Stopping Criterion for Complexity Reduction of Sphere Decoding

Zheng Ma, Bahram Honary, Pingzhi Fan, and Erik G. Larsson

***Abstract*—Maximum-likelihood detection in MIMO communications amounts to solving a least-squares problem with a constellation (alphabet) constraint. One popular method that can be used to solve this problem is sphere decoding. We show in this letter that by employing a simple stopping criterion, it is possible to significantly reduce the complexity of sphere decoding over a wide range of SNRs, without a noticeable performance degradation. Specifically, simulation results demonstrate that a 10%–90% reduction of the average complexity could be achieved.**

***Index Terms*—Lattice reduction, MIMO, sphere decoding.**

I. Introduction

MULTIPLE-input multiple-output (MIMO) technology is used in many emerging wireless communication systems. One main concern with MIMO is the high detection and decoding complexity at the receiver. Complexity considerations typically prohibit the use of optimum maximum-likelihood detection (MLD) when the number of antennas or the modulation size is large. Some sub-optimum detection algorithms, such as zero-forcing (ZF) and successive interference cancellation (SIC) can perform detection with much lower complexity, but at a significant performance degradation. Achieving MLD performance in a MIMO system amounts to solving a so-called closest-vector problem (CVP). The CVP itself is generally NP-hard, but by employing the enumeration algorithm proposed by Fincke and Pohst [1], one can often solve it with a reasonable computational effort. This algorithm, and its relatives, are often collectively referred to as "sphere decoding" in the communications literature. We use this terminology, and whenever we refer to sphere decoding in this paper, we assume that the detector takes the constellation boundary into account.

Manuscript received January 16, 2009. The associate editor coordinating the review of this letter and approving it for publication was J. van de Beek.

Z. Ma and P. Fan are with the Provincial Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu, 610031, China (e-mail: {zma, p.fan}@ieee.org). Z. Ma is also with the Department of Communications Systems, Lancaster University, UK and State Key Lab of Mobile Communications, Southeast University, China.

B. Honary is with the Dept. of Communication Systems, Lancaster University, UK (e-mail: honary@exchange.lancs.ac.uk).

E. G. Larsson is with the Dept. of Electrical Engineering, Linköping University, SE-581 83, Linköping, Sweden (e-mail: erik.larsson@isy.liu.se).

The work of Z. Ma and P. Fan was supported by the Sichuan Youth Science Foundation (No.08ZQ026-015), the Specialized Research Fund for the Doctoral Program of Higher Education (No.20070613007), the open research fund of National Mobile Communications Research Laboratory (Southeast University)(No.W200803), the Sino-Swedish Sci & Tech Cooperation Program under the grant No. 2008DFA12160, the 863 High-Tech R&D Program (No.2009AA01Z238), and the 111 project (No.111-2-14). The work of E. Larsson was supported by VINNOVA, the Swedish Research Council (VR), and the Swedish Foundation for Strategic Research (SSF). E. Larsson is a Royal Swedish Academy of Sciences (KVA) Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

Digital Object Identifier 10.1109/LCOMM.2009.090112

There are two main directions of research on CVP problems in communications. One is to combine lattice reduction [2] with sub-optimum detection algorithms to achieve sub-optimum performance with very low complexity [3], [4]. This approach is also known as lattice-reduction-aided detection. In [3] the performance gap to MLD is within 2.5–5 dB. A popular algorithm for lattice reduction is that of Lenstra, Lenstra and Lovász (LLL) [2] and it has been successfully used with ZF and SIC detectors [6]. The other research direction is to develop improved versions of sphere decoding [7], [8]. Although many such algorithms can approach MLD performance, they have very high complexity. In particular, in the scenarios relevant to us, the average complexity of sphere decoding grows exponentially with the number of antennas (see [5] for precise statements regarding complexity).

In iterative algorithms, one sometimes uses *stopping criteria* as a means for complexity reduction. Mow [3] argued that stopping criteria could be used to reduce the complexity of sphere decoding, and suggested a simple test that could be used to halt the decoding process. Zhao [9] proposed stopping criteria based on probabilistic assumptions, but this work did not exploit lattice reduction. In this letter, we propose a new stopping-criteria-aided MIMO detection algorithm that makes combined use of lattice reduction, SIC, and sphere decoding and which achieves near-MLD performance at much lower complexity than conventional sphere decoding.

