forum acusticum 2026
8-12 September · Graz, Austria

ADAPTIVE LEARNED FISTA FOR ROBUST ACOUSTIC SOURCE MAPPING

Adam Kujawski[1], Jan Christian Riedel[2], Ennes Sarradj[1]

[1] *Department of Engineering Acoustics, Technische Universität Berlin, Germany*
[2] *Communications and Information Theory, Technische Universität Berlin, Germany*

Abstract

CMF estimates acoustic source strengths on a predefined focus grid by fitting a modeled microphone Cross-Spectral Matrix (CSM) to the sample CSM and is often formulated as a nonnegative LASSO problem. Classical algorithms for the LASSO need model selection and many iterations for accurate reconstructions. Algorithm unfolding reduces this effort by learning selected algorithmic parameters from ground-truth source maps. Unfolded algorithms for CMF use a dictionary from an analytic propagation model in each unfolded iteration. In measurements, the unknown true propagation model may differ from this analytic model. Supervised training creates a second mismatch because ground-truth source maps are unavailable for measured array data, so training must rely on synthetic data generated with an analytic propagation model. This work therefore compares supervised training using the MSE objective with unsupervised training, focusing on different objectives including LASSO, Cross-Validation (CV), and a smooth approximation of the Bayesian Information Criterion (BIC). In the matched setting, CV and BIC can reach validation NMSEs similar to supervised training using the MSE objective. Under propagation model mismatch, direct unsupervised training on sample CSMs generated from measured transfer functions from MIRACLE-A1 yields sparser source maps than supervised training on perturbed synthetic data.

Keywords: *covariance matrix fitting, deep unfolding, sparse recovery, dictionary mismatch, microphone array*

1. Introduction

Unfolded algorithms [1] have shown favorable accuracy and runtime compared with classical LASSO algorithms, such as ISTA and FISTA [2], for Covariance Matrix Fitting (CMF) [3], [4]. However, many unfolded algorithms are trained with supervised objectives, which require ground-truth source maps [5], [6]. In acoustic measurements, such source maps are usually unavailable, so supervised training must rely on synthetic source maps and simulated propagation models. This can introduce mismatch between training data and measured array data. Unsupervised training avoids ground-truth source maps, but existing approaches mainly optimize the original LASSO objective [5], [6]. This objective still requires choosing the regularization parameter. Model-selection criteria such as Cross-Validation (CV) and Bayesian Information Criterion (BIC) address this choice, and smooth information criteria make such criteria differentiable [7]. This work compares supervised and unsupervised objectives under a matched synthetic propagation model and under propagation model mismatch using sample Cross-Spectral Matrices (CSMs) generated from measured transfer functions of the MIRACLE-A1 dataset.

The remainder of this paper is organized as follows. Section 2 introduces CMF, the nonnegative LASSO formulation, and the unfolded FISTA algorithms. Section 3 defines the supervised and unsupervised training objectives, and Sec. 4 presents the experimental settings and results.

2. Theory

2.1. Covariance Matrix Fitting

CMF estimates source strengths by fitting a modeled CSM to the sample CSM [4]. At a single frequency, the sample CSM of c microphone channels is

$$\widehat{C} = \frac{1}{l}\sum_{b=1}^{l} p_b p_b^{\mathsf{H}}, \qquad p_b \in \mathbb{C}^c, \tag{1}$$

where p_b contains the complex microphone pressures at that frequency and l is the number of snapshots.

***Corresponding author**: adam.kujawski@tu-berlin.de.*

For a focus grid with n points, the propagation matrix $H \in \mathbb{C}^{c\times n}$ is defined by

$$H_{i,j} = \frac{r_{0,j}}{r_{i,j}} \exp\!\big(-\mathrm{i}\tfrac{\omega}{c_0}(r_{i,j} - r_{0,j})\big), \tag{2}$$

where $r_{i,j}$ is the distance from focus point j to microphone channel i, $r_{0,j}$ is the distance to the reference microphone, and c_0 is the speed of sound. For uncorrelated sources, the source CSM $Q = \mathrm{diag}(q)$ is diagonal with source strengths $q \in \mathbb{R}^n_+$. The modeled microphone CSM is then

$$C = HQH^{\mathsf{H}} = \sum_{j=1}^{n} q_j H_{:,j} H_{:,j}^{\mathsf{H}}. \tag{3}$$

