Learning-based Sparse Recovery for Joint Activity Detection and Channel Estimation in Massive Random Access Systems

Sreeshma Shiv U K, Srikrishna Bhashyam, Chirag Ramesh Srivatsa and Chandra R. Murthy

***Abstract*—We consider the problem of joint activity detection and channel estimation in massive random access. When the receiver has multiple antennas, this is a joint sparse recovery problem with multiple measurement vectors (MMV). For the general setting where the channels could be correlated across antennas, we first develop a modified minimum mean squared error (MMSE) shrinkage function to be used in the Trainable Iterative Soft Thresholding Algorithm (TISTA). Then, we learn this MMSE shrinkage function using a model-based neural network. In the simulation results, the proposed learning-based method, L-MMSE-MMV-TISTA, offers a 30-40% reduction in preamble length requirement compared to TISTA. We also compare L-MMSE-MMV-TISTA with the state-of-the-art MMV sparse Bayesian learning (M-SBL) method. While M-SBL can provide better performance at the cost of higher complexity in highly measurement-constrained settings, LMMSE-MMV-TISTA provides a significant complexity advantage when only a slightly larger number of measurements are available.**

I. Introduction

Massive random access plays an important role in Internet-of-Things (IoT) and machine-type-communications (MTC) applications [1], [2]. In massive random access, a base station (BS) receives data from a large number of sporadically active devices. In such scenarios, sparse recovery algorithms are effective for detecting user activity and estimating the channels [1], [3]. Specifically, the joint activity detection and channel estimation problem with a multi-antenna BS can be formulated as a multiple-measurement-vector (MMV) or a joint-sparse recovery problem. Most of the existing algorithms for joint sparse recovery are variants of Iterative Soft Thresholding Algorithm (ISTA) [4], [5], Approximate Message Passing (AMP) [1], [6], [7], Alternating Direction Method of Multipliers (ADMM) [8], [9], and Sparse Bayesian Learning (SBL) [10].

Recently, sparse recovery algorithms that can learn from data while taking advantage of model-based signal processing have been developed based on deep unfolding [11], algorithm unrolling [12], or, more generally, model-based deep learning [13]. One of the first such methods is Learned ISTA (LISTA) [14]. In the context of joint user activity detection and channel estimation in massive random access, the most relevant references amongst the learning-based methods are [5], [15], [16]. Trainable ISTA (TISTA), proposed in [5] for sparse recovery from a single measurement vector, comprises a linear estimation step, an error variance estimation step, and a minimum mean squared error (MMSE) shrinkage step using the estimated error variance. TISTA was shown to outperform LISTA, AMP and Orthogonal AMP [17]. In [15], a deep unfolded version of an improved version of ADMM for the MMV setting [9] was presented. In [16], deep unfolding was used to develop a learned version of the vector AMP method with an MMSE denoiser in [1] for massive random access.

In this work, we enhance TISTA in the following ways: (i) we replace the MMSE shrinkage function by a network that can learn a more appropriate shrinkage function, (ii) we generalize the MMV setting to account for possibly correlated channel coefficients across the BS antennas. We obtain a modified Bayesian MMSE shrinkage function for the prior associated with correlated channel coefficients, in the MMV setting. Then, in order to replace the MMSE shrinkage function in TISTA, we develop a model-based approach inspired by [18] for learning the MMSE estimator. The model-based network for the shrinkage function learns from the training data and potentially overcomes inaccuracies in the model assumptions used to derive the MMSE shrinkage function. We numerically show that these enhancements provide a significant reduction in the training overhead required for user activity detection and channel estimation, compared to a direct application of TISTA and the modified ADMM-based approach in [15]. In terms of comparison with AMP-based unfolding approaches, the deep unfolded version of the MMV-ADMM method in [9] and the vector AMP-based method in [1] are observed to be comparable in performance in [16]. The modified-ADMM in [15] improves upon the standard MMV-ADMM method [9] before unfolding and provides better performance. We also compare against the MMV sparse Bayesian learning (M-SBL) method. While the M-SBL method can provide better performance at the cost of higher complexity when the number of measurements is very low, the proposed scheme provides a significant complexity advantage in less measurement-constrained regimes.

