

ГЕОМЕТРИЧЕСКИЙ СМЫСЛ МЕТОДА НЬЮТОНА

М.В. Пчелинцев, Н.А. Скоркин

Обнаруженный геометрический смысл метода Ньютона для решения систем нелинейных уравнений (в бесконечномерном случае - операторных уравнений) полностью проясняет его механизм. В практико-прикладном направлении это позволяет объяснить эмпирически наблюдаемые эффекты, получить единую характеристику метода и его модификаций, общую теорему локальной сходимости и ясное видение геометро-динамической природы проблемы сходимости в целом. Результаты демонстрируются модельным примером.

Ключевые слова: метод Ньютона, риманово пространство, численные методы, дифференциальные уравнения.

1. Концептуальные соображения

Метод Ньютона представляет собой формулу перехода от k -го приближения x_k численного решения уравнения $F(x) = 0$ соответственно к $k+1$ -му приближению x_{k+1} , которая имеет вид

$$x_{k+1} = x_k - [DF(x_k)]^{-1} F(x_k). \quad (1)$$

Хорошо известный в конечномерном случае геометрический смысл формулы состоит в том, что точка x_{k+1} получается как точка пересечения определенных гиперплоскостей, ортогональных к нормальям (градиентам) гиперповерхностей уровня компонент $F(x)$ в точке x_k [1,2].

Посмотрим на функцию F как на отображение из R^n в R^n :

$$F: X \subset R^n \rightarrow Y \subset R^n \sim y = F(x) \sim y^i = f_i(x^1, x^2, \dots, x^n) \quad (i=1, 2, \dots, n),$$

$$X = \left\{ x: x = (x^1, x^2, \dots, x^n)^T \right\}, \quad Y = \left\{ y: y = (y^1, y^2, \dots, y^n)^T \right\}, \quad F(x) = (f_1(x), f_2(x), \dots, f_n(x))^T.$$

В общем случае метрический тензор $g_{ij} = \left(\frac{\partial F}{\partial x^i}, \frac{\partial F}{\partial x^j} \right) = \sum_{k=1}^n \frac{\partial f_i}{\partial x^k} \frac{\partial f_j}{\partial x^k}$ не задает риманову структуру на всем X . Не вступая в противоречие с этим обстоятельством и в силу необходимости, мы воспользуемся понятиями и конструкциями риманова пространства. То есть во всех случаях, когда это необходимо, будет предполагаться, что отображение F непрерывно дифференцируемо нужное число раз (не на всем X , а в некоторых окрестностях точек и кривых).

Всюду в дальнейшем: x_* – неизвестное решение уравнения $F(x) = 0$; x_k ($x_k \neq x_*$) – очередное приближение к x_* ; $y_k = F(x_k)$; $\|\cdot\|, |\cdot|$ – норма и евклидова длина в R^n , определенные скалярным произведением, действующим на векторах стандартного базиса по правилу $(e_i, e_j) = \delta_{ij}$ (δ_{ij} – символ Кронекера); $|\cdot|_{\text{РГ}}$ – длина в смысле римановой метрики [3]; l – прямая в Y , состоящая из множества точек вида $y = (1-t)y_k$ ($0 \leq t < \alpha \leq 1$), которую будем называть идущей из точки y_k в 0; l_* – геодезическая линия в X в смысле римановой метрики, идущая из x_k в x_* , моделирующая прямую l (в частности $F(l_*) = l$, $|l_*|_{\text{РГ}} = |l|$).

Естественный вопрос – почему бы не попробовать двигаться из x_k в x_* по кривой l_* , которая является прямой в римановом смысле?

Уравнения геодезических в натуральной параметризации (а также в любой другой параметризации, пропорциональной натуральной) суть уравнения $\ddot{x}^i = -\Gamma_{jk}^i \dot{x}^j \dot{x}^k$ ($i=1, 2, \dots, n$) [3] (точки над символом обозначают дифференцирование по параметру, буква для обозначения которого

опущена, Γ_{jk}^i - коэффициенты связности и по повторяющимся индексам предполагается суммирование). Так как уравнения 2-го порядка, то для их интегрирования (хоть аналитического, хоть численного) требуется знать не только начальное положение, но и, например, начальный вектор скорости, причем, направленный вдоль геодезической от x_k к x_* . Очевидно, что таковым, готовым к применению является вектор $a_k = [DF(x_k)]^{-1}(-y_k) = -[DF(x_k)]^{-1}y_k = -[DF(x_k)]^{-1}F(x_k)$ (ибо, так как вектор $-y_k$, как вектор касательного пространства $T_{y_k}Y$, является касательным к прямой l в точке y_k и направлен от y_k к 0 , то линейным отображением $[DF(x_k)]^{-1}: T_{y_k}Y \rightarrow T_{x_k}X$ он переводится в вектор, касательный к кривой l_* в точке x_k и направленный вдоль l_* от x_k к x_*).

