CHALMERS

The block diagram shows a chain of elements: T, a, SPM, Loss (X_1), amplifier, $\oplus$ with N_1 (R_1), SPM, Loss (X_2), amplifier, $\oplus$ with N_2 (R_2), R.

Expectation Propagation-based Receiver for Coherent Optical Communication

Master of Science Thesis in the Master Degree Programme,

Communication Engineering

SHENZHOU HOU

Department of Signals and Systems
Division of Communication Systems and Information Theory
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2011
Master's Thesis 2011:1

Abstract

This thesis presents a novel maximum a posteriori receiver based on factor graphs and expectation propagation (EP) technologies for the coherent optical communication system. The receiver deals with the linear (AWGN) and nonlinear distortions in the desired system. Comparison has also been made on SER between the EP-based receiver and other methods. The discussions and analysis are given for the result in the end.

Keywords: optical, factor graphs, expectation propagation, linear, nonlinear

Acknowledgements

I would like to thank all those people who always offer me kind help and make my two years' master study an exciting and inspiring journey.

First of all, I want to thank my supervisor and examiner Henk Wymeersch. You always give me the confidence to conquer the unexpected problems during my master thesis. Without your qualified suggestion and powerful support, I cannot imagine that I could on day finish the thesis.

Nan and Yan, thanks a lot for your kind assistance on my thesis work. I have learned a lot from the discussion with you. There is no doubt that we have experienced a wonderful cooperation in the past few months. Meanwhile, I should also like to thank my classmates who gave me good advices during these two years in Chalmers.

Thanks should also be given to Lu. Thanks for your unconditional support. Without you, lots of fun will be undoubtedly missing. I will also thank all the members of Jiangyou party. I will never forget our precious experience for the rest of my life and hope to see your guys again.

Finally, I will give my deeply appreciation to my parents. It is my big fortune to be your child. This moment, words are no longer enough to express how I love you. Thank you very much for everything in these passed years. No matter where I am, my heart will be always with you.

<div align="right">Shenzhou Hou, Göteborg, Sweden, Sept 19, 2011</div>

Contents

Chapter 1

Introduction

1.1 Background

This section will give a brief introduction to coherent optical communication system and also the corresponding techniques used in the thesis.

1.1.1 Coherent optical communication system

In coherent optical communication system, the multi-level quadrature amplitude modulation (M-QAM) is employed because of its capability to provide a high data rate and spectral efficiency compared to traditional binary optical communication system. M-QAM often needs a high signal power to hold an acceptable signal-error-rate (SER) due to its more densely packed constellation. Nevertheless, the optical system gives a limitation on the performance while the input power increase, which is mostly due to the Kerr nonlinearity effect. The impairments from the Kerr effect contain deterministic and stochastic aspect, such as intrachannel four-wave mixing and the self-phase modulation (SPM) respectively [1, 2].

Additive white Gaussian noise (AWGN) which is often used to simulate the background noise of the channel can be also applied to the coherent optical communication system in the thesis.

1.1.2 Factor graphs and Expectation propagation

Factor graphs and expectation propagation are two fundamental techniques employed in this thesis. Factor graphs are a kind of graph models which can be used to transform the factorized functions into the form of graph models. The advantage of factor graphs is that it offers an efficient way to calculate the marginal distributions of the function [3].

Expectation Propagation (EP) is an algorithm which can be used together with factor graphs. EP can iteratively make approximations for the desired distributions.

1.2 Purpose

The purpose of this thesis is to design and develop a novel receiver which is based on factor graphs and expectation propagation for the coherent optical communication system with

16-QAM constellation. The receiver aims to solve the linear (AWGN) and nonlinear (SPM) distortions happened in the communication system. The simulations will be made in MATLAB and the results will be compared with other existing methods. Comments will be also provided in the end.

1.3 Structure of the thesis

This thesis is organized in a problem-solving structure. In chapter 2, details of the system and the corresponding problem will be introduced. A MAP receiver is also suggested in the end of chapter 2. Chapter 3 will explain the definition and details of factor graphs and show how to use factor graphs to build the MAP receiver mentioned in chapter 2. Moreover, the problem happened while using factor graphs will be also discussed in the end of chapter 3. In chapter 4, expectation propagation (EP) will be introduced to solve the problem in chapter 3. In the end of chapter 4, a general EP algorithm will be given for the optical communication system. Chapter 5 explains the details on how to implement the EP algorithm for linear and nonlinear system and a simplified EP algorithm will be given in the end of chapter 5. In chapter 6, comparison is made between the EP and other existing methods. The analysis on the results will be also provided. The last Chapter 7 talks about the conclusion of the thesis and the suggested work which could be carried out in the future.

Chapter 2

Coherent optical communication system

2.1 Introduction to the coherent optical communication model

As it's mentioned in the last chapter, because of the special property of the coherent optical communication system, the signal passing through the system will be distorted by a kind of nonlinear interference called self-phase modulation (SPM) [1] which is due to Kerr effect. And AWGN is also existed as the background noise of the system.

When considering long haul transmission in the optical fiber system, the system can be divided into a number of blocks called Spans. Each span contains a single mode fiber (SMF), a dispersion-compensating fiber, and an amplifier.

The thesis will focus on the interferences coming from the SPM and AWGN, which are also considered to represent the nonlinear and linear noise respectively in the system model. The figure below gives an example for a system with two spans.

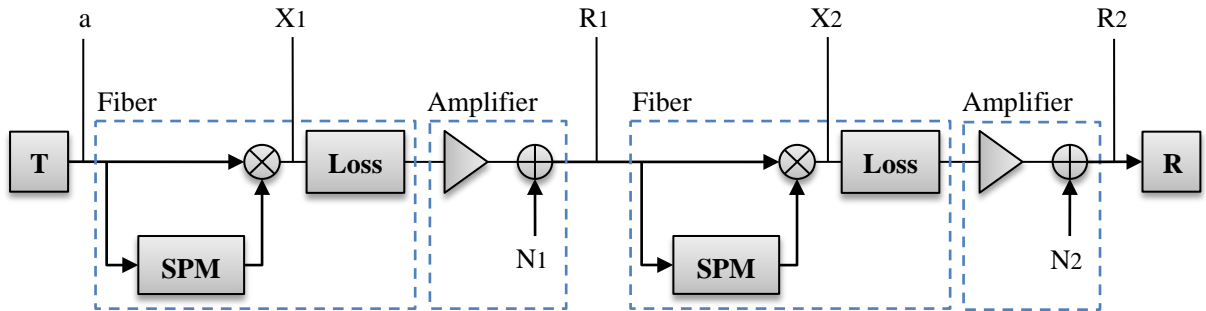Figure 2.1 System model for optical communication system with two spans

2.2 SPM, AWGN and Polarization

SPM: In figure 2.1, the function for SPM block is:

$$SPM(R_{i-1}) = exp(j * \gamma L_{eff} * ||R_{i-1}||^2) \tag{2.1}$$

Here γ is the nonlinearity parameter and L_{eff} is the effective lengthen factor of the single mode fiber (SMF) for each fiber. From the system model the signal after the SPM can be

obtained:

$$X_i = R_{i-1} * exp(j * \gamma L_{eff} * ||R_{i-1}||^2)$$
(2.2)

AWGN: In the system, AWGN is considered to be a two dimensional Gaussian noise [4]:

$$N \sim N(X; M, \Sigma)$$
(2.3)

Therefore, the relationship between the input and the output of a span is:

$$R_i = R_{i-1} * exp\left(j * \gamma L_{eff} * ||R_{i-1}||^2\right) + N_i$$
(2.4)

Polarization: The optical fiber communication discussed in the thesis is polarization multiplexed, which means that more than one set of signals (more than one ray of light) can be transmitted simultaneously in the system. Under this condition, the signal will be affected by all the signals transmitted at the same time in SPM block. Take dual polarization for an instance, the formula is as follows:

$$X^1_i = R^1_{i-1} * exp(j * \gamma L_{eff} * \left[||R^1_{i-1}||^2 + ||R^2_{i-1}||^2 \right])$$
(2.5)

$$X^2_i = R^2_{i-1} * exp(j * \gamma L_{eff} * \left[||R^1_{i-1}||^2 + ||R^2_{i-1}||^2 \right])$$
(2.6)

Here, X^1_i and X^2_i denote two simultaneously transmitted signals.

2.3 The MAP Receiver

According to the system model, it can be expected that, when passing through the optical system, the signals will be iteratively rotated and added Gaussian noise. The picture below gives an example of a distorted 16-QAM constellation after passing through the system with 17 spans and 5dbm average input power.
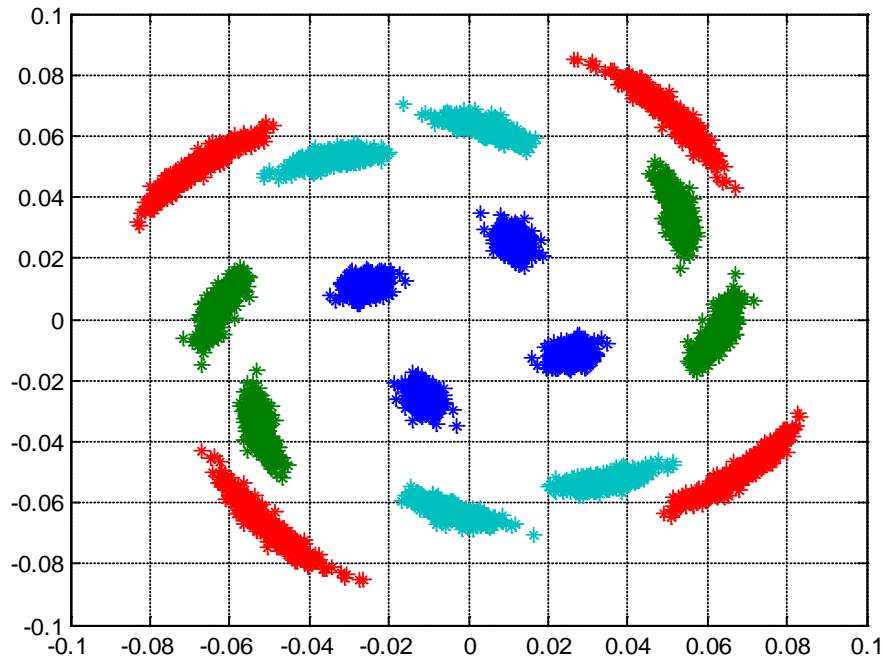
Figure 2.2 Distorted constellation of 16-QAM with 17 spans and average input power 5 dbm

To solve this problem and recover the distorted signals in the end, the thesis tries to design a MAP receiver:

$$\hat{a}(R_N) = \arg MAX\big(P(a|R_N)\big); \ a \in \Theta \qquad (2.7)$$

Here Θ is the constellation of the input signal, R_N is the observation and $\hat{a}$ is the decision made based on R_N.

Hence, the computation of the posterior $P(a|R_N)$ turns out to be the main task. The thesis introduces two technologies --- factor graphs and expectation propagation to deal with this problem in next chapters.

Chapter 3

Factor graphs

3.1 Introduction to Factor graphs

Factor graphs are a particular kind of graph model which can be used to represent the factorized functions in the form of graphs. Moreover, factor graphs enable an efficient way to compute the marginal distributions of the function with the algorithm called --- sum-product algorithm [3].

This section will introduce the basic theory of the factor graphs and sum-product algorithm with examples.

3.1.1 Marginals, Function factorizations, Factor graphs

Marginal: For function: $f(X_1, X_2,,, X_{n-1}, X_n)$, the marginal of variable x_n can be calculated by summing out all other variables:

$$B_{x_n}(X_n) = \sum_{\{\backslash X_n\}} f(X_1, X_2,,, X_{n-1}, X_n);\qquad(3.1)$$

Here $B_{x_n}(X_n)$ is the marginal of X_n and $\{\backslash X_n\}$ means the set of all possible values of all variables expect X_n.

