

模型、数据与知识驱动地球物理反问题及智能计算

王彦飞^{1,2,3*}

1. 中国科学院地质与地球物理研究所深层油气理论与智能勘探开发重点实验室, 北京 100029

2. 中国科学院大学地球与行星科学学院, 北京 100049

3. 中国科学院地球科学研究院, 北京 100029

* 通讯作者, E-mail: yfwang@mail.iggcas.ac.cn

2025-04-25收稿, 2026-01-05修回, 2026-02-04接受, 2026-05-26网络版发表

国家自然科学基金项目(42450232、42342062)和国家重点研发计划项目(2024YFA1012304)资助

摘要 地球物理反问题是利用地表观测数据反推地下介质的物性和结构, 广泛应用于资源能源探测、地质调查和环境监测等领域。传统的地球物理反演依赖于数学物理模型的构建与高效数值计算。然而, 由于数学问题的不适定性, 以及地下介质的复杂性, 仅依赖于模型驱动的计算方法通常带来多解性和计算不稳定等问题。为应对这些挑战, 数据驱动方法利用大量观测数据, 通过统计分析和机器学习技术揭示数据中的潜在模式, 提高反演结果的稳定性和精度。随着计算技术的进步, 知识驱动的智能计算方法通过结合领域知识和先验信息, 将物理模型与数据分析结合, 提升反问题求解的稳定性并增强反演结果的物理合理性和计算效率。此外, 量子计算(QC)的兴起也为地球物理反问题提供了新的解决途径。QC利用量子力学的特性, 能够在处理复杂计算任务时实现并行计算和指数级加速。对于地球物理反问题, QC的潜力在于提高计算效率和处理大规模数据集的能力, 从而实现更快速和更精确的反演。模型、数据与知识驱动的地球物理反问题及智能计算代表了一种综合创新的研究范式, 充分利用物理模型、观测数据和领域知识, 通过智能计算技术实现对地下结构的精准探测。这不仅推动了地球物理反问题研究的进展, 也为实际应用提供了更为可靠和高效的解决方案。

关键词 反问题, 地球物理, 正则化, 最优化, 统计策略, 数据驱动, 智能计算, 量子计算

1 前言

反问题在整个自然科学领域具有特殊重要性, 它是与数学、物理、化学、地学、生物学、金融、商业、生命科学、计算技术与工程相关的交叉学科。反问题及其高效数值计算是近年来的研究热点, 是应用学科与数学建模结合的新兴交叉学科方向。地球物理反演是在地球物理学中利用地球表面观测到的物理现象推测地球内部介质物理状态的空间变化及物性结构的一个分支。地球物理的核心问题之一是反演, 即利用地表或地层的地球物理资料来研究地球内部的结构, 并且相关的技术在油气勘探领域有广泛的应用(刘光鼎, 2018)。

数学问题的适定性条件: (1) 解是存在的; (2) 解是

唯一的; (3) 解是稳定的。反之, 若上述3个条件中, 至少有一个不能满足, 则称其为不适定的。反问题通常是不适定的, 即给定问题解未必存在; 解未必唯一; 以及解的连续性/稳定性不一定能够保证。反问题的不适定性由Hadamard于1923年提出; 1943年, A. N. Tikhonov在《苏联科学院院刊》中发表了著名的文章“关于反问题的稳定性”, 该成果是现代不适定问题理论的开始。

求解不适定问题的基本技巧是利用Tikhonov正则化思想衍生的各种方法, 包括基于模型驱动的变分正则化、离散正则化、迭代正则化方法、最优化计算方法、统计推理, 以及基于数据驱动的人工智能方法(特别是基于神经网络的深度学习等)。

目前, 人工智能在地球物理学中的应用已取得了一

中文引用格式: 王彦飞. 2026. 模型、数据与知识驱动的地球物理反问题及智能计算. 中国科学: 地球科学, doi: 10.1360/SSTe-2025-0113

英文引用格式: Wang Y. 2026. Model, data, and knowledge-driven geophysical inverse problems and intelligent computing. Science China Earth Sciences, <https://doi.org/10.1007/s11430-025-1868-y>

些重要进展。其中,最显著的是在地震学领域(Araya-Polo等, 2018; Ross等, 2018; Wang J等, 2019; Zhu和Beroza, 2019; Mousavi和Beroza, 2020; Kuang等, 2021; Jo和Ha, 2024)。地震探测是一种重要的资源勘探手段,以地震波传播和反射为基础,通过分析地震数据来推断地下结构。利用人工智能技术对地震数据进行处理和解释,可以提高勘探效率和精度。例如,通过深度学习算法可以对地震资料进行自动识别、定位和拾取,同时还可以实现地震相分类和特征提取等功能,减轻人工勘探的工作量;通过卷积神经网络(CNN)和循环神经网络(RNN)的结合,可以有效地对地震波形数据进行分类和分析;通过U-Net网络深度学习(Ronneberger等, 2015)可以高效地进行土地覆盖分类、道路和建筑物提取及地震剖面、岩相或断层检测与分割;还有研究者利用生成对抗网络(GAN)等方法进行地震数据合成和模拟,以实现更真实的地下结构表达;此外,基于神经网络的地震数据预测模型也能够在一定程度上帮助探测地震的发生时间和强度。除了地震勘探,人工智能技术还广泛应用于其他地球物理数据处理、重建和反演等方面,为勘探和解释提供更精确的结果。比如,在电磁勘探中,基于CNN的方法被应用于电磁数据处理、特征提取和解释,从而实现更准确的地下结构识别和解释(Puzyrev, 2019; Wu X等, 2022; Sun H F等, 2023);还有研究者将CNN用于重磁或电磁数据的联合反演,进一步提高勘探的准确性(Wei, 2023; Bai等, 2024; 王彦飞等, 2026)。人工智能技术还扩展到了其他领域,如地热勘探和地质灾害预测等(Wang等, 2018)。

从计算角度而言,由于地球物理反演往往涉及海量的数据分析和矩阵操作,对算力的要求极高,所以一直是高性能计算领域的一大挑战。作为科技创新的基本工具之一,高性能计算为科学研究与技术应用提供必不可少的算力保障,广泛应用于油气勘探、气象分析、灾害预报、宇宙探索、材料/药物设计等众多领域。传统的超级计算机主要依赖于大规模集成电路技术。然而,由于摩尔定律的限制,集成电路可容纳的晶体管数目已接近饱和,从而极大限制了超算算力的进一步提升。此外,超级计算机的高能耗及其导致的温室气体排放也是技术发展的重大瓶颈。以目前国内最快的超算“神威·太湖之光”为例,其一年的用电量达到15兆瓦(<https://www.top500.org>)。因此,世界各国亟需能够克服这些瓶颈的下一代高性能计算技术。

量子计算(QC)是一个非常有前景的技术应用领域,可以用来求解反问题。QC理论诞生于20世纪80年代初

(Benioff, 1980; Feynman, 1982)。与以比特(bit)为基础的传统计算机(电子计算机)不同, QC系统以符合量子力学性质的量子比特(qubit)为基础,能够以指数级的速度进行计算,从而成为下一代高性能计算最有希望的候选者之一。2019年,美国Google团队宣布利用53个量子比特的悬铃木(Sycamore)超导处理器,在200秒内完成了当时最快超算需要10000年才能完成的随机采样任务,从而揭开了对QC优越性的挑战(Arute等, 2019)。2024年12月9日, Google量子AI公司推出了全新量子芯片Willow, 拥有105个量子比特,通过将多个物理量子比特组合成一个逻辑量子比特,为实现实用QC提供了途径,且随着量子比特数量增加,逻辑错误率呈指数级下降,该公司还同时利用量子回声算法证实Willow芯片具有可验证的量子优势(Abanin等, 2025; Acharya等, 2025)。中国在玻色采样计算方面,相较于传统计算机,利用76个光子和113个光子在九章光子计算机上分别实现了 10^{14} 和 10^{24} 倍的加速(Zhong等, 2020, 2021)。

地球物理反演是根据地球物理观测数据来推断地下地质体的物理参数和几何结构的过程。在人工智能的背景下,传统的反演方法得到了迅速改进。例如,基于贝叶斯推理的反演方法,结合了概率论和先验知识,能够更全面地考虑地球物理数据的不确定性。在地震波数据的自动识别和分类中, CNN可以有效地提取地震波的波形特征。通过对大量标注的地震波数据(如不同类型的地震事件、噪声等)进行训练, CNN能够自动学习地震波的特征模式,从而实现对新地震波数据的准确分类。此外,大模型领域广泛应用的Transformer模型在地球物理反演中也展现出显著的应用潜力。其自注意力机制能够有效捕捉复杂地质数据的长程依赖关系,提高数据解释的准确性。在反演过程中, Transformer可以处理多尺度(multiscale, MS)、多模态的地球物理数据(如地震波形数据、电磁数据等),实现更精确的地下结构成像。此外, Transformer的并行计算能力显著加速了大规模数据的处理,使实时反演成为可能。

本文将从模型驱动的地球物理反问题、数据驱动的地球物理反问题、一般反问题的计算、智能计算、QC等多个角度展开综述,最后给出结论与发展展望。文章的整体框架图如下图1所示。

2 模型驱动的地球物理反问题

2.1 积分方程

重、磁反演方法是利用矿石和围岩之间的密度或磁性等差异,引起重力、磁场强度变化而定性和定量地预

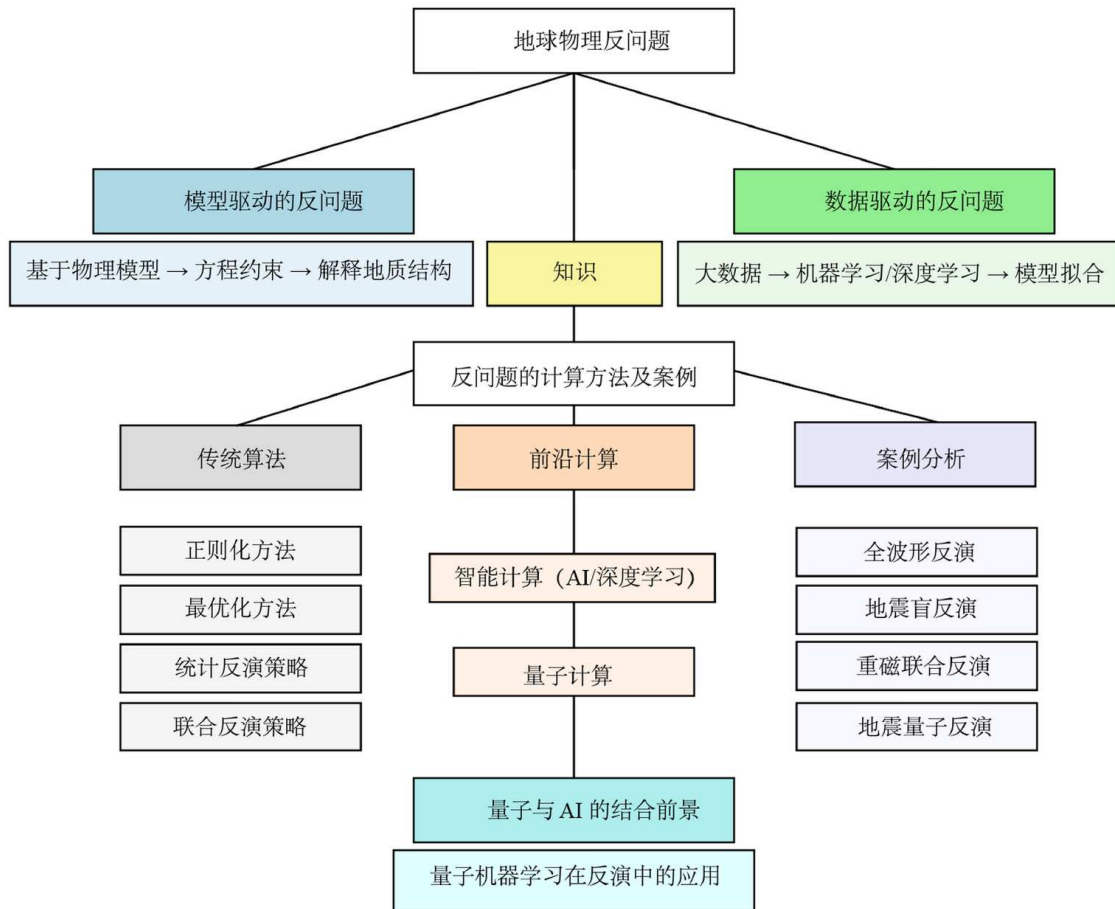


图1 地球物理反问题计算方法框架示意图

测地下岩矿石分布的方法。在模型驱动的重磁数据联合反演过程中，核心方法大致一致，主要通过优化目标函数来迭代求解最佳的地下模型。在处理具体的地球物理问题时，可以通过一组特定的参数来定量描述某个地质体，这包括但不限于密度、磁性、弹性或电性等物性参数，这些参数被统称为模型参数。这些模型参数中，有许多并不能被直接测量。对于那些必需但无法直接测得的参数，我们通常先测量一些其他可观测的参数，并确保这些可观测参数与所需的模型参数之间存在可靠的定量关联。这种方法不仅能够帮助我们间接地获取难以直接测量的物性参数，还能通过建立的参数关系模型，进一步精确地反演出地下结构的详细特征。重磁正问题的积分方程可以统一描述为第一类Fredholm算子方程(Zhdanov, 2002; 王彦飞等, 2011)

$$\iiint_V K(x, y, z; \xi, \eta, \varsigma) m(\xi, \eta, \varsigma) d\xi d\eta d\varsigma = d(x, y, z) \quad (1)$$

其中， $K(x, y, z; \xi, \eta, \varsigma)$ 为积分方程的核，是定义在自变量集合 V 上的连续函数； $d(x, y, z)$ 为已知的连续函数，该函

数定义了方程(1)的非齐次性(如果 $d \equiv 0$ ，那么方程被称为齐次方程)； $m(\xi, \eta, \varsigma)$ 为待求的连续函数。

第一类Fredholm算子方程描述的问题是不适定的(ill-posed)。反问题指的是给定观测数据 d 和核函数 K ，求取未知模型或参数 m 。由于不适定性(解的存在性、唯一性及稳定性至少一条不被满足)，通常用一簇与原问题相临近的适定问题的解，去逼近原问题的解，即Tikhonov正则化(肖庭延等, 2003)。

不同的观测系统，可以由不同的积分方程来描述。以磁法勘探为例，观测数据可以是总场数据(total magnetic intensity)、矢量场数据(vector field)或全张量磁梯度场数据(full tensor magnetic gradient)。我们以全张量磁梯度场数据反演的数学建模做一个简要的说明。全张量磁梯度场数据($\mathbf{B}_{ij}$)对应的积分表达式为(Wang Y F等, 2019; 王彦飞等, 2021)

$$\mathbf{B}_{\text{场偶极子}} = \frac{\mu_0}{4\pi} \iiint_V \left[\frac{3(\mathbf{M} \cdot \mathbf{r})\mathbf{r}}{r^5} - \frac{\mathbf{M}}{r^3} \right] d\mathbf{v} \quad (2)$$

$$\mathbf{B}_{ii} = \frac{\mu_0}{4\pi} \iint_V \left[\frac{6m_i(i-p_i)}{r^5} + \frac{3(\mathbf{M} \cdot \mathbf{r})}{r^5} - \frac{15(\mathbf{M} \cdot \mathbf{r})(i-p_i)(i-p_i)}{r^7} \right] dv \quad (3)$$

$$\mathbf{B}_{ij} = \frac{\mu_0}{4\pi} \iiint_V \left[\frac{3m_i(j-p_j)}{r^5} + \frac{3m_j(i-p_i)}{r^5} - \frac{15(\mathbf{M} \cdot \mathbf{r})(i-p_i)(j-p_j)}{r^7} \right] dv \quad (4)$$

令 $\mathbf{M} \in W_2^2(P)$, $\mathbf{B} \in L_2(Q)$, 并设算子 $\mathbf{A}$ (其核函数为 $\mathbf{K}$) 是一对一的映射。定义右端和待求问题解的范数分别为

$$\|\mathbf{B}\|_{L_2} = \sqrt{\|B_x\|_{L_2}^2 + \|B_y\|_{L_2}^2 + \|B_z\|_{L_2}^2 + \|B_{xx}\|_{L_2}^2 + \|B_{xy}\|_{L_2}^2 + \|B_{xz}\|_{L_2}^2 + \|B_{yz}\|_{L_2}^2 + \|B_{zz}\|_{L_2}^2}$$

$$\|\mathbf{M}\|_{W_2^2} = \sqrt{\|M_x\|_{W_2^2}^2 + \|M_y\|_{W_2^2}^2 + \|M_z\|_{W_2^2}^2}$$

记 $\bar{\mathbf{B}}$ 和 $\mathbf{A}$ 为不带误差的观测和观测系统 (算子), 它们的逼近值分别为 $\mathbf{B}_\delta$ 和 $\mathbf{A}_h$, 使得 $\|\mathbf{B}_\delta - \bar{\mathbf{B}}\|_{L_2} \leq \delta$,

$$\|\mathbf{A} - \mathbf{A}_h\|_{W_2^2 \rightarrow L_2} \leq h.$$

Tikhonov 正则化方法指的是极小化下面的泛函:

$$J^\alpha[\mathbf{M}] = \|\mathbf{A}_h \mathbf{M} - \mathbf{B}_\delta\|_{L_2}^2 + \alpha \|\mathbf{M}\|_{W_2^2}^2 \quad (5)$$

其中, $\alpha > 0$ 是正则化参数, 上述吉洪诺夫泛函有唯一的极小值 $\mathbf{M}_\eta^\alpha$, $\eta = \{\delta, h\}$, 使得 $J^\alpha[\mathbf{M}]$ 达到最小。具体的数值实现可以通过第5节描述的计算方法得到。

2.2 波动方程

2.2.1 声波、弹性波及随机波动方程

地震正演模拟就是在已知地下结构模型的情况下, 根据地震波在地下介质中的传播机理, 通过数值模拟方法求解该地下结构模型在假定激发源作用下的地震响应。声波方程通常表达为下面方式 (Aki 和 Richards, 2002; Carcione, 2007; Liang 等, 2018):

$$\frac{1}{K} \frac{\partial^2 p}{\partial t^2} = \frac{\partial}{\partial x} \left(\frac{\partial p}{\rho \partial x} \right) + \frac{\partial}{\partial y} \left(\frac{\partial p}{\rho \partial y} \right) + \frac{\partial}{\partial z} \left(\frac{\partial p}{\rho \partial z} \right) + src(t) \quad (6)$$

其中, ρ 表示密度, p 表示压强, K 是体积模量, 且有 $K = \lambda + 2\mu = \rho v^2$, 这里 v 是波速, λ 和 μ 是拉梅常数 (Lamé constants), $src(t)$ 为震源函数。有限差分法以其原理简单、需要的内存小、健壮性强、容易并行、计算时间短、能适应剧烈变化的地下介质状况等优点被广泛应用于地震正演、反演中 (Alford 等, 1974; Kelly 等, 1976; Virieux, 1986; Levander, 1988; Chen, 2007)。地震模拟软件中常用的方法是采用交错网格的形式对公式 (6) 差分离散化, 比如在 x 、 y 和 z 方向采用相同的有限差分 (finite difference) 系数, 并将 z 轴的向下方向设为正方向; 对于时间求

导, 可通过时间导数的二阶近似得到。一般来讲, 差分系数方程的求解可以采用以下几种形式: 伪谱法、Taylor 展开法, 优化的方法 (如最小二乘法、模拟退火方法等) (Etgen, 2007; Liu 和 Sen, 2009; Virieux 等, 2011; Zhang 和 Yao, 2013; Liang 等, 2018; Teixeira 等, 2023), 以及 Tikhonov 正则化方法 (Wang 等, 2014)。

弹性动力学三大基本方程分别描述了位移、应力和应变两两之间的关系:

$$(\lambda + 2\mu) \nabla (\nabla \cdot \mathbf{u}) - \mu \nabla \times (\nabla \times \mathbf{u}) + \mathbf{f} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (7)$$

其中, 位移 $\mathbf{u}$ 表示为 (u_x, u_y, u_z) , ρ 是介质的密度, λ 和 μ 是拉梅常数, 描述了材料的弹性性质, $\mathbf{f}$ 是震源函数, ∇ 是梯度算子, $\nabla \cdot$ 是位移场的散度, $\nabla \times$ 是位移场的旋度。实际地下介质是三维情况, 直接利用二维弹性波方程无法完全模拟地震波在实际地层的传播的情况。同样, 通过交错网格有限差分近似, 可以获得三维弹性波场的数值模拟 (孙卫涛, 2009)。

数值稳定性和网格频散是有限差分方法中的两个最基本的问题, 与时间离散的步长和空间离散的网格大小有关, 直接影响到波动方程数值模拟的应用和效果。时间和空间频散是由对时间和空间偏导数的网格化造成的, 相速度变成了网格间距的函数, 这会导致地震波的数值相速度不等于地球介质的真实相速度, 使得波场模拟的精度降低。一般来说, 如果存在时间频散, 则高频的相速度增大; 如果存在空间频散, 则高频的相速度减小 (Dablain, 1986)。为了压制频散, 可以采用时间-空间域优化的方法, 或线性正则化方法确定交错网格有限差分系数。

随机波动方程 (stochastic wave equation) 是一种特殊的随机偏微分方程 (stochastic partial differential equation, SPDE)。用于描述具有随机性或不确定性的波动现象, 比如随机源、随机介质、局部扰动环境中的声波/弹性波

传播(Bao等, 2010; Zhang和Zhang, 2013)。随机波动方程一般形式可以写成

$$\partial_t^2 u - v^2 \Delta u = F(u, \partial_t u) + \sigma(u, \partial_t u)(\dot{W})(t, \mathbf{x}) \quad (8)$$

其中, $u(\mathbf{x}, t)$ 为波场, v 为波传播速度, Δ 为空间Laplace算子, $F(\cdot, \cdot)$ 和 $\sigma(\cdot, \cdot)$ 分别为可能依赖于 u 或其导数的外力作用或噪声强度, $\dot{W}$ 为时空噪声 ($W(t, \mathbf{x})$ 表示一个随机过程, 如Brownian场)。

由于地球内部介质的复杂性和不均匀性, 波速和介质特性常常具有随机性, 具体体现在以下4个方面:

(1) 介质不均匀性: 地下介质是一个不同尺度的散射体(如裂缝、孔洞等小尺度随机结构), 其属性(如密度和弹性参数)很少是均匀的。这些属性可以在空间和时间上随机变化, 从而影响声波的速度和衰减。

(2) 环境扰动: 温度、压力和湿度等环境条件会在声波的传播中引入随机性。这在海洋勘探中是常见的现象。目前陆上地质勘探走向深层-超深层, 特别是在超深层, 温度梯度会迅速变化, 从而导致波速的随机变化。

(3) 震源或介质的不确定性: 在地球物理应用中, 地震源或地下特性的不确定性会显著影响波的传播。随机声波方程有助于对这些不确定性进行建模, 从而提供更准确的波行为预测。

(4) 复杂的边界条件: 地质环境通常具有复杂的边界, 可能会以不可预测的方式散射和反射波。比如在新能源勘探中, 地质氢气(或天然氢气)的边界条件就很难界定。在波动方程中引入随机项可以对这些复杂交互作用的影响进行建模。

上述诸多不确定性因素表明, 对系统认知的不完备性必须通过不确定性建模。随机技术的引入能更有效地耦合不同尺度间的相互作用, 尤其在尺度跨度较大时。

举一个简单的例子。设 $\eta(\mathbf{x}, t)$ 是一个随机过程(如白噪声或更复杂的随机场), 通常用来表示外部随机扰动。则随机波动方程可以形式化为

$$\frac{1}{c^2(\mathbf{x})} \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} - \nabla^2 u(\mathbf{x}, t) = \text{src}(\mathbf{x}, t) + \eta(\mathbf{x}, t) \quad (9)$$

其中, $u(\mathbf{x}, t)$ 是位移或压力场, $c(\mathbf{x}) = c_0 + \delta c(\mathbf{x})$ 表示随机环境扰动下的速度模型(c_0 为背景速度, $\delta c(\mathbf{x})$ 为描述介质不均匀性的小扰动项, 通常假定为零均值的随机场, 且可由一定的自相关函数或谱密度来刻画其统计特性), $\text{src}(\mathbf{x}, t)$ 是震源, $\eta(\mathbf{x}, t)$ 如上所述是随机扰动。上述表达式反映的是介质大尺度不均匀性(如沉积层速度变化、构造变形等), 它主要控制相位延迟和射线路径弯曲, 是几何光学级别的走时扰动。

在地震探测中, 如果有其他随机扰动因素(如局部微

结构、随机散射中心或与地层品质因子 Q 相关的吸收效应), 类似量子力学薛定谔方程中的势场项, 这时会有一个额外的随机势场加入。这时, 随机波动方程可以表示为

$$\frac{1}{c^2(\mathbf{x})} \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} - \nabla^2 u(\mathbf{x}, t) + V(\mathbf{x})u(\mathbf{x}, t) = \text{src}(\mathbf{x}, t) + \eta(\mathbf{x}, t) \quad (10)$$

其中, $V(\mathbf{x})$ 是一个额外的随机势场, 表示介质中存在的随机散射特性或其他非均匀扰动, 是波动级别的散射、干涉或衰减, 形式上可写成

$$V(\mathbf{x}) = V_0(\mathbf{x}) + \delta V(\mathbf{x}) \quad (11)$$

其中, $V_0(\mathbf{x})$ 表示已知的背景势(有时可简化为零), $\delta V(\mathbf{x})$ 表示随机扰动部分。假如 $V(\mathbf{x})$ 是零均值随机场, 即 $\mathbb{E}[V(\mathbf{x})] = 0$, $\mathbb{E}[V(\mathbf{x})V(\mathbf{x}')] = \sigma_V^2 \gamma(|\mathbf{x} - \mathbf{x}'|; \ell_c)$, 其中 σ_V^2 为方差, ℓ_c 为相关长度, $\gamma(\cdot)$ 为相关函数(如高斯、指数衰减等), 则可以通过Karhunen-Loève展开、快速傅里叶变换(fast Fourier transform, FFT)频域滤波法, 或者数值卷积法生成满足特定功率谱密度的随机场(Levy和Rubinstein, 1999; Grigoriu, 2006)。对于地球深部小尺度介质, 可以近似为大量微小散射体组成(数学上按某种分布(如泊松、分形)随机布置), 这些散射体的局部波速扰动经过尺度化可等效为点势或有限扩展势, 叠加得到 $V(\mathbf{x})$ 。吸收与品质因子 Q 通常会引起波幅衰减与相位变化, 可将其等效到一个复数形式的势场, $V(\mathbf{x}) = V_R(\mathbf{x}) + iV_I(\mathbf{x})$, $V_I > 0$ 表示吸收, $Q^{-1}(\mathbf{x})$ 的变化会直接决定 $V_I(\mathbf{x})$ 的分布强度与尺度。对于随机势场参数敏感性, 从总体上讲, 随机势场的引入会影响波的振幅、波形、衰减、散射和传播速度; 从频率角度上讲, 高频波对小尺度扰动更敏感, 低频波只受大尺度的统计特征影响。更一般的随机势场参数敏感性, 需要从势场强度、相关长度、势场的频谱特性及各向同性与各向异性(如裂缝或方向性结构等复杂介质)4个方面考虑。当然, 数值上还得从全局和局部敏感性方面分析。

关于随机波动方程的正问题的求解, 在确定性数值方法中嵌入随机采样(如对随机场进行概率空间离散), 与通常的偏微分方程(partial differential equation, PDE)求解一样, 采用时间和空间离散化的有限差分法、有限元法(需求解刚度矩阵和质量矩阵)、高精度的谱元法、边界积分法或Monte Carlo模拟(通过大量随机样本的统计分析来估计波动响应)来求解。

2.2.2 分数阶微分方程

分数阶微积分在黏弹性介质、流体力学、反常扩散,

以及软物质研究等领域都有广泛应用。与整数阶导数相比,分数阶导数具有全局相关性,能较好地体现系统函数发展的历史依赖过程,并且分数阶导数模型往往使用较少几个参数就可以与实验结果很好地吻合。因此,对于复杂多场多相衰减地层介质,应用分数阶微分方程可以获得更加精确的模拟。

(1) 时间分数阶波动方程

时间分数阶黏声波动方程是由常 Q 模型准确推导得来的,定义为

$$\frac{\partial^{2-2\gamma}\sigma}{\partial t^{2-2\gamma}} = c^2 \omega_0^{-2\gamma} \nabla^2 \sigma \quad (12)$$

