Variational Bayesian Learning for Piecewise Sparse Signal Recovery from Quantized Measurements

Sai Subramanyam Thoota and Chandra R. Murthy

Abstract—We develop variational Bayesian (VB) learning procedures to infer the posterior distribution of *piecewise* joint-sparse vectors from both unquantized and quantized compressive multiple measurement vectors. We overcome the analytical intractability in computing the exact posterior by imposing a structure on the posterior distribution and choosing an appropriate piecewise sparsity promoting conjugate prior. This results in an iterative VB algorithm that converges to a local optimum or stationary point of the underlying optimization problem from any initialization. We apply the piecewise sparse recovery algorithm to a massive multiple-input-multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) channel estimation problem. We evaluate the normalized mean-squared error (NMSE) and run-time performance of the piecewise sparse recovery procedures for the unquantized and quantized cases, and benchmark them against the state-of-the-art. We empirically show that utilizing the inherent piecewise sparse structure significantly reduces the computational complexity without compromising on the performance.

Index Terms—Compressed sensing, MIMO, OFDM, quantization, posterior distribution, variational Bayes.

I. INTRODUCTION

Sparse signal recovery refers to the problem of recovering high-dimensional vectors with mostly zeros as its components, from a set of noisy underdetermined linear measurements [1], [2]. Piecewise sparsity arises when each sub-block of a sparse signal is itself sparse [3], [4], so that the nonzero entries in a piecewise sparse vector are generally "spread out", and do not occur in clusters. This structure is important in practical applications such as the massive multiple-input-multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) channel estimation in 5G and beyond communication systems. Here, the concatenated channel impulse responses of all the users is sparse in the lag domain. Further, each user's channel is individually sparse, which results in a piecewise sparse structure. When the training signals from the users are received using multiple antennas at the base station (BS), the channels to different antennas have a common support, which leads to the *multiple measurement vector* (also called joint-sparse or row-sparse in the literature) *piecewise sparse recovery* problem.

Several greedy and Bayesian algorithms exist in the literature to exploit the different sparsity structures in the signals, such asblock sparsity, joint/row sparsity, intra-vector correlation [5]–[14]. However, piecewise sparsity has received less attention in the compressed sensing literature [3], [4]. Further, quantized compressed sensing is another important research topic which also finds applications in wireless communications. In massive MIMO systems, low resolution quantization (1 to 3 bits) results in large cost and power savings, especially when there are tens or hundreds of antennas in the base station [15]. Piecewise sparse signal recovery using noisy and low resolution quantized compressive measurements that can cater to such applications has not been considered in the literature, making the problem both timely and important.

In this letter, we consider the estimation of a piecewise joint-sparse $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T] \in \mathbb{C}^{N \times T}$ from noisy low-dimensional quantized measurements $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_T] \in \mathbb{C}^{M \times T}$, where $M < N$. The measurements are obtained as

$$\mathbf{Q} = \mathcal{Q}_b(\mathbf{Y}) = \mathcal{Q}_b(\mathbf{\Phi}\mathbf{X} + \mathbf{W}), \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ is a known measurement matrix, $\mathbf{Y} \in \mathbb{C}^{M \times T}$ is the unquantized signal, and $\mathbf{W} \in \mathbb{C}^{M \times T}$ is the additive noise matrix whose entries are independent and identically distributed (i.i.d.) as circularly symmetric complex Gaussian random variables with mean 0 and variance σ^2. The estimand $\mathbf{X}$ is joint-sparse, i.e., the indices of the nonzero entries of each column of $\mathbf{X}$ are the same. In addition, it has a *piecewise sparse* structure. That is, if we divide $\mathbf{X}$ into K sub-matrices $\mathbf{X}[1], \ldots, \mathbf{X}[K]$, where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}[1] \\ \vdots \\ \mathbf{X}[K] \end{bmatrix}, \mathbf{X}[k] = \begin{bmatrix} \mathbf{x}_1[k] & \ldots & \mathbf{x}_T[k] \end{bmatrix} \in \mathbb{C}^{L \times T}, \tag{2}$$

for $k = 1, \ldots, K$, (here, $N = LK$) then each $\mathbf{X}[k]$ is itself joint-sparse. Also, $\mathcal{Q}_b(\cdot)$ denotes an element-wise scalar b-bit quantization operation of the real and imaginary components of its argument. The quantizer is deterministic and known. In our experiments, we use a uniform scalar quantizer, however, the algorithm developed in the next section is applicable to any deterministic quantization function.