II. System Model

Consider a MIMO system with N_T transmit antennas and N_R receive antennas ($N_T \leqslant N_R$). With I/Q modulation, the baseband signal vector transmitted during each symbol period can be represented by a complex-valued vector of length N_T, where each component of the vector is drawn from a complex constellation. When implementing the detectors, it is convenient to work with a real-valued formulation of the problem, at least for separable constellations such as M-QAM (herein we will limit our discussion to such signal constellations). By appropriately rearranging the real and imaginary parts of the transmitted symbols and of the received samples (see, e.g., [10, Sec. II.A] for details) we can write

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

where $\boldsymbol{x}$ is a real-valued $2N_T$-length transmitted vector whose elements belong to a finite alphabet $\mathcal{S}$, $\boldsymbol{H}$ is a real-valued $2N_R \times 2N_T$ channel matrix, $\boldsymbol{y}$ is a real-valued $2N_R$-vector, and $\boldsymbol{w}$ is additive white Gaussian noise (AWGN) with i.i.d. $N(0, \sigma_w^2)$ elements.

If $\boldsymbol{H}$ is perfectly known at the receiver, the aim of the MLD is to solve

$$\hat{\boldsymbol{x}}_{\mathrm{MLD}} = \underset{x_k \in \mathcal{S}}{\operatorname{argmin}} ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 \tag{2}$$

TABLE I
NUMBER OF VECTOR ERRORS (VE) FOR DIFFERENT SNRS, $N_T = 4$, $N_R = 4$, AND QPSK MODULATION, LRA-SIC DETECTOR

SNR	BER	average number of vector errors per 1000 vectors
10	1.16×10^{-1}	488.7
15	2.63×10^{-2}	104.7
20	1.61×10^{-3}	5.37
25	3.21×10^{-5}	0.09

The idea of lattice reduction in MIMO detection is to transform the coordinate system so that the lattice is closer to orthogonal, and consequently so that low-complexity, sub-optimal methods perform better [6].

Among the low-complexity detectors, SIC has superior performance compared to the ZF detector, but it has almost the same complexity as ZF. Hence we will use the SIC detector in our discussion. To obtain MLD performance, sphere decoding with enumeration can be used. The lowest-complexity enumeration process for sphere decoding known is the Schnorr-Euchner improved version of the Fincke-Pohst algorithm [8]. The basic idea is to enumerate all points inside a sphere whose radius is no more than specific distance from the query point to the Babai point, and then choose the closest one as decoding result. Normally, the point found by lattice-reduction aided SIC is called the Babai point, and it is often taken as the starting point by the sphere decoder. The Babai point is likely to be the closest point itself [1].

Note that our discussion assumes that the receiver perfectly knows $\boldsymbol{H}$. The problem of optimal detection for the case when $\boldsymbol{H}$ is only partially known has the same form as (2) in many relevant cases (e.g., see [10, Sec. V.A]).

III. STOPPING CRITERIA FOR SPHERE DECODING

Linear detectors (such as ZF) have low complexity but offer worse performance than the MLD does. At the same time, sphere decoding offers MLD or near-MLD performance but still has very high complexity. The question is whether one can obtain MLD performance at lower complexity. To illuminate this question, we performed numerical experiments with the lattice-reduction-aided SIC (LRA-SIC) detector.

Table I shows the results, obtained by simulating 10^7 vectors. We say that a vector error (VE) occurs when at least one symbol in the detected vector is incorrect. At high SNR (e.g., 25 dB), almost all of the vectors can be correctly detected by lattice-reduction aided SIC, and there is no need to perform sphere decoding. To obtain MLD performance at SNR=15 dB, about 90% of the vectors would not need be processed by the sphere decoder. Hence, there is significant potential for complexity reduction. Note that the first step of the sphere decoder is to find the Babai point, which is precisely what the SIC detector delivers. If we could determine on-line whether the SIC did find the correct solution or not, then we could use this as a stopping criterion and halt the decoding so that the sphere decoder is run only when SIC was unable to find the correct solution. This would reduce complexity significantly while guaranteeing optimum (MLD) performance.

It is known that, one of the problems with lattice-reduction-aided linear detection is that some symbols of the detected vector may lie outside the constellation $\mathcal{S}$. When this occurs, we say that the resulting vector is invalid. In numerical experiments, we have noticed that the occurrence of a VE is highly correlated with the occurrence of an invalid vector. Table II shows empirical results for a $N_T = N_R = 8$ MIMO system with 16-QAM modulation. Here, $P(V)$ is the probability of a VE, and $P(O)$ is the probability that at least one symbol in the detected vector lies outside $\mathcal{S}$, i.e., that an invalid vector was obtained. The probability $P(V, \tilde{O})$ denotes the joint probability that a VE occurs but the detected vector is valid (i.e., all its elements belong to $\mathcal{S}$). We see that $P(O)$ is always slightly larger than $P(V)$. This is some because some out-of-modulation-alphabet vectors could be recovered by applying the modulation alphabet constraint [3]. However the joint probability $P(V, \tilde{O})$ is zero in all cases.[1] We can find cases where $P(V, \tilde{O})$ is nonzero, for example when N_T is small and the SNR is low (see Table III for the case of $N_T = 2$ and $SNR < 10$ dB). However, when we compare $P(V, \tilde{O})$ to the relevant target error probabilities, it is at least two orders of magnitude smaller. Hence, $P(V, \tilde{O})$ appears to be so small that it may be considered to be zero for practical purposes. Under the assumption that $P(V, \tilde{O}) = 0$, we have that $P(V|\tilde{O}) = P(V, \tilde{O})/P(\tilde{O}) = 0$ (because $P(\tilde{O}) \neq 0$). This means that a vector is almost always correctly detected if all its symbols belong to $\mathcal{S}$. This is the observation that serves as stopping criteria for our proposed improved sphere decoding. We summarize as follows:

