

广东工业大学硕士学位论文

(工学硕士)

贝叶斯推论快速近似算法研究

邹秋云

二〇一九年五月

分类号:

学校代号: 11845

UDC:

密级:

学号: 2111604051

广东工业大学硕士学位论文

(工学硕士)

# 贝叶斯推论快速近似算法研究

邹秋云

校内导师姓名、职称: 张浩川副教授

学科(专业)或领域名称: 控制科学与工程

学生所属学院: 自动化学院

论文答辩日期: 2019年5月29日

A Dissertation Submitted to Guangdong University of Technology for the Degree of Master  
(Master of Engineering Science)

Research on Fast Approximate Algorithm based Bayesian Inference

Candidate: Zou Qiuyun

Supervisor: A/Prof. Zhang Haochuan

May 2019

School of Automation

Guangdong University of Technology

Guangzhou, Guangdong, P. R. China,

## 符号对照表

$\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m \times n}$	一维实数, $n$ 维实向量, $m \times n$ 维实矩阵
$\mathbb{C}, \mathbb{C}^n, \mathbb{C}^{m \times n}$	一维复数, $n$ 维复向量, $m \times n$ 维复矩阵
$\triangle$	定义
$\mapsto$	映射
$\mathbb{R}^p \mapsto \mathbb{R}^n$	从 $p$ 维空间到 $n$ 维实数空间的映射
$\mathbb{C}^p \mapsto \mathbb{C}^n$	从 $p$ 维空间到 $n$ 维复空间的映射
$(\cdot)^T$	转置
$(\cdot)^H$	共轭转置
$ \cdot $	矩阵行列式
$\mathbf{H}$	矩阵
$\mathbf{x}$	向量
$\mathbf{I}_n$	$n \times n$ 维单位矩阵
$\mathbb{E}[\cdot]$	数学期望
$\text{Var}[\cdot]$	方差
$\exp(\cdot)$	以自然常数 $e$ 为底的指数函数
$\mathcal{N}(a, A)$	均值为 $a$ , 方差为 $A$ 的高斯概率密度函数
$\mathcal{N}(x a, A)$	自变量为 $x$ , 均值为 $a$ , 方差为 $A$ 的高斯概率密度函数
$\mathcal{N}(\mathbf{a}, \mathbf{A})$	均值为 $\mathbf{a}$ , 协方差为 $\mathbf{A}$ 的矢量高斯概率密度函数
$\mathcal{N}(\mathbf{x} \mathbf{a}, \mathbf{A})$	自变量为 $\mathbf{x}$ , 均值为 $\mathbf{a}$ , 协方差为 $\mathbf{A}$ 的矢量高斯概率密度函数
$\propto$	正比于
$\mathcal{N}_c(\cdot)$	复高斯分布
$\Sigma$	求和
$\text{Proj}[\cdot]$	投影, 如 $q(x) = \text{Proj}[p(x)]$
$\sim$	服从分布, 如 $X \sim p_X(x)$
$\odot$	逐点乘
$\oslash$	逐点除

$\text{diag}(\mathbf{A})$	将 $\mathbf{A}$ 对角线元素抽出，形成列向量
$\text{Diag}(\mathbf{a})$	将 $\mathbf{a}$ 的元素作为对角线元素，形成矩阵，非对角线元素为0
$\langle \cdot \rangle$	算术平均或概率期望，如 $\langle \mathbf{x} \rangle = \frac{1}{N} \sum_{i=1}^N x_i$ , $\langle x \rangle_q = \int x q(x) dx$
$\  \cdot \ _p$	$p$ -范数，若无指明默认为2-范数

## 缩写语

ADC	analog to digital converter, 模数转换器
ADF	assumed density filter, 假设密度滤波
AMP	approximate message passing, 近似消息传递
AWGN	additive white Gaussian noise, 加性高斯白噪声
BP	belief propagation, 置信传播
BS	base station, 基站
CLT	central limit theorem, 中心极限定理
CRLB	Cramer-Rao lower bound, 克拉美-罗下限
CS	compressive sensing, 压缩感知
DAMP	denosing-based approximate message passing, 基于降噪的近似消息传递
EC	expectation consistent, 期望一致
EP	expectation propagation, 期望传播
FG	factor graph, 因子图
GAMP	generalized approximate message passing, 广义近似消息传递
GEC	generalized expectation consistent, 广义期望一致
GEC-SR	generalized expectation consistent-signal recovery, 广义期望一致信号重构
GLM	generalized linear model, 广义线性模型
Gr-VAMP	generalized vector approximate passing, 广义矢量消息传递
LASSO	least absolute shrinkage and selection operator, 最小绝对值收缩和选择算子
LBP	loopy belief propagation, 循环置信传播
LMMSE	linear minimum mean square error, 线性最小均方误差
LS	least square, 最小二乘
MAP	maximum a posterior, 最大后验概率
MIMO	multi-input and multi-output, 多入多出

---

ML-AMP	multi-layer approximate message passing, 多层近似消息传递
ML-GEC	multi-layer generalized expectation consistent, 多层广义期望一致
ML-GLM	multi-layer generalized linear model, 多层广义线性模型
MSE	mean square error, 均方误差
MP	message passing, 消息传递
OAMP	orthogonal approximate message passing, 正交近似消息传递
PDF	probability distribution function, 概率密度函数
QPSK	quadrature phase shift keying, 正交相位键控
RBP	relaxed belief propagation, 松弛置信传播
RIP	restricted isometry property, 有限等距原理
RM	replica method, 副本方法
RV	random variable, 随机变量
SE	state evolution, 状态更新
SISO	single-input and single-output, 单入单出
SLM	standard linear model, 标准线性模型
SNR	signal-to-noise rate, 信噪比
SVD	singular value decomposition, 奇异值分解
TAP-lize	Thouless-Anderson-Palmer-lize, TAP化
VAMP	vector approximate message passing, 矢量近似消息传递
VMP	varitional message passing, 变分消息传递

## 摘要

本篇论文主要研究了贝叶斯算法及其在逆问题中用于重构目标信号的应用。更确切地说, 本文归纳推导了两类前沿的算法, 分别是近似消息传递类算法 (approximate message passing-like, AMP-like) 以及期望传播类算法 (expectation propagation-like, EP-like), 其中AMP-like算法包括: 近似消息传递 (approximate message passing, AMP)、广义近似消息传递 (generalized approximate message passing, GAMP)、多层广义近似消息传递 (multi-layer generalized approximate message passing, ML-GAMP) 等; EP-like算法包括: 期望传播 (expectation propagation, EP)、期望一致 (expectation consistent, EC)、广义期望一致信号重构 (generalized expectation consistent signal recovery, GEC-SR)、多层广义期望一致 (multi-layer generalized expectation consistent, ML-GEC) 等。

AMP算法最早由Donoho等人提出, 用于解决压缩感知领域的标准线性逆问题的信号重构。随后, Rangan等人扩展了AMP对噪声分布的限制, 提出了GAMP算法, 使之适用于广义线性模型。AMP和GAMP都是源于因子图的Sum-Product算法, 因其具有低复杂度、性能优异的特点而受到广泛关注。实验部分也验证了AMP-like算法在信号重构方面的杰出性能。

EP算法最早见于Minka的博士论文, 用于贝叶斯网络的因子可分解的概率近似应用。EP改进了假设密度滤波算法消息更新的方式, 并且与变分法有着很强的联系。更确切地说, EP与变分法具体区别在于对KL散度的使用上。此外, EP与AMP也存在着很强的联系, 通过用图来表示EP, 即Sum-Product算法是一种完全因子化的EP, 说明了AMP是EP的一类算法。与AMP相似, EP在广义线性模型、多层广义线性模型也有着相应地扩展。GEC和GEC-SR将EP扩展到广义线性模型, 用于解决广义线性逆问题的目标信号的恢复。GEC算法虽然适用于广义线性模型, 但是其算法稳定性极差。GEC-SR通过加入线性空间 $\delta(z - Hx)$ 模块, 解决了GEC的收敛问题。

本文推导归纳了, AMP-like算法和EP-like算法, 分别是

- 提出了一种基于分列式因子图的AMP推导方法。分列式因子图相对于传统因子图而言, 具有可扩展到多层模型的优势。
- 提出了一种新的消息更新规则用于推导VAMP算法。该消息更新规则结合分列式因



子图，对于多层模型同样适用。

- 提出了一种基于EP的GAMP推导方法。该方法极大地简化了GAMP的推导过程，并且将GAMP扩展到复数领域。此外，该方法建立了EP与GAMP之间的直接联系。
- 提出了一种基于分列式因子图的GAMP推导方法。该方法相对于原始的GAMP推导过程极大简化了推导步骤，同时该方法对于复数和实数运算同样适用。
- 补充了GEC-SR的证明。

**关键字：**信号重构、贝叶斯估计、近似消息传递、期望传播

## Abstract

This thesis mainly studies Bayesian algorithm and its application in inverse problem for recovering the signal of interest. Specifically, this thesis has derived two state-of-the-art algorithm which are approximate message passing like (AMP-like) algorithm and expectation propagation (EP-like) algorithm, respectively. AMP-like algorithm includes approximate message passing (AMP), generalized approximate message passing (GAMP), and multi-layer generalized approximate message passing (ML-GAMP) etc. EP-like algorithm comprises expectation propagation (EP), expectation consistent (EC), generalized expectation consistent (GEC), generalized expectation consistent signal recovery (GEC-SR), and multi-layer generalized expectation consistent (ML-GEC), etc.

The AMP algorithm was initially proposed by Donoho et al. to solve the signal recovery of standard linear model in compressive sensing. Subsequently, Rangan et al. broke the restriction of the noise distribution in AMP and proposed GAMP algorithm, which makes it apply to generalized linear model. Both AMP and GAMP are derived from Sum-Product algorithm of factor graph, which are widely concerned by their low complexity and excellent performance. The part of experiment has also verified the excellent performance of AMP-like algorithm in signal recovery.

EP was initially found in Minka's doctoral thesis, which was applied to approximate the factorable distribution in Bayesian network. EP develops the form of message update of assumed density filtering. In addition, there is a strong relationship between variational inference and EP. To be more precise, the difference between EP and variational inference is the form of KL-divergence. Moreover, there is also a strong connection between EP and AMP. Using the factor graph to represent the EP, it can be verified that Sum-Product algorithm is a class of EP. That also states that AMP is a class of expectation propagation (EP) techniques in fully factorable case. Similar to AMP, EP also has extension in generalized linear model and multi-layer generalized linear model. GEC and GEC-SR extend EP to generalized linear model for recovering the signal of interest in generalized inverse problem. Although GEC applies to generalized linear model, its convergence is poor. With a linear space, GEC-SR overcomes this problem perfectly.

Totally, this thesis has derived AMP-like algorithm and EP-like algorithm including

- Using a split factor graph to derive AMP. Comparing the traditional factor graph, this factor graph is easy to extend to multi-layer model.

- We have proposed a novel message passing rules called “moment-matched message passing”. With this message update rules, it is natural to obtain the derivation of VAMP algorithm. In addition, combining the rules and the split factor graph is also appropriate to multi-layer model.
- We have given a derivation of GAMP based on EP. It reduces the steps of GAMP’s derivation sharply. Beside that, it also covers the complex domain and builds a bridge between EP and GAMP.
- We have presented a derivation of GAMP based on the split factor graph. Comparing the original derivation of GAMP, it reduces the steps of GAMP sharply and both covers the real domain and complex domain. More essentially, it is easy to extend to multi-layer model.
- We have given a proof of GEC-SR.

**Key Words:** signal recovery, Bayesian estimation, approximate message passing, expectation propagation.

# 目录

摘要 .....	V
Abstract .....	VII
第一章 绪论 .....	1
1.1 研究背景和意义 .....	1
1.2 国内外研究现状 .....	3
1.3 章节安排 .....	5
1.4 主要贡献 .....	5
第二章 数学基础 .....	7
2.1 矩阵论与概率论 .....	7
2.1.1 矩阵论 .....	7
2.1.2 概率论 .....	8
2.2 贝叶斯估计原理 .....	12
2.2.1 最小均方误差估计 .....	13
2.2.2 最大后验概率估计 .....	14
2.3 因子图 .....	15
2.3.1 因子图表示 .....	15
2.3.2 消息传递算法 .....	16
第三章 标准线性逆问题的快速解法 .....	18
3.1 标准线性逆问题 .....	18
3.2 传统估计方法 .....	20
3.2.1 线性最小均方误差估计器 .....	20
3.2.2 最小二乘估计器 .....	22
3.3 近似消息传递算法 .....	23
3.3.1 算法推导 .....	23
3.3.2 状态更新方程 .....	31
3.4 矢量消息近似传递 .....	33
3.4.1 算法推导 .....	34
3.4.2 状态更新方程 .....	39

3.5 应用: Massive MIMO 信号检测.....	40
第四章 广义线性逆问题的快速解法.....	47
4.1 广义线性逆问题.....	47
4.2 传统估计方法.....	49
4.2.1 标准线性模型的构造.....	49
4.2.2 传统估计方法.....	50
4.3 广义近似消息传递算法.....	51
4.3.1 算法推导.....	51
4.4 广义期望一致信号重构.....	61
4.4.1 概述.....	61
4.4.2 算法推导.....	63
4.4.3 状态更新方程.....	67
4.5 应用: 带低精度 ADC 的 Massive MIMO 信号检测.....	68
结论与展望.....	77
参考文献.....	78
附录.....	82
攻读学位期间发表的学术论文.....	84
学位论文独创性声明.....	85
致谢.....	86

# Contents

Abstract (Chinese) .....	V
Abstract (English).....	VII
Chapter 1 Introduction .....	1
1.1 Background and Significance .....	1
1.2 Research Status .....	3
1.3 Chapters' arrangement.....	5
1.4 Major Contributions .....	5
Chapter 2 Fundamentals of Mathematics.....	7
2.1 Theory of Matrices and Probability.....	7
2.1.1 Theory of Matrices .....	7
2.1.2 Theory of Probability .....	8
2.2 Principle of Bayesian Estimation .....	12
2.2.1 Minimum Mean Square Error Estimation.....	13
2.2.2 Maximum A Posterior Estimation .....	14
2.3 Factor Graph .....	15
2.3.1 Representation using Factor Graph.....	15
2.3.2 Message Passing .....	16
Chapter 3 Fast Recovery of Standard Linear Inverse Problem.....	18
3.1 Standard Linear Inverse Problem .....	18
3.2 Conventional Estimation.....	20
3.2.1 Linear Minimum Mean Square Error Estimation .....	20
3.2.2 Least Square Estimation .....	22
3.3 Approximate Message Passing.....	23
3.3.1 Derivation .....	23
3.3.2 State Evolution.....	31
3.4 Vector Approximate Message Passing .....	33
3.4.1 Derivation .....	34
3.4.2 State Evolution.....	39
3.5 Application in Massive MIMO Signal Detection .....	40
Chapter 4 Fast Recovery of Generalized Linear Inverse Problem .....	47
4.1 Generalized Linear Inverse Problem .....	47
4.2 Conventional Estimation.....	49
4.2.1 Construction of Standard Linear Model.....	49
4.2.2 Conventional Estimation.....	50
4.3 Generalized Approximate Message Passing .....	51

4.3.1 Derivation .....	51
4.4 Generalized Expectation Consistent Signal Recovery .....	61
4.4.1 Introduction .....	61
4.4.2 Derivation .....	63
4.4.3 State Evolution.....	67
4.5 Application in Massive MIMO Signal Detection with Low-Resolution ADCs.....	68
Conclusion and Future Work .....	77
References .....	78
Appendix.....	82
Publication.....	84
Statement .....	85
Acknowledgements.....	86

# 第一章 绪论

## 1.1 研究背景和意义

信号重构理论是信号处理领域的一个重要课题，在通信信号处理<sup>[1]-[5]</sup>、模式识别<sup>[6]-[10]</sup>、压缩感知<sup>[11]-[13]</sup>等领域有着重要的应用。信号重构的实质是从观测向量 $\mathbf{y}$ 中恢复出目标信号 $\mathbf{x}$ ，该过程用映射 $g(\cdot)$ 表示，当为实数运算时， $g(\cdot) : \mathbb{R}^M \mapsto \mathbb{R}^N$ ，当为复数运算时， $g(\cdot) : \mathbb{C}^M \mapsto \mathbb{C}^N$ 。观测数据的形成以及信号检测的过程，可以通过图1-1来表示。这种问题，称之为逆问题<sup>[14]</sup>（inverse problem），即系统中无法观测到目标信号，但可以从观测向量中通过代数运算求解得到目标信号。

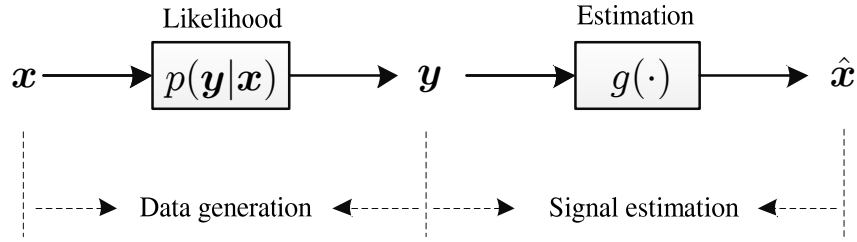


图1-1 观测数据生成及信号检测

在观测数据生成过程中，似然函数 $p(\mathbf{y}|\mathbf{x})$ 描述了目标信号与观测值的统计特性，也称转移概率。更确切地说，观测值 $\mathbf{y}$ 与目标信号 $\mathbf{x}$ 的直接关系通过函数 $f(\cdot)$ 表示，其中 $f(\cdot)$ 可以是线性函数，也可以是非线性函数（如广义线性模型时， $f(\cdot)$ 为非线性函数）。在众多信号处理的实际应用中，标准线性模型（standard linear model, SLM）有着较为广泛的应用<sup>[2]-[14]</sup>。SLM包括一个线性混合部分 $\mathbf{z} = \mathbf{H}\mathbf{x}$ 以及一个加性噪声扰动部分 $\mathbf{y} = \mathbf{z} + \mathbf{w}$ 。作为SLM的扩展，广义线性模型<sup>[15]</sup>（generalized linear model, GLM）用转移概率 $p(\mathbf{y}|\mathbf{z})$ 替换了加性噪声扰动。进一步地，多层广义线性模型<sup>[16][17]</sup>（multi-layer generalized linear model, ML-GLM）对GLM进行了层数的扩宽。

### ■ 标准线性模型

标准线性模型的形式如下

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (1.1)$$

其中 $\mathbf{y} \in \mathbb{R}^M$ 是 $M$ 维观测向量， $\mathbf{H} \in \mathbb{R}^{M \times N}$ 为线性观测矩阵， $\mathbf{w} \in \mathbb{R}^M$ 为噪声扰动项，



$\mathbf{x} \in \mathbb{R}^N$ 为目标信号。线性逆问题所研究的是从参数 $(\mathbf{y}, \mathbf{H})$ 中重构出目标信号 $\mathbf{x}$ 来。该问题可以归纳为LASSO问题<sup>[12][18]</sup>

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (1.2)$$

其中 $\|\cdot\|_p$ 表示 $p$ -范数。

在压缩感知领域（compressive sensing, CS）， $\mathbf{H}$ 为感知矩阵。CS所研究的是稀疏信号重构问题。这种情况下，观测向量的维度远小于目标信号的维度，即  $M \ll N$ （方程个数远小于未知数个数）。通常，这种情况下无法得到精确的 $\mathbf{x}$ 的估计。若目标信号 $\mathbf{x}$ 满足稀疏特性，这使得从少量观测数据中重构出目标信号成为可能。此外，稀疏信号重构的另外一个条件是感知矩阵 $\mathbf{H}$ 满足有限等距原理<sup>[19]</sup>（restricted isometry property, RIP）。

在其他领域，如通信信号处理领域，通常 $M > N$ ，即观测向量的长度大于目标信号的长度。此时目标信号为稠密信号。本文3.4节将详细介绍SLM在通信信号处理中的具体应用。

## ■ 广义线性模型

GLM扩展了SLM的适用范围，更确切来说，SLM是GLM的特例。GLM模型表示如下

$$\mathbf{x} \rightarrow \mathbf{z} = \mathbf{H}\mathbf{x} \rightarrow p(\mathbf{y}|\mathbf{z}) \rightarrow \mathbf{y} \quad (1.3)$$

与SLM所不同，GLM用一个逐点映射的转移概率 $p(\mathbf{y}|\mathbf{z}) = \prod_{a=1}^M p(y_a|z_a)$ 替换了SLM中的加性噪声项。转移概率 $p(\mathbf{y}|\mathbf{z})$ 刻画了中间变量 $\mathbf{z}$ 与观测向量 $\mathbf{y}$ 的统计特性，两者之间的直接关系，可以通过一个函数映射 $Q(\cdot): \mathbb{R}^M \rightarrow \mathbb{R}^M$ 来表示（在复数域为 $Q(\cdot): \mathbb{C}^M \rightarrow \mathbb{C}^M$ ），通常 $Q(\cdot)$ 是非线性的，如带低精度ADC量化的Massive MIMO系统（见4.5节）。特别地，当转移概率为高斯分布时，GLM退化为SLM。

将广义线性模型看成是一个模块，通过级联多个广义线性模型，可以构成ML-GLM。GLM是ML-GLM当 $L=1$ 时的特例，而SLM是GLM当转移概率 $p(\mathbf{y}|\mathbf{z})$ 为高斯分布时的特例，见图1-2。目前，ML-GLM在诸如：Massive MIMO中继信号检测、贝叶斯网络<sup>[16]</sup>等领域有着重要应用。

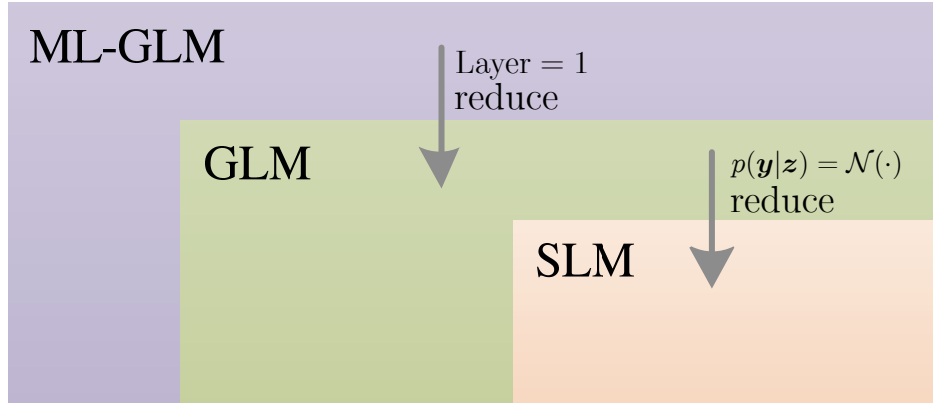


图1-2 三种模型的关系

## 1.2 国内外研究现状

标准线性逆问题所研究的是给定观测矩阵 $\mathbf{H}$ ，噪声分布 $p(\mathbf{w})$ ，以及目标信号的先验分布 $p(\mathbf{x})$ ，从观测向量 $\mathbf{y}$ 中解出目标信号 $\mathbf{x}$ 。目前，针对标准线性逆问题，现有文献中存在着多种算法，其中有两类算法在学术界有广泛关注，一类是基于似然函数的估计方法，如稀疏贝叶斯学习<sup>[7]</sup>；一类是贝叶斯算法，如近似消息传递算法（approximate message passing, AMP）、期望传播算法（expectation propagation, EP）。表1-1归纳了目前几种较为前沿的贝叶斯算法。

### ■ AMP-like

基于低复杂度因子图的消息传递<sup>[20]</sup>（message passing, MP）算法，也称置信传播（belief propagation, BP）算法，由于其具有低复杂度、稳定性好、性能高效等特点逐渐受到广大学者的青睐。对于树形结构因子图，BP算法被证明是严格收敛的<sup>[6]</sup>。基于高斯近似，文献[21]提出了松弛置信传播（relaxed belief propagation, RBP）算法。Donoho等人于2010年提出AMP<sup>[22]</sup>算法，该算法是对RBP算法的进一步简化。此外，该文献指出，AMP的性能可以通过状态更新方程进行预测。[23]利用统计物理的条件技巧给出了AMP收敛性的严格证明。现有文献中，有很多改进AMP的算法，如正交近似传播算法（Orthogonal AMP, OAMP）<sup>[24]</sup>、基于AMP的降噪算法（denoising-based AMP, D-AMP）<sup>[25]</sup>、S-AMP<sup>[26]</sup>等。上述所描述的算法所适用的模型均为标准线性模型。

为了扩展AMP的适用范围，Rangan等人提出了广义近似消息传递算法（generalized AMP, GAMP）<sup>[15]</sup>。相对于AMP而言，GAMP算法可以同时适用于标准线性模型和广义线性模型。文献[27]提出一种统一的框架，基于该框架，原本仅仅适用于SLM的算法可

以扩展到GLM。该框架的实质是用EP来处理转移概率 $p(\mathbf{y}|\mathbf{z})$ 。[17]提出ML-GAMP算法用于解决多层广义线性逆问题。该算法是GAMP在多层广义线性模型的扩展。

## ■ EP-like

EP是贝叶斯推论的一种算法，它通过最小化KL散度来近似可因子化的概率分布，特别是在贝叶斯网络和统计学习中被用来近似后验概率。EP最早见于Minka的博士论文[9]。该文献指出，EP是假设密度滤波的改进。[10]指出，EP与变分推论有很强的联系，两者的区别在于KL-散度的具体使用上。此外，该文献还指出，EP可以用因子图来表示，从而验证了，LBP是EP的一种特例。与EP极其相似，Rangan提出了矢量消息传递算法（vector approximate message passing）<sup>[28]</sup>。EP和VAMP的区别在于，首先EP所适用的模型比VAMP要广，其次VAMP具有自成体系的消息更新规则和因子图，且其因子图可以是矢量的，而传统的因子图是标量的。特别地，对于标准线性模型，EP和VAMP的性能是完美重合的。此外，EP算法与期望一致（expectation consistent, EC）算法同样存在着很强的联系，更确切的说，EP是单循环EC。

与AMP的情况很相似，EP在广义线性模型也有相应的扩展。[29]提出了一种GLM-VAMP算法用于解决广义逆问题，其实质是构造出一个标准线性模型。[27]提出了一种统一的框架，用于将SLM的算法扩展到GLM，因此相应的有Gr-VAMP算法。然而，该结构的合理性并未得到严谨的证明。与之所不同，[30]提出了一种GEC-SR算法，该算法的性能与根据Replica Method<sup>[31]</sup>所得到的不动点方程完美匹配，间接证明了GEC-SR的最优性。对于多层广义线性模型，EP也有对应的算法：ML-VAMP<sup>[16]</sup>算法。Fletcher等人，提出了ML-VAMP算法用于解多层广义线性逆问题。然而，该文献并未给出ML-VAMP的算法的推导过程。

表1-1 消息传递类算法归纳

	标准线性模型	广义线性模型	多层广义线性模型
EP-like	1. EP 2. EC 3. VAMP ...	1. GLM-VAMP 2. Gr-VAMP 3. GEC-SR ...	1. ML-VAMP
AMP-like	1. AMP 2. OAMP 3. Str-AMP ...	1. GAMP 2. Gr-AMP ...	1. ML-GAMP

### 1.3 章节安排

本文主要研究贝叶斯重构算法及其应用,针对目前几个应用较为广泛的系统模型,总结归纳现有的几个较为前沿的算法。本文的章节安排如下

- 第一章介绍了信号重构理论及其应用。针对几个应用较为广泛的系统模型,分析归纳了国内外现有的工作。
- 第二章为基础知识,罗列了本文所需要的数学知识点以及专业知识点,如随机过程、信号处理、矩阵理论、概率论、贝叶斯估计理论、因子图等。
- 第三章介绍了线性逆问题的快速解法,推导近似消息传递算法和矢量消息传递算法。为了验证算法的性能,本章应用部分给出了Massive MIMO信号检测器的应用,根据仿真总结算法性能。
- 第四章广义线性逆问题,该类问题是对线性逆问题的扩展。本章提出了基于EP的GAMP推导方案,并将GAMP扩展到复数领域。针对特定的无线通信领域问题,给出了具有低精度ADC的Massive MIMO信号检测器的应用,根据仿真总结算法性能。

### 1.4 主要贡献

本文的主要贡献主要集中在第三章和第四章

#### ■ 第三章:

- 提出了一种基于分列式因子图的AMP推导方法。分列式因子图相对于传统因子

图而言，具有可扩展到多层模型的优势。

- 提出了一种新的消息更新规则用于推导VAMP算法。该消息更新规则结合分列式因子图，对于多层模型同样适用。

## ■ 第四章

- 提出了一种基于EP的GAMP推导方法。该方法极大地简化了GAMP的推导过程，并且将GAMP扩展到复数领域。此外，该方法建立了EP与GAMP之间的直接联系。
- 提出了一种基于分列式因子图的GAMP推导方法。该方法相对于原始的GAMP推导过程极大简化了推导步骤，同时该方法对于复数和实数运算同样适用。
- 补充了GEC-SR的推导。[30]提出了一种GEC-SR算法，但并没有给出详细的证明过程。本文补充了该证明过程。

以下是本文部分章节内容与我们已经发表论文的对应关系

## ■ 第三章

-Zhang H, **Zou Q\***, Zhao X, Wen C K, et al. Moment matched message passing part I: theory[J]. IEEE Transactions on Signal Processing, 2019. （第二作者，在审）

-Zhang H, **Zou Q\***, Zhao X, Wen C K, et al. Moment matched message passing part II: applications[J]. IEEE Transactions on Signal Processing, 2019. （第二作者，在审）

## ■ 第四章

-**Zou Q**, Zhang H\*, Wen C K, et al. Concise derivation for generalized approximate message passing using expectation propagation[J]. IEEE Signal Processing Letters, 2018, 25(12): 1835-1839. （第一作者，SCI收录）

## 第二章 数学基础

本章给出了本文其余章节所用到的现有数学知识。这些知识点包括：常用的矩阵论和概率论的现有结论、贝叶斯估计原理以及因子图。

### 2.1 矩阵论与概率论

#### 2.1.1 矩阵论

##### ■ 矩阵奇异值分解

**定义：** 设矩阵  $\mathbf{H} \in \mathbb{R}^{m \times n}$ ，且  $\text{rank}(\mathbf{H}) = r$ ，则存在  $m$  阶正交矩阵  $\mathbf{U}$  和  $n$  阶正交矩阵  $\mathbf{V}$ ，使得

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad (2.1)$$

其中  $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ ， $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$ ，并且  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$ 。[32]

**证：** 由于  $\text{rank}(\mathbf{H}) = r$ ，因此设  $\mathbf{H}^T \mathbf{H}$  的特征值为

$$\lambda_1^2 \geq \dots \geq \lambda_r^2 \geq 0, \quad \lambda_{r+1}^2 = \lambda_n^2 = 0 \quad (2.2)$$

由于  $\mathbf{H}^T \mathbf{H}$  为对称矩阵，因此必可相似对角化，即存在正交矩阵  $\mathbf{V}$  使得

$$\mathbf{V}^T (\mathbf{H}^T \mathbf{H}) \mathbf{V} = \text{Diag}(\lambda_1^2, \dots, \lambda_r^2, \underbrace{0, \dots, 0}_{n-r}) \quad (2.3)$$

记  $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$ ，其中  $\mathbf{V}_1$  为  $n \times r$  矩阵， $\mathbf{V}_2$  为  $n \times (n - r)$  矩阵，因此上式可以写为

$$\mathbf{H}^T \mathbf{H} [\mathbf{V}_1, \mathbf{V}_2] = [\mathbf{V}_1, \mathbf{V}_2] \begin{pmatrix} \mathbf{\Lambda}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (2.4)$$

则有

$$\mathbf{H}^T \mathbf{H} \mathbf{V}_1 = \mathbf{V}_1 \mathbf{\Lambda}^2, \quad \mathbf{H}^T \mathbf{H} \mathbf{V}_2 = \mathbf{0} \quad (2.5)$$

对(2.5)第一式右乘矩阵  $\mathbf{\Lambda}^{-1}$ ，有

$$\mathbf{H}^T \mathbf{H} \mathbf{V}_1 \mathbf{\Lambda}^{-1} = \mathbf{V}_1 \mathbf{\Lambda} \quad (2.6)$$

定义  $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$ ，其中  $\mathbf{U}_1$  是  $m \times r$  矩阵， $\mathbf{U}_2$  为  $m \times (m - r)$  矩阵，令  $\mathbf{U}_1 = \mathbf{H} \mathbf{V}_1 \mathbf{\Lambda}^{-1}$ ，则有  $\mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I}_r$ ，即  $\mathbf{U}_1$  为列正交规范矩阵。

取  $\mathbf{U}_2$ ，使得  $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$  为正交矩阵，则有

$$\mathbf{U}_2^T \mathbf{H} \mathbf{V}_1 = \mathbf{U}_2^T \mathbf{U}_1 \mathbf{\Lambda} = \mathbf{0} \quad (2.7)$$

因子

$$\mathbf{U}^T \mathbf{H} \mathbf{V} = \begin{pmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{pmatrix} \mathbf{H} [\mathbf{V}_1, \mathbf{V}_2] = \begin{pmatrix} \mathbf{U}_1^T \mathbf{H} \mathbf{V}_1 & \mathbf{U}_1^T \mathbf{H} \mathbf{V}_2 \\ \mathbf{U}_2^T \mathbf{H} \mathbf{V}_1 & \mathbf{U}_2^T \mathbf{H} \mathbf{V}_2 \end{pmatrix} = \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (2.8)$$

上式分别左乘 $\mathbf{U}$ ，右乘 $\mathbf{V}^T$ 得

$$\mathbf{H} = \mathbf{U} \Sigma \mathbf{V}^T \quad (2.9)$$

### 2.1.2 概率论

#### ■ 标量高斯分布

设随机变量 $X$ 服从均值为 $a$ ，方差为 $A$ 的高斯分布，即

$$p_X(x) = \frac{1}{\sqrt{2\pi A}} \exp \left[ -\frac{(x-a)^2}{2A} \right] \quad (2.10)$$

记作 $X \sim \mathcal{N}(x|a, A)$ 。

若为复高斯分布，则为

$$p_X(x) = \frac{1}{\pi A} \exp \left( -\frac{\|x-a\|^2}{A} \right) \quad (2.11)$$

记作 $X \sim \mathcal{N}_c(x|a, A)$ 。

**引理 2-1 (标量高斯相乘引理)**<sup>[1]</sup>: 假设 $\mathcal{N}(x|a, A)$ ， $\mathcal{N}(x|b, B)$ ，则有

$$\mathcal{N}(x|a, A) \mathcal{N}(x|b, B) = \mathcal{N}(0|a-b, A+B) \mathcal{N} \left( x \left| \frac{\frac{a}{A} + \frac{b}{B}}{\frac{1}{A} + \frac{1}{B}}, \frac{1}{\frac{1}{A} + \frac{1}{B}} \right. \right) \quad (2.12)$$

**证**: 标量高斯相乘引理的证明分为两个部分，分别是指数部分和常数部分。

指数部分:

$$\begin{aligned} \mathcal{N}(x|a, A) \mathcal{N}(x|b, B) &\propto \exp \left[ -\frac{(x-a)^2}{2A} - \frac{(x-b)^2}{2B} \right] \\ &\propto \exp \left[ -x^2 \left( \frac{1}{2A} + \frac{1}{2B} \right) + x \left( \frac{a}{A} + \frac{b}{B} \right) \right] \\ &\propto \exp \left[ -\left( \frac{1}{2A} + \frac{1}{2B} \right) \left( x - \frac{\frac{a}{A} + \frac{b}{B}}{\frac{1}{A} + \frac{1}{B}} \right)^2 \right] \exp \left[ -\frac{(a-b)^2}{2(A+B)} \right] \\ &\propto \mathcal{N} \left( x \left| \frac{\frac{a}{A} + \frac{b}{B}}{\frac{1}{A} + \frac{1}{B}}, \frac{1}{\frac{1}{A} + \frac{1}{B}} \right. \right) \mathcal{N}(0|a-b, A+B) \end{aligned} \quad (2.13)$$

常数部分:

$$\frac{1}{\sqrt{2\pi A}} \cdot \frac{1}{\sqrt{2\pi B}} = \frac{1}{2\pi \sqrt{(A+B) \frac{1}{1/A+1/B}}} = \frac{1}{\sqrt{2\pi(A+B)}} \cdot \frac{1}{\sqrt{2\pi \frac{1}{1/A+1/B}}} \quad (2.14)$$

需要说明的是，该定理对于复数情况仍适用。

### ■ 多元高斯分布

设 $N$ 维随机矢量 $\mathbf{X}$ 服从均值为 $\mathbf{a}$ ，协方差矩阵为 $\mathbf{A}$ 的矢量高斯分布，即

$$p(\mathbf{x}) = |\mathbf{A}|^{-\frac{1}{2}} (2\pi)^{-\frac{N}{2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{a})^T \mathbf{A}^{-1} (\mathbf{x} - \mathbf{a}) \right] \quad (2.15)$$

