



申请代码	A0405
接收部门	
收件日期	
接收编号	1247013023



国家自然科学基金 申 请 书

(2024 版)

资助类别：	面上项目		
亚类说明：			
附注说明：			
项目名称：	大规模线性逆问题的分裂-融合-共识算法及在数据价值计算中的应用		
申 请 人：	赵云彬	BRID：	07663.00.05927
办公电话：	0755-84273747		
依托单位：	深圳市大数据研究院		
通讯地址：	广东省深圳市龙岗区龙翔大道2001号涂辉龙楼RB302		
邮政编码：	518172	单位电话：	0755-84273867
电子邮箱：	yunbinzhao@sribd.cn		

国家自然科学基金委员会



基本信息

申请人信息	姓名	赵云彬	性别	男	出生年月	1967年04月	民族	汉族
	学位	博士	职称	教授				
	是否在站博士后	否		电子邮箱	yunbinzhao@sribd.cn			
	办公电话	0755-84273747		国别或地区	英国			
	申请人类别	依托单位全职						
	工作单位	深圳市大数据研究院/深圳国际工业与应用数学中心						
	主要研究领域	运筹学，数学优化，科学计算，信号处理，数据科学						
依托单位信息	名称	深圳市大数据研究院						
	联系人	冯政		电子邮箱	fengzheng@sribd.cn			
	电话	0755-84273867		网站地址	www.sribd.cn			
合作研究单位信息	单位名称							
	北京工业大学 北京科技大学							
项目基本信息	项目名称	大规模线性逆问题的分裂-融合-共识算法及在数据价值计算中的应用						
	英文名称	Split-Fusion-Consensus Algorithms for Large-Scale Linear Inverse Problems and Application to Data Valuation						
	资助类别	面上项目				亚类说明		
	附注说明							
	申请代码	A0405. 连续优化				A0607. 数据科学中的数学理论与方法		
	研究期限	2025年01月01日 -- 2028年12月31日				研究方向：稀疏优化		
	申请直接费用	51.0000万元						
	研究属性	自由探索类基础研究						
中文关键词		压缩感知；线性逆问题；稀疏优化；数据价值；分裂-融合-共识算法						
英文关键词		Compressive Sensing; Linear Inverse Problem; Sparse Optimization ; Data Valuation; Split-Fusion-Consensus Method						



中文摘要	<p>大规模线性逆问题是描述和解决众多实际问题的通用数学模型。随着实际问题规模的不断增大，传统的求解方法已陷入了难以克服的瓶颈，其局限性主要体现在计算复杂度高、数值不稳定、内存需求量大、无视问题的分布式结构等若干方面。突破传统算法的瓶颈，发展能够应对大规模线性逆问题的专门算法技术是现代科技领域的一个重大需求。本项目旨在针对大规模线性逆问题设计高效、稳定、新型的计算方法，拟设计三类具体的算法模式：1) 正交子矩阵级联问题的分裂-融合算法；2) 大规模非结构问题的分裂-融合算法；3) 超大规模问题的分裂-融合-共识算法。同时，将这些算法拓展到新兴的数据市场的一个热点方向—大数据价值计算，打开数据质量评估的实用算法设计的一个新渠道。本项目的研究具有双重科学价值：引领线性逆问题算法的高级发展；实现大规模数据价值计算方法的根本性突破。</p>
英文摘要	<p>The large-scale Linear inverse problem is a common mathematical model for many practical problems in science and engineering. With the problem scale continues its growth, traditional algorithms for linear inverse problems are facing difficulty for solving problems due to the curse of dimensionality. The main limitations of existing algorithms are reflected in high computational complexity, numerical instability, large storage requirements, and disregard of the distributed structure of the underlying practical problem. Breaking through the bottlenecks of traditional algorithms and developing specific algorithm technologies that can truly cope with large-scale linear inverse problems is a major request in the field of science and technology. This project aims to develop a class of new algorithms that can solve large-scale linear inverse problems. Specifically, this project aims at developing three types of algorithms: 1) split-fusion algorithms for the problems with measurement matrices being the concatenation of orthogonal matrices; 2) split-fusion algorithms for large-scale non-structural inverse problems; 3) split-fusion-consensus algorithms for massive linear inverse problems. In the meantime, this project aims at applying these algorithms for the first time to large-scale dataset valuation, a research frontier of emerging data market, and thus setting up a new channel for the development of efficient methods for data quality assessment. The impact of this project are twofold: Leading a further development of the algorithm for linear inverse problems, and achieving a fundamental progress in the development of efficient methods for large-scale data valuation.</p>



主要参与者（注：主要参与者不包括项目申请人）

编号	姓名	出生年月	性别	职 称	学 位	工作单位	项目分工	办公电话	证件号码
1	张海斌	1965-06-03	男	教授	博士	北京工业大学	算法设计，理论分析	17801072076	1*****4
2	曾奕程	1992-11-03	男	副研究员	博士	深圳市大数据研究院	随机采样，降维建模	0755-84273613	4*****2
3	徐加樑	1989-12-30	男	讲师	博士	北京科技大学	数值分析，理论分析	13821885168	1*****2

总人数统计（注：包括项目申请人、主要参与者及其他参与人员；勿重复计数）

总人数	高级职称	中级职称	初级职称	博士后	博士生	硕士生	本科生及其他学生	其他
7	3	1			3			



国家自然科学基金项目资金预算表

项目申请号：1247013023 项目负责人：赵云彬 金额单位：万元

序号	科目名称	金额
1	一、项目直接费用合计	51.0000
2	1、设备费	0.0000
3	其中：设备购置费	0.0000
4	2、业务费	33.0000
5	3、劳务费	18.0000
6	二、其他来源资金	0.0000
7	三、合计	51.0000

注：请按照项目研究实际需要合理填写各科目预算金额。



预算说明书

(请按照《国家自然科学基金项目申请书预算表编制说明》等的有关要求,按照政策相符性、目标相关性和经济合理性原则,实事求是编制项目预算。填报时,直接费用应按设备费、业务费、劳务费三个类别填报,每个类别结合科研任务按支出用途进行说明。对单价 ≥ 50 万元的设备详细说明,对单价 < 50 万元的设备费用分类说明,对合作研究单位资质及资金外拨情况、自筹资金进行必要说明。)

(一) 设备费 (无)

深圳市大数据研究院提供研究场所,研究设备及所需计算资源,无需额外设备。

(二) 业务费 (共33万元)

1) 国际(境外)合作与交流费(共7.5万元):

- 出访共2万元: 申请人出访英国1次(与国际著名学者研讨项目有关研究问题),为期1个月,费用2万元(机票0.8万,当地交通和食宿费共1.2万)。
- 来访共2.5万元: 邀请香港理工大学专家孙德锋在项目执行期间来访研究院1次,为期1月,费用2.5万(交通0.3万/月,食宿2.2万/月)。
- 参加国际会议共3万元: 项目组成员参加国际交流会共2人次。会议平均费用为1.5万/人次(机票0.9万,会务费0.3万,食宿费0.3万)。共 $2 \times 1.5 = 3$ 万。

2) 国内会议(共10万元):

- 组织国际研讨会1次共4万元: 计划邀请国外2名专家来华参会,以1.5万/人计算(包括专家的机票和3天食宿);承担国内2名专家参会费用(0.5万/人)。(会议其他费用将由深圳国际工业与应用数学中心承担,本项目只承担上述部分费用)。
- 组织国内小型研讨会3次共6万元: 每次研讨会计划消费2万元(用于外单位专家交通和食宿,以及会议组织等相关费用)。

3) 国内差旅(共4万元):

- 在项目执行期间,项目组4个核心成员计划参加国内会议共8人次,共计4万元(每次会议费用以0.5万元计算,用于交通、会务费和住宿)。

4) 出版/文献/信息传播/知识产权事务费(共11.5万元):

- 版面费3.2万元(平均每年支付2-3篇论文版面费,共计不超过0.8万元/年,其他论文无版面费)。
- 文献资料费共2万元(订购急需的论文、原版专著等,以0.5万/年计算)。
- 成果宣传,打印装订费等共0.8万元(以0.2万/年计算);
- 至少3个专利,2个软件著作权和其他知识产权相关事务费共3万元: 国内专利0.8万/件,软著800元/项,其他相关费用共0.44万。
- 出版1本专著的相关费用共计2.5万元。

(三) 劳务费(共18万元)

- 劳务性费用共14.4万元: 博士生每人每年工作6个月,劳务费为2000元/人月。因此,3名博士生4年劳务费一共为 $4 \times 3 \times 1.2 = 14.4$ 万。
- 专家咨询费共3.6万元: 计划每年咨询3名具有高级专业技术职称的专家,平均咨询费为0.3万元(以平均咨询1.5天计算)。总咨询费为 $4 \times 3 \times 0.3 = 3.6$ 万。

申明: 经深圳市大数据研究院与北京工业大学及北京科技大学友好协商,本项目的经费不外拨,由深圳市大数据研究院负责经费的统一支配管理。



报告正文

参照以下提纲撰写，要求内容翔实、清晰，层次分明，标题突出。
请勿删除或改动下述提纲标题及括号中的文字。

（一）立项依据与研究内容（建议 8000 字以下）：

1. 项目的立项依据（研究意义、国内外研究现状及发展动态分析，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）；

1.1 研究背景

许多重要的实际问题常常通过逆问题进行描述。逆问题是科学研究的一个常用范式，它通过间接观测数据来探寻不可触之处的事物内部的影响因素与变化规律。例如，石油勘探就是一个典型的逆问题[1]。人们无法直接获得地层下的油田储量和位置信息，但可以通过地面爆炸向地下发射地震波，然后通过地层反射波信号成像来探测地层的地貌形态，因为地层反射信号包含了地层的密度、声速等结构信息。利用这些观察信号，通过求解逆问题，就可以成功获得几千米地层下的油储分布信息。由于逆问题的重要作用，我国著名数学家冯康把它列为计算数学的四大（正、逆、逼近和代数）问题之一。逆问题的范围非常宽阔，本项目只集中探讨线性逆问题（不涉及非线性问题）。线性逆问题的应用领域十分广泛，其数学模型和方法在系统识别、医学成像、信号处理、计算机视觉、地球物理学、遥距感应、无损检测、过程控制、无线通讯及人工智能等举不胜数的范畴中产生了持续深远的影响[1-4]。

目前，线性逆问题仍然是一个非常活跃的研究领域，它的基本数学模型可以表示成（欠定）代数方程或者等价的数学优化模型。我们以 X-射线计算机层析成像（CT）为例简单介绍这一基本数学模型。CT



是线性逆问题经典应用中最成功的例子之一。其基本思想是通过对穿透人体的 X-射线信号进行采集，重建人体内部的结构信息，生成医疗图像。为了降低图像重建的时间成本和减少病人在检测过程中面临的风险，医疗上要求减少 X-光辐射量，并缩短 CT 扫描时间。也就是说，需要解决如何用少量的观察数据重构质量保证的医学图像。因此，早期通用的滤波反投影算法逐渐被基于线性逆问题的重建方法取代。以二维场景为例，先将二维域离散成 n 个像素，然后利用能在每个像素上取得常数值函数来逼近衰减函数 w （代表病人身体轴向横切面的射线衰减系数），因此 w 可以表示成有限维空间 R^n 中的一个向量。每次测量均会得出 w 的线性函数，考虑误差和干扰的影响，经过共 m （远远小于 n ）次测量后，得到一个大规模欠定线性方程组：

$$Aw = b + e \quad (1)$$

其中 A 是测量矩阵， b 代表观察数据（观察信号）， e 代表测量误差和干扰。这个模型可称为线性逆问题的基本数学模型。

模型（1）的最根本特征是欠定性，即观察数据 b 的维数低于问题的变量数目。与上述 CT 场景类似，其他许多应用场景产生的线性逆问题模型也普遍具备欠定特征。实际上，当获得观察数据的成本很高时，极易导致观察数据的稀缺和不完整。例如很多疾病数，地震信号数据，以及时间序列数据的收集常常是不完整的，也无法在短时间内获得足够多的观察数据用于探求事物背后隐藏的真实信息。实际上，现代科技的很多领域包括通讯系统的多用户检测[3]，无线传感器网络压缩数据采集[4]，以及大数据的低秩表示[5]等均涉及只能使用少量或者欠缺的观察数据进行信号处理。

如果没有其他关于问题解 w 的额外信息，从具有无穷多解的欠定系统(1)中辨识背后的真实信息 w 是不可能的。但是，如果问题的解具有某种代数或者几何结构的额外信息可利用，就有可能从上述线性系



统的无穷多解中将原信号 w 辨识出来。Tikonov 正则化提供了有效地利用额外信息辨识问题真实解的方法[6]。利用正则化方法，可以将线性逆问题建模为如下数学优化问题：

$$\min \{ \psi(b - Aw) + \lambda \phi(w) : w \in C \}, \quad (2)$$

其中 ψ 是一个距离或者尺度函数， $\phi(w)$ 是一个正则化因子， C 是一个约束集。 $\phi(w)$ 和 C 共同刻画关于 w 的额外信息。

目前普遍采用的额外信息是问题解的“稀疏性结构”，**稀疏性结构反映了大多数物理场景普遍具备的特征或者相近的特征**。采用稀疏性结构的技术已经渗透到工程与科学的许多领域，成为解决欠定线性逆问题的一个闪亮的现代数学技术。稀疏性结构的发现和利用驱动了压缩感知理论与算法的迅速崛起，并持续发展成为信息技术领域的一个重要工具[4, 7, 8]。毫不夸张地说，压缩感知技术是最近 20 年里信号处理领域所取得的最耀眼的一项成就[9-11]。

然而，随着应用范围和问题复杂度的不断攀升，线性逆问题的规模也變得越来越大。分布式计算系统的出现，深度神经网络[12]以及最近大模型理论与技术的出现[13]，提供了人们探索求解更大规模的线性逆问题的机会与可能性，这些问题包含的变量与参数的数目可以达到亿级的规模。传统的计算方法[10, 11, 14-17]已经无法应对比比皆是的大规模线性逆问题（包括它的各种凸和非凸优化、稀疏优化、随机优化的变种问题）。因此设计和开发能够地有效地求解这些大规模线性逆问题的算法技术是工程和科技领域的一个重大需求。另外，即使有些大数据领域的计算问题表面上不是线性逆问题[18, 19]，但通过适当重新建模仍然可以表述成大规模线性逆问题，使得针对逆问题的求解方法（包括压缩感知及稀疏优化算法）也能够为大数据计算问题提供新的视角、理论及有效的算法技术。



总之，大规模线性逆问题是描述和解决众多实际问题的通用数学模型。突破传统算法的计算瓶颈，发展能够高效、稳定地求解大规模线性逆问题的新型算法技术是众多应用领域的一个重大需求。本项目拟针对大规模线性逆问题设计一类高效、新型的计算方法，并将算法拓展到新兴的数据市场的一个热点方向—大数据价值计算[18, 19]。