Вектор a_k есть в точности вектор смещения из точки x_k в точку x_{k+1} в методе Ньютона (1). И мы тем самым обнаруживаем, что шаг по методу Ньютона (1) - это прямолинейный шаг по касательной к геодезической линии l_* из точки x_k на всю длину вектора a_k .

О величине шага. Очевидно, имеют место быть равенства $|a_k|_{\mathbb{R}} = |-y_k| = |l| = |l_*|_{\mathbb{R}} = \|F(x_k)\|$, из которых следует, что метод Ньютона пытается одним прямолинейным шагом по касательной к линии l_* преодолеть всю ее (риманову) длину, равную $\|F(x_k)\|$, целиком. Нетрудно видеть, что настрой на такой «прыжок» наследуется из евклидова пространства Y . И в общем случае он слишком велик. Совершенно очевидно, что в этом кроется одна из причин зачастую быстрой потери сходимости метода при удалении начального приближения от искомого решения.

Обращая теперь внимание на тот факт, что векторное поле $a(x) = -[DF(x)]^{-1}F(x)$ в каждой точке x направлено по касательной к геодезической l_* , проходящей через точку x , замечаем, что эти линии являются интегральными кривыми поля $a(x)$. То есть они с аналитической точки зрения являются решениями автономной системы дифференциальных уравнений первого порядка:

$$\dot{x} = -[DF(x)]^{-1}F(x). \quad (2)$$

Соответственно, точки x_* являются стационарными решениями этой системы.

В общем случае решения системы (2) не являются решениями уравнений для геодезических (из-за различия в параметризациях). Причем если линии l_* как решения уравнений для геодезических всегда определены на конечных промежутках изменения параметра, то те же линии как решения системы (2) могут быть определены как на конечных, так и на бесконечных промежутках.

Если $x(t)$ - решение системы (2) с начальным условием $x(0) = x_0$ и $y(t) = F(x(t))$, то так как $\dot{y}(t) = \frac{d}{dt}F(x(t)) = [DF(x(t))]\dot{x}(t) = -F(x(t)) = -y(t)$, то $y(t) = y_0 e^{-t}$ - прямая в Y , идущая из точки $y_0 = y(0) = F(x(0)) = F(x_0)$ в ноль. В частности, отсюда следует, что для любых двух точек, соединяемых линией $x = x(t)$, она является кратчайшей. То есть она является экстремалью функционала длины [3], и, как говорят в таком случае ([4, с. 199]), с точностью до перепараметризации является геодезической. В дальнейшем слова «с точностью до перепараметризации» будут опускаться.

Если рассматривать линии l_* в натуральной (в римановом смысле) параметризации, то векторное поле, порождающее их должно иметь риманову длину 1. Очевидно, что таковым является поле $\tilde{a}(x) = |a(x)|_{\mathbb{R}}^{-1}a(x) = \|F(x)\|^{-1}a(x) = -\|F(x)\|^{-1}[DF(x)]^{-1}F(x)$. То есть в этой параметризации линии l_* являются интегральными кривыми автономной системы уравнений

$$\dot{x} = -\|F(x)\|^{-1}[DF(x)]^{-1}F(x). \quad (3)$$

2. К вопросу о локальной сходимости

Нетрудно догадаться, учитывая выбор направления движения по линиям l_* , что решения x_* уравнения $F(x) = 0$ должны быть устойчивыми узлами системы уравнений (2).

Теорема. Пусть функция $F(x)$ дважды непрерывно дифференцируема хотя бы в некоторой окрестности решения x_* уравнения $F(x) = 0$ и линейное отображение $DF(x_*)$ не вырождено. Тогда x_* является асимптотически устойчивым узлом системы уравнений (2).

