Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230776

BP-algorithm based on factor diagram about massive MIMO system

Ling Jiaqian

Southeast University, Nanjing, Jiangsu Province, China

213202717@seu.edu.cn

Abstract. In order to meet the growing demand for high-speed mobile data business, there has been a constant concern for the fifth-generation mobile communication system. 5G large-scale Multiple Input Multiple Output (MIMO) technique has significantly improved the system spectral efficiency and energy efficiency. In order to study the signal detection method under the large-scale MIMO system and improve the existing methods, the article describes the background of signal detection. It introduces the development of the belief propagation (BP) algorithm. There are four main BP algorithms: Original-BP algorithm, RDF-BP algorithm, EBRDF-BP algorithm, and GAI-BP algorithm. One algorithm is proposed based on these fundamental algorithms to improve the whole detection performance. This algorithm reduces complexity while not increasing the error rate too much, which means that it has certain feasibility and practicability.

Keywords: MIMO, factor graph, belief propagation (BP), signal detection

1. Introduction

Multiple input and multiple output system, known as a MIMO system, is an antenna system that uses multiple antennas to increase channel capacity greatly. Both the transmit and receive end use massive antennas, which forms massive channels. Because it can significantly increase the system's data rate and channel capacity, in the future, 5G and subsequent standards will also provide further technical support for the large-scale MIMO system [1], [2]. There is no doubt that the large-scale MMIO will become a critical technique in 5G unlimited communication technology.

Large-scale MIMO system was first proposed by T. L. Marzetta and other researchers in Bell Laboratory in the United States [3], [4]. The study found that negative effects such as Rayleigh fading and additive Gaussian white noise producing can be ignored completely with the increasing number of base station antennas. This situation improves the data transmission rate greatly. In massive MIMO system, base stations are equipped with a large number of antennas [5]. The number of antennas is usually tens, hundreds, or even thousands, which means that it’s a problem referring to the computational complexity.

A detector based on zero-forcing (ZF) filtering is proposed based on the multi-user communication literature [6]. It has relatively low complexity but it is significantly sub-optimal. Maximum a posteriori (MAP) or maximum likelihood (ML) detectors are optimal, but the complexity grows exponentially with an increasing number of transmit antennas. In many cases, it is unacceptable [7]. That means finding a better method that can be computed more easily and be optimal is necessary. Fortunately, J.

DOI: 10.54254/2755-2721/6/20230776

HU proposes an original belief propagation algorithm based on factor graph, known as the Original BP algorithm [7]. Then RDF and EBRDF are proposed to improve the detection performance. But those algorithms are still complicated and hard to apply to large-scale antenna scenarios. P. Som and some other researchers present an approximate way to approximate the multi-antenna interference as Gaussian. The Gaussian approximation of the interference, known as GAI-BP, greatly reduces computational complexity while achieving good results when detecting large-scale MIMO signals [8]. In the same year, the SE-BP detector is proposed and achieved great performance [9]. Based on GAI-BP, the BP detection on the real domain is first proposed, and proposed adaptive BP detection also reduces much complexity in large-scale MIMO [10]. Furthermore, BP algorithm is also connected with deep neural network (DNN). It is demonstrated that DNN improves robustness against various channel conditions with similar complexity and assures high efficiency simultaneously [11].

In this paper, the development of the belief propagation algorithm is sorted. Four BP algorithms: Original-BP algorithm, RDF-BP algorithm, EBRDF-BP algorithm, and GAI-BP algorithm, are mainly introduced. Based on GAI-BP, it presents a way to reduce complexity while maintaining low error rate by judging whether the message converges. Combined with mechanical learning, the judgment method is expected to show greater performance. The rest of this paper is organized as follows: the system model is presented in the next section. Section 3 presents a detailed discussion on various BP algorithms. Section 4 provides a brief description about computational complexity. Section 5 gives simulation results and shows simple analyze. Finally, Section 6 summarizes the paper and concludes the whole work.

