國 立 交 通 大 學

電信工程學系

碩 士 論 文

粒子群趨動交叉熵法於多輸入多輸出信號偵測之應用

A Particle-Swarm-Driven Cross-Entropy Method for Multiple-Input-Multiple-Output Signal Detection

研 究 生：王惇琳

指導教授：蘇育德 博士

中 華 民 國 九十八 年 八 月

粒子群趨動交叉熵法於多輸入多輸出信號偵測之應用

A Particle-Swarm-Driven Cross-Entropy Method for

Multiple-Input-Multiple-Output Signal Detection

研究生：王惇琳　　Student : Chun-Lin Wang

指導教授：蘇育德 博士　　Advisor : Dr. Yu T. Su

國 立 交 通 大 學

電 信 工 程 學 系 碩 士 班

碩 士 論 文

A Thesis

Submitted to Department of Communications Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Master of Science

in Communications Engineering

August 2009

Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 八 年 八 月

粒子群趨動交叉熵法於多輸入多輸出信號偵測之應用

研究生：王惇琳　　指導教授：蘇育德 博士

國立交通大學電信工程學系碩士班

中文摘要

對多輸入多輸出(MIMO)無線通訊系統而言，接收端如何由收到的信號分辨出不同傳送天線的資料，並正確地還原其內容是最基本的設計議題。較常見的有 逼零 (zero-forcing)、最小均方誤差 (minimum-mean-square-error)、球形解碼(sphere decoding)等檢測法及利用通道矩陣的特殊分解之垂直分層空時碼(V-BLAST)或晚近的格狀削減(lattice reduction)等。這些方法及其變形或改進版本陸續有推陳出新的成果出現，但總有性能不理想或複雜度過高的缺失。

本文試圖提出一類非傳統的有效多輸入多輸出信號偵測方法。我們藉由環繞著接收信號點的隨機取樣來估計傳送信號的可能分佈，並逐步找到最大似然 (maximum likelihood)的傳送信號 。我們運 用所謂的 交叉 熵(cross-entropy)法來估計這個機率分佈 。交叉熵法為一種近似於蒙地卡羅(Monte-Carlo)的迭代式最佳化問題解法。藉由尋找與最佳機率密度函數最接近（即 Kullback-Leibler 距離最小）的權重取樣分佈機率 (importance sampling density)，即可求出最佳信號點之所在。應用此法於多輸入多輸出信號之偵測，其位元錯誤率(BER)在低信雜比(SNR)時與最大似然法之表現幾乎完全一致，然而在高信號雜信比範圍卻存著錯誤階位(error floor)。為改善此現象，我們引入粒子群驅動演算法(particle swarm algorithm)的主要概念以便在尋找權重取樣分佈機率 過程中加入一股驅動力量使其不致陷於區域最佳解。我們稱此種修正的偵測法為粒子群趨動交叉熵法偵測法，其性能於高信雜比範圍有極顯著之改善並於低信雜比時維持原有優勢。最後，我們加入了通道估計誤差之考量並依此修正前述偵測法，電腦模擬的數值顯示經此修正後，我們的偵測法有相當穩健的性能表現。

A Particle-Swarm-Driven Cross-Entropy Method for Multiple-Input-Multiple-Output Signal Detection

Student : Chun-Lin Wang Advisor : Yu T. Su

Department of Communications Engineering

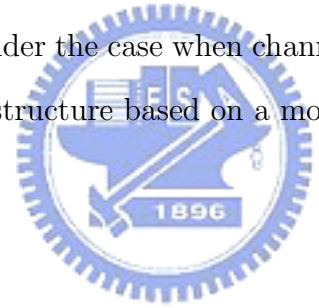
National Chiao Tung University

Abstract

Many solutions for detecting signals transmitted over flat-faded multiple input multiple output (MIMO) channels have been proposed, e.g., the zero-forcing (ZF), minimum mean squared error (MMSE), lattice reduction and V-BLAST algorithms, to name a few. However, these approaches suffer from either unsatisfactory performance or high complexity.

We present an alternative method for detecting quadrature amplitude modulated (QAM) MIMO signals. This method tries to estimate the probability distribution of the candidate signal location by sampling over a neighborhood of the received waveform. The proposed random sampling based iterative distribution estimator is similar to the class of Monte-Carlo based optimization approach and if the distance used in measuring the distance between a tentative distribution and the optimal distribution is the Kullback-Leibler distance (cross entropy) then our solution is identical to the one known as the Cross-Entropy (CE) method. The CE method is motivated by the search for an efficient rare-event simulation solution. The problem is equivalent to finding the optimal importance sampling density. The desired density is obtained by iterative random search in the space of exponential distributions with the CE metric.

The proposed CE-based detector yields bit-error-rate (BER) performance which is close to that achievable by the Maximum-Likelihood (ML) detector when the signal-to-noise ratio (SNR) is relatively low. Unfortunately the performance curves exhibit error floors in high SNR region. To improve the performance in high SNR region, we borrow the concept of particle swarm optimization (PSO) in designing our detector. PSO is a population-based iterative search algorithm which moves a number of particles through the feasible solution space towards the optimal solution with the information obtained in previous iterations. The modified iterative detector incorporates extra terms, which are generated by a PS-like process and represent a driving force to pull the iterative optimization process from being trapped in local minimums, in updating of the importance density and is called the particle-swarm-driven cross-entropy (PSD-CE) MIMO detector. The PSD-CE detector gives significant BER performance improvement in medium-to-high SNR region. We also consider the case when channel state information is imperfect and suggest a robust detector structure based on a modified score function.



誌　　謝

感謝我的指導教授　蘇育德老師這兩年來在研究上 不吝於指導和鼓勵，同時也在生活上時時給予關心，並帶給我許多人生的啟發。感謝口試委員鐘嘉德老師、蘇炫榮老師及李大嵩老師給予寶貴意見，以補足論文中的缺失與不足。感謝親愛的家人，漫長的求學生涯中家人無論在心靈上或生活上一直是我最大的支柱，由於你們無怨無悔的付出使我能無後顧之憂地完成學業。感謝實驗室同學、學長姐的幫忙鼓勵和學弟妹的陪伴，讓這兩年的碩班生活既充實又多采多姿。感謝一直以來在我身邊陪伴我、關心我的朋友們，在我沮喪時適時地為我加油打氣，和我一起分享生活上的快樂，一同經歷學習與生活中的酸甜苦辣。

感謝老師、家人、朋友和同學們的支持，是你們給我力量和勇氣克服每一次的困難，因為有你們這篇論文才得以順利完成，謝謝你們。

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Wireless communications are impaired predominantly by multipath fading [2]. However, in a richly scattered fading environment, fading can be beneficially as it promises a multiple-input-multiple-output (MIMO) system to achieve significant capacity gain through independent spatial modes [1]. For this reason, the MIMO technology has gained enormous popularity and attracted much research interest over the past decade [2]. Depending on the operation environment, a MIMO system possesses the potential to obtain (1) array gain, (2) spatial diversity gain, (3) spatial multiplexing gain and (4) interference reduction capability [1, 2].

Array gain refers to the increase in receive SNR through spatial processing at the receive antenna array and/or spatial pre-processing at the transmit antenna array. It improves the coverage and the range of a wireless network by raising resistance to noise. *Spatial diversity gain* alleviates fading of the signal level by providing the receiver with multiple copies of the transmitted signal in space, frequency or time. Ideally, these copies are independent and the number of independent copies is called the *diversity order*. The quality and reliability of the received signal improves as diversity order increases. A MIMO channel with N_T transmit antennas and N_R receive antennas offers a spatial diversity order of $N_T N_R$. Transmitting multiple and independent data streams within the operating bandwidth increases data rates and capacity of a wireless network, which is known as *Spatial multiplexing gain* [2]. In general, the number of data streams that

can be provided by a MIMO channel equals the minimum of the number of transmit antennas and the number of receive antennas, i.e. $\min\{N_T, N_R\}$. Exploiting the spatial dimension in a MIMO system might alleviate interference caused by sharing resources. In addition, directing signal energy towards the intended user rather than other users would avoid the impact of the interference. *Interference reduction and avoidance* improve the coverage and range of a wireless network [2].

Although these advantages can not exist simultaneously due to conflicting demands on the spatial degrees of freedom, some combinations across a wireless network would improve the system performance such as capacity and reliability.

Encouraged by the collective behavior of social animals such as fish schooling and the colony of ants, many genetic algorithms have been studied. In the work of Kennedy and Eberhart [3], *Particle Swarm Optimization* is motivated by bird-flocking behavior and is an iterative algorithm based on social-psychological model of social influence and social learning [4]. In the original model, individuals in a particle swarm follow a simple behavior. The collective behavior that emerges is that of discovering optimal regions of a high dimensional search space. At each iteration, each individual determines its nearest neighbor and replaces its velocity with that of its neighbor. To further extend the model, the "rooster" concept of Heppner and Grenander [5] was added, in the form of a memory of previous best and neighborhood best positions. The previous (personal) best position of each individuals is the best position found by the individual since the first simulation to the current iteration. The neighborhood best position is the best position found by the neighborhood. These two best positions serve as the attractor and the resulting model was referred to as particle swarm optimization. The swarm algorithm exhibits adaptive behavior since the state changes when personal best and global best position change.

The *Cross-Entropy* (CE) method is a general Monte Carlo approach to solve combinatorial and continuous multi-extremal optimization problems. The name is derived

from the cross-entropy distance (or the Kullback-Leibler distance) [6] which defines a distance between two probability density functions. This method was animated by an adaptive algorithm including the idea of minimizing variance for estimating probabilities of rare events in complex stochastic networks [7]. Soon after the first exploration, the fact was found that a simply modified version could be used not only for estimating probabilities of rare events but for solving difficult combinatorial optimization problems as well. This is done by translating an original deterministic optimization problem into a related stochastic estimation problem and applying the rare-event simulation mechanism to it [8].

The CE method is an iterative procedure and each iteration involves two phases [9] :

1. Generate a set of random samples according to a specified mechanism.

2. Update the parameters of the random mechanism based on the generated data in order to produce a better set of samples in the next iteration.

The advantage of the CE method is that it provides a simple adaptive procedure for estimating the optimal reference parameters. The fact that the updating rules are simple, explicit and optimal in some well-defined mathematical sense makes the CE method powerful and desirable. It provides a unifying approach to simulate and optimization and has great potential for exploring new search areas in the solution space.

The rest of this thesis is organized as follows. The ensuing chapter provides brief summary of the assumptions and models for the channel and system of concern. In Chapter 3, we review some general MIMO detection schemes including linear and non-linear detection methods. Chapter gives a detailed description of the Particle Swarm Algorithm. In the following chapter, the concepts of the cross-entropy-based MIMO detection method as well as the particle-swarm-driven cross-entropy MIMO detection method are proposed. Chapter 6 shows simulation performance of these algorithms are provided. Finally, our work is concluded. The notations used in this thesis are as follows.

Vectors and matrices are denoted by symbols in bold face. $(\cdot)^T$ and $(\cdot)^H$ represent transpose and Hermitian transpose, respectively. $\mathbb{E}\{\cdot\}$ denotes the statistical expectation. $\mathrm{tr}(\cdot)$ denotes the trace of an square matrix.