Function factorization: Assume the function $f(X_1, X_2,,, X_{n-1}, X_n)$ can be factorized as a product of K sub-functions (factors), therefore:

$$f(X_1, X_2,,, X_{n-1}, X_n) = \prod_{K=1}^{K} f_k(V_k);\qquad(3.2)$$

Here V_k is the set of variables and $V_k \subseteq \{X_1, X_2,,, X_{n-1}, X_n\}$. One thing should be mentioned is each function may have more than one factorization form.

Factor graphs: Factor graphs provide a one-to-one mapping from a factorized function. In factor graphs, neighborhood is an important concept which indicates that with a factorized function:

➢ The neighborhood of the factor f_k: $N(f_k)$, means a set of variables which appear in f_k.

➢ The neighborhood of the variable X_n: $N(X_n)$, means a set of factors which include X_n.

For an instance, a function can be factorized as:

$$f(X_1, X_2, X_3, X_4, X_5) = f_1(X_1, X_2) * f_2(X_2, X_3) * f_3(X_1, X_4, X_5) \qquad (3.3)$$

In this case, the neighborhood of the variable X_1: $N(X_1) = \{f_1, f_3\}$, while the neighborhood of the factor f_2: $N(f_2) = \{X_2, X_3\}$.

When mapping a given factorized function such as formula (3.2) into the corresponding factor graph, the process can be carried out as follows:

➢ Draw a vertex for each variable X_n, which is called variable node.
➢ Draw a vertex for each function f_k, which is called function node.
➢ Draw an edge between the function node f_k and every variable node $X_n \subseteq N(f_k)$.

The following example will show the process of mapping the factorized function into factor graph:

Assume the function can be factorized into 4 factors:

$$f(X_1, X_2, X_3, X_4, X_5, X_6) = f_1(X_1, X_2, X_3) * f_2(X_1, X_4) * f_3(X_1, X_5, X_6) * f_4(X_2); \qquad (3.4)$$
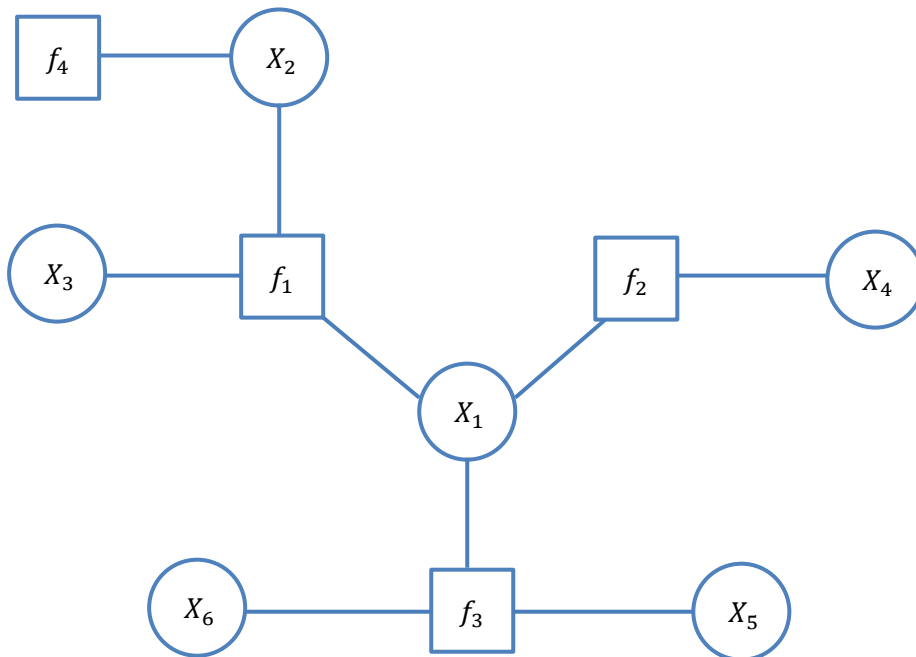
Then the corresponding factor graph is:

Figure 3.1 the factor graph of the function 3.4

The factor graph obviously shows that the neighborhood of X_1: $N(X_1) = \{f_1, f_2, f_3\}$ and the neighborhood of f_3: $N(f_3) = \{X_1, X_5, X_6\}$.

3.1.2 Sum-product algorithm in Factor graphs

The sum-product algorithm introduces a concept named 'Messages' which are passing over the edges between the variable nodes and function nodes in factor graphs and can be used to compute the marginals of the variables. In factor graphs the messages are functions of the related variable.

➢ The message from the function node f_k to the variable node $X_n \subseteq N(f_k)$ is denoted as $M_{f_k \to X_n}(X_n)$;
➢ The message from the variable node X_n to the function node $f_k \subseteq N(X_n)$ is denoted as $M_{X_n \to f_k}(X_n)$;

The messages can be calculated as follows:

$$M_{f_k \to X_n}(X_n) = \sum_{\backslash \{X_n\}} f_k \left(\{X_m = x_m\}_{X_m \subseteq N(f_k)} \right) * \prod_{X_m \subseteq N(f_k) \backslash \{X_n\}} M_{X_m \to f_k}(X_m)$$

(3.5)

$$M_{X_n \to f_k}(X_n) = \prod_{f_r \subseteq N(X_n) \backslash \{f_k\}} M_{f_r \to X_n}(X_n)$$
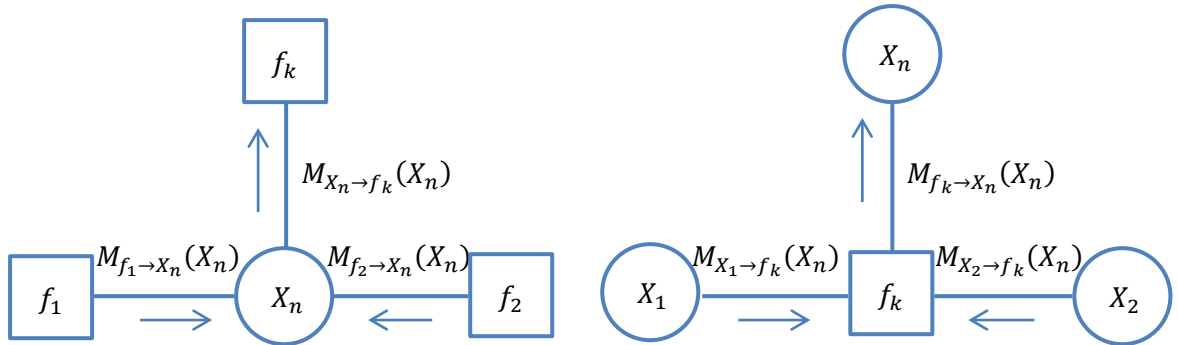
(3.6)

Figure 3.2 Messages passing from variable to function node and function to variable node

In factor graphs the marginal of a particular variable $B_{x_n}(X_n)$ can be obtained by the multiplication of the in-coming and out-going messages between the variable node X_n and the function node $f_k \subseteq N(X_n)$:

$$B_{x_n}(X_n) = M_{f_k \to X_n}(X_n) * M_{X_n \to f_k}(X_n)$$

(3.7)

Here $B_{x_n}(X_n)$ can also be called the belief of X_n.

Finally, a general sum-product algorithm can be:

➢ **Initialization:**
Each leaf function node f_k send the message to its neighborhood variable node X_n with $M_{f_k \to X_n}(X_n) = f_k(X_n)$. While each leaf variable node X_n send the message to its neighborhood function node f_k with $M_{X_n \to f_k}(X_n) = 1$. Notice that leaf node only has one neighborhood node.

➢ **Compute the message:**
Compute all the messages passing through the graph by using the formula (3.5) and (3.6).

➢ **Compute the Marginal:**
Finally, the Marginal of the variable X_n can be obtained by formula (3.7).

Take figure 3.1 for an example. The algorithm can be processed as follows:

Initialization:
Assume the message coming from the leaf function nodes:

$$M_{f4 \to X2}(X_2) = f_4(X_2)$$

Assume the message coming from the leaf variable nodes:

$$M_{X3 \to f1}(X_3) = 1$$
$$M_{X4 \to f2}(X_4) = 1$$
$$M_{X5 \to f3}(X_5) = 1$$
$$M_{X6 \to f3}(X_6) = 1$$

Compute all the messages:
From the figure 3.1, following messages can be obtained:

$$M_{X2 \to f1}(X_2) = M_{f4 \to X2}(X_2)$$

$$\boldsymbol{M_{f_1 \to X_1}(X_1)} = \sum_{\backslash\{X_1\}} f_1(X_1, X_2, X_3) * M_{X_2 \to f_1}(X_2) * M_{X_3 \to f_1}(X_3)$$

$$M_{f_2 \to X_1}(X_1) = \sum_{\backslash\{X_1\}} f_2(X_1, X_4) * M_{X_4 \to f_2}(X_4)$$

$$M_{f_3 \to X_1}(X_1) = \sum_{\backslash\{X_1\}} f_3(X_1, X_5, X_6) * M_{X_5 \to f_3}(X_5) * M_{X_6 \to f_3}(X_6)$$

$$M_{X1 \to f1}(X_1) = M_{f3 \to X1}(X_1) * M_{f2 \to X1}(X_1)$$

Compute the marginal:

Hence, the marginal of variable X_1:

$$B_{X1}(X_1) = M_{f1 \to X1}(X_1) * M_{X1 \to f1}(X_1)$$

3.2 Factor graphs for the optical communication system

This section will apply the factor graphs and sum-product algorithm to the optical communication model discussed in the last chapter. And some problems happened in the message computation will be mentioned in the end.

3.2.1 Factorization and Graphs

According to the last section, the factor graphs can be applied to the optical communication system described in figure 2.1. The distribution $P(a, X_1, R_1, X_2|R_2)$ can be factorized:

$$P(a, X_1, R_1, X_2|R_2) = Z_p * P(a) * P(X_1|a) * P(R_1|X_1) * P(X_2|R_1) * P(R_2|X_2) \quad (3.8)$$

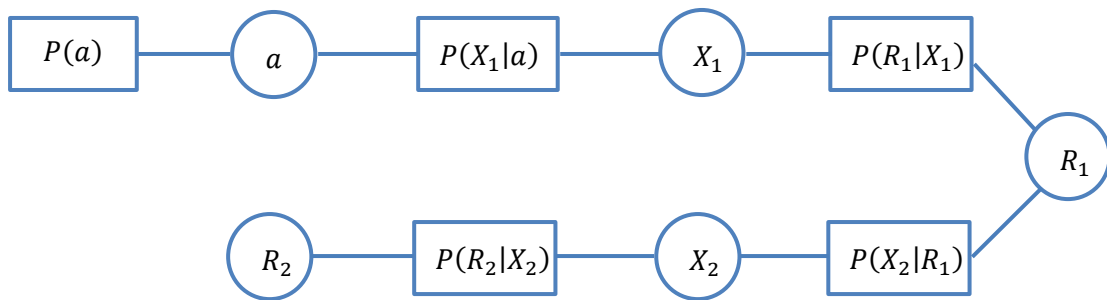Here Z_p is a constant. The factor graphs can be draw as follows:

Figure 3.3 Factor graphs for optical communication system

Here R_2 is the observation and a denotes the input signal. The task is to get the marginal of variable a based on the observation R_2. One thing should be mentioned is formula (3.8) and the factor graphs can be easily extended according to the number of spans in the system.

3.2.2 Sum-product algorithm

As discussed in section 3.2, the messages passing over the graph in figure 3.3 can be computed. For simplification, only one span is considered this time. Hence the factor graph is:
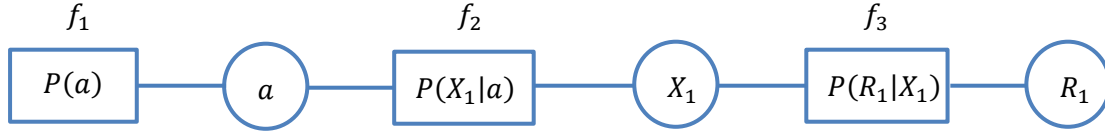
Figure 3.4 Factor graphs for optical communication system with one span

According to formula (3.5) and (3.6), the message computation can be simplified to:

➢ For variable nodes, the outgoing message is equal to the incoming message.
➢ For function nodes, the outgoing message is an integral of the product of the function and the incoming message.

