

Optimized filter functions for filtered back projection reconstructions*

Matthias Beckmann[†] and Judith Nickel[‡]

Abstract. The method of filtered back projection (FBP) is a widely used reconstruction technique in X-ray computerized tomography (CT), which is particularly important in clinical diagnostics. To reduce scanning times and radiation doses in medical CT settings, enhancing the reconstruction quality of the FBP method is essential. To this end, this paper focuses on analytically optimizing the applied filter function. We derive a formula for the filter that minimizes the expected squared L^2 -norm of the difference between the FBP reconstruction, given infinite noisy measurement data, and the true target function. Additionally, we adapt our derived filter to the case of finitely many measurements. The resulting filter functions have a closed-form representation, do not require a training dataset, and, thus, provide an easy-to-implement, out-of-the-box solution. Our theoretical findings are supported by numerical experiments based on both simulated and real CT data.

Key words. X-ray tomography, image reconstruction, filtered back projection, filter design, Gaussian noise.

AMS subject classifications. 44A12, 65R32, 94A12.

1. Introduction. The method of *filtered back projection* (FBP) is a commonly used reconstruction algorithm in X-ray computerized tomography (CT) [38], which aims at determining the internal structure of objects under investigation by measuring the attenuation of X-rays. Until today, CT is widely employed in medical diagnostics and non-destructive testing. The measured data can be modelled as line integral values of the *attenuation function* $f \equiv f(x, y)$ of the object, formally represented via its *Radon transform* $\mathcal{R}f \equiv \mathcal{R}f(s, \varphi)$, defined by

$$\mathcal{R}f(s, \varphi) = \int_{\{x \cos(\varphi) + y \sin(\varphi) = s\}} f(x, y) \, d(x, y) \quad \text{for } (s, \varphi) \in \mathbb{R} \times [0, \pi).$$

With this, we can formulate the *CT reconstruction problem* as determining the attenuation function f from its Radon data

$$\{\mathcal{R}f(s, \varphi) \mid s \in \mathbb{R}, \varphi \in [0, \pi)\}.$$

Hence, one needs to invert the Radon transform $\mathcal{R}$. Based on the work of Johann Radon [45], its analytical inversion was proven to be performed by the FBP formula [18, 35], given by

$$f(x, y) = \frac{1}{2} \mathcal{B}(\mathcal{F}^{-1}[|\sigma| \mathcal{F}(\mathcal{R}f)(\sigma, \varphi)])(x, y) \quad \text{for all } (x, y) \in \mathbb{R}^2,$$

*The authors are listed in alphabetical order. Corresponding author: Judith Nickel.

Funding: This work was supported by the Deutsche Forschungsgemeinschaft (DFG) - project numbers 530863002 and 281474342/GRK2224/2 - as well as the Bundesministerium für Bildung und Forschung (BMBF) - funding code 03TR07W11A. The responsibility for the content of this publication lies with the authors.

[†]Center for Industrial Mathematics, University of Bremen, Germany & Department of Electrical and Electronic Engineering, Imperial College London, UK (research@mbeckmann.de).

[‡]Center for Industrial Mathematics, University of Bremen, Germany (judith.nickel@uni-bremen.de).

where $\mathcal{F}h$ denotes the univariate Fourier transform acting on the radial variable s of a function $h \equiv h(s, \varphi)$, i.e.,

$$\mathcal{F}h(\sigma, \varphi) = \int_{\mathbb{R}} h(s, \varphi) e^{-is\sigma} ds \quad \text{for } (\sigma, \varphi) \in \mathbb{R} \times [0, \pi),$$

and $\mathcal{B}h$ is the back projection of $h \equiv h(s, \varphi)$, defined by

$$\mathcal{B}h(x, y) = \frac{1}{\pi} \int_0^\pi h(x \cos(\varphi) + y \sin(\varphi), \varphi) d\varphi \quad \text{for } (x, y) \in \mathbb{R}^2.$$

The FBP formula is highly sensitive to noise as the involved filter $|\sigma|$ particularly amplifies high-frequency components. Since, in practice, only noisy Radon data g^ε can be measured, it is necessary to stabilize the FBP formula. To this end, the filter $|\sigma|$ is typically replaced by a compactly supported low-pass filter $A_L \equiv A_L(\sigma)$ with bandwidth $L > 0$ [36], leading to the *approximate FBP reconstruction formula*

$$f_L^\varepsilon(x, y) = \frac{1}{2} \mathcal{B}(\mathcal{F}^{-1}[A_L(\sigma) \mathcal{F}g^\varepsilon(\sigma, \varphi)])(x, y) \quad \text{for } (x, y) \in \mathbb{R}^2.$$

This approximate reconstruction formula is computational efficient, especially when compared to iterative reconstruction approaches such as the algebraic reconstruction technique (ART) and simultaneous iterative reconstruction technique (SIRT).

However, the reconstruction quality deteriorates with noisy measurement data, in particular when using standard filter functions (cf. [42]). Additionally, in medical CT, reducing scanning times and radiation doses is crucial due to the harmful effects of X-radiation, which degrades the reconstruction quality even further. Consequently, comprehensive research has focused on improving the reconstruction quality of the FBP method. One approach is to optimize the introduced filter function A_L of the approximate FBP formula. This topic has recently regained attention in the research community, with several studies focusing on enhancing reconstruction quality by adapting the filter function. These approaches can be divided into two categories: analytical methods that define desirable properties of the filter and data-driven methods that involve learning a filter function. However, most methods in the literature do not provide a closed-form solution for the filter. Instead, they require solving an optimization problem or rely on datasets that are challenging to obtain, particularly in the context of medical CT. Moreover, noise in the measured data is often only implicitly addressed in the design of the filter function. For further details, we refer to Section 2.

As opposed to this, in this work, we derive an analytical formula for an optimized filter function enhancing the reconstruction quality of the approximate FBP formula. To achieve this, we model the noise in the measured data as a generalized stochastic process that approximates Gaussian white noise, which is grounded on observations in the literature. Based on this noise model, we construct our filter function by minimizing the expectation of the squared L^2 -norm of the difference between the approximate FBP reconstruction f_L^ε and the true attenuation function f . We derive a formula for the resulting filter function in the context of complete measurement data and provide an adaptation for the case of a finite number of measurements. Our proposed filters are as straightforward to implement as standard filters, quite fast to compute and do not require a training dataset, making them readily applicable.

The manuscript is organized as follows. Section 2 is devoted to a more detailed review of related result in the literature. In Section 3, we develop a suitable model for the measurement noise based on generalized stochastic Gaussian white noise processes. We then derive our optimized filter function in the continuous setting in Section 4.1 and adapt it to finite measurement data in Section 4.2. We validate the performance of our filter function through numerical experiments in Section 5 by comparing it with filter functions from the literature. Finally, Section 6 concludes with a discussion of our findings and future research directions.

2. Related results. The FBP method remains one of the most commonly used reconstruction algorithms in CT and extensive research has focused on improving its reconstruction quality. In the following, we describe various approaches that have been developed in the literature to address this task. One idea involves developing pre- and post-processing methods, either for denoising the measured data or enhancing the computed reconstructions. Besides classical methods like Wiener filtering and wavelet decomposition, recent works have explored the use of deep learning techniques. For more details on pre- and post-processing methods in the context of CT, we refer the reader to the overview articles [2, 12, 37, 47, 46].

In addition, the FBP method allows for a direct way to improve the reconstruction quality by changing the involved filter function A_L , which has been the focus of several studies. For noiseless measurement data, [3] noted that the classical Ram-Lak filter is the best choice w.r.t. the L^2 -norm of the reconstruction error given complete measurement data g . In [49], the authors adapt the Ram-Lak filter for noiseless finite measurements g_D by expressing the filter function through its Fourier series expansion and determining the coefficients by solving an optimization problem derived from a reformulation of the FBP method.

In real-world applications, however, noise in the measured data has to be considered when designing filter functions. In [42], a data-dependent filter for the FBP method is introduced, which minimizes the squared difference between the Radon transform of the reconstruction and the noisy data g_D^ε so that a minimization problem must be solved for each measurement independently. The authors call the resulting method the minimum residual FBP method, abbreviated as MR-FBP. They also add a regularization term to the minimization problem, resulting in the so-called MR-FBP_{GM} method. The regularization term incorporates a constraint on the horizontal and vertical discrete gradients of the reconstruction and can be controlled by an additional hyperparameter $\lambda > 0$.

Recently, the authors in [25] aim at choosing the filter function by minimizing the true risk w.r.t. the squared L^2 -error between the FBP reconstruction and the ground truth, i.e., they consider the minimization problem

$$\min_{A_L} \mathbb{E} \left(\left\| f - \frac{1}{2} \mathcal{B}(\mathcal{F}^{-1} A_L * g^\varepsilon) \right\|_{L^2(\mathbb{R}^2)}^2 \right),$$

where the expectation is w.r.t. the ground truth f and the noisy measurement data g^ε . Since calculating the true risk is infeasible in practice due to an unknown data distribution, they propose to minimize the empirical risk instead by replacing the expectation by the mean over a finite dataset consisting of noiseless measurements and additive noise samples. We refer to this as the empirical risk minimization FBP method, in short ERM-FBP.

Another approach for improving reconstruction quality is to mimic other reconstruction

methods, such as iterative reconstruction techniques. This approach is demonstrated in [43], where the filter is designed in such a way that the FBP method approximates the algebraic simultaneous iterative reconstruction technique, resulting in the SIRT-FBP method. The SIRT-FBP filter depends solely on the discretization parameters and can be pre-calculated before reconstructing. This filter is applied to a large-scale real-world dataset in [44].

The authors of [23] design filters that integrate both the filtering process with ideal ramp filter and an interpolation scheme, which is commonly employed when applying the FBP method, to reduce errors caused by interpolation. The authors discovered that their filters substantially enhance reconstructions only for interpolation schemes of up to second order.

Besides these analytical approaches, there have been recent developments in learning-based selection of filter functions. As this paper focuses on analytically optimizing the filter function, we will refrain from discussing it here. Instead, the interested reader is encouraged to consult [30, 41, 50, 52, 60] for various methods involving neural networks.

3. White noise and technical lemmata. In applications, the Radon data $\mathcal{R}f$ is always corrupted by noise caused, for example, by scattered radiation or the conversion of the measured X-ray photons into a digital signal by the detectors, cf. [10, 14, 22, 26, 59]. Due to the various sources of measurement errors and unknown system calibrations in real-world applications, it is in general not feasible to analytically model the probability distribution of the noise, see [19, 31]. Instead, one needs to approximate the distribution based on measurement data, which is dealt with in several papers, e.g., [31, 32, 33, 57]. The findings in the literature indicate that the measurement noise can be modelled by independent identically distributed (iid) additive Gaussian random variables, even in low intensity regimes. This process is commonly referred to as Gaussian white noise.

Its formal definition is based on the classical Wiener process $W : \Omega' \times \Omega \rightarrow \mathbb{R}$, also called Brownian motion process, with $\Omega' \subseteq \mathbb{R}^n$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which describes the path of a particle without inertia caused by Brownian motion. Gaussian white noise, typically denoted by H , is then defined as the distributional derivative of the Wiener process and characterizes the velocity of a particle without inertia due to Brownian motion. To be more precise, $H : \mathcal{D}(\Omega') \times \Omega \rightarrow \mathbb{R}$ is given by

$$H(\phi, \omega) = -\langle W(\cdot, \omega), \phi' \rangle = - \int_{\Omega'} W(x, \omega) \phi'(x) \, dx$$

for all $\phi \in \mathcal{D}(\Omega')$ and $\omega \in \Omega$, where $\mathcal{D}(\Omega')$ denotes the space of real-valued test functions. Note that, for simplicity, we consider the space of real-valued test functions $\mathcal{D}(\Omega')$ as well as the space of real-valued Schwartz functions $\mathcal{S}(\mathbb{R}^n)$ throughout this work. Our results, however, can be easily modified to complex-valued functions. The properties of the Wiener process imply that $H(\phi, \cdot)$ is a Gaussian random variable for all $\phi \in \mathcal{D}(\Omega')$ with

$$\mathbb{E}(H(\phi, \cdot)) = 0 \quad \text{and} \quad \mathbb{E}(H(\phi, \cdot)H(\psi, \cdot)) = (\phi, \psi)_{L^2(\Omega')}$$

for all $\phi, \psi \in \mathcal{D}(\Omega')$. For a precise definition of a Wiener process and a detailed derivation of the above properties we refer the reader to [21, Chapter III] and [51, Chapter 1]. Since a Gaussian process is uniquely defined by its mean and covariance, cf. [21, Chapter III.2.3],

we obtain the following equivalent definition, which is more commonly used in the setting of inverse problems, see, e.g., [27, p.3].

Definition 3.1 (Continuous white noise). Continuous white noise is a function

$$H : \mathcal{D}(\Omega') \times \Omega \rightarrow \mathbb{R}$$

such that $H(\cdot, \omega) \in \mathcal{D}'(\Omega')$ for all $\omega \in \Omega$ and $H(\phi, \cdot) : \Omega \rightarrow \mathbb{R}$ is a Gaussian random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for all $\phi \in \mathcal{D}(\Omega')$ with

$$\mathbb{E}(H(\phi, \cdot)) = 0 \quad \text{and} \quad \mathbb{E}(H(\phi, \cdot)H(\psi, \cdot)) = (\phi, \psi)_{L^2(\Omega')}$$

for all $\phi, \psi \in \mathcal{D}(\Omega')$. In addition, continuous white noise $H^\varepsilon : \mathcal{D}(\Omega') \times \Omega \rightarrow \mathbb{R}$ with noise level $\varepsilon > 0$ is defined by $H^\varepsilon = \varepsilon H$.

We remark that continuous white noise is a so-called *generalized stochastic process*, as introduced in [21, Chapter III]. Moreover, one can prove that $H(\cdot, \omega) \in \mathcal{S}'(\mathbb{R}^n)$ for almost all $\omega \in \Omega$, see [11, 15] and, thus, H can be equivalently defined as a mapping from $\mathcal{S}(\mathbb{R}^n) \times \Omega$ to $\mathbb{R}$ such that $H(\cdot, \omega)$ has support in $\Omega' \subseteq \mathbb{R}^n$ in the sense of distributions.

The definition of continuous white noise H implies that the amount of power is the same for all frequencies, i.e., the power spectrum of the process is flat, cf. [20, Chapter 1.5.4]. Such a process cannot exist in reality as it would require an infinite amount of energy. Instead, continuous white noise can be understood as an idealised model of a process with approximately independent values at every point. To circumvent this shortcoming of white noise, physicists often approximate white noise by assuming that points very close to each other are correlated, see [20, Chapter 4.1] and [51, Chapter I.10]. This motivates the following definition of approximate white noise.

Definition 3.2 (Approximate continuous white noise). Approximate continuous white noise $H_a : \mathcal{S}(\mathbb{R}^n) \times \Omega \rightarrow \mathbb{R}$ with support in $\Omega' \subseteq \mathbb{R}^n$ is defined as

$$H_a(\phi, \omega) = \int_{\mathbb{R}^n} h_a(x, \omega) \phi(x) \mathbb{1}_{\Omega'}(x) \, dx,$$

where $\mathbb{1}_{\Omega'} : \mathbb{R}^n \rightarrow \{0, 1\}$ is the characteristic function of Ω' and $h_a : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$ for $a > 0$ is a jointly-measurable function with the property that, for any $k \in \mathbb{N}$, $(h_a(x_1, \cdot), \dots, h_a(x_k, \cdot))$ with $x_1, \dots, x_k \in \mathbb{R}^n$ is a (multivariate) Gaussian random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with

$$\mathbb{E}(h_a(x, \cdot)) = 0 \quad \text{and} \quad \mathbb{E}(h_a(x, \cdot)h_a(y, \cdot)) = \delta_a(x - y) \quad \text{for all } x, y \in \mathbb{R}^n,$$

where $(\delta_a)_{a>0}$ is a Dirac sequence with even $\delta_a : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ satisfying $\delta_a \rightarrow \delta$ for $a \rightarrow \infty$. In addition, approximate continuous white noise $H_a^\varepsilon : \mathcal{S}(\mathbb{R}^n) \times \Omega \rightarrow \mathbb{R}$ of noise level $\varepsilon > 0$ is defined by $H_a^\varepsilon = \varepsilon H_a$.

In the following, we restrict our investigations to $H_a = H_a^1$. Note, however, that all results also hold for the general case of H_a^ε with $\varepsilon > 0$. Let us first observe that for fixed $\omega \in \Omega$, the realization $H_a(\cdot, \omega) \in \mathcal{S}'(\mathbb{R}^n)$ is a tempered distribution. Moreover, for fixed $x \in \mathbb{R}^n$, the variance $\mathbb{E}(h_a(x, \cdot)h_a(x, \cdot)) = \delta_a(0)$ diverges for $a \rightarrow \infty$.

In the case that $\delta_a(t) = \frac{a}{2} \exp^{-a|t|}$ for $t \in \mathbb{R}$, h_a corresponds to a stationary *Ornstein-Uhlenbeck* process with zero mean [20, Chapter 3.8.4 & 4.1]. The Ornstein-Uhlenbeck process was first introduced as an improved model of the velocity of a particle caused by Brownian motion [55] and is commonly used as an approximation of Gaussian white noise in various use cases, see, e.g., [8], [20, Chapter 4.1], [24] and [51, Chapter I.10]. In this sense, Definition 3.2 can be seen as a generalization of the Ornstein-Uhlenbeck process, allowing for a wider variety of processes.