Chapter 2

System Model

2.1 Perfect Channel Estimation

Consider a MIMO system with N_T transmit antennas and N_R receive antennas with $N_R \geqslant N_T$. Input data is demultiplexed into N_T substreams which are mapped onto sequences of M-QAM symbols. The set of candidate signals in the constellation is denoted as $\mathcal{A}_M$. These substreams are transmitted simultaneously and received synchronously. For convenience but without loss of generality, we present a time-discrete complex baseband model for one time slot only.

Let x_i and y_j denote the complex valued signal transmitted by the ith antenna and those received by the jth receive antenna, $i = 1, \cdots, N_T$, $j = 1, \cdots, N_R$, respectively. Denote $\mathbf{H}$ as the overall channel matrix and assume, for the time being, that $\mathbf{H}$ is perfect known to the receiver. In other words, there is no channel estimation error at the receiver. The (j, i)th element of $\mathbf{H}$, $h_{j,i}$, is the channel response between the ith transmit antenna and the jth receive antenna. The MIMO system model just described is shown in Fig. 2.1. The received signal (vector) expressed in matrix form is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{2.1}$$

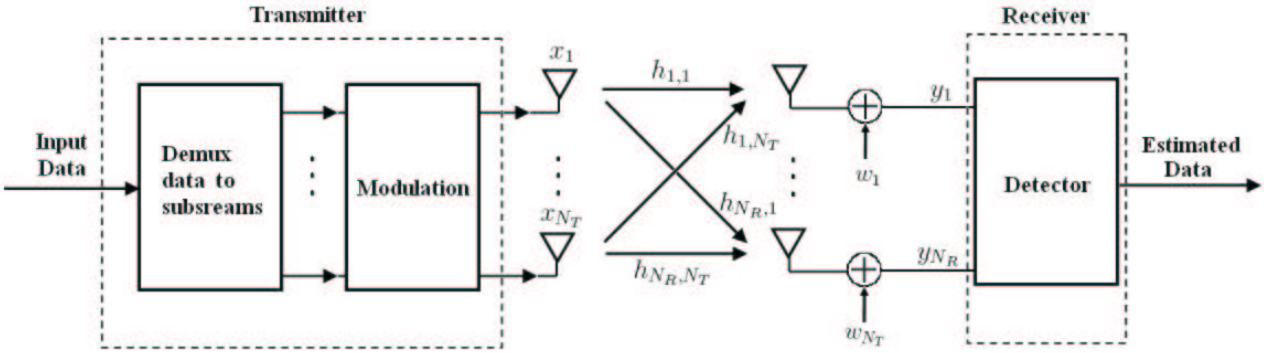


Figure 2.1: A MIMO system model.

where

$$\mathbf{x} = [x_1, \cdots, x_{N_T}]^T \in \mathcal{A}_M^{N_T}, \tag{2.2}$$

$$\mathbf{y} = [y_1, \cdots, y_{N_R}]^T \in \mathcal{C}^{N_R}, \tag{2.3}$$

$$\mathbf{w} = [w_1, \cdots, w_{N_R}]^T \in \mathcal{C}^{N_R}, \tag{2.4}$$

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,N_T} \\ \vdots & \ddots & \vdots \\ h_{N_R,1} & \cdots & h_{N_R,N_T} \end{bmatrix} \in \mathcal{C}^{N_R \times N_R}. \tag{2.5}$$

and $\mathcal{C}$ denotes the complex-valued domain. Every element of $\mathbf{H}$ is a complex Gaussian fading gain with unit variance, i.e. $\sigma_h^2 = 1$. The vector $\mathbf{w}$ represents the complex additive white Gaussian noise with zero mean and variance $\sigma_{\mathbf{w}}^2$. Each noise observed at the different antenna is assumed to be uncorrelated, i.e. $\mathbb{E}\{\mathbf{w}\mathbf{w}^H\} = \Sigma_{\mathbf{w}} = \sigma_{\mathbf{w}}^2 \mathbf{I}_{N_R}$ where $\mathbf{I}_{N_R}$ is an identity matrix of size $N_R \times N_R$. Throughout our work, the average transmit power of each antenna is also normalized to 1. In other words, $\mathbf{x}$ has a covariance matrix $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \sigma_{\mathbf{x}}^2 \mathbf{I}_{N_T}$ with $\sigma_{\mathbf{x}}^2 = 1$.

2.2 Pilot-based Channel Estimation

In practice, the assumption that $\mathbf{H}$ is perfect known by the receiver is not valid. There usually exists differences between the exact channel matrix and the estimated channel matrix due to channel estimation errors.

In order to estimate the channel matrix $\mathbf{H}$ by the receiver, a number of pilot symbols are sent prior to data symbols. Denote by $\mathbf{s}_i$ the $N_P \times 1$ pilot vector for ith transmit antenna and constitute the $N_T \times N_P$ matrix $\mathbf{S}$ as

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1^T \\ \vdots \\ \mathbf{s}_{N_T}^T \end{bmatrix}. \tag{2.6}$$

The receive vector $\mathbf{y}_P$ follows to

$$\mathbf{y}_P = \mathbf{HS} + \mathbf{w}_P \tag{2.7}$$

where $\mathbf{w}_P$ is AWGN noise with zero mean and covariance matrix $\sigma^2_{\mathbf{w}_P}\mathbf{I}_{N_R}$.

Define the channel matrix estimation error as $\Delta\mathbf{H}$, the channel known at the receiver can be written as

$$\hat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}, \tag{2.8}$$

and the linear least-square (LS) estimate of $\mathbf{H}$ is given by

$$\hat{\mathbf{H}} = \mathbf{y}_P\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}. \tag{2.9}$$

We thus have

$$\Delta\mathbf{H} = \mathbf{w}_P\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1} \tag{2.10}$$

It is known that mutual orthogonal pilot sequences will obtain the best channel estimate with uncorrelated estimation errors. Therefore, the rows of $\mathbf{S}$ are chosen to be orthogonal, i.e.

$$\mathbf{S}\mathbf{S}^H = N_P E_P \mathbf{I}_{N_T} \tag{2.11}$$

where E_P is the average power of the training symbols defined as

$$E_P = \frac{1}{N_P N_T}\mathrm{tr}(\mathbf{S}\mathbf{S}^H). \tag{2.12}$$

From [10], the conditional probability density function (PDF) of $\hat{\mathbf{H}}$ given $\mathbf{H}$ is a complex Gaussian distribution with mean $\mathbf{H}$ and covariance matrix $\mathbf{\Sigma}_{\Delta\mathbf{H}} = \sigma^2_{\Delta\mathbf{H}}\mathbf{I}_{N_T}$. Using the PDFs of H and $(\hat{\mathbf{H}}|\mathbf{H})$ with the conditions of mutually orthogonal pilot sequences and i.i.d. channel coefficients, we derive the PDF of $(\mathbf{H}|\hat{\mathbf{H}})$ as [10]

$$p(\mathbf{H}|\hat{\mathbf{H}}) = \mathcal{CN}(\delta\hat{\mathbf{H}}, \delta\sigma^2_{\Delta\mathbf{H}}\mathbf{I}_{N_T} \otimes \mathbf{I}_{N_R}) \tag{2.13}$$

with

$$\delta = \frac{\sigma_h^2}{\sigma_h^2 + \sigma^2_{\Delta\mathbf{H}}} \tag{2.14}$$

where $\mathcal{CN}(\cdot)$ denotes complex Gaussian distribution and $\otimes$ represents the Kronecker product.

Chapter 3

General MIMO Detection Schemes

We briefly survey four classes of popular MIMO detection schemes in this chapter for the convenience of subsequent discussions. These detectors are classified into two categories, namely, the linear and nonlinear detection schemes.

3.1 Linear Detectors

Using a $N_T \times N_R$ matrix $\mathbf{P}$ to linearly combine the elements of the received signal $\mathbf{y}$ is a straightforward approach to estimate the transmit signal $\mathbf{x}$. *Zero-Forcing* (ZF) and *Minimum-Mean-Square-Error* (MMSE) are the most two common methods in linear MIMO detection schemes.

3.1.1 Zero Forcing Dector

In a ZF detector, the interference introduced by the channel matrix is nulled out by multiplying the received signal vector $\mathbf{y}$ with the Moore-Penrose pseudo-inverse [11] of the channel matrix, i.e. $\mathbf{P}_{ZF} = \mathbf{H}^{+} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$. The transmit signal $\mathbf{x}$ is estimated by quantizing every element of the filter output vector to an element of the symbol alphabet,

$$\hat{\mathbf{x}} = \mathcal{Q}\{\mathbf{H}^{+}\mathbf{y}\} = \mathcal{Q}\{\mathbf{x} + (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{w}\}, \tag{3.1}$$

where $\hat{\mathbf{x}}$ is estimated transmit signal and $\mathcal{Q}$ denotes the quantization operation.

A drawback of a ZF detector is that nulling out interference without considering the noise might amplify the noise power significantly. For an orthogonal channel matrix, ZF is identical to the optimum detector, Maximum Likelihood (ML) detector. However, since the channel matrix is not ideal or orthogonal in practice, ZF leads to a noise enhancement problem generally.

3.1.2 Minimum Mean Squared Error Detector

To solve the noise enhancement problem in ZF detectors, MMSE detectors take the noise term into account and minimize the mean square error between the transmit signal and the estimated transmit signal, $J(\mathbf{P}) = E\{(\mathbf{x}-\hat{\mathbf{x}})^H(\mathbf{x}-\hat{\mathbf{x}})\}$, with respect to $\mathbf{P}$. The optimum matrix $\mathbf{P}_{MMSE}$ is given by

$$\mathbf{P}_{MMSE} = \left(\mathbf{H}^H\mathbf{H} + \frac{\sigma_{\mathbf{w}}^2}{\sigma_{\mathbf{x}}^2}\right)^{-1}\mathbf{H}^H. \tag{3.2}$$

Similar to ZF detectors, each element of the filter output is mapped by a minimum distance quantization so as to get the estimated transmit signal.

3.2 Nonlinear Detectors

3.2.1 Maximum Likelihood Detector

The maximum likelihood (ML) detector estimates the transmit signal $\hat{\mathbf{x}}$ by finding the one which minimizes the distance between the received signal vector and the estimated signal vector, i.e.

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{A}_M^{N_T}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||. \tag{3.3}$$

The problem can be solved by exhaustively searching over all possible $\mathbf{x}$ and choose the one that makes the distance smallest. However, although ML has the best performance among all MIMO detection algorithms, the number of all possible $\mathbf{x}$ increases with

N_T exponentially. The computational complexity will become prohibitively high when transmit antenna N_T and constellation size M increase.

