SPATIO-TEMPORAL TOMOGRAPHIC IMAGING

BY

BERK ISKENDER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Doctoral Committee:

Professor Yoram Bresler, Chair
Professor Minh Do
Assistant Professor Saurabh Gupta
Professor Farzad Kamalabadi
Associate Professor Zhizhen Zhao

Abstract

Spatio-temporal imaging (also known as dynamic imaging) refers to the inverse problem of the recovery of an underlying time-varying object f at each time instant t from its undersampled measurements g obtained by a known and possibly time-varying measurement operator R_t. The problem appears in various imaging modalities, such as computed tomography (CT) and magnetic resonance imaging (MRI).

In this thesis, we particularly focus on the problem of dynamic tomography with time-sequential measurements, which is significantly ill-posed due to only having a single projection for each time instant t. Performing a direct recovery using this inconsistent set of projections is not possible. Also, the acquisition strategy (also called the angular sampling order) is another design parameter that plays an important role in the accuracy and stability of the developed methods. We formulate various methods to tackle this problem. Several of the proposed methods employ partially-separable models (PSM) to represent the underlying spatio-temporal object. PSM representation introduces a particular low-rank structure, leading to the factorization of the spatial and temporal effects. It is also parsimonious and preserves interpretability. Firstly, we propose a projection-domain PSM and analyze conditions for a unique and stable solution to this ill-posed inverse problem. The special projection-domain PSM also allows a quantitative analysis and prediction of the performance of different time-sequential acquisition schemes. Then, to enable the incorporation of a spatial regularizer, we propose an object-domain recovery algorithm using a variational formulation with PSM enforced as a soft constraint. This method uses the temporal components of the PSM recovered by the fast projection-domain method for initialization to improve and accelerate convergence. Thirdly, we propose another object-domain method that efficiently combines the PSM and the popular Regularization-by-Denoising (RED) frameworks for the first time. Convergence of the proposed algorithm is improved and accelerated by initializing the method with the spatial and temporal components of the fast projection-domain PSM. The method includes a convergence analysis. Finally, the last method proposed in this thesis combines the neural fields (NFs), or implicit neural representations (INR), with the RED framework for the first time to recover the underlying spatio-temporal object with improved

accuracy compared to its low-rank alternative, and another deep-prior-based method. The proposed optimization algorithm avoids costly gradient computations through the deep denoiser for RED updates.

To my parents and my sister, for their encouragement, love, and support.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Yoram Bresler. In addition to providing invaluable help and guidance for my research, he enlightened me with his perspectives and insights on many different topics in life. He devoted a tremendous amount of time to helping me grow as a student and as a researcher throughout this journey. Furthermore, he has always been understanding and supportive during the challenges of both Ph.D. studies and life. I hope I have been able to absorb even a fraction of what he has to offer.

I thank the past and the present members of our research group: Yuqi Li, Luke Pfister, and Ufuk Soylu for all the fruitful discussions and fun times that we had. I extend my gratitudes to my colleague and friend, Ulaş Kamacı, and to many other dear friends at UIUC, who made my time here truly memorable.

I would like to acknowledge the support for my research from the Los Alamos National Laboratories (LANL) under Subcontract No. 599416/CW13995. I especially thank Dr. Marc Klasky for making this funding possible and for his collaboration in various projects.

I am deeply grateful to my doctoral committee members Prof. Minh Do, Prof. Farzad Kamalabadi, Prof. Zhizhen Zhao, and Prof. Saurabh Gupta for their feedback on my work, as well as the knowledge and guidance provided through their courses.

I also thank Peggy Wells from the Coordinated Science Laboratory for her kind assistance with numerous matters whenever needed.

Last of all, my heartfelt and endless thanks go to my parents Ayşe and Feridun İskender, and to my sister Selin İskender. Their love, understanding, and unconditional support throughout my life have been the foundation that made this work and many other endeavors possible.

Table of contents

List of Abbreviations

PSM	Partially Separable Model
NF	Neural Field
RED	Regularization by Denoising
ADMM	Alternating Direction Method of Multipliers
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
dCT	Dynamic Computed Tomography
dMRI	Dynamic Magnetic Resonance Imaging
DIP	Deep Image Prior
CNN	Convolutional Neural Network
MLP	Multi-layer Perceptron
DNN	Deep Neural Network
MSE	Mean Squared Error
SGD	Stochastic Gradient Descent
HFEN	High Frequency Error Norm
SSIM	Structural Similarity Index
MAE	Mean Absolute Error
PSNR	Peak Signal-to-Noise Ratio

Chapter 1

Introduction

1.1 Overview

In Chapter 2, we present a novel unsupervised projection-domain recovery algorithm called ProSep for dynamic tomography where we utilize a particular partially-separable model with harmonic spatial representations. The method employs bit-reversed time-sequential sampling and the factors affecting the uniqueness and stability of the solution are analyzed. We compare the proposed method with a recent deep-prior-based approach for video reconstruction and demonstrate its advantages.

In Chapter 3, we propose an object-domain variational formulation to incorporate explicit spatial priors. This formulation enforces the partially separable model (PSM) spatio-temporal prior as a soft constraint, and employs an accurate and fast temporal initialization obtained by ProSep. The results are promising even with a simple spatial regularization technique, but we anticipate even better performance with improved regularization.

In Chapter 4, we combine the *regularization by denoising* (RED), which offers a flexible framework to exploit the powerful performance of state-of-the-art denoisers for numerous inverse problems, with PSM techniques. To the best of our knowledge, this is the first PSM-based approach using a pre-trained, learned (RED-based) spatial prior. This method considers PSM as a hard constraint, and the recovery algorithm is formulated using a bilinear ADMM. We also propose a patch-based version of the proposed algorithm. This version eliminates the need to store the complete object in the memory. Thus, the method becomes scalable for high-resolution and/or 3D dynamic objects. This change leads to negligible to no loss in reconstruction accuracy. Also, the analysis shows that a minor variation of the implemented algorithm converges to a stationary point. The computational advantages and the reconstruction accuracy improvements of the proposed method over a recent DIP-based technique are shown.

In Chapter 5, we represent the spatio-temporal object using neural fields (NF), and similarly combine the NF representation with the RED-based spatial prior. As in Chapter 4, this is the first time that the RED-based spatial prior is combined with NFs. Likewise, the NF representation is considered as a hard constraint and the resulting objective is minimized using an efficient ADMM-based algorithm. The proposed method is shown to provide improved reconstruction accuracy compared to its alternatives, including the method proposed in Chapter 4.

1.2 Linear inverse problems: Variational formulation

Many imaging inverse problems are formulated as

$$g = Rf + \eta \tag{1.1}$$

where $R : \mathbb{R}^N \to \mathbb{R}^M$ is a linear measurement operator associated with the imaging modality, where $M \ll N$, $f \in \mathbb{R}^N$ is the object to be reconstructed, and $g \in \mathbb{R}^M$ is the set of measurements corrupted by additive noise $\eta \in \mathbb{R}^M$.

The associated recovery problem is

$$\arg\min_f \mathcal{H}(f, g) \quad \text{where} \quad \mathcal{H}(f, g) = \mathcal{L}(f, g) + \rho(f), \tag{1.2}$$

where $\mathcal{L}$ is the data fidelity term measuring the fit between available measurements and the ones obtained from the estimated object, often in the form

$$\mathcal{L}(f, g) = \|g - Rf\|_2^2. \tag{1.3}$$

When the noise is modeled as additive white Gaussian, this term is proportional to the negative log-likelihood of the data, and otherwise it is simply a squared deviation term. Function ρ is a regularization term containing prior knowledge about the object f.

1.3 Problem Statement: Dynamic CT

The dynamic tomography is a challenging ill-posed inverse problem that addresses the recovery of a time-varying object from projections acquired sequentially at specific time instants. Since the object evolves in time, and too few projections (only one, in the extreme case) are acquired at any time instant, they are insufficient to reconstruct the object at any time. The problem arises in the fields of myocardial perfusion imaging [1], thoracic CT [2], imaging of fluid flow

processes [3], [4], imaging of material samples undergoing compression [5], [6], and certain microscopic tomography tasks [7]. Also, it is closely related to the dynamic MRI (dMRI) problem, which typically arises in cardiac imaging [8].

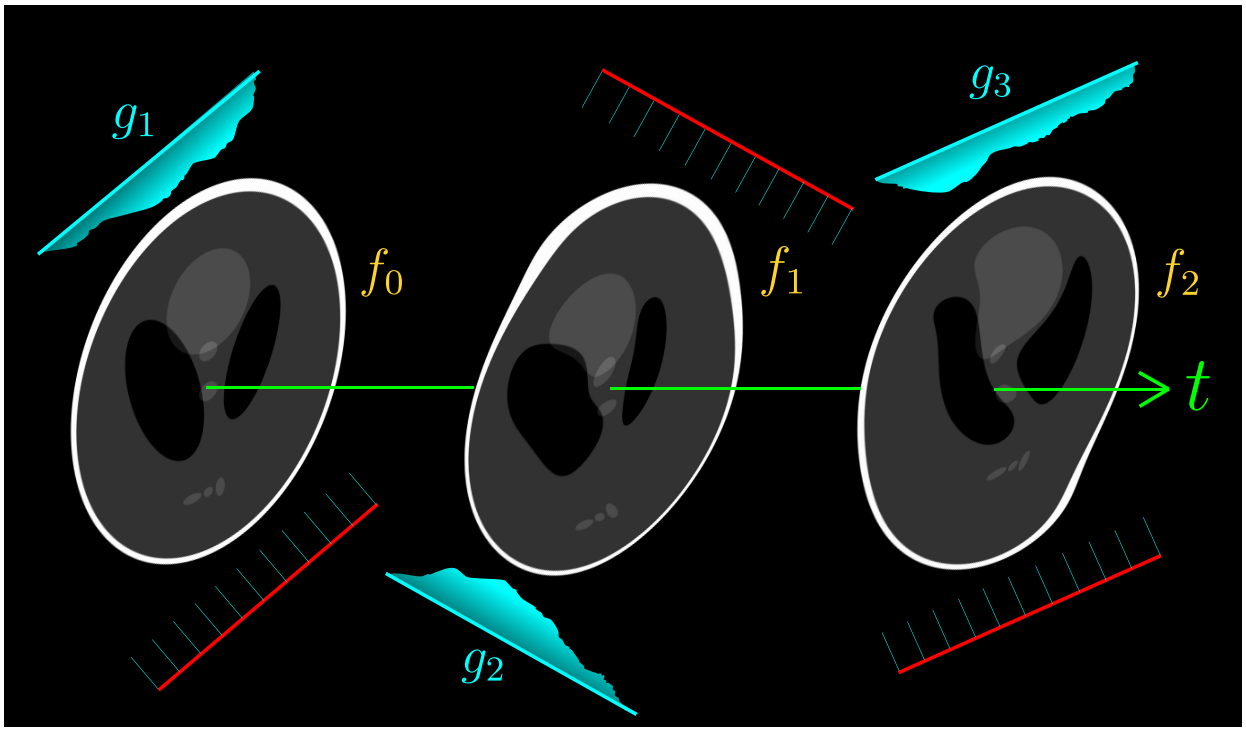


Figure 1.1: Projection acquisition geometry for time-varying tomography of the object f_t with single measurement g_t at each time instant for $t \in \{0, 1, 2\}$.

In a 2D setting, the goal in dynamic tomography is to reconstruct a time-varying object $f(\mathbf{x}, t)$, $\mathbf{x} \in \mathbb{R}^2$ vanishing outside a disc of diameter D, from its projections

$$g(\cdot, \theta, t) = \mathcal{R}_\theta\{f(\mathbf{x}, t)\} + \eta(\cdot, \theta, t) \tag{1.4}$$

obtained using the Radon transform operator $\mathcal{R}_\theta$ at angle θ where η is the additive noise. We consider time-sequential sampling of a single θ at each time, assuming that the acquisition of a projection is fast enough that the object is essentially static within the sampling time, but the variation between samples cannot be ignored. Assuming sampling uniform in time, the acquired data is

$$\{g(s, \theta_p, t_p)\}_{p=0}^{P-1}, \ \forall s, t_p = p\Delta_t, \tag{1.5}$$

where s is the offset of the line of integration from the origin (i.e., detector position), and P is the total number of projections (and temporal samples) acquired. The s variable is uniformly sampled to $\{s_j\}_{j=1}^{J}$, and we assume that this sampling is fine enough to not affect

the accuracy of the reconstruction, and therefore suppress it in the notation unless relevant. We consider the sequence $\{\theta_p\}_{p=0}^{P-1}$, with $\theta_p \in [0, 2\pi]$, which we call the angular sampling scheme, to be a free design parameter [1].

The objective is to reconstruct a time-sequence, a movie of the object $\{f(\mathbf{x}, t_p)\}_{p=0}^{P-1}$ from the time-sequential projections in (1.5). For a static object, $f(\mathbf{x})$ can be reconstructed from g using the inverse Radon transform $\mathcal{R}^{-1}$, which is well-approximated in practice by the filtered backprojection (FBP) algorithm for P large enough [9]. However, time-sequentially acquired projections (1.5) are inconsistent because different projections correspond to different objects, and a direct reconstruction using $\mathcal{R}^{-1}$ leads to severe artifacts.

1.4 Previous Work

We limit our discussion of previous work in this section to dynamic CT and dMRI, although similar or analogous methods have been or can be applied in other dynamic imaging modalities. The same can be said of the methods proposed in this thesis.

1.4.1 Dynamic MRI (dMRI)

In spite of major advances in hardware and signal acquisition methods, MRI has remained a relatively slow modality, posing significant challenges for dynamic imaging. This, coupled with the high flexibility in the acquisition strategy, has motivated extensive work on dMRI, with hundreds of papers on the subject. We highlight some key developments in the algorithm-based dMRI techniques.

Fundamentally, algorithmic approaches to dMRI are based on identifying and exploiting redundancy in the dynamic object, which enable its recovery from an incomplete, and usually sparse set of samples acquired in the joint k-space and time domain – the k-t-space. Many of the proposed methods have extensions to parallel imaging, which introduces additional such redundancy by additional hardware.

A large set of approaches rely on modeling the support of the object in the domain dual to the k-t-domain – the (x-f) domain (the space – temporal frequency domain). These approaches may be classified into three groups. The first involves generic modeling of the (x-f) support as in reduced field-of-view (FOV) techniques [10]–[12]. The second involves sparse and unknown support modeling in a transform domain, and sampling and recovery using the methods of compressed sensing [13]–[16]. The third involves adaptive sparse support

[1] An arbitrary angular sampling scheme is easily implemented in radial MRI, but also in CT systems such as micro-CT and industrial CT where the factor limiting acquisition speed is the x-ray exposure or sensor readout, rather than the rotation of the object or source.

modeling [17]–[20], where the support information is adapted to the dynamics of the imaged slice, and the theory of time-sequential sampling [21], [22] is used to design an optimal k-t sampling and reconstruction scheme. Other related approaches model prior information such as motion periodicity and related harmonic structure [23], [24], or the object's spatial and temporal correlation functions [25].

Since its introduction, the partially-separable model (PSM) [26], with a non-parametric factorized low-rank form separating the spatial and temporal structure, has been used extensively in many works on dMRI to represent the underlying spatio-temporal object. A group of methods recover the PSM in two steps: (a) estimate the temporal subspace using data (e.g., a navigator signal) sampled at low spatial, but high temporal resolution; and (b) estimate the spatial subspace using the entire acquired data set. Such methods include e.g., [26]–[30], with followup versions promoting sparsity of the object in a transform domain (e.g., [31]). Other PSM approaches use a low-rank matrix recovery formulation to estimate spatial and temporal components jointly (free of conditions or additional data acquisition enabling a separate reconstruction of the temporal subspace), e.g., [32], [33]), also with versions promoting sparsity [34], [35].

Manifold models have also been considered for dMRI. Recent such work [36] surveys previous related works, and proposes a new method that is related to PSM. The method partitions the dynamic object into temporal subsets each lying on a low-dimensional manifold, approximated by its tangent subspace. As a result, the object is represented by a temporally partitioned PSM. A simple sparsity penalty is used for the spatial representation. The spatial basis functions and linear transformations aligning the temporal subsets are determined by optimizing a non-convex objective using an ADMM algorithm. The subset temporal partitioning and temporal bases are estimated using an initial reconstruction obtained from training data. In some applications, such data may not be available.

Another set of methods, including [37]–[40], promotes low-rankness of the entire spatio-temporal object matrix or of patches thereof [41], [42] implicitly by using the nuclear norm or Schatten-p functional with $p < 1$, often promoting sparsity in a transform domain as well. Methods combining motion estimation and compensation with other priors are also used widely for dMRI, e.g., [43]–[46].

A different approach in dMRI decomposes the object into the sum of low-rank and sparse components, with the low-rankness encouraged implicitly using nuclear or a Schatten-p functional [47]–[50]. A recent related fast method [51] decomposes the object representation into the sum of three components: the mean signal; a low-rank PSM; and a residual sparse in the Fourier domain. Although benchmark competing methods are either faster or more accurate on some of the diverse data sets used for the comparison in [51], the method provides

the lowest reconstruction error and runtimes when these quantities are averaged over these data sets.

Recent approaches to dMRI reconstruction have used deep neural networks (DNNs). A majority of these methods, such as [52]–[65], are supervised, requiring a large set of high-resolution artifact-free images for network training. However, acquiring such training data can be difficult for some dMRI applications, and may be impossible for many other dynamic imaging applications, where clinical dMRI strategies for freezing motion such as ECG synchronization and breath-hold by a cooperative patient, are not available.

The need for a training data set is overcome by recent object-domain deep image prior (DIP)-based DNN algorithms for dMRI [66], [67], which are unsupervised. Providing impressive results, DIP-based algorithms such as [66] suffer from overfitting and usually require handcrafted early stopping criteria during the optimization of generator network parameters. To overcome these drawbacks, [67] includes regularization constraining the geodesic distances between generated objects. However, this requires the computation of the Jacobian of the generator network at each iteration of the update of the weights, significantly increasing the computation and run time.

A different unsupervised approach [68] combines factorized low rank (i.e., partially separable) with generative models. Although it combines the PSM with the recent DIP framework, this method has the following limitations: (i) The spatial generator is an artifact removal network taking the full-sized spatial basis functions as input. Since this prevents a patch-based implementation, it may be computationally problematic for high-resolution 3D+temporal settings; (ii) The CNN structural prior for natural images that is used in the spatial generator may not be useful as a prior for the individual spatial basis functions, since the least-squares optimal spatial basis functions are the left singular vectors of the complete object, and as such may not have the structure of natural images; (iii) As with the other DIP-based methods, if the additional penalties on the generator parameters are insufficient, this method can be prone to overfitting. These various limitations are overcome by the proposed approach.

Lately, neural fields (NFs), or implicit neural representations, have been proposed as deep learning-based continuous representations of spatio-temporal objects. Since NF representation is essentially resolution-free, the number of parameters to represent, and hence the memory required to store the underlying object, does not scale with the spatio-temporal resolution. This provides an advantage over the previously proposed alternatives to provide a parsimonious representation and reduced memory requirements.

Similar to PSM or DIP-based techniques, most of the NF-based algorithms are optimized in an unsupervised manner, using only the measurement data to estimate the network

parameters.

For dMRI, in [69], the object is represented by NF in the object-domain and the NF parameters are optimized using a fidelity term in the k-space combined with temporal total variation (TV) and low-rankness regularization terms. The method in [70] uses the NF representation in k-space, requiring an inverse Fourier transform to obtain the estimated object. Another method [71] combines the idea in [70] with an additional convolutional correction layer taking the NF outputs of the spatially neighboring coordinates as input and provides a corrected output for the coordinate of interest. This layer is optimized with the low-frequency k-space data estimated by the NF. Again using a k-space NF, [72] proposes a k-space-based self-consistency regularization for recovery. In [73], the authors introduce Fourier positional encodings [74] to shift the spectral bias [75] to higher frequencies.

1.4.2 Dynamic CT (dCT)

CT is faster than MRI, thanks to acquiring an entire projection at one view angle at once. However, because of the angular scanning required in almost all systems, in spite of significant advances in hardware, dCT is still a relatively slow modality. It is also the problem that motivated much of our work reported in this thesis.

In perhaps the earliest algorithmic approaches to dCT, the time-sequential sampling problem in d-CT was formulated and studied in [76], and [77], [78] provided a solution using time-sequential sampling of bandlimited spatio-temporal signals including an optimal view angle order for the scan and theoretical guarantees for unique and stable reconstruction. However, the approach is limited by its bandlimitedness assumptions.

A class of algorithms for dCT are based on modeling the time variation by a motion field, and motion-compensated reconstruction. Many of these algorithms, e.g., [79]–[82] alternate between estimating the motion field and the time-varying object. Some algorithms, e.g., [83] use the optical flow (mass continuity) PDE. A different method [84] estimates the motion field separately by evolving a linear elastic deformation by the Navier-Cauchy PDE, assuming known boundary evolution of the object and initial density distribution. However, these methods assume the total mass or density to be preserved over time, which may be a limiting assumption. As also stated in [85], another issue with the motion field-based methods is their inability to represent topological changes which require discontinuity in the representation. For instance, flow of a contrast agent into the imaged volume, or imaging a fixed slice of a 3D time-varying object under compression with cross-slice motion may violate either of these assumptions. As a hardware-based, problem-specific approach, optical tracking of fiducial marks has been used for motion correction [86].

Various methods approach dCT reconstruction as Bayesian estimation. These include methods [87]–[90] that employ a state-space formulation and use Kalman filter techniques to approximate an MMSE estimate. A different Bayesian method [91] incorporates a spatiotemporal Markov random field object prior, and models measurement imperfections. An iterative algorithm is used to compute a MAP estimate. An interlaced projection acquisition scheme is also proposed.

Several dCT methods motivated by compressed sensing use a total-variation penalty along both spatial and temporal coordinates, as in [92], while other methods, e.g. [93], propose efficient sparse representations for the dynamic object. Incorporating low-rank modeling [94] decomposes the object into low-rank+sparse components, promoting the low rankness implicitly using the nuclear norm. The method relies on the object being approximately static for groups of projections, enabling an approximate reconstruction from each such group.

The PSM model has been introduced into dynamic tomography [95], [96] by carrying the PSM to the projection domain using the harmonic representation of projections, and estimating the spatial and temporal basis functions jointly. The work provides a theoretical analysis of uniqueness and stability and the choice of time-sequential angular scan scheme for the problem. Despite the advantages of this approach, its performance is still limited by the null space of the measurement operator.

To help address the underdetermined problem, the method of [97] improves over [95], [96] by combining a PSM with basic spatial regularization by total variation. Like [95], [96] this method too estimates the temporal and spatial PSM components jointly. This makes it applicable to the time-sequential acquisition scenario considered in this paper , where only a single projection is available per time frame. It imposes low rankness as a soft constraint, using a hybrid biconvex objective with a data fidelity term and a TV regularizer expressed in terms of the unfactorized object, and a penalty for object-PSM mismatch. It uses [95], [96] for fast initialization. The method [97] can be considered a precursor to the method of this paper, which combines a hard-rank constrained PSM with a sophisticated learned spatial regularizer, and provides theoretical convergence analysis.

Recent methods [98]–[101] introduce deep learning into dCT. These methods require an initial reconstruction from the subsets of data from which the motion can be estimated accurately. However, without a periodicity assumption, this is not applicable to the scenarios in this paper, where only a single projection is available at a time. Another method [102] for cone-beam dCT proposes a multi-agent consensus equilibrium [103]-based technique using separate deep denoisers for axial, sagittal, and coronal slices of the object.

For dCT, method [104] combines the learning of NFs for static templates, with motion field estimation to warp the learned template for dynamic reconstruction. This framework has been

applied to generic video frames [105] and, by incorporating a differentiable forward operator and using available time-sequential projections, it was translated to dynamic tomography [104]. A deep-learning-based method [106] for dynamic photoacoustic tomography (PACT) used the partition of unity network (POUnet) [107] as the NF architecture to represent the dynamic object. To handle the dynamics, the input spatial coordinates to the NF were augmented by the temporal coordinate. The POUnet output is expressed in terms of cubic polynomials which are functions of spatio-temporal coordinates whose coefficients are determined by the POU NF, which are weighted by a learned matrix. These basis functions are defined as polynomials of a total degree of 3. To regularize the objective, a Frobenius norm penalty on the interpolation coefficients and a smoothed p-norm penalty $p \in (0, 1)$ on the NF outputs were used with decaying weights. The proposed NF structure and the alternating minimization strategy are claimed to offer robust optimization despite the problem nonconvexity, and a good alternative for representing piecewise smooth functions. The method outperforms its comparison benchmarks only in some of the projection-per-frame settings. Recently, [108] uses proximal splitting to separate updates involving the imaging operator from the NF parameter updates by solving additional embedding problems. The method uses a modified PSM-like POUnet form, where the NF is only dependent on the temporal variable, and the results are demonstrated on PACT. Another NF-based dynamic cone-beam CT (CBCT) method [109] decouples the reconstruction into recovering a reference volume via a spatial NF, and the motion field using principal component analysis (PCA)-based patient-specific models where the component weights are estimated by a temporal NF. The method requires pre-acquired dynamic data to extract a motion model. A follow-up work [110], proposes a faster version that does not require pre-acquired dynamic data. Similar to [110], a two-stage dynamic CBCT model [111] combines NF representation for the object with a neural deformation field, having the limitations mentioned for the methods that model time variation by a deformation field applied to a static template.

Chapter 2

Dynamic Tomography Reconstruction by Projection-Domain Separable Modeling

.

2.1 Introduction

In this chapter, we present a new unsupervised dynamic tomographic reconstruction algorithm, dubbed ProSep, that uses a special bilinear partially-separable model for the projections of a time-varying object. The newly proposed framework also enables the quantitative analysis of the effect of the different sampling schemes. Based on this analysis of time-sequential acquisition schemes, we consider a particular object-independent angular sampling order of the projections for this model and analyze factors affecting the uniqueness and stability of the solution. ProSep does not use any spatial prior for the object, but in numerical experiments shows performance superior to the recently proposed GMLR [112] - a deep image prior model for video reconstruction.

2.2 Problem Statement

We consider the time-sequential dynamic tomography problem described in Chapter 1.3 assuming sampling uniform in time.

Material of this chapter has previously appeared in [95], [96] and is adapted here with permission. Code link: https://github.com/berkiskender/ProSep

The objective is to recover the spatio temporal object $\{f(\boldsymbol{x}, t_p)\}_{p=0}^{P-1}$. We wish to reconstruct the time series of the dynamic object using minimal and verifiable assumptions, and to analyze the effect of various problem parameters, including the sampling scheme, on the uniqueness and stability of the reconstruction. We use a high benchmark for accuracy: FBP-based reconstruction of each frame $f(\mathbf{x}, t_p)$ from the complete set of P projections *taken simultaneously,* i.e., using a total of P^2 rather than just P projections.

2.3 Partially Separable Model (PSM)

2.3.1 Partially separable model in the object domain

The representation of a dynamic object $f(\mathbf{x}, t)$ by a K-th order partially separable model (PSM) is the series expansion

$$f(\mathbf{x}, t) = \sum_{k=0}^{K-1} f_k(\mathbf{x}) \psi_k(t), \tag{2.1}$$

which is known to be dense in $\mathcal{L}_2$ [113], meaning that any finite energy object can be approximated arbitrarily well by such a model of sufficiently high order. Empirically, it is found that even modest values of K provide high accuracy in applications to MR cardiac imaging [32], [114], [115], however we have not found a quantitative analysis of this phenomenon. Our analysis (see Section 2.7) shows that for a spatially bandlimited object, a time-varying affine transformation (i.e, combination of time-varying translation, scaling, and rotation) of bounded magnitude leads to a good approximation by a low order PSM model. Because our benchmark is FBP-based tomographic reconstruction of a static object from P simultaneous projections, which is inherently spatially bandlimited, the above analysis lends support to the use of the PSM with modest K. This reduces the number of degrees of freedom in the dynamic object, enabling reconstruction with less data.

2.3.2 Partially separable model in the projection domain

While the PSM has been used for dynamic imaging in MRI [32], [114], [115], we gain additional insight into its role in tomography by carrying it into the projection domain, and the harmonic representation of the projections. For real-valued time-varying projections, the circular harmonic expansion is

$$g(s, \theta, t) = \sum_{n \in \mathbb{Z}} h_n(s, t) e^{jn\theta}, \quad h_n = h_{-n}^*, \tag{2.2}$$

and we have the following result.

Theorem 1. *Let an object $f(\mathbf{x}, t)$ vanish outside a disk of diameter D and have the partially-separable representation*

$$f(\mathbf{x}, t) = \sum_{k=0}^{K-1} f_k(\mathbf{x})\psi_k(t) + \gamma_f(\mathbf{x}, t), \tag{2.3}$$

with error term bounded as $||\gamma_f||_2^2 \leq \epsilon_f$, for some $\epsilon_f > 0$. Then the projections admit the representation

$$g(s, \theta, t) = \sum_{n \in \mathbb{Z}} h_n(s, t)e^{jn\theta} + \gamma_g(s, \theta, t) \tag{2.4}$$

with $||\gamma_g||_2^2 \leq \pi D\epsilon_f$ and $h_n(s, t)$ represented by the special partially separable model

$$h_n(s, t) = \sum_{k=0}^{K-1} \beta_{n,k}(s)\psi_k(t). \tag{2.5}$$

Remark: The PSM (2.5) is special, in that all the $h_n(s, t)$ are expanded in a common temporal "basis" $\{\psi_k(t)\}$ - independently of n. Moreover, the approximation error of the projections using this special PSM for the expansion coefficients is bounded explicitly in terms of corresponding error in the object PSM (2.3). This can be used to show that the special PSM (2.4) is dense in the space of $\mathcal{L}_2$ functions $h_n(s, t)$.

The special PSM for the projections compresses their representation from $\approx PD^2$ parameters to $\approx KD^2 + KP$, with $K \ll P$. We introduce further compression to $\approx KD^2 + Kd$, by modeling each temporal function by $d \ll P$ parameters.

2.4 The recovery problem: Analysis

2.4.1 Representing the Sampled Projections

Substituting (2.5) into (2.2), it can be shown that the projections are given by

$$\mathbf{g}(s) = (\Theta \bullet \Psi)\beta(s) \tag{2.6}$$

where $\mathbf{g}(s) \in \mathbb{R}^P$, $\Psi \in \mathbb{R}^{P\times(K+1)}$, $\Theta \in \mathbb{C}^{P\times 2N+1}$, and $\beta(s) \in \mathbb{R}^{(2N+1)(K+1)}$ are defined as

$$\begin{aligned}
\mathbf{g}(s) &= \Big[g(s,\theta_1,t_1),\, g(s,\theta_2,t_2),\, \dots\, g(s,\theta_P,t_P)\Big]^T, \\
\Psi &= \begin{bmatrix} \psi_0(t_1) \dots \psi_K(t_1) \\ \vdots \quad \vdots \quad \vdots \\ \psi_0(t_P) \dots \psi_K(t_P) \end{bmatrix}, \quad \widetilde{\Theta} = \begin{bmatrix} 1 & e^{j\theta_1} & \dots & e^{j2N\theta_1} \\ \dots & \dots & \dots & \dots \\ 1 & e^{j\theta_P} & \dots & e^{j2N\theta_P} \end{bmatrix} \\
\Theta &= \mathrm{diag}([e^{-jN\theta_1} \;\; \dots \;\; e^{-jN\theta_P}])\widetilde{\Theta} \\
\beta(s) &= \Big[\beta_{-N,0}(s) \dots \beta_{-N,K}(s) \dots \beta_{N,0}(s) \dots \beta_{N,K}(s)\Big]^T.
\end{aligned} \tag{2.7}$$

and $\bullet$ denotes the Face-splitting product [116], where the p-th row of $\Theta \bullet \Psi$ is given by the Kronecker product of the respective rows, $\Theta_p \otimes \Psi_p$.

Next, as also applied for X-ray CT scatter correction in [117]–[119], we leverage the π-symmetry of flipped projections, $g(-s,\theta) = g(s,\theta+\pi)$, to double the number of equations for the same number of unknowns. Replacing θ_p by $\theta_p + \pi$ in (2.7) corresponds to post-multiplication of Θ by a diagonal matrix with entries $e^{jn\pi}, n = -N, \dots N$. Combining the resulting equations results in expanding model (2.6)-(2.7) to

$$\begin{aligned}
&\widehat{\mathbf{g}}(s) = [\mathbf{g}^T(s), \mathbf{g}^T(-s)]^T, \widehat{\Psi} = [\Psi^T, \Psi^T]^T, \widehat{\Theta} = [\Theta^T, \bar{\Theta}^T]^T, \\
&\widehat{\mathbf{g}}(s) = (\widehat{\Theta} \bullet \widehat{\Psi})\beta(s), \quad \bar{\Theta} = \Theta\ \mathrm{diag}\left\{(-1)^n\right\}_{n=-N}^{N}.
\end{aligned} \tag{2.8}$$

2.4.2 Low Dimensional Model for Temporal Functions Ψ

Suppose that a given fixed number of K sampled temporal functions $\psi_k \in \mathbb{R}^P$ reside in a d-dimensional subspace where $d > K$, but $d \ll P$. They can be expressed as $\psi_k = Uz_k$ where $U \in \mathbb{R}^{P\times d}$ is a fixed interpolator and $z_k \in \mathbb{R}^d$. Without loss of generality (wlog), we assume that $U^TU = I_d$, i.e. U has orthonormal columns. Then, the forward model (2.8) becomes

$$\widehat{\mathbf{g}}(s) = (\widehat{\Theta} \bullet (\widehat{U}Z))\beta(s), \quad \widehat{U} = [U^T, U^T]^T. \tag{2.9}$$

2.4.3 From Recovered Representation to Reconstructed Object

Once (2.9) is solved for the unknowns Z and $\beta(s)\,\forall s$, the reconstructed time series of the object is obtained by first forming the temporal functions $\Psi = UZ$ and then obtaining the harmonic expansion coefficients using (2.5). This allows to obtain the projections at all $\theta \in [0, 2\pi]$ for each t using (2.2). Finally, reconstructions at $\{t_p\}_{p=0}^{P-1}$ of the estimated projections are performed using FBP.

2.4.4 Bilinear problem and necessary conditions for recovery

The problem of recovering Z and β from the sampled projections in (1.5) using (2.9) is the bilinear problem

$$\begin{aligned} &\forall s \,\text{find}\, \beta(s) \in \mathcal{B}, Z \in \mathcal{Z} \\ &\text{s.t. } (\widehat{\Theta} \bullet (\widehat{U} Z))\beta(s) = \widehat{\mathbf{g}}(s), \end{aligned} \tag{2.10}$$

where $\mathcal{B}$ and $\mathcal{Z}$ are appropriate constraint sets. We wish to study conditions for uniqueness of the solution to problem (2.10) and its stability to perturbations in the data and the model. Since $\beta(s)$ and Z appear in product form, there is an inherent scaling ambiguity [120] in (2.10). To remove it, we impose wlog the constraint that $Z^T Z = I$, or equivalently, that $\mathcal{Z}$ is the d-dimensional Stiefel manifold in $\mathbb{R}^P$.

Next, we investigate conditions for uniqueness and stability of the solution to the bilinear problem. As shown in [120], a necessary condition is that when one of the two variables is fixed to a valid solution, the solution for the other is unique. This motivates the study of the linear inverse problems defined by fixing, in turn, one of the two variables $\beta(s)$ $\forall s$ or Z in (2.10).

(i) Model Linear in $\beta(s)$. Defining

$$L_1(Z) \triangleq \widehat{\Theta} \bullet \widehat{\Psi} = \widehat{\Theta} \bullet (\widehat{U} Z)$$

yields

$$\widehat{g}(s) = L_1(Z)\beta(s) \tag{2.11}$$

with the i-th row of $L_1(Z)$ given by

$$L_1(Z)_{i:} = \mathbf{z}^T (A_i \otimes \widehat{U}_{i:}^T) \tag{2.12}$$

where $\mathbf{z} \triangleq \vec{(}Z) \in \mathbb{R}^d(K+1)$ and

$$A_i \triangleq \widehat{\Theta}_{i:}^T \otimes I_{K+1} \in \mathbb{C}^{(K+1)\times(2N+1)(K+1)}. \tag{2.13}$$

For given $\widehat{\Psi}$ or Z, (2.11) is a linear inverse problem in $\beta(s)$.

We have the following result for the full-rankness of L_1.

Theorem 2. *Let $L_1(Z) = \widehat{\Theta} \bullet \widehat{\Psi}$, where $\widehat{\Theta}$ and $\widehat{\Psi}$ are defined in (2.8). Suppose that Ψ with orthonormal columns is random drawn from an absolutely continuous probability distribution on the Stiefel manifold $V_{(K+1)}(\mathbb{R}^P)$. If $2P \geq (K+1)(2N+1)$ and the P view angles*

$\theta_i \in [0, \pi], i = 1, \ldots, P$ *are distinct, then* L_1 *has full column rank w.p. 1.*

Theorem 2 implies that if $2P \geq (K+1)(2N+1)$, then for almost all (i.e., generic) Ψ with orthonormal columns, L_1 will have full column rank. In practice, we implement the low-dimensional model $\Psi = UZ$ of Section 2.4.2 with a fixed U satisfying $U^TU = I_d$ and $Z \in \mathbb{R}^{d\times(K+1)}$ with $d \geq K+1$ and $Z^TZ = I_{K+1}$. It then follows that $\Psi^T\Psi = I_{K+1}$. Numerical experiments in Section 2.6 for such structured Ψ resulted not only in full-rank $L_1(Z)$, but, in fact, in low condition number for appropriately chosen view-angle sampling scheme.

This provides confidence that this necessary full-rank condition for the uniqueness of the solution of (2.10) will be satisfied.

(ii) Model Linear in Z. Here we introduce explicitly the discretization of the detector positions s to J values. For fixed β it is shown in Section 2.8.2 that (2.10) reduces to a linear inverse problem in $\mathbf{z} \triangleq \mathrm{vec}(Z)$

$$\widehat{\mathbf{g}} = \mathbf{L}_2(\beta)\mathbf{z} \tag{2.14}$$

where vector $\widehat{\mathbf{g}} \in \mathbb{R}^{2JP}$ includes the stacked elements of $g(s)_i$, $s = \{s_j\}_{j=1}^J$, $i = 1, \ldots, 2P$ and $\mathbf{L}_2(\beta) \in \mathbb{R}^{2JP\times d(K+1)}$, with i-th row given by

$$L_2(\boldsymbol{\beta}(s))^{(i)} \triangleq \beta(s)^T A_i^T (I_{K+1} \otimes \widehat{U}_{i:}) \tag{2.15}$$

Problem (2.14) will have a unique solution for fixed $\boldsymbol{\beta}$ if and only if $\mathbf{L}_2(\boldsymbol{\beta})$ has full column rank. Furthermore, the stability of the solution is governed by the condition number $\kappa(\mathbf{L}_2(\boldsymbol{\beta}))$.

Theorem 3. *Let* $\Gamma \triangleq \sum_{j=1}^J \boldsymbol{\beta}(s_j)\boldsymbol{\beta}(s_j)^T \succ 0$. *Then,* $\mathbf{L}_2(\boldsymbol{\beta})$ *is full column rank, and*

$$\kappa(\mathbf{L}_2(\boldsymbol{\beta})) \leq \sqrt{\kappa(\Gamma)}. \tag{2.16}$$

Clearly, $\Gamma \succ 0$, i.e., $\kappa(\Gamma) < \infty$ establishes full column rank of $\mathbf{L}_2(\boldsymbol{\beta})$ and uniqueness of the solution to (2.14) for fixed $\boldsymbol{\beta}$. This requires $J \geq (2N+1)(K+1)$ and that there are at least $(2N+1)(K+1)$ linearly independent vectors in the set $\{\beta(s_j)\}_{j=1}^J$. This condition can be somewhat restrictive, but (2.16) is only an upper bound on $\kappa(\mathbf{L}_2(\boldsymbol{\beta}))$, which may not be tight. In practice, $\kappa(\mathbf{L}_2(\boldsymbol{\beta}))$ takes on small (< 10) values even if $\Gamma \succ 0$ is not satisfied. Furthermore, $\kappa(\Gamma)$ typically decreases with increasing spatial resolution J of the projections.

2.4.5 Condition numbers and view angle sampling schemes

Evaluation of the condition numbers $\kappa(L_1)$ and $\kappa(\mathbf{L}_2(\boldsymbol{\beta}))$ in the separate problems (2.11) and (2.14) enables explicit analysis of the effects of the view angle sampling scheme on stability. We consider three schemes: (i) progressive with $\Delta_\theta = \frac{2\pi}{P}$; (ii) random, $\theta_p \sim \mathcal{U}[0, 2\pi), \forall p$; and

(iii) bit-reversed in $[0, 2\pi)$, with angles obtained by the reversal of the binary representations of the progressive scheme [121]. The intervals are selected as $[0, \pi)$ when π-symmetry is exploited. We study the dependence of the condition numbers $\kappa(L_1(Z))$, and $\kappa(\mathbf{L}_2(\boldsymbol{\beta}))$ on view-angle sampling scheme, for fixed K, N, P. For the study of L_1, the temporal functions in Ψ were set using orthonormal Legendre polynomials of increasing order. (Because their version sampled uniformly at P points is not exactly orthonormal, it was orthonormalized by Gram-Schmidt.) $\mathbf{L}_2(\boldsymbol{\beta})$, the interpolator matrix is set to $U \in \mathbb{R}^{2P\times d}$ with elements $U_{ij} \sim \mathcal{N}(0, 1)$ independent and identically distributed (iid) and $\beta_i \sim \mathcal{N}(0, 1)$, iid.

Table 2.1: Condition numbers for different angular sampling schemes for K=5, N=28, P=512. For the random scheme, $\kappa(L_1)$ and $\kappa(\mathbf{L}_2)$ are the best out of 1000 different realizations.

	Progressive	Random	Bit-reversed
$\kappa(L_1)$ no symm./symm.	4.2e+16 / 1.8e+16	103.2 / 8.3	**11.7** / **3.0**
$\kappa(\mathbf{L}_2)$	1.2	1.2	1.2

Table 2.1 shows that for the given set of parameters, although $\kappa(\mathbf{L}_2)$ is not affected by the sampling scheme, $\kappa(L_1)$ depends on it significantly, and a naive progressive scheme results in a practically singular model, whereas the bit-reversed scheme improves substantially over the random case.

