MIMO DETECTION BY VARIATIONAL POSTERIOR INFERENCE

Junbin Liu, Mingjie Shao, and Wing-Kin Ma

Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR of China

ABSTRACT

In this paper we examine the application of variational inference (VI) to MIMO detection. Our study is motivated by the recent interest in applying machine learning concepts to signal processing. VI is an approach for providing friendly approximations of certain intractable posterior probabilities in statistics, and it has been popularly used in machine learning. In MIMO detection we also have a similar problem; specifically, we want to evaluate the posterior symbol probabilities for detection, but they are computationally too expensive to evaluate when the problem size and/or the constellation size are large. By approximating the discrete symbol prior by a continuous Gaussian mixture model, we show how the notion of VI can be used to derive an iterative MIMO detector. Interestingly, the detector resembles the MMSE detector in structure. The performance of the proposed detector is demonstrated by simulations.

Index Terms— MIMO detection, variational inference.

1. INTRODUCTION

MIMO detection is a classical problem in MIMO communications and has been studied for decades. We have seen a strikingly rich variety of MIMO detection methods, such as sphere decoding [1, 2], convex relaxations [3, 4], lattice reduction [5, 6], and various soft detection methods [7–11]. The topic has recently been reinvigorated by the application of deep learning [12–15]. While MIMO detection has been extensively studied, some open challenges remain. In particular, achieving low symbol-error probability with low computational complexity is a significant challenge in the regime of large transmit dimensions and higher-order possibly non-constant modulus symbol constellations.

The rise of machine learning has recently led researchers to revisit signal processing problems from a machine learning perspective; the application of deep learning to MIMO detection was exactly motivated by such interest [12–15]. In this paper, we are interested in variational inference (VI) [16–18]. In statistics and machine learning, VI is an approach for approximating the posterior probability for a class of inference problems. As examples, VI appeared in quantum mechanics in the 1990's [19], in the famous latent Dirichlet allocation for topic modeling in the 2000's [20], and in the very popular variational auto-encoding for unsupervised deep learning in recent years [21]. The key aspect in these contexts is to obtain the expressions of the posterior probabilities, but unfortunately, they are intractable integrals. The idea of VI is to approximate the posterior probability by a restricted, but tractable, family of distributions; optimization is used to find the best approximate distribution.

In this paper, we explore how VI can be applied to MIMO detection. Specifically, we study the approximate inference of the posterior symbol probabilities via VI. In this connection, it is worth noting that the problem of calculating the posterior symbol probabilities has been extensively studied in, and at the heart of, soft MIMO detection [7–11]. But it seems that VI has not been considered, to our knowledge. As we will see, the VI MIMO detection method developed by us resembles an iterative MMSE detector. Moreover, our simulation results suggest that the proposed method works reasonably well.

2. SIGNAL MODEL

We consider a standard problem statement for multiple-input multiple-output (MIMO) detection. Namely, we have a model:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{C}^M$ is the received signal; $\boldsymbol{x} \in \mathcal{S}^N$ is the transmitted signal with every element x_i drawn from a symbol constellation set $\mathcal{S}$ in a uniform, independent, and identically distributed (i.i.d.) fashion; $\boldsymbol{H} \in \mathbb{C}^{M \times N}$ is the channel matrix; $\boldsymbol{n} \in \mathbb{C}^M$ is element-wise i.i.d circular complex Gaussian noise with zero mean and variance σ^2. The problem is to detect the symbol vector $\boldsymbol{x}$ given the observation $\boldsymbol{y}$ and the channel matrix $\boldsymbol{H}$.

