# 稀疏感知导论

Introduction to Sparse Sensing

李廉林　李　芳　著

非外借

科学出版社

# 稀疏感知导论

# Introduction to Sparse Sensing

李廉林　李　芳　著

学院图书馆
藏书专用章

科学出版社

## 内 容 简 介

在大数据时代，信息科学必须发展信息表征、获取及复原的新理论、新方法。基于信号的稀疏性，“稀疏感知”用少量的测量数据实现高质量的信号复原，缓解大数据信息问题的压力。

实现稀疏感知的关键是：(1)有效地获取数据，(2)有效地处理数据，实现信息复原。对于数据获取，本书研究了信号稀疏性与信号采样之间的关系，讨论了压缩感知、矩阵填充、稀疏反卷积和相位复原等稀疏感知问题，从三个不同的角度讨论了测量矩阵的设计。对于数据处理和信息复原，本书研究了梯度迭代优化算法、Bayesian 算法和信息传递算法；特别地针对大数据处理问题，本书研究了乘子交替迭代优化算法、随机坐标优化算法和随机梯度优化算法等；本书还讨论了若干贪婪算法。

本书从稀疏感知的基本概念、基本问题出发，重点关注那些与实际应用问题密切相关的理论要点和算法核心技术，为相关领域的读者和研究人员提供必要的参考和指导。

**图书在版编目(CIP)数据**

稀疏感知导论/李廉林，李芳著. —北京：科学出版社，2018.1

ISBN 978-7-03-053026-4

Ⅰ. ①稀… Ⅱ. ①李… ②李… Ⅲ. ①信号处理 Ⅳ. ①TN911.7

中国版本图书馆 CIP 数据核字 (2017) 第 119106 号

责任编辑：钱 俊 胡庆家／责任校对：杨 然

责任印制：张 伟／封面设计：无极书装

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

北京建宏印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2018 年 1 月第 一 版 开本：720×1000 1/16

2018 年 1 月第一次印刷 印张：19 3/4

字数：393 000

**定价：128.00 元**

(如有印装质量问题，我社负责调换)

# 前　言

我们有幸生活在全球科技发生许多重大突破的时代，信息大爆炸、大数据一定程度上反映了当今时代的特点。21 世纪初 “稀疏感知”（Sparse Sensing）在信息科学掀起一场 “革命”。“稀疏感知” 利用信号的稀疏特性，用低于甚至远低于 Nyquist 采样率获取的信号离散样本，通过高性能的优化算法实现信号的完美重建。“稀疏感知” 使信息的获取摆脱了传统信息论的 “桎梏”，在信息论、图像处理、模式识别、微波成像、地球科学、天文观测、光学、无线通信等诸多领域受到极大关注，被美国的权威科技评论杂志评为 2007 年度十大科技进展。时代要求有创新力的科学家和工程技术人员必须掌握先进的数据处理的理论和方法，有能力从海量数据中高效率地提取所需要的目标信息。我们在应用稀疏感知方法研究电磁成像中深刻认识到，稀疏感知为复杂、多参数、电磁大数据问题的研究打开了新的窗口，并决心构建稀疏电磁感知的体系，为解决现有的电磁成像体制与信息处理中的困难开辟新途径。

稀疏感知是传统经典信息理论与方法的发展，但又有它自己全新的概念和方法。我们整理了十多年来在学习、工作、研究中积累的有关内容，参考国外的有关教材、专著和近年一些新的研究成果，形成本书。这里重点关注那些与实际应用密切相关的理论要点和算法核心技术，希望能有益于从事或将从事以及有兴趣于此领域的读者，帮助他们能以比较容易理解的方式进入，并消化、吸收、应用这个新的信息感知的理论和技术。

作者十分感谢得到上海交通大学 “李政道图书馆” 的惠允，提供抽象画《流光》原稿的扫描件作为本书的封面图。该画由诺贝尔物理奖获得者、著名物理学家李政道先生与中央工艺美术学院吴冠中教授合作，为 1996 年中国高等科学技术中心举办的 “复杂性与简单性” 国际学术研讨会而作。这幅具现代风格的抽象画，吴冠中美术大师 “以点、线挥洒神韵，千变万化，化静为动，犹如乾旋坤转”*, 在大师眼中 “复杂的天地万物形态各异，运动不息，又互有联系，有其各自和共同的规律”，“复杂行为背后存在着简单的规律和描述方法”*，其科学内涵与本书理义相通。

---

* 李政道，柳怀祖. 科学与艺术. 上海：上海科学技术出版社, 2000.

