

Photoacoustic Reconstruction Using Sparsity in Curvelet Frame: Image versus Data Domain

Bolin Pan, Simon R. Arridge, Felix Lucka, Ben T. Cox, Nam Huynh, Paul C. Beard, Edward Z. Zhang and Marta M. Betcke

Abstract—Curvelet frame is of special significance for photoacoustic tomography (PAT) due to its sparsifying and microlocalisation properties. We derive a one-to-one map between wavefront directions in image and data spaces in PAT which suggests near equivalence between the recovery of the initial pressure and PAT data from compressed/subsampled measurements when assuming sparsity in Curvelet frame. As the latter is computationally more tractable, investigation to which extent this equivalence holds conducted in this paper is of immediate practical significance. To this end we formulate and compare DR, a two step approach based on the recovery of the complete volume of the photoacoustic data from the subsampled data followed by the acoustic inversion, and $p_0\mathbf{R}$, a one step approach where the photoacoustic image (the initial pressure, p_0) is directly recovered from the subsampled data. Effective representation of the photoacoustic data requires basis defined on the range of the photoacoustic forward operator. To this end we propose a novel *wedge-restriction* of Curvelet transform which enables us to construct such basis. Both recovery problems are formulated in a variational framework. As the Curvelet frame is heavily overdetermined, we use reweighted ℓ_1 norm penalties to enhance the sparsity of the solution. The data reconstruction problem DR is a standard compressed sensing recovery problem, which we solve using an ADMM-type algorithm, SALSA. Subsequently, the initial pressure is recovered using time reversal as implemented in the k-Wave Toolbox. The p_0 reconstruction problem, $p_0\mathbf{R}$, aims to recover the photoacoustic image directly via FISTA, or ADMM when in addition including a non-negativity constraint. We compare and discuss the relative merits of the two approaches and illustrate them on 2D simulated and 3D real data in a fair and rigorous manner.

Index Terms—Sparsity, compressed sensing, Curvelet frame, ADMM methods, ℓ_1 minimization, photoacoustic tomography.

I. INTRODUCTION

SPARSITY is a powerful property closely related to the information content carried by a signal. Sparsity of a signal is defined with respect to a given basis¹, and signals may be highly sparse in some bases while not sparse in others. Thus the choice of the right basis is paramount to achieve effective sparse representation and hence to the success of sparsity enhanced signal recovery also known as *compressed sensing*. The

Bolin Pan, Simon R. Arridge and Marta M. Betcke are with the Department of Computer Science, University College London, UK, WC1E 6BT London, UK, e-mail: m.betcke@ucl.ac.uk

Felix Lucka is with Department of Computational Imaging, Centrum Wiskunde & Informatica, Netherlands, P.O. Box 94079 1090 GB Amsterdam, Netherlands.

Ben T. Cox and Nam Huynh and Edward Z. Zhang and Paul C. Beard are with Department of Medical Physics, University College London, UK, WC1E 6BT London, UK.

Nam Huynh and Edward Z. Zhang and Paul C. Beard are with Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, UK, WC1E 6BT London, UK.

¹We refer to a tight frame as a normalised (overdetermined) basis

theoretical underpinning of the field of compressed sensing was laid in a series of seminal papers by Candès, Tao [1], Romberg [2], [3], and Donoho [4].

A. Motivation

This work exploits the observation that in photoacoustic tomography, similar to X-ray tomography, both the photoacoustic *image* (initial pressure) and the photoacoustic *data* (pressure time series recorded by the detector) exhibit sparsity in their respective domains. The photoacoustic image can be thought of as a function with singularities along \mathcal{C}^2 smooth curves for which Curvelets provide a nearly optimally sparse representation [5]. On the other hand, in the regime of geometrical optics under the action of the wave equation, Curvelets have been shown to essentially propagate along the trajectories of the underlying Hamiltonian system (the projections on the physical space of the solutions of the underlying Hamiltonian system) [6]. Assuming the initial pressure to be a single Curvelet, the wave field induced by such initial pressure at any given point in time can be approximated by simply propagating the corresponding Curvelet along the geometrical ray initiating at the centre of the Curvelet with the initial direction perpendicular to its wavefront. Consequently, if the initial pressure is sparse in the Curvelet frame the acoustic propagation effectively preserves the sparsity of the initial pressure. However, this argument does not go far enough to claim sparsity of the photoacoustic data where the acoustic wavefront group propagating in physical space (\mathbb{R}^d) is mapped to a function of time and detector coordinate (data space) via the dispersion relation given by the wave equation. In Section III-A we derive a one-to-one map between the physical space (and by Hamiltonian flow argument also the image space) and the data space therewith essentially inducing the same sparsity of the initial pressure onto the photoacoustic data. Motivated by these results, we hypothesise near equivalence of the sparse representation of both the initial pressure (photoacoustic image) and the photoacoustic data in Curvelet frame.

B. Contribution

In this paper we focus on the problem of image reconstruction from subsampled photoacoustic data measured using a Fabry-Pérot planar detector scanner. An example of such a system is the single point multiwavelength backward-mode scanner [7], [8]. The proposed methodology also directly applies to compressed measurements, provided the sensing patterns form a (scrambled) unitary transform e.g. scrambled

Hadamard transform. A PAT system capable of compressed measurements acquisition was introduced in [9], [10], [11].

As outlined in the motivation, we argue the near optimal sparsity of the Curvelet representation of both the entire volume of the photoacoustic data and the photoacoustic image (initial pressure) and stipulate a near equivalence of both recovery problems based on a one-to-one wavefront direction map which we derive. To capitalise on the sparsity of data, we deploy a *two step approach* which first solves a basis pursuit problem to recover the full photoacoustic data volume from subsampled/compressed measurements, followed by a standard acoustic inversion to obtain the initial pressure (**DR**). We demonstrate that efficient and robust representation of the photoacoustic data volume requires basis which is defined on the range of the forward operator and propose a novel *wedge-restriction* of the standard Curvelet transform which allows us to construct such basis. On the other hand, motivated by the sparsity of the photoacoustic image in the Curvelet frame, we consider a *one step approach*, which recovers the photoacoustic image p_0 directly from the subsampled measurements ($p_0\mathbf{R}$).² In contrast to prior work [12], our new data reconstruction method builds upon the one-to-one wavefront direction map and captures the relationship between the time-steps. Therefore, it has the potential to perform on a par with the $p_0\mathbf{R}$ approach *without the need to iterate with the expensive forward operator*. To verify this hypothesis we rigorously compare these two strategies on an appropriately constructed simulated 2D example and real 3D PAT data. In both cases the sparse recovery is formulated in a variational framework, utilising an iteratively reweighted ℓ_1 norm to more aggressively pursue sparsity in an overdetermined frame, as appropriate for the Curvelets basis used herein. We discuss appropriate choices of algorithms for the solution of the two resulting optimisation problems. Finally, while we focus the presentation on photoacoustic tomography, the methodology applies to other imaging modalities where the forward operator admits microlocal characterization, most prominently X-ray computed tomography.

C. Related work

A number of one and two step approaches have been previously proposed in the context of photoacoustic tomography. To the best of our knowledge [13] presents the first one step approach for a circular geometry with focused transducers for 2D imaging using sparse representation in Wavelets, Curvelets (with Wavelets at the finest scale), and Fourier domains, and first derivative (total variation). The 3D reconstruction problem for PAT with an array of ultrasonic transducers using sparsity in Wavelet frame was presented in [14]. An iterative reconstruction approach based on an algebraic adjoint with sparse sampling for 2D and 3D photoacoustic images was first introduced in [15]. Two analytical adjoints have been subsequently proposed, one in a BEM-FEM setting [16] and the other in k-space setting [17], and exploited/refined in a series of papers [18], [19], [20]. The one step approach in the

present work builds upon [17], [18] and extends it to sparsity in the Curvelet frame in 2D and 3D.

The two step approach suggested in [12] recovers the photoacoustic data in every time step independently from pattern measurements under assumption of sparsity of the photoacoustic pressure on the detector at a fixed time in a 2D Curvelet frame. It has the advantage that each of the time step recovery problems can be solved independently, allowing trivial parallelization. The disadvantage is that the correspondence between the time steps is lost: *i.e.* the recovery does not take into account that the photoacoustic data corresponds to the solution to the same wave equation evaluated at the detector at different time steps. In another approach proposed in [21] the authors explore the temporal sparsity of the entire photoacoustic data time series using a custom made transform in time. Then the initial pressure is reconstructed via the universal back-projection (UBP) method [22].

In the two step approach proposed here, we use a variational formulation akin to the one in [12] while we represent the entire volume of data, *i.e.* sensor location and time, akin to [21], but in the Curvelet basis. This approach combines the best of both worlds: representing the entire data volume naturally preserves the connection between the time steps, while, as indicated above and discussed in Section III-D2, essentially optimal data sparsity is achieved using such 2D/3D Curvelet representation.

D. Outline

The remainder of this paper is structured as follows. In Section II we provide the background and formulation of the forward, adjoint and inverse problems in PAT. Section III discusses the details of sparse Curvelet representation of PAT image and PAT data. First, we derive a one-to-one map between wavefronts in ambient space (and hence image space) and data space which motivates use of Curvelet transform also for the representation of the PAT data volume. Then we briefly recall the Curvelet transform and propose a concept of *wedge restriction* which allows us to formulate a tight frame of *wedge restricted* Curvelets on the range of the PAT forward operator for PAT data representation. Finally, we consider details specific to PAT data such as symmetries, time oversampling and their consequences. In Section IV we formulate the reconstruction problem from subsampled measurements via **DR** and $p_0\mathbf{R}$ approaches in a variational framework, and propose tailored algorithms for their solution. In Section V the relative performance of both approaches is rigorously evaluated on a specially constructed 2D example and on experimental 3D real data of a human palm vasculature. We also compare our one step approach using Curvelets with one step approach using Wavelet bases and against state of the art total variation regularisation. We evaluate robustness of both one and two step approaches to noise and more aggressive subsampling. We summarize the results of our work and point out future research directions in Section VI.

II. PHOTOACOUSTIC TOMOGRAPHY

Photoacoustic tomography (PAT) has been steadily gaining in importance for biomedical and preclinical applications. The

²A flow chart diagram can be found in supplementary material

fundamental advantage of the technique is that optical contrast is encoded on to acoustic waves, which are scattered much less than photons. It therefore avoids the depth and spatial resolution limitations of purely optical imaging techniques that arise due to strong light scattering in tissue: with PAT, depths of a few centimetres with scalable spatial resolution ranging from tens to hundreds of microns (depending on depth) are achievable. However, obtaining high resolution 3D PAT images with a Nyquist limited spatial resolution of the order of 100 μm requires spatial sampling intervals on a scale of tens of μm over cm scale apertures. This in turn necessitates recording the photoacoustic waves at several thousand different spatial points which results in excessively long acquisition times using sequential scanning schemes, such as the Fabry–Pérot based PA scanner, or mechanically scanned piezoelectric receivers. In principle, this can be overcome by using an array of detectors. However, a fully sampled array comprising several thousand detector elements would be prohibitively expensive. Hence, new methods that can reconstruct an accurate image using only a subset of this data would be highly beneficial in terms of acquisition speed or cost.

A PAT system utilises a near infrared laser pulse which is delivered into biological tissue. Upon delivery some of the deposited energy will be absorbed and converted into heat which is further converted to pressure. This pressure is then released in the form of acoustic pressure waves propagating outwards. When this wave impinges on the ultrasound detector $\mathcal{S} \subset \mathbb{R}^d$, it is recorded at a set of locations $\mathbf{x}_S \in \mathcal{S}$ over time constituting the photoacoustic data $g(t, \mathbf{x}_S)$ which is subsequently used for recovery of the initial pressure p_0 .

With several assumptions on the tissues properties [23], the photoacoustic forward problem can be modelled as an initial value problem for the free space wave equation

$$\underbrace{\left(\frac{1}{c^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} - \nabla^2 \right)}_{:=\square^2} p(t, \mathbf{x}) = 0, \quad (t, \mathbf{x}) \in (0, T) \times \mathbb{R}^d, \quad (\mathbf{A})$$

$$p(0, \mathbf{x}) = p_0(\mathbf{x}), \quad p_t(0, \mathbf{x}) = 0,$$

where $p(t, \mathbf{x})$ is time dependent acoustic pressure in $C^\infty((0, T) \times \mathbb{R}^d)$ and $c(\mathbf{x}) \in C^\infty(\Omega)$ is the speed of sound in the tissue.

