



西安交通大学
XIAN JIAOTONG UNIVERSITY

稀疏性与 $L_{1/2}$ 正则化

徐宗本

西安交通大学信息与系统科学研究所

2009.10.28



西安交通大学

XIAN JIAOTONG UNIVERSITY

纲要

- 稀疏性: 问题及应用
- $L_{1/2}$ 正则化理论
- 结论



纲要

● 稀疏性: 问题及应用

- 什么是稀疏性问题?
- 如何解决稀疏性问题?
- 挑战与问题

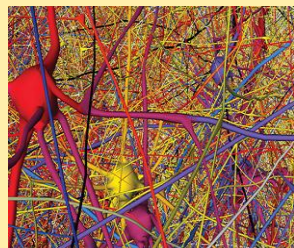
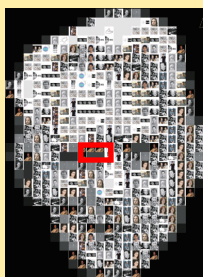
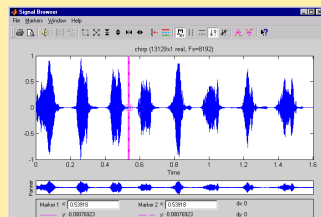
● $L_{1/2}$ 正则化理论

● 结论

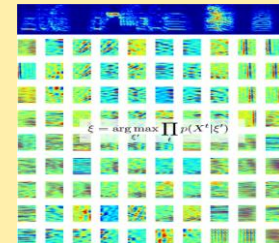
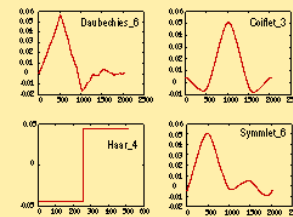
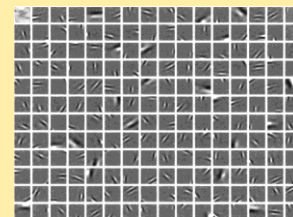


什么是稀疏性？

稀疏性(Sparsity): 信息表示的普遍属性。指信息的表示常常由大量疑似因素中的少数因素所决定的现象。一个问题的稀疏性依赖于表示方式与度量手段。



$$y = f(x_1, x_2, \dots, x_n)$$



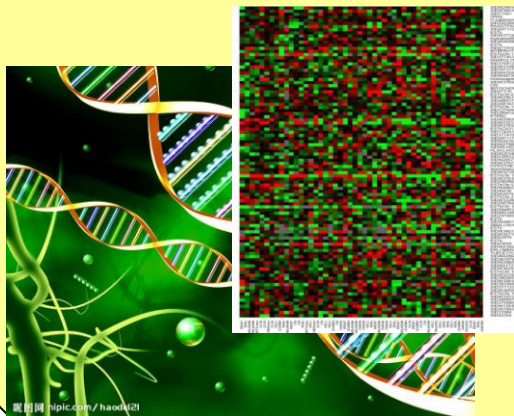
信息观测

信息表示



什么是稀疏性 (变量选择问题)

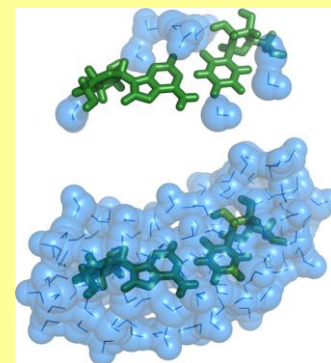
变量选择: 在线性表示的意义上下, 从具有大量疑似因素 (变量) 中决定决策变量的问题。



$$y = \sum_{i=1}^N x_i^T \beta_i$$



- 中药配方问题
- 基因选择问题
- 证券组合问题
- 病因分析问题
-



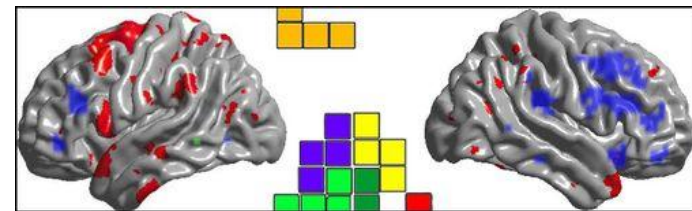


什么是稀疏性 (视觉编码问题)

视觉编码: 根据神经生物学家们的发现, 视觉神经元对视觉输入的相应是稀疏的。初级视皮层及V1区的大部分神经元对视觉信息提供线性特征表示(线性特征检测器), 而V1区的其他部分及其它区域神经元对视觉信息提供非线性特征整合表示。



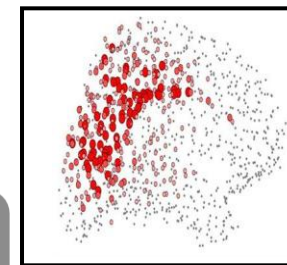
视觉细胞群对视觉输入
响应具有稀疏性



外界信息

(Olshausen & Field, *Nature*, 1996)

$$\min_{a, \phi} E = \sum_{xy} [I(x, y) - \sum_i a_i \phi_i(x, y)]^2 + \lambda \sum_i S\left(\frac{a_i}{\sigma}\right)$$



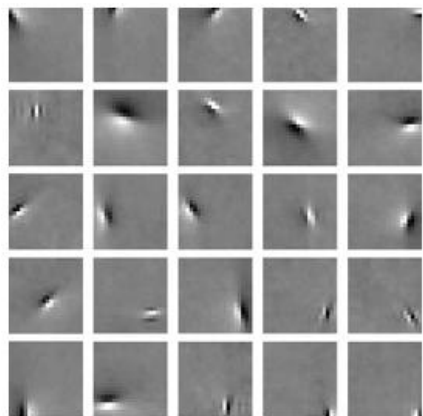
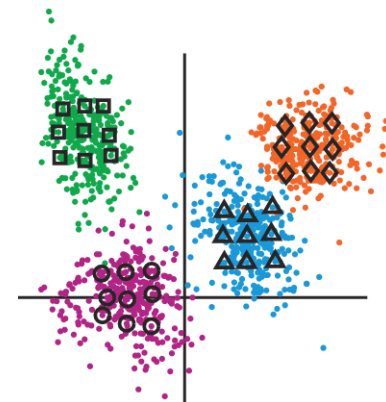
视觉皮层



什么是稀疏性 (视觉编码问题)

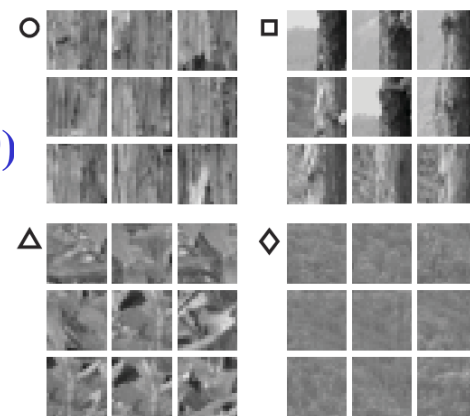


复杂细胞的特征整合功能，提供对视觉信息的稀疏非线性表示



(Karklin & Lewicki, *Nature*, 2009)

$$\min_{y,w,b} E(y,w,b)$$

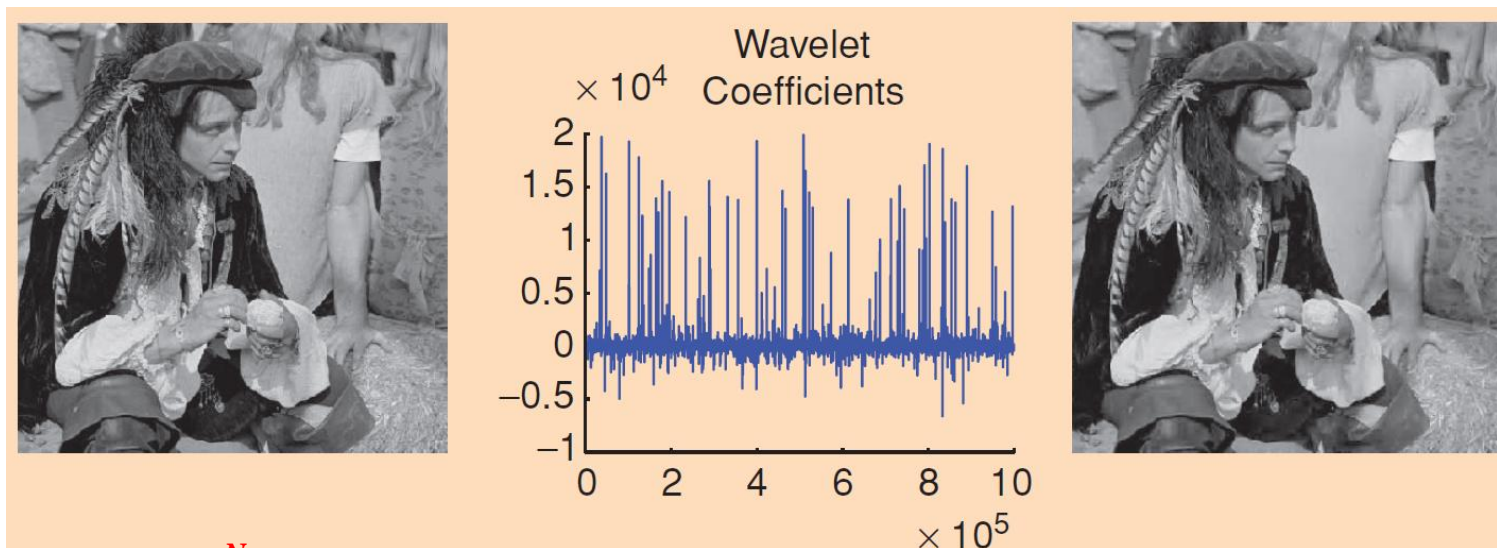


$$E(y,w,b) = \frac{1}{2} \sum_{jk} y_j w_{jk} + \frac{1}{2} \mathbf{x}^T (\exp(-\sum_{jk} y_j w_{jk} \mathbf{b}_k \mathbf{b}_k^T)) \mathbf{x} + \lambda \sum_j |y_j|$$



