Box Constrained Optimization for Detection of 16-QAM Signaling in MIMO Channels

Soonchul Park, Hongchao Zhang, William W. Hager, and Dong Seog Han, Member, IEEE,

Abstract—We develop a computationally efficient approximation of the maximum likelihood (ML) detector for 16 quadrature amplitude modulation (16-QAM) in multiple-input multiple-output (MIMO) systems. The detector is based on solving a convex relaxation of the ML problem by an affine-scaling cyclic Barzila-Borwein method for box constrained relaxation (BCR) problem. Simulation results in a random MIMO system show that this proposed approach outperforms the conventional decorrelator detector including the semidefinite relaxation (SDR) detector. We also note that the complexity of the proposed approach is less than that of those detectors. In the case of 8 antennas and 4 users, about 99% fewer computations are required when compared to the SDR and ML detectors.

Index Terms—Maximum likelihood (ML) detection, MIMO systems, affine-scaling, cyclic Barzila-Borwein (CBB) method, box constrained optimization.

I. INTRODUCTION

MULTIPLE-input multiple-output (MIMO) systems divide a data stream into multiple unique streams, each of which is modulated and transmitted through a different radio-antenna chain at the same time in the same frequency channel. By taking advantage of multipath, reflections of the signals, each MIMO receive antenna-radio chain is a linear combination of the multiple transmitted data streams. The data streams are separated at the receiver using MIMO algorithms that rely on estimates of all channels between each transmitter and each receiver. Hence, detection in MIMO systems is one of the fundamental problems.

The optimal model to minimize the joint probability error of detecting all the symbols simultaneously is the maximum likelihood (ML) detector [10]. It can be implemented using a brute-force search over all of the possible transmitted vectors or using more efficient search algorithms, for instance, the sphere decoder [1]. However, it has been shown that the expected computational complexity is exponential and impractical, for instance, there are $16^K$ vectors to be evaluated for $K$ users with 16-QAM. Consequently, there has been much interest in implementing suboptimal detection algorithms. The most common suboptimal detectors are the linear receivers, i.e., the matched filter (MF), the decorrelator or zero forcing (ZF), the minimum mean-squared error (MMSE) detectors, decision feedback equalization (DFE), and the semidefinite relaxation (SDR) detector. There are many other suboptimal detection schemes ranging from lattice-based algorithms, alternating variable methods, to expectation maximization and many more [10].

The ZF algorithm is a straightforward approach to signal detection. The receiver with the ZF detector uses the estimated channel matrix to detect the transmitted signal as follows:

$$\hat{s} = (H^H H)^{-1} H^H y = H^+ y,$$

where $y$ is the received signal, $H^H$ and $H^+$ denote the Hermitian conjugate and the pseudo inverse respectively, and $s$ is the transmitted symbols. Then each element of $\hat{s}$ is moved to the nearest constellation point.

One of the promising suboptimal detection strategies is the SDR detector [11]. The main reason for the high computational complexity of the ML detector is due to the fact that it is a non-convex combinatorial optimization problem. The approach of SDR is to formulate the ML problem in a higher dimension and then relax the non-convex constraints. Even though the SDR method solves the ML problem in polynomial time, it actually is not really practical for 16-QAM.

We propose a suboptimal detection algorithm, box constrained relaxation (BCR) which is much faster than the existing algorithms for ML problem. We apply the proposed algorithm for detection of 16-QAM signaling in MIMO system and compare it with the ZF detector and the SDR detector in [11].

The rest of this paper is organized as follows. In section 2, the 16-QAM ML detection problem is presented along with our quadratic programming relaxations involving both box constraints and a penalization term. In Section 3, we introduce the affine-scaling cyclic Barzilai-Borwein (AS_CBB) method we use to solve the quadratic programming problems. The simulation results are provided to compare the performance of the proposed box constrained relaxation detector with other detectors in terms of symbol-error-rate (SER) in Section 4. Conclusions are made in Section 5.