The approximate white noise H_a is again a Gaussian random variable for fixed $\phi \in \mathcal{S}(\mathbb{R}^n)$, whose first and second moments depend on the function h_a . This is characterized in the following lemma.

Lemma 3.3. *Let $H_a : \mathcal{S}(\mathbb{R}^n) \times \Omega \rightarrow \mathbb{R}$ be an approximate continuous white noise with support in $\Omega' \subseteq \mathbb{R}^n$. Then, $H_a(\phi, \cdot) : \Omega \rightarrow \mathbb{R}$ is a Gaussian random variable for all $\phi \in \mathcal{S}(\mathbb{R}^n)$ with*

$$\mathbb{E}(H_a(\phi, \cdot)) = 0 \quad \text{and} \quad \mathbb{E}(H_a(\phi, \cdot)H_a(\psi, \cdot)) = \int_{\Omega'} \int_{\Omega'} \delta_a(x - y) \phi(x) \psi(y) \, dx \, dy$$

for all $\psi, \phi \in \mathcal{S}(\mathbb{R}^n)$.

Proof. Let us start with the calculation of the expectation of $H_a(\phi, \cdot)$ for $\phi \in \mathcal{S}(\mathbb{R}^n)$. To this end, observe that $H_a(\phi, \cdot)$ is a random variable as h_a is jointly-measurable and

$$\int_{\Omega} \int_{\Omega'} |h_a(x, \omega)| |\phi(x)| \, dx \, dP(\omega) = \sqrt{\frac{2\delta_a(0)}{\pi}} \int_{\Omega'} |\phi(x)| \, dx < \infty,$$

which follows from Fubini's Theorem and $\mathbb{E}(|h_a(x, \cdot)|) = \sqrt{2\delta_a(0)/\pi}$ for all $x \in \mathbb{R}^n$ due to [54]. Hence, applying Fubini's Theorem again results in

$$\mathbb{E}(H_a(\phi, \cdot)) = \int_{\Omega} \int_{\Omega'} h_a(x, \omega) \phi(x) \, dx \, dP(\omega) = \int_{\Omega'} \mathbb{E}(h_a(x, \cdot)) \phi(x) \, dx = 0.$$

For the calculation of the covariance of $H_a(\phi, \cdot)$ and $H_a(\psi, \cdot)$ for $\phi, \psi \in \mathcal{S}(\mathbb{R}^n)$ note that

$$\begin{aligned} & \int_{\Omega} \int_{\Omega'} \int_{\Omega'} |h_a(x, \omega) h_a(y, \omega)| |\phi(x) \psi(y)| \, dx \, dy \, dP(\omega) \\ &= \int_{\Omega'} \int_{\Omega'} \mathbb{E}(|h_a(x, \cdot) h_a(y, \cdot)|) |\phi(x) \psi(y)| \, dx \, dy \\ &\leq \int_{\Omega'} \int_{\Omega'} \mathbb{E}(|h_a(x, \cdot)|^2)^{1/2} \mathbb{E}(|h_a(y, \cdot)|^2)^{1/2} |\phi(x) \psi(y)| \, dx \, dy \end{aligned}$$

by Fubini's Theorem and Cauchy Schwarz's inequality so that

$$\int_{\Omega} \int_{\Omega'} \int_{\Omega'} |h_a(x, \omega) h_a(y, \omega)| |\phi(x) \psi(y)| \, dx \, dy \, dP(\omega) \leq \delta_a(0) \int_{\Omega'} \int_{\Omega'} |\phi(x) \psi(y)| \, dx \, dy < \infty.$$

Consequently, we obtain

$$\begin{aligned}\mathbb{E}(H_a(\phi, \cdot)H_a(\psi, \cdot)) &= \int_{\Omega} \int_{\Omega'} \int_{\Omega'} h_a(x, \omega)h_a(y, \omega) \phi(x)\psi(y) \, dx \, dy \, dP(\omega) \\ &= \int_{\Omega'} \int_{\Omega'} \mathbb{E}(h_a(x, \cdot)h_a(y, \cdot)) \phi(x)\psi(y) \, dx \, dy \\ &= \int_{\Omega'} \int_{\Omega'} \delta_a(x - y) \phi(x)\psi(y) \, dx \, dy.\end{aligned}$$

In order to prove that $H_a(\phi, \cdot) : \Omega \rightarrow \mathbb{R}$ is a Gaussian random variable for all $\phi \in \mathcal{S}(\mathbb{R}^n)$, we consider the set

$$\mathcal{H} = \overline{\text{span}\{h_a(x, \cdot) : \Omega \rightarrow \mathbb{R} \mid x \in \mathbb{R}^n\}},$$

where the closure is w.r.t. the canonical norm of the Hilbert space of square integrable random variables on $(\Omega, \mathcal{F}, P)$. As h_a is a Gaussian process by assumption, $\mathcal{H}$ consists of Gaussian random variables. The aim now is to show that $H_a(\phi, \cdot) \in \mathcal{H}$, where we have $\mathcal{H} = (\mathcal{H}^\perp)^\perp$ since $\mathcal{H}$ is closed. We have already seen that $H_a(\phi, \cdot)$ is a square-integrable random variable. Thus, it remains to prove that $\mathbb{E}(XH_a(\phi, \cdot)) = 0$ for all $X \in \mathcal{H}^\perp$. To this end, note that

$$\begin{aligned}\int_{\Omega} \mathbb{E}(|X| |h_a(x, \cdot)|) |\phi(x)| \, dx &\leq \int_{\Omega} \mathbb{E}(|X|^2)^{1/2} \mathbb{E}(|h_a(x, \cdot)|^2)^{1/2} |\phi(x)| \, dx \\ &= \mathbb{E}(|X|^2)^{1/2} \delta_a(0)^{1/2} \int_{\Omega} |\phi(x)| \, dx < \infty\end{aligned}$$

due to Cauchy Schwarz's inequality. As a result, Fubini's Theorem yields

$$\mathbb{E}(XH_a(\phi, \cdot)) = \int_{\Omega'} \mathbb{E}(Xh_a(x, \cdot))\phi(x) \, dx = 0$$

because $h_a(x, \cdot) \in \mathcal{H}$ for all $x \in \mathbb{R}^n$ and the proof is complete. ■

These properties of the approximate continuous white noise H_a imply its convergence to white noise H as specified in the following lemma.

Lemma 3.4. *Let $H_a : \mathcal{S}(\mathbb{R}^n) \times \Omega \rightarrow \mathbb{R}$ be an approximate continuous white noise with support in $\Omega' \subseteq \mathbb{R}^n$. Then, for all $\phi \in \mathcal{D}(\Omega')$, the sequence of random variables $(H_a(\phi, \cdot))_{a>0}$ converges weakly¹ to a Gaussian random variable $H_0(\phi, \cdot) : \Omega \rightarrow \mathbb{R}$ with*

$$(3.1) \quad \mathbb{E}(H_0(\phi, \cdot)) = 0 \quad \text{and} \quad \mathbb{E}(H_0(\phi, \cdot)H_0(\psi, \cdot)) = (\phi, \psi)_{L^2(\Omega')}$$

for all $\psi \in \mathcal{D}(\Omega')$.

Proof. First, observe that Lemma 3.3 yields $\lim_{a \rightarrow \infty} \mathbb{E}(H_a(\phi, \cdot)) = 0$ and

$$\lim_{a \rightarrow \infty} \mathbb{E}(H_a(\phi, \cdot)H_a(\psi, \cdot)) = (\phi, \psi)_{L^2(\Omega')}$$

¹A sequence of random variables $(X_n)_n$ converges weakly to a random variable X with cumulative distribution function F if the corresponding cumulative distribution functions $(F_n)_n$ converge pointwise to F at all points of continuity of F , cf. [9, Chapter 5].

for all $\phi, \psi \in \mathcal{D}(\Omega')$ as $(\delta_a)_{a \in \mathbb{R}_+}$ is a Dirac-sequence. The method of moments [9, Theorem 30.2] in combination with the property that the moments of a Gaussian random variable are uniquely determined by its first and second moment [39, Chapter 5.4] now imply that $H_a(\phi, \cdot)$ converges weakly to a Gaussian random variable $H_0(\phi, \cdot) : \Omega \rightarrow \mathbb{R}$ for all $\phi \in \mathcal{D}(\Omega')$ with expectation zero and variance $(\phi, \phi)_{L^2(\Omega')}$. For the covariance, observe that

$$\begin{aligned} \mathbb{E}(H_0(\phi, \cdot)H_0(\psi, \cdot)) &= \frac{1}{2} (\mathbb{E}((H_0(\phi, \cdot) + H_0(\psi, \cdot))^2) - \mathbb{E}(H_0(\phi, \cdot)^2) - \mathbb{E}(H_0(\psi, \cdot)^2)) \\ &= \frac{1}{2} ((\phi + \psi, \phi + \psi)_{L^2(\Omega')} - (\phi, \phi)_{L^2(\Omega')} - (\psi, \psi)_{L^2(\Omega')}) \\ &= (\phi, \psi)_{L^2(\Omega')}, \end{aligned}$$

where we used that $H_0(\phi, \cdot) + H_0(\psi, \cdot)$ is the weak limit of $H_a(\phi, \cdot) + H_a(\psi, \cdot)$ and the same arguments as above prove that the variance of $H_0(\phi, \cdot) + H_0(\psi, \cdot)$ is given by $(\phi + \psi, \phi + \psi)_{L^2(\Omega')}$. This completes the proof. $\blacksquare$

The last lemma shows that our definition of approximate white noise is sensible in the sense that its weak limit has the same first and second moments as white noise H .

An important characteristic of white noise is the regularity of its realizations. There are several works addressing the regularity of white noise H , e.g., [1, 56, 16, 17]. For instance, [1] shows that the realizations of white noise $H : \mathcal{S}(\mathbb{R}^n) \times \Omega \rightarrow \mathbb{R}$ belong to the weighted Besov space $B_p^\tau(\mathbb{R}^n, \rho)$ with smoothness $\tau \in \mathbb{R}$, integrability $p \in \mathbb{N}$ and weight exponent $\rho \in \mathbb{R}$, cf. [1, [53, Chapter 1.2.3], almost surely if and only if $\tau < -n/2$ and $\rho < -n/p$. Similar arguments as in [1] show that the realizations of white noise $H : \mathcal{S}(\Omega') \times \Omega \rightarrow \mathbb{R}$ with compact support $\Omega' \subset \mathbb{R}^n$ belong to $B_p^\tau(\mathbb{R}^n, \rho)$ almost surely if and only if $\tau < -n/2 + n/p$ and $\rho \in \mathbb{R}$. Consequently, embedding results of weighted Besov spaces imply that the realizations of white noise with compact support belong to the Sobolev space of fractional order $H^\alpha(\mathbb{R}^n)$ almost surely if $\alpha < 0$, see [7, Theorem 6.4.4], [29] and [34]. Here, the Sobolev space $H^\alpha(\mathbb{R}^n)$ of fractional order $\alpha \in \mathbb{R}$ is defined as

$$H^\alpha(\mathbb{R}^n) = \{f \in \mathcal{S}'(\mathbb{R}^n) \mid \|f\|_{H^\alpha(\mathbb{R}^n)} < \infty\}$$

with

$$\|f\|_{H^\alpha(\mathbb{R}^n)}^2 = \int_{\mathbb{R}^n} (1 + \|\xi\|_{\mathbb{R}^n}^2)^\alpha |\mathcal{F}f(\xi)|^2 d\xi.$$

In contrast, approximate continuous white noise $H_a : \mathcal{S}(\Omega') \times \Omega \rightarrow \mathbb{R}$ with compact support $\Omega' \subset \mathbb{R}^n$ exhibits a higher regularity. This is shown in the subsequent lemma.

Lemma 3.5. *Assume that $H_a : \mathcal{S}(\mathbb{R}^n) \times \Omega \rightarrow \mathbb{R}$ is an approximate continuous white noise with compact support in $\Omega' \subset \mathbb{R}^n$. Then, $H_a(\cdot, \omega) \in L^2(\mathbb{R}^n)$ for almost all (f.a.a.) $\omega \in \Omega$. In addition, if, for some $\alpha \in \mathbb{R}$,*

$$\int_{\mathbb{R}^n} (1 + \|\xi\|_{\mathbb{R}^n}^2)^\alpha \int_{\Omega'} \int_{\Omega'} \delta_a(x - y) e^{-i(x-y)\xi} dx dy d\xi < \infty,$$

then $H_a(\cdot, \omega) \in H^\alpha(\mathbb{R}^n)$ f.a.a. $\omega \in \Omega$.

Proof. For the expectation of the L^2 -norm of H_a we obtain

$$\mathbb{E} \left(\|H_a\|_{L^2(\mathbb{R}^n)}^2 \right) = \int_{\Omega} \int_{\Omega'} |h_a(x, \omega)|^2 dx dP(\omega) = \int_{\Omega'} \delta_a(0) dx < \infty$$

by Fubini's Theorem and the compactness of Ω' . As a result, $H_a(\cdot, \omega) \in H^0(\mathbb{R}^n)$ f.a.a. $\omega \in \Omega$.

We proceed with determining the expectation of the H^α -norm of H_a with $\alpha \in \mathbb{R}$. To this end, observe that the expectation of $|\mathcal{F}h_a(\xi, \cdot)|^2$ for fixed $\xi \in \mathbb{R}^n$ can be rewritten as

$$\begin{aligned} \mathbb{E} (|\mathcal{F}h_a(\xi, \cdot)|^2) &= \int_{\Omega} \mathcal{F}h_a(\xi, \omega) \overline{\mathcal{F}h_a(\xi, \omega)} dP(\omega) \\ &= \int_{\Omega'} \int_{\Omega'} \int_{\Omega} h_a(x, \omega) h_a(y, \omega) dP(\omega) e^{-i(x-y)\xi} dx dy \\ &= \int_{\Omega'} \int_{\Omega'} \delta_a(x-y) e^{-i(x-y)\xi} dx dy \end{aligned}$$

due to Fubini's Theorem, which is applicable in this setting as

$$\begin{aligned} \int_{\Omega'} \int_{\Omega'} \int_{\Omega} |h_a(x, \omega) h_a(y, \omega)| dP(\omega) dx dy &= \int_{\Omega'} \int_{\Omega'} \mathbb{E} (|h_a(x, \cdot) h_a(y, \cdot)|) dx dy \\ &\leq \int_{\Omega'} \int_{\Omega'} \mathbb{E} (|h_a(x, \cdot)|^2)^{1/2} \mathbb{E} (|h_a(y, \cdot)|^2)^{1/2} dx dy \\ &= \int_{\Omega'} \int_{\Omega'} \delta_a(0) dx dy < \infty. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{E} \left(\|H_a\|_{H^\alpha(\mathbb{R}^n)}^2 \right) &= \int_{\Omega} \int_{\mathbb{R}^n} (1 + \|\xi\|_{\mathbb{R}^n}^2)^\alpha |\mathcal{F}h_a(\xi, \omega)|^2 d\xi dP(\omega) \\ &= \int_{\mathbb{R}^n} (1 + \|\xi\|_{\mathbb{R}^n}^2)^\alpha \int_{\Omega'} \int_{\Omega'} \delta_a(x-y) e^{-i(x-y)\xi} dx dy d\xi < \infty \end{aligned}$$

by assumption, which completes the proof. ■

The last result shows that the smoothness of approximate white noise is solely determined by the covariance δ_a . For example, realizations of the stationary Ornstein-Uhlenbeck process with zero mean and $\delta_a(t) = \frac{a}{2} \exp^{-a|t|}$ are almost surely continuous but nowhere differentiable, cf. [13, Theorem 1.2] in combination with the equivalent characterization of the Ornstein-Uhlenbeck process therein and [40].

