$ and find $\tilde{\Theta}$ to be the angle of most directional trends in the sub-image $\tilde{z}$. We refer to $\tilde{\Theta}$ as the reference angle of the training image. Similarly, we compute $\Theta = \arg\max_\theta v_\theta$ for the unknown image x. Then the rotation is approximately the difference between the angles, i.e., $\Theta - \tilde{\Theta}$.

To test this claim, let us choose 200×200 test images – similar to the D53 test image given in Fig. 12 – rotated by [5°, 10°, 30°, 45°, 60°, 90°], making six test images. We consider a training image with no rotation, i.e., with rotation angle 0° of size 200×200. In our first computational test, to find the correct

rotation angle, we consider a tomographic scenario with a full data set, i.e., projections from all possible angles. The $N_{\mathrm{p}} = 180$ projections are sampled with equidistant steps over $[0°, 180°]$, moreover we consider $N_{\mathrm{r}} = 283$ and 1% noise in the data.

Figure 14 shows the variance plots of the sinograms of our training image and rotated test images with different orientations. The sinogram of the reference training image with no rotation is noise free, while noise is present in the sinograms of the rotated test images. Note that the variance of the projections has two local maxima at 90° and 179° for the reference training image with no rotation. The local maximum at 179° is narrower compared with the local maximum at 90°, because there are more straight lines along 179°. Hence 179° is the reference orientation.



(a) 0° rotation (b) 5° rotation (c) 10° rotation

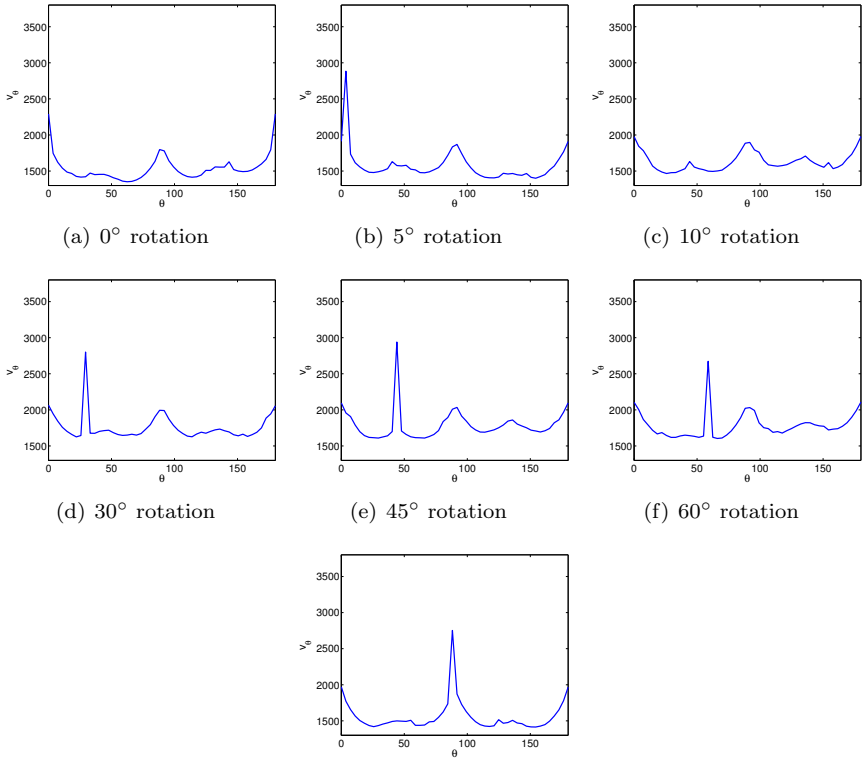
(d) 30° rotation (e) 45° rotation (f) 60° rotation

(g) 90° rotation

Figure 14: The variance of sinograms from the 200 × 200 D53 six test images with different rotation angles with full tomographic data comparing to a similar training image with no rotation of 200 × 200 size. Note how the maximum in the variance plots changes as the rotation degrees varies.

Given the plots in Fig. 14, we calculate the rotation degrees by finding the angle with the maximum variance in each plot, the difference to the original orientation in the reference training image gives the correct rotation. The estimations based on the full tomographic data are accurate and we obtain all the rotation angles, i.e., 5°, 10°, 30°, 45°, 60°, and 90°.

We now consider tomographic data with data from few projections of the same directional D53 images, we use 50 projections with uniform angular sampling in $[0°, 180°]$ and with relative noise level 1%, i.e., the same noise level as above. The variances of the sinograms of the training image and the test images with rotational angles [5°, 10°, 30°, 45°, 60°, 90°] are given in Figure 15.



