An Efficient Generalized Sphere Decoder for Rank-deficient MIMO Systems

Tao Cui and Chintha Tellambura,
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, AB, Canada T6G 2V4
Email: {taocui, chintha}@ece.ualberta.ca

***Abstract*— We derive a generalized sphere decoder (GSD) for rank-deficient Multiple Input Multiple Output (MIMO) systems using N transmit antennas and M receive antennas. This problem arises when $N > M$ or when the channel gains are strongly correlated. The upper triangular factorization of the Grammian yields an under-determined system and the standard sphere decoding (SD) fails. For constant modulus constellations, we modify the maximum likelihood (ML) cost metric so that the equivalent Grammian is rank N. The resulting GSD algorithm has significantly lower complexity than previous algorithms. A method to handle non-constant modulus constellations is also developed.**

Keywords– Sphere decoding, MIMO, Rank deficiency.

I. Introduction

Space-time wireless technology using multiple antennas provides very high data rates with low error probability. In this paper, we consider ML detection (MLD) for rank-deficient multi-antenna systems with N transmit antennas and M receive antennas, denoted by MIMO (N, M). MLD of the N symbols transmitted per symbol period over the MIMO channel is an integer least-squares (LS) problem, which is known to be NP-hard. Common, reduced-complexity algorithms such as V-BLAST [1] perform much worse than the true MLD. Sphere decoder (SD) algorithms [2]–[4] however achieve near exact ML performance by providing an efficient way for generating all candidate solutions that lie inside a hypersphere defined by the channel matrix and the received signal vector. In the high signal-to-noise ratio (SNR) region, the radius of the sphere can be chosen small enough so that only a few candidates are found inside the sphere. This search space is therefore drastically smaller than the MLD search space which consists of q^N points for a q-ary signal constellation. Thus, the number of operations (complexity) is polynomial in N in the high SNR region [5].

The QR decomposition of the $M \times N$ channel matrix or, equivalently, the Cholesky decomposition of the grammian is required. When $M \geq N$, these matrices have rank N and the standard SD algorithm is readily applicable. However, for $N > M$ (i.e., the downlink detection in MIMO systems) or for highly correlated MIMO channels [6], the matrix rank is less than the number of symbols to be estimated. A generalized SD (GSD) algorithm for this case has been derived in [4], [7]. The signal vector (to be estimated) and the channel matrix are partitioned and M symbols are estimated using SD for all q^{N-M} combinations of the remaining $N - M$ symbols. The complexity is exponential in $(N - M)$ regardless of the SNR. Recently, [8] also presented a fast GSD for rank-deficient MIMO systems by partitioning the subspace of the remaining $N-M$ symbols for an efficient search. In this paper, we utilize a different concept where the ML metric can be advantageously modified for a constant modulus constellation. We first derive the new algorithm assuming the transmit symbols to be constant modulus. We later show how MIMO systems with non-constant modulus constellations can be adapted so that our algorithm is applicable.

Notation: Bold symbols denote matrices or vectors. $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, $(\cdot)^\dagger$ denotes pseudo-inverse. The set of all complex $K \times 1$ vectors is denoted by $\mathcal{C}^K$. For 2^q-ary phase shift keying (PSK), the signal constellation $\mathcal{Q}_{2^q} = \{e^{j2\pi k/2^q}, k = 0, 1, \cdots, 2^q - 1\}$ and all PSK $N \times 1$ vectors are denoted by $\mathcal{Q}_{2^q}^N$. 4-ary PSK is commonly known as quadrature PSK (QPSK). If x and y are Gaussian with $E[x] = \mu_x$, $E[y] = \mu_y$ and $E\left[(x - \mu_x)^2\right] = E\left[(y - \mu_y)^2\right] = \sigma^2/2$, then $z = x + jy$ (where $j = \sqrt{-1}$) is Complex Gaussian. We write $z \sim \mathcal{CN}(\mu_x + j\mu_y, \sigma^2)$ in this case. The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$.

