A Variational Bayesian Perspective on MIMO Detection with Low-Resolution ADCs

Ly V. Nguyen*, A. Lee Swindlehurst†, and Duy H. N. Nguyen*,‡
*Computational Science Research Center, San Diego State University, CA, USA
†Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, USA
‡Department of Electrical and Computer Engineering, San Diego State University, CA, USA
Email: vnguyen6@sdsu.edu, swindle@uci.edu, duy.nguyen@sdsu.edu

2022 56th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5906-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEECONF56349.2022.10052059

***Abstract*—This paper proposes data detection methods for massive multiple-input multiple-output (MIMO) systems with low-resolution analog-to-digital converters (ADCs) based on the variational Bayes (VB) inference framework. We derive matched-filter quantized VB (MF-QVB) and linear minimum mean-squared error quantized VB (LMMSE-QVB) detection methods assuming the channel state information (CSI) is available. Unlike conventional VB-based detection methods that assume knowledge of the second-order statistics of the additive noise, we propose to float the noise variance/covariance matrix as an unknown random variable that is used to account for both the noise and the residual inter-user interference. Finally, we show via numerical results that the proposed VB-based methods provide robust performance and also significantly outperform existing methods.**

I. Introduction

Implementing massive antenna arrays, or massive multiple-input multiple-output (MIMO), for 5G-and-beyond wireless systems has been adopted and considered as one of the core technologies for providing high spectral and energy efficiency, especially in high-frequency bands like mmWave and THz where the radio path loss is severe. However, scaling up existing radio frequency (RF) technologies to very large arrays becomes complex, expensive, and demands high power consumption. A practical and promising solution is to use low-resolution analog-to-digital converters (ADCs). This is because the low-resolution ADCs have very simple structures and low power consumption. Unfortunately, the use of low-resolution quantization makes the system severely non-linear and significantly distorts the received signals, and thus requires special signal processing methods for data detection.

There has been a plethora of data detection studies for massive MIMO systems with low-resolution ADCs. For example, one-bit ML and near-ML methods were proposed in [1]. The Bussgang decomposition was used to derive different linear data detectors in [2]–[4]. While the ML and near-ML methods are either too complicated for practical implementation or non-robust at high signal-to-noise ratios (SNRs), the linear Bussgang-based receivers have lower complexity and are more robust, but they have limited performance.

Recently, machine learning for low-resolution MIMO data detection has gained interest and there has also been numerous results reported in the literature. In particular, the work in [5] shows how support vector machine (SVM) models can be applied to one-bit massive MIMO channel estimation and data detection. The authors of [3], [6] exploit a deep neural network (DNN) framework to develop a special model-driven detection approach that outperforms the SVM-based methods in [5]. The work in [7] proposed another DNN-based detector but its computational complexity is high since the detection network must be retrained for each new channel realization. Several learning-based blind detection methods were proposed in [8]–[10] but they are restricted to small-scale systems. The authors in [11] developed a bilinear generalized approximate message passing (BiGAMP) algorithm [12] to solve the joint channel estimation and data detection (JED) problem for few-bit MIMO systems. Another JED method was proposed in [13] based on the variational Bayesian (VB) inference framework, and it was shown to outperforms the BiGAMP-based method in [11] for soft symbol decoding. In a recent work [14], VB inference was also shown to be very efficient in MIMO data detection with infinite-resolution (perfect) ADCs.

In this paper, we develop a VB framework for data detection for massive MIMO systems with low-resolution ADCs. While conventional machine learning models such as SVM and DNN only provide a point estimate of the signal of interest, e.g., the data symbols, the VB approach can provide the posterior distribution of the estimate, which is important in subsequent signal processing steps such as channel decoding. Another advantage of VB is that it does not require a training process like DNNs which often suffer from performance degradation due to mismatch between the actual model and that used during training. The contributions of this paper are summarized as follows:

- We first devise a matched-filter quantized VB (MF-QVB) detection method for few-bit MIMO systems. Unlike the VB-based detection method in [13] that assumes a known noise variance, the proposed MF-QVB method floats the noise variance as a latent variable and uses it to also account for residual inter-user interference. This latent variable is jointly estimated with the transmitted data symbol vector.
- We then develop a linear minimum mean-squared error quantized VB (LMMSE-QVB) detector that treats the noise covariance matrix as a latent variable, rather than simply assuming the noise covariance is a scaled identity matrix. The LMMSE-QVB detector offers performance

similar to MF-QVB for independent and identically distributed (i.i.d.) channels, but significantly outperforms MF-QVB for spatially correlated channels. We show via numerical results that the proposed VB detection algorithms provide much lower symbol error rates (SERs) compared to the conventional VB-based methods in [13].

II. System Model

We consider an uplink massive MIMO system with K single-antenna users and an M-antenna base station (BS). Let $\mathbf{x} \in \mathbb{C}^{K\times 1}$ and $\mathbf{H} \in \mathbb{C}^{M\times K}$ denote the transmitted signal vector and the channel matrix, respectively, then the linear uplink MIMO system can be modeled as $\mathbf{r} = \mathbf{H}\mathbf{x}+\mathbf{n}$, where $\mathbf{r}$ is the unquantized received signal vector and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}_M)$ models the i.i.d. additive white Gaussian noise. The channel vector $\mathbf{h}_i$ from user-i to the BS is assumed to be distributed as $p(\mathbf{h}_i) = \mathcal{CN}(\mathbf{h}_i; \mathbf{0}, \mathbf{C}_i)$ where $\mathbf{C}_i \triangleq \mathbb{E}[\mathbf{h}_i\mathbf{h}_i^H]$ is the covariance matrix. Finally, we assume $\mathbb{E}[\mathbf{h}_i\mathbf{h}_j^H] = \mathbf{0}$ if $i \neq j$.