2. System model

Firstly, consider using the architecture with $\mathrm{N_t}$ transmit antennas and $\mathrm{N_t}$ receive antennas. Demultiplex the data sequence into $\mathrm{N_t}$ bit streams. Then modulate the bit streams separately and send them to each transmit antenna. After the fading MIMO channel, the channel detector is used to recover the transmitted data on the receiver side. Specifically, the channel model can be expressed as:

$$y = \sqrt{\rho/N_t} Hs + z \tag{1}$$

Here, $\mathrm{s} = \left[\mathrm{s}_0, \dots, \mathrm{s}_{\mathrm{N_t}-1}\right]^{\mathrm{T}}$ is a transmission signal vector with unit energy each antenna. $H = \left[h_0, \dots, h_{N_r-1}\right]^T$ represents channel coefficients (each entry is a complex Gaussian random variable with mean of 0 and variance of 0.5 in each dimension) and it is a $N_r \times N_t$ matrix. h_i is the ith row vector of the matrix H, representing the channel coefficients corresponding to the ith receive antenna. $z = \left[z_0, \dots, z_{N_r-1}\right]^T$ is the Gaussian noise vector with mean of 0 and variance of 0.5 in each dimension. $y = \left[y_0, \dots, y_{N_r-1}\right]^T$ is the receive signal vector. ρ is signal-to-noise ratio (SNR) at each receive antenna. The cardinality of the signal constellation set is $|A| = 2^M$, where M is the number of bits that are allocated by each symbol. Assume the channel is block-fading, which indicates that within a data block, the channel coefficients remain constant and it can change from one block to the next. Usually, take $\rho/\mathrm{Nt} = 1$.

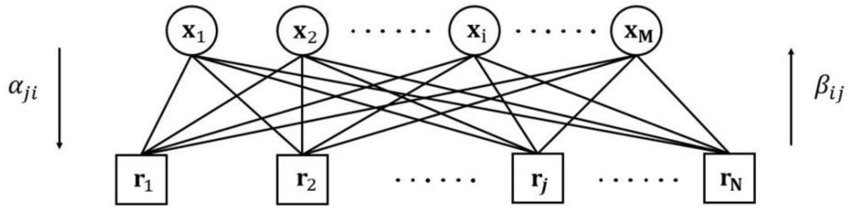

Figure 1. Factor graph in a large-scale MIMO

3. BP algorithm

3.1. Original-BP algorithm

For each iteration, there is two information flow with opposite directions, which is equivalent to two half-iterations. For the first process, the information is passed from the factor node in place, which is also called β information. For the second half of iteration, the information is sent in the opposite direction, and the information is called α information. The specific iteration process is to be introduced and how decision is made is to be explained.

3.1.1. *From factor node to Bit nodes:* This is the first half of the process. On each factor node, the transmit messages of the connected bit nodes are used as a prior information. Calculate each connected bit node's output message (external information) according to the channel model, noise channel observation, and other connected bit nodes. At the Lth iteration, the β message can be expressed as:

$$\begin{aligned}\beta_{ij}^{(l)} \approx \max\nolimits_{s:x_j=1} -|y_i - h_i s|^2 + \sum\nolimits_{k:x_k=1,k\neq j} \alpha_{ki}^{(l-1)} \\ - \max\nolimits_{s:x_j=0} -|y_i - h_i s|^2 + \sum\nolimits_{k:x_k=1,k\neq j} \alpha_{ki}^{(l-1)}\end{aligned} \tag{2}$$

3.1.2. From Bit node to factor node: This process is performed after the first half of the process, which means that the beta information is completely updated. The α message can be expressed as:

$$\begin{gathered}\alpha_{ji}^{(l)} = \sum\nolimits_{k=0,k\neq i}^{N_r-1} \beta_{kj}^{(l-1)}, \\ i = 0,1,\dots,N_r-1, j = 0,1,\dots,MN_t-1.\end{gathered} \tag{3}$$

3.1.3. Soft output calculation: The BP detector is a sort of SISO (Soft Input and Soft Output) detector. After several iterations (here is defined as L), the final soft output calculation is:

$$L(x_j) = \sum_{k=0}^{N_r-1} \beta_{kj}^{(Q_L)}, \quad j = 0,1,\dots,MN_t-1 \tag{4}$$

3.2. RDF algorithm

Based on the belief propagation idea, the RDF-BP algorithm is then proposed. In this algorithm, the same constraints are applied to all factor nodes. Fix each factor node r_j and sort N_t channel coefficients corresponding to the factor nodes according to the absolute values and select d_f larger coefficients. At the same time, the bit nodes and the factor nodes are disconnected. Consider the jth factor node r_j. Define