II. System Model

The complex baseband input-out relationship of a MIMO(N, M) system can be described in vector notation as

$$\mathbf{y} = \mathbf{H}\mathbf{s}_t + \mathbf{n} \tag{1}$$

where $\mathbf{s}_t \in \mathcal{Q}^N$ is the transmit vector, $\mathbf{y} \in \mathcal{C}^M$ is the receive vector, $\mathbf{H} = [h_{ij}]$ is the $M \times N$ channel matrix and $h_{ij} \sim \mathcal{CN}(0, 1)$ for $i = 1, \cdots, M$ and $j = 1, \cdots, N$ and $\mathbf{n} \in \mathcal{C}^M$ is the additive white Gaussian noise (AWGN) vector. The receiver has perfect knowledge of the random channel matrix $\mathbf{H}$, which remains constant over one or more symbol periods (slow fading). The transmitted vector over the linear MIMO channel (1) is ML detected as

$$\hat{\mathbf{s}}_t = \arg\min_{\mathbf{s} \in \mathcal{Q}^N} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2. \tag{2}$$

III. Sphere decoding algorithm

The original SD algorithm by Fincke and Phost [2] is proposed to solve the shortest vector problem (SVP). SD restricts the search space to the lattice points within a sphere instead of searching all the lattice points. Each time a valid lattice point is found, the search space is restricted further

0-7803-8521-7/04/$20.00 © 2004 IEEE

by updating the radius. Although its worst case complexity was shown to be exponential, FP has been widely used in the closest vector problem (CVP) due to its efficiency on many communication problems (refer to the references in the introduction and references therein).

In the context of communication systems, the maximum-likelihood estimator yields the following integer least squares problem

$$\min_{\mathbf{s}\in\mathcal{A}^M} \|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2 \tag{3}$$

where $\mathbf{y}\in\mathcal{R}^N$, $\mathbf{H}\in\mathcal{R}^{N\times M}$ and $\mathcal{A}=\{-(2P-1),\cdots,2P+1\}$ is a Pulse Amplitude Modulation (PAM) constellation of P elements. The primary difficulty with the optimization in (3) is that $\mathbf{s}$ is a *discrete* vector, and hence widely-available multidimensional continuous optimization strategies cannot be applied. A brute force algorithm must search all P^M possible values of $\mathbf{s}$. Hence, complexity is exponential in M (the number of variables to be estimated). Minimizing (3) is equivalent to

$$\begin{aligned}\hat{\mathbf{s}} &= \arg\min_{\mathbf{s}\in\mathcal{A}^M} \|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2\\ &= \arg\min_{\mathbf{s}\in\mathcal{A}^M} (\mathbf{H}^\dagger\mathbf{y}-\mathbf{s})^H\mathbf{R}^H\mathbf{R}(\mathbf{H}^\dagger\mathbf{y}-\mathbf{s})\\ &= \arg\min_{\mathbf{s}\in\mathcal{A}^M} (\mathbf{s}-\hat{\mathbf{s}})^H\mathbf{R}^H\mathbf{R}(\mathbf{s}-\hat{\mathbf{s}})\end{aligned} \tag{4}$$

where $\mathbf{R}$ is an upper triangular matrix such that $\mathbf{R}^H\mathbf{R}=\mathbf{H}^H\mathbf{H}$. If (3) is relaxed so that $\mathbf{s}\in\mathcal{R}^N$, the solution of the real least squares problem is $\hat{\mathbf{s}}=\mathbf{H}^\dagger\mathbf{y}$. In general, $\hat{\mathbf{s}}$ is not the optimal solution of (3). Furthermore, if the element of $\hat{\mathbf{s}}$ are quantized to the nearest element in $\mathcal{A}$, we obtain the so-called zero-forcing or decorrelation detectors. These are suboptimal. $\hat{\mathbf{s}}$ can be viewed as the center of the SD. The lattice point lies inside a hypersphere with radius r when

$$r^2 \geq (\mathbf{s}-\hat{\mathbf{s}})^H\mathbf{R}^H\mathbf{R}(\mathbf{s}-\hat{\mathbf{s}}) \tag{5}$$