3.2.2 V-BLAST Algorithm

Vertical Bell Labs layered space time algorithm (V-BLAST) is a popular nonlinear combining approach for detecting MIMO signals [12]. V-BLAST detection employs an ordered serial detect-and-cancel strategy similar to that of decision-feedback equalizer. At the receiver, a successive interference cancellation concept is applied such that each substream is considered as the desired signal in turn while the others are regarded as interferers. With a linear combinatorial nulling technique, nulling is performed in a particular detection order by linear weighting the received signals to satisfy a specific criterion such as ZF or MMSE. Taking ZF nulling for example, the weighting vectors $\mathbf{z}_i$ of size $N_R \times 1$ for $i = 1, \cdots, N_T$ are chosen to meet

$$\mathbf{z}_i^T (\mathbf{H})_j = \delta_{ij}. \tag{3.4}$$

where $(\mathbf{H})_j$ is the jth column of $\mathbf{H}$ and δ is the Kronecker delta. And the ith substream is estimated by

$$\hat{x}_i = Q(\mathbf{z}_i^T \mathbf{y}). \tag{3.5}$$

Interference from already-detected components of $\mathbf{x}$ is subtracted out from the received signal. Therefore, the received vector is modified iteratively and the interferers are less than that at the previous iteration. Due to the symbol cancellation used in this algorithm, the weighting vectors for ZF nulling can be revised as

$$\mathbf{z}_i^T (\mathbf{H})_j = \begin{cases} 0, & \text{if } \ j \geq i \\ 1, & \text{if } \ j = i \end{cases} \tag{3.6}$$

Denote the detection order set by $\mathcal{O} = \{s_1, \cdots, s_{N_T}\}$. The general detection process is described in Table 3.1.

Step 1 :	**Set $k = 1$.**
Step 2 :	**Choose the nulling vector $\mathbf{z}_{s_k}$ for s_kth substream in accordance with equation (3.6).**
Step 3 :	**Calculate the decision statistic for s_kth substream and quantize it to obtain $\hat{x}_{s_k}$ as equation (3.5)**
Step 4 :	**Assume $\hat{x}_{s_k} = x_{s_k}$, and suppress the component of x_{s_k} from the received signal $\mathbf{y}$. Modify the received vector as $\mathbf{y} = \mathbf{y} - \hat{x}_{s_k}(\mathbf{H})_j$.**
Step 5 :	**Stop at iteration $k = N_T$; otherwise, let $k = k + 1$ and go back to Step 2.**

Table 3.1: The V-BLAST detection algorithm.

The detection order plays an important role in this algorithm. An improper order will induce a error propagation problem. It is proved in [12] that the optimal detection order is determined to maximize the minimum post-detection signal-to-noise ratio (SNR) of all data streams. Therefore, selecting the "best" (smallest) post-detection SNR at each stage in the process leads to the global optimum ordering. The post-detection SNR for detected component x_{s_k} of $\mathbf{x}$ is obtained by

$$\rho_{s_k} = \frac{\mathbb{E}\{|x_{s_k}|^2\}}{\sigma_{\mathbf{w}}^2 ||\mathbf{z}_{s_k}||^2} \tag{3.7}$$

where the expectation is taken over the constellation set.

Chapter 4

Particle Swarm Optimization

4.1 Particle Swarm Algorithm

Inspired by the movements of birds flocking, the particle swarm algorithm is an optimization technique for a real-valued multidimensional solution space [4, 14]. It adjusts the trajectories of a population of "particles" (or samples) through the feasible solution space iteratively. Every particle is evolved based on the information about each particle's previous best performance and the best previous best performance of its neighborhood. They traverse a solution space where a quality measure, fitness can be evaluated. Through cooperation and competition among these particles over several iterations, all particles can move towards the optimal position. This approach is attractive due to its advantages of the simple mathematical model, resistance to being trapped in a local optimum and faster convergence. For MIMO detection, binary particle swarm optimization (PSO) are applied.

The coordinates of every particle represent a possible solution associated with two vectors, the **position** and the **velocity**. In binary PSO, the elements in velocity vector is squashed by a sigmoid function, $Sig(x) = \frac{1}{1+e^{-x}}$, to the range $(0, 1)$ and is used to determine whether the corresponding elements in position vector is either 0 or 1. In K-dimensional search space, the ith particle is represented by the position vector

$\mathbf{x}_i = [x_{i1}, \cdots, x_{iK}]^T$ and the velocity vector $\mathbf{v}_i = [v_{i1}, \cdots, v_{iK}]^T$. In a MIMO detection problem, $\mathbf{x}_i$ is regarded as a candidate solution. Applying the PS algorithm to this problem, the **fitness function** is defined as

$$f(\mathbf{x}) = ||\mathbf{y} - H\mathbf{x}||^2. \tag{4.1}$$

A fitness value of a particle is assigned to the particle's current location by using the coordinates of the particle. Also, let $\mathbf{g}^b$ and $\mathbf{p}_i^b$ denote the position vector of the particle with the best performance among its neighbors so far and the position vector of ith particle with the best performance along its previous fitness values, respectively.

Velocity vector $\mathbf{v}_i$ at tth iteration is updated according to the following equation:

$$\begin{aligned} &v_{ik}(t) = v_{ik}(t-1) + \phi_1[p_{ik}^b - x_{ik}(t-1)] + \phi_2[g_k^b - x_{ik}(t-1)] \\ &\text{with} \quad v_{ik} \in \{-v_{max}, v_{max}\} \end{aligned} \tag{4.2}$$

for $k = 1, \cdots, K$, where ϕ_1 and ϕ_2 are positive random numbers drawn from a uniform distribution with a predefined upper limit. In binary PSO, the limit is arbitrary but the sum of ϕ_1 and ϕ_2 are usually set to be less than 4 [14]. The first term in equation (4.2) indicate the impact of the velocity at $(t-1)$th iteration. The last two terms are considered as cognitive part and social part, respectively. The cognitive part denotes the effect of the evolution of ith particle itself. The social part represents the interaction between ith particle and its neighborhood.

In continuous PSO, the position vector is updated by the velocity vector as [15]

$$\mathbf{x}_i(t) = \mathbf{x}_i(t-1) + \mathbf{v}_i(t-1). \tag{4.3}$$

where t is the index of iteration. Vector representation for relationship between velocity and position is shown in Figure 4.1. In binary PSO [15], the elements in velocity vector after being squashed represents the probabilities of the elements in position vector taking the value 1. To update the position vector $\mathbf{x}_i$ at tth iteration, the binary decision rule

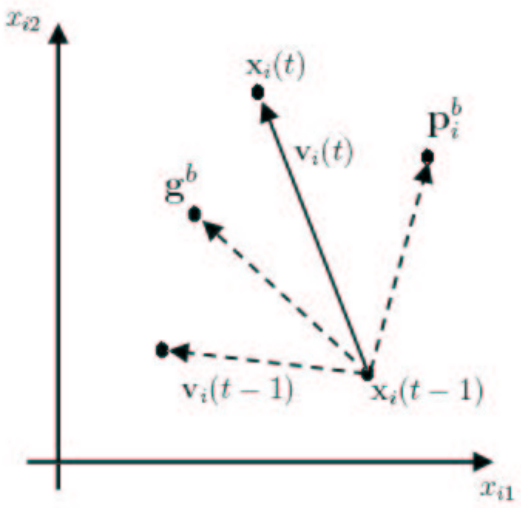


Figure 4.1: Vector representation for relationship between velocity and position with $k = 2$.

shown below is followed:

$$\begin{aligned} &\text{If} \quad \text{rand}() < Sig(v_{ik}(t)), \quad \text{then} \quad x_{ik}(t) = 1 \\ &\text{else}, \quad x_{ik}(t) = 0 \end{aligned} \tag{4.4}$$

for $k = 1, \cdots, K$, where rand$(\cdot)$ is a random number selection function from a uniform distribution in $[0, 1]$. With these new position vectors, fitness values of the particles are evaluated. In the meanwhile, $\mathbf{g}^b$ and $\mathbf{p}_i^b$ are updated again, and so on. This procedure is repeated until maximum number of iterations is reached and the estimated transmit signal is decided by $\mathbf{g}^b$. The flow diagram of the Particle Swarm algorithm is shown in Figure 4.2.

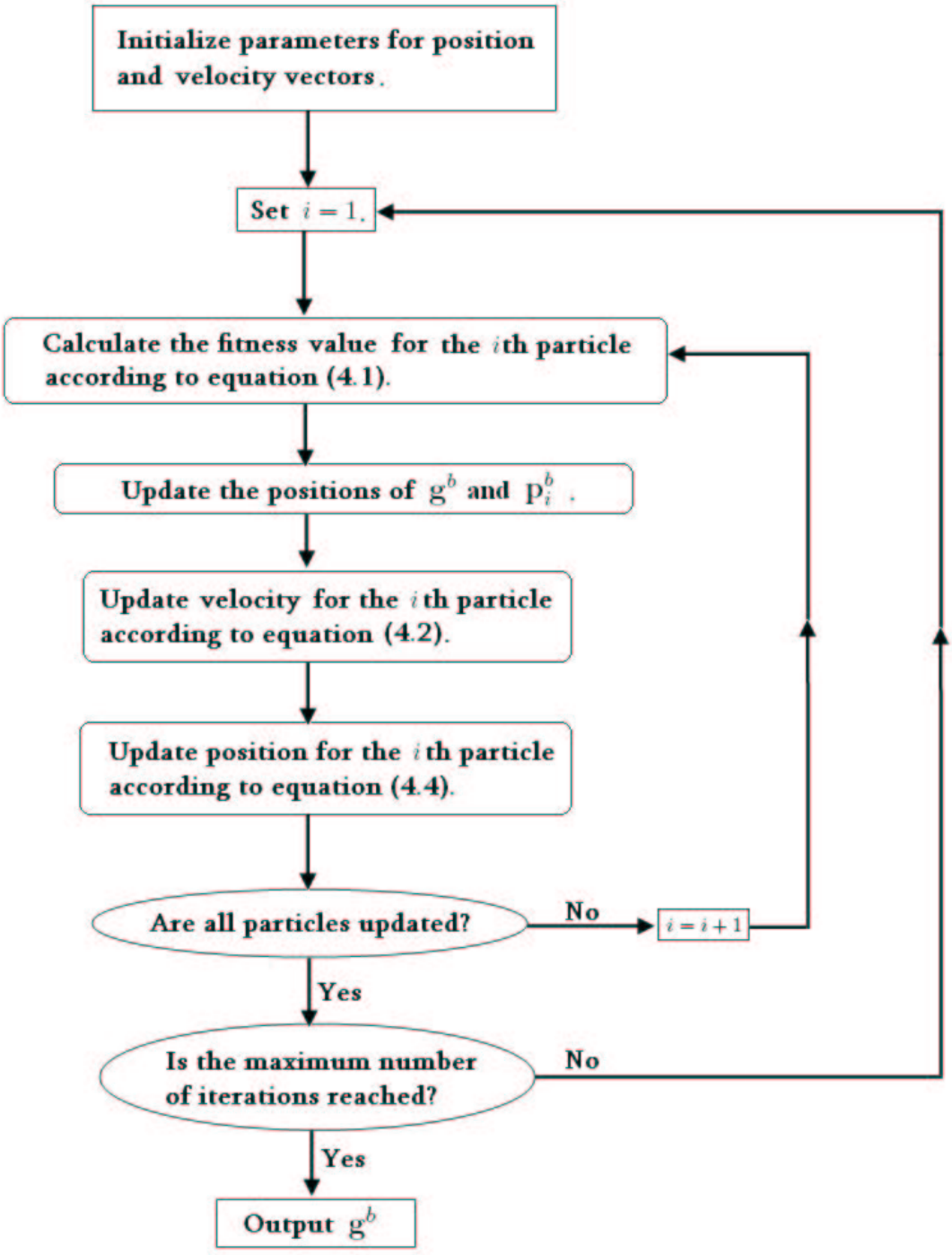


Figure 4.2: The Particle Swarm algorithm flow chart.

