REDUCING COMPLEXITY OF GENERALIZED MINIMUM MEAN SQUARE ERROR DETECTION

Tharwat Morsy, Jürgen Götze

Information Processing Lab., TU Dortmund University
Otto-Hahn Str. 4, 44221, Dortmund, Germany
phone: + (49) 2317552091, fax: + (49) 2317553251, email: (tharwat.morsy,juergen.goetze)@tu-dortmund.de
web: www.dt.e-technik.uni-dortmund.de

ABSTRACT

The generalized Minimum Mean Squared Error (GMMSE) detector has a bit error rate performance, which is similar to the MMSE detector. The advantage of the GMMSE detector is, that it does not require the knowledge of the noise power. However, the computational complexity of the GMMSE detector is significantly higher than the computational complexity of the MMSE detector. In this paper the idea of using the structure of the system matrix (Toeplitz) is combined with a convex relaxation of the detection problem to reduce the computational complexity of GMMSE detector. Furthermore, by using circular approximation of this structure an approximate GMMSE detector is presented, whose computational complexity is only slightly higher than MMSE, i.e. only an iterative gradient descent algorithm based on the inversion of diagonal matrices is required additionally.

1. INTRODUCTION

The maximum likelihood (ML) detection problem can be written as a quadratic optimization problem with integer constraints [1]. Unfortunately this problem is in general nondeterministic polynomial hard (NP-hard) [2]. This observation resulted in the development of many receivers that have reasonable complexity [3,4], e.g. the well-known least squares (LS) and minimum mean squared error (MMSE) detectors [5,6] as the most simple cases.

Recently convex programming has been successfully employed to suboptimally solve such detection problems. Using this kind of relaxation converts the discrete optimization problem into a continuous one which can be solved iteratively [7]. Generalized minimum mean squared error detector is one important detector that uses convex programming to solve the detection problem using unconstrained gradient descent algorithm [8]. The advantages of this detector are that it has a BER performance which is similar to the MMSE detector, and it does not require the knowledge of the noise power (MMSE detector needs this knowledge). Therefore, it can be used in scenarios where the noise power is changing rapidly or it is unknown. Associated with these advantages of GMMSE detector there is the disadvantage, that it has a significantly higher computational complexity compared to MMSE detector.

In order to decrease the GMMSE computational complexity the structure of the system matrix is used in this paper. First the Toeplitz structure of the channel convolution matrix is taken into consideration. In this case computing the solution of the GMMSE detector requires the EVD of Toeplitz matrix but it significantly reduces the effort for the iterative gradient descent algorithm. Nevertheless, computing the EVD of the Toeplitz matrix (using e.g Lanczos algorithm) is still computationally demanding. Therefore we approximate the banded Toeplitz matrix by a circular matrix as shown in [9]. In this case the MMSE/GMMSE solution is obtained by computing the EVD of the circular matrix using FFT/IFFT, such that the required EVD implies no additional effort. Furthermore, this procedure also significantly reduces the effort for the gradient descent algorithm since now the iteration steps of this algorithm are based on the diagonal matrix containing the eigenvalues. Therefore, the circular approximation is advantageous for both parts of the GMMSE solution.

This paper is organized as follows: In section 2 the system model for the detection problem and its convex relaxations are introduced. LS and MMSE detectors are described from the convex programming point of view in section 3. GMMSE detector is described in section 4. In section 5 we introduce our new detector which is derived from the GMMSE detector taking into account the structure of the channel matrix which leads to a reduced computational complexity. Simulation results are used to compare bit error rate (BER) of the different detectors in section 6 and the computational complexity is discussed in section 7. Conclusions are drawn in section 8.

2. DETECTION PROBLEM AND ITS RELAXATIONS

Consider the system model in matrix form as

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

The vector $\mathbf{r} \in \mathbb{R}^m$ is the received signal vector, the matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the convolution matrix, and the vector $\mathbf{n} \in \mathbb{R}^m$ is additive white Gaussian noise with noise power σ^2. The transmitted symbols $\mathbf{x} \in \mathbb{R}^n$ are drawn from Binary Phase Shift Keying (BPSK) constellation, i.e. $\mathbf{x} \in \{-1,+1\}^n$.

