

压缩感知和稀疏优化简介

文再文¹ 印卧涛^{2†} 刘 歆³ 张 寅²

摘要 介绍压缩感知和稀疏优化的基本概念、理论基础和算法概要. 压缩感知利用原始信号的稀疏性, 从远少于信号元素个数的测量出发, 通过求解稀疏优化问题来恢复完整的原始稀疏信号. 通过一个小例子展示这一过程, 并以此说明压缩感知和稀疏优化的基本理念. 接着简要介绍用以保证 ℓ_1 凸优化恢复稀疏信号的零空间性质和 RIP 条件. 最后介绍求解稀疏优化的几个经典算法.

关键词 压缩感知, 稀疏优化, 零空间性质, 受限正交条件, 紧缩算子, 线性化近似点算法, 分裂 Bregman 方法和交替方向增广拉格朗日函数法, Bregman 方法和增广拉格朗日函数法

中图分类号 O224, O29

数学分类号 49M29, 65K05, 90C25, 90C06

Introduction to compressive sensing and sparse optimization

WEN Zaiwen¹ YIN Wotao^{2†} LIU Xin³ ZHANG Yin²

Abstract We briefly introduce the basic principle and theory of compressive sensing and sparse optimization. Compressive sensing is a new paradigm of signal acquisition, which senses a sparse signal by taking a set of incomplete measurements and recovers the signal by solving an optimization problem. This article first illustrates the compressive sensing paradigm through a synthetic example. Then we describe two sufficient conditions, the null space property and restricted isometry principle, for ℓ_1 convex minimization to give the sparsest solution. Finally, we summarize a few typical algorithms for solving the optimization models arising from compressive sensing.

Keywords compressive sensing, sparse optimization, null space property, RIP, shrinkage, prox-linear algorithms, split Bregman/alternating direction augmented Lagrangian method, Bregman/augmented Lagrangian method

Chinese Library Classification O224, O29

2010 Mathematics Subject Classification 49M29, 65K05, 90C25, 90C06

收稿日期: 2012 年 7 月 6 日

1. 上海交通大学数学系, 自然科学研究院和科学工程计算教育部重点实验室, 上海, 200240; Department of Mathematics, Institute of Natural Sciences and MOE-LSC, Shanghai Jiaotong University, Shanghai 200240, China

2. 美国莱斯大学计算与应用数学系, 休斯顿 77005, 德克萨斯州, 美国; Department of Computational and Applied Mathematics, Rice University, Houston 77005, TX, USA

3. 中国科学院数学与系统科学研究院, 北京, 100190; Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

† 通讯作者 Corresponding author

0 背景介绍

“稀疏性”一词最近频频出现在科学及工程学科的各个领域的出版物中。稀疏是指信号的绝大多数元素为 0。与同样长度的普通信号相比, 一个稀疏信号包含较少的信息。因此, 稀疏信号可以被充分地压缩, 从而节约储存空间、减少传输量。最近 30 年左右, 稀疏性已经在信息的处理、分析和传输领域内展现了其强大的力量。特别在最近, 新兴的研究方向压缩感知 (也称作压缩采样和压缩传感, Candés and Tao^[1] and Donoho^[2]) 通过利用信号、信息的稀疏性, 大大提高了信号采集能力、信息处理能力, 缓解了海量数据的采样、存储、传输和分析负担。包括压缩感知在内, 大量的跨学科工作利用了稀疏性。在这些工作中, 稀疏优化起到了核心的作用, 得到了大量的关注。

稀疏优化旨在通过求解优化模型而获得稀疏解。它的理论、模型和算法正在迅猛发展。稀疏优化是最优化领域中相对年轻的一个分支。它的特点在于未知解的稀疏性。与之不同的是, 经典优化技术对稀疏性的利用着重于输入数据, 利用它们的稀疏性等结构进行计算加速和存储空间减少。在稀疏优化中, 由于解是未知的, 如何利用解的稀疏性进行快速和准确地计算是一个崭新的、富有挑战的研究方向。

本文将通过一些小例子及简单的推导让读者对稀疏优化和压缩感知有一个基本的、直观的了解。由于篇幅限制, 我们无法对很大一部分国内外同行的精彩相关工作进行介绍, 更无法像教科书那样将所有细节陈述清楚。我们将主要介绍最基本的和最有助理解的结果。

0.1 压缩感知的应用案例

压缩感知将一个长度为 n 的原始稀疏信号线性编码为一个长度缩短至 m 的压缩信号, m 大致和原始信号的非零元素个数 k 以及长度的对数的多项式成比例, 即 $m = O(k \text{ poly}(\log(n)))$ 。因此, 压缩后的长度 m 可以远小于原始长度 n 。然而, 解码过程需要求解一个稀疏优化问题或使用稀疏解码算法。这个过程与传统的感知方法中, 先采集信号、然后压缩、再解码的过程非常不同。因此, 压缩感知的主要特点可以归纳为 (编码兼采样) 感知少、(稀疏解码) 计算多。由于感知少的特点, 压缩感知为很多感知难题带来了重大突破, 也为一些已有的感知应用带来了显著改进。

我们通过一个简单的例子说明压缩感知的过程。假设我们要采集图 1(a) 中长度为 $n = 200$ 的稀疏信号 $x^o \in \mathbb{R}^n$ (上标 o 代表 original 即原始信号)。令 $k = \|x^o\|_0$ 为 x^o 非零元的个数 (这里 $k = 10$)。首先, 将 x^o 用线性变换 R 编码为一个压缩信号 $b = Rx^o \in \mathbb{R}^m$, $m < n$ 。由于信号尚未被采集转换为数字格式, 我们这里的“编码”通常不是由计算机完成, 而是通过物理手段直接得到 (如光学叠加、电磁波叠加等)。注意到 x^o 尚且未知, R 的设计必须独立于 x^o , 却要求对至少绝大多数可能的稀疏信号 x^o 均有效。在本例中, $R \in \mathbb{R}^{m \times n}$ 为从 n 维离散余弦变换 (DCT) 矩阵 Φ 中随机选取的 $m = 80$ 行。数 m 被称为采样值。图 1(b) 和 (c) 分别描述了 Φx^o 和 Rx^o 。缺失的测量, 即在 Φx^o 而不在 Rx^o 中的元素在图 1(c) 中由零代替。在压缩信号 $b = Rx^o$ 由传感器采集后, 我们需要恢复原始信号。尽管由于 $m < n$ 线性方程组 $Rx = b$ 有无穷多个解, 利用 x^o 中的非零元个数 $\|x^o\|_0$ 很小这个性质, 我们依旧通过求解最小 ℓ_0 问题

$$\min_{x \in \mathbb{R}^n} \{\|x\|_0 : Ax = b\} \quad (0.1)$$

来得到 x^o , 其中 $A = R$, x 的 ℓ_0 伪范数是指 x 非零元素的个数. 如果 $Rx = b$ 不存在另一个和 x^o 同样稀疏甚至更稀疏的解, x^o 将是问题 (0.1) 的解.

然而, 对于一般情况, ℓ_0 问题 (0.1) 是 NP 难的组合优化问题 [3]. 一个更可行的方案是求解凸优化 ℓ_1 问题 (也被称为基追踪问题):

$$\min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = b\}. \quad (0.2)$$

我们将在下节中指出, 问题 (0.2) 与 ℓ_0 问题在某些关于稀疏度和编码矩阵的条件下, 具有相同的解.