记作 $\mathbf{X} \sim \mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{A})$ 。

若为复高斯分布，则为

$$p(\mathbf{x}) = \frac{1}{|\pi \mathbf{A}|} \exp \left[ -(\mathbf{x} - \mathbf{a})^H \mathbf{A}^{-1} (\mathbf{x} - \mathbf{a}) \right] \quad (2.16)$$

记作 $\mathbf{X} \sim \mathcal{N}_c(\mathbf{x}|\mathbf{a}, \mathbf{A})$ 。

多元高斯分布的概率密度函数可以由标量高斯分布推导得到，具体如下。

证：设随机变量 $X, Y$ 独立同分布，且 $X \sim \mathcal{N}(0, 1)$ 。则 $X$ 与 $Y$ 的联合概率密度为

$$p(x, y) = p(x)p(y) = \frac{1}{2\pi} \exp(-\frac{1}{2}(x^2 + y^2)) \quad (2.17)$$

设 $\mathbf{z} = [x, y]^T$ ，则有

$$p(\mathbf{z}) = \frac{1}{2\pi} \exp(-\frac{1}{2}\mathbf{z}^T \mathbf{z}) \quad (2.18)$$

作变量替换 $\mathbf{z} = \mathbf{B}(\mathbf{x} - \mathbf{a})$ ，代入(2.18)中，该线性变量的雅克比行列式为 $J = |\mathbf{B}|$ 。

$$p(\mathbf{x}) = \frac{|\mathbf{B}|}{2\pi} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{a})^T \mathbf{B}^T \mathbf{B} (\mathbf{x} - \mathbf{a}) \right] \quad (2.19)$$

令 $\mathbf{A}^{-1} = \mathbf{B}^T \mathbf{B}$ ，则

$$p(\mathbf{x}) = \frac{1}{|\mathbf{A}|^{\frac{1}{2}} 2\pi} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{a})^T \mathbf{A}^{-1} (\mathbf{x} - \mathbf{a}) \right] \quad (2.20)$$

若 $\mathbf{x}$ 为 $N$ 维高斯变量，则

$$p(\mathbf{x}) = |\mathbf{A}|^{-\frac{1}{2}} (2\pi)^{-\frac{N}{2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{a})^T \mathbf{A}^{-1} (\mathbf{x} - \mathbf{a}) \right] \quad (2.21)$$

**引理 2-2 (矢量高斯相乘引理)：**假设 $\mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{A})$ ， $\mathcal{N}(\mathbf{x}|\mathbf{b}, \mathbf{B})$ ，则有

$$\mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{A}) \mathcal{N}(\mathbf{x}|\mathbf{b}, \mathbf{B}) = \mathcal{N}(\mathbf{0}|\mathbf{a} - \mathbf{b}, \mathbf{A} + \mathbf{B}) \mathcal{N}(\mathbf{x}|\mathbf{c}, \mathbf{C}) \quad (2.22)$$

其中 $\mathbf{C} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}$ ， $\mathbf{c} = \mathbf{C} \cdot (\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b})$ 。

证：

指数部分



$$\begin{aligned}
 \mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{A})\mathcal{N}(\mathbf{x}|\mathbf{b}, \mathbf{B}) &\propto \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{a})^T \mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}) - \frac{1}{2}(\mathbf{x} - \mathbf{b})^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{b}) \right] \\
 &\propto \exp \left[ -\frac{1}{2}\mathbf{x}^T(\mathbf{A}^{-1} + \mathbf{B}^{-1})\mathbf{x} - 2\mathbf{x}^T(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b}) \right] \\
 &\propto \mathcal{N}(\mathbf{x}|\mathbf{c}, \mathbf{C})
 \end{aligned} \tag{2.23}$$

系数部分

$$\begin{aligned}
 |\mathbf{A}||\mathbf{B}| &= |\mathbf{A} + \mathbf{B}|(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}| \\
 \Rightarrow |\mathbf{A}\mathbf{B}(\mathbf{A}^{-1} + \mathbf{B}^{-1})| &= |\mathbf{A} + \mathbf{B}| \\
 \Rightarrow |\mathbf{A} + \mathbf{B}| &= |\mathbf{A} + \mathbf{B}|
 \end{aligned} \tag{2.24}$$

需要特别说明的是该定理在复数域同样适用。

## ■ KL散度

KL散度或者称KL距离（Kullback–Leibler divergence, KL-divergence）用于刻画两个概率密度之间的距离，在信息理论中也被称为相对熵（relative entropy）<sup>[10]</sup>。从概率分布 $p(x)$ 到 $q(x)$ 的KL散度，通常定义为

$$\mathcal{D}_{\text{KL}}(p||q) = \int q(x) \log \frac{q(x)}{p(x)} dx \tag{2.25}$$

在机器学习领域， $\mathcal{D}_{\text{KL}}(p||q)$ 用于表示由 $q(x)$ 取代 $p(x)$ 所带来的信息损失。在具体应用中， $p(x)$ 表示真实分布，但该分布通常难以计算得到。为此，需要找到一个分布 $q(x)$ 来近似 $p(x)$ 。概率分布 $q(x)$ 的选取主要考虑两个方面：其一， $q(x)$ 的函数形式方便计算；其二，能够尽可能接近真实分布，或者说尽可能保持 $p(x)$ 的真实信息。此时，KL散度就用来衡量这种近似的效果，KL散度越小，近似效果越好。

特别需要说明的是KL散度不满足交换律，即 $\mathcal{D}_{\text{KL}}(p||q) \neq \mathcal{D}_{\text{KL}}(q||p)$ 。本文的后续部分会介绍到，期望传播算法及利用期望传播算法推导的算法所用的是 $\mathcal{D}_{\text{KL}}(p||q)$ ，而变分消息传递算法用的则是 $\mathcal{D}_{\text{KL}}(q||p)$ 。

## ■ PDF-to-RV引理

**引理2-3**（PDF-to-RV引理）<sup>[12]</sup>：设 $p_{\mathbf{W}}(\mathbf{w})$ 为随机矢量 $\mathbf{W} \in \mathbb{R}^p$ 的分布，以及映射 $g: \mathbb{R}^p \rightarrow \mathbb{R}$ ，则 $V = \frac{1}{\beta}g(\mathbf{W})$ 等价于

$$p_V(v) \propto \int \delta(\beta v - g(\mathbf{w})) p_{\mathbf{W}}(\mathbf{w}) d\mathbf{w} \tag{2.26}$$

证：若  $V = \frac{1}{\beta}g(\mathbf{W})$ ，则有

$$P(V \leq v) = P(g(\mathbf{W}) \leq \beta v) = \int u(\beta v - g(\mathbf{w})) p_{\mathbf{W}}(\mathbf{w}) d\mathbf{w} \quad (2.27)$$

其中  $u(\cdot)$  表示单位阶跃函数，因此

$$p_V(v) = \frac{d}{dv} \int u(\beta v - g(\mathbf{w})) p_{\mathbf{W}}(\mathbf{w}) = \beta \int \delta(\beta v - g(\mathbf{w})) p_{\mathbf{W}}(\mathbf{w}) d\mathbf{w} \quad (2.28)$$

反之，若(2.26)成立，由概率的归一性，可得

$$\beta \int_v \int_{\mathbf{w}} \delta(\beta v - g(\mathbf{w})) p_{\mathbf{W}}(\mathbf{w}) d\mathbf{w} dv = 1 \quad (2.29)$$

则有  $V = \frac{1}{\beta}g(\mathbf{W})$ 。

### ■ 中心极限定理

**定理 2-1（中心极限定理）：** 设随机变量  $X_1, \dots, X_K$  独立同分布，且都具有相同的均值和方差： $\mathbb{E}[X_k] = \mu$ ， $\text{Var}[X_k] = \sigma^2$ ，则随机变量之和的归一化变量

$$Y = \frac{\sum_{k=1}^K X_k - \mathbb{E}[\sum_{k=1}^K X_k]}{\sqrt{\text{Var}[\sum_{k=1}^K X_k]}} = \frac{\sum_{k=1}^K X_k - K\mu}{\sqrt{K}\sigma} \quad (2.30)$$

的分布函数为

$$\lim_{K \rightarrow \infty} P_Y(x) = \lim_{K \rightarrow \infty} P\left\{ \frac{\sum_{k=1}^K X_k - K\mu}{\sqrt{K}\sigma} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x) \quad (2.31)$$

即  $\lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K X_k - K\mu}{\sqrt{K}\sigma} \sim \mathcal{N}(0, 1)$ 。

证：设  $Z_k = X_k - \mu$ ，则  $Z_k (k \in [K])$  相互独立，且  $\mathbb{E}[Z_k] = 0$ ， $\text{Var}[Z_k] = \sigma^2$ ，则有

$$Y = \sum_{k=1}^K \left( \frac{Z_k}{\sqrt{K}\sigma} \right) \quad (2.32)$$

设  $Z_k$  的特征函数为  $\phi_{Z_k}(t)$

$$\phi_{Z_k}(t) = \int e^{jtz} p_{Z_k}(z) dz \quad (2.33)$$

根据特征函数的性质<sup>1</sup>有

$$\phi_{Y_n}(t) = \prod_{k=1}^K \left[ \phi_{Z_k} \left( \frac{t}{\sqrt{K}\sigma} \right) \right] \quad (2.34)$$

<sup>1</sup> 相互独立的随机变量之和的特征函数，等于每个随机变量特征函数的乘积。

对 $\phi_{Z_k}(t)$ 在 $t = 0$ 处泰勒展开, 有

$$\phi_{Z_k}(t) = \phi_{Z_k}(0) + t\phi'_{Z_k}(t)|_{t=0} + \frac{1}{2}t^2\phi''_{Z_k}|_{t=0} + o \quad (2.35)$$

其中

$$\begin{aligned} \phi_{Z_k}(0) &= \int_{-\infty}^{+\infty} p_{Z_k}(z)dz = 1 \\ \phi'_{Z_k}(t)|_{t=0} &= \left[ \int_{-\infty}^{+\infty} jze^{jtz} p_{Z_k}(z)dz \right]_{t=0} = 0 \\ \phi''_{Z_k}(t)|_{t=0} &= - \left[ \int_{-\infty}^{+\infty} z^2 e^{jtz} p_{Z_k}(z)dz \right]_{t=0} = -\sigma^2 \end{aligned} \quad (2.36)$$

故

$$\phi_{Z_k}(t) = 1 - \frac{\sigma^2}{2}t^2 + o \quad (2.37)$$

根据特征函数的性质<sup>2</sup>, 有

$$\begin{aligned} \lim_{K \rightarrow \infty} \phi_Y(t) &= \lim_{K \rightarrow \infty} \left[ 1 - \frac{1}{2K}t^2 + o \right]^K \\ &= \lim_{K \rightarrow \infty} \left( 1 - \frac{1}{2K}t^2 \right)^{\frac{2K}{t^2} \times \frac{t^2}{2}} \\ &\stackrel{(a)}{=} e^{-t^2/2} \end{aligned} \quad (2.38)$$

其中上述步骤(a)成立, 利用了极限公式 $\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = e$ 。根据特征函数的性质: 具有相同特征函数的概率分布相同。而标准正态的特征函数为:  $\phi(t) = e^{-t^2/2}$ 。从而证明 $\lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K X_k - K\mu}{\sqrt{K}\sigma} \sim \mathcal{N}(0, 1)$ 。

## 2.2 贝叶斯估计原理

对于未知参数是随机变量时, 贝叶斯估计方法相对于统计估计理论的经典方法(最小二乘、线性无偏估计、最小方差无偏估计等)加入了未知参数的先验分布去估计其特定的实现<sup>[33]</sup>。该方法基于贝叶斯公式, 因此称为贝叶斯方法。运用该方法的考量主要有两方面的因素。首先, 在未知参数的先验分布已知的前提下, 通过加入先验概率来提高估计精度; 其次, 在最小方差无偏估计器不能存在的情况下, 贝叶斯方法提供了一个可供选择的方案。

<sup>2</sup> 设  $Z = aX$ , 则  $\phi_Z(t) = \phi_X(at)$ 。

### 2.2.1 最小均方误差估计

贝叶斯方法寻找未知参数 $\mathbf{x}$ 的估计量 $\hat{\mathbf{x}}$ ，使得均方误差（mean square error, MSE）最小，定义如下

$$\text{MSE} = \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})^2] \quad (2.39)$$

其中期望对后验概率 $p(\mathbf{x}, \mathbf{y})$ 取， $\mathbf{y}$ 为观测数据。

$$\text{MSE} = \int_{\mathbf{y}} \left[ \int_{\mathbf{x}} (\mathbf{x} - \hat{\mathbf{x}})^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \right] p(\mathbf{y}) d\mathbf{y} \quad (2.40)$$

由于任意 $\mathbf{y}$ 都有 $p(\mathbf{y}) \geq 0$ ，因此仅使得括号内 $\int_{\mathbf{x}} (\mathbf{x} - \hat{\mathbf{x}})^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$ 最小即可。为了确定贝叶斯估计器 $\hat{\mathbf{x}}$ 的表达式形式，对内层积分求偏导如下

$$\frac{\partial}{\partial \hat{\mathbf{x}}} \int_{\mathbf{x}} (\mathbf{x} - \hat{\mathbf{x}})^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = 2 \int_{\mathbf{x}} (\mathbf{x} - \hat{\mathbf{x}}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (2.41)$$

并令该偏导数为0，得到

$$\hat{\mathbf{x}}_{\text{MMSE}} = \int_{\mathbf{x}} \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (2.42)$$

该估计器的形式是后验概率的均值，由于该估计器使得均值方误差最小，因此，我们称其为最小均方误差（minimum mean square error, MMSE）估计器。 $\mathbf{x}$ 的MMSE估计器由各分量的MMSE估计组成，即

$$\hat{\mathbf{x}} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_i \\ \vdots \\ \hat{x}_N \end{bmatrix} = \begin{bmatrix} \int_{x_1} x_1 p(x_1|\mathbf{y}) dx_1 \\ \vdots \\ \int_{x_i} x_i p(x_i|\mathbf{y}) dx_i \\ \vdots \\ \int_{x_N} x_N p(x_N|\mathbf{y}) dx_N \end{bmatrix} \quad (2.43)$$

其中 $p(x_i|\mathbf{y})$ 表示边缘后验概率，表示为

$$p(x_i|\mathbf{y}) = \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_N} p(\mathbf{x}|\mathbf{y}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N \quad (2.44)$$

注意，边缘后验概率 $p(x_i|\mathbf{y})$ 是由联合概率 $p(\mathbf{x}|\mathbf{y})$ 经过 $(N-1)$ 重积分得到。

为了研究一般的贝叶斯估计量，引入风险函数和代价函数（cost function）来泛化衡量误差的标准。风险函数定义为

$$\mathcal{R} \triangleq \mathbb{E}[\mathcal{C}(\epsilon)] = \int \mathcal{C}(\epsilon) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (2.45)$$

其中， $\epsilon = \mathbf{x} - \hat{\mathbf{x}}$ ，函数 $\mathcal{C}(\epsilon)$ 称为代价函数。

代价函数的选择，需要从数学角度尽可能简单出发。通常考虑二次误差形、绝对误

差、成功失败误差三种情况，如图2-1所示。显然，最小均方误差所选取的代价函数为二次误差型（Quadratic error）。

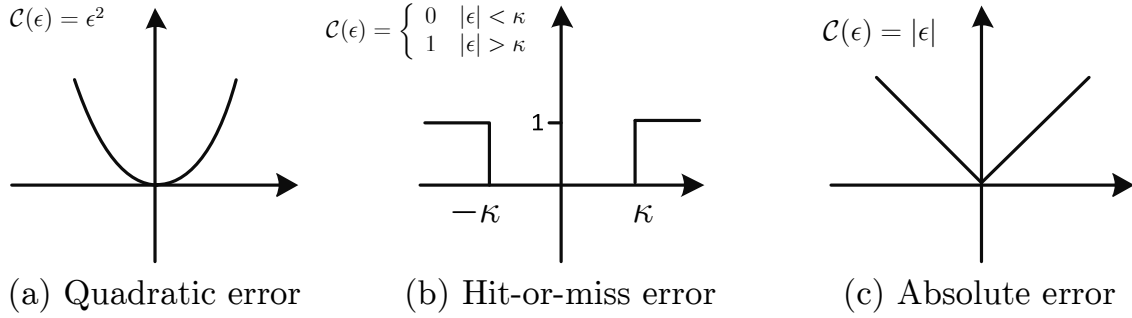


图2-1 常用代价函数

### 2.2.2 最大后验概率估计

当选择代价函数为成功-失败型（Hit-or-miss error）函数时，如图2-1（b），此时风险函数为

$$\mathcal{R} = \mathbb{E}[\mathcal{C}(\epsilon)] = \int_{\mathbf{y}} \left[ \int_{\mathbf{x}} \mathcal{C}(\epsilon) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \right] d\mathbf{y} \quad (2.46)$$

同理，当内层积分最小时，风险函数最小。定义内层积分为

$$g(\epsilon) \triangleq \int_{\mathbf{x}} \mathcal{C}(\epsilon) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (2.47)$$

对于成功-失败型函数，有

$$\begin{aligned} g(\epsilon) &= \lim_{\kappa \rightarrow 0} \left[ \int_{\hat{\mathbf{x}}+\kappa}^{+\infty} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} + \int_{\hat{\mathbf{x}}-\kappa}^{-\infty} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \right] \\ &= 1 - \lim_{\kappa \rightarrow 0} \int_{\hat{\mathbf{x}}-\kappa}^{\hat{\mathbf{x}}+\kappa} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \end{aligned} \quad (2.48)$$

为使得 $g(\epsilon)$ 最大，需使得 $\lim_{\kappa \rightarrow 0} \int_{\hat{\mathbf{x}}-\kappa}^{\hat{\mathbf{x}}+\kappa} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$ 最大，因此选择使得 $p(\mathbf{x}|\mathbf{y})$ 最大的 $\mathbf{x}$ 作为 $\mathbf{x}$ 的估计器。由于估计器 $\hat{\mathbf{x}}$ 选取使得后验概率最大时的 $\mathbf{x}$ 作为估计值，我们称这种估计器为最大后验概率（maximum a posterior, MAP）估计。

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \quad (2.49)$$

需要注意的是，不管是MMSE估计还是MAP估计，其前提都是需要计算后验概率。

当选取绝对值误差（absolute error）代价函数时，如图2-1（c）。其风险函数为

$$\mathcal{R} = \int_{\mathbf{y}} \left[ \int_{\mathbf{x}} |\mathbf{x} - \hat{\mathbf{x}}| p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \right] d\mathbf{y} \quad (2.50)$$

该贝叶斯风险函数的内层积分对 $\hat{x}$ 求偏导,有

$$\begin{aligned}\frac{\partial}{\partial \hat{x}} g(\epsilon) &= -\frac{\partial}{\partial \hat{x}} \int_{-\infty}^{\hat{x}} (x - \hat{x}) p(x|y) dx + \frac{\partial}{\partial \hat{x}} \int_{\hat{x}}^{+\infty} (x - \hat{x}) p(x|y) dx \\ &= \int_{-\infty}^{\hat{x}} p(x|y) dx - \int_{\hat{x}}^{+\infty} p(x|y) dx\end{aligned}\quad (2.51)$$

通过令偏导为0, 得

$$\int_{-\infty}^{\hat{x}} p(x|y) dx = \int_{\hat{x}}^{+\infty} p(x|y) dx \quad (2.52)$$

即 $\hat{x}$ 为后验概率 $p(x|y)$ 的中值点, 即 $P[x \leq \hat{x}] = 1/2$ 。

Remark:

- 贝叶斯估计器选取代价函数为二次误差型、成功-失败型、绝对值误差型分别对应后验概率的均值点、最大值点、中值点。
- 通常, 后验概率的中值点很难计算得到。因此, 常用的贝叶斯估计器主要为最小均方误差估计和最大后验概率估计。

## 2.3 因子图

本章的前面两节中, 我们介绍了两种常用的贝叶斯估计方法: 最小均方误差估计、最大后验概率估计。两种估计器都是基于边缘后验概率的估计, 而求边缘后验概率需要进行 $N - 1$ 重的积分, 其计算量将随着系统规模的增大呈指数型增加, 在处理大规模数据中, 其计算量是在灾难性的。因此, 最为迫切的问题是如何高效地计算边缘后验概率。幸运地是, 若联合概率可以因式化, 通过用因子图来表示联合概率, 此时, 边缘后验概率可以通过因子图的消息传递算法计算得到<sup>[20]</sup>。

### 2.3.1 因子图表示

我们通过一个简单的例子来介绍因子图 (factor graph, FG) 及其消息传递 (message passing, MP) 算法。给定联合概率函数 $p(x_1, x_2, x_3, x_4)$ 具有如下的因式可分解结构

$$p(x_1, x_2, x_3, x_4) = p_1(x_1, x_2) p_2(x_1, x_3, x_4) p_3(x_3) \quad (2.53)$$

该联合概率密度的因子图表示, 见图2-2。其中, 圆圈表示变量节点 (variable node), 方框表示因子节点 (factor node), 也称函数节点 (function node)。与因子节点相连的变量节点为函数的自变量, 如图2-2中红色框, 与因子节点 $p_1$ 相连接的是变量节点 $x_1$ 和

$x_2$ ，因此其所表示的就是 $p_1(x_1, x_2)$ 。各因子节点所组成的函数相乘得到联合概率密度函数。

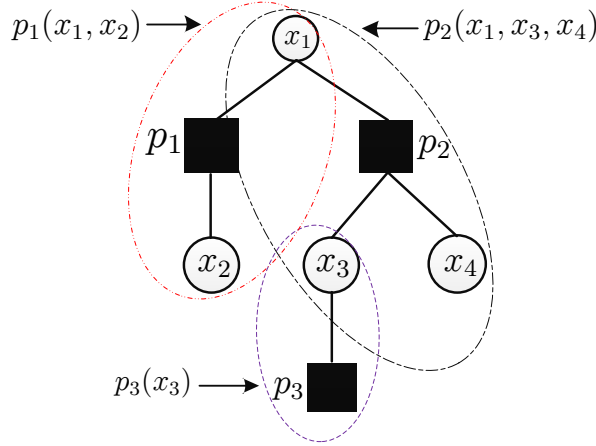


图2-2 联合概率的因子图表示

### 2.3.2 消息传递算法

联合概率密度 $p(x_1, x_2, x_3, x_4)$ 的边缘概率密度 $p(x_1)$ 可以表示如下

$$p(x_1) = \int p_1(x_1, x_2) p_2(x_1, x_3, x_4) p_3(x_3) dx_2 dx_3 dx_4 \quad (2.54)$$

为了计算该边缘概率密度，定义如图2-3的消息更新规则<sup>[20]</sup>。

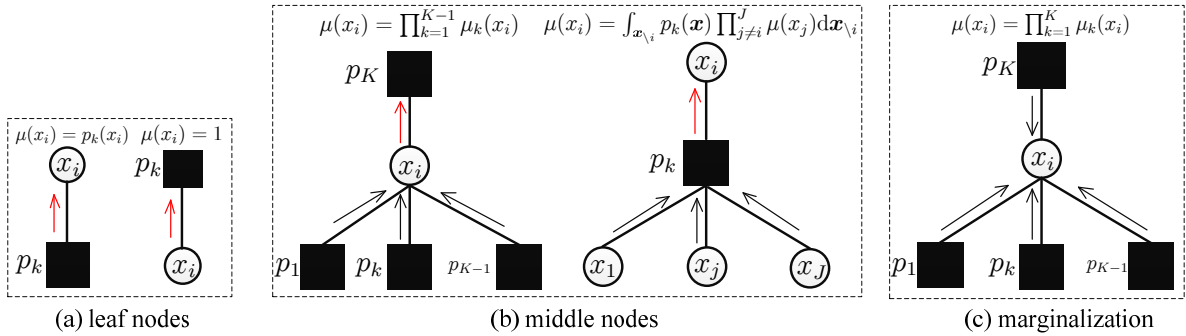


图2-3 因子图消息更新规则

- **初始化：**消息的传递从叶子节点（仅有一条边连接的节点）开始。叶子节点为因子节点，所传递的消息为 $p_k(x_i)$ ，如图2-3（a）左边；叶子节点为变量节点，所传递的消息为1，如图2-3（a）图右边。
- **中间节点：**若中间节点为变量节点 $x_i$ ，则变量节点 $x_i$ 向因子节点 $p_K$ 所传递的消息为

$$\mu(x_i) = \prod_{k=1}^{K-1} \mu_k(x_i) \quad (2.55)$$

若中间节点为因子节点，则因子节点 $p_k$ 向变量节点 $x_i$ 所传递的消息为

$$\mu(x_i) = \int_{\mathbf{x}_{\setminus i}} p_k(x_1, \dots, x_J) \prod_{j \neq i} \mu(x_j) d\mathbf{x}_{\setminus i} \quad (2.56)$$

其中， $\mathbf{x}_{\setminus i}$ 表示除第 $i$ 元素外， $\mathbf{x}$ 其余所有元素所组成的向量。

■ **边缘化：** $x_i$ 的边缘概率表示为，与变量节点 $x_i$ 相连接的所有因子节点向该变量节点传递的消息的乘积，如图2-3 (c)，即

$$\mu(x_i) = \prod_{k=1}^K \mu_k(x_i) \quad (2.57)$$

为了验证图2-3所定义的更新规则是否能够求解边缘概率，我们进行如下验证。考虑因子化联合概率密度函数(2.54)。根据图2-4，很容易验证边缘后验概率可以通过消息传递算法得到。

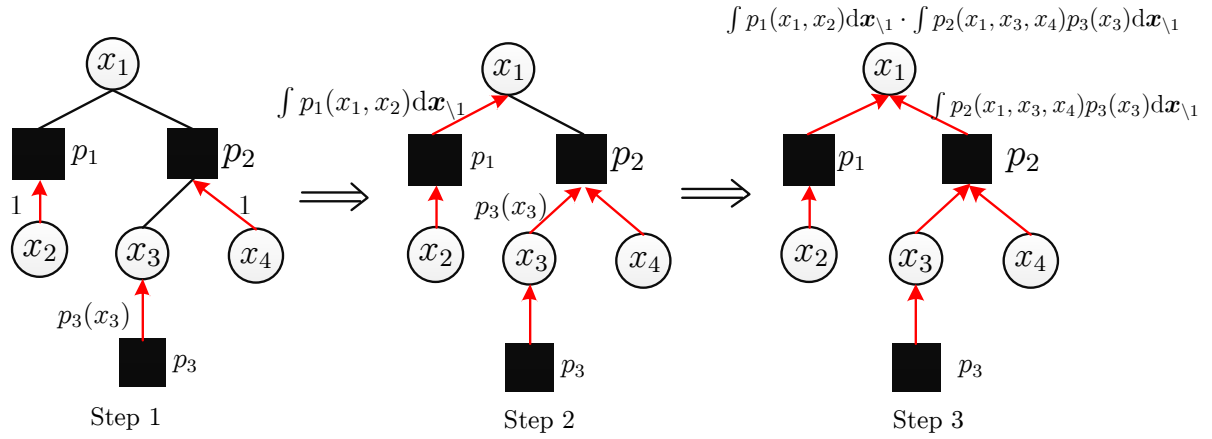


图2-4 边缘概率密度更新步骤



### 第三章 标准线性逆问题的快速解法

在上一章中，我们介绍了贝叶斯估计方法、因子图及其消息更新规则等知识点。本章主要研究标准线性逆问题，针对该问题介绍两种基于因子图（Factor Graph, FG）的快速重构算法，分别是近似消息传递算法（approximate message passing, AMP）和矢量近似消息传递算法（vector approximate message passing, VAMP）。本章的主要贡献在于

- 提出了一种基于分列式因子图的AMP推导方法。分列式因子图相对于传统因子图而言，具有可扩展到多层模型的优势。
- 提出了一种新的消息更新规则：矩匹配消息更新规则，并用于推导VAMP算法。此外该消息更新规则与分列式因子图对于多层模型同样适用。

本章的部分内容，已整理成论文，已向IEEE Transactions on Signal Processing期刊投稿，详见成果列表。

#### 3.1 标准线性逆问题

如图3-1所示，线性逆问题研究如下模型

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (3.1)$$

其中 $\mathbf{y} \in \mathbb{R}^M$ 是观测矢量， $\mathbf{x} \in \mathbb{R}^N$ 表示原始信号， $\mathbf{w} \in \mathbb{R}^M$ 为观测噪声矢量，以及 $\mathbf{H} \in \mathbb{R}^{M \times N}$ 是测量矩阵（Measurement Matrix）或观测矩阵。在压缩感知（Compressive Sense, CS）领域，其所研究的是 $M < N$ 的情况下目标信号的重构，即欠定方程组的求解。估计器需要从少量观测数据中重构出目标信号。这种欠定方程对观测矩阵 $\mathbf{H}$ 以及原始信号 $\mathbf{x}$ 都有要求。当观测矩阵 $\mathbf{H}$ 满足有限等距原理（restricted isometry property, RIP）时可以大概率重构出稀疏信号 $\mathbf{x}$ 。在其他领域，如通信领域，通常观测向量的长度大于目标信号长度，即 $M > N$ ，为超定方程组。

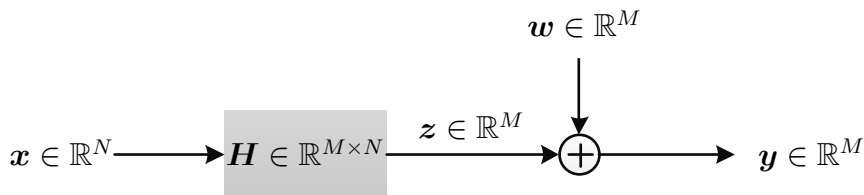


图3-1 标准线性可逆模型

在众多信号重构算法中，在缺少先验知识时，最大似然估计（Maximum Likelihood, ML）精度最高，该估计器的方差可以达到克拉美-罗下限（Cramer-Rao Lower Bound, CRLB）<sup>[33]</sup>。然而，最大似然估计的计算复杂度为 $\mathcal{O}(|\mathcal{S}|^N)$ 。显然，随着矩阵维度的增加，最大似然估计器的计算量将呈指数型增加。

$$\hat{\mathbf{x}}_{\text{ML}} = \arg \max_{\mathbf{x} \in \mathcal{S}} p(\mathbf{y}|\mathbf{x}) \quad (3.2)$$

为了降低最大似然估计的计算量并且在性能上尽可能逼近最大似然估计，目前已经存在很多基于最大似然估计或第二类最大似然估计的算法，如稀疏贝叶斯学习<sup>[7]</sup>、基于AMP的稀疏贝叶斯学习<sup>[34]</sup>，以及EM-稀疏贝叶斯学习<sup>[11]</sup>等。该类算法在先验信息未知时有很大优势，而对于先验分布已知时，贝叶斯算法有着巨大优势。

近年来，贝叶斯推论逐渐受到学者的广泛关注。贝叶斯估计根据代价函数的不同主要可以分为最大后验概率估计（Maximum a posterior, MAP）和最小均方误差估计（Minimum Mean Square Error, MMSE），其中最大后验概率估计的代价函数为“成功-失败”（hit-or-miss）型代价函数，而最小均方误差估计的代价函数为二次函数。因此，最大后验概率估计为后验概率的最大值点，而最小均方误差估计为后验概率的均值点。特别地，先验概率为高斯分布时两种估计器性能一致。

无论是最小均方误差估计还是最大后验概率估计，都需要计算后验概率。最小均方误差估计是选择后验概率的均值作为估计器

$$\hat{x}_i = \mathbb{E}[x_i|\mathbf{y}] \quad (3.3)$$

上式的期望是对边缘后验概率 $p(x_i|\mathbf{y})$ 取。进一步地，边缘后验概率表示为

$$p(x_i|\mathbf{y}) = \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_N} p(\mathbf{x}|\mathbf{y}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N \quad (3.4)$$

其中 $p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$ 。

最小均方误差估计在推导上相较于最大后验概率更符合代数体系，因此，本文主要考虑最小均方误差估计。在本文所介绍的算法中，最小均方误差估计与最大后验概率估计的区别主要体现在具体算法的输出函数（output function）上。进一步地，根据目标信号的先验分布，最小均方误差估计可以具体分为以下两种情况

- 目标信号服从高斯分布，即 $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{xx})$ 。已知观测信号 $\mathbf{y}$ ，测量矩阵 $\mathbf{H}$ 以及噪声分布 $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ 。这种情况下，存在着最小均方误差估计的闭式解（close-form solution）。通过高斯相乘引理（引理2-2），很容易得到MMSE估计器

$$\hat{\mathbf{x}}_{\text{MMSE}} = (\mathbf{C}_{xx}^{-1} + \sigma^{-2} \mathbf{H}^T \mathbf{H})^{-1} (\sigma^{-2} \mathbf{H}^T \mathbf{y}) \quad (3.5)$$

- 目标信号的先验分布不是高斯分布。这种情况，目标信号的MMSE估计器需要分别计算每一个分量的MMSE估计  $\hat{x}_i = \mathbb{E}[x_i | \mathbf{y}]$ ，这里期望是对边缘后验概率  $p(x_i | \mathbf{y})$  取，而边缘后验概率的计算需要经过  $(N - 1)$  的积分

$$p(x_i | \mathbf{y}) \propto \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_N} p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N \quad (3.6)$$

注意，若  $\mathbf{x}$  为离散分布时，边缘后验概率通过对联合概率的求和得到，而求和可以通过  $\delta$  积分来表示。

## 3.2 传统估计方法

如(3.6)所示，当先验概率为非高斯分布时边缘后验概率的计算需要通过  $(N - 1)$  重积分的计算。当系统规模很大时该计算量呈指数型增加。因此，传统的方法并不能以低复杂度的代价达到贝叶斯最优。为此，传统的估计考虑次优估计，如线性最小均方误差估计、最小二乘估计。次优估计牺牲了计算精度用以降低计算复杂度。注意，最小二乘估计并不属于贝叶斯方法的范畴。

### 3.2.1 线性最小均方误差估计器

线性最小均方误差（linear minimum mean square error, LMMSE）估计限定估计器模型为观测值  $\mathbf{y}$  的线性模型，即

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \mathbf{A} \mathbf{y} + \mathbf{b} \quad (3.7)$$

为了得到系统模型的LMMSE估计器的具体表达形式，我们需要进一步确定参数  $(\mathbf{A}, \mathbf{b})$ 。而参数的求解，是通过最小化如下均方误差实现

$$\text{MSE} = \mathbb{E}[(\hat{\mathbf{x}}_{\text{LMMSE}} - \mathbf{x})^2] \quad (3.8)$$

其中期望对联合概率分布  $p(\mathbf{x}, \mathbf{y})$  取。

如(3.7)所示，LMMSE估计器的结构如下

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \begin{pmatrix} \mathbf{A}_{1,:} \mathbf{y} + b_1 \\ \vdots \\ \mathbf{A}_{i,:} \mathbf{y} + b_i \\ \vdots \\ \mathbf{A}_{N,:} \mathbf{y} + b_N \end{pmatrix} \quad (3.9)$$

其中  $\mathbf{A}_{i,:}$  表示  $\mathbf{A}$  的第  $i$  行元素所组成的行向量。因此，为了求解参数  $(\mathbf{A}, \mathbf{b})$ ，我们需要分

别确定参数 $(\mathbf{A}_{i,:}, b_i), i = 1, \dots, N$ 。而参数对 $(\mathbf{A}_{i,:}, b_i)$ 的求解，可以抽象成问题：

$$\hat{x} = \mathbf{a}^T \mathbf{y} + b \quad (3.10)$$

其中 $\mathbf{a}^T$ 表示 $\mathbf{A}$ 中任意行向量。

计算均方误差函数对 $b$ 的偏导，并令导数为0，有

$$\frac{\partial}{\partial b} \mathbb{E} [(x - \mathbf{a}^T \mathbf{y} - b)^2] = -2\mathbb{E} [x - \mathbf{a}^T \mathbf{y} - b] = 0 \Rightarrow b = \mathbb{E}[x] - \mathbf{a}^T \mathbb{E}[\mathbf{y}] \quad (3.11)$$

将(3.11)代入MSE(3.8)中

$$\begin{aligned} \text{MSE} &= \mathbb{E} [(x - \mathbf{a}^T \mathbf{y} - \mathbb{E}[x] + \mathbf{a}^T \mathbb{E}[\mathbf{y}])^2] \\ &= \mathbb{E} [(\mathbf{a}^T (\mathbf{y} - \mathbb{E}[\mathbf{y}]) - (x - \mathbb{E}[x]))^2] \\ &= \mathbf{a}^T \mathbf{C}_{\mathbf{y}\mathbf{y}} \mathbf{a} - \mathbf{a}^T \mathbf{C}_{\mathbf{y}x} - \mathbf{C}_{x\mathbf{y}} \mathbf{a} + C_{xx} \end{aligned} \quad (3.12)$$

