ON RADIUS CONTROL OF TREE-PRUNED SPHERE DECODING

Byonghyo Shim[*]

Korea University
Anam-dong, Seongbuk, Seoul, Korea

Insung Kang

Qualcomm Inc.
5775 Morehouse Dr., San Diego, CA

ABSTRACT

In this paper, we propose a novel radius control strategy for sphere decoding referred to as *inter search radius control* that provides further improvement of the computational complexity with minimal extra cost and negligible performance penalty. The proposed method focuses on the sphere radius control strategy when a candidate lattice point is found. For this purpose, the dynamic radius update strategy as well as the lattice independent radius selection scheme are jointly exploited. From simulations in multiple-input and multiple-output (MIMO) channels, it is shown that the proposed method provides a substantial improvement in complexity with near-ML performance.

Index Terms— maximum likelihood, tree pruning, sphere radius, multiple input multiple output (MIMO), sphere decoding

1. INTRODUCTION

Maximum likelihood (ML) detection of the sequence of finite alphabet symbols requires an exhaustive search for the entire block of symbols. Although the ML solution is optimal for achieving the minimum probability of error, it did not received attention until recently due to the NP-hard nature of the problem [2, 4]. After the rediscovery of Fincke and Pohst's work [1], an efficient search algorithm called *Sphere Decoding* (SD) has received much attention these days. Since the heart of the SD algorithm lies on the choice of sphere radius within which the search space is limited, several approaches controlling the sphere radius have been proposed including *increase radius search* (IRS) [3], *improved increasing radius search* (IIRS) [6], and *increasing radii algorithm* (IRA) [5]. It has been shown that these algorithms improve the computational complexity at the expense of negligible performance loss. In [7], we have proposed an algorithm relaxing the strict ML search to gain the benefit in computation referred to as *probabilistic tree pruning SD* (PTP-SD). We have shown that an addition of the probabilistic noise constraint into the path metric, generated by the probabilistic model of the unvisited nodes, expedites the tree pruning. Since the sphere constraint is loose for most layers of the tree being searched, by placing the probabilistic pruning condition on top of the sphere condition, we could achieve considerable savings in computation.

In this paper, we propose a sphere radius control strategy that provides further improvement of the computational complexity with minimal extra cost and code change. While the PTP-SD tries to fortify the structural weakness of the sphere search by tightening the sphere radius per layer, proposed method focuses on the sphere radius control strategy when a candidate lattice point is found. In this respect, we can view the PTP-SD as *intra-search radius control* and the proposed method as *inter-search radius control* (ISRC). Note, despite the fact that the SD shrinks the volume of hypersphere by dynamically updating the radius whenever a new candidate is found, it does not guarantee the fast converge to the ML point. Hence, a strategy providing an aggressive radius control is required for achieving further reduction in search space. In fact, the ISRC exploits the competition between the dynamically updated radius and the statistically designed set of radii. As a result, number of unpromising lattice points is excluded from the search process while the performance close to the ML is maintained.

The rest of this paper is organized as follows. In section II, we briefly review the SD and the PTP-SD. In section III, we present the proposed ISRC algorithm. The simulation results and discussion are provided in section IV.

2. SD AND PTP-SD

2.1. SD Algorithm

Consider the ML detection of a real-valued linear system described by $\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{v}$ where $\mathbf{s}$ is the transmitted symbol vector whose components are elements of a finite set $\mathcal{F}$, $\mathbf{r}$ is the received signal vector, $\mathbf{v}$ is the i.i.d. Gaussian noise vector, and $\mathbf{H}$ is a channel matrix. Under the assumption that $\mathbf{H}$ is given, the ML estimate becomes

$$\mathbf{s}_{ml} = \arg \min_{\mathbf{s} \in \mathcal{F}^m} ||\, \mathbf{r} - \mathbf{H}\mathbf{s} \,||^2. \qquad (1)$$

Instead of searching all lattice points $\mathbf{H}\mathbf{s}$, the SD algorithm searches the lattice points inside the hypersphere with radius $\sqrt{r_0}$, i.e., $||\, \mathbf{r} - \mathbf{H}\mathbf{s} \,||^2 < r_0$. In fact, actual search is performed

[*]This work is supported by a research grant from Qualcomm Inc., KOSEF R01-2008-000-20292-0, and the second BK 21 project.