Each received analog signal is then quantized by a pair of b-bit ADCs to produce the quantized received signals $\Re\{\mathbf{y}\} = \mathcal{Q}_b(\Re\{\mathbf{r}\})$ and $\Im\{\mathbf{y}\} = \mathcal{Q}_b(\Im\{\mathbf{r}\})$, where $\mathcal{Q}_b(\cdot)$ denotes the b-bit ADC operation which is applied separately to every element of its matrix or vector argument. It is assumed that $\mathcal{Q}_b(\cdot)$ performs b-bit uniform scalar quantization, which is characterized by a set of $2^b - 1$ thresholds denoted as $\{d_1, \ldots, d_{2^b-1}\}$. For a quantization step size Δ, the quantization thresholds are given by $d_k = (-2^{b-1} + k)\Delta$, for $k \in \mathcal{K} = \{1, \ldots, 2^b-1\}$. The quantized output q is then defined as

$$q = \mathcal{Q}_b(r) = \begin{cases} d_k - \frac{\Delta}{2}, & \text{if } r \in (d_{k-1}, d_k] \text{ with } k \in \mathcal{K} \\ (2^b - 1)\frac{\Delta}{2}, & \text{if } r \in (d_{2^b-1}, d_{2^b}]. \end{cases}$$

We also define $q^{\text{low}} = d_{k-1}$ and $q^{\text{up}} = d_k$ as lower and upper thresholds of the quantization bin to which q belongs.

III. VB for Approximate Inference

This section presents a brief background on the VB method for approximate inference that will be developed for solving the problems of interest in this paper. We are interested in finding a posterior distribution $p(\mathbf{x}|\mathbf{y})$, where $\mathbf{x}$ is comprised of the latent variables and parameters, and $\mathbf{y}$ is the observed variables. Since obtaining a closed-form expression of $p(\mathbf{x}|\mathbf{y})$ is often intractable due to numerous reasons, e.g., high dimensionality or model sophistication, it is crucial to develop numerical methods that can efficiently approximate $p(\mathbf{x}|\mathbf{y})$. The VB inference method amounts to finding a density function $q(\mathbf{x})$ that is as close as possible to $p(\mathbf{x}|\mathbf{y})$ by minimizing their Kullback-Leibler (KL) divergence as follows [15]:

$$\begin{aligned} q(\mathbf{x}) &= \arg\min_{q(\mathbf{x})\in\mathcal{Q}} \mathrm{KL}\big(q(\mathbf{x})\|p(\mathbf{x}|\mathbf{y})\big) \\ &= \arg\min_{q(\mathbf{x})\in\mathcal{Q}} \mathbb{E}_{q(\mathbf{x})}[\ln q(\mathbf{x})] - \mathbb{E}_{q(\mathbf{x})}[\ln p(\mathbf{x},\mathbf{y})] + \ln p(\mathbf{y}), \end{aligned}$$

where $\mathcal{Q}$ is the mean-field variational family of $q(\mathbf{x})$ satisfying $q(\mathbf{x}) = \prod_{i=1}^{K} q_k(x_i)$. A general expression for the optimal solution for $q_i(x_i)$ can be obtained as [15]

$$q_i(x_i) \propto \exp\{\langle \ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x})\rangle_{-x_i}\} \quad (1)$$

Here, $\langle\cdot\rangle_{-x_i}$ denotes the expectation with respect to all latent variables except x_i. In the following, if $\langle\cdot\rangle$ is used, it means the variational expectation is taken w.r.t. all the latent variables in the argument. By sequentially updating $q_i(x_i)$ until convergence, an estimate of $p(\mathbf{x}|\mathbf{y})$ can be obtained as $q(\mathbf{x})$.

In the following, we present a theorem on the variational posterior mean of multiple random variables that will be applied repeatedly later in the paper.

Theorem 1. *Let $\mathbf{A}$, $\mathbf{y}$, and $\mathbf{x}$ of size $m \times n$, $m \times 1$, and $n \times 1$ be three independent random matrices (vectors) w.r.t. a variational distribution $q_{\mathbf{A},\mathbf{y},\mathbf{x}}(\mathbf{A},\mathbf{y},\mathbf{x}) = q(\mathbf{A})q(\mathbf{y})q(\mathbf{x})$. It is assumed that $\mathbf{A}$ is column-wise independent and let $\langle\mathbf{a}_i\rangle$ and $\mathbf{\Sigma}_{\mathbf{a}_i}$ be the variational mean and covariance matrix of the ith column of $\mathbf{A}$. Let $\langle\mathbf{x}\rangle$ and $\mathbf{\Sigma}_{\mathbf{x}}$ (and $\langle\mathbf{y}\rangle$ and $\mathbf{\Sigma}_{\mathbf{y}}$) be the variational mean and covariance matrix of $\mathbf{x}$ (and $\mathbf{y}$), respectively. For an arbitrary Hermitian matrix $\mathbf{B}$, let $\big\langle(\mathbf{y} - \mathbf{A}\mathbf{x})^H\mathbf{B}(\mathbf{y}-\mathbf{A}\mathbf{x})\big\rangle$ be the expectation of $(\mathbf{y}-\mathbf{A}\mathbf{x})^H\mathbf{B}(\mathbf{y}-\mathbf{A}\mathbf{x})$ w.r.t. $q_{\mathbf{A},\mathbf{y},\mathbf{x}}(\mathbf{A},\mathbf{y},\mathbf{x})$. We have*