Доказательство. Так как отображение $DF(x_*)$ не вырождено, то по теореме об обратном отображении [5] $F(x)$ является локальным диффеоморфизмом. Пусть \tilde{x} ($\tilde{x} \neq x_*$) – произвольная точка окрестности, в которой диффеоморфизм реализуется, и $\tilde{y} = F(\tilde{x})$. Прямую l , идущую из \tilde{y} в 0, можно задать параметрическими уравнениями $y = \tilde{y} - t\tilde{y}$ ($0 \leq t < 1$). Кривая $l_* = F^{-1}(l)$ является гладкой кривой, идущей из \tilde{x} в x_* , определяемой параметрическими уравнениями $x(t) = F^{-1}(\tilde{y} - t\tilde{y})$, и $|l_*|_{\text{гр}} = |l|$. Для любой другой кривой l_{**} , идущей из \tilde{x} в x_* , имеющей конечную риманову длину, ее образ $F(l_{**})$ является кривой, имеющей конечную евклидову длину, идущей из \tilde{y} в 0. Так как $|l_{**}|_{\text{гр}} = |F(l_{**})| \geq |l| = |l_*|_{\text{гр}}$, то кривая l_* является геодезической и, следовательно, в некоторой параметризации является интегральной кривой системы уравнений (2). Непосредственно проверяется, что таковой является параметризация τ , которая связана с t -параметризацией монотонной зависимостью $\tau = -\ln(1-t) \Leftrightarrow t = 1 - e^{-\tau}$, $\tau \in [0; +\infty) \Leftrightarrow t \in [0; 1)$. По теореме существования и единственности решения [6] это единственная интегральная кривая системы (2), выходящая из точки \tilde{x} . Так как $\tau \rightarrow +\infty \Leftrightarrow t \rightarrow 1 - 0$, то по определению и по непрерывности имеем $\lim_{\tau \rightarrow +\infty} x(t(\tau)) = \lim_{t \rightarrow 1-0} x(t) = \lim_{t \rightarrow 1-0} F^{-1}(\tilde{y} - t\tilde{y}) = F^{-1}\left(\lim_{t \rightarrow 1-0} (\tilde{y} - t\tilde{y})\right) = F^{-1}(0) = x_*$.

Теорема доказана.

Если $x(s)$ – решение системы (3) с начальным условием $x(0) = \tilde{x}$, где \tilde{x} – точка из доказательства теоремы, то $\lim_{s \rightarrow \|F(\tilde{x})\|_0} x(s) = x_*$. То есть в условиях доказанной теоремы x_* является изолированной точкой притяжения для системы (3) (не в смысле устойчивости по Ляпунову).

Из доказанной теоремы непосредственно вытекает общая теорема о локальной сходимости.

Теорема. Если хотя бы в некоторой окрестности решения уравнения $F(x) = 0$ итерационная схема аппроксимирует систему уравнений (2) или (3) и вычислительно устойчива, то определяемый ею итерационный процесс является локально сходящимся.

Отметим, в частности, что если аппроксимирующий характер метода Ньютона (1) для системы (2) внешне не угадывается, то для системы (3) он легко устанавливается.

Легко также видеть, что для упрощенного метода Ньютона $x_{k+1} = x_k - [DF(x_0)]^{-1} F(x_k)$ аналогом уравнения (2) является уравнение $\dot{x} = -[DF(x_0)]^{-1} F(x)$, имеют место аналогичные теоремы, причем от функции $F(x)$ достаточно потребовать, чтобы она была непрерывно дифференцируемой или была непрерывной и удовлетворяющей условию Липшица.

3. К вопросу о глобальной сходимости

Можно показать, что излагаемая ниже картина справедлива и в том случае, когда отображение F лишь непрерывно дифференцируемо (не обязательно на всем X). При этом интегральные кривые системы (2) будут геодезическими l_* , как стационарные кривые функционала длины [3].

Реальная картина такова.

Вся область X , за исключением подмножеств особых точек, покрывается непродолжаемыми интегральными кривыми уравнения (2) с соответствующим направлением движения по ним, определяемым векторным полем $a(x)$. Эти кривые являются (направленными) геодезическими l_* в римановой метрике.

Для всякого решения x_* исходного уравнения $F(x) = 0$ можно определить *область сходимости* к нему *по геодезическим*, как множество всех таких точек в X , для которых существуют проходящие через них линии l_* , идущие в x_* и достигающие x_* . Области, состоящие из точек, лежащих на линиях l_* , которые ведут в решения x_* , но не достигают их, можно соответственно

назвать областями несходимости. Подмножества особых точек являются "перегородками" между указанными областями двух типов.

Для всякой точки области сходимости по геодезическим к решению x_* , рассматриваемой как начальное приближение к x_* , можно добиться сходимости из нее к x_* . В крайнем случае, - движением мелкими шагами по одной геодезической. Представляется, что в рамках ньютоновской методологии это, по сути, пока единственная стратегия (в контексте задачи глобальной сходимости), гарантирующая положительный конечный результат.