什么是稀疏性 (图像表达问题)

图像表达: 自然图像在给定的一组基底 (如小波基) 下常呈现出稀疏性 (即表达式中的诸多系数为零或近似为零)。 \longrightarrow 压缩编码



$$\mathbf{x} = \sum_{i=1}^N s_i \psi_i$$

$$\mathbf{x} = \sum_{i \in V} s_i \psi_i, \mathbf{V} = \{i : |s_i| > \epsilon\}, |V| \ll N$$

自然图像在小波基下的稀疏表达

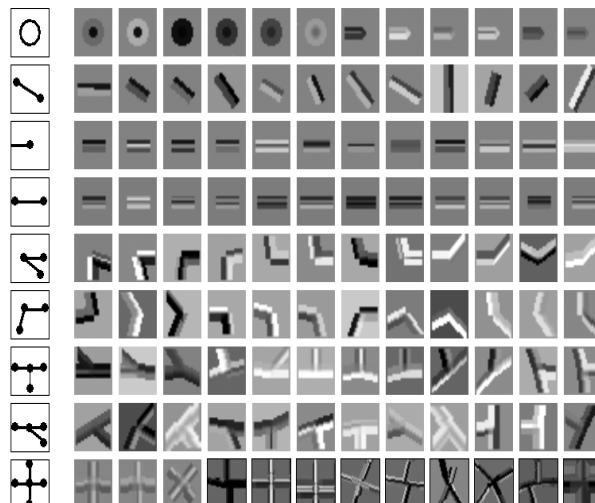


什么是稀疏性 (图像表达问题)

图像结构信息可以通过一组过完全图像基元稀疏表达。



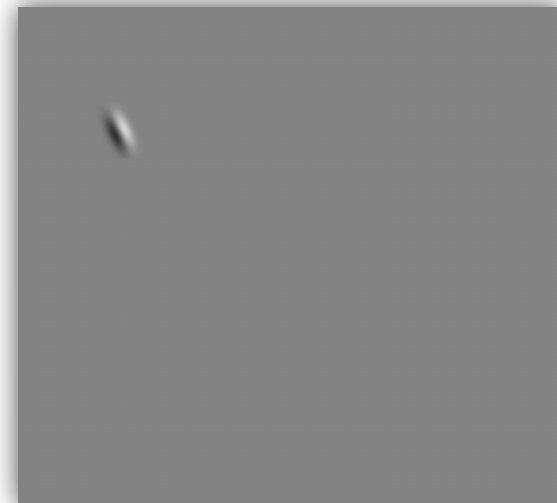
原图像



(a)

(b)

图像结构信息的基元



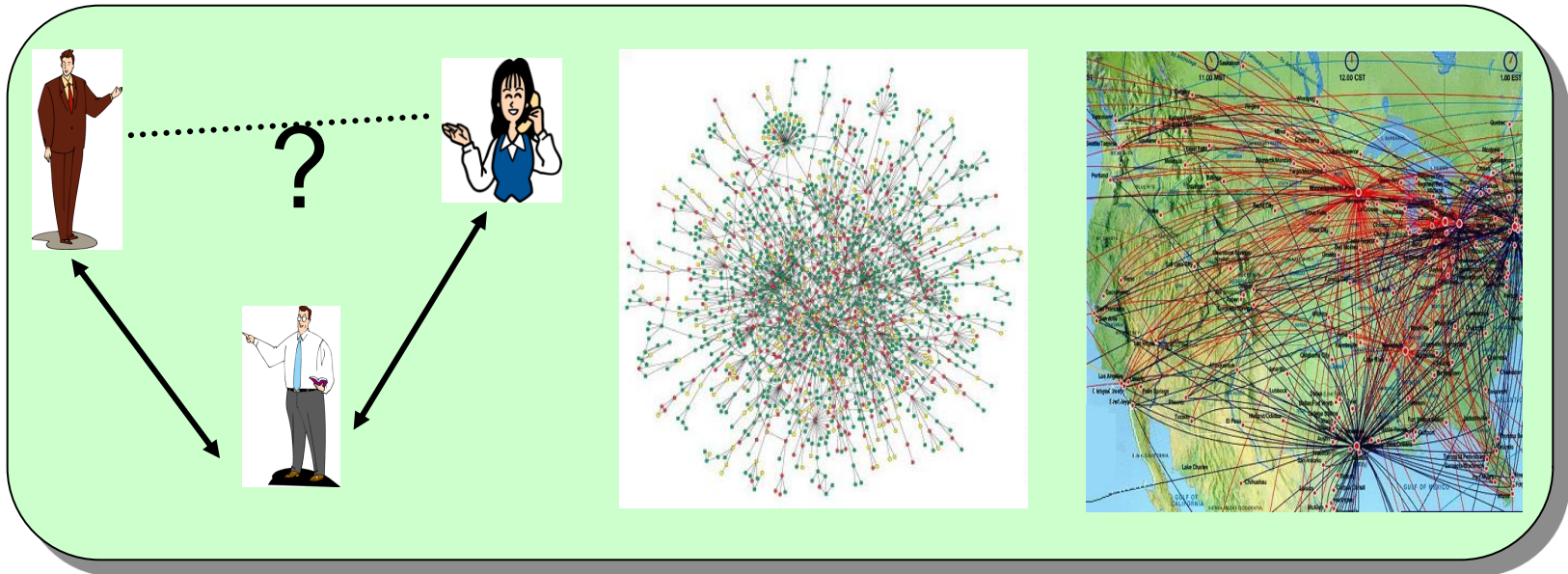
重构图像

自然图像在结构基元下的稀疏表达



什么是稀疏性 (复杂网络问题)

复杂网络问题: 复杂网络结构常常可通过稀疏连接来充分表达完整结构 (例如: 六度空间理论, 蛋白质网络结构, 地理空间结构网络)。



六度空间理论, **Psychology Today**, 1967

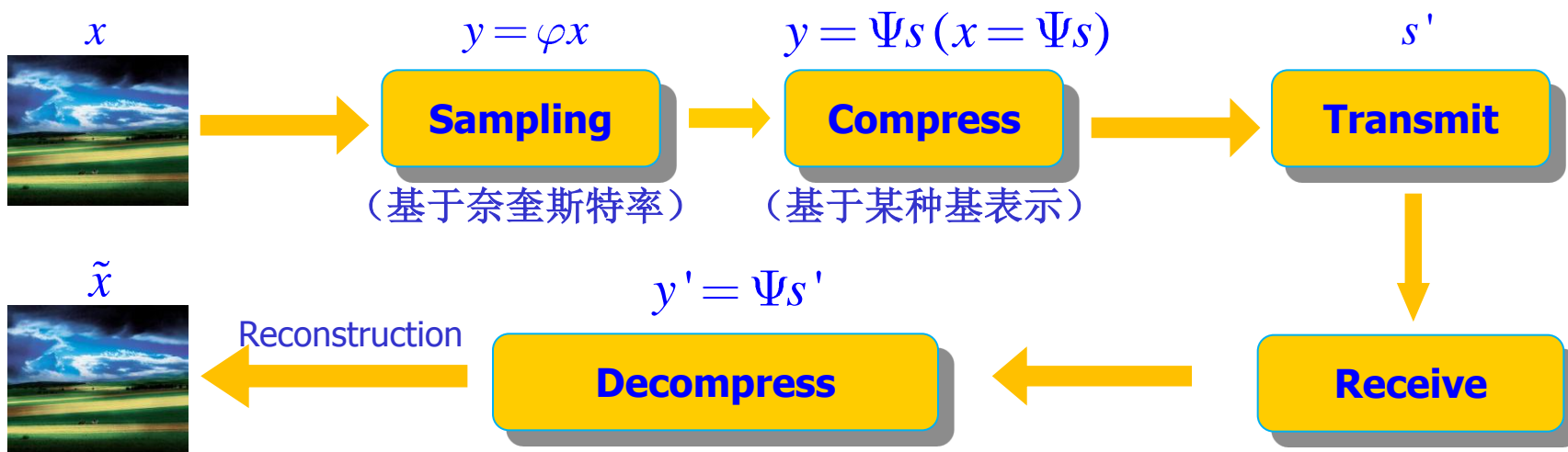
蛋白质网络结构. **Nature** 411, (2001).