TABLE II
EMPIRICAL RESULTS FOR $N_T = 8, N_R = 8$ AND 16-QAM MODULATION, LRA-SIC DETECTOR

SNR	BER	$P(V)$	$P(O)$	$P(V, \tilde{O})$
10	3.11×10^{-1}	0.9996	1	≈ 0
15	2.26×10^{-1}	0.9915	0.9997	≈ 0
20	1.47×10^{-1}	0.8700	0.9246	≈ 0
25	3.84×10^{-2}	0.2233	0.2439	≈ 0
30	7.26×10^{-2}	0.0028	0.0029	≈ 0

TABLE III
EMPIRICAL RESULTS FOR $N_T = 2, N_R = 2$ AND QPSK MODULATION, LRA-SIC DETECTOR

SNR	BER	$P(V)$	$P(O)$	$P(V, \tilde{O})$
5	1.49×10^{-1}	4.31×10^{-1}	4.51×10^{-1}	5.41×10^{-3}
10	7.05×10^{-2}	2.12×10^{-1}	2.32×10^{-1}	1.60×10^{-3}
15	2.15×10^{-2}	6.37×10^{-2}	6.43×10^{-2}	2.25×10^{-4}
20	4.40×10^{-3}	1.24×10^{-2}	1.27×10^{-2}	2.93×10^{-5}
25	6.89×10^{-4}	6.89×10^{-4}	6.90×10^{-4}	5.29×10^{-7}

Conjecture 1: For lattice-reduction aided SIC detection, a VE is most likely to occur if the vector is invalid (i.e., it has elements outside the alphabet). Conversely, if the detected vector is valid (all its elements belong to the alphabet), there is little chance that the vector is in error. □

We do not have a proof of Conjecture 1 other than the simulation results documented in Tables II–III. However, we can understand it intuitively as follows. The lattice-reduction performed prior to running the SIC detector is a highly nonlinear transformation and it will transform the signal constellation in a way such that points that were neighbors before the transformation end up being far away from each other after

[1] $P(V, \tilde{O})$ is not exactly zero, but we did not encounter any event $(V, \tilde{O})$ in this simulation. Therefore we write "≈ 0" in the table.

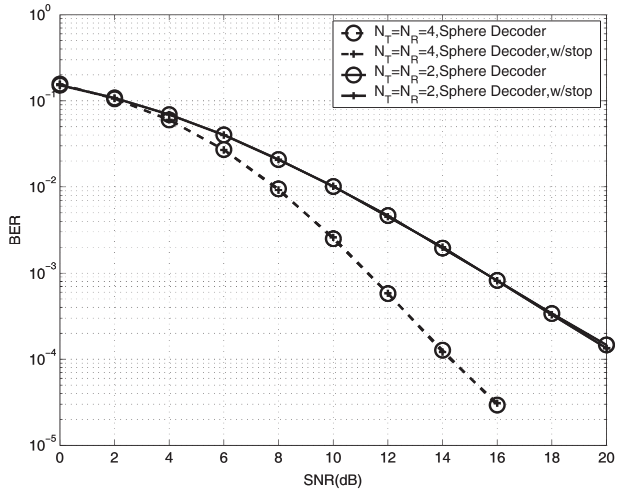

Fig. 1. Performances of sphere decoding (with infinite initial search radius) and of the proposed algorithm using sphere decoding with a stopping criterion. Results are shown for an $N_T = N_R = 2$ and an $N_T = N_R = 4$ MIMO system with QPSK modulation. "W/stop" stands for "with proposed stopping criterion."