其中 $\mathbf{C}_{\mathbf{y}\mathbf{y}}$ 表示 $\mathbf{y}$ 的协方差矩阵， $\mathbf{C}_{x\mathbf{y}}$ 是 $1 \times M$ 互协方差矢量，且 $\mathbf{C}_{x\mathbf{y}} = \mathbf{C}_{\mathbf{y}x}^T$ ， $C_{xx}$ 为 $x$ 的方差。

计算均方误差对 $\mathbf{a}$ 的偏导，并令偏导数为0，有

$$\frac{\partial \text{MSE}}{\partial \mathbf{a}} = 2\mathbf{C}_{\mathbf{y}\mathbf{y}} \mathbf{a} - 2\mathbf{C}_{\mathbf{y}x} = 0 \Rightarrow \mathbf{a} = \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{C}_{\mathbf{y}x} \quad (3.13)$$

联合(3.11)、(3.13)得到LMMSE估计器模型

$$\begin{aligned} \hat{x} &= \mathbf{C}_{x\mathbf{y}} + \mathbb{E}[x] - \mathbf{C}_{x\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} \mathbb{E}[\mathbf{y}] \\ &= \mathbb{E}[x] + \mathbf{C}_{x\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{y} - \mathbb{E}[\mathbf{y}]) \end{aligned} \quad (3.14)$$

根据(3.14)，即 $\mathbf{x}$ 的LMMSE估计器的元素表达式，可构造得到 $\mathbf{x}$ 的LMMSE估计器

$$\begin{aligned} \hat{\mathbf{x}} &= \begin{pmatrix} \mathbb{E}[x_1] \\ \mathbb{E}[x_2] \\ \vdots \\ \mathbb{E}[x_N] \end{pmatrix} + \begin{pmatrix} \mathbf{C}_{x_1\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{y} - \mathbb{E}[\mathbf{y}]) \\ \mathbf{C}_{x_2\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{y} - \mathbb{E}[\mathbf{y}]) \\ \vdots \\ \mathbf{C}_{x_N\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{y} - \mathbb{E}[\mathbf{y}]) \end{pmatrix} \\ &= \mathbb{E}[\mathbf{x}] + \mathbf{C}_{x\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{y} - \mathbb{E}[\mathbf{y}]) \end{aligned} \quad (3.15)$$

其中

$$\begin{aligned} \mathbf{C}_{\mathbf{y}\mathbf{y}} &= \mathbf{H} \mathbf{C}_{xx} \mathbf{H}^T + \mathbf{C}_w \\ \mathbf{C}_{x\mathbf{y}} &= \mathbf{C}_{xx} \mathbf{H}^T \end{aligned} \quad (3.16)$$

从而得到

$$\begin{aligned} \hat{\mathbf{x}}_{\text{LMMSE}} &= \mathbb{E}[\mathbf{x}] + \mathbf{C}_{xx} \mathbf{H}^T (\mathbf{H} \mathbf{C}_{xx} \mathbf{H}^T + \mathbf{C}_w)^{-1} (\mathbf{y} - \mathbf{H} \mathbb{E}[\mathbf{x}]) \\ &= \mathbb{E}[\mathbf{x}] + (\mathbf{C}_{xx}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{y} - \mathbf{H} \mathbb{E}[\mathbf{x}]) \end{aligned} \quad (3.17)$$

标准线性逆问题中，原始信号通常考虑零均值，即 $\mathbb{E}[\mathbf{x}] = \mathbf{0}$ 。因此，得到

$$\hat{\mathbf{x}}_{\text{LMMSE}} = (\mathbf{C}_{xx}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{y} \quad (3.18)$$

值得注意的是，通过详细推导过程得到的线性最小均方误差估计器表达式，与强

制设定先验分布为高斯分布 $p(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \mathbf{C}_{xx})$ 所得到的结果一致，注意这里 $\mathbb{E}[\mathbf{x}] = \mathbf{0}$ 。根据高斯相乘引理，很容易得到线性最小均方误差的表达式

$$\hat{\mathbf{x}}_{\text{LMMSE}} = (\mathbf{C}_{xx}^{-1} + \sigma_w^{-2} \mathbf{H}^T \mathbf{H})^{-1} (\sigma_w^{-2} \mathbf{H}^T \mathbf{y}) \quad (3.19)$$

这里，考虑噪声独立同分布的情况，即 $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I})$ 。

特别地，当原始信号 $\mathbf{x}$ 满足元素功率归一化条件时，则对应的估计器表达式变为

$$\hat{\mathbf{x}}_{\text{LMMSE}} = (\sigma_w^2 \mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} \quad (3.20)$$

Remark:

- LMMSE的实质是忽略先验概率 $p(\mathbf{x})$ 的真实分布，通过构造先验分布为高斯分布 $p(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \text{Var}[\mathbf{x}])$ ，所得到的MMSE估计。因此，当真实的先验分布为高斯时，LMMSE与MMSE是等价的。
- 由于改变了先验分布 $p(\mathbf{x})$ 的形式，因此必然存在信息的忽略，而所忽略的信息则具体体现在两者性能上的差异。特别地，当先验分布 $p(\mathbf{x})$ 服从高斯分布时，LMMSE与MMSE一致。

### 3.2.2 最小二乘估计器

最小二乘估计器考虑损失函数

$$J = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \quad (3.21)$$

由于该损失函数为凸函数[32]，因此通过计算损失函数对 $\mathbf{x}$ 偏导，有

$$\frac{\partial J}{\partial \mathbf{x}} = -2\mathbf{H}^T \mathbf{y} + 2\mathbf{H}^T \mathbf{H} \mathbf{x} \quad (3.22)$$

通过令导数为0，可得该模型的最小二乘估计器

$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} \quad (3.23)$$

通过观察(3.20)与(3.23)，可以发现：最小二乘估计器相对于最小均方误差而言，忽略了噪声的影响，因此最小二乘估计器对噪声非常敏感。换言之，当信噪比很小时，最小二乘估计器的性能相对最小均方误差估计器会有明显的劣势。当信噪比很大时，两者性能会逐渐重合。

此外，最小二乘估计还可以从几何学角度理解[33]。如图3-2所示，由 $\mathbf{H}$ 张成的超平面用 $\mathcal{C}$ 表示，最小化损失函数 $J = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$ ，就是要找到观测向量 $\mathbf{y}$ 在超平面上的正交投影。

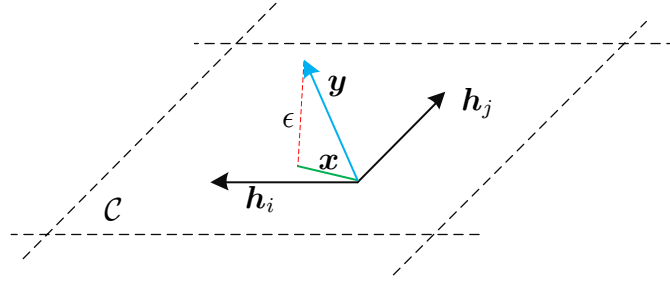


图3-2 LS的几何解释

### 3.3 近似消息传递算法

#### 3.3.1 算法推导

实际应用中，通常目标信号的先验分布 $p(\mathbf{x})$ 是非高斯的（例如，QPSK星座点）。如公式(3.6)所述，当先验分布 $p(\mathbf{x})$ 为非高斯分布时，计算边缘后验概率需要对联合分布 $p(\mathbf{x}, \mathbf{y})$ 进行 $(N - 1)$ 重的积分。随着系统维度的增加，这个计算量是灾难性的。幸运的是，相对于传统估计方法，AMP可以用较低的计算复杂度来实现贝叶斯最优。更确切的说，AMP的计算复杂度为 $\mathcal{O}(T(M + N))$ ，详见3.4节。如图3-3所示，AMP源于基于因子图的Sum-product算法。首先是利用因子图来表示联合分布 $p(\mathbf{x}, \mathbf{y})$ ，这样边缘后验概率 $p(x_i|\mathbf{y})$ 可以通过置信传播（belief propagation, BP）算法迭代得到，也称消息传递（message passing, MP）算法。当变量个数 $N \rightarrow \infty$ 时，基于中心极限定理（central limit theorem, CLT）对部分消息用高斯分布替换，得到松弛置信传播（relaxed belief propagation, RBP）算法。最后经过TAP-lize步骤，忽略无穷小项，得到AMP算法。



图3-3 从解析式到AMP算法的推导结构

首先假设联合概率 $p(\mathbf{x}, \mathbf{y})$ 可以完全因子化，即

$$p(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^N p(x_i) \prod_{a=1}^M p(y_a|\mathbf{x}) \quad (3.24)$$

该联合概率用因子图表示如图3-4（a）。通过该因子图，边缘后验概率可以通过消息传递算法进行计算。

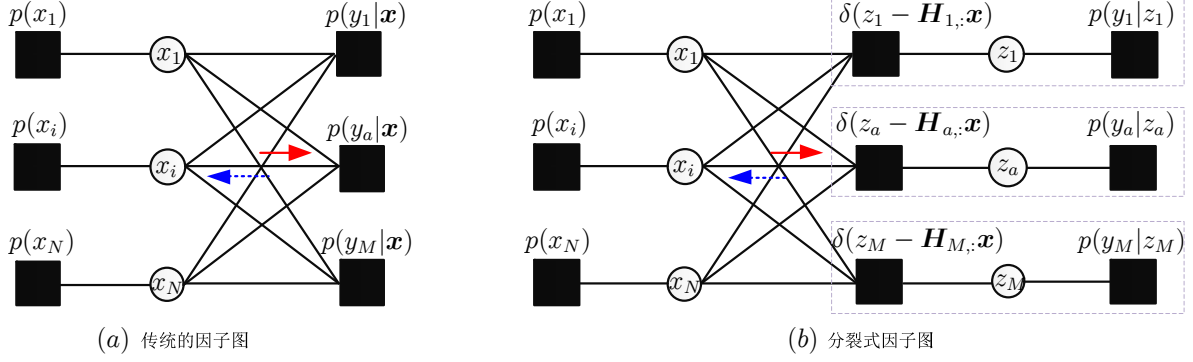


图3-4 标准线性模型的因子图

该模型的消息传递（message passing, MP）算法表示为

$$\begin{aligned}\mu_{i \rightarrow a}^{t+1}(x_i) &= \frac{1}{C_{i \rightarrow a}^{t+1}} p(x_i) \prod_{b \neq a} \mu_{i \leftarrow b}^t(x_i) \\ \mu_{i \leftarrow a}^t(x_i) &= \frac{1}{C_{i \leftarrow a}^t} \int p(y_a | \mathbf{x}) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) d\mathbf{x}_{\setminus i}\end{aligned}\quad (3.25)$$

其中  $\mu_{i \leftarrow a}^t(x_i)$ ，表示第  $t$  迭代，从第  $a$  个因子节点到第  $i$  个变量节点的消息，如图3-4中蓝色虚线箭头所示  $((i, a) \in [N] \times [M])$ ； $\mu_{i \rightarrow a}^{t+1}(x_i)$  表示第  $t+1$  次迭代从第  $i$  个变量节点到第  $a$  个因子节点的消息，如图3-4中红色实线箭头所示。此外， $C_{i \leftarrow a}^t$  和  $C_{i \rightarrow a}^{t+1}$  为归一化常数。因子图消息的传递是通过因子节点与变量节点之间的边（edge）进行的。最终，边缘概率  $p(x_i)$  可以表示为

$$p(x_i) = \mu_{i \leftarrow a}^t(x_i) \cdot \mu_{i \rightarrow a}^t(x_i) \quad (3.26)$$

对似然函数做进一步地拆分

$$p(y_a | \mathbf{x}) = \int_{z_a} p(y_a | z_a) \delta(z_a - \mathbf{H}_{a,:} \mathbf{x}) dz_a \quad (3.27)$$

其中， $\mathbf{H}_{a,:}$  表示矩阵  $\mathbf{H}$  第  $a$  行元素所组成的向量。相应地，我们将(3.25)重写为

$$\begin{aligned}\mu_{i \rightarrow a}^{t+1}(x_i) &= \frac{1}{C_{i \rightarrow a}^{t+1}} p(x_i) \prod_{b \neq a} \mu_{i \leftarrow b}^t(x_i) \\ \mu_{i \leftarrow a}^t(x_i) &= \frac{1}{C_{i \leftarrow a}^t} \int_{\mathbf{x}_{\setminus i}} \int_{z_a} \delta(z_a - \mathbf{H}_{a,:} \mathbf{x}) p(y_a | z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) dz_a d\mathbf{x}_{\setminus i}\end{aligned}\quad (3.28)$$

我们发现，该消息传递算法所对应的因子图可以表示为图3-4（b）。注意，由于图3-4（a）表示的联合概率为  $p(\mathbf{x}, \mathbf{y})$ ，而图3-4（b）表示的联合概率为  $p(\mathbf{x}) \delta(\mathbf{z} - \mathbf{H} \mathbf{x}) p(\mathbf{y} | \mathbf{z})$ ，因此两种因子图是不等价的。但是边缘分布  $p(x_i | \mathbf{y})$  都可以通过两种因子图的消息更新得到。本节后续推导AMP算法，都是基于图3-4（b）的分列式因子图。

消息传递算法，通过用图的形式来表示联合概率，进而通过消息更新规则迭代得到

联合概率的边缘化。然而，消息的更新（传递）是以概率密度的形式，而刻画一个概率密度需要其各阶距（充分统计量），也就是说 $\mu_{i \leftarrow a}^t(x_i)$ 和 $\mu_{i \rightarrow a}^{t+1}(x_i)$ 的计算存在着很大的计算量，尤其在大维度系统中。

当务之急是找到一类分布去近似 $\mu_{i \leftarrow a}^t(x_i)$ 和 $\mu_{i \rightarrow a}^{t+1}(x_i)$ ，这类分布主要需要满足两个条件：其一，仅仅需要有限个参数就可以确定分布的具体表达式；其二，消息迭代计算方便。幸运地是，高斯概率密度就是这样一类分布。高斯分布的确定仅仅需要均值和方差两个参数，此外高斯概率分布的计算相对简单。

从因子图出发，最终得到AMP算法，整个流程，见图3-3。

**步骤1：**消息的高斯近似。我们首先从化简 $\mu_{i \leftarrow a}^t(x_i)$ 开始。

$$\begin{aligned} \mu_{i \leftarrow a}^t(x_i) &\propto \int_{\mathbf{x}_{\setminus i}} \int_{z_a} \delta(z_a - \sum_{j \neq i} H_{aj} x_j - H_{ai} x_i) p(y_a | z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) dz_a d\mathbf{x}_{\setminus i} \\ &\stackrel{(a)}{\propto} \int_{\mathbf{x}_{\setminus i}} \int_{z_a} \delta \left[ x_i - \frac{1}{H_{ai}} (z_a - \sum_{j \neq i} H_{aj} x_j) \right] p_{\zeta_a}(z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) dz_a d\mathbf{x}_{\setminus i} \end{aligned} \quad (3.29)$$

其中(a)成立，根据Dirac函数的性质 $\delta(at) = \frac{1}{|a|} \delta(t)$ 以及定义 $p_{\zeta_a}(z_a) \triangleq \frac{p_{y_a|z_a}(y_a, z_a)}{\int p_{y_a|z_a}(y_a, z_a) dz_a}$ ，这里 $p_{y_a|z_a}(y_a, z_a) = p(y_a | z_a)$ 。显然， $\int p_{y_a|z_a}(y_a, z_a) dz_a$ 是关于 $y_a$ 的函数，由于 $y_a$ 是确定的观测值，因此 $\int p_{y_a|z_a}(y_a, z_a) dz_a$ 是一个常数。

利用PDF-to-RV引理<sup>[12]</sup>计算 $\mu_{i \leftarrow a}^t(x_i)$ 。首先，定义随机变量 $\xi_{j \rightarrow a}^t \sim \mu_{j \rightarrow a}^t(x_j)$ ， $\zeta_a \sim p_{\zeta_a}(\zeta_a)$ ，以及 $\xi_{i \leftarrow a}^t \sim \mu_{i \leftarrow a}^t(x_i)$ 。此外，定义随机矢量 $\mathbf{X} \triangleq \{\zeta_a, \{\xi_{j \rightarrow a}^t\}_{j \neq i}\}$ ，则该随机矢量的概率密度表示为

$$p_{\mathbf{X}}(z_a, \{x_j\}_{j \neq i}) = p_{\zeta_a}(z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) \quad (3.30)$$

定义函数 $g(z_a, \{x_j\}_{j \neq i}) \triangleq z_a - \sum_{j \neq i} H_{aj} x_j$ ，以及 $\beta = \frac{1}{H_{ai}}$ 。应用PDF-to-RV引理（见引理2-3），可得

$$\xi_{i \leftarrow a}^t = \frac{1}{H_{ai}} \left( \zeta_a - \sum_{j \neq i} H_{aj} \xi_{j \rightarrow a}^t \right) \quad (3.31)$$

根据中心极限定理，当 $N \rightarrow \infty$ ，有 $\sum_{j \neq i} \xi_{j \rightarrow a}^t \rightarrow \mathcal{N}(\sum_{j \neq i} H_{aj} m_{j \rightarrow a}^t, \sum_{j \neq i} H_{aj}^2 v_{j \rightarrow a}^t)$ ，其中 $m_{j \rightarrow a}^t$ 和 $v_{j \rightarrow a}^t$ 分别表示 $\mu_{j \rightarrow a}^t(x_j)$ 的均值和方差。此外， $\zeta_a \sim p_{\zeta}(z_a)$ 是均值为 $y_a$ ，方差为 $\sigma^2$ 的高斯随机变量。因此， $\mu_{i \leftarrow a}^t(x_i)$ 可以通过高斯分布 $\mathcal{N}(x_i | m_{i \leftarrow a}^t, v_{i \leftarrow a}^t)$ 来近似，其中



$$\begin{aligned}
 m_{i \leftarrow a}^t &= \frac{1}{H_{ai}} \left( y_a - \sum_{j \neq i} H_{aj} m_{j \rightarrow a}^t \right) \\
 v_{i \leftarrow a}^t &= \frac{1}{H_{ai}^2} \left( \sigma^2 + \sum_{j \neq i} H_{aj}^2 v_{j \rightarrow a}^t \right)
 \end{aligned} \tag{3.32}$$

此外，通过图3-4，可以更好的理解步骤1。图3-4中红色线段表示 $\mu_{i \leftarrow a}^{t+1}(x_i)$ ，为需要简化的消息。通过PDF-to-RV引理，将概率密度的运算转换到随机变量的运算，当变量个数 $N \rightarrow \infty$ ，即系统规模足够大时，通过中心极限定理，利用高斯概率密度来近似 $\mu_{i \leftarrow a}(x_i)$ 。

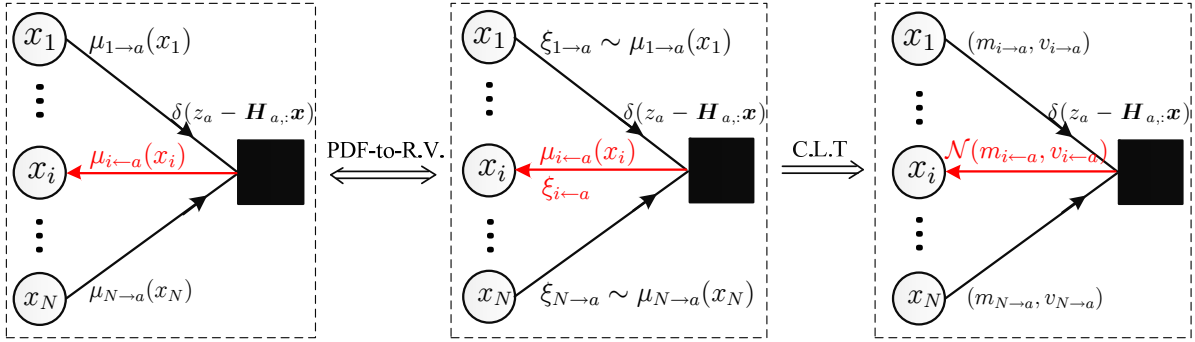


图3-4  $\mu_{i \leftarrow a}(x_i)$  的化简流程

**步骤2:** 简化 $\mu_{i \rightarrow a}^{t+1}(x_i)$ 。由于用高斯分布 $\mathcal{N}(x_i | m_{i \leftarrow a}^t, v_{i \leftarrow a}^t)$ 来近似 $\mu_{i \leftarrow a}^t(x_i)$ ，可以进一步简化 $\mu_{i \rightarrow a}^{t+1}(x_i)$ 。

$$\begin{aligned}
 \mu_{i \rightarrow a}^{t+1}(x_i) &\propto p(x_i) \prod_{b \neq a} \mu_{i \leftarrow b}^t(x_i) \\
 &\approx p(x_i) \prod_{b \neq a} \mathcal{N}(x_i | m_{i \leftarrow b}^t, v_{i \leftarrow b}^t) \\
 &\propto p(x_i) \mathcal{N}(x_i | \check{m}_{i \rightarrow a}^t, \check{v}_{i \rightarrow a}^t)
 \end{aligned} \tag{3.33}$$

上式最后一个步骤，是基于高斯相乘引理（引理2-1），以及定义

$$\begin{aligned}
 \check{v}_{i \rightarrow a}^t &\triangleq \left( \sum_{b \neq a} \frac{1}{v_{i \leftarrow b}^t} \right) \\
 \check{m}_{i \rightarrow a}^t &\triangleq \check{v}_{i \rightarrow a}^t \sum_{b \neq a} \frac{m_{i \leftarrow b}^t}{v_{i \leftarrow b}^t}
 \end{aligned} \tag{3.34}$$

为了简化概念，定义

$$\mathcal{M}(x|a, A) \triangleq \frac{p(x) \mathcal{N}(x|a, A)}{\int p(x) \mathcal{N}(x|a, A) dx} \tag{3.35}$$

以及该概率密度的均值和方差，表示为

$$\begin{aligned}\eta(a, A) &\triangleq \int x \mathcal{M}(x|a, A) dx \\ \phi(a, A) &\triangleq \int x^2 \mathcal{M}(x|a, A) dx - \eta^2(a, A)\end{aligned}\tag{3.36}$$

则  $\mu_{i \rightarrow a}^{t+1}(x_i)$  的均值和方差表示为  $\eta(\check{m}_{i \rightarrow a}^t, \check{v}_{i \rightarrow a}^t)$ ,  $\phi(\check{m}_{i \rightarrow a}^t, \check{v}_{i \rightarrow a}^t)$ 。最终的边缘后验概率表示为

$$\tilde{p}(x_i|\mathbf{y}) = \frac{\mathcal{M}(x_i|\check{m}_{i \rightarrow a}^T, \check{v}_{i \rightarrow a}^T) \mathcal{N}(x_i|m_{i \leftarrow a}^T, v_{i \leftarrow a}^T)}{\int \mathcal{M}(x_i|\check{m}_{i \rightarrow a}^T, \check{v}_{i \rightarrow a}^T) \mathcal{N}(x_i|m_{i \leftarrow a}^T, v_{i \leftarrow a}^T) dx_i}\tag{3.37}$$

通过步骤1和步骤2的化简，如图3-4，消息传递算法中消息传递的形式，由原来的概率密度，转化为参数传递，即所得为松弛传播<sup>[21]</sup>算法 (relaxed-belief propagation, RBP)。但是，由于  $(i, a) \in [N] \times [M]$ ，也就是说每一次迭代都需要更新  $\mathcal{O}(NM)$  个参数。对于大型系统而言，这样的计算量仍然是非常庞大的。

**步骤3: TAP-lize。** 为了简化一次迭代中更新消息的数量，定义如下变量

$$Z_a^t \triangleq \sum_i H_{ai} m_{i \rightarrow a}^t, \quad V_a \triangleq \sum_i |H_{ai}|^2 v_{i \rightarrow a}^t\tag{3.38}$$

$$\check{v}_i^t \triangleq \left( \sum_a \frac{1}{v_{i \leftarrow a}^t} \right)^{-1}, \quad \check{m}_i^t \triangleq \check{v}_i^t \sum_a \frac{m_{i \leftarrow a}^t}{v_{i \leftarrow a}^t}\tag{3.39}$$

为了与边缘后验概率一致，定义

$$\hat{m}_i^{t+1} \triangleq \eta(\check{m}_i^t, \check{v}_i^t), \quad \hat{v}_i^{t+1} \triangleq \phi(\check{m}_i^t, \check{v}_i^t)\tag{3.40}$$

由二元泰勒公式<sup>3</sup>，有

$$\phi(\check{m}_{i \rightarrow a}^t, \check{v}_{i \rightarrow a}^t) = \phi(\check{m}_i^t, \check{v}_i^t) + \Delta m \left. \frac{\partial \phi(m, \check{v}_i^t)}{\partial m} \right|_{m=\check{m}_i^t} + \Delta v \left. \frac{\partial \phi(\check{m}_i^t, v)}{\partial v} \right|_{v=\check{v}_i^t}\tag{3.41}$$

其中

<sup>3</sup>  $f(x + \Delta x, y + \Delta y) = f(x, y) + \Delta x f'_x(x, y) + \Delta y f'_y(x, y) + o$

$$\begin{aligned}
 \Delta m &= \check{m}_{i \rightarrow a}^t - \check{m}_i^t \\
 &= \frac{\frac{1}{v_{i \leftarrow a}^t} \left( \sum_b \frac{m_{i \leftarrow b}^t}{v_{i \leftarrow b}^t} \right) - \frac{m_{i \leftarrow a}^t}{v_{i \leftarrow a}^t} \left( \sum_b \frac{1}{v_{i \leftarrow b}^t} \right)}{\left( \sum_b \frac{1}{v_{i \leftarrow b}^t} \right) \left( \sum_{b \neq a} \frac{1}{v_{i \leftarrow b}^t} \right)} \\
 &= \frac{\check{m}_i^t \check{v}_{i \rightarrow a}^t}{v_{i \leftarrow a}^t} - \left( \frac{m_{i \leftarrow a}^t / v_{i \leftarrow a}^t}{\sum_b \frac{1}{v_{i \leftarrow b}^t}} \right) \left( 1 - \frac{\frac{1}{v_{i \leftarrow a}^t}}{\sum_b \frac{1}{v_{i \leftarrow b}^t}} \right)^{-1} \\
 &= \frac{H_{ai}^2 \check{m}_i^t \check{v}_{i \rightarrow a}^t}{\sigma^2 + \sum_{j \neq i} H_{aj}^2 v_{j \rightarrow a}^t} - \left( \frac{m_{i \leftarrow a}^t / v_{i \leftarrow a}^t}{\sum_b \frac{1}{v_{i \leftarrow b}^t}} \right) \left( 1 - \frac{H_{ai}^2 \check{v}_i^t}{\sigma^2 + \sum_{j \neq i} H_{aj}^2 v_{j \rightarrow a}^t} \right)^{-1} \\
 &\approx -\frac{m_{i \leftarrow a}^t \check{v}_i^t}{v_{i \leftarrow a}^t}
 \end{aligned} \tag{3.42}$$

这里, 最后一个式子是通过忽略含有  $H_{ai}^2$  的高阶项得到, 由于  $H_{ai}^2 = O(\frac{1}{N}) \rightarrow 0$ , 并且  $\check{m}_i^t$ ,  $\check{v}_{i \rightarrow a}^t$ , 以及  $\check{v}_i^t$  是有界常量。

$$\begin{aligned}
 \Delta v &= \check{v}_{i \rightarrow a}^t - \check{v}_i^t \\
 &= \frac{\frac{1}{v_{i \leftarrow a}^t}}{\left( \sum_b \frac{1}{v_{i \leftarrow b}^t} \right) \left( \sum_{b \neq a} \frac{1}{v_{i \leftarrow b}^t} \right)} \\
 &= \frac{H_{ai}^2 \check{v}_i^t \check{v}_{i \rightarrow a}^t}{\sigma^2 + \sum_{j \neq i} H_{aj}^2 v_{j \rightarrow a}^t} \\
 &= 0
 \end{aligned} \tag{3.43}$$

整理为

$$v_{i \rightarrow a}^{t+1} = \phi(\check{m}_i^t, \check{v}_i^t) - \frac{m_{i \leftarrow a}^t \check{v}_i^t}{v_{i \leftarrow a}^t} \frac{\partial \phi(m, \check{v}_i^t)}{\partial m} \Big|_{m=\check{m}_i^t} \tag{3.44}$$

同理, 可得

$$\begin{aligned}
 m_{i \rightarrow a}^{t+1} &= \eta(\check{m}_i^t, \check{v}_i^t) - \frac{m_{i \leftarrow a}^t \check{v}_i^t}{v_{i \leftarrow a}^t} \frac{\partial \eta(m, \check{v}_i^t)}{\partial m} \Big|_{m=\check{m}_i^t} \\
 &\stackrel{(a)}{=} \eta(\check{m}_i^t, \check{v}_i^t) - \frac{m_{i \leftarrow a}^t}{v_{i \leftarrow a}^t} \check{v}_i^{t+1}
 \end{aligned} \tag{3.45}$$

其中步骤(a)成立, 是根据指数族分布的性质

$$\begin{aligned}
 \frac{\partial \eta(m, v)}{\partial m} &= \frac{\partial}{\partial m} \frac{\int x p(x) \mathcal{N}(x|m, v) dx}{\int p(x) \mathcal{N}(x|m, v) dx} \\
 &= \frac{1}{v} \frac{\int \frac{x(x-m)}{v} p(x) \mathcal{N}(x|m, v) dx}{\int p(x) \mathcal{N}(x|m, v) dx} \\
 &\quad - \frac{1}{v} \frac{\int x p(x) \mathcal{N}(x|m, v) dx \cdot \int \frac{x-m}{v} p(x) \mathcal{N}(x|m, v) dx}{\left[ \int p(x) \mathcal{N}(x|m, v) dx \right]^2} \\
 &= \frac{1}{v} \phi(m, v)
 \end{aligned} \tag{3.46}$$

定义

$$Z_{i \leftarrow a}^t \triangleq \sum_{j \neq i}^N H_{ai} m_{j \rightarrow a}^t, \quad V_{i \leftarrow a}^t \triangleq \sum_{j \neq i} H_{aj} v_{j \rightarrow a}^t \quad (3.47)$$

联立(3.38)和(3.47)，通过忽略高阶项（包含 $H_{ai}^2$ 的项）

$$V_{i \leftarrow a}^t = V_a^t - H_{ai}^2 v_{i \rightarrow a}^t \approx V_a^t \quad (3.48)$$

将(3.48)，(3.32)代入(3.39)中，有

$$\check{v}_i^t = \left( \sum_a \frac{H_{ai}^2}{\sigma^2 + V_a^t} \right)^{-1} \quad (3.49)$$

将(3.38)，(3.45)以及(3.48)代入(3.47)，有

$$\begin{aligned} Z_{i \leftarrow a}^t &= Z_a^t - H_{ai} \left( \hat{m}_i^t - \frac{H_{ai}(y_a - Z_{i \leftarrow a}^{t-1})}{\sigma^2 + V_a^{t-1}} \right) \\ &= Z_a^t - H_{ai} \hat{m}_i^t \end{aligned} \quad (3.50)$$

将(3.50)代入(3.45)，得

$$\begin{aligned} m_{i \rightarrow a}^{t+1} &= \hat{m}_i^t - \frac{H_{ai} \hat{v}_i^{t+1}}{\sigma^2 + V_a^t} (y_a - Z_a^t + H_{ai} \hat{m}_i^t) \\ &= m_i^{t+1} - v_i^{t+1} \frac{H_{ai}(y_a - Z_a^t)}{\sigma^2 + V_a^t} \end{aligned} \quad (3.51)$$

将(3.51)代入(3.38)，有

$$\begin{aligned} Z_a^t &= \sum_i H_{ai} \left( \hat{m}_i^t - \hat{v}_i^t \frac{H_{ai}(y_a - Z_a^{t-1})}{\sigma^2 + V_a^{t-1}} \right) \\ &= \sum_i H_{ai} \hat{m}_i^t - V_a^t \frac{y_a - Z_a^{t-1}}{\sigma^2 + V_a^{t-1}} \end{aligned} \quad (3.52)$$

简化 $V_a^t$ ，将(3.44)代入(3.38)，有

$$\begin{aligned}
 V_a^t &\stackrel{(a)}{=} \sum_i H_{ai}^2 \left( \phi(\check{m}_i^t, \check{v}_i^t) - \frac{\check{v}_i^t}{v_{i \leftarrow a}^t} O(1) \right) \\
 &= \sum_i H_{ai} \hat{v}_i^t - \sum_i H_{ai}^2 \frac{(\sum_b v_{i \leftarrow b}^t)^{-1}}{v_{i \leftarrow a}^t} O(1) \\
 &= \sum_i H_{ai} \hat{v}_i^t - \sum_i H_{ai}^4 \frac{\left( \sum_b \frac{H_{bi}^2}{\sigma^2 + V_b^t} \right)^{-1}}{\sigma^2 + V_a^t} O(1) \\
 &\stackrel{(b)}{=} \sum_i H_{ai} \hat{v}_i^t - \sum_i H_{ai}^4 \frac{\left( \frac{1}{\sigma^2 + V_a^t} \sum_b H_{bi}^2 \right)^{-1}}{\sigma^2 + V_a^t} \\
 &= \sum_i H_{ai} \hat{v}_i^t - O\left(\frac{1}{N}\right) \\
 &= \sum_i H_{ai} \hat{v}_i^t
 \end{aligned} \tag{3.53}$$

其中(a)成立, 由于  $\left. \frac{\partial \phi(m, \check{v}_i^t)}{\partial m} \right|_{m=\check{m}_i^t}$  以及  $m_{i \leftarrow a}^t$  为有界常量, 详细证明见附录A。(b)成立, 由于  $\sum_b \frac{H_{bi}^2}{\sigma^2 + V_b^t} = \left( \frac{1}{\sigma^2 + V_a^t} \sum_b H_{bi}^2 \right) \cdot O(1)$ 。

将(3.50), (3.32)代入(3.39)中, 有

$$\begin{aligned}
 \check{m}_i^t &= \check{v}_i^t \sum_a \frac{H_{ai}(y_a - Z_{i \leftarrow a}^t)}{\sigma^2 + V_a^t} \\
 &= \check{v}_i^t \sum_a \frac{H_{ai}}{\sigma^2 + V_a^t} (y_a - Z_a^t + H_{ai} \hat{m}_i^t) \\
 &\stackrel{(c)}{=} \hat{m}_i^t + \check{v}_i^t \sum_a \frac{H_{ai}(y_a - Z_a^t)}{\sigma^2 + V_a^t}
 \end{aligned} \tag{3.54}$$

其中, (c)成立基于  $\check{v}_i^t = (\sum_a H_{ai}^2 / (\sigma^2 + V_a^t))^{-1}$ 。

结合变量  $(Z_a^t, V_a^t, \check{v}_i^t, \check{m}_i^t)$  可得AMP的迭代式, 见表3-1。

#### Remark:

- 通过忽略高阶项, RBP算法进一步降解为AMP算法。相对于RBP而言, AMP算法每一次迭代需要更新的变量个数, 由原来的  $O(MN)$  减少为  $O(M+N)$ 。对于大型系统而言, 减少的计算量是非常可观的。
- 此外, 随着系统规模的增加, AMP算法的性能将无限逼近RBP的性能。本文所介绍的AMP推导过程相对比较繁琐, 文献[43]介绍了一种用EP规则来推导AMP的方法。利用该方法, 可以极大简化AMP的推导过程。本文第四章, 扩展了这种方法, 利用EP规则来推导广义近似消息传递算法。

表3-1 近似消息传递算法

**算法 1: 近似消息传递**

1. 初始化:  $\hat{m}_i^1 = \int x_i p(x_i) dx_i$ ,  $\hat{v}_i^1 = \int (x_i - \hat{x}_i^1)^2 p(x_i) dx_i$ ,  $V_a^0 = 1$ ,  $Z_a^0 = y_a$

2. 迭代 ( $t = 1 : T$ )

$$\begin{aligned} V_a^t &= \sum_i |H_{ai}|^2 \hat{v}_i^t \\ Z_a^t &= \sum_i H_{ai} \hat{m}_i^t - \frac{V_a^t (y_a - Z_a^{t-1})}{\sigma^2 + V_a^{t-1}} \\ \check{v}_i^t &= \left[ \sum_a \frac{|H_{ai}|^2}{\sigma^2 + V_a^t} \right]^{-1} \\ \check{m}_i^t &= \hat{m}_i^t + \sum_a \frac{H_{ai} (y_a - Z_a^t)}{\sigma^2 + V_a^t} \\ \hat{m}_i^{t+1} &= \eta(\check{m}_i^t, \check{v}_i^t) \\ \hat{v}_i^{t+1} &= \theta(\check{m}_i^t, \check{v}_i^t) \end{aligned}$$

