Applied Mathematics for Modern Challenges
Vol. 7, 2026, pp. 1–21 (March 13, 2026)
doi:10.3934/ammc.2025018

DATA-DRIVEN FILTER DESIGN FOR FLEXIBLE AND NOISE-ROBUST TOMOGRAPHIC IMAGING

Hamid Fathi*[1,2], Alexander Skorikov[1] and Tristan van Leeuwen[1,2]

[1]Centrum Wiskunde & Informatica (CWI),
Science Park 123, 1098 XG, Amsterdam, The Netherlands

[2]Mathematics Institute, Utrecht University,
Budapestlaan 6, 3584 CD, Utrecht, The Netherlands

(Communicated by Jennifer Mueller)

Abstract. While filtered back-projection (FBP) is still the method of choice for fast tomographic reconstruction, its performance degrades noticeably in the presence of noise, incomplete sampling, or non-standard scan geometries. We propose a data-driven approach for learning FBP filters and projection weights directly from training data, with the goal of improving robustness without sacrificing computational efficiency. The resulting reconstructions adapt naturally to the noise level and acquisition geometry, while retaining the speed and simplicity of classical back-projection. The proposed method can be formulated as a regularized optimization problem for a linear inverse operator, which allows us to establish existence, uniqueness, and stability of the learned solution. From a spectral viewpoint, the learned filters act as data-adaptive gain functions that explicitly balance noise amplification and bias, in close analogy to a regularized pseudo-inverse. Experiments in both 2D and 3D show consistent improvements over conventional FBP and Feldkamp–Davis–Kress (FDK) in different case studies. Finally, we show that filters trained on synthetic laminography data generalize well to real-world measurements, delivering image quality comparable to advanced iterative methods without the high computational cost.

1. **Introduction.** X-ray computed tomography (CT) is widely used in industrial applications for non-destructive evaluation of objects. The goal is to reconstruct two- or three-dimensional images from X-ray measurements acquired at multiple viewing angles [14, 10, 12]. Well-known examples include medical CT and cone-beam CT. Another important example is X-ray laminography, a technique particularly well suited for imaging flat or layered specimens such as electronic components, where full rotation is often infeasible due to size or mechanical constraints.

Image reconstruction methods for CT generally fall into two categories: *algebraic approaches*, which iteratively solve a system of linear equations, and *analytical approaches*, which provide a direct solution through filtering and backprojection.

2020 *Mathematics Subject Classification.* Primary: 65R32; Secondary: 65F22, 92C55.

Key words and phrases. Computed tomography, computed laminography, learned reconstruction, filtered back-projection, filter design, data-driven learning.

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska–Curie grant agreement No. 956172 (xCTing).

*Corresponding author: Hamid Fathi.

In the algebraic approach, X-ray measurements are represented through a system of linear equations. Various iterative methods have been developed to solve the linear system [14]. A key advantage of these techniques is their flexibility in incorporating prior information or regularization terms to suppress artifacts caused by noise, incomplete data, or other acquisition imperfections [16, 2]. However, they are often computationally intensive, requiring significant time and memory resources for high-resolution reconstructions [7]. Commonly employed iterative algorithms in algebraic approaches include the Algebraic Reconstruction Technique (ART) [9], the Simultaneous Iterative Reconstruction Technique (SIRT) [8], Nesterov accelerated gradient descent (NAG) [19] and more advanced regularized iterative methods such as Total Variation (TV)-regularized reconstruction [16], which aim to improve image quality and robustness at the expense of additional computation.

In contrast, analytical approaches are attractive due to their simplicity and computational efficiency [24]. They provide a fast, non-iterative solution to the reconstruction problem [10]. A widely used example is the Feldkamp–Davis–Kress (FDK) algorithm [6], which is highly effective in conventional cone-beam CT configurations, particularly when the measurements are of high quality. However, its performance can degrade significantly in the presence of noise [31]. In FDK, standard filters such as Shepp–Logan, Hann, and Hamming use fixed window functions to attenuate high frequencies and suppress artifacts, however they are not tailored to the specific noise characteristics of the measured data [4]. Furthermore, the classical FDK algorithm is derived under the assumption of a circular source trajectory, which limits its direct applicability to more general scanning geometries. Although analytical derivations for laminography and related geometries have been presented [17, 18, 36], they are mathematically involved, highly dependent on specific acquisition setups, and often lead to strong artifacts due to the missing-cone issue.

These significant challenges with fixed, pre-defined filters have motivated a different line of research: optimizing the filter itself. This approach seeks to combine the computational efficiency of the filtered back-projection (FBP) framework with the superior image quality of iterative methods. The core idea is to move beyond fixed filters (like Shepp-Logan) and instead design filters that are better adapted to the specific geometry and noise statistics.

This concept of an optimized FBP is well-supported in the literature. A theoretical bridge between algebraic and analytical methods was presented in [22], and it was shown that any linear shift-invariant algorithm can be represented as an FBP method with some associated filter [5]. This led to methods that explicitly derive an FBP filter from an iterative algorithm, allowing FBP to approximate the results of iterative reconstruction [37, 1, 25]. Others have designed geometry-specific filters for each frequency component in tomosynthesis [20] or proposed data-dependent filtering strategies [26] to reconstruct the image in limited data or noisy scenarios better than classical FBP method. Finally, some works [3] proposed to apply a convolution kernel derived from a specific geometry to deblur the reconstructed image.

We propose a data-driven framework to optimize the filter and (when needed) projection weights in FBP, with a particular focus on 3D cone-beam circular laminography, where limited-angle sampling makes reconstruction especially challenging. The learning problem is posed as a regularized optimization of a linear inverse operator, which enables existence, uniqueness, and stability guarantees for the learned solution and yields an interpretable spectral-gain view analogous to a regularized

pseudo-inverse. We instantiate the framework with geometry-specific parameterizations (parallel-beam, fan-beam, elliptical trajectories, and 3D circular laminography) and demonstrate consistent improvements over classical FBP/FDK, including validation on experimental laminography measurements (LEGO dataset) showing generalization beyond the synthetic training setting.

The outline of the paper is as follows. In Section 2, we introduce the notation and briefly review the inverse problem formulation and fast approximate inversion via FBP. In Section 3, we present our learning framework for a regularized linear inverse operator, discuss its theoretical properties (existence, uniqueness, and stability), provide a spectral interpretation of the learned operator, and introduce geometry-informed parameterizations together with implementation aspects. In Section 4, we evaluate the proposed approach across different case studies, including 2D parallel-beam and fan-beam CT, a 2D elliptical trajectory, and 3D circular laminography on both synthetic and experimental data. In Section 5, we conclude the paper and summarize key findings.

2. Notation and preliminaries.

2.1. Notation.

Throughout, scalars and vector elements are denoted by lower-case italic letters, vectors by lower-case boldface letters, and matrices by upper-case boldface letters. The notation $\cdot^\top$ stands for the transpose of either a vector or matrix. By $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}$ we denote the ℓ_2-norm of a vector. The Frobenius norm of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is denoted by $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}$. $\operatorname{diag}(\boldsymbol{a})$ represents a diagonal matrix with elements of $\boldsymbol{a}$ on the main diagonal. $\mathbb{E}[\cdot]$ denotes the expectation. $\operatorname{Tr}(\mathbf{A})$ denotes the trace of a square matrix $\mathbf{A}$, defined as the sum of its diagonal elements. The symbol $\otimes$ represents the Kronecker product, and $\operatorname{vec}(\mathbf{A})$ is the vectorization operator that stacks the columns of a matrix $\mathbf{A}$ into a single column vector.

2.2. The inverse problem and regularization.

The X-ray tomographic acquisition process can be modeled as,

$$\mathbf{A}\mathbf{x} = \mathbf{y}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the forward operator, $\mathbf{x} \in \mathbb{R}^N$ represents the object, and $\mathbf{y} \in \mathbb{R}^M$ is the measured data. The structure of $\mathbf{A}$ reflects the geometry and ordering of the data acquisition. In a scan comprising m projection angles, each projection is measured by a detector with d pixels, yielding a total of $M = m \cdot d$ rows in $\mathbf{A}$. Each row corresponds to a single X-ray path, modeled as a line integral through the object $\mathbf{x}$ from the source to a specific detector pixel. Rows can be grouped into blocks of d consecutive rows, each block associated with a fixed projection angle.

The goal is to find an $\mathbf{x}$ that best solves this system. This inverse problem is typically ill-posed, as $\mathbf{A}$ is ill-conditioned and often there are fewer measurements than unknowns (i.e., $M \leq N$). A basic approach to stabilize the problem is through Tikhonov regularization, which approximates the solution through a regularized pseudo-inverse

$$\widehat{\mathbf{x}} = \mathbf{A}_\lambda^\dagger \mathbf{y}.$$

The operator $\mathbf{A}_\lambda^\dagger$ can be expressed in terms of the singular value decomposition (SVD) of $\mathbf{A}$. Writing

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top,$$

the regularized pseudo-inverse is given by

$$\mathbf{A}_\lambda^\dagger = \mathbf{V}_r \mathbf{\Sigma}_r \left(\mathbf{\Sigma}_r^2 + \lambda \mathbf{I}\right)^{-1} \mathbf{U}_r^\top,$$

where $r = \operatorname{rank}(\mathbf{A})$, $\mathbf{U}_r \in \mathbb{R}^{M\times r}$ contains the first r left singular vectors of $\mathbf{A}$, $\mathbf{\Sigma}_r \in \mathbb{R}^{r\times r}$ is a diagonal matrix containing the r largest singular values of $\mathbf{A}$, and $\mathbf{V}_r \in \mathbb{R}^{N\times r}$ contains the corresponding right singular vectors of $\mathbf{A}$. Furthermore $\lambda \geq 0$ is the regularization parameter. In practice, it is not feasible to explicitly compute the regularized pseudo-inverse and instead the regularized solution is approximated by iteratively solving a regularized least-squares problem [29].

