Filter-based reconstruction methods for tomography

Pelt, D.M.

Author: Pelt D.M.
Title: Filter-based reconstruction methods for tomography
Issue Date: 2016-05-03

1 Introduction and outline

In this chapter, an introduction is given to tomography and the problem of tomographic reconstruction. A mathematical formulation of the problem is given, and several standard solution methods are explained. We will discuss the need for developing new efficient and accurate methods for use in modern tomographic experiments where the standard methods fail to produce sufficiently accurate results, and more advanced methods have computational costs that are too high to be used in practice. Finally, an overview is given of the main results of this thesis, in which several efficient and accurate methods are introduced.

1.1 Tomography

In many applications, it is useful to have a way of looking inside an object without destroying it. For example, in medical applications, being able to examine the internals of a patient without needing surgery is helpful for diagnosis. *Radiography* is often used in hospitals to perform this task. In radiography, a patient is briefly illuminated by a source of X-rays, and the rays passing through the patient are collected by a detector. Since different parts of the body absorb different amounts of X-rays that pass through, the resulting image on the detector shows the internal structure of the body. Bones, for example, are highly absorbing, and as such are usually clearly visible in the radiograph. An example of a radiograph is shown in Fig. 1.1.

The result of radiography is a two-dimensional (2D) image of the three-dimensional (3D) internal structure of the patient. Information about the structure in the direction of the X-rays is lost: in a radiograph, it is impossible to see whether a certain feature is located at the front or the back of the patient. Specialists are able to correctly diagnose patients using a 2D radiograph in many cases, since a lot of information is known about the human body. In other cases, however, the complete 3D information about the

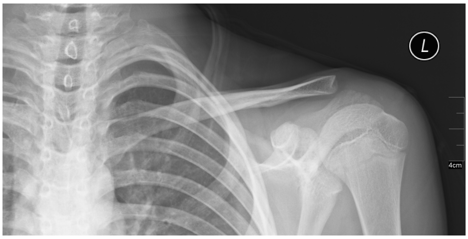
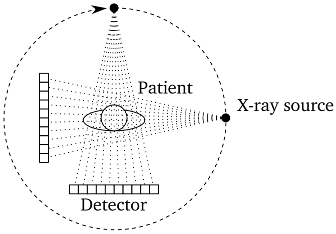


Figure 1.1: Radiograph of a human shoulder (© Nevit Dilmen).



(a)

(b)

Figure 1.2: (a) Schematic overview of a computed tomography scanner. (b) A single slice of a reconstructed tomographic dataset (source: James Heilman, MD).

patient is needed. For example, when diagnosing or treating a tumor, it is important to know the exact position of the tumor in the body. *Computed tomography* is often used in these cases to acquire a 3D view of the internals of a patient.

In computed tomography, multiple radiographs are acquired while rotating the X-ray source and detector around the patient. In this way, X-ray images, called *projections* in this context, are acquired for several angles. A schematic overview of a computed tomography scanner is shown in Fig. 1.2a. The acquired 2D projections can be used to compute the 3D internal structure of the patient by a mathematical process called *tomographic reconstruction*. As an example, a single slice of a reconstructed 3D image of a patient's internals is shown in Fig. 1.2b. Various algorithms can be used to reconstruct tomographic data, which will be explained in more detail in Section 1.2. Tomography is used routinely in many applications other than medical diagnosis. Examples include material science [Sal+03], biomedical research [Lov+13], and industrial applications [Sip93]. A wide variety of radiation sources and scanning devices can be used, with length scales ranging from the nanoscale in electron tomography [Sco+12] to the astronomical scale in astrotomography [BSC01].