The photoacoustic inverse problem recovers initial pressure $p_0(\mathbf{x}) \in C_0^\infty(\Omega)$ in the region of interest Ω on which $p_0(\mathbf{x})$ is compactly supported, from time dependent measurements

$$g(t, \mathbf{x}_S) = \omega(t)p(t, \mathbf{x}), \quad (t, \mathbf{x}_S) \in (0, T) \times \mathcal{S}$$

on a surface \mathcal{S} , e.g. the boundary of Ω , where $\omega \in C_0^\infty(0, T)$ is a temporal smooth cut-off function. It amounts to a solution of the following initial value problem for the wave equation with constraints on the surface [24]

$$\begin{aligned} \square^2 p(t, \mathbf{x}) &= 0, & (t, \mathbf{x}) &\in (0, T) \times \Omega, \\ p(0, \mathbf{x}) &= 0, \quad p_t(0, \mathbf{x}) &= 0, & (\mathbf{TR}) \\ p(t, \mathbf{x}) &= g(T - t, \mathbf{x}_S), & (t, \mathbf{x}_S) &\in (0, T) \times \mathcal{S}. \end{aligned}$$

also referred to as a *time reversal*. For non-trapping smooth $c(\mathbf{x})$, in 3D, the solution of **(TR)** is the solution of the PAT inverse problems if T is chosen large enough so that

$g(t, \mathbf{x}) = 0$ for all $\mathbf{x} \in \Omega, t \geq T$ and the wave has left the domain Ω . Assuming that the measurement surface \mathcal{S} surrounds the region of interest Ω containing the support of initial pressure p_0 , the wave equation **(TR)** has a unique solution. When no complete data is available (meaning that some singularities have not been observed at all as opposed to the case where partial energy has been recorded, see e.g. [25]), p_0 is usually recovered in a variational framework, including some additional penalty functional corresponding to prior knowledge about the solution. Iterative methods for solution of such problems require application of the adjoint operator which amounts to the solution of the following initial value problem with time varying mass source [17]

$$\begin{aligned} \square^2 q(t, \mathbf{x}) &= \begin{cases} \frac{\partial}{\partial t} g(T - t, \mathbf{x}_S), & (t, \mathbf{x}_S) \in (0, T) \times \mathcal{S} \\ 0 & \text{everywhere else} \end{cases} \\ q(0, \mathbf{x}) &= 0, \quad q_t(0, \mathbf{x}) = 0 \end{aligned} \quad (\mathbf{A}^*)$$

evaluated at T , $q(T, \mathbf{x})$. All the equations **(A)**, **(TR)** and **(A*)** can be solved using the pseudospectral method implemented in the k-Wave toolbox for heterogeneous media in two and three dimensions [26]; see [17] for implementation of the adjoint **(A*)** in this setting.

III. MULTISCALE REPRESENTATION OF PHOTOACOUSTIC DATA AND INITIAL PRESSURE

The presentation in this section assumes a cuboid domain Ω with a flat panel sensor \mathcal{S} at one of its sides and constant speed of sound $c(\mathbf{x}) = c$.

A. Wavefront direction mapping

The solution of wave equation **(A)** for a constant speed of sound $c(\mathbf{x}) = c$, assuming half space symmetry as per planar sensor geometry, can be explicitly written in Fourier domain as [27]

$$\hat{p}(t, \mathbf{k}) = \cos(c|\mathbf{k}|t)\hat{p}_0(\mathbf{k}), \quad (1)$$

where \hat{p} denotes the Fourier transform of p (in all or some variables) and \mathbf{k} is the Fourier domain wave vector.

Using (1), the following mapping can be obtained between the initial pressure p_0 and the PAT data $p(ct, \mathbf{x}_S)$ in Fourier domain (note we use c normalised time/frequencies, the non-normalised formulation contains additional factor $1/c$ on the right hand side) [28], [29], [12]

$$\hat{p}(\omega/c, \mathbf{k}_S) = \frac{\omega/c}{\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}} \hat{p}_0\left(\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}, \mathbf{k}_S\right). \quad (2)$$

Equation (2) describes change from Cartesian to hyperbolic coordinates including scaling by the Jacobian determinant. For this mapping to be real and bounded, the expression under the square root has to be positive bounded away from 0 for all non-zero wave vectors, i.e. $\mathcal{R}_A = \{(\omega/c, \mathbf{k}_S) \in \mathbb{R}^d : \omega/c > |\mathbf{k}_S|\} \cup \mathbf{0}$, resulting in a bow-tie shaped range of the photoacoustic operator in the sound-speed normalised Fourier space as illustrated by the contour plot of $\mathbf{k}_\perp = \text{sign}(\omega)\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}$, $\mathbf{k}_\perp = \text{const}$, in Figure 1.

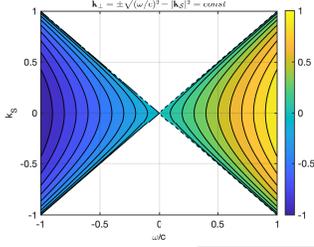


Fig. 1. Contour plot of $\mathbf{k}_\perp = \text{sign}(\omega) \sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2} = \text{const}$ over the bow-tie shaped frequency domain range of the photoacoustic forward operator, $\mathcal{R}_A = \{(\omega/c, \mathbf{k}_S) \in \mathbb{R}^d : \omega/c > |\mathbf{k}_S|\} \cup \mathbf{0}$.

The change of coordinates described by equation (2) (which is a consequence of the dispersion relation in the wave equation $(\omega/c)^2 = |\mathbf{k}|^2 = |\mathbf{k}_S|^2 + |\mathbf{k}_\perp|^2$) defines a one-to-one map between the frequency vectors $\mathbf{k} \in \mathbb{R}^d$ and $(\omega/c, \mathbf{k}_S) \in \mathcal{R}_A$

$$\begin{aligned} \mathbf{k} &= (\mathbf{k}_\perp, \mathbf{k}_S) \leftarrow \left(\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}, \mathbf{k}_S \right), \\ (\omega/c, \mathbf{k}_S) &\leftarrow \left(|\mathbf{k}| = \sqrt{|\mathbf{k}_\perp|^2 + |\mathbf{k}_S|^2}, \mathbf{k}_S \right). \end{aligned} \quad (3)$$

This mapping is illustrated below. Figure 2 shows the wavefronts and the corresponding wavefront vectors in the ambient space $\hat{\theta} = (-\cos \theta, \sin \theta)$ on the left and in the data space $\hat{\beta} = (\cos \beta, -\sin \beta)$ on the right. Note that these wavefront vectors have the same direction as their frequency domain counterparts $\mathbf{k}_{\hat{\theta}}, \mathbf{k}_{\hat{\beta}}$ which are depicted in Figure 3.

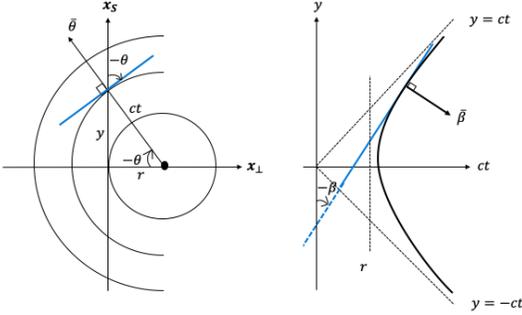


Fig. 2. Wavefront direction mapping between the ambient space $\hat{\theta}$ (left), and the data space, $\hat{\beta}$ (right) on an example of spherical wave in 2D (which can be seen as a cross-section of a 3D problem with rotational symmetry around ct axis). $-\theta$ and $-\beta$ are the angles that the wavefronts make with the sensor plane. Their signs follow the convention of right handed coordinate system $(\mathbf{x}_\perp, \mathbf{x}_S)$. The respective wavefronts (their tangent planes) are depicted in blue.

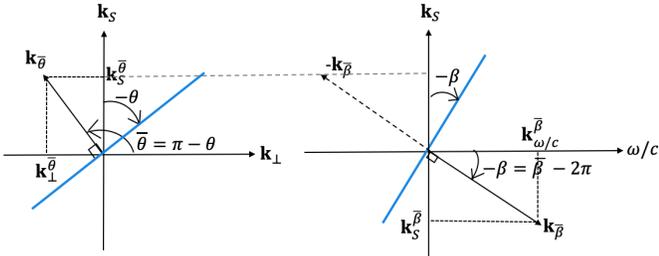


Fig. 3. Frequency domain visualization of the example in Figure 2. Fourier domain wavefront vector mapping between the ambient space, $\mathbf{k}_{\hat{\theta}}$ (left), and the data space, $\mathbf{k}_{\hat{\beta}}$ (right).

Formula (2) has an implicit assumption of even symmetry of $p_0(\mathbf{x}_\perp, \mathbf{x}_S)$, $p(ct, \mathbf{x}_S)$ in \mathbb{R}^d w.r.t. $\mathbf{x}_\perp = 0$, $ct = 0$, respectively, which is inherited by the map (3)

$$\begin{aligned} \hat{p}_0(-\mathbf{k}_\perp, \mathbf{k}_S) &= \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S), \\ \hat{p}(-\omega/c, \mathbf{k}_S) &= \hat{p}(\omega/c, \mathbf{k}_S). \end{aligned} \quad (4)$$

We note, that the even symmetry mirrors the spherical wave source (and consequently the data) in Figures 2, 3 w.r.t. the sensor plane. As a result we actually have two wavefronts synchronously impinging at the same point of the detector with angles $-\theta$ (original p_0) and θ (mirrored p_0). The first is mapped to wavefront with angle $-\beta$ (original data $p(\omega/c, \mathbf{x}_S)$) and β (mirrored data $p(-\omega/c, \mathbf{x}_S)$).

In contrast, the symmetric Curvelet transform introduced in Section III-B below assumes Hermitian symmetry which holds for Fourier transforms of real functions. Thus computing symmetric Curvelet transform of un-mirrored real functions³: initial pressure $p_0(\mathbf{x}_\perp, \mathbf{x}_S)$ (supported on $\mathbf{x}_\perp \geq 0$ and 0 otherwise) and data $p(\omega/c, \mathbf{x}_S)$ (supported on $\omega/c \geq 0$ and 0 otherwise), we instead continue the map (3) with Hermitian symmetry into the whole of \mathbb{R}^d

$$\begin{aligned} \hat{p}_0(-\mathbf{k}_\perp, \mathbf{k}_S) &= \text{Conj}(\hat{p}_0(\mathbf{k}_\perp, -\mathbf{k}_S)) \\ \hat{p}(-\omega/c, \mathbf{k}_S) &= \text{Conj}(\hat{p}(\omega/c, -\mathbf{k}_S)). \end{aligned} \quad (5)$$

Here $\text{Conj}(\cdot)$ denotes complex conjugate. The same Hermitian symmetry underpins the wavefront vectors which orientation is arbitrary i.e. $\mathbf{k}_{\hat{\theta}}$ vs. $-\mathbf{k}_{\hat{\theta}}$ or $\mathbf{k}_{\hat{\beta}}$ vs. $-\mathbf{k}_{\hat{\beta}}$ (the latter is highlighted in Figure 3(right)). We note, that neither the mirroring nor the Hermitian symmetry affect the wavefront angle. While mirroring duplicates the wavefronts in a sense explained above, this is not the case for Hermitian symmetry as the wavefront vector is defined up to the Hermitian symmetry. In conclusion, both formulations coincide on $\omega/c \geq 0$, $\mathbf{k}_\perp \geq 0$ and consequently wavefront angle mapping derived from mirrored formulation holds valid for the un-mirrored formulation on $\omega/c \geq 0$, $\mathbf{k}_\perp \geq 0$ but with Hermitian instead of even symmetry.

With the notation in Figures 2 and 3 we immediately write the map between the wavefront vector angles⁴ (which are the same in space and frequency domains)

$$\tan \bar{\beta} = \frac{\mathbf{k}_S^{\bar{\beta}}}{\mathbf{k}_{\omega/c}^{\bar{\beta}}} = \frac{-\mathbf{k}_S^{\bar{\theta}}}{|\mathbf{k}^{\bar{\theta}}|} = \frac{-\mathbf{k}_S^{\bar{\theta}}}{\sqrt{(|\mathbf{k}_\perp^{\bar{\theta}}|^2 + |\mathbf{k}_S^{\bar{\theta}}|^2)}} = \sin(-\bar{\theta}). \quad (6)$$

Expressing (6) in terms of the detector impingement angles using their relation with wavefront vector angles $\bar{\theta} = \pi - \theta$ and $\bar{\beta} = 2\pi - \beta$ and basic trigonometric identities we obtain the corresponding map between θ and β (note that actual impingement angles are negatives)

$$\beta = \arctan(\sin \theta). \quad (7)$$

While the derivation was done in 2D for clarity it is easily lifted to 3D. In 3D we have two angles describing the wavefront in ambient space $\theta = (\theta^1, \theta^2)^T$ (in planes spanned by \mathbf{k}_\perp and each detector coordinate axis $\hat{\mathbf{k}}_S^i$) and analogously in data space $\beta = (\beta^1, \beta^2)^T$, with the map (7) and all statements holding component wise.

We would like to highlight that this map only depends on the angle θ (not the distance r nor translation along the detector). The wavefront direction mapping is universal across the entire

³To keep the notation light, we do not explicitly distinguish between mirrored and un-mirrored quantities as it should be clear from the context.

⁴For simplicity we use $\hat{\theta}, \hat{\beta}$ to represent both the spatial domain vectors and their angles depending on the context.

data volume regardless of the position of the source w.r.t. the detector. For the un-mirrored initial pressure, as θ sweeps the angle range $(-\pi/2, \pi/2)$, β sweeps $(-\pi/4, \pi/4)$. For mirrored initial pressure, these angles can be simply negated which corresponds to mirroring them around the sensor plane. This is consistent with the hyperbolic envelope, which restricts the range of the PAT forward operator in frequency domain to a bow-tie region \mathcal{R}_A . We will make use of this observation in Section III-D2 for Curvelet representation of the PAT data.

B. Curvelets

Curvelet transform is a multiscale pyramid with different directions and positions. Figure 4 exemplifies a typical 2D Curvelet tiling of the frequency domain underpinning the transform introduced in [30]. For each scale $j \geq j_0$ and angle θ_l , $\theta_l \in [-\pi, \pi)$, in spatial domain the Curvelet envelope $\tilde{U}_{j,\theta_l}(\mathbf{x})$ is aligned along a ridge of length $2^{-j/2}$ and width 2^{-j} . In the frequency domain, a Curvelet is supported near a trapezoidal wedge with the orientation θ_l . The Curvelet wedges become finer with increasing scales which makes them capable of resolving singularities along curves.