Our goal is to exploit the piecewise sparse structure to estimate the high-dimensional signal $\mathbf{X}$ using the quantized measurements $\mathbf{Q}$ and the measurement matrix $\mathbf{\Phi}$. The unquantized sparse signal recovery problem is a special case of (1) where the observations $\mathbf{Q}$ and $\mathbf{Y}$ are the same.

II. QUANTIZED VARIATIONAL BAYESIAN LEARNING ALGORITHM

In order to develop an algorithm to estimate $\mathbf{X}$, we treat it as a latent variable with an appropriately chosen piecewise sparsity-promoting prior, and infer its posterior distribution from the quantized observations. Exact computation of the posterior distribution is computationally intractable due to the

The authors are with the Department of ECE, Indian Institute of Science, Bangalore, India 560 012 (Emails: thoota@iisc.ac.in; cmurthy@iisc.ac.in).

high dimensional integrals involved in obtaining the partition function. Therefore, we adopt an alternative approximate inference technique called variational Bayes (VB), where we replace the exact posterior with another easily computable probability distribution, to obtain an analytically and computationally tractable solution [16].

In [14], we developed a VB procedure to solve the sparse signal recovery problem using quantized measurements. However, that work does not account for the piecewise sparse structure of the signal, due to which, the resulting algorithm was computationally expensive due to an $N \times N$ matrix inversion operation in each iteration of the algorithm. Here, we show how to adapt the VB algorithm to explicitly account for the piecewise sparse structure, which leads to a significant lowering of the computational complexity (the matrix to be inverted is of size $L \times L$, and $L < N$), especially under quantized measurements.

We now present our VB-based quantized piecewise sparse signal recovery algorithm. Similar to the structure in (2), we divide the measurement matrix into K submatrices as $\mathbf{\Phi} \triangleq \begin{bmatrix} \mathbf{\Phi}_1 & \dots & \mathbf{\Phi}_K \end{bmatrix}$, $\mathbf{\Phi}_k \in \mathbb{C}^{M \times L}$, $k = 1, \dots, K$. We impose a two-stage hierarchical complex Gaussian prior on each column of $\mathbf{X}[k]$ with mean $\mathbf{0}_L$ and parameterized by a Gamma distributed diagonal precision matrix $\mathbf{P}[k] \triangleq \text{diag}(\boldsymbol{\alpha}[k]) \in \mathbb{R}_+^{L \times L}$, $\boldsymbol{\alpha}[k] = \begin{bmatrix} \alpha_1[k] & \dots & \alpha_L[k] \end{bmatrix}^T$. Such a two-stage prior is known to be sparsity-promoting. Specifically, the use of a hierarchical Gaussian prior and a common Gamma hyperprior on the precision matrix for each column of $\mathbf{X}$ results in a Student's t distributed marginalized prior, which promotes joint sparsity. The prior structure also reduces the dimension of the sparse signal to be estimated in each iteration of the algorithm, which helps in lowering the computational complexity. Mathematically,

$$p(\mathbf{x}_t[k]|\mathbf{P}[k]) = \frac{\det(\mathbf{P}[k])}{\pi^L} \exp\left(-\mathbf{x}_t^H[k]\mathbf{P}[k]\mathbf{x}_t[k]\right), \quad (3)$$

$$p(\boldsymbol{\alpha}[k]; a, r) = \prod_{\ell=1}^{L} \frac{r^a}{\Gamma(a)} \alpha_\ell^{a-1}[k] \exp\left(-r\alpha_\ell[k]\right), \quad (4)$$

