Enhancing Compressed Sensing 4D Photoacoustic Tomography by Simultaneous Motion Estimation[*]

Felix Lucka[†], Nam Huynh[‡], Marta Betcke[§], Edward Zhang[‡], Paul Beard[‡], Ben Cox[‡], and Simon Arridge[§]

Abstract. A crucial limitation of current high-resolution 3D photoacoustic tomography (PAT) devices that employ sequential scanning is their long acquisition time. In previous work, we demonstrated how to use compressed sensing techniques to improve upon this: images with good spatial resolution and contrast can be obtained from suitably subsampled PAT data acquired by novel acoustic scanning systems if sparsity-constrained image reconstruction techniques such as total variation regularization are used. Now, we show how a further increase of image quality can be achieved for imaging dynamic processes in living tissue (4D PAT). The key idea is to exploit the additional temporal redundancy of the data by coupling the previously used spatial image reconstruction models with sparsity-constrained motion estimation models. While simulated data from a 2D numerical phantom will be used to illustrate the main properties of this recently developed joint-image-reconstruction-and-motion-estimation framework, measured data from a dynamic experimental phantom will also be used to demonstrate its potential for challenging, large-scale, real-world, 3D scenarios. The latter only becomes feasible if a carefully designed combination of tailored optimization schemes is employed, which we describe and examine in more detail.

Key words. photoacoustic tomography, dynamic imaging, compressed sensing, simultaneous motion estimation, variational regularization

AMS subject classifications. 92C55, 65R32, 94A08, 94A12, 65K10

DOI. 10.1137/18M1170066

1. Introduction.

1.1. Compressed sensing photoacoustic tomography. Optical absorption of biological tissues is a desirable source of image contrast for a variety of clinical and preclinical applica-

[†]Computational Imaging, Centrum Wiskunde & Informatica (CWI), 1098 XG Amsterdam, The Netherlands, and Department of Computer Science, University College London, WC1E 6BT London, UK (Felix.Lucka@cwi.nl, http://felixlucka.github.io/).

[‡]Department of Medical Physics and Bioengineering, University College London, WC1E 6BT London, UK (n.huynh@ucl.ac.uk, e.zhang@ucl.ac.uk, paul.beard@ucl.ac.uk, b.cox@ucl.ac.uk).

[§]Department of Computer Science, University College London, WC1E 6BT London, UK (m.betcke@ucl.ac.uk, s.arridge@ucl.ac.uk).

tions. In particular, its wavelength dependence provides spectroscopic (chemical) information on the absorbing molecules (chromophores). *Photoacoustic tomography* (PAT) is an "imaging from coupled physics" technique [3] that employs laser-generated ultrasound (US) to obtain optical absorption images with the high spatial resolution of US. For recent reviews on the physical principles, technical realizations, and (pre)clinical applications of PAT, we refer the reader to [63, 4, 48, 66].

In [1], we discussed the particular challenges of acquiring high quality three-dimensional (3D) photoacoustic (PA) images with sequential scanning schemes, such as the *Fabry–Pérot-based PA scanner* (FP scanner): To reach a spatial resolution less than one hundred μm, acoustic waves containing frequencies up to a few tens of MHz have to be sampled over centimeter-scale apertures. For a scanning pattern to satisfy the spatial Nyquist criterion, sampling intervals in the order of tens of μm have to be chosen, which leads to several thousand detection points, and thereby long acquisition times. This imposes a severe limit for dynamic PAT (4D PAT), i.e., imaging dynamic anatomical and physiological events in high resolution in real time, an area of research of increasing interest [19]. The key observation to overcome this limitation is that the Nyquist criterion is often too conservative because it guarantees perfect recovery of the broad class of images that are band limited but otherwise arbitrary. However, images of absorbing tissue structures come from a much smaller subclass of images, as they typically also have a rather low spatial complexity (or a high *sparsity*). Therefore, data recorded in a conventional, regularly sampled fashion, satisfying the Nyquist criterion, is often highly redundant. *Compressed sensing* (CS) [12, 22, 28] techniques exploit this fact by combining subsampling schemes that try to maximize the nonredundancy of the data with image reconstruction approaches that employ sparsity constraints. In [1], we demonstrated the implementation of CS techniques to accelerate 3D PAT acquisition by using spatial sparsity constraints. In the context of 4D PAT, such techniques can be employed to reconstruct each temporal frame separately, i.e., as a *frame-by-frame* (fbf) image reconstruction method.

1.2. Spatio-temporal image reconstruction. In this work, we show that another significant acceleration can be obtained by also accounting for the temporal evolution of the target within a full spatio-temporal reconstruction scheme. A wide range of such approaches have been proposed for different applications and dynamics. If the dynamics between separate frames are sufficiently simple (e.g., affine deformations), low-dimensional parametric models can often be used to efficiently constrain the image reconstruction in time. An application to PAT is demonstrated in [15], and theoretical analysis of such approaches can be found in [35, 34, 33]. In such situations, the aim is often rather to compensate for the motion (see, e.g., [45] for an overview on compensating for respiratory motion) than to resolve it, which is our main aim here. Several approaches rely on extending popular spatial constraints into time. Incorporating ℓ_2 regularization of the temporal differences between frames is examined in [55, 56], and recently extending ℓ_1 functionals such as total variation functional and its higher order variants to spatio-temporal settings have been proposed and shown to work well for certain dynamics; see, e.g., [37, 54]. In the Bayesian approach to inverse imaging problems, spatio-temporal methods are commonly referred to as *Kalman filtering or smoothing*: Filtering refers to reconstructing each image frame based only on measured data up to that point in time, most often done via updating the previous image frame based on the most recent

Table 1
List of commonly occurring abbreviations.

Abbreviation	Meaning	Reference
ACS	alternate convex search	sect. 4.2
ADMM	alternating direction method of multipliers	sect. 4.4, Alg. 4.2
fbf	frame-by-frame	sect. 2.2
FP	Fabry–Pérot	sect. 1.1
mIP	maximum intensity projection	Fig. 2
NNLS	nonnegative least squares	sect. 5.1
(Q)PAT	(quantitative) photoacoustic tomography	sect. 1
PDHG	primal dual hybrid gradient	sect. 4.3, Alg. 4.1
TV	total variation regularization	sect. 3.2
TVTVL2	Joint image reconstruction and motion estimation approach	sect. 3.2, (3.5)

data. While this is the only option for real-time or *online* image reconstruction, it is also popular in *offline* image reconstruction due to its lower computational complexity compared to smoothing, which refers to estimating each image frame based on the whole set of measured data. See section 4 in [42] for a general introduction and further references to Kalman filtering and [57] for recent work on this topic. In the context of compressed sensing applications, low-rank-type models have been examined extensively; see, e.g., [36, 62, 60, 52]. These models rely on strong spatio-temporal decomposition assumptions which are very effective when fulfilled but not appropriate for all dynamics.

In this work, we adopt a very general spatio-temporal modeling framework introduced in [10] that can encode a priori information about a wide range of dynamics: It formulates an explicit PDE model for the image dynamics and then jointly estimates the image sequence and the corresponding motion field by minimizing a variational energy. An overview of similar approaches to joint image reconstruction and motion estimation can be found in the introduction of [10], which also contains theoretical analysis of this approach. While it was used for 2D dynamic computed tomography reconstruction in [9], we present the first application to a challenging, large-scale 3D dynamic problem with experimental data, which also requires the development of tailored numerical optimization schemes.

1.3. Structure. The remainder of the paper is organized as follows: Section 2 introduces the mathematical modeling of dynamic PAT and illustrates the limitations of reconstruction approaches that only account for spatial sparsity. Based on this, a variational spatio-temporal image reconstruction framework based on joint motion estimation is presented in section 3. Section 4 discusses the numerical solution of the optimization problems that originate from the variational approach, and in section 5 we present results with a simple 2D scenario with simulated data and a challenging 3D scenario with experimental data. Finally, we discuss the results of our work and point to future directions of research in section 6. Table 1 lists all commonly occurring abbreviations for reference.

2. Background and previous work.

2.1. Sequential acquisition of compressed dynamic PAT. Let us denote the biological tissue to be imaged by $\Omega \in \mathbb{R}^d$ ($d = 2, 3$), the space variable by $r \in \Omega$, the measurement interval by $[0, \mathcal{T}]$, and the (continuous) time variable by $\tau \in [0, \mathcal{T}]$. A reasonable mathemat-

ical model of dynamic PAT has to make certain assumptions about the different time scales involved in signal generation and measurement, in particular if the PA signal is scanned in a sequential manner. First, as described in more detail in section 1.1 of [2], the PA effect is only significant if the laser pulse duration, photon transport, photon absorption by chromophores, and subsequent thermalization take place sufficiently fast, i.e., within a few nanoseconds. The induced local pressure increase $p : \Omega \to \mathbb{R}_+$ initiates a broadband acoustic pulse that travels through Ω within a few microseconds. Therefore, this part of the signal generation is commonly modeled as an initial value problem for the wave equation

$$(2.1) \qquad (\partial_{\tau\tau} - c^2 \Delta)\breve{p}(r, \tau) = 0 \ , \qquad \breve{p}(r, \tau = 0) = p \ , \qquad \partial_\tau \breve{p}(r, \tau = 0) = 0 \ .$$

This approximates the whole optical part as instantaneous, which is equivalent to assuming the tissue remains at rest until the thermalization is complete. Sequential scanning systems can only measure a single spatial projection of $\breve{p}(r, \tau)$ over a sensor surface $\mathcal{S} \subset \partial\Omega$ for each pulse of the excitation laser:

$$(2.2) \qquad f_{m,l} = \int_{[0,\mathcal{T}]} \int_{\mathcal{S}} \breve{p}(r, \tau)\phi_m(r)\psi_l(\tau) \ dr \ d\tau, \quad m = 1, \ldots, M, \quad l = 1, \ldots, M_\tau \ .$$

where $\phi_m(r)$ describes the spatial window function used for the measurement associated with the mth laser pulse, and ψ_l is the lth temporal window function (we will only consider equidistant temporal point sampling in the following).

A single pressure-time series is recorded within a few microseconds and can therefore be regarded as instantaneous if we are interested in imaging dynamics taking place on the scale of a few seconds or even minutes. However, as described in more detail in [1], to form high-resolution 3D images, the spatial Nyquist criterion necessitates that several thousand such time series be recorded. As the pulse repetition rates of conventional excitation lasers are typically limited to tens of Hz, this means that the scanning process and the image dynamics interfere—the image is moving while the scanning is taking place—and neglecting this by assuming an instantaneous measurement can lead to severe motion blurring in the reconstructed images. A summary of the relevant time scales is depicted in Figure 1.

A fully continuous modeling encompassing all the different and interfering spatio-temporal processes described above is of only limited practical value and will not be pursued here. Instead, we assume that a temporal binning of the sequential acquisitions (2.2) into temporal frames, $t = 1, \ldots, T$, is chosen in such a way that the initial pressure can be assumed to be static during one frame. We then model the linear mapping of the discretized initial pressure $p_t \in \mathbb{R}^N$ to fully sampled, discrete data $f_t \in \mathbb{R}^{MM_\tau}$ via (2.1) and (2.2) by a time-independent, i.e., instantaneous, operator A. In this context, "fully sampled" refers to an ideal scanning scheme that samples $\mathcal{S}$ as demanded by the spatial Nyquist criterion, although our measurement setup might practically not allow for doing that within the duration of a single temporal bin. The real measurement is modeled by applying a time-dependent subsampling or *compression* operator $C_t \in \mathbb{R}^{M_c M_\tau \times MM_\tau}$ to f_t:

$$(2.3) \qquad\qquad f_t^c = C_t f_t = C_t A p_t + \varepsilon_t \ , \qquad\qquad t = 1, \ldots, T \ ,$$
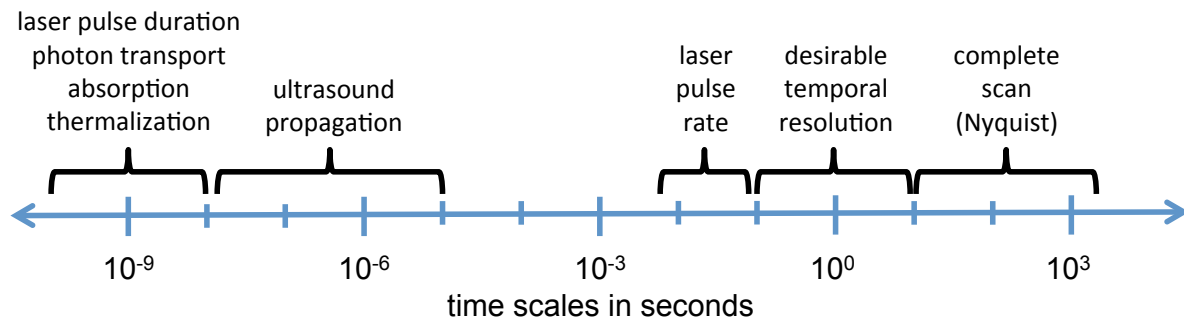
Figure 1. *Sketch of the relevant time scales in high-resolution 4D PAT with sequential acquisition: PAT is particularly suited to image dynamic processes in living tissue that are related to blood circulation. For this, one would ideally like to obtain a temporal resolution close to the heart beat (∼1s for humans, ∼0.1s for mice). As the processes that contribute to a single PA signal take place within nanoseconds to microseconds, the main temporal limitation of sequential acquisition systems is given by the excitation laser pulse rate. Lasers with sufficiently high pulse energies are currently limited to ∼200Hz, which for a complete scan typically leads to longer acquisition times than desired. For instance, scanning 20 000 locations with a 20Hz laser takes 1000s, while using a 200Hz laser in combination with a multibeam readout system as described in [41] takes 12.5s. In the latter case, applying compressed sensing with a sufficiently high subsampling factor would yield the desired temporal resolution.*

where ε_t accounts for additive measurement noise, which we assume can be modeled as an independent and identically distributed standard normal distributed after suitable data pre-processing is carried out.

We will mainly use s-periodic sequences C_t, $t = 1, 2, \ldots$, such that, for any $t_0 \geqslant 1$,

$$(2.4) \qquad \bar{C}_{t_0} = \begin{pmatrix} C_{t_0} \\ C_{t_0+1} \\ \vdots \\ C_{t_0+s} \end{pmatrix}$$

is invertible and can be transformed into $\bar{C}_1$ by row-permutation. This amounts to splitting a conventional, full scanning pattern $\bar{C} \in \mathbb{R}^{MM_\tau \times MM_\tau}$ consisting of M spatial projections into smaller temporal bins comprising disjoint subsets of M_c spatial projections and allows for an intuitive definition of the subsampling factor as $M_{\mathrm{sub}} = M/M_c$. However, the methods presented here can be used for any sequence $\{C_t\}_t^T$.