We assume the initial radius r is large enough so that the hypersphere (5) contains the ML solution. Let the entries of R be denoted by r_{ij}, $i\leq j$. To this end, we assume that $N\geq M$ so that $\mathbf{H}^H\mathbf{H}$ has full rank (positive definite). The diagonal terms of $\mathbf{R}$ are non-zero ($r_{ii}\neq 0$). We will discuss the case $N<M$ further below. Eq. (5) can be rewritten as

$$\begin{aligned}&(\mathbf{s}-\hat{\mathbf{s}})^H\mathbf{R}^H\mathbf{R}(\mathbf{s}-\hat{\mathbf{s}})\\ &=\sum_{i=1}^{M} r_{ii}^2\left[s_i-\hat{s}_i+\sum_{j=i+1}^{M}\frac{r_{ij}}{r_{ii}}(s_j-\hat{s}_j)\right]^2\\ &=\sum_{i=1}^{M} r_{ii}^2\left[s_i-\rho_i\right]^2 \leq r^2\end{aligned} \tag{6}$$

where

$$\rho_i=\hat{s}_i-\sum_{j=i+1}^{M}\frac{r_{ij}}{r_{ii}}(s_j-\hat{s}_j) \tag{7}$$

Since each term in the above equation is nonnegative, a necessary condition for $\mathbf{s}$ to lie inside the hypersphere is that

$$r_{MM}^2\left[s_M-\hat{s}_i\right]^2\leq r^2 \tag{8}$$

or equivalently

$$\left\lceil \hat{s}_M-\frac{r}{r_{MM}}\right\rceil \leq s_M \leq \left\lfloor \hat{s}_M+\frac{r}{r_{MM}}\right\rfloor \tag{9}$$

$\lceil\cdot\rceil$ denotes the smallest integer greater than or equal to its argument. $\lfloor\cdot\rfloor$ denotes the largest integer less than or equal to its argument. For each candidate s_M satisfying the above bound, we define $r_{M-1}=r^2-(s_M-\hat{s}_i)^2$. When we look at s_{M-1}, we can get the following inequality

$$\begin{aligned}&r_{M-1,M-1}^2\left[s_{M-1}-\hat{s}_{M-1}+\frac{r_{M-1,M}}{r_{MM}}(s_M-\hat{s}_M)\right]^2\\ &+r_{MM}^2\left[s_M-\hat{s}_i\right]^2\leq r^2\end{aligned} \tag{10}$$

which leads to the following bound

$$\begin{aligned}&\left\lceil \hat{s}_{M-1}-\frac{r_{M-1,M}}{r_{MM}}(s_M-\hat{s}_M)-\frac{r_{M-1}}{r_{M_1,M_1}}\right\rceil \leq s_{M-1}\\ &\leq \left\lfloor \hat{s}_{M-1}-\frac{r_{M-1,M}}{r_{MM}}(s_M-\hat{s}_M)+\frac{r_{M-1}}{r_{M_1,M_1}}\right\rfloor\end{aligned} \tag{11}$$

The sphere decoder can choose a candidate for s_{M-1} from the above region. We can continue in the same process for s_{M-2} and so on. The bounds for s_m are

$$\lceil \rho_m-\xi_m\rceil \leq s_m \leq \lfloor \rho_m+\xi_m\rfloor \tag{12}$$

where $\xi_m=[r^2-\sum_{i=m+1}^{M} r_{ii}^2(s_i-\rho_i)^2]^{1/2}/r_{mm}$. If there is no lattice point within the region say s_m, the SD come back to s_{m+1} and choose another candidate value from the corresponding region for s_{m+1}. If SD reaches s_1, the SD finds a candidate lattice point $\mathbf{s}'$ within the hypersphere of radius r. SD checks the value of $\|\mathbf{y}-\mathbf{H}\mathbf{s}'\|^2$. If this value is less than r, we update the radius r which means the search space is limited by the new radius. The above process continues until no further lattice points is found within the hypersphere. The lattice point achieves the smallest value of (3) within the hypersphere is deemed as the ML solution. If no point in the sphere is found, the sphere is empty and the search fails. In this case, the initial search radius r must be increased and the search is restarted with a new squared radius. In [5], they analyzed the complexity based on the statistical property of the problem, which shows that the expected complexity of FP strategy is polynomial.