In this work we are interested in computing the posterior symbol probability $p(x_i|\boldsymbol{y})$ for MIMO detection:

$$p\left(x_i|\boldsymbol{y}\right) = \sum_{\tilde{\boldsymbol{x}}_{-i} \in \mathcal{S}^{N-1}} p\left(x_i, \tilde{\boldsymbol{x}}_{-i}|\boldsymbol{y}\right), \tag{2}$$

where $\boldsymbol{x}_{-i} \in \mathbb{C}^{N-1}$ is a subvector of $\boldsymbol{x}$ that has x_i excluded; $p(\boldsymbol{x}|\boldsymbol{y})$ is the probability of $\boldsymbol{x}$ given $\boldsymbol{y}$, and it can be written as:

$$\begin{aligned} p\left(\boldsymbol{x}|\boldsymbol{y}\right) &= p\left(\boldsymbol{y}|\boldsymbol{x}\right) p(\boldsymbol{x})/p(\boldsymbol{y}), \\ p(\boldsymbol{y}) &= \sum_{\tilde{\boldsymbol{x}} \in \mathcal{S}^N} p\left(\boldsymbol{y}|\tilde{\boldsymbol{x}}\right) p(\tilde{\boldsymbol{x}}). \end{aligned} \tag{3}$$

Here, $p\left(\boldsymbol{y}|\boldsymbol{x}\right) = \mathcal{CN}\left(\boldsymbol{y}; \boldsymbol{H}\boldsymbol{x}, \sigma^2\boldsymbol{I}\right)$ is the distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$; $\mathcal{CN}\left(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ denotes a circular complex Gaussian distribution function with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$; $p(\boldsymbol{x})$ is the prior distribution of $\boldsymbol{x}$ and is given by:

$$\begin{aligned} p(x_i) &= \frac{1}{|\mathcal{S}|}\mathbb{1}_{\mathcal{S}}(x_i), \forall i, \\ p(\boldsymbol{x}) &= \prod_{i=1}^{N} p(x_i), \end{aligned} \tag{4}$$

where $\mathbb{1}_{\mathcal{S}}(x)$ is the indicator function, i.e., $\mathbb{1}_{\mathcal{S}}(x) = 1$ if $x \in \mathcal{S}$ and $\mathbb{1}_{\mathcal{S}}(x) = 0$ if $x \notin \mathcal{S}$. With $p(x_i|\boldsymbol{y})$, we can perform either hard decision $\left(\text{e.g, } \hat{x}_i = \arg\max_{x_i \in \mathcal{S}} p(x_i|\boldsymbol{y})\right)$ or soft decision (typically by further computing the bit-level posterior log-likelihood ratio (LLR) [7, 22]). The challenge of calculating $p(x_i|\boldsymbol{y})$ lies in

This work was supported by a General Research Fund (GRF) of the Research Grant Council (RGC), Hong Kong, under Project ID CUHK 142017318.

the complexity of evaluating the summations in (2) and (3), which scales exponentially with the transmit dimension N. In the MIMO detection literature, there has been much interest in finding effecient approximation schemes for computing $p(x_i|\boldsymbol{y})$ or the closely-related bit-level posterior LLR. For example, in max-log-approximation, the rationale is to approximate (2) by

$$\begin{aligned} p(x_i|\boldsymbol{y}) &\approx \max_{\tilde{\boldsymbol{x}}_{-i}\in\mathcal{S}^{N-1}} p\left(x_i, \tilde{\boldsymbol{x}}_{-i}|\boldsymbol{y}\right) \\ &\propto \max_{\tilde{\boldsymbol{x}}\in\mathcal{S}^N, \tilde{x}_i = x_i} -\|\boldsymbol{y}-\boldsymbol{H}\tilde{\boldsymbol{x}}\|^2, \end{aligned}$$

which allows us to use hard MIMO detection methods to approximate the posterior symbol probability. Approximations based on linear receivers and partial marginalizations are also worth noting [8]. Probabilistic data association (PDA) considers some form of Gaussian approximations with MIMO interference, and it approximately computes the $p(x_i|\boldsymbol{y})$'s recursively [10]. Approximate message passing, as suggested by its name, employs message passing to process $p(x_i|\boldsymbol{y})$; it also assumes a large-scale i.i.d. Gaussian $\boldsymbol{H}$ to enable certain approximations [9].

