Deep Unfolded Recovery of Sub-Nyquist Sampled Ultrasound Images

Alon Mamistvalov and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*— The most common technique for generating B-mode ultrasound (US) images is delay-and-sum (DAS) beamforming, where the signals received at the transducer array are sampled before an appropriate delay is applied. This necessitates sampling rates exceeding the Nyquist rate and the use of a large number of antenna elements to ensure sufficient image quality. Recently, we proposed methods to reduce the sampling rate and the array size relying on image recovery using iterative algorithms based on compressed sensing (CS) and the finite rate of innovation (FRI) frameworks. Iterative algorithms typically require a large number of iterations, making them difficult to use in real time. In this article, we propose a reconstruction method from sub-Nyquist samples in the time and spatial domain, which is based on unfolding the iterative shrinkage thresholding algorithm (ISTA), resulting in an efficient and interpretable deep network. The inputs to our network are the subsampled beamformed signals after summation and delay in the frequency domain, requiring only a subset of the US signal to be stored for recovery. Our method allows reducing the number of array elements, sampling rate, and computational time while ensuring high-quality imaging performance. Using *in vivo* data, we demonstrate that the proposed method yields high-quality images while reducing the data volume traditionally used up to 36 times. In terms of image resolution and contrast, our technique outperforms previously suggested methods as well as DAS and minimum-variance (MV) beamforming, paving the way to real-time applicable recovery methods.**

***Index Terms*— Beamforming, compressed sensing (CS), deep networks, sub-Nyquist reconstruction, ultrasound (US) imaging, unfolding.**

I. Introduction

MEDICAL ultrasound (US) imaging has become the method of choice in a variety of clinical cases due to its nonradiating and noninvasive nature while being of low cost compared to other medical imaging modalities. In classic US imaging, an output image is built line by line, by transmitting from an array of transducer elements a series of acoustic pulses along a narrow beam. While propagating through the tissue, the pulses are scattered after encountering acoustic impedance perturbations, resulting in echoes that are detected by the same transducer array. The collected data are then processed digitally, resulting in an image line, in a method referred to as beamforming [1]. The beamforming process involves applying an appropriate time delay to the received signals and subsequently averaging them, resulting in a B-mode US image.

Having a relatively low computational load, the most widely used method for beamforming is delay-and-sum (DAS) [2], [3]. In DAS, the delays are implemented digitally after the data are sampled. However, to avoid artifacts and to obtain satisfactory time delay resolution, US signals are traditionally sampled at a rate 4–10 times higher than their actual Nyquist rate [4], which is the rate required for perfect reconstruction of the beamformed US signal [5]. These high rates result in a massive amount of data that need to be stored and processed.

Generating high-resolution US images typically requires that the beampattern of the beamformer will consist of a narrow main lobe and low sidelobes [6]. To this end, adaptive beamforming methods, such as minimum variance (MV), or regularized beamformers [7] can be used [8]. Although MV offers better image resolution, its computational cost is relatively high and increases the data volume. Therefore, in clinical scenarios, where real-time imaging is required, the method of choice is DAS beamforming, leading to decreased image quality. Another possible approach for achieving high-resolution US images is considering larger transducer arrays, with more receiving elements, while keeping the pitch below half the wavelength to avoid grating sidelobes [9]. However, a large set of receiving channels, each consisting of high rate samplers, leads to an enormous amount of data to store and process. Therefore, it is of great importance to consider a US imaging method that will generate high-quality images out of partial data, sampled at a low rate.

To reduce the data size, easing the high computational load, but yet reconstructing high-quality US images, several techniques have been studied recently. Wagner *et al.* [10] and Chernyakova and Eldar [11] considered the translation of the process of time alignment to the frequency domain. They showed that DAS beamforming can be implemented equivalently in the Fourier domain, leading to a method called frequency-domain beamforming (FDBF). Based on FDBF,

Manuscript received March 1, 2021; accepted June 26, 2021. Date of publication June 29, 2021; date of current version November 23, 2021. This work was supported in part by the Igel Manya Center for Biomedical Engineering and Signal Processing, in part by the Benoziyo Endowment Fund for the Advancement of Science, in part by the Estate of Olga Klein–Astrachan, and in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 646804-ERC-COG-BNYQ. *(Corresponding author: Alon Mamistvalov.)*

The authors are with the Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: alon.mamistvalov@weizmann.ac.il; yonina.eldar@weizmann.ac.il).

Digital Object Identifier 10.1109/TUFFC.2021.3093507

they considered sub-Nyquist sampling while reconstructing the signals using compressed sensing (CS) techniques [12] based on the finite rate of innovation (FRI) structure of the beamformed signal [13].

Other approaches for reducing the number of receiving elements were studied in the literature. These methods consider the usage of sparse arrays [14]–[17] where some of the elements are removed. In [18], a technique for reducing the data size using a sparse array with a new beamforming method called convolutional beamforming algorithm (COBA) is presented. Cohen and Eldar showed that by choosing an array that satisfies appropriate conditions, the achieved beampattern of the virtual array, based on the array's sum coarray [19]–[21], is equivalent to the full uniform linear array (ULA). Thus, using a subset of array elements, whose sum coarray is the full ULA, results in better resolution and contrast using fewer receiving elements.

In [22] and [23], we combined the ideas of spatial and sampling rate reduction. High-quality US images were produced by sampling the signals below their Nyquist rate, using a sparse set of array elements, in a method called compressed frequency-domain COBA (CFCOBA). For reconstructing the signals, the FRI structure of the convolutionally beamformed signal was exploited [22], leading to the recovery of the signal based on solving an optimization problem using CS techniques and iterative algorithms.

Many other works, inspired by the impressive performance of deep learning [24], in a variety of fields, including medical imaging [25], combined the ideas of deep learning with US imaging [26]. To the best of our knowledge, all previously suggested deep learning-based methods in US imaging use and store full channel data of all transmitting elements, leading to an enormous size of information to store. In addition, general deep networks for US imaging are traditionally more complex as they need to process high-dimensional tensors. Zhuang and Chen [27], for example, combined MV beamforming with deep learning for contrast improvement. Their idea is to use an ensemble of networks, operating over subbands in the frequency domain, before or after computing the adaptive weights. Luijten *et al.* [28] considered applying adaptive weights, which are calculated using a deep network with a fully connected deep neural network. Kessler and Eldar [29] used temporally and spatially subsampled data to generate high-quality US images using a deep neural network that is based on a UNet architecture [30]. However, they reached only an 11-fold reduction in data size. Moreover, their recovery method required the storage of the diluted data from all of the channels, while in practice, as we will show, only a subset of the Fourier coefficients of the beamformed signal itself, after the summation over the elements, is needed, leading to further reduction of the data size.

Our main goal in this work is to present an efficient recovery method of a beamformed signal from a small subset of its samples based on a deep unfolded neural network. The proposed method, as opposed to other complicated deep learning approaches for beamforming, consists of fast and efficient training and testing steps since only a small fraction of the samples of the beamformed signal needs to be stored for training the network and for full image recovery. Namely, while previous works needed data from all receiving channels, we suggest using subsamples of the beamformed signal. Furthermore, the suggested technique for recovery is based on a relatively small pretrained deep network, leading to much faster reconstruction when compared to iterative approaches. The method we propose is based on previously suggested frameworks, including inverse problems, FRI structure, and algorithm unrolling; we show that combining these ideas yields high-quality experimental results.

We begin by introducing the model of the beamformed signal, which follows an FRI structure, and formulate a recovery problem from the subsampled beamformed signal. In previous works, the problem was solved using CS techniques with iterative shorthand for Nesterov's algorithm (NESTA) [31], [32]. Here, we present a deep network architecture, based on learned iterative shrinkage thresholding algorithm (LISTA) [33], [34], which unrolls the iterative ISTA algorithm [35]. The same unrolling approach was shown to be successful for super-resolution in US [26]. The basic architecture of our network is the same, namely, cascading several iterations of ISTA, resulting in a deep network. However, we slightly modified the network by adding a layer for final recovery. In addition, we changed the training process to better fit the characteristics of the US signal by choosing an appropriate loss function that is different from standard mean squared error (MSE), traditionally used for LISTA. In our method, channel data are acquired using a sparse set of array elements, each sampled below its effective Nyquist rate. The input to our network is the subsamples of the beamformed signal after delay is applied in the frequency domain following [11]. The signals are then summed according to FDBF or CFCOBA, resulting in the input to our network. We train our network with DAS beamformed targets, where training is performed for each image line separately. Despite the data size reduction that reaches 36-times fewer data and the simplification of the training process compared to other deep learning-based approaches, our network consists of only 30 layers that are only half the iterations needed for reconstruction in an iterative manner using NESTA, which incorporates more complex steps. Specifically, we present recovery in computation time of $O(n)$ achieving a reduction on the order of square root the operations needed for recovery in NESTA, $O(n^2)$. Moreover, our method shows higher robustness to data size shrinkage by yielding better results in cases of higher data size reduction. Thus, our approach offers much faster and higher quality imaging than those of existing methods.