II. System Model

We consider a massive random access system consisting of one M-antenna BS and N single-antenna IoT devices. Each device becomes active independently with a probability p. The set of active devices is denoted by $\mathcal{S} = \{1, 2, \ldots, S\}$. Each active user transmits an L-length pilot sequence followed by

Sreeshma Shiv U K (ee19s005@smail.iitm.ac.in) and Srikrishna Bhashyam (skrishna@ee.iitm.ac.in) are with the Dept. of EE, IIT Madras, Chennai 600036, India. Chirag Ramesh Srivatsa (chiragramesh@iisc.ac.in) and Chandra R. Murthy (cmurthy@iisc.ac.in) are with the Dept. of ECE, Indian Institute of Science, Bangalore 560012, India.

data. Let $\boldsymbol{a}_n \in \mathbb{C}^{L\times 1}$ denote the L-length training sequence employed by user n. While there are several ways of designing $\boldsymbol{a}_n$ (for example, using Zadoff-Chu sequences), we consider sequences of i.i.d. complex Gaussian entries with zero mean and unit variance for simplicity and as an illustrative example. The complex baseband channel vector between user n and the BS is represented as $\boldsymbol{h}_n^T \in \mathbb{C}^{1\times M}$. The channel could be correlated across antennas, e.g., $\boldsymbol{h}_n^T \sim \mathcal{CN}(0, \mathbf{C}_h)$, where $\mathbf{C}_h$ is the covariance matrix.[1] The received training signal $\boldsymbol{Y} \in \mathbb{C}^{L\times M}$ is modeled as

$$\boldsymbol{Y} = \sum_{i\in\mathcal{S}} \boldsymbol{a}_i \boldsymbol{h}_i^T + \boldsymbol{N},$$

where the noise $\boldsymbol{N}$ has i.i.d. $\mathcal{CN}(0,\sigma^2)$ entries. For each user n, let $\lambda_n = 0$ if the user is inactive and $\lambda_n = 1$ if the user is active. Then, we can rewrite the above as

$$\boldsymbol{Y} = \sum_{n=1}^{N} \lambda_n \boldsymbol{a}_n \boldsymbol{h}_n^T + \boldsymbol{N} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{N}, \tag{1}$$

where $\boldsymbol{X} = [\lambda_1\boldsymbol{h}_1, \lambda_2\boldsymbol{h}_2, \cdots, \lambda_N\boldsymbol{h}_N]^T \in \mathbb{C}^{N\times M}$ and $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N]$. Since the number of active users S is much smaller than N, matrix $\boldsymbol{X}$ is a *row sparse* matrix. Thus, we have a joint sparse recovery problem, where the receiver observes $\boldsymbol{Y}$ and needs to estimate a row sparse matrix $\boldsymbol{X}$.

A. Trainable ISTA (TISTA)

We briefly present TISTA for the single measurement vector case ($M = 1$) [5]. The following equations describe one iteration (or layer) of TISTA. Here, t denotes the iteration number, and the estimate of $\boldsymbol{X}$ (note that, when $M = 1$, $\boldsymbol{X}$ is a vector) at iteration t is $\boldsymbol{s}_t$.

$$\boldsymbol{r}_t = \boldsymbol{s}_t + \gamma_t \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{s}_t), \tag{2}$$

$$\boldsymbol{s}_{t+1} = \eta_{\text{MMSE}}(\boldsymbol{r}_t; \tau_t^2), \tag{3}$$

$$v_t^2 = \max\left\{ \frac{\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{s}_t\|_2^2 - L\sigma^2}{\mathsf{tr}(\boldsymbol{A}^T\boldsymbol{A})}, \epsilon \right\}, \tag{4}$$

$$\tau_t^2 = \frac{v_t^2}{N}(N + (\gamma_t^2 - 2\gamma_t)L) + \frac{\gamma_t^2\sigma^2}{N}\mathsf{tr}(\boldsymbol{W}\boldsymbol{W}^T), \tag{5}$$

where the matrix $\boldsymbol{W}$ is the pseudo-inverse matrix of the sensing matrix $\boldsymbol{A}$, and ϵ is a small real constant.[2] The initial condition $\boldsymbol{s}_0$ can be set to zero, and the final estimate of $\boldsymbol{X}$ after T iterations is $\boldsymbol{s}_T$. The scalar variables γ_t for $t = 0, 1, \ldots, T-1$, are learnable and are tuned in a training process. $\eta_{\text{MMSE}}(\cdot;\cdot)$ is the MMSE shrinkage function that is derived using an assumed prior on $\boldsymbol{X}$. In [5], $\boldsymbol{X}$ is a vector of i.i.d. Bernoulli Gaussian random variables, i.e., each entry of $\boldsymbol{X}$ is $\mathcal{N}(0,\alpha^2)$ with probability p and zero with probability $1-p$. Furthermore, in each iteration, $\boldsymbol{r}_t$ is modelled as $\boldsymbol{r}_t = \boldsymbol{X} + \boldsymbol{Z}$, where $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \tau^2\boldsymbol{I})$. For this model, the MMSE denoiser is an element-wise denoiser given by