$$\begin{aligned} &\big\langle(\mathbf{y} - \mathbf{A}\mathbf{x})^H\mathbf{B}(\mathbf{y} - \mathbf{A}\mathbf{x})\big\rangle \\ &= \big(\langle\mathbf{y}\rangle - \langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\big)^H\mathbf{B}\big(\langle\mathbf{y}\rangle - \langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\big) + \mathrm{Tr}\{\mathbf{B}\mathbf{\Sigma}_{\mathbf{y}}\} \\ &\quad + \langle\mathbf{x}\rangle^H\mathbf{D}\langle\mathbf{x}\rangle + \mathrm{Tr}\big\{\mathbf{\Sigma}_{\mathbf{x}}\mathbf{D}\big\} + \mathrm{Tr}\big\{\mathbf{\Sigma}_{\mathbf{x}}\langle\mathbf{A}^H\rangle\mathbf{B}\langle\mathbf{A}\rangle\big\}, \quad (2) \end{aligned}$$

where $\mathbf{D} = \mathrm{diag}\big(\mathrm{Tr}\{\mathbf{B}\mathbf{\Sigma}_{\mathbf{a}_1}\}, \ldots, \mathrm{Tr}\{\mathbf{B}\mathbf{\Sigma}_{\mathbf{a}_n}\}\big)$.

Proof: The proof of this theorem is similar to the proof of Theorem 1 in [14], except that $\mathbf{y}$ is now a random vector. Details of the proof are given in Appendix A. ■

We note that if any of $\mathbf{A}$, $\mathbf{y}$, and $\mathbf{x}$ is deterministic, the corresponding covariance matrices $\{\mathbf{\Sigma}_{\mathbf{a}_i}\}$, $\mathbf{\Sigma}_{\mathbf{y}}$, and $\mathbf{\Sigma}_{\mathbf{x}}$ will be set to $\mathbf{0}$ and the expectation of $(\mathbf{y} - \mathbf{A}\mathbf{x})^H\mathbf{B}(\mathbf{y} - \mathbf{A}\mathbf{x})$ given in (2) can be simplified accordingly.

IV. Proposed VB for Data Detection in Low-Resolution MIMO Systems

In this section, we develop new VB-based algorithms for solving the data detection problem in low-resolution MIMO systems with known channel $\mathbf{H}$.

A. Proposed MF-QVB

The VB-based methods proposed in [13] assume prior information about the noise variance N_0. However, in practice, N_0 is not known *a priori* and may need to be estimated. Furthermore, using the known noise variance, the conventional VB methods in [13] do not take into account the residual inter-user interference. Here, we consider the residual interference-plus-noise as an unknown parameter N_0^{post}, which is postulated by the estimation in the VB framework [14]. For ease of computation, we use $\gamma = 1/N_0^{\text{post}}$ to denote the precision to be estimated.

The joint distribution $p(\mathbf{y},\mathbf{r},\mathbf{x};\gamma,\mathbf{H})$ of the observed variable $\mathbf{y}$ and the latent variables $\mathbf{r}$ and $\mathbf{x}$ can be factored as

$$\begin{aligned} p(\mathbf{y},\mathbf{r},\mathbf{x};\gamma,\mathbf{H}) &= p(\mathbf{y}|\mathbf{r})p(\mathbf{r}|\mathbf{x};\gamma,\mathbf{H})p(\mathbf{x}) \\ &= \left[\prod_{m=1}^{M} p(y_m|r_m)\right] p(\mathbf{r}|\mathbf{x};\gamma,\mathbf{H}) \prod_{i=1}^{K} p(x_i), \end{aligned} \quad (3)$$

where $p(y_m|r_m) = \mathbb{1}\big(r_m \in [y_m^{\text{low}}, y_m^{\text{up}}]\big)$ and $p(\mathbf{r}|\mathbf{x};\gamma,\mathbf{H}) = \mathcal{CN}(\mathbf{r};\mathbf{Hx},\gamma^{-1}\mathbf{I}_M)$. Here, $\mathbb{1}(\cdot)$ denotes the indicator function which equals one if the argument holds true, or zero otherwise. We note that the random vector $\mathbf{r}$ is comprised of conditional independent elements due to the same noise variance being imposed on the M receive antennas.

In the ***E-step***, for a currently fixed estimate $\hat{\gamma}$ of γ, we aim to derive the mean field variational distribution $q(\mathbf{r},\mathbf{x})$ of $\mathbf{r}$ and $\mathbf{x}$ given $\mathbf{y}$ such that

$$p(\mathbf{r},\mathbf{x}|\mathbf{y};\hat{\gamma},\mathbf{H}) \approx q(\mathbf{r},\mathbf{x}) = q(\mathbf{r})\left[\prod_{i=1}^{K} q(x_i)\right]. \quad (4)$$

1) Updating $\mathbf{r}$. The variational distribution $q(\mathbf{r})$ is obtained by taking the expectation of the conditional in (3) w.r.t. $q(\mathbf{x})$:

$$\begin{aligned} q(\mathbf{r}) &\propto \exp\Big\{\big\langle \ln p(\mathbf{y}|\mathbf{r}) + \ln p(\mathbf{r}|\mathbf{x};\hat{\gamma},\mathbf{H})\big\rangle_{-\mathbf{r}}\Big\} \\ &\propto \exp\Big\{\big\langle \ln \mathbb{1}\big(\mathbf{r}\in[\mathbf{y}^{\text{low}},\mathbf{y}^{\text{up}}]\big) - \hat{\gamma}\|\mathbf{r}-\mathbf{Hx}\|^2\big\rangle_{-\mathbf{r}}\Big\} \\ &\propto \mathbb{1}\big(\mathbf{r}\in[\mathbf{y}^{\text{low}},\mathbf{y}^{\text{up}}]\big)\times\mathcal{CN}\big(\mathbf{r};\mathbf{H}\langle\mathbf{x}\rangle,\hat{\gamma}^{-1}\mathbf{I}_M\big). \end{aligned} \quad (5)$$