Under the white Gaussian noise assumption the ML detector of $\mathbf{x}$ is given by

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}\in\{-1,+1\}^n} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|_2^2. \tag{2}$$

The ML problem in (2) can be equivalently written as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}\in\{-1,+1\}^n} \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{r}^T\mathbf{H}\mathbf{x}. \tag{3}$$

Substituting the value of the matched filter output

$$\mathbf{y} = \mathbf{H}^T\mathbf{r}$$

into (3), we get

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}\in\{-1,+1\}^n} \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{y}^T\mathbf{x}. \quad (4)$$

This problem is NP hard and solving (4) by exhaustive search has a complexity which grows with 2^n [2]. This makes computationally less complex solutions of (4) interesting.

We use the benefits of convex programming as an important mathematical tool to solve problem (4) by relaxing its constraint set. The constraint set $\mathbf{x} \in \{-1,+1\}^n$which contains only the corners of the unit hypercube is not a convex set. Therefore we relax this constraint set using two relaxations which yield a convex set.

The first relaxation of the constraint in (4) is the whole space or by other words, there is no constraints and the second relaxation is the sphere which covers this unit hypercube. The solution in each case can be mapped to the feasible set of the original problem by taking the sign of each component of the relaxed solution vector.

It is worthwhile to note that the relaxation also works for higher constellations like QPSK, 8PSK, and 16PSK. We refer to these higher constellations as M-PSK. In our case we have $M = 2$ as BPSK. In M-PSK constellations, problem (4) can be written as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in S^n} \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{y}^T\mathbf{x}, \quad (5)$$

where S contains the PSK constellation points.
For M-PSK modulation, the constellation points take the form

$$e^{j\alpha_i}, \alpha_i = 2\pi i/M, \forall i = 1,...,M.$$

The discrete nature of the set α_i makes problem (5) intractable. As a result, we propose a continous relaxation of the set α_i to contain all possible angles in $[0,2\pi]$. In other words, we can relax the constraint $\mathbf{x} \in S^n$ to $\mathbf{x} \in U^n = \{\mathbf{x} : |x_i| = 1, \forall i = 1,...,n\}$, so problem (5) becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in U^n} \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{y}^T\mathbf{x}. \quad (6)$$

This relaxation is a quadratic optimization problem that can easily be relaxed to a convex optimization problem [7].

3. LEAST SQUARES AND MMSE DETECTORS

We first discuss the LS and MMSE solution from the convex programming point of view. Relaxing the constraint set to be the whole space, problem (4) takes the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{y}^T\mathbf{x}. \quad (7)$$

The following theorem stated in [7] describes LS and MMSE solution from the convex programming point of view.

Theorem 1 *Suppose that the objective function f in an unconstrained convex optimization problem is differentiable, so the well known necessary and sufficient optimality condition is*

$$\nabla f = 0. \quad (8)$$

Applying condition (8) to problem (7), which has an objective function

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{y}^T\mathbf{x},$$

the necessary and sufficient optimality condition gives the solution

$$\mathbf{x}^* = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{y}, \quad (9)$$

which is the well known least squares solution.

When the noise power σ^2 is known $\mathbf{H}^T\mathbf{H}$ is replaced by $\mathbf{H}^T\mathbf{H} + \sigma^2\mathbf{I}$, then using the same relaxation (the whole space) we get the minimum mean square error solution

$$\mathbf{x}^* = \left(\mathbf{H}^T\mathbf{H} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{y}. \quad (10)$$

4. GENERALIZED MMSE DETECTOR

If we relax the constraint set to be the sphere which contains the unit hypercube, then our detection problem takes the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}^T\mathbf{x}\leq n} \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{y}^T\mathbf{x}. \quad (11)$$

Since problem (11) has a convex objective function over a convex constraint set, i.e. it is a convex optimization problem and it has a unique minimum [7]. The convex duality theorem guarantees that no duality gap exists and one can solve for the dual problem instead [10]. Problem (11) has a single constraint such that there is only one dual variable and a simple iterative algorithm can be employed to solve this dual problem.