(a) 0° rotation (b) 5° rotation (c) 10° rotation

(d) 30° rotation (e) 45° rotation (f) 60° rotation

(g) 90° rotation

Figure 15: The variance of sinograms from the six 200×200 D53 test images with different rotation angles with limited tomographic data compared to a similar training image with no rotation of 200×200 size. Note how the maximum in the variance plots changes as the rotation degrees varies. With the limited tomographic data, the maximum disappears when rotating the reference image with 10°.

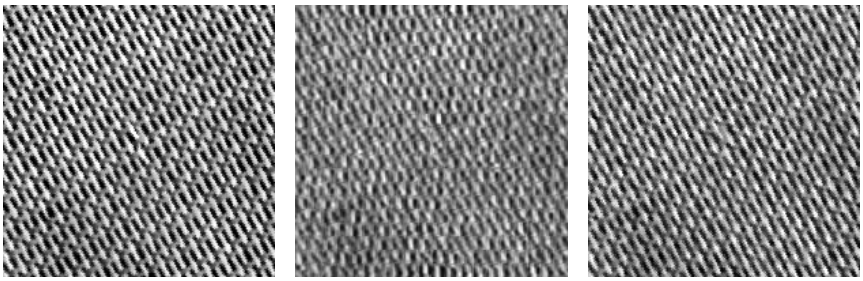
The variance plots in Fig. 15 indicate that with limited tomographic data where the projection data along some directions are missing and the information of the variances along all the directions are not attainable, we may not be able to find the correct orientation of the directional textures in the image. Note how the peak in the variance plot with the 10° rotation is missing. We find the rotation angles to be

$$[3.67°, 180°, 29.39°, 44.08°, 58.78°, 88.16°].$$

We observe that the method fails to find the correct orientation for the image with 10° rotation. One possible way to compensate for the missing projection

data and construct new data points for these missing projections from the known ones, is to use interpolation of the tomographic data in the sinogram. Using linear 2D interpolation for gridded data, we approximate the rotated angles as before, where we obtain [4°, 0°, 29°, 44°, 59°, 88°] as the rotations. Although we still can not achieve the correct orientation for the image with 10° rotation, in the presence of noise we can still approximate other rotation angles with a small error.

To complete this picture, we consider a tomographic problem where an exact image is given in Fig. 16. This exact image is rotated by 30° from the reference training image. We consider the same tomographic scenario with 50 projections in $[0°, 180°]$ and 1% noise. By the above method for the noisy sinogram we approximate the rotation angle to be 29°. A dictionary from 20×20 patches from the training image, i.e., $D \in \mathbb{R}^{400\times 800}$, is computed; each dictionary image is rotated by 29° and then 10×10 dictionary elements are extracted from the rotated 20×20 dictionary basis images, and then 300 dictionary elements are randomly chosen from these 800 rotated dictionary images. Now a new rotated dictionary such that $\bar{D} \in \mathbb{R}^{100\times 300}$ is at hand. We reconstruct the image using the rotated dictionary $\bar{D}$ and compare it with a reconstruction obtained using 10×10 dictionary elements and $s = 300$, obtained from the reference training image with 0° rotation. The results are illustrated in Fig. 16 which shows clearly how using a correctly rotated dictionary can improve the reconstruction significantly.



(a) x^{exact} (b) RE= 0.3354 (c) RE= 0.2507

Figure 16: Left: The 30° rotated exact image. Middle: The tomographic reconstruction using a dictionary obtained from our reference training image without any knowledge of rotation. Right: The reconstructed solution with a rotated dictionary where the degree of rotation is approximated from the noisy sinogram of the tomographic data.

5 Conclusions

This work is an extension of the work initialized in [22]. Our approach is similar to the work in [22] where we discussed the use of training images as prior information in the tomographic image reconstruction problem in a two-stage framework, while here, we continued this work by numerically investigating the sensitivity of the reconstruction formulation to the representation by the dictionary as well as inconsistency in scale and rotation of the unknown image to the

training images. In addition, we suggested algorithms to determine the correct scale and rotation degree of the unknown image from the tomographic sinogram. Numerical examples showed that both methods can be advantageous in obtaining the correct scale and rotation of the unknown image from the measurement data, however future work concerning approximating the correct scale of unknown textural images from the given sinogram where the proposed method fails should be considered.

Acknowledgments

The author acknowledges collaboration with Per Christian Hansen and Martin S. Andersen from DTU Compute. We would like to thank professor Samuli Siltanen from University of Helsinki for providing the high-resolution images of the peppers and matches.

