Learning Structured Compressed Sensing with Automatic Resource Allocation

Han Wang[*,‡], Eduardo Pérez[*,‡], Iris A. M. Huijben[†], Hans van Gorp[†], Ruud van Sloun[†], Florian Römer[*,‡]

[*]*Applied AI Signal Processing and Data Analysis, Fraunhofer Institute for Nondestructive Testing IZFP*, Germany
[†]*Department of Electrical Engineering, Eindhoven University of Technology*, The Netherlands
[‡]*Electronic Measurements and Signal Processing, Ilmenau University of Technology*, Germany
Email: han.wang@izfp.fraunhofer.de

Abstract—**Multidimensional data acquisition often requires extensive time and poses significant challenges for hardware and software regarding data storage and processing. Rather than designing a single compression matrix as in conventional compressed sensing, structured compressed sensing yields dimension-specific compression matrices, reducing the number of optimizable parameters. Recent advances in machine learning (ML) have enabled task-based supervised learning of subsampling matrices, albeit at the expense of complex downstream models. Additionally, the sampling resource allocation across dimensions is often determined in advance through heuristics. To address these challenges, we introduce Structured COmpressed Sensing with Automatic Resource Allocation (SCOSARA) with an information theory-based unsupervised learning strategy. SCOSARA adaptively distributes samples across sampling dimensions while maximizing Fisher information content. Using ultrasound localization as a case study, we compare SCOSARA to state-of-the-art ML-based and greedy search algorithms. Simulation results demonstrate that SCOSARA can produce high-quality subsampling matrices that achieve lower Cramér Rao Bound values than the baselines. In addition, SCOSARA outperforms other ML-based algorithms in terms of the number of trainable parameters, computational complexity, and memory requirements while automatically choosing the number of samples per axis.**

Index Terms—**Unsupervised Learning, Information Theory, Compressed Sensing, Subsampling.**

I. INTRODUCTION

Many practical problems involve multidimensional data with several axes that represent different domains [1], e.g. space, time, Doppler shift, color channels, azimuth and elevation angles, etc. [2], [3], [4]. Such data is information-rich, but quickly yields large data volumes and may involve time-consuming measurement procedures.

Compressed sensing (CS) [5] addresses the data volume and, if done judiciously, reduces the measurement time by measuring only a small number of linear projections of the data. The process of collecting these projections is represented through a rectangular matrix referred to as a compression matrix. Such compression matrices are classically designed as random matrices [6]. Such designs are optimal, but only asymptotically for large matrices or in expectation, and are also difficult to implement in hardware. This motivates the systematic design of optimal subsampling matrices, which is the goal of this paper.

Subsampling matrices are a particular case of compression matrices with *one-hot* rows, whereby only a subset of the original samples are measured [7], [8], [9], [10]. Although this results in loss of information when compared to random linear projections, the result is a CS scheme that is easily implemented through omission, planning, or reprogramming of the original measurement procedure without the need for additional hardware [11]. However, the design of subsampling matrices involves two combinatorial optimization subproblems: (I) "choosing k out of n items", and (II) "distributing k items into q bins". The present work addresses these problems on several fronts.

Problem (I) naturally arises when subsampling data. Advances in ML architectures and task-based learning have made it possible to use ML to learn a subsampling matrix[8], [12], [13], [14]. Due to the typically large volume of multidimensional data, task-based ML can be time-intensive or prohibitively resource-intensive. For example, a commonly-used deep unrolled neural network LISTA [15], [16] grows linearly with the number of unrolled iterations and also scales with the product of the dimension sizes of the input data (e.g. transmitters $\times$ receivers $\times$ time domain samples $\times$ size of the region of interest).

Problem (II) appears when structured CS of multidimensional data is applied [12], where each axis is compressed separately, turning a large subsampling problem into several smaller ones. More explicitly, the number of design parameters is reduced from the product of data dimension lengths to the sum of the data dimension lengths. However, the number of samples must now be distributed among the dimensions of the data. This resource allocation problem is often addressed through heuristics [8], [14], [12].