We can express the Lagrange dual function as

$$L(x,\lambda) = \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} - 2\mathbf{y}^T\mathbf{x} + \lambda\left(\mathbf{x}^T\mathbf{x} - n\right), \quad (12)$$

which is minimized over $\mathbf{x}$ and maximized over $\lambda \geq 0$. Solving for $\mathbf{x}$ in terms of λ and substituting back, we obtain

$$\max_{\lambda\geq 0} -\mathbf{y}^T\left(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}\right)^{-1}\mathbf{y} - \lambda n. \quad (13)$$

This problem has the advantage, that it is a one dimensional optimization problem so it is easier to solve this problem instead of problem (11). Problem (13) can be solved by different iterative algorithms [11]. A simple unconstrained gradient descent algorithm given by

$$\bar{\lambda}(t+1) = \bar{\lambda}(t) + \mu\left(\mathbf{y}^T\left(\mathbf{H}^T\mathbf{H} + \bar{\lambda}(t)\mathbf{I}\right)^{-2}\mathbf{y} - n\right), \quad (14)$$

converges to $\bar{\lambda}$ for a reasonable choice of the step size μ. The solution of (13) is given by

$$\lambda^* = max(0,\bar{\lambda}). \quad (15)$$

Then, the unique minimizer of (11) is

$$\mathbf{x}^* = \left(\mathbf{H}^T\mathbf{H} + \lambda^*\mathbf{I}\right)^{-1}\mathbf{y}. \quad (16)$$

This solution, looks familiar because of its similarity to the MMSE detector. When

$$\lambda^* = \sigma^2,$$

the GMMSE detector reduces to the MMSE detector. Therefore this detector which depends on the value of the optimum dual solution λ^* is named *Generalized MMSE detector*. The advantages of the GMMSE detector are, that it improves the BER performance (as compared to LS) and it does not require the knowledge of the noise power σ^2. Because of these advantages it can be used in scenarios where the noise power is changing rapidly or it is unknown. According to the nature of λ^* which is a function of $\mathbf{y}$ the GMMSE solution results in a nonlinear detector in contrast to the MMSE detector. However GMMSE detector has the disadvantage that it requires a higher computational complexity than MMSE detector.

5. STRUCTURED PROBLEM

In this section we use the GMMSE detector combined with the circular approximation of the banded Toeplitz structure which was presented in [9] in order to decrease the computational complexity. We achieve this aim in two steps . In the first step the Toeplitz structure of the channel convolution matrix $\mathbf{H}$ is used. In this step we express the matrix $\mathbf{H}^T\mathbf{H}$ of problem (11) by its eigenvalue decomposition

$$\mathbf{H}^T\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^T,$$

where $\mathbf{V}$ is the matrix whose columns are the eigenvectors of $\mathbf{H}^T\mathbf{H}$ and Λ is a diagonal matrix that contains the corresponding eigenvalues as its diagonal elements. Problem (11) can be rewritten as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}^T\mathbf{x}\leq n} \mathbf{x}^T\left(\mathbf{V}\Lambda\mathbf{V}^T\right)\mathbf{x} - 2\mathbf{y}^T\mathbf{x}. \quad (17)$$

The dual problem for the problem (17) takes the form

$$\max_{\lambda\geq 0} -\mathbf{y}^T\left(\left(\mathbf{V}\Lambda\mathbf{V}^T\right)+\lambda\mathbf{I}\right)^{-1}\mathbf{y} - \lambda n. \quad (18)$$

The unconstrained gradient descent algorithm takes the form

$$\bar{\lambda}(t+1) = \bar{\lambda}(t) + \mu\left(\mathbf{y}^T\mathbf{V}\left(\Lambda+\bar{\lambda}(t)\mathbf{I}\right)^{-2}\mathbf{V}^T\mathbf{y} - n\right) \quad (19)$$

and the GMMSE solution will be

$$\mathbf{x}^* = \mathbf{V}\left(\Lambda+\lambda^*\mathbf{I}\right)^{-1}\mathbf{V}^T\mathbf{y}. \quad (20)$$

Besides computing $\mathbf{V}^T\mathbf{y}$ only diagonal matrices must be converted in (19) and (20), which significantly simplifies the computations . We can also make use of the Toeplitz structure of $\mathbf{H}^T\mathbf{H}$ when computing the EVD by using the Lanczos algorithm [12]. Although this approach significantly reduces the computational complexity of the GMMSE detector (the iterations of (19) on the diagonal matrices are only of $O(n)$), it is still much more complex than MMSE because of the required EVD.