From now on, any reference to time is with respect to the image and measurement dynamics (indexed by t), not to the acoustic wave propagation (indexed by τ). Furthermore, we will often ease the notation when dealing with spatio-temporal quantities: Dropping the temporal index t refers to the whole sequence as a vector, e.g., $p \in \mathbb{R}^{NT}$. When spatial operators like the gradient ∇ are applied to such a vectorized dynamic quantity, it is understood as an fbf application, i.e., ∇p means $(I_T \otimes \nabla) p$, where I_T is the T-dimensional identity matrix.

2.2. Previous work. In [1], we focused on fbf image reconstruction techniques for (2.3); i.e., we reconstructed each p_t separately, without taking into account any temporal relationship

in the data f_t^c. In particular, we showed that variational approaches,

$$(2.5) \qquad \hat{p}_t = \operatorname*{argmin}_{p_t \geqslant 0} \left\{ \frac{1}{2} \|C_t A\, p_t - f_t^c\|_2^2 + \alpha \mathcal{J}(p_t) \right\}, \qquad \alpha > 0 \,,$$

that use the regularization functional $\mathcal{J}(p)$ to impose sparsity constraints that encode a priori knowledge that the images mainly consist of structures of low spatial complexity outperform linear reconstructions such as *time-reversal* or other *back-projection*-type approaches [26, 64, 2]. Similar studies by others confirm these results [51, 32, 65, 46, 47, 39, 6, 23]. As (2.5) has to be solved by iterative optimization schemes, fbf image reconstruction is appealing from a computational perspective. However, as it can only encode spatial a priori information, its ability to obtain good quality images from subsampled dynamic data (2.3) is limited. With data from an experimental phantom that will be described in more detail in section 5.2, we were able to show in [1] that while fbf reconstructions with $M_{\text{sub}} = 8$ still give acceptable results, using $M_{\text{sub}} = 16$ leads to reconstructions too heavily impaired by missing-data artifacts and noise. However, an inspection of consecutive frames as shown in Figure 2 reveals that the temporal correlation between both noise and artifacts differs strongly from the smooth spatio-temporal evolution of the target. Consequently, noise and artifacts should be effectively removed when using an appropriate smooth spatio-temporal image model. This is the key observation we will utilize to enhance dynamic compressed sensing PAT, either to improve the image quality compared to fbf reconstructions or to allow for higher subsampling factors M_{sub}.

3. Joint image reconstruction and motion estimation.

3.1. Simultaneous motion estimation.
A full nonparametric, spatio-temporal variational scheme reads

$$(3.1) \qquad \hat{p} = \operatorname*{argmin}_{p \geqslant 0} \left\{ \sum_t^T \frac{1}{2} \|C_t A\, p_t - f_t^c\|_2^2 + \mathcal{R}(p) \right\},$$

where the regularization $\mathcal{R}(p)$ is now a function of the whole image sequence $p \in \mathbb{R}^{NT}$ that cannot be decomposed over frames, i.e., $\mathcal{R}(p) \neq \sum_t \mathcal{J}_t(p_t)$. Here, we choose a particular construction of such a scheme introduced in [10]. For our time-discrete dynamic PAT problem, it is given as

$$(3.2) \qquad (\hat{p}, \hat{v}) = \operatorname*{argmin}_{(p \geqslant 0, v)} \left\{ \sum_t^T \frac{1}{2} \|C_t A\, p_t - f_t^c\|_2^2 + \alpha \mathcal{J}(p_t) + \beta \mathcal{H}(v_t) + \gamma \mathcal{M}(p, v) \right\}.$$

Here, each $v_t \in \mathbb{R}^{dN}$ describes a d-dimensional vector field describing the motion between p_t and p_{t+1}; $\mathcal{J}(p_t)$ and $\mathcal{H}(v_t)$ are spatial regularization terms on image and motion field, respectively; and α, β, γ are nonnegative regularization parameters. The key term is $\mathcal{M}(p, v)$, which enforces a relation between image sequence p and related motion field sequence v by measuring how well they fulfill a (discretized) motion PDE chosen to model a priori information about the underlying image dynamics. Note that (3.1) can be obtained from (3.2) by dropping $\hat{v}$ from the left-hand side and replacing the argmin over v with a minimization over v.
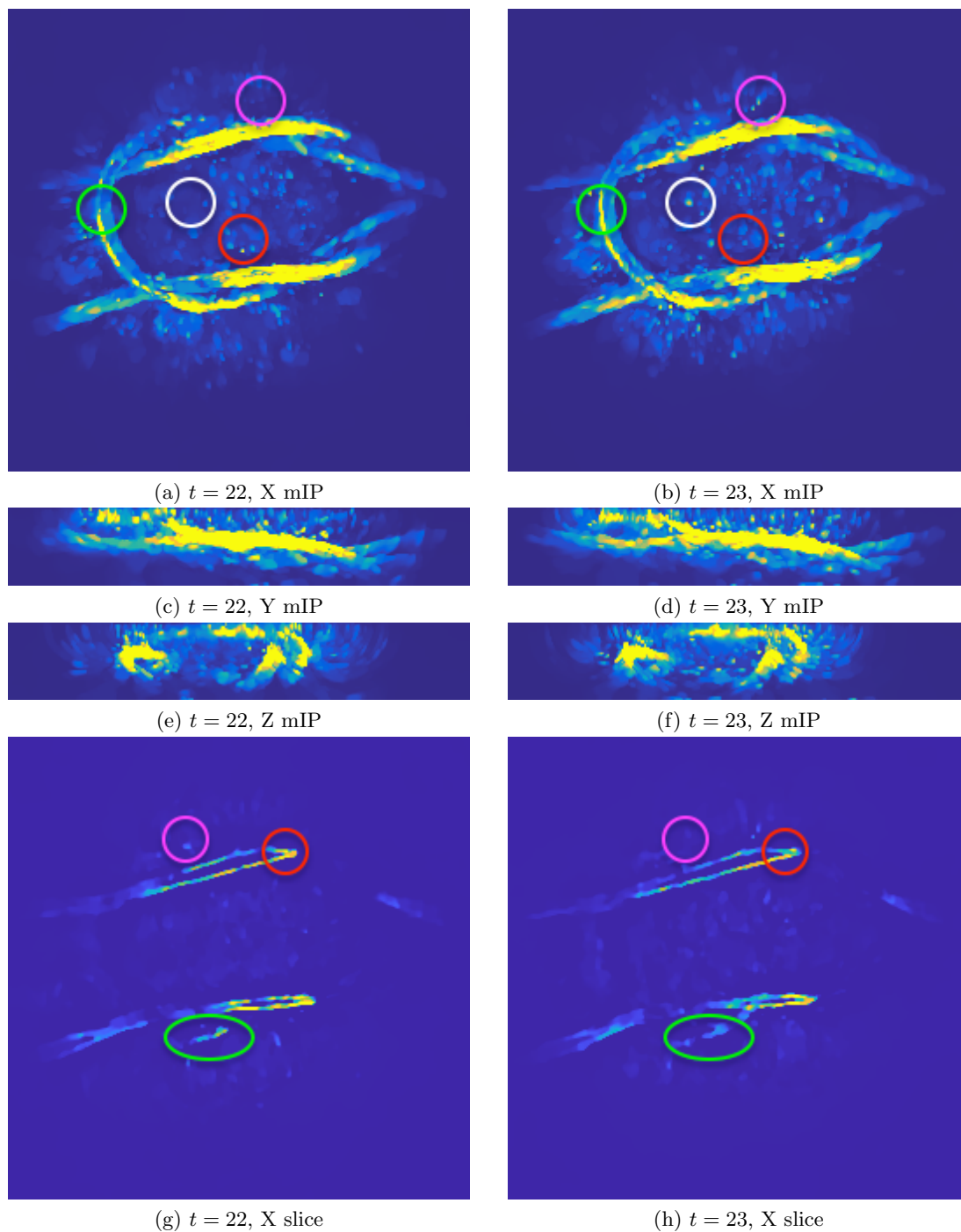
(a) $t = 22$, X mIP

(b) $t = 23$, X mIP

(c) $t = 22$, Y mIP

(d) $t = 23$, Y mIP

(e) $t = 22$, Z mIP

(f) $t = 23$, Z mIP

(g) $t = 22$, X slice

(h) $t = 23$, X slice

Figure 2. *Limitations of applying fbf image reconstruction* (2.5) *to a dynamic PAT data set (details given in section* 5.2*) when using a high subsampling factor* $M_{\mathrm{sub}} = 16$. *Figures in the left and right columns show maximum intensity projections (mIP) along different directions (*(a)–(f)*) and slice view visualizations (*(g)–(h)*) of the results* p_{22} *and* p_{23}, *respectively. One can easily see that image artifacts are not correlated between the two subsequent time frames (same colored circles in left and right images highlight examples), while the target's motion is.*

3.2. Optical flow constraints. The purpose of this work is a proof-of-concept study to show that a more sophisticated spatio-temporal approach like (3.2) can generally improve upon simpler fbf reconstruction. Therefore, we stick to rather generic choices of $\mathcal{J}$, $\mathcal{H}$, and $\mathcal{M}$ and leave the examination of problem-specific regularizers encoding more detailed information about image and dynamics for future work. For $\mathcal{J}$ and $\mathcal{H}$ we choose the popular (isotropic) *total variation* (TV) functional also used in [1]. The motion term $\mathcal{M}(p, v)$ should enforce a simple continuity equation, known as the *optical flow* equation [38] in the field of computer vision:

$$(3.3) \qquad \partial_\tau p(r, \tau) + (\nabla_r p(r, \tau)) \cdot v(r, \tau) = 0 \ .$$

One way to achieve this is to let $\mathcal{M}$ measure the least-squares error of a forward difference discretization of (3.3) in time:

$$(3.4) \qquad \mathcal{M}(p, v) = \sum_{t}^{T-1} \frac{1}{2} \left\| p_{t+1} - p_t + (\nabla p_t) \cdot v_t \right\|_2^2 \ .$$

In total, this leads to the variational scheme

$$
\begin{aligned}
(\hat{p}, \hat{v}) = \operatorname*{argmin}_{p \geqslant 0, v} \Big\{ \mathcal{E}(p, v) \Big\} :=& \operatorname*{argmin}_{p \geqslant 0, v} \Bigg\{ \sum_{t}^{T} \frac{1}{2} \left\| C_t A\, p_t - f_t^c \right\|_2^2 \\
(3.5) \qquad & + \alpha \| \nabla^+ p_t \|_1 + \beta \sum_{i}^{d} \| \nabla^+ v_{x_i, t} \|_1 + \frac{\gamma}{2} \left\| p_{t+1} - p_t + (\nabla^\pm p_t) \cdot v_t \right\|_2^2 \Bigg\} ,
\end{aligned}
$$

where we define $p_{T+1} := p_T$, $v_T := 0$ to simplify the formula. The spatial gradients in the TV terms are implemented with forward differences (denoted by ∇^+) as described in Appendix A in [1]. We chose to implement the TV of the motion field as a sum over the TV of the single components here and leave other possible choices for future work. The spatial gradient ∇p_t in the optical flow term is discretized using central differences, denoted by $\nabla^\pm$. As the scheme is solved implicitly for p given v, this gives a stable discretization of (3.3). More details on the discretization can be found in [20]. We will refer to (3.5) as the *TVTVL2* model.

4. Optimization. The TVTVL2 model (3.5) leads to a large-scale, nonsmooth, biconvex optimization problem in p and v involving the computationally intensive acoustic propagation operator A applied to T image frames. We will therefore decompose it into several subproblems to disentangle its most complicated components. All subproblems will be solved with iterative first order techniques. For general introductions to numerical optimization suited for imaging applications, we refer the reader to [11, 14].

4.1. Forward-backward splitting. Computing each matrix-vector product Ap_t or $A^* f_t$ involves the numerical solution of a potentially inhomogeneous 3D wave equation (2.1) with high spatial and temporal resolution. For this, we will use the *k*-space pseudospectral time domain method [44, 18, 59] implemented in the *k-Wave* MATLAB Toolbox [58]. With this implementation, each matrix-vector product with A or A^* has the complexity $\mathcal{O}(M_\tau N \log(N))$ [2] and typically $M_\tau > N$. In contrast, all linear operators in the regularization terms in (3.5)

have complexity $\mathcal{O}(N)$. For this reason, we build the outermost iteration (index i) of our scheme by decoupling the smooth, convex data term containing A from all other terms and the nonnegativity constraints on p by a *proximal forward-backward splitting/proximal gradient descent* scheme (see [30] for an extensive overview). For this, we need to define the *proximal operator* of a functional $\mathcal{J}(x)$ as

$$(4.1) \qquad \mathrm{prox}_{\alpha\mathcal{J}}(y) := \underset{x}{\mathrm{argmin}} \left\{ \alpha\mathcal{J}(x) + \frac{1}{2}\|x - y\|_2^2 \right\}.$$

Furthermore, as v is not part of the data term, it appears only in the second step of the iterative scheme:

$$(4.2a) \qquad \tilde{p}_t = p_t^i - \eta A^* C_t^* \left(C_t A p_t^i - f_t^c \right) \ \forall t = 1, \ldots, T \qquad \text{(forward step)},$$

$$(4.2b) \qquad \left(p^{i+1}, v^{i+1} \right) = \mathrm{prox}_{\eta\mathcal{R}}\left(\tilde{p}\right) \qquad\qquad\qquad \text{(backward step)},$$

where $\mathcal{R}(p, v)$ combines all regularization terms on p and v from (3.5):

$$(4.3) \qquad \mathcal{R}(p, v) := \sum_t^T \alpha\|\nabla^+ p_t\|_1 + \beta\sum_i^d \|\nabla^+ v_{x_i,t}\|_1 + \frac{\gamma}{2} \left\| p_{t+1} - p_t + (\nabla^\pm p_t) \cdot v_t \right\|_2^2.$$

In (4.2a)–(4.2b), we initialize $p^0 = 0$ and set the step size η to $1.5/\max_t L_t$. L_t is an approximation of the Lipschitz constant of $A^* C_t^* C_t A$ which can be precomputed for a given setting and subsampling scheme with a simple power iteration. The basic scheme (4.2a)–(4.2b) is extended by a gradient extrapolation step (*accelerated or fast gradient methods*) which will lead to an asymptotic convergence rate of $\mathcal{O}(1/i^2)$. For this, we use the *FISTA* extrapolation [5] with restart whenever an increase in the total energy $\mathcal{E}$ occurs.