4. Optimized filter function. The aim of this section is to derive optimized filter functions for Radon data corrupted by white noise. To this end, we consider the approximate FBP formula

$$f_L^\varepsilon(x, y) = \frac{1}{2} \mathcal{B}(\mathcal{F}^{-1}[A_L(\sigma) \mathcal{F}g^\varepsilon(\sigma, \varphi)])(x, y) \quad \text{for } (x, y) \in \mathbb{R}^2$$

with noisy measurements $g^\varepsilon : \mathbb{R} \times [0, \pi) \rightarrow \mathbb{R}$ of noise level $\varepsilon \geq 0$, low-pass filter $A_L : \mathbb{R} \rightarrow \mathbb{R}$, given by

$$A_L(\sigma) = |\sigma| W(\sigma/L) \quad \text{for } \sigma \in \mathbb{R},$$

fixed bandwidth $L > 0$ and even window $W \in L^\infty(\mathbb{R})$ with compact support in $[-1, 1]$. Observe that the assumptions on W imply $A_L \in L^2(\mathbb{R})$ with

$$\|A_L\|_{L^2(\mathbb{R})}^2 = \int_{-L}^L |\sigma|^2 |W(\sigma/L)|^2 d\sigma \leq \|W\|_{L^\infty(\mathbb{R})}^2 \int_{-L}^L |\sigma|^2 d\sigma = \frac{2}{3} L^3 \|W\|_{L^\infty(\mathbb{R})}^2 < \infty.$$

In addition, if $g^\varepsilon \in L^2(\mathbb{R} \times [0, \pi])$ has compact support, i.e., there exists an $R \geq 0$ with

$$g^\varepsilon(s, \varphi) = 0 \quad \forall |s| > R, \varphi \in [0, \pi),$$

the approximate FBP reconstruction $f_L^\varepsilon \in L^2(\mathbb{R}^2)$ is defined almost everywhere on $\mathbb{R}^2$ and can be rewritten as

$$f_L^\varepsilon = \frac{1}{2} \mathcal{B}(\mathcal{F}^{-1} A_L * g^\varepsilon) \quad \text{almost everywhere on } \mathbb{R}^2,$$

where

$$(\mathcal{F}^{-1} A_L * g^\varepsilon)(s, \varphi) = \int_{\mathbb{R}} \mathcal{F}^{-1} A_L(s - S, \varphi) g^\varepsilon(S, \varphi) dS \quad \text{for } (s, \varphi) \in \mathbb{R} \times [0, \pi).$$

Moreover, the (distributional) Fourier transform of f_L^ε is given by

$$(4.1) \quad \mathcal{F} f_L^\varepsilon(\sigma \cos(\varphi), \sigma \sin(\varphi)) = W(\sigma/L) \mathcal{F} g^\varepsilon(\sigma, \varphi) \quad \text{for almost all } \sigma \in \mathbb{R}, \varphi \in [0, \pi).$$

For details, we refer the reader to [4, Section 8.1].

4.1. Continuous setting. In Section 3, we argued that a suitable model for the measurement noise in X-ray CT is given by approximate additive Gaussian white noise. Based on this model, we now derive an optimized filter function by minimizing the expectation of the squared L^2 -error of the approximate FBP reconstruction f_L^ε . More precisely, as the filter A_L is uniquely determined by its window W , we consider the minimization problem

$$W^* = \arg \min_{\text{supp}(W) \subseteq [-1, 1]} \mathbb{E} \left(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2 \right),$$

where the real-valued target function $f \in L^2(\mathbb{R}^2)$ is assumed to be compactly supported with

$$\text{supp}(f) \subseteq B_R(0) \quad \text{for some } R > 0.$$

Note that this assumption is not very restrictive as objects under investigation are usually of finite extend, which implies that the Radon transform $\mathcal{R}f$ of f belongs to $L^2(\mathbb{R} \times [0, \pi])$ with

$$\mathcal{R}f(s, \varphi) = 0 \quad \forall |s| > R, \varphi \in [0, \pi).$$

Moreover, we assume that the Radon transform of f is corrupted by additive approximate white noise $H_a^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ of noise level $\varepsilon \geq 0$ with support in $\Omega' = [-R, R] \times [0, \pi)$ resulting in measured data $G_a^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ defined as

$$(4.2) \quad G_a^\varepsilon(\phi, \omega) = \int_0^\pi \int_{-R}^R g_a^\varepsilon(s, \varphi, \omega) \phi(s, \varphi) ds d\varphi \quad \text{for } \phi \in \mathcal{S}(\mathbb{R}^2), \omega \in \Omega$$

with $g_a^\varepsilon(s, \varphi, \omega) = \mathcal{R}f(s, \varphi) + h_a^\varepsilon(s, \varphi, \omega)$ for all $s \in [-R, R]$, $\varphi \in [0, \pi)$, $\omega \in \Omega$, where $h_a^\varepsilon = \varepsilon h_a$. Observe that we allow the case $\varepsilon = 0$, which corresponds to noiseless measured data and degenerate approximate white noise taking the value 0 with probability 1. Moreover, as $\mathcal{R}f$ is 2π -periodic in the angular variable φ , it would be more appropriate to consider the measured data G_a^ε as a function from $\mathcal{S}(\mathbb{R} \times \mathbb{S}^1) \times \Omega$ to $\mathbb{R}$. This, however, does not change our results, but only affects the notation. Lemma 3.5 in combination with $\mathcal{R}f \in L^2(\mathbb{R} \times [0, \pi))$ implies that $G_a^\varepsilon(\cdot, \omega) \in L^2(\mathbb{R} \times [0, \pi))$ for almost all $\omega \in \Omega$ and, thus, $F_L^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ with

$$(4.3) \quad F_L^\varepsilon(\phi, \omega) = \int_{\mathbb{R}^2} f_L^\varepsilon(x, y, \omega) \phi(x, y) \, d(x, y) \quad \text{for } \phi \in \mathcal{S}(\mathbb{R}^2), \omega \in \Omega,$$

where

$$f_L^\varepsilon(x, y, \omega) = \frac{1}{2} \mathcal{B}(\mathcal{F}^{-1} A_L(s) * g_a^\varepsilon(s, \varphi, \omega))(x, y) \quad \text{for almost all } (x, y) \in \mathbb{R}^2,$$

is well-defined with $F_L^\varepsilon(\cdot, \omega) \in L^2(\mathbb{R}^2)$ for almost all $\omega \in \Omega$. Moreover, $F_L^\varepsilon(\phi, \cdot) : \Omega \rightarrow \mathbb{R}$ is a random variable for fixed $\phi \in \Omega$. In what follows, we identify F_L^ε with f_L^ε as well as G_a^ε with g_a^ε to improve readability.

With these preliminaries, we now calculate the expectation of the squared $L^2(\mathbb{R}^2)$ -norm of the error $f_L^\varepsilon - f$.

Lemma 4.1. *Let $f \in L^2(\mathbb{R}^2)$ with $\text{supp}(f) \subseteq B_R(0)$ for fixed $R > 0$ and $A_L \in L^2(\mathbb{R})$ for $L > 0$. Moreover, assume that the measured data $G_a^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ is defined as in (4.2). Then,*

$$\begin{aligned} \mathbb{E} \left(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2 \right) &= \frac{L^2}{4\pi^2} \int_{\mathbb{R}} |\sigma| \left[(W(\sigma) - 1)^2 \int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 \, d\varphi \right. \\ &\quad \left. + \pi \varepsilon^2 W(\sigma)^2 \mathbb{E} (|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) \right] \, d\sigma \end{aligned}$$

with

$$\mathbb{E} (|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) = \int_{-R}^R \int_{-R}^R \delta_a(s - \hat{s}, 0) e^{-i(s-\hat{s})L\sigma} \, ds \, d\hat{s}.$$

Proof. Let us start with determining the $L^2(\mathbb{R}^2)$ -norm of $f_L^\varepsilon(\cdot, \omega) - f$ with fixed $\omega \in \Omega$. To this end, recall that $f_L^\varepsilon(\cdot, \omega) \in L^2(\mathbb{R}^2)$ for almost all $\omega \in \Omega$ and $f \in L^2(\mathbb{R}^2)$ by assumption. As a result, the Rayleigh-Plancherel theorem implies

$$\begin{aligned} \|f_L^\varepsilon(\cdot, \omega) - f\|_{L^2(\mathbb{R}^2)}^2 &= \frac{1}{4\pi^2} \|\mathcal{F}(f_L^\varepsilon(\cdot, \omega)) - \mathcal{F}f\|_{L^2(\mathbb{R}^2)}^2 \\ &= \frac{1}{4\pi^2} \int_0^\pi \int_{\mathbb{R}} |\mathcal{F}(f_L^\varepsilon(\cdot, \omega)) - \mathcal{F}f|(\sigma \cos(\varphi), \sigma \sin(\varphi))|^2 |\sigma| \, d\sigma \, d\varphi \end{aligned}$$

for almost all $\omega \in \Omega$ by transitioning to polar coordinates. Observe that $f \in L^1(\mathbb{R}^2)$ as $f \in L^2(\mathbb{R}^2)$ with $\text{supp}(f) \subseteq B_R(0)$ and, as a consequence, the classical Fourier slice theorem

$$\mathcal{F}(\mathcal{R}f)(\sigma, \varphi) = \mathcal{F}f(\sigma \cos(\varphi), \sigma \sin(\varphi)) \quad \forall \sigma \in \mathbb{R}, \varphi \in [0, \pi)$$

and (4.1) yield

$$\begin{aligned} \|f_L^\varepsilon(\cdot, \omega) - f\|_{L^2(\mathbb{R}^2)}^2 &= \frac{1}{4\pi^2} \int_0^\pi \int_{\mathbb{R}} |W(\sigma/L) \mathcal{F}g_a^\varepsilon(\sigma, \varphi, \omega) - \mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 |\sigma| \, d\sigma \, d\varphi \\ &= \frac{1}{4\pi^2} \int_0^\pi \int_{\mathbb{R}} |\sigma| \left[W(\sigma/L)^2 |\mathcal{F}g_a^\varepsilon(\sigma, \varphi, \omega)|^2 - W(\sigma/L) \mathcal{F}(\mathcal{R}f)(\sigma, \varphi) \overline{\mathcal{F}g_a^\varepsilon(\sigma, \varphi, \omega)} \right. \\ &\quad \left. - W(\sigma/L) \overline{\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)} \mathcal{F}g_a^\varepsilon(\sigma, \varphi, \omega) + |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 \right] d\sigma \, d\varphi. \end{aligned}$$

By exploiting $g_a^\varepsilon = \mathcal{R}f + h_a^\varepsilon$ with $h_a^\varepsilon = \varepsilon h_a$, we obtain

$$\begin{aligned} \|f_L^\varepsilon(\cdot, \omega) - f\|_{L^2(\mathbb{R}^2)}^2 &= \frac{1}{4\pi^2} \int_0^\pi \int_{\mathbb{R}} |\sigma| \left[(W(\sigma/L) - 1)^2 |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 + W(\sigma/L)^2 |\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \omega)|^2 \right. \\ &\quad \left. + (W(\sigma/L)^2 - W(\sigma/L)) [\mathcal{F}(\mathcal{R}f)(\sigma, \varphi) \overline{\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \omega)} \right. \\ &\quad \left. + \overline{\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)} \mathcal{F}h_a^\varepsilon(\sigma, \varphi, \omega)] \right] d\sigma \, d\varphi \end{aligned}$$

for almost all $\omega \in \Omega$. Now, for the expectation of $\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2 : \Omega \rightarrow \mathbb{R}$ follows that

$$\begin{aligned} \mathbb{E} \left(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2 \right) &= \int_{\Omega} \|f_L^\varepsilon(\cdot, \omega) - f\|_{L^2(\mathbb{R}^2)}^2 \, dP(\omega) \\ &= \frac{1}{4\pi^2} \int_{\Omega} \int_0^\pi \int_{\mathbb{R}} |\sigma| \left[(W(\sigma/L) - 1)^2 |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 \right. \\ &\quad \left. + W(\sigma/L)^2 |\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \omega)|^2 \right. \\ &\quad \left. + (W(\sigma/L)^2 - W(\sigma/L)) [\mathcal{F}(\mathcal{R}f)(\sigma, \varphi) \overline{\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \omega)} \right. \\ &\quad \left. + \overline{\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)} \mathcal{F}h_a^\varepsilon(\sigma, \varphi, \omega)] \right] d\sigma \, d\varphi \, dP(\omega) \\ &= \frac{1}{4\pi^2} \int_0^\pi \int_{\mathbb{R}} |\sigma| \left[(W(\sigma/L) - 1)^2 |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 \right. \\ &\quad \left. + W(\sigma/L)^2 \mathbb{E}(|\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot)|^2) \right. \\ &\quad \left. + (W(\sigma/L)^2 - W(\sigma/L)) [\mathcal{F}(\mathcal{R}f)(\sigma, \varphi) \mathbb{E}(\overline{\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot)}) \right. \\ &\quad \left. + \overline{\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)} \mathbb{E}(\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot))] \right] d\sigma \, d\varphi. \end{aligned}$$

The expectation of $\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot)$ is zero as

$$\mathbb{E}(\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot)) = \varepsilon \int_{-R}^R \mathbb{E}(h_a(s, \varphi, \cdot)) e^{-is\sigma} \, ds = 0 \quad \text{for all } s \in \mathbb{R}, \varphi \in [0, \pi)$$

due to the definition of h_a and Fubini's Theorem, which also implies that

$$\mathbb{E}(\overline{\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot)}) = \overline{\mathbb{E}(\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot))} = 0 \quad \text{for all } s \in \mathbb{R}, \varphi \in [0, \pi).$$

Similarly, for the variance of $\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot)$ follows that

$$\begin{aligned}\mathbb{E}(|\mathcal{F}h_a^\varepsilon(\sigma, \varphi, \cdot)|^2) &= \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(\sigma, \varphi, \cdot)|^2) \\ &= \varepsilon^2 \int_{-R}^R \int_{-R}^R \mathbb{E}[h_a(s, \varphi, \cdot) h_a(\hat{s}, \varphi, \cdot)] e^{-i(s-\hat{s})\sigma} ds d\hat{s} \\ &= \varepsilon^2 \int_{-R}^R \int_{-R}^R \delta_a(s - \hat{s}, 0) e^{-i(s-\hat{s})\sigma} ds d\hat{s}\end{aligned}$$

for all $\sigma \in \mathbb{R}$ and $\varphi \in [0, \pi)$. As a result, we finally obtain

$$\begin{aligned}\mathbb{E}(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2) &= \frac{L^2}{4\pi^2} \int_{\mathbb{R}} |\sigma| \left[(W(\sigma) - 1)^2 \int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi \right. \\ &\quad \left. + \pi \varepsilon^2 W(\sigma)^2 \int_{-R}^R \int_{-R}^R \delta_a(s - \hat{s}, 0) e^{-i(s-\hat{s})L\sigma} ds d\hat{s} \right] d\sigma,\end{aligned}$$

which completes the proof. ■

We now proceed to determining a window function $W^* \in L^\infty(\mathbb{R})$ with $W^*(\sigma) = W^*(-\sigma)$ for all $\sigma \in \mathbb{R}$ and $\text{supp}(W^*) \subseteq [-1, 1]$ that minimizes

$$\mathbb{E}(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2) = \frac{L^2}{4\pi^2} \int_{\mathbb{R}} k(\sigma, W(\sigma)) d\sigma$$

with $k : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ defined as

$$(4.4) \quad k(\sigma, w) = |\sigma|(w - 1)^2 \int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi + \pi \varepsilon^2 |\sigma| w^2 \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2)$$

for $\sigma, w \in \mathbb{R}$. To this end, we minimize $k(\sigma, \cdot)$ pointwise for fixed $\sigma \in \mathbb{R}$ and use the monotonicity of the integral to obtain a candidate for W^* . Since the window W is assumed to have compact support in $[-1, 1]$, we obtain

$$(4.5) \quad \mathbb{E}(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2) = \frac{L^2}{4\pi^2} \int_{-1}^1 k(\sigma, W(\sigma)) d\sigma + c,$$

where $c \in \mathbb{R}$ does not depend on W , and, therefore, it suffices to consider k as a function on $[-1, 1] \times \mathbb{R}$. Hence, in the following lemma, we derive a formula for the global minimizer of $k(\sigma, \cdot) : \mathbb{R} \rightarrow \mathbb{R}$ for $\sigma \in [-1, 1]$.

Lemma 4.2. *Let $f \in L^2(\mathbb{R}^2)$ with $\text{supp}(f) \subseteq B_R(0)$ for $R > 0$ and let $H_a^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ be approximate white noise with support in $[-R, R] \times [0, \pi)$ and noise level $\varepsilon \geq 0$. Then, for $\sigma \in [-1, 1]$, the global minimizer $w_\sigma^* \in \mathbb{R}$ of $k(\sigma, \cdot) : \mathbb{R} \rightarrow \mathbb{R}$ in (4.4) is given by*

$$w_\sigma^* = \begin{cases} \frac{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi}{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi + \pi \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2)} & \text{for } \sigma \in D \\ 1 & \text{for } \sigma \in [-1, 1] \setminus D \end{cases}$$

with

$$D = \left\{ \sigma \in [-1, 1] \setminus \{0\} \mid \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) > 0 \right\},$$

which is unique for $\sigma \in D$.

Proof. The proof can be divided into three cases:

Case $\sigma \in D$: The partial derivative of k w.r.t. the second variable w is given by

$$\frac{\partial k}{\partial w}(\sigma, w) = 2|\sigma|(w-1) \int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi + 2\pi\varepsilon^2 |\sigma| w \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) \quad \text{for } w \in \mathbb{R}$$

and, thus, the necessary condition of a minimum w_σ^* of $k(\sigma, \cdot)$ reads

$$\frac{\partial k}{\partial w}(\sigma, w_\sigma^*) = 0 \iff w_\sigma^* = \frac{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi}{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi + \pi\varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2)}.$$

Moreover, the second partial derivative

$$\frac{\partial^2 k}{\partial w^2}(\sigma, w) = 2|\sigma| \int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi + 2\pi\varepsilon^2 |\sigma| \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) \quad \text{for } w \in \mathbb{R}$$

is strictly larger than zero for all $\sigma \in D$ so that $k(\sigma, \cdot)$ is strictly convex and w_σ^* is the unique global minimizer of $k(\sigma, \cdot)$ on $\mathbb{R}$.

Case $\sigma = 0$: In this case, we have $k(\sigma, w) = 0$ for all $w \in \mathbb{R}$ and, consequently, $w_\sigma^* = 1$ is a global minimizer of $k(\sigma, \cdot)$ on $\mathbb{R}$.

