A Hybrid RTS-BP Algorithm for Improved Detection of Large-MIMO M-QAM Signals

Tanumay Datta, N. Srinidhi, A. Chockalingam, and B. Sundar Rajan
Department of ECE, Indian Institute of Science, Bangalore 560012, INDIA

Abstract—Low-complexity near-optimal detection of large-MIMO signals has attracted recent research. Recently, we proposed a local neighborhood search algorithm, namely reactive tabu search (RTS) algorithm, as well as a factor-graph based belief propagation (BP) algorithm for low-complexity large-MIMO detection. The motivation for the present work arises from the following two observations on the above two algorithms: i) Although RTS achieved close to optimal performance for 4-QAM in large dimensions, significant performance improvement was still possible for higher-order QAM (e.g., 16-, 64-QAM). ii) BP also achieved near-optimal performance for large dimensions, but only for \(|\pm1|\) alphabet. In this paper, we improve the large-MIMO detection performance of higher-order QAM signals by using a hybrid algorithm that employs RTS and BP. In particular, motivated by the observation that when a detection error occurs at the RTS output, the least significant bits (LSB) of the symbols are mostly in error, we propose to first reconstruct and cancel the interference due to bits other than LSBS at the RTS output and feed the interference cancelled received signal to the BP algorithm to improve the reliability of the LSBS. The output of the BP is then fed back to RTS for the next iteration. Simulation results show that the proposed algorithm performs better than the RTS algorithm, and semi-definite relaxation (SDR) and Gaussian tree approximation (GTA) algorithms.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) systems with large number (e.g., tens) of transmit and receive antennas, referred to as ‘large-MIMO systems,’ are of interest because of the high capacities/spectral efficiencies theoretically predicted in these systems [1]. Research in low-complexity receive processing (e.g., MIMO detection) techniques that can lead to practical realization of large-MIMO systems is both nascent as well as promising. Evolution of WiFi standards (evolution from IEEE 802.11n to IEEE 802.11ac to achieve multi-gigabit rate transmissions in 5 GHz band) now considers 16 × 16 MIMO operation; see 16 × 16 MIMO indoor channel sounding measurements at 5 GHz reported in [2] for consideration in WiFi standards.

In the context of large-MIMO signal detection, in our recent works, we have shown that certain algorithms from machine learning/artificial intelligence achieve near-optimal detection performance in large-MIMO systems at low complexities [3]-[9]. In [3]-[5], a local neighborhood search based algorithm, namely, a likelihood ascent search (LAS) algorithm, was proposed and shown to achieve close to maximum-likelihood (ML) performance in MIMO systems with several tens of antennas (e.g., 32 × 32, 64 × 64 MIMO). Subsequently, in [6],[7], another local search algorithm, namely, reactive tabu search (RTS) algorithm, which performed better than LAS through the use of a local minima exit strategy was presented.

In [8], near-ML performance in a 50 × 50 MIMO system was demonstrated using a Gibbs sampling based detection algorithm, where the symbols take values from \({\pm1}\). More recently, we, in [9], proposed a factor graph based belief propagation (BP) algorithm for large-MIMO detection, where we adopted a Gaussian approximation of the interference (GAI).

The motivation for the present work arises from the following two observations on the RTS and BP algorithms in [6],[7] and [9]: i) RTS works for general M-QAM. Although RTS was shown to achieve close to ML performance for 4-QAM in large dimensions, significant performance improvement was still possible for higher-order QAM (e.g., 16- and 64-QAM). ii) BP also was shown to achieve near-optimal performance for large dimensions, but only for \({\pm1}\) alphabet. In this paper, we improve the large-MIMO detection performance of higher-order QAM signals by using a hybrid algorithm that employs RTS and BP. In particular, we observed that when a detection error occurs at the RTS output, the least significant bits (LSB) of the symbols are mostly in error. Motivated by this observation, we propose to first reconstruct and cancel the interference due to bits other than the LSBS at the RTS output and feed the interference cancelled received signal to the BP algorithm to improve the reliability of the LSBS. The output of the BP is then fed back to the RTS for the next iteration. Simulation results show that the proposed algorithm performs better RTS, and semi-definite relaxation (SDR) and Gaussian tree approximation (GTA) algorithms in [14],[15].