2.3. **Fast approximate inversion.** For tomographic imaging, FBP offers a fast alternative to the above-mentioned reconstruction methods. They are based on the fact that for many geometries the inverse (on a continuous level) can be expressed as a back projection of filtered data [23]. Applying the filter through a fast Fourier transform then yields an efficient approximate inverse:

$$\mathbf{A}^\dagger \approx \mathbf{A}^\top \mathbf{F}^{-1} \mathbf{D} \mathbf{F},$$

where $\mathbf{F} \in \mathbb{C}^{M\times M}$ represents a discrete Fourier transform (which can be efficiently applied to vectors using the FFT) and $\mathbf{D} \in \mathbb{R}^{M\times M}$ is a diagonal matrix containing the filter weights. Note that in case $M \leq N$ and $\mathbf{A}$ has full rank we have $\mathbf{A}^\dagger = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}$. Regularization is achieved in these methods by designing a filter that cuts off high-frequency contributions. Aside from hand-crafted filters, learned filters have also been considered in the literature [32, 33].

3. A learned inverse operator for tomographic reconstruction.

3.1. **Learning a regularized pseudo-inverse.** In this section, we formalize the process of learning a filter. In contrast to existing learned filter approaches such as [32, 33], we cast this task as the learning of a regularized linear inverse mapping. Moreover, by allowing the parametrization of the learned operator to depend on the acquisition geometry, the framework naturally extends to non-standard settings. We therefore start from the following more general problem. Given $\mathbf{A} \in \mathbb{R}^{M\times N}$ with $M \leq N$, the aim is to learn the linear inverse mapping parametrized by $\mathbf{B} \in \mathbb{R}^{M\times M}$,

$$\min_{\mathbf{B} \in \mathbb{R}^{M\times M}} \mathbb{E}\|\mathbf{A}^\top \mathbf{B}\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{B}), \tag{2}$$

where we take the expectation over training pairs $(\mathbf{x}, \mathbf{y})$,$\lambda \geq 0$, and $\rho : \mathbb{R}^{M\times M} \to \mathbb{R}$ is a regularization function. The training pairs typically consist of ground truth objects $\mathbf{x}$ and corresponding noisy measurements $\mathbf{y} \approx \mathbf{A}\mathbf{x}$.

3.1.1. *General assumptions.* To ensure the well-posedness of the problem above, we make the following standing assumptions throughout the analysis.

- **(A1)** We assume that the underlying distribution of the training data has finite second moments, i.e., $\mathbb{E}\|\mathbf{x}\|_2^2 < \infty$, $\mathbb{E}\|\mathbf{y}\|_2^2 < \infty$, and the second-order moment matrices $\mathbf{\Sigma_{xx}} = \mathbb{E}\mathbf{x}\mathbf{x}^\top$ and $\mathbf{\Sigma_{yy}} = \mathbb{E}\mathbf{y}\mathbf{y}^\top$ are positive definite.
- **(A2)** The matrix $\mathbf{A}$ has full row rank, implying $\mathbf{A}\mathbf{A}^\top \succ 0$.
- **(A3)** The function ρ is μ-strongly convex.

With these assumptions in place, we can address the well-posedness of (2).

3.1.2. *Existence, uniqueness, and stability.* Under Assumptions (A1)-(A3), the objective function in (2) is strongly convex with respect to $\mathbf{B}$ and hence has a unique solution. Moreover, the optimal solution $\mathbf{B}^*$ satisfies the following matrix equation:

$$\mathbf{A}\mathbf{A}^\top\mathbf{B}\mathbf{\Sigma}_{\mathbf{yy}} + \lambda\nabla\rho(\mathbf{B}) = \mathbf{A}\mathbf{\Sigma}_{\mathbf{xy}}. \tag{3}$$

with $\mathbf{\Sigma}_{\mathbf{xy}} = \mathbb{E}\mathbf{x}\mathbf{y}^\top$.

Having established existence and uniqueness of the learned operator, we now analyze the stability of the solution with respect to perturbations in the data distribution. Let $\mathbf{B}^*_\pi$ and $\mathbf{B}^*_{\pi'}$ be the minimizers of (2) under distributions $(\mathbf{x},\mathbf{y}) \sim \pi$ and $(\mathbf{x},\mathbf{y}) \sim \pi'$, respectively. Given Assumptions (A1)-(A3), there exists a constant $C > 0$ such that:

$$\|\mathbf{B}^*_\pi - \mathbf{B}^*_{\pi'}\|_F \leq \frac{C}{\lambda\mu}\, W_2(\pi,\pi'),$$

where W_2 denotes the 2-Wasserstein distance. A proof is given in Appendix A.

3.1.3. *Interpretation as a regularized pseudo-inverse.* A more detailed interpretation of the learned operator can be obtained by considering a concrete example. In addition to assumptions (A1)-(A3), we let

- Additive Gaussian noise: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ with $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$,
- Tikhonov regularization: $\rho(\mathbf{B}) = \|\mathbf{B}\|_F^2$.

We then have

$$\mathbf{\Sigma}_{\mathbf{yy}} = \mathbf{A}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{A}^\top + \sigma^2\mathbf{I}, \qquad \mathbf{\Sigma}_{\mathbf{xy}} = \mathbf{\Sigma}_{\mathbf{xx}}\mathbf{A}^\top. \tag{4}$$

Combining (3) and (4) we then find that $\mathbf{B}^*$ satisfies the following matrix equation

$$\mathbf{A}\mathbf{A}^\top\mathbf{B}\mathbf{A}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{A}^\top + \sigma^2\mathbf{A}\mathbf{A}^\top\mathbf{B} + \lambda\mathbf{B} = \mathbf{A}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{A}^\top. \tag{5}$$

Introducing the auxiliary matrices

$$\mathbf{M} = \mathbf{A}\mathbf{A}^\top,\ \mathbf{N} = \mathbf{A}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{A}^\top, \tag{6}$$

we can re-write this as

$$(\mathbf{N}\otimes\mathbf{M} + \sigma^2\mathbf{I}\otimes\mathbf{M} + \lambda\mathbf{I})\,\mathrm{vec}(\mathbf{B}) = \mathrm{vec}(\mathbf{N}), \tag{7}$$

We can immediately see that when $\sigma = 0$, $\lambda = 0$ we learn the regular pseudo-inverse: $\mathbf{B}^* = (\mathbf{A}\mathbf{A}^\top)^{-1}$.

To further interpret the learned operator and its relation to classical analytical filters, we analyze the structure of (7) in the spectral domain. Substituting the SVD of $\mathbf{A}$ into (7) and premultiplying and postmultiplying by $\mathbf{U}^\top$ and $\mathbf{U}$, respectively, yields

$$\mathbf{\Lambda}\,\mathbf{H}\,\widetilde{\mathbf{N}} + (\sigma^2\mathbf{\Lambda} + \lambda\mathbf{I})\mathbf{H} = \widetilde{\mathbf{N}}, \tag{8}$$

where

$$\mathbf{\Lambda} = \mathrm{diag}(\sigma_1^2,\dots,\sigma_M^2), \quad \mathbf{H} = \mathbf{U}^\top\mathbf{B}\mathbf{U}, \quad \widetilde{\mathbf{N}} = \mathbf{U}^\top(\mathbf{A}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{A}^\top)\mathbf{U}.$$

To simplify the spectral analysis, we make the additional assumption that the object distribution is isotropic, i.e., $\mathbf{\Sigma}_{xx} = \mathbf{I}$. For concreteness, one may think of $\mathbf{x} \sim \mathcal{N}(0,\mathbf{I})$. Under this assumption $\widetilde{\mathbf{N}}$ is diagonal and the optimal $\mathbf{H}$ is diagonal as well. In this case, the spectral components decouple, and (8) reduces entry-wise to

$$\sigma_i^4 H_{ii} + (\sigma^2\sigma_i^2 + \lambda)H_{ii} = \sigma_i^2, \tag{9}$$

leading to the closed-form solution

$$H_{ii} = \frac{\sigma_i^2}{(\sigma_i^2 + \sigma^2)\sigma_i^2 + \lambda}. \tag{10}$$

Each H_{ii} thus acts as a spectral gain that modulates the contribution of the i-th mode according to its singular value, and noise variance. It provides an interpretable, data-adaptive filtering mechanism. Modes with small σ_i are damped by the regularization term λ, effectively suppressing noise amplification. This interpretation highlights that the learned operator $\mathbf{B}$ behaves as a *regularized pseudo inverse* in the spectral domain, adapting its frequency response based on the data and noise statistics. This expression closely resembles classical frequency-domain filters, such as the ramp filter in FBP but with an additional data-adaptive weighting that depends on the signal and noise variances.

3.1.4. *Reconstruction error.* Given a solution $\mathbf{B}^*$ to (2), and under the same assumptions as in Section 3.1.3, we can decompose the resulting reconstruction error as (see Appendix B)

$$\begin{aligned}\mathbb{E}\|\mathbf{A}^\top\mathbf{B}^*\mathbf{y}-\mathbf{x}\|_2^2 \le\ & \mathbb{E}\|\mathbf{A}^\top\mathbf{B}^*\mathbf{n}\|_2^2 \\ &+ \mathbb{E}\|(\mathbf{A}^\top\mathbf{B}^*-\mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\|_2^2 \\ &+ \mathbb{E}\|(\mathbf{A}^\dagger\mathbf{A}-\mathbf{I})\mathbf{x}\|_2^2.\end{aligned} \tag{11}$$

In the upper bound, the first term is interpreted as the variance (the influence of noise on the reconstruction) while the second term is interpreted as the bias (the influence of regularization on the learned inverse operator). The last term indicates a null-space error that we cannot get rid of, as the learned inverse operator is restricted to act within the column space of $\mathbf{A}$. In the following, we investigate the first two error terms in the spectral domain.