IV. Generalized Sphere Decoding Algorithm

Sphere decoding [2]–[4] provides an efficient recursive method for computing all $\mathbf{s}$ such that $\|\mathbf{R}(\mathbf{H}^\dagger\mathbf{y}-\mathbf{s})\|^2<r^2$. If $M\geq N$, $\mathbf{H}^H\mathbf{H}$ is full rank (positive definite). The SD is applicable due to the non-zero diagonal terms of $\mathbf{R}$.

However, if $M<N$, $\mathbf{H}^H\mathbf{H}$ is positive semidefinite. The Cholesky factor $\mathbf{R}$ of $\mathbf{H}^H\mathbf{H}$ is not full rank and only the first M rows of $\mathbf{R}$ are non-zero. The GSD developed in [4], [7] can solve this case. Our alternative idea is to solve $\|\mathbf{R}(\mathbf{H}^\dagger\mathbf{y}-\mathbf{s})\|^2+\alpha N<r_1^2$, where $r_1=\sqrt{\alpha N+r^2}$. Since the elements of $\mathbf{s}$ are of constant modulus, the product $\alpha\mathbf{s}^H\mathbf{s}$ is equal to αN

and we get an equivalent minimization problem as

$$\begin{aligned}\hat{\mathbf{s}} &= \arg\min_{\mathbf{s}\in Q^N} \|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2 + \alpha \mathbf{s}^H\mathbf{s} \\ &= \arg\min_{\mathbf{s}\in Q^N} \mathbf{y}^H\mathbf{y} - \mathbf{y}^H\mathbf{H}\mathbf{s} - \mathbf{s}^H\mathbf{H}^H\mathbf{y} + \mathbf{s}^H(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_N)\mathbf{s}.\end{aligned} \tag{13}$$

Let the positive definite matrix $\mathbf{G} = \mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N$ be Cholesky factorized as $\mathbf{G} = \mathbf{D}^H\mathbf{D}$, where $\mathbf{D}$ is an upper triangular matrix (i.e., $\mathbf{D} = [d_{ij}]$ and $d_{ij} = 0$ for $i > j$). Define $\rho = \mathbf{G}^{-1}\mathbf{H}^H\mathbf{y}$ and add $\rho^H\mathbf{D}^H\mathbf{D}\rho$ to the both sides of (13). An equivalent optimization problem can be represented as

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}\in Q^N} \|\mathbf{D}(\rho-\mathbf{s})\|^2. \tag{14}$$

Now the diagonal terms of $\mathbf{D}$ (rank N) are all non-zero and the original sphere decoding can be applied to (14). Note (2) and (14) are equivalent only for constant modulus signals. A flowchart of the proposed generalized sphere decoder, where C denotes the initial sphere radius, is given in Fig.4.

Remarks

- A tree of N levels and $\frac{q^{N+1}-1}{q-1}$ nodes can represent the entire solution space Q^N. The basic SD approach is to prune as many nodes of the tree as possible without eliminating the optimal solution. The key difference between the GSD (14) and that of [7] is as follows. First, the derived gram matrix $\mathbf{D}$ is positive definite, though the initial search radius in (14) is increased to $\sqrt{\alpha N + r^2}$. Second, the original GSD [7] cannot prune any nodes in the first $N-M$ levels while the proposed GSD (14) can prune from the first level. For large $N-M$, the new GSD can prune more nodes and hence has lower complexity.
- The complexity of the new GSD depends on the choice of α. If $\alpha = 0$, the new GSD reverts to the GSD in [7]. There exists an optimal α that leads to the lowest complexity. It appears impossible to derive a closed-form expression for the optimal α, which depends on the variance of the additive noise and $N-M$. In simulation, we fix $\alpha = 1$. We find even with this arbitrary choice of α the new GSD has less complexity than the previous GSDs [7], [8].
- Recently, Hassibi and Vikalo [5] showed that SD has polynomial complexity in the high SNR region with $M \geq N$. This is not true for any previous GSDs [7], [8]. The GSD (5) also has complexity exponential in $N-M$.
- The proposed GSD does not depend on the rank of matrix $\mathbf{H}$. Thus, the original SD algorithm or its derivatives can be used regardless of $N \leq M$ or $N > M$.
- As the proposed GSD exploits the constant modulus property of the transmit symbols and other algorithms do not make use of this information, we expect the proposed GSD to have low complexity.
- Like the GSD in [7], the proposed GSD can also be used for vector quantization and multiuser detection.