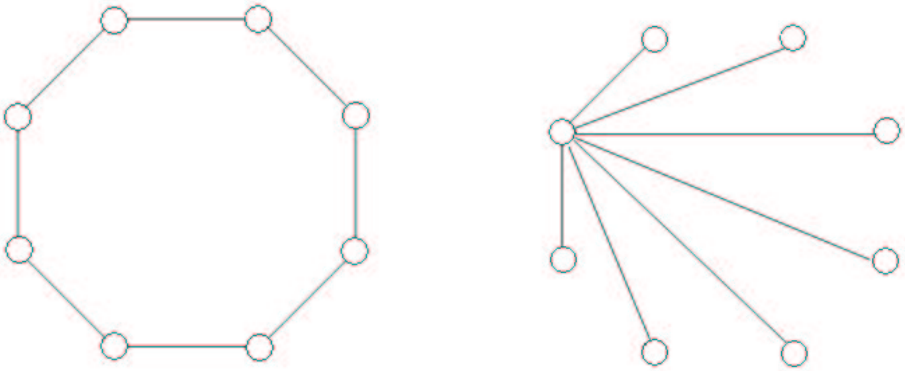
4.2 Variations of Particle Swarm Algorithm

In the PS algorithm, there are some parameters related to convergence speed and performance, such as the number of particles in the swarm, the neighborhood topology and the acceleration coefficients [4, 14]. The influences of these basic parameters are discussed in this section.

The more number of particles in the swarm, also called the *swarm size*, the larger the initial diversity of the swarm. A wider range of search space will be explored at each iteration if the swarm size is larger. Furthermore, it is more possible that more particles would reach the optimal solution within fewer iterations. Nevertheless, the computational complexity increase while a large swarm size is used and the swarm degrades to a parallel random search. Generally, fewer particles are needed for a smooth search space than that for a rough surface. The optimal swarm size is problem-dependent and is suggested to be optimized for each problem through cross-validation methods instead the heuristics found in publications.

In the PS algorithm, $\mathbf{g}^b$ denote the position vector of the particle with the best performance among its neighbors. The neighborhood size represents the quantity of social interaction occurring within the swarm. Smaller neighborhood sizes leads to a slow convergence and are insensitive to a local minimum, thus there are more dependable convergence to the optimal solution. To ensure an initial high diversity with faster convergence, the search can start with small neighborhoods. As the number of iterations increases, the size of neighborhood should be enlarged so as to move all particles to a promising search area. Two extreme cases for neighborhood topology are presented in Figure. 4.3. In Figure. 4.3(a) the neighborhood of every particle are the ones next to it while the neighbors for every particle in Figure. 4.3(b) are all other particles in the swarm .

In equation (4.2), the two random numbers ϕ_1 and ϕ_2 are regarded as *acceleration coefficients* which control the stochastic effects of the cognitive and social components



(a) (b)

Figure 4.3: Two examples for neighborhood topology.

over the procedure.

Particles move through smooth trajectories if ϕ_1 and ϕ_2 are small. They might explore more good areas rather than being stocked in a good region found before. On the contrary, large values for ϕ_1 and ϕ_2 lead to more acceleration, and particles move towards past good regions with hasty movements.

Initialization for ϕ_1 and ϕ_2 are important because an improper initialization may lead to divergent or cyclic behavior over the algorithm. In general, ϕ_1 and ϕ_2 are optimized with the values ensuring convergence first and remain static during the procedure.

Chapter 5

Cross-Entropy-based MIMO Signal Detection

5.1 Generic Cross-Entropy Method

5.1.1 Importance Sampling

The *Cross Entropy* (CE) method attempts to solve an optimization problem by relating it to a rare event simulation problem [9]. Usually, the sample size for estimating a rare event probability is very large. **Important Sampling** (IS) is a well-known technique used to reduce the variance by simulating a system under a different set of parameters (reference parameters) or under a different probability distribution. With this technique, the rare event is much more likely to occur so that the sample size of the rare event simulation can be reduced. In conventional IS, the optimal reference parameters are difficult to obtain. The CE method provides a simple and fast procedure to estimate the optimal reference parameters used in IS. More specifically, the CE method is an iterative Monte-Carlo based approach to find the IS density, i.e. the **importance distribution**, which is closest to the optimal important distribution in the Kullback-Leibler sense. The Kullback-Leibler distance $D(g, h)$ is a distance measure of

two different distributions $g(\mathbf{x})$ and $h(\mathbf{x})$ which defines as follows:

$$D(g,h) = \int g(\mathbf{x}) \ln \frac{g(\mathbf{x})}{h(\mathbf{x})} \mu(d\mathbf{x}). \tag{5.1}$$

A distance metric must satisfy the three rules below:

(1) The distance is non-negative.

(2) Symmetric property: The distance between two points is the same while measuring from either direction.

(3) Triangle inequality: Considering a triangle formed by three points, the sum of any two edges is larger than the third edge.

Therefore, the K-L distance is not a true distance metric since it is not symmetric and does not satisfy the triangle inequality.

5.1.2 A Generalized CE Method for Optimization

In this subsection, we give a brief description to the relationship between IS and the CE method for optimization problem. The CE method attempts to solve the following optimization function

$$\arg\max_{\omega \in \Omega} S(\omega) \tag{5.2}$$

where Ω is the domain of variable ω and S is the score function of ω defined on Ω. Applying IS to this problem, we find another set of parameter, e.g. v, instead of ω. To find the optimal importance distribution within a class of densities $f(\omega; v)$, we adapts the parameter v iteratively so that the Kullback-Leibler distance (i.e. the cross entropy) between the associated density and the optimal importance distribution is minimized. In general, a generic CE method can be described by the following steps :

1. Generate samples according to the importance distribution determined at the previous iteration.

2. Calculate the scores to the generated samples according to a specific score function.

3. Update the importance distribution by the samples with comparatively better scores.

4. Repeat the above steps until the stopping criterion is reached.

At the very beginning, we give a initial distribution to the importance distribution and generate a set of samples depending on it. Then we compute the scores for every sample individually. For an optimization problem, the value of the objective function are usually regarded as the score for each sample. The samples with better scores are called **elite samples** and the set composed of elite samples is defined as the **elite set**. We choose those elite samples to update the importance distribution. The updated importance distribution is a linear combination of the original importance distribution and the distribution determined by the elite samples. Again, new samples are generated according to the updated importance distribution and the same steps mentioned above are repeated. This procedure is processed iteratively until the stopping criterion is reached.

5.2 Cross-Entropy-based MIMO Signal Detection

5.2.1 A CE-based MIMO detection algorithm

To apply the CE method to a MIMO system, we first define a score function

$$S(\mathbf{x}) = ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2 \tag{5.3}$$

under the assumption of perfect channel estimation for solving the following optimization problem,

$$\arg \min_{\mathbf{x} \in \mathcal{A}_M^{N_T}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||. \tag{5.4}$$

Following the procedures of the CE method, the importance distribution of $\mathbf{x}$ with relatively smaller scores are estimated by minimizing the distance between the initial distribution and the optimal importance distribution. The estimated transmit signal $\hat{\mathbf{x}}$ is the symbol that is the most likely to occur according to the distribution or the sample with the smallest score during the whole process. Intuitively, an estimated transmit signal vector $\hat{\mathbf{x}}$ is regarded as a sample and the score will be calculated for the sample. However, there are M^{N_T} possible candidates if we treat a vector as a sample unit. Thus a large sample size may be required to cover a wider search region so that the computing complexity would inevitably increase. In order to avoid this problem, the importance distribution of every element in a transmit signal $\mathbf{x}$ is estimated separatively.

Let $f^{(k)}(x_i)$ denote the importance distribution of the ith element for $i = 1, \cdots, N_T$, where the superscript k is the index of iteration. U samples, $x_{i,u}^k$ for $u = 1, \cdots, U$, are generated at kth iteration in accordance with $f^{(k)}(x_i)$ for the ith element. To calculate the scores for these samples, a vector set $\{\mathbf{x}_u^k\}_{u=1}^U$ is constructed where $\mathbf{x}_u^k = [x_{1,u}^k, \cdots, x_{i,u}^k, \cdots, x_{N_T,u}^k]^T$ represents the uth sample vector at kth iteration.

Given a specific quantile probability ρ, there are infinite number of thresholds such that the probability of the scores less or equal to these thresholds are larger or equal to ρ. To select elite samples, we choose the threshold at kth iteration γ^k satisfying

$$\gamma^k = \arg\min_{\gamma} P(S(\mathbf{Z}) \leq \gamma) \geq \rho \quad \text{for} \quad \mathbf{Z} \in \{\mathbf{x}_u^k\}_{u=1}^U \tag{5.5}$$

And the elite samples are those whose scores satisfies $S(\mathbf{x}_u^k) \leq \gamma^k$. The distributions of elite samples are calculated as

$$f_s^{(k)}(x_i = a) = \frac{\sum_{u=1}^U \mathbf{I}_{\{S(\mathbf{x}_u^k) \leq \gamma^k\}} \mathbf{I}_{\{x_{i,u}^k = a\}}}{\sum_{u=1}^U \mathbf{I}_{\{S(\mathbf{x}_u^k) \leq \gamma^k\}}} \tag{5.6}$$

where $a \in \mathcal{A}_M$ for $i = 1, \cdots, N_T$.

Based on these elite samples, the importance distributions $f^{(k)}(x_i)$ for $i = 1, \cdots, N_T$ are updated according to

$$f^{(k+1)}(x_i = a) = \alpha f_s^{(k)}(x_i = a) + (1 - \alpha) f^{(k)}(x_i = a) \tag{5.7}$$

where α is the weighting factor and $0 \leq \alpha < 1$. The updated importance distributions are linear combinations of the original importance distributions and the distribution of the elite samples.

The procedure described above is repeated iteratively until the stopping criterion is met. For example, the pre-defined number of iterations is reached or the importance distributions converges. This algorithm is listed as shown in Table 5.1.

Step 1 :	**Initialize the importance distributions $f^{(k)}(x_i)$ with uniform distribution for $i = 1, \cdots, N_T$, respectively. And set $k = 0$.**
Step 2 :	**Generate U samples $x_{i,u}^k$ from $f^k(x_i)$ for $u = 1, \cdots, U$. Construct the set $\{\mathbf{x}_u^k\}_{u=1}^U$ where $\mathbf{x}_u^k = [x_{1,u}^k, \cdots, x_{i,u}^k, \cdots, x_{N_T,u}^k]^T$.**
Step 3 :	**Calculate the set of scores $\{S(\mathbf{x}_u^k)\}_{u=1}^U$ according to equation (5.3).**
Step 4 :	**Set a quantile parameter ρ such that there is a γ^k satisfying equation (5.5).**
Step 5 :	**Calculate the distribution of elite samples in accordance with equation (5.6).**
Step 6 :	**Update the importance distributions according to equation (5.7).**
Step 7 :	**Stop at iteration $k = K$ if the pre-defined stopping criterion is met; otherwise, let $k = k + 1$ and go back to Step 2.**

Table 5.1: The cross-entropy-based MIMO detection algorithm.