$\psi_i = (J_{i0}, J_{i1}, \dots, J_{iN_t-1})$ as the index set of the bit node connected with the jth factor node r_j. Define $\phi_i = (K_{j0}, K_{j1}, \dots, K_{jd_f-1})$ as the index set of d_f larger channel coefficients corresponding to the ith receive antenna. Define $\overline{\phi_i}$ as the index set of $N_t - d_f$ remaining channel coefficients. The messages transmitted between the ith factor node and the jth bit node is:

$$\alpha_{ji}^{(l)} = \sum_{k \in \Omega_{j \setminus i}} \beta_{kj}^{(l-1)} \tag{5}$$

The eventual output on the jth node is:

$$L(x_j) = \sum\nolimits_{k \in \Omega_j} \beta_{kj}^{(Q_L)} \tag{6}$$

3.3. EBRDF algorithm

First of all, consider the ith factor node. In the RDF algorithm, only Md_f edges update message. Each edge uses $Md_f - 1$ external information from adjacent edges and send updated information to Md_f connected bits nodes. Because the remaining $M(N_t - d_f)$ edges are disconnected, $M(N_t - d_f)$ bit nodes don't get any information from the ith factor node, which will poor performance. To make the message update more accurate, EBRDF algorithm is proposed. In the EBRDF algorithm, the core concept is the same as the RDF algorithm, which is choosing d_f large channel coefficients and updating Md_f. However, remaining $M(N_t - d_f)$ edges will be updated this time.

To be specific, for deleted edges $r_j \to x_i$, use h_{ik} to replace the channel coefficients corresponding to the edges $r_j \to x_i$. Here, h_{ik} is the smallest value in the d_f selected channel coefficients. In the meanwhile, update the information from edges $r_j \to x_i$ by incorporating information from other $Md_f - 1$ edges. By doing this, information from all MN_t bit nodes is updated. And owe to Gaussian approximation, low complexity has been maintained on the factor node. In general, EBRDF algorithm has slightly higher complexity, but it avoids the disconnection of the factor diagram, which helps to produce better performance.

3.4. GAI-BP

EBRDF-BP algorithm has good balance between complexity and performance. Nevertheless, it is still hard to apply it to large-scale antenna scenarios, mainly due to exponential complexity. Gaussian approximation iteration, known as GAI-BP, is proposed by P.Son[8], which is still useful nowadays. It is noticeable that BP algorithm is always discussed on the bit domain until Yang Junmei extends the algorithm to the real domain in 2017[10], which improves performance further. Here, GAI-BP on the real domain is discussed in order to make it easier to understand channel model.

As is shown, the fundamental channel model can be expressed as $y = Hx + n$. Through GAI-BP, the formula can be further expressed as:

$$r_j = h_{j,i}x_i + \underbrace{\sum\nolimits_{k=1,k\neq i}^{2N} h_{j,k}x_k}_{\text{interference}} + n_j = h_{j,i}x_i + z_{j,i} + n_j \tag{7}$$

Here, noise $z_{j,i}$ is modeled as $N\left(\mu_{z_{j,i}}, \sigma^2_{z_{j,i}}\right)$. And we have:

$$\begin{gathered} \mu_{z_{j,i}} = \sum\nolimits_{k=1,k\neq i}^{2M} h_{j,k} E\{x_k\} \\ \sigma^2_{z_{j,i}} = \sum\nolimits_{k=1,k\neq i}^{2M} h^2_{j,k} Var\{x_k\} + \sigma^2 \end{gathered} \tag{8}$$

Use the following equation to calculate the likelihood probability:

$$p\left(r_j \mid x, H\right)=\frac{1}{\sqrt{2\pi}\sigma_{z_{j,i}}} \exp\left\{\frac{\left(r_j-\mu_{z_{j,i}}-h_{j,i}x_i\right)^2}{2\sigma_{z_{j,i}}^2}\right\} \tag{9}$$

As for alpha message and beta message, it is similar to the analytical process of Original-BP.