To avoid redundant computations, only the upper triangular entries of the sample CSM, including the diagonal, are retained. The data vector $y \in \mathbb{R}^m$ stacks the corresponding real parts and the imaginary parts of the strict upper triangular entries, with $m = c^2$. The dictionary $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m\times n}$ is constructed with the same reduced representation, yielding

$$y = Aq + \varepsilon, \tag{4}$$

where ε is an unknown residual term that may include effects of finite acquisition time, sensor noise, and propagation-model mismatch.

2.2. LASSO

Source strengths are estimated with the nonnegative LASSO criterion

$$\mathcal{C}_{\mathrm{lasso}}(x,\lambda) = \frac{1}{2m}\|Dx - y\|_2^2 + \lambda\|x\|_1, \quad x \geq 0. \tag{5}$$

Here, x is the normalized source-strength vector and λ controls the solution sparsity. The factor $1/m$ keeps the least-squares term and the regularization scale comparable when the number of measurements changes. The normalized source-strength estimate is $\hat{x} \in \arg\min_{x\geq 0} \mathcal{C}_{\mathrm{lasso}}(x)$, and the corresponding source-strength estimate is $\hat{q} = U^{-1}\hat{x}$. The dictionary is normalized columnwise as $D = AU^{-1}$, where $U = \mathrm{diag}(\|a_1\|_2, \ldots, \|a_n\|_2)$, so that $x = Uq$. Equiregularized normalization is used throughout this work. For an unnormalized measurement $\bar{y}$, define

$$\lambda_{\max}(\bar{y}) = \frac{1}{m}\max\left\{0, \max_{j=1,\ldots,n} d_j^{\top}\bar{y}\right\}, \tag{6}$$

which is the smallest regularization value for which the nonnegative LASSO solution is zero. Thus, $\lambda_{\max}(y) = 1$, and $\lambda \in [0,1]$ has a consistent relative scale across problems. Measurements are rescaled as $y = \bar{y}/\lambda_{\max}(\bar{y})$, and the corresponding target source-strength vector is rescaled as $q = \bar{q}/\lambda_{\max}(\bar{y})$.

2.3. Unfolded FISTA

The Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) minimizes the criterion in Eq. (5) using proximal gradient descent with momentum [2]. Algorithm unfolding parameterizes selected parts of an optimization algorithm and learns these parameters from data [1]. Starting from $t^{(0)} = 1$ and $x^{(0)} = x^{(-1)} = 0$, the unfolded FISTA algorithms of this work use the iteration

$$t^{(k+1)} = \frac{1 + \sqrt{1 + 4(t^{(k)})^2}}{2}, \tag{7}$$

$$z^{(k)} = x^{(k)} + \frac{t^{(k)} - 1}{t^{(k+1)}}\big(x^{(k)} - x^{(k-1)}\big), \tag{8}$$

$$x^{(k+1)} = S^{+}_{\gamma^{(k)}\lambda}\left(z^{(k)} - \frac{\gamma^{(k)}}{m}W(Dz^{(k)} - y)\right). \tag{9}$$

Here, $S^{+}_{\eta}(u) = \max(u - \eta, 0)$ is applied componentwise, $\gamma^{(k)}$ denotes the iteration-dependent step size, and $W \in \mathbb{R}^{n\times m}$ denotes the effective weight matrix. The scalar sequence $t^{(k)}$ determines the momentum coefficient $(t^{(k)} - 1)/t^{(k+1)}$. Classical FISTA is recovered for $W = D^{\mathsf{T}}$ and constant $\gamma^{(k)}$.

Both unfolded FISTA variants use neural augmentation following [8], where a small LSTM determines the problem- and iteration-dependent step size $\gamma^{(k)}$. Depending on the training objective, λ is either fixed or learned, and the shrinkage threshold is $\theta^{(k)} = \gamma^{(k)}\lambda$. The variants differ only in the choice of the effective weight matrix W.