3. 输出:  $\hat{m}_i^{T+1}$ ,  $\tilde{p}(x_i | \mathbf{y}) = \frac{p(x_i) \mathcal{N}(x_i | \hat{m}_i^{T+1}, \hat{v}_i^{T+1})}{\int p(x_i) \mathcal{N}(x_i | \hat{m}_i^{T+1}, \hat{v}_i^{T+1}) dx_i}$ 。

### 3.3.2 状态更新方程

文献[18]给出了AMP的具体推导过程,然而在数理逻辑上,该推导并不严谨,仅仅作为一种启发式的推导。为了证明AMP的性能,[23]利用自旋玻璃体系中的条件技巧 (condition trick) 给出了AMP收敛性能的具体证明,该文献指出,在观测矩阵 $\mathbf{H}$ 的元素是独立同分布的高斯变量时,AMP算法的性能可以通过被称为状态更新方程 (state evolution, SE) 来预测。所谓状态更新,即系统的MSE (状态) 可以通过参数的更新得到。SE的本质,就是消除算法对观测值和观测矩阵的依赖。但是,该方法理解起来很费劲。本文从算法角度出发,提供了一种相对简单的证明思路。

正如我们所见,(3.40)中 $\hat{m}_i^t$ 为概率分布 $\tilde{p}(x_i | \mathbf{y}) = \frac{p(x_i) \mathcal{N}(x_i | \hat{m}_i^t, \hat{v}_i^t)}{\int p(x_i) \mathcal{N}(x_i | \hat{m}_i^t, \hat{v}_i^t) dx_i}$ 的均值。在大维系统中,该分布可以通过如下的标量模型来等效

$$\check{M} = X + N \sim \mathcal{N}(0, \check{v}) \quad (3.55)$$

其中 $x \sim p(x)$ 。该模型的MMSE估计表示为

$$\begin{aligned}\mathbb{E}[x|\check{M}] &= \int xp(x|\check{m})dx \\ &= \int x \frac{p(x)\mathcal{N}(x|\check{m}, \check{v})}{\int p(x)\mathcal{N}(x|\check{m}, \check{v})dx} dx\end{aligned}\quad (3.56)$$

相应的MMSE估计器的均方误差为

$$\text{MSE}(\check{v}) = \mathbb{E}[(X - \mathbb{E}[X|\check{m}])^2] \quad (3.57)$$

其中，外层期望是对 $p(x, \check{m})$ 取。

在介绍状态更新方程之前，我们首先介绍经验分布收敛于概率分布。对于 $k > 1$ ，我们说函数 $\phi: \mathbb{R}^P \rightarrow \mathbb{R}$ 是 $k$ 阶pseudo-Lipshitz连续的，当且仅当，对于任意 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ ，有

$$|\phi(\mathbf{x}) - \phi(\mathbf{y})| \leq L(1 + \|\mathbf{x}\|^{k-1} + \|\mathbf{y}\|^{k-1})\|\mathbf{x} - \mathbf{y}\| \quad (3.58)$$

记作： $\phi \in \text{PL}(k)$ 。

**定理 3-1**<sup>[23]</sup>: 令 $\mathbf{H}(n)$ 为随机矩阵 $\mathbf{H} \in \mathbb{R}^{m \times N}$ ，索引为 $n$ 的样本序列，矩阵 $\mathbf{H}$ 元素独立同分布， $H_{ai} \sim \mathcal{N}(0, \frac{1}{N})$ 。我们进一步考虑样本序列 $\mathbf{x}(n)$ ，其经验分布 $p_{\hat{\mathbf{x}}^0}$ 弱收敛于真实概率分布 $p_{\mathbf{x}^0}$ ，其中 $p_{\mathbf{x}^0}$ 具有有界的 $(2k-2)$ 阶距。并且假设，当 $k \leq 2$ 以及 $N \rightarrow \infty$ 时，有 $\mathbb{E}_{\hat{p}_{\mathbf{x}^0}(N)}[\mathbf{X}_0^{2k-2}] \rightarrow \mathbb{E}_{p_{\mathbf{x}^0}}[\mathbf{X}_0^{2k-2}]$ 。我们同样假设 $\mathbf{w}$ 中元素相互独立，服从分布 $p_{\mathbf{w}}$ ，并且具有有限的 $(2k-2)$ 阶距。则，对于任意的pseudo-Lipshitz连续函数 $\psi: \mathbb{R}^2 \rightarrow \mathbb{R}$ ，有

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \psi(\hat{m}_i^t, x_i) \stackrel{\text{a.s.}}{=} \mathbb{E}_{Z, X} [\psi(\eta(X + \sqrt{\check{v}^t}Z, \check{v}^t), X)] \quad (3.59)$$

处处成立。其中 $Z$ 表示高斯随机变量，详细证明见[23]。定理3-1所表示的是，随着样本数的增加，样本的复合函数的均值趋于概率的复合函数的均值。

根据定理3-1，设 $\psi$ 为二次函数，则可得

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\hat{m}_i^{t+1} - x_i)^2 = \mathbb{E}_{Z, X} [(\eta(X + \sqrt{\check{v}^t}Z, \check{v}^t) - X)^2] \quad (3.60)$$

显然，上式左边表示AMP算法的MSE。由(3.60)可知，多入多出模型（multiple-input multiple-out, MIMO） $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ 的AMP估计器的MSE等效为一个单入单出（single-input single-out, SISO）模型 $Y = X + N \sim \mathcal{N}(0, \check{v}^t)$ 。此外，为计算MSE还需确定 $\check{v}^t$ 的表达式。根据 $H_{ai}^2 = O(\frac{1}{N})$ ，可得 $V_a^t = \text{MSE}(\check{v}^t)$ 。代入(3.49)，有

$$\check{v}^t = \left( \sum_a \frac{H_{ai}^2}{\sigma^2 + V_a^t} \right)^{-1} = \frac{\sigma^2 + \text{MSE}(\check{v}^t)}{\alpha} \quad (3.61)$$

其中  $\alpha = \frac{M}{N}$ 。

综上，可以AMP的SE

$$\check{v}^t = \frac{1}{\alpha} (\sigma^2 + \text{MSE}(\check{v}^t)) \quad (3.62)$$

其中  $\text{MSE}(\check{v}^t)$  表示SISO模型  $Y = X + N \sim \mathcal{N}(0, \check{v}^t)$  的MSE。

若设  $H_{ai}^2 = O(\frac{1}{M})$ ，则有  $V_a^t = \frac{1}{\alpha} \varepsilon^t$ ，代入(3.49)中，可得

$$\check{v} = \left( \sum_a \frac{H_{ai}^2}{\sigma^2 + V_a^t} \right)^{-1} = \sigma^2 + \frac{1}{\alpha} \text{MSE}(\check{v}^t) \quad (3.63)$$

此时AMP的SE为

$$\check{v} = \sigma^2 + \frac{1}{\alpha} \text{MSE}(\check{v}^t) \quad (3.64)$$

很明显，(3.64)所表示的SE与[23]的SE是一致的。

#### Remark:

- AMP的SE还可以通过副本方法（replica method, RM）得到。RM是起源于统计物理的一种系统分析工具。RM通过最小化自由能条件下，分析系统模型所能达到的不动点方程。文献[31]中，在副本匹配的情况下，通过RM所得到的系统不动点方程与文献[23]通过条件技巧所得到的SE方程一致。也可以说明，AMP算法是贝叶斯最优的。
- 在大规模系统中  $(M, N) \rightarrow \infty$ ，SE指出一个MIMO系统可以由一个SISO系统进行等效，等效的基础是MSE等效。

### 3.4 矢量消息近似传递

对于标准线性逆问题，还存在着另外一种在计算复杂度和性能上表现优异的算法：矢量近似消息传递算法<sup>[28]</sup>（vector approximate message passing, VAMP）。VAMP[36]由Rangan等人提出，用于解决标准线性逆问题。在性能上，VAMP算法与期望传播（expectation propagation, EP）算法在处理标准线性模型是一致的，尤其是对角限制期望一致（diagonal restricted expectation consistent）<sup>[35]</sup>。两者的区别在于，首先，EP所适



用的模型比VAMP要广，其次，对于同样适用的模型，VAMP相对于EP更加规范，VAMP具有自成体系的消息更新规则，并且其性能可以用SE进行预测。与VAMP算法相比，AMP算法是基于循环因子图（loopy factor graph）所推导得到的消息传递算法，其节点是标量的，而VAMP算法是根据Rangan等人所提出的非循环的因子图所推导得到，并且该因子图可以是矢量的。VAMP通过引入 $\delta(\cdot)$ 中间因子节点（见[28]图1），利用消息更新的规则，来实现算法的收敛。

VAMP算法和AMP算法存在着很强的联系。首先，两种算法的性能都可以由SE方程进行预测；其次，由Replica Method根据两者所对应的观测矩阵，可以推导得到与两种算法性能所匹配的不动点方程，也就是说VAMP和AMP是在不同观测矩阵下，达到最小化自由能的算法。此外，Bishop等人，在[10]的第10章介绍了一种基于图的EP，该文献指出，完全因子化的EP可以转化为循环置信传播（loopy belief propagation, LBP），换言之，LBP是一种特殊化的EP算法，亦说明了AMP是一种完全因子化的EP算法。

Rangan等人提出了消息更新规则对于算法的提出起到了至关重要的作用，但是正是由于这套消息更新规则，使得VAMP算法受限于标准线性模型，很难扩展到广义线性模型、多层广义线性模型。为了克服这个困难，受Rangan等人所提出的消息更新规则启发，本文提出一种新的消息更新规则：矩匹配消息传递（moment-matched message passing）。应用该规则，在标准线性模型中，同样可以推导得到VAMP算法，此外，这套规则同样适用于广义线性模型和多层广义线性模型。

需要注意的是，与AMP推导过程不同，由于高斯相乘引理对复数运算同样成立，因此本文所给出的VAMP推导对复数和实数都成立。为了使用VAMP算法适用于复数域，以下所给的推导都是针对复数。

### 3.4.1 算法推导

#### ■ 基于矩匹配消息更新规则的VAMP推导

在介绍消息更新规则以及算法推导之前，首先介绍投影技巧。将概率 $p(x)$ 投影到一个分布族 $\Phi$ 上的操作，定义为

$$\text{Proj}_{\Phi}[p(x)] = \text{KL}(p(x)||q(x)) \quad (3.65)$$

其中 $\text{KL}(\cdot)$ 表示KL散度，KL散度用于刻画两个概率密度之间的距离，在信息理论里也称为相对熵。

消息更新的规则，如图3-5所示。消息更新规则是基于因子图的，为此需要给出标

准线性模型的矢量因子图，如图3-6。注意，该因子图相对于LBP而言，是没有环的。

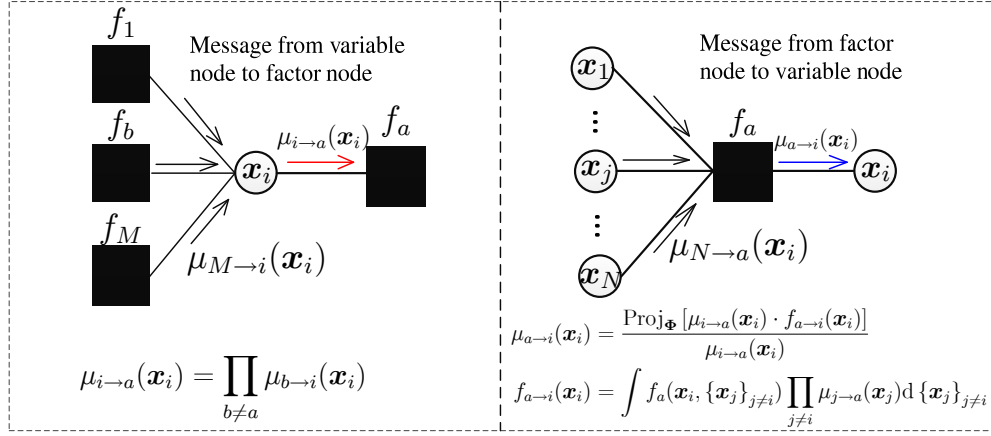


图3-5 矩匹配消息更新规则

注意，图3-6采用了上标来表示消息传递的方向：“+”表示前向；“-”表示反向。这样做的目的是为了使得符号体系与其在多层广义线性模型中一致。

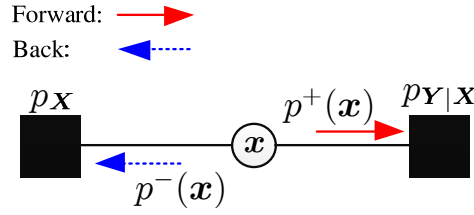


图3-6 标准线性模型的矢量因子图

以下给出VAMP详细的推导过程：

**步骤1：**由于逆向信息的处理包含观测值 $\mathbf{y}$ ，因此规定消息传递的初始方向为逆向。同时，初始化消息 $p^+(\mathbf{x})$ 为高斯分布，即 $p^+(\mathbf{x}) = \mathcal{N}_c(\mathbf{x}|\mathbf{x}^+, \mathbf{v}^+)$ 。

**步骤2：**计算因子节点 $p_{Y|X}$ 的后验分布 $\mathcal{N}_c(\mathbf{x}|\hat{\mathbf{x}}^-, \hat{\mathbf{v}}^-)$ ，以及外部消息（extrinsic messages） $p^-(\mathbf{x})$ 。

$$\begin{aligned}\hat{\mathbf{x}}^- &= \mathbb{E}[\mathbf{x}|\mathbf{x}^+, \mathbf{v}^+] \\ \hat{\mathbf{v}}^- &= \text{Var}[\mathbf{x}|\mathbf{x}^+, \mathbf{v}^+]\end{aligned}\quad (3.66)$$

其中，期望对分布 $\frac{p(\mathbf{y}|\mathbf{x})\mathcal{N}_c(\mathbf{x}|\mathbf{x}^+, \mathbf{v}^+)}{\int p(\mathbf{y}|\mathbf{x})\mathcal{N}_c(\mathbf{x}|\mathbf{x}^+, \mathbf{v}^+)d\mathbf{x}}$ 取。对于标准线性模型，由于 $p(\mathbf{y}|\mathbf{x})$ 是高斯分布，即 $p(\mathbf{y}|\mathbf{z}) = \mathcal{N}_c(\mathbf{y}|\mathbf{H}\mathbf{x}, \sigma^2\mathbf{I})$ 。因此，可以直接计算得到 $(\hat{\mathbf{x}}^-, \hat{\mathbf{v}}^-)$ 的解析表达式。

$$\begin{aligned}& \mathcal{N}_c(\mathbf{y}|\mathbf{H}\mathbf{x}, \sigma^2\mathbf{I})\mathcal{N}_c(\mathbf{x}|\mathbf{x}^+, \text{Diag}(\mathbf{v}^+)) \\ & \propto \mathcal{N}_c(\mathbf{x}|(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}, (\sigma^{-2}\mathbf{H}^H\mathbf{H})^{-1})\mathcal{N}_c(\mathbf{x}|\mathbf{x}^+, \text{Diag}(\mathbf{v}^+)) \\ & \propto \mathcal{N}_c(\mathbf{x}|\hat{\mathbf{x}}^-, \hat{\mathbf{v}}^-)\end{aligned}\quad (3.67)$$

其中

$$\begin{aligned} \mathbf{Q}_x^- &= (\sigma^{-2} \mathbf{H}^H \mathbf{H} + \text{Diag}(\mathbf{1} \oslash \mathbf{v}^+))^{-1} \\ \hat{\mathbf{x}}^- &= \mathbf{Q}_x^- (\sigma^{-2} \mathbf{H}^H \mathbf{y} + \text{Diag}(\mathbf{x}^+ \oslash \mathbf{v}^+)) \\ \hat{\mathbf{v}}^+ &= \text{diag}(\mathbf{Q}_x^-) \end{aligned} \quad (3.68)$$

(3.67)最后一步，可以通过高斯相乘引理（引理2-2）得到。

由于 $p^+(\mathbf{x})$ 被初始化为高斯的，以及因子节点 $p(\mathbf{y}|\mathbf{x})$ 的后验分布投影成高斯分布 $\mathcal{N}_c(\mathbf{x}|\hat{\mathbf{x}}^-, \hat{\mathbf{v}}^-)$ 。根据高斯相乘引理，可知外部信息 $p^-(\mathbf{x})$ 也是高斯的

$$\begin{aligned} p^-(\mathbf{x}) &\propto \frac{\mathcal{N}_c(\mathbf{x}|\hat{\mathbf{x}}^-, \hat{\mathbf{v}}^-)}{\mathcal{N}(\mathbf{x}|\mathbf{x}^+, \mathbf{v}^+)} \\ &\propto \mathcal{N}_c(\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-) \end{aligned} \quad (3.69)$$

其中

$$\begin{aligned} \mathbf{v}^- &= \mathbf{1} \oslash (\mathbf{1} \oslash \hat{\mathbf{v}}^- - \mathbf{1} \oslash \mathbf{v}^+) \\ \mathbf{x}^- &= \mathbf{v}^- \odot (\hat{\mathbf{x}}^- \oslash \hat{\mathbf{v}}^- - \mathbf{x}^+ \oslash \mathbf{v}^+) \end{aligned} \quad (3.70)$$

**步骤3:** 计算因子节点 $p(\mathbf{x})$ 的后验分布 $\mathcal{N}_c(\mathbf{x}|\hat{\mathbf{x}}^+, \hat{\mathbf{v}}^+)$ 以及外部信息 $p^+(\mathbf{x})$

$$\begin{aligned} \hat{\mathbf{x}}^+ &= \mathbb{E}[\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-] \\ \hat{\mathbf{v}}^+ &= \text{Var}[\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-] \end{aligned} \quad (3.71)$$

其中，期望对分布 $\frac{p(\mathbf{x})\mathcal{N}_c(\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-)}{\int p(\mathbf{x})\mathcal{N}_c(\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-)d\mathbf{x}}$ 取。这里需要根据先验分布 $p(\mathbf{x})$ 的具体表达式，得到参数 $(\hat{\mathbf{x}}^+, \hat{\mathbf{v}}^+)$ 。

根据高斯相乘引理，外部信息 $p^+(\mathbf{x})$ 的形式是高斯的

$$\begin{aligned} p^+(\mathbf{x}) &\propto \frac{\mathcal{N}_c(\mathbf{x}|\hat{\mathbf{x}}^+, \hat{\mathbf{v}}^+)}{\mathcal{N}(\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-)} \\ &\propto \mathcal{N}_c(\mathbf{x}|\mathbf{x}^+, \mathbf{v}^+) \end{aligned} \quad (3.72)$$

其中

$$\begin{aligned} \mathbf{v}^+ &= \mathbf{1} \oslash (\mathbf{1} \oslash \hat{\mathbf{v}}^+ - \mathbf{1} \oslash \mathbf{v}^-) \\ \mathbf{x}^+ &= \mathbf{v}^+ \odot (\hat{\mathbf{x}}^+ \oslash \hat{\mathbf{v}}^+ - \mathbf{x}^- \oslash \mathbf{v}^-) \end{aligned} \quad (3.73)$$

基于上述推导，可得VAMP的迭代表达式，见表3-2。

Remark:

- 相对于AMP而言，VAMP放松了对观测矩阵 $\mathbf{H}$ 的要求。AMP要求观测矩阵 $\mathbf{H}$ 是独立同分布，且服从次高斯分布（sub Gaussian），而VAMP则要求 $\mathbf{H}$ 是右正交可逆（right-rotationally invariant）。此外，VAMP对于 $\mathbf{H}$ 的均值不敏感，而AMP要求观测矩阵为均值为0的随机矩阵<sup>[36]</sup>。

- VAMP的推导，也可以通过变分消息传递<sup>[6][10]</sup>（varitional message passing, VMP）得到。这种推导方法，提供了另外一种视角，即VAMP算法是在指数族中寻找一个分布函数来最大化一个下界函数。

### ■ 期望传播算法

期望传播（expectation propagation, EP）算法，最早在Minka的博士论文[9]中提出。EP改进了假设密度滤波（assumed density filtering）消息更新的方式。EP通过最小化KL-散度来近似可因子化的概率密度。在信息理论中，KL-散度也被称为相对熵。

设观测向量 $\mathbf{y}$ 与待估计信号 $\mathbf{x}$ 的联合分布具有因子可分解结构，即

$$p(\mathbf{y}, \mathbf{x}) = p(\mathbf{y}|\mathbf{x}) \prod_{i=1}^N p(x_i) \quad (3.74)$$

利用函数 $q(\mathbf{x}) = \mathcal{N}_c(\mathbf{x}|\mathbf{m}, \mathbf{v})$ 代替 $p(\mathbf{x})$ ，得到后验概率的近似

$$\begin{aligned} \tilde{p}(\mathbf{x}|\mathbf{y}) &\propto \mathcal{N}_c(\mathbf{x}|\mathbf{m}, \mathbf{v}) \mathcal{N}_c(\mathbf{y}|\mathbf{H}\mathbf{x}, \sigma^2 \mathbf{I}) \\ &\propto \mathcal{N}_c(\mathbf{x}|\mathbf{m}, \mathbf{v}) \mathcal{N}_c(\mathbf{x}|(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{y}, (\sigma^{-2} \mathbf{H}^H \mathbf{H})^{-1}) \\ &\propto \mathcal{N}_c(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{aligned} \quad (3.75)$$

其中

$$\begin{aligned} \boldsymbol{\Sigma} &= (\sigma^{-2} \mathbf{H}^H \mathbf{H} + \text{Diag}(\mathbf{1} \oslash \mathbf{v}))^{-1} \\ \boldsymbol{\mu} &= \boldsymbol{\Sigma} (\sigma^{-2} \mathbf{H}^H \mathbf{y} + \text{Diag}(\mathbf{m} \oslash \mathbf{v})) \end{aligned} \quad (3.76)$$

对近似联合概率 $\tilde{p}(\mathbf{x}|\mathbf{y})$ 对角化，得到边缘近似后验概率

$$\tilde{p}(x_i|\mathbf{y}) = \mathcal{N}_c(x_i|\mu_i, \Sigma_{ii}) \quad (3.77)$$

由高斯相乘引理计算得到边缘似然函数的近似 $q^{\setminus i}(x_i)$

$$q^{\setminus i}(x_i) = \frac{\tilde{p}(x_i|\mathbf{y})}{q(x_i)} \propto \mathcal{N}_c(x_i|m_i^{\text{tem}}, v_i^{\text{tem}}) \quad (3.78)$$

其中

$$\begin{aligned} v_i^{\text{tem}} &= \left( \frac{1}{\Sigma_{ii}} - \frac{1}{v_i} \right)^{-1} \\ m_i^{\text{tem}} &= v_i^{\text{tem}} \left( \frac{\mu_i}{\Sigma_{ii}} - \frac{m_i}{v_i} \right) \end{aligned} \quad (3.79)$$

基于边缘似然函数的近似 $q^{\setminus i}(x_i)$ 以及先验分布 $p(x_i)$ ，可构造得到更新的边缘后验概率分布，用于更新 $\tilde{p}(x_i|\mathbf{y})$

$$\tilde{p}^{\text{new}}(x_i|\mathbf{y}) = \arg \max_{\tilde{p}(x_i|\mathbf{y})} \mathcal{D}_{\text{KL}} \left( \frac{1}{C} p(x_i) q^{\setminus i}(x_i) \| \tilde{p}(x_i|\mathbf{y}) \right) \quad (3.80)$$

其中KL散度用于衡量两个概率密度的距离，由于两个概率密度均为指数族分布[10]，因此，当左右两边的概率密度函数矩匹配时，达到KL散度最小。这里，为了与 $\tilde{p}(x_i|\mathbf{y})$ 相区别，我们用 $\tilde{p}^{\text{new}}(x_i|\mathbf{y})$ 表示更新之后的 $\tilde{p}(x_i|\mathbf{y})$ ，其均值和方差分别为

$$\begin{aligned}\hat{m}_i^t &= \mathbb{E}[x_i|m_i^{\text{tem}}, v_i^{\text{tem}}] \\ \hat{v}_i^t &= \text{Var}[x_i|m_i^{\text{tem}}, v_i^{\text{tem}}]\end{aligned}\quad (3.81)$$

其中期望对概率分布 $\frac{p(x_i)\mathcal{N}_c(x_i|m_i^{\text{tem}}, v_i^{\text{tem}})}{\int p(x_i)\mathcal{N}_c(x_i|m_i^{\text{tem}}, v_i^{\text{tem}})dx_i}$ 取。

边缘后验概率更新之后，进而更新近似先验分布 $q(x_i)$ 。由高斯相乘引理，知

$$q(x_i) = \frac{\tilde{p}(x_i|\mathbf{y})}{q^{\setminus i}(x_i)} \propto \mathcal{N}_c(x_i|m_i, v_i) \quad (3.82)$$

其中

$$\begin{aligned}v_i &= \left( \frac{1}{\hat{v}_i} - \frac{1}{v_i^{\text{tem}}} \right)^{-1} \\ m_i &= \left( \frac{\hat{m}_i}{\hat{v}_i} - \frac{m_i^{\text{tem}}}{v_i^{\text{tem}}} \right)\end{aligned}\quad (3.83)$$

联合参数 $(\Sigma_{ii}, \mu_i, m_i^{\text{tem}}, v_i^{\text{tem}}, \hat{m}_i, \hat{v}_i, m_i, v_i)$ 即可得到EP的迭代方程组。

通过观察EP和VAMP的迭代式，可知对于标准线性模型的情况，EP和VAMP在伪代码层面是等价的。

表3-2 矢量近似消息传递算法

---

**算法2: 矢量近似消息传递算法 (VAMP)**

---

1. 初始化:  $t = 1, \mathbf{x}^+ = \mathbf{0}, \mathbf{v}^+ = \mathbf{1}$

2. 迭代 ( $t < T$ )

$$\begin{aligned}\mathbf{Q}_x^- &= (\sigma^{-2} \mathbf{H}^H \mathbf{H} + \text{Diag}(\mathbf{1} \odot \mathbf{v}^+))^{-1} \\ \hat{\mathbf{x}}^- &= \mathbf{Q}_x^- (\sigma^{-2} \mathbf{H}^H \mathbf{y} + \text{Diag}(\mathbf{x}^+ \odot \mathbf{v}^+)) \\ \hat{\mathbf{v}}^+ &= \text{diag}(\mathbf{Q}_x^-) \\ \mathbf{v}^- &= \mathbf{1} \odot (\mathbf{1} \odot \hat{\mathbf{v}}^+ - \mathbf{1} \odot \mathbf{v}^+) \\ \mathbf{x}^- &= \mathbf{v}^- \odot (\hat{\mathbf{x}}^- \odot \hat{\mathbf{v}}^- - \mathbf{x}^+ \odot \mathbf{v}^+) \\ \hat{\mathbf{x}}^+ &= \mathbb{E}[\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-] \\ \hat{\mathbf{v}}^+ &= \text{Var}[\mathbf{x}|\mathbf{x}^-, \mathbf{v}^-] \\ \mathbf{v}^+ &= \mathbf{1} \odot (\mathbf{1} \odot \hat{\mathbf{v}}^+ - \mathbf{1} \odot \mathbf{v}^-) \\ \mathbf{x}^- &= \mathbf{v}^+ \odot (\hat{\mathbf{x}}^+ \odot \hat{\mathbf{v}}^+ - \mathbf{x}^- \odot \mathbf{v}^-)\end{aligned}$$

3. 输出:  $\hat{\mathbf{x}}^+, \tilde{p}(x_i|\mathbf{y}) = \frac{p(x_i)\mathcal{N}(x_i|x_i^-, v_i^-)}{\int p(x_i)\mathcal{N}(x_i|x_i^-, v_i^-)dx_i}$ , ( $x_i^-$ 和 $v_i^-$ 分别表示 $\mathbf{x}^-$ 和 $\mathbf{v}^-$ 的第 $i$ 个元素)。

---

### 3.4.2 状态更新方程

与AMP算法相似，VAMP算法也有对应的SE方程。VAMP算法的MSE性能可以通过SE来进行预测。为了方便理解，这里只考虑匹配情况下的SE。正如(3.71)所见， $\hat{x}^+$ 为概率分布 $\frac{p(x)\mathcal{N}_c(x|x^-,v^-)}{\int p(x)\mathcal{N}_c(x|x^-,v^-)dx}$ 的均值。本文3.3.2中介绍了，在大规模系统中，可以通过等效的标量模型来替代某些变量。即

$$X^- = X + N \sim \mathcal{N}_c(0, v^-) \quad (3.84)$$

该模型的MMSE估计表示为 $\hat{x} = \int xp(x|x^-)dx$ ，其中 $p(x|x^-) = \frac{p(x)\mathcal{N}_c(x|x^-,v^-)}{\int p(x)\mathcal{N}_c(x|x^-,v^-)dx}$ 。对应的均值误差表示为

$$\text{MSE}(v^-) = \mathbb{E}[|\hat{x} - X|^2] \quad (3.85)$$

其中，期望是对 $p(x, x^-)$ 取。

通过上述表示，可知 $X^-$ 可以通过样本 $x$ （由分布 $p_X(x)$ 产生）迭代高斯噪声得到，且噪声的方差为 $v^-$ 。此外，我们需要关注哪些变量对方差 $v^-$ 会产生影响。根据表3-2给出的算法可以看到，变量 $\hat{v}^+ = \langle \hat{v}^+ \rangle$ 中包含了观测矩阵 $\mathbf{H}$ 。为了消除观测矩阵的影响，基于矩阵奇异值分解 $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ ，我们对 $\hat{v}^+$ 进行简化。这里为了简化符号，定义

$$\gamma^+ \triangleq \frac{1}{v^+}, \quad \gamma^- \triangleq \frac{1}{v^-} \quad (3.86)$$

则有

$$\begin{aligned} \hat{v}^+ &= \frac{1}{N} \text{Tr}(\sigma^{-2} \mathbf{H}^H \mathbf{H} + \gamma^+ \mathbf{I})^{-1} \\ &= \frac{1}{N} \text{Tr}(\sigma^{-2} \mathbf{V}^H \mathbf{\Lambda}^H \mathbf{\Lambda} \mathbf{V} + \gamma^+ \mathbf{I})^{-1} \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma^{-2} \lambda_i + \gamma^+} \end{aligned} \quad (3.87)$$

其中 $\lambda_i$ 表示 $\mathbf{H}^H \mathbf{H}$ 的第 $i$ 个特征值。

整理，可以得到VAMP在匹配情况下的SE方程如下：

$$\begin{aligned} \hat{v}^+ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma^{-2} \lambda_i + \gamma^+} \\ \gamma^- &= \frac{1}{\hat{v}^+} - \gamma^+ \\ \gamma^+ &= \frac{1}{\text{MSE}(\gamma^-)} - \gamma^- \end{aligned} \quad (3.88)$$

特别的，对于特定的先验分布， $\text{MSE}(\gamma^-)$ 有解析解，详见3.5小节。

Remark:

- [36]给出了VAMP的SE的详细证明过程，该证明方法借鉴了[23]中的条件技巧，将其扩展到矢量形式。通过数学归纳法，得到VAMP的SE方程，该SE方程包括匹配和不匹配两种情况。但是，本身条件技巧是否适用于矢量形式，仍需证明。因此，[36]给出的SE证明过程是不严谨的。但这并不妨碍，我们理解VAMP以及其SE。
- VAMP的SE的另外一种证明思路，是通过副本分析方法实现。[38]利用RM给出了广义线性模型在最小化自由能时的不动点方程。该不动点方程与GEC-SR由算法所得到的不动点方程一致，间接证明了GEC-SR的最优性。而VAMP是GEC-SR的一种特例。因此，VAMP的SE也可以通过RM进行证明。
- SE分析的实质是消除算法对观测值 $\mathbf{y}$ 和观测矩阵 $\mathbf{H}$ 的依赖。

### 3.5 应用：Massive MIMO信号检测

#### ■ 系统模型

考虑图3-6场景，图中单小区单基站（base station, BS），基站装备 $M$ 根天线，服务 $N$ 个终端用户，每个终端装备1根天线。对于大规模MIMO上行链路，该场景可以通过如下抽象模型进行表示

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (3.89)$$

其中，用户数据 $\mathbf{x} \in \mathbb{C}^N$ 经过信道矩阵 $\mathbf{H} \in \mathbb{C}^{M \times N}$ （假设信道估计步骤中信道被完美估计）后，叠加高斯白噪声 $\mathbf{w} \sim \mathcal{N}_c(\mathbf{w}|\mathbf{0}, \sigma^2 \mathbf{I})$ ，得到接收信号 $\mathbf{y} \in \mathbb{C}^M$ 。本章考虑未编码情况，给定信道矩阵 $\mathbf{H}$ 以及噪声分布 $p_{\mathbf{w}}$ ，接收机需要从接收信号 $\mathbf{y}$ 中检测出用户数据 $\mathbf{x}$ 。

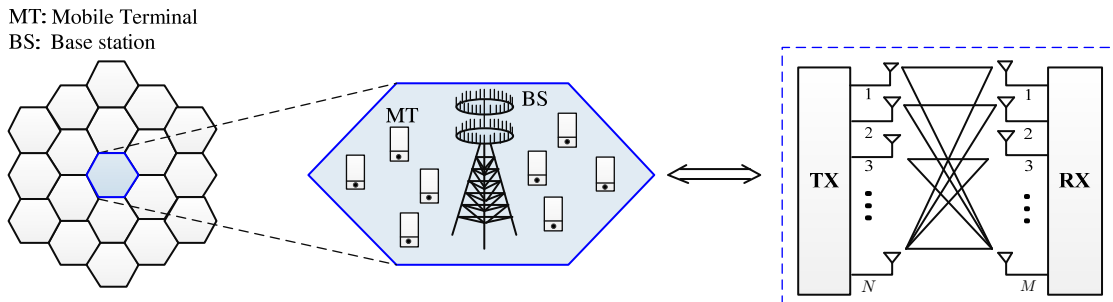


图3-6 Massive MIMO系统

#### ■ 检测器计算复杂度

根据现有信息（先验分布，噪声分布），信号检测器需要从观测向量 $\mathbf{y}$ 中估计出原始信号 $\mathbf{x}$ ，这个过程通过函数形式 $\hat{\mathbf{x}} = g(\mathbf{y})$ ，其中 $g(\cdot) : \mathbb{C}^M \mapsto \mathbb{C}^N$ 。对于传统估计方法（如LMMSE、LS）而言，其所对应的 $g(\cdot)$ 是线性映射。线性映射的优势在于计算量相对简单（LMMSE、LS的主要计算量是矩阵求逆，因此计算量为 $\mathcal{O}(N^3)$ ）。线性映射的缺陷也很明显，当先验分布为非高斯（如高斯分布，混合高斯分布）分布时，线性映射的性能是很糟糕的，这一点将在仿真实验部分得到验证。

对于Massive MIMO系统模型，传统估计算法的解析表达式为

$$\begin{aligned}\hat{\mathbf{x}}_{\text{LMMSE}} &= (\sigma^2 \mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} \\ \hat{\mathbf{x}}_{\text{LS}} &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}\end{aligned}\quad (3.90)$$

显然，LMMSE和LS的计算复杂度主要集中在矩阵求逆，大概是矩阵维度的3次方，即 $\mathcal{O}(N^3)$ 。

另外一种传统估计算法为最大似然估计。最大似然估计器表达式如下

$$\hat{\mathbf{x}}_{\text{ML}} = \arg \max_{\mathbf{x} \in \mathcal{S}} p(\mathbf{y}|\mathbf{x}) \quad (3.91)$$

该估计器需要遍历 $\mathbf{x}$ 的所有组合。该算法的优势在于计算精度最好，其缺点也很突出，即计算复杂度极其之高，为 $\mathcal{O}(|\mathcal{S}|^N)$ ，其中 $|\mathcal{S}|$ 表示集合的长度。因此，实际检测中，通常不会采用最大似然算法，特别对于大规模系统的信号检测问题。

对于消息传递类算法，由于其通过迭代的方式得到MMSE解，因此并不具有解析表达式。如表3-1所示，AMP算法的迭代式均为标量运算，假设迭代次数为 $T$ ，则其计算复杂度为 $\mathcal{O}(T(M+N))$ 。VAMP算法，见表3-2，其计算量主要集中在矩阵求逆部分，假设迭代次数为 $T$ ，则其计算复杂度为 $\mathcal{O}(TN^3)$ 。对比各检测器的计算复杂度，得到表3-3。

表3-3 检测器计算复杂度对比图

检测器类型	计算复杂度
ML	$\mathcal{O}( \mathcal{S} ^N)$
LS	$\mathcal{O}(N^3)$
LMMSE	$\mathcal{O}(N^3)$
VAMP/EP	$\mathcal{O}(TN^3)$
AMP	$\mathcal{O}(T(N+M))$