We note that variational distribution $q(\mathbf{r})$ is inherently separable as $\prod_{m=1}^{M} q(r_m)$ without enforcing the mean field approximation on $q(\mathbf{r})$. Thus, the variational mean and variance can be obtained concurrently for all the elements of $\mathbf{r}$. We see in (5) that $q(r_m)$ is the truncated complex normal distribution obtained from bounding $r_m \sim \mathcal{CN}\big(s_m,\hat{\gamma}^{-1}\big)$, where $s_m = \mathbf{H}_{m,:}\langle\mathbf{x}\rangle$, to the interval $(y_m^{\text{low}}, y_m^{\text{up}})$.

2) Updating x_i. The variational distribution $q(x_i)$ is obtained by taking the expectation of the conditional in (3) w.r.t. $q(\mathbf{r})\prod_{j\neq i} q(x_j)$:

$$\begin{aligned} q(x_i) &\propto \exp\big\{\big\langle \ln p(\mathbf{r}|\mathbf{x};\hat{\gamma},\mathbf{H}) + \ln p(x_i)\big\rangle_{-x_i}\big\} \\ &\propto p(x_i)\exp\big\{-\hat{\gamma}\big\langle\|\mathbf{r}-\mathbf{Hx}\|^2\big\rangle_{-x_i}\big\} \\ &\propto p(x_i)\ \exp\big\{-\hat{\gamma}\|\mathbf{h}_i\|^2\big(|x_i|^2 - 2\,\Re\{x_i^* z_i\}\big)\big\} \\ &\propto p(x_i)\ \exp\big\{-\hat{\gamma}\|\mathbf{h}_i\|^2|x_i-z_i|^2\big\} \\ &\propto p(x_i)\,\mathcal{CN}\big(z_i;x_i,1/(\hat{\gamma}\|\mathbf{h}\|^2)\big), \end{aligned} \quad (6)$$

where we define

$$z_i = \frac{\mathbf{h}_i^H}{\|\mathbf{h}_i\|^2}\Big(\langle\mathbf{r}\rangle - \sum_{j\neq i}^{K}\mathbf{h}_j\langle x_j\rangle\Big) = \langle x_i\rangle + \frac{\mathbf{h}_i^H}{\|\mathbf{h}_i\|^2}\big(\langle\mathbf{r}\rangle - \mathbf{H}\langle\mathbf{x}\rangle\big) \quad (7)$$

with $\langle x_i\rangle$ being the *currently fixed* nonlinear estimate of $x_i, \forall i$. We can see in (6) that the mean field VB approximation decouples the few-bit MIMO system into an AWGN channel $z_i = x_i + \mathcal{CN}\big(0,1/(\hat{\gamma}\|\mathbf{h}\|^2)\big)$ for user-i. The variational distribution $q(x_i)$ can be realized by normalizing $p(x_i)\,\mathcal{CN}\big(z_i;x_i,1/(\hat{\gamma}\|\mathbf{h}\|^2)\big)$.

In the ***M-step***, the estimate of γ is updated to maximize $\ln p(\mathbf{y},\mathbf{r},\mathbf{x};\gamma,\mathbf{H})$ w.r.t. $q(\mathbf{r},\mathbf{x})$, i.e.,

$$\begin{aligned} \hat{\gamma} &= \arg\max_{\gamma}\ \big\langle \ln p(\mathbf{r}|\mathbf{x};\gamma;\mathbf{H})\big\rangle \\ &= \arg\max_{\gamma}\ M\ln\gamma - \gamma\big\langle\|\mathbf{r}-\mathbf{Hx}\|^2\big\rangle. \end{aligned} \quad (8)$$

Applying Theorem 1 to evaluate the expectation $\big\langle\|\mathbf{r}-\mathbf{Hx}\|^2\big\rangle$, the new estimate of γ is given by

$$\hat{\gamma} = \frac{M}{\|\langle\mathbf{r}\rangle - \mathbf{H}\langle\mathbf{x}\rangle\|^2 + \sum_{m=1}^{M}\tau_{r_m} + \sum_{i=1}^{K}\tau_{x_i}\|\mathbf{h}_i\|^2}. \quad (9)$$

By iteratively optimizing $q(\mathbf{r})$, $\{q(x_i)\}$, and updating $\hat{\gamma}$, we obtain the variational Bayes expectation-maximization (VBEM) algorithm for estimating $\mathbf{r}$, $\mathbf{x}$, and γ. Similar to our previous work [14], we refer this scheme to as the ***MF-QVB algorithm*** due to the use of the matched-filter $\mathbf{h}_i^H/\|\mathbf{h}_i\|^2$ to obtain the linear estimate z_i of x_i in (7). If γ is fixed to N_0^{-1}, the ***MF-QVB algorithm*** will be referred to as the ***conv-QVB algorithm***, that was investigated as the QVB-CSIR algorithm in [13].