We begin from the first term of the upper bound in (11). Using decompositions introduced for $\mathbf{A}$ and $\mathbf{B}$ in Section 3.1.3, we can write $\mathbf{A}^\top\mathbf{B}$ as,

$$\mathbf{A}^\top\mathbf{B} = \mathbf{V}\mathbf{\Sigma}\mathbf{H}\mathbf{U}^\top = \mathbf{V}\,\mathrm{diag}\left(\frac{\sigma_i^3}{(\sigma_i^2+\sigma^2)\sigma_i^2+\lambda}\right)\mathbf{U}^\top, \tag{12}$$

where, we define the spectral variance factor as,

$$\zeta_i = \frac{\sigma_i^3}{(\sigma_i^2+\sigma^2)\sigma_i^2+\lambda}, \tag{13}$$

then, it can be written as,

$$\mathbb{E}\|\mathbf{A}^\top\mathbf{B}\mathbf{n}\|_2^2 = \sigma^2\cdot\sum_{i=1}^{M}\zeta_i^2 \tag{14}$$

Above derivation shows that the variance has a direct relation with the noise variance. Moreover, it goes to zero as the regularization parameter λ becomes very large.

The bias term of the upper bound in (11) can be investigated in the spectral domain as well. To do so, we start with:

$$\begin{aligned}(\mathbf{A}^\top\mathbf{B}-\mathbf{A}^\dagger)\mathbf{A} &= \Big(\mathbf{V}\mathbf{\Sigma}\mathbf{H}\mathbf{U}^\top - \mathbf{V}\,\mathbf{\Sigma}^\dagger\mathbf{U}^\top\Big)\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \\ &= \mathbf{V}\,\mathrm{diag}\Big(\frac{\sigma_i^4}{(\sigma_i^2+\sigma^2)\sigma_i^2+\lambda}-1\Big)\mathbf{V}^\top.\end{aligned} \tag{15}$$

Then, we define the spectral bias factor as

$$\beta_i = 1-\frac{\sigma_i^4}{(\sigma_i^2+\sigma^2)\sigma_i^2+\lambda}$$

$$= \frac{\sigma^2\sigma_i^2 + \lambda}{\sigma_i^4 + \sigma^2\sigma_i^2 + \lambda}. \tag{16}$$

To further analyze the bias term in the spectral domain, we continue under the additional simplifying assumption that $\mathbf{\Sigma}_{xx} = \mathbf{I}$ (for example, when $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I})$). Then:

$$\mathbb{E}\left\|(\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\right\|_2^2 = \sum_{i=1}^{M} \beta_i^2. \tag{17}$$

It indicates that the bias term vanishes for very large singular values. Furthermore, for very small σ_i's, it goes to 1 and introduces bias error.

Finally, we have an error-component that we cannot get rid of as the learned inverse operator is fundamentally restricted to producing solutions in the range of $\mathbf{A}^\top$ and thus cannot fill in components in the null-space.

3.2. **Parameterizing the filter for tomographic reconstruction.** For computational efficiency we want to learn $\mathbf{B}$ as a composition of filtering and weighting steps, parametrized by a set of parameters $\mathbf{p}$. The form of $\mathbf{B}(\mathbf{p})$ can be informed by the acquisition geometry, as we will see below. The choice for ρ will be discussed in more detail as well.

The well-posedness results in the previous section are proved for the unconstrained problem over the full matrix space, i.e., $\mathbf{B} \in \mathbb{R}^{M\times M}$ with $\lambda > 0$. When introducing a parametrization $\mathbf{B} = \mathbf{B}(\mathbf{p})$, we restrict the admissible operators to

$$\mathcal{S} = \{\mathbf{B}(\mathbf{p}) : \mathbf{p} \in \mathcal{P}\},$$

and consequently solve a constrained problem over $\mathcal{S}$ rather than over all of $\mathbb{R}^{M\times M}$.

Existence. Existence of a minimizer typically carries over under mild assumptions, e.g., if $\mathbf{p} \mapsto \mathbf{B}(\mathbf{p})$ is continuous and $\mathcal{P}$ is compact (or the objective is coercive in $\mathbf{p}$), then the continuous objective function attains its minimum on a compact set [28].

Uniqueness and stability. Uniqueness and stability do not automatically carry over in parameter space. Even if the minimizer is unique in operator space (i.e., the best $\mathbf{B}$ within $\mathcal{S}$ is unique), the corresponding parameter $\mathbf{p}$ may be non-unique unless the parametrization is identifiable [27]. Moreover, if $\mathbf{B}(\mathbf{p})$ depends nonlinearly on $\mathbf{p}$ (e.g., through bilinear compositions such as *weights* $\times$ *filtering*), then the objective is generally nonconvex in $\mathbf{p}$, so global uniqueness is typically lost and only local guarantees can be expected.

Independently of well-posedness, restricting to $\mathcal{S}$ introduces an approximation aspect: the parametrized solution is the best admissible operator in $\mathcal{S}$, which coincides with the unconstrained optimizer $\mathbf{B}^*$ only if $\mathbf{B}^* \in \mathcal{S}$ (or can be well-approximated by elements of $\mathcal{S}$).

In summary, the well-posedness results of Section 3.1.2 remain a useful reference at the operator level $\mathbf{B}$. In the parametrized setting, existence is typically preserved under mild conditions, whereas uniqueness and stability in the parameters depend on structural properties of the map $\mathbf{p} \mapsto \mathbf{B}(\mathbf{p})$, for instance whether it preserves convexity (e.g., affine dependence on $\mathbf{p}$) and whether it is identifiable.

Below, we specify geometry-informed parameterizations of $\mathbf{B}(\mathbf{p})$ for the acquisition settings considered in this work.

3.2.1. *2D parallel beam.* For the parallel-beam geometry with m angles and a detector with d pixels (so $M = m \cdot d$), $\mathbf{B}$ is expressed as the 1D convolution along the detector-pixel dimension with a specified filter. Specifically, we can present it as

$$\mathbf{B}(\mathbf{p}) = \mathbf{I}_m \otimes \left(\mathbf{F}_d^{-1} \operatorname{diag}(\mathbf{p}) \mathbf{F}_d\right), \tag{18}$$

where $\mathbf{I}_m$ represents the $m \times m$ identity matrix, $\mathbf{F}_d$ the $d-$dimensional DFT and $\mathbf{p} \in \mathbb{R}^d$ represent the filter coefficients. As analytically derived in the literature [10], in the ideal continuous and noise-free scenario, $\mathbf{p}$ corresponds to a high-pass ramp filter.

3.2.2. *2D circular fan-beam.* In a circular fan-beam geometry with m angles and a detector with d pixels (so $M = m \cdot d$), an additional weighting factor is needed to compensate for the non-uniform ray coverage [14]. We associate the learnable parameter $\mathbf{p}$ with a filter $\mathbf{p}_{\text{filter}} \in \mathbb{R}^d$ and weight $\mathbf{p}_{\text{weight}} \in \mathbb{R}^d$. We then have

$$\mathbf{B}(\mathbf{p}) = \mathbf{I}_m \otimes \left(\mathbf{F}_d^{-1} \operatorname{diag}(\mathbf{p}_{\text{filter}}) \mathbf{F}_d \operatorname{diag}(\mathbf{p}_{\text{weight}})\right) \tag{19}$$

3.2.3. *2D elliptical fan-beam.* This acquisition scheme, where the source and detector rotate around the object along an elliptical trajectory, is commonly used to accommodate objects with large aspect ratios. As the geometry changes per projection angle, the kernel $\mathbf{B}$ requires a parametrization for the filter and weighting vector that is slightly different from the circular fan beam case. Namely, we consider angle-dependent filters and weights and represent $\mathbf{B}$ as below,

$$\mathbf{B}(\mathbf{p}) = \left(\mathbf{I}_m \otimes \mathbf{F}_d^{-1}\right) \operatorname{diag}(\mathbf{p}_{\text{filter}}) \left(\mathbf{I}_m \otimes \mathbf{F}_d\right) \operatorname{diag}(\mathbf{p}_{\text{weight}}), \tag{20}$$

where $\mathbf{p}_{\text{filter}} \in \mathbb{R}^M$, $\mathbf{p}_{\text{weight}} \in \mathbb{R}^M$ represent an angle-dependent filter and weight.

3.2.4. *3D circular laminographic trajectory.* Laminography is an imaging technique tailored for flat, layered objects. Several acquisition trajectories are used in practice depending on sample size and hardware constraints [21]. Here we focus on the *circular* laminography trajectory, which is well suited for large, flat samples: a point source and a flat detector rotate on opposite sides of the sample with a fixed laminography tilt angle ϕ (the central ray is tilted by ϕ with respect to the sample plane); see Fig. 1. Throughout the rotation, the central ray is aligned to pass through a fixed point in the sample (typically the center of the middle layer), ensuring a consistent geometry.

In this setup, the detector data per projection angle are two-dimensional, hence we parametrize the kernel $\mathbf{B}$ using angle-dependent 2D filters and 2D weights. Let $\mathbf{F}_{d_1}$ and $\mathbf{F}_{d_2}$ denote the 1D Fourier transforms acting along the two detector dimensions. Exploiting the separability of the 2D Fourier transform,

$$\mathbf{F}_{d_1 \cdot d_2} = \mathbf{F}_{d_1} \otimes \mathbf{F}_{d_2},$$

the 2D Fourier transform acting independently on each view can be written as a Kronecker product of 1D Fourier operators.