Case $\sigma \in [-1, 1] \setminus (D \cup \{0\})$: In this setting, $k(\sigma, w)$ reads

$$k(\sigma, w) = |\sigma|(w-1)^2 \int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi \quad \text{for } w \in \mathbb{R}$$

with $k(\sigma, w) \geq 0$ for all $w \in \mathbb{R}$ and $k(\sigma, 1) = 0$. Therefore, $w_\sigma^* = 1$ is a global minimizer. ■

The preceding lemma, in combination with the monotonicity property of the integral, shows that a candidate for an optimal window function $W^* : \mathbb{R} \rightarrow \mathbb{R}$ with $\text{supp}(W^*) \subseteq [-1, 1]$ is given by

$$(4.6) \quad W^*(\sigma) = \begin{cases} \frac{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi}{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi + \pi\varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2)} & \text{for } \sigma \in D \\ 1 & \text{for } \sigma \in [-1, 1] \setminus D \\ 0 & \text{else} \end{cases}$$

with

$$D = \left\{ \sigma \in [-1, 1] \setminus \{0\} \mid \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) > 0 \right\}.$$

It remains to prove that the candidate W^* is indeed a window, i.e., $W^* \in L^\infty(\mathbb{R})$ is even.

Lemma 4.3. *Assume that $f \in L^2(\mathbb{R}^2)$ is real-valued with $\text{supp}(f) \subseteq B_R(0)$ for $R > 0$ and let $H_a^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ be approximate white noise with support in $[-R, R] \times [0, \pi)$ and noise level $\varepsilon \geq 0$. Then, $W^* : \mathbb{R} \rightarrow \mathbb{R}$ in (4.6) is well-defined, even and in $L^\infty(\mathbb{R})$.*

Proof. First, observe that

$$\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi = \int_0^\pi |\mathcal{F}f(L\sigma \cos(\varphi), L\sigma \sin(\varphi))|^2 d\varphi \leq \pi \|\mathcal{F}f\|_\infty^2 < \infty$$

and

$$\begin{aligned} \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) &\leq \int_{-R}^R \int_{-R}^R \mathbb{E}(|h_a(s, 0, \cdot) h_a(\hat{s}, 0, \cdot)|) ds d\hat{s} \\ &\leq \int_{-R}^R \int_{-R}^R \mathbb{E}(|h_a(s, 0, \cdot)|^2)^{1/2} \mathbb{E}(|h_a(\hat{s}, 0, \cdot)|^2)^{1/2} ds d\hat{s} = 4R^2 \delta_a(0) < \infty \end{aligned}$$

for all $\sigma \in \mathbb{R}$ due to the Fourier slice theorem, the Riemann-Lebesgue lemma, Fubini's theorem and Cauchy Schwarz's inequality. In addition,

$$|W^*(\sigma)| = \frac{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi}{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 d\varphi + \pi \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2)} \leq 1 \quad \text{for } \sigma \in D$$

and $W^*(\sigma) = 1$ for $\sigma \in [-1, 1] \setminus D$. Consequently, W^* is well-defined for all $\sigma \in \mathbb{R}$ and satisfies $W^* \in L^\infty(\mathbb{R})$. In order to prove that W^* is an even function, note that

$$|\mathcal{F}g(\sigma, \varphi)|^2 = \left| \int_{\mathbb{R}} g(s, \varphi) e^{-is\sigma} ds \right|^2 = \left| \int_{\mathbb{R}} \overline{g(s, \varphi)} e^{-is(-\sigma)} ds \right|^2 = |\mathcal{F}g(-\sigma, \varphi)|^2$$

for all real-valued $g \in L^2(\mathbb{R} \times [0, \pi))$ and all $\varphi \in [0, \pi)$. This implies that

$$|\mathcal{F}(\mathcal{R}f)(L\sigma, \varphi)|^2 = |\mathcal{F}(\mathcal{R}f)(-L\sigma, \varphi)|^2 \quad \text{and} \quad \mathbb{E}(|\mathcal{F}h_a(L\sigma, 0, \cdot)|^2) = \mathbb{E}(|\mathcal{F}h_a(-L\sigma, 0, \cdot)|^2)$$

for all $\sigma \in \mathbb{R}$ and $\varphi \in [0, \pi)$. Consequently, we have $\sigma \in D$ if and only if $-\sigma \in D$ and $W^*(\sigma) = W^*(-\sigma)$ for all $\sigma \in \mathbb{R}$. ■

The above Lemmata show that W^* is an optimized window function for the approximate FBP reconstruction in the sense that it minimizes the expectation of the squared $L^2(\mathbb{R}^2)$ -norm of $f_L^\varepsilon - f$ for window functions with compact support in $[-1, 1]$. This is summarized in the following theorem.

Theorem 4.4 (Optimized filter with compact support). *Assume that $f \in L^2(\mathbb{R}^2)$ is real-valued with $\text{supp}(f) \subseteq B_R(0)$ for fixed $R > 0$. In addition, assume that the Radon transform of f is corrupted by additive approximate white noise $H_a^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ with support in $[-R, R] \times [0, \pi)$ and noise level $\varepsilon \geq 0$, i.e., the measured data is given by $G_a^\varepsilon : \mathcal{S}(\mathbb{R}^2) \times \Omega \rightarrow \mathbb{R}$ with*

$$G_a^\varepsilon(\phi, \omega) = \int_0^\pi \int_{-R}^R g_a^\varepsilon(s, \varphi, \omega) \phi(s, \varphi) ds d\varphi \quad \text{for } \phi \in \mathcal{S}(\mathbb{R}^2), \omega \in \Omega,$$

where $g_a^\varepsilon(s, \varphi, \omega) = \mathcal{R}f(s, \varphi) + \varepsilon h_a(s, \varphi, \omega)$ for all $s \in [-R, R]$, $\varphi \in [0, \pi)$, $\omega \in \Omega$. Then, $A_L^* : \mathbb{R} \rightarrow \mathbb{R}$, defined by

$$A_L^*(\sigma) = \begin{cases} \frac{|\sigma| \int_0^\pi |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 d\varphi}{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 d\varphi + \pi \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2)} & \text{for } \sigma \in D_L \\ |\sigma| & \text{for } \sigma \in [-L, L] \setminus D_L \\ 0 & \text{else} \end{cases}$$

with

$$D_L = \left\{ \sigma \in [-L, L] \setminus \{0\} \mid \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2) > 0 \right\}$$

and

$$\mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2) = \int_{-R}^R \int_{-R}^R \delta_a(s - \hat{s}, 0) e^{-i(s-\hat{s})\sigma} ds d\hat{s},$$

is the optimized filter function with compact support that minimizes

$$\mathbb{E}(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2)$$

for filter functions $A_L(\cdot) = |\cdot| W(\cdot/L)$ with $\text{supp}(A_L) \subseteq [-L, L]$ and fixed $L > 0$.

Proof. The statement is a direct consequence of Lemmata 4.1, 4.2, 4.3, the monotonicity property of the integral and the identity $A_L^*(\cdot) = |\cdot| W^*(\cdot/L)$. $\blacksquare$

Observe that the optimized filter function depends on the covariance δ_a of h_a , the noise level ε and the true Radon transform $\mathcal{R}f$ of the target function f . The latter, however, is typically not known in applications, which makes the filter function difficult to use. We will address this in Chapter 5, where we present a possible adaptation of the above filter to the case of unknown true data.

In the case of noise-free measurement data, i.e., $\varepsilon = 0$, the optimized filter function reduces to

$$A_L^*(\sigma) = \begin{cases} |\sigma| & \text{for } \sigma \in [-L, L] \\ 0 & \text{else.} \end{cases}$$

Consequently, in this case, the optimized filter function is equal to the Ram-Lak filter. Similar observations were made in [4, 5].

To close this section, we consider the example of the above mentioned stationary Ornstein-Uhlenbeck process with zero mean and $\delta_a(t) = \frac{a}{2} \exp^{-a|t|}$ with $t \in \mathbb{R}$, which we extend to $\mathbb{R} \times [0, \pi)$ by setting

$$\delta_a(t, \varphi) = \frac{a^2}{4} \exp^{-a|t|} \exp^{-a|\varphi|} \quad \text{for } t \in \mathbb{R}, \varphi \in [0, \pi).$$

Standard computations then give

$$\mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2) = a^2 \left(\frac{aR}{a^2 + \sigma^2} + \frac{\frac{1}{2}(e^{-2aR}\cos(2R\sigma) - 1)(a^2 - \sigma^2) - a\sigma e^{-2aR}\sin(2R\sigma)}{a^4 + 2a^2\sigma^2 + \sigma^2} \right)$$

for all $\sigma \in \mathbb{R}$ and $a > 0$. Hence, in the limit $a \rightarrow \infty$ we obtain

$$\mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2) \xrightarrow{a \rightarrow \infty} \infty \quad \text{for all } \sigma \in \mathbb{R}$$

and the corresponding optimized filter function $A_L^*(\sigma)$ converges to 0 for all $\sigma \in \mathbb{R}$ and $\varepsilon > 0$. This behaviour is reasonable as H_a converges weakly to white noise for $a \rightarrow \infty$, cf. Lemma 3.4, with $\mathbb{E}(h_a(x, \cdot)^2) = \delta_a(0) \xrightarrow{a \rightarrow \infty} \infty$. In this sense, white noise has an infinite variance and corresponds to a process requiring an infinite amount of energy as already highlighted in Section 3. To compensate for this, the filter function must suppress all frequencies and, thus, $A_L^* \equiv 0$. The same behaviour can be observed for a larger variety of covariance functions. To this end, consider fixed $\varepsilon^2 > 0$ and assume that the covariance $\delta_a : \mathbb{R}^2 \rightarrow \mathbb{R}$ of the approximate white noise is of the form

$$\delta_a(s, \varphi) = a^2 \phi(as) \phi(a\varphi) \quad \text{for } s, \varphi \in \mathbb{R},$$

where $\phi \in L^1(\mathbb{R})$ with $\phi(x) \geq 0$ for all $x \in \mathbb{R}$, $\phi(0) > 0$ and $\int_{\mathbb{R}} \phi(x) dx = 1$. Then, δ_a is a Dirac sequence and

$$\mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2) = a^2 \phi(0) \int_{-R}^R \int_{-R}^R \phi(a(s - \hat{s})) e^{-i(s - \hat{s})\sigma} ds d\hat{s} \quad \text{for all } \sigma \in \mathbb{R}.$$

Observe that

$$\int_{-R}^R \int_{-R}^R a \phi(a(s - \hat{s})) e^{-i(s - \hat{s})\sigma} ds d\hat{s} \xrightarrow{a \rightarrow \infty} 2R > 0$$

uniformly for $\sigma \in [-L, L]$, which follows from Lebesgue's dominated convergence theorem, the uniform convergence of the complex exponential function on any bounded interval and the properties of ϕ . As a result, there exists an $a_0 > 0$ with $\mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2) > 0$ for all $\sigma \in [-L, L]$ and all $a \geq a_0$. Consequently,

$$A_L^*(\sigma) = \frac{|\sigma| \int_0^\pi |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 d\varphi}{\int_0^\pi |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 d\varphi + \pi \varepsilon^2 \mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2)} \quad \text{for } \sigma \in [-L, L], \quad a \geq a_0,$$

and in the limit $a \rightarrow \infty$, we obtain

$$A_L^*(\sigma) \xrightarrow{a \rightarrow \infty} 0 \quad \text{for all } \sigma \in \mathbb{R},$$

which follows from

$$\mathbb{E}(|\mathcal{F}h_a(\sigma, 0, \cdot)|^2) = a \phi(0) \int_{-R}^R \int_{-R}^R a \phi(a(s - \hat{s})) e^{-i(s - \hat{s})\sigma} ds d\hat{s} \xrightarrow{a \rightarrow \infty} \infty.$$

4.2. Discrete setting. The optimized filter function derived in the last section is based on the assumption that the measured data is known for every line in the plane. In real-world applications, however, this is not the case, since only finitely many measurements are available. Consequently, the approximate FBP formula and with this the optimized filter function are not applicable in these settings and need to be adapted to discrete measurement data. For simplicity, we assume that the data is measured using a parallel scanning geometry. To be more precise, the discrete measured data $g_D^\varepsilon : \{-M, \dots, M\} \times \{0, \dots, N_\varphi - 1\} \times \Omega \rightarrow \mathbb{R}$ with $M, N_\varphi \in \mathbb{N}$ is given by

$$g_D^\varepsilon(i, j, \omega) = \mathcal{R}f(s_i, \varphi_j) + h_a^\varepsilon(s_i, \varphi_j, \omega) \quad \text{for } -M \leq i \leq M, 0 \leq j \leq N_\varphi - 1, \omega \in \Omega$$

with $s_i = ih$ for $h > 0$, $\varphi_j = j\pi/N_\varphi$ and $g_a^\varepsilon : [-R, R] \times [0, \pi) \times \Omega \rightarrow \mathbb{R}$ as in (4.2), where the radial discretization is chosen to cover the entire interval, i.e., $Mh \geq R$. Moreover, we assume that h_a is sufficiently close to white noise such that $g_D^\varepsilon(i, j, \cdot) : \Omega \rightarrow \mathbb{R}$ and $g_D^\varepsilon(k, l, \cdot) : \Omega \rightarrow \mathbb{R}$ appear to be uncorrelated if $i \neq k$ or $j \neq l$. More precisely, the covariance of h_a has to satisfy

$$(4.7) \quad \mathbb{E}(h_a(s_i, \varphi_j, \cdot)h_a(s_k, \varphi_l, \cdot)) = \delta_a(s_i - s_k, \varphi_j - \varphi_l) = 0 \quad \text{if } i \neq k \text{ or } j \neq l$$

for all $-M \leq i, k \leq M$ and $0 \leq j, l \leq N_\varphi - 1$. This is for example the case if

$$\delta_a(s, \varphi) = \begin{cases} a^2 & \text{if } |s| \leq \frac{1}{2a}, |\varphi| \leq \frac{1}{2a}, \\ 0 & \text{else} \end{cases}$$

with $a > \max\left(\frac{1}{2h}, \frac{N_\varphi}{2\pi}\right)$. In this setting, the discrete measured data g_D^ε can be equivalently represented as

$$g_D^\varepsilon(i, j, \omega) = \mathcal{R}f(s_i, \varphi_j) + \xi_{i,j}^{\varepsilon_a}(\omega) \quad \text{for } -M \leq i \leq M, 0 \leq j \leq N_\varphi - 1, \omega \in \Omega$$

with independent and identically distributed Gaussian random variables $\xi_{i,j}^{\varepsilon_a} : \Omega \rightarrow \mathbb{R}$ with expectation zero and variance $\varepsilon_a^2 = \varepsilon^2 \delta_a(0, 0)$, which corresponds to the classical definition of discrete additive Gaussian white noise in the literature. Hence, in the following we use this more common definition of g_D^ε . In addition, we assume that the target function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is real-valued and bounded with

$$\text{supp}(f) \subseteq B_R(0) \quad \text{for some } R > 0.$$

To apply the approximate FBP formula to discrete measurement data, we follow a standard approach [36] and discretize (4.3) using the composite trapezoidal rule, analogous to [5, 6], resulting in

$$(4.8) \quad f_{L,D}^\varepsilon(x, y, \omega) = \frac{1}{2} \mathcal{B}_D(\mathcal{F}^{-1}A_L *_D g_D^\varepsilon(\cdot, \cdot, \omega))(x, y) \quad \text{for } (x, y) \in \mathbb{R}^2, \omega \in \Omega$$

with

$$\mathcal{B}_D(\mathcal{F}^{-1}A_L *_D g_D^\varepsilon(\cdot, \cdot, \omega))(x, y) = \frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} (\mathcal{F}^{-1}A_L *_D g_D^\varepsilon(\cdot, \cdot, \omega))(x \cos(\varphi_j) + y \sin(\varphi_j), \varphi_j)$$

for $(x, y) \in \mathbb{R}^2$ and

$$(\mathcal{F}^{-1} A_L *_D g_D^\varepsilon(\cdot, \cdot, \omega))(s, \varphi_j) = h \sum_{i=-M}^M \mathcal{F}^{-1} A_L(s - s_i) g_D^\varepsilon(i, j, \omega)$$

for $s \in \mathbb{R}$ and $0 \leq j \leq N_\varphi - 1$. Note that $f_{L,D}^\varepsilon$ is well-defined with $f_{L,D}^\varepsilon(\cdot, \cdot, \omega) \in L_{\text{loc}}^2(\mathbb{R}^2)$ for all $\omega \in \Omega$ if the filter function $A_L \in L^2(\mathbb{R})$ has finite bandwidth $L > 0$.

