压缩感知理论(CS)在弱信号量子测量与蛋白质结构计算中的应用

彭绪彪^{1*}, 赵清^{1*}, 葛墨林^{1,2}

1. 北京理工大学物理学院量子技术研究中心, 北京 100081;

2. 南开大学陈省身数学所, 天津 300071

*联系人, 彭绪彪, E-mail: xubiaopeng@bit.edu.cn; 赵清, E-mail: qzhaoyuping@bit.edu.cn

收稿日期: 2020-05-03; 接受日期: 2020-06-15; 网络出版日期: 2020-08-10

国家自然科学基金(编号: 11675014)资助项目

摘要 压缩感知(Compressive Sensing, CS)理论自2006年被提出至今已15年有余, 在许多领域如单像素相机、CT扫描成像、雷达等经典系统中获得了成功应用与充分的发展, 但在量子系统和生物系统中还有很大的发展空间. 本文首先简单回顾了CS的基本知识及其应用的发展历史, 然后重点介绍了北京理工大学量子技术研究中心多年来将CS(以及它的新发展理论)应用于量子精密测量和生物系统上取得的研究成果, 具体包括极微弱光信号下的测量与成像、利用CS理论进行多体量子态层析密度矩阵的计算以及根据由CS理论发展的矩阵填充理论进行蛋白质结构计算等方面, 说明CS理论不论是在弱信号探测还是量子计算甚至在生物大分子研究方面都具有极大的应用前景和优势, 最后对CS理论的应用和发展进行总结和展望.

关键词 压缩感知理论, 单光子计数单像素成像, 极弱信号探测, 蛋白质结构计算, 量子态层析

PACS: 42.30.Va, 42.30.Wb, 76.60.Pc, 03.65.Wj

1 引言

压缩感知(Compressive Sensing, CS)理论是最近15年来信息处理理论方面的一个革命性突破, 最初由Candès, Donoho, 陶哲轩等人于2006年正式提出^[1-3], 之后该理论被广泛应用到了各个领域并取得不同程度的进展. CS理论针对稀疏或可压缩的信号, 从数学上证明了只需要利用少量的测量采样就可以恢复出较高精度的图像. 该方法突破了传统信息恢复理论中基本

定律——Nyquist-Shannon采样定理, 能够极大地减少数据的测量, 一经提出即引起了信息界的轰动并很快在图像获取领域得到了极大应用. CS理论最初是针对稀疏或可压缩信号提出的, 但它普遍适用于各种自然信号的恢复, 其根本原因在于自然信号通常具有一定结构, 即便信号在某个域不满足稀疏性, 也总可通过特定的基变换变成稀疏的, 这一过程为信号的稀疏表达. 通俗地讲, 对于未知信号的测量, 如果提前知道信号的相关结构信息(先验知识), 就可减少所需采样数

引用格式: 彭绪彪, 赵清, 葛墨林. 压缩感知理论(CS)在弱信号量子测量与蛋白质结构计算中的应用. 中国科学: 物理学 力学 天文学, 2020, 50: 090003
Peng X B, Zhao Q, Ge M L. Application of CS in weak signal quantum measurement and protein structure calculation (in Chinese). Sci Sin-Phys Mech Astron, 2020, 50: 090003, doi: [10.1360/SSPMA-2020-0158](https://doi.org/10.1360/SSPMA-2020-0158)

目. 例如, 在胸透时, 病变部分只占胸部整个面积的很小部分, 即稀疏信号. 有了这种先验知识, 就可能不必对每个部分都进行采样. CS理论要求信号是稀疏的或者在某种变换下有稀疏表达, 数学上常见的稀疏基变换包括哈尔小波变换、离散余弦变换以及傅里叶变换等多种变换. 利用稀疏基进行稀疏表达得到稀疏信号后, 便可通过对信号进行随机亚采样, 最小化零范数, 并将之转化为一个凸优化问题来重建稀疏表达的信号^[3], 最后反变换回原信号.

具体地讲, CS理论^[1]提出, 假设一个信号可以稀疏表达, 那么可以将压缩与采样合并进行, 仅需对该信号进行少量(采样率远低于Nyquist-Shannon采样定理要求)的非自适应的线性含噪测量, 便能根据相应的重构算法尽可能精确地重建信号. 对于一个稀疏信号, 传统的Nyquist-Shannon采样数要求为 N , 其中 N 要足够大, 如带宽的2倍. 而根据CS理论则只需采样数 $\log N$, 便可利用最优化方法重构原始信号. 以傅里叶变换为例, 已知信号功率谱(Ω 空间), 自然立即可通过逆变换得到原函数, 这是因为其变换矩阵是个方阵. 但如果采样数少了, 变换矩阵变成长方阵, 没有逆矩阵, 这种情况下如何唯一地恢复原函数, 就是CS要解决的问题. 由于CS理论中信号的投影测量数据量($\log N$)远远小于传统方法的数据量(N), CS极大提高了采样效率, 使得高分辨率信号的采集成为可能. 压缩传感包括3个基本步骤: 信号的稀疏表示, 编码测量和利用重构算法由测量值及投影矩阵重构原始信号. 目前, 对压缩传感基础的研究主要是从这三个步骤出发, 并且已经获得最优化的方案.

对于基于基稀疏分解的CS理论, 其重构算法可以分为三大类. (1) 贪婪算法. 利用迭代的方法不断从测量矩阵中寻找合适的支持域, 通过最小化误差函数分离出稀疏信号的成分, 逐步逼近待重构的稀疏信号矢量, 具体算法包括匹配追踪算法、正交匹配追踪算法^[4,5]及其一些相应的改进算法. (2) 凸松弛算法. 将最小化0范数的非凸问题放宽成1范数的凸优化问题, 并利用凸优化方法(如线性规划、二次规划等)进行求解. 主要包括梯度投影法、基追踪法^[6]、最小角度回归法^[7]等. 一般来说, 凸松弛算法比贪婪算法准确, 但同时计算复杂度也较高. (3) 组合类算法^[8,9]. 该算法主要应用于联合检测(Combinatorial Group Testing)方面, 如用来识别有缺陷的产品、检测病变组织等, 在这些案

例中 $x_i \neq 0$ 代表异常的元素, $x_i = 0$ 代表正常元素. 组合算法与前两种算法相比一个很重要的区别是, 它可以自由选择那些能够减少计算复杂度的测量矩阵, 如经常将测量矩阵设计成具有很少非零值的稀疏矩阵. 此外, 迭代阈值法由于较易实现, 计算量适中, 也得到了广泛的应用, 但缺点在于该方法对阈值和初值的选取比较敏感, 且难以保证解的稀疏性.

自2007年以来, 在国际上形成研究CS热潮以后, CS在理论和实际应用方面都取得了巨大进展, 被认为是一场信息理论的革命. 其特点是利用目标先验信息大大减少测量所需的样本数. 如基于CS理论的单像素相机^[10], 其灵敏度能够突破传感器单个像元灵敏度的限制, 测量维度低、可进行高通量和亚采样测量, 在传感器网络方面具有独特的优势.

过去十几年中, 由于CS理论巨大的技术应用价值, 这方面开拓性文章的被引用次数已经达到上万余次, 并在实际技术中获得了应用. 国外的研究最早可追溯到CS理论提出后的第二年, 许多知名大学(如麻省理工学院、斯坦福大学等)就纷纷开展专门的课题研究. 两年后, CS理论的研究和应用得到了许多知名公司和美国空军实验室的重视, 并专门组织了定期的CS研讨会, 邀请包括信号处理、遥感和数学等各行业专家报告, 如Coifman, McClellan, Li, De Vore等^[11]. 包括加州理工大学的Candes组、莱斯大学的Romberg组、佐治亚理工学院Davenport研究组、威斯康星大学的陈光红组等在内的各个研究组, 不但完善了CS的理论体系^[12-18], 而且将此理论迅速应用于快速MRI成像^[19]、动态心脏CT扫描新原理^[20]、数码相机新原理(单像素多次曝光)^[10,21-23]、合成孔径雷达成像^[24]、快速STORM超分辨成像^[25]、量子关联成像^[26,27]、微光成像技术^[28]等众多领域, 并不断向其他领域进一步扩大. 例如, 利用CS理论发展的动态CT扫描新技术, 于2007年就被美国通用电气公司采纳制成心脏CT扫描成像设备, 行销世界. CS的引入不但加快了成像速度, 而且还可以提高成像分辨率, 减少伪影, 已在数码相机和照相机的改进方面应用成功, 此外美国已经投资将CS理论引入雷达以及其相关领域的研究. 在国内, 北京理工大学(简称北理工)研究组与中国科学院空间科学与应用研究中心(简称中科院空间中心)研究组密切合作, 率先推动了CS理论在实际中的应用^[11]. 中国科学院上海光学精密机械研究所利用CS成像实现了光

学雷达成像^[29].

CS理论在物理领域的应用也有很大进展. 美国学者将统计物理中的参量与CS理论中信号的稀疏度和所需的采样数相对应, 加深了对CS理论的理解, 认识到要对物理系统有更准确的描述, 还需要进一步改进CS理论. 量子物理方面, CS理论的引入将会提供一套处理多体量子系统的有效方法, 有可能将量子物理的研究引入一个新的更高的领域. 目前, 结合这种处理方法, 如信号用完备的基函数(Hilbert空间)展开等思想方法的研究, 人们逐渐认识到应用CS理论处理物理问题的重要性, 尤其是对多体量子物理系统的研究和发展将有重要的推动作用. 以CS理论为基础, 在物理中广泛应用, 是创新的交叉领域. 本研究是物理与信息理论交叉的很好的例子.

在分子生物学方面, CS理论也有一些重要应用, 如利用CS方法推测蛋白质与DNA分子相互作用的结合位点^[30]及估计微观蛋白丝的特征参数^[31]; 将CS理论应用于蛋白质分子结构的多维NMR测量上以提高NMR测量精度^[32]等. 除此之外, CS还被应用于各种基因组问题中, 包括DNA微阵列技术、RNA干扰技术及宏基因组学等^[33].

2 CS在物理应用研究方面的新进展

近15年来, CS理论方面的应用领域众多, 尤其是在诸如成像相关领域(雷达、显微镜、图片处理等), 生物医疗(如基因传感、DNA微阵列等), 通信系统(如无线传感器网络、天线阵列等), 模式识别(如人脸识别、手势识别等), 声音和视频处理, 流形学习(人工智能和机器学习的重要算法), 微纳电子学和超大规模集成电路(如纳米管中氧化情况的可视化)等方面, 具体例子参考文献^[34]. 这些应用主要利用CS理论可以进行亚采样的优势, 设计合适的测量矩阵, 缩短信号采集时间, 减少测量对目标造成的损伤, 减弱噪声影响, 降低高维数据的计算复杂度, 提高恢复信号的精度和时间分辨率. 本文主要侧重于介绍CS理论在弱信号量子测量和蛋白质结构计算这些新领域中的应用. 在实用方面, 以北理工为先导, 国内取得了大范围的发展和开拓, 特别是结合机器学习之后, 获得了更进一步的发展空间. 北理工研究团队于2010年与中科院空间中心研究组密切合作将CS理论成功地应用到单光子

计数成像技术中^[35], 承担了国家仪器专项, 在几乎全部国产元件基础上, 成功实现了极弱光下的高分辨率光子计数成像, 并进一步提出在DMD双臂进行可见光和红外光的同步测量以及将该成像系统用于天文暗星观测, 并于2018年研制出具有自主知识产权的基于单光子探测的单像素相机. 与此同时, 还开展了在极微弱信号下单像素成像的相关应用的研究, 并将CS理论扩展到了红外波段的探测与成像、多体量子态层析中密度矩阵的计算以及基于NMR测量数据的蛋白质结构计算方面的研究, 以下就这些科研成果做了详细介绍.