其中, $c^2 = c_0^2 \cos^2(\pi\gamma/2)$, ∇^2 为拉普拉斯算子,根据 $Q=1/\tan(\pi\gamma)$, $2-2\gamma$ 可以为介于1到2之间的任意实数($0 < \gamma < 0.5$, $Q > 0$)。利用其进行地震波场模拟的关键问题是时间分数阶导数的离散计算,式(12)中的空间整数阶导数可以用FFT或者有限差分法计算。根据分数阶理论,可以采用不同的定义方法来逼近分数阶导数。[Carcione等\(2002\)](#)使用Grunwald-Letnikov有限差分法计算时间分数阶导数,采用具有二阶精度的中心差分方法,其形式如下:

$$\frac{\partial^\beta f(t)}{\partial t^\beta} = f_h^{(\beta)}(t) \approx \frac{1}{h^\beta} \sum_{j=0}^J (-1)^j C(\beta, j) f\left[t + \left(\frac{\beta}{2} - j\right)h\right] \quad (13)$$

其中, β 是分数阶导数的阶次,且 $\beta=2-2\gamma$, $1 < \beta < 2$, h 是时间采样间隔,求和上限 $J=t/h-1$, $C(\beta, j)$ 是分数二项式系数,根据伽马函数性质 $\Gamma(x+1)=x \cdot \Gamma(x)$ 可以展开为

$$\begin{aligned} C(\beta, j) &= \frac{\beta(\beta-1)(\beta-2)\cdots(\beta-j+1)}{j!} \\ &= \frac{\Gamma(\beta+1)}{j! \Gamma(\beta-j+1)} = \frac{\Gamma(\beta+1)}{\Gamma(j+1) \Gamma(\beta-j+1)} \end{aligned} \quad (14)$$

利用式(13)计算时间分数阶导数需要保存过去所有的 f 值,即使对记忆长度 J 进行截断,依然会有较大的内存和计算量需求。

将式(13)带入式(12)可以得到时间分数阶波动方程在 $t=n\Delta t$ 时刻的离散格式为

$$\sigma^{n+\beta/2} = h^\beta c^2 \omega_0^{-2\gamma} \nabla^2 \sigma^n - \sum_{j=1}^J (-1)^j C(\beta, j) \sigma^{n-j+\beta/2} \quad (15)$$

其中, $\sigma^{n+\beta/2}$ 的上角标代表时间为 $t=(n+\beta/2) \cdot h$,计算该时刻的波场需要 $n\Delta t$ 时刻的波场 σ^n ,由于 $\beta \approx 2$,可以由 $\sigma^{n-1+\beta/2}$ 近似代替,空间拉普拉斯(Laplace)算子采用FFT计算。

(2) 空间分数阶波动方程

公式(12)中时间分数阶导数的计算会消耗大量内存和计算量,为避免这个问题,可以将时间分数阶导数转换为空间分数阶导数。[Carcione\(2010\)](#)提出了含分数阶拉普拉斯的黏声波动方程,如(16)所示

$$\omega_0^{2-2\beta} c_0^{2\beta} \left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial z^2} \right)^\beta p + s = \frac{\partial^2 p}{\partial t^2} \quad (16)$$

其中, P 为压力场, s 为震源, ω_0 为参考频率, $\gamma = \arctan(1/Q)/\pi$ 且 $\gamma=1-1/\beta$, $1 \leq \beta \leq 2$ 。 $\beta=1$ 时,式(16)退化为完全弹性声波方程; $\beta=2$ 时,式(16)为无限衰减介质。描述黏弹性介质时, β 一般略大于1。时间离散可以采用有限差分法,空间分数阶拉普拉斯可以采用快速傅里叶变换离散计算,这有效减少了计算量和内存。

[Carcione\(2010\)](#)针对空间分数阶波动方程(16)进行数值计算,其中分数阶拉普拉斯算子 $(\partial^2/\partial x^2 + \partial^2/\partial z^2)^\beta$ 可以利用傅里叶变换的微分性质进行计算,即:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} \right)^\beta p(x, z) = F^{-1} \left\{ (-1)^\beta (k_x^2 + k_z^2)^\beta F[p(x, z)] \right\} \quad (17)$$

其中, F 和 F^{-1} 分别代表傅里叶正变换和傅里叶反变换,采用中心差分法进行时间二阶导数计算,可以得到在 $t=n\Delta t$ 时,式(16)的离散格式为

$$\begin{aligned} p^{n+1} &= \omega_0^{2-2\beta} c_0^{2\beta} h^2 F^{-1} \left\{ (-1)^\beta (k_x^2 + k_z^2)^\beta F[p^n] \right\} \\ &\quad + 2p^n - p^{n-1} \end{aligned} \quad (18)$$

其中, k_x 和 k_z 分别为 x 和 z 方向的波数。

[Zhu和Harris\(2014\)](#)通过频散关系式近似的方法,提出了振幅衰减和速度频散项解耦的空间分数阶拉普拉斯黏声波动方程。即把时间分数阶导数转换为空间分数阶拉普拉斯算子,时间偏导数为二阶,可以导出一个近似常 Q 的黏声波动方程:

$$\begin{cases} \frac{\partial v(\mathbf{x}, t)}{\partial t} = \frac{1}{\rho(\mathbf{x})} \nabla \sigma(\mathbf{x}, t) + f \\ \frac{\partial}{\partial t} e(\mathbf{x}, t) = \nabla \cdot v(\mathbf{x}, t) \\ \sigma(\mathbf{x}, t) = M_0(x) \left[\eta(\mathbf{x}) (-\nabla^2)^{\gamma(\mathbf{x})} \right. \\ \quad \left. + \tau(\mathbf{x}) \frac{\partial}{\partial t} (-\nabla^2)^{\gamma(\mathbf{x})-1/2} \right] e(\mathbf{x}, t) \end{cases} \quad (19)$$

在振幅衰减和频散解耦的波动方程(19)中,一阶时间偏导数通过有限差分法计算,仅需要保存前一个时刻的波场值。一阶空间偏导数使用伪谱法进行计算,空间分数阶拉普拉斯算子使用广义的傅里叶变换进行计算,即:

$$\begin{aligned} (-\nabla^2)^\gamma e(\mathbf{x}, t) &= F^{-1} \{ k^{2\gamma} F[e(\mathbf{x}, t)] \} \\ (-\nabla^2)^{\gamma-1/2} \frac{\partial}{\partial t} e(\mathbf{x}, t) &= F^{-1} \left\{ k^{2\gamma-1} F \left[\frac{\partial}{\partial t} e(\mathbf{x}, t) \right] \right\} \end{aligned} \quad (20)$$

其中, $k = \sqrt{k_x^2 + k_z^2}$ 为波数,一阶时间和空间导数同时被

离散在交错网格上，波动方程(19)的离散格式为(Ning和Wang, 2021)

$$\begin{cases} v_x^{n+1/2} = v_x^{n-1/2} + \frac{dt}{\rho} F^{-1} \{ i k_x \cdot e^{i k_x \cdot dx/2} \cdot F[\sigma^n] \} \\ v_z^{n+1/2} = v_z^{n-1/2} + \frac{dt}{\rho} F^{-1} \{ i k_z \cdot e^{i k_z \cdot dz/2} \cdot F[\sigma^n] \} \\ \varepsilon^{n+1} = \varepsilon^n + dt \left\{ F^{-1} \{ i k_x \cdot e^{-i k_x \cdot dx/2} \cdot F[v_x^{n+1/2}] \} \right. \\ \quad \left. + F^{-1} \{ i k_z \cdot e^{-i k_z \cdot dz/2} \cdot F[v_z^{n+1/2}] \} \right\} \\ \sigma^{n+1} = M_0 [\eta F^{-1} \{ k^{2\gamma} \cdot F[\varepsilon^{n+1}] \} \\ \quad + \tau F^{-1} \{ k^{2\gamma-1} \cdot F[(\varepsilon^{n+1} - \varepsilon^n) / dt] \}] \end{cases} \quad (21)$$

其中， dx 和 dz 是 x 和 z 方向的空间采样间隔， dt 是时间采样间隔，上角标 n 代表波场在 $n \cdot dt$ 时刻；应力 σ 和应变 ε 被离散在整网格点， v_x 和 v_z 离散在半网格点。式(21)相比于式(15)虽然形式复杂，但仅需要保留当前时刻和过去一个时刻的波场信息即可，减少了计算量和内存需求。

2.2.3 波动方程反问题

波动方程反问题主要指的是通过反演地下岩性物性参数进行地下构造成像。其中，偏移(migration)是重要的一类成像方法。偏移成像问题本质上属于反问题的范畴，通过正演算子的伴随算子来充分逼近其逆算子，比如，把正演算子 L 的伴随算子 L^T 作用到观测数据 d_{obs} ，这本身就可以看作不动点迭代算法的第一步。逆时偏移(reverse time migration, RTM)是非常精确的一类成像方法。在叠前RTM中，正演算子与其伴随算子的核心就是波动方程的数值求解方法，而有限差分法又是数值求解方法中最常用的。当采用精度不够的有限差分法求解波场时，必然造成其伴随算子中产生严重的误差，不能充分逼近其逆算子。因此，从反问题角度理解偏移成像问题，需要有高精度的伴随算子。采用高精度的地震波有限差分数值模拟方法可以有效改善伴随算子的计算精度，从而提高RTM成像的效果。互相关成像条件具有简单易于实现的特点，是目前使用最广泛的成像条件(Claerbout, 1971; Guitton等, 2007)。其定义为

$$I(\mathbf{x}) = \sum_{s=1}^{ns} \int_0^{t_{\max}} dt \sum_{g=1}^{ng} p_s(\mathbf{x}, t; \mathbf{x}_s) p_g(\mathbf{x}, t; \mathbf{x}_g) \quad (22)$$

其中， $I(\mathbf{x})$ 是在点 $r=[x, y, z]$ 处的成像值， $p_s(\mathbf{x}, t; \mathbf{x}_s)$ 和 $p_g(\mathbf{x}, t; \mathbf{x}_g)$ 分别代表了正向波场和逆向波场在 $\mathbf{x}$ 点 t 时刻的值， $\mathbf{x}_s$ 和 $\mathbf{x}_g$ 代表炮点和检波点的位置。 ns 和 ng 分别代表炮和检波器的个数。

全波形反演理论上是利用来自地下的全波场信息

(走时、相位、振幅，以及透射波、反射波、回折波)来反演地下介质参数，相比于传统的初至波走时层析、叠加速度分析，以及偏移速度分析等方法，具有更高的分辨率。全波形反演通过极小化实际观测数据 $\mathbf{d}_{obs}$ 和模拟数据 $\mathbf{d}_{cal}$ 之间的残差来更新迭代速度模型。目标函数为

$$E(\mathbf{m}) = \frac{1}{2} \sum_{ns} \sum_{nr} \sum_{nt} \|\mathbf{d}_{cal} - \mathbf{d}_{obs}\|^2 \quad (23)$$

其中， $\mathbf{m}$ 为地下介质参数(如速度 v_p)， ns 为总的炮点数目， nr 为每一炮的检波器数目， nt 为每一道的时间采样点数， $\mathbf{d}_{obs}$ 和 $\mathbf{d}_{cal}$ 分别为观测数据和模拟数据。以声波为例，在二维情况下， $\mathbf{d}_{cal}$ 由下述的声波方程模拟得到：

$$\frac{1}{v_p(\mathbf{x})} \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} - \nabla^2 u(\mathbf{x}, t) = \delta(x_x, x_z) f(\mathbf{x}, t) \quad (24)$$

其中， u 为声压场， $\nabla^2 = \partial^2 / \partial x^2 + \partial^2 / \partial z^2$ 为拉普拉斯算子， (x_x, x_z) 为震源位置， $f(\mathbf{x}, t)$ 为震源函数。

对于弹性波方程，地震数据对不同参数的敏感性在特定的传播角度上具有相似性，因此，在多参数同时反演过程中，参数之间存在相互干涉(crosstalk)，并相互耦合，这增加了反演问题的非线性(Pan等, 2020)。针对参数间的串扰问题，Tarantola(1986)利用弹性波方程在不同参数化下辐射模式的分析给出了一种有效的弹性波反演参数化策略：对近偏移距数据(短波长成分)来说，阻抗是相对较好的参数化方式，而对远偏移距数据(长波长成分)来说，速度是比较好的参数化方式。

对于随机波动方程反演，与确定性反问题相比，随机反问题在既有挑战之外还存在更多本质性困难，这主要源于随机性和不确定性。例如，经典反问题的解具有确定性特征，而随机反问题的解是随机函数，因此针对随机性特定实现求解的意义有限。常用的蒙特卡罗模拟在计算统计量时，其计算资源需求往往比相应经典反问题高出数个数量级。因此，相关应用领域亟需发展新的模型与方法论。比如，使用多种神经网络模型生成初始速度模型；使用基于神经网络的PDEs求解器(Raissi等, 2019)，但如何提高物理信息引导的神经网络(physics-informed neural networks, PINNs)方法的计算效率仍然值得探讨；将随机微分方程的求解问题转化为随机优化问题求解(deep BSDEs)(E等, 2017; Han等, 2018)；考虑扩散过程(diffusion process)的随机微分方程反问题建模及计算(Meng和Kabashima, 2022; Boys等, 2023; Song等, 2023; Dou和Song, 2024; Zhang等, 2025)；以及考虑融合朗之万动力学(Langevin dynamics)和牛顿法的SNIPS算法(Kawar等, 2021)等。

2.3 麦克斯韦方程

电磁方法勘探(electromagnetic exploration)是一种基于地下岩层的电导率、磁导率、电化学反应和介电差异的地球物理找矿技术(何继善和鲍力知, 1990; 李貅, 2002; 汤井田和何继善, 2005)。它通过观测人工或天然电磁场的时间变化及其空间分布特征来进行资源勘查和探测地球内部结构。电磁勘探可分为两大类: 时域电磁(TDEM)和频域电磁(FDEM)。瞬变电磁法(TEM)是一种时域的人工源电磁技术(Sun等, 2022; Sun H F等, 2023)。其基本原理是用不接地线圈或接地电极作为电源发送脉冲电流作为激励信号。地下导电介质在激励下产生初级脉冲场, 随后由于电磁感应产生时变的二次感应场。二次场的响应速度与波形与地质体的形态、结构及电学参数有关。通过在主脉冲之间的时间段测量纯二次场的衰减响应, 可以获取地下介质的电性分布(何继善和鲍力知, 1990)。由于TEM对低阻体高度敏感且体积效应小, 它在金属矿产勘查、煤矿水文地质调查与工程勘察中具有重要价值(李貅, 2002)。尤其在地下流体动力学监测方面, TEM展现出强大的应用潜力——垃圾填埋场渗滤液泄漏、浅层地下水系统恢复、海水入侵、采矿诱发的地下水渗流, 以及页岩气压裂监测(Ray等, 2013)。

电磁法勘探所依赖的理论基础就是麦克斯韦方程组(Maxwell equations), 它描述了电场 $\mathbf{E}$ 与磁场 $\mathbf{H}$ (或磁感应强度 $\mathbf{B}$)随时间和空间的变化关系, 以及它们与电荷分布和导电介质性质的联系。在电磁法勘探中, 通常采用频率域或时域形式的麦克斯韦方程组, 并在地球物理介质假设下作一定简化(如准静态近似)。考虑磁偶极子作为激发源, 在麦克斯韦方程中引入磁电流密度 $\mathbf{J}_m$ 和磁电荷密度 ρ_m 。磁电流密度 $\mathbf{J}_m$ 表示由磁偶极子产生的等效磁电流, 其方向沿着偶极子轴线。麦克斯韦方程组可以写作

$$\begin{cases} \sigma \mathbf{E} + \frac{\partial \mathbf{D}}{\partial t} = \nabla \times \mathbf{H} \\ -\mu \frac{\partial \mathbf{H}}{\partial t} - \mathbf{M}_0 = \nabla \times \mathbf{E} \\ \nabla \cdot \mathbf{B} = 0, \nabla \cdot \mathbf{D} = 0 \\ \mathbf{B} = \mu \mathbf{H}, \mathbf{D} = \varepsilon \mathbf{E} \end{cases} \quad (25)$$

其中, $\mathbf{E}$ 是电场强度(V/m), $\mathbf{H}$ 是磁场强度(A/m), $\mathbf{B}$ 是磁通密度(Wb/m²), $\mathbf{D}$ 是电位移矢量(C/m²), σ 是电导率(S/m)。在各向同性介质中, 物质关系为 $\mathbf{D}=\varepsilon \cdot \mathbf{E}$ 、 $\mathbf{B}=\mu \cdot \mathbf{H}$, 其中 ε 是介电常数(F/m), μ 是磁导率(H/m)。

上述方程组的正演计算方面, 可以通过对时间偏导数的二阶中心差分和对空间变量偏导数的四阶中心差分格式实现。其中, 电场分量和磁场分量在空间上分别交

错排列(如图2)。在边界条件处理上, 地面和空气边界可以使用简单的向上连续边界条件, 其他边界则应用吸收边界条件, 以确保电磁场模拟的准确性。

电磁场反问题有很多种数据类型, 比如直流电阻率(DC resistivity)、TEM、大地电磁法(magnetotellurics, MT)、感应极化(IP)、电磁测深(FDEM)、井中电磁测井等。反演实现上大体分TDEM/FDEM反问题两大类。以时域TEM为例, 可以通过最小化Wasserstein度量来建立TEM的反演方案, 实现最优传输, 核心是求解下面的最优化问题(Sun等, 2022):

$$\begin{aligned} W_2^2(f, g) &= \frac{1}{2} \inf_{T \in \mathcal{M}} \int_X \|x - T(x)\|^2 f(x) dx \\ \text{s.t., } \det(D^2 u(x)) &= \frac{f_X(x)}{g_Y(T(x))} \\ \nabla u(X) &= Y = T_{\text{opt}} \end{aligned} \quad (26)$$

其中, 公式(26)的第一式为二次Wasserstein度量, 第二式就是所谓的蒙日-安培(Monge-Ampère)方程, $u(x)$ 是Brenier势函数(未知的凸函数), $D^2 u(x)$ 表示 u 的Hessian矩阵, $\det(D^2 u(x))$ 是Hessian的行列式; $f(x)$ 是源分布密度函数, $g(y)$ 是目标分布密度函数(对应着观测数据), $\nabla u(x)$ 给出最优传输映射。该问题中的Monge-Ampère方程微分运算通过中心差分方法解决, Wasserstein度量的Fréchet梯度是通过应用伴随方程法和反向传播电磁场理论计算的。

可以证明, 在欧几里得空间的凸域 $\mathbb{R}^+$ 上, 存在一个凸函数 u (称为Brenier势函数), 它依赖于分布函数 f 和 $g(f: X \rightarrow \mathbb{R}^+, g: Y \rightarrow \mathbb{R}^+)$ 。函数 u 的梯度给出了最优传输的映射解 T_{opt} , 并且它同时满足Monge-Ampère方程(Brenier, 1991)。数值上, 基于Wasserstein度量的梯度迭代算法具有非单调收敛特征, 这一点与确定性优化的非单调梯度下降法非常类似, 因而都具有加快收敛的效果。

比如考虑一个三维地质体, 观测点和源位于地表和坐标原点。在该模型中, 有一个低电阻率异常体, 其真实电阻率为10 Ω m。异常体在空间中的位置如图3A所示。背景电阻率为1000 Ω m。三维网格划分为21×21×10, 完美匹配层网格为10。单位网格长度为10 m, 初始模型为半空间, 电阻率为1000 Ω m。模拟数据 E_x , E_y 和 H_z 通过前述的有限差分法获得。利用上述方法, 可以给出反演结果(图3B)和收敛性曲线(图3C), 验证了基于Wasserstein度量极小化的有效性。

2.4 小结

模型驱动的地球物理反问题是通过物理模型、观测数据和数学算法来反演地下介质的结构与特性, 广泛应用于石油矿产资源探测等领域。这类反问题通常涉及复

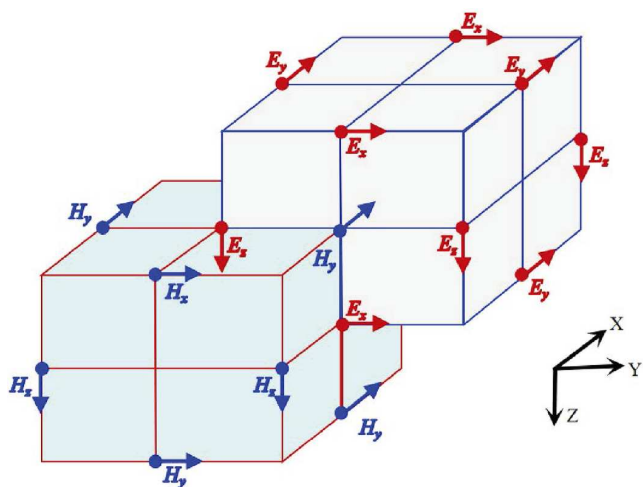


图2 麦克斯韦方程组交错网格离散化, 据Sun等(2022)

杂的数学模型和数值计算。建模中的范数选取方面除了常见的 L_2 、 L_1 、 L_q 和全变差(TV)度量, 还可以考虑 W_p^m 和Cauchy/Huber/熵/KL散度/Wasserstein度量等。如果考虑到问题的特殊性(如地质特征、图像特征), 还可以考虑几何正则化(如图、流形等, 这个在机器学习中经常用到)。表1给出了前述的几类反问题的概述。未来, 随着计算技术和数据处理方法的不断进步, 反演的精度将进一步扩展。如何处理不确定性、噪声及数据缺失问题, 仍将是该领域的研究重点。

3 数据驱动地球物理反问题

近年来, 随着计算机技术和算法研究的飞速发展, 基于数据驱动的地球物理反演方法, 特别是机器学习相

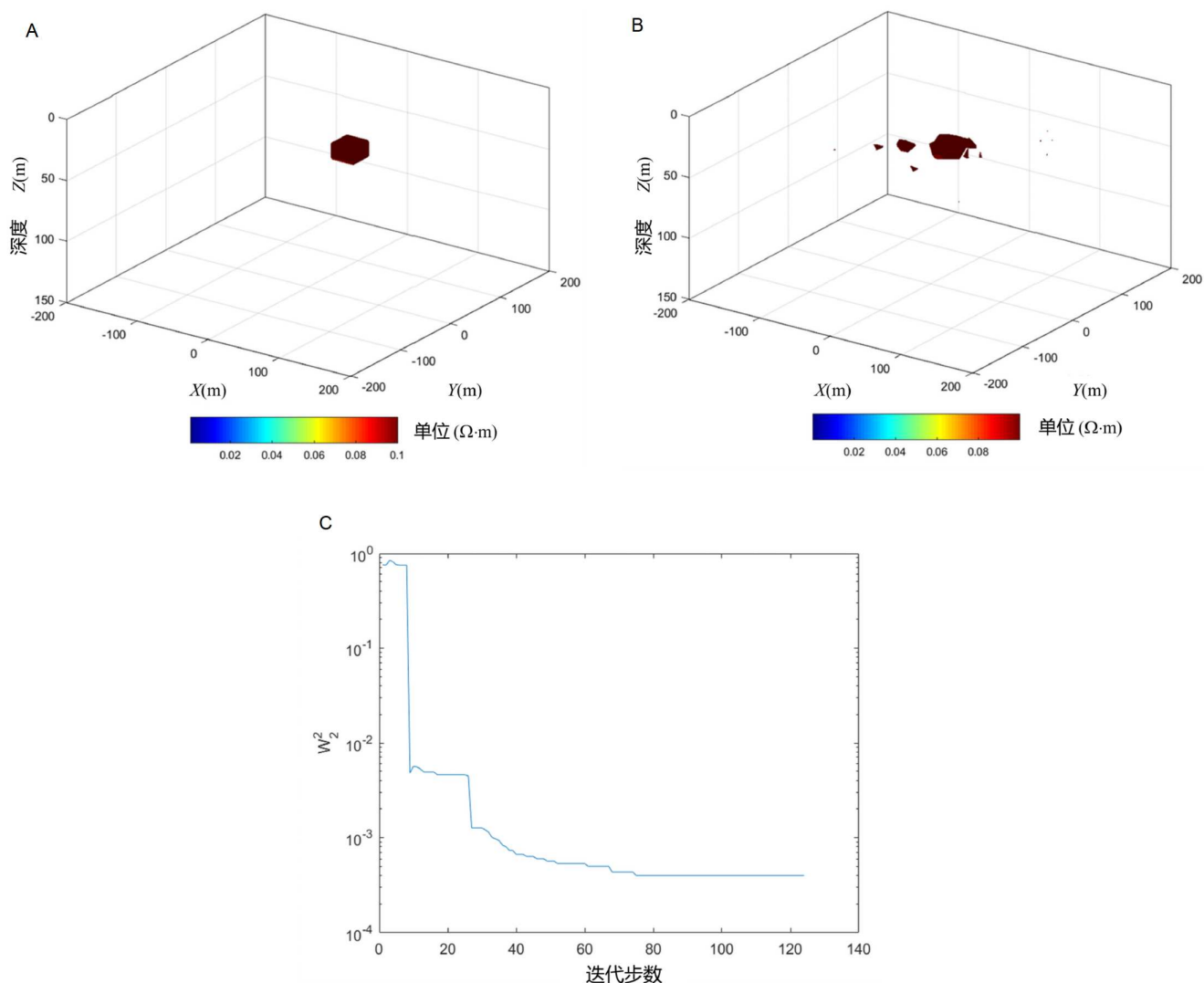


图3 (A) 三维地质模型; (B) 基于Wasserstein度量极小化的反演结果; (C) 目标泛函的收敛曲线。改自Sun等(2022))

表1 模型驱动地球物理反问题

基于模型的反问题	方法/技术	优点	缺点	计算复杂度
积分方程反问题	重磁勘探	大尺度地下结构反演, 处理复杂的几何与介质	反演精度可能较低	低
声波方程反问题	地震勘探	数学模型简单, 计算效率高, 地下空间探测、矿产勘探等	对于深层复杂结构, 受噪声干扰影响较大	中
弹性波方程反问题	深层地震勘探	处理复杂介质, 精度较高, 适用于超深层油气等的勘探	数值求解难度较大, 计算资源需求高	高
随机波动方程反问题	随机介质建模、非确定性地质反演	处理不确定性与复杂介质, 适用于考虑随机性的地下勘探	数值解法较为复杂	高
分数阶微分方程反问题	非均匀介质反演	描述不均匀、分形、复杂的地下介质行为, 如天然气水合物	数值求解难度大, 算法较复杂	高
麦克斯韦方程反问题	电磁勘探	处理复杂电磁数据, 时频域结合可实现高精度探测	求解过程计算量大, 需要高效的数值求解方法	高

关的方法, 开始受到越来越多学者的关注。这也为地球物理反问题的解决提供了新途径和新方法。与传统基于模型的反演方法相比, 基于机器学习的地球物理反演方法不需要复杂的先验信息, 一旦网络模型经过良好的训练之后, 网络在预测阶段能够以极快的速度推导出较为准确的反演结果。然而, 这些方法的一个主要挑战是如何获取足够的高质量训练数据集, 以确保模型训练的准确性和可靠性。

地学数据具有大数据的特征: (1) 数据采集昂贵, 数据内容“不可完全重复”(value); (2) 这些数据在不同地学学科中存在较大的差异(variety); (3) 数据具有高度不确定性(观测手段、环境等各种影响)(veracity); (4) 数据分析具有高度计算复杂性(不同物性、不同属性), 数据还具有高维特性(volume); (5) 数据速率也变化较大, 如地震数据、电磁数据、随季节变化的遥感遥测数据等(velocity)。

从广义的反问题角度来讲, 从上述大数据中挖掘知识就是一类特殊的反问题。基于数据驱动的地球物理反问题主要为分类/识别和预测/回归。分类/识别任务通常采用如CNN等深度学习模型来对地球物理数据进行特征提取和分类; 预测/回归任务通常采用回归网络(如全连接神经网络、深度神经网络等)来进行地下结构或物理属性学习和预测。近年来, 利用人工智能技术, 基于数据驱动的固体地球物理及地球系统科学得到了越来越多的重视(Bergen等, 2019; Reichstein等, 2019; Wu等, 2020; Wang和Wang, 2022; Yan等, 2025)。特别是Wu等(2019, 2020)给出了用于地震数据特征分类识别的测试数据集, 极大方便了科研人员和学生在数据资源稀缺条件下开展机器学习的算法验证、模型对比与性能评估, 为后续相关方法的开发与推广提供了重要的基础条件。基于灰度图像的分类和回归任务在计算机图形学领域已

发展得相对成熟, 下面以3个例子来说明地球物理数据驱动的反问题。

3.1 地震数据断层识别的网络架构

用于高分辨率地震断层解释的网络架构包括: (1) 一个分割模块, 在地震图像中执行断层识别; (2) 一个特征融合模块, 用于将目标特征(断层)和背景信息(水平反射)融合为全局特征图; (3) 一个鉴别器模块, 用于识别全局特征图中的目标特征是来自分割模块的预测结果还是人工标注的断层标签。分割模块和鉴别器都是深度CNN, 而特征融合模块是一个特征增强处理流程, 用于解决常见的地质断层分割问题, 如边界模糊、断层连续性差和将倾斜地层误判为断层等。分割模块和鉴别器以对抗的方式进行训练, 在博弈中提高彼此的性能。比如图4给出了一个有效的断层特征识别神经网络(Wang和Wang, 2022)。

