Regularization with Optimal Space-Time Priors

Regularization with optimal space-time priors

Tatiana A. Bubba*, Tommi Heikkilä†, Demetrio Labate‡, and Luca Ratti§

Abstract. We propose a variational regularization approach based on a multiscale representation called cylindrical shearlets aimed at dynamic imaging problems, especially dynamic tomography. The intuitive idea of our approach is to integrate a sequence of separable static problems in the mismatch term of the cost function, while the regularization term handles the nonstationary target as a spatio-temporal object. This approach is motivated by the fact that cylindrical shearlets provide (nearly) optimally sparse approximations on an idealized class of functions modeling spatio-temportal data and the numerical observation that they provide highly sparse approximations even for more general spatio-temporal image sequences found in dynamic tomography applications. To formulate our regularization model, we introduce cylindrical shearlet smoothness spaces, which are instrumental for defining suitable embeddings in functional spaces. We prove that the proposed regularization strategy is well-defined, and the minimization problem has a unique solution (for $p > 1$). Furthermore, we provide convergence rates (in terms of the symmetric Bregman distance) under deterministic and random noise conditions, within the context of statistical inverse learning. We numerically validate our theoretical results using both simulated and measured dynamic tomography data, showing that our approach leads to an efficient and robust reconstruction strategy.

Key words. Cylindrical shearlet, dynamic tomography, regularization, statistical inverse learning, smoothness spaces, cylindrical cartoon-like functions

MSC codes. 47A52, 42C40, 65J20, 65J22

1. Introduction.

Traditional X-ray tomography is typically used to determine the interior structure of an unknown (static) object measuring the total attenuation of X-rays from many orientations. This inverse problem is known to become more severely ill-posed as the number of measurements (i.e., *projections*) decreases. Regularization theory of inverse problems is widely used to overcome errors and limitations caused by very few measurements or too coarse modelling. The underpinning idea is that in order to obtain realistic approximate solutions from limited and noisy data, suitable prior information needs to be built into the problem by means, e.g., of a penalty term, leading to a variational regularization formulation of the reconstruction problem. A wide range of different priors have been proposed to enforce desired structures and properties on the reconstructions. Often, in limited data problems, total variation (TV), which enforces gradient sparsity, is used as a simple, yet effective prior [24, 59, 66, 94, 93, 97], as well as its variants, e.g., total generalized variation [79]. Alternative strategies have employed wavelets [51, 58, 68, 78, 85, 38], curvelets [17, 36], shearlets [11, 25, 47, 86, 101] and combination of those [39] as priors.

However, noise and limited data are not the only cause of errors in tomographic reconstructions. In many applications, the target (or parts of it) is dynamic, i.e., it can change between the recording of two consecutive projection images. As a consequence, movement or

*Department of Mathematics and Computer Science, University of Ferrara, Italy (tatiana.bubba@unife.it).
†School of Engineering Sciences, LUT University, Finland (tommi.heikkila@lut.fi).
‡Department of Mathematics, University of Houston, USA (dlabate@math.uh.edu).
§Department of Mathematics, University of Bologna, Italy (luca.ratti5@unibo.it).

changes of the target can cause reconstruction artefacts [71, 74], unless carefully accounted [3] or compensated for [27, 57, 48], either prior [69], during [100, 77, 13, 22, 70, 65] or after the reconstruction process [87, 43].

Dynamic tomography, that is, the study of tomographic image reconstructions of non-stationary targets, is the motivating example of this work. An extensive review of dynamic imaging problems (including tomography) is provided, e.g., in [91, 52, 53]. Adopting the classical viewpoint of regularization theory of static inverse problems, we are interested in establishing a variational regularization framework where both spatial structure and temporal evolution of the solution are restricted by some predetermined model or representation system. As compared to those approaches aiming at accurately modelling the uncertainties coming from the non-stationary target in the forward operator, variational approaches turn modelling into a quest for a suitable prior for the penalty term. For example, changes over time can be controlled by penalizing the total variation in time [82, 83], possibly coupled with optical flow [13], separating the dynamic parts from the static background via low-rank constraints [37, 81, 105, 1], Bayesian estimation [33, 75, 64], Kalman filters [89, 49] or by sparse multivariate systems [96, 9, 8, 60].

Here we introduce a variational regularization approach based on *cylindrical shearlets* [31], a recently proposed direction-aware representation system aimed at efficiently approximating spatio-temporal data, that is, temporal sequences (or videos) of 2D images. Compared to wavelets or even classical shearlets [62], the geometry of cylindrical shearlets is better suited to the characteristics of spatio-temporal data, as such data are typically dominated by edge discontinuities in the spatial coordinates only. We are motivated by the recent observation that cylindrical shearlets provide nearly optimally sparse approximations for a class of functions that is useful to model spatio-temporal image sequences and videos [31, 8] where they provide $\mathcal{N}$-term approximations with error rate of order $O(\mathcal{N}^{-2})$. In contrast, conventional 3D wavelets and 3D shearlets only achieve approximation rates that are $O(\mathcal{N}^{-\frac{1}{2}})$ [28] and $O(\mathcal{N}^{-1})$ [45], respectively.

To formulate our regularization model, we also need to define the corresponding *cylindrical shearlet smoothness spaces* (in subsection 2.3) by extending the work of one of the authors [63]. This definition is based on the theory of decomposition spaces introduced by K. Gröbner and H. G. Feichtinger [35, 34] and later adapted into the theory of smoothness spaces by L. Borup and M. Nielsen [5]. Building cylindrical shearlet smoothness spaces is instrumental for deriving suitable embeddings in functional spaces. This allows us to express the norm of the target's cylindrical shearlet coefficients as the norm of the target in a suitable cylindrical shearlet smoothness space, similarly to the classical case of wavelet coefficients and Besov norms [26].

This sets the foundation for stating our regularised model for dynamic X-ray tomography:

$$\operatorname*{argmin}_{f\in X}\left\{\frac{1}{2}\,\|\mathcal{A}f-g^{\delta}\|_{Y}^{2}+\frac{\alpha}{p}\|\,\mathcal{SH}(f)\|_{\ell^{p}}^{p}\right\},$$

where $\mathcal{A}$ is the time-dependent X-ray transform, $\mathcal{SH}$ the cylindrical shearlet transform, X the cylindrical shearlet smoothness space, Y a Bochner space (integrating over time), and we consider $1<p<2$ (see section 3 for all the details). This formulation is certainly reminiscent of a classical regularized formulation of static imaging problems. However, there is a twist in our model: while the mismatch term essentially integrates a sequence of separable, static

tomographic problems, the regularization term sees the non-stationary target as a spatio-temporal object, tying inextricably the temporal dimension to the spatial ones.

This approach is seemingly straightforward, both in its formalization (constrained analysis formulation) and conceptualization (treating time as an integral dimension of the target). However, it proves to be very rich to explore and analyze from a theoretical viewpoint and surprisingly effective in its practical numerical implementation. Indeed, we are able to prove that this regularization strategy is well-defined, and the solution of the minimization problem exists and is unique for $p > 1$ (see subsection 3.2). Furthermore, we provide convergence rates (in terms of the symmetric Bregman distance) under deterministic and random noise conditions (see subsection 3.3). These results build upon the vast literature on regularization of inverse problems, in particular on the derivation of convergence rates, in various error metrics. The most relevant to our framework are [67], which considers $p \in [0, 2]$, [41] where $p \in [1, 2]$, [15, 40, 42, 50], which focus on $p = 1$, and [56, 107] where Tikhonov regularization in Besov spaces is deeply analysed, both in the context of deterministic and random noise.

We remark that the bounds we derive in Proposition 3.2 for the dynamic case are the same of those derived in the literature for the static one. Finally, these estimates are also analysed under the lens of the statistical inverse learning framework [2] (see subsection 3.4), extending to the dynamic setting the work in [7, 12], where convergence rates for p-homogeneous regularization functional are established, and numerically verified in the context of tomographic imaging with randomized projections. The latter estimates could be relevant for (dynamic) tomography scenarios where the exact projection angles can not be determined a priori by the operator. For example, in really fast tomography, where hardware limits rotation and measurement speed [88]; or [104], where gating is subject to irregular movement, making the chosen angles irregular too; or in cryo electron tomography where many copies of similar targets are taken at unknown orientation, which can be viewed as having many projections of just one target from unknown directions [80].

To show that the derived theoretical convergence rates are attained in practice, we include several numerical experiments using both simulated and measured dynamic tomographic data (see section 4). Our numerical experiments validate the theoretical approximation rates and show that our theoretical framework leads to a practical and robust reconstruction strategy. While the numerical experiments are restricted to a (2+time)-dimensional setup for computational feasibility, both the theoretical and computational frameworks are extendable to higher dimensions, the only limitation being the computational times. In the spirit of reproducible research, our numerical codes are made available on GitHub [10].

Finally, we remark that, while the results derived in this work are motivated by the case of dynamic computed tomography, our approach is very general since quite general assumptions are required on the dynamic operator, making it extendable to many dynamic imaging modality, e.g., magnetic resonance imaging (MRI) [81], magnetic particle imaging (MPI) [6] or positron emission tomography (PET) [90].

2. Cylindrical shearlets. Cylindrical shearlets were introduced to provide an efficient representation system of spatio-temporal data, assuming that edge discontinuities are mostly located along the spatial coordinates while along the last coordinate axis the geometry is notably simpler. We briefly recall below their construction from [31].

2.1. Construction. Similar to traditional bandlimited shearlets in $L^2(\mathbb{R}^3)$, the construction of cylindrical shearlets consists first in splitting the frequency domain $\mathbb{R}^3$ into symmetric pairs of conical sections, followed by partitioning into rectangular annuli before applying directional filtering. Hence, we first define the cones:

$$P_1 = \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 : \left|\frac{\xi_2}{\xi_1}\right| \leq 1\}, \quad P_2 = \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 : \left|\frac{\xi_1}{\xi_2}\right| \leq 1\}.$$

Then, a cone-adapted *cylindrical shearlet system* is a collection of functions

$$\Psi^{(\iota)} = \{\psi^{(\iota)}_{j,k,\boldsymbol{m}} : j \geq 0, |k| \leq 2^j, \boldsymbol{m} \in \mathbb{Z}^3\}, \tag{2.1}$$

defined separately for the two cones $P_\iota, \iota = 1, 2$. Denoting the Fourier transform of ψ by $\widehat{\psi}$, the cylindrical shearlet elements of (2.1) are defined in the Fourier domain as

$$\widehat{\psi}^{(\iota)}_{j,k,\boldsymbol{m}}(\xi) = \left|\det A_{(\iota)}\right|^{-\frac{j}{2}} W(2^{-2j}\xi) V_{(\iota)}\left(B^{-k}_{(\iota)} A^{-j}_{(\iota)} \xi\right) e^{2\pi i B^{-k}_{(\iota)} A^{-j}_{(\iota)} \xi \cdot \boldsymbol{m}}, \tag{2.2}$$

where the functions W and $V_{(\iota)}$ (defined below) provide the scale- and direction-based filtering and the matrices are

$$A_{(1)} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \; B_{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; A_{(2)} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \; B_{(2)} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.3}$$

Note that the shear matrices $B_{(\iota)}$ are transposed compared to the usual definitions since we work mostly in the Fourier domain.

To define our Fourier subbands, we start by choosing $\phi \in L^2(\mathbb{R}^3)$ such that $\widehat{\phi} \in \mathcal{C}^\infty_0$ (i.e., the class of infinitely differentialble functions with compact support), with $0 \leq \widehat{\phi}(\xi) \leq 1$ and

$$\widehat{\phi}(\xi) = 1 \text{ for } \xi \in \Big[-\frac{1}{16}, \frac{1}{16}\Big]^3 \text{ and } \widehat{\phi}(\xi) = 0 \text{ for } \xi \in \mathbb{R}^3 \setminus \Big[-\frac{1}{8}, \frac{1}{8}\Big]^3. \tag{2.4}$$

Setting $W^2(\xi) = \widehat{\phi}^2\Big(\frac{\xi}{4}\Big) - \widehat{\phi}^2\big(\xi\big)$, we obtain the following smooth partition of unity:

$$\widehat{\phi}^2(\xi) + \sum_{j=0}^{\infty} W^2\left(2^{-2j}\xi\right) = 1 \text{ for } \xi \in \mathbb{R}^3, \tag{2.5}$$

where each function $W\left(2^{-2j}\xi\right)$ is supported on a rectangular annulus $\left[-2^{2j-1}, 2^{2j-1}\right]^3 \setminus \left[-2^{2j-4}, 2^{2j-4}\right]^3$. This construction partitions the Fourier space into partially overlapping rectangular annuli outside a central cube. Next, to produce directional filtering, we let

$$V_{(1)}(\xi) = v\left(\tfrac{\xi_2}{\xi_1}\right), \; V_{(2)}(\xi) = v\left(\tfrac{\xi_1}{\xi_2}\right),$$

where $v \in \mathcal{C}^\infty(\mathbb{R})$ is such that $\operatorname{supp}(v) \subset [-1, 1]$ and

$$|v(z-1)|^2 + |v(z)|^2 + |v(z+1)|^2 = 1, \text{ for } z \in [-1, 1], \tag{2.6}$$

which implies that $\sum_{k\in\mathbb{Z}} |v(z+k)|^2 = 1$, for any $z \in \mathbb{R}$.

By the definitions of W and v, we see that each element $\widehat{\psi}^{(1)}_{j,k,\boldsymbol{m}}$ is contained in the set

$$Q^{(1)}_{j,k} = \left\{ \xi \in \left[-2^{2j-1}, 2^{2j-1}\right]^3 \setminus \left[-2^{2j-4}, 2^{2j-4}\right]^3 : \left| \frac{\xi_2}{\xi_1} - 2^{-j}k \right| \le 2^{-j} \right\}.$$

The set $Q^{(2)}_{j,k}$ is defined similarly by switching ξ_1 and ξ_2 around. Notice also that $Q_{j,k} \subset B^k A^j Q_{0,0}$ and for any j, k we have:

$$M_{\leftrightarrow} Q^{(1)}_{j,k} = Q^{(2)}_{j,k}, \qquad \text{where } M_{\leftrightarrow} = M_{\leftrightarrow}^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is also easy to choose an affine transformation M_0 such that $\det(M_0) = 1$ and $\left[-2^{-3}, 2^{-3}\right]^3 \subset M_0 Q^{(1)}_{0,0}$, to include the central low-frequency cube. For example, a simple shift $M_0\xi = \xi + (\frac{1}{3}, 0, 0)$ suffices. Notice that we do not always specify the index of the sets Q and matrices A and B when $\iota = 1$ if it is otherwise clear from context.

The wedge-shaped Fourier-domain support of the cylindrical shearlet system is illustrated side-by-side with the partition of traditional 3D shearlets in Figure 1, showing that cylindrical shearlets tile the Fourier domain differently compared to traditional 3D shearlets.

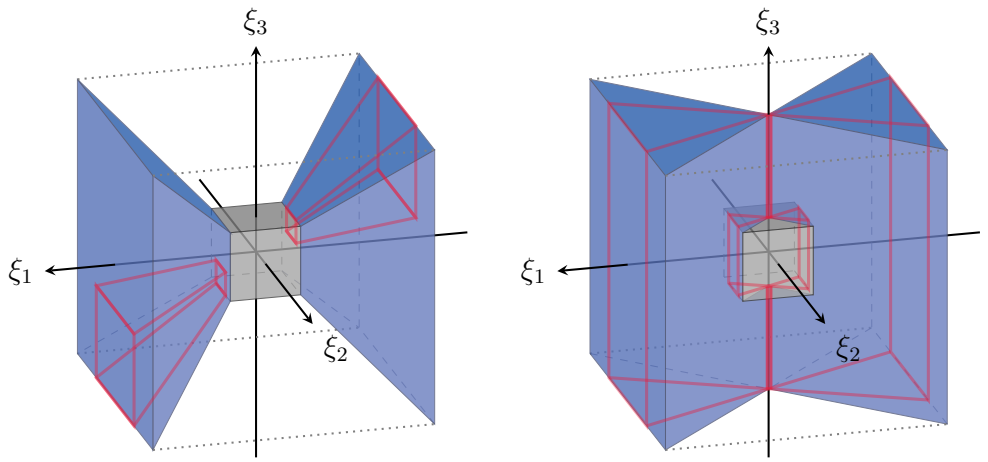

Figure 1. *Left: illustration of one pyramid-like section of the traditional 3D shearlet system in the Fourier domain and (in red) the outline of the Fourier support of a shearlet element $\psi^{(\iota)}_{j,k,\boldsymbol{m}}$, $\iota = 1, j = 0, k = (0,1)$ and $\boldsymbol{m} \in \mathbb{Z}^3$. Right: illustration of one cone-like section of the 3D cylindrical shearlet system in the Fourier domain and (in red) the outline of the Fourier support of a cylindrical shearlet element $\psi^{(\iota)}_{j,k,\boldsymbol{m}}$, $\iota = 1, j = 0, k = 0$ and $\boldsymbol{m} \in \mathbb{Z}^3$. In both illustrations, the low frequency cube is shown in gray.*

We obtain a smooth Parseval frame of cylindrical shearlets for $L^2(\mathbb{R}^3)$ by combining the two conical systems $\Psi^{(1)}$ and $\Psi^{(2)}$, given by (2.1), with a coarse scale system generated by

the integer translates of ϕ, given by (2.4). Additionally, to ensure that all elements of this combined system have good decay properties, we modify the elements of the shearlet system overlapping the boundaries of the regions P_1 and P_2 in the frequency domain. In brief, the shearlet elements from $\Psi^{(1)}$ and $\Psi^{(2)}$ overlapping the lines $\xi_1 = \pm\xi_2$ are merged together to ensure a smooth transition between the two pyramidal regions in the Fourier domain. We refer to [46] for details about the construction of these *boundary shearlets*. Hence, we obtain the *discrete cylindrical shearlet system*:

$$\Psi := \{\psi_{-1,\boldsymbol{m}} := \phi(x-\boldsymbol{m}) \ : \ \boldsymbol{m} \in \mathbb{Z}^3\} \cup \{\psi^{(\iota)}_{j,k,\boldsymbol{m}} : j \geq 0, |k| < 2^j, \boldsymbol{m} \in \mathbb{Z}^3, \iota = 1,2\}$$
$$\cup\{\tilde{\psi}^{(\iota)}_{j,k,\boldsymbol{m}} : j \geq 0, k = \pm 2^j, \boldsymbol{m} \in \mathbb{Z}^3, \iota = 1,2\}, \tag{2.7}$$

where the functions $\psi^{(\iota)}_{j,k,\boldsymbol{m}}$ are given by (2.2) and the functions $\tilde{\psi}^{(\iota)}_{j,k,\boldsymbol{m}}$ are obtained by a simple modification of the functions $\psi^{(\iota)}_{j,k,\boldsymbol{m}} \in \Psi^{(\iota)}$ (we refer to [31] for all the details on the construction). We recall the following result.