The rest of this article is organized as follows. Section II discusses several previously proposed beamforming methods. Section III reviews methods for beamformed signal recovery from subsampled data based on iterative algorithms and algorithm unrolling. In Section IV, we present our method and discuss its properties. The performance of the suggested technique is validated in Section V using phantom and *in vivo* scans. This article is concluded in Section VI.

II. US Beamforming Techniques

Using multiple transducer elements for transmission and reception of acoustic pulses, a US image is built for each direction θ, line by line. In transmission, a pulse is transmitted by each element in the array, where the pulses are delayed and scaled so that their sum creates a directional beam propagating at a specific direction through the tissue. The real-time computational complexity in the transmission is negligible since the relevant parameters per angle are calculated offline and saved in tables. On the other hand, beamforming in reception is more challenging. A line in the image is formed by dynamically delaying the signals received at each element of the array and then averaging them, which leads to signal-to-noise-ratio (SNR) enhancement. Here, we describe both the standard beamforming method and techniques that reduce the number of samples required for imaging.

A. Time-Domain Beamforming

Consider a ULA U comprised of M transducer elements aligned along the x-axis. Let m_0 be the reference element, at the origin, and δ_m denote the distance between m_0 and the mth element. At $t = 0$, a pulse is transmitted by each transducer element, resulting in a beam propagating at direction θ through the tissue, with the speed of sound in the tissue, c. Reflectors in the tissue cause echoes that are received by all array elements at times that depend on their location. Let $\phi_m(t)$ be the received signal at the mth element. The process of beamforming involves averaging the signals received by different array elements while applying appropriate time shifts for alignment due to the differences in arrival time. The geometry of the imaging setup defines the delay that needs to be applied. The aligned signal, $\hat{\phi}_m(t;\theta)$, results from applying an appropriate delay to $\phi_m(t)$ and aligning it to m_0 [10], [36]

$$\hat{\phi}_m(t;\theta) = \phi_m(\tau_m(t,\theta))$$
$$\tau_m = \frac{1}{2}(t + \sqrt{t^2 - 4(\delta_m/c)t\sin\theta + 4(\delta_m/c)^2}). \quad (1)$$

These signals are then averaged to form the beamformed signal for direction θ

$$\Phi_{\text{DAS}}(t;\theta) = \frac{1}{M}\sum_{m=1}^{M}\hat{\phi}_m(t;\theta). \quad (2)$$

DAS beamforming is performed digitally. To achieve delays on the order of nanoseconds, the required sampling rate is on the order of hundreds of megahertz [37], a requirement that is not feasible. Therefore, US signals are sampled at lower rates, on the order of tens of megahertz, and fine delay resolution is obtained by digital interpolation. However, these lower rates are still much higher than the Nyquist rate of the signal, which is defined by twice its bandwidth [4]. In general, a well-known rule of thumb is that the sampling rate of the signal should be 4–10 times the transducer's central frequency. These sampling rates lead to a huge amount of data that need to be stored and processed.

B. Frequency-Domain Beamforming

To reduce the required sampling rate, applying the delay in the frequency domain was presented in [11]. In addition, it was shown that using a small number of Fourier coefficients of the received signals, beamforming can be performed efficiently. Let $c[k]$ denote the kth Fourier series coefficient of the beamformed signal, $\Phi(t;\theta)$ as described in (2), and let $\hat{c}_m[k]$ be the kth Fourier series coefficient of $\hat{\phi}_m(t;\theta)$. Then,

$$c[k] = \frac{1}{M}\sum_{m=1}^{M}\hat{c}_m[k] \quad (3)$$

where $\hat{c}_m[k]$ is obtained by calculating

$$\hat{c}_m[k] = \frac{1}{T}\int_0^T I_{[0,T_B(\theta))}(t)\hat{\phi}_m(t;\theta)e^{\frac{-2\pi j}{T}kt}dt. \quad (4)$$

The indicator function, $I_{[a,b)}$, is equal to 1 when $a \le t < b$. Here, $T_B(\theta) < T$ is defined in [10] and T is the pulse penetration depth, and the beam is supported on $[0, T_B(\theta))$.

Following [10] and [11], one can write the Fourier coefficients of the delayed signal as

$$\hat{c}_m[k] = \sum_{n=-\infty}^{\infty} c_m[k-n]Q_{k,m;\theta}[n] \quad (5)$$

where $c_m[k]$ are the Fourier coefficients of the received signal in the mth channel before the delay is applied. The variables $Q_{k,m;\theta}[n]$ are the Fourier coefficients of a distortion function, determined solely by the geometry of the imaging setup and can be computed offline once and stored in memory. The delay in the frequency domain is obtained by using the Fourier coefficients of the distortion function. The decay properties of $\{Q_{k,m;\theta}[n]\}$ and the fact that most of the energy of this set is centered around the dc component lead to the option of replacing the summation in (5) with a relatively small finite summation. Thus, assuming that $\{Q_{k,m;\theta}[n]\}$ decays rapidly for $n < -N_1$, $n > N_2$, where $N_1, N_2 \in \mathbb{N}$, it leads to

$$\hat{c}_m[k] = \sum_{n=-N_1}^{N_2} c_m[k-n]Q_{k,m;\theta}[n]. \quad (6)$$

Combining (6) and (3) leads to FDBF

$$c[k] = \frac{1}{M}\sum_{m=1}^{M}\sum_{n=-N_1}^{N_2} c_m[k-n]Q_{k,m;\theta}[n]. \quad (7)$$

With appropriate zero padding and applying an inverse Fourier transform to $\{c[k]\}$, the time-domain beamformed signal is obtained.

The main advantage of FDBF is that only the nonzero Fourier coefficients of the received signal are required. Those coefficients are obtained from sub-Nyquist samples of the signal at each receiving element, leading to significant reduction in the sampling rate. To obtain the sub-Nyquist samples of the signal, we follow the mechanism proposed and discussed in [13] and [38]. The output is sampled at its effective Nyquist rate, and the required Fourier coefficients are the Fourier transform of the output; using these coefficients, the beamformed signal is recovered, as will be discussed in Section II-E.

C. Convolutional Beamforming

Using a sparse set of array elements, following COBA [18], [20], can lead to substantial data reduction while producing images at least as good as those obtained with DAS. As was shown in [18], using COBA, one can achieve the effective beampattern as would have been obtained using the sum coarray of the given array of transducers [39].

The beamformed signal in COBA is

$$\Phi_{\mathrm{COBA}}(p;\theta)=\sum_{n=-(N-1)}^{N-1}\sum_{m=-(N-1)}^{N-1}u_n(p;\theta)u_m(p;\theta) \tag{8}$$

where

$$u_m(p;\theta)=e^{j\angle\hat{\phi}_m(p;\theta)}\sqrt{|\hat{\phi}_m(p;\theta)|} \tag{9}$$

with $\hat{\phi}_m(p,\theta)$ the pth sample of the delayed signal at channel m, sampled at sampling intervals $T_s=(1/f_s)$, with $\angle\hat{\phi}_m(p,\theta)$ and $|\hat{\phi}_m(p,\theta)|$ being its phase and the magnitude, respectively. Applying (9) ensures that the product of COBA will have the same order of the received signal. An equivalent form of (8) is

$$\Phi_{\mathrm{COBA}}(p;\theta)=\sum_{n=-2(N-1)}^{2(N-1)}s_n(p;\theta) \tag{10}$$

where

$$s_n(p;\theta)=\sum_{i,j\in U,i+j=n}u_i(p;\theta)u_j(p;\theta) \tag{11}$$

with $n=-2(N-1),\ldots,2(N-1)$. Therefore, the vector s can be calculated by applying lateral convolution on u with itself, after zero padding it to length $2|M|-1$, i.e., the entries s_n can be calculated by

$$s(p;\theta)=u(p;\theta)\underset{s}{*}u(p;\theta) \tag{12}$$

where $*$ stands for convolution over the spatial dimension.

The beampattern of DAS beamforming is given by

$$H_{\mathrm{DAS}}(\theta)=\sum_{n=-(N-1)}^{N-1}\exp\left(-j\omega_0\frac{\delta_n\sin\theta}{c}\right) \tag{13}$$

where ω_0 is the central frequency of the transducer and δ is the distance between two adjacent array elements such that $\delta_n=n\delta$. Following [18] and based on (8), we see that $H_{\mathrm{COBA}}=H_{\mathrm{DAS}}H_{\mathrm{DAS}}$, and it can be written as a single polynomial

$$H_{\mathrm{COBA}}(\theta)=\sum_{n=-2(N-1)}^{2(N-1)}a_n\exp\left(-j\omega_0\frac{\delta\sin\theta}{c}n\right) \tag{14}$$

where a_n are intrinsic apodization weights given by $\boldsymbol{a}=\mathbb{I}_M\underset{s}{*}\mathbb{I}_M$. Here, $\mathbb{I}_M$ is a binary vector whose mth entry is 1 if $m\in M$.

