

Spatiotemporal image reconstruction to enable high-frame-rate dynamic photoacoustic tomography with rotating-gantry volumetric imagers

Refik Mert Cam,^{a,†} Chao Wang,^{b,†} Weylan Thompson^{b,c}, Sergey A. Ermilov^{b,c},
Mark A. Anastasio^{b,d} and Umberto Villa^{b,e,*}

^aUniversity of Illinois Urbana-Champaign, Department of Electrical and Computer Engineering, Urbana, Illinois, United States

^bNational University of Singapore, Department of Statistics and Data Science, Singapore

^cPhotoSound Technologies Inc., Houston, Texas, United States

^dUniversity of Illinois Urbana-Champaign, Department of Bioengineering, Urbana, Illinois, United States

^eThe University of Texas at Austin, Oden Institute for Computational Engineering and Sciences, Austin, Texas, United States

ABSTRACT. **Significance:** Dynamic photoacoustic computed tomography (PACT) is a valuable imaging technique for monitoring physiological processes. However, current dynamic PACT imaging techniques are often limited to two-dimensional spatial imaging. Although volumetric PACT imagers are commercially available, these systems typically employ a rotating measurement gantry in which the tomographic data are sequentially acquired as opposed to being acquired simultaneously at all views. Because the dynamic object varies during the data-acquisition process, the sequential data-acquisition process poses substantial challenges to image reconstruction associated with data incompleteness. The proposed image reconstruction method is highly significant in that it will address these challenges and enable volumetric dynamic PACT imaging with existing preclinical imagers.

Aim: The aim of this study is to develop a spatiotemporal image reconstruction (STIR) method for dynamic PACT that can be applied to commercially available volumetric PACT imagers that employ a sequential scanning strategy. The proposed reconstruction method aims to overcome the challenges caused by the limited number of tomographic measurements acquired per frame.

Approach: A low-rank matrix estimation-based STIR (LRME-STIR) method is proposed to enable dynamic volumetric PACT. The LRME-STIR method leverages the spatiotemporal redundancies in the dynamic object to accurately reconstruct a four-dimensional (4D) spatiotemporal image.

Results: The conducted numerical studies substantiate the LRME-STIR method's efficacy in reconstructing 4D dynamic images from tomographic measurements acquired with a rotating measurement gantry. The experimental study demonstrates the method's ability to faithfully recover the flow of a contrast agent with a frame rate of 10 frames per second, even when only a single tomographic measurement per frame is available.

Conclusions: The proposed LRME-STIR method offers a promising solution to the challenges faced by enabling 4D dynamic imaging using commercially available volumetric PACT imagers. By enabling accurate STIRs, this method has the potential to significantly advance preclinical research and facilitate the monitoring of critical physiological biomarkers.

*Address all correspondence to Umberto Villa, uvilla@oden.utexas.edu

†Refik Mert Cam and Chao Wang are equal contributors to this work and designated as co-first authors.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.29.S1.S11516](https://doi.org/10.1117/1.JBO.29.S1.S11516)]

Keywords: photoacoustic computed tomography; optoacoustic tomography; spatiotemporal image reconstruction; dynamic imaging; small animal imaging; low-rank matrix estimation

Paper 230303SSR received Sep. 30, 2023; revised Nov. 22, 2023; accepted Dec. 20, 2023; published Jan. 19, 2024.

1 Introduction

Photoacoustic computed tomography (PACT), also referred to as optoacoustic tomography, is an emerging and promising imaging modality with broad applications in the field of biomedical imaging.¹⁻³ By combining the high spatial resolution of ultrasound imaging with the high soft tissue contrast of optical imaging, PACT offers unique advantages for imaging biological structures while avoiding the use of ionizing radiation. In PACT, a fast laser pulse in the near infrared range illuminates the object. The absorption of optical energy by various molecules within the object (chromophore) induces a localized increase of acoustic pressure through the photoacoustic effect. The acoustic wavefield propagating through the object and coupling medium (water) is subsequently detected by ultrasonic transducers. The measured wavefield data can then be utilized to reconstruct an image that depicts the initial induced pressure distribution within the object.

In preclinical and clinical research, the ability to monitor dynamic physiological processes is of utmost importance for comprehending the progression of diseases and developing new treatments.⁴⁻⁶ For example, tumor vascular perfusion is a dynamic process that is critical in the study of cancers. High vascular perfusion is indicative of angiogenesis, a well-established hallmark of cancerous growth.^{7,8} Due to its noninvasive nature and the combination of optical contrast and spatial resolution at depths beyond the optical diffusion limit, PACT represents a promising imaging modality for monitoring critical dynamic physiological processes in preclinical and clinical research.⁹⁻¹³

Despite its considerable promise, current dynamic PACT technologies suffer from fundamental limitations. They often target two-dimensional (2D) spatial imaging due to shorter data acquisition times and computationally less demanding image reconstruction compared with 3D imaging.^{11,14-18} Most existing 3D PACT imagers developed to date utilize a rotating measurement geometry in which the tomographic data are sequentially acquired^{2,19-21} as opposed to being acquired simultaneously at all views. This design is advantageous because it reduces system costs by employing a limited number of acoustic transducers and associated electronics. However, data-acquisition times for a complete tomographic scan can be tens of seconds. Due to the relatively slow rotational speed, the temporal resolution is significantly limited. Although enhancing temporal resolution using sparsely sampled tomographic data is possible, the associated dynamic image reconstruction problem becomes ill-posed and highly challenging.

Previous studies on dynamic PACT^{10,11,14-18,22} have primarily focused on scenarios in which the sufficiently sampled tomographic data can be rapidly acquired. In such cases, a straightforward approach is to employ a frame-by-frame image reconstruction (FBFIR) method.^{10,11,14-18} These techniques utilize conventional static image reconstruction methods to estimate a sequence of images from sufficiently sampled tomographic data. The temporal resolution is limited by the duration of the complete data acquisition process. Rapid data acquisition is feasible either with 2D PACT imaging or by leveraging a dense, albeit expensive, static transducer array in 3D imaging. However, for volumetric imagers with sequential scanning strategy, FBFIR methods are not applicable, primarily due to the extended time required to accumulate the complete set of tomographic measurements. This limitation arises because conventional static image reconstruction techniques require densely sampled tomographic measurements for accurate object estimates; if sparsely sampled measurement data are used, the reconstructions suffer from severe artifacts.²³⁻²⁵

On the other hand, spatiotemporal image reconstruction (STIR) methods estimate a sequence of images simultaneously instead of frame-by-frame, and they have demonstrated their

efficacy in accurately reconstructing dynamic objects from sparsely sampled data in various medical imaging modalities, including computed tomography,²⁶ positron emission tomography,²⁷ single photon emission computed tomography,²⁸ and magnetic resonance imaging.²⁹ Although a few STIR techniques have been proposed for PACT, some of them still presume the availability of sufficiently sampled tomographic data and strive for enhanced accuracy and/or reduced computational complexity compared with FBFIR techniques.³⁰ The STIR methods,^{31–33} considering sparsely sampled tomographic data, rely on the principles of compressed sensing and require specific data sampling schemes that are different than the sequential sampling schemes employed by currently available volumetric imagers.

This study introduces a novel STIR method based on low-rank matrix estimation (LRME-STIR), which is applicable to currently available sequential 3D PACT imaging systems without requiring hardware modifications. By employing the LRME-STIR technique, it becomes possible to overcome the challenges caused by the sparsely sampled tomographic data. The proposed approach holds the potential to advance the field by enabling accurate and efficient STIR, thereby facilitating the monitoring of dynamic physiological changes using PACT.

The remainder of the paper is organized as follows. Section 2 presents an imaging model for sequential volumetric imagers and introduces the inverse problem formulation. The proposed LRME-STIR method is described in Sec. 3. The conducted numerical and experimental studies, as well as their results, are provided in Secs. 4 and 5, respectively. Finally, the paper concludes with a discussion in Sec. 6.

2 Imaging Model for Sequential Volumetric Imagers and Inverse Problem Formulation

In the context of PACT, the sequential data acquisition strategy commonly involves utilizing one or more rotating or translating ultrasonic transducer arrays within single or multiple acoustic probes.^{2,19,20} This approach facilitates data collection by rotating the probes along a fixed axis, resulting in the acquisition of a few tomographic measurements at each step, as depicted in Fig. 1. These measurements are accumulated sequentially to form a complete tomographic measurement set.

During each step of sequential data acquisition, the object can be considered to be static (quasi-static assumption), which is justified by the negligible data acquisition time during each step (on the order of 10^{-4} s), significantly shorter than the time between consecutive measurements (on the order of 10^{-1} s). An object frame is defined as the short period of time when the object is considered static. The sequence of object frames constitutes the dynamic object. Essentially, each data acquisition step, hereafter referred to as an imaging frame, corresponds to an object frame. Although object and imaging frames may seem interchangeable, the distinction lies in the fact that the set of imaging frames is essentially a subset of the set of object frames because the dynamic object might not be imaged throughout all object frames. Under the quasi-static assumption, the time-dependent object function at the k -th imaging frame, specifically the dynamic induced initial pressure distribution, is represented as $f_k(\mathbf{r}) = f(\mathbf{r}, k\Delta t)$ for

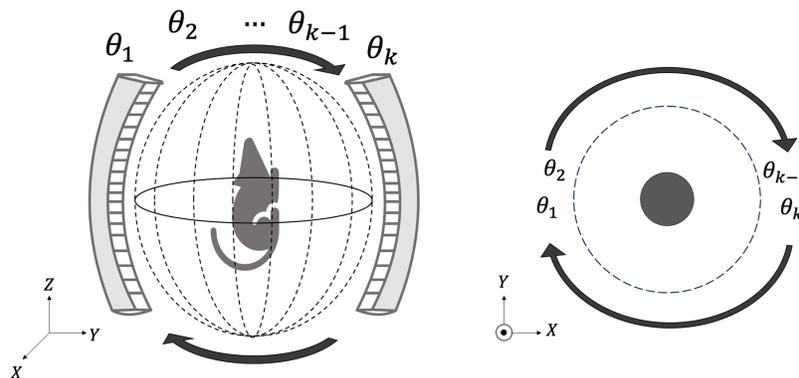


Fig. 1 Schematic illustrating the sequential scanning strategy employing a rotating acoustic probe around the object.

$k = 1, \dots, K$. Here, K represents the number of imaging frames, and Δt is the time interval between consecutive imaging frames, which is equal to the laser repetition rate. Finally, the rotation speed of the gantry determines the angular spacing between imaging frames.