Перспективу попадания из точек областей несходимости в какое-либо решение x_* можно охарактеризовать словами "как повезет". Эксперименты показывают, что, как правило, не везет. Вопрос о том, как узнать для конкретной точки, попадает ли она в какую-либо область сходимости по геодезическим или нет, остается открытым.

О существовании решения исходного уравнения $F(x) = 0$. Во-первых, существование решения может следовать из физических соображений. Во-вторых, если таких соображений нет, то в распоряжении имеется известное векторное поле $a(x)$ системы уравнений (2), которое указывает решения x_* (оно, образно говоря, отовсюду в области сходимости по геодезическим "течет" в одну точку). В двухмерном и трехмерном случаях все особенности портрета интегральных кривых созерцаемы непосредственно.

4. Характеризация метода Ньютона и его модификаций и еще о проблеме сходимости

Выше была установлена геометрическая картина итерационных шагов поиска x_* по методу Ньютона, которая заключается в следующем. Из точки x_k по касательной к l_* делается прямолинейный шаг на всю (евклидову) длину вектора $a_k = -[DF(x_k)]^{-1} F(x_k)$. Поскольку $|a_k|_{\mathbb{R}} = |l_*|_{\mathbb{R}} = \|F(x_k)\|$, то в римановом смысле длина этого шага равна (римановой) длине линии l_* , которая есть $\|F(x_k)\|$. Следующее приближение x_{k+1} является концевой точкой вектора a_k , отложенного соответственно от x_k . Из точки x_{k+1} (если она не вылетела из области сходимости в область несходимости) выходит своя линия l_* , идущая в решение x_* (которое уже может быть не тем же самым, что и для точки x_k , если произошел перескок в другую область сходимости). Далее итерационная процедура повторяется из x_{k+1} и т. д.

Представляется, что такая картина итерационных шагов является характерной для метода Ньютона и его модификаций (методов ньютоновского типа). То есть общий геометрический смысл их состоит в том, что делается шаг по направлению «вдоль» геодезической и он имеет как евклидову, так и определенную риманову длину. При этом передвижение "вдоль" геодезической будет буквальным, если величина шага мала. Если она не мала, то возникают отклонения, причем каждая модификация имеет свое собственное специфическое отклонение, обусловленное поведением геодезической в данной точке.

Сделанный вывод подтверждается двумя обстоятельствами: 1) методы ньютоновского типа имеют одну и ту же методологическую природу (в частности, метод Ньютона (1) является главной линейной частью итерационного шага); 2) метод Ньютона и его модификации обеспечивают сходимость к решению x_* уравнения $F(x) = 0$ только в том случае, когда начальные приближения достаточно близки к x_* [1]. Объясним, почему второе обстоятельство является подтверждающим, а также выясним его природу.

Точка x_k может находиться в непосредственной близости от неизвестного решения x_* . Линия l_* может иметь в x_k большую евклидову скорость a_k и в то же время (в евклидовом смысле) проходит мимо x_* . Пройдя мимо, кривая l_* в какой-то момент, и будучи (в принципе как угодно) далеко от x_* , начнет разворачиваться, развернется и направится либо уже непосредственно в x_* , либо перед этим еще попетляет. Из-за большой скорости полный шаг по методу Ньютона из точки x_k (для полноты которого, как мы скоро увидим, имеются как бы все основания) приводит к явлению, которое на практике получило общее название «прострел» [1]. То есть, образно говоря,

вектор a_k «пробрасывает» точку x_{k+1} мимо x_* в даль (хотя для потери сходимости хватает, как правило, и не очень далеко), из которой итерации уже никогда не вернуться обратно.

С другой стороны. Так как точка x_k находится в непосредственной близости от x_* , то величина $\|F(x_k)\|$ может быть достаточно малой и в силу этого сигнализирующей о том, что искомое решение x_* где-то близко. То есть для достижения цели имеются все основания сделать решительно полный ньютоновский шаг a_k !!!

Отметим также следующее. Из-за наличия в непосредственной близости от x_* точек «протрельного» типа шаровая область сходимости (с центром в x_*) может оказаться очень малой (в принципе, сколь угодно малой). В то же время реальная область сходимости может быть заметно большего размера, но по своей конфигурации сильно отличающейся от шаровой. Этот факт (экспериментально проверенный) является теоретическим объяснением давно сделанного на практике вывода о том, что классические результаты, в том числе свежие, о сходимости метода Ньютона [7 - 10] имеют преимущественным образом лишь теоретическое значение.