1.2 国内外研究现状及发展动态分析

基于额外信息的不同利用方法，线性逆问题可以被建模成不同的数学模型。本项目关注该问题的数学优化模型（2）。即使属于同一类型的优化模型也可能由于使用了不同的距离函数 ψ 和正则化因子 ϕ ，算法的设计也相应有所不同。在稀疏性条件下，由于正则化因子或者问题的约束条件对解的稀疏性结构进行刻画，问题（2）常常被称为稀疏优化问题。稀疏性被广泛使用的原因是实际应用场景中影响某种物理现象的关键因数常常并非很多。例如，一般图像类别中的自由度数量是远小于图像中的总像素数量。有效地利用稀疏性，就有可能在只有部分测量或不完整测量的条件下，通过求解优化模型（2）获得线性逆问题的真实解。由于压缩感知（CS）问题是一个典型的线性逆问题[4, 9-11, 14, 20]，对CS的研究极大推动了线性逆问题的进一步发展和应用。CS的研究大约始于2004年，历经20左右年的发展，CS基本理论与算法已经取得了重大进展，被视为自从Nyquist理论以来信号处理领域最耀眼的一项数学成就。压缩感知、稀疏优化与线性逆问题三者相辅相成，协同发展，时至今日仍然是十分活跃的研究与应用领域，尤其是针对大规模线性逆问题的研究目前仍然处于初级阶段，尚缺真正实用的算法技术。

线性逆问题在CS场景下，其理论发展十分迅速，已经取得了丰硕的研究成果，已发现的理论分析工具包括NSP [10]，Coherence [11]，



RIP [21]和 RSP [22-24], 其中 RIP 是目前使用最广泛的算法分析工具。在算法方面, L_1 -极小化也许是最早用于求解线性逆问题的优化方法, 其理论有效性在 2014 年左右通过 RIP 获得了几乎完全的刻画[25], 该方法的计算效能也被大量的应用所证实。但是, 随着大规模问题的不断涌现, L_1 -极小化算法已经无法满足求解实际需求的需求。将 L_1 -极小化模型推广到矩阵优化模型更加面临计算时间上的挑战[26, 27]。大多数凸优化算法的共性特点是基于内点算法。当矩阵维数较高时, 内点算法在迭代过程中需要反复求解大规模稠密的正规线性方程组才能获得下降搜索方向, 因此不适合求解大规模线性逆问题。

为了缓解这一计算瓶颈, Beck 和 Teboulle 于 2009 年提出了 ISTA 迭代算法[28], 其基本原理是采用梯度型的迭代公式与软 (soft) 型的阈值技术相结合求解线性逆问题, 同时基于 Nesterov 加速技术[29]提出了著名的 FISTA 算法, 该算法已成功应用于许多线性逆问题的求解。FISTA 算法产生的迭代点列依赖于软阈值参数的选取, 当矩阵维数十分庞大 (百万或者亿级) 时迭代过程也同样遭遇“维数灾”挑战。另一方面, 硬 (Hard) 阈值型算法的研究自 2008 以来也获得了巨大成功 [10, 30-32], 其中针对著名的 IHT, HTP, CoSaMP 和 SP 几类算法的研究目前已非常深刻。例如, 在 RIP 假设下, 我们于 2022 年证明了 IHT 的最紧 $3k$ 阶 RIP 收敛界限, 并将 CoSaMP 算法的理论收敛性界限也同时刷新[33]; 目前 HTP 算法的最好理论结果是由 Foucas 在[34]中完成。

由于硬型阈值算子可能抛弃大量包含重要信息的梯度分量, 因此可能导致算法收敛速度的减缓。这类算法对迭代步长的选取也十分敏感, 对一类问题或者某规模的问题有效的步长, 对另一类问题或者当规模发生变化时, 算法可能很容易失效。因此这类算法的普遍缺点是稳定性差, 也可能出现数值振荡或者完全失败。因此, 我们最近提出



了最优阈值 (ROTPw) 的概念和相应的算法, 这类算法具有良好的数值稳定性和找到问题真实解的强劲性[35, 36]。但这类算法的计算复杂度较高, 迭代过程需要求解二次凸优化的数据压缩子问题, 因此不适合求解大规模线性逆问题。我们进一步提出了基于部分牛顿方向和部分梯度型的改进方法[37, 38]和基于重球加速技术的改进方法[39, 40]。虽然这类改进算法在处理中、小规模问题时确实比原最优阈值算法在计算时间上有明显的优势, 但同样在求解大规模问题时遇到挑战。目前, 针对最优阈值算法的改进版本中, 最成功的也许是我们最近提出的自然阈值 (NT) 类算法[41], 这种方法避免了成本较高的最优阈值运算, 使得 NT 算法的计算时间与精度能够到目前国际上最好的算法水准。但是 NT 算法与目前硬型阈值类方法的一个共同局限性是除了对参数 (迭代步长) 的选取仍然较敏感外, 算法还需要存储完整的测量矩阵, 也不利于应对超大规模的问题。

还有一类主流算法—正交匹配跟踪算法 (OMP), 以缓慢的方式构建线性逆问题的稀疏解[42], 其中的最小二乘子问题的规模随着迭代过程缓慢增大, 这一特征不同于其他任何稀疏优化算法。但 OMP 的缺点是一旦在某一步选错了解的非零元的指标, 错选的指标常常会停留在所选的指标集当中, 后续迭代很难纠正该错误; 而且当大规模问题解的非零元素个数较多时, 收敛速度会相当慢, 错选指标集的概率也会增大, 导致算法的失败率上升。目前 OMP 算法已有几款改进形式包括 gOMP [43], StOMP [44] 以及最近我们提出的动态正交匹配算法(DOMP) 和它的加强形式 (EDOMP) [45], 其计算效果明显比传统的 OMP 好。但改进的 OMP 型算法在求解大规模线性逆问题时仍然有局限性, 随着指标集在迭代过程中逐步递增, 算法后续迭代过程会相当耗时。

现代科技领域产生的线性逆问题的规模越来越庞大, 传统算法已经陷入了难以克服的瓶颈, 催生了对具有低复杂性, 低内存, 稳定性,



的新一代算法技术的迫切需求。为了应对这种需求，有两个基本的算法发展趋势：

- 传统算法的 AI 化。将传统的稀疏优化迭代算法与深度学习/强化学习相结合，形成线性逆问题的求解技术。例如，最近兴起的 Learn-To-Optimize 方法（学习优化）就是属于这类算法[46]。实际上，稀疏优化的一些主流迭代算法（IHT，HTP 等）通过展开（Unfolding）技术与深度神经网络相融合已获得了新的发展[47，48]。目前的研究表明 AI 化的优化算法确实对某些线性逆问题比传统算法在计算时间上具有一定的优势，但是目前基于 AI 的稀疏优化/CS 算法的发展仍然很不成熟，对问题求解的有效性等诸多方面还需要大量，深入的研究投入。
- 发展完全崭新的直接针对大规模线性逆问题的分解-融合（共识）型算法技术。将大规模问题通过分解转化成一系列中、小规模的子问题进行求解，并基于问题之间的结构将子问题的解通过某种融合、共识技术形成原问题的解。本项目主要沿着这个方向进行算法的研发，同时将研发的算法应用于解决目前大数据领域的一个热点问题——基于 Shapley Value (SV) 的大数据价值估计问题。

下面，我们介绍数据价值计算领域的发展状况。大规模、高质量的数据集是人工智能时代最核心的资源。人类社会正在加速步入“数据商品化”的时代。目前已出现提供“数据访问”的市场（例如 DAWEX、IOTA、Xignite [49-51]），“数据运行关系查询”市场（Google BigQuery [52]）以及机器学习模型市场 [53] 等。新兴的数据市场带来了前所未有的商业契机和新的技术挑战：如何在多个数据提供者（贡献者）之间公平分配利润？例如，构建机器学习模型的数据通常具有多样性，由多方提供，而且来自不同提供者的数据质量千差万别，利益相关者



关注的一个关键问题是：如何将学习模型产生的利润公平地分配给数据提供者？为了合理、公平地分配利润，数据市场需要进行数据质量的评估，弄清各个提供者的数据价值。最近几年，数据价值评估问题吸引了许多研究者的兴趣。例如，Koutris 和他的合作人在其开创性工作[54]中探讨了对数据运行关系查询赋予价值的问题；Chen 等人[53]对“模型定价”开展了研究；数据价值的研究也受到构建基于隐私保护的数据市场的驱动[55, 56]。

幸运的是“博弈论”提供了解决数据价值评估问题的一种自然方法，可以将数据市场中的每个数据贡献者建模为协作博弈（Cooperative Game）中的参与者。最近的研究发现由诺贝尔经济学奖获得者 Shapley 提出的收益分配方法—Shapley Value (SV) 可以满足数据价值概念的诸多需求[18, 19, 57, 58]。通过 SV 来定义“数据的相对价值”似乎具有公平性、合理性和去中心化性的属性。在用于数据价值估计之前，SV 的应用领域早已十分广泛，包括经济学、反恐、机器学习、供应链、云计算等等。然而，SV 的计算十分挑战，通常需要指数时间来完成：要获得 n 个数据点的 SV 值，需要至少 $O(2^n)$ 模型评估进行计算，即使寻求对 SV 的近似计算，目前至少也需要 $O(n \log n)$ 阶的计算复杂度。如何在数百万甚至数亿个数据点上计算或近似计算 SV 已经成为十分挑战的数学问题，而这种数据量级在当今的机器学习模型训练场合都是司空见惯的。不幸的是在许多数据场合 SV 的计算涉及到效用 (Utility) 函数的计算，它本身的计算已经十分昂贵，很多时候需要借助于机器学习任务来实现，因此想利用目前的算法对大规模数据集的 SV 进行快速计算显然都是不切实际。本项目旨在探索从一个新的技术角度研发能够有效地近似计算 SV 的方法，拟建立大规模线性逆问题和数据价值计算两个看似不相关问题的关联模型，在此基础上借助本项目研发的针对大规模线性逆问题的分解-融合型算法获得



数据集的 SV 的实用、快速的近似计算方法。

1.3 项目的意义与应用前景

大规模线性逆问题是工程和科学领域的常见数学问题，是人类探索、认识自然规律常用的“据果求因”的数学方法。随着问题规模的不断变大，传统的求解方法普遍面临瓶颈（计算复杂度高，内存量需求大，计算不稳定，无视问题的分布式结构等）。本项目旨在突破传统方法的局限性，设计能够求解大规模线性逆问题的有效数值方法。本项目的研究具有重要的科学价值和应用前景。

首先，本项目拟对大规模线性逆问题提出基于测量矩阵分裂的新型计算技术，突破传统优化和迭代算法的理论与设计框架，拟设计的方法拥有两大优点：1) 通过矩阵的降维分裂减少算法每步迭代所需的内存和计算量；2) 增强算法的数值稳定性。一方面，矩阵分裂降维有助于增强算法的分布式计算能力，同时通过矩阵分裂相对增加每个“子逆问题”的测量数据，经典的 CS 关于测量次数与信号重构理论肯定了这种设计的理论依据。另一方面，由于测量误差被分散到子问题，每个子问题的测量误差相对较小，理论上可以增强算法求解逆问题的稳定性和成功率。

其次，本项目研发的算法除了可以直接用于求解各种大规模信号、图像重构的工程问题，还可以用于解决其他十分挑战的大数据计算问题。目前，大模型训练（GPT 等）和机器学习（DNN 等）的一个关键问题是训练数据的质量问题。随着 2022 年以来 ChatGPT 掀起的大模型训练的科技热浪，数据质量问题被推到了大数据和人工智能研究的前沿，对数据质量问题的研究很大可能将延申到长远的未来。如何快速进行数据的质量评估和生成高质量的数据，是在人工智能领域十分关键的需求。数据的 SV 提供了评估数据质量的一个方法。本项目将数据 SV



的计算这个十分挑战的问题建模成大规模线性逆问题，通过这种思路为大数据质量分析提供实用有效的算法支撑。可以预计本项目研发的新型算法技术在这个新兴的数据科学领域将有很好的应用前景。

本项目具有典型的“需求牵引、前沿创新、突破瓶颈”特征。因此完成本项目计划的研究内容，具有重大的理论与实际意义。不仅能够为解决工程和科学领域产生的大规模线性逆问题提供新型的求解方法，而且还能够将算法应用到一个与传统算法应用领域完全不同的其他领域—大数据价值计算，打开数据质量评估算法的一个新设计渠道。

参考文献：

- [1] M. Richter, *Inverse Problems: Basics, Theory and Applications in Geophysics*, Birkhäuser, 2022.
- [2] C. Groetsch, *Linear Inverse Problems*. In: Scherzer, O. (eds) *Handbook of Mathematical Methods in Imaging*. Springer, New York, NY, 2015.
- [3] H. Djelouat, et al., Spatial correlation aware compressed sensing for user activity detection and channel estimation in massive MTC, *IEEE Trans. Wireless Commun.*, 21 (8), 6402-6416, 2022.
- [4] M. Leinonen, et al., Compressed sensing with applications in wireless networks, *Found. & Trends in Signal Process.*, 13 (1-2), 1-282, 2019.
- [5] H. Chen, X. et al., Low-rank representation with adaptive dimensionality reduction via manifold optimization for clustering, *ACM Trans. Knowledge Discovery from Data*, 17 (9), Art no. 128, 2023.
- [6] C. Clason, *Regularization of Inverse Problems*, arXiv, 2021.
- [7] S. Gregson, et al., Compressive sensing in massive MIMO array testing: A practical guide, *IEEE Trans. Anten. Prop.*, 70 (9), 7978-7988, 2022.
- [8] X. Tong, et al., Joint multi-user communication and sensing exploiting both signal and environment sparsity, *IEEE J. Sel. Topics Signal Process.*, 15(6), 1409-1422, 2021.
- [9] G. Kutyniok, et al., *Compressed Sensing in Information Processing*, Birkhäuser, 2022.
- [10] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, Springer, 2013.



