Polynomial-Constrained Detection Using a Penalty Function and a Differential-Equation Algorithm for MIMO Systems

Tao Cui, *Student Member, IEEE,* and Chintha Tellambura, *Senior Member, IEEE*

***Abstract*—In this letter, we develop a family of approximate maximum-likelihood (ML) detectors for multiple-input multiple-output systems by relaxing the ML detection problem using constellation-specific polynomial constraints. The resulting constrained optimization problem is solved using a penalty function approach. Moreover, to escape from the local minima, which improves the detection performance, a differential equation algorithm using classical mechanics is proposed. Simulation results show that the polynomial constrained detector performs better than least-squares (LS) detector.**

***Index Terms*—Data detection, maximum likelihood (ML), multiple-input multiple-output (MIMO).**

I. Introduction

THE enormous capacity and remarkably high spectral efficiencies promised by multiple-input multiple-output (MIMO) wireless systems in rich scattering multipath environments have stimulated notable research endeavors. Along with the theoretical studies, detection algorithms have been proposed to exploit the capacity provided by MIMO systems. The vertical Bell Laboratories layered space time (V-BLAST) architecture is one of the first to be developed [1].

The V-BLAST detection algorithm or equivalently zero-forcing (ZF) decision feedback detection (DFD) is based on nulling and interference cancellation with optimal ordering. However, the V-BLAST detector performs much worse than the maximum-likelihood detector (MLD), which achieves the minimum error probability for independent and identically distributed (i.i.d.) random data symbols. In [2], the sphere decoder (SD), offering near-optimal performance, is proposed, which attains low complexity in high signal-to-noise ratio (SNR). Its complexity is nevertheless high in low SNR or for large systems. The large gap in both performance and complexity between the MLD and the V-BLAST detector has motivated the search for alternative detectors.

The relaxation approach has previously been applied to code-division multiple-access (CDMA) and orthogonal frequency division multiplexing (OFDM)/spatial-division multiple-access (SDMA) systems. In [3], a generalized minimum mean-square error (GMMSE) detector is proposed for CDMA, where the binary phase-shift keying (BPSK) vectors are relaxed to lie inside the smallest hypersphere that contains the unit hypercube. A tighter relaxation for OFDM/SDMA systems is used in [4] by restricting the binary vectors to be on the hypersphere rather than the inside. In [5], semidefinite relaxation (SDR) has been applied to CDMA systems with BPSK, and SDR has been extended to general M-PSK constellations in [6]. These contributions motivate further search for tighter and universal relaxations applicable for any constellation.

In this letter, we develop a family of approximate MLDs for MIMO systems by relaxing the ML detection problem using constellation-specific polynomial constraints. The ML MIMO detection problem is hence reformulated as an equality-constrained minimization problem. It is solved using a penalty function with the Newton method. Since the Newton method may be trapped by local minima, a differential-equation (DE) algorithm using classical mechanics is proposed to improve the detection performance.

Notation: Bold symbols denote matrices or vectors. $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ denote transpose, conjugate transpose, and conjugate, respectively. $(\cdot)^\dagger$ denotes pseudo-inverse. $\Re\{x\}$ and $\Im\{x\}$ denote the real part and imaginary part of x, respectively. $\|(\cdot)\|^2$ is the two-norm of $(\cdot)$. The set of all complex $K \times 1$ vectors is denoted by $\mathcal{C}^K$. A circularly complex Gaussian variable with mean μ and variance σ^2 is denoted by $z \sim \mathcal{CN}(\mu, \sigma^2)$.

II. MIMO System Model

A MIMO system with n transmit antennas and m receive antennas is considered. We focus on spatial multiplexing systems, where the signals are spatially independent rather than jointly encoded. Source data are mapped into complex symbols from a finite constellation $\mathcal{Q}$ of size s and $\mathcal{Q} = \{q_1, q_2, \ldots, q_s\}$. The input data stream is demultiplexed into n substreams that are simultaneously transmitted through the n antennas over a rich scattering channel. We assume that the MIMO channel is flat fading. Each receive antenna receives signals from all the n transmit antennas. The discrete-time baseband received signals can thus be written as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x} = [x_1, \ldots, x_n]^T$, $x_i \in \mathcal{Q}$ is the transmitted signal vector, $\mathbf{r} = [r_1, \ldots, r_m]^T$, $r_i \in \mathcal{C}$ is the received signal vector, $\mathbf{H} = [h_{i,j}] \in \mathcal{C}^{m \times n}$ is the channel matrix, and $\mathbf{n} = [n_1, \ldots, n_m]^T$, $n_i \sim \mathcal{CN}(0, \sigma_n^2)$ is an additive white Gaussian noise (AWGN) vector. The components of $\mathbf{n}$ are thus