空中航线 **Université Paris-Sud, France.**



什么是稀疏性 (压缩传感问题)

传统信号处理过程: 根据香农采样定理, 要完全重建原始信号, 必须要以奈奎斯特率采样信息。它是现今常见信息获取与压缩(如JPEG2000)的基础, 其主要利用了数据的带宽先验。

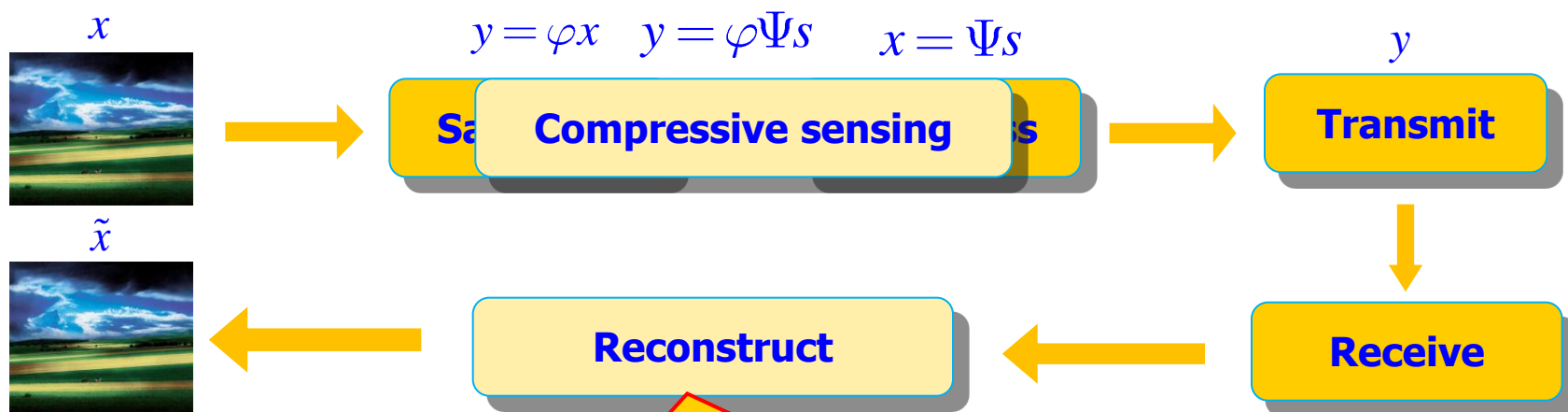


基于奈奎斯特率的传感



什么是稀疏性 (压缩传感问题)

压缩传感 (Compressive sensing): 一种同时传感与压缩的框架。利用信号表示的稀疏性, 在完全重建的前提下, 通过设计巧妙的采样方式 (如以随机线性采样) 来显著减少采样数 (突破奈奎斯特率限制), 以实现信息获取方式的突破。



$\tilde{x} = \Psi s$, 其中 s 为优化问题 $\min \|s\|_0, s.t. y = \varphi \Psi s$ 的解



如何解决稀疏性问题 (L_0 框架)

稀疏性问题的数学表述: L_0 问题, 呈现三种相互等价的形式:

压缩传感形式

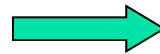
$$\min_x \|x\|_0$$

$$s.t. Ax = y$$

变量选择形式

$$\min_x \|Ax - y\|_2$$

$$s.t. \|x\|_0 \leq t$$



$$\min_x \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_0 \right\}$$

正则化形式

- L_0 问题为NP组合难问题, 对较大规模数据无法直接求解;
- 存在两种直接求解 L_0 问题的算法
 - 贪婪算法: **OMP (1993, 2007), Chaining Pursuit (2006), ROMP(2007), CoSaMp (2008)**
 - 门限算法: **Hard (2007)**
- 存在的问题:
 - 贪婪算法时间代价过高, 无法保证收敛到全局最优;
 - 门限算法时间代价低, 但对数据噪声十分敏感, 解不具有连续性; 不能保证收敛到全局极小.



如何求解稀疏性问题 (L_1 框架)

稀疏性问题的求解: 为使求解稀疏性问题变得可行, 当前一般性的作法是将 L_0 问题放松到 L_1 问题, 从而将一个组合优化问题放松到一个凸优化问题来求解。

L_1 问题:
$$\min_x \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_1 \right\}$$

选择 L_1 框架

- 距离 L_0 最近的凸
- 具有解析解
- 能获得较为合理的稀疏解
- 有快速计算方法
- 在一定条件下通过解 L_1 能获得 L_0 问题的解

Candes 等(Candes, Romberg, Tao, 2006)严格证明, 当原信号是 k 稀疏时, 仅使用 $O(k \log(N/k))$ 采样信息, 原始信号即可通过求解 L_1 问题获得完全重建。

- Soft (2004),
- Adaptive Lasso (2006)

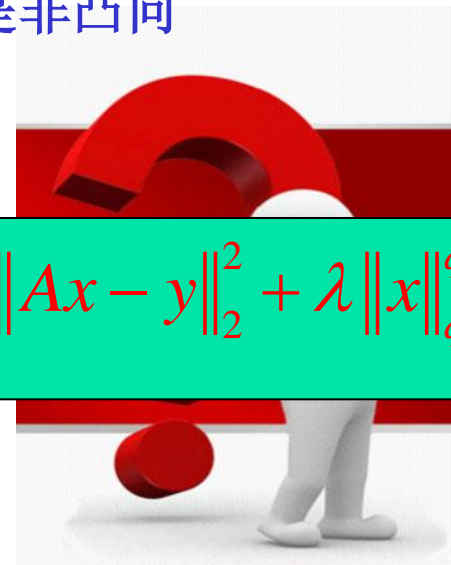
... ..



挑战与问题

- 只有在很严格的条件下才有“ L_1 问题与 L_0 问题求解结果的等价性”；
- L_1 求解框架仍不能保证获得满意的稀疏解，它往往与真实稀疏解(L_0 的解)的差距甚大；
- 一个自然的改进方法是使用 L_q 框架($0 < q < 1$)，但 L_q 是非凸问题，理论分析与求解都非常困难；
- 如何从 $0 < q < 1$ 中选择一个最优的 q ？

$$L_q \text{ 问题: } \min_x \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_q^q \right\}$$





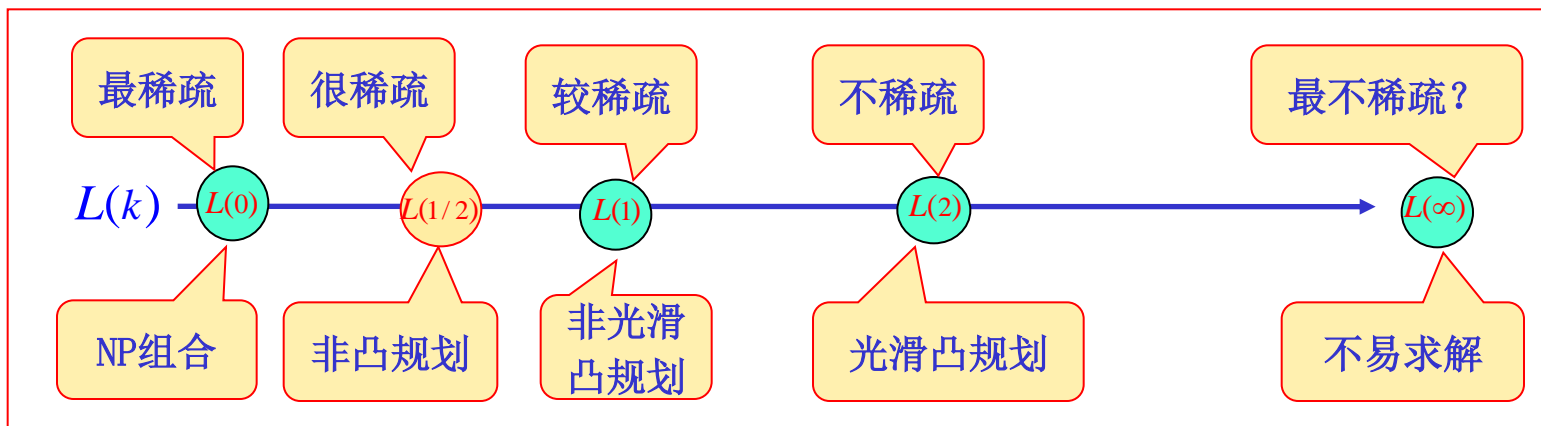
纲要

- 稀疏性: 问题及应用
- $L_{1/2}$ 正则化理论
 - 为什么有 $L_{1/2}$?
 - $L_{1/2}$ 好吗?
 - $L_{1/2}$ 能快速算吗?
 - $L_{1/2}$ 应用效果如何?
- 结论



为什么会有 $L_{1/2}$ (空间理论启示)