## ■ 均方误差

在3.3.2节以及3.42节中,我们介绍了AMP算法和VAMP算法的MSE可以通过SE进行跟踪预测。SE指出一个多维信道可以通过一个标量信道进行等效,见(3.55)、(3.84)。为了泛化信道,我们用如下标量信道表示

$$M = X + N \sim \mathcal{N}(0, v) \quad (3.92)$$

其中 $X \sim p_X(x)$ 。该标量信道模型的PME表示为

$$\hat{x}_{\text{MMSE}} = \int x p(x|m) dx \quad (3.93)$$

其中条件概率 $p(x|m) = \frac{p(x)p(m|x)}{\int p(x)p(m|x)dx}$ 。进一步的, PME估计器的MSE表示为

$$\text{MSE} = \int_m \int_x |\hat{x} - x|^2 p(x, m) dx dm \quad (3.94)$$

特别地,对于某些特定的先验分布,MSE存在解析表达式。我们列举了一些常用先验分布,并给出了该先验分布时,PME估计器的MSE表达式:

- QPSK (quadrature phase shift keying) 星座点输入<sup>[37]</sup>。如果先验分布 $p(x)$ 为QPSK星座点情况时,即 $p(x) = \frac{1}{4}I(x - \Omega)$ ,其中 $I(\cdot)$ 表示示信函数

$$I(x) = \begin{cases} 1, & x = 0 \\ 0, & \text{otherwise} \end{cases} \quad (3.95)$$

此时,MSE表示为

$$\text{MSE}(v) = 1 - \int \tanh(1/v + \sqrt{1/v}z) Dz \quad (3.96)$$

其中 $Dz = \mathcal{N}(z|0, 1)dz$ ,  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ 。

- 混合高斯分布 (Gaussian mixture) <sup>[3]</sup>。若先验分布 $p(x)$ 为混合高斯分布时,即

$$p(x) = \rho_k \mathcal{N}_c(x|0, \sigma_k^2) \quad (3.97)$$

为保证概率归一化和功率归一化,参数 $(\rho_k, \sigma_k^2)$ 需满足条件

$$\sum_k^K \rho_k = 1, \quad \sum_k^K \rho_k \sigma_k^2 = 1 \quad (3.98)$$

则后验概率表示为

$$\begin{aligned}
 p(x|m) &= \frac{p(x)\mathcal{N}_c(x|m, v)}{\int p(x)\mathcal{N}_c(x|m, v)dx} \\
 &= \frac{\sum_{k=1}^K \rho_k \mathcal{N}_c(m|0, \sigma_k^2 + v) \mathcal{N}_c\left(x \left| \frac{m\sigma_k^2}{\sigma_k^2 + v}, \frac{\sigma_k^2 v}{\sigma_k^2 + v} \right.\right)}{\sum_{k=1}^K \rho_k \mathcal{N}_c(m|0, \sigma_k^2 + v)}
 \end{aligned} \tag{3.99}$$

计算得到后验概率的一阶距和二阶距分别为

$$\begin{aligned}
 f_a(x|m, v) &\triangleq \mathbb{E}[x|m, v] \\
 &= \frac{\sum_{k=1}^K \rho_k \frac{m\sigma_k^2}{\sigma_k^2 + v} \mathcal{N}_c(m|0, \sigma_k^2 + v)}{\sum_{k=1}^K \rho_k \mathcal{N}_c(m|0, \sigma_k^2 + v)} \\
 f_b(x|m, v) &\triangleq \mathbb{E}[|x|^2|m, v] \\
 &= \frac{\sum_{k=1}^K \rho_k \left[ \frac{\sigma_k^2 v (\sigma_k^2 + v) + |m|^2 \sigma_k^4}{(\sigma_k^2 + v)^2} \right] \mathcal{N}_c(m|0, \sigma_k^2 + v)}{\sum_{k=1}^K \rho_k \mathcal{N}_c(m|0, \sigma_k^2 + v)}
 \end{aligned} \tag{3.100}$$

相应地，得到该分布的方差如下

$$f_c(x|m, v) = f_b(x|m, v) - |f_a(x|m, v)|^2 \tag{3.101}$$

由于  $X \sim \sum_{k=1}^K \rho_k \mathcal{N}_c(x|0, \sigma_k^2)$ ，由卷积公式<sup>4</sup>，可得

$$p_M(m) = \sum_{k=1}^K \rho_k \mathcal{N}_c(m|0, v + \sigma_k^2) \tag{3.102}$$

因此MMSE估计器的MSE表示为

$$\begin{aligned}
 \text{MSE} &= \int_{x, m} |\hat{x} - x|^2 p(x, m) dx dm \\
 &= \int_m \left( \int_x |\hat{x} - x|^2 p(x|m) dx \right) p(m) dm \\
 &= \int_m f_c(x|m, v) p(m) dm
 \end{aligned} \tag{3.103}$$

结合  $f_c(x|m, v)$  与  $p(m)$  的具体表达式，可得

<sup>4</sup> 设  $X \sim p(x)$ ,  $Y \sim p(y)$ ,  $(X, Y) \sim p(x, y)$ 。令  $Z = X + Y$ ，则有

$$p(z) = \int_{-\infty}^{+\infty} p(x, z-x) dx = \int_{-\infty}^{+\infty} p(z-y, y) dy$$

若  $X$  与  $Y$  相互独立，则有

$$p(z) = p_X * p_Y = \int_{-\infty}^{+\infty} p_X(x) p_Y(z-x) dx = \int_{-\infty}^{+\infty} p_X(z-y) p_Y(y) dy$$

$$\begin{aligned} \text{MSE}(v) = \sum_{k=1}^K \frac{\sigma_k^2 v + \sigma_k^4}{\sigma_k^2 + v} - \int \left| \frac{\sum_{k=1}^K \rho_k \frac{m \sigma_k^2}{\sigma_k^2 + v} \mathcal{N}_c(m|0, \sigma_k^2 + v)}{\sum_{k=1}^K \rho_k \mathcal{N}_c(m|0, \sigma_k^2 + v)} \right|^2 \\ \times \sum_{k=1}^K \rho_k \mathcal{N}_c(m|0, \sigma_k^2 + v) dm \end{aligned} \quad (3.104)$$

- 高斯伯努利分布（Gaussian-Bernoulli）。若先验分布为高斯伯努利分布时，即

$$p(x) = (1 - \rho)\delta(x) + \rho \mathcal{N}_c(x|0, \rho^{-1}) \quad (3.105)$$

很容易验证，该分布满足零均值和功率归一化条件。此时，标量信道的后验概率

表示为

$$p(x|m) = \frac{(1 - \rho)\delta(x)\mathcal{N}_c(x|m, v) + \rho \mathcal{N}_c(0|m, \rho^{-1} + v)\mathcal{N}_c\left(x|\frac{m}{\rho v + 1}, \frac{v}{\rho v + 1}\right)}{(1 - \rho)\mathcal{N}_c(0|m, v) + \rho \mathcal{N}_c(0|m, \rho^{-1} + v)} \quad (3.106)$$

计算得到后验概率的均值和方差分别为

$$\begin{aligned} \mathbb{E}[x|M] &= C \frac{m}{\rho v + 1} \\ \text{Var}[x|M] &= C \left( \frac{v}{\rho v + 1} + \left| \frac{m}{\rho v + 1} \right|^2 \right) - |\mathbb{E}[x|M]|^2 \end{aligned} \quad (3.107)$$

其中

$$C = \frac{\rho \mathcal{N}_c(0|m, \rho^{-1} + v)}{(1 - \rho)\mathcal{N}_c(0|m, v) + \rho \mathcal{N}_c(0|m, \rho^{-1} + v)} \quad (3.108)$$

该先验分布下估计器的MSE[30]表示为

$$\text{MSE}(v) = 1 - \frac{\rho}{\rho v + 1} \int \frac{|z|^2}{\rho + (1 - \rho)e^{-\frac{|z|^2}{\rho v}} \left( \frac{1}{\rho v} + 1 \right)} Dz \quad (3.109)$$

其中  $Dz = \mathcal{N}_c(z|0, 1)dz$ 。

## ■ 仿真实验

(1) 图3-7给出了迭代次数与算法MSE的曲线图。参数设置：用户数据为QPSK星座点信号，发送天线  $N = 512$ ，接收天线  $M = 1024$ ，信噪比（signal-to-noise rate, SNR） $\text{SNR} = 8\text{dB}$ ，实验次数  $10^4$ 。从图中，可以看出，

- VAMP的收敛速度比AMP要快，VAMP曲线在  $t = 3$  时，达到收敛，而AMP的收敛

速度相对较慢，需要 $t = 5$ 达到算法收敛。

- 另外，图3-7给出了两种算法的SE曲线图，可以看到两种算法的SE曲线收敛速度上有明显的区别，但是收敛的性能是一致的。换言之，当系统维度趋于无穷大时，VAMP和AMP最终达到收敛的性能是一致的。在该系统维度设定下，VAMP与AMP算法最终的收敛的性能是一致的。

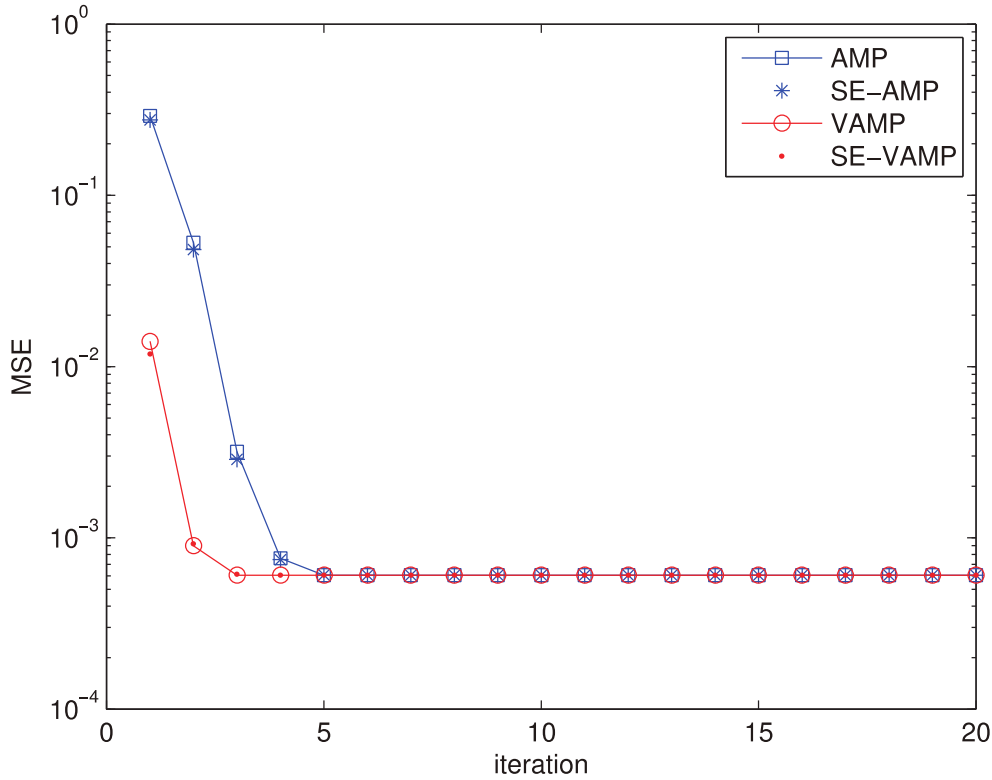


图3-7 Iter v.s. MSE

(2) 图3-8给出了SNR与算法BER的曲线图。参数设定：用户数据为QPSK星座点信号，发送天线 $N = 256$ ，接收天线 $M = 512$ ，每个信噪比点实验次数 $10^4$ 。本次实验比较了两种贝叶斯算法与传统信号重构算法的性能差异。根据图3-8所示，

- 两种传统算法与消息传递类算法之间存在着明显的差距，当性能曲线达到 $10^{-3}$ 时，VAMP的信噪比为6.94dB，AMP为7.22dB，LMMSE为9.42dB，LS为9.82dB。LS与VAMP的BER达到到 $10^{-3}$ 时，所需要的SNR相差了1.88dB。
- 此外，VAMP算法相对于AMP算法存在着微弱的优势。当 $\text{SNR} \leq 5\text{dB}$ 时，VAMP的性能与AMP几乎重合，随着SNR的增大，两者之间有微弱的差距。由图3-7可知，存在这种性能间隔的原因是矩阵的维度不够大。

- 另一方面，对于两种传统算法，随着SNR的增大，LS算法的性能逐渐逼近LMMSE的性能，这一结论也可以从估计器的表达式上得到。

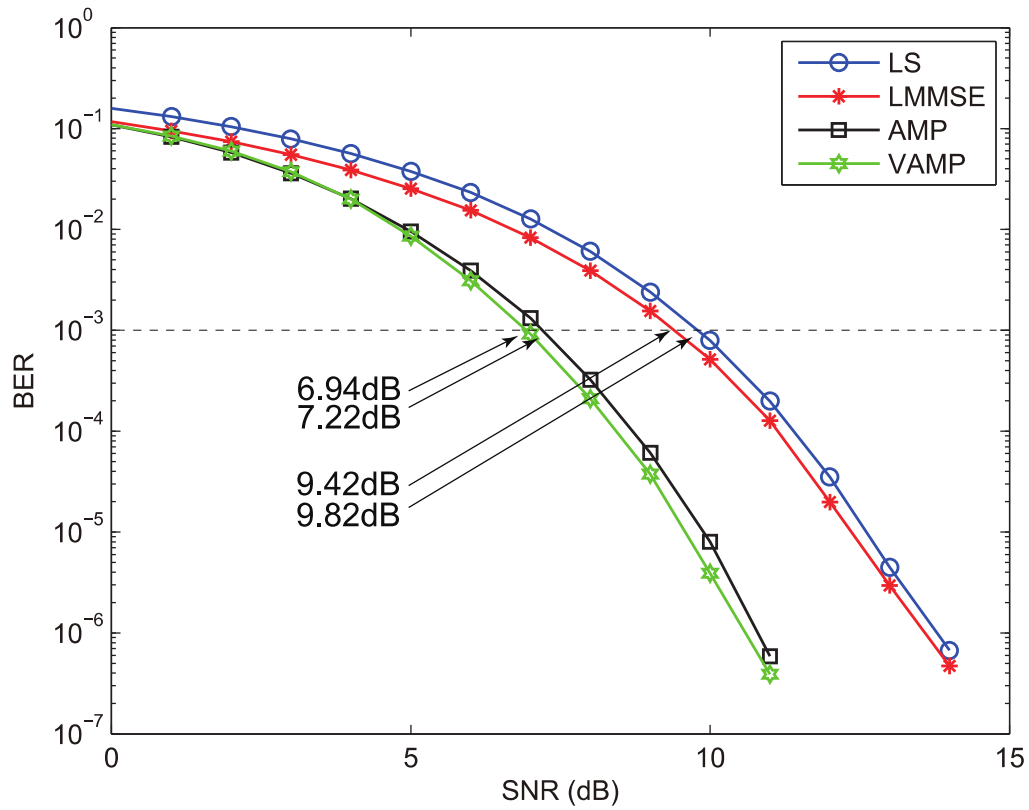


图3-8 SNR v.s. BER

## 第四章 广义线性逆问题的快速解法

广义线性逆问题是线性逆问题的扩展。广义线性逆问题在压缩感知、通信信号处理、模式识别等领域有着重要应用。本章主要研究广义线性逆问题，针对该问题介绍两种算法：广义近似消息传递（generalized approximate message passing, GAMP）和广义期望一致信号重构（generalized expectation consistent-signal recovery, GEC-SR）。我们的贡献主要有：

- 提出了一种基于EP的GAMP推导方法。该方法极大地简化了GAMP的推导过程，并且将GAMP扩展到复数领域。此外，该方法建立了EP与GAMP之间的直接联系。
- 提出了一种基于分列式因子图的GAMP推导方法。该方法相对于原始的GAMP推导过程极大简化了推导步骤，同时该方法对于复数和实数运算同样适用。此外，该方法同样适用于多层广义线性模型。
- 补充了GEC-SR的推导。[30]提出了一种GEC-SR算法，但并没有给出详细的证明过程。本文补充了该证明过程。

目前，本章的部分内容，已经发表于期刊IEEE Signal Processing Letters上，详细见成果列表。

### 4.1 广义线性逆问题

如图4-1所示，广义线性逆问题研究如下模型

$$\mathbf{x} \rightarrow \mathbf{z} = \mathbf{H}\mathbf{x} \rightarrow p(\mathbf{y}|\mathbf{z}) \rightarrow \mathbf{y} \quad (4.1)$$

其中， $\mathbf{x} \in \mathbb{R}^N$ 是目标信号，经过线性变换矩阵 $\mathbf{H} \in \mathbb{R}^{M \times N}$ （信道矩阵已知）后，得到中间变量 $\mathbf{z} \in \mathbb{R}^M$ 。经过一个分量方式的映射（componentwise mapping） $p(\mathbf{y}|\mathbf{z}) = \prod_{a=1}^M p(y_a|z_a)$ 后，得到观测向量或者观测值 $\mathbf{y}$ 。

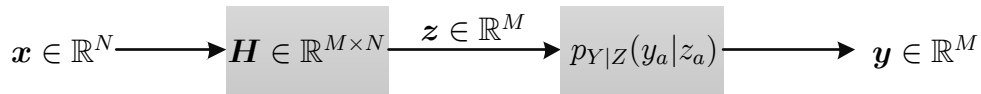


图4-1 广义线性模型

广义线性逆问题所研究的是给定观测矩阵 $\mathbf{H}$ ，信号的先验分布 $p(\mathbf{x})$ ，以及转移概率 $p(\mathbf{y}|\mathbf{z})$ ，如何从观测向量 $\mathbf{y}$ 中重构出原始信号 $\mathbf{x}$ 。在线性逆问题中，转移概率 $p(\mathbf{y}|\mathbf{z})$ 为高斯分布。对于转移概率为非高斯分布的情况，这无疑加深信号恢复的难度<sup>[15]</sup>。而，广义线性逆问题在通信<sup>[2]</sup>、压缩感知<sup>[30]</sup>、矩阵分解<sup>[39]</sup>等领域有着重要应用，因此迫切需要一种性能出众且计算复杂度低的信号重构算法。

如本文3.1节中所介绍的一样，对于广义线性模型，同样考虑先验概率是否已知。当先验概率未知时，通常考虑最大似然估计

$$\begin{aligned}\hat{\mathbf{x}}_{\text{ML}} &= \arg \max_{\mathbf{x} \in \mathcal{S}} p(\mathbf{y}|\mathbf{x}) \\ &= \arg \max_{\mathbf{x} \in \mathcal{S}} \int p(\mathbf{y}|\mathbf{z}) \delta(\mathbf{z} - \mathbf{H}\mathbf{x}) d\mathbf{z}\end{aligned}\quad (4.2)$$

其中 $\mathcal{S}$ 表示 $\mathbf{x}$ 的取值集合。特别的，当输入为星座点分布时， $\mathcal{S}$ 为离散点集合。对于具体的转移概率 $p(\mathbf{y}|\mathbf{z})$ ，最大似然估计存在着很大的计算量。如，当为星座点输入时，最大似然估计的计算量为 $\mathcal{O}(|\mathcal{S}|^N)$ 。显然，随着矩阵维度的增加，最大似然估计器的计算量将呈指数型增加。为了降低最大似然估计的计算量，有很多基于最大似然估计的算法，如广义稀疏贝叶斯学习<sup>[41]</sup>、EM-GAMP<sup>[40]</sup>等。

当先验概率已知时，贝叶斯估计相对于最大似然估计有优势。由于贝叶斯估计考虑了先验概率来提高估计的精度<sup>[33]</sup>。贝叶斯估计根据代价函数的不同主要分为最大后验概率估计（Maximum a posterior, MAP）和最小均方误差估计（minimum mean square error, MMSE），其中最大后验概率估计选取的代价函数为“成功-失败”型函数，而最小均方误差估计选取的代价函数为二次误差型函数。两者所对应的分别是后验概率的最大值点和均值点，这一点在本文2.3节有详细的介绍。特别地，当后验概率为高斯分布时，两者性能一致。

与最大似然函数不同，贝叶斯估计是基于后验概率的估计方法。因此，对于贝叶斯估计而言，最为核心的是计算后验概率。由于最小均方误差估计器更符合代数逻辑，因此本文主要考虑最小均方误差估计。广义线性模型的最小均方误差估计器表示为

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_N \end{pmatrix} = \begin{pmatrix} \mathbb{E}[x_1|\mathbf{y}] \\ \vdots \\ \mathbb{E}[x_N|\mathbf{y}] \end{pmatrix}\quad (4.3)$$

其中期望是对边缘后验概率 $p(x_i|\mathbf{y})$  ( $i = 1, \dots, N$ )取。边缘后验概率表示为

$$\begin{aligned}
 p(x_i|\mathbf{y}) &= \int_{\mathbf{x}_{\setminus i}} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}_{\setminus i} \\
 &\propto \int_{\mathbf{x}_{\setminus i}} p(\mathbf{x}) p(\mathbf{y}|\mathbf{x}) d\mathbf{x}_{\setminus i}
 \end{aligned} \tag{4.4}$$

其中 $\mathbf{x}_{\setminus i}$ 表示 $\mathbf{x}$ 中除第 $i$ 个元素外，所有元素组成的矢量。上式最后一步正比符号‘ $\propto$ ’取代了常数项 $p(\mathbf{y})$ 。这里由于 $\mathbf{y}$ 是已知的观测向量，因此对于任意的概率形式， $p(\mathbf{y})$ 都是确定值。

如(3.6)所示，边缘后验概率的计算需要通过 $(N-1)$ 重积分。特别地，当转移概率 $p(\mathbf{y}|\mathbf{z})$ 为高斯分布时，广义线性模型退化为标准线性模型，进一步地，当先验分布为高斯时，我们可以写出系统的最小均方误差表达式，见(3.5)。但当转移概率为非高斯分布时，只能通过 $(N-1)$ 重积分的方式得到边缘后验概率。当系统规模很大时，这个计算量是极其庞大的。

## 4.2 传统估计方法

精确的信号重构算法往往伴随着巨大的计算量，如最小均方误差估计。最小均方误差估计的前提是计算后验概率的边缘化，其计算量是 $(N-1)$ 重积分。在大规模系统中，求边缘后验概率的计算量是灾难性的。传统的估计方法，以牺牲计算精度的代价，来降低计算复杂度，即次优估计。本节，主要考虑线性最小均方误差估计和最小二乘估计器。注意，最小二乘估计器并不属于贝叶斯估计的范畴。

### 4.2.1 标准线性模型的构造

线性最小均方误差估计考虑估计器 $\hat{\mathbf{x}}_{\text{LMMSE}}$ 与观测值 $\mathbf{y}$ 呈线性关系。与标准线性模型所不同的是，广义线性模型的转移概率 $p(\mathbf{y}|\mathbf{z})$ 为非高斯分布。因此，[33]中第11章中所介绍的推导步骤，并不适用与广义线性模型。为此，我们需要构造出一个标准线性模型来近似真实模型。为了方便理解，标准线性模型的构造流程如图4-2。



图4-2 标准线性模型的构造



首先, 根据中心极限定理, 在大规模系统中, 中间变量 $\mathbf{z}$ 服从渐进高斯分布, 即

$$p(\mathbf{z}) \rightarrow \mathcal{N}(\mathbf{z} | \mathbf{z}^{\text{tem}}, \mathbf{v}_z^{\text{tem}}) \quad (4.5)$$

其中 $\mathbf{z}^{\text{tem}} = \mathbf{H}\mathbb{E}[\mathbf{x}]$ ,  $\mathbf{v}_z^{\text{tem}} = \mathbf{H}\text{Var}[\mathbf{x}]\mathbf{H}^T$ 。因此, 计算得到 $\mathbf{z}$ 的后验均值估计为

$$\begin{aligned} \mathbf{z}^{\text{post}} &= \mathbb{E}[\mathbf{z} | \mathbf{z}^{\text{tem}}, \mathbf{v}_z^{\text{tem}}] \\ \mathbf{v}_z^{\text{post}} &= \text{Var}[\mathbf{z} | \mathbf{z}^{\text{tem}}, \mathbf{v}_z^{\text{tem}}] \end{aligned} \quad (4.6)$$

其中期望对 $\frac{p(\mathbf{y}|\mathbf{z})\mathcal{N}_c(\mathbf{z}|\mathbf{z}^{\text{tem}}, \mathbf{v}_z^{\text{tem}})}{\int p(\mathbf{y}|\mathbf{z})\mathcal{N}_c(\mathbf{z}|\mathbf{z}^{\text{tem}}, \mathbf{v}_z^{\text{tem}})d\mathbf{z}}$ 取。

基于上述推导, 构造如下标准线性模型

$$\tilde{\mathbf{y}} = \mathbf{H}\mathbf{x} + \tilde{\mathbf{w}} \quad (4.7)$$

其中 $\tilde{\mathbf{y}} = \mathbf{z}^{\text{post}}$ ,  $\tilde{\mathbf{w}} \sim \mathcal{N}_c(\mathbf{0}, \tilde{\sigma}^2\mathbf{I})$ ,  $\tilde{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M v_{z,i}^{\text{post}}$ 。

Remark:

- 通过构造标准线性模型, 我们采用标准线性模型的方法来解出原始信号 $\mathbf{x}$ 。换言之, 基于构造标准线性模型, 所有适用于标准线性模型的方法如: 传统估计方法(线性最小均方误差估计、最小二乘估计), 消息传递算法(AMP、VAMP等), 都可以应用于广义线性模型。
- 值得注意的是, 通过构造的标准线性模型解出来的 $\hat{\mathbf{x}}$ 并未更新现有的标准线性模型。换言之, 该方法模型与估计值并未形成反馈回路。这一部分工作, 在[27]中得到了完善。

## 4.2.2 传统估计方法

### ■ 线性最小均方误差估计

根据已构造的标准线性模型 $\tilde{\mathbf{y}} = \mathbf{H}\mathbf{x} + \tilde{\mathbf{w}}$ , 通过强制设定先验概率为高斯分布, 可以得到最小均方误差估计器的表达式如下

$$\hat{\mathbf{x}}_{\text{LMMSE}} = (\tilde{\sigma}^2\mathbf{I} + \mathbf{H}^T\mathbf{H})^{-1} \mathbf{H}^T \tilde{\mathbf{y}} \quad (4.8)$$

详细步骤见本文3.2节。

### ■ 最小二乘估计

如上一小节所述, 通过构造标准线性模型 $\tilde{\mathbf{y}} = \mathbf{H}\mathbf{x} + \tilde{\mathbf{w}}$ , 传统的次优方法可以应用于广义线性模型。本文3.2.2节, 详细介绍了最小二乘估计器的推导过程, 这里就不再赘述。通过构造标准线性模型 $\tilde{\mathbf{y}} = \mathbf{H}\mathbf{x} + \tilde{\mathbf{w}}$ , 则该模型的最小二乘估计表示为

$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{H}^T\mathbf{H})^{-1} \mathbf{H}^T \tilde{\mathbf{y}} \quad (4.9)$$

### 4.3 广义近似消息传递算法

基于构造线性模型的传统估计方法，虽然在计算复杂度上有了明显的改观，但是在性能上与贝叶斯最优相去甚远。为此，迫切需要一种性能优异（逼近或者达到贝叶斯最优）同时具有低复杂度的算法。基于因子图的广义近似消息传递（generalized approximate message passing, GAMP）算法就是其中一种。GAMP算法源于AMP算法，并且扩展了AMP的适用范围。换言之，AMP算法是GAMP算法的一种特例。

#### 4.3.1 算法推导

文献[14]给出了GAMP的详细推导。该方法从因子图出发，利用泰勒公式和中心极限定理完成消息的近似。该推导方法，由于包含泰勒近似环节，当用于复数信号时，推导过程中的变量个数将成倍增加<sup>[1]</sup>。

##### ■ 基于传统因子图的GAMP推导

本文提出一种基于EP的GAMP推导方法，在保留原有的因子图上（传统因子图），通过定义新的消息更新规则，完成GAMP的推导。该推导思路，极大简化了GAMP的推导过程，并且使得GAMP算法很自然适用于复数领域。此外，基于EP的推导思路，建立了EP和GAMP两类算法的直接联系。目前，该成果已经发表在[1]上。GAMP推导过程中假设联合概率是完全因式化（fully factorable），即

$$p(\mathbf{x}, \mathbf{y}) = \prod_{a=1}^N p(x_i) \prod_{a=1}^M p(y_a | \mathbf{x}) \quad (4.10)$$

该联合概率用因子图表示如图4-3。注意，这里广义线性模型的因子图与标准线性模型的因子图一致，如图3-3。

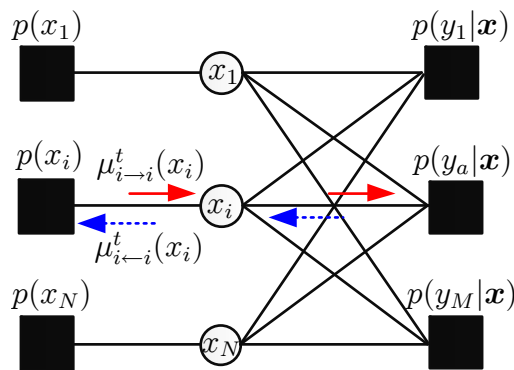


图4-3 广义线性模型因子图

根据图3-5给出的消息更新规则，可以得到如图4-3所表示的广义线性模型的消息传递算法为

$$\begin{aligned}
 \mu_{i \leftarrow a}^{t+1}(x_i) &\propto \frac{\text{Proj}_{\Phi}[\mu_{i \rightarrow a}^t(x_i) \cdot f_{i \leftarrow a}^t(x_i)]}{\mu_{i \rightarrow a}^t(x_i)} \\
 \mu_{i \leftarrow i}^{t+1}(x_i) &\propto \prod_b \mu_{i \leftarrow b}^{t+1}(x_i) \\
 \mu_{i \rightarrow i}^{t+1}(x_i) &\propto \frac{\text{Proj}_{\Phi}[p(x_i) \mu_{i \leftarrow i}^{t+1}(x_i)]}{\mu_{i \leftarrow i}^t(x_i)} \\
 \mu_{i \rightarrow a}^{t+1}(x_i) &\propto \mu_{i \rightarrow i}^{t+1}(x_i) \prod_{b \neq a} \mu_{i \leftarrow b}^t(x_i)
 \end{aligned} \tag{4.11}$$

其中

$$f_{i \leftarrow a}^t(x_i) \triangleq \int p(y_a | \mathbf{x}) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) d\mathbf{x}_{\setminus i} \tag{4.12}$$

$\text{Proj}_{\Phi}[\cdot]$ 表示投影操作，见3.4节。此外， $\mu_{i \leftarrow a}^{t+1}(x_i)$ 表示第 $t+1$ 次迭代，从因子节点 $p(y_a | \mathbf{x})$ 到变量节点 $x_i$ 消息，如图4-2第二根蓝色虚线； $\mu_{i \leftarrow i}^{t+1}(x_i)$ 表示第 $t+1$ 次迭代，从变量节点 $x_i$ 到因子节点 $p(x_i)$ 的消息。 $\mu_{i \rightarrow i}^{t+1}(x_i)$ 表示第 $t+1$ 迭代，从因子节点 $p(x_i)$ 到变量节点 $x_i$ 的消息； $\mu_{i \rightarrow a}^{t+1}(x_i)$ 表示第 $t+1$ 次迭代，从变量节点 $x_i$ 到因子节点 $p(y_a | \mathbf{x})$ 的消息。

观察(4.11)，合并后三个公式，得到

$$\begin{aligned}
 \mu_{i \leftarrow a}^{t+1}(x_i) &\propto \frac{\text{Proj}_{\Phi}[\mu_{i \rightarrow a}^t(x_i) \cdot f_{i \leftarrow a}^t(x_i)]}{\mu_{i \rightarrow a}^t(x_i)} \\
 \mu_{i \rightarrow a}^{t+1}(x_i) &\propto \frac{\text{Proj}_{\Phi}[p(x_i) \prod_b \mu_{i \leftarrow b}^{t+1}(x_i)]}{\mu_{i \leftarrow a}^{t+1}(x_i)}
 \end{aligned} \tag{4.13}$$

最终的边缘后验概率分布表示为 $\tilde{p}(x_i | \mathbf{y}) = \mu_{i \rightarrow a}(x_i)^T \cdot \mu_{i \leftarrow a}^T(x_i)$ 。注意，与LBP所不同的，基于矩匹配消息更新规则所传递的消息均为高斯分布。

以下给出详细的推导步骤

**步骤1：**简化 $f_{i \leftarrow a}^t(x_i)$ 。

$$\begin{aligned}
 f_{i \leftarrow a}^t(x_i) &\propto \int_{\mathbf{x}_{\setminus i}} \int_{z_a} \delta(z_a - \sum_{j \neq i} H_{aj} x_j - H_{ai} x_i) p(y_a | z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) dz_a d\mathbf{x}_{\setminus i} \\
 &\stackrel{(a)}{\propto} \int_{\mathbf{x}_{\setminus i}} \int_{z_a} \delta\left(x_i - \frac{1}{H_{ai}}(z_a - \sum_{j \neq i} H_{aj} x_j)\right) p_{\zeta_a}(z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) dz_a d\mathbf{x}_{\setminus i}
 \end{aligned} \tag{4.14}$$

其中(a)，依据 $\delta(ax) = \frac{1}{|a|}\delta(x)$ 以及定义 $p_{\zeta_a}(z_a) \triangleq \frac{p(y_a | z_a)}{\int p(y_a | z_a) dz_a}$ 。

定义随机变量 $\xi_{j \rightarrow a}^t \sim f_{j \rightarrow a}^t(x_j)$ ， $\zeta_a \sim p_{\zeta_a}(z_a)$ ，以及 $\xi_{i \leftarrow a}^t \sim \mu_{i \leftarrow a}^t(x_i)$ 。此外定义随机变量 $\mathbf{X} \triangleq \{\zeta_a, \{\xi_{j \rightarrow a}^t\}_{j \neq i}\}$ ，则该随机变量的概率密度函数表达为

$$p_{\mathbf{X}}(z_a, \{x_j\}_{j \neq i}) = p_{\zeta_a}(z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) \quad (4.15)$$

为了应用 PDF-to-RV 引理，定义  $g(z_a, \{x_j\}_{j \neq i}) \triangleq z_a - \sum_{j \neq i} H_{aj} x_j$  及  $\beta = \frac{1}{H_{ai}}$ 。由 PDF-to-RV 引理（引理2-3），有

$$\xi_{i \leftarrow a}^t = \frac{1}{H_{ai}} \left( \zeta_a - \sum_{j \neq i} H_{aj} \xi_{j \rightarrow a}^t \right) \quad (4.16)$$

根据中心极限定理，当  $N \rightarrow \infty$ ，有  $\sum_{j \neq i} \xi_{j \rightarrow a}^t \rightarrow \mathcal{N}_c(\sum_{j \neq i} H_{aj} m_{j \rightarrow a}^t, \sum_{j \neq i} |H_{aj}|^2 v_{j \rightarrow a}^t)$ ，其中  $m_{j \rightarrow a}^t$  和  $v_{j \rightarrow a}^t$  分别表示  $\mu_{j \rightarrow a}^t(x_i)$  的均值和方差。注意，与标准线性模型不同，这里  $\zeta_a$  并不是一个高斯随机变量。

定义

$$Z_{i \leftarrow a}^t = \sum_{j \neq i} H_{aj} m_{j \rightarrow a}^t, \quad V_{i \leftarrow a}^t = \sum_{j \neq i} |H_{aj}|^2 v_{j \rightarrow a}^t \quad (4.17)$$

由卷积公式<sup>5</sup>及随机变量的比例变换性质<sup>6</sup>，可得

$$f_{i \leftarrow a}^t(x_i) \propto \mathbb{E}_{\zeta_a} \left[ \mathcal{N}_c \left( x_i \left| \frac{\zeta_a - Z_{i \leftarrow a}^t}{H_{ai}}, \frac{V_{i \leftarrow a}^t}{|H_{ai}|^2} \right) \right) \right] \quad (4.18)$$

其中，期望对  $p_{\zeta_a}(z_a)$  取。注意， $f_{i \leftarrow a}^t(x_i)$  并不是高斯概率密度，因为  $\zeta_a$  通常是非高斯的。若  $\zeta_a$  是高斯，则可以通过高斯相乘引理（引理2-1）得到  $f_{i \leftarrow a}^t(x_i)$  也是高斯的，见3.3节。对于广义线性模型，通常  $\zeta_a$  是非高斯的。

**步骤2：** 计算投影  $\text{Proj}_{\Phi}[\mu_{i \rightarrow a}^t(x_i) \cdot f_{i \leftarrow a}^t(x_i)]$ 。

$$\text{Proj}_{\Phi}[\mu_{i \rightarrow a}^t(x_i) f_{i \leftarrow a}^t(x_i)] = \text{Proj}_{\Phi} \left\{ \mathbb{E}_{\zeta_a} \left[ \mu_{i \rightarrow a}^t(x_i) \mathcal{N}_c \left( x_i \left| \frac{\zeta_a - Z_{i \leftarrow a}^t}{H_{ai}}, \frac{V_{i \leftarrow a}^t}{|H_{ai}|^2} \right) \right) \right] \right\} \quad (4.19)$$