Manuscript received July 25, 2005; revised November 8, 2005. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, in part by the Informatics Circle of Research Excellence, and in part by the Alberta Ingenuity Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jitendra K. Tugnait.

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: taocui@ece.ualberta.ca; chintha@ece.ualberta.ca).

Digital Object Identifier 10.1109/LSP.2005.862619

i.i.d. complex Gaussian. Likewise, the elements of $\mathbf{H}$ are i.i.d. with $h_{i,j} \sim \mathcal{CN}(0,1)$, $\forall i, j$. We assume that the channel is perfectly known to the receiver and $n \leq m$. If $n > m$, we can readily transform the rank-deficient system into a full-rank system using the algorithm in [7]. Note that the linear model (1) can also be applied to certain spatially coded MIMO systems, single antenna systems over time and frequency-selective channels, intersymbol interference (ISI) channels, and multiuser systems. Consequently, the relaxation approach proposed in this letter can also be used for such applications as multiuser detection for CDMA.

Assuming uncorrelated noise and transmitted signals, the MLD that minimizes the average error probability is given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{Q}^n} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2. \tag{2}$$

Equation (2) is an NP-hard problem, and the complexity of brute-force search is exponential in n.

III. Polynomial Constraint and Penalty Function Method

Due to the finite alphabet nature of $\mathcal{Q}$, each $x_i \in \mathcal{Q}$ satisfies the following equation:

$$f(x_i) = \prod_{k=1}^{s} (x_i - q_k) = 0. \tag{3}$$

The ML detection problem (2) can thus be relaxed as

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{C}^n} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \\ &\text{s.t. } f(x_i) = \prod_{k=1}^{s} (x_i - q_k) = 0, \quad i = 1, \ldots, n. \end{aligned} \tag{4}$$

Both the objective function and constraints are polynomial in $\mathbf{x}$. For example, for BPSK, $f(x) = x^2 - 1$. To avoid the complex operation, equation (4) can be transformed into a real problem as

$$\begin{aligned} &\min_{\tilde{\mathbf{x}}\in\mathcal{R}^{2n}} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 \\ &\text{s.t. } g^r(\Re\{x_i\}, \Im\{x_i\}) = 0, \text{ and } g^i(\Re\{x_i\}, \Im\{x_i\}) = 0 \\ &\qquad i = 1, \ldots, n \end{aligned} \tag{5}$$

where

$$\tilde{\mathbf{r}} = \begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix}, \quad \tilde{\mathbf{x}} = \begin{bmatrix} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{bmatrix} \tag{6}$$

and

$$\tilde{\mathbf{H}} = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix}. \tag{7}$$

$g^r(\Re\{x_i\}, \Im\{x_i\}) = \Re\{f(x_i)\}$ and $g^i(\Re\{x_i\}, \Im\{x_i\}) = \Im\{f(x_i)\}$, which are also polynomial in $\Re\{x_i\}$ and $\Im\{x_i\}$. Specifically for decouplable constellations, i.e., QAM, (5) can be simplified as

$$\begin{aligned} &\min_{\tilde{\mathbf{x}}\in\mathcal{R}^{2n}} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 \\ &\text{s.t. } g(\tilde{x}_i) = 0, \quad i = 1, \ldots, 2n \end{aligned} \tag{8}$$

where $\tilde{x}_i$ is the ith element of $\tilde{\mathbf{x}}$. For example, for 16-QAM, $g(x) = (x+3)(x-3)(x+1)(x-1) = x^4 - 10x^2 + 9$.