We alleviate the large-scale nature of problem (I) by replacing task-based learning with Fisher information maximization and exploiting the structure in the data. The Fisher Information Matrix (FIM) provides an alternative to task-based optimization [17], as commonly used in multistatic localization [18]. The FIM and its inverse, the Cramér-Rao Bound (CRB), are known to accurately describe the performance of CS methods when the signal-to-noise ratio is high and quantization error is low [19], [20], [21], [22], [23]. This makes the unsupervised maximization of the trace of the FIM an attractive alternative to task-based approaches in optimizing subsampling matrices. We address problem (II) by formulating the structured subsampling problem

so that the number of samples per axis is learned automatically during training.

In this work, we propose a framework for the design of structured subsampling matrices by means of ML-based unsupervised maximization of the trace of the FIM. The framework, dubbed Structured COmpressed Sensing with Automatic Resource Allocation (SCOSARA), automatically distributes the samples among the axes of multidimensional data.

II. STRUCTURED SUBSAMPLING

Data acquisition is performed across q different dimensions, and the number of samples in the ith dimension is N_i (with $i = 1, 2, \ldots, q$). The q-dimensional data array can be expressed as $\mathbf{Y} \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_q}$, which can be vectorized as $\mathbf{y} \in \mathbb{C}^{N_\Pi}$, with $N_\Pi = \Pi_{i=1}^q N_i$ the total number of samples. The sampling process is commonly represented as a linear transformation from the to-be-estimated signal $\mathbf{x} \in \mathbb{C}^{N_s}$ to this measurement:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathbb{C}^{N_\Pi}$ represents circularly-symmetric Additive White Gaussian Noise (AWGN) following $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{A} \in \mathbb{C}^{N_\Pi \times N_s}$ is the signal model.

We introduce a subsampling matrix $\mathbf{S} \in \{0, 1\}^{M_\Pi \times N_\Pi}$ that selects M_Π out of N_Π data samples thanks to its row-wise one-hot structure, resulting in subsampled measurement $\mathbf{y}_s \in \mathbb{C}^{M_\Pi}$ through

$$\mathbf{y}_s = \mathbf{S}\mathbf{y}. \tag{2}$$

Structured COmpressed Sensing (SCOS) compresses data along each ith dimension separately. Structured subsampling – the focus of this work – is an instance of SCOS, where each ith axis is *subsampled* separately. To this end, we introduce subsampling matrices $\mathbf{S}_i \in \{0, 1\}^{M_i \times N_i}$, with $M_i < N_i$. This yields the following formulation via the Kronecker product:

$$\mathbf{y}_s = (\mathbf{S}_1 \otimes \mathbf{S}_2 \otimes \cdots \otimes \mathbf{S}_q)\mathbf{y} = \mathbf{S}\mathbf{y}. \tag{3}$$

This reduces the problem of choosing M_Π out of N_Π samples into q problems where M_i out of N_i must be chosen. In general, $(M_\Sigma = \sum_{i=1}^q M_i) < M_\Pi$ and so the structured subsampling problem is simpler than the original one. However, the single hyperparameter M_Π is replaced with q hyperparameters M_i. In the next section, we elaborate on how a sampling budget is distributed across the axis, i.e. how we set all M_i in SCOS.

III. AUTOMATIC RESOURCE ALLOCATION

Given a total sampling budget $M_\Sigma = \sum_{i=1}^q M_i$ and the number of dimensions q in the data, the samples can be allocated in $\binom{M_\Sigma - 1}{q - 1}$ number of ways, resulting in a combinatorial resource allocation problem. Instead of heuristically specifying the active elements M_i per axis, we automatically learn how to distribute the sampling budget. We achieve this by modifying the ML-based subsampling approach proposed in [24] so that it automatically performs resource allocation while learning which samples to preserve.

In [24], the design of subsampling matrices is treated as sampling without replacement from a categorical distribution.