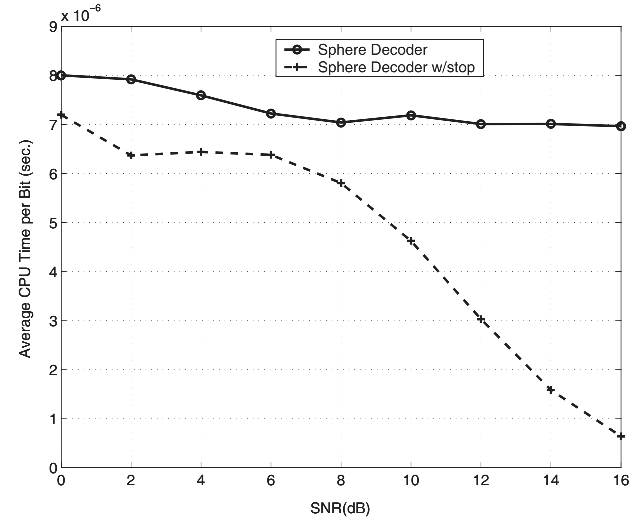

Fig. 2. The computational complexity of the different detectors, for the $N_T = N_R = 4$ MIMO system with QPSK, with and without the proposed stopping criterion, as a function of the SNR.

the transformation. In particular, a single detection error in the transformed space is not unlikely to translate into a large error in the original space, and even yield a point that lies outside the signal constellation. A similar observation was made in [6].

In summary we propose the following algorithm based on sphere decoding with a stopping criterion:

Proposed algorithm: (i) First run the lattice-reduction aided SIC detector to find the Babai point. (ii) If this results in a valid vector (all elements are inside the alphabet), then stop and deliver this vector as result. Otherwise, feed the so-obtained vector to the sphere decoder. □

IV. Simulation Results

We compare the performance and complexity with and without the proposed stopping criterion by means of Monte-Carlo simulation. An idealized model for a fast Rayleigh fading channel was used. Specifically, the channel matrix had i.i.d. circularly symmetric complex Gaussian elements with zero mean, and a new realization of the channel was drawn for each received vector. We simulated sufficient amounts of data to count 1000 bit errors at each SNR point.

Fig. 1 shows the BER versus SNR for an $N_T = N_R = 2$ and an $N_T = N_R = 4$ MIMO system with QPSK modulation, with and without the proposed stopping criterion.[2] We see that for the $N_T = N_R = 2$ system, the stopping criterion does not introduce any noticeable performance degradation at all relative to the MLD. The complexities with and without the stopping criterion for the $N_T = N_R = 4$ system are compared in Fig. 2. This figure shows the average CPU time per bit. (The CPU time measure was also used in [3]. Clearly different computers give different results, but the relative comparison should give a ballpark estimate.) At low SNR (SNR$\lesssim 8$ dB), the complexity reduction ratio is not significant. However at higher SNR, the proposed stopping criterion saves about 90% of the computations. We should point out that worst-case complexity is unaffected by the proposed stopping criterion (the worst-case complexity is dictated by the complexity of the sphere decoder, which is occasionally run even with the stopping criterion).

[2]Without the stopping criterion, the sphere decoder used in Fig. 1 is an exact implementation of the MLD.

V. Conclusion

A simple but effective stopping criterion for sphere decoding has been proposed. The idea is to first run a lattice-reduction aided SIC detector. If this results in a valid vector, with all elements within the symbol alphabet, the algorithm stops. Otherwise, it proceeds by running the sphere decoder. Our numerical results indicate significant average complexity savings compared to conventional sphere decoding.

References

[1] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comp.* vol. 44, pp. 463-471, 1995.

[2] A. K. Lenstra, H. W. Lenstra, Jr.,and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, no. 261, pp. 513-534, 1983.

[3] W. H. Mow, "Universal lattice decoding: principles and recent advances," *Wireless Communications and Mobile Computing*, special issue on coding and its applications in wireless CDMA systems, vol. 3, pp. 553-569, Aug. 2003.

[4] –, "Maximum likelihood sequence estimation from the lattice view-point," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1591-1600, Sept. 1994.

[5] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474-1484, Apr. 2005.

[6] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE GLOBECOM*, 2002.

[7] C. P. Schnorr and M. Euchner, "Lattice basis reduction: improved practical algorithms and solving subset sum problems," *Mathematical Programming*, vol. 66, pp. 181-191, 1994.

[8] A. M. Chan and I. Lee, "A new reduced-complexity sphere decoder for multiple antenna systems," in *Proc. IEEE ICC*, 2002.

[9] W. Zhao, and G. B. Giannakis, "Reduced complexity closest point decoding algorithms for random lattices," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 101–111, Jan. 2006.

[10] E. G. Larsson and J. Jaldén, "Soft MIMO detection at fixed complexity via partial marginalization," *IEEE Trans. Signal Processing*, vol. 56, pp. 3397–3407, Aug. 2008.