

第1章

全波形反演的意义和基本原理

1.1 引言

全波形反演(full-waveform inversion, FWI)是一种基于波动方程的高分辨率地球物理参数反演技术。它通过数值求解波动方程得到全波场数据(包含直达波、反射波、折射波、散射波等全部波动信息),并将其与观测数据匹配优化,迭代修正地下介质的速度、密度或其他弹性参数模型,使得在目标函数定义的范数意义下实现对观测数据的最佳拟合,最终获得与真实地质结构高度吻合的地下物性模型。其核心目标是利用观测波形的振幅、相位及走时等完整信息,最大限度地恢复地下介质的精细结构和物理性质。

FWI 的概念和基本原理框架是由法国地球物理学教授 Tarantola^[1]于 1984 年提出的。Tarantola 运用泛函分析的方法推导了在声波近似下体积模量、密度和震源子波的梯度公式。该公式应用到了伴随状态(adjoint-state)法,因此可以通过反传伴随源(adjoint source)计算梯度,从而巧妙地规避了求解雅可比矩阵(Jacobian matrix),极大地减少 FWI 所需要的计算机内存和计算量,使 FWI 成为一种可行的技术。这是 FWI 的第一个里程碑。

FWI 的第二个里程碑是 1999 年地球物理学家 Pratt^[2]将 FWI 成功地应用到基于物理模型,模拟井间地震的超声数据上。在理论上,Pratt 用矩阵代数重新描述了 Tarantola 基于泛函分析的 FWI 理论,使 FWI 的反演理论更加通俗易懂,并指出 FWI 的极限分辨率是半波长的分辨率^①,远高于层析成像的分辨率,即第一菲涅耳(Fresnel)带宽度。在当时仍然受限于计算机的速度和内存的容量,Pratt 通过在频率域求解波动方程实现了二维声波波场的高效模拟^[3]。相对于时间域的波场模拟,在频率域进行波场模拟有两个优势:首先,只需要分解波动方程的阻抗矩阵和进行一次求逆,就可以快速地实现多炮模拟;其次,在频率域里,可以通过引入

^① 这里的半波长分辨率是指反演模型的最小波长是地震信号映射到模型域的波长的一半。当考虑“瑞利判据”时,则极限分辨率为 1/4 波长,和通常地震勘探中所说的“地震分辨率为 1/4 波长”吻合。

复数速度方便地表达地层品质因子 Q ^[4]。但是,目前的 FWI 基本使用时间域的波场模拟。这主要有两个原因:首先,频率域模拟需要显式构建阻抗矩阵,三维模拟需要的内存远超目前的硬件条件;其次,频率域的单频地震数据的预处理过程(如切除等)较为烦琐,不够便捷。

FWI 的第三个里程碑是 2009 年英国石油公司(British Petroleum, BP)的地球物理学家 Sirgue 博士等^[5]将三维频率域 FWI 成功地应用到了 Valhall 油田的三维海底电缆(ocean bottom cable, OBC)数据上,这是首次用工业界的地震数据展示出了 FWI 的强大能力:既能修正背景速度模型,又能恢复高分辨率的构造信息(参见文献[6]中图 14)。从此,FWI 也开始得到了学术界和工业界的广泛关注和大量研发投入。直接用数值方法求解三维频率域波动方程非常困难, Sirgue 博士巧妙地使用时间域显式有限差分方法求解波动方程,在每一个时间步上通过傅里叶级数提取反演需要的几个离散频率,从而间接地实现频率域的三维波场模拟^[7]。之后, Warner 等^[8]将时间域的三维各向异性声波方程应用到 FWI 中,成功地反演了北海 Tommeliten 油田的 OBC 数据。相较于传统的射线层析,FWI 更精确地刻画了气云的边界,并显著提升了后续进行的深度偏移成像中背斜形态的准确性和连续性。