Here we introduce Curvelet transform straight away in \mathbb{R}^d following the continuous presentation of the Curvelet transform in [30]. Utilising Plancherel's theorem, Curvelet transform is computed in frequency domain. For each scale j , orientation $\theta_l = 2\pi/L \cdot 2^{-\lfloor j/2 \rfloor} \cdot l$, $l \in \mathbb{Z}^{d-1}$ such that $\theta_l \in [-\pi, \pi)^{d-1}$ ($L \cdot 2^{\lfloor j_0/2 \rfloor}$ is the number of angles at the second coarsest scale $j = j_0 + 1$, j_0 even), and translation $\mathbf{a} \in \mathbb{Z}^d$, the Curvelet coefficients of $u : \mathbb{R}^d \rightarrow \mathbb{R}$, are computed as a projection of its Fourier transform $\hat{u}(\mathbf{k})$ on the corresponding basis functions $\tilde{U}_{j,\theta_l}(\mathbf{k}) \exp(-i\mathbf{k} \cdot \mathbf{x}_\mathbf{a}^{(j,l)})$ with $\tilde{U}_{j,\theta_l}(\mathbf{k})$ a trapezoidal frequency window with scale j and orientation θ_l , and spatial domain centre at $\mathbf{x}_\mathbf{a}^{(j,l)}$ corresponding to the sequence of translation parameters $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{Z}^d$: $\mathbf{a} = (a_1 \cdot 2^{-j}, a_2 \cdot 2^{-j/2}, \dots, a_d \cdot 2^{-j/2})$

$$C_{j,l}(\mathbf{a}) = \int_{\mathbb{R}^d} \hat{u}(\mathbf{k}) \tilde{U}_{j,\theta_l}(\mathbf{k}) e^{i\mathbf{x}_\mathbf{a}^{(j,l)} \cdot \mathbf{k}} d\mathbf{k}. \quad (\text{C})$$

We would like to highlight the effect of the parabolic scaling on the translation grid which is reflected in the first dimension being scaled differently $a_1 \cdot 2^{-j}$ than the remaining dimensions $a_i \cdot 2^{-j/2}$, $i = 2, \dots, d$ (the spatial translation grid is denser along the dimension 1) which is of significance for the wedge restricted Curvelet transform proposed in the next section. As we are only interested in representing real functions u , we use the symmetrised version of the Curvelet transform which forces the symmetry (w.r.t. origin) of the corresponding frequency domain wedges θ_l and $\theta_l + \pi$. The coarse scale Curvelets are isotropic Wavelets corresponding to translations of an isotropic low pass window $\tilde{U}_0(\mathbf{k})$. The window functions at all scales and orientations form a smooth partition of unity on the relevant part of \mathbb{R}^d .

There are different ways to numerically efficiently evaluate the integral. In this paper, we use the digital coronisation with shear rather than rotation and its implementation via wrapping introduced in [30]. The resulting discrete transform $\Psi : \mathbb{R}^n \rightarrow \mathbb{C}^N$ (or \mathbb{R}^N when symmetric) has a finite number of discrete angles θ_ℓ per scale (refined every second scale according to

parabolic scaling) and a fixed (per quadrant and scale) finite grid of translations $\mathbf{x}_\mathbf{a}^{(j)} = (a_1 \cdot 2^{-j}, a_2 \cdot 2^{-j/2}, \dots, a_d \cdot 2^{-j/2})$ as opposed to the continuous definition where the grids are rotated to align with the Curvelet angle. This fast digital Curvelet transform is a numerical isometry. For more details we refer to [30] and the 2D and 3D implementations in the Curvelab package⁵.

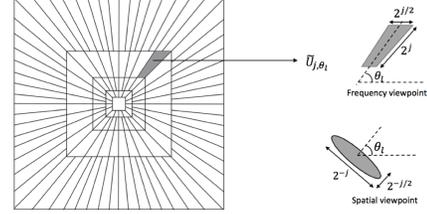


Fig. 4. An example of 2D frequency domain Curvelet tiling using 5 scales. The highlighted wedge, magnified view in top right, corresponds to the frequency window near which the Curvelet \tilde{U}_{j,θ_l} at the scale j with orientation θ_l is supported. The orientation of the envelope of the corresponding Curvelet function in the spatial domain is shown in bottom right corner.

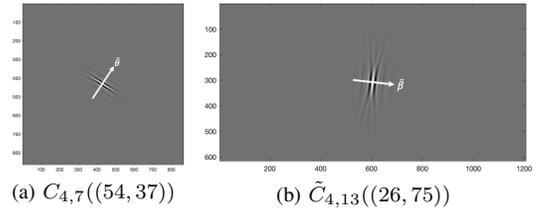


Fig. 5. An example of a single Curvelet (a) in image domain, $C_{4,7}((54, 37))$, (b) and a corresponding one in the data domain, $\tilde{C}_{4,13}((26, 75))$.

C. Wedge restricted Curvelets

We now introduce a modification of the Curvelet transform to a subset of orientations in an abstract setting first, before we tailor it to represent the PAT data in the next section. We recall that the Curvelet orientation is parametrised by a vector of angles $\theta \in [-\pi, \pi)^{d-1}$. Let \mathbf{k}_θ denote the Curvelet wavefront vector and $\hat{\mathbf{k}}^i$ the vector along the i th coordinate direction. Then the components of θ are angles θ^i , $i = 1, \dots, d-1$ between $\hat{\mathbf{k}}^1$ and \mathbf{k}_θ measured in planes spanned by $\hat{\mathbf{k}}^1$ and each of the remaining coordinate vectors $\hat{\mathbf{k}}^{i+1}$, $i = 1, \dots, d-1$, respectively. $\theta = \mathbf{0}$ corresponds to Curvelets with orientation $\mathbf{k}_\theta = \hat{\mathbf{k}}^1$. In 2D $\theta \in [-\pi, \pi)$ is just a single angle $\angle(\hat{\mathbf{k}}^1, \mathbf{k}_\theta)$ and in 3D we have two angles.

With this notation the *wedge restricted Curvelet transform* amounts to computing the Curvelet coefficients i.e. evaluating (C) only for a wedge-shaped subset of angles $\theta_l \in [\theta_b, \theta^t] \subset [-\pi, \pi)^{d-1}$. We furthermore define a symmetrised *wedge restricted Curvelet transform* by restricting the Curvelet orientations to the symmetric double wedge (bow-tie) $\theta_l \in [-\theta^w, \theta^w] \cup [-\theta^w + \pi, \theta^w + \pi]$ with $\theta^w \in (0, \pi/2)^{d-1}$. An example of a symmetric wedge restricted transform is illustrated in Figure 6 for 2D (a) and 3D (c), where the grey region corresponds to the Curvelet orientations inside the symmetric double wedge and the white to orientations outside it which are excluded from the computation.

⁵<http://www.curvelet.org/software.html>

Formally, we can write the wedge restricted Curvelet transform $\tilde{\Psi} : \mathbb{R}^n \rightarrow \mathbb{R}^{\tilde{N}}$, $\tilde{N} < N$ as a composition of the standard Curvelet transform $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^N$ and a projection operator $P_W : \mathbb{R}^N \rightarrow \mathbb{R}^{\tilde{N}}$, $\tilde{\Psi} = P_W \Psi$, where P_W projects on the directions inside the bow-tie shaped set $W = [-\theta^w, \theta^w] \cup [-\theta^w + \pi, \theta^w + \pi]$ with $\theta^w \in (0, \pi/2)^{d-1}$. As the coarse scale, $j = j_0$, Curvelets are isotropic Wavelets, they are non-directional and thus are mapped to themselves. Here, we focused on the symmetric transform but definitions for complex Curvelets follow analogously.

Bearing in mind the definition of the (digital) Curvelet transform where the angles $\theta_{j,l}$ are quantised to a number of discrete orientations (which essentially correspond to the centres of the wedges), the projector P_W maps all Curvelets $C_{j,l}$ with orientations $\theta_{j,l} \in W$ to themselves and all $C_{j,l}$ with orientations $\theta_{j,l} \notin W$ to $\mathbf{0} \in \mathbb{R}^{\tilde{N}}$

$$P_W : \mathbb{R}^N \rightarrow \mathbb{R}^{\tilde{N}} \quad (8)$$

$$P_W(C_{j,l}(\mathbf{a})) = \begin{cases} C_{j,l}(\mathbf{a}), & \theta_{j,l} \in W, j > j_0 \\ C_{j,l}(\mathbf{a}), & j = j_0 \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

We note that the underlying full Curvelet transform is an isometry i.e. its left inverse $\Psi^\dagger = \Psi^T$. The same holds for the wedge restricted transform when defined on the restriction to the range of P_W i.e. $\tilde{\Psi} : \mathbb{R}^n|_{\text{range}(P_W)} \rightarrow \mathbb{R}^{\tilde{N}}$ (assuming the resulting $\tilde{\Psi}$ is over-determined) as, by definition, on its range P_W is an identity operator. The range of P_W corresponds to the set $W \cup \tilde{U}_0$ in the Fourier domain, union of the wedge W and \tilde{U}_0 the low frequency window at coarse scale j_0 . Intuitively, the restriction to range of P_W , $\mathbb{R}^n|_{\text{range}(P_W)}$, corresponds to images which only contain high frequency features with wave front vectors in W .

A simple implementation obtained via the P_W projector which does not require changes to Curvelab can be downloaded from Github⁶. A more efficient implementation would avoid the computation of the out-of-wedge Curvelets all together but it would require making changes to Curvelab code.

D. Curvelets and PAT

1) *Curvelet representation of initial pressure*: Curvelets were shown to provide an almost optimally sparse representation if the object is smooth away from (piecewise) \mathcal{C}^2 singularities [5], [31], [32]. More precisely, the error of the best s -term approximation (corresponding to taking the s largest in magnitude coefficients) in Curvelet frame of an image g decays as [5]

$$\|g - g_s\|_2^2 \leq \mathcal{O}((\log s)^3 \cdot s^{-2}). \quad (9)$$

Motivated by this result, we investigate the Curvelet representation of the initial pressure p_0 . Here we can directly apply the symmetrised Curvelet transform (C) to un-mirrored p_0 which is a real function $\mathbb{R}^d \rightarrow \mathbb{R}$. In this case the frequency \mathbf{k} in (C) corresponds directly to the frequency \mathbf{k} in the ambient space. In practice, p_0 is discretised on an n point grid in \mathbb{R}^d , $p_0 \in \mathbb{R}^n$, and we apply the symmetric discrete Curvelet transform $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^N$.

2) *Wedge restricted Curvelet representation of PAT data*: The acoustic propagation is nearly optimally sparse in the Curvelet frame and so is well approximated by simply translating the Curvelet along the Hamiltonian flow [6]. This implies the sparsity of the initial pressure in the Curvelet frame is essentially preserved by acoustic propagation. In Section III-A we derived the one to one mapping between the wavefront direction in ambient space and the corresponding direction in data domain (6). We can use this relation to essentially identify a Curvelet in the ambient space (and hence image space) with the induced Curvelet in the data space, which in turn implies that the sparsity of the initial pressure is translated into the sparsity of the PAT data. Motivated by this result we consider sparse representation of the full volume of photoacoustic data in the Curvelet frame.

The bow-tie shaped range of the PAT forward operator together with Hermitian symmetry of real functions in Fourier domain suggest the *symmetrised wedge restricted Curvelet transform* introduced in III-C for PAT data volume representation.

For the PAT data representation, we reinterpret the notation in (C) underlying the symmetrised wedge restricted Curvelet transform as follows

- \mathbf{k} as a data domain frequency $(\omega/c, \mathbf{k}_S)$;
- $\theta_{j,l}$ as a discrete direction $\beta_{j,l}$ of the data domain trapezoidal frequency window $\tilde{U}_{j,\beta_{j,l}}(\omega/c, \mathbf{k}_S)$. We note that we do not use the wavefront angle map (7) to compute these angles explicitly, only to argue the sparsity of the data. The tiling is computed using standard Curvelet transform applied to rectangular domain, followed by the projection on the range P_W (see Sections III-D3, III-D4). The corresponding angles $\beta_{j,l}$ are obtained via (11) for each quadrant;
- W the restriction of $\beta_{j,l}$ to those in the bow-tie $W = (-\pi/4, \pi/4)^{d-1} \cup (3/4\pi, 5/4\pi)^{d-1}$ effects the projection on the range of PAT forward operator \mathcal{R}_A (to be precise on a set containing the range due to isotropic coarse scale functions \tilde{U}_0 but we use this term loosely throughout the paper);
- $(\mathbf{a}_t, \mathbf{a}_S)$ the grid of translations along the normalised time axis ct and the detector \mathcal{S} , respectively. Because we are using the wrapping implementation of fast digital Curvelet transform, this grid is the same for all angles within the same quadrant \mathbb{T} at the same scale;

This results in the following formula for wedge restricted Curvelet coefficients

$$\tilde{C}_{j,l}(\mathbf{a}_t, \mathbf{a}_S) \quad \beta_{j,l} \in W \quad (\text{WRC})$$

$$= \iint_{\mathbb{R}^d} \hat{u}\left(\frac{\omega}{c}, \mathbf{k}_S\right) \tilde{U}_{j,\beta_{j,l}}\left(\frac{\omega}{c}, \mathbf{k}_S\right) e^{i(\mathbf{x}_a \cdot (\omega/c, \mathbf{k}_S))} d\frac{\omega}{c} d\mathbf{k}_S$$

with $\mathbf{x}_a = (2^{-j} \mathbf{a}_t, 2^{-j/2} \mathbf{a}_S)$.