3. VARIATIONAL INFERENCE

As described in the Introduction, the idea of VI is to find an approximate distribution $q(\boldsymbol{x})$ of $p(\boldsymbol{x}|\boldsymbol{y})$ by optimization. We begin by considering the minimization of the Kullback-Leibler divergence:

$$\min_{q\in\mathcal{Q}} \mathbb{E}_{\boldsymbol{x}\sim q}\left[\log\left(\frac{q(\boldsymbol{x})}{p(\boldsymbol{x}|\boldsymbol{y})}\right)\right], \tag{P1}$$

where $\mathcal{Q}$ is the set of distributions whose support is the same as that of $p(\boldsymbol{x})$, which is $\mathcal{S}^N$. By Jensen's inequality, the objective value of (P1) attains its minimum if $q(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{y})$. Now, we want to approximate $p(\boldsymbol{x}|\boldsymbol{y})$ by a $\mathbb{C}^N$-supported dstribution. To this end, we apply a $\mathbb{C}^N$-supported approximation of $p(\boldsymbol{x})$:

$$\hat{p}(\boldsymbol{x}) = \prod_{i=1}^N \hat{p}(x_i) = \prod_{i=1}^N \left[\frac{1}{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|}\mathcal{CN}(x_i; s_j, \delta_i)\right], \tag{5}$$

for a given and small $\delta_i > 0$; $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$; each s_j represents a point in $\mathcal{S}$. Specifically, (5) considers a Gaussian mixture model with respect to each symbol x_i. Plugging (5) into (3) and (P1), we get the following approximation of (P1):

$$\min_{q\in\mathcal{Q}} \mathbb{E}_{\boldsymbol{x}\sim q}\left[\log\left(\frac{q(\boldsymbol{x})}{p(\boldsymbol{y}|\boldsymbol{x})\hat{p}(\boldsymbol{x})}\right)\right], \tag{P2}$$

where $\mathcal{Q}$ is now the set of all $\mathbb{C}^N$-supported continuous distributions, and we omit $p(\boldsymbol{y})$ in (P2) since it does not depend on $\boldsymbol{x}$.

Problem (P2) does not appear to have a tractable objective function due to the expectation. We deal with this by variational approximation. By Jensen's inequality, (5) can be expressed as:

$$\log\hat{p}(x_i) = \max_{\boldsymbol{w}_i\in\Delta} g(\boldsymbol{w}_i, x_i), \tag{6}$$

where

$$g(\boldsymbol{w}_i, x_i) = \sum_{j=1}^{|\mathcal{S}|} w_{ij}\log\left(\frac{\mathcal{CN}(x_i; s_j, \delta_i)}{w_{ij}}\right) + \log\frac{1}{|\mathcal{S}|},$$

$$\Delta = \left\{\boldsymbol{\omega}\in\mathbb{R}_{++}^{|\mathcal{S}|} \mid \mathbf{1}^T\boldsymbol{\omega} = 1\right\}.$$

It is worth noting that (6) resembles expectation maximization for Gaussian mixture models [23]. Also, as a key step, we have

$$\mathbb{E}_{\boldsymbol{x}\sim q}[\max_{\boldsymbol{w}_i\in\Delta} g(\boldsymbol{w}_i, x_i)] \geq \max_{\boldsymbol{w}_i\in\Delta}\mathbb{E}_{\boldsymbol{x}\sim q}[g(\boldsymbol{w}_i, x_i)]. \tag{7}$$

Applying (7) to (P2), we obtain the following upper-bound approximation of (P2):

$$\min_{\substack{q\in\mathcal{Q},\\ \boldsymbol{w}_i\in\Delta, i=1,\ldots,N}} f\left(q, \boldsymbol{W}\right), \tag{P3}$$

where

$$\begin{aligned} &\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_N]; \\ &f\left(q, \boldsymbol{W}\right) = \mathbb{E}_{\boldsymbol{x}\sim q}\left[\log q(\boldsymbol{x}) - \sum_{i=1}^N g(\boldsymbol{w}_i, x_i) - \log p(\boldsymbol{y}|\boldsymbol{x})\right]. \end{aligned} \tag{8}$$