同年,国际勘探地球物理学家学会(Society of Exploration Geophysicists, SEG)在阿曼举行的关于 FWI 的研讨会打出的广告词为“Full-waveform inversion has emerged as the final and ultimate solution to the Earth resolution and imaging objective”,突出地强调 FWI 是终极反演、成像方法。这种观点也引来了地球物理领域研究其他成像方法同行的质疑。如休斯敦大学的 Weglein 教授认为声波全波形反演运用了“错误的数学模型”(声波方程)、“错误数据”(纯 P 波数据)和“错误的方法”(迭代的方法)反演更为复杂的地球模型^[9]。在某种意义上, Weglein 教授对早期的 FWI 理论,即 Tarantola 在 1984 年提出的理论^[1]的质疑具有一定的合理性。Tarantola 在 1984 年的理论是基于声波方程利用梯度迭代法来反演纯 P 波数据,以实现 P 波速度和密度的恢复。然而,地球是一个弹性体,除了 P 波速度和密度,还包括 S 波速度和一些其他地震参数,因此声波方程不能准确地描述实际地震波的传播过程。如果要反演 P 波、S 波速度和密度,仅依靠纯 P 波数据是不充分的。另外,由于 FWI 的非线性,基于梯度的迭代方法可能使 FWI 陷入局部极小点。Weglein 教授的质疑也为 FWI 指明了主要研究方向。

经过过去十多年的工业化发展,FWI 在工业界已经取得了巨大的成功。特别是在海洋油气勘探领域,FWI 技术取得了突出的效果,如墨西哥湾的盐丘复杂构造^[10-11]、北海气云^[5, 8]以及我国的南海泥底辟^[12]和潜山^[13]等。这些成功归功于:①地震数据采集技术的进步,包括海底节点(ocean bottom node, OBN)^[14-16]和宽/全方位长拖缆采集^[17];②FWI 算法的进步^[18-20]。但是,FWI 在陆地地震数据的应用还存在很大的挑战,其核心原因在于复杂近地表的影响,导致波动方程模

拟的地震数据和观测到的地震数据不匹配,即两者之间的差异无法用波动方程描述。

本章后续内容的结构安排如下:1.2节概述 FWI 的意义;1.3节概述 FWI 的数学原理,并以各向同性声波方程为例分别采用“矩阵法”和“拉格朗日乘法”推导 FWI 的伴随状态方程和梯度公式;1.4节概述 FWI 的程序设计,包括并行程序结构、波场重构和反演效率的优化。

1.2 全波形反演的意义

在全波形反演技术实现工业化应用之前,工业界地震数据建模的主要方法是基于射线的初至波走时层析(first-arrival travel-time tomography)和反射波走时层析(reflection travel-time tomography),地震数据的成像方法是偏移成像(migration imaging)。层析建模和偏移成像是两个独立的过程,各自存在一定的局限性。而全波形反演技术的出现成功克服了这些局限性,为地震数据的成像提供了一种“终极方法”。全波形反演技术的意义在于:

1) 填补了层析和偏移的分辨率缺口

在射线层析技术中,地震波在透射过程中的走时是通过射线路径来计算的。射线是对地震波的高频近似,其假设地震波的频率无限高,而实际的勘探地震波的频率在 5~100 Hz(注:工业界在努力拓展地震激发和接收的频带)。当有限频带的地震波透射过速度异常体后,受地震波衍射的作用,只有当速度异常体尺度大于第一菲涅耳带(first Fresnel zone)的宽度时,在检波器处接收到的地震波才有走时响应。否则,检波器的记录无响应。这种现象也叫“波前愈合”(wavefront healing)。因此,透射地震波的分辨率为第一菲涅耳带宽度。虽然极高频射线的第一菲涅耳带宽度为零,对地下异常体的分辨率为无穷大,但是透射地震波本身的分辨率限制了所有层析方法的分辨率也为第一菲涅耳带宽度。如图 1-1 所示,实际地震数据的层析结果的等效频率小于 2 Hz。

地震偏移成像利用反射地震数据,通过成像条件(如零延迟互相关),运算正传震源子波的震源波场和反传地表记录的检波器波场,得到偏移成像。如图 1-2 所示,偏移成像的波数为震源波场和检波器波场的波数矢量和,即

$$|\mathbf{k}| = |\mathbf{k}_s + \mathbf{k}_r| = \frac{2\omega}{v} \cos \frac{\theta}{2} \quad (1-1)$$

其中, \mathbf{k}_s 、 \mathbf{k}_r 和 \mathbf{k} 分别表示震源波场、检波器波场和偏移成像的波数矢量; ω 表示地震波的角频率; v 表示速度; θ 是震源波场和检波器波场的波数矢量之间的夹角。由于 θ 的取值范围为 $0^\circ \sim 180^\circ$,因此,偏移成像的波长 $2\pi/|\mathbf{k}|$ 的值域为 $0.5\lambda \sim \infty$ ($\lambda = 2\pi v/\omega$ 为波长)。考虑到“瑞利判据”(Rayleigh criterion),偏移成像的分辨率为成像波长的一半,即为 $\frac{\lambda}{4} \sim \infty$ 。当夹角 θ 为 0 时,达到成像分辨率的极限 $\lambda/4$,