本书基于我们多年的学习笔记和研究工作，凝聚了我们的心血；但自知管见所及有限，不足之处难免，诚望读者朋友及各位专家、学者不吝赐教。

最后，在书稿将成之际，铭记在成书过程中受到多位同事、亲人朋友的帮助和北京市教育局“优秀博士导师”项目资助，在此一并致谢。

作　者

2017 年 8 月于北京

# 符号使用和约定

1. 常用符号

$x$, $\boldsymbol{x}$, $\boldsymbol{A}$ 分别表示标量、向量和矩阵 (对确定性量)
$X, x$ 分别表示随机变量和它的某次随机实现 (对随机量)
$\boldsymbol{I}_M$ 表示大小为 $M \times M$ 的单位矩阵
$\mathbf{1}$ 表示元素均为 1 的向量或矩阵 (视具体情况而定)
$\mathbf{0}$ 表示元素均为 0 的向量或矩阵 (视具体情况而定)
$\boldsymbol{x}_\Omega$ 表示支撑于集合 $\Omega$ 上 $\boldsymbol{x}$ 的子向量
$\boldsymbol{A}(:, i)$ 表示矩阵 $\boldsymbol{A}$ 的第 $i$ 列
$\boldsymbol{A}(i, :)$ 表示矩阵 $\boldsymbol{A}$ 的第 $i$ 行
$A_{i,j}$ 表示矩阵 $\boldsymbol{A}$ 的第 $(i, j)$ 个元素
$\boldsymbol{A}'$ 表示矩阵 $\boldsymbol{A}$（复矩阵）的共轭转置，或矩阵 $\boldsymbol{A}$（实矩阵）的转置
$\boldsymbol{x}'$ 表示向量 $\boldsymbol{x}$（复向量）的共轭转置，或向量 $\boldsymbol{x}$（实向量）的转置
$\boldsymbol{A}^{\mathrm{T}}$ 表示矩阵 $A$ 的转置
$\boldsymbol{A}^{\mathrm{H}}$ 表示矩阵 $A$ 的共轭转置
$\mathbb{R}^N$ 表示 $N$ 维实值列向量
$\mathbb{R}^{M,N}$ 表示 $M \times N$ 大小的实值矩阵
$\mathbb{C}^N$ 表示 $N$ 维实值列向量
$\mathbb{C}^{M,N}$ 表示 $M \times N$ 大小的复矩阵
$\mathrm{Trace}(\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}$ 的迹
$\mathrm{Rank}(\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}$ 的秩
$E_p[\,]$ 表示在概念 $p$ 意义下 (或者相对随机变量 $p$) 的数学期望
$\mathrm{Var}_p[\,]$ 表示在概率 $p$ 意义下 (或者相对随机变量 $p$) 的方差
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}'\boldsymbol{y}$
$\boldsymbol{x} \circ \boldsymbol{y}$ 表示 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的 Hadamard 乘积
$\boldsymbol{A} \circ \boldsymbol{B}$ 表示 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的 Hadamard 乘积
$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{Trace}\left(\boldsymbol{A}'\boldsymbol{B}\right)$
$|\boldsymbol{A}|$ 表示矩阵 $\boldsymbol{A}$ 的行列式
$|T|$ 表示集合 $T$(非空集合) 中元素的个数
$o(a)$ 和 $O(a)$ 分别表示 $\boldsymbol{a}$ 的高阶无穷小量和同阶无穷小量

2. 向量的加权范数

$$\|\boldsymbol{x}\|_A = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$$

3. 矩阵的范数

算子范数: $\|\boldsymbol{A}\|_p = \sup_{\|\boldsymbol{x}\|_p} \|\boldsymbol{A}\boldsymbol{x}\|_p$

Schatten $p$-范数: $\|\boldsymbol{A}\|_p = \text{Trace}\left[\left(\boldsymbol{A}'\boldsymbol{A}\right)^{p/2}\right]$

混合 $(p,q)$-范数: $\|\boldsymbol{A}\|_{p,q} = \text{Trace}\left[\left(A'A\right)^{p/2}\right]$