The beampattern given by (14) is equivalent to that of a DAS beamformer, using the sum coarray of the original ULA, U. Due to the fact that effectively the array used for imaging is bigger, improved imaging performance is obtained. Therefore, Cohen and Eldar [18] suggested using a sparse array, given by the original array after removing several elements, for generating a desired beamformer beampattern, in a method called sparse COBA (SCOBA). The use of fractal arrays [40]–[44] was discussed in [20]. The geometry of such arrays is given by

$$\begin{aligned}W_0&=0\\ W_{r+1}&=\cup_{n\in\mathbb{G}}(W_r+nL^r),\quad r\in\mathbb{N}\end{aligned} \tag{15}$$

where r is the array order, $\mathbb{G}$ is the generator array in fractal terminology, and $\min(\mathbb{G})=0$. The factor L is given by $L=2\max(\mathbb{G})+1$ and is referred to as the translation factor. This choice of geometry leads to the resulting array W, which after applying convolutional beamforming provides an effective beampattern with better image resolution and contrast.

D. Compressed Frequency-Domain Convolutional Beamforming

Further reduction of the data volume was discussed in [22], where both the time and the spatial dilution are applied. The method is referred to as CFCOBA, and it combines the ideas of FDBF and SCOBA using a fractal array geometry. In that way, the main advantages of both FDBF and SCOBA are obtained, and the beamformed signal is recovered from only a partial set of samples, acquired by a sparse set of array elements while preserving the same image quality.

Following the steps in [22], we let N_{sN} be the number of samples acquired by sub-Nyquist sampling of the received signal [13]. The compressed frequency-domain signal is defined by

$$\hat{\Phi}(n_s)_{\mathrm{CFCOBA}}=\sum_{n\in U}\sum_{m\in U}\hat{\phi}_n(n_s)\hat{\phi}_m(n_s) \tag{16}$$

where $n_s=0,\ldots,N_{\mathrm{sN}}-1$, $N_{\mathrm{sN}}=|\mu|$ are the discrete time samples with μ being the set of Fourier series coefficients obtained using the Xampling mechanism [13], [45]–[47], and $\hat{\phi}_m(n_s)$ and $\hat{\phi}_n(n_s)$ are the delayed signals at channels m and n, respectively, with the delay applied in the frequency domain. Based on the derivations in [22], by calculating the Fourier series coefficients of both sides of (16), a relation between the subsampled Fourier coefficients of the received signal and the Fourier coefficients of the convolutionally beamformed signal is given by

$$\begin{aligned}\hat{c}[k]_{\mathrm{CFCOBA}}&=N_{\mathrm{sN}}\sum_{l\in U}\sum_{m\in U}\sum_{p+q=k}\hat{c}_m[p]\hat{c}_l[q]\\ &=N_{\mathrm{sN}}\sum_{l\in U}\sum_{m\in U}(\hat{c}_m*\hat{c}_l)[k]\end{aligned} \tag{17}$$

where the last equation is obtained by setting $q=k-p$, and $\hat{c}_m[p]$ and $\hat{c}_l[q]$, the Fourier coefficients at channels m and l, are calculated by (6).

To efficiently calculate the relation in (17), it was shown in [22] that due to the linear operations in the temporal dimension, the relation between the Fourier coefficients of the received signals and those of the convolutionally beamformed signal is given by

$$\hat{c}[k]_{\mathrm{CFCOBA}}=N_{\mathrm{sN}}\sum_{l\in S_{\tilde{U}}}(\hat{c}\underset{s}{*}*\hat{c})_l[k] \tag{18}$$

where $\tilde{U}$ is a sparse array based on a sparse set of elements from the ULA, U, that was previously defined, and $S_{\tilde{U}}$ is its corresponding sum coarray [39]. The calculation in (18) denotes a 2-D convolution operation, one over the temporal dimension and one over the spatial dimension. The calculation is performed efficiently using inverse fast Fourier transform (IFFT) and appropriate zero padding based on the convolution theorem.

In [22], the FRI structure of the convolutionally beamformed signal was proved, and it was shown that it consists of time-shifted replicas of the square of the known transmitted pulse. This leads to the possibility of reconstructing the signal based on a portion of its Fourier coefficients.

In Section II-E, we discuss a possible iterative recovery approach of the subsampled signal in both methods, FDBF and CFCOBA.

E. CS Recovery for Further Reduction

When sampling below the effective Nyquist rate of the signal, applying IFFT is insufficient. In [11] and [22], CS techniques were suggested to fully recover the signal based on its FRI structure. We follow the steps in [11], but the same steps are performed in [22] using the square of the known transmitted pulse.

According to [10], a beamformed signal obeys an FRI model, which means that it can be written as a sum of shifted replicas of the known transmitted pulse, $h(t)$, with unknown amplitudes and delays

$$\hat{\Phi}(t) = \sum_{s=1}^{S} a_s h(t - t_s) \tag{19}$$

where S is the number of scattering elements in the tissue in a certain direction θ and $\{a_s\}$ and $\{t_s\}$ are the unknown amplitudes of the reflections and the times at which the reflection from the sth scatterer arrived at the receiving element, respectively. Based on the FRI model and the known pulse shape, the beamformed signal is completely defined by $2S$ parameters. For a duration T, the Fourier series expansion of (19) can be written as

$$\begin{aligned}\hat{c}[k] &= \frac{1}{T}\int_0^T \sum_{s=1}^{S} a_s h(t - t_s) e^{j((2\pi)/T)kt} dt \\ &= H\left(\frac{2\pi k}{T}\right) \sum_{s=1}^{S} a_s e^{-j((2\pi)/T)kt_s}\end{aligned} \tag{20}$$

where $\hat{c}[k]$ are the Fourier coefficients of the beamformed signal and $H(\omega)$ is the continuous-time Fourier transform of $h(t)$. Writing (20) in matrix form results in

$$\hat{c} = \boldsymbol{H}\boldsymbol{V}\boldsymbol{a}, \tag{21}$$

where $\hat{\boldsymbol{c}}$ is a vector of length $|\mu|$ with $\{\hat{c}[k]\}_{k\in\mu}$ as its entries, $\boldsymbol{H}$ is a diagonal matrix of size $|\mu| \times |\mu|$ with $H(2\pi k/T)$ on its diagonal, $\boldsymbol{V}$ is a Fourier matrix of size $|\mu| \times S$ with the (k, s)th element $e^{-j((2\pi)/T)kt_s}$, and $\boldsymbol{a}$ is a vector of size S with the amplitudes, $\{a_s\}$, as its entries. If the set of k is comprised consecutive indices, then the matrix V has a Vandermonde form and has full column rank as long as $|\mu| \geq S$ and the time delays $t_s \neq t_{s'}$, $\forall s \neq s'$. Moreover, if $|\mu| \geq 2S$, then this problem can be solved for the unknown delays and amplitudes using methods, such as an annihilating filter [48].

In practice, the recovery problem is solved using CS methods. By quantizing the delays with step size $T_s = (1/f_s)$, such that $t_s = q_s T_s$, where f_s is the sampling rate for traditional DAS beamforming, and letting $N = \lfloor T/T_s \rfloor$, the Fourier coefficients can be written as

$$\hat{c}[k] = H\left(\frac{2\pi k}{T}\right) \sum_{s=0}^{N-1} \tilde{a}_s e^{-j((2\pi)/N)ks}. \tag{22}$$

The vector $\tilde{\boldsymbol{a}}$ of length N is defined as

$$\tilde{a}_s = \begin{cases} a_s, & \text{if } s = q_s \\ 0, & \text{else.} \end{cases} \tag{23}$$

The recovery problem then reduces to determining the S-sparse vector $\tilde{\boldsymbol{a}}$ from

$$\hat{c} = \boldsymbol{H}\boldsymbol{D}\tilde{\boldsymbol{a}} = \boldsymbol{A}\tilde{\boldsymbol{a}} \tag{24}$$

where $\boldsymbol{D}$ is a $|\mu| \times N$ matrix, formed by taking the set μ of rows from an $N \times N$ FFT matrix. By choosing $|\mu| \geq CL(\log N)^4$ rows uniformly at random for some constant $C > 0$, the matrix $\boldsymbol{A}$ obeys the restricted isometry property (RIP) with high probability [49]. The proposed formulation has a standard CS structure.