B. Proposed LMMSE-QVB

We now develop the LMMSE-QVB method for low-resolution MIMO detection that uses a postulated noise covariance matrix $\mathbf{C}^{\text{post}}$ instead of the postulated noise variance N_0^{post} in the MF-QVB method. The idea of using a postulated noise covariance matrix $\mathbf{C}^{\text{post}}$ was proposed in [14] but for infinite-resolution ADCs. For ease of computation, we use $\mathbf{\Gamma} = (\mathbf{C}^{\text{post}})^{-1}$ as the precision matrix to be estimated.

The joint distribution $p(\mathbf{y},\mathbf{r},\mathbf{x};\mathbf{\Gamma},\mathbf{H})$ of the observed variable $\mathbf{y}$ and the latent variables $\mathbf{r}$ and $\mathbf{x}$ at time slot t can be factored as

$$\begin{aligned} p(\mathbf{y},\mathbf{r},\mathbf{x};\mathbf{\Gamma},\mathbf{H}) &= p(\mathbf{y}|\mathbf{r})p(\mathbf{r}|\mathbf{x};\mathbf{\Gamma},\mathbf{H})p(\mathbf{x}) \\ &= \left[\prod_{m=1}^{M} p(y_m|r_m)\right] p(\mathbf{r}|\mathbf{x};\mathbf{\Gamma},\mathbf{H})\prod_{i=1}^{K} p(x_i), \end{aligned} \quad (10)$$

where $p(\mathbf{r}|\mathbf{x},\mathbf{\Gamma};\mathbf{H}) = \mathcal{CN}\big(\mathbf{r};\mathbf{Hx},\mathbf{\Gamma}^{-1}\big)$. We note that the random vector $\mathbf{r}$ is no longer comprised of conditional independent elements, since the noise covariance matrix $\mathbf{\Gamma}^{-1}$ is in general non-diagonal.

In the ***E-step***, for a currently fixed estimate $\hat{\mathbf{\Gamma}}$ of $\mathbf{\Gamma}$, we aim to derive the mean field variational distribution $q(\mathbf{r},\mathbf{x})$ of $\mathbf{r}$ and $\mathbf{x}$ given $\mathbf{y}$ such that

$$p(\mathbf{r},\mathbf{x}|\mathbf{y};\hat{\mathbf{\Gamma}},\mathbf{H}) \approx q(\mathbf{r},\mathbf{x}) = \left[\prod_{m=1}^{M} q(r_m)\right]\left[\prod_{i=1}^{K} q(x_i)\right]. \quad (11)$$

1) Updating r_m. The variational distribution $q(r_m)$ is obtained by taking the expectation of the conditional in (10) w.r.t. $q(\mathbf{x})\prod_{n\neq m} q(r_n)$:

$$\begin{aligned} q(r_m) &\propto \exp\Big\{\big\langle \ln p(y_m|r_m) + \ln p(\mathbf{r}|\mathbf{x};\hat{\mathbf{\Gamma}},\mathbf{H})\big\rangle_{-r_m}\Big\} \\ &\propto \exp\Big\{\ln \mathbb{1}\big(r_m\in[y_m^{\text{low}},y_m^{\text{up}}]\big) \\ &\qquad - \big\langle(\mathbf{r}-\mathbf{Hx})^H\hat{\mathbf{\Gamma}}(\mathbf{r}-\mathbf{Hx})\big\rangle_{-r_m}\Big\} \\ &\propto \mathbb{1}\big(r_m\in[y_m^{\text{low}},y_m^{\text{up}}]\big)\times\exp\big\{-\hat{\gamma}_{mm}|r_m-s_m|^2\big\} \\ &\propto \mathbb{1}\big(r_m\in[y_m^{\text{low}},y_m^{\text{up}}]\big)\times\mathcal{CN}\big(r_m;s_m,\hat{\gamma}_{mm}^{-1}\big), \end{aligned} \quad (12)$$

where s_m is now defined as

$$s_m = \mathbf{H}_{m,:}\langle \mathbf{x}\rangle - \hat{\gamma}_{mm}^{-1} \sum_{n\neq m}^{M} \hat{\gamma}_{mn}\big(\langle r_n\rangle - \mathbf{H}_{n,:}\langle \mathbf{x}\rangle\big)$$
$$= \langle r_m\rangle - \frac{\hat{\mathbf{\Gamma}}_m^H}{\hat{\gamma}_{mm}}\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big), \tag{13}$$

$\langle r_m\rangle$ is the *currently fixed* nonlinear estimate of r_m and $\hat{\mathbf{\Gamma}}_m$ is the mth column of the Hermitian matrix $\hat{\mathbf{\Gamma}}$. We can see in (12) that the variational distribution $q(r_m)$ is the truncated complex normal distribution obtained from bounding $r_m \sim \mathcal{CN}(s_m, \hat{\gamma}_{mm}^{-1})$ to the interval $(y_m^{\text{low}}, y_m^{\text{up}})$.