Therefore, we parametrize $\mathbf{B}$ as

$$\mathbf{B}(\mathbf{p}) = \left(\mathbf{I}_m \otimes \mathbf{F}_{d_1}^{-1} \otimes \mathbf{F}_{d_2}^{-1}\right) \operatorname{diag}(\mathbf{p}_{\text{filter}}) \left(\mathbf{I}_m \otimes \mathbf{F}_{d_1} \otimes \mathbf{F}_{d_2}\right) \operatorname{diag}(\mathbf{p}_{\text{weight}}), \tag{21}$$

where $\mathbf{p}_{\text{filter}} \in \mathbb{R}^M$ and $\mathbf{p}_{\text{weight}} \in \mathbb{R}^M$ are vectors of concatenated 2D filter and 2D weight coefficients, with $M = m \cdot d_1 \cdot d_2$.

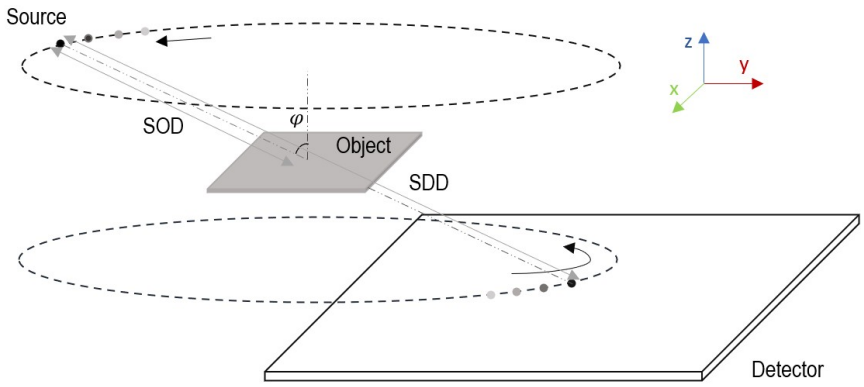


FIGURE 1. Sketch of acquisition geometry in a circular laminography setup.

3.3. **Implementation.** For given training set $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^K$ we now need to solve

$$\min_{\mathbf{p}} K^{-1} \sum_{i=1}^{K} \|\mathbf{A}^\top \mathbf{B}(\mathbf{p})\mathbf{y}_i - \mathbf{x}_i\|_2^2 + \lambda\rho(\mathbf{p}), \tag{22}$$

where $\mathbf{p} = (\mathbf{p}_{\text{filter}}, \mathbf{p}_{\text{weight}})$ contains the parameters for the filter and weight. In contrast to (2), the parametrized formulation is, in general, no longer strictly convex. While the objective is strictly convex in the operator $\mathbf{B}$, substituting $\mathbf{B} = \mathbf{B}(\mathbf{p})$ changes the geometry of the problem: if $\mathbf{B}(\mathbf{p})$ depends non-affinely on $\mathbf{p}$, i.e., $\mathbf{B}$ is formed by composition or products of parameters, then the resulting objective becomes nonconvex in $\mathbf{p}$. Consequently, global uniqueness in parameter space is not guaranteed and optimization typically converges to a first-order critical point of the objective. The resulting learned operator therefore corresponds to a (possibly local) solution within the admissible family $\mathcal{S} = \{\mathbf{B}(\mathbf{p})\}$.

3.3.1. *Regularization.* We generally want to enforce smoothness on both the filter and spatial weights, which we achieve by setting

$$\rho(\mathbf{p}) = \|\mathbf{L}_{\text{freq}}\mathbf{p}_{\text{filter}}\|_2^2 + \|\mathbf{L}_{\text{space}}\mathbf{p}_{\text{weight}}\|_2^2,$$

with $\mathbf{L}_{\text{freq}}, \mathbf{L}_{\text{space}}$ representing a finite-difference approximation of the derivative or gradient along the frequency and spatial axes. In addition, we enforce symmetry of the filter, which guarantees that its Fourier transform is real-valued.

3.3.2. *Implementation details.* The forward operator $\mathbf{A}$ is implemented using Tomosipo [11], which provides a PyTorch-compatible interface to the ASTRA toolbox [34]. During training, we use PyTorch's automatic differentiation to backpropagate through the operator $\mathbf{B}(\mathbf{p})$. The entire pipeline, from the frequency-domain parameters to the final reconstruction error, is differentiable and is optimized with Adam to update the filter coefficients and spatial weights simultaneously [15].

4. **Case studies.** In this section, we investigate the data-driven filter design method in several practical scenarios. We begin with noisy measurements in 2D

parallel-beam and fan-beam geometries. We then learn filters and weights for a 2D elliptical trajectory and for 3D circular laminography.

4.1. **2D parallel-beam geometry.** We focus on learning a filter for noisy measurements in parallel-beam geometry. Specifically, we investigate how the filter in (18) can be adapted to noisy scenarios by making use of training data. To this end, we generate synthetic 2D phantoms of size 400×400 consisting of circles with random sizes and positions with binary values. The simulated acquisition consists of 360 uniformly spaced projections with a resolution of 512 pixels. The pixel size is set to 0.002 for both the phantom and the detector. To ensure stability during training, we include a small smoothness regularization with the coefficient of 0.001 to suppress abrupt variations in the filter.

Figure 2 shows the filters learned from datasets corrupted with different levels of additive Gaussian noise, along with the standard Ram-Lak filter, which is typically used in FBP. While the Ram-Lak filter exhibits a linear increase in amplitude with frequency, the learned filters follow a similar trend at low frequencies but begin to deviate and attenuate at higher frequencies across all noise scenarios, with the inflection point shifting toward lower frequencies for higher noise levels. This behavior suggests that the learned filters adapt to the noise characteristics of the training data, effectively acting as modified high-pass filters that suppress noise-induced artifacts while the unmodified Ram-Lak filter might otherwise amplify. Notably, the same behavior with respect to noise is commonly observed or enforced in other adaptive and fixed filter types [26, 30].

We evaluate the performance of the learned filters on separate validation datasets with the same noise levels. Table 1 presents the MSE and SSIM results across various noise conditions. From these results, we observe that the standard FBP method becomes less effective at higher noise levels, while at lower noise levels its performance becomes comparable with the learned method. Moreover, the MSE results in Table 1 are the best when the validation noise level matches the training noise level, although this trend does not hold consistently for SSIM. In particular, SSIM may favor reconstructions that are smoother and contain fewer high-frequency fluctuations, whereas MSE penalizes bias and oversmoothing more directly. Consequently, a model trained at a different noise level can achieve higher SSIM on a given test set if it suppresses high-frequency artifacts even when its MSE is not optimal.

TABLE 1. Learned reconstruction performance for different training/validation SNR levels in parallel-beam geometry (T/V: train SNR / validation SNR). Entries are reported as **MSE** (mean ± std) $\times 10^{-3}$ and **SSIM** (mean ± std).

T/V	**20**	**25**	**30**
20	5.3 ± 0.6 / 0.305 ± 0.034	4.3 ± 0.4 / 0.499 ± 0.035	4.0 ± 0.3 / 0.646 ± 0.031
25	6.2 ± 1.0 / 0.197 ± 0.023	3.9 ± 0.4 / 0.369 ± 0.037	3.2 ± 0.3 / 0.564 ± 0.039
30	10.1 ± 2.1 / 0.132 ± 0.0098	4.6 ± 0.7 / 0.247 ± 0.026	2.9 ± 0.3 / 0.444 ± 0.043
FBP	39 ± 9 / 0.080 ± 0.001	12 ± 2 / 0.120 ± 0.005	4 ± 1 / 0.220 ± 0.022

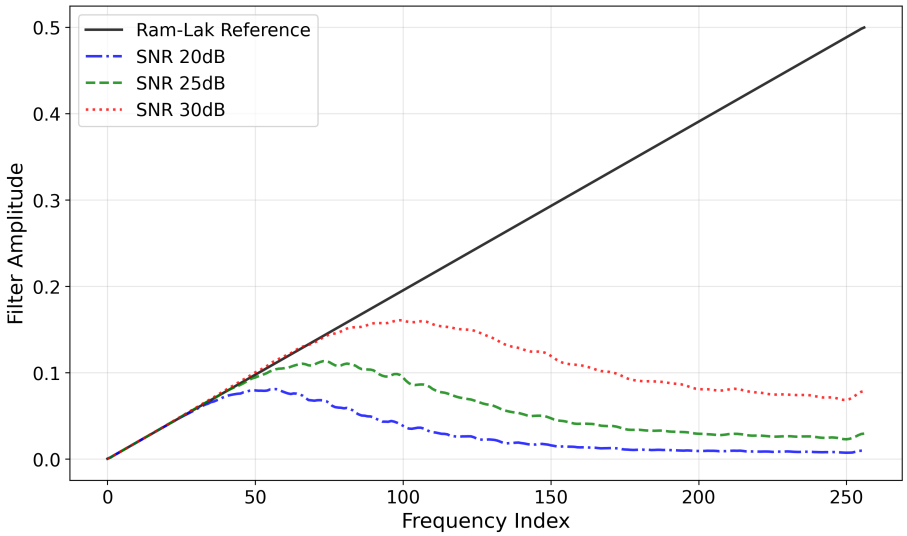
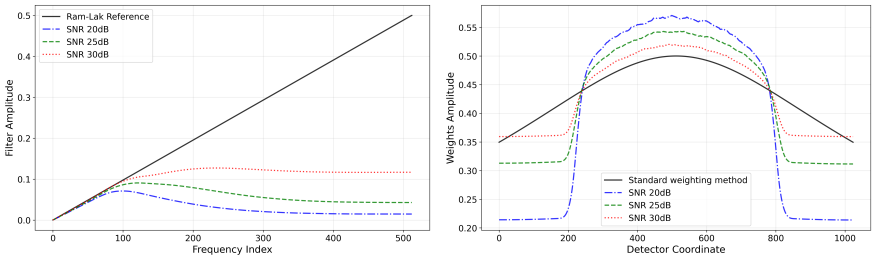


FIGURE 2. Learned filters for different noisy datasets in a parallel beam geometry versus Ram-Lak filter shown in the frequency domain.