Next, we show how the GSD (14) can also detect M-ary Quadrature Amplitude Modulation (M-QAM). Any M-QAM ($M = 2^n$) constellation can be represented as a weighted sum of $n/2$ QPSK constellations when n is an even number [9]. That is, for $z \in M$-QAM and $z_i \in$ QPSK, $0 \leq i < n/2$, we have

$$z = \sum_{i=0}^{\frac{n}{2}-1} 2^i \left(\frac{\sqrt{2}}{2}\right) z_i. \tag{15}$$

For brevity, we only show how to decompose 16QAM so that our algorithm is applicable. Other M-QAM cases can be derived similarly. Using (15), the 16QAM transmit vector $\mathbf{x}$ can be expressed as

$$\mathbf{s} = \sqrt{2}\mathbf{s}_1 + \frac{\sqrt{2}}{2}\mathbf{s}_2 \tag{16}$$

where $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{Q}_4^N$. Eq. (1) can then be represented as

$$\begin{aligned}\mathbf{y} &= \begin{bmatrix} \sqrt{2}\mathbf{H} & \frac{\sqrt{2}}{2}\mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} + \mathbf{n} \\ &= \tilde{\mathbf{H}}\tilde{\mathbf{s}} + \mathbf{n}.\end{aligned} \tag{17}$$

That is, an MIMO(N, M) system with 16QAM is equivalent to MIMO($2N, M$) with QPSK. The new GSD (14) can be used with (8).

V. Numerical Results

We consider a frequency flat fading MIMO channel defined by $\mathbf{H} = [h_{ij}]$. We incorporate the so-called Schnorr-Euchner strategy [3], [4] into the GSD in [7] and the new GSD with $\alpha = 1$. Fig. 1 compares the computational complexity of the proposed sphere decoder with those of [7] and [8] with different depths for fixed $N = 15$ and QPSK at an SNR of 20dB. We simulate this system using MATLAB V5.3 on a PC with an Intel Pentium-4 processor at 1.8GHz. The matlab command "flops" is used to count the number of flops. Only the flops of the search algorithm are counted without accounting for the preprocessing stage. The Depth i GSD in Fig. 1 denotes the GSD in [8]. The new sphere decoder has significantly lower decoding complexity than the GSD in [7] and [8], especially when the difference between N and M is large. For $N-M = 8$, the average number of flops are 6.76×10^6 for the GSD in [7] and 2.35×10^5 for GSD (5). Fig. 2 shows the complexity comparison between the new GSD with that of [7] for fixed $N = 8$ and 16QAM at an SNR of 20dB. Even though the new GSD enlarges the dimension (17), when $N-M$ is more than 3, the new GSD reduces the complexity significantly.

Fig. 3 shows the performance of a MIMO system with 4 antennas at the base station and 3 antennas at the mobile. The two sets of antennas serve as both transmitting and receiving. We consider both QPSK and 16QAM modulations. The proposed GSD is used for both up and down links. The uplink detection is compared with that of V-BLAST optimal order detection [1], which can not be used for the downlink. The proposed GSD achieves the ML performance and outperforms V-BLAST detection.