实现上述网络参数优化的核心数学模型如下:

(1) 平衡交叉熵损失函数

$$\ell_{\text{pixel}}(P, y) = -\beta_{|y|=1} \sum_{i \in M} \log(P_{i|y=1}) - \beta_{|y|=0} \sum_{i \in M} \log(1 - P_{i|y=0}) \quad (27)$$

(2) 分类损失函数

$$\ell_D(P, z) = -\sum_{i=1}^{2N} z_{1i} \log(P_{1i}) - \sum_{i=1}^{2N} z_{2i} \log(P_{2i}) \quad (28)$$

(3) 失配函数

$$\ell_{\text{image}}(\mathbf{C}) = \frac{-\sum_{i=1}^N 0 \cdot \log(C_{1i}) - \sum_{i=1}^N 1 \cdot \log(C_{2i})}{N} \quad (29)$$

(4) 正则化损失函数

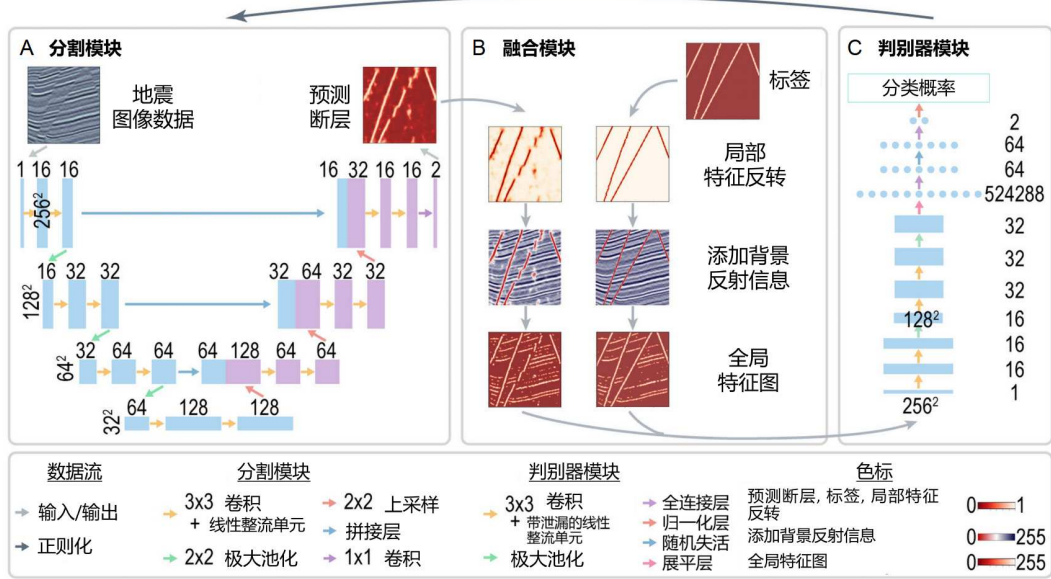


图4 断层识别的网络架构。(A) 分割模块(segmentation)；(B) 特征融合模块(feature fusion)；(C) 判别器模块(discriminator)。图中，(A)和(B)中的方块表示按尺寸和卷积核数量标注的特征图；(C)中的点表示全连接层的神经元，作为分类器使用；其中中文常对应的计算机术语有：Input or output-输入/输出，Regularization-正则化，Convolution-卷积，ReLU-线性修正单元，Maxpooling-极大池化，Upsampling-上采样，Concatenation-串联，LeakyReLU-带泄露的修正线性单元，Dense-全连接，Dropout-随机失活，Flatten-展平。改自Wang和Wang(2022)

$$\ell_S(P, y, C) = \ell_{\text{pixel}}(P, y) + \lambda \cdot \ell_{\text{image}}(C) \quad (30)$$

其中， M 是输入地震图像中的像素总数， y_i 是二进制标签（断层像素标签为1，非断层像素标签为0）， P_i 是分割模块输出的断层预测概率（ $0 \leq P_i \leq 1$ ）， $\beta_{y_i=1} = M_{y_i=1}/M$ 是断层像素数 $M_{y_i=1}$ 和像素总数 M 的比值， $\beta_{y_i=0} = M_{y_i=0}/M$ 是非断层像素数 $M_{y_i=0}$ 和像素总数 M 的比值。 $2N$ 为图像数量， z_{1i} 和 z_{2i} 是类别标签，即(1, 0)表示目标特征是来自分割模块的预测结果，而(0, 1)表示目标特征是来自人工标注的训练标签， P_{1i} 和 P_{2i} 分别是鉴别器对于上述输入给出的预测类别概率。 $C = (C_1, C_2)$ 是鉴别器的输出张量， C_{1i} 和 C_{2i} 分别是指第 i 个图像来自分割模块和断层标签的概率， ℓ_{image} 表示分割模块对鉴别器的欺骗能力， λ 是超参数（正则化参数）。损失函数 $\ell_S(P, y, C)$ 是像素级（局部）预测误差 $\ell_{\text{pixel}}(P, y)$ 和图像级（全局）预测误差 $\ell_{\text{image}}(C)$ 的加权和。具体地震断层理论数据和实际数据示例结果，见(Wang和Wang, 2022)。

3.2 完全数据驱动的波形反演

传统的全波形反演方法依赖于较好的初始模型并且需要大量的计算时间，Yang和Ma(2019)提出采用神经网络来实现原始的炮集地震记录到地下速度模型的非线性映射，该方法的依据是神经网络的万能逼近定理(Hornik, 1991)，网络的权重学习可以通过解下面的最优化问题得到：

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \frac{1}{mN} \sum_{n=1}^N L(v_n, \operatorname{Net}(u_n; \Theta)) \quad (31)$$

其中， Θ 为神经网络待学习的参数， $\operatorname{Net}(\cdot)$ 为基于全CNN的非线性映射， N 代表训练模型数量， v_n 代表第 n 个模型的真实地下速度值， m 为速度模型的总网格数， u_n 为观测的地震记录， $L(\cdot)$ 代表 v_n 和预测值之间的误差。

基于完全数据驱动的全波形反演方法忽略了传统方法中的正传和伴随算子，直接实现从炮集域(x, t)到成像域(x, z)的映射。由于输入和输出空间的不一致，预测结果的质量依赖于训练数据的多样性。这种方法的抗噪能力和泛化能力需要进一步提升。一种解决方案是将扩散生成过程与声波/弹性波方程相结合，通过学习观测地震数据约束下地下速度模型的概率分布来获得反问题概率意义下的解。

3.3 基于数据融合的反演/识别/预测

数据融合是指将来自多源数据的信息进行整合与优化，以形成更加全面、准确和一致的数据表达形式。作为一种重要的技术手段，数据融合在多个领域中得到了广泛应用，包括图像处理、医学诊断和自动驾驶等。在地球物理领域，单一观测手段往往难以全面揭示复杂的地下结构特征，而数据融合技术则提供了一种有效的解决方案。通过整合多源地球物理数据（如重力、磁力、地震波、地电等），不仅可以显著提高数据利用效率，还能

够生成高质量的数据集。这些数据集包含多种物性信息，能够更加全面地反映地下物性特征，为地球物理反演提供更丰富的信息源和更可靠的支撑(李红蕾等，2021；Bai等，2024)。

数据融合可以通过数据级、特征级、决策级融合，时序数据同化和神经网络融合实现(Meng等，2020)。基于控制理论和统计估计的数据同化可以将不同时间、空间分布且含有噪声的观测数据 $\mathbf{y}$ ，通过考虑观测误差和模型背景误差的协方差矩阵，利用贝叶斯估计或变分优化方法(如递推的卡尔曼滤波、3D/4D-Var)，求得系统状态的“最大后验概率估计(MAP)”或“最小均方差估计(MMSE)”： $\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}(\mathbf{x}_b))$ ($\mathbf{x}_a$ 为同化融合后的最优状态估计向量， $\mathbf{x}_b$ 为背景值， $\mathbf{H}$ 为观测算子， $\mathbf{K}$ 为权重由观测误差和模型背景误差的协方差矩阵计算得到)，该方法发展相对成熟(Kalnay, 2003；Evensen, 2009；Liang等，2013；Li等，2024)，具体实现方式不再赘述。这里以新一代神经网络融合技术为例。为实现观测数据 A 和 B 的特征融合处理，可设计如下深度学习网络：输入为尺寸为 $N_i \times N_i$ 的数据 A (data A)和 B (data B)，输出为同样尺寸的 $N_i \times N_i$ 融合数据。该网络通过两个独立的卷积分支对输入数据进行特征提取，并通过特征拼接与融合卷积层实现联合特征提取，最终生成单通道的融合结果。在经典的CNN中，卷积层后通常会接入池化层。然而，池化操作是不可学习的确定性过程，会丢失部分信息。为此，可设计网络为均采用卷积层构成，并以步进卷积(strided convolution)替代传统的“卷积+池化”组合。在网络中，数据尺寸的变化用 N_i 表示，其大小由卷积核尺寸、步长和填充3个超参数共同决定。网络中各层的通道数由该层卷积核的数量决定，反映了所学习特征的深度与复杂度。通道数的变化遵循“先扩张，后压缩”的策略，先逐步增加卷积核个数，经过拼接层(concatenate)后，再逐步减少卷积核个数，以实现两种不同来源数据的有效融合。图5

展示了观测数据 A 和 B 的步进CNN融合架构示意图。

更一般的情况下，在需要动态建模多源或多模态数据间复杂关系的场景中，可以采用Transformer网络架构中常用的自注意力机制(self-attention mechanism)，通过计算元素间的相似度(query-key匹配)，自动学习不同数据部分的权重分配，生成融合后的特征。以地球物理数据融合为例，基于Transformer网络架构，利用其自注意力机制对来自地震勘探、航空磁测、电磁探测、重力测量等不同传感器的异构数据进行时空特征对齐与关联建模。基本思路是通过独立编码器对各源数据进行特征提取与维度映射；在多头自注意力层中同时建立不同数据源间的全局依赖关系，实现跨模态信息交互与权重自适应分配；融合层将多源信息综合为统一表征，最终在解码器或任务特定模块中输出用于反演、分类、预测或异常探测的结果(如图6)。Transformer网络架构的核心技术概括为

(1) 多源特征对齐：将不同物理场(如波场、电磁场、位场的物性)映射到统一的高维嵌入空间，通过注意力权重(比如自适应加权策略)动态建模跨模态关联；

(2) 时空尺度兼容：利用位置编码(positional encoding)处理不同观测尺度(如地震与重力)的时空差异，这可以通过直接串联空间与时间编码、独立位置编码后相加或者联合嵌入处理空间和时间信息的相互作用的编码方式实现；

(3) 非线性耦合：通过多头注意力(multi-head attention)从多个子空间并行学习捕捉数据间复杂的非线性耦合关系(如岩石物性参数与地震反射特征的隐式关联)，突破了传统方法依赖固定物理公式或统计模型的局限性。

图6所示为多源地球物理数据融合的Transformer架构示意图，用于说明多源数据融合结果在目标区中如何用于反演、识别及预测。真实的场景通常需依据工区实际条件选择地球物理数据，以确保任务执行的有效性。

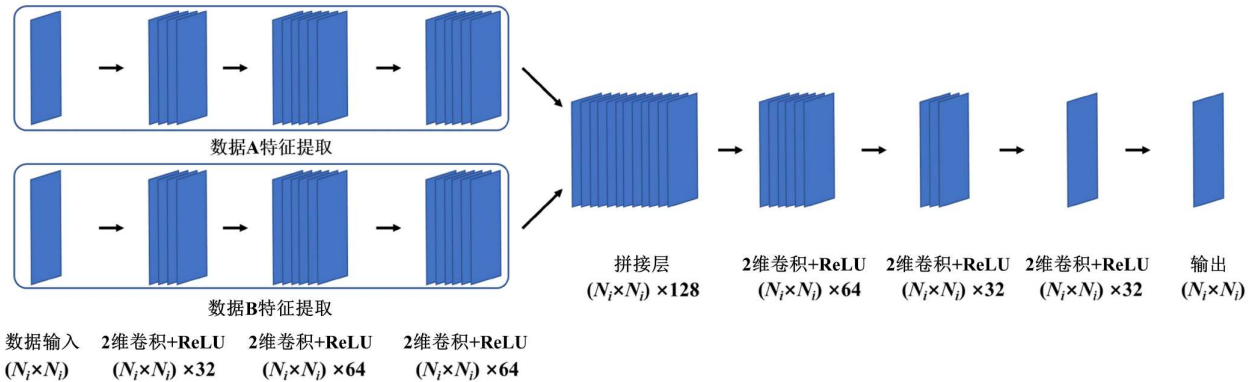


图5 观测数据A和B的神经网络融合架构示意图

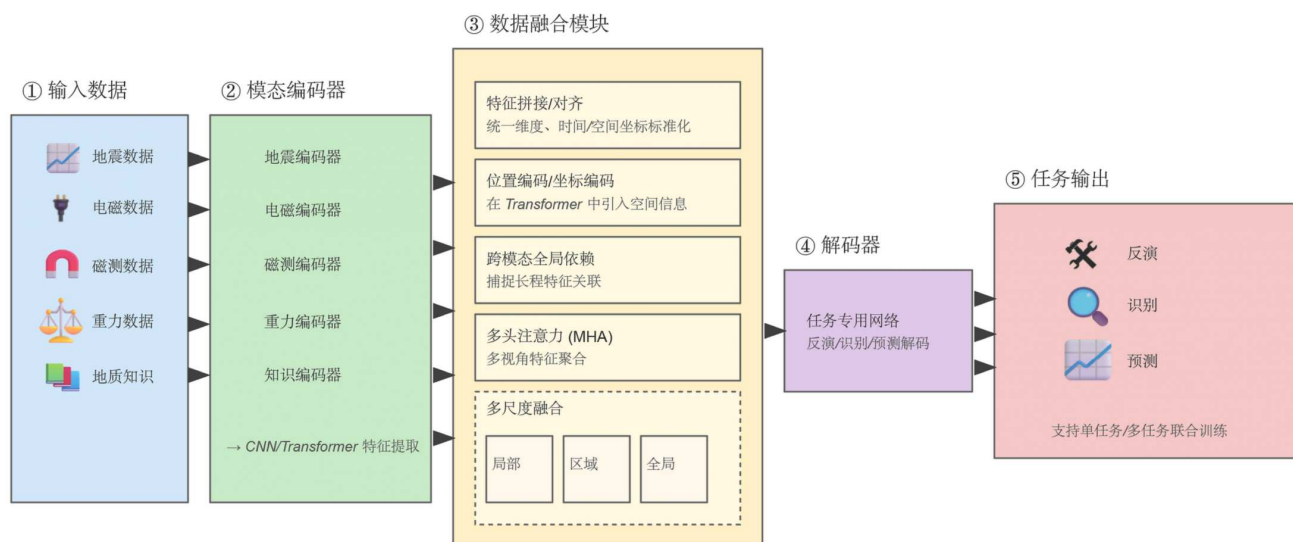


图6 多源地球物理数据融合Transformer架构示意图

3.4 进一步思考

传统地球物理反问题主要依赖于物理模型和数值计算，而数据驱动方法则通过人工智能算法挖掘大量观测数据中的潜在规律，进行更高效的反演和预测。随着多源、跨尺度数据融合的深入，以及MS人工智能计算方法的发展，有望解决数据噪声和不确定性带来的问题，提升反演的精度和效率，从而为资源探测提供更加全面和准确的地下信息。

4 一般反问题的计算方法

本节概要介绍一般反问题的计算方法，包括经典的Tikhonov正则化方法及其各种变种、最优化计算方法、统计反演策略，以及多源数据情形下的联合反演策略。本节没有介绍无导数优化和其他随机或启发式搜索方法，如粒子群优化、遗传算法、模拟退火、差分进化、灰狼优化及鱼群算法等(Kirkpatrick等, 1983; Dorigo和Stützle, 2004; Price等, 2005)。所有反问题的计算方法必须在数学严格证明了“收敛性、稳定性、正则性”的前提下才能放心应用于工程实践。

4.1 正则化方法

正则化方法在反问题求解中具有重要作用，其中Tikhonov变分正则化方法作为经典方法，最初在泛函空间中提出，通过构造一簇适定问题的解来逼近原问题的解，其关键在于正则化参数的选取，通常分为先验策略(基于算子性质、专家知识)和后验策略(基于误差水平，更具实用性)，如Morozov偏差准则、拟最优准则、L-Curve准则和广义交叉准则等(王彦飞, 2007; Tikhonov, 2009)。

另一重要方向是迭代正则化方法，研究表明迭代步数可类比正则参数，其终止准则对应参数选择方法，典型代表包括Landweber(1951)和Fridman(1956)提出的线性迭代格式，该方法因稳定性好、抗噪能力强而被广泛研究，后扩展至非线性不适定问题。Nashed和Votrubá(1976)基于代数理论提出广义逆方法(内逆法(inner inverse)和外逆法(outer inverse))，而Kirsch(1996)和Engl等(1996)分别研究了Tikhonov正则化在无限维空间中的收敛性及非线性问题的稳定性。King(1992)提出多层次Tikhonov算法，Scherzer等(1993)则探讨后验参数选择策略的最优性。Hettlich(1998)应用Landweber迭代求解逆散射问题，Tautenhahn和Jin(2003)比较了先验与后验策略的收敛性。2000年前后，后验选取正则参数的高阶优化算法也相继被不同的学者提出(肖庭延等, 2003)。对于求解非线性不适定问题的Tikhonov正则化方法，通过弱化假设条件，利用后验选取方法，也可以给出优化的正则化参数选取策略。此外，众多学者针对Tikhonov正则化方法、Landweber、Kaczmarz型正则化方法的正则解的误差估计，考虑了Hilbert空间和Banach空间上的分析结果(Bauer等, 2007; Hein和Kazimierski, 2010; Leitão和Alves, 2012)。

非线性正则化方法主要分为迭代Tikhonov、Landweber-Kaczmarz和梯度算子方法3类格式。Haltmeier等(2007)用Kaczmarz方法求解非线性反问题，并给出了收敛性分析。Pradeep和Rajan(2016)提出基于Lavrentiev序列的迭代方法。而梯度类方法(如共轭梯度法、最速下降法等)也被证明适用于带噪声问题，并可以写成Tikhonov正则化的形式(肖庭延等, 2003; 王彦飞, 2007)。在统计反问题中，Bauer等(2007)和Blanchard和Mücke(2018)分

别研究了广义正则化的收敛率。此外，地球物理领域发展了非光滑Tikhonov正则化、多种先验知识约束下的贝叶斯推理等新方法，具有与传统Tikhonov正则化的等价性(王彦飞等, 2011, 2021; Wang等, 2020)，形成从理论到应用的完整体系。

4.2 最优化方法

最优化方法主要用来求解非线性反问题。修正后的Gauss-Newton方法已经可用于求解非线性问题(Kaltenbacher, 1997)，Kangro等(2013)验证了其双参数变体的收敛速率。经典的Levenberg-Marquardt方法也可以用来求解反问题(Hanke, 1997; 肖庭延等, 2003; 王彦飞, 2007)。Wang和Yuan(2005)研究了非线性反问题的优化算法，特别是针对信赖域算法是不是正则化的疑问，指出标准的信赖域算法也是一种正则化方法，理论上证明了求解反问题的信赖域算法的稳定性、正则性和收敛性，为应用信赖域算法求解不适定问题奠定了理论基础。在结合优化算法方面，如正则化Gauss-Newton法、混合Newton-Tikhonov方法、Frozen-Landweber迭代法、非单调梯度下降算法(Barzilai-Borwein, BB)，非线性共轭梯度法、Newton-Landweber型迭代法、谱投影梯度法(spectral projected gradient methods)、Split-Bregman算法、交替方向法、交替方向乘子法(alternating direction method of multipliers, ADMM)等相继被提出研究(戴或虹和袁亚湘, 2000; 袁亚湘, 2008; Boyd等, 2011; 王彦飞等, 2011, 2021; 中国科学院, 2020)，这些方法较经典的Newton方法和最速下降法有明显的计算效率优势。特别是近10年，结合压缩感知思想(Candes等, 2006; Donoho, 2006; Candes和Wakin, 2008; Herrmann和Hennenfent, 2008)，稀疏优化方法在稀疏信号处理、压缩重构等领域也得到了新的应用，比如 L_1 范数约束条件下的信赖域算法在地震信号压缩感知中得到了应用，进一步推广的 L_q - L_p 范数条件下的地震信号压缩感知成像问题，以及稀疏矩阵优化算法(Figueiredo等, 2007; Goldstein和Osher, 2009; 王彦飞等, 2011; Wang和Wang, 2018; Xu和Wang, 2018)。这些稀疏优化问题大部分都可以写成复合优化的形式

$$\min(\Phi(u) + \Psi(v)), \quad \text{s.t.} \quad \mathbf{A}u + \mathbf{B}v = z \quad (32)$$

其中， $u \in \mathbb{R}^n$ 和 $v \in \mathbb{R}^m$ 是优化变量，等式约束中 $\mathbf{A} \in \mathbb{R}^{p \times n}$ ， $\mathbf{B} \in \mathbb{R}^{p \times m}$ ， $z \in \mathbb{R}^p$ ， Φ 和 Ψ 都是凸函数。结合对偶分解和乘子法，可以建立如下的无约束目标函数

$$L_p(u, v, y) = \Phi(u) + \Psi(v) + y^T(\mathbf{A}u + \mathbf{B}v - z) + \frac{\rho}{2} \|\mathbf{A}u + \mathbf{B}v - z\|_2^2 \quad (33)$$

其中， ρ 代表步长。在迭代次数 k 中，我们先对 u 进行最小化，然后对 v 进行最小化，最后更新 y ，在每次最小化期间保持其他变量不变，这就构成了ADMM(Boyd等, 2011)。由于ADMM的核心思想是将整体的优化问题拆分为子问题进行迭代求解，因而特别适用于大规模并行优化问题。

4.3 统计反演策略

用香农(Shannon)信息量刻画方式获得极大似然估计可能是不稳定的。贝叶斯(Bayes)理论通过先验分布和似然函数来表示模型参数的后验概率密度分布，结合已知的先验信息和观测数据获得模型参数的最佳拟合解，同时用概率分布描述可信度的反演结果可以让我们更好地认识反演多解性。假设模型参数 m ， d 为观测数据， G 为模型空间到数据空间的已知映射。基于贝叶斯理论的统计反演，是将所求的反演参数 m 看作是服从某一概率分布 $p(m)$ 的随机变量通过最大化后验概率密度函数 $p(m|d)$ 获得相应的解。其中， $p(m)$ 称为先验概率，比如在勘探地震学上，模型参数的先验概率通常都是通过测井或地质统计等方法获得的，这相当于在反演中引入了额外的约束信息，类似Tikhonov的思想，起到了正则化项的作用，能够降低反演的多解性。同时， $p(d|m)$ 称为似然函数，描述了在给定模型参数 m 时观测数据真值为 d 的条件概率密度，其代表了在给定模型参数 m 的情况下，通过正演计算出的模型值与真值间的近似程度。把模型参数的先验信息(先验分布)作为反演约束条件，并取不同光滑度的似然函数，则可以导出不同形式的贝叶斯反演格式，该反演格式具有与Tikhonov正则化方法的一致性(Wang等, 2008)。

统计推理同时也是量子理论的核心，这是因为量子系统的基本概率性质。贝叶斯方法在经典统计推理中得到了广泛的应用。近年来，量子贝叶斯规则(quantum Bayesian rule)在理论物理中出现了诸多讨论(Fuchs, 2010; von Baeyer, 2016; Caves等, 2002; Healey, 2023)。它结合了贝叶斯统计的概念与量子力学的原理，在量子状态的更新过程中进行合理的概率推断。当我们对一个量子系统进行测量时，测量结果会影响系统的状态。量子贝叶斯规则提供了一种框架，用于根据测量结果更新量子态的概率分布。在量子力学中，系统的状态由密度矩阵或量子态矢量描述，量子贝叶斯规则帮助我们理解如何根据测量结果调整这些描述。这种规则特别有用，因为量子测量会导致系统状态的坍缩，而量子贝叶斯规则可以帮助描述这一过程的概率性和信息更新。贝叶斯网络是一种用于表示和推理概率信息的图模型，

在AI中得到了广泛的应用。它在QC、量子信息处理，以及量子机器学习等领域具有潜在应用，帮助优化量子系统的控制和信息处理策略。希望这种图模型可以用于对量子世界中大型系统的行为进行推理。

4.4 联合反演策略

4.4.1 基于物性经验关系的联合反演

基于相同物性参数的联合反演是利用不同类型观测对同一组物性参数(如地下介质速度 V_p 、 V_s)的互补敏感性，通过联合建模与计算以减小非唯一性并提高模型分辨率(Han等, 2022)。通常的处理手段包括：共参数(common-parameter)联合反演，即直接在同一参数空间中同时拟合所有数据，建立不同物理场到参数的正演算子与统一的数据误差权重；也可以建立结构耦合(structural coupling)或交叉梯度(cross-gradient)方法，通过约束不同参数场在空间结构(界面或梯度)上相互匹配达到高分辨率反演效果，适用于不同物理量尺度与关系复杂的情形。

基于不同物性的地球物理观测数据之间的联合反演方面：在重、磁异常由同一场源引起的前提条件下，Guillen和Menichetti(1984)研究了使用广义反演法来实现重、磁联合反演，减小了重磁反演的多解性问题。Rasmussen和Hohmann(1991)用重力数据和瞬变电磁测深联合反演，取得了一定的效果。Chavez等(1987)在假设重、磁观测数据的响应为同一场源所引起的前提下，研究了密度和磁化强度的经验关系，构建了联合反演模型，进行了二维的重力-磁力联合反演实验。

基于不同物性参数之间的联合反演及其反演方法不同可以进一步分为：第一类是建立在岩石物性之间存在经验公式的相关性基础上的联合反演，第二类是不考虑岩石物性之间的相关性，利用同一地质体的不同物性参数具有相同的空间分布结构作为约束。基于岩石物性关系耦合的联合反演方法的前提和基础是：对于地下同一场源的地质体模型，它的不同岩石物性参数之间存在联系，它们之间的这种联系可以用经验公式显式地表示出来。比如，Moorkamp等(2011)利用电阻率、速度和密度之间的经验关系，实现了电磁法数据、地震数据和重力数据的二维联合反演。然而，基于岩石物性关系耦合的联合反演方法具有一定的局限性，在某些特定的地质环境中，不同物性之间存在一定的经验关系，但是在不同区域，它们对应的岩石物性关系可能是不同的，我们很难对所有的区域都使用统一的物性关系函数。而且并不是所有的物性参数之间都会相互联系，更不用考虑存在显式的经验关系，此时这种方法就不再可行。因此，这

种基于岩石物性关系耦合的联合反演方法具有一定的局限性。

数值实现上可采用确定性最小二乘/正则化框架或贝叶斯/采样方法进行不确定度量化。其中，初始模型选择、MS参数化、正则化项设计与数据加权在实际问题中需要精细考虑。比如，可以先单独反演获取初始模型，再进行联合反演。联合反演虽然会带来计算方面的开销，但可以增强深度分辨率、约束界面深度。

4.4.2 交叉梯度约束联合反演

交叉梯度函数联合反演本质上是一种基于结构耦合思想的联合反演方法，该方法能有效地改善反演结果，提高反演精度，在地球物理的很多领域都有应用(Gallardo和Meju, 2003; Gao和Zhang, 2018)。

三维的交叉梯度函数定义为

$$\mathbf{t}(x, y, z) = \nabla m_1(x, y, z) \times \nabla m_2(x, y, z) \quad (34)$$

其中， m_1 和 m_2 代表了不同的物性参数，比如密度、磁化率或慢度参数， ∇ 表示对该物性参数求梯度，整个交叉梯度函数是关于两两物性参数之间的梯度的叉乘，这是一个关于物性参数的非线性函数。

其在三维空间内可写为三分量的形式：

$$\mathbf{t}(x, y, z) = t_x(x, y, z)\mathbf{i} + t_y(x, y, z)\mathbf{j} + t_z(x, y, z)\mathbf{k} \quad (35)$$

其中， $\mathbf{i}$ 、 $\mathbf{j}$ 和 $\mathbf{k}$ 为三维空间轴向单位向量， $\mathbf{t}$ 在三个方向上可以分布进行展开：

$$\begin{aligned} t_x &= \frac{\partial m_1}{\partial y} \cdot \frac{\partial m_2}{\partial z} - \frac{\partial m_1}{\partial z} \cdot \frac{\partial m_2}{\partial y} \\ t_y &= \frac{\partial m_1}{\partial z} \cdot \frac{\partial m_2}{\partial x} - \frac{\partial m_1}{\partial x} \cdot \frac{\partial m_2}{\partial z} \\ t_z &= \frac{\partial m_1}{\partial x} \cdot \frac{\partial m_2}{\partial y} - \frac{\partial m_1}{\partial y} \cdot \frac{\partial m_2}{\partial x} \end{aligned} \quad (36)$$