2.5 Recovery Algorithms

In view of unavoidable noise and model approximations, we solve (2.11) in the least-squares sense. Define the loss function and the least-squares problem as

$$\mathcal{L}(\beta(s), Z) = ||\hat{g}(s) - L_1(Z)\beta(s)||_2^2 \tag{2.17}$$

$$\min_{\substack{Z\in\mathbb{R}^{d\times(K+1)},\\ \beta(\cdot):\mathbb{R}\to\mathbb{R}^{(2N+1)(K+1)}}} \sum_s \mathcal{L}(\beta(s), Z). \tag{2.18}$$

Problem (2.18) can be solved using the Variable Projection method (VarPRO) [122]. First, because the summed term $\mathcal{L}(\beta(s), Z)$ is nonnegative, for fixed Z, (2.18) is minimized by minimizing (2.17) pointwise w.r.t $\beta(s)$, so that (2.18) becomes

$$\min_{Z\in\mathbb{R}^{d\times(K+1)}} \sum_s \left[\min_{\beta(s)\in\mathbb{R}^{(2N+1)(K+1)}} \mathcal{L}(\beta(s), Z) \right]. \tag{2.19}$$

Now, the inner minimization in (2.19) is a least squares problem by $\beta(s)$, optimized for

$\beta^*(s) = L_1^\dagger(Z)\hat{g}(s)$. Inserting this into the objective function and simplifying, we obtain

$$\mathcal{L}(\beta^*(s), Z) = \mathrm{tr}\{P_{R^\perp(L_1(Z))}\hat{g}(s)\hat{g}(s)^T\}, \tag{2.20}$$

where $P_{R^\perp(L_1(Z))}$ is the orthogonal projection matrix onto the orthocomplement of the range space $R\left(L_1(Z)\right)$. of $L_1(Z)$. Then (2.19) reduces to

$$\min_{Z\in\mathbb{R}^{d\times(K+1)}} \mathrm{tr}\{P_{R^\perp(L_1(Z))}\Xi(\hat{\boldsymbol{g}})\}, \tag{2.21}$$

where $P_{R^\perp(L_1(Z))}$ is the orthogonal projection matrix onto the orthocomplement of the range space $R\left(L_1(Z)\right)$ and where $\Xi(\hat{\boldsymbol{g}}) \triangleq \sum_s \hat{g}(s)\hat{g}(s)^T$ and $\Xi(\hat{\boldsymbol{g}}) \in \mathbb{R}^{P\times P}$ without π-symmetry, or $\Xi(\hat{\boldsymbol{g}}) \in \mathbb{R}^{2P\times 2P}$ with it. In this form, (2.21) reveals an important fact: the optimum Z, which determines the temporal functions $\psi_k(t)$, does not depend on the detailed measurements $\hat{g}(s)$. Instead, it only depends on the data matrix $\Xi(\hat{\boldsymbol{g}})$, which aggregates all the measurement information.

As written, (2.21) does not have a unique solution w.r.t. Z, because $R\left(L_1(Z)\right)$ is invariant to scaling of the columns of Z. We remove this ambiguity by constraining Z to have orthonormal columns, $Z^TZ = I$. In the implementation, we use a penalized form and gradient descent for minimization.

2.6 Experiments

We compare ProSep with the recent GMLR method [112], an extension of deep image prior [123] to video. GMLR uses a generative model to map latent codes ζ_t to images f_t to reconstruct a video from incomplete measurements. It performs joint optimization of latent codes and the generator parameters to match the measured data, while enforcing smoothness on the sequence of ζ_t to achieve temporal smoothness in f_t. In our application of GMLR to dynamic tomography it solves

$$\min_{\zeta_1,\ldots,\zeta_T;\gamma} \mathcal{L}(\boldsymbol{\zeta},\gamma) \quad \text{s.t.} \quad \mathrm{rank}(\boldsymbol{\zeta}) = r$$

$$\text{where} \quad \mathcal{L}(\boldsymbol{\zeta},\gamma) = \lambda \sum_{t=1}^{P} \left\| g_t - \mathcal{R}_{\theta(t)} G_\gamma\left(\zeta_t\right)\right\|_2^2 + (1-\lambda)\sum_{t=1}^{P-1} \|\zeta_{t+1} - \zeta_t\|_2^2$$

where $g_t = \mathcal{R}_{\theta(t)} f_t$, G_γ is the generator with parameters γ, $\boldsymbol{\zeta} = [\zeta_1\ \zeta_2\ \ldots \zeta_P]$, and λ controls the similarity of consecutive latent codes. In [112] the architecture of G_γ is the DCGAN [124] for a 64×64 image with minor changes. Since our experiments consider $f \in \mathbb{R}^{128\times 128}$,

another upsampling layer was added to the original G_γ configuration [125].

We present the reconstruction results for ProSep with and without leveraging the π-symmetry. Both settings use the bit-reversed angular scheme between $[0, \pi]$ and $[0, 2\pi]$, respectively. The interpolator U is set to a cubic spline interpolator for both cases. The Adam [126] algorithm was used for optimizing Z with learning rate of 0.2. For GMLR, the learning rates for γ and ζ were kept as in the posted code [125]. For each P, the model was trained for $6(10^4)$ steps with $\lambda = 0.5$ and rank $r = 4$.

In all experiments, we use the synthetic dynamic 128x128 CT slice of a walnut shown in Fig. 2.1 from the publicly available *static* 3D walnut CT dataset [127]. We create the dynamic test object by synthetically warping the central axial slice of one of the walnut objects using a sinusoidal piecewise-affine time-varying warp [128]. To be precise, the image is divided into a $N \times N$ uniformly spaced rectangular grid, and the following vertical displacement is applied on each row separately to drive the temporally varying warp $\Delta_{n,t} = -C(t)\sin(3\pi n/N),\ n \in \{0, \ldots, N-1\}$, where $C(t)$ is a linearly increasing function of t and $C(0) = 0$.

Table 2.2 shows the reconstruction PSNR (in dB), SSIM and mean absolute error (MAE) values. Using π-symmetry for P views between $[0, \pi)$ allows larger N and K for stable recovery and thus improves the accuracy significantly for each P. GMLR, ProSep, and ProSep-Symm, are compared in Fig. 2.2. Both error figures and quantitative metrics show competitive performance of ProSep relative to GMLR, and better performance of ProSep-Symm. Furthermore, the ProSep and ProSep-Symm reconstructions of the arc-shaped feature in the zoom-in images are sharper.

(a) (b) (c)

Figure 2.1: The object undergoing pointwise affine transform at time instances (a) $t = 0$, (b) $t = P/2$ and (c) $t = P$.

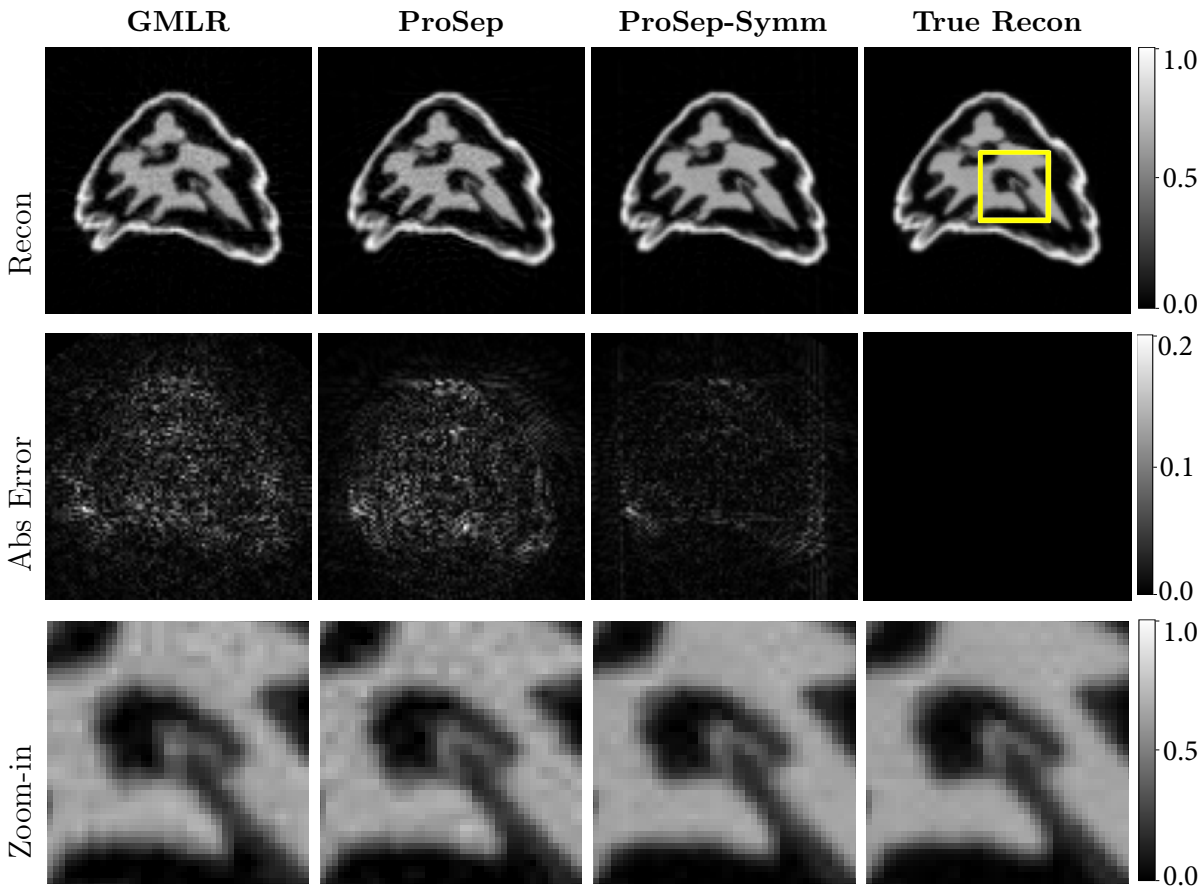


Figure 2.2: Row 1: reconstructions of the object in Fig. 2.1 at $t = 990/1024$ for different methods; Row 2: absolute deviations from "True Recon"; Row 3: zoom-in of the yellow box region.

Table 2.2: Average reconstruction accuracies for the complete 3D time-varying phantom in Fig. 2.1 for different P. "ProSep symm" employs the π-opposite projection symmetry.

P	Method	K	N	d	PSNR (dB)	SSIM	MAE
256	GMLR	-	-	-	27.3	0.783	0.027
	ProSep	3	24	4	26.9	0.894	0.022
	ProSep symm.	5	30	6	**30.4**	**0.928**	**0.015**
512	GMLR	-	-	-	31.1	0.876	0.017
	ProSep	5	28	6	30.5	0.944	0.014
	ProSep symm	7	48	8	**35.1**	**0.959**	**0.010**
1024	GMLR	-	-	-	36.7	0.925	0.009
	ProSep	7	48	8	36.8	0.979	0.007
	ProSep symm.	9	56	10	**39.5**	**0.980**	**0.006**

2.7 Analysis of Partially-Separable Model for Affine Motion

Throughout this section, we consider a nominal object $f_0(\mathbf{x}) \in \mathcal{L}_2$, with $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, vanishing outside a disk of radius L, and essentially bandlimited to spatial bandwith B radians, that is

$$f_0(\mathbf{x}) = 0 \quad \text{for} \quad \|\mathbf{x}\|_2 > L \tag{2.22}$$

$$\int_{\|\boldsymbol{\omega}\|>B} |F_0(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \leq \epsilon_f \tag{2.23}$$

where F_0 is the 2D Fourier transform of f_0, and $\epsilon_f \ll \|f_0\|_2^2$. We consider motions of the nominal object resulting in a time-varying object $f(\mathbf{x}, t)$, which we assume remains bounded in the disk of radius L, that is, $f(\mathbf{x}, t) = 0 \, \forall t$ for $\|\mathbf{x}\|_2 > L$.

Time-Varying Translation

Consider the nominal object translating in time along trajectory $\mathbf{c}(t) = (c_1(t), c_2(t))$ of bounded extent, $\|\mathbf{c}(t)\|_2 \leq c_{\max} \, \forall t$, resulting in the time-varying object

$$f(\mathbf{x}, t) = f_0(\mathbf{x} - \mathbf{c}(t)) \tag{2.24}$$

$$F(\boldsymbol{\omega}, t) = F_0(\boldsymbol{\omega}) e^{-j(\boldsymbol{\omega}^T \mathbf{c}(t))}. \tag{2.25}$$

Using the assumptions and the Cauchy-Schwartz inequality, it follows that $|\boldsymbol{\omega}^T \mathbf{c}(t)| \leq B c_{\max}$, suggesting the use of a K-th order Taylor series expansion of the complex exponential in

(2.25). To bound the remainder term, we use the integral form of the remainder in Taylor's theorem,

$$e^{jx} = \sum_{k=0}^{K} \frac{(jx)^k}{k!} + \frac{j^{K+1}}{K!} \int_0^x (x-s)^K e^{js} ds \tag{2.26}$$

which yields

$$\begin{aligned} \left| e^{jx} - \sum_{k=0}^{K} \frac{(jx)^k}{k!} \right| &\leq \frac{|x|^{K+1}}{(K+1)!} \\ &\approx \sqrt{2\pi(K+1)} \left(\frac{e|x|}{K+1} \right)^{K+1}, \end{aligned} \tag{2.27}$$

where the approximate form follows by Striling's approximation of the factorial. It follows that the remainder is exponentially decaying for $K > e|x| - 1$. Applying (2.27) to the complex exponential in (2.25) we obtain that the remainder in its $K+1$ term expansion is bounded by

$$\begin{aligned} \left| e^{-j\boldsymbol{\omega}^T \mathbf{c}(t)} - \sum_{k=0}^{K} \frac{(-j\boldsymbol{\omega}^T \mathbf{c}(t))^k}{k!} \right| &\leq \frac{|Bc_{\max}|^{K+1}}{(K+1)!} \\ &\approx \sqrt{2\pi(K+1)} \left(\frac{Bc_{\max}}{K+1} \right)^{K+1}. \end{aligned} \tag{2.28}$$

This remainder is exponentially decaying for $K > Bc_{\max} - 1$. Because it holds pointwise for each $\boldsymbol{\omega}$, it implies that the corresponding K-th order expansion for $F(\boldsymbol{\omega}, t)$ has relative rms error bounded by the same remainder bound.

The resulting expansion leads to the following approximation in the spatial domain

$$f(\mathbf{x}, t) \approx \sum_{\substack{(\alpha_1+\alpha_2) \leq K \\ \alpha_1, \alpha_2 \geq 0}} (-1)^{(\alpha_1+\alpha_2)} c(t)^\alpha \left(\frac{\mathcal{D}^\alpha}{\alpha_1! \alpha_2!} f_0(\mathbf{x}) \right) \tag{2.29}$$

where we use the index notation

$$\begin{aligned} \mathbf{c}^\alpha &\triangleq c_1^{\alpha_1} c_2^{\alpha_2} \\ \mathcal{D}^\alpha &\triangleq \frac{\partial^{(\alpha_1+\alpha_2)}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}} \end{aligned} \tag{2.30}$$

By Parseval's identity, the relative rms truncation error of the expansion in (2.29) is bounded by the right-hand side of (2.28). Hence, f admits a partially separable expansion with $(K+1)(K+2)/2$ terms, with possibly moderate K.

The number of significant terms in the expansion depends on the spatial bandwidth B of

f_0 and the magnitude $c_{\max}$ of the motion.

Time-Varying Scaling

Consider the nominal object $f_0(\mathbf{x})$ undergoing scaling by diagonal matrix $\mathbf{C}(t) = \text{diag}\,(c_1(t), c_2(t))$ to $f(\mathbf{x}, t) = f_0(\mathbf{C}(t)\mathbf{x})$. This case, too, can be represented using the partially-separable model. To see this, we use the 2D Mellin transform [129]

$$\hat{f}(s,\tau) \triangleq \int_0^\infty \int_0^\infty x_1^{s-1} x_2^{\tau-1} f(\mathbf{x}) dx_1 dx_2 \tag{2.31}$$

which is well-defined and has an inverse if the integral

$$\int_0^\infty \int_0^\infty |f(\mathbf{x})| x_1^{s-1} x_2^{\tau-1} dx_1 dx_2 \tag{2.32}$$

converges for all s s.t $\alpha_1 < \Re(s) < \beta_1$ and all τ s.t $\alpha_2 < \Re(\tau) < \beta_2$ for some $\alpha_1 < \beta_1$ and $\alpha_2 < \beta_2$.

Since $f(\mathbf{x}, t)$ is defined on a centered disk, we cannot apply the Mellin transform, which is only defined for non-negative arguments, to it directly. Instead, we consider the part of $f(\mathbf{x}, t)$ supported in the first quadrant, and to simplify the notation we still denote it by $f(\mathbf{x}, t)$ and $f_0(\mathbf{x})$. The extension to include the other parts is discussed later. For our setting, since f is supported on a disk of radius L and $f \in \mathcal{L}_2$, this condition holds for $0.5 < \Re(s) < \infty$ and $0.5 < \Re(\tau) < \infty$.

The Mellin transform $\hat{f}$ of the object is given by

$$\hat{f}(s,\tau,t) = c_1(t)^{-s} c_2(t)^{-\tau} \hat{f}_0(s,\tau) \tag{2.33}$$

Again, using a truncated series expansion for the exponential term, we obtain

$$\begin{aligned}\hat{f}(s,\tau,t) &\approx \sum_{k=0}^{K} \frac{(-1)^k}{k!} \left(s \ln c_1(t) + \tau \ln c_2(t)\right)^k \hat{f}_0(s,\tau) \\ &= \sum_{\substack{(\alpha_1+\alpha_2)\le k \\ \alpha_1,\alpha_2 \ge 0}} \frac{(-1)^{\alpha_1+\alpha_2}}{\alpha_1!\alpha_2!} \left(s \ln c_1(t)\right)^{\alpha_1} \left(\tau \ln c_2(t)\right)^{\alpha_2} \hat{f}_0(s,\tau)\end{aligned}$$

Using the differentiation property for the 2D Mellin transform

$$\left(\mathbf{x}_1 \frac{d}{d\mathbf{x}_1}\right)^{\alpha_1} \left(\mathbf{x}_2 \frac{d}{d\mathbf{x}_2}\right)^{\alpha_2} f(\mathbf{x}) \leftrightarrow (-1)^{\alpha_1+\alpha_2} s^{\alpha_1} \tau^{\alpha_2} \hat{f}(s,\tau), \tag{2.34}$$

yields

$$f(\mathbf{x},t) \approx \sum_{\substack{(\alpha_1+\alpha_2)\leq k \\ \alpha_1,\alpha_2\geq 0}} \ln^{\alpha}\mathbf{C}(t)\left(\frac{1}{\alpha_1!\alpha_2!}(\mathbf{x}\mathcal{D})^{\alpha} f_0(\mathbf{x})\right) \tag{2.35}$$

using the notation in (2.30) with

$$\ln^{\alpha}\mathbf{C} = \ln^{\alpha_1} c_1(t)\ln^{\alpha_2} c_2(t). \tag{2.36}$$

In addition to the spatial bandwidth of f_0 and the log-magnitude of the motion components $c_1(t)$ and $c_2(t)$, the number of significant terms in the expansion depends also on the radius L of the support of the object f_0. Time-varying scaling factors $c_1(t)$ and $c_2(t)$ being closer to 1, lower spatial bandwidth, and a smaller radius L would result in fewer significant terms in the expansion.

To incorporate the parts of the object in the remaining quadrants to the analysis, we use the following steps. The object f_0 is decomposed into the sum of 4 pieces as $f_0(\mathbf{x}) = \sum_{m,n=0}^{1} f_{mn}(\mathbf{x})$ where f_{mn} is supported in the mn-th quadrant in binary notation. Then, we represent the versions of these decomposed parts of the object after reflection into the first quadrant as $\tilde{f}_{mn}(\mathbf{x}) = f_{mn}((-1)^m x_1,(-1)^n x_2)$ and apply (2.35) on each $\tilde{f}_{mn}$. The object is finally reconstituted from its subparts as $f(\mathbf{x},t) = \sum_{m,n=0}^{1} \tilde{f}_{mn}((-1)^m x_1,(-1)^n x_2,t)$ by incorporating the necessary sign changes in the term $(\mathbf{x}\mathcal{D})^{\alpha}$ as $\left((-1)^m \mathbf{x}_1 \frac{d}{d\mathbf{x}_1}\right)^{\alpha_1}\left((-1)^n \mathbf{x}_2 \frac{d}{d\mathbf{x}_2}\right)^{\alpha_2}$.

Although this partition creates discontinuities at $x_1 = 0$ and $x_2 = 0$, since the derivatives are multiplied with x_1 and x_2 at these points, these discontinuities do not constitute a problem. Furthermore, the composition of the object from the four quadrants does not increase the number of terms in the expansion (2.35), because the expansions for all four quadrants share the same temporal functions $\ln^{\alpha}\mathbf{C}(t)$.

Time-Varying Rotation

Consider the nominal object $f_0(x)$ rotated by angle $\theta(t)$. Using the polar representation of $f(\mathbf{x},t)$

$$f_{\text{pol}}(r,\phi,t) = f_{0,\text{pol}}(r,\phi-\theta(t)) \tag{2.37}$$

the rotation angle $\theta(t)$ acts as a translation in the angular coordinate. Therefore, we adopt the same approach used for a time-varying translation in Section 2.7 1). Computing the

Fourier transform of f_{pol} which is 2π-periodic in ϕ, w.r.t. ϕ, yields

$$F_{pol}(r, \omega_\phi, t) = F_{0,pol}(r, \omega_\phi, t) e^{-j\omega_\phi \theta(t)}. \tag{2.38}$$

Expanding the complex exponential using a k-th order Taylor series as in Section 2.7 1), we obtain a similar bound on the remainder term

$$\left| e^{j\omega_\phi \theta(t)} - \sum_{k=0}^{K} \frac{1}{k!} (j\omega_\phi \theta(t))^k \right| \leq \frac{|\omega_\phi \theta(t)|^{K+1}}{(K+1)!} \tag{2.39}$$

$$\leq \frac{|B_\phi \theta_{\max}|^{K+1}}{(K+1)!} \tag{2.40}$$

where $|\theta(t)| \leq \theta_{\max}\ \forall t$ and $B_\phi = BL$ is the angular bandlimit of the object $f_{0,pol}$.

Finally, the resulting expansion in the spatial domain is given by

$$f_{pol}(r, \phi, t) \approx \sum_{k=0}^{K} (-1)^k \theta^k(t) \left(\frac{1}{k!} \frac{d^k}{d\phi^k} f_{0,pol}(r, \phi) \right). \tag{2.41}$$

It follows that in the case of rotation, f_{pol} admits a partially separable representation with $K+1$ significant terms, whose number depends on the bandwidth of the object, its radius L, and the magnitude of the rotation. Again, the bound on the truncation error decays exponentially with decreasing values of these parameters.

2.8 Proofs

In this section, we establish some of the proofs of the theorems and derivations that are used in this chapter.

2.8.1 Proof of Theorem 1

Before proving the result we state the following lemma.

Lemma 4. *([9], Proposition 6.6.1.) Suppose that object $f \in \mathcal{L}_2(\mathbb{R}^2)$ vanishes outside a disk of radius L. Then, for each θ, we have the estimate*

$$\int_{-\infty}^{\infty} |(\mathcal{R}f)(s, \theta)|^2 ds \leq 2L ||f||_2^2.$$

Denote the first term in (2.3) by $\tilde{f}(\mathbf{x},t)$. Then its Radon transform is given by

$$(\mathcal{R}\tilde{f})(s,\theta,t) = \sum_{k=0}^{K-1} q_k(s,\theta)\psi_k(t) \tag{2.42}$$

where $q_k \triangleq \mathcal{R}f_k$. Next, expanding q_k in a Fourier series yields

$$q_k(s,\theta) = \sum_n \beta_{n,k}(s)e^{jn\theta}. \tag{2.43}$$

Substituting (2.43) into (2.42) and switching the order of summation yields

$$(\mathcal{R}\tilde{f})(s,\theta,t) = \sum_n h_n(s,t)e^{jn\theta} \tag{2.44}$$

where $h_n(s,t)$ is given by (2.5).

Next, we define the Radon transform of the second term in (2.3) as $\gamma_g(s,\theta,t) = (\mathcal{R}\gamma_f)(s,\theta,t)$. Then, applying Lemma 4 to $\gamma_f(.,t)$ for fixed t yields

$$\int_{-\infty}^{\infty} |\gamma_g(s,\theta,t)|^2 ds \leq 2L \int_{-\infty}^{\infty} |\gamma_f(\mathbf{x},t)|^2 d\mathbf{x}$$

and integrating both sides over θ and t

$$||\gamma_g||_2^2 \leq 2\pi L||\gamma_f||_2^2 \leq 2\pi L\epsilon_f. \tag{2.45}$$

Combining (2.44) and (2.45) shows that the projections admit the representation (2.4) with an error term bounded as $||\gamma_g||_2^2 \leq 2\pi L\epsilon_f$ and $h_n(s,t)$ represented by the *special* partially separable model (2.5). ⊠

2.8.2 Derivation of $\mathbf{L}_2(\boldsymbol{\beta})$

To formulate the model linear in Z of Sec. 2.4.4 we manipulate the expression for the i-th element of $\hat{g}(s)$,

$$\begin{aligned} \hat{g}(s)_i = \hat{g}(s)_i^T &= \beta(s)^T L_{1_{i:}}^T(Z) \\ &= \beta(s)^T (A_i^T \otimes \widehat{U}_{i:})\mathbf{z} \\ &= \beta(s)^T A_i^T (I_{K+1} \otimes \widehat{U}_{i:})\mathbf{z}. \end{aligned} \tag{2.46}$$

where A_i is defined in (2.13).

Define $\hat{g}^{(i)} \in \mathbb{R}^J$ by $\hat{g}_s^{(i)} \triangleq \hat{g}(s)_i$, that is, $\hat{g}^{(i)}$ is a vector containing the i-th elements of all

$\hat{g}(s)$. This vector can be obtained as

$$\hat{g}^{(i)} = L_2(\beta)^{(i)}\mathbf{z} \quad i = 1, \ldots, P \tag{2.47}$$

where matrix $\boldsymbol{\beta} \in \mathbb{R}^{(2N+1)(K+1)\times J}$ contains $\beta(s_j),\ j \in \{1, \ldots, J\}$ as columns and $L_2(\beta)^{(i)} \in \mathbb{R}^{J\times d(K+1)}$ is the stacking in s of row vectors

$$L_2(\boldsymbol{\beta}(s))^{(i)} \triangleq \beta(s)^T A_i^T (I_{K+1} \otimes \widehat{U}_{i:}). \tag{2.48}$$

Then, stacking the vectors $\hat{g}^{(i)}$ and operators $L_2(\beta)^{(i)}$ for $i \in \{1, \ldots, 2P\}$, we obtain the problem (2.14) linear in Z where $\hat{g}$ and $\mathbf{L}_2(\boldsymbol{\beta})$ are defined as

$$\hat{\mathbf{g}} = \begin{bmatrix} \hat{g}^{(1)} \\ \vdots \\ \hat{g}^{(2P)} \end{bmatrix} \quad \mathbf{L}_2(\boldsymbol{\beta}) = \begin{bmatrix} L_2(\beta(s_1))^{(1)} \\ \vdots \\ L_2(\beta(s_J))^{(2P)} \end{bmatrix}. \tag{2.49}$$

2.8.3 Proof of Theorem 2

We first state and prove some preliminary results. We use the property of Kruskal rank of matrix M, or krank(M), defined as the maximal number k such that any subset of k columns of M are linearly independent.

Theorem 5. *Suppose $V \in \mathbb{C}^{2P\times n}$ and V^T has Kruskal rank* krank$(V^T) = n$*, and $\Psi \in \mathbb{R}^{P\times L}$ has elements independently and identically distributed as $N(0,1)$. Let matrix $\widehat{\Psi} \triangleq [\Psi^T, \Psi^T]^T$ have columns $\widehat{\psi}_i$. Then, if $2P \geq nL$, the matrix $M = \begin{bmatrix} \text{diag}(\widehat{\psi}_1)V & \ldots & \text{diag}(\widehat{\psi}_L)V \end{bmatrix}$ has full column rank w.p.1.*

Proof. Throughout the proof we assume $2P = nL$, so that M is a square $nL \times nL$ matrix, and note that because adding rows to a matrix does not decrease column rank, the obtained results hold for $2P \geq nL$. We also assume that n is even, and consider the odd n case at the end.

We denote the set of integers $[1, 2, \ldots, N]$ by $[N]$. To show the desired result, we follow an approach similar to the proof of Theorem 5.2 in [130].

Partition $[P]$ into L disjoint sets $\{J_l \subset [P]\}_{l=1}^L$ of equal size $|J_l| = n/2$, and divide the matrix V into two parts $V^T = [V^{(1)T}, V^{(2)T}]^T$ where $V^{(1)}, V^{(2)} \in \mathbb{C}^{P\times n}$. Denote the matrices containing the rows of $V^{(1)}$ and $V^{(2)}$ indexed by elements of J_l by $V^{(1)}_{J_l,:}, V^{(2)}_{J_l,:} \in \mathbb{C}^{(n/2)\times n}$. Then

the condition $\operatorname{krank}(V^T) = n$ yields

$$\operatorname{rank}\left(\begin{bmatrix} V^{(1)}_{J_l,:} \\ V^{(2)}_{J_l,:} \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} V^{(1)}_{J_l,:} \\ V^{(2)}_{J_l,:} \end{bmatrix}^T\right) = n. \tag{2.50}$$

Now, $\operatorname{rank}(M) = nL \iff \det M \neq 0$. Since $\det M$ is a multivariate polynomial in the $PL = nL^2/2$ entries of Ψ with coefficients dependent only on the entries of V, it is either identically zero or its zero set is an affine algebraic set and thus a nowhere dense set of measure zero in $\mathbb{C}^{nL^2/2}$. Thus, it suffices to show $\det M \neq 0$ for a single choice of Ψ [131]–[133], demonstrating that the polynomial $\det M$ does not vanish identically.

Permuting the rows of M produces the following matrix

$$\begin{bmatrix} D^{(1)}_{J_1} V^{(1)}_{J_1,:} & \dots & D^{(L)}_{J_1} V^{(1)}_{J_1,:} \\ D^{(1)}_{J_1} V^{(2)}_{J_1,:} & \dots & D^{(L)}_{J_1} V^{(2)}_{J_1,:} \\ \vdots & \ddots & \vdots \\ D^{(1)}_{J_L} V^{(1)}_{J_L,:} & \dots & D^{(L)}_{J_L} V^{(1)}_{J_L,:} \\ D^{(1)}_{J_L} V^{(2)}_{J_L,:} & \dots & D^{(L)}_{J_L} V^{(2)}_{J_L,:} \end{bmatrix} \tag{2.51}$$

where $D^{(l)}_{J_k} \triangleq \operatorname{diag}(\Psi_{J_k,l})$. Setting $\Psi_{J_l,l} = \mathbf{1}_{n/2}$ and thus $D^{(l)}_{J_l} = I_{n/2}$, and $\Psi_{J_k,l} = \mathbf{0}_{n/2}$ and thus $D^{(l)}_{J_k} = 0$ for $l \neq k$ yields

$$\begin{bmatrix} V^{(1)}_{J_1,:} & \dots & 0 \\ V^{(2)}_{J_1,:} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & V^{(1)}_{J_L,:} \\ 0 & \dots & V^{(2)}_{J_L,:} \end{bmatrix} \tag{2.52}$$

which is a block diagonal matrix with each $n \times n$ block along the diagonal being full rank by the assumption $\operatorname{krank}(V^T) = n$. Thus, M is full rank for this choice of Ψ, and hence has full rank for almost all Ψ (i.e., generically) and w.p. 1 for the random Ψ.

For the odd n case, if L is even, we choose $2P = nL$. Then, consider two different complementary partitions of $[P]$ into L subsets, $[P] = \bigcup_{k=1}^{L} J^{(i)}_k, i = 1, 2$ such that for each k, $|J^{(1)}_k| + |J^{(2)}_k| = n$. We apply the $i = 1$ partition to the top P rows of matrix M and to $V^{(1)}$ and the $i = 2$ partition to the bottom P rows of matrix M and to $V^{(2)}$. Repeating the previous argument involving permutation of the rows of M, the resulting matrix analogous

to (2.51) will have $k-l$ block given by

$$\begin{bmatrix} D^{(l)}_{J_k^{(1)}} V^{(1)}_{J_k,:} \\ D^{(l)}_{J_k^{(2)}} V^{(2)}_{J_k,:} \end{bmatrix} = \text{block.diag}\left(D^{(l)}_{J_k^{(1)}}, D^{(l)}_{J_k^{(2)}} \right) \begin{bmatrix} V^{(1)}_{J_k,:} \\ V^{(2)}_{J_k,:} \end{bmatrix} \tag{2.53}$$

where $D^{(l)}_{J_k^{(i)}} \triangleq \text{diag}(\Psi_{J_k^{(i)},l})$. Setting $[\Psi^T_{J_k^{(1)},l}, \Psi^T_{J_k^{(2)},l}]^T = \delta[k-l]\mathbf{1}_n$ and thus block.diag$(D^{(l)}_{J_k^{(1)}}, D^{(l)}_{J_k^{(2)}}) = \delta[k-l]I_n$, yields again a full-rank block diagonal matrix for the permuted M as in (2.52), establishing the result for n odd and L even.

When both n and L are odd, we choose $2P = nL+1$ (which is the smallest integer value satisfying $2P \geq nL$), and the additional row is discarded before partitioning $[P]$ and $[P-1]$ into L sets for $V^{(1)}$ and $V^{(2)}$ as in the even nL case since the extra row does not affect the full rankness of matrix M. ⊠

Corollary 1. *Theorem 5 also holds if instead of being a random normal matrix, Ψ is drawn at random from an absolutely continuous probability distribution on the Stiefel manifold $V_{(L)}\left(\mathbb{R}^P\right)$.*

Proof. There exists a column permutation matrix Π_M such that

$$M\Pi_M = \begin{bmatrix} \text{diag}(v_1)\widehat{\Psi} & \dots & \text{diag}(v_{2N+1})\widehat{\Psi} \end{bmatrix}. \tag{2.54}$$

Consider the SVD $\Psi = U\Sigma\Xi^*$, and let $S \in \mathbb{R}^{L\times L}$ be the block-diagonal matrix with $\Xi\Sigma^{-1}$ repeated n times on its diagonal, i.e.

$$S \triangleq \text{block.diag}(\Xi\Sigma^{-1}, \dots, \Xi\Sigma^{-1}). \tag{2.55}$$

Then,

$$M\Pi_M S = \begin{bmatrix} \text{diag}(v_1)\begin{bmatrix} U \\ U \end{bmatrix} & \dots & \text{diag}(v_{2N+1})\begin{bmatrix} U \\ U \end{bmatrix} \end{bmatrix}. \tag{2.56}$$

Applying another column permutation matrix Π_S yields

$$M\Pi_M S\Pi_S = \begin{bmatrix} \text{diag}(\widehat{u}_1)V & \dots & \text{diag}(\widehat{u}_{K+1})V \end{bmatrix}, \tag{2.57}$$

where $\widehat{u}_l \triangleq [u_l^T, u_l^T]^T$. Now, because S is an invertible matrix, and column permutations do not change the rank, we have

$$\text{rank}(M\Pi_M S\Pi_S) = \text{rank}(M\Pi_M S) = \text{rank}(M). \tag{2.58}$$

By Theorem 5 M has full column rank w.p. 1 when Ψ has elements i.i.d. distributed as $N(0,1)$. Now, for the SVD of $\Psi = U\Sigma\Xi^*$, it is well-known that the left singular vectors U are distributed uniformly on the Stiefel manifold, because the distribution of an iid Gaussian matrix is invariant to rotations (on both left and right). Then, using the second identity in (2.58), matrix $M\Pi_M S$ has also full rank w.p. 1 for U drawn uniformly at random from the Stiefel manifold $V_{(L)}\left(\mathbb{R}^P\right)$. Thus, selecting $\Psi = U$, and using the first identity in (2.58) establishes the corollary for the uniform distribution on the Stiefel manifold. However, this implies the same for any distribution that is absolutely continuous with respect to the latter. ⊠

Now, having stated the preliminary results of Theorem 5 and Corollary 1, we can prove the Theorem 2.

Recall that $L_1 = \widehat{\Theta} \bullet \widehat{\Psi}$, where $\widehat{\Theta} \in \mathbb{C}^{2P\times(2N+1)}$, and $\widehat{\Psi} \in \mathbb{R}^{2P\times(K+1)}$. Hence $L_1 \in \mathbb{C}^{2P\times(2N+1)(K+1)}$, and the stated condition $2P \geq (2N+1)(K+1)$ is necessary for L_1 to have full column rank. Because $\mathrm{rank}(L_1 L_1^T) = \mathrm{rank}(L_1)$, we consider $L_1 L_1^T$, which is more convenient to analyze.

$$\begin{aligned} L_1 L_1^T &= (\widehat{\Theta} \bullet \widehat{\Psi})(\widehat{\Theta} \bullet \widehat{\Psi})^T \\ &= (\widehat{\Theta} \bullet \widehat{\Psi})(\widehat{\Theta}^T \star \widehat{\Psi}^T) \end{aligned} \tag{2.59}$$

Using the mixed product property $(A \bullet B)(C \star D) = (AC) \odot (BD)$ [134],

$$\begin{aligned} L_1 L_1^T &= \widehat{\Theta}\widehat{\Theta}^T \odot \widehat{\Psi}\widehat{\Psi}^T \\ &= R \odot \tilde{R} \end{aligned} \tag{2.60}$$

where $R = \widehat{\Theta}\widehat{\Theta}^T$ and $\tilde{R} = \widehat{\Psi}\widehat{\Psi}^T$.

Thanks to the assumption that the P view angles $\theta_i \in [0,\pi], i = 1,\ldots,P$ are distinct, it follows that the $2P$ view angles $\theta_i, \theta_i + \pi, i = 1,\ldots,P$ are all distinct modulo 2π, and thus the $2P$ exponentials $e^{j\theta_i}, e^{j(\theta_i+\pi)}, i = 1,\ldots,P$ defining the rows of $\widehat{\Theta}^T = [\Theta^T,\ \bar{\Theta}^T]$ are all distinct. Now, up to scaling by a full-rank diagonal matrix, matrix $\widehat{\Theta}^T$ is a Vandermonde matrix with distinct bases, and therefore has full Kruskal rank [135]. Thus, $\mathrm{krank}(\widehat{\Theta}) = 2N+1$.

This implies $\widehat{\Theta}$ has full column rank, hence $\mathrm{rank}(R) = \mathrm{rank}(\widehat{\Theta}) = 2N+1$. Let the eigendecomposition of $R \succcurlyeq 0$ be

$$R = V\Lambda V^T = \sum_{k=1}^{2N+1} \lambda_k v_k v_k^T \tag{2.61}$$

Note that under the assumption on Ψ in Theorem 2 we have $(1/\sqrt{2})\widehat{\Psi}^T(1/\sqrt{2})\widehat{\Psi} = I$. It

therefore follows that the eigendecomposition of $\tilde{R} \succcurlyeq 0$ is

$$\tilde{R} = \widehat{\Psi}(0.5I)\widehat{\Psi}^T \tag{2.62}$$

Then, using the eigendecomposition property of the Hadamard product yields

$$L_1 L_1^T = R \odot \tilde{R} = 0.5 \sum_{k=1}^{2N+1} \sum_{l=1}^{K+1} \lambda_k (v_k \odot \widehat{\psi}_l)(v_k \odot \widehat{\psi}_l)^T \tag{2.63}$$

which implies that $\text{rank}(L_1)$ is equal to the number of $v_k \odot \widehat{\psi}_l$ that are linearly independent, upper bounded by $(2N+1)(K+1)$. Therefore, we have the following

$$\begin{aligned} \text{rank}(L_1) &= \text{rank}\left[\text{diag}(\widehat{\psi}_1)V \ldots \text{diag}(\widehat{\psi}_{K+1})V\right] \\ &\leq (2N+1)(K+1), \end{aligned} \tag{2.64}$$

where the matrix on the right hand side has dimensions $2P \times (2N+1)(K+1)$.

To apply Corollary 1 to (2.64), we require $\text{krank}(V^T)$. Recall that V contains the left singular vectors of $\widehat{\Theta}$. Consider the SVD $\widehat{\Theta}^T = U\Sigma V^T$. Because U and Σ are invertible, we have $\text{krank}(V^T) = \text{krank}(\widehat{\Theta}^T) = 2N+1$. Then, since Ψ is drawn randomly from the Stiefel manifold $V_{(K+1)}\left(\mathbb{R}^P\right)$ we may apply Corollary 1 with $n = 2N+1$ and $L = K+1$ to (2.64), to conclude that $\text{rank}(L_1) = (2N+1)(K+1)$ is satisfied w.p. 1, completing the proof of Theorem 2. ⊠

2.8.4 Proof of Theorem 3

To study the rank and condition number of $\mathbf{L}_2(\boldsymbol{\beta})$ we consider the Gram matrix $\mathbf{Y} \triangleq \mathbf{L}_2(\boldsymbol{\beta})^T\mathbf{L}_2(\boldsymbol{\beta})$ and then use the facts that $\text{rank}(A) = \text{rank}(A^TA),\ \forall A \in \mathbb{C}^{m\times n}$, and $\kappa(A^TA) = \kappa^2(A)$.

Using 2.15, we can write $L_2(\boldsymbol{\beta})^{(i)T}L_2(\boldsymbol{\beta})^{(i)}$ as

$$L_2(\boldsymbol{\beta})^{(i)T}L_2(\boldsymbol{\beta})^{(i)} = J\tilde{U}_i^T A_i \Gamma A_i^T \tilde{U}_i. \tag{2.65}$$

where $\tilde{U}_i = I_{K+1} \otimes \widehat{U}_{i:}$ and A_i is defined in (2.13).

Let $\lambda_1 \geq \ldots \geq \lambda_{d(K+1)}$ be the eigenvalues of Γ. Since $\lambda_1 I \succcurlyeq \Gamma \succcurlyeq \lambda_{d(K+1)} I$, we have

$$\lambda_{d(K+1)}\tilde{U}_i^T\tilde{U}_i \preccurlyeq \frac{L_2(\boldsymbol{\beta})^{(i)T}L_2(\boldsymbol{\beta})^{(i)}}{J(2N+1)} \preccurlyeq \lambda_1\tilde{U}_i^T\tilde{U}_i, \tag{2.66}$$

where $A_iA_i^T = ||\widehat{\Theta}_{i:}||_2^2 I_{K+1}$ using the mixed product identity, i.e. $(A \otimes B)(C \otimes D) =$

$(AC) \otimes (BD)$, and for all i, $||\widehat{\Theta}_{i:}||_2^2 = 2N + 1$.