2.1 CS算法在弱光探测与光子计数成像中的应用

2.1.1 CS极弱光探测(单光子成像)的应用

单光子成像系统的组成包括光源、DMD芯片^[36]、透镜、单光子计数器、计算机, 其实物图(中科院空间中心提供)如图1(a)所示, 其原理图如图1(b)所示.

单光子成像系统与单像素相机类似, 只是把光电探测器换成了单光子探测器, 通过对光子进行计数得到测量值, 利用极弱光探测成像. 具体地, 光线通过上方透镜在DMD芯片上成像, 经DMD芯片反射通过下方的透镜聚焦在单光子探测器上, 单光子探测器对光子进行计数. 其中DMD的状态控制原理与单像素相机相同, 测量 m 次可以得到 m 个测量. 最后根据测量值, 利用CS理论重建图像. 实验对印有“BIT”字符的胶条进行以上所述的单光子计数成像测试, 对图像的每一行随机采样600次, 共 600×70 个测量值, 每次测量用于成像的光子能量为0.0387 nJ, DMD单位像素上平均光子数约为1.1106. 最后重建出一幅分辨率为 1024×79 的图像, 结果如图2所示, 可以恢复出完整清晰的图像.

微弱信号的精密测量在许多研究及应用领域都具有重要意义, 尤其在国防科技方面. 然而, 在CS理论提出之前, 这方面的发展存在如下根本性的困难.

(1) 如何消除噪声. 任何一个系统必然存在噪声, 当信号非常微弱, 甚至比噪声小几个数量级时, 传统办法将失效, 必须发展新的方法来提取被噪声淹没的极微弱信号.

(2) 如何提高测量精度.

(3) 如何缩短测量时间.

现在随着CS理论和单光子探测技术的发展, 还有

图1 (网络版彩图)单光子成像系统实物图(a)及其原理图(b)

Figure 1 (Color online) The picture (a) and schematic diagram (b) of the single photon imaging system.

图2 (网络版彩图)极弱信号成像下的“BIT”伪彩色图. 像素 1024×79 , 用于成像的光子能量为 0.0387 nJ , DMD单位像素上的平均光子数 1.1006 count/s

Figure 2 (Color online) The pseudo-color image of the “BIT” under extremely weak signal. The resolution is 1024×79 , and the photon energy for imaging is 0.0387 nJ . There are 1.1006 count/s photons in average in each pixel on DMD.

其他相关的理论研究和实验手段, 以及相关器件性能的提高, 逐渐具备了解决这些问题的条件.

2.1.2 基于CS的单光子计数时间分辨成像光谱仪

以点扫描为主要方式的传统成像光谱方案存在诸多缺陷, 如均分光通量、测量冗余、系统稳定性差等. 针对上述问题, 可将线阵探测器和双DMD调制分别与光子计数成像结合, 从而形成两种基于CS的压缩成像光谱方案. 前者以线阵的方式实现成像光谱同时测量, 而后者以点探测的方式完成同步探测, 并且二者都不需要扫描, 从而可以减少测量的次数、维度以及时间, 使系统的灵敏度得以提高, 能够稳定而高效的运行. 此外, 在双DMD调制方案的基础上, 还可以通过在第一块DMD上加以设计, 使之能够同时测量可见波段和近红外波段的成像光谱. 这两种新方案可以避免由于扫描而导致的时间维度冲突, 故而在时间分辨测量方面有很高的应用价值, 可实现针对非周期变化长时序物体和周期变化荧光物体两类系统的时间分辨测量, 基于双DMD调制的系统又可以分别基于等时间间

隔、门宽延迟以及光子到达时间进行时间分辨测量. 值得一提的是, 自然光量子关联成像的初步实验测试可以证实时间分辨测量是可行的. 图3为双DMD调制压缩成像方案的原理图和样机.

在算法重建方面, 可以采用3种方法解决由存储和计算复杂度带来的问题, 以便在减少测量次数和降低硬件资源的情况下仍然可以进行实时的大图抗噪重建. 这3种方法分别是基于稀疏矩阵的CS改进算法、基于Hadamard矩阵和排列矩阵的CS恢复算法以及基于小波树和稀疏表达的自适应压缩成像. 尤其值得注意的是, 第二种方法极大地减少了存储空间和计算时间, 降低了硬件要求, 从而使准实时的大图重建得以实现, 目前已经开始进行工程化代码编写, 有望使该算法硬件化.

通过实验验证和理论分析对上述两种基于CS的压缩成像光谱方案进行比较, 可以看出在第一种方案中, 光谱分辨能力和空间分辨能力是相互制约的, 其中的光子线阵探测器目前正在自主研发期, 产品性能还不够完善, 然而由于能够通过多通道计数得到光谱, 理论上获得的测量结果相对会更加准确. 对于第二种方案, 重建的成像光谱质量与两块DMD的测量数有关, 只有在选取合理的情况下, 才会重建出质量很好的成像光谱. 该方案以单点模式获得成像光谱, 其光通量、灵敏度以及量子效率都会更高, 还可以使测量维度和测量数大大减少, 同时还覆盖了更广的光谱范围. 然而由于无法忽视DMD掩膜的加载时间, 第二种方案的测量时间将会比第一种长, 从而对重建光谱的准确性造成影响, 或者对某一波长的二维光谱图像质

图3 (网络版彩图)极弱光探测与成像原理图(a)与样机(b)

Figure 3 (Color online) The schematic diagram (a) and prototype equipment (b) for the measurement and imaging system of extremely weak signals.

量造成影响. 因此, 两种方案各有利弊, 应当针对不同的情况进行合理抉择. 由于二者都具备超灵敏特性, 使其在生物显微、天文观测、材料检测等领域拥有非常理想的应用前景.

2.1.3 单像素光子计数成像的堆积效应研究

本小节主要介绍了堆积效应对单像素光子计数成像的影响. 堆积效应普遍存在于计数探测(或事件探测)系统中, 如光子计数、粒子计数、辐射计数等, 在时间分辨光子计数系统中, 堆积效应通常会造成光子时间分布曲线向短光子到达时间方向漂移, 而在此单像素激光雷达中, 我们研究发现堆积效应还会带来特殊的“正负像”现象, 并对此进行分析和提出改进方案. 在此单像素激光雷达中, 照明光源非量子态, 探测方式使用光子时间分辨计数形式, 具有量子探测的特点.

光子计数成像通常用于极弱光探测和激光雷达成像探测, 其本质是利用信号光子触发APD的雪崩击穿效应, 输出离散的电信号, 把光信号以数字化的形式记录, 避免了热噪声的影响^[37]. 典型的光子计数系统如图4所示.

由于光电探测器存在堆积效应, 在基于CS的单像素光子计数红外激光雷达系统中发现, 由于堆积效应会导致“正负像”的情况出现, 即在雷达系统探测场景中一个深度位置的切片重建图像里, 会出现其他深度位置处物体形貌灰度反转的图像. 如图5(a)所示, 重建“N”型物体的图像中出现了一个灰度反转“T”. 同样地, 如图5(b)所示, 在重建的“T”型物体的图像中有一个负像“N”, 并且只有光子计数率接近探测器的饱和计数率时才会出现负像.

图4 光子计数系统示意图

Figure 4 The schematic diagram on photon counting system.

为了避免探测器发生堆积效应, 单像素光子计数成像系统应在非常低的光子计数模式下工作, 这种方法效率低, 探测时间长, 与单像素成像高光子通量的特点相冲突. 所以可利用Coates^[38]提出的方法来纠正强信号情况下由堆积效应导致的误差. 红外激光雷达光子计数成像系统, 针对出现由堆积效应引起的“正负像”的情况进行修正. 修正第 i 次测量中“N”型物体反射的光子计数为

$$\hat{y}_N^{(i)} = -\ln\left(1 - y_N^{(i)} / (N_P - y_P^{(i)})\right), \quad (1)$$

式中, $y_P^{(i)}$ 是第 i 次测量模式中未校正的“T”型物体反射的光子计数, N_P 是激发脉冲在一个测量矩阵的测量区间内的强度.

图6给出了将校正算法应用于图5(a)的实验数据中的实验结果. 对于两个在较大激光脉冲重复频率探测的实验中, 重建的“N”型物体在图5(a)中可以观察到由堆积效应引起的“正负像”现象; 通过应用校正算法进行修正后, 完全消除了图5(a)中的负像.

图5 不同深度处重建的物体. (a) 深度位置为1.98 m处重建的物体“N”; (b) 2.13 m深度重建的物体“T”. 图片来源文献[39]

Figure 5 The reconstructed images of objects at different depths. (a) The reconstructed object “N” at depth of 1.98 m; (b) the reconstructed object “T” at depth of 2.13 m. Reprinted from ref. [39] with permission.

图6 将校正算法应用于图5(a)所得修正图像. (a) 光子计数率为655 kHz时的修正图; (b) 光子计数率为536 kHz时的修正图. 图片来源文献[39]

Figure 6 The refined images after the refinement algorithm on Figure 5(a). (a) The refined image at photon counting rate 655 kHz; (b) the refined image at photon counting rate 536 kHz. Reprinted from ref. [39] with permission.

同样针对单像素红外激光雷达系统, 测量矩阵的强度调制也会影响探测器的计数率. 可采用稀疏采样的调制方式来校正堆积效应的影响. 通过将稀疏采样的方式应用到单像素红外激光雷达系统中可得到如图7的结果, 其中测量矩阵大小为64像素×64像素. 可以观察到“T”型的负像随着“1”的个数减少而消失. 每组的测量次数为2000次, 小于重建图像的总像素数的1/2,

曝光时间为200 s. CS中测量矩阵的设计应满足受限等距的特性, 矩阵应具有均值为0的特性. 稀疏矩阵包含二值0和1, 平均值高于0. 然而, 由于矩阵中1的数目很少, 稀疏矩阵的平均值相当接近0. 例如, 用于实现图7(d)的稀疏矩阵的平均值为0.012. 因此这种稀疏矩阵测量方法可以有效地对图像进行调制, 并且有利于重建图像.

