

SUBMITTED VERSION

Suwichaya Suwanwimolkul, Lei Zhang, Damith C. Ranasinghe, Qinfeng Shi
One-step adaptive Markov random field for structured compressive sensing
Signal Processing, 2019; 156:116-144

© 2018 Elsevier B.V. All rights reserved.

Published at: <http://dx.doi.org/10.1016/j.sigpro.2018.10.020>

PERMISSIONS

<https://www.elsevier.com/about/policies/sharing>

Preprint

- Authors can share their preprint anywhere at any time.
- If accepted for publication, we encourage authors to link from the preprint to their formal publication via its Digital Object Identifier (DOI). Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
- Authors can update their preprints on arXiv or RePEc with their accepted manuscript .

Please note:

- Some society-owned titles and journals that operate double-blind peer review have different preprint policies. Please check the journals Guide for Authors for further information
- Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles.

6 May 2021

<http://hdl.handle.net/2440/118396>

One-step Adaptive Markov Random Field for Structured Compressive Sensing

Suwichaya Suwanwimolkul^{*,a}, Lei Zhang^{a,b}, Damith C. Ranasinghe^a, Qinfeng Shi^a

^a*School of Computer Science, The University of Adelaide, South Australia*

^b*School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, China*

Abstract

Recently, Markov random fields (MRFs) have gained much success in sparse signal recovery. One of the challenges is to adaptively estimate the MRF parameters from a few compressed measurements in compressive sensing (CS). To address this problem, a recently developed method proposes to estimate the MRF parameters based on the point estimation of sparse signals. However, the point estimation cannot depict the statistical uncertainty of the latent sparse signal, which can result in inaccurate parameters estimation; thus, limiting the ultimate performance. In this study, we propose a one-step MRF based CS that estimates the MRF parameters from the given measurements through solving a maximum marginal likelihood (MML) problem. Since the marginal likelihood is obtained from averaging over the latent sparse signal population, it offers better generalization over all the latent sparse signals than the point estimation. To solve the MML problem effectively, we approximate the MRF distribution by the product of two simpler distributions, which enables to produce closed-form solutions for all unknown variables with low computational cost. Extensive experiments on a synthetic and three real-world datasets demonstrate the effectiveness of the proposed method in recovery accuracy, noise tolerance, and runtime.

^{*}Corresponding author

Key words: image processing, compressive sensing, signal reconstruction

2000 MSC: 94A08, 68U10

1. Introduction

Compressive sensing (CS) provides an advanced sampling strategy to acquire a high-dimensional signal at a sub-Nyquist rate. It has been the core of new signal acquisition and compression systems, *e.g.* imaging and radar systems, data
5 compression, and telecommunications. To realize the sub-Nyquist rate sampling, CS aims at recovering a sparse signal \mathbf{x} of high dimensionality N from a few noisy, linear measurements \mathbf{y} of size M (*e.g.*, $M \ll N$), *i.e.* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ where $\mathbf{A} \in \mathcal{R}^{M \times N}$ is a measurement matrix and \mathbf{n} is the small perturbation. The sparse
10 signal recovery often requires an appropriate prior, such as the sparsity structure of the sparse signal, to achieve good reconstruction [1, 2]. Recently, Markov random fields (MRFs) have been used in the recovery of sparse signal in CS to achieve state-of-the-art performance [1–10]. In these works, MRFs represent the structure of signals with a graphical model where a Boltzmann machine (BM) is used as the probability distribution because of its ability to model different signal
15 distributions. The parameters of the BM and the underlying graph structure of the MRFs are estimated from extensive training examples. However, the performance of these MRFs is constrained by the amount of information in these training examples, as the BM parameters and the underlying graph of the MRF cannot adapt to new signal structures.

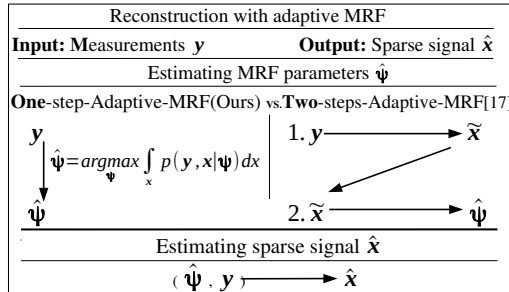
20 To address the lack of adaptiveness, a line of research [11–21] resorts to data-adaptive models without the necessity for training. The majority of these approaches employ clustered sparsity models [11–19] where a mixture model such as beta-Bernoulli is used to model signal distribution, because the closed-form updates for the model parameters are available. However, these works are limited

25 to signals with the assumed clustered structure only, *i.e.*, the signal coefficients group in clusters; hence, the clustered sparsity models are not as flexible as the MRFs. The other approaches [20, 21] consider developing data-adaptive MRF priors. In particular, the work in [20] models signal structure with an MRF, but the MRF contains only the pairwise potentials of the BM. Although the parameters of pairwise potentials can be adaptively estimated, the underlying graph of this MRF is fixed and cannot adapt for new structures.

Recently, the work in [21] proposes to adaptively adjust both the BM parameters and the underlying graph of the MRF where the full BM with both pairwise and unary potentials is employed to model signal distribution. Consequently, this adaptive MRF-based approach offers the higher flexibility to capture and adapt to any signal structure, compared to all the previous approaches [3–20]. To adaptively estimate an MRF for a signal structure, this method employs two major estimation steps—i) sparse signal estimation, and ii) based on the resulting sparse signal, the MRF parameters estimation which includes the *BM parameters* and the *underlying graph* of MRF estimations. Hence, we refer to this method as *Two-steps-Adaptive MRF*. However, Two-steps-Adaptive MRF has two main problems :

- (i) The estimated MRF parameters do not always capture the underlying structure of the latent sparse signals: The MRF parameter estimation is only based on the point estimation of the latent sparse signal population. However, the point estimation cannot depict the statistical uncertainty of the latent sparse signals.
- (ii) High computational cost: The Two-steps-Adaptive MRF iteratively performs the two estimation steps, MRF estimation and signal estimation, until convergence. Thus, the total cumulative computational cost is high.

To address these problems, we propose to take a Bayesian approach to provide



$\tilde{\mathbf{x}}$ denotes the intermediate estimate of the sparse signal;
 $\hat{\boldsymbol{\psi}}$ is the estimated MRF parameters.

Figure 1: Comparison between the two frameworks. Our One-step-Adaptive MRF directly estimates the parameters from measurements based on Bayesian estimation, while the Two-step-Adaptive MRF [21] estimates the parameters based on the point estimation of sparse signal.

a better generalization over the population of latent sparse signals. This process is shown in Figure 1. Our approach captures the statistical uncertainty by considering the marginal likelihood for the MRF parameters given the measurements.

55 The marginal likelihood is obtained by integrating out all the unknowns, which can be seen as *weighted averaging* with the probability of each variation of sparse signals. Thus, this offers better generalization over the latent sparse signals than the point estimation. As the latent sparse signals are integrated out, the MRF parameters are estimated directly from the measurements in one-step. Thus, our
 60 method is referred to as *One-step-Adaptive MRF*.

To implement this, we, first, approximate the BM with a new MRF distribution which is the product of two simpler priors to enable a closed-form update for MRF parameter estimation. The two priors are Bernoulli model [16] and pairwise MRF [20]. The Bernoulli model represents the bias toward zero for each
 65 signal coefficient, while pairwise MRF represents the correlation between these coefficients. Then, the parameters of the new MRF distribution are estimated from solving a maximum marginal likelihood (MML) problem. More importantly, the estimation of all the unknown variables resulted from the MML problem gains closed-form updates with low computational cost.

70 Figure 2 compares the effectiveness in signal recovery and the MRF param-

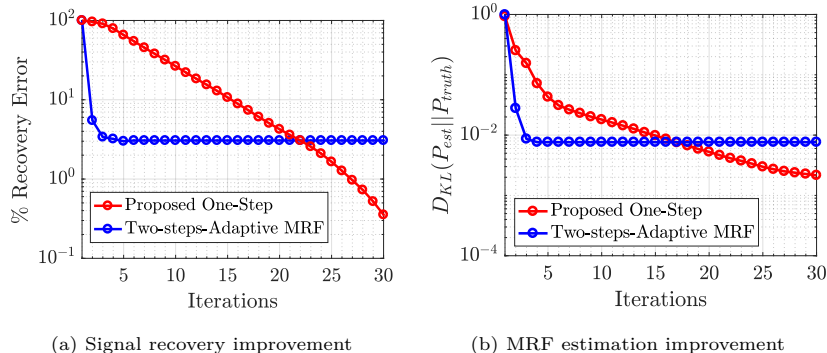


Figure 2: Performance comparison between (proposed) One-step Adaptive MRF and Two-steps Adaptive MRF in (A) signal recovery and (B) MRF parameter estimation. The performance is measured by the KL-divergence with respect to the ground truth distribution.

eters estimation between them. The effectiveness is evaluated based on 1000 synthesized sparse signals sampled from a known distribution. The accuracy of MRF parameters estimation is measured by the KL-divergence with respect to the ground truth. As our One-step-Adaptive MRF takes the Bayesian approach, it offers a better generalization over latent sparse signals. Therefore, it can further minimize the recovery error and KL-divergence than Two-steps-Adaptive MRF. Meanwhile, Two-steps-Adaptive MRF is based on the point estimation of sparse signals; thus, it lacks the generalization over the latent sparse signals, which can result in early convergence. Extensive experiments demonstrate the superior performance of our method (see Section 6).

In summary, this study makes the following contributions:

1. We propose a new MRF distribution that approximates the Boltzmann machine (BM) of MRFs to enable closed-form updates for the MRF parameters with a low computational cost. To achieve this, our distribution is the product between a *Bernoulli* model and a *pairwise MRF*. We provide both theoretical and empirical results showing that our proposed MRF distribution can well approximate the BM (see Section 5.1) The proposed MRF distribution can achieve the best approximation to BM as compared

to using the Bernoulli model [16] or the pairwise MRF [20] alone (see
90 Section 6.4).

2. With our MRF distribution, we propose One-step-adaptive MRF to better generalize the latent sparse signals, by solving the maximum marginal likelihood (MML) problem to obtain the MRF parameters from given measurements. The marginal likelihood offers the better generalization over
95 the latent sparse signals. We employ a variational expectation maximization (EM) [22] to efficiently solve the MML problem. Thus, we improve (i) the generalization in MRF estimation and (ii) the runtime as the estimation for all the unknowns gains closed-form updates (see Section 6.5).
3. We demonstrate state-of-the-art recovery performance on three benchmark
100 datasets: i) MNIST, ii) CMU-IDB, and iii) CIFAR-10 images in terms of recovery accuracy, noise tolerance, and runtime performance (see Section 6).

2. Related works

Our one-step-adaptive MRF is related to the following lines of research, but moves beyond their respective limitations.

105 **MRF-based approaches** [3–10] employ Markov random fields (MRF) for their flexibility to model various types of signal structures. In these approaches, BM parameters and the underlying graph of MRFs are learned from training data. By using the trained MRF as a prior in signal recovery, support estimation is computationally demanding in general. Different approaches [3–10] aim to
110 reduce the computation in the support estimation. Nevertheless, the trained MRFs are effective only when the structure of testing data are similar to those of training data. Unlike these approaches, our work has the mechanism to adapt the MRFs to effectively exploit different signal structures.

Clustered-sparsity based approaches [11–19] use clustered sparsity mod-
 els to capture the structure of sparse signals. This sparsity models represent
 115 the clustering of non-zero coefficients in sparse signals with probabilistic mod-
 els such as Gaussian-Bernoulli [11, 13, 16] or Gaussian-inverse Gamma [14, 15]
 [17, 18]. These probabilistic models often employ Gaussian distribution to model
 the distribution of the amplitude of sparse signal coefficients given the structure
 120 of non-zero coefficients. Meanwhile, the structure of non-zero coefficients is cap-
 tured by a prior probability distribution such as inverse Gamma and Bernoulli
 distribution. The model parameters of these mixture models can be estimated
 with EM algorithms which result in closed-form formulations. However, due to
 the limited structure assumption, the underlying structure of clustered sparsity
 125 models is fixed and cannot adapt to an actual sparse signal structure. On the
 contrary, our method uses an MRF able to capture any structures, and its
 underlying graph can be adapted for actual signals. **Adaptive MRF based
 approaches** [20, 21] improve the performance of MRF-based approaches [3–10]
 to be able to adapt the MRF for any signal structure. Wang *et al.* [20] employs
 130 an MRF, but the BM of the MRF contains only pairwise potentials to enable
 closed-form formulations to update the model parameters. This MRF and the
 method in [20] are called *pairwise MRF* in this paper. However, the pairwise
 MRF has two limitations: (i) it is not as flexible as the MRF with the full BM;
 and (ii) its underlying graph is fixed and cannot adapt to a new structure. Two-
 135 steps-Adaptive MRF [21] is more flexible than other models [11–20], because it
 uses an MRF with a full BM and adaptively adjust both the BM parameters and
 the underlying graph. With many unknown parameters, this method iteratively
 performs sparse signal estimation and MRF parameter estimation based on the
 point estimate of the sparse signal, until convergence. However, this can lead to
 140 high computational cost. Also, the point estimation cannot depict the statistical

uncertainty of the latent sparse signals. Our research objective shares a similar spirit to [21]; however, we avoid these problems by enabling close-form updates for all the unknown variables and improving generalization over the latent sparse signals.

145 In the following, we present the observation model for graphical CS in Section 3. Signal modelling with our proposed MRF distribution is presented in Section 3.1. We present One-step-Adaptive MRF to estimate the MRF parameters in Section 4. Then, we provide the theoretical result to show how well the proposed MRF distribution can approximate the BM and the algorithm
150 complexity in Section 5. To this end, we evaluate the performance of the proposed method and the proposed MRF distribution by experiments in Section 6.

3. Graphical compressive sensing

Inspired by [11–14, 16], we decompose the sparse signal $\mathbf{x} \in \mathcal{R}^N$ into a support vector $\mathbf{s} \in \{0, 1\}^N$ with a scale vector $\mathbf{t} \in \mathcal{R}^N$, which can be denoted as $\mathbf{x} = \mathbf{t} \odot \mathbf{s}$. The support vector \mathbf{s} indicates the position of non-zero coefficients in the sparse signal \mathbf{x} . Thus, our goal is to recover \mathbf{t} and \mathbf{s} from the linear observation model

$$\mathbf{y} = \mathbf{A}(\mathbf{t} \odot \mathbf{s}) + \mathbf{n}. \quad (1)$$

The small perturbation \mathbf{n} is modelled with additive Gaussian white noise with the noise precision σ_n^{-1} . Thus, the corresponding observation likelihood is

$$p(\mathbf{y}|\mathbf{t}, \mathbf{s}; \sigma_n) = \mathcal{N}(\mathbf{A}(\mathbf{t} \odot \mathbf{s}), \sigma_n^{-1} \mathbf{I}), \quad (2)$$

where \mathbf{I} is an identity matrix with proper size. Generally, given appropriate priors, $p(\mathbf{s})$ and $p(\mathbf{t})$, the latent \mathbf{s} and \mathbf{t} are inferred by solving the MAP problem

$$\{\hat{\mathbf{s}}, \hat{\mathbf{t}}\} = \max_{\mathbf{s}, \mathbf{t}} p(\mathbf{t}, \mathbf{s} | \mathbf{y}) \propto (\mathbf{y} | \mathbf{t}, \mathbf{s}) p(\mathbf{t}) p(\mathbf{s}). \quad (3)$$

To represent the flexible distribution of the sparse signal coefficients, a Markov random field (MRF) prior is imposed on the support \mathbf{s} . In connection with this, a statistical model is imposed on the signal scale \mathbf{t} . In the following section, we will discuss the models $p(\mathbf{s})$ and $p(\mathbf{t})$, respectively.

3.1. Markov random field based support prior

This section, first, reviews Boltzmann machine, the commonly used probability distribution in MRFs and, then, presents our MRF distribution.

Since MRFs are flexible and expressive enough to model complex dependency, the majority of the existing works [3–10, 20, 21] employ the MRF to capture the underlying structure of a sparse representation through its support \mathbf{s} . The MRF represents the dependency between support coefficients by defining the probability distribution over an undirected graph. Let $\mathcal{G} = \{V, E\}$ denotes the underlying undirected graph of the MRF, where V and E is the set of nodes and undirected edges in \mathcal{G} . Each coefficient is mapped one-to-one to a node in the graph \mathcal{G} . The probability distribution is defined as Boltzmann machine (BM):

$$p(\mathbf{s}) = \frac{1}{Z} \prod_c \prod_{i \in \mathbb{N}_c} \exp(s_i \delta_i^c + s_i \sum_{j \in \mathcal{E}_i} \gamma_{ij}^c s_j) \quad (4)$$

where $Z(\cdot)$ is a normalizing constant; $\{\delta_i^c, \gamma_{ij}^c\}$ are local parameters that model the interaction among signal coefficients. δ_i^c defines bias toward zero for each s_i . γ_{ij}^c weights the dependency between s_i and its adjacent s_j which is defined by the local edge set \mathcal{E}_i where the edge set $E = \{\mathcal{E}_i\}_{i \in V}$. The neighborhood set \mathbb{N}_c defines how these parameters are shared among support coefficients.

An important key for applying the MRFs is to estimate the parameters $\{\delta_i^c, \gamma_{ij}^c\}$ in Eq. (4). Generally, the parameters of the MRFs are learned from

the training data, but the learned model cannot adapt for new signal structures. The work [21] proposes to adaptively estimate the parameters based on a point estimate of the sparse signal. However, both the parameter and sparse signal
170 estimations are performed in every iteration, which requires high computation.

To address this problem, we propose to approximate the BM Eq.(4) with a new probability distribution. Inspired by [23], we assume conditional independence between each node given its adjacent nodes. Thus, the joint distribution is written as the products of conditional probabilities. Then, we approximate each conditional probability distribution with the product of two simpler distributions. Each of them corresponds to the unary and pairwise potentials in the BM distribution. The proposed MRF distribution for support \mathbf{s} is given as

$$p(\mathbf{s}) = \prod_c \prod_{i \in \mathbb{N}_c} p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^c), \quad (5)$$

$$\text{where } \log(p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^c)) \propto \phi_u(s_i | \theta_i^u) + \phi_p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^p).$$

$\phi_u(s_i | \theta_i^u) = \log(p_u(s_i | \theta_i^u))$ and $\phi_p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^p) = \log(p_p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^p))$. Here, $p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^c)$ is the conditional distribution of a support s_i given $\mathbf{s}_{\mathcal{E}_i}$ where $\mathbf{s}_{\mathcal{E}_i} = [s_j]_{j \in \mathcal{E}_i}$ contains the support coefficients connected to the node s_i with the edges specified by \mathcal{E}_i . It is approximated with the product of $p_u(s_i | \theta_i^u)$ and $p_p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^p)$
175 which are associated with the unary $\phi_u(\cdot)$ and pairwise $\phi_p(\cdot)$ potentials. In the following, we will introduce the specific forms of $p_u(s_i | \theta_i^u)$ and $p_p(s_i | \mathbf{s}_{\mathcal{E}_i}, \theta_i^p)$.

Unary potential. To control local sparsity in a fixed-size neighboring region, we employ the Bernoulli model [16] where every support coefficient in the neighboring region shares a common parameter b_c , *i.e.*,

$$p_u(s_i | b_i) = \text{Bernoulli}(s_i | b_i) \quad \text{with} \quad b_i = b_c \sim \text{Beta}(\alpha, \beta) \quad \forall i \in \mathbb{N}_c. \quad (6)$$

b_c defines the tendency toward non-zero according to the setting of α and β .

The distribution $p_u(s_i)$ alone reflects the bias toward zero on support coefficients within a neighborhood, but it cannot reflect the interaction between them. The
180 connection between Bernoulli distribution and the BM is explained in details in .1.