Under the quasi-static assumption, the data acquisition process at each imaging frame is described by a continuous-to-discrete (C-D) imaging model as^{30,34}

$$[\mathbf{g}_k]_{(q-1)P+p} = \frac{1}{\Omega_{qk}} \int_{\Omega_{qk}} d\mathbf{r}'_{qk} \frac{1}{4\pi} \int_V d\mathbf{r} \frac{f_k(\mathbf{r})}{c_0^2} \frac{d}{d\tau} \frac{\delta\left(\tau - \frac{|\mathbf{r}'_{qk} - \mathbf{r}|}{c_0}\right)}{|\mathbf{r}'_{qk} - \mathbf{r}|} \Bigg|_{\tau=p\Delta T}, \quad \begin{matrix} p = 1, 2, \dots, P \\ q = 1, 2, \dots, Q \end{matrix}, \quad (1)$$

where τ denotes the fast-time (i.e., the arrival time of acoustic signals to the transducer) and ΔT corresponds to the fast-time sampling interval. The object function at the k 'th imaging frame, $f_k(\mathbf{r})$, is assumed to be bounded and contained within volume V . The scalar c_0 denotes the speed of sound, which is assumed to be constant throughout volume V . The quantities \mathbf{r} and \mathbf{r}'_{qk} specify the spatial coordinates within V and the location of the q 'th transducer at the k 'th imaging frame, respectively. The vector $\mathbf{g}_k \in \mathbb{R}^{PQ}$ represents lexicographically ordered pressure traces measured by the transducers at the k 'th imaging frame. Here, Q stands for the number of transducers, and P represents the number of electrical signals recorded by each transducer. The notation $[\mathbf{g}_k]_{(q-1)P+p}$ refers to the $(q-1)P+p$ -th entry of the measurement vector \mathbf{g}_k . Here, the integer-valued indices q and p denote the transducer index and temporal sample, respectively, and Ω_{qk} denotes the detection area of the q 'th transducer at the k 'th imaging frame. When the transducer size is small and/or the object is located near the center of a relatively large measurement geometry, an idealized point-like transducer model can be assumed, and the surface integral over Ω_{qk} can be neglected.³⁰

To facilitate the implementation of an iterative image reconstruction algorithm, a discrete-to-discrete (D-D) imaging model is defined as follows. The spatially continuous object functions f_k corresponding to the k 'th imaging frame are approximated using a finite linear combination of spatial expansion functions $\{\psi_n(\mathbf{r})\}_{n=1}^N$ and are given as

$$f_k^{(N)}(\mathbf{r}) = \sum_{n=1}^N \alpha_n^k \psi_n(\mathbf{r}), \quad k = 1, \dots, K, \quad (2)$$

where N denotes the number of spatial expansion functions. In this study, the expansion functions are piecewise trilinear Lagrangian functions defined on a uniform Cartesian grid. Their expression is given as³⁴

$$\psi_n(\mathbf{r}) = \begin{cases} \left(1 - \frac{|x-x_n|}{\Delta s}\right) \left(1 - \frac{|y-y_n|}{\Delta s}\right) \left(1 - \frac{|z-z_n|}{\Delta s}\right), & \text{if } |x-x_n|, |y-y_n|, |z-z_n| \leq \Delta s \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where $\mathbf{r} = (x, y, z)$ denotes the spatial coordinate and $\mathbf{r}_n = (x_n, y_n, z_n)$ specifies the location of the n 'th node of the uniform Cartesian grid. The parameter Δs indicates the distance between adjacent grid points.

The coefficients $\{\alpha_n^k\}_{n=1, k=1}^{NK}$ are organized into a matrix $\mathbf{F} \in \mathbb{R}^{N \times K}$ with entries $[\mathbf{F}]_{nk} \equiv \alpha_n^k$. The k 'th column of \mathbf{F} is denoted by $\mathbf{f}_k \in \mathbb{R}^N$ and represents the discrete approximation of the object function at the k 'th imaging frame, that is,

$$\mathbf{F} = \sum_{k=1}^K \mathbf{f}_k \otimes \mathbf{e}_k = \begin{bmatrix} \alpha_1^1 & \dots & \alpha_1^K \\ \vdots & \ddots & \vdots \\ \alpha_N^1 & \dots & \alpha_N^K \end{bmatrix}, \quad (4)$$

where $\mathbf{e}_k \in \mathbb{R}^K$ represents the k 'th column of the identity matrix in \mathbb{R}^K and \otimes denotes the vector outer product. Correspondingly, a frame-dependent D-D imaging model accounting for measurement noise is expressed as

$$\mathbf{g}_k = \mathbf{H}_k \mathbf{f}_k + \boldsymbol{\eta}_k, \quad k = 1, 2, \dots, K, \quad (5)$$

where \mathbf{g}_k represents the (noisy) tomographic measurements at the k 'th imaging frame and $\boldsymbol{\eta}_k$ accounts for the measurement noise and the modeling and discretization errors. The operator

$\mathbf{H}_k \in \mathbb{R}^{Q^P \times N}$ stems from discretization of the C-D PACT imaging operator associated with the k 'th imaging frame. In particular, the vector $\mathbf{g}_k \in \mathbb{R}^{Q^P}$ representing the action of \mathbf{H}_k on \mathbf{f}_k is defined as in Eq. (1) but with the continuous in space object function $f_k(\mathbf{r})$ replaced by its finite dimensional approximation $f_k^{(N)}(\mathbf{r})$ defined in Eq. (2).

Given the data matrix $\mathbf{G} = \sum_{k=1}^K \mathbf{g}_k \otimes \mathbf{e}_k \in \mathbb{R}^{PQ \times K}$ and the set of imaging operators \mathbf{H}_k ($k = 1, \dots, K$) corresponding to each imaging frame, the goal of dynamic PACT image reconstruction is to find a matrix $\hat{\mathbf{F}} = \sum_{k=1}^K \hat{\mathbf{f}}_k \otimes \mathbf{e}_k \in \mathbb{R}^{N \times K}$, with its column $\hat{\mathbf{f}}_k$ representing an object estimate at the k 'th imaging frame; for simplicity, hereafter $\hat{\mathbf{f}}_k$ is referred to as the k 'th frame of the spatiotemporal object estimate. Given the sparse tomographic measurements acquired for each imaging frame, the task of dynamic image reconstruction constitutes a significantly ill-posed inverse problem. A unique estimator, $\hat{\mathbf{F}}$, is obtained by solving the following penalized least squares optimization problem:

$$\hat{\mathbf{F}} = \arg \min_{\mathbf{F} \in \mathbb{R}^{N \times K}} \mathbf{J}(\mathbf{F}) := \mathbf{L}(\mathbf{F}) + \mathbf{R}(\mathbf{F}) = \sum_{k=1}^K \mathbf{L}_k(\mathbf{F}) + \mathbf{R}(\mathbf{F}), \quad (6)$$

where $\mathbf{R}(\mathbf{F})$ is the regularization term, which is convex but possibly non-smooth. The total data fidelity term $\mathbf{L}(\mathbf{F}) = \sum_{k=1}^K \mathbf{L}_k(\mathbf{F})$ is the sum of data fidelity terms $\mathbf{L}_k(\mathbf{F})$ associated with the k 'th imaging frame. These quantities are defined as

$$\mathbf{L}(\mathbf{F}) = \frac{1}{2} \left\| \left(\sum_{k=1}^K (\mathbf{H}_k \mathbf{f}_k) \otimes \mathbf{e}_k \right) - \mathbf{G} \right\|_F^2 \quad \text{and} \quad \mathbf{L}_k(\mathbf{F}) = \frac{1}{2} \|\mathbf{H}_k \mathbf{f}_k - \mathbf{g}_k\|_2^2, \quad (7)$$

respectively, where $\|\cdot\|_F$ denotes the Frobenius norm.

In addition to the inherent ill-posed nature of the considered dynamic image reconstruction problem, significant implementation challenges exist. First, unlike the relatively straightforward task of static image reconstruction in which a single image is estimated, STIR involves dealing with a considerably higher computational burden as all frames are reconstructed simultaneously, particularly when each frame is 3D in space. Second, as the number of frames increases, there is a growing need for memory space during computation, which necessitates effective computational strategies with minimal memory usage.

3 Low-rank Matrix Estimation-based Spatiotemporal Image Reconstruction

In numerous biomedical applications, the spatiotemporal object function of interest has been demonstrated to be effectively approximated by a small set of weights σ_r and functions $u_r(\mathbf{r})$ and $v_r(t)$, depending only on the space or time variables, known as the semiseparable approximation.^{29,35-40} Consequently, the object function is expressed as $f(\mathbf{r}, t) \approx \sum_{r=1}^R \sigma_r u_r(\mathbf{r}) v_r(t)$.^{29,35-40} Leveraging the semiseparable approximation, the spatiotemporal reconstruction problem can be reduced to the problem of estimating R weights and R spatial and temporal functions. In a discretized formulation, this is algebraically equivalent to enforcing a low-rank structure on the matrix \mathbf{F} .

A penalty scheme can then be imposed on the nuclear norm of the spatiotemporal reconstruction matrix to promote low-rankness. Specifically, the regularization term stemming from the nuclear norm of \mathbf{F} is defined as

$$\mathbf{R}^{nn}(\mathbf{F}) = \|\mathbf{F}\|_* = \sum_{r=1}^{\min(N,K)} \sigma_r, \quad (8)$$

where σ_r ($r = 1, \dots, \min(N, K)$) represent the singular values of \mathbf{F} . This approach not only effectively regularizes the ill-posed inverse problem⁴¹ but also reduces memory demands without sacrificing accuracy. Rather than explicitly storing \mathbf{F} in memory, its truncated singular value decomposition (SVD) $\mathbf{U}_R \mathbf{\Sigma}_R \mathbf{V}_R^T$ is stored, resulting in a decreased memory requirement. Here, $R < \min(N, K)$ denotes the truncation index, $\mathbf{\Sigma}_R$ is the diagonal matrix comprising the largest R singular values, and \mathbf{U}_R and \mathbf{V}_R are matrices with orthonormal columns collecting the left and right singular vectors, respectively, corresponding to the R largest singular values.

This approach not only facilitates an efficient approximation of \mathbf{F} but also enforces the space-time semiseparability. Specifically, the columns of \mathbf{U}_R and \mathbf{V}_R are the algebraic counterpart of functions $u_r(\mathbf{r})$ and $v_r(t)$, respectively.

To further regularize the inverse problem, under the assumption that the object undergoes a smooth and slow temporal change, another regularization scheme penalizing the difference between two consecutive frames can be employed. This is accomplished by penalizing the squared Frobenius norm of the temporal difference matrix, which is expressed through the temporal (forward) difference operator, denoted as $\mathbf{D} \in \mathbb{R}^{K \times (K-1)}$, being applied to \mathbf{F} , given as

$$\mathbf{R}^t(\mathbf{F}) = \frac{1}{2} \|\mathbf{F}\mathbf{D}\|_F^2, \quad (9)$$

where \mathbf{D} is defined as

$$\mathbf{D} = [\mathbf{d}_1 \quad \cdots \quad \mathbf{d}_k \quad \cdots \quad \mathbf{d}_{K-1}] = \begin{bmatrix} -1 & 0 & 0 & 0 & \cdots \\ 1 & -1 & 0 & 0 & \cdots \\ 0 & 1 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & -1 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}, \quad (10)$$

where \mathbf{d}_k corresponds to the k 'th column of the temporal (forward) difference operator, \mathbf{D} . Within this column, the k 'th and $(k+1)$ 'th elements possess values of -1 and 1 , respectively, and all other elements are set to 0 .

