

СЛОЖНОСТНО–СТРУКТУРНЫЙ ПОДХОД К ИССЛЕДОВАНИЮ ОБЛАСТИ ПРИМЕНИМОСТИ АЛГОРИТМА PROFIT

Татьяна Ступина, Иван Кулаков

Аннотация: В работе представлена формальная постановка сложностно - структурного подхода к исследованию области применимости алгоритма PROFIT (PROfile Forward and Inverse Tomographic modeling). Основная идея состоит в определении кластера моделей среды (разреза геологического строения земной коры), представленных изображениями, которые с допустимой погрешностью восстанавливаются алгоритмом. Метрика в пространстве изображений задается специальным образом. Она учитывает изменение поля скоростей и аномальные включения, количественный и качественный состав которых отражает этапы усложнения моделей сред.

Ключевые слова: сейсмическое профилирование, томография, прямое моделирование, синтетическая модель.

ACM Classification: G1.10. Mathematics of Computing - Applications

Conference: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.

Введение

Подход к обработке экспериментальных данных зависит от специфики решаемой задачи в конкретной прикладной области и конечной цели, которая ставится в задаче. В различных областях знаний, целью которых является обнаружение причинно-следственных связей какого-либо процесса или явления, необходимо решать обратную задачу, которая, как правило, является неустойчивой как относительно входных данных, так и самого оператора инверсии. Для оценки качества полученного решения также необходимо уметь решать и прямую задачу. На каждом этапе возникают свои трудности, с которыми необходимо справляться, используя, предложенные к обработке данные и основываясь на выбранной модели. Оценить же качество выбранной модели независимо от качества обрабатываемых данных возможно лишь теоретически при известной «истинной» модели [Вапник, 1984].

В данной работе рассматривается алгоритм PROFIT прямого и обратного моделирования профильных данных в сейсмической томографии [Кулаков, 2007]. Преимущества данного алгоритма при решении задач на реальных и синтетических данных подробно отражены в работах И.Ю. Кулакова. В работе представлены основные этапы этого алгоритма, подчеркивающие многопараметрическую сложность и специфику решаемой задачи. Важной компонентой предложенного алгоритма является возможность построения реконструированной модели геологической среды, т.е. глубинного разреза, что является достаточно интересным и важным для задач геологической интерпретации в целях изучения строения земной коры и поиска полезных ископаемых. Для того чтобы оценить множество правдоподобных, относительно экспертных знаний о моделях сред, реконструкций предлагается сложностно - структурный подход.

В работе предложена формальная постановка сложностно - структурного подхода к исследованию области применимости алгоритма PROFIT. Основная идея состоит в определении кластера моделей среды, представленных изображениями глубинных разрезов, которые с допустимой погрешностью восстанавливаются алгоритмом. Элементы кластера формируются в метрическом пространстве

изображений относительно некоторого экспертного шаблона (модель, допустимая по смыслу с точки зрения строения земной коры в данном регионе). Другими словами, задается критерий, определяющий допустимую реконструированную модель среды, построенную алгоритмом PROFIT для данного географического региона и фиксированной системы наблюдений.

Основные понятия

В общем виде задача сейсмической томографии [Нолет, 1990] заключается в определении поля скоростей $v(r)$ по множеству измерений времени на поверхности

$$T_i = \int_{S_i} \frac{ds}{v(r)}, \quad i = 1, \dots, N.$$

Задача является достаточно сложной, поскольку неизвестная функция $v(r)$ в неявном виде присутствует в определении лучевой траектории S_i . Поэтому задача решается в линеаризованной постановке применением принципа Ферма. Обозначим время, предсказываемое начальной (референтной, опорной) моделью

$$T_i^\circ = \int_{S_i^\circ} \frac{ds}{v_\circ(r)}. \quad (1)$$

Где S_i° - траектория луча в начальной модели, тогда время задержки определяется следующим образом

$$\begin{aligned} \delta T_i = T_i - T_i^\circ &= \int_{S_i} \frac{ds}{v} - \int_{S_i^\circ} \frac{ds}{v_\circ} \approx \int_{S_i^\circ} \left(\frac{1}{v} - \frac{1}{v_\circ} \right) ds, \\ \delta T_i &= - \int_{S_i^\circ} \frac{\delta v(r)}{v_\circ^2(r)} ds. \end{aligned} \quad (2)$$