4.2. **2D fan-beam geometry.** This section extends the learning of parameters (filter and weights) for the FBP approach to fan-beam geometry under noisy conditions. To learn the required FBP parameters using clean training data, it is sufficient to incorporate equation (19) into the optimization problem of (22). The training and validation datasets consist of synthetic 2D phantoms of size 400×400 containing circles with random sizes and positions with pixel size of 0.005. The simulated acquisition includes 360 projections with a resolution of 1024 pixels and a pixel size of 0.01, while the source-to-object distance (SOD) and source-to-detector distance (SDD) are set to 2.5 and 5, respectively. To ensure stable training, a small smoothness regularization term with a coefficient of 0.0001 is added to independently suppress abrupt variations in both the filter and the weight vector.

Figure 3 illustrates the learned filter and fan-beam weighting for different noise levels, together with the standard choices used in conventional FBP. As in the parallel-beam case, the learned filters deviate from the Ram-Lak at higher frequencies, with stronger attenuation at lower SNR, reflecting noise-adaptive spectral shaping. Moreover, the learned weights mimic the standard weighting method with noise-dependent deviations that are largest at low SNR. Since the fan-beam parametrization is factored into a filter and weights, the representation is subject to a global scale ambiguity (a rescaling of the filter can be compensated by an inverse rescaling of the weights). Following common practice in bilinear/factored models, we fix this gauge by normalizing the learned weights to the standard weighting profile and rescaling the learned filter accordingly [13].

Quantitative results for the fan-beam case, presented in Table 2, further confirm the advantage of the learned filters over FBP method. While FBP performs poorly across all noise levels, the learned models consistently achieve lower MSE and higher SSIM values. As in the parallel-beam experiments, the best MSE results occur when



(A) (B)

FIGURE 3. Learned fan-beam (A) filter and (B) weights for different noise levels (SNR 20/25/30 dB), compared to the Ram-Lak filter and the standard fan-beam weighting used in conventional FBP. The weights are normalized to the standard profile to remove the filter–weight scaling ambiguity; the filter is rescaled accordingly.

the training and validation noise levels are matched. However, this pattern is not always reflected in SSIM, similarly to the parallel beam case.

TABLE 2. Learned reconstruction performance for different training/validation SNR levels in fan-beam geometry (T/V: train SNR / validation SNR). Entries are reported as **MSE** (mean ± std) $\times 10^{-3}$ and **SSIM** (mean ± std).

T/V	**20**	**25**	**30**
20	5.0 ± 0.5 / 0.319 ± 0.028	4.1 ± 0.4 / 0.504 ± 0.041	3.8 ± 0.3 / 0.629 ± 0.027
25	6.1 ± 0.8 / 0.217 ± 0.019	4.0 ± 0.5 / 0.381 ± 0.040	3.4 ± 0.3 / 0.525 ± 0.030
30	9.9 ± 1.7 / 0.147 ± 0.009	4.9 ± 0.8 / 0.262 ± 0.031	3.4 ± 0.4 / 0.401 ± 0.031
FBP	38.5 ± 7.7 / 0.090 ± 0.001	13.8 ± 3.2 / 0.133 ± 0.008	6.3 ± 1.1 / 0.205 ± 0.016

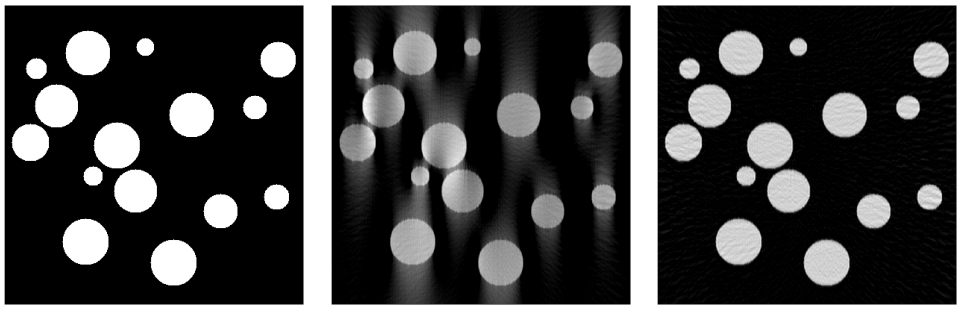
4.3. **2D elliptical trajectory with fan beam.** In this case, we learn filter and weights for FBP in a 2D elliptical trajectory. It requires incorporating (20) into (22) and minimizing over the filter and weights. The training dataset includes synthetic 2D phantoms of size 400×400 with random circles in terms of size and position. The measurements are acquired from 360 projections over a full rotation, each consisting of 512 pixels with a pixel size of 0.015. The source and detector rotate around the object along an elliptical trajectory, where the source-to-object distance (SOD) and source-to-detector distance (SDD) are 4.5 and 9, respectively, along the semi-minor axis, and 12 and 24 along the semi-major axis.

Table 3 compares the performance of the learned reconstruction method with the FBP algorithm, which is not inherently designed for elliptical trajectories. The quantitative metrics show that the learned reconstruction method significantly outperforms FBP in both MSE and SSIM, as the standard algorithm struggles to handle the non-static acquisition geometry. The qualitative results in figure 4 further highlight this improvement: while the standard FBP reconstruction (middle) exhibits significant artifacts, the learned approach (right) effectively compensates

TABLE 3. Reconstruction quality of the learned method against standard FBP for elliptical trajectory on the test set (mean ± std).

Method	MSE	SSIM
Learned reconstruction	0.0032 ± 0.0002	0.404 ± 0.0127
FBP	0.0535 ± 0.0073	0.122 ± 0.0098

for the trajectory's geometry to produce a result nearly identical to the ground truth (left).



Ground Truth Conventional FBP Learned Method

FIGURE 4. Qualitative comparison showing reconstructions of a 2D phantom for the ground truth, FBP, and the learned method, respectively.

4.4. **3D circular laminography: synthetic data.** In the following, we optimize the filter and weights for FBP in a 3D circular laminography setup, illustrated in figure 1. This is achieved using the parameterization introduced in (21), incorporated into the optimization problem defined in (22). The training data consist of synthetic 3D flat phantoms of size $40 \times 400 \times 400$ with three distinct layers. Each layer contains circles with randomly sampled positions and sizes. The acquisition comprises 100 projections over a full 360° rotation, each with a resolution of 512×512 pixels and the pixel size is set to 0.02. The laminography angle is $\phi = 45°$, with a SOD of 5 and a SDD of 20.

Table 4 evaluates the reconstruction performance of the learned method compared to the conventional FDK algorithm quantitatively. It shows that the learned approach consistently outperforms FDK across the validation dataset, achieving significantly better MSE and SSIM distributions. These results are confirmed by the visual comparisons of a test phantom in figure 5. While the conventional FDK reconstructions (middle column) exhibit severe artifacts in both lateral and coronal views, the learned reconstructions (right column) demonstrate substantial improvements in structural clarity, closely matching the ground truth (left column).

Figure 6 evaluates the generalizability and robustness of the learned reconstruction parameters, which were trained at a 45° laminography angle (ϕ) and subsequently applied to measurements from different angles. The quantitative plots (top row) show that while the SSIM slightly declines as the acquisition angle deviates from the training setup, the MSE improves (decreases) at higher angles due to the

TABLE 4. Evaluation of the learned method for cone-beam circular laminography against FDK results. It shows mean ± standard deviation.

Method/Metric	MSE	SSIM
Learned reconstruction	0.094 ± 0.002	0.246 ± 0.004
FDK	0.160 ± 0.004	0.059 ± 0.003

enhanced depth information. The qualitative results (middle and bottom rows) visualize these effects through central slices in the lateral and coronal planes for $40°, 45°$, and $50°$, demonstrating stable performance across the tested range.

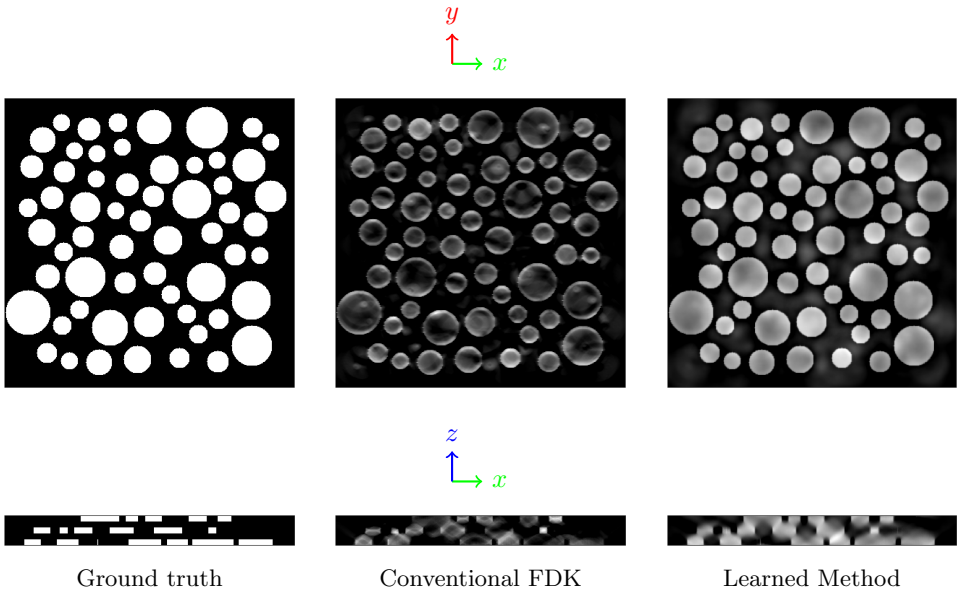


FIGURE 5. Qualitative comparison of lateral and coronal slices in learned reconstruction (right column), respectively, against Ground Truth (left column) and conventional FDK (middle column).