Theorem 2.1 (Theorem 1, [31]). *Let $\phi \in L^2(\mathbb{R}^3)$ be chosen such that $\hat{\phi} \in \mathcal{C}^\infty_0(\mathbb{R}^3)$ and (2.4) hold, and $v \in C^\infty_0(\mathbb{R})$ be chosen so that* $\mathrm{supp}(v) \subset [-1,1]$ *and (2.6) hold. Then the discrete cylindrical shearlet system (2.7) is a Parseval frame for $L^2(\mathbb{R}^3)$. Furthermore, the elements of this system are $\mathcal{C}^\infty$ and compactly supported in the Fourier domain.*

With a slight abuse of notation we denote the full discrete cylindrical shearlet parameter group as:

$$\Lambda = \mathcal{T} \times \mathbb{Z}^3 \sim \{(j,k,\iota;\boldsymbol{m}) : j \in \mathbb{N}\cup\{0,-1\}, |k| \leqslant 2^j, \iota = 1,2\,;\, \boldsymbol{m} \in \mathbb{Z}^3\}.$$

Separating the translations $\boldsymbol{m} \in \mathbb{Z}^3$ from the other parameters (in $\mathcal{T}$) is needed later in subsection 2.3.2. With this notation, we define the cylindrical shearlet transform:

$$\mathcal{SH} : L^2(\mathbb{R}^3) \longrightarrow \ell^2(\Lambda), \qquad \text{where} \quad \mathcal{SH}(f) = (\langle f, \psi_\lambda\rangle)_{\lambda\in\Lambda}\,. \tag{2.8}$$

In subsection 2.3, we show that this function system belongs to a class of decomposition spaces associated with a special partition of the Fourier domain. Let us first motivate why we choose cylindrical shearlets to represent spatio-temporal data.

2.2. Cylindrical cartoon-like functions and cartoon-like videos. In this section, we recall the definition of cylindrical cartoon-like functions from [31], used to model a simple class of spatio-temporal sequences with discontinuities in the spatial plane only.

Following [29], we start by defining a class of indicator functions of sets with $\mathcal{C}^2$ boundaries. In polar coordinates, let $\rho(\theta) : [0,2\pi) \to [0,1]$ be a radius function and define the set $S = \{x \in \mathbb{R}^2 : |x| \leq \rho(\theta), \theta \in [0,2\pi)\}$. In particular, the boundary ∂S is given by the curve in $\mathbb{R}^2$:

$$\beta(\theta) = \begin{pmatrix} \rho(\theta)\cos(\theta) \\ \rho(\theta)\sin(\theta) \end{pmatrix}, \quad \theta \in [0,2\pi). \tag{2.9}$$

For any fixed constant $Z > 0$, we say that a set $S \in STAR^2(Z)$ if $S \subset [0,1]^2$ and S is a translate of a set satisfying (2.9) where

$$\sup \left|\rho''(\theta)\right| \leq Z, \quad \rho \leq \rho_0 < 1. \tag{2.10}$$

Hence, denoting with $\mathcal{C}_0^2([0,1]^2)$ the class of twice differentiable functions with support in $[0,1]^2$, the set of **cartoon-like images** $\mathcal{E}^2(Z) \subset L^2(\mathbb{R}^2)$ was defined in [29] as the set of functions $h = h_0\,\chi_S$ where $h_0 \in \mathcal{C}_0^2([0,1]^2)$, $S \in STAR^2(Z)$ and $\sum_{|\alpha|\leq 2} \|D^\alpha h\|_\infty \leq 1$.

Using the above notation and definition, some of the authors defined in [31] the class of **cylindrical cartoon-like functions** as follows.

Definition 2.2 ([31, Equation 12]). *Given a constant $Z > 0$, the class of* **cylindrical cartoon-like functions** *is the set of functions $\mathcal{U}^2(Z) \subset L^2(\mathbb{R}^3)$ of the form*

$$f(x) = f(x_1, x_2, x_3) = h(x_1, x_2)\,\chi_S(x_1, x_2)\,z(x_3), \tag{2.11}$$

where $S \in STAR^2(Z)$, $h \in \mathcal{C}_0^2([0,1]^2)$, $z \in \mathcal{C}_0^2([0,1])$ and $\|D^\alpha h\|_\infty, \|D^\alpha z\|_\infty \leq 1$.

Note that for any cylindrical cartoon-like function f and constant $c \in \mathbb{R}$, the function $\tilde{f}_c(x) = f(x_1, x_2, c)$ is a cartoon-like image, where the edge boundary ∂S is the same for all x_3.

2.2.1. Sparse approximations. Cylindrical shearlets were shown to provide (nearly) optimally sparse approximations, in a precise sense, for the class of cylindrical cartoon-like functions defined by (2.11).

Let $\Psi = \{\psi_\lambda : \lambda \in \Lambda\}$ be the Parseval frame of discrete cylindrical shearlets given by (2.7). For any function $f \in L^2(\mathbb{R}^3)$, the *cylindrical shearlet coefficients* are the elements of the sequence $\mathcal{SH}(f) = \{\langle f, \psi_\lambda\rangle : \lambda \in \Lambda\}$. We denote by $f_\mathcal{N}$ the $\mathcal{N}$-term approximation of f obtained from the $\mathcal{N}$ largest coefficients of its cylindrical shearlet expansion, namely:

$$f_\mathcal{N} = \sum_{\lambda \in \mathcal{I}_\mathcal{N}} \langle f, \psi_\lambda\rangle\,\psi_\lambda,$$

where $\mathcal{I}_\mathcal{N} \subset \Lambda$ is the set of indices corresponding to the $\mathcal{N}$ largest entries of the absolute-value sequence $\{|\langle f, \psi_\lambda\rangle| : \lambda \in \Lambda\}$. We have the following result from [31].

Theorem 2.3 ([31, Theorem 3]). *Given any $Z > 0$, let $f \in \mathcal{U}^2(Z)$, the set of cylindrical cartoon-like functions, and $f_\mathcal{N}$ be the $\mathcal{N}$-term approximation to f defined above. Then, for $\mathcal{N} \in \mathbb{N}$ there exists a constant c independent of λ in $\mathcal{SH}(f)$ and $\mathcal{N}$ such that:*

$$\|f - f_\mathcal{N}\|_{L^2}^2 \leq c\,\mathcal{N}^{-2}(\ln(\mathcal{N}))^3.$$

Remark 2.4. The decay rate in Theorem 2.3 is nearly optimal, in the sense that, up to a logarithmic factor, no other representation system satisfying polynomial depth search can achieve an asymptotic approximation rate faster than $\mathcal{N}^{-2}$ for the class of functions $\mathcal{U}^2(Z)$ as shown in [31]. By contrast, separable 3D wavelets and conventional 3D shearlets, up to a logarithmic factor, only achieve approximation rates that are $O(\mathcal{N}^{-1/2})$ [28] and $O(\mathcal{N}^{-1})$ [45], respectively. The notion of representation systems *satisfying polynomial depth search* includes any Parseval frame and was introduced to include function representations that are more general than orthonormal bases but not as large as to make the approximation problem computationally intractable [18, 45]. The significance of Theorem 2.3 is that sparsity-promoting *cylindrical shearlet norms are expected to provide a sparsity-inducing prior for data in the class of cylindrical cartoon-like functions.*

2.2.2. Cartoon-like videos. Cylindrical cartoon-like functions are a rather coarse model of spatio-temporal data, as they exhibit the same discontinuous edge at the same location over the entire temporal sequence. We introduce here a generalization of this class of functions providing a more realistic model of spatio-temporal data found in applications.

Definition 2.5. *Given a constant $Z > 0$, we define the class of* **cartoon-like videos** *as the functions $\mathcal{G}^2(Z) \subset L^2(\mathbb{R}^3)$ of the form*

$$f(x) = h(x_1 - q_1(x_3), x_2 - q_2(x_3))\, \chi_S(x_1 - q_1(x_3), x_2 - q_2(x_3))\, z(x_3), \tag{2.12}$$

where $S \in STAR^2(Z)$, $h \in \mathcal{C}_0^2([0,1]^2)$, $z \in \mathcal{C}_0^2([0,1])$, $\|D^\alpha h\|_\infty, \|D^\alpha z\|_\infty \leq 1$ and $q_1, q_2 \in \mathcal{C}_0^2([0,1])$ with $q_1(0) = q_2(0) = 0$.

By construction, for any x_3, discontinuities in cartoon-like videos occur on the boundary of the sets:

$$S_{x_3} = \{(x_1, x_2) \in \mathbb{R}^2 : |(x_1 - q_1(x_3), x_2 - q_2(x_3))| \leq \rho(\theta), \theta \in [0, 2\pi)\},$$

where the radius function $\rho(\theta)$ is given above and $S = S_0$; i.e., we have the boundary curves:

$$\theta \mapsto (\rho(\theta)\cos(\theta) + q_1(x_3), \rho(\theta)\sin(\theta) + q_2(x_3))\,.$$

Hence, we can write (2.12) as:

$$f(x) = h(x_1 - q_1(x_3), x_2 - q_2(x_3))\, \chi_{S_{x_3}}(x_1, x_2)\, z(x_3).$$

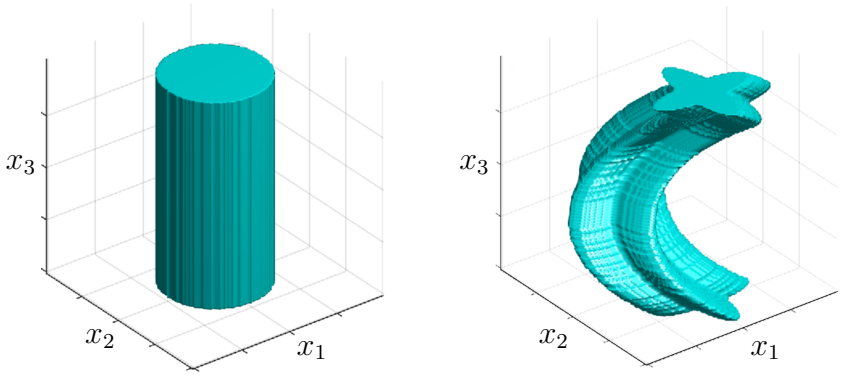

Figure 2. *Example of a cylindrical cartoon-like function (left) and a cartoon-like video (right) in $L^2(\mathbb{R}^3)$.*

Figure 2 shows an example of a cylindrical cartoon-like function (left) and an example of a cartoon-like video (right) defined by displacing a starshape region $S \subset \mathbb{R}^2$ as a function of the x_3 coordinate. Note that the region S is allowed to gradually disappear if $z(x_3) = 0$ at some x_3.

Remark 2.6. Based on our numerical experiments reported below, we conjecture that the sparse approximation result in Theorem 2.3 is also valid for the class of cartoon-like videos. However, no proof of this conjecture is available at this time.

2.3. Decomposition and smoothness spaces. We introduce the class of *cylindrical shearlet smoothness spaces* by adapting the construction of the shearlet decomposition spaces [63] associated with traditional shearlets [62, 44] to cylindrical shearlet systems. We start by recalling the basic definitions of the decomposition space theory from [35, 34, 5]. For brevity, we write $f \lesssim g$ if there exists a constant $C > 0$ independent of f and g such that $f \leq Cg$. We denote $f \lesssim g \lesssim f$ by $f \simeq g$.

2.3.1. Constructing decomposition spaces. An *admissible covering* $\{Q_i \subset \mathbb{R}^d : i \in I\}$ is any collection of bounded and measurable sets such that $\bigcup_{i\in I} Q_i = \mathbb{R}^d$ and there exists a uniform bound on how many of the sets Q_i can overlap at any point in $\mathbb{R}^d$. Given any admissible covering a *bounded admissible partition of unity* (BAPU) is a family of functions $\{\gamma_i : i \in I\}$ such that, each γ_i is supported only in Q_i; $\sum_i \gamma_i(\xi) = 1$ for any $\xi \in \mathbb{R}^d$; and $\sup_i |Q_i|^{\frac{1}{p}-1} \|\check{\gamma}_i\|_{L^p} < \infty$ for any $p \in (0, \infty]$, where $\check{\gamma}$ is the inverse Fourier transform of γ.

We are interested in more structured coverings. Let $\mathcal{T} = \{\mathcal{T}_k : x \mapsto A_k x + b_k : k \in \mathbb{N}\}$ be a collection of invertible affine transformations on $\mathbb{R}^d$ and, to ease the notation, we omit the index k, if possible. Consider two bounded open sets $P, Q \subset \mathbb{R}^d$ such that P is compactly contained in Q and the collections $\mathcal{Q} = \{\mathcal{T}Q : \mathcal{T} \in \mathcal{T}\}$ and $\mathcal{P} = \{\mathcal{T}P : \mathcal{T} \in \mathcal{T}\}$ are both admissible coverings. If, in addition there is a constant C such that:

$$(A_k Q + b_k) \cap (A_l Q + b_l) \neq \emptyset \quad \Longrightarrow \quad \|A_l^{-1} A_k\|_{\ell^\infty} < C,$$

then we call $\mathcal{Q}$ a *structured admissible covering* and $\mathcal{T}$ a *structured family of affine transformations*. Then, we know that $\mathcal{Q}$ and $\mathcal{T}$ are also associated with a *squared BAPU* $\{\varphi_{\mathcal{T}} : \mathcal{T} \in \mathcal{T}\}$, i.e., the second condition is replaced by $\sum_{\mathcal{T}} |\varphi_{\mathcal{T}}(\xi)|^2 = 1$ and the third one only needs to hold for any $p \in (0, 1]$.

Denote $|\mathcal{T}| = |\det A|$ and let $\mathcal{Q}$ and $\mathcal{T}$ be as defined above. Suppose $K_a = [-a, a]^d$, for $a > 0$, is a cube in $\mathbb{R}^d$ containing Q. Then we can define the system:

$$\{\eta_{\boldsymbol{m},\mathcal{T}} \ : \ \widehat{\eta}_{\boldsymbol{m},\mathcal{T}} = \varphi_{\mathcal{T}} e_{\boldsymbol{m},\mathcal{T}}, \ \boldsymbol{m} \in \mathbb{Z}^d, \mathcal{T} \in \mathcal{T}\},$$

where $\varphi_{\mathcal{T}}$ is a squared BAPU and

$$e_{\boldsymbol{m},\mathcal{T}}(\xi) = \frac{\chi_{K_a}(\mathcal{T}^{-1}\xi) e^{i\frac{\pi}{a}\boldsymbol{m}\cdot\mathcal{T}^{-1}\xi}}{\sqrt{|2a|^d |\mathcal{T}|}}, \ \boldsymbol{m} \in \mathbb{Z}^d, \mathcal{T} \in \mathcal{T}.$$

It follows that the system of $\eta_{\boldsymbol{m},\mathcal{T}}$'s is a Parseval frame of $L^2(\mathbb{R}^d)$ [5]. Then, if we denote:

$$\eta^{(p)}_{\boldsymbol{m},\mathcal{T}} = |\mathcal{T}|^{\frac{1}{2}-\frac{1}{p}} \eta_{\boldsymbol{m},\mathcal{T}},$$

we can characterize all of the decomposition spaces $D(\mathcal{Q}, L^p, \mathcal{Y}_w)$ using the associated frame coefficients $\{\langle f, \eta^{(p)}_{\boldsymbol{m},\mathcal{T}}\rangle : \boldsymbol{m}, \mathcal{T}\}$ in the way specified by the following result.

Proposition 2.7 ([5, Proposition 3]). *Let $\mathcal{Q} = \{\mathcal{T}Q : \mathcal{T} \in \mathcal{T}\}$ be a structured admissible covering, $\mathcal{Y}_w$ a solid (quasi-)Banach sequence space on $\mathcal{T}$ with a $\mathcal{Q}$-moderate weight[1] w. For*

[1] A strictly positive function w is $\mathcal{Q}$-moderate if exists $C > 0$ such that $w(x) \leqslant Cw(y)$ for all $x, y \in Q_i \in \mathcal{Q}$ and all $i \in I$. A $\mathcal{Q}$-moderate *weight* on I is derived as the sequence $(w_i = w(x_i))_{i\in I}$, where each $x_i \in Q_i$.

$0 < p \leqslant \infty$ *we have the characterization:*

$$\|f\|_{D(\mathcal{Q},L^p,\mathcal{Y}_w)} \simeq \left\| \left(\Big(\sum_{\boldsymbol{m}\in\mathbb{Z}^d} |\langle f, \eta^{(p)}_{\boldsymbol{m},\mathcal{T}}\rangle|^p \Big)^{\frac{1}{p}} \right)_{\mathcal{T}\in\mathcal{T}} \right\|_{\mathcal{Y}_w}.$$

This indicates that there is a coefficients space $d(\mathcal{Q},\ell^p,\mathcal{Y}_w)$ associated with the decomposition space and we can define it as the set of coefficients $c = (c_{\boldsymbol{m},\mathcal{T}})_{\boldsymbol{m}\in\mathbb{Z}^d,\mathcal{T}\in\mathcal{T}}$ which satisfy:

$$\|c\|_{d(\mathcal{Q},\ell^p,\mathcal{Y}_w)} = \left\| \left(\Big(\sum_{\boldsymbol{m}\in\mathbb{Z}^d} |c_{\boldsymbol{m},\mathcal{T}}|^p \Big)^{\frac{1}{p}} \right)_{\mathcal{T}\in\mathcal{T}} \right\|_{\mathcal{Y}_w} < \infty.$$

If we define the *coefficient operator* $\mathcal{K} : D(\mathcal{Q},L^p,\mathcal{Y}_w) \to d(\mathcal{Q},\ell^p,\mathcal{Y}_w)$ and the *reconstruction operator* $\mathcal{K}^* : d(\mathcal{Q},\ell^p,\mathcal{Y}_w) \to D(\mathcal{Q},L^p,\mathcal{Y}_w)$ in the obvious way:

$$\mathcal{K}(f) = \Big(\langle f, \eta^{(p)}_{\boldsymbol{m},\mathcal{T}}\rangle\Big)_{\boldsymbol{m},\mathcal{T}}, \qquad \text{and} \qquad \mathcal{K}^*(c) = \sum_{\boldsymbol{m},\mathcal{T}} c_{\boldsymbol{m},\mathcal{T}}\, \eta^{(p)}_{\boldsymbol{m},\mathcal{T}},$$

then both operators are bounded and $\mathcal{K}^*\mathcal{K} = \mathrm{Id}_{D(\mathcal{Q},L^p,\mathcal{Y}_w)}$ [5]. From now on we shall use the more familiar names *analysis* and *synthesis operator* when referring to the coefficient $\mathcal{K}$ and reconstruction $\mathcal{K}^*$ operators, respectively. [2]

Finally, as shown in [5], by choosing $\mathcal{T}$ to be a structured family of affine transformations (where $\mathcal{T}x := A_{\mathcal{T}}x + b_{\mathcal{T}}$), $\mathcal{Y}_w = \ell^q_{w^\beta}(\mathcal{T})$, with w a $\mathcal{Q}$-moderate function, $\beta \in \mathbb{R}$ and the weights defined by $w^\beta = \big(w^\beta(\mathcal{T})\big)_{\mathcal{T}\in\mathcal{T}}$, the decomposition spaces form certain *smoothness spaces*, which we denote by $S^\beta_{p,q}(\mathcal{T},w)$. For example, with wavelet transform-like dilations $\mathcal{T}$ we can identify the smoothness space $S^\beta_{p,q}$ with a suitable Besov space [63].

In the next section, we construct a similar setting for the cylindrical shearlets.