2.3.1. Neurally Augmented Step-Learned FISTA

Learned or pre-optimized weight matrices can improve reconstruction for unfolded ISTA under supervised training [9]. However, the unsupervised LASSO setting motivates using $W = D^{\mathsf{T}}$. For ISTA, deep layers of a convergent unfolded algorithm trained with a LASSO objective reduce to learned step-size iterations under suitable assumptions [5]. Neurally Augmented Step-Learned FISTA (NA-SLFISTA) therefore sets $W = D^{\mathsf{T}}$ and uses the LSTM to determine $\gamma^{(k)}$, with $\theta^{(k)} = \gamma^{(k)}\lambda$.

2.3.2. Adaptive Neurally Augmented Step-Learned FISTA

Adaptive Neurally Augmented Step-Learned FISTA (Ada-NA-SLFISTA) follows Ada-LISTA, which introduces a learnable weight matrix $B \in \mathbb{R}^{m\times m}$ while keeping the dictionary explicit. This allows one learned algorithm to handle varying dictionaries [10]. Under suitable coherence and sparsity assumptions, convergence guarantees for a clean dictionary D can extend to noisy dictionaries $\tilde{D} = D + E$ [10]. Motivated by this adaptive formulation, Ada-NA-SLFISTA adds a learnable weight matrix $B \in \mathbb{R}^{m\times m}$, initialized as the identity, and uses the effective weight matrix

$$W = D^{\mathsf{T}}B^{\mathsf{T}}. \tag{10}$$

3. Optimization Objectives

This section defines the loss functions used to train the unfolded algorithms. For the sth sample prob-

lem with measurements $y_s \in \mathbb{R}^m$ and dictionary $D \in \mathbb{R}^{m\times n}$, let $\hat{x}_s^{(K)} \in \mathbb{R}_+^n$ denote the estimate after K unfolded iterations. Training minimizes the empirical loss

$$\mathcal{L} = \frac{1}{S}\sum_{s=1}^{S} \mathcal{C}\left(\hat{x}_s^{(K)}\right), \tag{11}$$

over a batch of S sample problems, where $\mathcal{C}$ is the objective criterion.

3.1. Supervised Learning

3.1.1. Mean-Squared Error (MSE)

When ground-truth source maps q_s are available, training can use supervised objectives. The MSE is commonly used to train unfolded algorithms and is defined as

$$\mathcal{C}_{\text{mse}}(\hat{x}_s^{(K)}) = \frac{1}{n}\|q_s - \hat{q}_s^{(K)}\|_2^2. \tag{12}$$

3.2. Unsupervised Learning

In the unsupervised setting, ground-truth source maps are not provided, and the training dataset consists only of $\{y_s\}_{s=1}^S$.

3.2.1. Least Absolute Shrinkage and Selection Operator (LASSO)

A natural choice is the LASSO criterion from Eq. (5), evaluated at the solution $x = \hat{x}_s^{(K)}$ and $y = y_s$. This criterion matches the objective minimized by the classical algorithm that the unfolded algorithm is based on. A disadvantage is that λ must be specified. This creates an outer model-selection problem because the optimal λ is unknown. In addition, the LASSO objective is not fully differentiable because $\lambda\|\hat{x}_s^{(K)}\|_1$ is nondifferentiable at zero. It therefore requires subgradient handling, which modern autodifferentiation frameworks support.

3.2.2. Cross-Validation (CV)

CV does not require λ to be specified beforehand. In CV, the scalar entries of each measurement vector y_s are randomly split into training and validation subsets based on the row-index sets T_s and V_s. The split defines $\{(A_{s,t}, A_{s,v}, y_{s,t}, y_{s,v})\}_{s=1}^S$, with 80% of the entries assigned to T_s and 20% assigned to V_s. Here, $m_t = |T_s|$ and $m_v = |V_s|$ denote the numbers of scalar entries in the training and validation subsets, respectively. Training and validation dictionaries are normalized with $U_{s,t} = \operatorname{diag}(\|a_{s,t,1}\|_2, \ldots, \|a_{s,t,n}\|_2)$, giving $D_{s,t} = A_{s,t}U_{s,t}^{-1}$ and $D_{s,v} = A_{s,v}U_{s,t}^{-1}$. Both measurement subsets are normalized by $\lambda_{s,t,\max}$. This value is computed from Eq. (6) using $D_{s,t}$, $\bar{y}_{s,t}$, and m_t, so that $y_{s,t} = \lambda_{s,t,\max}^{-1}\bar{y}_{s,t}$ and $y_{s,v} = \lambda_{s,t,\max}^{-1}\bar{y}_{s,v}$. The CV objective minimizes the mean-squared residual on the validation entries for reconstructions $\hat{x}_{s,t}^{(K)}$ computed from the training subset:

$$\mathcal{C}_{\text{cv}}(\hat{x}_{s,t}^{(K)}) = \frac{1}{2m_v}\|D_{s,v}\hat{x}_{s,t}^{(K)} - y_{s,v}\|_2^2. \tag{13}$$

3.2.3. Bayesian Information Criterion (BIC)

Here, BIC is used as a training objective. For the estimate $\hat{x}_s^{(K)}$, the objective is

$$\mathcal{C}_{\text{bic}}(\hat{x}_s^{(K)}) = m\log\left(\frac{\|D\hat{x}_s^{(K)} - y_s\|_2^2}{m}\right) + \log(m)\|\hat{x}_s^{(K)}\|_{0,\epsilon_s}. \tag{14}$$

The first term corresponds to the Gaussian linear-model log-likelihood with unknown variance, using $\|D\hat{x}_s^{(K)} - y_s\|_2^2/m$ as the variance estimate. The second term smooths the number of nonzero coefficients used as an unbiased degrees-of-freedom estimate for the LASSO [11]. Since the ℓ_0-norm is nondifferentiable, the cardinality term is replaced by a smooth approximation [7], [12],

$$\|\hat{x}\|_{0,\epsilon} = \sum_{j=1}^{n}(1 - \phi_\epsilon(\hat{x}_j)), \quad \lim_{\epsilon\to 0}\phi_\epsilon(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases}. \tag{15}$$

Thus, $1 - \phi_\epsilon(x)$ approximates the indicator function smoothly. This work uses the Gaussian family,

$$\phi_\epsilon(x) = \exp\left(\frac{-x^2}{2\epsilon^2}\right) \tag{16}$$

The smoothing threshold ϵ_s is chosen from the current reconstruction rather than fixed globally. For each sample, ϵ_s is set 20 dB below the largest entry of the reconstructed normalized source-strength vector,

$$\epsilon_s = \alpha\|\hat{x}_s^{(K)}\|_\infty, \qquad \alpha = 10^{-2}. \tag{17}$$

This makes the smoothed cardinality term depend on each reconstructed source map. Entries far below the largest reconstructed source-strength coefficient are therefore treated as negligible, while the threshold remains comparable across normalized sample problems with different source levels and source counts. Other smooth approximations, including rational, triangular, and truncated hyperbolic families, are discussed in [12]. Comparing these alternatives is beyond the scope of the present study.

4. Experiments

The unfolded algorithms are evaluated in two experimental settings. First, the experiment with a matched synthetic propagation model in Sec. 4.2 tests whether unfolded algorithms trained with unsupervised objectives can perform comparably to supervised training using the MSE objective. Second, the experiment in Sec. 4.3 tests how robust Ada-NA-SLFISTA is under propagation model mismatch depending on the objective.

4.1. Experimental Setup and Data Generation

Fig. 1 shows the phyllotaxis-based microphone array and the focus grid used in all experiments [13]. The aperture $d_a = 0.648\,\text{m}$ is the maximum pairwise mi-

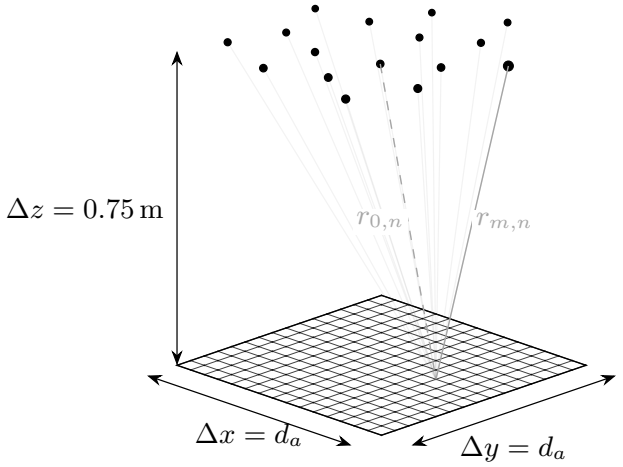

Figure 1: Phyllotaxis microphone array with 16 microphones and the 16×16 focus grid used for reconstruction.