4.5. **3D circular laminography: real data.** To further validate the proposed method, we test it on an experimental cone-beam laminographic scan of a LEGO brick acquired by Waygate Technologies. The detector has a pixel size of 100μm and a native resolution of 3000×3000 pixels, downscaled by a factor of three to reduce computational cost. The acquisition comprises 72 projections with imaging geometry defined by $SOD = 360$, $SDD = 1300$, and $\phi = 17.1°$.

The model was first trained on the previously described synthetic dataset containing circle phantoms of size $125 \times 1000 \times 1000$ voxels while using experimental geometry. Then, learned filters and weights were applied to the LEGO brick dataset to evaluate generalization. For qualitative comparison, we present representative slices of the reconstructed volume alongside results obtained with conventional FDK

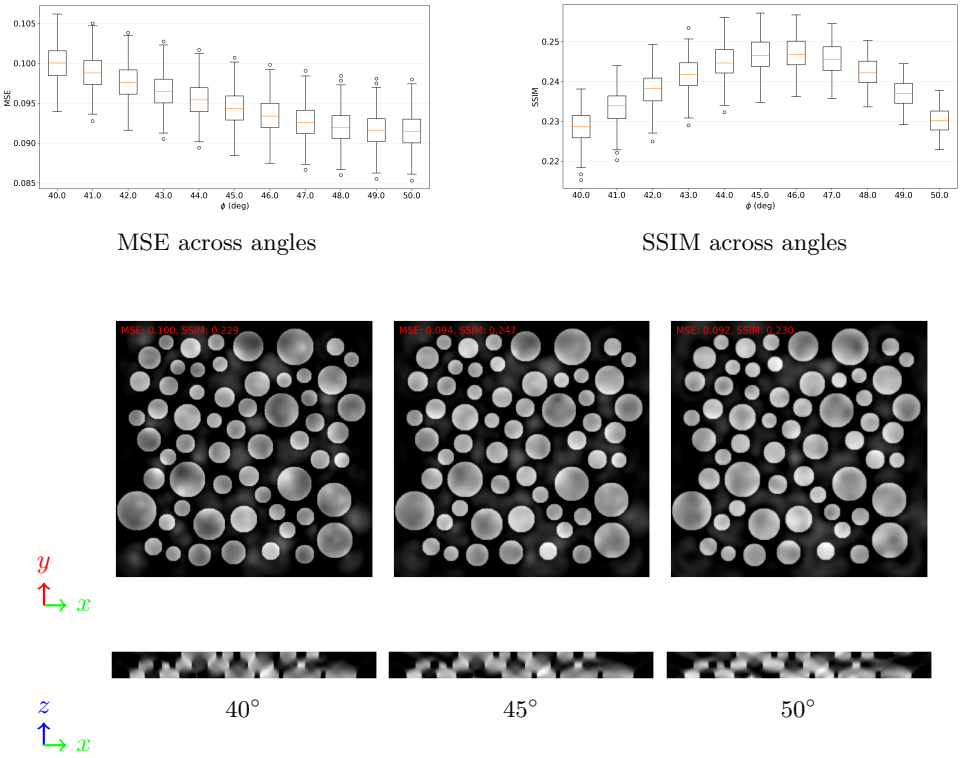


FIGURE 6. Generalizability of the learned parameters across varying laminography angles. **Top row:** Quantitative assessment via MSE and SSIM boxplots for variations up to ±5°. **Middle and bottom rows:** Qualitative lateral and coronal cross-sections, respectively, showing reconstructions at 40°, 45°, and 50°.

and Nesterov-accelerated gradient descent (NAG), which is an iterative reconstruction method [19].

Figure 7 shows axial views of the top layer reconstructed with the three methods. Treating the NAG results as a reference, it can be observed that conventional FDK fails to recover several structures, whereas the learned reconstruction preserves those features, closely matching the NAG solution. A similar observation is made for the bottom layer, shown in figure 8. Coronal views of the central slice are presented in figure 9, where FDK performs poorly compared to both the learned and iterative reconstructions. Finally, the sagittal view in figure 10 indicates that although FDK appears more comparable in this orientation, the learned reconstruction still demonstrates clear suppression of streak artifacts and achieves image quality on par with NAG.

Importantly, this result is obtained without iterative refinement at the reconstruction step, resulting in a comparable computational cost to the conventional FDK. The learned reconstruction not only outperforms analytical FDK on simulated phantoms but also generalizes effectively to experimental data, producing reconstructions comparable to iterative algorithms while retaining the computational efficiency of a direct method.

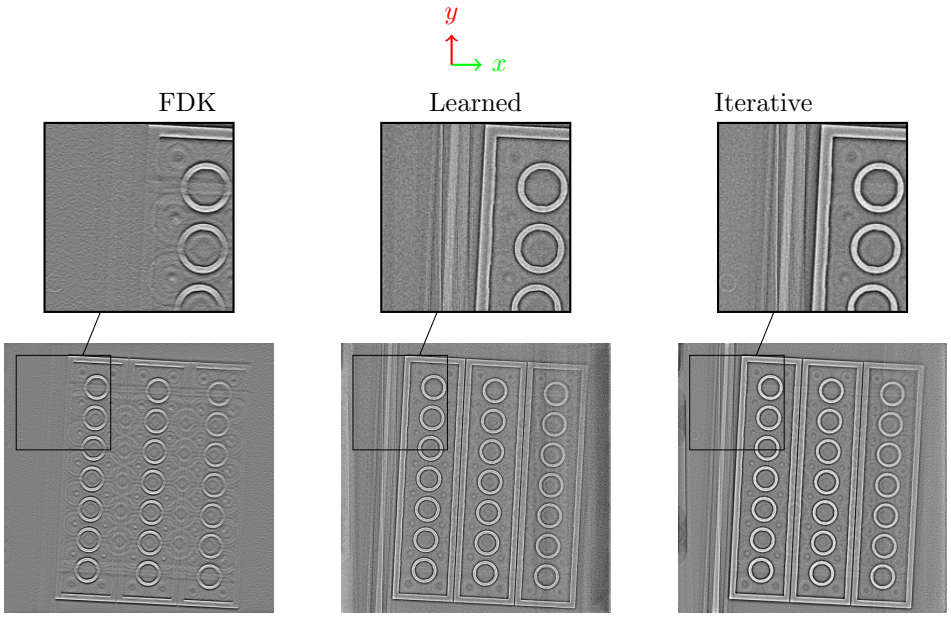


FIGURE 7. Comparison of the top axial slice through the reconstructions of experimental LEGO brick dataset. The coordinate system (top) applies to all images, with x (red), y (green), and z (blue) axes. Each method shows the same ROI zoomed above the image.

5. **Conclusion.** This paper introduces a data-driven framework to optimize filters and weighting factors for the FBP method, specifically targeting the limitations of traditional analytical techniques in challenging imaging conditions, such as with non-conventional geometries, significant noise and undersampling. By formulating a regularized optimization problem, the approach learns noise and geometry-adapted parameters directly from training data to address the ill-posedness of the reconstruction. Unlike the fixed filters used in conventional FBP and FDK, the learned filters act as spectral gains that adaptively attenuate high frequencies to suppress noise-induced artifacts while preserving structural details. The proposed method demonstrates significant performance gains in both 2D and 3D scenarios, including parallel-beam, fan-beam, and complex non-circular trajectories like elliptical paths and circular laminography. Quantitative results show that these learned angle-dependent filters and weights consistently lead to lower reconstruction errors and higher structural similarity compared to classical FBP and FDK methods. The framework enables the effective handling of incomplete sampling and acquisition geometries that deviate from standard CT configurations, which typically lead to strong artifacts in analytical reconstructions. Furthermore, the approach proves its practical utility by successfully generalizing from synthetic training data to real-world experimental measurements, such as the LEGO brick dataset. Ultimately, this framework offers a flexible and robust alternative that retains the real-time efficiency of direct back-projection while achieving the superior image quality associated with advanced iterative methods.

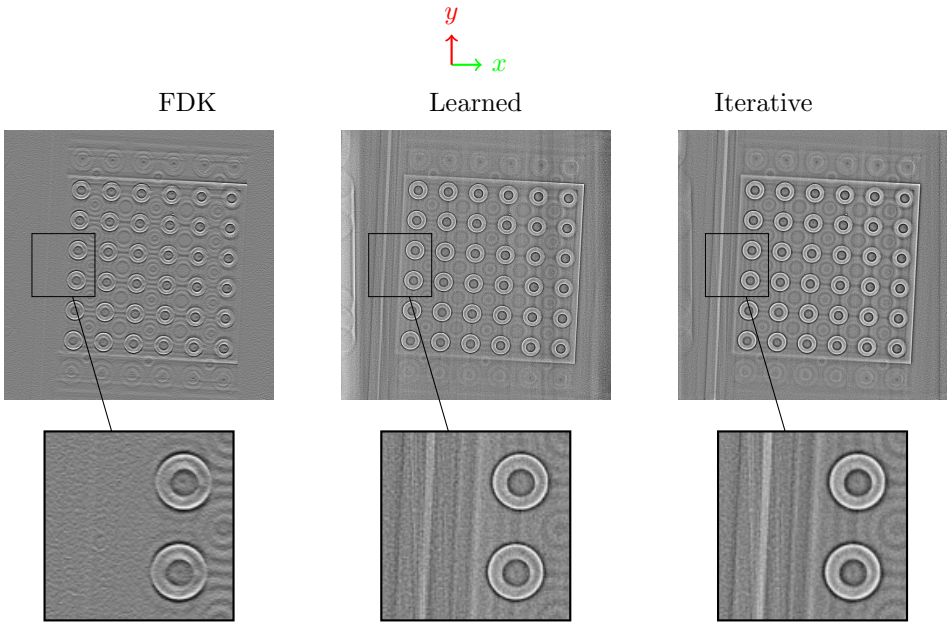


FIGURE 8. Comparison of the bottom axial slice through the reconstructions of experimental LEGO brick dataset. Each method (FDK, learned, iterative) shows the same ROI zoomed above the image.