In this way, the sought after estimate of the dynamic reconstruction problem with consideration of a maximum rank constraint and temporal and nuclear norm penalties is formulated as

$$\hat{\mathbf{F}} = \arg \min_{\mathbf{F} \in \mathbb{R}_{R_{\max}}^{N \times K}} \mathbf{J}(\mathbf{F}) := \mathbf{L}(\mathbf{F}) + \gamma \mathbf{R}^t(\mathbf{F}) + \lambda \mathbf{R}^{nn}(\mathbf{F}), \quad (11)$$

where $\mathbb{R}_{R_{\max}}^{N \times K}$ denotes the set of all N -by- K matrices of rank R_{\max} at most and the cost function $\mathbf{J}(\mathbf{F})$ is comprised of the data fidelity term $\mathbf{L}(\mathbf{F})$ in Eq. (7), the temporal regularization term $\mathbf{R}^t(\mathbf{F})$ in Eq. (9), and the nuclear norm regularization term $\mathbf{R}^{nn}(\mathbf{F})$ in Eq. (8). Here, the parameters $\gamma \geq 0$ and $\lambda \geq 0$ control the strength of the temporal and nuclear norm regularization terms, respectively. Similar to the data fidelity computation in Eq. (7), the temporal regularization term $\mathbf{R}^t(\mathbf{F}) = \sum_{k=1}^K \mathbf{R}_k^t(\mathbf{F})$ is written as the sum of contributions \mathbf{R}_k^t from each imaging frame. Each term is defined as

$$\mathbf{R}_k^t(\mathbf{F}) := \begin{cases} \|\mathbf{F}\mathbf{d}_k\|_2^2 = \|\mathbf{f}_{k+1} - \mathbf{f}_k\|_2^2 & \text{if } k < K \\ \|\mathbf{F}\mathbf{d}_K\|_2^2 = 0 & \text{if } k = K, \end{cases} \quad (12)$$

Where, for uniformity of notation, $\mathbf{d}_K \in \mathbb{R}^K$ is the zero vector. To highlight the contribution from each frame, the minimization problem in Eq. (11) is then rewritten as

$$\begin{aligned} \hat{\mathbf{F}} &= \arg \min_{\mathbf{F} \in \mathbb{R}_{R_{\max}}^{N \times K}} \sum_{k=1}^K \mathbf{L}_k(\mathbf{F}) + \gamma \sum_{k=1}^K \mathbf{R}_k^t(\mathbf{F}) + \lambda \mathbf{R}^{nn}(\mathbf{F}) \\ &= \arg \min_{\mathbf{F} \in \mathbb{R}_{R_{\max}}^{N \times K}} \sum_{k=1}^K \|\mathbf{H}_k \mathbf{f}_k - \mathbf{g}_k\|_2^2 + \frac{\gamma}{2} \sum_{k=1}^K \|\mathbf{F}\mathbf{d}_k\|_2^2 + \lambda \|\mathbf{F}\|_*. \end{aligned} \quad (13)$$

In the formulated minimization problem, the data fidelity and temporal penalty terms are convex and smooth, whereas the nuclear norm penalty term is convex but non-smooth. Accordingly, the minimization problem can be solved using a proximal gradient descent (PGD) method.^{42,43}

The convergence speed of the PGD method can be improved with momentum schemes.^{44,45} To further improve the convergence speed, especially in the early iterations, an ordered subsets (OS) approach⁴⁶ can be incorporated with momentum schemes,^{47,48} despite the lack of theoretical guarantees of convergence. In addition to acceleration, applying the OS approach to find an approximate solution to Eq. (13) offers several other significant benefits. Notably, it substantially

reduces memory requirements by a factor proportional to the number M of subsets used. The gradient with respect to all imaging frames requires $\mathcal{O}(NK)$ memory usage, whereas the gradient corresponding to each subset only requires $\mathcal{O}(NK/M)$ storage. Furthermore, it preserves the low-rank structure of the reconstructed object function estimate at each step of the proximal gradient iteration.

For the optimization problem given by Eq. (11), the OS-based cost function is expressed as^{47,48}

$$\mathbf{J}_{\mathcal{K}_j}(\mathbf{F}) = M \sum_{k \in \mathcal{K}_j} \mathbf{L}_k(\mathbf{F}) + M\gamma \sum_{k \in \mathcal{K}_n} \mathbf{R}_k^t(\mathbf{F}) + \lambda \mathbf{R}^{nn}(\mathbf{F}), \quad (14)$$

where \mathcal{K}_j represents the set of frame indices in the j 'th ordered subset. The OS approach relies on the ‘‘subset balance’’ approximation,^{47,48} implying that $\mathbf{J}_{\mathcal{K}_j}(\mathbf{F}) \approx \mathbf{J}(\mathbf{F})$. The update procedure in PGD with OS consists of two main steps. Initially, a gradient descent step is performed; it moves in the negative gradient of the smooth components of the OS-based cost function, which yields

$$\begin{aligned} \mathbf{F}_{j+\frac{1}{2}} &= \mathbf{F}_j - \eta \nabla \left(M \sum_{k \in \mathcal{K}_j} \mathbf{L}_k(\mathbf{F}_j) + M\gamma \sum_{k \in \mathcal{K}_j} \mathbf{R}_k^t(\mathbf{F}_j) \right) \\ &= \mathbf{F}_j - \eta M \sum_{k \in \mathcal{K}_j} \left((\mathbf{H}_k^T (\mathbf{H}_k \mathbf{f}_k - \mathbf{g}_k)) \otimes \mathbf{e}_k + \gamma \mathbf{F}_j (\mathbf{d}_k \otimes \mathbf{d}_k) \right). \end{aligned} \quad (15)$$

Here, η is the step size, and \mathbf{F}_j denotes the spatiotemporal object estimate at the beginning of the j 'th update. The first component of the gradient, associated with the data fidelity term for the k 'th frame, results in an outer product that produces a rank-1 matrix. Similarly, the second component of the gradient, related to the temporal regularization term for the k 'th frame, also yields a rank-1 matrix. Thus, the maximum rank of the gradient of the OS-based cost function $\mathbf{J}_{\mathcal{K}_j}$ is bounded by $2|\mathcal{K}_j|$, where $|\mathcal{K}_j|$ denotes the number of imaging frames in the ordered subset.

Following the gradient descent step, the proximal step is executed by applying the proximal operator to account for the nonsmooth component of the objective function. The proximal step corresponds to the solution of the following minimization problem:

$$\mathbf{F}_{j+1} = \text{prox}_{\eta\lambda\|\cdot\|_*} \left(\mathbf{F}_{j+\frac{1}{2}} \right) := \underset{\mathbf{F} \in \mathbb{R}^{N \times K}}{\text{argmin}} \frac{1}{2} \left\| \mathbf{F} - \mathbf{F}_{j+\frac{1}{2}} \right\|_2^2 + \eta\lambda \|\mathbf{F}\|_*, \quad (16)$$

the solution of which can be efficiently implemented via a truncated SVD factorization of $\mathbf{F}_{j+\frac{1}{2}}$ and the application of the soft-thresholding operator, $\mathcal{S}(\cdot)$, to its singular values $\{\sigma_i\}$.⁴⁹ The solution of the Eq. (16) is expressed as follows:

$$\mathbf{F}_{j+1} = \mathbf{U}_{j+\frac{1}{2}} \mathcal{S}_{\eta\lambda}(\boldsymbol{\Sigma}_{j+1/2}) \mathbf{V}_{j+\frac{1}{2}}^T = \mathbf{U}_{j+\frac{1}{2}} \tilde{\boldsymbol{\Sigma}}_{j+1/2} \mathbf{V}_{j+\frac{1}{2}}^T, \quad (17)$$

where $\mathbf{U}_{j+\frac{1}{2}}$, $\boldsymbol{\Sigma}_{j+\frac{1}{2}}$, and $\mathbf{V}_{j+\frac{1}{2}}$ stem from the truncated SVD of $\mathbf{F}_{j+\frac{1}{2}}$ with maximum rank R_{\max} and $\tilde{\boldsymbol{\Sigma}}_{j+\frac{1}{2}} := \mathcal{S}_{\eta\lambda}(\boldsymbol{\Sigma}_{j+1/2})$. The soft-thresholding operator, $\mathcal{S}(\cdot)$, is defined (component-wise) as

$$\mathcal{S}_{\eta\lambda}(\sigma_i) = \tilde{\sigma}_i = \begin{cases} \sigma_i - \eta\lambda & \text{if } \sigma_i > \eta\lambda \\ 0, & \text{if } \sigma_i \leq \eta\lambda. \end{cases} \quad (18)$$

The soft-thresholding is computationally efficient and enforces a low-rank structure that effectively attenuates the singular values.

Algorithm 1 summarizes the proposed accelerated PGD algorithm, integrating both momentum and ordered subset techniques to efficiently find an approximate solution to the minimization problem in Eq. (11). The algorithm takes the following parameters as input: the maximum allowed rank R_{\max} ; the threshold ε for the stopping criterion; the regularization parameter γ for temporal regularization; the regularization parameter λ for nuclear-norm regularization; the step size η ; and the number M of subsets. At the beginning of each iteration, the sequence of frame indices undergoes random shuffling. Within the subsequent inner loop, these shuffled indices are partitioned into M subsets. The gradient pertaining to both the data fidelity and temporal regularization components is computed for these subset frame indices, and the gradient descent step is executed. Then, the proximal mapping associated with nuclear norm regularization is evaluated efficiently using randomized SVD⁵⁰ and soft thresholding. Subsequent to this step, the fast

Algorithm 1 LRME-STIR.**Input:** $R_{\max}, \epsilon, \gamma, \lambda, \eta, M$ **Output:** F **Initialization:** $F_1^{(0)} = \mathbf{0}, t_1^{(0)} = 1, b = \lceil K/M \rceil, i = 0$ **while** $\|F^{(i)} - F^{(i-1)}\|_F^2 / \max_{l \leq i} \|F^{(l)} - F^{(l-1)}\|_F^2 \leq \epsilon$ **do**

▷ Convergence check

 $\mathcal{K} = \text{shuffle}(1, 2, \dots, K)$

▷ Randomly shuffle frames

for $j = 1$ **to** M **do** $\mathcal{K}_j = \mathcal{K}[(j-1)b + 1 : \min(jb, K)]$

▷ Select the current subset of frames

Compute gradient descent for the current subset of frames:

$$F_{j+\frac{1}{2}}^{(i)} = \bar{F}_j^{(i)} - \eta \nabla \left(M \sum_{k \in \mathcal{K}_j} \mathbf{L}_k(\bar{F}_j^{(i)}) + M \gamma \sum_{k \in \mathcal{K}_j} \mathbf{R}_k^t(\bar{F}_j^{(i)}) \right)$$

▷ $\text{rank}(F_{j+\frac{1}{2}}^{(i)}) \leq 2R_{\max} + 2b$ **Perform proximal operator:**⁴⁹

$$U_{j+\frac{1}{2}}^{(i)}, \Sigma_{j+\frac{1}{2}}^{(i)}, V_{j+\frac{1}{2}}^{(i)} = \text{SVD}(F_{j+\frac{1}{2}}^{(i)}, R_{\max})$$