In this thesis, we will focus mainly on tomographic datasets acquired at a synchrotron radiation facility and datasets acquired with an electron microscope. Synchrotron radiation facilities, or *synchrotrons*, accelerate electrons in a circular path that can be several kilometers in length by applying strong magnetic fields at specific points along the path. At these points, a highly intense beam of photons is produced by the interaction of the high-energy electrons with the magnetic field. The photon beam has many useful properties for tomographic experiments, such as a high energy, flux, brilliance, and stability, and synchrotrons are routinely being used for advanced high-resolution tomographic experiments at the μm scale. Electron microscopes use a beam of electrons instead of photons to produce images. Since the wavelength of electrons is much smaller than the wavelength of X-rays, electron microscopes are able to image much smaller features compared to X-ray scanners. To perform *electron tomography*, i.e. tomographic experiments with electron microscopes, the sample is usually mounted on a holder that can rotate the sample within the beam. Because of physical restrictions of the system, the angular range for which projections can be acquired is typically limited in electron tomography. In most tomographic experiments performed with either synchrotrons or electron microscopes, the X-ray or electron beam does not diverge significantly when passing through the object, and can be regarded as a *parallel* beam. Mathematically, the fact that the beam is non-diverging has advantageous properties that we will exploit in this thesis.

1.2 Tomographic reconstruction

An important part of tomography is the reconstruction of the 3D structure using the acquired 2D projection images. In this section, we will define the reconstruction problem mathematically and give a brief overview of standard tomographic reconstruction methods. We will restrict the explanation to 2D parallel-beam problems, where the goal is to reconstruct a 2D image from one-dimensional (1D) projections, also called sinograms in this context. Note that 3D parallel-beam problems can usually be regarded as a collection of separate 2D parallel-beam problems for each slice.

Mathematically, we model the scanned object as a finite and integrable function $f : \mathbb{R}^2 \to \mathbb{R}$ with bounded support. We can define a single ray passing through the object as a line $l_{\theta,t}$ with the characteristic equation $t = x\cos\theta + y\sin\theta$. The line integral of f over a single line $l_{\theta,t}$, written as a function $P_\theta : \mathbb{R} \to \mathbb{R}$, is given by:

$$P_\theta(t) = \int_{l_{\theta,t}} f(x,y)\,\mathrm{d}s \tag{1.1}$$

$$= \iint_{\mathbb{R}^2} f(x,y)\delta(x\cos\theta + y\sin\theta - t)\,\mathrm{d}x\mathrm{d}y \tag{1.2}$$

Here, δ is the Dirac delta function. The goal of reconstruction in 2D parallel beam tomography is to recover the unknown function f given its line integrals $P_\theta(t)$ for different combinations of θ and t. The geometry is shown graphically in Fig. 1.3.

In practice, projection data are acquired only for a finite number of angles θ, using a finite set of detectors that each detect a single ray t per angle. If N_θ angles are used with N_d detectors per angle, the acquired dataset consists of $N_\theta N_d$ line integrals, one for each combination of angle $\theta \in \Theta = \{\theta_0, \ldots, \theta_{N_\theta-1}\}$ and detector $d \in \{0, \ldots, N_d - 1\}$. The position τ_d of a detector d relative to the central detector is given by $\tau_d = s(d-(N_d-1)/2)$, and the set of detector positions by $T = \{\tau_0, \ldots, \tau_{N_d-1}\}$.

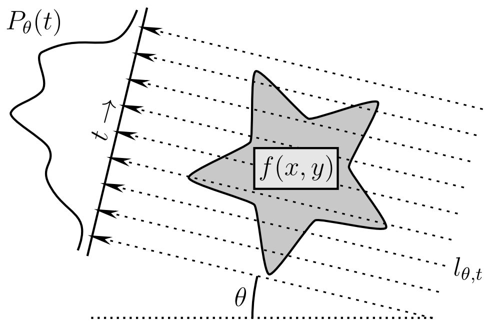


Figure 1.3: The two-dimensional parallel-beam tomography model. Parallel lines, rotated by angle θ, pass through the object f. A line $l_{\theta,t}$ has the characteristic equation $t = x\cos\theta + y\sin\theta$, and a projection $P_\theta(t)$ of f is given by the line integral of f over the line $l_{\theta,t}$.

In many cases, the unknown function f is reconstructed on a grid of $N \times N$ pixels, where N is often chosen to be equal to N_d.