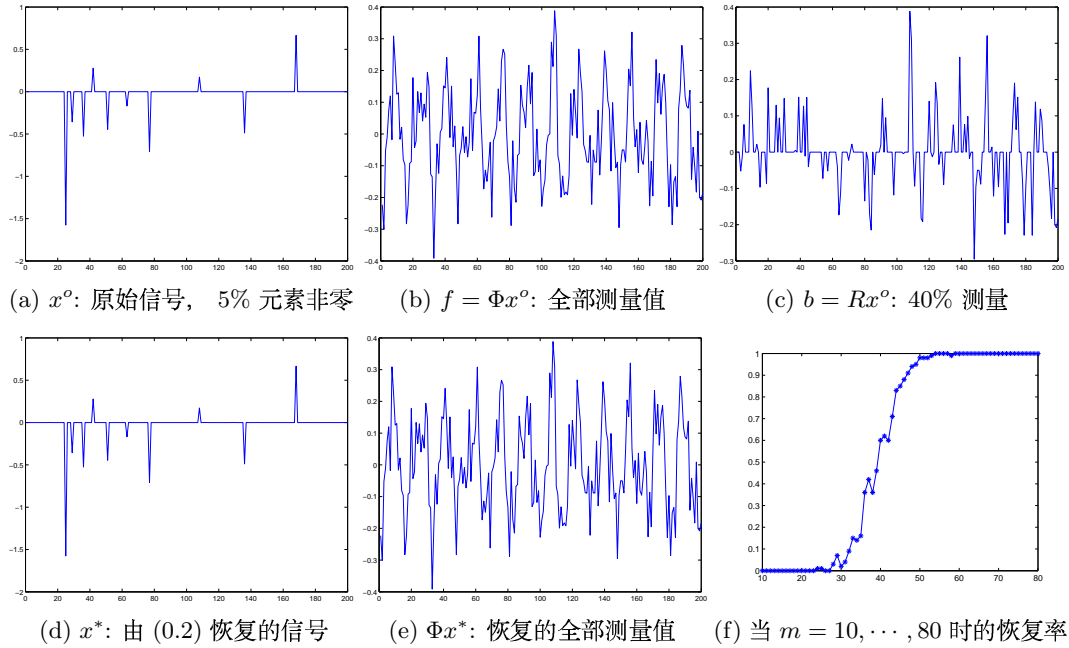


图 1 降采样时的信号恢复, 其中 Φ 是离散余弦变换

因为信号中的非零元素数量 k 代表了信号中的信息量, 因此我们自然希望只进行 $m = k \equiv \|x^o\|_0$ 次测量. 然而, 由于我们不知道 x^o 中非零元素的位置 (一共有 C_k^n 种可能性), 作为代价, 我们必须进行额外的采样. 文献 [1,4,5] 指出, 当 R 是某些随机矩阵时, 用 $m = O(k \log(n/k))$ 次测量, 通过求解 (0.2) 恢复原始信号, 能够以很大的概率得到 x^o . 虽然测量次数 m 大于 k , 只要 $k \ll n$, 这样的 m 依旧远小于 n . 为了通过图例展示这一点, 我们对每一个 $m = 10, 11, \dots, 80$, 进行 100 次独立的测试. 图 1(e) 描述了成功恢复的几率. 它表明对我们所测试的 k, n 和 A 的组合, 通常来讲取 $m > 6k = 60$ 就会得到极高的成功率.

0.2 隐式稀疏

除了空间和时间上稀疏的信号, 压缩感知还适用于很多其它的问题. 比如, 大部分图像的像素元素并不稀疏, 但在傅立叶或小波变换下具有 (近似) 稀疏表达. 令向量 u^o 表示一个图像. JPEG2000 等压缩算法将寻找一个可逆矩阵 Ψ , 即小波基, 使得向量 $x^o = \Psi u^o$

有相对少量的大绝对值元素. 令 \tilde{x} 为仅取 x^o 的大绝对值元素的稀疏向量. 则 u^o 可以比较精确的由 $\Psi^{-1}\tilde{x}$ 近似. 由于 Ψu^o 是稀疏的, 人们可以从 $b = Ru^o$ 通过求解

$$\min_u \{\|\Psi u\|_1 : Ru = b\}, \quad (0.3)$$

来恢复 u^o . 上式等价于 (0.2) 中令 $x = \Psi u$ 和 $A = R\Psi^{-1}$.

和图像一样, 我们关心的许多信号都具有规律性, 它们在某些线性的表达下的系数是 (近似) 稀疏的. 在压缩感知中, 如果信号的稀疏性被有效地利用, 该信号便能从其相对少量的测量中被恢复. 然而, 对于给定信号, 找到一个好的稀疏表达通常并不容易. 最近, 在低光成像、医学成像中的核磁共振成像 (MRI) 和计算机断层成像 (CT)、红外线传感、生物传感、雷达信号处理、多传感器网络、分布式传感和模拟信息转换等多个领域的信号有了不少进展. 感兴趣的读者可以访问 Rice 大学压缩感知网站 [6].

0.3 传统感知与压缩感知的差别

传统的感知方式是先采样后处理. 今天的数码相机就是如此, 感光元件 CCD 或者 CMOS 先将光子转化为数字图像, 计算元件再将数字图像压缩成图像文件. 与传统的感知手段相比, 压缩感知采样更少但计算更多. 采样少意味着采样更快或采用更小更便宜的传感器. 比如, 在核磁共振和断层成像等医学成像方法中, 使用压缩感知可以减少图像采集的时间, 使得诊断过程更安全、更舒适, 费用更合理. 另一方面, 由于传统感知采样更多, 后继压缩过程可以更灵活、更有效、达到更高的压缩比, 因此能更有效地降低存储空间和传输时间. 这两种感知方式各有自己的优点, 均不能代替彼此. 它们适用于不同的需求, 达到不同的效果, 并且相互补充. 但是, 因为压缩感知是一种新的感知方式, 许多研究者和工程师对它的理论和计算还不是很熟悉. 本文接下来的部分对压缩感知的基本理论和算法做一个简要的介绍.

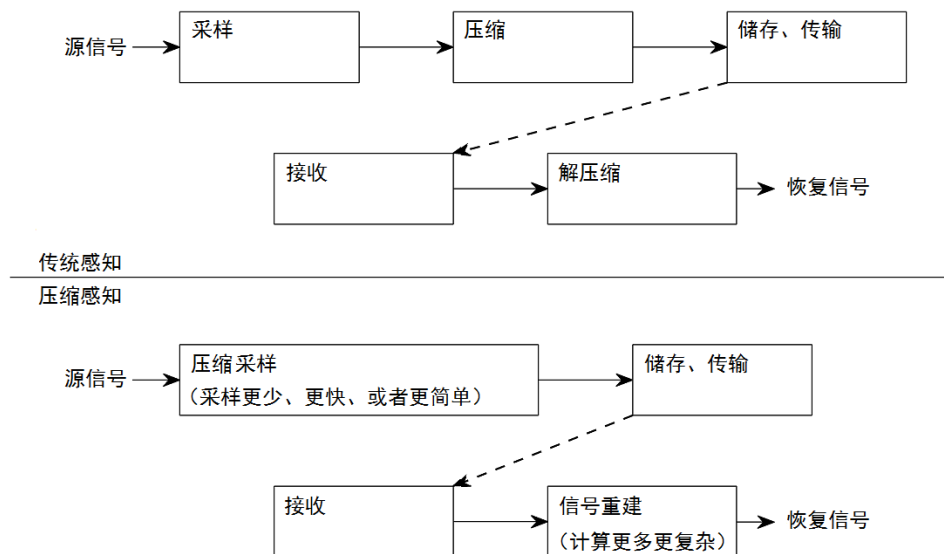


图 2 传统感知与压缩感知的差别

1 ℓ_0 和 ℓ_1 问题等价条件

如果原始信号 x^o 充分稀疏并且我们的测量是精确的, 也即 $b = Ax^o$, 则在某些条件下, 如零空间性质 (Null Space Property) 或受限正交条件 (Restricted Isometry Property, RIP) 等满足时, x^o 就是基追踪问题 (0.2) 的唯一解. 这些性质各有长处, 它们保证了存在一大类感知矩阵既满足压缩感知的要求又有可恢复性的保证. 这些矩阵有元素是由高斯分布、傅立叶变换、Walsh-Hadamard 等采样得到的矩阵; 随机 Toeplitz 或循环矩阵等等. 这一节我们将给大家概要性地介绍这些条件.