crophone distance. The square focus grid is parallel to the array plane at $z = 0.75\,\mathrm{m}$, spans $d_a \times d_a$, and contains 16×16 points ($n = 256$). The reference microphone is closest to the array center.
Source strengths are sampled following [3]. For each realization, the source strength is sampled as

$$\begin{aligned} q_j &= \mathbf{1}_{\{r_j < J/n\}} u_j, \\ r_j &\sim \mathcal{U}(0,1), \quad u_j \sim \mathcal{R}(1), \\ j &= 1, \ldots, n. \end{aligned} \tag{18}$$

Here, $J = 10$ is the expected number of active sources, and $\mathcal{R}(1)$ denotes the Rayleigh distribution with unit scale. When the source grid and focus grid coincide, q serves as the ground-truth source map.
For the active set $\mathcal{I} = \{j : q_j > 0\}$, $H_{\mathcal{I}}$ contains the corresponding columns of the propagation matrix, and $Q_{\mathcal{I}} = \mathrm{diag}(q_{\mathcal{I}})$ denotes the source CSM. Finite acquisition time is modeled by drawing the sample source CSM from a scaled complex Wishart distribution,

$$\widehat{Q}_{\mathcal{I}} \sim \frac{1}{l} \mathcal{W}_{\mathbb{C}}(l, Q_{\mathcal{I}}), \qquad \mathbb{E}\{\widehat{Q}_{\mathcal{I}}\} = Q_{\mathcal{I}}, \tag{19}$$

where l is the equivalent number of independent snapshots. Spatially white sensor noise is drawn as

$$\widehat{\Sigma}_n \sim \frac{1}{l} \mathcal{W}_{\mathbb{C}}\big(l, \sigma^2 I_c\big), \qquad \sigma^2 = \frac{\sum_{j \in \mathcal{I}} |H_{0,j}|^2 q_j}{10^{\mathrm{SNR}/10}}. \tag{20}$$

All experiments use $l = 128$ and $\mathrm{SNR} = 20\,\mathrm{dB}$. The sample CSM is then

$$\widehat{C} = H_{\mathcal{I}} \widehat{Q}_{\mathcal{I}} H_{\mathcal{I}}^{\mathsf{H}} + \widehat{\Sigma}_n. \tag{21}$$

All experiments use the nominal Helmholtz numbers

$$\mathrm{He} = \frac{f d_a}{c_0}, \tag{22}$$

with $\mathrm{He} \in \{16, 8\}$, speed of sound $c_0 = 343\,\mathrm{m/s}$, and frequency f in Hz.

4.2. Matched Synthetic Propagation Model

Sample CSMs are generated as described in Sec. 4.1, using the analytic propagation matrix. The source grid equals the focus grid, so all simulated sources lie on the reconstruction grid. For each combination of nominal Helmholtz number and objective, separate instances of NA-SLFISTA and Ada-NA-SLFISTA are trained. The objectives are CV, BIC, LASSO with $\lambda = 0.1\lambda_{\max}$ or $0.5\lambda_{\max}$, and MSE as the supervised reference. For LASSO, λ is fixed to the stated value. For the other objectives, λ is trained as a nonnegative scalar initialized at $0.01\lambda_{\max}$. The algorithms are trained incrementally up to $K = 15$ unfolded iterations. For each trained variant and unfolded depth K, training uses $1,000$ mini-batches of 512 generated sample problems with a learning rate of 10^{-4}. For each K, the current unfolded algorithm is evaluated every 100 optimization updates on one validation batch of 256 generated sample problems. After training, test evaluation uses $1,024$ generated sample problems. The resulting validation Normalized Mean-Squared Error (NMSE) curves over optimization updates are shown in Fig. 2.