2.3.2. Cylindrical shearlet decomposition spaces. We fix the structured admissible covering by using exactly those affine transformations defined for the cylindrical shearlet system (2.1) in subsection 2.1 and write:

$$\mathcal{T} := \{\mathcal{T} = M^{\iota-1}_{\leftrightarrow} B^k A^j \,:\, \iota = 1,2, |k| \leq 2^j, j \in \mathbb{N}\} \cup \{M_0\} \qquad \text{and} \qquad \mathcal{Q} := \{\mathcal{T}Q_{0,0} \,:\, \mathcal{T} \in \mathcal{T}\}.$$

Now, a squared BAPU can be obtained by cancelling out the normalization term in the cylindrical shearlet system (2.7), that is:

$$\varphi_{\mathcal{T}_{j,k,\iota}}(\xi) = \big|\det A_{(\iota)}\big|^{\frac{j}{2}}\, \widehat{\psi}^{(\iota)}_{j,k,0}(\xi) := |\mathcal{T}|^{\frac{1}{2}} \widehat{\psi}_{\mathcal{T},0}(\xi). \tag{2.13}$$

Hence we define the cylindrical shearlet smoothness space $S^\beta_{p,q}(\mathbb{R}^3)$, for $\beta \in \mathbb{R}, 0 < p,q \leqslant \infty$ and $w(\mathcal{T}) = 2^j$, as the set of functions $f \in \mathcal{S}'(\mathbb{R}^3)$ such that:

$$\|f\|_{S^\beta_{p,q}} \simeq \Big(\sum_{\mathcal{T}\in\mathcal{T}} \Big(2^{\beta j} \sum_{\boldsymbol{m}\in\mathbb{Z}^3} \big|\langle \widehat{f}, |\mathcal{T}|^{\frac{1}{2}-\frac{1}{p}} \widehat{\psi}_{\mathcal{T},\boldsymbol{m}}\rangle\big|^p \Big)^{\frac{q}{p}} \Big)^{\frac{1}{q}}$$

[2] Historically it seems that the names coefficient and reconstruction operator were used in the theory of atomic decompositions, whereas analysis and synthesis originate from the more classical frame theory of Hilbert spaces. The theory of Banach frames lies at their intersection (see [20, 19] for more details and context).

$$= \Big(\sum_{\mathcal{T}\in\mathcal{T}} \Big(2^{\beta j} |\mathcal{T}|^{\frac{p}{2}-1} \sum_{\boldsymbol{m}\in\mathbb{Z}^3} \big| \langle \widehat{f}, \widehat{\psi}_{\mathcal{T},\boldsymbol{m}} \rangle \big|^p \Big)^{\frac{q}{p}} \Big)^{\frac{1}{q}} < \infty. \tag{2.14}$$

In particular for $q = p$ and if we denote $\tilde{w} = \tilde{w}_{\beta,p}(\mathcal{T}) := w^\beta(\mathcal{T})\,|\mathcal{T}|^{\frac{p}{2}-1}$ we can define the cylindrical shearlet smoothness space norm:

$$\|f\|_{S^\beta_{p,p}} \simeq \|\,\mathcal{SH}(f)\|_{\ell^p(\tilde{w})} = \Big(\sum_{\lambda\in\Lambda} \tilde{w}(\lambda) |\langle f, \psi_\lambda \rangle|^p \Big)^{\frac{1}{p}} < \infty \quad \forall f \in S^\beta_{p,p}, \tag{2.15}$$

where the cylindrical shearlet transform $\mathcal{SH}$ from (2.8) acts as the analysis operator. Essentially, we expressed the (weighted) ℓ^p-norm of the cylindrical shearlet coefficients of f as the norm of f in the cylindrical shearlet smoothness space $S^\beta_{p,p}$. Moreover, since the cylindrical shearlet system (2.7) is a (smooth) Parseval frame (Theorem 2.1), we have the following result.

Proposition 2.8. *Let $0 < p \le 2$ and $\|f\|_{L^2} \ge 1$. For any weight $\tilde{w}$ such that $\inf_{\lambda\in\Lambda} \tilde{w}(\lambda) > 0$, there is $C > 0$ such that*

$$\|f\|_{L^2} \le \|\,\mathcal{SH}(f)\|_{\ell^p} \le C \|\,\mathcal{SH}(f)\|_{\ell^p(\tilde{w})}. \tag{2.16}$$

In particular, the cylindrical shearlet smoothness space norm is coercive in $L^2(\mathbb{R}^3)$.

Proof. First, we use the fact that for $0 < p \le 2$ we have:

$$\|u\|^2_{\ell^2} \le \sum_{i=1}^{\infty} |u_i|^p = \|u\|^p_{\ell^p} \qquad \forall u \in \ell^2 : \|u\|_{\ell^2} \le 1, \tag{2.17}$$

which follows since $|u_i|^2 \le |u_i|^p \le 1$ for all i. Next, notice that for any $\|v\|_{\ell^2} > 1$ we have:

$$\begin{aligned} 1 < \|v\|^2_{\ell^2} = \|v\|^2_{\ell^2} \big\| \tfrac{v}{\|v\|_{\ell^2}} \big\|^2_{\ell^2} \le \|v\|^2_{\ell^2} \big\| \tfrac{v}{\|v\|_{\ell^2}} \big\|^p_{\ell^p} = \|v\|^{2-p}_{\ell^2} \|v\|^p_{\ell^p} \quad &\iff \quad \|v\|^p_{\ell^2} \le \|v\|^p_{\ell^p} \\ &\iff \quad \|v\|_{\ell^2} \le \|v\|_{\ell^p}. \end{aligned} \tag{2.18}$$

Since cylindrical shearlets form a Parseval frame, we get the unweighted version of (2.16) by using (2.18), i.e., $\|f\|_{L^2} = \|\,\mathcal{SH}(f)\|_{\ell^2} \le \|\,\mathcal{SH}(f)\|_{\ell^p}$. Then, given any (not necessarily $\mathcal{Q}$-moderate) weight $\tilde{w}$, and denoting $c = \inf_{\lambda\in\Lambda} \tilde{w}(\lambda)$, it follows that:

$$\|f\|_{L^2} \le \frac{1}{c^{1/p}} c^{1/p} \|\,\mathcal{SH}(f)\|_{\ell^p} \le \frac{1}{c^{1/p}} \|\,\mathcal{SH}(f)\|_{\ell^p(\tilde{w})}.$$

This proves the coercivity and inequality (2.16) for $C = \frac{1}{c^{1/p}}$. ∎

Since in section 3 we are interested in weights of the form $\tilde{w}(\mathcal{T}) = w^\beta(\mathcal{T})\,|\mathcal{T}|^{\frac{p}{2}-1}$ with $w(\mathcal{T}) = 2^j$, we conclude this section proving the next result.

Corollary 2.9. *For $\tilde{w} = 2^{j\beta}\,|\mathcal{T}|^{\frac{p}{2}-1}$, if $\beta \ge \frac{5}{4}(2-p)$, then $\|\,\mathcal{SH}(f)\|_{\ell^p(\tilde{w})}$ is coercive.*

Proof. We need to check when $\inf \tilde{w} > 0$. Since $\det(M_0) = 1$ and $\det(M_{\leftrightarrow}) = 1$, we can focus on the case when $j \ge 1$. Hence the sequence of weights is always monotone $\tilde{w}(j) = 2^{\beta j}\, 2^{\frac{5}{2} j \left(\frac{p}{2}-1\right)} = 2^{j\beta + \frac{5}{2} j \left(\frac{p}{2}-1\right)}$. It is clear that $\inf_{j\ge 0} \tilde{w}(j) = \lim_{j\to\infty} \tilde{w}(j) = 0$ only when the exponent is negative. Therefore, the necessary condition for β is given by:

$$j\beta + \frac{5}{2} j \left(\frac{p}{2} - 1 \right) \ge 0 \quad \iff \quad \beta \ge \frac{5}{4}(2-p).$$

∎

Remark 2.10. The usual choice for unweighted ℓ^p regularization of the (cylindrical) shearlet coefficients actually corresponds to the limit case $\beta = \frac{5}{4}(2-p)$, i.e., $w^\beta(j) = 2^{\frac{5}{4}(2-p)j}$ in order to have $\tilde{w}(\mathcal{T}) = 1$ for all $\mathcal{T} \in \mathcal{T}$.

For the remainder of the paper, we only consider the case $p > 1$ to carry out our theoretical analysis. When $p > 1$, the smoothness spaces $S^\beta_{p,p}(\mathbb{R}^3)$ are reflexive Banach spaces, since the spaces $L^p(\mathbb{R}^3)$ and $\mathcal{Y}_w = \ell^p_{w^\beta}(\mathcal{T})$ are both reflexive Banach spaces (cf. [103, Theorem 3.21] and [35, Theorem 2.8]). Note also that, by inequality (2.17) and the fact that the cylindrical shearlet system is a Parseval frame, it follows that convergence in $S^\beta_{p,p}$ implies convergence in L^2.

3. Regularizing a dynamic inverse problem with cylindrical shearlets.

In this section, we formulate the inverse problem we want to tackle using cylindrical shearlet regularization. To this end, we first introduce the related notation and the necessary assumptions, along with a motivating example (i.e., the inverse problem arising from dynamic Computed Tomography (CT)) which will be used in the numerical simulations in section 4. Then we derive convergence rates considering two scenarios: full measurements and (randomly) subsampled measurements.

3.1. Dynamic inverse problems.

Consider the following (dynamic) inverse problem:

$$\text{(3.1)} \quad \text{given} \quad g^\delta \in Y \ : \quad g^\delta(t) = (\mathcal{A}f^\dagger)(t) + \delta\epsilon(t) \quad \text{a.e. } t \in (0,T), \qquad \text{recover} \quad f^\dagger \in X.$$

We can formulate the problem in a rather general framework assuming that X is a reflexive Banach space, $X \subset L^2(\Omega \times (0,T))$, being $\Omega \subset \mathbb{R}^2$ a bounded set, and $Y = L^2(0,T;\tilde{Y})$, where $\tilde{Y}$ is a Hilbert space. The notation $L^2(0,T;\tilde{Y})$ is used for the Bochner space:

$$L^2(0,T;\tilde{Y}) = \left\{ g\colon [0,T] \to \tilde{Y} \text{ measurable s.t. } \int_0^T \|g(t)\|^2_{\tilde{Y}}\, dt < \infty \right\}.$$

For later purposes (see subsection 3.4), it is convenient to interpret the elements of $\tilde{Y}$ as functions, whose variable lives in the compact set $\mathcal{U}$ and whose values range in a suitable Hilbert space $\mathcal{V}$: i.e., $\tilde{Y} = L^2(\mathcal{U};\mathcal{V})$. We assume that $\mathcal{A}$ is a bounded linear operator from X to Y. In the case of our motivating example (i.e., dynamic CT), the operator $\mathcal{A}$ is the Radon transform, which we briefly introduce here. This definition is the first step towards introducing our final model, which is detailed in the following subsections.

Motivating Example, Step I (Radon transform):
For a function $f \in L^2(\Omega)$, the Radon transform $\mathcal{R}f$ is defined as follows:

$$\mathcal{R}f(\theta, s) = \int_{\mathbb{R}} f(s\theta + \sigma\theta^\perp) d\sigma \qquad \theta \in S^1, s \in \mathbb{R}\,.$$

The function $\mathcal{R}f$ (the so-called sinogram) belongs to the space $H^{1/2}([0,2\pi) \times (-\bar{s},\bar{s}))$ for a suitable $\bar{s} > 0$ depending on Ω (see [76, Theorem 2.10]). In particular, the range of $\mathcal{R}$ is a subset of $L^2([0,2\pi) \times (-\bar{s},\bar{s}))$, which we may interpret as $L^2(\mathcal{U};\mathcal{V})$ being $\mathcal{U} = [0,2\pi)$ and $\mathcal{V} = L^2(-\bar{s},\bar{s})$. The dynamic operator $\mathcal{A}\colon L^2(\Omega \times (0,T)) \to Y$ associated with the Radon transform is then defined by $(\mathcal{A}f)(t) = \mathcal{R}(f(\cdot,t))$ for almost every time t.

Next, going back to problem (3.1), for the error term $\delta\epsilon$, with $\delta > 0$, we consider two different scenarios, for which we formulate different assumptions:

- *Deterministic* noise. In this case, $\epsilon \in Y$ is a generic (unknown) function perturbing the measurement, for which we only assume that

$$\|\epsilon\|_Y \leq 1. \tag{3.2}$$

  This is the most common environment in which classical regularization techniques for inverse problems have been introduced and developed (see [32]).
- *Statistical* noise. In this scenario, the measurement acquisition is modeled as a random experiment, and the uncertainty ϵ is modeled as a random process in time such that, at almost every t, $\epsilon(t)$ is a random variable taking values in $\tilde{Y}$. In this case, analogously to [7], we assume that at almost every time t the variable $\|\epsilon(t)\|_{\tilde{Y}}$ is sub-Gaussian (see [102, Definition 2.5.6]) and

$$\mathbb{E}[\epsilon(t)] = 0 \quad \text{and} \quad \left\|\|\epsilon(t)\|_{\tilde{Y}}\right\|_{\mathrm{sG}} \leq 1 \qquad \text{a.e. } t \in (0,T). \tag{3.3}$$

It is clear that the deterministic case can be interpreted as a single realization of the statistical one, provided that the noise distribution is bounded. In subsection 3.2 and subsection 3.3, our analysis will consider both scenarios, whereas in subsection 3.4 we will restrict only to the statistical one in the case of (randomly) sampled measurements. We point out that, in the statistical noise scenario, the assumption that $\epsilon(t)$ is a random variable on $\tilde{Y}$ can be restrictive: for example, as shown in [61], Gaussian white noise cannot be interpreted as a random variable on the infinite-dimensional space $L^2(\Omega)$. In order to encompass a larger family of noise models, the reader is referred to the approach presented in [14].

3.2. Regularization with cylindrical shearlets. In this section, we show how cylindrical shearlets can be employed to define regularization strategies for dynamic inverse problems. Our inspiration comes from applications in the non-dynamic setting [21, 85]: indeed, in many static, imaging problems it is common practice to assume *a priori* that the target is sparse with respect to some wavelet representation. This is modeled as a variational problem in a suitable Besov space (i.e., employing Besov norms as a regularizer).

In our setup, motivated by the observation that cylindrical cartoon-like functions (i.e., "dynamic images") are optimally sparse with respect to cylindrical shearlet frames (see subsection 2.2), we consider the smoothness spaces $S^{\beta}_{p,p}(\mathbb{R}^3)$ as the natural embedding space for this class of problems.

In principle, the theoretical analysis in section 2 would hold for any $p > 0$, but we are only interested here in the case where $p < 2$ (to enforce sparsity) and $p > 1$ (to leverage strong convexity). Notice that in place of a smoothness space as the ones introduced in subsection 2.3.2 we can consider its subspace consisting of those functions which are supported on the compact set $\overline{\Omega} \times [0,T]$. We therefore set

$$X = S^{\beta}_{p,p}(\mathbb{R}^3) \cap \{f : \mathrm{supp}(f) \subset \overline{\Omega} \times [0,T]\}, \tag{3.4}$$

with $1 < p < 2$ and $\beta \geq \frac{5}{4}(2-p)$. As remarked above, any converging sequence in X must converge also in L^2-norm implying that the limit must also be supported on the compact set

$\overline{\Omega} \times [0,T]$. Hence X is a closed subspace of $S^{\beta}_{p,p}(\mathbb{R}^3)$. By Corollary 2.9 X is also continuously embedded in $L^2(\Omega \times (0,T))$, and we can define the following minimization problem:

$$f^{\delta}_{\alpha} = \operatorname*{argmin}_{f \in X} J^{\delta}_{\alpha}(f) := \operatorname*{argmin}_{f \in X} \left\{ \frac{1}{2} \, \|\mathcal{A}f - g^{\delta}\|^2_Y + \alpha R(f) \right\}, \tag{3.5}$$

where $J^{\delta}_{\alpha} \colon X \to \mathbb{R}$ and we choose

$$R(f) = \frac{1}{p}\|f\|^p_X = \frac{1}{p}\| \mathcal{SH}(f)\|^p_{\ell^p(\tilde{w})}, \tag{3.6}$$

with the weights $\tilde{w}_\lambda = \tilde{w}(j) = 2^{j\beta + \frac{5}{2} j\left(\frac{p}{2}-1\right)}$, being $\lambda \in \Lambda$ the single index describing the shearlet elements, and $j = |\lambda|$ its scale.

We first observe that problem (3.5) is well-posed, and in particular that the minimizer f^{δ}_{α} always exists and is unique. The next result proves this for both proposed noise scenarios.

Proposition 3.1. *Let X be as in (3.4) with $1 < p < 2$ and $\beta \geq \frac{5}{4}(2-p)$, and let $\mathcal{A}$ be a bounded operator from X to Y. Define $g^{\delta} = \mathcal{A}f^{\dagger} + \delta\epsilon$, with $\delta > 0$ and $\alpha > 0$. Then:*

- *If $\epsilon \in Y$ satisfies (3.2), there exists a unique solution f^{δ}_{α} of (3.5).*
- *If ϵ is a random process on Y satisfying (3.3), the solution f^{δ}_{α} of (3.5) is a uniquely defined random variable on X.*

Proof. Let us first consider the deterministic scenario. In this case, the existence of a minimizer is guaranteed by standard arguments, since the functional $J^{\delta}_{\alpha} \colon X \to \mathbb{R}$ is lower semi-continuous and coercive (thanks to Corollary 2.9), thus its sublevel sets are bounded and, since X is a reflexive Banach space, they are also compact with respect to the weak topology. Moreover, J^{δ}_{α} is the sum of a convex quadratic term and a strictly convex term (since $p > 1$): as a consequence, its minimizer is also unique.
In the statistical scenario, at each time the noise $\epsilon(t)$ is a random variable on $\tilde{Y}$, namely, a measurable function on a suitably defined probability space. We can use the same argument as above to show that, for every realization of the process ϵ, a minimizer of J^{δ}_{α} exists in X. The function mapping the probability space to the minimizer f^{δ}_{α} is also measurable, thanks to Aumann's selection principle (see [95, Lemma A.3.18]). ∎

The next step is showing that (3.5) represents a regularization strategy for the inverse problem (3.1). To do so, we shall prove that if the noise δ reduces, there exists a selection rule $\alpha(\delta)$ such that the solution $f^{\delta}_{\alpha(\delta)}$ of the variational problem converges, in a suitable metric, towards the solution $f^{\dagger}$. This is the content of the next section, where we formally state this result, providing also convergence rates.

3.3. Full measurements: convergence rates. In this section, we provide quantitative estimates regarding the convergence of the regularized solution f^{δ}_{α} to the true solution $f^{\dagger}$, according to a suitable a priori choice of parameter $\alpha(\delta)$. Our approach shares many similarities with the one in [14] (e.g., leading to Corollary 3.12). The main difference with respect to their result is that we are considering a simpler model for the noise, namely that it belongs to Y (deterministic noise) or is a random process in Y(statistical noise). One should compare our results with the one in [14, Corollary 3.12] by setting $r_2 = 0$ in their estimate.