References

[1] Aharon, M., Elad, M., and Bruckstein, A.M.: *K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation.* IEEE Trans. Signal Process. **54**, 4311–4322 (2006)

[2] Bar L. and Sapiro, G.: *Hierarchical invariant sparse modeling for image analysis.* in Proc. IEEE Int. Conf. Image Process., 2397–2400 (2011)

[3] Barthelemy, Q., Larue, A., Mayoue, A., Mercier, D., and Mars, J.I.: Shift and 2-D rotation invariant sparse coding for multivariate signals. IEEE Trans. Signal Process. **60**(4), 1597–1611 (2012)

[4] Becker, S., Candès, E.J., and Grant, M.: *Templates for convex cone problems with applications to sparse signal recovery.* Math. Prog. Comp. **3**, 165–218 (2011)

[5] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J.: *Distributed optimization and statistical learning via the alternating direction method of multipliers.* Found. Trends Mach. Learn. **3**, 1–122 (2011)

[6] Bruckstein, A.M., Donoho, D.L., and Elad, M.: *From sparse solutions of systems of equations to sparse modeling of signals and images.* SIAM Review **51**(1), 34–81 (2009)

[7] Chen, Y.-C., Sastry, C.S., Patel, V.M., Phillips, P.J., and Chellappa, R.: *In-plane rotation and scale invariant clustering using dictionaries.* IEEE Trans. Image Process. **22**(6) (2013)

[8] Elad, M.,: *Sparse and Redundant Representations, From Theory to Applications in Signal and Image Processing.* Springer, New York (2010)

[9] Elad, M. and Aharon, M.: *Image denoising via sparse and redundant representations over learned dictionaries.* IEEE Trans. Image Process. **15**, 3736–3745 (2006)

[10] Hansen, P.C. and Saxild-Hansen, M.: *AIR Tools - A MATLAB package of algebraic iterative reconstruction methods.* J. Comput. Appl. Math. **236**, 2167–2178 (2012)

[11] Hotta, K., Kurita, T., and Mishima, T.: *Scale invariant face recognition method using spectral features of Log-Polar image.* Proceedings of SPIE **3808**, 33–43 (1999)

[12] Hoyer, P.O.: *Non-negative matrix factorization with sparseness constraints.* J. Mach. Learn. Res. **5**, 1457–1469 (2004)

[13] Jafari Khouzani, K. and Soltanian Zadeh, H.: *Radon Transform Orientation Estimation for Rotation Invariant Texture Analysis.* IEEE Trans. Pattern Anal. Mach. Intell. **27** (6), 1004–1008 (2005)

[14] Mairal, J., Bach, F., Ponce, J., and Sapiro, G.,: *Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res.* **11**, 19–60 (2010)

[15] Mairal, J., Sapiro, G., and Elad, M.: *Learning multiscale sparse representations for image and video restoration.* SIAM Multiscale Model. Simul. **7**, 214–241 (2008)

[16] Mairal, J., Sapiro, G., and Elad, M.: *Multiscale sparse image representation with learned dictionaries.* IEEE International Conference on Image Processing (ICIP), (2007)

[17] Mueller, J.L. and Siltanen S.: *Linear and nonlinear inverse problems with practical applications.* SIAM, (2012)

[18] Liu, Q., Liang, D., Song, Y., Luo, J., Zhu, Y., and Li, W.: *Augmented Lagrangian-based sparse representation method with dictionary updating for image deblurring.* SIAM J. Imaging Sci. **6**, 1689–1718 (2013)

[19] Lowe, D.G.: *Object recognition from local scale-invariant features.* Proceedings of the International Conference on Computer Vision (ICCV)**2**, 1150–1157 (1999)

[20] Olshausen, B.A. and Field, D.J.: *Emergence of simple-cell receptive field properties by learning a sparse code for natural images.* Nature, **381** (1996), 607–609.

[21] Ophir, B., Lustig, M., and Elad, M.: *Multi-scale dictionary learning using wavelets.* IEEE Journal of Selected Topics in Signal Processing (J-STSP) **5**(5), 1014–1024 (2011)

[22] Soltani, S., Andersen, M.S., and Hansen, P.C.: *Tomographic image reconstruction using dictionary priors.* arxiv.org/abs/1503.01993, submitted to Inverse Probl. Imaging (2014)

[23] Tošsić, I., Jovanović, I., Frossard, P., Vetterli, M., and Duric, N.: *Ultrasound tomography with learned dictionaries*, IEEE Int. Conf. Acoust., Speech, Signal Processing, 5502–5505 (2010)

[24] Tropp, J.A., Wright, S.J. *Computational methods for sparse solution of linear inverse problems.* Proc. IEEE, **98**(6), 948–958, (2010)

[25] Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P.: *Image quality assessment: From error visibility to structural similarity.* IEEE Trans. Image Process. **13**(4), 600–612 (2004)

[26] Xu, Q., Yu, H., Mou, X., Zhang, L., Hsieh, J., and Wang, G.: *Low-dose X-ray CT reconstruction via dictionary learning.* IEEE Trans. Med. Imag. **31**, 1682–1697 (2012)

[27] http://multibandtexture.recherche.usherbrooke.ca/normalized_brodatz.html