2.1.4 基于离散正交Krawtchouk矩变换的实时重构光子成像

Krawtchouk矩属于离散正交矩, 其定义域能够与图像的坐标空间保持一致, 计算的过程中避免了空间变换, 离散化过程不会产生计算方面的误差, 具有优良的压缩能量的特性, 可以非常好地表达出图像的信息, 并且其噪声敏感度很小^[40], 还可以从图像的任意感兴趣区域提取局部特征. 借助Krawtchouk矩变换实现了单像素成像, 进一步与单光子探测技术结合, 提出了一种基于离散正交Krawtchouk矩变换的单光子成像算法(Krawtchouk-SPI). Krawtchouk-SPI要求率先获取目标物体的Krawtchouk矩, 随后通过反变换Krawtchouk矩重新建立目标物体图像.

首先对两组128像素×128像素的图像进行仿真, 在不同阶数下分别计算图像的Krawtchouk矩, 分别对应5%, 10%, 25%, 50%和100%的采样率, 仿真参数 $p_1=p_2=0.5$. 仿真结果如图8所示, 即便5%的采样率也能重建出图像的主要特征, 且在采样率大于25%之后重建图像的质量明显改善. 该方法能够有效地实现目标的单像素成像, 从而使利用较少的测量次数重建低阶Krawtchouk矩图像的策略成为可能. 重构图像的质量随着采样率的增加而提高, 这是因为随着Krawtchouk矩的阶数增加, 其重构误差减小.

图7 稀疏采样完成的64像素×64像素图像重建. (a) 每种模式的“1”数目为1000; (b) 每种模式的“1”数目为500; (c) 每种模式的“1”数目为100; (d) 每种模式中的“1”数为50. 图片来源文献[39]

Figure 7 The reconstructed 64×64 images from sparse sampling. (a) The number of “1” in each model is 1000; (b) the number of “1” in each model is 500; (c) the number of “1” in each model is 100; (d) the number of “1” in each model is 50. Reprinted from ref. [39] with permission.

图8 (网络版彩图)在不同的采样率辣椒和摄影师的Krawtchouk矩和重建图像(KM: Krawtchouk矩; RI: 重建图像). (a) 辣椒图篇; (b) 摄影师图片. 图片来源文献[41]

Figure 8 (Color online) The Krawtchouk moments and the reconstructed images of the Peppers and Cameraman at different sampling ratio (KM: Krawtchouk moment; RI: reconstructed image). (a) The image of Peppers; (b) the image of Cameraman. Reprinted from ref. [41] with permission.

Krawtchouk矩的仿真参数 p_1 和 p_2 控制局部区域的提取, 参数 p_1 用于控制垂直移动感兴趣区域, 而参数 p_2 用于控制水平移动感兴趣区域. 仿真选取的原始测试图像如图9(a)所示, 图像由4个英文字母组成, 128像素×128像素. 参数 p_1 与 p_2 分别设置为(0.5,0.5), (0.2,0.2), (0.2,0.8), (0.9,0.1)和(0.9,0.9). 采样过程使用较低的(29+29)阶, 对应采样率为5%, 仿真结果如图9(b)–(f)所示. 参数 p_1 与 p_2 设置为(0.5,0.5)指感兴趣区域位于图像的中心区域, 而4个字母分散在图像的4个角上, 图像的中心区域没有有效目标, 因此图9(b)中的图像质量并不令人满意. 如果参数 $p_1=p_2=0.2$, 则感兴趣的区域位于图像的左上角. 如图9(c)所示, Krawtchouk-SPI算法可以成功地提取左上角的字母K, 重建图像质量良好. 类似地, 当 p_1 与 p_2 设置为(0.2,0.8), (0.9,0.1)和(0.9,0.9)时, 感兴趣的区域分别位于图像的右上角、左下角和右下角, 相应的仿真结果如图9(d)–(f)所示, 其中位于3个对应角的字母M, S和I分别被清晰地提取出来. 仿真实验表明, 当目标集中分布区域可以作为一项先验知识时, 通过调整参数 p_1 与 p_2 的值, 可以直接对感兴趣目标区域进行重点采样, 在仅执行5%采样率的低阶采样情况下, 就可以非常清晰地重建目标. 由于Krawtchouk矩具有特殊的特征提取能力, Krawtchouk-SPI算法特别适

图9 参数 p_1 和 p_2 不同取值情况下感兴趣区域重建结果. (a) 原始图; (b) $p_1=0.5, p_2=0.5$; (c) $p_1=0.2, p_2=0.2$; (d) $p_1=0.2, p_2=0.8$; (e) $p_1=0.9, p_2=0.1$; (f) $p_1=0.9, p_2=0.9$. 图片来源文献[41]

Figure 9 The reconstructed results of the interested region at different values of p_1 and p_2 . (a) The original image; (b) $p_1=0.5, p_2=0.5$; (c) $p_1=0.2, p_2=0.2$; (d) $p_1=0.2, p_2=0.8$; (e) $p_1=0.9, p_2=0.1$; (f) $p_1=0.9, p_2=0.9$. Reprinted from ref. [41] with permission.

用于目标物体在场景中分散分布的情况.

基于Krawtchouk-SPI算法的工作原理, 可在实验室搭建用于验证Krawtchouk-SPI算法局部区域成像的光学实验平台, 图10为实验原理图. 实验中用左下角位置包含一个英文字母K的胶片作为成像目标, 为了证

图10 (网络版彩图) Krawtchouk-SPI局部区域光子成像实验原理图. 图片来源文献[41]

Figure 10 (Color online) The schematic diagram of the Krawtchouk-SPI local region photon imaging experiment. Reprinted from ref. [41] with permission.

明Krawtchouk-SPI在极低采样率下的成像效果, 采用三项递推算法, 基于阶数(13+13), (18+18), (23+23)和(29+29), 分别计算生成了4组128像素×128像素的灰度Krawtchouk基模式, 分别对应于1%, 2%, 3%和5%的采样率, 将参数 p_1 和 p_2 设置为(0.9,0.1). 对灰度Krawtchouk基模进行差分和二值量化处理, 量化位数 $b=6$.

计算4个采样率下测量的强度, 以便产生物体的Krawtchouk矩, 利用其反变换重建目标图像, 重建的图像规模是128像素×128像素, 重建结果如图11所示. 光子成像实验结果表明, 在基于DMD的二值差分调制下, Krawtchouk-SPI算法可以有效地实现目标的局部区域成像. 即使在采样率仅为1%的情况下, 利用该方法仍然能够清晰地成像, 从而证明了Krawtchouk-SPI算

法独特的特征提取能力, 尤其适用于场景中物体只占据局部区域的情况. 此研究成果在国防应用方面具有巨大的潜在应用背景, 如可以应用于雷达全方位搜索确定目标大致范围后的定向扫描目标等实际应用场景.

2.1.5 并行红外CS在远距离成像上的应用

红外并行CS成像系统原理在理论上是 4×4 个DMD上的像素被1个CCD像素采集, 为了确定实际实验时出现的是均匀成像还是非均匀成像, 可用大小为 40×40 的黑白小方块拼接成的大区域进行模拟, 其中黑色区域像素值为0, 白色区域像素值为1. 如图12所示. 图12(a)为模拟的 4×4 对应 1×1 的完美成像结果, 图12(b)为非均匀的不完美成像结果, 从两图对比可以发现, 图12(b)更为模糊, 且每一小方块中出现了非0非1的像素值.

在实际实验中, 将上述模拟中的黑白大区域加载到实际光路的DMD上, 并用CCD采集图像的实验光路是非完美成像, 即非均匀成像. 针对非均匀成像, 利用卷积的方法来进行解决与优化.

对于上述所采用的并行压缩成像理论进行数字模拟, 分别模拟了在加入2%的噪声的情况下, 使用0-1随机掩膜与不同 H 矩阵的哈达玛矩阵掩膜, 比较不同掩膜对恢复原始物体的影响.

如图13所示, 图13(a)为模拟的原始物体, 图13(b)为加载0-1矩阵的随机掩膜、 H 为 4×4 的矩阵、下采样系数为16、噪声为2%的情况下, 全采样恢复的结果. 其中, 全采样是指采样次数等于下采样系数, 即采样16次. 图13(c)为加载0-1矩阵的随机掩膜、 H 为 6×6 的矩阵、下采样系数为16、噪声为2%的情况下, 全采样

图11 重建图像的单像素压缩成像实验. (a) 傅里叶SPI法重建图像; (b) KM-SPI方法重建图像. 图片来源文献[41]

Figure 11 The single pixel compress imaging experiment. (a) Images reconstructed with Fourier-SPI method; (b) images reconstructed with KM-SPI. Reprinted from ref. [41] with permission.

图12 红外并行CS成像模拟结果. (a) 完美成像结果; (b) 不完美成像结果

Figure 12 The simulation result of the parallel compressed sensing with infrared light. (a) Perfect imaging results; (b) imperfect imaging results.

图13 0-1随机掩膜与不同H矩阵在2%的噪声下对恢复结果的影响. (a) 原始物体; (b) H为4×4的矩阵; (c) H为6×6的矩阵

Figure 13 The effects of the random 0, 1 masks and different H matrices on the recovered results under 2% noise. (a) The original image; (b) H matrix is 4×4; (c) H matrix is 6×6.

恢复的结果. 通过对比可以发现, 图13(b)更加清晰, 且相对误差较小, 这是由于恢复图像时, 使用了匹配的测量矩阵与矩阵 H , 从而减小了误差.

如图14所示, 图14(a)为模拟的原始物体, 图14(b)为加载-1-1哈达玛矩阵的掩膜、 H 为2×2的矩阵、下采样系数为4、噪声为2%的情况下, 全采样恢复的结果. 图14(c)为加载-1-1哈达玛矩阵的掩膜、 H 为6×6的矩阵、下采样系数为4、噪声为2%的情况下, 全采样恢复的结果. 同样, 当 H 矩阵与测量矩阵匹配时, 还原效果最好.

最后进行简单物体的计算机模拟, 如图15所示. 图15(a)为物体原图, 图15(b)为模拟加载-1-1的哈达玛矩阵、测量率为2倍、噪声为2%、 H 为4×4的矩阵的还原结果, 图15(c)为 H 为6×6的矩阵的还原结果.

通过上述模拟试验可以得出, 在并行CS成像中, 只有确定了系统的非均匀系数矩阵 H , 才能进一步得

到相匹配的测量矩阵 A_H , 从而得到清晰且误差较小的还原结果.

在数字物体的非均匀并行压缩成像实验中, 可将数字虚拟物体加载在DMD上, 进行实际光路的虚拟物体模拟实验. 加载10张大小为400×400的随机0-1矩阵作为掩膜, 将CCD上收集的10张100×100的图像拉成一列作为 y 值, 在实际测试中, H 矩阵选择4×4以及5×5两种作为比较, 结果如图16所示, 图16(a)为 H 矩阵为4×4的还原结果, 图16(b)为 H 矩阵为5×5的还原结果, 因此, 可选择 H 为5×5的矩阵来作为实验光路的非均匀矩阵.