In [10], an l_0 regularizer was suggested, assuming that the US signal is sparse. However, this suggestion produced worse results when compared to an l_1 regularizer [11]. The reason for these results is that in practice, due to speckle, the coefficient vector $\boldsymbol{a}$, defined in (23), is only approximately sparse. Hence, to reconstruct $\boldsymbol{a}$, we use the l_1 norm, leading to

$$\min_{\tilde{a}} \|\tilde{\boldsymbol{a}}\|_1 \quad \text{s.t } \|\boldsymbol{A}\tilde{\boldsymbol{a}} - \hat{\boldsymbol{c}}\|_2 \leq \epsilon \tag{25}$$

with ϵ being an appropriate noise level. This optimization problem can be solved using various known techniques, such as interior-point methods [50] or iterative shrinkage techniques [35], [51]. For solving this problem, Chernyakova and Eldar [11] suggested using the NESTA algorithm [31], [32]. This algorithm was shown to be highly suitable for a compressible signal with a high dynamic range, which is true in the case of US signals. However, applying NESTA to signals sampled at very low rates leading to a small number of samples fails to fully recover the signal, and artifacts, such as high granularity, are present in the resulting image. Moreover, NESTA's computational complexity is relatively high compared to other iterative algorithms, such as ISTA. As will be discussed in Section III-A, ISTA consists of only two matrix multiplications and one nonlinear operation per iteration; one iteration of NESTA includes more than twice the operations in ISTA, making it less efficient. For easing the computational load, one may suggest using ISTA for recovering the signal; however, as shown in Fig. 1, even after a relatively high number of 100 iterations, the algorithm fails to recover the beamformed signal and many artifacts are present in the image.

To overcome the difficulties of recovering the signal from a very small set of samples and the computational load,

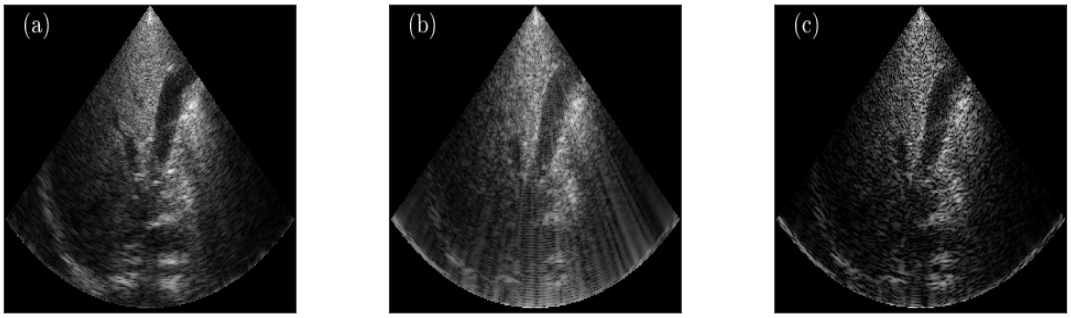

Fig. 1. *In vivo* images produced with (a) DAS - 1920 samples, (b) FDBF + ISTA recovery—130 samples and 100 iterations, and (c) FDBF + NESTA recovery—130 samples and 60 iterations.

we next propose a deep unfolded recovery. In Section V, we show that for several reduction factors in data size, ranging from 8- to 36-fold dilution, the proposed method performs better than previously suggested iterative algorithms in terms of image quality, such as resolution and contrast. Moreover, the suggested technique is more efficient than the iterative techniques and requires only 30 layers of a deep network for recovering the beamformed signal. Finally, when comparing our method to traditional end-to-end DNNs, which typically require storing the complete channel data prior to DAS, in our method, we suggest storing only a small set of samples of the beamformed signal, leading to a considerable reduction in data size.

III. Model-Based Recovery via a Deep Unfolded Network

Iterative algorithms that consider the structure and the model of the beamformed signal, such as ISTA and NESTA, run iteratively until a certain threshold condition is met. However, they suffer from high computational load as well as artifacts and low image quality when the number of samples is very low. By unfolding the iterations of these algorithms as deep network layers, in a method that is referred to as algorithm unfolding or algorithm unrolling [26], [33], [34], the quality of the recovery is enhanced. In this section, we briefly present the ISTA algorithm for sparse recovery and discuss a possible way of cascading its iteration to a deep network resulting in LISTA. The network proposes the usage of an architecture dedicated to sparse recovery, which consists of a relatively small number of layers compared to the number of iterations needed for ISTA and NESTA [26].

A. Iterative Shrinkage Threshold Algorithm

The ISTA algorithm is a widely used algorithm for sparse inference. The convex optimization problem presented in (25) can be written in an equivalent form as

$$\min_{x} \ \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_1 \tag{26}$$

for a given input vector $\boldsymbol{y} \in \mathbb{C}^n$, where $\boldsymbol{A} \in \mathbb{C}^{n\times m}$, with $m \geq n$, is the measurement matrix and $\boldsymbol{x} \in \mathbb{C}^m$ is the sparse code representation based on the measurement matrix $\boldsymbol{A}$ and λ is a regularization parameter that controls the sparseness of the solution. The block diagram of ISTA is presented in Fig. 2. The main part of ISTA is recursively iterating, until convergence, the following iterations:

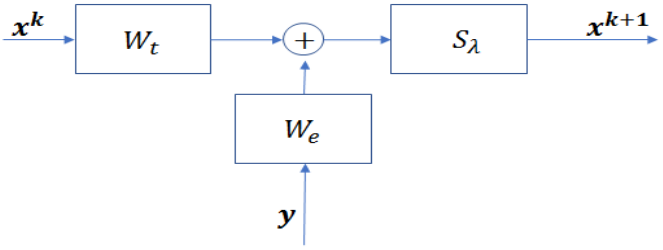

Fig. 2. Block diagram of the kth iteration of the standard ISTA algorithm.

$$\boldsymbol{x}^{k+1}=S_\lambda(\boldsymbol{W}_e\boldsymbol{y}+\boldsymbol{W}_t\boldsymbol{x}^k)), \quad \boldsymbol{x}^0=0 \tag{27}$$

where $\boldsymbol{x}^{k+1}, \boldsymbol{x}^k$ describe the kth iteration approximation for $\boldsymbol{x}$, $\boldsymbol{W}_e=\mu\boldsymbol{A}^T$, $\boldsymbol{W}_t=\boldsymbol{I}-\mu\boldsymbol{A}^T\boldsymbol{A}$ with step size μ, and S_λ is a soft-thresholding operator with threshold λ.

B. Learned Iterative Shrinkage Threshold Algorithm

To overcome the disadvantages of the iterative algorithms, such as high computational load and the difficulty of recovering the signal from a small number of samples, one can unfold the algorithm's iterations into network layers, with each layer replacing the standard algorithm iteration. Unfolding was found to be beneficial when the measurement matrix is not known precisely, which can be the case in US imaging when the scanning depth is changed [33], and was also shown to be useful in super-resolution tasks [26].

LISTA is based on the ISTA algorithm whose iteration is shown in Fig. 2. The iteration can be easily converted into a single deep network layer since it comprises a series of analytic operations (summation, matrix–vector multiplication, and soft thresholding), which are of the same nature as basic operations of a deep neural network. Therefore, executing K iterations of ISTA can be converted to a network of K such cascaded layers. In the unfolded network of LISTA, each layer k of the network consists of trainable convolutions $\boldsymbol{W}_e^k$ and $\boldsymbol{W}_t^k$ and a trainable shrinkage parameter λ_k, corresponding to the threshold in ISTA. The block diagram of LISTA is shown in Fig. 3, within the dashed yellow rectangle.

Training is performed with a sequence of input vectors, $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^N \in \mathbb{R}^n$ and their corresponding ground-truth sparse

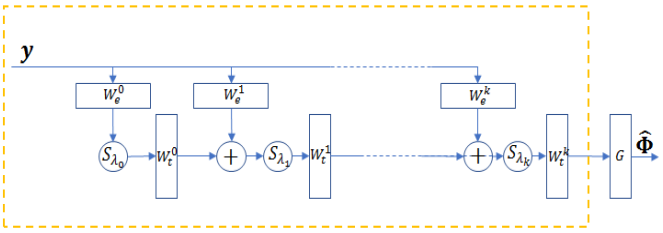

Fig. 3. LISTA block diagram. In the proposed network, the layer $\boldsymbol{G}$ is added for calculating the output beamformed signal $\boldsymbol{\Phi}$. All rectangles represent 5×1 1-D trainable convolution layers, and S_{λ_k} is a nonlinear layer with trainable parameter λ_k, given by (30) in the proposed architecture.

codes $\tilde{\boldsymbol{x}}^1, \ldots, \tilde{\boldsymbol{x}}^N$. The training process is done by feeding an input, $\boldsymbol{y}^i$, $i \in 1, \ldots, N$, to the network and retrieving its corresponding predicted sparse code, $\hat{\boldsymbol{x}}^i(\boldsymbol{y}^i; \boldsymbol{W}_t, \boldsymbol{W}_e, \lambda)$. The output is then compared to the ground-truth sparse code, $\tilde{\boldsymbol{x}}^i$, using a loss function

$$L(\boldsymbol{W}_t, \boldsymbol{W}_e, \lambda) = \frac{1}{N} \sum_{i=1}^{N} \|\hat{\boldsymbol{x}}^i(\boldsymbol{y}^i; \boldsymbol{W}_t, \boldsymbol{W}_e, \lambda) - \tilde{\boldsymbol{x}}^i\|_2^2. \quad (28)$$

Training is performed through loss minimization using standard methods such as stochastic gradient descent [52] to learn the weights $\boldsymbol{W}_t^k, \boldsymbol{W}_e^k, \lambda_k$ for each layer. After training is completed, the predicted sparse code is obtained by inserting an input from a dataset different from the training set.