From factor node to Bit nodes: β message is expressed as：

$$\beta_{j,i}^{(l)}(s_k) \approx max_{x:x_i=s_k}\left(\sum_{\substack{T:x_t=s_k \\ t\neq i}} \alpha_{t,j}^{(l-1)}(s_k)-\frac{(r_j-\mu_{z_{j,i}}^{(l-1)}-h_{j,i}x_i)^2}{2(\sigma_{z_{j,i}}^2)^{(l-1)}}\right)-max_{x:x_i=s_0}\left(\sum_{\substack{(:x_t=s_0 \\ t\neq i}} \alpha_{t,j}^{(l-1)}(s_k)-\frac{(r_j-\mu_{z_{j,i}}^{(l-1)}-h_{j,i}x_i)^2}{2(\sigma_{z_{j,i}}^2)^{(l-1)}}\right) \tag{10}$$

From Bit node to factor node: α message is expressed as:

$$\alpha_{i,j}^{(l)}(s_k)=\sum_{t=1,t\neq j}^{2N} \beta_{t,i}^{(l-1)}(s_k) \tag{11}$$

A posteriori probability can be expressed as:

$$p_{i,j}^{(l)}(s_k)=\frac{\exp\left(\alpha_{i,j}^{(l)}(s_k)\right)}{\sum_{m=0}^{\sqrt{Q}-1} \exp\left(\alpha_{i,j}^{(l)}(s_m)\right)} \tag{12}$$

In this case, the multi-antenna interference is approximated as Gaussian, which simplifies calculation process greatly. Inspired by this idea, we propose one method to reduce complexity further as we will see later.

3.5. MMSE detector

Although GAI-BP reduces much complexity, the performance is not as good as Original-BP. So, it’s still a problem to improve GAI-BP. An algorithm based on the false a prior information obtained from the linear detector is proposed. Feichi Long gets inspiration from this idea and cascade only one MMSE detector before the BP detector [9]. In this way, false a prior information is used to improve the accuracy of signal detection and reduce the bit error rate The MMSE detector is a linear filter, which is defined as:

$$S_{MMSE}=(H^*H+\sigma^2\cdot I)^{-1}\cdot H^*\cdot y \tag{13}$$

Here, I is a $N_t\times N_t$ unit matrix. When L (iteration) is 0, the prior distribution of x is approximate to:

$$p^{(0)}(x_i) \propto \exp\left(-\frac{|x_i-S_{MMSE_{k(i)}}|^2}{2\sigma_{MMSE_{k(i)}}^2}\right) \tag{14}$$

Here, $\boldsymbol{S_{MMSE_{k(i)}}}$ is the $k(i)th$ element in $\boldsymbol{S_{MMSE}}$. $\sigma_{MMSE_{k(i)}}^2$ is the $k(i)th$ row and $k(i)th$ column element in the noise covariance matrix, which is defined as:

$$\boldsymbol{K_{MMSE}}=(H^*H+\sigma^2\cdot I)^{-1} \tag{15}$$

A prior information is defined as:

$$\alpha_{i,j}^{(0)} = \frac{2\mathcal{R}\left(s_{MMSE_{k(i)}}\right)}{\sigma_{MMSE_{k(i)}}^{2}} \tag{16}$$

3.6. Node judgement based on GAI-BP

In GAI - BP algorithm, the α and β information is updated according to the factor diagram, which means that information is updated after one iteration. Although this method guarantees information accuracy, it increases unnecessary complexity. Therefore, we propose node judgement to improve this situation.

During the update, the α and β information at different nodes don't converge to the same degree. The complexity of the system would increase if the information transferred by some nodes is close to the final value and it is still updated according to the fully connected factor graph. Here, we're going to think about making a judgment at the end of each update to see if the corresponding message converges.

For α information, we can make use of the detection characteristic of GAI-BP, which means we can see whether the relationship between corresponding probabilities of different symbols satisfy one condition. If they satisfy, then we can set the corresponding probability value of the symbol to 1, which is the highest probability. The probability value of the remaining symbol of the node is set to zero and update stops.

So, it refers to what the condition is. After some tests, we finally reckon that we can select two symbols with the highest probability from the same node. If they satisfy inequation (17), we can end judgement. And it's noticeable that this discussion is according to the same node. For different nodes, we just repeat the judgement process.

$$p_{max1} - p_{max2} > TH \tag{17}$$

Here, p_{max1} is the highest probability. p_{max2} is the second highest probability. TH is judgment threshold and the range of value is from 0 to 1. According to the different situations (different channel models, different electromagnetic environment and so on), the value of TH can be different.