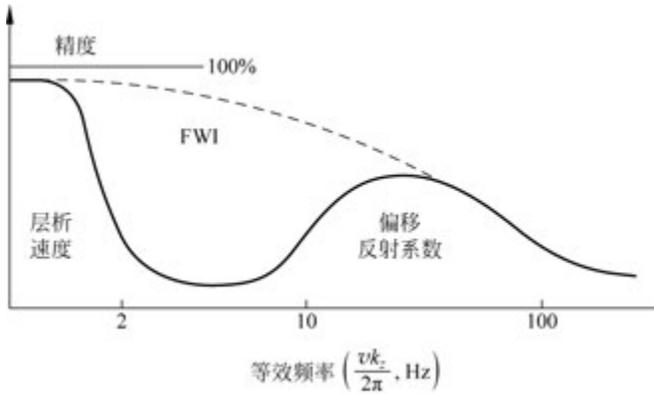


图 1-1 层析、偏移和全波形反演的精确度和分辨率区间(根据 Claerbout(1985)修改^[21])

黑色实线给出了层析和偏移结果的精确度和分辨率区间。它们的分辨率在 2~10 Hz 区间存在一个空缺；而通过利用大偏移距的折射波和低频信号，全波形反演结果的分辨率能够填补这个空缺^[22]

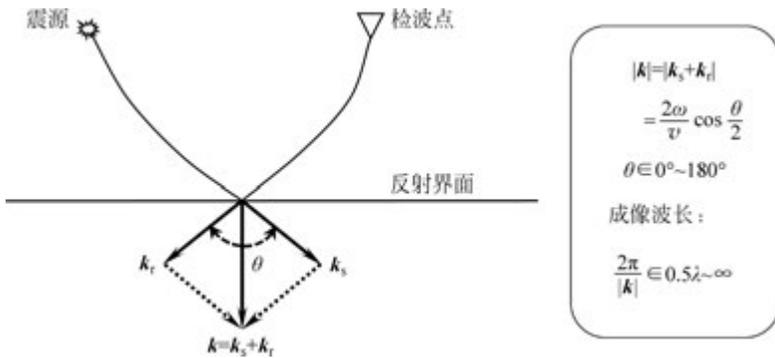


图 1-2 偏移成像的分辨率

当夹角 θ 为 180° 时，无成像分辨能力。由上面的分析可知，随着地震波入射角(在各向同性介质中为 $\theta/2$)的增加，成像分辨率逐渐降低，成像的轴向宽度变大，形成“拉伸畸变”。在实际成像处理中，为了减少这种畸变的影响，会对拉伸畸变明显的偏移距进行切除。因此，偏移成像只使用了中小偏移距的数据。另外，地震数据的频带较为有限，大多数陆地地震资料的有效频率下限为 10 Hz。因此，如图 1-1 所示，偏移成像的等效频段与层析的等效频段存在一个缺口，即 2~10 Hz。

另外，“两宽一高”(宽频、宽方位、高密度)地震采集技术的出现，为地震勘探提供了更丰富的低频和大入射角数据。因此，FWI 结合更丰富的地震数据能够填补层析和偏移的分辨率缺口，使地震成像的频谱更加完整。这是 FWI 的第一个重要意义。

2) 构建高精度的地震模型

由上述分析可知，FWI 在一定程度上填补了层析和偏移的频率空缺。然而，

这也说明了,在只有数据驱动的情况下,FWI 并没有将层析和偏移的频率范围向低频或高频拓展,即在本质上没有直接提升它们的分辨率。但由上述分析和图 1-3 可知,初至走时层析主要利用初至波,反射走时层析则主要利用中小角度的反射波,介于图 1-3(a)中两虚线间的大角度反射波、折射波以及透射回传波则不能被两种层析所利用。由于偏移拉伸,这一部分数据也不能被偏移成像有效利用。这些地震波中的大角度反射波、折射波以及透射回传波等波现象携带了模型的信息,能被波动方程模拟,而 FWI 利用波动方程进行波场模拟,因此 FWI 能够利用这些不能被层析和偏移利用的数据进行反演。这使得 FWI 系统具有比层析和偏移更多的数据约束,从而能给出更高精度的地震模型。图 1-4 展示了利用图 1-3 的数据分别进行反射层析和 FWI 的速度建模结果,以及这些速度的偏移成像。从图 1-4(b) 中可以看出,利用更多数据的 FWI 给出了更精确的速度模型。