4.2. Biconvex optimization. Combining (4.3) and (4.1), we see that solving the proximal operator in (4.2b) amounts to solving the following *TVTVL2-regularized denoising problem*:

$$(4.4) \qquad \left(p^{i+1}, v^{i+1} \right) = \mathrm{prox}_{\eta\mathcal{R}}\left(\tilde{p}\right) = \underset{p \geqslant 0, v}{\mathrm{argmin}} \left\{ \tilde{\mathcal{E}}(p, v) \right\} := \underset{p \geqslant 0, v}{\mathrm{argmin}} \left\{ \sum_t^T \frac{1}{2}\|p_t - \tilde{p}_t\|_2^2 \right.$$
$$\left. + \eta\alpha\|\nabla^+ p_t\|_1 + \eta\beta\sum_i^d \|\nabla^+ v_{x_i,t}\|_1 + \frac{\eta\gamma}{2}\left\| p_{t+1} - p_t + (\nabla^\pm p_t) \cdot v_t \right\|_2^2 \right\}.$$

The main difficulty here is the motion term. The product $(\nabla^\pm p_t) \cdot v_t$ renders it *biconvex*; i.e., $\tilde{\mathcal{E}}(p, v)$ is convex in each of the single variables p or v once the other is fixed, but nonconvex as a function of both variables. As such, biconvex problems are global optimization problems that can have a large number of local minima. An overview of biconvex optimization can be found in [31]. Compared to general global optimization problems, the convex substructures can be utilized to design efficient optimization schemes with certain global convergence properties. A popular approach is given by the *alternate convex search* (ACS) method, which alternates between minimizing $\tilde{\mathcal{E}}(p, v)$ for one variable while keeping the other fixed. Applied to (4.4),

the ACS iteration (index j) reads

$$(4.5a) \qquad p^{j+1} = \underset{p \geqslant 0}{\operatorname{argmin}} \left\{ \sum_t^T \frac{1}{2} \|p_t - \tilde{p}_t\|_2^2 + \tilde{\alpha} \|\nabla^+ p_t\|_1 + \frac{\tilde{\gamma}}{2} \left\| p_{t+1} - p_t + (\nabla^{\pm} p_t) \cdot v_t^j \right\|_2^2 \right\},$$

$$(4.5b) \qquad v^{j+1} = \underset{v}{\operatorname{argmin}} \left\{ \sum_t^T \tilde{\beta} \sum_i^d \|\nabla^+ v_{x_i,t}\|_1 + \frac{\tilde{\gamma}}{2} \left\| p_{t+1}^{j+1} - p_t^{j+1} + (\nabla^{\pm} p_t^{j+1}) \cdot v_t \right\|_2^2 \right\},$$

where we defined $\tilde{\alpha} = \eta\alpha$, $\tilde{\beta} = \eta\beta$, $\tilde{\gamma} = \eta\gamma$. The first problem (4.5a) is a denoising problem for p with a regularization consisting of a TV and a transport term. Problem (4.5b) for v is an optical flow estimation problem with TV regularization. Note that it is separable in t; i.e., it can be solved fbf. Both subproblems are convex, and therefore approximate solutions can be found reasonably fast by iterative first order schemes (iteration index k). The next two sections will present two different approaches for each subproblem. First, we will repeat how to apply the *primal dual hybrid gradient* (PDHG) [50, 13] algorithm as already proposed in [10]. While this will be sufficient for treating small-scale 2D problems such as examined in section 5.1, we will then introduce tailored *alternating direction method of multipliers* (ADMM) (see, e.g., [7]) schemes that will be shown to be sufficiently efficient to also treat large-scale 3D problems as encountered in the real-data scenarios examined in section 5.2.

However, as (4.5a) and (4.5b) are nonsmooth, both PDHG and ADMM rely on dual or primal dual formulations and can therefore not guarantee a monotonous decay of the iterates energy $\tilde{\mathcal{E}}(p, v)$. This leads to a potential problem: While ACS will still converge in objective value $\tilde{\mathcal{E}}(p, v)$ if we do not solve the subproblems exactly but only find fast approximate solutions, we need to guarantee that $\tilde{\mathcal{E}}(p, v)$ decreases in every step. Therefore, we will need to track the energies of all iterates and allow subroutines to run long enough to ensure a sufficient decay. In addition, we will warm-start the subroutines with all the variables from their last call, even though this will lead to an increased memory consumption.

4.3. Solution of convex subproblems by PDHG. The PDHG algorithm has become the de facto standard template for solving convex, nonsmooth optimization problems in a vector space $\mathcal{X}$ involving complicated linear operators $K : \mathcal{X} \to \mathcal{Y}$ for which matrix-vector products with K and K^* can be computed. The idea is to formulate the problem in the *primal form* as

$$(4.6) \qquad \min_{x \in \mathcal{X}} \mathcal{E}(x) = \min_{x \in \mathcal{X}} \mathcal{G}(x) + \mathcal{F}(Kx),$$

with proper, convex functionals $\mathcal{G}$ and $\mathcal{F}$, and to then switch to the equivalent *primal dual formulation*,

$$(4.7) \qquad \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \langle Kx, y \rangle + \mathcal{G}(x) - \mathcal{F}^*(y),$$

which involves the *convex conjugate* $\mathcal{F}^*$ of $\mathcal{F}$. The advantage of this formulation over (4.6) is that the operator K does not show up in the nonlinear terms anymore. The PDHG algorithm then solves the saddle-point problem (4.7) by basically alternating a gradient descent

Algorithm 4.1. Primal dual hybrid gradient scheme (PDHG).

Given $\mu > 0$, $\nu > 0$, $\theta \in [0,1]$, $\hat{x}^0$, y^0, iterate for $k = 1, 2, \ldots$:

$$(4.8a) \qquad y^{k+1} = \operatorname{prox}_{\nu F^*}\left(y^k + \nu K \hat{x}^k\right) \qquad \text{(prox-grad step in } y)\,,$$

$$(4.8b) \qquad x^{k+1} = \operatorname{prox}_{\mu G}\left(x^k - \mu K^* y^{k+1}\right) \qquad \text{(prox-grad step in } x)\,,$$

$$(4.8c) \qquad \hat{x}^{k+1} = x^{k+1} + \theta\left(x^{k+1} - x^k\right) \qquad \text{(overrelaxation)}\,.$$

in the primal variable and a gradient ascent in the dual variable. In addition, it performs an overrelaxation step in one of the variables (here, the primal one); see Algorithm 4.1.

To apply this to (4.5a), i.e., $x = p$, we choose

$$(4.9a) \qquad Kp = \begin{bmatrix} \nabla^+ \\ \partial_t^+ + \sum_i^d v_{x_i}^j \partial_{x_i}^\pm \end{bmatrix} p := \begin{bmatrix} \nabla^+ \\ D_{v^j} \end{bmatrix} p\,, \qquad K^* y = -\nabla^+ \cdot y_1 + D_{v^j}^* y_2\,,$$

$$(4.9b) \qquad \mathcal{G}(p) = \chi_+(p) + \sum_t^T \frac{1}{2}\|p_t - \tilde{p}_t\|_2^2\,, \qquad \chi_+(p) := \begin{cases} 0 & \text{if } p_i \geqslant 0\ \forall\, i\,, \\ \infty & \text{else}\,, \end{cases}$$

$$(4.9c) \qquad \mathcal{F}(y) = \mathcal{F}\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \tilde{\alpha}\|y_1\|_1 + \frac{\tilde{\gamma}}{2}\|y_2\|_2^2\,,$$

where $y_1 \in \mathbb{R}^{NTd}$ represents a d-dimensional spatio-temporal vector field resulting from applying ∇^+ to every frame of a d-dimensional dynamic image p ($d = 2, 3$ here). The explicit form of the proximal operators needed to implement Algorithm 4.1 with these choices are listed in Appendix A. We can use the PDHG scheme to solve (4.5b), i.e., $x = v$, by choosing

$$(4.10a) \qquad Kv = \left(I_d \otimes \nabla^+\right)\begin{bmatrix} v_{x_1} \\ \vdots \\ v_{x_d} \end{bmatrix}\,, \qquad K^* y = \begin{bmatrix} -\nabla^+ \cdot y_1 \\ \vdots \\ -\nabla^+ \cdot y_d \end{bmatrix}\,,$$

$$(4.10b) \qquad \mathcal{G}(v) = \frac{\tilde{\gamma}}{2}\sum_t^T \left\|p_{t+1}^{j+1} - p_t^{j+1} + (\nabla^\pm p_t^{j+1})\cdot v_t\right\|_2^2\,,$$

$$(4.10c) \qquad \mathcal{F}(y) = \mathcal{F}\left(\begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix}\right) = \sum_t^T \tilde{\beta}\sum_i^d \|y_{i,t}\|_1\,.$$

Here, $y \in \mathbb{R}^{NTd^2}$ represents a $(d \times d)$-dimensional spatio-temporal tensor field with components $y_i \in \mathbb{R}^{NTd}$ representing the spatial Jacobian of every frame of a d-dimensional dynamic vector field v. Again, the solution of the involved proximal operators is shifted to Appendix A. Note that as (4.5b) can be solved fbf, i.e., for each v_t separately, the PDHG algorithm sketched above can be parallelized over t. While this is an appealing option, for its use within ACS one has to implement it such that the overall energy $\tilde{\mathcal{E}}(p,v)$ (which is summed over t) deceases sufficiently; cf. section 4.2.

The overrelaxation parameter θ in both PDHG schemes is chosen as 1. Furthermore, we need to choose the step sizes μ, ν in dependence on K to ensure convergence (cf. [13, 14]). Due to the complicated structure of (4.9a), we use the extension of PDHG by diagonal preconditioning proposed as in [49] (the α parameter in [49] is set to 1), which is easy to compute for our problem and was found to work well compared to standard choices based on estimates of $\|K\|_{2,2}$. The operator K in (4.10a) has a simple structure, and it can be shown that $\|K\|_{2,2} \leqslant 4d$ [13]. As such, the choice $\mu = 1/2d$, $\nu = 1/2$ fulfills $\mu\nu \|K\|_{2,2}^2 \leqslant 1$ and leads to convergence (this balancing between μ and ν was found empirically).

4.4. Solution of convex subproblems by ADMM. In the ADMM approach, the unconstrained but coupled convex problem (4.6) is first converted into an equality-constrained but uncoupled convex problem by introducing an auxiliary variable $y = Kx$,

$$(4.11) \qquad (4.6) \iff \min_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathcal{G}(x) + \mathcal{F}(y) \quad \text{such that} \quad y = Kx \,,$$

which is then solved by a combination of *dual ascent*, *augmented Lagrangian techniques* and the *method of multipliers*. The final ADMM scheme is described in Algorithm 4.2.

Algorithm 4.2. Alternating direction method of multipliers (ADMM).

Given $\rho > 0$, y^0, w^0, iterate for $k = 1, 2, \ldots$:

$$(4.12) \qquad x^{k+1} = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \mathcal{G}(x) + \frac{\rho}{2} \left\| Kx - y^k + w^k \right\|_2^2 \right\},$$

$$(4.13) \qquad y^{k+1} = \operatorname*{argmin}_{y \in \mathcal{Y}} \left\{ \mathcal{F}(y) + \frac{\rho}{2} \left\| Kx^{k+1} - y + w^k \right\|_2^2 \right\},$$

$$(4.14) \qquad w^{k+1} = w^k + Kx^{k+1} - y^{k+1} \,.$$

The crucial difference to the PDHG schemes is that the update of x, (4.12), is now implicit, and we will choose the split $y = Kx$ such that it will be given as the solution of a least-squares problem involving all linear operators.[1] This can be advantageous in cases where K suffers from bad conditioning, but only leads to a computationally efficient scheme if the corresponding normal equations can be solved quickly. Fortunately, ADMM still converges if the subproblems (4.12) and (4.13) are solved approximately but with accuracy increasing with k (see [25] and references therein for a precise statement). Therefore, warm-started iterative linear solvers with carefully chosen stop conditions can be used. For problem (4.5a), i.e.,

---

[1]The ADMM and PDHG schemes are actually very closely related, although most introductions of the two methods do not immediately imply this, and our short overview here cannot cover it. See [11, 14] for an extensive discussion.

$x = p$, we realize the ADMM iteration by

$$(4.15a) \qquad Kp = \begin{bmatrix} \nabla^+ \\ I_N \end{bmatrix} p \quad \Rightarrow \quad K^* \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = -\nabla^+ \cdot y_1 + y_2 \,,$$

$$(4.15b) \qquad \mathcal{G}(p) = \sum_t^T \frac{1}{2} \|p_t - \tilde{p}_t\|_2^2 + \frac{\tilde{\gamma}}{2} \left\| p_{t+1} - p_t + v_t^j \cdot (\nabla p_t) \right\|_2^2$$

$$= \frac{1}{2} \left\| \begin{bmatrix} I_N \\ \sqrt{\tilde{\gamma}} D_{v^j} \end{bmatrix} p - \begin{bmatrix} \tilde{p} \\ 0 \end{bmatrix} \right\|_2^2 \,,$$

$$(4.15c) \qquad \mathcal{F}(y) = \mathcal{F}\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) = \tilde{\alpha}\|y_1\|_1 + \chi_+(y_2) \,,$$

where $y_1 \in \mathbb{R}^{NTd}$ represents a d-dimensional spatio-temporal vector field resulting from applying ∇^+ to a d-dimensional dynamic image p, and $y_2 \in \mathbb{R}^{NT}$ accounts for the nonnegativity constraints. We will denote the corresponding parts of w by w_1 and w_2 as well. For these choices, the update (4.12) is given by

$$(4.16) \quad p^{k+1} = \underset{p}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \begin{bmatrix} I_N \\ \sqrt{\tilde{\gamma}} \, D_{v^j} \\ \sqrt{\rho} \, \nabla^+ \\ \sqrt{\rho} \, I_N \end{bmatrix} p - \begin{bmatrix} \tilde{p} \\ 0 \\ \sqrt{\rho} \left(y_1^k - w_1^k\right) \\ \sqrt{\rho} \left(y_2^k - w_2^k\right) \end{bmatrix} \right\|_2^2 \right\}$$

$$= \left((1 + \rho)I_N + \tilde{\gamma} D_{v^j}^* D_{v^j} + \rho \Delta^+ \right)^{-1} \left(\tilde{p} + \rho \nabla^+ \cdot \left(y_1^k - w_1^k \right) + \rho \left(y_2^k - w_2^k \right) \right) \,.$$

All the linear operators can easily be implemented in a matrix-free way, and so (4.16) can be solved with a standard *conjugate gradient* (CG) implementation. As in the PDHG schemes, update (4.13) can be solved explicitly using proximal operators (Appendix A).