$$\eta(y;\tau^2) = \frac{y\alpha^2}{\alpha^2+\tau^2} \cdot \frac{pF(y;\alpha^2+\tau^2)}{(1-p)F(y;\tau^2) + pF(y;\alpha^2+\tau^2)},$$

where $F(z;v) = \frac{1}{\sqrt{2\pi v}}\exp\left(\frac{-z^2}{2v}\right)$. Here, $\{\gamma_t\}$, p and α^2 are learnt from the training data in TISTA, while σ^2 is assumed to be known. Note that τ^2 is estimated in each iteration.

III. Proposed Solutions

A. MMSE-MMV-TISTA

We now consider the MMV model in (1) with $M > 1$ antennas. In order to exploit the row sparse structure of $\boldsymbol{X}$ and the correlation structure of $\boldsymbol{h}_n$, the MMSE shrinkage step in (3) needs to be modified. Before we discuss this, we briefly mention how the other two steps of TISTA are adapted to our problem. The linear estimation step in (2) is modified to

$$\boldsymbol{R}_t = \boldsymbol{S}_t + \gamma_t \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}_t), \tag{6}$$

where $\boldsymbol{R}_t$ and $\boldsymbol{S}_t$ are now $N\times M$ matrices. The error variance estimation in (4) can be used either with the vectorized model, or simply with the observation from any one of the antennas.

The MMSE shrinkage function is derived as follows. Let $\boldsymbol{x}_i^T$ be the ith row of $\boldsymbol{X}$. This corresponds to the received signal vector from the ith user at the M receive antennas. Thus, $\boldsymbol{x}_i$ is distributed as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{C}_h)$ with probability p (if user i is active) and is $\boldsymbol{0}$ with probability $1-p$ (if user i is not active). Here, the correlation matrix $\boldsymbol{C}_h$ is assumed to arise due to the mutual coupling between receive antennas, and therefore independent of i. Further, the channel vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are independent for $i \neq j$. Therefore, the MMSE denoiser, given by

$$\boldsymbol{S}_{t+1} = \eta_{\text{MMSE}}(\boldsymbol{R}_t; \tau_t^2), \tag{7}$$

is a row-wise denoiser, i.e., the ith row of $\boldsymbol{S}_{t+1}$ is the MMSE estimate of $\boldsymbol{x}_i$ from the ith row of $\boldsymbol{R}_t$. Let $\boldsymbol{y}_i^T$ be the ith row of $\boldsymbol{R}_t$, and $\hat{\boldsymbol{x}}_i^T$ be the ith row of $\boldsymbol{S}_{t+1}$. In order to obtain the MMSE estimate of $\boldsymbol{x}_i$, we use the model $\boldsymbol{y}_i = \boldsymbol{x}_i + \boldsymbol{z}_i$, where $\boldsymbol{z}_i \sim \mathcal{N}(\boldsymbol{0}, \tau_t^2\boldsymbol{I})$ and $\boldsymbol{x}_i$ is a Bernoulli-Gaussian random vector as described above. Now, using [18, Lemma 1], we get

$$\hat{\boldsymbol{x}}_i = \boldsymbol{W}(\hat{\boldsymbol{C}})\boldsymbol{y}_i, \quad \text{with} \quad \hat{\boldsymbol{C}} = \frac{1}{\tau_t^2}\boldsymbol{y}_i\boldsymbol{y}_i^T, \tag{8}$$

$$\boldsymbol{W}(\hat{\boldsymbol{C}}) = \frac{p\,\exp\left[\mathsf{tr}(\boldsymbol{W}_h\hat{\boldsymbol{C}}) + \log|\boldsymbol{I} - \boldsymbol{W}_h|\right]\boldsymbol{W}_h}{(1-p) + p\,\exp\left[\mathsf{tr}(\boldsymbol{W}_h\hat{\boldsymbol{C}}) + \log|\boldsymbol{I} - \boldsymbol{W}_h|\right]}, \tag{9}$$

and $\boldsymbol{W}_h = \boldsymbol{C}_h(\boldsymbol{C}_h + \tau_t^2\boldsymbol{I})^{-1}$. In summary, the proposed MMSE-MMV-TISTA method is specified by equations (6), (7), (4) and (5), where the observations from one of the antennas[3] are used in (4). The details of (7) are in (8)-(9). The active users are detected from the estimate $\hat{\mathbf{X}} = \mathbf{S}_T$ by comparing the norm of each row with a threshold that is a constant times the noise variance σ^2. The value of the constant is chosen empirically, but the performance is not sensitive to the specific value used, since the row norms typically nicely separate into low and high values.