We can write the acquired projections as a vector $\boldsymbol{p} \in \mathbb{R}^{N_\theta N_d}$ with $N_\theta N_d$ elements. Similarly, the reconstructed image can be written as a vector $\boldsymbol{x} \in \mathbb{R}^{N^2}$ with N^2 elements. Using these definitions, we can write the tomographic acquisition process as a linear equation:

$$\boldsymbol{p} = \boldsymbol{W}\boldsymbol{x} \tag{1.3}$$

Here, $\boldsymbol{W}$ is a $N_\theta N_d \times N^2$ matrix that corresponds to computing the line integrals of the object $\boldsymbol{x}$. In the matrix $\boldsymbol{W}$, the element w_{ij} specifies the contribution of pixel j to detector i. The multiplication of an image $\boldsymbol{x}$ by $\boldsymbol{W}$ is called the *forward projection* of $\boldsymbol{x}$, and the multiplication of a vector $\boldsymbol{p}$ by $\boldsymbol{W}^T$ is called the *backprojection* of $\boldsymbol{p}$. Note that for typical sizes of tomographic datasets, the dimensions of $\boldsymbol{W}$ will be too large to store it in computer memory. Instead, forward projections are usually computed on-the-fly by computing the line integrals of an image $\boldsymbol{x}$ directly [PBS11], with backprojections computed on-the-fly as well. The projection operations can be computed efficiently using graphic processor units (GPUs), which helps to reduce reconstruction times in practice [XM05; MXN07].

Two types of methods are commonly used to reconstruct the unknown object from the acquired projections: *analytical* reconstruction methods and *algebraic* reconstruction methods.

Analytical reconstruction

Analytical reconstruction methods are based on taking the continuous model of tomographic acquisition (Eq. (1.2)) and inverting it to find an expression for $f(x, y)$. The result of this approach for 2D parallel-beam tomography is the *filtered backprojection* method (FBP). FBP starts by convolving the acquired projection data $P_\theta(t)$ with a filter $h : \mathbb{R} \to \mathbb{R}$:

$$q_\theta(t) = \int_{-\infty}^{\infty} h(\tau) P_\theta(t - \tau) \mathrm{d}\tau \tag{1.4}$$

This convolution operation can be efficiently performed in Fourier space, with $\hat{h}$ and $\hat{P}_\theta$ being the Fourier transforms of h and P_θ:

$$q_\theta(t) = \int_{-\infty}^{\infty} \hat{h}(u)\hat{P}_\theta(u)e^{2\pi\imath ut}\mathrm{d}u \tag{1.5}$$

By taking $\hat{h}(u) = |u|$, we obtain an expression for $f(x, y)$ [KS01]:

$$f(x, y) = \int_0^{\pi} q_\theta(x\cos\theta + y\sin\theta)\mathrm{d}\theta \tag{1.6}$$

In practice, projections are only acquired for a finite number of angles θ and detectors τ, as explained above. Therefore, we have to discretize Eq. (1.6) to be able to obtain the filtered backprojection method:

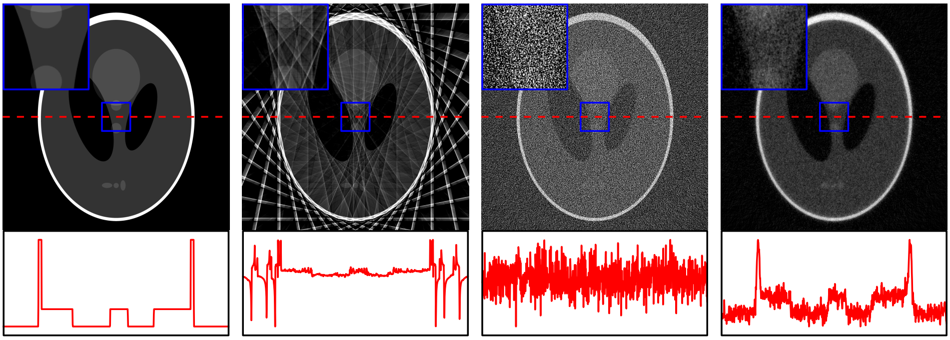
$$f(x, y) \approx FBP_h(x, y) = \sum_{\theta_d\in\Theta}\sum_{\tau_p\in T} h(\tau_p)P_{\theta_d}(t - \tau_p) \tag{1.7}$$