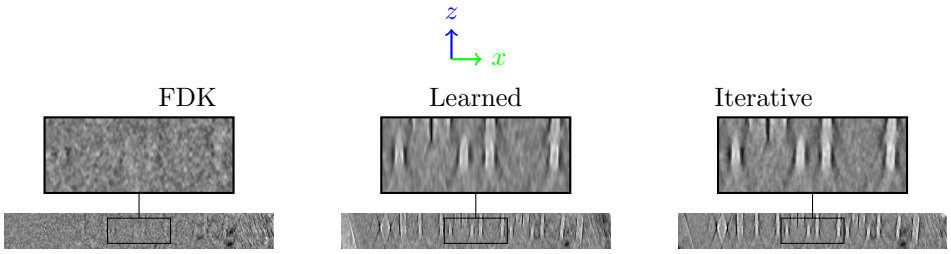


FIGURE 9. Comparison of the central coronal slice through the reconstructions of experimental LEGO brick dataset. Each method (FDK, learned, iterative) shows the same ROI zoomed above the image.

Acknowledgments. The authors thank Dr. Alexander Suppes (Waygate Technologies) for acquiring the LEGO brick laminography dataset.

Appendix A. **Stability.**

Proof. Let π and π' be two joint distributions of $(\mathbf{x}, \mathbf{y})$ satisfying Assumptions (A1)–(A3), and set

$$\mathbf{\Sigma_{yy}}(\pi) = \mathbb{E}_\pi[\mathbf{y}\mathbf{y}^\top], \qquad \mathbf{\Sigma_{xy}}(\pi) = \mathbb{E}_\pi[\mathbf{x}\mathbf{y}^\top],$$

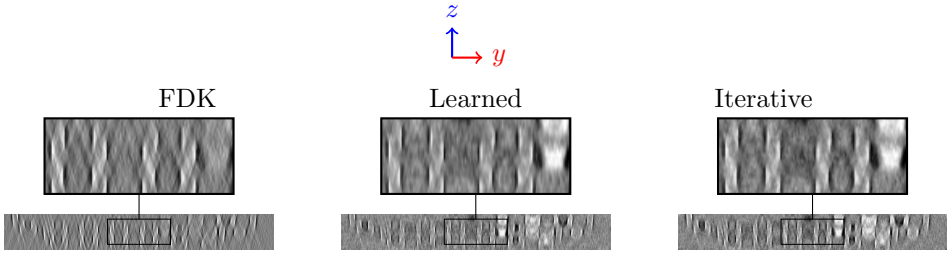


FIGURE 10. Comparison of the central sagittal slice through the reconstructions of experimental LEGO brick dataset. Each method (FDK, learned, iterative) shows the same ROI zoomed above the image.

with analogous definitions for π'. Denote the objective in (2) by $F_\pi(\mathbf{B})$. Its gradient reads

$$\nabla F_\pi(\mathbf{B}) = 2\,\mathbf{A}\mathbf{A}^\top \mathbf{B}\,\mathbf{\Sigma_{yy}}(\pi) - 2\,\mathbf{A}\,\mathbf{\Sigma_{xy}}(\pi) + \lambda \nabla \rho(\mathbf{B}). \tag{23}$$

Let $\mathbf{B}^*_\pi$ and $\mathbf{B}^*_{\pi'}$ be minimizers of F_π and $F_{\pi'}$, hence $\nabla F_\pi(\mathbf{B}^*_\pi) = \nabla F_{\pi'}(\mathbf{B}^*_{\pi'}) = \mathbf{0}$.

Step 1 (strong convexity). Under (A1)–(A2) the data term is strongly convex in $\mathbf{B}$, and since ρ is μ-strongly convex, F_π is $(\gamma + \lambda\mu)$-strongly convex for some $\gamma > 0$ [19]. Therefore, for any $\mathbf{B}, \mathbf{B}'$,

$$(\gamma + \lambda\mu)\,\|\mathbf{B} - \mathbf{B}'\|_F^2 \le \big\langle \nabla F_\pi(\mathbf{B}) - \nabla F_\pi(\mathbf{B}'),\, \mathbf{B} - \mathbf{B}' \big\rangle.$$

With $\mathbf{B} = \mathbf{B}^*_\pi$ and $\mathbf{B}' = \mathbf{B}^*_{\pi'}$ and using $\nabla F_\pi(\mathbf{B}^*_\pi) = \mathbf{0}$, we obtain

$$(\gamma + \lambda\mu)\,\|\mathbf{B}^*_\pi - \mathbf{B}^*_{\pi'}\|_F^2 \le -\big\langle \nabla F_\pi(\mathbf{B}^*_{\pi'}),\, \mathbf{B}^*_\pi - \mathbf{B}^*_{\pi'} \big\rangle. \tag{24}$$

Step 2 (difference of gradients). Using (23) and $\nabla F_{\pi'}(\mathbf{B}^*_{\pi'}) = \mathbf{0}$ gives

$$\nabla F_\pi(\mathbf{B}^*_{\pi'}) = 2\,\mathbf{A}\mathbf{A}^\top \mathbf{B}^*_{\pi'}(\mathbf{\Sigma_{yy}}(\pi) - \mathbf{\Sigma_{yy}}(\pi')) - 2\,\mathbf{A}(\mathbf{\Sigma_{xy}}(\pi) - \mathbf{\Sigma_{xy}}(\pi'))\,.$$

Hence,

$$\begin{aligned}\|\nabla F_\pi(\mathbf{B}^*_{\pi'})\|_F \le{}& 2\|\mathbf{A}\mathbf{A}^\top\|_F\,\|\mathbf{B}^*_{\pi'}\|_F\,\|\mathbf{\Sigma_{yy}}(\pi) - \mathbf{\Sigma_{yy}}(\pi')\|_F \\ &+ 2\|\mathbf{A}\|_F\,\|\mathbf{\Sigma_{xy}}(\pi) - \mathbf{\Sigma_{xy}}(\pi')\|_F.\end{aligned} \tag{25}$$

Step 3 (moment bounds via W_2). Using the coupling characterization of W_2 [35] together with the uniform moment bounds in (A1), there exists $L > 0$ such that

$$\|\mathbf{\Sigma_{yy}}(\pi) - \mathbf{\Sigma_{yy}}(\pi')\|_F \le L\,W_2(\pi,\pi'), \qquad \|\mathbf{\Sigma_{xy}}(\pi) - \mathbf{\Sigma_{xy}}(\pi')\|_F \le L\,W_2(\pi,\pi').$$

Moreover, by strong convexity and the uniform moment bounds, $\|\mathbf{B}^*_{\pi'}\|_F$ is uniformly bounded by some $C_0 > 0$. Substituting these bounds into (25) yields

$$\|\nabla F_\pi(\mathbf{B}^*_{\pi'})\|_F \le 2L\big(\|\mathbf{A}\mathbf{A}^\top\|_F\,C_0 + \|\mathbf{A}\|_F\big)\,W_2(\pi,\pi') =: C_1\,W_2(\pi,\pi').$$

Step 4 (conclusion). From (24) and Cauchy–Schwarz,

$$(\gamma + \lambda\mu)\,\|\mathbf{B}^*_\pi - \mathbf{B}^*_{\pi'}\|_F^2 \le \|\nabla F_\pi(\mathbf{B}^*_{\pi'})\|_F\,\|\mathbf{B}^*_\pi - \mathbf{B}^*_{\pi'}\|_F,$$

and therefore

$$\|\mathbf{B}^*_\pi - \mathbf{B}^*_{\pi'}\|_F \le \frac{\|\nabla F_\pi(\mathbf{B}^*_{\pi'})\|_F}{\gamma + \lambda\mu} \le \frac{C_1}{\gamma + \lambda\mu}\,W_2(\pi,\pi').$$

Since $\gamma > 0$ is fixed (for given $\mathbf{A}$ and the moment bounds), we may absorb it into the constant and write

$$\|\mathbf{B}^*_\pi - \mathbf{B}^*_{\pi'}\|_F \le \frac{C}{\lambda\mu}\, W_2(\pi, \pi'),$$

for some $C > 0$ independent of λ. □

Appendix B. Reconstruction error bound.

Proof. Let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where the noise satisfies $\mathbb{E}[\mathbf{n} \mid \mathbf{x}] = \mathbf{0}$. Assume that $\mathbf{A}$ has full row rank (**A2**), so that the Moore–Penrose pseudoinverse $\mathbf{A}^\dagger = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}$ is well defined and that $\mathbf{A}^\dagger\mathbf{A}$ is the orthogonal projector onto the range of $\mathbf{A}^\top$.