1.1 零空间性质

非常自然地, 我们知道 x^o 是问题 (0.2) 的唯一解的充分必要条件是

$$\|x^o + h\|_1 > \|x^o\|_1, \quad \forall h \in \text{Null}(A) \setminus \{0\}, \quad (1.1)$$

这里 $\text{Null}(A) := \{y \mid Ay = 0\}$ 是指 A 的零空间.

如果我们假设 $\mathcal{S} := \{i \mid x_i^o \neq 0\}$, 则 $\mathcal{S}^c := \{i \mid x_i^o = 0\}$. 则通过简单推导, 有

$$\begin{aligned} \|x^o + h\|_1 &= \|x_{\mathcal{S}}^o + h_{\mathcal{S}}\|_1 + \|0 + h_{\mathcal{S}^c}\|_1 \\ &= \|x^o\|_1 + (\|h_{\mathcal{S}^c}\|_1 - \|h_{\mathcal{S}}\|_1) + (\|x_{\mathcal{S}}^o + h_{\mathcal{S}}\|_1 - \|x_{\mathcal{S}}^o\|_1 + \|h_{\mathcal{S}}\|_1). \end{aligned}$$

不难看出, 上式最后一项根据三角不等式是大于或等于零的, 于是, 如果 $\|h_{\mathcal{S}^c}\|_1 > \|h_{\mathcal{S}}\|_1$, 则 (1.1) 成立. 这就是我们定义如下零空间性质的初衷.

定义 1.1 如果条件

$$\|h_{\mathcal{S}}\|_1 < \|h_{\mathcal{S}^c}\|_1, \quad (1.2)$$

对于所有 A 的零空间中非零向量 h 和所有满足 $|\mathcal{S}| \leq k$ 的坐标集合 $\mathcal{S} \subset \{1, 2, \dots, n\}$ 都成立, 则矩阵 A 满足 k 阶零空间性质.

直观上讲, 我们可以将零空间性理解解为 A 零空间中所有非零向量的非零元分布均匀, 质量不会集中在任意 k 个元素中.

利用零空间性质, 且注意到三角不等式可以相等, 我们有如下定理成立.

定理 1.1(零空间性质条件)^[7-9] 问题 (0.2) 从测量值 $b = Ax^o$ 唯一恢复所有 k -稀疏解 x^o 当且仅当 A 满足 k 阶零空间性质.

通过一些简单的分析, 我们可以得到基于零空间性质但更直观的条件, 它们可以为我们设计合适的感知矩阵提出有建设性的指导. 我们这里简单介绍其中一个.

我们知道 (1.1) 等价于 $\|h\|_1 > 2\|h_{\mathcal{S}}\|_1$. 考虑到 $h_{\mathcal{S}}$ 的长度等于集合 \mathcal{S} 的大小, 并且等于 $\|x^o\|_0$, 以及 $h_{\mathcal{S}}$ 是 h 的一个子向量, 有

$$\|h_{\mathcal{S}}\|_1 \leq \sqrt{|\mathcal{S}|} \|h_{\mathcal{S}}\|_2 \leq \sqrt{\|x^o\|_0} \|h\|_2.$$

进而我们就得到充分条件

$$\sqrt{\|x^o\|_0} < \frac{1}{2} \frac{\|h\|_1}{\|h\|_2}, \quad \forall h \in \text{Null}(A) \setminus \{0\}. \quad (1.3)$$

这个条件对原始信号 x^o 的稀疏性作出要求. 它保证 (1.1) 成立.

关于 (1.3) 的右端项, Kashin, Garnaev 和 Gluskin 曾分别给出如下定理.

定理 1.2(Kashin-Garnaev-Gluskin 定理)^[10-11] 对任意 $p < n$, 存在 p 维子空间 $V_p \subset \mathbb{R}^n$, 使得

$$\frac{\|h\|_1}{\|h\|_2} \geq \frac{C\sqrt{n-p}}{\sqrt{\log(n/(n-p))}}, \forall h \in V_p \setminus \{0\} \quad (1.4)$$

成立, 其中 C 是一个与维数 p 和 n 无关的常数.

根据统计中的经典理论^[12], 我们可以进一步证明任意由 p 个服从独立同分布的随机向量张成的 p 维子空间都以极大的概率满足 (1.4). 在压缩感知中, 如果用独立同分布的标准高斯随机数产生 A , 令定理 1.2 中的 V_p 为 A 的零空间, p 为 $n-m$, 那么 V_p 以极大的概率满足 (1.4). 结合我们需要的条件 (1.3), 我们就得到了关于信号稀疏度的充分条件

$$\|x^o\|_0 < \frac{C^2}{4} \frac{m}{\log(n/m)}. \quad (1.5)$$

该条件等价于观测次数的条件 $m \geq O(k \log(n/k))$. 在这些条件下, x^o 是模型 (0.2) 的唯一解. 实际上, 零空间条件 (1.3) 还保证用模型 (0.2) 重建信号是稳定的^[18](稳定性讨论见下节).

1.2 受限正交条件 (RIP)

现在描述关于 A 正交性的性质. 这个性质比零空间性质相对更为有限, 很多基于 RIP 的结果是通过建立零空间性质而获得的. 另一方面, 它受到了很多研究者的关注. 很多算法, 包括优化、贪婪、迭代硬阈值算法, 它们的许多分析是基于 RIP 假设的. 有不少随机矩阵以很大的概率满足 RIP 性质.

定义 1.2^[13] 如果对所有 k -稀疏的向量 $x \in \mathbb{R}^n$, 满足

$$(1 - \delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k)\|x\|_2^2 \quad (1.6)$$

的最小常数 δ_k 小于 1, 则称矩阵 A 满足 RIP 性质. 该常数 δ_k 被称为 RIP 常数.

注意到只要在缩放后 A 满足 (1.6), 我们通常就说 A 满足 RIP. 由于压缩感知使用的矩阵 A 的列数大于行数, A 并不能在通常意义下正交. 另一方面, RIP 只要求 A 限制在 k -稀疏的向量的集合上近似正交. 如果该性质对距正交性有 δ_{2k} 最大偏差的 $2k$ -稀疏的向量满足, 则 (0.2) 能保证恢复所有 k -稀疏的向量. 文献证明了 $\delta_{2k} < 0.4142$ 就足够了, 后来提高到 $\delta_{2k} < 0.4531$ (见 [14]), $\delta_{2k} < 0.4652$ (见 [15]), $\delta_{2k} < 0.4721$ (见 [16]), 以及 $\delta_{2k} < 0.4931$ (见 [17]).