The log probabilities, or logits, of the categorical distribution are learned in a gradient-based fashion by employing the straight-through Gumbel estimator [25], [26]. In the forward direction, the Gumbel-max trick is employed, i.e. Gumbel noise Gumbel$(0, 1)$ is added to the logits, followed by the application of argmax to sample from the distribution over elements in the subsampling matrix. During backpropagation, 'soft samples' are drawn from the Gumbel-softmax distribution by relaxing the non-differentiable argmax function with a softmax, allowing the usage of backpropagation for the optimization of the logits.

Based on this procedure, the subsampling matrices $\mathbf{S}_i$ can be learned based only on the total budget M_Σ, without specifying each M_i separately. To this end, we introduce a single vector $\boldsymbol{\phi}$ containing $N_\Sigma = \sum_{i=1}^q N_i$ logits. The entries of $\boldsymbol{\phi}$ are ordered in the same order the dimensions of $\mathbf{Y}$ are vectorized in. The Gumbel-softmax trick is then used to obtain a differentiable approximation of the process of sampling M_Σ items without replacement out of the total N_Σ. To do so, Gumbel noise $\mathbf{g} \in \mathbb{R}^{N_\Sigma}$, $\mathbf{g} \sim \text{Gumbel}(0,1)$ is added to the logits $\boldsymbol{\phi}$, yielding the perturbed logits vector $\tilde{\boldsymbol{\phi}} \in \mathbb{R}^{N_\Sigma}$. Next, the softmax function is applied to $\tilde{\boldsymbol{\phi}}$. To obtain M_Σ samples without replacement, the largest entry of $\tilde{\boldsymbol{\phi}}$ is replaced by $-\infty$ (i.e. the probability of choosing the same entry again is set to 0), and then softmax is applied again [24]. This process is repeated until an auxiliary matrix $\tilde{\boldsymbol{\Phi}}$ of size $M_\Sigma \times N_\Sigma$ is obtained. The same effect can be achieved through the repetition of the perturbed logits vector $\tilde{\boldsymbol{\phi}}$ M_Σ times, followed by the addition of a *masking matrix* $\mathbf{W}$ of size $M_\Sigma \times N_\Sigma$ whose entries are taken from $\{-\infty, 0\}$.

Since the entries of the logits vector $\boldsymbol{\phi}$ are ordered according to the vectorization of $\mathbf{Y}$, the auxiliary matrix $\tilde{\boldsymbol{\Phi}}$ exhibits the structure

$$\tilde{\boldsymbol{\Phi}} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{N}_{1,2} & \cdots & \mathbf{N}_{1,q} \\ \mathbf{N}_{2,1} & \mathbf{S}_2 & \cdots & \mathbf{N}_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{N}_{q,1} & \mathbf{N}_{q,2} & \cdots & \mathbf{S}_q \end{bmatrix}. \tag{4}$$

The subsampling matrices $\mathbf{S}_i \in \mathbb{R}^{M_i \times N_i}$ appear along the diagonal blocks of $\tilde{\boldsymbol{\Phi}}$, whereas the off-diagonal blocks $\mathbf{N}_{i,j} \in \mathbb{R}^{M_i \times N_j}, 1 \leq i, j \leq q$, contain only nuisance terms that are nonzero due to the usage of softmax.

Two crucial observations must be made regarding $\tilde{\boldsymbol{\Phi}}$. First, the sizes of the diagonal blocks are unknown because the M_i are unknown. Second, since the entries of $\tilde{\boldsymbol{\phi}}$ change when the noise $\mathbf{g}$ is added and when $\boldsymbol{\phi}$ is modified through backpropagation, the matrix $\tilde{\boldsymbol{\Phi}}$ is not available in practice. Instead, the order the samples are drawn in, is unknown and only $\boldsymbol{\Pi}\tilde{\boldsymbol{\Phi}}$ can be obtained, where $\boldsymbol{\Pi} \in \mathbb{R}^{M_\Sigma \times M_\Sigma}$ is an unknown permutation matrix. Both problems are addressed by computing $\tilde{\boldsymbol{\Psi}} = (\boldsymbol{\Pi}\tilde{\boldsymbol{\Phi}})^{\mathrm{T}}(\boldsymbol{\Pi}\tilde{\boldsymbol{\Phi}}) = \tilde{\boldsymbol{\Phi}}^{\mathrm{T}}\tilde{\boldsymbol{\Phi}} \in \mathbb{R}^{N_\Sigma \times N_\Sigma}$, yielding a matrix of the general form