For instance, in figure 3.3:

$$M_{f2 \to X1}(X_1) = M_{X1 \to f3}(X_1) \tag{3.9}$$

$$M_{f2 \to a}(a) = \int f_2(a, X_1) * M_{X1 \to f2}(X_1) dx_1 \tag{3.10}$$

After all the messages have been obtained, the marginal of a can be calculated as

$$P(a|R_1) = M_{f1 \to a}(a) * M_{a \to f1}(a) \tag{3.11}$$

3.2.3 Challenges

As is known in chapter 2, the optical fiber will introduce SPM and AWGN when signals are passing through. For factor graph in figure 3.3, the function node f_2 and f_1 represents the nonlinear effect and Gaussian noise separately:

$$f_2(a, X_1) = \delta\left(X_1 - a * exp\left(j * \gamma L_{eff} * ||a||^2\right)\right) \tag{3.12}$$

$$f_1(X_1, R_1) = C * exp\left(-\frac{1}{N_0}||R_1 - X_1||^2\right) \tag{3.13}$$

Hence, with formula (3.10) the message $M_{f2 \to a}(a)$ can be computed as:

$$M_{f2 \to a}(a) = \int f_2(a, X_1) * M_{X1 \to f2}(X_1) dx_1$$

$$= \int \delta(X_1 - a * exp(j * \gamma L_{eff} * ||a||^2)) * M_{X1 \to f2}(X_1) dx_1$$

$$= M_{X1 \to f2}(a * exp(j * \gamma L_{eff} * ||a||^2))$$

What's more, because of the Gaussian function node f_1 the message $M_{X1 \to f2}(X_1)$ also belongs to Gaussian family:

$$M_{X1 \to f2}(X_1) = C * exp\left(-\frac{1}{N_0} ||X_1 - R_1||^2\right)$$

Here the observation R_1 can be seen as a constant. Finally the message $M_{f2 \to a}(a)$ is:

$$M_{f2 \to a}(a) = C * exp\left(-\frac{1}{N_0} \left|\left|a * exp(j * \gamma L_{eff} * ||a||^2) - R_1\right|\right|^2\right) \qquad (3.14)$$

It is obviously that the function of $M_{f2 \to a}(a)$ is complicated and one can imagine that after a number of spans the messages will be too hard to compute and the final result of the marginal will be too complicated to use. That is why Expectation Propagation is introduced in the next chapter.

Chapter 4

Expectation Propagation

4.1 Introduction to Expectation Propagation

Expectation Propagation (EP) is an algorithm which can iteratively make approximations for the desired distributions. This section will discuss the basic theory of EP [6, 12] and give an instance to show the general process of EP [7].

4.1.1 Why Expectation Propagation

In the end of the last chapter a problem of computing messages has been thrown out. In a more general model, assume X is continuous unknown sample which need to be decided, $A = \{Y_1, Y_2, \ldots, Y_{n-1}, Y_n\}$ denote observations of the system and all belong to i.i.d..

According to Bayes' rule:

$$p(X|A) \propto p(X) \prod_i p(Y_i|X) \tag{4.1}$$

Here $p(X|A)$ is already normalized. Assume $p(X)$ and $p(Y_i|X)$ are prior and likelihood function respectively. It can be seen that it will be unpractical to directly multiply a number of likelihoods. For instance, assume $p(Y_i|X)$ is the mixture of two different distributions: $p(Y_i|X) = G_1 + G_2$, hence the complexity of the calculation will dramatically increase along with the growing of the number of observations. Actually, 500 observations need 2^{500} times of multiplication of two distributions. Therefore, the result formula for $p(X|A)$ (normalized) is hard to access by directly computation. One reasonable method to deal with this problem is to make an approximation $h(X)$ which could be close to the true $p(X|A)$:

$$p(X|A) \approx h(X) \tag{4.2}$$

This is the main strategy of EP for solving this kind of problem as in the last chapter.

4.1.2 Exponential family and Kullback-Leibler divergence

When making approximations, one may consider which kinds of the distributions can be selected to make the approximation and how to do it. These concern two basic concepts of EP: Exponential family and Kullback-Leibler divergence.

Exponential family: Exponential family is often used to make approximations in EP. The exponential family has the form:

$$r(\boldsymbol{x}|\theta) = h(\boldsymbol{x})\exp\left(\sum_{i=1}^{s}\eta_i(\theta)T_i(\boldsymbol{x}) - A(\theta)\right) \tag{4.3}$$

Here θ is called the parameter of the family. Function $\eta_i(\theta)$ presents the weight of the potential function $T_i(\boldsymbol{x})$. Function $A(\theta)$ is a normalizing function which ensures that the value of the integral or summation of $r(\boldsymbol{x}|\theta)$ results to be one.

Exponential family is often used because it contains many famous members such as Gaussian, Poisson, Bernoulli, etc., which is suitable to be applied to many different kind of models. What is more, the use of exponential family can simplify the computation when making the approximation. Because of the property of the exponential family, the multiplication or division of two distributions from the same exponential family results a distribution belonging to the same family with a normalizing factor. An example of the multiplication and division of two Gaussian distribution is given below [5]:

$$N(x;\,m_1,\sigma_1) * N(x;\,m_2,\sigma_2) = N(m_1;\,m_2,\sigma_1+\sigma_2) * N(x;\,m_3,\sigma_3) \tag{4.4}$$

$$\sigma_3 = (\sigma_1{}^{-1} + \sigma_2{}^{-1})^{-1}$$
$$m_3 = \sigma_3\sigma_1{}^{-1}m_1 + \sigma_3\sigma_2{}^{-1}m_2$$

$$\frac{N(x;\,m_1,\sigma_1)}{N(x;\,m_2,\sigma_2)} = \frac{\sigma_2}{(\sigma_2-\sigma_1)N(m_1;\,m_2,\sigma_2-\sigma_1)} * N(x;\,m_3,\sigma_3) \tag{4.5}$$

$$\sigma_3 = (\sigma_1{}^{-1} - \sigma_2{}^{-1})^{-1}$$
$$m_3 = \sigma_3\sigma_1{}^{-1}m_1 - \sigma_3\sigma_2{}^{-1}m_2$$

The formula above also can be extended to two-dimensional Gaussian distribution.

Kullback-Leibler (KL) divergence: When approximating the distribution to a certain exponential family, the Kullback-Leibler (KL) divergence is often used. The KL divergence can measure the difference between the existing distribution and the desired exponential family.

A general form of KL divergence is:

$$KL\big(p(X)||p'(X)\big) = \int p(X) * \log\left(\frac{p(X)}{p'(X)}\right)dx \tag{4.6}$$

Here $p'(X)$ is the approximation of $p(X)$.

In order to make the approximation, one should try to minimize the KL divergence:

$$\theta' = arg_\theta \ min \ KL\big(p(X)||p'(X)\big) \tag{4.7}$$

Moment matching is considered to be an optimal way to minimize the KL divergence:

$$\int p(X) * T(\boldsymbol{x}) \, dx = \int p'(X) * T(\boldsymbol{x}) \, dx \tag{4.8}$$

The formula can also be rewritten as:

$$\mathbb{E}_{\boldsymbol{P}}[T(\boldsymbol{x})] = \mathbb{E}_{\boldsymbol{p'}}[T(\boldsymbol{x})] \tag{4.9}$$

Hence, it can be seen that minimizing the KL divergence is equal to mapping the desired moments of the existing distribution to a certain exponential family. Take Gaussian distribution for an example. The Minimization of the KL divergence is just computing the mean and covariance of the existing distribution and using the results to be the mean and covariance for the Gaussian distribution.

When talking about KL-divergence, another related concept --- Projection is often used, which can denote the process of moment matching:

$$p'(x) = \boldsymbol{Projection}[p(x)] \tag{4.10}$$

Here $p'(x)$ is the approximated distribution with the same moments mapped from $p(x)$.

4.1.3 Expectation Propagation

It has been said that EP makes approximations 'iteratively'. Why 'iteratively'?

Rewrite formula (4.1):

$$p(X|\boldsymbol{A}) \propto p(X) \prod_i p(Y_i|X) = \prod_i r_i(X) \tag{4.11}$$

Here $r_0(X) = p(X)$, $r_1(X) = p(Y_1|X)$,,,, $r_i(X) = p(Y_i|X)$. An approximation $h(X)$ needed to be made for $p(X|\boldsymbol{A})$. As it already known in the last section, a projection can be made to get the approximation:

$$h(X) = \boldsymbol{Projection}[p(X|\boldsymbol{A})] \tag{4.12}$$

Here $h(X)$ is the normalized version of $H(X)$. Because $p(X|\boldsymbol{A})$ consists of a bunch of

distributions $r_i(X)$, the moment matching is not easy to carry out directly. EP introduces a method to get the final approximation $h(X)$ iteratively.

Assume approximation $H(X)$ also consists of a bunch of sub-approximations $\tilde{r}_i(X)$:

$$H(X) = \prod_i \tilde{r}_i(X) \tag{4.13}$$

$$h(X) = \frac{\prod_i \tilde{r}_i(X)}{\int \prod_i \tilde{r}_i(X)\, dx} \tag{4.14}$$

Each $\tilde{r}_i(X)$ is an approximated distribution for $r_i(X)$. Hence, after all $\tilde{r}_i(X)$ have been obtained, the final approximation can be calculated by formula (4.13).

EP algorithm firstly initializes all $\tilde{r}_i(X)$ to an exponential family. In each iteration, one can choose any of $\tilde{r}_i(X)$ to make a new approximation and then refine that $\tilde{r}_i(X)$. When approximating $\tilde{r}_i(X)$, EP makes a projection on the multiplication of the true $r_i(X)$ and the term $h^{\backslash i}(X)$ which denotes all sub-approximations except $\tilde{r}_i(X)$:

$$h^{\backslash i}(X) \propto \frac{h(X)}{\tilde{r}_i(X)} \tag{4.15}$$

EP will continuously process the iteration until all $\tilde{r}_i(X)$ converge.

Hence, a typical EP process can be:

1. Initialization of the sub-approximation term $\tilde{r}_i(X)$:

2. Calculate the posterior of x:

$$h(X) = \frac{\prod_i \tilde{r}_i(X)}{\int \prod_i \tilde{r}_i(X)\, dx}$$

3. Do iterations until all $\tilde{r}_i(X)$ converge:

   ➢ Pick up any $\tilde{r}_i(X)$ for updating
   ➢ Divide $h(X)$ by $\tilde{r}_i(X)$ and normalize to get the 'lack' posterior $h^{\backslash i}(X)$:

$$h^{\backslash i}(X) \propto \frac{h(X)}{\tilde{r}_i(X)}$$

   ➢ Make a projection on the Multiplication of $h^{\backslash i}(X)$ and $r_i(X)$, which will produce the new mean, covariance and a normalizing factor K_i for the new $h(X)$:

$$h(X) = \textbf{\textit{Projection}}[h^{\backslash i}(X) * r_i(X)] \tag{4.16}$$

➢ Update $\tilde{r}_i(X)$ with the new $h(X)$:

$$\tilde{r}_i(X) = \frac{K_i * h(X)}{h^{\backslash i}(X)} \tag{4.17}$$

4. Compute the final approximation for $p(X)$:

$$p(X) \approx h(X) \propto H(X) = \prod_i \tilde{r}_i(X) \tag{4.18}$$

4.1.4 Potential problem

One problem may happen when making the projection: The moments of the desired distribution may not be easy to calculate mathematically. Then other technique such as Monte Carlo sampling needed be introduced to assist the approximation. This will be discussed later in the thesis.

4.2 An example of Expectation Propagation: clutter problem

This section presents the process of Expectation Propagation with factor graphs by solving a classic problem – clutter problem [7].