where $t = x\cos\theta + y\sin\theta$. Since $t - \tau_p$ is not necessarily equal to one of the acquired detector positions, some interpolation is needed to find the value of $P_{\theta_d}(t - \tau_p)$. Linear interpolation is often used, since projection data are usually reasonably smooth. The filter h is only needed at a finite number of discrete positions $h(\tau_p)$, and is often defined as a vector $\boldsymbol{h} \in \mathbb{R}^{N_d}$. Several standard filters are commonly used in practice, such as the Ram-Lak (ramp), Shepp-Logan, and Hann filters [Far+97; WWH05]. One of the most popular filters is the Ram-Lak filter, which is obtained by discretizing the theoretical $\hat{h}(u) = |u|$ filter. In the matrix and vector notation, the FBP method can be written as:

$$\boldsymbol{x}_{FBP} = FBP_{\boldsymbol{h}}(\boldsymbol{p}) = \boldsymbol{W}^T\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{p} \tag{1.8}$$

Here, $\boldsymbol{C}_{\boldsymbol{h}}$ is a convolution operation that convolves each 1D array of detector values, taken at a single rotation angle, with the 1D filter $\boldsymbol{h}$.

The filtered backprojection method is computationally efficient, since the convolution operation can be computed using the Fast Fourier Transform method, and only a single backprojection is needed to reconstruct an image. The reconstruction quality, however, depends on how well the discretized equation (Eq. (1.7)) approximates the continuous inverse equation (Eq. (1.6)). The continuous inverse equation assumes that noise-free projection data are available for an infinite number of rotation angles between 0 and π, which is not possible to achieve in practice. If only a limited number of projections are acquired or when there is noise present in the acquired projections, reconstructions computed with the FBP method tend to contain significant artifacts that can make further analysis impossible. Examples of typical artifacts found in FBP reconstructions are shown in Fig. 1.4. When sufficiently many projections are available and the noise in them is sufficiently small, however, the FBP method is usually able to produce reconstructions that are accurate enough for analysis. For this reason, and because of its computational efficiency, filtered backprojection remains the most popular reconstruction method in many applications of tomography [PSV09].



(a) (b) (c) (d)

Figure 1.4: Three reconstructions of the Shepp-Logan head phantom (a), showing artifacts that occur with imperfect data. In (b) and (c), FBP was used to reconstruct the images. In (b), data from only 16 projections were used, and severe streak artifacts are present in the result. In (c), data from 1024 projections were used, but a large amount of Poisson noise was present, resulting in severe noise in the resulting image. In (d), SIRT, an algebraic method, was used to reconstruct the image using the same noisy projection data as in (c). The algebraic reconstruction image (d) has less noise compared to the FBP reconstruction (c). Under each image the line profile of the middle row is shown, and a small section is shown enlarged in the upper left of each image.

Algebraic reconstruction

Algebraic reconstruction methods are based on the discrete matrix representation of the tomographic reconstruction problem (Eq. (1.3)). Specifically, the reconstruction problem is written as a system of linear equations, which is then solved by an optimization method. Most algebraic methods minimize the difference between the forward projection of the reconstructed image with the acquired projection data, which is called the *projection error*. A popular choice is to minimize the ℓ_2-norm of the projection error, in which case we can write the algebraic approach as the following optimization problem:

$$\boldsymbol{x}_{alg} = \underset{\boldsymbol{x}\in\mathbb{R}^{N^2}}{\operatorname{argmin}} \|\boldsymbol{p} - \boldsymbol{W}\boldsymbol{x}\|_2 \tag{1.9}$$