交叉梯度函数中的梯度和叉乘运算，具有如下性质：

(1) 梯度是一个矢量，函数在某点处的梯度表示沿着该梯度方向的变化最快，变化率最大。

(2) 叉乘具体可由两个向量的模乘以 $\sin\theta$ 计算，其中， θ 代表两个向量在 (x, y, z) 点处之间的夹角。当 $\theta=0^\circ$ 或 180° 时， $\sin\theta=0$ ，此时两个向量的叉乘值为零。

这两种性质在交叉梯度中体现为：

(1) $\nabla m(x, y, z)$ 为模型参数在 (x, y, z) 处的梯度，梯度方向为模型参数在该点变化最快的方向，变化率最大。

(2) $\nabla m_1(x, y, z)$ 与 $\nabla m_2(x, y, z)$ 的叉乘为零时，表示在 (x, y, z) 点处，两种物性参数的梯度方向的夹角为 $\theta=0^\circ$ 或 180° ，那么表示两种物性参数的梯度变化方向一致或其中一种物性参数梯度方向没有变化。

在地球物理中，对于地下同一地质体，其边界处不

同物性参数的变化往往是一致的，也就是在边界处物性参数的梯度变化方向是一致的，它们之间的叉乘值为零，我们可以利用交叉梯度函数趋于零作为约束条件，应用到地球物理联合反演中。通过结构约束，最小化包含交叉梯度项的目标函数，实现针对同一地质体的不同物性参数结构一致性反演。

5 智能计算

5.1 人工智能反演方法

在实际勘探和解释过程中，地球物理数据量庞大、信息复杂且受噪声干扰，传统方法往往面临难以处理的问题。而人工智能技术(特别是深度学习技术(LeCun等, 2015))则通过自主学习和模式识别能力，具备了处理大规模、高维度数据，以及探测潜在关联性的能力，为地球物理学研究带来了新的机遇。人工智能地球物理(artificial intelligence geophysics)是将人工智能技术应用于地球物理学领域的研究方向，旨在利用人工智能技术解决地球物理数据处理和解释中的挑战(Geng和Wang, 2020a; He和Wang, 2021; Zhang和Gao, 2021; Wu等, 2023; Fu等, 2023; Liu等, 2023; Wang和Wang, 2023a, 2023b; Li等, 2025)。随着人工智能的发展和应用广泛性的不断扩大，它也逐渐在地球物理研究中展现出巨大的潜力。下面以地震反演和重磁反演为例，给出人工智能处理方法。

5.1.1 物理信息引导的神经网络反演

PINNs是一种融合深度学习和物理规律的创新方法(Raissi等, 2019)，在数据稀疏或噪声较大的情况下，仍能保持高精度的预测和建模能力。其核心思想是将控制物理系统的PDE直接嵌入神经网络的损失函数，使模型在训练过程中同时满足数据拟合和物理规律约束。相较于传统数值方法，PINNs可以实现无网格计算，特别适用于复杂集合结构的建模；PINNs适用于多物理场耦合和先验知识融合，以提高模型的可解释性；同时，在实验数据有限的情况下，物理方程的引入能有效减少过拟合风险，从而实现物理意义下的对抗学习。这种数值计算模式在地震勘探领域也得到了广泛的关注(Rasht-Behesht等, 2022; Song等, 2022; 邹婧博等, 2023; Li等, 2025)。

Li等(2025)提出了一种概率地质信息约束的叠前多道地震反演方法，通过将储层非均质性权重(RH)和地质模式约束权重(GPC)融入深度学习框架来克服传统反演中远离井位时解非唯一性增强的问题。基本思路是采用

2D残差UNet网络，结合克里金插值生成的参考模型与两种概率权重，构建能平衡井数据与地震数据约束的三维概率标签系统。方法在保持空间连续性的同时，提高非均质储层中弹性参数变化的表征精度。文中通过在3D合成模型和实际致密砂岩储层数据测试，验证了该方法的抗噪能力和地质体边界识别能力。该方法同样适用于碳酸盐岩等横向变化显著的复杂储层，为整合地震、测井与地质信息提供了新的思路。

针对一般形式的PDE: $\mathcal{L}(u) = 0$, $u \in \Omega$, $\mathcal{L}$ 是微分算子, u 是未知函数, Ω 是计算区域, PINNs的目的是找到满足上述微分方程及其边界条件的解。PINNs基本思路是构建一个神经网络 $u_\theta(x)$, 其输入是位置或时间变量 x , 网络参数为 θ , 输出为解 $u(x)$ 。核心是极小化如下的目标函数:

$$L_{\text{total}} = L_{\text{PDE}} + L_{\text{BC}} + L_{\text{IC}} \quad (37)$$

其中, L_{PDE} 、 L_{BC} 、 L_{IC} 分别表示微分方程约束、边界条件约束、初始条件约束。

(1) 微分方程约束，即通过神经网络输出 $u_\theta(x)$ 对微分方程 $\mathcal{L}(u) = 0$ 进行建模，并对其进行最小化:

$$L_{\text{PDE}} = (1 / N_{\text{PDE}}) \sum_{i=1}^{N_{\text{PDE}}} \left| \mathcal{L}(u_\theta(x_i)) \right|^2。$$

(2) 边界条件约束，即根据问题的边界条件(如Dirichlet边界条件或Neumann边界条件)，最小化边界上的

$$\text{误差: } L_{\text{BC}} = (1 / N_{\text{BC}}) \sum_{i=1}^{N_{\text{BC}}} \left| u_\theta(x_{\text{BC},i}) - u_{\text{exact}}(x_{\text{BC},i}) \right|^2。$$

(3) 初始条件约束，即对于时间依赖问题，使用初始条件约束神经网络输出与初始状态的误差:

$$L_{\text{IC}} = (1 / N_{\text{IC}}) \sum_{i=1}^{N_{\text{IC}}} \left| u_\theta(x_{\text{IC},i}) - u_{\text{initial}}(x_{\text{IC},i}) \right|^2。$$

为了稳定地求解PDEs，建立形如目标函数式(37)的形式是不够的，一般还需要额外引入正则化约束: $L = L_{\text{total}} + L_{\text{regu}}$, L_{regu} 可以为超参数或超参数微分的模量，类似声波方程的正则化那样(Wang等, 2014; Su等, 2025)。

Rasht-Behesht等(2022)开发了适用于声波方程的PINN算法，并通过正演模型和FWI(full waveform inversion)案例验证其有效性。针对远震平面波和地震点源两种场景设计的合成案例表明，该方法能适应不同结构复杂度需求。PINNs的无网格特性使其能灵活实现波动方程和各类边界条件，例如自动满足传统波传播求解器难以处理的吸收边界条件，同时支持将地下结构先验知识直接编码到数学表达式中。PINNs通过偏微分控制定律在数

据有限条件下成功求解波传播和FWI问题，其三大核心优势体现在：(1) 多样化约束的无缝整合能力，(2) 无网格特性及简易实现，(3) 为地球物理反演提供了新一代求解框架。这种灵活架构可同时融合多类型数据与成像结构的先验知识，为联合反演和贝叶斯反演开辟了新途径，不仅适用于地震全波形反演，还可扩展至MT、重力等其他地球物理数据领域。

单连璞等(2025)以频域声波传播问题为例，研究了采用PINNs格式求解带有PML(perfectly matched layers)的Helmholtz方程，其控制方程的一般形式为

$$(\nabla^2 + \omega^2 m(\mathbf{x}))u(\mathbf{x}) = s(\mathbf{x}) \quad (38)$$

其中， u 为复值位移场，空间坐标 $\mathbf{x}=(x, y)$ 分别对应水平与深度方向， ∇^2 为拉普拉斯算子， $s(\mathbf{x})$ 为震源项， $m(\mathbf{x})=v^{-2}(\mathbf{x})$ 为介质平方慢度， ω 为角频率。为抑制有限计算域边界反射，采用了PML边界条件。未引入基于监督学习的数据损失项，只考虑了包含方程残差项的物理损失项，定义损失函数为

$$\mathcal{L}(\mathbf{x}; \theta) = \frac{1}{N} \sum_{i=1}^N \left| \frac{\partial}{\partial x} \left(A_i \frac{\partial u(\mathbf{x})}{\partial x} \right) + \frac{\partial}{\partial y} \left(B_i \frac{\partial u(\mathbf{x})}{\partial y} \right) + C_i \omega^2 m(\mathbf{x}) u(\mathbf{x}) - s(\mathbf{x}) \right|^2 \quad (39)$$

其中， A_i, B_i, C_i 为复数拉伸因子的参数组合； θ 为网络所有权重矩阵与偏置向量构成可训练参数集；训练过程通过最小化损失函数 $\mathcal{L}(\mathbf{x}; \theta)$ 确定最优参数 θ^* ：

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(\mathbf{x}; \theta) \quad (40)$$

图7给出了一般的PINNs求解Helmholtz方程框架示意图。

5.1.2 重参数化地震全波形反演

全波形反演就是求解一个PDE约束的反问题，在地震速度反演和医学成像中得到了广泛应用(Wu和McMechan, 2019; Guasch等, 2020; Pan和Wang, 2020; Pan等, 2020; He和Wang, 2021; 杨顶辉等, 2025)，目标函数定义为

$$\begin{cases} f(m) = \frac{1}{2} \| Ru(m) - d_{\text{obs}} \|_2^2 \\ F(u, m) = s \end{cases} \quad (41)$$

其中， $\|\cdot\|_2$ 表示 L_2 范数， R 是一个投影运算符，其作用是仅在每个震源的接收器位置选择波场， m 代表参数化地下的物理参数， s 为震源函数， F 为非线性算子(如声波或弹性波算子)， d_{obs} 是观测数据， $u(m)$ 代表模拟数据。杨顶辉等(2025)系统描述了FWI成像理论和方法，指出该方

法不仅在揭示地球内部构造、动力学演化和地震发生机理等方面发挥重要的作用，而且在地球资源探测、医学成像、工程检测、二氧化碳地质封存、地震灾害预测等领域显示出广泛的应用潜力。对于上述等式约束的目标函数极小化问题，可以使用梯度迭代法或拟牛顿方法求解，在第 $k+1$ 次迭代中，针对这些优化方法的参数的更新公式为

$$m_{k+1} = m_k + \alpha_k p_k \quad (42)$$

其中， p_k 表示该步的下降方向， α_k 为步长。例如，选择 $p_k = -\nabla f(m_k)$ 便得到最速下降法，选择 $p_k = -B_k^{-1} \nabla f(m_k)$ 便产生牛顿类型的方法，式中 B_k 是黑塞(Hessian)算子 H_k 的对称正定近似。

上述全波形反演问题还可以利用神经网络参数化表达并通过深度学习实现(He和Wang, 2021; Wang W等, 2021; Zhu等, 2022; Sun J等, 2023; Jin等, 2024)。这种深度学习反演方法的主要思想是：使用深度神经网络的权重对物理参数进行重新参数化，将原反演问题转化为对网络参数的学习问题。该反演方法的一个优势是：网络的稀疏表示性质使得该方法具有正则化的作用，因此其比较适合求解不适定的非线性反问题。此外，由于该方法仅需要一阶导数信息，因此具有很高的计算效率。此外，通过使用多个图形处理单元(GPU)和中央处理单元(CPU)进行权重更新和波场正演模拟，使其能够充分利用硬件对算法进行加速。

根据万能逼近定理，只要给予深度神经网络隐藏层足够的神经元，神经网络能以任意的精度逼近任意博雷尔(Borel)可测函数。对于FWI问题的反演参数 m ，均可以找到一个深度神经网络 $G(w)$ ，其可以给定的精度近似逼近反演参数化 m 。换句话说，模型参数 m 可以通过利用神经网络参数 w 进行表示(或重参数化)，简记为

$$m(x) = G(w)(x) := G_L(w_{L-1}, \dots, G_1(w_1))(x) \quad (43)$$

进一步简洁地表示为 $m(w)=G(w)$ ，以明确其依赖于权重 w 。通过 m 的重新参数化，式(41)中的目标函数变为

$$J(w) = f(m(w)) = \frac{1}{2} \| Ru(G(w)) - d_{\text{obs}} \|_2^2 \quad (44)$$

在这种表述中，FWI的重构问题转化为重建网络 $G(w)$ 的权重 w 。这种重新参数化策略的一个优点是，它通过引入特殊的层(如卷积层)来提供模型的稀疏表示，以正则化的方式减轻FWI局部极值问题。此外，根据物理参数的先验信息，可以通过构建网络的特殊层来提取一些特定特征。换句话说，在DNN(deep neural networks)框架下的反演可以看作是一种隐式正则化方法，这对于解决FWI问题的不适性至关重要。此外，可以通过成熟高

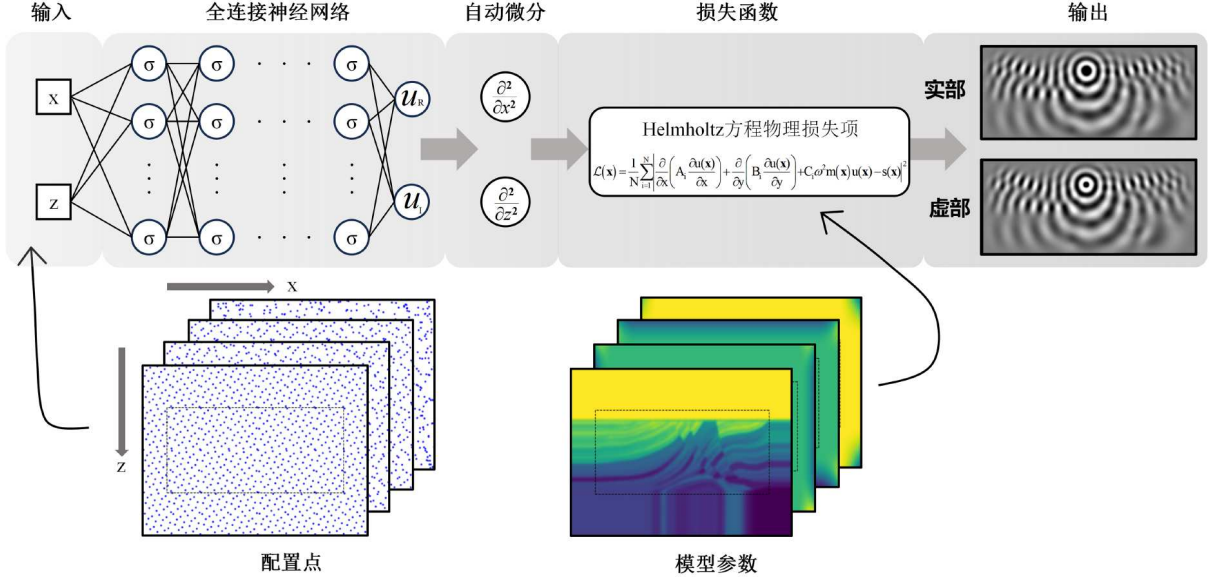


图7 PINNs求解亥姆霍兹方程框架示意图, 改自单连璞等(2025)

效的神经网络库, 例如TensorFlow、PyTorch或PaddlePaddle, 使用GPU来加速DNN的训练过程(He和Wang, 2021)。

为了使用梯度型的方法极小化目标函数式(44), 应该有效地计算目标函数关于学习参数 w 的梯度。对方程(44)进行微分并应用链式规则, 我们有

$$\begin{aligned} \nabla J(w) &= \left(\frac{\partial m(w)}{\partial w} \right)^T \frac{\partial f(m(w))}{\partial m} \\ &= \left(\frac{\partial G(w)}{\partial w} \right)^T \nabla f(m(w)) \end{aligned} \quad (45)$$

其中, 第一项 $\partial m(w)/\partial w$ 是网络的雅可比矩阵, 可以使用反向传播算法有效地进行估算。因子 $\nabla f(m(w))$ 是常规FWI相对于 m 的梯度。其也可以通过伴随方法有效地进行计算。

在进行DNN-FWI训练之前, 必须参数化给定的初始模型 m_0 ; 选取合适的网络架构, 比如PyTorch, 通过输入数组(张量) $\mathbf{a}$, 以使网络学习该先验信息。我们称这种参数化处理为“预训练”。为了参数化或学习初始模型, 我们需要最小化以下目标函数, 该函数度量了 $G(w)$ 和 m_0 之间的距离:

$$w^* = \underset{w}{\operatorname{argmin}} J_{pre}(w) = \| G(w) - m_0 \|_1 \quad (46)$$

其中, $\|\cdot\|_1$ 表 l_1 范数。 l_1 范数可以捕获初始模型的主要特征, 与使用 l_2 范数相比, 捕捉异常点值的能力更强大。

在预训练后, 开始DNN-FWI的训练过程。常规训练和重新参数化FWI训练之间的主要区别是相对于权重的 m 梯度计算。对于常规训练, 仅将网络输出与给定标记数据之间的差异从输出层向后传播到输入层, 以计算梯度。

对于DNN-FWI, 向后传播梯度 $\nabla f(m(w))$ 以计算相对于权重 w 的梯度。因此, DNN-FWI通过合成记录的数据以确保数据的一致性, 可以利用PDE约束的正演模型。此外, 可以分别在CPU和GPU上对 $\nabla f(m(w))$ 和网络 $G(w)$ 的偏导数进行加速。

与传统的迭代正则化算法的收敛性、稳定性证明方式类似, 重参数化的DNN-FWI算法亦可以证明其收敛性, 具体可参考非线性优化问题的正则化证明思路(Wang和Yuan, 2005)。

5.1.3 基于偏移成像的深度学习

针对深度学习缺少大量训练速度模型和全波形反演无法得到准确的深层速度模型的问题, 基于卷积神经网络的反射波反演(CNN-RWI)是一个不错的选择, 因为该方法能够有效利用初始速度模型的背景速度信息(低波数成分)和相应的偏移图像的反射界面信息(高波数成分), 准确预测出高清晰度的地震速度模型(Wu Y L等, 2022)。

在CNN-RWI算法中, CNN不断地通过初始速度模型和偏移图像预测速度模型。初始速度模型 $V_0(\mathbf{x})$ 可以通过层析成像或者速度分析来获得。偏移图像 $I_0(\mathbf{x})$ 可以通过RTM(Baysal等, 1983; McMechan, 1983; Whitmore, 1983)计算得出:

$$I_0(\mathbf{x}) = \int_0^T U_s(\mathbf{x}, t) U_r(\mathbf{x}, t) dt \quad (47)$$

其中, $U_s(\mathbf{x}, t)$ 和 $U_r(\mathbf{x}, t)$ 分别是在时刻 t , 位置 $\mathbf{x}$ 的震源波场和检波点波场, T 是总时间。

当初速度模型 $V_0(\mathbf{x})$ 和相应的偏移图像 $I_0(\mathbf{x})$ 计算出

来后,二者将被CNN-RWI反复输入到CNN中去预测未知的真实速度模型。实现上通过内外二层循环得到。外循环的主要目的是在不改变CNN预测时输入数据(即初始速度模型 $V_0(\mathbf{x})$ 相应的偏移图像 $I_0(\mathbf{x})$)的条件下,通过自适应反馈改善CNN训练集从而改善CNN的拟合能力且降低过拟合问题,继而实现CNN更加准确的模型预测。外循环包含4部分:准备训练集里的真实速度模型、CNN训练、CNN预测和误差评估。内循环是指CNN在训练阶段不断地更新CNN的参数去拟合训练标签。

CNN-RWI算法主要包括数据预处理(获得初始速度模型和偏移图像)、聚类分割(生成训练速度模型)、训练速度模型的数据预处理(获得初始速度模型和偏移图像)、CNN的训练和预测,以及自适应反馈机制。下面给出实现算法,具体案例见Wu Y L等(2022)。

(1) 获取初始速度模型和偏移图像。常规的地震勘探预处理流程,比如层析成像和速度分析可以基于观测地震记录获得初始速度模型。偏移成像技术(如RTM)可以基于初始速度模型和真实地震记录获得模型空间下的偏移图像。考虑到CNN具有空间平移不变性,CNN-RWI试图基于地震勘探预处理所得到的模型域的先验信息(即初始速度模型和偏移图像),来预测未知的真实速度模型。

(2) 生成训练速度模型。采用基于 k 均值(k -means)的自上而下的层次聚类算法。该算法能够从先验速度模型(如初始速度模型和CNN预测的速度模型)构建速度模型的特征集合,并基于该特征集合的统计分布等信息随机生成速度模型样本库。基于先验信息生成的随机速度模型样本库比不依赖于先验信息的随机速度模型更好地代表了未知的真实速度模型,从而有效降低了CNN训练过拟合的问题。

(3) 训练速度模型的数据预处理。在生成了随机速度模型作为训练标签后,相应训练集的输入(初始速度模型和偏移图像)可以通过训练速度模型的数据预处理获得。初始速度模型直接通过对随机训练速度模型进行平滑处

理得到。平滑处理极大地节省了地震勘探预处理(如层析成像和速度分析)的计算成本,但也会让训练集中初始速度模型的特征与基于真实地震记录预处理所得到的初始模型特征有所不同,从而影响训练集对真实速度模型,以及相应的先验信息表征的准确度。不过,当CNN-RWI所采用的初始速度模型较为精准时,训练集中经过平滑处理后得到的初始速度模型与先验的初始速度模型的特征差异可以忽略不计。与此同时,训练集中偏移图像的获取采用与真实地震数据预处理一致的偏移成像算法(如RTM)。

(4) CNN的训练和预测。CNN的训练与预测采用传统的监督学习下的CNN的训练和预测,主要区别在于训练集是基于先验信息生成的。在训练阶段,通过输入训练库中的初始速度模型和偏移图像到CNN,再迭代式地不断更新CNN参数来拟合相应的训练速度模型。在预测阶段,将基于真实地震记录得到的初始速度模型和偏移图像输入到CNN,获得CNN预测的速度模型。预测的速度模型将会被自适应地反馈到模型切割步骤中用于生成新的训练库。

(5) 自适应反馈机制。通过不断地聚类分割CNN预测的速度模型获得更准确、更有代表性的训练集,降低CNN的过拟合,提高CNN预测的速度模型的准确度。同时,自适应反馈机制避免了传统FWI的地震记录拟合过程。

5.1.4 基于ResNet深度模型的分数阶黏声波动方程反演

在解耦的黏声波动方程式(19)中,两个分开的分数阶拉普拉斯项分别表示相位频散和振幅衰减,其中 $\eta \cdot (-\nabla^2)^{\gamma+1} \cdot \sigma$ 为频散主导项, $\tau \cdot \partial/\partial t ((-\nabla^2)^{\gamma+1/2} \cdot \sigma)$ 为振幅衰减主导项。因此,深度学习采用的高质量的数据 $\mathbf{I}(\mathbf{v}, Q)$ 可以由 Q 补偿RTM实现(Zhang和Gao, 2021)。

采用深度学习的思想,构建神经网络 A_θ^k ,基于解耦的分数阶黏声波动方程,分别对地质模型的速度 $\mathbf{v}$ 和品质因子 Q 进行反演,其迭代过程可以表示为

$$\begin{cases} \mathbf{v}_{k_1+1} = A_\theta^{k_1}[\mathbf{v}_{k_1}, I(\mathbf{v}_{\text{back}}, Q_{\text{back}}), I_{\text{stack}}(\mathbf{v}_{\text{back}}, Q_{\text{back}}), \nabla R(\mathbf{v}_{k_1})], \\ Q_{k_2+1} = A_\theta^{k_2}[Q_{k_2}, I(\mathbf{v}_{\text{back}}, Q_{\text{back}}), I_{\text{stack}}(\mathbf{v}_{\text{back}}, Q_{\text{back}}), \nabla R(Q_{k_2})], \end{cases} \quad \begin{matrix} (k_1 = 0, 1, 2, \dots, N_1 - 1) \\ (k_2 = 0, 1, 2, \dots, N_2 - 1) \end{matrix} \quad (48)$$

其中, N_1 和 N_2 分别为速度 $\mathbf{v}$ 和品质因子 Q 迭代学习的次数,并且 $\mathbf{v}_0 = \mathbf{v}_{\text{back}}$, $Q_0 = Q_{\text{back}}$,正则化项取为Dirichlet能量函数,即:

$$R(m) = \frac{1}{2} \|\nabla m\|_2^2 \quad (49)$$

对式(49)求梯度可以得到:

$$\nabla R(m) = \nabla \cdot \nabla m \quad (50)$$

其中, $\nabla \cdot$ 为散度算子。公式(50)即为当前模型的拉普拉斯表达,有助于提取边界信息。

神经网络架构 A_θ^k 有多种模式,我们提出采用多尺度卷积残差网络结构(ResNet^{MS}),如下图8所示。

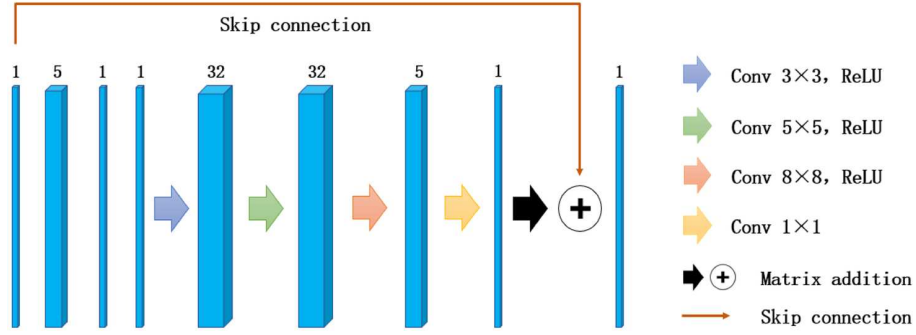


图8 多尺度卷积残差网络结构(ResNet^{MS}): Skip connection-跳跃连接, Conv-卷积, Matrix addition-矩阵加法

下面给出对ResNet^{MS}网络的说明:立方体上方的数字代表特征通道数,输入模型的第一个通道为速度 v 或者品质因子 Q ,接下来5通道的数据体为叠前RTM成像数据,其余两个通道的数据体分别为叠后RTM成像数据和正则化项的梯度,之后对输入数据依次进行 3×3 , 5×5 和 8×8 的卷积,提取数据不同尺度的信息,每个卷积后跟随一个ReLU激活函数,再通过 1×1 卷积将其映射为残差值,并与输入速度或者品质因子相加得到本轮迭代更新的结果,之后继续迭代直至达到相应的预设迭代次数。

对初始ResNet^{MS}进行迭代训练时的损失函数表示为 $L = \frac{1}{2} \left\| A_{\theta}^k[\mathbf{m}_k, \mathbf{I}'_1, \mathbf{I}'_2, \nabla R(\mathbf{m}_k)] - \mathbf{m}_{\text{true}} \right\|_2^2$,即均方根误差。其中, A_{θ}^k 表示第 k 次迭代时的ResNet^{MS}, θ 表示ResNet^{MS}的网络参数; $\mathbf{m}_k$ 表示第 k 次迭代的目标参数,目标参数表示:速度或者品质因子, $\mathbf{I}'_1$ 表示训练叠前RTM成像数据, $\mathbf{I}'_2$ 表示训练叠后RTM成像数据, $\nabla R(\mathbf{m}_k)$ 表示目标参数的正则化项 $R(\mathbf{m}_k)$ 的梯度, $\mathbf{m}_{\text{true}}$ 表示样本地层模型中的目标参数真实值。

为了对网络进行训练,我们基于White和有效介质模型随机生成了229个含水合物模型,模型的大小为 411×411 ,空间采样间隔为5 m。将生成的229个含水合物模型中的195个模型作为训练样本,其余34个模型作为测试样本。深度学习算法要求训练集和未来的测试集有相似的特征空间和分布,但是在实际的应用中包含全部特征样本的训练集是难以实现的。因此,在实际应用中,还得利用一部分目标数据进行迁移学习。下面给出针对非常规油气(天然气水合物)的深度学习反演结果分析。复杂地质构造背景是中国南海神狐海域水合物地质模型(杨胜雄等, 2017)。图9和10分别为预测的速度和品质因子模型,可以看到相比于光滑背景模型,水合物的位置基本得到了恢复,但是水合物轮廓的精细程度和数值准确性有待提高,同时深部的一些层状结构未能得到预测。

5.1.5 基于Transformer架构地震盲反演

Transformer架构自2017年提出后迅速成为深度学习领域的重要模型之一。在自然语言处理(NLP)中,Transformer及其变体推动了大语言模型(LLM)的发展(如OpenAI的ChatGPT、Google的Gemini、DeepSeek、Qwen等);同时,基于Transformer的方法也被成功应用到图像、语音和多模态学习等任务(如Vision Transformer等)。Transformer的核心包括多头自注意力机制、位置编码、逐位置前馈网络,以及残差连接与层归一化等模块。由于其并行化能力和强大的表示学习能力,Transformer及其在强化学习、蒸馏、微调等训练技术上的结合(特别是DeepSeek的巨大成功),成为近年来的研究热点和工程实践重点。以波阻抗反演为例来说明该网络的强大推理能力。