4. 矩阵的常用特性

$\boldsymbol{A}^{-1} = \dfrac{1}{|\boldsymbol{A}|}\boldsymbol{A}^*$, 其中 $\boldsymbol{A}^*$ 表示 $\boldsymbol{A}$ 的伴随矩阵。

$$\frac{\mathrm{d}\,|\boldsymbol{A}|}{\mathrm{d}\boldsymbol{A}} = |\boldsymbol{A}|\,\boldsymbol{A}^{-1}$$

$$|\boldsymbol{A}|^{-1} = \left|\boldsymbol{A}^{-1}\right|$$

$$\text{Trace}\left(\boldsymbol{A}\boldsymbol{B}\right) = \text{Trace}\left(\boldsymbol{B}\boldsymbol{A}\right)$$

$$\text{Trace}\left(\boldsymbol{A}'\right) = \text{Trace}\left(\boldsymbol{A}\right)$$

$$\left(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D}+\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}$$

$$\left(\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}'\right)^{-1} = \boldsymbol{A}^{-1} - \frac{\left(\boldsymbol{A}^{-1}\boldsymbol{u}\right)\left(\boldsymbol{v}'\boldsymbol{A}^{-1}\right)}{\boldsymbol{v}'\boldsymbol{A}^{-1}\boldsymbol{u}}$$

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1}+\boldsymbol{F}\boldsymbol{E}^{-1}\boldsymbol{F}' & -\boldsymbol{F}\boldsymbol{E}^{-1} \\ -\boldsymbol{E}\boldsymbol{F}' & \boldsymbol{E}^{-1} \end{bmatrix}$$

其中, $\boldsymbol{E} = \boldsymbol{D} - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}, \boldsymbol{F} = \boldsymbol{A}^{-1}\boldsymbol{B}$。

$$\left|\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & \boldsymbol{D} \end{bmatrix}\right| = |\boldsymbol{A}|\left|\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right| = |\boldsymbol{D}|\left|\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right|$$

5. 距离

函数 $\text{dist}(\cdot,\cdot)$ 表示一个距离度量 (Distance Measure)，满足基本性质：

(1) 非负性：$\text{dist}(\boldsymbol{x}_i,\boldsymbol{x}_j) \geqslant 0$, 当且仅当 $\boldsymbol{x}_i = \boldsymbol{x}_j$ 取等号；

(2) 对称性：$\text{dist}(\boldsymbol{x}_i,\boldsymbol{x}_j) = \text{dist}(\boldsymbol{x}_j,\boldsymbol{x}_i)$；

(3) 三角不等性：$\text{dist}(\boldsymbol{x}_i,\boldsymbol{x}_j) \leqslant \text{dist}(\boldsymbol{x}_j,\boldsymbol{x}_k) + \text{dist}(\boldsymbol{x}_i,\boldsymbol{x}_k)$。

**例**　给定两个向量 $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathbb{R}^N$ 的闵可夫斯基距离 (Minkowski Distance) 是

$$\text{dist}_{mk}(\boldsymbol{x}_i,\boldsymbol{x}_j) = \left(\sum_{n=1}^{N}\left|x_i[n]-x_j[n]\right|^p\right)^{1/p}$$

其中，$p \geqslant 1$。当 $p=1$ 时，闵可夫斯基距离为曼哈顿距离 (Manhattan Distance)；

当 $p=2$ 时，闵可夫斯基距离为欧氏距离 (Euclidean Distance)；当 $p=\infty$ 时，闵可夫斯基距离为切比雪夫距离 (Chebyshev Distance)。

6. 高斯分布及其特性

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{N/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],\quad \boldsymbol{x}\in\mathbb{R}^N$$

高斯分布具有如下性质：

(1) 设 $\boldsymbol{x}_1\in\mathbb{R}^{N_1}$ 和 $\boldsymbol{x}_2\in\mathbb{R}^{N_2}$ 是两个高斯随机向量：

$$\boldsymbol{x}_1\sim\mathcal{N}(\boldsymbol{x}_1|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11}),\quad \boldsymbol{x}_2\sim\mathcal{N}(\boldsymbol{x}_2|\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22})$$

它们的联合概率密度是

$$\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}\sim\mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}\middle|\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

那么条件概率仍是高斯概率：

$$p(\boldsymbol{x}_1|\boldsymbol{x}_2)=\mathcal{N}\left(\boldsymbol{x}_1|\boldsymbol{\mu}_{1|2},\boldsymbol{\Sigma}_{1|2}\right)$$