The reconstruction error reads

$$\mathbf{A}^\top\mathbf{B}\mathbf{y} - \mathbf{x} = (\mathbf{A}^\top\mathbf{B}\mathbf{A} - \mathbf{I})\mathbf{x} + \mathbf{A}^\top\mathbf{B}\mathbf{n}.$$

Adding and subtracting $\mathbf{A}^\dagger\mathbf{A}\mathbf{x}$ yields

$$\mathbf{A}^\top\mathbf{B}\mathbf{y} - \mathbf{x} = \mathbf{A}^\top\mathbf{B}\mathbf{n} + (\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x} + (\mathbf{A}^\dagger\mathbf{A} - \mathbf{I})\mathbf{x}.$$

Since both $\mathbf{A}^\top\mathbf{B}\mathbf{n}$ and $(\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}$ belong to the range of $\mathbf{A}^\top$, while $(\mathbf{A}^\dagger\mathbf{A} - \mathbf{I})\mathbf{x} = -(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{x}$ belongs to the null-space of $\mathbf{A}$, the third term is orthogonal to the sum of first two and the Pythagorean theorem gives

$$\|\mathbf{A}^\top\mathbf{B}\mathbf{y} - \mathbf{x}\|_2^2 = \|\mathbf{A}^\top\mathbf{B}\mathbf{n} + (\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\|_2^2 + \|(\mathbf{A}^\dagger\mathbf{A} - \mathbf{I})\mathbf{x}\|_2^2.$$

Expanding the first term and taking expectation,

$$\begin{aligned}\mathbb{E}\|\mathbf{A}^\top\mathbf{B}\mathbf{y} - \mathbf{x}\|_2^2 = {}& \mathbb{E}\|\mathbf{A}^\top\mathbf{B}\mathbf{n}\|_2^2 + \mathbb{E}\|(\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\|_2^2 \\ & + 2\,\mathbb{E}\big[\langle \mathbf{A}^\top\mathbf{B}\mathbf{n}, (\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\rangle\big] + \mathbb{E}\|(\mathbf{A}^\dagger\mathbf{A} - \mathbf{I})\mathbf{x}\|_2^2.\end{aligned}$$

Using conditional expectation and $\mathbb{E}[\mathbf{n} \mid \mathbf{x}] = 0$, the cross term vanishes:

$$\mathbb{E}\big[\langle \mathbf{A}^\top\mathbf{B}\mathbf{n}, (\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\rangle\big] = \mathbb{E}_{\mathbf{x}}\big[\langle \mathbf{A}^\top\mathbf{B}\mathbb{E}[\mathbf{n} \mid \mathbf{x}], (\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\rangle\big] = 0.$$

Consequently, under these assumptions one obtains the identity

$$\mathbb{E}\|\mathbf{A}^\top\mathbf{B}\mathbf{y} - \mathbf{x}\|_2^2 = \mathbb{E}\|\mathbf{A}^\top\mathbf{B}(\mathbf{y} - \mathbf{A}\mathbf{x})\|_2^2 + \mathbb{E}\|(\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\|_2^2 + \mathbb{E}\|(\mathbf{A}^\dagger\mathbf{A} - \mathbf{I})\mathbf{x}\|_2^2.$$

If the noise is not assumed to be centered or independent of $\mathbf{x}$, the mixed term above cannot be guaranteed to vanish; in that case the same derivation yields the inequality

$$\mathcal{E}(\mathbf{B}) \;\le\; \mathbb{E}\|\mathbf{A}^\top\mathbf{B}(\mathbf{y} - \mathbf{A}\mathbf{x})\|_2^2 + \mathbb{E}\|(\mathbf{A}^\top\mathbf{B} - \mathbf{A}^\dagger)\mathbf{A}\mathbf{x}\|_2^2 + \mathbb{E}\|(\mathbf{A}^\dagger\mathbf{A} - \mathbf{I})\mathbf{x}\|_2^2,$$

which is the stated bound. □

REFERENCES

[1] K. J. Batenburg and L. Plantagie, Fast approximation of algebraic reconstruction methods for tomography, *IEEE Transactions on Image Processing*, **21** (2012), 3648-3658.

[2] B. Bilgic, I. Chatnuntawech, A. P. Fan, K. Setsompop, S. F. Cauley, L. L. Wald and E. Adalsteinsson, Fast image reconstruction with l2-regularization, *Journal of Magnetic Resonance Imaging*, **40** (2014), 181-191.

[3] T. Blumensath, Backprojection inverse filtration for laminographic reconstruction, *IET Image Processing*, **12** (2018), 1541-1549.

[4] T. Buzug, *Computed Tomography: From Photon Statistics To Modern Cone-Beam CT*, Springer Berlin Heidelberg, 2014.

[5] R. Clack, Towards a complete description of three-dimensional filtered backprojection, *Physics in Medicine Biology*, **37** (1992), 645.

[6] L. A. Feldkamp, L. C. Davis and J. W. Kress, Practical cone-beam algorithm, *Journal of the Optical Society of America A*, **1** (1984), 612-619.

[7] L. L. Geyer, U. J. Schoepf, F. G. Meinel, J. W. Nance Jr, G. Bastarrika, J. A. Leipsic, N. S. Paul, M. Rengo, A. Laghi and C. N. De Cecco, State of the art: Iterative CT reconstruction techniques, *Radiology*, **276** (2015), 339-357.
[8] P. Gilbert, Iterative methods for the three-dimensional reconstruction of an object from projections, *Journal of Theoretical Biology*, **36** (1972), 105-117.
[9] R. Gordon, R. Bender and G. T. Herman, Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and X-ray photography, *Journal of Theoretical Biology*, **29** (1970), 471-481.
[10] P. C. Hansen, J. S. Jørgensen and W. R. B. Lionheart (eds.), *Computed Tomography: Algorithms, Insight, and Just Enough Theory*, Society for Industrial and Applied Mathematics, 2021.
[11] A. A. Hendriksen, D. Schut, W. J. Palenstijn, N. Viganó, J. Kim, D. M. Pelt, T. Van Leeuwen and K. Joost Batenburg, Tomosipo: fast, flexible, and convenient 3d tomography for complex scanning geometries in python, *Optics Express*, **29** (2021), 40494-40513.
[12] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2^{nd} edition, Springer Publishing Company, Incorporated, 2009.
[13] M. Jin, S. Roth and P. Favaro, Normalized blind deconvolution, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 668-684.
[14] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Society for Industrial and Applied Mathematics, 2001.
[15] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *CoRR*, **abs/1412.6980**.
[16] H. Kudo, F. Yamazaki, T. Nemoto and K. Takaki, A very fast iterative algorithm for TV-regularized image reconstruction with applications to low-dose and few-view CT, in *Developments in X-Ray Tomography X*, **9967** (2016), 31-46.
[17] G. Lauritsch and W. H. Härer, Theoretical framework for filtered back projection in tomosynthesis, in *Medical Imaging 1998: Image Processing*, **3338** (1998), 1127-1137.
[18] A. Myagotin, A. Voropaev, L. Helfen, D. Hänschke and T. Baumbach, Efficient volume reconstruction for parallel-beam computed laminography by filtered backprojection on multi-core clusters, *IEEE Transactions on Image Processing*, **22** (2013), 5348-5361.
[19] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, 1^{st} edition, Springer Publishing Company, Incorporated, 2014.
[20] T. Nielsen, S. Hitziger, M. Grass and A. Iske, Filter calculation for x-ray tomosynthesis reconstruction, *Physics in Medicine & Biology*, **57** (2012), 3915.
[21] N. S. O'Brien, R. P. Boardman, I. Sinclair and T. Blumensath, Recent advances in X-ray cone-beam computed laminography, *Journal of X-ray Science and Technology*, **24** (2016), 691-707.
[22] J. K. Older and P. C. Johns, Matrix formulation of computed tomogram reconstruction, *Physics in Medicine Biology*, **38** (1993), 1051.
[23] V. Palamodov, *Reconstruction from Integral Data*, Chapman & Hall/CRC Monographs and Research Notes in Mathematics, CRC Press, 2016.
[24] X. Pan, E. Y. Sidky and M. Vannier, Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?, *Inverse Problems*, **25** (2009), 123009.
[25] D. M. Pelt and V. De Andrade, Improved tomographic reconstruction of large-scale real-world data by filter optimization, *Advanced Structural and Chemical Imaging*, **2** (2016), 17.
[26] D. M. Pelt and K. J. Batenburg, Improving filtered backprojection reconstruction by data-dependent filtering, *IEEE Transactions on Image Processing*, **23** (2014), 4750-4762.
[27] M. Petrica and I. Popescu, A note on identifiability for inverse problem based on observations, preprint, `arXiv:2408.14616`.
[28] P. Rusnock and A. Kerr-Lawson, Bolzano and uniform continuity, *Historia Mathematica*, **32** (2005), 303-311.
[29] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2^{nd} edition, Society for Industrial and Applied Mathematics, 2003.
[30] L. A. Shepp and B. F. Logan, The Fourier reconstruction of a head section, *IEEE Transactions on Nuclear Science*, **21** (1974), 21-43.
[31] E. Y. Sidky and X. Pan, Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization, *Physics in Medicine and Biology*, **53** (2008), 4777.

[32] Y. Sun, L.-S. Schneider, F. Fan, M. Thies, M. Gu, S. Mei, Y. Zhou, S. Bayer and A. Maier, Data-Driven Filter Design in FBP: Transforming CT Reconstruction with Trainable Fourier Series, in *CT-Meeting*, 2024.
[33] X. Tan, X. Liu, K. Xiang, J. Wang and S. Tan, Deep filtered back projection for CT reconstruction, *IEEE Access*, **12** (2024), 20962-20972.
[34] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg and J. Sijbers, Fast and flexible X-ray tomography using the astra toolbox, *Opt. Express*, **24** (2016), 25129-25147.
[35] C. Villani et al., *Optimal Transport: Old and New*, vol. 338, Springer, 2008.
[36] M. Yang, G. Wang and Y. Liu, New reconstruction method for X-ray testing of multilayer printed circuit board, *Optical Engineering*, **49** (2010), 056501-056501.
[37] G. L. Zeng, A filtered backprojection algorithm with characteristics of the iterative landweber algorithm, *Medical Physics*, **39** (2012), 603-607.

Received for publication February 9, 2026.