II. RTS AND BP ALGORITHMS FOR LARGE-MIMO DETECTION

Consider a \(N_r \times N_t\) V-BLAST MIMO system whose received signal vector, \(y_c \in \mathbb{C}^{N_r}\), is of the form

\[
y_c = H_c x_c + n_c, \tag{1}
\]

where \(x_c \in \mathbb{C}^{N_t}\) is the symbol vector transmitted, \(H_c \in \mathbb{C}^{N_r \times N_t}\) is the channel gain matrix, and \(n_c \in \mathbb{C}^{N_r}\) is the noise vector whose entries are modeled as i.i.d \(\mathcal{CN}(0,\sigma^2)\). Assuming rich scattering, we model the entries of \(H_c\) as i.i.d

---

1 Similar algorithms have been reported earlier in the context of multiuser detection in large CDMA systems.

2 In [5],[7], we compared the performance and complexities of LAS and RTS algorithms with those of the sphere decoding (SD) variants in [10] and [11], and showed that, while LAS and RTS scale well for large dimensions, these SD variants do not scale well for large dimensions.
\( \mathcal{CN}(0, 1) \). Each element of \( \mathbf{x} \)\( \mathbf{c} \) is an \( M \)-PAM or \( M \)-QAM symbol. \( M \)-PAM symbols take values from \( \{ A_m, m = 1, 2, \cdots, M \} \), where \( A_m = (2m - 1 - M) \), and \( M \)-QAM is nothing but two PAMs in quadrature. As in [4], we convert (1) into a real-valued system model, given by

\[
\mathbf{y} = \mathbf{Hx} + \mathbf{n},
\]

where \( \mathbf{H} \in \mathbb{R}^{2N_t \times 2N_r}, \mathbf{y} \in \mathbb{R}^{2N_r}, \mathbf{x} \in \mathbb{R}^{2N_t}, \mathbf{n} \in \mathbb{R}^{2N_r} \).

For \( M \)-QAM, \( \{ x_1, \cdots, x_{2N_t} \} \) can viewed to be from an underlying \( M \)-PAM signal set, and so is \( \{ x_{N_t+1}, \cdots, x_{2N_t} \} \). Let \( \mathcal{A}_i \) denote the \( M \)-PAM signal set from which \( x_i \) takes values, \( i = 1, 2, \cdots, 2N_t \). Defining a \( 2N_t \)-dimensional signal space \( \mathcal{S} \) to be the Cartesian product of \( \mathcal{A}_1 \) to \( \mathcal{A}_{2N_t} \), the ML solution vector, \( \mathbf{x}_{ML} \), is given by

\[
\mathbf{x}_{ML} = \arg \min_{\mathbf{x} \in \mathcal{S}} \| \mathbf{y} - \mathbf{Hx} \|^2,
\]

whose complexity is exponential in \( N_t \). The RTS algorithm in [6],[7] is a low-complexity algorithm, which minimizes the ML metric in (3) through a local neighborhood search.

A. RTS Algorithm

A detailed description of the RTS algorithm for large-MIMO detection is available in [6],[7]. The per-symbol complexity of RTS for detection of \( M \)-PAM signals is \( O(N_t N_r) \). Here, we present the 16- and 64-QAM performance of the RTS algorithm that motivates the current work.

1) Motivation of Current Work: Figure 1 shows the uncoded BER performance of RTS using the algorithm parameters optimized through simulations for 4-, 16-, and 64-QAM in a 32 \( \times \) 32 V-BLAST system. As lower bounds on the error performance in MIMO, the SISO AWGN performance for 4-, 16-, and 64-QAM are also plotted. It can be seen that, in the case of 4-QAM, the RTS performance is just about 0.5 dB away from the SISO AWGN performance at 10\(^{-3} \) BER. However, the gap between RTS performance and SISO AWGN performance at 10\(^{-3} \) BER widens for 16-QAM and 64-QAM; the gap is 7.5 dB for 16-QAM and 16.5 dB for 64-QAM. This gap can be viewed as a potential indicator of the amount of improvement in performance possible further. A more appropriate indicator will be the gap between RTS performance and the ML performance. Since simulation of sphere decoding (SD) of 32 \( \times \) 32 V-BLAST with 16- and 64-QAM (64 real dimensions) is computationally intensive, we do not show the SD (ML) performance. Nevertheless, the widening gap of RTS Performance from SISO AWGN performance for 16- and 64-QAM seen in Fig. 1 motivated us to explore improved algorithms to achieve better performance than RTS performance for higher-order QAM.