Где $\delta v(r) = v(r) - v_\circ(r)$. Здесь была произведена замена неизвестной лучевой траектории на траекторию для референтной модели, поэтому в прямую задачу внесена поправка второго порядка. Обратная же задача может быть плохо обусловлена и нет уверенности, что ошибка второго порядка не приведёт к большой ошибке в решении. На практике очень часто система наблюдений не отвечает идеальным условиям, когда невозможно обеспечить достаточную лучевую плотность, поэтому обратная задача решается численными методами. С этой целью определённым образом задается параметризация по базису $h_j(r)$

$$\delta v(r) = \sum_{j=1}^M \gamma_j h_j(r).$$

Тогда уравнение временных задержек переписывается в следующем виде

$$\delta T_i = \sum_{j=1}^M \left(- \int_{S_i^\circ} \frac{1}{v_\circ^2(r)} \gamma_j h_j(r) ds \right) = \sum_j A_{ij} \gamma_j.$$

Где A_{ij} - матрица первых производных по параметрам, $A_{ij} = \frac{\partial T_i}{\partial \gamma_j} = - \int_{S_i^\circ} \frac{h_j(r)}{v_\circ^2(r)} ds$.

Таким образом, задача сводится к решению СЛАУ

$$Ay = \delta T. \quad (3)$$

В общем случае такая система является плохо обусловленной, поэтому при её решении применяются методы регуляризации и итеративные подходы обращения сильно разреженных матриц большого размера. В алгоритме PROFIT реализованы одни из таких методов, позволяющие за достаточно приемлемое время получать хорошие решения.

Принцип работы алгоритма PROFIT

Программа PROFIT состоит из двух основных компонент. Первая содержит алгоритмы томографической инверсии и может быть использована и в качестве отдельного и независимого программного кода для обработки сейсмических рефрагированных данных. Вторая компонента подключает к исполнению совместно прямое и обратное моделирование и предназначена для построения реконструированной модели среды. Код программы реализован на языке Фортран-90 и разработан под операционной системе Windows. Исходные и исполняемые коды программы находятся в свободном доступе.

Основная структура алгоритма томографической инверсии в программе основывается на общепризнанных вычислительных методах компьютерной томографии [Нолет, 1990]. Вычисления производятся итеративно в соответствии со следующими шагами:

1. Трассировка лучей в 2 D скоростной модели (стартовая модель для первой итерации или пересчитанная скоростная модель после предыдущих итераций);
2. Построение сеточной параметризации (только для первой итерации);
3. Вычисление матрицы и обращение;
4. Обновление скоростной модели. Переход на первый шаг.

Более подробно опишем наиболее важные моменты для каждого из этих шагов.

Лучевое трассирование. Лучевое трассирование, применяемое в представляемом алгоритме, основывается на принципе Ферма и состоит в нахождении пути, обеспечивающим минимальное время между источником и приёмником. Эта идея является основой для расчета изгиба лучевого трассирования, которая широко применялась в последние десятилетия и внедрялась как основа в различных алгоритмах томографии локальных землетрясений, так и сейсмического моделирования.

В представленном алгоритме реализована альтернативная версия алгоритма расчёта изгиба, которая схематично показана на рис. 1. Поиск пути с минимальным временем состоит из последовательно выполняемых нескольких шагов вычисления изгиба луча B , первоначальная траектория которого описывается тригонометрической функцией

На первом шаге (Фигура А) точка на морском дне (точка b) располагается сразу под источником S . Мы начинаем с прямой линии между точкой b и приёмником r , потом последовательно изменяем её до получения минимального времени прохода луча. В первом приближении отклонение $A(s)$ относительно начального прямого пути вычисляется в соответствии со следующей формулой:

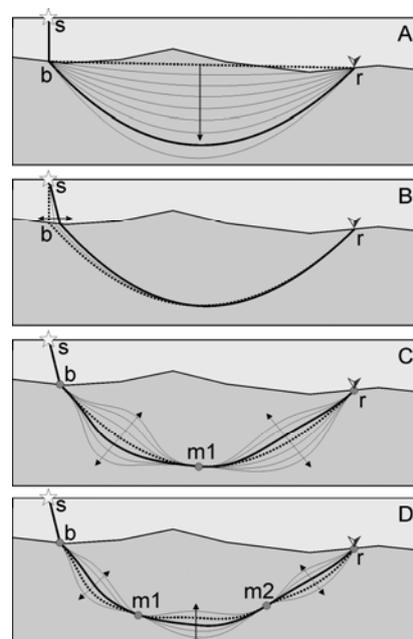


Рис. 1. Иллюстрация лучевого трассирования

$$A(s) = B \cos\left(\pi \frac{s - D/2}{D}\right), \quad (4)$$

где B – вычисленное значение изгиба, s – расстояние вдоль начального пути, D – длина начального пути. Значение параметра B настраивается таким образом, чтобы кривая S_i обеспечивала минимальное значение интеграла (1), т.е. $B^* = \arg \min_B T_i(B)$, где $v(s)$ – распределение скорости вдоль луча S_i .