- [11] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, NY, 2010.
- [12] J. D. Kelleher, *Deep Learning*, MIT Press, 2019.
- [13] H. Naveed, et al., A comprehensive overview of large language models, arXiv, 20 Feb, 2024.
- [14] J. A. Tropp and S. Wright. Computational methods for sparse solution of linear inverse problems. *Proceeding of the IEEE*, 98: 948–958, 2010.
- [15] **Y.-B. Zhao** and D. Li, Reweighted L1-minimization for sparse solutions to underdetermined linear systems, *SIAM J. Optim.*, 22, 1065-1088, 2012.
- [16] **Y.-B. Zhao** and Z.-Q. Luo, Constructing new weighted l1-algorithms for the sparsest points of polyhedral sets, *Math. Oper. Res.*, 42, 57-76, 2017
- [17] **Y.-B. Zhao** and M. Kocvara, A new computational method for the sparsest solutions to systems of linear equations, *SIAM J. Optim.*, 25, 1110-1134, 2015.
- [18] R. Jia, et al., Efficient task-specific data valuation for nearest neighbor algorithms, *Proceedings of the VLDB Endowment*, 12 (11), 1610-1623, 2019.
- [19] S. Schoch, et al., CS-SHAPLEY: Class-wise shapley values for data valuation in classification, NeurIPS, 2022.
- [20] G.S. Alberti, et al., Compressed sensing for inverse problems and the sample complexity of the sparse radon transform, arXiv:2302.03577v1, 2023.
- [21] E.J. Candes and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), 4203-4215.
- [22] **Y.-B. Zhao**, RSP-based analysis for sparsest and least l1-norm solutions to underdetermined linear systems, *IEEE Trans. Signal Process.*, 61, 5777-5788, 2013.
- [23] **Y.-B. Zhao**, *Sparse Optimization Theory and Methods*, CRC Press, FL, 2018.
- [24] **Y.-B. Zhao**, et al., Weak stability of ℓ_1 -minimization methods in sparse data reconstruction, *Math. Oper. Res.*, 44(2), 173-195, 2019.
- [25] J. Andersson and J. O. Stromberg. On the theorem of uniform recovery of structured random matrices. *IEEE Trans. Inform. Theory*, 60: 1700–1710, 2014.
- [26] I. Waldspurger, *Lecture notes on non-convex algorithms for low-rank matrix recovery*, arXiv:2105.10318, 2021.
- [27] S. Huang, H. Wolkowicz, Low-rank matrix completion using nuclear norm minimization and facial reduction. *J Glob Optim.*, 72, 5–26, 2018。
- [28] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for



- linear inverse problems. *SIAM J. Imaging Sci.*, 2:183–202, 2009.
- [29] Y. Nesterov, Gradient methods for minimizing composite functions. *Math. Program.* 140, 125–161, 2013.
- [30] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14:629–654, 2008.
- [31] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27:265–274, 2009.
- [32] D. Needell and J.A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26:301–321, 2009.
- [33] **Y.-B. Zhao** and Z.-Q. Luo, Improved RIP-based bounds for guaranteed performance of two compressed sensing algorithms. *Sci. China Math.* 66, 1123–1140, 2023.
- [34] S. Foucart. Hard thresholding pursuit: An algorithm for compressed sensing. *SIAM J. Numer. Anal.*, 49: 2543–2563, 2011.
- [35] **Y.-B. Zhao**, Optimal k-thresholding algorithms for sparse optimization problems, *SIAM J. Optim.*, 30, 31–50, 2020.
- [36] **Y.-B. Zhao**, and Z.-Q. Luo, Analysis of optimal thresholding algorithms for compressed sensing, *Signal Process.*, 187, 108148, 2021.
- [37] N. Meng, **Y.-B. Zhao**, et al., Partial gradient optimal thresholding algorithms for a class of sparse optimization problems *J. Global Optim.*, 84, 393–412, 2022.
- [38] N. Meng and **Y.-B. Zhao**, Newton step based hard thresholding algorithms for sparse signal recovery, *IEEE Trans. Signal Process.*, 68, 6594–6606, 2020.
- [39] Z. Sun, J. Zou, **Y.-B. Zhao**, et al., Heavy-ball-based hard thresholding algorithms for sparse signal recovery, *J. Comput. and Appl. Math.*, 430, art. No. 115264, 2023.
- [40] Z. Sun, J. Zhou, **Y.-B. Zhao**, Heavy-ball-based optimal thresholding algorithms for sparse linear inverse problems, *Journal of Scientific Computing*, 96 (3), Art. no. 93, 2023.
- [41] **Y.-B. Zhao**, Z.-Q. Luo, Natural thresholding algorithm for signal reconstruction, *IEEE Open J. Signal Process.*, 3, 417–431, 2022.
- [42] J. A. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53: 4655–4666, 2007.
- [43] J. Wang, S. Kwon, and B. Shim, Generalized orthogonal matching pursuit, *IEEE Trans. Signal Process.*, 60(12), 6202–6216, 2012.



- [44] D. Donoho, Y. Tsaig, I. Drori, et al., Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit, *IEEE Trans. Inf. Theory*, 58 (2), 1094–1121, 2012.
- [45] **Y.-B. Zhao** and Z.-Q. Luo, Dynamic orthogonal matching pursuit for signal reconstruction, *IEEE Open J. Signal Process*, 4, 242-256, 2023.
- [46] T. Chen, et al., Learning to optimize: A prime and a benchmark, *J. Machine Learning Res.*, 23, 1-59, 2022.
- [47] J. Tachella, D. Chen, M. Davies, Sensing theorems for unsupervised learning in linear inverse problems, *J. Machine Learning Res.* 24, 1-45, 2023.
- [48] A. Machidon and V. Pejović, Deep learning for compressive sensing: a ubiquitous systems perspective, *Artif. Intell. Rev.*, 56, 3619–3658, 2023。
- [49] Dawex, <https://www.dawex.com/en/>.
- [50] Iota, <https://data.iota.org/#/>.
- [51] Xignite, <https://apollomapping.com/>.
- [52] Google bigquery, <https://cloud.google.com/bigquery/>.
- [53] L. Chen, P. Koutris and A. Kumar, Model-based pricing for machine learning in a data marketplace, SIGMOD '19, June 30-July 5, 2019.
- [54] P. Koutris, et al., Query-based data pricing, *Journal of the ACM*, 62, art.no. 43, 2015.
- [55] N. Hynes, et al., A demonstration of sterling: a privacy preserving data marketplace, *PVLDB*, 11, 2086–2089, 2018.
- [56] M. Goldblum, et al., Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks and Defenses, *IEEE Trans. Pattern Anal. & Machine Intelligence*, 45 (2), 1563-1580, 2023.
- [57] R. Jia, et al., Scalability vs. Utility: Do we have to sacrifice one for the other in data importance quantification? *CVPR*, 2021.
- [58] K. Bojan, et al., Data debugging with Shapley importance over end-to-end machine learning pipelines, *arXiv:2204.11131v2*, 2022.

2. 项目的研究内容、研究目标，以及拟解决的关键科学问题（此部分为重点阐述内容）；

即使在理想（即精确测量）状态下，搜索欠定线性方程组的最稀



求解已经是一个 NP-难问题。因此，在稀疏性条件下，求解大规模线性逆问题（和等价的稀疏优化问题）是十分挑战的。事实上，只有在测量矩阵满足非常苛刻的条件下，即满足所谓的 NSP, RIP 或 RSP 性质时，或者问题的解 w 的稀疏度也要低到一定程度（比如，低于矩阵 Spark 值的一半），线性逆问题才可能是一个相对容易的问题（多项式时间内可解）。然而大量的实际问题都无法满足这些理想状态，而且“大规模性”更进一步加剧了问题的求解困难。传统算法所面临的瓶颈包括 1) 应对大规模问题时对内存需求量大；2) 算法对参数的选取比较敏感，容易导致数值振荡；3) 当问题的稀疏度相对于测量矩阵的维数 (m, n) 较高时（我们称这种情形为线性逆问题的“强困难域”），目前的算法遇到这种问题时，其求解能力均会急剧下降，算法的失败率会迅速攀升。因此，本项目旨在设计大规模线性逆问题的新型数值方法，拟在计算效率、维数灾难应对和“强困难域”问题的求解能力等几个方面优于传统的算法。下面具体介绍本项目的四个主要研究内容。

2.1 正交子矩阵级联的线性逆问题的分裂-融合算法

在数据稀疏表示和信号处理领域，线性逆问题中的测量矩阵可以是级联正交矩阵。不失一般性，矩阵可以写成如下形式：

$$A = [A_1, A_2, \dots, A_J],$$

其中每一子块 A_i 内的列向量是相互正交的，每个子块不一定是方阵。传统的算法如 L_1 -极小化，阈值方法，OMP 等没有充分利用这种问题的特殊结构，因此这些算法的潜能自然没有获得充分的实现。本项目的第一个任务是充分利用问题的这种正交级联结构设计大规模问题的新型求解方法。

2.1.1 研究内容



研究如何充分利用问题的上述特殊结构，设计高效、稳定、抗干扰性强的算法技术，称为分裂-融合算法。本项目主要考察基于矩阵的两种不同的分裂模式进行算法的设计：

1) 矩阵的“极大列正交分裂”。矩阵列分裂原则是：每一个分裂的子块 A_i 是极大列正交的（也就是 A_i 具有最多个数的正交列数）。在整个迭代过程中，矩阵的列的这种分裂方式保持恒定不变。这种分裂的优势在于子问题的求解能充分利用正交性结构，在迭代过程中算法只需要执行矩阵子块与低维向量的乘积运算，因此降低了算法的内存需求。

2) 矩阵的“非极大正交分裂”。对每一极大列正交块可以被进一步划分为更小一点的子矩阵，从而获得规模更小的子问题。这种情形下，子问题变成典型的过度（overdetermined）测量型问题，也就是子问题的观察数据多于它的变量个数。因此传统针对欠定问题的所有稀疏优化算法均不再适合求解这种子问题。最小二乘方法是求解该子问题的通用方法，但我们将研究基于问题正交结构的最小二乘法方法，以及基于正交结构的主成分分析（PCA）方法。

3) 融合阈值策略。完成子问题的求解后，需要将子问题的所有解进行合并，再进行适当的阈值处理，从而产生原问题的可行解。这一过程称为“融合”过程。最后，对设计的算法进行理论分析并证明它的稳定性和收敛性。传统算法的理论分析是基于测量矩阵 A 作为一个整体，而本项目对算法的分析将利用矩阵子块的关联性和相对独立性，可以预测我们的分析方法和结果将与传统的算法有巨大的区别。

2.1.2 研究目标

1) 充分利用问题矩阵的正交级联结构，设计首个分裂-融合计算方法，实现线性逆问题的稳定和快速求解；避免使用传统的凸优化算



法求解大规模问题时所面临的维数灾难，以及传统的阈值型算法所面临的“对参数的依赖性和敏感性”的劣势。

2) 基于问题的正交结构，在温和的假设条件下完成对算法的理论分析，建立算法的首个稳定性和收敛性结果。

2.1.3 拟解决的关键科学问题

传统的稀疏优化算法的一个共性弱点是：算法忽视了问题的关键性结构（比如正交性结构），采用通用、整体、直接的求解范式，从而导致了算法耗时、不稳定的局限性。本项目拟解决的关键科学问题是：如何充分利用问题的正交级联结构，通过测量矩阵的正交块分裂，确立观察数据在子问题之间的分配策略，制定合理的子问题解之间的融合新技术，从而形成大规模线性逆问题的首个基于问题结构的高效、稳定的分裂-融合算法？

2.2 大规模非结构线性逆问题的分裂-融合算法

大多数应用场景产生的线性逆问题的测量矩阵不具有明显的结构。例如，在应用压缩感知技术时，测量矩阵可以是大型稠密的 Gaussian 矩阵、Bernoulli 矩阵或者其他分布的随机矩阵。当问题规模较大时，传统的基于内点算法的凸优化方法的每步迭代需要求解大型稠密的正规线性方程组，因此凸优化方法不适合求解大规模线性逆问题。同样，OMP 类算法在后续迭代步也需要反复求解规模逐渐增大的一系列大规模子问题，也不适合求解大规模线性逆问题；软和硬阈值型的迭代算法虽然每步的计算量比 OMP 和凸优化算法低，但是这类算法仍然需要处理整个大型测量矩阵以及算法对参数的敏感性等共性问题，导致算法求解问题的成功率和精度在很多情况下往往比凸优化方法逊色。

2.2.1 研究内容



针对一般非结构的大规模线性逆问题，设计两类分裂-融合型算法，并分析算法的有效性。算法的设计分别基于两种不同的矩阵分裂方式（固定分裂和自适应分裂）。具体研究内容包括：

1) 设计基于矩阵固定分裂策略的算法。特点是：算法在迭代过程中保持矩阵的列分裂方式恒定不变，且每个子块保持欠定性，因此对应的子问题均为欠定型。（这种算法结构的优点在于通过展开迭代过程，有利于获得算法的“学习优化”形式，但本项目不探讨学习优化，但可以作为项目的一个后续研究工作）。

2) 设计基于矩阵的自适应分裂策略的算法。研究如何根据当前迭代步产生的迭代点信息（即解的估计），制定算法在下一步迭代过程中的矩阵分裂方式。从而对子问题大小和个数进行自动调整，使算法更快、更有弹性地向问题的真实解逼近，形成求解非结构大规模线性逆问题的自适应矩阵分裂-融合方法。

3) 研究上述两个算法的稳定性和收敛性。在合适的假设条件（Coherence, RIP 或者 RSP）下刻画算法的理论性质。同时从数值模拟角度（以计算时间，求解精度或者达到所需精度需要的迭代步数等作为量化指标），研究矩阵划分的个数和算法的有效性二者之间的平衡。我们推测增加矩阵分块数，可以减少每步的计算量和所需内存，但很可能增加算法的发散风险。因此，本项目拟考察矩阵的分裂块数对算法数值性能的影响。

2.2.2 研究目标

1) 针对非结构测量矩阵问题，提出首个基于矩阵固定分裂方式的分解-融合算法，形成初步算法软件，使得算法能够克服传统算法的计算瓶颈，真正有效地求解大规模线性逆问题。尤其是求解在观察数据稀缺、昂贵、十分不完整的场景下产生的大规模线性逆问题。



2) 提出首个非结构问题的自适应分解-融合算法。通过自适应调整矩阵的列分裂,使算法在收敛速度方面超越上述固定型的矩阵分裂方法;同时在稳定性、抗干扰性、重构信号的成功率等指标上,使自适应分解-融合算法比固定分裂型算法具有优势。

3) 从理论和数值模拟角度,建立算法的分裂块数与算法有效性的理论或者经验关系,为算法的有效实现提供理论与经验支撑。

2.2.3 拟解决的关键科学问题

很多场合观察数据的收集十分困难且收集成本极高,导致观察数据十分稀缺和不完整,而事物内部的变化因数(自由度)却十分庞大。这种情况下,如何利用稀缺少量的观察数据探求事物内部的变化因数?这是人们长期关注和拟解决的关键科学问题。把该问题描述成大规模线性逆问题(其维数满足 $m \ll n$,且相比之下 m 的值并不算太大),那么上述问题正好是本项目拟解决的关键科学问题:能否设计有效的计算方法求解大规模的欠定线性逆问题?分解-融合算法的思想提供了解决这一问题的可行方向,同时也带来另一个更为具体的关键问题。如前所述,矩阵的分裂会减少每步计算所需内存,但也可能影响算法的求解成功率。因此,本项目拟解决如何达成矩阵分裂和算法的求解效能两者之间的平衡问题,这也是设计和应用算法时值得解决的一个具体的关键科学问题。