由于地震子波在地层传播过程中,幅值、相位、频滤都会发生变化,即地震子波是一种时变子波,取 $\alpha(\tau, \omega) = \exp\left(-\pi \cdot \omega \cdot \int_0^{\tau} (1/Q(t)) dt\right)$,其中, ω 表示频率, $Q(t)$ 为 t 时刻的地层衰减(或滤波)因子。假设子波在横向不变,多道非平稳地震记录可以表示成如下形式(Zou等, 2022):

$$\mathbf{S} = \mathbf{W} \mathbf{\Lambda}_Q \mathbf{D}_t \mathbf{Z} + \mathbf{N} \quad (51)$$

其中, $\mathbf{W}$ 是地震子波的离散矩阵形式, $\mathbf{S}=[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_M]$ 是非平稳地震记录, $\mathbf{Z}=[\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M]$ 代表对数形式离化后的反射系数矩阵, $\mathbf{N}=[\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_M]$ 代表噪声矩阵, $\mathbf{\Lambda}_Q$ 表示衰减函数 α 的离散化形式, $\mathbf{D}_t$ 表示关于时间的差分矩阵, M 代表地震数据的道数。

传统的求解方法是在 $\mathbf{S}$ 和 $\mathbf{D}_t$ 已知的情况下,由式(51)反演 $\mathbf{Z}$ 和 $\mathbf{W}$ 。地震盲反演问题公式表述如下:

$$\underset{\mathbf{Z}, \mathbf{W}}{\operatorname{argmin}} J(\mathbf{Z}, \mathbf{W}) \quad \text{subject to} \quad \left\| \mathbf{W} \mathbf{\Lambda}_Q \mathbf{D}_t \mathbf{Z} - \mathbf{S} \right\|_{\text{F}}^2 \leq \varepsilon \quad (52)$$

其中, $J(\mathbf{Z}, \mathbf{W})=J_1(\mathbf{Z})+J_2(\mathbf{W})$, J_1 和 J_2 分别表示反射系数矩阵 $\mathbf{Z}$ 和子波矩阵 $\mathbf{W}$ 的目标函数,定义为 $J_1(\mathbf{Z})=\lambda|\mathbf{Z}|_{\text{WATV}}+$

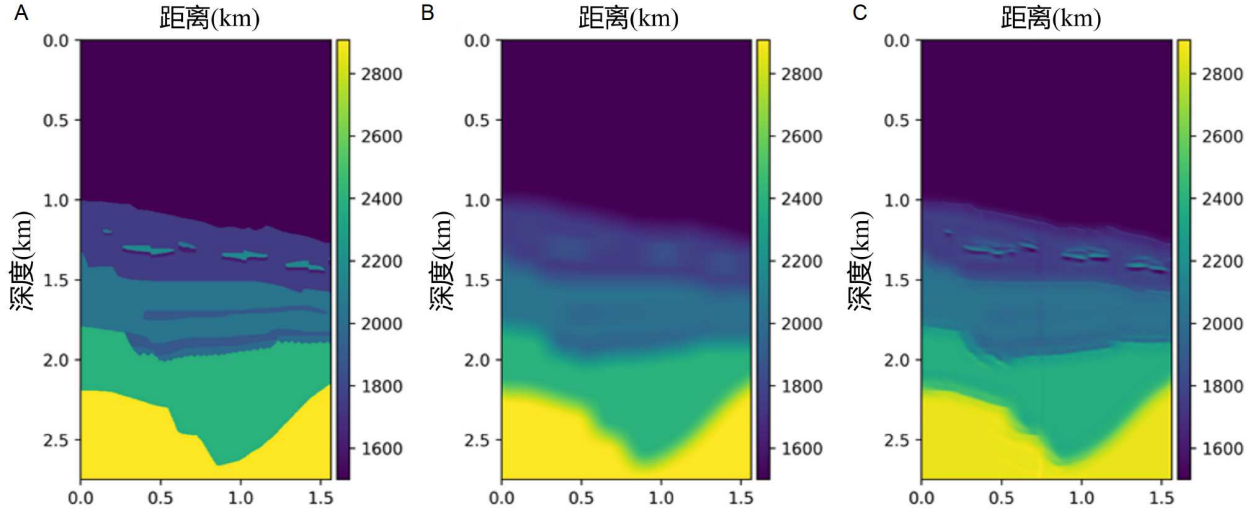


图9 复杂地质构造背景下含水合物模型的速度预测结果。(A) 真实速度模型；(B) 光滑的背景速度模型；(C) 预测速度模型，速度单位m/s。改自宁亚鑫(2022)、王彦飞和宁亚鑫(2025)

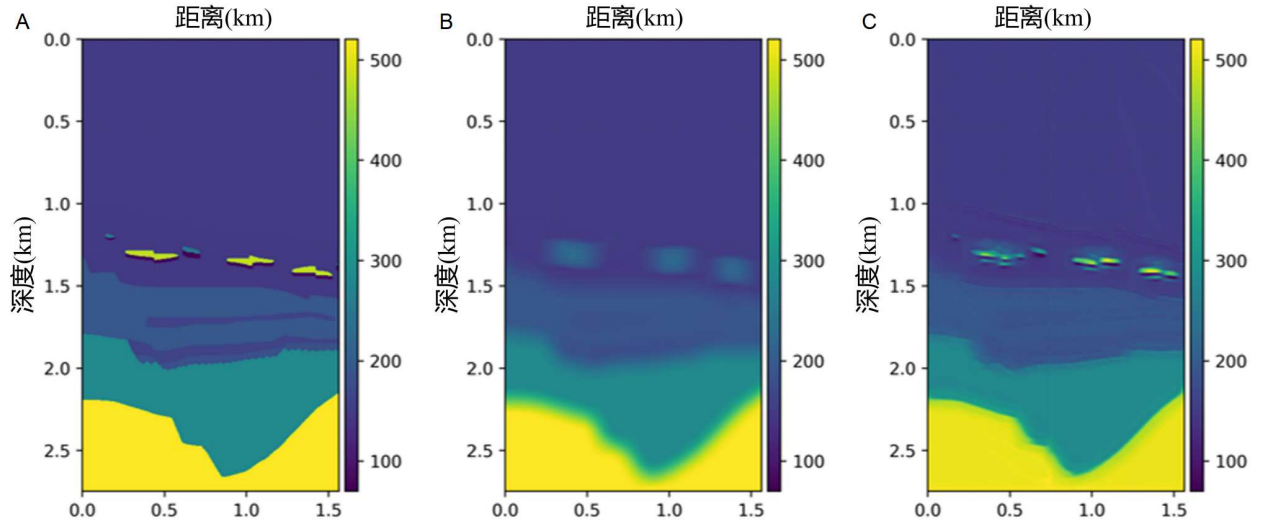


图10 复杂地质构造背景下含水合物模型的品质因子预测结果。(A) 真实品质因子模型；(B) 光滑的背景品质因子模型；(C) 预测品质因子模型。改自宁亚鑫(2022)、王彦飞和宁亚鑫(2025)

$$\frac{\beta}{2} \|\hat{\mathbf{Z}}_{\text{low}} - \mathcal{L} \circ \hat{\mathbf{Z}}\|_2^2, \quad J_2(\mathbf{W}) = \sigma\zeta(\mathbf{W}) + \gamma|\mathbf{W}|_{\text{ATV}} + \frac{\mu}{2} \|\mathbf{W}\|_2^2;$$

$\hat{\mathbf{Z}}_{\text{low}}$ 和 $\hat{\mathbf{Z}}$ 分别为 $\mathbf{Z}_{\text{low}}$ 和 $\mathbf{Z}$ 的二维傅里叶变换, $\mathcal{L}$ 为傅里叶域高斯低通滤波算子, λ 和 $\beta > 0$ 为正则化参数, 符号“ $\circ$ ”为 Hadamard 乘积; $|\cdot|_{\text{WATV}}$ 表示加权各向异性总差分正则化项, 定义为 $|\mathbf{Z}|_{\text{WATV}} = \sum_{i,j} (|\nabla_t^w \mathbf{Z}|_{i,j} + |\nabla_x^w \mathbf{Z}|_{i,j})$, $\nabla_t^w \mathbf{Z} = \mathbf{W}_t \circ (\mathbf{D}_t \mathbf{Z})$, $\nabla_x^w \mathbf{Z} = \mathbf{W}_x \circ (\mathbf{D}_x \mathbf{Z})$; $\mathbf{D}_t$ 和 $\mathbf{D}_x$ 分别为时间和空间方向上的一阶差分, $\mathbf{W}_t$ 和 $\mathbf{W}_x$ 是两个加权矩阵, 其元素为

$$w_{i,j}^t = 1 / \left(\phi \log_2 \left(1 + \exp \left((\mathbf{D}_t \mathbf{Z})_{i,j} / \phi \right) \right) \right),$$

$$w_{i,j}^x = 1 / \left(\phi \log_2 \left(1 + \exp \left((\mathbf{D}_x \mathbf{Z})_{i,j} / \phi \right) \right) \right),$$

ϕ 是一个小于 1 的正常数, 可以人为设定。

最小化问题(式(52))的求解, 可通过经典的 ADMM、快速梯度类方法或子空间优化方法实现(袁亚湘, 2008; Wang 等, 2011; 刘浩洋等, 2020; 王彦飞等, 2021)。

从式(51)可以看出, 该盲反演过程处理的是带有地质特征的时间序列, 因而可以采用 LLM 中的 Transformer 架构, 并施加地球物理约束求解。Transformer 由自注意力机制和前馈神经网络组成, 其网络结构也成为了近年来的研究热点。邹安祺(2022)研究了深度学习在地震反

射系数盲反演中的应用, 提出采用Transformer网络对地震反射系数和地震子波进行深度学习, 实验结果表明, 采用Transformer网络可以学习到地震反射系数和地震子波的特征, 并比传统算法具有更高的精度和更好的抗噪性能。Fu等(2023)针对地震阻抗反演中子波未知和观测数据频带受限导致的高分辨率问题, 提出了CNN与Transformer联合的深度学习框架, 其中, CNN负责提取局部细节, Transformer用于建模长程依赖与全局信息, 二者通过残差连接融合特征以增强表示能力。

(1) LLM Transformer架构: 由Vaswani等(2017)提出, 主要由3部分组成。(i) 自注意力机制: Transformer使用自注意力机制来处理输入数据, 该机制允许模型在处理每个单词时, 考虑到输入序列中的所有单词, 从而能够捕捉序列内复杂的依赖关系。(ii) 无需循环和卷积结构: 与传统的RNN或CNN不同, Transformer完全由注意力层和全连接层组成, 这使得它在处理长序列数据时, 可以更有效地并行计算。(iii) 层次化结构: Transformer模型通常包含多个相同的层, 每层包含两个主要的子层: 多头注意力(multi-head attention)层和前馈神经网络(feed-forward neural network)层。每个子层后面都跟有层归一化(layer normalization)和残差连接, 以促进深层网络的训练。

(2) 反射地震盲反演Transformer架构: (i) 地震数据输入嵌入: 地震数据蕴含了关键的地层反射信息; (ii) 位置编码: 由于模型本身不处理序列的顺序, 位置编码向每个输入嵌入添加位置信息, 这些位置信息记录了地震子波的时空特征和地震反射系数的时空特征; (iii) 解码: 核心运算是多头注意力与前馈神经网络。解码器由反射系数序列解码器和子波解码器组成。在我们的设计中, 每种解码器由2个相同层组成。每层中包含3个子层, 第一个子层: 多头自注意机制(“编码器-解码器关注”层), 查询(Q)来自前一层解码器, 键(K)和值(V)来自编码器的输出; 第二个子层: 全连接层组成的前馈网络。解码器还插入了第3个子层(masked multi-head attention), 其核心是: 在计算每一个元素时, 并不是和之前一样考虑所有的输入元素, 而是只考虑当前计算元素之前的元素, 并不考虑还未计算的元素。与编码器类似, 我们在每个子层周围残差网络的short-cut连接, 然后进行层归一化。(iv) 物理约束: 将上述两个解码器的输出输入到一个地震记录褶积模型正演层, 反射系数与地震子波的褶积, 生成合成地震记录。这个模型正演层是基于物理原理的, 因而网络具有很强的泛化能力, 是一种物理原理约束下的深度学习技术。(v) 反演结果输出: 分别输出的反射系数序列和地震子波。

(3) 标注数据集: 给定地震反射系数模型(如层状介质模型或Marmousi模型), 根据地震正演模型公式(51), 生成带有标注的合成地震数据集。

(4) 建立目标函数与模型训练: 网络训练所采用的损失目标函数 J 定义如下:

$$J(s, r, w, r_{\text{Label}}, w_{\text{Label}}) = \text{MSE}(r, r_{\text{Label}}) + \text{MSE}(w, w_{\text{Label}}) + \text{MSE}(s, \text{input}) \quad (53)$$

其中, s 表示地震记录即网络的输出, r 表示网络输出的反射系数序列, w 表示网络输出的子波, r_{Label} 和 w_{Label} 分别代表了反射系数序列的标签和子波的标签, $\text{MSE}(\cdot)$ 代表均方误差(mean square error, MSE)损失函数(构成了物理约束)。在训练中, 上述3个MSE函数也可以配以不同的加权系数。

我们将上述步骤(1)–(4)建立的网络叫作TBIS-net(transform-based blind inversion seismic networks)。作为一个简单的示例说明, 子波选用15–55 Hz的Ricker子波, 间隔0.2 Hz, 并对子波进行0°–40°的相位旋转, 相位旋转的间隔为0.1°, 子波长度为101。地震反射系数模型采用Marmousi模型的数据, 用0°、6°、12°、18°、24°、30°入射角生成地震反射系数序列, 并对地震记录添加10 dB–40 dB的随机噪声, 反射系数序列长度为1024。共计生成7万个带有标注的数据集, 并对数据集按照5:1:1划分为训练集、验证集和测试集。利用Transformer神经网络模型, 采用Mini-batch方法进行训练, 并采用Adam优化方法进行误差反传播的优化计算, 反射系数、地震子波的反演结果分别如图11和12所示, 具体的优化方法细节见第5.3节。常规盲反褶积结果是针对式(52)采用前面所述的ADMM算法获得的。结果表明, 基于Transformer神经网络可以获得与常规方法相媲美的结果: 地震反射系数的分辨率高、抗噪性好; 同时Transformer网络可以学习到地震子波, 与真实子波匹配度良好。

为了进一步提高Transformer在大规模地震数据处理、反演中的应用, 可以考虑基于大规模强化学习的推理技术路线(Mnih等, 2013; Sutton和Barto, 2018)。

5.1.6 物理信息引导的非地震数据深度学习

非地震地球物理勘探方法通常依赖于地面测量数据, 如重力、磁力、电阻率、诱导极化等, 基于物理信息引导的深度学习是利用物理方程(如位场积分方程、电磁波传播方程等)为网络的训练过程提供额外的约束, 能够有效提高地下反演的精度和可靠性。结合人工智能技术的地球物理探测方法, 在矿产资源勘查、环境监测、地下水资源评估等领域, 将发挥越来越重要的作用。

以位场探测为例。为解决三维重力与磁异常的联合反演问题并提高反演精度，可以采用基于模糊C均值(fuzzy C-means, FCM)聚类的约束损失函数，用于三维重磁数据的联合反演。该损失函数由两个部分组成：(1) MSE项，用于衡量预测模型与真实模型之间的差异，它是用来评价模型预测精度的指标；(2) FCM聚类约束项：通过数据驱动的分类约束，对地下介质单元进行聚类，从而提升地下模型的可靠性与稳定性。机器学习过程中极小化的目标函数如下：

$$J = \frac{1}{N} \sum_{i=1}^N \left(u_{\text{pred}_i} - u_{\text{true}_i} \right)^2 + \lambda \varphi_{\text{fcm}} \quad (54)$$

$$\varphi_{\text{fcm}} = \sum_{k=1}^C \sum_{i=1}^N u_{ki}^m d^2(u_{\text{pred}_i}, v_k)$$

其中， u_{pred_i} 表示第*i*个地下模型单元的预测值(如密度、磁化率等)； u_{true_i} 表示第*i*个地下模型单元的真实值或标签(由已知模型、测井数据或合成模型给出)；*N*为所有模型单元的总数，即样本总数； φ_{fcm} 刻画了基于FCM聚类的约束项(*C*为聚类数目，*m*为控制聚类的模糊程度)，用于引导模型结果在物性参数上更接近实际地质分区； $d^2(\cdot, \cdot)$ 表示加权距离函数(如加权Euclidean距离或加权Mahalanobis距离)； λ 为正则化因子(权重)，用于平衡MSE项与FCM约束项的重要程度； v_k 是第*k*类的聚类中心， u_{ki} 是 u_{pred_i} 对第*k*类的隶属度，表示 u_{pred_i} 属于第*k*类的程度，且 $u_{ki} \in [0, 1]$ ；对每一个样本*i*，有 $\sum_{k=1}^C u_{ki} = 1$ 。图13给出了物理信息引导的重磁联合反演深度学习框架示意图。

5.2 人工智能数据分析

5.2.1 数据生成策略

由于地下结构的复杂性和不可直接观察性，传统的地球物理测量方法常常受到数据质量、分辨率及各种噪声的影响。为此，生成高质量的地球物理数据，特别是通过模型模拟或数据驱动的方法，成为研究和实际应用中的重要问题。

传统的地球物理数据生成方式是基于物理模型的地球物理数据生成，比如通过建立地下介质的物理模型(如地震波速模型、电磁场模型等)，可以模拟不同条件下的地球物理响应。传统地球物理勘探需要大量采样点和长时间观测，这不仅成本高昂，还会带来大量冗余数据。因此可以基于压缩感知(Donoho, 2006)，利用稀疏性、高效采样和稀疏优化算法来降低数据获取成本，同时保持高质量的数据重建能力(Herrmann和Hennenfent, 2008; Wang等, 2011)。随着深度学习技术的发展，利用生成模

型(如GAN、扩散模型等)生成仿真地震数据，并用于训练地震反演网络或作为缺失数据的补充，是一个不错的选择。

借鉴于图像领域的生成技术，可以基于神经网络的生成模型，如GAN和变分自编码器(VAEs)，来生成地球物理数据。GAN通过两个神经网络(生成器和判别器)对抗训练生成模拟数据，旨在提升生成数据的真实性。数学上需要求解如下的随机优化问题，即

$$L := E_{x \sim p_{\text{data}}} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(G(z)))] \rightarrow \min \quad (55)$$

其中， $D(x)$ 是判别器对样本*x*是否为真实数据的判断， $G(z)$ 是生成器通过随机噪声*z*生成的数据。通过对抗训练，生成器能够不断优化，生成与真实数据分布相似的数据。

扩散模型(diffusion models)是一类基于概率生成的深度学习方法，近年来被广泛应用于图像与信号生成。核心思想是通过逐步添加噪声和去噪来学习数据分布，从而在给定随机噪声条件下生成符合地下真实结构的数据。主要包括前向过程(扩散)和反向过程(生成)两个步骤：

(1) 前向过程：逐步向真实地球物理数据 $x(0)$ 添加噪声，直到其接近高斯分布。数学上表示成

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (56)$$

其中， β_t 为噪声强度参数， x_t 表示在第*t*步的信号。

(2) 反向过程(生成)：训练一个参数化模型(如深度神经网络)去学习逆过程，从随机噪声逐步恢复出地球物理数据。数学上表示成

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (57)$$

模型通过学习 μ_θ 、 Σ_θ 来逼近真实分布。

统计物理学中的朗之万动力学(Langevin dynamics)是一种利用随机微分方程(stochastic differential equation)进行数据采样的策略。其迭代公式为

$$x_{k+1} = x_k - \frac{\eta}{2} \nabla_x E(x_k) + \sqrt{\eta} \xi_k \quad (58)$$

其中， $E(x)$ 为数据的能量函数(通常与概率密度函数 $p(x)$ 相关)， $E(x) = -\log p(x)$ ； η 为迭代步长； $\xi_k \sim \mathcal{N}(0, I)$ 表示高斯噪声。

在扩散模型的框架下，可以将朗之万动力学作为反向采样的一部分，用于在高维复杂地球物理数据空间中进行高效采样，即反向操作不单纯依赖神经网络预测均值，而是将预测梯度与随机扰动结合，模拟能量最小化采样过程。这样既能引入噪声保持多样性，又能引导采样朝着高概率区域收敛。比如，在扩散模型的每一步反向去噪时，用神经网络预测梯度 $\nabla_x \log p_\theta(x_t)$ ，再嵌入朗之万更新公式

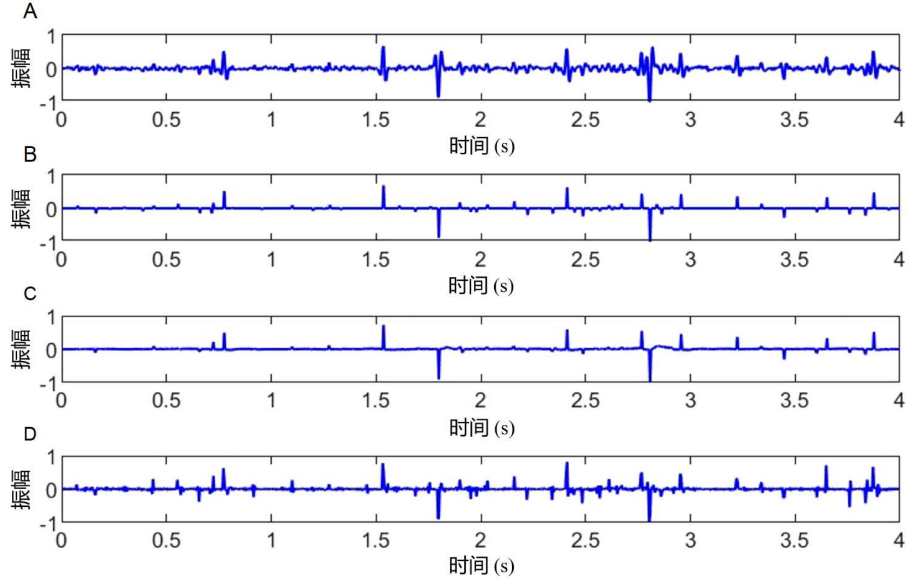


图 11 (A) 合成地震记录; (B) 真实反射系数; (C) 利用Transformer神经网络模型TBIS-Net学习结果; (D) 常规盲反褶积结果。改自邹安祺(2022)、王彦飞和邹安祺(2024)

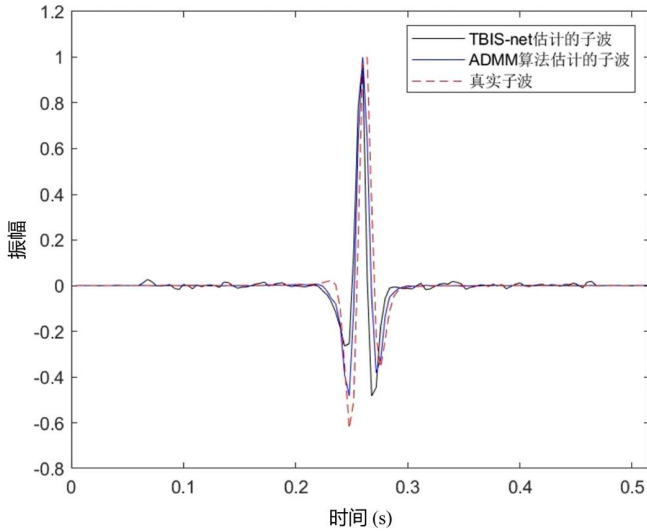


图 12 真实地震子波、传统算法(ADMM)估计的子波与Transformer神经网络模型学习(TBIS-net)的子波结果对比。改自邹安祺(2022)、王彦飞和邹安祺(2024)

$$x_{t-1} = x_t + \eta \nabla_x \log p_\theta(x_t) + \sqrt{2\eta} \xi_t \quad (59)$$

上述实现方式表明，朗之万动力学为扩散模型提供了一种基于能量的迭代采样机制，使得反向生成过程不仅依赖学习到的均值估计，还通过能量梯度与随机扰动探索更广泛的高概率区域，从而避免模型塌缩，使生成数据更接近真实地下结构分布。

5.2.2 轻量化神经网络深度学习

大模型通常具有更多的参数和更深的网络结构，在

处理复杂的任务时能提供更高的精度。但在传统深度神经网络中，深层卷积结构往往带来巨大的计算负担和存储负担。因为大模型通常需要大量的计算资源进行训练和推理，需要大量的标注数据进行训练，随着模型的复杂度增加，过拟合风险增加导致泛化性降低，导致科研人员对模型的理解和预测过程变得困难。要想在深度学习中实现“小模型”，通常涉及结构优化、参数压缩和高效推理等多方面技术。英伟达研究院和佐治亚理工大学的研究人员最近发表了一篇文章，指出在多智能体模型(multi-agent model)领域，小语言模型(SLM)才是AI智能体的未来，而不是那些庞大的大型语言模型(Belcak等, 2025)。

轻量化神经网络(lightweight neural networks)旨在保持较高精度的同时显著减少模型参数数量和计算量，适用于算力受限的地球物理应用、移动设备、边缘计算等场景，其目标是在保证模型精度的前提下，显著降低模型的参数数量和计算复杂度，从而实现低功耗、快速推理和易于部署的“小模型”。轻量化的核心技术主要包括：网络结构优化、模型压缩、高效推理与硬件优化。

在网络结构优化方面，典型方法是引入深度可分离卷积(depthwise separable convolution)、分组卷积(group convolution)与瓶颈结构(bottleneck)，代表性网络有MobileNet、ShuffleNet与EfficientNet-Lite，它们在网络层级设计上减少冗余计算(Sandler等, 2018; Tan和Le, 2019; Qin等, 2024)。

模型压缩方面，常用方法包括剪枝(pruning)、量化

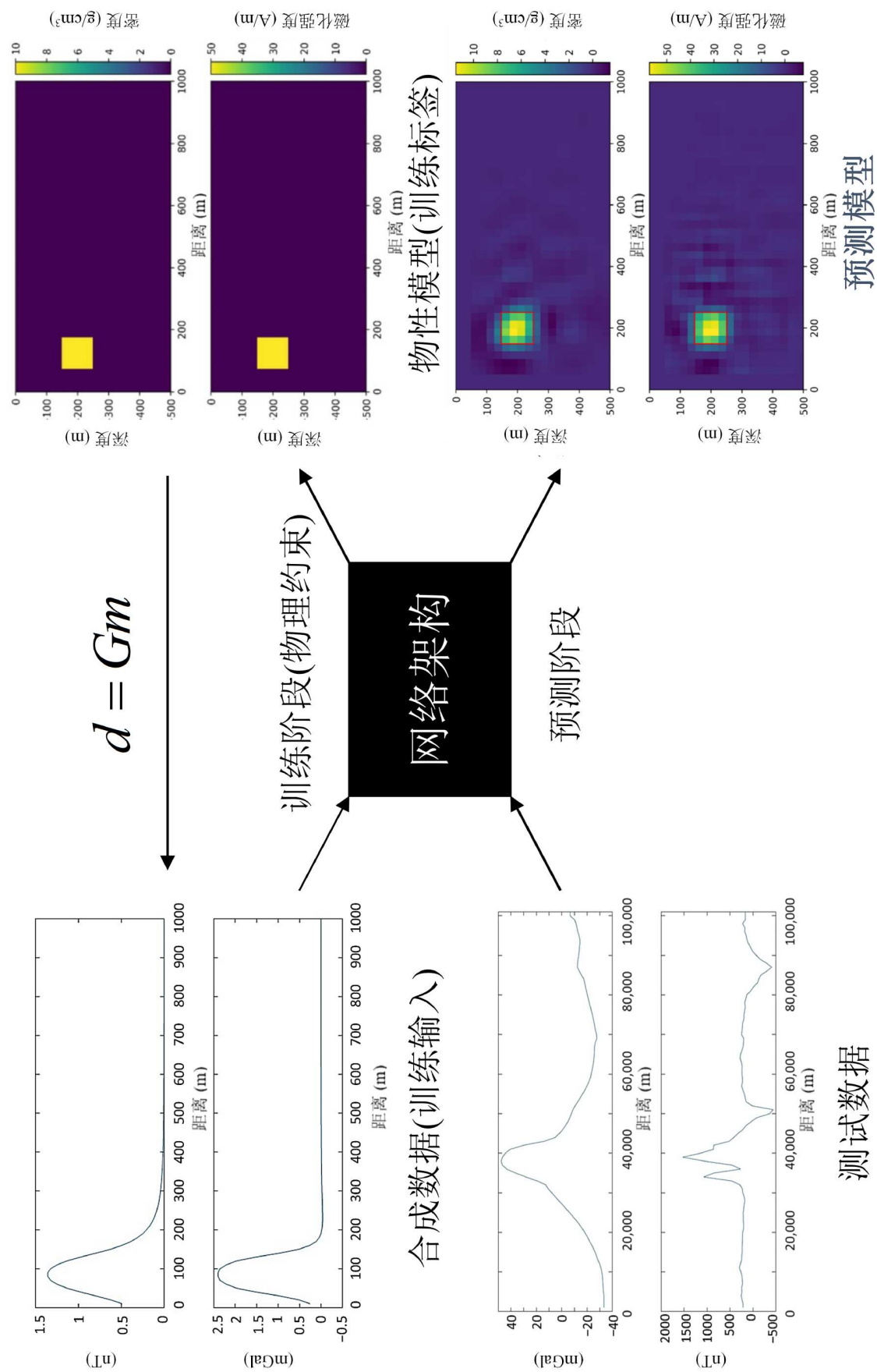


图 13 物理信息引导的重磁联合反演深度学习示意图，图中磁场单位为纳特(nT)，重力单位为毫伽(mGal)