▷ Compute the truncated SVD with maximum rank R_{\max}

$$\tilde{\Sigma}_{j+\frac{1}{2}}^{(i)} = \mathcal{S}_{\eta}(\Sigma_{j+\frac{1}{2}}^{(i)})$$

▷ Soft thresholding for nuclear norm regularization

$$F_{j+1}^{(i)} = U_{j+\frac{1}{2}}^{(i)} \tilde{\Sigma}_{j+\frac{1}{2}}^{(i)} (V_{j+\frac{1}{2}}^{(i)})^T$$

▷ $\text{rank}(F_{j+1}^{(i)}) \leq R_{\max}$ **Implement FISTA acceleration:**⁴⁴

$$t_{j+1}^{(i)} = \frac{1 + \sqrt{1 + 4(t_j^{(i)})^2}}{2}$$

▷ Update momentum term

$$\bar{F}_{j+1}^{(i)} = F_{j+1}^{(i)} + \frac{t_j^{(i)} - 1}{t_j^{(i)}} (F_{j+1}^{(i)} - F_j^{(i)})$$

▷ FISTA update for the current iteration;
 $\text{rank}(\bar{F}_{j+1}^{(i)}) \leq 2R_{\max}$ **end for**

$$F^{(i+1)} = F_{N+1}^{(i)}$$

$$\bar{F}_1^{(i+1)} = \bar{F}_{N+1}^{(i)}$$

$$t_1^{(i+1)} = t_{N+1}^{(i)}$$

$$i = i + 1$$

▷ Increment outer loop iteration

end while

iterative shrinkage-thresholding algorithm⁴⁴ (FISTA) acceleration scheme is deployed to enhance the convergence rate. The algorithm terminates when the squared Frobenius norm of the difference between two successive iterations $\|F^{(i)} - F^{(i-1)}\|_F^2$, normalized by its maximum $\max_{l \leq i} \|F^{(l)} - F^{(l-1)}\|_F^2$ over the previous iterations, falls below the threshold ϵ defined by the user. This metric is equivalent to monitoring the norm of the gradient in smooth optimization.⁴³

4 Study Description

4.1 3D PACT Imaging System Specifications for Experimental and Numerical Studies

The TriTom preclinical imaging system^{19,51} developed by PhotoSound was employed for the experimental studies and emulated for the numerical in-silico studies. It integrates photoacoustic (PA) and fluorescence (FL) imaging modalities, harnessing their individual strengths. The system comprises a central rotary scanning stage, an optical excitation setup for PA and FL imaging, a curvilinear 96-element PA transducer array, and a fluorescence-enabled scientific complementary metal-oxide-semiconductor (sCMOS) camera. The configuration of the TriTom setup is depicted in Fig. 2(a), and the imaging chamber is shown in Fig. 2(b).

During imaging, the object is immersed in water and continuously rotated while being optically stimulated by a short laser pulse with a 10 Hz repetition rate. The maximum rotation speed is 10 degrees per second; therefore, completing a full 360-deg scan requires 36 s. Four optical fiber bundles are located on the outer circumference of the cylindrical imaging chamber, perpendicular to the scanning plane of the PA array. The laser emits pulses in the 670 to 1064 nm wavelength range at a frequency of 10 Hz, with each pulse lasting 5 ns. Acoustic waves generated by the laser excitation are detected by the PA transducer array, which comprises piezoelectric transducer elements, each measuring $1.3 \times 1.3 \text{ mm}^2$. The center frequency of the transducer elements is $6 \text{ MHz} \pm 10\%$ (at -6 dB) with bandwidth $\geq 55\%$. The array is vertically oriented and cylindrically focused, with the central element positioned 65 mm from the center of the imaging chamber. For fluorescence imaging, an sCMOS camera with a 2048×2040 pixel resolution and a $40 \times 40 \text{ mm}^2$ field-of-view is placed outside the imaging chamber.

In the acoustic modeling of the TriTom imaging system, the transducer array was assumed to consist of idealized point-like transducers positioned at the central locations of the transducer elements. The acoustic simulation was implemented with a GPU-accelerated D-D imaging model³⁴ assuming an acoustically homogeneous medium. Although the TriTom system has a single transducer arc, other existing 3D PACT designs (e.g., Refs. 20 and 52) feature multiple transducer arcs; thus the numerical studies also explored scenarios in which multiple tomographic measurements were acquired per imaging frame. Specifically, measurement configurations in which two transducer arcs separated by 90 deg and four transducer arcs separated by 45 deg were considered, as illustrated in Fig. 3. The associated compression ratios for sparse sampling in these scenarios are 1/360, 1/180, and 1/90 for the measurement configuration with 1, 2, and 4 transducer arcs, respectively. The sampling rate for both the experimental and numerical studies was set to 31.25 MHz, with 2048 temporal samples collected per imaging frame.

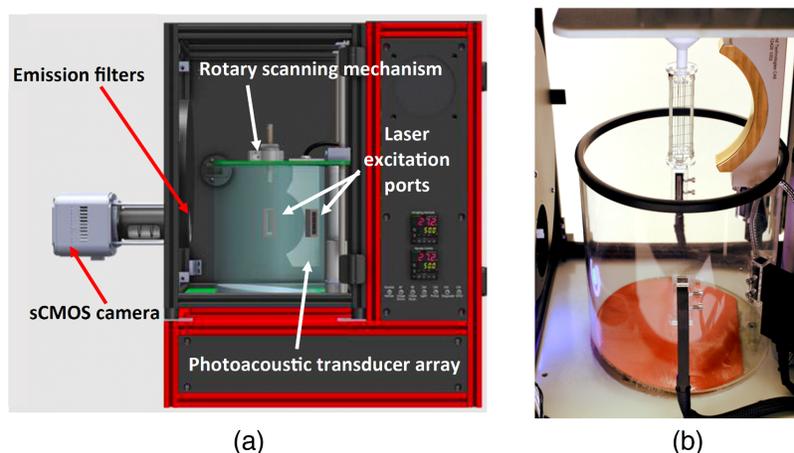


Fig. 2 Illustration of the TriTom imaging system (a) and a picture of the imaging chamber (b).

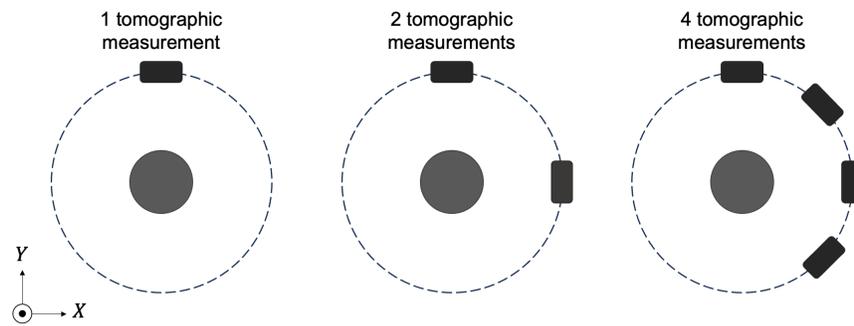


Fig. 3 Top view of the virtual imaging systems collecting 1, 2, or 4 tomographic measurements per imaging frame.

4.2 Inverse Crime Validation Study

To verify that Algorithm 1 was correctly implemented, an inverse crime validation study was conducted in silico in which a simple rank-4 dynamic phantom was employed. The phantom consisted of $40 \times 40 \times 3$ spatial voxels and 360 object frames, with a voxel size of $0.4 \times 0.4 \times 0.4 \text{ mm}^3$. The induced pressure in the phantom was assumed constant along the z -axis and piecewise constant within the xy -plane at each object frame. The phantom was structured into four distinct regions, each characterized by varying temporal activities as shown in Fig. 4. Figure 4(a) illustrates the central z -slice of the phantom at the 120th object frame, and Fig. 4(b) displays the time activity curves (TACs) corresponding to the numbered regions. For each imaging frame (360 in total), four tomographic measurements separated by 45 deg were considered, as depicted in Fig. 3(b). Simulated acoustic pressure data were generated using a grid voxel size of $0.4 \times 0.4 \times 0.4 \text{ mm}^3$, assuming the phantom was centered within the imaging system.³⁴ The speed of sound was assumed constant at 1495 m/s, and no noise was added to the simulated measurement data.

During image reconstruction, the same computational grid that was used for generating the measurement data was employed. In the algorithm, the maximum allowed rank, R_{\max} , was set to 4, and no temporal or nuclear norm penalties were applied ($\lambda = \gamma = 0$). The step-size, η , was tuned empirically to ensure convergence. To explore the impact of the number of OS used in the randomized evaluation of the data fidelity term, three different numbers of OS were investigated: $M \in \{1, 2, 6\}$.

4.3 Numerical Phantom Study

A dynamic numerical phantom was utilized to evaluate the performance of the proposed method through in silico experiments. The phantom consisted of 360 object frames and contained four convex ellipsoidal blobs within a larger ellipsoidal blob, along with a vasculature mimicking structure, as shown in Fig. 5(d). The size of the numerical phantom was $40 \times 40 \times 30 \text{ mm}^3$, with a voxel size of $0.4 \times 0.4 \times 0.4 \text{ mm}^3$. The time activity at each voxel was designed to mimic

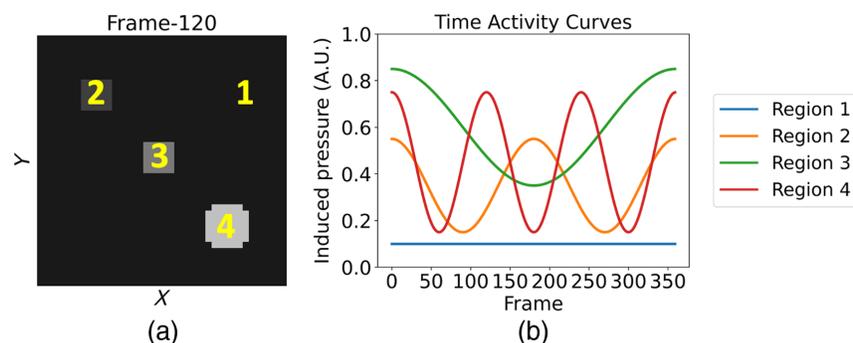


Fig. 4 Central z -slice of the simple rank-4 phantom at the 120th object frame (a), TACs at the numbered regions in the z -slice (b). Region 1 denotes a static background.

