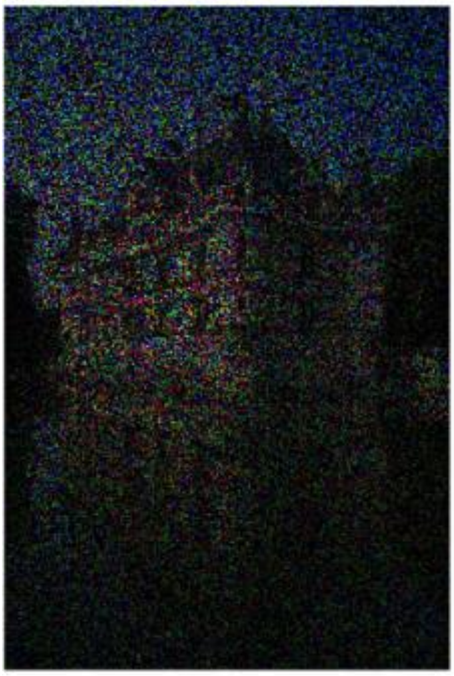
稀疏表达是近年来 SP, ML, PR, CV 领域中的一大热点，文章可谓是普天盖地，令人目不暇给。老板某门课程的课程需要大纲，我顺道给扩展了下，就有了这个上中下三篇介绍性质的东西。遗憾的是，我在绝大多数情况下实在不算是一个勤快的人，这玩意可能充满 bug，更新也可能断断续续，尽请诸位看官见谅了。顺道一提，ICCV09 有一个相关的 tutorial 。

据传博文里公式数量和其人气是成反比例关系的，一个公式可以驱散 50%的读者，我写完这个（上）之后点了点公式数量，觉得大约是要无人问津了。所以，在介绍稀疏表达之前，让我们先来展示下其在 computer vision 中的应用，吸引下眼球。

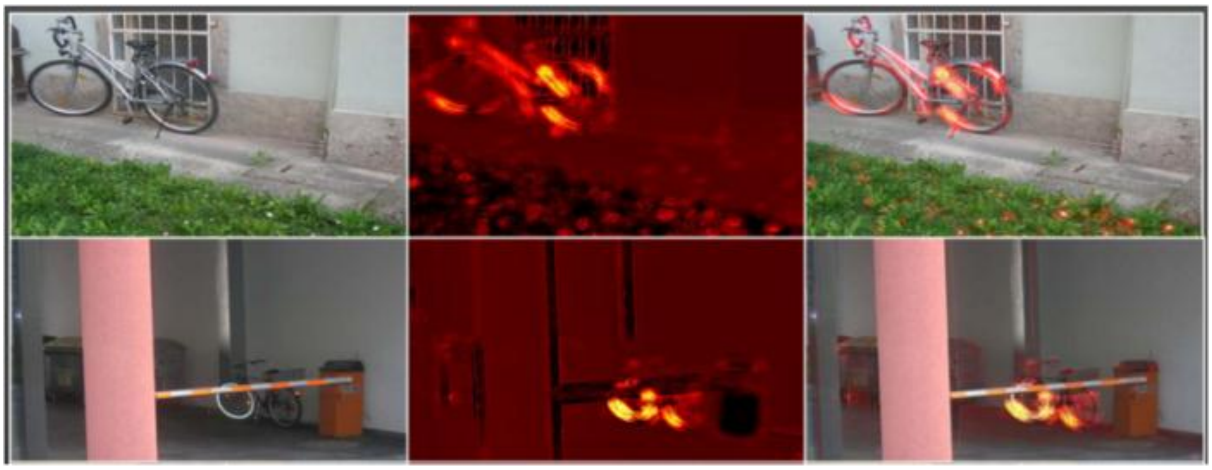
首先是图像恢复（以前有人贴过 Obama 还记得不），由左侧图像恢复出右侧结果



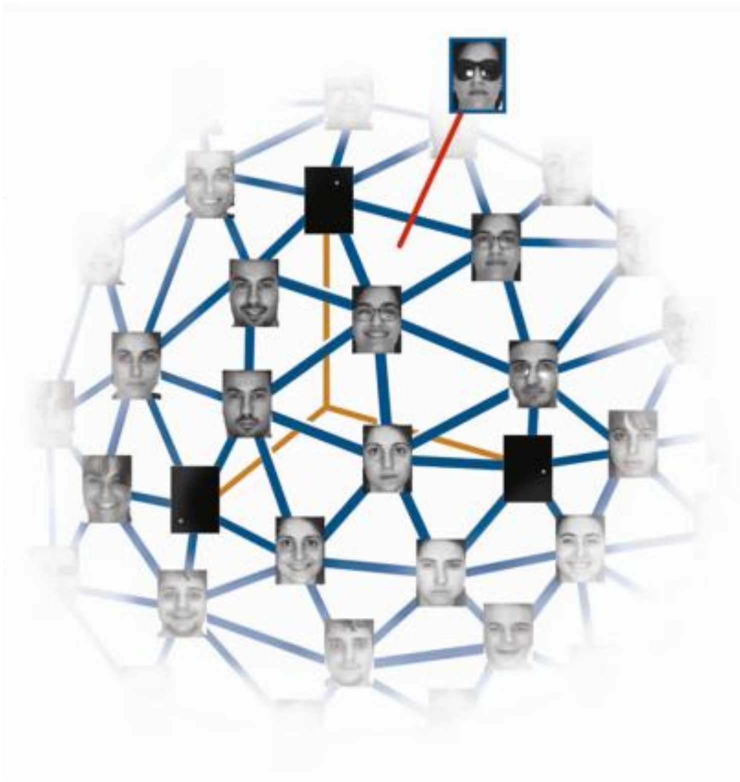
然后是类似的图像 inpainting

然后是图像去模糊，左上为输入模糊图像，右下为输出清晰图像及估计的相机运动（其实是 PSF），中间均为迭代过程：

再然后是物体检测（自行车），左侧输入图像，中间为位置概率图，右侧为检测结果



当然我个人还推荐 Yi Ma 的 sparse face，这个在对抗噪声的效果上很棒，比如下图中左侧的那张噪声图像（你能辨认是哪位不？这方法可以！）



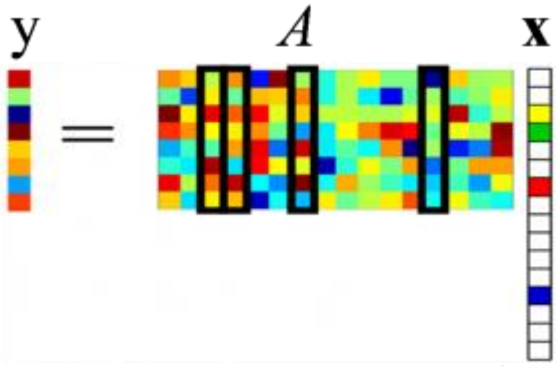
且说 sparse representation 这个概念，早在 96-97 年的时候就火了一把。最著名的大约要数 Nature 上的某篇文章，将稀疏性加入 least square 的 regularization，然后得到了具有方向特性图像块（basis）。这样就很好的解释了初级视皮层（V1）的工作机理，即对于线段的方向选择特性。几乎同一时期，著名的 LASSO 算法也被发表在 J. Royal. Statist. Soc B。Lasso 比较好的解决了 least square (l2 norm) error + l1 norm regularization 的问题。然而，这个时候绝大多数人没有意识到（或者没法解决）这 l1 norm 和稀疏性之间的联系。其实早在这之前，Osher 等人提出的 Total Variation （TV）已经包含了 l1 norm 的概念了，只不过 TV 原本是连续域上的积分形式。（啥？你不知道 Osher…想想 Level Set 吧）

