Метод Ньютона и его роль в оптимизации и вычислительной математике

Б. Т. Поляк

Институт системного анализа
Российской академии наук (ИСА РАН)

Метод Ньютона является фундаментальным инструментом в численном анализе, исследовании операций, оптимизации и управлении. У него есть множество приложений к инженерным, финансовым и статистическим задачам. Его роль в оптимизации невозможно переоценить: большинство наиболее эффективных методов в линейном и нелинейном программировании строятся на его основе. Например, важнейший полиномиальный алгоритм внутренней точки в выпуклой оптимизации основан на методе Ньютона. В работе описаны базовые идеи метода, история его создания, основные теоретические результаты о сходимости, а также различные приложения. Представлены новейшие разработки в этой области и наиболее современные версии метода.

Ключевые слова. Нелинейное программирование, метод Ньютона, сходимость, глобальное поведение, методы внутренней точки.

1. Введение

В настоящей работе делается попытка обзора основных идей метода Ньютона в историческом плане, а также современных исследований и приложений. Ее цель — обсудить наиболее перспективные задачи в этой области и дать по возможности подробный список литературы.

Статья построена следующим образом. В разделе 2 описываются основные идеи метода и история его развития. Основные результаты о сходимости и скорости сходимости приведены в разделе 3. Метод Ньютона в исходном виде обладает лишь локальной сходимостью; глобальное поведение и способы достичь глобальной сходимости обсуждаются в разделах 4 и 5. Случай недоопределенных систем уравнений заслуживает специального рассмотрения (раздел 6). Первоначально метод Ньютона предназначался для решения уравнений. Однако он может быть применен для решения задач безусловной (раздел 7) и условной (раздел 8) оптимизации. Некоторые обобщения метода и направления для будущих исследований описаны в разделе 9.

2. Идея и история метода

Основная идея *метода Ньютона* очень проста — это идея линеаризации. Предположим что $F : R^1 \to R^1$ дифференцируемая функция и мы решаем уравнение

$$F(x) = 0. \tag{1}$$

Начав с точки x_0 мы можем построить линейную аппроксимацию $F(x)$ в окрестности $x_0 : F(x_0 + h) \approx F(x_0) + F'(x_0)h$ и решить получающееся линейное уравнение $F(x_0)+F'(x_0)h = 0$. Так мы приходим к итеративному методу

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \quad k = 0, 1, \dots\ . \tag{2}$$

Это и есть метод, предложенный Ньютоном в 1669. Более точно, Ньютон оперировал только с полиномами; в выражении для $F(x + h)$ он отбрасывал члены более высокого порядка по h, чем линейные. Метод иллюстрировался им на примере $F(x) = x^3 - 2x - 5 = 0$. Начальное приближение для корня было $x = 2$. Тогда $F(2 + h) = h^3 + 6h^2 + 10h - 1$, отбрасывая члены более высокого порядка Ньютон получал уравнение $10h - 1 = 0$. Таким образом следующее приближение $x = 2 + 0.1 = 2.1$ и процесс может быть продолжен из этой точки. Рис. 1 показывает, что сходимость к корню в этом случае очень быстрая. Ученик Ньютона Рафсон в 1690 г. предложил общую форму метода (2) (т. е. не предполагалось что $F(x)$ обязательно полином и использовалось понятие производной), поэтому часто говорят о *методе Ньютона—Рафсона*.

Дальнейшее развитие исследований связано с именами таких известных математиков, как Фурье, Коши и другие. Например, Фурье доказал в 1818 г. что метод сходится квадратично в окрестности корня, а Коши (1829, 1847) предложил многомерное обобщение метода (2) и использовал метод для доказательства существования решения уравнения. Существенный вклад внесли Файн [17] и Беннет [3]; их статьи опубликованы в одном томе журнала *Proceedings of National Academy of Sciences USA* в 1916 г. Файн доказал сходимость в n-мерном случае без предположения о существовании решения. Беннет обобщил результат на бесконечномерный случай; это поразительно, так как основы функционального анализа на то время не были еще заложены. В 1948 г., Л. В. Канторович опубликовал важную