Banach 空间启示: 在 $1/p+1/q=1$ 对偶框架下, L_p ($1 \leq p \leq \infty$) 形成一个以 L_2 为中心(自对偶)的局部凸空间体系。在 $p+q=1$ 的对偶框架下, L_q ($0 \leq q \leq 1$) 会形成一个以 $L_{1/2}$ 为中心的非局部凸空间体系。 $L_{1/2}$ 应该有特殊性! [$L_{1/2}$ 正则子, 中国科学, 2009]



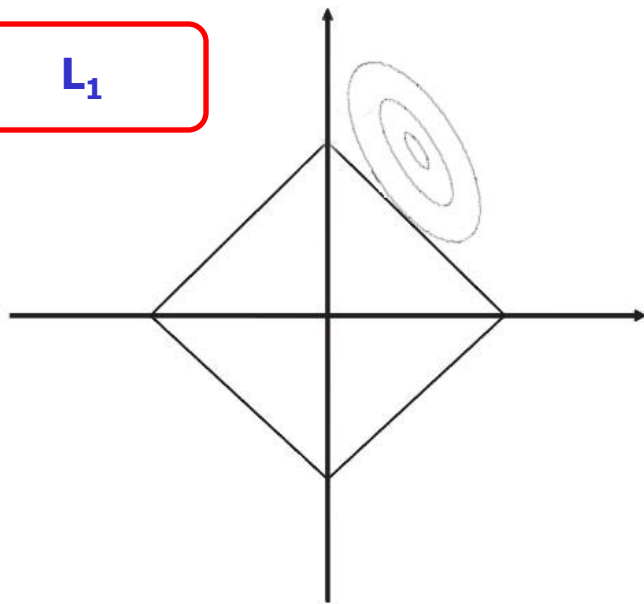
$L_{1/2}$ 问题:
$$\min_x \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_{1/2}^{1/2} \right\}$$



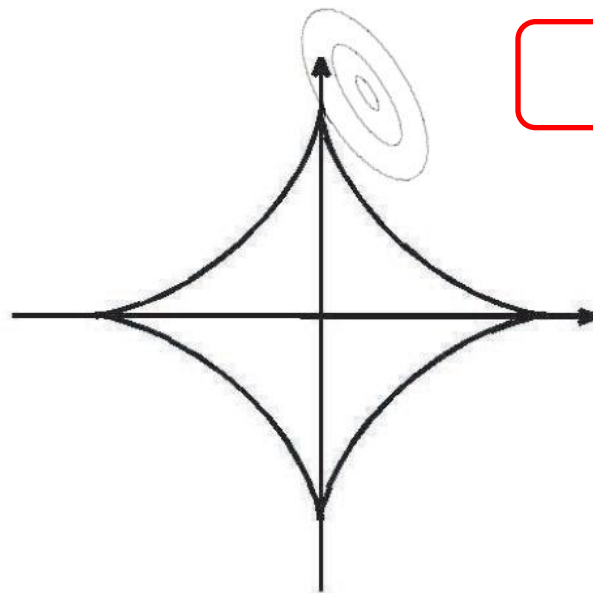
为什么会有 $L_{1/2}$ (几何启示)

约束优化问题解的启示：从单位球的形状可看出， $L_{1/2}$ 正则子所产生的解会比 L_1 正则子产生的解更具有稀疏性。

L_1



$L_{1/2}$





$L_{1/2}$ 会更好吗？

- 具有比 L_1 的更好的理论性质
 - 比 L_1 更好的Oracle统计性质
 - 比 L_1 更宽泛的可恢复性条件
- 解的可解析表达性 (新型 $L_{1/2}$ 算法的基础)
- 对 L_q , $0 \leq q \leq 1$, 的代表性
- 有快速的求解算法



$L_{1/2}$ 会更好吗 (统计性质)

- 一个好的变量选择模型应具有良好的Oracle统计性质，即有**稀疏性、无偏性、渐近正态性且满足Oracle不等式** (Fan, JASA, 2001; Candès, et al., AS, 2007)， L_1 具有无偏性与渐近正态性，但并不具有稀疏性与Oracle不等式 (Zou, JASA, 2006)；
- 我们证明 $L_{1/2}$ 具备这些良好统计性质。

➤ **稀疏性**：假设有真实模型 $\beta^* = (\beta_1^*, \beta_2^*)^T, \beta_2^* = 0$ ， $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$ 为估计模型，这里n为样本个数，则 $\lim_{n \rightarrow \infty} \Pr(\hat{\beta}_2 = 0) = 1$ 。该性质说明：当真实模型是稀疏情形， $L_{1/2}$ 能找到这个稀疏解。

➤ **无偏性**： $\lim_{n \rightarrow \infty} E(\hat{\beta}_n) = \beta^*$

➤ **渐近正态性**： $\sqrt{n}(\hat{\beta}_n - \beta^*) \rightarrow_d N(0, \Sigma)$

➤ **Oracle不等式**：在一定条件下，

$$\|\hat{\beta} - \beta^*\|_2 \leq O(\sigma + r_k^{(2)}(\beta^*) + r_k^{(1)}(\beta^*)(l^{(-1/2)}))$$



$L_{1/2}$ 会更好吗(可恢复性)

研究框架及结果: 在RIP及Coherence框架下, 一系列作者证明了一定条件下的“ L_1 与 L_0 问题的等价性”

- RIP框架: $(1-\delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\|x\|_2^2$, 其中 $\|x\|_0 \leq k$
- Coherence框架: $\mu(A) = \max_{i \neq j} |\langle a_i, a_j \rangle|$, 其中 a_i, a_j 为 l_2 规范化后 A 的列向量

已知最好结果

$$RIP: \delta_{2k} < \sqrt{2} - 1 \approx 0.414 \quad (L_1, \text{Candes}, 2008)$$

$$RIP: \delta_{2k} < 0.4531 \quad (L_q, \text{Foucart, et.al}, 2009)$$

$$COH: \mu(A) < \frac{1}{2k-1} \quad (L_1, \text{Tony, et.al}, 2009)$$

我们的结果

$$RIP: \delta_{2k} < 0.5 \implies \|x - x^*\|_2 = O(s^{1/2-1/q} \|x - x_s\|_q) \quad (\downarrow q)$$

$$COH: \mu(A) < \left(\frac{1}{2k-1}\right)^{1/q} \implies n \geq \frac{N}{(N-1) 2k-1^{-2/q} + 1} \quad (\downarrow q) \quad (L_q)$$



$L_{1/2}$ 会更好吗 (解的可解析性)

$L_{1/2}$ 问题的解:

$$x = \arg \min_x \|Ax - b\|_2^2 + \lambda \|x\|_{1/2}^{1/2} \longrightarrow A^T(Ax - b) + \lambda \partial(\|x\|_{1/2}^{1/2}) = 0$$

$$x = I(\cdot) + \lambda \mu \partial(\|\cdot\|_{1/2}^{1/2})^{-1} B_\mu(x) \longleftarrow x + \lambda \mu \partial(\|x\|_{1/2}^{1/2}) = x - \mu A^T(Ax - b) := B_\mu(x)$$

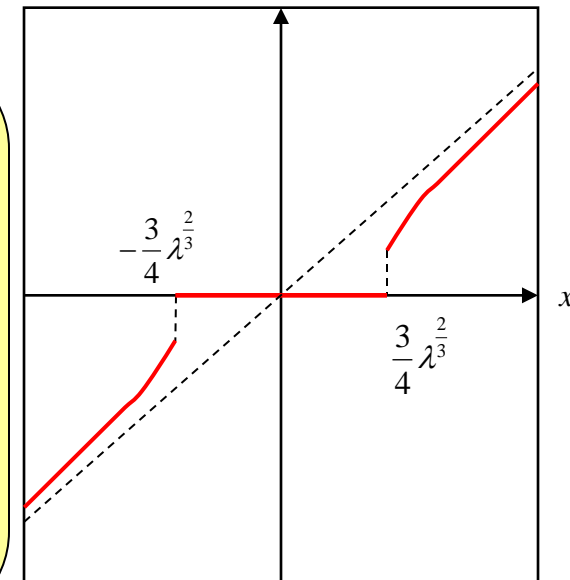
记 $H_{1/2,\lambda} = I(\cdot) + \lambda \partial(\|\cdot\|_{1/2}^{1/2})^{-1}$ 。

定理: $H_{1/2,\lambda}$ 为对角非线性阈值算子, 且可表达为

$$H_{1/2,\lambda} = h_{1/2,\lambda}(x_1), h_{1/2,\lambda}(x_2), \dots, h_{1/2,\lambda}(x_n), \text{ 其中}$$

$$h_{1/2,\lambda}(x) = \begin{cases} \frac{2}{3}x(1 + \cos(\frac{2\pi}{3} - \frac{2\phi}{3})), & |x| > \frac{3}{4}\lambda^{2/3}, \\ 0, & \text{otherwise} \end{cases}$$

$$\phi = \arccos(\frac{\lambda}{8}(\frac{|x_i|}{3})^{-\frac{3}{2}}). \text{ (核心: 三次方程卡当公式)}$$