IV. LISTA for Subsampled US Image Reconstruction

Here, we discuss the recovery of the beamformed signal from a sparse set of its Fourier coefficients. Since the signal is sampled below the Nyquist rate, it suffers from aliasing in the time domain, after applying the inverse Fourier transform. Our goal is to determine the recovered beamformed signal from the subsampled signal using a deep network. Our output is a high-quality B-mode US image, consisting of recovered image lines, one for each θ, the propagating direction.

A. Problem Formulation

In Section II-E, we have seen that a beamformed signal obeys an FRI model, and therefore, its recovery from a subset of Fourier coefficients is reduced to determining a sparse vector from (24). Here, we suggest solving (24) in the time domain using LISTA. Therefore, the Fourier coefficients are zero-padded to maintain the appropriate temporal resolution of the beamformed signal, as described in Section IV-C, and transformed back to the time domain, leading to

$$\boldsymbol{\Phi} = \boldsymbol{G}\tilde{\boldsymbol{a}}. \quad (29)$$

Here, $\boldsymbol{\Phi} \in \mathbb{R}^{N_{\text{st}}}$ is a vector of samples of the beamformed signal, whose Fourier coefficients are the subsampled coefficients, $\boldsymbol{G} \in \mathbb{R}^{N_{\text{st}} \times N_{\text{st}}}$ is a known convolution matrix with time samples of the known transmitted pulse as its entries, and $\tilde{\boldsymbol{a}} \in \mathbb{R}^{N_{\text{st}}}$ is the L-sparse vector such that $N_{\text{st}} \gg L$. We insert the vector $\boldsymbol{\Phi}$ to LISTA, as the input $\boldsymbol{y}$ and result in $\hat{\boldsymbol{x}}$ which is the predicted beamformed signal.

B. Network Architecture

The main goal of our deep unfolded network is to recover the beamformed signal from temporal and spatial subsampled data. Although the final output is an image, the training is performed line by line, and the output is the beamformed signal corresponding to each direction in the image.

The basic architecture of our network is similar to that discussed in Section III-B. The input to our network is the 1-D vector of the subsampled beamformed signal, $\boldsymbol{\Phi}$. The output of the proposed network is an estimated vector of the beamformed line, of the same length as the input, denoted by $\hat{\boldsymbol{\Phi}}$. To better suit our recovery problem, we made two modifications to LISTA. First, a convolution layer is added to the output layer. The new output layer is designed to act as the final conversion between the sparse code, generated by the traditional LISTA, denoted by $\tilde{\boldsymbol{a}}$, and actual replicas of the transmitted pulse, namely it mimics $\boldsymbol{G}$ in (29). The second change is choosing the soft-thresholding layer, S_{λ_k}, to be of the form [53]

$$S_{\lambda_k}(x) = \frac{x}{1 + e^{-(|x| - \lambda_k)}} \quad (30)$$

where λ_k is learned and is different in each layer $k \in \{0, \ldots, K\}$ of the network. This term was chosen after empirically yielding good results.

The rest of the network consists of 1-D operations designed to replace ISTA, and cascaded to overall 30 layers of a deep network, smaller networks that consist of fewer layers were tested; however, they empirically failed to produce fine results in the case of the lowest sampling rate, specifically we checked ten-layer architecture. The final size of the network of 30 layers was finally chosen due to its ability to produce both visually and quantitatively good results. Similar to LISTA presented in Sections I and III-B, the learned parameters are weights of the convolution layers $\boldsymbol{W}_e^k$ and $\boldsymbol{W}_t^k$, which are of size 5×1, and a shrinkage parameter λ_k as described in (30). Therefore, the output of the network is $\hat{\boldsymbol{\Phi}}(\boldsymbol{\Phi}; \boldsymbol{W}_t, \boldsymbol{W}_e, \lambda)$.

C. Data Subsampling and Preprocessing

We evaluated our method on three different datasets. Prior to training and testing, the data is transferred through a preprocessing pipeline consisting of several steps. Two of the datasets consist of temporal subsampled data using the Xampling mechanism [13] with different reduction factors. The third dataset consists of spatial reduction as well, following [18], [20], [22]. The subsampled Fourier coefficients are calculated as previously described in Section II-B, for the first two datasets, and following Section II-D for the third dataset. After calculating the Fourier coefficients, each dataset is transformed back to the time domain, by restoring its negative side spectrum, and zero padding with appropriate zero-vector size, to maintain the same temporal resolution in time as the DAS beamformed signal. Finally, applying the inverse Fourier transform results in the subsampled signal in the time domain, which is the input to our proposed network, described in Section IV-B.

The entire flow of the proposed method is shown in Fig. 4. For convenience and clarity of presentation, the figure depicts

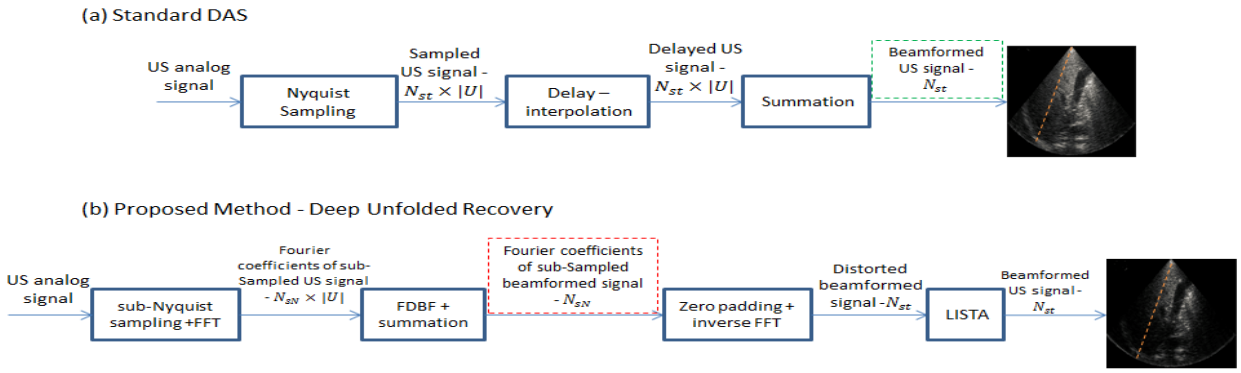

Fig. 4. Beamforming methods algorithm flow. (a) Standard DAS beamforming. Nyquist sampling acquires $N_{\text{st}} \times |U|$ samples in $|U|$ receiving elements, the signal is delayed via linear interpolation and summed to obtain an image line of size N_{st}. The green dashed line denotes the size of the vector stored in DAS. (b) Proposed technique. US signal is subsampled to size $N_{\text{sN}} \times |U|$, and the signal is then delayed in the frequency domain, resulting in the Fourier coefficients of the beamformed signal. Before inserting the signal into the network, the temporal resolution is preserved by zero padding, and the signal is transformed back to the time domain before inserted into the LISTA network. The red dashed line denotes the size of the vector where $N_{\text{st}} \gg N_{\text{sN}}$.

the flow of reduction in sampling rate without reducing the number of elements, i.e., for the first two datasets. One can adjust the sizes to fit the case of reduction in receiving elements and adding the spatial convolution block after delaying the signal in the frequency domain, following (18).

D. Loss Function

In US imaging, the data are characterized by high dynamic range and is typically compressed after beamforming to generate the final image. Therefore, we follow [28], [29], replace the MSE function used for LISTA traditionally (28), and use signed-mean-squared-logarithmic error (SMSLE) as the loss function, which is given by

$$\begin{aligned} L_{\text{SMSLE}}(\boldsymbol{W}_t, \boldsymbol{W}_e, \lambda) &= \frac{1}{2}\left\| \log_{10}\left(\hat{\Phi}^{+}_{\text{LISTA}}\right) - \log_{10}\left(\Phi^{+}_{\text{DAS}}\right)\right\| \\ &\quad + \frac{1}{2}\left\| \log_{10}\left(\hat{\Phi}^{-}_{\text{LISTA}}\right) - \log_{10}\left(\Phi^{-}_{\text{DAS}}\right)\right\| \end{aligned} \tag{31}$$

where $\hat{\Phi}^{+}_{\text{LISTA}}$ and $\hat{\Phi}^{-}_{\text{LISTA}}$ are the positive and the negative parts of the network's prediction respectively, and $\hat{\Phi}^{+}_{\text{DAS}}$ and $\hat{\Phi}^{-}_{\text{DAS}}$ are the DAS target's positive and negative parts, respectively.