5.2.2 Weighting Factors

In equation (3.6), the weighting factor α effects the **exploration** and **exploitation** ability. Exploration is an ability to explore more different regions in the search space to find the global optimum. On the contrary, exploitation is an ability to concentrate the search around the a specific region in order to refine a candidate solution. If α is

larger, the component of updated distribution depends on more information from elite samples. Therefore, it is more likely to exploit than to explore and the convergence speed is faster. However, it is also much more possible to trap in one area since the elite samples generated at first several iterations may lead to a local minimum. If α is small, it takes more iterations to converge but has a higher probability to locate the global optimum.

Usually, to balance the exploration and exploitation, the weighting of the original importance distribution, i.e. the importance distribution at the previous iteration, is about twice that of the distribution of elite samples from some experiences of simulations. Therefore, α is chosen to be about 0.3 in our work. For other optimization problems, it depends.

5.2.3 Simulation Result of the CE-based Detection Method

Figure 5.1 shows the BER performance of ML detection and the CE-based detection method under a 4×4 MIMO system using 4-QAM.

As shown in the figure, our simulation indicates that the resulting BER performance is close to that of ML detector at low SNR region. However, there exists an error floor when SNR is larger than about 10dB since the estimated importance distributions do not converge uniformly, i.e., some of the importance distributions do converge but not all the N_T importance distributions.

5.3 Particle-Swarm-Driven Cross-Entropy Methods

5.3.1 A PS-Driven CE MIMO Detection Algorithm

Inspired by the evolution concept of swarm algorithm, the updating formula of importance distribution, i.e. equation (5.7), in the CE-based MIMO detection method

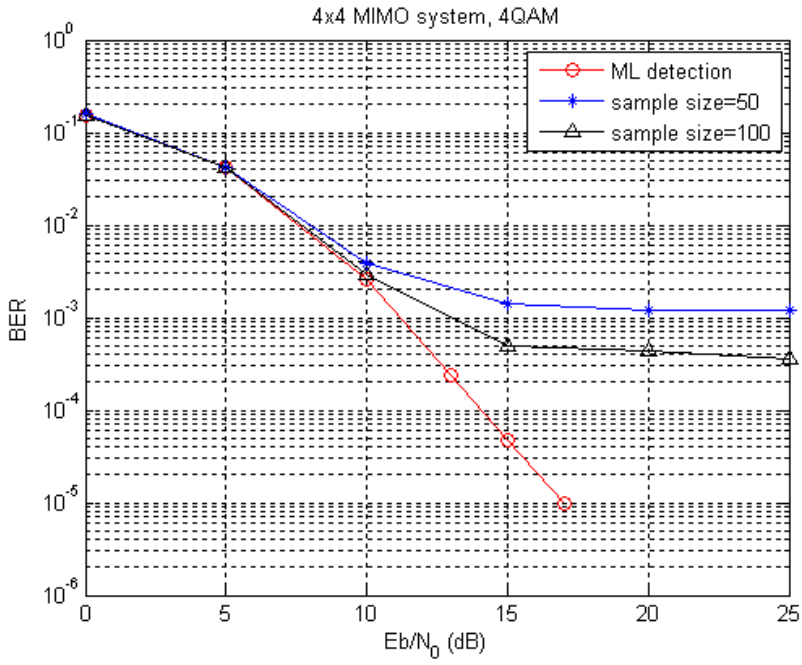


Figure 5.1: BER performance comparison of ML detection and the CE-based detection with $\alpha = 0.3$.

is modified to solve the problem of nonuniform convergence mentioned in the previous section. The main idea is to keep the newly generated samples close to the generated samples with the best scores. Except for the elite samples generated in the current iteration, we enhance the effect of sample vectors with the best score in the current iteration and those among all iterations so far. The updated importance distribution thus become a mixture of distributions determined by the current elite set, the current best sample vector(s) and the overall best sample vector(s).

Denote by $\mathbf{x}_{g(1)} = [x_{g(1),1}, \cdots, x_{g(1),N_T}]^T$ the best sample vector from the first iteration to the current iteration where the index "1" represents the rank of score. Similarly, $\mathbf{x}_{p(1)} = [x_{p(1),1}, \cdots, x_{p(1),N_T}]^T$ is defined as the best sample vector in the current iteration. At each iteration, $\mathbf{x}_{g(1)}$ and $\mathbf{x}_{p(1)}$ are recorded and updated. New added distributions at kth iteration $f^{(k)}_{g(1)}(x_i)$ and $f^{(k)}_{p(1)}(x_i)$ for $i = 1, \cdots, N_T$, are given based on these two

vectors respectively as

$$f_{g(1)}^{(k)}(x_{g(1),i}=a)=\begin{cases}\lambda p, & \text{if } a\in\mathcal{A}'_M\\ p\ , & \text{if } a\notin\mathcal{A}'_M\end{cases} \tag{5.8}$$

and

$$f_{p(1)}^{(k)}(x_{p(1),i}=a)=\begin{cases}\lambda p, & \text{if } a\in\mathcal{A}'_M\\ p\ , & \text{if } a\notin\mathcal{A}'_M\end{cases} \tag{5.9}$$

subject to the constraints

$$\sum_{a\in\mathcal{A}_M} f_{g(1)}^{(k)}(x_{g(1),i}=a)=1 \tag{5.10}$$

and

$$\sum_{a\in\mathcal{A}_M} f_{p(1)}^{(k)}(x_{p(1),i}=a)=1, \tag{5.11}$$

where $p\in(0,1)$ denotes a probability, $\mathcal{A}'_M$ is a subset of $\mathcal{A}_M$ containing all neighbors of a, and λ is a constant positive integer. In our work, $\mathcal{A}'_M$ is defined as all nearest points of a. Taking $a=1+i$ in 16-QAM for example, the region of $\mathcal{A}'_M$ is shown in Figure 5.2. Even though a is not the global optimal point exactly, the true optimum may occur within its neighbors intuitively. Hence, higher probabilities are assigned to the four points nearest a and a itself.

Including these components into the updating of the CE-based MIMO detection method, the updating formula of importance distributions are revised as follows,

$$\begin{aligned} f^{(k+1)}(x_i=a) &= \alpha_1 f_s^{(k)}(x_i=a)+\alpha_2 f_{g(1)}^{(k)}(x_i=a)+\alpha_3 f_{p(1)}^{(k)}(x_i=a)\\ &\quad+\left(1-\sum_{i=1}^{3}\alpha_i\right) f^{(k)}(x_i=a), \end{aligned} \tag{5.12}$$

where α_i are weighting factors with $0\le\alpha_i<1$ for $i=1,2,3$ and $\sum_{i=1}^{3}\alpha_i<1$.

The algorithm of the Particle-Swarm-driven CE MIMO detection method is listed in Table 5.2. As shown in the simulations, the PS-driven CE MIMO detection method provides a significant improvement in the BER performance compared with the CE-based MIMO detection method.

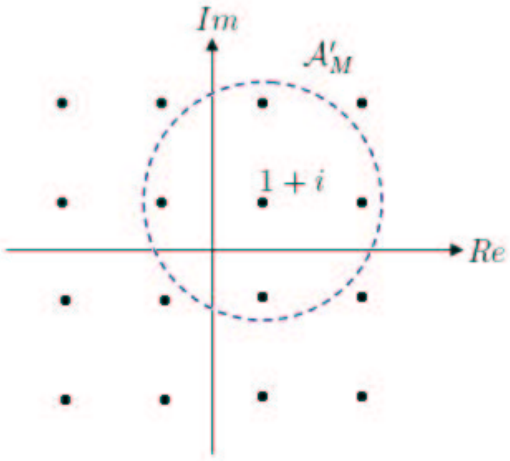


Figure 5.2: An example of region of $\mathcal{A}'_M$ as $a = 1 + i$ with 16-QAM.

Step 1 : Initialize the importance distributions $f^{(k)}(x_i)$ with uniform distribution for $i = 1, \cdots, N_T$, respectively. And set $k = 0$.

Step 2 : Generate U samples $x^k_{i,u}$ from $f^{(k)}(x_i)$ for $u = 1, \cdots, U$. Construct the set $\{\mathbf{x}^k_u\}^U_{u=1}$ where $\mathbf{x}^k_u = [x^k_{1,u}, \cdots, x^k_{i,u}, \cdots, x^k_{N_T,u}]^T$.

Step 3 : Calculate the set of scores $\{S(\mathbf{x}^k_u)\}^U_{u=1}$ according to equation (5.3).

Step 4 : Set a quantile parameter ρ such that there is a γ^k satisfying equation (5.5).

Step 5 : Calculate the distribution of elite samples in accordance with equation (5.6).

Step 6 : Update the sample vector with the best score overall (from the 1st to the kth iteration), $\mathbf{x}^k_{g(1)}$, and the best sample vector in the current iteration, $\mathbf{x}^k_{p(1)}$.

Step 7 : Calculate the component importance distributions $f^{(k)}_{g(1)}$ and $f^{(k)}_{p(1)}$ according to equation (5.8) and (5.9).

Step 8 : Update the importance distributions following the revised updating equation (5.12).

Step 9 : Stop at iteration $k = K$ if the pre-defined stopping criterion is met; otherwise, let $k = k + 1$ and go back to Step 2.

Table 5.2: The Particle-Swarm-driven CE MIMO detection algorithm.

As mention in the previous section, the weighting factors follow the rule of that in the CE-based MIMO detection method. The weighting for the distribution of the previous iteration remains most part of the updating equation.

5.3.2 Modifications for the PS-Driven CE Method

Enhancement of the effects of sample vectors in the current iteration and those among all iterations leads to a uniform convergence. However, it may encounter the problem of being trapped in a local minimum. To draw a higher probability to explore the global minimum during the procedure, the samples vectors with the second best score are also considered.

Let $\mathbf{x}_{g(2)}$ and $\mathbf{x}_{p(2)}$ denote the sample vectors with the second best score among all iterations and in the current iteration, respectively. Similar to $f^{(k)}_{g(1)}(x_i)$ and $f^{(k)}_{p(1)}(x_i)$, $f^{(k)}_{g(2)}(x_i)$ and $f^{(k)}_{p(2)}(x_i)$ are given depending on these two vectors. Including the two new added distributions to the updating formula of importance distributions, equation (5.12) can be modified as

$$\begin{aligned} f^{(k+1)}(x_i = a) &= \alpha_1 f^{(k)}_s(x_i = a) + \alpha_2 f^{(k)}_{g(1)}(x_i = a) + \alpha_3 f^{(k)}_{g(2)}(x_i = a) \\ &\quad + \alpha_4 f^{(k)}_{p(1)}(x_i = a) + \alpha_5 f^{(k)}_{p(2)}(x_i = a) \\ &\quad + \left(1 - \sum_{i=1}^{5} \alpha_i\right) f^{(k)}(x_i = a), \end{aligned} \tag{5.13}$$

According to equation(5.13), any one component can be dropped if the corresponding weighting factor is assigned to zero. Thus there are several combinations of the composite components for the updating formula. To further extend this model, the sample vector with the third best score can even be considered. However, the sample vector with too poor score will degrade the performance. Therefore, to enhance the exploration ability of this algorithm, the sample vectors used should have relatively better scores but not

too worse at all. The discussion of different combinations for those component is given in Chapter 6 with their BER performances.