式中

$$\begin{aligned}\boldsymbol{\mu}_{1|2}=&\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\\ =&\boldsymbol{\mu}_1-\boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\\ =&\boldsymbol{\Sigma}_{1|2}(\boldsymbol{\Lambda}_{11}\boldsymbol{\mu}_1-\boldsymbol{\Lambda}_{12}(\boldsymbol{x}_2-\boldsymbol{\mu}_2))\\ \boldsymbol{\Sigma}_{1|2}=&\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}=\boldsymbol{\Lambda}_{11}^{-1}\end{aligned}$$

其中，$\boldsymbol{\Lambda}=\boldsymbol{\Sigma}^{-1}=\begin{bmatrix}\boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12}\\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22}\end{bmatrix}$。

(2) 设两个高斯随机概率密度函数 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})$ 和 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22})$，它们的乘积为

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22})=Z^{-1}\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

其中，$\boldsymbol{\mu}=\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2\right)$，$\boldsymbol{\Sigma}=\left(\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{22}^{-1}\right)^{-1}$。

7. Kullback-Leibler 散度 (KL 距离，相对熵)

KL 散度用于衡量两个概率密度函数 $p$ 和 $q$ 的相似程度，其定义是

$$\mathrm{KL}(p\|q)=\int p(x)\ln\frac{p(x)}{q(x)}\mathrm{d}x$$

KL 散度可进一步表示为

$$\mathrm{KL}(p\parallel q)=\int p(x)\ln p(x)\,\mathrm{d}x-\int p(x)\ln q(x)\,\mathrm{d}x$$

$$= -H(p) + H(p,q)$$

其中，$H(p) = -\int p(x)\ln p(x)\,\mathrm{d}x$ 表示 $p$ 的熵 (Information Entropy)，$H(p,q) = -\int p(x)\ln q(x)\,\mathrm{d}x$ 表示 $p$ 和 $q$ 的交叉熵 (Cross Entropy)。

KL 散度具有一个重要性质：$\mathrm{KL}(p \parallel q) \geqslant 0$，当且仅当 $p = q$ 时，$\mathrm{KL}(p \parallel q) = 0$。

值得注意的是，$\mathrm{KL}(p \parallel q)$ 关于 $q$ 和 $q$ 不对称，$\mathrm{KL}(p \parallel q) \neq \mathrm{KL}(q \parallel p)$；它的一个对称表示形式是所谓的 Jensen-Shannon 散度：

$$\mathrm{JS}(p_1, p_2) = \mathrm{KL}(p_1 \parallel q)/2 + \mathrm{KL}(p_2 \parallel q)/2$$

其中，$q = \dfrac{p_1}{2} + \dfrac{p_2}{2}$。

# 缩 写 词 表

ADMM (Alternating Direction Method of Multipliers) 乘子交替迭代优化方法

AMP (Approximate Message Passing) 近似信息传递算法

ART (Algebraic Reconstruction Technique) 代数重建算法 (或 Kaczmarz 算法)

BP (Basis Pursuit) 基追踪

CD (Coordinate Descent Algorithm) 坐标下降算法

CoSaMP (Compressed Sampling based Matching Pursuit) 压缩采样匹配追踪算法

CV (Cross Validation) 交叉验证方法

EM (Expectation Maximization) 最大期望算法

FISTA (Fast Iterative Soft-threshold Algorithm) 加速迭代软门限算法

GMRF (General Markov Random Field) 广义马尔可夫随机场

IHT (Iterative Hard Threshold) 迭代硬门限算法

KKT (Karush-Kuhn-Tucker) 处理凸优化问题的 KKT 条件

KL (Kullback-Leibler) KL 散度 (KL 距离)