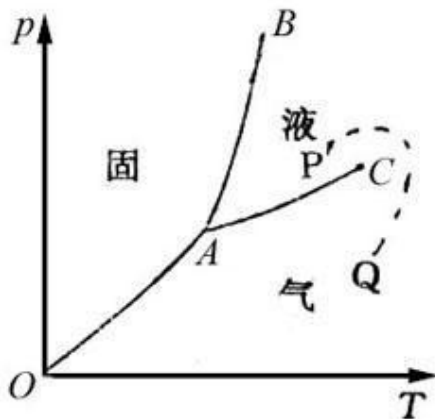
$h_{1/2,\lambda}(x)$



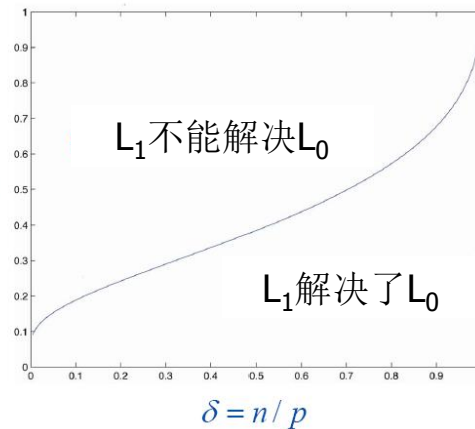
$L_{1/2}$ 会更好吗 (对 L_q 的代表性)

相变图: 相变是一个物理学概念。Donoho等(2006)借用相变边界曲线来精细刻画 L_1 与 L_0 的等价性条件。我们利用 L_q 在若干应用问题上的相变图来说明 $L_{1/2}$ 对于 L_q , $0 < q \leq 1$, 的代表性。

物理相变图



$$\rho = k/n$$



信号恢复性相变图

- 相变空间坐标 (ρ, δ)

- ◇ $\rho = k/n$: 稀疏程度
- ◇ $\delta = n/p$: 欠定程度

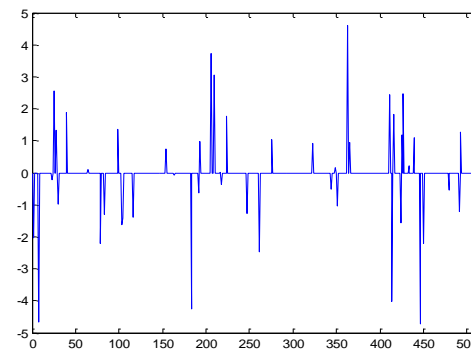
- 相变空间中点 (ρ, δ) 处的像素值对应于相应情形下 L_1 解与 L_0 解的相对误差 (以一定的误差接受准则接受)。



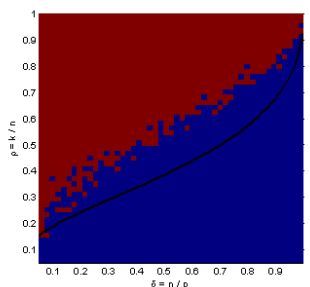
$L_{1/2}$ 会更好吗 (对 L_q 的代表性)

L_q 对信号恢复问题的相变图:

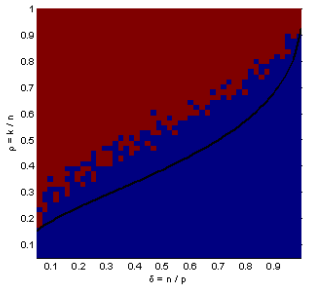
$$Ax = y$$



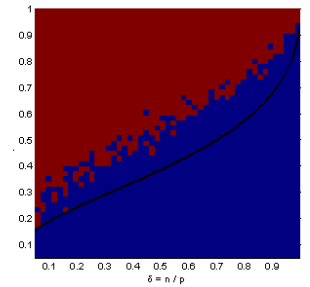
$$\min_x \|x\|_0, s.t. Ax = y$$



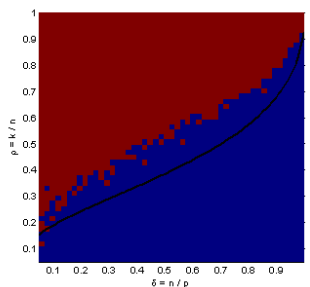
q=0.1



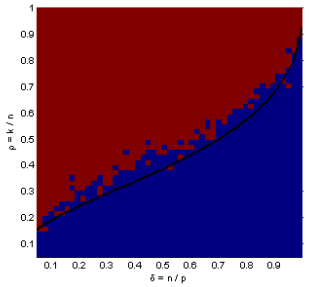
q=0.3



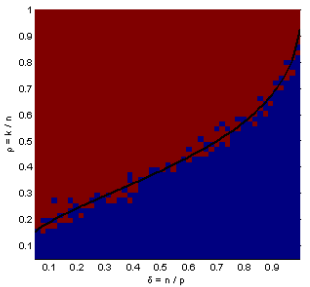
q=0.5



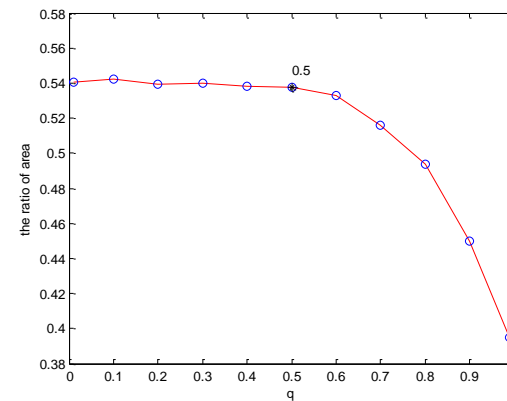
q=0.7



q=0.9



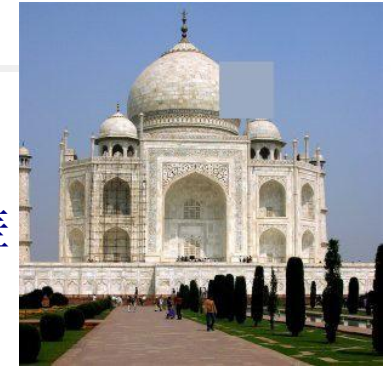
q=1.0



L_q 对信号的可恢复比例



$L_{1/2}$ 会更好吗 (代表性)



$$\min(\|\varepsilon\|_q^q)$$

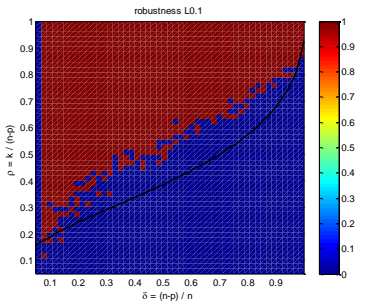
$$s.t., \tilde{y} = Fy = F\varepsilon$$

L_q 对误差校正问题的相变图:

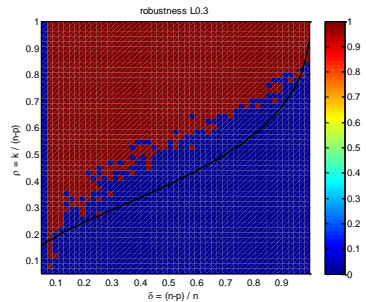
$$\min(\|Ax - y\|_q^q)$$

$$s.t., y = Ax + \varepsilon, \varepsilon \text{为误差}$$

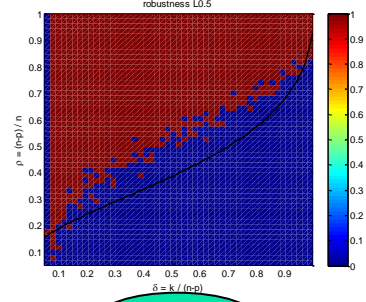
引进 $FA = 0$ \rightarrow



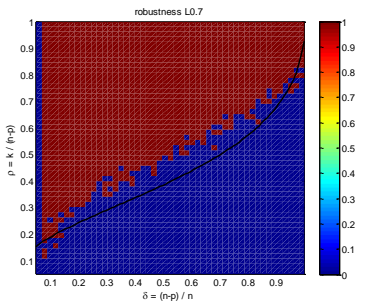
q=0.1



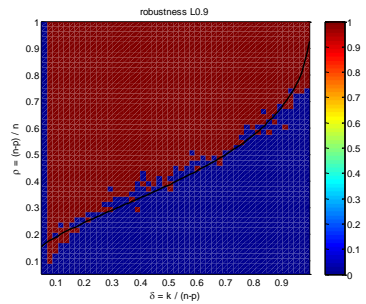
q=0.3



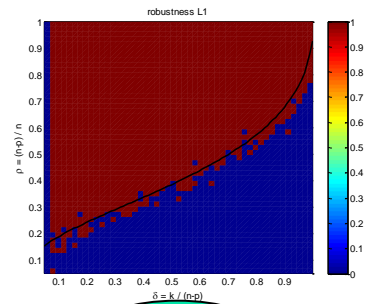
q=0.5



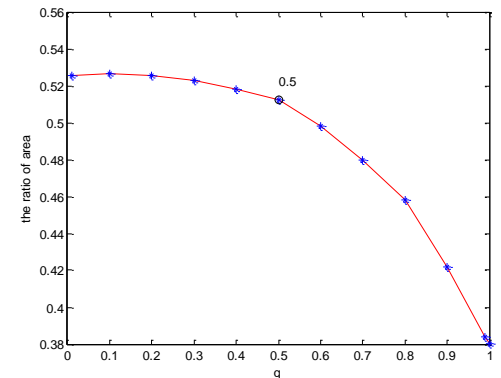
q=0.7



q=0.9



q=1.0



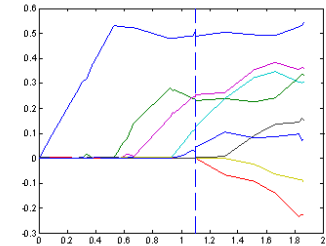
L_q 对信号的可恢复比例



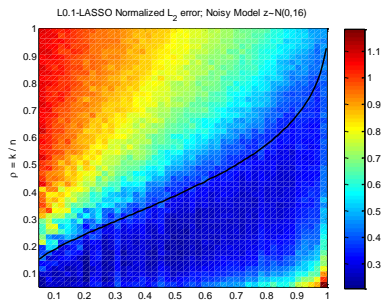
$L_{1/2}$ 会更好吗 (代表性)