在进入现代的压缩感知、稀疏表示这一课题前，让我们来首先回顾下这一系列问题的核心，即线性方程组

其中矩阵$A \in R^{m\times n}$ $m \ll n$，通常而言是满秩的。向量$x \in R^n$ $y \in R^m$。

现在已知y，A，求解x。学过线性代数的同学可能都会说：这个不难啊，因为$m \ll n$，故而这个方程组是欠定的，所以有无穷多组解啊，咱还可以算算基础解系啥的…

但是如果我们希望其解x尽可能的稀疏：比如$||x||_0$（即x中非零元个数）尽可能的小。那么问题就会变得比较微妙了，下图给出了问题的形象示意。



换言之给定 m 维空间中一组过完备的基$A \in R^{m \times n}$，如何选择最少个数的基向量，重构给定向量$y \in R^m$，其严格定义可以写成

$$\min \|x\|_0 \quad s.t. \quad Ax" = "y$$

时光之轮播快到 2003~2004 年，Donoho & Elad 做了一个很漂亮的证明，如果矩阵A满足某种条件，具体而言：

$$\sigma(A) \geq 2\|x\|_0$$

那么上文提及的 0 范数优化问题具有唯一的解。这里的$\sigma(A)$是个比较诡异（请允许我使用这词）的定义：最小的线性相关的列向量集所含的向量个数（吐槽：明白了么，我做 TA 的时候就被这个问题问倒了）。本来想在这个概念上唠叨两句，后来发现了 Elad 的一个 talk，清晰明了。

即便是唯一性得到了证明，求解这个问题仍然是 NP 难的。科研的车轮滚滚向前，转眼到了 2006 年，传奇性的华裔数学家 Terrence Tao 登场了，Tao 和 Donoho 的弟子 Candes 合作证明了在 RIP 条件下，0 范数优化问题与以下 1 范数优化问题具有相同的解：

$$\min \|x\|_1 \quad s.t. \quad Ax" = "y$$

其中 RIP 条件，即存在满足某种条件的（与 N 相关）常数:μ_N

$$(1-\mu_N)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1\mu_N)\|x\|_2^2 \quad \forall x \quad \|x\|_0 \leq N$$

RIP 条件是对于矩阵A列向量正交性的一种衡量（此处咱就不细说了）。其实早在 1993 年 Mallat 就提出过 Mutual Coherence 对于正交性进行度量，并提出了下文还要提及的 matching pursuit 方法。

实际上以上的 1 范数优化问题是一个凸优化，故而必然有唯一解，至此 sparse representation 的大坑初步成型。总结一下：

1. 如果矩阵满足$\sigma(A) \geq 2\|x\|_0$，则 0 范数优化问题有唯一解。
2. 进一步如果矩阵A满足 RIP 条件，则 0 范数优化问题和 1 范数优化问题的解一致。
3. 1 范数优化问题是凸优化，故其唯一解即为 0 范数优化问题的唯一解。

进一步可以考虑含噪声情况，即

$$\min \|x\|_0 \quad s.t. \quad \|Ax-y\|_2^2 \leq \varepsilon$$

可以得到相似的结果，有兴趣的同学可以查阅相关文献。理论坑只有大牛能挖，但一般人也能挖挖这个优化算法啊，于是 SP、ML、CV 邻域里都有做这个优化算法的，这个出招可就

真是五花八门了。据我所知，大致可以分为三大流派：

1. 直接优化

$$\min \|x\|_0 \quad s.t. \quad \|Ax - y\|_2^2 \le \varepsilon$$

一般的方法是 greedy algorithm，代表有 Matching Pursuit, Orthogonal Matching Pursuit

2. 优化

$$\min \|Ax - y\|_2^2 \quad s.t. \quad \|x\|_1 \le \varepsilon$$

还记得上面提到的 LASSO 么，这就是它的模型。

3. 如果已知拉格朗日乘子，优化无约束凸优化问题

$$\min_x \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$$

解这个的方法现在基本上 soft thresholding 的方法一统天下，常见的有 coordinate descent, Bregman Iteration (又是 Osher)等

4. 如果未知拉格朗日乘子，优化

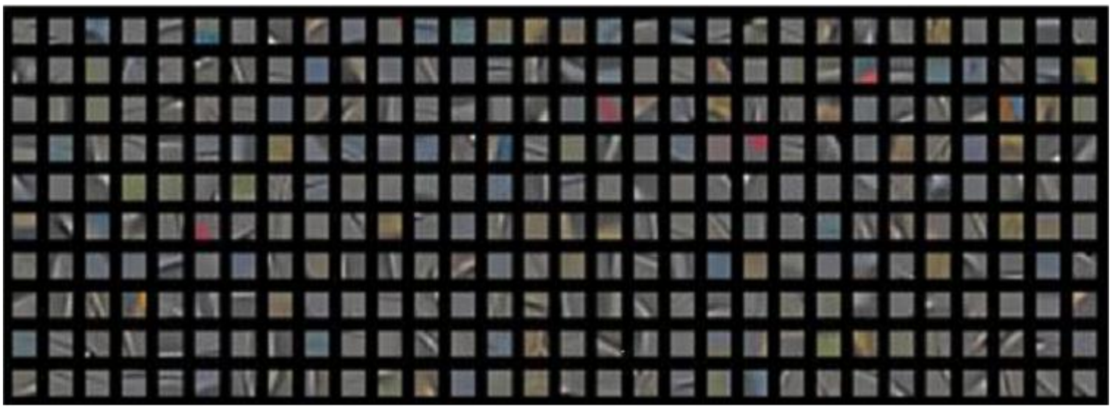
$$\min_{x,\lambda} \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$$

这类方法又叫 Homotopy，可以认为是 3 的扩展。核心出发点是 objective function 是λ的分段线性函数。

除此之外，还有利用 p 范数逐次逼近 0 范数的方法等等，此处不再赘述。顺道说一句，稀疏表示在不同的领域连名称都不同，搞信号的管这个叫 basis pursuit，搞统计的叫 l1 regularization….然后，让我们把话题拉回到 Nature 的那篇文章：如果我们不知道矩阵A，只知道一堆向量$\{y_i\}$。我们应当如何构造A，使得在这一字典（矩阵）下$\{y_i\}$的表示最稀疏？类比以上过程，这个问题被称为 Dictionary Learning，可以写成以下优化问题：

$$\min_{x_i,A} \sum_i \frac{1}{2}\|y_i - Ax_i\|_2^2 + \lambda\|x_i\|_1$$

这个东西可就相对麻烦了，最关键的是这个优化不是凸的（优化变量相乘）。所以一般的想法是 block descent：首先固定A，优化x_i（相当于多个独立的 1 范数优化问题）；其次将计算出的x_i固定，优化A，这就是一个（可能带约束）的 least square 问题。如此反复，直到算法收敛到某个（局部）极小值。实际上解这个问题的方法目前有三种：efficient sparse coding algorithm NIPS 06; K-SVD tsp 06; Online dictionary learning for sparse coding, ICML 09 & JMLR 10。前两种都是 batch 的方法，后一种是 online 的，据个人测试最后一种的方法比前两者要快很多很多….下面这个是我利用 ICML09 的方法从 1200 张彩色图像中训练出一组过完备基，具有比较好的方向特性。



最后，还记得本文开头的那些 demo 么？INRIA 做了一个 sparse representation 的 matlab 工具包 SPAMS，虽然不开源，但其效率（大部分时候）是现有公开工具包之冠（底层用了 intel 的 MKL），利用这个工具包，几行简单的 matlab 代码就可以几乎实现以上提及的所有 demo 了….大家有兴趣的话，欢迎尝试^_^

下期预告: 借着 collaborative filter 的东风, Candes 在 08 年又挖出了 matrix completion 的新坑。于是，当向量的 1 范数推广到矩阵的迹范数（trace norm）之后…..