2) Updating x_i. The variational distribution $q(x_i)$ is obtained by taking the expectation of the conditional in (3) w.r.t. $q(\mathbf{r})\prod_{j\neq i} q(x_j)$:

$$q(x_i) \propto \exp\Big\{\big\langle \ln p(\mathbf{r}|\mathbf{x};\hat{\mathbf{\Gamma}},\mathbf{H}) + \ln p(x_i)\big\rangle_{-x_i}\Big\}$$
$$\propto p(x_i)\exp\Big\{-\big\langle(\mathbf{r}-\mathbf{H}\mathbf{x})^H\hat{\mathbf{\Gamma}}(\mathbf{r}-\mathbf{H}\mathbf{x})\big\rangle_{-x_i}\Big\}$$
$$\propto p(x_i)\exp\big\{-\mathbf{h}_i^H\hat{\mathbf{\Gamma}}\mathbf{h}_i|x_i - z_i|^2\big\}$$
$$\propto p(x_i)\,\mathcal{CN}\big(z_i; x_i, 1/(\mathbf{h}_i^H\hat{\mathbf{\Gamma}}\mathbf{h}_i)\big), \tag{14}$$

where z_i is a linear estimate of x_i that is now defined as

$$z_i = \frac{\mathbf{h}_i^H\hat{\mathbf{\Gamma}}}{\mathbf{h}_i^H\hat{\mathbf{\Gamma}}\mathbf{h}_i}\Big(\langle \mathbf{r}\rangle - \sum_{j\neq i}^{K}\mathbf{h}_j\langle x_j\rangle\Big)$$
$$= \langle x_i\rangle + \frac{\mathbf{h}_i^H\hat{\mathbf{\Gamma}}}{\mathbf{h}_i^H\hat{\mathbf{\Gamma}}\mathbf{h}_i}\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big), \tag{15}$$

and $\langle x_i\rangle$ is the *current* nonlinear estimate of x_i. Here, z_i is the LMMSE estimate of x_i using the LMMSE filter $\mathbf{h}_i^H\hat{\mathbf{\Gamma}}/(\mathbf{h}_i^H\hat{\mathbf{\Gamma}}\mathbf{h}_i)$. The variational distribution $q(x_i)$ can be realized by normalizing $p(x_i)\,\mathcal{CN}\big(z_i; x_i, 1/(\mathbf{h}_i^H\hat{\mathbf{\Gamma}}\mathbf{h}_i)\big)$.

In the ***M-step***, the estimate of $\mathbf{\Gamma}$ is updated to maximize $\ln p(\mathbf{y},\mathbf{r},\mathbf{x};\mathbf{\Gamma},\mathbf{H})$ w.r.t. $q(\mathbf{r},\mathbf{x})$, i.e.,

$$\hat{\mathbf{\Gamma}} = \arg\max_{\mathbf{\Gamma}}\ \big\langle \ln p(\mathbf{r}|\mathbf{x};\mathbf{\Gamma},\mathbf{H})\big\rangle$$
$$= \arg\max_{\mathbf{\Gamma}}\ \ln|\mathbf{\Gamma}| - \big\langle(\mathbf{r}-\mathbf{H}\mathbf{x})^H\mathbf{\Gamma}(\mathbf{r}-\mathbf{H}\mathbf{x})\big\rangle. \tag{16}$$

By applying Theorem 1, we have

$$\big\langle(\mathbf{r}-\mathbf{H}\mathbf{x})^H\mathbf{\Gamma}(\mathbf{r}-\mathbf{H}\mathbf{x})\big\rangle$$
$$= \mathrm{Tr}\Big\{\Big[\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big)\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big)^H + \mathbf{\Sigma}_\mathbf{r} + \mathbf{H}\mathbf{\Sigma}_\mathbf{x}\mathbf{H}^H\Big]\mathbf{\Gamma}\Big\},$$

where $\mathbf{\Sigma}_\mathbf{r} = \mathrm{diag}(\tau_{r_1},\ldots,\tau_{r_M})$, $\mathbf{\Sigma}_\mathbf{x} = \mathrm{diag}(\tau_{x_1},\ldots,\tau_{x_K})$. Note that τ_{r_i} and τ_{x_i} are the variational variance of r_i and x_i, respectively. Thus, a new estimate of $\mathbf{\Gamma}$ is given by

$$\hat{\mathbf{\Gamma}} = \Big(\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big)\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big)^H + \mathbf{\Sigma}_\mathbf{r} + \mathbf{H}\mathbf{\Sigma}_\mathbf{x}\mathbf{H}^H\Big)^{-1}. \tag{17}$$

We note that the matrix inversion in (17) often results in numerical errors due to the rank deficiency of $\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big)\big(\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\big)^H + \mathbf{\Sigma}_\mathbf{r} + \mathbf{H}\mathbf{\Sigma}_\mathbf{x}\mathbf{H}^H$. Similar to the approach in [14], we propose to use the following estimator

$$\hat{\mathbf{\Gamma}} = \left(\frac{\|\langle \mathbf{r}\rangle - \mathbf{H}\langle \mathbf{x}\rangle\|^2}{M}\mathbf{I}_M + \mathbf{\Sigma}_\mathbf{r} + \mathbf{H}\mathbf{\Sigma}_\mathbf{x}\mathbf{H}^H\right)^{-1}. \tag{18}$$

for the precision matrix $\mathbf{\Gamma}$.

By iteratively optimizing $\{q(r_m)\}$, $\{q(x_i)\}$, and $\hat{\mathbf{\Gamma}}$, we obtain the VBEM algorithm for estimating $\mathbf{r}$, $\mathbf{x}$, and $\mathbf{\Gamma}$. We refer to this scheme as the ***LMMSE-QVB algorithm*** due to the use of the LMMSE filter $\mathbf{h}_i^H\hat{\mathbf{\Gamma}}/(\mathbf{h}_i^H\hat{\mathbf{\Gamma}}\mathbf{h}_i)$ to obtain the linear estimate z_i of x_i in (15).