Since the variational regularization problem (3.5) involves the minimization of a convex functional, we employ tools from convex analysis to address it. In particular, we use the symmetric Bregman distance of the functional $R(f) = \frac{1}{p}\|f\|_X^p$ to measure the distance between f_α^δ and $f^\dagger$. We briefly recall its definition. First, for a convex functional R, the subdifferential $\partial R(f)$ at a point $f \in X$ consists of:

$$\partial R(f) = \{r \in X^* : R(\tilde{f}) \geq R(f) + \langle r, \tilde{f} - f\rangle_* \quad \forall \tilde{f} \in X\},$$

where $\langle\cdot,\cdot\rangle_*$ denotes the standard pairing between the (reflexive) Banach space X and its dual X^*. If the functional R is differentiable (as in our case, since the norm in a reflexive Banach space is Fréchet differentiable, see [92, Remark 2.38]), then $\partial R(f) = \{\nabla R(f)\}$. For $f, \tilde{f} \in X$, we can define the symmetric Bregman distance between f and $\tilde{f}$ as:

$$D_R(\tilde{f}, f) = \langle \nabla R(\tilde{f}) - \nabla R(f), \tilde{f} - f\rangle_*.$$

The purpose of the next result is to show that $D_R(f_\alpha^\delta, f^\dagger)$ converges to 0, also providing some convergence rates. To do so, we need to introduce some additional assumption on the solution $f^\dagger$. Within this framework, we limit ourselves to the simplest and most classical assumption, namely the so-called *source condition*:

$$\exists\, w^\dagger \in Y \;:\; \nabla R(f^\dagger) = \mathcal{A}^* w^\dagger. \tag{3.7}$$

The adjoint operator $\mathcal{A}^*\colon Y \to X^*$ is bounded thanks to the assumptions on $\mathcal{A}$, and is defined by the equality $\langle \mathcal{A}^* f, g\rangle_{X^*} = \langle f, \mathcal{A} g\rangle_Y$, for all $f \in Y$ and $g \in X$.

Proposition 3.2. *Let X be as in (3.4) with $1 < p < 2$ and $\beta \geq \frac{5}{4}(2-p)$, and let $\mathcal{A}$ be a bounded operator from X to Y. Suppose that $f^\dagger$ satisfies the source condition (3.7). Let $g^\delta = \mathcal{A} f^\dagger + \delta\epsilon$, $\delta > 0$, and f_α^δ the solution of (3.5) with $\alpha > 0$. Then, if $\epsilon \in Y$ satisfies (3.2), it holds:*

$$D_R(f_\alpha^\delta, f^\dagger) \leq 2\alpha\|w^\dagger\|_Y^2 + \frac{2\delta^2}{\alpha}. \tag{3.8}$$

If ϵ is a random process on Y satisfying (3.3) it holds:

$$\mathbb{E}[D_R(f_\alpha^\delta, f^\dagger)] \leq 2\alpha\|w^\dagger\|_Y^2 + \frac{2\delta^2}{\alpha} \tag{3.9}$$

Proof. Let us first consider the deterministic noise, $\epsilon \in Y$. Since the functional J_α^δ in (3.5) is differentiable and strictly convex, the minimizer f_α^δ satisfies the optimality condition:

$$\mathcal{A}^*(\mathcal{A} f_\alpha^\delta - g^\delta) + \alpha \nabla R(f_\alpha^\delta) = 0$$

Taking a duality product with $f_\alpha^\delta - f^\dagger$ and using the expression of g^δ, we get

$$\|\mathcal{A}(f_\alpha^\delta - f^\dagger)\|_Y^2 + \alpha\langle \nabla R(f_\alpha^\delta), f_\alpha^\delta - f^\dagger\rangle_* = \delta\langle \mathcal{A}^*\epsilon, f_\alpha^\delta - f^\dagger\rangle_*.$$

Subtracting $\alpha\langle \nabla R(f^\dagger), f_\alpha^\delta - f^\dagger\rangle_*$ on both sides, we get

$$\|\mathcal{A}(f_\alpha^\delta - f^\dagger)\|_Y^2 + \alpha D_R(f_\alpha^\delta, f^\dagger) = \alpha\langle \nabla R(f^\dagger), f_\alpha^\delta - f^\dagger\rangle_* + \delta\langle \mathcal{A}^*\epsilon, f_\alpha^\delta - f^\dagger\rangle_*.$$

Now, using the source condition and the definition of the adjoint,

$$\|\mathcal{A}(f_\alpha^\delta - f^\dagger)\|_Y^2 + \alpha D_R(f_\alpha^\delta, f^\dagger) = \langle \alpha w^\dagger, \mathcal{A}(f_\alpha^\delta - f^\dagger)\rangle_Y + \langle \delta\epsilon, \mathcal{A}(f_\alpha^\delta - f^\dagger)\rangle_Y.$$

Finally, using Young's inequality on both terms on the right-hand side, we get

$$\frac{1}{2}\|\mathcal{A}(f_\alpha^\delta - f^\dagger)\|_Y^2 + \alpha D_R(f_\alpha^\delta, f^\dagger) \leq 2\alpha^2\|w^\dagger\|_Y^2 + \delta^2\|\epsilon\|_Y^2, \tag{3.10}$$

from which (3.8) immediately follows. In the statistical noise scenario, (3.9) holds for every realization of ϵ: by taking the expectation on both sides, (3.9) follows immediately from (3.3) and from the bound on the moments of sub-Gaussian random variables in terms of their sub-Gaussian norm, see [102, Proposition 2.5.2]. ∎

In both estimates, by choosing $\alpha = \alpha(\delta) = \frac{\delta}{\|w^\dagger\|}$, we get:

$$D_R(f_\alpha^\delta, f^\dagger) \leq 2\|w^\dagger\|_Y\delta \qquad \text{and} \qquad \mathbb{E}[D_R(f_\alpha^\delta, f^\dagger)] \leq 2\|w^\dagger\|_Y\delta. \tag{3.11}$$

Note that the rates in Proposition 3.2 do not depend on p, but the metric D_R does.

Remark 3.3 (non-negativity constraint). In many (static or dynamic) imaging applications, including X-ray tomography, it is known *a priori* that the desired solution $f^\dagger$ is non-negative and including this information in the variational problem (3.5) is fundamental to obtain superior reconstruction results. This leads to the following minimization problem:

$$\operatorname*{argmin}_{f\in X}\left\{\frac{1}{2}\ \|\mathcal{A}f - g^\delta\|_Y^2 + \alpha R(f) + \iota_+(f)\right\}, \tag{3.12}$$

where ι_+ is the indicator function of the non-negative orthant in $X \subset L^2(\Omega\times(0,T))$, namely:

$$\iota_+(f) = \begin{cases} 0 & \text{if } f \geq 0 \text{ a.e. in } \Omega\times(0,T) \\ +\infty & \text{otherwise.} \end{cases}$$

The statement of Proposition 3.2 still holds in the case of the minimization problem (3.12): it is easy to check that the Bregman distance associated with $R(\cdot)$ is the same of the Bregman distance associated with $R(\cdot) + \iota_+(\cdot)$, when the arguments are all non-negative.

3.4. Sampled measurements: convergence rates. In many practical instances, full measurements are not at disposal, and only a subsampled version of g^δ is available.
Therefore, we now focus on the case of subsampled measurements, and in particular those sampled according to a random procedure. In detail, we assume that, at almost every time t, instead of the full measurement $(\mathcal{A}f)(t)$ which should be a function from $\mathcal{U}$ to $\mathcal{V}$, and particularly in $\tilde{Y} = L^2(\mathcal{U};\mathcal{V})$, we only have N point evaluations of it, corresponding to the values $u_1(t),\dots,u_N(t)$ which are randomly sampled. Since the point evaluation of a function in $\tilde{Y}$ is not well-defined, we introduce the following assumptions on the operator $\mathcal{A}$:

(O1) $\mathcal{A} : X \to \mathcal{Z}$ where $\mathcal{Z} = L^2(0,T;\mathcal{C}(\mathcal{U},\mathcal{V}))$.
(O2) $\mathcal{A}$ is continuous from X to $\mathcal{Z}$ and $\|\mathcal{A}\|_{X\to\mathcal{Z}} \leq 1$.

Assumptions (O1) allows to evaluate $(\mathcal{A}f)(t)$ at any point $u \in \mathcal{U}$, for $f \in X$, whereas (O2) will be useful later.

Let us consider again our motivating example, introduced in subsection 3.1 (see Step I, Radon transform). Our goal is to deal with a subsampled version of the sinogram $\mathcal{R}f$, in which the angles are randomly selected.

Motivating Example, Step II (Semi-discrete Radon transform): *The Radon transform $\mathcal{R}$ does not allow for a continuous point-wise sampling in the variable θ. Following [7, 12], a possible alternative strategy is to consider a semi-discrete version of the Radon transform (by introducing a discretization of the variable s), which ensures higher regularity with respect to the angular variable θ than the operator $\mathcal{R}$. In particular, we set the variable s in a discrete space which corresponds to modeling the X-ray attenuation measurements with a finite-accuracy detector, consisting of N_{dtc} cells. To this end, we introduce a uniform partition $\{I_1, \dots, I_{N_{dtc}}\}$ of the interval $(-\bar{s}, \bar{s})$, where we denote by s_i the midpoint of each interval I_i and take a smooth positive function ϱ of compact support within $(-1, 1)$ such that $\int_{-1}^{1} \varrho(x)dx = 1$. The semi-discrete Radon transform is then an operator $\mathcal{R}^{sd}\colon L^2(\Omega) \to L^2([0, 2\pi); \mathbb{R}^{N_{dtc}})$ such that, for any f and θ, each component of the vector $(\mathcal{R}^{sd} f)(\theta) \in \mathbb{R}^{N_{dtc}}$ can be written as:*

$$[(\mathcal{R}^{sd} f)(\theta)]_i = \int_{I_i} \int_{\mathbb{R}} f(s\theta + \sigma\theta^{\perp}) \varrho\left(\frac{s - s_i}{|I_i|}\right) d\sigma ds = \int_{\mathbb{R}^2} f(x) \varrho_i(x, \theta) dx, \tag{3.13}$$

being $\varrho_i(x, \theta) = \varrho\left(\frac{x_1 \cos(\theta) + x_2 \sin(\theta) - s_i}{|I_i|}\right)$. It is easy to check (see [7, Section 6.1]) that if $f \in L^2(\Omega)$, then its semi-discrete sinogram $\mathcal{R}^{sd} f$ belongs to the space $\mathcal{C}((0, 2\pi); \mathbb{R}^{N_{dtc}})$.

Finally, when considering the dynamic operator $\mathcal{R}^{sd}$ associated with the semi-discrete Radon transform $\mathcal{R}^{sd}$, we introduce local averages in time, to ease the sampling process. To this end, we introduce another smooth, positive, real-valued function $\overline{\varrho}$ such that $\operatorname{supp}(\overline{\varrho}) \subset (-\overline{\eta}, \overline{\eta})$ and $\int_{-\overline{\eta}}^{\overline{\eta}} \overline{\varrho}(\tau)d\tau = 1$, and define $\mathcal{R}^{sd}\colon L^2(\Omega \times (0, T)) \to Y$ as:

$$(\mathcal{R}^{sd} f)(t) = \int_0^T \mathcal{R}^{sd}(f(\cdot, \tau))\overline{\varrho}(t - \tau)d\tau = \mathcal{R}^{sd}\left(\int_0^T f(\cdot, \tau)\overline{\varrho}(t - \tau)d\tau\right), \quad t \in (0, T). \tag{3.14}$$

It is straightforward to verify that $\mathcal{R}^{sd}$ satisfies the assumptions (O1) and (O2) by setting $\mathcal{U} = [0, 2\pi)$ and $\mathcal{V} = \mathbb{R}^{N_{dtc}}$ for any $X \subset L^2(\Omega \times (0, T))$.

Note that we introduce the smooth partitions of the spatial offset s and time t for mainly technical reasons. However, in practice both of these quantities are always averaged as real detector elements have finite resolution and require some time to collect sufficient photon statistics to make a measurement.

We now specify the statistical measurements model for both sampling and noise. In this section, we only consider a statistical noise scenario. We introduce two random processes (for the sake of clarity, we also show their dependence on the random event ω living in a suitable probability space). The first random process, $\mathbf{u}(t, \omega) = (u_n(\omega, t))_{n=1}^N$, is defined in $(0, T)$ and takes values in $\mathcal{U}_N$; the second one, $\boldsymbol{\epsilon}(t, \omega) = (\epsilon_n(\omega, t))_{n=1}^N$, is defined in $(0, T)$ and takes values in $\mathcal{V}_N$. Here, $\mathcal{U}_N$ and $\mathcal{V}_N$ denote the product spaces of $\mathcal{U}$ and $\mathcal{V}$ with themselves N times. We

consider the following scalar product in $\mathcal{V}_N$:

$$\langle \mathbf{v}, \tilde{\mathbf{v}} \rangle_{\mathcal{V}_N} = \frac{1}{N} \sum_{n=1}^{N} \langle v_n, \tilde{v}_n \rangle_{\mathcal{V}}.$$

Furthermore, we assume the following:

(M1) $\boldsymbol{\epsilon}(\omega, \cdot) \in L^2(0,T;\mathcal{V}_N)$ for a.e. ω.
(M2) For a.e. $t \in (0,T)$ and for $n = 1, \dots, N$, the random variables $\epsilon_n(t)$ are independent and identically distributed as ν, a random variable on $\mathcal{V}$.
(M3) ν is such that $\mathbb{E}[\nu] = 0$ and ν is sub-Gaussian with $\big\| \|\nu\|_{\mathcal{V}} \big\|_{\mathrm{sG}} \leq 1$.
(M4) $\mathbf{u}(\omega, \cdot)$ is a bounded and measurable function from $(0,T)$ to $\mathcal{U}_N$, for a.e. ω.
(M5) For a.e. $t \in (0,T)$ and for $n = 1, \dots, N$, the random variables $u_n(t)$ are independent and identically distributed as μ, a random variable on $\mathcal{V}$.
(M6) Random variables ν and μ are independent.

Finally, we can state an analogue of the inverse problem (3.1) for the case of random subsampled measurements:

$$\text{(3.15)} \quad \text{given } g_N^\delta \in L^2(0,T;\mathcal{V}_N) \ : \ g_N^\delta(t) = (\mathcal{A}_{\mathbf{u}} f^\dagger)(t) + \delta \boldsymbol{\epsilon}(t) \ \text{ a.e. } t \in (0,T), \quad \text{recover } f^\dagger \in X.$$

In particular, the sampling operator:

$$\mathcal{A}_{\mathbf{u}} \colon X \to L^2(0,T;\mathcal{V}_N), \qquad (\mathcal{A}_{\mathbf{u}} f)(t) = \big((\mathcal{A} f)(t)(u_n(t)) \big)_{n=1}^N \quad \text{a.e. } t \in (0,T)$$

is well-defined due to our assumptions. We can regard this operator as the composition $\mathcal{A}_{\mathbf{u}} = S_{\mathbf{u}} \mathcal{A}$, being $S_{\mathbf{u}} \colon \mathcal{Z} \to L^2(0,T;\mathcal{V}_N)$ the sampling operator associated with the vector $\mathbf{u} \in \mathcal{U}_N$:

$$\text{(3.16)} \qquad (S_{\mathbf{u}} g(t))_n = g(t)(u_n) \quad \text{a.e. } t \in (0,T).$$

To solve (3.15) we want to use the following regularization strategy:

$$\text{(3.17)} \qquad f_{\alpha,N}^\delta = \operatorname*{argmin}_{f \in X} J_{\alpha,N}^\delta(f) := \operatorname*{argmin}_{f \in X} \left\{ \frac{1}{2} \, \|\mathcal{A}_{\mathbf{u}} f - g_N^\delta\|_{L^2(0,T;\mathcal{V}_N)}^2 + \alpha R(f) \right\},$$

with R as in (3.6). The existence and uniqueness of $f_{\alpha,N}^\delta$ can be proved following the same procedure as in Proposition 3.1. In particular, thanks to a more refined selection principle suited for product spaces (see [95, Lemma 6.23]), it can be shown that $f_{\alpha,N}^\delta$ is a random variable in $L^2(0,T;\mathcal{V}_N)$, from which the proof follows.

Similarly to the fully sampled case, we are interested in deriving convergence estimates for the regularized problem (3.17), specifically for the quantity $\mathbb{E}[D_R(f_{\alpha,N}^\delta, f^\dagger)]$. In this context, we can take into account two asymptotic behaviours: when $\delta \to 0$ and when $N \to \infty$. In contrast to the (deterministic) inverse problem literature, where one only expects convergence of $f_{\alpha,N}^\delta$ to $f^\dagger$ when the noise level δ is reduced, in the statistical scenario it is also possible to recover $f^\dagger$ for a non-vanishing noise level, by increasing the number of point evaluations. This is common in the statistical learning literature and is essentially due to the fact that the

noise affecting each evaluation is (zero-mean, additive, and) independent: thus, increasing the number of measurements helps cancel the effect of noise.

In analogy to the fully sampled case, we look for an *a priori* choice rule for α of the form $\alpha = \alpha(\delta, N)$. Unfortunately, following closely the simple strategy adopted to prove Proposition 3.2 leads to convergence rates that are provably suboptimal. Hence we propose a strategy which relies more closely on the convexity of the functional $R(f) = \frac{1}{p}\|f\|_X^p$, where X is as in (3.4), and follows the one in [7, 12]. Indeed, the minimization problem (3.17) is analogous to the one considered in [7, 12], the main difference being the presence of the Bochner norm $\|\cdot\|_{L^2(0,T;\mathcal{V}_N)}$, which integrates the mismatch in time. As a consequence, we can take advantage of the results proved in [7, 12], with slight formal modifications. For an in-depth comparison in a non-dynamic setting, we refer to those publications (cf. also Appendix A where the ancillary results building up to the main result, Theorem 3.4, are stated).

Recall that X is a reflexive Banach space, continuously embedded in $L^2((0,T)\times\Omega)$ (see Proposition 2.8) and that the convex conjugate $R^\star$ is defined by:

$$R^\star\colon X^* \to \mathbb{R}, \quad R^\star(y) = \sup_{x\in X} \{\langle y, x\rangle_* - R(x)\}.$$

Next, we define the following quantity, which plays a crucial role to prove the convergence rates and it is strictly connected to source conditions:

$$\mathcal{R}(\bar\beta, \mathbf{u}; f^\dagger) = \inf_{\bar w \in L^2(0,T;\mathcal{V}_N)} \left\{ R^\star\left(r^\dagger - \mathcal{A}_{\mathbf{u}}^* \bar w\right) + \frac{\bar\beta}{2}\|\bar w\|^2_{L^2(0,T;\mathcal{V}_N)} \right\}, \tag{3.18}$$

being $r^\dagger = \nabla R(f^\dagger)$. This quantity naturally arises in one of the auxiliary results (see the right-hand side of (A.2) in Proposition A.1).

We are now ready to state the main result, establishing the convergence rates, for $N \to \infty$, of the expected value of the Bregman distance associated with the optimal choice of α and for $1 < p < 2$. In particular, the following result has been proved (in the non-dynamic case) in [7, Theorem 4.10], where one has to consider $Q = q/2$. We report it here by first isolating the general estimate (3.20), which is then specified into the two main convergence rates (3.21) and (3.22). To simplify the notation, we use the symbols $\lesssim$ and $\simeq$ to denote that an inequality and an equality hold up to a multiplicative constant independent of α, δ, N.