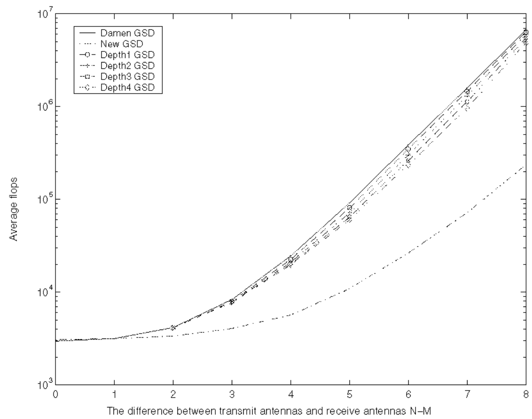

Fig. 1. The average complexity versus $N - M$ in a BPSK MIMO system

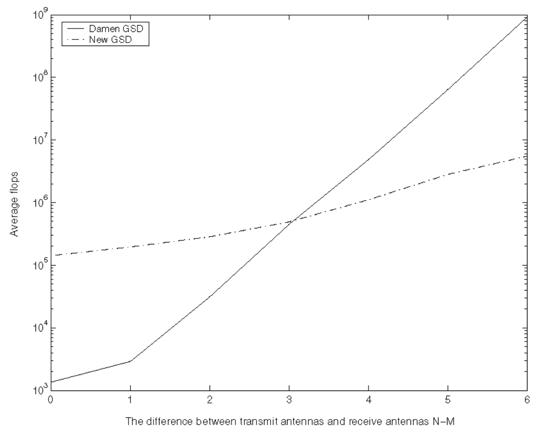

Fig. 2. The average complexity versus $N - M$ in a 16QAM MIMO system

VI. Conclusion

For rank-deficient multiple antenna systems, we have developed a new sphere decoder assuming constant modulus signalling alphabets. This assumption is not limiting because a QAM constellation can be represented as a linear sum of QPSK points. The new sphere decoder is ML and has significantly lower decoding complexity than the GSD [4], [7] and the improved GSD [8]. The proposed GSD can also be extended to frequency selective MIMO channels.

Acknowledgment

The authors would like to thank Pranav Dayal from the Department of Electrical & Computer Engineering, University of Colorado at Boulder for providing simulation code for the GSD in [8].

References

[1] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, Jan. 1999.

[2] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, inlcuding a complexity analysis," *Math. Computation*, vol. 44, pp. 463–471, Apr. 1985.

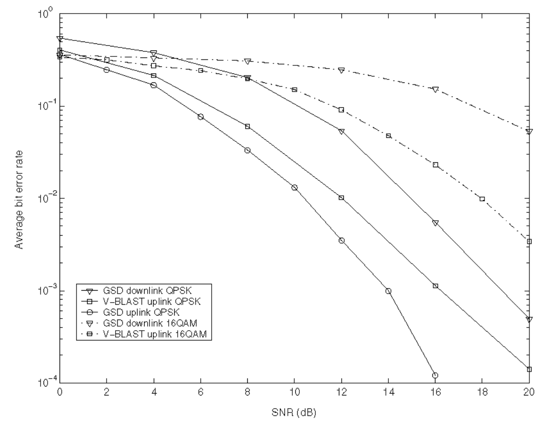

Fig. 3. Uplink and downlink BER over a Rayleigh fading MIMO channel

[3] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Programming*, vol. 66, pp. 181–191, 1994.

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

[5] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm: Part I, the expected complexity," *IEEE Transactions on Signal Processing*, submitted.

[6] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: models and performance prediction," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[7] M. Damen, K. Abed-Meraim, and J.-C. Belfiore, "Generalised sphere decoder for asymmetrical space-time communication architecture," *Electronics Letters*, vol. 36, no. 2, pp. 166–167, Jan. 2000.

[8] P. Dayal and M. K. Varanasi, "A fast generalized sphere decoder for optimum decoding of under-determined MIMO systems," *Proc. of 41st Annual Allerton Conf. on Comm. Control, and Comput.,*, Oct. 2003.

[9] B. Tarokh and H. Sadjadpour, "Construction of OFDM M-QAM sequences with low peak-to-average power ratio," *IEEE Transactions on Communications*, vol. 51, no. 1, pp. 25 – 28, Jan. 2003.