where a and r are the shape and rate parameters, $\Gamma(a) \triangleq \int_0^\infty t^{a-1} \exp(-t) \, dt$ is the Gamma function evaluated at $a > 0$, $t = 1, \dots, T$ and $k = 1, \dots, K$. Note that the hyperparameters do not depend on the column index of $\mathbf{X}$, which ties the sparsity patterns of the columns together. We define $\mathbf{P} = \begin{bmatrix} \mathbf{P}[1] & \dots & \mathbf{P}[K] \end{bmatrix}$ and $\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}^T[1] & \dots & \boldsymbol{\alpha}^T[K] \end{bmatrix}^T$.

We express the logarithm of the joint probability distribution of $\mathbf{Q}, \mathbf{Y}, \mathbf{X}, \boldsymbol{\alpha}$ as

$$\ln p(\mathbf{Q}, \mathbf{Y}, \mathbf{X}, \boldsymbol{\alpha}; \mathbf{\Phi}, \sigma^2, a, r)$$
$$= \ln p(\mathbf{Q} \mid \mathbf{Y}) + \ln p(\mathbf{Y} \mid \mathbf{X}; \mathbf{\Phi}, \sigma^2)$$
$$+ \sum_{k=1}^{K} \sum_{t=1}^{T} \ln p(\mathbf{x}_t[k]|\mathbf{P}[k]) + \sum_{k=1}^{K} \ln p(\boldsymbol{\alpha}[k]; a, r), \quad (5)$$

where the prior distributions are given by (3) and (4). We set a and r to small values (say, 10^{-6}) which makes the hyperprior $p(\boldsymbol{\alpha}[k]; a, r)$, $\forall k$ non-informative. We impose a factorized structure on the posterior distribution

$p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\alpha} \mid \mathbf{Q}; \mathbf{\Phi}, \sigma^2, a, r)$ of the latent variables as:

$$p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\alpha} \mid \mathbf{Q}; \mathbf{\Phi}, \sigma^2, a, r)$$
$$\approx q_{\mathbf{Y}}(\mathbf{Y}) q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \prod_{k=1}^{K} q_{\mathbf{X}[k]}(\mathbf{X}[k])$$
$$= \prod_{t=1}^{T} q_{\mathbf{y}_t}(\mathbf{y}_t) \prod_{k=1}^{K} \prod_{\ell=1}^{L} q_{\alpha_\ell}(\alpha_\ell[k]) \prod_{t=1}^{T} \prod_{k=1}^{K} q_{\mathbf{x}_t[k]}(\mathbf{x}_t[k]), \quad (6)$$

where $\mathbf{Y} \triangleq [\mathbf{y}_1, \dots, \mathbf{y}_T]$. The conditional probability distributions of the observations and latent variables required to compute the posterior distributions are given by

$$p(\mathbf{Q} \mid \mathbf{Y}) = \prod_{t=1}^{T} \prod_{m=1}^{M} \mathbb{1}\left(\Re(y_{mt}) \in \left(\Re(y_{mt}^{(\text{lo})}), \Re(y_{mt}^{(\text{hi})})\right)\right)$$
$$\times \mathbb{1}\left(\Im(y_{mt}) \in \left(\Im(y_{mt}^{(\text{lo})}), \Im(y_{mt}^{(\text{hi})})\right)\right)$$
$$\triangleq \prod_{t=1}^{T} \mathbb{1}\left(\mathbf{y}_t \in \left(\mathbf{y}_t^{(\text{lo})}, \mathbf{y}_t^{(\text{hi})}\right)\right),$$

$$p(\mathbf{Y} \mid \mathbf{X}; \mathbf{\Phi}, \sigma^2) = \prod_{t=1}^{T} \frac{1}{(\pi\sigma^2)^M} \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y}_t - \mathbf{\Phi}\mathbf{x}_t\|^2\right),$$

where y_{mt} is the $(m, t)^{\text{th}}$ entry of $\mathbf{Y}$, $\mathbb{1}(\cdot)$ is the indicator function, $y_{mt}^{(\text{lo})}$ and $y_{mt}^{(\text{hi})}$ are the lower and upper quantization thresholds corresponding to the $(m, t)^{\text{th}}$ entry of $\mathbf{Q}$.