Theorem 3.4 ([7, Theorem 4.10 with $Q = q/2$]). *Suppose that assumptions (O1)-(O2) are satisfied and it holds that:*

$$\mathbb{E}\left[\mathcal{R}(\bar\beta, \mathbf{u}; f^\dagger)\right] \lesssim \bar\beta + N^{-\frac{q}{2}} \qquad \textit{and} \qquad \mathbb{E}\left[R^\star(\mathcal{A}_{\mathbf{u}}^*\boldsymbol{\epsilon})\right] \lesssim N^{-\frac{q}{2}}, \tag{3.19}$$

where N is the number of randomly subsampled measurements. Then, as $N \to \infty$, we have the following estimate:

$$\mathbb{E}\left[D_R(f^\delta_{\alpha,N}, f^\dagger)\right] \lesssim \alpha + \frac{\delta^2}{N\alpha^3} + N^{-1} \tag{3.20}$$

which, depending on the asymptotic regime, reads as:

- *If $\delta N \to \infty$ (and $\delta^2/N \to 0$), then:*

$$\mathbb{E}\left[D_R(f^\delta_{\alpha,N}, f^\dagger)\right] \lesssim \left(\frac{\delta^2}{N}\right)^{\frac{1}{3}} \quad \textit{for} \quad \alpha \simeq \left(\frac{\delta^2}{N}\right)^{\frac{1}{3}}. \tag{3.21}$$

- *If δN is bounded, then:*

$$\mathbb{E}\left[D_R(f^\delta_{\alpha,N}, f^\dagger)\right] \lesssim N^{-1} \quad \textit{for} \quad \alpha \simeq N^{-1}. \tag{3.22}$$

Remark 3.5 (non-negativity constraint). As in the case of full measurements, it is advantageous to incorporate in the variational problem (3.17) the information that the desired solution $f^\dagger$ is non-negative. This leads to the following minimization problem:

$$\operatorname*{argmin}_{f \in X} \left\{ \frac{1}{2} \, \|\mathcal{A}_{\mathbf{u}} f - g^\delta_N\|^2_{L^2(0,T;\mathcal{V}_N)} + \alpha R(f) + \iota_+(f) \right\}, \tag{3.23}$$

where ι_+ is the indicator function of the non-negative orthant. The statement of Theorem 3.4 still holds in the case of the minimization problem (3.23): the proof is analogous to the one outlined in Section 3 of [12].

In order to verify the conditions (3.19) of Theorem 3.4, we need to introduce the following assumptions on the exact solution $f^\dagger$ and on the operator $\mathcal{A}$. First, we have to define an operator $\mathcal{A}^*_\mu$ associated with the probability distribution governing the sample procedure in $\mathcal{U}$. Let Y_μ be the Hilbert space $Y = L^2(0,T,\tilde{Y}_\mu)$, where $\tilde{Y}_\mu$ is the space $\tilde{Y} = L^2(\mathcal{U};\mathcal{V})$, having replaced its inner product with:

$$\langle f, g \rangle_\mu = \int_{\mathcal{U}} \langle f(u), g(u) \rangle_{\mathcal{V}} \, d\mu(u).$$

Then, $\mathcal{A}^*_\mu$ is the adjoint of $\mathcal{A}$ in Y_μ. Notice in particular that, if $\mathcal{U}$ is bounded and μ is a uniform distribution on $\mathcal{U}$, then μ is a continuous distribution associated with a (constant) density equal to $1/|\mathcal{U}|$. Therefore, the inner products of Y and Y_μ only differ by a multiplicative constant $1/|\mathcal{U}|$ and $\mathcal{A}_\mu = 1/|\mathcal{U}|\mathcal{A}$. Next, we introduce the following assumptions:

(S1) The ground truth satisfies the source condition:

$$\exists \, w^\dagger \in \mathcal{Z} = L^2(0,T;\mathcal{C}(\mathcal{U},\mathcal{V})) \qquad \text{s.t.} \qquad \nabla R(f^\dagger) = \mathcal{A}^*_\mu w^\dagger.$$

Notice that this condition is essentially equivalent to the one introduced for the full-measurement case, (3.7). In particular, there is no need to verify a source condition for every sampled operator $\mathcal{A}_{\mathbf{u}}$, and the random sampling procedure only affects (3.7) by replacing the original operator $\mathcal{A}^*$ with its representation $\mathcal{A}^*_\mu$ in Y_μ.

(S2) Let $(\psi_\lambda)_{\lambda \in \Lambda}$ be the cylindrical-shearlet (Parseval) frame introduced in (2.7) and, for each $\lambda \in \Lambda$, let $X \ni \overline{\psi}_\lambda = \psi_\lambda|_{\Omega \times (0,T)}$. Then the operator $\mathcal{A}$ satisfies:

$$\|(\|\mathcal{A}\overline{\psi}_\lambda\|_{\mathcal{Z}})_\lambda\|_{\ell^q(\hat{w})} = \sum_{\lambda \in \Lambda} \tilde{w}_\lambda^{-q/p} \|\mathcal{A}\overline{\psi}_\lambda\|^q_{\mathcal{Z}} = C_q < \infty,$$

where we denoted by $\hat{w}$ the weight sequence such that $\hat{w}_\lambda = \tilde{w}_\lambda^{-q/p}$.

These assumptions allow to verify the hypotheses of Theorem 3.4, as shown by the next result.

Proposition 3.6. *The assumptions (S1) and (S2) imply that* (3.19) *is satisfied.*

Proof. The proof of Proposition 3.6 is analogous to the ones of Proposition 4.2 and 4.3 in [12], with significant but technical modifications due to the time-dependent framework. To balance self-containment and clarity, we only prove that the first bound in (3.19) holds.

In the definition (3.18) of $\mathcal{R}$, we consider $\bar{w} = S_{\mathbf{u}} w^\dagger$, being $w^\dagger$ the source condition element in (S1) and $S_{\mathbf{u}}$ the sampling operator defined in (3.16). Thus, we obtain:

$$\begin{aligned}\mathcal{R}(\bar{\beta}, \mathbf{u}; f^\dagger) &\leq R^\star(f^\dagger - \mathcal{A}_{\mathbf{u}}^* S_{\mathbf{u}} w^\dagger) + \frac{\bar{\beta}}{2}\|S_{\mathbf{u}} w^\dagger\|^2_{L^2(0,T;\mathcal{V}_N)} \\ &\leq \|\mathcal{A}_\mu^* w^\dagger - \mathcal{A}_{\mathbf{u}}^* S_{\mathbf{u}} w^\dagger\|_{X^*} + \frac{\bar{\beta}}{2}\|S_{\mathbf{u}} w^\dagger\|^2_{L^2(0,T;\mathcal{V}_N)},\end{aligned}$$

where we used (S1) and (A.5). The expectation of the second term on the right-hand side can be computed as follows:

$$\begin{aligned}\mathbb{E}_{\mathbf{u}\sim\mu^N}\left[\frac{\bar{\beta}}{2}\|S_{\mathbf{u}} w^\dagger\|^2_{L^2(0,T;\mathcal{V}_N)}\right] &= \frac{\bar{\beta}}{2}\int_{\mathcal{U}} \|S_{\mathbf{u}} w^\dagger(t,\cdot)\|^2_{L^2(0,T;\mathcal{V}_N)} d\mu^N(\mathbf{u}) \\ = \frac{\bar{\beta}}{2N}\int_{\mathcal{U}}\int_0^T \sum_{n=1}^N \|w^\dagger(t,u_n)\|^2_{\mathcal{V}} dt\, d\mu(u_n) &= \frac{\bar{\beta}}{2N}\sum_{n=1}^N \int_0^T \|w^\dagger(t,\cdot)\|^2_{Y_\mu} = \frac{\bar{\beta}}{2}\|w^\dagger\|^2_{L^2(0,T;Y_\mu)},\end{aligned}$$

and gives rise to the first term in the right-hand side of the first bound in (3.19). Instead, for the first term on the right-hand side, recalling (A.5), we have:

$$\begin{aligned}\|\mathcal{A}_\mu^* w^\dagger - \mathcal{A}_{\mathbf{u}}^* S_{\mathbf{u}} w^\dagger\|_{X^*} &= \sum_{\lambda\in\Lambda} \tilde{w}_j^{-q/p} \left|\langle (\mathcal{A}_\mu^* w^\dagger - \mathcal{A}_{\mathbf{u}}^* S_{\mathbf{u}} w^\dagger), \overline{\psi}_\lambda\rangle_{X^*\times X}\right|^q \\ = \sum_{\lambda\in\Lambda} \tilde{w}_j^{-q/p} \left|\langle w^\dagger, \mathcal{A}_\mu \overline{\psi}_\lambda\rangle_{Y_\mu} - \langle S_{\mathbf{u}} w^\dagger, \mathcal{A}_{\mathbf{u}}\overline{\psi}_\lambda\rangle_{L^2(0,T;\mathcal{V}_N)}\right|^q &= \sum_{\lambda\in\Lambda} \tilde{w}_j^{-q/p} \left|\frac{1}{N}\sum_{n=1}^N \zeta_n^\lambda\right|^q\end{aligned}$$

where we have set: $\zeta_n^\lambda = \langle w^\dagger, \mathcal{A}_\mu\overline{\psi}_\lambda\rangle_{Y_\mu} - \langle S_{u_n} w^\dagger, S_{u_n}\mathcal{A}\overline{\psi}_\lambda)\rangle_{L^2(0,T;\mathcal{V})}$.

It is easy to show that the real-valued random variables ζ_n^λ are zero-mean and also independent and identically distributed. Moreover, they are bounded random variables. Indeed, in view of (O2) and of the Cauchy–Schwarz inequality, for any $u_n \in \mathcal{U}$ we have:

$$\begin{aligned}\langle S_{u_n} w^\dagger, S_{u_n}\mathcal{A}\overline{\psi}_\lambda)\rangle_{L^2(0,T;\mathcal{V})} &= \int_0^T \langle w^\dagger(t,u_n(t)), (\mathcal{A}\overline{\psi}_\lambda)(t)(u_n(t))\rangle_{\mathcal{V}} dt \\ &\leq \int_0^T \|w^\dagger(t,\cdot)\|_{\mathcal{C}(\mathcal{U},\mathcal{V})}\|\mathcal{A}\overline{\psi}_\lambda(t)\|_{\mathcal{C}(\mathcal{U},\mathcal{V})} dt \leq \|w^\dagger\|_{\mathcal{Z}}\|\mathcal{A}\overline{\psi}_\lambda\|_{\mathcal{Z}}.\end{aligned}$$

Therefore, each ζ_n^λ takes values in the interval of size $2\|w^\dagger\|_{\mathcal{Z}}\|\mathcal{A}\psi_\lambda\|_{\mathcal{Z}}$ and we can apply the Hoeffding's inequality for bounded random variables to get:

$$\mathbb{P}\left(\left|\sum_{n=1}^N \zeta_n^\lambda\right| > t\right) \leq 2\exp\left(-\frac{t^2}{2\|w^\dagger\|^2_{\mathcal{Z}}\|\mathcal{A}\overline{\psi}_\lambda\|^2_{\mathcal{Z}}}\right).$$

In conclusion:

$$\begin{aligned}
\mathbb{E}\left[\|\mathcal{A}_\mu^* w^\dagger - \mathcal{A}_{\mathbf{u}}^* S_{\mathbf{u}} w^\dagger\|_{X^*}\right] &= \frac{1}{q}\sum_{\lambda\in\Lambda} \tilde{w}_j^{-q/p} N^{-q} \mathbb{E}\left[\left|\sum_{n=1}^{N} \zeta_n^\lambda\right|^q\right] \\
&\leq \frac{1}{q}\sum_{\lambda\in\Lambda} \tilde{w}_j^{-q/p} N^{-q} \int_0^\infty t^{q-1}\mathbb{P}\left(\left|\sum_{n=1}^{N} \zeta_n^\lambda\right| > t\right) dt \\
&\leq \frac{2}{q}\sum_{\lambda\in\Lambda} \tilde{w}_j^{-q/p} N^{-q} \int_0^\infty t^{q-1}\exp\left(-\frac{t^2}{2\|w^\dagger\|_{\mathcal{Z}}^2 \|\mathcal{A}\overline{\psi}_\lambda\|_{\mathcal{Z}}^2}\right) dt \\
&\leq \frac{2}{q}\sum_{\lambda\in\Lambda} \tilde{w}_j^{-q/p} N^{-\frac{q}{2}} \|w^\dagger\|_{\mathcal{Z}}^q \|\mathcal{A}\overline{\psi}_\lambda\|_{\mathcal{Z}}^q \int_0^\infty s^{q-1}\exp\left(-\frac{1}{2}s^2\right) ds,
\end{aligned}$$

which can be bounded by a constant, depending on q and on C_q in (S2), multiplying $N^{-\frac{q}{2}}$. ■

Thanks to Proposition 3.6, we know that Theorem 3.4 holds true if (S1) and (S2) are satisfied. The source condition (S1) depends on the unknown solution $f^\dagger$ and it is in general difficult to check. Instead, the property (S2) only depends on the operator $\mathcal{A}$ and on the frame $\{\psi_\lambda\}_\lambda$. For example, this is verified in the case of our motivating example.

Motivating Example, Step III ($\mathcal{R}^{\text{sd}}$ and cylindrical shearlets satisfy (S2)): *The following lemma shows that, for a cylindrical shearlet frame, condition (S2) is satisfied if, e.g., the forward operator $\mathcal{A}$ is the semi-discretized Radon transform* (3.14).

Lemma 3.7. *The cylindrical shearlet frame truncated to $\Omega\times(0,T)$, $\{\overline{\psi}_\lambda\}_\lambda$, and the operator $\mathcal{R}^{sd}$ defined by* (3.14) *satisfy*

$$\sum_{\lambda\in\Lambda} \hat{w}_\lambda \|\mathcal{R}^{sd}\overline{\psi}_\lambda\|_{\mathcal{Z}}^q < \infty. \tag{3.24}$$

Proof. First of all, we remark that, since $q \geq 2$ and $\hat{w}_\lambda = \tilde{w}_\lambda^{-q/p}$, being $\tilde{w}_\lambda$ uniformly bounded from below, it holds:

$$\sum_{\lambda\in\Lambda} \|\mathcal{R}^{\text{sd}}\overline{\psi}_\lambda\|_{\mathcal{Z}}^2 < \infty \quad \Rightarrow \quad \sum_{\lambda\in\Lambda} \hat{w}_\lambda \|\mathcal{R}^{\text{sd}}\overline{\psi}_\lambda\|_{\mathcal{Z}}^q < \infty.$$

We thus prove the convergence of the series on the left-hand side. Since $\mathcal{U} = [0, 2\pi)$ is one-dimensional, we can take advantage of the Sobolev embedding:

$$\|\mathcal{R}^{\text{sd}}\overline{\psi_\lambda}\|_{\mathcal{Z}}^2 = \int_0^T \|(\mathcal{R}^{\text{sd}}\overline{\psi_\lambda})(t)\|_{C((0,2\pi);\mathbb{R}^{N_{\text{dtc}}})}^2 dt \lesssim \int_0^T \|(\mathcal{R}^{\text{sd}}\overline{\psi_\lambda})(t)\|_{H^1((0,2\pi);\mathbb{R}^{N_{\text{dtc}}})}^2 dt.$$

We can therefore compute:

$$\sum_{\lambda\in\Lambda} \|\mathcal{R}^{\text{sd}}\overline{\psi_\lambda}\|_{\mathcal{Z}}^2 \lesssim \sum_{\lambda\in\Lambda} \int_0^T \sum_{i=1}^{N_{\text{dtc}}} \left(\|[(\mathcal{R}^{\text{sd}}\overline{\psi_\lambda})(t)]_i\|_{L^2(0,2\pi)}^2 + \|\partial_\theta[(\mathcal{R}^{\text{sd}}\overline{\psi_\lambda})(t)]_i\|_{L^2(0,2\pi)}^2\right) dt$$

and, more specifically, using the definition of $\mathcal{R}^{\mathrm{sd}}$ in (3.14):

$$\begin{aligned}\sum_{\lambda\in\Lambda}\int_0^T\sum_{i=1}^{N_{\mathrm{dtc}}}\|[(\mathcal{R}^{\mathrm{sd}}\,\overline{\psi_\lambda})(t)]_i\|^2_{L^2(0,2\pi)} &= \sum_{\lambda\in\Lambda}\int_0^T\sum_{i=1}^{N_{\mathrm{dtc}}}\int_0^{2\pi}\left|\int_0^T\int_\Omega \psi_\lambda(x,s)\varrho_i(x,\theta)\overline{\varrho}(t-s)dxds\right|^2 d\theta dt\\ &= \int_0^T\sum_{i=1}^{N_{\mathrm{dtc}}}\int_0^{2\pi}\sum_{\lambda\in\Lambda}|\langle\psi_\lambda,\varrho_i(\cdot,\theta)\overline{\varrho}(t-\cdot)\rangle|^2 d\theta dt\\ &\le \int_0^T\sum_{i=1}^{N_{\mathrm{dtc}}}\int_0^{2\pi}\|\varrho_i(\cdot,\theta)\overline{\varrho}(t-\cdot)\|^2_{L^2(\Omega\times(0,T))}d\theta dt<\infty,\end{aligned}$$

and analogously:

$$\sum_{\lambda\in\Lambda}\int_0^T\sum_{i=1}^{N_{\mathrm{dtc}}}\|\partial_\theta[(\mathcal{R}^{\mathrm{sd}}\,\overline{\psi_\lambda})(t)]_i\|^2_{L^2(0,2\pi)} \le \int_0^T\sum_{i=1}^{N_{\mathrm{dtc}}}\int_0^{2\pi}\|\partial_\theta\varrho_i(\cdot,\theta)\overline{\varrho}(t-\cdot)\|^2_{L^2(\Omega\times(0,T))}d\theta dt<\infty.$$

We now have all the ingredients to numerically verify the expected convergence rates proven in Theorem 3.4 for the two asymptotic behaviours of the noise.

4. Numerical tests. To show that the theoretical convergence rates of subsection 3.3 and subsection 3.4 can be attained in practice, in this section we conduct some numerical experiments (similar to those in [12]) which aim at validating the rates. In addition, we provide some of the resulting reconstructions to highlight that while our perspective is relatively theoretical, it also leads to a practical and robust reconstruction algorithm. Since a number of repeated runs of the algorithm is needed to validate the results in expected value (see subsection 4.3), the experiments were carried out on a computational cluster[3] using 4 Intel 3.2 GHz CPU cores with 1 GB of RAM each, running MATLAB 2022a on RHEL version 8.