2.2 CS算法在量子态层析中的应用

2.2.1 基于CS-Phaselift的量子态层析和保真度估计

基于CS优化模型, 将其向量恢复推广至矩阵恢复, 可以得到如下密度矩阵恢复凸优化模型^[42]:

图14 -1-1随机掩膜与不同 $\mathbf{H}$ 矩阵在2%的噪声下对恢复结果的影响. (a) 原始物体; (b) $\mathbf{H}$ 为 2×2 的矩阵; (c) $\mathbf{H}$ 为 6×6 的矩阵

Figure 14 The effects of the random -1-1 masks and different $\mathbf{H}$ matrices on the recovered result under 2% noise. (a) The original image; (b) $\mathbf{H}$ matrix is 2×2 ; (c) $\mathbf{H}$ matrix is 6×6 .

图15 简单物体的计算机模拟结果. (a) 原始物体; (b) $\mathbf{H}$ 为 4×4 的矩阵; (c) $\mathbf{H}$ 为 6×6 的矩阵

Figure 15 The simulation results on the simple objects. (a) The original image; (b) $\mathbf{H}$ matrix is 4×4 ; (c) $\mathbf{H}$ matrix is 6×6 .

$$\text{Minimize } \|A(\rho) - b\|_1, \text{ subject to } \rho \geq 0, \quad (2)$$

式中, b_{ij} 是相对频率, 对应该测量基下测量量子态的实际概率, 理想情况下, 当量子态全同副本的个数接近无限时, 相对频率等于理论计算概率, 即平均值; 在实验中, 当副本数量足够多时, 相对频率接近平均值. 根据实验符合计数统计, 可以估算这个相对频率. $\mathbf{M}_{i,j}$ 是测量矩阵, i 是测量装置, j 是相对应的测量基. 在具体的求解过程中, 一阶、二阶梯度下降法可以求得密度矩阵.

实验中设计出了高度纠缠的六比特薛定谔猫态^[43], 这个态 ρ_{exp} 可以用一个 64×64 的密度矩阵表示, 其与理想量子态 ψ 的保真度可以被定义为^[44]

$$F_{\text{exp}}(|\psi\rangle) = \langle \psi | \rho_{\text{exp}} | \psi \rangle. \quad (3)$$

数值模拟^[44]表明, 估算子式(3)可以高精度恢复未

知量子态的密度矩阵, 且对于任意秩为1的密度矩阵均有效. 如果我们只考虑无噪音情形, 测量矩阵从 S 集合中随机选出, 当测量次数很大时, 可以近似认为每一个测量基底的测量平均值等于相对频率. 从重建结果可以看出, 恢复的密度矩阵的保真度为1, 均方差为0. 假定从集合 S 中提取不同正算符值测量算符(POVM)的概率均为 $1/3136$, 那么当采样数 m 大于380时即可高精度恢复出密度矩阵.

接下来需要对更加一般的量子态进行数值模拟测试, 这里考虑秩为1的随机密度矩阵, 借助估算子式(3)恢复了密度矩阵, 计算了估算密度矩阵的保真度, 如图17所示. 当 m 足够大时, Phaselift就能够保证成功恢复, 这就表明采用估计子来重建密度矩阵是合理的.

此外实验收集所有POVM基底的光子的偶然符合

图16 (网络版彩图)实际光路下的虚拟物体模拟结果. (a) H 为 4×4 的矩阵; (b) H 为 5×5 的矩阵

Figure 16 (Color online) The simulation result of the dummy objects in real light path. (a) H matrix is 4×4 ; (b) H matrix is 5×5 .

计数的统计结果, 并将其转换为斯托克斯参数, 输入到估算子式(3)即可以高精度恢复出密度矩阵, 如图18所示^[45], 与 $|HHHHHH\rangle|HHHHHH\rangle$ 和 $|VVVVVV\rangle|VVVVVV\rangle$ 分别为密度矩阵主对角线上的两个数值最大的元素, 实部分别是0.3816(对应 $|HHHHHH\rangle|HHHHHH\rangle$)和0.3402(对应 $|VVVVVV\rangle|VVVVVV\rangle$). 与 $|HHHHHH\rangle|VVVVVV\rangle$ 和 $|VVVVVV\rangle|HHHHHH\rangle$ 分别对应于密度矩阵反对角线上的两个最大的元素, 实部均为0.2544.

量子信息中薛定谔猫态、W态^[46]以及Cluster态^[47]等因为其特殊的纠缠结构、纠缠特性, 在实际的应用中受到广泛关注. 结合Phase-lift模型, 以及Nestrov算法^[48], 通过模拟重建不同比特数的不同量子态^[49](如图19所示), 其中包括GHZ态、W态、Cluster态, 发现Nestrov对于多比特量子系统, 重建时精度明显提高,

图17 保真度随着样本数变化. 图片来源文献^[45]

Figure 17 The relation between fidelity and the number of sampling. Reprinted from ref. ^[45] with permission.

图18 (网络版彩图)密度矩阵实部的重建, 其中采样率为 $(49 \times 26)/(26)^2 \approx 0.7656$. 图片来源文献^[45]

Figure 18 (Color online) The reconstruction of the real part in density matrix, where the sampling ratio is $(49 \times 26)/(26)^2 \approx 0.7656$. Reprinted from ref. ^[45] with permission.

而且相比于传统的量子态层析方法, 该方法不仅在精度上有明显提升, 而且重建时间大大减少, 而且值得一提的是, 这种方法只需要部分测量数据即可高精度恢复密度矩阵, 这在实际应用中可减少各种资源消耗.

最后通过将Nestrov算法与CVX^[50]等其他密度矩阵恢复算法进行比较, 如图20所示, 可以发现, 对于多比特系统, Nestrov重建精度明显较高.

2.2.2 基于CS-PhaseLift的多步优化量子态层析

自适应量子态层析的流程如下. 首先将制备好的量子态全同副本分为两部分, 其中一部分用来估算密度矩阵, 另外一部分用来自适应修正估算的密度矩阵. 也就是说, 粗略估算一个未知量子态的密度矩阵, 相当于将一部分量子态全同副本平均分配在所有的测量基底之上, 其余部分只分别在部分测量基的之上, 这样通过测算每个测量基底的相对频率, 借助这个相对频率, 我们可以估算出密度矩阵.

本小节基于自适应量子态层析模型^[51,52], 提出一种改进的自适应机制, 在每一步测量之前选出最优测量基底, 同时借助相位提升估算子重建密度矩阵. 具体操作流程如下: 首先数值模拟泡利测量过程, 具体地对于 n 比特系统, 测量基底的总数为 3^n , $f(1)_{\mu,\nu}$ 表示第一轮测量中第 μ 个基底上的第 ν 个基对应的相对频率,

图19 不同比特量子态密度重建结果. (a)–(c) 重建误差随采样率的变化, 从左到右分别为GHZ态、Cluster态和W态. (d)–(f) 重建时间随采样率的变化, 从左到右分别为GHZ态、Cluster态和W态. 图片来源文献[49]

Figure 19 The reconstructions of the quantum state densities for different qubits. (a)–(c) The MSE values with sampling rate increasing, where the quantum states are GHZ state, Cluster state and W-state from left to right. (d)–(f) The reconstruction time cost with sampling rate increasing, where the quantum states are GHZ state, Cluster state and W-state from left to right. Reprinted from ref. [49] with permission.

(1)表示第一轮测量, μ 是一个整数, 介于 $1-3^n$ 之间. 要得到这个相对频率, 可以通过在相应的POVM上测量 $N \cdot R/3n$ 个 ρ_T 态复本获得. 借助这个相对频率, 先估算得到一个大致的密度矩阵 ρ_{E0} . 接着引入新的自适应机制, 采用最优测量基底进而将所剩的量子态全副本合理地分配至下一轮测量过程, 同样地, 根据不同基底上的光子统计分别相应得到相对频率 $f(2)_{\mu,v}$. 这样将 $f(1)_{\mu,v}$ 和 $f(2)_{\mu,v}$ 线性组合可得到新的相对频率, 即修正的相对频率, 接下来借助这个修正的相对频率, 用相位提升的方法可以高精度恢复出密度矩阵.

最后分别恢复重建了薛定谔猫态(SC)(如图21(a)所示)和随机量子态的密度矩阵, 其中随机量子态满足迹等于1, 重建结果如图21(b)所示.

图22是Phase-lift重建密度矩阵的均方误差(MSE). 红线是传统的量子态层析重建结果, 黑色虚线是自适应模型重建密度矩阵的结果. 可以发现自适应模型重建的密度矩阵的MSE明显低于传统的量子态层析重建的结果.

近年来, 在量子力学领域, 人们分别在量子纠缠、光子统计、CS等方面的研究取得了极大的成果, 以上

研究中的相位提升模型和自适应量子态层析模型将分别在量子信息领域有效解决量子态纠缠验证, 并且在量子态层析过程中大幅度减少测量次数以及量子态全副本数, 极大地减少了量子态层析的物力成本和时间成本, 最重要的是, 重建精度也得到明显提高. 我们希望以上方法可以应用到量子信息应用领域, 推动量子信息领域的发展.

2.3 基于CS的蛋白质结构计算与预测的应用

确定一个生物大分子的结构对了解其物理、化学性质和生物功能非常重要, 这也是结构生物学所关注的重点, 尤其是对于蛋白质、核酸这样的生物分子, 因此计算蛋白质的三维结构成为当前结构生物学和生物信息学中广泛研究的课题之一. 蛋白质的化学结构研究已经取得杰出的成就, 将其结构理论发展到量子层次是必然的趋势. 但蛋白测试取样量通常很大, 耗资耗时. CS理论, 包括矩阵填充(Matrix Completion, MC)等数学工具可以在结构分析中发挥其特殊的效用. 本节作为CS理论在蛋白质结构分析方面的初步尝试, 主要集中在利用CS理论发展的矩阵填充方法从核

图20 不同算法重建密度矩阵结果. (a), (b) 重建误差随量子比特数的变化, (a)为GHZ态, (b)为随机量子态. (c), (d) 重建时间随量子比特数的变化, (c)为GHZ态, (d)为随机量子态. (a)中插图为量子数为3, 4, 5, 6时Nesterov算法在采样率为0.3和0.6时的重建误差, (c)中插图为对应的重建时间. 图片来源文献[49]

Figure 20 The reconstruction of the density matrix with different algorithms. (a), (b) The MSE values with number of qubits increasing, (a) is the GHZ state and (b) is the random states. (c), (d) The reconstruction time cost with number of qubits increasing, (c) is the GHZ state and (d) is the random states. The inset of panel (a) shows the MSE of 3, 4, 5, 6 qubits for Nesterov algorithm at sampling rates 0.3 and 0.6, respectively. The inset of panel (c) is the corresponding time cost. Reprinted from ref. [49] with permission.