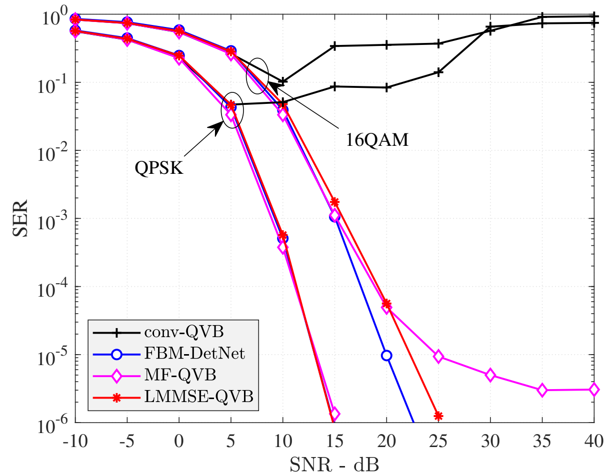

Fig. 1: SER comparison for i.i.d. channels with $b = 3$ bits, $K = 16$, $M = 32$ and $M = 64$ for QPSK and 16QAM, respectively.

V. Numerical Results

This section presents numerical results comparing the performance of the proposed VB-based methods with the conventional quantized VB-based method, denoted as conv-QVB, in [13] and FBM-DetNet in [6], which are the most recent and related methods to the work in this paper. The maximum number of iterations is set to 50 for all the iterative algorithms. The covariance matrices $\mathbf{C}_i$ are normalized such that their diagonal elements are 1, which implies $\mathbb{E}[\|\mathbf{h}_i\|^2] = M$, $\forall i$. The noise variance N_0 is set according to the operating SNR, which is defined as $\text{SNR} = \mathbb{E}[\|\mathbf{H}\mathbf{x}\|^2]/\mathbb{E}[\|\mathbf{n}\|^2] = K/N_0$. For i.i.d. channels, we set $\mathbf{C}_i = \mathbf{I}$, $\forall i$. For spatially correlated channels, we use the typical urban channel model in [16] where the power angle spectrum of the channel model follows a Laplacian distribution with an angle spread of $10°$. The covariance matrix $\mathbf{C}_i$ is obtained according to [17, Eq. (2)].

Results for i.i.d. and spatially correlated channels are shown in Fig. 1 and Fig. 2, respectively. It can be seen that, for both i.i.d. and correlated channels, the conv-QVB method in [13] is outperformed by all other methods and its performance is severely degraded at high SNRs. This is because conv-QVB does not take into account the residual inter-user interference and often encounters the catastrophic cancellation issue at high SNR. For i.i.d. channels, FBM-DetNet, MF-QVB, and

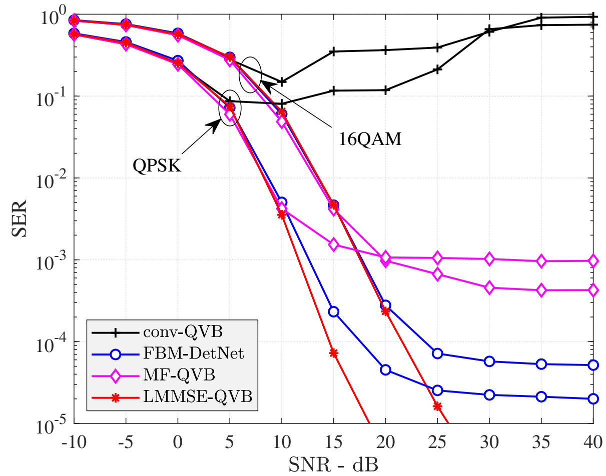

Fig. 2: SER comparison for spatially correlated channels with $b = 3$ bits, $K = 16$, $M = 32$ and $M = 64$ for QPSK and 16QAM, respectively.

LMMSE-QVB all yield the same performance for QPSK signals, while for 16QAM FBM-DetNet and LMMSE-QVB are similar and both outperform MF-QVB. For spatially correlated channels, LMMSE-QVB provides a significantly lower SER than FBM-DetNet and MF-QVB due to its estimation of the precision matrix $\mathbf{\Gamma}$ which can better represent the effect of the residual inter-user interference.

VI. Conclusion

We have proposed MF-QVB and LMMSE-QVB data detection methods for MIMO detection with low-resolution ADCs. The proposed methods exploit the VB framework to efficiently estimate the posterior distribution of the transmitted signal given the channel and the received signal. The proposed methods do not require prior information about the noise power and properly compensates for the residual inter-user interference plus noise, and therefore significantly outperforms other existing methods.

Appendix A
Proof of Theorem 1

Expanding $\left\langle \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \right\rangle$ and taking into account the independence between $\mathbf{A}$, $\mathbf{y}$, and $\mathbf{x}$, we have

$$\begin{aligned}
&\left\langle (\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{B} (\mathbf{y} - \mathbf{A}\mathbf{x}) \right\rangle \\
&= \langle \mathbf{y}^H \mathbf{B} \mathbf{y} \rangle - 2\,\Re\left\{ \langle \mathbf{y}^H \mathbf{B} \mathbf{A} \mathbf{x} \rangle \right\} + \langle \mathbf{x}^H \mathbf{A}^H \mathbf{B} \mathbf{A} \mathbf{x} \rangle \\
&= (\langle \mathbf{y} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle)^H \mathbf{B} (\langle \mathbf{y} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle) + \langle \mathbf{y}^H \mathbf{B} \mathbf{y} \rangle \\
&\quad - \langle \mathbf{y}^H \rangle \mathbf{B} \langle \mathbf{y} \rangle + \langle \mathbf{x}^H \mathbf{A}^H \mathbf{B} \mathbf{A} \mathbf{x} \rangle - \langle \mathbf{x}^H \rangle \langle \mathbf{A}^H \rangle \mathbf{B} \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle.
\end{aligned} \tag{19}$$