In the VB procedure, we compute the posterior distribution of latent variables by finding the expectations of the logarithm of the joint distribution (5) with respect to the posterior distributions of all the other latent variables. We present the posterior distribution $q_{\mathbf{X}[k]}(\mathbf{X}[k])$ in the following Lemma.

Lemma 1 (Computation of $q_{\mathbf{X}[k]}(\mathbf{X}[k])$, $k = 1, \dots, K$):
The posterior distribution $q_{\mathbf{x}_t[k]}(\mathbf{x}_t[k])$ is complex Gaussian with the covariance matrix and mean given by

$$\mathbf{\Sigma}_{\mathbf{X}[k]} = \left(\frac{1}{\sigma^2} \mathbf{\Phi}_k^H \mathbf{\Phi}_k + \langle \mathbf{P}[k] \rangle\right)^{-1}, \quad (7)$$

$$\langle \mathbf{X}[k] \rangle = \frac{1}{\sigma^2} \mathbf{\Sigma}_{\mathbf{X}[k]} \mathbf{\Phi}_k^H \left(\langle \mathbf{Y} \rangle - \sum_{\substack{k'=1 \\ k' \neq k}}^{K} \mathbf{\Phi}_{k'} \langle \mathbf{X}[k'] \rangle\right). \quad (8)$$

Here, $\langle \mathbf{P}[k] \rangle = \text{diag}(\langle \boldsymbol{\alpha}[k] \rangle)$, and $\langle \mathbf{Y} \rangle$ and $\langle \boldsymbol{\alpha}[k] \rangle$ are the posterior means of $q_{\mathbf{Y}}(\mathbf{Y})$ and $q_{\boldsymbol{\alpha}[k]}(\boldsymbol{\alpha}[k])$, respectively. ∎

The posterior covariance matrix $\mathbf{\Sigma}_{\mathbf{X}[k]}$ of $q_{\mathbf{X}[k]}(\mathbf{X}[k])$ contains a $L \times L$ matrix inverse for each k. Therefore, the computational complexity per iteration scales as $\mathcal{O}(KL^3)$ instead of $\mathcal{O}(K^3L^3)$ for the algorithm in [14]. This results in a significant reduction in the convergence time of the developed piecewise sparse signal recovery procedure.

We present the quantized VB piecewise (QVB-PW) sparse signal recovery procedure in Algorithm 1. Here, $\mathbf{Y}^{(\text{lo})}$ and $\mathbf{Y}^{(\text{hi})}$ are the lower and upper quantization levels corresponding to the observation $\mathbf{Q}$, respectively, and $\langle \mathbf{X} \rangle$ is the posterior mean of $q_{\mathbf{X}}(\mathbf{X})$. Also, $f(\cdot)$ and $F(\cdot)$ are the PDF and CDF of a standard Gaussian random variable, respectively, computed element-wise on the real and imaginary parts of the argument. The division operation to compute $\langle \mathbf{Y} \rangle$ is performed element-wise. $x_{\ell t}[k]$ is the $(\ell, t)^{\text{th}}$ element of $\mathbf{X}[k]$,

Algorithm 1 QVB-PW Sparse Signal Recovery

Input: $\mathbf{Q}$, $\mathbf{\Phi}$, σ, K, L
Output: $\langle\mathbf{X}\rangle$, $\langle\boldsymbol{\alpha}[1]\rangle, \ldots, \langle\boldsymbol{\alpha}[K]\rangle$
 1: Initialize $\langle\mathbf{Y}\rangle$, $\langle\boldsymbol{\alpha}[1]\rangle, \ldots, \langle\boldsymbol{\alpha}[K]\rangle$, a, r
 2: **repeat**
 3:   **for** $k = 1$ to K **do**
 4:     $\langle\mathbf{P}[k]\rangle = \text{diag}(\langle\boldsymbol{\alpha}[k]\rangle)$
 5:     Compute $\mathbf{\Sigma}_{\mathbf{X}[k]}$ using (7)
 6:     Compute $\langle\mathbf{X}[k]\rangle$ using (8)
 7:     **for** $\ell = 1$ to L **do**
 8:       Compute $\langle\alpha_\ell[k]\rangle = \frac{a+T}{r+\sum_{t=1}^{T}\langle|x_{\ell t}[k]|^2\rangle}$.
 9:     **end for**
10:   **end for**
11:   Compute $\langle\mathbf{Y}\rangle = \mathbf{\Phi}\langle\mathbf{X}\rangle$

