Variational Bayesian Symbol Detection for Massive MIMO Systems with Symbol-Dependent Transmit Impairments

Wang, P.; Koike-Akino, T.; Orlik, P.V.; Pajovic, M.; Kim, K.J.

TR2019-139 December 09, 2019

Abstract

In this paper, we propose a variational Bayesian inference approach for a low-complexity symbol detection for massive MIMO systems with symbol-dependent transmit-side impairments. This study is motivated by observations that realworld communication transceivers are often affected by the hardware impairments, such as non-linearities of power amplifiers, I/Q imbalance, phase drifts due to non-ideal oscillators, and carrier frequency offsets. Particularly, symbol-dependent perturbations are fully accounted into the designed hierarchical signal model as unknown model parameters. The developed variational Bayesian symbol detector is able to learn the unknown perturbations in an iterative fashion. Numerical evaluation confirms the effectiveness of the proposed approach.

IEEE Global Communications Conference (GLOBECOM)

Variational Bayesian Symbol Detection for Massive MIMO Systems with Symbol-Dependent Transmit Impairments

Pu Wang, Toshiaki Koike-Akino, Philip V. Orlik, Milutin Pajovic, Kyeong Jin Kim
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA 02139, USA
e-mail: pwang@merl.com

***Abstract*—In this paper, we propose a variational Bayesian inference approach for a low-complexity symbol detection for massive MIMO systems with symbol-dependent transmit-side impairments. This study is motivated by observations that real-world communication transceivers are often affected by the hardware impairments, such as non-linearities of power amplifiers, I/Q imbalance, phase drifts due to non-ideal oscillators, and carrier frequency offsets. Particularly, symbol-dependent perturbations are fully accounted into the designed hierarchical signal model as unknown model parameters. The developed variational Bayesian symbol detector is able to learn the unknown perturbations in an iterative fashion. Numerical evaluation confirms the effectiveness of the proposed approach.**

***Index Terms*—Massive MIMO systems, hardware impairments, BPSK, variational Bayesian inference.**

I. Introduction

Massive multiple-input multiple-output (MIMO) is a promising technology to meet the ever growing demands for higher throughput and better quality-of-service of next-generation wireless communication systems. Massive MIMO systems are equipped with a large number of antennas at base station (BS) simultaneously serving a number of single-antenna users sharing the same time-frequency slot. By exploiting the asymptotic orthogonality among channel vectors associated with different users, the massive MIMO system can achieve almost perfect inter-user interference cancelation with a simple linear precoder and receive combiner, and thus have the potential to enhance the spectrum efficiency by several orders of magnitude [1].

Despite all these benefits, the massive MIMO system poses new challenges for system design and hardware implementation. For example, the hardware cost and power consumption become prohibitively high as the number of antennas at the BS is large and high-resolution analog-to-digital convertors (ADCs) are employed. Extensive studies have considered channel estimation and symbol detection for the massive MIMO system with low-resolution ADCs (e.g. 1-3 bits) [2], [3]. In addition, transmit impairments have been recently received attention for the massive MIMO system. In fact, real-world communication transceivers are often affected by the hardware impairments such as non-linearities of power amplifiers, I/Q imbalance, phase drifts due to non-ideal oscillators, and carrier frequency offsets [4]–[17].

In this paper, *symbol-dependent* perturbations due to transmit impairments are modeled as unknown parameters [4], [6],

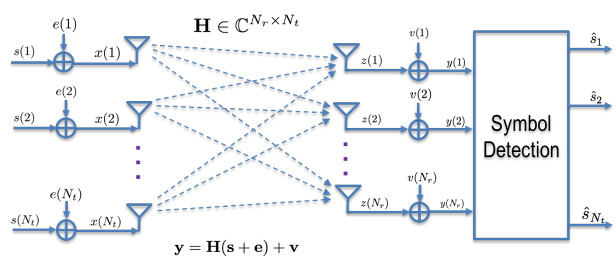


Fig. 1. The signal model of symbol detection for massive MIMO systems with transmit-side impairments.

[11], [18]. Specially, we introduce a truncated Gaussian mixture prior distribution to effectively transmitted symbols. As shown in Section IV, this prior distribution has the capability to push the solution towards unknown but fixed boundaries which can be used for later symbol detection. As a result, a hierarchical Gaussian mixture signal model on the perturbed transmitted symbol is used to enforce the finite alphabet nature, and the framework of variational Bayesian inference is employed to develop an iterative detection algorithm. One key challenge here is that, to update the model parameters such as the unknown boundaries, we need to compute the expectation of the logarithm of the normalization factor over the posterior distribution, which is difficult to derive in a closed-form expression. To address this issue, we propose an approximate, closed-form updating rule by finding the optimal the adjustment of the two boundaries for the next iteration. The performance is numerically evaluated by using the Monte-Carlo simulation.

The remainder of this paper is organized as follows. In Section II, the signal model of the massive MIMO system is introduced to account for the hardware impairment. We briefly review the boxed-LASSO approach in Section III. The iterative Bayesian approach with a new scheme of updating boundary parameters is proposed in Section IV. Simulation results are provided in Section V, followed by the conclusion in Section VI.