Then, using the mixed product identity again,

$$\begin{aligned} \sum_{i=1}^{2P} \tilde{U}_i^T \tilde{U}_i &= \sum_{i=1}^{2P} I_{K+1} \otimes \widehat{U}_{i:}^T \widehat{U}_{i:}. \\ &= I_{K+1} \otimes (\widehat{U}^T \widehat{U}) \\ &= I_{K+1} \otimes I_d = I_{d(K+1)}. \end{aligned} \tag{2.67}$$

and combining (2.66) and (2.67) yields

$$\lambda_{d(K+1)} I_{d(K+1)} \preccurlyeq \frac{\mathbf{Y}}{J(2N+1)} \preccurlyeq \lambda_1 I_{d(K+1)}. \tag{2.68}$$

This result indicates that the condition number of $\mathbf{Y}$ is bounded by $\lambda_1/\lambda_{d(K+1)}$, i.e.

$$\kappa(\mathbf{Y}) \leq \kappa(\Gamma) = \lambda_1/\lambda_{d(K+1)}. \tag{2.69}$$

It follows that

$$\kappa(\mathbf{L}_2(\boldsymbol{\beta})) \leq \sqrt{\kappa(\Gamma)}, \tag{2.70}$$

establishing the upper bound on the condition number of $\mathbf{L}_2(\boldsymbol{\beta})$.

Next we state that based on the assumption $\Gamma \succ 0$, we have $\lambda_{d(K+1)} > 0$, and $\mathbf{Y}$ and $\mathbf{L}_2(\boldsymbol{\beta})$ are full column rank. Since $\Gamma \in \mathbb{R}^{(2N+1)(K+1)\times(2N+1)(K+1)}$, a necessary condition for $\Gamma \succ 0$ is $J \geq (2N+1)(K+1)$. A sufficient condition is that there are at least $(2N+1)(K+1)$ linearly independent vectors in the set $\{\beta(s_j), j = 1, \ldots, J\}$. $\boxtimes$

Chapter 3

Factorized Projection-domain Spatio-temporal Regularization for Dynamic Tomography

3.1 Introduction

As covered in Chapter 2.1, previous work on dynamic tomography [77], [78] provided an optimal sampling pattern, recovery algorithm, and theoretical guarantees for unique and stable reconstruction, however, was restricted to objects localized in space and in the spatio-temporal spectrum. Several methods [136], [137] recover the underlying motion field by incorporating Radon transform properties, but require a priori knowledge of the initial static object information. Other algorithms [79] alternate iteratively between recovering the motion field and the object. None of these iterative algorithms come with performance analysis, let alone uniqueness guarantees.

In dynamic MRI, there are several object-domain deep learning-based techniques [67], [138] that use a partially-separable model (PSM) for the dynamic object, and a deep image prior (DIP)-based algorithm [66]. Other methods for dynamic MRI [32], [114], [115] utilize low-rank matrix recovery algorithms. However, these methods were not directly tailored for dynamic tomography, and the solution is constrained by the PSM model.

Recent unsupervised dynamic tomography reconstruction algorithm with uniqueness and stability analysis [95] in Chapter 2 uses a special projection-domain bilinear PSM for dynamic

Material of this chapter has previously appeared in [97] and is adapted here with permission.

tomography reconstruction. In this work we use this algorithm as a component, extending the formulation to include spatial regularization, relax the PSM assumption, and obtain substantially improved accuracy of the recovery.

3.1.1 Contributions:

1. A novel method for dynamic tomography which, instead of costly initial temporal function estimates, incorporates an accurate temporal initialization from a fast projection-domain partially separable model. This leads to accelerated and improved convergence of the algorithm.

2. Provide additional flexibility to the reconstruction by not directly representing the object by a partially separable model, but using it for regularization instead.

3. The use of the spatio-temporal PSM and its relation to the projection domain PSM provide quantitative interpretability of recovery conditions and criteria for the selection of the time-sequential acquisition scheme for the projections.

4. The proposed method achieves better results with a simple spatial regularization, and we anticipate that the use of better spatial priors will further improve the method.

3.2 Object Domain Recovery Algorithm

Considering the same problem statement in Chapter 2.2, we propose the following variational formulation

$$\min_{f} \min_{\{\Lambda_k\},\{z_k\}} \sum_{t} \Big(||R_{\theta(t)} f_t - g_t||_2^2 + \lambda D(f_t) \Big) + \gamma ||f - \sum_{k} \Lambda_k \psi_k^T||_F^2 \tag{3.1}$$

where the spatio-temporal object f is estimated by solving the optimization problem. The first term in (3.1), where $R_{\theta(t)}$ denotes the Radon transform operator at view angle $\theta(t)$, $g_t = g(\cdot, \cdot, t)$, and $f_t = f(\cdot, t)$, equals to the squared error between the acquired projections and the projections of the estimated object, and is a standard data fidelity term. The second term in (3.1) is a conventional regularizer, which we choose in the experiments in this paper to be a simple spatial total-variation regularizer applied on each image frame. The last term in (3.1) represents the PSM spatio-temporal prior, which is key to the interpretability and performance improvements of the proposed algorithm. Here $\{\Lambda_k\}$ represent the set of $K+1$ spatial basis functions, and $\psi_k = U z_k$ with the interpolation operator $U \in \mathbb{R}^{P \times d}$ are temporal basis functions. The second term then expresses the prior that the object can be approximated by a PSM. Furthermore, to the extent that this approximation is good,

the analysis in Chapter 2.4.4 provides necessary conditions for unique recovery (without a regularizer D) and a quantitative tool to analyze the effect of the angular sampling pattern. Finally, constants γ and λ are used to adjust the weight of each term in the objective.

Note that for given $\{\psi_k\}$ the optimum $\{\Lambda_k\}$ can be determined in close-form. The last term then reduces to $||P_{R(\{\psi_k\})^\perp} f||_2^2$ where $P_{R(\{\psi_k\})^\perp}$ is the orthogonal projector onto the orthocomplement of the range space $R(\{\psi_k\})$. This term then encapsulates the same temporal evolution functions of the object f that are used in the projection-domain PSM, and eliminates the need for additional regularization to constrain the solution to a temporally smooth manifold as in other methods. Given $K+1$ latent temporal representations z_k obtained using the ProSep-Symm algorithm [95], as described in Chapter 2.5; these are used to initialize the optimization in (3.1) and speed up convergence.

3.3 Experiments

The proposed method is compared with the projection-domain partially separable model ProSep [95] in Chapter 2, and a generic deep prior method for video GMLR [112] which was also covered in Chapter 2.6. While optimizing the latent codes ξ_t to learn images of the object f_t from undersampled measurements, GMLR achieves temporal smoothness by constraining the set of ξ_t to be low-rank. When applied to dynamic tomography problem in our case, GLMR solves

$$\min_{\zeta_1,\ldots,\zeta_T;\gamma} \mathcal{L}(\boldsymbol{\zeta},\gamma) \quad \text{s.t.} \quad \operatorname{rank}(\boldsymbol{\zeta}) = r \quad \text{where} \quad \mathcal{L}(\boldsymbol{\zeta},\gamma) =$$
$$\lambda \sum_{t=1}^{P} \left\| g_t - \mathcal{R}_{\theta(t)} G_\gamma\left(\zeta_t\right) \right\|_2^2 + (1-\lambda) \sum_{t=1}^{P-1} \left\| \zeta_{t+1} - \zeta_t \right\|_2^2$$

where $g_t = \mathcal{R}_{\theta(t)} f_t$, G_γ is the generator with parameters γ, $\boldsymbol{\zeta} = [\zeta_1\ \zeta_2\ \ldots \zeta_P]$, and λ controls the similarity of consecutive latent codes.

3.3.1 Algorithm Settings

The Adam [126] optimization algorithm is used for all methods in the comparison. For GMLR the learning rates were selected as posted in the available implementation [125]. As in the comparison for [95], for each P, the GMLR model was optimized for $6(10)^4$ steps with $\lambda = 0.5$ and the low-rank constraint of $r = 4$.

Both GMLR and the proposed method produce a reconstructed object as output. In the case of ProSep, since it produces instead an estimate of the complete set of projections for

each t, we show, for each experimental setting, the respective FBP reconstruction.

In ProSep, the interpolation of the low-dimensional representations of the temporal functions is done using a cubic spline interpolator. The ProSep Symm algorithm leverages the π-symmetry of flipped projections in parallel beam CT to improve its estimates. Thus, we use it to provide the estimates of the temporal functions ψ_k used in the proposed algorithm.

The spatial regularizer for the proposed method was selected as the ℓ_1-norm of the discrete 2D total variation computed on the estimated object at each time instant, $D(f_t) = \mathrm{TV}(f(.,t))$. Also, the object f is initialized as the FBP reconstruction of the estimated projections of the ProSep-Symm algorithm. For walnut experiments, $\lambda = 0.05$ and $\gamma = 10^3$ for $P = 256$ and $\gamma = 10^4$ for $P = 512$ were used. For the compressed object, $\lambda = 10$ and $\gamma = 5(10^4)$ for $P = 256$ and $\gamma = 5(10^5)$ for $P = 512$ were used.

3.3.2 Tomographic Acquisition Scheme

All methods use the bit-reversed angular sampling scheme, over the range $[0, \pi]$ for ProSep, and over the range $[0, 2\pi)$ for all other methods.

3.3.3 Test Data

We used two different data sets, collected on a Carl Zeiss Xradia 520 Versa.

Walnut dataset: As in Chapter 2, the first is the slice of a CT reconstruction of walnut from the dataset [127], which is synthetically warped using a piecewise-affine time-varying warp [128] defined in Chapter 2.6 to produce a simulated ground-truth dynamic sequence of images at P time instants.

Compressed object dataset: The compressed object data set is obtained from a materials science experiment [139] with a sequence of nine increasing compression (loading) steps applied to an object, with a full set of radiographic projections collected (using Carl Zeiss Xradia 520 Versa) and reconstructed by the instrument's software at each step. To create a 2D+time test phantom using this quasi-static data set, a fixed axial slice is extracted from each of the nine reconstructions. These nine extracted slices corresponding to nine time points are interpolated to P time frames using a recent deep learning-based video interpolation algorithm [140]. This serves as the ground truth for our second phantom.

Three slices of this dynamically compressed object and the warped walnut at different t are shown in Figure 3.1. Note that during the materials science experiment, the object structure first distorts and then collapses, creating a significant and rapid change.

We note that both the warping algorithm used to create the walnut dynamic phantom and the interpolation algorithm used to generate the compressed object ground-truth data

are black box methods for all the dynamic tomography algorithms that we compare - that is, the algorithms are agnostic to these particular methods of generating the test data.

Time-sequential projection data for each experimental setting was simulated by numerical parallel-beam reprojection of the dynamic phantoms with a single 128-pixel projection at each of the P time instants, using the bit-reversed angular sampling scheme. The measurements were then corrupted by additive white noise with standard deviation $\sigma = 5(10^{-3}) \max(g)$. At this noise level, an FBP reconstruction (with a RAM-LAK filter) using a full set of P projections at each time instant has a PSNR of about 46 dB.

3.3.4 Run Time Comparison

All methods were run on a workstation with an Intel(R) Xeon(R) Gold 5320 CPU and NVIDIA RTX A6000 GPU. For P=512 the proposed method with K=7 is approximately three times faster per iteration than GMLR. Because GMLR with randomly initialized network typically requires 5 - 10× more iterations than the proposed method for convergence, the latter is about 15 - 30× faster than the GMLR.

3.3.5 Results

Figure 3.2 compares reconstructions for various methods for different objects and time instants. Despite providing sharper features, GMLR suffers from speckle-like artifacts and dark streaks. This may be due to lacking accurate spatial prior for the specific problem. However, it is difficult to detect the exact reason due to the complex structure and large number of hyperparameters of the algorithm. On the other hand, the proposed method further suppresses the spatial artifacts caused by the harmonic expansion in ProSep Symm. Table 3.1 reports PSNR (in dB), SSIM and mean absolute error (MAE) values for the estimated objects for the various methods. The results show consistent improvement by the proposed method over the other compared algorithms. To further explore GMLR, we use an ad-hoc method to suppress its high-frequency artifacts: we reproject the image frames created by it to a full set of P projections, and reconstruct them using FBP. This is abbreviated as GMLR-FBP in Table 3.1, showing improved metrics, but still below the proposed method, which also has the advantages of speed, interpretability, and theoretical analysis. Also, since metrics over the time interval may hide worst-case performances at individual t, we report worst per frame PSNR in Table 3.2. This again highlights the improvement introduced by the proposed algorithm.

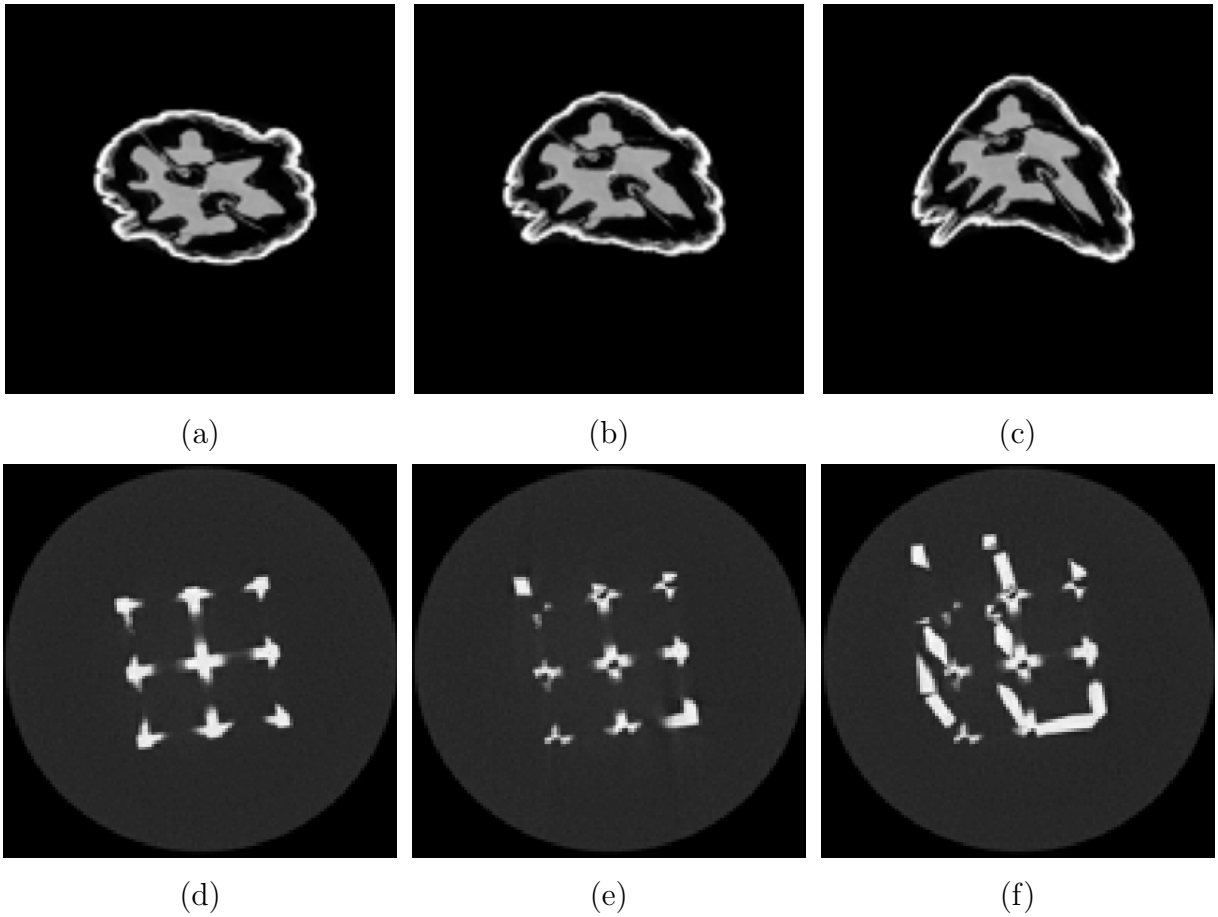


Figure 3.1: The synthetically warped walnut undergoing pointwise affine transform (a)-(c) and the compressed object (d)-(f) at time instances (a), (d) $t = 0$, (b), (e) $t = P/2$ and (c), (f) $t = P$.

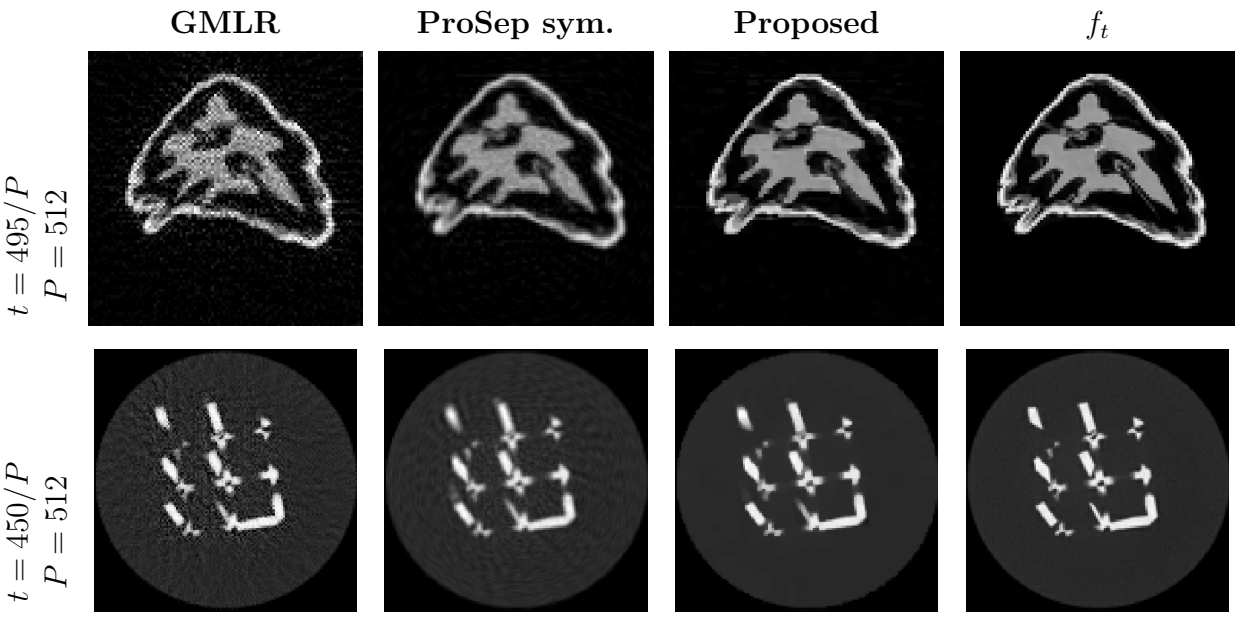


Figure 3.2: Comparison of reconstructions of the object in Fig. 3.1 at different time instances for different methods.

Table 3.1: Average accuracies for the complete 3D time-varying objects in Fig. 3.1 for different P.

P	Method	K	N	d	PSNR(dB)	SSIM	MAE
256	GMLR	-	-	-	23.1	0.76	0.039
	GMLR FBP	-	-	-	25.2	0.79	0.019
	ProSep symm.	5	30	6	26.5	0.88	0.022
	Proposed	5	-	-	**29.1**	**0.97**	**0.013**
512	GMLR	-	-	-	24.2	0.80	0.035
	GMLR FBP	-	-	-	27.1	0.86	0.015
	ProSep symm.	7	48	8	28.6	0.89	0.018
	Proposed	7	-	-	**32.3**	**0.97**	**0.011**

(a) Walnut

P	Method	K	N	d	PSNR(dB)	SSIM	MAE
256	GMLR	-	-	-	26.7	0.78	0.026
	GMLR FBP	-	-	-	29.6	0.84	0.017
	ProSep symm.	5	30	6	28.4	0.84	0.020
	Proposed	5	-	-	**31.8**	**0.96**	**0.010**
512	GMLR	-	-	-	28.9	0.80	0.020
	GMLR FBP	-	-	-	32.0	0.89	0.012
	ProSep symm.	7	48	8	30.1	0.85	0.017
	Proposed	7	-	-	**33.6**	**0.97**	**0.008**

(b) Compressed object

Table 3.2: Worst per frame PSNR values over the middle 80% of $P = 512$ frames. The first and the last 10% of the time interval are not included, to avoid the effect of algorithm transients.

Object	GMLR	ProSep Symm	Proposed
Walnut	23.7	27.6	**31.0**
Compressed	25.8	27.5	**29.9**

Chapter 4

RED-PSM: Regularization by Denoising of Partially Separable Models for Dynamic Imaging

4.1 Introduction

Time-varying or dynamic tomography involves the reconstruction of a dynamic object using its projections acquired sequentially in time.

The problem is a challenging ill-posed inverse problem: since the measurements are inconsistent due to the evolving object, traditional reconstruction algorithms lead to significant artifacts. As discussed in Chapter 1.4, although numerous methods have been proposed to address this problem, they suffer from various limitations, or provide less than satisfactory reconstruction quality, especially in the scenario of main interest in this chapter, when at each time instant only one projection is available.

4.1.1 Proposed Approach

We propose an approach, RED-PSM [97], [141], [142], which combines for the first time two powerful techniques to address this challenging imaging problem. The first technique, are non-parametric *factorized low-rank* object models, popularized under the name *partially separable models (PSMs)* [26]. PSM was combined with low-rank matrix recovery in [32]–[34], and

Material of this chapter has previously appeared in [141]–[144] and is adapted here with permission. Code is available at `https://github.com/berkiskender/RED-PSM`.

used since then to represent or motivate a *low-rank object prior* in tens of works on dynamic imaging. The second technique is the recent *Regularization by Denoising (RED)* [145], which provides a flexible framework to exploit the impressive performance of state-of-the-art image denoising algorithms, for various inverse problems.

We propose a non-convex partially separable objective with RED for learning-regularized low-rank spatio-temporal object recovery and a computationally efficient and scalable optimization scheme with bi-convex ADMM.

Theoretical analysis proves the convergence of our objective to a value corresponding to a stationary point satisfying the first-order optimality conditions. Convergence is accelerated by a particular projection-domain-based initialization.

We demonstrate the performance and computational improvements of our proposed RED-PSM with a learned image denoiser by comparing it to a recent deep-prior-based method known as TD-DIP [66], and to spatial and spatiotemporal total variation (TV) regularized versions of PSM with a bi-convex objective.

Although the main focus is on dynamic tomography, we also show performance advantages of RED-PSM in a cardiac dynamic MRI setting.

RED was originally inspired by the Plug-and-Play (PnP) approach [146], which was the first to exploit powerful denoisers to represent the prior on the object in inverse problems but without explicitly defining the regularizer. Instead, in PnP the denoiser is incorporated by replacing a proximal mapping. Both PnP and RED have shown strong empirical performance in various applications. A recent survey [147] provides a detailed discussion of the main results and different applications of PnP and the relation to RED. For the algorithmic variations PnP-ADMM and PnP-ISTA, there are convergence results to unique fixed points. Moreover, for PnP-ISTA, similar to our result, convergence is shown to a stationary point of a possibly non-convex objective when the denoiser is an MMSE estimator [148]. This result requires the prior to be non-degenerate, i.e., not lie on a lower dimensional manifold, which, unfortunately, would be violated by the low-rank bilinear representation in our PSM. We believe that a PnP-based method similar to RED-PSM can be formulated for our problem, but a proof of convergence to a stationary point may not directly follow, and a different approach and analysis may be needed.

Our motivation for using the original RED formulation in RED-PSM rather than more recent formulations and analyses (e.g., as in [149] and the references therein) are its simple gradient expression, and explicit regularizer expression facilitating implementation and theoretical analysis. The objective is to demonstrate the large improvement in performance that this simple learned regularizer can bring to the PSM scheme, and the improvement over much more complex and slower state-of-the-art methods. Recent variations on RED such as

in [149], and PnP methods (as covered in [147]) have improved over the original methods in reconstruction quality in various static imaging problems and/or in theoretical guarantees. However, theoretical guarantees for the case of a non-convex data fidelity term have been provided recently for only one such non-convex scenario - phase-retrieval [150]. As this is the first work to extend RED to spatio-temporal imaging with a PSM model, and to provide a convergence guarantee for the method in the non-convex bilinear scenario, we chose to limit our attention in this chapter to the original RED version [145], [151].

4.1.2 Contributions

1. To the best of our knowledge, RED-PSM is the first PSM-based approach to dynamic imaging that incorporates a particular (RED-based [145]) spatial prior that is learned and pre-trained.

2. We are not aware of any prior work that uses RED with an explicit low-rankness constraint. In RED-PSM, we achieve parsimonious representation for the dynamic object using PSM, and reduce the data requirements further by incorporating the RED prior.

3. Unlike supervised learning-based methods for spatio-temporal imaging [52]–[65], [152]–[154] that learn a spatio-temporal model from training data, RED-PSM does not require ground-truth spatio-temporal training data, which is often not available or expensive to acquire. Thus RED-PSM offers the best of both worlds: a learned spatial model using readily available *static* training data, and unsupervised single-instance adaptation to the spatio-temporal measurement.

4. A novel and effective ADMM algorithm for the resulting new minimization problem enforces PSM as a hard constraint and incorporates the learned RED spatial regularizer.

5. The method is supported by theoretical analysis: we show convergence of the proposed PSM-based objective to a stationary point, and do so for a highly non-trivial learned regularizer. To date, the convergence of RED has been analyzed in the nonlinear measurement setting only for the phase retrieval problem [150]. Our analysis of RED-PSM is the first convergence result of RED in the bilinear (bi-convex), non-convex scenario. This convergence guarantee is of practical importance. First, it provides a mathematical justification to terminate the algorithm after sufficiently many iterations. Second, it eliminates concerns about the solution degrading after too many iterations as in the case of DIP-based methods.

6. Compared to a recent DIP-based [66] algorithm, RED-PSM achieves better reconstruction accuracy with orders of magnitude faster run times.

7. To improve and speed up the empirical convergence of RED-PSM, we use a particular fast projection-domain PSM initialization scheme [95], [96] for the spatial and temporal basis functions. The accelerated and reliable convergence with this initialization scheme is important for applications that require fast turn-around between acquisition and reconstruction.

8. A version of the approach with a patch-based regularizer is shown to provide almost equivalent reconstruction accuracy. This makes the proposed method conveniently scalable to high-resolution 3D or 4D settings.

9. We also show that RED-PSM is robust to a decrease in the number of available distinct projection angles, which is important for settings in which projections can be acquired nearly simultaneously and repeatedly by a few sources at a small set of fixed view angles.

4.2 Proposed Method: RED-PSM

The proposed approach mainly tackles the ill-posed dynamic tomography problem with time-sequential projections covered in Chapter 2.2. In this approach, we also extend our experiments to a dMRI scenario and a dCT scenario where multiple projections at a time are obtained.

The overall RED-PSM framework for recovering dynamic objects explained in this Chapter is illustrated in Figure 4.1.

4.2.1 Variational Formulation

We use a discretized version of the PSM (2.1) for the dynamic object, with the object $f(\cdot, t)$ at each time instant $t = 0, 1, \ldots, P-1$, represented by a $N \times N$-pixel image (a "time frame", or "snapshot"). Vectorizing these images to vectors $f_t \in \mathbb{R}^{N^2}$, the entire dynamic object is the $N^2 \times P$ matrix $f = [f_0 \ldots f_{P-1}]$. It will also be useful to extract individual time frames from f. Denoting the t-th column of the $P \times P$ identity matrix by e_t, we have $fe_t = f_t$, i.e., e_t extracts the t-th column of f. Applying the PSM model, we assume $f = \Lambda\Psi^T \in \mathbb{R}^{N^2 \times P}$, where the columns Λ_k and Ψ_k of $\Lambda \in \mathbb{R}^{N^2 \times K}$ and $\Psi \in \mathbb{R}^{P \times K}$ are the discretized spatial and temporal basis functions for the PSM representation, respectively.

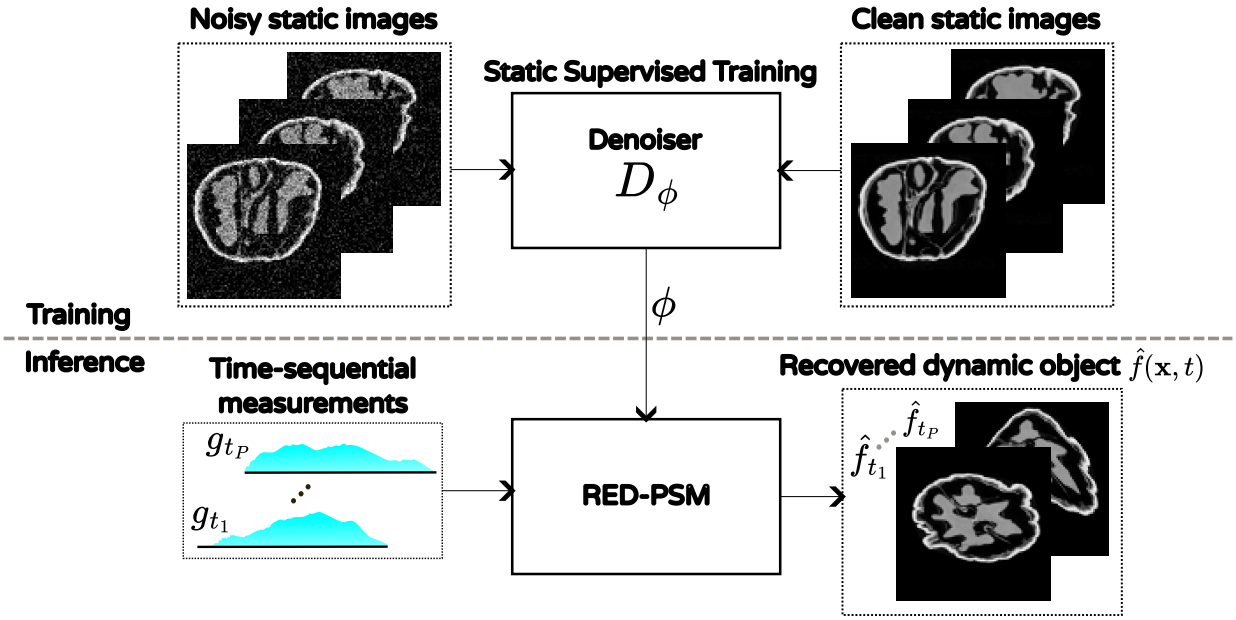


Figure 4.1: The RED-PSM framework. The deep denoiser D_ϕ is trained on slices of *static* objects similar to the object of interest, and the learned spatial prior is used at inference time.

Assuming that the x-ray detector has N bins, the projection of the object at time t is $g_t = g(\cdot, \theta_t, t) = R_{\theta_t} f_t \in \mathbb{R}^N$, where the measurement matrix $R_{\theta_t} \in \mathbb{R}^{N \times N^2}$ computes the projection at view angle θ_t.

We formulate the recovery of f as the solution $\hat{f} = \hat{\Lambda}\hat{\Psi}^T$ to the following variational problem

$$(\hat{\Lambda}, \hat{\Psi}) = \arg\min_{\Lambda, \Psi} \sum_{t=0}^{P-1} \left(\left\| R_{\theta_t} \Lambda \Psi^T e_t - g_t \right\|_2^2 + \lambda \rho(\Lambda \Psi^T e_t) \right) + \xi \|\Psi\|_F^2 + \xi \|\Lambda\|_F^2 \quad \text{where} \quad \Psi = UZ. \tag{4.1}$$

The first term is the data fidelity term measuring the fit between available undersampled measurements g_t of the true object and the measurements obtained from the estimated object $f = \Lambda \Psi^T \in \mathbb{R}^{N^2 \times P}$. The second term with weight $\lambda > 0$ is a *spatial* regularizer injecting relevant spatial prior to the problem. It is applied to the PSM $\Lambda \Psi^T$ column by column, that is, to individual temporal image frames. The last two terms with weight $\xi > 0$ prevent the bilinear factors from growing without limit. [1]

[1] Inclusion of these Frobenius norm terms also happens to lead to connections with nuclear norm, but, as discussed in detail in Section 4.3, our problem formulation (even if the RED regularizer is disregarded) is *not* equivalent to a formulation encouraging a low-rank solution by a nuclear norm or other Schatten norm

Finally, the identity $\Psi = UZ$ is an implicit *temporal* regularizer that restricts the temporal basis functions Ψ to a d-dimensional subspace of $\mathbb{R}^P$ spanned by a fixed basis $U \in \mathbb{R}^{P\times d}$ where $d \geq K$. The action of U on columns of Z may be interpreted as interpolation from d samples to the P-sample temporal basis functions. In practice, we incorporate this identity by explicit substitution (reparametrization of Ψ in terms of the free variable $Z \in \mathbb{Z}^{d\times K}$) into the objective, and the minimization in (4.1) over Ψ is thus replaced by minimization over Z. This reduces the number of degrees of freedom in Ψ to a fixed number dK, independent of the number P of temporal sampling instants. For notational conciseness, we do not display this constraint/reparametrization in the sequel, but it is used throughout.

4.2.2 Incorporating Regularization by Denoising

For the spatial regularizer $\rho(\cdot)$ we consider "Regularization by Denoising (RED)" [145]. RED proposes a recovery method using a denoising operator $D : \mathbb{R}^{N^2} \to \mathbb{R}^{N^2}$ in an explicit regularizer of the form

$$\rho(f_t) = \frac{1}{2} f_t^T (f_t - D(f_t)). \tag{4.2}$$

Recent studies using this regularizer provide impressive performances for various *static* image reconstruction tasks including high-dimensional cases [155] and phase retrieval [156]. This regularizer was also combined with a DIP-based fidelity term in [157]. However, we avoid this due to disadvantages related to speed, overfitting, and convergence guarantees.

While it provides significant flexibility for the type of denoisers that can be used, RED still requires D to be differentiable and locally homogeneous, and to satisfy the passivity condition $\|D(f_t)\| \leq \|f_t\|$, for the theoretical analysis of RED to apply. [2]

Next, we consider the optimization in (4.1). For the conventional variational formulation

$$\hat{f}_t = \arg\min_{f_t} \|R_{\theta_t} f_t - g_t\|_2^2 + \lambda\rho(f_t),$$

an efficient choice are iterative algorithms [145] that use the standard "early termination" approach [158], and only require a single use of the denoiser per iteration. However, the regularized PSM objective in (4.1) does not allow to propagate the RED updates on f to the respective basis functions efficiently. To overcome this difficulty, we perform a bilinear

penalty such as used in [37], [39], [40]. Instead, the factorized form $f = \Lambda\Psi^T$ (as in other PSM-based matrix recovery works such as [32], [33]) introduces this rank constraint explicitly.

[2] While many powerful denoisers have been demonstrated to satisfy these conditions in [145], recent work [151] provides another framework to explain the good performance of RED with denoisers not satisfying them.

variable splitting $f = \Lambda\Psi^T$ and obtain our final formulation

$$\min_{f,\Lambda,\Psi} H(f,\Lambda,\Psi) \quad \text{s.t.} \quad f = \Lambda\Psi^T \tag{4.3}$$

$$\text{where } H(f,\Lambda,\Psi) = \sum_{t=0}^{P-1} \left(\left\| R_{\theta_t} \Lambda\Psi^T e_t - g_t \right\|_2^2 + \lambda\rho(fe_t) \right) + \xi\|\Lambda\|_F^2 + \xi\|\Psi\|_F^2. \tag{4.4}$$

Since the PSM is enforced as a hard constraint, the estimated object f is constrained to have $\text{rank}(f) \leq K$. Problem (4.3) is non-convex even if ρ is convex, because of the product between unknowns Λ and Ψ.

We propose an algorithm based on ADMM to solve (4.3). To this end, we form the augmented Lagrangian in the scaled form, [159], [160]

$$\mathcal{L}_\beta[\Lambda,\Psi,f;\gamma] = \sum_t \left(\left\| R_{\theta_t} \Lambda\Psi^T e_t - g_t \right\|_2^2 + \lambda\rho(fe_t) \right) + \xi\|\Psi\|_F^2 + \xi\|\Lambda\|_F^2 - \frac{\beta}{2}\|\gamma\|_F^2 + \frac{\beta}{2}\|\Lambda\Psi^T - f + \gamma\|_F^2, \tag{4.5}$$

where $\gamma \in \mathbb{R}^{P\times N^2}$ represents the scaled dual variable associated with the constraint $f = \Lambda\Psi^T$ and $\beta > 0$ is the penalty parameter.

Then, ADMM can be used to solve (4.5) as in Algorithm 1.

Algorithm 1 RED-PSM

input: $\Lambda^{(0)}$, $\Psi^{(0)}$, $\gamma^{(0)}$, $f^{(0)} = \Lambda^{(0)}\Psi^{(0)T}$, $\beta > 0$, $\lambda > 0$, $\xi > 0$

1: **for** $i \in \{1,\dots,I\}$ **do**
2: $\quad \Lambda^{(i)} = \arg\min_\Lambda \{\sum_t \|R_{\theta_t}\Lambda\Psi^{(i-1)T}e_t - g_t\|_2^2 + \frac{\beta}{2}\|\Lambda\Psi^{(i-1)T} - f^{(i-1)} + \gamma^{(i-1)}\|_F^2 + \xi\|\Lambda\|_F^2\}$
3: $\quad \Psi^{(i)} = \arg\min_\Psi \{\sum_t \|R_{\theta_t}\Lambda^{(i)}\Psi^T e_t - g_t\|_2^2 + \frac{\beta}{2}\|\Lambda^{(i)}\Psi^T + \gamma^{(i-1)} - f^{(i-1)}\|_F^2 + \xi\|\Psi\|_F^2\}$
4: $\quad \forall t: \; f_t^{(i)} = \arg\min_{f_t} \{\lambda\rho(f_t) + \frac{\beta}{2}\|(\Lambda^{(i)}\Psi^{(i)T} + \gamma^{(i-1)})e_t - f_t\|_2^2\}$
5: $\quad \gamma^{(i)} = \gamma^{(i-1)} + \Lambda^{(i)}\Psi^{(i)T} - f^{(i)}$
6: **end for**

Line 4 in Algorithm 1 then corresponds to the variational denoising for all t of the "pseudo image frame" $(\Lambda^{(i)}\Psi^{(i)T} + \gamma^{(i-1)})e_t$ with regularization $\lambda\rho(f_t)$. Instead of solving this denoising problem by running an iterative algorithm to convergence, we follow the RED approach [145], [151], and for each t replace the f_t update in Step 4 by a single fixed-point iteration step

using the approach of early stopping [158] (Sec. 4.3.2), taking advantage of the gradient rule

$$\nabla\rho(f_t) = f_t - D_\phi(f_t) \tag{4.6}$$

where D_ϕ is the denoiser. This results in Algorithm 2, which requires only single use of the denoiser per iteration of ADMM. Furthermore, as in Algorithm 1, the modified Line 4 in Algorithm 2 can be performed in parallel for all t.

Algorithm 2 RED-PSM with efficient f step

Notes: Inputs, and Lines 1-3 and 5-6 are the same as Algorithm 1. The f step is applied $\forall t$.

4: $\forall t: \quad f_t^{(i)} = \frac{\lambda}{\lambda+\beta} D_\phi(f_t^{(i-1)}) + \frac{\beta}{\lambda+\beta}\left(\Lambda^{(i)}\Psi^{(i)T} + \gamma^{(i-1)}\right)e_t$

4.2.3 Regularization Denoiser

The regularization denoiser D_ϕ has a deep neural network CNN (DnCNN) [161] architecture and is trained in a supervised manner on a training set of 2D slices $f_i \in \mathbb{R}^{N^2}, i = 1, \ldots N$ of one or more *static* objects similar to the object of interest, assuming that such data will be available in the settings of interest. Thus, the RED steps are agnostic to the specific motion type. The training objective for the denoiser is

$$\min_\phi \sum_i \|f_i - D_\phi(\tilde{f}_i)\|_F^2 \quad s.t. \quad \tilde{f}_i = f_i + \eta_i, \ \forall i, \tag{4.7}$$

where the injected noise $\eta_i \sim \mathcal{N}(0, \sigma_i^2 I)$ has noise level $\sigma_i \sim U[0, \sigma_{\max}]$ spanning a range of values, so that the denoiser learns to denoise data with various noise levels.

4.2.4 Computational Cost

Space (memory requirements) and time (operation count) complexities for two variants of the RED-PSM bilinear ADMM Algorithm are shown in Table 4.1. The operation counts are for a single outer iteration. As the number of outer iterations typically has a weak dependence on the size of the problem, the scaling shown tends to determine the run time. Furthermore, thanks to its structure, the algorithm also offers many opportunities for easy parallelization, so that actual runtime can be proportionally reduced by allocating greater computational resources. See Section 4.6.2 for the detailed analysis of computational requirements.

The complexities for the proposed Algorithm 2 are given in the first column of Table 4.1. Space complexity is dominated by the storage of Λ, and scales proportionally to image cross-

section size in pixels N^2, and the order K of the PSM. The time complexity is dominated by the computations of the gradient with respect to Λ, and scales proportionally to the size N^2P of the spatio-temporal object, PSM order K, and the number M_i of inner iterations used to solve optimization subproblems (Lines 2 and 3 in Algorithm 2).

Finally, in the second column of Table 4.1, we investigate the patch-based version of the proposed algorithm described in Section 4.4.4. Given a patch size $N_B \ll N$ and stride $s \leq N_B$, this alternative increases the operation count by $(\frac{N_B}{s})^2$ since it operates on overlapping patches. For example, for the settings in our experiments, $s = \frac{N_B}{2}$, increasing the operation count by a factor of 4. This is a modest price to pay, because this alternative also reduces the space complexity by a factor of $(\frac{N_B}{N})^2$. For example, for N_B=8 and N=128 as used in our experiments, this corresponds to a reduction by a factor of 144 in space complexity. Thus, this variant of RED-PSM enables scaling for high-dimensional and high-resolution settings.

Table 4.1: Time and space complexities for two variants of the RED-PSM algorithm for a single outer iteration of the bilinear ADMM. M_i is the number of inner iterations for each iteratively solved subproblem, and $N_B \ll N$, $s \leq N_B$.