开始正文之前，咱首先得说明一下，这篇东西偏向于理论，各位看官可以自行跳过某些部分。这方面的工作奠基人同样也是 compressive sensing 的大牛之一 E.J Candes（Donoho 的得意门生），以及 Candes 的学生 Ben Recht，前者刚从 caltech 被挖到 stanford，后者目前刚到 wisconsin 做 AP。Candes 大牛，stanford 统计系出生，师从 Donoho。Candes 原来的主要工作集中在小波分析上（实际上 C 牛非常多产），比如著名的 curvelets 以及 ridgelets，04 年左右开始和 Tao 合作从事 compressive sensing 的理论工作，这里有他的简要介绍。

继续唠叨，上回说到借着 collaborative filtering 的东风，矩阵的稀疏表示受到了广泛的关注。说到矩阵的稀疏性，大部分看官可能有所误解。这个矩阵稀疏表示严格而言可以分为两种：

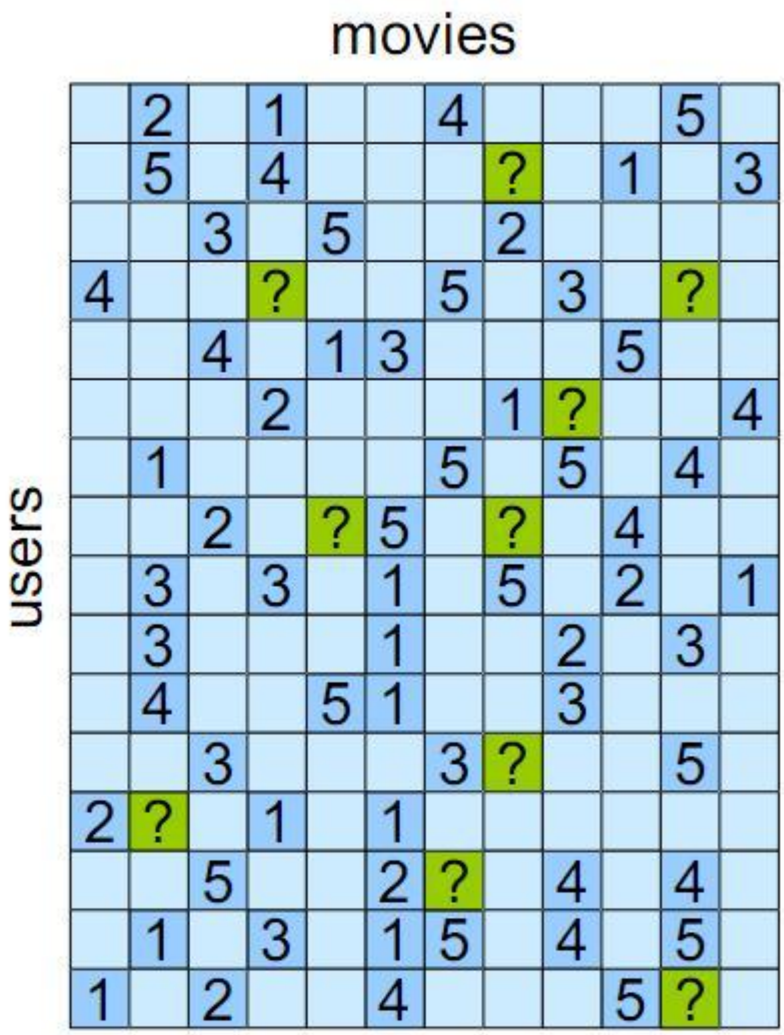
1. 矩阵元素的稀疏性，即矩阵非 0 元个数相对较少。参照向量的范数，同样可以定义矩阵的 0 范数，并将其松弛到矩阵的 1 范数的优化问题。
2. 矩阵奇异值的稀疏性，即矩阵奇异值中非 0 元的个数（即矩阵的秩）相对较少。仿照向量情况下 0 范数与 1 范数的关系，同样可以将其松弛的到迹范数（trace norm）的优化问题。

咱下面会分别聊聊这两个问题。首先，咱的出发点是 machine learning 中的 collaborative filtering，这个概念并不是啥新东西了，最早大约可以追朔到 1992 的某篇同名文章。这玩意是做啥的呢，通俗的说，每次你在淘宝上闲逛的时候，下面都会有一行推荐商品。这些个网络服务商（淘宝，Amazon, Ebay）就在想了，如果这个推荐系统做的足够好，

那么消费者（比如你我）的购物欲望就会得到刺激，这个销量也就上去了。实际上，这和超市里玲琅满目的货架是一个道理。

这里就得提提 [Netflix Prize]这件事了，话说 netflix 是家在线 dvd 租赁公司，这公司就抱了同样的想法。不过这家公司想了个主意：该公司提供数据，出资 100 万美刀，奖励研发这个推荐系统算法的小组，并要求这些算法发表在学术会议或期刊之上。这可以算是现实版的百万富翁了（学术和 money 两不误），于是 collaborative filtering 着实火了一把（比如 SIGKDD 上的不少文章）。最终历时两年，由 AT&T 实验室成员组成的 BellKor's Pragmatic Chaos 赢得了这 100 万刀。顺到一提，国内也有不少家伙参与了这个 Prize，比如排名第二的 Ensemble 组里就能看到中科院某所学生的身影。

这个推荐系统咋做呢？我们先从简单的模型开始。以 netflix 为例，netflix 有个影评系统，在你租完 DVD 以后会让你打分（1-5 分）。当然不是所有人都会认真去打，实际上只有少数家伙会给打分（这世界上懒人何其之多）。同样，对每个用户而言，他也只可能给部分看过的 DVD 打分。假设现在有m个用户和n部电影，如果把所有评分列成一张大表，可以得到矩阵$D \in R^{m \times n}$。其中，每一行对应一个用户的评分，每一列对应一部电影的用户评价。可以想象，这个矩阵中只有少部分元素是已知的（图 1）。



从现有的用户数据，来预测未知的用户数据，这就是 collaborative filtering 了。那么这个东西怎么实现呢？解释起来难，做起来容易，这个模型放在在 topic model 里叫做 Probabilistic latent semantic analysis （PLSA），放在代数里叫做矩阵分解（Matrix Fatorization）或者矩阵填充（Matrix Completion），这里就只能形象的解释下。虽然用户千奇百怪、电影成千上万，但总可以归结为若干类型：比如有腐女向、宅男向电影之分，再比如有悲剧也有喜剧。如果把这些 latent factor 画成一个空间，那么不同的用户群体应当位于这个 latent factor 空间的不同位置，体现了不同用户的喜好。如果可以把用户喜好连同潜在的 latent factor 一同计算出来，预测也自然水到渠成了。从某种角度来看，奇异值分解过程也就是上述的剥离 latent factor 和用户喜好的过程 这其中的 philosophy 可以参见这篇文章。