We next show how to apply the penalty function method to (8). Equation (5) can be solved similarly. The most common method is the one that associates a penalty that is proportional to the square of the constraints. So, (8) is replaced by

$$\min_{\tilde{\mathbf{x}}\in\mathcal{R}^{2n}} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 + \frac{1}{2}c\sum_{i=1}^{2n} |g(\tilde{x}_i)|^d \tag{9}$$

where the positive scalar c controls the magnitude of the penalty, and d controls the acceleration of penalty. If $d = 2$, it reduces to the usual penalty function in [8]. In the following, we choose $d = 2$. Since (9) is a polynomial in $\tilde{\mathbf{x}}$, the Hessian matrix of (9) can be computed in a close form. The well-known Newton or quasi-Newton method can be used to solve (9). The initial point for the Newton method can be chosen as the least-squares (LS) or minimum mean-square error (MMSE) solution.

It seems logical to choose very large c to ensure that no constraint is violated. However, large c may lead to numerical difficulties or ill-conditioning, and the search would be trapped by the local minima corresponding to, for example, the LS or MMSE solution. The minimization thus should be initialized with a relatively small c, and c would be increased gradually. A typical value for $c^{(k+1)}/c^{(k)}$ is 5 [8]. If $\tilde{\mathbf{x}}$ is the true transmit vector, $\|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2$ is a sum of squares of noise terms. Consequently, the initial c can be set to a value proportional to σ_n^2.

To overcome the ill-conditioning, the penalty function method can be combined with the Lagrange multipliers. The so-called augmented Lagrangian function [8] is defined as

$$L(\tilde{\mathbf{x}}, \lambda_1, \ldots, \lambda_{2n}) = \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 + \sum_{i=1}^{2n} \lambda_{ig}(\tilde{x}_i) + \frac{1}{2}c\sum_{i=1}^{2n} |g(\tilde{x}_i)|^2. \tag{10}$$

The initial λ_i can be set to $\lambda_i^{(0)} = 0$. After minimizing (10), [8] suggests updating λ_i as

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} - c^{(k)} g\left(\tilde{x}_i^{(k)}\right) \tag{11}$$

where $\tilde{x}_i^{(k)}$ is the estimate of $\tilde{x}_i$ in the kth iteration.

To avoid the trap of local minima using the Newton method, we propose a DE algorithm inspired by the classical mechanics to improve the detection performance. Let the function to be minimized in (9) be denoted as $F(\mathbf{x})$, where $\tilde{}$ is omitted for brevity. We associate the following second-order DE with (9) [9]:

$$\mu\frac{d^2\mathbf{x}}{dt^2} + \beta(t)\frac{d\mathbf{x}}{dt} + \nabla F(\mathbf{x}) = \mathbf{0} \tag{12}$$

where μ is a positive constant, $\beta(t) > 0$ is a function, and $\nabla F(\mathbf{x})$ is the gradient of $F(\mathbf{x})$.

Equation (12) represents Newton's second law for a particle of mass μ moving in $\mathbb{R}^n$, subject to the force $-\nabla F(\mathbf{x})$ given by the potential $F(\mathbf{x})$ and the friction $-\beta(t)d\mathbf{x}/dt$, where $\beta(t)$ is the time-varying friction coefficient.

Let the initial values for (12) be

$$\mathbf{x}(0) = \mathbf{x}_0, \quad \frac{d\mathbf{x}}{dt} = \mathbf{w}_0. \tag{13}$$

Typically, $\mathbf{x}_0$ is chosen to be the LS or V-BLAST solution, and $\mathbf{w}_0 = \mathbf{0}$. We numerically integrate the DE (12) with the initial

conditions (13). Equation (12) can be rewritten as first-order equations as

$$\frac{d\mathbf{x}}{dt} = \mathbf{w}, \quad \mu\frac{d\mathbf{w}}{dt} = -\beta(t)\mathbf{w} - \nabla F(\mathbf{x}). \tag{14}$$

Applying numerical integration to (14), we have

$$\begin{aligned} h_n\mathbf{w}_{n+1} &= \mathbf{x}_{n+1} - \mathbf{x}_n \\ \mu(\mathbf{w}_{n+1} - \mathbf{w}_n) &= -h_n\beta_n\mathbf{w}_n - h_n\nabla F(\mathbf{x}_n) \end{aligned} \tag{15}$$

where h_n is the time integration step at time instant t_n. Note that the A-stable linearly implicit method [9] can be used. However, it needs to compute the Jacobian of $F(\mathbf{x})$, which has high complexity.