Complexity	PSM-fidelity	Patch-based
Space	$O(KN^2)$	$O(KN_B^2)$
Time	$O(KN^2PM_i)$	$O(KN^2PM_i(\frac{N_B}{s})^2)$

4.3 Convergence Analysis

In this section, we follow an approach similar to recent work on ADMM for a bilinear model [162] to analyze convergence. We show that under mild technical conditions, the objective in Algorithm 1 is guaranteed to converge (with increasing number I of iterations) to a value corresponding to a stationary point of the Lagrangian, that is, satisfying the necessary conditions for first order optimality.

In practice, Algorithm 2 with the efficient f-step version, which we implemented and used in the experiments reported in Section 4.4, has better run times, and rapid empirical convergence. However, its analysis requires additional steps, which are not particularly illuminating. Therefore we focus on the analysis of the nominal Algorithm 1.

In spite of the similarity, at a high level, to the problem and analysis in [162], our problem formulation and algorithm differ in several aspects, which require modifications in the analysis and proof. In particular, different to [162], where the bilinear form only appears in the constraint, the bilinear form appears both in our objective function and constraint. To satisfy

strong convexity for the respective subproblems, our objective also uses the Frobenius norm terms for the PSM factors, instead of the interim proximal terms in [162]. Moreover, to allow the efficient f-step in Algorithm 2, we use a different order of updates than [162]. Importantly, we account for the explicit RED regularization and its required properties in the proof of convergence. Our analysis includes a proof (similar to [163]) of the boundedness of the iterates to justify the existence of points of accumulation of the iterate sequence, which appears to have been inadvertently left out in [162]. However, different from [163], which has objectives with linear constraints, our objective (4.1) has a bilinear constraint.

To simplify the notation, we replace the separate computation of the projections of time t-frames in the data fidelity term by using the operator $\bar{R} : \mathbb{R}^{N^2\times P} \to \mathbb{R}^{N\times P}$ that computes the entire set of P projections at view angles θ_t of the image frames at times t, $t = 1, \ldots, P$ of dynamic image f, i.e, of each of its columns indexed by t, producing $g = \bar{R}f \in \mathbb{R}^{N\times P}$. When applied to the PSM, $\bar{R}$ performs $R_{\theta_t}\Lambda\Psi^T e_t$ for each t. We also aggregate the contribution of the RED regularizer into $\bar{\rho}(f) \triangleq \sum_{t=0}^{P-1} \rho(f_t) : \mathbb{R}^{N^2\times P} \to \mathbb{R}$, and the denoiser into $\bar{D} : \mathbb{R}^{N^2\times P} \to \mathbb{R}^{N^2\times P}$, which performs D for each column of f indexed by t. Then, the Lagrangian function with dual variable γ can be rewritten as

$$\mathcal{L}[\Lambda, \Psi, f; \gamma] = \|\bar{R}\Lambda\Psi^T - g\|_F^2 + \lambda\bar{\rho}(f) + \xi\|\Lambda\|_F^2 + \xi\|\Psi\|_F^2 + \beta\langle\gamma, (\Lambda\Psi^T - f)\rangle, \tag{4.8}$$

where the inner product is defined as $\langle A, B\rangle = \mathrm{Tr}(A^T B)$.

The corresponding augmented Lagrangian is

$$\begin{aligned}\mathcal{L}_\beta[\Lambda, \Psi, f; \gamma] &= \|\bar{R}\Lambda\Psi^T - g\|_F^2 + \lambda\bar{\rho}(f) \\ &+ \xi\|\Lambda\|_F^2 + \xi\|\Psi\|_F^2 + \beta\langle\gamma, (\Lambda\Psi^T - f)\rangle + \frac{\beta}{2}\|\Lambda\Psi^T - f\|_F^2.\end{aligned} \tag{4.9}$$

Then, we can state the subproblems with respect to each primal variable in Algorithm 1 as

$$\begin{aligned}S_\Lambda &= \|\bar{R}\Lambda\Psi^{(i-1)T} - g\|_F^2 + \xi\|\Lambda\|_F^2 \\ &\quad + \frac{\beta}{2}\|\Lambda\Psi^{(i-1)T} - f^{(i-1)} + \gamma^{(i-1)}\|_F^2 \qquad (4.10)\\ S_\Psi &= \|\bar{R}\Lambda^{(i)}\Psi^T - g\|_F^2 + \xi\|\Psi\|_F^2 \\ &\quad + \frac{\beta}{2}\|\Lambda^{(i)}\Psi^T - f^{(i-1)} + \gamma^{(i-1)}\|_F^2, \text{ and} \qquad (4.11)\\ S_f &= \lambda\bar{\rho}(f) + \frac{\beta}{2}\|\Lambda^{(i)}\Psi^{(i)T} - f + \gamma^{(i-1)}\|_F^2. \qquad (4.12)\end{aligned}$$

Algorithm 1 will be shown in Theorem 6 to converge to a stationary point of Problem

(4.3), which is defined below.

Definition 1 (Stationary solution for Problem (4.3)). The point $W^* = (\Lambda^*, \Psi^*, f^*, \gamma^*)$ is a stationary solution of the problem (4.3) if it satisfies the stationarity and primal feasibility conditions for the variables of the Lagrangian in (4.8):

$$\nabla_f \mathcal{L}(\Lambda^*, \Psi^*, f^*; \gamma^*) = \lambda \nabla \bar{\rho}(f^*) - \beta\gamma^* = 0; \tag{4.13a}$$

$$\nabla_\Lambda \mathcal{L}(\Lambda^*, \Psi^*, f^*; \gamma^*) = (2\bar{R}^T(\bar{R}\Lambda^*\Psi^{*T} - g) + \beta\gamma^*)\Psi^* + 2\xi\Lambda^* = 0; \tag{4.13b}$$

$$\nabla_\Psi \mathcal{L}(\Lambda^*, \Psi^*, f^*; \gamma^*) = (2\bar{R}^T(\bar{R}\Lambda^*\Psi^{*T} - g) + \beta\gamma^*)^T\Lambda^* + 2\xi\Psi^* = 0; \tag{4.13c}$$

$$f^* = (\Lambda\Psi^T)^{(*)}. \tag{4.13d}$$

By its definition, a stationary solution of (4.3) satisfies the necessary first-order conditions for optimality.

To state Theorem 6, we need to introduce some additional definitions pertaining to the denoiser D used in RED.

Definition 2 (Strong Passivity [145]). A function D is strongly passive if $\|D(f)\| \leq \|f\|$.

Moreover, we assume the gradient rule of RED (4.6). This is shown to hold true when the denoiser D is assumed locally homogeneous (LH) [145] and Jacobian symmetric (JS) [151]. It was shown [145] that several popular denoisers are practically LH, and that for (4.6) to hold for the explicit regularizer ρ in (4.2), JS is necessary [151].

In our analysis, we only require the gradient rule (4.6) to hold, rather than the explicit form (4.2) of the regularizer. Even for denoisers without LH and JS, the gradient rule can be used to define the fixed point iteration in the algorithm [151], and, more recent works on using denoisers for regularization [164]–[168] (in particular [166]) shows how to learn a denoiser jointly with the gradient so that the gradient rule is satisfied.

Finally, we will also assume Lipschitz continuity of the denoiser with some Lipschitz constant L_D, i,e., $\|D_\phi(f_1) - D_\phi(f_2)\|_2 \leq L_D\|f_1 - f_2\|_2$ for all images $f_1, f_2 \in \mathbb{R}^{N^2}$. This means that the denoiser has some finite gain L_D, which is satisfied by any reasonable denoiser.

Our main convergence result is the following.

Theorem 6. *Suppose that the denoiser D_ϕ satisfies the gradient rule in (4.6), and is Lipschitz continuous with some Lipschitz constant L_D, and strongly passive.*

Then, if $\beta > 2L$ where $L \triangleq \lambda(1 + L_D)$, Algorithm 1 converges globally (i.e., regardless of initialization) to a stationary solution (f^*, Λ^*, Ψ^*) of (4.3) in the following sense:

(i) The sequence of values of the objective H converges to a limit $H^* = H(f^*, \Lambda^*, \Psi^*)$; and (ii) The iterates generated by Algorithm 1 converge subsequentially to a stationary solution $(f^*, \Lambda^*, \Psi^*, \gamma^*)$ of (4.3), that is, any accumulation point (of which there is at least one) of the sequence $(f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}, \gamma^{(i)})$ is a stationary solution of (4.3).

The proof of Theorem 6 is provided in Section 4.6.4.

Regarding convergence to the globally optimal solution, more can be said. Problem (4.1) is non-convex even if $\bar{\rho}$ is convex, because of the product between unknowns Λ and Ψ. However, similar to Theorem 7.1 of [169] (which generalizes arguments in [170],[171]), thanks to the inclusion of the Frobenius norms of the factors Λ and Ψ in the bilinear form, when $\bar{\rho}$ is convex (i.e., when the gradient rule (4.6) is assumed to hold), the global minimum $\hat{f} = \hat{\Lambda}\hat{\Psi}^T$ of (4.1) can be shown to coincide with the global minimum in

$$\hat{f} = \arg\min_{f} \sum_{t=0}^{P-1} \|R_{\theta_t} f_t - g_t\|_2^2 + \lambda\bar{\rho}(f) + 2\xi\|f\|_* \quad \text{s.t. } \operatorname{rank}(f) \leq K, \tag{4.14}$$

where the penalty weight $\xi > 0$ is identical to the Frobenius norm penalty in our RED-PSM objective (4.3).

Importantly, Problem (4.14) is non-convex too, because of the rank constraint. In other words, it is *not* equivalent to a conventional nuclear norm penalized problem. Hence, although the Frobenius norm penalties on the PSM factors in our problem formulation (4.1) lead to a connection to nuclear norm, Problem (4.1) is *not equivalent* to a formulation encouraging a low-rank solution by a nuclear norm penalty.

However, returning to the question of determining global optimality of a candidate solution to (4.14), consider the convex version of Problem (4.14), without the rank constraint, and denote its solution by $\hat{f}^{\text{Convex}}$. If it so happens that $\operatorname{rank}(\hat{f}^{\text{Convex}}) \leq K$, then the low-rank constraint in Problem (4.14) is not active, and $\hat{f}^{\text{Convex}} = \hat{f}$. It then follows [169] that the optimality conditions of the convex problem (without the rank constraint) can be used to assess the global optimality of a candidate solution $\hat{f} = \hat{\Lambda}\hat{\Psi}^T$ produced by the algorithm solving (4.1).

4.4 Experiments

4.4.1 Datasets

Three categories of data sets are used in this work.

Walnut Dataset: We use the CT reconstructions of two different (static) walnut objects from the publicly available 3D walnut CT dataset [127]. As defined in Chapter 2.6, we create a dynamic test object by synthetically warping the central axial slice of one of the walnut objects using a sinusoidal piecewise-affine time-varying warp [128]. *Static* axial, coronal, and sagittal slices of the other walnut object are used to train the denoiser D_ϕ.

Compressed Object Dataset: As in Chapter 3.3, we use the same fixed axial slice of the quasi-static compressed material dataset [139] with 9 loading steps interpolated to P time frames using the deep learning-based method [140]. The denoiser D_ϕ for our experiments on this data set was trained using the axial slices of the *static* pre-compression and post-compression versions of the object, which would be available in actual dynamic loading experiments of this type.

MPM-based Deformation Dataset: We test the proposed method on dynamic CT data [172] synthesized using an accurate numerical method, known as the Material Point Method (MPM)-based deformation [173], [174]. The dataset includes 10 different 3D time-varying object shapes of size $80 \times 80 \times 80$, each deforming under various mechanical loading settings, resulting in a total of 157 different scenarios. For each scenario, the data set contains only 10 ground-truth temporal frames, uniformly spaced in time. For our experiments, we used data for two such scenarios of two different dynamic objects with distinct shapes and mechanical loadings. For each scenario, we extracted a single 80×80 fixed-position coronal slice from each of the $T = 10$ time instants. The ground-truth slices for five (every other) of these instances are shown in Figure 4.3. We note that the setup considered in our experiments is much more challenging than the one in [104], which uses for each scenario 90 projections uniformly spaced in time: our setup is temporally undersampled (by a factor of 9) and contains sharper transitions. Having just a handful (10) of ground-truth instants for each object, instead of a time-sequential acquisition, for each object we obtain, by the numerical projection of the ground-truth frames, *multiple projections* at a given time instant. We use the bit-reversed angular scheme for the projections (See Figure 4.4 for details).

We note that all algorithms compared in this work are agnostic to both the synthetic warp applied on the static walnut slice and to the data-driven interpolation method used for the compressed object.

Spatio-temporal projection data for each dataset is simulated by a parallel-beam projection with N=128 detector bins of the dynamic phantoms, a single projection at each of the P time instants. The sequence of projection angles $\{\theta_t\}_{t=0}^{P-1}$ (a free experimental design parameter) was chosen to follow the bit-reversed view angle sampling scheme, which has been shown [95] to provide an especially favorable conditioning of the recovery problem. The simulated measurements are corrupted using AWGN with standard deviation $\sigma = 5 \cdot 10^{-3}$. This noise

level leads to the FBP (with Ram-Lak filter) of the full set of *P=512 projections at each time instant* having a PSNR of approximately 46 dB. When, in the actual experiments with sequentially sampled data, only $1/P$ of this data is used, the PSNR of the reconstruction may be expected to be lower.

Ground-truth frames for $P = 4$ are shown in Figure 4.2.

Cardiac dMRI Dataset: For a more direct comparison with the setting and data used in dMRI works, we also test RED-PSM on the "retrospective" cardiac dMRI data in [66]. The data includes 23 distinct time frames for one cardiac cycle. Details of the data and experiments are in Section 4.4.4.

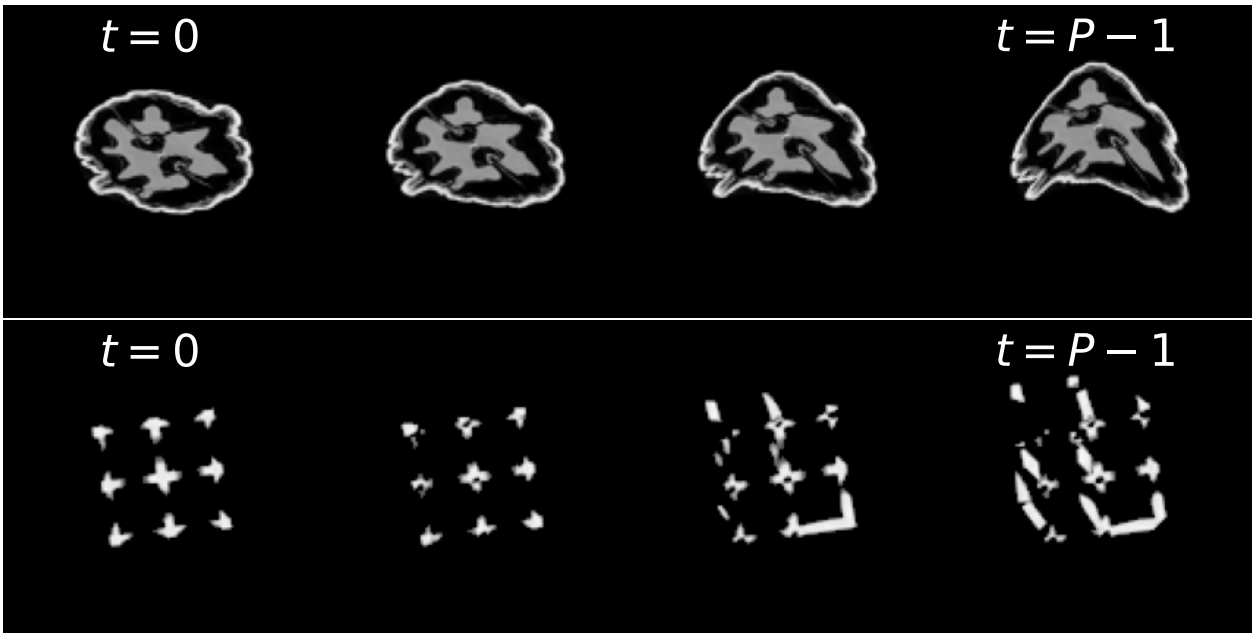


Figure 4.2: Ground-truth frames uniformly sampled in time for $P = 4$, for the time-varying walnut (top) and compressed object (bottom).

4.4.2 Comparison Benchmarks

PSM-TV: Similar to the proposed approach, this algorithm also uses a PSM to represent the object, but instead of the RED regularizer, the regularization penalizes the discrete 2D total variation of the temporal frames of f at each time instant. To this end, the constraint $f = \Lambda\Psi^T$ is implemented by substitution into the objective in (4.3), and the definition of ρ is changed to $\rho(\cdot) = \mathrm{TV}(\cdot)$, and the rest of the objective is kept the same. We consider spatial (PSM-TV-S) and spatiotemporal (PSM-TV-ST) alternatives of TV. The former computes TV only spatially in a single frame at each t, whereas the latter also computes differences between temporally adjacent pixels at $t-1$ and $t+1$. The unconstrained problem is then solved

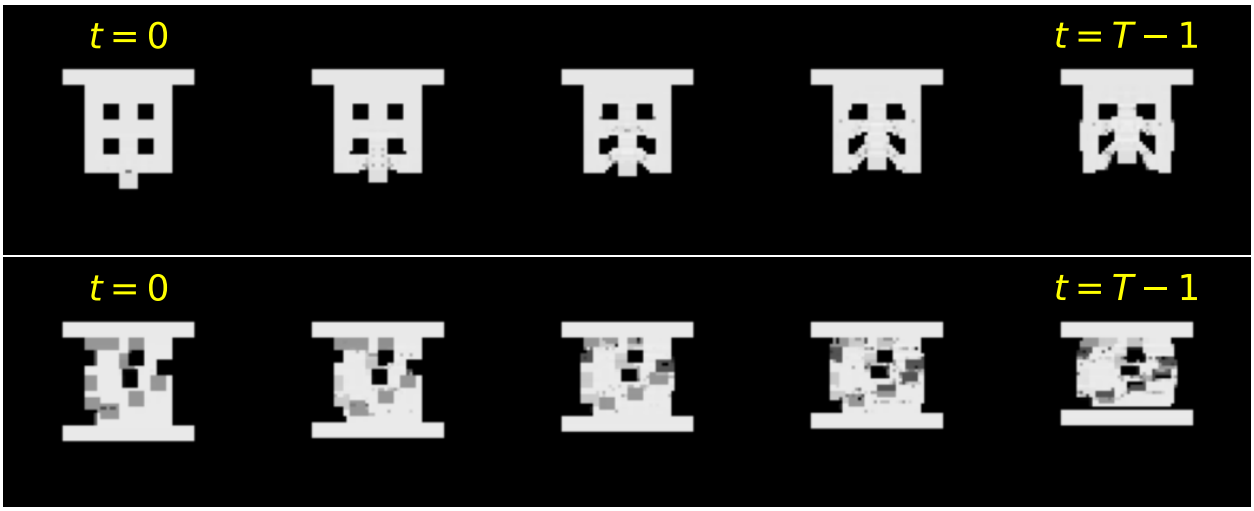


Figure 4.3: Ground-truth frames from the MPM-based deformation dataset uniformly sampled in time for $T = 5$, for Object #1 (top) and Object #2 (bottom).

for $\{\hat{\Lambda}, \hat{\Psi}\}$ (using Adam optimizer in Pytorch). Finally, the estimated object is obtained as $\hat{f} = \hat{\Lambda}\hat{\Psi}^T$.

TD-DIP [66]: TD-DIP is a recent method based on the Deep Image Prior (DIP) approach.[3] It uses a mapping network M_α and a generative model N_β in cascade to obtain the estimated object at each time instant f_t from fixed and handcrafted latent representations χ_t. Because TD-DIP was originally proposed for dynamic MRI, we modified the objective minimally for dynamic tomography as

$$\min_{\alpha,\beta} \sum_t \left\| g_t - R_{\theta(t)}((N_\beta \circ M_\alpha)(\chi_t)) \right\|^2. \tag{4.15}$$

For the comparisons in this work, the mapping network and generator architectures, latent representation dimensionality, optimizer, learning rate, and decay schemes are are identical to those in the available implementation [175]. The original work focuses on the beating heart problem and thus proposes a helix-shaped manifold for χ_t with number of cycles equal to the number of heartbeats during measurement acquisition. Since we do not have a repetition assumption for the motions of the walnut and compressed object, we use a linear manifold as explained in the original paper [66]. Thus, for clarity, in Section 4.4.4 the method is sometimes denoted as "TD-DIP (L)".

[3] It would be interesting to include yet another comparison benchmark (also developed for dMRI) – the DIP-based PSM approach [68]. However, as an implementation of this method was unavailable, and due to potential issues with replicating its performance and adapting to our CT problem (specific initialization and framework that use other MRI algorithms), we were unable to do so.

4.4.3 Experimental Settings

All methods are run on a workstation with an Intel(R) Xeon(R) Gold 5320 CPU and NVIDIA RTX A6000 GPU. In practice, we used a minor variation of Algorithm 2, where we combined the subproblems for Λ and Ψ, and minimized with respect to both basis functions simultaneously using gradient descent with Adam [126] optimizer.

Denoiser and training. Each convolutional layer in the denoiser network is followed by a ReLU nonlinearity except for the final single-channel output layer. We tested both direct and residual DnCNN denoisers, where the former predicts the denoised image and the latter estimates the noise from the input. We use the denoiser type that performs better for each object, but the differences are minor. Further architectural details for the denoisers in our experiments are in Table 4.8 in Section 4.6.1. We use three pre-trained denoisers, one for each of the three object types. In each case, the same pre-trained denoiser was used for all values of P.

The upper limit for noise level used in training the denoiser was set to $\sigma_{\max} = 5 \cdot 10^{-2}$. For the dynamic walnut object, the denoiser D_ϕ is trained on the central 200 axial, 200 sagittal, and 200 coronal slices of another static walnut CT reconstruction downsampled to size 128×128. For the compressed object, axial slices of pre-compression and post-compression *static* versions of the object, containing 462 slices in total, are used to train D_ϕ. For the cardiac MRI setting, the denoiser was trained on the *static* MRI training slices of the ACDC dataset [176]. For all datasets, D_ϕ is trained for 500 epochs using the Adam optimizer with a learning rate of $5 \cdot 10^{-3}$. As mentioned above, we evaluate both direct and residual DnCNN denoisers.

Temporal Basis. In compressed material and cardiac dMRI data experiments, we use the parametrization $\Psi = UZ$ with a fixed basis U that corresponds to a cubic spline interpolator, and for the warped walnut we use DCT-II, to interpolate the low-dimensional temporal representation Z to Ψ.

Initialization. Unless stated otherwise, the spatial and temporal basis functions are initialized using the SVD truncated to the rank of the dynamic object estimate produced by a recent projection-domain PSM-based method, "ProSep" [96]. If the ProSep estimate has rank smaller than K, the remaining basis functions are initialized as 0. Otherwise, all spatial basis functions are initialized as 0 and the temporal latent representations z_k are initialized randomly as $z_k \sim \mathcal{N}(0, I)$.

Tomographic Acquisition Scheme. All methods mentioned in this paper use the bit-reversed angular sampling scheme, over the range $[0, \pi]$. For time-sequential acquisition, the bit-reversed scheme was shown [95], [96] to provide favorable results via better conditioning of the forward model in comparison to alternatives. In a standard CT scanner, the speed of

rotation might have to be significantly increased to implement this scheme, possibly leading to greater motion blur. However, for several scenarios this would not be much of an issue. These include radial acquisition in MRI; a CT scanner with electronic beam deflection [177]; and settings where the acquisition time is dominated by the time to acquire each view rather than the rotation time, e.g., micro-CT, or imaging of dense materials.

To help address the challenge in implementing the bit-reversed scheme in other, more physically constrained settings, the number of distinct view angles $\hat{P}$ can be reduced and these views can be repeated periodically without a performance drop as also shown in Figure 4.14. With a reduced $\hat{P}$, the bit-reversed scheme can be implemented more conveniently, e.g., by multiple source-detector pairs, or by carbon nanotube sources [178], [179].

To apply the scheme to the multiple simultaneous projections per time frame considered for the MPM-based deformation dataset [172], we first consider the time-sequential bit-reversed scheme for P, and partition it into $T = 10$ subsets, each corresponding to a different time frame of the object. The number of simultaneous projections per time frame is 4 and 7 for $P = 32$ and $P = 64$, respectively. This number varies for the first and last time instants. The naive uniformly spaced progressive scheme is compared with the bit-reversed scheme ($\hat{P} = 64$) in Figure 4.4. The comparison also includes a periodic version of the bit-reversed scheme with $\hat{P}$ distinct view angles (the period) that are repeated $P/\hat{P}$ times.

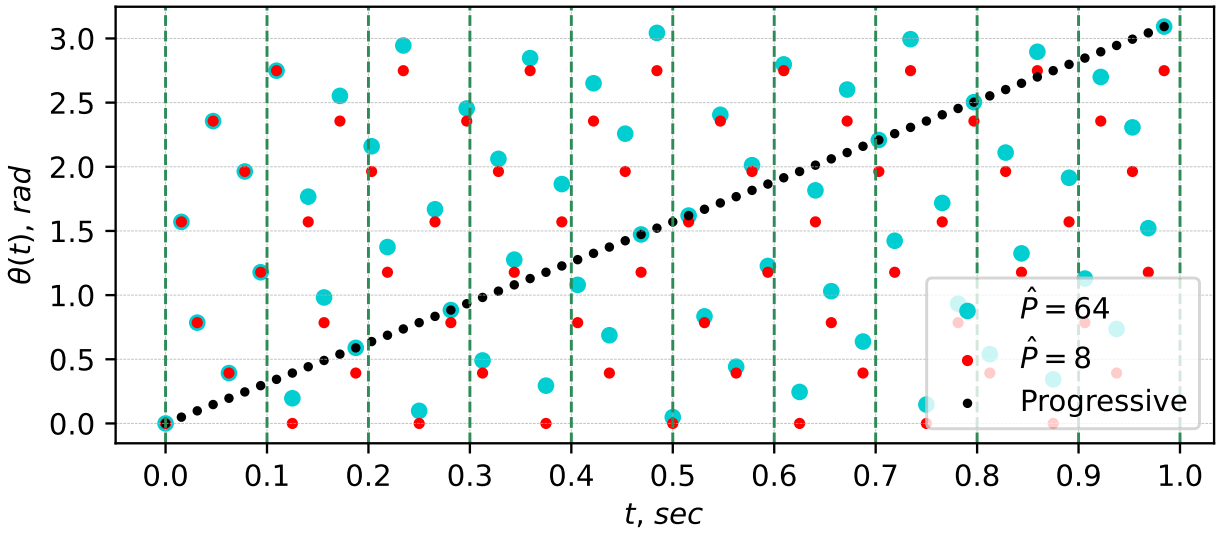


Figure 4.4: Progressive and bit-reversed acquisition schemes with different periods (the number of distinct view angles) $\hat{P}$ for $P = 64$. To facilitate comparison, for the three schemes sampling points are plotted for the *time-sequential* acquisition in $t \in [0, 1]$. Regions between dashed vertical lines show the $T = 10$ subsets of view angles used *simultaneously* for each time frame of the MPM-based deformation object in Figure 4.3.

Run Times. For $P = 256$ and using the specified computational resources and parameter settings, to reach the peak PSNR during optimization, RED-PSM with ProSep initialization requires $50 <$ iterations < 150 taking about 2 to 6 minutes, whereas TD-DIP with batch size P typically requires $>$ 30k steps, taking about 3.5 hours to complete. Hence, RED-PSM provides a speedup over TD-DIP by a factor of 35 to 105. Depending on the parameter configuration, the speedup factor may vary. However, the proposed method provides a significant run time reduction in all cases.

Evaluation Metrics. Four quantitative metrics were implemented for comparing different method performances: (i) the *peak signal-to-noise ratio* (PSNR) in dB; (ii) the *structural similarity index* (SSIM) [180]; (iii) the *mean absolute error* (MAE); and (iv) the *high-frequency error norm* (HFEN) [181]. The latter is defined as $\mathrm{HFEN}(f, f_r) = \|\mathrm{LoG}(f) - \mathrm{LoG}(f_r)\|_2$ where LoG is a rotationally symmetric Laplacian of Gaussian filter with a standard deviation of 1.5 pixels.

4.4.4 Results

Reconstruction accuracies for different P

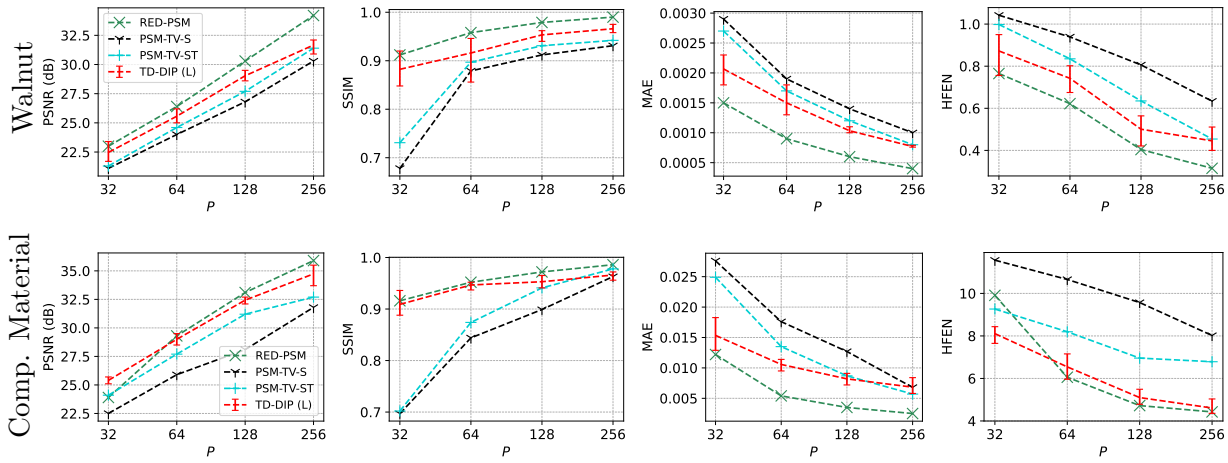


Figure 4.5: Reconstruction metrics for the time-varying walnut and compressed material vs. P using different methods. For TD-DIP, the metrics reported are assuming a "stopping oracle" that stops the iterations at the best PSNR reconstruction. The minimum and maximum values over three different runs with different random initial values are shown with bars, and the mean values are connected by dashed lines

RED-PSM is compared in Figure 4.5 and Table 4.2 with PSM-TV-S, PSM-TV-ST and TD-DIP (using, for both a-periodic objects, a linear latent representation). Parameter configurations for the experiments are provided in Table 4.6 in Section 4.6.1. While Figure 4.5 facilitates the assessment and comparison of trends in the metrics for different P for the

various methods, Table 4.2 provides the detail for a more precise quantitative comparison. In fact, since the plotted range of metrics such as PSNR in Figure 4.5 is very large due to varying P, Table 4.2 emphasizes important differences between methods for the same P.

Remarks:

1) The scale differences in the MAE and HFEN between the two objects are due to working with un-normalized densities.

2) In all TD-DIP experiments, optimization is stopped early to achieve the best PSNR reconstructions, assuming such a "stopping oracle" is available. In a more realistic setting, in the absence of this oracle, TD-DIP optimization with continued iterations suffers from overfitting, and produces in these experiments degraded results. For instance, for the warped walnut slice with 40k iterations, PSNR, SSIM, and MAE degrade significantly to 19.5 dB, 0.822, and $3.1 \cdot 10^{-3}$ for $P = 32$, and 24.1 dB, 0.824, and $2.1 \cdot 10^{-3}$ for $P = 64$. Thanks to the accurate spatial prior and the convergence properties, we do not encounter such a problem for RED-PSM.

3) Because the performance of TD-DIP varies with initialization, in each experiment we ran it three times, with different random initialization, each time using a "stopping oracle" to obtain the best PSNR for the given initialization. Table 4.2 reports the average of *the best PSNR reconstruction accuracies* for TD-DIP in these three runs. Figure 4.5, complements this information, by showing in addition to the average results, also the best and worst of these runs (still using the stopping oracle to get the best PSNR per run).

As expected, for all methods, the estimates improve with increasing P. In terms of PSNR, the proposed algorithm performs on par with or usually better than TD-DIP (for all but the lowest P on the compressed material object), and consistently better than PSM-TV-S and PSM-TV-ST. The PSNR improvement of RED-PSM over TD-DIP enhances with increasing P, reaching 2.4 dB for the time-varying walnut with $P = 256$. Moreover, for the other three metrics, SSIM, MAE, and HFEN, the improvement of RED-PSM over other algorithms is more significant. Specifically, the reduction in MAE reaches up to and exceeds %50 for both objects for $P = 256$. These observations are valid for both objects, however RED-PSM provides slightly greater improvement over TD-DIP in the walnut case.

We would like to emphasize the significance of these reconstruction quality improvements by comparing them to some representative examples. TD-DIP reports up to 3 dB PSNR and 0.005 SSIM improvement with respect to an *older* dMRI method in a *single* scenario. While providing comparable PSNR improvements (2.4 dB), we are able to provide four times larger SSIM (0.02) improvements *relative to TD-DIP itself*. Other recent dMRI [67] and dynamic photo-acoustic tomography [106] methods improve 1.5 dB, and at most 1 dB and often no improvement over their respective benchmarks, again in single imaging scenario for each

method.

Also, based on the run-time improvement (∼3x) of [67] with respect to TD-DIP, RED-PSM is still much (∼10x-30x) faster, than [67], and more interpretable.

Table 4.2: Reconstruction accuracies for for different P. Random (R) and ProSep (Pr) initialization for the PSM methods. For TD-DIP, the reported accuracies are for the best PSNR using a "stopping oracle", averaged over three runs with random initial conditions.

		(a) Walnut				(b) Compressed Material			
P	Method	PSNR (dB)	SSIM	MAE (1e-3)	HFEN	PSNR (dB)	SSIM	MAE (1e-2)	HFEN
32	PSM-TV-S (R)	21.1	0.678	2.9	1.04	22.5	0.697	2.9	11.6
	PSM-TV-ST (R)	21.3	0.731	2.6	0.99	24.1	0.702	2.5	9.3
	TD-DIP (L)	22.5	0.882	2.1	0.87	**25.4**	0.909	1.5	**8.1**
	RED-PSM (Pr)	**22.8**	**0.911**	**1.5**	**0.78**	23.9	**0.916**	**1.2**	9.9
64	PSM-TV-S (R)	24.0	0.879	1.9	0.94	25.6	0.845	1.8	10.7
	PSM-TV-ST (R)	24.6	0.897	1.7	0.83	27.7	0.874	1.3	8.2
	TD-DIP (L)	25.6	0.916	1.5	0.74	29.0	0.947	1.1	6.5
	RED-PSM (Pr)	**26.4**	**0.958**	**0.9**	**0.57**	**29.3**	**0.952**	**0.5**	**6.0**
128	PSM-TV-S (R)	26.8	0.912	1.4	0.78	29.2	0.907	1.1	9.6
	PSM-TV-ST (R)	27.7	0.931	1.2	0.63	31.2	0.942	0.9	7.0
	TD-DIP (L)	29.0	0.953	1.0	0.50	32.4	0.953	0.8	5.1
	RED-PSM (Pr)	**30.3**	**0.979**	**0.6**	**0.40**	**33.1**	**0.972**	**0.4**	**4.7**
256	PSM-TV-S (R)	30.1	0.934	1.0	0.63	31.9	0.963	0.7	8.0
	PSM-TV-ST (R)	31.4	0.942	0.8	0.45	32.7	0.978	0.6	6.8
	TD-DIP (L)	31.7	0.966	0.8	0.45	34.7	0.966	0.7	4.6
	RED-PSM (Pr)	**34.2**	**0.989**	**0.4**	**0.32**	**35.9**	**0.986**	**0.3**	**4.4**

Figure 4.6 compares the reconstructions for both objects at two different values of t for $P = 256$. As expected, PSM-TV-S performs the worst among the compared methods and provides, for both objects, blurry reconstructions lacking finer details. The TD-DIP reconstructions improve somewhat over PSM-TV-S and PSM-TV-ST, but contain visible noise-like artifacts on the piecewise constant regions of the walnut object, which are alleviated by RED-PSM. This is manifested also in the absolute difference figures, with error for TD-DIP distributed throughout the interior regions of the walnut. Also, around the shell of the walnut, RED-PSM is further able to preserve sharper details. For the compressed material, in comparison to TD-DIP, RED-PSM shows reduced absolute error almost uniformly over the object. This difference is more prominent around the highly dynamic regions of the object, emphasizing the advantage of the proposed method. For a better understanding of

RED-PSM results, we also display the reconstructed spatial and temporal basis Λ and Ψ for the time-varying walnut scenario in Section 4.6.3.

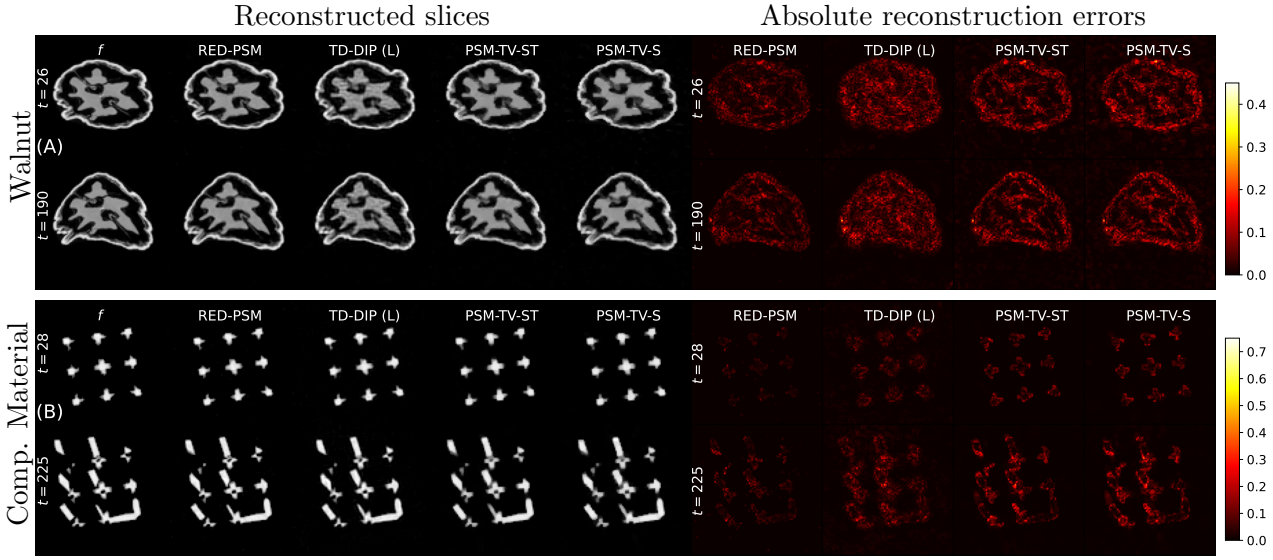


Figure 4.6: Comparison of reconstructed object frames at two time instants using different methods for P=256, and the corresponding normalized absolute reconstruction errors for (A) the time-varying walnut, and (B) compressed object.

Reconstructed x-t "slices" through the dynamic walnut are compared in Figure 4.7. The location of the x-t slice is highlighted on the $t = 0$ static x-y frame by a yellow line. Consistent with the comparison in Figure 4.6, RED-PSM provides reduced absolute error values throughout the respective x-t slice. Also, as more apparent on the error figures, TD-DIP leads to higher background errors.

Finally, a zoomed-in comparison of the time-varying walnut object for another time instant for $P = 256$ is provided in Figure 4.8. The comparison shows the better performance of RED-PSM at recovering the finer details clearly.

PSNR vs. t Comparisons

To complement the cumulative metrics in Figure 4.5 and Table 4.2 and the "snapshot" qualitative comparisons in Figure 4.6, we study how the reconstructed image frame PSNRs vary over the reconstructed time interval. The per-frame PSNRs (in dB) of the walnut and compressed object reconstructed with $P = 256$ by the different methods are shown in Figure 4.9 as a function of t. For TD-DIP, the best PSNR obtained using a "stopping oracle" is reported, with the red shading indicating, for each t, the interval between the highest and lowest PSNR in three runs with different random initial conditions. For the warped walnut object, RED-PSM provides consistently better PSNR than the best-case TD-DIP for all t.

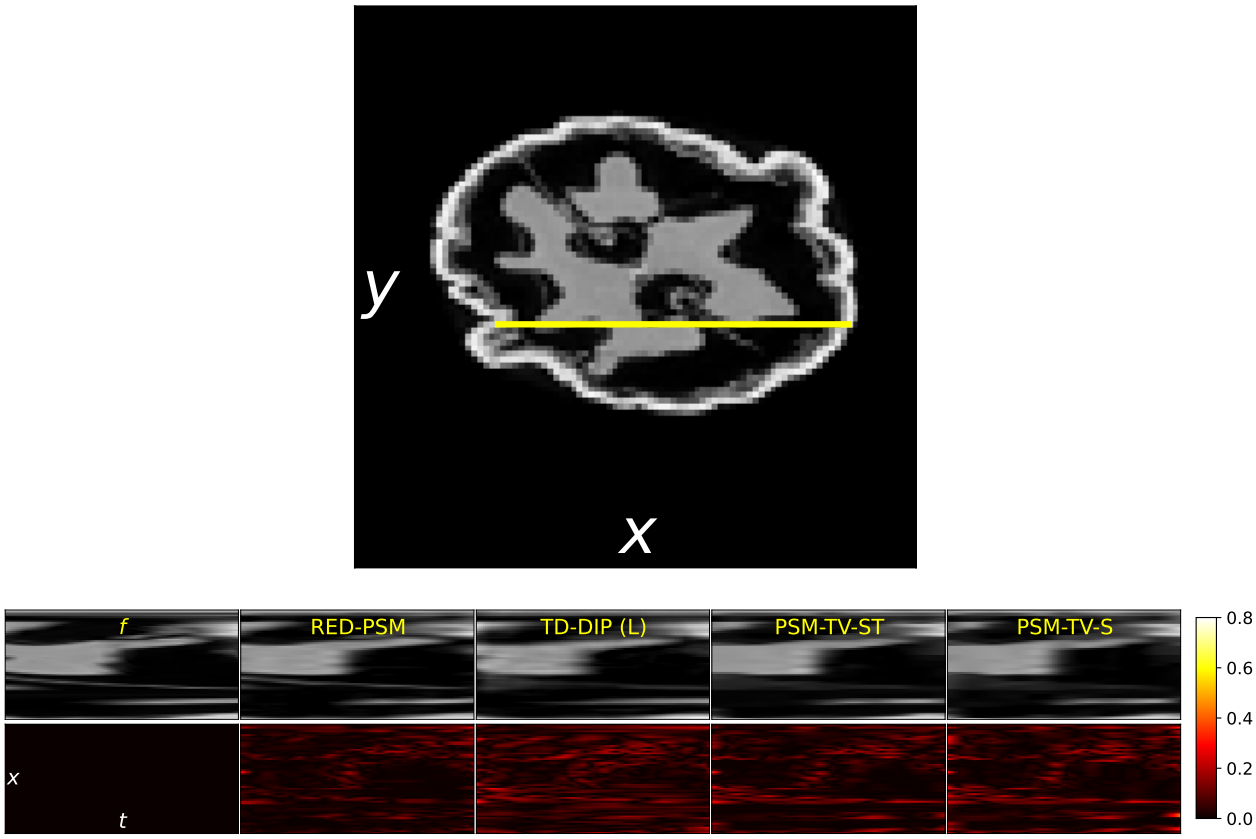


Figure 4.7: Comparison of reconstructed x-t slices (top) and corresponding normalized absolute error (bottom) using different methods for P=256. The cross-section location is indicated on the static $t = 0$ object with a yellow line. The x-y-t coordinates are indicated in white text on the static object and bottom left absolute error figure.