2.3 基于矩阵网格分裂的大规模线性逆问题的分裂-融合-共识算法

“研究内容 2.1”探讨具有正交结构的问题的计算方法;“研究内容 2.2”探讨非结构问题的算法,问题测量矩阵的行数 m 远远低于其列数 n ,但 m 相对不算巨大,也就是说每个子块 $m \times n_i$, $i = 1, \dots, J$ 的规模对可用的计算资源来说是可接受的。因此这种情况下,只需要基



于测量矩阵的列分裂进行分裂-融合算法的设计就能在已有计算资源条件下完成问题的求解。

但是当 m, n 巨大时, 单一的列分裂策略仍然面临巨大的计算困难。因为只对矩阵进行列分裂, 产生的欠定子矩阵维数 (m, n_i) 仍然十分庞大, 迭代过程中使用传统算法反复求解这样的子问题仍然是不现实的。更为困难的情形是当 m, n 巨大, $m < n$, 且 m 与 n 的差距不大时 (也就是 m 并非远远小于 n), 这种情况下如果只进行矩阵的列分裂容易导致 $m > n_i$ 的非结构超定子问题, 或者导致 $m \approx n_i$ 的问题 (即处于“强困难域”中的问题)。目前的稀疏优化方法不适合求解这些问题。因此, 本项目将探索基于测量矩阵的网格分裂的算法, 如图 2.1 所示:

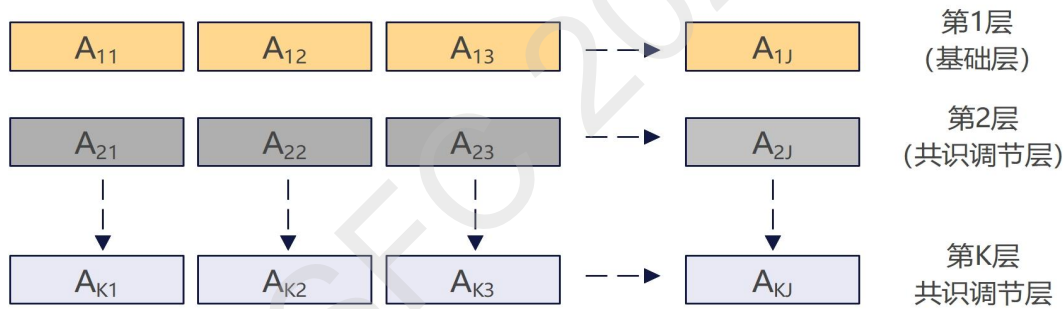


图 2.1: 矩阵网格分裂图

矩阵 A 被分裂成水平方向上 J 块和纵向方向上 K 层。本项目拟称第一层为“基础层”, 其他层称为“共识调节层”。我们认为: 基于矩阵的网格分裂研发逆问题的新型计算方法反映了压缩感知、线性逆问题和稀疏优化在经过 20 年左右的蓬勃发展后迈入了更高级的研究阶段—从小规模问题的求解时代, 发展到中规模, 最后集成小规模问题的求解方法走向大规模和超大规模问题的计算时代。

2.3.1 研究内容

1) 基于固定网格的分裂-融合-共识算法。针对超大规模线性逆问



题，研发基于矩阵的固定的行列同时分裂（即网格分裂）的算法技术。该技术包含三个关键、缺一不可的要素：矩阵网格分裂；子问题解的融合（fusion）；行层子问题之间的共识（consensus）。分裂的目的是将问题化小，融合的目的是恢复原问题的解，共识的目的是维持行层子问题解之间的一致性。

2) 研究自适应网格分裂-融合-共识算法。基于迭代点提供的信息，以及行层子问题之间的共识程度（拟通过共识误差来定义或度量）等信息，制定矩阵的网格分裂的调整规则，使算法在迭代过程中能够自适应地调整矩阵的网格划分。

3) 网格分裂-融合-共识算法的理论分析。在经典的 RIP 或者可检验的 Coherence 假设条件下分析算法的收敛性、稳定性及算法的误差界限。分析算法的计算性能（收敛性、收敛速度、稳定性、误差界限）与矩阵的网格分裂方式、观察数据在子问题之间的分配方式、行层之间的共识方法及共识精度等几大因数之间的联系，为有效地实现算法提供理论依据和可解释性。

2.3.2 研究目标

1) 搭建超大规模线性逆问题的网格分裂-融合-共识算法的一般框架，提供低内存需求的迭代算法设计的一个新范式。研发两款具体的算法模式：固定矩阵网格分裂的算法和自适应矩阵网格分裂的算法。

2) 建立网格分裂方式与算法计算效能指标之间联系。具体建立理论收敛速度、迭代复杂性界限、稳定性控制、误差下降速率等算法特性与网格分裂方法之间的某种理论关系。为算法的通用和垂直应用提供自适应调整策略，也为将来进一步探讨更智能化的算法提供理论基础与数值经验。



2.3.3 拟解决的关键科学问题

传统的算法已经无法有效地求解超大规模的线性逆问题。本项目拟解决的关键科学问题是：如何有效地进行矩阵的网格分裂将超大规模问题转化成一系列中、小规模子问题进行求解，通过子问题解之间（水平方向）的融合和（垂直方向）共识，最终达到求解超大规模问题的目的？为了解决这个问题，本项目的主要攻坚难点包括以下三个具体方面：首先，制定子问题之间观察数据的分配原则；其次，确定子问题的共识机理以保证算法的整体计算效能（速度，进度，精度，稳定性等）。能否从理论上或者数值模拟角度揭示所采用的共识方法与算法效能之间的某种联系？值得探讨的另一个具体的关键问题是：能否从理论上或者通过数值模拟的方式刻画测量矩阵的网格分裂块数对算法真实效能的影响？

2.4 数据价值的新型、低复杂度计算方法

本项目针对大规模线性逆问题提出的分解-融合（-共识）算法有广阔的应用前景。首先可以直接应用于大规模数据的稀疏表示/压缩，信号与图像去噪/重构，传感器网路中数据压缩采集，通讯系统信道信息重构，无线网络多用户监测，盲源信息分离，时间序列数据的在线跟踪，雷达图像的实时处理，地震信号的辨识等众多应用领域。为了展示算法跨领域的应用前景，本项目拟将分裂-融合-共识算法拓展到基于 Shapley Value (SV) 的数据价值计算问题，一个与传统的信号、图像处理完全不同的新兴大数据计算领域。

用于模型训练的数据质量以及学习模型的质量（或者价值）是机器学习(ML)的两个最核心的要数。例如，给定一个包含数百万数据点的数据集 D ，数据消费者支付 Y 元来训练 D 上的ML模型，自然产生如下问题：应该如何将这 Y 元公平地分配给每个数据点以反映其数据的



“价值”？如前所述，可以使用数据点的 SV 来度量其价值。也就是说，可以将每个数据点建模为合作博弈（Cooperative Game）中的“参与者”，其价值可以通过所谓的效用函数来表征。具体来说，令数据集 $D = \{D_1, D_2, \dots, D_n\}$ 由 n 个数据点 $D_i, i = 1, \dots, n$ 构成。以机器学习模型为例，用 $Z: 2^D \rightarrow R$ 表示在数据点子集上训练 ML 模型的效用函数，数据点 D_i 的 SV 精确计算公式如下

$$S_i = \frac{1}{n} \sum_{S \subseteq I \setminus \{i\}} \frac{1}{\binom{n-1}{|S|}} (Z(S \cup \{i\}) - Z(S)),$$

其中 $I = \{1, \dots, n\}$ 。可见完成 n 个数据点的 SV 计算，需要指数级 2^n 的效用评估。当用 SV 作为 ML 模型价值度量时，每一次效用函数的计算还需要通过 ML 模型训练，对计算资源（时间，软件，硬件等）的需求量可想而知。虽然最近几年出现了关于数据的 SV 计算的少量探索，但是目前算法的共性问题仍然是计算成本太高而不切实际。这对使用 SV 进行数据价值评估提出了根本性的挑战——我们如何有效地计算数百万甚至数亿的数据点的 SV？这种规模的数据集对于中、大规模的机器学习场景却是非常司空见惯的。

2.4.1 研究内容

针对 SV 计算的挑战，本项目拟开展以下两个方面的研究：

1) 将大规模数据的 SV 的精确和近似计算问题建模为大规模线性逆问题。为此，首先需要探索 SV 在某种变换下的稀疏性，其次采用随机化方法（Randomized Method）将确定性的数据 SV 计算问题建模为基于统计模型的 SV 估计框架，从而可实现对数据集的 SV 进行少量随机采样，形成超大规模的线性逆问题模型。

2) 针对建立的 SV-线性逆问题模型，将本项目提出的通用分解-融合（-共识）算法技术拓展到数据的 SV 计算问题；同时根据该逆问



题模型的结构对通用算法进行微调；并考察采用结构化和非结构的测量矩阵对 SV 计算的影响。重点考察包括级联正交列矩阵（其优势在于算法对内存需求量低，计算过程只用小规模矩阵与向量的乘积运算）和 Bernoulli 矩阵作为测量矩阵时对数据的 SV 计算的影响。

2.4.2 研究目标

1) 搭建基于博弈论 SV 计算的数据价值评估的大规模线性逆问题模型，从而将数据集的 SV 计算与几个前沿的科学计算领域（稀疏优化，压缩感知，机器学习等）直接关联起来。

2) 研发实用的大数据价值计算(质量评估)的低复杂度、低内存的新型计算方法，克服目前数据集的 SV 计算瓶颈，实现对大规模数据集的 SV 进行有效的近似计算。

2.4.3 拟解决的关键科学问题

数据集的 SV 计算随着数据规模的攀升成指数增长，这对使用 SV 进行数据评估提出了根本性的挑战。如何建立 SV 的近似计算模型，并设计能够在多项式时间内完成对 SV 的估计或者近似的高效计算方法是本项目拟解决的关键科学问题。本项目拟通过建立 SV 计算的线性逆问题模型，揭示这两个计算问题之间的联系。在此基础上，将本项目提出的分解-融合型算法以及其他最新的稀疏优化理论与算法首次应用于数据价值计算领域，解决目前 SV 算法所面临的“维数灾”困境。

3. 拟采取的研究方案及可行性分析（包括研究方法、技术路线、实验手段、关键技术等说明）；

3.1 人员分工



项目组共 7 人（4 个核心研究人员和 3 个与香港中文大学（深圳）联合培养的博士生）。赵云彬负责整个项目的全面协调，承担主要算法的设计，理论分析，论文撰写，应用指导与专利申报；张海斌（北工大）承担部分算法设计及分析，并在项目需要更多的研究助理时提供备选学生；徐加樑（北科大）承担部分数值实验和理论分析；曾奕程承担部分理论分析，尤其是利用统计和随机化方法对 SV 重新建模部分的研究及分析，并协助负责人管理项目；3 名博士生主要承担数值实验，部分理论分析及初步算法软件，并协助核心成员申报专利和软件著作权等相关事宜。

3.2 项目的总技术路线图

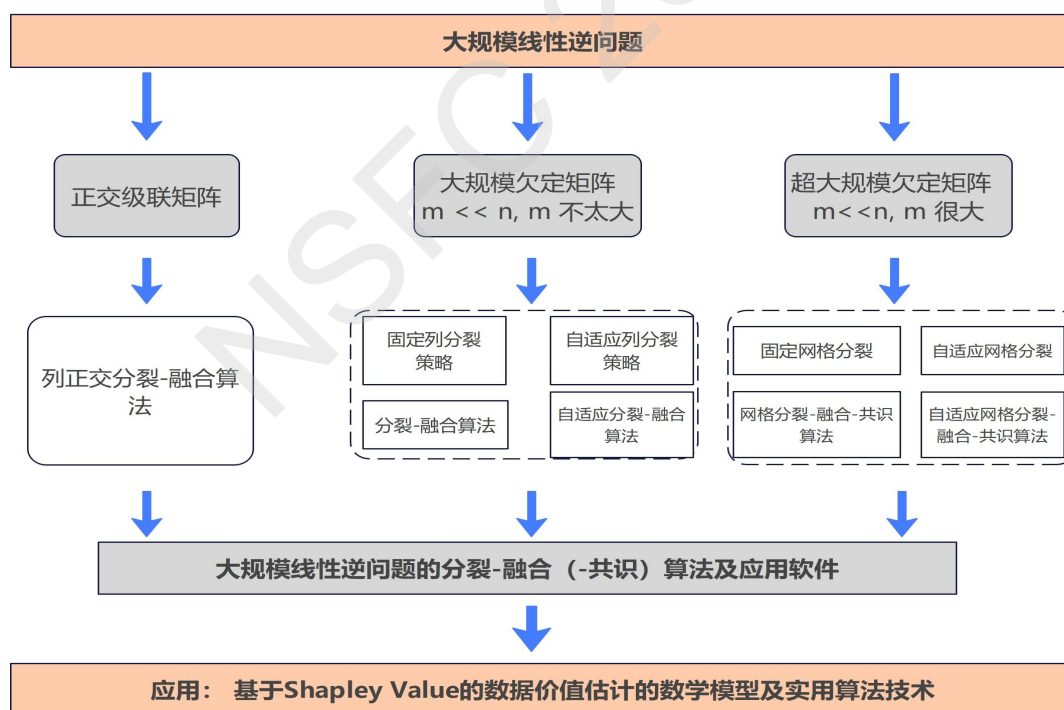


图 3.1：技术路线总图

大规模欠定线性逆问题根据其矩阵的结构可以分为三类：1）带正



交级联矩阵的问题；2) 中等 m 值的大规模问题；3) 巨大 m 值的超大规模问题。第一类是结构特殊的问题，第二和第三类是实际应用中普遍遇到的问题。后面两类问题由于规模的不同，设计算法的方式也不同。第三类问题的算法研发是十分挑战的，是本项目的主要攻坚点。总的研究方案是从第一类问题的算法的设计开始，然后将算法设计的基本思路拓展到第二类问题 ($m \ll n$, m 值不太大)，再进一步引入共识方法将第二类问题的研究思路拓展到第三类问题 ($m \ll n$, m 值巨大)。最后探讨所设计的算法在大数据计算领域中的应用—数据价值估计。下面具体介绍各阶段研究方法的可行性、每一类问题研究的具体技术路线以及涉及的关键技术。