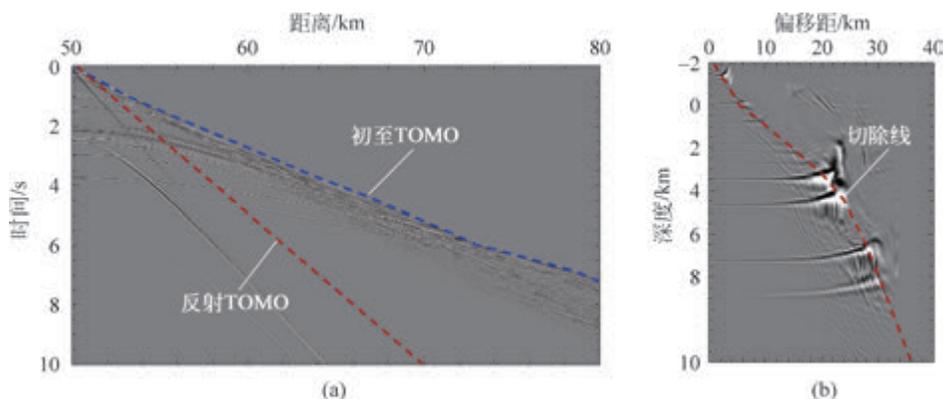


图 1-3 “盲测”数据的初至层析、反射层析使用数据信息

(a) 炮集示意图,初至层析使用初至波的走时信息,反射层析使用红色虚线以下小角度的反射波走时信息;(b) 成像道集示意图,由于大偏移距存在拉伸畸变,无法准确拾取剩余曲率(反射层析的信息来源),因此,红色曲线示意了反射层析的数据切除线

另外,FWI 能反演波动方程中包含的所有模型参数,因此 FWI 也能给出比层析和偏移成像更丰富的地震模型。这是 FWI 的第二个重要意义。

3) 缩短数据处理周期

FWI 利用波动方程模拟地震数据,地震波中的各种波现象都能被模拟,因此,FWI 能够直接利用所有能被波动方程模拟的数据。然而,反射层析和偏移成像主要使用零相位的一次反射波,因此,它们需要通过预处理消除其他数据,包括多次波、鬼波、气枪鸣震、折射波等。然而,这些数据都能被 FWI 所利用,实现“变废为宝”。这不但能提升 FWI 的稳定性,而且可以简化 FWI 的数据预处理。通常情况下,FWI 的数据预处理仅包括去噪,以消除那些不能被波动方程模拟的数据。