B. BP Algorithm Based on GAI

In [9], we presented a detection algorithm based on BP on factor graphs of MIMO systems. In (2), each entry of the vector \( \mathbf{y} \) is treated as a function node (observation node), and each symbol, \( x_i \in \{ \pm 1 \} \), as a variable node. A key ingredient in the BP algorithm in [9], which contributes to its low complexity, is the Gaussian approximation of interference (GAI), where the interference + noise term, \( z_{ik} \), in

\[
y_i = h_{ik} x_k + \sum_{j=1,j\neq k}^{2N_t} h_{ij} x_j + n_i,
\]

is modeled as \( \mathcal{CN}(\mu_{z_{ik}}, \sigma^2_{z_{ik}}) \) with

\[
\mu_{z_{ik}} = \sum_{j=1,j\neq k}^{N_r} h_{ij} \mathbb{E}(x_j),
\]

and

\[
\sigma^2_{z_{ik}} = \sum_{j=1,j\neq k}^{2N_t} |h_{ij}|^2 \text{Var}(x_j) + \frac{\sigma^2}{2},
\]

where \( h_{ij} \) is the \( (i,j) \)th element in \( \mathbf{H} \). With \( x_i \)’s \( \in \{ \pm 1 \} \), the log-likelihood ratio (LLR) of \( x_k \) at observation node \( i \), denoted by \( \Lambda_i^k \), is

\[
\Lambda_i^k = \log \left( \frac{p(y_i | \mathbf{H}, x_k = 1)}{p(y_i | \mathbf{H}, x_k = -1)} \right)
= \frac{2}{\sigma^2_{z_{ik}}} \Re \left( h_{ik}^* (y_i - \mu_{z_{ik}}) \right).
\]

The LLR values computed at the observation nodes are passed to the variable nodes. Using these LLRs, the variable nodes compute the probabilities

\[
p_i^{k+} = p_i(x_k = +1 | \mathbf{y}) = \frac{\exp(\sum_{i \neq i} \Lambda_i^k)}{1 + \exp(\sum_{i \neq i} \Lambda_i^k)},
\]

and pass them back to the observation nodes. This message passing is carried out for a certain number of iterations. At the end, \( x_k \) is detected as

\[
\hat{x}_k = \text{sgn} \left( \sum_{i=1}^{2N_r} \Lambda_i^k \right).
\]
behavior,’ where the bit error performance improves with increasing number of dimensions. In Fig. 1, the uncoded BER performance of this BP algorithm for 4-QAM (input data vector of size 2N_t with elements from \{±1\}) in 32 × 32 V-BLAST is also plotted. We can see that the performance is almost the same as that of RTS. In terms of complexity, the BP algorithm has the advantage of no need to compute an initial solution vector and \(H^T H\), which is required in RTS. The per-symbol complexity of the BP algorithm for detection in V-BLAST is \(O(N_t)\). A limitation with this BP approach is that it is not for general \(M\)-QAM. However, its good performance with \{±1\} alphabet at lower complexities than RTS can be exploited to improve the higher-order QAM performance of RTS, as proposed in the following section.

III. PROPOSED HYBRID RTS-BP ALGORITHM FOR LARGE-MIMO DETECTION

In this section, we highlight the rationale behind the hybrid RTS-BP approach and present the proposed algorithm.

Why Hybrid RTS-BP?

The proposed hybrid RTS-BP approach is motivated by the the following observation we made in our RTS BER simulations. We observed that, at moderate to high SNRs, when an RTS output vector is in error, the least significant bits (LSB) of the data symbols are more likely to be in error than other bits. An analytical reasoning for this behavior can be given as follows.