定理 1.3(精确和稳定恢复的 RIP 条件)^[17] 考虑模型:

$$b = Ax^o + w, \quad (1.7)$$

其中 w 是噪音或者测量误差. 如果 A 满足 $\delta_{2k} \leq 0.4931$ 的 RIP 条件, 则 (0.2) 的解 x^* 满足

$$\|x^* - x^o\|_1 \leq C_1 \cdot \sqrt{k}\|w\|_2 + C_2 \cdot \sigma_{[k]}(x^o), \quad (1.8)$$

$$\|x^* - x^o\|_2 \leq \bar{C}_1 \cdot \|w\|_2 + \bar{C}_2 \cdot \sigma_{[k]}(x^o)/\sqrt{k}, \quad (1.9)$$

其中 C_1, C_2, \bar{C}_1 和 \bar{C}_2 都是常数, $\sigma_{[k]}(x^o)$ 是将 x^o 按照元素 (绝对值或者模) 的大小排序后最小的 $n - k$ 个元素的 ℓ_1 模 ($\sigma_{[k]}(x^o)$ 也称作最近 k 项近似的误差).

这个结果表明只要感知矩阵 A 满足 RIP, 则 ℓ_1 恢复对于噪声和最佳 k - 项近似的误差是稳定的. 当 $w = 0$ 且 x^o 是 k - 稀疏的, 不等式 (1.8) 和 (1.9) 给出了 $x^* = x^o$ 的精确恢复条件. 当 $w \neq 0$, 即使 x^o 的最大 k 项的位置已知, 最小二乘法的误差为 $O(\|w\|_2)$, 因此右端项的误差 $O(\|w\|_2)$ 不能被显著地减小. 当 $\sigma_{[k]}(x^o) \neq 0$, 误差 $O(\sigma_{[k]}(x^o))$ 在以幂方衰减的信号中达到, 所以以上误差估计已经是最紧的了. 如果 x^o 的衰减服从幂方衰减: x^o 的第 i 大的元素的大小服从 $|x_{(i)}| = C \cdot i^{-r}$, $r > 1$, 则有 $\sigma_{[k]}(x^o)/\sqrt{k} = O(k^{-r+1/2})$. 另一方面, 即使令 $x^* = x_{[k]}^o$, 我们仍然得到 $\|x^* - x^o\|_2 = O(k^{-r+1/2})$. 因此, 除非对信号和噪音做进一步假设, (1.8) 和 (1.9) 里的界是紧的. 另外, RIP 的第一个条件 (1.6) 对于在任意的噪声 w 下稳定恢复是必要的. 然而, RIP 对近似稀疏信号的稳定恢复不是必要的, 从比 RIP 更松的 SSP 条件^[18–19] 就可得到近似稀疏信号的稳定恢复. 当 $m > O(k * \log(n/k))$ 时, 高斯随机矩阵以及其它一些随机矩阵能以很大概率满足 RIP 条件.

2 稀疏优化算法

2.1 稀疏优化模型

上一节介绍了稀疏优化的模型及其可恢复性的条件, 这一节我们介绍求解这些问题的算法. 当介绍算法时, 我们着重考虑如下 ℓ_1 问题:

$$\min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = b\}, \quad (2.1)$$

$$\min_{x \in \mathbb{R}^n} \|x\|_1 + \frac{\mu}{2} \|Ax - b\|_2^2, \quad (2.2)$$

$$\min_{x \in \mathbb{R}^n} \{\|x\|_1 : \|Ax - b\|_2 \leq \sigma\}, \quad (2.3)$$

其中 $A \in \mathbb{R}^{m \times n}$. 当 ℓ_1 范数被相应的正则函数代替时, 这些模型有很多变形, 比如, 模型

$$\min_{s \in \mathbb{R}^n} \{\|s\|_1 : A\Psi s = b\}, \quad (2.4)$$

$$\min_{u \in \mathbb{R}^n} \{TV(u) : \mathcal{A}(X) = b\}, \quad (2.5)$$

$$\min_{X \in \mathbb{R}^{m \times n}} \{\|X\|_{2,1} : \mathcal{A}(X) = b\}, \quad (2.6)$$

$$\min_{X \in \mathbb{R}^{m \times n}} \{\|X\|_* : \mathcal{A}(X) = b\}, \quad (2.7)$$

分别被广泛应用到求解在基 Ψ 下稀疏的向量、分段常数的信号、由联合稀疏向量组成的矩阵和低秩矩阵等等. 这里, $TV(u)$ 是 u 的全变差, $\|X\|_{2,1}$ 是 X 每一行的二范数之和, $\|X\|_*$ 是 X 的奇异值之和, $\mathcal{A}(X) = [A_1 \bullet X, A_2 \bullet X, \dots, A_M \bullet X]$, 其中 $Y \bullet X = \sum_{m=1}^M \sum_{n=1}^N y_{mn} x_{mn}$.

本节的算法能被很自然地推广到这些变形.

问题 (2.1)–(2.7) 为非光滑的凸问题. 当对偶问题有比较好的性质时, 求解对偶问题可能比直接求解原始问题更有效. 我们这里简单的写出几个对偶形式. 比如问题 (2.1) 的对偶问题是

$$\max_y \{b^T y : \|A^T y\|_\infty \leq 1\}. \quad (2.8)$$

而问题 (2.7) 的对偶问题为

$$\max_y \left\{ b^T y : \left\| \sum_{m=1}^M y_m A_m \right\|_2 \leq 1 \right\},$$

其中 $\|X\|_2 = \sigma_1(X)$ 是 X 的最大奇异值.

上述问题中大部分能直接转换成凸锥规划的标准形式. 比如问题 (2.1) 等价于

$$\min_{x_1, x_2} \{ \mathbf{1}^T (x_1 + x_2) : A(x_1 - x_2) = b, x_1 \geq 0, x_2 \geq 0 \}, \quad (2.9)$$

其中 x_1, x_2 分别对应于 x 的正和负的部分. 假设 (x_1^*, x_2^*) 是 (2.9) 的解, 则 $x^* = x_1^* - x_2^*$ 是 (2.1) 的解. 问题 (2.9) 是线性规划, 可以用单纯形法和内点法来求解. 问题 (2.2) 和 (2.3) 可以转换为二次锥规划问题, 问题 (2.7) 能转换为半定规划, 均可以由内点法求解. 虽然单纯形法和内点法被广泛使用, 许多稀疏优化问题, 特别是图像和高维数据处理中的数据量, 对它们而言规模太大. 因此, 相关算法软件包往往运行得很慢或者报告内存不够.

与单纯形法和内点法相比, 本节介绍的算法不需要对维数很大的矩阵求逆或者进行分解. 它们通常只需要计算矩阵和向量的乘积, 以及对某些简单的子问题进行显式求解. 对于有稀疏解的问题, 这些算法通常比标准的单纯形法和内点法在 CPU 和内存上更有效率. 由于篇幅限制, 本文不能涉及大多数稀疏优化的算法. 甚至对提到的算法, 我们也仅仅给出他们的背景和基本性质. 详细的分析请参考相关文献.