Аналогично обстоят дела и для модификаций метода. Как уже отмечалось, каждая модификация имеет на итерационном шаге свое собственное специфическое отклонение от геодезической, обусловленное ее поведением. То есть для каждой модификации есть благоприятные точки, из которых шаг делается как бы по направлению к решению, и есть неблагоприятные точки, из которых шаг делается совсем в другом направлении (итерации расходятся). Причем последние точки могут быть сколь угодно близки к искомому решению. Вопрос о том, как узнать для конкретной точки, благоприятна она или нет, остается открытым. Вытекающая из этого непредсказуемость итерационного шага с точки зрения дальнейшей перспективы есть еще одна характерная черта методов ньютоновского типа.

Отсутствие глобальной сходимости у методов ньютоновского типа считается одним из недостатков [2]. Теперь же видно, что это не недостаток, а их естественное свойство.

Легко объясняется и большая локальная скорость сходимости, которая на практике зачастую выше, нежели устанавливаемая теорией. Вблизи от x_* линии l_* мало отличаются от прямых, равенство римановых длин a_k и l_* влечет с пренебрежимой погрешностью равенство их евклидовых длин. Ньютоновский шаг практически становится шагом прямо в x_* .

Итак, можно сделать вывод, что проблема сходимости в рамках ньютоновской методологии (хоть локальной, хоть глобальной) напрямую связана с *геометрической структурой интегральных кривых уравнения (2) и динамикой движения по ним*.

Краткое объяснение природы больших евклидовых скоростей движения по линиям l_* таково. Если для некоторого участка прямой l , имеющего небольшую длину, его прообраз на l_* напротив сильно растянут в евклидовом смысле (а это растяжение, в принципе, может быть сколь угодно большим), то на таком участке кривой l_* и будет большая скорость движения.

Проблема глобальной сходимости с геометрической точки зрения - это проблема ориентации, проблема обнаружения направления местоположения искомого решения. Соответственно в практическом плане - это проблема геометрически осмысленного выбора величины и направления шага (в частности, отхода в нужный момент от направления движения вдоль геодезической). Проблема очень непростая, поскольку имеет не локальный характер.

5. Модификация метода Ньютона и некоторые общие замечания

Системой уравнений (2) можно воспользоваться для модификации метода Ньютона (1) в сторону повышения его порядка аппроксимации и соответственно увеличения скорости сходимости за счет дополнительных поправочных членов высших порядков малости. Поправку второго порядка малости можно получить также исходя непосредственно из уравнений для геодезических.

Пусть $x(t)$ - параметрические уравнения линии l_* как интегральной кривой системы уравнений (2) с начальным условием $x(0) = x_k$, $x(s)$ - параметрические уравнения l_* в натуральной (в римановом смысле) параметризации как интегральной кривой системы уравнений (3) с начальным условием $x(0) = x_k$. И пусть параметры t и s отвечают одной и той же точке $x(t) = x(s)$.

Нетрудно вывести локальную связь $t = s \|F(x_k)\|^{-1}$. Так как s есть риманова длина участка геоде-

зической l_* от x_k до $x(s)$, а $\|F(x_k)\| = |l_*|_{gr}$, то полученная связь означает, что параметр t определяет долю натурального параметра s относительно римановой длины кривой l_* .

Разложение функции $x(s)$ в ряд Тейлора в окрестности точки x_k с условием $x(0) = x_k$, как показывают вычисления, имеет вид $x(s) = x_k + a_k s \|F(x_k)\|^{-1} + 0,5 b_k s^2 \|F(x_k)\|^{-2} + \dots$, где $b_k^i = -\Gamma_{lm}^i(x_k) a_k^l a_k^m$, $(a_k^1, \dots, a_k^n)^T$ - компоненты вектора $a_k = -[DF(x_k)]^{-1} F(x_k)$, а коэффициенты связности $\Gamma_{lm}^i(x_k)$ вычисляются непосредственно по определению $\frac{\partial^2 F}{\partial x^l \partial x^m} = \text{представляя}$ в этом разложении справа первые три слагаемых, полагая $x(s) = x_{k+t}$, учитывая связь между t и s и добавляя управляющие параметры α_k, β_k , получим формулу

$$x_{k+1} = x_k + \alpha_k a_k t + 0,5 \beta_k b_k t^2, \quad (4)$$

которая представляет собой модифицированный метод Ньютона, имеющий (при $\alpha_k = \beta_k = 1$) третий порядок аппроксимации. Параметр t ($0 \leq t \leq 1$) определяет длину итерационного шага от x_k до x_{k+1} относительно римановой длины геодезической l_* . При $t = 1$ - шаг на всю длину, при $t = 0,5$ - шаг на полдлины и т. д. Параметры α_k и β_k предназначены для дополнительной корректировки шага и, в первую очередь, его направления.