Problem (P3) is our main problem to tackle, and it has some nice features in algorithm design. First, recall that $p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{CN}\left(\boldsymbol{y}; \boldsymbol{H}\boldsymbol{x}, \sigma^2\boldsymbol{I}\right)$. It can be shown that, given $\boldsymbol{W}$,

$$f\left(q, \boldsymbol{W}\right) \propto \mathbb{E}_{\boldsymbol{x}\sim q}\left[\log\left(\frac{q(\boldsymbol{x})}{\mathcal{CN}\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)}\right)\right], \tag{9}$$

where

$$\boldsymbol{\Sigma} = \left(\frac{1}{\sigma^2}\boldsymbol{H}^H\boldsymbol{H} + \mathrm{Diag}\left(\frac{1}{\delta_1}, \ldots, \frac{1}{\delta_N}\right)\right)^{-1}; \tag{10a}$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}\left(\boldsymbol{\alpha} + \frac{1}{\sigma^2}\boldsymbol{H}^H\boldsymbol{y}\right); \tag{10b}$$

$$\boldsymbol{\alpha} = \left(\frac{1}{\delta_1}\boldsymbol{s}^T\boldsymbol{w}_1, \ldots, \frac{1}{\delta_N}\boldsymbol{s}^T\boldsymbol{w}_N\right)^T; \tag{10c}$$

$\mathrm{Diag}(\boldsymbol{x})$ denotes a diagonal matrix with diagonal elements given by $x_1, x_2, \ldots, x_N$; $\boldsymbol{s} = [s_1, \ldots, s_{|\mathcal{S}|}]^T$ lists all the points in $\mathcal{S}$. By Jensen's inequality, the minimization of f over $q\in\mathcal{Q}$ and given $\boldsymbol{W}$ is:

$$q(\boldsymbol{x}) = \mathcal{CN}\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \tag{11}$$

Second, the minimization of f over $\boldsymbol{W}$ (with $\boldsymbol{w}_i \in \Delta$ for all i) amounts to

$$\min_{\boldsymbol{w}_i\in\Delta} -\sum_{i=1}^N \mathbb{E}_{\boldsymbol{x}\sim q}\big[g(\boldsymbol{w}_i, x_i)\big], \;\; i = 1, 2, \ldots, N. \tag{12}$$

Given $q(\boldsymbol{x}) = \mathcal{CN}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, we have

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{x}\sim q}\big[g(\boldsymbol{w}_i, x_i)\big] \\ &\propto \mathbb{E}_{\boldsymbol{x}\sim q}\left[\sum_{j=1}^{|\mathcal{S}|} w_{ij}\log\left(\frac{\mathcal{CN}(x_i; s_j, \delta_i)}{w_{ij}}\right)\right]. \end{aligned} \tag{13}$$

By Jensen's inequality, the solutions to (12) are

$$w_{ij} = \frac{\exp\left(-\frac{1}{\delta_i}|\mu_i - s_j|^2\right)}{\sum_{k=1}^{|\mathcal{S}|}\exp\left(-\frac{1}{\delta_i}|\mu_i - s_k|^2\right)}, \;\; \forall i, \; j. \tag{14}$$

The above results suggest that we can optimize (P3) by alternating minimization — one time for q and one time for $\boldsymbol{W}$. By doing so, we have closed-form expressions for each alternating minimization step. This alternating minimization algorithm is summarized in Algorithm 1. Note that computing $\boldsymbol{\Sigma}$ requires matrix inversion, which is the most computationally expensive step in the algorithm. It is interesting to observe that line 6 of Algorithm 1 resembles the MMSE detector.

Algorithm 1 Alternating Optimization of (P3)

Input:

$\boldsymbol{\mu}^{(0)} \in \mathbb{C}^N$, the initial mean of $p(\boldsymbol{x})$.

$\boldsymbol{\delta} \in \mathbb{R}_+^N$, Gaussian approximation parameters.

1: $\boldsymbol{\Sigma} = \left(\frac{1}{\sigma^2}\boldsymbol{H}^H\boldsymbol{H} + \mathrm{Diag}\left(\frac{1}{\delta_1}, ..., \frac{1}{\delta_N}\right)\right)^{-1}$.

2: $t = 0$.