We will demonstrate in the results section that the above restriction to the range of PAT data representation is absolutely essential to effectively eliminate artefacts associated with subsampling of the data, which align along the out of range directions.

⁶https://github.com/BolinPan/Wedge_Restricted_Curvelet

3) *The effect of time oversampling*: Any practical realisation of (wedge restricted) Curvelet transform for PAT data representation needs to account for PAT data being commonly oversampled in time w.r.t. to the spatial resolution resulting in data volumes elongated in the time dimension. As the Curvelet transform is agnostic to the underlying voxel size, a.k.a. grid spacing, to apply the wedge restriction correctly we need to calculate the speed of sound per voxel, c_v . To this end we first estimate \mathbf{x}_{\max} , the longest path between any point in the domain $\Omega \in \mathbb{R}^d$ and the detector \mathcal{S} (the length of the diagonal for the cuboid Ω) and the corresponding longest travel time $T_{\max} = \mathbf{x}_{\max}/c$. We note, that when simulating data using k-Wave, T_{\max} coincides with the time over which the data is recorded. However, for a number of reasons, the real data is often measured over a shorter period of time, $T < T_{\max}$.

Assuming an idealised data domain $(0, T_{\max}) \times \mathcal{S}$, we can calculate the speed of sound per voxel⁷, c_v , as a ratio of the spatial grid points to temporal steps to travel \mathbf{x}_{\max}

$$c_v = \frac{\mathbf{x}_{\max}/h_{\mathbf{x}}}{T_{\max}/h_t} = \frac{ch_t}{h_{\mathbf{x}}}. \quad (10)$$

Here $h_{\mathbf{x}}$ and h_t are the sensor/spatial grid spacing (we assume the same grid spacing in each dimension in space and coinciding with the grid at the sensor), and the temporal step, respectively. Plausibly, c_v is independent of the number of time steps in the data i.e. it is the same for any measurement time T .

So far, we have been consistently using normalised time i.e. ct instead of t (frequency ω/c instead of ω) which corresponds to the range of PAT forward operator $\beta \in (-\pi/4, \pi/4)^{d-1} \cup (3\pi/4, 5\pi/4)^{d-1}$. To illustrate the effect of time oversampling we will now shortly resort to the non-normalised formulation. This is for illustrative purpose only and exclusively the normalised formulation is used in computations i.e. the angle β always refers to the normalised formulation.

In the non-normalised formulation the speed of sound per voxel c_v determines the slope of the asymptotes in the voxelized domain, $\pm c_v t$. Taking the Fourier transform in each dimension, we obtain the corresponding asymptotes in frequency domain as $\pm\omega/c_v$. Figure 6 illustrates the effect of the time oversampling i.e. $c_v < 1$ on the Curvelet discretized range of the PAT forward operator in two and three dimensions.

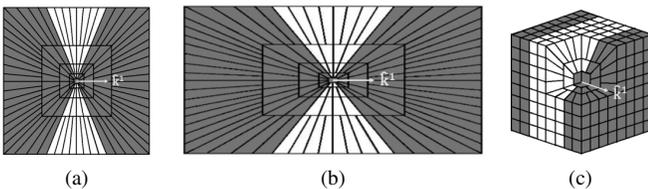


Fig. 6. An illustration of the bow-tie range of the PAT forward operator mapped over a 2D Curvelet tiling in (a) non-normalised data frequency space $(\omega, \mathbf{k}_{\mathcal{S}})$ (note asymptotes $\pm\omega/c_v$), (b) normalised data frequency space $(\omega/c_v, \mathbf{k}_{\mathcal{S}})$ restricted to the sensor frequency range (this form underlies the discrete Curvelet transform [33]). The Curvelets in the range are depicted in grey and these out of range in white. We stress that due to our definition of the angle β the corresponding wavefront vector in our coordinate system is $\hat{\beta} = (\cos\beta, -\sin\beta)$ which sweeps the grey region while the angle β sweeps $(-\pi/4, \pi/4) \cup (3/4\pi, 5/4\pi)$; (c) 3D equivalent of (a). Visualisation corresponds to $c_v = 0.5$.

4) *Discretisation of the orientation $\beta_{j,l}$* : Figure 6(b) depicts the tiling underpinning the discrete Curvelet transform on a rectangular domain [30]. As this tiling is based on the normalised data frequency space (w.r.t. the sound speed per voxel c_v), the slopes of the asymptotes remain ± 1 and the range of the PAT forward operator remains unaltered in this representation, $\beta \in (-\pi/4, \pi/4)^{d-1} \cup (3/4\pi, 5/4\pi)^{d-1}$.

Taking standard Curvelet transform as defined in [30] of time-oversampled discretised data volume corresponds to discretising the normalised data frequency domain $(\omega/c_v, \mathbf{k}_{\mathcal{S}})$ with maximal frequency $|\omega^{\max}/c_v| = 1/(Tc_v)$ along time and $|\mathbf{k}_{\mathcal{S}}^{\max,i}|$, $i = 1, \dots, d-1$ along the detector coordinates. In effect, the highest represented frequency in the discretised data volume along the time axis $|\omega^{\max}/c_v| = 1/(Tc_v)$ is $1/c_v$ times higher than the corresponding highest frequency along the sensor coordinates $|\mathbf{k}_{\mathcal{S}}^{\max,i}|$, $i = 1, \dots, d-1$. This affects the sampling of the discrete angles $\beta_{j,l}$ because it changes the ratio of the sides of the cuboid domain.

The wedge orientations $\beta_{j,l}$ in the discrete Curvelet transform [30] are obtained via equisampling the tangents of their corresponding angles $\hat{\beta}_{j,l}^{\mathbb{T}}$. In 2D for each of the four quadrants (in 3D equivalently 6 pyramids [33]): north, west, south and east, defined by the corners and the middle point of the rectangular domain, for each wedge angle $\beta_{j,l}$, $\hat{\beta}_{j,l}^{\mathbb{T}}$ denotes its version measured from the central axis of the corresponding quadrant $\mathbb{T} = \{\mathbb{N}, \mathbb{W}, \mathbb{S}, \mathbb{E}\}$. Consequently, $\tan \hat{\beta}_{j,l}^{\mathbb{N},\mathbb{S}}$ in the north/south quadrant are scaled with $1/c_v$ while $\tan \hat{\beta}_{j,l}^{\mathbb{W},\mathbb{E}}$ in the west/east triangle are scaled with c_v

$$\hat{\beta}_{j,l}^{\mathbb{N},\mathbb{S}} = \arctan \frac{1}{c_v} \frac{l-1/2}{L_j/8}, \quad l = -L_j/8 + 1, \dots, L_j/8, \quad (11)$$

$$\hat{\beta}_{j,l}^{\mathbb{W},\mathbb{E}} = \arctan c_v \frac{l-1/2}{L_j/8}, \quad l = -L_j/8 + 1, \dots, L_j/8,$$

where $L_j = L \cdot 2^{\lfloor j/2 \rfloor}$ is the total number of angles at the scale j which doubles every second scale. Note that the slope of the boundary between two neighboring quadrants is not ± 1 but $\tan 1/c_v$ or $\tan c_v$, respectively. The side effect of this scaling is that the sampling of the angles in the range is more dense than the angles out of range (one in four angles $\beta_{j,l}$ are out of the range at each scale in Figure 6), which plays in favour of our method.

IV. OPTIMIZATION

Let $\Phi \in \mathbb{R}^{m \times n}$ be the sensing matrix the rows of which are the respective sensing vectors ϕ_i^T , $i = 1, \dots, m$ with $m \ll n$ and $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^N$ be the isometric sparsifying transform with $\Psi^\dagger = \Psi^T$ its left inverse. We model the compressed measurements $b \in \mathbb{R}^m$ as

$$b = \Phi g + e = \Phi \Psi^\dagger f + e, \quad (12)$$

where e is the measurement noise (with $\|e\|_2 \leq \varepsilon$ for some small ε), and f is the sparse representation of the original signal $g \in \mathbb{R}^n$ in the basis Ψ , i.e. $f = \Psi g$. Under conditions specified in [3], [34], the sparse signal f can be recovered from the compressed measurements b via the basis pursuit (ℓ_1 minimization problem)

$$\min_{f \in \mathbb{R}^N} \|f\|_1 \quad \text{s.t.} \quad \|\Phi \Psi^\dagger f - b\|_2 \leq \varepsilon. \quad (13)$$

⁷ c_v is usually close to 0.3 in experiments in soft tissue

Furthermore, an (approximately) s -sparse signal f can be robustly recovered with overwhelming probability if $m \geq C \cdot \log(n) \cdot s$ for some positive constant C [2], [3].

A. Iteratively reweighted ℓ_1 norm

Efficient representation in an overdetermined basis requires highly sparse solutions. In contrast to ℓ_0 semi-norm where each non-zero coefficient is counted regardless of its magnitude, the ℓ_1 penalty is proportional to the coefficient's magnitude resulting in small coefficients being penalised less than the large coefficients and consequently in less sparse solutions. We address this issue using weighted formulation of ℓ_1 norm penalty, $\|\Lambda f\|_1$, where Λ is a diagonal matrix with positive weights and f is the sparse representation of the solution in a basis Ψ . Reweighted ℓ_1 penalty was shown in many situations to outperform the standard ℓ_1 norm in the sense that substantially fewer measurements are needed for exact recovery [35], [36], [37], [38].

To enhance the sparsity of the solution, the weights can be chosen as inversely proportional to the magnitudes of the components of the true solution. As the latter is not available, the current approximation f^k is used in lieu [35]

$$\Lambda^k = \text{diag}(1/(|f^k| + \epsilon)). \quad (14)$$

Initialisation choosing $\Lambda^1 = \text{diag}(\mathbf{1})$ i.e. with the non-weighted ℓ_1 seems to work well in practice. We follow an adaptive scheme for the choice of ϵ proposed in [35] (another approach is suggested in [36]). In each iteration k , we first normalize the magnitude of f^k , denoted \bar{f}^k . Let $\rho(\bar{f}^k)$ be the non-increasing arrangement of \bar{f}^k , and $\rho(\bar{f}^k)_S$ be the S -th element after reordering. Then we update ϵ as

$$\epsilon = \max\{\rho(\bar{f}^k)_S, 10^{-4}\}. \quad (15)$$

Such choice of ϵ is geared towards promoting S -sparse solutions. As Curvelet frame is heavily overdetermined, we propose to use reweighted ℓ_1 for both data reconstruction and p_0 reconstruction. Recall the s -sparse signals can be recovered exactly if $m \geq C \cdot \log(n) \cdot s$, and so the sparsity S is chosen to be as $m/(C \cdot \log(n))$ for some positive constant C . A version of ℓ_1 reweighting used in our later algorithms is summarized in Algorithm **W**.

W $\Lambda \leftarrow \text{weight}(f, S)$ [35]

- 1: **Input:** f, S
 - 2: **Output:** Λ
 - 3: $\bar{f} = |f| / \max(|f|)$
 - 4: $\epsilon = \max(\rho(\bar{f})_S, 10^{-4})$
 - 5: $\Lambda = \text{diag}(1/(|f| + \epsilon))$
-

B. Data reconstruction

In this two step approach we propose to first reconstruct the full volume photoacoustic data from subsampled data as described below, followed by the recovery of initial pressure (photoacoustic image) using time reversal via a first order method implemented in the k-Wave Toolbox [26].

We formulate the data reconstruction problem as an iteratively reweighted ℓ_1 minimization problem

$$f^{k+1} \stackrel{k}{\leftarrow} \arg \min_f \frac{1}{2} \|\Phi \tilde{\Psi}^\dagger f - b\|_2^2 + \tau \|\Lambda^k f\|_1, \quad (\mathbf{DR})$$

where $\Phi \in \mathbb{R}^{m \times n_b}$ is the subsampling matrix, $b \in \mathbb{R}^{n_b}$ is the subsampled PAT data, $\tilde{\Psi}^\dagger : \mathbb{R}^{N_b} \rightarrow \mathbb{R}^{n_b}$ is the left inverse (and adjoint) of the wedge restricted Curvelet transform of the PAT data $\tilde{\Psi}$ (see III-C, III-D2, III-D3, III-D4), f are the coefficients of the full PAT data in $\tilde{\Psi}$ and Λ^k is a diagonal matrix of weights updated at each iteration using procedure **W**.