(quantization)和知识蒸馏(knowledge distillation)。知识蒸馏利用大模型作为“教师”为小模型提供监督信号,提高小模型的泛化能力,比如DeepSeek采用的这种技术达到了惊人的学习效率。

高效推理与硬件优化方面,结合Facebook和微软2017年发布的ONNX(Open Neural Network Exchange)以及TensorRT等工具,或利用专用硬件(NPU、TPU)结合量化和混合精度计算,以及稀疏矩阵加速,使小模型在实际应用中兼顾性能与速度。

基于数据驱动的自动搜索策略,一个轻量化的网络是SeismicPatchNet(0.725697 M训练参数),该网络架构由拓扑模块和MS特征融合单元组成,用于对地震数据进行分类(Geng和Wang, 2020b)。该网络具有以下特点:(1)各网络层中的各卷积核的规格均不相同,取值受限于指定的搜索空间,防止特征图的无限增长并降低计算性能。(2)并行的MS卷积核种类丰富,但在各拓扑融合层输出时聚合为一个输出张量,作为下一网络层的输入。(3)在各拓扑融合层中,卷积核的尺寸和各种单元的数量均不相同,表明融合层并不是按常规网络方式将具有相同类型的层直接堆叠在一起。有趣的是,卷积核的数量与尺寸从较低融合层向较高融合层方向呈现出逐渐增长的趋势,表明更高的层可学习到更高级的抽象模式。(4)只考虑了常规且最常用的操作,例如卷积与特征采样,用于显著降低计算资源消耗并提升计算速度。现如今,一种可直接部署于微控制器的超微型智能模型也已经建立(Geng等, 2025)。与传统的大型深度学习模型相比,这种超微型模型不仅计算资源需求极低(~10 KB内存),而且能够在算力(~200 MHz)与成本极低的通用微控制器上高效运行。

引入知识的深度神经网络相比经典神经网络具有更大的优势:便于提取数据中的稀疏信息,采用很少的神经单元(节点)可逼近各种复杂函数,以及提升对未知数据的推理能力。深度学习是一个特殊的反问题,通常引入正则化项(如 l_p 、TV、KL、Wasserstein或熵正则化)来提升模型的预测泛化性能,增加网络参数的稀疏性(Bach等, 2012; Geng等, 2025)。特别地,在地球与环境科学探测领域,大模型对算力和数据量的要求过高,限制了其在极端应用场景中的可行性;因而计算资源需求极低、能在算力与成本极低的通用微控制器上高效运行、能部署于边缘端进行智能监测与自主决策的超微型模型(micro-model)将是未来发展的方向。进一步约束解空间增加可解释性是基于物理定律约束下的引导训练过程,通过向目标函数中引入物理相关的损失项实现(Karpatne等, 2017; Nabian等, 2021)。在基于物理定律引导的训练过

程中,如果网络预测结果偏离了物理指示,则对模型参数采取更大的更新步长。目标函数的损失值将向与常规优化方法不同的收敛方向递减,通常可能在相对更高的损失值处收敛,防止对原始数据的过拟合,并得到一个由物理定律约束的解。

5.3 智能计算中的数学优化

5.3.1 数学优化

机器学习是一门多领域交叉的学科,其中最优化理论和算法是机器学习的支柱学科,一直在机器学习理论和技术中发挥着重要作用(Shu等, 2023; Zhang等, 2024; Chen等, 2025; 郭田德等, 2025; 李晨毅和文再文, 2025)。网络模型在训练中相当于在做一个最优化过程,网络中的参数是损失函数的自变量。一个好的优化算法可以让损失函数有一个较快的收敛,目前比较典型的策略主要有随机梯度下降(stochastic gradient descent)、动量优化法(momentum optimization)、自适应学习率算法等。随机梯度下降包含多个变种,一般是指根据数据生成分布抽取其中几个小批量样本进行学习,在机器学习领域应用广泛,它是单一样本梯度下降和全量批梯度下降法的折中,在数据量较大的时候比全量批梯度下降法每轮的训练时间少,由于每次选择多组训练样本又规避了单一样本梯度下降方法中容易向非最优方向优化的问题,其中关键的参数是学习率 ϵ_k 的

选取: $\sum_{k=1}^{\infty} \epsilon_k = \infty, \sum_{k=1}^{\infty} \epsilon_k^2 < \infty$ 。

动量优化法主要是解决一般梯度下降法下降比较慢的情况,主要引入类比物体运动时候的动量调整前进的方向,运动物体受到外力的前进方向与当前的动量有关,在物理中动量是质量乘以速度,可以设置一个超参数表示之前的梯度对现在方向的影响,在很多情况下动量法可以解决黑塞矩阵(Hessian)病态的问题。

自适应学习算法有AdaGrad(Duchi等, 2011)、Adam(Kingma和Ba, 2015)等。以LLM为代表的大模型算法主要采用Adam随机优化算法。通常以MSE为目标函数来调整网络的参数或衡量模型在测试集上的预测值:

$$J(w) = \frac{1}{M} \sum_{i=1}^M L(f(x^{(i)}; w), y^{(i)}) \quad (60)$$

其中, $f(\cdot; w)$ 定义了一个比较广泛的函数族, w 为待定参数(学习得到), (x, y) 表示输入 x -输出 y 的数据对,比如以分类为例,输出 y 是分类器给出的输入 x 的一个映射。分类问题往往使用的是交叉熵损失函数,这是一种数据真实分布和估计分布相似性的度量方式。

动量法在数学上相当于每次选择更新下降方向的时候，都考虑了上一步下降方向，

$$v \leftarrow \alpha v - \tau \Delta_w \left[\frac{1}{M} \sum_{i=1}^M L(f(x^{(i)}), y^{(i)}) \right] \quad (61)$$

$$w = w + v \quad (62)$$

动量法有很多有名的变种比如Nesterov加速梯度算法(Nesterov, 1983; Sutskever等, 2013)等。Adam算法在Unet网络中也被经常使用，Adam算法的优势在于迭代中依次修正从原点初始化的一阶矩和二阶矩的估计，并且被认为对超参数的选择比较稳健。

5.3.2 Adam优化算法

Adam优化算法分为3步：

第一步：初始化

(1) 给定：步长(学习率) $\alpha > 0$ ；矩估计的指数衰减速率，超参数 $\rho_1, \rho_2 \in [0, 1]$ ；以及用于数值稳定的小常数 $\varepsilon > 0$ ；

(2) 输入初始参数 x_0 ；初始化一阶和二阶矩变量： $m_0=0, v_0=0$ ；

(3) 初始迭代指数： $t=0$ 。

第二步：对迭代指数 t 执行下列运算

(1) $t=t+1$ ；

(2) 从数据训练集中采用包含 p 个样本 $\{x^{(1)}, \dots, x^{(p)}\}$

的小批量，并计算梯度： $g_t = \frac{1}{p} \sum_{i=1}^p \nabla f_i(x_t^{(i)})$ ；

(3) 更新有偏一阶矩估计： $m_t = \rho_1 m_{t-1} + (1 - \rho_1) g_t$ ；

(4) 更新有偏二阶原始矩估计： $v_t = \rho_2 v_{t-1} + (1 - \rho_2) g_t \odot g_t$ ；

(5) 修正一阶矩： $\tilde{m}_t = \frac{m_t}{1 - \rho_1^t}$ ；

(6) 修正二阶矩： $\tilde{v}_t = \frac{v_t}{1 - \rho_2^t}$ ；

(7) 更新变量： $x_t = x_{t-1} - \alpha \frac{\tilde{m}_t}{\sqrt{\tilde{v}_t} + \varepsilon}$ 。

第三步：若满足收敛性条件，则输出 x_t 。

Adam的一个缺点是对权重衰减处理不当，可能导致比带动量的SGD(stochastic gradient descent)更差的结果。为了解决这一问题，提出了AdamW，它将权重衰减与Adam的参数更新解耦，不再作为参数更新的一部分，从而在更新过程中保持更高的稳定性并减少过拟合(Loshchilov和Hutter, 2017)。在构造梯度下降方向的时候，如果能利用Hessian矩阵或者Hessian矩阵的逼近，有时候会获得更好的加速收敛效果。比如，令 $s_k = w_{k+1} - w_k$ ，

$p_k = g_{k+1} - g_k$ ， H_0 为单位矩阵，通过迭代对构造的Hessian矩阵进行估计

$$H_{k+1} = \left(I - \frac{s_k p_k^T}{s_k^T p_k} \right) H_k \left(I - \frac{p_k s_k^T}{s_k^T p_k} \right) + \frac{s_k s_k^T}{s_k^T p_k} \quad (63)$$

则可以构造周期步长

$$\tau_k = \begin{cases} \tau_k^{\text{SD}}, & \text{if } \text{mod}(k, T) < 2 \\ \tau_{k-1}^{\text{F}}, & \text{if } \text{mod}(k, T) = 2 \\ \tau_{k-1}, & \text{else} \end{cases} \quad (64)$$

其中， T 是循环迭代周期，可以自由设定， $\text{mod}(k, T)$ 是迭代步数 k 对周期 T 的余数。其中 τ_k^{SD} 和 τ_k^{F} 有多种选取方式：

$$\begin{cases} \tau_k^{\text{SD}} = \frac{g_k^T g_k}{g_k^T H g_k} \\ \tau_k^{\text{F}} = F(\tau_k^{\text{SD}}, g_k) \end{cases} \quad (65)$$

其中， $F(\cdot)$ 为一个线性或非线性函数。 $F(\cdot)$ 的选择要使得构造的迭代公式满足梯度下降算法的收敛性(袁亚湘, 2008; Wang等, 2011)。

周期步长选取的随机梯度下降算法会导致迭代目标函数出现上下随机跳跃，即使是非凸函数，也可能会使优化方向从当前的局部极小值点跳到另一个更好的局部极小值点，随着迭代步数的增加，最终收敛于一个较好的局部极值点，甚至全局极值点。

5.4 知识涌现与尺度率

知识涌现(knowledge emergence)是指在复杂系统中，通过个体之间的相互作用和协同作用，产生出新的、整体性的知识或行为模式。这种涌现现象通常是不可预测的，仅通过分析个体行为难以推测出整体特性。知识涌现在多个领域中都有应用，如生物学中的群体行为、社会学中的社会动态，以及人工智能中的群体智能等。尺度定率/规模定律(scaling law)则是指在某些系统中，某些物理量随系统大小或尺度的变化呈现出一定的幂律关系，可以帮助人们理解不同尺度下系统的行为特征，并用于预测和分析复杂系统的行为。

LLM遵循着尺度定律(Kaplan等, 2020)。其描述的是模型的性能 L ，模型的参数量大小 N ，训练模型的数据大小 D ，以及训练模型使用的计算量 C 之间的关系： $N_{\text{opt}}(C), D_{\text{opt}}(C) = \underset{N, D, \text{Flops}(N, D)=C}{\text{argmin}} L(N, D)$ 。这里的尺度定律默认要求LLM使用的是Transformer的解码器结构。Kaplan等(2020)建立了 N 、 D 和性能 L 的联合幂律关系为：

$$L(N, D) = \left[(N_C / N)^{\frac{\alpha_N}{\alpha_D}} + D_C / D \right]^{\alpha_D}, \text{ 其中, } N_C, D_C, \alpha_N \text{ 和}$$

α_D 的具体数值取决于词汇表大小和分词方式,因此并不具备根本性的意义。Hoffmann等(2022)建立了 N 、 D 和性能 L 的联合幂律关系为: $L(N, D) = E + A/N^\alpha + B/D^\beta$,其中, α 和 β 为常数, E 是数据集自身的不可约误差, A 是模型容量(参数量 N)引起的近似误差的尺度因子, B 为样本有限(训练数据量 D)引起的估计误差的尺度因子。第一项捕捉了数据分布上理想生成过程的损失,应于自然文本的熵;第二项捕捉了一个拥有 N 个参数的完美训练的Transformer在理想生成过程下的表现不足;最后一项捕捉了Transformer未完全训练到收敛的事实,因为我们只在数据集分布的一个样本上进行了有限次数的优化步骤。

对于地球科学而言,地球探测涉及海量多模态数据(如卫星遥感、地震波、位场、电磁波、应力应变、地质样本等),通过大规模数据训练AI模型或部署分布式传感器网络,可能从复杂系统中涌现出新的规律,揭示出单一数据源难以发现的地质结构和动态过程,从而更好地理解地球内部的复杂过程。例如利用深度学习从数万张遥感图像中自动识别地质断层或预测气候变化趋势,这也可以看作一种知识涌现现象。模定率则体现在数据量、计算资源与探测精度的正向关联中——随着传感器密度、模型参数或算力的提升,系统性能(如地震预警时效性、矿产资源预测准确率)可能遵循对数或指数规律增强。典型案例如NASA(National Aeronautics and Space Administration)与IBM(International Business Machines Corporation)合作,结合几种最新Transformer架构的思想,训练了23亿参数,创建了天气和气候的AI基础模型Prithvi WxC,为人工智能天气预报勾画了美好的前景光明(Nguyen等, 2023)。

然而,无论是知识涌现还是尺度规律,对于地学而言都面临多重瓶颈:地球数据的噪声、碎片化特性可能干扰知识涌现的可靠性;规模扩展受限于计算成本(如训练全球高分辨率模型需千亿级参数)和硬件约束;涌现的“黑箱”结论也需与领域知识交叉验证。未来需突破自适应规模定率(如针对稀疏地质数据的优化模型)、人机协同推理(专家经验嵌入AI训练)等关键技术。若实现“数据-算力-领域知识”三角闭环,地球探测或能涌现出跨尺度规律,例如,通过百万级智能传感器实时监测地磁异常,结合超大规模模拟揭示板块运动的深层机制。

5.5 进一步思考

随着物理信息引导的深度学习技术的不断进步,未来地球物理反问题计算将更加精确和高效。通过将先验知识、物理机理与数据相结合,可以在提高反演精度的

同时,减少对大规模数据集的依赖,这对于数据稀缺或噪声较多的环境尤为重要。从计算框架角度讲,智能计算中的数学优化和轻量化神经网络将使得反演过程更加高效,尤其是在实时监测和在线反演中具有重要应用。对于地质智能体、智能终端建设而言,迫切需要发展模型即服务(model as a service, MAAS)和智能体即服务(agent as a service, AAAS),这对于深层地下资源精准探测和复杂灾害预警十分必要。在地球物理反演中,下一代前沿方法技术有可能是智能原生(AI native),即一种新的计算范式,以实时感知,充分数字化为基础,以通用人工智能(artificial general intelligence, AGI)为中心,综合多种算法进行智能决策,在整个过程中遵守智能优先原则。该范式通过自适应神经网络模型,从大数据中自主发现规律、生成知识,并根据环境或任务的变化进行动态调整、自我优化,不仅提高反演精度,还能在面临新的、复杂的地下环境时快速做出响应。

6 量子计算

6.1 量子计算机简介

量子计算机按功能划分主要为两种:通用量子计算机和量子退火机。

通用量子计算机基于量子门电路建造,是全面取代电子计算机的下一代高性能计算解决方案(Arute等, 2019; Zhong等, 2020, 2021; Wu等, 2021)。然而,通用量子计算机的建造还存在很多技术挑战,尤其是量子纠错技术还很不成熟,制约着应用级别通用量子计算机的建造。IBM(International Business Machines Corporation)于2021年11月发布了127量子比特的Eagle处理器(IBM; Newsroom, 2021),是一款大规模的通用量子计算机处理器,但距离实现百万量子比特的应用级别通用量子计算机尚有比较远的距离。2024年12月9日,Google量子AI公司推出了全新量子芯片Willow,拥有105个量子比特,通过将多个物理量子比特组合成一个逻辑量子比特,为实现实用QC提供了途径,且随着量子比特数量增加,逻辑错误率呈指数级下降(Abanin等, 2025)。2025年3月3日,中国科学院直属的中国科学技术大学与合作伙伴组成的科研团队在随机量子电路采样领域取得重大突破,该团队利用“祖冲之3号”超导量子计算原型机(配备105个量子比特和182个耦合器)进行实验,其运算速度比当前最强超级计算机快 10^{15} 倍,较谷歌最新公布成果快一百万倍(Gao等, 2025)。这一系列激动人心的进展显著提升了业界对于QC技术作为下一代高性能计算解决方案的信心。

与上面提到的通用型量子门计算机不同，量子退火机为专用型量子计算机，专门为解决NP(non-deterministic polynomial)复杂度的组合优化问题而研发(Finnila等, 1994; Kadowaki和Nishimori, 1998)。此方案可以有效地规避通用量子计算机面临的一些挑战，比如尚不成熟的量子纠错技术，利用现有的技术建造应用级别的量子退火机以高效快速地解决组合优化问题。目前，加拿大温哥华的D-Wave公司是唯一一家研发制造量子退火机的企业(<https://www.dwavequantum.com/>)。2011年，第一代商用量子退火机，128量子比特的D-Wave One系统面世。其后，经历10年的发展，最新的D-Wave Advantage系统(第五代系统)拥有5640个量子比特，在此基础上利用“量子+经典”混合计算技术可支持高达100万个参数的组合优化计算，初步达到工业应用级别(<https://www.dwavequantum.com/solutions-and-products/systems/>)。由于该量子系统设计的专用性，基于量子退火技术和量子退火机的QC渴望在解决地球科学大规模计算中发挥作用。

比特(Bit)是传统计算机使用的最小信息单位，表示为一个0或1的数据单元。而量子比特(Qubit)可以处于的状态除了 $|0\rangle$ 和 $|1\rangle$ ，还可以处于 $|0\rangle$ 和 $|1\rangle$ 的线性叠加态，表示为

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (66)$$

量子体系中除了这种态，还有一种不能用两个直积而生成的态，如贝尔(Bell)态：

$$|\psi_{\text{Bell}}\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle \quad (67)$$

无论如何没法分解成两个量子态的直积，称这样的态为量子的纠缠态。

在封闭体系中，一个量子态的演化可以用一个酉变换来描述。一个 t_1 时刻的量子态 $|\psi(t_1)\rangle$ 通过一个酉算子 U (Unitary operator)可以变换到 $|\psi(t_2)\rangle$ ，这个酉变换仅取决于时间 t_1 和 t_2 ：

$$|\psi(t_2)\rangle = U(t_2, t_1)|\psi(t_1)\rangle \quad (68)$$

类比于经典计算机，从一些门集合(如或门、与门和非门)出发，可以通过组合，完成任意的经典计算过程。

QC科学家研究了很多可以基于量子计算机的算法，往往可以把原本计算复杂度为NP或更高的问题转化为多项式复杂度，实现平方甚至指数级的加速。目前不少经典的机器学习问题，如主元素分析(PCA)、支持向量机(SVM)、GAN等都有了量子算法的理论加速版本，并且有的还在专用或通用的量子计算机中进行了原理性实验演示。现在人工智能若想要保持旺盛冲劲，必须克服摩

尔定律限制，结合最前沿的计算机计算途径和硬件性能去突破。QC的很多算法可以把AI程序涉及的计算复杂度变为多项式级，从根本上提升运算效率，这无疑是非常有吸引力的。

6.2 量子退火机

量子退火机D-Wave Advantage系统的能耗为25千瓦，与传统超级计算机相比，节能2–3个数量级，初步达到落地应用级别。2016年Google团队报道了D-Wave 2X系统(第三代系统，1152量子比特)在某些组合优化问题上相比单核CPU运行模拟退火算法和量子蒙特卡罗算法达到1亿倍的提速(Denchew等, 2016)。购买D-Wave系统的客户包括Lockheed Martin、Google、NASA、Los Alamos National Laboratory、Jülich Supercomputing Centre等众多企业及科研机构(https://en.wikipedia.org/wiki/D-Wave_Systems)。此外，普通用户可通过D-Wave Leap云平台接入D-Wave的量子退火服务系统(<https://www.dwavequantum.com/solutions-and-products/cloud-platform/>)。目前D-Wave Leap已经在北美和欧洲大面积开放，其他地区开放的国家包括澳大利亚、印度、日本、新加坡等。量子退火机以统计物理中的伊辛模型(Ising model)作为设计基础，这意味着它只能执行一类特殊的量子退火算法。

Kirkpatrick等(1983)将模拟退火(simulated annealing)算法引入组合优化问题的求解。模拟退火过程描述为在给定初始温度条件下，通过一定的进度缓慢地降低温度，从而使得模拟退火算法在搜索空间里能够找到最优解。与模拟退火利用热波动搜索最优解不同，量子退火(quantum annealing)算法利用量子系统波动产生的量子隧穿效应(quantum tunneling effect)来使算法摆脱局部最优解，从而以更大概率搜索到全局最优解(Finnila等, 1994; Kadowaki和Nishimori, 1998)。量子退火算法克服了传统模拟退火算法的降温速度慢、耗时久、计算量大等缺点，是一类新颖的高效求解组合优化问题的算法。

量子退火算法的流程概括如下：首先，定义一个含时哈密顿(Hamilton)量 $H(t)$ ：

$$H(t) = \left(1 - \frac{t}{T}\right)H_0 + \frac{t}{T}H_1 \quad (69)$$

其中， H_0 是系统的初始状态。假设系统有足够长的特征时间标度 T ，从而使得系统随薛定谔方程的演化是一个绝热过程。量子退火与模拟退火的对比示意图见图14。

6.2.1 量子退火算法

对于D-Wave量子退火计算机，其初始和终止哈密顿(Hamilton)量分别可以如下表示：

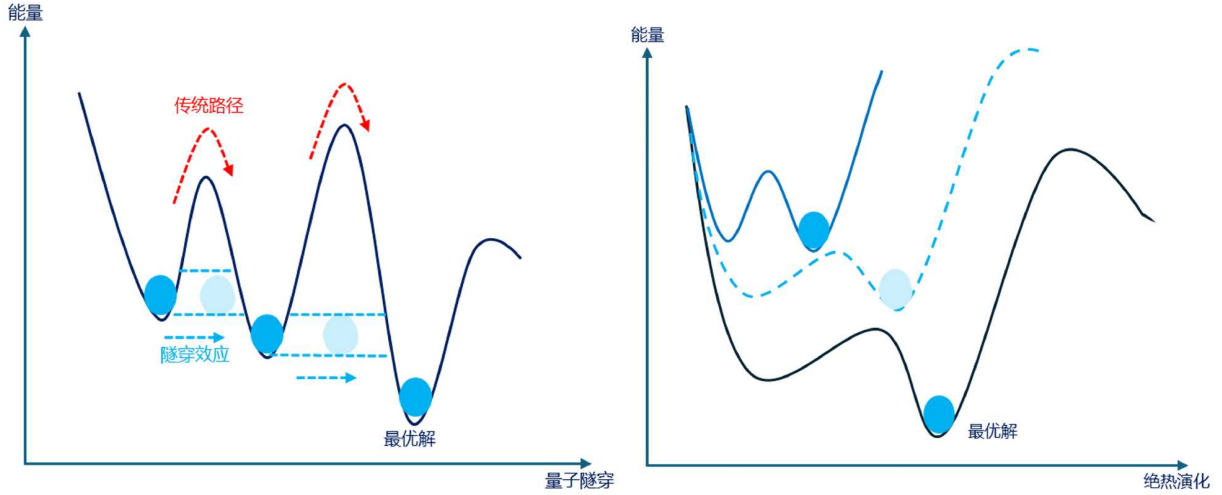


图 14 量子退火与模拟退火示意图

$$H_{\text{initial}} = \sum_i^n \hat{\sigma}_x^{(i)} \quad (70)$$

$$H_{\text{final}} = \sum_i^n h_i \hat{\sigma}_z^{(i)} + \sum_{i < j} J_{i,j} \hat{\sigma}_z^{(i)} \hat{\sigma}_z^{(j)}$$

其中, $\hat{\sigma}_l^{(i)}$, $l=x,y,z$ 表示作用在第 i 个量子比特上的泡利矩阵 (Pauli matrix), h_i 和 $J_{i,j}$ 分别表示偏置 (bias) 和耦合强度 (coupling strength)。当退火结束时, 量子比特退化到经典系统, 则 $\hat{\sigma}_z^{(i)}$ 可以通过自旋变量 $s_i = \pm 1$ 代替。此时系统可以通过经典的 Ising 模型 (Kadowaki 和 Nishimori, 1998) 来描述:

$$E_{\text{Ising}}(\mathbf{s}) = \sum_i^n h_i s_i + \sum_{i < j} J_{i,j} s_i s_j \quad (71)$$

这种表示形式等价于二次无约束二元最优化问题 (quadratic unconstrained binary optimization, QUBO), 形式如下:

$$E_{\text{QUBO}}(\mathbf{q}) = \sum_i^n Q_{i,i} q_i + \sum_{i < j} Q_{i,j} q_i q_j = \mathbf{q}^T \mathbf{Q} \mathbf{q} \quad (72)$$

其中, 变量 $\mathbf{q} = [q_1, q_2, \dots, q_n]^T$, $q_i \in \{0, 1\}$, $\mathbf{Q}$ 是由 h_i 与 $J_{i,j}$ 构成的矩阵。

使用量子退火求解线性系统, 需要将问题转化为 QUBO 形式, 首先要对待求变量 $\mathbf{x}$ 进行编码近似。受限于量子比特数量, 我们需要使用迭代方法逐步逼近, 以消除直接求解所带来的误差。对于第 k 次迭代, 令 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta \mathbf{x}$, 将目标函数

$$J(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \quad (73)$$

转化为

$$J(\Delta \mathbf{x}) = \|\mathbf{A}\Delta \mathbf{x} + \mathbf{A}\mathbf{x}^{(k)} - \mathbf{b}\|_2^2 \quad (74)$$

对于待求变量 $\Delta \mathbf{x}$ 按照下式进行近似编码近似:

$$\Delta \mathbf{x} = L\mathbf{B}\mathbf{q} \quad (75)$$

其中, L 是一个尺度因子, 补足编码范围。 $\mathbf{B} = \mathbf{I} \otimes \boldsymbol{\eta}_{\text{bv}}^T$, $\mathbf{I}$ 为单位矩阵, $\boldsymbol{\eta}_{\text{bv}} = [-2^{j_0-1}, 2^{j_0-2}, \dots, 2^{j_0-R}]^T$ 表示每一位比特所对应的编码值 (下标 bv 是 bit value 的缩写), R 表示编码长度, 这样做使得解的范围在合适的解空间。由此, 我们的目标函数便由式 (73) 转换为

$$J(\mathbf{q}) = \|\mathbf{M}\mathbf{q} - \mathbf{d}\|_2^2 \quad (76)$$

其中, $\mathbf{M} = L\mathbf{A}\mathbf{B}$, $\mathbf{d} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}$ 。

将式 (76) 转化为 QUBO 问题形式, 并使用量子退火器采样得到满足下式的 $\mathbf{q}^*$:

$$\mathbf{q}^* = \underset{\mathbf{q} \in \{0,1\}^{R \times m}}{\text{argmin}} \mathbf{q}^T \mathbf{Q} \mathbf{q} + C \quad (77)$$

其中, $\mathbf{Q} = \mathbf{M}^T \mathbf{M} - 2\text{diag}(\mathbf{d}^T \mathbf{M})$, $C = \mathbf{d}^T \mathbf{d}$, $\text{diag}(\cdot)$ 表示将矢量转化为对角阵。

6.2.2 量子退火算法求解正问题

以声波方程正演为例 (Wang 等, 2014)。利用有限差分离散化, 得到差分系数求解的线性系统: $\mathbf{A}\mathbf{x} = \mathbf{b}$ 。我们把上述量子退火算法用于差分系数的求解 (Wen 和 Wang, 2025)。