$$\tilde{\boldsymbol{\Psi}} = \begin{bmatrix} \mathbf{S}_1^{\mathrm{T}}\mathbf{S}_1 + \mathbf{N}_1' & \mathbf{N}_{1,2}' & \cdots & \mathbf{N}_{1,q}' \\ \mathbf{N}_{2,1}' & \mathbf{S}_2^{\mathrm{T}}\mathbf{S}_2 + \mathbf{N}_2' & \cdots & \mathbf{N}_{2,q}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{N}_{q,1}' & \mathbf{N}_{q,2}' & \cdots & \mathbf{S}_q^{\mathrm{T}}\mathbf{S}_q + \mathbf{N}_q' \end{bmatrix}. \tag{5}$$
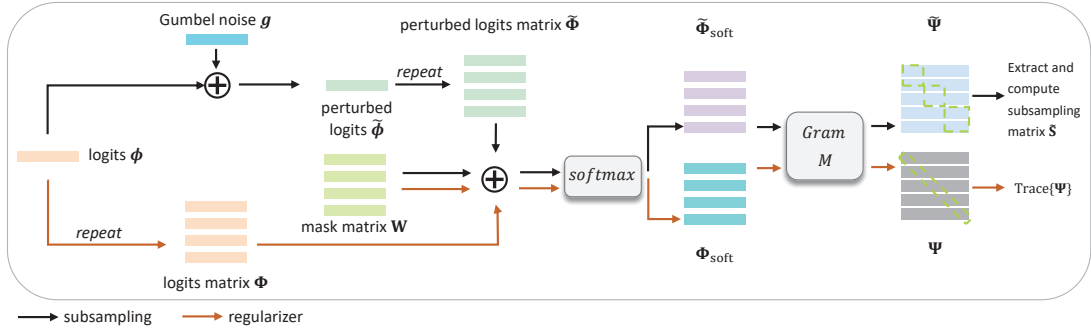
Fig. 1: Flowchart of automatic resource allocation in SCOSARA. There are two parallel ways to obtain the subsampling matrices and the regularizer: the top path perturbs the logits with Gumbel noise to approximate the process of sampling from a categorical distribution, while the lower path directly uses the logits to produce the regularizer.

The block matrices along the diagonal of $\tilde{\mathbf{\Psi}}$, i.e. $\tilde{\mathbf{S}}_i = \mathbf{S}_i^{\mathrm{T}}\mathbf{S}_i + \mathbf{N}_i' \in \mathbb{R}^{N_i \times N_i}$, can then be used to construct

$$\tilde{\mathbf{S}} = \tilde{\mathbf{S}}_1 \otimes \tilde{\mathbf{S}}_2 \otimes \cdots \otimes \tilde{\mathbf{S}}_q \in \mathbb{R}^{N_\Pi \times N_\Pi}. \qquad (6)$$

The structure of $\tilde{\mathbf{\Psi}}$ also motivates the following observation. Consider the unperturbed logits ϕ and apply the Gumbel-softmax trick without the noise $\mathbf{g}$ so as to obtain a new auxiliary matrix $\mathbf{\Phi}$, and from it, $\mathbf{\Psi} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$. Since the rows of $\mathbf{\Phi}$ are non-negative and sum to one due to the softmax function, maximizing $\mathrm{Trace}(\mathbf{\Psi})$ (equivalently, $\|\mathbf{\Phi}\|_{\mathrm{F}}^2$) reduces the magnitude of the off-diagonal elements of $\tilde{\mathbf{\Psi}}$, thereby aligning the forward and backward passes of the straight-through estimator. This term can thus be used as a regularizer when optimizing the logits.