Pairwise potential. To reflect interaction between support coefficients, we employ the pairwise MRF [20], where the connection between the i^{th} support coefficient and the other coefficients is defined by \mathcal{E}_i . The pairwise MRF is

$$p_p(s_i|\mathbf{s}_{\mathcal{E}_i}; w_i) = \frac{1}{Q(w_i, \mathbf{s}_{\mathcal{E}_i})} \exp(s_i \sum_{j \in \mathcal{E}_i} w_i s_j), \quad (7)$$

where w_i weights the dependency between s_i and other coefficients $\mathbf{s}_{\mathcal{E}_i}$, and $Q(w_i, \mathbf{s}_{\mathcal{E}_i}) = 2 \cosh(w_i \sum_{j \in \mathcal{E}_i} s_j)$. The edge set $E = \{\mathcal{E}_i\}$ defines the pairwise connection between nodes in the entire underlying graph \mathcal{G} .

With the defined probability distributions associated with the unary and pairwise potentials, we represent our MRF distribution of \mathbf{s} as

$$p(\mathbf{s}|\mathbf{b}, \mathbf{w}) \propto \prod_c \prod_{i \in \mathbb{N}_c} p_u(s_i; b_i) p_p(s_i|\mathbf{s}_{\mathcal{E}_i}; w_i) = \prod_c p(\mathbf{s}_{\mathbb{N}_c} | b_c, \mathbf{w}_{\mathbb{N}_c}), \quad (8)$$

185 where $p_u(s_i; b_i) = \text{Bernoulli}(s_i|b_i)$ with $b_i = b_c \sim \text{Beta}(\alpha, \beta) \forall i \in \mathbb{N}_c$ and $p_p(s_i|\mathbf{s}_{\mathcal{E}_i}; w_i) = \frac{1}{Z(w_i, \mathbf{s}_{\mathcal{E}_i})} \exp(s_i \sum_{j \in \mathcal{E}_i} w_i s_j)$. $\mathbf{s}_{\mathbb{N}_c} = [s_i]_{i \in \mathbb{N}_c}$ and $\mathbf{w}_{\mathbb{N}_c} = [w_i]_{i \in \mathbb{N}_c}$ represent the vector of support coefficients and pairwise parameters in \mathbb{N}_c .

Because the distributions associated with the unary and pairwise potentials are separately modelled in Eq. (8), their parameters can be separately estimated.
190 This benefits simplify the following MRF parameter estimation with using a variational expectation-maximization (EM) in Section 4. The parameters of the Bernoulli model obtain a closed-form solution in inference, and the parameters of pairwise MRF are obtained by solving an maximum marginal likelihood (MML) problem, which is also resulted in a closed-form formulation. More details will

195 be further clarified in Section 4.

The proposed MRF distribution Eq. (8) can be viewed as a surrogate for the BM Eq. (4) where $\delta_i^c = \delta_c \forall i \in \mathbb{N}_c$ and $\gamma_{ij}^c = \gamma_i \forall j \in \mathcal{E}_i$. The effectiveness of our MRF distribution in approximating the BM can be measured by the Kullback-Leibler (KL) divergence between them. We can demonstrate the KL-
 200 divergence can be bounded under a small constant, under the assumption that the proposed MRF is used to model sparse signals (see Section 5.1). We also provide empirical performance in approximating the BM where the empirical KL-divergence of the proposed MRF is compared it with that of some existing approximation schemes [16, 20] in Section 6.4. The KL-divergence of our MRF
 205 distribution is smaller than those of the existing schemes. Then, we evaluate the similarity/difference between the learned MRF parameters of our proposed MRF distribution (Eq. (8)) and those of the BM (Eq. (4)) based on real data. This shows that Our MRF distribution Eq. (8) can well approximate the BM.

3.2. The signal scale prior

In connection with the support model, we impose statistical models to the signal scale coefficients in each neighborhood site. Specifically, let $\mathbf{t}_{\mathbb{N}_c} = [t_i]_{i \in \mathbb{N}_c}$ be a vector of scale coefficients in \mathbb{N}_c . We impose an iid Gaussian distribution as a prior of the scale coefficients $\mathbf{t}_{\mathbb{N}_c}$ as follows

$$p(\mathbf{t}_{\mathbb{N}_c}; \sigma_{t,c}) = \prod_{i \in \mathbb{N}_c} \mathcal{N}(t_i | 0, \sigma_{t_i}^{-1} \mathbf{I}), \quad \text{where } \sigma_{t_i} = \sigma_{t,c} \sim \text{Gamma}(\varpi, \xi). \quad (9)$$

210 $\sigma_{t,c}$ is the signal precision shared among scale coefficients in \mathbb{N}_c . Here, Gamma distribution is used a hyperprior over the hyperparameter $\sigma_{t,c}$ where ϖ and ξ are constant with appropriate settings [20, 24]. This model weakly imposes structure among $\{t_i\}_{i \in \mathbb{N}_c}$ to help controls sparsity level in addition to the unary potentials.

3.3. The hyperprior for noise precision

215 To facilitate the inference for σ_n , the Gamma prior is imposed on σ_n since the perturbation \mathbf{n} is assumed to be Gaussian white noise, *i.e.*, $\sigma_n \sim \text{Gamma}(\varpi_0, \xi_0)$.

4. One-step-Adaptive-MRF Inference

With the hyperpriors $p(b_i; \alpha, \beta)$, $p(\sigma_{t_c}; \varpi, \xi)$, and $p(\sigma_n; \varpi_0, \xi_0)$, the posterior of the unknown sparse signal scale \mathbf{t} and support \mathbf{s} , given measurements \mathbf{y} is

$$\begin{aligned} p(\mathbf{t}, \mathbf{s} | \mathbf{y}, \Theta) &\propto p(\mathbf{y} | \mathbf{t}, \mathbf{s}, \Theta) p(\mathbf{t}, \mathbf{s} | \Theta) p(\Theta) \\ &= p(\mathbf{y} | \mathbf{t}, \mathbf{s}, \sigma_n) \prod_c p(\mathbf{t}_{\mathbb{N}_c}, \mathbf{s}_{\mathbb{N}_c} | \sigma_{t_c}, b_c, \mathbf{w}_{\mathbb{N}_c}) p(\sigma_{t_c}; \varpi, \xi) p(b_c; \alpha, \beta) p(\sigma_n; \varpi_0, \xi_0) \end{aligned} \quad (10)$$

where $p(\mathbf{t}_{\mathbb{N}_c}, \mathbf{s}_{\mathbb{N}_c} | \sigma_{t_c}, b_c, \mathbf{w}_{\mathbb{N}_c}) = p(\mathbf{t}_{\mathbb{N}_c} | \sigma_{t_c}) p(\mathbf{s}_{\mathbb{N}_c} | b_c, \mathbf{w}_{\mathbb{N}_c})$;

220 $\Theta = \{\sigma_n, \sigma_t, \mathbf{b}, \mathbf{w}\}$; $\sigma_t = [\sigma_{t_c}]$, $\mathbf{b} = [b_c]$, $\mathbf{w} = [\mathbf{w}_{\mathbb{N}_c}]$. Most existing MRF-based methods [3–10] estimate these unknown Θ with training samples. However, the resulting Θ cannot adapt for actual sparse signals. The two-steps method in [21] adaptively estimates Θ based on the point estimation of sparse signals, which, however, cannot depict the statistical uncertainty of the latent sparse signal. To address this problem, we estimate Θ directly from the measurements 225 with a statistical inference process described in the following. Then, given Θ , the sparse signal is estimated by solving MAP Eq.(3).

4.1. Adaptive MRF parameter estimation with variational EM

Our objective is to adaptively estimate the unknown parameters Θ directly from measurements \mathbf{y} . With the hyperprior imposed on σ_n , σ_t , and \mathbf{b} , these unknowns can be considered as the unknown random variables; meanwhile, \mathbf{w} is the only unknown parameter. Thus, we aims to solve the following maximum

marginal likelihood (MML) problem

$$\max_{\mathbf{w}} \ln p(\mathbf{y}|\mathbf{w}) \propto \int \ln p(\mathbf{y}, \mathbf{\Lambda}|\mathbf{w}) d\mathbf{\Lambda}. \quad (11)$$

where $\mathbf{\Lambda} = \{\mathbf{t}, \mathbf{s}, \sigma_n, \sigma_t, \mathbf{b}\}$ is the set of all unknown variables. To solve this MML problem, all the unknown variables in $\mathbf{\Lambda}$ are to be integrated out. Since
 230 calculating the integral in Eq. (11) is intractable, we resort to the variational EM [22] to estimate the unknown parameters.

In the variational EM, the unknown parameters are obtained by solving a bound maximization problem. The lower bound of the log-likelihood function $\ln p(\mathbf{y}|\mathbf{w})$ is obtained by introducing a pseudo probability over the unknown variables. Here, we denote the pseudo probability as $q(\mathbf{\Lambda})$. With the pseudo probability $q(\mathbf{\Lambda})$, Eq. (11) can be written as [22]:

$$\ln p(\mathbf{y}|\mathbf{w}) = F(q, \mathbf{w}) + KL(q||p), \quad (12)$$

where $F(q, \mathbf{w}) = \int q(\mathbf{\Lambda}) \ln \frac{p(\mathbf{y}, \mathbf{\Lambda}|\mathbf{w})}{q(\mathbf{\Lambda})} d\mathbf{\Lambda}$ is the lower bound of $\ln p(\mathbf{y}|\mathbf{w})$, and
 $KL(q||p) = - \int q(\mathbf{\Lambda}) \ln \frac{p(\mathbf{\Lambda}|\mathbf{y}, \mathbf{w})}{q(\mathbf{\Lambda})} d\mathbf{\Lambda}$ is the Kullback-Leibler divergent between
 235 $p(\mathbf{\Lambda}|\mathbf{y}, \mathbf{w})$ and $q(\mathbf{\Lambda})$. Since it is always that $KL(q||p) \geq 0$, it holds that $F(q, \mathbf{w})$ is the lower bound of $\ln p(\mathbf{y}|\mathbf{w})$. This suggests that if the pseudo probability $q(\mathbf{\Lambda})$ has a very close approximation to $p(\mathbf{\Lambda}|\mathbf{y}, \mathbf{w})$, $KL(q||p)$ approaches zero. Therefore, we turn to maximize the lower bound $F(q, \mathbf{w})$, by iteratively performing [22]:

(i) Expectation: $q(\mathbf{\Lambda})$ is assumed to have a factorized form with respect to each partition: $q(\mathbf{\Lambda}) = q(\sigma_n)q(\mathbf{t})q(\mathbf{s}) \prod_c q(\sigma_{t_c}) \prod_c q(b_c)$. The optimal distribution of a latent variable Λ_p follows

$$\hat{q}(\Lambda_p) = \langle p(\mathbf{y}, \mathbf{\Lambda}|\mathbf{w}) \rangle_{q(\mathbf{\Lambda} \setminus \Lambda_p)}. \quad (13)$$

(ii) **Maximization:** Given $\hat{q}(\mathbf{\Lambda})$ from the expectation step, the unknown parameter \mathbf{w} is estimated by solving the following problem:

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} F(\hat{q}(\mathbf{\Lambda}), \mathbf{w}), \quad (14)$$

where $\langle f(\cdot) \rangle_{\mathbf{\Lambda} \setminus \Lambda_p}$ represents the expectation of $f(\cdot)$ with respect to the distribution $q(\mathbf{\Lambda} \setminus \Lambda_p)$ where $\mathbf{\Lambda} \setminus \Lambda_p$ represents the set $\mathbf{\Lambda}$ without Λ_p .
 240

As a result, each unknown variable in $\mathbf{\Lambda} = \{\mathbf{t}, \mathbf{s}, \sigma_n, \sigma_t, \mathbf{b}\}$ is calculated through approximating the true posterior $p(\mathbf{y}, \mathbf{\Lambda} | \mathbf{w})$ in Eq. (13) (Expectation step). As \mathbf{t} and \mathbf{s} are estimated in the Expectation step, there is no needed to solve MAP Eq. (3). The updating rule for each parameter in \mathbf{w} is calculated by maximizing the lower bound $F(\mathbf{q}, \mathbf{w})$ Eq. (14) (Maximization step). Due to the
 245 conditional independence assumption, each w_i can be estimated separately.

4.2. Optimization details

In this part, we give the optimization details for all unknown variables. In the following, the updates from 4.2.1 to 4.2.5 belong to the expectation step, while the update in 4.2.6 is the maximization step. To adapt for a new signal
 250 structure, we update the underlying graph which is the edge set $E = \{\mathcal{E}_i\}$ in 4.2.7. Here, we employ the graph update technique from [21] since it requires low computation.

4.2.1. Update for \mathbf{t}

Given the update parameters and variables (*i.e.*, $\hat{\sigma}_t$, $\hat{\mathbf{s}}$, and $\hat{\sigma}_n$), and according to Eq.(13), we obtain the following update equation for \mathbf{t} :

$$\hat{q}(\mathbf{t}) \propto \langle p(\mathbf{y} | \mathbf{t}, \mathbf{s}, \sigma_n) p(\mathbf{t} | \sigma_t) \rangle_{q(\mathbf{\Lambda} \setminus \mathbf{t})}. \quad (15)$$

Substituting the prior of coefficient scale \mathbf{t} Eq. (9) and the observation

likelihood Eq. (2) into Eq. (15), we obtain $\mathcal{N}(\mathbf{u}_t, \mathbf{C}_t^{-1})$ with

$$\mathbf{u}_t = \hat{\sigma}_n \mathbf{C}_t^{-1} \hat{\mathbf{S}} \mathbf{A}^T \mathbf{y}; \quad \mathbf{C}_t = \hat{\Sigma}_t + \hat{\sigma}_n \langle \hat{\mathbf{S}} \mathbf{A}^T \mathbf{A} \hat{\mathbf{S}} \rangle_{q(\mathbf{s})}, \quad (16)$$

where $\hat{\mathbf{S}} = \text{diag}(\hat{\mathbf{s}})$; $\hat{\mathbf{s}}$ is the update value of \mathbf{s} from previous iteration;
 $\langle \hat{\mathbf{S}} \mathbf{A}^T \mathbf{A} \hat{\mathbf{S}} \rangle_{q(\mathbf{s})} = \hat{\mathbf{S}} \mathbf{A}^T \mathbf{A} \hat{\mathbf{S}} + (\mathbf{A}^T \mathbf{A} \odot \text{diag}(\hat{\mathbf{s}} \odot (1 - \hat{\mathbf{s}})))$; and $\hat{\Sigma}_t = \text{diag}(\hat{\sigma}_t)$.
The calculation of means and covariance of \mathbf{t} are similar to those of [11–14, 16],
as the support and the scale vector are decomposed from the sparse signal in
the same manner. Thus, the update for \mathbf{t} :

$$\hat{\mathbf{t}} = \mathbf{u}_t. \quad (17)$$

255 4.2.2. Update for \mathbf{s}

Given $\hat{\sigma}_t$, $\hat{\sigma}_n$, $\hat{\mathbf{t}}$ and $\hat{\mathbf{w}}$, and $\hat{\mathbf{s}}$, the log posterior probability of each s_i is

$$\hat{q}(s_i) \propto \langle p(\mathbf{y}|\mathbf{t}, \mathbf{s}, \sigma_n) p(\mathbf{s}|\mathbf{w}, \mathbf{b}) \rangle_{q(\mathbf{A} \setminus s_i)}. \quad (18)$$

The probability when $s_i = 1$ is given as

$$\begin{aligned} \ln \hat{q}(s_i = 1) &\propto -\frac{\sigma_n}{2} (\mathbf{y}^T \mathbf{y} + \langle t_i^2 \rangle \mathbf{a}_i^T \mathbf{a}_i - 2\hat{t}_i \mathbf{a}_i^T (\mathbf{y} - \sum_{j \neq i} \mathbf{a}_j \hat{t}_j \hat{s}_j)) + w_i \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j \\ &\quad - \ln(2 \cosh(w_i \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j)) + \langle \ln(p_u(s_i = 1|b_c)) \rangle_{q(b_c)}, \end{aligned} \quad (19)$$

The probability when $s_i = 0$ is given as

$$\ln \hat{q}(s_i = 0) \propto -w_i \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j - \ln(2 \cosh(w_i \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j)) + \langle \ln(p_u(s_i = 0|b_c)) \rangle_{q(b_c)}, \quad (20)$$

where $\hat{z}_i = 2\hat{s}_i - 1$, and $\langle t_i^2 \rangle \propto \hat{t}_i^2 + \text{var}(\hat{t}_i)$. $\text{var}(\hat{t}_i)$ is the variance of \hat{t}_i which is from Eq. (16), *i.e.*, $\text{var}(\hat{t}_i) = \text{diag}\{\text{inv}(\mathbf{C}_t)\}_{i,i}$. The update for $\langle \ln(p_u(s_i = 1)) \rangle_{q(b_i)}$

and $\langle \ln(p_u(s_i = 0)) \rangle_{q(b_i)}$ are in Eq. (23). The update for s_i is:

$$\hat{s}_i = \frac{\hat{q}(s_i = 1)}{\hat{q}(s_i = 1) + \hat{q}(s_i = 0)} \quad (21)$$

Then, update $\hat{z}_i = 2\hat{s}_i - 1$ and $\hat{\mathbf{x}} = \hat{\mathbf{t}} \odot \hat{\mathbf{s}}$.

4.2.3. Update for $p_u(s_i|b_c)$

$$\hat{q}(b_c) \propto \langle \prod_j \prod_{i \in \mathbb{N}_j} p_u(s_i|b_j)p(b_j|\alpha, \beta) \rangle_{q(\mathbf{\Lambda} \setminus b_j=c)} \propto \text{Beta}(\hat{\alpha}, \hat{\beta}), \quad (22)$$

which performs expectation over all unknown random variables, except every term that involves with b_c . Since Bernoulli and Beta distributions are conjugate pair, the posterior hyperparameters in Eq. (22) are given as $\hat{\alpha} = \alpha + \sum_{i \in \mathbb{N}_c} \hat{s}_i$, and $\hat{\beta} = \beta + |\mathbb{N}_c| - \sum_{i \in \mathbb{N}_c} \hat{s}_i$. Thus, we have

$$\begin{aligned} \ln(p_u(s_i = 1|b_c))_{q(b_c)} &= \psi(\hat{\alpha}) - \psi(\hat{\alpha} + \hat{\beta}) \\ \langle \ln(p_u(s_i = 0|b_c)) \rangle_{q(b_c)} &= \psi(\hat{\beta}) - \psi(\hat{\alpha} + \hat{\beta}), \end{aligned} \quad (23)$$

where $\psi(x) = (d/dx) \ln \Gamma(x)$.

4.2.4. Update for σ_{t_c}

The update of σ_{t_c} is obtained as follows:

$$q(\sigma_{t_c}) \propto \langle \prod_j \prod_{i \in \mathbb{N}_j} p(t_i|\sigma_{t_j})p(\sigma_{t_j}|\varpi, \xi) \rangle_{q(\mathbf{\Lambda} \setminus \sigma_{t_j=c})} \propto \text{Gamma}(\hat{\varpi}, \hat{\xi}). \quad (24)$$

As Gaussian and Gamma distributions are conjugate pair, the posterior hyperparameters in Eq. (24) are given as $\hat{\varpi} = \varpi + \frac{|\mathbb{N}_c|}{2}$, and $\hat{\xi} = \xi + \frac{\sum_{i \in \mathbb{N}_c} (t_i^2 + \text{var}(t_i))}{2}$.