II. Signal Model

Consider a massive MIMO system with N_t transmit and N_r receive antennas. The transmitted symbols take values from a finite constellation set $\mathcal{A}$ (e.g., PSK or QAM). Without loss of generality, we assume $\mathcal{A} = \{\pm 1\}$. Let $\mathbf{s} \in \mathcal{A}^{N_t}$ denote the nominal transmitted vector and $\mathbf{H} \in \mathcal{C}^{N_r \times N_t}$ denote

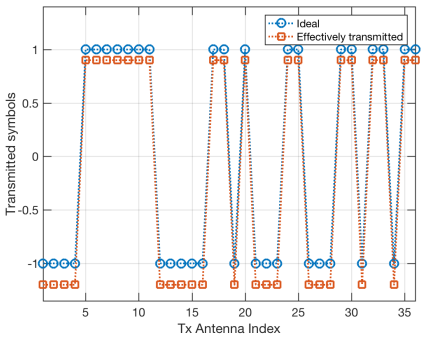


Fig. 2. An example of symbol-dependent perturbations on the transmitted symbol due to transmit-side impairments on BPSK modulation, where $u_1 = -0.1$ and $u_2 = -0.2$.

the channel gain matrix, whose entries are assumed to be independent and identically distributed (i.i.d.) Gaussian with zero mean and unit variance. The received vector $\mathbf{y} \in \mathcal{C}^{N_r \times 1}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v} = \mathbf{H}(\mathbf{s} + \mathbf{e}) + \mathbf{v}, \tag{1}$$

where $\mathbf{x}$ is the effectively transmitted vector due to the transmit impairment, $\mathbf{e}$ is the transmit perturbation vector, and $\mathbf{v}$ is the Gaussian distributed noise with zero mean and an unknown variance β^{-1}, i.e., $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I}_{N_t})$. Moreover, the nominal transmitted symbol $s_n \in \{-1, 1\}$ follows a Bernoulli distribution,

$$p(s_n; \pi) = (\pi)^{(1+s_n)/2} (1-\pi)^{(1-s_n)/2}, \tag{2}$$

where $\pi = 0.5$. Moreover, the transmit hardware impairment introduces a *symbol-dependent* permutation vector $\mathbf{e}$,

$$e_n = \begin{cases} u_1 & s_n = 1 \\ u_2 & s_n = -1 \end{cases}, \tag{3}$$

where u_1 and u_2 are unknown variables.

The problem of interest is, given the received vector $\mathbf{y}$, to detect the symbol $\mathbf{s}$ by taking into account the binary nature of $\mathbf{s}$ and in the case of unknown symbol-dependent transmit-impairments $\mathbf{e}$.

III. Prior Arts

In the following, we briefly review existing approaches for symbol detection for massive MIMO systems.

A. The Maximum-Likelihood Decoder

The maximum likelihood (ML) decoder maximizes the probability of error (assuming $\{1, -1\}$ are equally likely) is given by

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \{\pm 1\}^{N_t}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2, \tag{4}$$

which is often computationally intractable, especially when the dimension N_t is large.

B. Decorrelator

A simple relaxation of the ML decoder is to relax the feasible set to the N dimensional space $\mathbb{R}^{N_t}$

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^{N_t}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2, \tag{5}$$

which essentially removes the constraints and converts the discrete optimization problem into a continuous one. It is easy to show that

$$\mathbf{z} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y} \tag{6}$$

and the decorrelator takes the sign of the above solution as

$$\hat{\mathbf{s}} = \text{sign}\{\mathbf{z}\} \tag{7}$$

C. Box Relaxation

The constraint set of (4) consists of corner points of the unit hypercube (box). Another solution is to relax the constraint set to cover the whole hypercube and convert (4) to a convex programming problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in [-1,1]^{N_t}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \tag{8}$$

Both the cost function and the constraint set in (8) are convex. Thus, it has a unique minimum. However, the optimum point does not have a closed form and one should use iterative methods to find the solution. Then the solution to (8) (denoted as $\hat{\mathbf{x}}_{\text{BR}}$) is hard-thresholded to produce the final binary estimate.

$$\hat{\mathbf{s}} = \text{sign}\{\hat{\mathbf{x}}_{\text{BR}}\}. \tag{9}$$

Numerous low-complexity implementations of (8) have been proposed in the literature.

IV. Proposed Variational Bayesian Symbol Detection

In this section, we propose a low-complexity symbol detector for the massive MIMO system with symbol-dependent transmit impairments. Specifically, we utilize the variational Bayesian inference (VBI) framework to recover $\mathbf{x}$ with a special design of the hierarchical prior on the binary vector $\mathbf{x}$ and introduce a new scheme to update unknown boundaries of the effectively transmitted symbols.

A. Hierarchical Signal Model

It is noted that the element of $\mathbf{x}$ takes either of the binary values $\{v_2 = -1 + u_2, v_1 = 1 + u_1\}$. In order to explore this binary nature, we impose independent truncated Gaussian mixture prior distributions on the elements of $\mathbf{x}$,

$$\begin{aligned} & p(x_n|\alpha_{n1}, \alpha_{n2}, c_n; v_1, v_2), \quad x_n \in [v_2, v_1], \\ & = \left[\frac{\mathcal{N}(x_n; v_1, \alpha_{n1}^{-1})}{\eta_{n1}}\right]^{c_n} \cdot \left[\frac{\mathcal{N}(x_n; v_2, \alpha_{n2}^{-1})}{\eta_{n2}}\right]^{1-c_n}, \end{aligned} \tag{10}$$

where $c_n \in \{0, 1\}$ is a binary label variable for the n-th element x_n, and $\eta_{n1} = 0.5 - \Phi\left(-2v\sqrt{\alpha_{n1}}\right)$ and $\eta_{n2} = -0.5 + \Phi\left(2v\sqrt{\alpha_{n2}}\right)$ are the normalization factors with $v = 1 + (u_1 - u_2)/2$ and $\Phi(\cdot)$ denoting the cumulative distribution function of the standard normal distribution.