For solving (4.5b), i.e., $x = v$, choose exactly the same split as in the corresponding PDHG scheme, i.e., (4.10a)–(4.10c), and make use of the fact that the optimization can be solved for each t separately: For each t, the update (4.12) is given by

$$(4.17) \qquad v_t^{k+1} = \left(\tilde{\gamma} E^* E + \rho I_d \otimes \Delta^+ \right)^{-1} \left(\tilde{\gamma} E^* \left(-p_{t+1}^{j+1} + p_t^{j+1} \right) + \rho K^* \left(y^k - w^k \right) \right) \,,$$

where $E = E(p_t^{j+1})$ is an $N \times dN$ matrix implementing the pointwise multiplication and summation of the components of a vector field with the spatial gradients of p_t^{j+1}, i.e., $Ev = \sum_i^d v_{x_i} \partial_{x_i}^{\pm} p_t^{j+1}$ as

$$(4.18) \qquad E = \begin{bmatrix} \operatorname{diag}\left(\partial_{x_1}^{\pm} p_t^{j+1} \right) & \dots & \operatorname{diag}\left(\partial_{x_d}^{\pm} p_t \right) \end{bmatrix} \,.$$

For $d = 2$, we take a closer look at the structure of the matrix to invert in (4.17):

$$(4.19) \qquad \left(\tilde{\gamma} E^* E + \rho I_d \otimes \Delta^+ \right) = \tilde{\gamma} \begin{bmatrix} T_{11} & T_{12} \\ T_{11} & T_{22} \end{bmatrix} + \rho \begin{bmatrix} \Delta^+ & 0 \\ 0 & \Delta^+ \end{bmatrix} \,,$$

where $T_{kl} := \text{diag}\big(\partial_{x_k}^{\pm} p_t^{j+1} \cdot \partial_{x_l}^{\pm} p_t^{j+1}\big)$. While one can easily implement matrix-free iterative solvers for this system, we chose to explicitly build this very sparse matrix to be able to use efficient preconditioning techniques. Within the ADMM iteration, this comes with little overhead as only the right-hand side in system (4.17) changes during the iteration. We will examine different combinations of preconditioners and iterative solvers in the numerical studies [53]. As preconditioners, we consider the following:

- IC(0): incomplete Cholesky preconditioner with zero-fill as implemented in MATLAB (R2016a).
- ICT: incomplete Cholesky preconditioner with threshold dropping (threshold: 1e-3) as implemented in MATLAB (R2016a).
- AMG: algebraic multigrid W-cycle preconditioner based on the implementation in [43], which uses a modification of Ruge–Stuben coarsening, two-point interpolation (using at most two connected coarse nodes), and a direct solver on the coarsest level.

As iterative solvers, we examine the standard CG method and the *minimum residual method* (MINRES). Further details will be discussed in the next section. The iterative solvers are warm-started with the previous solution p^k or v_t^k; we perform at least 3 iterations and stop when the relative residual norm is below $\text{tol}(k) = 10^{-3}/k^{3/2}$, i.e., we progressively increase the precision to which we solve subproblem (4.12). Update (4.13) can be solved as for the corresponding PDHG scheme (Appendix A).

While ADMM converges for all $\rho > 0$, its choice has a crucial impact on the speed of convergence and other properties of the iterates, e.g., the monotonicity of the energy $\mathcal{E}\big(x^k\big)$, which is important for using ADMM inside of an ASC (cf. section 4.2). For using ADMM on the p update (4.5a), we use the adaptation strategy described in section 3.4.1 of [7] during the steps $k = 1, \dots, 25$ and fix it thereafter. For the first p update (4.5a) within ASC we initialize $\rho = 1$ and then always warm-start the following p update with the adapted ρ. In the ADMM scheme for the v update (4.5b), we fix $\rho = 10^{-1}$ for $d = 2$ and 10^2 for $d = 3$, first to avoid a recomputation of the matrices and their preconditioners (see above), and second to enforce a fast transition to the regime of monotonous energy decay. As with any alternating optimization, the ADMM scheme can benefit from overrelaxation. We use the technique discussed in section 3.4.3 of [7], which consists of replacing the quantity Kx^{k+1} in Algorithm 4.2 by $sKx^{k+1} + (1 - s)y^k$. Throughout the experiments, we use $s = 1.8$.

Although we limited our presentation here to the most important features, it already became apparent that, compared to PDHG, ADMM schemes are more difficult to design and parametrize. Also note that ADMM with the specific type of split that we used here is equivalent to the *split Bregman method* [29, 25], which derives Algorithm 4.2 from a different perspective.

5. Results. In this section, we first demonstrate the main features of the proposed methods on a simple numerical phantom in 2D before we discuss their realization for experimental data in 3D. As we can only show snapshots for a few time frames of the reconstructions here, movies of all reconstructions can be found in the supplementary material (M117006_01.pdf [local/web 153KB]). For computing the results presented, we used ADMM in both the p update (4.5a) and the v update (4.5b) as described in the previous section. In the v update, AMG-CG was used as a least-squares solver. In section 5.3, we compare this choice to possible alternatives in more detail.
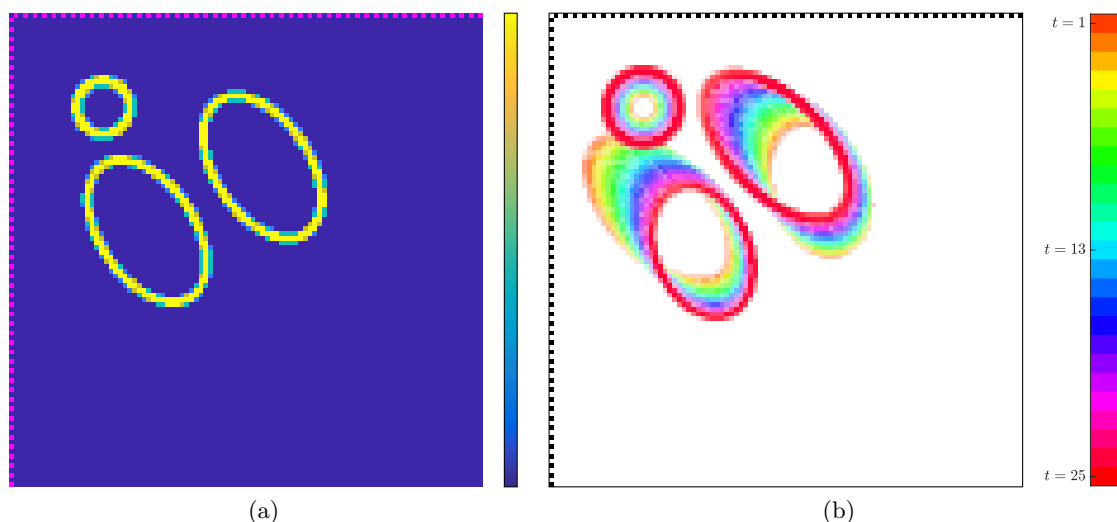
(a)   (b)

Figure 3. *The* 2D *numerical phantom.* (a) *A snapshot at* $t = 13$ *and the corresponding color bar. The sensor locations are shown as pink pixels (left and top edge).* (b) *A visualization of all* $T = 25$ *frames as a color-coded RGB overlay. The color bar displays which colors are assigned to which time frame; the sensor pixels are shown as black pixels.*

All routines have been implemented as part of a MATLAB toolbox for PAT image reconstruction which will be made available in the near future. The toolbox relies on the k-Wave toolbox (see [58], http://www.k-wave.org/) to implement A and A^*, which allows us to use highly optimized C++ and CUDA code to compute the 3D wave propagation on parallel CPU or GPU architectures.

5.1. Numerical 2D phantom. The computational domain is a square of length 20mm which is divided into $N = 100 \times 100$ pixels. Its acoustic properties are assumed homogeneous with $c = 1500 \mathrm{m\,s}^{-1}$. The conventionally scanned (fully sampled) measurement data (referred to as "cnv") is acquired at $M = 100$ sensors sampled at $M_\tau = 472$ time steps with $\delta_\tau = 40$ns. The sensors are arranged in two orthogonal lines, which corresponds to a 2D version of a scanning system using two orthogonal Fabry–Pérot (FP) sensors [24, 27]. This way, reconstructions from the fully sampled sensor array will not suffer from severe limited view artifacts, and we can concentrate on the effects of subsampling.

The dynamical phantom consists of three tubes that change center position, orientation, and size smoothly over $T = 25$ frames and should loosely resemble the dynamics of the X-slices of the experimental phantom in Figure 2. Figure 3 shows different visualizations of the phantom. White Gaussian noise with a standard deviation of $\sigma = 5 \cdot 10^{-3}$ was added to the simulated pressure time series, leading to an average signal-to-noise ratio of 20.65 dB. To subsample the data, we now assume that in each frame, we can acquire data at a subset of 4 out of the 100 original sensor locations that have been chosen randomly but disjointly, such that after $T = 25$ frames, each location has been scanned once. This means that the subsampling factor $M_{\mathrm{sub}} = M/M_c$ (cf. section 2.1) equals T; i.e., we acquire all 25 frames with the same scanning time as a single frame in the full data setup. The operators C_t can
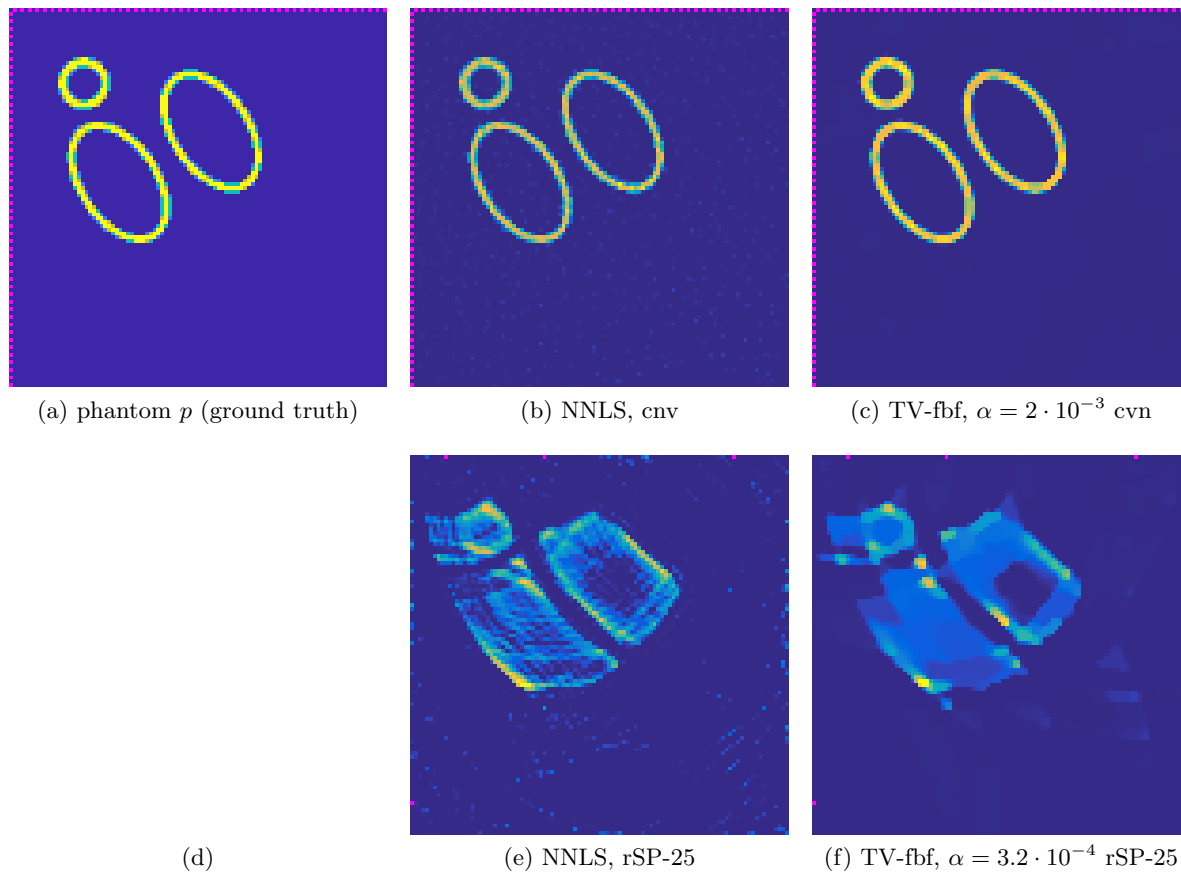
(a) phantom p (ground truth)   (b) NNLS, cnv   (c) TV-fbf, $\alpha = 2 \cdot 10^{-3}$ cvn

(d)   (e) NNLS, rSP-25   (f) TV-fbf, $\alpha = 3.2 \cdot 10^{-4}$ rSP-25

Figure 4. *Snapshots at $t = 13$ of the results of fbf image reconstruction methods (2.5) for full (cvn) and subsampled (rSP-25) data.*

thus be written as $C_t = I_{M_\tau} \otimes \tilde{C}_t$, with $\tilde{C}_t$ being a binary 4×100 matrix with 4 ones on the main diagonal and all zeros otherwise, and

$$(5.1) \qquad \bar{C} = \begin{bmatrix} C_1 \\ \vdots \\ C_T \end{bmatrix}$$

is a row-permutation of I_M. We will denote this subsampling strategy by rSP-25.