B. L-MMSE-MMV-TISTA

The MMSE shrinkage function in TISTA depends on the prior used for $\boldsymbol{X}$. Our row-sparse prior model with possible

[1] The algorithm in Sec. III-A requires knowledge of $\mathbf{C}_h$, while the algorithm in Sec. III-B learns $\mathbf{C}_h$ from the data.

[2] The sparse recovery problem with complex vectors and matrices in (1) can be converted to an equivalent real-valued problem as done in several other works [15], [19], [20] that use deep learning based optimizers [21] which work with real-valued parameters. Hence, we present all the algorithms for the real-valued case in this paper.

[3] An estimate of the variance of the error in entries of $\mathbf{S}_t$ can be obtained from (4). Empirically, we observed that using one column of $\mathbf{Y}$ and $\mathbf{S}_t$ provides similar results as using all the observations. So, we choose to use only one column to simplify the computation.

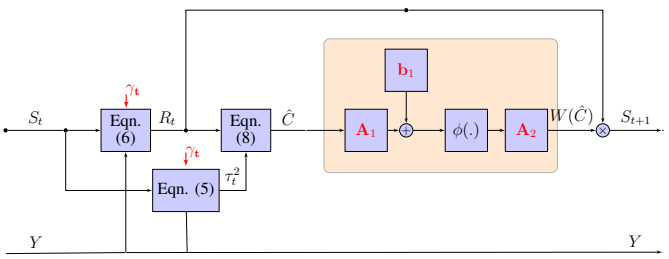

Fig. 1: One iteration of L-MMSE-MMV-TISTA.

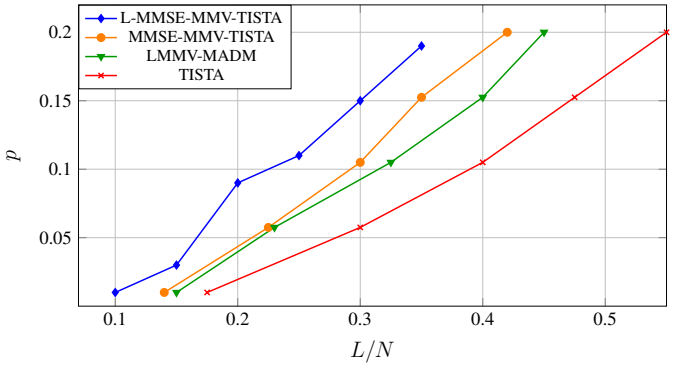

Fig. 2: $N = 500$, $M = 4$, SNR $= 30$ dB, $\rho = 0.5$.

antenna correlation is different from the sparse vector model in [5]. An open problem mentioned in [5] is the possibility of learning this shrinkage function using a neural network. In this section, we present a learned version of MMSE-MMV-TISTA, where the shrinkage function or denoiser is learned using a model-based neural network. The structure of the neural network is based on (8)-(9) derived earlier. The learned version, named L-MMSE-MMV-TISTA, has the following advantages: (i) knowledge of $\boldsymbol{C}_h$ and p are not required, as they are learned implicitly during training, and (ii) inaccuracies in the model assumptions for the estimate from the linear estimation step, i.e., $\boldsymbol{R}_t = \boldsymbol{X} + \boldsymbol{Z}$, could potentially be alleviated during the training process.