Если в формуле (4) оставить только линейную часть приращения (положить $\alpha_k = 1, \beta_k = 0$), то, как и должно быть, получится формула численного интегрирования системы уравнений (2) явным методом Эйлера $x_{k+1} = x_k + t a_k = x_k - t [DF(x_k)]^{-1} F(x_k)$. Формально заменяя букву t на ω_k , получим формулу наиболее простой известной модификации метода Ньютона, позволяющей реализовать на итерационном шаге требование уменьшения нормы $\|F(x_{k+1})\| \leq \|F(x_k)\|$ [1, с. 185]. Следует отметить, что такое требование в общем случае недостаточно для сходимости итераций к какому-либо решению. Так происходит, например, при движении по одной геодезической l_* , которая не достигает решение (модельный пример, рис. 5). Геометрический смысл параметра ω_k ясен, и стандартный метод Ньютона (1) получается, как и должно быть, при значении $t = \omega_k = 1$.

6. Модельный пример

В качестве модельного примера рассматривалась задача решения системы двух нелинейных уравнений

$$f_1(x) = \left(\frac{x^1}{7}\right)^2 + \left(\frac{x^2}{2}\right)^2 - 1 = 0, \quad f_2(x) = \left(\left(\frac{x^1}{6,5}\right)^2 + \left(\frac{x^2}{2,1}\right)^2 - 1\right) \cdot \exp\sqrt{\left(\frac{x^1}{6,5}\right)^2 + \left(\frac{x^2}{2,1}\right)^2} = 0.$$

Система является симметричной относительно координатных осей и начала координат. Решений четыре - по одному в каждой координатной четверти. В первой координатной четверти решение есть точка с координатами $x_*^1 = 4,24559788\dots$, $x_*^2 = 1,59014638\dots$

Результаты численных экспериментов представлены на рис. 1-7.

На рис. 1 приведено поле направлений вектора $a(x)$ для первой и четвертой координатных четвертей. Хорошо просматривается наличие двух точек притяжения, областей отталкивания и притяжения на координатных осях. Например, на оси Ox интервалы OC, DE - области отталкивания, отрезок CD и луч от E вправо - притягивающие множества.

На рис. 2 приведен портрет интегральных кривых системы уравнений (2) (геодезических линий l^*). Интегральные кривые строились численно с использованием метода (4), который за счет повышенного порядка аппроксимации позволяет использовать более крупный шаг с сохранением достаточной гладкости. Направление движения по кривым указано стрелками. Область сходимости по геодезическим для решения в первой четверти имеет границей контур $ADEF$. Все остальное пространство в первой четверти относится к областям несходимости. Переходы ньютоновских итераций, в том числе с регулируемой величиной шага, из областей несходимости в области сходимости не получаются. Для метода (4) такие переходы возможны (рис. 5).

На рис. 3 приведены типичные примеры «прострелов» метода Ньютона для двух начальных приближений, причем точки x_1 «вылетают» за пределы области сходимости к x_* по геодезическим. Хорошо видно, что ньютоновский шаг делается по касательной к линии l_* .

На рис. 4 приведена довольно интересная картина поведения последовательных приближений по методу Ньютона (с небольшой релаксацией - уменьшением шага, поскольку при полном шаге расходимость начинается с первой же итерации). Итерационный процесс как бы пытается, образно выражаясь, зацепиться за решение, но это ему не удается. Для сравнения приведены итерации $\tilde{x}_1, \tilde{x}_2, \dots$ по методу (4) при значениях $\alpha_k = \beta_k = t = 1$, которые сходятся за 3 шага. Метод (4) в целом имеет более высокую локальную скорость сходимости, чем метод Ньютона (1).