II. 16-QAM ML DETECTION AND BOX OPTIMIZATION

Consider the standard MIMO channel

$$y = Hs + w$$

where $y$ is the received signal of length $N$, $H$ is an $N \times K$ channel matrix, $\bar{s}$ is the length $K$ vector of transmitted
symbols, and $\bar{w}$ is a length $N$ complex random vector with normal distribution of zero mean and covariance $\sigma^2I$. The symbols of $\bar{s}$ belongs to some known complex constellation. In this paper, we consider a 16-QAM constellation, i.e., the real part and the imaginary part of $\bar{s}_i$ for $i = 1, \ldots, K$ belong to the set $\{\pm 1, \pm 3\}$.

In order to avoid handling complex-valued variables, it is more convenient to use the following decoupled model:

$$y = Hs + w$$  \hspace{1cm} (2)

where

$$y = \begin{bmatrix} Re\{\bar{y}\} \\ Im\{\bar{y}\} \end{bmatrix} ; s = \begin{bmatrix} Re\{\bar{s}\} \\ Im\{\bar{s}\} \end{bmatrix} ; w = \begin{bmatrix} Re\{\bar{w}\} \\ Im\{\bar{w}\} \end{bmatrix} ;$$

$$H = \begin{bmatrix} Re\{H\} & -Im\{H\} \\ Im\{H\} & Re\{H\} \end{bmatrix}$$

Using these definitions, we can formulate the ML-detector of the transmitted symbols as

$$\text{ML: } \begin{cases} \min \|y - Hs\|^2 \\ \text{subject to } s_i \in \{\pm 1, \pm 3\}, i = 1, \ldots, 2K. \end{cases}$$  \hspace{1cm} (3)

The ML detector is a combinatorial problem and can be solved in a brute-force fashion by searching over all the $4^{2K} = 16^K$ possibilities. Clearly, as $K$ increases, the brute-force search becomes prohibitively expensive.

As an alternative to a brute-force search, we compute an initial approximation to a solution of ML by solving the following continuous (relaxed) box-constrained optimization problem:

$$\text{RML: } \begin{cases} \min f(s) := \|y - Hs\|^2 \\ \text{subject to } -3 \leq s_i \leq 3, i = 1, \ldots, 2K. \end{cases}$$  \hspace{1cm} (4)

In RML, we ignore the integer constraints in ML and only require that $x_i$ lies between $-3$ and $+3$. Let $A = H^T H$, $b = H^T y$, and $x = s$. Then, the relaxation of the ML problem RML is equivalent to the following quadratic programming problem with box constraints,

$$\min \frac{1}{2}x^TAx - b^Tx$$  \hspace{1cm} subject to \hspace{0.5cm} \begin{align*} -3 & \leq x_i \leq 3, \quad i = 1, \ldots, 2K. \end{align*}$$  \hspace{1cm} (5)

We will solve the relaxation (5) by the AS_CBB method, which is described in the next section. Since a solution of (5) typically has noninteger components, we need to transform our noninteger solutions to a point feasible for ML. We move closer to a feasible point for ML by solving another quadratic programming problem with a penalty term. Let $y$ denote an optimal solution to (5); let $y_i^+$ denote the smallest integer $-1$ or $+1$ or $+3$ which is greater than or equal to $y_i$, let $y_i^-$ denote the largest integer $-3$ or $-1$ or $+1$ which is less than or equal to $y_i$. We consider the penalized problem:

$$\min \frac{1}{2}x^TAx - b^Tx + p \sum_{i=1}^{2K} (y_i^+ - x_i)(x_i - y_i^-)$$  \hspace{1cm} subject to \hspace{0.5cm} \begin{align*} -3 & \leq x_i \leq 3, \quad i = 1, \ldots, 2K. \end{align*}$$  \hspace{1cm} (6)

In our numerical experiments, the penalty parameter $p$ was $.5$ times the largest diagonal element of $A$. When we solve the penalized problem (6), the optimal $x$ must make both the original ML quadratic small while keeping the components of $x$ near feasible points for ML.

The solution to (6) is typically closer to a feasible point for ML than the solution to (5). Nonetheless, a solution to $z$ (6) is typically fractional. We move each component of $z$ to a feasible point for ML by a rounding process. In other words, each component of $z_i$ is replaced by quantize($z_i$) where quantize($\alpha$) rounds $\alpha$ to the nearest element in the set $\{\pm 1, \pm 3\}$.