3: **repeat**

4: $\quad w_{ij}^{(t+1)} = \frac{\exp\left(-\frac{1}{\delta_i}|\mu_i^{(t)} - s_j|^2\right)}{\sum_{k=1}^{|\mathcal{S}|}\exp\left(-\frac{1}{\delta_i}|\mu_i^{(t)} - s_k|^2\right)}$, for all i, j.

5: $\quad \boldsymbol{\alpha}^{(t+1)} = \left(\frac{1}{\delta_1}\boldsymbol{s}^T\boldsymbol{w}_1^{(t+1)}, ..., \frac{1}{\delta_N}\boldsymbol{s}^T\boldsymbol{w}_N^{(t+1)}\right)^T$.

6: $\quad \boldsymbol{\mu}^{(t+1)} = \boldsymbol{\Sigma}\left(\boldsymbol{\alpha}^{(t+1)} + \frac{1}{\sigma^2}\boldsymbol{H}^H\boldsymbol{y}\right)$.

7: $\quad t = t + 1$.

8: **until** some stopping criterion is satisfied.

Output: $\boldsymbol{\mu}^{(t-1)}$.

4. PRACTICAL IMPLEMENTATION

We have one remaining issue: the choice of the δ_i's. On one hand, we want δ_i's to be as small as possible such that $\hat{p}(\boldsymbol{x})$ is nearly the same as $p(\boldsymbol{x})$. On the other hand, having the δ_i's too small makes the problem more ill-conditioned as suggested by our empirical experience. We tackle this issue by adopting the idea of homotopy optimization [24, 25]. We start with large δ_i's and run Algorithm 1 to get a solution; then, we gradually reduce δ_i's and run Algorithm 1 again, warm-started by the previous solution. This process goes on until the δ_i's are small enough. The algorithm with homotopy optimization is shown in Algorithm 2. There, the fifth line is the rule of updating the δ_i's in each iteration. The intuition behind this is that we may change δ_i according to the distance between an estimated symbol value and a nearby constellation point: the closer the distance, the smaller increment of δ_i is. Our simulation results suggest that this idea works well in practice. We call our method as Variational Inference of Posterior (VIP).

Algorithm 2 Variational Inference of Posterior $p(x_i|\boldsymbol{y})$

Input:

$\boldsymbol{\mu}^{(0)}, \boldsymbol{\delta}^{(0)}$, and momentum β.

1: $t = 0, \tau = 0$.

2: **repeat**

3: $\quad$ Run Algorithm 1 with input $\left(\boldsymbol{\mu}^{(\tau)}, \boldsymbol{\delta}^{(\tau)}\right)$ and get output $\boldsymbol{\mu}$.

4: $\quad \boldsymbol{\mu}^{(\tau+1)} = \beta\boldsymbol{\mu} + (1-\beta)\boldsymbol{\mu}^{(\tau)}$.

5: $\quad \frac{1}{\delta_i^{(\tau+1)}} = \frac{1}{\delta_i^{(\tau)}} + \left(\frac{\left|s_{k_i^\star}\right| - \left|\mu_i - s_{k_i^\star}\right|}{\left|s_{k_i^\star}\right|}\right)^2$,

$\quad$ where $k_i^\star = \arg\min_{j \in |\mathcal{S}|} |\mu_i - s_j|$.

6: $\quad \tau = \tau + 1$.

7: **until** some stopping criterion is satisfied.

Output: $\boldsymbol{\mu}^{(\tau-1)}, \boldsymbol{\Sigma}^{(\tau-1)}$.