2.2 紧缩算子

紧缩算子 (Shrinkage) 和它的扩展在许多稀疏优化算法中被广泛使用. 虽然 ℓ_1 范数不可微, 但它对于每个元素来讲是可分的, 即 $\|x\|_1 = \sum_{i=1}^N |x_i|$. 因此, 当目标函数还有其它元素可分的函数时, 很容易被极小化. 特别地, 当 $\tau > 0$ 时, 问题

$$\min_x \|x\|_1 + \frac{1}{2\tau} \|x - z\|_2^2 \quad (2.10)$$

的显式解为

$$(\text{shrink}(z, \tau))_i = \begin{cases} z_i - \tau, & z_i \geq \tau, \\ 0, & -\tau < z_i < \tau, \\ z_i + \tau, & z_i \leq -\tau. \end{cases} \quad (2.11)$$

令 $Z = U\Sigma V^T$ 是矩阵 Z 的奇异值分解, $\hat{\Sigma}$ 是对角元为

$$\text{diag}(\hat{\Sigma}) = \text{shrink}(\text{diag}(\Sigma), \tau) \quad (2.12)$$

的对角矩阵, 则我们有矩阵形式的紧缩算子

$$U\hat{\Sigma}V^T = \arg \min_X \|X\|_* + \frac{1}{2\tau} \|X - Z\|_F^2. \quad (2.13)$$

图像处理中经常用到的另外一个扩展与全变差有关,

$$\min_u TV(u) + \frac{1}{2\tau} \|u - z\|_2^2. \quad (2.14)$$

虽然问题 (2.14) 没有显式解, 但它有非常有效的求解方法, 比如图分割算法^[20-21], 原始—对偶算法^[22], 以及 [23-24].

2.3 线性化近似点算法

问题 (2.2) 可以写成广义形式

$$\min_x r(x) + f(x), \quad (2.15)$$

其中 r 是正则函数, f 是数据拟合函数. 线性化近似点算法有着很长的历史, 比如 [25-27]. 它的基本思想是在迭代点 x^k 处线性化函数 $f(x)$ 并添加近似项, 得到新的迭代格式

$$x^{k+1} = \arg \min_x r(x) + f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2\delta_k} \|x - x^k\|_2^2. \quad (2.16)$$

最后一项使得 x^{k+1} 靠近 x^k , 而 δ_k 是步长因子. 保留 r 而线性化 f 的主要目的是子问题 (2.16) 的解容易计算.

例 2.1(稀疏解恢复) 当 $r = 0$ 时, (2.16) 得到的迭代格式就是最速下降法. 假设 $r(x) = \|x\|_1$, $f(x)$ 是可微函数 (如 (2.2) 中的 $\frac{\mu}{2} \|Ax - y\|_2^2$), 则 (2.16) 变成了

$$x^{k+1} = \text{shrink}(x^k - \delta_k \nabla f(x^k), \delta_k^{-1}). \quad (2.17)$$

可以看出, 在第 k 步的主要计算量是 $\nabla f(x^k)$ 的计算 (如 (2.2) 中的 $A^T A x^k$). 算法 FPC^[28] 就是基于 (2.17).

如果 $r(x) = \|\Psi x\|_1$, 而且 Ψ 是一个正交线性变换, 则 (2.16) 给出:

$$x^{k+1} = \Psi^T \text{shrink}(\Psi(x^k - \delta_k \nabla f(x^k)), \delta_k^{-1}). \quad (2.18)$$

例 2.2(低秩矩阵恢复) 令 $r(X) = \|X\|_*$ 和 $f(X) = \frac{\mu}{2} \|A(X) - b\|_2^2$. 则 (2.16) 迭代格式为

$$X^{k+1} = \text{shrink}_*(X^k - \delta_k \nabla f(X^k), \delta_k^{-1}), \quad (2.19)$$

其中 shrink_* 由 (2.12) 和 (2.13) 定义. 算法 FPCA^[29] 就是基于 [31].

线性化近似点算法通常和如 continuation 子空间方法等技巧一起使用, 达到显著的加速效果.

2.4 Bregman 方法和增广拉格朗日函数法

本小节介绍求解问题 (2.1) 的 Bregman 算法, 并阐述它与增广拉格朗日函数法的关系.

定义 2.1^[30] 由凸函数 $r(\cdot)$ 诱导出来的 Bregman 距离定义为

$$D_r(x, y, p) = r(x) - r(y) - \langle p, x - y \rangle, \quad \text{其中 } p \in \partial r(y). \quad (2.20)$$

从 $k = 0$ 和 $(x^0, p^0) = (0, 0)$ 开始, 原始的 Bregman 迭代有两步:

$$x^{k+1} = \arg \min_x D_r(x, x^k; p^k) + \frac{\delta}{2} \|Ax - b\|_2^2, \quad (2.21a)$$

$$p^{k+1} = p^k - \delta A^T(Ax^{k+1} - b). \quad (2.21b)$$

由 (2.21a) 的一阶最优条件:

$$0 \in \partial r(x^{k+1}) - p^k + \delta A^T(Ax^{k+1} - b), \quad (2.22)$$

步 (2.21b) 满足 $p^{k+1} \in \partial r(x^{k+1})$.

Bregman 算法的一个有意思的变形为“残差回加”迭代:

$$b^{k+1} = b + (b^k - Ax^k), \quad (2.23a)$$

$$x^{k+1} = \arg \min_x r(x) + \frac{\delta}{2} \|Ax - b^{k+1}\|_2^2, \quad (2.23b)$$

其中 $k = 0$ 和 $(x^0, b^0) = (0, 0)$. 可以证明 (2.21) 和 (2.23) 产生的迭代点列相同, 并且 $p^k = -\delta A^T(Ax^k - b^k)$.

另一方面, (2.23) 实际上是增广拉格朗日函数法的一种形式. 增广拉格朗日函数法常用的另一种形式是:

$$x^{k+1} = \arg \min_x r(x) - (y^k)^T(Ax - b) + \frac{\delta}{2} \|Ax - b\|_2^2, \quad (2.24a)$$

$$y^{k+1} = y^k + \delta(b - Ax^{k+1}). \quad (2.24b)$$

其中, y 称为拉格朗日乘子. 可以看出 $y^k = \delta b^{k+1}$, 因此算法 (2.23) 与 (2.24) 等价.

通常 Bregman 算法中步 (2.21a) 没有显式解, 需要使用如线性化近似点等算法来求解. 因此, 线性 Bregman 算法将线性化应用到 (2.21a) 中的 $\frac{\delta}{2} \|Ax - b\|_2^2$, 添加近似点项, 并得到如下迭代:

$$x^{k+1} \leftarrow \arg \min_x D_r(x, x^k; p^k) + \langle \delta A^T(Ax^k - b^k), x \rangle + \frac{1}{2\alpha} \|x - x^k\|_2^2, \quad (2.25a)$$

$$p^{k+1} \leftarrow p^k - \delta A^T(Ax^k - b^k) - \frac{1}{\alpha}(x^{k+1} - x^k). \quad (2.25b)$$

2.5 交替方向增广拉格朗日函数法及分裂 Bregman 方法

Bregman 方法和增广拉格朗日函数法中的子问题没有显式解. 本节讨论交替方向增广拉格朗日函数法. 分裂 Bregman 方法是对于类似 (2.2) 的无约束模式而设计的. 在特定条件下, 它等价于交替方向增广拉格朗日函数法. 这些方法中, 子问题通常可以写出显式解, 因此比 Bregman 方法和增广拉格朗日函数法有速度优势.