咱首先要谈的是矩阵奇异值的稀疏性，为此先来回忆下奇异值分解$M" = "USV$。

1. 奇异值非负，即$\sigma_i(M) \geq 0$

2. 奇异值非 0 元的个数即为矩阵M的秩（rank）

如果把奇异值写成对角矩阵S的形式（比如 SVD 分解的标准形式），其对角元为$\sigma_i(M) \geq 0$。进一步，矩阵M的迹范数（trace norm）$\|M\|_{tr}$定义为矩阵奇异值之和，即有

$$\|M\|_{tr}" = "\sum_i$$

现在我们可以把 collaborative filtering 的基本问题回顾一下，给定一张推荐数据表$D \in R^{m\times n}$，已知其下标子集Ω中的元素（也就是有评分的部分），如何恢复这个矩阵？这就是 matrix completion 的问题了…

乍眼一看，这基本就是 mission impossible 了，即使只有一个元素未知，这个矩阵也不可能唯一。但是如果我们加一些限制条件，这个问题就变得有趣起来了。Candes 考虑的是这么一个问题：

$$min_{\tilde{M}} \quad rank(\tilde{M}) \quad s.t. \quad P_\Omega(\tilde{M} - M)" = "0$$

其中$P_\Omega()$表示在子集Ω上的投影（即只取子集上的对应元素）。实际上，同样的问题可以有不同的表达形式，如果把这个优化问题稍作调整，可以得到相对容易解释的模型：

$$min_{\tilde{M}} \quad \|P_\Omega(\tilde{M} - M)\|_F^2 \quad s.t. \quad rank(\tilde{M}) \leq R_c$$

其中 Frobenius 范数也就是矩阵的 2 范数。从这个式子来看，我们希望找到这么一个矩阵$\tilde{M}$，使得其在已有的数据上和用户评分尽可能的一致（2 范数意义下），同时具有比较低的秩（受到上限R_c的约束）。这里对于秩的约束，很多时候是为了降低模型自身的复杂度（比如 collaborative filtering，multiple instance learning）。当然，这里也可以看成是一个 fidelity term + regulariztion term 的经典形式。

实际上矩阵的 rank 是一个不那么友好的函数，rank 自身是非凸、不连续的，最后的结果就是对于 rank 的优化问题是 NP 难的。类比 0 范数与 1 范数的关系，矩阵M的秩（rank）

相当于这个对角阵的 0 范数；矩阵M的迹范数（trace norm）$\|M\|_{tr}$相当于这个对角矩阵的 1 范数。为此，如果这个对角矩阵足够稀疏，即矩阵M的秩$r \ll min(m,n)$，那么可参照向量的稀疏表示，利用矩阵的迹范数(trace norm)代替矩阵的秩（rank）。

$$min_{\tilde{M}} \quad \|\tilde{M}\|_{tr} \quad s.t. \quad P_\Omega(\tilde{M}-M)" = "0$$

同样，由于迹范数(trace norm)是凸的，上式是一个凸优化问题，故而必有唯一的最优解。如果这种近似是可以接受的，那么这个问题自然也就解决了。

这种近似靠谱么？这就是 Candes 和 Recht 回答的关键问题。Candes 从 random orthogonal model 出发，证明了在此假设下从某个秩为r的真实矩阵$M \in R^{m\times n}$中均匀抽取k个元素，且满足(这里不妨设$m \leq n$，反之只需要转置即可)

$$k \geq Cn^{\frac{6}{5}} r \log n$$

则凸优化问题的唯一最优解M^*至少以概率$1-cn^{-3}$逼近原始矩阵M，即有

$$P(M^*=M) \geq 1-cn^{-3}$$

其中c, C均为某常数。更进一步，如果矩阵的秩r足够小，对于元素数量k的要求会进一步降低。

咱来聊聊这个结果，这说明在 random orthogonal model 假设成立的条件下，如果r相对于$min(m,n)$比较小，那么只需要知道这个矩阵中约$O(n^{\frac{5}{4}})$个元素，就可以很高的概率恢复出这个矩阵。举例而言，如果我们有一个$4096*4096$秩为 10 的矩阵，那我们大致只需要从中随机抽取约 270 万个元素就可以（以很高概率）恢复出原始矩阵了(当然 270 万貌似也是一个很大的数，但原始矩阵约含有 1700 万个元素…）。实际上，这是一个相对保守的界，Recht 在此基础上还进行了一系列的理论工作。自从出现了这个之后，under mild condition，大家都把 rank 直接放成 trace norm 了…从实用的角度来说，Candes 告诉我们用凸优化去近似一个 NP 问题，可能得到很好的解。从实验结果来看（代码见此），这种近似有时候效果一流，但有时候也根本不 work（违背了假设条件），故而具体问题还得具体对待。

虽然早在 04 年 NIPS 上，就有人提出了类似的优化方法（MMMF），用 trace norm 代替 rank，并且 ML 领域中也确实有不少类似的工作。但是，Candes 的工作解决了根本的理论问题，并为一系列的 rank minimization 的问题指了一条出路。这里有一个比较有意思的地方是，MMMF 是从构造最大间隔线性分类器的角度出发来考虑 matrix factorization 的问题，并且用的是 low norm，但和 matrix completion 的模型本质上是差不多的，两者关系大家可以自行推导下。

咱接着要讨论的是矩阵元素的稀疏性，这个工作也和 Candes 有着很大的关系。咱先把上面的公式照着 copy 一遍：

$$min_{\tilde{M}} \quad \|P_\Omega(\tilde{M}-M)\|_F^2 \quad s.t. \quad rank(\tilde{M}) \leq R_c$$

如果咱已知矩阵M的全部元素，这个东西类似很常见的 PCA 了：

$$min_{\tilde{M}} \quad \|\tilde{M} - M)\|_F^2 \quad s.t. \quad rank(\tilde{M}) \leq R_c$$

这样问题就变成了去噪+降维。进一步把 F 范数（2 范数）改写为 0 范数：

$$min_{\tilde{M}} \quad \|\tilde{M} - M)\|_0 \quad s.t. \quad rank(\tilde{M}) \leq R_c$$

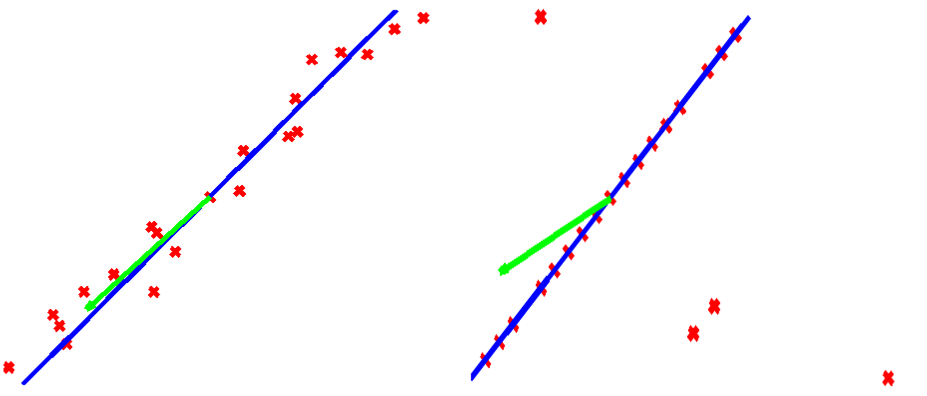
为啥是 0 范数呢，这是基于这么一种假设：误差相对于总体样本而言总是稀疏的。于是，我们可以引入辅助变量表示误差S，并把上式稍作改写：

$$min_{\tilde{M},S} \quad rank(\tilde{M}) + \lambda\|S\|_0 \quad s.t. \quad \tilde{M} + S" = "M$$