Note that $\langle \mathbf{x}\mathbf{x}^H \rangle = \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^H + \mathbf{\Sigma}_{\mathbf{x}}$ and $\langle \mathbf{y}^H \mathbf{B} \mathbf{y} \rangle = \mathrm{Tr}\{\mathbf{B} \langle \mathbf{y}\mathbf{y}^H \rangle\} = \langle \mathbf{y}^H \rangle \mathbf{B} \langle \mathbf{y} \rangle + \mathrm{Tr}\{\mathbf{B}\mathbf{\Sigma}_{\mathbf{y}}\}$. In addition, we have

$$\left[\langle \mathbf{A}^H \mathbf{B} \mathbf{A} \rangle \right]_{ij} = \langle \mathbf{a}_i^H \mathbf{B} \mathbf{a}_j \rangle = \begin{cases} \langle \mathbf{a}_i^H \rangle \mathbf{B} \langle \mathbf{a}_i \rangle + \mathrm{Tr}\{\mathbf{B}\mathbf{\Sigma}_{\mathbf{a}_i}\}, & \text{if } i = j \\ \langle \mathbf{a}_i^H \rangle \mathbf{B} \langle \mathbf{a}_j \rangle, & \text{otherwise.} \end{cases}$$

It thus follows that $\langle \mathbf{A}^H \mathbf{B} \mathbf{A} \rangle = \langle \mathbf{A}^H \rangle \mathbf{B} \langle \mathbf{A} \rangle + \mathbf{D}$, and

$$\begin{aligned}
\langle \mathbf{x}^H \mathbf{A}^H \mathbf{B} \mathbf{A} \mathbf{x} \rangle &= \mathrm{Tr}\left\{ \langle \mathbf{A}^H \rangle \mathbf{B} \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle \langle \mathbf{x}^H \rangle \right\} + \langle \mathbf{x} \rangle^H \mathbf{D} \langle \mathbf{x} \rangle \\
&\quad + \mathrm{Tr}\left\{ \mathbf{\Sigma}_{\mathbf{x}} \mathbf{D} \right\} + \mathrm{Tr}\left\{ \mathbf{\Sigma}_{\mathbf{x}} \langle \mathbf{A}^H \rangle \mathbf{B} \langle \mathbf{A} \rangle \right\}.
\end{aligned}$$

The statement (2) thus follows by removing the duplicated terms in (19). Note that $\langle \mathbf{x} \rangle^H \mathbf{D} \langle \mathbf{x} \rangle + \mathrm{Tr}\left\{ \mathbf{\Sigma}_{\mathbf{x}} \mathbf{D} \right\}$ can also be written as $\langle \mathbf{x} \rangle^H \mathbf{D} \langle \mathbf{x} \rangle + \mathrm{Tr}\left\{ \mathbf{\Sigma}_{\mathbf{x}} \mathbf{D} \right\} = \sum_{i=1}^{n} \langle |x_i|^2 \rangle \mathrm{Tr}\{\mathbf{B}\mathbf{\Sigma}_{\mathbf{a}_i}\}$.

Acknowledgment

This work was supported by the U.S. National Science Foundation under Grants ECCS-2146436 and CCF-2225576.

References

[1] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[2] A. S. Lan, M. Chiang, and C. Studer, "Linearized binary regression," in *Proc. Annual Conf. on Inform. Sciences and Systems*, Princeton, NJ, USA, Mar. 2018.

[3] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "Linear and deep neural network-based receivers for massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7333–7345, Nov. 2021.

[4] N. Kolomvakis, T. Eriksson, M. Coldrey, and M. Viberg, "Quantized uplink massive MIMO systems with linear receivers," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, June 2020.

[5] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "SVM-based channel estimation and data detection for one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2086–2099, 2021.

[6] L. V. Nguyen, D. H. Nguyen, and A. L. Swindlehurst, "Deep learning for estimation and pilot signal design in few-bit massive MIMO systems," *IEEE Trans. Wireless Commun. (Early Access)*, 2022.

[7] S. Khobahi, N. Shlezinger, M. Soltanalian, and Y. C. Eldar, "LoRD-Net: Unfolded deep detection network with low-resolution receivers," *IEEE Trans. Signal Process.*, vol. 69, pp. 5651–5664, 2021.

[8] Y. Jeon, S. Hong, and N. Lee, "Supervised-learning-aided communication framework for MIMO systems with low-resolution ADCs," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7299–7313, Aug. 2018.

[9] L. V. Nguyen, D. T. Ngo, N. H. Tran, A. L. Swindlehurst, and D. H. N. Nguyen, "Supervised and semi-supervised learning for MIMO blind detection with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2427–2442, Apr. 2020.

[10] S. Kim, J. Chae, and S.-N. Hong, "Machine learning detectors for MU-MIMO systems with one-bit ADCs," *IEEE Access*, vol. 8, pp. 86 608–86 616, Apr. 2020.

[11] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[12] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing–Part I: Derivation," vol. 62, no. 22, pp. 5839–5853, 2014.

[13] S. S. Thoota and C. R. Murthy, "Variational Bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low resolution ADCs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3467–3481, May 2021.

[14] D. H. Nguyen, I. Atzeni, A. Tölli, and A. L. Swindlehurst, "A variational Bayesian perspective on massive MIMO detection," *arXiv preprint arXiv:2205.11649*, 2022.

[15] C. M. Bishop, *Pattern recognition and machine learning*. New York, NY, USA: Springer, 2006.

[16] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[17] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, June 2015.