The update for σ_{t_c} is therefore: $\forall i \in \mathbb{N}_c$,

$$\hat{\sigma}_{t_i} = \hat{\sigma}_{t_c} = \frac{\hat{\varpi}}{\hat{\xi}}. \quad (25)$$

260 Then, $\hat{\Sigma}_t = \text{diag}(\hat{\sigma}_t)$ where $\hat{\sigma}_t = [\hat{\sigma}_{t1}, \dots, \hat{\sigma}_{tN}]$.

4.2.5. Update for σ_n

Given $\hat{\mathbf{t}}$ and $\hat{\mathbf{s}}$, the update for σ_n is obtained according to Eq. (14)

$$\hat{q}(\sigma_n) \propto \langle p(\sigma_n | \varpi_0, \xi_0) p(\mathbf{y} | \mathbf{t}, \mathbf{s}, \sigma_n) \rangle_{q(\mathbf{A} \setminus \sigma_n)} \propto \text{Gamma}(\hat{\varpi}_0, \hat{\xi}_0) \quad (26)$$

where the hyperparameters of the posterior distribution are given as $\hat{\varpi}_0 = \varpi_0 + \frac{M}{2}$ and $\hat{\xi}_0 = \xi_0 + \frac{\langle \|\mathbf{y} - \mathbf{A}(\hat{\mathbf{t}} \odot \hat{\mathbf{s}})\|^2 \rangle_{q(\hat{\mathbf{t}}, q(\hat{\mathbf{s}}))}}{2}$. Here, the expectation $\langle \|\mathbf{y} - \mathbf{A}(\hat{\mathbf{t}} \odot \hat{\mathbf{s}})\|^2 \rangle_{q(\hat{\mathbf{t}}, q(\hat{\mathbf{s}}))} = \mathbf{y}^T \mathbf{y} - 2(\hat{\mathbf{t}} \odot \hat{\mathbf{s}})^T \mathbf{A}^T \mathbf{y} + \mathbf{1}^T [\langle \mathbf{s} \mathbf{s}^T \rangle \odot \langle \mathbf{t} \mathbf{t}^T \rangle \odot \langle \mathbf{A}^T \mathbf{A} \rangle] \mathbf{1}$, where $\langle \mathbf{s} \mathbf{s}^T \rangle = \hat{\mathbf{s}} \hat{\mathbf{s}}^T + \text{diag}(\hat{\mathbf{s}} \odot (\mathbf{1} - \hat{\mathbf{s}}))$ and $\langle \mathbf{t} \mathbf{t}^T \rangle = \hat{\mathbf{t}} \hat{\mathbf{t}}^T + \hat{\Sigma}_t$, and $\hat{\Sigma}_t = \text{diag}(\hat{\sigma}_t)$. The update for σ_n is therefore:

$$\hat{\sigma}_n = \frac{\hat{\varpi}_0}{\hat{\xi}_0}. \quad (27)$$

4.2.6. Update for \mathbf{w}

Give the updated $\hat{\mathbf{z}}$, w_i is estimated by solving

$$\begin{aligned} \hat{w}_i &= \arg \max_{w_i} \langle \ln p_p(z_i | z_{\hat{\mathcal{E}}_i}, w_i) \rangle_{q(z_i)} \\ &\equiv \hat{z}_i w_i \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j - \ln(\exp(w_i \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j) + \exp(-w_i \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j)). \end{aligned} \quad (28)$$

Take gradient with respect to w_i , and equate it to zero, the update of w_i is

$$\hat{w}_i = \frac{1}{2 \sum_{j \in \hat{\mathcal{E}}_i} \hat{z}_j} \ln \left(\frac{1 + \hat{z}_i}{1 - \hat{z}_i} \right). \quad (29)$$

4.2.7. Update edge set \hat{E} :

Inspired by [16, 21], we can update the underlying graph (*i.e.*, edges set) constructed based on the non-zero coefficients. Since \hat{s}_i has a continuous value, *i.e.* $\hat{s}_i \in [0, 1]$, the binary support vector for $\hat{\mathbf{s}}$ is obtained by performing thresholding over $\hat{\mathbf{s}}$. Let \mathbf{d} represent a binary vector corresponding to $\hat{\mathbf{s}}$. We assign binary

Algorithm 1: Update edge set \hat{E}

Initialization: $\hat{E} = \emptyset$.

Input : Binary support vector \mathbf{d}

1. Set each entries in \mathbf{d} being nodes in a graph \mathcal{G}
2. For every node from $i = 1$ to N , find the adjacent nodes of the i^{th} node located within its neighborhood \mathbb{N} that are non-zero;
3. Establish edges from the i^{th} node to these non-zero nodes, and then, collect these edges in $\hat{\mathcal{E}}_i$. If adjacent nodes are all zeros, $\hat{\mathcal{E}}_i$ is empty.

Output : $\hat{E} = \{\hat{\mathcal{E}}_i\}_{i=1}^N$.

Algorithm 2: One-step-Adaptive MRF (OA-MRF)

Input : A measurement signal \mathbf{y} , \mathbf{A} , $\{\mathcal{E}_i\}_{initialized}$.

Initialization: $\mathbf{\Sigma}_t = \mathbf{I}_{N \times N}$, $\sigma_n = 1$, $\mathbf{s} = \mathbf{1}$, $\mathbf{w} = \mathbf{0}$ and $\mathbf{t} = \mathbf{0}$;

while a stopping criterion is not satisfied **do**

1. Update $\hat{\mathbf{t}}$ as Eq. (17) ;
2. Update $\hat{\mathbf{s}}$ by Eq. (21) ;
3. Update $\hat{\mathbf{b}}$ as Eq. (23) ;
4. Update $\hat{\sigma}_t$ as Eq. (25) ;
5. Update $\hat{\sigma}_n$ as Eq. (27) ;
6. Update $\hat{\mathbf{w}}$ as Eq. (29) ;
7. Update the edge set $\{\hat{\mathcal{E}}_i\}_{i=1}^N$;

end

Output : Recovered $\mathbf{x} = \mathbf{t} \odot \mathbf{s}$

value to \mathbf{d} by Eq. 30 where the logic '0' is assigned to d_i , if the coefficient \hat{s}_i has a negligible value. The logic '1' is assigned to d_i , if otherwise. The value of s_i is considered as being negligible, if absolute value of s_i is less than a threshold T_s . Here, we set T_s to a mean absolute value of coefficients in $\hat{\mathbf{s}}$, *i.e.* $T_s = \frac{1}{N} \sum_1^N \text{abs}(s_i)$, where $\text{abs}(s_i)$ denotes the absolute value of s_i . We update \mathbf{d} as follows:

$$\hat{d}_i = \begin{cases} 1, & \text{if } \text{abs}(\hat{s}_i) > T_{\hat{\mathbf{s}}} \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

With the binary vector \mathbf{d} , each of the binary coefficients is mapped to a node in the graph \mathcal{G} , and each edge is connected from one node to other non-zero nodes within a neighboring region \mathbb{N} . The procedure is summarized in Algorithm 1.

Solving Eq. (11) with the EM [22] is summarized in Algorithm 2. The update rules for calculating Expectation-Maximization steps are performed iteratively until convergence. In most cases, the convergence of the EM is guaranteed.

270 **5. Theoretical result.**

5.1. *How well the proposed MRF distribution can approximate the BM.*

In this section, we provide a theoretical result to demonstrate how well the proposed MRF distribution (Eq. (8)) can approximate the original BM (Eq. (4)) based on the KL-divergence. We can show that the KL-divergence between the
275 proposed MRF distribution and the BM can be bounded under a small value, under the assumption that the proposed MRF distribution models sparse signals.

According to [6], a sparse coefficient can be modelled with the BM Eq. (8) under an assumption that $p(s_i = 1) \ll 0$. Therefore, the unary parameters b_i is negative, and the number of edges in the underlying graph of the BM Eq. (8) is often small [7]. Let q and p denote the probability of a support with the BM Eq. (8) and the proposed MRF distribution Eq. (4). Then, we have that

$$\begin{aligned}
 q_1 &= \frac{\exp(b_i + \sum_{j \in \mathcal{E}_i} w_j s_j)}{\exp(b_i + \sum_{j \in \mathcal{E}_i} w_j s_j) + 1}; \text{ and } q_0 = \frac{1}{\exp(b_i + \sum_{j \in \mathcal{E}_i} w_j s_j) + 1}. \\
 p_1 &= \frac{\exp(b_i + \sum_{j \in \mathcal{E}_i} w_j s_j)}{(\exp(b_i) + 1)(\exp(\sum_{j \in \mathcal{E}_i} w_j s_j) + 1)}; \text{ and} \\
 p_0 &= \frac{1 + \exp(b_i) + \exp(\sum_{j \in \mathcal{E}_i} w_j s_j)}{(\exp(b_i) + 1)(\exp(\sum_{j \in \mathcal{E}_i} w_j s_j) + 1)}.
 \end{aligned}$$

The KL-divergence between the proposed MRF distribution and the BM is

$$\begin{aligned}
KL(p||q) &= p_1 \log\left(\frac{p_1}{q_1}\right) + p_0 \log\left(\frac{p_0}{q_0}\right) \\
&= p_1 \log\left(\frac{1 + \exp(b_i + \sum_j w_j s_j)}{(\exp(b_i) + 1)(\exp(\sum_{j \in \mathcal{E}_i} w_j s_j) + 1)}\right) + \\
&\quad p_0 \log\left(1 + \exp(b_i) + \exp\left(\sum_{j \in \mathcal{E}_i} w_j s_j\right)\right).
\end{aligned} \tag{31}$$

The first term is negative; thus, the KL-divergence is bounded by the second term, *i.e.*,

$$KL(p||q) \leq p_0 \log\left(1 + \exp(b_i) + \exp\left(\sum_{j \in \mathcal{E}_i} w_j s_j\right)\right) \tag{32}$$

With the sparsity assumption, the unary parameters b_i is negative, and the number of edges in the underlying graph of the BM Eq. (4) is often small [7]. Thus, it is always that $\exp(-|b_i|) \leq 1$. $\exp(\sum_{j \in \mathcal{E}_i} w_j s_j)$ can be bounded to a small constant under the assumption that the number of edges are small. For
280 example, if $|\mathcal{E}_i| = 0$, then $\exp(\sum_{j \in \mathcal{E}_i} w_j s_j) = 1$. Therefore, $KL(p||q)$ can be bounded by a small constant.

It is worth mentioning that we can draw a similar conclusion from the experiment result in Section 6.4 where we studied the empirical KL-divergence
285 of the proposed MRF distribution with respect to the BM. Our experiment shows that the empirical KL-divergence increases with the edges and the sparsity levels. Nevertheless, the KL-divergence of the proposed MRF Eq. (8) is much smaller in comparison with the KL-divergence of the existing approximation schemes [16, 20] that use the Bernoulli or the pairwise MRF model alone.

290 *5.2. Computational complexity*

All the update steps in Algorithm 2 are in closed-form solutions. Most of which requires matrix-vector productions. The matrix inversion in Eq. (16) with the computational cost of $\mathcal{O}(N^3)$ dominates other costs. This cost can be reduced to $\mathcal{O}(M^3)$ where $M \ll N$ by applying matrix a inverse property. Thus, Eq. (16) can be rewritten as

$$\mathbf{C}_t^{-1} = \mathbf{P}^{-1} - \mathbf{P}^{-1} \hat{\mathbf{S}}^T \mathbf{A}^T (\sigma_n^{-1} \mathbf{I} + \mathbf{A} \hat{\mathbf{S}} \mathbf{P}^{-1} \hat{\mathbf{S}}^T \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{S} \mathbf{P}^{-1}, \quad (33)$$

where $\mathbf{P} = \boldsymbol{\Sigma}_t + \sigma_n \left(\text{diag}(\hat{\mathbf{s}} \odot (1 - \hat{\mathbf{s}})) \odot (\mathbf{A}^T \mathbf{A}) \right)$ is a diagonal matrix whose inverse can be easily computed. The matrix inversion complexity is reduced to $\mathcal{O}(M^3)$. Our complexity is much less than Two-steps-Adaptive MRF [21], *i.e.*, $\mathcal{O}(c_1 2M^3 + c_2 |G| + c_3 C(\mathcal{G}))$ that includes sparse recovery $c_1 2M^3$, support estimation $c_2 |G|$, and MRF parameter estimation $c_3 C(\mathcal{G})$ with using [23], where $|G|$ denotes the size of a graph \mathcal{G} . c_1 and c_2 , and c_3 are the maximum iteration numbers for executing the three subroutines.

6. Experimental Result

The performance of our One-step-Adaptive MRF is studied through three different experiments to evaluate: (i) the effectiveness of our MRF distribution in Section 6.4; (ii) the effectiveness of MRF parameter estimation in Section 6.5; and (iii) the performance evaluation with state-of-the-art algorithms in Section 6.6. The details for datasets, settings, and comparison methods are as follows.

6.1. Datasets

The performance of our One-step-Adaptive MRF is evaluated on three benchmark datasets: (i) MNIST [25], a handwritten image dataset; (ii) CMU-
IDB [26], a face image dataset; and (iii) CIFAR-10 [27], a natural image dataset.

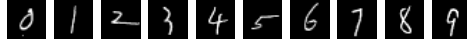


Figure 3: MNIST. The ground truth handwritten digit images.

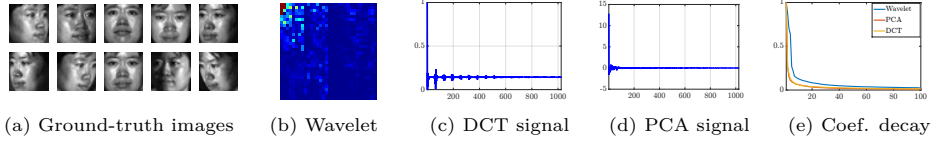


Figure 4: CMU-IDB. (a) The ground truth natural images; (b) examples of wavelet (c) DCT, and (d) PCA signal; and (e) the decay of signal coefficients in wavelet, DCT, and PCA domain.

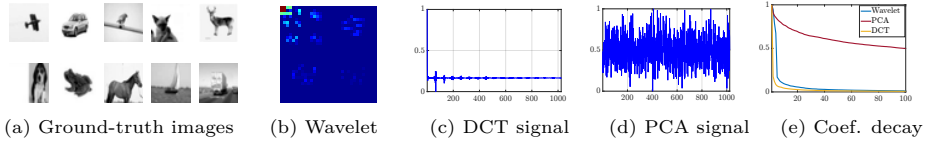


Figure 5: CIFAR-10. (a) The ground truth natural images; (b) examples of wavelet (c) DCT, and (d) PCA signal; and (e) the decay of signal coefficients in wavelet, DCT, and PCA domain.

The images selected from each dataset are shown in Figure 3, 4a, 5a. The MNIST images are strictly sparse; thus, the compression is applied onto the images directly. The sparse representations of CMU-IDB and CIFAR-10 images are obtained by using wavelet transform, discrete cosine transform (DCT), and principal component analysis (PCA). The examples of these representations are in Figure 4b, 4c, 4d and Figure 5b, 5c, 5d, for CMU-IDB and CIFAR-10. Notice that CIFAR-10 images are not sparse in PCA domain. Meanwhile we report the signal recovery performance evaluated based on these image signals, our discussion will focus only on MNIST images, CMU-IDB images in PCA domain, and CIFAR-10 images in wavelet domain, whose the representations are sparse (see Figure 4e, 5e); thus, they can better reflect the signal recovery performance.

6.2. Settings

Experiment setting: In compression, the sparse signal \mathbf{x} is sampled by a random Bernoulli matrix \mathbf{A} to generate the linear measurements \mathbf{y} . The sampling rate (M/N) are set to 0.2-0.4 to test the performance across different measurement sizes. To simulate the noise corruption in \mathbf{y} , three levels of Gaussian white noise are added into \mathbf{y} , which results in the signal to noise ratio (SNR) of \mathbf{y} to

325 be 30dB, 20dB, 10dB. The lowest SNR (10dB) indicates highest noise corruption.

Evaluation criterion: The performance is evaluated based on the recovery accuracy, which is measured by peak signal to noise ratio (PSNR) and structural similarity index (SSIM), and the runtime performance which is the total runtime.

330 **Algorithm setting:** One-step-Adaptive MRF is initialized as follows: The hyperparameters ϖ and ξ in Eq. (24) and ϖ_0 and ξ_0 in Eq. (26) are set to 10^{-6} . The initial value for α and β in Eq. (22) is set according to [16]. Notice that all the images from these datasets are gray-scale images. The compression process is applied on these images directly. To efficiently perform image processing on
335 these images, we normalize the pixel value to a value between zero and one. The proposed method can be applied to reconstruct 2D and 1D-vectorized image signal. In our experiment, we reconstruct 1D-vectorized image. Regardless of whether the reconstructed image signal is 1D or 2D, spatial correlation can always be captured by configuring the edge set $E = \{\mathcal{E}_i\}$. To capture spatial
340 correlation in 2D signals, *i.e.*, handwritten and wavelet images, \mathbb{N}_c and \mathbb{N} are set to cover 8-neighbors of each node. For 1D signals, *i.e.*, PCA and DCT signals, \mathbb{N}_c and \mathbb{N} are set to cover two adjacent nodes. At the first iteration, the edge set $E = \{\mathcal{E}_i\}$ is initialized as an empty set. The algorithm stops when the minimum update difference, *i.e.* $\frac{\|\mathbf{x}^{prev} - \mathbf{x}^{new}\|_2}{\|\mathbf{x}^{prev}\|_2}$, is less than 10^{-3} , or when the iteration
345 reaches 200.

6.3. Comparison methods

Our method is compared with 7 state-of-the-art competitors: (i) **Adaptive-MRF based methods:** Two-step-Adaptive MRF (TA-MRF) [21] and Pairwise MRF[20]¹ ; (ii) **MRF-based methods (Non-Adaptive):** MAP-OMP [7] and

¹For TA-MRF, Pairwise MRF, and Bernoulli model, we use the same setting for neighboring set \mathbb{N}_c , as described in Algorithm setting in Section 6.2

350 Gibbs [4]²; (iii) **Cluster sparsity-based methods**: Bernoulli model [16]
¹; (iv) **Sparsity-based methods**: RLPHCS[28] and OMP[29]. All of the
comparison methods, except Pairwise MRF [20], are implemented by the code of
the authors with tuned parameters to the best performance. The Pairwise MRF
is coded by ourselves and uses the same setting for \mathbb{N} and terminating criterion
355 to our work.

To show the best possible result using ground truth support, we use **the oracle estimator** in [7] that uses *the ground truth support* to estimate the signal via Eq. (17) with homogeneous noise and signal parameters. All the other methods *do not have* the access to the ground truth support.

360 6.4. Effectiveness of our MRF distribution

In the following paragraphs, we (i) study the effectiveness of our MRF distribution in approximating the BM Eq. (4) and (ii) evaluate the difference/similarity between our MRF distribution Eq. (8) and the BM Eq. (4) based on real data.

6.4.1. Effectiveness of our MRF distribution in approximating the BM

365 This section demonstrates the effectiveness of our MRF distribution Eq. (8) in approximating the Boltzmann machine (BM) Eq. (4) by measuring the KL-divergence between them. The effectiveness is evaluated based on modelling sparse signals with different levels of *sparsity* and *sparse coefficients dependency*. The proposed distribution is compared with some of the existing approximation
370 schemes: the Bernoulli model [16] Eq. (6) and the Pairwise MRF [20] Eq. (7). Our evaluation is based on one thousand synthesized sparse signals that are generated from a handcrafted BM distribution in order to control the level of sparsity and sparse coefficients dependency in sparse signals. One thousand BM distributions are synthesized and used in our experiments.

²The graphical model, noise and signal variance parameters provided to MAP-OMP and Gibbs is from training data.

Sparsity (k)	Averaged KL-divergence wrt. the BM Eq. (4)		
	Our dist. Eq. (8)	Bernoulli [16] Eq. (6)	Pairwise [20] Eq. (7)
10	0.0020	0.0281	3.0179
20	0.0025	0.0617	2.1777
30	0.0026	0.1103	1.4552

(a) Approximating the BM Eq. (4) across different sparsity levels.