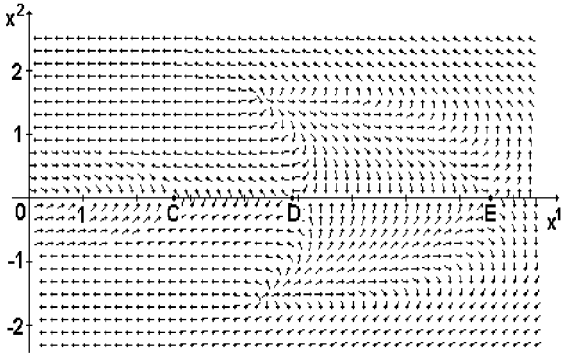


Рис. 1. Поле направлений вектора $a(x)$

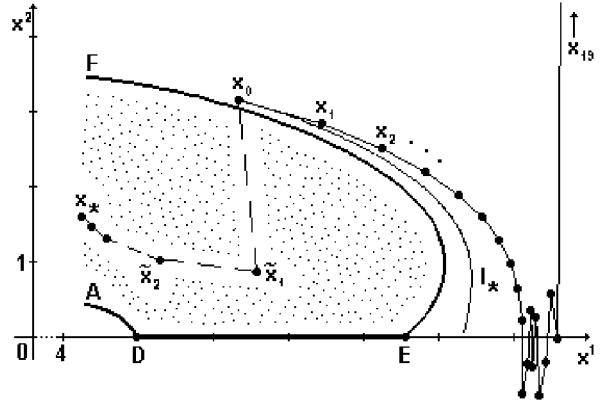


Рис. 5. Попадание на перегородку из особых точек

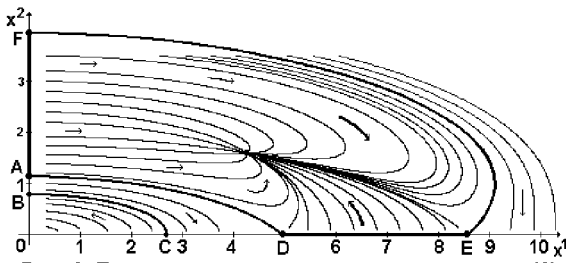


Рис. 2. Портрет интегральных кривых системы (2)

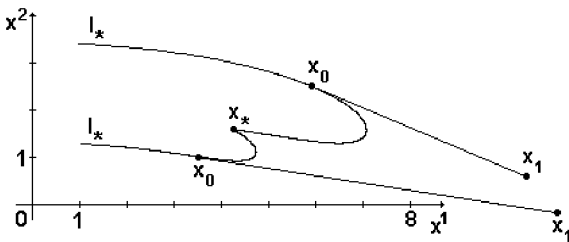


Рис. 3. «Прострелы»