The overall procedure described throughout this section is illustrated in Figure 1. The upper branch of the figure uses Gumbel noise $\mathbf{g}$ to construct the matrix $\tilde{\mathbf{\Psi}}$ used for sampling, while the lower branch is noiseless and yields the regularizing matrix $\mathbf{\Psi}$.

IV. Optimization Target

Replacing task-based optimization because of its time and resource-intensive nature, we take FIM and CRB into account. The CRB is a lower bound for the variance of unbiased estimators [17]. However, CS methods are biased, the sources of the bias being amplitude errors [21], parameter discretization errors [22], and the enforcing of sparse support [23]. In spite of this, the cited works illustrate that the CRB correctly describes the behavior of practical CS estimators when the signal-to-noise ratio is high and the sampling grid is fine enough.

The CRB is computed as the inverse of FIM, which quantifies the amount of information that an observable random variable carries about the parameters of its distribution. The FIM is also closely related to the Restricted Isometry Property (RIP) [19]. The inter-column coherence of a model matrix is inversely proportional to the eigenvalues of the FIM [23], [20]. Similarly, the Restricted Isometry Constant (RIC) involved in the RIP is smaller when the eigenvalues of the FIM are large.

Therefore, instead of the downstream task such as directly computing the signal recovery, we try to maximize the trace of FIM for two reasons. First, it largely reduces the network and

training complexity since it is estimator-agnostic. Second, the A-optimality criterion, which refers to the minimization of the trace of the CRB, is a commonly employed optimization target in the design of experiments. However, in the present setting, directly using the CRB would require multiple matrix inversions during training. The maximization of the trace of FIM can be seen as a heuristic for the CRB, as it can be formulated in terms of the harmonic mean of the eigenvalues of the CRB matrix and is therefore closely related to A-optimality.

In the derivation of the cost function, we let $\mathbf{Ax}$ obey a differentiable parametric model $\mathbf{b}$ with a real-valued parameter vector $\boldsymbol{\xi}$, so that the subsampled data $\mathbf{y}_{\mathrm{s}}$ can be reformulated as:

$$\mathbf{y}_{\mathrm{s}} = \mathbf{S}(\mathbf{Ax} + \mathbf{n}) = \mathbf{S}(\mathbf{b}(\boldsymbol{\xi}) + \mathbf{n}) = \boldsymbol{\mu} + \tilde{\mathbf{n}}, \qquad (7)$$

where $\mathbf{b}$ denotes the noiseless fully sampled data and $\boldsymbol{\mu}$ corresponds to the noiseless compressed data. The noise $\mathbf{n}$ is circularly-symmetric AWGN characterized by the distribution $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{\Sigma} = \sigma^2\mathbf{I}$. Instead, since we can only compute $\tilde{\mathbf{S}}$ from SCOSARA, following the notation in (6), we rewrite this as:

$$\mathbf{y}_{\mathrm{s}} = \tilde{\mathbf{S}}(\mathbf{b}(\boldsymbol{\xi}) + \mathbf{n}) = \tilde{\boldsymbol{\mu}} + \tilde{\mathbf{n}}, \qquad (8)$$

where the covariance of $\tilde{\mathbf{n}}$ is approximately given by $\sigma^2\mathbf{SS}^{\mathrm{T}}$. Based on the Slepian-Bangs formalism, the FIM can be written as

$$\mathcal{J} \approx \frac{2}{\sigma^2}\mathfrak{R}\left\{\left(\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\xi}}\right)^{\mathrm{H}}\left(\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\xi}}\right)\right\} = \frac{2}{\sigma^2}\mathfrak{R}\left\{\boldsymbol{J}_b^{\mathrm{H}}\tilde{\mathbf{S}}^{\mathrm{T}}\tilde{\mathbf{S}}\boldsymbol{J}_b\right\} \quad (9)$$

where $(\cdot)^{\mathrm{H}}$ is the conjugate transpose, $\mathfrak{R}$ refers to extracting the real-valued part and $\boldsymbol{J}_b$ stands for the Jacobian matrix of data $\mathbf{b}$ with respect to parameters $\boldsymbol{\xi}$. Note that in the case when multiple targets coexist, the CRB of a single target can also adequately describe the estimation problem when they are well-spaced. In summary, the FIM-driven SCOSARA aims to maximize Fisher information, resulting in the following cost function:

$$\mathcal{L} = \underset{\phi}{\mathrm{argmin}} \; -\mathrm{Trace}(\mathcal{J}) - \mathrm{Trace}(\mathbf{\Psi}). \qquad (10)$$