4.1. Discretization of the mathematical model. The problem setting for the numerical experiments is the one of the motivating example, i.e., a sparse angle dynamic tomography problem with time steps $\{t_1,\dots,t_\kappa\}\subset[0,T]$, very similar to the one in [9, 8]. To set the notation, we briefly explain how to derive a fully discrete counterpart of (3.17) in the case $\mathcal{A}=\mathcal{R}^{\mathrm{sd}}$. We discretize objects of X by means of vectors in $\mathbb{R}^{\kappa N_{\mathrm{pxl}}}$, where N_{pxl} denotes the total number of pixels involved in the discretization of Ω and κ is the number of time steps. We consider the following discrete model:

$$\boldsymbol{g}_N^\delta = \boldsymbol{g}_N^\dagger + \delta\boldsymbol{\epsilon}_N = \boldsymbol{\mathcal{R}}_{\boldsymbol{\theta}}\,\boldsymbol{f}^\dagger + \delta\boldsymbol{\epsilon}_N \tag{4.1}$$

with

$$\boldsymbol{f}^\dagger = \begin{bmatrix}\boldsymbol{f}^\dagger_{t_1}\\ \vdots\\ \boldsymbol{f}^\dagger_{t_\kappa}\end{bmatrix},\quad \boldsymbol{\mathcal{R}}_{\boldsymbol{\theta}} = \begin{bmatrix}\boldsymbol{\mathcal{R}}_{\boldsymbol{\theta},t_1} & & \\ & \ddots & \\ & & \boldsymbol{\mathcal{R}}_{\boldsymbol{\theta},t_\kappa}\end{bmatrix},\quad \boldsymbol{g}_N^\delta = \begin{bmatrix}\boldsymbol{g}^\delta_{N,t_1}\\ \vdots\\ \boldsymbol{g}^\delta_{N,t_\kappa}\end{bmatrix}.$$

[3] https://wiki.helsinki.fi/xwiki/bin/view/it4sci/

where, for $\tau = t_1, \dots, t_\kappa$, $\boldsymbol{f}_\tau^\dagger \in \mathbb{R}^{N_{\mathrm{pxl}}}$ denotes the (unknown) discrete and vectorized image at time τ, $\boldsymbol{\mathcal{R}}_{\boldsymbol{\theta},\tau} \in \mathbb{R}^{N_{\mathrm{dtc}}N \times N_{\mathrm{pxl}}}$ represents the sampled version of the Radon operator corresponding to the N randomly sampled angles $\boldsymbol{\theta}$ at time τ, $\boldsymbol{g}_{N,\tau}^\dagger \in \mathbb{R}^{N_{\mathrm{dtc}}N}$ is the subsampled sinogram at time τ and $\boldsymbol{\epsilon}_{N,\tau} \in \mathbb{R}^{N_{\mathrm{dtc}}N}$ is the noise at time τ. In the implementation, for each time step $\tau = t_1, \dots, t_\kappa$, we consider a normal distribution for the noise vector, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbb{1}_{N_{\mathrm{dtc}}N})$, where $\mathbb{1}_{N_{\mathrm{dtc}}N}$ is the identity matrix in $\mathbb{R}^{N_{\mathrm{dtc}}N \times N_{\mathrm{dtc}}N}$. To form each $\boldsymbol{\mathcal{R}}_{\boldsymbol{\theta},\tau}$, we pick $N \in \mathbb{N}$ projection angles randomly from a uniform distribution, independently for each time step τ.
Then, a regularized solution $\boldsymbol{f}_{\alpha,N}^\delta \in \mathbb{R}_+^{\kappa N_{\mathrm{pxl}}}$ is obtained by considering regularizers of the form:

$$\boldsymbol{R}(\boldsymbol{f}) = \frac{1}{p} \|\boldsymbol{K}\boldsymbol{f}\|_p^p \tag{4.2}$$

where $1 < p < 2$ and $\boldsymbol{K} \in \mathbb{R}^{c\kappa N_{\mathrm{pxl}} \times \kappa N_{\mathrm{pxl}}}$, with $c \geq 1$, depends on the sparsifying transform. Our main focus is cylindrical shearlet-based regularization where $\boldsymbol{K} = \boldsymbol{\mathcal{SH}}$ is the matrix representation of the 3D cylindrical shearlet transform detailed in subsection 2.1. In particular, thanks to the theoretical framework developed in subsection 2.3, $\boldsymbol{R}(\boldsymbol{f})$ is the discrete counterpart to the norm of the smoothness space $S_{p,p}^\beta(\mathbb{R}^3)$, where for simplicity we choose the parameter β as in Remark 2.10 to have unitary weights in the coefficients. For comparison, we also consider wavelet-based regularization, where $\boldsymbol{K} = \boldsymbol{W}$ is the matrix representation associated with the 3D wavelet transform. In this case, $\boldsymbol{R}(\boldsymbol{f})$ is equivalent to the norm of $B_p^s(\mathbb{R}^3)$, provided that $s = d\left(\frac{1}{p} - \frac{1}{2}\right)$.

Even though in this work we do not explicitly derive convergence rates for wavelet-based regularization for the dynamic setting, in view of the results in [7, 12], we expect the theoretical rates to be the same for any orthogonal wavelets. The comparison with wavelet-based regularization is motivated by its wide application in many inverse problems, including tomography, despite being suboptimal for approximating signals of dimension 2 and higher.

Finally, with the notations introduced above, the discrete counterpart of (3.17) reads as:

$$\boldsymbol{f}_{\alpha,N}^\delta = \operatorname*{argmin}_{\boldsymbol{f} \in \mathbb{R}^{\kappa N_{\mathrm{pxl}}}} \left\{ \frac{1}{2N} \left\| \boldsymbol{\mathcal{R}}_{\boldsymbol{\theta}}\, \boldsymbol{f} - \boldsymbol{g}_N^\delta \right\|_2^2 + \alpha \boldsymbol{R}(\boldsymbol{f}) \right\}. \tag{4.3}$$

To solve (4.3), we use the variable metric inexact line-search algorithm (VMILA) [4] and the full framework including a list of all the necessary toolboxes is available on Github [10]. In addition, the appendix of [12] contains a detailed explanation of a numerical implementation process which is almost directly applicable to the dynamic problem in (3.17) as well. The main difference in this case (with respect to the implementation detailed in [12]) are the implementation of the forward operator and the corresponding data, which consists of multiple time steps. However, those can be computed separately using standard implementations. In particular, the discrete Radon transform is implemented using the ASTRA Toolbox [99, 98]. Then, we use the cylindrical shearlet transform $\boldsymbol{\mathcal{SH}}$ from [31], whose implementation is available on Github [30], with a bandlimited generator function, and three scales, with $2^3, 2^3$ and 2^4 directions per scale, using zero-padding boundary conditions. Compared to the original implementation, we chose to precompute the directional filters before starting any of the iterations, which is more efficient. For wavelet-based regularization, the implementation of the

3D discrete wavelet transform $\boldsymbol{W}$ is from the Matlab wavelet toolbox [73], using Daubechies-2 filters (with symmetric, i.e., mirrored, boundary conditions) and three scales.

4.2. Test data. Our regularized reconstruction approach is tested on both simulated and measured data. Specifically, we consider the following two sets of data.

• **Cartoon phantom**: these are simulated data of a custom ellipsoid phantom (similar to the one used in [8]). The ground truth at some relevant time steps is shown in Figure 3. In particular, the intensity values of the two larger ellipsoids change linearly in the interval $[0, 1]$, while the intensities of the multiple smaller ellipsoids follow a periodic pattern. The boundaries of the phantom remain fixed over time. The spatial dimensions of the phantom are 128×128 and we simulated in total $\kappa = 32$ sparse angle sinograms (i.e., corresponding to 32 different time steps). For each time step $\tau = t_1, \dots, t_{32}$, all data projections of $\boldsymbol{g}^{\delta}_{N,\tau}$ are simulated using parallel-beam geometry from the same reference object $\boldsymbol{f}^{\dagger}_{\tau}$ at twice the spatial resolution, and the sinogram is then binned to avoid inverse crime.

• **STEMPO data set**: these are real X-ray tomography data measured with a motorized device. During the measurement process, a square section of the phantom is translated in a straight line across the fixed circular region. The data are available in Zenodo (cf. [55]) and a detailed description is reported in [54]. In particular, here we use the `stempo_seq8x180_2d_b8` data, which consists of 1440 projections in fan-beam geometry (8 full rotations with 2° equispaced angles), each from a unique time step. We approximate $\kappa = 16$ sinograms by first dividing the data into 16 batches (corresponding to 90 equispaced projections from just a half revolution or a 180° range each) and then by linearly interpolating the N random projection angles from each batch. Although the theory presented in section 3 pertains the case of parallel-beam geometry, the extension to the fan-beam case would be a matter of geometrical reparametrization. In practice, the 180° range seems to be enough even for fan-beam and allows double the temporal resolution compared to using full 360°. The spatial dimensions of the final reconstructions are 280×280.

To verify the convergence rates in expected value of Theorem 3.4, a ground truth is needed. To obtain a reliable "ground truth" in this case, we use the original (densely sampled) sinogram batches to compute a reconstruction using a variational method similar to the one in the original documentation [54]. To avoid bias (towards either cylindrical shearlets or real-valued wavelets priors), we use 3D dual-tree complex wavelets [23] for the sparsifying transform in the regularization term. The regularization parameter is chosen automatically [84] and it is not used in any of the other numerical experiments. These reconstructions are shown in Figure 4 and, for selected time steps, on the rightmost column of Figure 5 and Figure 6.

Remark 4.1 (Verification of source conditions and cylindrical cartoon-like function assumptions). The (nearly) optimal approximation result reported in section 2 holds under the assumption that the target belongs to the class of cylindrical cartoon-like functions (cf. Definition 2.5). In addition, to derive the convergence rates in section 3, a source condition assumption is needed (cf. (S1)). However, for more realistic simulations, the model of cartoon-like videos (cf. (2.12)) offers a more natural setting for the phantoms presented above. Although there is no theoretical guarantee that the same approximation rate holds in this case, we argue that the result of section 2 still offers a valuable guideline and the numerical results provide an indication that a result similar to Theorem 2.3 is expected to hold in a more general set-

ting. Regarding condition (S1) for the cartoon phantom, it was already observed in [12] that a version of a simulated phantom generated by numerically enforcing the source condition is visually almost indistinguishable from the original. This suggests that the numerical verification of the convergence rates does not differ significantly between the original phantom and its version satisfying the source condition. Instead, for STEMPO data, the source condition cannot be established or enforced because of the experimental nature of the data. Finally, a relationship has been established between a (different) class of source conditions (in particular, variational source conditions) and the theory of optimally sparse approximations for orthogonal wavelet systems. In [72], it was shown that optimally sparse nonlinear wavelet approximations of cartoon-like images automatically satisfy a variational source condition. Although this observation does not apply directly to our setting, it is plausible that a similar relationship could be established also in our case. This topic is beyond the scope of this work and is left for future investigation.

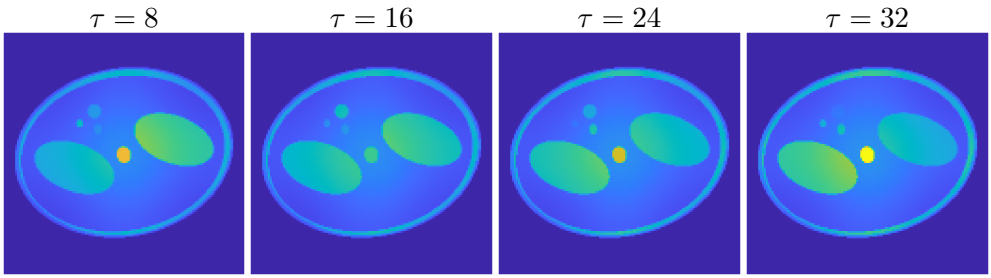

Figure 3. *Cartoon phantom ground truth at different time steps* τ*. The reconstruction size is* $128\times128\times32$.

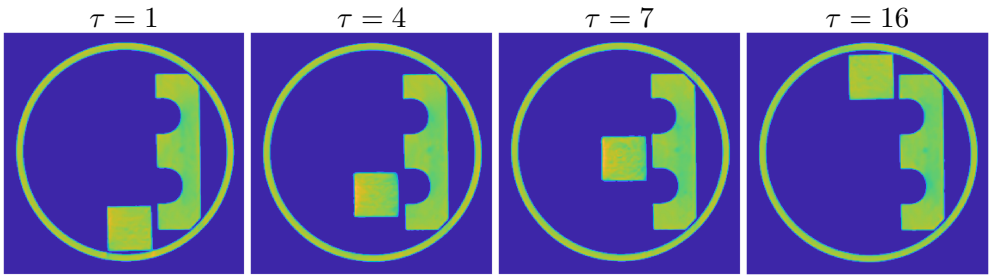

Figure 4. *Ground truth reconstruction (*$280\times280\times16$*) of the STEMPO phantom at different time steps* τ.

4.3. Results. In this subsection, we verify the expected convergence rates proven in Theorem 3.4 for the sparse dynamic CT problem (4.2)-(4.3). We focus on the statistical noise regime (cf. (3.3)), for which similarly to [7, 12] we consider the following two scenarios:

• **Decreasing noise**: in particular, $\delta \simeq N^{-1}$. In this case, the optimal parameter choice is $\alpha \simeq N^{-1}$. In particular, we choose $\delta = c_\delta N_{\min} \| \boldsymbol{\mathcal{R}} \boldsymbol{f}^\dagger \|_\infty N^{-1}$, where $c_\delta = 0.6$ for the cartoon phantom, and $c_\delta = 0.45$ for the STEMPO data. Here, $N_{\min}$ is the minimum value of N considered for the numerical experiments and it is different for the two tested data. Notice that as

N increases from $N_{\min}$ to $N_{\max}$, the corresponding noise level δ decreases from $c_\delta \| \boldsymbol{\mathcal{R}} \boldsymbol{f}^\dagger \|_\infty$ to $c_\delta N_{\min}/N_{\max} \| \boldsymbol{\mathcal{R}} \boldsymbol{f}^\dagger \|_\infty$. Finally, we set $\alpha = c_\alpha N^{-1}$, where c_α is heuristically determined for each choice of data and regularization method. The specific values of c_α and c_δ are in Table 1.

• **Fixed noise**, i.e., $\delta > 0$ constant. Since $\delta N \to \infty$, we take $\alpha \simeq N^{-1/3}$. In particular, we choose $\delta = c_\delta \| \boldsymbol{\mathcal{R}} \boldsymbol{f}^\dagger \|_\infty$, where $c_\delta = 0.03$ for the cartoon phantom and $c_\delta = 0.05$ for the STEMPO data. Then, we let $\alpha = c_\alpha N^{-1/3}$, where c_α is heuristically determined for each choice of regularization method and data. The specific values of c_α and c_δ are in Table 1.

In the cartoon phantom experiments, we test 7 distinct values of N between $N_{\min} = 24$ and $N_{\max} = 240$, and the projection angles are sampled uniformly from the interval $[0, 2\pi)$. In the STEMPO data experiments, we test 6 distinct values of N between $N_{\min} = 9$ and $N_{\max} = 60$, and the projection angles are sampled uniformly from the interval $[0, \pi)$ or $[\pi, 2\pi)$ in an alternating fashion based on the time step τ. Indeed, due to the limited amount of STEMPO data, each of the sinogram $\boldsymbol{g}^\delta_{N,\tau}$ corresponds to just 180° range of projection angles and (unlike parallel-beam) in fan-beam geometry projections from opposite directions (i.e., θ and $\theta + 180°$) are unique.

Finally, the expected values appearing in Theorem 3.4 are approximated by sample averages, computed using 5 random realizations. This means that, for each number of angles N, the reconstruction is performed 5 times, each time with a different set of N drawn angles (sampled using Matlab's `rand`) and a new realization of the noise vector.

4.3.1. Validation: $p = 3/2$. We start by considering the case $p = \frac{3}{2}$, which is fully supported by the theory developed in section 3. We compute the value of the expected Bregman distance $\mathbb{E}\left[D_{\boldsymbol{R}}(\boldsymbol{f}^\delta_{\alpha,N}, \boldsymbol{f}^\dagger)\right]$ as a function of N and report the slope of the decay rate in Table 1. In computing the Bregman distance, $\boldsymbol{f}^\delta_{\alpha,N}$ and $\boldsymbol{f}^\dagger$ are considered as 3D objects. In order to provide a quantitative assessment of the decay, we compare the theoretically predicted decay with the experimental one, which is obtained by computing the best monomial approximation cN^b of the resulting curves.

In the supplementary material (SM) we have included plots showing the decay rates for both decreasing and fixed noise regimes, using the simulated cartoon phantom in Figure SM1 and the measured STEMPO data in Figure SM2. In each plot, the value of the expected Bregman distance is indicated by a blue solid line and its monomial approximation by a black dashed line. We also report the standard deviation error bars (that is, the shaded region in each plots). Notice also that, with cylindrical shearlet regularization, there are basically no oscillations around the mean, contrary to the wavelet-based case where, especially with a lower number of projections, the variance is more pronounced. All of the predicted and numerically confirmed decay rates are also collected in to Table 1, along with the corresponding regularization and noise parameters c_α and c_δ. The first 4 rows (for $p = \frac{3}{2}$) show that the theoretical behaviour is numerically verified: $\mathbb{E}\left[D_{\boldsymbol{R}}(\boldsymbol{f}^\delta_{\alpha,N}, \boldsymbol{f}^\dagger)\right]$ decays as $N^{-1/3}$ in the fixed noise scenario, and as N^{-1} in the decreasing noise one. The results are particularly noteworthy for the STEMPO data considering that the Bregman distance is computed from an estimated ground truth object, obtained from fixed and relatively limited data with some unavoidable measurement error and whose (randomized) projections have been further interpolated.

Finally, to complement the convergence rates analysis and show that the proposed strategy leads to a practical and robust reconstruction algorithm, we report in Figure 5 some

Table 1

Theoretical and numerically approximated convergence rates, and different regularization and noise parameter values for each choice of p, noise scenario, data set, and regularization method used in the numerical tests.

p	Noise scenario	Data	Theoretical rate b $[N^b]$	Cylindrical shearlet		Daubechies-2 wavelet		c_δ
				Approx. rate	c_α	Approx. rate	c_α	
$\frac{3}{2}$	**Decreasing**	Cartoon	-1	-1.190	0.03	-1.001	0.12	0.6
		STEMPO		-0.943	0.005	-0.938	0.015	0.45
	Fixed	Cartoon	$-\frac{1}{3}$	-0.315	0.03	-0.381	0.12	0.03
		STEMPO		-0.332	0.005	-0.394	0.015	0.05
1	**Decreasing**	Cartoon	N/A	-1.053	0.001	-0.996	0.004	0.6
		STEMPO		-0.920	0.0001	-0.864	0.0003	0.45
	Fixed	Cartoon	N/A	-0.320	0.001	-0.373	0.004	0.03
		STEMPO		-0.340	0.00005	-0.314	0.0003	0.05

reconstructions from the STEMPO data for the fixed noise scenario (using both cylindrical shearlets and wavelets). All the reconstructions are qualitatively good, especially considering the very limited data. Considering the low data regime, both regularization methods suffer from some sparse-angle artifacts and produce slightly blurry images. However, wavelet-based reconstructions are distinctly noisier. Overall, the cylindrical shearlet-based approach is able to produce clearer reconstructions, even in the limit case of only 9 projection views: the thin circular boundary of the phantom, in particular, is notably sharper.

Some reconstructions of the simulated data are shown in Figure SM5 of the supplementary material, including also the relative ℓ^2 error and SSIM [106] values. Since the dynamics are less severe, these reconstruction do not show as strong movement artefacts but the quality trends (for the two regularization methods and different number of projections) match those of the real data.

4.3.2. Extension: $p = 1$. Next, we take a leap of faith and extend the numerical experiments outside the theoretically guaranteed range and set $p = 1$, which corresponds to the widespread regularization approach by minimization of the ℓ^1-norm of wavelet or shearlet coefficients to enforce sparsity. The numerical machinery from the previous subsection can be still applied, however the Bregman distance is no longer uniquely defined nor is a metric. To circumvent the first issue, as in [12], we follow the strategy in [16] to pick an element in the subdifferential ∂R. The approximate decay rates are listed in Table 1. Despite $D_{\boldsymbol{R}}$ still not being a metric, similarly to the numerical tests in [12], it seems that our choice of $D_{\boldsymbol{R}}(\boldsymbol{f}^\delta_{\alpha,N}, \boldsymbol{f}^\dagger)$ is informative enough to be able to capture the desired convergence properties of the expected Bregman distance $\mathbb{E}\left[D_{\boldsymbol{R}}(\boldsymbol{f}^\delta_{\alpha,N}, \boldsymbol{f}^\dagger)\right]$.

The plots showing the resulting decay rates are included in the supplementary material: in Figure SM3 (simulated cartoon phantom) and in Figure SM4 (measured STEMPO data).