In Algorithm 2, we need to calculate $\boldsymbol{\Sigma}$ which involves matrix inversion. In the massive MIMO regime where N is large, computing $\boldsymbol{\Sigma}$ is expensive. To alleviate the computational burdens, we propose the following modification. Let $\delta_i = \delta$ be identical for all i. Then, the update of $\boldsymbol{\Sigma}$ in (10a) can be done by

$$\boldsymbol{\Sigma} = \boldsymbol{V}\left(\frac{1}{\sigma^2}\boldsymbol{\Lambda} + \delta\boldsymbol{I}\right)^{-1}\boldsymbol{V}^H, \tag{15}$$

where $\boldsymbol{H}^H\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H$ is the eigenvalue decomposition of $\boldsymbol{H}^H\boldsymbol{H}$ and can be pre-computed. Then, in each iteration, we only need to compute the inverse of a diagonal matrix. We use the following modification of the updating rule to make all δ_i identical:

$$\frac{1}{\delta^{(\tau+1)}} = \frac{1}{\delta^{(\tau)}} + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|s_{k_i^\star}| - |\mu_i - s_{k_i^\star}|}{|s_{k_i^\star}|}\right)^2. \tag{16}$$

We call this simplified VIP method SVIP.

5. SIMULATIONS

In this section, we provide simulation results to evaluate the performance of the proposed algorithms. We adopt the symbol-error rate (SER) as the performance metric. We compare our methods, VIP and SVIP, with the lattice reduction-aided MMSE DF detector (MMSE LRA-DF) [5], an AMP-based detector called LAMA [9], and the bound constrained semidefinite relaxation method (BC-SDR) [26]. We also show the performance when there is no MIMO interference, as a benchmark. The signal-to-noise ratio (SNR) is defined as $\mathbb{E}\left[\|\boldsymbol{H}\boldsymbol{x}\|_2^2\right] / \mathbb{E}\left[\|\boldsymbol{n}\|_2^2\right]$. All the results are obtained by averaging over 5,000 Monte-Carlo trials. In each trial, the channel is independently and randomly generated and so do the transmitted symbols.

The settings of our algorithms are as follows. The momentum factor β in Algorithm 2 is set as 0.7. The vector $\boldsymbol{\mu}$ is initialized as $\boldsymbol{0}$, and we set $\delta_i^{(0)} = 50$ for all i.

5.1. i.i.d. Gaussian Channels

We first consider simulations based on i.i.d. complex circular Gaussian $\boldsymbol{H}$, with zero mean and unit variance. Figs. 1, 2, and 3 show the SER performance, where the receive dimension is $M = 60$ and transmit dimension is $N = 50$. It is seen that VIP, SVIP and BC-SDR achieve comparable SER performance; they are close to the no-interference lower bound with no more than 3dB SNR gap at the SER level of 10^{-5}. MMSE LRA-DF performs worse than these three methods. LAMA works well when the SNR is low to medium, but becomes unstable for high SNRs.

Next, we test the algorithms under a more challenging setting in Fig. 4, where the channel matrix is square and has a larger dimension, namely, $M = N = 100$. In this case, VIP and SVIP slightly outperform BC-SDR. LAMA has similar performance trends.

The runtime performance of the different algorithms is summarized in Table 1; the SNR is fixed as 20dB. First, it is seen that although VIP, SVIP and BC-SDR achieve comparable SER performance in Fig. 1, VIP and SVIP are much faster than BC-SDR. Second, with the modification (15), SVIP saves approximately 25% of the runtime compared to VIP. Third, LAMA is also computationally very efficient, and is faster than VIP and SVIP.

Table 1. Average runtime performance (in sec.), SNR=20dB

Channel Size (M, N)	$(50, 40)$	$(60, 50)$	$(70, 60)$
MMSE LRA-DF	0.0005	0.0006	0.0007
LAMA	0.0011	0.0014	0.0015
BC-SDR	0.2014	0.3024	0.4370
VIP	0.0210	0.0236	0.0314
SVIP	0.0154	0.0171	0.0201

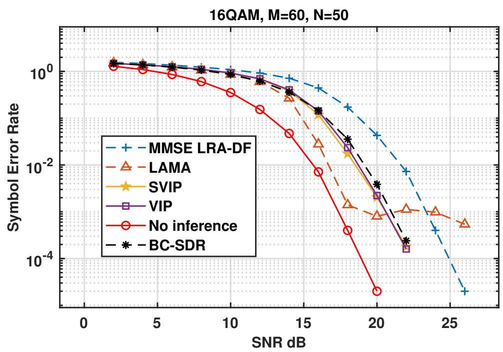

Fig. 1. Symbol error rate under i.i.d Gaussian channel, 16-QAM.