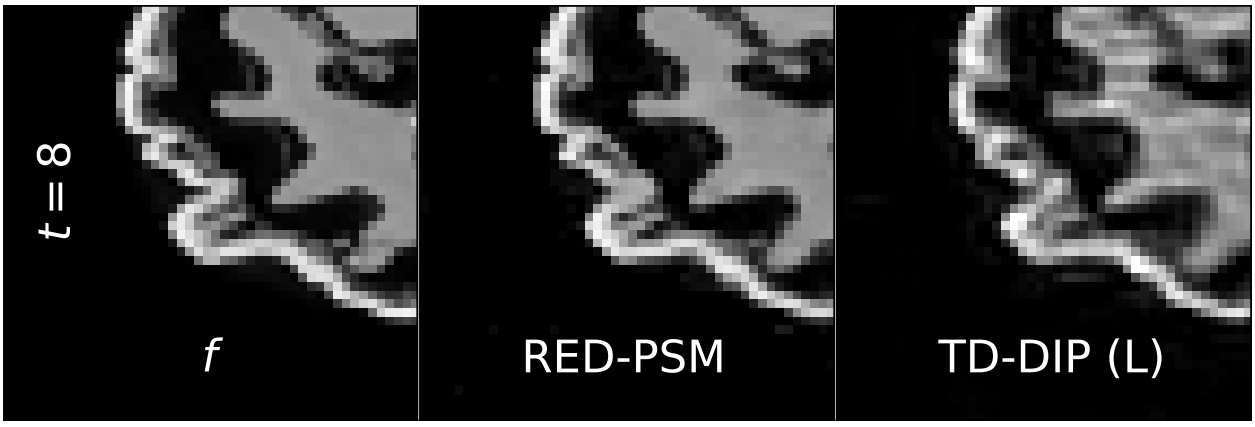


Figure 4.8: Zoomed-in walnut reconstruction comparison for P=256.

For the compressed object, the same is true at about 70% of t points. Figure 4.9 also shows transient effects at the beginning and the end for both objects and all methods. In scenarios such as the object compression experiment, in which the initial and final state are static and could be measured using multiple projections, such transients could be eliminated. Similarly, in quasi-periodic scenarios such as cardiac imaging, the effect of such transients would be minimal.

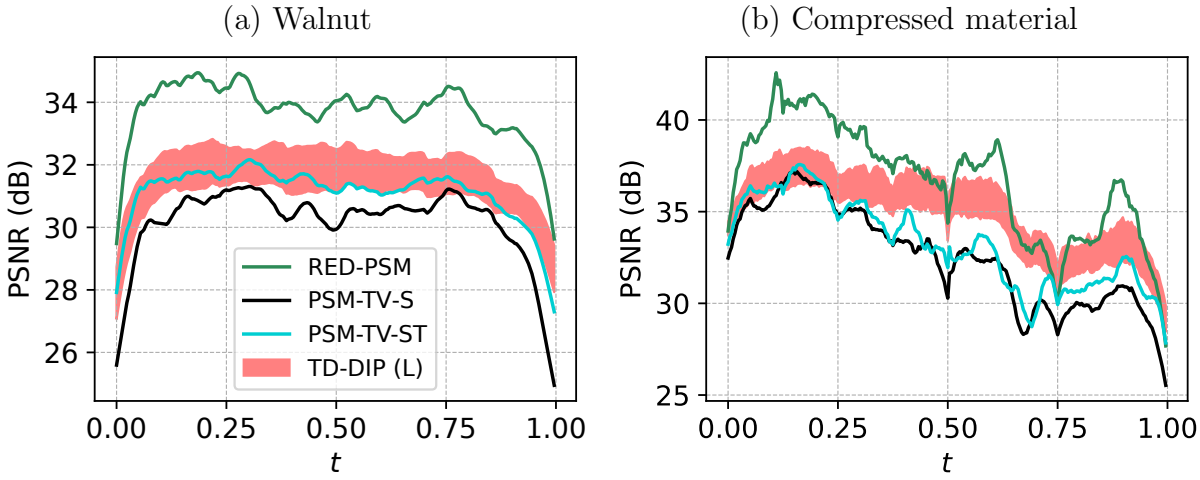


Figure 4.9: Reconstruction PSNR vs. t for the (a) time-varying walnut, and (b) compressed material for P=256. The red shading for TD-DIP indicates, for each t, the interval between the best and worst PSNR with the early stopping oracle explained in Section 4.4.4 in three runs with different random initial conditions.

Effect of Initialization

The initialization of Λ, Ψ, and f plays an important role in the performance and convergence speed of RED-PSM. We observe significant speed-up when rather than a random initialization, we initialize the algorithm with ProSep [95] estimated reconstruction. Figure 4.10 shows PSNR vs. iterations comparison for different initialization techniques for the dynamic walnut object with $P = 256$. The rest of the parameters were selected identical to those indicated in Section 4.6.2 Table 4.6. This experiment highlights the advantages of initializing with ProSep estimated basis functions: eliminating the need for multiple runs for a best-case result, and speeding up convergence considerably.

Combined also with the convergent algorithm eliminating the need for an unrealistic stopping oracle and the theoretical analysis, RED-PSM provides improved reliability which is of practical significance.

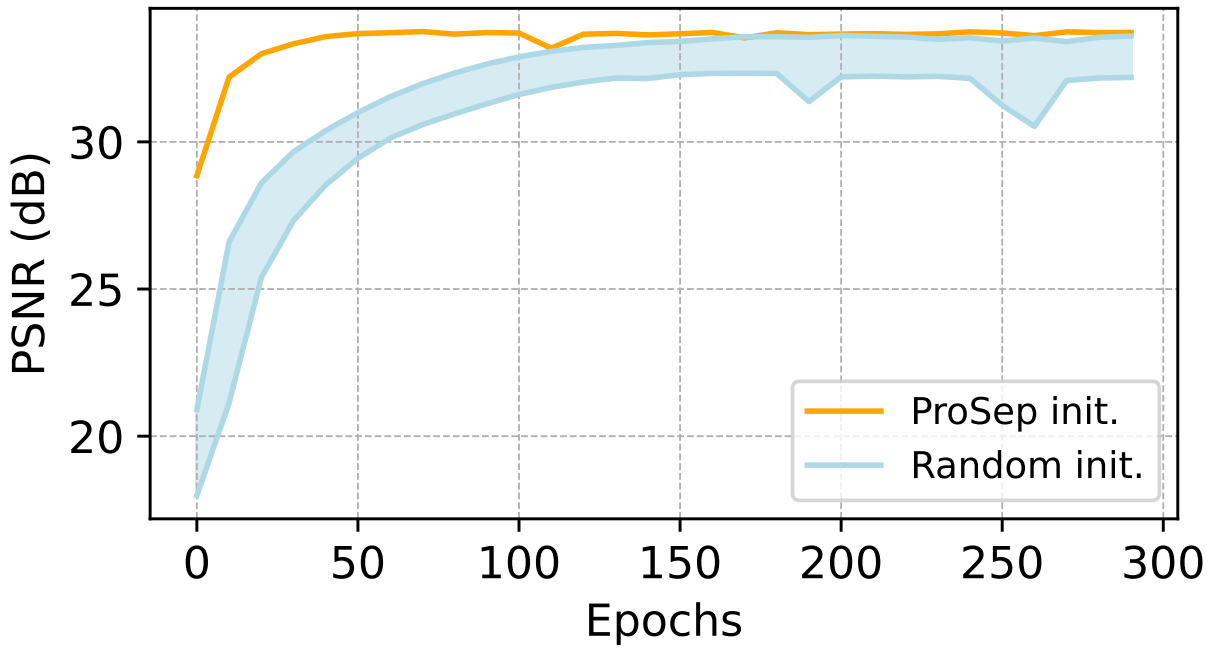


Figure 4.10: Advantage of ProSep-based vs. random initialization of RED-PSM (see Section 4.4.3). For random initialization, the area in blue between the best and the worst PSNR for each iteration highlights the varying performances of five different runs with different random initializations.

Cardiac dMRI data experiments

In this setting, different to time-sequential acquisition experiments, we used 4 k-space radial lines ("spokes") per frame at the bit-reversed angles. We used 1.4 and 2.8 cardiac cycles, with $23 \times 4 = 92$ spokes/cycle, for a total of P=128 and P=256 spokes. The problem is still severely undersampled compared to the experiment in [66] where 13 spokes are used per frame for 13 cycles, for a total of $13 \times 23 \times 13 = 3,887$ spokes.

Since the data is periodic, we also tested the helix latent scheme (H) for TD-DIP.

The metrics in Table 4.3 and the qualitative comparison in Figure 4.11 with zoomed-in reconstructions and absolute error maps, show that RED-PSM performs better than both versions of TD-DIP.

Table 4.3: Reconstruction accuracies for RED-PSM and TD-DIP for the retrospective dMRI data [66].

P	Method	PSNR (dB)	SSIM	MAE (1e-2)	HFEN
128	TD-DIP (L)	34.6	0.923	1.7	3.38
	TD-DIP (H)	34.6	0.928	1.7	3.32
	RED-PSM	**36.6**	**0.939**	**1.4**	**2.73**
256	TD-DIP (L)	36.2	0.948	1.4	3.47
	TD-DIP (H)	36.4	0.947	1.4	3.44
	RED-PSM	**38.4**	**0.962**	**1.1**	**2.99**

In [67], the authors compare their method to an *older* version of TD-DIP without the improved latent representation prior scheme for cardiac dMRI and report 1.5 dB PSNR improvement. In RED-PSM, we exceed this improvement on the latest version of TD-DIP in multiple scenarios.

MPM-based deformation data experiments

The proposed method is compared to TD-DIP [66] and to the 2D total-variation regularized version of the hard-constraint PSM objective, denoted as PSM-TV. The weights λ and β were typically on the order of 10^{-3} to 10^{-2}. Table 4.4 shows the reconstruction accuracies for three different metrics: (i) the peak signal-to-noise ratio (PSNR) in dB, (ii) structural similarity index (SSIM), and (iii) the mean absolute error (MAE). In these experiments, projections are degraded by AWGN with standard deviation 10^{-2}, and the order of the PSM is chosen as $3 \leq K \leq 6$ for PSM-TV, and $6 \leq K \leq 8$ for RED-PSM.

It is seen that Object #2 is more challenging than Object #1 for all compared methods, perhaps due to its greater deformation. However, for both objects and all compared methods,

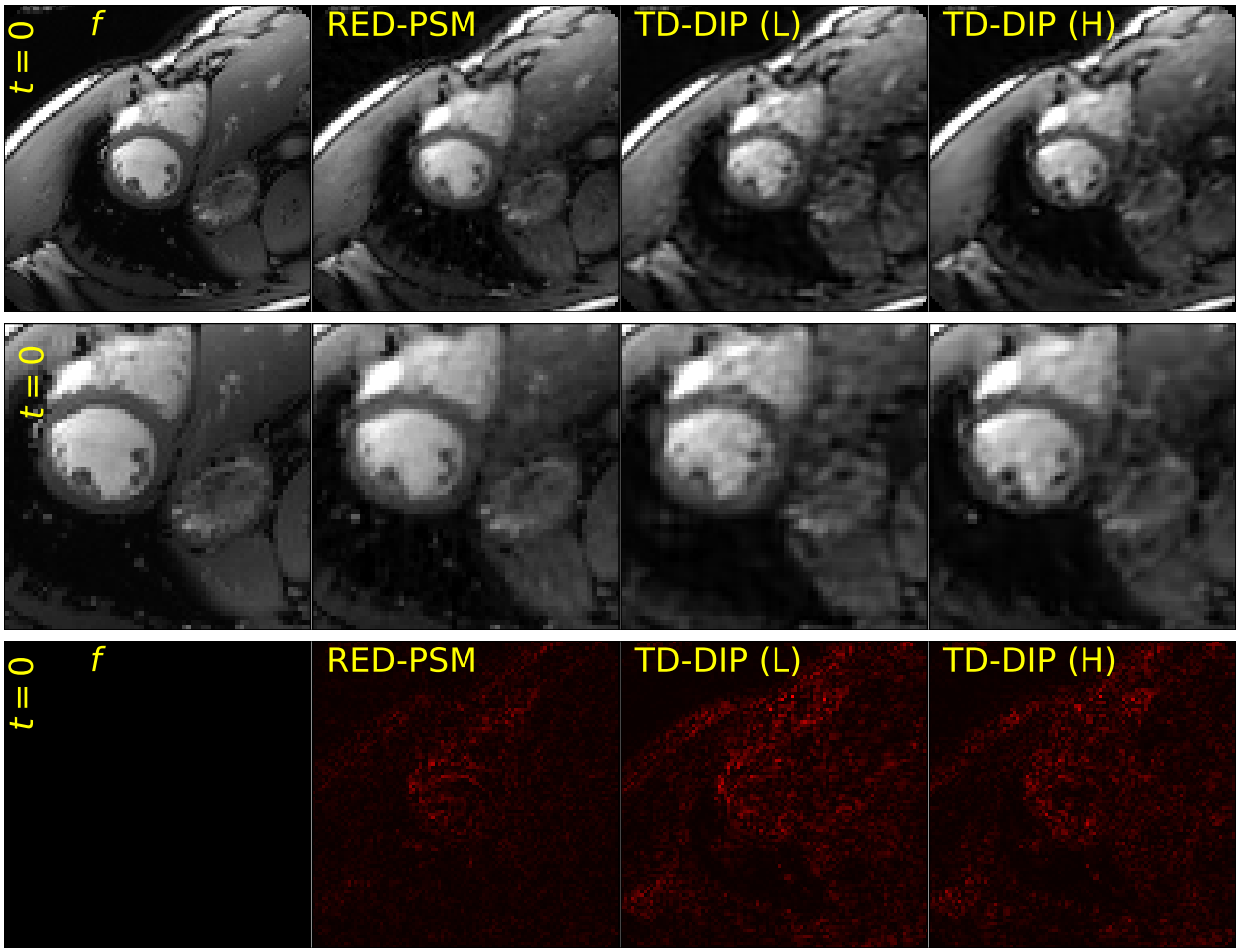


Figure 4.11: Reconstructed frames for P=256 for retrospective dMRI data [66] with zoomed-in frames (middle row), and absolute reconstruction errors (last row).

Table 4.4: Reconstruction accuracies for different P where the number distinct view angles $\hat{P} = P$ for MPM-based deformation objects. For TD-DIP, the reported accuracies are for the best PSNR using a "stopping oracle", averaged over three runs with random initial conditions.

		(a) Object #1			(b) Object #2		
P	Method	PSNR (dB)	SSIM	MAE (1e-3)	PSNR (dB)	SSIM	MAE (1e-3)
32	PSM-TV	23.6	0.836	3.7	21.0	0.726	5.3
	TD-DIP	25.0	0.853	2.8	21.2	0.822	4.4
	RED-PSM	**27.6**	**0.928**	**1.1**	**22.3**	**0.858**	**3.0**
64	PSM-TV	29.2	0.868	1.7	23.5	0.806	3.6
	TD-DIP	28.6	0.898	2.0	23.0	0.882	3.7
	RED-PSM	**30.4**	**0.950**	**0.8**	**24.4**	**0.914**	**2.0**

accuracies improve consistently with an increasing total number of projections P. RED-PSM provides improved reconstruction accuracies for each setting for each considered metric. While PSNR improvements are more significant for the first object, SSIM and MAE enhance more uniformly for both objects.

All our reported TD-DIP experiments use a "stopping oracle" for early termination of the model optimization to obtain the best PSNR reconstruction. Due to overfitting, without this oracle, the performance of TD-DIP degrades significantly. For instance, for Object #1 and $P = 64$, the PSNR, SSIM, and MAE degrade to 24.0 dB, 0.685, and $4.9 \cdot 10^{-3}$ for 40k iterations, and further to 20.0 dB, 0.633, and $8.1 \cdot 10^{-3}$ for 80k iterations. Even greater degradation is observed for $P = 32$.

The reconstructions and absolute error maps of the two objects at particular times are compared across the different methods in Figure 4.12. For both objects RED-PSM suppresses the reconstruction errors better than the other methods. In particular, the TD-DIP reconstructions contain horizontal and vertical line artifacts. As expected, PSM-TV smoothes finer details excessively, resulting in reduced resolution.

Since obtaining multiple projections from different angles simultaneously (or in a time sufficiently short that the object can be assumed static) can be physically challenging, we also test the performance of RED-PSM with an acquisition scheme that may be easier to realize physically: keeping the same total number $P = 64$ of projections, these are taken at a smaller number $\hat{P}$ (called "period") of *distinct* view angles. In this ablation study, all other hyperparameters of RED-PSM are kept constant and were not fine-tuned for each period $\hat{P}$. Figure 4.13 shows that both the reconstruction PSNR and MAE experience little change down to $\hat{P} = 8$, but degrade at $\hat{P} = 4$. On the other hand, the SSIM fluctuates around a mean value for the different $\hat{P}$. These results suggest that RED-PSM is robust to the decrease (down to a low threshold) of distinct view angles in the acquisition scheme, and thus it can be used in more constrained settings without a performance drop.

Acquisition with smaller number of distinct view angles

Since obtaining time-sequential projections from different angles in a sufficiently short time period can be physically challenging, we also test the performance of RED-PSM with an acquisition scheme that may be easier to realize physically: keeping the same total number of P projections, but taken at a smaller number $\hat{P}$ (called "period") of *distinct* view angles, which are also obtained using the bit-reversed angular sampling scheme. The comparison in Figure 4.14 shows that it is possible to use up to 1/8-th of the distinct view angles without performance loss for RED-PSM.

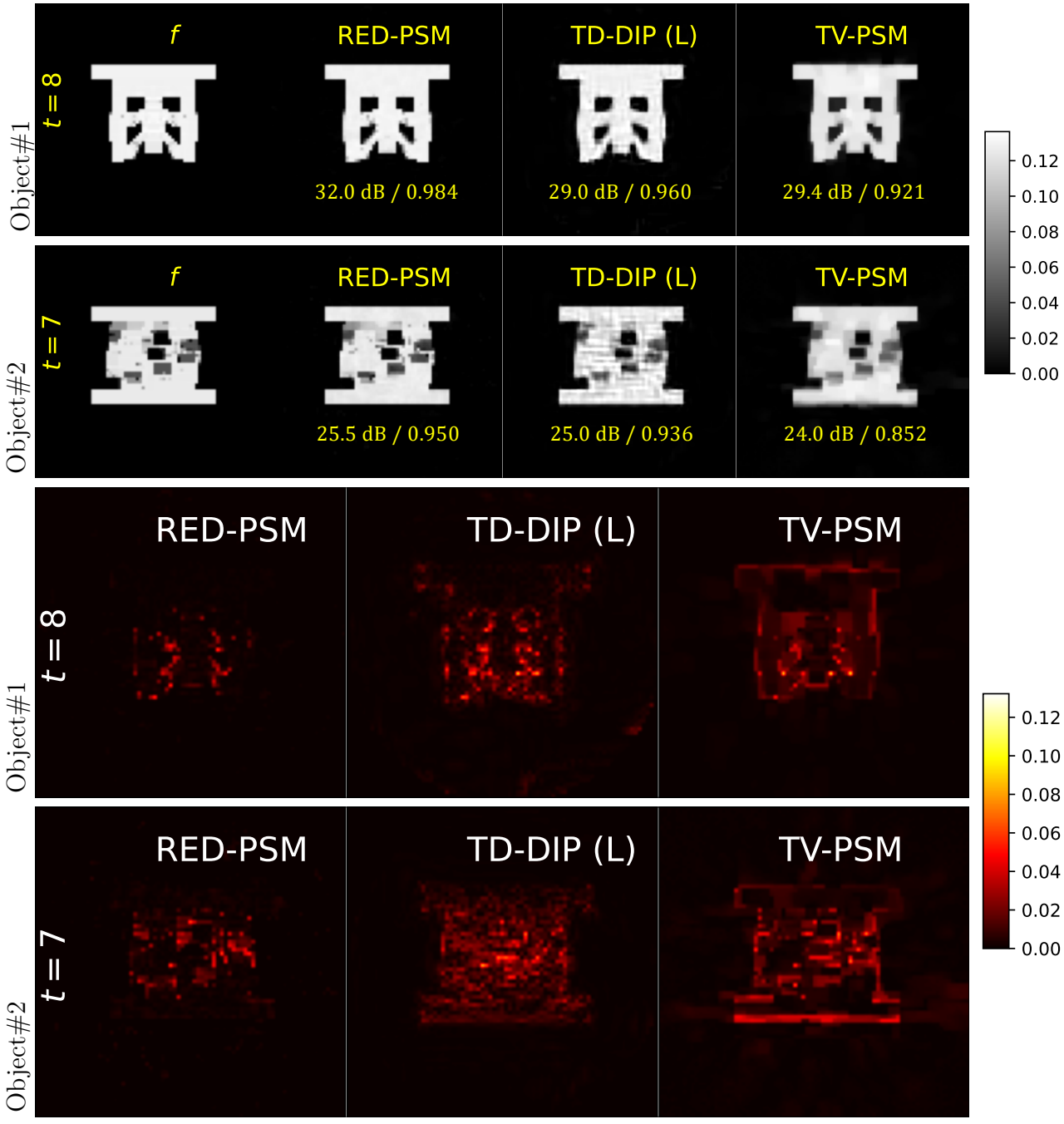


Figure 4.12: Reconstructed frames for the MPM-based deformation Object 1 (top) and Object 2 (bottom) at different time instants for $P = \hat{P} = 64$ (left) with projection AWGN standard deviation of $5 \cdot 10^{-3}$, and the corresponding absolute reconstruction errors (right) using different methods. The TD-DIP reconstructions are for the experiment with the median PSNR value out of three runs, and use a "stopping oracle" (see Section 4.4.4). PSNR (in dB) and SSIM values for the *complete* spatio-temporal objects are reported under each reconstructed frame.

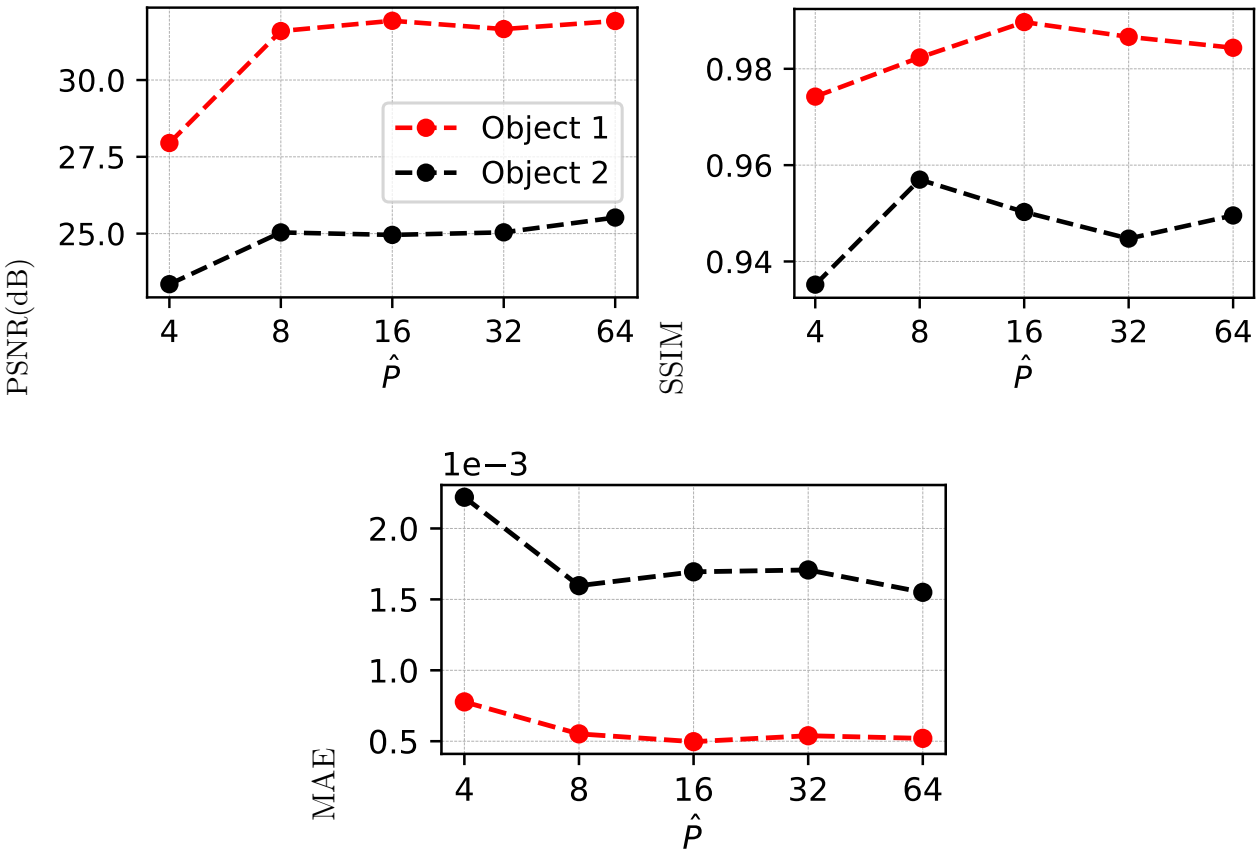


Figure 4.13: Reconstruction metrics for the two MPM-based deformation objects vs. number of distinct view angles $\hat{P}$ using RED-PSM with $P = 64$ and projection AWGN standard deviation of $5 \cdot 10^{-3}$.

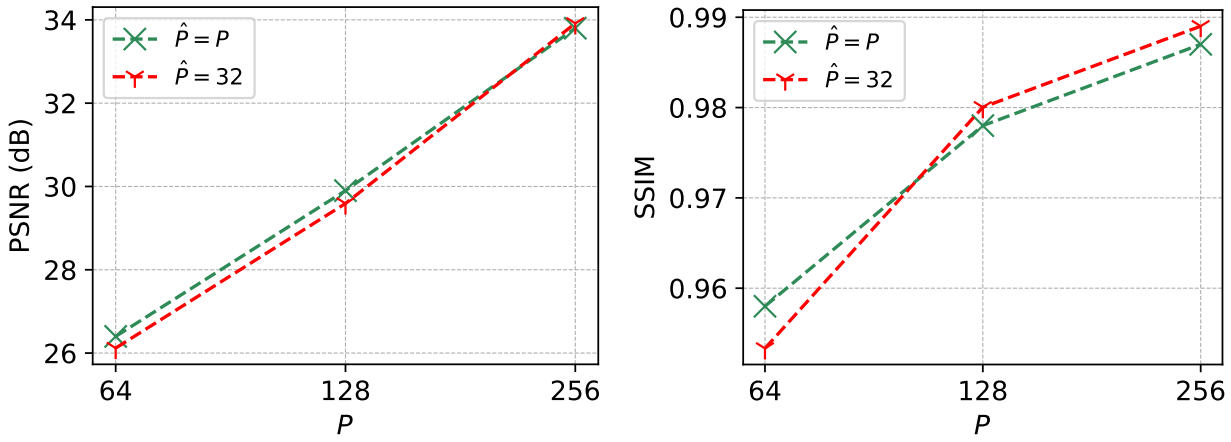


Figure 4.14: Reconstruction PSNR and SSIM for the time-varying walnut vs. the number of distinct view angles $\hat{P}$ using RED-PSM with different total number of views P.

Patch-based RED denoiser

To improve the scalability of the method to higher resolution and/or 3D dynamic object, conveniently, the objective in (4.3) can be manipulated to operate on the patches of temporal image frames of the time-varying object. This circumvents the need to store the complete image frame at a given time, and also enables the denoiser D_ϕ to be both trained and to operate on patches of image frames, having the potential to incorporate well-known as well as recent techniques [182]–[185]. To showcase the potential of the suggested scheme, we replace the full-size D_ϕ with a patch-based counterpart in the RED step and compare the performance with the originally proposed method for 2D dynamic objects.

The patch-based denoiser for RED updates is trained using

$$\min_{\phi} \sum_i \sum_l \|B_l f_i - D_\phi(B_l \tilde{f}_i)\|_F^2 \quad \text{s.t.} \quad \tilde{f}_i = f_i + \eta_i,\ \forall i,$$

where η_i, σ_i are set as in (4.7), and B_l, with $l \in \{0, \ldots, L-1\}$, is the operator to extract the l-th patch of the image. The denoiser D_ϕ operates separately on each patch.

To train the patch-based D_ϕ, uniformly random rotations (multiples of $\frac{\pi}{2}$), and random horizontal/vertical flips, each with $\frac{1}{2}$ probability, were used for data augmentation. The patch size was chosen 8×8 with a stride of 2.

Table 4.5 compares the results for the two denoiser types for both objects, using the same denoiser training policy as in Section 4.4.3, and experimental configurations of Table 4.7 in Section 4.6.1. The results show little difference between the patch-based denoiser and full frame-based denoiser RED-PSM variants, thus verifying the effectiveness of the patch-based version. The analysis of the computational requirements of the patch-based RED-PSM variant in Section 4.2.4 shows its potential for a highly scalable implementation.

We note that in a divergent beam scenario, the contribution of a patch to a projection will be position-dependent and this would require accurate bookkeeping. Details for doing so can be found in tile-based methods for fan-beam [186] and cone-beam [187] tomography.

4.5 Conclusions

We proposed RED-PSM, the first PSM-based approach to dynamic imaging using a pre-trained and learned (RED-based) spatial prior. The objective in the proposed variational formulation is optimized using a novel and effective bi-convex ADMM algorithm, which enforces the PSM as a hard constraint. Unlike existing PSM-based techniques, RED-PSM is supported by theoretical analysis, with a convergence guarantee to a stationary point of the objective. The

Table 4.5: Performance comparison for different denoiser types for RED-PSM. Patch-based denoisers also use DnCNN and have the same configuration and training policy as the full image denoiser.

Object	P	Denoiser	PSNR(dB)	SSIM	MAE(1e-3)	HFEN
Walnut	256	Full-image	33.8	0.987	0.4	0.32
		Patch-based	33.7	0.989	0.4	0.35
Material	256	Full-image	35.9	0.986	2.5	4.43
		Patch-based	35.6	0.981	2.7	4.56

results of the numerical experiments show better reconstruction accuracy and considerably faster run times compared to a recent DIP-based algorithm. A patch-based regularizer version of RED-PSM provides almost equivalent performance with massive reduction of storage requirements, indicating the potential of our framework for dynamic high-resolution 2D or 3D settings.

Possible directions for future work include the application of RED-PSM to different imaging scenarios other than tomography and MRI, and robust denoiser training for RED framework since the deep denoisers encounter varying artifact distributions during optimization. This could also improve the generalizability of the framework to different input types.

4.6 Proofs and Supplementary Material

4.6.1 Experimental configurations

The PSM-TV and PSM-RED parameter selections for the experiments listed in Table 4.2 are provided in Table 4.6. Likewise, Table 4.7 shows the parameter configurations for the denoiser type comparison experiments for RED in Table 4.5. Finally, the architectural information for the DnCNN denoisers used throughout this work is in Table 4.8.

4.6.2 Time and space complexity analyses

In this section, we analyze the operation count and storage requirements for the proposed algorithm and its patch-based version.

In this analysis, M_i represents the number of inner iterations for iteratively solved subproblems, assumed common to the different subproblems, and we assume $K \ll N$. Also, $R_{\theta(t)}$ and $R_{\theta(t)}^T$ are implemented as operators, each requiring $O(N^2)$ when applied to a $N \times N$ image or a projection of size N.

Table 4.6: The parameter selections for the reconstructions in Table 4.2. The latent penalty weight was selected as ξ=10^{-1} for the walnut, and ξ=10^{-3} for the compressed object experiments.

		(a) Walnut					(b) Comp. Material				
P	Method	K	d	λ	$\tilde{\lambda}$	β	K	d	λ	$\tilde{\lambda}$	β
32	PSM-TV-S (R)	3	4	5e-2	-	-	3	4	10	-	-
32	PSM-TV-ST (R)	4	5	5e-2	5e-2	-	3	4	1e1	1e2	-
32	PSM-RED (P)	3	7	1e-4	-	1e-4	3	9	5.12e-2	-	1.6e-2
64	PSM-TV-S (R)	3	4	5e-2	-	-	5	6	10	-	-
64	PSM-TV-ST (R)	4	5	5e-2	5e-2	-	5	6	1e1	1e2	-
64	PSM-RED (P)	4	7	1e-4	-	5e-4	5	9	32e-4	-	4e-3
128	PSM-TV-S (R)	5	7	5e-2	-	-	8	9	5	-	-
128	PSM-TV-ST (R)	6	7	5e-2	1e-1	-	8	9	5	5e1	-
128	PSM-RED (P)	6	9	2e-4	-	2e-4	8	11	4e-4	-	2e-3
256	PSM-TV-S (R)	10	11	5e-2	-	-	11	13	10	-	-
256	PSM-TV-ST (R)	10	11	5e-2	5e-2	-	11	13	1e1	1e2	-
256	PSM-RED (P)	10	11	5e-5	-	1e-4	11	13	1e-4	-	1e-3

Table 4.7: Parameter configurations for the denoiser type study experiments in Table 4.5.

		(a) Walnut				(b) Comp. Material			
P	Denoiser	K	d	λ	β	K	d	λ	β
256	Full-image	10	11	5e-5	1e-4	11	13	1e-4	1e-3
256	Patch-based	10	11	5e-5	1e-4	11	13	1e-4	1e-3

Table 4.8: Denoiser DnCNN configurations for different datasets.

Dataset	# of layers	# of channels	Denoising
Walnut	6	64	Direct
Compressed material	3	32	Residual
Cardiac dMRI	6	64	Residual

Storage requirements for the input quantities

When calculations are performed sequentially in t as in the following sections, the storage requirements for the primal and dual variables are calculated for f_t, γe_t, $\Psi^T e_t$ and Λ. The storage of the spatial basis functions Λ dominates the overall storage cost at $O(KN^2)$. The space complexity analyses of different subproblems in the following sections consider terms other than these input quantities.

RED-PSM

Expanding the Λ and Ψ subproblems in Algorithm 2 in t, we have

$$\min_{\Lambda} \sum_t \left(\|R_{\theta(t)} \Lambda \Psi^T e_t - g_t\|_2^2 + \beta \|(\Lambda \Psi^T - f + \gamma) e_t\|_2^2 \right) + \xi \|\Lambda\|_F^2, \tag{4.16}$$

and

$$\min_{\Psi} \sum_t \|R_{\theta(t)} \Lambda \Psi^T e_t - g_t\|_2^2 + \xi \|\Psi^T e_t\|_2^2 + \beta \|(\Lambda \Psi^T - f + \gamma) e_t\|_2^2. \tag{4.17}$$

We first analyze the operation counts for gradients with respect to Λ and $\Psi^T e_t$, respectively.

We note that in these gradient computations, the term $R_{\theta_t}^T g_t$ is pre-computed and stored before the start of the algorithm, and $(\gamma - f)e_t$ and $(R_{\theta(t)}^T g_t + (\gamma - f)e_t)$ are computed and stored at each bilinear ADMM outer iteration so that they do not contribute to the respective costs in each inner iteration, i.e, their cost does not scale with M_i.

Starting with the gradient of (4.17) with respect to $\Psi^T e_t$, we have

$$2((R_{\theta(t)}\Lambda)^T (R_{\theta(t)}\Lambda) + \Lambda^T \Lambda + \xi I)\Psi^T e_t - 2(R_{\theta(t)}\Lambda)^T g_t + 2\Lambda^T (\gamma - f) e_t. \tag{4.18}$$

The overall operation count for this term is $O(KN^2P)$ due to the computation of $R_{\theta(t)}\Lambda^T$ and $\Lambda^T(\gamma - f)e_t$ at each t. These terms do not need to be computed for each inner iteration.

When $\Psi = UZ$ is used, the gradient for Z requires the multiplication of (4.18) with $U^T e_t$, adding a minor operation count $O(dK)$.

The most efficient implementation in terms of space complexity is the sequential computation of the gradients for each t without increasing the operation count. It is also possible to parallelize the operations along t with the trade-off between the run time and the storage requirements.

Considering sequential computation for each t, the additional storage requirement is $O(NK)$ due to $R_{\theta(t)}\Lambda$.

Next, we consider the gradient of (4.16) with respect to Λ

$$2\xi\Lambda + \sum_t 2(R_{\theta(t)}^T R_{\theta(t)} + I)\Lambda(\Psi^T e_t)(e_t^T \Psi) + (R_{\theta(t)}^T g_t + (\gamma - f)e_t)(e_t^T \Psi). \tag{4.19}$$

The operation count for this gradient term is $O(KN^2PM_i)$ due to $R_{\theta(t)}^T R_{\theta(t)}\Lambda(\Psi^T e_t)(e_t^T\Psi)$.

Assuming sequential computations in t, the additional storage requirement is determined by the storage of the terms $R_{\theta(t)}^T R_{\theta(t)}\Lambda$, $R_{\theta(t)}^T R_{\theta(t)}\Lambda(\Psi^T e_t)(e_t^T\Psi)$, and $(R_{\theta(t)}^T g_t + (\gamma - f)e_t)(e_t^T\Psi)$, each requiring $O(KN^2)$.

The operation count for the efficient implementation of the f-step in Line 4 of Algorithm 2 is dominated by the single forward computation of the denoiser for each t, leading to $O(PC_D)$ where C_D is the operation count of the denoiser D_ϕ for a single frame. In addition to the input variables, the only storage requirement is due to the term $\Lambda\Psi^T e_t$ which is $O(N^2)$.

Patch-based RED-PSM

The patch-based variant of RED-PSM uses a reduced spatial size N_B for its inputs where $N_B \ll N$. Compared to the originally proposed RED-PSM formulation in (4.3), the operation count is also dependent on the stride s of the patch extraction operators and the depth of the deep denoiser architecture. Assuming equal-depth or shallower deep convolutional denoisers for patch-based inputs, the total operation count due to the efficient denoiser step stays the same or decreases. Thus, the operation count is again determined by the gradient of the data fidelity term with respect to the patch-based spatial basis functions, and the result approximately scales by the increase in the total stored number of pixels as $O(KN^2PM_i(\frac{N_B}{s})^2)$ where the stride $s \leq N_B$.

The gradients with respect to the data fidelity term require the computation of projections of multiple spatial patches for a given view angle at time t. However, these projections can be computed separately and accumulated. Thus, the space complexity reduces to $O(KN_B^2)$ since we only need to store a single spatial patch of Λ at a given time.

A summary comparison of the two variants of RED-PSM of this section is provided in Section 4.2.4 of the manuscript.

4.6.3 Sample reconstructed spatial and temporal basis functions

To help interpret the operation of the proposed RED-PSM algorithm, we display the reconstructed spatial and temporal basis functions Λ and Ψ for the time-varying walnut scenario with $P = 256$ and DCT-II latent temporal basis U in Figure 4.15. The energy corresponding to each basis pair $E_k = \|\Lambda_k \Psi_k^T\|_F^2$ is shown in Figure 4.16.

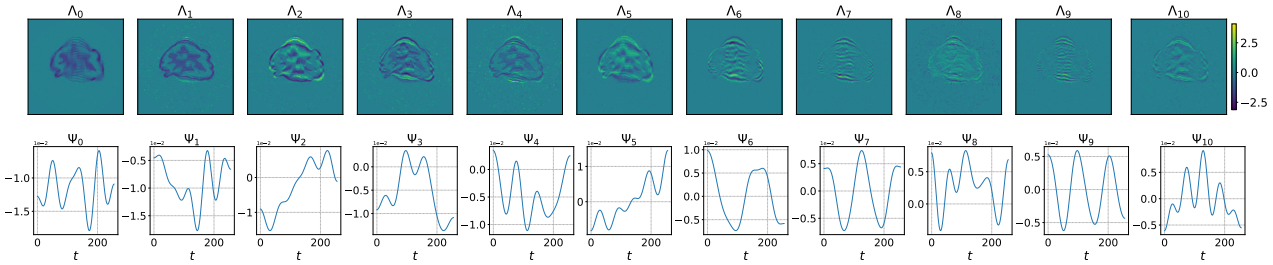


Figure 4.15: Reconstructed $K = 11$ spatial (Λ) and temporal (Ψ) basis functions for the time-varying walnut with $P = 256$ and U as DCT-II basis.