Num. of edges ($ E $)	Averaged KL-divergence wrt. the BM Eq. (4)		
	Our dist. Eq. (8)	Bernoulli [16] Eq. (6)	Pairwise [20] Eq. (7)
$2^\dagger N^*$	0.0036	0.0671	0.7409
$10^\dagger N^*$	0.0125	0.0781	0.5337
$20^\dagger N^*$	0.0743	0.1219	0.1039

$^\dagger 2, 10,$ and 20 are the number of pairwise edges connecting to each node.

$^* N$ is the signal dimension

(b) Approximating the BM Eq. (4) across different number of edges.

Table 1: Effectiveness of our MRF distribution Eq. (8) in approximating the BM Eq. (4) vs. existing approximation schemes: the Bernoulli model [16] Eq. (6), and the pairwise MRF [20] Eq. (7) across (a) different sparsity levels and (b) number of edges.

375 **Ground truth BM.** To generate sparse signals with sparsity levels (k) of
10, 20, and 30, we configure the unary parameters of the BMs (Eq. (4)). The
unary parameters are selected from three $\mathcal{N}(\mu_b, 1)$ with $\mu_b = -2.5, -2, -1.5$
corresponding to three sparsity levels. The pairwise parameters are selected from
 $\mathcal{N}(-0.1, 1)$. The number of edges are fixed to 200. To synthesize sparse signals
380 with different levels of dependency, we vary the number of pairwise edges ($|E|$)
of the ground truth BMs, *i.e.* $|E| = 2N, 10N,$ and $20N$ where $N = 200$. The
unary and pairwise parameters are randomly selected from $\mathcal{N}(\cdot, 1)$ with mean
of -1 and -0.3. We measure the empirical KL-divergence based on the sparse
signals modelled by the proposed MRF and those ground truth sparse signals.
385 One thousand ground truth sparse signals are generated from each BM. The
KL-divergence is then averaged over one thousand BMs.

Experimental Results. Table 1a and Table 1b demonstrate the approxi-
mation to the BM with (i) different sparsity levels (k) and (ii) number of edges
($|E|$). In Table 1a, the KL-divergence of our MRF distribution is less than

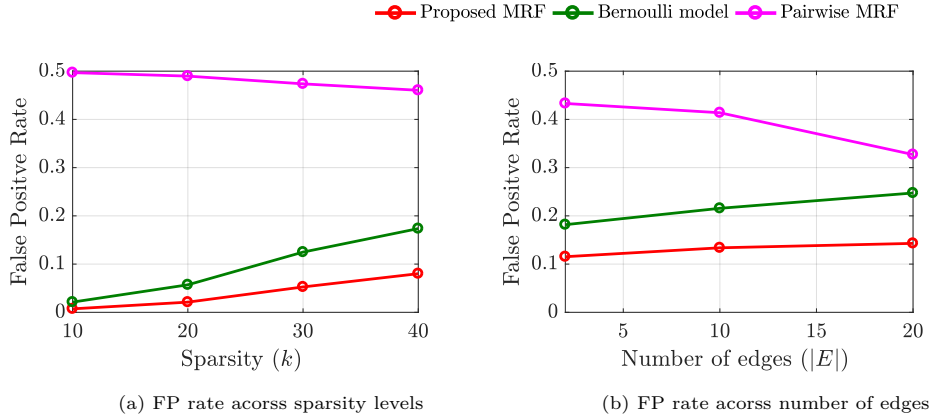


Figure 6: False positive rates (FPR) comparison: (a) FPR across different sparsity (k), *i.e.* $k = 10, 20, 30, 40$, and (b) FPR across different number of edges ($|E|$), *i.e.* $|E| = 2N, 10N, 20N$.

390 *one order* and *three order* of magnitude in comparison with Bernoulli model and Pairwise MRF. In Table 1b, the KL-divergence of our MRF distribution is smaller than the others. When $|E| < 20N$, the KL-divergence of the proposed MRF which is 0.0125 is at most 17% of the Bernoulli model that is 0.0781. The KL-divergence of the proposed MRF is at most 3% of the Pairwise MRF's, which
 395 is 0.5337. When $|E| = 20N$, our KL-divergence is approximately 60% and 70% of Bernoulli model and Pairwise MRF. With both unary and pairwise parts, our MRF distribution (Eq. (8)) can best approximates the BM (Eq. (4)) across different configurations with the smallest KL divergence.

It is noticeable that the KL-divergence of the proposed MRF distribution
 400 increases with the sparsity levels and number of edges which agrees with our theoretical result in Section 5.1. The KL-divergence of the Bernoulli model follows a similar pattern to the proposed MRF; however, this is not the case for Pairwise MRF whose KL-divergence increases when the sparsity levels and number of edges decrease. To further investigate, Figure 6a and 6b compare the
 405 false positive rates of each model in associated with Table 1a and 1b.

Figure 6a provides the false positive rate of each model across different

sparsity levels. The false positive rate of the proposed MRF and Bernoulli increase with the sparsity levels. However, the false positive rate of Pairwise MRF increases as the sparsity level decreases. This is due to that Pairwise MRF does not have a unary part to control the sparsity of the modeled sparse signal. Figure 6b shows the false-positive rate across different number of pairwise edges. Although Pairwise MRF has the pairwise part to control the dependency level according to the number of edges, it does not have a unary term to model the sparsity of signals when number of edges is low. Because the proposed MRF distribution contains both unary and pairwise parts, its false positive rate is much lower than the Bernoulli model and Pairwise MRF.

6.4.2. Differences analysis of our proposed MRF to the BM

In this experiment, we evaluate (i) the differences/similarity of unary and pairwise parts of the proposed MRF distribution (Eq. (8)) and those of the BM (Eq. (4)) on 10 MNIST images and (ii) the quality of the supports estimated from each probabilistic model. Structural similarity (SSIM) index is used to measure the differences. Thus, in the former case (i), the SSIM calculated with respect to the learned pairwise and unary of the BM; whereas in the latter case (ii), the SSIM is calculated with respect to the ground truth images. Figure 7 and 8 compares pairwise and unary potentials of the proposed MRF and the BM. The SSIM between the learned pairwise of the proposed MRF and those of the BM are higher than 0.94, and the SSIM between the learned unary potentials of the proposed MRF and those of the BM are higher than 0.98. Figure 9 compares the quality of the estimated supports from the proposed MRF and the BM. It is clear that the supports estimated from the proposed MRF are similar to those of the BM. These estimated supports accurately represent the ground truth supports with the SSIM value equal to 1. Thus, the results from Figure 7, 8, and 9 suggest that the proposed MRF Eq. (8) functions similarly to the BM Eq. (4).

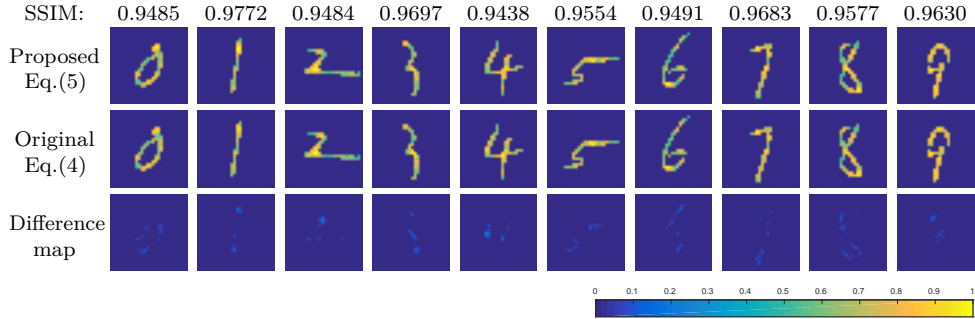


Figure 7: Sum of the pairwise potentials at each image pixel. Comparison between the learned pairwise potentials of the proposed MRF Eq. (8) and those of the BM Eq. (4), and the difference map between them. SSIM is calculated with respect to learned pairwise potentials of the BM Eq. (4).

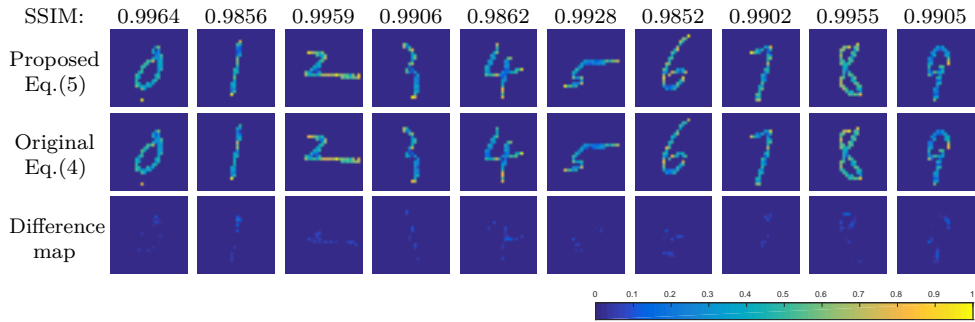


Figure 8: Unary potentials at each image pixel. Comparison between the learned unary potentials of the proposed MRF Eq. (8) and those of the BM Eq. (4), and the difference map between them. SSIM is calculated with respect to learned unary potentials of the BM Eq. (4).

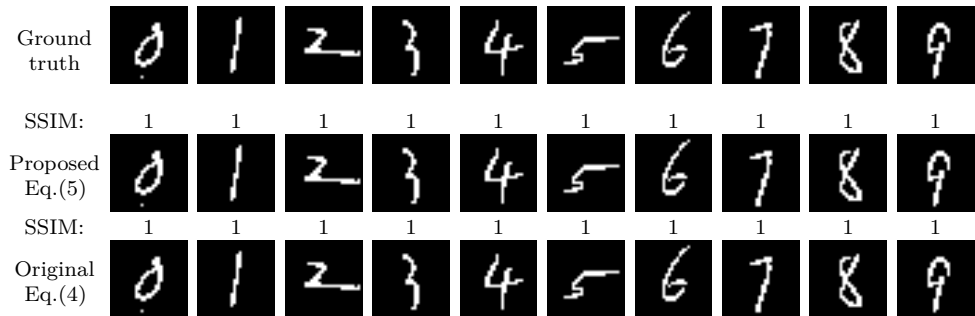


Figure 9: Estimated support. Comparison between those estimated support using the proposed MRF Eq.(5) and the original BM Eq.(4). (SSIM is calculated with respect to the ground truth images.)

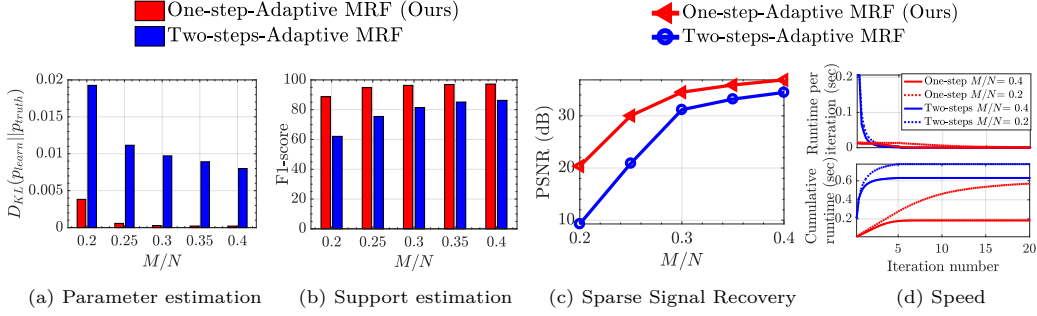


Figure 10: *Compressibility*. Comparison between our One-step-Adaptive MRF and Two-steps-Adaptive MRF [21] in (a) quality of MRF parameter estimation, (b) accuracy of support estimation, (c) accuracy of sparse signal recovery, and (d) average runtime across different sampling rates. Noise level (SNR) is 30 dB.

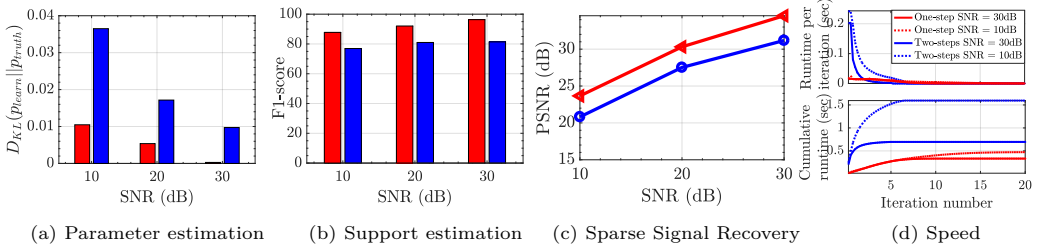


Figure 11: *Noise tolerance*. Comparison between our One-step-Adaptive MRF and Two-steps-Adaptive MRF [21] in (a) quality of MRF parameter estimation, (b) accuracy of support estimation, (c) accuracy of sparse signal recovery, and (d) average runtime across different noise levels. Sampling rate is 0.3.

6.5. Effectiveness of MRF parameters estimation: One-step vs Two-step

435 We compare the effectiveness of our One-step-Adaptive MRF versus Two-step-Adaptive MRF [21] in estimating MRF parameters for 10,000 signals sampled from 10 distribution. The evaluation is based on the parameter estimation measured by KL-divergence and the final performance by F1-score, recovery accuracy, and runtime. Figure 10 and 11 show the results across different sampling rates (M/N) and noise levels (in SNR). In Figure 10, the KL divergence of our
 440 One-step-Adaptive MRF is at most 25% of the Two-steps-Adaptive MRF. Our method also yields higher F1-score³ by at least 5%, higher accuracy by at least 2 dB, and less runtime as all the update equations are in closed form. In Figure 11,

³For our algorithm, F1-score is calculated from the binary support obtained from Eq. (30).

the KL divergence of our method is at most 30% of Two-steps-Adaptive MRF.
445 Our method yields higher F1-score ³ and accuracy, by at least 5% and 3 dB,
with less runtime. With the improved MRF estimation, our method is more
efficient.

6.6. Performance Evaluation

This section provides the performance of our One-step-Adaptive in comparison
450 with the state-of-the-art algorithms on MNIST, CMU-IDB, and CIFAR-10.

Compressibility. Figure 12 and 13 shows the average PNSR and SSIM
curves across different sampling rates (M/N). The noise level is 30 dB. In
Figure 12, the proposed One-step-Adaptive MRF offers the highest performance
in most cases. The proposed One-step-Adaptive MRF exceeds the second best
455 method by at least 2 dB on MNIST, when $M/N > 0.25$. For CMU-IDB, it
exceeds the second best method by at least 1 dB, 0.5 dB, and 2 dB in wavelet,
PCA, and DCT domains. For CIFAR-10, it exceeds the second best method
by at least 0.25 dB and 2 dB in wavelet and DCT domains. In Figure 13, our
One-step-Adaptive MRF offers the highest SSIM to the ground truth images
460 across different datasets. On MNIST images and PCA signals of CMU-IDB,
our One-step-Adaptive MRF performs well across different sampling rates. On
CMU-IDB, it exceeds the second best method by at least 0.1 and 0.5 in DCT and
wavelet domains. For CIFAR-10, it exceeds the second best method by at least
0.25 and 0.2 in wavelet and DCT domains. To further examine visual results,
465 Figure 16a, 17a, 18a, and Figure 16b, 17b, 18b provide the images reconstructed
from top five algorithms when $M/N = 0.3$ and 0.2. Our One-step-Adaptive
MRF (OA-MRF) offers the best visual results that contains more details and
less noise in most cases. We provide all the reconstructed images from each
dataset in .3.

470 **Noise tolerance.** Figure 14 provides the average PNSR curves across

different noise levels (in SNR). $M/N = 0.3$. Our method achieves the superior noise tolerance over the other methods across different datasets: On MNIST images, our One-step-Adaptive MRF outperforms the second best method by at least 2 dB, when $\text{SNR} > 5$ dB. For CMU-IDB, it exceeds the second best method
475 by at least 2 dB, 1 dB, and 1 dB in wavelet, PCA, and DCT domains. For CIFAR-10, it exceeds the second best method by at least 1 dB and 2 dB in wavelet and DCT domains. Note that for the recovery of MNIST images and CIFAR-10 images in DCT domain, our One-step-Adaptive MRF even outperforms the oracle estimator. This is because our One-step-Adaptive MRF enables heterogeneous
480 noise parameters obtained from the adaptive noise estimation. Meanwhile, the oracle estimator uses the homogeneous noise parameters obtained from the training data. This indicates that the adaptive mechanism provides a good prior to separate signal information from noise.

Runtime. Figure 15 provides average runtime curves across different sam-
485 pling rate (M/N). The noise level is 30 dB. Our method requires a moderate runtime in most cases. On MNIST, the average runtime of our One-step-Adaptive MRF is moderate compared with the others. It is faster than Two-step-Adaptive MRF, MAP-OMP, and Pairwise-MRF; it is comparable to RLPHCS, but is slower than Bernoulli and OMP. For CMU-IDB and CIFAR-10, the runtime per-
490 formance of our One-step-Adaptive MRF is much better than many structured CS methods. The runtime performance is similar across the wavelet, DCT, and PCA domains: our One-step-Adaptive MRF is faster than Two-step-Adaptive MRF, MAP-OMP, Bernoulli, and Pairwise MRF. Its runtime is comparable with RLPHCS and only slower than OMP. Note that OMP and RLPHCS require

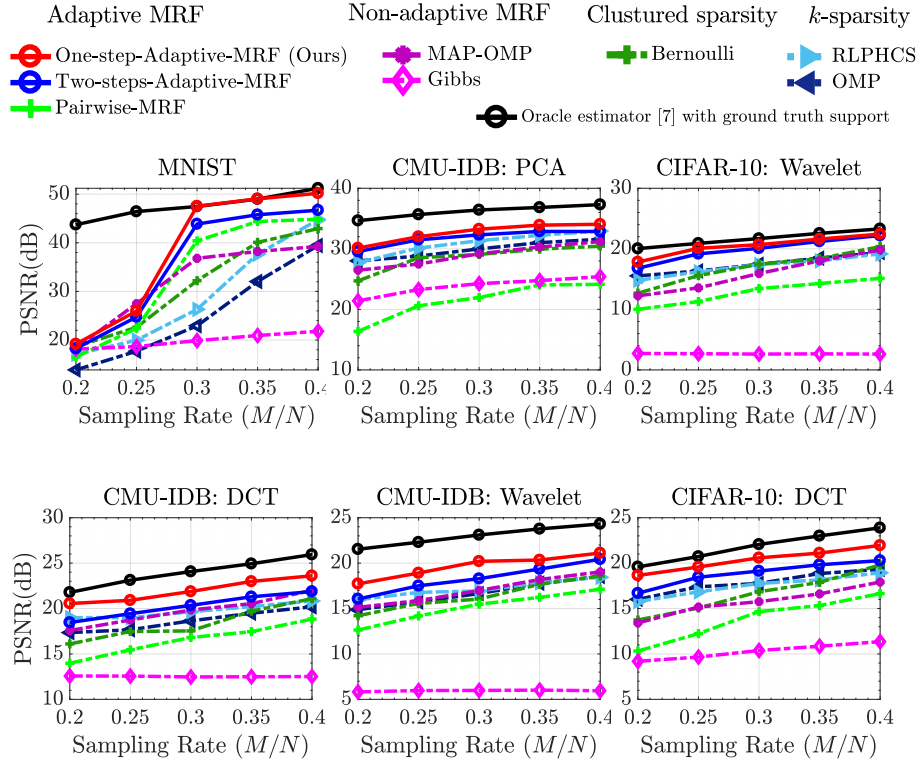


Figure 12: *Compressibility*. The PSNR curves across different sampling rates on three datasets. Noise level (SNR) is 30 dB.