图21 (网络版彩图)两比特薛定谔猫态的密度矩阵(a)和随机产生的密度矩阵(b). 图片来源文献[52]

Figure 21 (Color online) The density matrix of the 2-qubit Schrödinger cat state (a) and the randomly generated density matrix (b). Reprinted from ref. [52] with permission.

磁共振测量值重建蛋白质结构的应用上.

迄今为止, X射线晶体衍射(XRD)法和核磁共振

(NMR)光谱法是两种最主要的确定蛋白质结构的方法.

对于一些不容易结晶的蛋白质, NMR方法具有天

图22 (网络版彩图)未知量子态不同数量的复本测量下对应的均方误差(MSE). 图片来源文献[52]

Figure 22 (Color online) The mean squared error (MSE) values at different number of replicas for the unknown quantum states. Reprinted from ref. [52] with permission.

然的优势, 因为NMR实验可以直接在溶液中进行, 而且这种状态也更加接近蛋白质能够行使其功能的自然状态。

蛋白质结构的NMR测量方法始于20世纪80年代,

由于存在核奥弗豪泽效应(Nuclear Overhauser Effect, NOE), 核磁共振光谱学提供了一个空间上邻近原子之间的距离测量网络, 而蛋白质结构模型可由此建立. 典型的基于NMR测定蛋白质结构的流程通常包括以下几个步骤: 核磁共振波谱峰的提取、化学位移赋值(谱赋值)、几何约束赋值、结构计算, 如图23所示。

其中从核磁共振数据中获得的部分距离约束计算分子构象主要有两种方法: 一种方法是利用模拟退火方法最优化含距离约束的蛋白质能量函数, 从而得到分子结构, 目前蛋白质数据库(PDB)中大多数NMR蛋白用的都是这种方法, 常见计算软件包括ARIA, CYANA和UNIO等; 另一种方法是使用距离几何方法解析大分子构象, 软件如EMBED, DISGEO, DGSOL.

近年来矩阵填充作为一种新兴的技术在模型降阶、模式识别和机器学习等领域引起了众多研究者的关注. 这项技术是CS的一个分支, 根据矩阵元素之间的依赖关系寻找与已知项一致的最低秩矩阵算法。

在蛋白质结构的问题中所用的矩阵是欧氏距离矩阵, 对于给定的 N 个欧氏空间中的点, 其距离矩阵是一

图23 (网络版彩图)蛋白质NMR结构确定方法流程图

Figure 23 (Color online) The workflow of the NMR protein structure determination.

个以点距作为元素的 $N \times N$ 的对称矩阵. 图24给出了以6个点为例的距离矩阵. 图24(a)是由a, b, c, d, e, f共6个点组成的点集, 图24(b)是点集所对应的距离矩阵, 矩阵元代表点集中相应的行指标与列指标所对应的点之间的距离, 不难发现距离矩阵是一个对称矩阵, 而且其对角元素为零. 距离矩阵还有一个特点是要满足三角不等式约束, 即在点集中任意三个点满足: $d_{ij} \leq d_{ik} + d_{kj}, \forall i, j, k$.

由于欧式距离矩阵的低秩性, 可以通过矩阵填充技术解析蛋白质结构: 首先基于NMR实验数据的距离信息, 蛋白质中一些先验的距离信息(如蛋白质中共价键和共面原子的间距)以及三角不等式估算方法建立初始矩阵; 然后通过使用矩阵填充算法将这个不完整的初始矩阵恢复成低秩完整的距离矩阵, 从而获得整个的蛋白质结构. 关于矩阵填充应用于蛋白结构计算的具体工作介绍如下.

2.3.1 APG算法计算蛋白质结构

本小节使用一个典型的NMR蛋白质结构模型来建立初始距离矩阵, 详见文献[53]中的算法2, 其中假设所有小于5 Å的H-H距离都已由NMR实验得到, 然后使用APG算法实现蛋白质原子距离矩阵的填充, 并计算三维结构. 该方法在8种蛋白质上进行了测试.

为了验证算法的可行性, 首先考虑初始距离估计值完全正确的理想情况. 结果如表1和图25所示, 其中RMSD表示重建结构与目标结构之间的相似度. 从中可以看出, 重建蛋白与目标蛋白完美重合, 且蛋白质中原子的数量越多, 距离矩阵的维数越大, 恢复结构所需的采样率越低, 但同时程序的运行时间将随着矩阵维数的增加而增加.

由于实验的NOE距离约束具有一定的误差, 可采用三角不等式将误差引入初始距离矩阵, 并限制在距离的10%以内. 这种情况下的计算结果如表2和图26所

图24 (网络版彩图)由6个点组成的点集和对应的距离矩阵. (a) 点集; (b) 距离矩阵

Figure 24 (Color online) The set and the corresponding distance matrix with six points. (a) The set points; (b) the distance matrix.

表1 根据理想距离矩阵数据进行结构预测

Table 1 The structure prediction results based on the ideal distances

蛋白质	描述	原子数	残基数	采样率	RMSD (Å)	时间 (s)
2M5Z	Antimicrobial protein	598	44	0.0673	1.2×10^{-3}	7.36
1B4R	PKD domain1 from human polycystein-1	1114	80	0.0449	9.6×10^{-4}	10.77
1G6J	Ubiquitin	1228	76	0.0407	9.2×10^{-4}	15.33
2KT6	Outer membrane usher protein papC	1283	85	0.0390	2.1×10^{-3}	12.72
1CN7	Yeast ribosomal protein L30	1648	104	0.0303	1.3×10^{-3}	22.86
2KTS	Heat shock protein hs1J	1784	117	0.0280	3.4×10^{-4}	27.45
2K49	UPF0339 protein SO3888	1823	118	0.0247	6.7×10^{-4}	29.27
2KTE	Bacillus subtilis	2380	152	0.0210	1.0×10^{-3}	38.16

图25 (网络版彩图)原始PDB结构(红色)与理想距离矩阵数据预测结构(蓝色)的重叠情况. (a) 2M5Z; (b) 1B4R; (c) 1G6J; (d) 2KT6; (e) 1CN7; (f) 2KTS; (g) 2K49; (h) 2KTE. 图片来源文献[53]
Figure 25 (Color online) The superimposition between the original PDB structure (red) and the reconstructed structure (blue) from the ideal distance data. (a) 2M5Z; (b) 1B4R; (c) 1G6J; (d) 2KT6; (e) 1CN7; (f) 2KTS; (g) 2K49; (h) 2KTE. Reprinted from ref. [53] with permission.

示. 结果显示, 重构蛋白质的RMSD值仍都小于 $2.0 \text{ \AA}$, 符合高分辨率结构的标准. 因此, 利用APG矩阵填充算法根据部分距离信息重建蛋白质结构是可行的.

2.3.2 基于ScaledASD算法蛋白质结构计算

近年来, 一种新的矩阵填充算法——交替最速下降(Alternating Steepest Descent, ASD)算法及其变体 ScaledASD算法被成功地应用到图像处理中并得到了很好的结果, 因此可将该算法应用在蛋白质结构计算中来. 与2.3.1小节中利用已知结构模型建立初始矩阵不同, 本节利用实际中能够测到的真实的NMR实验数据建立初始矩阵, 并对重构出的蛋白质结构进行后处

理精修, 以得到分辨率更高的结构.

后处理过程如下. 首先采用BFGS (Broyden-Fletcher-Goldfarb-Shanno)算法来最小化距离违规, 其中包括3种不同距离约束条件: 共价键和共面原子间距为等式约束; 估算值的三角不等式和实验数据的NOE距离为上限约束, 范德瓦尔斯半径决定的距离为下限约束. 然后, 由于蛋白质和氨基酸有特定的手性偏好, 而距离矩阵无法反映其手性, 需要对重建的结构进行手性修正. 最后将蛋白质放入溶液环境并进行能量最小化得到最终蛋白质结构.

选取10个不同大小、不同拓扑结构的蛋白进行重建, 结果如表3和图27所示^[54]. 表3显示与PDB中的蛋白

表2 用三角形约束估算核磁共振结构的计算结果

Table 2 The reconstructed structures from the triangle constraints estimation

蛋白质	描述	原子数	残基数	采样率	RMSD ($\text{\AA}$)	时间 (s)
2M5Z	Antimicrobial protein	598	44	0.0673	0.826	17.36
1B4R	PKD domain1 from human polycystein-1	1114	80	0.0449	1.074	41.63
1G6J	Ubiquitin	1228	76	0.0407	0.747	56.55
2KT6	Outer membrane usher protein papC	1283	85	0.0390	0.818	66.35
1CN7	Yeast ribosomal protein L30	1648	104	0.0303	0.786	80.23
2KTS	Heat shock protein hs1J	1784	117	0.0280	1.668	83.72
2K49	UPF0339 protein SO3888	1823	118	0.0247	0.887	90.59
2KTE	Bacillus subtilis	2380	152	0.0210	1.320	116.83

图26 (网络版彩图)原始PDB结构(红色)与三角形约束预测结构(蓝色)的重叠情况. (a) 2M5Z; (b) 1B4R; (c) 1G6J; (d) 2KT6; (e) 1CN7; (f) 2KTS; (g) 2K49; (h) 2KTE. 图片来源文献[53]

Figure 26 (Color online) The superimposition of the original PDB structures (red) and structures estimated from triangle constraints (blue). (a) 2M5Z; (b) 1B4R; (c) 1G6J; (d) 2KT6; (e) 1CN7; (f) 2KTS; (g) 2K49; (h) 2KTE. Reprinted from ref. [53] with permission.

质结构比较, 利用矩阵填充重建的蛋白质结构的RMSD在2 Å左右, 说明该重建总体上是准确的, 另外一个评价蛋白质结构相似性的指标TM-score大部分在0.8以上, 同样也说明重建方法的可行性. 图27则展示了相应的蛋白结构, 可以看到, 计算结构和参考结构之间有很高的重合. 通过进一步分析了重建蛋白的主链原子在Ramachandran图上的分布(表4, 图28(a)和(b)), 可以发现重建蛋白的原子大部分都处在Ramachandran图的可利区域, 说明了重建蛋白结构的合理性. 最后通过分析矩阵填充方法重建的蛋白质的二级结构, 发现与PDB的参考结构具有高度的相似性(如图28(c)所示), 反映了计算结构的稳定性. 至此, 通过一个小的测试集, 说明了MC算法可用于真实NMR测量的蛋白质结构重建, 且重建效果与PDB的传统重建方法相当.

2.3.3 基于Riemannian方法的蛋白质结构计算

基于黎曼理论的矩阵填充算法由Jawanpuria等人^[55]提出, 本小节将之应用于NMR测量的蛋白质结构重建. 基本计算流程与基于ScaledASD算法蛋白质结构计算类似. 不同的是, 本小节采用更大的测试集和更合理的测试方法来验证评价MC方法在蛋白质结构重建中的有效性: 随机选择43对兼有NMR结构和X射线晶体结构的蛋白质, 其大小从5到23 kDa不等, 用Riemannian方法(包括两种算法Cg和Tr)进行结构重建.