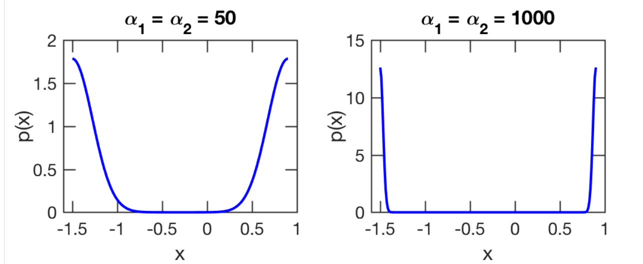


Fig. 3. Truncated Gaussian mixture prior $p(x_n)$ for the effectively transmitted symbol $x_n \in [v_2 = -1.5, v_1 = 0.9]$ with different prior precisions α_1 and α_2, when $u_1 = -0.1$, $u_2 = -0.5$ and $\pi = 0.5$.

In addition, the binary label vector $\mathbf{c} = [c_1, \ldots, c_N]^{\mathrm{T}}$ follows an *i.i.d.* Bernoulli distribution with parameter π,

$$p(c_n; \pi) = (\pi)^{c_n} (1-\pi)^{1-c_n}. \tag{11}$$

With (10) and (11), the prior distribution of x_n is given as

$$\begin{aligned} & p(x_n | \alpha_{n,1}, \alpha_{n,2}; v_1, v_2) \\ & = \sum_{c_n \in \{0,1\}} p(x_n | \alpha_{n,1}, \alpha_{n,2}, c_n; v_1, v_2) \, p(c_n; \pi) \\ & = \pi \frac{\mathcal{N}\left(x_n; v_1, \alpha_{n1}^{-1}\right)}{\eta_{n1}} + (1-\pi) \frac{\mathcal{N}\left(x_n; v_2, \alpha_{n2}^{-1}\right)}{\eta_{n2}} \end{aligned} \tag{12}$$

where $x_n \in [v_2, v_1]$. Fig. 3 shows the truncated Gaussian mixture prior $p(x)$ when the symbol-dependent perturbations are given as $u_1 = -0.1$ and $u_2 = -0.5$ (resulting in $[v_2, v_1] = [-1.5, 0.9]$) with different prior precisions α_1 and α_2. As shown in the figure, larger prior precisions (α_1 and α_2) push the prior distribution of x_n towards its boundaries and hence it captures the binary nature of x_n.

Furthermore, we treat the perturbation precision, α_{n1} and α_{n2}, as *i.i.d.* random variables and specify the the Gamma distribution as hyperpriors over these precision variables,

$$p(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2; \zeta_1, \zeta_2) = \prod_{i=1}^{2} \prod_{n=1}^{N} \mathrm{Gamma}(\alpha_{ni} | \zeta_1, \zeta_2), \tag{13}$$

where $\boldsymbol{\alpha}_1 = [\alpha_{11}, \ldots, \alpha_{N1}]^{\mathrm{T}}$, $\boldsymbol{\alpha}_2 = [\alpha_{12}, \ldots, \alpha_{N2}]^{\mathrm{T}}$, and

$$\mathrm{Gamma}(\alpha | \zeta_1, \zeta_2) = \Gamma(\zeta_1)^{-1} \zeta_2^{\zeta_1} \alpha^{\zeta_1 - 1} e^{-\zeta_2 \alpha} \tag{14}$$

with ζ_1 and ζ_2 are set as small values, e.g., $\zeta_1 = \zeta_2 = 10^{-6}$, for non-informative hyperpriors on $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. Overall, the hierarchical truncated Gaussian mixture model can be described in a graphical representation shown in Fig. 4, where hidden random variables (red circles) are given as $\{\mathbf{x}, \mathbf{c}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$, unknown model parameters include the two unknown boundary parameters $\{v_1, v_2\}$ and noise variance β^{-1}, and pre-determined hyperparameters include the prior symbol probability of $\pi = 0.5$ and the hyper-prior parameters $\zeta_1 = \zeta_2 = 10^{-6}$.

B. Variational Bayesian Symbol Detection

According to the above hierarchical signal model, we utilize the variational Bayesian inference for the posterior distributions of hidden random variables $\{\mathbf{x}, \mathbf{c}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$ and updating

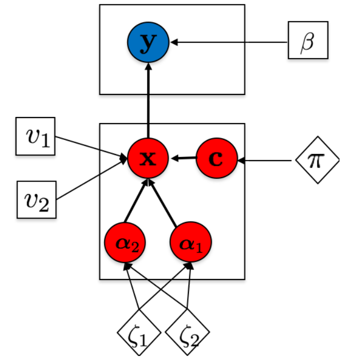


Fig. 4. A graphical representation of the signal model for the massive MIMO system with symbol-dependent impairments. $\mathbf{y}$, the measurement at the receiver side, is an observable random variable denoted by blue circles. Red circles represent hidden random variables including the effectively transmitted symbol $\mathbf{x}$, the nominal symbol $\mathbf{c}$ and the precision parameters of impairment-induced perturbation $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. Squares denote the unknown deterministic model parameters including the perturbation boundaries $\{v_1, v_2\}$ and the noise variance β^{-1}. And diamonds denote pre-determined hyperparameters, i.e., the prior probability of the nominal symbol $\pi = 0.5$ and the hyper-prior parameters $\zeta_1 = \zeta_2 = 10^{-5}$.

rules for unknown model parameters, i.e., the deterministic perturbation parameters $\{v_1, v_2\}$ and noise variance β^{-1}.