5.4 Score Function under Imperfect Channel Estimation

So far, the channel state information is assumed perfect known at the receiver. Thus the score function used in the proposed algorithms is defined as

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

Recalling a detector in the ML sense and define the likelihood function as $L(\mathbf{y}|\mathbf{H}, \mathbf{x})$, the ML detector estimates $\mathbf{x}$ by maximizing the likelihood function under i.i.d. Gaussian noise:

$$\begin{aligned} \hat{\mathbf{x}}_{ML} &= \arg \max_{\mathbf{x} \in \mathcal{A}_M^{N_T}} L(\mathbf{y}|\mathbf{H}, \mathbf{x}) && (5.14) \\ &= \arg \min_{\mathbf{x} \in \mathcal{A}_M^{N_T}} (-\log L(\mathbf{y}|\mathbf{H}, \mathbf{x})) && (5.15) \\ &= \arg \min_{\mathbf{x} \in \mathcal{A}_M^{N_T}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||. && (5.16) \end{aligned}$$

However, when imperfect channel estimation is taken into account, the score function must be modified in the presence of channel estimation errors. Denote $L_m(\mathbf{y}|\hat{\mathbf{H}}, \mathbf{x})$ the modified likelihood function which is obtained by averaging $L(\mathbf{y}|\mathbf{H}, \mathbf{x})$ over all estimation errors as [10]

$$\begin{aligned} L_m(\mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}) &= \int_{\mathbf{H} \in \mathcal{C}^{N_R \times N_T}} L(\mathbf{y}|\mathbf{H}, \mathbf{x}) p(\mathbf{H}|\hat{\mathbf{H}}) \mathrm{d}\mathbf{H} && (5.17) \\ &= \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \left[L(\mathbf{y}|\mathbf{H}, \mathbf{x}) | \hat{\mathbf{H}} \right]. && (5.18) \\ &= \mathcal{CN}(\delta \hat{\mathbf{H}}\mathbf{x}, \mathbf{\Sigma}_{\mathbf{w}} + \delta \mathbf{\Sigma}_{\Delta \mathbf{H}} ||\mathbf{x}||^2) && (5.19) \end{aligned}$$

Using equation(5.19), the ML estimate under imperfect channel estimation is revised

as

$$\hat{\mathbf{x}}_{ML} = \arg \min_{\mathbf{x} \in \mathcal{A}_M^{N_T}} \left(-\log L_m(\mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}) \right) \tag{5.20}$$

$$= \arg \min_{\mathbf{x} \in \mathcal{A}_M^{N_T}} N_R \log \pi(\sigma_{\mathbf{w}}^2 + \delta\sigma_{\Delta\mathbf{H}}^2 ||\mathbf{x}||^2) + \frac{||\mathbf{y} - \delta\hat{\mathbf{H}}\mathbf{x}||^2}{\sigma_{\mathbf{w}}^2 + \delta\sigma_{\Delta\mathbf{H}}^2 ||\mathbf{x}||^2}. \tag{5.21}$$

Hence the modified score function in the presence of imperfect channel estimation is defined as

$$f_m(\mathbf{x}, \mathbf{y}, \hat{\mathbf{H}}) \triangleq -\log L_m(\mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}) \tag{5.22}$$

$$= \frac{||\mathbf{y} - \delta\hat{\mathbf{H}}\mathbf{x}||^2}{\sigma_{\mathbf{w}}^2 + \delta\sigma_{\Delta\mathbf{H}}^2 ||\mathbf{x}||^2} + N_R \log \pi(\sigma_{\mathbf{w}}^2 + \delta\sigma_{\Delta\mathbf{H}}^2 ||\mathbf{x}||^2) \quad . \tag{5.23}$$

The first term in equation (5.23) is similar to the original score function. It also indicates that the modified score function only includes the term $\sigma_{\Delta\mathbf{H}}^2$ instead of other detailed information about estimations errors.

Chapter 6

Simulation Results

In this chapter, we examine the performance of various proposed MIMO detectors. A MIMO system with $N_T = 4$ transmit antennas and $N_R = 4$ receive antennas with 4-QAM modulation is considered. Let E_b be the average received energy per information bit and denote by N_0 the noise power density. The signal-to-noise (SNR) ratio is defined as

$$\frac{E_b}{N_0} = \frac{N_R}{\log_2(M)\sigma_{\mathbf{w}}^2}. \tag{6.1}$$

6.1 Perfect Channel Estimation

The channel matrix is assumed perfect known by the receiver in this section. Fig. 6.1 shows the bit error rate (BER) of the CE-based detection method with $\alpha = 0.3$ and the PS-driven CE MIMO detection method with $\alpha_1 = \alpha_2 = 0.1$,$\alpha_3 = 0.2$ and $\beta = 0.3$. It is obvious to see that the proposed PS-driven CE detection algorithm can effectively solve the problem of nonuniform convergence in the CE-based detection algorithm. The error floor caused by the CE-based detection is eliminated at high SNR regions.

The BER performance comparison between some existing detection methods such as ZF, MMSE, ZF-VBLAST and the PS-driven CE detection algorithm is shown in Fig. 6.2. Due to noise enhancement, the performance of ZF is poor in comparison to ML even with V-BLAST algorithm. MMSE offers a slight improvement compared with ZF.

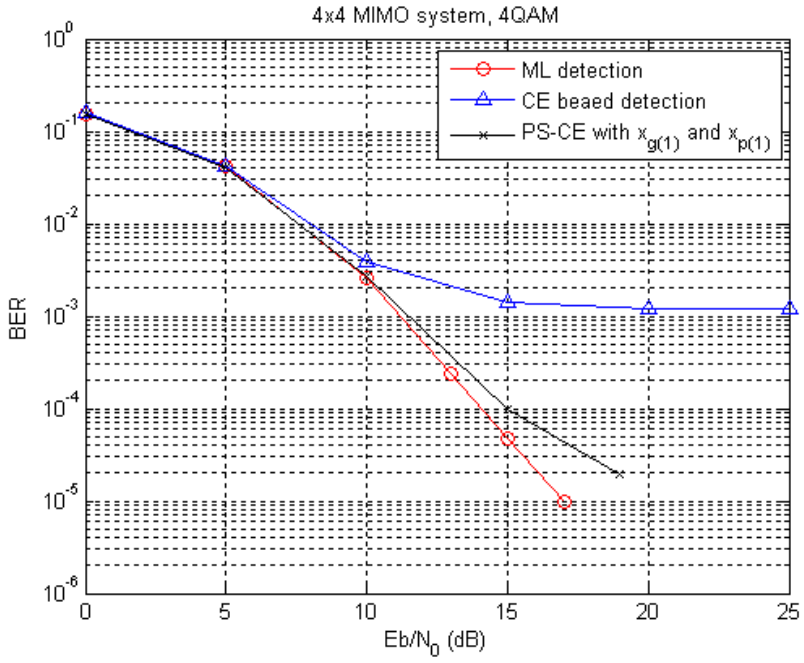


Figure 6.1: BER performance of the two proposed detectors.

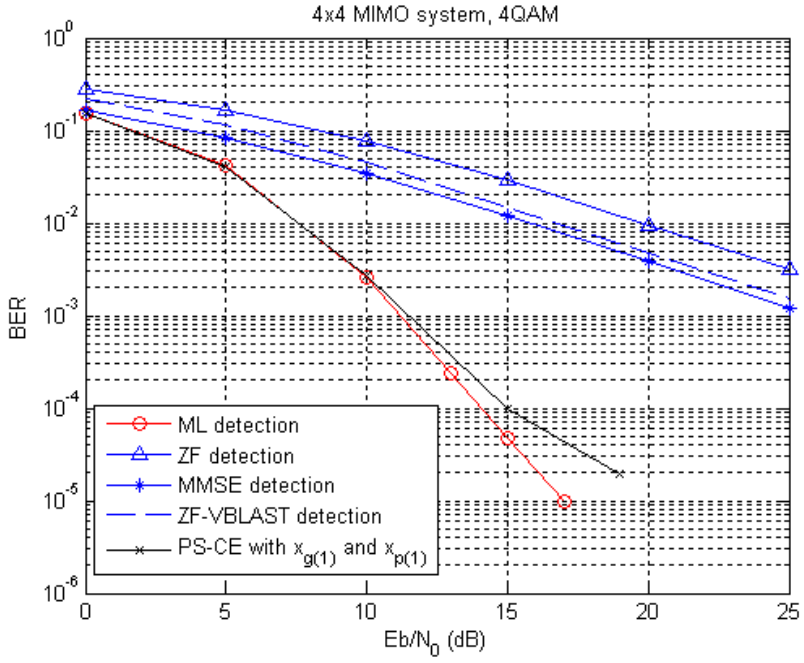


Figure 6.2: BER performance comparison of different detectors.

The proposed algorithm provides a near-ML performance and outperforms the MMSE detector by more than 12dB at BER=10^{-4}.

Next, we discuss some modifications for the PS-driven CE detection methods. BER performances are shown for the PS-driven CE detection method with different combinations in the updating formula, equation (5.13). Fig. 6.3 compares the difference between the one using $\mathbf{x}_{g(1)}$ only and the one using both $\mathbf{x}_{g(1)}$ and $\mathbf{x}_{p(1)}$. As shown in this figure, the one using both $\mathbf{x}_{g(1)}$ and $\mathbf{x}_{p(1)}$ reaches lower BER when SNR is larger than 10dB since $\mathbf{x}_{p(1)}$ provides another more reliable position at each iteration in addition to the best position among overall iterations.

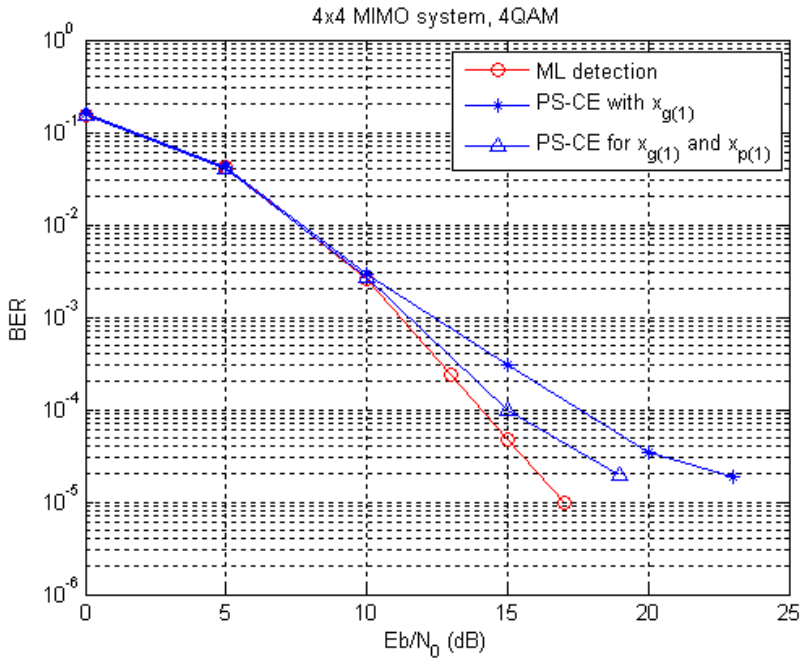


Figure 6.3: BER comparison of the PS-driven CE detector using(1) $\mathbf{x}_{g(1)}$ only; (2) both $\mathbf{x}_{g(1)}$ and $\mathbf{x}_{p(1)}$.