V. Evaluation Results

In this section, we evaluate our results both quantitatively and visually. For quantitative evaluation, we discuss its benefits compared to standard DAS and iterative techniques. For evaluating axial and lateral resolutions, full-width at half-maximum (FWHM) is calculated for a phantom scan and *in vivo* data.

For contrast evaluation, contrast-to-noise ratio (CNR) [54] is calculated. CNR is evaluated from two regions in each image, the cyst mimicking part and its background. The value is calculated after envelope detection and log compression and is given by

$$\text{CNR} = 20\log_{10}\left(\frac{|\mu_c - \mu_b|}{\sqrt{\sigma_c^2 + \sigma_b^2}}\right) \tag{32}$$

where μ_b, μ_c, σ_b, and σ_c denote the means and the STD of the cyst and the background, respectively.

The proposed method was evaluated using *in vivo* data from healthy volunteers for training and testing. The set for testing included tissue-mimicking phantoms Gammex 403GSLE and 404GSLE, as well as different body parts, such as bladder, kidney, and liver.

The acquisition was performed using the Verasonics Vantage 256 System, using the 64-element phased array transducer P4-2v. The frequency response of this probe is centered at 2.72 MHz, and a sampling rate of 10.8 MHz was used, leading to 1920 samples per image line. Two temporally subsampled datasets were generated from it, using the Xampling mechanism, and one dataset of both temporal and spatial dilution. In the first, 230 samples out of 1920 were used per image line, leading to an 8-fold reduction in data size. In the second dataset, only 130 samples were used, leading to a 15-fold reduction in the volume of the data. The third dataset was generated by diluting both time and space data based on the fractal array geometry, where a generator array $\mathbb{G} = \{0, 1\}$ with array order 4 was taken, leading to 15 elements, and 230 samples at each, resulting in a 36-fold total reduction.

Data from two healthy volunteers are used for training, and data from a third volunteer are used for testing. The training data consist of 72 frames of bladders, and each consists of 128 image lines of size 1920. The testing is applied to a wide range of organs and not solely to the bladder, to check the robustness of the proposed method. Targets for training were generated using standard DAS beamforming. DAS was chosen over MV beamforming [55] after empirically yielding better results.

The network was implemented using Keras based on Tensorflow backend. The training was done separately to each of the datasets, based on the Adam optimizer, reaching convergence in less than 120 epochs. The initial learning rate was $1e^{-}3$, and the learning rate was set to decrease each time the loss is not improved over few epochs, to make sure that the minimum is reached and not skipped due to high step size. The input vector

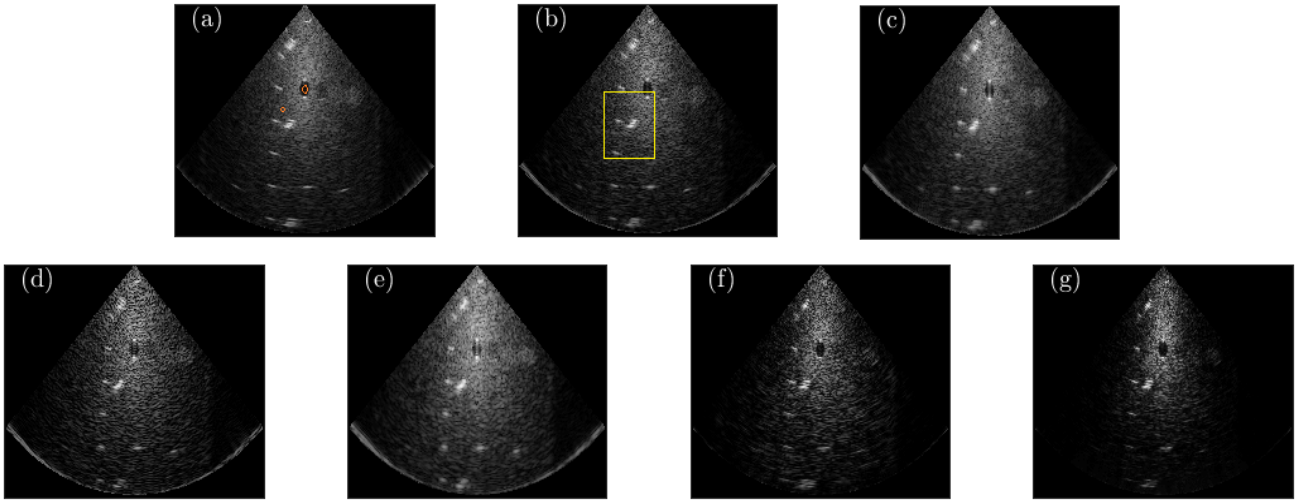

Fig. 5. Phantom images produced with (a) DAS–1920 samples, (b) FDBF + NESTA recovery—230 samples and eightfold reduction, (c) FDBF + LISTA recovery—230 samples and eightfold reduction, (d) FDBF + NESTA recovery—130 samples and 15-fold reduction, (e) FDBF + LISTA recovery—130 samples and 15-fold reduction, (f) CFCOBA + NESTA recovery—230 samples, 15 channels, and 36-fold reduction, (g) CFCOBA + LISTA recovery—230 samples, 15 channels, and 36-fold reduction. The regions used for calculating CNR are marked with orange circles in (a). The top region denotes the cyst and the bottom region denotes the background. In (b), the yellow rectangle marks the zoomed in region in Fig. 6.

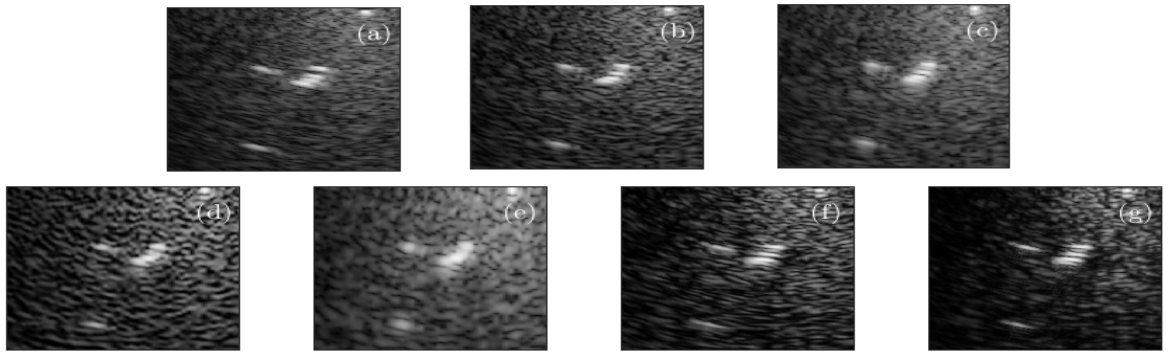

Fig. 6. Zoomed-in phantom images produced with (a) DAS—1920 samples, (b) FDBF + NESTA recovery—230 samples and eightfold reduction, (c) FDBF + LISTA recovery—230 samples and eightfold reduction, (d) FDBF + NESTA recovery—130 samples and 15-fold reduction, (e) FDBF + LISTA recovery—130 samples and 15-fold reduction, (f) CFCOBA + NESTA recovery—230 samples, 15 channels, and 36-fold reduction, and (g) CFCOBA + LISTA recovery—230 samples, 15 channels, and 36-fold reduction.

size to the network was 1920×1 as described previously. The initialization of the weights was based on the Glorot uniform initialization [56].

To demonstrate the performance of the proposed reconstruction method, we compare it quantitatively to DAS and NESTA, the iterative algorithm used in previous works.

Table I shows a quantitative evaluation of the lateral and axial resolutions of the compared reconstruction methods. The resolution is calculated from the *in vivo* scans and averaged over the three organs and for phantom scan data. The results were estimated by computing FWHM for each lateral and axial cut per frame and averaging the results. For 230 samples at 64 channels, LISTA recovery outperforms NESTA. The recovery for 130 samples using the full ULA resulted in slightly worse resolution than produced by NESTA, but NESTA caused high granularity in the resulting image. The evaluated resolution of the proposed method over the third dataset is much better than that produced by NESTA and DAS, and it is actually better than MV beamforming.