Let $\mathsf{vec}(\boldsymbol{X})$ denote the column vector obtained by stacking the columns of $\boldsymbol{X}$. Then, simplifying (9), we get

$$\mathsf{vec}(\boldsymbol{W}(\hat{\boldsymbol{C}})) = \boldsymbol{A}_{\mathrm{W}} \frac{\exp(\boldsymbol{A}_{\mathrm{W}}^T \mathsf{vec}(\hat{\boldsymbol{C}}) + \boldsymbol{b})}{\boldsymbol{1}^T \exp(\boldsymbol{A}_{\mathrm{W}}^T \mathsf{vec}(\hat{\boldsymbol{C}}) + \boldsymbol{b})}, \qquad (10)$$

where $\boldsymbol{A}_{\mathrm{W}} = [\mathsf{vec}(\boldsymbol{W}_h), \boldsymbol{0}]$ is an $M^2 \times 2$ matrix, $\boldsymbol{b} = [b_1, b_2]^T$, $b_1 = \log|I - \boldsymbol{W}_h| + \log p$, and $b_2 = \log(1-p)$. In our setting, each row of $\boldsymbol{X}$ is Gaussian with two possible covariance matrices: zero with probability $1-p$ and $\boldsymbol{C}_h$ with probability p. This is different from [18] where the case with N different equally likely covariance matrices has been considered. We obtain (10) with unequal probabilities for the different possible covariance matrices, by adding the $\log p_i$ term in b_i, where p_i is the probability of covariance matrix i. The structure of (10) is that of a feed-forward neural network with two linear layers connected by a *softmax* activation function. The $\boldsymbol{A}_{\mathrm{W}}^T \mathsf{vec}(\hat{\boldsymbol{C}}) + \boldsymbol{b}$ term in (10) corresponds to the first linear layer. The softmax activation function computes the ratio $\frac{\exp(\boldsymbol{A}_{\mathrm{W}}^T \mathsf{vec}(\hat{\boldsymbol{C}}) + \boldsymbol{b})}{\boldsymbol{1}^T \exp(\boldsymbol{A}_{\mathrm{W}}^T \mathsf{vec}(\hat{\boldsymbol{C}}) + \boldsymbol{b})}$. The multiplication by $\boldsymbol{A}_{\mathrm{W}}$ corresponds to the second linear layer. Thus, this denoiser structure (shown inside a box in Fig. 1) reduces to (10) for $\boldsymbol{A}_1 = \boldsymbol{A}_{\mathrm{W}}^T$, $\boldsymbol{b}_1 = \boldsymbol{b}$, and $\boldsymbol{A}_2 = \boldsymbol{A}_{\mathrm{W}}$. The input to this network is $\mathsf{vec}(\hat{\boldsymbol{C}})$ and the output is $\mathsf{vec}(\boldsymbol{W}(\hat{\boldsymbol{C}}))$.

Figure 1 shows the computations in one iteration of the proposed learnt MMSE-MMV-TISTA method. The complexity of one iteration of L-MMSE-MMV-TISTA is of the order $O(LNM)$. The trainable parameters in each iteration, namely γ_t, $\boldsymbol{A}_1$, $\boldsymbol{b}_1$, and $\boldsymbol{A}_2$, are shown in red color. Note that the same network is used to denoise each row of $\boldsymbol{R}_t$ with the $\hat{\boldsymbol{C}}$ computed as in (8) for each row. Therefore, the total number of trainable parameters in one iteration is $4M^2 + 3$. This is to be expected since the unknown covariance $\boldsymbol{C}_h$ is an $M \times M$ matrix. The number of parameters can be reduced in the following ways: (i) Since one of the covariances is zero (when user is not active), this information can be used by inserting the structure of $\boldsymbol{A}_{\mathrm{W}}$ in (10); (ii) If the covariance has a Toeplitz structure (which is common in many practical settings), the number of parameters can be reduced to be of order M [18].

IV. Simulation Results and Discussion

In this section, we first evaluate the performance of the proposed methods, MMSE-MMV-TISTA and L-MMSE-MMV-TISTA, in the massive random access setting. We compare with the LMMV-MADM method in [15] and a direct implementation of TISTA by vectorizing the MMV problem to an SMV problem. In subsection IV-A, we compare with the M-SBL method [10]. The deep unfolded version of the MMV-ADMM method in [9] and the vector AMP-based method in [1] are observed to be comparable in performance in [16]. The modified-ADMM in [15] improves upon the standard MMV-ADMM method [9] before unfolding and provides better performance. Therefore, we compare with [15].