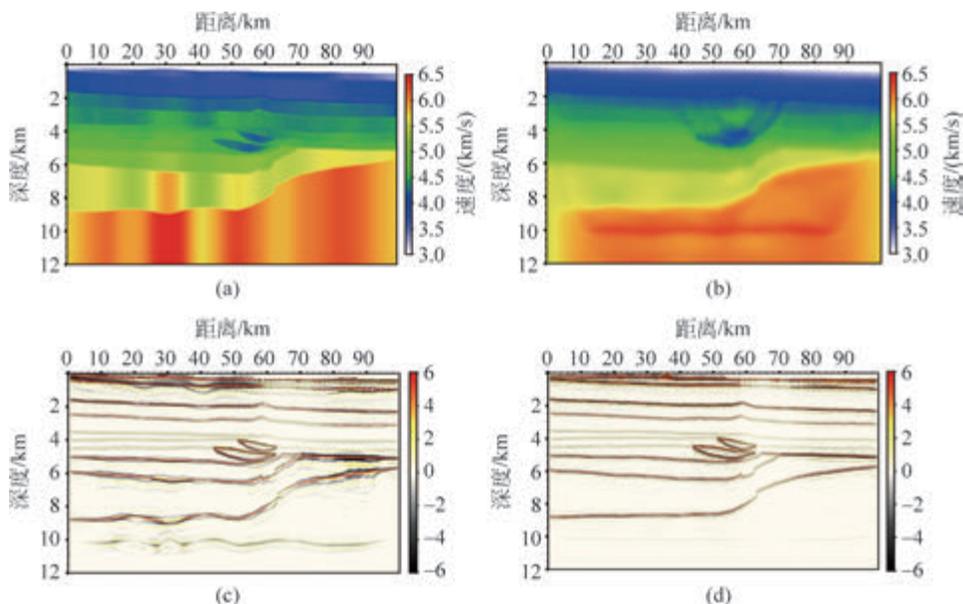


图 1-4 “盲测”数据速度建模

(a) 反射层析结果；(b) FWI 结果，反演 8 个频率^①，即 3-4-5-6-7-8-9-10 Hz，每个频率迭代 5 次，共计 40 次迭代；(c) 和 (d) 分别为使用 (a) 和 (b) 的偏移成像。由于只使用了中小偏移距的反射数据，因此反演系统信息不足，导致反演模型中存在“条竖状”假异常，也导致了 (c) 中反射轴扭曲。(a) 中的界面是沿层反射层析拾取的地层界面。而 FWI 使用了更多数据，反演系统更稳定，消除了反射层析中的“条竖状”假异常，使偏移成像的同相轴位置更准确，更符合实际构造形态

另外，常规的速度建模，包括射线层析，都存在人机交互的步骤。这使得虽然层析等建模的计算量相对 FWI 小，但是整个建模周期并不短。相比之下，FWI 是一个自动化的过程，人工干预少，因此，在算力足够的情况下，FWI 建模的周期短于常规的速度建模，如图 1-5 所示。这是 FWI 的第三个重要意义。

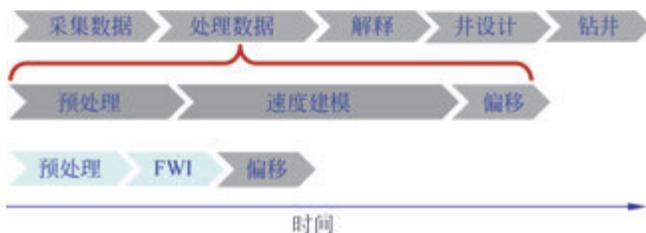


图 1-5 FWI 简化了地震数据处理流程、缩短处理周期

^① 本反演是在时间域内进行的，一个频率是指通过带通滤波提取的以该频率为中心的一个频段。如果仅为单一频率，则在时间域内会形成强烈的振荡。

1.3 全波形反演的数学原理

本节将分别采用“矩阵法”和“拉格朗日乘子法”推导 FWI 的梯度和伴随状态方程。由于使用了伴随状态方程,FWI 避免了显式计算雅可比矩阵,从而使其成为一种在计算上可行的技术。

1.3.1 基于“矩阵法”的数学推导

1. 波动方程

FWI 使用的最简单的波动方程为各向同性变密度方程,其形式如下:

$$\frac{1}{\rho v^2} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - \nabla \cdot \left(\frac{1}{\rho} \nabla p(\mathbf{x}, t) \right) = \delta(\mathbf{x} - \mathbf{x}_s) s(t) \quad (1-2)$$

其中, \mathbf{x} 表示空间位置,二维模型包含 (x, z) 两个分量而三维模型包含 (x, y, z) 三个分量; t 表示时间; p 表示压强; v 和 ρ 分别表示速度和密度,它们是空间的函数; s 表示震源子波,通过狄拉克 δ 函数加载在 \mathbf{x}_s 位置。以上波动方程考虑了密度,主要目的是更精确地模拟地震波的振幅。这一点和逆时偏移成像不同,逆时偏移成像中通常使用常密度波动方程,因为其只要求较精确地模拟出透射波和反射波的走时。

在波动方程(1-2)中,波场变量 p 与速度 v 或密度 ρ 是非线性关系,但是与震源子波 s 是线性关系。因此,对波动方程进行离散(如有限差分离散)后,波动方程可以表示成“矩阵-向量”形式的线性方程组,即

$$\mathbf{A} \mathbf{p} = \mathbf{s} \quad (1-3)$$

其中, \mathbf{p} 和 \mathbf{s} 是列向量,表示在空间和时间离散点的波场和震源; \mathbf{A} 是矩阵,表示数值离散实现的波动方程算子

$$\frac{1}{\rho v^2} \frac{\partial^2}{\partial t^2} - \nabla \cdot \left(\frac{1}{\rho} \nabla \right) \quad (1-4)$$