in the QR-transformed domain given by

$$J(\mathbf{s}) = ||\,\mathbf{y} - \mathbf{R}\mathbf{s}\,||^2 \leq d_0 \tag{2}$$

where $\mathbf{H} = [\mathbf{Q}\ \ \mathbf{U}][\mathbf{R}^T\ \ \mathbf{0}^T]^T$, $\mathbf{y} = \mathbf{Q^T r}$, and $d_0 = r_0 - ||\mathbf{U^T r}||^2$. Since $\mathbf{R}$ is an upper triangular matrix, (2) becomes

$$\sum_{k=1}^{m}(\,y_k - \sum_{l=k}^{m} r_{k,l}s_l\,)^2 \leq \quad d_0. \tag{3}$$

Emphasizing that each term in the left side is a function of $s_k, \cdots, s_m$ (henceforth denote s_k^m), (3) can be expressed as

$$B_1(s_1^m) + B_2(s_2^m) + \cdots + B_m(s_m^m) \leq \quad d_0 \tag{4}$$

where $B_k(s_k^m) = (\,y_k - \sum_{l=k}^{m} r_{k,l}s_l\,)^2$ is the branch metric at layer $m-k+1$. In the first layer (bottom row in the matrix structure), s_m satisfying $B_m(s_m) \leq d_0$ is found. Once s_m is chosen, we move on to the next layer to find s_{m-1} satisfying $B_{m-1}(s_{m-1}^m) + B_m(s_m) \leq d_0$. By repeating this step and updating a radius whenever a new lattice point $\mathbf{Rs}$ is found, the SD algorithm outputs the ML point $\mathbf{s}_{ml}$ for which the cost function $J(\mathbf{s})$ is minimized.

2.2. Probabilistic Tree Pruning (PTP) SD

Although the SD algorithm should test the condition described in (4), due to the causality of the search, the actual condition to be checked is

$$P_k^m(s_k^m) = B_k(s_k^m) + \cdots + B_m(s_m^m) \leq d_0 \tag{5}$$

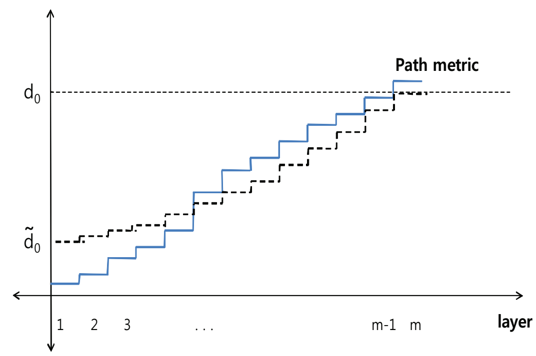
where $P_k^m(s_k^m)$ is the path metric that is an accumulation of branch metrics from layers 1 to $m-k+1$. The key idea behind the probabilistic tree pruning is to use (4) instead of (5) throughout all layers in the search. Since the branch metrics $B_1, \cdots, B_{k-1}$ are unavailable at layer $m-k+1$, assuming perfect decoding, we model them as Gaussian noise, i.e., $B_l(s_l^m) = (\,y_l - \sum_{j=l}^{m} r_{l,j}\,s_j\,)^2 = v_l^2$ where v_l is the l-th component of $\mathbf{v}$. With this setup, the new necessary condition becomes

$$\sum_{l=1}^{m} B_l(s_l^m) = P_k^m(s_k^m) + \sum_{l=1}^{k-1} v_l^2 \leq d_0. \tag{6}$$

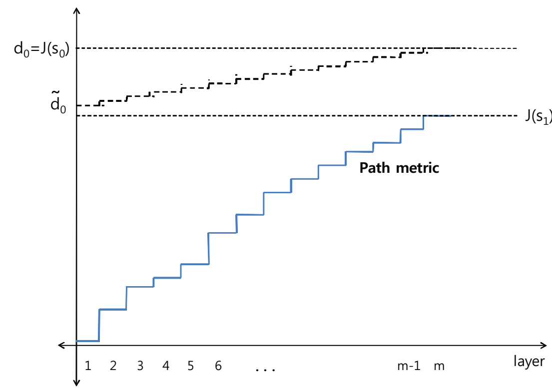
Since $v_1, \cdots, v_{k-1}$ are values from i.i.d. Gaussian distribution, $\sum_{l=1}^{k-1} v_l^2$ becomes the χ^2-random variable (RV) with $k-1$ degrees of freedom (DOF). Denoting $\Phi_{k-1} = \sum_{l=1}^{k-1} v_l^2$, (6) becomes

$$P_k^m(s_k^m) + \Phi_{k-1} \leq d_0. \tag{7}$$

In order to obtain the pruning condition, a notion of pruning probability is introduced. On each node visited, we examine the probability that the rest of the tree is decoded perfectly so that the remaining portion is a pure noise contribution. If