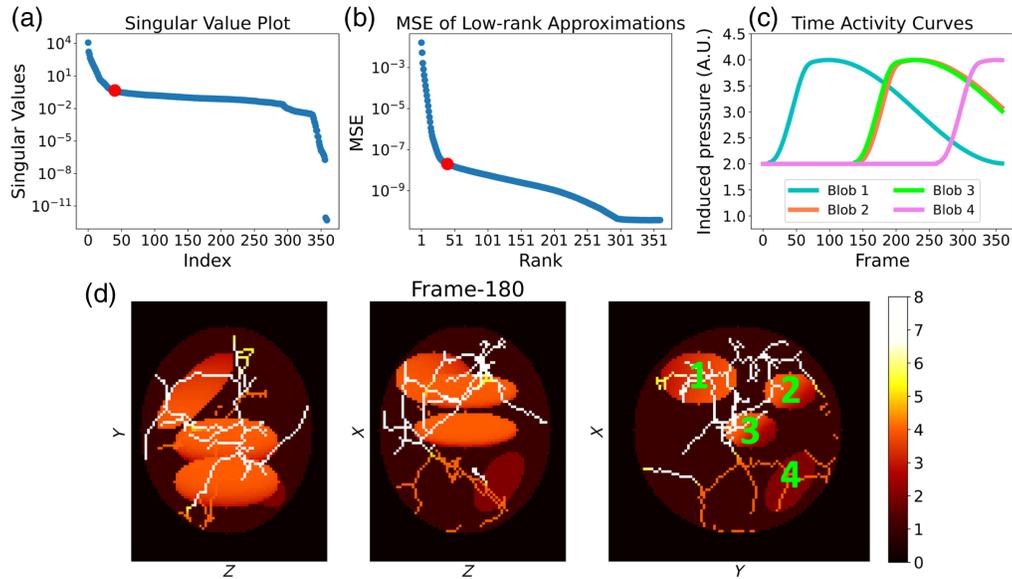


Fig. 5 (a) Singular value plot of the dynamic numerical phantom, (b) MSE of low-rank approximations, (c) TACs at the center of each ellipsoidal blob, and (d) maximum intensity projection (MIP) images of the 180th object frame of the dynamic phantom (bottom). The ellipsoidal blobs are numbered in the MIP along the z-axis image. The red dot in panels (a) and (b) denotes the index $R_{\max} = 40$ used as a maximum rank constraint in the reconstruction studies.

a contrast agent's flow along the paths from ellipsoidal blobs 1 to 2 and 3 to 4. Figure 5(c) illustrates the time activity at the center of each ellipsoidal blob, and the blobs are numbered in Fig. 5(d). The singular values of the dynamic numerical phantom are shown in Fig. 5(a), where a rapid singular value decay is observed. The rapid singular decay indicates that the phantom can be accurately approximated with a low-rank representation, as demonstrated by the mean squared error (MSE) versus rank plot depicted in Fig. 5(b).

To generate the synthetic measurement data, the grid voxel size was set to $0.2 \times 0.2 \times 0.2 \text{ mm}^3$, and 360 imaging frames (corresponding to a complete rotation of the system) were simulated. The speed of sound was assumed to be constant at 1495 m/s. Zero-mean Gaussian noise with a standard deviation equivalent to a specified percentage of the maximum value of the simulated data (1%, 3%, or 5%) was added to the pressure signals.

The primary objective of this numerical study was to investigate the effect of the following physical factors on the proposed method.

- Number of tomographic measurements per imaging frame: in this study, the noise level was fixed at 1%, and three scenarios were examined with 1, 2, and 4 tomographic measurements per imaging frame, as illustrated in Fig. 3.
- Measurement noise level: in this study, the number of tomographic measurements per imaging frame was set to 2, and three different noise levels were considered: 1%, 3%, and 5%.

4.4 Numerical Study

For image reconstruction, the maximum allowed rank in the algorithm was set to $R_{\max} = 40$ as it allows for an accurate approximation of the dynamic numerical phantom with an MSE of around 10^{-8} , as shown in Fig. 5(d). To avoid the discretization inverse crime, a coarser grid with voxel size $0.4 \times 0.4 \times 0.4 \text{ mm}^3$ was used for the reconstruction. The speed of sound value was the same as that used for simulating the data. The threshold ε for the stopping criterion of Algorithm 1 was set to 2.5×10^{-1} . The step-size, η , was tuned empirically to ensure convergence. The number of subsets, M , was set to 18. To select appropriate regularization parameters, the balancing principle⁵³ was employed to reduce the number of tunable regularization parameters from two to one. The balancing principle rescales each regularization term by an estimate of their value at the object function F^{true} . An iterative procedure was proposed in Ref. 53 to estimate such values; however, for simplicity, this work assumes direct knowledge of the actual nuclear norm and

the Frobenius norm of the temporal difference of F^{true} . Importantly, this assumption is specific to the numerical study, which examines the method's performance under controlled conditions. In practical applications in which direct knowledge of the true values is unavailable, as in the experimental study (see Sec. 4.5), empirical parameter sweeps can be used to tune regularization parameters for dynamic image reconstruction. Specifically, in this study, the solution to the dynamic image reconstruction problem is defined as

$$\hat{F} = \arg \min_F \frac{1}{2} \sum_{k=1}^K \|H_k f_k - g_k\|^2 + \kappa \left(\frac{\sum_{k=1}^{K-1} \|f_{k+1} - f_k\|^2}{\sum_{k=1}^{K-1} \|f_{k+1}^{true} - f_k^{true}\|^2} + \frac{\|F\|_*}{\|F^{true}\|_*} \right). \quad (19)$$

For each case, four different values of the parameter κ were explored: $\kappa \in \{10^{-4} \|G\|_F^2, 5 \times 10^{-4} \|G\|_F^2, 2.5 \times 10^{-3} \|G\|_F^2, 1.25 \times 10^{-2} \|G\|_F^2\}$. The κ parameter yielding the best average normalized squared error (nSE) over frames was selected.⁵⁴ The nSE for each frame was computed as

$$\text{nSE} = \|f_k^{true} - \hat{f}_k\|_2^2 / \max_k \|f_k^{true}\|_2^2. \quad (20)$$

Specifically, the regularization parameter value that yielded the optimal results was $\kappa = 5 \times 10^{-4} \|G\|_F^2$ when four tomographic measurements per frame were available and $\kappa = 2.5 \times 10^{-3} \|G\|_F^2$ for all other cases.

4.5 Experimental Study

An open-ended dynamic flow phantom was constructed by bending a silicone tube into a U-shaped structure, as depicted in Fig. 6. Figure 6(a) provides an illustration of the phantom within the imaging chamber, and Fig. 6(b) shows an actual photograph of the physical phantom. The inner diameter of the silicone tube was 0.0635 cm, and the outer diameter was 0.1194 cm. The dynamic phantom, positioned at the center of the imaging system, was illuminated by a laser pulse with a wavelength of 770 nm and an energy of 100 mJ (before entering the fiber optic light delivery unit). PACT data were acquired with the TriTom imaging system during the injection of a photoacoustic-fluorescent contrast agent (PAiR⁵⁵) through one end of the tube. Over a span of 36 s, the dynamic phantom underwent scanning, which resulted in 360 imaging frames in total, with 10 imaging frames per second and a single tomographic measurement per imaging frame. Two-dimensional fluorescence images were concurrently gathered, which serve as reference for the time evolution of the contrast agent concentration within the tube.

Based on the region illuminated by laser, the reconstruction volume was set to a region of $30 \times 30 \times 30 \text{ mm}^3$ located at the center of the imaging system. The voxel size was set to $0.4 \times 0.4 \times 0.4 \text{ mm}^3$, resulting in $75 \times 75 \times 75$ spatial voxels. The maximum allowed rank, R_{\max} , was set to 40, and the number of subsets, M , was set to 18. For the stopping criterion of Algorithm 1, the threshold ϵ was set to 5×10^{-1} . The step-size, η , was tuned empirically to

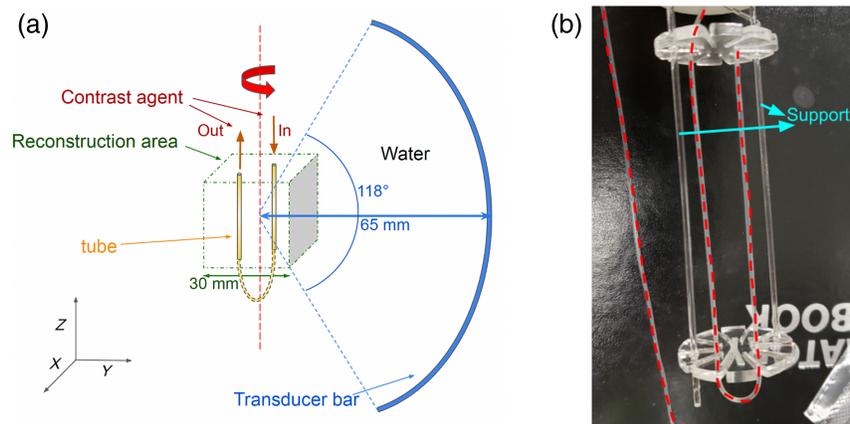


Fig. 6 Illustration (a) and a picture (b) of the experimental dynamic phantom.

ensure convergence. To calibrate the speed of sound, multiple static estimates of the dynamic object were reconstructed using the universal back-projection algorithm⁵⁶ assuming speed of sound values in the range of 1480 to 1520 m/s, with 5 m/s increments. The speed of sound value of 1495 m/s resulted in the best visual appearance and was selected to perform the dynamic image reconstruction. To choose the regularization parameters, nine dynamic image reconstructions were performed for all possible combinations of $\gamma \in \{5.5 \times 10^0, 5.5 \times 10^1, 5.5 \times 10^2\}$ and $\lambda \in \{10^{-1}, 10^0, 10^1\}$. Ultimately, the spatiotemporal object estimate that most closely captured the observed dynamic changes in the reference fluorescence images, as determined through visual examination, was selected. The corresponding regularization parameters were $\gamma = 5.5 \times 10^1$ and $\lambda = 10^0$.

5 Results

5.1 Inverse Crime Validation Study Results

In this inverse crime study, the step sizes η for $M = 1, 2, 6$ were empirically tuned to ensure convergence, resulting in values of $\{10^{-4}, 10^{-4}, 3.3 \times 10^{-5}\}$, respectively. The algorithm was allowed to run for 2500 iterations in each case to ensure convergence of the solution up to numerical precision. For the purpose of this validation study, Algorithm 1 was modified to re-evaluate the total data fidelity term $\mathbf{L}(\mathbf{F}_i)$, which accumulates the contributions from all time frames, at each iteration.

The results of the validation study are depicted in Fig. 7. Figure 7(a) exhibits the data fidelity $\mathbf{L}(\mathbf{F}_i)$ versus the iteration count, and Fig. 7(b) presents the average nSE versus the iteration count. For all values of M , a significant decrease of approximately 11 and 13 orders of magnitude can be observed in the data fidelity term and average nSE, respectively. This indicates that the proposed method using momentum-acceleration in combination with OS (case $M \in \{2, 6\}$), although lacking a theoretical guarantee of convergence, can achieve (up to machine precision) to the same object estimate produced by the momentum-accelerated PGD without the subsampling method ($M = 1$). It is also evident that a larger number of subsets improves the convergence speed, particularly in the early iterations. This also translates into a possibly faster time to solution as the cost of each iteration is dominated by the evaluation of the imaging operator, whereas the time spent in performing the truncated SVD factorization using the randomized method is negligible. In the numerical studies presented here, the computational time required per iteration was approximately 15 min on a workstation (AMD EPYC 7702P 64-Core processor, 32 GB RAM, one Nvidia Geforce RTX 2080 graphic processing unit), independent of the number M of subsets used.