4.2.1 The Problem

A bunch of signals are transmitted through an added Gaussian noise channel which can cause the clutter problem. Assume that the distribution of the observed signals on the condition of the sent signals is a Gaussian distribution mixed with another Gaussian distribution caused by the clutter:

$$p(Y|X) = \alpha * N(Y; X, \sigma_1 * \mathbf{I_d}) + (1 - \alpha) * N(Y; 0, \sigma_2 * \mathbf{I_d}) \tag{4.19}$$

$$N(Y; M, \Sigma) = 2\pi|\Sigma|^{-\frac{1}{2}} * exp\left(-\frac{1}{2}(Y - M)\,\Sigma^{-1}\,(Y - M)^T\right) \tag{4.20}$$

In formula (4.19) Y and X are observed and sent signal respectively. The first part of the formula is useful for later research while the second part presents the influence of the clutter. Finally, parameter α shows the ratio between the first and second part in the finally distribution. Moreover, $p(Y|X)$ also can be considered into Gaussian family because it is a

mixture of two Gaussian distributions.

Suppose a set of independent signals $\mathbf{A} = \{Y_1, Y_2, \,,\,,\,,\, Y_{n-1}, Y_n\}$ has been observed in the end. Hence, the joint distribution of observation A and X is:

$$p(\mathbf{A}, X) \propto p(X) \prod_i p(Y_i|X) = \prod_i r_i(X) \qquad (4.21)$$

Here $r_0(X) = p(X)$, $r_1(X) = p(Y_1|X)\,,\,,,\, r_i(X) = p(Y_i|X)$. Finally, the task becomes to compute the probability density function $p(\mathbf{A}, X)$ while $r_i(X)$ is known.

4.2.2 Factor Graph

According to Chapter 3, formula (4.21) can be represented into factor graph below:
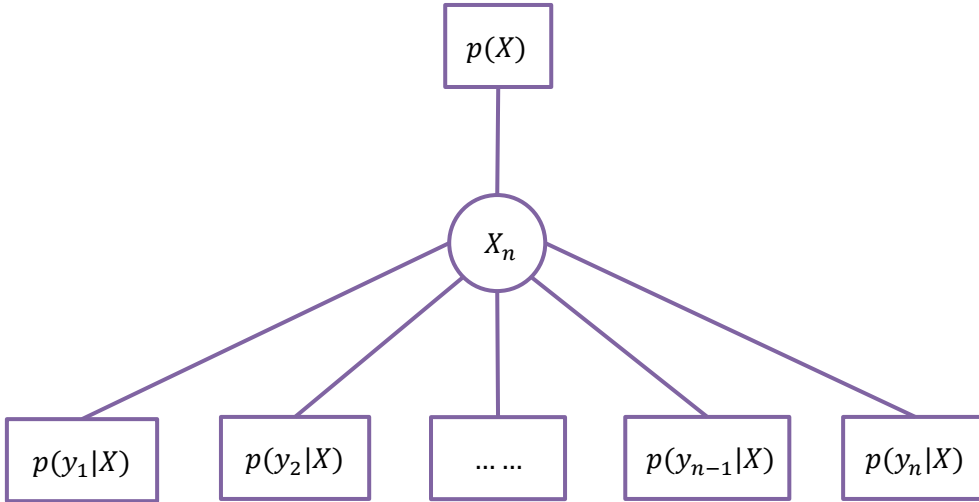
Figure 4.1 Factor Graph of clutter problem

4.2.3 Processing Expectation Propagation

From the last section, the following formula can be obtained:

$$\prod_i r_i(X) = p(X) \prod_i p(Y_i|X) \propto p(\mathbf{A}, X) \approx h(X) \propto H(X) = \prod_i \tilde{r}_i(X) \quad (4.22)$$

It can be seen that $r_0 = p(X) = \tilde{r}_0$ and $r_i(X)$, $i = \{1,2,3,\,,\,,n-1, n\}$ consists of Gaussian distributions. Therefore the approximation component $\tilde{r}_i(X)$ can also be approximated to be Gaussian-like:

$$\tilde{r}_i(X) = C_i * exp\left(-\frac{1}{2}(X - M_i)\,\Sigma_i^{-1}\,(X - M_i)^T\right) \qquad (4.23)$$

What is more, because the result of the multiplication of Gaussian distributions is still an Gaussian distribution (not normalized). Hence, $p(A, X)$(normalized) can be approximated to be a Gaussian distribution. The formulas for Gaussian multiplication and division can be found in formula (4.4) and (4.5).

According to section 4.2.3, the EP can process as follows:

1. Initialization of the mean and covariance matrix of the approximation term $\tilde{r}_i(X)$.

➢ Assume $M_0 = \mathbf{0}, \Sigma_0 = 10\mathbf{I_d}, C_0 = 2\pi|\Sigma_0|^{-\frac{1}{2}}$ for the $\tilde{r}_0(X)$ term.

➢ Because the rest of the approximation term are unknown at the beginning, the mean and covariance matrix are initialized to be $M_i = \mathbf{0}, \Sigma_i = 1000\mathbf{I_d} \to \infty, C_i = 1$ to make the probability of $\tilde{r}_i(X)$ be approximately identical in the data space at first.

2. According to formula (4.14) and (4.23):

$$h(X) = N(X; \boldsymbol{m_x}, \sigma_x * \boldsymbol{I_d}), \boldsymbol{m_x} = \boldsymbol{m_0}, \sigma_x = \sigma_0.$$

3. Doing loop from $\tilde{r}_1, \dots, \tilde{r}_n$ until all $\tilde{r}_i$ converge:

➢ According to formula (4.15) and (4.5), Dividing $h(X)$ by $\tilde{r}_i(X)$ and normalizing to get the mean and covariance of the 'lack' posterior $h^{\backslash i}(X)$:

$$\sigma_x^{\backslash i} = (\sigma_x^{-1} - \sigma_i^{-1})^{-1}$$

$$m_x^{\backslash i} = \sigma_x^{\backslash i}\sigma_x^{-1}m_x - \sigma_x^{\backslash i}\sigma_i^{-1}m_i$$

➢ Multiply $h^{\backslash i}(X)$ and $r_i(X)$ which equals to formula (4.23). Make a projection and get the new mean, covariance and a normalizing factor K_i for the new $h(X)$:

$$h(X) = \boldsymbol{Projection}[h^{\backslash i}(X) * r_i(X)]$$

$$K_i = (1 - \alpha) * N(Y_i; 0, \sigma_2 * \boldsymbol{I_d}) + \alpha * N(Y_i; m_x^{\backslash i}, (\sigma_x^{\backslash i} + \sigma_1)\boldsymbol{I_d})$$

$$m_x = m_x^{\backslash i} + \frac{\alpha * \sigma_x^{\backslash i} * \left(Y_i - m_x^{\backslash i}\right) * N(Y_i; m_x^{\backslash i}, (\sigma_x^{\backslash i} + \sigma_1)\boldsymbol{I_d})}{K_i * (\sigma_x^{\backslash i} + 1)}$$

$$\sigma_x = \sigma_x^{\backslash i} - \frac{\alpha * (\sigma_x^{\backslash i})^2 * N(Y_i; m_x^{\backslash i}, (\sigma_x^{\backslash i} + \sigma_1)\boldsymbol{I_d})}{K_i * \left(\sigma_x^{\backslash i} + 1\right)} * [1$$

$$- \frac{(1 - \alpha) * N(Y_i; 0, \sigma_2 * \boldsymbol{I_d}) * \sigma_x^{\backslash i} * |Y_i - m_x^{\backslash i}|^2}{K_i * \left(\sigma_x^{\backslash i} + 1\right)}]$$

➢ According to formula (4.5) and (4.17), update $\tilde{r}_i(X) = K_i * h(X)/h^{\backslash i}(X)$:

$$\sigma_i = \left(\sigma_x^{-1} - (\sigma_x^{\backslash i})^{-1}\right)^{-1}$$

$$m_i = \sigma_i \sigma_x^{-1} m_x - \sigma_i (\sigma_x^{\backslash i})^{-1} m_x^{\backslash i}$$

$$C_i = \frac{(2\pi)^{-1}|\sigma_i * \boldsymbol{I_d}|^{-\frac{1}{2}} * \sigma_x^{\backslash i}}{(\sigma_x^{\backslash i} - \sigma_x)N\left(m_x; m_x^{\backslash i}, \sigma_x^{\backslash i} - \sigma_x\right)} * K_i$$

➢ Compute the approximated distribution --- $p(\boldsymbol{A}, X)$:

$$p(\boldsymbol{A}, X) \propto \prod_i \tilde{r}_i(X)$$

4.3 Expectation Propagation for the optical communication system

In the last chapter, the factor graphic model and message computation method has been introduced to deal with the optical communication model. This section will firstly extend EP algorithm to the message computing in factor graphs, then develop an method with EP for the message computing in chain-like model such as optical communication system.

4.3.1 Expectation Propagation for message computing

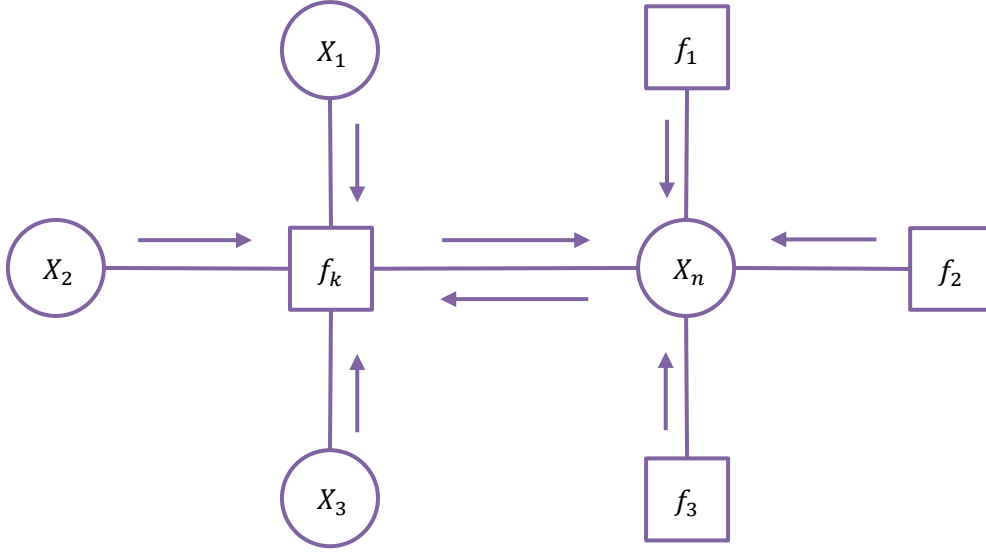An example will be given below to show how to deal with the message computing with EP.

Figure 4.2 Message computing with EP

According to the sum-product algorithm, the belief of X_n is:

$$B_{x_n}(X_n) = M_{f_k \to X_n}(X_n) * M_{X_n \to f_k}(X_n)$$

$$= \prod_{f_r \subseteq N(X_n)} M_{f_r \to X_n}(X_n) \tag{4.24}$$

$$b_{x_n}(X_n) \propto \prod_{f_r \subseteq N(X_n)} M_{f_r \to X_n}(X_n) \tag{4.25}$$

Here $b_{x_n}(X_n)$ is normalized from $B_{x_n}(X_n)$. Now the task is to compute $b_{x_n}(X_n)$. It can be seen that $b_{x_n}(X_n)$ is the multiplication of a number of distributions, as it's discussed before, make computation or approximation on $b_{x_n}(X_n)$ directly is not reasonable. EP algorithm can do approximation iteratively.

Assume:

$$\prod_{f_r \subseteq N(X_n)} \widetilde{M}_{f_r \to X_n}(X_n) \propto h(X) \approx b_{x_n}(X_n) \propto \prod_{f_r \subseteq N(X_n)} M_{f_r \to X_n}(X_n) \tag{4.26}$$

Here, $\widetilde{M}_{f_r \to X_n}(X_n)$ is the sub-approximation for each message.

As in EP algorithm, do iteration until all messages converged. In each iteration, choose one message to refine by EP algorithm as discussed in section 4.1.3:

For example, move $\widetilde{M}_{f_k \to X_n}(X_n)$ in figure 4.2:

$$h^{\backslash f_k}(X) \propto \frac{h(X)}{\widetilde{M}_{f_k \to X_n}(X_n)} \tag{4.27}$$

$$h(X) = \boldsymbol{Projection}\left[h^{\backslash f_k}(X) * M_{f_k \to X_n}(X_n)\right] \tag{4.28}$$

$$\widetilde{M}_{f_k \to X_n}(X_n) = \frac{K_i * h(X)}{h^{\backslash f_k}(X)} \tag{4.29}$$

After all messages converged, then $b_{x_n}(X_n)$ can be obtained:

$$b_{x_n}(X_n) \approx h(X) \propto \prod_{f_r \subseteq N(X_n)} \widetilde{M}_{f_r \to X_n}(X_n) \tag{4.30}$$

Obviously, the example in figure 4.1 in section 4.2 can also be translated into message computing where approximation and computation are made for $b_{x_n}(X_n)$.