$$+\frac{\sigma}{\sqrt{2}}\frac{f\left(\frac{\mathbf{Y}^{(\text{lo})}-\mathbf{\Phi}\langle\mathbf{X}\rangle}{\sigma/\sqrt{2}}\right) - f\left(\frac{\mathbf{Y}^{(\text{hi})}-\mathbf{\Phi}\langle\mathbf{X}\rangle}{\sigma/\sqrt{2}}\right)}{F\left(\frac{\mathbf{Y}^{(\text{hi})}-\mathbf{\Phi}\langle\mathbf{X}\rangle}{\sigma/\sqrt{2}}\right) - F\left(\frac{\mathbf{Y}_n^{(\text{lo})}-\mathbf{\Phi}\langle\mathbf{X}\rangle}{\sigma/\sqrt{2}}\right)}.$$

12: **until** stopping condition is met

---

and $\langle|x_{\ell t}[k]|^2\rangle = |\langle x_{\ell t}[k]\rangle|^2 + \mathbf{\Sigma}_{\mathbf{X}[k]}[\ell,\ell]$. The details involved in computing $q_{\mathbf{Y}}(\mathbf{Y})$, $q_{\alpha_\ell[k]}(\alpha_k)$, and their posterior statistics are presented in [14].

The VB algorithm proceeds iteratively by randomly initializing the posteriors and alternately computing each of the posterior distributions until a suitable convergence condition is satisfied. As VB falls into the category of minorization-maximization algorithms, it is known to converge to a stationary point of the original optimization problem from any initialization [17]. Once the algorithm converges, we use the posterior mean from (8) as our final estimate of $\mathbf{X}$.

III. APPLICATION TO MASSIVE MIMO-OFDM CHANNEL ESTIMATION

We evaluate the unquantized and quantized VB piecewise sparse signal recovery algorithms in the context of massive MIMO-OFDM sparse channel estimation [18]. We consider the uplink (UL) of a single cell massive MIMO-OFDM system with N_r receive antennas at the BS and K single transmit antenna user equipments (UEs), where $N_r \geq K$. Each UE transmits τ_p pilot OFDM symbols consisting of N_c subcarriers each. The unquantized complex baseband received signal at the n_r^{th} receive antenna in the nth symbol interval ($0 \leq n \leq N_c - 1$) within the tth pilot is

$$y_{n_r}[t][n] = \sum_{k=1}^{K}\sum_{\ell=0}^{L-1} h_{n_r,k}[\ell]\bar{s}_k[t][n-\ell] + w_{n_r}[t][n], \quad (9)$$

where $t \in \{1, \ldots, \tau_p\}$, $h_{n_r,k}[\ell]$ is the complex channel gain of the ℓth delay tap of the channel between the kth UE and the n_r^{th} receive antenna at the BS, distributed as $\mathcal{CN}(h_{n_r,k}[\ell]; 0, \beta_{k\ell})$, where $\beta_{k\ell}$ is the large scale fading coefficient (LSFC), L is the total number of delay taps of the frequency selective channel, $\bar{s}_k[t] = [\bar{s}_k[t][0], \ldots, \bar{s}_k[t][N_c - 1]]$ is the pilot symbol transmitted in the time domain by the kth UE in the tth OFDM symbol, and $w_{n_r}[t][n]$ is the complex additive white
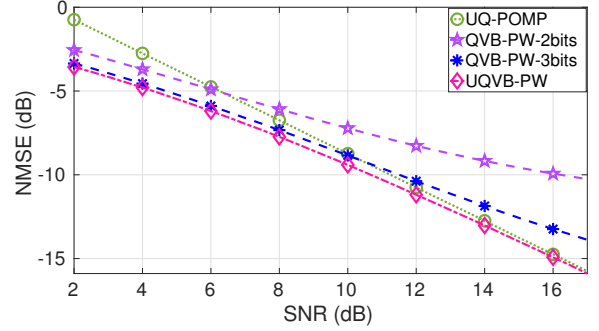

