

АЛГОРИТМ РАСПОЗНАВАНИЯ ПОВЕРХНОСТЕЙ МАКСИМАЛЬНОГО ГРАДИЕНТА ДЛЯ ТРЕХМЕРНЫХ МОДЕЛЕЙ РАСПРЕДЕЛЕНИЯ ФИЗИЧЕСКИХ ПАРАМЕТРОВ

В настоящее время при обработке и интерпретации геофизических данных удается получить трехмерные пространственные модели распределения физических параметров [1]. Распознавание поверхностей максимального градиента для полученных трехмерных моделей позволяет выделить границы однородных областей и осуществить построение структурной трехмерной модели изучаемой территории.

Представленная работа посвящена рассмотрению проблемы разработки алгоритма распознавания поверхностей максимального градиента вектора гравитации.

При разработке алгоритма распознавания и его программной реализации предъявлялись следующие сформулированные требования:

- высокая эффективность распознавания;
- оптимизация времени работы алгоритма;
- оптимизация программной реализации по времени выполнения и сложности реализации.

Эффективность распознавания в данном случае определяется традиционно – алгоритм должен по возможности надежно определять существующие поверхности максимума градиента в трехмерной модели распределения физических параметров и минимизировать возможность определения ложных поверхностей.

Требование к оптимизации времени работы алгоритма и его программной реализации обусловлено тем, что предполагаемый для обработки объем данных может достигать значительных размеров (порядка миллиона точек измерений).

Сложность программной реализации алгоритма обусловлена в основном двумя причинами.

Во-первых, большие объемы данных накладывают существенные ограничения на объем рабочих структур данных программного комплекса. Неудачный выбор этих структур может в худшем случае привести к переполнению оперативной памяти. Но даже если этого не происходит, то большой объем рабочих структур может вызвать постоянное

свопирование оперативной памяти, что замедляет работу программы в десятки и сотни раз.

Во-вторых, реализация сложного алгоритма приводит к большим временным затратам процесса отладки. Это обусловлено существованием большого количества различных частных ситуаций, определяемое конкретными обрабатываемыми данными. В неизбежном случае программных ошибок очень трудно и времяземко выделение этих ситуаций с целью обнаружения и устранения неточностей алгоритма и ошибок программной реализации.

Соглашения

Для описания алгоритма определим ряд соглашений по постановке решаемой задачи используемой терминологии.

Пространство, в котором находятся узлы сетки, определено в прямоугольной декартовой системе координат (ПДСК). Значения исследуемого поля заданы в узлах прямоугольной сетки.

Термин искомая поверхность определим как множество узлов, удовлетворяющих следующим условиям:

- количество узлов множества не меньше наперед заданного значения (критерий величины поверхности);
- множество узлов связано, т.е. каждый узел имеет хотя бы одного непосредственного соседа в сетке, принадлежащего данному множеству (критерий связанности);
- для каждого узла множества, если он имеет более чем одного соседа, значение модуля ВГГ лежит в пределах диапазона, определяемого наперед заданным количеством среднеквадратичных отклонений величины модуля ВГГ, вычисленным по значениям соседних узлов (критерий доверительности);
- для каждого узла множества пространственный угол между соседними узлами не превышает наперед заданного значения (критерий пространственного угла).

Цель алгоритма

Целью алгоритма является нахождение всех поверхностей, удовлетворяющих вышеприведенным критериям. Решением поставленной задачи является набор множеств узлов, принадлежащих к распознанным поверхностям.

Анализ возможных путей решения

При решении сформулированной выше задачи рассмотрено несколько подходов. Назовем их условно: подходы сечений, кластерный и векторный.

В двух последних подходах задача решается в два этапа. На первом из полного множества узлов выделяется по некоторым критериям подмножество (или набор подмножеств) узлов, в дальнейшем считаемые узлами, которые могут принадлежать искомым поверхностям. Отброшенные узлы не рассматриваются на этапе построения поверхностей. Данный принцип обеспечивает резкое сокращение участвующих в анализе узлов, что обеспечивает существенное (в десятки и сотни раз) повышение скорости получения результата.

Подход, основывающийся на сечениях

Данный подход наиболее естественный и, как показывает практика обсуждения со специалистами, наиболее часто предлагаемый при первоначальном анализе проблемы. Алгоритм восстановления пространственных поверхностей в этом случае основывается на сравнительном анализе сечений исследуемой области несколькими плоскостями. Наиболее очевидным и просто реализуемым при таком подходе является сечение исследуемого пространства с помощью рабочих плоскостей параллельных плоскостям ПДСК. На каждой из осей ПДСК выбираются проекции узлов на данную ось и через точки проекций строятся рабочие плоскости, перпендикулярные данной оси. Т.е. по сути, строится множество рабочих плоскостей, представляющих собой все возможные сечения, проходящие через узлы сетки и перпендикулярные одной из осей ПДСК.