Moreover, we need to discretize the optimized filter function derived in the last section. For this, we again use the composite trapezoidal rule, which leads to

$$\int_0^\pi |\mathcal{F}(\mathcal{R}f)(\sigma, \varphi)|^2 d\varphi \approx \frac{\pi}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D(\mathcal{R}f)(\sigma, j)|^2 \quad \text{for } \sigma \in \mathbb{R}$$

with

$$\mathcal{F}_D(\mathcal{R}f)(\sigma, j) = h \sum_{i=-M}^M \mathcal{R}f(s_i, \varphi_j) e^{-is_i\sigma} \quad \text{for } \sigma \in \mathbb{R}, 0 \leq j \leq N_\varphi - 1,$$

and

$$\mathcal{F}h_a(\sigma, 0, \omega) \approx \mathcal{F}_Dh_a(\sigma, 0, \omega) = h \sum_{i=-M}^M h_a(s_i, 0, \omega) e^{-is_i\sigma} \quad \text{for } \sigma \in \mathbb{R}, \omega \in \Omega$$

with

$$\begin{aligned} \mathbb{E}(|\mathcal{F}_Dh_a(\sigma, 0, \cdot)|^2) &= h^2 \sum_{i=-M}^M \sum_{j=-M}^M \mathbb{E}(h_a(s_i, 0, \cdot) h_a(s_j, 0, \cdot)) e^{-i(s_i - s_j)\sigma} \\ &= h^2 \sum_{i=-M}^M \delta_a(0, 0) = h^2 (2M + 1) \delta_a(0, 0) \quad \text{for all } \sigma \in \mathbb{R}, \end{aligned}$$

where we used (4.7). Replacing the continuous expressions in Theorem 4.4 with the discretized expressions yields an optimized filter for discrete measurements in parallel beam geometry.

Definition 4.5 (Optimized filter function with compact support for discrete measurements).

For a real-valued, bounded function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ with $\text{supp}(f) \subseteq B_R(0)$ for $R > 0$ and discrete measurements $g_D^\varepsilon : \{-M, \dots, M\} \times \{0, \dots, N_\varphi - 1\} \times \Omega \rightarrow \mathbb{R}$ for fixed $M, N_\varphi \in \mathbb{N}$, given by

$$g_D^\varepsilon(i, j, \omega) = \mathcal{R}f(s_i, \varphi_j) + \xi_{i,j}^{\varepsilon_a}(\omega)$$

with $s_i = ih$ for $h > 0$ so that $Mh \geq R$, $\varphi_j = j\pi/N_\varphi$ and discrete Gaussian white noise ξ^{ε_a} , the optimized filter $A_{L,D}^* : \mathbb{R} \rightarrow \mathbb{R}$ of bandwidth $L > 0$ is defined by

$$A_{L,D}^*(\sigma) = \begin{cases} \frac{|\sigma| \frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D(\mathcal{R}f)(\sigma, j)|^2}{\frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D(\mathcal{R}f)(\sigma, j)|^2 + h^2 \varepsilon_a^2 (2M + 1)} & \text{for } \sigma \in [-L, L] \\ 0 & \text{for } \sigma \notin [-L, L]. \end{cases}$$

Note that the optimized filter for discrete measurements depends on the target function f , the discretization parameters M , N_φ and h as well as on the variance ε_a^2 , which imposes practical limitations on the applicability of $A_{L,D}^*$ because the objective function and variance are rarely known. We address this issue in the next chapter. As in the continuous setting, the optimized filter $A_{L,D}^*$ for noise-free discrete measurements coincides with the Ram-Lak filter.

In contrast to the optimized filter function for continuous measurement data derived in Theorem 4.4, $A_{L,D}^*$ does not minimize

$$(4.9) \quad \mathbb{E} \left(\|f_{L,D}^\varepsilon - f\|_{L^2(K)}^2 \right)$$

with compact $K \subset \mathbb{R}^2$, which is illustrated in the following remark.

Remark 4.6. The reconstruction error can be bounded above by

$$\begin{aligned} \|f_{L,D}^\varepsilon(\cdot, \cdot, \omega) - f\|_{L^2(K)}^2 &\leq \|f_{L,D}^\varepsilon(\cdot, \cdot, \omega) - 1/2 \mathcal{B}_D(\mathcal{F}^{-1}(|\cdot| \mathbb{1}_{[-L,L]}(\cdot) \mathcal{F}_D(\mathcal{R}f)))\|_{L^2(K)}^2 \\ &\quad + \|1/2 \mathcal{B}_D(\mathcal{F}^{-1}(|\cdot| \mathbb{1}_{[-L,L]}(\cdot) \mathcal{F}_D(\mathcal{R}f))) - f\|_{L^2(K)}^2 \end{aligned}$$

for fixed $\omega \in \Omega$ and compact $K \subset \mathbb{R}^2$. The first summand can be interpreted as a combination of the error induced by the filter function A_L and the measurement process. In contrast, the second summand can be considered as an approximation error, which is independent of the chosen filter function. The first summand can be bounded above by

$$\begin{aligned} &\|f_{L,D}^\varepsilon(\cdot, \cdot, \omega) - 1/2 \mathcal{B}_D(\mathcal{F}^{-1}(|\cdot| \mathbb{1}_{[-L,L]}(\cdot) \mathcal{F}_D(\mathcal{R}f)))\|_{L^2(K)}^2 \\ &\leq \text{diam}(K)^2 \frac{2LN_\varphi}{16\pi^2 N_\varphi^2} \sum_{j=0}^{N_\varphi-1} \int_{-L}^L |\sigma|^2 |W(\sigma/L) \mathcal{F}_D g_D^\varepsilon(\sigma, j, \omega) - \mathcal{F}_D(\mathcal{R}f)(\sigma, \varphi_j)|^2 d\sigma. \end{aligned}$$

Now, the optimized filter function for discrete measurement data defined in Definition 4.5 minimizes this upper bound, as can be shown by arguments similar to those in Section 4.1. To derive a filter function that minimizes (4.9), one needs a different optimization approach that is more involved since the relations used in the continuous setting do not carry over to the discrete case.

5. Numerical Experiments. In this section, we finally present selected numerical experiments to illustrate and validate our theoretical findings. To this end, recall that we aim at recovering the target function $f \in L^2(\mathbb{R}^2)$ with $\text{supp}(f) \subseteq B_R(0)$ from noisy measurements $g_D^\varepsilon : \{-M, \dots, M\} \times \{0, \dots, N_\varphi - 1\} \times \Omega \rightarrow \mathbb{R}$ given by

$$g_D^\varepsilon(i, j, \omega) = \mathcal{R}f(s_i, \varphi_j) + \xi_{i,j}^{\varepsilon_a}(\omega) \quad \text{for } -M \leq i \leq M, 0 \leq j \leq N_\varphi - 1, \omega \in \Omega$$

with $s_i = ih$ for $h \geq R/M$, $\varphi_j = j\pi/N_\varphi$ and noise $\xi_{i,j}^{\varepsilon_a} : \Omega \rightarrow \mathbb{R}$ with expectation zero and variance ε_a^2 , which is assumed to be known or can be estimated. For simplicity, by rescaling the noise level we set $\varepsilon_a = \varepsilon$. Moreover, we from now on suppress the dependence on $\omega \in \Omega$ and simply write

$$g_D^\varepsilon(i, j) = \mathcal{R}f(s_i, \varphi_j) + \xi_{i,j}^\varepsilon \quad \text{for } -M \leq i \leq M, 0 \leq j \leq N_\varphi - 1.$$

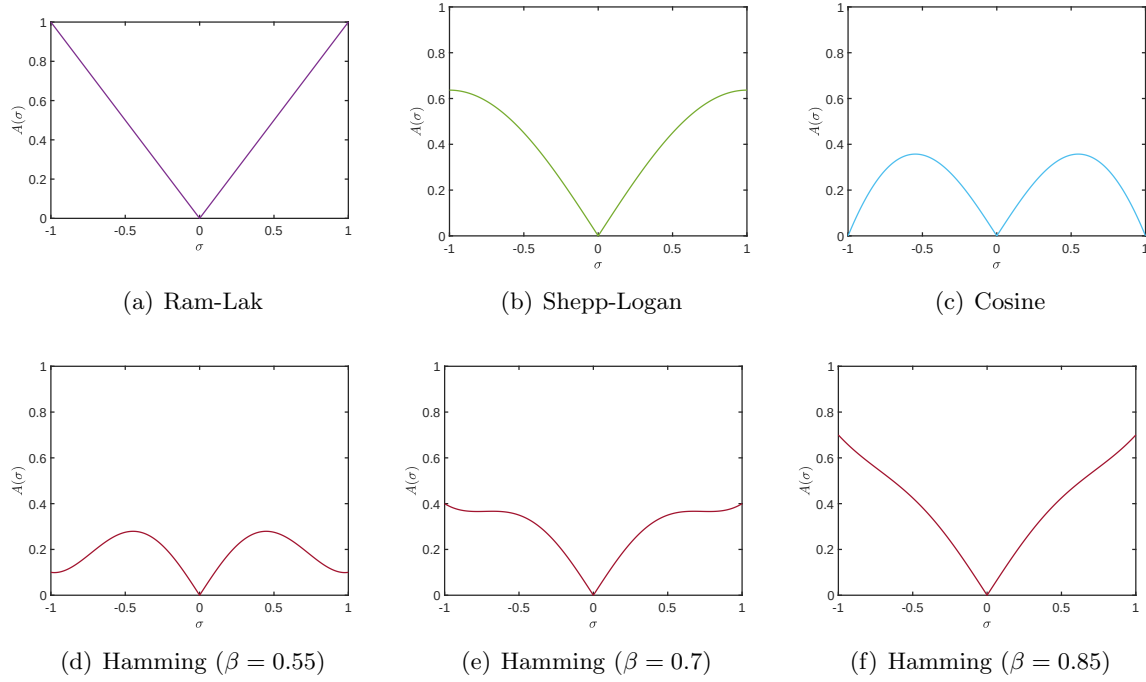


Figure 1. Selection of classical low-pass filters used in our numerical experiments.

We wish to apply the discretized approximate FBP formula (4.8), i.e.,

$$f_{L,D}^\varepsilon = \frac{1}{2} \mathcal{B}_D(\mathcal{F}^{-1} A_L *_{\mathcal{D}} g_D^\varepsilon).$$

The evaluation of $f_{L,D}^\varepsilon$, however, requires the computation of the values

$$(\mathcal{F}^{-1} A_L *_{\mathcal{D}} g_D^\varepsilon)(x \cos(\varphi_j) + y \sin(\varphi_j), \varphi_j) \quad \forall 0 \leq j \leq N-1$$

for each reconstruction point $(x, y) \in \mathbb{R}^2$, which is computationally expensive. To reduce the computational costs, a standard approach is to evaluate the function

$$s \mapsto (\mathcal{F}^{-1} A_L *_{\mathcal{D}} g_D^\varepsilon)(s, \varphi_j)$$

only at the points $s = s_l$, $l \in I$, for a sufficiently large index set $I \subset \mathbb{Z}$ and interpolate its function value for $s = x \cos(\varphi_j) + y \sin(\varphi_j)$ using an interpolation method $\mathcal{I}$. This leads us to the *discrete FBP reconstruction formula*

$$(5.1) \quad f_{\text{FBP}} = \frac{1}{2} \mathcal{B}_D(\mathcal{I}[\mathcal{F}^{-1} A_L *_{\mathcal{D}} g_D^\varepsilon]),$$

where linear or cubic spline interpolation is typically used depending on the regularity of f . Note, however, that the errors incurred by the utilized interpolation method are not taken into account in our theoretical derivation of optimized filter functions. Moreover, we choose the

Table 1

Window functions of classical low-pass filters, where $W(\sigma) = 0$ for all $|\sigma| > 1$ in all cases.

Name	$W(\sigma)$ for $ \sigma \leq 1$	Parameter
Ram-Lak	1	-
Shepp-Logan	$\text{sinc}(\pi\sigma/2)$	-
Cosine	$\cos(\pi\sigma/2)$	-
Hamming	$\beta + (1 - \beta) \cos(\pi\sigma)$	$\beta \in [1/2, 1]$

bandwidth $L > 0$ of the filter A_L and the discretization parameters $h > 0$, $M \in \mathbb{N}$ optimally depending on the number of angles $N_\varphi \in \mathbb{N}$ according to [36] via

$$M = \left\lfloor \frac{N_\varphi}{\pi} \right\rfloor, \quad L = \frac{\pi M}{R}, \quad h = \frac{\pi}{L},$$

where we assume that $\text{supp}(f) \subseteq B_R(0)$ with $R \in \mathbb{N}$.

The application of our optimized filter for discrete measurements from Definition 4.5,

$$A_{L,D}^*(\sigma) = \begin{cases} \frac{|\sigma| \frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D(\mathcal{R}f)(\sigma, j)|^2}{\frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D(\mathcal{R}f)(\sigma, j)|^2 + h^2 \varepsilon^2 (2M + 1)} & \text{for } \sigma \in [-L, L] \\ 0 & \text{for } \sigma \notin [-L, L], \end{cases}$$

requires the exact knowledge of the noiseless Radon samples $\mathcal{R}f(s_i, \varphi_j)$ for all $i = -M, \dots, M$ and $j = 0, \dots, N_\varphi - 1$. As this information is typically not available in applications, we investigate two options to overcome this bottleneck. In the first approach, we simply replace the Radon samples $\mathcal{R}f(s_i, \varphi_j)$ by the available measurements $g_D^\varepsilon(i, j)$ leading to the filter function

$$A_{L,D}^\varepsilon(\sigma) = \begin{cases} \frac{|\sigma| \frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D g_D^\varepsilon(\sigma, j)|^2}{\frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D g_D^\varepsilon(\sigma, j)|^2 + h^2 \varepsilon^2 (2M + 1)} & \text{for } \sigma \in [-L, L] \\ 0 & \text{for } \sigma \notin [-L, L]. \end{cases}$$

In the second approach, we first convolve the given measurements g_D^ε with a Wiener filter, which minimizes the expected squared error between convolved data $\hat{g}_D^\varepsilon$ and true data $\mathcal{R}f$. Thereon, we replace the Radon samples $\mathcal{R}f(s_i, \varphi_j)$ by the denoised measurements $\hat{g}_D^\varepsilon(i, j)$ yielding the filter function

$$\hat{A}_{L,D}^\varepsilon(\sigma) = \begin{cases} \frac{|\sigma| \frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D(\hat{g}_D^\varepsilon)(\sigma, j)|^2}{\frac{1}{N_\varphi} \sum_{j=0}^{N_\varphi-1} |\mathcal{F}_D(\hat{g}_D^\varepsilon)(\sigma, j)|^2 + h^2 \varepsilon^2 (2M + 1)} & \text{for } \sigma \in [-L, L] \\ 0 & \text{for } \sigma \notin [-L, L], \end{cases}$$

where the kernel size of the Wiener filter acts as hyperparameter. Let us stress that alternative denoising techniques can be applied like, e.g., data-driven approaches based on neural networks, which is beyond the scope of this work and asks for more in-depth future research.

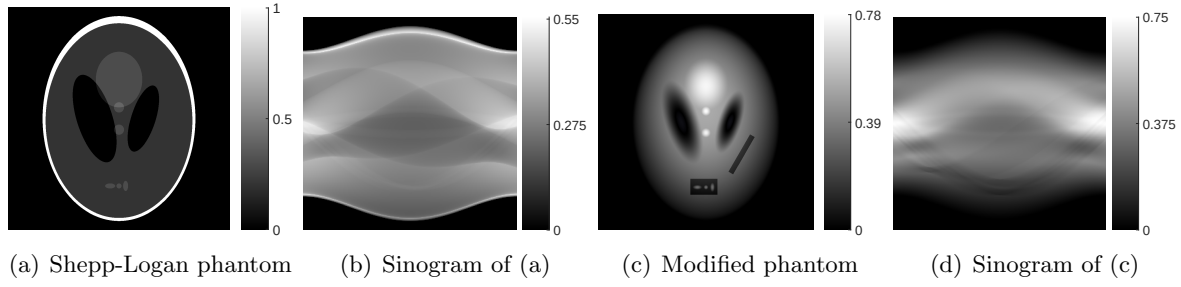


Figure 2. *Phantoms used in our numerical experiments along with their sinograms (Radon data).*

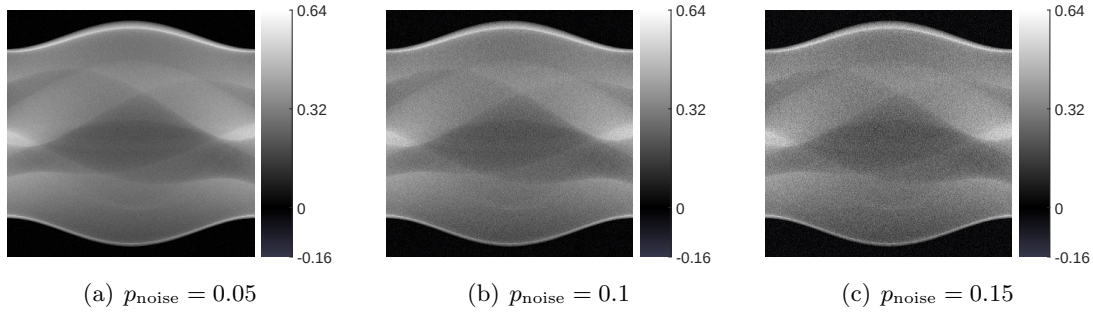


Figure 3. *Noisy sinograms of the Shepp-Logan phantom with different noise levels.*

In our numerical experiments we compare the reconstruction performance of our optimized filter function with classical low-pass filters, which are illustrated in Figure 1 and whose window functions are listed in Table 1, as well as the recently proposed filters from [25, 42, 43].