表3 10种蛋白质在well-defined区域Ca RMSD值和TM-score值

Table 3 The values of Ca RMSD and TM-score in well-defined region for 10 proteins

蛋白质	RMSD_well-defined (Å)	TM-score
1G6J	0.99±0.13	0.87±0.02
1B4R	1.57±0.16	0.84±0.02
2K62	1.81±0.14	0.86±0.02
1CN7	1.55±0.19	0.78±0.02
2K49	1.55±0.14	0.85±0.02
2L3O	2.08±0.23	0.70±0.02
2GJY	1.64±0.12	0.80±0.01
2K7H	1.96±0.20	0.90±0.02
2YT0	0.92±0.09	0.81±0.01
2L7B	2.25±0.19	0.89±0.01

同时, 基于相同的约束条件使用最优秀的传统重建算法之一——CYANA程序重建结构, 通过比较两种方法得到的结构相对于与X射线晶体结构的相似度, 验证了基于Riemannian方法的蛋白质结构计算的有效性和优势.

对相同的蛋白质, 尽管解析条件不同, 核磁共振结构与X射线晶体所决定的结构基本相似, 基于X射线结构通常是更可靠的设定, 因此可通过比较NMR重建结构与其X射线对应结构的相似性来评估重建方法的有

图27 (网络版彩图)原始PDB结构(红色)与三角形约束预测结构(蓝色)的重叠情况. (a) 1G6J; (b) 1B4R; (c) 2K62; (d) 1CN7; (e) 2K49; (f) 2L3O; (g) 2GJY; (h) 2K7H, (i) 2YT0; (j) 2L7B. 图片来源文献[54]

Figure 27 (Color online) The superimposition of original PDB structure (red) and the structures from triangle constraints estimation (blue). (a) 1G6J; (b) 1B4R; (c) 2K62; (d) 1CN7; (e) 2K49; (f) 2L3O; (g) 2GJY; (h) 2K7H, (i) 2YT0; (j) 2L7B. Reprinted from ref. [54] with permission.

效性. 以X射线结构为参照, 分别计算Riemannian和CYANA方法重建结构的TM-score/GDT-TS值. TM-score差值($\text{TM-score}_{\text{Rieman}} - \text{TM-score}_{\text{CYANA}}$)和累积分布如图29所示. 可以看出, 与大多数蛋白质的CYANA结构相比, 黎曼方法提高了TM-score评分. 在图29(b)中,

重建后的结构TM-score均在0.54–0.96的范围内, 说明它们在SCOP/CATH折叠中具有相同的性能. Tr算法的曲线总是低于其他算法, 表明Tr算法重构了更多高TM-score的蛋白质结构. 类似地, GDT-TS结果如图29(c)和(d)所示, 可以看出Tr算法的性能优于CYANA算法和Cg算法.

同样, 以X射线结构为标准, 分别计算Riemannian和CYANA方法重建结构的RMSD值, 并以其差值($\text{RMSD}_{\text{Rieman}} - \text{RMSD}_{\text{CYANA}}$)来表征Riemannian方法相对CYANA在重建结构上的优劣性. 蛋白质主链在蛋白质well-defined区域 and 所有区域的RMSD差异及其对应的累积分布图如图30所示. 总的来说, 与CYANA方法相比, 黎曼方法重构的蛋白质更接近于对应的X射线晶体学的结构. 图31显示, 4种蛋白(1BVM, 1TOF, 2HFI, 2K5P)用Tr算法重建的结构(洋红色)、CYANA重建的结构(绿色)和相应的X射线晶体结构(蓝色)重合, 可以看到它们的结构都有很好的重叠. 总之, 通过测量相对于X射线结构的相似度(特别是所有区域的主干RMSD), 证明了黎曼MC算法(特别是其Tr算法)可以与CYANA相比, 而且对于某些特定的蛋白质有时甚至更优.

后通过分析Riemannian方法在某些蛋白质上表现较差的可能原因, 发现Riemannian方法的重建效果与已知约束的比例之间存在显著的正相关关系: 已知约束数越多, Riemannian方法的重建效果越好. 这是因为Riemannian方法(即MC算法)主要是基于几何的, 而CYANA是基于能量的. 最后得出的结论是, 当存在更

表4 10种蛋白质结构在Ramachandran图中的有利区域和不利区域的百分比(%)

Table 4 The percentages (%) in the favored and disallowed regions in Ramachandran plots for the 10 proteins

蛋白质	Favorable (reference)	Favorable (calculation)	Disallowed (reference)	Disallowed (calculation)
1G6J	92.23±2.40	94.66±2.42	0.63±0.91	1.08±1.36
1B4R	83.14±1.87	81.41±4.19	3.91±1.14	6.73±2.69
2K62	87.69±2.46	88.25±1.82	0.93±0.56	2.64±1.15
1CN7	79.51±2.56	85.49±3.39	4.02±1.71	3.14±1.51
2K49	94.57±1.66	92.68±2.26	0.30±0.42	1.69±1.16
2L3O	74.18±2.02	70.57±2.80	8.86±1.85	13.07±1.81
2GJY	75.99±3.12	79.57±2.86	7.29±1.96	6.31±1.95
2K7H	97.29±1.08	91.71±2.80	1.10±0.24	1.55±1.07
2YT0	85.55±2.76	79.55±2.77	1.15±0.56	6.82±1.94
2L7B	75.69±1.58	73.24±1.72	6.41±1.19	9.19±1.76

图28 (网络版彩图)以1CN7为例, 计算结构与参考结构Ramachandran图及二级结构区域对比. (a) 重建结构的Ramachandran图; (b) 参考结构的Ramachandran图; (c) 二级结构对比. 图片来源文献[54]

Figure 28 (Color online) The comparison between the calculated structure and reference structure in Ramachandran plot and secondary structures, taking 1CN7 as an example. (a) The Ramachandran plot for the reconstructed structure; (b) the Ramachandran plot for the reference structure; (c) the comparison of secondary structure. Reprinted from ref. [54] with permission.

多的已知距离信息时, Riemannian方法可能表现得更好, 这可以通过改进核磁共振测量和化学位移分配技术来实现.

通过比较Riemannian方法与CYANA方法在同时具有NMR和X射线晶体结构的蛋白质测试集上的不同指标证明, 黎曼方法, 特别是Tr算法, 通常具有与CYANA方法相当的性能, 对于一些距离约束信息较多的蛋白质, Tr算法的结果更好. 因此, 用基于Riemannian方法的矩阵填充算法从NMR实验数据重建蛋白质结构是可行的.

2.3.4 基于矩阵填充技术从核磁共振测量数据中恢复晶体结构中的缺失区域

本小节提出了一种借助矩阵填充理论恢复蛋白质晶体结构模型中缺失部分的方法. 所用到的数据是晶体结构的已知(未缺失)部分和与未知(缺失)部分相关的NMR实验数据. 本小节在PDB数据库中挑选了10个蛋白进行了测试, 并将计算结果跟当前比较流行的

MODELLER算法的计算结果进行比较. 这种新方法在用分子动力学研究蛋白折叠过程时特别有用, 可用来建立用于模拟的合理的初始结构.

该方法说明如下. 首先建立一个初始距离矩阵. 由于所要完成的是把蛋白质晶体结构中的缺失区域恢复完整, 因此晶体结构中的未缺失结构可以作为已知信息添加到初始矩阵中. 但距离矩阵中对应的缺失结构的位置仍出现大片空白, 即缺失部分对应的整行整列元素都是空白的, 这显然不符合矩阵填充中均匀采样的条件. 因此, 在剩下的空白区域中, 可将蛋白质结构特性相关的键长、键角、共面性等距离信息添加到距离矩阵中. 除此之外, 也将跟缺失结构相关的NMR实验数据加入到距离矩阵中. 为了达到距离矩阵均匀采样的条件, 将剩下的未知距离用三角不等式估算出来, 然后进行随机采样得到初始矩阵.

然后用ScaledASD算法填充距离矩阵并进行后处理操作以减小误差. 需要注意, 此处的后处理过程与之前的蛋白质NMR结构计算里的后处理过程略有不同:

图29 (网络版彩图) Riemannian结构和CYANA结构的TM-score/GDT-TS数值之差图和累积分布图. (a) Riemannian结构和CYANA结构的TM-score之差; (b) Riemannian结构和CYANA结构的TM-score累积分布图; (c) Riemannian结构和CYANA结构的GDT-TS数值之差; (d) Riemannian结构和CYANA结构的GDT-TS累积分布图. 图片来源文献[56]

Figure 29 (Color online) The plots of differences and accumulative distributions on TM-score/GDT-TS between the Riemannian structures and CYANA structures. (a) The TM-score differences between the Riemannian structures and CYANA structures; (b) the accumulative distributions of the TM-score for Riemannian structures and CYANA structures; (c) the GDT-TS differences between the Riemannian structures and CYANA structures; (d) the accumulative distributions of the GDT-TS for Riemannian structures and CYANA structures. Reprinted from ref. [56] with permission.

用BFGS算法进行距离约束的修正时, 与之前的计算相比, 此处所用约束限制还需将晶体结构中所有已知部分的距离考虑在内; 能量优化时需将对原晶体结构的已知部分进行约束. 最后获得一个合理完整的蛋白质结构.

为检验方法的可行性, 从PDB库中挑选了10个既有晶体结构又有NMR结构的蛋白, 见表5. 在这10个晶体结构中, 既有末端的缺失情况也有在中间区域的缺失情况, 而且缺失部分的氨基酸残基链也长短不一, 这样的测试样本有利于算法的普适性检验.

最后用Ca原子的RMSD作为结构相似性的判定. 在对比过程中, 以当前补全缺失结构的流行方法——MODELLER计算的结构作为对照组. 每一次的对比中, 这两种方法各取5个模型进行比较. 分别计算缺失结构的RMSD (见表5第4列)和已知结构的RMSD (见

表5第5列). 可以看到, 矩阵填充的方法和MODELLER在已知结构区域的RMSD都非常小, 而且两种方法的结果也非常一致, 这说明这两种方法在恢复缺失结构的过程中对已知的结构基本上都没有很大的改动, 这也是所期望的. 在缺失结构的RMSD比较中, 以NMR结构的第一个模型作为参考对象, 计算氨基酸长度大于3的缺失结构的RMSD. 从表5数据中可以看出, 用MC方法所计算出来的结构的RMSD比MODELLER的更小(见图32), 也就是说, MC方法计算出来的缺失结构部分更加接近NMR结构. 因此, 利用MC补全缺失结构的方法兼顾了晶体结构和NMR结构的特点, 比MODELLER的补全方法更准确. 两种方法恢复出来的三维结构见图33. 图33中, 天蓝色部分为晶体中已知结构部分, 红色代表缺失部分的NMR参考结构, 其他色彩为计算出来的缺失部分的不同模型, 奇数列的结构

图30 (网络版彩图) Riemannian结构和CYANA结构的RMSD数值之差和累积分布. (a) Riemannian结构和CYANA结构well-defined区域的RMSD数值之差; (b) Riemannian结构和CYANA结构well-defined区域的RMSD累积分布图; (c) Riemannian结构和CYANA结构所有区域的RMSD数值之差; (d) Riemannian结构和CYANA结构所有区域的RMSD累积分布图. 图片来源文献[56]

Figure 30 (Color online) The RMSD differences between the Riemannian structures and CYANA structures. (a) The RMSD difference in well-defined region between the Riemannian structures and CYANA structures; (b) the accumulative distributions of the RMSD in well-defined region for Riemannian structures and CYANA structures; (c) the RMSD difference in all region between the Riemannian structures and CYANA structures; (d) the accumulative distributions of the RMSD in all region for Riemannian structures and CYANA structures. Reprinted from ref. [56] with permission.