Decoupled Transmit-Channel-Based Likelihood Function: The receiver-channel-based likelihood function of $\mathbf{y}$ is given by

$$p(\mathbf{y} | \mathbf{x}; \beta) = \frac{1}{(2\pi\beta^{-1})^{N_r/2}} e^{-\frac{\beta \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2}{2}}, \tag{15}$$

where the measurement y_m at each receive channel includes contributions from all effectively transmitted symbols $\{x_n\}_{n=1}^{N_t}$ due to the mixing channel matrix $\mathbf{H}$. In order to derive the posterior distributions, it is necessary to factorize the receiver-channel-based likelihood function to a decoupled transmit-channel-based likelihood function. This can be done by using the GAMP framework which approximates the likelihood function as a product of approximate marginal likelihoods:

$$p(\mathbf{y} | \mathbf{x}; \beta) \approx \prod_{n=1}^{N_t} p(x_n | \hat{r}_n, \hat{\tau}_n) = \prod_{n=1}^{N_t} \frac{1}{\sqrt{2\pi \hat{\tau}_n}} e^{-\frac{(x_n - \hat{r}_n)^2}{2\hat{\tau}_n}}. \tag{16}$$

As a result, the receiver-channel-based likelihood function is approximately decoupled in the transmit-channel sense (with respect to the transmitting antenna index n). For each transmit-channel, we have an equivalent Gaussian marginal likelihood with mean $\hat{r}_n$ and variance $\hat{\tau}_n$. The detailed derivation of mean and variance can be found in Appendix. It is worth noting that this decoupling process of (16) has been used in the massive MIMO symbol detection [1], [14], [19] and the peak-to-average power ratio reduction for MIMO-OFDM systems [20].

Posterior Distributions of Hidden Random Variables: Then, with the decoupled likelihood function of (16), the variational expectation-maximization (EM) algorithm is used to derive the posterior distribution of the hidden random variables, i.e., $\{\mathbf{x}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \mathbf{c}\}$.

Posterior of the effectively transmitted symbol $\mathbf{x}$: We first start with the derivation of the posterior distribution of the effectively transmitted symbol vector $\mathbf{x}$. With (10) and (16) and by only keeping terms related to x_n, we have

$$\begin{aligned}\ln q(\mathbf{x}) =& \langle \ln p(\mathbf{y},\mathbf{x},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c};\boldsymbol{\theta})\rangle_{q(\boldsymbol{\alpha}_1)q(\boldsymbol{\alpha}_2)q(\mathbf{c})} + \text{const}\\ =&\langle \ln[p(\mathbf{y}|\mathbf{x};\beta)p(\mathbf{x}|\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c};v_1,v_2)]\rangle_{q(\boldsymbol{\alpha}_1)q(\boldsymbol{\alpha}_2)q(\mathbf{c})} + \text{const}\\ =& -\frac{1}{2}\sum_{n=1}^{N_t}\left[\hat{\tau}_n^{-1}(x_n-\hat{r}_n)^2 + \langle c_n\rangle\langle\alpha_{n1}\rangle(x_n-v_1)^2\right.\\ &\left. + (1-\langle c_n\rangle)\langle\alpha_{n2}\rangle(x_n-v_2)^2\right] + \text{const}\\ =& -\frac{1}{2}\sum_{n=1}^{N_t}\left(\hat{\tau}_n^{-1} + \langle c_n\rangle\langle\alpha_{n1}\rangle + (1-\langle c_n\rangle)\langle\alpha_{n2}\rangle\right)x_n^2\\ &-2\left(\hat{\tau}_n^{-1}\hat{r}_n + \langle c_n\rangle\langle\alpha_{n1}\rangle v_1 + (1-\langle c_n\rangle)\langle\alpha_{n2}\rangle v_2\right)x_n\\ &+ \text{const}, \qquad \text{if } x_n\in[v_2,v_1],\end{aligned} \tag{17}$$

where $v_1 = v+b$ and $v_2 = -v+b$. This implies the posterior distribution of $\mathbf{x}$ can be factorized into independent truncated Gaussian distribution,

$$q(x_n) = \begin{cases}\frac{\mathcal{N}(\tilde{\mu}_n,\tilde{\sigma}_n^2)}{\tilde{\eta}_n} & x_n\in[v_2,v_1]\\ 0 & \text{elsewhere}\end{cases}, \tag{18}$$

where the posterior mean $\tilde{\mu}_n$ and variance $\tilde{\sigma}_n^2$ are given as

$$\tilde{\mu}_n = \left(\hat{\tau}_n^{-1}\hat{r}_n + \langle c_n\rangle\langle\alpha_{n1}\rangle v_1 + (1-\langle c_n\rangle)\langle\alpha_{n2}\rangle v_2\right)\tilde{\sigma}_n^2, \tag{19}$$

$$\tilde{\sigma}_n^2 = \left(\hat{\tau}_n^{-1} + \langle c_n\rangle\langle\alpha_{n1}\rangle + (1-\langle c_n\rangle)\langle\alpha_{n2}\rangle\right)^{-1}, \tag{20}$$

with the normalization factor

$$\tilde{\eta}_n = \Phi\left(\tilde{\sigma}_n^{-1}(v_1-\tilde{\mu}_n)\right) - \Phi\left(\tilde{\sigma}_n^{-1}(v_2-\tilde{\mu}_n)\right) \tag{21}$$

