УДК 528.2
DOI:10.30533/GiA-2023-002

Решение дифференциального уравнения Лапласа в виде глубокой нейросети как единый алгоритм приближённого решения задач физической геодезии в локальном районе

Ю.М. Нейман[1,2], **Л.С. Сугаипова**[1,2]

[1] Московский государственный университет геодезии и картографии, Москва, Россия
[2] Федеральный научно-технический центр геодезии, картографии и инфраструктуры пространственных данных, Москва, Россия

yuney@miigaik.ru

ЦИТИРОВАНИЕ Нейман Ю.М., Сугаипова Л.С. Решение дифференциального уравнения Лапласа в виде глубокой нейросети как единый алгоритм приближенного решения задач физической геодезии в локальном районе // Известия вузов «Геодезия и аэрофотосъемка». 2023. Т. 67. № 1. С. 18–25. DOI:10.30533/GiA-2023-002.

КЛЮЧЕВЫЕ СЛОВА дифференциальное уравнение Лапласа, геопотенциал, глубокие нейросети

АННОТАЦИЯ Одна из основных задач физической геодезии состоит в определении внешнего гравитационного поля Земли (ГПЗ). На данный момент можно полагать, что низкочастотная часть ГПЗ хорошо изучена и достаточно надежно описывается рядами по шаровым функциям. Актуальной остается задача локального моделирования высокочастотной части ГПЗ. Указанную задачу предлагается трактовать как решение дифференциального уравнения Лапласа в виде глубокой искусственной нейросети (ИНС). При этом роль псевдо начально-краевых условий играют имеющиеся в локальном районе результаты измерения различных трансформант геопотенциала. Приведена блок-схема алгоритма решения задачи в указанной постановке. В качестве глубокой ИНС используется многослойный персептрон обратного распространения. Возникающая при этом проблема вычисления частных производных решается с помощью хорошо известной в настоящее время техники автоматического компьютерного дифференцирования. Алгоритм адаптивного оценивания стохастических моментов (Adam) позволяет надежно осуществить процедуру модификации параметров ИНС. В заключении указываются теоретические проблемы, подлежащие дальнейшему исследованию: достаточность исходных данных для единственности решения, оценивание точности результатов и др.

1 Введение

Одним из новых направлений в вычислительной математике и практике создания сложных технических систем являются искусственные нейросети (ИНС), способные выполнять самые разнообразные операции, в том числе недоступные для традиционной математики. Область использования ИНС в настоящее время чрезвычайно широка — от диагностики заболеваний и автоматического анализа документов до управления динамическими системами и создания искусственного интеллекта. Поэтому естественно желание выяснить возможности нового мощного математического метода и при решении задач геодезического профиля, тесно связанного с проблемами аппроксимации и оптимизации.

Известно, что потенциал $V(x,y,z)$ сил притяжения внешнего гравитационного поля Земли выражается интегралом Ньютона [1]:

1
$$V(x,y,z) = G\iiint\limits_{\Omega} \frac{\mu(x',y',z')dx'dy'dz'}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}},$$

где (x,y,z) — точки внешнего пространства, (x',y',z') — текущие внутренние точки тела Земли Ω, $\mu(x',y',z')$ — плотность масс в текущих точках, G — гравитационная постоянная.

Понятно, что непосредственно воспользоваться соотношением (1) не представляется возможным, и только фундаментальные теории Дж. Стокса и М.С. Молоденского указали практические пути вычисления гравитационного потенциала и его трансформант на поверхности Земли и в ее внешнем пространстве без знания плотности $\mu(x',y',z')$.

Но для нас сейчас важно осознать, что согласно (1), в природе реально существует некоторая функция $V = f(x,y,z)$, которая устанавливает соответствие между произвольной точкой внешнего пространства и значением в этой точке ньютоновского потенциала. При этом современная теория ИНС доказывает [2, 3], что с помощью набора определенных операций можно построить такую ИНС, которая аппроксимирует любую непрерывную функцию с определенной наперед заданной точностью. Другими словами, ИНС при соответствующем выборе ее структуры можно сделать универсальным аппроксиматором, то есть какую бы функцию нам ни предстояло вычислить, мы знаем, что существует нейросеть, способная сделать это. В данной работе построена подобная ИНС в виде решения дифференциального уравнения Лапласа, краевое условие в котором заменено требованием удовлетворять имеющемуся набору результатов измерений различных трансформант геопотенциала в локальном районе. В настоящее время низкочастотная часть гравитационного поля Земли (ГПЗ) определена уже достаточно надежно в виде ряда по шаровым функциям, единого для всей планеты. Поэтому практический интерес представляет лишь остаточная, высокочастотная часть ГПЗ, заведомо зависящая от индивидуальных характеристик конкретного района вычислений. Следствием является необходимость использовать различные ИНС-решения для различных локальных районов. Таким образом, ограниченность района вычислений определяется не только количеством и качеством имеющихся в районе гравиметрических измерений разного рода, но и сутью задачи о локализации и моделировании именно остаточной высокочастотной части ГПЗ. Искать решение дифференциального уравнения Лапласа именно в виде глубокой ИНС представляется вполне естественным, поскольку такую модель с большим количеством параметров сравнительно легко адаптировать так, чтобы удовлетворять оператору Лапласа, сохраняя при этом требование о совпадении с заданными результатами измерений. Подобные идеи широко используются в настоящее время при решении различных дифференциальных уравнений в частных производных (см., например, [4]).