At each time t_n, we save the current potential $F(\tilde{\mathbf{x}}_n)$, kinetic $\mu\mathbf{w}_n^2/2$, and the corresponding location $\mathbf{x}_n$. The integration is stopped when the particle stops moving or the maximum kinetic of K time steps is less than a threshold. The point with the minimum potential on the trajectory is output as the solution for (9). If the maximum of $|g(\tilde{x}_i)|^2$ is larger than a threshold, c in (9) is increased gradually, and we search again until the condition is satisfied. Due to the existence of the inertial term or the second-order term in (12), local minima of $F(\mathbf{x})$ may be overpassed. However, this algorithm does not guarantee the global minimum.

Given the initial value β_0 and $\beta_m > \beta_0$, the friction coefficient β_n is kept constant for the first ten steps and then is doubled at each step until $2\beta_n > \beta_m$. If $2\beta_n > \beta_m$, β_n is set to β_m, and it remains constant during the rest of the integration.

Given the initial value h_0, the value of h_n is updated by a factor of γ. If the total mechanical energy E_n is increased, we choose $\gamma < 1$ or $\gamma > 1$. In the simulation, we choose γ from 1.6 or 0.6.

To keep the total complexity of the DE algorithm constant, we can set the maximum number of time integration steps $N_{\max}$ in the algorithm. The integration stops after reaching $N_{\max}$.

To ensure a tight relaxation, we utilize the maximum and minimum absolute values of the constellation elements $\rho_{\max}$ and $\rho_{\min}$, resulting in additional constraints

$$|x_i| \le \rho_{\max}^2, \quad |x_i| \ge \rho_{\min}^2. \tag{16}$$

Hence, we can further add two penalty terms to (9)

$$\left\{\min\left(0, \rho_{\max}^2 - |x_I|\right)\right\}^2 + \left\{\min\left(0, |x_i| - \rho_{\min}^2\right)\right\}^2. \tag{17}$$

This may give tighter relaxation and result in better performance.

IV. Simulation Results

The error rates of our proposed constrained detector is simulated for an MIMO system over a flat Rayleigh fading channel. We assume the receiver has perfect channel state information (CSI), and the initial value of c is chosen proportional to the noise variance. Our polynomial constrained detector is denoted as PCD. The detector without the maximum number of time steps constraint is denoted as PCD-Op , or it is denoted as PCD-X, where X is $N_{\max}$. The LS detector and the SD are used as benchmark detectors. In PCD, we set $\mu = 1$, $\beta_0 = 0.1$, $\beta_m = 3$, and $h_0 = 0.03$.

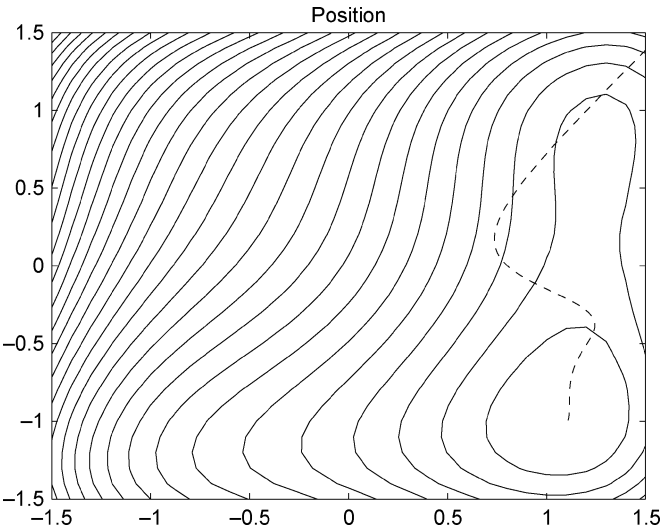

Fig. 1. Trajectory of a particle for a 2 × 2 MIMO system with BPSK and 5 dB.