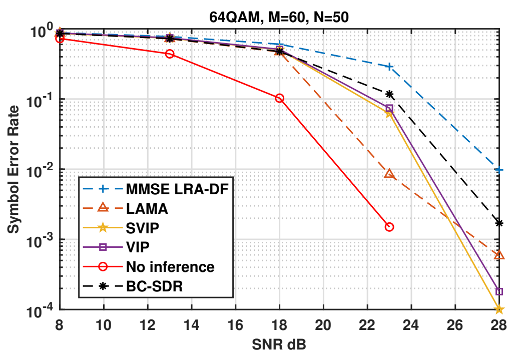

Fig. 2. Symbol error rate under i.i.d Gaussian channel, 64-QAM.

5.2. Multi-path Channels

Next, we test the algorithms under multi-path channels. The channel matrix is generated according to

$$\mathbf{H} = \sqrt{1-\varepsilon}\mathbf{H}_m + \sqrt{\varepsilon}\boldsymbol{H}_g, \tag{17}$$

where $\boldsymbol{H}_g$ follows the i.i.d complex circular Gaussian channel model;

$$\mathbf{H}_m = \frac{1}{N_P}\sum_{\ell=1}^{N_\mathrm{P}} \alpha_\ell \mathbf{a}_\mathrm{R}\left(\theta_{\mathrm{R},\ell}\right)\mathbf{a}_\mathrm{T}^H\left(\theta_{\mathrm{T},\ell}\right),$$

follows a multi-path channel model. Herein, N_p is the number of paths and we set $N_\mathrm{p} = 10$ in the simulation; the complex channel

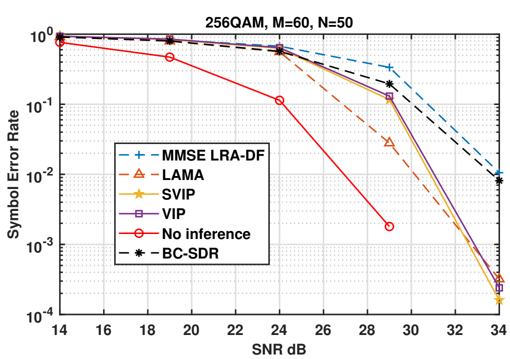

Fig. 3. Symbol error rate under i.i.d Gaussian channel, 256-QAM.

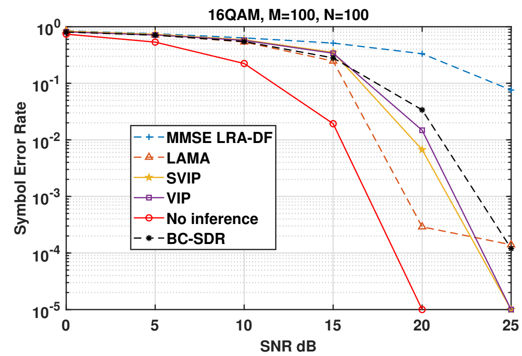

Fig. 4. Symbol error rate under large-size i.i.d Gaussian channel.

gains α_ℓ's are zero-mean unit-variance complex circular Gaussian; and $\mathbf{a}_\mathrm{R}\left(\theta_{\mathrm{R},\ell}\right), \mathbf{a}_\mathrm{T}\left(\theta_{\mathrm{T},\ell}\right)$ are given by

$$\mathbf{a}(\theta) = \left[1, e^{-j\frac{2\pi d}{\lambda}\sin\theta}, \ldots, e^{-j\frac{2\pi d}{\lambda}\sin\theta(N-1)}\right]^T,$$

with $d = \lambda/2$; $\theta_{\mathrm{R},\ell}$ and $\theta_{\mathrm{T},\ell}$ are uniformly sampled from $[-\pi/2, \pi/2]$.

The simulation results are presented in Fig.5; we set $\varepsilon = 0.2$ in (17). VIP and BC-SDR show comparable performance, while SVIP performs worse in the low SNR regime. LAMA in this case is not able to yield reasonable performance; this may be due to the fact that LAMA assumes i.i.d. Gaussian channels, and model mismatch happens here.