In summary, this validation study demonstrates the correct implementation, efficiency, and robust convergence, despite a lack of theoretical guarantees, of the proposed method.

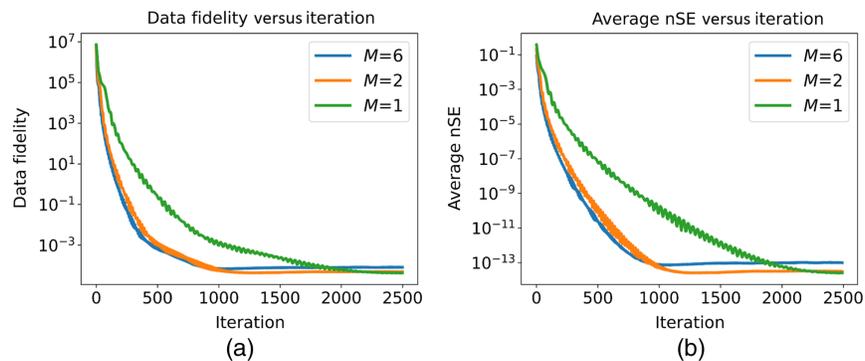


Fig. 7 Data fidelity versus the iterations (a), and the average nSE versus iterations (b). The plots confirm the correct implementation of the algorithm and illustrate how increasing the number of subsets can accelerate the reduction of the data fidelity term. Additionally, the plots illustrate the similar order-of-magnitude reduction in both data fidelity and average nSE values for every number of subsets $M \in \{1, 2, 6\}$.

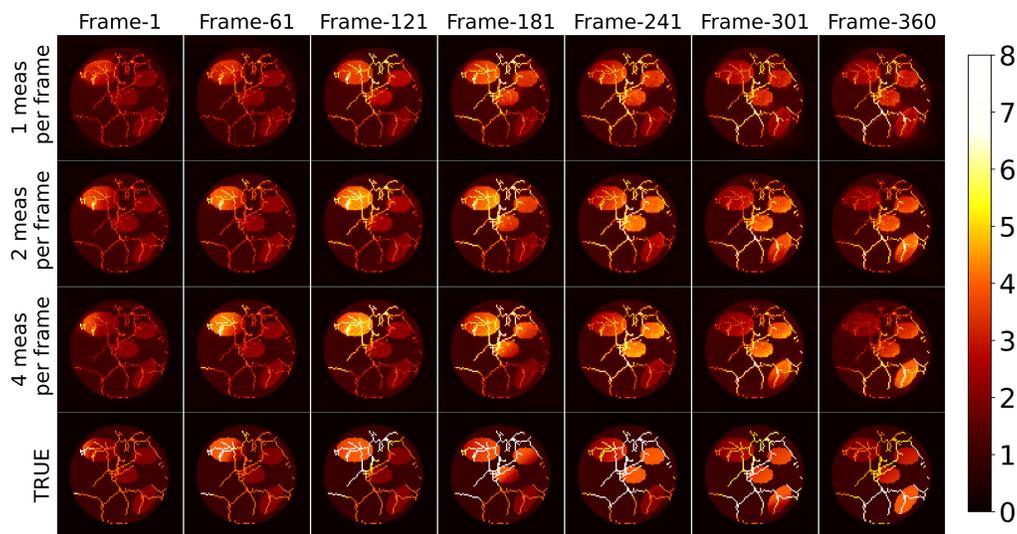


Fig. 8 Selected MIP images (along the z-axis) depicting the numerical phantom and corresponding spatiotemporal estimates from simulated data with varying numbers of tomographic measurements per imaging frame. A close examination reveals improved reconstruction accuracy with an increased number of tomographic measurements. Notably, the reconstructions using two and four tomographic measurements per imaging frame more accurately capture the intensity change in blob-4 (located at the bottom right) compared with the one using a single tomographic measurement per imaging frame. This observation is further supported by the TACs shown in Fig. 9. A video featuring the spatiotemporal evolution of the numerical phantom and its reconstructed estimates is available as [Video 1](https://doi.org/10.1117/1.JBO.29.S1.S11516.s1) (Video 1, MP4, 10 MB [URL: <https://doi.org/10.1117/1.JBO.29.S1.S11516.s1>]).

5.2 Numerical Phantom Study Results

5.2.1 Sensitivity to the number of tomographic measurements per imaging frame

Figure 8 displays a selection of MIP images depicting the numerical phantom and the spatiotemporal estimates from simulated data with varying numbers of tomographic measurements per imaging frame. A video featuring the spatiotemporal evolution of the numerical phantom and its corresponding estimates is available as [Video 1](https://doi.org/10.1117/1.JBO.29.S1.S11516.s1). Upon careful inspection, it is apparent that increasing the number of tomographic measurements per imaging frame leads to more accurate estimates of the dynamic object. For instance, object estimates reconstructed from data with two and four measurements per imaging frame correctly capture the increase in intensity of blob-4 after frame-301. However, this intensity change is less prominent in the dynamic estimate reconstructed from data with only one tomographic measurement per imaging frame, which also exhibits artifacts around blob-4. In addition, the estimate reconstructed using four measurements per imaging frame better captures the intensity difference between the two ends of blob-1 in frame-1. These observations are more evident in Fig. 9, which shows the TACs at each ellipsoidal blob's center in both the numerical phantom and the spatiotemporal estimates from data with different numbers of tomographic measurements per imaging frame. Notably, the TACs for the reconstruction from four tomographic measurements per imaging frame closely align with those of the numerical phantom. The nSE versus the imaging frame plot for the reconstructions with varying numbers of tomographic measurements per imaging frame is depicted in Fig. 11(a). Both the TACs and nSE versus frame plots confirm the observation that a higher number of tomographic measurements per imaging frame leads to more accurate spatiotemporal reconstructions.

5.2.2 Sensitivity to measurement noise

Figure 10 displays the TACs at each ellipsoidal blob's center in both the numerical phantom and its spatiotemporal estimates from data with varying noise levels, when the number of

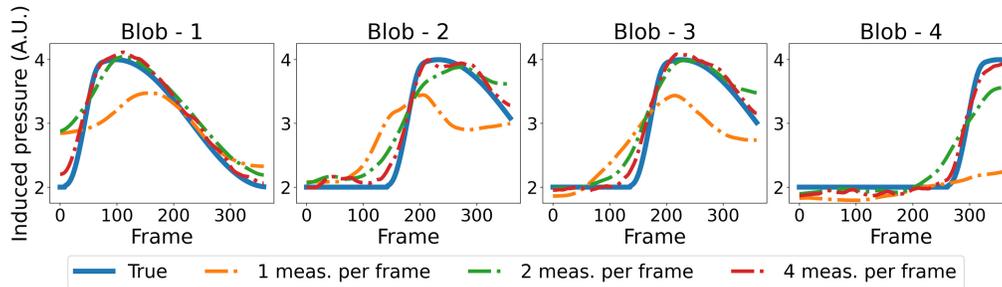


Fig. 9 TACs at each ellipsoidal blob's center, comparing the numerical phantom with its estimate reconstructed from different numbers of tomographic measurements per imaging frame. The noise level was kept at 1%. The curves demonstrate that increasing the number of measurements enhances the fidelity of temporal activities in the reconstructions.

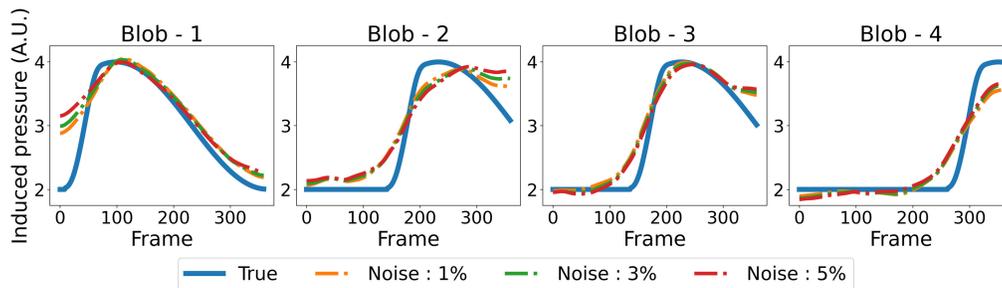


Fig. 10 TACs at each ellipsoidal blob's center, comparing the true phantom with reconstructions from data with different noise levels; the number of tomographic measurements per frame was fixed at 2. It is seen that the recovered TACs are close to each other, which shows the robustness of the algorithm against the noise level.

tomographic measurements per frame is kept at two. Notably, the estimated TACs remain similar across all noise levels, highlighting the algorithm's robustness in capturing dynamic changes even in the presence of increased noise. Figure 11(b) shows the nSE versus imaging frame plot for estimates reconstructed from data with varying noise levels. As anticipated, the nSE exhibits an upward trend with increasing noise levels.

The numerical phantom study results underscore the effectiveness of the proposed method in accurately estimating dynamic changes, when limited tomographic measurements per frame are available. The study reveals that the ill-posed nature of the dynamic reconstruction problem

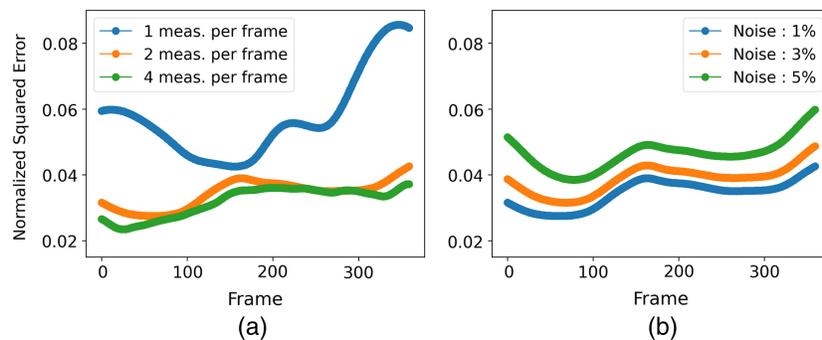


Fig. 11 (a) nSE versus the imaging frame number for reconstructions with varying numbers of tomographic measurements per imaging frame; the noise level was fixed at 1%. A direct correlation between an increase in measurements and reduction in nSE is observable, thus highlighting the improved reconstruction accuracy. (b) nSE versus the imaging frame number for reconstructions from data with varying noise levels; the number of tomographic measurements per frame was fixed at 2. As anticipated, the nSE exhibits an upward trend with increasing noise levels.

diminishes significantly as the number of tomographic measurements per imaging frame increases. Notably, the method's robustness in faithfully capturing the temporal dynamics of the object remains evident even in the presence of increased noise levels. These findings collectively underscore the algorithm's reliability and potential significance in addressing challenges associated with dynamic imaging scenarios.