LASSO (Least Absolute Shrinkage and Selection Operate) 最小绝对收缩和选择操作

LDA (Linear Discriminant Analysis) 线性判别分析

LISA (Learned Iterative Soft-threshold Algorithm) 基于学习的迭代软门限算法

MAP (Maximum A Posterior ) 最大后验概率估计

MMSE (Minimum Mean Squared Error) 最小均方差

MP (Matching Pursuit) 匹配追踪

MRF (Markov Random Field) 马尔可夫随机场

mRIP (Matrix Restricted Isometric Property) 矩阵的受限等容条件

MUSIC (MUltiple SIgnal Classification) 多重信号分类

OMP (Orthogonal Matching Pursuit) 正交匹配追踪

PA (Sparse Analysis) 稀疏分析

PCA (Principal Component Analysis) 主成分分析

PS (Sparse Synthesis) 稀疏综合

RBM (Restricted Boltzmann Machine)　受限 Boltzmann 机

RIP (Restricted Isometric Property)　受限等容性条件

SA (Stochastic Approximation)　随机梯度优化法

SAMP (Sparsely Adaptive Matching Pursuit)　稀疏自适应匹配追踪技术

SPSA (Simultaneous-Perturbation Stochastic Algorithm)　同步扰动随机逼近算法

SURE (Stein Unbiased Risk Estimator)　Stein 无偏性风险估计

SVD (Singular Value Decomposition)　奇异值分解

# 目　　录

**前言**
**符号使用和约定**
**缩写词表**
**绪论** ····· 1
参考文献 ····· 5
**第 1 章　信号采样、表征与稀疏感知** ····· 8
1.1　Nyquist-Shannon 采样定理 ····· 8
1.2　信号表征 ····· 15
1.2.1　信号的确定性表征 ····· 15
1.2.2　信号的一般性统计描述 ····· 17
1.2.3　白化信号的统计表征 ····· 19
1.3　稀疏信号与稀疏感知 ····· 22
1.3.1　信号的稀疏性与表征 ····· 22
1.3.2　稀疏感知问题 ····· 24
附录 1A　广义信号采样方法 ····· 27
参考文献 ····· 29
**第 2 章　稀疏感知的若干数学问题** ····· 31
2.1　压缩感知 ····· 31
2.2　低秩矩阵感知 ····· 41
2.3　稀疏卷积感知 ····· 45
2.4　相位复原 ····· 49
附录 2A　三个常用的概率不等式 ····· 52
参考文献 ····· 53
**第 3 章　RIP 分析与 $L1$-正则化优化** ····· 55
3.1　广义 RIP 定义及其特性分析 ····· 55
3.2　广义 RIP 与 $L1$-最小化 ····· 64
3.3　广义 RIP 与 $L1/L2$-最小化 ····· 67
3.4　统计 RIP 与 $L1$ 最优化 ····· 70
3.5　最优测量矩阵设计 (1) ····· 75
附录 3A　$L1$ 优化估计的无偏性分析 ····· 77

参考文献 ····· 81
**第 4 章　贪婪算法** ····· 82
4.1　匹配追踪算法 ····· 82
4.1.1　正交匹配追踪算法 ····· 82
4.1.2　CoSaMP 算法 ····· 85
4.2　迭代硬门限算法 ····· 92
4.3　低秩矩阵感知的迭代硬门限算法 ····· 99
4.3.1　低秩矩阵的硬门限投影 ····· 99
4.3.2　迭代硬门限方法复原低秩矩阵 ····· 103
附录 4A　SURE 估计 ····· 104
参考文献 ····· 106
**第 5 章　梯度类凸优化方法** ····· 108
5.1　凸优化的有关概念 ····· 109
5.1.1　凸函数的定义及基本性质 ····· 109
5.1.2　拉格朗日乘子法 ····· 114
5.1.3　Fenchel 共轭函数 ····· 116
5.1.4　Bregman 距离 ····· 118
5.2　基于 Nesterov 光滑化方法的梯度优化方法 ····· 120
5.2.1　Nesterov 光滑化 ····· 120
5.2.2　梯度迭代算法的一般性描述 ····· 122
5.2.3　加速梯度迭代优化方法 ····· 131
5.3　邻近算子方法 ····· 137
5.3.1　邻近算子 ····· 138
5.3.2　迭代软门限方法 ····· 141
5.3.3　加速迭代软门限方法 ····· 145
5.4　亚梯度与 Bregman 算法 ····· 146
附录 5A　Wirtinger 导数 ····· 150
附录 5B　Pareto 曲线 ····· 151
附录 5C　基于深度神经网络的迭代软门限算法 ····· 153
附录 5D　最优测量矩阵设计 (2) ····· 155
参考文献 ····· 156
**第 6 章　面向大数据的优化方法** ····· 158
6.1　乘子交替迭代优化方法 ····· 158
6.1.1　稀疏优化问题的拉格朗日方法 ····· 158
6.1.2　ADMM 算法 ····· 161