(a) Early pruning typical in low to mid SNR



(b) Pruning is not effective in high SNR

Fig. 1. Illustration of PTP-SD.

the probability of this event is too small and thus less than a pre-specified threshold, we regard this event as unlikely one and prune the subtree starting from the node. This condition is summarized as

$$P_r(\Phi_{k-1} \leq d_0 - P_k^m(s_k^m)) \ < \ P_\epsilon \tag{8}$$

where P_ϵ is the pre-specified pruning probability. Since the left side of (8) is the CDF of χ^2-RV, we have

$$d_0 - P_k^m(s_k^m) \ < \ \beta_{k-1}. \tag{9}$$

where $\beta_{k-1} = F_\Phi^{-1}(P_\epsilon\,;\,k-1)$. The interpretation of (9) is that if the path metric in layer $m-k+1$ is larger than $d_0 - \beta_{k-1}$, then the rest of search is unlikely to satisfy the sphere condition even for the best scenario (the remaining nodes are detected perfectly and their contributions are noises only). Hence, whenever a path s_k^m meets this condition, we remove all its children from the tree (refer Fig. 1(a)).

In [7], we reported a significant reduction of complexity of the PTP-SD over the SD in low and mid SNR regimes. However, the benefit of the PTP-SD vanishes as the SNR increases so that the complexity of the PTP-SD converges asymptotically to the SD complexity in the high SNR regime. The main reason is that if the cost function difference $(J(\mathbf{s}_0) - J(\mathbf{s}_1))$ is larger than $\beta_{\max}$, as illustrated in Fig. 1(b), the pruning of the PTP-SD is no more useful. Indeed, as the distance between cost functions increases, the efficiency of PTP-SD disappears. Hence, it might be better to consider lattice independent radius selection strategy for this scenario.

3. INTER-SEARCH RADIUS CONTROL (ISRC)

3.1. Lattice Independent Radius Selection

As a lattice independent radius selection scheme, *increase radius search* (IRS) has been proposed [3]. The assumption of IRS is that when the detection is done perfectly, branch metrics would contain the noise contribution only. Although it is an ideal scenario but provides a clue to choose the initial radius square d_0. With this assumption, $||\,\mathbf{y} - \mathbf{Rs}\,||^2 = \sum_{i=1}^{n} v_i^2$ becomes the χ^2-RV with n DOF. By denoting $\Phi_n = \sum_{i=1}^{n} v_i^2$ and setting a threshold probability P_{th} (say $P_{th} = 0.01$), a condition for the initial radius is obtained as $F_\Phi(d_0\,;\,n) = 1 - P_{th}$. Taking the inverse of χ^2-CDF, we directly get $d_0 = F_\Phi^{-1}(1 - P_{th}\,;\,n)$. Due to the fact that the radius is chosen by the noise statistics only, this approach has an advantage of skipping lots of unnecessary lattice points in the initial search. However, if lattice points are densely packed, this method might not be effective (e.g., low SNR scenario). In addition, when d_0 is chosen to be smaller than the ML distance ($d_0 < J(\mathbf{s}_{ml})$), the sphere search fails. In this case, d_0 should be re-computed with smaller P_{th} and the search should be started over so that an additional loop is needed for the implementation.

3.2. Inter Search Radius Control

By the marriage between the dynamic radius update and the lattice independent (noise statistics based) radius selection, we can maximize the benefit in complexity without sacrificing the performance. The salient feature of the proposed method is to speed-up the search by choosing a smaller sphere radius than the cost function of the lattice point found.

In a normal SD operation, the detection error occurs when the last candidate $\mathbf{s}_f$ which always equals $\mathbf{s}_{ml}$ is not equal to the transmitted symbol vector $\mathbf{s}_{tx}$

$$\begin{aligned} P_{err}(\text{ML}) &= P(\mathbf{s}_{ml} \neq \mathbf{s}_{tx}) \\ &= P(J(\mathbf{s}_{ml}) < J(\mathbf{s}_{tx}) = ||\mathbf{v}||^2). \end{aligned} \quad (10)$$