R-SALSA Reweighted Split Augmented Lagrangian Shrinkage Algorithm [39]

- 1: **Initialization:** $y^1 = \mathbf{0}, w^1 = \mathbf{0}, f^1 = \mathbf{1}, \Lambda^1 = \text{diag}(\mathbf{1}), \mu > 0, \tau > 0, \eta > 0$ and K_{\max}
 - 2: $k := 1$
 - 3: **Repeat**
 - 4: $f^{k+1} = \arg \min_f \|\Phi \tilde{\Psi}^\dagger f - b\|_2^2 + \mu \|f - y^k - w^k\|_2^2$
 - 5: $y^{k+1} = \arg \min_y \tau \|\Lambda^k y\|_1 + \frac{\mu}{2} \|f^{k+1} - y - w^k\|_2^2$
 - 6: $w^{k+1} = w^k - (f^{k+1} - y^{k+1})$
 - 7: $\Lambda^{k+1} = \text{weight}(f^{k+1}, S)$
 - 8: $k = k + 1$
 - 9: **Until** $\|f^k - f^{k-1}\|_2 / \|f^{k-1}\|_2 < \eta$ or $k > K_{\max}$
-

At each iteration of the iteratively reweighted ℓ_1 scheme for **DR**, the minimisation problem for a fixed Λ^k is solved with an ADMM type scheme, SALSA [39], resulting in a reweighted version of SALSA summarized in Algorithm **R-SALSA**. The sub-problem in **R-SALSA** line 4 is a quadratic minimization problem; its solution can be obtained explicitly via the normal equations

$$f^{k+1} = (\tilde{\Psi} \Phi^T \Phi \tilde{\Psi}^\dagger + \mu I)^{-1} (\tilde{\Psi} \Phi^T b + \mu(y^k + w^k)) \quad (16)$$

and using the Sherman–Morrison–Woodbury formula, the inverse in (16) can be reformulated as

$$(\tilde{\Psi} \Phi^T \Phi \tilde{\Psi}^\dagger + \mu I)^{-1} = \frac{1}{\mu} \left(I - \frac{1}{\mu + 1} \tilde{\Psi} \Phi^T \Phi \tilde{\Psi}^\dagger \right). \quad (17)$$

The solution of the proximal w.r.t. $\frac{\tau}{\mu} \Lambda^k \|\cdot\|_1$ sub-problem in line 5 of **R-SALSA** amounts to applying to $\tilde{y}^k = f^{k+1} - w^k$ the following soft thresholding operator with vector valued threshold $S_{\frac{\tau}{\mu} \Lambda^k}$

$$S_{\frac{\tau}{\mu} \Lambda^k}(\tilde{y}^k) = \text{sign}(\tilde{y}^k) \odot \left(|\tilde{y}^k| - \frac{\tau}{\mu} \Lambda^k \mathbf{1} \right), \quad (18)$$

where $|\cdot|$ is taken pointwise and \odot denotes pointwise multiplication.

C. p_0 reconstruction

In this one step approach we reconstruct the initial pressure p_0 (photoacoustic image) directly from subsampled photoacoustic data. We formulate the p_0 reconstruction as an iteratively reweighted ℓ_1 minimization problem

$$f^{k+1} \stackrel{k}{\leftarrow} \arg \min_f \frac{1}{2} \|\Phi A \Psi^\dagger f - b\|_2^2 + \tau \|\Lambda^k f\|_1, \quad (p_0 \mathbf{R})$$

where $\Phi \in \mathbb{R}^{m \times n_b}$ and $b \in \mathbb{R}^{n_b}$ are, as before, the subsampling matrix and the corresponding subsampled PAT data, $A : \mathbb{R}^{n_{p_0}} \rightarrow \mathbb{R}^{n_b}$ is the photoacoustic forward operator, $\Psi^\dagger : \mathbb{R}^{N_{p_0}} \rightarrow \mathbb{R}^{n_{p_0}}$ is the left inverse (and adjoint) of the standard Curvelet transform Ψ (see III-B, III-D1), f are the coefficients of p_0 in Ψ and Λ^k is a diagonal matrix of weights updated at each iteration using procedure **W**.

R-FISTA Reweighted Fast Iterative Shrinkage Thresholding Algorithm [40]

- 1: **Initialization:** $y^1 = \mathbf{0}$, $f^1 = \mathbf{1}$, $\Lambda^1 = \text{diag}(\mathbf{1})$, $\alpha^1 = 1$, $\mu = \frac{1}{L}$, $\tau > 0$, $\eta > 0$ and K_{\max}
 - 2: $k := 1$
 - 3: **Repeat**
 - 4: $\tilde{z}^k = y^k - \mu \Psi A^T \Phi^T (\Phi A \Psi^\dagger y^k - b)$
 - 5: $f^{k+1} = \arg \min_f \mu \tau \|\Lambda^k f\|_1 + \frac{1}{2} \|f - \tilde{z}^k\|_2^2$
 - 6: $\alpha^{k+1} = (1 + \sqrt{1 + 4(\alpha^k)^2})/2$
 - 7: $y^{k+1} = f^{k+1} + \frac{\alpha^k - 1}{\alpha^{k+1}} (f^{k+1} - f^k)$
 - 8: $\Lambda^{k+1} = \text{weight}(f^{k+1}, S)$
 - 9: $k = k + 1$
 - 10: **Until** $\|f^k - f^{k-1}\|_2 / \|f^{k-1}\|_2 < \eta$ or $k > K_{\max}$
-

Due to a high cost of application of the PAT forward A and adjoint A^T operators, we propose to solve $p_0 \mathbf{R}$ with FISTA [40] which avoids inner iteration with A and A^T . At each iteration of FISTA we solve the proximal Λ^k weighted ℓ_1 problem and subsequently update the weights as detailed in Algorithm **R-FISTA** (lines 5, 8). In the **R-FISTA** algorithm, L is an approximation to $\|\Psi A^T \Phi^T \Phi A \Psi^\dagger\|_2$, which can be pre-computed for given A , Ψ and a subsampling scheme Φ , using a simple power iteration.

D. p_0 reconstruction with non-negativity constraint

Considering the non-negativity constraint as additional prior information on photoacoustic image, we formulate the non-negativity constraint version of p_0 reconstruction

$$f^{k+1} \stackrel{k}{\leftarrow} \arg \min_f \frac{1}{2} \|\Phi A \Psi^\dagger f - b\|_2^2 + \tau \|\Lambda^k f\|_1, \quad (p_0 \mathbf{R}+)$$

s.t. $\Psi^\dagger f \geq 0$.

To incorporate the non-negativity constraint into the objective function we rewrite $p_0 \mathbf{R}+$ with $p = \Psi^\dagger f$

$$p^{k+1} \stackrel{k}{\leftarrow} \arg \min_{p \geq 0} \frac{1}{2} \|\Phi A p - b\|_2^2 + \tau \|\Lambda^k \Psi p\|_1$$

$$= \arg \min_p \frac{1}{2} \|\Phi A p - b\|_2^2 + \tau \|\Lambda^k \Psi p\|_1 + \chi_+(p), \quad (19)$$

where $\chi_+(p)$ is the characteristic function of non-negative reals applied pointwise. The non-negativity constraint p_0 reconstruction in a form (19) is amiable to solution with ADMM [41] inside an iteratively reweighted ℓ_1 minimization scheme. We first introduce an auxiliary variable

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \Psi \\ I_{n_{p_0}} \end{pmatrix} p = E p. \quad (20)$$

Then (19) is equivalent to

$$\arg \min_{p, y} G(p) + H(y) \quad \text{s.t. } y = E p, \quad (21)$$

where the functions G and H are

$$G(p) = \frac{1}{2} \|\Phi A p - b\|_2^2, \quad (22)$$

$$H(y) = \tau \|\Lambda^k y_1\|_1 + \chi_+(y_2).$$

The corresponding dual variable $w = (w_1, w_2)^T$ follows an analogous split. A reweighted version of ADMM used in this work is summarized in Algorithm **R-ADMM**.

Taking advantage of the isometry of the Curvelet transform Ψ , the update in line 4 of Algorithm **R-ADMM** can be reformulated as a solution of a large scaled least-squares problem which is solved using the *Conjugate Gradient Least Square (CGLS)* method [42]. The update in line 5 of Algorithm **R-ADMM** can be further split into two sub-problems, each of which admits an analytical solution

$$y_1^{k+1} = \arg \min_{y_1} \frac{1}{2} \|\Psi p^{k+1} - y_1^k + w_1^k\|_2^2 + \frac{\tau}{\mu} \|\Lambda^k y_1^k\|_1$$

$$= S_{\frac{\tau}{\mu} \Lambda^k} (\Psi p^{k+1} + w_1^k), \quad (23)$$

$$y_2^{k+1} = \arg \min_{y_2} \frac{\rho}{2} \|p^{k+1} - y_2^k + w_2^k\|_2^2 + \chi_+(y_2^k)$$

$$= \max(0, p^{k+1} + w_2^k).$$

R-ADMM Reweighted Alternating Direction Method of Multipliers Algorithm [41]

- 1: **Initialization:** $y^1 = \mathbf{0}$, $w^1 = \mathbf{0}$, $f^1 = \mathbf{1}$, $\Lambda^1 = \text{diag}(\mathbf{1})$, $\mu > 0$, $\tau > 0$, $\eta > 0$ and K_{\max}
 - 2: $k := 1$
 - 3: **Repeat**
 - 4: $p^{k+1} = \arg \min_p G(p) + \frac{\mu}{2} \|E p - y^k + w^k\|_2^2$
 - 5: $y^{k+1} = \arg \min_y H(y) + \frac{\mu}{2} \|E p^{k+1} - y + w^k\|_2^2$
 - 6: $w^{k+1} = w^k + E p^{k+1} - y^{k+1}$
 - 7: $f^{k+1} = \Psi p^{k+1}$
 - 8: $\Lambda^{k+1} = \text{weight}(f^{k+1}, S)$
 - 9: $k = k + 1$
 - 10: **Until** $\|f^k - f^{k-1}\|_2 / \|f^{k-1}\|_2 < \eta$ or $k > K_{\max}$
-

V. PHOTOACOUSTIC RECONSTRUCTION

We present our results of the data reconstruction **DR** and p_0 reconstructions $p_0 \mathbf{R}$, $p_0 \mathbf{R}+$, using 25% randomly subsampled measurements on a simulated data and a real 3D Fabry-Pérot raster scan data. The time reversal **TR** and the PAT forward **A** and adjoint **A*** operators are implemented with the k-Wave Toolbox [26]. If non-negativity of initial pressure is not explicitly enforced by the algorithm, negative values are clipped to 0 in a post-processing step.

A. 2D vessel phantom: rigorous comparison

In an attempt to rigorously compare the two approaches **DR** and $p_0 \mathbf{R}$ on a realistic scenario we created a 2D vessel phantom in Figure 7. This phantom is a 2D idealisation of a typical flat panel detector geometry setup where the depth of the image, n_{x_\perp} , is smaller comparing to the sensor size ($\sim n_S^{1/(d-1)}$), where n_S is the cardinality of the regular grid on the sensor), resulting in an image resolution of $n_{x_\perp} \times n_S = 42 \times 172$ with uniform voxel size $h_x = 11.628 \mu\text{m}$. We assume homogeneous speed of sound $c = 1500 \text{m/s}$ and medium density 1000kg/m^3 . The pressure time series recorded every $h_t = 29.07 \text{ns}$ by a line sensor on top of the object with image matching resolution, i.e. $h_x = 11.628 \mu\text{m}$, constitutes the PAT data which we endow with additive white noise with standard deviation $\sigma = 0.01$. We henceforth interpret this data as a time-steps \times sensors shaped volume.

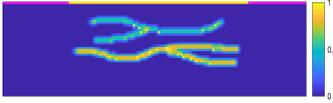


Fig. 7. Vessel phantom. Position of the line sensor is marked in purple and a window with higher probability of being sampled is highlighted in yellow.

1) *Fair comparison:* We note that the PAT data volume is always larger than the original initial pressure for two reasons: the measurement time T_{\max} needs to be long enough for signals to arrive from the farthest corner of the domain (using notation in Section III-D3, the diagonal of the cube Ω in voxels is $\lceil \mathbf{x}_{\max}/h_{\mathbf{x}} \rceil$) and it is a common practice to oversample the PAT data in time. Thus to compare **DR** and $p_0\mathbf{R}$, $p_0\mathbf{R}+$ approaches in a realistic scenario fairly we upscale the reconstructed initial pressure as follows. We recall the definition of c_v in Section III-D3, $c_v = (\mathbf{x}_{\max}/h_{\mathbf{x}})/(T_{\max}/h_t)$, and note that $1/c_v$ is essentially the time oversampling factor resulting in the number of time sampling points $n_t = \lceil T_{\max}/h_t \rceil = \lceil (1/c_v)\mathbf{x}_{\max}/h_{\mathbf{x}} \rceil$ and a measured PAT data volume size $n_t \times n_s$. We calculate a uniform upscaling factor α for the regular reconstruction grid to match the number of data points in the measured PAT data⁸, $\alpha^d n_{\mathbf{x}_{\perp}} n_s = n_t n_s$, yielding $\alpha = \sqrt[d]{\frac{n_t}{n_{\mathbf{x}_{\perp}}}}$. An alternative approach would be to downscale the data during its reconstruction by representing it in a low frequency Curvelet frame introduced in [12] which also allows the wedge restriction.

For the vessel phantom we choose $c_v = 0.3$ resulting in PAT volume size 591×172 and leading to $\alpha \approx 3.75$ and reconstruction grid 158×645 with spacing $h_{\mathbf{x}} \approx 3.1\mu\text{m}$. For each of these point matched grids we construct a Curvelet transform with the same number of scales and angles to as close as possible mimic the one to one wavefront mapping. For the reconstructed p_0 image of size 158×645 we use a Curvelet transform with 4 scales and 128 angles (at the 2nd coarsest level). For the data grid $n_t \times n_s = 591 \times 172$, we start with a standard Curvelet transform with 4 scales and 152 angles (at the 2nd coarsest level). This choice, after removing the out of range angles $\beta_l \notin (-\pi/4, \pi/4) \cup (3/4\pi, 5/4\pi)$, results in a wedge restricted Curvelet transform with 128 angles matching the number of angles of image domain Curvelets.