L_q 对变量选择问题的相变图:

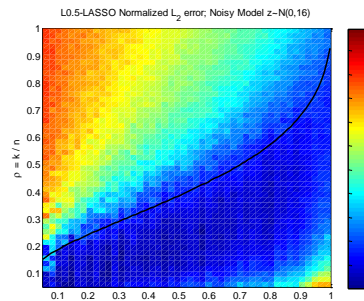
$$y = x^T \beta + \varepsilon, \beta \text{ 稀疏}$$



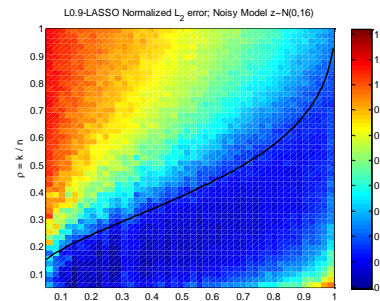
$$\min(\|y - x^T \beta\|_2^2 + \lambda \|\beta\|_q^q)$$



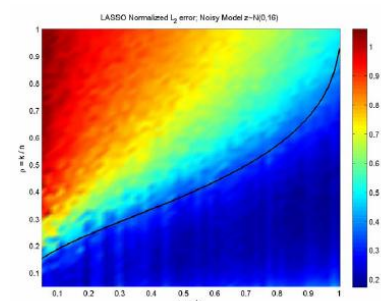
$q=0.1$



$q=0.5$



$q=0.9$



$q=1.0$

结论:

- ◆ L_q 解的稀疏性在 $q \in [0, 1)$ 时明显优于 L_1 解;
- ◆ L_q 解的稀疏性在 $q \in [0, 1/2)$ 时无显著差异, 而当 $q \in (1/2, 1]$ 时, 随着 q 的减少而显著加强;
- ◆ $L_{1/2}$ 在 $L_q, q \in [0, 1]$ 中表现最好, 因此具有代表性。



$L_{1/2}$ 能快速算吗(算法一)

- **$L_{1/2}$ 问题:** $\min_x \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_{1/2} \right\}$
- **基本思想:** 用 $|x_0|^{1/2} + \frac{1}{2|x|^{1/2}}(x - x_0)$ 近似 $|x|^{1/2}$, 并取分母部分为当前值;
- **正则化参数设置:** L_1 中已有较成熟的方法, 典型的如: 交叉验证方法、AIC 准则, BIC 准则等;
- **算法收敛性** (Xu et.al., 中国科学, 2009);
- **算法复杂性:** $O(TmN^2)$ 。

Reweighted-L1 算法

Step I: 令 $t = 0$, $W^0 = I$, 给定算法最大迭代步数 T .

Step II: 求解加权 L1 极小化问题

$$x^{(t)} = \arg \min (\|y - Ax\|_2^2 + \lambda_t \|W^{(t)} x\|_1);$$

Step III: 令 $W^{t+1} = \text{diag}(W_1^{t+1}, W_2^{t+1}, \dots, W_p^{t+1})$, 其中

$$W_i^{(t+1)} = \frac{1}{\sqrt{|x_i^{(t)}|} + \varepsilon};$$

Step IV: 若 $t+1 = T$, 算法终止; 否则, 置 $t := t+1$ 并转 Step II。



$L_{1/2}$ 如能快速算吗 (算法二)

- **基本思想**: 对 $L_{1/2}$ 解的解析形式 $x = I(\cdot) + \lambda\mu\partial(\|\cdot\|_{1/2}^{1/2})^{-1} B_\mu(x)$ 使用逐次逼近; (等价于 Daubechies 的解耦技术);
- **正则化参数设置**: 假定问题的稀疏度 K 已知, 由优化理论可证明 $\lambda^* \leq 4/3(B(x^*)(K))^{3/2}$ 。采用逐次近似, 我们取 $\lambda_t = 4/3\mu B_t(K+1)^{3/2}$ 来逼近 λ^* ;
- **算法收敛性**: 对于足够小的 μ , 可证 $\|X_{t+1} - X_t\|_2 \rightarrow 0$ 。
- **算法复杂性**: $O(mN^2)$ 。

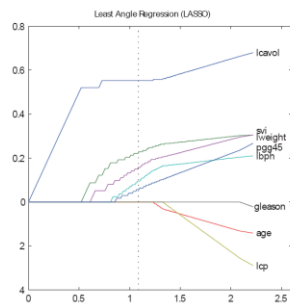
Half算法

- Step I:** 随机选取 $X_0 \in R^p$; 设 $B_0 = X_0 + A^T(y - AX_0)$, $\lambda_0 = \mu 4/3 B_0(K+1)^{3/2}$, 其中 K 为问题稀疏度的预估值, $\hat{X}_0(K+1)$ 为 X_0 的第 $K+1$ 大的分量, 取定 $0 < \mu < 1$ 和误差容限 ε 。置 $t := 0$;
- Step II:** 计算
$$X_{t+1} = H_{1/2, \lambda_t}(X_t + A^T(y - AX_t));$$
- Step III:** 令 $B_t = X_t + A^T(y - AX_t)$, $\lambda_{t+1} = 4/3\mu B_t(K+1)^{3/2}$;
- Step IV:** 若 $\|X_{t+1} - X_t\|_2 < \varepsilon$, 算法终止; 否则, 置 $t = t+1$, 转 Step II。

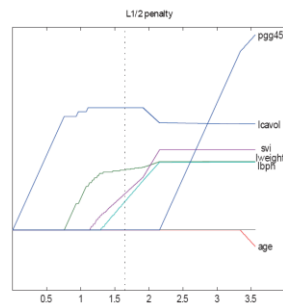


$L_{1/2}$ 应用效果如何 (变量选择问题)

- **前列腺癌数据**: 数据集取自 Stamey 等人 1989 年的研究, 主要反映了 97 个前列腺病人的前列腺特殊抗原 (PSA) 水平与 8 个临床指标 $lcavol$, svi , $lweight$, $pgg45$, $lbph$, $gleason$, age , lcp 之间的联系。人们期望知道哪些指标是影响 PSA 高低的关键指标, 以便有针对性的治疗与预测新病人样本的 PSA。



L_1 解



$L_{1/2}$ 解

Table: 不同变量选择方法的结果比较

	Lasso	Ridge	最佳子集	Boosting	HL0.5
$lcavol$	0.545	0.389	0.740	0.560	0.805
$lweight$	0.237	0.238	0.367	0.220	0.391
age	0	-0.029	0	0	0
$lbph$	0.098	0.159	0	0.090	0.187
svi	0.165	0.217	0	0.150	0.213
lcp	0	0.026	0	0	0
$gleason$	0	0.042	0	0	0
$pgg45$	0.059	0.123	0	0.060	0
预测误差	0.491	0.540	0.574	0.479	0.468



$L_{1/2}$ 应用效果如何 (信号恢复问题)

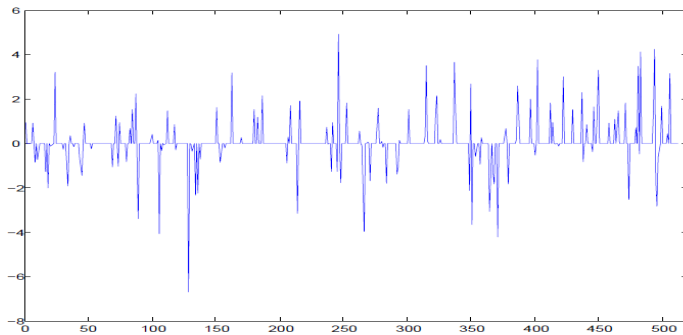


Table: 五种不同正则化方法随不同采样数的信号恢复结果

n	Method	MSE	Time(s)
330	L1	5.02E-05	1.905
	soft	8.91E-04	137.64
	hard	4.20E-07	0.1789
	RL1	9.67E-06	34.07
	half	1.12E-06	12.08
250	L1	6.7428 (fail)	11.01
	soft	6.9936 (fail)	116.64
	hard	10.0409 (fail)	0.5027
	RL1	1.27E-05	56.56
	half	2.71E-07	49.98
239	L1	1.3541 (fail)	1.41
	soft	5.6182 (fail)	88.21
	hard	11.9221 (fail)	0.6121
	RL1	2.60E-05	65.68
	half	3.56E-07	63.13
238	L1	10.4323	1.42
	soft	10.1627	54.06
	hard	16.3918	0.8111
	RL1	9.8711	78.5
	half	12.4422	66.24