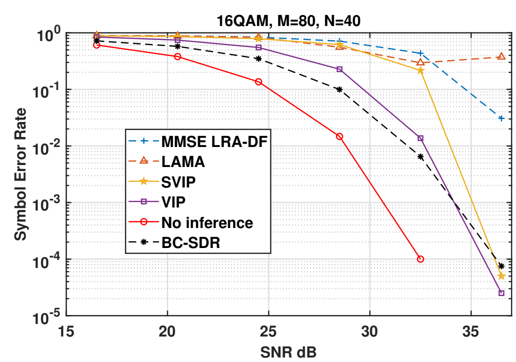

Fig. 5. Symbol error rate under multi-path channels.

6. CONCLUSION

In this paper, we studied the application of VI for approximate posterior inference in MIMO detection. The result is an iterative MMSE-like detector, and simulation results show that the VI-based detector works reasonably well for large-scale MIMO detection with higher-order QAM constellations. The current endeavor largely focused on the demonstration of the VI idea in MIMO detection. As future work, it would be interesting to explore and draw connections between the VI approach and the state-of-art methods in soft MIMO detection.

7. REFERENCES

[1] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, 1999.

[2] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.

[3] P. H. Tan and L. K. Rasmussen, "The application of semidefinite programming for detection in CDMA," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1442–1449, 2001.

[4] W.-K. Ma, T. N. Davidson, K. M. Wong, Z.-Q. Luo, and P.-C. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912–922, 2002.

[5] D. Wübben, D. Seethaler, J. Jaldén, and G. Matz, "Lattice reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, 2011.

[6] C. Windpassinger and R. F. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Inf. Theory Workshop*, pp. 345–348, IEEE, 2003.

[7] B. M. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, 2003.

[8] E. G. Larsson and J. Jaldén, "Fixed-complexity soft MIMO detection via partial marginalization," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3397–3407, 2008.

[9] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimal data detection in large MIMO," *arXiv preprint arXiv:1811.01917*, 2018.

[10] J. Luo, K. R. Pattipati, P. K. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 361–363, 2001.

[11] P. Fertl, J. Jaldén, and G. Matz, "Performance assessment of MIMO-BICM demodulators based on mutual information," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1366–1382, 2011.

[12] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, 2019.

[13] N. Shlezinger, R. Fu, and Y. C. Eldar, "Deepsic: Deep soft interference cancellation for multiuser MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1349–1362, 2020.

[14] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, 2020.

[15] M. Shao and W.-K. Ma, "Binary MIMO detection via homotopy optimization and its deep adaptation," *IEEE Trans. Signal. Process.*, vol. 69, pp. 781–796, 2020.

[16] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine Learning*, vol. 37, no. 2, pp. 183–233, 1999.

[17] H. Attias, "A variational baysian framework for graphical models," in *Proc. Neural Inf. Process. Syst.*, vol. 12, 1999.

[18] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Statist. Assoc.*, vol. 112, no. 518, pp. 859–877, 2017.

[19] J. J. Sakurai and E. D. Commins, *Modern Quantum Mechanics*. American Association of Physics Teachers, 1995.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[21] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Int. Conf. on Learn. Representations*, 2013.

[22] D. Seethaler, G. Matz, and F. Hlawatsch, "An efficient MMSE-based demodulator for MIMO bit-interleaved coded modulation," in *Proc. IEEE Global Telecommun. Conf.*, vol. 4, pp. 2455–2459, 2004.

[23] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[24] D. M. Dunlavy and D. P. O'Leary, "Homotopy optimization methods for global optimization," *Report SAND2005-7495, Sandia National Laboratories*, 2005.

[25] E. L. Allgower and K. Georg, *Introduction to Numerical Continuation Methods*. SIAM, 2003.

[26] N. D. Sidiropoulos and Z.-Q. Luo, "A semidefinite relaxation approach to MIMO detection for high-order QAM constellations," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 525–528, 2006.