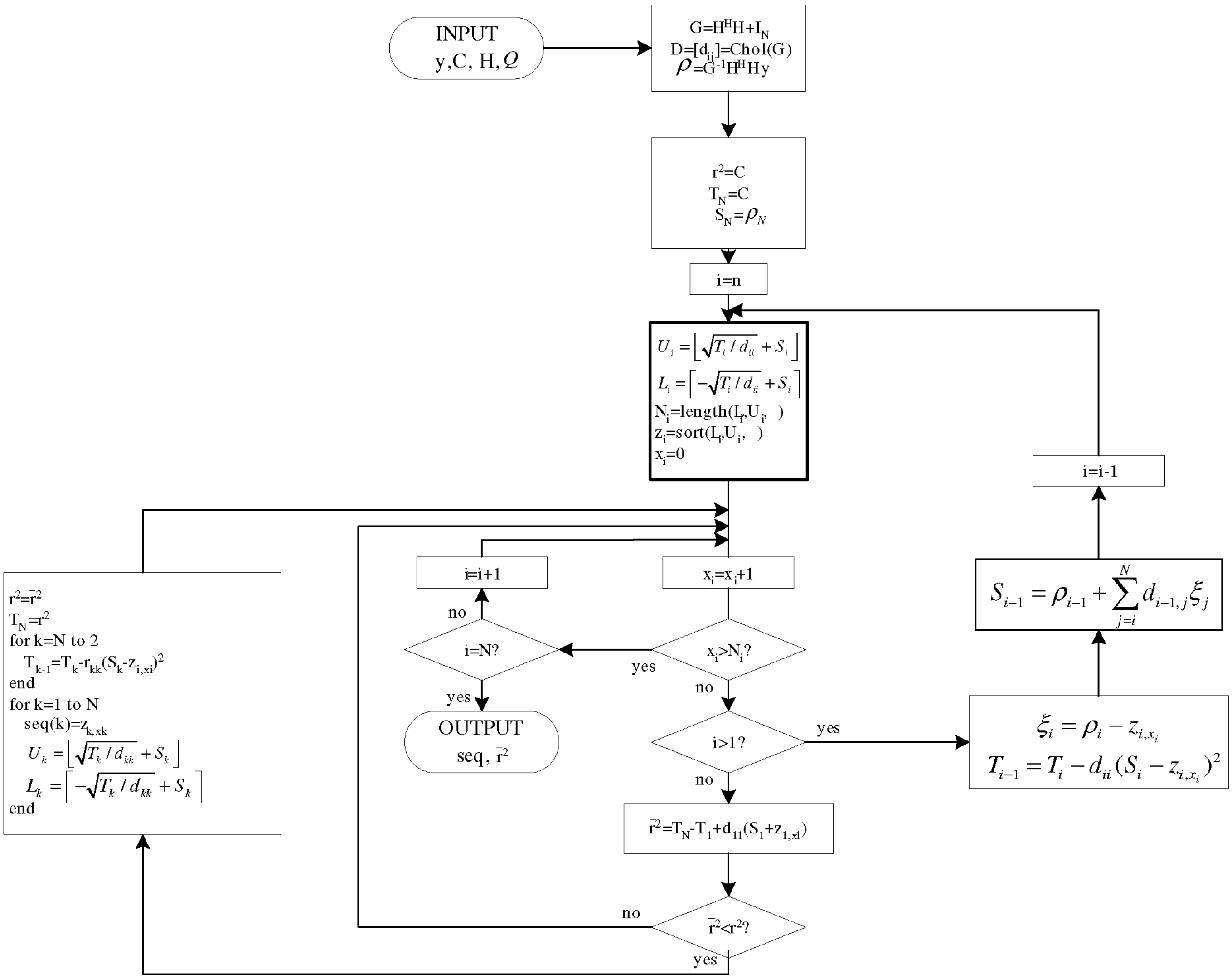

Fig. 4. Flowchart of the generalized sphere decoder algorithm