Posterior of precision variables $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$: With (13) and (16) and by only keeping terms related to α_{n1}, we have

$$\begin{aligned}\ln q(\boldsymbol{\alpha}_1) =& \langle \ln p(\mathbf{y},\mathbf{x},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c};\boldsymbol{\theta})\rangle_{q(\mathbf{x})q(\boldsymbol{\alpha}_2)q(\mathbf{c})} + \text{const}\\ =& \langle \ln[p(\mathbf{x}|\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c};v_1,v_2)p(\boldsymbol{\alpha})]\rangle_{q(\mathbf{x})q(\boldsymbol{\alpha}_2)q(\mathbf{c})} + \text{const}\\ =& \sum_{n=1}^{N_t}\left[(0.5\langle c_n\rangle + \zeta_1 - 1)\ln\alpha_{n1}\right.\\ &\left. - \left(0.5\langle c_n\rangle\langle(x_n-v_1)^2\rangle + \zeta_2\right)\alpha_{n1} - \langle c_n\rangle\ln\eta_{n1}\right] + \text{const},\\ \approx& \sum_{n=1}^{N_t}\left[(0.5\langle c_n\rangle + \zeta_1 - 1)\ln\alpha_{n1}\right.\\ &\left. - \left(0.5\langle c_n\rangle\langle(x_n-v_1)^2\rangle + \zeta_2\right)\alpha_{n1}\right] + \text{const},\end{aligned}$$

where we have used the updated value $\ln\eta_{n1}^{(t)}$ to replace $\ln\eta_{n1}$ and make it irrelevant to the posterior distribution of α_{n1}. As a result, the posterior distribution of $\boldsymbol{\alpha}_1$ can be factorized into independent Gamma distribution, i.e.

$$q(\alpha_{n1}) = \text{Gamma}\left(\alpha_{n1}|\tilde{a}_{n1},\tilde{b}_{n1}\right), \tag{22}$$

where

$$\tilde{a}_{n1} = \zeta_1 + 0.5\langle c_n\rangle, \tag{23}$$

$$\tilde{b}_{n1} = \zeta_2 + 0.5\langle c_n\rangle\langle(x_n-v_1)^2\rangle. \tag{24}$$

Similarly, the posterior distribution of $\boldsymbol{\alpha}_2$ can be factorized into independent Gamma distribution, i.e.

$$q(\alpha_{n2}) = \text{Gamma}\left(\alpha_{n2}|\tilde{a}_{n2},\tilde{b}_{n2}\right), \tag{25}$$

where

$$\tilde{a}_{n2} = \zeta_1 + 0.5(1-\langle c_n\rangle), \tag{26}$$

$$\tilde{b}_{n2} = \zeta_2 + 0.5(1-\langle c_n\rangle)\langle(x_n-v_2)^2\rangle. \tag{27}$$

Posterior of the label variable $\mathbf{c}$: For the last class of hidden variables, the binary label variable $\mathbf{c}\in\{0,1\}$, its posterior distribution can be inferred as

$$\begin{aligned}\ln q(\mathbf{c}) =& \langle \ln p(\mathbf{y},\mathbf{x},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c};\boldsymbol{\theta})\rangle_{q(\mathbf{x})q(\boldsymbol{\alpha}_1)q(\boldsymbol{\alpha}_2)} + \text{const}\\ =&\langle \ln[p(\mathbf{x}|\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c};v_1,v_2)p(\mathbf{c})]\rangle_{q(\mathbf{x})q(\boldsymbol{\alpha}_1)q(\boldsymbol{\alpha}_2)} + \text{const}\\ =&\sum_{n=1}^{N_t}(\ell_{n1}-\ell_{n2})c_n + \text{const},\end{aligned} \tag{28}$$

where $\ell_{n1} = \ln\pi + 0.5\langle\ln\alpha_{n1}\rangle - \langle\ln\eta_{n1}\rangle - 0.5\langle\alpha_{n1}\rangle\langle(x_n - v_1)^2\rangle$ and $\ell_{n2} = \ln(1-\pi) + 0.5\langle\ln\alpha_{n2}\rangle - \langle\ln\eta_{n2}\rangle - 0.5\langle\alpha_{n2}\rangle\langle(x_n-v_2)^2\rangle$. The computation of the posterior quantities $\langle\ln\eta_{n1}\rangle$ and $\langle\ln\eta_{n2}\rangle$ is quite involved and can be replaced by their updated values $\ln\eta_{n1}^{(t)}$ and $\ln\eta_{n2}^{(t)}$ from the previous iteration. As a result, $\mathbf{c}$ has independent posterior Bernoulli distribution with the parameter $\tilde{\pi}_n = (1+e^{\ell_{n2}-\ell_{n1}})^{-1}$

$$q(c_n;\tilde{\pi}_n) = (\tilde{\pi}_n)^{c_n}(1-\tilde{\pi}_n)^{1-c_n}. \tag{29}$$

Compared with the prior distribution of c_n in (11), the posterior distribution is no longer identical since the parameter $\tilde{\pi}_n$ is now dependent on the index n.