Let \(\mathbf{x}\) be the transmit vector and \(\hat{\mathbf{x}}\) be the corresponding output of the RTS detector. Let \(A = \{a_1, a_2, \ldots, a_M\}\) denote the \(M\)-PAM alphabet that \(x_i\)'s take values from. Consider the symbol-to-bit mapping, where we can write the value of each entry of \(\hat{\mathbf{x}}\) as a linear combination of its constituent bits as

\[
\hat{x}_i = \sum_{j=0}^{N-1} 2^j \hat{g}^{(j)}_i, \quad i = 1, \ldots, 2N_t, \tag{8}
\]

where \(N = \log_2 M\) and \(\hat{g}^{(j)}_i \in \{±1\}\). We note that the RTS algorithm outputs a local minima as the solution vector. So, \(\hat{\mathbf{x}}\), being a local minima, satisfies the following conditions:

\[
\|\mathbf{y} - H\hat{\mathbf{x}}\|^2 \leq \|\mathbf{y} - H(\hat{\mathbf{x}} + \lambda_i \mathbf{e}_i)\|^2, \quad \forall i = 1, \ldots, 2N_t, \tag{9}
\]

where \(\lambda_i = (a_q - \hat{x}_i), q = 1, \ldots, M\), and \(\mathbf{e}_i\) denotes the \(i\)th column of the identity matrix. Defining \(\mathbf{F} = H^T H\), \(\mathbf{r} = H\hat{\mathbf{x}}\), and denoting the \(i\)th column of \(\mathbf{H}\) as \(\mathbf{h}_i\), the conditions in (9) reduce to

\[
2\lambda_i \mathbf{h}_i^T \mathbf{h}_i \leq 2\lambda_i r^T \mathbf{h}_i + \lambda_i^2 f_{ii}, \tag{10}
\]

where \(f_{ij}\) denotes the \((i, j)\)th element of \(\mathbf{F}\). Under moderate to high SNR conditions, ignoring the noise, (10) can be further reduced to

\[
2(\mathbf{x} - \hat{\mathbf{x}})^T f_i \text{sgn}(\lambda_i) \leq \lambda_i f_{ii} \text{sgn}(\lambda_i), \tag{11}
\]

where \(f_i\) denotes the \(i\)th column of \(\mathbf{F}\). For Rayleigh fading, \(f_{ii}\) is chi-square distributed with \(2N_t\) degrees of freedom with mean \(N_t\). Approximating the distribution of \(f_{ij}\) to be normal with mean zero and variance \(\frac{N_t}{N_t - 1}\), we can drop \(\text{sgn}(\lambda_i)\) in (11). Using the fact that the minimum value of \(|\lambda_i|\) is 2, (11) can be simplified as

\[
\sum_{x_j \neq \hat{x}_j} \Delta_j f_{ij} \leq f_{ii}, \tag{12}
\]

where \(\Delta_j = x_j - \hat{x}_j\). Also, if \(x_i = \hat{x}_i\), by the normal approximation in the above

\[
\sum_{x_j \neq \hat{x}_j} \Delta_j f_{ij} \sim \mathcal{N}\left(0, \frac{N_t}{4} \sum_{x_j \neq \hat{x}_j} \Delta_j^2\right). \tag{13}
\]

Now, the LHS in (12) being normal with variance proportional to \(\Delta_j^2\) and the RHS being positive, it can be seen that \(\Delta_j, \forall i\) take smaller values with higher probability. Hence, the symbols of \(\hat{\mathbf{x}}\) are nearest Euclidean neighbors of their corresponding symbols of the global minima with high probability.\(^3\)

Now, because of the symbol-to-bit mapping in (8), \(\hat{x}_i\) will differ from its nearest Euclidean neighbors certainly in the LSB position, and may or may not differ in other bit positions. Consequently, the LSBs of the symbols in the RTS output \(\hat{\mathbf{x}}\) are least reliable.

The above observation then led us to consider improving the reliability of the LSBs of the RTS output using the BP algorithm in [9], and iterate between RTS and BP as follows.

**Proposed Hybrid RTS-BP Algorithm:**

Figure 2 shows the block schematic of the proposed hybrid RTS-BP algorithm. The following four steps constitute the proposed algorithm.

- **Step 1:** Obtain \(\hat{\mathbf{x}}\) using the RTS algorithm. Obtain the output bits \(\hat{b}^{(j)}_i, i = 1, \ldots, 2N_t, j = 0, \ldots, N - 1\), from \(\hat{\mathbf{x}}\) and \(\mathbf{y}\).
- **Step 2:** Using the \(\hat{b}^{(j)}_i\)'s from Step 1, reconstruct the interference from all bits other than the LSBs (i.e., interference from all bits other than \(\hat{b}^{(0)}_i\)'s) as

\[
\mathbf{I} = \sum_{j=1}^{N-1} 2^j H \hat{b}^{(j)}, \tag{14}
\]

where \(\hat{b}^{(j)} = [\hat{b}^{(j)}_1, \hat{b}^{(j)}_2, \ldots, \hat{b}^{(j)}_{2N_t}]^T\). Cancel the reconstructed interference in (14) from \(\mathbf{y}\) as

\[
\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{I}. \tag{15}
\]