注意  $\mu_{i \rightarrow a}^t(x_i)$  已经初始化为高斯分布，即  $\mu_{i \rightarrow a}^t(x_i) = \mathcal{N}_c(x_i | m_{i \rightarrow a}^t, v_{i \rightarrow a}^t)$ 。根据高斯相乘引理（引理2-1），有

<sup>5</sup> 令  $X \sim p_X(x)$ ， $Y \sim p_Y(y)$ ，设  $X, Y$  相互独立，令  $Z = X + Y$ ，则

$$p_Z(z) = p_X * p_Y = \int_{-\infty}^{+\infty} p_X(x) p_Y(z - x) dx = \int_{-\infty}^{+\infty} p_X(z - y) p_Y(y) dy$$

<sup>6</sup> 令  $X \sim p_X(x)$ ，定义  $Y = aX$ ，则有

$$p_Y(y) = \frac{1}{a} p_X\left(\frac{y}{a}\right)$$

$$\begin{aligned}
 & \mu_{i \rightarrow a}^t(x_i) \mathcal{N}_c \left( x_i \left| \frac{\zeta_a - Z_{i \leftarrow a}^t}{H_{ai}}, \frac{V_{i \leftarrow a}^t}{|H_{ai}|^2} \right. \right) \\
 &= \mathcal{N}_c \left( 0 \left| m_{i \rightarrow a}^t - \frac{\zeta_a - Z_{i \leftarrow a}^t}{H_{ai}}, v_{i \rightarrow a}^t + \frac{V_{i \leftarrow a}^t}{|H_{ai}|^2} \right. \right) \mathcal{N}_c(x_i | \tilde{m}_{i \leftarrow a}^t, \tilde{v}_{i \leftarrow a}^t) \\
 &\propto \mathcal{N}_c(\zeta_a | H_{ai} m_{i \rightarrow a}^t + Z_{i \leftarrow a}^t, |H_{ai}|^2 v_{i \rightarrow a}^t + V_{i \leftarrow a}^t) \mathcal{N}_c(x_i | \tilde{m}_{i \leftarrow a}^t, \tilde{v}_{i \leftarrow a}^t)
 \end{aligned} \tag{4.20}$$

其中

$$\begin{aligned}
 \tilde{v}_{i \leftarrow a}^t &= \left( \frac{1}{v_{i \rightarrow a}^t} + \frac{|H_{ai}|^2}{V_{i \leftarrow a}^t} \right)^{-1} \\
 \tilde{m}_{i \leftarrow a}^t &= \tilde{v}_{i \leftarrow a}^t \left( \frac{m_{i \rightarrow a}^t}{v_{i \rightarrow a}^t} + \frac{H_{ai}^*(\zeta_a - Z_{i \leftarrow a}^t)}{V_{i \leftarrow a}^t} \right)
 \end{aligned} \tag{4.21}$$

定义

$$Z_a^t \triangleq H_{ai} m_{i \leftarrow a}^t + Z_{i \leftarrow a}^t, \quad V_a^t \triangleq |H_{ai}|^2 v_{i \rightarrow a}^t + V_{i \leftarrow a}^t \tag{4.22}$$

对(4.20)关于 $x_i$ 归一化，得到PDF（关于 $x_i$ 的PDF）

$$\begin{aligned}
 & \frac{\mathbb{E}_{\zeta_a} [\mathcal{N}_c(\zeta_a | Z_a^t, V_a^t) \mathcal{N}_c(x_i | \tilde{m}_{i \leftarrow a}^t, \tilde{v}_{i \leftarrow a}^t)]}{\mathbb{E}_{\zeta_a} [\mathcal{N}_c(\zeta_a | Z_a^t, V_a^t)]} \\
 &= \frac{\int p(y_a | z_a) \mathcal{N}_c(z_a | Z_a^t, V_a^t) \mathcal{N}_c(x_i | \tilde{m}_{i \leftarrow a}^t, \tilde{v}_{i \leftarrow a}^t) dz_a}{\int p(y_a | z_a) \mathcal{N}_c(z_a | Z_a^t, V_a^t) dz_a}
 \end{aligned} \tag{4.23}$$

上式最后一个公式结合 $\zeta_a = \frac{p(y_a | z_a)}{\int p(y_a | z_a) dz_a}$ 的定义得到。为了进一步简化，定义随机变量

$$\eta_a^t \sim \frac{p(y_a | z_a) \mathcal{N}_c(z_a | Z_a^t, V_a^t)}{\int p(y_a | z_a) \mathcal{N}_c(z_a | Z_a^t, V_a^t) dz_a} \tag{4.24}$$

其均值和方差分别为

$$\tilde{z}_a^t = \mathbb{E}_{\eta_a^t} [\eta_a], \quad \tilde{v}_a^t = \text{Var}_{\eta_a^t} [\eta_a] \tag{4.25}$$

利用该定义，则(4.23)的均值为

$$\begin{aligned}
 m_i^t &= \int x_i \int_{z_a} p_{\eta_a^t}(z_a) \mathcal{N}_c(x_i | \tilde{m}_{i \leftarrow a}^t, \tilde{v}_{i \leftarrow a}^t) dz_a dx_i \\
 &= \int_{z_a} p_{\eta_a^t}(z_a) \left( \frac{1}{v_{i \rightarrow a}^t} + \frac{|H_{ai}|^2}{V_{i \leftarrow a}^t} \right)^{-1} \left( \frac{m_{i \rightarrow a}^t}{v_{i \rightarrow a}^t} + \frac{H_{ai}^*(z_a - Z_{i \leftarrow a}^t)}{V_{i \leftarrow a}^t} \right) dz_a \\
 &= \left( \frac{1}{v_{i \rightarrow a}^t} + \frac{|H_{ai}|^2}{V_{i \leftarrow a}^t} \right)^{-1} \left( \frac{m_{i \rightarrow a}^t}{v_{i \rightarrow a}^t} + \frac{H_{ai}^*(\tilde{z}_a^t - Z_{i \leftarrow a}^t)}{V_{i \leftarrow a}^t} \right)
 \end{aligned} \tag{4.26}$$

同理可得(4.23)的方差

$$v_i^t = \frac{v_{i \rightarrow a}^t V_{i \leftarrow a}^t V_a^t + (v_{i \rightarrow a}^t)^2 |H_{ai}|^2 \tilde{v}_a^t}{(V_a^t)^2} \tag{4.27}$$

因此

$$\text{Proj}_{\Phi} [\mu_{i \rightarrow a}^t(x_i) \cdot f_{i \leftarrow a}^t(x_i)] = \mathcal{N}_c(x_i | m_i^t, v_i^t) \tag{4.28}$$

其中参数 $(m_i^t, v_i^t)$ 由(4.26)、(4.27)给出。

**步骤3:** 计算 $\mu_{i \leftarrow a}^{t+1}(x_i)$ 。由高斯相乘引理（引理2-1），分别得到 $\mu_{i \leftarrow a}^{t+1}(x_i)$ 的均值和方差

$$\begin{aligned} v_{i \leftarrow a}^{t+1} &= \left( \frac{1}{v_i^t} - \frac{1}{v_{i \rightarrow a}^t} \right)^{-1} = \frac{(V_a^t)^2 - v_{i \rightarrow a}^t (V_a^t - \tilde{v}_a^t)}{|H_{ai}|^2 (V_a^t - \tilde{v}_a^t)} \\ m_{i \leftarrow a}^{t+1} &= v_{i \leftarrow a}^{t+1} \left( \frac{m_i^t}{v_i^t} - \frac{m_{i \rightarrow a}^t}{v_{i \rightarrow a}^t} \right) = \frac{H_{ai}^* (\tilde{z}_a^t - Z_a^t) V_a^t + |H_{ai}|^2 m_{i \rightarrow a}^t (V_a^t - \tilde{v}_a^t)}{|H_{ai}|^2 (V_a^t - \tilde{v}_a^t)} \end{aligned} \quad (4.29)$$

为了与GAMP算法相对照，定义变量

$$\begin{aligned} \tau_a^t &\triangleq \frac{V_a^t - \tilde{v}_a^t}{(V_a^t)^2} \\ s_a^t &\triangleq \frac{\tilde{z}_a^t - Z_a^t}{V_a^t} \end{aligned} \quad (4.30)$$

通过忽略高阶项（包含 $|H_{ai}| = O(\frac{1}{N})$ 项），进一步化简得到

$$\begin{aligned} v_{i \leftarrow a}^{t+1} &= \frac{1}{|H_{ai}|^2 \tau_a^t} \\ m_{i \leftarrow a}^{t+1} &= \frac{H_{ai}^* s_a^t + |H_{ai}|^2 \tau_a^t m_{i \rightarrow a}^t}{|H_{ai}|^2 \tau_a^t} \end{aligned} \quad (4.31)$$

**步骤4:** 计算投影 $\text{Proj}_{\Phi}[p(x_i) \prod_a \mu_{i \leftarrow a}^{t+1}(x_i)]$ 。由高斯相乘引理（引理2-1），得

$$\prod_{a=1} \mu_{i \leftarrow a}^{t+1}(x_i) \propto \mathcal{N}_c(x_i | R_i^{t+1}, \Sigma_i^{t+1}) \quad (4.32)$$

其中

$$\begin{aligned} \Sigma_i^{t+1} &\triangleq \left( \sum_{a=1}^M |H_{ai}|^2 \tau_a^t \right)^{-1} \\ R_i^{t+1} &\triangleq \Sigma_i^{t+1} \sum_{a=1}^M (H_{ai}^* s_a^t + |H_{ai}|^2 \tau_a^t m_{i \rightarrow a}^t) \end{aligned} \quad (4.33)$$

则 $\text{Proj}_{\Phi}[p(x_i) \prod_a \mu_{i \leftarrow a}^{t+1}(x_i)] = \mathcal{N}_c(x_i | \hat{m}_i^{t+1}, \hat{v}_i^{t+1})$ ，其中

$$\begin{aligned} \hat{m}_i^{t+1} &= \mathbb{E}[x_i | R_i^{t+1}, \Sigma_i^{t+1}] \\ \hat{v}_i^{t+1} &= \text{Var}[x_i | R_i^{t+1}, \Sigma_i^{t+1}] \end{aligned} \quad (4.34)$$

这里期望对 $\frac{p(x_i) \mathcal{N}_c(x_i | R_i^{t+1}, \Sigma_i^{t+1})}{\int p(x_i) \mathcal{N}_c(x_i | R_i^{t+1}, \Sigma_i^{t+1}) dx_i}$ 取。

**步骤5:** 计算 $\mu_{i \rightarrow a}^{t+1}(x_i)$ 。由高斯相乘引理（引理2-1），计算得到 $\mu_{i \rightarrow a}^{t+1}(x_i)$ 的均值和方差分别为

$$\begin{aligned} v_{i \rightarrow a}^{t+1} &= \left( \frac{1}{\hat{v}_i^{t+1}} - \frac{1}{v_{i \leftarrow a}^{t+1}} \right)^{-1} = \left( \frac{1}{\hat{v}_i^{t+1}} - |H_{ai}|^2 \tau_a^t \right)^{-1} = \hat{v}_i^{t+1} \\ m_{i \rightarrow a}^{t+1} &= v_{i \rightarrow a}^{t+1} \left( \frac{\hat{m}_i^{t+1}}{\hat{v}_i^{t+1}} - \frac{m_{i \leftarrow a}^{t+1}}{v_{i \leftarrow a}^{t+1}} \right) = \hat{m}_i^{t+1} - H_{ai}^* s_a^t \hat{v}_i^{t+1} \end{aligned} \quad (4.35)$$

**步骤6:** 简化参数( $Z_a^t, V_a^t, R_i^t$ )。将(4.35)代入(4.22), 得

$$\begin{aligned} V_a^t &= \sum_{i=1}^N |H_{ai}|^2 \hat{v}_i^t \\ Z_a^t &= \sum_{i=1}^N H_{ai} \hat{m}_i^t - s_a^{t-1} V_a^t \end{aligned} \quad (4.36)$$

将(4.35)第2式与(4.33)第1式, 代入(4.33)第2式, 得

$$R_i^t = \hat{m}_i^t + \sum_{a=1}^M H_{ai}^* s_a^t \quad (4.37)$$

结合参数( $Z_a^t, V_a^t, \tau_a^t, s_a^t, \hat{m}_i^{t+1}, \hat{v}_i^{t+1}$ )构成GAMP的迭代方程式, GAMP算法见表4-1。

## ■ 基于分列式因子图的GAMP推导

广义线性模型的分列式因子图, 如图4-4所示。与传统因子图相比, 分列式因子图将似然函数节点 $p(y_a|\mathbf{x})$ 分裂成了一个转移概率节点 $p(y_a|z_a)$ 和一个 $\delta(\cdot)$ 节点, 即

$$p(y_z|\mathbf{x}) = \int \delta(z_a - \mathbf{H}_{a,:}\mathbf{x}) p(y_a|z_a) dz_a \quad (4.38)$$

注意, 两种因子图是不等价的, 但是两种因子图根据LBP算法所得到的边缘后验概率是一致的, 这一点在3.3节中有详细的叙述。

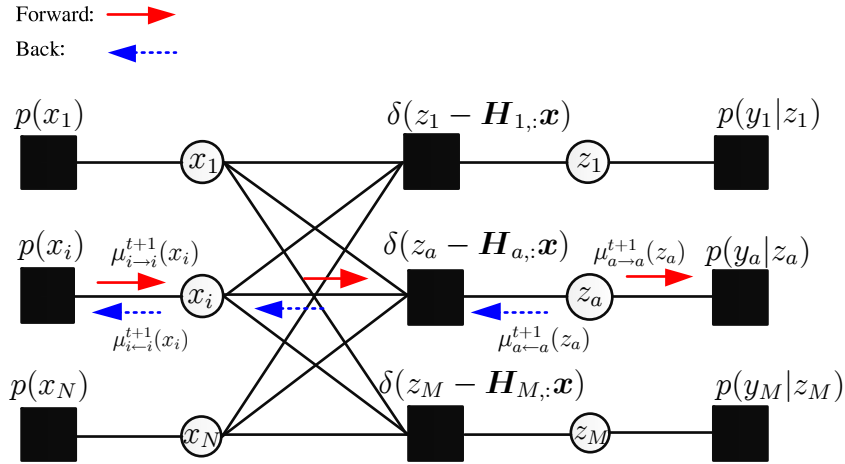


图4-4 广义线性模型的分列式因子图

如图4-4所示, 消息( $\mu_{a \rightarrow a}^t(z_a), \mu_{a \leftarrow a}^t(z_a), \mu_{i \rightarrow i}^{t+1}(x_i), \mu_{i \leftarrow i}^{t+1}(x_i)$ )已经在图中标识, 此外 $\mu_{i \rightarrow a}^{t+1}(x_i)$ 表示第 $t+1$ 次迭代中从因子节点 $\delta(z_a - \mathbf{H}_{a,:}\mathbf{x})$ 传递到变量节点 $x_i$ 的消息。 $\mu_{i \leftarrow a}^{t+1}(x_i)$ 表示第 $t+1$ 次迭代中从变量节点 $x_i$ 传递到因子节点 $\delta(z_a - \mathbf{H}_{a,:}\mathbf{x})$ 的消息。

基于分列式因子图以及匹配消息更新规则, 可以得到如下消息更新方程

$$\begin{aligned}
 \mu_{a \leftarrow a}^t(z_a) &\propto \frac{\text{Proj}_{\Phi}[\mu_{a \rightarrow a}^t(z_a)p(y_a|z_a)d]}{\mu_{a \rightarrow a}^t(z_a)} \\
 \mu_{i \leftarrow a}^t(x_i) &\propto \frac{\text{Proj}_{\Phi}[\mu_{i \rightarrow a}^t(x_i)f_{i \leftarrow a}^t(x_i)]}{\mu_{i \rightarrow a}^t(x_i)} \\
 \mu_{i \leftarrow a}^t(x_i) &\propto \prod_{b=1}^N \mu_{i \leftarrow b}^t(x_i) \\
 \mu_{i \rightarrow i}^{t+1}(x_i) &\propto \frac{\text{Proj}_{\Phi}[p(x_i)\mu_{i \leftarrow i}^t(x_i)]}{\mu_{i \leftarrow i}^t(x_i)} \\
 \mu_{i \rightarrow a}^{t+1}(x_i) &\propto \mu_{i \rightarrow i}^{t+1}(x_i) \prod_{b \neq a} \mu_{i \leftarrow b}^t(x_i) \\
 \mu_{a \rightarrow a}^{t+1}(z_a) &\propto \frac{\text{Proj}_{\Phi}[\mu_{a \leftarrow a}^t(z_a)f_{a \rightarrow a}^{t+1}(z_a)]}{\mu_{a \leftarrow a}^t(z_a)}
 \end{aligned} \tag{4.39}$$

合并(4.39)第3、4、5式，有

$$\begin{aligned}
 \mu_{a \leftarrow a}^t(z_a) &\propto \frac{\text{Proj}_{\Phi}[\mu_{a \rightarrow a}^t(z_a)p(y_a|z_a)]}{\mu_{a \rightarrow a}^t(z_a)} \\
 \mu_{i \leftarrow a}^t(x_i) &\propto \frac{\text{Proj}_{\Phi}[\mu_{i \rightarrow a}^t(x_i)f_{i \leftarrow a}^t(x_i)]}{\mu_{i \rightarrow a}^t(x_i)} \\
 \mu_{i \rightarrow a}^{t+1}(x_i) &\propto \frac{\text{Proj}_{\Phi}[p(x_i) \prod_b \mu_{i \leftarrow b}^t(x_i)]}{\mu_{i \leftarrow a}^t(x_i)} \\
 \mu_{a \rightarrow a}^{t+1}(z_a) &\propto \frac{\text{Proj}_{\Phi}[\mu_{a \leftarrow a}^t(z_a)f_{a \rightarrow a}^{t+1}(z_a)]}{\mu_{a \leftarrow a}^{t+1}(z_a)}
 \end{aligned} \tag{4.40}$$

其中

$$\begin{aligned}
 f_{i \leftarrow a}^t(x_i) &= \int \delta(z_a - \mathbf{H}_{ai}x_i) \mu_{a \leftarrow a}^t(z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) d\mathbf{x}_{\setminus i} dz_a \\
 f_{a \rightarrow a}^{t+1}(z_a) &= \int \delta(z_a - \mathbf{H}_{a,:}\mathbf{x}) \prod_{i=1}^N \mu_{i \rightarrow a}^{t+1}(x_i) d\mathbf{x}
 \end{aligned} \tag{4.41}$$

基于(4.40)给出的消息更新方程，以下给出详细化简步骤：

**步骤1：**初始化前向消息 $\mu_{a \rightarrow a}^t(z_a)$ ， $\mu_{i \rightarrow a}^t(x_i)$ 。消息更新的方向规定是从负向开始。相应地，初始化正向的消息为高斯分布

$$\mu_{a \rightarrow a}^t(z_a) = \mathcal{N}(z_a|Z_a^t, V_a^t), \quad \mu_{i \rightarrow a}^t(x_i) = \mathcal{N}_c(x_i|m_{i \rightarrow a}^t, v_{i \rightarrow a}^t) \tag{4.42}$$

**步骤2：**计算投影 $\text{Proj}_{\Phi}[\mu_{a \rightarrow a}^t(z_a)p(y_a|z_a)]$ 以及外部信息 $\mu_{a \leftarrow a}^{t+1}(z_a)$ 。

定义

$$\mathcal{N}_c(z_a|\tilde{z}_a^t, \tilde{v}_a^t) = \text{Proj}_{\Phi}[\mu_{a \rightarrow a}^t(z_a)p(y_a|z_a)] \tag{4.43}$$

其中

$$\tilde{z}_a^t \triangleq \mathbb{E}[z_a|Z_a^t, V_a^t], \quad \tilde{v}_a^t = \text{Var}[z_a|Z_a^t, V_a^t] \tag{4.44}$$



这里期望对概率分布  $\frac{p(y_a|z_a)\mathcal{N}_c(z_a|Z_a^t, V_a^t)}{\int p(y_a|z_a)\mathcal{N}_c(z_a|Z_a^t, V_a^t)dz_a}$  取。

根据高斯相乘引理，知外部信息  $\mu_{a \leftarrow a}^t(z_a)$  亦为高斯分布，即

$$\mu_{a \leftarrow a}^t(z_a) = \mathcal{N}_c(z_a | \tilde{Z}_a^t, \tilde{V}_a^t) \quad (4.45)$$

其中

$$\begin{aligned} \tilde{V}_a^t &= \left( \frac{1}{\tilde{v}_a} - \frac{1}{V_a^t} \right)^{-1} \\ \tilde{Z}_a^t &= \tilde{V}_a^t \left( \frac{\tilde{z}_a^t}{\tilde{v}_a^t} - \frac{Z_a^t}{V_a^t} \right) \end{aligned} \quad (4.46)$$

**步骤3:**  $f_{i \leftarrow a}^t(x_i)$  的高斯近似。

$$\begin{aligned} f_{i \leftarrow a}^t(x_i) &= \int \delta(z_a - \mathbf{H}_{ai}x_i) \mu_{a \leftarrow a}^t(z_a) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) d\mathbf{x}_{\setminus i} dz_a \\ &\propto \int \delta \left( x_i - \frac{1}{H_{ai}} \left( z_a - \sum_{j \neq i} H_{aj}x_j \right) \right) \mathcal{N}_c(z_a | Z_a^t, V_a^t) \prod_{j \neq i} \mu_{j \rightarrow a}^t(x_j) d\mathbf{x}_{\setminus i} dz_a \end{aligned} \quad (4.47)$$

定义随机变量  $\zeta_a^t \sim \mathcal{N}_c(z_a | Z_a^t, V_a^t)$ ,  $\xi_{j \rightarrow a}^t \sim \mu_{j \rightarrow a}^t(x_j)$ ,  $\xi_{i \leftarrow a}^t \sim f_{i \leftarrow a}^t(x_i)$ 。应用PDF-to-RV引理，有

$$\xi_{i \leftarrow a}^t = \frac{1}{H_{ai}} \left( \zeta_a^t - \sum_{j \neq i} H_{aj} \xi_{j \rightarrow a}^t \right) \quad (4.48)$$

因此，可以得到  $f_{i \leftarrow a}^t(x_i)$  亦为高斯分布，由于  $f_{i \leftarrow a}^t(x_i)$ 、 $\mu_{i \rightarrow a}^t(x_i)$  均为高斯分布，从而  $\mu_{i \leftarrow a}^t(x_i) = f_{i \leftarrow a}^t(x_i)$ ，即  $\mu_{i \leftarrow a}^t(x_i) = \mathcal{N}_c(x_i | m_{i \leftarrow a}^t, v_{i \leftarrow a}^t)$ ，其中

$$\begin{aligned} m_{i \leftarrow a}^t &= \frac{1}{H_{ai}} \left( \tilde{Z}_a^t - \sum_{j \neq i} H_{aj} m_{j \rightarrow a}^t \right) \\ v_{i \leftarrow a}^t &= \frac{1}{|H_{ai}|^2} \left( \tilde{V}_a^t - \sum_{j \neq i} |H_{aj}|^2 v_{j \rightarrow a}^t \right) \end{aligned} \quad (4.49)$$

为了简化符号，定义

$$Z_{i \leftarrow a}^t = \sum_{j \neq i} H_{aj} m_{j \rightarrow a}^t, \quad V_{i \leftarrow a}^t = \sum_{j \neq i} |H_{aj}|^2 v_{j \rightarrow a}^t \quad (4.50)$$

**步骤4:** 计算投影  $\text{Proj}_{\Phi}[p(x_i) \prod_b \mu_{i \leftarrow b}^t(x_i)]$  及外部信息  $\mu_{i \rightarrow a}^{t+1}(x_i)$ 。由高斯相乘引理，有

$$\prod_b \mu_{i \leftarrow b}^t(x_i) = \mathcal{N}_c(x_i | R_i^t, \Sigma_i^t) \quad (4.51)$$

其中

$$\begin{aligned}\Sigma_i^t &= \left( \sum_b \frac{1}{v_{i \leftarrow b}^t} \right)^{-1} \\ R_i^t &= \Sigma_i^t \left( \sum_b \frac{m_{i \leftarrow b}^t}{v_{i \leftarrow b}^t} \right)\end{aligned}\tag{4.52}$$

因此，计算投影可得

$$\mathcal{N}_c(x_i | \hat{m}_i^{t+1}, \hat{v}_i^{t+1}) = \text{Proj}_{\Phi} \left[ p(x_i) \prod_b \mu_{i \leftarrow b}^t(x_i) \right]\tag{4.53}$$

其中

$$\begin{aligned}\hat{m}_i^{t+1} &= \mathbb{E}[x_i | R_i^t, \Sigma_i^t] \\ \hat{v}_i^{t+1} &= \text{Var}[x_i | R_i^t, \Sigma_i^t]\end{aligned}\tag{4.54}$$

由高斯相乘引理，计算得到外部消息  $\mu_{i \rightarrow a}^{t+1}(x_i)$

$$\begin{aligned}\frac{1}{v_{i \rightarrow a}^{t+1}} &= \frac{1}{\hat{v}_i^{t+1}} - \frac{|H_{ai}|^2}{\tilde{V}_a^t + V_{i \leftarrow a}^t} \\ m_{i \rightarrow a}^{t+1} &= v_{i \rightarrow a}^{t+1} \left( \frac{\hat{m}_i^{t+1}}{\hat{v}_i^{t+1}} - \frac{H_{ai}^*(\tilde{Z}_a^t - Z_{i \leftarrow a}^t)}{\tilde{V}_a^t + V_{i \leftarrow a}^t} \right)\end{aligned}\tag{4.55}$$

由于  $|H_{ai}|^2 = O(\frac{1}{N})$ ,  $(i, a) \in [N] \times [M]$ , 忽略高阶项有

$$v_{i \rightarrow a}^{t+1} = \hat{v}_i^{t+1}\tag{4.56}$$

将(4.56)代入(4.55)第2式，有

$$m_{i \rightarrow a}^{t+1} = \hat{m}_i^{t+1} - \hat{v}_i^{t+1} \frac{H_{ai}^*(\tilde{Z}_a^t - Z_{i \leftarrow a}^t)}{\tilde{V}_a^t + V_{i \leftarrow a}^t}\tag{4.57}$$

**步骤5:**  $f_{a \rightarrow a}^{t+1}(z_a)$ 的高斯近似。与处理  $f_{i \leftarrow a}^t(x_i)$ 类似，由高斯相乘引理，知  $f_{a \rightarrow a}^{t+1}(z_a)$ 为高斯分布，因此  $\mu_{a \rightarrow a}^{t+1} = f_{a \rightarrow a}^{t+1}(z_a)$ ，即

$$f_{a \rightarrow a}^{t+1}(z_a) \propto \mathcal{N}_c(z_a | Z_a^{t+1}, V_a^{t+1})\tag{4.58}$$

其中

$$\begin{aligned}Z_a^{t+1} &= \sum_i H_{ai} m_{i \rightarrow a}^{t+1} \\ V_a^{t+1} &= \sum_i |H_{ai}|^2 v_{i \rightarrow a}^{t+1} = \sum_i |H_{ai}|^2 \hat{v}_i^{t+1}\end{aligned}\tag{4.59}$$

至此，我们完成了(4.40)迭代方程组的简化。

**步骤6:** 简化参数  $(Z_a^t, \Sigma_i^t, R_i^t)$ 。将(4.59)代入(4.57)中，有

$$m_{i \rightarrow a}^{t+1} = \hat{m}_i^{t+1} - \hat{v}_i^{t+1} \frac{H_{ai}^*(\tilde{Z}_a^t - Z_a^t)}{\tilde{V}_a^t + V_a^t}\tag{4.60}$$

将(4.60)代入(4.59)第一式，有

$$Z_a^t = \sum_i H_{ai} \hat{m}_i^t - V_a^t \frac{\tilde{Z}_a^{t-1} - Z_a^{t-1}}{\tilde{V}_a^{t-1} + V_a^{t-1}} \quad (4.61)$$

将(4.49)第2式，(4.50)，(4.59)第2式，代入(4.52)第1式中，有

$$\Sigma_i^t = \left( \sum_a \frac{|H_{ai}|^2}{\tilde{V}_a^t + V_a^t} \right) \quad (4.62)$$

联立(4.49)、(4.50)、(4.52)、(4.61)、(4.60)有

$$\begin{aligned} R_i^t &= \Sigma_i^t \left( \sum_a \frac{H_{ai}^* (\tilde{Z}_a^t - Z_{i \leftarrow a}^t)}{\tilde{V}_a^t + V_{i \leftarrow a}^t} \right) \\ &= \Sigma_i^t \left( \sum_a \frac{H_{ai}^* (\tilde{Z}_a^t - Z_a^t + |H_{ai}|^2 \hat{m}_i^t)}{\tilde{V}_a^t + V_a^t} \right) \\ &= \hat{m}_i^t + \Sigma_i^t \left( \sum_a \frac{H_{ai}^* (\tilde{Z}_a^t - Z_a^t)}{\tilde{V}_a^t + V_a^t} \right) \end{aligned} \quad (4.63)$$

为了与Rangan的GAMP迭代方程对应，定义

$$\begin{aligned} \tau_a^t &\triangleq \frac{1}{\tilde{V}_a^t + V_a^t} = \frac{V_a^t - \hat{v}_a^t}{(V_a^t)^2} \\ s_a^t &\triangleq \frac{\tilde{Z}_a^t - Z_a^t}{\tilde{V}_a^t - V_a^t} = \frac{\tilde{z}_a^t - Z_a^t}{V_a^t} \end{aligned} \quad (4.64)$$

相应地，参数 $(Z_a^t, \Sigma_i^t, R_i^t)$ 变为

$$\begin{aligned} Z_a^t &= \sum_i H_{ai} \hat{m}_i^t - V_a^t s_a^{t-1} \\ \Sigma_i^t &= \left( \sum_a |H_{ai}|^2 \tau_a^t \right)^{-1} \\ R_i^t &= \hat{m}_i^t + \Sigma_i^t \sum_a H_{ai}^* s_a^t \end{aligned} \quad (4.65)$$

结合参数 $(Z_a^t, V_a^t, \tau_a^t, s_a^t, \hat{m}_i^{t+1}, \hat{v}_i^{t+1})$ 构成GAMP的迭代方程式，GAMP算法见表4-1。

Remark:

- 本节提出了两种GAMP的推导方法，包括：一种基于传统因子图和矩匹配消息更新规则的GAMP推导方法，另外一种为基于分列式因子图和矩匹配消息传递消息更新规则。
- 第一种推导方法相对于第二种而言，有进一步消息的省略，见(4.29)、(4.31)。而第

二种推导方法可以完美推导GAMP算法。

- 此外，第二种推导方法可以扩展到多层情况，用于推导多层广义近似消息传递[17]算法，这也是我们未来的工作重心。

表4-1 广义近似消息传递算法

### 算法 3: 广义近似消息传递算法

1. 初始化:  $\hat{x}_i^1 = \int x_i p(x_i) dx_i$ ,  $\hat{v}_i^1 = \int (x_i - \hat{x}_i^1)^2 p(x_i) dx_i$ ,  $V_a^0 = 1$ ,  $Z_a^0 = y_a$

2. 迭代 ( $t = 1 : T$ )

$$\begin{aligned} V_a^t &= \sum_i |H_{ai}|^2 \hat{v}_i^t \\ Z_a^t &= \sum_i H_{ai} \hat{m}_i^t - s_a^{t-1} V_a^t \\ \tilde{z}_a^t &= \mathbb{E}_{\eta_a^t}[\eta_a^t] \\ \tilde{v}_a^t &= \text{Var}_{\eta_a^t}[\eta_a^t] \\ s_a^t &= \frac{\tilde{z}_a^t - Z_a^t}{V_a^t} \\ \tau_a^t &= \frac{V_a^t - \tilde{v}_a^t}{(V_a^t)^2} \\ \Sigma_i^t &= \left( \sum_a |H_{ai}|^2 \tau_a^t \right)^{-1} \\ R_i^t &= \hat{m}_i^t + \Sigma_i^t \sum_a H_{ai}^* s_a^t \\ \hat{m}_i^{t+1} &= \mathbb{E}[x_i | R_i^t, \Sigma_i^t] \\ \hat{v}_i^{t+1} &= \text{Var}[x_i | R_i^t, \Sigma_i^t] \end{aligned}$$

3. 输出:  $\hat{m}_i^{T+1}$ ,  $\tilde{p}(x_i | \mathbf{y}) = \frac{p(x_i) \mathcal{N}(x_i | R_i^T, \Sigma_i^T)}{\int p(x_i) \mathcal{N}(x_i | R_i^T, \Sigma_i^T) dx_i}$ 。

## 4.4 广义期望一致信号重构

### 4.4.1 概述

#### ■ 期望一致算法

广义期望一致<sup>[42]</sup> (generalized expectation consistent, GEC) 算法，是期望一致 (expectation consistent, EC) [35] 算法的扩展。Oppor等人于2005年提出EC算法。EC考虑典型的场景，即联合概率分布  $p(\mathbf{u})$  可以分解为两个函数之积，即

$p(\mathbf{u}) = \frac{1}{Z} f(\mathbf{u}) = \frac{1}{Z} f_r(\mathbf{u}) f_q(\mathbf{u})$ , 其中  $f_r(\mathbf{u})$  具有复杂的函数形式,  $Z$  为配分函数。EC 采用指数族函数替换  $f_r(\mathbf{u})$ , 其配分函数为  $Z_q$

$$\begin{aligned} Z &= Z_q \frac{Z}{Z_q} = Z_q \frac{\int d\mathbf{u} f_r(\mathbf{u}) f_q(\mathbf{u}) \exp(\boldsymbol{\lambda}_q - \boldsymbol{\lambda}_q)^T g(\mathbf{u})}{\int d\mathbf{u} f_q(\mathbf{u}) \exp(\boldsymbol{\lambda}_q^T g(\mathbf{u}))} \\ &= Z_q \langle f_r(\mathbf{u}) \exp(-\boldsymbol{\lambda}_q^T g(\mathbf{u})) \rangle_q \end{aligned} \quad (4.66)$$

其中  $\langle \cdot \rangle_q$  表示对  $q(\mathbf{u}) = \frac{f_q(\mathbf{u}) \exp(\boldsymbol{\lambda}_q^T g(\mathbf{u}))}{\int f_q(\mathbf{u}) \exp(\boldsymbol{\lambda}_q^T g(\mathbf{u})) d\mathbf{u}}$  取期望。此外, EC 利用指数族函数全部替换  $f_r(\mathbf{u}) f_q(\mathbf{u})$ , 即所得分布为  $s(\mathbf{u}) = \frac{1}{Z_r(\boldsymbol{\lambda})} \exp(\boldsymbol{\lambda}_s^T g(\mathbf{u}))$ , 并用  $s(\mathbf{u})$  替换  $q(\mathbf{u})$ , 得

$$Z^{\text{EC}} = Z_q \langle f_r(\mathbf{u}) \exp(-\boldsymbol{\lambda}_q^T g(\mathbf{u})) \rangle_s \quad (4.67)$$

取对数, 有

$$-\ln Z^{\text{EC}}(\boldsymbol{\lambda}_q, \boldsymbol{\lambda}_s) = -\ln Z_q - \ln \langle f_r(\mathbf{u}) \exp(-\boldsymbol{\lambda}_q^T g(\mathbf{u})) \rangle_s \quad (4.68)$$

计算  $Z^{\text{EC}}(\boldsymbol{\lambda}_q, \boldsymbol{\lambda}_s)$  关于  $(\boldsymbol{\lambda}_q, \boldsymbol{\lambda}_s)$  的驻点

$$\begin{aligned} \frac{\partial \ln Z^{\text{EC}}}{\partial \boldsymbol{\lambda}_q} &= 0 \Rightarrow \langle g(\mathbf{u}) \rangle_q = \langle g(\mathbf{u}) \rangle_r \\ \frac{\partial \ln Z^{\text{EC}}}{\partial \boldsymbol{\lambda}_s} &= 0 \Rightarrow \langle g(\mathbf{u}) \rangle_r = \langle g(\mathbf{u}) \rangle_s \end{aligned} \quad (4.69)$$

其中

$$\begin{aligned} q(\mathbf{u}) &= \frac{1}{Z_q(\boldsymbol{\lambda}_q)} f_q(\mathbf{u}) \exp(\boldsymbol{\lambda}_q^T g(\mathbf{u})) \\ r(\mathbf{u}) &= \frac{1}{Z_r(\boldsymbol{\lambda}_r)} f_r(\mathbf{u}) \exp(\boldsymbol{\lambda}_r^T g(\mathbf{u})) \\ s(\mathbf{u}) &= \frac{1}{Z_r(\boldsymbol{\lambda})} \exp(\boldsymbol{\lambda}_s^T g(\mathbf{u})) \end{aligned} \quad (4.70)$$