2. 目标函数

全波形反演的核心目的是找到一个地下模型,如式(1-2)中的速度和密度,使得通过求解波动方程产生的预测地震数据和观测地震数据之间的差异最小化。数学上,通常使用范数来衡量这种差异,最常见的是 L_2 -范数。因此,我们可以运用预测地震数据和观测地震数据之差的 L_2 -范数的平方来建立目标函数(objective function),即

$$f(\mathbf{m}) = \frac{1}{2} \|\delta \mathbf{d}\|_2^2 = \frac{1}{2} \|\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}}\|_2^2 = \frac{1}{2} \sum_{n_s} \sum_{n_r} \sum_{n_t} (d_{\text{pre}} - d_{\text{obs}})^2 \quad (1-5)$$

其中, f 表示目标函数,也称为代价函数(cost function)、误差函数或失配函数(misfit

function)或者泛函(functional); \mathbf{d}_{pre} 表示正演模拟的地震记录向量; \mathbf{d}_{obs} 表示观测的地震记录向量; $\delta\mathbf{d}$ 表示数据残差; \mathbf{m} 表示求解的模型参数,如速度或密度; n_s 、 n_r 和 n_t 分别表示地震数据集中的炮数、每炮中的地震道数和每道中的时间样点数。通过最小化目标函数 $f(\mathbf{m})$ (注意:对于有些目标函数,如基于互相关的目标函数,反演的过程是最大化目标函数),对应的模型参数 \mathbf{m} 便是全波形反演的结果。式(1-5)包含的系数 $1/2$ 是为了简化梯度表达式。在式(1-5)中,将地震数据集中所有的地震记录排成一列就构成了记录向量 \mathbf{d} ,将二维或三维模型排成一列就构成了模型参数向量 \mathbf{m} 。正演模拟的地震记录 \mathbf{d}_{pre} 为检波器从地震波场 \mathbf{p} 中抽取的波场值,这个过程可以表示为

$$\mathbf{d}_{\text{pre}} = \mathbf{D}\mathbf{p} \quad (1-6)$$

其中, \mathbf{D} 为采集(算子)矩阵。对于单点采集检波器, \mathbf{D} 矩阵的每行只有一个元素为 1,该元素对应检波器在波场 \mathbf{p} 中的位置,其余元素为 0。由于检波器只是稀疏地分布在模型的一些元素点,因此, \mathbf{D} 矩阵是一个宽矩阵(wide/fat matrix),而其转置 \mathbf{D}^T 矩阵是一个高矩阵(tall/thin matrix)。

目标函数(1-5)受波动方程(1-3)和采集方程(1-6)约束,构成了 FWI 系统。因为地震数据是模型的非线性函数,所以 FWI 实质上是一个非线性反问题,可基于一个初始模型 \mathbf{m}_0 通过迭代修正的方式,使得目标函数收敛至最小。目标函数的最小化是一种最优化计算过程,可以采用全局优化方法或局部优化方法来实现。全局优化方法(例如:模拟退火算法、粒子群算法及遗传算法等)因其具有更大的搜索空间,能够使得目标函数收敛于全局最小。全局搜索的方法涉及很多次迭代,每次迭代中都需要多次求解波动方程,因此计算量巨大,目前的计算机速度无法满足其大规模应用。与之相对,局部优化算法可以通过几十次的迭代实现目标函数的收敛,因此计算量较小,目前被广泛采用。但因其搜索解的空间范围也较小,容易收敛至局部极小而非全局最小,因此需要采用一些方法或技巧避免反演陷入局部极小。这些方法包括:地震数据分频的多尺度反演^[23]、更凸的目标函数(如,最优运输目标函数^[24-26]、自适应目标函数^[18])或一个更精确的初始模型。图 1-6 展示了一个三层模型基于式(1-5)的目标函数值,其中存在一个全局最小值和多个局部极小值。

3. 局部反演

采用局部迭代反演方案的 FWI 首先从一个假定为足够接近真实模型的初始模型 \mathbf{m}_0 出发,然后寻求对该模型进行一系列改进,逐步将目标函数减小至接近零。因此,我们需要考虑一个初始模型 \mathbf{m}_0 和一个更新模型 $\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}$ 的目标函数。