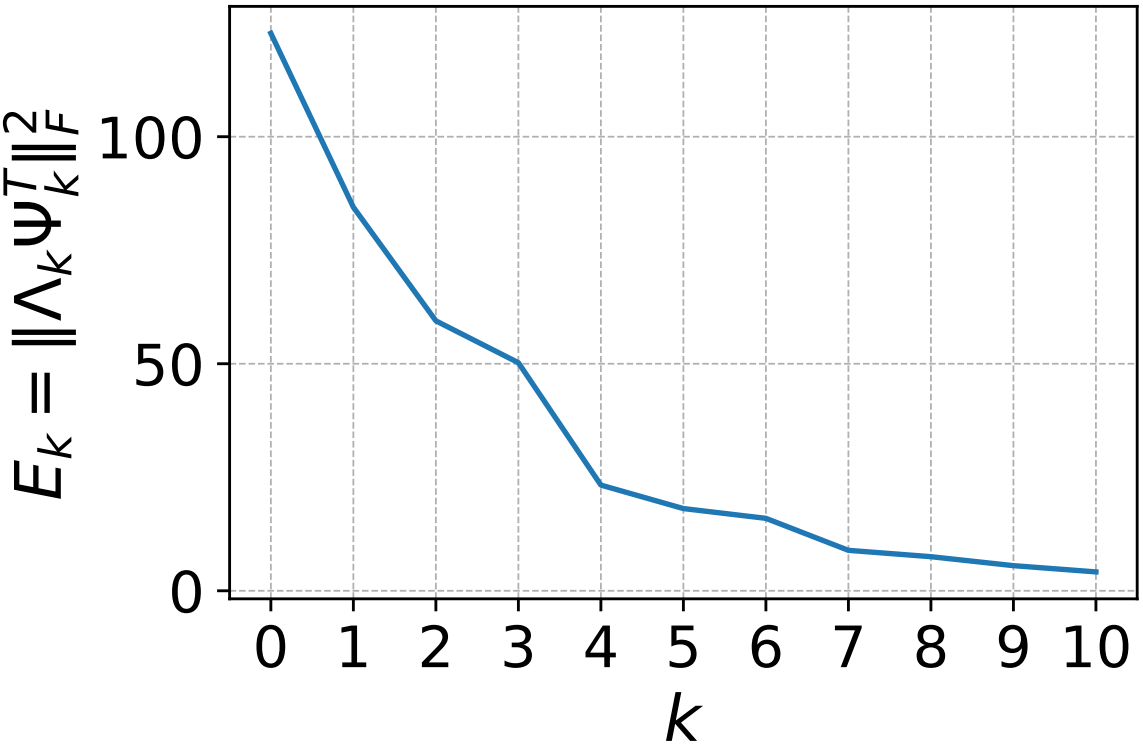


Figure 4.16: $E_k = \|\Lambda_k \Psi_k^T\|_F^2$ for each spatiotemporal basis pair for the setting described in Section 4.6.3.

4.6.4 Proof of Theorem 6

In this section, we provide the detailed steps for the proof of Theorem 6.

We begin by establishing some simple consequences of the assumed properties of the denoiser. First, the objective $H(f, \Lambda, \Psi)$ (4.4) is lower bounded by $\underline{\mathrm{H}} \in \mathbb{R}$ because the RED regularizer $\bar{\rho}(f)$ is non-negative, as follows from the strong passivity assumption on D_ϕ.

$$\begin{aligned}\bar{\rho}(f) &= \langle f, (f - \bar{D}_\phi(f))\rangle = \|f\|_F^2 - \langle f, \bar{D}_\phi(f)\rangle \\ &= \sum_t \|f_t\|_2^2 - f_t^T D_\phi(f_t) \qquad (4.20) \\ &\geq \sum_t \|f_t\|_2^2 - \|f_t\|_2 \|D_\phi(f_t)\|_2 \geq 0. \qquad (4.21)\end{aligned}$$

Second, we have the following result.

Lemma 7. *If the denoiser D_ϕ satisfies the gradient rule in (4.6) and L_D Lipschitz continuous, then the regularizer $\bar{\rho}(f)$ is gradient Lipschitz continuous with gradient Lipschitz constant $L_{\nabla\rho} = 1 + L_D$.*

Proof. Let $f_{(1)}$ and $f_{(2)}$ denote two different spatio-temporal objects. By the gradient rule of ρ, (4.6), and the assumed L_D Lipschitz continuity of the denoiser D_ϕ we have

$$\begin{aligned}&\|\nabla\bar{\rho}(f_{(1)}) - \nabla\bar{\rho}(f_{(2)})\|_F \\ &= \|f_{(1)} - f_{(2)} - (\bar{D}_\phi(f_{(1)}) - \bar{D}_\phi(f_{(2)}))\|_F \\ &\leq \|f_{(1)} - f_{(2)}\|_F + \|\bar{D}_\phi(f_{(1)}) - \bar{D}_\phi(f_{(2)})\|_F \\ &\leq (1 + L_D)\|f_{(1)} - f_{(2)}\|_F.\end{aligned}$$

⊠

Third, we establish strong convexity of the objectives in the different subproblems in Algorithm 2. Clearly, for any choice of positive constant ξ the objectives S_Λ (4.10) and S_Ψ (4.11) in the subproblems for Λ and Ψ are strongly convex with moduli $\alpha_\Lambda \geq \xi$ and $\alpha_\Psi \geq \xi$, respectively.

As we show next, with the assumed $\beta > L$, the objective S_f in (4.12) is strongly convex too with modulus $\alpha_f \geq \beta - L > 0$.

To prove this, we rewrite $S_f = \omega(f) + \frac{1}{2}(\beta - L)\|f\|_F^2$ where

$$\begin{aligned}\omega(f) \triangleq\ & \lambda\bar{\rho}(f) + \frac{L}{2}\|f\|_F^2 - \beta\langle f, \Lambda^{(i)}\Psi^{(i)T} + \gamma^{(i-1)}\rangle \\ &+ \frac{\beta}{2}\|\Lambda^{(i)}\Psi^{(i)T} + \gamma^{(i-1)}\|_F^2\end{aligned}$$

and show that $\omega(f)$ is a convex function. Since $\frac{1}{2}(\beta - L)\|f\|_F^2$ is convex quadratic for $\beta > L$, it then follows (by one of the alternative definitions of strong convexity) that S_f is a strongly convex function with modulus $\alpha_f \geq \beta - L$.

All that remains, is to show that $\omega(f)$ is a convex function. This follows immediately by recalling Lemma 7, that $\lambda\bar{\rho}(f)$ is gradient Lipschitz continuous with gradient Lipschitz constant $L = \lambda(1 + L_D)$, and applying Lemma 8 below to the first two terms of $\omega(f)$.

Lemma 8. (Thm. 2.1 of [188]) *If $\bar{\rho}(f) : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuously differentiable on a convex set C with some Lipschitz constant L, then $\phi(f) = \bar{\rho}(f) + \frac{1}{2}\beta\|f\|_F^2$ is a convex function on C for every $\beta \geq L$.*

With these preliminaries established, our proof of Theorem 6 follows steps similar to [162](Sec. 4.2), but for our own Algorithm 1:

1. Bounding the size of successive differences of the dual variables by those of the primal ones.

2. Showing that $\mathcal{L}_\beta[f^i, \Lambda^i, \Psi^i; \gamma^i]$, the augmented Lagrangian, is a lower-bounded decreasing function.

3. Combining the first two steps and showing convergence to a stationary solution.

The following result upper bounds the successive differences of the dual variable by those of the primal.

Lemma 9. *Using the update rules in Algorithm 1, the following holds:*

$$\|\gamma^{(i)} - \gamma^{(i-1)}\|_F \leq \frac{L}{\beta}\|f^{(i)} - f^{(i-1)}\|_F, \ \forall i \geq 1,$$

where $L = \lambda(1 + L_D)$.

Proof. Since $f^{(i)}$ is the optimal solution for Line 4 of Algorithm 1, it should satisfy the optimality condition

$$\begin{aligned} &\lambda\nabla\bar{\rho}(f^{(i)}) - \beta(\gamma^{(i-1)} - \Lambda\Psi^{(i)T} + f^{(i)}) = 0 \\ &\lambda\nabla\bar{\rho}(f^{(i)}) - \beta\gamma^{(i)} = 0 \end{aligned} \tag{4.22}$$

where (4.22) is due to the dual variable update Step 5 in Algorithm 1. Then, using Lemma 7, we have

$$\begin{aligned} \|\gamma^{(i)} - \gamma^{(i-1)}\|_F &= \frac{\lambda}{\beta}\|\nabla\bar{\rho}(f^{(i)}) - \nabla\bar{\rho}(f^{(i-1)})\|_F \\ &\leq \frac{\lambda}{\beta}(1 + L_D)\|f^{(i)} - f^{(i-1)}\|_F = \frac{L}{\beta}\|f^{(i)} - f^{(i-1)}\|_F. \end{aligned}$$

⊠

Next, using Lemma 9, we show that the augmented Lagrangian is decreasing and bounded below.

Lemma 10. *In Algorithm 1, if $\beta > 2L$, then the following two propositions hold true.*
1. Successive differences of the augmented Lagrangian function (4.9) are bounded above by

$$\begin{aligned}
&\mathcal{L}_\beta[f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}; \gamma^{(i)}] - \mathcal{L}_\beta[f^{(i-1)}, \Lambda^{(i-1)}, \Psi^{(i-1)}; \gamma^{(i-1)}] \\
&\quad \leq -C_f \|f^{(i)} - f^{(i-1)}\|_F^2 - C_\Psi \|\Psi^{(i)} - \Psi^{(i-1)}\|_F^2 \\
&\qquad - C_\Lambda \|\Lambda^{(i)} - \Lambda^{(i-1)}\|_F^2,
\end{aligned} \tag{4.23}$$

where $C_f = \frac{\alpha_f}{2} - \frac{L^2}{\beta}$, $C_\Lambda = \frac{\alpha_\Lambda}{2}$, and $C_\Psi = \frac{\alpha_\Psi}{2}$ are positive constants. That is, the augmented Lagrangian is monotone decreasing.

2. There exists an $\underline{\mathcal{L}_\beta}$ such that

$$\mathcal{L}_\beta[f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}; \gamma^{(i)}] \geq \underline{\mathcal{L}_\beta}.$$

That is, the augmented Lagrangian is lower bounded.

Proof. **Part 1.**

For conciseness define $W^{(i)} = (\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i)})$. Then the successive difference of the augmented Lagrangian can be expressed by adding and subtracting the term $\mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}]$,

$$\begin{aligned}
&\mathcal{L}_\beta[W^{(i)}] - \mathcal{L}_\beta[W^{(i-1)}] \\
&\quad = \mathcal{L}_\beta[W^{(i)}] - \mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}] \\
&\quad + \mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}] - \mathcal{L}_\beta[W^{(i-1)}].
\end{aligned} \tag{4.24}$$

For the first two terms on the RHS, using Lemma 9 we have

$$\begin{aligned}
&\mathcal{L}_\beta[W^{(i)}] - \mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}] \\
&\quad = \beta \langle \gamma^{(i)} - \gamma^{(i-1)}, (\Lambda \Psi^T)^{(i)} - f^{(i)} \rangle \\
&\quad = \|\gamma^{(i)} - \gamma^{(i-1)}\|_F^2 \leq L^2 \|f^{(i)} - f^{(i-1)}\|_F^2.
\end{aligned} \tag{4.25}$$

where the first equality in (4.25) follows from the dual variable update in Step (5) of Algorithm

1. The last two terms on the RHS of (4.24) are further split into

$$\begin{aligned}
&\mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}] - \mathcal{L}_\beta[W^{(i-1)}] \\
&= \mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}] - \mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}] \\
&\quad + \mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}] \\
&\quad - \mathcal{L}_\beta[\Psi^{(i-1)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}] \\
&\quad + \mathcal{L}_\beta[\Psi^{(i-1)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}] - \mathcal{L}_\beta[W^{(i-1)}]
\end{aligned} \tag{4.26}$$

The first two terms on the RHS of (4.26) are upper bounded using the strong convexity of S_f with modulus α_f:

$$\begin{aligned}
&\mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}] - \mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}] \\
&\quad \leq \langle \nabla_f S_f[\Psi^{(i)}, \Lambda^{(i)}, f^{(i)}; \gamma^{(i-1)}], f^{(i)} - f^{(i-1)} \rangle \\
&\qquad - \frac{\alpha_f}{2} \|f^{(i-1)} - f^{(i)}\|_F^2 = -\frac{\alpha_f}{2} \|f^{(i-1)} - f^{(i)}\|_F^2,
\end{aligned} \tag{4.27}$$

where the last line of (4.27) is due to the optimality condition of Line 4 of Algorithm 1 for $f^{(i)}$.

The next two terms on the RHS of (4.26) are upper bounded using the strong convexity of S_Ψ with modulus α_Ψ as

$$\begin{aligned}
&\mathcal{L}_\beta[\Psi^{(i)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}] - \mathcal{L}_\beta[\Psi^{(i-1)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}] \\
&= S_\Psi(\Lambda^{(i)}, \Psi^{(i)}, f^{(i-1)}; \gamma^{(i-1)}) \\
&\quad - S_\Psi(\Lambda^{(i)}, \Psi^{(i-1)}, f^{(i-1)}; \gamma^{(i-1)}) \\
&\leq \langle \nabla_\Psi S_\Psi[\Psi^{(i)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}], \Psi^{(i)} - \Psi^{(i-1)} \rangle \\
&\quad - \frac{\alpha_\Psi}{2} \|\Psi^{(i)} - \Psi^{(i-1)}\|_F^2 = -\frac{\alpha_\Psi}{2} \|\Psi^{(i)} - \Psi^{(i-1)}\|_F^2.
\end{aligned}$$

Repeating similar steps for the last two terms on the RHS of (4.26) using the objective S_Λ leads to the upper bound of $-\frac{\alpha_\Lambda}{2} \|\Lambda^{(i)} - \Lambda^{(i-1)}\|_F^2$.

Finally, we establish the positivity of the constants in (4.23). We have $c_\Psi = \alpha_\Psi \geq \xi > 0$ and $c_\Lambda = \alpha_\Lambda \geq \xi > 0$. Next, because $\alpha_f \geq \beta - L$, we have $C_f = \frac{\alpha_f}{2} - L^2 \geq \frac{\beta - L}{2} - L^2 > 0$, where the last inequality follows by the assumption $\beta > 2L$.

Combining these results, we achieve the inequality in (4.23), which concludes Part 1.

Part 2. For Part 2, we need to show that the augmented Lagrangian is lower bounded.

$$
\begin{aligned}
&L_\beta[W^{(i)}] = H(f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}) \\
&+ \beta\langle \gamma^{(i)}, (\Lambda\Psi^T)^{(i)} - f^{(i)}\rangle + \frac{\beta}{2}\|(\Lambda\Psi^T)^{(i)} - f^{(i)}\|_F^2 && (4.28)\\
&\geq H(f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}) + \langle \lambda\nabla\bar{\rho}(f^{(i)}), (\Lambda\Psi^T)^{(i)} - f^{(i)}\rangle \\
&+ \frac{L}{2}\|(\Lambda\Psi^T)^{(i)} - f^{(i)}\|_F^2 && (4.29)\\
&\geq H(f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}) + \lambda\bar{\rho}((\Lambda\Psi^T)^{(i)}) - \lambda\bar{\rho}(f^{(i)}) && (4.30)\\
&= H(\Lambda^{(i)}\Psi^{(i)T}, \Lambda^{(i)}, \Psi^{(i)}). && (4.31)
\end{aligned}
$$

where (4.29) follows by using (4.22) to replace the second term of (4.28) by that of (4.29) and assuming $\beta > L$, and (4.30) is due to the gradient Lipschitz continuity of the regularizer $\bar{\rho}(\cdot)$,

$$\lambda\bar{\rho}(\Lambda\Psi^T) \leq \lambda\bar{\rho}(f) + \langle \lambda\nabla_f\bar{\rho}(f), \Lambda\Psi^T - f\rangle + \frac{L}{2}\|\Lambda\Psi^T - f\|_F^2.$$

Then, since we initially showed that our minimization objective in (4.4) is lower bounded by $\underline{\mathrm{H}}$, we have

$$H(\Lambda^{(i)}\Psi^{(i)T}, \Lambda^{(i)}, \Psi^{(i)}) \geq \underline{\mathrm{H}},$$

which leads, by combining with (4.31), to $L_\beta[W^{(i)}] \geq \underline{\mathcal{L}_\beta} = \underline{\mathrm{H}}$ and completes the proof of Lemma 10. ⊠

By Lemma 10, the sequence $L_\beta[W^{(i)}]$ of augmented Lagrangian values converges. What remains is to address the convergence of the objective H and the sequence of *iterates* $(f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}, \gamma^{(i)})$ to a stationary solution of (4.3), which we do in the third and final step of the proof of Theorem 6. We start by showing that the duality gap shrinks to zero.

Lemma 11. *Consider Algorithm 1 for solving problem (4.3). If $\beta > 2L$, then the duality gap goes to zero, i.e.*

$$\lim_{i\to\infty} \|(\Lambda\Psi^T)^{(i)} - f^{(i)}\|_F \to 0.$$

Proof. Lemma 10 implies that when $\beta > 2L$, the sequence of augmented Lagrangian values is convergent. Thus, the LHS of (4.23) converges to zero. This also means that each of the norms on the RHS converges to zero to satisfy the inequality. Hence, the following successive

differences converge to zero,

$$f^{(i)} - f^{(i-1)} \to 0; \Lambda^{(i)} - \Lambda^{(i-1)} \to 0; \Psi^{(i)} - \Psi^{(i-1)} \to 0.$$

Furthermore, applying Lemma 9, we have $\gamma^{(i)} - \gamma^{(i-1)} \to 0$. Finally, using the update equation for the dual variable γ, we obtain

$$(\Lambda\Psi^T)^{(i)} - f^{(i)} \to 0, \tag{4.32}$$

which completes the proof of Lemma 11. ⊠

It follows from Lemma 11 that the gap $L_\beta[W^{(i)}] - H(\Lambda^{(i)}\Psi^{(i)T}, \Lambda^{(i)}, \Psi^{(i)})$ between the augmented Lagrangian and the objective converges to zero, and, therefore, thanks to the convergence of $L_\beta[W^{(i)}]$, the objective H too converges. This establishes Part (i) of Theorem 6.

Turning to Part (ii) of Theorem 6, we first show that the iterates are bounded.

Lemma 12. *The iterates* $(f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}, \gamma^{(i)})$ *of Algorithm 1 for solving* (4.3) *are bounded for* $i > 0$.

Proof. We use the following facts:

(F1) $\mathcal{L}_\beta[W^{(i)}]$ is a non-increasing function as i increases for all $i > 0$, as shown in Lemma 10.

(F2) The objective $H(f, \Lambda, \Psi)$ is coercive with respect to the second and the third variables.

It follows from (F1) that $\mathcal{L}_\beta[W^{(i)}] \leq \mathcal{L}_\beta[W^{(0)}],\ \forall i \geq 0$. Combining this with (4.31), yields

$$H((\Lambda\Psi^T)^{(i)}, \Lambda^{(i)}, \Psi^{(i)}) \leq \mathcal{L}_\beta[W^{(i)}] \leq \mathcal{L}_\beta[W^{(0)}]. \tag{4.33}$$

Now, note that by (4.3), the first variable $(\Lambda\Psi^T)^{(i)}$ in $H((\Lambda\Psi^T)^{(i)}, \Lambda^{(i)}, \Psi^{(i)})$ only appears as the argument of $\bar{\rho}(\cdot)$, and by (4.21) $\bar{\rho}(\cdot) \geq 0$. This implies by (F2) that $H((\Lambda\Psi^T)^{(i)}, \Lambda^{(i)}, \Psi^{(i)})$ is coercive with respect to $\Lambda^{(i)}$ and $\Psi^{(i)}$. Combining this with the boundedness of H in (4.33) implies the boundedness of these iterates.

Next, the boundedness of $f^{(i)}$ follows from Lemma 11, since we showed that $\Lambda^{(i)}$ and $\Psi^{(i)}$, thus $(\Lambda\Psi^T)^{(i)}$ are all bounded, and the duality gap shrinks to zero as the iterates progress.

Finally, since all the primary variables are bounded, the augmented Lagrangian is bounded both from below (by Part 2 of Lemma 10) and from above by (4.33), and the dual variable iterates $\gamma^{(i)}$ appear only in the linear Lagrangian term, they are also bounded.

This concludes the proof of Lemma 12. ⊠

It follows that the sequences of iterates $\{f^{(i)}, \Lambda^{(i)}, \Psi^{(i)}, \gamma^{(i)})\}$ has at least one accumulation point $(f^*, \Lambda^*, \Psi^*, \gamma^*)$. Next, we establish the properties of all such accumulation points.

Taking the limit over i in (4.22) yields (4.13a).

Next, since $\Lambda^{(i)}$ is optimal for the subproblem in Line 2 of Algorithm 1, we have

$$\begin{aligned}
&\nabla_\Lambda S_\Lambda(\Psi^{(i-1)}, \Lambda^{(i)}, f^{(i-1)}; \gamma^{(i-1)}) = 0 \\
&2\bar{R}^T(\bar{R}\Lambda^{(i)}\Psi^{(i-1)T}\Psi^{(i-1)} - g\Psi^{(i-1)}) \\
&\quad + \beta(\Lambda^{(i)}\Psi^{(i-1)T} - f^{(i-1)} + \gamma^{(i-1)})\Psi^{(i-1)} + \xi\Lambda^{(i)} = 0
\end{aligned}$$

Taking the limit over i yields (4.13b). Similarly, taking the limit in the optimality condition with respect to $\Psi^{(i)}$, $\nabla_\Psi S_\Psi(\Psi^{(i)}, \Lambda^{(i-1)}, f^{(i-1)}; \gamma^{(i-1)}) = 0$ yields (4.13c). Finally, taking the limit with respect to i in (4.32) verifies (4.13d). These results complete the proof of Theorem 6. ⊠

Chapter 5

RED-NF: Neural Field Regularization by Denoising for Dynamic Imaging

5.1 Introduction

Dynamic imaging considers the ill-posed problem of reconstructing a time-varying object from its undersampled measurements. In the extreme case, these measurements are time-sequential, meaning that only a single of them is available at any given time instant.

Similar to algorithms covered in Chapters 2, 3, 4, the proposed method in this chapter also tackles the ill-posed time-sequential dynamic CT problem [77], [80], [91], [95]–[97], [99], [100], [102], [141]–[144], described in detail in Chapter 1.3. Although the experiments in this work mainly consider the dCT, this problem arises in many areas of science and engineering, including dynamic MRI (dMRI) [26], [31], [32], [51], [66]–[68], photoacoustic CT (PACT) [106], [108], and dynamic 3D scene representation [85], [105], [189]–[193].

To address this challenging problem in dMRI and dCT, different representation alternatives with reduced degrees of freedom have been proposed for the underlying dynamic object. A detailed overview of the proposed techniques is provided in Chapter 1.4.

5.1.1 Proposed Approach

Although the low-rank method with pre-learned spatial priors proposed in Chapter 4 was shown to provide considerable improvements over the Deep Image Prior (DIP)-based techniques [66] and to other low-rank alternatives with simpler priors, it may not provide sufficiently accurate reconstructions when the underlying dynamic object is inherently high-rank.

To overcome this limitation while utilizing learned spatial priors, we propose the recon-

struction algorithm RED-NF, which uses a neural field to provide a highly expressive yet parsimonious representation for the spatio-temporal object combined with a pre-learned spatial prior in the form of RED [145]. RED-NF allows to exploit state-of-the-art denoiser performance across a range of different inverse problems. To enable recovery in critically undersampled scenarios, RED-NF uses in addition a smoothness prior for temporal regularization. To solve the proposed non-convex optimization problem, we use an ADMM with an efficient fixed-point update for the spatial regularizer. The proposed RED-NF leads to performance advantages over the low-rank RED-PSM algorithm proposed in Chapter 4, and another DIP-based scheme, TD-DIP [66], in dynamic CT reconstruction problem.

The method does not require ground-truth dynamic data for incorporating prior to NF scheme as in [109]. In contrast to the NF-based algorithms such as [104], [105], [110], [111] that recover a static template and the corresponding motion field, the proposed method does not assume a fixed total density which may be limiting. For example, injecting a contrast agent to the field-of-view or cross-slice motion of a compressed object while imaging a fixed slice can violate this assumption. As also stated in [85], another issue with the motion field-based methods is their inability to represent topological changes which require discontinuity in the representation. Finally, as compared with the NF-based approach in [106], which uses simpler regularizers, the proposed objective is regularized with more accurate, pre-learned spatial priors.

5.1.2 Contributions

The main contributions of the proposed method are:

- To the best of our knowledge, RED-NF is the first method to combine NF-based object representation with a pre-learned, RED-based [145] spatial prior.
- RED-NF does not require spatio-temporal training data.
- The proposed optimization scheme avoids costly gradient computations through the deep denoiser for RED updates.

5.2 Background and Related Work

5.2.1 Neural Fields (NF)

Neural fields (NFs), also known as implicit neural representations (INRs), have recently become popular as a resolution-free, coordinate-based method for representing objects and

natural scenes using deep neural networks [194]. NFs compute the d_o-dimensional output features $\mu_{\boldsymbol{\nu}} \in \mathbb{R}^{d_o}$ (e.g. density, color, etc.), that represent a d_i-dimensional object f at a given input coordinate $\boldsymbol{\nu} \in \mathbb{R}^{d_i}$ as

$$f^{\alpha}(\boldsymbol{\nu}) = \mu_{\boldsymbol{\nu}} \tag{5.1}$$

where $f^{\alpha} : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d_o}$ is a deep neural network with an MLP-based architecture parametrized by α.

Applications of NFs include, but are not limited to, solving partially differentiable equations (PDEs) [195], 3D static [196]–[199] and dynamic scene representation and novel view synthesis.

Current NF methods for spatio-temporal scenes can be categorized, for the most part, into different groups as follows: (i) augmenting the temporal coordinate to the input [189], [191]; (ii) decomposing the object into a static template and a temporal deformation field applied to the static template by warping its input spatial coordinates as a function of time [105], [190], [192], [193], [200], [201]; (iii) representing the object as a so-called conditional NF, whereby a hyper network is used to represent the mapping from the time coordinate to the spatial NF parameters [202]; (iv) lifting into a higher dimension and representing the spatio-temporal object as level sets [85].

Neural fields are also used in various computational imaging applications such as static CT [199], [203]–[205] and MRI reconstruction [199], dynamic CT [104], [109], [111] and dynamic MRI [69]–[71], [73] reconstruction as covered in Chapter 1.4, dynamic photoacoustic tomography [106], [108], cryo-EM [206], addressing measurement coordinate uncertainty [207], and medical imaging in general [208].

As mentioned previously, MLP-based architectures are used in NFs. However, the regular MLP with ReLU nonlinearity leads to degraded representation accuracy, missing spatial high-frequency details in different applications due to spectral bias [75]. To avoid this issue, positional encodings using fixed transforms to increase the input dimensionality and to introduce high-frequency features [74], [203], [209], [210], and different nonlinearity types in the activation functions [195], [211] were used. Such nonlinearities include sinusoidal [195] and wavelet-based [211] alternatives. The analysis in [212] provides bounds on the expressivity of wavelet-based nonlinearity [211] whereas in [213] specifically addresses the case of sinusoidal positional encodings with polynomial nonlinearities, stating that sinusoidal activations can be approximated by a low order polynomial.

Despite being resolution-free and the improved representation accuracy with reduced degrees of freedom, standard NF implementations may suffer from slow optimization speeds

when minimizing a standard loss function such as mean-squared error using gradient descent. To this end, several methods were proposed to speed up the optimization process for various applications [108], [214]–[217]. For 3D scene representation, in [215], caching is used to speed up the rendering process significantly, [216] performs a low-rank tensor decomposition, and [217] uses multi-resolution hash encodings. Differently, [108] proposes a proximal splitting method separating the NF parameter optimization as an embedding task from the updates that call the imaging operator, claiming advantages for high-resolution or high-dimensional object reconstruction.

To incorporate additional prior information to the reconstruction tasks using NFs in various applications, different techniques are used. These include conditioning NFs which provides a way to express prior using a latent code. The latent codes are used as an input to a hypernetwork to output a subset of NF parameters [218]–[221]. The algorithms differ in the mapping type or the parameters they map.

Another alternative is the gradient-based meta-learning [222]. Assuming the availability of observations from a particular distribution, the NF is initialized with a parameter configuration sufficiently close to configurations that represent different objects in the same distribution. Then, the NF is fine-tuned with test-time optimization with few gradient descent steps [223]–[225].

In this work, different from previous approaches, we propose a pre-learned RED-based spatial prior for the NF scheme.

5.2.2 Regularization-by-Denoising (RED)

As also covered in further detail in Chapter 4.2.2, *Regularization-by-Denoising* was proposed as a regularization technique using a denoising operator $D : \mathbb{R}^N \to \mathbb{R}^N$ [145].

RED was originally inspired by the Plug-and-Play (PnP) method [147], [226] which uses denoisers to replace a proximal mapping without explicitly defining the regularizer. Differently, RED proposes an explicit regularizer

$$\rho(f_t) = \frac{1}{2} f_t^T (f_t - D(f_t)), \tag{5.2}$$

leading to a simple gradient expression, and an easy implementation.

Combined with different representation techniques for the object, the regularizer has been widely used for *static* reconstruction tasks [155]–[157], and recently applied for the first time to the dynamic imaging setting [141], [142].

Despite providing a very flexible framework for different types of denoisers, for the original theoretical analysis in [145] to apply, D is required to be differentiable, locally homogeneous,

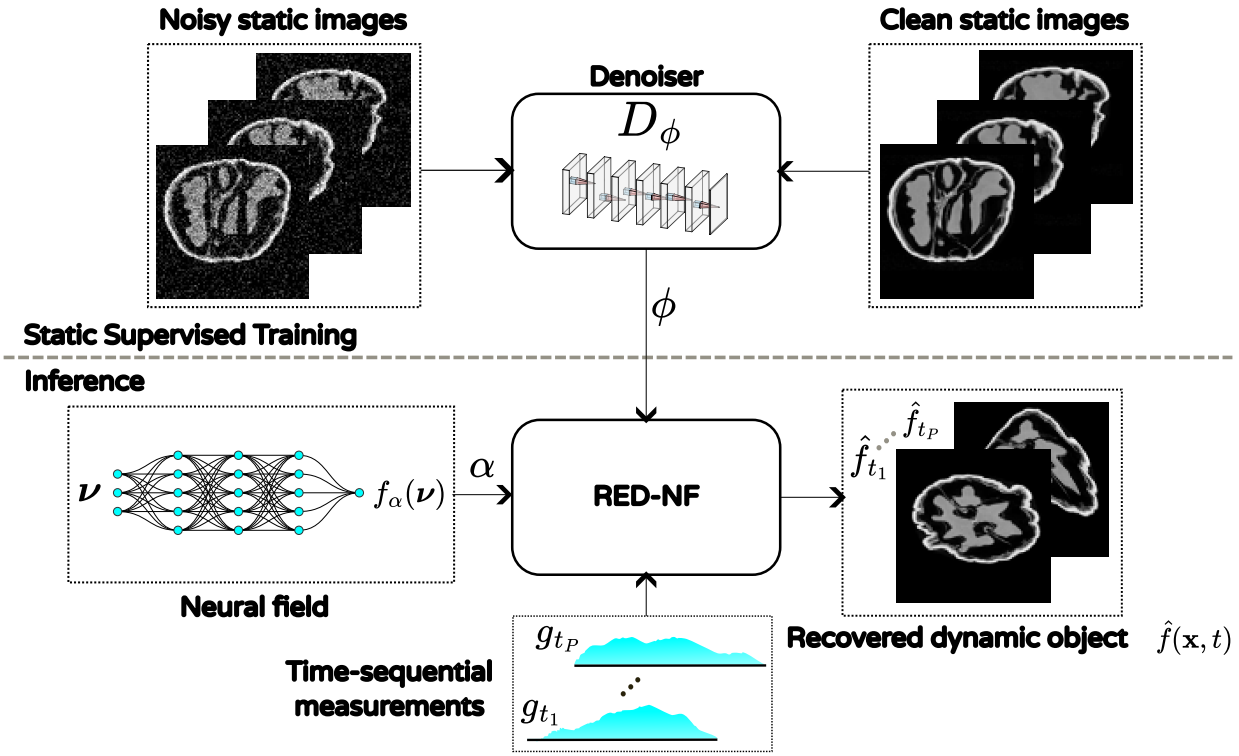


Figure 5.1: The RED-NF framework. The deep denoiser D_ϕ is trained on slices of *static* objects similar to the object of interest, and the learned spatial prior is used at inference time.

strongly passive, and as shown in [151], Jacobian symmetric. Although these conditions are not simultaneously satisfied for many denoisers, [151] provides a score-matching-based framework to explain the good performance of RED in such cases.

5.3 Proposed Method: RED-NF

5.3.1 Neural field representation for the object f

To represent the complete 2D dynamic object f, we use a neural field $N_\alpha : \mathbb{R}^3 \to \mathbb{R}$ where the temporal coordinate is also added to the input as $\boldsymbol{\nu} = (x, t)$ with all coordinates normalized such that $\boldsymbol{\nu} \in [0, 1]^3$, and α represents the learnable parameters. The NF has a MLP architecture using fixed Fourier positional encodings $\gamma(\boldsymbol{\nu}) \in \mathbb{R}^{L\times 6}$ with L linearly increasing frequencies [203], where the l-th row is defined as

$$\gamma(\boldsymbol{\nu})_l = (\sin(\pi l/2\boldsymbol{\nu}), \cos(\pi l/2\boldsymbol{\nu})) \tag{5.3}$$

where $l \in \{1, \ldots, L\}$, and sin and cos are applied to coordinate vectors element-wise. The positional encodings are followed by a MLP network $M_\alpha : \mathbb{R}^{L\times 6} \to \mathbb{R}$ mapping the encoded coordinates to the estimated output densities. Thus, the overall NF model computes the density values $\mu_{\boldsymbol{\nu}} \in \mathbb{R}$ at a given spatio-temporal coordinate $\boldsymbol{\nu} \in \mathbb{R}^3$ as

$$f^\alpha(\boldsymbol{\nu}) = M_\alpha(\gamma(\boldsymbol{\nu})) = \mu_{\boldsymbol{\nu}}. \tag{5.4}$$

5.3.2 Variational Formulation

The variational objective to be optimized is

$$\begin{aligned} \arg\min_{\alpha,\bar{f}} \sum_t \|g_{\theta(t),t} - R_{\theta(t)} f_t^\alpha\|_2^2 + \lambda\rho(\bar{f}_t) + \xi\rho_\tau(f^\alpha) \\ \text{s.t. } f^\alpha = \bar{f} \end{aligned} \tag{5.5}$$

where $g_{\theta(t),t} \in \mathbb{R}^J$ is the true time-sequential projection at time t and view angle $\theta(t)$ corrupted with AWGN, $f^\alpha \in \mathbb{R}^{J^2\times P}$ is the spatially-vectorized neural field representation of the object rendered on a fixed $J^2 \times P$ grid aligned with the space-time coordinates, and $f_t^\alpha \in \mathbb{R}^{J^2}$, $t = 0, \ldots P-1$ is its t-th time instant ("frame"), $R_{\theta(t)} : \mathbb{R}^{J^2} \to \mathbb{R}^J$ is the discretized Radon transform operator at view angle $\theta(t)$. The temporal regularizer ρ_τ is a finite difference approximation to the energy in the second order derivative with respect to time $\|\partial^2 f/\partial t^2\|_F^2$,

$$\rho_\tau(f^\alpha) = \sum_{t=2}^{P-1} \|f_{t-1}^\alpha - 2f_t^\alpha + f_{t+1}^\alpha\|_2^2. \tag{5.6}$$

The spatial regularizer ρ with weight $\lambda > 0$ is the RED term

$$\rho(\bar{f}_t) = \bar{f}_t^T(\bar{f}_t - D_\phi(\bar{f}_t)) \tag{5.7}$$

where D_ϕ is a denoiser that operates on a single time frame f_t of the object. Note that in (5.5) and (5.7), the spatial regularizer ρ is applied to auxiliary variable $\bar{f}$, rather than directly to the NF output f^α. Using the NF representation for the RED updates in solving the optimization problem (5.5) would require costly gradient computations through the deep denoiser. To avoid this, we perform the variable splitting $f^\alpha = \bar{f}$. The split variable $\bar{f} \in \mathbb{R}^{J^2\times P}$ is constrained to equal the rendered NF representation f^α.

To solve the resulting problem, we use an ADMM framework similar to [141], [142]. This time the variable split is performed on the NF representation instead of on the low-rank

bilinear PSM. The resulting augmented Lagrangian in the scaled form [159], [160] is

$$\mathcal{L}_\beta[f^\alpha, \bar{f}; \gamma] = \sum_t \left(\left\| R_{\theta_t} f^\alpha e_t - g_t \right\|_2^2 + \lambda \rho(\bar{f} e_t) \right) + \xi \rho_\tau(f^\alpha) \quad - \frac{\beta}{2} \|\gamma\|_F^2 + \frac{\beta}{2} \|f^\alpha - \bar{f} + \gamma\|_F^2, \tag{5.8}$$

where $\gamma \in \mathbb{R}^{J^2 \times P}$ is the dual variable and $\beta > 0$ is the augmented Lagrangian weight.

The proposed RED-NF algorithm minimizes the augmented Lagrangian (5.8) with respect to the primal variables f^α and $\bar{f}$ and performs a dual ascent step for γ. The overall method is described in Algorithm 3.

Algorithm 3 RED-NF

input: $\alpha^{(0)}$, $\bar{f}^{(0)} = f_\alpha^{(0)}$, $\gamma^{(0)}$, $\beta > 0$, $\lambda > 0$, $\xi > 0$

1: **for** $i \in \{1, \ldots, I\}$ **do**
2: $\quad \alpha^{(i)} = \arg\min_\alpha \{ \sum_t \|g(\cdot, \theta(t), t) - R_{\theta(t)} f_t^\alpha\|_2^2 + \frac{\beta}{2} \|f_t^\alpha + \gamma_t^{(i-1)} - \bar{f}_t^{(i-1)}\|_F^2 + \xi \rho_\tau(f^\alpha) \}$
3: $\quad \forall t: \ \bar{f}_t^{(i)} = \arg\min_{\bar{f}_t} \{ \lambda \rho(\bar{f}_t) + \frac{\beta}{2} \|(f^{\alpha(i)} + \gamma^{(i-1)}) e_t - \bar{f}_t\|_2^2 \}$
4: $\quad \gamma^{(i)} = \gamma^{(i-1)} + f^{\alpha(i)} - \bar{f}^{(i)}$
5: **end for**

As in Chapter 4.2.2, the third step can be replaced by a fixed-point update with early stopping. Using the gradient rule

$$\nabla \rho(\bar{f}_t) = \bar{f}_t - D_\phi(\bar{f}_t), \tag{5.9}$$

this leads to the efficient single use of the denoiser D_ϕ at each outer iteration.

Algorithm 4 RED-NF with efficient f-step

Notes: Inputs, and Lines 1-2 and 4-5 are the same as Algorithm 3. The f-step is applied $\forall t$.

4: $\forall t: \ \bar{f}_t^{(i)} = \frac{\lambda}{\lambda+\beta} D_\phi(\bar{f}_t^{(i-1)}) + \frac{\beta}{\lambda+\beta} \left(f_t^{\alpha(i)} + \gamma_t^{(i-1)} \right)$

5.3.3 Regularization Denoiser

As in Chapter 4.2.3 and [155], the regularization denoiser D_ϕ has a DnCNN architecture [161]. The denoiser is pre-trained on a training set of 2D slices of similarly distributed *static* objects. In the absence of such data, the training is performed using pre and post-motion static slices of the same object. This strategy is agnostic to the specific dynamic behavior of

the object of interest. The variational objective for the supervised pretraining of the denoiser is

$$\min_{\phi} \sum_{i} \|f_i - D_\phi(\tilde{f}_i)\|_F^2 \quad s.t. \quad \tilde{f}_i = f_i + \eta_i, \ \forall i, \tag{5.10}$$

where to allow the denoiser to denoise various levels of artifacts, the injected noise $n_i \sim \mathcal{N}(0, \sigma_i^2 I)$ has noise level $\sigma_i \sim U[0, \sigma_{\max}]$.

5.4 Experiments

5.4.1 Datasets

Compressed polymer

The dataset of a porous polymer under compression was generated at the Los Alamos National Laboratory by physics-based numerical simulation as the evolution over time of the coordinates of finite element nodal positions. The material is a polymer with a cellular structure under quasi-static compression where the complete temporal evolution consists of 1001 time frames. To obtain a voxel-based representation, we first create a 4D $128 \times 128 \times 128 \times 1001$ spatio-temporal grid with the object at $t = 0$ at the center, and map each node in all time frames to the corresponding voxel. The mapping is binary, such that the voxel density is 1 if there are one or more nodes that are mapped to it, and 0 otherwise. Then, we conduct our experiments on the $128 \times 128 \times 1001$ fixed 48th axial slice of the 3D dynamic object shown in Figure 5.2.

Walnut dataset

We use the test object defined in Chapter 2.6. The RED denoiser D_ϕ is trained using the static axial, coronal and sagittal slices of the other static walnut reconstruction in the dataset [127] using the same policy in Chapter 4.4.