这里的λ用于平衡矩阵的秩和误差的稀疏性。同样，rank 和 0 范数什么的都是相当讨厌的东西，于是咱松弛一下，就有

$$min_{\tilde{M},S} \quad \|\tilde{M}\|_{tr} + \lambda\|S\|_1 \quad s.t. \quad \tilde{M} + S" = "M$$

这就是 Robust Principle Component Analysis (RPCA) 或者 Principle Component Pursuit 的核心模型了。这幅图很好的说明了 RPCA 和 PCA 的区别（转自 Yi Ma 主页）。



PCA RPCA

说起 RPCA，这里岔开两句，这个东西原来是 Yi Ma 的学生 John Wright 发在 NIPS09 上的一篇文章。结果接收之后，被 Candes 指出了一个 bug（审稿人没看出来），于是 Candes 对这个问题进行了考虑，从而就有了一篇叫做《Robust Principal Component Analysis?》的文章（preprint）。Candes 证明了在同 matrix completion 基本相同的假设下，这种近似以很高的概率恢复精确结果（详细结果可见 RPCA 的论文）。特别的，此时可以简单选择参数。Matrix Completion（MC）和 RPCA 在 Yi Ma 的主页上有一个简单的介绍，上面列举了相关文献与代码的链接。

MC 和 RPCA 在 computer vision 上有啥用呢？John Wright 在 NIPS 的文章里做了两个实验：背景建模，人脸阴影去除。大家有兴趣可以查查 cvpr 10 的 paper，有用 MC 来做 video denoising 的，有用 RPCA 来做人脸对齐的…还有那篇 best paper 也是紧密相关。咱本来还想聊聊这些模型的优化算法，鉴于篇幅所限，就只能留到（下）篇去了。

分析压缩感知基本理论

1．引言

信号采样是模拟的物理世界通向数字的信息世界之必备手段。多年来，指导信号采样的理论基础一直是着名的 Nyquist 采样定理。定理指出，只有当采样速率达到信号带宽的两倍以上时，才能由采样信号精确重建原始信号。可见，带宽是 Nyquist 采样定理对采样的本质要求。但是，对于超宽带通信和信号处理、核磁共振成像、雷达遥感成像、传感器网络等实际应用，信号的带宽变得越来越大，人们对信号的采样速率、传输速度和存储空间的要求也变得越来越高。为了缓解对信号传输速度和存储空间的压力，当前常见的解决方案是信号压缩，如基于小波变换的 JPEG2000 标准。但是，信号压缩实际上是一种严重的资源浪费，因为大量的采样数据在压缩过程中被丢弃了，而它们对于信号来说是不重要的或者只是冗余信息。从这个意义而言，我们得到以下结论：带宽不能本质地表达信号的信息，基于信号带宽的 Nyquist 采样机制是冗余的或者说是非信息的。

一个很自然的问题是：是否存在或者能否提出一种基于信息的采样理论框架，使得采样过程既能保持信号信息，又能只需远少于 Nyquist 采样定理所要求的采样数目就可精确或近似精确重建原始信号？简言之，能否同时实现信号的采样与压缩？与信号带宽相比，稀疏性能够直观地而且相对本质地表达信号的信息。事实上，稀疏性在现代信号处理领域一直起着至关重要的作用，例如基于稀疏性的逼近、基于稀疏性的估计、基于稀疏性的压缩、基于稀疏性的降维等。不同于 Nyquist 信号采样机制，Candès、Tao、Romberg、Donoho 等人。

近年来基于信号稀疏性提出一种称为压缩感知(compressed sensing)或压缩采样(compressivesampling)的新兴采样理论，成功实现了信号的同时采样与压缩。

简单地说，压缩感知理论指出：当信号在某个变换域是稀疏的或可压缩的，可以利用与变换矩阵非相干的测量矩阵将变换系数线性投影为低维观测向量，同时这种投影保持了重建信号所需的信息，通过进一步求解稀疏最优化问题就能够从低维观测向量精确地或高概率精确地重建原始高维信号。在该理论框架下，采样速率不再取决于信号的带宽，而在很大程度上取决于两个基本准则：稀疏性和非相干性，或者稀疏性和等距约束性。当前，压缩感知理论主要涉及三个核心问题：

(1) 具有稀疏表示能力的过完备字典设计；

(2) 满足非相干性或等距约束性准则的测量矩阵设计；

(3) 快速鲁棒的信号重建算法设计。

上迅速兴起的热门研究方向。目前，学者们已经在模拟-信息采样、合成孔径雷达成像、遥感成像、核磁共振成像、深空探测成像、无线传感器网络、信源编码、人脸识别、语音识别、探地雷达成像等诸多领域对压缩感知展开了广泛的应用研究。

值得注意的是，Rice 大学已经成功设计出了一种基于压缩感知的新型单像素相机，在实践中为取代传统相机迈出了实质性的一步。

目前，压缩感知理论的相关工作尚有很多亟待解决的问题，尤其是国内关于压缩感知理论的基础研究基本处于空白。为此，本文围绕稀疏字典设计、测量矩阵设计、重建算法设计三个核心问题，对压缩感知的基本理论和实现方法进行了系统阐述，同时指出了压缩感知有待解决的若干理论推广和关键技术。本文结构安排如下：第 2 部分基于非相干性和和等距约束性准则系统阐述了压缩感知的基本理论；第 3 部分系统介绍了压缩感知的三个层面的核心技术，即稀疏字典设计、测量矩阵设计、重建算法设计；第 4 部分指出压缩感知有待解决的若干关键问题；第 5 部分对全文作了总结。

2．压缩感知理论的基本框架

2.1 信号的稀疏性定义 (可压缩信号) 称信号 N x ∈ ?? 在 Ψ 域是可压缩的，如果变换向量 s 大部分分量的取值很小，只有少部分分量的取值很大；或者说只要少部分取值大的分量就能很好地逼近原始信号 x 。

例如，尽管很多信号自身取值都是非零的，但是在小波正交基下，信号大部分小波系数的取值都很小，只有少量的小波系数取值很大，这些大系数承载了信号的绝大部分信息。小波变换的这种稀疏性或可压缩性已被成功地用于现代图像压缩标准—JPEG2000。这种通过稀疏变换实现压缩的方法称为变换编码[1]。变换编码在现代数据获取系统中一直发挥着重要的作用，例如数码相机、数码摄相机。但是，这种采样再压缩的数据获取过程造成了严重的资源浪费，尤其对于核磁共振成像、雷达遥感成像等特殊应用。例如，百万级像素的传感器只使用了压缩后的几百 Kbyte 数据。

2.2 压缩感知问题描述考虑一般的采样问题[23]:

2.3 压缩感知基本理论

在压缩感知理论中，采样速率不再取决于信号带宽，而在很大程度上取决于两个基本准则，即稀疏性和非相干性，或者，稀疏性和等距约束性。

2.3.1 基于稀疏性和非相干性

准则的压缩感知理论定理 1 显示了非相干正交基对(Φ% ,Ψ)对于压缩感知的重要性。不等式指出，μ (Φ% ,Ψ)越小，压缩采样所需的测量个数就会越少，意味着压缩测量 yk 包含 x 的信息就会越多。特别地，当 μ (Φ% ,Ψ)趋近于 1 时，压缩感知只需 O(K ? log N)个压缩测量就能以大概率精确重建原始信号。从信号重建过程来看，只需求解约束的 l1 范数最小化这个凸最优化问题；从数值解的精度来看，压缩采样的 O(K ? log N)个压缩测量在很大概率上没有损失原始信号的信息。