As in the previous subsection, some reconstructions from the STEMPO data for the fixed noise scenario are shown in Figure 6 (using both cylindrical shearlets and wavelets). Despite the very limited data, the reconstructions are relatively sharp: both regularization methods are able to reconstruct well the static sections of the phantom and, with cylindrical shearlets,

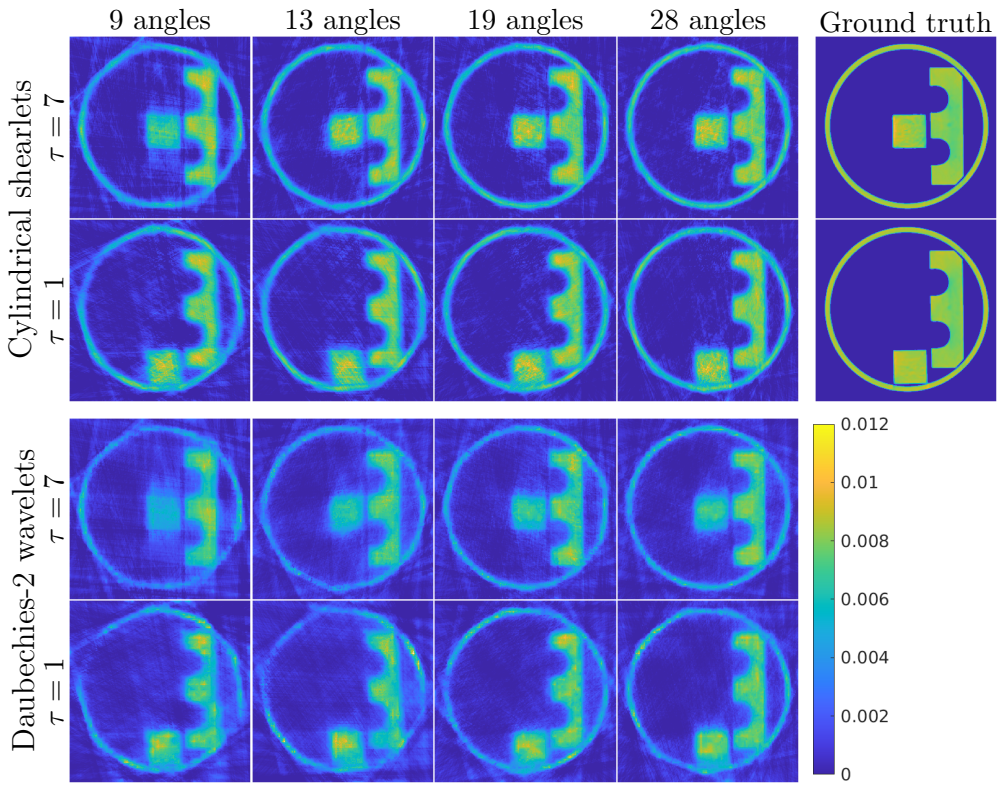

Figure 5. *Cylindrical shearlet and wavelet reconstructions from selected time steps and angular samples, fixed noise,* $p = \frac{3}{2}$*. STEMPO data.*

there are notably less noise and sparse-angle artifacts in the reconstructions. Instead, the wavelet regularized images show point-like artifacts even with a relatively large number of projections. The cylindrical shearlets produce better results across all N, even though the translated square section produces slightly visible trajectory artifacts.

Finally, in Figure 7 the pixelwise variance of the different reconstruction samples is shown for both regularization methods and values of p. Only one time step ($\tau = 7$) is considered, for the sole fixed noise scenario. This illustrates the effect of randomly picking the projection angles for each measurement setting: as one should expect, with only 9 angles there is more variance whereas with 28 projections the reconstruction process is more robust and the results vary less. Similar results with 13 and 19 angles are available in the supplementary material (Figure SM6). For $p = 3/2$ (first and second column) the differences are relatively small and overall there is less variance in every test case compared to same instances with $p = 1$. For $p = 1$ (third and fourth column) the differences are more noticeable. Cylindrical shearlet-based reconstructions exhibit a larger variance for the dynamic object (rectangle in the middle) while the static parts are more uniform. Instead, wavelet-based reconstructions show significant differences in both the dynamic and static parts of the reconstructions. This was somewhat expected given the larger oscillations of the wavelet samples in Figures SM1 to SM4.

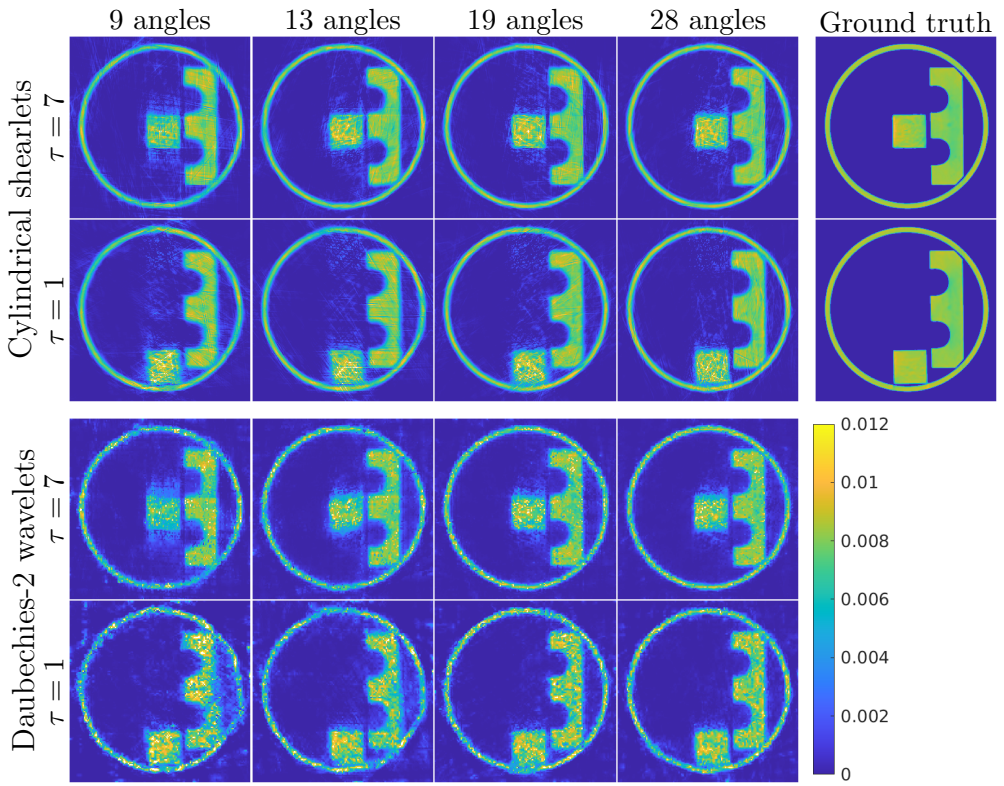

Figure 6. *Cylindrical shearlet and wavelet reconstructions from selected time steps and angular samples, fixed noise, $p = 1$. STEMPO data.*

5. Discussion and conclusions. We have demonstrated the application of the cylindrical shearlet transform in variational regularization of spatio-temporal data to solve a class of ill-posed dynamic tomography problems. Our choice is motivated by the fact that cylindrical shearlets provide (nearly) optimally sparse approximations for the class of cylindrical cartoon-like functions, which is suitable to model our desired application. We also introduce the class of cartoon-like videos, to model more realistic movements, but a proof of (nearly) optimally sparse approximations is left to future work. Moreover, we extended the theory of shearlet decomposition spaces to include cylindrical shearlets in order to use their decomposition space norms as well-defined regularization functionals with a variety of weights and any $p > 0$. We examined these regularization strategies and characterized their convergence properties (including in the statistical inverse learning framework) for dynamic tomography problems. In the full measurements case, we provided convergence rates as the noise level goes to zero, considering both deterministic and random noise conditions. Then, we considered the situation where only a limited number of imperfect samples are obtained though a random process, in our case a limited number of random X-ray projections of the changing target, and proved bounds for the error decay as the number of measurements grows. We provided rates for two distinct cases: where the amount of noise remains fixed and where its severity decreases with more samples. Our current methods of obtaining the convergence rates are limited to

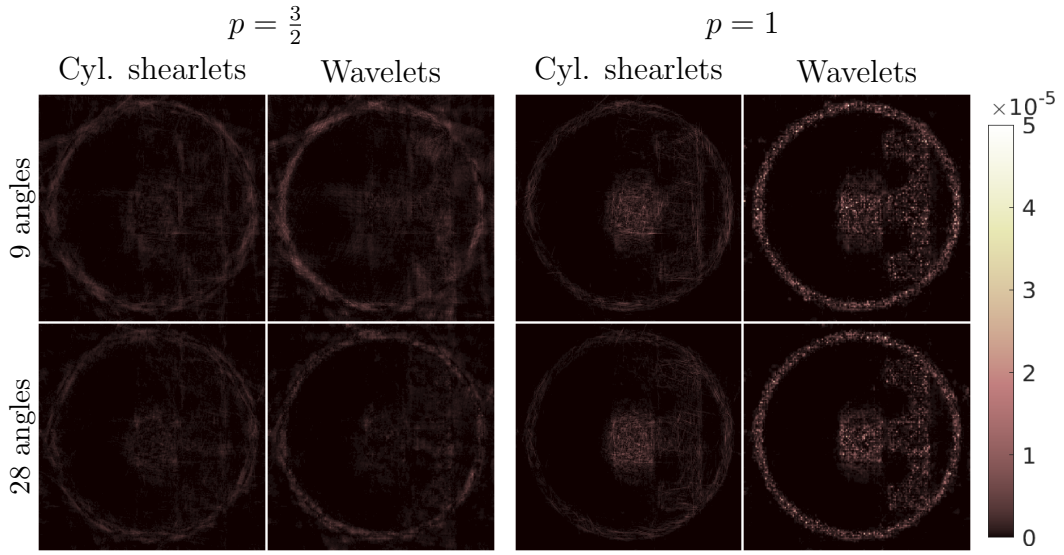

Figure 7. *Pixelwise variance between 5 reconstructions samples of the STEMPO data. Columns show different regularization methods and p-norms, rows show different number of projection angles N. Only the results from time step $\tau = 7$ and fixed noise are shown.*

$p > 1$ and an interesting future development would be to show similar rates for $p = 1$ (i.e., the usual choice for sparsity promoting regularization). However, in addition to confirming the theoretical rates, our numerical studies indicate that a similar behavior is likely to be true for $p = 1$.

It is particularly interesting how well most of the numerical experiments follow the predicted theoretical rates with the measured STEMPO data, even if we do not have access to arbitrarily many, truly random and noise free projections nor the ground truth object. In addition, the actual reconstruction algorithm seems robust and effective. Perhaps similar convergence rates are viable also to researchers applying imaging methods (even other than tomography) and not just purely theoretical results.

Although possible, we do no comparisons with traditional shearlets. Firstly, as observed after Theorem 2.3, their approximation properties are expected to fall in between wavelets and cylindrical shearlets but at a higher computational cost; secondly, the relatively small temporal resolution of the data noticeably limits their reliability in practice (especially for the STEMPO data). With cylindrical shearlets this is not an issue as the directional filtering is only performed on the larger, spatial dimensions.

Acknowledgments. The authors thank the Finnish Computing Competence Infrastructure (FCCI) for supporting this project with computational and data storage resources. TAB and LR are members of INdAM-GNCS and are funded by the European Union - NextGeneration EU through the Italian Ministry of University and Research as part of the PNRR – M4C2, Investment 1.3 (MUR Directorial Decree no. 341 of 03/15/2022), FAIR "Future Partnership Artificial Intelligence Research", Proposal Code PE00000013 - CUP J33C22002830006). TH acknowledges the support of the Vilho, Yrjö and Kalle Väisälä Foundation, the Finnish Foun-

dation for Technology Promotion, the Centre of Excellence in Inverse Modelling and Imaging and the Flagship of Advanced Mathematics for Sensing, Imaging and Modelling. DL acknowledges support of Simons Foundation grant MPS-TSM-00002738.

Appendix A. Ancillary results for section 3.

We report here ancillary results leading to Theorem 3.4. While their statement is analogous to those in [7, 12], we state them here for the specific case of the dynamic setting. We only sketch the proofs since they follow closely the arguments in [7].

The starting point to derive the converge rates in Theorem 3.4 is the deterministic setting, i.e., by fixing a realization of the noise $\boldsymbol{\epsilon}$ and of the sampling pattern $\mathbf{u}$. We begin with the optimality criterion associated with (3.5), which reads as:

$$\mathcal{A}_{\mathbf{u}}^*(\mathcal{A}_{\mathbf{u}} f_{\alpha,N}^{\delta} - \mathbf{g}_N^{\delta}) + \alpha \nabla R(f_{\alpha,N}^{\delta}) = 0. \tag{A.1}$$

Since the simple strategy adopted in (the proof of) Proposition 3.2 would deliver suboptimal convergence rates, we use instead the Fenchel-Young's inequalities. Notice that we do not require any source condition on $f^\dagger$ yet. This leads to the following result, which is derived exactly as in [7, proposition 3.2], where for simplicity we consider the choice $\Gamma_1 = \widetilde{\gamma}_1 \,\mathrm{Id}$ and $\Gamma_2 = \widetilde{\gamma}_2 \,\mathrm{Id}$, where $\widetilde{\gamma}_1, \widetilde{\gamma}_2 \in \mathbb{R}$ are positive constants and Id is the identity operator.

Proposition A.1 ([7, Proposition 3.2]). *The regularized solution $f_{\alpha,N}^{\delta}$ given by (3.5) satisfies*

$$\begin{aligned} D_R(f_{\alpha,N}^{\delta}, f^\dagger) \leq \inf_{\bar{w} \in L^2(0,T;\mathcal{V}_N)} & \left(R^\star\left(\frac{1}{\widetilde{\gamma}_1}(r^\dagger - \mathcal{A}_{\mathbf{u}}^* \bar{w}) \right) + \frac{\alpha}{2} \|\bar{w}\|^2_{L^2(0,T;\mathcal{V}_N)} \right) \\ & + R(\widetilde{\gamma}_1(f^\dagger - f_{\alpha,N}^{\delta})) + \frac{1}{\alpha}\left(R^\star\left(\frac{\delta}{\widetilde{\gamma}_2} \mathcal{A}_{\mathbf{u}}^* \boldsymbol{\epsilon} \right) + R\left(\widetilde{\gamma}_2 (f^\dagger - f_{\alpha,N}^{\delta}) \right) \right), \end{aligned} \tag{A.2}$$

where $r^\dagger = \nabla R(f^\dagger)$ and $\widetilde{\gamma}_1, \widetilde{\gamma}_2 \in \mathbb{R}$ are positive constants.

Now, due to the definition of the norm $\|f\|_X = \|\mathcal{SH} f\|_{\ell^p(\tilde{w})}$, we can deduce that, as the space ℓ^p, X is p-smooth and 2-convex (see [92, Definitions 2.32-2.33]). Hence, to estimate the second and the last term on the right hand side of (A.2), as in [7, Section 4], we rely on the p-homogeneity of the functional R, as well as on tools from convex analysis, such as the Xu-Roach's inequalities [108] for the 2-convex Banach space X. This leads to the next lemma.

Lemma A.2 ([7, Lemma 4.1]). *Let $f, \tilde{f} \in X$. For $1 < p < 2$ it holds that*

$$\widetilde{\gamma}^p R(f - \tilde{f}) \leq C\left(1 - \frac{p}{2}\right) \widetilde{\gamma}^{\frac{2p}{2-p}} \max\left\{ R(f), R(\tilde{f}) \right\} + \frac{p}{2} D_R(f, \tilde{f}), \tag{A.3}$$

for some $C > 0$ depending on p with any $\widetilde{\gamma} > 0$.

The first term on the right hand side of (A.3) can be treated through the following estimate.

Proposition A.3 ([7, Proposition 3.1, second estimate]). *The unique minimizer $f_{\alpha,N}^{\delta}$ of $J_{\alpha,N}^{\delta}$ satisfies, for some constant $C > 0$,*

$$R(f_{\alpha,N}^{\delta}) \leq C\left(R(f^\dagger) + \left(\frac{\delta}{\alpha}\right)^{\frac{p}{p-1}} R^\star(\mathcal{A}_{\mathbf{u}}^* \boldsymbol{\epsilon}) \right). \tag{A.4}$$

The following result combines all the previous estimates to provide a deterministic bound on the Bregman distance between $f^{\delta}_{\alpha,N}$ and $f^{\dagger}$. It relies also on the fact that, for the choice of R outlined in (3.6), it holds that:

$$R^{\star}(y) \leq \frac{1}{q}\|y\|^{q}_{X^*} = \frac{1}{q}\sum_{\lambda\in\Lambda} \hat{w}_{\lambda}|\langle y, \overline{\psi}_{\lambda}\rangle_{X^*\times X}|^{q} \ , \tag{A.5}$$

where $\overline{\psi}_{\lambda} = \psi_{\lambda}|_{\Omega\times(0,T)}$ are the truncated shearlet frame elements, q is the Hölder conjugate of p and $\hat{w}_{\lambda} = (\tilde{w}_{\lambda})^{-q/p}$. This can be proved via [7, Proposition 4.1], using the fact that, in this case, the shearlet transform $\mathcal{SH}$ is associated with a Parseval frame, hence $\mathcal{SH}^{\dagger} = \mathcal{SH}^{*}$.

Theorem A.4 ([7, Theorem 4.3, (i)]). *The regularized solution $f^{\delta}_{\alpha,N}$ given by (3.5) satisfies*

$$D_R(f^{\delta}_{\alpha,N}, f^{\dagger}) \leq \widetilde{C}_p \left[\widetilde{\gamma}_1^{-q}\mathcal{R}(\alpha\widetilde{\gamma}_1^{q}, \mathbf{u}; f^{\dagger}) + H(\alpha,\delta,\widetilde{\gamma}_1,\widetilde{\gamma}_2)R^{\star}(\mathcal{A}^{*}_{\mathbf{u}}\boldsymbol{\epsilon}) + \left(\widetilde{\gamma}_1^{p} + \frac{\widetilde{\gamma}_2^{p}}{\alpha}\right)^{\frac{2}{2-p}} R(f^{\dagger})\right]$$

for arbitrary $\widetilde{\gamma}_1, \widetilde{\gamma}_2 > 0$, where $\widetilde{C}_p > 0$ is a constant independent of α, δ, N and

$$H(\alpha,\delta,\widetilde{\gamma}_1,\widetilde{\gamma}_2) = \frac{\delta^{q}}{\alpha\widetilde{\gamma}_2^{q}} + \left(\widetilde{\gamma}_1^{p} + \frac{\widetilde{\gamma}_2^{p}}{\alpha}\right)^{\frac{2}{2-p}} \left(\frac{\delta}{\alpha}\right)^{q}.$$

The statement (and the proof) of Theorem 3.4 follows from the last result by taking the expectation of Bregman distance associated with the optimal choice of α.

REFERENCES

[1] S. Arridge, P. Fernsel, and A. Hauptmann, *Joint reconstruction and low-rank decomposition for dynamic inverse problems*, Inverse Problems and Imaging, 16 (2022), pp. 483–523.

[2] G. Blanchard and N. Mücke, *Optimal rates for regularization of statistical inverse learning problems*, Foundations of Computational Mathematics, 18 (2018), pp. 971–1013.

[3] S. E. Blanke, B. N. Hahn, and A. Wald, *Inverse problems with inexact forward operator: iterative regularization and application in dynamic imaging*, Inverse Problems, 36 (2020), p. 124001.

[4] S. Bonettini, I. Loris, F. Porta, and M. Prato, *Variable metric inexact line-search-based methods for nonsmooth optimization*, SIAM journal on optimization, 26 (2016), pp. 891–921.