5.4.2 Comparison Benchmarks

RED-PSM

The method described in Chapter 4, RED-PSM [141], [142], [144], combines an object-domain low-rank PSM with the RED [145] framework, and the resulting optimization objective is minimized using an effective bilinear ADMM algorithm with convergence guarantees. As

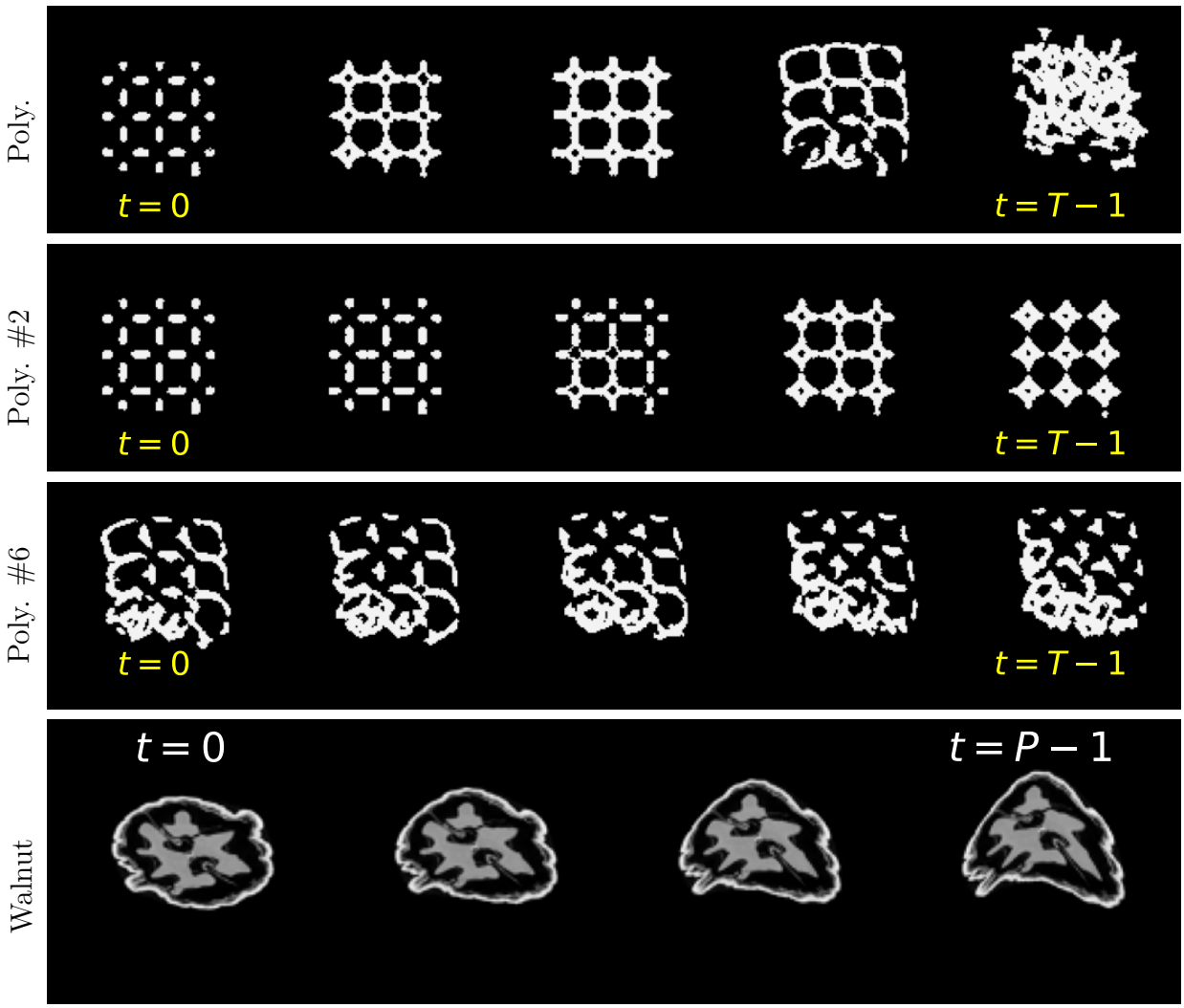


Figure 5.2: Ground-truth frames of different dynamic objects: (i)-(iii) the compressed polymer uniformly sampled in time: (i) Complete spatio-temporal object for T=1001, (ii) non-overlapping subinterval #2 of length T=128, (iii) non-overlapping subinterval #6 of length T=128; (iv) warped walnut for $T = 128$.

shown in Chapter 4.4.4, to accelerate and improve convergence, the method can use a fast projection-domain PSM [95], [96] estimate for initialization.

TD-DIP

The Time-Dependent Deep Image Prior for Dynamic MRI (TD-DIP) [66] is an unsupervised DIP-based algorithm, proposed for dynamic MRI. As in Chapter 4, for d-CT, we replace the forward operator with the 2D Radon transform $R_{\theta(t)}$. TD-DIP uses fixed low-dimensional latent representations to incorporate motion-specific prior and subsequent two-stage deep neural network architecture to generate the estimated object at each time instant.

TV-PSM

This algorithm also uses a low-rank PSM as in the case of RED-PSM [141]–[143] but replaces the RED spatial regularizer in the objective with spatio-temporal total variation (TV) by computing TV spatially and temporally between adjacent frames. The resulting objective is minimized for the spatial and temporal PSM components using the Adam optimizer in Pytorch.

5.4.3 Experimental Settings

All of the methods are run on a workstation with an Intel(R) Xeon(R) Gold 5320 CPU and NVIDIA RTX A6000 GPU. The optimization subproblem for the NF parameters in Algorithm 3 was solved using SGD with the Adam [126] optimizer. For each update, the mini-batches were of size $P/8$ where the projections from the entire set of P projections were sampled at random (with replacement).

Denoiser and Pre-training

As in [141]–[143], each convolutional layer in DnCNN is followed by a ReLU nonlinearity, except for the final single output channel layer. A single denoiser was pre-trained for each of the objects and used in all concerning that object. The upper limit for noise level used in training the denoiser was set to $\sigma_{\max} = 5 \cdot 10^{-2}$. For the dynamic walnut object, the denoiser D_ϕ is trained on the central 200 axial, 200 sagittal, and 200 coronal slices of another static walnut CT reconstruction downsampled to size 128×128. For the compressed polymer, a total of 422 axial, sagittal, and coronal slices of the pre-compression and post-compression objects were used to train the denoiser. Although we conduct experiments on the temporal subintervals of the full dynamic object, we do not assume the availability of the beginning and end frames of each subinterval for training.

Tomographic Acquisition Scheme

The methods used in this paper employ the bit-reversed angular sampling scheme over the range $[0, \pi]$. Experiments in [95], [96] have shown that for the time-sequential acquisition scheme, the bit-reversed scheme provides better conditioning of the forward operator compared to other alternatives.

Although the bit-reversed scheme requires increased rotation speed in practical scenarios, in certain applications such as MRI with radial acquisition, CT scanner with electronic beam deflection [177], or in scenarios where the acquisition time of a single projection dominates the rotation time (e.g., because of low source photon flux or a dense material), this does not constitute a problem. Also, in some physically constrained settings, the number of distinct views can be reduced and these views can be repeated periodically using multiple source-detector pairs, or by carbon nanotube sources [178], [179].

NF architecture and initialization

Unless stated otherwise, the neural field used in the experiments consists of $L = 10$ different Fourier positional encodings for each of the input coordinates with linearly increasing frequencies in the interval $[\pi/2, L\pi/2]$ [203] followed by a MLP architecture with 7 hidden layers of width 64. The final nonlinearity is a sigmoid layer. The MLP parameters α are initialized randomly.

The number of hidden layers, considering the number of sinusoidal encodings L hence the maximum encoding frequency $L\pi/2$, and the resolution $J = 128$ of the sampling grid, is sufficient for achieving the maximum frequency on the grid without aliasing. This is related to the expressivity analysis in [213] which leads to a requirement of $\log_2(\frac{J\pi}{L\pi/2}) < 5$ layers for this condition.

Evaluation Metrics

To evaluate the performances of the compared methods, we use the peak signal-to-noise ratio (PSNR) in dB, the structural similarity index (SSIM) [180], the mean absolute error (MAE), and the high-frequency error norm (HFEN) [181] as $\mathrm{HFEN}(f, f_r) = \|\mathrm{LoG}(f) - \mathrm{LoG}(f_r)\|_2$ where LoG is the rotationally symmetric Laplacian of Gaussian filter with a standard deviation of 1.5 pixels.

5.4.4 Results

Embedding experiments

In this subsection, we compare the representation accuracies of the proposed neural field (NF) model with the partition of unity NF network (POUnet) architecture [106] and low-rank PSM for different numbers of degrees of freedom (parameters in the NF model). We perform this comparison on the full data (not projection measurements) for different subintervals of the compressed polymer data set. For the PSM representation, we show the rank K Casorati matrix accuracies obtained by truncating the SVD of the object f to include only the first K largest singular values for different K. On the other hand, for the two NF representations, we conduct embedding experiments for MLP architectures with different numbers of layers and number of channels per layer. To optimize the NF parameters, we use Adam optimizer with 125K iterations with a learning rate of $5(10^{-3})$.

Qualitative comparison of representations at different time instants using the low-rank PSM and the proposed NF architecture are shown in Figure 5.3 for the compressed polymer subintervals 2 and 6. In these comparisons, the NF uses 7 layers and 64 channels per layer, and the low-rank PSM representation has a rank of 3 which leads to a similar number of free parameters for both alternatives. As can be seen in the absolute error figures, the NF provides substantial accuracy improvements for both objects, whereas the PSM leads to over-smoothed representations, especially for Subinterval #6, which has faster dynamics.

We also provide PSNR (in dB) comparisons for various representations with different number of degrees of freedom in Figure 5.4. The results verify that for similar numbers of free parameters, NF-based representations with various widths and depths provide significant improvements over the PSM. The NF architecture used in this work, and the POUnet alternative of [106] provide similar representation accuracies for similar number of free parameters. However, for Subinterval #2, when the number of degree of freedom is small, the proposed architecture provides improvement over the POUnet.

Another aspect in the comparison of the PSM vs. NF approaches, is that for each K, the accuracy for the Casorati matrix truncated to rank K constitutes an upper bound on the ℓ_2 sense performance of the rank-k PSM. Conversely, for the NF embeddings, changes in the optimization scheme may enable further accuracy improvements.

Dynamic CT reconstruction accuracies for different P

We compare the proposed RED-NF with three different benchmarks: (i) the previously proposed low-rank RED-PSM method [141]–[143]; (ii) another PSM with a simpler TV-based prior, TV-PSM; and (iii) TD-DIP [66]. This comparison is done on the warped walnut and

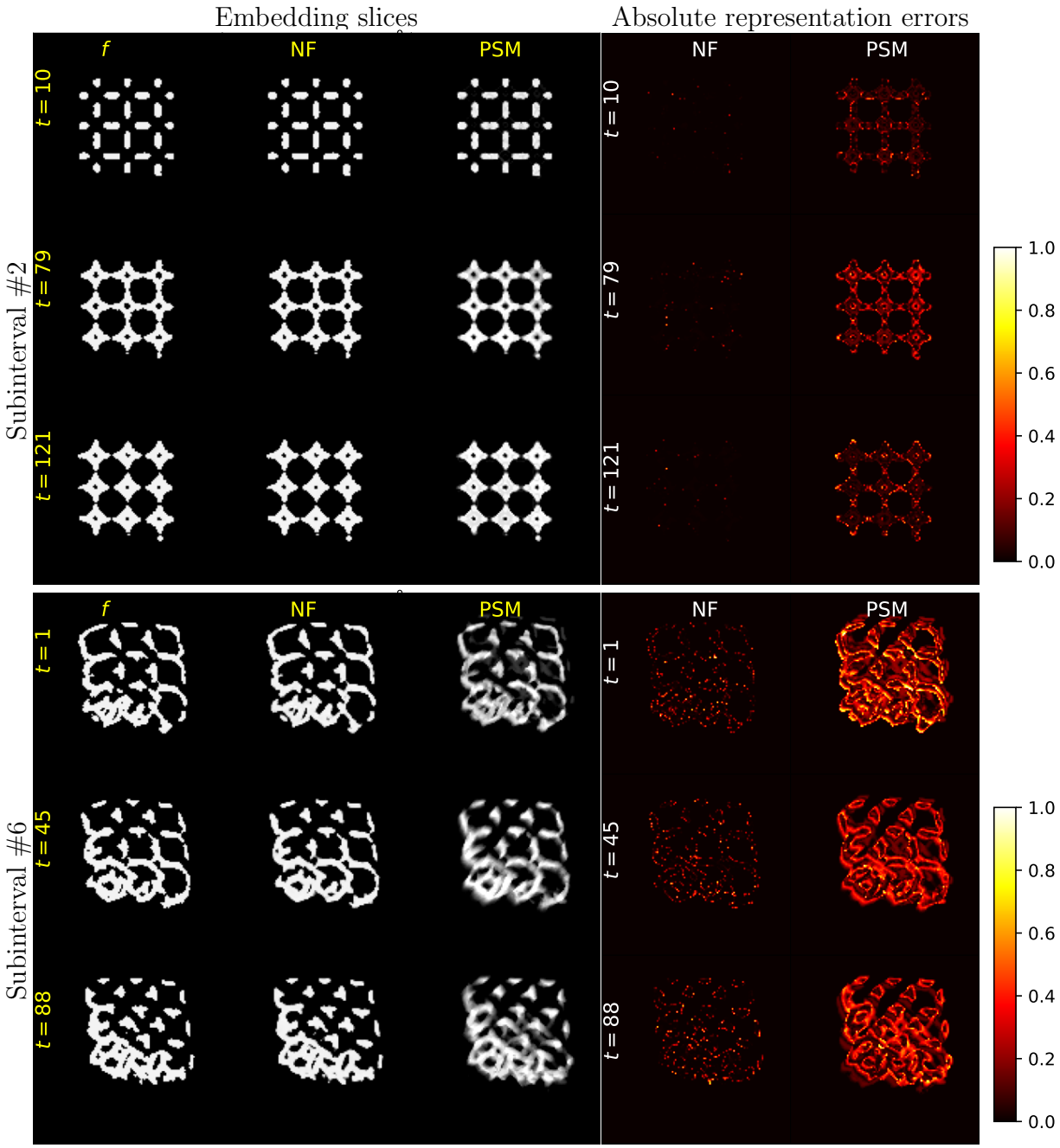


Figure 5.3: Comparison of representations at three time instants for NF embedding with 7 layers and 64 channel width per layer and low-rank PSM Casorati estimate of rank K=3 where P=128, and the corresponding normalized absolute errors for the compressed polymer object subintervals #2 and #6. The PSM rank and the NF architecture are selected such that the number of free parameters in PSM is approximately $\sim 1.5\times$ more than NF.

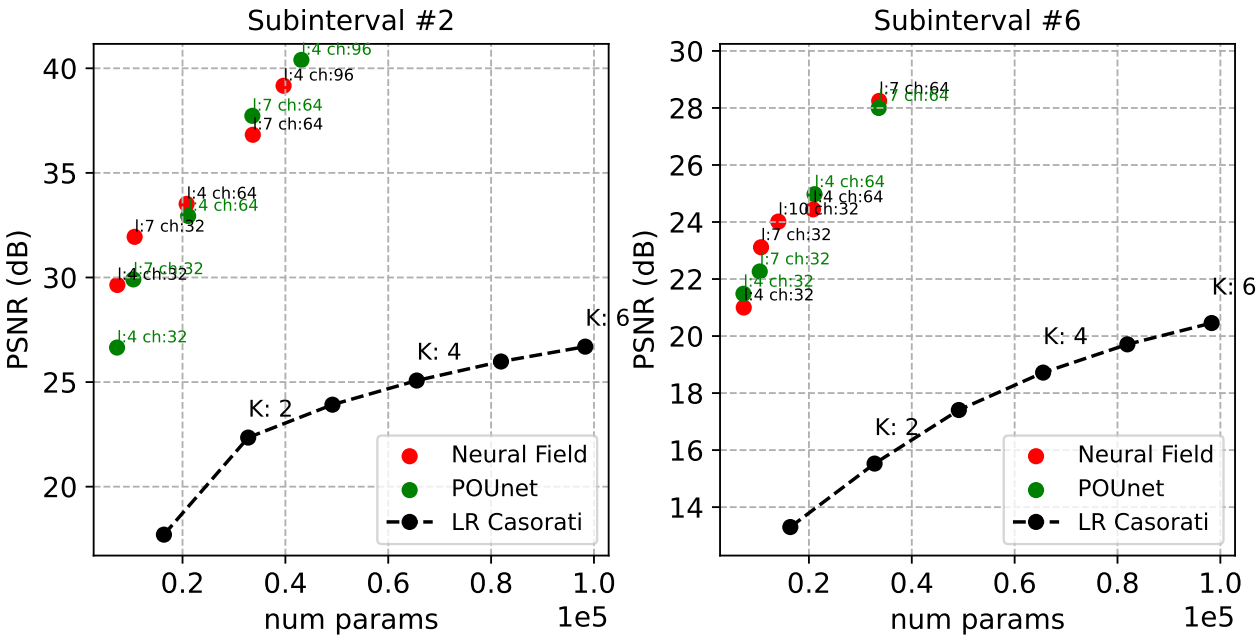


Figure 5.4: Representation PSNR (in dB) comparisons for different number of free parameters for NF embeddings and low-rank PSM Casorati representations of rank K.

polymer Subinterval #2, using the time-sequential projection data. The parameter configurations for each of these techniques for the various methods are provided in Supplementary Material Table 5.5. The accuracy metrics vs. number of time-sequential measurements P are shown in Figure 5.5. Since the range of metrics for different P usually spans a large interval in Figure 5.5, Tables 5.1 and 5.2 enable a more precise comparison of the plotted accuracies. For both objects and for various P, the RED-NF reconstructions have substantially improved metrics over the other three compared methods.

In Figure 5.6, we provide a qualitative comparison of reconstructions for three different time instants $P = 128$ for the dynamic walnut. To facilitate comparison, we include only the two best-performing methods in the figure, RED-NF and RED-PSM. Compared to the low-rank RED-PSM, RED-NF suppresses the errors further considerably. Specifically, the high-frequency features around the shell of the walnut are recovered with greater accuracy for RED-NF, whereas the low-rank method suffers from over-smoothing. Likewise, Figure 5.7 shows a comparison of reconstructions for three different time instants $P = 128$ for two subintervals of the polymer data: Subintervals #2 and #6. Again, we include only the two methods in the figure, RED-NF and RED-PSM. As expected from the embedding accuracies in Figure 5.4, both methods perform better for subinterval #2 than for subinterval #6, which has more complicated dynamic behavior based on visual examination , with RED-NF being considerably more accurate than RED-PSM for both subintervals. For Subinterval #6,

the errors are concentrated on the edges of the object for RED-NF, whereas they are more uniformly spread to the inner parts and with larger magnitudes for RED-PSM.

Table 5.1: Reconstruction accuracies for different P for dynamic walnut. For TD-DIP, the reported accuracies are for the best PSNR using a "stopping oracle", averaged over three runs with random initial conditions.

P	Method	PSNR (dB)	SSIM	MAE (1e-3)	HFEN
32	PSM-TV-S	21.1	0.678	2.9	1.04
	PSM-TV-ST	21.3	0.731	2.6	0.99
	TD-DIP (L)	22.5	0.882	2.1	0.87
	RED-PSM	22.8	0.911	1.5	0.78
	RED-NF	**25.9**	**0.962**	**1.1**	**0.51**
64	PSM-TV-S	24.0	0.879	1.9	0.94
	PSM-TV-ST	24.6	0.897	1.7	0.83
	TD-DIP (L)	25.6	0.916	1.5	0.74
	RED-PSM	26.4	0.958	0.9	0.57
	RED-NF	**31.1**	**0.987**	**0.6**	**0.28**
128	PSM-TV-S	26.8	0.912	1.4	0.78
	PSM-TV-ST	27.7	0.931	1.2	0.63
	TD-DIP (L)	29.0	0.953	1.0	0.50
	RED-PSM	30.3	0.979	0.6	0.40
	RED-NF	**33.9**	**0.993**	**0.4**	**0.26**

Table 5.2: Reconstruction accuracies for different P for compressed polymer subinterval #2. For TD-DIP, the reported accuracies are for the best PSNR using a "stopping oracle", averaged over three runs with random initial conditions.

P	Method	PSNR (dB)	SSIM	MAE (1e-3)	HFEN
32	TD-DIP (L)	21.1	0.850	0.0309	9.67
	RED-PSM	20.2	0.865	0.0242	10.92
	RED-NF	**23.6**	**0.974**	**0.0099**	**6.32**
64	TD-DIP (L)	23.5	0.915	0.0213	8.59
	RED-PSM	22.7	0.917	0.0151	9.74
	RED-NF	**26.0**	**0.985**	**0.0061**	**6.22**
128	TD-DIP (L)	24.3	0.849	0.0241	10.58
	RED-PSM	25.3	0.940	0.0098	9.68
	RED-NF	**30.2**	**0.995**	**0.0021**	**5.22**

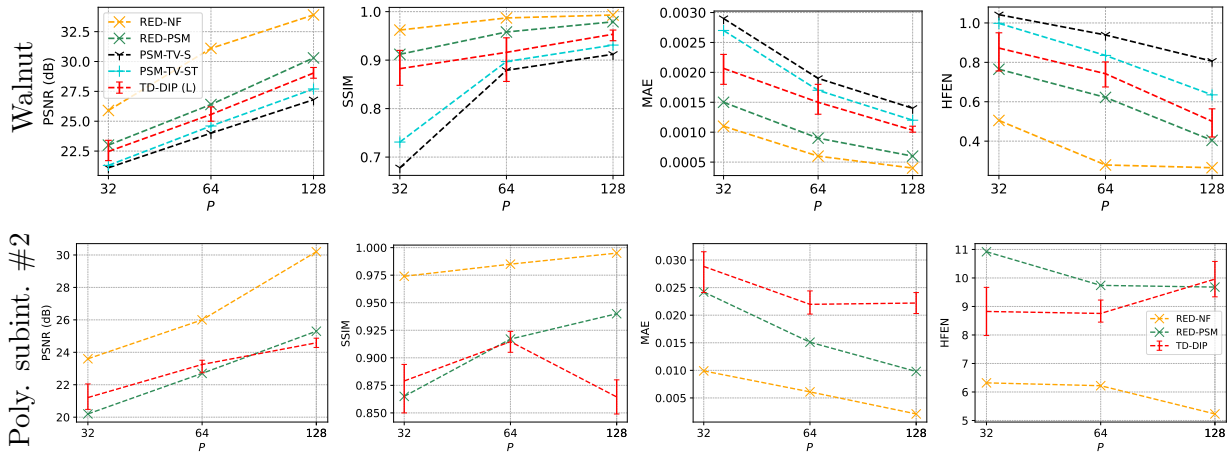


Figure 5.5: Reconstruction metrics for the time-varying walnut and compressed polymer vs. P using different methods. For TD-DIP, the metrics reported are assuming a "stopping oracle" that stops the iterations at the best PSNR reconstruction.

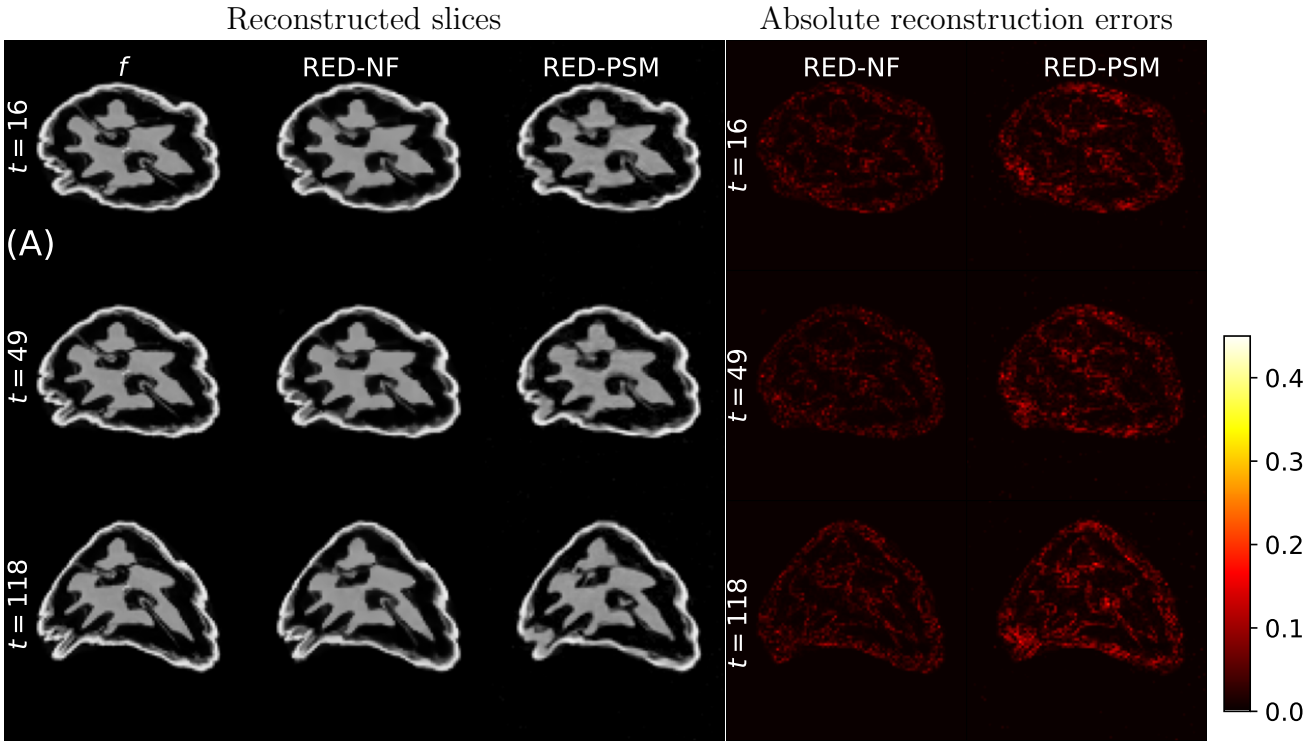


Figure 5.6: Comparison of reconstructed object frames for dynamic walnut at three time instants using different methods for P=128, and the corresponding normalized absolute reconstruction errors.

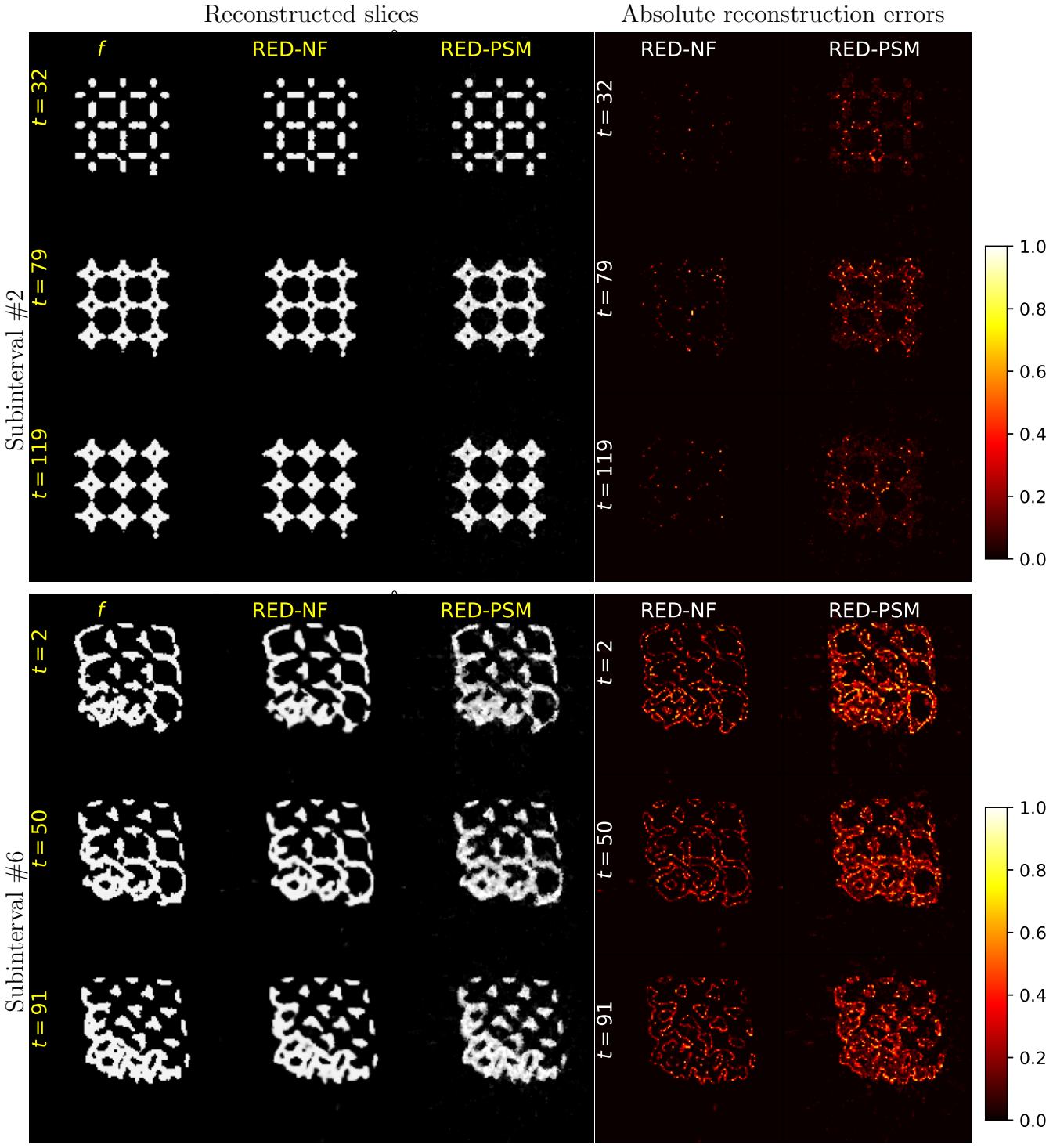


Figure 5.7: Comparison of reconstructed object frames at two time instants using different methods for P=128, and the corresponding normalized absolute reconstruction errors for compressed polymer object subintervals #2 and #6.

Effect of denoiser pre-training

The experiments in this subsection seek answers to two questions related to the proposed semi-supervised denoiser pretraining strategy for D: Is training the denoiser on representative data helpful, and, can a generic prior be as effective as the proposed strategy?

Thus, we compare the proposed strategy with different alternatives. These include: (i) an unsupervised wavelet-based denoiser [227] using Daubechies 1 wavelet and applying an adaptive soft-thresholding for each subband [228]; (ii) the DnCNN-based denoiser pre-trained on static datasets with significantly different distributions; and (iii) the proposed pre-training strategy in Section 5.3.3. In these experiments, all other hyperparameters of the method were kept the same.

The reconstruction accuracies are reported in Table 5.3. The accuracy metrics for the dynamic walnut RED-NF reconstructions with a deep denoiser highlight the importance of the *static* pre-training dataset distribution. As expected, for both $P = 32$ and $P = 128$ pre-training on static slices of another walnut provides the best performance. Switching instead to pre-training on the pre- and post-compression polymer object slices ("wrong distribution") leads to a substantial decline in the accuracy, especially for $P = 128$. Finally, for both $P = 32$ and $P = 128$, the unsupervised generic wavelet-based denoiser alternative ranks second-best. However, the improvement provided by the well-matched spatial prior introduced by the static pre-training over the generic wavelet-based alternative is greater for $P = 32$ than for $P = 128$: 1.7 dB vs. 0.8 dB in PSNR, and 0.02 vs. 0.001 in SSIM. This might be related to the fact that at $P = 32$ the measurements are more severely undersampled.

In contrast to the walnut case, for the compressed polymer, although the polymer pre-training of the RED denoiser performs the best, the improvement it provides is relatively small, except for a considerable $\sim 20\%$ reduction in MAE for $P = 32$ over the wavelet-based denoiser. Otherwise, the reconstruction metrics for various alternatives are quite similar. We attribute this difference between the two objects to the simplicity of the binary spatial density distribution of the polymer, and the accurate additional prior for such distribution induced by the final sigmoid layer of the NF.

Effect of RED spatial prior on few view static CT

How much improvement does RED spatial prior provide when it is combined with the NF representation? To delve into this question, we conduct few view *static* CT reconstruction experiments with and without the RED prior. These experiments consider the walnut object and the compressed polymer subinterval #6 at $t = 0$. A total number of P projections are obtained with uniform angular spacing in the range $[0, \pi)$. Using the same NF architecture,

Table 5.3: Reconstruction accuracies for RED-NF when $P = 128$ using different denoiser types or pre-training strategies where "walnut" and "polymer" indicate the pre-training datasets for the DnCNN denoiser, and "wavelet" is the wavelet-based unsupervised denoiser [227], [228].

Object	Denoiser Type	PSNR (dB)	SSIM	MAE (1e-3)	HFEN
Walnut P=32	Polymer pre-trained	23.1	0.934	1.5	0.76
	Wavelet-based	24.2	0.941	1.3	0.68
	Walnut pre-trained	**25.9**	**0.962**	**1.1**	**0.51**
Walnut P=128	Polymer pre-trained	23.3	0.953	1.3	1.48
	Wavelet-based	33.1	0.992	0.5	0.28
	Walnut pre-trained	**33.9**	**0.993**	**0.4**	**0.27**
Poly. #2 P=32	Walnut pre-trained	23.2	0.965	12.5	6.79
	Wavelet-based	23.2	0.968	12.0	6.70
	Polymer pre-trained	**23.6**	**0.974**	**9.9**	**6.32**
Poly. #2 P=128	Walnut pre-trained	29.6	0.993	2.8	5.60
	Wavelet-based	29.7	0.993	2.5	5.48
	Polymer pre-trained	**30.2**	**0.995**	**2.1**	**5.23**

the following variational objective is minimized with coordinate inputs $\boldsymbol{\nu} \in \mathbb{R}^2$

$$\arg\min_{\alpha,\bar{f}} \sum_i \|g_{\theta_i} - R_{\theta_i} f^\alpha\|_2^2 + \lambda\rho(\bar{f}) \text{ s.t. } f^\alpha = \bar{f}, \tag{5.11}$$

where $i \in \{1, \ldots, P\}$.

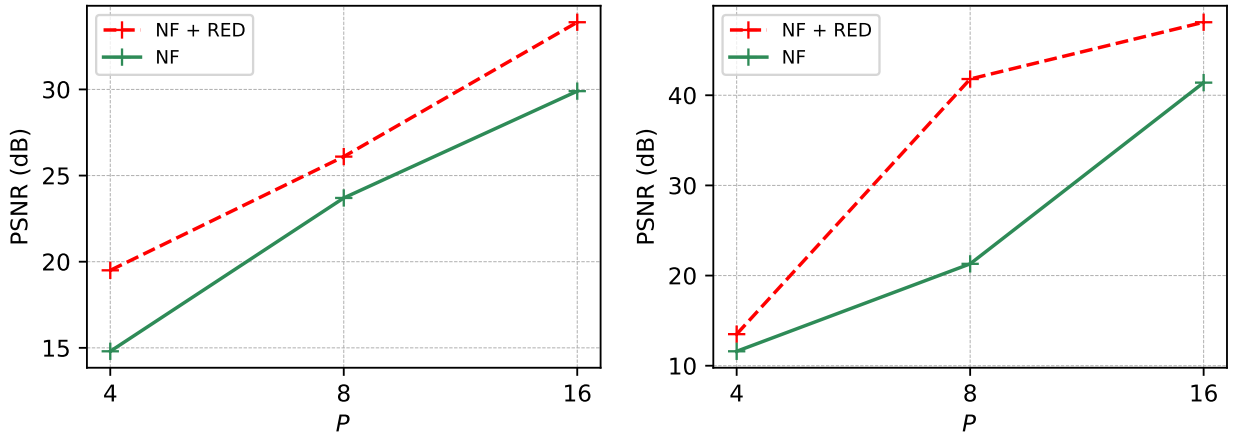
The reconstruction accuracies are shown in Table 5.4 and Figure 5.8. The results show consistent PSNR improvements of up to 5 dB for the walnut object for various P. On the other hand, for the polymer subinterval #6, we see a threshold effect with the increasing number of views. For $P = 4$, neither alternative is able to provide a reasonable reconstruction due to measurement scarcity. For $P = 8$, however, the spatial prior introduced by RED leads to a 20 dB increase in PSNR over the prior introduced by only the NF, which fails to provide an accurate reconstruction. Finally, for $P = 16$, both alternatives perform well due to the presence of sufficient number of projections, but RED-NF still improves on NF by almost 7 dB.

Effect of RED spatial prior and temporal regularization on dynamic CT

Another important question considers the individual contributions of RED spatial regularizer ρ and temporal regularizer ρ_τ in the time-sequential dynamic CT framework. While keeping all other parameters constant, we test three alternatives on the warped walnut object for

Table 5.4: Reconstruction accuracies for the proposed neural field model for the static object at time t with P uniformly spaced views between $[0, \pi)$ for different spatial regularization alternatives.

Object	P	Prior	PSNR (dB)	SSIM	MAE (1e-3)
Walnut	4	None	14.8	0.639	4.1
		RED	**19.5**	**0.825**	**2.7**
	8	None	23.4	0.928	1.5
		RED	**26.1**	**0.958**	**1.0**
	16	None	29.9	0.978	0.7
		RED	**33.9**	**0.987**	**0.4**
Poly. #6	4	None	11.6	0.720	82.1
		RED	**13.5**	**0.742**	**80.9**
	8	None	21.3	0.963	13.5
		RED	**41.8**	**0.987**	**1.8**
	16	None	41.4	0.999	1.1
		RED	**48.1**	**0.999**	**0.5**



(a) Walnut

(b) Polymer subinterval #6

Figure 5.8: Few view static CT reconstruction PSNR values for varying number of uniformly spaced views P for (i) only NF representation where $\lambda = 0$, and (ii) NF representation with RED spatial prior for (a) the walnut and (b) the compressed polymer subinterval #6 at $t = 0$.

varying P in this ablation study: (i) only using the NF representation with $\lambda = \xi = 0$; (ii) NF with temporal regularization only with $\lambda = 0$; and (iii) the proposed RED-NF method.

The results in Figure 5.9 indicate that using the NF by itself exhibits a threshold-like behavior, as it fails to reconstruct the object until the problem is sufficiently measurement-rich with $P = 128$. Note though that this scenario still represents measurement undersampling by a factor of about 128, as only a single projection is available per 128×128 time frame.

Hence, the NF inductive bias is quite strong by itself. Temporal regularization is crucial for the proposed NF structure to work with smaller P. Finally, the RED spatial prior provides a modest PSNR refinement of the results at $P = 32$ or $P = 64$ when the measurements are not critically undersampled, but not at $P = 128$ when the NF alone provides very good recovery. On the other hand, the SSIM improvement introduced by the RED spatial prior increases monotonically with decreasing P.

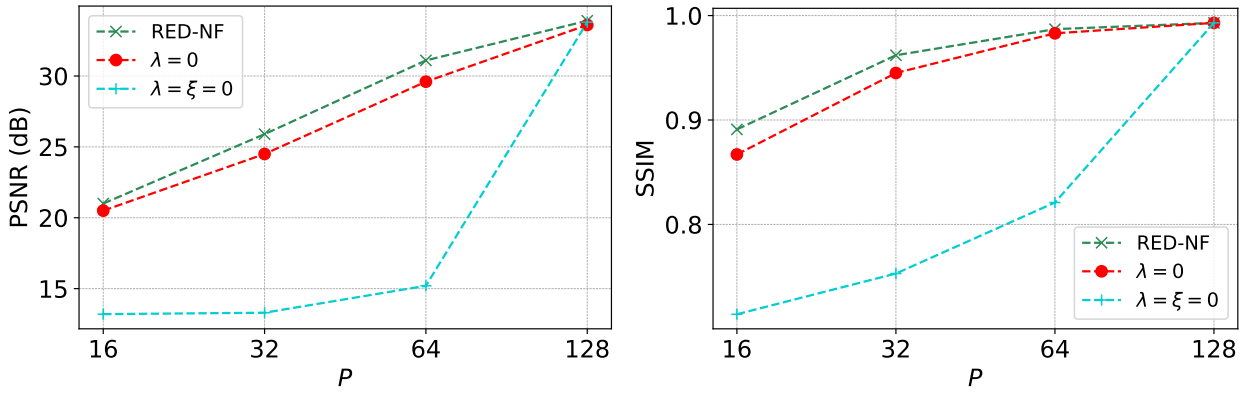


Figure 5.9: Dynamic CT reconstruction PSNR and SSIM values for the warped walnut object for (i) RED-NF, (ii) only temporal regularization with $\lambda = 0$, and (iii) no regularization $\lambda = \xi = 0$ cases for varying number of time-sequential projections P.

5.5 Conclusions

In this work, we proposed RED-NF, the first NF-based dynamic imaging method with pre-learned RED-based spatial prior. The proposed ADMM algorithm for optimization avoids costly gradient computations through the deep denoiser for RED updates. The experiments show significantly better embedding performance for similar number of degrees of freedom compared to a low-rank representation, and improved reconstruction accuracies compared to a recent PSM-based alternative RED-PSM, and a DIP-based method. Ablation studies were performed to quantify the effect of spatial and temporal regularization, and NF inductive bias on dynamic CT, and the effect of RED prior in static CT. The dynamic CT study highlights the necessity of temporal regularization for the method to work when the problem is severely undersampled, and the modest improvement introduced by the RED spatial prior when it is not critically undersampled or sufficiently measurement-rich. On the other hand, the static CT studies demonstrate significant improvements when the RED spatial prior is combined with the NF representation.

Future work may include exploration of different initialization schemes, comparison with

recent NF-based methods proposed for different inverse problems adapted for dynamic tomography , and application of RED-NF to different imaging scenarios.

Supplementary Material

5.5.1 Experimental configurations

The parameter selections for the experiments listed in Tables 5.1 and 5.2 are provided in Table 5.5. Also, the architectural information for the DnCNN denoisers used throughout this work is in Table 5.6.

Table 5.5: The parameter selections for the reconstructions in Tables 5.1 and 5.2. The RED-PSM latent penalty weight was selected as 10^{-1} for the dynamic walnut, and 10^{-3} for the compressed polymer experiments.

		(a) Walnut						(b) Comp. polymer					
P	Method	K	d	λ	$\tilde{\lambda}$	β	ξ	K	d	λ	$\tilde{\lambda}$	β	ξ
32	PSM-TV-S	3	4	5e-2	-	-	-	-	-	-	-	-	-
32	PSM-TV-ST	4	5	5e-2	5e-2	-	-	-	-	-	-	-	-
32	RED-PSM	3	7	1e-4	-	1e-4	-	6	7	2e-4	-	2e-3	-
32	RED-NF	-	-	1e-5	-	5e-6	1e4	-	-	1e-5	-	5e-6	1e2
64	PSM-TV-S	3	4	5e-2	-	-	-	-	-	-	-	-	-
64	PSM-TV-ST	4	5	5e-2	5e-2	-	-	-	-	-	-	-	-
64	RED-PSM	4	7	1e-4	-	5e-4	-	8	9	1e-4	-	4e-3	-
64	RED-NF	-	-	1e-4	-	2.5e-5	1e4	-	-	4e-5	-	4e-5	1e4
128	PSM-TV-S	5	7	5e-2	-	-	-	-	-	-	-	-	-
128	PSM-TV-ST	6	7	5e-2	1e-1	-	-	-	-	-	-	-	-
128	RED-PSM	6	9	2e-4	-	2e-4	-	12	13	2e-4	-	2e-3	-
128	RED-NF	-	-	1e-5	-	5e-6	1e-2	-	-	4e-5	-	8e-5	1e2

Table 5.6: Denoiser DnCNN configurations for different datasets.

Dataset	# of layers	# of channels	Denoising
Walnut	6	64	Direct
Compressed polymer	6	64	Residual

References

[1] I. Danad, J. Szymonifka, J. Schulman-Marcus, and J. K. Min, "Static and dynamic assessment of myocardial perfusion by computed tomography," *European Heart Journal - Cardiovascular Imaging*, vol. 17, no. 8, pp. 836–844, Mar. 2016, ISSN: 2047-2404. DOI: 10.1093/ehjci/jew044. eprint: https://academic.oup.com/ehjcimaging/article-pdf/17/8/836/17465289/jew044.pdf. [Online]. Available: https://doi.org/10.1093/ehjci/jew044.