Figure 1. NMSE (dB) as a function of SNR (dB) for $K = 8$, $\tau_p = 1$.

Gaussian noise, with mean 0 and variance σ^2. We define the signal-to-noise-ratio (SNR) as $1/\sigma^2$.

We reformulate the received signal model above to utilize the lag-domain sparsity for channel estimation. We denote the channel sparsity, i.e., the maximum number of nonzero delay taps in the channel, by L_{sp}, where $L_{\text{sp}} \ll L$. After vectorizing and stacking the unquantized received pilot signal in (9) for the τ_p OFDM symbols and N_r receive antennas, and utilizing the diagonalization property of circulant matrices, we get

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1[1] & \cdots & \mathbf{y}_{N_r}[1] \\ \vdots & \ddots & \vdots \\ \mathbf{y}_1[\tau_p] & \cdots & \mathbf{y}_{N_r}[\tau_p] \end{bmatrix}$$
$$= \begin{bmatrix} (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H)\mathbf{S}[1](\mathbf{I}_K \otimes \mathbf{F}_{N_c,L}) \\ \vdots \\ (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H)\mathbf{S}[\tau_p](\mathbf{I}_K \otimes \mathbf{F}_{N_c,L}) \end{bmatrix}\mathbf{H} + \mathbf{W}$$
$$\triangleq \mathbf{\Phi}\mathbf{H} + \mathbf{W} \in \mathbb{C}^{\tau_p N_c \times N_r}, \quad (10)$$

where $\mathbf{y}_{n_r}[t] = \begin{bmatrix} y_{n_r}[t][1] & \cdots & y_{n_r}[t][N_c] \end{bmatrix}^T \in \mathbb{C}^{N_c \times 1}$, $\mathbf{H} \in \mathbb{C}^{KL \times N_r}$ is the piecewise joint-sparse signal containing the lag-domain sparse channels of the users as its submatrices, $\otimes$ denotes the matrix Kronecker product operator, $\mathbf{1}_K$ is the all-ones vector of size $K \times 1$, $\mathbf{F}_{N_c} \in \mathbb{C}^{N_c \times N_c}$ is the discrete Fourier transform (DFT) matrix, $\mathbf{F}_{N_c,L} \in \mathbb{C}^{N_c \times L}$ is the L column truncated DFT matrix, and $\mathbf{S}[t] \in \mathbb{C}^{KN_c \times KN_c}$ is a diagonal matrix containing the pilot symbols of the K users during the tth OFDM symbol.

We quantize the real and imaginary parts of each entry of (10) using b-bit ADCs. We set the dynamic range of the real and imaginary parts of the quantizer using the expected received signal power, P_R, as $\delta_0 = -2.5\sqrt{P_R/2}$, $\delta_B = 2.5\sqrt{P_R/2}$. The quantized received signal is

$$\mathbf{Q} = \mathcal{Q}_b(\mathbf{Y}) = \mathcal{Q}_b(\mathbf{\Phi}\mathbf{H} + \mathbf{W}) \in \mathbb{C}^{\tau_p N_c \times N_r}. \quad (11)$$

Now, we apply the quantized VB piecewise sparse recovery Algorithm 1, to estimate $\mathbf{H}$, given $\mathbf{Q}$ and $\mathbf{\Phi}$. Note that, the number of measurements $M = \tau_p N_c$, the dimension of the sparse vector $N = KL$, and the number of measurement vectors $T = N_r$. Also, if $\tau_p N_c < KL$, (11) represents an underdetermined system of equations.