In the following we use an approximation of the Toeplitz case to further reduce the computational complexity. A banded Toeplitz structured convolution matrix $\mathbf{H}$ is approximated to a circular matrix $\mathbf{H}^T$ by adding $L-1$ columns to the Toeplitz matrix, where L is the length of the channel impulse response. This is shown in the following example for $L = 2$:

$$\mathbf{H} = \begin{bmatrix} h_1 & 0 \\ h_2 & h_1 \\ 0 & h_2 \end{bmatrix} \rightarrow \begin{bmatrix} h_1 & 0 & h_2 \\ h_2 & h_1 & 0 \\ 0 & h_2 & h_1 \end{bmatrix} = \tilde{\mathbf{H}}.$$

If the channel matrix $\mathbf{H}$ in problem (11) is approximated by the circular matrix $\tilde{\mathbf{H}}$ we obtain

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}^T\mathbf{x}\leq n} \mathbf{x}^T\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\mathbf{x} - 2\mathbf{y}^T\mathbf{x}. \quad (21)$$

We can express the matrix $\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}$ by its eigenvalue decomposition

$$\tilde{\mathbf{H}}^T\tilde{\mathbf{H}} = \mathbf{F}^T\Lambda\mathbf{F},$$

where $\mathbf{F}$ is the discrete Fourier transform matrix (computed by FFT) and

$$\Lambda = diag\left(\mathbf{F}\cdot\tilde{\mathbf{H}}(:,1)\right).$$

In that case problem (11) can be written as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}^T\mathbf{x}\leq n} \mathbf{x}^T\left(\mathbf{F}^T\Lambda\mathbf{F}\right)\mathbf{x} - 2\mathbf{y}^T\mathbf{x}. \quad (22)$$

The dual problem for problem (22) is

$$\max_{\lambda\geq 0} -\mathbf{y}^T\left(\left(\mathbf{F}^T\Lambda\mathbf{F}\right)+\lambda\mathbf{I}\right)^{-1}\mathbf{y} - \lambda n \quad (23)$$

and the gradient descent algorithm in the circular case takes the form

$$\bar{\lambda}(t+1) = \bar{\lambda}(t) + \mu\left(\mathbf{y}^T\mathbf{F}^T\left(\Lambda+\bar{\lambda}(t)\mathbf{I}\right)^{-2}\mathbf{F}\mathbf{y} - n\right). \quad (24)$$

After getting the optimal value λ^*, the GMMSE solution in the circular case is

$$\mathbf{x}^* = \mathbf{F}^T\left(\Lambda+\lambda^*\mathbf{I}\right)^{-1}\mathbf{F}\mathbf{y}. \quad (25)$$

Again, besides computing $\mathbf{F}\mathbf{y}$ (IFFT) only diagonal matrices must be inverted in (24) and (25). Most important, no EVD computation is required in the circular case, since the EVD of a circular matrix is easily obtained using FFT/IFFT. Therefore, in this case the additional effort (compared to MMSE) given by the iteration of (24) is moderate, i.e. inversions of diagonal matrices and scalar products.

6. SIMULATION RESULTS

The BER performance of the different detectors is discussed for BPSK modulation. In the simulation we compare the BER performance for LS, MMSE, and GMMSE detectors, taking into account that we have two different structures, Toeplitz and circular approximation. We applied this simulation for four different simulation scenarios:

- $L = 5$, $n = 100$
- $L = 5$, $n = 1000$
- $L = 15$, $n = 100$
- $L = 15$, $n = 1000$

The first two scenarios are shown in figure (1) and the last two scenarios are shown in figure (2).

Figures (1) and (2) show that GMMSE detector has almost the same performance as MMSE detector but it has the advantage that it does not require the knowledge of σ^2. Furthermore, we see that the circular approximation only slightly degrades the performance of the detectors.