4.3.2 Message computing for chain-like model with Expectation Propagation

In last section, one thing should be noticed that, according to the sum-product algorithm, the message: $M_{f_k \to X_n}(X_n)$ in formula (4.28) is equal to:

$$M_{f_k \to X_n}(X_n) = \sum_{\backslash \{X_n\}} f_k \left(\{X_m = x_m\}_{X_m \subseteq N(f_k)}\right) * \prod_{X_m \subseteq N(f_k) \backslash \{X_n\}} M_{X_m \to f_k}(X_m)$$

$$= \sum_{\backslash \{X_n\}} f_k \left(X_1, X_2, X_3, X_n\right) * M_{X_1 \to f_k}(X_1) * M_{X_2 \to f_k}(X_2) * M_{X_3 \to f_k}(X_3)$$

$$\tag{4.31}$$

From formula (4.31), it can be seen that in a chain-like model, more messages ($M_{X_1 \to f_k}(X_1), M_{X_2 \to f_k}(X_2), M_{X_3 \to f_k}(X_3)$ in this case) need to be computed or approximated before the message: $M_{f_k \to X_n}(X_n)$ can be used to make the approximation. Hence, in a chain-like model [8] (optical communication) the EP algorithm will be modified from the one discussed in beginning of this chapter.

Take the optical communication model with one span for an instance.
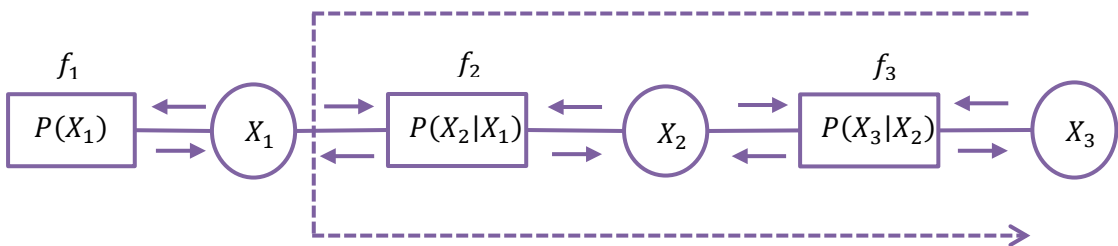

Figure 4.3 Factor graph for optical communication model with one span
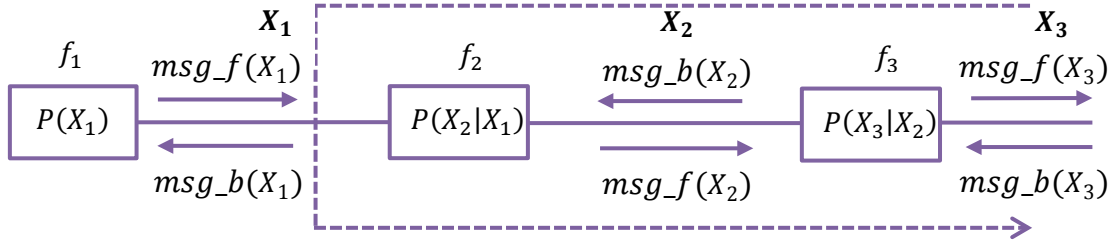
Figure 4.3 can be simplified as:

Figure 4.4 Simplified factor graph for optical communication model with one span

In figure 4.4, the direction from f_1 to f_3 is define as 'forward' and message passing in this direction is written as: $msg\_f(X_n)$. While the opposite direction is 'backward' and the message is: $msg\_b(X_n)$.

In figure 4.4, the target is to get the belief of X_1:

$$B_{X_1}(X_1) = msg\_f(X_1) * msg\_b(X_1) \tag{4.32}$$

Take figure 4.4 for an instance, the EP process can be carried out as follows:

1. Initialize the forward and backward messages passing through the factor graph.

2. Do a loop until all the messages converged:

➢ Make approximations for the messages of the last function block in the end of the model.

➢ Move to the backward function block and obtain the corresponding messages block by block until finish the message approximation for the second function block of the model (block f_2).

➢ From block f_2, move to the forward function block and approximate the messages block by block until reach the last function block in the end of the model.

3. Calculate the belief of X_1:

$$B_{X_1}(X_1) = msg\_f(X_1) * msg\_b(X_1) \tag{4.33}$$

This process can also be seen in Figure 4.4. What's more, according to the algorithm discussed above, it can be seen that the two outgoing messages computed in each function block can update the backward and forward incoming message of its left and right function

block respectively in the model. Hence, the messages passing through the model can be refined block by block in iterations.

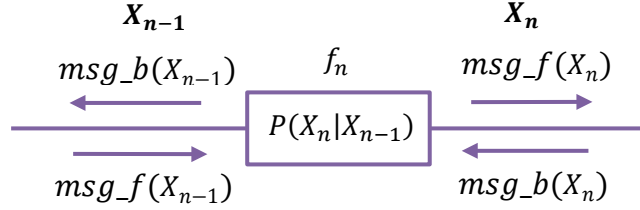The method for message computation and approximation for each function block is described below:

Figure 4.5 Message computation and approximation for one function block

According to figure 4.5, the task is to get the outgoing messages $msg\_b(X_{n-1})$ and $msg\_b(X_{n-1})$, while the incoming messages $msg\_f(X_{n-1})$ and $msg\_b(X_n)$ are given.

Take the outgoing message $msg\_f(X_n)$ for an example:

According to the sum-production algorithm discussed in section 3.2.2, the belief of X_n can be calculated as:

$$B_{X_n}(X_n) = \int P(X_n|X_{n-1}) * msg\_f(X_{n-1}) * msg\_b(X_n) \, dx_{n-1} \qquad (4.34)$$

A projection can be made on $B_{X_n}(X_n)$, which results an approximated distribution $\tilde{b}(X_n)$ and a normalizing factor K_n.

$$\tilde{b}(X_n) = \boldsymbol{Projection}\big[B_{X_n}(X_n)\big] \qquad (4.35)$$

The outgoing message of X_n can be computed as:

$$\widehat{msg}\_f(X_n) = \frac{K_n * \tilde{b}(X_n)}{msg\_b(X_n)} \qquad (4.36)$$

The outgoing message $msg\_b(X_{n-1})$ can be obtained in the same way. One thing should be mentioned that the algorithm above is possible to be simplified in some particular case, which will be discussed in the next chapter.

Chapter 5

Implementation of Expectation Propagation

5.1 Optical communication model with linear block

In this section, the optical communication system model involving linear (AWGN) function block will be discussed. The details of the message approximation for linear block will be talked about. The linear system model and the corresponding factor graph for two spans are as follows:
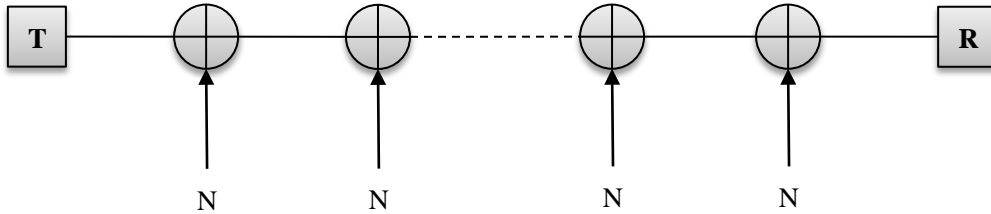

Figure 5.1 System model with only linear block
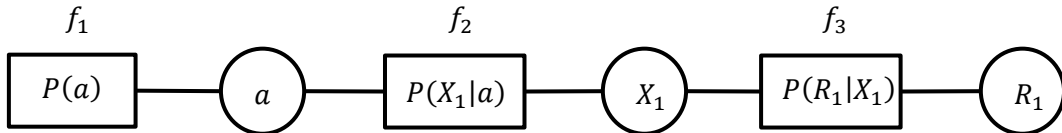
Figure 5.2 Factor graph for model with only linear block

5.1.1 System setup

As discussed in chapter 4, the initialization should be done before the main process will be carried out. For this system, choose two dimensional Gaussian distribution from the exponential family as the approximation term for the messages:

$$\tilde{r}_i(X) = C_i * exp\left(-\frac{1}{2}(X - M_i)\, \Sigma_i^{-1}\,(X - M_i)^T\right) \tag{5.1}$$

Obviously, the message passing through factor graph is the mean and covariance matrix of the Gaussian distribution.

Initial all the forward incoming messages of each function block to have:

$$Mean: M = [0,0], Covariance: \Sigma = \begin{bmatrix} \infty & 0 \\ 0 & \infty \end{bmatrix}$$

Considering the implementation, make ∞ to be 100. What's more, the back incoming message from the observation R to the last function block in the end of the model can be seen as a Gaussian distribution with:

$$Mean: M = [X_R, Y_R], Covariance: \Sigma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

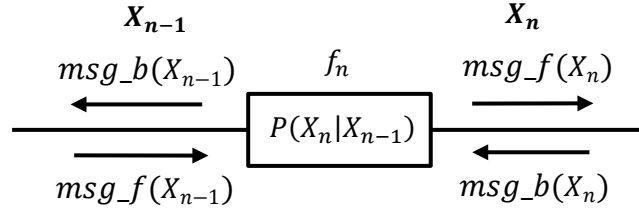### 5.1.2 Message approximation for linear block

Figure 5.3 Message computation and approximation for linear function block

According to figure 5.3, the task is to get the outgoing message: $msg\_b(X_{n-1})$ and $msg\_f(X_n)$ by using the incoming message: $msg\_f(X_{n-1})$ and $msg\_b(X_n)$. And the function f_n is:

$$f_n(X_n, X_{n-1}) = K * exp\left(-\frac{1}{N_0}||X_n - X_{n-1}||^2\right) \tag{5.2}$$

$msg\_f(X_n)$: For the reason that $msg\_f(X_{n-1})$ has:

$$Mean: M = [0,0], Covariance: \Sigma = \begin{bmatrix} \infty & 0 \\ 0 & \infty \end{bmatrix}$$

This makes the Gaussian distribution actually to be a uniform distribution. Consider that the function f_n is Gaussian-like, therefore the message should be still a uniform distribution after passed the function. Hence, in this case:

$$msg\_f(X_n) = msg\_f(X_{n-1}) \tag{5.3}$$

$msg\_b(X_{n-1})$: Refer to the formula (4.34):

$$B_{X_{n-1}}(X_{n-1}) = \int P(X_n|X_{n-1}) * msg\_f(X_{n-1}) * msg\_b(X_n) \, dx_n$$

Because the components involved in the formula above are all Gaussian-like, hence the message can be computed directly without any approximation.

$$msg\_b(X_{n-1}) = \int P(X_n|X_{n-1}) * msg\_b(X_n) \, dx_n \tag{5.4}$$

According to formula (5.1) and (5.2)

$$msg\_b(X_{n-1}) = \int N(X_n; X_{n-1}, N_0) * N(X_n; M_n, \Sigma_n) \, dx_n \tag{5.5}$$

With formula (4.4), one can obtain:

$$msg\_b(X_{n-1}) = \int N(X_{n-1}; M_n, N_0 + \Sigma_n) * N(X_n; M, \Sigma) \, dx_n \tag{5.6}$$

$$= N(X_{n-1}; M_n, N_0 + \Sigma_n)$$

5.2 Optical communication model with linear and nonlinear block

This section will introduce the nonlinear block (SPM) into the system discussed above. Monte Carlo method will be discussed to solve the message approximation for the nonlinear block. Finally, solutions will be given respectively for the optical communication system with single and dual polarization. The system model and factor graph for one span are as given below:
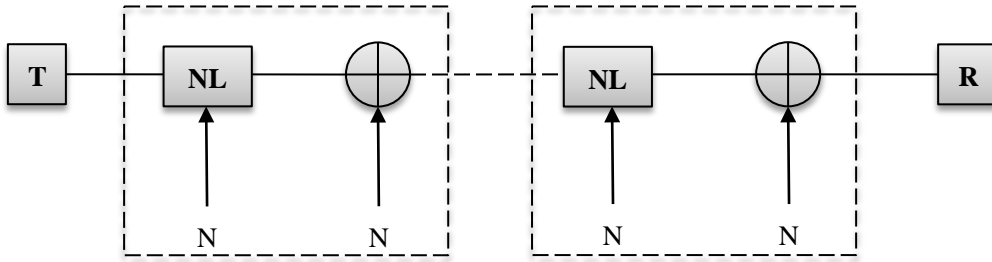

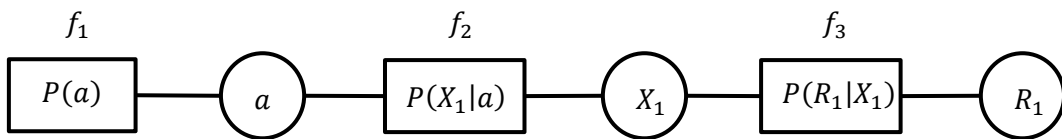Figure 5.4 System model with nonlinear and linear block

Figure 5.5 Factor graph for model with linear and nonlinear block

In figure 5.5, f_2 and f_3 represent the nonlinear and linear block respectively in one span.