V. Evaluation

As an illustrative example of multidimensional data acquisition, we detail the application of the SCOS within the context of

TABLE I: Model parameters summary

Parameter	N_{T}	N_{R}	N_{F}	$N_{\sum}$	N_{Π}	$M_{\sum}$
Value	64	64	113	241	462848	$30 \sim 220$

multichannel ultrasound localization. The simulation scenarios and experimental settings are similar to [12], but we use a larger data model whose parameters are summarized in TABLE I. The task is to localize a single scatterer in the Region of Interest (ROI) whose parameter vector can be formulated by $\boldsymbol{\xi} = [x, z, a, \varphi]$, which corresponds to its two-dimensional coordinates, reflectivity coefficient, and phase. The dataset is generated by assuming a single scatterer with a varying reflectivity coefficient is randomly located in a given ROI.

We design/learn three subsampling matrices for reducing transmitters, receivers, and Fourier coefficients, denoted $\mathbf{S}_{\mathrm{T}} \in \mathbb{R}^{M_{\mathrm{T}} \times N_{\mathrm{T}}}$, $\mathbf{S}_{\mathrm{R}} \in \mathbb{R}^{M_{\mathrm{R}} \times N_{\mathrm{R}}}$, and $\mathbf{S}_{\mathrm{F}} \in \mathbb{R}^{M_{\mathrm{F}} \times N_{\mathrm{F}}}$. They compress all three axes while retaining as much information about scatterer locations as possible. To demonstrate the performance of **SCOSARA**, we compare it against **Uniform Compression** and other SOTA algorithms, including ML-based methods **DPS-topK** [8] and **J-DPS** [12], and a **Greedy Search** algorithm. To provide comprehensive and fair comparisons, we consider the following points in the evaluation:

- We trained SCOSARA first and then used its resource allocation scheme for other algorithms for two reasons: the baseline methods provide no resource allocation scheme, and the compression factor M_{Π}/N_{Π} should be consistent.
- We test 20 values of $M_{\sum}$ from 30 to 220 at intervals of 10, each has a distinct compression factor.

Baselines: The Uniform Compression scheme implies the equally spaced allocation of M_{T} choices among N_{T}, and so forth. The DPS-TopK refers to optimizing a vector of logits $\phi_{\Pi} \in \mathbb{R}^{N_{\Pi}}$ and using the Gumbel-topK trick in the forward pass to select $M_{\sum}$ elements set to 1. The J-DPS divides the long vector into three smaller dimension-specific ones. The Greedy Search algorithm iteratively discards the entry of ϕ which reduces $\mathrm{Trace}(\mathcal{J})$ the least until the desired compression factor is achieved.

Results: We used an NVIDIA A100 GPU node to run all experiments, the DPS-TopK failed to operate due to computational constraints because it optimizes a large logits vector ϕ_{Π}, which requires the generation of large matrices of shape $(M_{\Pi} \times N_{\Pi})$ in the computational graph. After running other algorithms given all the values of $M_{\sum}$, at each compression factor, we apply the four SCOS schemes to the evaluation dataset and compute the CRB values based on their compressed data. The resulting four curves are illustrated in Fig. 2, showing that SCOSARA outperforms other methods, achieving the lowest CRB across all compression factors.