6.1.3 Scaled-ADMM 算法 ····· 163
6.1.4 ADMM 算法的收敛性 ····· 165
6.2 随机梯度优化方法 ····· 169
6.3 随机坐标优化算法 ····· 176
6.3.1 随机坐标优化算法及收敛性分析 ····· 176
6.3.2 加速随机坐标优化算法 ····· 181
6.4 Robust 优化方法 ····· 183
6.5 维度约化 ····· 186
6.5.1 主成分分析 ····· 186
6.5.2 线性判别分析 ····· 188
6.5.3 流形学习 ····· 193
附录 6A 增强拉格朗日乘子法在矩阵分解中的应用 ····· 195
参考文献 ····· 196
**第 7 章 贝叶斯分析** ····· 198
7.1 贝叶斯分析的基本概念 ····· 198
7.1.1 贝叶斯建模 ····· 200
7.1.2 贝叶斯方法与确定性方法的关系 ····· 212
7.2 最大期望算法 ····· 216
7.3 Laplace EM-贝叶斯分析 ····· 220
7.3.1 Laplace 信号建模 ····· 221
7.3.2 Lapalce 模型的 EM-贝叶斯算法 ····· 223
7.4 最大期望-变分贝叶斯算法 ····· 227
7.5 混合高斯模型的 EM-贝叶斯分析 ····· 234
7.5.1 标准 EM-贝叶斯算法 ····· 235
7.5.2 基于分层模型的 EM-贝叶斯算法 ····· 239
7.6 基于蒙特卡罗的贝叶斯分析 ····· 242
7.6.1 蒙特卡罗采样的 Metropolis 算法 ····· 242
7.6.2 限制 Boltzmann 机 ····· 246
7.6.3 对比散度算法 ····· 247
附录 7A 常用的概率密度函数表 ····· 250
附录 7B 贝叶斯分析在盲反卷积中的应用例 ····· 251
附录 7C 最优测量矩阵设计 (3) ····· 254
附录 7D 稀疏高斯随机过程 ····· 254
附录 7E 重要性采样 ····· 256
参考文献 ····· 259

**第 8 章　信息传递算法** ····· 262
8.1　信息传递算法基本概念 ····· 262
8.2　求解 $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{n}$ 的信息传递算法 ····· 267
8.2.1　Sum-Product 近似信息传递算法 ····· 270
8.2.2　Max-Product 近似信息传递 ····· 275
8.3　Gaussian-Bernoulli 稀疏感知近似信息传递算法 I：Krzakala 方法 ··· 280
8.4　Gaussian-Bernoulli 稀疏感知近似信息传递算法 II：Schniter 方法 ···· 289
附录 8A　对 Max-Product 传递模式和 Sum-Product 传递模式的进一步讨论 ····· 296
参考文献 ····· 298

# 绪　论

获取信息、从获取的信息中认识大千世界，这是人类文明发展史中的主旋律。在 21 世纪，各种传感器技术快速发展：种类和数量不断增长、硬件成本不断降低、精度和分辨率不断增强，数据量呈现爆炸式增长，我们进入一个密集型数据挖掘和知识发现的新时代——**大数据时代**。大数据洪流给传统的数据存储、传输、处理方法带来前所未有的压力，只有发展**新型的信息表征**、**获取及存储方法**，才能突破传统数据处理、传输和存储技术的瓶颈。触手可及的海量数据促使新思维模式的产生：**从数据中学习**，形成信号表征、信息获取与复原的新概念和新方法。

1. 信号表征

自然界的大多数信号具有一定的结构性，结构性信号的自由度要远低于信号本身的维度。海量数据的典型特征是信息冗余，也就是说海量数据仅携带非常有限的信息量，我们用**稀疏**来描述这个特性 (本书第 1 章给出稀疏的严格数学定义)。稀疏为海量数据的表征与处理提供了极大便利，为更有效的数据解译提供了可能。稀疏的概念远追溯至著名的 Prony 方法 [1]：在噪声干扰情况下从稀疏复指数函数的线性组合中估计复指数参数的方法 (1795 年)。1809 年，法国数学家 Pierre-Simon Laplace(1749~1827) 提出了“$L_1$-范数测度”，将稀疏性概念引入线性代数领域 [2]。Gauss 对 Laplace 此工作的评价是 “Laplace made use of another principle for the solution of linear equations, the number of which is greater than the number of unknown quantities ……”。20 世纪末，美国斯坦福大学的 Tibshirani 等 [3] 和 Chen 等 [4] 分别从机器学习和信号处理的角度几乎同时提出了 $L_1$ 正则化的稀疏线性模型，揭示了稀疏性在信号表征中的重要作用。