5.2.1 System setup

The initialization for the system with nonlinear block is the same as in section 5.1.1.

5.2.2 Message approximation for nonlinear system with single polarization

Consider the picture in figure 5.3 again, the function block f_n now is:

$$f_n(X_n, X_{n-1}) = \delta\big(X_n - X_{n-1} * exp(j * \gamma L_{eff} * ||X_{n-1}||^2)\big) \tag{5.7}$$

$\boldsymbol{msg\_f(X_n)}$**:** As discussed in section 5.1.2, the incoming message: $msg\_f(X_{n-1})$ is actually a uniform distribution. Meanwhile the effect of function f_n is a rotation for the incoming message according to message computation method discussed in chapter 3. Hence, formula (5.3) can be still applied in this case.

$\boldsymbol{msg\_b(X_{n-1})}$**:** Refer to the formula (4.34):

$$\begin{cases} B_{X_{n-1}}(X_{n-1}) = \int P(X_n|X_{n-1}) * msg\_f(X_{n-1}) * msg\_b(X_n)\, dx_n \\[2mm] P(X_n|X_{n-1}) = \delta\big(X_n - X_{n-1} * exp(j * \gamma L_{eff} * ||X_{n-1}||^2)\big) \end{cases}$$

$$B_{X_{n-1}}(X_{n-1}) = msg\_f(X_{n-1}) * msg\_b\left(X_{n-1} * \exp\left(j * \gamma L_{eff} * \left|\left|X_{n-1}\right|\right|^2\right)\right) \tag{5.8}$$

According to the EP algorithm discussed in section 4.2.2, $B_{X_{n-1}}(X_{n-1})$ should be projected into a Gaussian distribution:

$$N(X_{n-1}; M_{n-1}, \Sigma_{n-1}) \sim \tilde{b}(X_{n-1}) = \boldsymbol{Projection}\big[B_{X_{n-1}}(X_{n-1})\big] \tag{5.9}$$

In formula (5.10), the mean: M_{n-1} and covariance: Σ_{n-1} matrix of $B_{X_{n-1}}(X_{n-1})$ should be calculated for the approximation term: $\tilde{b}(X_{n-1})$. In this case, the mean and covariance matrix is not easy to figure out mathematically, hence, the importance sampling and Monte Carlo integration [9, 10] are introduced to deal with the problem.

Assume:

$$B_X(X) = msg\_f(X) * msg\_b\left(X * \exp\left(j * \gamma L_{eff} * \left|\left|X\right|\right|^2\right)\right)$$

$$= N_1(X) * N_2\left(X * exp\left(j * \gamma L_{eff} * ||X||^2\right)\right)$$

$$= N_1(X) * Q(X) \tag{5.10}$$

1. Draw L samples from $Q(X)$: $\left\{x^{(1)},,,,,, x^{(l)}\right\}$

   ➢ Draw L samples from $N_2(X)$: $\left\{y^{(1)},,,,,, y^{(l)}\right\}$

   ➢ Rotate the samples: $\quad x^{(l)} = y^{(l)} * exp\left(-j * \gamma L_{eff} * \left\|y^{(l)}\right\|^2\right) \tag{5.11}$

2. Compute weights for samples: $\left\{x^{(1)},,,,,, x^{(l)}\right\}$:

$$w^{(l)} = N_1\left(x^{(l)}\right) \tag{5.12}$$

3. Normalize the weights:

$$v^{(l)} = \frac{w^{(l)}}{\sum_L w^{(l)}} \tag{5.13}$$

The set: $\left\{x^{(l)}, v^{(l)}\right\}_{l=1}^L$ is called the representation of $B_X(X)$.

4. Calculate the mean and covariance by Monte Carlo integration:

$$\begin{cases} \widetilde{m} = \sum_{l=1}^L v^{(l)} * x^{(l)} \\ \widetilde{v} = \sum_{l=1}^L v^{(l)} * \left(x^{(l)} - \widetilde{m}\right) * \left(x^{(l)} - \widetilde{m}\right)^T \end{cases} \tag{5.14}$$

After the mean and covariance matrix have been approximated the message: $msg\_b(X_{n-1})$ can be made as in formula (4.36).

5.2.3 Message approximation for nonlinear system with dual polarization

In the system with dual polarization, two signals are transmitted simultaneously. The rotation phenomena caused by the nonlinear block for each signal is not only dependent on the signal itself but also on the other one. The formulas are given below:

$$X^1{}_n = X^1{}_{n-1} * exp\left(j * \gamma L_{eff} * \left[||X^1{}_{n-1}||^2 + ||X^2{}_{n-1}||^2\right]\right) \tag{5.15}$$

$$X^2{}_n = X^2{}_{n-1} * exp\left(j * \gamma L_{eff} * \left[||X^1{}_{n-1}||^2 + ||X^2{}_{n-1}||^2\right]\right) \tag{5.16}$$
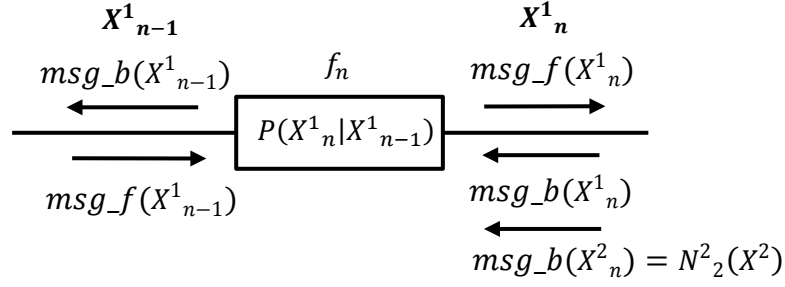
Figure 5.6 Message computation and approximation for nonlinear function block, dual polarization

According to figure 5.6, it is obvious that the analysis method and the result of the message: $\boldsymbol{msg\_f(X^1_n)}$ is the same as in the last section. The only difference is in the message: $\boldsymbol{msg\_b(X^1_{n-1})}$ when processing the importance sampling and Monte Carlo integration in step 1:

1. Draw L samples from $Q(X): \{\boldsymbol{x^{(1)}},,,,,,\boldsymbol{x^{(l)}}\}$

    ➢ Draw L samples from $N_2(X): \{\boldsymbol{y^1}^{(1)},,,,,,\boldsymbol{y^1}^{(l)}\}$

    ➢ Draw L samples from $N^2_2(X^2): \{\boldsymbol{y^2}^{(1)},,,,,,\boldsymbol{y^2}^{(l)}\}$

    ➢ Rotate the samples:

$$\boldsymbol{x^1}^{(l)} = \boldsymbol{y^1}^{(l)} * exp\left(-j * \gamma L_{eff} * \left[\left\|\boldsymbol{y^1}^{(l)}\right\|^2 + \left\|\boldsymbol{y^2}^{(l)}\right\|^2\right]\right) \qquad (5.17)$$

The following steps are the same as discussed in the last section. And the message: $\boldsymbol{msg\_f(X^2_n)}$, $\boldsymbol{msg\_b(X^2_{n-1})}$ can be obtained in the same way above.

5.3 Simplification on EP algorithm

From the discussion on initialization and message computation and approximation method above, it can be easily computed that all the forward messages passing through the system have:

$$Mean: M = [0,0], Covariance: \Sigma = \begin{bmatrix} \infty & 0 \\ 0 & \infty \end{bmatrix}$$

Therefore, the messages (mean and covariance matrix) will be converged in only one way from the end to the beginning of the model without iteration. Hence, the EP algorithm discussed in section 4.3 is simplified:

4. Initialize the forward and backward messages.

5. Make approximations for the messages of the last function block in the end of the model. Move to the backward function block and obtain the corresponding messages block by block until finish the message approximation for the second function block of the model (block f_2).

6. Calculate the belief of X_n:

$$B_{X_n}(X_n) = msg\_f(X_n) * msg\_b(X_n) \tag{5.18}$$

Note that the algorithm can be easily extended to dual polarization system.

Chapter 6

Data analysis and comparison

6.1 Introduction to ML, SBP, BP method

In the following sections of this chapter, the results from the EP implementation will be analyzed together with some other methods such as ML (Maximum Likelihood), SBP (Stochastic Back Propagation) [11] and BP (Back Propagation) [11]. This section will give a short introduction of these technologies. Moreover, some simulation parameters will be also mentioned in the end.

6.1.1 ML, SBP and BP

ML: ML (Maximum Likelihood) is a common used method for the receiver design. It simply compares the 'distance' between the observation r_N and every possible original signal $\mathbf{s}$, finds the smallest 'distance' and makes decision:

$$\hat{\mathbf{s}} = arg\,min\,\|r_N - \mathbf{s}\|^2,\ \mathbf{s} \in \Omega^2 \tag{6.1}$$

SBP: SBP (Stochastic Back Propagation) [11] is a new method developed for solving the problem in optical communication system. It is based on the factor graphs and particle technology which draws a number of samples (particles) for each observation in the end. Then all the particles are propagated backward to the beginning of the model and try to compensate for the distortions when passing through the system model. Finally, Monte Carlo integration is applied. And the result is compared with all possible original signals in order to make a decision.

BP: BP (Back Propagation) is similar to SBP, the main difference is that BP does not consider the AWGN in the model and does not use particles while SBP dose.

$$\hat{\mathbf{s}} = arg\,min\,\left\|r_N - \mathbf{s} * exp(j * \gamma L_{eff} * N * \|\mathbf{s}\|^2)\right\|^2,\ \mathbf{s} \in \Omega^2 \tag{6.2}$$

6.1.2 Simulation parameters

During the simulation, the system has: 16-QAM constellation, 22 spans, $N_0 = 4.89 * 10^{-7}\,W/Hz$ for AWGN block, $\gamma = 1.25 W^{-1} km^{-1}$ and $L_{eff} = 17.36\,km$ for the nonlinear block, $Num = 100$ for importance sampling.