这里  $\boldsymbol{\lambda}_r = \boldsymbol{\lambda}_s - \boldsymbol{\lambda}_q$ 。计算。通过解(4.69)方程, 得到EC的迭代式。EC算法包括单循环算法和双循环算法, 其中单循环算法被证明与EP算法等价。

## ■ 广义期望一致

[42]提出了广义期望一致 (generalized expectation consistent, GEC) 算法, 该算法扩展了EC的适用范围。此外, 该文献分析了GEC所能达到的不动点方程, 然而GEC的收敛性很差。为了克服GEC收敛困难的问题, [30]提出了广义期望一致信号重构 (generalized expectation consistent, GEC-SR) 算法。GEC-SR可以达到GEC的性能。除此之外, GEC-SR比GEC具有更快收敛的特点, 而GEC在某些特定的场景存在曲线发散的现象。[38]利用RM分析广义线性模型的不动点方程, 从不动点方程反推出, 通过加入

一个线性空间模块，来提高GEC的稳定性。其本质是，将似然函数 $p(\mathbf{y}|\mathbf{x})$ 分裂成一个线性空间模块 $\delta(\mathbf{z} - \mathbf{H}\mathbf{x})$ 和一个分量式模块 $p(\mathbf{y}|\mathbf{z})$ 。但是，[38]、[30]并未给出GEC-SR的推导过程。本文根据矩匹配消息更新规则和分列式因子图给出GEC-SR的详细推导过程。

#### 4.4.2 算法推导

广义线性模型的矢量因子图，如图4-5所示

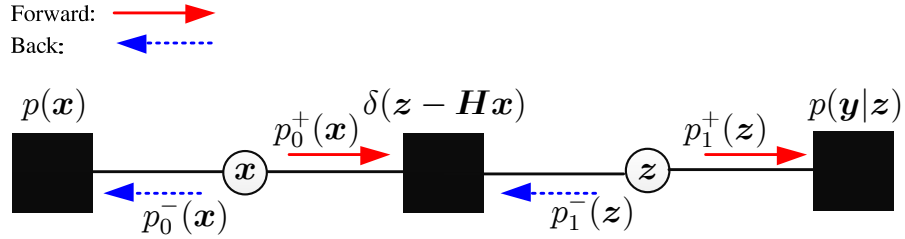


图4-5 广义线性模型的矢量因子图

图中消息的上标表示方向，‘+’表示正向，‘-’表示负向；图中下标表示消息所在的层数，这里根据变量节点进行分层，如广义线性模型包含变量节点 $\mathbf{x}, \mathbf{z}$ ，因此，广义线性模型的矢量因子图有2层。由于负方向消息的更新包含观测值 $\mathbf{y}$ ，因此，消息更新的初始方向规定为负向。消息更新的规则，采用图3-4所示的矩匹配消息更新规则。

以下为GEC-SR的详细推导步骤：

**步骤1：**初始化。GEC-SR消息的更新方向规定是负向。在进行消息更新之前，对正向的消息初始化为高斯分布，即 $p_1^+(\mathbf{z}) = \mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+)$ ， $p_0^-(\mathbf{x}) = \mathcal{N}_c(\mathbf{x}|\mathbf{m}_0^+, \mathbf{v}_0^+)$ 。

**步骤2：**计算因子节点 $p(\mathbf{y}|\mathbf{z})$ 的PME以及外部消息 $p_1^-(\mathbf{z})$ 。

$$\mathcal{N}_c(\mathbf{z}|\hat{\mathbf{z}}^-, \mathbf{v}_z^-) = \text{Proj}_z [p(\mathbf{y}|\mathbf{z})\mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+)] \quad (4.71)$$

其中

$$\begin{aligned} \hat{\mathbf{z}}^- &= \mathbb{E}[\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+] \\ \mathbf{v}_z^- &= \text{Var}[\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+] \end{aligned} \quad (4.72)$$

这里期望对分布 $\frac{p(\mathbf{y}|\mathbf{z})\mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+)}{\int p(\mathbf{y}|\mathbf{z})\mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+)d\mathbf{z}}$ 取。

根据高斯相乘引理，知外部信息 $p_1^-(\mathbf{z})$ 亦为高斯分布，即

$$\mathcal{N}_c(\mathbf{z}|\mathbf{m}^-, \mathbf{v}_z^-) \propto \frac{\mathcal{N}_c(\mathbf{z}|\hat{\mathbf{z}}^-, \mathbf{v}_z^-)}{\mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+)} \quad (4.73)$$

其中

$$\begin{aligned} \mathbf{v}_1^- &= \mathbf{1} \oslash (\mathbf{1} \oslash \mathbf{v}_z^- - \mathbf{1} \oslash \mathbf{v}_1^+) \\ \mathbf{m}_1^- &= \mathbf{v}_1^- \odot (\hat{\mathbf{z}}^- \oslash \mathbf{v}_z^- - \mathbf{m}_1^+ \oslash \mathbf{v}_1^+) \end{aligned} \quad (4.74)$$

**步骤3:** 计算线性空间负向输出以及外部信息 $p_0^-(\mathbf{x})$ 。

$$\begin{aligned} &\mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}^-, \text{Diag}(\mathbf{v}_x^-)) \\ &= \text{Proj}_{\mathbf{x}} \left[ \int p(\mathbf{z}|\mathbf{x}) \mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \text{Diag}(\mathbf{v}_0^+)) \mathcal{N}(\mathbf{z}|\mathbf{m}_1^-, \text{Diag}(\mathbf{v}_1^-)) d\mathbf{z} \right] \end{aligned} \quad (4.75)$$

经化简

$$\begin{aligned} &\int_{\mathbf{z}} \delta(\mathbf{z} - \mathbf{H}\mathbf{x}) \mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \text{Diag}(\mathbf{v}_0^+)) \mathcal{N}(\mathbf{z}|\mathbf{m}_1^-, \text{Diag}(\mathbf{v}_1^-)) d\mathbf{z} \\ &= \mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \text{Diag}(\mathbf{v}_0^+)) \mathcal{N}(\mathbf{H}\mathbf{x}|\mathbf{m}_1^-, \text{Diag}(\mathbf{v}_1^-)) \\ &\propto \mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \text{Diag}(\mathbf{v}_0^+)) \\ &\quad \times \mathcal{N}(\mathbf{x} | (\mathbf{H}^H \text{Diag}(\mathbf{1} \oslash \mathbf{v}_1^-) \mathbf{H})^{-1} \mathbf{H}^H \text{Diag}(\mathbf{1} \oslash \mathbf{v}_1^-) \mathbf{m}_1^-, (\mathbf{H}^H \text{Diag}(\mathbf{1} \oslash \mathbf{v}_1^-) \mathbf{H})^{-1}) \end{aligned} \quad (4.76)$$

由高斯相乘引理，进而得到

$$\begin{aligned} \mathbf{Q}_x^- &= (\mathbf{H}^H \text{Diag}(\mathbf{1} \oslash \mathbf{v}_1^-) \mathbf{H} + \text{Diag}(\mathbf{1} \oslash \mathbf{v}_0^+))^{-1} \\ \hat{\mathbf{x}}^- &= \mathbf{Q}_x^- (\mathbf{H}^H \text{Diag}(\mathbf{1} \oslash \mathbf{v}_1^-) \mathbf{m}_1^- + \mathbf{m}_0^+ \oslash \mathbf{v}_0^+) \\ \mathbf{v}_x^- &= \text{diag}(\mathbf{Q}_x^-) \end{aligned} \quad (4.77)$$

由高斯相乘引理，可知外部信息 $p_0^-(\mathbf{x})$ 亦为高斯分布，即

$$\mathcal{N}(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-) \propto \frac{\mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}^-, \mathbf{v}_x^-)}{\mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \mathbf{v}_0^+)} \quad (4.78)$$

其中

$$\begin{aligned} \mathbf{v}_0^- &= \mathbf{1} \oslash (\mathbf{1} \oslash \mathbf{v}_x^- - \mathbf{1} \oslash \mathbf{v}_0^+) \\ \mathbf{m}_0^- &= \mathbf{v}_1^- \odot (\hat{\mathbf{x}}^- \oslash \mathbf{v}_x^- - \mathbf{m}_0^+ \oslash \mathbf{v}_0^+) \end{aligned} \quad (4.79)$$

**步骤4:** 计算因子节点 $p(\mathbf{x})$ 的PME以及外部信息 $p_0^+(\mathbf{x})$ 。

$$\mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}^+, \mathbf{v}_x^+) = \text{Proj}_{\mathbf{x}} [p(\mathbf{x}) \mathcal{N}(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)] \quad (4.80)$$

其中

$$\begin{aligned} \hat{\mathbf{x}}^+ &= \mathbb{E}[\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-] \\ \mathbf{v}_x^+ &= \text{Var}[\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-] \end{aligned} \quad (4.81)$$

这里期望对概率分布 $\frac{p(\mathbf{x}) \mathcal{N}(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)}{\int p(\mathbf{x}) \mathcal{N}(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-) d\mathbf{x}}$ 取。

由高斯相乘引理，知外部信息 $p_0^+(\mathbf{x})$ 也是高斯分布，即

$$\mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \mathbf{v}_0^+) \propto \frac{\mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}^+, \mathbf{v}_x^+)}{\mathcal{N}(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)} \quad (4.82)$$

其中

$$\begin{aligned} \mathbf{v}_0^+ &= \mathbf{1} \oslash (\mathbf{1} \oslash \mathbf{v}_x^+ - \mathbf{1} \oslash \mathbf{v}_0^-) \\ \mathbf{m}_0^+ &= \mathbf{v}_0^+ \odot (\hat{\mathbf{x}}^+ \oslash \mathbf{v}_x^+ - \mathbf{m}_0^- \oslash \mathbf{v}_0^-) \end{aligned} \quad (4.83)$$

**步骤5:** 计算线性空间的正向输出及外部信息 $p_1^+(\mathbf{z})$ 。

$$\begin{aligned} &\mathcal{N}(\mathbf{z}|\hat{\mathbf{z}}^+, \text{Diag}(\mathbf{v}_z^+)) \\ &= \text{Proj}_{\mathbf{z}} \left[ \int p(\mathbf{z}|\mathbf{x}) \mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \text{Diag}(\mathbf{v}_0^+)) \mathcal{N}(\mathbf{z}|\mathbf{m}_1^-, \text{Diag}(\mathbf{v}_1^-)) d\mathbf{x} \right] \end{aligned} \quad (4.84)$$

简化

$$\begin{aligned} &\int_{\mathbf{x}} p(\mathbf{z}|\mathbf{x}) \mathcal{N}(\mathbf{x}|\mathbf{m}_0^+, \text{Diag}(\mathbf{v}_0^+)) \mathcal{N}(\mathbf{z}|\mathbf{m}_1^-, \text{Diag}(\mathbf{v}_1^-)) d\mathbf{x} \\ &= \int_{\mathbf{x}} \delta(\mathbf{z} - \mathbf{H}\mathbf{x}) \mathcal{N}_c(\mathbf{x}|\mathbf{m}^+, \mathbf{Q}_x^+) d\mathbf{x} \end{aligned} \quad (4.85)$$

其中利用到关系式 $\mathbf{z} = \mathbf{H}\mathbf{x}$ 以及

$$\begin{aligned} \mathbf{Q}_x^+ &= (\mathbf{H}^H \text{Diag}(\mathbf{1} \oslash \mathbf{v}_1^-) \mathbf{H} + \text{Diag}(\mathbf{1} \oslash \mathbf{v}_0^+))^{-1} \\ \mathbf{m}_x^+ &= \mathbf{Q}_x^+ (\mathbf{H} \text{Diag}(\mathbf{1} \oslash \mathbf{v}_1^-) \mathbf{m}_1^- + \mathbf{m}_0^+ \oslash \mathbf{v}_0^+) \end{aligned} \quad (4.86)$$

根据PDF-to-RV引理（引理2-3），有

$$\begin{aligned} \hat{\mathbf{z}}^+ &= \mathbf{H}\mathbf{m}_x^+ \\ \mathbf{v}_z^+ &= \text{diag}(\mathbf{H}\mathbf{Q}_x^+ \mathbf{H}^H) \end{aligned} \quad (4.87)$$

根据高斯相乘引理（引理2-1），知 $p_1^+(\mathbf{z})$ 亦为高斯分布，即

$$\mathcal{N}(\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+) \propto \frac{\mathcal{N}(\mathbf{z}|\hat{\mathbf{z}}^+, \mathbf{v}_z^+)}{\mathcal{N}(\mathbf{z}|\mathbf{m}_1^-, \mathbf{v}_1^-)} \quad (4.88)$$

其中

$$\begin{aligned} \mathbf{v}_1^+ &= \mathbf{1} \oslash (\mathbf{1} \oslash \mathbf{v}_z^+ - \mathbf{1} \oslash \mathbf{v}_1^-) \\ \mathbf{m}_1^+ &= \mathbf{v}_1^+ \odot (\hat{\mathbf{z}}^+ \oslash \mathbf{v}_z^+ - \mathbf{m}_1^- \oslash \mathbf{v}_1^-) \end{aligned} \quad (4.89)$$

基于上述推导（步骤1-步骤5），可以得到GEC-SR算法，如表4-2所示。此外，为了方便理解GEC-SR算法，本文给出了GEC-SR的系统框图，如图4-5所示。

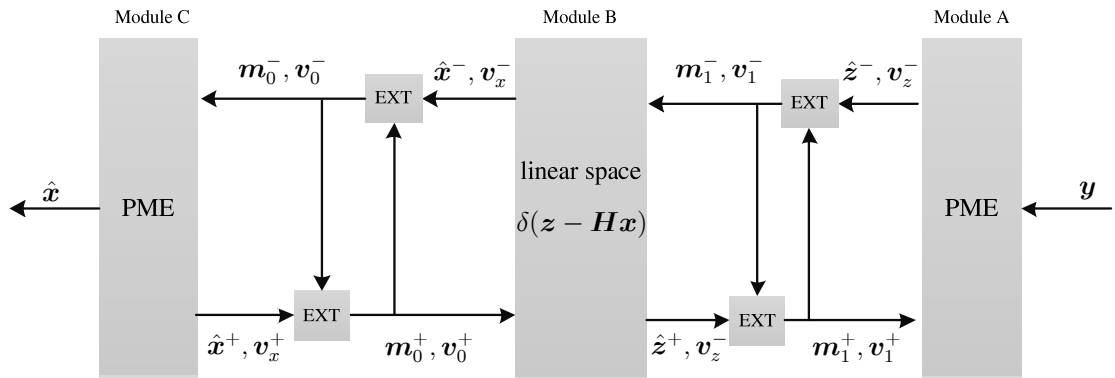


图4-6 GEC-SR系统框图



观测向量 $\mathbf{y}$ , 进入模块A, 计算得到投影值 $\mathcal{N}_c(\mathbf{z}|\hat{\mathbf{z}}^-, \mathbf{v}_z^-)$ 。随后计算扣除 $\mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+)$ , 得到外部信息 $\mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^-, \mathbf{v}_1^-)$ 。 $\mathcal{N}_c(\mathbf{z}|\mathbf{m}_1^-, \mathbf{v}_1^-)$ 和 $\mathcal{N}_c(\mathbf{m}_0^+, \mathbf{v}_0^+)$ 作为模块B的输入, 投影得到输出 $\mathcal{N}_c(\mathbf{x}|\hat{\mathbf{x}}, \mathbf{v}_x^-)$ 。扣除 $\mathcal{N}_c(\mathbf{x}|\mathbf{m}_0^+, \mathbf{v}_0^+)$ , 得到外部信息 $\mathcal{N}_c(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)$ 。随后,  $\mathcal{N}_c(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)$ 作为模块C的输入, 计算估计值 $\hat{\mathbf{x}}$ , 这里概率分布 $\mathcal{N}_c(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)$ 可以看作是对似然函数的近似。同理, 按照这样的消息传递方式, 完成正向的消息的更新。总而言之, GEC-SR的信号流程 $A \rightarrow B \rightarrow C \rightarrow B \rightarrow A \cdots$ 。

表4-2 广义期望一致信号重构

**算法4: GEC-SR**

1. 初始化:  $t = 1, \mathbf{m}_1^+ = \mathbf{0}, \mathbf{v}_1^+ = \mathbf{1}, \mathbf{m}_0^+ = \mathbf{0}, \mathbf{v}_0^+ = \mathbf{1}$

2. 迭代 ( $t < T$ )

$$\begin{aligned}
 \hat{\mathbf{z}}^- &= \mathbb{E}[\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+] \\
 \mathbf{v}_z^- &= \text{Var}[\mathbf{z}|\mathbf{m}_1^+, \mathbf{v}_1^+] \\
 \mathbf{v}_1^- &= \mathbf{1} \odot (\mathbf{1} \odot \mathbf{v}_z^- - \mathbf{1} \odot \mathbf{v}_1^+) \\
 \mathbf{m}_1^- &= \mathbf{v}_1^- \odot (\hat{\mathbf{z}}^- \odot \mathbf{v}_z^- - \mathbf{m}_1^+ \odot \mathbf{v}_1^+) \\
 \mathbf{Q}_x^- &= (\mathbf{H}^H \text{Diag}(\mathbf{1} \odot \mathbf{v}_1^-) \mathbf{H} + \text{Diag}(\mathbf{1} \odot \mathbf{v}_0^+))^{-1} \\
 \hat{\mathbf{x}}^- &= \mathbf{Q}_x^- (\mathbf{H}^H \text{Diag}(\mathbf{1} \odot \mathbf{v}_1^-) \mathbf{m}_1^- + \mathbf{m}_0^+ \odot \mathbf{v}_0^+) \\
 \mathbf{v}_x^- &= \text{diag}(\mathbf{Q}_x^-) \\
 \mathbf{v}_0^- &= \mathbf{1} \odot (\mathbf{1} \odot \mathbf{v}_x^- - \mathbf{1} \odot \mathbf{v}_0^+) \\
 \mathbf{m}_0^- &= \mathbf{v}_0^- \odot (\hat{\mathbf{x}}^- \odot \mathbf{v}_x^- - \mathbf{m}_0^+ \odot \mathbf{v}_0^+) \\
 \hat{\mathbf{x}}^+ &= \mathbb{E}[\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-] \\
 \mathbf{v}_x^+ &= \text{Var}[\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-] \\
 \mathbf{v}_0^+ &= \mathbf{1} \odot (\mathbf{1} \odot \mathbf{v}_x^+ - \mathbf{1} \odot \mathbf{v}_0^-) \\
 \mathbf{m}_0^+ &= \mathbf{v}_0^+ \odot (\hat{\mathbf{x}}^+ \odot \mathbf{v}_x^+ - \mathbf{m}_0^- \odot \mathbf{v}_0^-) \\
 \mathbf{Q}_x^+ &= (\mathbf{H}^H \text{Diag}(\mathbf{1} \odot \mathbf{v}_1^-) \mathbf{H} + \text{Diag}(\mathbf{1} \odot \mathbf{v}_0^+))^{-1} \\
 \mathbf{m}_x^+ &= \mathbf{Q}_x^+ (\mathbf{H} \text{Diag}(\mathbf{1} \odot \mathbf{v}_1^-) \mathbf{m}_1^- + \mathbf{m}_0^+ \odot \mathbf{v}_0^+) \\
 \hat{\mathbf{z}}^+ &= \mathbf{H} \mathbf{m}_x^+ \\
 \mathbf{v}_z^+ &= \text{diag}(\mathbf{H} \mathbf{Q}_x^+ \mathbf{H}^H) \\
 \mathbf{v}_1^+ &= \mathbf{1} \odot (\mathbf{1} \odot \mathbf{v}_z^+ - \mathbf{1} \odot \mathbf{v}_1^-) \\
 \mathbf{m}_1^+ &= \mathbf{v}_1^+ \odot (\hat{\mathbf{z}}^+ \odot \mathbf{v}_z^+ - \mathbf{m}_1^- \odot \mathbf{v}_1^-)
 \end{aligned}$$

3. 输出:  $\hat{\mathbf{x}}^+, \tilde{p}(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})\mathcal{N}(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)}{\int p(\mathbf{x})\mathcal{N}(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)d\mathbf{x}}$

### 4.4.3 状态更新方程

GEC-SR的渐进性能可以被SE方程进行预测。[38]利用RM分析系统的不动点方程，发现，利用RM分析得到的不动点方程与算法对观测值和观测矩阵的平均所得到的迭代式一致。此外，通过仿真验证，两者的性能完全重合。从而证明了GEC-SR的最优性。然而，该文献给出的RM推导过程存在着多处错误以及笔误，因此，该推导方案较为晦涩。为了方便理解，本文给出了从算法角度出发，通过对观测矩阵和观测值的平均，得到SE方程。

如(4.81)所描述，原始信号 $\mathbf{x}$ 的PME表示为

$$\hat{\mathbf{x}} = \int \mathbf{x} \frac{p(\mathbf{x})\mathcal{N}_c(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)}{\int p(\mathbf{x})\mathcal{N}_c(\mathbf{x}|\mathbf{m}_0^-, \mathbf{v}_0^-)d\mathbf{x}} d\mathbf{x} \quad (4.90)$$

在大规模系统中，该分布可以通过一个等效的标量模型来表示

$$M_0^- = X + N \sim \mathcal{N}_c(0, v_0^-) \quad (4.91)$$

该模型的MMSE估计表示为 $\hat{x} = \int x p(x|M_0^-)dx$ ，其中 $p(x|M_0^-) = \frac{p(x)\mathcal{N}_c(x|m_0^-, v_0^-)}{\int p(x)\mathcal{N}_c(x|m_0^-, v_0^-)dx}$ 。对应的均方误差表示为

$$\text{MSE}(v_0^-) = \mathbb{E} [|\hat{x} - X|^2] \quad (4.92)$$

其中期望对概率分布 $p(x, m_0^-)$ 取。该MSE还可以理解为估计器的方差对观测值的平均，即

$$\text{MSE} = \mathbb{E}_{M_0^-} [\hat{v}_x] = \int_{m_0^-} \left( \int_x (x - \hat{x})^2 p(x|m_0^-) dx \right) p(m_0^-) dm_0^- \quad (4.93)$$

通过(4.91)的标量模型表示，则估计器的MSE仅仅取决于方差 $v_0^-$ 。为此，我们需要从算法4（表4-2）中提取出于 $v_0^-$ 有关的变量。从表中可以观察到，跟 $v_0^-$ 有关的变量为 $(\mathbf{v}_x^+, \mathbf{v}_0^+, \mathbf{v}_z^+, \mathbf{v}_1^+, \mathbf{v}_z^-, \mathbf{v}_1^-, \mathbf{v}_x^-, \mathbf{v}_0^-)$ 。

为了简化符号，定义

$$\gamma_z^+ = \frac{1}{v_z^+}, \gamma_z^- = \frac{1}{v_z^-}, \gamma_x^+ = \frac{1}{v_x^+}, \gamma_x^- = \frac{1}{v_x^-} \quad (4.94)$$

为了消除算法对矩阵 $\mathbf{H}$ 的依赖，基于SVD分解，有 $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ 。应用SVD分解，可以得到

$$\begin{aligned} v_z^+ &= \frac{1}{M} \text{Tr} \left( \mathbf{H} (\gamma_1^- \mathbf{H}^H \mathbf{H} + \gamma_0^+ \mathbf{I})^{-1} \mathbf{H}^H \right) \\ &= \frac{1}{M} \sum_i^N \frac{\lambda_i}{\gamma_1^- \lambda_i + \gamma_0^+} \end{aligned} \quad (4.95)$$

同理，可得

$$v_x^- = \frac{1}{N} \sum_{i=1}^N \frac{1}{\gamma_1^- \lambda_i + \gamma_0^+} \quad (4.96)$$

这些参数中，较难处理得是变量 $v_z^-$ 。该该变量中， $\mathbf{y}$ 和 $\mathbf{m}_1^+$ 均与观测向量有关。因此

$$v_z^- = \mathbb{E}_{M_1^+, Y} [\text{Var}[z|M_1^+, \gamma_1^+]] \quad (4.97)$$

其中，外层期望对联合概率 $p(\mathbf{m}_1^+, y)$ 取。而，该概率分布与具体的转移概率 $p(y|z)$ 直接相关。对于特定的场景有解析解，如低精度ADC信号检测。在本章的4.5小节中将介绍低精度ADC信号检测中，关于 $p(\mathbf{m}_1^+, y)$ 的具体操作。

根据上述推导，可以得到广义期望一致信号重构算法的SE方程如下

$$\begin{aligned} v_z^- &= \mathbb{E}_{M_1^+, Y} [\text{Var}[z|M_1^+, \gamma_1^+]] \\ \gamma_1^- &= \frac{1}{v_z^-} - \gamma_1^+ \\ v_x^- &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\gamma_1^- \lambda_i + \gamma_0^+} \\ \gamma_0^- &= \frac{1}{v_x^-} - \gamma_0^+ \\ \gamma_0^+ &= \frac{1}{\text{MSE}(\gamma_0^-)} - \gamma_0^- \\ v_z^+ &= \frac{1}{M} \sum_{i=1}^N \frac{\lambda_i}{\gamma_1^- \lambda_i + \gamma_0^+} \\ \gamma_1^+ &= \frac{1}{v_z^+} - \gamma_1^- \end{aligned} \quad (4.98)$$

#### 4.5 应用：带低精度ADC的Massive MIMO信号检测

本章应用部分研究特定的通信系统中的信号检测问题。与3.4节中描述的场景所不同的是，本节将考虑更为实际的带低精度ADC的Massive MIMO系统。我们考虑上行通信链路系统，该系统中发射机装备 $N$ 根发射天线，接收方装备 $M$ 根接收天线( $M \geq N$ )。该系统可以抽象成如下数学模型

$$\begin{aligned} \text{Linear part : } \mathbf{z} &= \mathbf{H}\mathbf{x} \\ \text{Non-linear part : } \mathbf{y} &= Q_c(\mathbf{z} + \mathbf{w}) \end{aligned} \quad (4.99)$$

该模型包括两个部分，第一部分部分：用户数据 $\mathbf{x} \in \mathbb{C}^N$ 经过信道 $\mathbf{H} \in \mathbb{C}^{M \times N}$ 得到中间变量 $\mathbf{z} \in \mathbb{C}^M$ 。叠加噪声后，接收天线接收到模拟信号 $\tilde{\mathbf{y}} = \mathbf{z} + \mathbf{w}$ 。经过模数转换后得观测值 $\mathbf{y}$ ，其中模数转换用函数 $Q_c(\cdot)$ 表示，即

$$y_a = Q_c(\tilde{y}_a) = Q(\text{Re}(\tilde{y}_a)) + Q(\text{Im}(\tilde{y}_a)), \quad a = 1, \dots, M \quad (4.100)$$

其中  $Q(\cdot) : \mathbb{R} \mapsto \mathcal{R}_B$ ,  $\mathcal{R}_B$  表示ADC的输出集合, 定义如下

$$\mathcal{R}_B \triangleq \left\{ \left( -\frac{1}{2} + b \right) S, \quad b = -\frac{2^B}{2} + 1, \dots, \frac{2^B}{2} \right\} \quad (4.101)$$

其中  $S$  表示量化步长,  $B$  表示ADC的比特数。换言之, 观测值  $\mathbf{y}$  的取值来自于集合  $\mathcal{R}_B$ 。

对于ADC的输出  $y_a$ , 所对应的ADC输入的上界和下界表示为

$$\begin{aligned} y_a^{\text{low}} &= \begin{cases} y_a - \frac{S}{2}, & \text{if } y_a \geq -\left(\frac{2^B}{2} - 1\right) S \\ -\infty, & \text{otherwise} \end{cases} \\ y_a^{\text{up}} &= \begin{cases} y_a + \frac{S}{2}, & \text{if } y_a \leq \left(\frac{2^B}{2} - 1\right) S \\ \infty, & \text{otherwise} \end{cases} \end{aligned} \quad (4.102)$$

换言之, 对于ADC输出  $y_a = Q(\tilde{y}_a)$ , 所对应的ADC输入  $\tilde{y}_a$  的取值范围为  $[y_a^{\text{low}}, y_a^{\text{up}}]$ 。

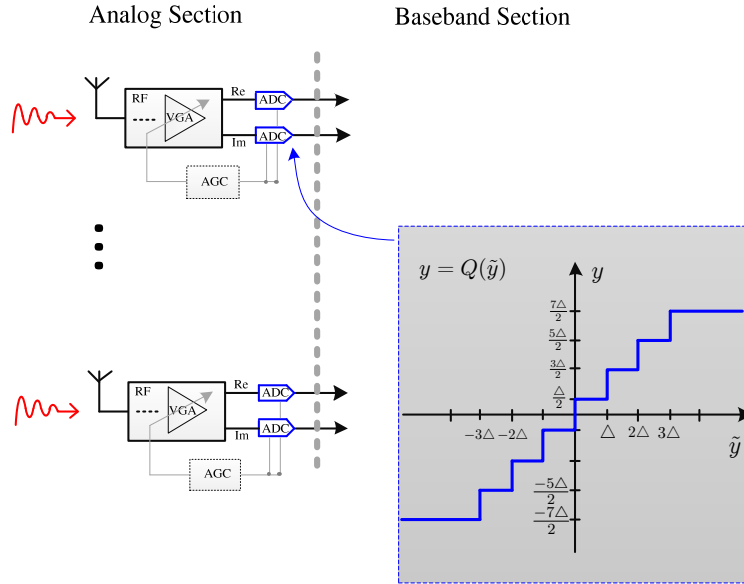


图4-6 ADC量化

图4-6给出了3-bit量化的具体过程, 包括一个自动增益放大器 (variable gain amplifier, VGA), 经过VGA之后, 将信号的实部和虚部分开处理。VGA的作用是确保信号在一个可处理的量程之内。

#### ■ $z$ 的后验均值估计

数目转化的过程是逐点进行, 因此转移概率  $p(\mathbf{y}|\mathbf{z})$  的分量式的 (componentwise), 即  $p(\mathbf{y}|\mathbf{z}) = \prod_{a=1}^M p(y_a|z_a)$ 。文献[2]给出了转移概率的具体表达式

$$p(y_a|z_a) = \Psi\left(y_{Ra}|z_{Ra}, \frac{\sigma_w^2}{2}\right) \Psi\left(y_{Ia}|z_{Ia}, \frac{\sigma_w^2}{2}\right) \quad (4.103)$$

这里为了简化符号体系，我们采用 $y_{Ra}$ 表示 $y_a$ 的实部， $y_{Ia}$ 表示 $y_a$ 的虚部。此外

$$\Psi(y|z, c^2) \triangleq \Phi\left(\frac{y^{\text{up}} - z}{c}\right) - \Phi\left(\frac{y^{\text{low}} - z}{c}\right) \quad (4.104)$$

其中， $\Phi(x) \triangleq \int_{-\infty}^x \mathcal{N}(t|0, 1)dt$ 。

给定 $p(y_a|z_a)$ 具体表达式之后，我们可以计算(4.44)、(4.72)的具体表达式。为了泛化符号体系，计算如下后验概率分布的均值和方差

$$\tilde{p}(z|y) = \frac{p(y|z)\mathcal{N}_c(z|m, v)}{\int p(y|z)\mathcal{N}_c(z|m, v)dz} \quad (4.105)$$

注意，这里计算其均值和方差应该实部和虚部分别求，然后再合并得到 $\tilde{p}(z|y)$ 的均值和方差。

经过代数运算得到（详见附录B）

$$\begin{aligned} \mathbb{E}[z|m, \frac{v}{2}] &= m - \frac{v}{\sqrt{2(\sigma_w^2 + v)}} \frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \\ \text{Var}[z|m, \frac{v}{2}] &= \frac{v}{2} - \frac{(v)^2}{2(\sigma_w^2 + v)} \left[ \frac{\eta_1\phi(\eta_1) - \eta_2\phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} + \left( \frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \right)^2 \right] \end{aligned} \quad (4.106)$$

其中

$$\begin{aligned} \eta_1 &\triangleq \frac{y^{\text{up}} - m}{\sqrt{(\sigma_w^2 + v)/2}} \\ \eta_2 &\triangleq \frac{y^{\text{low}} - m}{\sqrt{(\sigma_w^2 + v)/2}} \end{aligned} \quad (4.107)$$

这里期望对 $\frac{p(y|z)\mathcal{N}(z|m, v/2)}{\int p(y|z)\mathcal{N}(z|m, v/2)dz}$ 。

因此， $\tilde{p}(z|y)$ 的均值和方差分别表示为

$$\begin{aligned} \text{mean} &= \mathbb{E}[z_R|m_R, \frac{v}{2}] + j\mathbb{E}[z_I|m_I, \frac{v}{2}] \\ \text{variance} &= \text{Var}[z_R|m_R, \frac{v}{2}] + \text{Var}[z_I|m_I, \frac{v}{2}] \end{aligned} \quad (4.108)$$

## ■ SE

由(4.106)所示，估计器的方差 $\text{Var}[z|m, \frac{v}{2}]$ 中包含了 $y, m$ 两个参数，这两个参数均与观测向量 $y$ 有关。为了消除变量对观测量的依赖，需要计算 $\text{Var}[z|m, \frac{v}{2}]$ 对 $p(y, m)$ 的期望。由马尔科夫性，有

$$\begin{aligned} p(y, m) &= \int p(y|z, m)p(z, m)dz \\ &= \int p(y|z)p(z|m)p(m)dz \end{aligned} \quad (4.109)$$

SE的难点在于如何处理联合分布 $p(z, m)$ 。根据实验观察， $m$ 所对应的算法中变量 $m$ ，可以表示为

$$m = H\hat{x} + \epsilon \quad (4.110)$$

其中 $\hat{x}$ 为PME， $\epsilon$ 表示残差，该残差渐进趋于0。根据中心极限定理以及PME的正交性原理<sup>7</sup>，可知 $(z, m)$ 服从均值为0，协方差矩阵为 $\Omega$ 的联合高斯分布，其中

$$\Omega = \begin{pmatrix} T_z & T_z - v \\ T_z - v & T_z - v \end{pmatrix} \quad (4.111)$$

其中 $T_z = \mathbb{E}[\lambda]\mathbb{E}[|x|^2]$ ， $\lambda$ 为 $H^H H$ 的特征值。根据多元高斯分布的条件高斯<sup>8</sup>可得 $p(z|m)$ 和 $p(m)$ 的分布。为了简化符号，这里仅仅考虑实部，相应地，得到

$$p(z, m) = \mathcal{N}\left(z|m, \frac{v}{2}\right) \mathcal{N}\left(m\left|0, \frac{T_z - v}{2}\right.\right) \quad (4.112)$$

联立(4.109)、(4.112)可得

$$\begin{aligned} p(y, m) &= \int_z \Psi\left(y\left|z, \frac{\sigma^2}{2}\right.\right) \mathcal{N}\left(z\left|m, \frac{v}{2}\right.\right) \mathcal{N}\left(m\left|0, \frac{T_z - v}{2}\right.\right) dz \\ &= \int_z \int_{y^{\text{low}}}^{y^{\text{up}}} \mathcal{N}\left(x\left|z, \frac{\sigma^2}{2}\right.\right) dx \mathcal{N}\left(z\left|m, \frac{v}{2}\right.\right) \mathcal{N}\left(m\left|0, \frac{T_z - v}{2}\right.\right) dz \\ &\stackrel{(a)}{=} \Psi\left(y\left|m, \frac{v + \sigma^2}{2}\right.\right) \mathcal{N}\left(m\left|0, \frac{T_z - v}{2}\right.\right) \end{aligned} \quad (4.113)$$

其中(a)成立，基于高斯相乘引理。

为了计算 $v_z^{\text{post}} \triangleq \text{Var}[z|m, \frac{v}{2}]$ 对 $p(y, m)$ 求期望，其根本是分别计算 $\zeta$ 和 $\xi$ 对 $p(y, m)$ 的期望

$$v_z^{\text{post}} = \frac{v}{2} - \frac{v^2}{2(\sigma^2 + v)} \left[ \underbrace{\frac{\eta_1 \phi(\eta_1) - \eta_2 \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)}}_{\triangleq \zeta} + \underbrace{\left( \frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \right)^2}_{\triangleq \xi} \right] \quad (4.114)$$

分别计算得到

<sup>7</sup> 设 $\hat{x}$ 为 $x$ 的MMSE估计，则有 $\mathbb{E}[(x - \hat{x})\hat{x}^T] = 0$ ，其中分布对 $p(x, y)$ 取。