Because of the size of the matrix $\boldsymbol{W}$, it is often impossible to use direct methods such as singular value decomposition to find a solution to Eq. (1.9). Instead, methods are used in which the projection error is iteratively minimized. Different mathematical optimization methods can be used to minimize the projection error, leading to different algebraic reconstruction methods. For example, if the projection error is minimized by Landweber iteration [Lan51], i.e. using gradient-descent steps on the projection error, the result is the Simultaneous Iterative Reconstruction Technique (SIRT). Iterations of the SIRT method can be written as:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha \boldsymbol{W}^T \left(\boldsymbol{p} - \boldsymbol{W}\boldsymbol{x}^k\right) \tag{1.10}$$

Here, $\alpha \in \mathbb{R}$ is a parameter that influences the convergence rate and stability of the method. Other examples of standard algebraic methods are the CGLS method, where the projection error is minimized by a Conjugate Gradient method [HS52], and the Algebraic Reconstruction Technique (ART) [GBH70], where the error is minimized by the Kaczmarz method [Kac37].

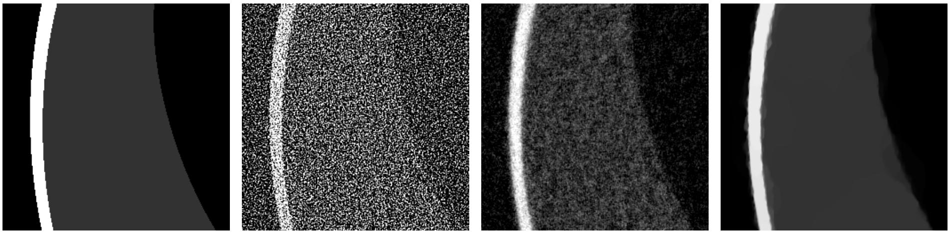
Since algebraic methods use a model of the tomographic reconstruction problem that includes only the projections that are actually acquired instead of assuming that an infinite number of projections are available, they tend to handle problems with a limited number of projections better than analytical methods. Furthermore, the effect of noise in the projection data can be limited in most algebraic methods by stopping the iteration process early, which is a form of implicit regularization. A comparison between a reconstruction computed by the algebraic SIRT method with a reconstruction computed by the analytical FBP method is shown in Fig. 1.4. Note that the artifacts caused by the noise are significantly reduced in the algebraic reconstruction, shown in Fig. 1.4d, compared to the analytical reconstruction, shown in Fig. 1.4c.

One of the main disadvantages of algebraic methods is their high computational costs. Several tens or hundreds of iterations are typically needed in algebraic methods to converge to an acceptable reconstruction image, with multiple projection operations per iteration. For the large datasets that are routinely acquired at experimental facilities, the high computational costs can be prohibitive in practice. For example, suppose we want to reconstruct a 1024^3 3D volume using 1024 projections of 1024×1024 pixels each, which would be a medium-sized problem in synchrotron tomography. Using a state-of-the-art GPU system, it would take around 80 seconds to reconstruct the full volume with the analytical FBP method, while reconstructing with 200 iterations of the algebraic SIRT method would take around one and a half hours. Since a single tomographic dataset can often be acquired in a few minutes or less, the reconstruction time of algebraic methods tends too be too long to routinely use them in practice, and FBP is used instead.

In advanced tomographic experiments, one is often restricted to acquiring only a very limited number of projections, and a large amount of noise can be present in each projection. In these cases, algebraic methods are often also unable to produce reconstructions that are sufficiently accurate for further analysis, similar to analytical methods. The reason for this is that the linear system that is solved in algebraic methods (Eq. (1.9)) is highly underdetermined if $N_\theta N_d \ll N^2$. The result is that there exist infinitely many reconstructions that have the same projection error, most of which are not suitable for analysis. It depends on the specific optimization method that is used which reconstruction is computed by an algebraic method. Furthermore, the projection matrix $\boldsymbol{W}$ is ill-conditioned in most applications of tomography. This means that even minor inconsistencies in the projection data, such as noise, can have a large effect on the reconstructed image.