2) *Representation/compression in image and data domain:* Let us first demonstrate that exclusion of the out of range angles in wedge restricted Curvelet transform $\tilde{\Psi}$ indeed does not impact the accuracy beyond numerical/discretisation errors. Figure 8 shows the effect of wedge restricted transform $\tilde{\Psi}$ in data and image domain comparing (a,b) the simulated data $g(t, \mathbf{x}_S)$ and its time reversal versus (c,d) the data projected on the range of $\tilde{\Psi}$ by applying forward and inverse wedge restricted transform $\hat{g}(t, \mathbf{x}_S) = \tilde{\Psi}^\dagger \tilde{\Psi} g(t, \mathbf{x}_S)$ and its time reversal. The projection error is of the order of 4% in both data (e) and image (f) domain and concentrates at the boundary of the bow-tie shaped range of PAT forward operator. In fact, (e) suggests that this is the effect of the angle discretisation with the artefacts corresponding to the wavefronts in the range of the continuous PAT forward operator but with $|\beta|$ larger than the largest discrete $|\beta_l|$ (for $\beta \in (-\pi/4, \pi/4)$) and symmetric

⁸alternatively, one could match the cardinalities of Curvelet coefficients

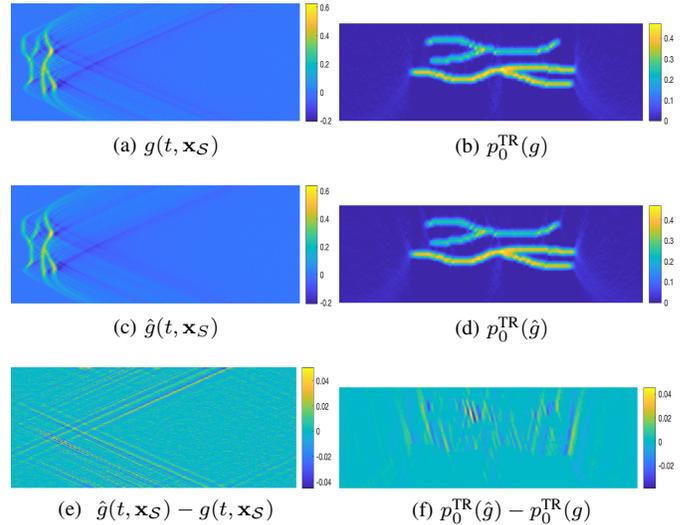


Fig. 8. Visualisation of accuracy of wedge restricted Curvelet transform $\tilde{\Psi}$ of vessel phantom data: (a) full simulated data $g(t, \mathbf{x}_S)$, (b) initial pressure via **TR** of $g(t, \mathbf{x}_S)$, (c) data projected on the range of $\tilde{\Psi}$, $\hat{g}(t, \mathbf{x}_S) = \tilde{\Psi}^\dagger \tilde{\Psi} g(t, \mathbf{x}_S)$, (d) initial pressure via **TR** of $\hat{g}(t, \mathbf{x}_S)$. (e) data error $\hat{g}(t, \mathbf{x}_S) - g(t, \mathbf{x}_S)$, (f) initial pressure error $p_0^{\text{TR}}(\hat{g}) - p_0^{\text{TR}}(g)$.

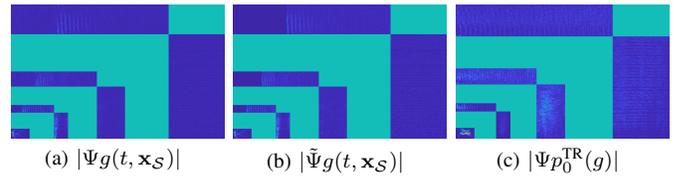


Fig. 9. Visualisation of image and data domain Curvelet coefficients (showing coarse scale overlaid over top-right quarter of higher scale scale-normalised coefficient magnitudes) of vessel phantom: (a) standard (4, 152)-Curvelet transform of $g(t, \mathbf{x}_S)$ (there are very small magnitude coefficients in the out of range wedge visible when zooming in), (b) wedge restricted (4, 128)-Curvelet transform $g(t, \mathbf{x}_S)$ (note that all coefficient out of range are 0), and (c) standard (4, 128)-Curvelet transform of $p_0^{\text{TR}}(g)$.

analogy for $\beta \in (3/4\pi/4, 5/4\pi)$). These wavefronts are mapped according to (7) to image domain resulting in artefacts visible in (f).

Finally, we compare the efficacy of compression in data and image domain. To this end in Figure 10(a) we plot the decay of the log amplitudes of $\Psi g(t, \mathbf{x}_S)$, $\tilde{\Psi} g(t, \mathbf{x}_S)$ and $\Psi p_0^{\text{TR}}(g)$ noting that the cardinalities of image and data domain Curvelet transforms (before restriction) are similar but not the same as a result of the upscaling of the reconstruction grid. We see that the magnitudes of the image domain Curvelet coefficients, $\Psi p_0^{\text{TR}}(g)$, decay faster than the data domain coefficients $\Psi g(t, \mathbf{x}_S)$ or $\tilde{\Psi} g(t, \mathbf{x}_S)$ hinting at a higher compression rate and a more economic representation by Curvelets in image domain. We attribute this mainly to data, in contrast to initial pressure, not being compactly supported and hence more Curvelets needed to represent the hyperbolic branches. A tailored discretisation (progressively coarsening as t increases) could somewhat counteract this issue but is out of scope of this paper.

3) *Subsampling:* For such a compact phantom, uniform random subsampling of the flat sensor is not optimal. Instead we choose to concentrate the sampled points above the vessels. Practically, this is accomplished by sampling the centre of the sensor (the yellow window in Figure 7) with probability 5 times higher, otherwise uniform, than the rest of the sensor.

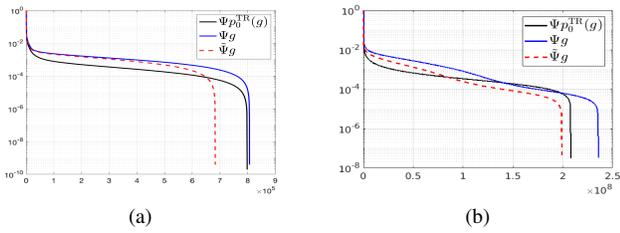


Fig. 10. The decay of log amplitudes of image and data Curvelet coefficients $\Psi p_0^{\text{TR}}(g)$, $\Psi g(\mathbf{x}_S, t)$ and $\tilde{\Psi} g(\mathbf{x}_S, t)$: (a) Vessel phantom, (b) Palm18.

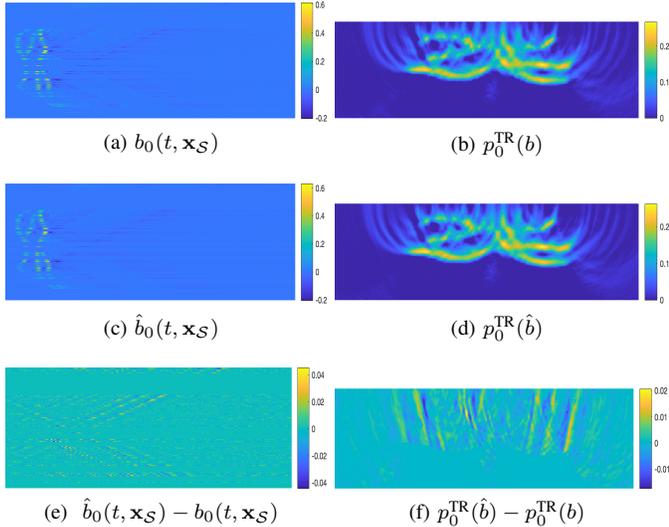


Fig. 11. Counterpart of Figure 8 for subsampled data: (a) 25% subsampled 0-filled data $b_0(t, \mathbf{x}_S)$, (b) initial pressure via **TR** of b , (c) data projected on the range of $\tilde{\Psi}$, $\tilde{b}_0(t, \mathbf{x}_S) = \tilde{\Psi}^T \tilde{\Psi} b_0(t, \mathbf{x}_S)$, (d) initial pressure via **TR** of $\tilde{b} = \Phi \tilde{b}_0(t, \mathbf{x}_S)$, (e) data error $\tilde{b}_0(t, \mathbf{x}_S) - b_0(t, \mathbf{x}_S)$, (f) initial pressure error $p_0^{\text{TR}}(\tilde{b}) - p_0^{\text{TR}}(b)$.

Figure 11 is the counterpart of Figure 8 for 25% subsampled data b visualised in Figure 11(a) with the missing data replaced by 0 (to which we henceforth refer as $b_0(t, \mathbf{x}_S)$, note that it holds $b = \Phi b_0$). The time reversed solutions⁹ in (b) and (d) illustrate the subsampling artefacts in the linear reconstruction which are visible in the error plot (f) along the “range discretisation” artefacts already present in Figure 8(f). Plotting the Curvelet coefficients of b_0 , Figure 12(a), reveals that the subsampling and 0-filling introduced out of range (non-physical) wavefronts perpendicular to the detector aligning with the jumps in $b_0(t, \mathbf{x}_S)$ due to the all-0 data rows. These unwanted coefficients would derail the sparsity enhanced data reconstruction but are effectively removed by the wedge restricted transform, Figure 12(b), as expected without any perceptible detrimental effect on the data, Figure 11(c,e).

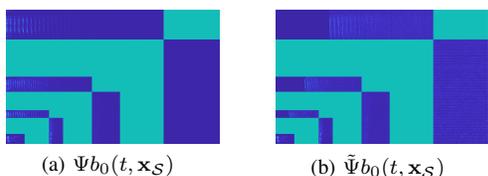


Fig. 12. Vessel phantom: Curvelet coefficients of subsampled, 0-filled data $b_0(t, \mathbf{x}_S)$ (see Figure 9 for explanation of coefficient visualisation): (a) standard (4,152)-Curvelets with clearly visible out of range (non-physical) coefficients, (b) wedge restricted (4,128)-Curvelets.

⁹We note that in **TR** we only impose the sampled sensor points i.e. b , not the 0-filled data b_0 , which leads to better results.

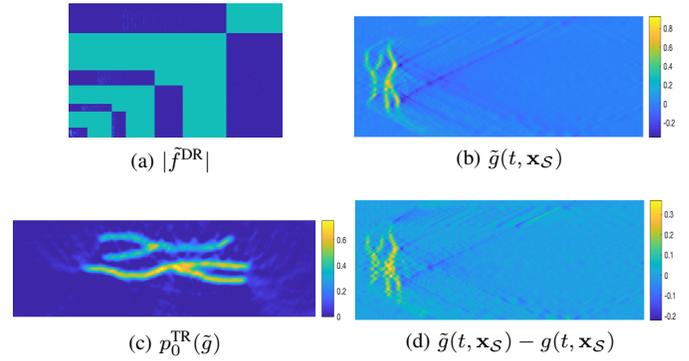


Fig. 13. Vessel phantom: two step method. (a) Recovered wedge restricted Curvelet coefficients \tilde{f}^{DR} (s.a. description in Figure 9), (b) the corresponding recovered full data $\tilde{g}(t, \mathbf{x}_S)$, (c) initial pressure obtained via **TR** of $\tilde{g}(t, \mathbf{x}_S)$, (d) data recovery error $\tilde{g}(t, \mathbf{x}_S) - g(t, \mathbf{x}_S)$.

4) *Reconstruction*: Finally, we compare performance of the two step and one step reconstruction methods on such 25% subsampled data. The intermediate step of the two step procedure, the PAT data recovered with **R-SALSA** ($\tau = 5 \cdot 10^{-5}$, $\mu = 1$, $C = 5$ and stopping tolerance $\eta = 5 \cdot 10^{-4}$ or after 100 iterations) is visualised in Figure 13: (a) shows the recovered Curvelet coefficients, (b) the corresponding PAT data $\tilde{g}(t, \mathbf{x}_S)$ and (d) the recovered data error $\tilde{g}(t, \mathbf{x}_S) - g(t, \mathbf{x}_S)$. The initial pressure $p_0^{\text{TR}}(\tilde{g})$ obtained from the recovered PAT data $\tilde{g}(t, \mathbf{x}_S)$ via time reversal is shown Figure 13(c). This can be compared against the one step procedure where the initial pressure p_0 is reconstructed directly from the subsampled data. Figure 14 illustrates the results of the one step method in different scenarios: (a,c) $p_0^{\text{R}}(b)$ along its Curvelet coefficients reconstructed with **R-FISTA** ($\tau = 10^{-3}$, $C = 5$ and stopping tolerance $\eta = 5 \cdot 10^{-4}$ or after 100 iterations); (b,d) $p_0^{\text{R}+}(b)$ along its Curvelet coefficients reconstructed with **R-ADMM** heeding the non-negativity constraint ($\tau = 10^{-4}$, $\mu = 0.1$, $C = 5$ and stopping tolerance $\eta = 5 \cdot 10^{-4}$ or after 100 iterations); (e,f) for comparison purposes we also show result of reconstructions with non-negativity constraint total variation **TV+**, obtained with an ADMM method described in [18], with two different values of the regularisation parameter 10^{-5} and $5 \cdot 10^{-4}$. The reconstructions obtained with two step method clearly exhibit more limited view artifacts. At least in part, we attribute this to the inefficiency of the data domain transform to represent exactly these peripheral wavefronts. The one step method produces overall superior results due to higher effectiveness of image domain Curvelet compression as well as iterated application of the forward and adjoint operators in the reconstruction process. The non-negativity constraint results in slightly smoother reconstructions. Despite having higher SNR, the **TV+** reconstructions struggle to find a sweet spot between too noisy and overly smooth images while enforcing Curvelet sparsity results in a more continuous appearance of the reconstructed vessels. Quantitative comparison of the one and two step approaches including extra one step reconstructions à la p_0^{R} with Haar and Daubechies-2 Wavelet bases is summarised in Table I. The reconstructed images with Haar and Daubechies-2 Wavelet bases can be found in supplementary material. Despite Wavelets resulting in higher SSIM, the appearance of, in particular, the top vessel is less continuous in comparison with Curvelet reconstruction. Both

Wavelet bases produce reconstructions with larger variation in contrast across the image hinting at isolated points with over-estimated pressure, which are not apparent for reconstructions with Curvelet basis (which is consistent with its highest PSNR and MSE). To evaluate robustness to noise of our one and two step methods we reconstructed the same subsampled data but with higher noise levels. The results are summarised in Table II while images are again deferred to supplementary material. The imaging metrics are split. Uniformly lower MSE and higher PSNR are consistent with smoother images produced by the two step approach **DR** while SSIM suggests that one step methods $p_0\mathbf{R}$, $p_0\mathbf{R}^+$ are better at preserving structure even for higher noise levels which is corroborated by visual inspection of the reconstructed images (with exception of $p_0\mathbf{R}^+$ for $\sigma = 0.08$).