<sup>8</sup> 假设 $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ 是均值为 $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ ，协方差为 $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ 。通过去相关操作，可以得到条件分布 $p(\mathbf{x}_1|\mathbf{x}_2) \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\Sigma}}_1)$ ，其中 $\tilde{\boldsymbol{\mu}}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$ ， $\tilde{\boldsymbol{\Sigma}}_1 = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ 。

$$\begin{aligned}
 \mathbb{E}[\zeta] &= \sum_{y \in \mathcal{R}_B} \int_m \frac{\eta_1 \phi(\eta_1) - \eta_2 \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \Psi\left(y \middle| m, \frac{T_z + \sigma^2}{2}\right) \mathcal{N}\left(m \middle| 0, \frac{T_z - v}{2}\right) dm \\
 &\stackrel{(a)}{=} \sum_{y \in \mathcal{R}_B} \int_m (\eta_1 \phi(\eta_1) - \eta_2 \phi(\eta_2)) \mathcal{N}\left(m \middle| 0, \frac{T_z - v}{2}\right) dm \\
 &\stackrel{(b)}{=} 0
 \end{aligned} \tag{4.115}$$

其中(a)成立由于 $\Psi\left(y \middle| m, \frac{v + \sigma^2}{2}\right) = \Phi(\eta_1) - \Phi(\eta_2)$ , (b)成立由于 $\lim_{\eta \rightarrow \infty} \eta \phi(\eta) = 0$ 。

另一方面

$$\begin{aligned}
 \mathbb{E}[\xi] &= \sum_{y \in \mathcal{R}_B} \int_m \left( \frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \right)^2 \Psi\left(y \middle| m, \frac{v + \sigma^2}{2}\right) \mathcal{N}\left(m \middle| 0, \frac{T_z - v}{2}\right) dm \\
 &= \sum_{y \in \mathcal{R}_B} \int_m \frac{[\phi(\eta_1) - \phi(\eta_2)]^2}{\Phi(\eta_1) - \Phi(\eta_2)} \mathcal{N}\left(m \middle| 0, \frac{T_z - v}{2}\right) dm
 \end{aligned} \tag{4.116}$$

令 $u = \sqrt{\frac{2}{T_z - v}}m$ , 有 $dm = \sqrt{\frac{T_z - v}{2}}du$ , 则(4.116)可以表示为

$$\mathbb{E}[\xi] = \sum_{y \in \mathcal{R}_B} \int_u \frac{[\phi(\varsigma_1) - \phi(\varsigma_2)]^2}{\Phi(\varsigma_1) - \Phi(\varsigma_2)} Du \tag{4.117}$$

其中 $Du = \mathcal{N}(u|0, 1)du$ , 以及

$$\begin{aligned}
 \varsigma_1 &\triangleq \frac{y^{\text{up}} - \sqrt{\frac{T_z - v}{2}}u}{\sqrt{\frac{\sigma_w^2 + v}{2}}} \\
 \varsigma_2 &\triangleq \frac{y^{\text{low}} - \sqrt{\frac{T_z - v}{2}}u}{\sqrt{\frac{\sigma_w^2 + v}{2}}}
 \end{aligned} \tag{4.118}$$

基于上述推导, 则GEC-SR在ADC量化场景下的SE方程表示为

$$\begin{aligned}
\alpha &= \sum_{y \in \mathcal{R}_B} \int_u \frac{[\phi(\varsigma_1) - \phi(\varsigma_2)]^2}{\Phi(\varsigma_1) - \Phi(\varsigma_2)} Du \\
v_z^- &= v_1^+ - \frac{(v_1^+)^2}{\sigma_w^2 + v_1^+} \alpha \\
\gamma_1^- &= \frac{1}{v_z^-} - \gamma_1^+ \\
v_x^- &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\gamma_1^- \lambda_i + \gamma_0^+} \\
\gamma_0^- &= \frac{1}{v_x^-} - \gamma_0^+ \\
\gamma_0^+ &= \frac{1}{\text{MSE}(\gamma_0^-)} - \gamma_0^- \\
v_z^+ &= \frac{1}{M} \sum_{i=1}^N \frac{\lambda_i}{\gamma_1^- \lambda_i + \gamma_0^+} \\
\gamma_1^+ &= \frac{1}{v_z^+} - \gamma_1^-
\end{aligned} \tag{4.119}$$

## ■ 仿真

(1) 图4-7给出了GAMP的MSE与迭代次数的曲线图。参数设定为：用户数据为QPSK信号，发送天线 $N = 256$ ，接收天线 $M = 512$ ，系统信噪比 $\text{SNR} = 12\text{dB}$ 。如图4-7所示：

- 当bit数目为1时，GAMP算法的性能极差。当bit数目为2时，相对于bit=1性能上有很大的飞跃。
- 当bit数目为3时，GAMP的MSE曲线已经很接近GAMP在AWGN（无穷bit数）时的曲线。这也说明了，我们可以以少量的ADC量化bi数（如3bit）来逼近AWGN的性能。



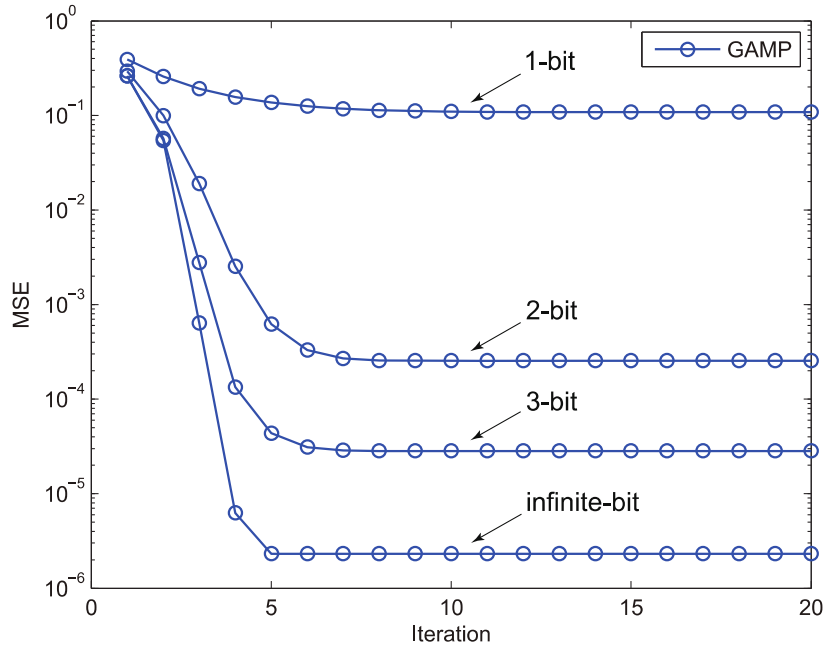


图4-7 Iter vs MSE (GAMP)

(2) 图4-8给出了GEC算法的迭代次数与MSE的曲线图。参数设置为：用户数据为QPSK信号，发射天线 $N = 512$ ，接收天线数 $M = 512$ ，信噪比 $\text{SNR} = 12\text{dB}$ 。如图4-8所示：

- 当bit=1时，GEC-SR的性能很差，基本不工作。随着bit数的增加，GEC-SR的性能逐渐改善。特别是，当bit=2时相对于bit=1，GEC-SR在性能上有了巨大的提升。
- 此外，从图中可以看出GEC-SR的MSE曲线与SE (Analytical) 很好匹配。

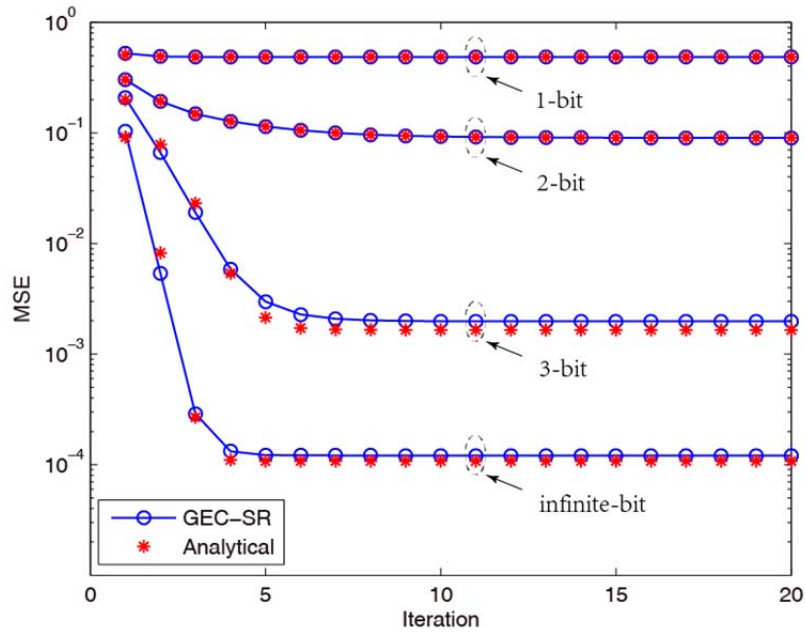


图4-8 Iteration vs MSE

(3) 图4-9比较了GEC-SR、GAMP算法与传统估计算法如LS、LMMSE的性能。参数设定为：用户数据为QPSK信号，发送天线 $N = 256$ ，接收天线 $M = 512$ ，量化bit = 3。由实验可知：

- GEC-SR、GAMP相较于传统估计算法LS、LMMSE存在着巨大的优势，当 $BER = 10^{-3}$ 时，GEC-SR需要的信噪比为 $SNR = 8.09dB$ ，GAMP为 $SNR = 8.22dB$ ，而LMMSE为 $SNR = 13.88dB$ ，LS为 $SNR = 15.01dB$ 。其中，当 $BER = 10^{-3}$ 时，GAMP所需的SNR相对于LMMSE小了5.66dB，而GEC-SR相对于LS则小了6.92dB。
- 在该参数设置下，GEC-SR相对于GAMP存在着微弱的优势，并且这种优势随着SNR的增加，逐渐增大。如图4-9黑线与绿线。
- LMMSE相较于LS有优势，且这种优势，并不随着信噪比的增加而减小。原理上，若对于精确的标准线性模型，随着SNR的增加，LS的性能曲线将逐渐与LMMSE重合，见图。而图4-9中，LMMSE与LS曲线始终保持着一定的间隔，并且两者之差有增大的趋势，可以认为是构造标准线性模型所带来的误差引起。

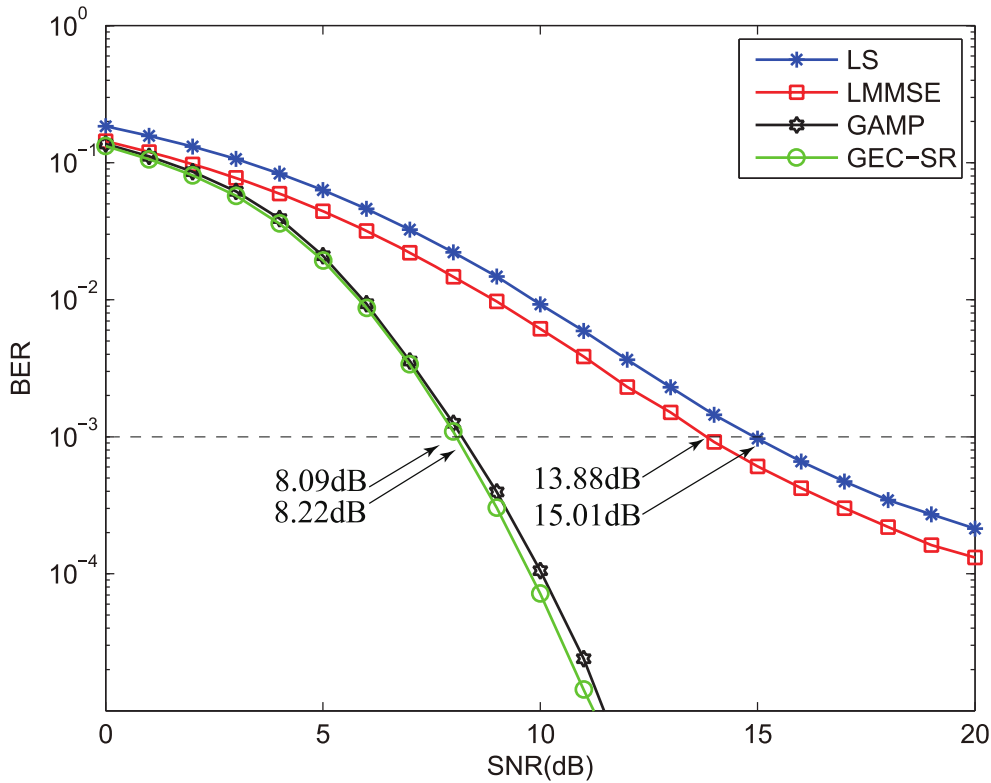


图4-9 GEC-SR、GAMP与传统估计方法性能比较

(4) 图4-10测试了系统维度对GEC-SR与GAMP算法性能的影响。参数设定：用户数据为QPSK信号,  $\text{bit}=3$ , 绘制了三组曲线, 分别是天线数  $32 \times 32$ ,  $128 \times 128$ , 以及  $512 \times 512$ 。从图中可以看出:

- 随着天线数的增加, GEC-SR和GAMP算法的性能都逐渐提高。
- 可以明显看到, 当天线配置为  $32 \times 32$  与天线配置为  $128 \times 128$  的曲线存在着明显的差距。我们可以从基于中心极限定理的高斯近似角度去理解, 当变量个数很小时, 这种近似是不合理的, 因而算法性能很差。
- 此外, 天线配置为  $512 \times 512$  的曲线与天线配置为  $128 \times 128$  的曲线, 差距不是很明显。这也说明了, 当系统的维度达到一定的规模, 在其他参数不变的情况下, 增大系统维度, 在性能上并不会会有太大改善。

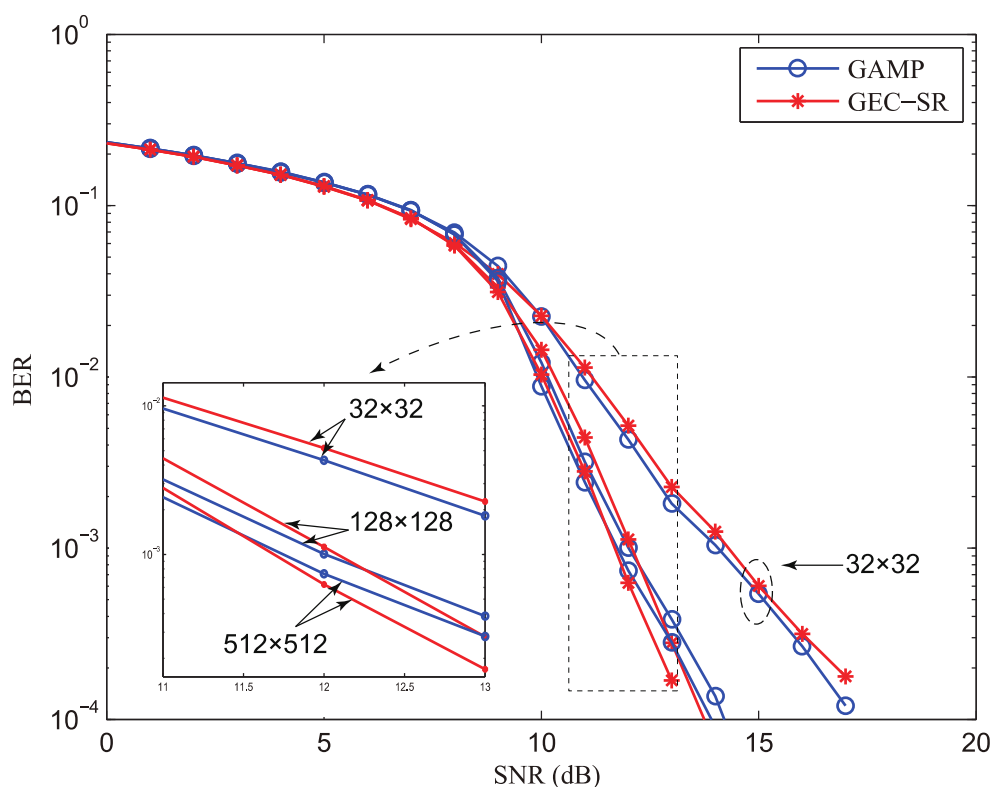


图4-10 系统维度对算法影响

## 结论与展望

本文研究了两类比较前沿的贝叶斯估计算法，用于求解逆问题。更确切地说，这种问题是从 $M$ 维观测向量 $\mathbf{y}$ 中恢复出 $N$ 维目标向量 $\mathbf{x}$ 。作为贝叶斯算法中的杰出算法，AMP算法是对Sum-Product算法的完美近似。在大规模系统中，基于中心极限定理，使得这种近似所存在的误差是可忽略的。本文给出了两种AMP推导方法，分别是基于分列式因子图的推导方法和基于EP的AMP推导方法。实验部分给出了AMP算法的收敛性实验和算法性能实验。

GAMP算法放松了AMP对噪声分布的限制。本文给出了两种GAMP推导方法，分别是基于传统因子图的GAMP推导和基于分列式因子图的GAMP推导。两种推导方法相对于传统基于Max-Sum算法的GAMP推导，极大化简了推导步骤，并且结论适用于复数信号。

另外方面，EP作为贝叶斯算法中另外一类杰出算法，主要用于近似可因子化的概率分布。本文给出了两种EP在SLM场景下的推导方法，分别是基于矢量因子图的VAMP推导以及基于变分法的EP推导。

GEC和GEC-SR算法扩展了EP的使用范围，其中GEC算法存在着收敛性问题。GEC-SR通过引入了一个线性空间 $\delta(\mathbf{z} - \mathbf{H}\mathbf{x})$ 模块，很好得解决了这个问题。本文补充了GEC-SR的推导。

总而言之，EP-like算法与AMP-like算法在性能、收敛速度、稳定性上各有千秋。并且，除SLM、GLM、ML-GLM外，在其他模型如双线性模型、广义双线性模型、多层广义双线性模型，有着相应地扩展。

### 未来工作

- 将GEC-SR扩展到广义双线性模型。广义双线性模型相对于GLM，观测矩阵 $\mathbf{H}$ 是未知的，类似于盲信号估计。
- 推导ML-GAMP算法并利用Replica Method分析ML-GLM的不动点方程。
- 完善ML-GEC的推导，完成ML-GEC的仿真。

## 参考文献

- [1] Zou Q, Zhang H, Wen C K, et al. Concise derivation for generalized approximate message passing using expectation propagation[J]. IEEE Signal Processing Letters, 2018, 25(12): 1835-1839.
- [2] Wen C K, Wang C J, Jin S, et al. Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs[J]. IEEE Transactions on Signal Processing, 2016, 64(10): 2541-2556.
- [3] Wen C K, Jin S, Wong K K, et al. Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning[J]. IEEE Transactions on Wireless Communications, 2015, 14(3): 1356-1368.
- [4] Meng X, Wu S, Kuang L, et al. Multi-user detection for spatial modulation via structured approximate message passing[J]. IEEE Communications Letters, 2016, 20(8): 1527-1530.
- [5] Wen C K, Zhang J, Wong K K, et al. On sparse vector recovery performance in structurally orthogonal matrices via LASSO[J]. IEEE Transactions on Signal Processing, 2016, 64(17): 4519-4533.
- [6] Winn J, Bishop C M. Variational message passing[J]. Journal of Machine Learning Research, 2005, 6(Apr): 661-694.
- [7] Tipping M E. Sparse Bayesian learning and the relevance vector machine[J]. Journal of machine learning research, 2001, 1(Jun): 211-244.
- [8] Prasad R, Murthy C R. Cramér-Rao-type bounds for sparse Bayesian learning[J]. IEEE Transactions on Signal Processing, 2013, 61(3): 622-632.
- [9] Minka T P. A family of algorithms for approximate Bayesian inference[D]. Massachusetts Institute of Technology, 2001.
- [10] Bishop C M. Pattern recognition and machine learning[M]. springer, 2006.
- [11] Wipf D P, Rao B D. Sparse Bayesian learning for basis selection[J]. IEEE Transactions on Signal processing, 2004, 52(8): 2153-2164.
- [12] Pfister H D. Compressed Sensing Supplemental Material for Graphical Models and Inference[J]. 2014.
- [13] Parker J T, Cevher V, Schniter P. Compressive sensing under matrix uncertainties: An approximate message passing approach[C]. 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR). IEEE, 2011:

804-808.

- [14] Andersen M R. Sparse inference using approximate message passing[J]. Technical University of Denmark, Department of Applied Mathematics and Computing, 2014.
- [15] Rangan S. Generalized approximate message passing for estimation with random linear mixing[C]. 2011 IEEE International Symposium on Information Theory Proceedings. IEEE, 2011: 2168-2172.
- [16] Fletcher A K, Rangan S, Schniter P. Inference in deep networks in high dimensions[C]. 2018 IEEE International Symposium on Information Theory (ISIT). IEEE, 2018: 1884-1888.
- [17] Manoel A, Krzakala F, Mézard M, et al. Multi-layer generalized linear estimation[C]. 2017 IEEE International Symposium on Information Theory (ISIT). IEEE, 2017: 2098-2102.
- [18] Donoho D L, Maleki A, Montanari A. How to design message passing algorithms for compressed sensing[J]. preprint, 2011.
- [19] Davenport M A, Duarte M F, Eldar Y C, et al. Introduction to compressed sensing[J]. Compressed sensing: theory and applications, 2012, 105: 106.
- [20] Richardson T, Urbanke R. Modern coding theory[M]. Cambridge university press, 2008.
- [21] Guo D, Wang C C. Asymptotic mean-square optimality of belief propagation for sparse linear systems[C]. 2006 IEEE Information Theory Workshop-ITW'06 Chengdu. IEEE, 2006: 194-198.
- [22] Donoho D L, Maleki A, Montanari A. Message-passing algorithms for compressed sensing[J]. Proceedings of the National Academy of Sciences, 2009, 106(45): 18914-18919.
- [23] Bayati M, Montanari A. The dynamics of message passing on dense graphs, with applications to compressed sensing[J]. IEEE Transactions on Information Theory, 2011, 57(2): 764-785.
- [24] Ma J, Ping L. Orthogonal amp[J]. IEEE Access, 2017, 5: 2020-2033.
- [25] Metzler C A, Maleki A, Baraniuk R G. From denoising to compressed sensing[J]. IEEE Transactions on Information Theory, 2016, 62(9): 5117-5144.
- [26] Cakmak B, Winther O, Fleury B H. S-AMP: Approximate message passing for general matrix ensembles[C]. 2014 IEEE Information Theory Workshop (ITW 2014). IEEE, 2014: 192-196.
- [27] Meng X, Wu S, Zhu J. A unified Bayesian inference framework for generalized linear models[J]. IEEE Signal Processing Letters, 2018, 25(3): 398-402.

- 
- [28]Rangan S, Schniter P, Fletcher A K. Vector approximate message passing[C]. 2017 IEEE International Symposium on Information Theory (ISIT). IEEE, 2017: 1588-1592.
  - [29]Schniter P, Rangan S, Fletcher A K. Vector approximate message passing for the generalized linear model[C]. 2016 50th Asilomar Conference on Signals, Systems and Computers. IEEE, 2016: 1525-1529.
  - [30]He H, Wen C K, Jin S. Generalized expectation consistent signal recovery for nonlinear measurements[C]. 2017 IEEE International Symposium on Information Theory (ISIT). IEEE, 2017: 2333-2337.
  - [31]Guo D, Verdú S. Randomly spread CDMA: Asymptotics via statistical physics[J]. arXiv preprint cs/0503063, 2005.
  - [32]Boyd S, Vandenberghe L. Convex optimization[M]. Cambridge university press, 2004.
  - [33]Kay S M. Fundamentals of statistical signal processing[M]. Prentice Hall PTR, 1993.
  - [34]Al-Shoukairi M, Rao B. Sparse Bayesian learning using approximate message passing[C]. 2014 48th Asilomar Conference on Signals, Systems and Computers. IEEE, 2014: 1957-1961.
  - [35]Opper M, Winther O. Expectation consistent approximate inference[J]. Journal of Machine Learning Research, 2005, 6(Dec): 2177-2204.
  - [36]Rangan S, Schniter P, Fletcher A K. Vector Approximate Message Passing[J]. 2016.
  - [37]Guo D, Wu Y, Shitz S S, et al. Estimation in Gaussian noise: Properties of the minimum mean-square error[J]. IEEE Transactions on Information Theory, 2011, 57(4): 2371-2385.
  - [38]He H, Wen C K, Jin S. Bayesian optimal data detector for hybrid mmwave MIMO-OFDM systems with low-resolution ADCs[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(3): 469-483.
  - [39]Oxvig, C. S., Arildsen, T., & Larsen, T. (2017). Generalized Approximate Message Passing: Relations and Derivations. DOI: 10.5278/VBN.GAMPTechReport
  - [40]Vila J, Schniter P. Expectation-maximization Bernoulli-Gaussian approximate message passing[C]. 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR). IEEE, 2011: 799-803.
  - [41]Zhu J, Han L, Meng X. An AMP-Based Low Complexity Generalized Sparse Bayesian Learning Algorithm[J]. IEEE Access, 2019, 7: 7965-7976.
  - [42]Fletcher A, Sahraee-Ardakan M, Rangan S, et al. Expectation consistent approximate inference: Generalizations and convergence[C]. 2016 IEEE International Symposium on Information Theory (ISIT). IEEE, 2016: 190-194.

- [43] Meng X, Wu S, Kuang L, et al. An expectation propagation perspective on approximate message passing[J]. IEEE Signal Processing Letters, 2015, 22(8): 1194-1197.



## 附录

### A (3.53)

$$\phi(m, v) = \frac{\int x^2 p(x) \mathcal{N}(x|m, v) dx}{\int p(x) \mathcal{N}(x|m, v) dx} - \left( \frac{\int x p(x) \mathcal{N}(x|m, v) dx}{\int p(x) \mathcal{N}(x|m, v) dx} \right)^2 \quad (\text{A.1})$$

为简化符号, 定义  $f(x) = p(x) \mathcal{N}(x|m, v)$ , 则有

$$\begin{aligned} \frac{\partial \phi(m, v)}{\partial m} &= \frac{\int \frac{x^2(x-m)}{v} f(x) dx \cdot \int f(x) dx - \int x^2 f(x) dx \cdot \int \frac{x-m}{v} f(x) dx}{\left( \int f(x) dx \right)^2} \\ &\quad - 2 \left( \frac{\int \frac{x(x-m)}{v} f(x) dx \cdot \int f(x) dx - \int x f(x) dx \cdot \int \frac{x-m}{v} f(x) dx}{\left( \int f(x) dx \right)^2} \right) \quad (\text{A.2}) \\ &= \frac{1}{v} (\mathbb{E}[X^3] - \mathbb{E}[X^2] \mathbb{E}[X]) - \frac{2}{v} (\mathbb{E}[X^2] - (\mathbb{E}[X])^2) \end{aligned}$$

其中期望对概率分布  $\frac{f(x)}{\int f(x) dx}$  取。当  $v > 0$  时, 显然  $\frac{\partial \phi(m, v)}{\partial m}$  是有界的。当  $v \rightarrow 0$  时, 有

$$\begin{aligned} \lim_{v \rightarrow 0} \frac{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}{v} &= \lim_{v \rightarrow 0} \frac{\frac{\int x^2 p(x) \exp\left(-\frac{(x-m)^2}{2v}\right) dx}{\int p(x) \exp\left(-\frac{(x-m)^2}{2v}\right) dx} - \left( \frac{\int x p(x) \exp\left(-\frac{(x-m)^2}{2v}\right) dx}{\int p(x) \exp\left(-\frac{(x-m)^2}{2v}\right) dx} \right)^2}{v} \quad (\text{A.3}) \\ &\stackrel{(a)}{=} \lim_{v \rightarrow 0} \frac{\frac{m^2 p(m)}{p(m)} - \left( \frac{m p(m)}{p(m)} \right)^2}{v} \\ &\stackrel{(b)}{=} 0 \end{aligned}$$

其中(a)、(b)成立, 基于

- (1)  $\delta(x) = \lim_{b \rightarrow \infty} b e^{-\pi(bx)^2}$ ;
- (2)  $\delta(x)$  函数的性质  $\int h(x) \delta(x-m) dx = h(m)$ ;
- (3)  $\exp(-\frac{1}{v})$  比  $v$  更快趋于 0。

同理, 可以得到  $\lim_{v \rightarrow \infty} \frac{1}{v} (\mathbb{E}[X^3] - \mathbb{E}[X^2] \mathbb{E}[X]) = 0$ , 因此  $\frac{\partial \phi(m, v)}{\partial m}$  是有界的。

### B (4.106)

定义  $Z$  的后验均值及其方差为

$$\begin{aligned} z^{\text{post}} &= \int z \frac{p(y|z) \mathcal{N}_c(z|m, v)}{\int p(y|z) \mathcal{N}_c(z|m, v) dz} dz \\ v_z^{\text{post}} &= \int |z^{\text{post}} - z|^2 \frac{p(y|z) \mathcal{N}_c(z|m, v)}{\int p(y|z) \mathcal{N}_c(z|m, v) dz} dz \quad (\text{B.1}) \end{aligned}$$

由于 $Z$ 的实部和虚部相互独立, 因此仅考虑实部情况, 为了简化符号, 我们采用 $(y, z, m)$ 来表示参数 $(y_R, z_R, m_R)$ 。基于该表述, 转移概率 $p(y|z)$ 为 $p(y|z) = \Psi(y|z, \frac{\sigma^2}{2})$ 。

计算

$$\begin{aligned} \int_z p(y|z) \mathcal{N}(z|m, \frac{\sigma^2}{2}) dz &\stackrel{(a)}{=} \int_z \int_{y^{\text{low}}}^{y^{\text{up}}} \mathcal{N}(x|z, \frac{\sigma^2}{2}) dx \mathcal{N}(z|m, \frac{\sigma^2}{2}) dz \\ &\stackrel{(b)}{=} \int_{y^{\text{low}}}^{y^{\text{up}}} \mathcal{N}(x|m, \frac{\sigma^2}{2}) dx \\ &= \Phi(\eta_1) - \Phi(\eta_2) \triangleq C \end{aligned} \quad (\text{B.2})$$

其中(a)成立, 基于积分变量替换<sup>9</sup>。(b)成立, 基于高斯相乘引理。此外, 包含定义变量如下

$$\begin{aligned} \eta_1 &= \frac{y^{\text{up}} - m}{\sqrt{\frac{\sigma^2 + v}{2}}} \\ \eta_2 &= \frac{y^{\text{low}} - m}{\sqrt{\frac{\sigma^2 + v}{2}}} \end{aligned} \quad (\text{B.3})$$

对 $\int p(y|z) \mathcal{N}(z|m, \frac{\sigma^2}{2}) dz = C$ 等式两边, 分别对 $m$ 求偏导, 得

$$\int \left( \frac{z - m}{v/2} \right) p(y|z) \mathcal{N}(z|m, \frac{v^2}{2}) dz = -\sqrt{\frac{2}{\sigma^2 + v}} (\phi(\eta_1) - \phi(\eta_2)) \quad (\text{B.4})$$

上式成立, 利用了 $\frac{\partial \Phi(x)}{\partial m} = \phi(x) \frac{\partial x}{\partial m}$ 以及 $\phi(x) = \mathcal{N}(x|0, 1)$ 。

结合以上结论, 可得

$$z^{\text{post}} = m - \frac{v}{\sqrt{2(\sigma^2 + v)}} \frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \quad (\text{B.5})$$

利用指数族的性质<sup>10</sup>, 进一步有

$$v_z^{\text{post}} = \frac{v}{2} - \frac{v^2}{2(\sigma^2 + v)} \left[ \frac{\eta_1 \phi(\eta_1) - \eta_2 \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} - \left( \frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \right)^2 \right] \quad (\text{B.6})$$

<sup>9</sup> 对于 $\Phi(\frac{y^{\text{up}} - z}{\sigma}) = \int_{-\infty}^{y^{\text{up}}} \mathcal{N}(t|0, 1) dt$ , 定义辅助变量 $x = \sigma t + z$ , 则有 $dx = \sigma dt$

$$\int_{-\infty}^{\frac{y^{\text{up}} - z}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \int_{-\infty}^{y^{\text{up}}} \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-z)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{y^{\text{up}}} \mathcal{N}(x|z, \sigma^2) dx.$$

<sup>10</sup>  $\frac{\partial}{\partial m} \frac{\int p(x) \mathcal{N}(x|m, v) dx}{\int p(x) \mathcal{N}(x|m, v) dx} = \frac{1}{v} \frac{\int (x - \hat{x})^2 p(x) \mathcal{N}(x|m, v) dx}{\int p(x) \mathcal{N}(x|m, v) dx}$ , 其中 $p(x)$ 为任意非负有界函数。

## 攻读学位期间发表的学术论文

### ■ 已发表论文

-**Zou Q**, Zhang H, Wen C K, et al. Concise derivation for generalized approximate message passing using expectation propagation[J]. IEEE Signal Processing Letters, 2018, 25(12): 1835-1839. (第一作者, SCI收录)

### ■ 在审论文

-Zhang H, **Zou Q**, Zhao X, Wen C K, et al. Moment matched message passing part I: theory[J]. IEEE Transactions on Signal Processing, 2019. (通讯作者, 在审)

-Zhang H, **Zou Q**, Zhao X, Wen C K, et al. Moment matched message passing part II: applications[J]. IEEE Transactions on Signal Processing, 2019. (通讯作者, 在审)

## 学位论文独创性声明

本人郑重声明：所呈交的学位论文是我个人在导师的指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明，并表示了谢意。本人依法享有和承担由此论文所产生的权利和责任。

论文作者签名：郭秋云 日期：2019年5月29日

## 学位论文版权使用授权声明

本学位论文作者完全了解学校有关保存、使用学位论文的规定，同意授权广东工业大学保留并向国家有关部门或机构送交该论文的印刷本和电子版本，允许该论文被查阅和借阅。同意广东工业大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可授以采用影印、缩印、扫描或数字化等其他复制手段保存和汇编本学位论文。保密论文在解密后遵守此规定。

论文作者签名：郭秋云 日期：2019年5月29日

指导教师签名：张浩川 日期：2019年5月29日

## 致谢

时光匆匆，恍然之间已经是研究生生涯的第三个年头。三年，说长不长，说短也不短。回首这三年，有过热泪盈眶，也有过不知所措，有过歧路徘徊，也有过有的放矢。不悔经行处，只恨太匆匆！

这一路走来，需要感谢的人太多太多了。首先，非常感谢我的导师张浩川老师。刚踏入研究生阶段，我就开始跟张老师一起做科研。张老师扎实的学术功底和平易近人的处事风格让我印象深刻。一开始，我的学术底子很薄，在科研过程中经常遇到一些难以解决的问题，这时候，张老师总是很热心地帮我解答。每当我遇到挫折而犹豫徘徊时，是您的鼓励和支持，让我克服重重困难。您不仅教会我科研的本领，更教会我很多做人的道理。以前，我对科研比较懒散，而且也很抵触别人对我所认知的反驳，是您不厌其烦的提醒，让我慢慢地改掉了这些坏毛病。可以说，我能在学术道路上不断前进探索，完全离不开您的帮助和鼓励。张老师给了我很多机会，我也很珍惜，希望在博士阶段能够取得更多的成果，以不辜负老师之期许。此时此刻，我想对张老师说：谢谢您！谢谢您对我的信任和教导！

此外，非常感谢广州大学的范立生老师。我与范老师是在操场上正式相识。范老师为人洒脱不羁，有时候，我和范老师之间，更像是朋友一般。在准备考研时，每当复习进度不佳，心情低落的时候，和范老师跑跑步聊聊天，总会感觉到满满的动力。谢谢您的鼓励和教诲，陪伴我走过那段艰辛而充实的岁月。在研究生阶段，每当科研遇到挫折时，您总会以您高屋建瓴的认知，鼓励、帮助我克服重重困难。范老师，谢谢您对我的认可，谢谢！

我还要感谢国立中山大学的温朝凯老师。温老师博学渊识，治学严谨，每次在专业上遇到问题向温老师请教，温老师都是很耐心、逐条细致回复邮件。非常感谢温老师在专业上对我的指导。非常感谢北京邮电大学的杨鸿文老师给了我这么一个宝贵的攻读博士的机会，我会好好珍惜这个来之不易的机会，争取在博士阶段做出更多高质量的成果。另外，感谢大团队的余荣老师，对我工作的支持。

感谢中南大学的庄翠芳博士，山东大学的李志鹏博士，中山大学的赖夏智博士，西南交通大学的蔡东洪博士，南京邮电大学的朱敏博士给予我的帮助。

感谢江丰、悦城、灼鸿、梓荣、书山、武斌、佳宁几位老友，在闲暇之余跟我聊

天谈心。此外，我还要感谢王颖同学对我的认可和鼓励。在生命中，能遇到你们，是我莫大的荣幸。

最后，我要感谢下我的父母。虽然与你们很少交流自己的想法，但是我知道你们一直在默默地支持我。有太多太多东西，不需要表达，但一直都在。

一万年太久，只争朝夕！