The approximation quantize($z$) to a solution of ML may not be a local minimum in the discrete ML problem. Hence, to obtain a discrete local minimizing approximation to a solution of ML, we perform cycles of discrete coordinate descent where the starting point $x$ is a quantized solution to (6). The discrete coordinate descent algorithm is the following:

for $i = 1 : 2K$

$$\hat{t} = \arg \min \{f(x_i(t)) : t \in \{-3, -1, 1, 3\}\}$$
$$x_i(t) = \hat{t} \text{ (replace } i\text{-th component of } x \text{ by } \hat{t})$$
end

where

$$x_i(t) = (x_1, \ldots, x_{i-1}, \hat{t}, x_{i+1}, \ldots, x_{2K}).$$

Since $f$ is a quadratic function, the computational complexity of the this loop is essentially on the order of the computational cost of one function evaluation of (5). Hence, its computational expense is negligible compared with solving (5).

In summary, our quadratic programming algorithm for obtaining an approximation to a ML solution is the following:

1. Compute a solution $\hat{x}$ to box constrained quadratic programming problem (5).
2. Using $\hat{x}$ as a starting guess, compute a solution $z$ to the penalized, quadratic box constrained optimization problem (6).
3. Apply the quantization operator to each component of $z$.
4. Perform discrete coordinate descent starting from $x = z$.

### III. The Affine-Scaling Cyclic BB Method

In this section, we describe the optimization algorithm AS_CBB [4]–[7] that we use to solve either (5) or (6). The CBB method can be applied to an unconstrained problem

$$\min f(x), \quad x \in \mathbb{R}^n,$$

where $f$ is continuously differentiable, and $\mathbb{R}^n$ denotes Euclidean $n$ space. Suppose that $x_0$ is an initial point, $x_k$ is the current point, and $g_k$ is the gradient of $f$ at $x_k$, then gradient methods calculates the next point from

$$x_{k+1} = x_k - \alpha_k g_k,$$

where $\alpha_k$ is a stepsize computed by some line search algorithm. In the steepest descent (SD) method, the stepsize is chosen such that $f(x_k)$ is minimized along the search direction $g_k$:

$$\alpha_k^{SD} \in \arg \min_{\alpha \in \mathbb{R}} f(x_k - \alpha g_k).$$

It is well-known that steepest descent method can be very slow when the Hessian of $f$ is ill-conditioned at a local minimum.
On the other hand, it has been shown that if the exact steepest descent step is reused in a cyclic fashion, then the convergence is accelerated. Given an integer \( m \geq 2 \), which we call the cyclic length, cyclic steepest descent (CSD) method can be expressed as:

\[
\alpha_{ml+i} = c_{ml+1}^{SD} \quad \text{for } i = 1, \ldots, m,
\]

\( l = 0, 1, \ldots \), where \( c_{ml+1}^{SD} \) is the stepsize of steepest descent.

The basic idea of Barzilai-Borwein (BB) method is to regard the matrix \( \mathbf{D}(\alpha_k) = \frac{1}{\alpha_k} \mathbf{I} \), where \( \mathbf{I} \) denotes an identity matrix, as an approximation of the Hessian \( \nabla^2 f(x_k) \) and impose a quasi-Newton property on \( \mathbf{D}(\alpha_k) \):

\[
\alpha_k^{BB} = \arg\min_{\alpha \in \mathbb{R}} \| \mathbf{D}(\alpha) \mathbf{t}_{k-1} - \mathbf{v}_{k-1} \|_2, \quad (7)
\]

where \( \mathbf{t}_{k-1} = x_k - x_{k-1}, \mathbf{v}_{k-1} = g_k - g_{k-1} \), and \( k \geq 2 \). The proposed stepsize, obtained from (7), is

\[
\alpha_k^{BB} = \frac{\mathbf{t}_{k-1}^T \mathbf{t}_{k-1}}{\mathbf{t}_{k-1}^T \mathbf{v}_{k-1}}. \quad (8)
\]

Similar to the cyclic steepest descent, the cyclic BB method (CBB) developed in [5] reuses the previous stepsizes:

\[
\alpha_{ml+i} = c_{ml+1}^{BB} \quad \text{for } i = 1, \ldots, m,
\]

where \( m \geq 2 \) is again the cycle length. In [3] Dai established the R-linear convergence of CBB for a strongly convex quadratic. Furthermore, in [5] the local R-linear convergence for the CBB method at a local minimizer of a general nonlinear function is proved.