First, we compute fbf reconstructions (2.5) without any regularizer $\mathcal{J}(p)$, i.e., *nonnegative least squares* (NNLS), and then using a TV functional (denoted as *TV-fbf*). For this, we use 100 iterations of the accelerated proximal gradient descent introduced in section 4.1. The proximal step (4.2b) is simply a projection onto the positive orthant for NNLS, while it amounts to solving a TV-regularized denoising problem in the case of TV (for details, see [1]). The results are shown in Figure 4 and again demonstrate that while we can obtain a good reconstruction with fbf methods for full data, they fail for severely subsampled data, similarly to the motivating example shown in Figure 2.
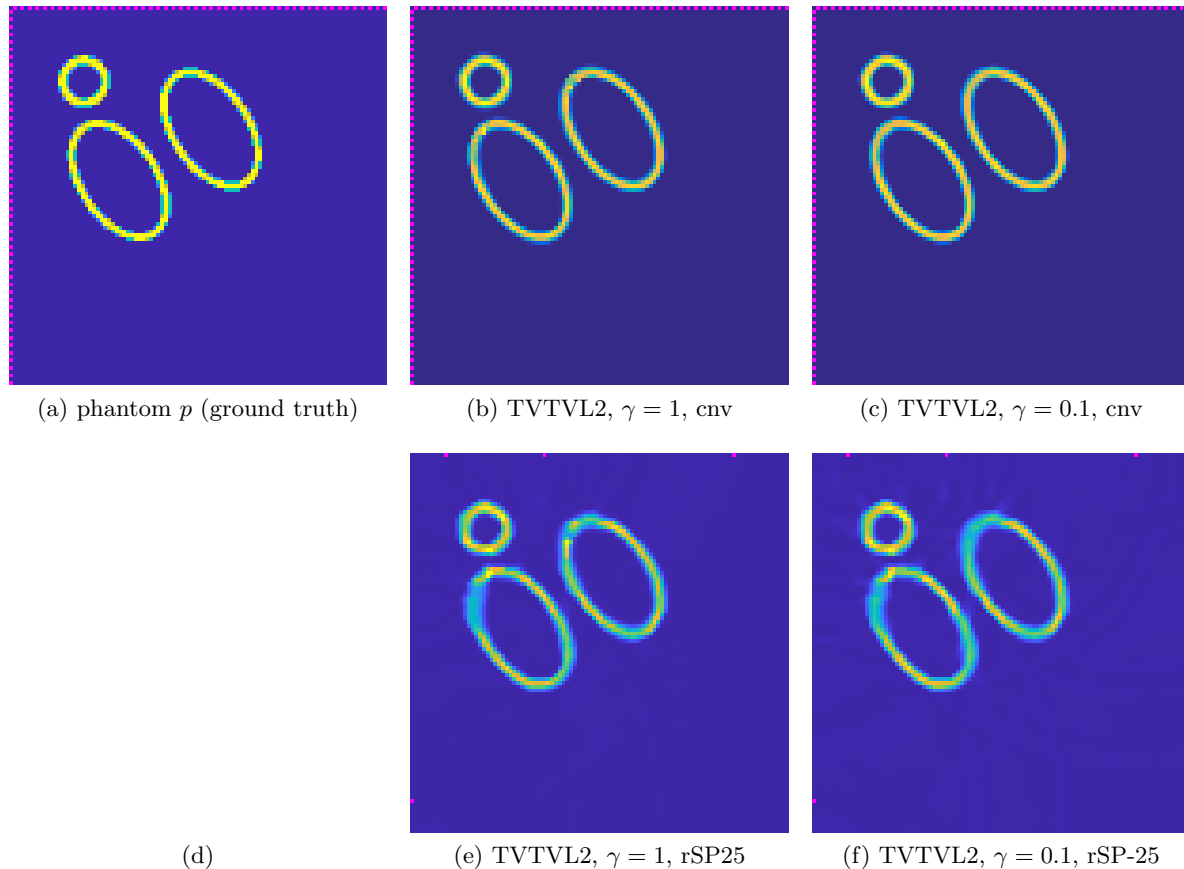
(a) phantom p (ground truth)   (b) TVTVL2, $\gamma = 1$, cnv   (c) TVTVL2, $\gamma = 0.1$, cnv

(d)   (e) TVTVL2, $\gamma = 1$, rSP25   (f) TVTVL2, $\gamma = 0.1$, rSP-25

Figure 5. *Snapshots at $t = 13$ of the results of the TVTVL2 image reconstruction* (3.5) *for full (cvn) and subsampled (rSP-25) data. The parameters α and β were set to the corresponding value of the α used for the TV-fbf reconstructions in Figure* 4, *and γ was set to* 1 *or* 0.1.

Next we compute reconstructions with the TVTVL2 model (3.5). An apparent challenge of this more sophisticated spatio-temporal model that we did not discuss up to now is that it relies on three regularization parameters: α, β, and γ. For TV-fbf, it is easy to fix the single parameter α manually: We computed reconstructions for different α for a single frame, and then used the smallest α that visually removed most noise for all frames, which we will denote as $\hat{\alpha}$. For the TVTVL2 model, we start by simply setting $\alpha = \beta = \hat{\alpha}$ and $\gamma = 1$. Figures 5(b) and 5(e) show the results of this naive parameter choice. Although the reconstructions for the subsampled data still suffer from some blurring and artifacts, one can clearly see a significant improvement compared to the fbf reconstructions in Figure 4. We then varied (α, β, γ) around this first guess. Figures 5(c) and 5(f) show the effect of decreasing γ to 0.1, which has by far the biggest positive impact. Figure 6 illustrates the effects of also varying α and β, which leads to trade-offs between oversmoothing and artifact reduction. We leave a more detailed parameter study for future work and instead investigate the estimated motion fields. Figure 7 shows that main features of the motion fields can be reconstructed even from subsampled data. In particular, the motion fields facilitate the distinction and tracking of different moving objects.
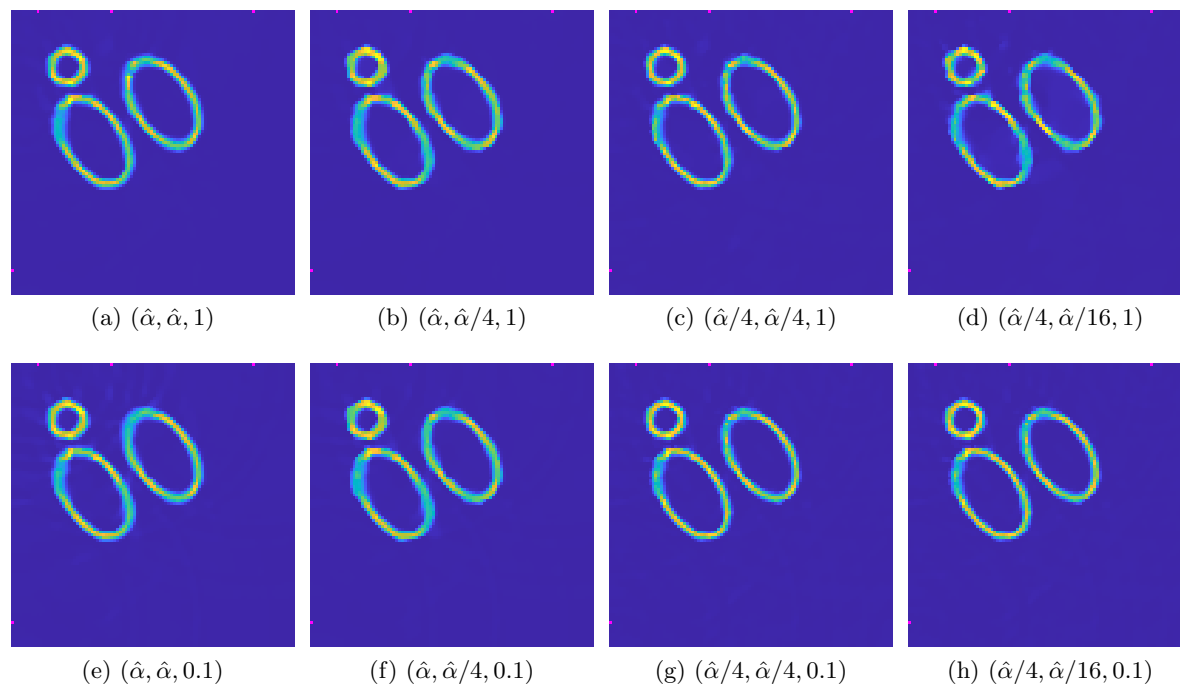
(a) $(\hat{\alpha}, \hat{\alpha}, 1)$ (b) $(\hat{\alpha}, \hat{\alpha}/4, 1)$ (c) $(\hat{\alpha}/4, \hat{\alpha}/4, 1)$ (d) $(\hat{\alpha}/4, \hat{\alpha}/16, 1)$

(e) $(\hat{\alpha}, \hat{\alpha}, 0.1)$ (f) $(\hat{\alpha}, \hat{\alpha}/4, 0.1)$ (g) $(\hat{\alpha}/4, \hat{\alpha}/4, 0.1)$ (h) $(\hat{\alpha}/4, \hat{\alpha}/16, 0.1)$

Figure 6. *Snapshots at $t = 13$ of the results of the TVTVL2 image reconstruction* (3.5) *for subsampled (rSP-25) data for different combinations of (α, β, γ). Here, $\hat{\alpha}$ corresponds to the value of the regularization parameter used for the TV-fbf reconstructions in Figure* 4.

5.2. Experimental 3D phantom. Now, we examine the performance of the methods on high-resolution 3D reconstructions from dynamic experimental phantom data. As outlined in the motivation in section 2.2, we use the same data as in [1] to investigate whether the methods described here can improve upon the fbf reconstructions for $M_{\mathrm{sub}} = 16$ (cf. Figure 2). However, for this article to be self-contained, we first briefly recap the setup and preprocessing used.

5.2.1. Setup and preprocessing. The phantom consists of two polythene tubes filled with 100% and 10% ink immersed in a 1% intralipid solution with de-ionized water. The tubes were interleaved to form a knot with 4 open ends. As shown in Figure 8, while three of the ends are fixed, one is tied to a motor shaft. We then acquired PA data using an FP scanner in a stop-motion style: With the whole arrangement at rest, a full, conventional scan was performed. Then the motor shaft was turned by a fixed angle, which caused the knot to both move towards the motor and tighten, and the new arrangement was scanned again. In total, $T = 45$ frames were acquired. The excitation laser pulses were delivered at a rate of 20Hz, and had a wavelength of 1064nm and an energy of around 20mJ. For a full, conventional scan, pressure time courses at 134×133 locations on a spatial grid with grid size 150μm were measured for $M_t = 625$ time points with a temporal resolution of 12ns. For preprocessing, the data was first clipped to 132×132 locations. Then we preformed baseline-correction, band-pass filtering (0.5–20MHz), and noisy-channel exclusion and clipped the time courses to the
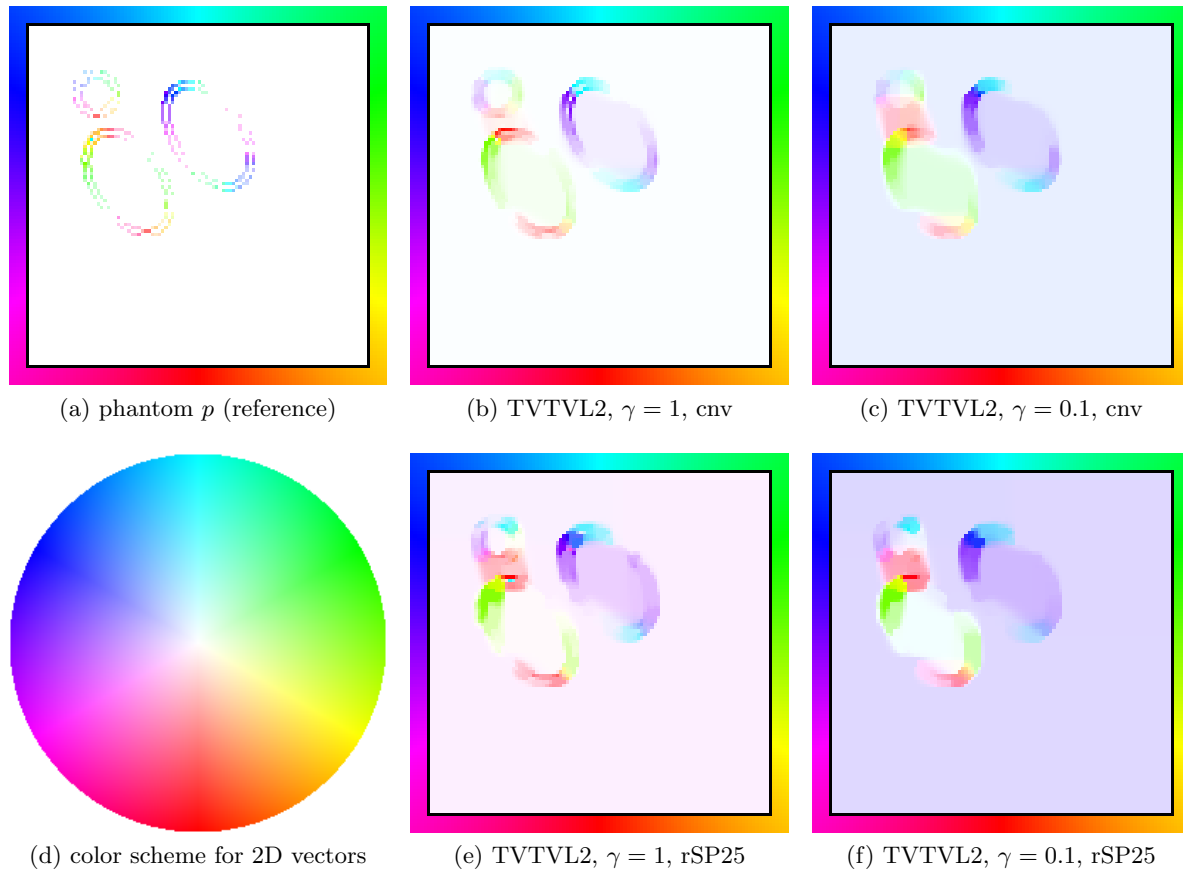
(a) phantom p (reference)   (b) TVTVL2, $\gamma = 1$, cnv   (c) TVTVL2, $\gamma = 0.1$, cnv

(d) color scheme for 2D vectors   (e) TVTVL2, $\gamma = 1$, rSP25   (f) TVTVL2, $\gamma = 0.1$, rSP25

Figure 7. *Illustration of the reconstructed motion fields v for the results shown in Figure 5. First, each vector field is rescaled such that $\max_i \|v_i\|_2 = 1$. Then each pixel i is colored by mapping direction and norm of v_i to the color scheme displayed in (d). To further ease the visualization, a colored frame was added to each motion image to depict the colors corresponding to a vector pointing from the middle of the image to the pixels of the frame. The motion field for the phantom (a) was computed by solving (4.5b) with the true p as input and $\gamma = 1$, $\beta = 10^{-6}$.*

time points 10–400. More details can be found in [1]. Note that the signal recorded by the FP sensor is only proportional to the acoustic pressure. To obtain absolute pressure values, one would need to calibrate it with an ultrasound transducer prior to the measurement. While this is necessary to perform quantitative, spectroscopic inference in a second analysis step [17, 27], we did not do it here and all images shown can be considered in arbitrary units.