В каждый узел рабочей плоскости заносится характеристика поля. Здесь возможны два варианта. Первый – на каждой из рабочих плоскостей для каждого узла сетки, лежащего в рабочей плоскости, определяется величина проекция вектора поля на данную плоскость. Второй – на каждой из рабочих плоскостей для каждого узла сетки, лежащего в рабочей плоскости, определяется модуль вектора поля. И в первом и во втором варианте оперирование в дальнейшем производится с определенным соответствующим образом значением поля.

При таком подходе решения поставленной задачи сводится к необходимости определения корреляции между множеством сечений. При очевидной простоте описанного этапа данного подхода получение результата в этом случае не является простой проблемой и требуется большой объем исследовательской работы. Простой иллюстрацией возникающих сложностей является пример, в котором существует максимум в виде хребта, ориентированного по направлению одной из осей координат.

Кластерный подход

Данный подход основывался на кластеризации величины модуля ВГГ. На первом этапе проводилась кластеризация модуля ВГГ на некоторое количество задаваемых кластеров. Объектами кластеризации служат значения модуля ВГГ. Мера близости определяется, например, как абсолютная величина разности значения модуля ВГГ и среднего значения кластера. Несомненное достоинство данного подхода в том, что на этапе кластеризации достаточно быстро формируются группы элементов (кластеры) объединенные по основному для решаемой задачи признаку – величина модуля ВГГ. По завершению этапа кластеризации восстановление поверхностей производится итерационно.

На первом шаге итерации выбираются узлы, принадлежащие к кластеру с наибольшим средним значением. Из узлов данного множества строятся возможные поверхности.

На последующих этапах в качестве обрабатываемого кластера выбирается из множества необработанных кластеров кластер с наибольшим средним значением модуля ВГГ. Из множества узлов текущего кластера выбираются узлы, примыкающие к найденным ранее поверхностям, и добавляются в поверхности. Из оставшихся узлов кластера, если таковые существуют, делается попытка построения новых поверхностей.

Это общая схема подхода, так как, естественно, параллельно ведется контроль «толщины» поверхности, отслеживание выхода поверхности за границы заданной сеткой области и многое другое, что будет обсуждено ниже.

Несомненное достоинство данного подход – достаточно хорошее время распознавания за счет того, то анализ начинается с узлов, наиболее вероятно входящих в искомые поверхности. И в каждой последующей итерации участвуют необработанные узлы наиболее вероятно входящие в поверхности.

Практическая реализация данного подхода подтвердила его достоинства. Но выявила существенный недостаток. Не удалось выработать эффективного критерия

окончания процесса итерации. А это приводит к тому, что найденные поверхности начинают с какого-то момента объединяться в одну. Вроде бы очевидный на первый взгляд критерий окончания – наперед заданное допустимое пороговое значение модуля ВГГ не приемлем. Например, имеется две поверхности с большим значением модуля ВГГ и одна с малым. И минимальное значение модуля ВГГ в двух первых больше, чем максимальное в третьей поверхности. Тогда если установить порог обрезания выше, чем максимальное значение третьей, то возможно добиться разделения двух первых поверхностей, но третья будет безвозвратно потеряна. Установка порога обрезания ниже максимального значения третьей поверхности позволит ее обнаружить, но может привести к слиянию двух первых.

Разумного компромисса между этими противоречивыми условиями найти не удалось.

Векторный подход

Векторный подход, как отмечалось выше, состоит из нескольких этапов. Назовем их соответственно: *отсечка фоновых значений узлов, нахождение базовых опорных узлов, определение дополнительных опорных узлов, построение поверхностей, заполнения «дыр», удаление маловероятных поверхностей.*

На первом этапе – **отсечка фоновых значений узлов**, производится отбор из полного множества узлов подмножества узлов, у которых значение модуля ВГГ превышает наперед заданное значение. Выбранное подмножество узлов является исходными данными для всех последующих этапов обработки за исключением этапа заполнения «дыр», о чем будет сказано ниже. Такой отбор естественен, т.к. существуют ошибки измерения величины модуля вектора градиента, и рассматривать значения меньшие и сравнимые с величиной этой ошибки – операция, не имеющая смысла.

Одна из основных оригинальных идей описываемого подхода лежит в основе этапа **нахождения базовых опорных узлов.**

Введем понятие *направляющая кривая вектора градиента гравитации*, определяемое как кривая в пространстве в каждой точке которой направление вектора градиента гравитации либо совпадает с ней, либо является касательной к этой кривой.