Computation of Posterior Quantities: To update the above posterior distributions, we need to compute the following posterior quantities:

$$\begin{aligned}\langle x_n\rangle &= \tilde{\mu}_n - \frac{\tilde{\sigma}_n}{\tilde{\eta}_n}\left[\phi\left(\frac{v_1-\tilde{\mu}_n}{\tilde{\sigma}_n}\right) - \phi\left(\frac{v_2-\tilde{\mu}_n}{\tilde{\sigma}_n}\right)\right],\\ \langle x_n^2\rangle &= \tilde{\mu}_n\langle x_n\rangle + \tilde{\sigma}_n^2\\ &\quad - \frac{\tilde{\sigma}_n}{\tilde{\eta}_n}\left[v_1\phi\left(\frac{v_1-\tilde{\mu}_n}{\tilde{\sigma}_n}\right) - v_2\phi\left(\frac{v_2-\tilde{\mu}_n}{\tilde{\sigma}_n}\right)\right],\\ \langle\alpha_{n1}\rangle &= \tilde{a}_{n1}/\tilde{b}_{n1}, \quad \langle\alpha_{n2}\rangle = \tilde{a}_{n2}/\tilde{b}_{n2}\\ \langle\ln\alpha_{n1}\rangle &= \psi(\tilde{a}_{n1}) - \ln\tilde{b}_{n1}, \quad \langle\ln\alpha_{n2}\rangle = \psi(\tilde{a}_{n2}) - \ln\tilde{b}_{n2}\\ \langle c_n\rangle &= \frac{1}{1+e^{\ell_{n2}-\ell_{n1}}},\end{aligned}$$

where $\phi(x)$ is the standard normal probability density function at the value of x, and $\psi(a) = \partial\ln\Gamma(a)/\partial a$ is the digamma function [21].

Update of Deterministic Model Parameters: In the following, we obtain the updating rules for three deterministic parameters $\boldsymbol{\theta} = \{\beta, v_1, v_2\}$. The general rule is to maximize the Q-function with respect to the unknown parameters [22]

$$\left\{\boldsymbol{\theta}^{\text{NEW}}\right\} = \underset{\boldsymbol{\theta}}{\text{argmax}}\,\langle\ln p(\mathbf{y},\mathbf{x},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c};\boldsymbol{\theta})\rangle_{q(\mathbf{x})q(\boldsymbol{\alpha}_1)q(\boldsymbol{\alpha}_2)q(\mathbf{c})}$$

where the Q-function is obtained as the expectation of the logarithm of the complete likelihood function $(\mathbf{y},\mathbf{x},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\mathbf{c})$

with respect to the posterior distributions of all hidden variables $\{\mathbf{x}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \mathbf{c}\}$.

First, the corresponding Q-function of β can be expressed as

$$Q(\beta, \beta^{(k)}) = \sum_{m=1}^{N_r} \langle \ln p(y_m|d_m;\beta)\rangle_{p(d_m|\mathbf{y};\beta)} + \text{const}, \quad (30)$$

where $y_m = d_m + v_m$ and d_m is the m-th element of the noiseless measurement $\mathbf{d} = \mathbf{H}\mathbf{x}$ whose posterior distribution $p(d_m|\mathbf{y};\beta)$ can be found in Step 2 of the Appendix. In other words, the equivalent variable $\mathbf{d}$ can summarize all contributions from the hidden variables $\{\mathbf{x}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \mathbf{c}\}$. Then it is straightforward to show that

$$Q(\beta, \beta^{(k)}) = -\frac{1}{2}\sum_{m=1}^{N_r} [\ln \beta + \beta\langle (y_m - d_m)^2\rangle] + \text{const}, \quad (31)$$

which yields

$$\beta^{(k+1)} = \frac{N_r}{\sum_{m=1}^{N_r} \langle (y_m - d_m)^2\rangle}, \quad (32)$$

where the expectation is taken over the posterior distribution of d_m.

Next, to update the two unknown boundary values $\{v_1, v_2\}$, the corresponding Q-function of $\{v_1, v_2\}$ is difficult to find a closed-form expression. Alternatively, we consider a least-square updating procedure. Specifically, we minimize the following cost function

$$[\Delta\hat{b}, \Delta\hat{v}] = \arg\min_{\Delta b, \Delta v} \|\mathbf{y} - \mathbf{A}(\hat{\mathbf{x}}^{(k)} + \Delta b \mathbf{1} + \Delta v \mathbf{h})\|_2^2 \quad (33)$$

where $\mathbf{1}$ is the all-one vector and $h_n = \mathbf{h}(n) = 1$ if $x_n^{(k)} > \hat{b}^{(k)}$ or $h_n = -1$ if $x_n^{(k)} \leq \hat{b}^{(k)}$ with $\hat{b}^{(k)} = (\hat{v}_1^{(k)} + \hat{v}_2^{(k)})/2$ denoting the estimated middle point of the unknown interval at the k-th iteration. It can be seen that the updating rule of $\{v_1, v_2\}$ is converted to the updating of the middle point b and the marginal distance v to the current estimate of bounds by finding the optimal adjustments $\{\Delta b, \Delta v\}$. More precisely, (33) minimizes the data fitting error using the current estimate of $\mathbf{x}$, i.e., $\hat{\mathbf{x}}^{(k)}$, the adjustment of the mean Δb and its marginal distance Δv to the two boundaries. The exact solution of $\{\Delta b, \Delta v\}$ is given as