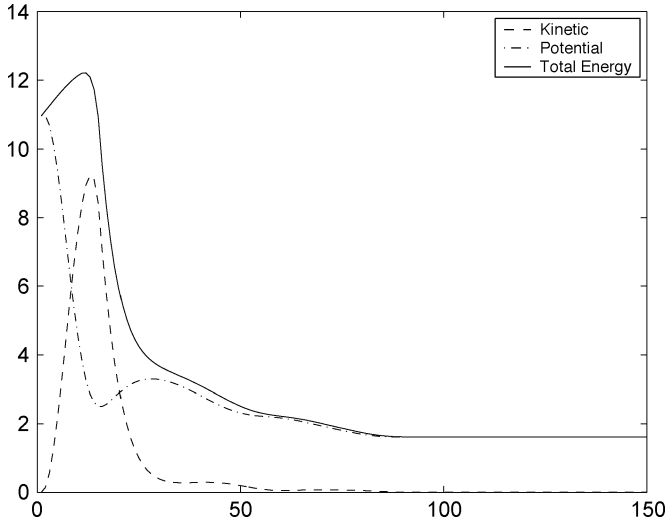

Fig. 2. Energy of a particle as a function of time steps for a 2 × 2 MIMO system with BPSK and 5 dB.

We first show a simple example of 2 × 2 BPSK system at an SNR of 5 dB. Fig. 1 shows the trajectory of the particle (dash line) on a contour graph. The initial point is set to $\mathbf{x}_0 = [1.5, 1.5]^T$. Clearly, we can see the particle is not trapped by the local minimum around [1,1], and it stops at the global minimum around $[1, -1]$. Fig. 2 shows the kinetic, potential, and the total mechanical energies as a function of the time steps. From the potential curve, we find the particle overpasses a local minimum at time step 20. The kinetic energy decreases toward zero as the particle reaches the global minimum.

Fig. 3 shows the BER performance of different detectors in a BPSK modulated system with eight transmit and eight receive antennas. The initial values for $\mathbf{x}_0$ are chosen as the LS solution. Our PCD has a significant performance gain over both V-BLAST and LS. At BER $= 10^{-2}$, PCD-Op has a 4-dB gain over V-BLAST, and the performance loss over ML is only 2.2 dB. When PCD-40 is used to achieve constant complexity, the

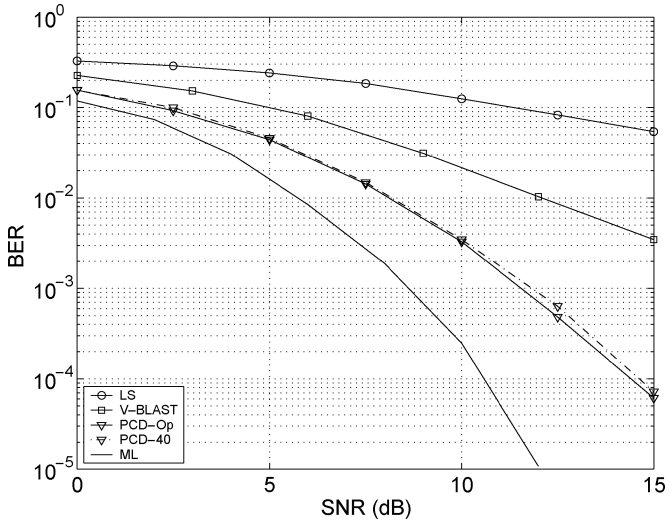

Fig. 3. Performance comparison of different detectors in an 8 × 8 MIMO system with BPSK.

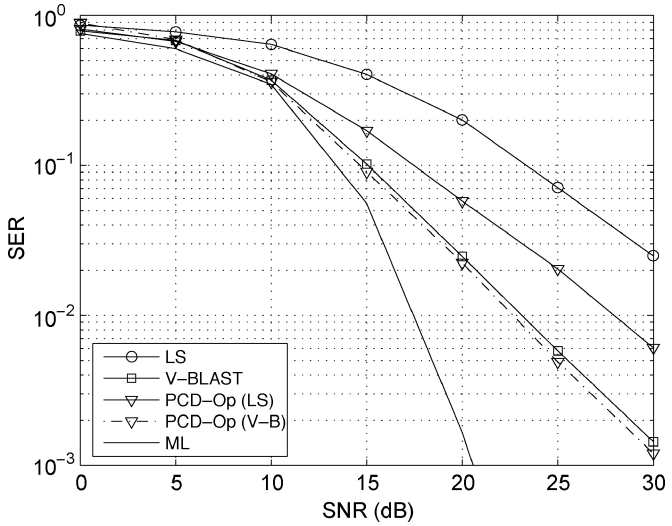

Fig. 4. Performance comparison of different detectors in an 4 × 4 MIMO system with 16QAM.

performance loss over PCD-Op is less than 0.1 dB at BER $= 10^{-2}$. However, the complexity of PCD-40 is roughly 15% of that of ML search.