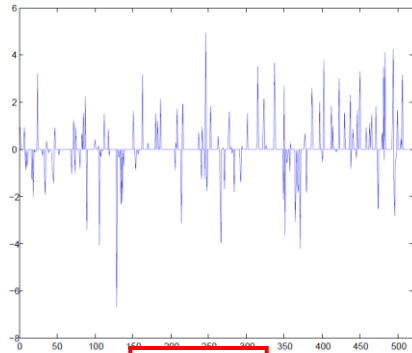
● **问题:** 基于压缩传感技术对稀疏信号进行恢复。给定长度为 512 的稀疏脉冲信号 $x = \Phi s$ ，其本质稀疏度为 $k = 130$ 。又给定 m 维采样 $y = \varphi x$ ，通过求解 $\min_x \|s\|_q, s.t. \varphi \Phi s = y$ ，来对原信号进行恢复。

● **结论:**

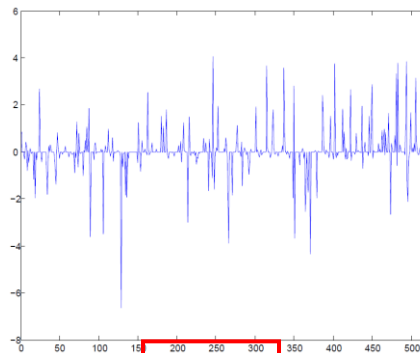
1. 在可恢复条件下， $L_{1/2}$ 总是能获得更高精度的解(恢复);
2. $L_{1/2}$ 所需的采样数少于 L_1 。



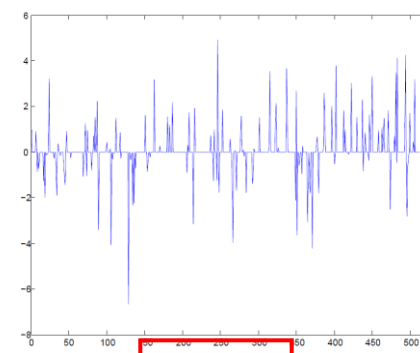
$L_{1/2}$ 应用效果如何 (信号恢复问题)



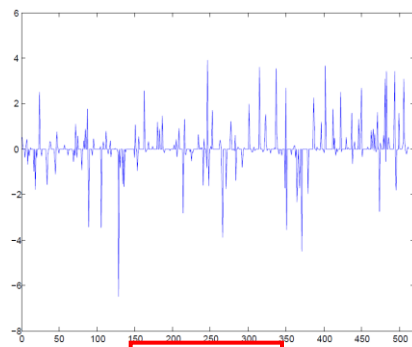
原信号



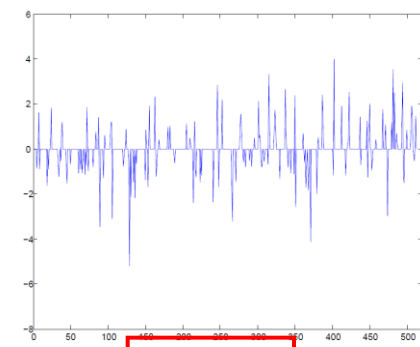
L1结果



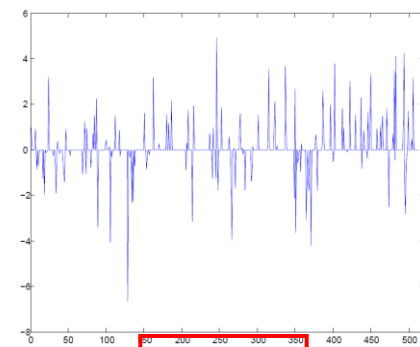
RL1结果



Soft结果



Hard结果



Half结果



L_{1/2}应用效果如何 (信号恢复问题)

- **问题:** 对上述信号恢复问题通过带噪声采样 $y = \phi x + \varepsilon$, 通过 $\min_x \|s\|_q, s.t. \|\phi \Phi s - y\|_2 \leq \varepsilon$ 对信号进行恢复; 噪声为 $\varepsilon \sim N(0, 0.1^2)$ 。

Table: 五种不同正则化方法随不同采样数的信号恢复结果

n	Method	MSE	Percentage	Time(s)
330	L1	3.71	1.4606	2.1361
	soft	3.33	1.311	34.95
	hard	2.86	1.126	0.4128
	RL1	3.58	1.4094	47.07
	HL0.5	2.60	1.0236	25.99
	Orcale	2.54		
300	L1	5.30	1.7785	1.8567
	soft	5.30	1.7785	51
	hard	8.35	2.802	0.4143
	RL1	4.04	1.3557	54.2262
	HL0.5	3.52	1.1812	52.29
	Orcale	2.98		
275	L1	6.10	1.7479	1.9061
	soft	6.33	1.8138	64.52
	hard	9.64	2.7622	0.4198
	RL1	5.05	1.447	61.98
	HL0.5	4.06	1.1633	50.42
	Orcale	3.49		
274	L1	6.34	1.8647	1.7519
	soft	6.14	1.8059	95.16
	hard	8.35	2.4559	0.4068
	RL1	6.10	1.7941	60.8344
	HL0.5	4.07	1.1971	40.49
	Orcale	3.40		

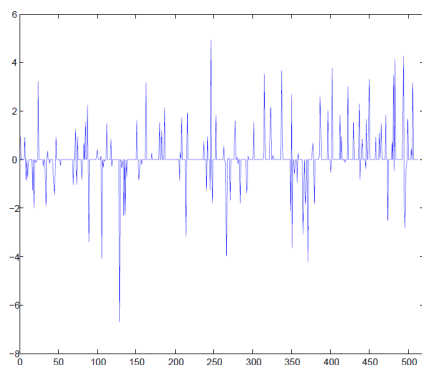
n	Method	MSE	Percentage	Time(s)
273	L1	6.87	1.9629	1.742
	soft	7.03	2.0086	64322
	hard	6.6766	1.9076	0.9239
	RL1	6.00	1.7143	63.6652
	HL0.5	6.21	1.7743	48.33
	Orcale	3.50		
239	L1	6.77	1.4685	1.4886
	soft	7.18	1.5575	86.16
	hard	13.58	2.9458	0.7719
	RL1	6.53	1.4165	54.0328
	HL0.5	9.84	2.1345	78.67
	Orcale	4.61		
238	L1	11.07	2.4655	1.3392
	soft	10.86	2.4187	50.45
	hard	15.83	3.5256	0.3629
	RL1	11.11	2.4744	49.7001
	HL0.5	12.90	2.8731	72.24
	Orcale	4.49		

观察:

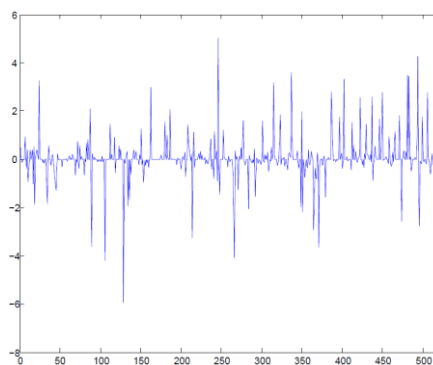
- 在同等噪声水平下, 发生相变之前 L_{1/2} 总能得到最好的信号恢复。



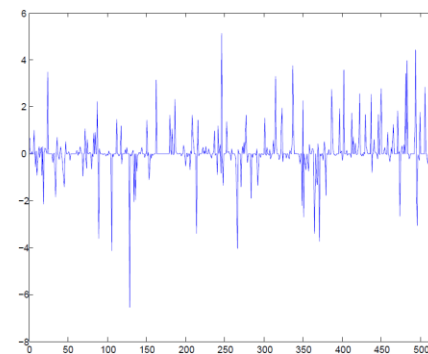
$L_{1/2}$ 应用效果如何 (信号恢复问题)