In the simulations, we set the number of users $N = 500$, number of antennas $M = 4$ or 10, and number of iterations (or layers after unfolding) of each algorithm to 12 (similar to [2], [5], [15], [22]). The SNR is defined to be $\frac{E[\|\boldsymbol{AX}\|_F^2]}{E[\|\boldsymbol{N}\|_F^2]}$. The covariance matrix $\boldsymbol{C}_h$ has (i,j)th entry equal to $\rho^{|i-j|}$ [23], [24]. The L-MMSE-MMV-TISTA network is trained layer-by-layer in a supervised manner using the ADAM optimizer [21].[4] For each layer, the learning rate is 0.04, batch size is 150 and the number of batches is 200. The NMSE is defined as $\frac{E[\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F^2]}{E[\|\boldsymbol{X}\|_F^2]}$, where $\hat{\boldsymbol{X}}$ is the estimate of $\boldsymbol{X}$. 7500 samples are used to compute the NMSE during testing.

Figure 2 shows the phase transition performance of the different algorithms. The phase transition characterizes the minimum required L/N (ratio of preamble length to the number of users) for a given activity probability p. The SNR is set to 30 dB, $\rho = 0.5$, $M = 4$, and the networks are then trained and tested at various activity values of $p \in [0.01, 0.2]$. The network is considered successful in recovery if the NMSE goes below -20 dB. We observe that LMMSE-MMV-TISTA is significantly better than TISTA in terms of the preamble length requirement. For example, at $p = 0.1$, the required

[4]Note that the training of the network parameters can be done offline, so that this does not affect the complexity in the testing phase.

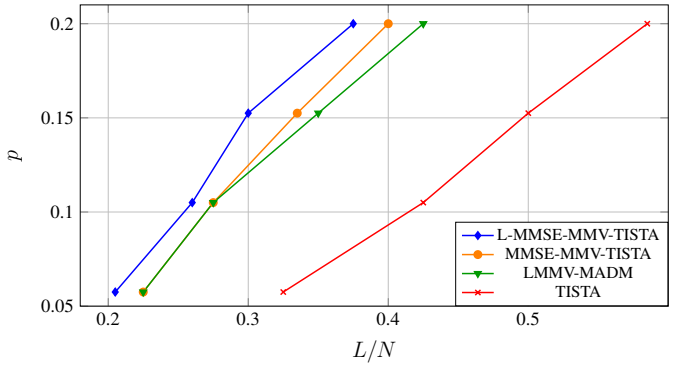

Fig. 3: $N = 500$, $M = 10$, SNR $= 20$ dB, $\rho = 0.5$.

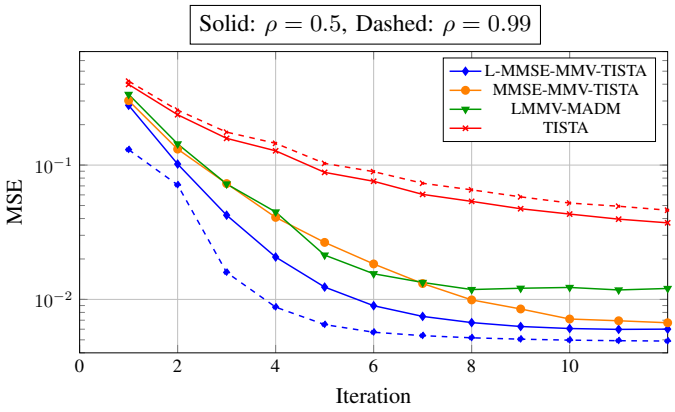

Fig. 4: MSE convergence: $N = 500$, $M = 4$, SNR $= 20$ dB, $p = 0.1$, $L/N = 0.35$.

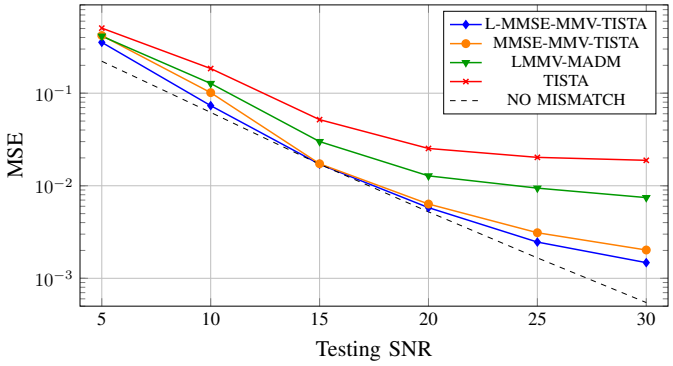

Fig. 5: MSE vs. Testing SNR: Training SNR $= 15$ dB, $N = 500$, $M = 4$, $\rho = 0.5$, $p = 0.1$, $L/N = 0.4$.