The symbol error rate (SER) of different detectors for a 4×4 system with 16QAM is shown in Fig. 4. PCD-Op (LS) denotes our PCD-Op with bfx_0 chosen as the LS solution. PCD-Op (V-B) denotes our PCD-Op with $\mathbf{x}_0$ chosen as the V-BLAST solution. Our PCD-Op (LS) has a 5.8-dB gain over LS at SER $= 1 \times 10^{-1}$. The gain reduces to 0.5 dB when V-BLAST is used for initialization at SER $= 1 \times 10^{-2}$. However, the performance improvement is reduced compared to that of BPSK. In high SNR, the performance gap between the PCD-Op and the MLD is large. However, the complexity of our PCD is only 1% of that of ML search. The average number of flops of V-BLAST is 12 158, while it is $C_{\text{initial}} + 558N_{\max}$, where C_{initial} is the complexity of finding the initial point in PCD, i.e., $C_{\text{initial}} = 12\,158$ for V-BLAST and $C_{\text{initial}} = 1524$ for LS. Therefore, the complexity of PCD in each iteration is much less than C_{initial} for V-BLAST. Even though PCD performance is not as good as the MLD in high SNR, it is computationally efficient in low SNR and can be readily parallelized, which is appealing for practical application. The diversity order (i.e., the negative slope of the BER curve in high SNR) of the PCD-Op appears to be one. When the constellation size is large, the performance gain by using our PCD decreases since the DE algorithm may also be trapped by local minima (but not the initial one).

V. Conclusion

In this letter, we have proposed an approximate relaxation approach for the maximum likelihood detection problem for uncoded MIMO systems. Using constellation-specific polynomial constraints, the detection problem was reformulated as an equality-constrained optimization problem. It was solved using a generalized penalty function method, along with the classical Newton method. Since the Newton method may be trapped by local minima, a DE algorithm inspired by the classical mechanics has been proposed. Simulation results show that our proposed relaxation detector always outperforms the LS detector, but its performance relative to that of the V-BLAST detector depends on the constellation size. Of course, the complexity of the proposed detector is significantly less than that of exact ML via exhaustive search. The proposed detector may also work for spatially coded MIMO systems, single antenna systems over time and frequency-selective channels, ISI channels, multiuser systems, and others. It would be interesting to study how it performs in those applications.

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," *Electron. Lett.*, vol. 35, no. 1, pp. 14–15, Jan. 1999.

[2] M. O. Damen, A. Chkeif, and J. C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161–163, May 2000.

[3] A. Yener, R. D. Yates, and S. Ulukus, "CDMA multiuser detection: A nonlinear programming approach," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1016–1024, Jun. 2002.

[4] S. Thoen, L. Deneire, L. V. der Perre, M. Engels, and H. D. Man, "Constrained least squares detector for OFDM/SDMA-based wireless networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 129–140, Jan. 2003.

[5] W. K. Ma, T. Davidson, K. M. Wong, Z.-Q. Luo, and P.-C. Ching, "Quasimaximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912–922, Apr. 2002.

[6] W.-K. Ma, P. C. Ching, and Z. Ding, "Semidefinite relaxation based multiuser detection for M-ary PSK multiuser systems," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2862–2872, Oct. 2004.

[7] T. Cui and C. Tellambura, "An efficient generalized sphere decoder for rank-deficient MIMO systems," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 423–425, May 2005.

[8] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic, 1982.

[9] F. Aluffi-Pentini, V. Parisi, and F. Zirilli, "A differential-equations algorithm for nonlinear equations," *ACM Trans. Math. Softw.*, vol. 10, no. 3, pp. 299–316, Sep. 1984.