TABLE I
RESOLUTION EVALUATION

Beamforming technique		*in-vivo* data		Phantom data	
		axial	lateral	axial	lateral
DAS	1920 samples, 64 elements	3.66	5.01	3.2	4.5
MV	1920 samples, 64 elements	2.77	4.06	2.64	4.2
NESTA	230 samples, 64 elements	4.89	5.65	4.82	5.56
NESTA	130 samples, 64 elements	5.6	5.57	6	5.8
NESTA	230 samples, 15 elements	4.89	5.65	3.6	4.6
LISTA	230 samples, 64 elements	3.37	4.38	4.38	5.37
LISTA	130 samples, 64 elements	8.51	7.36	9.5	7.9
LISTA	230 samples, 15 elements	1.92	2.77	1.43	2.4

In Table II, the quantitative evaluation of CNR for a tissue-mimicking phantom scan is evaluated. As can be seen, in all cases, our method outperforms NESTA. In addition,

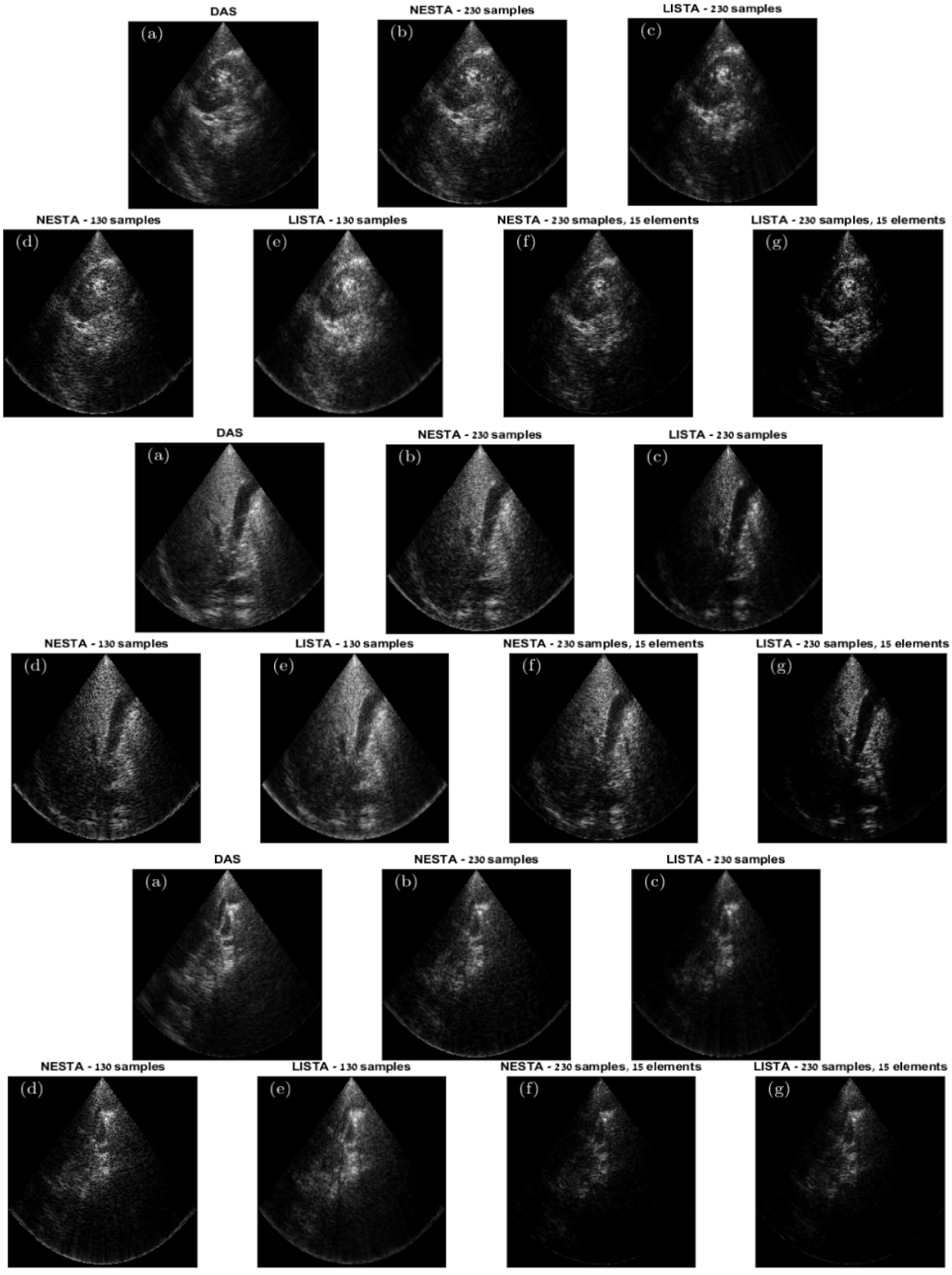

Fig. 7. *In vivo* images produced with (a) DAS—1920 samples, (b) FDBF + NESTA recovery—230 samples and 8-fold reduction, (c) FDBF + LISTA recovery—230 samples and 8-fold reduction, (d) FDBF + NESTA recovery—130 samples and 15-fold reduction, (e) FDBF + LISTA recovery—130 samples and 15-fold reduction, (f) CFCOBA + NESTA recovery—230 samples, 15 channels, and 36-fold reduction, and (g) CFCOBA + LISTA recovery—230 samples, 15 channels, and 36-fold reduction.

TABLE II
CNR EVALUATION

		CNR (dB)
DAS	1920 samples, 64 elements	8.35
NESTA	230 samples, 64 elements	5.9
NESTA	130 samples, 64 elements	1.9
NESTA	230 samples, 15 elements	6.1
LISTA	230 samples, 64 elements	8.7
LISTA	130 samples, 64 elements	7.7
LISTA	130 samples, 15 elements	7.9

the second dataset reconstructed using NESTA leads to the lowest CNR, which indicates that contrast between noises of the signal of interest almost does not exist. These results indicate that our method is much more robust to a reduction both in sampling rate and in the number of acquiring channels than NESTA reconstruction. The resulting images from the phantom scans are presented in Fig. 5. For visually better examining the results, zoomed-in images are given in Fig. 6. While in (a)–(c) all three options depict approximately the same visual quality. Comparing (d) and (e), one may claim that (e) exhibits worse resolution with regard to (d), and however, there are two major disadvantages in the latter. First, due to the relatively high reduction in the number of samples per image line, the recovered image shows high granularity. The second drawback is low contrast, as calculated and as can be seen, making it difficult to notice actual reflectors that are easily visible in (e). Finally, in (f)–(g), both contrast and resolution are visually proved to be much better in (g).

Images produced from *in vivo* scan data are presented in Fig. 7. It can be seen that the proposed method, over the first dataset of 230 samples, yields the results that are visually better than those produced with DAS or NESTA. Moreover, in the case of 130 samples per image line, the proposed method outperforms NESTA. The image reconstructed using NESTA suffers from high granularity that results from the high dilution factor and can be considered as an artifact. In addition, testing the proposed method over the spatially and temporally diluted data presents better image contrast in comparison to NESTA and DAS, a result that agrees with previously calculated CNR.

In addition to the visual and quantitative improvement of our method, we also managed to ease the computational load compared to the iterative approach used in previous works. While NESTA is an iterative process, our method is based on a pretrained network that mimics ISTA, a relatively simple algorithm, with weights that are saved before reconstruction, leading to a real-time applicable recovery technique. Moreover, our method offers a network of only 30 layers, while NESTA iterates over 60 times until convergence and produces poorer results when compared to the proposed method. To emphasize the total reduction in computation complexity of our technique, we estimate the total operations in each of the methods. The NESTA algorithm includes several matrix multiplications in each iteration, leading to an overall complexity of over $O(n^2)$, for a given input vector of size n and several lower complexity operations. In LISTA, each layer includes two convolutions with a kernel of size 5×1 and nonlinear operation, leading to a computational complexity of $O(n)$. This, combined with the fact that each iteration of NESTA consists of more operations than the iterative version of ISTA, leads to a highly efficient implementation of US image reconstruction from sub-Nyquist samples. These results validate that recovery of the beamformed signal using LISTA is much more efficient without degrading image quality compared to the iterative approach of NESTA.

Although we focus on line-based beamforming in this work, the proposed mechanism can be extended to other US imaging techniques, such as plane-wave imaging. In [57], Chernyakova *et al.* formulated the recovery from subsampled signals acquired via plane-wave US imaging. Using their inverse problem formulation, one can recover the beamformed signal using a similar architecture as the one described in this article.

VI. CONCLUSION

In this article, we presented a deep unfolding-based reconstruction method, for high-quality images, from temporally and spatially subsampled channel data based on the Xampling mechanism. We extended the proposed methods in [11] and [22] and showed that based on the FRI model, a US image can be recovered using a deep network that mimics the iterations of ISTA.

Our network consists of only 30 folds, leading to an efficient implementation, using a loss function that better fits the studied field of US than traditionally used MSE. After training the network on a dataset of only *in vivo* bladders, we validated our approach and tested it on phantom channel data and several *in vivo* scans different from bladders, demonstrating the high versatility of our network. The proposed method produced high-quality results both in quantitative and visual terms while using 36 times less data. The output of our technique yields better image resolution and CNR compared to NESTA, proving the robustness of our method and resulting in visually high-quality B-mode images.