Представим для начала идеальную модель исследуемой области предполагающую, что известен вектор градиента гравитации в каждой точке пространства и все измерения величин сделаны абсолютно точно и не имеют ошибки измерения. Построим направляющую кривую вектора градиента гравитации, проходящую через произвольно

выбранную точку внутри рассматриваемого объема. Двигаясь от одного конца кривой к другому, построим зависимость изменения модуля ВГГ от пройденного расстояния. Теперь, сведя проблему от трехмерного случая к двумерному, по полученной зависимости не составляет труда решить традиционную задачу нахождения локальных экстремумов значения модуля ВГГ. Отметим в пространстве соответствующие точки найденных локальных максимумов. Если построить все возможные кривые и отметить на них найденные локальные максимумы, то отмеченные точки в пространстве составят искомые поверхности.

В случае реальных измерений, естественно, ни одно из условий идеальной модели не выполняется. К основным отличиям реальной модели от идеальной относятся:

- дискретность определение поля – значение поля определено только в узлах сетки;
- наличие ошибки измерения величины модуля вектора градиента гравитации;
- наличие ошибок измерения координат вектора градиента гравитации.

Отсюда в реальном случае построение направляющих кривых вектора градиента гравитации возможно только опираясь на существующие значения вектора градиента гравитации в узлах сетки. По сути, реальная направляющая кривая ВГГ аппроксимируется ломанной прямой, проходящей через узлы сетки. И количество возможных для построения направляющих кривых ВГГ становится не бесконечным, как в идеальном случае, а ограничено сверху теоретически количеством узлов сетки. На практике данное количество существенно меньше указанного теоретического лимита, т.к. отдельная направляющая кривая объединяет некоторую совокупность узлов сетки. Вторым существенным отличием реального случая от идеального при построении направляющих кривых ВГГ является условие начала и окончания кривой. В реальном случае условием начала и конца кривой служат не только геометрические границы исследуемой области, как в идеальной модели, но и узлы, значения модуля градиента гравитации в которых ниже значения определенного на этапе отсечки фоновых значений узлов.

Кроме того, в результате дискретности измерений и наличия ошибок измерений возможны случаи вырождение аппроксимирующей ломанной вектора градиента гравитации в замкнутую линию. В этом случае, началом ее может считаться любой узел, а концом – узел, предшествующий начальному.

Найденные локальные максимумы по построенным направляющим кривым составляют множество опорных узлов.

В результате описанных причин построенные в реальном случае направляющие кривые вектора градиента гравитации с некоторой вероятностью отображают реальное поведение направляющих кривых. Чем выше точность проведенных измерений, а, главное, чем меньше шаг сетки, тем вероятность совпадения выше. Однако в целом, в случае достаточно хорошей дискретной аппроксимации исследуемого пространства, картина поведения идеальных и реальных кривых совпадает. И данные аппроксимирующие ломанные являются основой для нахождения локальных максимумов величины модуля ВГГ. Найденные по всем аппроксимирующим направляющим ВГГ локальные максимумы составляют множество первичных опорных узлов.

Но в реальном случае при нахождении локальных максимумов возникает вероятность появления ошибок первого и второго рода. Для того чтобы минимизировать количество ошибок первого рода, на этапе **определения дополнительных опорных узлов** производится расширение множества опорных узлов. Предлагается две стратегии этого процесса. Первая состоит в том, что для каждого узла множества первичных опорных узлов производится поиск узлов, имеющих большее значение модуля ВГГ, среди непосредственно окружающих его узлов. Если таковые существуют, то они включаются во множество опорных узлов. Вторая стратегия менее очевидна. Она состоит в том, что во множество опорных узлов включаются все узлы непосредственно окружающие узлы, принадлежащие множеству первичных опорных узлов. Но, учитывая существование ошибок измерения величины модуля ВГГ, становится очевидным, что и данная стратегия имеет право на существование и реально приводит к уменьшению вероятности ошибок первого рода.

Результатом описанных этапов алгоритма является множество опорных узлов, включающее с избытком наиболее вероятные узлы, составляющие искомые поверхности. Следующие этапы алгоритма направлены на создание поверхностей из найденного множества опорных узлов.

Следует отметить, что найденное множество опорных узлов существенно меньше общего количества узлов сетки. И практически все последующие этапы алгоритма используют только данное подмножество узлов, что значительно ускоряет процесс обработки исходных данных в целом.