考虑问题:

$$\min_{x,y} \{f(x) + h(y) : Ax + By = b\}. \quad (2.26)$$

该问题的增广拉格朗日函数为

$$L(x, y; \lambda) = f(x) + h(y) - \lambda^T(Ax + By - b) + \frac{c}{2} \|Ax + By - b\|_2^2,$$

增广拉格朗日函数法为

$$(x^{k+1}, y^{k+1}) = \arg \min_{x, y} L(x, y, \lambda^k), \quad (2.27a)$$

$$\lambda^{k+1} = \lambda^k - c(Ax^{k+1} + By^{k+1} - b). \quad (2.27b)$$

由于子问题 (2.27a) 需要同时求解 x 和 y , 计算比较复杂. 交替方向增广拉格朗日函数法 (ADM) 将步 (2.27a) 拆为两步, 得到迭代格式:

$$x^{k+1} = \arg \min_x L(x, y^k, \lambda^k), \quad (2.28a)$$

$$y^{k+1} = \arg \min_y L(x^{k+1}, y, \lambda^k), \quad (2.28b)$$

$$\lambda^{k+1} = \lambda^k - \gamma c(Ax^{k+1} + By^{k+1} - b), \quad (2.28c)$$

其中 $\gamma \in (0, (\sqrt{5} + 1)/2)$.

如果其中一个子问题, 比如关于 x 子问题 (2.28a), 仍然没有显式解, 则可以使用非精确 ADM:

$$x^{k+1} = x^k - \tau_k \nabla L(x^k, y^k, \lambda^k), \quad (2.29a)$$

$$y^{k+1} = \arg \min_y L(x^{k+1}, y, \lambda^k), \quad (2.29b)$$

$$\lambda^{k+1} = \lambda^k - \gamma c(Ax^{k+1} + By^{k+1} - b), \quad (2.29c)$$

其中 τ_k 是步长因子.

问题形式 (2.26) 通常出现在变量分裂中. 比如, 考虑问题

$$\min_x p(x) + q(Ax).$$

引入变量 $y = Ax$, 则得到分裂后的问题

$$\min_{x, y} \{p(x) + q(y) : Ax - y = 0\}.$$

该问题目标函数有两个可分项 $p(x)$ 和 $q(y)$. 因此, 交替方向增广拉格朗日函数法可以应用过来.

例 2.3 罚问题 (2.2) 可以写成等价形式

$$\min_x \left\{ \|x\|_1 + \frac{\mu}{2} \|y\|_2^2 : Ax + y = b \right\}. \quad (2.30)$$

将不精确交替方向法应用到 (2.30) 就得到迭代

$$\begin{aligned} y^{k+1} &= \frac{c}{\mu + c} (\lambda^k / c - (Ax^k - b)), \\ x^{k+1} &= \text{shrink}(x^k - \tau_k \nabla_x L(x^k, y^{k+1}; \lambda^k), \tau_k), \\ \lambda^{k+1} &= \lambda^k - \gamma c(Ax^{k+1} + y^{k+1} - b). \end{aligned}$$

例 2.4 对偶问题 (2.8) 的等价形式为

$$\max_y \{b^T y : s = A^T y, \|s\|_\infty \leq 1\}.$$

将精确交替方向法应用到上式就得到迭代:

$$\begin{aligned} y^{k+1} &= (AA^T)^{-1} ((Ax^k - b)/c + As^k), \\ s^{k+1} &= \text{Proj}_{[-1,1]}(A^T y^{k+1} + x^k/c), \\ x^{k+1} &= x^k + \gamma c(A^T y^{k+1} - s^{k+1}). \end{aligned}$$

该算法中需要计算逆矩阵 $(AA^T)^{-1}$. 事实上压缩感知里有很多矩阵满足 AA^T 是单位矩阵或对角矩阵, 因此, 该逆的计算很方便. 另一方面, 如果计算逆矩阵的代价很大, 则可以使用线性化的不精确 ADM. 更详细的推导以及相关的软件包, 请参看文献 [48].

2.6 非凸优化方法

基于非凸优化, 比如 ℓ_q 伪范数

$$\|x\|_q = \left(\sum_i |x_i|^q \right)^{1/q}, \quad 0 < q < 1,$$

和它的变形, 也有不少方法来恢复稀疏向量 [31–33] 和低秩矩阵 [34–36]. 伪范数 $\|x\|_q^q$, $0 < q < 1$, 比 $\|x\|_1$ 范数更接近于 $\|x\|_0$ 范数. 在假设某些 RIP 性质的条件下, 文献 [37] 指出稀疏向量 $x^0 \in \mathbb{R}^N$ 是 $Ax = b$ 的全局 ℓ_q 极小解, 其中 b 的观测数量比 ℓ_1 凸优化所需求的更少. 文献 [14,38,39] 推导了比已知 ℓ_1 优化恢复稀疏向量要求更弱的充分条件.

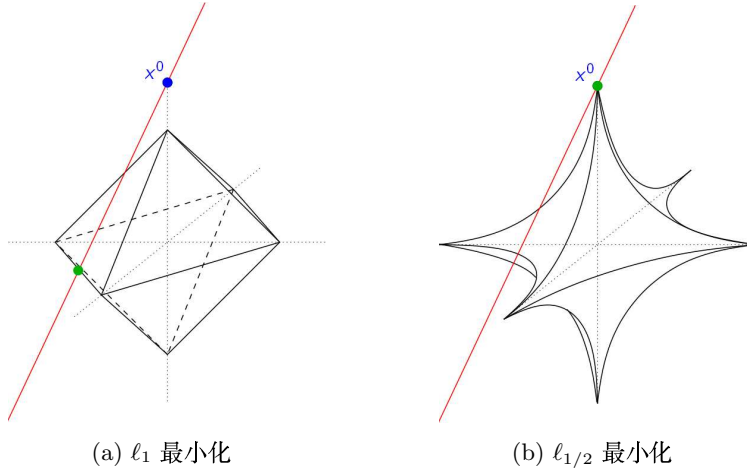


图 3 ℓ_1 和 $\ell_{1/2}$ 优化

图 3 给出了恢复 x^0 时 ℓ_1 优化失败而 $\ell_{1/2}$ 优化成功的例子. 在图 3(a) 中, 直线代表满足 $Ax = b$ 点的集合, 最小的 ℓ_1 球和这条线相交在有两个非零元的下方的圆点. 在图 3(b), 最小的 $\ell_{1/2}$ 球和这条线相交在只有一个非零元的点 x^0 .

然而, ℓ_q 伪范数非凸, 这样的 ℓ_q 优化通常是 NP-难 [40]. 直接极小化 ℓ_q 伪范数极有可能找到很多局部极小点之间的一个. 算法 [31–33] 求解一系列近似子问题. 特别地, 文献 [31] 求解一系列的重新加权 ℓ_1 子问题: 给定初始点 x^k , 该算法使用权重 $w_i := (\varepsilon + |x_i^k|)^{q-1}$ (或者和其它约束条件配合) 极小化 $\sum_i w_i |x_i|$ 和其它目标函数项的组合. 如果令 $x^k = x$, $\varepsilon = 0$,

以及定义 $0/0$ 为 0 , 则得到 $\sum_i w_i |x_i| = \sum_i |x_i|^q = \|x\|_q^q$. 同时, [32-33] 求解重新加权的最小二乘 (也称为重新加权 ℓ_2) 子问题: 即使用 $w_i := (\varepsilon^2 + |x_i^k|^2)^{q/2-1}$ 和 $\sum_i w_i |x_i|^2$ 来近似 $\|x\|_q^q$.