For the inversion, we assume a homogeneous sound speed of $1540 \mathrm{m\,s}^{-1}$ and use a 3D spatial grid of dimensions $44 \times 264 \times 264$ with grid size 75µm (the reason for this upsampling in space is the oversampling in time, explained in [1]). Reconstructions from the full, conventional data will again be denoted by "cnv" and will be used to provide a ground truth. The subsampled data is generated using the same scheme (5.1) as for the simulated data, except that $M_{\mathrm{sub}} = 16$. The subsampling operators are repeated periodically, i.e., $C_{M_{\mathrm{sub}}+i} = C_i$.
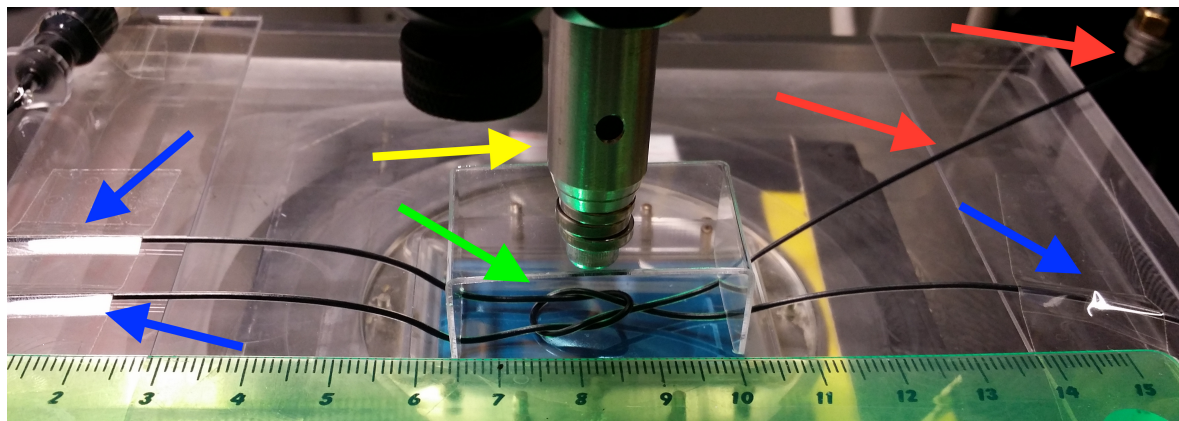
Figure 8. *Experimental setup for dynamic, stop-motion phantom: The two polythene tubes are immersed in a bath of intralipid solution placed on the FP sensor plane (green arrow). Three of the tube ends are fixed (blue arrows), while one is tied to a motor shaft (red arrows). The excitation laser (yellow arrow) is illuminating from the top.*

5.2.2. Experimental results. We used the same strategy to choose the regularization parameters as before. For the TVTVL2 model, we choose $\alpha = \beta = \hat{\alpha}$, $\gamma = 0.1$, where $\hat{\alpha}$ is the regularization parameter for TV-fbf that yields a good compromise between removing noise, subsampling, and image features (cf. Figure 2). Figure 9 shows the results after 20 iterations (index i) of the accelerated proximal gradient descent. Again, we can see a significant improvement when using the simultaneous motion estimation introduced by TVTVL2 compared to TV-fbf. The motion of our phantom has two dominant components: a translation component resulting from pulling the whole knot towards the motor shaft by one tube end, and a component describing the contraction resulting from the three other tube ends being fixed. To examine the later component, we suppress the translation by subtracting the mean motion vector in every frame $\bar{v}_t = N^{-1} \sum_i (v_t)_i$. Figure 10 shows the remaining parts of the motion. Both of the fields reconstructed from full and from subsampled data accurately describe the contraction. The coloring indicates that the tubes move towards each other, i.e., the knot contracts.

5.3. Optimization. As noted earlier, the results we showed up to now were computed using ADMM in both the p update (4.5a) and the v update (4.5b) in the TVTVL2-regularized denoising problem (4.4). In the v update, AMG-CG was used as a least-squares solver. In this section, we justify this choice retrospectively. Due to the large number of different parameters ACS, PDHG, and ADMM have, this is not an exhaustive comparison. We tuned all parameters we do not explicitly mention to best performance and made sure that all methods make best use of the computational platform we used (Intel Xeon CPU with 12 cores at 2.70GHz, 256GB RAM). Another problem is caused by the nonconvexity of (4.4), which adds an arbitrary element to such a comparison: In principle, one would need to test all methods on a large number of inputs and initializations and compare average performances. Again, we restrict ourselves here to the two concrete examples we presented in the previous two sections, and in each of those, we only examine the computation of the TVTVL2-regularized denoising problem
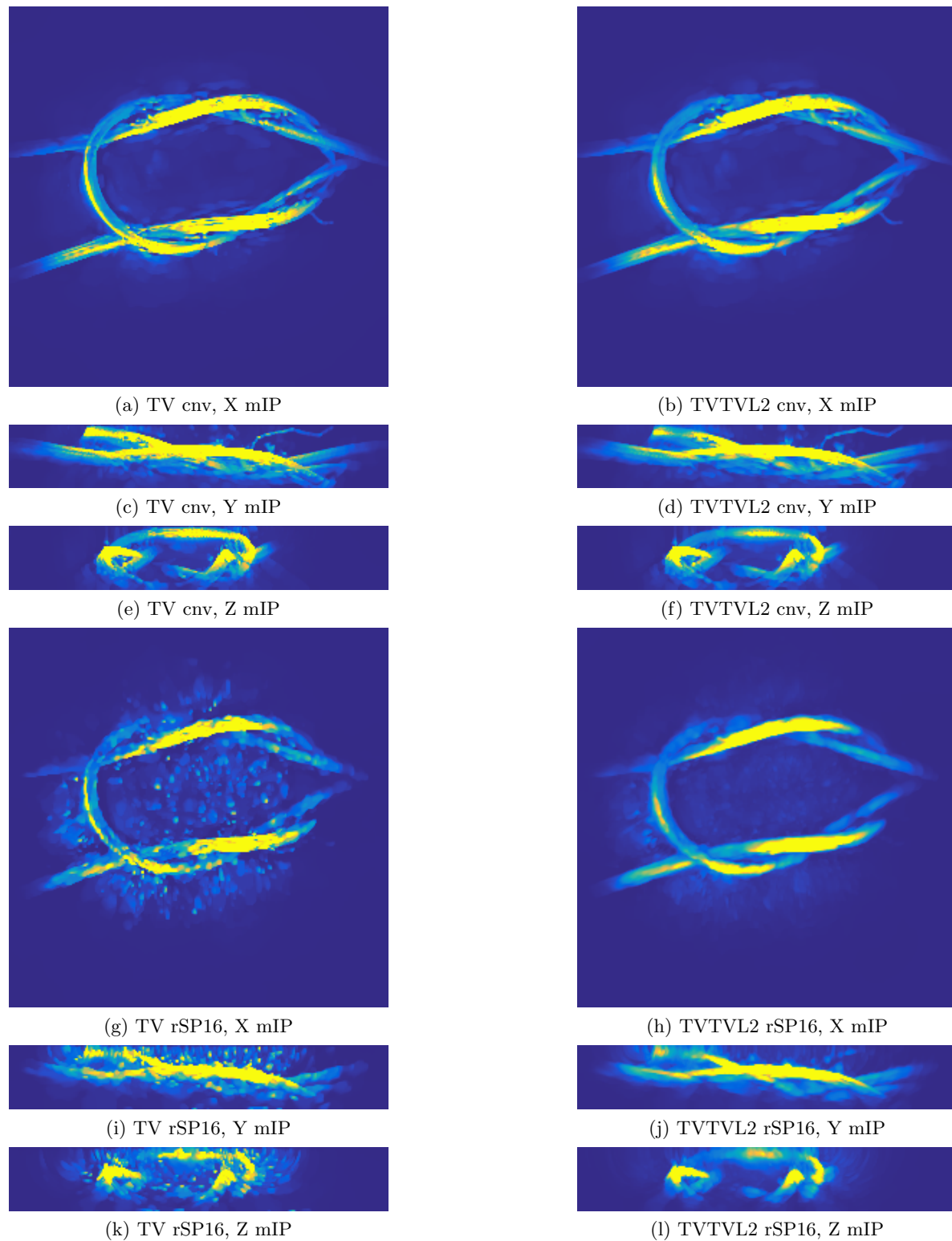
(a) TV cnv, X mIP

(b) TVTVL2 cnv, X mIP

(c) TV cnv, Y mIP

(d) TVTVL2 cnv, Y mIP

(e) TV cnv, Z mIP

(f) TVTVL2 cnv, Z mIP

(g) TV rSP16, X mIP

(h) TVTVL2 rSP16, X mIP

(i) TV rSP16, Y mIP

(j) TVTVL2 rSP16, Y mIP

(k) TV rSP16, Z mIP

(l) TVTVL2 rSP16, Z mIP

Figure 9. *Snapshots at $t = 23$ of the reconstructed pressure p for full (cvn) and subsampled (rSP-16) experimental data.*
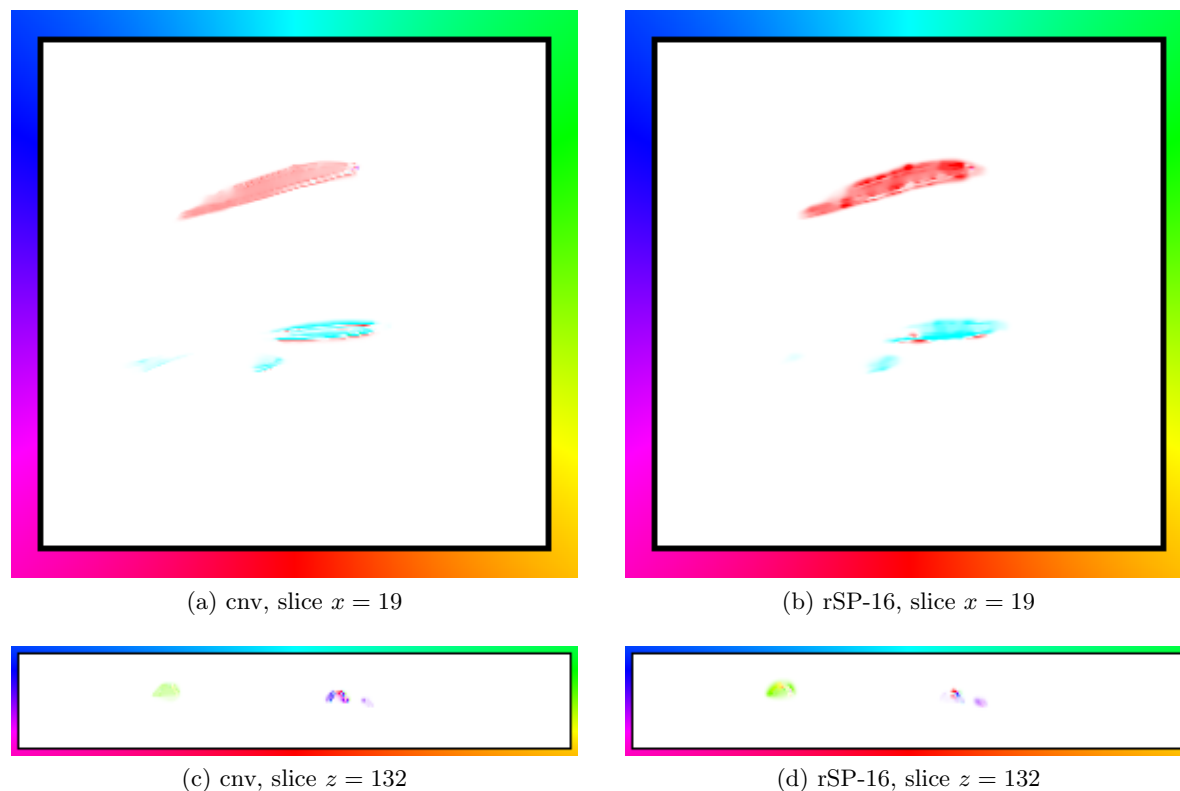
(a) cnv, slice $x = 19$

(b) rSP-16, slice $x = 19$

(c) cnv, slice $z = 132$

(d) rSP-16, slice $z = 132$

Figure 10. *Snapshots of the nontranslational part of the motion field $v^{\mathrm{NT}} = v - N^{-1} \sum_i v_i$ reconstructed by the TVTVL2 method for full (cvn) and subsampled (rSP-16) experimental data between frames $t = 22$ and $t = 23$. For each image, a single slice along a particular dimension was extracted, and only the components of the vector field in the remaining two dimensions is depicted here. The resulting 2D vector field is color-coded in the same way as in Figure 7.*

(4.4) arising from the first iteration, $i = 1$, of the forward-backward splitting (4.2a)–(4.2b). As $p^0 = 0$, this means we examine $\tilde{p}_t = \nu A^T C^T f_t^c$ as an input in (4.4). All other variables are initialized to 0. Figure 11 compares the decay of the denoising energy $\tilde{\mathcal{E}}(p, v)$ over computation time for the four different combinations of using PDHG and ADMM for each of the substeps. In 2D (Figure 11(a)), the convergence speed of the different combinations is quite similar, and the different energy levels they reach correspond to the different local minima they end up in. In 3D, the situation is quite different: Figure 11(b) shows that using PDHG for the v update (4.5b) leads to prohibitively long computations times. While PDHG performs well for the p update (4.5a) in this study, we also encountered scenarios where this is not the case. This observation was the main reason we considered using the more complicated ADMM methods in the first place: We started off by using PDHG for both subproblems like in [10] based on the corresponding code available on github.[2] While this worked for 2D scenarios, we encountered severe difficulties for 3D scenarios, which we were only able to overcome by implementing the
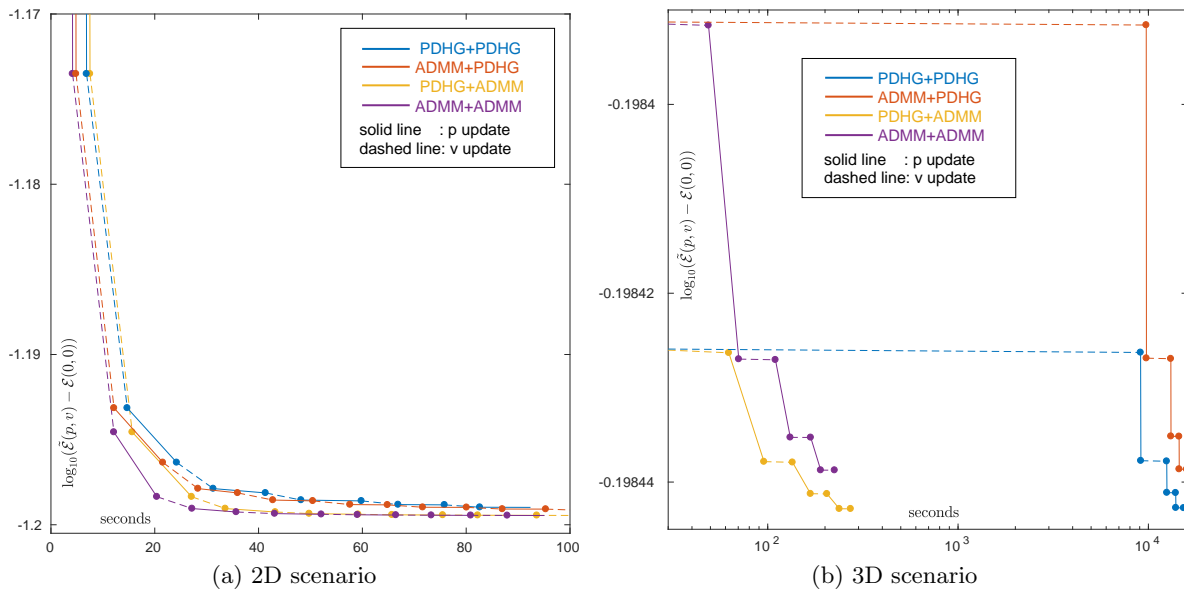
---

[2]https://github.com/HendrikMuenster/JointMotionEstimationAndImageReconstruction

(a) 2D scenario   (b) 3D scenario

Figure 11. *Comparison of different methods to solve the biconvex optimization problem* (4.4) *via ACS for* (a) *the* 2D *example described in section* 5.1 *and* (b) $T = 10$ *frames of the* 3D *scenario described in section* 5.2. *The plots display the decay in energy* $\tilde{\mathcal{E}}(p, v)$ *relative to the initialization with* $p = 0$, $v = 0$ *vs. computational time in seconds (in logarithmic scale in* (b)). *Solid parts of the line plots correspond to the* p *update* (4.5a), *and dashed parts to the* v *update* (4.5b). *A total of* 4 *ACS alternations is displayed. "ADMM-PDHG" refers to using ADMM for the* p *update and PDHG for the* v *update.*

tailored ADMM implementations presented here.