4. Computational Complexity

For MMSE detector, it has great performance in high SNRs but a higher complexity. The main calculated amount is operation of pseudoinverse. B. Hassibi has proposed one way to improve algorithms [12]. In general, the lowest computational complexity of MMSE is $O(N_t)$

For RD-GAI-BP algorithm, there are three complexity sources: 1) addition complexity. α and β message update at each node as (11) and (10), which requires $O(N_t N_r)$ complexity. 2) multiplication complexity. β message update at each node as (10) and the calculation of the likelihood probability as (9), which requires $O\left(\sqrt{M} N_t N_r\right)$ complexity. 3) exponent complexity. It is mainly due to the calculation of a posteriori probability as (12), which requires $O\left(\sqrt{M} N_t N_r\right)$ complexity. For node judgement method, it mainly reduces addition complexity. The frequency of α message update at each node decreases, which means calculation of (11) reduces. Therefore, it can reduce algorithm complexity.

5. Performance analysis

In this section, some simulation results for different BP algorithms in uncoded system will be shown. Here, we set L (iteration) is 10, N_t = 4, N_r = 8, which is for all simulation scenarios. And the results are all for 16QAM modulation.

Firstly, in Figure 2, we compare the performance of Original-BP, EBRDF-BP, GAI-BP (bit domain). Original-BP has the lowest bit error rate while GAI-BP is opposite. As is shown before, it is related to their complexity. In small-scale system, it seems that complexity is not so important that it's a better choice for Original-BP and EBRDF-BP.

DOI: 10.54254/2755-2721/6/20230776

Next, in Figure 3, we compare the performance of EBRDF-BP, GAI-BP (bit domain) and RD-GAI-BP (real domain). Compared with discussion on bit domain, RD-GAI-BP performs better and even has lower error rate than EBRDF-BP. Discussion on real domain can be generalized to other algorithms and it is adopted widely nowadays.

Finally, in Figure 4, we compare the performance of RD-GAI-BP, MMSE, Node Judgement. Here, we combine MMSE detector with node judgement method and set TH to 0.95. Although MMSE detector has higher error rate in low SNRs, it performs much better than any other method that we have mentioned in high SNRS. Additionally, good result is achieved by combining node judgement method with MMSE.