На втором шаге (Фигура В), мы передвигаем точку b до получения минимального значения интеграла (1). В случае наземных наблюдений этот шаг пропускается, поскольку точки S и b имеют одинаковые координаты.

На третьем шаге дальнейшее отклонение от пути между b и r вычисляется итеративно с использованием следующей формулы:

$$A(s) = \frac{B}{2} \cos\left(2\pi \frac{s - (D_2 - D_1)/2}{(D_2 - D_1)}\right) + \frac{1}{2} \quad (5)$$

где D_1 и D_2 соответствуют длинам вдоль путей начала и конца текущего сегмента.

На первой итерации вычисление значения изгиба производится для всего сегмента $b-r$, представленном на Фигуре А, по формуле (4). На второй итерации (Фигура 1) путь делится на два равных сегмента с длинами $b - m_1$ и m_1-r соответственно и для каждого из них пересчитывается длина изгиба в соответствии с формулой (5). После определения минимального времени, весь путь делится на три части (Фигура D), и пересчет значений изгиба осуществляется для трёх сегментов $b-m_1$, m_1-m_2 , и m_2-r . Последовательно эта процедура повторяется для четырёх, пяти, шести и более частей. Процедура заканчивается, когда длина сегмента становится меньше заданной величины.

Параметризация. В программе PROFIT 2D распределение скоростей задается по аналогии с ранее разработанным для 3D томографической инверсии подходом и реализованным в программе LOTOS-07 [Кулаков, 2007]. Скоростные аномалии (неоднородности) интерполируются билинейно между фиксированными узлами сетки $\{h_j(r)\}_{j=1,M}$, $r = (x, z)$, которые задаются на множестве вертикальных линий с фиксированным шагом Δx вдоль профиля. Вдоль каждой линии по направлению в глубину (по переменной z) вычисляется значение плотности лучей $\rho(S^j)$ как сумма длин лучей в единичном объёме. И тогда узлы j , $j = 1, \dots, M$, распределяются в соответствии с лучевой плотностью. В вертикальном направлении, во избежание чрезмерных узловых флуктуаций, задаётся минимальный шаг Δz между узлами. В областях с пониженной лучевой плотностью расстояние между узлами наибольшее. Узлы не устанавливаются в областях с лучевой плотностью меньше заданного значения (например, 0.1 от среднего). Хотелось бы отметить, узловые шаги в вертикальном и горизонтальном направлении не равны между собой, поскольку мы полагаем различное разрешение по вертикали и горизонтали. Узлы сетки устанавливаются только в первой итерации в соответствии с распределением лучей, протрассированных в стартовой модели. На последующих итерациях скоростные изменения Δv_j пересчитываются в этих же узлах.

При построении узлового распределения по лучевой плотности сеть можно определённым образом адаптировать в соответствии с некоторыми особенностями данных и пересчитать для выделенных изменений в лучевой плотности. Распределение лучей и плотность являются индивидуальными

параметрами для каждой области рефрагированных данных, зависящие не только от инструментальной настройки, но и от композиции и геометрии геологической поверхности и среды.

Отметим также, что хотя малый размер (порядок) сетки обеспечивает обнаружение скоростных аномалий, расположенных в изучаемой области в тоже время это приводит к увеличению размерности матрицы обращения.

Вычисление матрицы и обращение. Матрица первых производных вычисляется на основе лучевых траекторий S_i , которые определяются с помощью лучевого трассирования по 2D модели. Каждый элемент A_{ij} матрицы A в уравнении (3) равен временному изменению Δt_i вдоль S_i луча за единичное скоростное изменение Δv_j в j -ом узле параметризации. Элементы матрицы определяются численно.

Обращение матрицы A осуществляется с помощью итеративного LSQR алгоритма [Слуис, Ворст, 1987], который достаточно быстро сходится при хорошо подобранных параметрах регуляризации Am (амплитуда демпфирования) и Sm (весовой параметр сглаживания скоростных вариаций в соседних узлах параметризационной сетки). Итерационно решаем систему уравнений:

$$\begin{pmatrix} \mathbf{A} \\ am \mathbf{I} \\ sm \mathbf{C} \end{pmatrix} d\gamma = \begin{pmatrix} dT \\ 0 \\ 0 \end{pmatrix}. \quad (6)$$

Где \mathbf{I} – диагональная единичная матрица размера M , \mathbf{C} - прямоугольная матрица, каждая линия которой состоит из двух единичных элементов противоположного знака, являющихся соседними узлами в параметризационной сетке.