图31 (网络版彩图)蛋白质的X射线晶体结构(蓝色), CYANA结构(绿色)和Tr算法重建结构(紫红色)的重叠图. (a) 1BVM; (b) 1TOF; (c) 2HFI; (d) 2K5P. 图片来源文献[56]

Figure 31 (Color online) The superimposition between the X-ray crystal structures (blue), CYANA structures (green) and reconstructed structures from Tr algorithm (magenta). (a) 1BVM; (b) 1TOF; (c) 2HFI; (d) 2K5P. Reprinted from ref. [56] with permission.

由基于MC算法的方法所恢复, 偶数列的结构由MODELLER方法所恢复. 从图中可以直观地看到MC方法计算的结构比MODELLER计算的结构总体上更加地接近参考结构.

最后, 利用MPscore这个指标来分析结构的合理性, 其值越小说明结构越合理. 从计算结果(图34)可以看出^[57], 基于MC的方法重建的结构所对应的MPscore明显小于MODLLER的结果, 进一步证实了矩阵填充方法在此应用上的有效性.

对于CS理论在蛋白质结构计算上的应用, 除了以上所讨论的根据NMR实验测量的距离信息重建三维结构外, 还可以有以下两方面的后续研究方向: (1) 由于CS理论可应用于分子动力学模拟^[58], 可以考虑利用CS理论改进蛋白质动力学模拟的效率和所需时间; (2) 蛋白质的孤子模型表明, 仅用200个特定参数的孤子模型就可较高精度地覆盖92%的蛋白质数据库(PDB)中

表5 用本文方法和MODELLER方法恢复结构的RMSD和MPscore的比较

Table 5 The comparison of RMSD and MPscore between the structures from our method and the MODELLER method

蛋白质	方法	缺失结构	缺失结构RMSD (Å)	已知结构RMSD (Å)	MPscore
1CTR	本文方法	75–80	2.03±0.43	0.23±0.02	2.54±0.90
	MODELLER		3.08±0.14	0.10±0.01	4.07±0.02
2RKK	本文方法	65–75	3.19±0.18	0.18±0.05	1.55±0.14
	MODELLER		3.44±0.21	0.57±0.01	2.66±0.05
2SAK	本文方法	1–15	4.21±0.35	0.14±0.01	1.61±0.12
	MODELLER		5.47±0.75	0.03±0.01	1.79±0.24
1GSV	本文方法	1–3	0.16±0.04	0.13±0.01	1.17±0.48
	MODELLER		0.45±0.30	0.05±0.05	1.66±0.19
5WLB	本文方法	31–37	1.42±0.14	0.15±0.01	1.14±0.09
	MODELLER		1.88±0.48	0.05±0.01	1.33±0.10
2VRW	本文方法	178–184	1.46±0.24	0.11±0.01	1.22±0.21
	MODELLER		2.26±0.83	0.01±0.01	1.74±0.07
3SD6	本文方法	67–69	0.85±0.09	0.01±0.00	1.29±0.06
	MODELLER		0.86±0.14	0.04±0.01	2.22±0.16
1EY4	本文方法	1–5	1.60±0.22	0.14±0.01	1.75±0.04
	MODELLER		1.73±0.13	0.05±0.02	2.96±0.26
	本文方法	142–149	2.31±0.18	–	–
	MODELLER		2.75±0.52	–	–
2ZR4	本文方法	1–4	1.01±0.27	0.16±0.01	1.46±0.07
	MODELLER		0.98±0.38	0.17±0.01	2.02±0.10
	本文方法	40–44	0.81±0.17	–	–
	MODELLER		1.25±0.18	–	–
1AVS	本文方法	1–6	1.25±0.24	0.18±0.01	1.64±0.09
	MODELLER		3.32±0.72	0.08±0.05	3.03±0.64
	本文方法	88–90	0.38±0.09	–	–
	MODELLER		0.12±0.06	–	–

的蛋白质片段^[59], 因此蛋白质结构在孤子模型的表象上是稀疏的, 可以考虑将CS理论与蛋白质孤子模型结合对蛋白质结构进行模拟和计算, 从而提高蛋白质结构分析和模拟计算的效率.

3 总结和展望

压缩传感理论在经典物理等基础研究和CT成像等各种技术研究领域中已经被广泛应用且展现出明显的优势, 本文通过回顾北理工研究组在CS理论应用于量子精密测量和生物系统上取得的相关研究成果, 展示了CS理论在量子物理和生命科学中巨大的应用

前景.

(1) CS理论在单光子计数成像方面的成功应用, 让人们看到了它已经在量子物理研究中崭露头角并且会在处理量子系统问题方面有着光明的前景, 尤其会对多体量子物理系统的研究和发展将有重要的推动作用. 另外, 在此基础上发展的超灵敏时间分辨成像光谱仪, 可以实现成像和光谱的同时测量, 突破了探测仪器本身的灵敏度极限, 可为包括生命科学、材料、化学、分析仪器、天文、遥感在内的相关领域提供有力的科研分析手段. 如可用来探测蛋白质中某些氨基酸残基的弱荧光, 进而实时观测蛋白质微环境的动力学变化, 在分子生物学上具有重要意义.

图32 (网络版彩图)用矩阵填充方法和MODELLER方法恢复出来的结构的RMSD比较图. 蓝色代表本工作所用方法的结果, 红色代表MODELLER方法的结果. 图片来源文献[57]

Figure 32 (Color online) The RMSD comparison between the structures from matrix completion method and MODELLER method. Blue is from our method while red is from MODELLER. Reprinted from ref. [57] with permission.

(2) 在量子信息处理中, 其中关键步骤之一便是量子态层析, 即利用大量量子态拷贝测量重建未知的量子态的技术. 通过将CS理论应用于此技术中, 并结合相位提升和反馈机制, 有效解决了多比特量子态纠缠判定和量子态层析中测量算符数量需求问题, 减少了层析所需的量子态复本, 在提高重建精度的同时缩短了量子态重建时间. 以上方法有望应用到量子信息应用领域, 推动量子信息领域的发展.

(3) 在生命科学中, 蛋白质结构测定是从微观角度理解生命的基础. 本文基于蛋白质的NMR实验结果, 通过将CS理论的另一个分支——矩阵填充应用于结构模型的重建, 发现矩阵填充的重建效果与目前最优秀的蛋白质重建程序之一——CYANA不相伯仲, 而且在NMR测量数据较多时甚至可超越CYANA, 说明CS

图33 (网络版彩图) MC方法与MODELLER方法恢复出来的结构与NMR参考结构的重叠图. 其中1, 3列为MC方法重建结构, 2, 4列为MODELLER方法重建结构. 图片来源文献[57]

Figure 33 (Color online) The superimpositions for the structures recovered from MC and MODELLER methods with the corresponding NMR reference structure, respectively. The columns 1, 3 are from MC method, while columns 2, 4 are from MODELLER method. Reprinted from ref. [57] with permission.

图34 (网络版彩图)用矩阵填充方法和MODELLER方法恢复出来的结构的MPscore比较. 蓝色代表本工作所用方法的结果, 红色代表MODELLER方法的结果, 绿色为PDB中NMR结构的MPscore. 图片来源文献[57]

Figure 34 (Color online) The MPscore comparison between the structures from matrix completion method and MODELLER method. Blue is from our method, red is from MODELLER and green is the NMR structures from the PDB. Reprinted from ref. [57] with permission.

理论在蛋白质结构测定方面具有良好的应用前景. 同时针对蛋白质晶体结构中的缺失部分, 利用矩阵填充

算法补齐的缺失结构比传统方法更合理, 同样显示出矩阵填充技术在蛋白质结构计算中的优势.

值得注意的是, 以上仅是CS理论在量子 and 生命科学上应用的一部分, 除此之外, CS理论在量子通信和多体物理中的应用也越来越受到重视, 例如最近苏刚团队^[60]通过CS理论与张量网络和机器学习结合, 发展出了张量网络的CS理论; 另外, 随着深度学习和强化学习等人工智能(AI)技术的发展, CS理论与AI技术相结合也显示出巨大的优势, 如DeepMind提出的深度压缩感知框架显著提高了信号恢复的性能与速度^[61], 而斯坦福AI研究院根据统计压缩感知技术中设计出非确定性自编码器, 提高了编码器性能^[62]. 可以预见, CS理论与AI技术的结合是CS理论应用发展的一个重要趋势, 并将在各个领域显示出其独特的优势和应用前景.

总之, CS理论作为一个普适的针对稀疏信号的测量与重建的理论, 不仅能够给经典物理的基础研究和成像技术带来革命性发展, 在量子物理、多体物理和生命科学方面同样具有广阔的应用前景和独特的优势, 而且随着新的AI技术的发展会有更多潜在的应用.

致谢 感谢中国科学院国家空间科学中心翟光杰研究员、刘雪峰和姚旭日副研究员等仪器专项中所有合作者的帮助和支持, 感谢姚旭日副研究员对文章内容的讨论. 感谢张伟、韦仙、柴绪丹和靳晓鹏四位博士生为本文提供素材和整理资料.