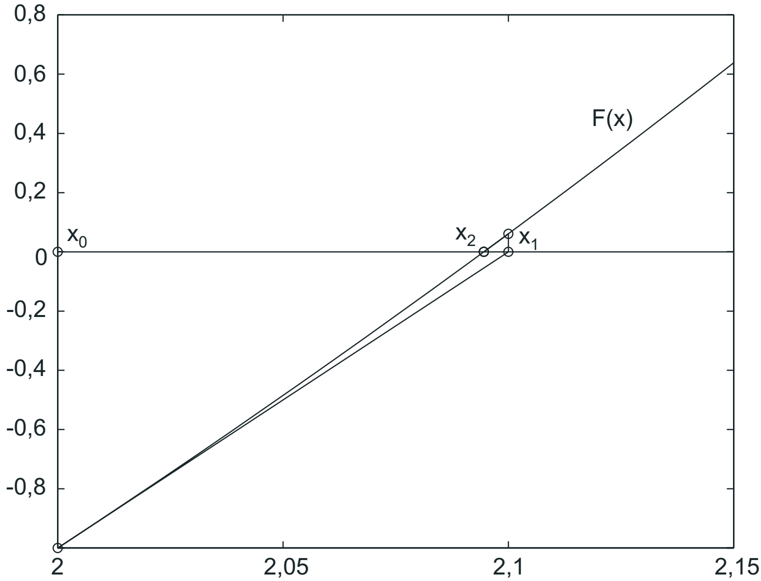

Рис. 1. Метод Ньютона

работу [25] где было дано обобщение метода для уравнений в функциональных пространства (*метод Ньютона—Канторовича*). Этот результат включен также в обзорную статью [26]; дальнейшие работы Канторовича по этой тематике — [27–31].

Основные результаты по методу Ньютона и его приложениям можно найти в монографиях Островского [47], Ортеги и Рейнболдта [48], Канторовича и Акилова [32, 33]. Современная библиография приведена в книге [13, 50, 60], обзорной статье [67] и на специальном веб-сайте, посвященном методу Ньютона [41].

3. Сходимость

Канторович [25] рассматривает то же уравнение, что и (1):

$$F(x) = 0, \tag{3}$$

но теперь $F : X \to Y$, где X, Y банаховы пространства. Предлагаемый метод записывается как и (2)

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \quad k = 0, 1, \dots, \tag{4}$$

где $F'(x_k)$ производная (по Фреше) нелинейного оператора $F(x)$ в точке x_k, а $F'(x_k)^{-1}$ — обратная к ней. Основной результат о сходимости из [25] формулируется так.

Теорема 1. *Пусть F определена и дважды непрерывно дифференцируема на шаре $B = \{x : ||x - x_0|| \leqslant r\}$, линейный оператор $F'(x_0)$ обратим, $||F'(x_0)^{-1}F(x_0)|| \leqslant \eta$, $||F'(x_0)^{-1}F''(x)|| \leqslant K, x \in B$ и*

$$h = K\eta < \frac{1}{2}, \quad r \geqslant \frac{1 - \sqrt{1 - 2h}}{h}\eta. \tag{5}$$

Тогда уравнение (3) *имеет решение $x^* \in B$, метод* (4) *определен и сходится к x^* с квадратичной скоростью*:

$$||x_k - x^*|| \leqslant \frac{\eta}{h2^k}(2h)^{2^k}. \tag{6}$$

Доказательство этой теоремы достаточно просто; главная новизна результата Канторовича заключается не в деталях доказательства, но в общности постановки задачи и использовании соответствующей техники функционального анализа. До работ Канторовича [25–27] не было понимания связи вычислительной математики и функционального анализа (стоит обратить внимание на заголовок [26]). Другой особенностью теоремы является отсутствие предположения о существовании решения — поэтому она может рассматриваться и как теорема существования.

Эти свойства результата Канторовича привели к его широкому использованию. Многочисленные нелинейные задачи — нелинейные интегральные, дифференциальные уравнения (обыкновенные или с частыми производными), вариационные задачи — все они могут рассматриваться в рамках уравнения (3) и метод (4) может быть применен. Различные примеры таких приложений приведены в книгах [32, 33]. Более того, метод Ньютона—Канторовича как способ доказательства существования решения был вскоре использован Колмогоровым, Арнольдом и Мозером [1] при построении знаменитой *КАМ-теории* в механике. Многие классические результаты в функциональном анализе обоснованы с помощью теорем типа вышеприведенной; типичным примером является *теорема Люстерника* о касательном подпространстве, см. исходную работу [38] или современное исследование [22].

Позже Канторович [28, 29] получи другое доказательство теоремы 1 и ее модификаций, основанное на *методе мажорант*. Его идея заключается в сравнении процесса (4) со скалярными итерациями, которые в некотором смысле мажорируют процесс и обладают свойствами сходимости. Этот подход обладает большей гибкостью; исходное доказательство есть его частный случай для квадратичных мажорант.

Существует множество вариантов теоремы 1, отличающихся предположениями и результатами. Приведем один из них — теорему Мысовских [43].