Оптимальные значения этих параметров зависят от многих факторов. Например, при увеличении размера данных параметр демпфирования должен увеличиваться, в то время как при увеличении числа узлов вследствие измельчения разбиения параметр демпфирования должен быть уменьшен. В случае же большого уровня шума в данных параметр демпфирования должен быть достаточно весомым для сохранения устойчивости решения. Процедура нахождения коэффициентов демпфирования до сих пор не формализована. К тому же соотношение между числом параметров, лучей и значениями коэффициентов амплитуд и сглаживания является нелинейным. Например, при удвоении числа лучей соответствующая амплитуда решения получается увеличением коэффициента демпфирования до 1.2. Для каждой сети наблюдений эти значения определяются индивидуально. В качестве основного критерием при нахождении весовых коэффициентов демпфирования используется оценка среднеквадратического среднего (RMS) на каждой итерации. Но он не является абсолютно объективным, поскольку не обеспечивает устойчивости решения. Альтернативным подходом, реализованным в программе PROFIT, к определению оптимальных значений параметров демпфирования является синтетическое моделирование на тестовых примерах с аномалиями фиксированного размера. Например, тесты с шахматной доской.

Скоростные аномалии, полученные после инверсии, вычисляются по регулярной сети и добавляются в скоростную модель, полученную на предыдущей итерации.

Формализация подхода к исследованию области применимости алгоритма

Для того чтобы оценить качество решения, получаемое алгоритмом, часто прибегают к тестированию на синтетических, тестовых или реальных данных. Выше представленный алгоритм имеет в своём коде компоненты генерации синтетических данных по моделям сред (решение прямой задачи). Обладая такой

инструментом, появляется возможность формально описать класс задач (множество геологических разрезов земной коры) хорошо решаемых данным алгоритмом.

Обозначим через $G = \langle (L^d, v_o^d, \kappa) \rangle$, $d \in N$, графическое цветное изображение поля скоростей модели среды с описаниями: L^d - d -ый «выпуклый» контур, v_o^d - скорость среды внутри контура, κ - количество связанных выпуклых контуров. Связанными будем называть контуры, имеющие общую границу. Пусть G^* - изображение синтетической модели среды, G^m - реконструированное изображение среды. Реконструированное изображение строится пользователем по томографическому изображению поля скоростей, являющегося решением алгоритма. При подборе параметров регуляризации и минимизации значения RMS, естественно предполагается, что синтетическая модель является неизвестной.

Зададим множество $\tilde{G}(G^*)$ допустимых моделей G^m относительно синтетической G^* следующим образом

$$\tilde{G}(G^*) = \{G^m : g(G^m, G^*) \leq \delta^*, \|T - T^o\| \leq \varepsilon\}, \quad (7)$$

Где $g(G^m, G^*)$ - «метрическая» разность изображений, построенная на основе симметрической разности топологий, ограниченных контурами L^d изображений и учитывающая допустимый диапазон изменения скорости Δv_o^d внутри контура; $\|T - T^o\|$ - RMS оценка качества решения при построении томографического изображения алгоритмом.

Определим множество (кластер) \mathfrak{S} как область применимости алгоритма следующим образом

$$\mathfrak{S} = \{G^* : \exists G \ g(G, G^*) \leq \delta, \|T - T^o\| \leq \varepsilon\}. \quad (8)$$

Если мощность этого множества достаточно большая, то можно говорить о большой области применимости алгоритма. Каждый элемент этого множества представляет собой существенно отличающиеся друг от друга изображения сред, например зоны субдукции и соляные купола.

Для того чтобы ранжировать изображения формализуем понятие сложности геологического изображения и зададим её парой $\mu = (\mu_1, \mu_2)$, $\mu_1 = d$, $\mu_2 = 1 - \min_d V(L^d)$, $V(L^d)$ - нормированный объём

(площадь) топологического множества, ограниченного контуром L^d .

В качестве примера на рис. 2,3,4 приведем изображения синтетической модели соляных куполов, являющиеся результатами работы алгоритма PROFIT.