信号表征的基本任务是描述信号组成的基本要素，揭示信号组织和生成的方式。离散余弦变换和小波变换等经典信号表征方法已经在信号、图像、视频分析等领域发挥着重要的作用，但它们仅描述信号的基本要素和简单线性组织方法，不能揭示信号深层次的组织和生成方式。除此之外，这些经典信号表征方法无法对特定应用 (例如，指纹、脑信号等) 的信号特征进行有效表征。在统计分析和机器学习的驱动下，尤其是深度学习 [5,6] 的驱动下，从数据中学习信号特征的表征方法开始崭露头角，信号表征进入了**深度表征**的时代。区别于传统的信息表征方式，深度表征以信息深度组织方式的形式进行信息表征 (图 0.1)，

并使信息表征与类别认知结合在一起，更具有结构性，更适合完成特定任务的需求。

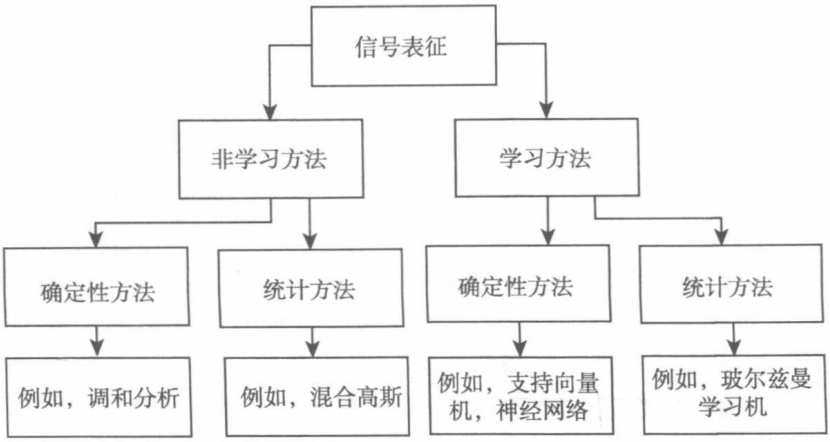



图 0.1　信号表征的一些方法

2. 信息复原

信息复原主要指信息复原算法以及与之相关的理论。本书第 4 章至第 8 章分别从确定和概率的角度研究多种稀疏感知的信号复原算法，重点研究面向大尺度稀疏感知问题的信号复原算法。

由贝叶斯 (Thomas Bayes，1701~1761) 公式可知

$$P(\text{信息}|\text{数据}) = \frac{P(\text{数据}|\text{信息})}{P\left(\text{数据}\right)} \times P\left(\text{先验}\right)$$

表明了信息的推断由数据和先验共同决定，前者反映客观推理，后者反映主观认知。从概率分析的角度看，经典信号复原算法旨在最大化似然函数 $\dfrac{P(\text{数据}|\text{信息})}{P\left(\text{数据}\right)}$，它仅处理测量数据。然而稀疏感知算法旨在分析处理$\dfrac{P(\text{数据}|\text{信息})}{P\left(\text{数据}\right)} \times P\left(\text{先验}\right)$，它不仅处理数据，也分析信号的先验信息 $P(\text{先验})$ 对结果的影响和作用。由于经典信号处理仅限于考虑目标信号本身，而没有考虑信号内在深层次的信息 (即使这些信息已被很好认知或将被认知)，因此经典的信号理论的许多结论过于保守和悲观。例如，经典的 Nyquist-Shannon 采样定理就过于狭隘，它仅利用了信号的带宽信息 [7]。再比如，经典的 Rayleigh 分辨率极限也是一个显得过于狭隘的结论 [8]，它也是只利用系统带宽信息的产物。稀疏感知信号复原算法不仅处理数据，而且挖掘利用信号的先验信息，将信号的先验认识有机融入信号的获取和处理中，为发展高性能信息感知奠定了基础。