定理 1 存在的不足之处是：只适用于 K -稀疏信号，而实际信号如自然图像往往是可压缩的；最优化问题的解有不精确的可能。

2.3.2 基于稀疏性和等距约束性

准则的压缩感知理论定理 3 通过收紧等距约束常数的范围，使得 Θ 的任意 2K 列组成的子矩阵更趋于近似正交，将压缩感知松弛转化为约束 1l 范数最小化的凸最优化问题，不仅保证了压缩感知的理论完备性，而且在数值计算上保证了压缩感知的可行性。当信号 x 在 Ψ 域是 K -稀疏的，真实解 s 只有 K 个非零项，则 s? = s。换句话说，当 Θ 的 2K 阶等距约束常数小于 0.414，约束最优化问题(10)能够利用 M ?? N 个压缩测量精确重建原始信号。定理 3 不仅适用于 K -稀疏信号，而且适用于可压缩信号。当信号是可压缩的，(11)式给出了近似解与真实解的误差范围，完全由某个固定常数 C 和 s 的 N ? K 个最小绝对值分量决定。同时，测量矩阵设计部分将指出，测量次数 M 与稀疏性 K 和信号长度 N 之间的具体关系将由稀疏矩阵 Ψ 和测量矩阵 Φ 共同决定。

3．压缩感知的核心问题

3.1 压缩感知的稀疏字典设计信号 x 的稀疏性或可压缩性是压缩感知的重要前提和理论基础。因此，压缩感知理论首要的研究任务就是信号的稀疏表示研究。稀疏字典设计是压缩感知的核心问题之一，在于：

只有选择合适的稀疏字典，才能保证表示系数具有足够的稀疏性或衰减性，才能在减少压缩测量的同时保证压缩感知的重建精度。

目前，稀疏字典主要包括正交基字典、紧框架字典、过完备字典。正交基字典主要是计算调和分析中的正交变换系统，如 Wavelet 变换；紧框架字典主要是以 Ridgelet、CurveletBandletContourlet 为代表的图像几何多分辨率表示或者称 Beyond Wavelet 变换；在过完备字典中，用于稀疏表示的不再是“单一基”，而是通过构造或学习得到的冗余原子库，通过提高变换系统的冗余性增强信号逼近的灵活性，提高对图像等复杂信号的稀疏表示能力。

1993 年，Mallat 和 Zhang 首次提出了基于过完备字典的稀疏分解思想，指出了过完备字典对于信号稀疏表示的必要性和重要性.基于过完备字典的稀疏分解依然是当前信号稀疏表示研究的热点和难点。过完备字典由称为过完备原子库的冗余系统构成，原子不必再是“单一基”函数。过完备字典的构造或学习应遵循基本准则：字典中的原子应能尽量匹配信号本身固有的各种不同特征。在这种准则下，稀疏字典必定是非正交的且是冗余的，正是通过增加原子个数提高变换系统的冗余性来增强信号逼近的灵活性，进而提高图像等复杂信号的稀疏表示能力.当字典中的原子个数大于信号维数 N 且包含 N 个线性无关向量张成整个信号空间时，字典称为过完备的。基于过完备字典的稀疏分解使得信号能量集中在极少数原子上， 正是这些具有非零系数的原子匹配了信号的不同特征。

设计适合特定信号的过完备字典，目前主要包括人工构造和训练学习两大类方法。基于构造方法的过完备字典设计是主流，主要包括：Wavelet 和局部 Cosine 函数的级联、各向同性的 Gabor 字典、各向异性的 Refinement-Gaussian 混合字典、各向异性的 Gabor 感知多成份字典等。虽然 Wavelet 能够稀疏表示信号中的点奇异特征、局部 Cosine 函数能够有效表征纹理特征，但是由于 Wavelet 的可分离性与各向同性，Wavelet 和局部 Cosine 函数的级联不能有效刻画图像中的边缘轮廓等线奇异特征。各向同性的 Gabor 字典能够有效刻画纹理特征。但是，由于 Gabor 原子的各向同性和单频带宽，也不适于有效刻画图像中的边缘轮廓等线奇异特征。各向异性的 Refinement-Gaussian 混合字典采用 Gauss 函数及其二阶导数作为原子的生成函数，能够有效表征图像中边缘轮廓结构，但是没有能够有效刻画纹理特征的原子。各向异性的 Gabor 感知多成分字典基于视觉感知的“有效编码假设”，以二维 Gabor 函数作为字典原子的生成函数，依据视觉皮层中神经元的响应特性和组织方式以及图像的多成分特性，约束生成函数中自由参数的取值范围，通过对生成函数进行平移、旋转、伸缩等几何变换生成一系列原子，遵循了过完备字典构造应该遵循的基本准则。基于学习的过完备字典是过完备字典设计问题的难点和热点，涌现的典型学习算法主要有：其中，K-SVD 这类学习算法具有代表性，稀疏表示效果好，计算复杂度低，但不足之处是缺乏严格的理论支撑.基于过完备字典的稀疏表示的另一个方面是设计快速有效的稀疏分解算法。由于压缩感知信号重建问题追求的同样是稀疏解，因此某种程度上这里的稀疏分解算法可以推广应用到压缩感知问题。为了避免重复，稀疏分解的相关算法将在下文予以介绍。国内关于稀疏表示也展开了广泛的理论和应用研究。例如，谢胜利等人基于稀疏表示思想开展自适应的盲分离算法研究;尹忠科等人利用快速傅立叶变换实现匹配追踪的快速算法研究。

3.2 压缩感知的测量矩阵设计测量

矩阵设计是压缩采样理论的核心，直接决定了压缩采样理论是否能够成功实现。由于压缩测量个数和信号重建精度以及信号稀疏性有着密切的联系，因此测量矩阵的设计应该与稀疏字典的设计统筹考虑。从原理的角度看，测量矩阵的设计要以非相干性或等距约束性为基本准则，既要减少压缩测量个数又要确保压缩感知的信号重建精度。从技术的角度看，测量矩阵的设计包括两个方面：一是测量矩阵的元素，Candès 等人给出了随机生成的设计策略；二是测量矩阵的维数，压缩测量个数 M 与信号稀疏性 K 和信号长度 N 应该满足一定的关系。

3.2.1 基于非相干性准则的测量

矩阵设计稀疏信号的非相干压缩感知定理要求非相干的正交基对(Φ% ,Ψ)。理想情况下，当正交基对具有最大非相干性，即 μ (Φ% ,Ψ) = 1，压缩感知信号重建问题(7)只需 O(K ? log N)个压缩测量就能以大概率精确重建原始信号。典型的例子是，正交基对(Φ% ,Ψ)由 Delta 基函数和 Fourier 基函数设计获得[2][4][6][23]。由定理 1 知，若信号 N x∈?? 本身是 K -稀疏的，当 k?% 取为 Fourier 基函数( ) 1/ 2 i 2 kj / Nk j N e π ? ? ? % = ，ψ j 取为 Delta 基函数 ψ j (k) = δ (j ? k) ，则只要在 Fourier 域任意随机均匀地选取 M = O(K ? log N)个变换系数 sΩ,|Ω|=M (即随机测量)，就能保证很大概率地精确重建原始信号；如果信号 N x ∈ ?? 在 Fourier 基函数下是 K -稀疏的，当 k?% 取为 Delta 基函数( ) ( ) k ?% j = δ j ? k，ψ j 取为 Fourier 基函数( ) 1/ 2 i 2 jk / Nj k N e π ψ ? = ，则只要在时域任意随机均匀地选取 M = O(K ? log N)个信号值，就能保证很大概率地精确重建原始信号。