在重新加权的 ℓ_1/ℓ_2 迭代中, 固定的参数 $\varepsilon > 0$ 不仅避免除以 0 的数值问题, 而且导致 $x^\varepsilon = \lim_{k \rightarrow \infty} x^k$ 使得解中有很少的绝对值大于 $O(|\varepsilon|)$ 的元素. 从这个意义上讲, x^ε 通常是一个靠近 x^0 的误差为 $O(|\varepsilon|)$ 近似. 为了从 $b = Ax$ 中恢复稀疏向量 x , 这些算法必须从一个相对较大的值开始, 慢慢地减小 ε .

2.7 贪婪算法

由于稀疏优化的贪婪算法没有一个整体目标函数, 严格来讲它们并不是优化算法. 但他们一步步的构造解或者解的支撑集, 得到的解的确是稀疏解.

一个最基本的贪婪算法称为正交匹配追踪 (Orthogonal Matching Pursuit, OMP)^[41]. 从初始点 x^0 , 支撑集 \mathcal{S}^0 和 $k = 1$ 开始, 依次迭代

$$r^k = b - Ax^{k-1}, \quad (2.31a)$$

$$\mathcal{S}^k = \mathcal{S}^{k-1} \cup \arg \min_{i \notin \mathcal{S}^{k-1}} \{\|\phi_i \alpha - r^k\|_2 : \alpha \in \mathbb{R}\}, \quad (2.31b)$$

$$x^k = \min_x \{\|Ax - b\|_2 : \text{supp}(x) \subseteq \mathcal{S}^k\}, \quad (2.31c)$$

直到 $\|r^k\|_2 \leq \varepsilon$ 满足. 步 (2.31a) 计算观测残量, 步 (2.31b) 添加更能表达残量的 A 的列到支撑集中, 步 (2.31c) 在新的支撑集上更新解的估计. 在 (2.31b) 中, 对每一个 i , 最优解是 $\alpha = \phi_i^T r^k / \|\phi_i\|_2$. 因此, 选择的 i 在所有 $i \in \mathcal{S}^{k-1}$ 中有最小的 $\|(\phi_i^T r^k / \|\phi_i\|_2) \phi_i - r^k\|_2$. 步 (2.31c) 需要求解一个最小二乘问题.

算法 OMP 有很多进一步的发展, 来提高解的质量或求解速度, 比如逐步 OMP^[42], 正则 OMP^[43], 子空间追踪^[44] 和 CoSaOMP^[45]. 最后两个方法通常比其它贪婪算法有更好的数值表现. 由于它们很相似, 我们这里只描述 CoSaOMP. 每次迭代有更多的步来预测和提炼中间解和支撑集. 从初始点 x^0 , 支撑集 \mathcal{S}^0 和 $k = 1$ 开始, CoSaOMP 依次迭代

$$r^k \leftarrow b - Ax^{k-1} \quad (\text{残量}) \quad (2.32a)$$

$$a \leftarrow A^* r^k \quad (\text{相关系数}) \quad (2.32b)$$

$$\mathcal{T} \leftarrow \text{supp}(x^{k-1}) \cup \text{supp}(a_{2s}) \quad (\text{合并非零元支撑集}) \quad (2.32c)$$

$$c \leftarrow \min_x \{\|Ax - b\|_2 : \text{supp}(x) \subseteq \mathcal{T}\} \quad (\text{最小二乘}) \quad (2.32d)$$

$$x^k \leftarrow c_s, \quad (\text{截断}) \quad (2.32e)$$

直到 $\|r^k\|_2 \leq \varepsilon$ 满足, 其中 $a_{2s} = \arg \min \{\|x - a\|_2 : \|x\|_0 \leq 2s\}$ 是 a 的最好 $2s$ 近似, c_s 是 c 的最好 s 近似. 在迭代过程中, 支撑集 $\text{supp}(x^k)$ 的更新始终保证它不超过 s 个元素. 因此, \mathcal{T} 最多不超过 $3s$ 个元素. 与 OMP 不同, CoSaMP 中支撑集是一块更新的.

3 总 结

本文简要介绍了压缩感知和稀疏优化的基本概念, 保证稀疏恢复的一些条件和求解相关模型的计算方法. 由于篇幅限制, 如最小列相关数 (spark), 非相关性 (incoherence), 球形割性质 (spherical section property), 不基于受限正交条件 (“RIPless” conditions) 等恢复条件, 如 (块) 坐标下降法, 迭代硬阈值算法 (iterative hard thresholding), 各类组合算法, 各类特殊形式矩阵在压缩感知中的性质, Barzilai-Borwein 梯度法, Nesterov 等算法, 低秩矩阵恢复, 鲁棒 (robust) 主成分分析, 稀疏逆协方差矩阵估计, 联合稀疏, 分散式稀疏优化等问题都没有涉及. 感兴趣的读者可以参考莱斯大学网站 <http://dsp.rice.edu/cs> 和博客 <http://nuit-blanche.blogspot.com>.

需要指出的是, 国内相关研究也取得丰富的成果, 如北京交通大学的修乃华教授和孔令臣教授、浙江大学的李松教授和莫群教授、中科院数学与系统科学研究院的许志强教授等在压缩感知的可恢复性方面的理论; 西安交通大学的徐宗本院士、香港理工大学的陈小君教授、上海交通大学的葛冬冬教授 (与美国斯坦福大学叶荫予教授合作) 在 $\ell_p (p \in (0, 1))$ 模优化的理论和算法; 南京大学的何炳生教授、香港浸会大学的袁晓明教授、南京大学杨俊锋教授、香港中文大学的马士谦教授等在稀疏优化的算法方面的工作; 国防科技大学的朱炬波教授、成礼智教授、张慧博士、黄石生博士在稀疏优化的理论及算法中的贡献; 中国科技大学凌青教授、西安电子科技大学的石光明教授和刘丹华教授、哈尔滨工业大学的马坚伟教授等在各自应用领域 (如电子、遥感、自动化、网络、石油勘探) 中的应用. 由于篇幅和撰写时间的限制, 这里没有包括国外和台湾地区学者的工作, 国内学者工作的收罗也难免有所疏漏, 对此作者深表歉意. 如果您对本文有任何的意见或建议, 请和作者联系, 我们将致以最衷心的感谢. 我们也将博客 <http://blog.sina.com.cn/appopt> 中即时更新这篇综述的勘正版本. 此外, 我们注意到西安电子科技大学石光明教授和刘丹华教授等^[46]及中国科学院数学与系统科学研究院的许志强教授^[47]最近分别撰写了关于压缩感知的综述性文章, 这两篇文章更强调压缩感知的应用及理论性部分, 与我们的着眼点正好互补, 感兴趣的读者可以进一步阅读以对压缩感知有一个更为全面的了解.

参 考 文 献

- [1] Candès E, Tao T. Near optimal signal recovery from random projections: universal encoding strategies [J]. *IEEE Transactions on Information Theory*, 2006, **52**(1): 5406-5425.
- [2] Donoho D. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, **52**, 1289-1306.
- [3] Natarajan B K. Sparse approximate solutions to linear systems [J]. *SIAM Journal on Computing*, 1995, **24**: 227-234.
- [4] Cohen A, Dahmen W, DeVore R A. Compressed sensing and best k -term approximation [J]. *Journal of the American Mathematical Society*, 2009, **22**: 211-231.
- [5] Rudelson M, Vershynin R. Geometric approach to error correcting codes and reconstruction of signals [J]. *International Mathematical Research Notices*, 2005, **64**: 4019-4041.
- [6] Compressive Sensing Resources. <http://dsp.rice.edu/cs>.
- [7] Donoho D, Huo X. Uncertainty principles and ideal atomic decompositions [J]. *IEEE Transactions on Information Theory*, 2001, **47**: 2845-2862.