5.3 Experimental Study Results

Figure 12 displays dynamic PACT images reconstructed from experimental data (bottom row), along with the corresponding reference fluorescence images (top row). The fluorescence images were processed to suppress the background, and the contrast agent is highlighted in yellow. This enhancement was accomplished through manual segmentation of the tube and contrast agent from the raw fluorescence images. The spatiotemporal PACT object estimates were visualized using ParaView.⁵⁷ To ensure a qualitatively close alignment of the field-of-view between the ParaView visualization of the spatiotemporal PACT object estimate and the 2D fluorescence images, the view angle in ParaView was adjusted manually at each frame; however, a slight misalignment remains. A video featuring the raw images, visually enhanced images through

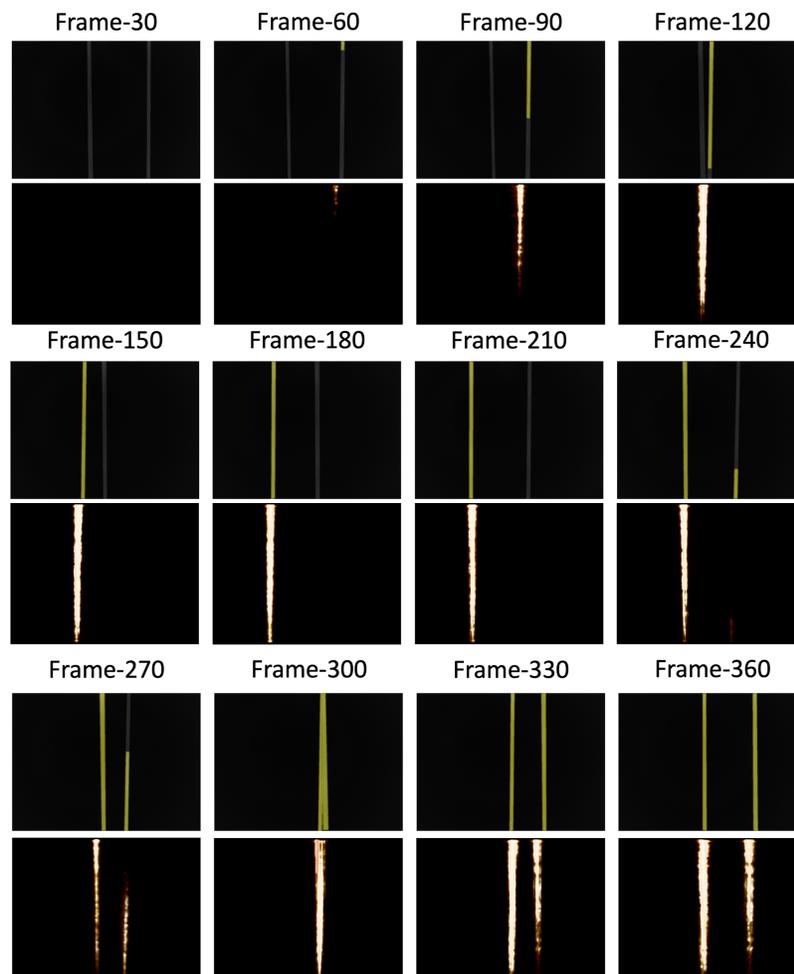


Fig. 12 Sample instances of the dynamic recovery from PACT data (at the bottom of each row) and reference fluorescence images (at the top of each row). The spatiotemporal evolution of the contrast agent inferred from the spatiotemporal estimates reconstructed from PACT data is in strong agreement with the reference images collected by the fluorescence-enabled sCMOS camera. A video featuring the raw images, visually enhanced images through the segmentation of the tube and contrast agent, and ParaView visualizations of the spatiotemporal PACT object estimate is provided in [Video 2](#) ([Video 2](#), MP4, 36.3 MB [URL: <https://doi.org/10.1117/1.JBO.29.S1.S11516.s2>]).

the segmentation of the tube and contrast agent, and ParaView visualizations of the spatiotemporal PACT object estimate is provided in [Video 2](#).

In particular, when examining frames 30 to 120, one can readily observe the accurate recovery of dynamic contrast flow within the tube. A similar trend is evident in frames 210 to 300, with the lower right section of frame-240 showing the presence of the contrast agent, albeit with reduced contrast. This reduced contrast is also noticeable in specific frames, such as 330 and 360, possibly due to light obstruction by other tube segments. Nevertheless, the overall effective recovery of dynamic flow remains apparent, as convincingly demonstrated in [Video 2](#). This highlights the effectiveness and practicality of the proposed method for experimental data beyond simulated measurements, even when only one tomographic measurement per imaging frame is available.

Furthermore, it is worth noting that the spatiotemporal reconstruction offers a frame rate that is equal to the laser pulse rate [10 frames per second (FPS) for the instrument used in the phantom studies]. By contrast, an FBFIR technique would provide a frame rate of 1/36 FPS, which is equal to the reciprocal of the time required for a complete tomographic measurement. This underscores the reconstruction method's potential for monitoring dynamic physiological processes that demand an enhanced frame rate. This can enhance the value of 3D PACT systems that involve rotating measurement gantries for preclinical research, enabling STIR from limited tomographic measurements per imaging frame.

6 Conclusion and Discussion

Volumetric PACT imagers commonly adopt a sequential data acquisition strategy utilizing rotating measurement gantries because of their cost-effectiveness and simplified hardware.^{2,19-21} However, the relatively slow data acquisition times associated with this approach pose significant challenges for dynamic imaging. Consequently, previous studies on dynamic PACT have predominantly focused on 2D dynamic imaging scenarios, leveraging the rapid data acquisition and computationally less demanding image reconstruction calculations of 2D systems. Recognizing that many commercially available volumetric imagers employ sequential data acquisition strategies with rotating gantries, it becomes imperative to advance dynamic image reconstruction techniques tailored to these widely adopted configurations. Addressing this need will enhance the versatility and practical utility of volumetric PACT imagers, allowing for more effective imaging of dynamic processes.

This study presented an accurate and computationally efficient LRME-STIR method for dynamic PACT attuned for commercially available volumetric imagers that employ a rotating measurement gantry in which the tomographic data are sequentially acquired. The implementation of the method was verified by an inverse-crime numerical validation study. The effect of varying number of tomographic measurements per imaging frame and the noise level on the method's accuracy was investigated in an in-silico numerical study. The numerical studies demonstrated that the proposed method is robust against increasing noise levels. They also demonstrated that, as expected, increasing the number of tomographic measurements per frame improves the reconstruction accuracy. Therefore, it is desirable and beneficial for dynamic imaging to design acquisition systems that, like the University of Twente breast imager,⁵² utilize multiple acoustic probes. The experimental study demonstrated the LRME-STIR method's ability to reconstruct the flow of a contrast agent at a frame rate of 10 FPS, even when only a single tomographic measurement per imaging frame was available. Numerical and experimental studies confirm the accuracy of the proposed technique. Thus, this work will potentially have an immediate and sustained positive impact by creating a new capacity to perform dynamic 3D PACT.

The LRME-STIR method proposed in this study employed an accelerated PGD method combined with an OS approach, distinguishing it from previously proposed LRME-STIR methods.^{29,35,36,40} It is important to note that previous studies have demonstrated that the combination of the OS approach with momentum schemes can lead to convergence instability, particularly as the number of subsets increases.^{47,48,58,59} This instability arises due to error accumulation in the momentum term, as the "subset balance" approximation does not perfectly hold.⁴⁷ The "subset balance" approximation implies that the cost function, $\mathbf{J}(\mathbf{F})$, defined in Eq. (11) can be approximated with the OS-based cost function, $\mathbf{J}_{\mathcal{K}_j}(\mathbf{F})$, defined in Eq. (14),

i.e., $\mathbf{J}_{\mathcal{J}_j}(\mathbf{F}) \approx \mathbf{J}(\mathbf{F})$. When a small number of subsets is employed, the “subset balance” approximation is more accurate; thus, the method exhibits more stable convergence characteristics; however, the acceleration from the ordered subset approach is limited in such cases.⁴⁷

To enhance stability, various strategies have been introduced in the literature. For instance, the relaxation of the momentum term has been proposed to mitigate convergence issues.⁴⁷ Another approach involves a monotonic adaptation of the FISTA acceleration, which utilizes objective function values to adaptively determine the subsequent update.⁴⁸ Additionally, an adaptive restarting method that resets the momentum terms based on a stability metric was introduced, further promoting convergence reliability.⁵⁹ Although combining OS and momentum schemes does not provide theoretical guarantees of convergence, in the numerical and experimental studies, the proposed approach demonstrated stable convergence behavior, given the proper selection of the step size and number of subsets. Future refinements of the proposed method may explore the different strategies from the literature to obtain a provably convergent method.^{48,58,59}

Subsequent investigations may focus on evaluating the method’s efficacy in imaging complex dynamic physiological processes and quantifying relevant parameters, such as wash-in and wash-out rates for tumor vascular perfusion.^{37,60} These evaluations might encompass both *in-vivo* and *in-silico* experiments, further advancing the understanding and application of the proposed method.

Disclosures

The authors S. Ermilov, W. Thompson, and M. Anastasio disclose ownership interests in Photosound Technologies, Inc. The other authors have no relevant financial interests and no potential conflicts of interest to disclose.

Code and Data Availability

Data and code are available upon request. Please contact Dr. Villa at uvilla@oden.utexas.edu.

Acknowledgments

The research reported in this publication was supported in part by the Office of The Director, National Institutes of Health (NIH; grant nos. R44OD023029 and R01EB031585). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. X. Wang et al., “Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain,” *Nat. Biotechnol.* **21**(7), 803–806 (2003).
2. H.-P. Brecht et al., “Whole-body three-dimensional optoacoustic tomography system for small animals,” *J. Biomed. Opt.* **14**(6), 064007 (2009).
3. S. Yang et al., “Functional imaging of cerebrovascular activities in small animals using high-resolution photoacoustic tomography,” *Med. Phys.* **34**(8), 3294–3301 (2007).
4. G. C. Kagadis et al., “In vivo small animal imaging: current status and future prospects,” *Med. Phys.* **37**(12), 6421–6442 (2010).
5. G. Loudos, G. C. Kagadis, and D. Psimadas, “Current status and future perspectives of in vivo small animal imaging using radiolabeled nanoparticles,” *Eur. J. Radiol.* **78**(2), 287–295 (2011).
6. B. L. Franc et al., “Small-animal SPECT and SPECT/CT: important tools for preclinical investigation,” *J. Nucl. Med.* **49**(10), 1651–1663 (2008).
7. P. Carmeliet, “Angiogenesis in life, disease and medicine,” *Nature* **438**(7070), 932–936 (2005).
8. P. Carmeliet and R. K. Jain, “Angiogenesis in cancer and other diseases,” *Nature* **407**(6801), 249–257 (2000).
9. J. Xia and L. V. Wang, “Small-animal whole-body photoacoustic tomography: a review,” *IEEE Trans. Biomed. Eng.* **61**(5), 1380–1389 (2014).
10. L. Li et al., “Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution,” *Nat. Biomed. Eng.* **1**(5), 0071 (2017).
11. P. K. Upputuri and M. Pramanik, “Dynamic in vivo imaging of small animal brain using pulsed laser diode-based photoacoustic tomography system,” *J. Biomed. Opt.* **22**(9), 090501 (2017).
12. S. Shi et al., “Thermosensitive biodegradable copper sulfide nanoparticles for real-time multispectral optoacoustic tomography,” *ACS Appl. Biomater.* **2**(8), 3203–3211 (2019).