4.2.1. Results and Discussion

The learnable weight matrix improves training convergence, as shown in Fig. 2. For Ada-NA-SLFISTA, BIC closely follows the supervised MSE reference at both Helmholtz numbers and reaches a similar final NMSE. The CV objective also benefits from the learnable weight matrix but converges more slowly, especially at the lower Helmholtz number. In contrast, LASSO with $0.5\lambda_{\max}$ underfits for all tested cases. Reducing the regularization parameter to $0.1\lambda_{\max}$ lowers the NMSE for LASSO across all tested unfolded algorithms, indicating that $0.5\lambda_{\max}$ yields overly sparse reconstructions. The smaller improvement for LASSO is consistent with the Step-LISTA (SLISTA) analysis [5]. For unfolded algorithms that converge to the LASSO solution, the deepest unfolded iterations reduce to ISTA-like updates with learned step sizes. Accordingly, when training minimizes the LASSO objective, the learnable weight matrix B is not expected to substantially change the solution reached after many unfolded iterations. The slower CV convergence at the lower Helmholtz number is consistent with the entry-wise split of a single vectorized sample CSM, which separates correlated entries rather than independent observations. This correlation is expected to increase at lower Helmholtz numbers and may contribute to the slower CV convergence.

4.3. Generalization under Propagation Model Mismatch

This experiment evaluates Ada-NA-SLFISTA under propagation model mismatch. Sample CSMs are generated as in Sec. 4.1, but the propagation matrix is assembled from measured transfer functions of the MIRACLE-A1 dataset [14]. Source positions are sam-

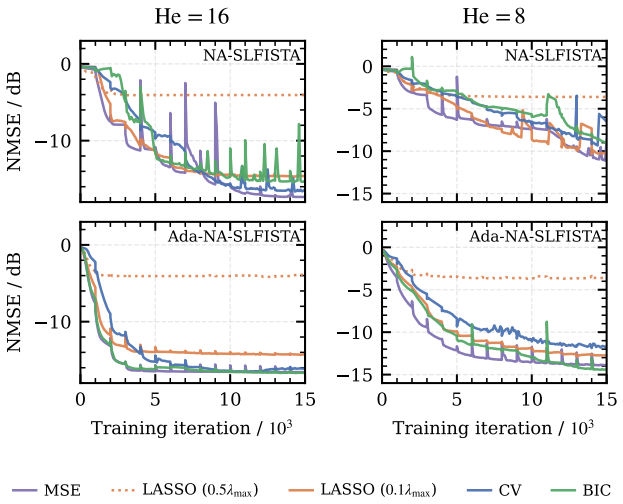

Figure 2: Validation NMSE in decibels during training of NA-SLFISTA and Ada-NA-SLFISTA under the matched synthetic propagation model.

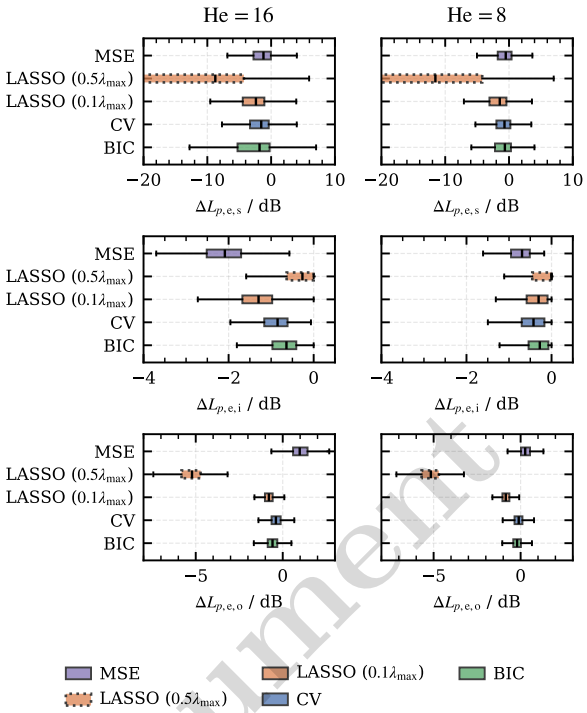

Figure 3: Boxplots of sourcemap reconstruction quality measures for Ada-NA-SLFISTA on the measured-transfer test set. Boxes indicate the interquartile range and median; whiskers follow the Tukey convention of 1.5 interquartile ranges, and outliers are omitted.