The main difficulty in both ADMM methods is in solving the least-squares problems (4.16) and (4.17) by a fast iterative method. As explained in section 4.4, the v update (4.17) can be solved frame by frame, which allows one to explicitly set up the system matrix and use efficient preconditioning techniques. To compare them, we set p_{t+1}^{j+1} and p_t^{j+1} in (4.17) to the TV-fbf solutions shown in Figure 2 (note that $E = E(p_t^{j+1})$). Figure 12 shows the results, which demonstrate that for linear systems arising from regularized 3D optical flow estimation, AMG-CG is a powerful solver. Note, however, that this comes with increased memory costs: the system matrix is 1.24GB large and the corresponding AMG preconditioner we chose here is 6.75GB large (there is also a little computational overhead in computing them, but as they do not change over the whole ADMM scheme, this is typically negligible).

6. Discussion, outlook, and conclusion.

6.1. Discussion and outlook.
The results for both simulated and experimental data clearly demonstrate that a significant improvement of image quality over fbf reconstructions (2.5) that only use spatial sparsity constraints can be obtained when using a generic spatio-temporal approach based on simultaneous, sparsity-constrained motion estimation (3.2). Furthermore, the reconstructed motion fields provide additional information on the dynamics that can be useful for subsequent analysis. While these dynamic parameters look qualitatively correct, even from subsampled data, further investigations have to examine whether
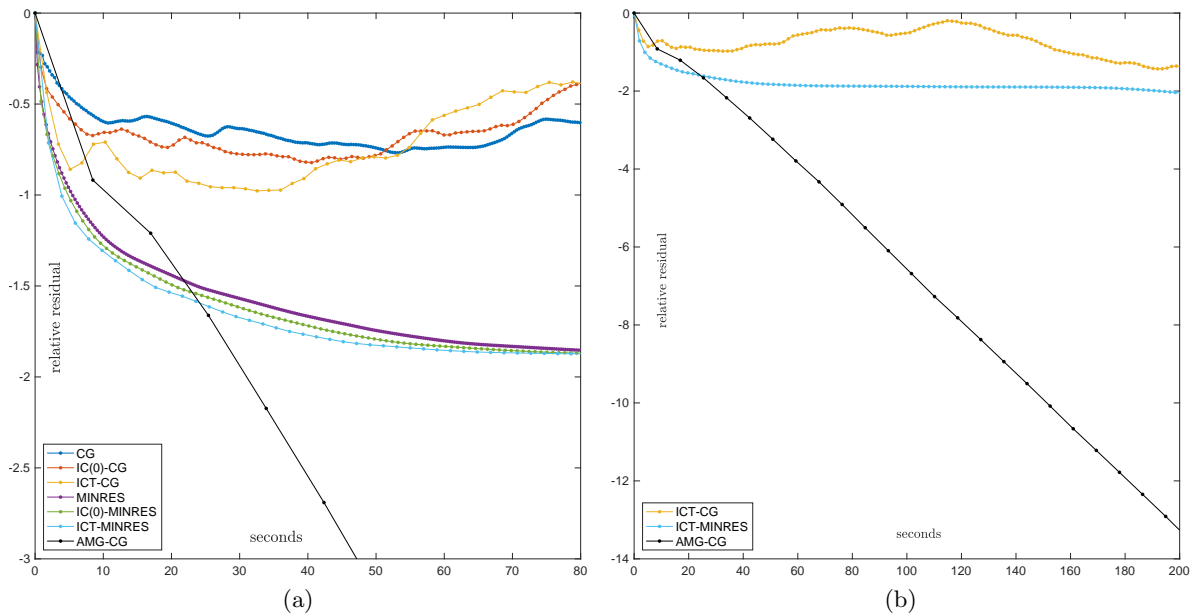
Figure 12. (a) *Comparison between different iterative methods and preconditioners to solve* (4.17) *(see section* 5.3 *for the details of the setup). The vertical axis shows the relative residual, while the horizontal axis shows the computation time on a single CPU core averaged over* 10 *repetitions.* (b) *The same plot with only a subset of the solvers and expanded axis.*

they are also quantitatively correct. For this first proof-of-concept study, we used very generic regularization functionals in space (TV) and a generic motion model based on a simple continuity equation (3.3). As we already obtained promising results with this rather unspecific model, we want to investigate the use of tailored motion models that better reflect the real physics of the underlying motion for a concrete application. In addition, we chose to measure the distance to the discretized motion PDE in the squared L_2-norm; cf. (3.4). While this is computationally advantageous, studies generalizing this to L_p-norms, e.g., for $p = 1$, have shown promising results and directions for future research [20, 10, 9].

The main drawback of the concrete TVTVL2 model we used here (3.5) is that it leads to a challenging, large-scale biconvex optimization problem. Even with the tailored ADMM schemes we developed (cf. section 4.4), computing the 3D reconstructions presented in section 5.2 took 4 days and 6 hours on a powerful workstation (Intel Xeon CPU with 12 cores at 2.70GHz, 256GB RAM, Tesla K40 GPU) compared to 6 hours and 34 minutes for the TV-fbf reconstruction. There are several possibilities to close this gap:

- As a simple block alternation, the ACS scheme can be modified by introducing techniques like overrelaxation, inertia methods, or line search.
- For solving substep (4.5b), developing an ADMM scheme that uses an algebraic multigrid preconditioner was crucial. However, the high memory demand of this approach limits the number of frames which can be computed in parallel. Using *geometric multigrid* preconditioning instead could keep the fast convergence (cf. Figure 12) while requiring much less memory [8].

- If the nonsmooth sparsity constraints are approximated by smooth functionals such as the Huber functional, fast, monotone solvers can be used to solve (4.5a) and (4.5b); see, e.g., [61].

Another potential problem is that the ACS scheme presented in section 4.2 will only converge to a local minimum of the biconvex variational energy (3.5). Figure 11 showed that already the choice of the convex optimization scheme to solve (4.5a) and (4.5b) can influence which local minimum is found. Other parameters like the accuracy with which these problems are solved, how the schemes are initialized, whether the scenario is 2D or 3D, etc., have an often nontrivial influence as well. In future work, we plan to examine these issues in a systematic way.

From a modeling perspective, the simple optical flow discretization (3.4) we chose here can only resolve small motions: If the support of $p_{t+1} - p_t$ and $\nabla^\pm p_t$ do not overlap, v_t cannot minimize $\|p_{t+1} - p_t + (\nabla p_t) \cdot v_t\|_2^2$. An extension of the framework to estimate large-scale motions is described in [21].

This article focused on the mathematical and computational aspects of 4D PAT. We therefore assumed here that there is a generic binning of the sequence of acoustic measurements (2.2) into temporal bins during which the target can be considered static (and used phantoms for which this holds true) and only compared image quality for a fixed subsampling factor. However, in reality, sequential scanners measure a single time pressure course for every pulse of the excitation laser. The temporal binning of this stream of acquisitions leads to a more complicated interplay between artifacts arising from subsampling, motion blur, and the spatio-temporal continuity imposed by the variational model. We will examine this issue more closely in forthcoming work that will focus on the technical and practical aspects of 4D PAT with novel acoustic scanners [41, 40]. For the application to in-vivo imaging, additional challenges need to be addressed, such as heterogeneous tissue properties.

6.2. Conclusion. In this work, we extended our earlier results on using compressed sensing techniques to accelerate high-resolution 3D PAT acquisition with sequential scanners [1]. We demonstrated that in the context of dynamic PAT, another substantial increase of image quality can be obtained by using a generic variational framework that couples sparsity-constrained image reconstruction and simultaneous, sparsity-constrained motion estimation. In particular, we considered a motion model based on the popular optical flow equation and used the total variation functional as sparsity constraints. For this, promising results for simulated and experimental data were obtained in a proof-of-concept study that justifies further research in this field. A major challenge for using these variational approaches for large-scale 4D inverse problems with complicated forward operators are the computational demands of the corresponding optimization routines. We described and examined a set of related methods that can be used as a starting point to implement similar strategies for other applications.

Appendix A. Proximal operators. An extensive overview on how to use and compute proximal operators (4.1) is given in [16]. The splits we use in this work have been introduced such that the functionals for which we have to compute the proximal operators decouple over space and time into the sum of 1- or d-dimensional functionals $\phi(x)$ or $\phi(x_1, \ldots, x_d)$. As such, all proximal operators can be computed explicitly and pointwise in space and time; i.e., for an image/vector field sequence $x \in \mathbb{R}^{NT}/x \in \mathbb{R}^{dNT}$, the proximal operators can be computed

by solving NT subproblems of dimension $1/d$ using explicit formulae.

For $\mathcal{G}(x)$ in (4.9b) this leads to

$$(A.1) \qquad \phi(x) = \chi_+(x) + (x - z)^2 \,, \qquad \mathrm{prox}_{\alpha\phi}(\tilde{x}) = \max\left(0, \alpha z + \frac{\tilde{x}}{\alpha + 1}\right) \,.$$

The proximal operator for the functional $\mathcal{G}(x)$ in (4.10b) is a d-dimensional quadratic problem:

$$(A.2) \qquad \phi(x) = \frac{1}{2}(z + c_1 x_1 + \cdots + c_d x_d)^2 \,,$$

$$(A.3) \qquad \mathrm{prox}_{\alpha\phi}(\tilde{x}) = \operatorname*{argmax}_{x \in \mathbb{R}^d}\left\{\frac{\alpha}{2}\left(z + \sum_i^d c_i x_i\right)^2 + \frac{1}{2}\sum_i^d (x_i - \tilde{x}_i)^2\right\}.$$

Its optimality condition leads to a d-dim linear system, which we show here for $d = 3$:

$$(A.4) \qquad \begin{bmatrix} (1 + \alpha c_1^2) & \alpha c_1 c_2 & \alpha c_1 c_3 \\ \alpha c_2 c_1 & (1 + \alpha c_2^2) & \alpha c_2 c_3 \\ \alpha c_3 c_1 & \alpha c_3 c_2 & (1 + \alpha c_3^2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 - \alpha c_1 z \\ \tilde{x}_2 - \alpha c_2 z \\ \tilde{x}_3 - \alpha c_3 z \end{bmatrix} \,.$$

It can be solved explicitly for $d = 2, 3$, and for its use within the PDHG scheme, most relevant terms can be precomputed.

The ℓ_1-norms involved in the isotropic TV terms are actually global ℓ_1-norms of the local ℓ_2-norms of the gradient vectors. For a gradient field of image z represented as $y \in R^{dN}$ indexed as $y_{x_i,j}$ for derivative direction and location index, respectively, we have

$$(A.5) \qquad \|\nabla^+ z\|_1 = \sum_j^N \sqrt{\sum_i^d y_{x_i,j}^2} \qquad \Longrightarrow \qquad \phi(x) = \sqrt{x_1 + \cdots + x_d} \,,$$

$$(A.6) \qquad \mathrm{prox}_{\alpha\phi}(\tilde{x}) = \begin{cases} \max(\phi(\tilde{x}) - \alpha, 0)\, \tilde{x}/\phi(\tilde{x}) & \text{if} \quad \phi(\tilde{x}) > 0 \,, \\ 0 & \text{else} \,. \end{cases}$$

With this, one can easily build the proximal operator for (4.10c) and (4.15c).

Next we need the convex conjugates $\mathcal{F}^*(y)$ and their proximal mappings in some places. For the isotropic TV term, we have

$$(A.7) \qquad \mathcal{F}(y) = \alpha\, \|y\|_1 = \alpha \sum_j^N \sqrt{\sum_i^d y_{x_i,j}^2} \,, \quad \mathcal{F}^*(y) = \sum_j^N \alpha \chi_{[0,1]}\left(\frac{1}{\alpha}\sqrt{\sum_i^d y_{x_i,j}^2}\right) \,,$$

which means that $\mathcal{F}^*(y)$ is 0 if all gradient vectors have an amplitude that is smaller than α, and ∞ otherwise; see, e.g., [13]. As such, the proximal operator is just a projection:

$$(A.8) \qquad \phi(y) = \chi_{[0,1]}\left(\frac{1}{\alpha}\sqrt{\sum_i^d y_{x_i}^2}\right) \quad \Longrightarrow \quad \mathrm{prox}_{\beta\phi}(\tilde{y}) = \frac{\tilde{y}}{\max\left(1, \frac{1}{\alpha}\sqrt{\sum_i^d \tilde{y}_{x_i}^2}\right)} \,.$$

The second part of $\mathcal{F}(y)$ in (4.10b) is $\frac{\tilde{\gamma}}{2}\|y_2\|_2^2$. Its convex conjugate is given by $\frac{1}{2\tilde{\gamma}}\|y_2\|_2^2$ and the proximal mapping can be computed using

$$(A.9) \qquad \phi(y) = \frac{y^2}{2\tilde{\gamma}} \quad \Longrightarrow \quad \text{prox}_{\alpha\phi}(\tilde{y}) = \frac{\tilde{\gamma}}{\tilde{\gamma} + \alpha}\,\tilde{y}\;.$$

REFERENCES

[1] S. Arridge, P. Beard, M. Betcke, B. Cox, N. Huynh, F. Lucka, O. Ogunlade, and E. Zhang, *Accelerated high-resolution photoacoustic tomography via compressed sensing*, Phys. Med. Biol., 61 (2016), 8908, http://stacks.iop.org/0031-9155/61/i=24/a=8908.