Although including the use of best sample vector in every iteration leads to an obvious improvement instead of using $\mathbf{x}_{g(1)}$ only, it is possible that the swarm is trapped in a local minimum since only the best sample vectors are considered. Replacing $\mathbf{x}_{p(1)}$ with $\mathbf{x}_{p(2)}$ or $\mathbf{x}_{p(3)}$, Fig. 6.4 plots BER performance comparison of these different detection

methods.

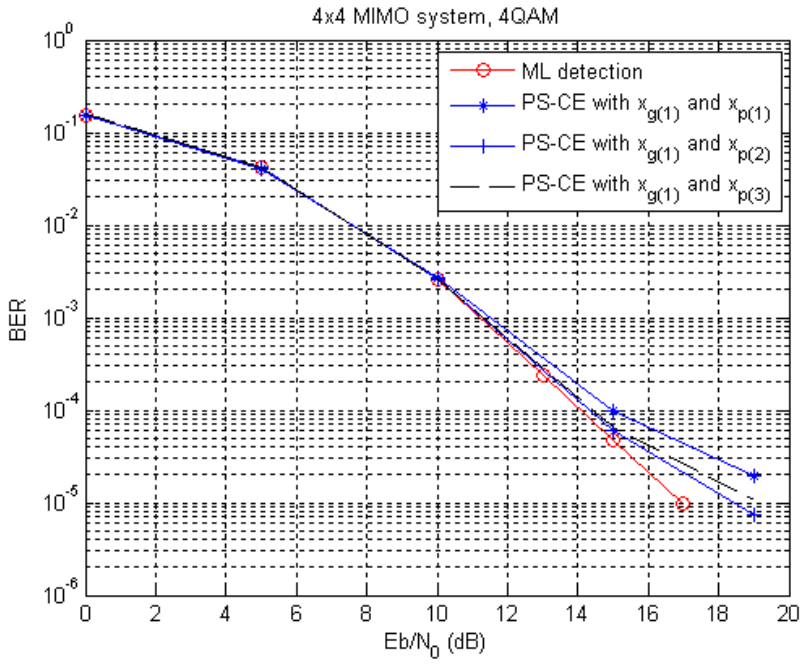


Figure 6.4: BER comparison of the PS-driven CE detector using $\mathbf{x}_{g(1)}$ and (1) $\mathbf{x}_{p(1)}$; (2) $\mathbf{x}_{p(2)}$; (3) $\mathbf{x}_{p(3)}$, respectively.

As expected, the one with $\mathbf{x}_{p(2)}$ performs best among these three scenarios. The case of $\mathbf{x}_{p(3)}$ provides a similar contribution with that of $\mathbf{x}_{p(2)}$. However, the third best sample vector may be too far away the optimum and can't afford good enough information. Fig. 6.5 shows results of adjustments for the weighting factor of $\mathbf{x}_{p(2)}$. The weighting for $f_{g(1)}^{(k)}(x_i)$ and $f_s^{(k)}(x_i)$ remain constant as 0.1, respectively. In Fig. 6.6, the combination of $\mathbf{x}_{g(1)}$, $\mathbf{x}_{g(2)}$ and $\mathbf{x}_{p(2)}$ is considered. It also indicates that the sample vectors used in each iteration plays an more important role for the algorithm rather than those among all iterations.

Figs. 6.7 and 6.8 verify the one with $\mathbf{x}_{p(2)}$ performs better than that with $\mathbf{x}_{p(1)}$. Fig. 6.8 plots the averaged minimum distance trajectory, i.e. $||\mathbf{y} - \mathbf{H}\mathbf{x}||^2$, at each iteration when SNR is 15 dB over 1,000,000 simulations and it exhibits that using $\mathbf{x}_{p(2)}$ exactly leads to a smaller distance.

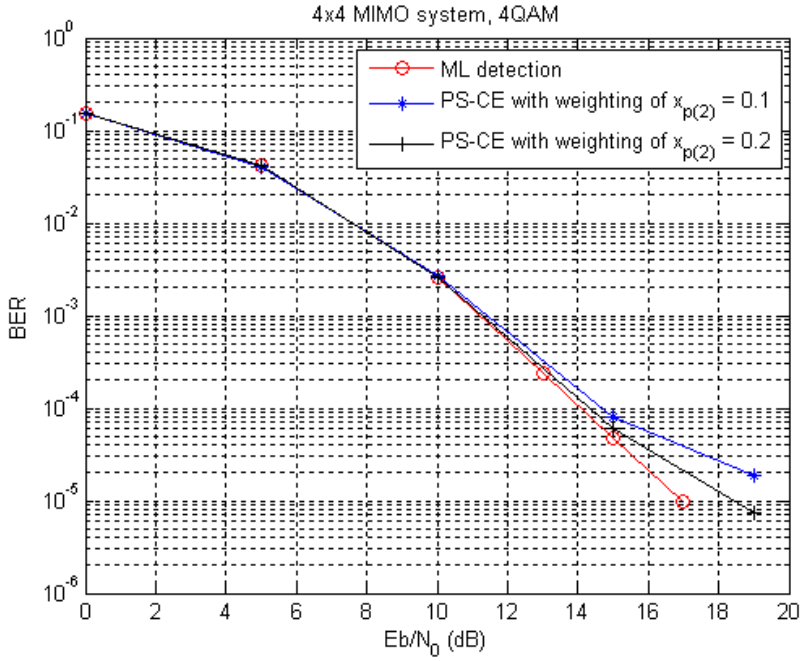


Figure 6.5: BER comparison of the PS-driven CE detector using $\mathbf{x}_{g(1)}$ and $\mathbf{x}_{p(2)}$ with different weighting of $\mathbf{x}_{p(2)}$.

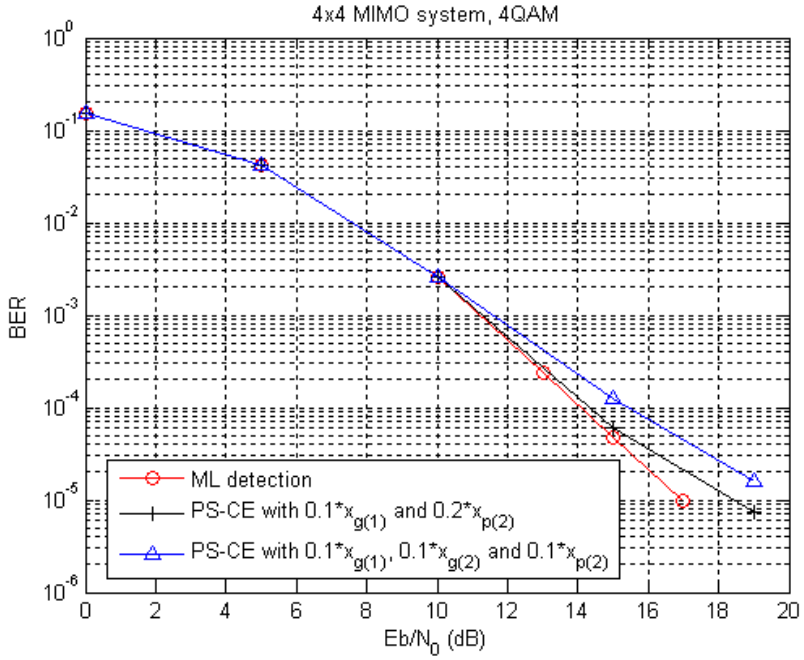


Figure 6.6: BER comparison of the PS-driven CE detector using (1)$\mathbf{x}_{g(1)}$ and $\mathbf{x}_{p(2)}$; (2)$\mathbf{x}_{g(1)}$, $\mathbf{x}_{g(2)}$ and $\mathbf{x}_{p(2)}$.

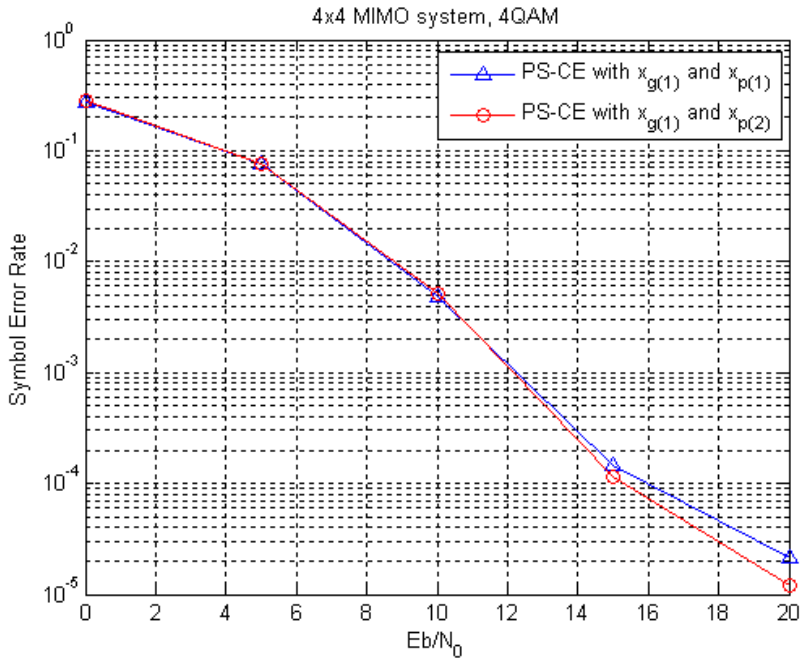


Figure 6.7: Symbol error rate performance of the PS-driven CE detector using $\mathbf{x}_{g(1)}$ and (1) $\mathbf{x}_{p(1)}$; (2) $\mathbf{x}_{p(2)}$, respectively.

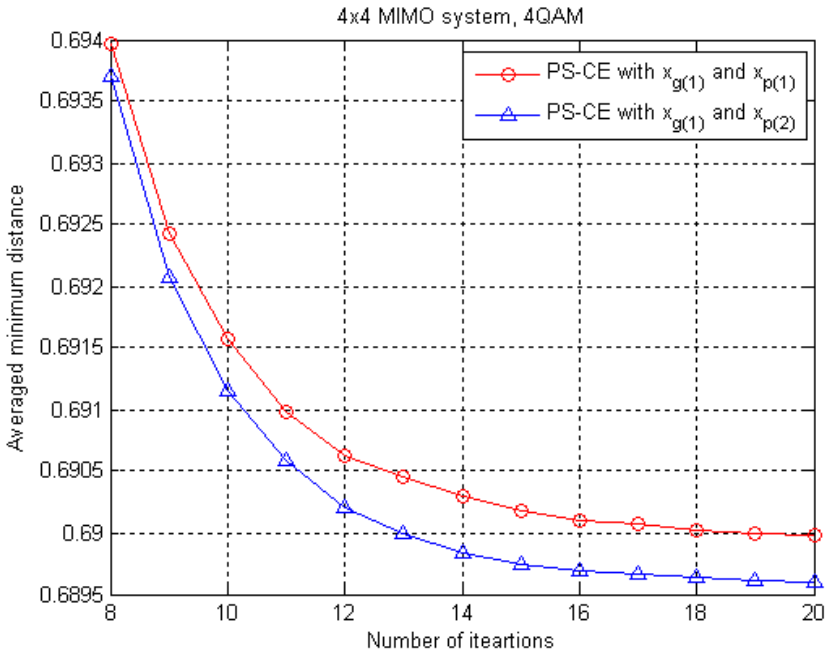


Figure 6.8: The averaged minimum distance at each iteration when SNR=15 dB.