3.3 研究方法 with 可行性

3.3.1 正交子矩阵级联的线性逆问题的分裂-融合算法

我们研究测量矩阵 A 是有限个列正交块矩阵 A_i 组成的级联矩阵： $A = [A_1, A_2, \dots, A_J]$ 。为了充分利用正交子矩阵的特性，在对 A 进行列分裂时，拟探讨两种基本的设计方案：第一种方案保证每一子块是极大正交的；第二种方案允许每一分裂形成的正交子块不一定是极大的。显然第一类算法是第二类算法的特例。

基于第一种矩阵分裂方案，在整个迭代过程中，矩阵的分裂方式保持不变（本项目把这种情形称为固定列分裂算法）。我们拟通过迭代点提供的信息制定观察向量 b 在子问题之间的分配量 $b_i, i = 1, \dots, J$ 。这种情况下，子问题通常是超定型的（变量的个数不超过观察数据 b_i 的维数）。关于超定子问题的求解，一种可行方案是采用通用的最小二乘法，另一个可行的方法是采用 Principal Component Analysis (PCA) 方法（PCA 可以尽量获得子问题具有一定稀疏度的解，可能对算法的总体收敛速度会有一定的加速作用）。不论是标准最小二乘或 PCA，还



是其他方法，充分利用矩阵块 A_i 的正交性，可以预见算法比一般情况下的计算要快得多）。我们预测本项目设计的算法是一类低内存、快速、稳定的计算方法。

第二类分裂方案对矩阵进行更细的分裂，将形成规模更小的超定子问题。传统的针对压缩感知和欠定稀疏优化问题的任何算法均不适合求解这类子问题。但 PCA 与最小二乘方法可以应用于这种场合。外，为了增加算法的数值稳定性，根据稀疏优化算法经验，在迭代过程中应用阈值技术产生问题可行解的同时，配合使用一次匹配追踪步能起到既加速又稳定算法的效果。

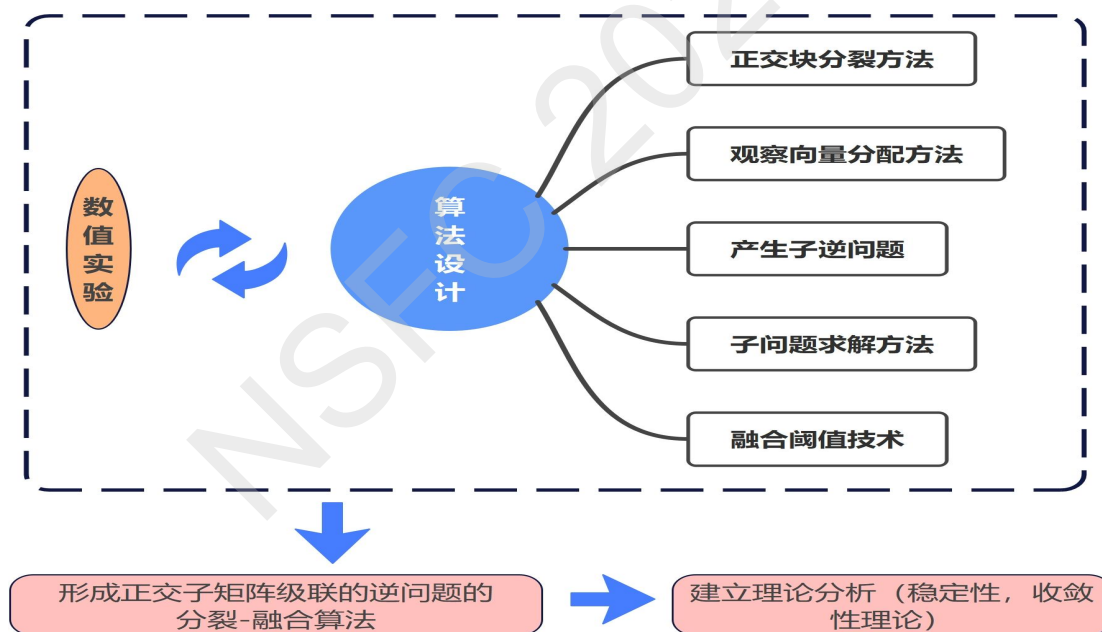


图 3.2：项目第一阶段的技术路线图

在理论分析方面，拟证明算法的两个关键性质（稳定性和收敛性）。传统分析方法把 A 作为一个整体，因此对测量矩阵整体的要求比较苛刻。由于本项目提出的分解算法不是将 A 作为一个不可拆分的整体，因此算法的理论分析手段自然与传统的方法不同。从 Coherence 角度



来看，矩阵的分裂，会有利于 Coherence 值的减少，经典的压缩感知理论显示这对信号（数据）的重构是有利的。但过度的分裂又有可能增加算法发散的风险。可以预见，虽然本项目拟建立的收敛理论与传统的理论有很大的不同，但基本的分析工具仍然可以采用 Coherence 和 RIP 或者类似的工具，这些经典的分析工具为本项目算法的分析提供了可靠的手段和良好的基础。图 3.2 是这一阶段的具体技术路线。

3.3.2 大规模非结构线性逆问题的分裂-融合算法

本项目这一阶段研究：1) 设计基于矩阵固定分裂方式的分裂-融合算法；2) 设计自适应分裂-融合算法；3) 建立两类算法的收敛性和稳定性理论，并探索矩阵分裂数与算法有效性的理论或数值关系。下面简单说明拟采用的研究方法的可行性。

如果 $w^* \in R^n$ 是线性逆问题的真实解，将矩阵分裂成 J 块，相应地将 w^* 分裂成 (w_1^*, \dots, w_J^*) ，那么观察数据对应于最优子问题的分配将是 (b_1^*, \dots, b_J^*) ，其中 $b_i^* = A_i w_i^* + e_i^*$ 可以被称为第 i 个子问题的最优观察数据分配。这说明逆问题的解确实可以写成若干个子问题的解的某种融合。因此，分裂-融合算法的设计思路从理论上说是合理可行的。

具体来说，本项目对矩阵固定分裂方法的研究拟采用如下的研究方案：首先采用去中心化（decentralization）的思想，根据测量矩阵的固定列分裂方式将观察数据（ b 向量）按某种策略分解成几个子向量。这样，原问题被分解成几个结构相似，相互关联的欠定子模型。子问题的计算可以直接调用任何一款稀疏优化算法（也可以针对子问题的结构进行专门的算法设计），通过对子问题解进行合并融合及阈值处理（例如采样硬型阈值技术），产生下一个迭代点。这是易于实现的设计方案。

本项目对自适应算法的设计拟考虑如下方案：一种方案是把算法



的当前迭代点（或者已产生的几个迭代点的）的稀疏结构作为问题解的稀疏结构的预测，重新调整算法在下一步迭代中的矩阵分裂方式；另一种方案是根据压缩感知的基本理论调节矩阵的分裂方式。例如，对具有稠密解的子问题增加其矩阵列数，加强该子问题的欠定性和解的稀疏度，从而有可能增强稀疏优化算法构建子问题的稀疏解的能力和速度。因此，本项目的自适应算法的研究方案既遵从压缩感知的基本理论，也将符合稀疏优化算法应用的实际经验。

显然，分裂块数与算法的数值表现也是息息相关的。可以推测块数太多可能会降低算法找到问题解的可能性。本项目拟采用 Coherence 或者 RIP 作为基本理论分析工具，建立算法的收敛和稳定的充分条件。我们推测这些条件与矩阵分裂数之间存在关系。这种关系为算法的最优设计和算法计算潜能的实现提供了可能性。这一研究方法也是基于项目组成员多年来对线性逆问题、压缩感知及稀疏优化的理论与算法的理解和丰富的实践经验。

项目这一阶段的具体技术路线可以简单描述如图 3.3。

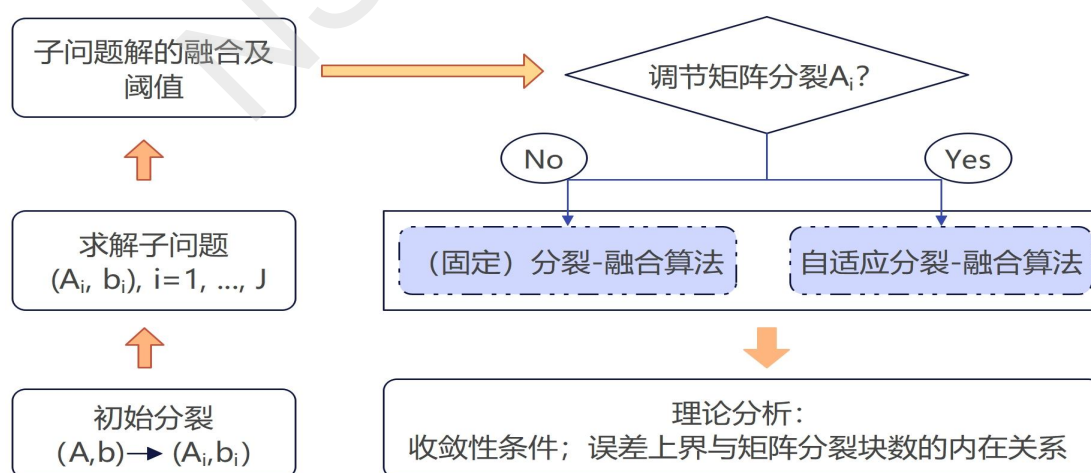


图 3.3：项目第二阶段的技术路线图

3.3.3 基于矩阵网格分裂的大规模线性逆问题的分裂-融合-共识算法



设计矩阵网格分裂-融合-共识算法的关键因数包括矩阵的行和列分裂准则，子问题之间的共识策略，以及子问题解的融合阈值策略。分裂准则决定迭代走向，影响后面的共识和融合步，从而最终影响算法的计算复杂度和求解问题的能力。本项目拟研究两个算法设计方案。

固定网格分裂

首先设计基于固定网格分裂的算法。为了利于子问题能够被传统的稀疏优化算法高效地求解，矩阵的列分裂应遵循压缩感知理论的指引。比如，分裂时尽量降低分裂子块的 Coherence 值，有助于提高稀疏优化方法找到子问题的稀疏解的机会。因此，如果采用使 Coherence 尽量下降的原则进行矩阵的分裂将是一个合理、可行的算法设计方案。同时，矩阵的行分裂策略与列分裂方法将共同决定子问题的规模。我们拟探索如下的初步设计思路：将矩阵网格分裂后形成的最顶层作为基础层，对基础层应用本项目第二阶段研发的分裂-融合算法，忽略其他层，获得计算的初步结果。在此基础上，把其他层作为共识层，引入子问题的共识调节器——通过引入正则 (Regularization) 项对子问题的解进行优化调节，促使低层问题的解与基础层尽量靠近，达成一定的“共识”。最近若干年，“共识”技术在分布式优化领域获得了高度重视，这一技术也为本项目的算法设计提供了初步思路 and 基础。

自适应网格分裂

1) 列分裂的自适应策略：拟通过考察当前迭代点的稀疏性结构调节矩阵的分裂方式。例如，当子问题具有稠密解时，通过增加其矩阵的列数，可以提高其解的稀疏度，压缩感知理论指出这种调节方案有利于加强稀疏优化算法获得子问题解的精度和稳定性。这种自适应调节方案也可能缩减上一步迭代使用的分裂块数，从而也起到调节子问题个数的目的。

2) 行分层的自适应策略：行分层会导致不同层间的迭代解的不一



致性，因此需要层之间达成一定的共识，自适应算法根据某种策略调节矩阵行分裂。本项目拟探索如何根据不同层间的共识误差进行矩阵的行分裂的调整。根据层间的共识误差大小，为下一步迭代制定矩阵网格分裂的调整规则，包括调节矩阵的分层数和各层的厚度。可以考察如下的方案：对误差较大的两层，拟采用增厚上一层，减薄下一层的手段（这样经过下一步迭代后，该两层之间的共识误差趋于减小）。这样，子问题模块随着算法迭代的进展将获得自适应调节。可以预测，当迭代点接近逆问题真实解时，由于其稀疏性结构被确定，列分裂的调节趋于固定，层间的共性误差将越来越小。

总之，本项目的上述算法设计方案是基于稀疏优化和压缩感知的基本原理和长期的算法应用与发展经验，即新颖又有理论的支撑，是合理可行的研究方案。另外，通过子矩阵的 Coherence 或 RIP，以及通过共识误差的控制策略，刻画算法的收敛性条件似乎也是十分合理可行的。图 3.4 简单描述了这一阶段的基本技术路线。

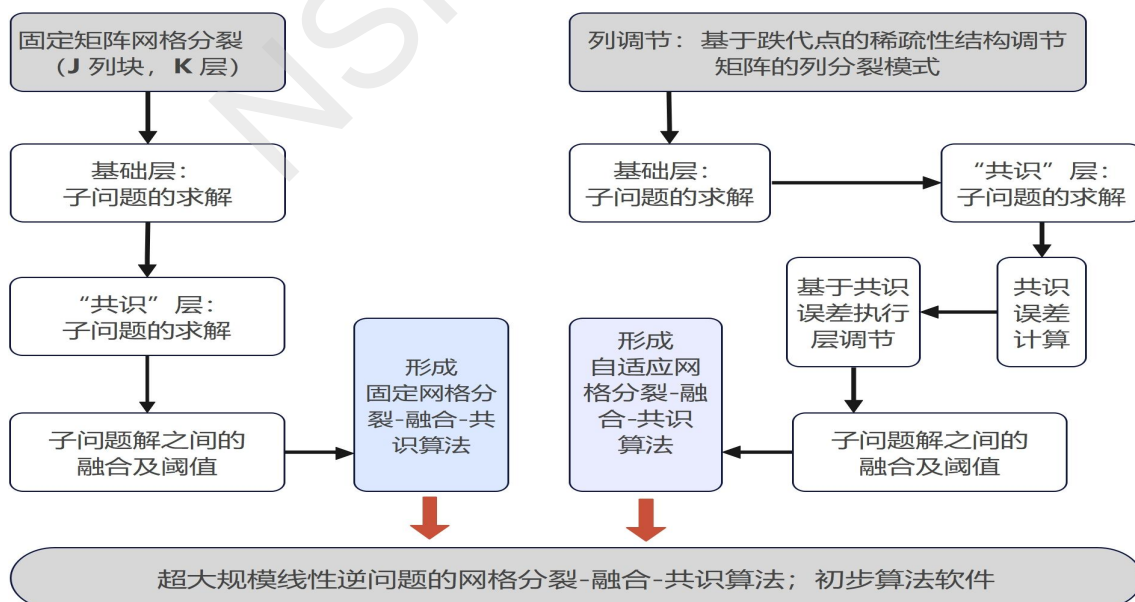


图 3.4：项目第三阶段的技术路线图



3.3.4 数据价值的新型、低复杂度计算方法

大规模数据集的 SV 计算是十分挑战的，其计算成本随着数据规模的递增成指数增长，哪怕仅仅对 $n = 1000$ 个数据点，要完成这些数据点的 SV 计算所需的效用函数计算次数已经高达 2^{1000} ，传统的算法早已无能为力了。这对使用 SV 进行数据质量和价值评估提出了根本性的挑战。本项目拟采用分解-融合-共识算法解决目前 SV 计算所面临的“维数灾”。具体的研究方案分为两大步：