[5] L. Borup and M. Nielsen, *Frame decomposition of decomposition spaces*, Journal of Fourier Analysis and Applications, 13 (2007), pp. 39–70.

[6] C. Brandt, T. Kluth, T. Knopp, and L. Westen, *Dynamic image reconstruction with motion priors in application to 3D magnetic particle imaging*, arXiv preprint arXiv:2306.11625, (2023).

[7] T. A. Bubba, M. Burger, T. Helin, and L. Ratti, *Convex regularization in statistical inverse learning problems*, Inverse Problems and Imaging, 17 (2023), pp. 1193–1225.

[8] T. A. Bubba, G. Easley, T. Heikkilä, D. Labate, and J. P. Rodriguez Ayllon, *Efficient representation of spatio-temporal data using cylindrical shearlets*, Journal of Computational and Applied Mathematics, 429 (2023), p. 115206.

[9] T. A. Bubba, T. Heikkilä, H. Help, S. Huotari, Y. Salmon, and S. Siltanen, *Sparse dynamic tomography: a shearlet-based approach for iodine perfusion in plant stems*, Inverse Problems, 36 (2020), p. 094002.

[10] T. A. Bubba, T. Heikkilä, and L. Ratti, *randDynTomo*, 2023, https://github.com/tommheik/randDynTomo. Accessed 21.11.2023.

[11] T. A. Bubba, F. Porta, G. Zanghirati, and S. Bonettini, *A nonsmooth regularization approach based on shearlets for Poisson noise removal in ROI tomography*, Appl. Math. Comput., 318 (2018), pp. 131 – 152.
[12] T. A. Bubba and L. Ratti, *Shearlet-based regularization in statistical inverse learning with an application to X-ray tomography*, Inverse Problems, 38 (2022), p. 054001.
[13] M. Burger, H. Dirks, L. Frerking, A. Hauptmann, T. Helin, and S. Siltanen, *A variational reconstruction method for undersampled dynamic x-ray tomography based on physical motion models*, Inverse Problems, 33 (2017), p. 124008.
[14] M. Burger, T. Helin, and H. Kekkonen, *Large noise in variational regularization*, Transactions of Mathematics and its Applications, 2 (2018), p. tny002.
[15] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse problems, 20 (2004), p. 1411.
[16] M. Burger, E. Resmerita, and L. He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing, 81 (2007), pp. 109–135.
[17] E. J. Candès and D. L. Donoho, *Curvelets and reconstruction of images from noisy Radon data*, in Proc. SPIE, Wavelet applications in signal and image processing VIII, vol. 4119, 2000, pp. 108–118.
[18] E. J. Candes, D. L. Donoho, et al., *Curvelets: A surprisingly effective nonadaptive representation for objects with edges*, Department of Statistics, Stanford University Stanford, CA, USA, 1999.
[19] D. Carando, S. Lassalle, and P. Schmidberg, *The reconstruction formula for Banach frames and duality*, Journal of Approximation Theory, 163 (2011), pp. 640–651.
[20] P. G. Casazza, D. Han, and D. R. Larson, *Frames for Banach spaces*, Contemporary Mathematics, 247 (1999), pp. 149–182.
[21] A. Chambolle, R. A. De Vore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process., 7 (1998), pp. 319–335.
[22] C. Chen, B. Gris, and O. Öktem, *A new variational model for joint image reconstruction and motion estimation in spatiotemporal imaging*, SIAM Journal on Imaging Sciences, 12 (2019), pp. 1686–1719.
[23] H. Chen and N. Kingsbury, *Efficient registration of nonrigid 3-d bodies*, IEEE Trans. Image Process., 21 (2011), pp. 262–272.
[24] Z. Chen, X. Jin, L. Li, and G. Wang, *A limited-angle CT reconstruction method based on anisotropic TV minimization*, Phys. Med. Biol., 58 (2013), pp. 2119–41.
[25] F. Colonna, G. Easley, K. Guo, and D. Labate, *Radon transform inversion using the shearlet representation*, Appl. Comput. Harmon. A., 29 (2010), pp. 232 – 250.
[26] I. Daubechies, M. Defrise, and C. D. Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.
[27] L. Desbat, S. Roux, and P. Grangeat, *Compensation of some time dependent deformations in tomography*, IEEE Trans. Image Process., 26 (2007), pp. 261–269.
[28] R. A. DeVore, *Nonlinear approximation*, Acta numerica, 7 (1998), pp. 51–150.
[29] D. L. Donoho, *Sparse components of images and optimal atomic decompositions*, Constructive Approximation, 17 (2001), pp. 353–382.
[30] G. Easley, T. Heikkilä, et al., *3d_cylind_shear – 3d Cylindrical Shearlet transform for Matlab*, 2023, https://github.com/tommheik/3d_cylind_shear. Accessed 21.11.2023.
[31] G. R. Easley, K. Guo, D. Labate, and B. R. Pahari, *Optimally sparse representations of cartoon-like cylindrical data*, The Journal of Geometric Analysis, 31 (2021), pp. 8926–8946.
[32] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375, Springer Science & Business Media, 1996.
[33] M. D. Fall, É. Barai, C. Comtat, T. Dautremer, et al., *Dynamic and clinical PET data reconstruction: A nonparametric Bayesian approach*, in 2013 IEEE ICIP, 2013, pp. 345–349.
[34] H. G. Feichtinger, *Banach spaces of distributions defined by decomposition methods, II*, Mathematische Nachrichten, 132 (1987), pp. 207–237.
[35] H. G. Feichtinger and P. Gröbner, *Banach spaces of distributions defined by decomposition methods, I*, Mathematische Nachrichten, 123 (1985), pp. 97–120.
[36] J. Frikel, *Sparse regularization in limited angle tomography*, Applied and Computational Harmonic Analysis, 34 (2013), pp. 117–141.

[37] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, *Robust principal component analysis-based four-dimensional computed tomography*, Physics in Medicine & Biology, 56 (2011), p. 3181.

[38] B. Goossens, D. Labate, and B. G. Bodmann, *Robust and stable region-of-interest tomographic reconstruction using a robust width prior*, Inverse Problems and Imaging, 14 (2020), pp. 291–316.

[39] S. Göppel, J. Frikel, and M. Haltmeier, *A complementary ℓ^1-TV reconstruction algorithm for limited data CT*, arXiv preprint arXiv:2304.11599, (2023).

[40] M. Grasmair, *Linear convergence rates for Tikhonov regularization with positively homogeneous functionals*, Inverse Problems, 27 (2011).

[41] M. Grasmair, M. Haltmeier, and O. Scherzer, *Sparse regularization with ℓ^q penalty term*, Inverse Problems, 24 (2008), p. 055020.

[42] M. Grasmair, O. Scherzer, and M. Haltmeier, *Necessary and sufficient conditions for linear convergence of ℓ^1-regularization*, Comm. Pure Appl. Math., 64 (2011), pp. 161–182.

[43] E. Gravier, Y. Yang, and M. Jin, *Tomographic reconstruction of dynamic cardiac image sequences*, IEEE transactions on image processing, 16 (2007), pp. 932–942.

[44] K. Guo and D. Labate, *Optimally sparse 3D approximations using shearlet representations*, Electronic Research Announcements, 17 (2010), p. 125.

[45] K. Guo and D. Labate, *Optimally Sparse Representations of 3D Data with C^2 Surface Singularities Using Parseval Frames of Shearlets*, SIAM Journal on Mathematical Analysis, 44 (2012), p. 851–886.

[46] K. Guo and D. Labate, *The construction of smooth parseval frames of shearlets*, Mathematical Modelling of Natural Phenomena, 8 (2013), p. 82–105, https://doi.org/10.1051/mmnp/20138106.

[47] K. Guo and D. Labate, *Optimal recovery of 3D X-ray tomographic data via shearlet decomposition*, Adv. Comput. Math., 39 (2013), pp. 227–255.

[48] B. N. Hahn, *Motion estimation and compensation strategies in dynamic computerized tomography*, Sensing and Imaging, 18 (2017), pp. 1–20.

[49] J. Hakkarainen, Z. Purisha, A. Solonen, and S. Siltanen, *Undersampled dynamic X-ray tomography with dimension reduction Kalman filter*, IEEE Transactions on Computational Imaging, 5 (2019), pp. 492–501.

[50] M. Haltmeier, *Stable signal reconstruction via ℓ^1-minimization in redundant, non-tight frames*, IEEE transactions on signal processing, 61 (2012), pp. 420–426.

[51] K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen, *Sparse Tomography*, SIAM J. Sci. Comput., 35 (2013), pp. B644–B665.

[52] U. Hampel, L. Babout, R. Banasiak, E. Schleicher, et al., *A review on fast tomographic imaging techniques and their potential application in industrial process control*, Sensors, 22 (2022), p. 2309.

[53] A. Hauptmann, O. Öktem, and C. Schönlieb, *Image reconstruction in dynamic inverse problems with temporal models*, Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, (2021), pp. 1–31.

[54] T. Heikkilä, *STEMPO–dynamic X-ray tomography phantom*, arXiv preprint arXiv:2209.12471, (2022).

[55] T. Heikkilä, *STEMPO - dynamic X-ray tomography phantom*, 2023, https://doi.org/10.5281/zenodo.8239013. v1.2.0 [data set].

[56] T. Hohage and P. Miller, *Optimal convergence rates for sparsity promoting wavelet-regularization in Besov spaces*, Inverse Problems, 35 (2019), p. 065005.

[57] A. A. Isola, A. Ziegler, T. Koehler, W. J. Niessen, and M. Grass, *Motion-compensated iterative cone-beam CT image reconstruction with adapted blobs as basis functions*, Physics in Medicine & Biology, 53 (2008), p. 6777.

[58] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, *GPU-based iterative cone-beam CT reconstruction using tight frame regularization*, Phys. Med. Biol., 56 (2011), pp. 3787–807.

[59] J. S. Jørgensen and E. Y. Sidky, *How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography*, Phil. Trans. R. Soc. A, 373 (2015), p. 20140387.

[60] A. Kadu, F. Lucka, and K. J. Batenburg, *Single-shot tomography of discrete dynamic objects*, arXiv preprint arXiv:2311.05269, (2023).

[61] H. Kekkonen, M. Lassas, and S. Siltanen, *Analysis of regularized inversion of data corrupted by white Gaussian noise*, Inverse problems, 30 (2014), p. 045009.

[62] G. Kutyniok and D. Labate, *Shearlets: multiscale analysis for multivariate data*, Birkhauser, Applied and Numerical Harmonic Analysis, 2012.

[63] D. Labate, L. Mantovani, and P. Negi, *Shearlet smoothness spaces*, Journal of Fourier Analysis and Applications, 19 (2013), pp. 577–611.

[64] S. Lan, M. Pasha, and S. Li, *Spatiotemporal Besov Priors for Bayesian Inverse Problems*, arXiv preprint arXiv:2306.16378, (2023).

[65] J. Liu, A. I. Aviles-Rivero, H. Ji, and C.-B. Schönlieb, *Rethinking medical image reconstruction via shape prior, going deeper and faster: Deep joint indirect registration and reconstruction*, Medical Image Analysis, 68 (2021), p. 101930.

[66] Y. Liu, J. Ma, Y. Fan, and Z. Liang, *Adaptive-weighted total variation minimization for sparse data toward low-dose X-ray computed tomography image reconstruction*, Phys. Med. Biol., 57 (2012), pp. 7923–56.

[67] D. A. Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, Journal of Inverse and Ill-posed Problems, 16 (2008).

[68] I. Loris, G. Nolet, I. Daubechies, and T. Dahlen, *Tomographic inversion using ℓ^1-norm regularization of wavelet coefficients*, Geophys. J. Int., 170 (2007), pp. 359–370.

[69] W. Lu and T. R. Mackie, *Tomographic motion detection and correction directly in sinogram space*, Physics in Medicine & Biology, 47 (2002), p. 1267.

[70] S. Lunz, A. Hauptmann, T. Tarvainen, C.-B. Schonlieb, and S. Arridge, *On learned operator correction in inverse problems*, SIAM Journal on Imaging Sciences, 14 (2021), pp. 92–127.

[71] P. Milanfar, *A model of the effect of image motion in the Radon transform domain*, IEEE Transactions on Image processing, 8 (1999), pp. 1276–1281.

[72] P. Miller and T. Hohage, *Maximal spaces for approximation rates in ℓ_1-regularization*, Numerische Mathematik, 149 (2021), pp. 341–374.

[73] M. Misiti, G. Oppenheim, J.-M. Poggi, Y. Misiti, et al., *Wavelet Toolbox*. Release 2022a, original release in 1997, https://www.mathworks.com/help/wavelet/. MathWorks inc.

[74] Y. D. Mutaf, J. A. Antolak, and D. H. Brinkmann, *The impact of temporal inaccuracies on 4DCT image quality*, Medical physics, 34 (2007), pp. 1615–1622.

[75] G. R. Myers, M. Geleta, A. M. Kingston, B. Recur, and A. P. Sheppard, *Improving dynamic tomography, through maximum a posteriori estimation*, in Developments in X-Ray Tomography IX, vol. 9212, SPIE, 2014, pp. 270–278.

[76] F. Natterer and F. Wübbeling, *Mathematical methods in image reconstruction*, SIAM, 2001.

[77] E. Niemi, M. Lassas, A. Kallonen, L. Harhanen, and other, *Dynamic multi-source X-ray tomography using a spacetime level set method*, J. Comput. Phys., 291 (2015), pp. 218–237.

[78] K. Niinimäki, S. Siltanen, and V. Kolehmainen, *Bayesian multiresolution method for local tomography in dental X-ray imaging*, Phys. Med. Biol., 52 (2007), pp. 6663–6678.

[79] S. Niu, Y. Gao, Z. Bian, J. Huang, et al., *Sparse-view x-ray CT reconstruction via total generalized variation regularization*, Phys. Med. Biol., 59 (2014), pp. 2997–3017.

[80] C. M. Oikonomou, Y.-W. Chang, and G. J. Jensen, *A new view into prokaryotic cell biology from electron cryotomography*, Nature Reviews Microbiology, 14 (2016), pp. 205–220.

[81] R. Otazo, E. Candes, and D. K. Sodickson, *Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components*, Magnetic resonance in medicine, 73 (2015), pp. 1125–1136.

[82] E. Papoutsellis, E. Ametova, C. Delplancke, G. Fardell, et al., *Core Imaging Library-Part II: multichannel reconstruction for dynamic and spectral tomography*, Philosophical Transactions of the Royal Society A, 379 (2021), p. 20200193.

[83] M. Pasha, A. K. Saibaba, S. Gazzola, M. I. Español, and E. de Sturler, *A computational framework for edge-preserving regularization in dynamic inverse problems*, Electronic Transactions on Numerical Analysis, 58 (2023), pp. 486–516.

[84] Z. Purisha, J. Rimpeläinen, T. A. Bubba, and S. Siltanen, *Controlled wavelet domain sparsity for x-ray tomography*, Measurement Science and Technology, 29 (2017), p. 014002.

[85] M. Rantala, S. Vänskä, S. Järvenpää, M. Kalke, M. Lassas, J. Moberg, and S. Siltanen, *Wavelet-based reconstruction for limited-angle X-ray tomography*, IEEE transactions on medical imaging, 25 (2006), pp. 210–217.

[86] N. A. B. Riis, J. Frøsig, Y. Dong, and P. C. Hansen, *Limited-data x-ray CT for underwater pipeline inspection*, Inverse Probl., 34 (2018), p. 034002.

[87] C. J. Ritchie, J. Hsieh, M. F. Gard, J. D. Godwin, Y. Kim, and C. R. Crawford, *Predictive respiratory gating: a new method to reduce motion artifacts on CT scans.*, Radiology, 190 (1994), pp. 847–852.
[88] A. Ruhlandt, M. Töpperwien, M. Krenkel, R. Mokso, and T. Salditt, *Four dimensional material movies: High speed phase-contrast tomography by backprojection along dynamically curved paths*, Scientific reports, 7 (2017), p. 6487.
[89] U. Schmitt, A. K. Louis, C. Wolters, and M. Vauhkonen, *Efficient algorithms for the regularization of dynamic inverse problems: II. Applications*, Inverse Problems, 18 (2002), p. 659.
[90] B. Schmitzer, K. P. Schäfers, and B. Wirth, *Dynamic Cell Imaging in PET With Optimal Transport Regularization*, IEEE Transactions on Medical Imaging, 39 (2020), pp. 1626–1635.
[91] T. Schuster, B. Hahn, and M. Burger, *Dynamic inverse problems: modelling – regularization – numerics*, Inverse Problems, 34 (2018), pp. 1–4.
[92] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski, *Regularization methods in Banach spaces*, vol. 10, Walter de Gruyter, 2012.
[93] E. Y. Sidky, C.-M. Kao, and X. Pan, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT*, J. X-Ray Sci. Technol., 14 (2006), pp. 119–139.
[94] E. Y. Sidky and X. Pan, *Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization*, Phys. Med. Biol., 53 (2008), pp. 4777–807.
[95] I. Steinwart and A. Christmann, *Support vector machines*, Springer Science & Business Media, 2008.
[96] S. Tan, Y. Zhang, G. Wang, X. Mou, G. Cao, Z. Wu, and H. Yu, *Tensor-based dictionary learning for dynamic tomographic reconstruction*, Physics in Medicine & Biology, 60 (2015), p. 2803.
[97] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, *Low-dose CT reconstruction via edge-preserving total variation regularization*, Phys. Med. Biol., 56 (2011), pp. 5949–67.
[98] W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, *Fast and flexible X-ray tomography using the ASTRA toolbox*, Optics express, 24 (2016), pp. 25129–25147.
[99] W. Van Aarle, W. J. Palenstijn, J. De Beenhouwer, et al., *The ASTRA toolbox: A platform for advanced algorithm development in electron tomography*, Ultramicroscopy, 157 (2015), pp. 35–47.
[100] G. Van Eyndhoven, J. Sijbers, and J. Batenburg, *Combined motion estimation and reconstruction in tomography*, in Computer Vision–ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7-13, 2012, Proceedings, Part I 12, Springer, 2012, pp. 12–21.
[101] B. Vandeghinste, B. Goossens, R. Van Holen, C. Vanhove, A. Pižurica, S. Vandenberghe, and S. Staelens, *Iterative CT reconstruction using shearlet-based regularization*, IEEE Transactions on Nuclear Science, 60 (2013), pp. 3305–3317.
[102] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*, vol. 47, Cambridge university press, 2018.
[103] F. Voigtlaender, *Embeddings of decomposition spaces*, vol. 287, American Mathematical Society, 2023.
[104] S. M. Walker, D. A. Schwyn, R. Mokso, M. Wicklein, T. Müller, M. Doube, M. Stampanoni, H. G. Krapp, and G. K. Taylor, *In vivo time-resolved microtomography reveals the mechanics of the blowfly flight motor*, PLoS biology, 12 (2014), p. e1001823.
[105] K. Wang, J. Xia, C. Li, L. V. Wang, and M. A. Anastasio, *Fast spatiotemporal image reconstruction based on low-rank matrix estimation for dynamic photoacoustic computed tomography*, Journal of biomedical optics, 19 (2014), pp. 056007–056007.
[106] 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.
[107] F. Weidling, B. Sprung, and T. Hohage, *Optimal convergence rates for Tikhonov regularization in Besov spaces*, SIAM Journal on Numerical Analysis, 58 (2020), pp. 21–47.
[108] Z.-B. Xu and G. F. Roach, *Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces*, Journal of Mathematical Analysis and Applications, 157 (1991), pp. 189–210.