pled from measured source positions within the focus-grid extent. These source positions lie at $z \approx 0.74\,\mathrm{m}$ and generally do not coincide with the 16×16 focus grid at $z = 0.75\,\mathrm{m}$. The training data differs by objective. Training with unsupervised objectives uses sample CSMs generated from measured transfer functions, because CV, LASSO, and BIC do not require grid-aligned ground-truth source maps. In contrast, supervised MSE training requires ground-truth source maps and is therefore performed on synthetic data whose source grid is identical to the focus grid, with propagation model perturbations. For these synthetic sample problems, the sound speed is drawn independently from $\mathcal{U}(340.5, 345.5)\,\mathrm{m/s}$. The microphone x- and y-coordinates are perturbed independently by zero-mean Gaussian noise with standard deviation 0.001 m. All trained variants are evaluated on the same test set of sample CSMs generated from measured transfer functions. During reconstruction, all variants use the analytic free-field propagation model from Sec. 4.2 with $c_0 = 343\,\mathrm{m/s}$. The training setup and final test evaluation are otherwise the same as in Sec. 4.2.

4.3.1. Results and Discussion

Fig. 3 compares the training objectives on the test set in terms of three sourcemap reconstruction quality measures [15]. Circular source sectors are centered at the nominal source positions with a radius of 1.5× the focus-grid increment. For each sector, $\Delta L_{p,\mathrm{e,s}}$ measures the difference between the target source strength and the reconstructed source strength within that sector. $\Delta L_{p,\mathrm{e,i}}$ compares the reconstructed source strength assigned to all sectors with the total reconstructed source strength in the full source map. Negative values therefore indicate spurious contributions outside the sectors. $\Delta L_{p,\mathrm{e,o}}$ measures over- or underestimation of the total reconstructed source strength.

$\Delta L_{p,\mathrm{e,s}}$ first assesses whether Ada-NA-SLFISTA trained with supervised MSE on synthetic propagation-model perturbations can recover source strengths on the test set generated from measured transfer functions. Supervised MSE training reaches median $\Delta L_{p,\mathrm{e,s}}$ values of about $-1.2\,\mathrm{dB}$ at $\mathrm{He} = 16$ and $-0.5\,\mathrm{dB}$ at $\mathrm{He} = 8$, similar to CV and BIC. In contrast, LASSO underestimates source levels, especially for $\lambda = 0.5\lambda_{\max}$, consistent with stronger ℓ_1 shrinkage. However, $\Delta L_{p,\mathrm{e,i}}$ and $\Delta L_{p,\mathrm{e,o}}$ reveal a limitation of supervised MSE training that is not visible from $\Delta L_{p,\mathrm{e,s}}$ alone. Supervised MSE training yields median $\Delta L_{p,\mathrm{e,i}}$ values of about $-2.1\,\mathrm{dB}$ at $\mathrm{He} = 16$ and $-0.7\,\mathrm{dB}$ at $\mathrm{He} = 8$. The positive median $\Delta L_{p,\mathrm{e,o}}$ values for supervised MSE training indicate overestimation of the total reconstructed source strength. In contrast, the mostly negative $\Delta L_{p,\mathrm{e,o}}$ values for the unsupervised objectives indicate underestimation, consistent with stronger shrinkage or suppression of off-sector contributions. The source maps in Fig. 4 illustrate the same trends. Ada-NA-SLFISTA trained with supervised MSE recovers the strongest source but shows incorrect source contributions outside the sectors, consistent with its negative $\Delta L_{p,\mathrm{e,i}}$. The source maps produced after training with unsupervised objectives are sparser, and the BIC-trained variant has $\Delta L_{p,\mathrm{e,s}}$ and $\Delta L_{p,\mathrm{e,i}}$ values closest to zero among the shown objectives in this example. The remaining spurious sources likely result from the mismatch between the propagation matrix used to