回忆一下,对于一个单变量标量函数的泰勒级数,将其截断到二阶项可以表示为

$$f(x) = f(x_0 + \delta x) = f(x_0) + \delta x \left. \frac{df}{dx} \right|_{x=x_0} + \frac{1}{2} \delta x^2 \left. \frac{d^2 f}{dx^2} \right|_{x=x_0} + O(\delta x^3) \quad (1-7)$$

同理,对于一个矢量模型 \mathbf{m} ,类似的泰勒级数展开可以表示为

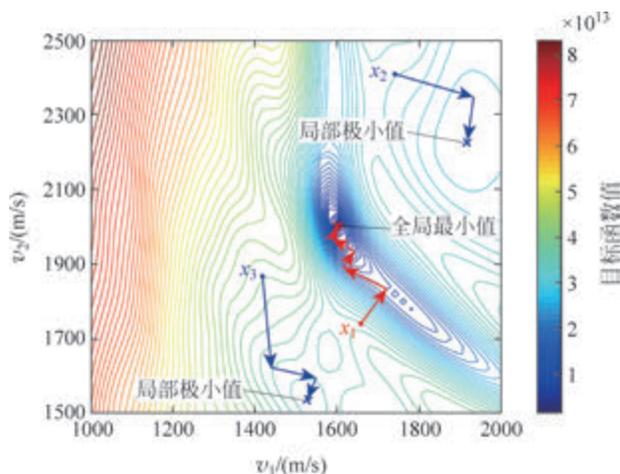


图 1-6 全波形反演目标函数示意图

计算使用了水平三层速度模型,第一层厚 200 m,第二层厚度 200 m,地表为自由表面,子波是主频为 15 Hz 的气枪震源子波,使用了 L_2 -范数目标函数。横轴为第一层的速度值,纵轴为模型第二层的速度值,第三层的速度固定为 3000 m/s,整个模型密度为 2000 kg/m^3

$$f(\mathbf{m}) = f(\mathbf{m}_0 + \delta\mathbf{m}) = f(\mathbf{m}_0) + \delta\mathbf{m}^T \left. \frac{\partial f}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0} + \frac{1}{2} \delta\mathbf{m}^T \left. \frac{\partial^2 f}{\partial \mathbf{m}^2} \right|_{\mathbf{m}=\mathbf{m}_0} \delta\mathbf{m} + O(\delta\mathbf{m}^3) \quad (1-8)$$

为了最小化目标函数 f ,可以通过求取目标函数关于 \mathbf{m} 的导数,并将导数设置为零来实现。由于 $\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}$,并且 \mathbf{m}_0 是常数,因此,目标函数关于 \mathbf{m} 的梯度等价于关于 $\delta\mathbf{m}$ 的梯度。因此,有

$$\left. \frac{\partial f(\mathbf{m})}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0+\delta\mathbf{m}} = 0 + \left. \frac{\partial f}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0} + \left. \frac{\partial^2 f}{\partial \mathbf{m}^2} \right|_{\mathbf{m}=\mathbf{m}_0} \delta\mathbf{m} + O(\delta\mathbf{m}^2) = 0 \quad (1-9)$$

忽略上式中的高阶项,可得模型的更新表达式为

$$\delta\mathbf{m} \approx - \left(\left. \frac{\partial^2 f}{\partial \mathbf{m}^2} \right|_{\mathbf{m}=\mathbf{m}_0} \right)^{-1} \left. \frac{\partial f}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0} \equiv -\mathbf{H}^{-1} \nabla_{\mathbf{m}} f \quad (1-10)$$

其中, $\mathbf{H} = \left. \frac{\partial^2 f}{\partial \mathbf{m}^2} \right|_{\mathbf{m}=\mathbf{m}_0}$ 表示海森(Hessian)矩阵; $\nabla_{\mathbf{m}} f = \left. \frac{\partial f}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0}$ 表示目标函数关于模型的梯度。需要注意的是,海森矩阵和梯度都是基于模型 \mathbf{m}_0 计算的。根据式(1-10),可得牛顿法的模型更新表达式为

$$\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m} = \mathbf{m}_0 - \mathbf{H}^{-1} \nabla_{\mathbf{m}} f \quad (1-11)$$

上式中,如果海森矩阵的逆矩阵不能精确求解,则需要对其添加步长搜索,即

$$\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m} = \mathbf{m}_0 - \alpha \mathbf{H}^{-1} \nabla_{\mathbf{m}} f \quad (1-12)$$