We now introduce the AS_CBB method [7], an iterative method for the constrained optimization problem

\[
\min \{ f(x) : x \geq 0 \}, \quad (9)
\]

where \( f : \mathbb{R} \rightarrow \mathbb{R} \) is a real-valued, continuously differentiable function defined over domain \( x \geq 0 \). AS_CBB is valid for problems with both upper and lower bound constraints, however, to simplify the discussion, we focus on the nonnegativity constraint \( x \geq 0 \).

The algorithm starts at a point \( x_1 \) in the interior of the feasible set, and generates a sequence \( x_k, k \geq 2 \), by the following rule:

\[
x_{k+1} = x_k + d_k \quad (10)
\]

where the \( i \)-th component of \( d_k \) is given by

\[
d_{ki} = -\left( \frac{1}{\lambda_k + g_{i}^T(x_k)/x_k} \right) g_{i}(x_k). \quad (11)
\]

Here \( \lambda_k \) is a positive scalar, \( g_i(x) \) is the \( i \)-th component of the gradient \( \nabla f(x) \), and \( t^+ = \max(0, t) \) for any scalar \( t \). We compute \( \lambda_k \) using a cyclic version [5] of the Barzilai-Borwein (BB) stepsize rule [2]. That is, we first define

\[
\lambda_k^{BB} := \arg\min_{\lambda \geq 0} \| \lambda \mathbf{t}_{k-1} - \mathbf{v}_{k-1} \|_2 = \max \left\{ \lambda_0, \frac{\mathbf{t}_{k-1}^T \mathbf{v}_{k-1}}{\mathbf{t}_{k-1}^T \mathbf{t}_{k-1}} \right\}, \quad (12)
\]

where \( k \geq 2 \) and \( \lambda_0 > 0 \) is a fixed parameter. The starting parameter value \( \lambda_k^{BB} \) can be chosen freely, subject to the constraint \( \lambda_1 \geq \lambda_0 \); for example,

\[
\lambda_k^{BB} = \max \{ \lambda_0, \| g_1 \|_\infty \}.
\]

The cyclic choice for \( \lambda_k \) is

\[
\lambda_{ml+i} = \lambda_{ml+1}^{BB} \quad \text{for } i = 1, \ldots, m,
\]

As shown in [7], this algorithm is obtained through an approximation to Newton’s method applied to the first-order optimality conditions where \( \nabla^2 f(x_k) \) is replaced by \( \lambda_k \mathbf{I} \).

For our box constrained problems (5) or (6), the generalization of the AS_CBB search direction (11) is

\[
d_{ki} = -\left( \frac{1}{\lambda_k + g_i(x_k)} \right) g_i(x_k),
\]

where

\[
X_i(x) = \begin{cases} 
3 - x_i & \text{if } g_i(x) \leq 0, \\
x_i + 3 & \text{otherwise}.
\end{cases}
\]

In general, it is shown that the AS_CBB method is a gradient based affine-scaling interior-point method [7]. And because of its relative efficiency, very low memory requirement, insensitivity to the noise, and the simplicity of the iteration (only gradients are required), it is very suitable for solving large-scale optimization problems. Although the CBB method has global convergence for strongly convex quadratic optimization, it is still not clear whether AS_CBB is also globally convergent for strongly convex quadratics. However, global convergence was established [7] for an implementation of AS_CBB based on a nonmonotone line search first introduced in [6].

First, let’s define

\[
f_k^{\max} = \max\{ f(x_{k-i}) : 0 \leq i \leq \min(k-1,M-1) \}. \quad (13)
\]

Here \( M > 0 \) is a fixed integer. The AS_CBB algorithm with a nonmonotone line search can be described as follows:

**Affine-scaling CBB algorithm with line search**

Initialize \( k = 1, x_1 = \text{starting guess}, \) and \( f_0^k = f(x_1) \).