Finally, we present the BER performance using the PS-driven CE detection method which uses the components of $\mathbf{x}_{g(1)}$ and $\mathbf{x}_{p(2)}$ in a 6×6 MIMO system.

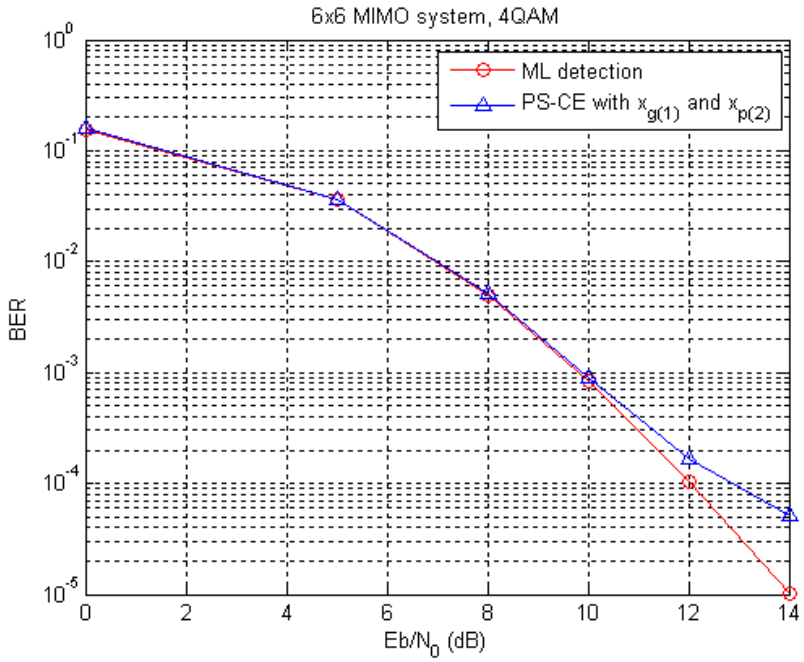


Figure 6.9: BER of a 6×6 MIMO system with 4-QAM under the PS-driven CE detection algorithm and ML detection.

6.2 Imperfect Channel Estimation

In this section, the channel estimation error is included. Orthogonal training sequences are generated from a perfect root-of-unity sequence (PRUS) [16] which can be constructed by the Frank-Zadoff-Chu-sequence [17] as

$$s(k) = \begin{cases} e^{j\pi Ck^2/N}, & \text{for } N \text{ is even} \\ e^{j\pi Ck(k+1)/N}, & \text{for } N \text{ is odd} \end{cases} \tag{6.2}$$

with $k = 0, \cdots, N-1$ where N denotes the length of sequence and C is a positive integer that is coprime to N. The updating formula for importance distributions is chosen to

be

$$f^{(k+1)}(x_i = a) = 0.1 f_s^{(k)}(x_i = a) + 0.1 f_{g(1)}^{(k)}(x_i = a) + 0.2 f_{p(2)}^{(k)}(x_i = a) + 0.6 f^{(k)}(x_i = a) \quad (6.3)$$

since it performs best according to the simulation results shown in the previous section. Fig. 6.10 shows the BER performance with $\sigma^2_{\Delta \mathbf{H}} = 0.03$ and $N_P = 8$ using the original score function and the modified one.

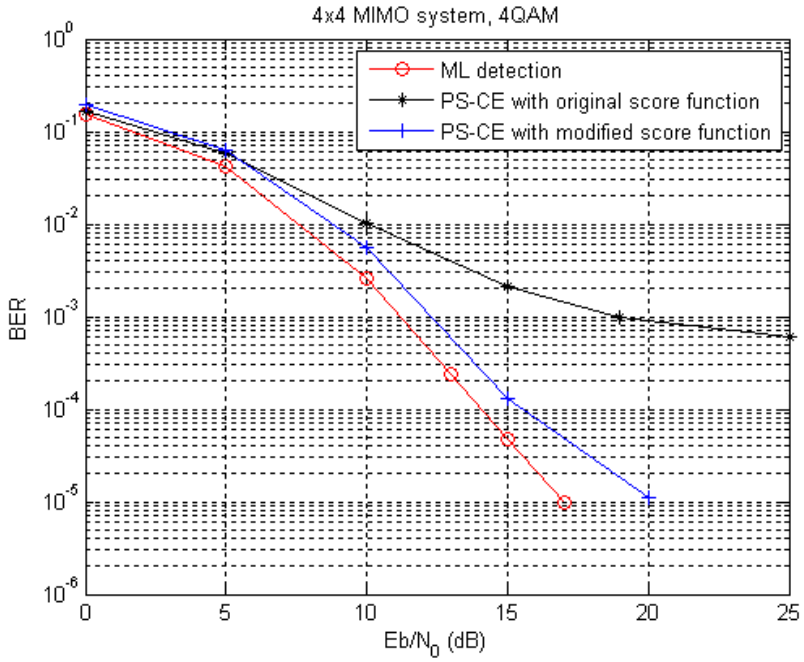


Figure 6.10: BER performance comparison of a 4×4 MIMO system with 4-QAM based on different score functions with $\sigma^2_{\Delta \mathbf{H}} = 0.03$.

6.3 Complexity Comparison

The computational complexity comparison of conventional detectors and the proposed detectors is listed in Table 6.1.

Although the complexity of the proposed detectors is higher than conventional detectors such as ZF, MMSE, ZF-VBLAST, the BER performance of the PSD-CE outperforms those detectors. Compared with the ML detector, the complexity is much lower

when N_T is large. Though the sample size U increases as the constellation size M increases, it has a very slight variation while N_T increases. Taking 4×4 and 6×6 MIMO systems for example, the complexity of the PSD-CE detector for these two cases is the same since U and N_{itr} remain constant in the two cases. However, the complexity of ML detector increase two orders when the number of antennas is raised from 4 to 6.

Detector	**Complexity**
ML	$(N_T N_R + N_R) \times M^{N_T}$
ZF	$4N_T^3 + 2N_T^2 N_R + N_T N_R$
MMSE	$4N_T^3 + 2N_T^2 N_R + N_T N_R + N_T$
ZF-VBLAST	$\sum_{i=0}^{N_T}(4i^3 + 2N_R i^2) + \sum_{i=0}^{N_T-1}[N_T(N_T - i) + 2N_T]$
CE	$[(N_T N_R + N_R) \times U + 3] \times N_{itr}$
PSD-CE	$[(N_T N_R + N_R) \times U + 5] \times N_{itr}$

Table 6.1: Complexity comparison between different detectors.

Chapter 7

Conclusion

The purpose of this thesis is to present alternative algorithms for detecting signals in a MIMO system. We first propose a detector which is based on the Cross-Entropy (CE) method. The CE-based MIMO detection method is a Monte-Carlo based approach to estimate the importance distribution of transmit signals by minimizing the distance between the provisional distribution derived at each iteration and the optimal importance distribution. Simulation result indicates that the BER performance of our detector is close to the ML detector when SNR is comparatively low. However, an error floor occurs at high SNR region due to nonuniform convergence of the CE approach.

To improve the CE-based MIMO detector and reduce the error floor, we include the ideas of the Particle Swarm Optimization (PSO) and propose the PS-driven CE detection algorithm. Updating the importance distributions by a mixture of densities determined by the elite set and the best sample vector among all iterations improves the detector performance. In addition, considering the best sample vector at each iteration in determining the updated importance distributions further improves the performance. Moreover, to extend the PS-driven CE MIMO detection algorithm, some modifications for the updating formula of importance distributions are also considered. As shown in the simulation results, PS aided CE approach can significantly eliminate the error floor, outperforming the conventional ZF or MMSE detector by more than 12dB at BER=10^{-4}.

To take into account the channel estimation errors, we propose a robust detector structure which is based a modified score function. The estimation errors are modelled as complex Gaussian distributed and the modified score function includes the variance of the channel estimation error as an extra term. Simulation results prove that the modified score function used by our detector helps to improve the BER performance when the practical imperfect channel estimation scenario is considered.

Bibliography

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[2] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj and H. V. Poor, *MIMO wireless communications*, Cambridge University Press, 2007.

[3] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", *Proc. of the IEEE Int. Joint Conf. on Neural Networks,* pp. 1942-1948, 1995.

[4] A. P. Engelbrecht, *Fundamentals of Computational Swarm Intelligence*, John Wiley & Sons, 2005.

[5] F. Heppner and U. Grenander, "A stochastic nonlinear model for coordinated bird flocks", *S. Krasner, editor, The Ubiquity of Chaos,* AAAS Publications, Washington, DC, 1990.

[6] J. N. Kapur and H. K. Kesavan, *Entropy Optimization Priciples with Applications*, Academic Press, New York, 1992.

[7] R. Y. Rubinstein, "Optimization of computer simulation models with rare events", *European Journal of Operational Research,* vol.99, pp. 89-112, Nov. 1997.

[8] R. Y. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization", *Methodology and Computing an Applied Probability,* vol.2, pp.127-190, 1999.

[9] R. Y. Rubinstein and D. P. Kroese, *The Cross-Entropy Method*, Springer, 2004.

[10] S. M. S. Sadough, M. A. Khalighi, and P. Duhamel, "Improved iterative MIMO signal detection accounting for channel estimation errors," IEEE Trans. on Vehicular Technology, accepted for future publication.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1983.

[12] P. W. Wolniansky, G. J. Foschini, G. D. Golden and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," *Proc. of IEEE Int. Symposium on Signals, Systems and Electronics (ISSSE'98)*, Pisa, Italy, Sep. 1998.

[13] A. A. Khan, M. Naeem and S. I. Shah, "A particle swarm algorithm for symbols detection in wideband spatial multiplexing systems," *Proc. of the Genetic and Evolutionary Computation Conf. (GECCO'2007),* London, U. K., pp. 63-70, Jul. 2007.

[14] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*, Morgan Kaufman publisher, 2001.

[15] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," *IEEE Int. Conf. on Systems, Man, and Cybernetics,* pp.4104-4108, 1997.

[16] O. Weikert, U. Zölzer, "Efficient MIMO channel estimation with optimal training sequences," *Proc. of 1st Workshop on Commercial MIMO-Components and-Systems (CMCS 2007),* Duisburg, Germany, Sep. 2007.

[17] D.C. CHu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. on Information Theory,* pp. 531-532, Jul. 1992.

作者簡歷

王惇琳，台南市人，1985 年生

國立交通大學電信工程學系系統組	2007.9~2009.8
國立交通大學電信工程學系	2003.9~2007.6
國立台南女子高級中學	2000.9~2003.6

Graduate Courses：

1. Coding Theory
2. Computer Communication Networks (I)
3. Detection and Estimation Theory
4. Digital Communications
5. Digital Signal Processing
6. Matrix Computation
7. Multimedia Communications
8. Stochastic Processes