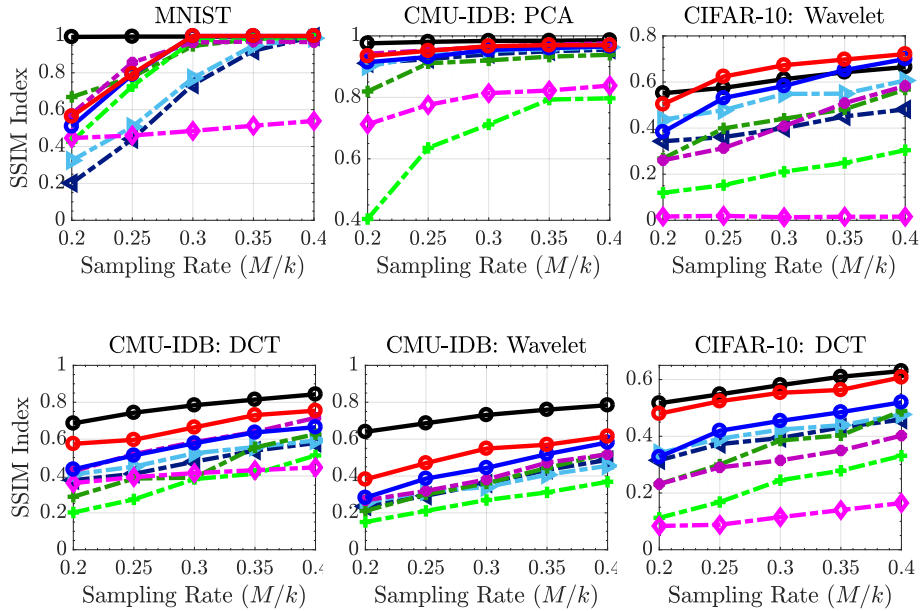


Figure 13: *Compressibility*. The SSIM indices across different sampling rates on three datasets. Noise level (SNR) is 30 dB.

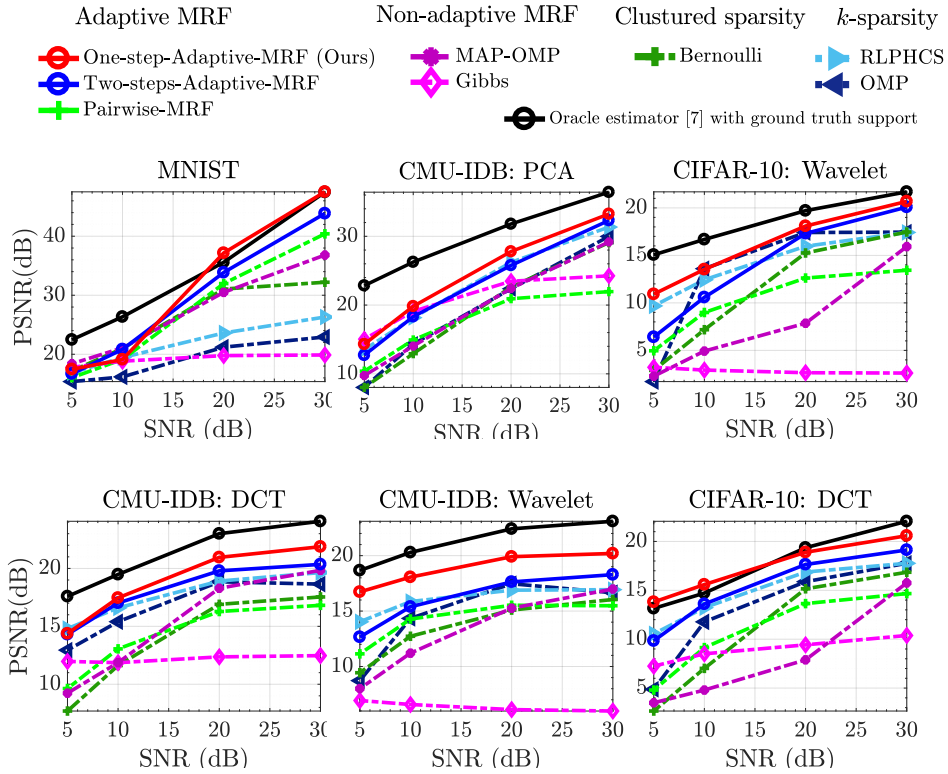


Figure 14: *Noise tolerance.* The PSNR curves across different noise levels (SNR) on three datasets. Sampling rate is 0.3.

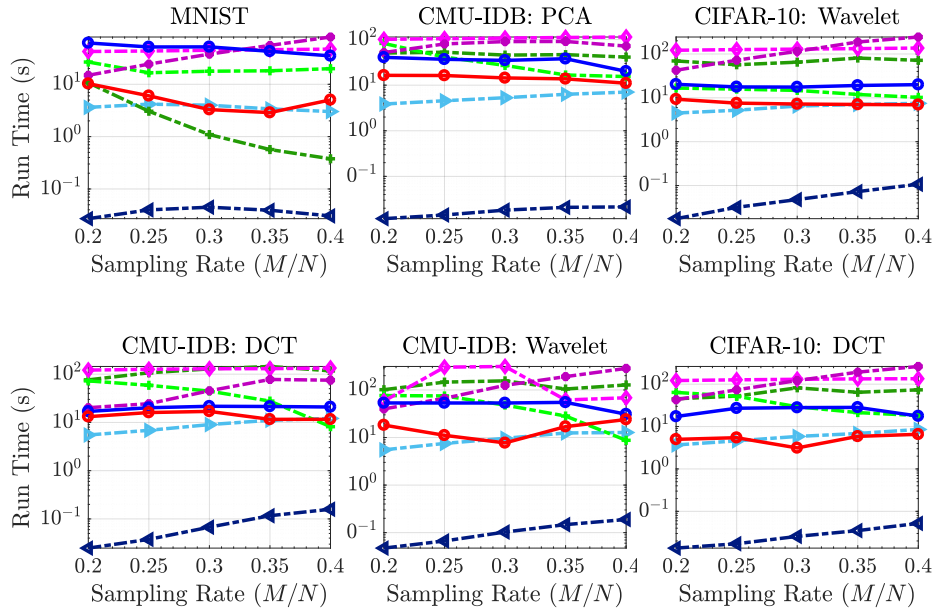
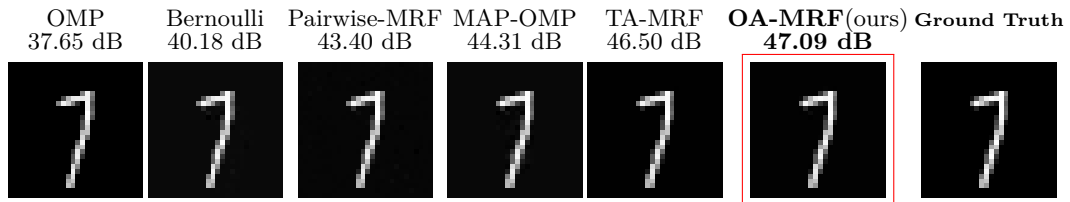
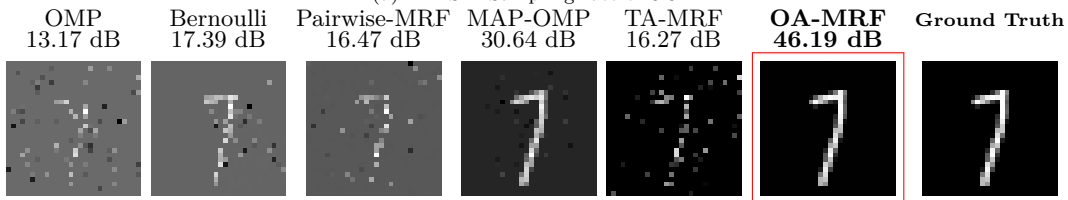


Figure 15: *Runtime performance.* Total runtime curves across different sampling rates on three datasets. Noise level (SNR) is 30 dB.

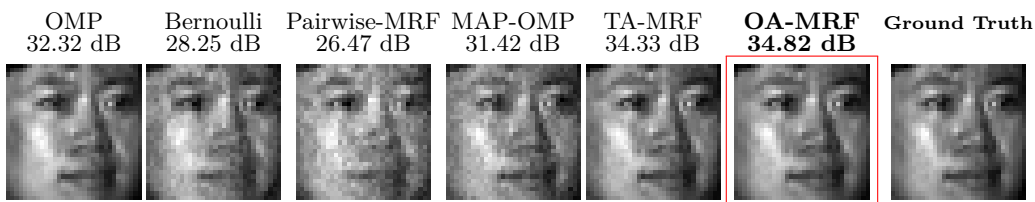


(a) MNIST: Sampling rate of 0.3.

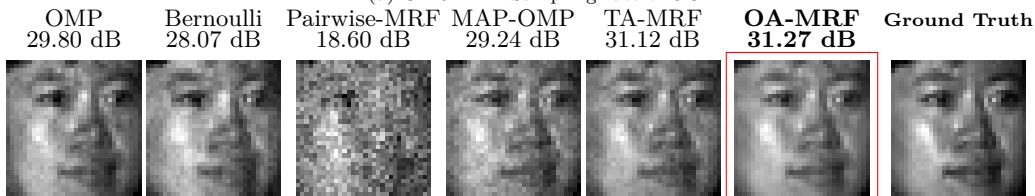


(b) MNIST: Sampling rate of 0.2.

Figure 16: MNIST: Visual results of selected handwritten images by top 6 algorithms when noise level (SNR) is 30dB at the sampling rate of (a) 0.3 and (b) 0.2.

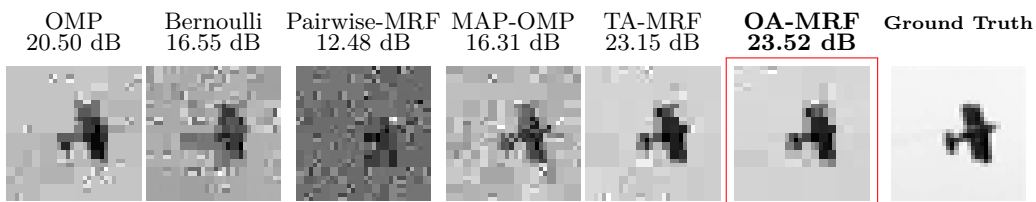


(a) CMU-IDB: Sampling rate of 0.3.

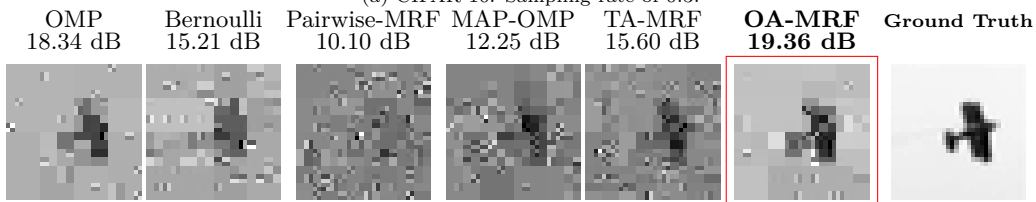


(b) CMU-IDB: Sampling rate of 0.2.

Figure 17: CMU-IDB: Visual results of selected face images recovered in PCA domain by top 6 algorithms when noise level (SNR) is 30dB at the sampling rate of (a) 0.3 and (b) 0.2.



(a) CIFAR-10: Sampling rate of 0.3.



(b) CIFAR-10: Sampling rate of 0.2.

Figure 18: CIFAR-10: Visual results of selected natural images recovered in wavelet domain by top 6 algorithms when noise level (SNR) is 30dB at the sampling rate of (a) 0.3 and (b) 0.2.

495 low computation in general because they do not exploit the structure in sparse
signal coefficients.

7. Conclusion

We have presented a novel one-step Markov random field (MRF) based
structured CS to estimate the parameters of an MRF from a few measurements.
500 A recent method estimates the MRF parameters based on a point estimation of
the sparse signals, but cannot depict the statistical uncertainty of the latent sparse
signals. Therefore, we propose to estimate the MRF parameters from solving
a maximum marginal likelihood (MML) problem to offer better generalization
over the latent sparse signals. A new MRF distribution is proposed to enable
505 closed-form formulations for all the unknowns in the MML problem. Extensive
experiment demonstrates the state-of-the-arts results of the proposed method.

Acknowledgment

This research was supported by grant funding from the Department of State
Development under the Collaboration Pathways Program (CPP39), Government
510 of South Australia.

Appendices

.1. Connection between Bernoulli distribution and the BM Eq. (4)

The Bernoulli model can well substitute the unary potentials s_i in the BM
Eq. (4). Without the pairwise part, the BM distribution becomes a Bernoulli
distribution as follows [6]

$$\begin{aligned} p(\mathbf{s}) &= \prod_c \prod_{i \in \mathbb{N}_c} \frac{1}{Z_i} \exp(s_i \delta_i^c) \\ &= \prod_c \prod_{i \in \mathbb{N}_c} \text{Bernoulli}(s_i) \end{aligned} \tag{34}$$

where $Z_i = \sum_{j \in \{0,1\}} \exp(s_j \delta_i^c) = \exp(\delta_i^c) + 1$ and

$$\text{Bernoulli}(s_i) = \begin{cases} \frac{\exp(\delta_i^c)}{1 + \exp(\delta_i^c)}, & \text{if } s_i = 1 \\ \frac{1}{1 + \exp(\delta_i^c)}, & \text{otherwise} \end{cases} \quad (35)$$

The proposed MRF contains both the unary and pairwise parts which are similar to those of the BM Eq. (4). The only difference between the proposed MRF Eq. (8) and the BM Eq. (4) is the denominator. In the next section, we will show how well the proposed MRF Eq. (4) can approximate the BM Eq. (4) based on the KL-divergence between them.

.2. Performance comparison with clustered sparsity-based approaches

In the following, we provide additional performance comparison with clustered sparsity-based approaches, *i.e.* pattern-coupled sparse Bayesian learning (PC-SBL) and Block-sparse Bayesian learning (B-SBL), on three datasets used in Section 6.6.

To evaluate the success rate, we define a successful trial as the one where its relative mean squared-error is in the lowest 25 percentiles. Here, we use the 25 percentiles of MSE because the different quality of reconstructed images from different datasets. The mean squared-error is also used to evaluate the performance across different sampling rates and noise levels. Here, we did not compare the performance with approximate message passing for block sparse signal (AMP-B-SBL) because of the difficulty in implementation: more than six parameters are needed to be specifically tuned for each image.

Figure 19 and 20 provide the curves of the success rate and the relative mean squared-error (in percentage) across different sampling rates (M/N), when the noise level is low. It is noticeable that our One-step-Adaptive MRF can work well across different datasets. Because our One-step-Adaptive MRF employs

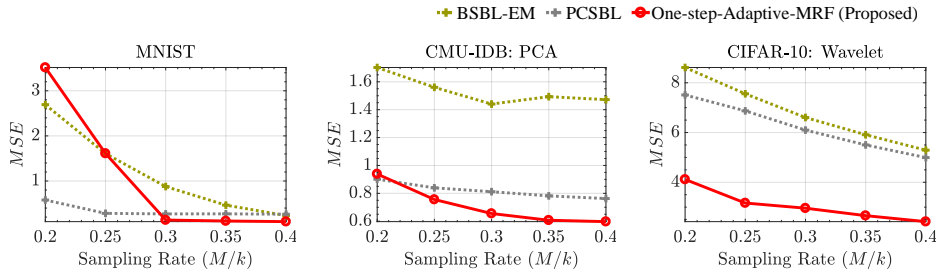


Figure 19: *Low noise cases.* The %MSE curves across different sampling rates on three datasets.

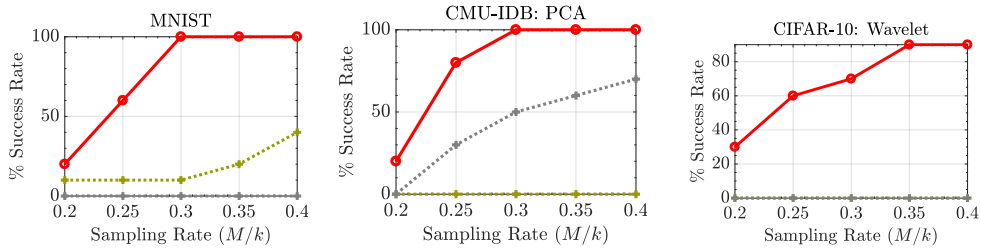


Figure 20: *Low noise cases.* The success rates curves across different sampling rates on three datasets.

535 the flexible MRF that can adapt for different types of signals, it achieves the highest success rate and the lowest MSE in most cases. Meanwhile, PC-SBL and BSBL-EM employ clustered sparsity models. They achieve a good performance on MNIST images whose sparse coefficients group in clusters. However, because the clustered sparsity models are fixed and cannot adapt for different sparse
540 signals, the performance of PC-SBL and BSBL-EM are limited, and thus, the resulting MSE are mostly higher than 25 percentile. As a result, they can achieve lower success rate compared to One-step-Adaptive MRF.

Figure 21 provide the curves of the MSE across different noise levels. Notice that the higher SNR signifies the less noise corruption. Our One-step-Adaptive
545 MRF offers the highest noise tolerance in most cases. This is due to the high flexible and adaptive MRF that can better represent the signals against noise.

This experiment has demonstrated that the proposed method can work well

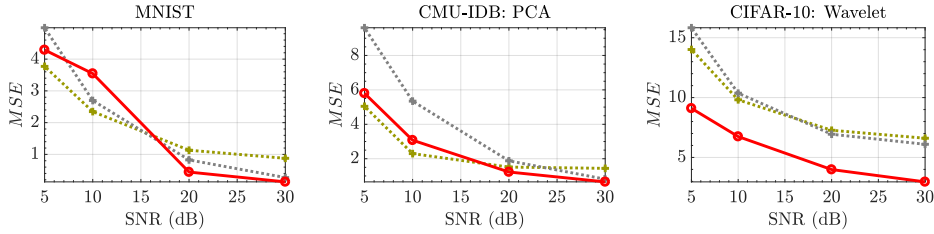


Figure 21: *Noisy cases.* The %MSE curves across different noise levels in SNR on three datasets. Sampling rate is 0.3.

across different types of sparse signals, it not the main objective. This conclusion complies with our performance comparison with various methods including the clustered sparsity models-based approaches in Section 6.

.3. Full image reconstruction results

This section provided all the reconstructed images from MNIST, CMU-1DB, and CIFAR-10 datasets.

MNIST. The reconstruction of the MNIST’s handwritten images are provided in Figure 22 and 23 when the sampling rate is 0.3 and 0.2, respectively. It is clear that the proposed One-step-Adaptive MRF achieves the best reconstruction results across different MNIST handwritten digit samples in most cases. It provides the highest PSNR improvement over the second most competitive method by 5.60 dB on digit no. 5 in Figure 22, and by 15.55 dB on digit no. 7 in Figure 23. This visual results are consistent with the numerical results of the PSNR in the Figure 12.

CMU-1DB. The reconstruction of CMU-1DB images in PCA, wavelet, and DCT domains are provided in Figure 24, 26, 28, and Figure 25, 27, 29 when the sampling rate is 0.3 and 0.2, respectively. The proposed One-step-Adaptive MRF can provide the best results on sparse signal recovery in most cases. On recovering CMU-1DB images in PCA domain, the proposed One-step-Adaptive MRF yields the highest PSNR improvement over the second most competitive

method by 2.4 dB in the 7th row from the top of Figure 24, and by 2.24 in the 3rd row of Figure 25 at the sampling rate of 0.3 and 0.2. In wavelet domain, the
570 proposed method yields the highest PSNR improvement over the second most competitive method by 2.54 dB in the 2nd row from the top of Figure 26, and by 2.46 dB in the 10th row from the top of Figure 27 at the sampling rate of 0.3 and 0.2. In DCT domain, the proposed method yields the highest PSNR improvement over the second most competitive method by 2.02 dB in the 8th
575 row from the top of Figure 28, and by 1.32 dB in the 5th row from the top of Figure 29 at the sampling rate of 0.3 and 0.2. This visual results are consistent with the numerical results of the PSNR in the Figure 12.