作为简单示例, 我们考虑了复杂地质体的正演模拟, 使用了 SEG (Society of Exploration Geophysicists) 盐速度模型 (BP2004) 的一部分 (Billette 和 Brandsberg-Dahl, 2005), 如图 15 所示。将空间采样间隔设置为 20 米, 时间步长为 1 毫秒, 选择主频为 20 Hz 的 Ricker 子波作为震源。传统方法、正则化方法和量子退火方法被用来找到长度为 $M=10$

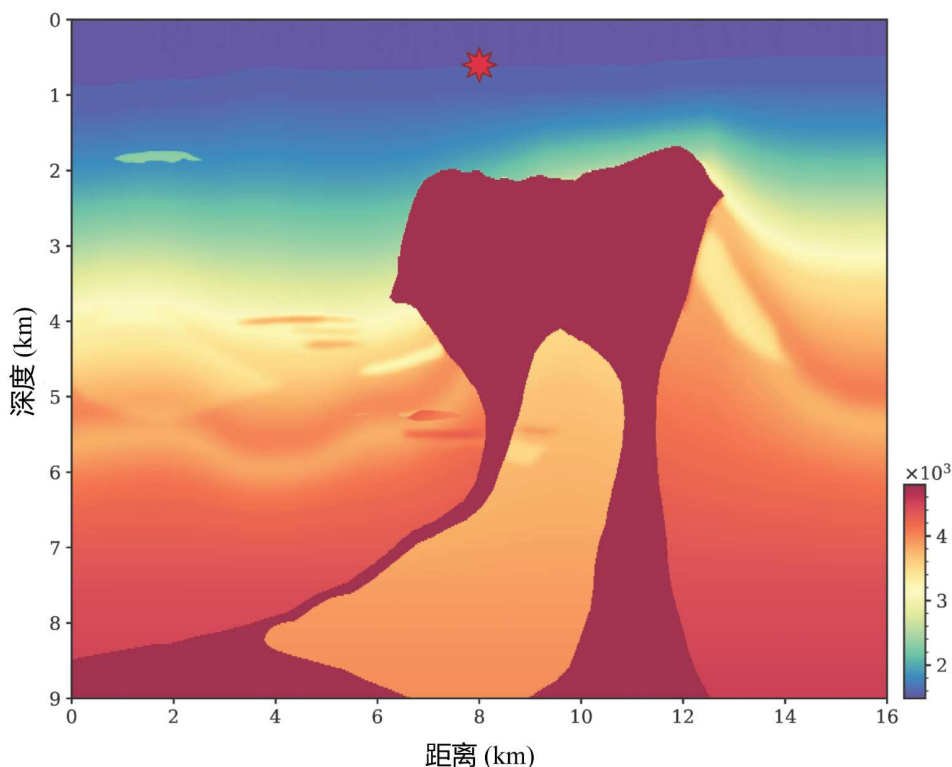


图15 BP2004盐丘模型和震源位置(红色八边形), 图中色标示意速度(m/s)。改自Wen和Wang(2025)

的差分算子。3种方法正演模拟的2公里处的地震记录如图16所示。

模型的速度变化范围是1486–4790 m/s。模型上部的速度较低, 导致地震波波长较短, 空间频散更为严重。对比发现, 量子退火方法的时空频散关系优于传统方法。其波形相位与正则化方法一致, 但频散能量更小, 波形更加清晰。

6.2.3 量子退火算法求解反问题

以层析成像为例(王彦飞和文玺翔, 2025)。地震层析成像(seismic tomography), 利用地震波在地球内部传播过程中的速度变化, 反演出地球内部的结构和物性参数。其本质上是对波动方程的高频近似, 因此假设在震源和接收器之间的波是沿着无限窄的射线路径传播。假设在 $\mathbf{x}$ 点位置的慢度为 $s(\mathbf{x})$, 那么地震射线传播经过的时间 t 便可通过与传播路径 l 的积分来表示: $t = \int_l s(\mathbf{x}) d\mathbf{l}$ 。积分离散化之后的等价于求解一个线性系统: $\mathbf{Ax}=\mathbf{b}$ 。

随机生成一个参考速度模型, 如图17所示。图中地表的倒三角表示震源的位置, 接收器布设在研究区域的底部, 接收器与震源间距相等。使用上文中提到的量子退火算法进行迭代求解, 选择本地CPU模拟退火器, 共

迭代50次。为了克服数据噪声带来的影响, 同样加入正则化方法。最终反演模型误差为 5.470×10^{-4} , 结果如图18所示。

6.3 量子编程

QC编程语言主要有以下几类: Qiskit、ProjectQ、D-Wave、TensorFlow Quantum和Qpanda。Qiskit以开源形式免费提供, 是一种采用Python语言的软件开发工具包。使用Qiskit时, 用户工作流程包括以下4个步骤: (1) 电路构建: 设计一个表示您正在考虑的问题的量子电路; (2) 电路编译: 为特定的量子服务编译电路, 例如量子系统或经典模拟器; (3) 线路运行: 在指定的量程服务上运行已编译的线路。这些服务可以是基于云的, 也可以是本地的; (4) 结果分析: 计算汇总统计数据并可视化实验结果。ProjectQ是苏黎世联邦理工学院启动的QC开源软件框架, 与IBM 5bit云上QC兼容。ProjectQ可以将程序转换为任何类型的后端, 无论是在经典计算机上运行的模拟器还是实际的量子芯片。ProjectQ是基于python编译、python/C++运行的, 使用时, 主要是用户直接编写门电路, 或者自定义函数编写门电路。由于是通过python编写, 开发比较友好; 执行时, ProjectQ会对输入的门电路编译优化执行。

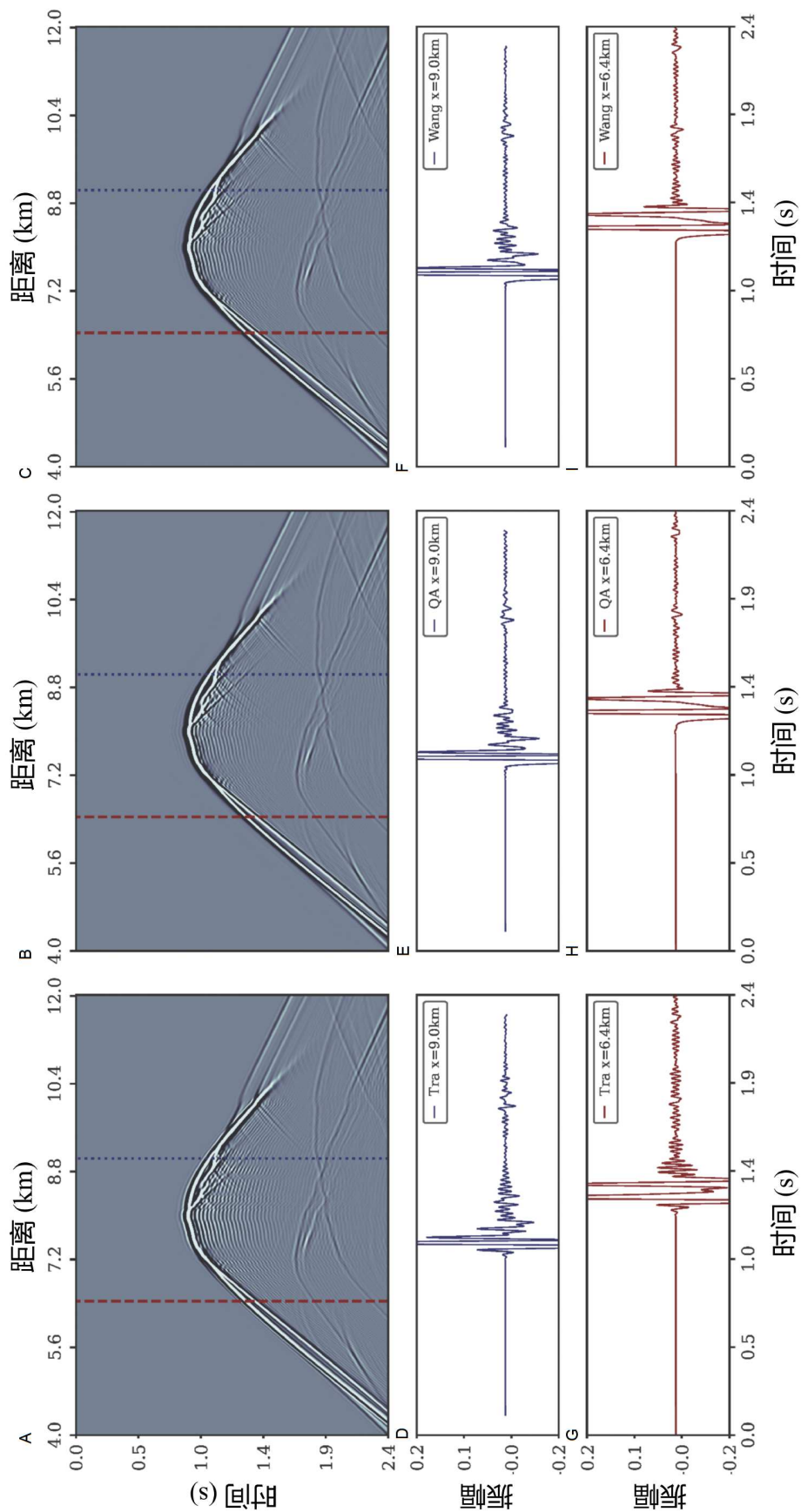


图 16 地震记录比较。(A-C) 在盐丘介质中使用常规有限差分算子(Tra)、量子退火算子(QA)、正则化方法(Wang)的二维声波方程地震记录；(D-F) 对应波场快照在9公里偏移处的切片；(G-I) 对应波场快照在6.4公里深度处的切片。改自Wen和Wang(2025)

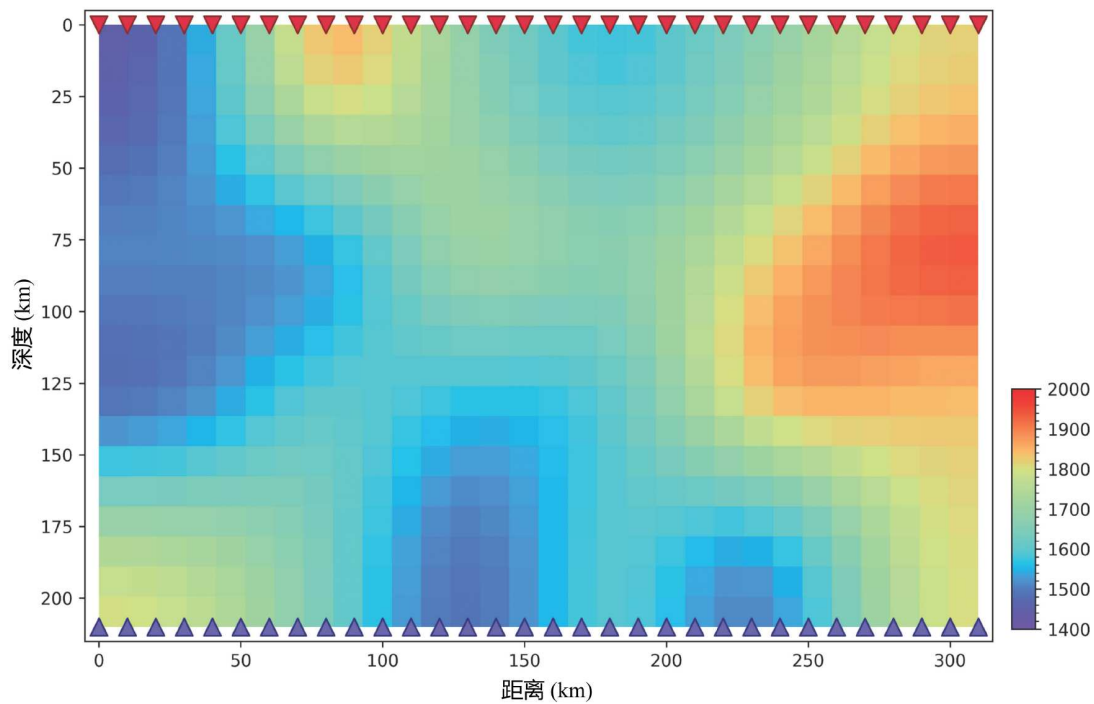


图 17 线性层析成像方法的参考速度模型，图中色标示意速度(m/s)。改自王彦飞和文玺翔(2025)

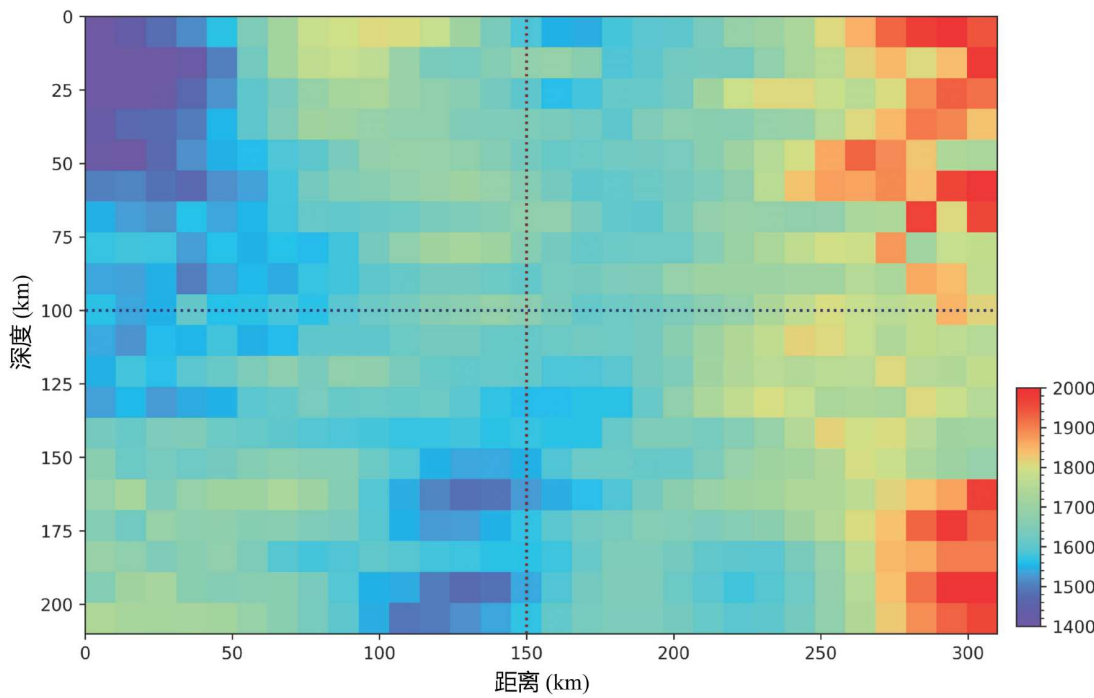


图 18 量子退火方法反演结果，图中色标示意速度(m/s)。改自王彦飞和文玺翔(2025)

6.4 其他量子算法

QC将在很大程度上有助于实现人工智能的工程目标，将其应用于各种AI系统，以加快计算过程，但设计

量子算法来解决某些人工智能问题确实非常困难，要达到同样的目的，QC的算法要比现有的经典算法更有效。

量子算法的研究一直是整个QC领域的驱动力，因为一些量子算法表明QC可以提供比经典计算更大的速度。

目前已经发展了3类量子算法，它们都比已知的经典算法有优势：

(1) 基于量子傅里叶变换的算法(QFT-based algorithms)，如用于分解和离散对数的Deutsch-Jozsa算法和Shor算法。Shor算法在量子计算机上分解整数 N 的时间复杂度为 $O((\log N)^3)$ ，几乎是对已知最有效的经典因数分解算法的 e 指数加速(Shor, 1997)，这种加速有可能在量子计算机上中断如RSA(Rivest-Shamir-Adleman)的现代加密机制(Rivest等, 1978)。此外，量子线性代数算法是研究得比较多的一类算法，特别是HHL(Harrow-Hassidim-Lloyd)算法是一个用量子计算机解决线性系统 $\mathbf{Ax}=\mathbf{b}$ 最优解的算法，可以用于处理和分析大规模数据集，被广泛应用于许多量子机器学习算法中(如SVM，主成分分析等)(Harrow等, 2009)。HHL算法有5个主要组成部分，即状态准备、量子相位估计(quantum phase estimation, QPE)、辅助位旋转、逆量子相位估计(inverse OPE)和测量。HHL算法的效率依赖于矩阵 $\mathbf{A}$ 的条件数 κ ，如果条件数很大，算法的效率会显著降低；此外，准确地制备初始量子态 $|b\rangle$ 和测量最终量子态 $|x\rangle$ 也是一个重要的问题。

(2) 非结构化量子搜索算法(quantum amplitude amplification)，即Grover算法及其扩展。Grover算法是一个非结构化的量子搜索算法(Grover, 1996)。这个算法是用于寻找能够通过一个黑箱函数(Oracle)产生特定输出对应最大可能的唯一输入态。该算法的时间复杂度为 $O(\sqrt{N})$ 与经典算法的时间复杂度 $O(N)$ 相比，Grover算法实现了计算的平方加速。

(3) 量子算法模拟量子系统(quantum simulation)，其基本思想可追溯至Feynman(1982)。其核心是直接利用量子系统模拟其他量子系统，比如第一性原理量子化学模拟等。

此外，随着含噪声中等规模量子(Noisy Intermediate-Scale Quantum, NISQ)时代的到来，变分量子算法(Variational Quantum Algorithm, VQA)也越来越多地被研究。这类算法采用经典-量子混合优化的结构，不仅为近期的量子化学模拟(如Variational Quantum Eigensolver, VQE)提供了可行路径，还被广泛拓展至经典组合优化(如Quantum Approximate Optimization Algorithm, QAOA)和量子机器学习领域(王彦飞和文玺翔, 2025)。

7 讨论

当下，AI for Science以及Science for AI已经深入人心(白春礼, 2021；符慧山等, 2025)，《中国科学院院刊》还在2024年第1期出版了一个相关专题“大力推进科研范

式变革”，对AI赋能的科研范式做了多方位的探讨。因此，跳出反问题的传统数理方程研究的范畴，在大数据的科学管理下，我们需要新的切入点，将云计算、智能计算和QC等技术应用到反问题中，研究复杂度可控的大数据统计分析算法。反问题求解的不适定性本质上决定了其解的不确定性和计算不稳定性，而人工智能特别是当前基于机器学习的弱人工智能成为关键解决手段。作为机器学习基石，数学优化方法贯穿学科发展全过程，其中随机梯度下降法因其计算高效性成为大数据时代的主流算法，尤其在深度学习中Robbins-Monro随机梯度法(SG)占据核心地位(Robbins和Monro, 1951)。然而SG方法存在并行性差、通信开销大等固有缺陷，难以适配新型计算架构。为此，需要重点发展二阶近似方法(如随机拟牛顿法、高斯牛顿法等)和小批量梯度下降法，这些方法兼具快速收敛性和抗非线性干扰能力，同时为构建新型并行分布式计算体系提供可能。

在机器学习的最优化过程中，最突出的挑战是问题的病态性。在深度神经网络中尤为明显，病态体现在随机梯度下降会“卡”在某些情况，此时即使很小的更新步长也会增加代价函数。并且尽管机器学习的很多方法都存在病态问题，但是有些适用于其他情况的解决病态的技术并不适用于神经网络。例如，牛顿法在解决带有病态条件的Hessian矩阵的凸优化问题时，是一个非常优秀的工具，但针对神经网络的优化问题时牛顿法还需要进行很大的改动。

深度学习面临的大多都是非凸优化问题，对于非凸函数，有可能会存在多个局部极小值。事实上，几乎所有的深度模型都会有非常多的局部极小值，并且这些局部极小值不是非凸性带来的问题。由于深度神经网络模型具有不可辨识性，这使得神经网络代价函数具有非常多，甚至无限多的局部极小值，且这些极小值具有相同的代价函数值，使得深度模型的局部极小值问题更为复杂。

在机器学习的高维非凸优化问题中，除了局部极值点外，鞍点是一种更为普遍存在的临界点，其特征是Hessian矩阵同时包含正负特征值。具体而言，在鞍点处，沿着正特征值对应特征向量方向的参数变化会导致代价函数值增大，而沿着负特征值方向则能获得更小的代价函数值。研究表明，这种临界点的分布特性与参数空间的维度密切相关：在低维空间中，局部极小值占据主导地位；随着维度升高，局部极小值变得极其稀少，而鞍点则成为最主要的临界点类型。值得注意的是，通过理论分析和实验验证，学者发现实际神经网络训练中遇到的鞍点往往对应着较高的代价函数值，这给优化过程带

来了显著挑战。

优化问题中除了局部极值和鞍点，还存在梯度与Hessian矩阵均为零的平坦区域。这种退化情形对所有数值优化算法都构成挑战：在凸优化中，平坦区域必然包含全局最优解；在非凸情况下，这类区域可能对应目标函数的高值平台，形成优化障碍。

深度学习网络由于层数过深，在反向传播过程中需要根据链式法则连续相乘各层梯度值，而每层梯度又由非线性激活函数的导数与该层权重共同决定，这使得非线性激活函数的导数值和初始权重的大小直接决定了梯度消失或爆炸的发生。梯度消失会导致参数更新方向难以确定，无法有效优化目标函数，而梯度爆炸则会造成训练过程不稳定。由于深层网络的梯度是各层梯度项的连乘积，当网络深度过大时，这种连乘效应会引发内在的不稳定性，使得深度神经网络的优化过程受到不稳定梯度问题的严重制约。此外，传统优化算法通常假设能够获得精确的梯度或Hessian矩阵信息，但在深度学习中，梯度计算往往依赖于随机采样的小批量训练样本，这些估计通常带有噪声，甚至偏差。这进一步加剧了深度学习优化问题的复杂性。

现有理论研究表明，部分结论仅适用于输出离散值的神经网络单元，而实际应用中大多数神经网络单元会输出连续光滑值，这使得基于梯度下降的局部搜索优化成为可能。虽然某些理论结果指出特定类型问题的计算是不可解的，但在实践中往往难以判断具体问题是否属于该类别；同时，尽管理论分析表明在给定网络规模下寻找可行解具有计算困难性，但实际训练时通过增加网络参数通常能够轻松找到可接受的解。值得注意的是，神经网络训练通常不追求损失函数的精确全局极小点，而是关注能否将其优化至足够低的水平以获得良好的泛化性能，然而从理论上分析优化算法能否实现这一目标仍极具挑战性。因此，探索优化算法在现实场景中的性能上界(而非理想条件下的理论极限)仍是机器学习领域的重要研究目标。

QC与量子人工智能(QAI)的结合代表了未来科技发展的前沿，无疑是下一代人工智能领域的重要研究方向，美国国家科学基金会(NSF)拨款7600万美元用于该领域的前沿研究(NSF, 2024)。2024年和2025年的诺贝尔物理学奖也分别授予了这两个领域，表明了科技发展的风向标。目前D-Wave Advantage系统的量子比特数量为5640个，这表明它在QC的硬件能力上具有相当大的潜力。量子比特的数量与QC的处理能力直接相关，但仅仅依赖量子比特的数量不足以完全反映QC的实际应用能力。对于地球物理反演任务而言，量子退火算法的应用

并不单纯依赖于量子比特的数量，还与算法设计、噪声容忍能力，以及QC系统在特定任务中的适用性有关。从理论上讲，与QC相比，经典计算机使用传统的冯·诺依曼架构进行计算，这使得其在处理大规模数据时的能耗相对较高，尤其在进行大量迭代计算(如地球物理反问题)时，能耗和计算时间都可能显著增加。目前的QC硬件，尽管有相对较高的量子比特数，但由于退相干时间、误差率等因素，QC的有效性会随着问题规模的增大而下降。因此，量子退火算法在地球物理反演中的应用仍局限于相对较小规模的问题。在当前阶段，QC硬件的“噪声耐受性”和“量子纠错”技术仍处于发展之中，无法处理非常复杂或大规模的反演任务。大规模的地球物理反问题通常需要处理 $O(10^7)$ 或更多参数，尤其当涉及复杂物理模型或大范围数据集时。量子退火算法在这类大规模问题上仍然面临挑战，因为需要大量的量子比特和量子门操作，而当前量子硬件的性能还难以有效支持这种级别的计算。

8 结论

本文较为系统地概述了地球物理反问题的多维度研究进展，涵盖了从模型驱动到数据驱动的反演方法、一般反问题的计算方法、智能计算与QC等多个角度。在模型驱动的地球物理反问题研究中，依赖于物理方程和理论模型，通过建立正演与反演的框架来连接观测数据与地下介质的物理参数。尽管此方法具有较强的物理依据，但也受到模型假设与数据不足的限制。数据驱动的地球物理反问题通过大数据与深度学习等技术，从观测数据中提取潜在的地球物理信息，展现了更高的灵活性与自动化能力，但也面临数据质量差和过拟合等问题的挑战。在实际落地应用中，人们往往面临众多不适定问题，这就需要充分利用各种先验信息对问题作适当形式的转换；同时受反问题中非线性问题的困扰，需要反复正反演迭代求解，这将导致巨大的计算量。因而，在反问题的发展中，我们应该从实际应用出发，充分研究问题的性质和特点，构造出精巧、快速的算法以适应生产的需求。

在一般反问题的计算方面，本文强调了正则化方法、最优化方法和统计反演的重要性，这些方法为解决反问题中的病态性和不适定性提供了可靠的计算数学理论支持。此外，随着人工智能技术的发展，智能计算逐渐成为地球物理反问题求解的有效工具。通过将机器学习与专家知识结合，智能计算为反问题的高效解法提供了新的途径和可能。

结合量子物理、人工智能、地球科学等方面的优势，我们计划利用QC和QAI技术对地学反演及相关的应用与

相关领域的科学家和公司企业开展合作研究。由于受到D-Wave系统规模的限制(~1000–2000量子比特)，地学反演均为小尺度系统(O'Malley, 2018; Greer和Malley, 2020; Souza等, 2022)。D-Wave Leap云平台中最新的可支持高达100万个参数的“量子+经典”混合计算技术使得突破小尺度系统的反演计算成为了可能。我们计划利用这一技术进行更大规模的反演计算，为落地应用做算法储备。QC作为前沿的计算方法，在大规模并行计算与复杂优化问题中展现出巨大潜力，可能会为未来的地球物理反问题的求解带来颠覆性的创新。

致谢 课题组的各位同事和研究生团队给予了讨论和协助，责任编辑和审稿人提出了宝贵的修改建议，在此一并致谢。

参考文献

- 白春礼. 2021. 科技革命与产业变革：趋势与启示. 科技导报, 39: 11–14
- 戴或虹, 袁亚湘. 2000. 非线性共轭梯度法. 北京: 科学出版社
- 符慧山, 赵亮, 李志伟, 等. 2025. 地球物理学、空间物理学和大地测量学与人类文明. 科学通报, 70: 1–9
- 郭田德, 李安琪, 韩丛英. 2025. 组合优化问题的机器学习求解方法. 中国科学: 数学, 55: 451–480
- 何继善, 鲍力知. 1990. 可控源音频大地电磁法. 长沙: 中南大学出版社
- 李晨毅, 文再文. 2025. 基于Lean的数学形式化简介. 计算数学, 47: 191–213
- 李红蕾, 陈石, 庄建仓, 等. 2021. 贝叶斯同化重力反演方法构建龙门山地壳密度模型. 地球物理学报, 64: 1236–1252
- 李貅. 2002. 瞬变电磁测深的理论与应用. 陕西: 陕西科学技术出版社
- 刘光鼎. 2018. 地球物理通论. 北京: 科学出版社
- 刘浩洋, 户将, 李勇锋, 等. 2020. 最优化: 建模、算法与理论. 北京: 高等教育出版社
- 宁亚鑫. 2022. 含水合物地层保幅成像方法及人工智能地震属性参数反演. 博士学位论文. 北京: 中国科学院大学
- 单连璞, 于彩霞, 王彦飞. 2025. 物理信息神经网络求解亥姆霍兹方程的配置点采样方法. 煤田地质与勘探, 53: 1–9
- 孙卫涛. 2009. 弹性波动方程的有限差分数值方法. 北京: 清华大学出版社
- 汤井田, 何继善. 2005. 可控源音频大地电磁法及其应用. 长沙: 中南大学出版社
- 王彦飞. 2007. 反演问题的计算方法及其应用. 北京: 高等教育出版社
- 王彦飞, 邹安琪. 2024. 地震反射系数和地震子波同时反演的方法及装置. CN202410831631.1
- 王彦飞, 宁亚鑫. 2025. 天然气水合物的位置确定方法、装置和电子设备. CN202311540847.4
- 王彦飞, 文玺翔. 2025. 反问题的量子计算方法. 数值计算与计算机应用, 46: 95–115
- 王彦飞, 斯捷潘诺娃 I E, 提塔连科 V N, 等. 2011. 地球物理数值反演问题. 北京: 高等教育出版社
- 王彦飞, 沃尔科夫 V T, 亚格拉 A G. 2021. 反问题基本理论: 变分分析及在地球科学中的应用. 北京: 科学出版社
- 王彦飞, 高静怀, 金双根, 等. 2026. 综合地球物理联合反演理论、方法及智能计算. 北京: 科学出版社
- 肖庭延, 于慎根, 王彦飞. 2003. 反问题的数值解法. 北京: 科学出版社
- 杨顶辉, 董兴朋, 黄建东, 等. 2025. 高分辨率全波形地震成像研究——进展、挑战与展望. 中国科学: 地球科学, 55: 319–347
- 杨胜雄, 梁金强, 陆敬安, 等. 2017. 南海北部神狐海域天然气水合物成藏特征及主控因素新认识. 地学前缘, 24: 1–14
- 袁亚湘. 2008. 非线性优化计算方法. 北京: 科学出版社
- 中国科学院. 2020. 中国学科发展战略-数学优化. 北京: 科学出版社
- 邹安琪. 2022. 波阻抗反演与人工智能分析理论与方法. 博士学位论文. 北京: 中国科学院大学
- 邹婧博, 刘财, 赵鹏飞. 2023. 地震波正演问题的物理信息神经网络算法研究进展. 地球物理学进展, 38: 430–448
- Abanin D A, Acharya R, Aghababaie-Beni L, et al. 2025. Observation of constructive interference at the edge of quantum ergodicity. *Nature*, 646: 825–830
- Acharya R, Abanin D A, Aghababaie-Beni L, et al. 2025. Quantum error correction below the surface code threshold. *Nature*, 638: 920–926
- Aki K, Richards P G. 2002. Quantitative Seismology. 2nd ed. Mill Valley: University Science Books
- Alford R M, Kelly K R, Boore D M. 1974. Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, 39: 834–842
- Araya-Polo M, Jennings J, Adler A, et al. 2018. Deep-learning tomography. *Leading Edge*, 37: 58–66
- Arute F, Arya K, Babbush R, et al. 2019. Quantum supremacy using a programmable superconducting processor. *Nature*, 574: 505–510
- Bach F, Jenatton R, Mairal J. 2012. Optimization with sparsity-inducing penalties. *Found Trends Mach Learn*, 4: 1–106
- Bai Z J, Wang Y F, Wang C Z, et al. 2024. Joint gravity and magnetic inversion using CNNs' Deep Learning. *Remote Sens*, 16: 1115
- Bao G, Chow S N, Li P J, et al. 2010. Numerical solution of an inverse medium scattering problem with a stochastic source. *Inverse Probl*, 26: 074014
- Bauer F, Pereverzev S, Rosasco L. 2007. On regularization algorithms in learning theory. *J Complexity*, 23: 52–72
- Baysal E, Kosloff D D, Sherwood J W C. 1983. Reverse time migration. *Geophysics*, 48: 1514–1524
- Belcak P, Heinrich G, Diao S Z, et al. 2025. Small language models are the future of agentic AI. \arXiv{2506.02153}
- Benioff P. 1980. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *J Stat Phys*, 22: 563–591
- Bergen K J, Johnson P A, de Hoop M V, et al. 2019. Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363: eaau0323
- Billette F, Brandsberg-Dahl S. 2005. The 2004 BP velocity benchmark. In: 67th EAGE Conference and Exhibition. Madrid.
- Blanchard G, Mücke N. 2018. Optimal rates for regularization of statistical inverse learning problems. *Found Comput Math*, 18: 971–1013
- Boyd S, Parikh N, Chu E, et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends Mach Le*, 3: 1–122
- Boys B, Girolami M, Pidstrigach J, et al. 2023. Tweedie moment projected diffusions for inverse problems. \arXiv{2310.06721}
- Brenier Y. 1991. Polar factorization and monotone rearrangement of vector-valued functions. *Comm Pure Appl Math*, 44: 375–417
- Candes E J, Wakin M B. 2008. An introduction to compressive sampling. *IEEE Signal Process Mag*, 25: 21–30
- Candes E J, Romberg J, Tao T. 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inform Theor*, 52: 489–509
- Carcione J M. 2007. Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic, Porous and Electromagnetic Media. 2nd ed. Amsterdam: Elsevier
- Carcione J M. 2010. A generalization of the Fourier pseudospectral method. *Geophysics*, 75: A53–A56
- Carcione J M, Cavallini F, Mainardi F, et al. 2002. Time-domain seismic modeling of constant-Q wave propagation using fractional derivatives.