6.2 System with linear block

The figure below presents the result SER along with the input power P_{in} in dBm for the linear system equipped with EP and ML receiver.
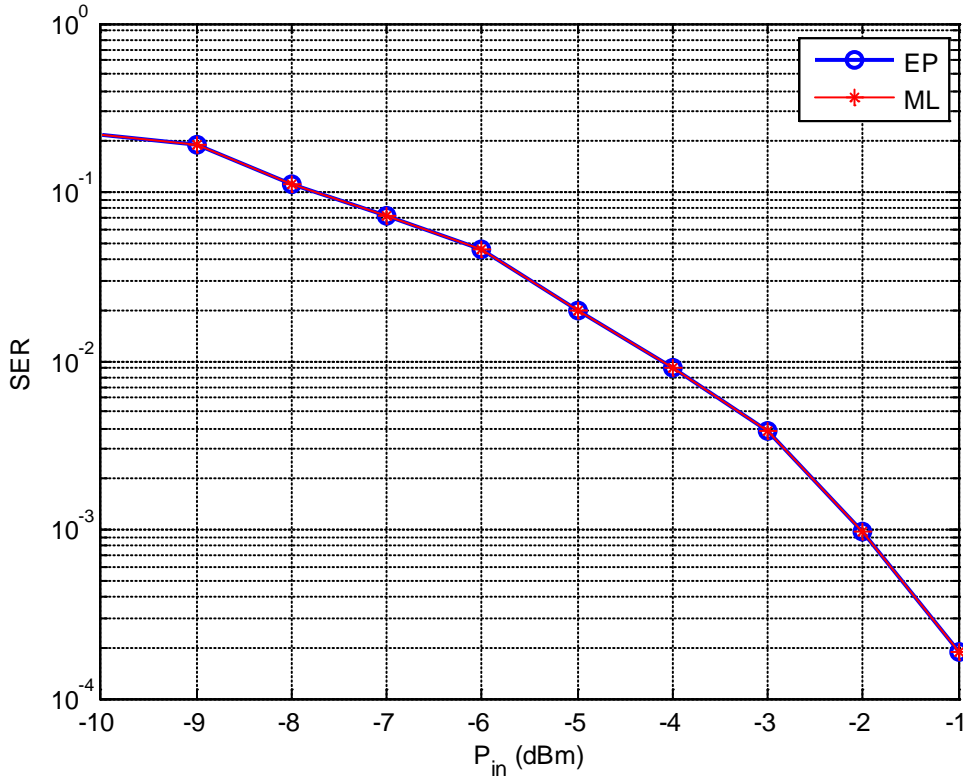

Figure 6.1 SER as a function of P_{in} with 16-QAM

The figure shows that the two curves coincide with each other. The EP and ML receiver has the same performance with each other. This is reasonable, because in EP process, the observation r_N is just the mean matrix of the first message. According to the message approximation method for linear block in section 5.1, the mean of the message won't change while passing through the factor graph while the covariance of the new message is the covariance of the old message plus N_0 (covariance of AWGN).

Hence, the final message used to make the decision has the mean matrix which is equal to r_N and covariance matrix which is equal to the sum of N_0 (covariance of AWGN). Therefore, the decision from EP is expected to be the same with the result from ML receiver which measures the 'distance' between the observation r_N and all possible original signals.

One more comment in this case can be given is that the process of EP is more efficient than SBP. The reason is that in SBP a number of samples should be generated and transmitted back through the model. For each linear block, every sample should be added an AWGN in order to compensate for the distortion caused by the linear block. And finally Monte Carlo integration

has to be used. While for EP, only the mean and covariance are transmitted and changed in the factor graph.

6.3 System with linear and nonlinear block

This section will firstly show how the message is changed when passing through the factor graph. And the influence of the number of samples in Monte Carlo in nonlinear block will be also discussed. The comparison between EP and other methods in SER is given at last.

6.3.1 Message changing through the factor graph

It has already been mentioned that the message passing though the factor graph is approximated into Gaussian distributions. The figures below displays how the message (Gaussian distribution) is changed from the end to the beginning of the factor graph in one polarized system.
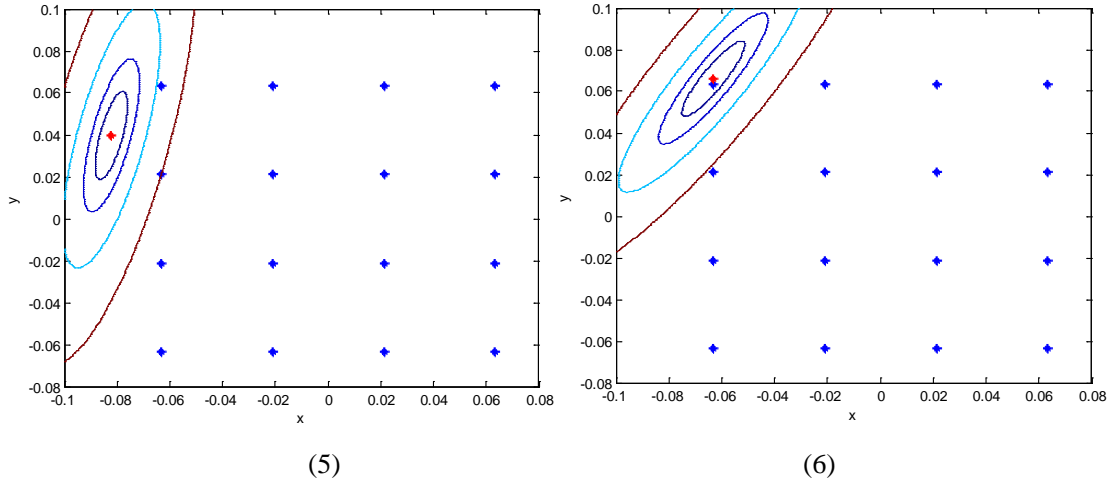

(1)

(2)

(4)

(3)

(5)   (6)
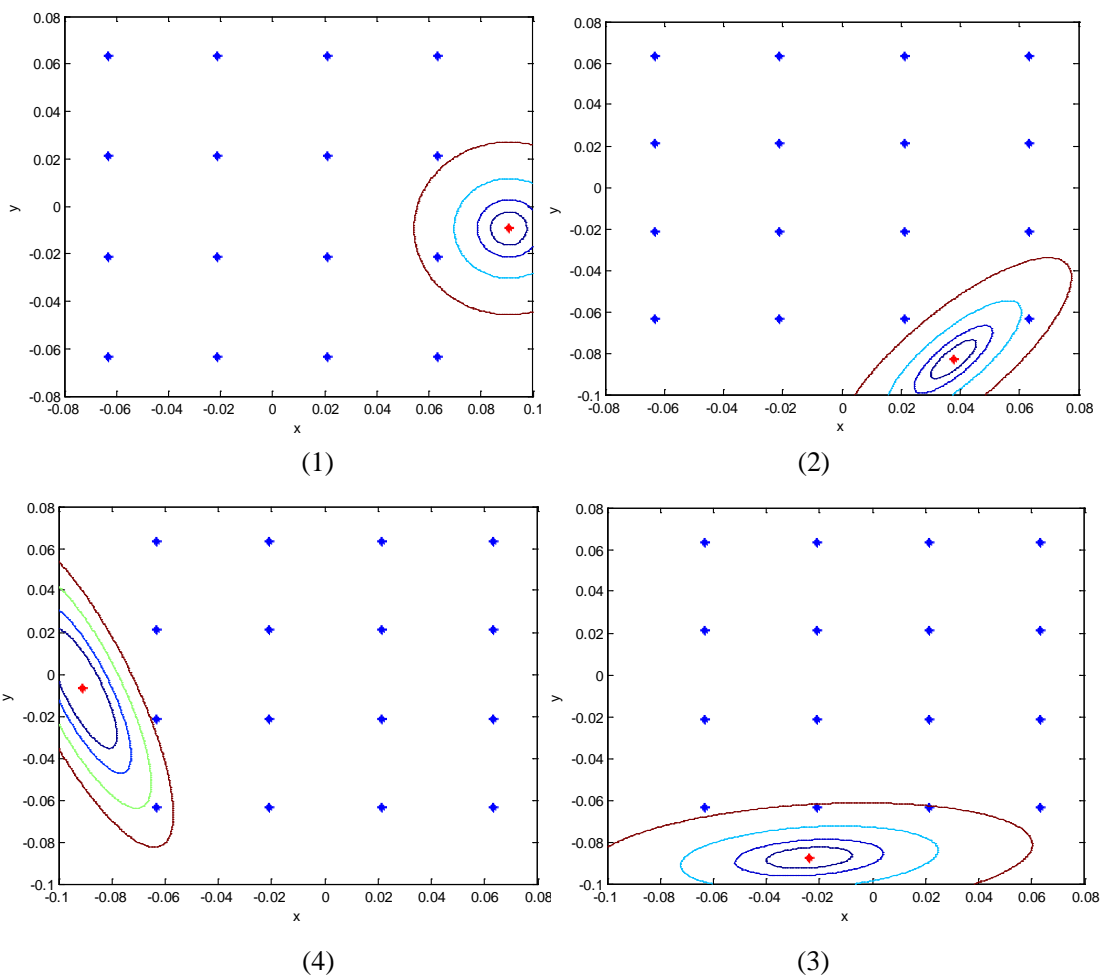
Figure 6.2 Message changing from the end to the beginning of the model

The Observation and original signal in the case of figure 6.2 are:

$$r_N = [0.0906, -0.0093], \qquad s = [-0.0634, 0.0634]$$

From the figure, one can see that the approximated Gaussian distribution finally reach the original signal step by step. Actually, a correct decision can be just made from the fifth subplot of figure 6.2 for the reason that the value of the Gaussian distribution for the original signal is already the maximum one among the 16-QAM constellation.

Note that the whole process has 44 steps for 22 spans. Figure 6.2 only shows 6 steps to roughly present the way EP works.

6.3.2 Effect of the number of samples in Monte Carlo

It has been already discussed that the Monte Carlo method has to be used to make the approximation for the messages for each nonlinear block. Hence, the number of samples drew by using Monte Carlo is expected to affect the final result. Figure 6.3 shows the relationship between the SER and the number of samples for the two polarized system with a fixed input power in 4 dBm.

The figure shows that the SER goes down dramatically at first and doesn't decrease much after the number of sample reaches certain value. Because the number of sample in Monte Carlo will obviously influence the speed of the process, a 'cost-effective' number of samples should be chosen when considering the efficiency of the process. Figure 6.3 displays that a number around 300 could be a reasonable one for the system. In this thesis, consider the condition of the simulation equipment and comparison, the number of 100 is chosen
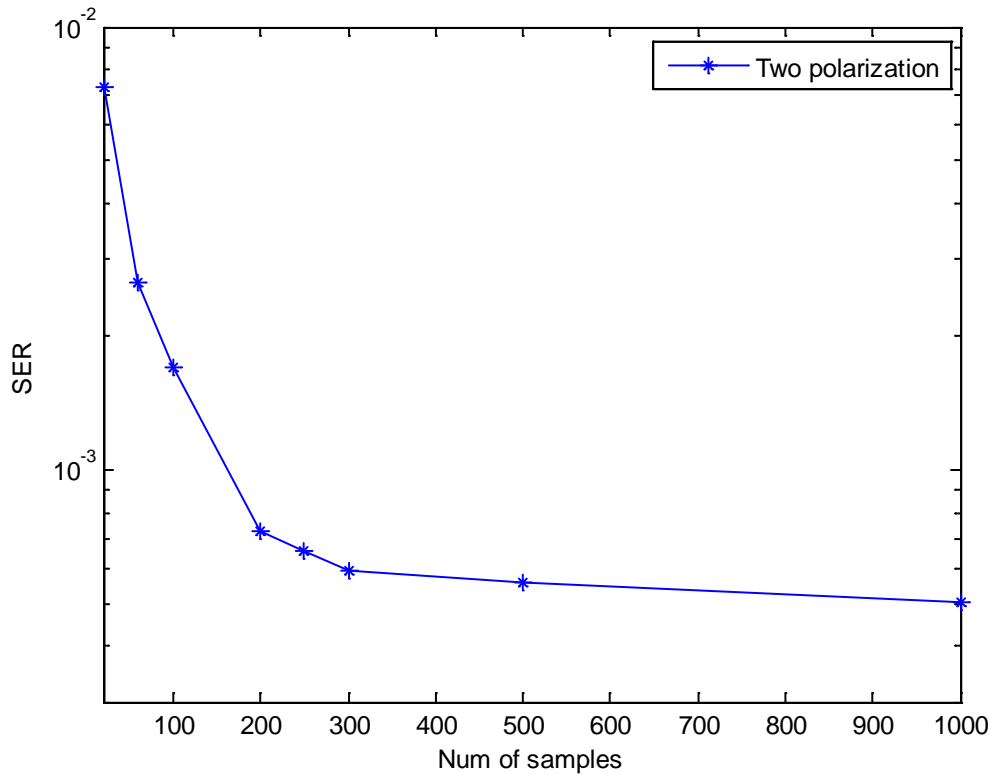
Figure 6.3 SER along with the number of samples for dual polarization

6.3.3 Comparison with other methods

In this section, a comparison has been made between EP and other methods aiming to analyze the different performance of these methods.