Furthermore, we also evaluate the recovery performance in the multi-scatterer case. Knowing the distance of two pixels is approximately equal to half the wavelength in the simulation, we define three pairs of scatterers whose distances are two, three, and four pixels, respectively. We applied the four SCOS schemes given $M_{\sum} = 120$ along with the 2000-iteration complex FISTA, obtaining the reconstructed images. We computed the MSE
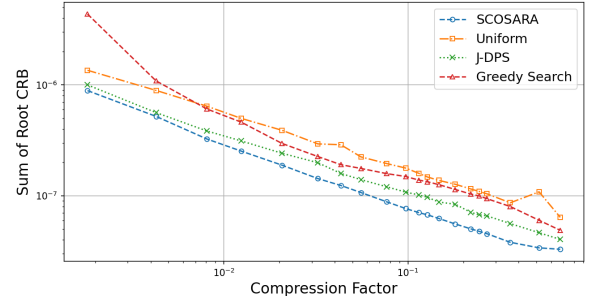

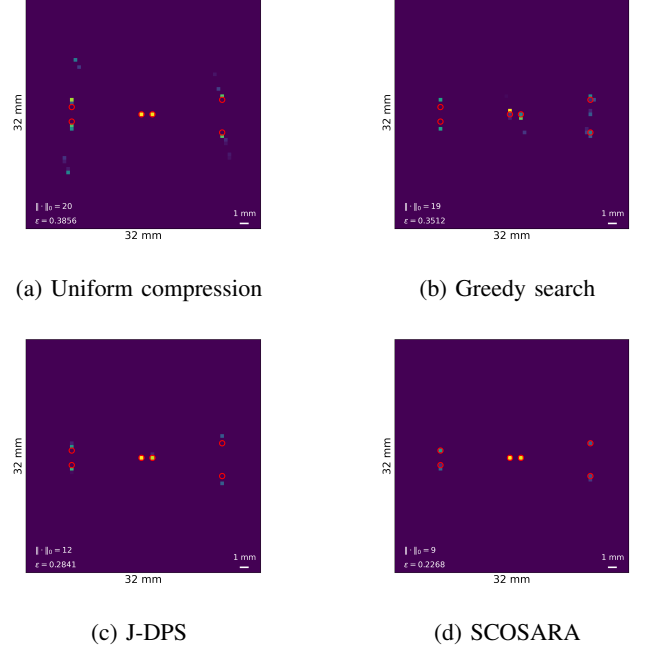Fig. 2: CRB as a function of the compression factor (higher factor denotes more selected samples).

(a) Uniform compression

(b) Greedy search

(c) J-DPS

(d) SCOSARA

Fig. 3: Reconstructed images comparison in the multi-scatterer cases when $M_{\sum} = 120$. The ground truth scatterers are represented by red circles.

ϵ and the number of nonzero elements $\| \cdot \|_0$ for quantitative comparison. As illustrated in Fig. 3, SCOSARA outperforms with respect to image quality and localization accuracy.

VI. CONCLUSION

In this work, we introduce the concept of FIM-driven SCOSARA, which can automatically allocate the sampling resources to each dimension while maximizing the Fisher information. Using ultrasound multichannel localization as a case study, we quantitatively compare SCOSARA with other baseline algorithms, where it shows superior performance in terms of CRB analysis and recovery performance. Preliminary results not included in this work show that the SCOSARA can be extended to allow preferential compression of user-specified axes by replacing the regularization term $\mathrm{Trace}\{\boldsymbol{\Psi}\}$ with $\mathrm{Trace}\{\mathbf{D}\boldsymbol{\Psi}\}$, where $\mathbf{D}$ is a diagonal matrix. These results will be presented along with experiments on measurement data and comparisons against task-based algorithms by involving convolutional sparse coding algorithms [27] in a future manuscript.

REFERENCES

[1] G. Dzemyda, O. Kurasova, and J. Zilinskas. Multidimensional data visualization. *Methods and applications series: Springer optimization and its applications*, 75(122):10–5555, 2013.

[2] J. Li and P. Stoica. *MIMO radar signal processing*. John Wiley & Sons, 2008.

[3] M. T. Vlaardingerbroek and J. A. Boer. *Magnetic resonance imaging: theory and practice*. Springer Science & Business Media, 2013.