- **Step 3:** Run the BP-GAI algorithm in Sec. II-B on the vector \(\tilde{\mathbf{y}}\) in Step 2, and obtain an estimate of the

\(^3\)Because \(x_i\)'s and \(\hat{x}_i\)'s take values from \(M\)-PAM alphabet, \(\hat{x}_i\) is said to be the Euclidean nearest neighbor of \(x_i\) if \(|x_i - \hat{x}_i| = 2\).
IV. PERFORMANCE OF THE HYBRID RTS-BP DETECTOR

Here, we present the uncoded and coded BER performance of the proposed RTS-BP algorithm evaluated through simulations. Perfect knowledge of $H$ is assumed at the receiver.

Performance in large V-BLAST Systems: Figure 3 shows the uncoded BER performance of $32 \times 32$ V-BLAST with 16- and 64-QAM. Performance of both RTS-BP as well as RTS are shown. It can be seen that, at an uncoded BER of $10^{-3}$, RTS-BP performs better than RTS by about 3.6 dB for 64-QAM and by about 1.6 dB for 16-QAM. This illustrates the effectiveness of the proposed hybrid RTS-BP approach. Also, this improvement in uncoded BER is found to result in improved coded BER as well, as illustrated in Fig. 4. In Fig. 4, we have plotted the turbo coded BER of RTS-BP and RTS in $32 \times 32$ V-BLAST with 64-QAM for rate-1/2 (96 bps/Hz) and rate-3/4 (144 bps/Hz) turbo codes. It can be seen that, at a coded BER of $3 \times 10^{-4}$, RTS-BP performs better than RTS by about 1.5 dB at 96 bps/Hz and by about 2.5 dB at 144 bps/Hz.

Comparison with SDR and GTA algorithms: In Fig. 5, we present a comparison of the symbol error rate (SER) performance the RTS and RTS-BP algorithms with those of other low-complexity detection algorithms employing semi-definite relaxation (SDR) in [14] and Gaussian tree approximation (GTA) in [15] in a $12 \times 12$ V-BLAST MIMO system with 16-QAM. Sphere decoding (SD) and ZF-SIC/MMSE-SIC performances are also shown. We see that, while the performance of SDR and GTA are far from the SD performance, RTS-BP achieves closer to SD performance at a much less complexity than SD.

Performance in frequency-selective large V-BLAST systems: We note that the performance plots in Figs. 3 to 6 are for frequency-flat fading, which could be the fading scenario in...
MIMO-OFDM systems where a frequency-selective fading channel is converted to frequency-flat channels on multiple subcarriers. RTS-BP, RTS, and LAS algorithms, being suited to work well in large dimensions, can be applied to equalize signals in frequency-selective channels in large-MIMO systems. Following the equivalent real-valued system model of the form in (2) for frequency-selective MIMO systems developed in [13], we evaluated the performance of RTS-BP, RTS, and LAS algorithms in $16 \times 16$ V-BLAST with 16-QAM on a frequency selective channel with $L = 6$ equal energy multipath components and $K = 64$ symbols per frame. Figure 7 shows the superior performance of RTS-BP algorithm over RTS and LAS algorithms in this frequency-selective $16 \times 16$ MIMO system with 16-QAM.

V. CONCLUSIONS

We proposed a hybrid algorithm that exploited the good features of the RTS and BP algorithms to achieve improved bit error performance and nearness to capacity performance for $M$-QAM signals in large-MIMO systems at practically affordable low complexities. We illustrated the performance gains of the proposed hybrid approach over the RTS algorithm in flat-fading as well as frequency-selective fading for large V-BLAST and large non-orthogonal STBC MIMO systems. We note (e.g., from the performance plots for 64-QAM in Figs. 1 and 4) that further improvement in performance beyond what is achieved by the proposed hybrid RTS-BP algorithm could be possible. Exploring other algorithms based on this observation, we have proposed other variants of RTS including a layered RTS (LTS) algorithm [16], and a random-restart RTS (R3TS) algorithm [17] which achieved even closer to ML performance than the RTS-BP algorithm.

REFERENCES