However, if an aggressive radius control is introduced, the search might be finished without reaching the ML point. In this case, the detection error probability is

$$\begin{aligned} & P_{err}(\text{Near-ML}) = P_{err}(\mathbf{s}_f = \mathbf{s}_{ml}) + P_{err}(\mathbf{s}_f \neq \mathbf{s}_{ml}) \\ & = P(\mathbf{s}_f \neq \mathbf{s}_{tx},\, \mathbf{s}_f = \mathbf{s}_{ml}) + P(\mathbf{s}_f \neq \mathbf{s}_{tx},\, \mathbf{s}_f \neq \mathbf{s}_{ml}). \end{aligned} \quad (11)$$

The first term in the right of (11) equals $P_{err}(\text{ml})$. The second term causing an additional increase in the error probability can be further expressed as

$$\begin{aligned} P(\mathbf{s}_f \neq \mathbf{s}_{tx},\, \mathbf{s}_f \neq \mathbf{s}_{ml}) &= P(J(\mathbf{s}_{ml}) < J(\mathbf{s}_f) < ||\mathbf{v}||^2) \\ &+ P(J(\mathbf{s}_{ml}) \leq ||\mathbf{v}||^2 < J(\mathbf{s}_f)). \end{aligned} \quad (12)$$

Since $J(\mathbf{s}_{ml})$ and $||\mathbf{v}||^2$ are equal or very close for the mid and high SNR regimes, the second term in (12) becomes a dominating factor and thus

$$P_{err}(\text{Near-ML}) - P_{err}(\text{ML}) \sim P(J(\mathbf{s}_{ml}) \leq ||\mathbf{v}||^2 < J(\mathbf{s}_f)). \quad (13)$$

In the sphere search, the event $J(\mathbf{s}_{ml}) \leq ||\mathbf{v}||^2 < J(\mathbf{s}_f)$ occurs when the sphere radius square d_0 is set aggressively to $d_0 < J(\mathbf{s}_{ml}) \leq ||\mathbf{v}||^2 < J(\mathbf{s}_f)$. Since our goal is to design the sphere radius reducing the complexity while maintaining the performance close to ML detection (say $P_{err}(\text{Near-ML}) - P_{err}(\text{ML})$ is within P_δ), we should have

$$P(d_0 < J(\mathbf{s}_{ml}) \leq ||\mathbf{v}||^2 < J(\mathbf{s}_f)) \leq P_\delta. \quad (14)$$

Again it is highly likely that $J(\mathbf{s}_{ml}) = ||\mathbf{v}||^2$ for the mid and high SNR regimes, so we approximately have

$$F_\Phi(J(\mathbf{s}_f)\,;\,n) - F_\Phi(d_0\,;\,n) \leq P_\delta \quad (15)$$

and directly we have $F_\Phi^{-1}(F_\Phi(J(\mathbf{s}_f)\,;\,n) - P_\delta\,;\,n) \leq d_0$. Hence, a natural choice of the sphere radius when a lattice point $\mathbf{s}$ is found is

$$d_0 = F_\Phi^{-1}(F_\Phi(J(\mathbf{s})\,;\,n) - P_\delta\,;\,n). \quad (16)$$

Employing (16), further shrinking of the search space can be achieved. In fact, when the lattice points are packed locally in their cost function, even with small P_δ, employment of (16) provides a good complexity gain. Regarding P_δ, it is clear that $F_\Phi(J(\mathbf{s})\,;\,n) - P_\delta > 0$ and thus we choose

$$P_\delta = \epsilon\, F_\Phi(J(\mathbf{s})\,;\,n) \quad (17)$$

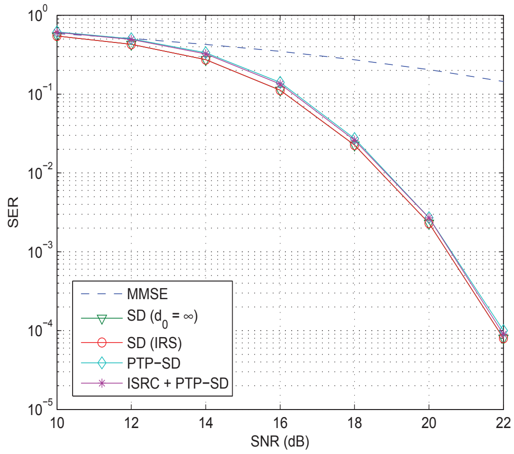
where $0 < \epsilon < 1$. As a rule of thumb, we might use relatively large ϵ (≈ 0.5) for a few initial candidates and small ϵ (≈ 0.1) for the rest of candidates.