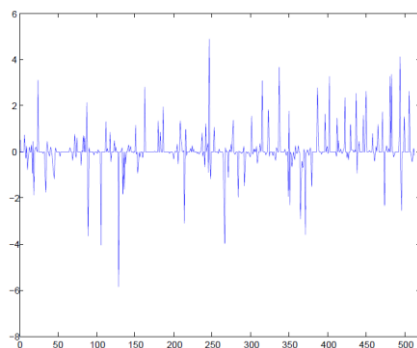
原信号



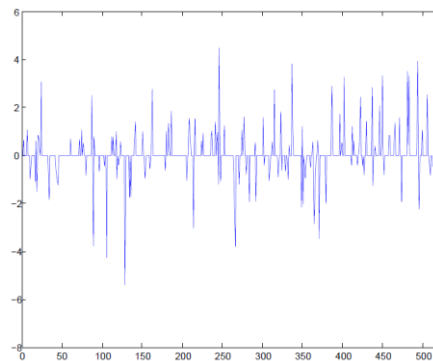
L1结果



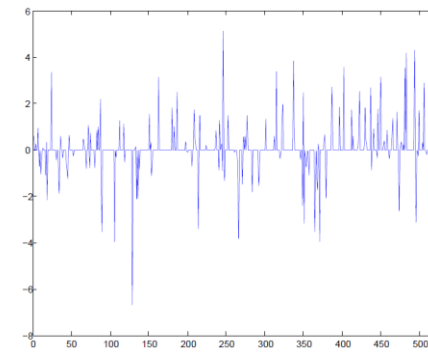
RL1结果



Soft结果



Hard结果



HL0.5结果

$L_{1/2}$ 应用效果如何 (图像重建问题)

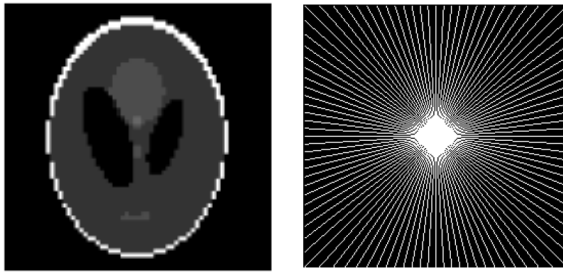


Table: 五种不同正则化方法随不同采样数的图像(32×32)恢复结果

L	Method	MSE	Time (s)
27	L1	fail	Inf
	soft	0.7348	80.11
	hard	7.58E-07	8.338
	RL1	fail	Inf
	half	1.63E-13	34.6663
35	L1	9.82E-05	53.14
	soft	9.35E-07	551.27
	hard	5.82E-07	3.0248
	RL1	1.18E-05	165.34
	half	1.35E-13	19.0638
52	L1	5.54E-06	33.2
	soft	4.93E-07	301.03
	Hard	3.69E-07	1.0515
	RL1	6.63E-06	108.3
	half	7.84E-13	64.93

- 问题:** 基于压缩传感技术对稀疏图像进行重建。给定原图像 $x = \Phi s$, 其中 s 具有本质 K 稀疏性, 又给定 x 的采样 $y = \varphi x$ (φ 为高斯随机传感矩阵), 通过求解 $\min_x \|s\|_q, s.t. \varphi \Phi s = y$ 来对图像进行重建;
- 采样方式:** 在原图像的 Fourier 频谱表示图上, 均匀取 L 条射线, 然后在每条射线上高斯采样。



$L_{1/2}$ 应用效果如何 (图像重建问题)

Table: 五种不同正则化方法随不同采样数的图像(64×64)恢复结果

L	Method	MSE	Time(s)
27	L1	fail	Inf
	soft	0.7348	80.11
	hard	7.58E-07	8.338
	RL1	fail	Inf
	half	1.63E-13	34.6663
35	L1	9.82E-05	53.14
	soft	9.35E-07	551.27
	hard	5.82E-07	3.0248
	RL1	1.18E-05	165.34
	half	1.35E-13	19.0638
52	L1	5.54E-06	33.2
	soft	4.93E-07	301.03
	hard	3.69E-07	1.0515
	RL1	6.63E-06	108.3
	half	7.84E-13	64.93



原图像



L1结果



RL1结果



Soft结果



Hard结果



Half结果



$L_{1/2}$ 应用效果如何 (图像重建问题)

Table: 五种不同正则化方法随不同采样数的图像(分辨率 256×256)恢复结果

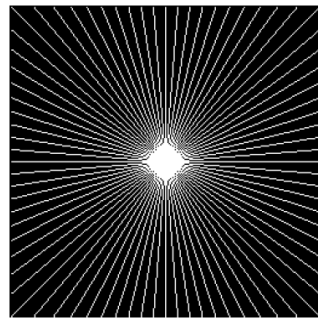
L	Method	MSE	Time
40	L1	fail	
	soft	8.26E+00	3.99E+03
	hard	1.54E+01	1.10E+03
	R11	fail	
	half	5.50E-12	1.42E+03
52	L1	9.3458(fail)	1008.72
	soft	1.0476(fail)	2.78E+03
	hard	8.35E-06	147.03
	R11	4.6881(fail)	3650.24
	half	3.26E-12	601.0603



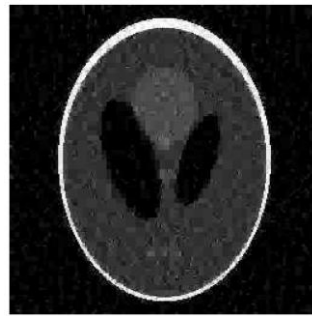
$L_{1/2}$ 应用效果如何 (图像重建问题)



原图像



采样数为40的
Fourier频谱图



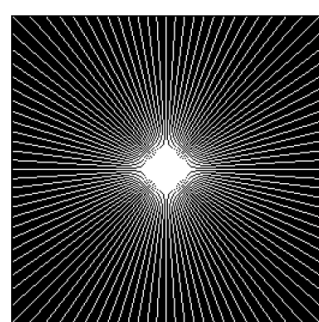
Soft结果:L=40



Hard结果:L=40



Half结果:L=40



采样数为52的
Fourier频谱图



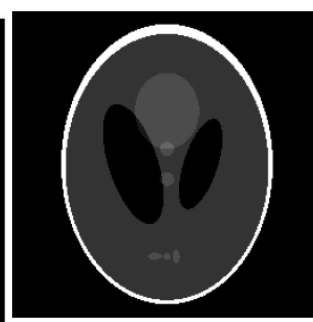
L1结果:L=52



RL1结果:L=52



Soft结果:L=52



Hard结果:L=52



Half结果:L=52



纲要

- 稀疏性
- $L_{1/2}$ 正则化
- 结论
 - 总结
 - 进一步的工作



总结

- 稀疏性问题是信息处理中的基本问题之一，当今解决稀疏性问题的普遍框架为 L_1 正则化。我们的研究比较完整地提出了解决稀疏性问题的一个全新框架—— $L_{1/2}$ 正则化理论。
 - ◆ 比 L_1 具有更好的理论性质
 - 比 L_1 更好的Oracle统计性质
 - 比 L_1 更宽泛的可恢复性条件
 - ◆ 解可解析表达
 - ◆ 在 L_q , $0 \leq q \leq 1$, 中具有代表性
 - ◆ 有快速的求解算法
- $L_{1/2}$ 正则化理论的正确性与方法的有效性已经在广泛的应用问题中得到验证。



进一步的研究问题

- 最小 $1/2$ 乘理论

$L_{1/2}$ 解是否具有某种广义逆形式？

- 压缩传感问题

现有的压缩传感框架假定了传感方式是线性（矩阵）的，而且是随机的，实际的信息获取方式未必是线性的和随机的，那么如何根据实际的信息获取方式设计合适的压缩传感策略？另外，现有的压缩传感理论又如何对现有的信息获取方式产生影响？



进一步的研究问题

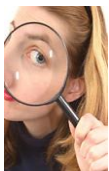
● $L_{1/2}$ 的数学基础问题

研究 L_q 正则化算法的基础是 L_q 空间的特征数量律。
如何推广**Xu-Roach**定理？

稀疏性问题
(L_p 正则化)



非欧氏框架下的
机器学习



$(a+b)^2 = a^2 + 2ab + b^2$
能有这样的公式
推演吗？

Xu-Roach定理 (类二项式公式)

$$\|a+b\|^p \geq \|a\|^p + p \langle J_p a, b \rangle + p \int_0^1 \frac{(\|a+tb\| \vee \|b\|)^p}{t} \delta_x \left(\frac{t\|b\|}{\|a+tb\| \vee \|b\|} \right) dt \quad \longleftrightarrow \text{一致凸}$$

$$\|a+b\|^p \leq \|a\|^p + p \langle J_p a, b \rangle + p \int_0^1 \frac{(\|a+tb\| \vee \|b\|)^p}{t} \rho_x \left(\frac{t\|b\|}{\|a+tb\| \vee \|b\|} \right) dt \quad \longleftrightarrow \text{一致平滑}$$

L_p 类二项式公式

$$\|a+b\|^2 > \|a\|^2 + 2 \langle Ja, b \rangle + (p-1) \|b\|^2 \Leftrightarrow p < 2$$

$$\|a+b\|^2 = \|a\|^2 + 2 \langle a, b \rangle + (p-1) \|b\|^2 \Leftrightarrow p = 2$$

$$\|a+b\|^2 < \|a\|^2 + 2 \langle Ja, b \rangle + (p-1) \|b\|^2 \Leftrightarrow p > 2$$



西安交通大学
XI'AN JIAOTONG UNIVERSITY

欢迎批评指正!