[2] S. Arridge, M. Betcke, B. Cox, F. Lucka, and B. Treeby, *On the adjoint operator in photoacoustic tomography*, Inverse Problems, 32 (2016), 115012, http://stacks.iop.org/0266-5611/32/i=11/a=115012.

[3] S. R. Arridge and O. Scherzer, *Imaging from coupled physics*, Inverse Problems, 28 (2012), 080201, http://stacks.iop.org/0266-5611/28/i=8/a=080201.

[4] P. Beard, *Biomedical photoacoustic imaging*, Interface Focus, 1 (2011), pp. 602–631, https://doi.org/10.1098/rsfs.2011.0028.

[5] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process., 18 (2009), pp. 2419–2434, https://doi.org/10.1109/TIP.2009.2028250.

[6] Y. E. Boink, M. J. Lagerwerf, W. Steenbergen, S. A. van Gils, S. Manohar, and C. Brune, *A framework for directional and higher-order reconstruction in photoacoustic tomography*, Phys. Med. Biol., 63 (2018), 045018, http://iopscience.iop.org/article/10.1088/1361-6560/aaaa4a/meta.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122, https://doi.org/10.1561/2200000016.

[8] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr, *A multigrid platform for real-time motion computation with discontinuity-preserving variational methods*, Int. J. Comput. Vision, 70 (2006), pp. 257–277, https://doi.org/10.1007/s11263-006-6616-7.

[9] M. Burger, H. Dirks, L. Frerking, A. Hauptmann, T. Helin, and S. Siltanen, *A variational reconstruction method for undersampled dynamic x-ray tomography based on physical motion models*, Inverse Problems, 33 (2017), 124008, http://iopscience.iop.org/article/10.1088/1361-6420/aa99cf.

[10] M. Burger, H. Dirks, and C.-B. Schönlieb, *A variational model for joint motion estimation and image reconstruction*, SIAM J. Imaging Sci., 11 (2018), pp. 94–128, https://doi.org/10.1137/16M1084183.

[11] M. Burger, A. Sawatzky, and G. Steidl, *First order algorithms in variational image processing*, in Splitting Methods in Communication, Imaging, Science, and Engineering, R. Glowinski, S. Osher, and W. Yin, eds., Springer, Cham, 2016, https://doi.org/10.1007/978-3-319-41589-5_10.

[12] E. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509, https://doi.org/10.1109/TIT.2005.862083.

[13] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[3]https://github.com/HendrikMuenster/JointMotionEstimationAndImageReconstruction

[14] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319, https://doi.org/10.1017/S096249291600009X.

[15] J. Chung and L. Nguyen, *Motion estimation and correction in photoacoustic tomographic reconstruction*, SIAM J. Imaging Sci., 10 (2017), pp. 216–242, https://doi.org/10.1137/16M1082901.

[16] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, New York, 2011, pp. 185–212, https://doi.org/10.1007/978-1-4419-9569-8_10.

[17] B. Cox, J. Laufer, S. Arridge, and P. Beard, *Quantitative spectroscopic photoacoustic imaging: A review*, J. Biomed. Optics, 17 (2012), 061202, https://doi.org/10.1117/1.JBO.17.6.061202.

[18] B. T. Cox, S. Kara, S. R. Arridge, and P. C. Beard, *k-space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics*, J. Acoust. Soc. Amer., 121 (2007), pp. 3453–3464.

[19] X. L. Dean-Ben, S. Gottschalk, B. Mc Larney, S. Shoham, and D. Razansky, *Advanced opto-acoustic methods for multiscale imaging of in vivo dynamics*, Chem. Soc. Rev., 46 (2017), pp. 2158–2198, https://doi.org/10.1039/C6CS00765A.

[20] H. Dirks, *Variational Methods for Joint Motion Estimation and Image Reconstruction*, Ph.D. thesis, Institute for Computational and Applied Mathematics, University of Muenster, 2015.

[21] H. Dirks, *Joint Large-Scale Motion Estimation and Image Reconstruction*, preprint, https://arxiv.org/abs/1610.09908, 2016.

[22] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[23] R. Ellwood, F. Lucka, E. Zhang, P. Beard, and B. Cox, *Photoacoustic imaging with a multi-view Fabry-Perot scanner*, Proc. SPIE, 10064 (2017), 100641F, https://doi.org/10.1117/12.2252728.

[24] R. Ellwood, O. Ogunlade, E. Zhang, P. Beard, and B. Cox, *Photoacoustic tomography using orthogonal Fabry-Perot sensors*, J. Biomed. Optics, 22 (2016), 041009, https://doi.org/10.1117/1.JBO.22.4.041009.

[25] E. Esser, *Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman*, tech. report, University of California, Irvine, 2009.

[26] D. Finch, S. K. Patch, and Rakesh, *Determining a function from its mean values over a family of spheres*, SIAM J. Math. Anal., 35 (2004), pp. 1213–1240, https://doi.org/10.1137/S0036141002417814.

[27] M. Fonseca, E. Malone, F. Lucka, R. Ellwood, L. An, S. Arridge, P. Beard, and B. Cox, *Three-dimensional photoacoustic imaging and inversion for accurate quantification of chromophore distributions*, Proc. SPIE, 10064 (2017), 1006415, https://doi.org/10.1117/12.2250964.

[28] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, Basel, 2013.

[29] T. Goldstein and S. Osher, *The split Bregman method for $L1$-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343, https://doi.org/10.1137/080725891.

[30] T. Goldstein, C. Studer, and R. Baraniuk, *A Field Guide to Forward-Backward Splitting with a FASTA Implementation*, preprint, https://arxiv.org/abs/1411.3406, 2014.

[31] J. Gorski, F. Pfeuffer, and K. Klamroth, *Biconvex sets and optimization with biconvex functions: A survey and extensions*, Math. Methods Oper. Res., 66 (2007), pp. 373–407, https://doi.org/10.1007/s00186-007-0161-1.

[32] Z. Guo, C. Li, L. Song, and L. V. Wang, *Compressed sensing in photoacoustic tomography in vivo*, J. Biomed. Optics, 15 (2010), 021311, https://doi.org/10.1117/1.3381187.

[33] B. N. Hahn, *Dynamic linear inverse problems with moderate movements of the object: Ill-posedness and regularization*, Inverse Probl. Imaging, 9 (2015), pp. 395–413, https://doi.org/10.3934/ipi.2015.9.395.

[34] B. N. Hahn, *Null space and resolution in dynamic computerized tomography*, Inverse Problems, 32 (2016), 025006, http://stacks.iop.org/0266-5611/32/i=2/a=025006.

[35] B. N. Hahn and E. T. Quinto, *Detectable singularities from dynamic radon data*, SIAM J. Imaging Sci., 9 (2016), pp. 1195–1225, https://doi.org/10.1137/16M1057917.

[36] 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, 2010, pp. 716–719, https://doi.org/10.1109/ISBI.2010.5490076.

[37] M. Holler and K. Kunisch, *On infimal convolution of TV-type functionals and applications to video*

and image reconstruction, SIAM J. Imaging Sci., 7 (2014), pp. 2258–2300, https://doi.org/10.1137/130948793.

[38] B. K. HORN AND B. G. SCHUNCK, *Determining optical flow*, Artificial Intelligence, 17 (1981), pp. 185–203, https://doi.org/10.1016/0004-3702(81)90024-2.

[39] C. HUANG, K. WANG, L. NIE, L. WANG, AND M. ANASTASIO, *Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media*, IEEE Trans. Med. Imaging, 32 (2013), pp. 1097–1110, https://doi.org/10.1109/TMI.2013.2254496.

[40] N. HUYNH, F. LUCKA, E. ZHANG, M. BETCKE, S. ARRIDGE, P. BEARD, AND B. COX, *Sub-sampled Fabry-Perot photoacoustic scanner for fast 3D imaging*, Proc. SPIE, 10064 (2017), 100641Y, https://doi.org/10.1117/12.2250868.

[41] N. HUYNH, O. OGUNLADE, E. ZHANG, B. COX, AND P. BEARD, *Photoacoustic imaging using an 8-beam Fabry-Perot scanner*, Proc. SPIE, 9708 (2016), 9708, https://doi.org/10.1117/12.2214334.

[42] J. P. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, Appl. Math. Sci. 160, Springer, New York, 2005, https://doi.org/10.1007/b138659.

[43] C. L, *iFEM: An Integrated Finite Element Method Package in MATLAB*, tech. report, University of California, Irvine, 2009.

[44] T. MAST, L. SOURIAU, D.-L. LIU, M. TABEI, A. NACHMAN, AND R. WAAG, *A k-space method for large-scale models of wave propagation in tissue*, IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, 48 (2001), pp. 341–354, https://doi.org/10.1109/58.911717.

[45] J. MCCLELLAND, D. HAWKES, T. SCHAEFFTER, AND A. KING, *Respiratory motion models: A review*, Med. Image Anal., 17 (2013), pp. 19–42, https://doi.org/10.1016/j.media.2012.09.005.

[46] J. MENG, L. V. WANG, D. LIANG, AND L. SONG, *In vivo optical-resolution photoacoustic computed tomography with compressed sensing*, Opt. Lett., 37 (2012), pp. 4573–4575, https://doi.org/10.1364/OL.37.004573.

[47] J. MENG, L. V. WANG, L. YING, D. LIANG, AND L. SONG, *Compressed-sensing photoacoustic computed tomography in vivo with partially known support*, Opt. Express, 20 (2012), pp. 16510–16523, https://doi.org/10.1364/OE.20.016510.

[48] L. NIE AND X. CHEN, *Structural and functional photoacoustic molecular tomography aided by emerging contrast agents*, Chem. Soc. Rev., 43 (2014), pp. 7132–70, https://doi.org/10.1039/c4cs00086b.

[49] T. POCK AND A. CHAMBOLLE, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in 2011 International Conference on Computer Vision, IEEE, 2011, pp. 1762–1769, https://doi.org/10.1109/ICCV.2011.6126441.

[50] T. POCK, D. CREMERS, H. BISCHOF, AND A. CHAMBOLLE, *An algorithm for minimizing the Mumford-Shah functional*, in 12th International IEEE Conference on Computer Vision, IEEE, 2009, pp. 1133–1140, https://doi.org/10.1109/ICCV.2009.5459348.

[51] J. PROVOST AND F. LESAGE, *The application of compressed sensing for photo-acoustic tomography*, IEEE Trans. Med. Imaging, 28 (2009), pp. 585–594, https://doi.org/10.1109/TMI.2008.2007825.

[52] S. RAVISHANKAR, B. E. MOORE, R. R. NADAKUDITI, AND J. A. FESSLER, *Low-rank and adaptive sparse signal (LASSI) models for highly accelerated dynamic imaging*, IEEE Trans. Med. Imaging, 36 (2017), pp. 1116–1128, https://doi.org/10.1109/TMI.2017.2650960.

[53] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003, https://doi.org/10.1137/1.9780898718003.

[54] M. SCHLOEGL, M. HOLLER, A. SCHWARZL, K. BREDIES, AND R. STOLLBERGER, *Infimal convolution of total generalized variation functionals for dynamic MRI*, Magnetic Resonance in Medicine, 78 (2017), pp. 142–155, https://doi.org/10.1002/mrm.26352.

[55] U. SCHMITT AND A. K. LOUIS, *Efficient algorithms for the regularization of dynamic inverse problems:* I. *Theory*, Inverse Problems, 18 (2002), pp. 645–658, http://stacks.iop.org/0266-5611/18/i=3/a=308.

[56] U. SCHMITT, A. K. LOUIS, C. H. WOLTERS, AND M. VAUHKONEN, *Efficient algorithms for the regularization of dynamic inverse problems:* II. *Applications*, Inverse Problems, 18 (2002), pp. 659–676, http://stacks.iop.org/0266-5611/18/i=3/a=309.

[57] A. SOLONEN, T. CUI, J. HAKKARAINEN, AND Y. MARZOUK, *On dimension reduction in Gaussian filters*, Inverse Problems, 32 (2016), 045003, http://stacks.iop.org/0266-5611/32/i=4/a=045003.

[58] B. E. TREEBY AND B. T. COX, *k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields*, J. Biomed. Optics, 15 (2010), 021314, https://doi.org/10.1117/1.3360308.

[59] B. E. TREEBY, E. Z. ZHANG, AND B. T. COX, *Photoacoustic tomography in absorbing acoustic media*

using time reversal, Inverse Problems, 26 (2010), 115003, http://stacks.iop.org/0266-5611/26/i=11/a=115003.

[60] B. TREMOULHEAC, N. DIKAIOS, D. ATKINSON, AND S. ARRIDGE, *Dynamic MR image reconstruction: Separation from undersampled* $(\mathbf{k}, t)$*-space via low-rank plus sparse prior*, IEEE Trans. Med. Imaging, 33 (2014), pp. 1689–1701, https://doi.org/10.1109/TMI.2014.2321190.

[61] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002, https://doi.org/10.1137/1.9780898717570.

[62] K. WANG, J. XIA, C. LI, L. V. WANG, AND M. A. ANASTASIO, *Fast spatiotemporal image reconstruction based on low-rank matrix estimation for dynamic photoacoustic computed tomography*, J. Biomed. Optics, 19 (2014), 056007.

[63] L. V. WANG, *Multiscale photoacoustic microscopy and computed tomography*, Nature Photonics, 3 (2009), pp. 503–509, https://doi.org/10.1038/nphoton.2009.157.

[64] Y. XU AND L. V. WANG, *Application of time reversal to thermoacoustic tomography*, in Biomedical Optics 2004, International Society for Optics and Photonics, 2004, pp. 257–263.

[65] Y. ZHANG, Y. WANG, AND C. ZHANG, *Total variation based gradient descent algorithm for sparse-view photoacoustic image reconstruction*, Ultrasonics, 52 (2012), pp. 1046–1055, https://doi.org/10.1016/j.ultras.2012.08.012.

[66] Y. ZHOU, J. YAO, AND L. V. WANG, *Tutorial on photoacoustic tomography*, J. Biomed. Optics, 21 (2016), 061007, https://doi.org/10.1117/1.JBO.21.6.061007.