5.1. Shepp-Logan phantom. In a first set of numerical simulations, we make use of the popular Shepp-Logan phantom, proposed in [48] as a simplistic model of a human head section. It consists of ten ellipses of different sizes, eccentricities and locations leading to a piecewise constant attenuation function $f_{\text{SL}} \in L^2(\mathbb{R}^2)$ with $\text{supp}(f_{\text{SL}}) \subseteq B_1(0)$, whose Radon transform can be computed analytically, cf. Figure 2 (a)-(b). The noisy measurements $g_D^\varepsilon(i, j)$ are simulated by adding white Gaussian noise with variance ε^2 to the Radon samples $\mathcal{R}f_{\text{SL}}(s_i, \varphi_j)$, where $\varepsilon = p_{\text{noise}} m_{\mathcal{R}f_{\text{SL}}}$ with $p_{\text{noise}} \in \{0.05, 0.1, 0.15\}$ depends on the arithmetic mean

$$m_{\mathcal{R}f_{\text{SL}}} = \frac{1}{(2M+1)N_\varphi} \sum_{i=-M}^M \sum_{j=0}^{N_\varphi-1} |\mathcal{R}f_{\text{SL}}(s_i, \varphi_j)|,$$

see Figure 3 for an illustration of the resulting noisy sinograms.

The results of our numerical experiments are depicted in Figure 4, where we plot the mean squared error (MSE) in logarithmic scale and the structural similarity index measure (SSIM) from [58] of the FBP reconstruction on an equidistant grid of 1024×1024 pixels as a function of the number of angles $N_\varphi \in \{90, 180, 270, 360, 450, 540, 630, 720\}$ for various filter functions. To account for the randomness, we computed averages over 50 reconstructions with different noise realizations. Involved filter parameters were optimized based on 20 independent samples.

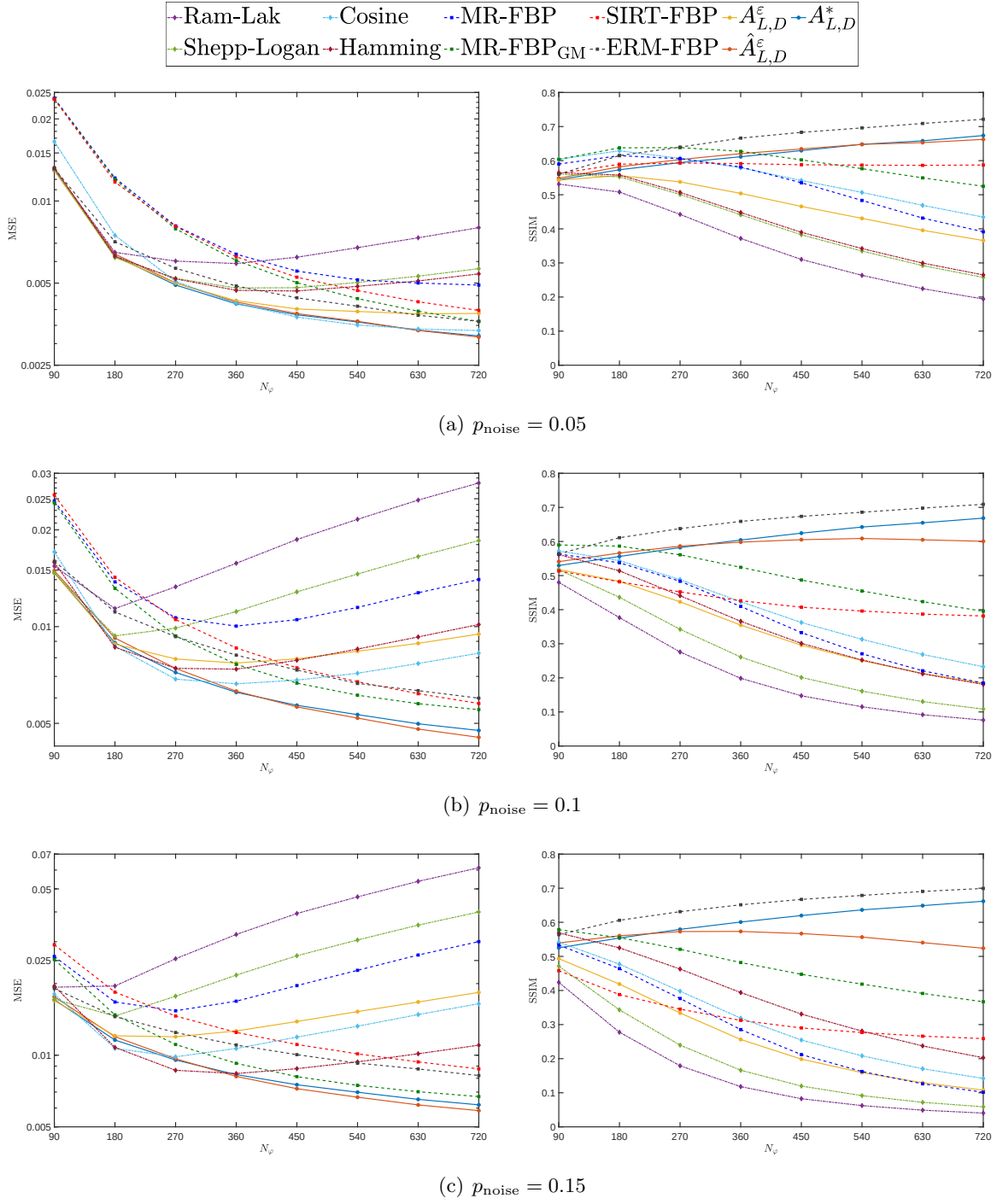


Figure 4. Plots of the MSE and SSIM of FBP reconstructions for the Shepp-Logan phantom.

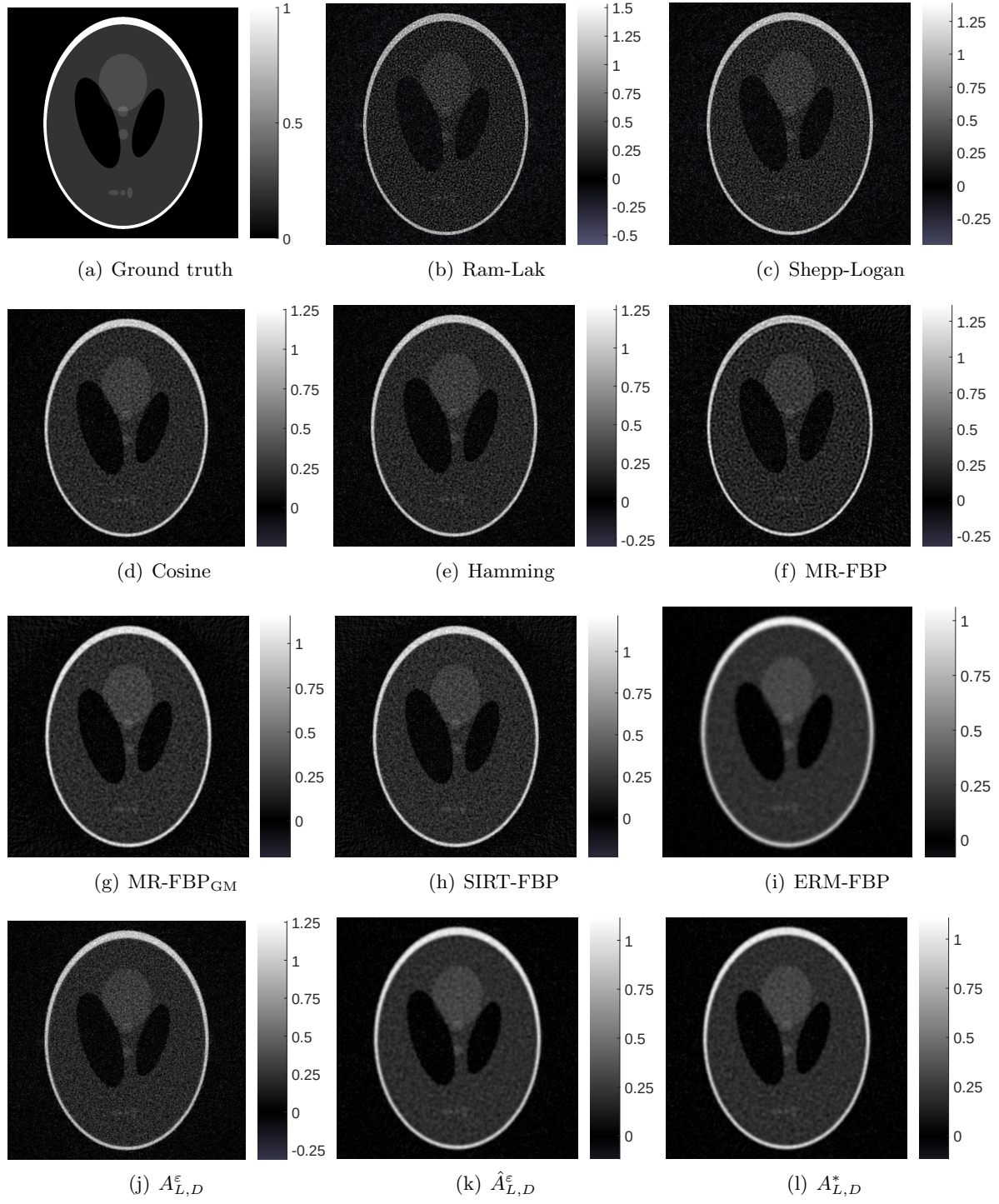


Figure 5. Reconstructions of Shepp-Logan phantom from noisy Radon data ($p_{\text{noise}} = 0.1$, $N_{\varphi} = 360$).

In terms of MSE, we observe that for the classical filters used (Ram-Lak, Shepp-Logan, Cosine, Hamming) the error first decreases, but then increases again with increasing N_φ . In contrast to this, for the MR-FBP_{GM} method from [42], the SIRT-FBP method from [43] and the filter from [25], which we refer to as ERM-FBP method, as well as for our proposed filter functions $A_{L,D}^*$ and $\hat{A}_{L,D}^\varepsilon$, the error continues to decrease for all chosen N_φ . The MR-FBP method from [42] and our simple workaround $A_{L,D}^\varepsilon$ give decent results for small noise, but also show an increasing error for increasing N_φ already for moderate noise. In total, we observe that our filters $A_{L,D}^*$ and $\hat{A}_{L,D}^\varepsilon$ perform best, which was expected since the MSE is the discrete counterpart of the objective we optimized in the derivation of our filter function. Moreover, we see that the difference in performance increases with increasing $p_{\text{noise}} \in \{0.05, 0.1, 0.15\}$. For illustration, Figure 6 shows $A_{L,D}^*$, $A_{L,D}^\varepsilon$ and $\hat{A}_{L,D}^\varepsilon$ for the noisy Shepp-Logan sinogram with $N_\varphi \in \{90, 180, 360\}$ and $p_{\text{noise}} \in \{0.05, 0.1, 0.15\}$.

In terms of SSIM, we observe that for the classical filters, our simple workaround $A_{L,D}^\varepsilon$, the MR-FBP and MR-FBP_{GM} method as well as the SIRT-FBP method the reconstruction quality decreases with increasing N_φ . For the ERM-FBP method as well as our optimized filter $A_{L,D}^*$, however, the SSIM increases with increasing N_φ . For its variant $\hat{A}_{L,D}^\varepsilon$ the behaviour of the SSIM depends on the amount of noise. For small noise, the reconstruction quality increases with increasing N_φ , but for moderate and high noise, the SSIM begins to decrease again for large values of N_φ . In total, we observe that now ERM-FBP performs best, closely followed by our filters $A_{L,D}^*$ and $\hat{A}_{L,D}^\varepsilon$. Note, however, that ERM-FBP requires access to training data comprising noise-free sinograms and noise samples, whereas our filters can be used off the shelf. As proposed in [25], in our simulations the ERM-FBP filter was trained on 10000 random ellipse phantoms and different realizations of noise. For illustration, Figure 5 shows the different reconstructions from noisy Radon data with $p_{\text{noise}} = 0.1$ and $N_\varphi = 360$.

5.2. Modified Shepp-Logan phantom. To study the generalization abilities of the filters, in our second set of numerical simulations we consider a modified version of the Shepp-Logan phantom comprising smoother as well as rougher components, see Figure 2 (c)-(d). To be more precise, we replace the characteristic function

$$\chi_{B_1(0)}(x, y) = \begin{cases} 1 & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{if } x^2 + y^2 > 1 \end{cases}$$

in the definition of the classical Shepp-Logan phantom by

$$p_\nu(x, y) = \begin{cases} (1 - x^2 - y^2)^\nu & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{if } x^2 + y^2 > 1 \end{cases}$$

with smoothness parameter $\nu = 1.5$ and add two rectangles of different sizes. Thereon, we rescale the resulting attenuation function $f_{\text{MSL}} \in L^2(\mathbb{R}^2)$ with $\text{supp}(f_{\text{MSL}}) \subseteq B_1(0)$ so that its Radon transform $\mathcal{R}f_{\text{MSL}}$ has the same arithmetic mean as $\mathcal{R}f_{\text{SL}}$, i.e., $m_{\mathcal{R}f_{\text{MSL}}} = m_{\mathcal{R}f_{\text{SL}}}$.

To investigate the effect of underestimating the amount of noise, the noisy measurements $g_D^\varepsilon(i, j)$ are simulated by adding white Gaussian noise with variance ε^2 to $\mathcal{R}f_{\text{MSL}}(s_i, \varphi_j)$, where ε is 20% larger than expected, i.e., $\varepsilon = 1.2 \cdot p_{\text{noise}} m_{\mathcal{R}f_{\text{MSL}}}$ with $p_{\text{noise}} \in \{0.05, 0.1, 0.15\}$ and we omit the factor 1.2 in the definition of our proposed filters $A_{L,D}^*$, $\hat{A}_{L,D}^\varepsilon$ and $A_{L,D}^\varepsilon$.

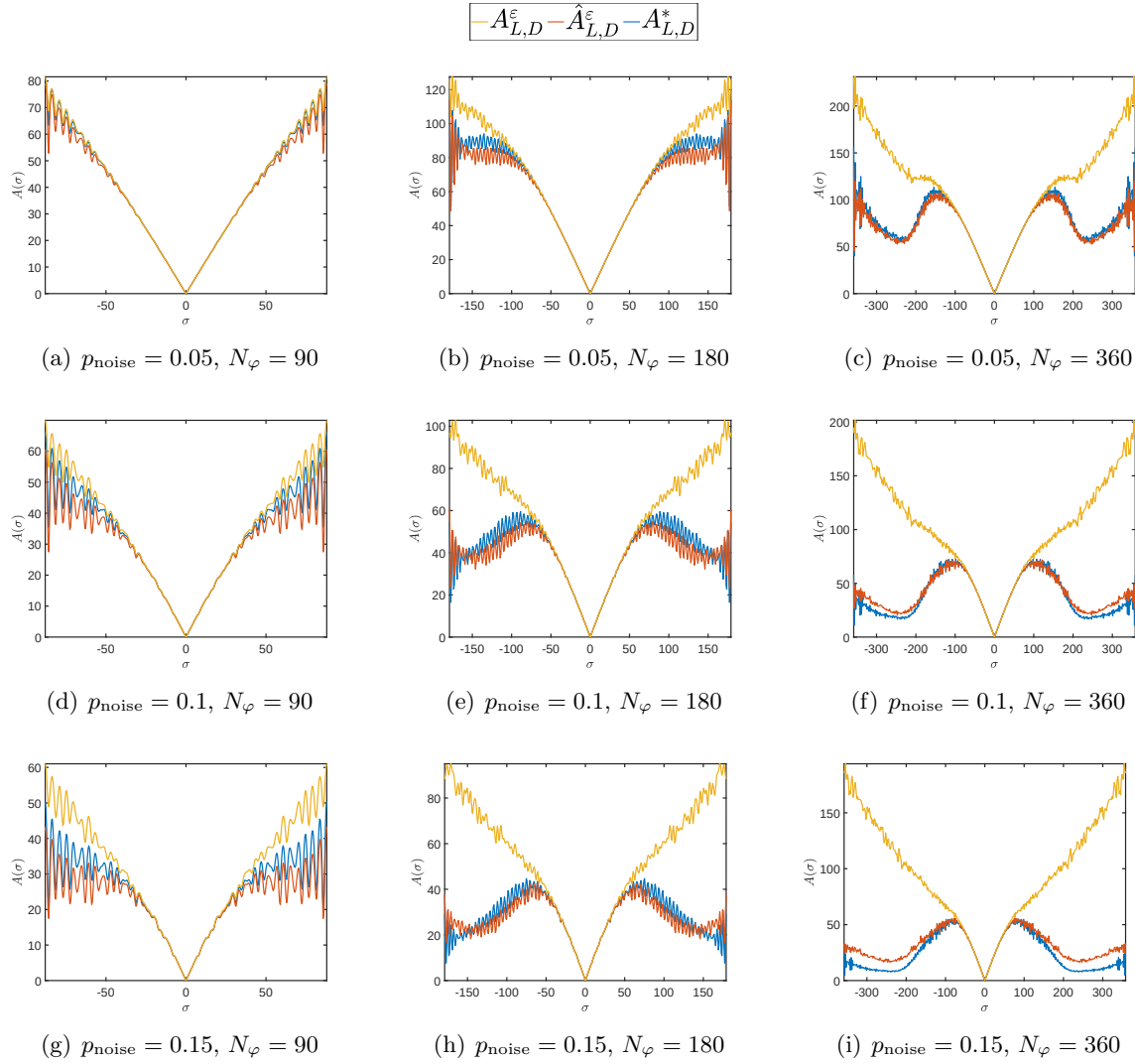


Figure 6. Plots of the optimized filter functions $A_{L,D}^*$, $A_{L,D}^\varepsilon$ and $\hat{A}_{L,D}^\varepsilon$ for the noisy Shepp-Logan sinogram with $N_\varphi \in \{90, 180, 360\}$ and $p_{\text{noise}} \in \{0.05, 0.1, 0.15\}$.