These results prove that our method can be easily plugged into previously suggested schemes, for recovery from sub-Nyquist sampling and sparse arrays, while improving image quality and using an efficient deep neural network instead of iterative algorithms.

ACKNOWLEDGMENT

The authors would like to thank Dr. Israel Aharony for voluntarily performing the scans that provided the data for testing and evaluating the proposed method.

REFERENCES

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
[2] K. E. Thomenius, "Evolution of ultrasound beamformers," in *Proc. IEEE Ultrason. Symp.*, vol. 2, Nov. 1996, pp. 1615–1622.
[3] B. D. Steinberg, "Digital beamforming in ultrasound," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 39, no. 6, pp. 716–721, Nov. 1992.
[4] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.
[5] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[6] K. Ranganathan and W. F. Walker, "A novel beamformer design method for medical ultrasound. Part I: Theory," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 50, no. 1, pp. 15–24, Jan. 2003.

[7] T. Szasz, A. Basarab, and D. Kouame, "Beamforming through regularized inverse problems in ultrasound medical imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 12, pp. 2031–2044, Dec. 2016.

[8] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[9] G. R. Lockwood, J. R. Talman, and S. S. Brunke, "Real-time 3-D ultrasound imaging using sparse synthetic aperture beamforming," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 45, no. 4, pp. 980–988, Jul. 1998.

[10] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4643–4657, Sep. 2012.

[11] T. Chernyakova and Y. C. Eldar, "Fourier-domain beamforming: The path to compressed ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, no. 8, pp. 1252–1267, Aug. 2014.

[12] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[13] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.

[14] E. Roux *et al.*, "Validation of optimal 2D sparse arrays in focused mode: Phantom experiments," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[15] J. T. Yen, J. P. Steinberg, and S. W. Smith, "Sparse 2-D array design for real time rectilinear volumetric imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 47, no. 1, pp. 93–110, Jan. 2000.

[16] A. Austeng and S. Holm, "Sparse 2-D arrays for 3-D phased array imaging–design methods," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, no. 8, pp. 1073–1086, Aug. 2002.

[17] S. S. Brunke and G. R. Lockwood, "Broad-bandwidth radiation patterns of sparse two-dimensional Vernier arrays," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 44, no. 5, pp. 1101–1109, Sep. 1997.

[18] R. Cohen and Y. C. Eldar, "Sparse convolutional beamforming for ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 12, pp. 2390–2406, Dec. 2018.

[19] R. Cohen and Y. C. Eldar, "Optimized sparse array design based on the sum coarray," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3340–3343.

[20] R. Cohen and Y. C. Eldar, "Sparse array design via fractal geometries," 2020, *arXiv:2001.01217*. [Online]. Available: http://arxiv.org/abs/2001.01217

[21] C.-L. Liu and P. P. Vaidyanathan, "Maximally economic sparse arrays and cantor arrays," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2017, pp. 1–5.

[22] A. Mamistvalov and Y. C. Eldar, "Compressed Fourier-domain convolutional beamforming for wireless ultrasound imaging," 2020, *arXiv:2010.13171*. [Online]. Available: http://arxiv.org/abs/2010.13171

[23] A. Mamistvalov and Y. C. Eldar, "Sparse convolutional beamforming for wireless ultrasound," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 9254–9258.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[25] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[26] R. J. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, no. 1, pp. 11–29, Jan. 2020.

[27] R. Zhuang and J. Chen, "Deep learning based minimum variance beamforming for ultrasound imaging," in *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*. Cham, Switzerland: Springer, 2019, pp. 83–91.

[28] B. Luijten *et al.*, "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3967–3978, Dec. 2020.

[29] N. Kessler and Y. C. Eldar, "Deep-learning based adaptive ultrasound imaging from sub-nyquist channel data," 2020, *arXiv:2008.02628*. [Online]. Available: http://arxiv.org/abs/2008.02628

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[31] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program.*, vol. 103, no. 1, pp. 127–152, May 2005.

[32] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1–39, Jan. 2011.

[33] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.

[34] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 399–406.

[35] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[36] J. A. Jensen, "Linear description of ultrasound imaging systems: Notes for the international summer school on advanced ultrasound imaging at the technical university of denmark," Tech. Univ. Denmark, Lyngby, Denmark, Tech. Rep., Jul. 1999.

[37] G. DeMuth, "Frequency domain beamforming techniques," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, May 1977, pp. 713–715,

[38] E. Baransky, G. Itzhak, N. Wagner, I. Shmuel, E. Shoshan, and Y. Eldar, "Sub-Nyquist radar prototype: Hardware and algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 809–822, Apr. 2014.

[39] R. T. Hoctor and S. A. Kassam, "The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging," *Proc. IEEE*, vol. 78, no. 4, pp. 735–752, Apr. 1990.

[40] C. Puente-Baliarda and R. Pous, "Fractal design of multiband and low side-lobe arrays," *IEEE Trans. Antennas Propag.*, vol. 44, no. 5, p. 730, May 1996.

[41] D. H. Werner, R. L. Haupt, and P. L. Werner, "Fractal antenna engineering: The theory and design of fractal antenna arrays," *IEEE Antennas Propag. Mag.*, vol. 41, no. 5, pp. 37–58, Oct. 1999.

[42] D. H. Wqrner and S. Ganguly, "An overview of fractal antenna engineering research," *IEEE Antennas Propag. Mag.*, vol. 45, no. 1, pp. 38–57, Feb. 2003.

[43] J. Feder, *Fractals*. Cham, Switzerland: Springer, 2013.

[44] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*. Hoboken, NJ, USA: Wiley, 2004.

[45] T. Michaeli and Y. C. Eldar, "Xampling at the rate of innovation," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1121–1133, Mar. 2012.

[46] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET Circuits, Devices Syst.*, vol. 5, no. 1, pp. 8–20, Jan. 2011.

[47] M. Mishali, Y. C. Eldar, and A. J. Elron, "Xampling: Signal acquisition and processing in union of subspaces," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4719–4734, Oct. 2011.

[48] P. Stoica, *Introduction to Spectral Analysis*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[49] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.

[50] E. Cands and J. Romberg, "L_1-magic: Recovery of sparse signals via complex programming," California Inst. Technol., Pasadena, CA, USA, Tech. Rep., 2007. [Online]. Available: https://www.acm.caltech.edu/l1magic/downloads/l1magic

[51] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing," Rice Univ., Houston, TX, USA, Tech. Rep. CAAM TR07-07, 2007, p. 44, vol. 43.

[52] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*. Cham, Switzerland: Springer, 2012, pp. 9–48.

[53] G. Dardikman-Yoffe and Y. C. Eldar, "Learned SPARCOM: Unfolded deep super-resolution microscopy," 2020, *arXiv:2004.09270*. [Online]. Available: http://arxiv.org/abs/2004.09270

[54] A. Rodriguez-Molares *et al.*, "The generalized contrast-to-noise ratio: A formal definition for lesion detectability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020.

[55] J. F. Synnevag, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 54, no. 8, pp. 1606–1613, Aug. 2007.

[56] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[57] T. Chernyakova *et al.*, "Fourier-domain beamforming and structure-based reconstruction for plane-wave imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 10, pp. 1810–1821, Oct. 2018.

Alon Mamistvalov received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 2017. He is currently pursuing the M.Sc. degree in mathematics and computer science with the Weizmann Institute of Science, Rehovot, Israel.

His research interests include signal processing, medical imaging, advanced signal processing methods for ultrasonic imaging, and deep learning for ultrasound imaging.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel Aviv University (TAU), Tel Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002.

She was a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair in Engineering. She was a Visiting Professor at Stanford University, Stanford, CA, USA. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She is also a Visiting Professor with MIT, a Visiting Scientist with the Broad Institute, Cambridge, and an Adjunct Professor with Duke University, Durham, NC, USA. Her research interests include the broad areas of statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology and optics.

Dr. Eldar is a member of the Israel Academy of Sciences and Humanities (elected in 2017) and a fellow of EURASIP. She has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award in 2013, the IEEE/AESS Fred Nathanson Memorial Radar Award in 2014, and the IEEE Kiyo Tomiyasu Award in 2016. She was a Horev Fellow of the Leaders in Science and Technology Program at the Technion and a fellow of Alon. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel and David Jacknow Award for Excellence in Teaching, and the Technion's Award for Excellence in Teaching (two times). She received several best paper awards and best demo awards together with her research students and colleagues, including the SIAM Outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award, and the IET Circuits, Devices, and Systems Premium Award. She was selected as one of the 50 most influential women in Israel. She was the co-chair and the technical co-chair of several international conferences and workshops. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor-in-Chief of *Foundations and Trends in Signal Processing* and a member of the IEEE Sensor Array and Multichannel Technical Committee. She serves on several other IEEE committees. In the past, she was a Distinguished Lecturer of the IEEE Signal Processing Society and a member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing Technical Committees. She served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*.