The reconstruction quality of algebraic methods can be improved by exploiting prior knowledge about the scanned object or scanning system. Mathematically, this prior knowledge is often encoded as an additional term $g : \mathbb{R}^{N^2} \to \mathbb{R}$ in the objective function that is minimized, resulting in *regularized iterative methods*:

$$\boldsymbol{x}_{reg} = \underset{\boldsymbol{x} \in \mathbb{R}^{N^2}}{\operatorname{argmin}} \left[\|\boldsymbol{p} - \boldsymbol{W}\boldsymbol{x}\|_2 + \lambda g(\boldsymbol{x}) \right] \tag{1.11}$$



(a) (b) (c) (d)

Figure 1.5: Zoomed-in reconstructions of the Shepp-Logan head phantom (a), showing the resulting images of three different reconstruction methods: (b) FBP, (c) SIRT, and (d) total variation minimization. The images were reconstructed on a 1024 × 1024 pixel grid, using projection data acquired with $N_d = 1024$ detectors and $N_\theta = 256$ projection angles, equally distributed in the interval $[0, \pi]$, and additional Poisson noise applied.

Here, $g(\boldsymbol{x})$ is a penalty function that penalizes undesired solutions that do not fit with the assumed prior knowledge, and the λ term controls how strongly the penalty function is weighted compared to the projection error term. For example, when it is assumed that the gradient of the reconstructed object is sparse, a popular choice is to use *Total Variation minimization* (TV-minimization) by setting $g(\boldsymbol{x}) = \|\nabla \boldsymbol{x}\|_1$, where ∇ is a discrete gradient operator [SP08]. If the assumed prior knowledge prior knowledge is appropriate for the acquired data, regularized iterative methods can be extremely successful in reconstructing objects from (highly) limited data [BS11; Kos+13]. A comparison between FBP, SIRT, and TV-minimization reconstructions is shown in Fig. 1.5 for noisy projection data. A major disadvantage of regularized iterative methods is their high computational costs, which tend to be even higher than those of algebraic methods. For example, when reconstructing the 1024^3 volume defined above on the same state-of-the-art GPU system, it would take more than a *day* to reconstruct the full volume using the FISTA method for TV-minimization [BT09a].

1.3 Overview of this thesis

As explained above, in many applications of tomography, analytical methods produce reconstructions that are not accurate enough for further analysis. More accurate reconstructions can be obtained by using (regularized) iterative methods, but these can have computational costs that are too high to be used in practice. In this thesis, new reconstruction methods are developed that combine the analytical and algebraic approaches, resulting in methods that are as computationally efficient as analytical methods, but with a reconstruction accuracy of algebraic methods. Analytical methods allow for changing the filter $\boldsymbol{h}$ without increasing the needed computation time. We will use this freedom in filter choice to develop new *filter-based* reconstruction methods, which are based on the FBP method with specific filters. The filters can be defined and computed in different ways, and can depend on the acquisition geometry, the scanned object, and/or a separate pre-computing step. Several filter-based methods are introduced in this thesis, and reconstruction results are compared with other popular methods. In the rest of this section, the contributions and results of each chapter are explained.

In Chapter 2, the MR-FBP method is introduced, which uses a filter that minimizes the projection error of the resulting FBP reconstruction. The filter can be computed using an approach that is similar to algebraic reconstruction methods, i.e. minimizing the ℓ_2-norm of the projection error:

$$\boldsymbol{h}^* = \underset{\boldsymbol{h} \in \mathbb{R}^{N_d}}{\operatorname{argmin}} \|\boldsymbol{p} - \boldsymbol{W} FBP_{\boldsymbol{h}}(\boldsymbol{p})\|_2 \tag{1.12}$$

Note that the filter that is computed depends on the acquired data. The results of Chapter 2 show that the method is able to produce reconstructions with similar reconstruction quality as the algebraic SIRT method, but is much faster at producing them. Furthermore, it is shown that some forms of prior knowledge can be exploited to improve reconstruction quality, and that the computed filters automatically adapt to the characteristics of the object and the scanning parameters.