[2] P. J. Keall *et al.*, "Acquiring 4D thoracic CT scans using a multislice helical method," *Phys. Med. Biol.*, vol. 49, no. 10, p. 2053, 2004.

[3] A. Scanziani, K. Singh, T. Bultreys, B. Bijeljic, and M. J. Blunt, "In situ characterization of immiscible three-phase flow at the pore scale for a water-wet carbonate rock," *Advances in Water Resources*, vol. 121, pp. 446–455, 2018.

[4] J. O'Connor, P. Tofts, K. Miles, L. Parkes, G. Thompson, and A. Jackson, "Dynamic contrast-enhanced imaging techniques: CT and MRI," *The British journal of radiology*, vol. 84, S112–S120, 2011.

[5] B. M. Patterson, N. L. Cordes, K. Henderson, *et al.*, "In situ x-ray synchrotron tomographic imaging during the compression of hyper-elastic polymeric materials," *Journal of materials science*, vol. 51, no. 1, pp. 171–187, 2016.

[6] B. M. Patterson, L. Kuettner, T. Shear, *et al.*, "Synchrotron ct imaging of lattice structures with engineered defects," *Journal of Materials Science*, vol. 55, no. 25, pp. 11 353–11 366, 2020.

[7] E. Maire, C. Le Bourlot, J. Adrien, A. Mortensen, and R. Mokso, "20 Hz X-ray tomography during an in situ tensile test," *International Journal of Fracture*, vol. 200, no. 1, pp. 3–12, 2016.

[8] M. Salerno, B. Sharif, H. Arheden, *et al.*, "Recent advances in cardiovascular magnetic resonance: Techniques and applications," *Circulation: Cardiovascular Imaging*, vol. 10, no. 6, e003951, 2017.

[9] C. Epstein, *Introduction to the Mathematics of Medical Imaging*. Pearson Education/Prentice Hall, 2003, ISBN: 9780130675484. [Online]. Available: https://books.google.com/books?id=vQYZAQAAIAAJ.

[10] X. Hu and T. Parrish, "Reduction of field of view for dynamic imaging," *Magnetic Resonance in Medicine*, vol. 31, no. 6, pp. 691–694, 1994, ISSN: 1522-2594. DOI: 10.1002/mrm.1910310618.

[11] M. E. Brummer *et al.*, "Noquist: Reduced field-of-view imaging by direct Fourier inversion," *Magnetic Resonance in Medicine*, vol. 51, no. 2, pp. 331–342, 2004, ISSN: 1522-2594. DOI: 10.1002/mrm.10694.

[12] B. Madore, G. H. Glover, and N. J. Pelc, "Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 813–828, 1999, ISSN: 1522-2594. DOI: 10.1002/(SICI)1522-2594(199911)42:5<813::AID-MRM1>3.0.CO;2-S.

[13] M. Lustig *et al.*, "K-T SPARSE: High Frame Rate Dynamic MRI Exploiting Spatio-Temporal Sparsity," in *Proceedings of the Annual Meeting of ISMRM*, Seattle, 2006, p. 2420.

[14] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 2, pp. 365–373, 2008, ISSN: 1522-2594. DOI: 10.1002/mrm.21477.

[15] H. Jung, J. C. Ye, and E. Y. Kim, "Improved k–t BLAST and k–t SENSE using FOCUSS," *Physics in Medicine & Biology*, vol. 52, no. 11, p. 3201, May 2007, ISSN: 0031-9155. DOI: 10.1088/0031-9155/52/11/018.

[16] H. Jung *et al.*, "K-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 103–116, 2009, ISSN: 1522-2594. DOI: 10.1002/mrm.21757.

[17] Q. Zhao, N. aggarwal, and Y. Bresler, "Dynamic Imaging of Time-Varying Objects," in *Proc. ISMRM 2001*, Glasgow, Scotland, Apr. 2001, p. 1776.

[18] N. Aggarwal, Q. Zhao, and Y. Bresler, "Spatio-temporal Modeling and Minimum Redundancy Adaptive Acquisition in Dynamic MRI," in *Isbi*, Washington, DC, Jul. 2002, pp. 737–740.

[19] N. Aggarwal and Y. Bresler, "Patient-adapted reconstruction and acquisition dynamic imaging method (PARADIGM) for MRI," *Inverse Problems*, vol. 24, no. 4, pp. 0 450 151–04 501 529, 2008, ISSN: 0266-5611.

[20] B. Sharif *et al.*, "Patient-adaptive reconstruction and acquisition in dynamic imaging with sensitivity encoding (PARADISE)," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 501–513, 2010.

[21] N. P. Willis and Y. Bresler, "Lattice-Theoretic Analysis of Time-Sequential Sampling of Spatio-Temporal Signals Part I," vol. 43, no. 1, pp. 190–207, Jan. 1997.

[22] N. P. Willis and Y. Bresler, "Lattice-Theoretic Analysis of Time-Sequential Sampling of Spatio-Temporal Signals Part II: Large Space-Bandwidth-Product Asymptotics," vol. 43, no. 1, pp. 208–220, Jan. 1997.

[23] Z.-P. Liang and P. Lauterbur, "An efficient method for dynamic magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 677–686, Dec. 1994, ISSN: 1558-254X. DOI: 10.1109/42.363100.

[24] Z.-P. Liang *et al.*, "Dynamic imaging by model estimation," *International Journal of Imaging Systems and Technology*, vol. 8, no. 6, pp. 551–557, 1997, ISSN: 1098-1098. DOI: 10.1002/(SICI)1098-1098(1997)8:6<551::AID-IMA7>3.0.CO;2-9.

[25] J. Tsao, P. Boesiger, and K. P. Pruessmann, "K-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations," *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 1031–1042, 2003, ISSN: 1522-2594. DOI: 10.1002/mrm.10611.

[26] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *2007 4th IEEE international symposium on biomedical imaging: from nano to macro*, IEEE, 2007, pp. 988–991.

[27] H. Pedersen *et al.*, "K-t PCA: Temporally constrained k-t BLAST reconstruction using principal component analysis," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 62, no. 3, pp. 706–716, 2009.

[28] C. Brinegar *et al.*, "Improving temporal resolution of pulmonary perfusion imaging in rats using the partially separable functions model," *Magnetic Resonance in Medicine*, vol. 64, no. 4, pp. 1162–1170, 2010.

[29] J. P. Haldar and Z.-P. Liang, "Low-rank approximations for dynamic imaging," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2011, pp. 1052–1055.

[30] B. Zhao *et al.*, "Further development of image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2011, pp. 1593–1596.

[31] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints," *IEEE transactions on medical imaging*, vol. 31, no. 9, pp. 1809–1820, 2012.

[32] J. P. Haldar and Z.-P. Liang, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Apr. 2010, pp. 716–719. DOI: 10.1109/ISBI.2010.5490076.

[33] B. Zhao *et al.*, "Low rank matrix recovery for real-time cardiac MRI," in *2010 ieee international symposium on biomedical imaging: From nano to macro*, IEEE, 2010, pp. 996–999.

[34] S. Goud, Y. Hu, and M. Jacob, "Real-time cardiac MRI using low-rank and sparsity penalties," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2010, pp. 988–991.

[35] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE transactions on medical imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.

[36] Y. Djebra *et al.*, "Manifold learning via linear tangent space alignment (LTSA) for accelerated dynamic MRI with sparse sampling," *IEEE Transactions on Medical Imaging*, 2022.

[37] S. G. Lingala *et al.*, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: Kt SLR," *IEEE transactions on medical imaging*, vol. 30, no. 5, pp. 1042–1054, 2011.

[38] A. Majumdar and R. K. Ward, "An algorithm for sparse MRI reconstruction by Schatten p-norm minimization," *Magnetic Resonance Imaging*, vol. 29, no. 3, pp. 408–417, Apr. 2011, ISSN: 0730-725X. DOI: 10.1016/j.mri.2010.09.001.

[39] A. Majumdar, R. K. Ward, and T. Aboulnasr, "Non-convex algorithm for sparse and low-rank recovery: Application to dynamic MRI reconstruction," *Magnetic resonance imaging*, vol. 31, no. 3, pp. 448–455, 2013.

[40] A. Majumdar and R. Ward, "Learning space-time dictionaries for blind compressed sensing dynamic MRI reconstruction," in *2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2015, pp. 4550–4554.

[41] J. Trzasko and A. Manduca, "Local versus Global Low-Rank Promotion in Dynamic MRI Series Reconstruction," in *Proc. Int. Soc. Mag. Reson. Med.*, vol. 24, 2011, p. 4371.

[42] J. D. Trzasko, "Exploiting local low-rank structure in higher-dimensional MRI applications," in *Wavelets and Sparsity XV*, vol. 8858, SPIE, Sep. 2013, pp. 551–558. DOI: 10.1117/12.2027059.

[43] S. G. Lingala, E. DiBella, and M. Jacob, "Deformation corrected compressed sensing (DC-CS): A novel framework for accelerated dynamic MRI," *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 72–85, 2014.

[44] X. Chen, M. Salerno, Y. Yang, and F. H. Epstein, "Motion-compensated compressed sensing for dynamic contrast-enhanced mri using regional spatiotemporal sparsity and region tracking: Block low-rank sparsity with motion-guidance (blosm)," *Magnetic resonance in medicine*, vol. 72, no. 4, pp. 1028–1038, 2014.

[45] H. Yoon *et al.*, "Motion adaptive patch-based low-rank approach for compressed sensing cardiac cine MRI," *IEEE Transactions on Medical Imaging*, vol. 33, no. 11, pp. 2069–2085, Nov. 2014, Conference Name: IEEE Transactions on Medical Imaging, ISSN: 1558-254X. DOI: 10.1109/TMI.2014.2330426.

[46] A. Tolouee, J. Alirezaie, and P. Babyn, "Nonrigid motion compensation in compressed sensing reconstruction of cardiac cine MRI," *Magnetic resonance imaging*, vol. 46, pp. 114–120, 2018.

[47] B. Trémoulhéac *et al.*, "Dynamic MR image reconstruction–separation from undersampled (k, t)-space via low-rank plus sparse prior," *IEEE transactions on medical imaging*, vol. 33, no. 8, pp. 1689–1701, 2014.

[48] M. Chiew *et al.*, "K-t FASTER: Acceleration of functional MRI data acquisition using low rank constraints," *Magnetic resonance in medicine*, vol. 74, no. 2, pp. 353–364, 2015.

[49] 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*, vol. 73, no. 3, pp. 1125–1136, 2015.

[50] S. Ravishankar *et al.*, "Low-rank and adaptive sparse signal (LASSI) models for highly accelerated dynamic imaging," *IEEE transactions on medical imaging*, vol. 36, no. 5, pp. 1116–1128, 2017.

[51] S. Babu, S. G. Lingala, and N. Vaswani, "Fast low rank column-wise compressive sensing for accelerated dynamic MRI," *IEEE Transactions on Computational Imaging*, 2023.

[52] J. Schlemper *et al.*, "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, Feb. 2018, ISSN: 1558-254X. DOI: 10.1109/TMI.2017.2760978.

[53] C. Qin *et al.*, "Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 280–290, Jan. 2019, ISSN: 1558-254X. DOI: 10.1109/TMI.2018.2863670.

[54] S. Biswas, H. K. Aggarwal, and M. Jacob, "Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM," *Magnetic Resonance in Medicine*, vol. 82, no. 1, pp. 485–494, 2019, ISSN: 1522-2594. DOI: 10.1002/mrm.27706.

[55] A. Hauptmann *et al.*, "Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease," *Magn. Reson. Med.*, vol. 81, no. 2, pp. 1143–1156, 2019.

[56] T. Küstner *et al.*, "CINENet: Deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions," *Scientific Reports*, vol. 10, no. 1, p. 13 710, 1 Aug. 2020, ISSN: 2045-2322. DOI: 10.1038/s41598-020-70551-8.

[57] Q. Huang *et al.*, "Dynamic MRI reconstruction with end-to-end motion-guided network," *Medical Image Analysis*, vol. 68, p. 101 901, Feb. 2021, ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101901.

[58] Z. Ke *et al.*, "Deep Manifold Learning for Dynamic MR Imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1314–1327, 2021, ISSN: 2333-9403. DOI: 10.1109/TCI.2021.3131564.

[59] Z. Ke *et al.*, "Learned Low-Rank Priors in Dynamic MR Imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3698–3710, Dec. 2021, ISSN: 1558-254X. DOI: 10.1109/TMI.2021.3096218.

[60] C. M. Sandino *et al.*, "Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction," *Magnetic Resonance in Medicine*, vol. 85, no. 1, pp. 152–167, 2021, ISSN: 1522-2594. DOI: 10.1002/mrm.28420.

[61] V. Ghodrati *et al.*, "Temporally aware volumetric generative adversarial network-based MR image reconstruction with simultaneous respiratory motion compensation: Initial feasibility in 3D dynamic cine cardiac MRI," *Magnetic Resonance in Medicine*, vol. 86, no. 5, pp. 2666–2683, 2021, ISSN: 1522-2594. DOI: 10.1002/mrm.28912.

[62] T. Küstner *et al.*, "LAPNet: Non-Rigid Registration Derived in k-Space for Magnetic Resonance Imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3686–3697, Dec. 2021, ISSN: 1558-254X. DOI: 10.1109/TMI.2021.3096131.

[63] H. Qi *et al.*, "End-to-end deep learning nonrigid motion-corrected reconstruction for highly accelerated free-breathing coronary MRA," *Magnetic Resonance in Medicine*, vol. 86, no. 4, pp. 1983–1996, 2021, ISSN: 1522-2594. DOI: 10.1002/mrm.28851.

[64] K. Hammernik *et al.*, "Motion-Guided Physics-Based Learning for Cardiac MRI Reconstruction," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, Oct. 2021, pp. 900–907. DOI: 10.1109/IEEECONF53345.2021.9723134.

[65] S. Wang *et al.*, "DIMENSION: Dynamic MR imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training," *NMR in Biomedicine*, vol. 35, no. 4, e4131, 2022, ISSN: 1099-1492. DOI: 10.1002/nbm.4131.

[66] J. Yoo, K. H. Jin, H. Gupta, J. Yerly, M. Stuber, and M. Unser, "Time-dependent deep image prior for dynamic mri," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3337–3348, 2021.

[67] Q. Zou, A. H. Ahmed, P. Nagpal, S. Kruger, and M. Jacob, "Dynamic imaging using a deep generative storm (gen-storm) model," *IEEE transactions on medical imaging*, vol. 40, no. 11, pp. 3102–3112, 2021.

[68] A. H. Ahmed, Q. Zou, P. Nagpal, and M. Jacob, "Dynamic imaging using deep bi-linear unsupervised representation (deblur)," *IEEE transactions on medical imaging*, vol. 41, no. 10, pp. 2693–2703, 2022.

[69] J. Feng, R. Feng, Q. Wu, Z. Zhang, Y. Zhang, and H. Wei, "Spatiotemporal implicit neural representation for unsupervised dynamic mri reconstruction," *arXiv preprint arXiv:2301.00127*, 2022.

[70] W. Huang, H. B. Li, J. Pan, G. Cruz, D. Rueckert, and K. Hammernik, "Neural implicit k-space for binning-free non-cartesian cardiac mr imaging," in *International Conference on Information Processing in Medical Imaging*, Springer, 2023, pp. 548–560.

[71] V. Spieker, W. Huang, H. Eichhorn, *et al.*, "ICoNIK: generating respiratory-resolved abdominal mr reconstructions using neural implicit representations in k-space," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2023, pp. 183–192.

[72] V. Spieker, H. Eichhorn, J. K. Stelter, *et al.*, "Self-supervised k-space regularization for motion-resolved abdominal mri using neural implicit k-space representation," *arXiv preprint arXiv:2404.08350*, 2024.

[73] J. F. Kunz, S. Ruschke, and R. Heckel, "Implicit neural networks with fourier-feature inputs for free-breathing cardiac mri reconstruction," *arXiv preprint arXiv:2305.06822*, 2023.

[74] M. Tancik, P. Srinivasan, B. Mildenhall, *et al.*, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[75] N. Rahaman, A. Baratin, D. Arpit, *et al.*, "On the spectral bias of neural networks," in *International conference on machine learning*, PMLR, 2019, pp. 5301–5310.

[76] N. P. Willis and Y. Bresler, "Tomographic imaging of time-varying distributions," in *Biomedical Image Processing*, SPIE, vol. 1245, 1990, pp. 111–123.

[77] P. Willis and Y. Bresler, "Optimal scan for time-varying tomography. I. Theoretical analysis and fundamental limitations," *IEEE Transactions on Image Processing*, vol. 4, no. 5, pp. 642–653, 1995. DOI: 10.1109/83.382498.

[78] N. Willis and Y. Bresler, "Optimal scan for time-varying tomography. II. Efficient design and experimental validation," *IEEE Transactions on Image Processing*, vol. 4, no. 5, pp. 654–666, 1995. DOI: 10.1109/83.382499.

[79] C. Jailin and S. Roux, "Dynamic tomographic reconstruction of deforming volumes," *Materials*, vol. 11, no. 8, 2018, ISSN: 1996-1944. DOI: 10.3390/ma11081395.

[80] G. Zang *et al.*, "Space-time tomography for continuously deforming objects," *ACM Transactions on Graphics*, 2018.

[81] C. Jailin *et al.*, "Projection-based dynamic tomography," *Phys. Med. Biol.*, vol. 66, no. 21, p. 215 018, 2021.

[82] S. Capostagno *et al.*, "Deformable motion compensation for interventional cone-beam CT," *Phys. Med. Biol.*, vol. 66, no. 5, p. 055 010, Feb. 2021, ISSN: 0031-9155. DOI: 10.1088/1361-6560/abb16e. (visited on 01/22/2024).

[83] 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*, vol. 33, no. 12, p. 124 008, 2017.

[84] B. N. Hahn *et al.*, "Using the Navier-Cauchy equation for motion estimation in dynamic imaging," *Inverse Problems and Imaging*, vol. 16, no. 5, p. 1179, 2022. DOI: 10.3934/ipi.2022018.

[85] K. Park, U. Sinha, P. Hedman, *et al.*, "Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields," *ACM Trans. Graph.*, vol. 40, no. 6, Dec. 2021.

[86] L. Brombal *et al.*, "Motion artifacts assessment and correction using optical tracking in synchrotron radiation breast CT," *Medical Physics*, vol. 48, no. 9, pp. 5343–5355, 2021. DOI: 10.1002/mp.15084.

[87] A. K. George *et al.*, "Time-resolved cardiac CT reconstruction using the ensemble Kalman Filter," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, May 2008, pp. 1489–1492. DOI: 10.1109/ISBI.2008.4541290. (visited on 01/22/2024).

[88] M. D. Butala *et al.*, "Tomographic Imaging of Dynamic Objects With the Ensemble Kalman Filter," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1573–1587, Jul. 2009, ISSN: 1941-0042. DOI: 10.1109/TIP.2009.2017996. (visited on 01/22/2024).

[89] J. Hakkarainen *et al.*, "Undersampled Dynamic X-Ray Tomography With Dimension Reduction Kalman Filter," *IEEE Transactions on Computational Imaging*, vol. 5, no. 3, pp. 492–501, Sep. 2019, ISSN: 2333-9403. DOI: 10.1109/TCI.2019.2896527. (visited on 01/22/2024).

[90] X. Cao *et al.*, "Undersampled dynamic tomography with separated spatial and temporal regularization," in *7th International Conference on Image Formation in X-Ray Computed Tomography*, vol. 12304, SPIE, Oct. 2022, pp. 288–295. DOI: 10.1117/12.2647508. (visited on 01/22/2024).

[91] K. A. Mohan *et al.*, "TIMBIR: A method for time-space reconstruction from interlaced views," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 96–111, 2015.

[92] F. Lucka *et al.*, "Enhancing Compressed Sensing 4D Photoacoustic Tomography by Simultaneous Motion Estimation," *SIAM J. Imaging Sci.*, vol. 11, no. 4, pp. 2224–2253, Jan. 2018. DOI: 10.1137/18M1170066. (visited on 01/22/2024).

[93] T. A. Bubba *et al.*, "Sparse dynamic tomography: A shearlet-based approach for iodine perfusion in plant stems," *Inverse Problems*, vol. 36, no. 9, p. 094 002, Sep. 2020, ISSN: 0266-5611. DOI: 10.1088/1361-6420/ab9c15.

[94] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, "Robust principal component analysis-based four-dimensional computed tomography," *Physics in Medicine & Biology*, vol. 56, no. 11, p. 3181, 2011.

[95] B. Iskender *et al.*, "Dynamic tomography reconstruction by projection-domain separable modeling," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, IEEE, 2022.

[96] B. Iskender *et al.*, "Dynamic tomography reconstruction by projection-domain separable modeling," *arXiv:2204.09935*, 2022.

[97] B. Iskender *et al.*, "Factorized projection-domain spatio-temporal regularization for dynamic tomography," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2023.

[98] S. Zhi *et al.*, "Artifacts reduction method for phase-resolved Cone-Beam CT (CBCT) images via a prior-guided CNN," in *Medical Imaging 2019: Physics of Medical Imaging*, vol. 10948, SPIE, Mar. 2019, pp. 562–567. DOI: 10.1117/12.2513128. (visited on 01/24/2024).

[99] X. Huang *et al.*, "U-net-based deformation vector field estimation for motion-compensated 4D-CBCT reconstruction," *Medical Physics*, vol. 47, no. 7, pp. 3000–3012, 2020, ISSN: 2473-4209. DOI: 10.1002/mp.14150. (visited on 01/25/2024).

[100] F. Madesta *et al.*, "Self-contained deep learning-based boosting of 4D cone-beam CT reconstruction," *Medical Physics*, vol. 47, no. 11, pp. 5619–5631, 2020, ISSN: 2473-4209. DOI: 10.1002/mp.14441. (visited on 01/25/2024).

[101] Z. Zhang *et al.*, "Deep learning-based motion compensation for four-dimensional cone-beam computed tomography (4D-CBCT) reconstruction," *Medical Physics*, vol. 50, no. 2, pp. 808–820, 2023, ISSN: 2473-4209. DOI: 10.1002/mp.16103. (visited on 01/22/2024).

[102] S. Majee *et al.*, "Multi-slice fusion for sparse-view and limited-angle 4D CT reconstruction," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 448–462, 2021.

[103] G. T. Buzzard *et al.*, "Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium," *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2001–2020, 2018.

[104] A. W. Reed, H. Kim, R. Anirudh, *et al.*, "Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2258–2268.

[105] K. Park, U. Sinha, J. T. Barron, *et al.*, "Nerfies: Deformable neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5865–5874.

[106] L. Lozenski, M. A. Anastasio, and U. Villa, "A memory-efficient self-supervised dynamic image reconstruction method using neural fields," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 879–892, 2022.

[107] K. Lee, N. A. Trask, R. G. Patel, M. A. Gulian, and E. C. Cyr, "Partition of unity networks: Deep hp-approximation," *arXiv preprint arXiv:2101.11256*, 2021.

[108] L. Lozenski, R. M. Cam, M. A. Anastasio, and U. Villa, "Proxnf: Neural field proximal training for high-resolution 4d dynamic image reconstruction," *arXiv preprint arXiv:2403.03860*, 2024.

[109] Y. Zhang, H.-C. Shao, T. Pan, and T. Mengke, "Dynamic cone-beam CT reconstruction using spatial and temporal implicit neural representation learning (STINR)," *Physics in Medicine & Biology*, vol. 68, no. 4, p. 045 005, 2023.

[110] H.-C. Shao, T. Mengke, T. Pan, and Y. Zhang, "Dynamic cbct imaging using prior model-free spatiotemporal implicit neural representation (pmf-stinr)," *ArXiv*, 2023.

[111] L. Birklein, E. Schömer, R. Brylka, U. Schwanecke, and R. K. W. Schulze, "Neural deformable cone beam CT," Eurographics, 2023.

[112] R. Hyder and M. S. Asif, "Generative models for low-dimensional video representation and reconstruction," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1688–1701, 2020, ISSN: 1941-0476. DOI: 10.1109/TSP.2020.2977256.

[113] In *Methods of Modern Mathematical Physics*, M. Reed and B. Simon, Eds., Academic Press, 1972, ISBN: 978-0-12-585001-8. DOI: https://doi.org/10.1016/B978-0-12-585001-8.50001-5.

[114] S. Ma, H. Du, Q. Wu, and W. Mei, "Dynamic MRI reconstruction exploiting partial separability and t-SVD," in *2019 IEEE 7th International Conference on Bioinformatics and Computational Biology (ICBCB)*, IEEE, 2019, pp. 179–184.

[115] S. Ma, Y. Fan, and Z. Li, "Dynamic MRI exploiting partial separability and shift invariant discrete wavelet transform," in *2021 6th International Conference on Image, Vision and Computing (ICIVC)*, IEEE, 2021, pp. 242–246.

[116] V. Slyusar, "A family of face products of matrices and its properties," *Cybernetics and Systems Analysis*, vol. 35, no. 3, pp. 379–384, 1999.

[117] B. Iskender and Y. Bresler, "X-ray CT scatter correction by a physics-motivated dnn with opposite view processing," in *Proc. 6th Int. Conf. Image Formation X-Ray Computed Tomogr.*, 2020, pp. 308–311.

[118] B. Iskender, "X-ray CT scatter correction by a physics-motivated deep neural network," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2020.

[119] B. Iskender and Y. Bresler, "Scatter correction in X-ray CT by physics-inspired deep learning," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 1074–1088, 2022.

[120] Y. Li, K. Lee, and Y. Bresler, "Identifiability in blind deconvolution with subspace or sparsity constraints," *IEEE Transactions on information Theory*, vol. 62, no. 7, pp. 4266–4275, 2016.

[121] R. W. Chan, E. A. Ramsay, E. Y. Cheung, and D. B. Plewes, "The influence of radial undersampling schemes on compressed sensing reconstruction in breast MRI," *Magnetic resonance in medicine*, vol. 67, no. 2, pp. 363–377, 2012.

[122] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on numerical analysis*, vol. 10, no. 2, pp. 413–432, 1973.

[123] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.

[124] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[125] R. Hyder, *Generative Models for Low-Rank Video Representation and Reconstruction*, https://github.com/CSIPlab/GMLR, 2019.

[126] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[127] H. Der Sarkissian, F. Lucka, M. van Eijnatten, G. Colacicco, S. B. Coban, and K. J. Batenburg, "A cone-beam x-ray computed tomography data collection designed for machine learning," *Scientific data*, vol. 6, no. 1, p. 215, 2019.

[128] *Piecewise Affine Transformation*, https://scikit-image.org/docs/dev/auto_examples/transform/plot_piecewise_affine.html.

[129] Y. A. Brychkov, *Multidimensional integral transformations*. CRC Press, 1992.

[130] L. Pfister, R. Bhargava, Y. Bresler, and P. S. Carney, "Composition-aware spectroscopic tomography," *Inverse Problems*, vol. 36, no. 11, p. 115 010, 2020.

[131] G. Harikumar and Y. Bresler, "Fir perfect signal reconstruction from multiple convolutions: Minimum deconvolver orders," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 215–218, 1998.

[132] G. Harikumar and Y. Bresler, "Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 202–219, 1999.

[133] T. Jiang, N. D. Sidiropoulos, and J. M. Ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1849–1859, 2001.

[134] V. Slyusar, "New operations of matrix products for application of radars," in *IEEE MTT/ED/AP West Ukraine Chapter DIPED - 97. Direct and Inverse Problems of Electromagnetic and Acoustic Theory (IEEE Cat. No.97TH8343)*, 1997, pp. 73–74. DOI: 10.1109/DIPED.1997.710918.

[135] B. Alexeev, J. Cahill, and D. G. Mixon, "Full spark frames," *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, pp. 1167–1194, 2012.

[136] D. Robinson and P. Milanfar, "Fast local and global projection-based methods for affine motion estimation," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 35–54, 2003.

[137] X. Xiong and K. Qin, "Linearly estimating all parameters of affine motion using radon transform," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4311–4321, 2014.

[138] A. H. Ahmed, P. Nagpal, S. Kruger, and M. Jacob, "Dynamic imaging using deep bilinear unsupervised learning (deblur)," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2021, pp. 1099–1102.

[139] A. for the review process, "Personal communication," 2022.

[140] F. Reda, J. Kontkanen, E. Tabellion, D. Sun, C. Pantofaru, and B. Curless, "Film: Frame interpolation for large motion," *arXiv preprint arXiv:2202.04901*, 2022.

[141] B. Iskender, M. L. Klasky, and Y. Bresler, "RED-PSM: Regularization by denoising of partially separable models for dynamic imaging," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2023, pp. 10 595–10 604.

[142] B. Iskender, M. L. Klasky, and Y. Bresler, "RED-PSM: Regularization by denoising of factorized low rank models for dynamic imaging," *arXiv preprint arXiv:2304.03483*, 2023.

[143] B. Iskender, M. L. Klasky, and Y. Bresler, "Dynamic tomography reconstruction via low-rank modeling with a red spatial prior," in *2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, IEEE, 2023, pp. 506–510.

[144] B. Iskender, M. L. Klasky, and Y. Bresler, "RED-PSM: Regularization by denoising of factorized low rank models for dynamic imaging," *IEEE Transactions on Computational Imaging*, pp. 1–16, 2024. DOI: 10.1109/TCI.2024.3402347.

[145] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[146] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948. DOI: 10.1109/GlobalSIP.2013.6737048.

[147] U. S. Kamilov *et al.*, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 85–97, 2023.

[148] X. Xu *et al.*, "Provable convergence of plug-and-play priors with MMSE denoisers," *IEEE Signal Processing Letters*, vol. 27, pp. 1280–1284, 2020.

[149] R. Cohen, M. Elad, and P. Milanfar, "Regularization by denoising via fixed-point projection (RED-PRO)," *SIAM Journal on Imaging Sciences*, vol. 14, no. 3, pp. 1374–1406, 2021.

[150] D. Xue *et al.*, "On the convergence of non-convex phase retrieval with denoising priors," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4424–4439, 2022, ISSN: 1941-0476. DOI: 10.1109/TSP.2022.3203228. (visited on 11/18/2023).

[151] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE transactions on computational imaging*, vol. 5, no. 1, pp. 52–67, 2018.

[152] X. Shi *et al.*, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.

[153] L. Wang *et al.*, "Spatial-temporal transformer for video snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[154] N. Ginio *et al.*, "Efficient machine learning method for spatio-temporal water surface waves reconstruction from polarimetric images," *Measurement Science and Technology*, vol. 34, no. 5, p. 055 801, 2023.

[155] Y. Sun, J. Liu, and U. Kamilov, "Block coordinate regularization by denoising," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[156] C. Metzler *et al.*, "prDeep: robust phase retrieval with a flexible deep network," in *International Conference on Machine Learning*, PMLR, 2018, pp. 3501–3510.

[157] G. Mataev, P. Milanfar, and M. Elad, "DeepRED: Deep image prior powered by RED," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[158] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[159] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & mathematics with applications*, vol. 2, no. 1, pp. 17–40, 1976.

[160] J. Eckstein and D. P. Bertsekas, "On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.

[161] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[162] D. Hajinezhad and Q. Shi, "Alternating direction method of multipliers for a class of nonconvex bilinear optimization: Convergence analysis and applications," *Journal of Global Optimization*, vol. 70, pp. 261–288, 2018.

[163] Y. Wang, W. Yin, and J. Zeng, "Global convergence of admm in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, pp. 29–63, 2019.

[164] R. Cohen *et al.*, "It has potential: Gradient-driven denoisers for convergent solutions to inverse problems," in *Advances in Neural Information Processing Systems*, vol. 34, Curran Associates, Inc., 2021, pp. 18 152–18 164. (visited on 11/18/2023).

[165] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient step denoiser for convergent Plug-and-Play," in *International Conference on Learning Representations, ICLR 2022*, 2021.

[166] R. Fermanian, M. Le Pendu, and C. Guillemot, "PnP-ReG: Learned regularizing gradient for plug-and-play gradient descent," *SIAM Journal on Imaging Sciences*, vol. 16, no. 2, pp. 585–613, 2023. DOI: 10.1137/22M1490843. eprint: https://doi.org/10.1137/22M1490843.

[167] A. Berk *et al.*, "Deep proximal gradient method for learned convex regularizers," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5. DOI: 10.1109/ICASSP49357.2023.10094632. (visited on 11/18/2023).

[168] J. R. Chand and M. Jacob, "Multi-scale energy (MuSE) plug and play framework for inverse problems," *arXiv:2305.04775*, 2023.

[169] M. Udell *et al.*, "Generalized low rank models," *Foundations and Trends® in Machine Learning*, vol. 9, no. 1, pp. 1–118, 2016.

[170] J. D. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 713–719.

[171] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[172] H. Kim, J. Kang, K. Champley, R. Anirudh, and A. Mohan, "LLNL D4DCT datasets: Dynamic 4DCT datasets using mpm-based deformation,"

[173] A. Sadeghirad, R. M. Brannon, and J. Burghardt, "A convected particle domain interpolation technique to extend applicability of the material point method for problems involving massive deformations," *International Journal for numerical methods in Engineering*, vol. 86, no. 12, pp. 1435–1456, 2011.

[174] A. Sadeghirad, R. M. Brannon, and J. Guilkey, "Second-order convected particle domain interpolation (CPDI2) with enrichment for weak discontinuities at material interfaces," *International Journal for numerical methods in Engineering*, vol. 95, no. 11, pp. 928–952, 2013.

[175] J. Yoo, *Time-Dependent Deep Image Prior for Dynamic MRI*, https://github.com/jaejun-yoo/TDDIP, 2021.

[176] O. Bernard *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, 2018.

[177] S. Kulkarni, J. A. Rumberger, and S. Jha, "Electron beam CT: A historical review," *American Journal of Roentgenology*, vol. 216, no. 5, pp. 1222–1228, May 2021, Publisher: American Roentgen Ray Society, ISSN: 0361-803X. DOI: 10.2214/AJR.19.22681. (visited on 11/17/2023).

[178] D. Spronk *et al.*, "Evaluation of carbon nanotube x-ray source array for stationary head computed tomography," *Med. Phys.*, vol. 48, no. 3, pp. 1089–1099, Mar. 2021, ISSN: 2473-4209. DOI: 10.1002/mp.14696.

[179] S. Xu *et al.*, "Volumetric computed tomography with carbon nanotube x-ray source array for improved image quality and accuracy," *Commun Eng*, vol. 2, no. 1, pp. 1–9, Oct. 7, 2023, Number: 1 Publisher: Nature Publishing Group, ISSN: 2731-3395. DOI: 10.1038/s44172-023-00123-x. (visited on 11/17/2023).

[180] 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*, vol. 13, no. 4, pp. 600–612, 2004.

[181] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE transactions on medical imaging*, vol. 30, no. 5, pp. 1028–1041, 2010.

[182] K. Dabov *et al.*, "Bm3d image denoising with shape-adaptive principal component analysis," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[183] S. Gu *et al.*, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2862–2869.

[184] M. H. Alkinani and M. R. El-Sakka, "Patch-based models and algorithms for image denoising: A comparative review between patch-based images denoising methods for additive noise reduction," *EURASIP journal on image and video processing*, vol. 2017, pp. 1–27, 2017.

[185] S. Liang, B. Iskender, B. Wen, and S. Ravishankar, "LABMAT: learned feature-domain block matching for image restoration," in *2021 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2021, pp. 1689–1693.

[186] S. Xiao, Y. Bresler, and D. C. Munson, "O (N/sup 2/log N) native fan-beam tomographic reconstruction," in *Proceedings IEEE International Symposium on Biomedical Imaging*, IEEE, 2002, pp. 824–827.

[187] Y. Bresler and J. Brokish, "Fast hierarchical backprojection for helical cone-beam tomography," in *Proceedings 2003 International Conference on Image Processing*, IEEE, vol. 2, 2003, pp. II–815.

[188] S. Zlobec, "On the liu–floudas convexification of smooth programs," *Journal of Global Optimization*, vol. 32, no. 3, pp. 401–407, 2005.

[189] W. Xian, J.-B. Huang, J. Kopf, and C. Kim, "Space-time neural irradiance fields for free-viewpoint video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9421–9431.

[190] Z. Li, S. Niklaus, N. Snavely, and O. Wang, "Neural scene flow fields for space-time view synthesis of dynamic scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6498–6508.

[191] Y. Du, Y. Zhang, H.-X. Yu, J. B. Tenenbaum, and J. Wu, "Neural radiance flow for 4d view synthesis and video processing," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE Computer Society, 2021, pp. 14 304–14 314.

[192] C. Gao, A. Saraf, J. Kopf, and J.-B. Huang, "Dynamic view synthesis from dynamic monocular video," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021.

[193] C. Wang, B. Eckart, S. Lucey, and O. Gallo, "Neural trajectory fields for dynamic novel view synthesis," *arXiv preprint arXiv:2105.05994*, 2021.

[194] Y. Xie, T. Takikawa, S. Saito, *et al.*, "Neural fields in visual computing and beyond," in *Computer Graphics Forum*, Wiley Online Library, vol. 41, 2022, pp. 641–676.

[195] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *arXiv*, 2020.

[196] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, and Y. Sheikh, "Neural volumes: Learning dynamic renderable volumes from images," *arXiv preprint arXiv:1906.07751*, 2019.

[197] L. Yariv, Y. Kasten, D. Moran, *et al.*, "Multiview neural surface reconstruction by disentangling geometry and appearance," *Advances in Neural Information Processing Systems*, vol. 33, pp. 2492–2502, 2020.

[198] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[199] L. Shen, J. Pauly, and L. Xing, "Nerp: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[200] E. Tretschk, A. Tewari, V. Golyanik, M. Zollhöfer, C. Lassner, and C. Theobalt, "Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 959–12 970.

[201] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-nerf: Neural radiance fields for dynamic scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 318–10 327.

[202] S. Pan, S. L. Brunton, and J. N. Kutz, "Neural implicit flow: A mesh-agnostic dimensionality reduction paradigm of spatio-temporal data," *Journal of Machine Learning Research*, vol. 24, no. 41, pp. 1–60, 2023.

[203] Y. Sun, J. Liu, M. Xie, B. Wohlberg, and U. S. Kamilov, "Coil: Coordinate-based internal learning for tomographic imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1400–1412, 2021.

[204] G. Zang, R. Idoughi, R. Li, P. Wonka, and W. Heidrich, "Intratomo: Self-supervised learning-based tomography via sinogram synthesis and prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1960–1970.

[205] B. Song, L. Shen, and L. Xing, "Piner: Prior-informed implicit neural representation learning for test-time adaptation in sparse-view ct reconstruction," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 1928–1938.

[206] E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis, "Cryodrgn: Reconstruction of heterogeneous cryo-em structures using neural networks," *Nature methods*, vol. 18, no. 2, pp. 176–185, 2021.

[207] S. Gupta, K. Kothari, V. Debarnot, and I. Dokmanić, "Differentiable uncalibrated imaging," *IEEE Transactions on Computational Imaging*, 2023.

[208] X. Wang, S. Hu, H. Fan, H. Zhu, and X. Li, "Neural radiance fields in medical imaging: Challenges and next steps," *arXiv preprint arXiv:2402.17797*, 2024.

[209] E. D. Zhong, T. Bepler, J. H. Davis, and B. Berger, "Reconstructing continuous distributions of 3d protein structure from cryo-em images," *arXiv preprint arXiv:1909.05215*, 2019.

[210] P.-S. Wang, Y. Liu, Y.-Q. Yang, and X. Tong, "Spline positional encoding for learning 3d implicit signed distance fields," *arXiv preprint arXiv:2106.01553*, 2021.

[211] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, "Wire: Wavelet implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 507–18 516.

[212] T. M. Roddenberry, V. Saragadam, M. V. de Hoop, and R. G. Baraniuk, "Implicit neural representations and the algebra of complex wavelets," *arXiv preprint arXiv:2310.00545*, 2023.

[213] G. Yüce, G. Ortiz-Jiménez, B. Besbinar, and P. Frossard, "A structured dictionary perspective on implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19 228–19 238.

[214] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5855–5864.

[215] S. J. Garbin, M. Kowalski, M. Johnson, J. Shotton, and J. Valentin, "Fastnerf: High-fidelity neural rendering at 200fps," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 14 346–14 355.

[216] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "Tensorf: Tensorial radiance fields," in *European Conference on Computer Vision*, Springer, 2022, pp. 333–350.

[217] Y. Wang, Q. Han, M. Habermann, K. Daniilidis, C. Theobalt, and L. Liu, "Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3295–3306.

[218] M. Yang, Y. Wen, W. Chen, Y. Chen, and K. Jia, "Deep optimized priors for 3d shape modeling and reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3269–3278.

[219] W. Jang and L. Agapito, "Codenerf: Disentangled neural radiance fields for object categories," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 949–12 958.

[220] J. Mu, W. Qiu, A. Kortylewski, A. Yuille, N. Vasconcelos, and X. Wang, "A-sdf: Learning disentangled signed distance functions for articulated shape representation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 001–13 011.

[221] V. Sitzmann, S. Rezchikov, B. Freeman, J. Tenenbaum, and F. Durand, "Light field networks: Neural scene representations with single-evaluation rendering," *Advances in Neural Information Processing Systems*, vol. 34, pp. 19 313–19 325, 2021.

[222] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 9, pp. 5149–5169, 2021.

[223] V. Sitzmann, E. Chan, R. Tucker, N. Snavely, and G. Wetzstein, "Metasdf: Meta-learning signed distance functions," *Advances in Neural Information Processing Systems*, vol. 33, pp. 10 136–10 147, 2020.

[224] M. Tancik, B. Mildenhall, T. Wang, *et al.*, "Learned initializations for optimizing coordinate-based neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2846–2855.

[225] J. Lee, J. Tack, N. Lee, and J. Shin, "Meta-learning sparse implicit neural representations," *Advances in Neural Information Processing Systems*, vol. 34, pp. 11 769–11 780, 2021.

[226] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, IEEE, 2013, pp. 945–948.

[227] *Wavelet Denoising*, https://scikit-image.org/docs/stable/auto_examples/filters/plot_denoise_wavelet.html.

[228] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE transactions on image processing*, vol. 9, no. 9, pp. 1532–1546, 2000.