1) 首先将 SV 计算问题建模成超大规模的线性逆问题模型。这种方案是合理可行的。一方面，目前的统计研究显示 SV 在某种（简单或者特殊）线性变换下具有一定的可压缩性或者稀疏性；另一方面，由于数据点的 SV 由该数据点的所有可能的效用边际的平均来决定，因此可以采用随机化（randomized）的方法进行边际估算，获取 SV 的压缩性采样“观察值”，为 SV 计算问题建模成线性逆问题提供了合理的手段。

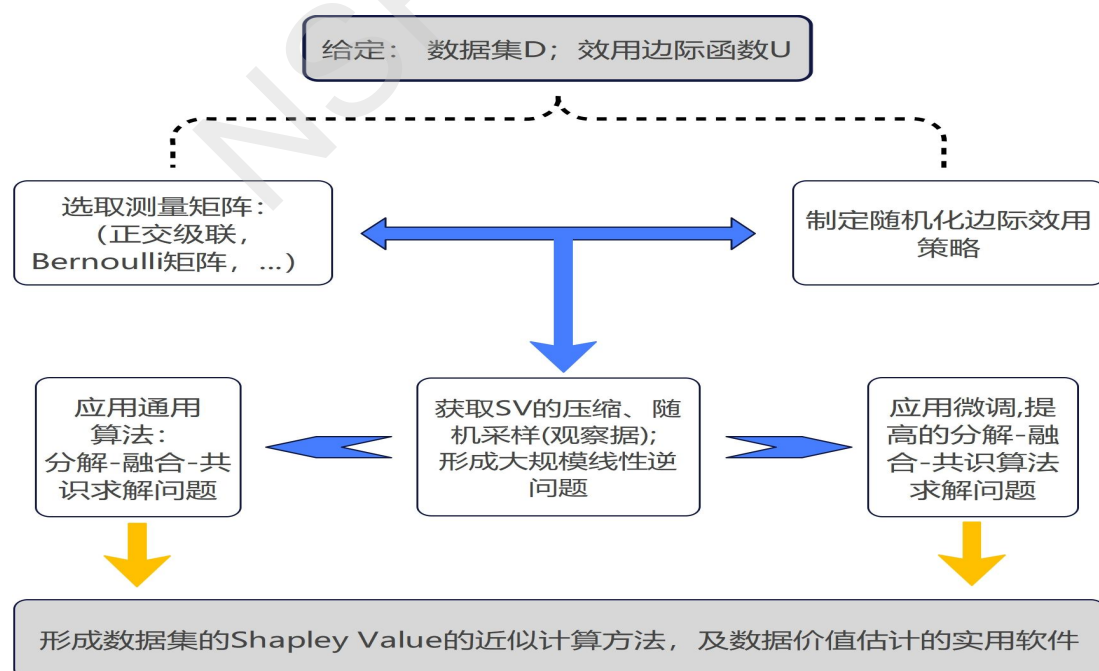


图 3.5：项目第四阶段的技术路线图



2) 其次, 基于建立的 SV 的超大规模线性逆问题模型, 本项目研发的大规模线性逆问题的求解方法自然可以用来求解 SV -线性逆问题。基于该问题的结构特征, 还可以对算法进行微调、修改及提高。由于分解-融合-共识算法提供了将大规模问题转化成一系列可操作、可求解的小规模问题进行求解, 因此也提供了可靠的手段用于攻克 SV 计算的维数灾问题, 为实现 SV 快速近似计算提供了极大的可能性。图 3.5 是这一研究阶段的技术路线图。

3.4 关键技术说明

正交子矩阵级联的线性逆问题的分裂-融合算法: 传统的稀疏优化算法(例如 L_1 -极小化方法)和梯度阈值迭代算法(IHT, HTP, NT, CoSaMP等)以及OMP算法类均不能充分利用测量矩阵子块的正交性。本项目充分利用子块的正交性结构设计算法, 因此与传统算法的设计有巨大的区别, 其中所使用的列正交分裂和相应子问题之间的观察向量分配技术是算法设计的关键技术。不同的分裂方法和观测向量的分配机理将直接影响算法的求解效能。

大规模非结构线性逆问题的分裂-融合算法: 1) 对于固定分裂模式的算法设计, 关键技术仍然是制定观察向量在子问题之间的分配策略, 由于分配方案的非唯一性, 探索有效的分配方案是算法设计的关键。2) 对于自适应算法的设计, 制定矩阵的自适应列分裂判据(规则)是整个设计的关键。本项目拟基于算法获得的迭代点的信息以及迭代点的稀疏性结构制定矩阵的自适应分裂策略。另外, 如何平衡矩阵分裂个数和算法求解问题的成功率也是算法设计的一个关键技术问题。建立算法的理论收敛性界限(其中包含矩阵分裂个数等信息), 通过优化这种上界解决上述平衡问题是本阶段算法设计的另一个关键技术。

基于矩阵网格分裂的大规模线性逆问题的分裂-融合-共识算法:



观察信号 b 的分配和矩阵分裂策略是算法设计的两个关键因数，决定迭代的走向及算法的收敛性和收敛速度。项目第二阶段关于观察向量的分配和矩阵的列分裂策略同样在第三阶段的研究中发挥关键作用，但第三阶段的算法设计更为挑战，还需要“矩阵的自适应行分层”这一关键技术。如前所述，为了保持不同层子问题的解之间的一致性，需要引入共识调节函数，与子问题结合形成共识优化子问题，通过它调节矩阵的分层数和层厚度，形成自适应矩阵网格分裂方法，智能地引导算法走向更高效的计算过程。

数据价值的新型、低复杂度计算方法：为了真正实现利用分解-融合-共识算法解决大规模数据的 SV 计算的目标，本项目拟采用的关键技术是通过边际效益随机化和随机压缩采样，将 SV 的（近似）计算问题表述为合理的大规模线性逆问题。也就是说通过模型转化搭建 SV 计算与逆问题的联系是本项目采用的关键技术。这样，大数据的 SV 计算问题可以转化为分解-融合-共识算法能够应对的实际问题，使得大数据的 SV 估计所面临的“维数灾”困境能够被有效地规避。

4. 本项目的特色与创新之处；

线性逆问题是科学与工程领域中的一个通用数学模型。压缩感知的发现再次将线性逆问题推到现代科技的聚光灯下，其理论与算法的发展正在迈向一个新的台阶。线性逆问题的求解规模历经了从过去的小规模，发展到中规模，最后集成小规模算法及新兴的算法技术走向大规模计算的新时代。

特色：本项目旨在通过稀疏优化、压缩感知、大数据计算的交叉数学技术，研发大规模线性逆问题的新型求解方法。本项目提出的算法旨在引领大规模线性逆问题求解的新一轮发展，其中的分解、融合、



共识是本项目算法设计的三个根本要数，区别于传统算法的设计范式。另一个特色是：将算法应用到一个新兴的大数据计算领域—数据价值计算，旨在开辟数据质量评估算法的一个崭新的设计渠道。本项目的研究内容具有典型的“需求牵引、前沿创新、突破瓶颈”特征。

创新之处：（a）对于大规模线性逆问题，传统的算法均陷入了难以克服的瓶颈。本项目首次设计基于分裂、融合、共识为一体的直接针对大规模线性逆问题的新型计算方法，与传统的算法有重大区别；（b）首次探索将所研发的分解-融合-共识算法应用于大数据价值计算这一前沿的热点领域，将大数据的 SV 计算问题转化为可实现的稀疏优化计算问题，有效地解决目前 SV 计算所面临的根本性困难，填补这一领域实用算法的空白。

5. 年度研究计划及预期研究结果（包括拟组织的重要学术交流活动、国际合作与交流计划等）。

项目将按照前述的 4 个“研究内容”执行。在执行过程中，根据实际进展情况，有些研究任务可以交叉或者同步进行。但总的计划为平均每年完成一项研究内容。具体年度计划与预期成果分别简述如下：

5.1 （2025/1-2025/12）执行第一阶段的研究任务“正交子矩阵级联的线性逆问题的分裂-融合算法”

- 1) **完成 3 个具体的研究任务：**T1 基于“极大列正交分裂”策略设计算法；T2 基于“非极大列正交分裂”策略设计算法；T3 理论分析、大量数值模拟以及算法的初步软件和应用。
- 2) **交流与合作：**申请人计划访问英国南安普顿大学（1 个月），与统计专家卢祖帝教授和压缩感知算法专家 T. Blumensath 教授探讨本



项目算法设计与应用问题，后者是著名的稀疏优化 IHT 算法的提出者；计划在 2025 年夏季组织 1 次线性逆问题算法相关的国内研讨会（20 人左右）；计划参加 1 次国际会议（IEEE、SIAM 或 INFORMS）和 2 次国内运筹优化学术会议；项目成员张海斌（北工大）和徐加樑（北科大）来访研究院各 1 个月，参与项目第一阶段的研究（这两位合作人来访相关费用由依托单位资助）。

- 3) **预期成果：**国际会议论文 1-2 篇；著名学术刊物（IEEE、SIAM 及其他）论文 2 篇；基于项目第一阶研究形成的初步计算软件 1 套；组织国内研讨会 1 次。

5.2 （2026/1-2026/12）执行第二阶段的研究任务“大规模非结构线性逆问题的分裂-融合算法”

- 1) **完成 3 个具体的研究任务：**T4 分裂-融合算法设计；T5 自适应分裂-融合算法设计；T6 理论分析，初步软件及应用。
- 2) **交流与合作：**邀请香港理工大学专家孙德锋教来访深圳市大数据研究院 1 个月；组织 1 次国内小型优化算法研讨会（20 人左右）；项目成员拟参加国内优化算法、大数据计算相关的会议 2 人次；项目成员张海斌（北工大）和徐加樑（北科大）来访研究院各 1 个月，参与项目第二阶段的合作研究（合作人来访的相关费用由依托单位资助）。
- 3) **预期成果：**会议论文 1-2 篇；著名刊物论文 2 篇；形成 2 款算法软件（固定矩阵分裂型和自适应分裂型）；申请算法相关的发明专利及软件著作权各 1 项；组织国内研讨会 1 次。

5.3 （2027/1-2028/12）执行第三阶的研究任务“基于矩阵网格分裂的大规模线性逆问题的分裂-融合-共识算法”



- 1) **完成 3 个具体的研究任务：**T7 固定网格分裂-融合-共识算法设计；T8 自适应网格分裂-融合-共识算法设计；T9 理论分析、软件形成及应用；T10 网格划分与算法性能关系研究。
- 2) **交流与合作：**项目成员计划参加数值分析与计算优化相关的国际会议 1 人次（比如由 Birmingham 主办的英国数值分析与优化年会），国内计算优化相关的会议 2 人次，并在“深圳国际应用数学中心”主办 1 场大数据、稀疏优化算法的国际研讨会（50 人左右）；项目成员张海斌（北工大）和徐加樑（北科大）来访研究院各 1 个月，继续参与项目第三阶段的研究（两位合作人来访的相关费用由依托单位资助）。
- 3) **预期成果：**著名刊物学术论文 2 篇；会议论文 1-2 篇；组织国际会议 1 场；2 款初步算法软件；申请与本阶段算法成果有关的发明专利 1 项。

5.4 （2028/1-2028/12）执行第四阶段的研究任务“数据价值的新型、低复杂度计算方法”

- 1) **完成 3 个具体的研究任务：**T11 建立 SV 计算的逆问题数学模型；T12 SV 的分解-融合-共识算法的微调及实现；T3 形成初步软件并完成与传统算法的比较。
- 2) **交流与合作：**提交 Springer 专著出版的提案，并撰写关于线性逆问题算法的英文专著初稿（计划在项目结题后正式出版）；参加国内优化、计算数学相关的会议 2 人次；组织最后 1 场国内研讨会议（参会 20 人左右），总结项目已取得的成果并探讨成果的应用前景和领域发展的新方向，完成项目结题与验收；最后，依托深圳市大数据研究院的通用软件开发平台发布大规模线性逆问题的算法软件；



项目成员张海斌（北工大）和徐加樑（北科大）来访研究院各1个月，参与项目最后阶段的研究（两位合作人来访的相关费用由依托单位资助）。

- 3) **预期成果：**发表学术论文2篇；会议论文1-2篇；形成实用的算法软件（深圳市大数据研究院通用软件研发部协助开发）；申请数据价值评估相关的专利1项及软件著作权1项；完成 Springer 英文专著1部；组织国内研讨会1次；完成项目综合科技报告1份。

项目总的时间、任务、进度安排大致见表 5.1 所示。

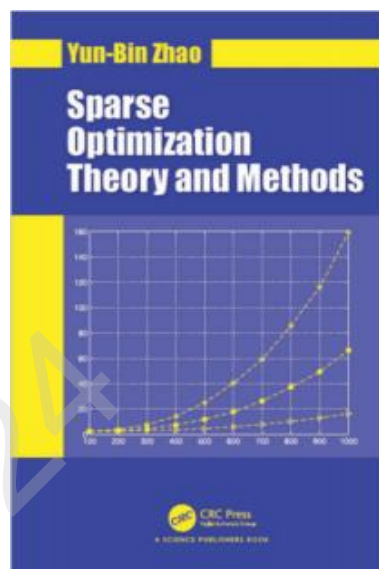
阶段	任务		2025/1-2025/12	2026/1-2026/12	2027/1-2027/12	2028/1-2028/12
1	T1	“极大列正交分裂”算法设计				
	T2	“非极大列正交分裂”算法设计				
	T3	理论分析，初步软件及应用				
2	T4	分裂-融合算法设计				
	T5	自适应分裂-融合算法设计				
	T6	理论分析，初步软件及应用				
3	T7	国定网格分裂-融合-共识算法				
	T8	自适应网格分裂-融合-共识算法				
	T9	理论分析，软件形成及应用				
	T10	网格划分与算法性能关系研究				
4	T11	SV 计算的逆问题模型				
	T12	SV 计算：分解算法微调及实现				
	T13	初步软件，与传统算法的比较				