13. S. Manohar and M. Dantuma, "Current and future trends in photoacoustic breast imaging," *Photoacoustics* **16**, 100134 (2019).
14. L. Lin et al., "Single-breath-hold photoacoustic computed tomography of the breast," *Nat. Commun.* **9**(1), 2352 (2018).
15. J. Gamelin et al., "A real-time photoacoustic tomography system for small animals," *Opt. Express* **17**(13), 10489–10498 (2009).
16. W. He et al., "Plasmonic titanium nitride nanoparticles for in vivo photoacoustic tomography imaging and photothermal cancer therapy," *Biomaterials* **132**, 37–47 (2017).
17. P. K. Upputuri et al., "A high-performance compact photoacoustic tomography system for in vivo small-animal brain imaging," *J. Visual. Exp.* (124), e55811 (2017).
18. T. Shan et al., "In-vivo hemodynamic imaging of acute prenatal ethanol exposure in fetal brain by photoacoustic tomography," *J. Biophotonics* **13**(5), e201960161 (2020).
19. H. P. Brecht et al., "A 3D imaging system integrating photoacoustic and fluorescence orthogonal projections for anatomical, functional and molecular assessment of rodent models," *Proc. SPIE* **10494**, 1049411 (2018).
20. L. Lin et al., "High-speed three-dimensional photoacoustic computed tomography for preclinical research and clinical translation," *Nat. Commun.* **12**(1), 882 (2021).
21. S. A. Ermilov et al., "Laser optoacoustic imaging system for detection of breast cancer," *J. Biomed. Opt.* **14**(2), 024007 (2009).
22. L. Xiang et al., "4-D photoacoustic tomography," *Sci. Rep.* **3**(1), 1–8 (2013).
23. A. Hauptmann et al., "Model-based learning for accelerated, limited-view 3-D photoacoustic tomography," *IEEE Trans. Med. Imaging* **37**(6), 1382–1393 (2018).
24. J. Schwab, S. Antholzer, and M. Haltmeier, "Learned backprojection for sparse and limited view photoacoustic tomography," *Proc. SPIE* **10878**, 1087837 (2019).
25. J. Friel and E. T. Quinto, "Artifacts in incomplete data tomography with applications to photoacoustic tomography and sonar," *SIAM J. Appl. Math.* **75**(2), 703–725 (2015).
26. J.-F. Cai et al., "Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study," *IEEE Trans. Med. Imaging* **33**(8), 1581–1591 (2014).
27. M. Wernick, E. Infusino, and M. Milosevic, "Fast spatio-temporal image reconstruction for dynamic PET," *IEEE Trans. Med. Imaging* **18**(3), 185–195 (1999).
28. Q. Ding, M. Burger, and X. Zhang, "Dynamic SPECT reconstruction with temporal edge correlation," *Inverse Probl.* **34**(1), 014005 (2017).
29. J. P. Haldar and Z.-P. Liang, "Spatiotemporal imaging with partially separable functions: a matrix recovery approach," in *IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, IEEE, pp. 716–719 (2010).
30. K. Wang et al., "Fast spatiotemporal image reconstruction based on low-rank matrix estimation for dynamic photoacoustic computed tomography," *J. Biomed. Opt.* **19**(5), 056007 (2014).
31. S. Arridge et al., "Accelerated high-resolution photoacoustic tomography via compressed sensing," *Phys. Med. Biol.* **61**(24), 8908 (2016).
32. A. Özbek, X. L. Deán-Ben, and D. Razansky, "Compressed optoacoustic sensing of volumetric cardiac motion," *IEEE Trans. Med. Imaging* **39**(10), 3250–3255 (2020).
33. F. Lucka et al., "Enhancing compressed sensing 4D photoacoustic tomography by simultaneous motion estimation," *SIAM J. Imaging Sci.* **11**(4), 2224–2253 (2018).
34. K. Wang et al., "Accelerating image reconstruction in three-dimensional optoacoustic tomography on graphics processing units," *Med. Phys.* **40**(2), 023301 (2013).
35. Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *4th IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, IEEE, pp. 988–991 (2007).
36. C. Brinegar et al., "Real-time cardiac MRI using prior spatial-spectral information," in *Annu. Int. Conf. IEEE Eng. in Med. and Biol. Soc.*, IEEE, pp. 4383–4386 (2009).
37. R. M. Cam et al., "Dynamic image reconstruction to monitor tumor vascular perfusion in small animals using 3D photoacoustic computed-tomography imagers with rotating gantries," *Proc. SPIE* **12379**, 123790F (2023).
38. R. M. Cam et al., "Low-rank matrix estimation-based spatiotemporal image reconstruction from few tomographic measurements per frame for dynamic photoacoustic computed tomography," *Proc. SPIE* **12463**, 124630R (2023).
39. L. Lozanski, M. A. Anastasio, and U. Villa, "A memory-efficient self-supervised dynamic image reconstruction method using neural fields," *IEEE Trans. Comput. Imaging* **8**, 879–892 (2022).
40. S. G. Lingala et al., "Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt SLR," *IEEE Trans. Med. Imaging* **30**(5), 1042–1054 (2011).
41. S. Gu et al., "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 2862–2869 (2014).

42. R. Tibshirani et al., “Proximal gradient descent and acceleration,” Lecture Notes, <http://www.stat.cmu.edu/ryantibs/convexopt-S15/lectures/08-prox-grad.pdf> (2010).
43. P. L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. Bauschke et al., Eds., pp. 185–212, Springer (2011).
44. A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009).
45. Y. Nesterov, “Gradient methods for minimizing composite functions,” *Math. Program.* **140**(1), 125–161 (2013).
46. H. M. Hudson and R. S. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” *IEEE Trans. Med. Imaging* **13**(4), 601–609 (1994).
47. D. Kim, S. Ramani, and J. A. Fessler, “Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction,” *IEEE Trans. Med. Imaging* **34**(1), 167–178 (2014).
48. V. Haase et al., “An improved combination of ordered subsets and momentum for fast model-based iterative CT reconstruction,” *Proc. SPIE* **11312**, 113121N (2020).
49. J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM J. Optim.* **20**(4), 1956–1982 (2010).
50. N. Halko, P.-G. Martinsson, and J. A. Tropp, “Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM Rev.* **53**(2), 217–288 (2011).
51. W. R. Thompson et al., “Characterizing a photoacoustic and fluorescence imaging platform for preclinical murine longitudinal studies,” *J. Biomed. Opt.* **28**(3), 036001 (2023).
52. S. M. Schoustra et al., “Twente Photoacoustic Mammoscope 2: system overview and three-dimensional vascular network images in healthy breasts,” *J. Biomed. Opt.* **24**(12), 121909 (2019).
53. K. Ito, B. Jin, and T. Takeuchi, “Multi-parameter Tikhonov regularization—an augmented approach,” *Chin. Ann. Math. Ser. B* **35**(3), 383–398 (2014).
54. S. Pereverzev and E. Schock, “On the adaptive selection of the parameter in regularization of ill-posed problems,” *SIAM J. Numer. Anal.* **43**(5), 2060–2076 (2005).
55. W. Thompson et al., “A preclinical small animal imaging platform combining multi-angle photoacoustic and fluorescence projections into co-registered 3D maps,” *Proc. SPIE* **11240**, 112400L (2020).
56. M. Xu and L. V. Wang, “Universal back-projection algorithm for photoacoustic computed tomography,” *Phys. Rev. E* **71**(1), 016706 (2005).
57. J. Ahrens, B. Geveci, and C. Law, *ParaView: An End-User Tool for Large Data Visualization*, pp. 717–731, Elsevier Butterworth-Heinemann, Burlington, MA, USA (2005).
58. D. Kim and J. A. Fessler, “Ordered subsets acceleration using relaxed momentum for x-ray CT image reconstruction,” in *IEEE Nucl. Sci. Symp. and Med. Imaging Conf. (2013 NSS/MIC)*, IEEE, pp. 1–5 (2013).
59. A. Sisniega et al., “Accelerated 3D image reconstruction with a morphological pyramid and noise-power convergence criterion,” *Phys. Med. Biol.* **66**(5), 055012 (2021).
60. C. W. Huppel et al., “A light-fluence-independent method for the quantitative analysis of dynamic contrast-enhanced multispectral photoacoustic tomography (DCE MSOT),” *Photoacoustics* **10**, 54–64 (2018).

Refik Mert Cam is a PhD candidate in electrical and computer engineering at the University of Illinois Urbana-Champaign. He obtained his BS degree in electrical and electronics engineering from the Middle East Technical University, Turkey, in 2020. His research interests include inverse problems, medical image reconstruction, and deep learning for medical imaging. He is a student member of SPIE and 2023–2024 Mavis Future Faculty Fellow at Grainger College of Engineering of the University of Illinois Urbana-Champaign.

Chao Wang received her BS degree from Jilin University in China. She obtained her PhD in computational mathematics from Peking University in 2019. She worked as a postdoc at the University of Illinois at Urbana-Champaign in 2020 and as a research fellow at the National University of Singapore from 2021 to 2023. Her research involves several aspects of computational mathematics, including PDE-constrained inverse problems, medical imaging, generative neural networks in operator learning, and modeling of proton therapy.

Weylan Thompson received his BS degree in physics from McGill University in 2015. He joined PhotoSound Technologies, Inc., in 2018 and has since been involved in developing and characterizing photoacoustic imaging systems and data acquisition devices.

Sergey A. Ermilov holds his master of biomedical engineering degree from Bauman Moscow State Technical University, Russia, and his PhD in bioengineering from Rice University,

Houston, Texas, United States. His scientific expertise and interests are in the field of biomedical photoacoustics and include development of theoretical aspects, system modeling, development of multimodality instruments, and data processing. He is an author of 24 peer-reviewed publications and 19 patents related to the field of photoacoustic imaging.

Mark A. Anastasio is a Donald Biggar Willett professor in engineering and the head of the Department of Bioengineering, University of Illinois at Urbana–Champaign, Urbana, Illinois, United States. He received his PhD from the University of Chicago, Chicago, Illinois, United States, in 2001. His research broadly addresses computational image science, inverse problems in imaging, and machine learning for imaging applications. He has contributed broadly to emerging biomedical imaging technologies that include photoacoustic computed tomography, ultrasound computed tomography, and x-ray phase-contrast imaging. His work has been supported by numerous NIH grants, and he served for 2 years as the chair of the NIH EITA Study Section. He is a fellow of SPIE, the American Institute for Medical and Biological Engineering, the International Academy of Medical and Biological Engineering, and the Institute of Electrical and Electronics Engineers.

Umberto Villa is a research scientist at Oden Institute for Computational Engineering and Science, University of Texas at Austin, Texas, United States. He received his BS and MS degrees in mathematical engineering from Politecnico di Milano, Milan, Italy, in 2005 and 2007, respectively, and his PhD in mathematics from Emory University, Atlanta, Georgia, United States, in 2012. His research interests lie in the computational and mathematical aspects of large-scale inverse problems, imaging science, and uncertainty quantification.