Figure 6.4 and 6.5 compare the SER along with the input power between ML, BP, SBP and EP method for single polarization and dual polarization respectively. 100 particles (samples) are used in SBP and EP in both single and dual polarization. From these two figures, following comment could be obtained:

1. All the methods except ML have the lowest SER in the middle area along the X axis and the performance roughly get worse while the input power decrease and increase from the lowest SER point.

2. All the methods have better performance in single polarization system.

3. EP has a greater degradation in performance than SBP and BP dose when the system extends from single polarization to dual polarization.
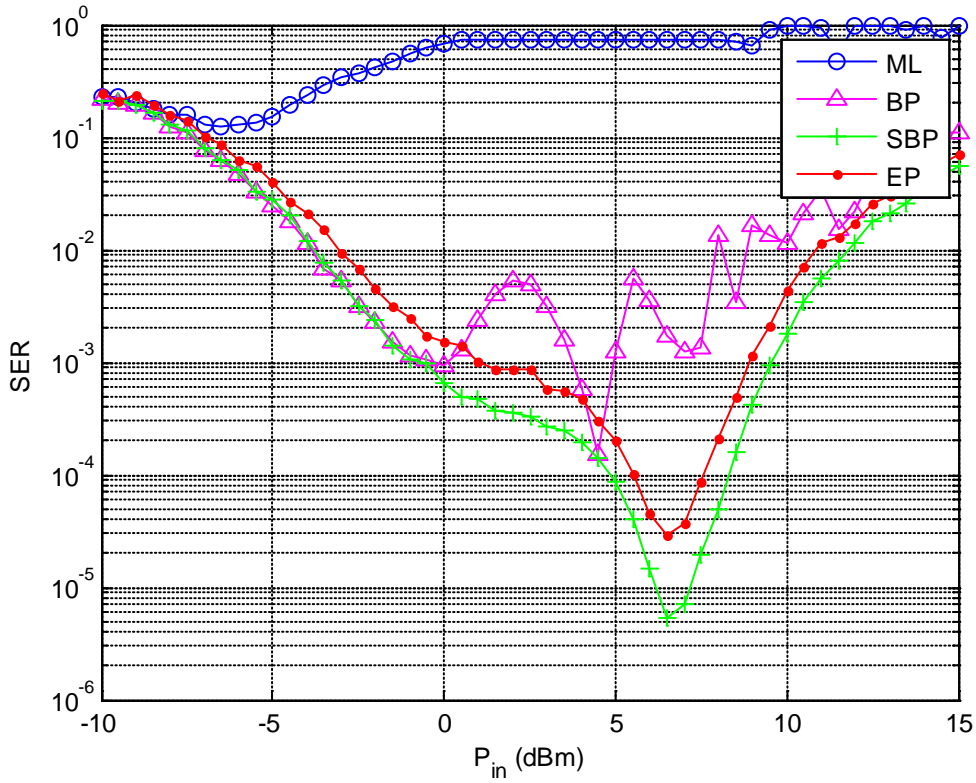
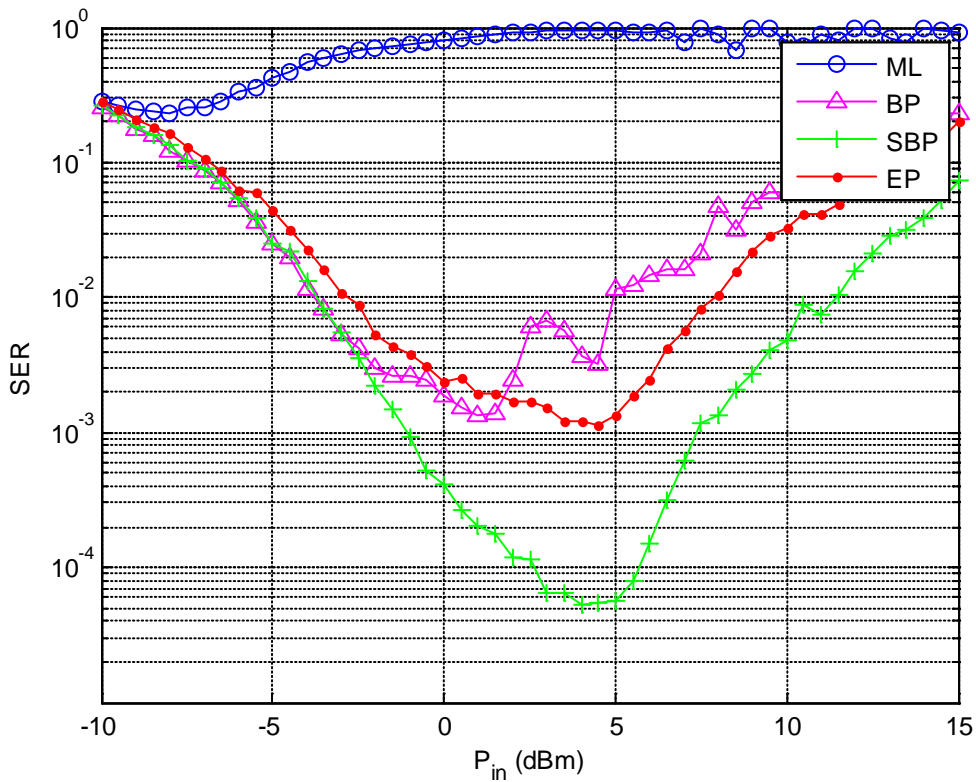Figure 6.4 SER for single polarization with 16-QAM

Figure 6.5 SER for dual polarization with 16-QAM

4. In both single and dual polarization system, SBP and ML have the best and worst results respectively. And BP and ML are comparable with each other.

5. The curve of BP method has oscillations in both single and dual polarization.

Firstly, because the ML method just neglects the nonlinear effect, so it is not strange that it has the worst performance while the input power grows. All of the rest three methods have taken nonlinear into consideration. AWGN and nonlinear effect are the main distortions for which BP, SBP and EP try to compensate. When the input power is pretty small the AWGN contribute the most to the distortion in the system. When the input power is getting bigger, the effect of the AWGN becomes smaller and the SER goes down until it reaches the lowest point. As the input power continues increasing, the nonlinear effect dominates the distortion and will increase the value of SER. That is why the SER curve has 'V' shape in the figure.

In dual polarization the rotation caused by the nonlinear block is based on both two signals transmitted simultaneously in the system (formula 2.5). Hence, the nonlinear effect in dual polarization is stronger than that in single polarization and the obtained worse results.

Moreover, it can be found that the performance of EP obviously has a bigger degradation than that of SBP and BP has from the single to dual polarization. The possible reason for this phenomenon is that: In the dual polarization system, the nonlinear rotation effect is greater than that in the single polarization case. Hence, the Gaussian distribution chosen to approximated the message in the system may be no longer a good approximation term. And it can be also expected that the performance of EP with Gaussian distribution as the approximation term will have a great drop when the system extends to triple or quadruple polarization system.

It can be seen from the figure that SBP give out the best performance in SER. The possible reason is that:

Compared with BP, SBP has taken care of AWGN while BP doesn't. In the nonlinear system model each AWGN block is followed by a nonlinear block. When the input power is high, the interaction between the AWGN and the nonlinear rotation also get bigger, which is supposed to be the reason for the poorer performance for BP in high input power compared with SBP and EP. Moreover, because of the natural uncertainty of AWGN, the uncertainty of the result SER for BP is also expected. That can explain the oscillation happened in the BP curve in the figure above.

Compared with BP, both SBP and EP have taken AWGN and nonlinear into consideration. Nevertheless, SBP uses particles and makes a stochastic process in the model while EP makes approximations block by block, in other words, EP process continues losing the accuracy while it makes approximations for the blocks in the model. That is why EP obtains the SER curve which has the same shape with SBP but a worse SER.

One more thing should be mentioned is that the strategy of EP is making a balance between the accuracy of the result and the efficiency of the process. In the linear system, EP successfully makes a more efficient algorithm than SBP does. However, because of the use of Monte Carlo method for the nonlinear block, the performance of EP in the nonlinear system is more or less the same with SBP when it comes to the efficiency.

Chapter 7

Conclusion and future work

7.1 Conclusion

The thesis works on factor graphs and expectation propagation (EP) and try to implement these technologies into the coherent optical communication system aiming to deal with the linear and nonlinear distortion existed in the system. The results indicate that factor graphs and EP works well on both linear (AWGN) and nonlinear system. Moreover, Gaussian distribution has also been proved a reasonable approximation term for EP to approximate the messages in the system.

Comparison has also been made between EP and other existing methods. Generally, the performance of EP is as expected.

7.2 Future work

According to the discussions in the thesis, the EP algorithm and the message computation method could be improved to make a more efficient algorithm. Moreover, besides the linear and nonlinear effect, other kinds of distortion and components in the system could also be take into consider.

7.2.1 Expectation propagation

As discussed in chapter 6, the Monte Carlo method used in the message approximation for nonlinear block makes EP algorithm less efficient. Hence, more work could be taken on how to make message approximation process mathematically or use other methods which can be more efficient than Monte Carlo.

What's more, discussion shows that, as an approximation term, the performance of Gaussian distribution degrades a lot in the multiplexed polarization system. Therefore, a better distribution could be chosen from the exponential family for the multiplexed polarization system.

7.2.2 Other problems

Besides the linear and nonlinear distortion discussed in the thesis, other problems could also be taken into consider.

Polarization dependent loss (PDL): PDL is also a significant distortion which influences the performance of the system. PDL can be calculated as the ratio of the maximum and minimum power amplitude to all the states of the polarization. In the optical system model, PDL exists in each span of fiber and the PDL of the system cannot be obtained by just adding all the PDL components together but should be considered separately [11].

More noise: This thesis only takes care of AWGN and nonlinear phase noise. In the future work, more noise could be taken into consider for the system, such as photoelectron noise, thermal noise and photon noise.

Moreover, the unitary channel [11] and the corresponding equalizer are also suggested to be added into the system model.

References

[1] S. Tan, H. Wymeersch, P. Johannisson, E. Agrell, P. Andrekson, and M. Karlsson, "An ML-based detector for optical communication in the presence of nonlinear phase noise," *International Conference on Communications*, 2010.

[2] Nan Jiang, Yan Gong, Johnny Karout, Henk Wymeersch, Pontus Johannisson, Magnus Karlsson, Erik Agrell, Peter A. Andrekson, *"Stochastic Backpropagation for Coherent Optical Communications,"* 37th European Conference and Exhibition on Optical Communication (ECOC), 2011.

[3] H. Wymeersch, *Iterative Receiver Design*. Cambridge University Press, 2007, pp. 36-75.

[4] Scott L. Miller, Donald G. Childers, *Probability and Random Processes with Applications to Signal Processing and Communications*. Elsevier Academic Press, 2004, pp. 148-232

[5] Peter Ahrendt, *The Multivariate Gaussian Probability Distribution*. IMM, Technical University of Denmark, 2005.

[6] Thomas P Minka, *"A family of algorithms for approximate Bayesian inference,"* Doctor's thesis, 2001.

[7] Thomas P Minka, *"Expectation Propagation for Approximate Bayesian Inference,"* Statistics Dept. Carnegie Mellon University, 2002.

[8] Tom Heskes, Onno Zoeter, *"Extended Version of "Expectation propagation for approximate inference in dynamic Bayesian networks","* SNN, University of Nijmegen, The Netherlands, 2003.

[9] David J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2005, pp. 357–382.

[10] Eric C. Anderson, *"Monte Carlo Methods and Importance Sampling,"* Lecture Notes for Stat 578C, Statistical Genetics, 1999

[11] Yan gong, Nan jiang, *"Stochastic Backpropagation for Coherent Optical Communication,"* Master's thesis in optical communication systems, 2011

[12] Shuang Wang, *"Expectation Propagation Algorithm,"* School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK, 74135, 2011.