2 Материалы и методы

2.1 Постановка задачи

Пусть в некоторой пространственной области D ограниченных размеров в точках $p_1, p_2, \ldots, p_n$ с известными координатами выполнены измерения какой-либо трансформанты геопотенциала, то есть получены значения соответствующего функционала F на геопотенциале $b_1, b_2, \ldots, b_n$. Ограничимся, для простоты, наличием результатов измерений только какой-нибудь одной трансформанты, хотя излагаемый ниже алгоритм вполне допускает использование результатов измерения и нескольких различных трансформант — чем больше измерений (в том числе, разнородных) можно привлечь, тем лучше результат. Требуется найти в области D такое частное решение $V(p)$ дифференциального уравнения Лапласа:

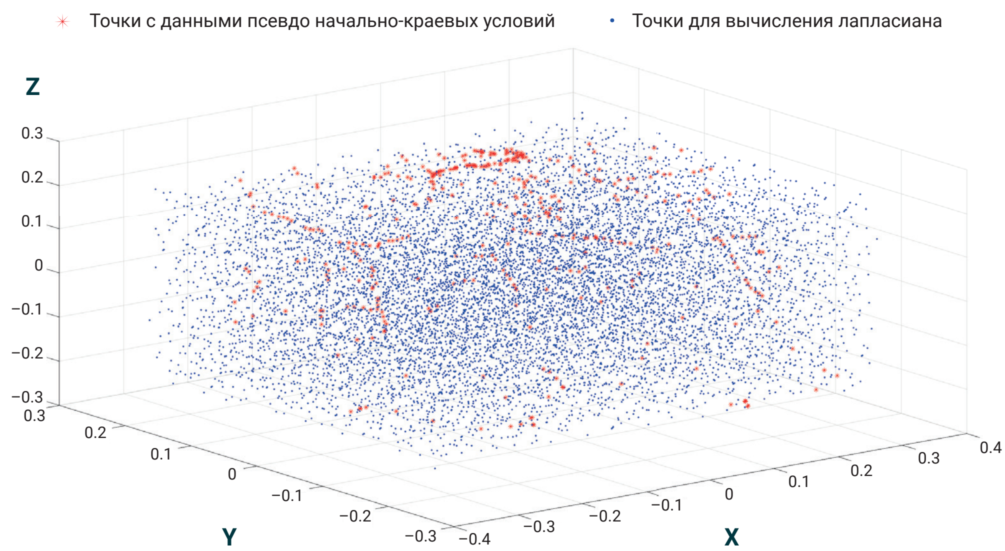
$$LapV(p) \equiv \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0, \forall p \in D, \tag{2}$$

которое удовлетворяет — после соответствующих преобразований — имеющимся значениям $b_1, b_2, \ldots, b_n$. Назовем эти значения псевдо начально-краевыми условиями (ПНК). Предполагается, что из них, с помощью упоминавшихся классических рядов по шаровым функциям, удалены известные низкочастотные составляющие. Под $LapV$ (в уравнении (2) и далее) понимается лапласиан (оператор Лапласа) функции $V(p)$, $p \in D$, а решение этого уравнения предлагается искать в виде глубокой искусственной нейросети. Последнее предложение имеет принципиальное значение, поскольку использование именно нейросетей с их уникальной способностью обучаться на конкретных данных позволяет решить поставленную задачу достаточно надежно. Понятно, что получить точное решение дифференциального уравнения Лапласа в указанных условиях не представляется возможным. Таким образом, речь идет, в отличие от классических краевых задач, лишь о приближённых оценках остаточного возмущающего геопотенциала, точность которых определяется той гравиметрической информацией, которой мы в данном районе располагаем. Но ведь, строго говоря, точные решения практических задач, опирающихся на реальные точечные, а не непрерывные измерения, вообще не достижимы.