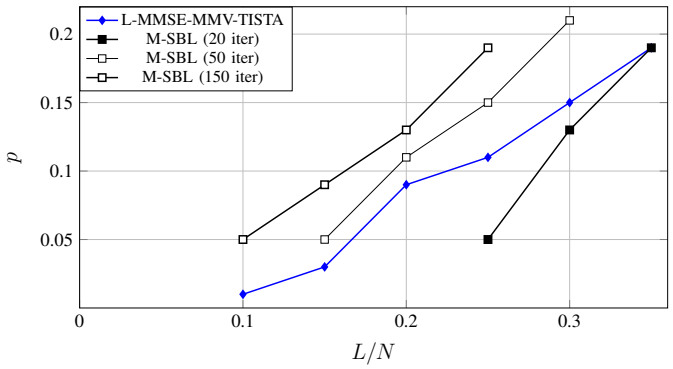

Fig. 6: $N = 500$, $M = 4$, SNR $= 30$ dB, $\rho = 0.5$.

preamble length is 35% lesser than for TISTA. Furthermore, the L-MMSE-MMV-TISTA performs better than the MMSE-MMV-TISTA. This is because the model $\boldsymbol{R}_t = \boldsymbol{X} + \boldsymbol{Z}$ used to derive the MMSE denoiser expression in MMSE-MMV-TISTA is not accurate. Similarly, some assumptions on the residual error at each stage are used in the error variance estimation step in TISTA. The neural network based approach is able to compensate for some of these inaccuracies by learning from the training data. Significant performance improvement is also seen at SNR $= 20$ dB, $M = 10$, in Figure 3. At $p = 0.1$, the reduction in preamble requirement with LMMSE-MMV-TISTA compared to TISTA is 39%. For the 4 and 20 antenna cases with the same parameters (not shown in the Figure), the reduction in preamble requirement was 33% and 40%, respectively. As expected, the gain is higher when the number of antennas is higher. Using the row sparsity and correlation structure provides more gains with larger number of antennas.

Next, we compare the MSE convergence performance of the algorithms in Figure 4. Here, we set $p = 0.1$, $L = 175$, $N = 500$, and $\rho = 0.5$. We see that L-MMSE-MMV-TISTA converges with fewer layers and also achieves a lower MSE after convergence than TISTA and LMMV-MADM. The faster convergence of the L-MMSE-MMV-TISTA for a high correlation of $\rho = 0.99$ is also shown in Figure 4.

In TISTA and our proposed extensions, the noise variance σ^2 is assumed to be known, while the LMMV-MADM method [15] does not require σ^2. In Figure 5, we plot the MSE achieved when the testing SNR is different from training SNR. The MSE achieved by L-MMSE-MMV-TISTA when there is no mismatch is plotted for reference. It can be observed that the proposed method shows significant improvement over existing methods even under this SNR mismatch scenario.

A. *Comparison with M-SBL*

In Fig. 6, we compare L-MMSE-MMV-TISTA and M-SBL [10] for the same setting as in Fig. 3. The L-MMSE-MMV-TISTA with 20 layers outperforms M-SBL with 20 iterations, i.e., when the computational complexity allowed is the same. However, M-SBL can perform better with higher number of iterations, i.e., at a higher computational complexity. It is worth noting that one iteration of M-SBL is more complex than one iteration/layer of L-MMSE-MMV-TISTA. The complexity of one iteration of M-SBL is $O(L^3NM)$, while that of L-MMSE-MMV-TISTA and LMMV-MADM are $O(LNM)$.

In Fig. 7, we compare the MSE convergence of L-MMSE-MMV-TISTA, LMMV-MADM and M-SBL. Here, we use 30 dB SNR with 10 antennas and set $N = 500$, $L = 150$, $\rho = 0$. The activity probability p is set to 0.05 or 0.13. For $p = 0.05$, all three algorithms converge to the same MSE, and L-MMSE-MMV-TISTA and LMMV-MADM can converge faster than M-SBL. Thus, L-MMSE-MMV-TISTA and LMMV-MADM can provide a lower complexity solution for this setting. The faster convergence is achieved mainly because the parameters for each layer are learnt from data. However, as we increase