CIFAR-10. The reconstruction of CIFAR-10 natural images in wavelet and DCT domains are provided in Figure 30, 32 and Figure 31, 33, when the sampling
580 rate is 0.3 and 0.2, respectively. The proposed One-step-Adaptive MRF can provide the highest reconstruction quality across different sparse representation in most cases. On recovering CIFAR-10 images in wavelet domain, the proposed One-step-Adaptive MRF yields the highest PSNR improvement over the second most competitive method by 1.16 dB in the 10th row from the top of Figure
585 30, and by 2.86 dB over the second best method in the 6th row from the top of Figure 31, at the sampling rate of 0.3 and 0.2, respectively. In DCT domain, the proposed method provides the highest PSNR improvement over the second most competitive method by 2.50 dB in the 3rd row from the top of Figure 32, and by 2.69 dB in the 2nd row from the top of Figure 33 at the sampling rate of 0.3 and
590 0.2, respectively. This visual results are consistent with the numerical results of the PSNR in the Figure 12.

In conclusion, these visual results (Figure 22- Figure 34) indicate that our One-step-Adaptive MRF offers the best results among all the methods in the most cases, both visually and numerically.

595 **References**

- [1] V. Cevher, P. Indyk, L. Carin, R. G. Baraniuk, Sparse signal recovery and acquisition with graphical models, *IEEE Signal Process. Mag.* 27 (6) (2010) 92–103. doi:10.1109/MSP.2010.938029.
- [2] R. G. Baraniuk, V. Cevher, M. B. Wakin, Low-dimensional models for
600 dimensionality reduction and signal recovery: A geometric perspective, *Proc. IEEE* 98 (6) (2010) 959–971.
- [3] P. J. Wolfe, S. J. Godsill, W. Ng, Bayesian variable selection and regularization for time frequency surface estimation, *J. R. Stat. Soc. Series B Stat. Methodol.* 66 (3) (2004) 575–589.
- [4] P. Garrigues, B. A. Olshausen, Learning horizontal connections in a sparse
605 coding model of natural images, in: *Adv. Neural Inf. Process. Syst. (NIPS)*, 2008, pp. 505–512.
- [5] V. Cevher, M. F. Duarte, C. Hegde, R. Baraniuk, Sparse signal recovery using markov random fields, in: *Adv. Neural Inf. Process. Syst. (NIPS)*,
610 2009, pp. 257–264.
- [6] A. Drémeau, C. Herzet, L. Daudet, Boltzmann machine and mean-field approximation for structured sparse decompositions, *IEEE Trans. Signal Process.* 60 (7) (2012) 3425–3438.
- [7] T. Peleg, Y. C. Eldar, M. Elad, Exploiting statistical dependencies in sparse
615 representations for signal recovery, *IEEE Trans. Signal Process.* 60 (5) (2012) 2286–2303.
- [8] J. Ren, J. Liu, Z. Guo, Context-aware sparse decomposition for image denoising and super-resolution, *IEEE Trans. Image Process.* 22 (4) (2013) 1456–1469. doi:10.1109/TIP.2012.2231690.

- 620 [9] M. Panić, J. Aelterman, V. Crnojević, A. Pizurica, Compressed sensing in mri with a markov random field prior for spatial clustering of subband coefficients, in: The 24th Eur. Signal Process. Conf. (EUSIPCO), 2016, pp. 562–566. doi:10.1109/EUSIPCO.2016.7760311.
- [10] R. Torkamani, R. A. Sadeghzadeh, Bayesian compressive sensing using wavelet based markov random fields, Signal Processing: Image Communication 58 (Supplement C) (2017) 65 – 72. doi:https://doi.org/10.1016/j.image.2017.06.004.
- [11] L. He, H. Chen, L. Carin, Tree-structured compressive sensing with variational bayesian analysis, IEEE Signal Processing Letters 17 (3) (2010) 233–236. doi:10.1109/LSP.2009.2037532.
- 630 [12] L. Yu, C. Wei, J. Jia, H. Sun, Compressive sensing for cluster structured sparse signals: Variational bayes approach, IET Signal Processing 10 (7) (2016) 770–779.
- [13] L. Yu, H. Sun, J.-P. Barbot, G. Zheng, Bayesian compressive sensing for cluster structured sparse signals, Signal Processing 92 (1) (2012) 259–269.
- 635 [14] Z. Zhang, B. D. Rao, Recovery of block sparse signals using the framework of block sparse bayesian learning, in: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 3345–3348. doi:10.1109/ICASSP.2012.6288632.
- 640 [15] M. Shekaramiz, T. K. Moon, J. H. Gunther, Hierarchical bayesian approach for jointly-sparse solution of multiple-measurement vectors, in: 2014 48th Asilomar Conference on Signals, Systems and Computers, 2014, pp. 1962–1966. doi:10.1109/ACSSC.2014.7094813.

- [16] L. Yu, H. Sun, G. Zheng, J. P. Barbot, Model based bayesian compressive
645 sensing via local beta process, *Signal Processing* 108 (2015) 259–271.
- [17] J. Fang, Y. Shen, H. Li, P. Wang, Pattern-coupled sparse bayesian learning
for recovery of block-sparse signals, *IEEE Trans. Signal Process.* 63 (2)
(2015) 360–372. doi:10.1109/TSP.2014.2375133.
- [18] J. Fang, L. Zhang, H. Li, Two-dimensional pattern-coupled sparse bayesian
650 learning via generalized approximate message passing, *IEEE Trans. Image
Process.* 25 (6) (2016) 2920–2930. doi:10.1109/TIP.2016.2556582.
- [19] M. Shekaramiz, T. K. Moon, J. H. Gunther, AMP-B-SBL: An algorithm
for clustered sparse signals using approximate message passing, *Ubiquitous
Comput Electron Mob Commun Conf (UEMCON) IEEE Annu* 2016.
- 655 [20] L. Wang, L. Zhao, G. Bi, C. Wan, Sparse representation-based isar imaging
using markov random fields, *IEEE J. Sel. Topics Appl. Earth Observ. in Re-
mote Sens.* 8 (8) (2015) 3941–3953. doi:10.1109/JSTARS.2014.2359250.
- [21] S. Suwanwimolkul, L. Zhang, D. Gong, Z. Zhang, C. Chen, D. C. Ranas-
inghe, Q. Shi, An adaptive markov random field for structured compressive
660 sensingSubmitted to *IEEE Trans. Image Proc.*
URL <https://arxiv.org/abs/1802.05395>
- [22] D. G. Tzikas, A. C. Likas, N. P. Galatsanos, The variational approximation
for bayesian inference, *IEEE Signal Process. Mag.* 25 (6) (2008) 131–146.
doi:10.1109/MSP.2008.929620.
- 665 [23] J. Besag, Statistical analysis of non-lattice data, *J. R. Stat. Soc., Ser. D
Stat.* 24 (3) (1975) 179–195.
URL <http://www.jstor.org/stable/2987782>

- [24] D. P. Wipf, B. D. Rao, An empirical bayesian strategy for solving the simultaneous sparse approximation problem, *IEEE Trans. Signal Process.* 55 (7) (2007) 3704–3716. doi:10.1109/TSP.2007.894265.
- [25] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [26] T. Sim, S. Baker, M. Bsat, The cmu pose, illumination, and expression database, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (12) (2003) 1615 – 1618.
- [27] A. Krizhevsky, Learning multiple layers of features from tiny images, Tech. rep., University of Toronto (2009).
- [28] L. Zhang, W. Wei, C. Tian, F. Li, Y. Zhang, Exploring structured sparsity by a reweighted laplace prior for hyperspectral compressive sensing, *IEEE Trans. Image Process.* 25 (10) (2016) 4974–4988. doi:10.1109/TIP.2016.2598652.
- [29] J. Tropp, A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory* 53 (12) (2007) 4655–4666.

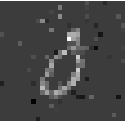

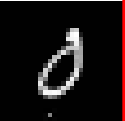
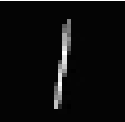

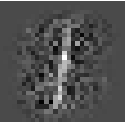

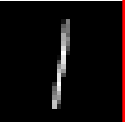
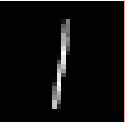
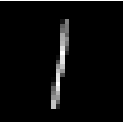


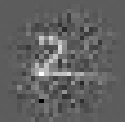





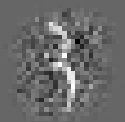


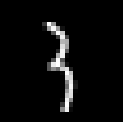
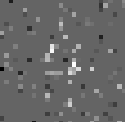

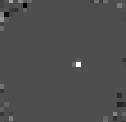
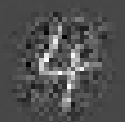


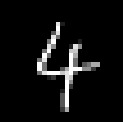


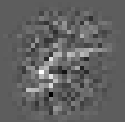

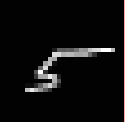
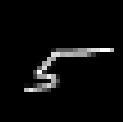




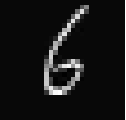
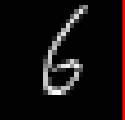
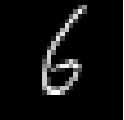
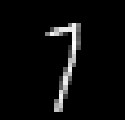




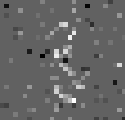


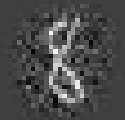








OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
								
12.49 dB	18.40 dB	30.26 dB	40.36 dB	20.50 dB	41.69 dB	42.22 dB	44.25 dB	
								
40.57 dB	47.75 dB	46.79 dB	46.00 dB	19.62 dB	46.24 dB	53.29 dB	54.76 dB	
								
36.52 dB	29.57 dB	39.37 dB	42.76 dB	17.78 dB	24.93 dB	44.36 dB	46.12 dB	
								
37.16 dB	40.06 dB	39.18 dB	44.63 dB	18.16 dB	44.66 dB	46.45 dB	48.50 dB	
								
12.99 dB	16.14 dB	27.62 dB	31.51 dB	17.85 dB	42.42 dB	43.55 dB	45.17 dB	
								
37.48 dB	43.82 dB	44.14 dB	42.48 dB	18.17 dB	25.96 dB	43.30 dB	49.74 dB	
								
34.00 dB	22.87 dB	23.59 dB	42.30 dB	20.68 dB	43.35 dB	45.56 dB	45.38 dB	
								
37.65 dB	41.30 dB	40.18 dB	43.40 dB	19.01 dB	44.31 dB	46.50 dB	47.09 dB	
								
12.24 dB	19.83 dB	22.85 dB	35.85 dB	18.00 dB	40.48 dB	40.12 dB	42.91 dB	
								
31.97 dB	27.78 dB	31.13 dB	34.91 dB	17.99 dB	40.81 dB	42.68 dB	44.86 dB	

Figure 22: Visual results of MNIST handwritten digit images (at $M/N = 0.3$, SNR = 30 dB).

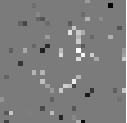
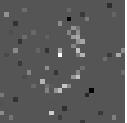
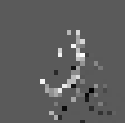
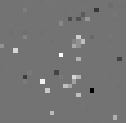
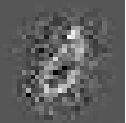
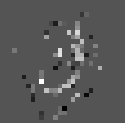
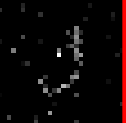
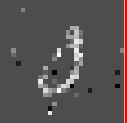
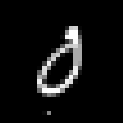
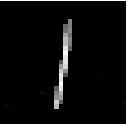
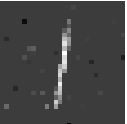
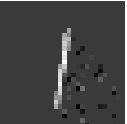
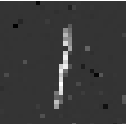
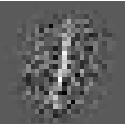
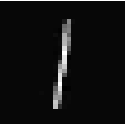
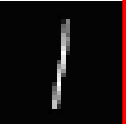
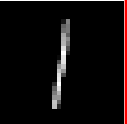
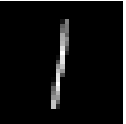
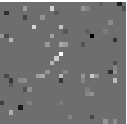
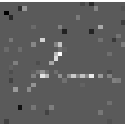
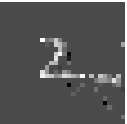
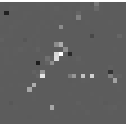
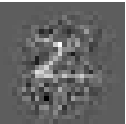
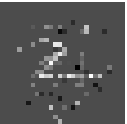
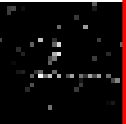
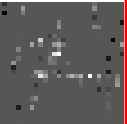

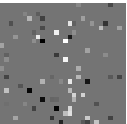
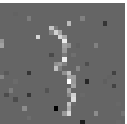

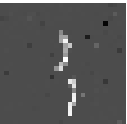
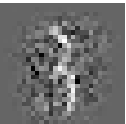
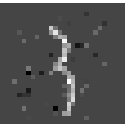

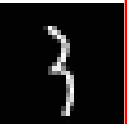
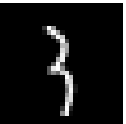
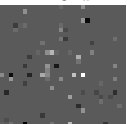
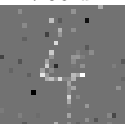


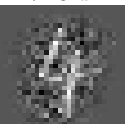
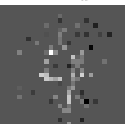


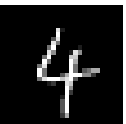
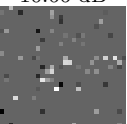
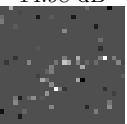

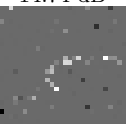
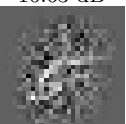
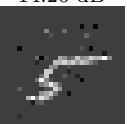

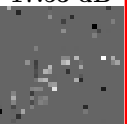
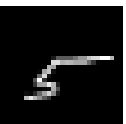
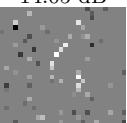
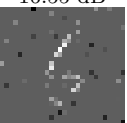
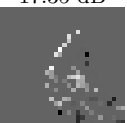
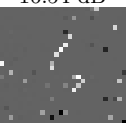
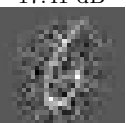
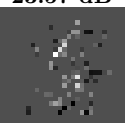
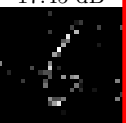
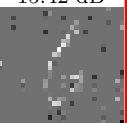
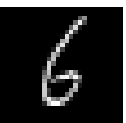
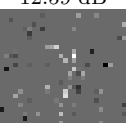
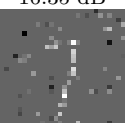
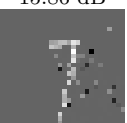
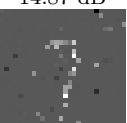
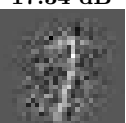

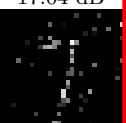
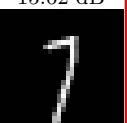
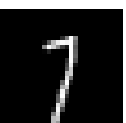
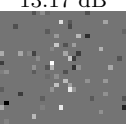
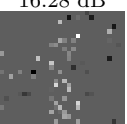
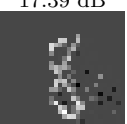
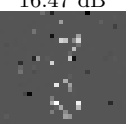
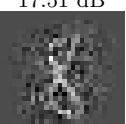
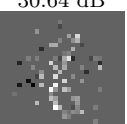
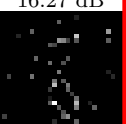

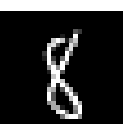
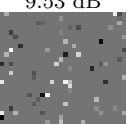
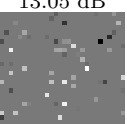
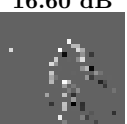
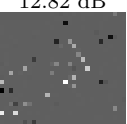
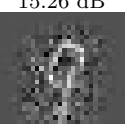
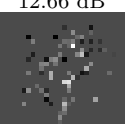
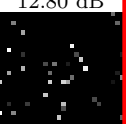
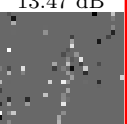
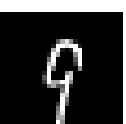
OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
								
11.89 dB	14.88 dB	15.10 dB	13.72 dB	17.01 dB	14.49 dB	15.34 dB	18.09 dB	
								
37.90 dB	26.06 dB	25.01 dB	24.15 dB	18.97 dB	47.68 dB	50.10 dB	52.78 dB	
								
12.03 dB	16.20 dB	18.89 dB	15.25 dB	16.85 dB	15.68 dB	16.07 dB	14.79 dB	
								
12.49 dB	17.96 dB	25.29 dB	18.05 dB	17.18 dB	22.21 dB	43.66 dB	41.04 dB	
								
10.66 dB	14.98 dB	16.24 dB	14.74 dB	16.63 dB	14.26 dB	15.88 dB	17.53 dB	
								
14.05 dB	16.55 dB	17.35 dB	16.54 dB	17.41 dB	23.57 dB	17.45 dB	15.42 dB	
								
12.59 dB	16.35 dB	15.86 dB	14.87 dB	17.34 dB	13.94 dB	17.04 dB	15.62 dB	
								
13.17 dB	16.28 dB	17.39 dB	16.47 dB	17.51 dB	30.64 dB	16.27 dB	46.19 dB	
								
9.53 dB	13.05 dB	16.60 dB	12.82 dB	15.26 dB	12.66 dB	12.80 dB	13.47 dB	
								
10.93 dB	12.69 dB	13.95 dB	12.09 dB	16.11 dB	12.99 dB	12.39 dB	11.92 dB	

Figure 23: Visual results of MNIST handwritten digit images (at $M/N = 0.2$, SNR = 30 dB).

OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
30.08 dB	30.85 dB	31.95 dB	20.55 dB	23.94 dB	28.94 dB	31.39 dB	32.46 dB	
32.32 dB	32.86 dB	28.25 dB	26.47 dB	26.40 dB	31.42 dB	34.33 dB	34.82 dB	
30.80 dB	32.25 dB	31.33 dB	24.03 dB	24.77 dB	29.56 dB	32.90 dB	33.26 dB	
30.37 dB	32.11 dB	27.65 dB	21.92 dB	25.23 dB	29.67 dB	32.91 dB	33.44 dB	
28.96 dB	30.63 dB	30.51 dB	20.11 dB	23.35 dB	28.22 dB	31.62 dB	32.20 dB	
29.99 dB	31.49 dB	28.84 dB	20.69 dB	23.24 dB	28.98 dB	32.27 dB	31.87 dB	
29.65 dB	31.47 dB	30.33 dB	22.61 dB	26.25 dB	28.45 dB	31.58 dB	33.98 dB	
28.19 dB	30.01 dB	25.79 dB	20.52 dB	23.39 dB	28.58 dB	31.48 dB	33.18 dB	
28.60 dB	30.12 dB	28.69 dB	20.80 dB	22.62 dB	28.56 dB	31.91 dB	33.52 dB	
31.68 dB	32.10 dB	29.14 dB	23.89 dB	24.13 dB	29.83 dB	33.07 dB	34.15 dB	

Figure 24: Visual results of CMU-IDB face images from PCA sparse signal reconstruction (at $M/N = 0.3$, SNR = 30 dB).

OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
								
29.33 dB	26.68 dB	26.81 dB	13.22 dB	21.19 dB	26.33 dB	30.25 dB	30.08 dB	
								
29.80 dB	30.46 dB	28.07 dB	18.60 dB	23.60 dB	29.24 dB	31.12 dB	31.27 dB	
								
27.92 dB	28.47 dB	25.15 dB	17.06 dB	21.83 dB	26.59 dB	28.72 dB	30.96 dB	
								
27.60 dB	27.95 dB	26.61 dB	16.08 dB	22.27 dB	26.89 dB	30.59 dB	30.91 dB	
								
27.86 dB	27.12 dB	22.60 dB	16.79 dB	20.79 dB	26.59 dB	29.99 dB	29.41 dB	
								
28.08 dB	27.18 dB	20.58 dB	16.10dB	20.54 dB	26.46 dB	29.24 dB	30.06 dB	
								
27.12 dB	27.63 dB	22.84 dB	18.81 dB	22.69 dB	25.32 dB	28.93 dB	31.05 dB	
								
26.37 dB	26.58 dB	26.17 dB	13.29 dB	20.57 dB	24.931 dB	27.38 dB	27.22 dB	
								
27.50 dB	26.59 dB	23.79 dB	16.91 dB	19.94 dB	24.91 dB	28.30 dB	28.22 dB	
								
29.02 dB	29.53 dB	27.60 dB	19.07 dB	21.26 dB	28.44 dB	32.00 dB	33.69 dB	

Figure 25: Visual results of CMU-1DB face images from PCA sparse signal reconstruction (at $M/N = 0.2$, SNR = 30 dB).

OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
17.83 dB	17.49 dB	16.96 dB	14.91 dB	5.78 dB	17.61 dB	19.42 dB	19.81 dB	
19.83 dB	20.95 dB	20.03 dB	19.13 dB	7.81 dB	19.06 dB	21.34 dB	23.88 dB	
15.08 dB	16.97 dB	14.56 dB	14.97 dB	6.69 dB	16.08 dB	17.18 dB	19.387 dB	
15.94 dB	15.65 dB	17.41 dB	14.10 dB	5.88 dB	16.66 dB	17.30 dB	19.01 dB	
16.61 dB	17.82 dB	18.25 dB	16.12 dB	6.19 dB	16.68 dB	18.97 dB	20.37 dB	
16.74 dB	16.52 dB	14.76 dB	15.85 dB	6.03 dB	17.49 dB	18.15 dB	18.70 dB	
16.91 dB	17.28 dB	19.50 dB	15.50 dB	4.94 dB	17.8 dB	18.91 dB	19.16 dB	
14.04 dB	16.06 dB	13.76 dB	13.81 dB	4.74 dB	14.48 dB	15.41 dB	17.53 dB	
16.91 dB	16.59 dB	16.02 dB	15.77 dB	5.36 dB	17.45 dB	18.44 dB	19.75 dB	
17.30 dB	17.60 dB	15.11 dB	15.63 dB	6.80 dB	17.30 dB	18.98 dB	20.38 dB	

Figure 26: Visual results of CMU-IDB face images from wavelet sparse signal reconstruction (at $M/N = 0.3$, SNR = 30 dB).

OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
16.23 dB	16.67 dB	17.22 dB	12.79 dB	5.26 dB	16.17 dB	16.21 dB	18.47 dB	
19.18 dB	20.25 dB	18.95 dB	17.48 dB	7.45 dB	18.61 dB	19.68 dB	22.47 dB	
14.81 dB	15.98 dB	14.21 dB	12.38 dB	6.65 dB	13.75 dB	15.65 dB	16.80 dB	
12.96 dB	14.71 dB	12.21 dB	11.60 dB	5.99 dB	13.66 dB	15.33 dB	15.76 dB	
14.97 dB	15.82 dB	16.38 dB	12.84 dB	6.25 dB	14.61 dB	15.99 dB	18.48 dB	
14.82 dB	15.47 dB	12.38 dB	10.93 dB	5.76 dB	15.42 dB	16.20 dB	17.72 dB	
16.41 dB	16.23 dB	14.90 dB	12.95 dB	4.78 dB	16.14 dB	15.79 dB	18.11 dB	
12.70 dB	14.07 dB	11.69 dB	11.29 dB	4.55 dB	13.18 dB	13.55 dB	14.65 dB	
12.91 dB	14.59 dB	11.43 dB	11.71 dB	5.34 dB	14.46 dB	15.99 dB	17.24 dB	
15.71 dB	17.30 dB	16.02 dB	13.98 dB	6.54 dB	16.68 dB	17.29 dB	19.76 dB	

Figure 27: Visual results of CMU-IDB face images from wavelet sparse signal reconstruction (at $M/N = 0.2$, $\text{SNR} = 30$ dB).

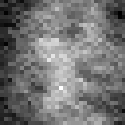
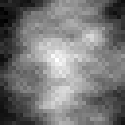
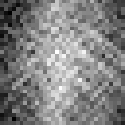
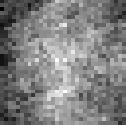
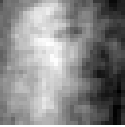
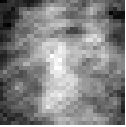
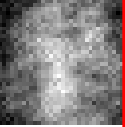
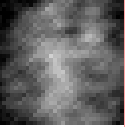

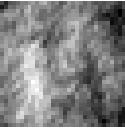
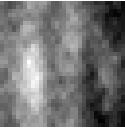
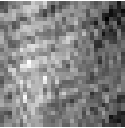
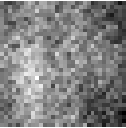
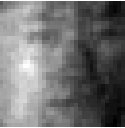
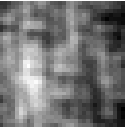
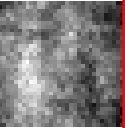
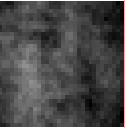

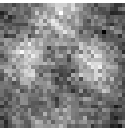
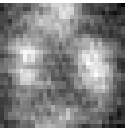
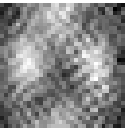
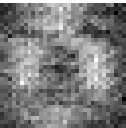
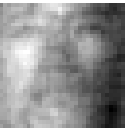
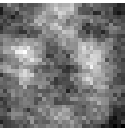
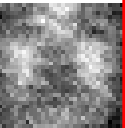
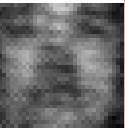

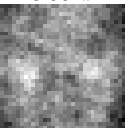
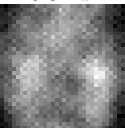
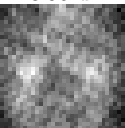
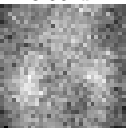
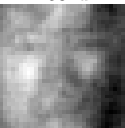
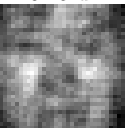
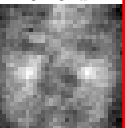
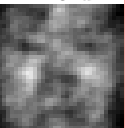

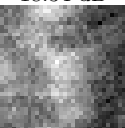
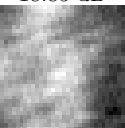
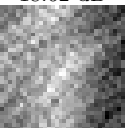
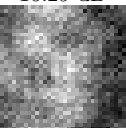
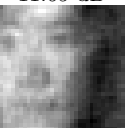
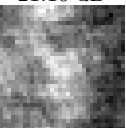
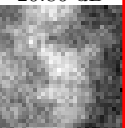
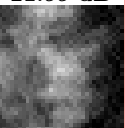

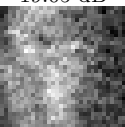
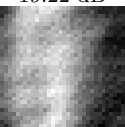
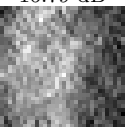
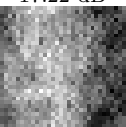

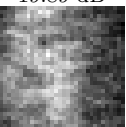
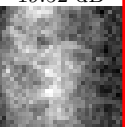
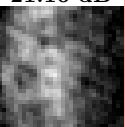

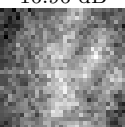
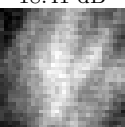
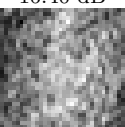
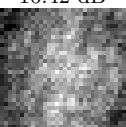
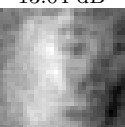
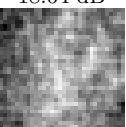
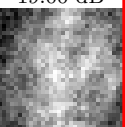


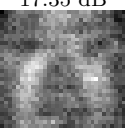
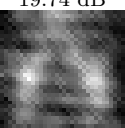
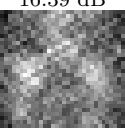
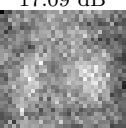
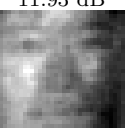
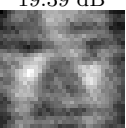
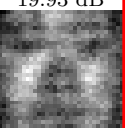


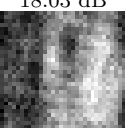
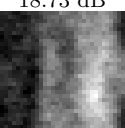
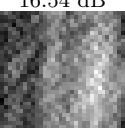
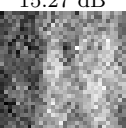
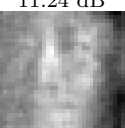
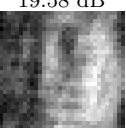
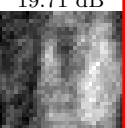
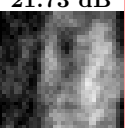

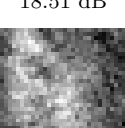
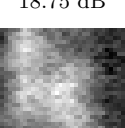
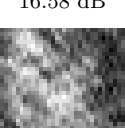
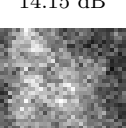
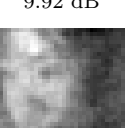
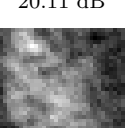
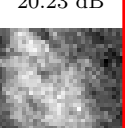
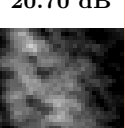

OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
								
19.65 dB	20.86 dB	17.46 dB	17.86 dB	11.95 dB	21.05 dB	21.14 dB	22.42 dB	
								
23.85 dB	24.40 dB	20.32 dB	21.08 dB	16.99 dB	24.97 dB	24.74 dB	25.42 dB	
								
18.50 dB	20.31 dB	18.95 dB	18.55 dB	12.59 dB	20.25 dB	20.29 dB	22.15 dB	
								
18.34 dB	18.85 dB	18.62 dB	16.29 dB	11.65 dB	21.10 dB	20.86 dB	21.99 dB	
								
19.05 dB	19.22 dB	16.79 dB	17.22 dB	12.74 dB	19.89 dB	19.52 dB	21.16 dB	
								
16.96 dB	18.41 dB	16.40 dB	16.42 dB	13.04 dB	18.64 dB	19.00 dB	20.27 dB	
								
17.35 dB	19.74 dB	16.39 dB	17.09 dB	11.93 dB	19.39 dB	19.93 dB	21.69 dB	
								
18.03 dB	18.73 dB	16.54 dB	15.27 dB	11.24 dB	19.58 dB	19.71 dB	21.73 dB	
								
18.51 dB	18.75 dB	16.58 dB	14.15 dB	9.92 dB	20.11 dB	20.23 dB	20.70 dB	
								
17.85 dB	18.61 dB	18.18 dB	18.18 dB	14.42 dB	20.0 dB	19.42 dB	21.19 dB	

Figure 28: Visual results of CMU-1DB face images from DCT sparse signal reconstruction (at $M/N = 0.3$, SNR = 30 dB).

OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
17.94 dB	19.85 dB	17.70 dB	15.10 dB	12.04 dB	17.90 dB	18.90 dB	19.51 dB	
22.59 dB	23.16 dB	21.34 dB	16.91 dB	16.64 dB	22.79 dB	23.07 dB	23.31 dB	
18.26 dB	18.62 dB	17.28 dB	13.96 dB	12.82 dB	18.61 dB	19.70 dB	20.48 dB	
17.58 dB	17.52 dB	16.64 dB	14.02 dB	11.68 dB	16.72 dB	18.20 dB	19.08 dB	
18.03 dB	18.89 dB	17.38 dB	12.94 dB	13.07 dB	17.79 dB	18.87 dB	20.21 dB	
15.65 dB	17.78 dB	14.40 dB	12.79 dB	12.92 dB	15.88 dB	16.85 dB	18.32 dB	
16.48 dB	18.62 dB	17.93 dB	14.43 dB	12.29 dB	17.12 dB	17.98 dB	19.18 dB	
15.31 dB	16.33 dB	11.12 dB	13.53 dB	11.33 dB	16.37 dB	16.73 dB	17.57 dB	
16.72 dB	16.51 dB	16.51 dB	13.47 dB	10.30 dB	17.33 dB	17.70 dB	17.61 dB	
16.85 dB	17.69 dB	14.58 dB	13.32 dB	14.02 dB	16.97 dB	18.14 dB	18.32 dB	

Figure 29: Visual results of CMU-IDB face images from DCT sparse signal reconstruction (at $M/N = 0.2$, SNR = 30 dB).

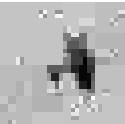
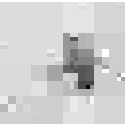
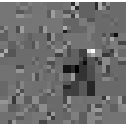
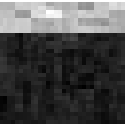
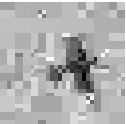




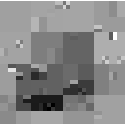
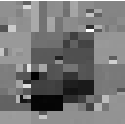
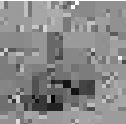
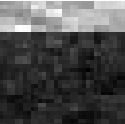
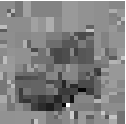


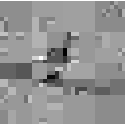


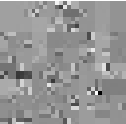
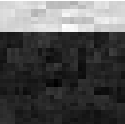

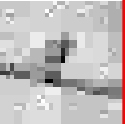





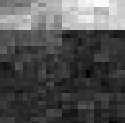




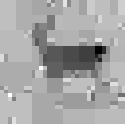


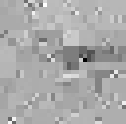
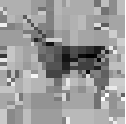
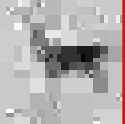
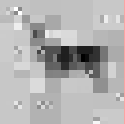
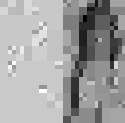

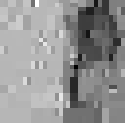
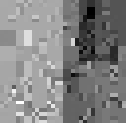

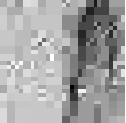
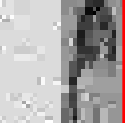
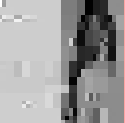
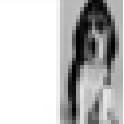
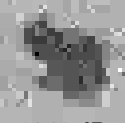



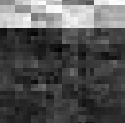
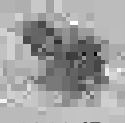
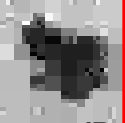





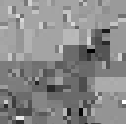
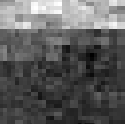

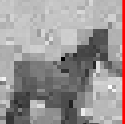





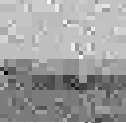

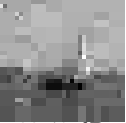


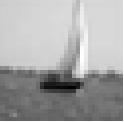









OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
 20.50 dB	 18.36 dB	 16.55 dB	 12.48 dB	 1.17 dB	 16.31 dB	 23.15 dB	 23.52 dB	
 16.32 dB	 15.37 dB	 15.15 dB	 12.79 dB	 1.98 dB	 16.31 dB	 18.35 dB	 18.80 dB	
 19.04 dB	 19.09 dB	 17.95 dB	 14.09 dB	 1.59 dB	 17.68 dB	 22.00 dB	 22.87 dB	
 18.19 dB	 16.52 dB	 19.06 dB	 13.25 dB	 2.67 dB	 14.25 dB	 19.29 dB	 18.638 dB	
 18.47 dB	 17.42 dB	 20.39 dB	 12.96 dB	 1.19 dB	 14.86 dB	 19.52 dB	 19.90 dB	
 18.37 dB	 17.84 dB	 17.07 dB	 14.40 dB	 1.58 dB	 15.87 dB	 19.27 dB	 19.86 dB	
 16.39 dB	 14.56 dB	 17.57 dB	 10.30 dB	 1.82 dB	 14.84 dB	 18.14 dB	 18.13 dB	
 15.02 dB	 15.05 dB	 13.25 dB	 11.93 dB	 4.26 dB	 13.66 dB	 17.96 dB	 18.233 dB	
 24.04 dB	 23.62 dB	 22.28 dB	 18.50 dB	 6.09 dB	 22.48 dB	 24.46 dB	 26.23 dB	
 19.32 dB	 19.79 dB	 18.72 dB	 16.20 dB	 5.29 dB	 18.65 dB	 21.12 dB	 22.28 dB	

Figure 30: Visual results of CIFAR-10 from wavlet sparse signal reconstruction (at $M/N = 0.3$, SNR = 30 dB).

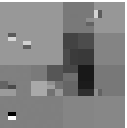
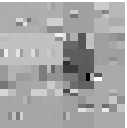
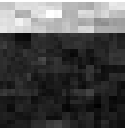


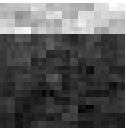
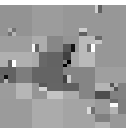

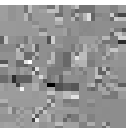
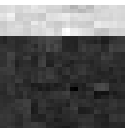
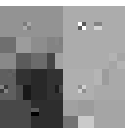



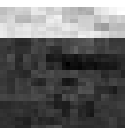


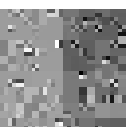







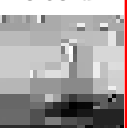
OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
 18.34 dB	 16.09 dB	 15.21 dB	 10.10 dB	 1.45 dB	 12.25 dB	 15.60 dB	 19.36 dB	
 15.67 dB	 13.26 dB	 11.83 dB	 9.46 dB	 2.29 dB	 11.47 dB	 15.32 dB	 16.38 dB	
 18.33 dB	 17.50 dB	 16.39 dB	 13.68 dB	 1.61 dB	 13.50 dB	 18.24 dB	 20.03 dB	
 15.64 dB	 14.02 dB	 11.22 dB	 8.28 dB	 2.29 dB	 10.83 dB	 16.33 dB	 17.90 dB	
 15.80 dB	 14.72 dB	 16.86 dB	 10.48 dB	 1.38 dB	 11.11 dB	 16.60 dB	 16.43 dB	
 15.37 dB	 15.65 dB	 14.00 dB	 10.12 dB	 1.54 dB	 11.70 dB	 15.68 dB	 18.54 dB	
 13.56 dB	 10.56 dB	 8.24 dB	 7.05 dB	 2.31 dB	 10.59 dB	 15.17 dB	 14.86 dB	
 13.49 dB	 12.21 dB	 8.28 dB	 9.82 dB	 4.61 dB	 11.00 dB	 16.27 dB	 14.68 dB	
 22.25 dB	 20.95 dB	 21.72 dB	 14.35 dB	 5.91 dB	 17.12 dB	 20.96 dB	 23.53 dB	
 18.63 dB	 17.41 dB	 11.35 dB	 9.20 dB	 5.20 dB	 14.84 dB	 19.59 dB	 20.08 dB	

Figure 31: Visual results of CIFAR-10 from wavlet sparse signal reconstruction (at $M/N = 0.2$, $\text{SNR} = 30$ dB).