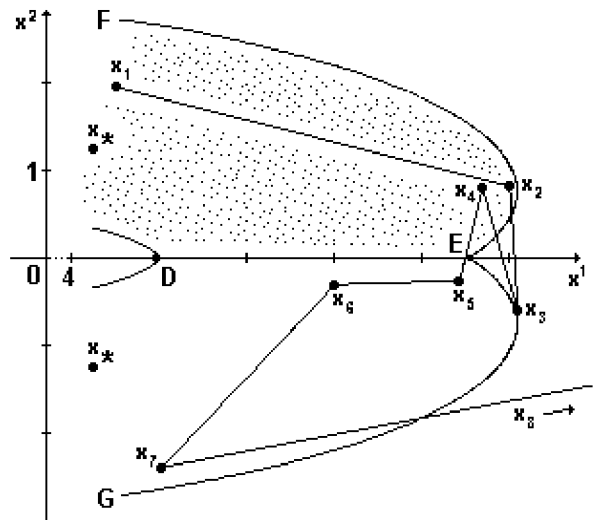


Рис. 6. Переходы из одной области сходимости по геодезическим в другую

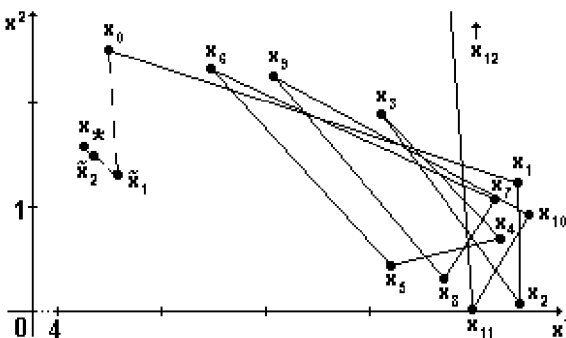


Рис. 4. Интересное поведение приближений

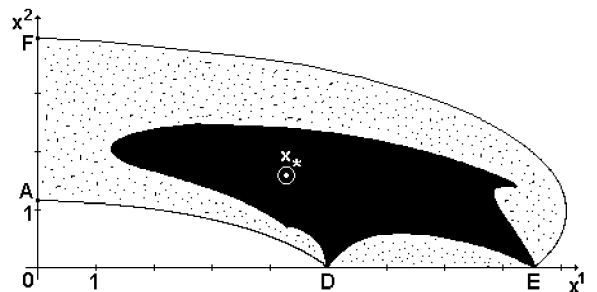


Рис. 7. Сравнение областей сходимости

На рис. 5 приведены последовательные приближения по методу Ньютона из начального приближения, находящегося вне области сходимости по геодезическим (граница области - кривая EF). Во избежание быстрой расходимости ньютоновский шаг уменьшен в десять раз. Хорошо видно, как последовательные точки, постепенно (и естественно) отклоняясь от геодезической l_* , соответствующей начальной точке, приближаются к перегородке из особых точек и после нескольких колебательных движений вверх-вниз вблизи него устремляются высоко вверх. Итерации $\tilde{x}_1, \tilde{x}_2, \dots$ по методу (4) сходятся (за 7 шагов) при уменьшении полного шага на 32 % ($t = 0,68, \alpha_k = \beta_k = 1$).

На рис. 6 приведена ситуация, когда последовательные приближения переходят из одной области сходимости по геодезическим в другую и обратно.

На рис. 7 приведены для сравнения различные области сходимости. Черным цветом закрашена область сходимости метода Ньютона (1). Белая окружность ограничивает шаровую область сходимости, которая устанавливается теорией и практически ничтожна по сравнению с истинной областью сходимости. Область сходимости по геодезическим закрашена и имеет границей контур ADEF. Видим, что в данном примере область сходимости по геодезическим относительно не на много больше области сходимости метода (1).

В заключение отметим, что основные концептуальные моменты, рассмотренные для конечного случая, сохраняют свой смысл и для нелинейных операторных уравнений в бесконечном случае.

Литература

1. Ортега, Дж. Итерационные методы решения нелинейных систем уравнений со многими неизвестными / Дж. Ортега, В. Рейнболдт. - М.: Мир, 1975. - 558 с.
2. Дэннис, Дж. Численные методы безусловной оптимизации и решения нелинейных уравнений / Дж. Дэннис, Р. Шнабель. - М.: Мир, 1988. - 440 с.
3. Математическая энциклопедия: в 5 т. - М.: Советская энциклопедия, 1984. - Т. 4. - с. 1004.
4. Постников, М.М. Риманова геометрия / М.М. Постников. - М.: Изд. «Факториал», 1998. - 495 с.
5. Спивак, М. Математический анализ на многообразиях / М. Спивак. - М.: Мир, 1968. — 320 с.
6. Понтрягин, Л.С. Обыкновенные дифференциальные уравнения / Л.С. Понтрягин. - 5 изд. - М.: Наука, 1982.-332 с.
7. Канторович, Л.В. О методе Ньютона / Л.В. Канторович // Труды мат. ин-та им. В.А. Стеклова. - 1949. - № 28. - С. 104-144.
8. Мысовских, И.П. К вопросу о сходимости метода Ньютона / И.П. Мысовских // Труды мат. ин-та им. В.А. Стеклова. - 1949. - № 28. - С. 145-147.
9. Канторович, Л.В. Принцип мажорант и метод Ньютона / Л.В. Канторович // Докл. АН СССР.- 1951.-Т. 76, № 1.-С. 17-20.
10. Михеев, СЕ. Сходимость метода Ньютона на различных классах функций / СЕ. Михеев // Вычислительные технологии. - 2005. - Т. 10, № 3. - С. 72-86.

Поступила в редакцию 21 февраля 2009 г.

GEOMETRICAL SENSE OF NEWTON METHODS

New geometrical sense of Newton methods for solving the system of nonlinear equations (in infinite-measuring case - nonlinear operational equations) found by us, clarifies completely its inner mechanism. From the point of view of application it enables to explain empirically observed effects, to get a unified characterization of the method and its modification, to get a general theorem on the problem of local convergence and to get a quite clear vision of geometrical-dynamic nature of convergence problem on the whole (both local and global). The results obtained are demonstrated on the model example.

Keywords: Newton method, Riemannian geometry, calculus of approximations, differentials equations.

Pchelintsev Mikhail Vasilyevich - Associate Professor, Snezhinsk State Academy of Physics and Technology.

Пчелинцев Михаил Васильевич - доцент, Снежинская государственная физико-техническая академия, г. Снежинск.

e-mail: pchelintsev@sfti.snz.ru

Skorkin Nikolai Andreevich - Dr. Sc. (Engineering), Professor, Branch of the South Ural State University in the city of Snezhinsk.

Скоркин Николай Андреевич - доктор технических наук, профессор, филиал Южно-Уральского государственного университета, г. Снежинск.

e-mail: n.a.scorkin@rambler.ru