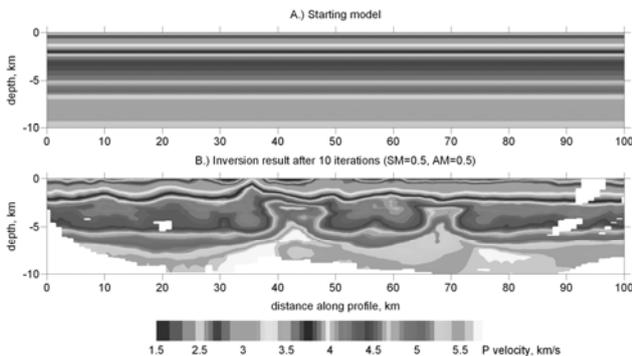


Рис. 2. Фигура В является томографическим изображением, представляющим результат инверсии с оптимально подобранными параметрами по временам, полученным по синтетической модели.

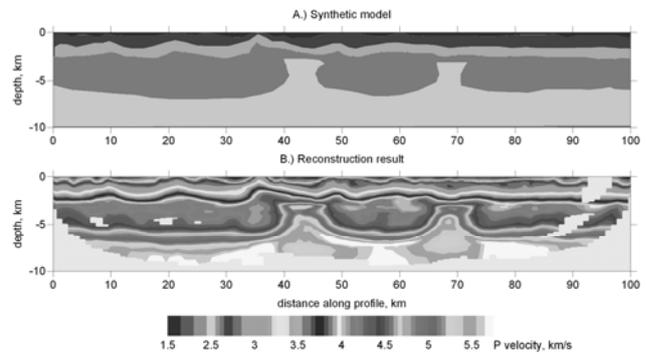


Рис. 3. Фигура В является томографическим изображением, представляющим результат инверсии по временам реконструированного изображения фигуры А.

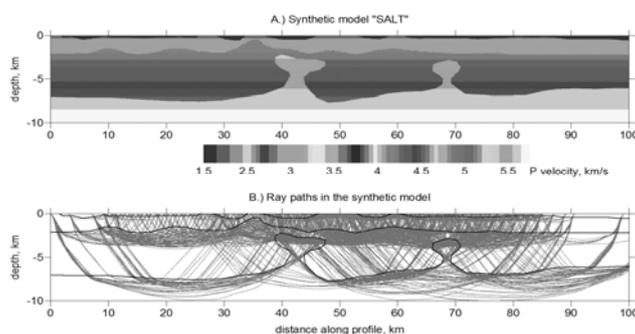


Рис. 4. Фигура А является синтетической моделью. Фигура В демонстрирует траектории лучевого трассирования.

Заключение

В работе предложена формальная постановка сложно-структурного подхода к исследованию области применимости алгоритма PROFIT. Это позволяет количественно оценить возможности алгоритма относительно числа («мощность» множества \mathfrak{S}) и сложности (μ) реконструируемых геологических изображений. Отметим, что к настоящему моменту при интерпретации результатов томографической инверсии в основном используется экспертная оценка «качества» получаемого изображения. Нами предпринята попытка некоторой формализации оценки. Более того такой подход может быть использован и при сравнении томографических изображений с некоторыми поправками на размытые границы контура. В заключении также отметим, что несмотря на то, что при решении прямой задачи и вычислении времен по синтетической модели применялся один метод лучевого трассирования, построение лучей происходит абсолютно независимо. В будущем в целях наиболее объективного оценивания качества планируется оценивать качество алгоритма по синтетическим моделям с временами, полученными другими методами.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

- [Вапник, 1984] В.Н. Вапник. Алгоритмы и программы восстановления зависимостей. Москва: Наука, 1984.
- [Кулаков, 2007] И.Ю. Кулаков. Геодинамические процессы в коре и верхней мантии земли по результатам региональной и локальной сейсмотомографии. Диссертация на соискание ученой степени доктора геолого-минералогических наук. ИГМ СО РАН, Новосибирск, 2007.
- [Нолет, 1990] Г. Нолет. Сейсмическая томография. Москва: Мир, 1990.
- [Слуис, Ворст, 1987] А. ван дер Слуис, Х. А. ван дер Ворст. Численное решение больших разреженных линейных алгебраических систем, возникающих в задачах томографии. В монографии [Г. Нолет, 1990].

Информация об авторах

Татьяна Ступина – Институт нефтегазовой геологии и геофизики СО РАН, проспект Коптюга, 3, Новосибирск, 630090, Россия; e-mail: stupinata@ipgg.nsc.ru

Иван Кулаков – Институт нефтегазовой геологии и геофизики СО РАН, проспект Коптюга, 3, Новосибирск, 630090, Россия; e-mail: KoulakovIY@ipgg.nsc.ru