The features of the ISRC strategy are summarized as

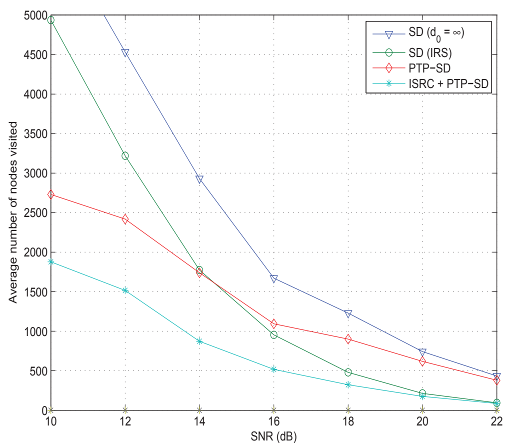
1. By the addition of the probabilistic radius control on top of the dynamic adjustment, tight sphere radius minimizing performance loss can be obtained.

2. ISRC strategy and intra search radius control (PTP-SD) can be effectively united. The former strategy is especially useful when the lattice points are spaced apart (e.g., high SNR) and the latter one is more effective for the densely packed lattice structure (e.g., low SNR).

3. The extra complexity of the intra search radius control is at most one subtraction per layer [7] and that for the ISRC is one compare operation when a candidate lattice point is found. Since the online computation of χ^2-CDF and inverse CDF is a bit cumbersome, lookup table might be a good option for computing d_0.



(a) SER



(b) Complexity

Fig. 2. Performance and complexity of SD algorithms for 8×8 MIMO system with 16-QAM modulation.

4. SIMULATIONS AND DISCUSSION

In order to observe the performance of the proposed method, we consider 16-QAM transmission over the MIMO channel with Rayleigh fading ($h_{ij} \sim CN(0, 1)$). For comparison, we employ linear MMSE estimation, reference SD ($d_0 = \infty$), IRS, PTP-SD, as well as the proposed ISRC+PTP-SD algorithm in our simulation, where we set $P_\epsilon = 0.1$ for the PTP-SD. As a measure for the performance and complexity, we employ the symbol error rate (SER) and the average number of nodes visited.

Fig. 2(a) provides the performance results for 8×8 MIMO systems with 16-QAM modulation (total number of lattice points is $16^8 \sim 4.3 \times 10^9$). While the performance difference is unnoticeable among the reference SD, IRS, PTP-SD, and the proposed method over the entire simulation range, we observe the clear distinction in complexity as shown in Fig. 2(b). Even though the PTP-SD achieves substantial complexity savings in the low SNR regime, as mentioned in Section 2.2, its benefit disappears as the SNR increases . In contrast, the IRS-based SD shows relatively large complexity reduction in the high SNR regime so that we observe the clear crossover point between them around 14 dB. Since the proposed method (ISRC+PTP-SD) adopts the advantage of both, it is no surprise that the proposed method outperforms both in complexity. In fact, the proposed method achieves the minimal computational complexity among all methods under test providing 3X reduction in complexity over the reference SD. Even compared to the PTP-SD, the proposed method achieves at least 30% reduction in the entire simulation range.

5. REFERENCES

[1] U. Fincke and M. Pohst, “Improved methods for calculating vectors of short length in a lattice, including a complexity analysis,” *Mathematics of Computation*, vol. 44, pp. 463-471, 1985.

[2] E. Agrell, T. Eriksson, A. Vardy, and K. Zegar, “Closet point search in lattices,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 2201-2214, Aug. 2002.

[3] B. Hassibi and H. Vikalo, “On the sphere-decoding algorithm I. Expected complexity,” *IEEE Trans. Signal Proc.*, vol. 53, pp. 2806-2818, Aug. 2005.

[4] M. O. Damen, H. E. Gamel, and G. Caire, “On maximum-likelihood detection and the search for the closest lattice point,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 2389-2402, Oct. 2003.

[5] R. Gowaikar and B. Hassibi, “Statistical Pruning for Near-Maximum Likelihood Decoding”, *IEEE Trans. Sig. Proc.*, vol. 55, no. 6, pp. 2661-2675, June 2007.

[6] W. Zhao and G. B. Giannakis, “Sphere decoding algorithms with improved radius search,” *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1104-1109, July 2005.

[7] B. Shim and I. Kang, “Sphere decoding with a probabilistic tree pruning”, *IEEE Trans. Signal Proc.* vol. 56, no. 10, pp. 4867-4878, Oct. 2008.