Процесс, реализуемый на этапе **построения поверхностей**, является итерационным. Он реализует стратегию последовательного подбора наиболее вероятных узлов для построения искомых поверхностей. При его реализации логически выделяются стадии *нахождения основы поверхности, расширения поверхности, слияние*

поверхностей, «латания дыр». Исходными данными для работы множество узлов, полученных на предыдущем этапе.

На стадии **нахождения основы поверхности** исходными данными являются множество опорных узлов и описываемые ниже параметры. Основой поверхности считается множество узлов, удовлетворяющих описанным при определении понятия поверхности критериям с количеством узлов равным наперед заданному значению.

При построении очередной основы поверхности вначале выбирается узел с максимальным значением модуля ВГГ из множества опорных узлов. Далее к нему делается попытка присоединить некоторое множество узлов, во-первых, необходимое для достижения наперед заданного для основы поверхности количества узлов, и, во-вторых, удовлетворяющее описанным при определении поверхности критериям поверхности. В случае достижения успеха, найденные узлы временно удаляются из множества опорных узлов.

Если не удалось построить ни одной основы поверхности, то считается, что все возможные поверхности уже определены.

В случае успеха, проводится нахождение следующей основы поверхности. Когда построено несколько основ поверхности из их множества выбирается как наиболее вероятная основа поверхности с наибольшим средним значением модуля ВГГ. Остальные основы поверхности не рассматриваются и узлы, принадлежащие к остальным найденным основам поверхности, возвращаются во множество опорных узлов.

Стадия **расширения поверхности** обеспечивает «наращивание» найденной основы поверхности и так же носит итерационный характер. Найденная основа поверхности изначально считается поверхностью. Рассматривается множество узлов, являющихся соседними с узлами поверхности. Из этого множества выбирается наиболее вероятный узел, удовлетворяющий критериям поверхности и имеющий наибольшее значение модуля ВГГ. Если такой узел найден – он заносится во множество узлов поверхности и исключается из множества опорных узлов. Если такой узел на очередном шаге итерации не обнаружен, то процесс построения очередной поверхности считается законченным.

На стадии **слияния поверхностей** производится попытка объединения построенных ранее поверхностей. Критерием принятия решения объединения поверхностей служат абсолютное и относительное количество соседних узлов в рассматриваемых областях и сохранение выполнения критериев поверхности для

объединенной поверхности. Рассматриваются все возможные пары поверхностей и делается попытка их объединения.

После завершения стадии слияния поверхностей алгоритм осуществляет попытку создания очередной основы новой поверхности, т.е. переходит к стадии нахождения основы поверхности.

После того, как на очередной итерации не удастся построить основу поверхности, считается, что все возможные при данных значениях критериев поверхности обнаружены. Производится финальная обработка найденных поверхностей.

Обработка найденных поверхностей заключается в «латании дыр» и удалении маловероятных поверхностей.

Стадия «латания дыр» предназначена для обнаружения и устранения небольших областей узлов, не принадлежащих поверхности, но расположенных внутри поверхности, т.е. узлов, окруженных узлами поверхности, но не принадлежащей ей. Как показала практика эта ситуация является достаточно типичной. Хотя в целом наличие «дыр» не имеет существенного влияния на конечный результат распознавания, но при последующей визуализации распознанных поверхностей наличие «дыр» ухудшает восприятие изображения.

Алгоритм поиска «дыры» основывается на положении, что узел поверхности имеет фиксированное количество соседних узлов в зависимости его положения на поверхности и его расположения относительно сетки. Например, узел, не принадлежащий краю поверхности, имеет восемь соседних узлов поверхности. Количество соседей узлов лежащих на границе сетки зависит от их положения и направления ВГГ. Основываясь на этом алгоритм «латания дыр» находит узлы поверхности, не имеющие достаточного количества соседей, и пробует подобрать недостающий узел, удовлетворяющий критериям распознаваемой поверхности. При этом выбор осуществляется из всех узлов не приписанный ни к одной из найденных поверхностей. В случае удачи найденный узел считается принадлежащим к поверхности и включается в соответствующую поверхность.

На стадии удаления маловероятных поверхностей производится удаление найденных поверхностей считающимися маловероятными. Критериями этого процесса могут служить количество узлов в найденных поверхностях и ряд статистических характеристик. К числу последних относятся среднее количество соседних узлов, среднее значение модуля ВГГ, среднее значение пространственного угла между двумя соседними поверхностями, среднеквадратичные отклонения последних двух величин.

Основное назначение стадии удаления маловероятных поверхностей – отбраковка поверхностей, появившихся в результате ошибок второго рода.

Выполненная программная реализация описанного алгоритма продемонстрировала его надежность и эффективность.

Результаты исследования эффективности и параметров алгоритма будут приведены в следующей статье.