一般地，信号重建问题(7)需要 M = O(μ 2 (Φ% ,Ψ) ? K ? log N)个随机压缩测量才能很大概率地精确重建原始信号[2][23]。例如，当 Φ% 取为 Noiselet，Ψ 取为 Harr Wavelet，μ (Φ% ,Ψ) = 2 ；Ψ 取为 Daubechies D4 Wavelet，μ (Φ% ,Ψ) = 2.2；Ψ 取为 Daubechies D8 Wavelet，μ (Φ% ,Ψ)=2.9。

事实上，当 Ψ 取为任意固定正交矩阵，Φ% 取为某种随机正交矩阵，Φ% 与 Ψ 在很大程度上是非相干的[23]。例如，在单位球上随机均匀独立地选取 N 个正交单位向量，通过这种随机方式获得的正交基 Φ% 与任意固定正交基 Ψ 的相干性大约为 2log N ；如果 Φ% 的所有元素都是独立同分布选取的，例如服从 Gaussian 分布或 Bernoulli 分布，Φ% 与 Ψ 的相干性也是非常小的。

3.2.2 基于等距约束性准则的测量

矩阵设计由定理 3、4 知，信息算子 Θ 或者测量矩阵 Φ 要满足 RIP 条件。注意的是，在这些定理中， Φ 既可以是随机矩阵也可以是确定性矩阵。从这个意义上讲，基于稀疏性和 RIP 准则的压缩感知理论具有一般性。这里主要介绍随机测量矩阵[5],[23]。

首先，对于(3.2.1)讨论的非相干正交基对(Φ% ,Ψ)，只要满足 M ≥ C ? K(log N)4，其中，C 为某个固定常数，则 Θ = Q Φ% Ψ 将很大概率地满足 RIP。如果希望信息算子不满足 RIP 的概率小于 O(n) ?β ，其中 β > 0，则要求 M ≥ C ? K(log N)5。

其次，(3.2.1)部分随机产生正交基 Φ% 的方法同样适用于测量矩阵 Φ，例如：

(1) 在?? M 的单位球上均匀独立地选取 N 正交单位向量生成 Φ ；

(2) 独立同分布地从均值为零方差为 1/M 的正态分布生成 Φ；

(3) 独立同分布地从取值为±1/ M 对称的伯努利(Bernoulli)分布生成 Φ 。

对于任意固定的正交基 Ψ 和随机产生的测量矩阵 Φ，如果 M ≥ C ? K log(N / K)，其中，C 为某个固定常数，则 Θ = Φ Ψ 很大概率地满足 RIP。可见，当 Ψ 为正交基时，

上述测量矩阵 Φ 在某种意义上具有普适性。主要在于：随机矩阵在正交变换下具有旋转不变性，Θ = Φ Ψ 的随机性没有因为 Φ 乘上 Ψ 而改变。对于定理 5、6，存在类似的结论，感兴趣的读者可以阅读参考文献。

3.2 压缩感知的重建算法设计

作为不适定的数学反问题，压缩感知信号重建在理论上存在着无数多个可行解。但是，上文压缩感知相关定理指出，非相干性或等距约束性准则为近似精确或精确重建提供了理论上的保证。压缩感知的第三个核心问题是重建算法的设计。重建算法的设计应该遵循如下基本准则：算法应该利用尽可能少的压缩测量快速、稳定、精确或近似精确地重建原始信号。

定理 2 指出，当信号在变换域是 K -稀疏的，如果 Θ 的 2K 阶约束等距常数小于 1，那么压缩感知的信号重建可以转化为约束 0l 范数最小化的非凸最优化问题求解。但是，由于 0l 范数的高度非凸性， 0l 范数最小化是个需要组合搜索的 NP-hard 问题。当 N 很大时，不仅在数值计算上无法有效实现，而且抗噪能力很差。为此，学者们陆续提出了多种近似等价的信号重建算法。简单地说，主要包括三类方法：松弛方法、贪婪方法、非凸方法。需要指出的是，由于稀疏表示追求的同样是稀疏解，因此这三类算法也适用于信号的稀疏表示问题。

松弛方法最典型的就是基于 1l 范数最小化。例如，定理 3 显示，当 Θ 的等距约束常数满足收紧的 RIP 条件，非凸 0l 范数最小化与松弛的 1l 范数最小化是等价的。典型的 1l 范数最小化求解方法是基于线性规划的基追踪算法(BP)。但是，BP 算法在实际应用中存在两个明显的问题：一方面，当压缩测量个数 M ≥ cK ，c ≈ log(N / K)时，计算复杂度的量级为(O N 3)；另一方面， 1l 范数不能区分指示稀疏系数的位置，将导致低尺度的能量迁移到高尺度的可能，在高频区域出现震荡等伪人工现象。为了降低计算复杂度，文献陆续报道了内点法、最小角回归(LARs)、梯度投影(GPSR、软/硬迭代阈值等多种稀疏重建算法。总的来说，此类方法但是重建精度高，需要的压缩测量个数少 O(K ? log(N / K))，但是计算复杂度相对较高。

第二类方法就是贪婪方法，基本思想是通过每次迭代时进行局部最优化寻找各个非零系数。主要包括：匹配追踪(MP)[49]、正交匹配追踪(OMP)[75-77]、近似 OMP 的梯度追踪(GP)、正则正交匹配追踪(ROMP)、树形匹配追踪(TMP)、分段匹配追踪(StOMP)、子空间追踪(SP)、压缩感知匹配追踪(CoSaMP)、稀疏性自适应匹配追踪(SAMP)[81]等。贪婪方法计算复杂度相对较低，但是与松弛方法相比，需要更多的压缩测量 O(K ? log N)，重建精度相对较低。例如，当压缩测量个数满足 M ≥ cK ，c ≈ 2 ? log N 时，OMP 能够以较高的概率重构信号，计算复杂度为 O (NK 2)。因此，与 BP 算法相比，OMP 是以较多的压缩测量换取较快的计算速度。又如，Donoho 等人提出的 StOMP 以牺牲计算精度为代价进一步提高 OMP 的计算速度。在上述 9 种贪婪算法中，由于 SAMP 不需要稀疏度先验，因此在实用性和有效性上由于其它贪婪算法。

第三类方法就是非凸方法。该类方法所需的压缩测量个数、计算复杂度、信号重构精度总体上介于松弛和贪婪两类方法之间。典型的正则算法有，基于 pl (0<p<1)范数的 FOCUSS 算法和迭代重新加权算法. 最近, Ji 等人基于 Gaussian 和 Gamma、Babacan 等人基于 Laplacian 和 Gamma 分别提出了多层 Bayesian CS 信号重建算法，通过第 II 类最大似然估计法求解相关参数和稀疏系数。