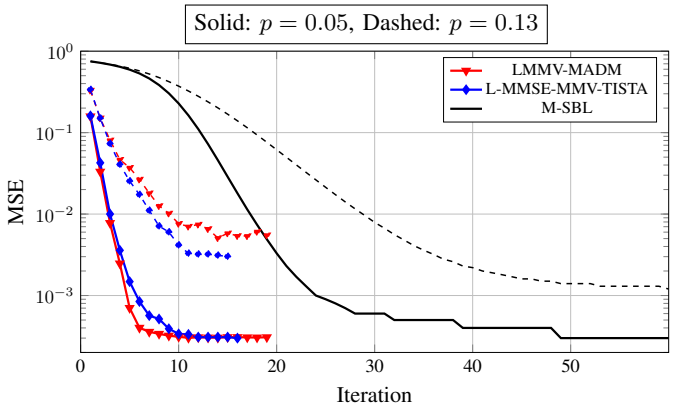

Fig. 7: MSE convergence: $N = 500$, $M = 10$, SNR $= 30$ dB, $\rho = 0$, $p = 0.05, 0.13$, $L/N = 0.3$.

p, we observe that M-SBL is able to achieve lower MSE for $p \geq 0.13$. This is because, for higher p, the number of measurements is severely constrained and M-SBL can provide a lower MSE at a higher computational cost. In Fig. 7, for $p = 0.13$, even though the convergence is faster for L-MMSE-MMV-TISTA, M-SBL is able to achieve a lower MSE after 50 iterations.

V. Conclusions

Trainable ISTA [5] is a powerful sparse recovery algorithm with good convergence and a fast training process. Addressing an open problem mentioned in [5], we developed an extension of TISTA, named L-MMSE-MMV-TISTA, where the MMSE shrinkage function is replaced by a model-based neural network that can learn an appropriate shrinkage function. We also considered a more general MMV setting that arises in a massive random access system with correlated antennas at the base station. We observed via simulations that the algorithms developed here provide gains even when there is a mismatch between training and testing SNRs. We also compared L-MMSE-MMV-TISTA with other existing methods like LMMV-MADM and M-SBL and identified settings where there are significant advantages in terms of reducing the overhead or complexity. Future work can consider extending this approach to the M-SBL framework, to obtain algorithms with the performance of M-SBL and the speed of L-MMSE-MMV-TISTA, especially in measurement constrained regimes.

References

[1] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, 2018.

[2] T. Jiang, Y. Shi, J. Zhang, and K. B. Letaief, "Joint activity detection and channel estimation for IoT networks: Phase transition and computation-estimation tradeoff," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6212–6225, 2019.

[3] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, 2018.

[4] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, 2003.

[5] D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3113–3125, 2019.

[6] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, 2010, pp. 1–5.

[7] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 340–354, 2013.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, p. 1–122, Jan. 2011. [Online]. Available: https://doi.org/10.1561/2200000016

[9] X. L. Hongtao Lu and J. Lv, "A fast algorithm for recovery of jointly sparse vectors based on the alternating direction methods," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, vol. 15, 2011, pp. 461–469.

[10] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[11] J. R. Hershey, J. L. Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," 2014. [Online]. Available: https://arxiv.org/abs/1409.2574

[12] 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, 2021.

[13] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," 2020. [Online]. Available: https://arxiv.org/abs/2012.08405

[14] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, p. 399–406.

[15] A. P. Sabulal and S. Bhashyam, "Joint sparse recovery using deep unfolding with application to massive random access," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5050–5054.

[16] J. Johnston and X. Wang, "Model-based neural networks for massive and sporadic connectivity," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 964–969.

[17] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[18] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, June 2018.

[19] M. Borgerding, P. Schniter, and S. Rangan, "Amp-inspired deep networks for sparse linear inverse problems," *IEEE Trans. Signal Process.*, vol. 65, no. 16, p. 4293–4308, Aug 2017.

[20] S. Li, W. Zhang, Y. Cui, H. V. Cheng, and W. Yu, "Joint design of measurement matrix and sparse support recovery method via deep auto-encoder," *IEEE Signal Process. Lett.*, vol. 26, no. 12, pp. 1778–1782, 2019.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[22] C. R. Srivatsa and C. R. Murthy, "Throughput analysis of PDMA/IRSA under practical channel estimation," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2019, pp. 1–5.

[23] B. Clerckx, G. Kim, and S. Kim, "Correlated fading in broadcast mimo channels: Curse or blessing?" in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, 2008, pp. 1–5.

[24] J. Choi and D. J. Love, "Bounds on eigenvalues of a spatial correlation matrix," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1391–1394, 2014.