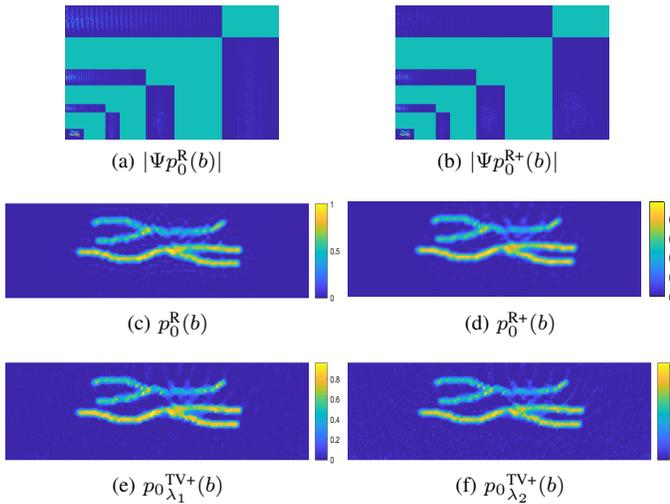


Fig. 14. Vessel phantom: one step methods $p_0\mathbf{R}$, $p_0\mathbf{R}^+$ (Curvelet sparsity), $p_0\mathbf{TV}^+$ (total variation). (a,c) $p_0^{\mathbf{R}}(b)$ along its Curvelet coefficients (s.a. description in Figure 9); (b,d) $p_0^{\mathbf{R}^+}(b)$ with non-negativity constraint along its Curvelet coefficients; (e,f) $p_0^{\mathbf{TV}^+}(b)$ with **TV** regularisation and non-negativity constraint for regularisation parameter values $\lambda_1 = 5 \cdot 10^{-4}$ and $\lambda_2 = 10^{-5}$.

B. 3D experimental data: palm18

Finally, we compare the two and one step methods using Curvelets on a 3D data set of palm vasculature (palm18) acquired with a 16 beam Fabry-Pérot scanner at 850nm excitation wavelength. The speed of sound and density are assumed homogeneous $c = 1570\text{m/s}$ and 1000kg/m^3 , respectively. The full scan consists of a raster of 144×133 of equispaced locations with $h_x = 106\mu\text{m}$, sampled every $h_t = 16.67\text{ns}$ resulting in $n_t = 390$ time points. The palm18 data can be thought of as a time \times sensor shaped volume with $390 \times 144 \times 133$ voxels with the sound speed per voxel $c_v = 0.2469$. Figure 15(a) shows the cross-sections of the palm18 data volume $g(\mathbf{x}_S, t)$ through time step $t = 85$ and scanning locations $y = 59, z = 59$.

Similarly as before, we upscale the reconstruction grid by the factor $\alpha = \sqrt[3]{n_t/n_{x_\perp}}$ to match the cardinality of the data as closely as possible. For the real data we calculate the image depth before upscaling as $n_{x_\perp} = n_t \cdot h_t \cdot c/h_x$ resulting in $\alpha = \sqrt[3]{h_x/(c \cdot h_t)} \approx 1.5941$ and the reconstruction grid $153 \times 229 \times 212$ with uniform spatial resolution of $66.5\mu\text{m}$. Figure 16(a) shows the the time reversal of the complete data

on this grid which we use as a reference solution for the subsampled data reconstructions.

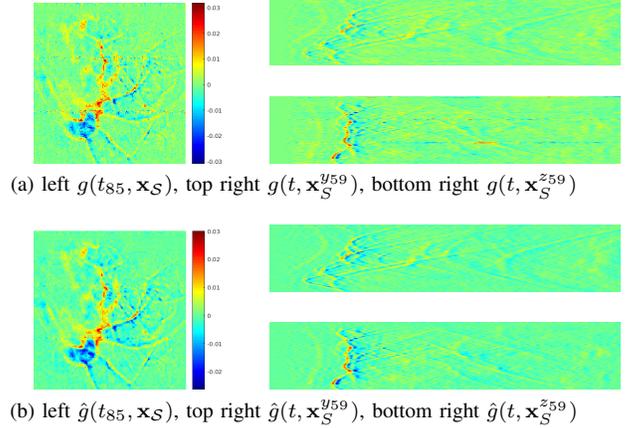


Fig. 15. 3D real data - palm18: cross-sections corresponding to time step $t = 85$, and scanning locations $y = 59, z = 59$ of (a) full data $g(t, \mathbf{x}_S)$, (b) full data projected on the range of $\tilde{\Psi}$, $\hat{g}(t, \mathbf{x}_S) = \tilde{\Psi}^\dagger \tilde{\Psi} g(t, \mathbf{x}_S)$.

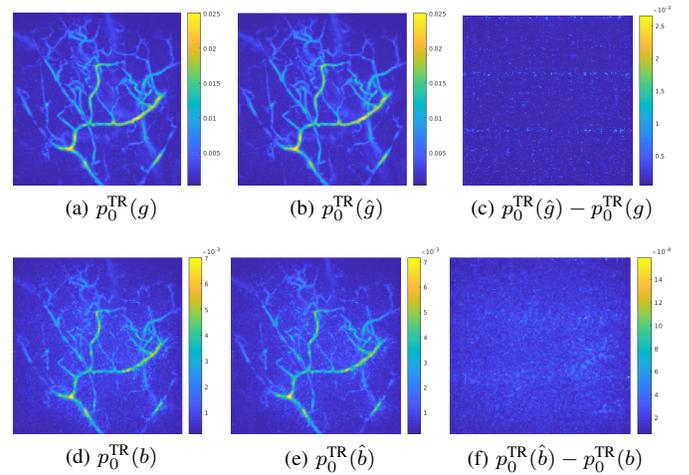


Fig. 16. 3D real data - palm18: maximum intensity projection of initial pressure obtained via time reversal of (a) the full data $g(t, \mathbf{x}_S)$, (b) the full data projected on the range of $\tilde{\Psi}$, $\hat{g}(t, \mathbf{x}_S) = \tilde{\Psi}^\dagger \tilde{\Psi} g(t, \mathbf{x}_S)$ and (c) their difference $p_0^{\text{TR}}(\hat{g}) - p_0^{\text{TR}}(g)$; (d) the subsampled data b , (e) the subsampled data projected on the range of $\tilde{\Psi}$, $\hat{b} = \Phi \tilde{\Psi}^\dagger \tilde{\Psi} b_0(t, \mathbf{x}_S)$ and (f) their difference $p_0^{\text{TR}}(\hat{b}) - p_0^{\text{TR}}(b)$.

We construct the image Curvelets Ψ and the data wedge restricted Curvelets $\tilde{\Psi}$ (after restriction) to both have 4 scales and 128 angles (at the 2nd coarsest level in each plane) and conduct analogous experiments as in 2D. Figure 15(b) shows the data projected on the range of $\tilde{\Psi}$, $\hat{g}(t, \mathbf{x}_S) = \tilde{\Psi}^\dagger \tilde{\Psi} g(t, \mathbf{x}_S)$. We note that the projection effectively removes the faulty sensors appearing as horizontal lines in the bottom plot on the right.

Figure 16(b) shows $p_0^{\text{TR}}(\hat{g})$ obtained via the time reversal of the projected data $\hat{g}(t, \mathbf{x}_S)$ and (c) the difference in image domain $p_0^{\text{TR}}(\hat{g}) - p_0^{\text{TR}}(g)$. The log magnitude plot of Curvelet coefficients in Figure 10(b) reveals that while the magnitudes of the image domain Curvelet coefficients $\Psi p_0^{\text{TR}}(g)$ decay the fastest, the magnitudes of wedge restricted Curvelet coefficients $\tilde{\Psi} g(t, \mathbf{x}_S)$ still decay at a faster rate than $\Psi g(t, \mathbf{x}_S)$, which was not the case in 2D.

Next, we uniformly randomly subsample the data down to 25% resulting in, after 0-filling, $b_0(t, \mathbf{x}_S)$ which cross-sections

TABLE I

VESSEL PHANTOM: IMAGING METRICS FOR p_0 RECONSTRUCTED USING DIFFERENT METHODS. THE METRICS ARE COMPUTED USING THE PHANTOM IMAGE UPSCALED TO THE RECONSTRUCTION GRID VIA BILINEAR INTERPOLATION AS GROUND TRUTH.

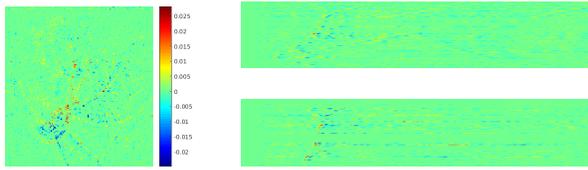
	Time reversal		DR	Variational reconstruction à la $p_0\mathbf{R}$ with different bases: $p_0^{\text{crv}}(b)$, $p_0^{\text{crv}+}(b)$ correspond to $p_0\mathbf{R}$, $p_0\mathbf{R}+$.							
	$p_0^{\text{TR}}(g)$	$p_0^{\text{TR}}(b)$	$p_0^{\text{TR}}(\tilde{g})$	$p_0^{\text{TV}+}_{\lambda_1=10^{-5}}(b)$	$p_0^{\text{TV}+}_{\lambda_2=5 \cdot 10^{-4}}(b)$	$p_0^{\text{crv}}(b)$	$p_0^{\text{crv}+}(b)$	$p_0^{\text{db}2}(b)$	$p_0^{\text{db}2+}(b)$	$p_0^{\text{haar}}(b)$	$p_0^{\text{haar}+}(b)$
MSE	0.0153	0.0261	0.0107	0.004	0.0035	0.0034	0.0032	0.0042	0.0037	0.0058	0.0041
SSIM	0.6683	0.5532	0.6207	0.5329	0.7591	0.8079	0.791	0.8238	0.7312	0.8288	0.6862
PSNR	18.1469	16.0234	18.7033	24.0295	24.5818	24.638	25.0053	23.7574	24.3298	22.3984	23.9174
SNR	18.8274	18.8169	22.8668	25.3583	25.2431	25.1421	25.2718	25.0987	25.2866	25.0487	25.3295

TABLE II

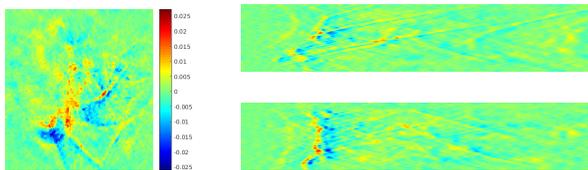
VESSEL PHANTOM: IMAGING METRICS FOR p_0 RECONSTRUCTED WITH \mathbf{DR} , $p_0\mathbf{R}$, $p_0\mathbf{R}+$ FOR DIFFERENT NOISE LEVELS. THE METRICS ARE COMPUTED USING THE PHANTOM IMAGE UPSCALED TO THE RECONSTRUCTION GRID VIA BILINEAR INTERPOLATION AS GROUND TRUTH.

	$\sigma = 0.04$			$\sigma = 0.08$		
	$p_0^{\text{TR}}(\tilde{g})$	$p_0^{\mathbf{R}}(b)$	$p_0^{\mathbf{R}+}(b)$	$p_0^{\text{TR}}(\tilde{g})$	$p_0^{\mathbf{R}}(b)$	$p_0^{\mathbf{R}+}(b)$
MSE	0.0219	0.0324	0.0321	0.0226	0.0334	0.0328
SSIM	0.5134	0.6117	0.6272	0.4163	0.4524	0.2421
PSNR	16.5901	14.8992	14.9406	16.4616	14.7628	14.8392
SNR	10.8592	13.1607	13.1989	4.9591	7.2005	7.3159

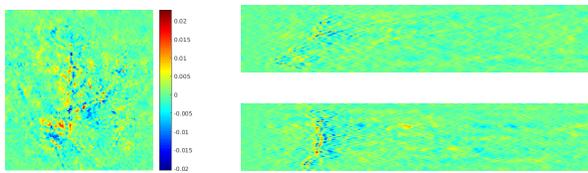
at time step $t = 85$ and scanning locations $y = 59$, $z = 59$ are shown in Figure 17(a). Figure 16(f) illustrates the noise-like difference between the initial pressure obtained via time reversal of the subsampled data b shown in Figure 16(d) and the subsampled, 0-filled data projected on the range of $\tilde{\Psi}$, $\hat{b} = \Phi\tilde{\Psi}^\dagger\tilde{\Psi}b_0(t, \mathbf{x}_S)$ shown in Figure 16(e).