2.2 Принципиальная схема алгоритма

Воспользуемся местной системой прямоугольных координат, и пусть П1 и П2 обозначают два прямоугольных параллелепипеда, грани которых параллельны координатным плоскостям. При этом П1 — минимальный по размерам прямоугольный параллелепипед, целиком охватывающий все множество точек $p_1,p_2,...,p_n$, а П2 является внешним по отношению к П1 так, что расстояния между гранями П1 и П2 по координатным осям определены как $h_x \geq 0$, $h_y \geq 0$, $h_z \geq 0$. Внутреннее пространство П2 будем обозначать D и называть локальным районом исследований (Рис. 1).

Рис. 1
Локальный район исследований D и расположение в нем точек для обучения ИНС.



Из множества заданных точек $p_1,p_2,...,p_n$ и соответствующих значений заданного функционала F на искомом потенциале V целесообразно выделить 15–20 %, которые не будут использоваться в обучении ИНС, но позволят оценить обобщающие способности уже обученной сети. Множество координат таких точек будем обозначать как XYZ_{test}, а оставшееся множество для обучения — как XYZ_{train}.

Далее в области D необходимо сгенерировать достаточно плотную пространственную регулярную сетку, шаг которой зависит от сложности ГПЗ в D и подбирается экспериментально. Количество узлов этой сетки должно значительно превышать (в 10–100 раз) количество точек с ПНК. Множество координат всех узлов созданной сетки обозначим как XYZ_{lap} и будем использовать для того, чтобы выход $V(p)$ из описанной далее ИНС удовлетворял уравнению (2).

Таким образом, в изучаемой области D имеется три массива точек, имеющих разное предназначение. При дальнейшем описании схемы алгоритма удобно различать эти массивы по номерам: $XYZ_{lap} = XYZ1$, $XYZ_{train} = XYZ2$, $XYZ_{test} = XYZ3$.

В качестве глубокой ИНС рекомендуется многослойный персептрон обратного распространения со $S \geq 3$ скрытыми взаимосвязанными слоями по N нейронов в каждом, с тремя каналами на входе (координаты точек) и одним каналом на выходе (текущее значение потенциала V). В качестве нелинейных передаточных функций, по-видимому, лучше всего пользоваться гиперболическим тангенсом:

$$th\,x = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (th\,x)' = 1 - th^2 x,$$

поскольку он (в отличие от широко используемой сигмоиды) симметричен относительно нуля (что удобно при решении задач аппроксимации), а его производная легко вычисляется.

Совокупность весов и смещений для всех слоев будем называть параметрами ИНС. Их первоначальные значения можно задавать по-разному, например, с помощью специального инициализатора He [5]. Но сейчас это не столь важно, и поэтому будем полагать, что параметры назначены просто в виде нормированных случайных чисел. В результате инициализации параметров ИНС можно трактовать как некоторую функцию:

3 $$f(x, y, z, parameters) = V.$$

Для компьютерной реализации такой функции написана специальная программа, которую будем называть $model$. На вход этой программы, как видно из (3), должны поступать текущие значения параметров ИНС и координаты точек из области D, а выходом служит некоторая функция V, которая подлежит обучению. Под обучением понимается разработка и реализация определенных обучающих рекомендаций по улучшению параметров ИНС. Эти рекомендации базируются на мере близости текущего значения V нашим целям: если входные координаты x,y,z взяты из массива $XYZ1$, то лапласиан соответствующей функции V должен равняться нулю; если входные координаты x,y,z взяты из массива $XYZ2$, то соответствующая функция V должна удовлетворять ПНК условиям, то есть совпадать (после соответствующих преобразований) с заданными значениями $b_1, b_2, \dots, b_n$. Поэтому введем функцию цели в виде:

$$aim = \frac{c_1}{k}\sum_{i=1}^{k}(LapV(p_i))^2 + \frac{c_2}{n}\sum_{i=1}^{n}(\widehat{F}(p_i) - F(p_i))^2 \Rightarrow minimum,$$

где c_1 и c_2 — весовые коэффициенты, k — количество точек в массиве $XYZ1$, n — количество точек в массиве $XYZ2$, $F(p_i) = b_i$ — заданное значение в точке p_i исследуемого функционала, $\widehat{F}\left(p_i\right)$ — соответствующее вычисленное значение этого функционала на текущем значении функции V.