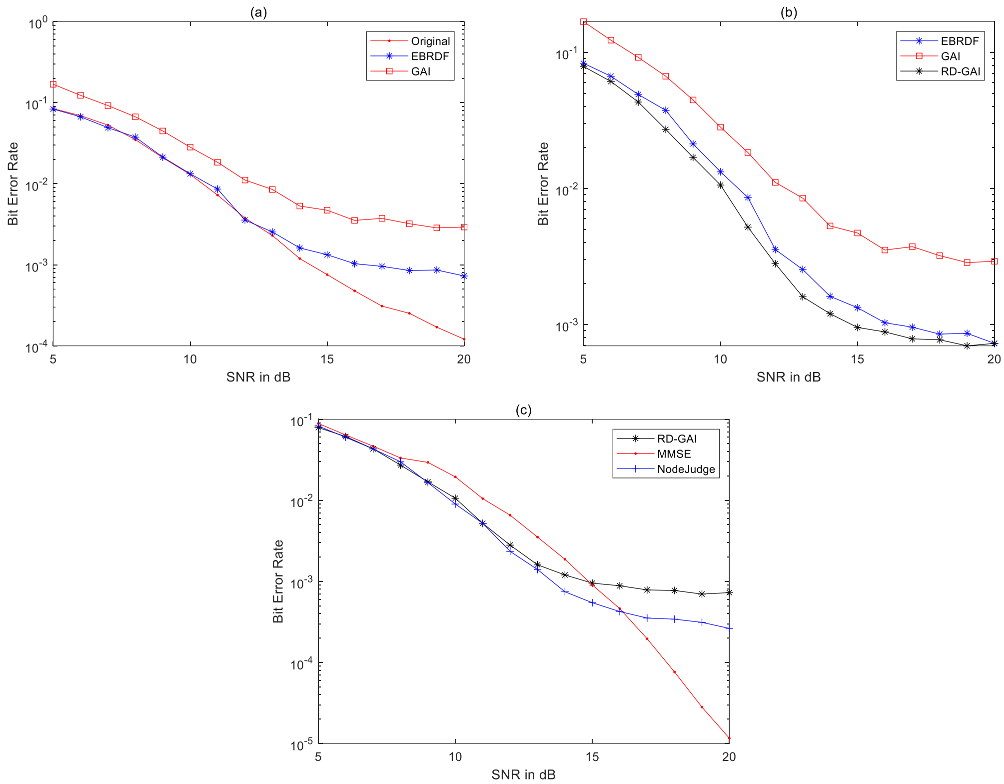

Fig.2. BER performance: (a) Original-BP, EBRDF-BP, GAI-BP with N_t = 4, N_r = 8 (16QAM) (b)EBRDF-BP, GAI-BP, RD-GAI-BP with N_t = 4, N_r = 8 (16QAM). (c) RD-GAI-BP, MMSE, Node Judgement with N_t = 4, N_r = 8 (16QAM)

6. Conclusion

The paper has concluded several important BP algorithms and compared their different performance. We also propose a method to judge message convergence based on the GAI-BP. Reducing the number of message update makes the whole complex lower than the original complexity. From the simulation results, we can see that if we MMSE detector we can get great results than RD-GAI-BP. Considering the complexity of MMSE, we combine MMSE with node judgment method to reduce complexity. It can be seen that this method can still get better result than RD-GAI-BP. In general, the idea has certain feasibility and practicability.

References

[1] C. -X. Wang et al., “Cellular architecture and key technologies for 5G wireless commun ication networks,” in IEEE Communications Magazine, vol. 52, no. 2, pp. 122-130, F ebruary 2014, doi: 10.1109/MCOM.2014.6736752.

[2] Boccardi, F., Heath, R., Lozano, A., Marzetta, T. and Popovski, P., 2014. Five disruptiv e technology directions for 5G. IEEE Communications Magazine, 52(2), pp.74-80.
[3] Marzetta, T., 2010. Noncooperative Cellular Wireless with Unlimited Numbers of Base S tation Antennas. IEEE Transactions on Wireless Communications, 9(11), pp.3590-3600.
[4] Rusek, F., Persson, D., Buon Kiong Lau, Larsson, E., Marzetta, T. and Tufvesson, F., 2 013. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. IEEE Signal Processing Magazine, 30(1), pp.40-60.
[5] J. Hoydis, S. ten Brink and M. Debbah, "Massive MIMO in the UL/DL of Cellular Net works: How Many Antennas Do This paper Need?," in IEEE Journal on Selected Ar eas in Communications, vol. 31, no. 2, pp. 160-171, February 2013, doi: 10. 1109/JS AC. 2013. 130205.
[6] Foschini, G., 2002. Layered space-time architecture for wireless communication in a fadi ng environment when using multi-element antennas. Bell Labs Technical Journal, 1(2), pp.41-59.
[7] Jun Hu and Duman, T., 2008. Graph-based detection algorithms for layered space-time a rchitectures. IEEE Journal on Selected Areas in Communications, 26(2), pp.269-280.
[8] P. Som, T. Datta, A. Chockalingam and B. S. Rajan, Improved large-MIMO detection b ased on damped belief propagation, 2010 IEEE Information Theory Workshop on Inf ormation Theory (ITW 2010, Cairo), 2010, pp. 1-5, doi: 10. 1109/ITWKSPS. 2010. 5 503188.
[9] F. Long, T. Lv, R. Cao and H. Gao, Single edge based belief propagation algorithms fo r MIMO detection, 34th IEEE Sarnoff Symposium, 2011, pp. 1-5, doi: 10. 1109/SAR NOF. 2011. 5876456.
[10] Yang, J., Song, W., Zhang, S., You, X. and Zhang, C., 2017. Low-Complexity Belief P ropagation Detection for Correlated Large-Scale MIMO Systems. Journal of Signal Pr ocessing Systems, 90(4), pp.585-599.
[11] Tan, X., Xu, W., Sun, K., Xu, Y., Be'ery, Y., You, X. and Zhang, C., 2020. Improvin g Massive MIMO Message Passing Detectors With Deep Neural Network. IEEE Tran sactions on Vehicular Technology, 69(2), pp.1267-1280.
[12] B. Hassibi, "An efficient square-root algorithm for BLAST," 2000 IEEE International Co nference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH371 00), 2000, pp. II737-II740 vol.2, doi: 10.1109/ICASSP.2000.859065.