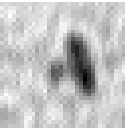
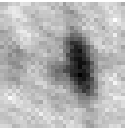
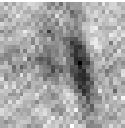
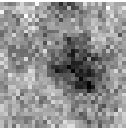
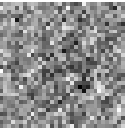
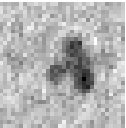
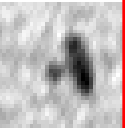
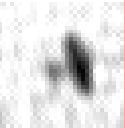

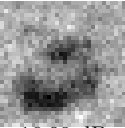
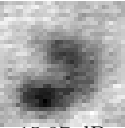
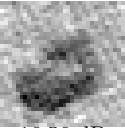
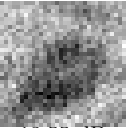
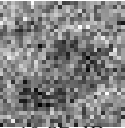
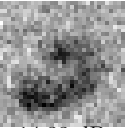
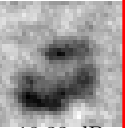


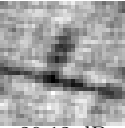
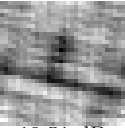
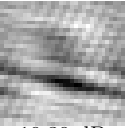
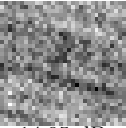
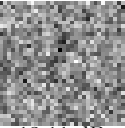
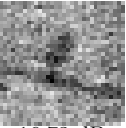
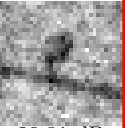
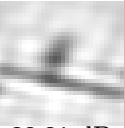

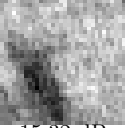
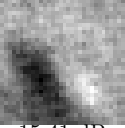
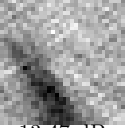
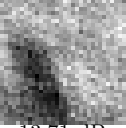
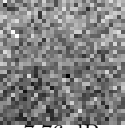
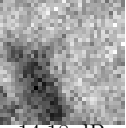
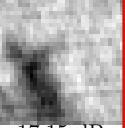
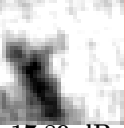

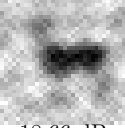
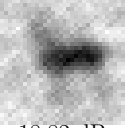
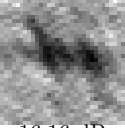
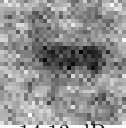
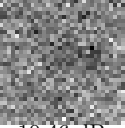
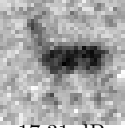
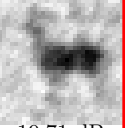


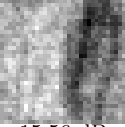
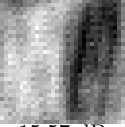
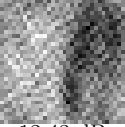
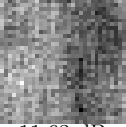
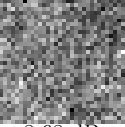
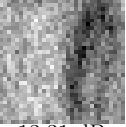
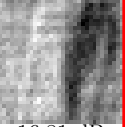


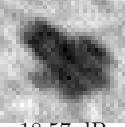
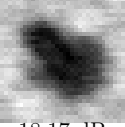
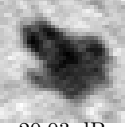
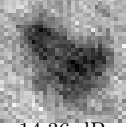
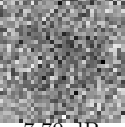
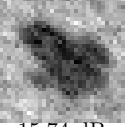
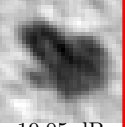


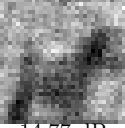
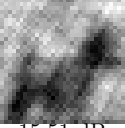

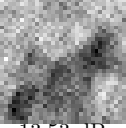
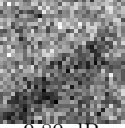
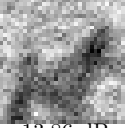
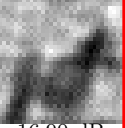


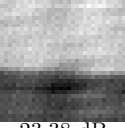
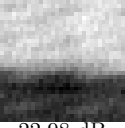
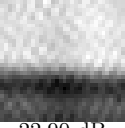
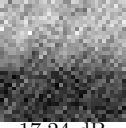
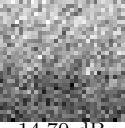
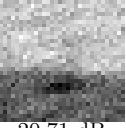
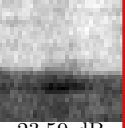


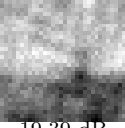
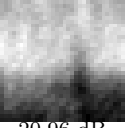

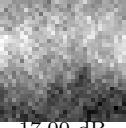
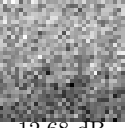
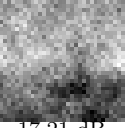
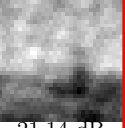


OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
 19.30 dB	 18.03 dB	 15.51 dB	 14.12 dB	 11.01 dB	 16.50 dB	 20.08 dB	 20.64 dB	
 16.09 dB	 15.87 dB	 16.56 dB	 13.55 dB	 10.57 dB	 14.33 dB	 18.33 dB	 19.64 dB	
 20.12 dB	 19.51 dB	 19.39 dB	 14.95 dB	 13.11 dB	 16.73 dB	 20.81 dB	 23.31 dB	
 15.32 dB	 15.41 dB	 13.47 dB	 13.71 dB	 7.76 dB	 14.18 dB	 17.15 dB	 17.89 dB	
 18.66 dB	 18.82 dB	 16.16 dB	 14.13 dB	 10.46 dB	 17.31 dB	 19.71 dB	 20.82 dB	
 15.56 dB	 15.57 dB	 12.43 dB	 11.63 dB	 8.68 dB	 13.31 dB	 16.81 dB	 18.90 dB	
 18.57 dB	 18.17 dB	 20.03 dB	 14.36 dB	 7.70 dB	 15.74 dB	 19.05 dB	 20.85 dB	
 14.77 dB	 15.51 dB	 17.95 dB	 13.53 dB	 9.80 dB	 13.86 dB	 16.90 dB	 19.59 dB	
 23.38 dB	 22.98 dB	 22.90 dB	 17.24 dB	 14.70 dB	 20.71 dB	 23.59 dB	 24.36 dB	
 19.39 dB	 20.96 dB	 19.19 dB	 17.00 dB	 12.68 dB	 17.21 dB	 21.14 dB	 23.24 dB	

Figure 32: Visual results of CIFAR-10 natural images from DCT sparse signal reconstruction (at $M/N = 0.3$, SNR = 30 dB).

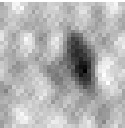
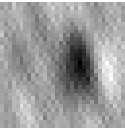
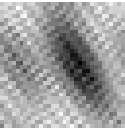
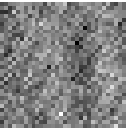
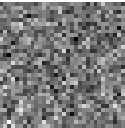
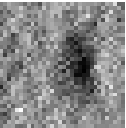
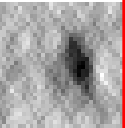
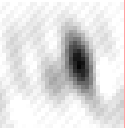

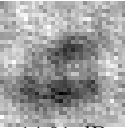
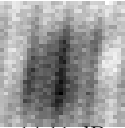
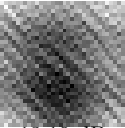
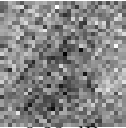
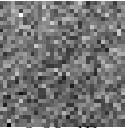
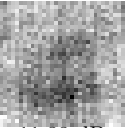
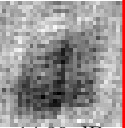
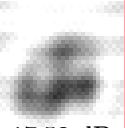

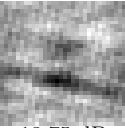
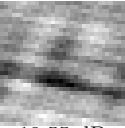
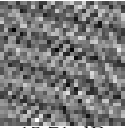
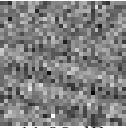
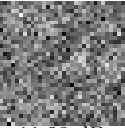
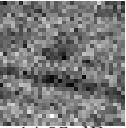
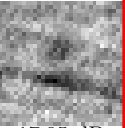


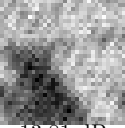
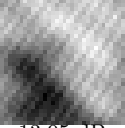
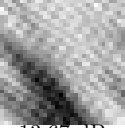
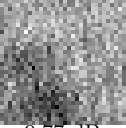
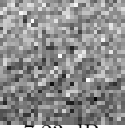
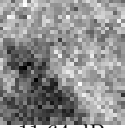
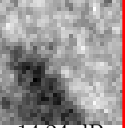
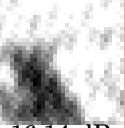

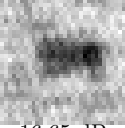
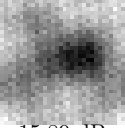
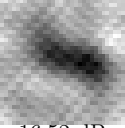
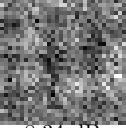
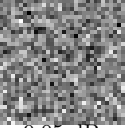
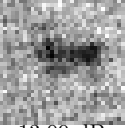
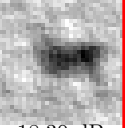


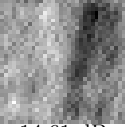
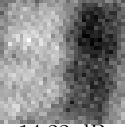
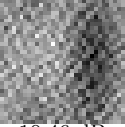
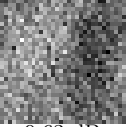
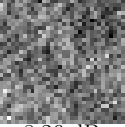
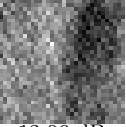
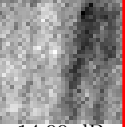


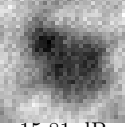
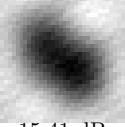
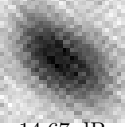
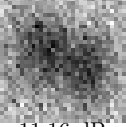
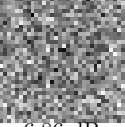
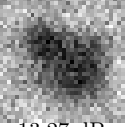
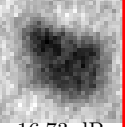


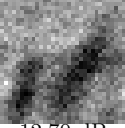
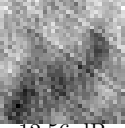
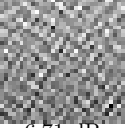
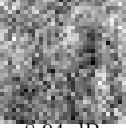
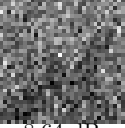
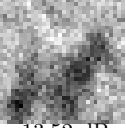
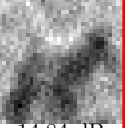


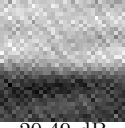
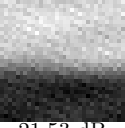
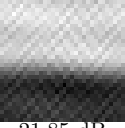
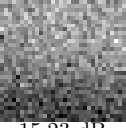
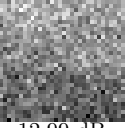
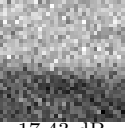
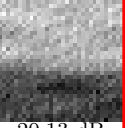


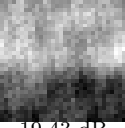


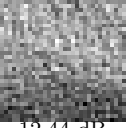
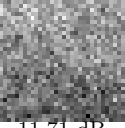
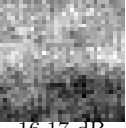
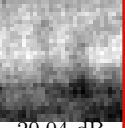


OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
								
17.35 dB	15.84 dB	15.33 dB	8.67 dB	9.15 dB	12.68 dB	17.08 dB	18.92 dB	
								
14.01 dB	14.11 dB	13.10 dB	9.28 dB	8.80 dB	11.88 dB	14.89 dB	17.58 dB	
								
18.75 dB	19.55 dB	15.71 dB	11.96 dB	11.08 dB	14.05 dB	17.95 dB	20.22 dB	
								
13.81 dB	13.95 dB	13.67 dB	9.77 dB	7.23 dB	11.64 dB	14.24 dB	16.14 dB	
								
16.65 dB	15.89 dB	16.52 dB	8.34 dB	9.05 dB	13.09 dB	18.30 dB	20.39 dB	
								
14.61 dB	14.22 dB	10.46 dB	9.63 dB	8.26 dB	12.06 dB	14.99 dB	17.53 dB	
								
15.81 dB	15.41 dB	14.67 dB	11.16 dB	6.86 dB	13.27 dB	16.73 dB	17.53 dB	
								
12.70 dB	12.56 dB	6.71 dB	8.94 dB	8.64 dB	13.52 dB	14.84 dB	16.74 dB	
								
20.49 dB	21.53 dB	21.85 dB	15.23 dB	12.99 dB	17.43 dB	20.13 dB	22.72 dB	
								
19.43 dB	18.39 dB	18.68 dB	12.44 dB	11.71 dB	16.17 dB	20.04 dB	20.87 dB	

Figure 33: Visual results of CIFAR-10 natural images from DCT sparse signal reconstruction (at $M/N = 0.2$, SNR = 30 dB).

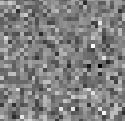
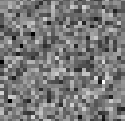
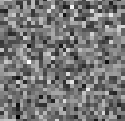
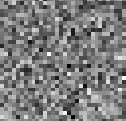
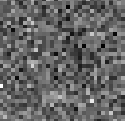
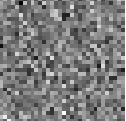
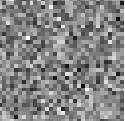
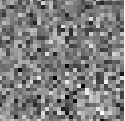

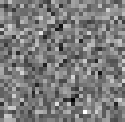
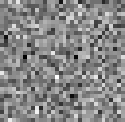
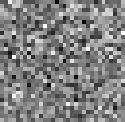
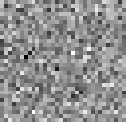
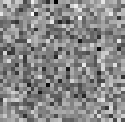
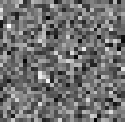
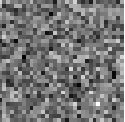
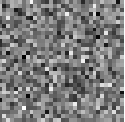

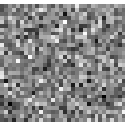
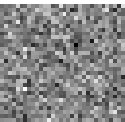
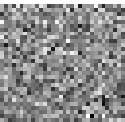
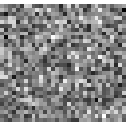
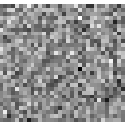
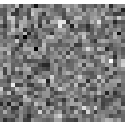
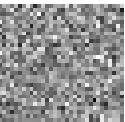
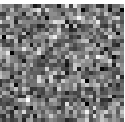

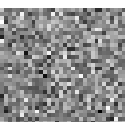
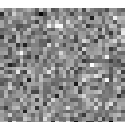
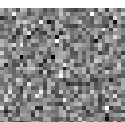
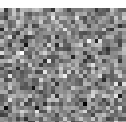
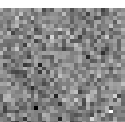
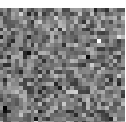
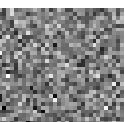
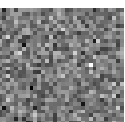

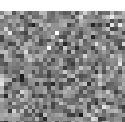
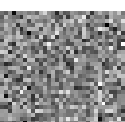
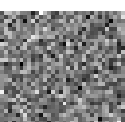
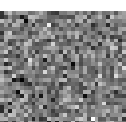
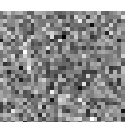
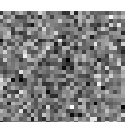
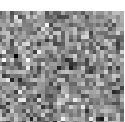
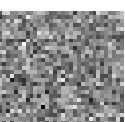

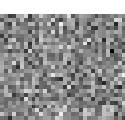
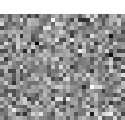
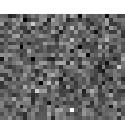
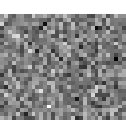
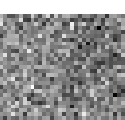
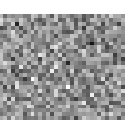
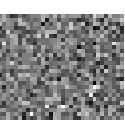
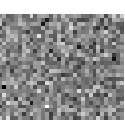

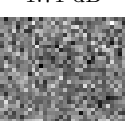
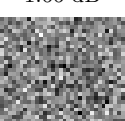
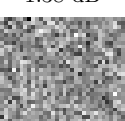
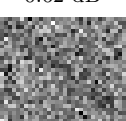
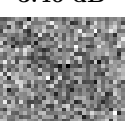
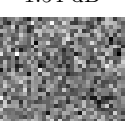
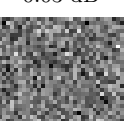
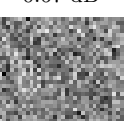

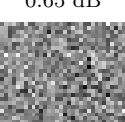
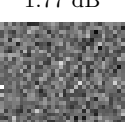
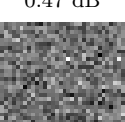
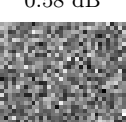
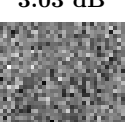
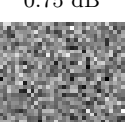
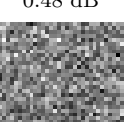
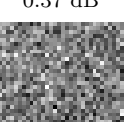

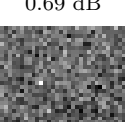
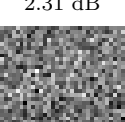
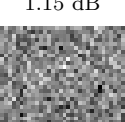
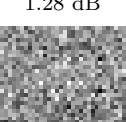
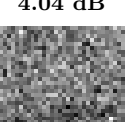
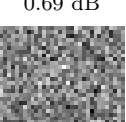
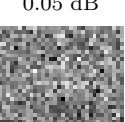
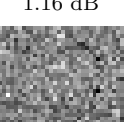

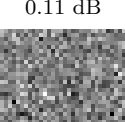
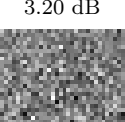
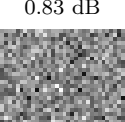
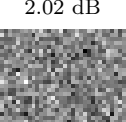
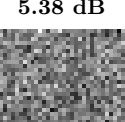
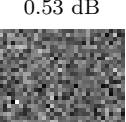
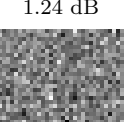
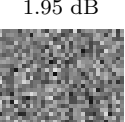

OMP	RLPHCS	Bernoulli	Pairwise MRF	Gibbs	MAP-OMP	TA-MRF	OA-MRF (Ours)	Ground Truth
 2.98 dB	 0.76 dB	 1.23 dB	 0.91 dB	 2.80 dB	 2.83 dB	 1.07 dB	 0.57 dB	 1.24 dB
 2.10 dB	 0.86 dB	 1.56 dB	 0.05 dB	 2.77 dB	 2.29 dB	 0.44 dB	 0.26 dB	 0.76 dB
 1.56 dB	 1.45 dB	 1.13 dB	 0.28 dB	 3.03 dB	 0.76 dB	 0.76 dB	 0.33 dB	 0.47 dB
 1.77 dB	 1.25 dB	 1.02 dB	 0.30 dB	 2.79 dB	 2.06 dB	 0.31 dB	 0.04 dB	 0.75 dB
 0.86 dB	 1.11 dB	 1.46 dB	 0.05 dB	 2.74 dB	 0.98 dB	 0.49 dB	 0.03 dB	 0.45 dB
 1.71 dB	 1.66 dB	 1.38 dB	 0.62 dB	 3.46 dB	 1.54 dB	 0.03 dB	 0.67 dB	 0.03 dB
 0.65 dB	 1.77 dB	 0.47 dB	 0.58 dB	 3.03 dB	 0.75 dB	 0.48 dB	 0.37 dB	 0.42 dB
 0.69 dB	 2.31 dB	 1.15 dB	 1.28 dB	 4.04 dB	 0.69 dB	 0.05 dB	 1.16 dB	 0.77 dB
 0.11 dB	 3.20 dB	 0.83 dB	 2.02 dB	 5.38 dB	 0.53 dB	 1.24 dB	 1.95 dB	 1.22 dB
 0.12 dB	 3.01 dB	 0.81 dB	 2.41 dB	 4.61 dB	 0.70 dB	 1.56 dB	 2.31 dB	 1.66 dB

Figure 34: Every algorithm fails in reconstruction of CIFAR-10 natural images in PCA domain (at $M/N = 0.3$, $\text{SNR} = 30$ dB).