Последовательность действий, составляющих одну эпоху обучения, показана на Рисунке 2. Предполагается, что сначала выполняются действия с массивом $XYZ1$, а затем — с массивом $XYZ2$. Операция $grad\ aim$ вычисляет производные функции цели относительно параметров модели и, таким образом, обеспечивает важнейший принцип обратного распространения ошибок и модификацию параметров по схеме:

4 $$\theta_{it+1} = \theta_{it} - \alpha \cdot grad\ aim(\theta_{it}),$$

где θ обозначает параметры ИНС, it — номер итерации, $\alpha > 0$ определяет скорость обучения.

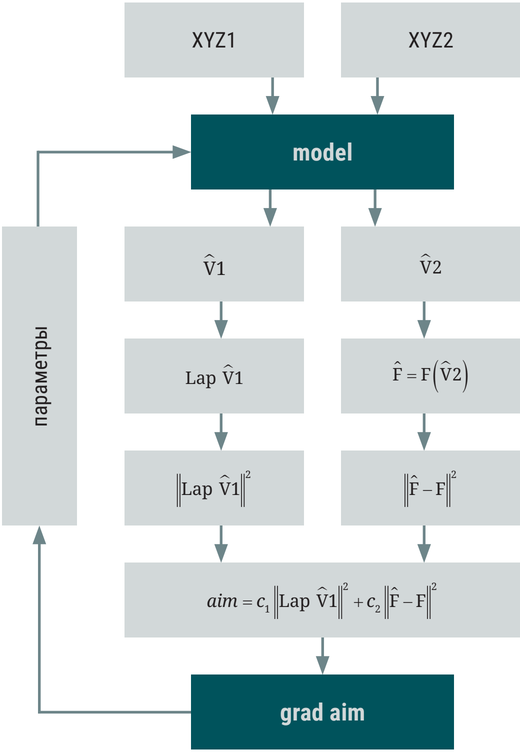


Рис. 2
Блок-схема предлагаемого алгоритма.

Модифицированные параметры порождают модифицированную функцию V и, следовательно, модифицированное значение функции цели. Такой циклический процесс повторяется до тех пор, пока функция цели не окажется меньше заданного малого числа $\varepsilon > 0$ или количество выполненных эпох не превысит заданное предельное значение. Качество обученной ИНС проверяется путем вычисления значения функции V при использовании не участвовавшего в обучении массива координат $XYZ3$, вычисления значений соответствующих функционалов на этой функции и сравнения с ПНК.

Таким образом, описанный алгоритм представляет собой некоторый вариант глобальной оптимизации определенной функции цели известным методом градиентного спуска. При этом важнейшей технической проблемой до недавнего времени являлась процедура компьютерного вычисления частных производных, необходимых при вычислении градиентов и при составлении лапласиана текущей функции V. Но сегодня эта техника уже хорошо освоена, и если удается представить сложную функцию в виде композиции более простых, то вычислить ее производную по любой переменной оказывается довольно просто (см., например, [6, 7]). Наиболее целесообразным является представление исследуемой сложной функции в виде графа вычислений, узлами которого являются достаточно простые функции (обычно взятые из заранее фиксированного набора), а ребра связывают функции со своими аргументами. В нейронных сетях в качестве базисных элементарных функций графа вычислений обычно используют функции, из которых получаются нейроны. Например, скалярного произведения векторов X^TY и сигмоиды σ достаточно для того, чтобы построить любую, даже самую сложную нейронную сеть, составленную из нейронов с функцией активации σ [7]. При программировании практически на любом современном языке можно воспользоваться готовыми пакетами, обеспечивающими компьютерное автоматическое дифференцирование.

Аналогичная ситуация имеет место и относительно другой важнейшей части предлагаемого алгоритма — блока $grad\ aim$. Возложенные на этот блок операции модификации параметров ИНС, приводящие к уменьшению значения функции цели, достаточно просто реализовать, пользуясь современными разработками. Мы рекомендуем пользоваться алгоритмом адаптивного оценивания стохастических моментов первых двух порядков, описанным в работе [8] под названием Adam (adaptive moment estimation). В отличие от стандартного градиентного спуска, где градиент в (4) вычисляется с использованием сразу всего обучающего множества, Adam работает в пакетном режиме, то есть вычисляет градиент и обновляет параметры ИНС последовательно с использованием частичных обучающих подмножеств (англ. — minibatch). Полная обработка одного такого подмножества и соответствующее уточнение параметров ИНС определяет одну итерацию. Таким образом, эпоха, соответствующая полной обработке всего обучающего множества, состоит из определенного количества итераций, которое назначается пользователем. Adam является градиентным спуском, по существу, стохастическим, поскольку обновленные параметры, рассчитанные с использованием мини-пакетов, представляют собой некоторые оценки того, что было бы получено в результате использования полного набора обучающих данных. При этом метод персонально уточняет скорости обучения для различных параметров по оценкам первого и второго моментов производных. При необходимости предусмотрена возможность использования скользящего среднего, что уменьшает неизбежно накапливающиеся шумы.