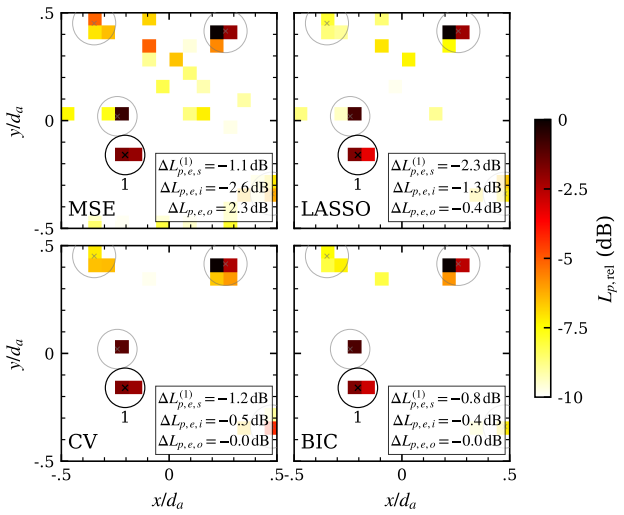

Figure 4: Example source maps obtained with Ada-NA-SLFISTA at He = 16 using sample CSMs generated from measured transfer functions. The LASSO result uses $\lambda = 0.1\lambda_{\max}$. Crosses mark measured source positions and circles mark the integration sectors.

generate the test CSMs and the analytic propagation matrix used for reconstruction.

5. Conclusion

Unfolded FISTA algorithms for CMF can be trained using unsupervised objectives. Because they do not require ground-truth source maps, these objectives could allow unfolded FISTA algorithms to be trained directly on measured array data instead of relying only on supervised training with synthetic data. CV and BIC avoided the regularization-parameter choice required by the LASSO objective, achieved validation errors comparable to supervised MSE training, and benefited from a trainable weight matrix. Future work should investigate unsupervised training with measured array data under limited data availability, including unsupervised fine-tuning of unfolded FISTA algorithms that were initially trained with supervised objectives.

6. Acknowledgments

This work was funded by the German Research Foundation (DFG), project 439144410.

References

[1] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel: Omnipress, 2010, pp. 399–406.

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

[3] C. Kayser, A. Kujawski, and E. Sarradj, "A fast data-driven method for inverse microphone array signal processing," *JASA Express Lett.*, vol. 3, no. 4, p. 042 401, 2023.

[4] T. Yardibi, J. Li, P. Stoica, and L. N. Cattafesta, "Sparsity constrained deconvolution approaches for acoustic source mapping," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 2631–2642, 2008.

[5] P. Ablin, T. Moreau, M. Massias, and A. Gramfort, "Learning step sizes for unfolded sparse coding," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[6] K. Nagahisa, R. Hayakawa, and Y. Iiguni, "Comparison between supervised and unsupervised learning in deep unfolded sparse signal recovery," *arXiv preprint arXiv:2509.01331*, 2025.

[7] M. O'Neill and K. Burke, "Variable selection using a smooth information criterion for distributional regression models," *Stat. Comput.*, vol. 33, no. 3, p. 71, 2023.

[8] F. Behrens, J. Sauder, and P. Jung, "Neurally Augmented ALISTA," in *International Conference on Learning Representations*, 2021.

[9] J. Liu, X. Chen, Z. Wang, and W. Yin, "ALISTA: Analytic weights are as good as learned weights in LISTA," in *International Conference on Learning Representations*, 2019.

[10] A. Aberdam, A. Golts, and M. Elad, "Ada-LISTA: Learned Solvers Adaptive to Varying Models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9222–9235, 2022.

[11] H. Zou, T. Hastie, and R. Tibshirani, "On the "Degrees of Freedom" of the lasso," *Ann. Statist.*, vol. 35, no. 5, pp. 2173–2192, 2007.

[12] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0 norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, 2008.

[13] E. Sarradj, "A Generic Approach To Synthesize Optimal Array Microphone Arrangements," in *Proceedings of the 6th Berlin Beamforming Conference*, February 29 – March 1, Berlin, Germany, 2016, pp. 1–12.

[14] A. Kujawski, A. J. R. Pelling, and E. Sarradj, "MIRACLE – A Microphone Array Impulse Response Dataset for Acoustic Learning," *EURASIP J. Audio Speech Music Process.*, pp. 1–16, 2024.

[15] G. Herold and E. Sarradj, "Performance analysis of microphone array methods," *J. Sound Vib.*, vol. 401, pp. 152–168, 2017.