While \( x_k \) is not a stationary point

1. Let \( d_k \) be given by (11).
2. Choose \( f_k^j \) so that \( f(x_k) = f_k^j \leq f_k^\max \leq \max\{ f_k^{j-1}, f_k^\max \} \) and \( f_k^j \leq f_k^\max \) infinitely often.
3. Let \( f_k^\max \) be either \( f_k^j \) or \( \min\{ f_k^\max, f_k^j \} \). If \( f(x_k + d_k) \leq f_k^\max + \delta g_1^T d_k \), then \( \epsilon_k = 1.5 \).
4. If \( f(x_k + d_k) > f_k^\max + \delta g_1^T d_k \), then \( \epsilon_k = \eta_j \) where \( j > 0 \) is the smallest integer such that

\[
f(x_k + \eta_j d_k) \leq f_k^\max + \eta_j \delta g_1^T d_k. \quad (14)
\]

5. Set \( x_{k+1} = x_k + \epsilon_k d_k \) and \( k = k + 1 \).

End

Here the parameters \( \delta \) and \( \eta \) used in the Armijo line search of Step 4 must satisfy \( \delta \in (0,1) \) and \( \eta \in (0,1) \).

If the iterates generated by AS_CBB method converge to a nondegenerate local minimizer and the second order sufficient optimality condition holds, the local convergence rate is
showed to be at least R-linear [7]. When the columns of $H$ are linearly independent, which is often true for our case (5) where $N$ is often much larger than $K$, $A$ is positive definite. Hence, the quadratic programming (5) has a unique global minimum at which the second order sufficient optimality condition holds. Thus, if this global minimum is nondegenerate, AS_CBB method converges to the global minimum at least R-linearly.

In summary, we have the following local complexity result for the AS_CBB method applied to the quadratic relaxation of (5).

**Proposition 1:** Suppose the columns of $H$ are linearly independent and the problem (5) is nondegenerate at its global minimum $x^*$. Then, there exists a $\delta > 0$ such that it takes at most

$$ M := O\left(\max\{-\log(\epsilon/\|x_0 - x^*\|), 0\}\right) $$

iterations to satisfy the condition

$$ \|x_k - x^*\| \leq \epsilon, \text{ for all } k \geq M, $$

for all $x_0 \in B_\delta(x^*)$, where $x_k, \ k \geq 0$, is the sequence generated by the AS_CBB method for the problem (5).

Since the AS_CBB method only needs $O(N \cdot K)$ (see [7]) computing flops for each iteration, under the condition of Proposition 1, the complexity of using AS_CBB to find an $\epsilon$ accurate solution of (5) or (6), for $p$ sufficiently small, is $O(M \cdot N \cdot K)$.

**IV. SIMULATION RESULTS**

In this section, we would like to compare the solutions obtained by our BCR approach with those obtained by other strategies. In our first simulation, we considered a MIMO system with $K = 6$ inputs and $N = 12$ outputs using 16-QAM signaling. The entries of the MIMO channel were chosen as independent and identically distributed, zero-mean, complex normal random variables. For each signal to noise ratio (SNR), we use 120 channels and 10000 symbol sets for each channel to estimate the average probability of the error in detecting the message vectors. The box relaxation problem was solved by the AS_CBB method [7]. For comparison, we also simulated the conventional linear ZF detector and the SDR detector in [11]. The results are provided in Fig. 1.

In the first simulation, we did not compare the performance of the BCR with that of the ML detector since the complexity of the ML detector is too high (16$^6$ possible vectors). In our second simulation, we consider a smaller system where a full ML search is impractical but possible, $K = 4$ and $N = 8$, i.e., 16$^4$ vectors. The rest of the parameters are as in the first simulation. The results are shown in Fig. 2. The gain of the proposed algorithm is about 0.9 dB and 2.3 dB over SDR and ZF, respectively, at an set error rate (SER) of $10^{-2}$ in Figs. 1 and 2. The loss of proposed algorithm is only about 0.1 dB over the ML. So the proposed algorithm is very effective with greatly reduced complexity.