其中, α 为步长,可以通过线性搜索得到,本章后续小节将介绍一种步长搜索方法。

如果一个二维模型的维度为 $n \times m$,则海森矩阵的维度为 $(n \times m) \times (n \times m)$ 。

因此,海森矩阵的维度非常巨大,构建其较为困难,求其逆则更加困难。因此,在实际 FWI 应用中,最速下降(steepest-descent)法或共轭梯度(conjugate-gradient)法使用更为广泛。其中,最速下降法可以表示为

$$\mathbf{m} = \mathbf{m}_0 - \alpha \nabla_{\mathbf{m}} f \quad (1-13)$$

对比式(1-12)可以看出,最速下降法将牛顿法中的海森矩阵的逆用单位矩阵进行了代替。由于模型参数的元素个数众多,且目标函数非线性,因此,牛顿法或最速下降法都需要多次迭代,即一次更新模型后,将更新后的模型 \mathbf{m} 视为 \mathbf{m}_0 开展下一次更新计算,每一次更新就是 FWI 中的一次迭代。

4. 伴随状态法求取梯度

通过观察式(1-13)可知,在 FWI 中,求取目标函数关于模型的梯度 $\nabla_{\mathbf{m}} f$ 是最速下降法的关键。根据目标函数(1-5)可知,梯度可以表示为

$$\nabla_{\mathbf{m}} f = \frac{\partial f}{\partial \mathbf{m}} = \frac{\partial}{\partial \mathbf{m}} \left(\frac{1}{2} \delta \mathbf{d}^T \delta \mathbf{d} \right) = \left(\frac{\partial \mathbf{d}}{\partial \mathbf{m}} \right)^T \delta \mathbf{d} = \mathbf{J}^T \delta \mathbf{d} \quad (1-14)$$

其中, \mathbf{J} 为雅可比(Jacobian)矩阵。如果数据向量 \mathbf{d} 的维度为 $n \times 1$,模型 \mathbf{m} 的维度为 $m \times 1$,则雅可比矩阵的维度为 $n \times m$ 。由式(1-14)可知,因为 $\partial \mathbf{d} / \partial \mathbf{m} \approx \delta \mathbf{d} / \delta \mathbf{m}$,所以雅可比矩阵表示了模型扰动引起的数据扰动,即散射波。扰动模型的一个元素,形成的散射波组成雅可比矩阵的一列。因此,如果数值计算雅可比矩阵,则需要解 $n_s \times m + n_s$ 次波动方程,其中 n_s 表示炮数。由于地震数据和反演模型都是高维度向量,因此,雅可比矩阵的维度非常巨大,直接构建的成本也很高。而在射线层析中,由于其雅可比矩阵维度较小,且计算代价小,因此可以显式构建雅可比矩阵进行模型更新。由于雅可比矩阵庞大的维度和求解波动方程的高计算量,因此在 FWI 中,显式构建雅可比矩阵和构建海森矩阵一样几乎是不可能实现的。为了计算梯度,Pratt^[2]巧妙地运用“矩阵法”实现了“伴随状态法”求取 FWI 的梯度。其推导过程如下:

对波动方程(1-3)两边求取关于模型 \mathbf{m} 的偏导数,可得

$$\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{p} + \mathbf{A} \frac{\partial \mathbf{p}}{\partial \mathbf{m}} = \frac{\partial \mathbf{s}}{\partial \mathbf{m}} = 0 \quad (1-15)$$

在式(1-15)中,考虑到震源子波不是模型 \mathbf{m} 的函数,因此,其关于模型的偏导数等于零。重新排列式(1-15),有

$$\frac{\partial \mathbf{p}}{\partial \mathbf{m}} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{p} \quad (1-16)$$

考虑到数据采集方程(1-6),有

$$\mathbf{J} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}} = \mathbf{D} \frac{\partial \mathbf{p}}{\partial \mathbf{m}} = -\mathbf{D} \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{p} \quad (1-17)$$

将式(1-17)代入式(1-14),可得

$$\nabla_{\mathbf{m}} f = \mathbf{J}^T \delta \mathbf{d} = \left(-\mathbf{D} \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{p} \right)^T \delta \mathbf{d} = -\mathbf{p}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \mathbf{A}^{-T} \mathbf{D}^T \delta \mathbf{d} \quad (1-18)$$