We evaluate the normalized mean-squared error (NMSE) of the unquantized and quantized VB piecewise sparse recovery algorithms in Figures 1 and 2. We set $N_r = 64$,
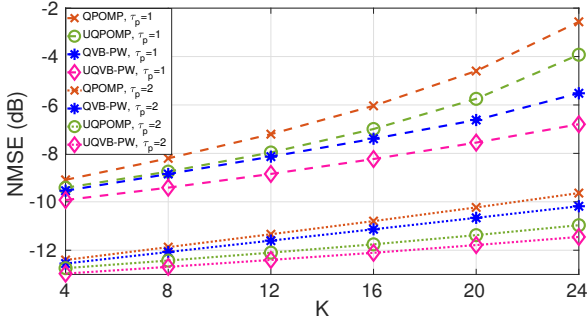
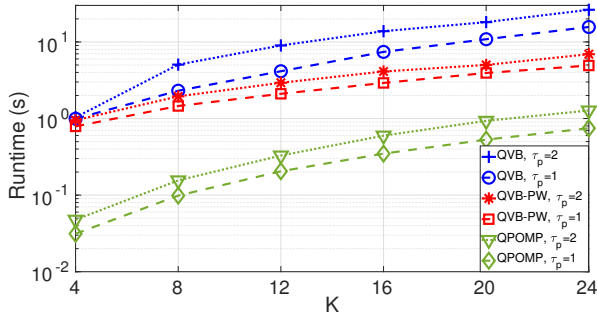Figure 2. NMSE (dB) as a function of K. SNR set to 10 dB.

Figure 3. Runtime (dB) as a function of K. SNR set to 10 dB.

$N_c = 256$, $L = 64$, $L_{sp} = 8$ for the simulations. We fix the large scale fading coefficients of all the users to 1 for convenience. We use the normalized squared ℓ_2 distance between the estimates of two successive iterations of the VB algorithm as the stopping condition, and set it to 10^{-6}. We also fix the maximum number of iterations of VB to 100. We compare the performance against the state-of-the-art piecewise orthogonal matching pursuit (POMP) algorithm, which recovers the sparse signal using unquantized compressive measurements [3]. We also benchmark the runtime performance against QVB algorithm in [14]. The NMSE performance of QVB algorithm in [14] overlaps with that of the QVB-PW procedure presented in this paper. Therefore, we do not include it to avoid clutter. The run-time plot empirically shows that the developed algorithm is not only high-performing, but also is computationally less complex.

Fig. 1 shows the NMSE (dB) of VB and POMP algorithms as a function of SNR (dB) for $K = 8$ and $\tau_p = 1$. We observe that the VB algorithm with the unquantized measurements outperforms the POMP algorithm by around 2 dB at an NMSE of around -7 dB. Moreover, the VB algorithm with only 2 and 3 bits quantization performs better than the POMP with unquantized measurements when the SNR is below 6 dB and 10 dB, respectively. At high SNRs, both the VB and POMP with unquantized measurements perform the same. At high SNR, POMP recovers the support accurately, and its MSE is Bayes' optimal conditioned on correct support recovery. Thus, empirically, we see that the VB algorithm with unquantized measurements is also Bayes' optimal at high SNRs.

Fig. 2 shows the NMSE (dB) of VB and POMP algorithms as a function of the number of users for $\tau_p = 1$ and 2,

and with 3 bits quantization. For the quantized case, we feed $\mathbf{Q}$ to the POMP algorithm. As the number of measurements increase ($\tau_p = 1$ to $\tau_p = 2$), the performance of both VB and POMP algorithms become better. However, in a measurement constrained regime ($\tau_p = 1$ case), the VB algorithm with just 3 bits quantized measurements outperforms even the POMP algorithm with unquantized measurements. Moreover, as K increases (which translates to increase in the dimension of the piecewise sparse vectors), the performance gap between the VB and POMP algorithms increases. This shows that, not only does the VB algorithm perform well, its performance also scales better than the POMP algorithm.