3 Результаты и выводы

Реализация предлагаемого алгоритма в виде соответствующей компьютерной программы позволяет единообразно решать следующий класс задач физической геодезии. Пусть в локальном районе даны значения одного или нескольких различных функционалов на остаточном возмущающем потенциале и координаты местоположений этих значений. Требуется найти функцию координат точек исследуемого района, оценивающую локальный остаточный возмущающий потенциал и/или значения какой-нибудь трансформанты (или значения нескольких различных трансформант) в указанных точках изучаемого района.

Полученные результаты, помимо свойств исходных данных, существенно зависят, конечно, от избранной архитектуры ИНС, меры ее глубины, выбора упоминавшихся в описании параметров и других факторов, разумно установить которые можно лишь в процессе экспериментов.

Если говорить об оставшихся многочисленных вопросах теоретического характера, то важнейшими из них являются, по-видимому, вопросы достаточности исходных данных для единственности решения, оценивания точности результатов и др.

БЛАГОДАРНОСТИ

Исследование выполнено в рамках федерального проекта «Поддержание, развитие и использование системы ГЛОНАСС» государственной программы Российской Федерации «Космическая деятельность России» на 2021–2030 гг., Регистрационный номер ЕГИСУ НИОКТР № 1210806000081-5 и в рамках госзадания Минобрнауки РФ FSFE-2023-0005 «Фундаментальные и поисковые исследования путей повышения эффективности комплексного использования разнородных пространственных данных в интересах ускорения цифровой трансформации экономики и обеспечения устойчивого развития территорий Российской Федерации».

The study was carried out within the Federal project “Maintenance, development and use of the GLONASS system” for the State Program of the Russian Federation “Space activities of Russia” for 2021–2030, EGISU NIOKTR registration No. 1210806000081-5 and within the State assignment of the Ministry of Education and Science of the Russian Federation, No. FSFE-2023-0005 “Fundamental research and exploration of the ways to improve the efficiency for the integrated use of heterogeneous spatial data in the interests of accelerating the digital transformation of economy and ensuring sustainable development of the Russian Federation territories”.

БИБЛИОГРАФИЯ

1. Гофман-Велленгоф Б., Мориц Г. Физическая геодезия. Перевод с англ. М.: МИИГАиК, 2007. 410 с.
2. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators // Neural Networks. 1989. № 2. p. 359–366.
3. Leshno M., Lin V., Pinkus A., et al. Multilayer feedforward networks with nonpolynomial activation function can approximate any function // Neural Networks. 1993. № 6. p. 861–867.
4. Raissi M., Perdikaris P., Karniadakis G. E. Physics informed deep learning (part 1): Datadriven solutions of nonlinear partial differential equations // Preprint arXiv.org, 2017. [Электронный ресурс]. Режим доступа: https://arxiv.org/abs/1711.10561 (дата обращения: 22.12.2022).
5. He K., Zhang X., Ren S., et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification // 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile. 2015. P. 1026–1034. DOI:10.1109/ICCV.2015.123.
6. Domke J. Automatic Differentiation and Neural Networks. [Электронный ресурс] Режим доступа: https://people.cs.umass.edu/~domke/courses/sml2011/08autodiff_nnets.pdf (дата обращения: 10.12.2022).
7. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. СПб: Питер, 2018. 480 с.
8. Kingma D. P., Ba J. Adam. A method for stochastic optimization // Preprint arXiv.org, 2014. [Электронный ресурс]. Режим доступа: https://arxiv.org/abs/1412.6980 (дата обращения: 01.12.2022).

АВТОРЫ

Нейман Юрий Михайлович
ФГБУ «Федеральный научно-технический центр геодезии, картографии и инфраструктуры пространственных данных», Москва, Россия;
ФГБОУ ВО «Московский государственный университет геодезии и картографии» (МИИГАиК), Москва, Россия
кафедра высшей математики
д-р техн. наук, профессор
0000-0002-7728-6567

Сугаипова Лейла Супьяновна
ФГБУ «Федеральный научно-технический центр геодезии, картографии и инфраструктуры пространственных данных», Москва, Россия;
ФГБОУ ВО «Московский государственный университет геодезии и картографии» (МИИГАиК), Москва, Россия
кафедра высшей математики
д-р техн. наук, доцент
0000-0002-3393-7823

Поступила 28.01.2023. Принята к публикации 20.02.2023. Опубликована 28.02.2023.