- Pure Appl Geophys, 159: 1719–1736
- Caves C M, Fuchs C A, Schack R. 2002. Quantum probabilities as Bayesian probabilities. *Phys Rev A*, 65: 022305
- Chavez R E, Bailey R C, Garland G D. 1987. Joint interpretation of gravity and magnetic data over axial symmetric bodies with application to the darnley bay anomaly, NWT Canada. *Geophys Prospect*, 35: 374–392
- Chen J B. 2007. High-order time discretizations in seismic modeling. *Geophysics*, 72: SM115–SM122
- Chen Y M, Zhang Y, Liu Y, et al. 2025. A memory efficient randomized subspace optimization method for training large language models. *arXiv {2502.07222}*
- Claerbout J F. 1971. Toward a unified theory of reflector mapping. *Geophysics*, 36: 467–481
- Dablain M A. 1986. The application of high-order differencing to the scalar wave equation. *Geophysics*, 51: 54–66
- Denchev V S, Boixo S, Isakov S V, et al. 2016. What is the computational value of finite-range tunneling? *Phys Rev X*, 6: 031015
- Donoho D L. 2006. Compressed sensing. *IEEE Trans Inform Theor*, 52: 1289–1306
- Dorigo M, Stützle T. 2004. Ant Colony Optimization. Cambridge: MIT Press
- Dou Z, Song Y. 2024. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In: Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024), Vienna, Austria
- Duchi J, Hazan E, Singer Y. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res*, 12: 257–269
- E W, Han J, Jentzen A. 2017. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun Math Stat*, 5: 349–380
- Engl H W, Hanke M, Neubauer A. 1996. Regularization of Inverse Problem. Dordrecht: Kluwer Academic Publishers
- Etgen J T. 2007. A Tutorial on Optimizing Time Domain Finite-Difference Scheme: “Beyond Holberg”. Stanford Exploration Project Report. 129: 33–43
- Evensen G. 2009. Data Assimilation: The Ensemble Kalman Filter. Berlin: Springer Science & Business Media
- Feynman R P. 1982. Simulating physics with computers. *Int J Theor Phys*, 21: 467–488
- Figueiredo M A T, Nowak R D, Wright S J. 2007. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J Sel Top Signal Process*, 1: 586–597
- Finnila A B, Gomez M A, Sebenik C, et al. 1994. Quantum annealing: A new method for minimizing multidimensional functions. *Chem Phys Lett*, 219: 343–348
- Fridman V. 1956. Methods of successive approximation for Fredholm integral equations of the first kind (in Russian). *Uspekhi Mat Nauk*, 11: 233–234
- Fu J C, Fan R, Cao J X, et al. 2023. Seismic impedance inversion using a joint deep learning model based on convolutional neural network and transformer. *IEEE J Sel Top Appl Earth Observations Remote Sens*, 16: 8913–8922
- Fuchs C A. 2010. QBism, the perimeter of quantum Bayesianism. *Quant Inf Comp*, 10: 616–623
- Gallardo L A, Meju M A. 2003. Characterization of heterogeneous near-surface materials by joint 2D inversion of dc resistivity and seismic data. *Geophys Res Lett*, 30: 183–196
- Gao D X, Fan D J, Zha C, et al. 2025. Establishing a new benchmark in quantum computational advantage with 105-qubit Zuchongzhi 3.0 Processor. *Phys Rev Lett*, 134: 090601
- Gao J, Zhang H J. 2018. An efficient sequential strategy for realizing cross-gradient joint inversion: Method and its application to 2-D cross borehole seismic traveltime and DC resistivity tomography. *Geophys J Int*, 213: 1044–1055
- Geng Z, Wang Y F. 2020a. Physics-guided deep learning for predicting geological drilling risk of wellbore instability using seismic attributes data. *Eng Geol*, 279: 105857
- Geng Z, Wang Y F. 2020b. Automated design of a convolutional neural network with multi-scale filters for cost-efficient seismic data classification. *Nat Commun*, 11: 3311
- Geng Z, Wang Y F, Pan W Y, et al. 2025. Real-time discrimination of earthquake signals by integrating artificial intelligence technology into IoT devices. *Commun Earth Environ*, 6: 73
- Goldstein T, Osher S. 2009. The split Bregman method for l_1 -regularized problems. *SIAM J Imag Sci*, 2: 323–343
- Greer S, Malley D O. 2020. An approach to seismic inversion with quantum annealing. SEG Technical Program Expanded Abstracts, 2845–2849
- Grigoriu M. 2006. Evaluation of Karhunen–L  ve, spectral, and sampling representations for stochastic processes. *J Eng Mech*, 132: 179–189
- Grover L K. 1996. A fast quantum mechanical algorithm for database search, STOC’96. In: Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing. Philadelphia. 212–219
- Guasch L, Calder  n A O, Tang M X, et al. 2020. Full-waveform inversion imaging of the human brain. *npj Digit Med*, 3: 28
- Guillen A, Menichetti V. 1984. Gravity and magnetic inversion with minimization of a specific functional. *Geophysics*, 49: 1354–1360
- Guitton A, Kaelin B, Biondi B. 2007. Least-squares attenuation of reverse-time-migration artifacts. *Geophysics*, 72: S19–S23
- Haltmeier M, Leit  o A, Scherzer O. 2007. Kaczmarz methods for regularizing nonlinear ill-posed equations I: Convergence analysis. *Inverse Problems Imag*, 1: 289–298
- Han J, Jentzen A, E W. 2018. Overcoming the curse of dimensionality: Solving high-dimensional partial differential equations using deep learning. *Proc Natl Acad Sci USA*, 115: 8505–8510
- Han S, Zhang H, Gao L, et al. 2022. Joint inversion of body wave arrival times, surface wave dispersion data and receiver functions: Method and application to south China. *J Geophys Res-Solid Earth*, 127: e2022JB024083
- Hanke M. 1997. Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. *Numer Funct Anal Optim*, 18: 971–993
- Harrow A W, Hassidim A, Lloyd S. 2009. Quantum algorithm for linear systems of equations. *Phys Rev Lett*, 103: 150502
- He Q L, Wang Y F. 2021. Reparameterized full-waveform inversion using deep neural networks. *Geophysics*, 86: V1–V13
- Healey R. 2023. Quantum-Bayesian and pragmatist views of quantum theory. The Stanford Encyclopedia of Philosophy
- Hein T, Kazimierski K S. 2010. Accelerated Landweber iteration in Banach spaces. *Inverse Probl*, 26: 1037–1050
- Hennessy J. 2018. The End of Moore’s Law & Faster General Purpose Computing, and a New Golden Age. In: Proceedings of DARPA Electronics Resurgence Initiative (ERI) Summit 2018. San Francisco
- Herrmann F J, Hennenfent G. 2008. Non-parametric seismic data recovery with curvelet frames. *Geophys J Int*, 173: 233–248
- Hettlich F. 1998. The Landweber iteration applied to inverse conductive scattering problems. *Inverse Probl*, 14: 931–947
- Hoffmann J, Borgeaud S, Mensch A, et al. 2022. Training compute-optimal large language models. *arXiv {2203.15556}*
- Hornik K. 1991. Approximation capabilities of multilayer feedforward networks. *Neural Netws*, 4: 251–257
- IBM Newsroom. 2021. IBM unveils breakthrough 127-qubit quantum processor. IBM Newsroom, Nov 16, 2021. <https://summit.quantum-computing.ibm.com/home>
- Jin P, Feng Y, Feng S, et al. 2024. An empirical study of large-scale data-

- driven full waveform inversion. *Sci Rep*, 14: 20034
- Jo J H, Ha W. 2024. Seismic traveltime tomography using transfer learning. *IEEE Trans Geosci Remote Sens*, 62: 1–14
- Kadowaki T, Nishimori H. 1998. Quantum annealing in the transverse Ising model. *Phys Rev E*, 58: 5355–5363
- Kalnay E. 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge: Cambridge University Press
- Kaltenbacher B. 1997. Some Newton-type methods for the regularization of nonlinear ill-posed problems. *Inverse Probl*, 13: 729–753
- Kangro I, Kangro R, Vaarmann O. 2013. Some approximate Gauss-Newton-type methods for nonlinear ill-posed problems. *P Est Acad Sci*, 62: 227–237
- Kaplan J, McCandlish S, Henighan T, et al. 2020. Scaling laws for neural language models. [\arXiv{2001.08361}](#)
- Karpatne A, Atluri G, Faghmous J H, et al. 2017. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE T Knowl Data En*, 29: 2318–2331
- Kawar B, Vaksman G, Elad M. 2021. SNIPS: Solving noisy inverse problems stochastically, NIPS’21. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. New York. 21757–21769
- Kelly K R, Ward R W, Treitel S, et al. 1976. Synthetic seismograms: A finite-difference approach. *Geophysics*, 41: 2–27
- King J T. 1992. Multilevel algorithms for ill-posed problems. *Numer Math*, 61: 311–334
- Kingma D P, Ba J. 2015. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego
- Kirkpatrick S, Gelatt Jr. C D, Vecchi M P. 1983. Optimization by simulated annealing. *Science*, 220: 671–680
- Kirsch A. 1996. An Introduction to the Mathematical Theory of Inverse Problems. New York: Springer-Verlag
- Kuang W, Yuan C, Zhang J. 2021. Real-time determination of earthquake focal mechanism via deep learning. *Nat Commun*, 12: 1432
- Landweber L. 1951. An iteration formula for Fredholm integral equations of the first kind. *Amer J Math*, 73: 615–624
- Leitão A, Alves M M. 2012. On Landweber–Kaczmarz methods for regularizing systems of ill-posed equations in Banach spaces. *Inverse Probl*, 28: 104008
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature*, 521: 436–444
- Levander A R. 1988. Fourth-order finite-difference P-SV seismograms. *Geophysics*, 53: 1425–1436
- Levy A, Rubinstein J. 1999. Hilbert-space Karhunen–Loève transform with application to image analysis. *J Opt Soc Am A*, 16: 28–35
- Li M X, Cao H, Yang Z F, et al. 2025. Intelligent prestack multitrace seismic inversion constrained by probabilistic geologic information. *Geophysics*, 90: IM15–IM34
- Li X, Liu F, Ma C, et al. 2024. Land data assimilation: Harmonizing theory and data in land surface process studies. *Rev Geophys*, 62: e2022RG000801
- Liang W Q, Wu X, Wang Y F, et al. 2018. A simplified staggered-grid finite-difference scheme and its linear solution for the first-order acoustic wave-equation modeling. *J Comput Phys*, 374: 863–872
- Liang S L, Li X, Xie X H. 2013. Land Surface Observation, Modeling and Data Assimilation. Singapor: World Scientific Publishing Company
- Liu B, Jiang P, Wang Q, et al. 2023. Physics-driven self-supervised learning system for seismic velocity inversion. *Geophysics*, 88: R145–R161
- Liu Y, Sen M K. 2009. A new time–space domain high-order finite-difference method for the acoustic wave equation. *J Comput Phys*, 228: 8779–8806
- Loshchilov I, Hutter F. 2017. Decoupled weight decay regularization. [\arXiv{1711.05101}](#)
- O’Malley D. 2018. An approach to quantum-computational hydrologic inverse analysis. *Sci Rep*, 8: 6919
- McMechan G A. 1983. Migration by extrapolation of time-dependent boundary values. *Geophys Prospect*, 31: 413–420
- Meng T, Jing X Y, Yan Z, et al. 2020. A survey on machine learning for data fusion. *Inf Fusion*, 57: 115–129
- Meng X, Kabashima Y. 2022. Diffusion model based posterior sampling for noisy linear inverse problems. [\arXiv{2211.12343}](#)
- Mnih V, Kavukcuoglu K, Silver D, et al. 2013. Playing Atari with deep reinforcement learning. In: Proceedings of NIPS Deep Learning Workshop. Nevada
- Moorkamp M, Jegen M D, Heincke B, et al. 2011. A framework for 3D joint inversion of mt, gravity and seismic refraction data (invited). *Geophys J Int*, 184: 477–493
- Mousavi S M, Beroza G C. 2020. A machine-learning approach for earthquake magnitude estimation. *Geophys Res Lett*, 47: e2019GL085976
- Nabian M A, Gladstone R J, Meidani H. 2021. Efficient training of physics-informed neural networks via importance sampling. *Comput Aided Civil Eng*, 36: 962–977
- Nashed M Z, Votruba G F. 1976. A unified operator theory of generalized inverses. Generalized Inverses and Applications, 1–109
- Nesterov Y. 1983. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Math Doklady*, 27: 372–376
- Nguyen T, Brandstetter J, Kapoor A, et al. 2023. ClimaX: A foundation model for weather and climate. [\arXiv{2301.10343}](#)
- Ning Y X, Wang Y F. 2021. Q-compensated reverse-time migration for natural gas hydrate using fractional viscoacoustic wave equation. *Energy Rep*, 7: 8505–8521
- NSF Funds New Crop of Big Research Centers. 2024. <https://ww2.aip.org/fyi/nsf-funds-new-crop-of-big-research-centers>
- Pan W, Wang Y F. 2020. On the influence of different misfit functions for attenuation estimation in viscoelastic full-waveform inversion: Synthetic study. *Geophys J Int*, 221: 1292–1319
- Pan W, Innanen K A, Wang Y F. 2020. Parameterization analysis and field validation of VTI-elastic full-waveform inversion in a walk-away vertical seismic profile configuration. *Geophysics*, 85: B87–B107
- Pradeep D, Rajan M P. 2016. A regularized iterative scheme for solving nonlinear ill-posed problems. *Numer Funct Anal Optimization*, 37: 342–362
- Price K E, Storn R, Lampinen J. 2005. Differential Evolution: A Practical Approach to Global Optimization. Berlin: Springer
- Qin D F, Lechner C, Delakis M, et al. 2024. MobileNetV4: Universal models for the mobile ecosystem. In: Proceedings of Computer Vision–ECCV 2024: 18th European Conference. Milan. 78–96
- Raissi M, Perdikaris P, Karniadakis G E. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys*, 378: 686–707
- Rasht-Behesht M, Huber C, Shukla K, et al. 2022. Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions. *J Geophys Res-Solid Earth*, 127: e2021JB023120
- Rasmussen H J, Hohmann G W. 1991. Cooperative inversion of transient electromagnetic and gravity data for depth determination in a basin. *SEG Tech Pro Expand Abst*, 1: 1646
- Ray A, Alumbaugh D L, Hoversten G M. 2013. Robust and accelerated Bayesian inversion of marine controlled-source electromagnetic data using parallel tempering. *Geophysics*, 78: 271–E280
- Reichstein M, Camps-Valls G, Stevens B, et al. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566: 195–204
- Rivest R L, Shamir A, Adleman L. 1978. A method for obtaining digital

- signatures and public-key cryptosystems. *Commun ACM*, 21: 120–126
- Robbins H, Monro S. 1951. A stochastic approximation method. *Ann Math Statist*, 22: 400–407
- Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention. Munich. 234–241
- Ross Z E, Meier M A, Hauksson E. 2018. P wave arrival picking and first-motion polarity determination with deep learning. *J Geophys Res-Solid Earth*, 123: 5120–5129
- Sandler M, Howard A, Zhu M, et al. 2018. MobileNetV2: Inverted residuals and linear bottlenecks. In: Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. 4510–4520
- Scherzer O, Engl H W, Kunisch K. 1993. Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. *SIAM J Numer Anal*, 30: 1796–1838
- Shor P W. 1997. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J Comput*, 26: 1484–1509
- Shu J, Meng D Y, Xu Z B. 2023. Learning an explicit hyperparameter prediction policy conditioned on tasks. *J Mach Learn Res*, 24: 1–74
- Song C, Alkhalifah T, Waheed U B. 2022. A versatile framework to solve the Helmholtz equation using physics-informed neural networks. *Geophys J Int*, 228: 1750–1762
- Song J, Vahdat A, Mardani M, et al. 2023. Pseudoinverse-guided diffusion models for inverse problems. In Proceedings of the International Conference on Learning Representations (ICLR 2023). Kigali, Rwanda
- Souza A M, Martins E O, Roditi I, et al. 2022. An application of quantum annealing computing to seismic inversion. *Front Phys*, 9: 748285
- Su H, Zhang L, Zhao J. 2025. SPIKE: Stable physics-informed kernel evolution method for solving hyperbolic conservation laws. *arXiv* {2510.18266}
- Sun H F, Zhang N Y, Li D R, et al. 2023. The first semi-airborne transient electromagnetic survey for tunnel investigation in very complex terrain areas. *Tunn Undergr Sp Tech*, 132: 104893
- Sun X M, Wang Y F, Yang X, et al. 2022. Three-dimensional transient electromagnetic inversion with optimal transport. *J Inverse Ill-posed Pro*, 30: 549–565
- Sun J, Innanen K, Zhang T, et al. 2023. Implicit seismic full waveform inversion with deep neural representation. *J Geophys Res-Solid Earth*, 128: e2022JB025964
- Sutton R S, Barto A G. 2018. Reinforcement Learning: An Introduction. 2nd ed. Cambridge: The MIT Press
- Sutskever I, Martens J, Dahl G, et al. 2013. On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta. 1139–1147
- Tarantola A. 1986. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51: 1893–1903
- Tan M X, Le Q. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning. California. 6105–6114
- Teixeira F L, Sarris C, Zhang Y, et al. 2023. Finite-difference time-domain methods. *Nat Rev Methods Primers*, 3: 75
- Tautenhahn U, Jin Q. 2003. Tikhonov regularization and *a posteriori* rules for solving nonlinear ill posed problems. *Inverse Probl*, 19: 1–21
- Tikhonov A N. 2009. Inverse and Improperly Posed Problems (Collection of Scientific Works in Ten Volumes, Vol. III). Moscow: Nauka.
- Vaswani A, Shazeer N M, Parmar N, et al. 2017. Attention is all you need. NIPS'17. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. California. 6000–6010
- Virieux J. 1986. P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 51: 889–901
- Virieux J, Calandra H, Plessix R É. 2011. A review of the spectral, pseudo-spectral, finite-difference and finite-element modelling techniques for geophysical imaging. *Geophys Prospect*, 59: 794–813
- Puzirev V. 2019. Deep learning electromagnetic inversion with convolutional neural networks. *Geophys J Int*, 218: 817–832
- von Baeyer H C. 2016. QBism: The future of quantum physics. Cambridge: Harvard University Press
- Wang F, Chen S C, Liu Y A. 2019. Deep learning for gravity and magnetic data interpolation. SEG Technical Program Expanded Abstracts, 5407: 1052–3812
- Wang J, Xiao Z W, Liu C, et al. 2019. Deep learning for picking seismic arrival times. *J Geophys Res-Solid Earth*, 124: 6612–6624
- Wang L, Lee E W M, Yuen R K K. 2018. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. *Appl Energy*, 228: 1740–1753
- Wang L P, Wang Y F. 2018. A joint matrix minimization approach for seismic wavefield recovery. *Sci Rep*, 8: 2188
- Wang S, Wang H, Perdikaris P. 2021. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Science Advance*, 7: eabi8605
- Wang T Q, Wang Y F. 2022. High-resolution seismic faults interpretation based on adversarial neural networks with a regularization technique. *Geophysics*, 87: IM207–IM219
- Wang W, McMechan G A, Ma J. 2021. Elastic isotropic and anisotropic full-waveform inversions using automatic differentiation for gradient calculations in a framework of recurrent neural networks. *Geophysics*, 86: R795–R810
- Wang Y, Wang Y F. 2023a. Spectral decomposition and multifrequency joint amplitude-variation-with-offset inversion based on the neural network. *Geophysics*, 88: R373–R383
- Wang Y, Wang Y F. 2023b. Quantitative evaluation of gas hydrate reservoir by AVO attributes analysis based on the Brekhovskikh equation. *Pet Sci*, 20: 2045–2059
- Wang Y F, Yuan Y X. 2005. Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems. *Inverse Probl*, 21: 821–838
- Wang Y F, Yang C C, Li X W. 2008. A regularizing kernel-based BRDF model inversion method for ill-posed land surface parameter retrieval using smoothness constraint. *J Geophys Res*, 113: D13101
- Wang Y F, Yagola A G, Yang C C. 2011. Optimization and Regularization for Computational Inverse Problems and Applications. New York: Springer
- Wang Y F, Liang W Q, Nashed Z, et al. 2014. Seismic modeling by optimizing regularized staggered-grid finite-difference operators using a time-space-domain dispersion-relationship-preserving method. *Geophysics*, 79: T277–T285
- Wang Y F, Rong L L, Qiu L Q, et al. 2019. Magnetic susceptibility inversion method with full tensor gradient data using low-temperature SQUIDS. *Pet Sci*, 16: 794–807
- Wang Y F, Leonov A S, Lukyanenko D V, et al. 2020. General Tikhonov regularization with applications in geoscience. *CSIAM-AM*, 1: 53–85
- Wen X X, Wang Y F. 2025. Seismic wave propagation simulation with quantum computing. *J Geophys Res Mach Learn Comput*, 2: e2024H000473
- Wei N Y. 2023. Joint 3D inversion of gravity and magnetic data using deep learning neural networks. In: Proceedings of Second International Meeting for Applied Geoscience & Energy. Houston
- Whitmore N D. 1983. Iterative depth migration by backward time propagation. SEG Technical Program Expanded Abstracts, 382–385
- Wu X, Xue G Q, Zhao Y, et al. 2022. A deep learning estimation of the earth resistivity model for the airborne transient electromagnetic observation. *J*

- Geophys Res-Solid Earth*, 127: e2021JB023185
- Wu X M, Liang L M, Shi Y Z, et al. 2019. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, 84: IM35–IM45
- Wu X M, Yan S S, Qi J, et al. 2020. Deep learning for characterizing paleokarst collapse features in 3-D seismic images. *J Geophys Res-Solid Earth*, 125: e2020JB019685
- Wu X M, Ma J W, Si X, et al. 2023. Sensing prior constraints in deep neural networks for solving exploration geophysical problems. *Proc Natl Acad Sci USA*, 120: e2219573120
- Wu Y L, Bao W S, Cao S R, et al. 2021. Strong quantum computational advantage using a superconducting quantum processor. *Phys Rev Lett*, 127: 180501
- Wu Y L, McMechan G A. 2019. Parametric convolutional neural network-domain full-waveform inversion. *Geophysics*, 84: R881–R896
- Wu Y L, McMechan G A, Wang Y F. 2022. Adaptive feedback convolutional-neural-network-based high-resolution reflection-waveform inversion. *J Geophys Res-Solid Earth*, 127: e2022JB024138
- Xu F M, Wang Y F. 2018. Recovery of seismic wavefields by an Lq-norm constrained regularization method. *Inverse Probl Imag*, 12: 1157–1172
- Yan B P, Zhao J Q, Peng K K, et al. 2025. 3D karst cave recognition using TransUnet with dual attention mechanisms in seismic images. *Geophysics*, 90: IM133–IM143
- Yang F, Ma J. 2019. Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics*, 84: R583–R599
- Zhang B, Chu W, Berner J, et al. 2025. Improving diffusion inverse problem solving with decoupled noise annealing. [\arXiv{2407.01521}](#)
- Zhang H W, Zhang B. 2013. A novel integral equation for scattering by locally rough surfaces and application to the inverse problem. *SIAM J Appl Math*, 73: 1811–1829
- Zhang J H, Yao Z X. 2013. Optimized explicit finite-difference schemes for spatial derivatives using maximum norm. *J Comput Phys*, 250: 511–526
- Zhang W, Gao J H. 2021. Deep-learning full-waveform inversion using seismic migration images. *IEEE T Geosci Remote*, 60: 1–18
- Zhang Y S, Rybin D, Luo Z Q. 2024. Finite horizon optimization: Framework and applications. [\arXiv{2412.21068}](#)
- Zhdanov M S. 2002. *Geophysical Inverse Theory and Regularization Problems*. Amsterdam: Elsevier
- Zhong H S, Wang H, Deng Y H, et al. 2020. Quantum computational advantage using photons. *Science*, 370: 1460–1463
- Zhong H S, Deng Y H, Qin J, et al. 2021. Phase-programmable Gaussian Boson sampling using stimulated squeezed light. *Phys Rev Lett*, 127: 180502
- Zhu T, Harris J M. 2014. Modeling acoustic wave propagation in heterogeneous attenuating media using decoupled fractional Laplacians. *Geophysics*, 79: T105–T116
- Zhu W, Beroza G C. 2019. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophys J Int*, 216: 261–273
- Zhu W, Xu K, Darve E, et al. 2022. Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification. *Geophysics*, 87: R93–R109
- Zou A Q, Wang Y F, Wang D H. 2022. Blind inversion of multichannel nonstationary seismic data for acoustic impedance and wavelet. *Pure Appl Geophys*, 179: 2147–2166

(责任编辑: 徐涛)