Fig. 3 shows the runtime (in seconds) of QVB, QVB-PW, and POMP algorithms as a function of the number of users for $\tau_p = 1$ and 2, and with 3 bits quantization. As the number of measurements increase ($\tau_p = 1$ to $\tau_p = 2$), the runtime of all the algorithms increase which is due to the increase in the number of mathematical operations to recover $\mathbf{X}$. The runtime of QVB-PW is lower than that of QVB even though both have the same NMSE performance. This illustrates that QVB-PW is computationally less complex than QVB due to the reduction in the size of the matrix inverse to be computed in the QVB-PW algorithm. Moreover, QVB-PW with $\tau_p = 2$ converges faster than QVB with $\tau_p = 1$. This shows that, when we utilize the inherent piecewise sparse structure, the resulting algorithm is both fast and high-performing.

The piecewise VB sparse recovery algorithm has a matrix inverse in each iteration whose complexity scales cubically with the number of channel taps L. The POMP algorithm also has a matrix inverse operation to recover one tap of each user's channel. However, its complexity increases progressively in each iteration. The maximum complexity of the POMP algorithm scales cubically with the sparsity of each user's channel. Therefore, at high SNRs where the POMP and VB perform equally well for the unquantized case, it is better to choose POMP than piecewise VB. However, when the measurements are quantized with large number of users and at low to medium SNRs, it is sensible to choose VB for its better NMSE performance than POMP.

IV. CONCLUSIONS

We presented VB based algorithms for recovering piecewise joint-sparse signals using unquantized and quantized compressive measurements. We imposed a factorized structure on the posterior distributions of the latent variables, which resulted in an iterative, computationally and analytically tractable solution. We benchmarked the NMSE and runtimes of the piecewise sparse VB algorithms with the state-of-the-art POMP and VB algorithms for piecewise sparse recovery, in the context of MIMO-OFDM channel estimation. We showed that the VB algorithms outperform the POMP algorithm in measurement constrained regimes as well as at low and medium SNRs. We also numerically showed that utilizing the inherent piecewise sparse structure assisted in reducing the computational complexity of the developed algorithm. We also empirically showed that it is Bayes' optimal in the MSE sense.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] K. Li, C. R. Rojas, T. Yang, H. Hjalmarsson, K. H. Johansson, and S. Cong, "Piecewise sparse signal recovery via piecewise orthogonal matching pursuit," in *Proc. ICASSP*, 2016, pp. 4608–4612.

[4] J. Zhong and C. Li, "Piecewise sparse recovery via piecewise greedy method," *J. Math. Res. with Appl.*, vol. 6, pp. 643–658, 2018.

[5] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[6] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[7] J. Tropp, A. Gilbert, and M. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. ICASSP*, vol. 5, 2005, pp. v/721–v/724 Vol. 5.

[8] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[9] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep. 2011.

[10] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.

[11] D. Prasanna and C. R. Murthy, "mmWave channel estimation via compressive covariance estimation: Role of sparsity and intra-vector correlation," *IEEE Trans. Signal Process.*, vol. 69, pp. 2356–2370, 2021.

[12] Y. Ding, S.-E. Chiu, and B. D. Rao, "Sparse recovery with quantized multiple measurement vectors," in *2017 51st Asilomar Conference on Signals, Systems, and Computers.* IEEE, 2017, pp. 845–849.

[13] ——, "Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog-digital processing and low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 499–513, Jun. 2018.

[14] S. S. Thoota and C. R. Murthy, "Massive MIMO-OFDM systems with low resolution ADCs: Cramér-Rao bound, sparse channel estimation, and soft symbol decoding," *IEEE Trans. Signal Process.*, 2022, Early Access.

[15] B. Murmann, *ADC Performance Survey 1997–2018*. [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html

[16] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer New York, 2006.

[17] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, 2004.

[18] A. Maltsev *et al.*, "Channel models for 60 GHz WLAN systems," IEEE, Piscataway, New Jersey, Tech. Rep. 802.11-09/0334r8, 2010.