- [8] Gribonval R, Nielsen M. Sparse representations in unions of bases [J]. *IEEE Transactions on Information Theory*, 2003, **49**(12): 3320-3325.
- [9] Zhang Y. A simple proof for recoverability of ℓ_1 -minimization: go over or under? [R]. *Rice University CAAM Technical Report TR05-09*, 2005.
- [10] Kashin B S. Diameters of certain finite-dimensional sets in classes of smooth functions [J]. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 1977, **41**(2): 334-351.
- [11] Garnaev A, Gluskin E D. The widths of a Euclidean ball [J]. *Dokl. Akad. Nauk SSSR*, 1984, **277**(5): 1048-1052.
- [12] Milman V D, Schechtman G. Asymptotic theory of finite dimensional normed spaces, with an appendix by M. Gromov [M]// *Lecture Notes in Mathematics*, Berlin: Springer, 2001.
- [13] Candès E, Tao T. Decoding by linear programming [J]. *IEEE Transactions on Information Theory*, 2005, **51**: 4203-4215.
- [14] Foucart S, Lai M J. Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$ [J]. *Applied and Computational Harmonic Analysis*, 2009, **26**(3): 395-407.
- [15] Foucart S. A note on guaranteed sparse recovery via ℓ_1 -minimization [J]. *Applied and Computational Harmonic Analysis*, 2010, **29**(1): 97-103.
- [16] Cai T T, Wang L, Xu G. Shifting inequality and recovery of sparse signals [J]. *IEEE Transactions on Signal Processing*, 2010, **58**(3): 1300-1308.
- [17] Mo Q, Li S. New bounds on the restricted isometry constant δ_{2k} [J]. *Applied and Computational Harmonic Analysis*, 2011, **31**(3): 460-468.
- [18] Zhang Y. Theory of compressive sensing via ℓ_1 -minimization: a non-RIP analysis and extensions [R]. Rice University, CAAM Technical Report TR08-11, 2008.
- [19] Vavasis S A. Derivation of compressive sensing theorems from the spherical section property [EB/OL] www.student.math.uwaterloo.ca/~co769/simplif.pdf. 2009.
- [20] Darbon J, Sigelle M. Image restoration with discrete constrained total variation, Part I: fast and exact optimization [J]. *Journal of Mathematical Imaging and Vision*, 2006, **26**(3): 261-276.
- [21] Goldfarb D, Yin W. Parametric maximum flow algorithms for fast total variation minimization [J]. *SIAM Journal on Scientific Computing*, 2009, **31**(5): 3712-3743.
- [22] Chan T F, Zhou H M, Chan R H. Continuation method for total variation denoising problems [J]. *Advanced Signal Processing Algorithms*, 1995, **2563**(1): 314-325.
- [23] Chambolle A. An algorithm for total variation minimization and applications [J]. *Journal of Mathematical Imaging and Vision*, 2004, **20**: 89-97.
- [24] Wohlberg B, Rodriguez P. An Iteratively Reweighted Norm Algorithm for Minimization of Total Variation Functionals [J]. *Signal Processing Letters, IEEE*, 2007, **14**(12): 948-951.
- [25] Figueiredo M, Nowak R. An EM algorithm for wavelet-based image restoration [J]. *IEEE Transactions on Image Processing*, 2003, **12**: 906-916.
- [26] De Mol C, Defrise M. A Note on Wavelet-based Inversion Algorithms [J]. *Contemporary Mathematics*, 2002, **313**: 85-96.
- [27] Bect J, Blanc-Feraud L, Aubert G, et al. A ℓ_1 -unified variational framework for image restoration [J]. *European Conference on Computer Vision, Prague, Lecture Notes in Computer Sciences* 3024, 2004: 1-13.
- [28] Hale E T, Yin W, Zhang Y. Fixed-Point Continuation for ℓ_1 -Minimization: Methodology and Convergence [J]. *SIAM Journal on Optimization*, 2008, **19**(3): 1107-1130.
- [29] Ma S Q, Goldfarb D, Chen L F. Fixed point and Bregman iterative methods for matrix rank minimization [J]. *Math. Program.*, 2011, **128**(1-2): 321-353.
- [30] Bregman L. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming [J]. *USSR Computational Mathematics and Mathematical Physics*, 1967, **7**: 200-217.

- [31] Candes E, Wakin M B, Boyd S P. Enhancing sparsity by reweighted ℓ_1 minimization [J]. *Journal of Fourier Analysis and Applications*, 2008, **14**(5): 877-905.
- [32] Chartrand R, Yin W. Iteratively Reweighted Algorithms for Compressive Sensing [C]// *International Conference on Acoustics, Speech, and Signal Processing*, 2008: 3869-3872.
- [33] Daubechies I, DeVore R, Fornasier M, et al. Iteratively reweighted least squares minimization for sparse recovery [J]. *Communications on Pure and Applied Mathematics*, 2010, **63**(1): 1-38.
- [34] Mohan K, Fazel M. Iterative reweighted least squares for matrix rank minimization [C]// *48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010: 653-661.
- [35] Fornasier M, Rauhut H, Ward R. Low-rank Matrix Recovery via Iteratively Reweighted Least Squares Minimization [J]. *SIAM Journal on Optimization*, 2011, **21**: 1614-1640.
- [36] Lai M J, Xu Y, Yin W. On Unconstrained Nonconvex Minimization for Sparse Vector and Low-Rank Matrix Recovery [J]. *Submitted to SIAM Journal on Numerical Analysis*, 2011.
- [37] Chartrand R, Staneva V. Restricted isometry properties and nonconvex compressive sensing [J]. *Inverse Problems*, 2008, **24**: 1-14.
- [38] Lai M J, Liu L Y. A New Estimate of Restricted Isometry Constants for Sparse Solutions [J]. *Applied and Computational Harmonic Analysis*, 2011, **30**: 402-406.
- [39] Sun Q. Recovery of sparsest signals via ℓ_q -minimization [J]. *Applied and Computational Harmonic Analysis*, 2012, **32**(3): 329-341.
- [40] Ge D, Jiang X, Ye Y. A note on complexity of L_p minimization [J]. *Mathematical Programming*, 2011, **129**: 285-299.
- [41] Tropp J, Gilbert A. Signal recovery from partial information via orthogonal matching pursuit [J]. *IEEE Transactions on Information Theory*, 2007, **53**(12): 4655-4666.
- [42] Donoho D, Tsaig Y, Drori I, et al. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit [R]. *Stanford Technical Report*, 2006.
- [43] Needell D, Vershynin R. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit [J]. *Selected Topics in Signal Processing, IEEE Journal of*, 2010, **4**(2): 310-316.
- [44] Dai W, Milenkovic O. Subspace pursuit for compressive sensing reconstruction [J]. *IEEE Transactions on Information Theory*, 2009, **55**(5): 2230-2249.
- [45] Needell D, Tropp J A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples [J]. *Applied and Computational Harmonic Analysis*, 2008, **26**(3): 301-321.
- [46] 石光明, 刘丹华, 高大化, 等. 电子学报压缩感知理论及其研究进展 [J]. *电子学报*, 2009, **37**(5): 1070-1081.
- [47] 许志强. 压缩感 [J]. *中国科学*, 已接收.
- [48] Yang J F, Zhang Y. Alternating direction algorithms for ℓ_1 -problems in compressive sensing [J]. *SIAM Journal on Scientific Computing*, 2011, **33**(1): 250-278.