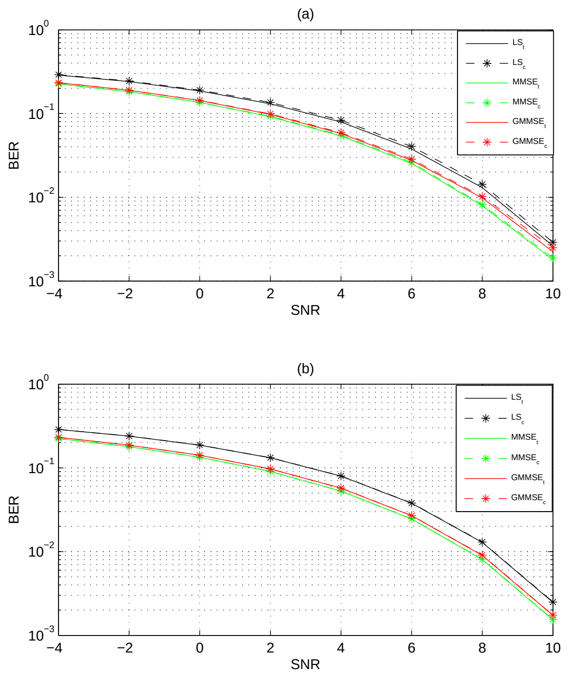

Figure 1: Performance analysis for Structured LS, MMSE, and GMMSE detectors with channel length $L = 5$, (a) $n = 100$, (b) $n = 1000$ (t: Toeplitz case; c: circular case)

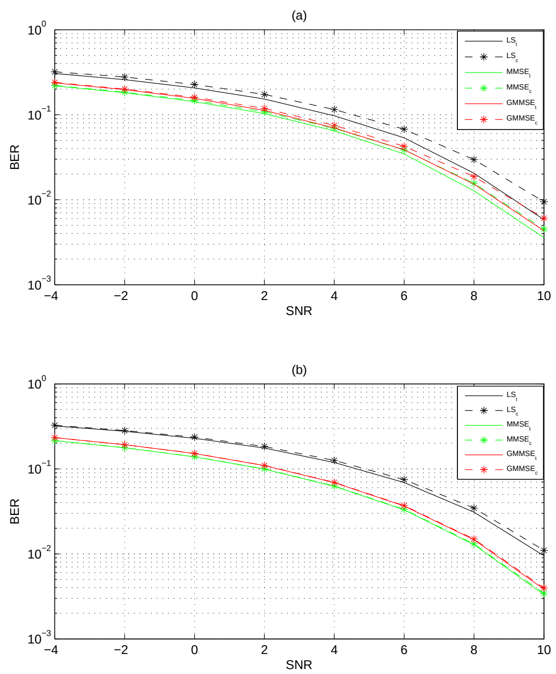

Figure 2: Performance analysis for Structured LS, MMSE, and GMMSE detectors with channel length $L = 15$, (a) $n = 100$, (b) $n = 1000$ (t: Toeplitz case; c: circular case)

7. COMPUTATIONAL COMPLEXITY

The computational complexity of the GMMSE detector is composed of two parts:

Part 1 The complexity of the solution of the system of equations ((16), (20) or (25)) which is the same as for LS and MMSE ((9) or (10)).

Part 2 The complexity of the iterations required for the gradient descent algorithm ((14), (19) or (24)).

In part 1, if there is no structure the solution is obtained by the Cholesky algorithm with complexity $n^3/3$. When there is a Toeplitz structure, the solution is given by the Levinson algorithm with complexity $4n^2$ and if we approximate this Toeplitz matrix to a circular structure, the solution is obtained using the FFT decomposition with complexity $3/2(n+L-1)\log_2(n+L-1)+(n+L-1)$. Therefore, the circular approximation results in a significantly reduced computational complexity.

In part 2 gradient descent algorithm adds some complexity. However, for the structured cases ((19) or (24)) the iterations of the gradient descent algorithm are only applied to diagonal matrices (Λ) such that the complexity is only of $O(n)$ per iteration. Figure (3) shows the mean number of required iterations for the Toeplitz case and circular case in our first two scenarios when the channel length is $L = 5$ and figure (4) shows it in our last two scenarios when the channel length is $L = 15$. Since the required number of iterations is quite small and the computational complexity is only of $O(n)$ per iteration, the complexity of the gradient descent algorithm is almost negligible compared to part(1).

System variables	GMMSE complexity		
	No structure	Toeplitz	Circular
$n = 100; L = 5$	$3.3 * 10^5$	$0.4 * 10^5$	1149
$n = 1000; L = 5$	$3.3 * 10^8$	$0.04 * 10^8$	11413
$n = 100; L = 15$	$3.3 * 10^5$	$0.4 * 10^5$	1282
$n = 1000; L = 15$	$3.3 * 10^8$	$0.04 * 10^8$	16194

Table 1: Computational complexity for GMMSE detector

The overall complexity (part(1) and part(2)) for all scenarios are shown in table (1).