（二）研究基础与工作条件

1. 研究基础（与本项目相关的研究工作积累和已取得的研究工作成绩）；

申请人在稀疏优化、压缩感知及线性逆问题的相关领域已取得了一系列具有国际影响力的研究成果，在国际著名的学术期刊上发表了多篇相关论文，并于 2018 年通过 CRC Press 出版了 1 本关于稀疏优化理论与算法的英文专著（见下面专著封面）。已有的工作积累为进一步开展本项目针对大规模线性逆问题的研究提供了有用的理论与算法涉及经验。



已有的主要工作积累简述如下：

1) 重新加权 L_1 -极小化的统一算法框架和对偶空间稠密化方法：

申请人（与香港中文大学合作）提出了稀疏优化的重新加权 L_1 -极小化方法的统一框架，并证明了统一框架算法的收敛性。成果发表在 *SIAM J. Optim.*，22 卷，2012。这类算法目前仍然是信号处理领域很多同行使用的算法之一。申请人进一步与伯明翰大学，剑桥大学和香港中文大学（深圳）的同行合作，提出了稀疏优化问题的“对偶空间稠密化算法”，一个原创性的凸优化型算法。主要结果发表在 *SIAM J. Optim.*，25 卷，2015 和 *Math. Oper. Res.*，42 卷，2017。

2) 压缩感知与稀疏优化理论研究：于 2013 年，申请人首次建立了在国际上被称为稀疏信号恢复的“RSP 压缩感知理论”，从转置感知矩阵的值空间角度刻画了几类凸优化算法的信号重构理论，主要工作发表于 *IEEE Trans. Signal Process.*，61 卷(22 期)，2013。后来又



将 RSP 理论推广到 1-比特(one-bit)压缩感知问题,发表在 *Sci. China Math.*, 2016, 59 卷, 2016。在此基础上, 申请人利用 RSP 建立了统计 Lasso 模型和 Dantzig Selector 等凸优化型算法的统一稳定性理论, 分别发表于 *Math. Oper. Res.*, 2019, 44 卷(第 1 期), 2019, *Optim. Method Software*, 35 卷(第 4 期), 2020, 以及 CRC Press 出版的专著(*Sparse Optimization Theory & Method*, 2018)。这些工作积累为完成本项目算法的理论分析部分供了一定的基础。

3) 最优阈值算法与加速技术: 申请人于 2019 年提出了最优阈值算法的概念和相应的算法类被称为 ROTPw, 打开了稀疏优化算法设计的一个新的研究方向。论文发表于 *SIAM J. Optim.*, 30 卷, 2020 和 *Signal Process.*, 187 卷, 2021。在此基础上, 申请人(与他人合作)对稀疏优化问题又提出了基于 Newton 型的迭代阈值算法, 发表于 *IEEE Trans. Signal Process.*, 68 卷, 2020; 最近(与他人合作)进一步研究了最优阈值算法的重球加速技术, 论文发表于 *Journal of Scientific Computing*, 96 卷 (no. 93), 2023。

4) 低复杂性自然阈值与匹配跟踪方法: 与香港中文大学(深圳)合作, 基于正则化的最优阈值模型和一阶近似方法, 构造了具有特殊结构的线性规来代替最优阈值算法中的二次凸优化子问题, 提出了一项原创性算法(称为“自然阈值算法”), 该算法大幅度地降低了计算复杂度。该成果发表于 *IEEE Open J. Signal Process.*, 3 卷, 2022。最近, 申请人还提出了动态的 OMP 算法, 突破了 OMP 的总迭代次数依赖于信号稀疏度的局限性, 克服了传统算法在迭代次数超过信号稀疏度时导致的过度匹配现象, 算法的求解性能明显优于 OMP 算法。成果发表于 *IEEE Open J. Signal Process.*, 4 卷, 2023。

申请人取得的上述相关成果为本项目的研究提供了良好的理论基础与算法设计经验。



其他工作基础：项目组成员之间也有一定的合作基础。申请人曾经与张海斌合作发表过一篇论文（Computational Optimization & Applications, 61 卷, 2015），也参加了张海斌主持的国家自然科学基金面上项目（“变量部分稀疏正则化算法设计与研究”，11771003, 2018/01-2021/12, 已结题）。申请人与徐加樑也合作发表过相关的 2 篇论文 (Optimization Methods & Software, 35 卷(4), 2020; Journal of Global Optimization, 81 卷, 2021)，因此也有良好的合作基础。

2. 工作条件（包括已具备的实验条件，尚缺少的实验条件和拟解决的途径，包括利用国家实验室、国家重点实验室和部门重点实验室等研究基地的计划与落实情况）；

深圳市大数据研究院具有一流的科研环境, 丰富的计算资源及强大的科研团队。在应用与计算数学、优化理论与算法、信号与图像处理、大数据与人工智能领域聚集了一批国际一流的专家和学者。在数学优化理论、算法及应用领域取得了具有国际水平的研究成果。研究院的研究人员获得了各种奖项, 包括 2 两位被当选为 SIAM 会士, 获 2 人次国际数学优化“P. Tseng”奖, 首届中国工业与应用数学学会“王选应用数学奖”, 以及华为重大应用成果奖。

深圳市大数据研究院拥有一流的计算资源。其计算平台中存储集群有 13 台存储服务器, 裸容量超过 2700TB SSD; CPU 集群的总 CPU 核数为 528 个, 内存超过 10TB, 单精度浮点算力能达到 60 TFLOPS; GPU 服务器 19 台, 其中有高性能显卡 A100 80 张和 V100S 40 张, GPU 单精度浮点算力达 2556 TFLOPS; AI 计算平台中 CPU 核数超过 416 个, 裸容量超过 730TB (SSD)。

研究院还拥有一支强大的数学优化算法通用软件的研发团队, 目



前已开发通用的 XOPT 优化求解器。我们的研究团队与腾讯、华为、中科院、鹏城实验室、和全国许多高校均有密切合作，还参与或者牵头成立了广东省大数据理论与计算重点实验室，中国与沙特阿拉伯联合实验室，以及与中科院数学与系统科学研究院和香港中文大学（深圳）三方联合成立了“刘徽实验室”（申请人担任实验室副主任）。同时，研究院还主办优化与数据科学相关领域的国际学术年会（例如 MIIS 年会等）。因此，研究院的一流资源和科研环境都为本项目的顺利开展提供了优越的工作条件和资源保证。

3. 正在承担的与本项目相关的科研项目情况（申请人和主要参与者正在承担的与本项目相关的科研项目情况，包括国家自然科学基金的项目和国家其他科技计划项目，要注明项目的资助机构、项目类别、批准号、项目名称、获资助金额、起止年月、与本项目的关系及负责的内容等）；

1) 申请人赵云彬主持国家自然科学基金面上项目, 12071307, “数据压缩与重构的新一代高性能优化算法”, 2021/01-2024/12。项目的研究内容包括：数据压缩算法在值域空间性质下的理论有效性；硬型阈值迭代算法和数据压缩子规划的加速技术；以及在矩阵空间上的推广及应用。负责的内容：在项目中全面负责算法的理论分析与算法设计，以及大部分数值分析。与本项目的关系：在研究内容上没有重叠部分，但取得的研究成果对本项目的算法设计部分有一定的启发作用。

2) 项目组成员曾奕程主持国家自然科学基金青年科学基金项目, 12301383, “高维数据的无监督型降维方法研究”, 2024/01-2026/12。项目的研究内容包括：运用随机矩阵理论对几类无监督降维方法对应的高维随机矩阵的特征值及特征向量进行分析；改进经典降维方法以



适用于高维数据；并分析新降维方法对若干后续统计分析的影响。负责的内容：全面负责无监督型降维方法的设计和理论分析。与本项目的关系：研究内容没有重叠部分，但处理高维统计数据的降维思想与方法可以对本项目的算法设计和SV的统计建模提供一定的理论指导。

3) 项目组成员徐加樑主持国家自然科学基金青年科学基金项目，12301393，“稀疏优化问题中的匹配追踪类和阈值类算法研究”、2024/01-2026/12。项目主要内容：基于一类复杂的非线性稀疏优化模型，设计基于对偶空间的匹配追踪类算法和加速类算法。负责的内容：全面负责算法设计及其应用。与本项目的关系：研究内容没有重叠部分，但其中的算法研究方法也为本项目的研究提供部分思路。

4. 完成国家自然科学基金项目情况（对申请人负责的前一个已资助期满的科学基金项目（项目名称及批准号）完成情况、后续研究进展及与本申请项目的关系加以详细说明。另附该项目的研究工作总结摘要（限500字）和相关成果详细目录）。

申请人负责的前一个结题项目名称：国家自然科学基金委，面上项目，“不确定数据优化问题的等价表示和计算复杂性”，10671199，2007/01-2009/12。该结题项目以及后续研究内容都与本申请项目没有关系（内容没有任何交叉重叠）。结题项目如期完成，预期的研究目标均已达到。

该结题项目工作总结摘要：主要研究成果包括（a）与加拿大多伦多大学(I. Averbakh)合作，解决了鲁棒二阶锥优化的凸松弛问题的有限维等价表示问题，为鲁棒二阶锥优化的近似求解提供了理论和算法支撑；（b）将不确定数据优化的方法引进样条函数的研究和计算，提出了插值节点位置不固定的样条函数的概念，与传统的样条函数相比



不再需要确定性的插值节点；(c) 受本基金资助，项目组成员王建方（中科院数学与系统科学研究院）于 2008-2009 年度通过科学出版社发行了一本中文专著，并研究了离散优化的 Incidence 有关的理论。

标注该项目资助的部分期刊文章如下：

- 1) I. Averbakh, Y.-B. Zhao, Explicit reformulation for robust mathematical optimization with general uncertainty sets, *SIAM J. Optim.*, 18, 1436-1466, 2008.
- 2) Y.-B. Zhao, Enlarging neighborhoods of interior-point algorithms for linear programming via least values of proximity functions, *Appl. Numer. Math.*, 57, 1033-1049, 2007.
- 3) Y.-B. Zhao, S.-C. Fang, J.E. Lavery, Geometric dual formulation for first derivative based univariate cubic L1 splines, *J. Global Optim.*, 40, 589-62, 2008.
- 4) Y.-B. Zhao, J. Hu, Global bounds for the distance to solutions of co-coercive variational inequalities, *Oper. Res. Lett.*, 35, 409-415, 2007.
- 5) I. Averbakh, S.-C. Fang, Y.-B. Zhao, Robust univariate cubic L2 splines: Interpolating data with uncertain positions of measurements, *J. Industrial & Manag. Optim.*, 5, 351-361, 2009.
- 6) I. Averbakh and Y.-B. Zhao, Relaxed robust second-order-cone programming, *Appl. Math. Comput.* 210, 387-397, 2009.
- 7) X. Chen, J. Hu, and X. Hu, The minimum risk spanning tree problem, *Lect. Note in Comput. Sci.*, 4616, 81-90, 2007.
- 8) Z. Zhang, B. Yao, J. Li, L. Liu, J. Wang, B. Xu, On incidence graphs. *Ars Combin.*, 87 (2008), 213-223.

后续影响：结题项目对后续关于计算凸分析方向的研究工作产生一定的影响：由于不确定数据优化的有限维等价优化问题涉及函数 Legendre-Fenchel 变换的计算，因此如何计算一个函数的 L-F 变换是



一个重要数学问题。负责人（赵云彬）后续开展了对这一变换的研究工作，其中最重要的一项后续成果是：成功地推导了乘积凸二次型的 Legendre-Fenchel 变换的显示公式，解决了优化与非线性分析领域（SIAM Rev., 2007 公布）的一个公开数学问题。该后续成果主要发表在 SIAM J Matrix Anal. & Appl., 31 卷, 1792-1811, 2010, Appl. Math. Optim., 62 卷, 411-434, 2010。这些结果受到法国数学家 J.-B. Hiriart-Urruty 的赞扬，被加拿大和国内一些学者使用并进一步对这些结果进行改进与提高。

（三）其他需要说明的情况

1. 申请人同年申请不同类型的国家自然科学基金项目情况（列明同年申请的其他项目的项目类型、项目名称信息，并说明与本项目之间的区别与联系）。

无

2. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在同年申请或者参与申请国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，申请或参与申请的其他项目的项目类型、项目名称、单位名称、上述人员在该项目中是申请人还是参与者，并说明单位不一致原因。

无

3. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在与正在承担的国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，正在承担项目的批准号、项目类型、项目名称、单位名称、起止年月，并说明单位不一致原因。

无

4. 其他。

无



赵云彬（BRID: 07663.00.05927）简历

2024版

深圳市大数据研究院，深圳国际工业与应用数学中心，教授

教育经历：

- (1) 1995-09 至 1998-03，中科院数学与系统科学研究院，运筹学与控制论，博士
- (2) 1989-09 至 1992-06，重庆大学，应用数学，硕士
- (3) 1985-09 至 1989-06，西北工业大学，应用数学，学士

博士后工作经历：

- (1) 1998-06 至 2001-03，中科院计算数学与科学工程计算研究所
- (2) 1999-01 至 2001-03，香港中文大学，系统工程系

科研与学术工作经历（博士后工作经历除外）：

- (1) 2020-02 至今，深圳市大数据研究院，深圳国际工业与应用数学中心，教授
- (2) 2013-04 至 2020-02，英国伯明翰大学，数学学院，副教授
- (3) 2007-09 至 2013-03，英国伯明翰大学，数学学院，讲师
- (4) 2003-04 至 2007-08，中科院数学与系统科学研究院，应用数学研究所，副研究员
- (5) 2001-04 至 2003-03，中科院数学与系统科学研究院，应用数学研究所，助理研究员

曾使用其他证件信息：

- (1) 6101*****618

近五年主持或参加的国家自然科学基金项目/课题：

- (1) 国家自然科学基金委员会，面上项目，12071307，数据压缩与重构的新一代高性能优化算法，2021-01-01 至 2024-12-31，52万元，在研，主持

近五年主持或参加的其他科研项目/课题（国家自然科学基金项目除外）：

无

代表性研究成果和学术奖励情况（填写代表性论文时应根据其发表时的真实情况如实规范列出所有作者署名，并对本人署名情况进行标注，包括：①作者署名按姓氏排序；②唯一第一作者；③共同第一作者；④唯一通讯作者；⑤共同通讯作者；⑥其他情况）：

一、代表性论著（请在“申请书详情”界面，点开“人员信息”-“代表性成果”卡片查看对应的全文）：

(1) Yun-Bin Zhao ; Optimal k-thresholding algorithms for sparse optimization problems, *SIAM Journal on Optimization*, 2020, 30(1): 31-55 (期刊论文) (本人标注：唯一第一作者)

(2) Yun-Bin Zhao; Zhi-Quan Luo ; Analysis of optimal thresholding algorithms for compressed sensing, *Signal Processing*, 2021, 187: 108148 (期刊论文) (本人标注：唯一第一作者，唯一通讯作者)