[4] C. Holmes, B. W. Drinkwater, and P. D. Wilcox. Post-processing of the full matrix of ultrasonic transmit–receive array data for non-destructive evaluation. *NDT & e International*, 38(8):701–711, 2005.

[5] Y. C. Eldar and G. Kutyniok. *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[6] D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[7] M. Lustig, D. Donoho, and J. M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[8] I. A. Huijben, B. S. Veeling, K. Janse, M. Mischi, and R. JG. van Sloun. Learning sub-sampling and signal recovery with applications in ultrasound imaging. *IEEE Transactions on Medical Imaging*, 39(12):3955–3966, 2020.

[9] J. Kirchhof, S. Semper, C. W. Wagner, E. Pérez, F. Römer, and G. Del Galdo. Frequency subsampling of ultrasound nondestructive measurements: acquisition, reconstruction, and performance. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(10):3174–3191, 2021.

[10] H. Wang, E. Pérez, and F. Römer. Data-driven subsampling matrices design for phased array ultrasound nondestructive testing. In *2023 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2023.

[11] E. Pérez, J. Kirchhof, F. Krieg, and F. Römer. Subsampling approaches for compressed sensing with ultrasound arrays in non-destructive testing. *Sensors*, 20(23):6734, 2020.

[12] H. Wang, E. Zhou, Y.and Pérez, and F. Römer. Jointly learning selection matrices for transmitters, receivers and Fourier coefficients in multichannel imaging. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8691–8695, 2024.

[13] S. Mulleti, H. Zhang, and Y. C. Eldar. Learning to sample: Data-driven sampling and reconstruction of FRI signals. *IEEE Access*, 2023.

[14] H. Wang, E. Pérez, and F. Römer. Deep learning-based optimal spatial subsampling in ultrasound nondestructive testing. In *2023 31st European Signal Processing Conference (EUSIPCO)*, pages 1863–1867. IEEE, 2023.

[15] V. Monga, Y. Li, and Y. C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.

[16] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning*, pages 399–406, 2010.

[17] S. M. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[18] G. Fatima, P. Stoica, A. Aubry, A. De Maio, and P. Babu. Optimal placement of the receivers for multistatic target localization. *IEEE Transactions on Radar Systems*, 2024.

[19] J. D. Blanchard, C. Cartis, and J. Tanner. Compressed sensing: How sharp is the restricted isometry property? *SIAM review*, 53(1):105–125, 2011.

[20] C. D. Austin, E. Ertin, J. N. Ash, and R. L. Moses. On the relation between sparse sampling and parametric estimation. In *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, pages 387–392. IEEE, 2009.

[21] C. D. Austin, J. N. Ash, and R. L. Moses. Dynamic dictionary algorithms for model order and parameter estimation. *IEEE Transactions on Signal Processing*, 61(20):5117–5130, 2013.

[22] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Transactions on Signal Processing*, 59(5):2182–2195, 2011.

[23] Z. Ben-Haim and Y. C. Eldar. The cramér-rao bound for estimating a sparse parameter vector. *IEEE Transactions on Signal Processing*, 58(6):3384–3389, 2010.

[24] I. A. Huijben, B. S. Veeling, and R. JG. van Sloun. Deep probabilistic subsampling for task-adaptive compressed sensing. In *8th International Conference on Learning Representations, ICLR 2020*, 2020.

[25] E. Jang, S. Gu, and B. Poole. Categorical Reparametrization with Gumble-Softmax. In *International Conference on Learning Representations (ICLR 2017)*. OpenReview. net, 2017.

[26] C. Maddison, A. Mnih, and Y. Teh. The concrete distribution: A continuous relaxation of discrete random variables. In *Proceedings of the International Conference on Learning Representations*. International Conference on Learning Representations, 2017.

[27] H. Wang, Kvich Y., E. Pérez, F. Römer, and Y. C. Eldar. Efficient convolutional forward modeling and sparse coding in multichannel imaging. In *2024 32st European Signal Processing Conference (EUSIPCO)*. IEEE, 2024.