A different approach is taken in Chapter 3, where the SIRT-FBP method is introduced. This method explicitly approximates the SIRT method by the FBP method with a specific filter. The approximation is achieved by first rewriting the SIRT method into a matrix form, and then approximating k iterations of SIRT by a single convolution operation. By comparing with the standard FBP method, it is possible to show that the approximated SIRT method is identical to the FBP method with a specific angle-dependent filter. The filter can be pre-computed for a certain scanning geometry by a single SIRT-like iteration method, after which it can be reused for datasets that are acquired with the same geometry. The results of Chapter 3 show that reconstructions computed with the SIRT-FBP method are virtually identical to standard SIRT reconstructions.

In Chapter 4, a method is introduced that allows one to approximate a slow regularized iterative method inside a (small) subvolume of the complete reconstruction volume. If one is only interested in a small subvolume of the scanned object, as is often the case, this method can significantly reduce the computation time that is needed to perform the required analyses. Note that regularized iterative methods generally need to compute the entire reconstruction volume, since they are based on minimizing a global objective function (Eq. (1.11)). The local approximation method is based on extending the SIRT-FBP method introduced in Chapter 3 to allow for different types of additional regularization terms. In the results of Chapter 4, we show that the reconstruction quality of the local approximations is almost identical to that of the much slower global regularized iterative methods for several popular types of prior knowledge, such as TV-minimization.

The NN-FBP method is introduced in Chapter 5. An NN-FBP reconstruction consists of a nonlinear combination of multiple FBP reconstructions, each with a different filter. The filters are pre-computed by using projection data and high-quality reconstructions of objects that are similar to the objects that will be reconstructed later. Methods from neural network theory are used to *train* the filters, after which they can be used to accurately reconstruct similar objects even when only a limited number of projections are acquired. The results of Chapter 5 show that the NN-FBP method is able to produce reconstructions with a significantly higher quality than both analytical and algebraic reconstruction methods. Also, it is shown that the method can be used to approximate slow regularized iterative methods in the case that it is not possible to acquire a large number of projections for use in training.

An application of the NN-FBP method in electron tomography is presented in Chapter 6. In electron tomography, acquiring a large number of projections is both time-consuming and labor-intensive. As a result, it is difficult to scan a large number of samples and obtain statistically significant results for certain sample characteristics. By lowering the number of projections that need to be acquired, the effort needed to obtain statistically significant results can be reduced. In Chapter 6, we use the NN-FBP method to obtain accurate reconstructions of gold nanoparticles using a very limited number of projections. The NN-FBP method is trained on a small number of nanoparticles that are scanned with a higher number of projections. Results show that the NN-FBP method is able to produce accurate reconstructions with only a few projections, and that the method can be used to obtain statistically significant results with reduced scanning time.

In Chapter 7, a different problem is addressed compared to the other chapters of this thesis. An additional reason that (regularized) iterative methods are not used routinely at experimental facilities is a practical one: it is often difficult to install and use software that can perform these reconstructions at the facilities. In Chapter 7, we improve this situation for the synchrotron community by combining a software toolbox that is specifically designed to be easy-to-use at synchrotrons, TomoPy [Gür+14], with a toolbox that can be used to develop advanced reconstruction methods, the ASTRA toolbox [Aar+15]. The result of this combination is that methods that are developed using the ASTRA toolbox can be easily installed and used at synchrotrons, with minimal changes needed in user scripts. In Chapter 7, some code examples are given, explaining the various capabilities and options of the integrated toolboxes. Furthermore, results are shown for the reduction of computation time that can be achieved for iterative methods when using the optimized GPU-based methods of the ASTRA toolbox compared to using CPU-based methods. Finally, an example for an experimental dataset is given, where the combination of the preprocessing methods of TomoPy and the reconstruction methods of the ASTRA toolbox is able to improve reconstruction quality in practice.