The results of our numerical experiments are depicted in Figure 7, where we use the same filter parameters as in the first set of experiments for the Shepp-Logan phantom in Section 5.1. In terms of MSE, we observe that for the classical filters, the MR-FBP method as well as the MR-FBP_{GM} method the error increases with increasing N_φ . The same is true for our simple workaround $A_{L,D}^\varepsilon$, which shows that this choice cannot deal with underestimated noise levels. For $\hat{A}_{L,D}^\varepsilon$ the behaviour depends on the amount of noise. For small noise, the MSE increases with increasing N_φ , but for moderate and high noise, the MSE begins to increase again for large values of N_φ . The SIRT-FBP method only slightly decreases with increasing N_φ , whereas ERM-FBP and $A_{L,D}^*$ show the best results with $A_{L,D}^*$ outperforming ERM-FBP. However, the difference in performance decreases with increasing p_{noise} .

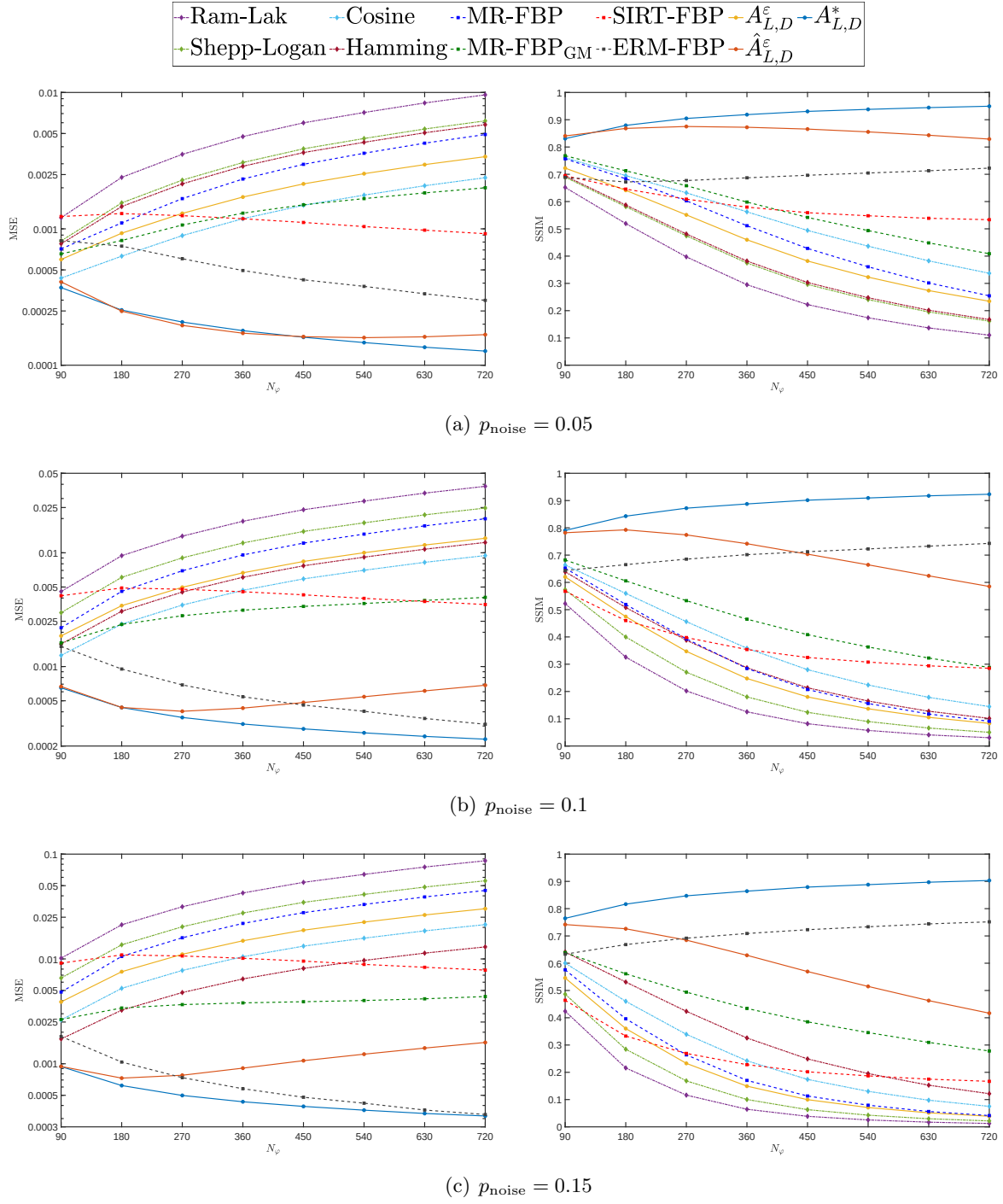


Figure 7. Plots of the MSE and SSIM of FBP reconstructions for the modified Shepp-Logan phantom.

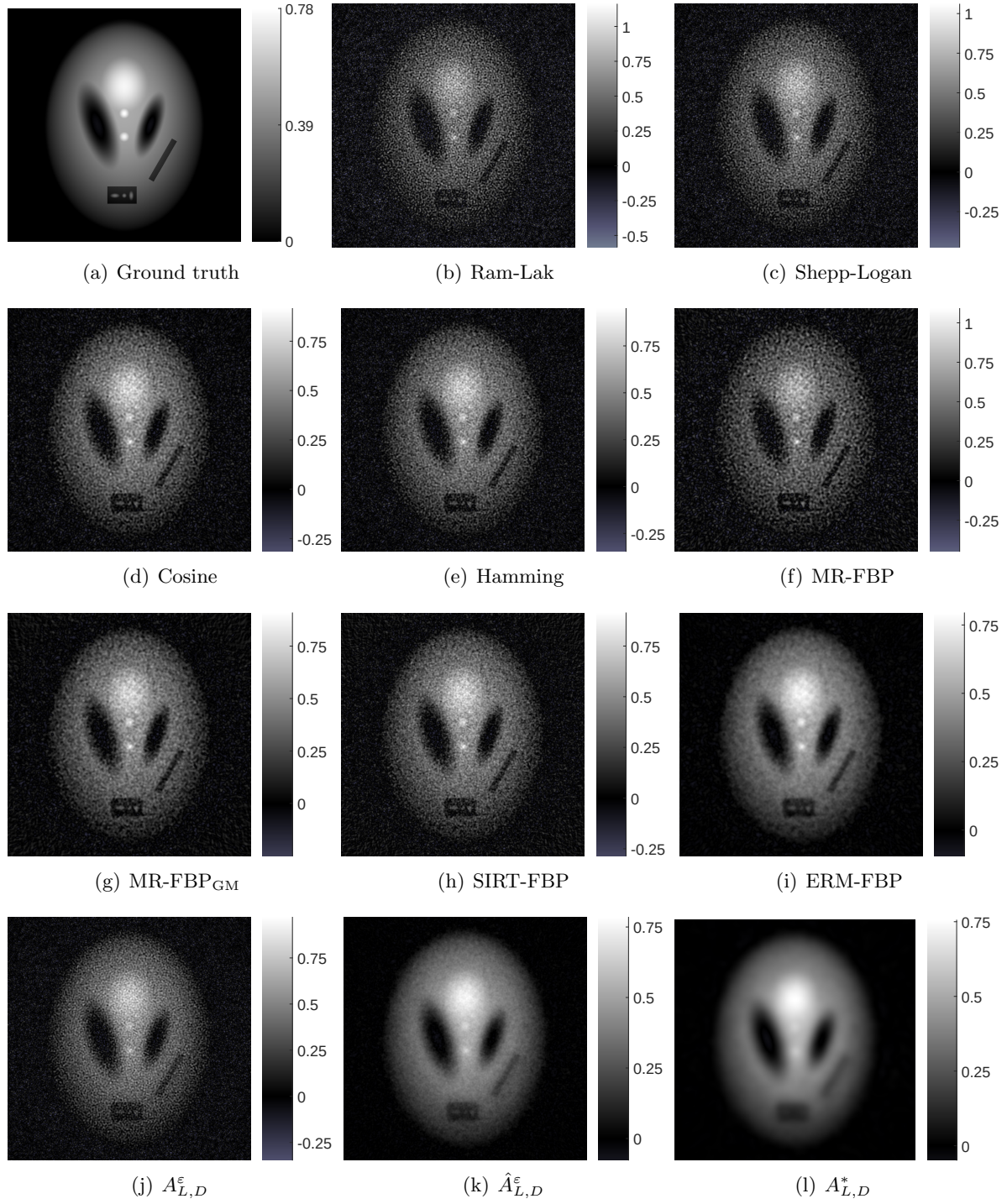


Figure 8. Reconstructions of modified Shepp-Logan phantom from noisy Radon data ($p_{\text{noise}} = 0.1$, $N_{\varphi} = 360$).

In terms of SSIM, we again observe that for the classical filters, our simple workaround $A_{L,D}^\varepsilon$, the MR-FBP and MR-FBP_{GM} method as well as the SIRT-FBP method the SSIM decreases with increasing N_φ . Also for $\hat{A}_{L,D}^\varepsilon$ the reconstruction quality deteriorates for large values of N_φ . In contrast to this, for the ERM-FBP method as well as our optimized filter $A_{L,D}^*$ the SSIM increases with increasing N_φ , where now $A_{L,D}^*$ performs best. This suggests that $A_{L,D}^*$ is able to compensate for an inaccurately estimated noise level, while ERM-FBP has problems in reconstructing target functions very different from the training samples.

For illustration, Figure 8 shows the different reconstructions of the modified Shepp-Logan phantom from noisy Radon data with $p_{\text{noise}} = 0.1$ and $N_\varphi = 360$. Although $A_{L,D}^*$ performs best in terms of MSE and SSIM, we observe that while suppressing the noise very well, it smooths out fine details in the reconstruction. As opposed to this, $\hat{A}_{L,D}^\varepsilon$ seems to provide a decent compromise between noise reduction and detail reconstruction, underpinning the potential of alternative denoising techniques for the noisy Radon samples $g_D^\varepsilon(i, j)$ in the definition of $A_{L,D}^\varepsilon$.

5.3. 2DeteCT. In our third and last set of numerical experiments, we evaluate the performance of the different filters on realistic Radon data. To this end, we make use of the 2DeteCT dataset [28] consisting of real X-ray CT measurements in three acquisition modes. Here, we focus on mode 1 yielding noisy low-dose fan-beam Radon data of slices of various test samples that produce similar image features as exhibited in medical abdominal CT scans. In addition to the 5000 slices of raw projection data, the dataset provides target reconstructions that serve as ground truth and were computed via an iterative reconstruction scheme to solve a non-negative least squares problem based on Nesterov accelerated gradient descent.

In our experiments, we first transform the provided fan-beam data into a parallel scanning geometry, which is needed for applying the discrete FBP reconstruction formula (5.1). Thereon, we subsample the sinograms with factor 2 in both the radial and angular variable, which serves as basis for the FBP reconstructions on an equidistant grid of 1024×1024 pixels.

Exemplary reconstructions are depicted in Figure 9, where we use the same filters as in our previous experiments. Note, however, that in this setting our optimized filter $A_{L,D}^*$ cannot be used as noise-free Radon data is not available. Also the ERM-FBP method from [25] cannot be applied due to the lack of training data consisting of noise-free sinograms and noise samples. For the other filters, involved parameters were optimized w.r.t. the mean MSE of the first 50 slices of the dataset. For the Hamming filter, the optimal parameter is $\beta = 1$ so that it agrees with the Ram-Lak filter. For $A_{L,D}^\varepsilon$, we use $\varepsilon = 0.045$ and for $\hat{A}_{L,D}^\varepsilon$ the optimal choice is $\varepsilon = 0.032$.

Visually, all filters yield comparable results that are close to the 2DeteCT target reconstruction, which is shown in Figure 9 (a). However, our filters $A_{L,D}^\varepsilon$ and $\hat{A}_{L,D}^\varepsilon$ seem to produce reconstruction with the best noise reduction. Also in terms of MSE, our filters show the best performance with $9.0792 \cdot 10^{-6}$ for $A_{L,D}^\varepsilon$ and $9.1616 \cdot 10^{-6}$ for $\hat{A}_{L,D}^\varepsilon$. The next best performance is achieved by the Shepp-Logan filter with an MSE of $9.1803 \cdot 10^{-6}$, followed by the Cosine filter with $9.5116 \cdot 10^{-6}$ and the MR-FBP_{GM} method with $9.8905 \cdot 10^{-6}$. The worst performance is achieved by the Ram-Lak filter with an MSE of $1.0703 \cdot 10^{-5}$, followed by the SIRT-FBP method with $1.0742 \cdot 10^{-5}$ and lastly the MR-FBP method with an error of $1.2231 \cdot 10^{-5}$.

Remarkably, although the measurement noise is not expected to be Gaussian, our simple

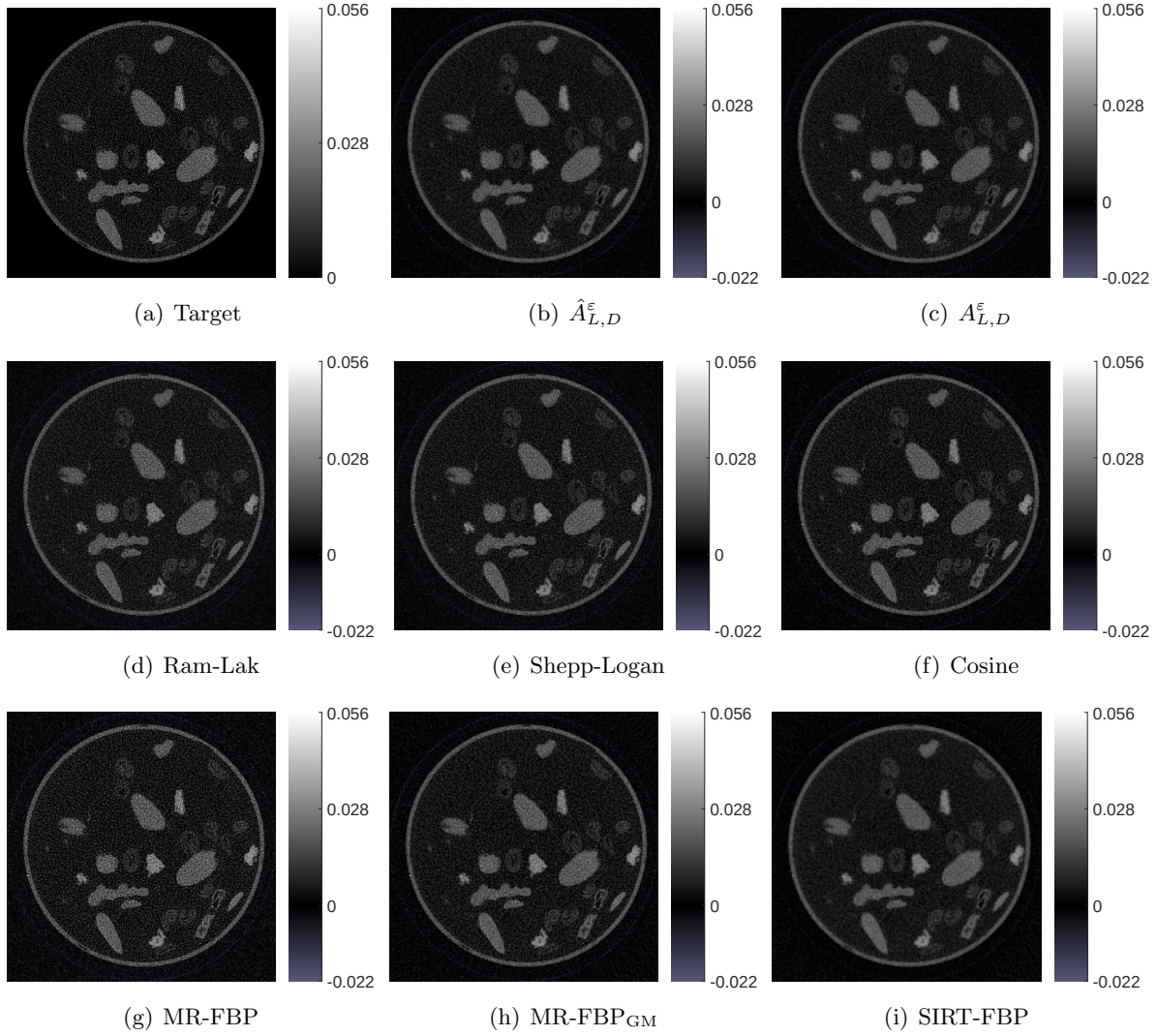


Figure 9. Reconstructions of 2DeteCT slice 1942 (mode 1).

workaround $A_{L,D}^\varepsilon$ performs best, closely followed by $\hat{A}_{L,D}^\varepsilon$, which are explicitly designed for Gaussian noise. We expect that the performance of $\hat{A}_{L,D}^\varepsilon$ can be further improved by utilizing a denoising scheme specifically tailored to the true noise in the Radon measurements.