Here, let us consider the complexities of the SDR, ZF and BCR detectors. The complexity of the SDR method in [11] is $O(R^{6.5}L^{6.5})$ where $L = O(K)$, $K$ is the maximum of the number of symbols and the number of antennas, and $R$ is the square root of the constellation [9]. It is easy to notice that the complexity of SDR depends not only the number of users and antennas, but also the order of the QAM. Since the ZF algorithm multiplies the received signal $y$ by the pseudo inverse of the channel matrix $H^*$, the complexity of the ZF detector is $O(L^3)$. The complexity of the box relaxation method is only $O(MNK)$, where $M$ is in (15). In our computer simulation, the computation of the solution of the box relaxed approximation is about 100 times faster than solution time for the SDP method in [11], and even faster than the ZF detector, as expected by the complexity analysis.

**V. CONCLUSION**

In this paper, we relaxed the constraints of the ML detector for 16-QAM signaling over MIMO channels. Even though the SDR detector showed good results, its complexity is high. Our box relaxed detector solved by the AS_CBB method outperformed the SDR detector and the ZF detector, moreover, it was more computationally efficient than other detectors.
REFERENCES


Soonchul Park received his B.S. degree in Mathematics from Yonsei University, Wonju, Korea, in 1992, and his Ph.D degrees in Mathematics from the University of Florida, Gainesville, in 1999. He worked at the University of Florida as a post-doctoral from 1999 to 2001. He was an Assistant Professor at the West Texas A&M University, Canyon, TX, from 2001 to 2003, an Visiting Professor at the University of Florida from 2003 to 2004, a Lecturer in the Department of Mathematics from 2005 to 2006, and a BK Associate Professor in the school of Electronic and Electrical Engineering at Kyungpook National University, Daegu, Korea, from 2006 to the present. His main research interests are optimization, signal processing, and resource allocation.

Hongchao Zhang received his B.S. degree in Mathematics from Shandong University, Jinan, P.R.China, in 1998 and his M.S. degree in Computational Mathematics from Institute of Computational Mathematics and Scientific/Engineering Computing of the Chinese Academy of Sciences, in 2001, respectively. He received his Ph.D. degree in Applied Mathematics from University of Florida, Gainesville, in 2006. Since 2006, he is a Post-Doctoral Fellow at Institute for Mathematics and its Applications (IMA) at University of Minnesota and IBM T.J. Watson Research Center, U.S.A. Hongchao Zhang has received the SIAM best student paper prize in 2006 and has been an Associate Editor of Computational Mathematics and Applications since 2006. His main research interests are nonlinear optimization theory and its applications.

William W. Hager received his B.S. degree in Mathematics from Harvey Mudd College in 1970. He was awarded a National Science Foundation Fellowship for graduate studies in mathematics at M.I.T. where he received his Masters and Ph.D degrees in 1971 and 1974, respectively. He was an Assistant Professor at the University of South Florida, Tampa, from 1974 to 1976, an Assistant Professor at Carnegie Mellon University, Pittsburgh, PA, from 1976 to 1980, an Associate Professor from 1980 to 1986 and a Professor from 1986 to 1988 at Pennsylvania State University, University Park, and a Professor at the University of Florida, Gainesville, from 1988 to the present. He has been editor-in-chief of Computational Optimization and Applications since 1992 and co-Director of the Center for Applied Optimization at the University of Florida since 1992. He was Program Director of the SIAM Activity Group on Control and System Theory from 1998 to 2001. Dr. Hager was co-chair of the 2001 SIAM Conference on Control and Its Applications.

Dong Seog Han received his B.S. degree in electronic engineering from Kyungpook National University (KNU), Daegu, Korea, in 1987, his M.S and Ph.D degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejon, Korea, In 1989 and 1993, respectively. From October 1987 to August 1996, he was with Samsung Electronics, Co. Ltd., where he developed the transmission systems for QAM HDTV and Grand Alliance HDTV receivers. Since September 1996, he has been with the School of Electronic and Electrical Engineering, Kyungpook National University. Currently he is an Associate Professor in the school of Electronic and Electrical Engineering, KNU. He worked as a courtesy Associate Professor in the electrical and computer engineering, University of Florida in 2004. His main research interests are digital communications and array signal processing.