4．有待研究的几个关键问题

压缩感知经过近年来的迅猛发展，已基本形成了自己的理论框架，包括基础理论、实现

方法和实际应用。但是，压缩感知理论还有很多亟待解决的问题，为此本文列出了压缩感知有待解决的几个关键问题。

4.1 基础理论层面

1. 非 1l 范数驱动的压缩感知信号重建理论。根据上文讨论， 1l 范数最小化已经成功应用于非相干性准则驱动的稀疏信号压缩感知理论和等距约束性准则驱动的可压缩信号压缩感知理论。具体表现在：当要精确或近似精确重建原始信号，基于 1l 范数最小化的相关定理对稀疏字典 Ψ 、测量矩阵 Φ 、压缩测量个数 M、信号重建精度等都有了明确的界定。基于贪婪方法和非凸方法的信号重建虽然有很多优点，但是在理论上一直存在诸多不完备之处。

突出表现在：在特定稀疏字典和测量矩阵下，大多贪婪和非凸重建算法没有给出对应信号重建精度所需的压缩测量个数 M。

2. 基于非正交稀疏字典的压缩感知信号重建理论。在等距约束性准则驱动的可压缩信号压缩感知定理中，关于稀疏字典 Ψ 和测量矩阵 Φ 仅要求两者乘积 Θ = Φ Ψ 满足 RIP。但是，测量矩阵设计部分关于压缩测量个数 M 的界定还额外附加了假设条件，即稀疏字典 Ψ 是正交基。当测量矩阵 Φ 依然通过三种方式生成，但是稀疏字典 Ψ 不再正交时，Θ = Φ Ψ 是否满足 RIP？压缩测量个数 M 的下限是否不变？由于过完备的稀疏字典才能保证表示系数具有足够的稀疏性或衰减性，进而能够在减少压缩测量的同时保证压缩感知的重建精度，所以需要设计鲁棒的测量矩阵 Φ 使之与过完备稀疏字典依然满足 RIP，同时需要重新估计压缩测量个数 M 的下限，这时所需的压缩测量定会减少。

3. 自然图像的自适应压缩感知信号重建理论。虽然基于线性投影的压缩感知理论能够直接应用于自然图像这样的复杂高维信号，但是由于没有考虑到自然图像的固有特性，诸如结构多成分性、高阶统计性等，对于自然图像压缩采样本身没有特殊的指导作用。事实上，相对于一维离散信号，自然图像的复杂性和高维性使之需要自适应的压缩采样和重建算法。

例如，基于图像多成分性的特点能够提高重建图像的峰值信噪比和视觉效果。注意到，压缩感知理论的大部分文献中，测量矩阵 Φ 都是线性的且设计好的，不需根据观测信号自适应地变化。对于自然图像，假如能够实现非线性自适应的压缩测量，压缩感知的压缩性能势必会获得大幅度的提高。目前，自然图像的自适应压缩感知信号重建理论基本空白。这项工作对压缩感知的理论推广和实际应用都具有重要意义。

4.2 实现方法层面 1. 基于学习的自然图像过完备字典设计。目前，基于构造方法的自然图像过完备字典设计具有很好的理论支撑，正则化几何方法、几何多尺度分析、基于信息论的“有效编码假设”为其奠定了坚实广阔的理论基础。但是，从国际上关于过完备字典设计的整体情况看，基于学习的自然图像过完备字典设计的工作非常少，主要在于：设计难度大、性能要求高，同时缺乏严格的理论支撑。这项工作对于稀疏字典和压缩感知都将是重要的理论完善。

2. 硬件易实现的确定性测量矩阵设计。在等距约束性准则驱动的可压缩信号压缩感知定理 3、4 中，要求稀疏字典 Ψ 和测量矩阵 Φ 的乘积 Θ = Φ Ψ 满足 RIP。其中，稀疏字典 Ψ 可以是正交的也可以是非正交的，测量矩阵 Φ 可以是随机的也可以是确定的。但是，面向应用且硬件易实现的测量矩阵应该具有以下基本特点：满足等距约束性、压缩测量个数少、采样计算成本低、存储矩阵的空间小、以及测量矩阵最好是确定性的。设计出硬件容易实现的测量矩阵和快速稳定的重建算法是将压缩感知理论推向实用的关键。

3. 噪声情形大尺度问题的快速鲁棒重建算法设计。快速稳定的信号重建算法是将压缩感知理论推向实用的关键技术之一，特别适用于纠错编码、核磁共振成像、NMR 波谱研究等大尺度问题。通常，基于 1l 最小化松弛算法的计算复杂度相对较高。因而，在非 1l 范数最小化驱动的压缩感知理论完善工作的基础上，希望能够基于稀疏性自适应的贪婪迭代和基于多层超先验建模的非凸迭代思想设计适于噪声情形大尺度问题的快速鲁棒重建算法。

5. 总结

压缩最初由美国威斯康星大学的 Mistretta 教授等人提出：能否通过减少采样数据缩短核磁共振成像的时间并且能够利用这些有限量的数据重建原始图像。对于这个问题，威斯康星大学的研究人员最初采用传统图像重建算法进行实验。结果显示[88]，重建图像的分辨率不仅低，而且边缘模糊，人工效应也很明显。完全出乎意料地，加利福尼亚科技学院的 Candès 教授及其团队仅仅基于惩罚的思想完美重建出了原始图像，同时证明了：只需随机选取信号 $M \geq 2K$ 个 Fourier 表示系数，就能唯一精确重建原始图像[3]。正是这个意外的发现触发了压缩感知理论的思想来源。压缩感知的诞生，堪称世纪之作。在这里，向奠定压缩感知基础理论的三位科学家表示致敬，他们分别是：加利福尼亚科技学院应用与计算数学系教授 Emmanuel J. Candès、加利福尼亚大学数学系教授 Terence Tao、乔治亚科技学院电子与计算机工程系的 Justin Romberg！

压缩感知理论自诞生之日就有着极强的生命力，已经对信号处理、理论数学、计算数学、计算机科学、信息论、概率论、电子工程、光学工程等诸多领域产生了重要影响。压缩感知的新颖性在于：只需远少于传统 Nyquist 采样定理所要求的采样数就能精确或高概率精确重建原始信号。采样速率不再取决于信号带宽，而在很大程度上取决于稀疏性和非相干性准则，或者稀疏性和等距约束性准则。本文围绕压缩感知的稀疏字典设计、测量矩阵设计、重建算法设计三个核心问题，对其基本理论和主要方法进行了系统阐述。同时，在基础理论和实现方法两个层面提出了压缩感知有待解决的若干理论问题与关键技术，具体包括：

(1) 非 1l 范数驱动的压缩感知信号重建理论；
(2) 基于非正交稀疏字典的压缩感知信号重建理论；
(3) 自然图像的自适应压缩感知信号重建理论；
(4) 基于学习的自然图像过完备字典设计；
(5) 硬件易实现的确定性测量矩阵设计；
(6) 噪声情形大尺度问题的快速鲁棒重建算法设计。

总之，不管是基础理论还是重建算法，都必须以把压缩感知理论推向实用为准则。至于压缩感知理论能否应用于某种实际领域，要看是否有将大量信息蕴含于少量采样数据的迫切需要，例如：提高应用系统的性能、缩短数据获取的时间、降低数据获取的耗能、减少数据的存储空间、提高数据的传输速度等等。最后，需要指出的是：压缩感知理论不是普适的。对于随机信号或者噪声信号等非结构性的信号，压缩感知理论肯定不适用。特别地，对于某些实际应用，Nyquist 采样还是首选有效的方法，因为目前压缩感知理论可能存在某些暂时不能解释也不能克服的局限性。