$$\begin{bmatrix} \Delta b^{(k+1)} \\ \Delta v^{(k+1)} \end{bmatrix} = \left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{Q}^T(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{(k)}) \quad (34)$$

where $\mathbf{Q} = \mathbf{A}[\mathbf{1}, \mathbf{h}]$. Finally, the two boundaries $\{v_1, v_2\}$ can be updated as

$$\begin{aligned} v_1^{(k+1)} &= v_1^{(k)} + (\Delta b^{(k+1)} + \Delta v^{(k+1)}), \\ v_2^{(k+1)} &= v_2^{(k)} + (\Delta b^{(k+1)} - \Delta v^{(k+1)}). \end{aligned} \quad (35)$$

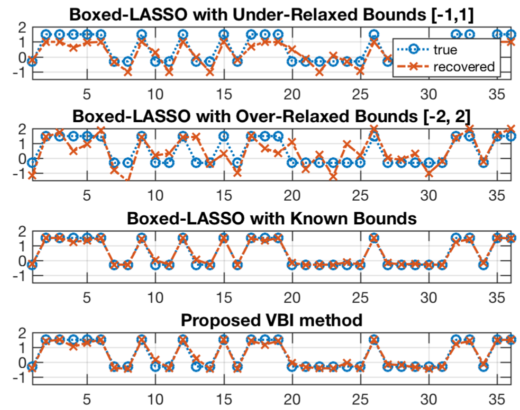
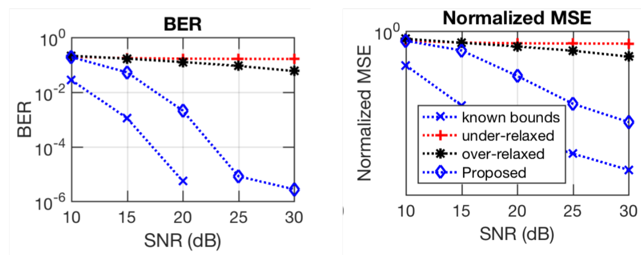


Fig. 5. Recovered transmitted symbol $x_n \in [-0.3, 1.5]$ with $u_1 = 0.5$, $u_2 = 0.7$ using various methods when SNR = 20 dB.



(a) (b)

Fig. 6. Performance comparison in terms of (a) BER and (b) normalized MSE as a function of SNRs.

V. Numerical Results

In this section, numerical results are provided to evaluate the proposed symbol detector. Specifically, we consider the a MIMO system of $N_t = N_r = 32$ transmit and receive antennas with BPSK modulation. The symbol-dependent perturbation is $u_1 = 0.5$ and $u_2 = 0.7$ which results in $[v_2, v_1] = [-0.3, 1.5]$. The channel matrix is generated as the Gaussian matrix with zero mean and unit variance. The SNR is defined on a basis of per receive-antenna, i.e., SNR $= \|\mathbf{A}\mathbf{x}\|^2/(M\sigma^2)$. We compare the proposed symbol detector with the boxed-LASSO approach of (8) with 1) under-relaxed bounds $[-1, 1]$, 2) over-relaxed bounds $[-2, 2]$, and 3) known bounds $[-0.3, 1.5]$ which serves as the performance benchmark for all methods.

Fig. 5 shows an illustrative example of the recovered transmitted signal x_n for all considered methods when SNR = 20 dB. It is shown that the boxed-LASSO approach with mismatched bounds gives more fluctuating estimates of x_n over the proposed estimates. On the other hand, the proposed variational Bayesian symbol detection gives similar results to those of the boxed-LASSO with known bounds.

Fig. 6 (a) shows the bit error rate (BER) for all considered methods when SNR varies from 10 dB to 30 dB. It is clear that the boxed-LASSO approach with known bounds provides the best performance while the ones with mismatched (under-relaxed and over-relaxed) bounds give worse performance. The proposed VBI approach gives better performance than the boxed-LASSO with either under-relaxed or over-relaxed bounds. Fig. 6 (b) shows the normalized MSE $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2/\|\mathbf{x}\|_2^2$

for all considered methods. Similar observations can be made from the normalized MSE criterion.

VI. Conclusion

In this paper, we proposed the variational Bayesian symbol detection for the massive MIMO system which is subject to symbol-dependent transmit-side impairments. Specifically, we imposed a truncated Gaussian mixture prior distribution to the perturbed transmitted symbol to capture the binary nature. With a hierarchical signal model, we obtained the posterior distributions of all hidden variables, e.g., the effectively transmitted symbols, and closed-form updating formulas for unknown model parameters, e.g., the unknown impairment-induced perturbation parameters.