(a) left $b_0(t_{85}, \mathbf{x}_S)$, top right $b_0(t, \mathbf{x}_S^{y59})$, bottom right $b_0(t, \mathbf{x}_S^{z59})$



(b) left $\tilde{g}(t_{85}, \mathbf{x}_S)$, top right $\tilde{g}(t, \mathbf{x}_S^{y59})$, bottom right $\tilde{g}(t, \mathbf{x}_S^{z59})$



(c) $\tilde{g}(t, \mathbf{x}_S) - \hat{g}(t, \mathbf{x}_S)$ at t_{85} , \mathbf{x}_S^{y59} and \mathbf{x}_S^{z59}

Fig. 17. 3D real data - palm18: cross-sections corresponding to time step $t = 85$, and scanning locations $y = 59$, $z = 59$ of (a) 25% uniformly randomly subsampled, 0-filled data $b_0(t, \mathbf{x}_S)$, (b) full data $\tilde{g}(t, \mathbf{x}_S)$ recovered with $\mathbf{R-SALSA}$. (c) $\tilde{\Psi}$ range restricted data recovery error $\tilde{g}(t, \mathbf{x}_S) - \hat{g}(t, \mathbf{x}_S)$.

The 3D data $\tilde{g}(t, \mathbf{x}_S)$ recovered with $\mathbf{R-SALSA}$ ($\tau = 1 \cdot 10^{-5}$, $\mu = 1$, $C = 5$ and stopping tolerance $\eta = 5 \cdot 10^{-4}$

or after 50 iterations) in the first step of the two step method is shown in Figure 17(b) and its range restricted error in (c). Time reversal of $\tilde{g}(t, \mathbf{x}_S)$ results in initial pressure $p_0^{\text{TR}}(\tilde{g})$ in Figure 18(a). The one step $p_0^{\mathbf{R}}(b)$ reconstruction with $\mathbf{R-FISTA}$ ($\tau = 2.5 \cdot 10^{-6}$, $C = 5$ and stopping tolerance $\eta = 5 \cdot 10^{-4}$ or after 50 iterations) is illustrated in Figure 18(b) and $p_0^{\mathbf{R}+}(b)$ its counterpart with non-negativity constraint reconstructed with $\mathbf{R-ADMM}$ ($\tau = 2.5 \cdot 10^{-6}$, $\mu = 0.1$, $C = 5$ and stopping tolerance $\eta = 5 \cdot 10^{-4}$ or after 50 iterations) in Figure 18(c). Again, for comparison in Figure 18(d,e) we plot one step p_0 reconstructions with non-negativity constraint and total variation regularisation $\text{TV}+$ (see [18]) for a range of regularization parameters.

More aggressive subsampling with 12.5% data results in a subtle deprecation of the quality of reconstructed images for all methods: higher background noise, smaller/thinner vessels brake up, more blur (see supplementary material).

VI. CONCLUSION

Summarising, we derived a one-to-one map between the wavefront directions in ambient (and hence image) space and data space which suggested a near equivalence of initial pressure and PAT data volume reconstructions assuming sparsity in Curvelet frame. We introduced a novel wedge restriction to the Curvelet transform which allowed us to define a tight Curvelet frame on the range of PAT forward operator which is crucial to successful full data reconstruction from subsampled data. Intrigued by this near equivalence we compared one and two step approaches to photoacoustic image reconstruction from subsampled data utilising sparsity in tailored Curvelet frames. The major benefit of any two step approach is that it decouples the nonlinear iterative compressed sensing recovery and the linear acoustic propagation problems, thus the expensive acoustic forward and adjoint operators are not iterated with. A further important advantage w.r.t. an earlier two step method [12] is exploitation of the relationship between the time steps (as it is the case for one step methods) by treating the entire PAT data as a volume which is then transformed with the wedge restricted Curvelet frame.

The main drawbacks of the data reconstruction are, that the PAT geometry results in a larger data volume than the original PAT image which is further exacerbated by the commonly employed time oversampling of the PAT data. In effect, we have to deal with a larger volume and consequently more coefficients in compressed sensing recovery. Furthermore flat sensor PAT data is effectively unbounded (only restricted by

sensor size) even for compactly supported initial pressures leading to less sparse/compressed representations. This is reflected in our experiments indicating that PAT data is less compressible in the wedge restricted Curvelet frame than PAT image in the standard Curvelet frame which for the same number of measurements yields lower quality data reconstructions than those obtained with one step methods.

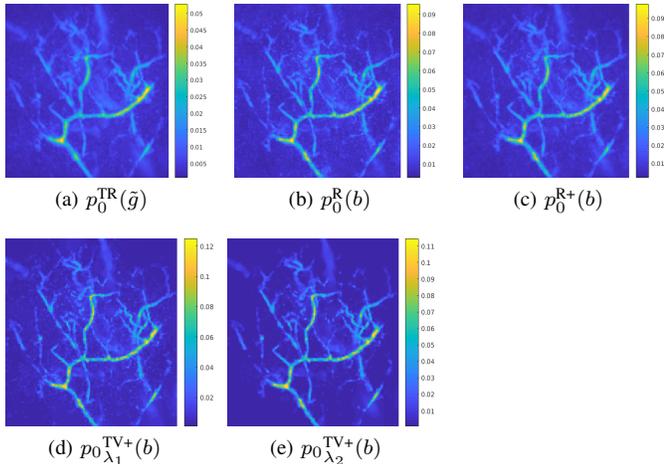


Fig. 18. 3D real data - palm18: maximum intensity projection of initial pressure (a) obtained via **TR** from recovered PAT data $\tilde{g}(\mathbf{x}_S, t)$, (b) $p_0^R(b)$ via one step method $p_0\mathbf{R}$, (c) $p_0^{R+}(b)$ via one step method $p_0\mathbf{R}+$, (d,e) $p_0^{\text{TV}+}(b)$ via one step method **TV+** for regularisation parameter values $\lambda_1 = 2.5 \cdot 10^{-4}$ and $\lambda_2 = 5 \cdot 10^{-4}$.

Comparing one step methods $p_0\mathbf{R}$, $p_0\mathbf{R}+$ to total variation regularisation **TV+** or à la $p_0\mathbf{R}$ reconstructions with Wavelet bases, Curvelet sparsity constraint results in smoother looking vessel images. The reason is that vasculature images commonly contain wide range of scales (vessel diameters) to which multiscale frames can easily adapt. While Wavelet bases share this property, they cannot, in contrast to Curvelets, represent curve like singularities which are common feature of vascular images. On the other hand, no single regularisation parameter value for total variation can effectively denoise larger features without removing/blurring finer ones. The main drawback of the one step methods is their computational cost due to iterative evaluation of the expensive forward and adjoint operators, in particular when using the pseudospectral method which cannot take advantage of sparse sampling. Other forward solvers based e.g. on ray tracing could potentially be used to alleviate some of the costs [43], [44].

Finally, we would like to note that Curvelets provide a Wavelet-Vagulette biorthogonal decomposition for the Radon transform and similar result was shown for PAT in [45]. The wavefront mapping in Section III-A provides a relationship between the image and data domain wavefronts and hence the corresponding Curvelets. However, the proposed reconstruction methods are not based on Wavelet-Vagulette decomposition which is subject of future work.

VII. ACKNOWLEDGEMENTS

The authors acknowledge financial support by EPSRC grant EP/K009745/1 and ERC grant 74119.

REFERENCES

- [1] E. J. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [2] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [3] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [4] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [5] E. J. Candès and D. L. Donoho, “New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities,” *Communications on Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, 2004.
- [6] E. J. Candès and L. Demanet, “The curvelet representation of wave propagators is optimally sparse,” *Communications on Pure and Applied Mathematics*, vol. 58, no. 11, pp. 1472–1528, 2005.
- [7] E. Zhang, J. Laufer, and P. Beard, “Backward-mode multiwavelength photoacoustic scanner using a planar Fabry-Perot polymer film ultrasound sensor for high-resolution three-dimensional imaging of biological tissues,” *Applied Optics*, vol. 47, no. 4, pp. 561–577, 2008.
- [8] 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,” in *Photons Plus Ultrasound: Imaging and Sensing 2017*, vol. 10064, 2017, p. 100641Y.
- [9] N. Huynh, E. Zhang, M. Betcke, S. Arridge, P. Beard, and B. Cox, “Patterned interrogation scheme for compressed sensing photoacoustic imaging using a Fabry Perot planar sensor,” in *Photons Plus Ultrasound: Imaging and Sensing 2014*, vol. 8943, 2014, p. 894327.
- [10] N. Huynh, E. Zhang, M. Betcke, S. R. Arridge, P. Beard, and B. Cox, “A real-time ultrasonic field mapping system using a Fabry Perot single pixel camera for 3D photoacoustic imaging,” in *Photons Plus Ultrasound: Imaging and Sensing 2015*, vol. 9323, 2015, p. 932310.
- [11] N. Huynh, E. Zhang, M. Betcke, S. Arridge, P. Beard, and B. Cox, “Single-pixel optical camera for video rate ultrasonic imaging,” *Optica*, vol. 3, no. 1, pp. 26–29, Jan 2016.
- [12] M. M. Betcke, B. T. Cox, N. Huynh, E. Z. Zhang, P. C. Beard, and S. R. Arridge, “Acoustic wave field reconstruction from compressed measurements with application in photoacoustic tomography,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 710–721, 2017.
- [13] J. Provost and F. Lesage, “The application of compressed sensing for photo-acoustic tomography,” *IEEE Transactions on Medical Imaging*, vol. 28, no. 4, pp. 585–594, 2009.
- [14] Z. Guo, C. Li, L. Song, and L. V. Wang, “Compressed sensing in photoacoustic tomography in vivo,” *Journal of Biomedical Optics*, vol. 15, no. 2, p. 021311, 2010.
- [15] C. Huang, K. Wang, L. Nie, L. V. Wang, and M. A. Anastasio, “Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1097–1110, 2013.
- [16] Z. Belhachmi, T. Glatz, and O. Scherzer, “A direct method for photoacoustic tomography with inhomogeneous sound speed,” *Inverse Problems*, vol. 32, no. 4, p. 045005, 2016.
- [17] S. R. Arridge, M. M. Betcke, B. T. Cox, F. Lucka, and B. E. Treeby, “On the adjoint operator in photoacoustic tomography,” *Inverse Problems*, vol. 32, no. 11, p. 115012, 2016.
- [18] 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,” *Physics in Medicine & Biology*, vol. 61, no. 24, p. 8908, 2016.
- [19] A. Javaherian and S. Holman, “A multi-grid iterative method for photoacoustic tomography,” *IEEE transactions on medical imaging*, vol. 36, no. 3, pp. 696–706, 2016.
- [20] M. Haltmeier and L. V. Nguyen, “Analysis of iterative methods in photoacoustic tomography with variable sound speed,” *SIAM Journal on Imaging Sciences*, vol. 10, no. 2, pp. 751–781, 2017.
- [21] M. Haltmeier, T. Berer, S. Moon, and P. Burgholzer, “Compressed sensing and sparsity in photoacoustic tomography,” *Journal of Optics*, vol. 18, no. 11, p. 114004, 2016.
- [22] M. Xu and L. V. Wang, “Universal back-projection algorithm for photoacoustic computed tomography,” *Physical Review E*, vol. 71, no. 1, p. 016706, 2005.

- [23] K. Wang and M. A. Anastasio, "Photoacoustic and thermoacoustic tomography: image formation principles," in *Handbook of Mathematical Methods in Imaging*. Springer, 2011, pp. 781–815.
- [24] D. Finch and S. K. Patch, "Determining a function from its mean values over a family of spheres," *SIAM Journal on Mathematical Analysis*, vol. 35, no. 5, pp. 1213–1240, 2004.
- [25] P. Stefanov and G. Uhlmann, "Thermoacoustic tomography with variable sound speed," *Inverse Problems*, vol. 25, no. 7, p. 075011, 2009.
- [26] B. E. Treeby and B. T. Cox, "k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of Biomedical Optics*, vol. 15, no. 2, p. 021314, 2010.
- [27] B. Cox and P. Beard, "Fast calculation of pulsed photoacoustic fields in fluids using k-space methods," *The Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3616–3627, 2005.
- [28] K. P. Köstli, M. Frenz, H. Bebie, and H. P. Weber, "Temporal backward projection of optoacoustic pressure transients using Fourier transform methods," *Physics in Medicine & Biology*, vol. 46, no. 7, p. 1863, 2001.
- [29] Y. Xu, D. Feng, and L. V. Wang, "Exact frequency-domain reconstruction for thermoacoustic tomography. I. Planar geometry," *IEEE Transactions on Medical Imaging*, vol. 21, no. 7, pp. 823–828, 2002.
- [30] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, 2006.
- [31] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.
- [32] D. L. Donoho, "Sparse components of images and optimal atomic decompositions," *Constructive Approximation*, vol. 17, no. 3, pp. 353–382, 2001.
- [33] L. Ying, L. Demanet, and E. Candes, "3D discrete curvelet transform," in *Wavelets XI*, vol. 5914, 2005, p. 591413.
- [34] M. Fornasier and H. Rauhut, "Compressive sensing," in *Handbook of Mathematical Methods in Imaging*. Springer, 2011, pp. 187–228.
- [35] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [36] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [37] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, 2009.
- [38] M. Lai and J. Wang, "An unconstrained ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems," *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 82–101, 2011.
- [39] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.
- [40] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [41] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [42] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software*, vol. 8, no. 1, pp. 43–71, 1982.
- [43] F. Rullan and M. M. Betcke, "Hamilton-Green solver for the forward and adjoint problems in photoacoustic tomography," *arXiv preprint arXiv:1810.13196*, 2018.
- [44] F. Rullan, "Photoacoustic tomography: flexible acoustic solvers based on geometrical optics," Ph.D. dissertation, UCL (University College London), 2020.
- [45] J. Frikel and M. Haltmeier, "Sparse regularization of inverse problems by operator-adapted frame thresholding," in *Mathematics of Wave Phenomena*. Springer, 2020, pp. 163–178.