参考文献

- 1 Candès E J, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inform Theor*, 2006, 52: 489–509
- 2 Donoho D L. Compressed sensing. *IEEE Trans Inform Theor*, 2006, 52: 1289–1306
- 3 Candès E J. Compressive sampling. In: *Proceedings of the International Congress of Mathematicians*. Madrid, 2006. 1433–1452
- 4 Davenport M A, Wakin M B. Analysis of orthogonal matching pursuit using the restricted isometry property. *IEEE Trans Inform Theor*, 2010, 56: 4395–4401
- 5 Mallat S G, Zhang S G. Matching pursuits with time-frequency dictionaries. *IEEE Trans Signal Process*, 1993, 41: 3397–3415
- 6 Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. *SIAM J Sci Comput*, 1998, 20: 33–61
- 7 Tibshirani R. Regression shrinkage and selection via the lasso. *J R Statist Soc B*, 1996, 58: 267–288
- 8 Du D Z, Hwang F K. *Combinatorial Group Testing and Its Applications*. Singapore: World Scientific, 2000
- 9 Erlich Y, Shental N, Amir A, et al. Compressed sensing approach for high throughput carrier screen. In: *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Monticello, 2009
- 10 Takhar D, Laska J N, Wakin M B, et al. A new compressive imaging camera architecture using optical-domain compression. In: *Proceedings of SPIE 6065, Computational Imaging IV*. San Jose, 2006. 606509

- 11 Ge M L, Zhao Q. The application of CS theory in measuring the extremely weak signals (in Chinese). In: Chinese Academy of Engineering, ed. *Developments of the Measuring Technology and Innovations of the Precision Instruments*. Beijing: Higher Education Press, 2016. 74–94 [葛墨林, 赵清. CS理论应用于极微弱信号测量. 见: 中国工程院, 主编. 测量技术发展与精密仪器创新. 北京: 高等教育出版社, 2016. 74–94]
- 12 Willett R M, Duarte M F, Davenport M A, et al. Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection. *IEEE Signal Process Mag*, 2014, 31: 116–126
- 13 Willett R, Raginsky M. Poisson compressed sensing. In: *Defense Applications of Signal Processing*. Coolum, 2011
- 14 Arguello H, Arce G R. Code aperture design for compressive spectral imaging. In: *Proceedings of the 18th European Signal Processing Conference*. Aalborg, 2010
- 15 Arguello H, Arce G R. Code aperture optimization for spectrally agile compressive imaging. *J Opt Soc Am A*, 2011, 28: 2400–2413
- 16 Wu Y H, Mirza I O, Arce G R, et al. Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Opt Lett*, 2011, 36: 2692–2694
- 17 Duarte M F, Baraniuk R G. Spectral compressive sensing. *Appl Comput Harmonic Anal*, 2013, 35: 111–129
- 18 Arce G R, Brady D J, Carin L, et al. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Process Mag*, 2014, 31: 105–115
- 19 Lustig M, Donoho D, Pauly J M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med*, 2007, 58: 1182–1195
- 20 Chen G H, Tang J, Leng S. Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Med Phys*, 2008, 35: 660–663
- 21 Romberg J. Imaging via compressive sampling. *IEEE Signal Process Mag*, 2008, 25: 14–20
- 22 Duarte M F, Davenport M A, Takhar D, et al. Single-pixel imaging via compressive sampling. *IEEE Signal Process Mag*, 2008, 25: 83–91
- 23 Chan W L, Charan K, Takhar D, et al. A single-pixel terahertz imaging system based on compressed sensing. *Appl Phys Lett*, 2008, 93: 121105
- 24 Baraniuk R G, Steeghs P. Compressive Radar Imaging. In: *Proceedings of the 2007 IEEE Radar Conference*. Boston, 2007. 128–133
- 25 Zhu L, Zhang W, Elnatan D, et al. Faster STORM using compressed sensing. *Nat Methods*, 2012, 9: 721–723
- 26 Howland G A, Howell J C. Efficient high-dimensional entanglement imaging with a compressive-sensing double-pixel camera. *Phys Rev X*, 2013, 3: 011013
- 27 Gross D, Liu Y K, Flammia S T, et al. Quantum state tomography via compressed sensing. *Phys Rev Lett*, 2010, 105: 150401
- 28 Marim M, Angelini E, Olivo-Marin J C, et al. Off-axis compressed holographic microscopy in low-light conditions. *Opt Lett*, 2011, 36: 79–81
- 29 Zhao C Q, Gong W L, Chen M L, et al. Ghost imaging lidar via sparsity constraints. *Appl Phys Lett*, 2012, 101: 141123
- 30 AlQuraishi M, McAdams H H. Direct inference of protein-DNA interactions using compressed sensing methods. *Proc Natl Acad Sci*, 2011, 108: 14819–14824
- 31 Mahrooghy M, Yarahmadian S, Menon V, et al. The use of compressive sensing and peak detection in the reconstruction of microtubules length time series in the process of dynamic instability. *Comput Biol Med*, 2015, 65: 25–33
- 32 Bostock M J, Holland D J, Nietlispach D. Improving resolution in multidimensional NMR using random quadrature detection with compressed sensing reconstruction. *J Biomol NMR*, 2017, 68: 67–77
- 33 Rao A, Deepthi P, Renumadhavi C H, et al. Compressed sensing methods for DNA microarrays, RNA interference, and metagenomics. *J Comput Biol*, 2015, 22: 145–158
- 34 Rani M, Dhok S B, Deshmukh R B. A systematic review of compressive sensing: Concepts, implementations and applications. *IEEE Access*, 2018, 6: 4875–4894
- 35 Yu W K, Liu X F, Yao X R, et al. Single photon counting imaging system via compressive sensing. arXiv: [1202.5866v1](https://arxiv.org/abs/1202.5866v1)
- 36 Yoder L A, Duncan W M, Koontz E M, et al. DLP (TM) technology: Applications in optical networking. In: *Proceedings of Spie the International Society for Optical Engineering*. San Diego, 2001. 4457: 54–61
- 37 Howland G A, Dixon P B, Howell J C. Photon-counting compressive sensing laser radar for 3D imaging. *Appl Opt*, 2011, 50: 5917
- 38 Coates P B. The correction for photon “pile-up” in the measurement of radiative lifetimes. *J Phys E: Sci Instrum*, 1968, 1: 878–879
- 39 Liu S, Yao X R, Liu X F, et al. Pile-up effect in an infrared single-pixel compressive LiDAR system. *Opt Express*, 2019, 27: 22138–22146
- 40 Yap P T, Paramesran R, Ong S H. Image analysis by Krawtchouk moments. *IEEE Trans Image Process*, 2003, 12: 1367–1377
- 41 Chen Y, Yao X R, Zhao Q, et al. Single-pixel compressive imaging based on the transformation of discrete orthogonal Krawtchouk moments. *Opt*

- Express, 2019, 27: 29838–29853
- 42 Eldar Y C, Mendelson S. Phase retrieval: Stability and recovery guarantees. *Appl Comput Harmonic Anal*, 2014, 36: 473–494
- 43 Gao W B, Lu C Y, Yao X C, et al. Experimental demonstration of a hyper-entangled ten-qubit Schrödinger cat state. *Nat Phys*, 2010, 6: 331–335
- 44 Flammia S T, Liu Y K. Direct fidelity estimation from few Pauli measurements. *Phys Rev Lett*, 2011, 106: 230501
- 45 Lu Y P, Liu H, Zhao Q. Quantum state tomography and fidelity estimation via Phaselift. *Ann Phys*, 2015, 360: 161–179
- 46 Somma R D, Chiaverini J, Berkeland D J. Lower bounds for the fidelity of entangled-state preparation. *Phys Rev A*, 2006, 74: 052302
- 47 Briegel H J, Raussendorf R. Persistent entanglement in arrays of interacting particles. *Phys Rev Lett*, 2001, 86: 910–913
- 48 Becker S, Bobin J, Candès E J. NESTA: A fast and accurate first-order method for sparse recovery. *SIAM J Imag Sci*, 2011, 4: 1–39
- 49 Chai X D, Lu Y P, Zhang A N, et al. Tomographic reconstruction of multiqubit states by phase lift with the Nesterov algorithm. *Phys Rev A*, 2019, 99: 042321
- 50 Boyd S, Vandenberghe L. *Convex Optimization*. New York: Cambridge University, 2004. 127
- 51 Lu Y P, Zhao Q. Minimum copies of Schrödinger’s cat state in the multi-photon system. *Sci Rep*, 2016, 6: 32057
- 52 Lu Y P, Zhao Q. Enhancing multi-step quantum state tomography by PhaseLift. *Ann Phys*, 2017, 384: 198–210
- 53 Li Z C, Li Y, Lei Q, et al. Protein structure estimation from NMR data by matrix completion. *Eur Biophys J*, 2017, 46: 525–532
- 54 Li Z C, Li S J, Wei X, et al. Scaled alternating steepest descent algorithm applied for protein structure determination from nuclear magnetic resonance data. *J Comput Biol*, 2019, 26: 1020–1029
- 55 Jawanpuria P, Mishra B. A unified framework for structured low-rank matrix learning. In: *Proceedings of the 35th International Conference on Machine Learning*, PMLR 80. Stockholm, 2018. 2254–2263
- 56 Wei X, Li Z C, Li S J, et al. Protein structure determination using a Riemannian approach. *FEBS Lett*, 2020, 594: 1036–1051
- 57 Li Z C, Li S J, Wei X, et al. Recovering the missing regions in crystal structures from the nuclear magnetic resonance measurement data using matrix completion method. *J Comput Biol*, 2020, 27: 709–717
- 58 Andrade X, Sanders J N, Aspuru-Guzik A. Application of compressed sensing to the simulation of atomic systems. *Proc Natl Acad Sci*, 2012, 109: 13928–13933
- 59 Krokhotin A, Niemi A J, Peng X. Soliton concepts and protein structure. *Phys Rev E*, 2012, 85: 031906
- 60 Ran S, Sun Z, Fei S, et al. Quantum compressed sensing with unsupervised tensor network machine learning. arXiv: 1907.10290, 2019
- 61 Wu Y, Rosca M, Lillicrap T, et al. Deep compressed sensing. In: *Proceedings of the International Conference on Machine Learning*. Long Beach, 2019. 6850–6860
- 62 Grover A, Ermon S. Uncertainty autoencoders: Learning compressed representations via variational information maximization. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics*. Naha, 2019. 2514–2524

Application of CS in weak signal quantum measurement and protein structure calculation

PENG XuBiao^{1*}, ZHAO Qing^{1*} & GE MoLin^{1,2}

¹ Center for Quantum Technology Research, School of Physics, Beijing Institute of Technology, Beijing 100081, China;

² Chern Institute of Mathematics, Nankai University, Tianjin 300071, China

It has been more than 15 years since the compressed sensing (CS) theory was put forward in 2006. The CS theory has been successfully applied and developed in many classical fields such as the single-pixel camera, CT imaging and radar imaging system, but there is still much space for its applications in quantum and biological system. In this paper, the basic knowledge of CS theory and its application/development history are briefly introduced. Then a review is given on a series of achievements in the CS theory applied in quantum precision measurement and biological system, mainly obtained by the Center for Quantum Technology Research in Beijing Institute of Technology (BIT). The review covers several different topics including the measurement and imaging in extremely weak signals, the density matrix calculation for multi-body quantum state tomography and the protein structure reconstruction from matrix completion theory, which is developed based on the CS theory. With the above reviews, we show that the CS theory has great potential for further application, not only in weak signal detection, but also in quantum calculation and even the research on bio-macromolecules. In the end, the summary and outlook are given on the future developments and applications of the CS theory.

compressed sensing theory, single photon counting single pixel imaging, extremely weak signal detection, protein structure calculation, quantum state tomography

PACS: 42.30.Va, 42.30.Wb, 76.60.Pc, 03.65.Wj

doi: [10.1360/SSPMA-2020-0158](https://doi.org/10.1360/SSPMA-2020-0158)