(3) Yun-Bin Zhao; Houyuan Jiang; Zhi-Quan Luo ; Weak stability of l1-minimization methods in sparse data reconstruction, *Mathematics of Operations Research*, 2019, 44(1): 173-195 (期刊论



文) (本人标注: 唯一第一作者)

(4) Nan Meng; **Yun-Bin Zhao** ; Newton-step-based hard thresholding algorithms for sparse signal recovery, *IEEE Transactions on Signal Processing*, 2020, 68: 6594-6606 (期刊论文) (本人标注: 唯一通讯作者)

(5) Zhong-Feng Sun; Jin-Chuan Zhou; **Yun-Bin Zhao** ; Heavy-ball-based optimal thresholding algorithms for sparse linear inverse problems, *Journal of Scientific Computing*, 2023, 96: 93 (期刊论文) (本人标注: 其他情况)

二、论著之外的代表性研究成果和学术奖励:

无

NSFC 2024



张海斌（BRID: 09663.00.31760）简历

2024版

北京工业大学，数学统计学与力学学院，教授

教育经历：

- (1) 1999-09 至 2002-06，中国农业大学管理工程学院，运筹学与最优化，博士
- (2) 1996-09 至 1999-07，北京工业大学应用数学系，运筹学与最优化，硕士
- (3) 1983-09 至 1987-07，山西师范大学数学系，数学，学士

博士后工作经历：

无

科研与学术工作经历（博士后工作经历除外）：

- (1) 2010-12 至今，北京工业大学，数学统计学与力学学院，教授
- (2) 2002-06 至 2010-12，北京工业大学，应用数理学院，副教授
- (3) 1987-07 至 1996-09，山西农业大学，基础部，讲师

曾使用其他证件信息：

无

近五年主持或参加的国家自然科学基金项目/课题：

无

近五年主持或参加的其他科研项目/课题（国家自然科学基金项目除外）：

无

代表性研究成果和学术奖励情况（填写代表性论文时应根据其发表时的真实情况如实规范列出所有作者署名，并对本人署名情况进行标注，包括：①作者署名按姓氏排序；②唯一第一作者；③共同第一作者；④唯一通讯作者；⑤共同通讯作者；⑥其他情况）：

一、代表性论著（请在“申请书详情”界面，点开“人员信息”-“代表性成果”卡片查看对应的全文）：

- (1) 张海斌；张凯丽；凸优化理论与算法，科学出版社，2020（学术专著）
- (2) Kai Tu; Haibin Zhang; Huan Gao; Junkai Feng; A hybrid Bregman alternating direction method of multipliers for the linearly constrained difference-of-convex problems, *Journal of Global Optimization*, 2020. 秋季, 76(4): 665-693（期刊论文）（本人标注：其他情况）
- (3) T. Giday Woldu; Haibin Zhang; Xin Zhang; Y. Hailu Fissuh; A Modified Nonlinear Conjugate Gradient Algorithm for Large-Scale Nonsmooth Convex Optimization, *Journal of Optimization Theory and Applications*, 2020, 185(1): 223-238（期刊论文）（本人标注：其他情况）
- (4) Yuqi Zhang; Haibin Zhang; Yingjie Tian; Sparse multiple instance learning with non-convex penalty, *Neurocomputing*, 2020, 391(1): 142-156（期刊论文）（本人标注：共同第一作者）
- (5) Jinyao Ma; Haibin Zhang; Shanshan Yang; Jiaojiao Jiang; A tight bound of modified iterative hard thresholding algorithm for compressed sensing, *Applications of Mathematics*, 2023, 68(5): 623-642（期刊论文）（本人标注：其他情况）



二、论著之外的代表性研究成果和学术奖励：

无

NSFC 2024



曾奕程（BRID: 01895.00.56381）简历

2024版

深圳市大数据研究院，深圳国际工业与应用数学中心，副研究员

教育经历：

- (1) 2016-09 至 2019-08, 香港浸会大学, 统计学, 博士
- (2) 2014-09 至 2016-06, 浙江大学, 统计学, 硕士
- (3) 2010-09 至 2014-07, 天津大学, 数学与应用数学, 学士

博士后工作经历：

- (1) 2019-11 至 2022-02, 多伦多大学

科研与学术工作经历（博士后工作经历除外）：

- (1) 2022-06 至今, 深圳市大数据研究院, 深圳国际工业与应用数学中心, 副研究员

曾使用其他证件信息：

无

近五年主持或参加的国家自然科学基金项目/课题：

- (1) 国家自然科学基金委员会, 青年科学基金项目, 12301383, 高维数据的无监督型降维方法研究, 2024-01-01 至 2026-12-31, 30万元, 在研, 主持

近五年主持或参加的其他科研项目/课题（国家自然科学基金项目除外）：

- (1) 深圳市科技创新委员会, 优秀科技创新人才培养（博士基础研究启动）, RCBS20221008093336086, 高维数据的充分降维方法研究, 2023-04 至 2025-04, 30万元, 在研, 主持

代表性研究成果和学术奖励情况（填写代表性论文时应根据其发表时的真实情况如实规范列出所有作者署名，并对本人署名情况进行标注，包括：①作者署名按姓氏排序；②唯一第一作者；③共同第一作者；④唯一通讯作者；⑤共同通讯作者；⑥其他情况）：

一、代表性论著（请在“申请书详情”界面，点开“人员信息”-“代表性成果”卡片查看对应的全文）：

- (1) Xin Chen; **Yicheng Zeng**; Siyue Yang; Qiang Sun ; Sketched Ridgeless Linear Regression: The Role of Downsampling, *The 40th International Conference on Machine Learning*, Honolulu, Hawaii, USA, 2023-07-23至2023-07-29 (会议论文) (本人标注: 共同第一作者)
- (2) **Yicheng Zeng**; Lixing Zhu ; Order determination for spiked type models, *Statistica Sinica*, 2022, 32: 1633-1659 (期刊论文) (本人标注: 唯一第一作者)
- (3) **Yicheng Zeng**; Lixing Zhu ; Order determination for spiked-type models with a divergent number of spikes, *Computational Statistics & Data Analysis*, 2023, 182: 107704 (期刊论文) (本人标注: 唯一第一作者)
- (4) Junshan Xie; **Yicheng Zeng**; Lixing Zhu ; Limiting laws for extreme eigenvalues of large-dimensional spiked Fisher matrices with a divergent number of spikes, *Journal of Multivariate Analysis*, 2021, 184: 104742 (期刊论文) (本人标注: 共同第一作者)
- (5) Fangchen Yu; **Yicheng Zeng**; Jianfeng Mao; Wenye Li ; Online estimation of similarity



matrices with incomplete data, *The Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, Pittsburgh, PA, USA, 2023-07-31至2023-08-04 (会议论文) (本人标注: 其他情况)

二、论著之外的代表性研究成果和学术奖励:

无

NSFC 2024



徐加樑 (BRID: 08192.00.31356) 简历

2024版

北京科技大学, 数理学院, 讲师

教育经历:

- (1) 2014-09 至 2019-02, 英国伯明翰大学, 运筹优化, 博士
- (2) 2012-09 至 2013-12, 英国伯明翰大学, 应用数学, 硕士
- (3) 2008-09 至 2012-06, 天津财经大学, 统计学, 学士

博士后工作经历:

- (1) 2019-11 至 2022-07, 中国科学院数学与系统科学研究院

科研与学术工作经历 (博士后工作经历除外):

- (1) 2022-08 至今, 北京科技大学, 数理学院, 讲师

曾使用其他证件信息:

无

近五年主持或参加的国家自然科学基金项目/课题:

- (1) 国家自然科学基金委员会, 青年科学基金项目, 12301393, 稀疏优化问题中的匹配追踪类和阈值类算法研究, 2024-01-01 至 2026-12-31, 30万元, 在研, 主持

近五年主持或参加的其他科研项目/课题 (国家自然科学基金项目除外):

- (1) 中国博士后科学基金会, 博士后基金, YJ20200197, 博士后国际交流计划引进项目, 2020-08 至 2022-08, 60万元, 结题, 主持

代表性研究成果和学术奖励情况 (填写代表性论文时应根据其发表时的真实情况如实规范列出所有作者署名, 并对本人署名情况进行标注, 包括: ①作者署名按姓氏排序; ②唯一第一作者; ③共同第一作者; ④唯一通讯作者; ⑤共同通讯作者; ⑥其他情况):

一、代表性论著 (请在“申请书详情”界面, 点开“人员信息”-“代表性成果”卡片查看对应的全文):

- (1) Jialiang Xu; Yun-Bin Zhao ; Stability analysis of a class of sparse optimization problems, *Optimization Methods and Software*, 2020, 35(4): 836-854 (期刊论文) (本人标注: 唯一第一作者)

- (2) Jialiang Xu; Yun-Bin Zhao ; Dual-density-based reweighted ℓ_1 -algorithms for a class of ℓ_0 -minimization problems, *Journal of Global Optimization*, 2021, 81(3): 749-772 (期刊论文) (本人标注: 唯一第一作者)

- (3) Jialiang Xu ; Nonuniqueness of Solutions of a Class of ℓ_0 -minimization Problems, *Journal of the Operations Research Society of China*, 2021, 9(4): 893-908 (期刊论文) (本人标注: 唯一第一作者, 唯一通讯作者)

- (4) Jialiang Xu; Xu Zhang ; Data-Time Tradeoffs for Optimal k-Thresholding Algorithms in Compressed Sensing, *2022 IEEE International Symposium on Information Theory (ISIT)*, Aalto University in Espoo, Finland, 2022-6-26至2022-7-1 (会议论文) (本人标注: 唯一第一作者)



二、论著之外的代表性研究成果和学术奖励：

无

NSFC 2024



附件信息

序号	附件名称	备注	附件类型
1	代表性论著1	PDF 全文	代表性论著
2	代表性论著2	PDF 全文	代表性论著
3	代表性论著3	PDF 全文	代表性论著
4	代表性论著4	PDF 全文	代表性论著
5	代表性论著5	PDF 全文	代表性论著

NSFC 2024



项目名称： 大规模线性逆问题的分裂-融合-共识算法及在数据价值计算中的应用
资助类型： 面上项目
申请代码： A0405. 连续优化

国家自然科学基金项目申请人和参与者承诺书

为了维护国家自然科学基金项目评审公平、公正，共同营造风清气正的科研生态，本人**在此郑重承诺**：严格遵守《中华人民共和国科学技术进步法》《国家自然科学基金条例》《关于进一步加强科研诚信建设的若干意见》《关于进一步弘扬科学家精神加强作风和学风建设的意见》《关于加强科技伦理治理的意见》以及科技部、自然科学基金委关于科研诚信建设有关规定和要求；申请材料信息真实准确，不含任何涉密信息或敏感信息，不含任何违反法律法规或违反科研伦理规范的内容；在国家自然科学基金项目申请、评审和执行全过程中，恪守职业规范和科学道德，遵守评审规则和工作纪律，杜绝以下行为：

- (一) 抄袭、剽窃他人申请书、论文等科研成果或者伪造、篡改研究数据、研究结论；
- (二) 购买、代写申请书；购买、代写、代投论文，虚构同行评议专家及评议意见；购买实验数据；
- (三) 违反成果发表规范、署名规范、引用规范，擅自标注或虚假标注获得科技计划等资助；
- (四) 在项目申请书中以高指标通过评审，在项目计划书中故意篡改降低相应指标；
- (五) 以任何形式打听或散布尚未公布的评审专家名单及其他评审过程中的保密信息；
- (六) 本人或委托他人通过各种方式和途径联系有关专家进行请托、游说、“打招呼”，违规到评审会议驻地窥探、游说、询问等干扰评审或可能影响评审公正性的行为；
- (七) 向工作人员、评审专家等提供任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包，或提供宴请、旅游、娱乐健身等任何可能影响评审公正性的活动；
- (八) 违反财经纪律和相关管理规定的行为；
- (九) 其他弄虚作假行为。

如违背上述承诺，本人愿接受国家自然科学基金委员会和相关部门做出的各项处理决定，包括但不限于撤销科学基金资助项目，追回项目资助经费，向社会通报违规情况，取消一定期限国家自然科学基金项目申请资格，记入科研诚信严重失信行为数据库以及接受相应的党纪政务处分等。

申请人签字：

编号	参与者姓名 / 工作单位名称（应与加盖公章一致） / 证件号码	签字
1	张海斌 / 北京工业大学 / 1*****4	
2	曾奕程 / 深圳市大数据研究院 / 4*****2	
3	徐加樑 / 北京科技大学 / 1*****2	
4		
5		
6		
7		
8		
9		



项目名称： 大规模线性逆问题的分裂-融合-共识算法及在数据价值计算中的应用
资助类型： 面上项目
申请代码： A0405. 连续优化

国家自然科学基金项目申请单位承诺书

为了维护国家自然科学基金项目评审公平、公正，共同营造风清气正的科研生态，**本单位郑重承诺**：申请材料中不存在违背《中华人民共和国科学技术进步法》《国家自然科学基金条例》《关于进一步加强科研诚信建设的若干意见》《关于进一步弘扬科学家精神加强作风和学风建设的意见》《关于加强科技伦理治理的意见》以及科技部、自然科学基金委关于科研诚信建设有关规定和要求的情况；申请材料符合《中华人民共和国保守国家秘密法》和《科学技术保密规定》等有关法律法规和规章制度要求，不含任何涉密信息或敏感信息；申请材料不含任何违反法律法规或违反科研伦理规范的内容；申请人符合相应项目的申请资格；依托单位、合作研究单位、申请人及主要参与者不在限制申报、承担或参与财政性资金支持的科技活动的期限内；在项目申请和评审活动全过程中，遵守有关评审规则和工作纪律，杜绝以下行为：

（一）以任何形式打听或公布未公开的项目评审信息、评审专家信息及其他评审过程中的保密信息，干扰评审专家的评审工作；

（二）组织或协助申请人/参与者向工作人员、评审专家等给予任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包等；宴请工作人员、评审专家，或组织任何可能影响科学基金评审公正性的活动；

（三）支持、放任或对申请人/参与者抄袭、剽窃、重复申报、提供虚假信息（含身份和学术信息）等不当手段申报国家自然科学基金项目疏于管理；

（四）支持或协助申请人/参与者采取“打招呼”“围会”等方式影响科学基金项目评审；

（五）其他违反财经纪律和相关管理规定的行为。

如违背上述承诺，本单位愿接受自然科学基金委和相关部门做出的各项处理决定，包括但不限于停拨或核减经费、追回项目已拨经费、取消本单位一定期限国家自然科学基金项目申请资格、记入科研诚信严重失信行为数据库以及主要责任人接受相应党纪政务处分等。

依托单位公章：

日期： 年 月 日

合作研究单位公章：

日期： 年 月 日

合作研究单位公章：

日期： 年 月 日