8. CONCLUSIONS

In this paper, it was shown that the circular approximation of the Toeplitz channel matrix is not only effective to significantly reduce the computational complexity of GMMSE detector using the gradient descent algorithm, but it also keeps the performance gain compared to LS detector (is almost the same as MMSE) without any requirement to know the noise power value (σ^2) .

In future work we will apply the presented technique to various practical problems and evaluate the performance depending on the channel length (L) and the dimension of the transmitted bit vector (n). We will also apply it to some common communication schemes like CDMA and OFDM.

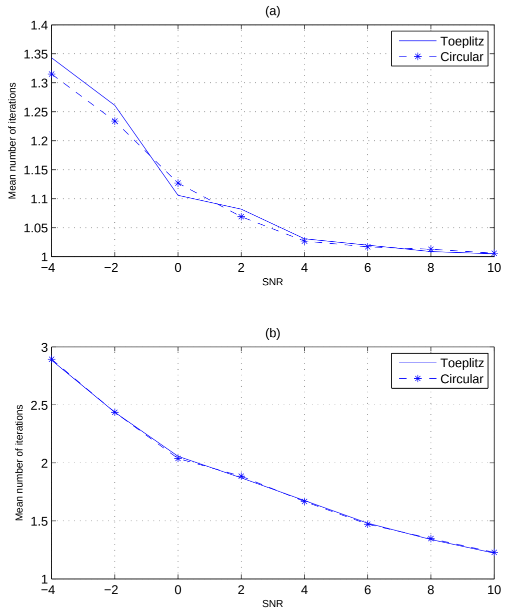

Figure 3: Iterations for gradient descent algorithm in Toeplitz and circular structure cases with channel length $L = 5$, (a) $n = 100$ and (b) $n = 1000$.

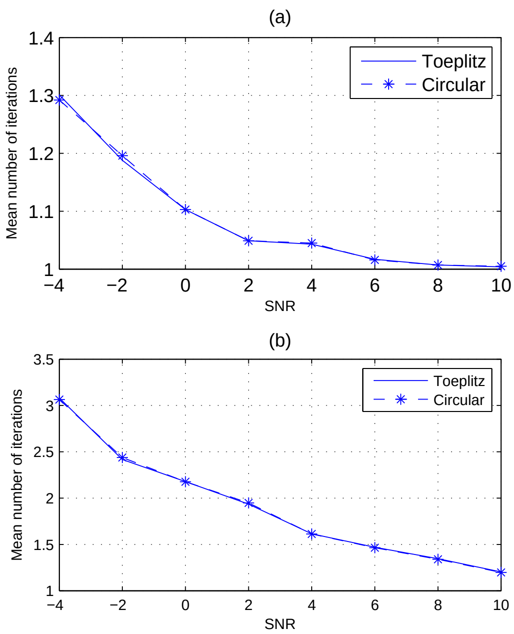

Figure 4: Iterations for gradient descent algorithm in Toeplitz and circular structure cases with channel length $L = 15$, (a) $n = 100$, and (b) $n = 1000$.

REFERENCES

[1] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[2] S. Verdu, "Computational complexity of multiuser detection," *Algorithmica*, vol. 4, no.4, pp. 303–312, 1989.

[3] P. H. Tan and L. K. Rasmussen, "The application of semidefinite programming for detection in CDMA," *IEEE journal on selected areas in communications*, vol. 19, no. 8, pp. 1442-1449, August. 2001.

[4] X.-W. Chang and Q. Han, "Solving box-constrained integer least squares problem," *IEEE Transactions on wireless communications*, vol. 7, no. 1, pp. 277-287, 2008.

[5] R. Lupas and S.Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.

[6] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

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

[9] M. Vollmer, M. Haardt, and J. Götze, "Comparative study of joint-detection techniques for TD-CDMA based mobile radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1461–1475, August. 2001.

[10] S. Nash and A. Sofer, *Linear and Nonlinear Programming*. New York: McGraw-Hill, 1996.

[11] P. Hansen, "Methods of nonlinear 0-1 programming, "*In Annals of Discrete Mathematics 5: Discrete Optimization II, P. L. Hammer, E. L. Johnson, and B. H. Korte, Eds.* Amsterdam, The Netherlands: North Holland, 1979.

[12] G. Golub and C. Loan, *Matrix Computations*. Johns Hopkins Univ. Press, 3rd edition 1996.