6. Discussion and Outlook. This work focuses on improving the reconstruction quality of the approximate filtered back projection reconstruction f_L^ε in computerized tomography by optimizing the applied low-pass filter A_L with bandwidth $L > 0$. The definition of our filter function is motivated by investigations of the noise behaviour of measured CT data, resulting in an approximate Gaussian white noise model. Based on this, our proposed filter minimizes the expected squared error $\mathbb{E}(\|f_L^\varepsilon - f\|_{L^2(\mathbb{R}^2)}^2)$ given infinite noisy measurement data with noiselevel $\varepsilon > 0$ for a fixed target attenuation function f . Our resulting filter A_L^* depends on

the noise approximation h_a , noiselevel ε and Radon data $\mathcal{R}f$. We extend this filter to handle finitely many measurements corrupted by standard discrete additive Gaussian white noise in a parallel scanning geometry. The discretized filter $A_{L,D}^*$ then depends on the variance of the discrete Gaussian white noise, the discretization parameters and the true Radon data $\mathcal{R}f$ of the target function f . To circumvent the dependency on $\mathcal{R}f$, which is rarely known in practice, we propose two adaptations. The first replaces $\mathcal{R}f$ by the measurements g_D^ε , resulting in the filter $A_{L,D}^\varepsilon$. The second replaces $\mathcal{R}f$ by Wiener filtered data $\hat{g}_D^\varepsilon$, resulting in the filter $\hat{A}_{L,D}^\varepsilon$.

In addition to our theoretical investigations, we conduct extensive numerical experiments to evaluate the performance of our optimized filter function and its adaptations compared to both standard and proposed filters from the literature. These experiments are performed on the conventional Shepp-Logan phantom, a modified Shepp-Logan phantom to examine the impact of underestimating noise, and the 2DeteCT dataset to test the reconstruction quality for real X-ray CT measurements. The results reveal that our proposed filter functions $A_{L,D}^*$ and $\hat{A}_{L,D}^\varepsilon$ significantly surpass all dataset-independent filter functions in terms of MSE and SSIM and yield reconstruction qualities comparable to the dataset-dependent ERM-FBP filter. Moreover, our proposed filters are easy to implement, similar to standard filter functions, thus providing an out-of-the-box solution with theoretical justification. Consequently, our filters bridge the gap in the literature between filters without closed-form representation – requiring iterative schemes in advance or minimization problems for each measurement – and filters requiring a training dataset, which is often challenging to obtain in practical applications.

We observe that our adaptation $\hat{A}_{L,D}^\varepsilon$ shows promising results in particular for the real 2DeteCT data, where $A_{L,D}^*$ is not applicable, and in scenarios with inaccurate noise level estimation, offering a good balance between noise reduction and the reconstruction of finer details. However, our chosen denoising technique for calculating $\hat{A}_{L,D}^\varepsilon$ might not be optimal for real-world data. Therefore, we propose exploring different denoising strategies in future works, where data-driven approaches seem promising, as they can learn the true noise distribution from the measured data. Besides using denoised measured data, one could also employ deep learning to overcome the dependence of the true Radon data $\mathcal{R}f$. This approach also allows for including penalty terms in the loss function to incorporate prior knowledge. Beyond these natural adaptations of our proposed filters, future work could also build on our theoretical investigations and modify the underlying minimization problem by, for instance, changing the loss function, adding additional penalty terms, or directly optimizing in the discrete setting.

In summary, our analytical derivations provide a solid foundation for future investigations, where we see particular potential in combining our theoretical results with neural networks to overcome the dependence of the filter on the target function and further improve the reconstruction quality.

REFERENCES

- [1] S. AZIZNEJAD AND J. FAGEOT, *Wavelet analysis of the Besov regularity of Lévy white noise*, Electronic Journal of Probability, 25 (2020), pp. 1–38, <https://doi.org/10.1214/20-EJP554>.
- [2] Z. A. BALOGH AND B. JANOS KIS, *Comparison of CT noise reduction performances with deep learning-based, conventional, and combined denoising algorithms*, Medical Engineering & Physics, 109 (2022), p. 103897, <https://doi.org/10.1016/j.medengphy.2022.103897>.

- [3] M. BECKMANN AND A. ISKE, *Error estimates for filtered back projection*, in IEEE International Conference on Sampling Theory and Applications (SampTA), 2015, pp. 553–557, <https://doi.org/10.1109/SAMPTA.2015.7148952>.
- [4] M. BECKMANN AND A. ISKE, *Error estimates and convergence rates for filtered back projection*, Mathematics of Computation, 88 (2019), pp. 801–835, <https://doi.org/10.1090/mcom/3343>.
- [5] M. BECKMANN AND A. ISKE, *Saturation rates of filtered back projection approximations*, Calcolo, 57 (2020), p. 12, <https://doi.org/10.1007/s10092-020-00360-y>.
- [6] M. BECKMANN, P. MAASS, AND J. NICKEL, *Error analysis for filtered back projection reconstructions in Besov spaces*, Inverse Problems, 37 (2021), p. 014002, <https://doi.org/10.1088/1361-6420/aba5ee>.
- [7] J. BERGH AND J. LÖFSTRÖM, *Interpolation Spaces. An Introduction*, Grundlehren der mathematischen Wissenschaften, Wiley, 1976, <https://doi.org/10.1007/978-3-642-66451-9>.
- [8] E. BIBBONA, G. PANFILO, AND P. TAVELLA, *The Ornstein–Uhlenbeck process as a model of a low pass filtered white noise*, Metrologia, 45 (2008), p. S117, <https://doi.org/10.1088/0026-1394/45/6/S17>.
- [9] P. BILLINGSLEY, *Probability and Measure*, Wiley, 1995.
- [10] T. M. BUZUG, *Computed Tomography*, Springer, 2010.
- [11] R. C. DALANG AND T. HUMEAU, *Lévy processes and Lévy white noise as tempered distributions*, The Annals of Probability, 45 (2017), pp. 4389–4418, <https://doi.org/10.1214/16-AOP1168>.
- [12] M. DIWAKAR AND M. KUMAR, *A review on CT image noise and its denoising*, Biomedical Signal Processing and Control, 42 (2018), pp. 73–88, <https://doi.org/10.1016/j.bspc.2018.01.010>.
- [13] J. L. DOOB, *The Brownian movement and stochastic equations*, Annals of Mathematics, 43 (1942), pp. 351–369.
- [14] C. EPSTEIN, *Introduction to the Mathematics of Medical Imaging*, SIAM, 2 ed., 2008.
- [15] J. FAGEOT, *On tempered discrete and Lévy white noises*, 2022, <https://arxiv.org/abs/2201.00797>.
- [16] J. FAGEOT, A. FALLAH, AND M. UNSER, *Multidimensional Lévy white noise in weighted Besov spaces*, Stochastic Processes and their Applications, 127 (2017), pp. 1599–1621, <https://doi.org/10.1016/j.spa.2016.08.011>.
- [17] J. FAGEOT, M. UNSER, AND J. P. WARD, *On the Besov regularity of periodic Lévy noises*, Applied and Computational Harmonic Analysis, 42 (2017), pp. 21–36, <https://doi.org/10.1016/j.acha.2015.07.001>.
- [18] T. FEEMAN, *The Mathematics of Medical Imaging: A Beginner’s Guide*, Springer, 2015.
- [19] L. FU, T. C. LEE, S. M. KIM, A. M. ALESSIO, P. E. KINAHAN, Z. CHANG, K. SAUER, M. K. KALRA, AND B. DE MAN, *Comparison between pre-log and post-log statistical models in ultra-low-dose CT reconstruction*, IEEE Transactions on Medical Imaging, 36 (2017), pp. 3290–3303, <https://doi.org/10.1109/TMI.2016.2627004>.
- [20] K. GARDINER, *Stochastic Methods. A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics, Springer, 2009.
- [21] I. M. GEL’FAND AND N. Y. VILENKIN, *Generalized Functions. Volume 4: Applications of Harmonic Analysis*, Academic Press, 1964.
- [22] G. HERMAN, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Advances in Pattern Recognition, Springer, 2 ed., 2009.
- [23] S. HORBELT, M. LIEBLING, AND M. A. UNSER, *Filter design for filtered back-projection guided by the interpolation model*, in SPIE Medical Imaging, 2002, pp. 806–813, <https://doi.org/10.1117/12.467227>.
- [24] S. HOTTOVY, A. MCDANIEL, AND J. WEHR, *A small delay and correlation time limit of stochastic differential delay equations with state-dependent colored noise*, Journal of Statistical Physics, 175 (2019), pp. 19–46, <https://doi.org/10.1007/s10955-019-02242-2>.
- [25] S. KABRI, A. AURAS, D. RICCIO, H. BAUERMEISTER, M. BENNING, M. MOELLER, AND M. BURGER, *Convergent data-driven regularizations for CT reconstruction*, Communications on Applied Mathematics and Computation, (2024), <https://doi.org/10.1007/s42967-023-00333-2>.
- [26] A. C. KAK AND M. SLANEY, *Principles of Computerized Tomographic Imaging*, no. 33 in Classics in Applied Mathematics, SIAM, 2001.
- [27] H. KEKKONEN, M. LASSAS, AND S. SILTANEN, *Analysis of regularized inversion of data corrupted by white Gaussian noise*, Inverse Problems, 30 (2014), p. 045009, <https://doi.org/10.1088/0266-5611/30/4/045009>.
- [28] M. B. KISS, S. B. COBAN, K. J. BATENBURG, T. VAN LEEUWEN, AND F. LUCKA, *2DeteCT - a large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning*, Scientific

- Data, 10 (2023), <https://doi.org/10.1038/s41597-023-02484-6>.
- [29] T. KUHN, H. G. LEOPOLD, W. SICKEL, AND L. SKRZYPCZAK, *Entropy numbers of embeddings of weighted Besov spaces*, Constructive Approximation, 23 (2005), pp. 61–77, <https://doi.org/10.1007/s00365-005-0594-9>.
 - [30] M. J. LAGERWERF, A. A. HENDRIKSEN, J.-W. BUURLAGE, AND K. J. BATENBURG, *Noise2Filter: fast, self-supervised learning and real-time reconstruction for 3D computed tomography*, Machine Learning: Science and Technology, 2 (2020), p. 015012, <https://doi.org/10.1088/2632-2153/abbd4d>.
 - [31] T. LI, X. LI, J. WANG, J. WEN, H. LU, J. HSIEH, AND Z. LIANG, *Nonlinear sinogram smoothing for low-dose X-ray CT*, IEEE Transactions on Nuclear Science, 51 (2004), pp. 2505–2513, <https://doi.org/10.1109/TNS.2004.834824>.
 - [32] Z. LIAO, S. HU, M. LI, , AND W. CHEN, *Noise estimation for single-slice sinogram of low-dose X-ray computed tomography using homogenous patch*, Mathematical Problems in Engineering, 2012 (2012), p. 696212, <https://doi.org/10.1155/2012/696212>.
 - [33] H. LU, I. T. HSIAO, X. LI, AND Z. LIANG, *Noise properties of low-dose CT projections and noise treatment by scale transformations*, in IEEE Nuclear Science Symposium Conference Record, 2001, pp. 1662–1666, <https://doi.org/10.1109/NSSMIC.2001.1008660>.
 - [34] M. MEYRIES AND M. VERAAR, *Sharp embedding results for spaces of smooth functions with power weights*, Studia Mathematica, 208 (2012), pp. 257–293, <https://doi.org/10.4064/sm208-3-5>.
 - [35] F. NATTERER, *The Mathematics of Computerized Tomography*, SIAM, 2001.
 - [36] F. NATTERER AND F. WÜBBELING, *Mathematical methods in image reconstruction*, SIAM Monographs on Mathematical Modeling and Computation, SIAM, 2001.
 - [37] A. A. OMER, O. I. HASSAN, A. I. AHMED, AND A. ABDELRAHMAN, *Denoising CT images using median based filters: a review*, in International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 2018, pp. 1–6, <https://doi.org/10.1109/ICCCEEE.2018.8515829>.
 - [38] 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), p. 123009, <https://doi.org/10.1088/0266-5611/25/12/123009>.
 - [39] A. PAPOULIS AND S. U. PILLAI, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 2002.
 - [40] R. PAYLEY, N. WIENER, AND A. ZYGMUND, *Note on random functions*, Mathematische Zeitschrift, 37 (1933), pp. 647–668.
 - [41] D. M. PELT AND K. J. BATENBURG, *Fast tomographic reconstruction from limited data using artificial neural networks*, IEEE Transactions on Image Processing, 22 (2013), pp. 5238–5251, <https://doi.org/10.1109/TIP.2013.2283142>.
 - [42] D. M. PELT AND K. J. BATENBURG, *Improving filtered backprojection reconstruction by data-dependent filtering*, IEEE Transactions on Image Processing, 23 (2014), pp. 4750–4762, <https://doi.org/10.1109/TIP.2014.2341971>.
 - [43] D. M. PELT AND K. J. BATENBURG, *Accurately approximating algebraic tomographic reconstruction by filtered backprojection*, in International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2015, pp. 158–161.
 - [44] 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), p. 17, <https://doi.org/10.1186/s40679-016-0033-y>.
 - [45] J. RADON, *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten*, Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften, 69 (1917), pp. 262–277.
 - [46] A. RAVISHANKAR, S. ANUSHA, H. K. AKSHATHA, A. RAJ, S. JAHNAVI, AND J. MADHURA, *A survey on noise reduction techniques in medical images*, in International Conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 385–389, <https://doi.org/10.1109/ICECA.2017.8203711>.
 - [47] R. T. SADIA, J. CHEN, AND J. ZHANG, *CT image denoising methods for image quality improvement and radiation dose reduction*, Journal of Applied Clinical Medical Physics, 25 (2024), p. e14270, <https://doi.org/10.1002/acm2.14270>.
 - [48] L. A. SHEPP AND B. F. LOGAN, *The Fourier reconstruction of a head section*, IEEE Transactions on

- Nuclear Science, 21 (1974), pp. 21–43, <https://doi.org/10.1109/TNS.1974.6499235>.
- [49] H. SHI AND S. LUO, *A novel scheme to design the filter for CT reconstruction using FBP algorithm*, BioMedical Engineering OnLine, 12 (2013), p. 50, <https://doi.org/10.1186/1475-925X-12-50>.
- [50] J. SHTOK, M. ELAD, AND M. ZIBULEVSKY, *Adaptive filtered-back-projection for computed tomography*, in IEEE Convention of Electrical and Electronics Engineers in Israel, 2008, pp. 528–532, <https://doi.org/10.1109/EEEI.2008.4736585>.
- [51] K. SOBCZYK, *Stochastic Differential Equations. With Applications to Physics and Engineering*, Mathematics and Its Applications (East European Series), Kluwer Academic Publishers, 1991.
- [52] C. SYBEN, B. STIMPEL, K. BREININGER, T. WÜRFL, R. FAHRIG, A. DÖRFLER, AND A. MAIER, *Precision learning: Reconstruction filter kernel discretization*, in International Conference on Image Formation in X-ray Computed Tomography, 2018, pp. 386–390.
- [53] H. TRIEBEL, *Function Spaces and Wavelets on Domains*, EMS Tracts in Mathematics, European Mathematical Society Publishing House, 2008.
- [54] M. TSAGRIS, C. BENEKI, AND H. HASSANI, *On the folded normal distribution*, Mathematics, 2 (2014), pp. 12–28, <https://doi.org/10.3390/math2010012>.
- [55] G. E. UHLENBECK AND L. S. ORNSTEIN, *On the theory of the Brownian motion*, Physical Review, 36 (1930), pp. 823–841, <https://doi.org/10.1103/PhysRev.36.823>.
- [56] M. VERAAR, *Regularity of Gaussian white noise on the d-dimensional torus*, 2010, <https://arxiv.org/abs/1010.6219>.
- [57] J. WANG, H. LU, Z. LIANG, D. EREMINA, G. ZHANG, S. WANG, J. CHEN, AND J. MANZIONE, *An experimental study on the noise properties of X-ray CT sinogram data in Radon space*, Physics in Medicine & Biology, 53 (2008), pp. 3327–3341, <https://doi.org/10.1088/0031-9155/53/12/018>.
- [58] Z. WANG, A. C. BOVIK, H. R. SHEIKH, AND E. P. SIMONCELLI, *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–612, <https://doi.org/10.1109/TIP.2003.819861>.
- [59] B. R. WHITING, P. MASSOUMZADEH, O. A. EARL, J. A. O’SULLIVAN, D. L. SNYDER, AND J. F. WILLIAMSON, *Properties of preprocessed sinogram data in X-ray computed tomography*, Medical Physics, 33 (2006), pp. 3290–3303, <https://doi.org/10.1118/1.2230762>.
- [60] J. XU, C. SUN, Y. HUANG, AND X. HUANG, *Residual neural network for filter kernel design in filtered back-projection for CT image reconstruction*, in German Workshop on Medical Image Computing, 2021, pp. 164–169.