VII. Appendix

To get the approximate likelihood function of (16), we need to compute the approximate mean $\hat{r}_n$ and variance $\hat{\tau}_n$, which can be obtained by using the GAMP algorithm [23] with inputs from the means $\hat{x}_n = \langle x_n \rangle_{q(x_n)}$, variances $\tau_n^x = \langle (x_n - \hat{x}_n)^2 \rangle_{q(x_n)}$, and the noise variance β^{-1}. Particularly, to compute the decoupled likelihoods $\mathcal{N}(x_n|\hat{r}_n, \hat{\tau}_n)$ and the posterior likelihood of the noiseless measurement $\mathcal{N}(d_m|\hat{d}_m, \hat{\tau}_m^w)$, we follow the steps below:

- Initialize $\hat{s}_m = 0, \quad m = 1, \cdots, N_r$;
- **Step 1**: for all $m = 1, \cdots, N_r$:

$$\hat{\tau}_m^x = \sum_n H_{mn}^2 \tau_n^x, \quad \hat{p}_m^x = \sum_n H_{mn}\hat{x}_n - \hat{\tau}_m^x \hat{s}_m,$$

  where H_{mn} is the (m,n)-th element of $\mathbf{H}$.
- **Step 2**: for all $m = 1, \cdots, N_r$, compute the posterior mean and variance of d_m with respect to $p(d_m|y_m, \hat{\tau}_m^x, \hat{p}_m^x)$, i.e.,

$$\hat{d}_m = \langle d_m \rangle_{p(d_m|y_m, \hat{\tau}_m^x, \hat{p}_m^x)},$$
$$\hat{\tau}_m^d = \langle (d_m - \hat{d}_m)^2 \rangle_{p(d_m|y_m, \hat{\tau}_m^x, \hat{p}_m^x)},$$

  and update

$$\hat{s}_m = \frac{\hat{d}_m - \hat{p}_m^x}{\hat{\tau}_m^x}, \quad \hat{\tau}_m^s = \frac{1 - \hat{\tau}_m^d / \hat{\tau}_m^x}{\hat{\tau}_m^x},$$

- **Step 3**: for all $n = 1, \cdots, N_t$, compute the mean and variance of the decoupled likelihood function

$$\hat{\tau}_n = \left(\sum_m H_{mn}^2 \hat{\tau}_m^s \right)^{-1}, \quad \hat{r}_n = \hat{x}_n + \hat{\tau}_n \sum_m H_{mn}\hat{s}_m.$$

References

[1] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1227–1231.

[2] P. Wang, J. Li, M. Pajovic, P. T. Boufounos, and P. V. Orlik, "On angular-domain channel estimation for one-bit massive MIMO systems with fixed and time-varying thresholds," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, Oct 2017, pp. 1056–1060.

[3] F. Liu, H. Zhu, J. Li, P. Wang, and P. V. Orlik, "Massive MIMO channel estimation using signed measurements with antenna-varying thresholds," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*, June 2018, pp. 188–192.

[4] B. Goransson, S. Grant, E. Larsson, and Z. Feng, "Effect of transmitter and receiver impairments on the performance of MIMO in HSDPA," in *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*, July 2008, pp. 496–500.

[5] H. Suzuki, T. V. A. Tran, I. B. Collings, G. Daniels, and M. Hedley, "Transmitter noise effect on the performance of a MIMO-OFDM hardware implementation achieving improved coverage," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 6, pp. 867–876, August 2008.

[6] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *2010 International ITG Workshop on Smart Antennas (WSA)*, Feb. 2010, pp. 189–196.

[7] E. Björnson, P. Zetterberg, and M. Bengtsson, "Optimal coordinated beamforming in the multicell downlink with transceiver impairments," in *2012 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2012, pp. 4775–4780.

[8] J. Qi and S. Aissa, "On the power amplifier nonlinearity in MIMO transmit beamforming systems," *IEEE Transactions on Communications*, vol. 60, no. 3, pp. 876–887, March 2012.

[9] B. Maham and O. Tirkkonen, "Transmit antenna selection OFDM systems with transceiver I/Q imbalance," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 865–871, Feb. 2012.

[10] E. Björnson, P. Zetterberg, M. Bengtsson, and B. Ottersten, "Capacity limits and multiplexing gains of MIMO channels with transceiver impairments," *IEEE Communications Letters*, vol. 17, no. 1, pp. 91–94, January 2013.

[11] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[12] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 711–723, Feb. 2015.

[13] M. Vehkaperä, T. Riihonen, M. A. Girnyk, E. Bjrnson, M. Debbah, L. K. Rasmussen, and R. Wichman, "Asymptotic analysis of SU-MIMO channels with transmitter noise and mismatched joint decoding," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 749–765, March 2015.

[14] R. Ghods, C. Jeon, A. Maleki, and C. Studer, "Optimal large-MIMO data detection with transmit impairments," in *The 53rd Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2015.

[15] P. Wang, M. Pajovic, P. V. Orlik, T. Koike-Akino, K. J. Kim, and J. Fang, "Sparse channel estimation in millimeter wave communications: Exploiting joint AoD-AoA angular spread," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[16] M. Pajovic, P. Wang, T. Koike-Akino, and P. Orlik, "Estimation of frequency unsynchronized millimeter-wave channels," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2017, pp. 1205–1209.

[17] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1123–1133, Feb 2018.

[18] M. Wenk, *MIMO-OFDM Testbed: Challenges, Implementations, and Measurement Results*, Series in microelectronics. ETH, 2010.

[19] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 902–915, Oct 2014.

[20] H. Bao, J. Fang, Z. Chen, H. Li, and S. Li, "An efficient Bayesian PAPR reduction method for OFDM-based massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4183–4195, 2016.

[21] M. Abramowitz and I. A.Ryzhik, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, NY, 1972.

[22] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.

[23] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *2011 IEEE International Symposium on Information Theory Proceedings*, July 2011, pp. 2168–2172.