Теорема 2. *Пусть F определена и дважды непрерывно дифференцируема на шаре $B = \{x : ||x - x_0|| \leqslant r\}$, линейный оператор $F'(x)$ обратим на B и $||F'(x)^{-1}|| \leqslant \beta$, $||F''(x)|| \leqslant K$, $x \in B$, $||F(x_0)|| \leqslant \eta$, а*

$$h = K\beta^2\eta < 2, \quad r \geqslant \beta\eta \sum_{n=0}^{\infty} (h/2)^{2^n - 1}. \tag{7}$$

Тогда уравнение (3) *имеет решение $x^* \in B$ и метод* (4) *сходится к x^* с квадратичной скоростью*:

$$||x_k - x^*|| \leqslant \frac{\beta\eta(h/2)^{2^k - 1}}{1 - (h/2)^{2^k}}. \tag{8}$$

Разница с теоремой 1 — в предположении об обратимости $F'(x)$ на шаре B, а не только в начальной точке x_0 и более слабые требования к h ($h < 2$ вместо $h < 1/2$). Родственные результаты можно найти в [10, 32, 33, 35, 47, 48].

Метод Ньютона (4) требует вычисления производных $F'(x)$ и решения линейного уравнения на каждой итерации; это может быть достаточно трудоемко. *Модифицированный метод Ньютона*

$$x_{k+1} = x_k - F'(x_0)^{-1}F(x_k), \quad k = 0, 1, \ldots \tag{9}$$

не обладает этим недостатком — матрица $F'(x_0)$ вычисляется и обращается только в начальной точке. Метод сходится при тех же условиях, что и (4), однако плата за упрощение высока — метод (9) сходится с линейной, а не квадратичной скоростью. Возможны и некоторые компромиссы — например, можно пересчитывать производные каждые несколько итераций.

4. Глобальное поведение

Критическими условиями сходимости являются (5) или (7). Они означают, что в начальной точке x_0 функция $||F(x_0)||$ должна быть достаточно мала, т. е. x_0 должна быть достаточно близка к решению. Таким образом, метод Ньютона сходится локально. Очень простые одномерные примеры показывают отсутствие глобальной сходимости даже для гладких монотонных $F(x)$. Есть много способов модифицировать метод, чтобы добиться глобальной сходимости (мы обсудим их позже), но представляет интерес глобальное поведение исходного метода. Конечно есть ряд простых ситуаций — например, когда для некоторой окрестности S решения условие $x_0 \in S$ влечет сходимость к решению (такая окрестность называется *областью притяжения*), а траектории, начинающиеся вне S, расходятся (уходят в бесконечность). Однако в случае неединственности решения область притяжения каждого решения может быть весьма сложна и даже иметь *фрактальную структуру*. Это было обнаружено Кэли еще в 1879 г.

Рассмотрим пример Кэли: решить уравнение $z^3 = 1$ методом Ньютона. Мы берем $F(z) = z^3 - 1$ и применяем метод (2):

$$z_{k+1} = z_k - \frac{z_k^3 - 1}{3z_k^2} = \frac{2z_k}{3} + \frac{1}{3z_k^2}.$$

Нужно отметить, что до сих пор мы формулировали метод Ньютона в вещественном пространстве, но он равным образом применим и для комплексных переменных; в рассматриваемом примере мы берем $z_k \in \mathbf{C}$. Уравнение имеет три корня

$$z_1^* = 1, \quad z_{2,3}^* = -\frac{1}{2} \pm i\frac{\sqrt{3}}{2}$$

и естественно ожидать, что вся плоскость $\mathbf{C}$ разобьется на три области притяжения

$$S_m = \{z_0 : z_k \to z_m^*\}, \quad m = 1, 2, 3,$$

расположенные вокруг каждого из корней. Однако в действительности ситуация гораздо более сложная. Прежде всего, есть лишь одна точка $z = 0$ где метод не определен. У нее есть три прообраза — точки z_0 такие что $z_1 = 0$, это $-\rho, \rho\ (1/2 \pm i\sqrt{3}/2)$, где $\rho = 1/\sqrt[3]{2}$. Вновь каждая из этих точек имеет три прообраза и так далее. Таким образом, существуют 3^k точек z_0, которые после k итераций отображаются в точку $z_k = 0$, и они порождают счетное множество начальных точек, для которых метод не определен на какой-либо итерации

$$S_0 = \{z_0 : z_k = 0 \quad \text{для некоторого } k\}.$$

Можно показать, что для всех остальных точек $z_0 \notin S_0$ метод сойдется к одному из корней (заметим, что если $|z_k| > 1$, тогда $|z_{k+1}| < |z_k|$) и поэтому

$$\mathbf{C} = \bigcup_{m=0}^{3} S_m.$$

Множества S_m имеют фрактальную структуру: S_0 является границей S_m, $m = 1, 2, 3$ и в любой окрестности любой точки $z \in S_0$ найдутся точки из S_m, $m = 1, 2, 3$. Множество S_1 показано на рис. 2, множество S_0 — на рис. 3. Множества начальных точек, из которых нет сходимости (подобные S_0) для итераций общих дробно-рациональных функций были

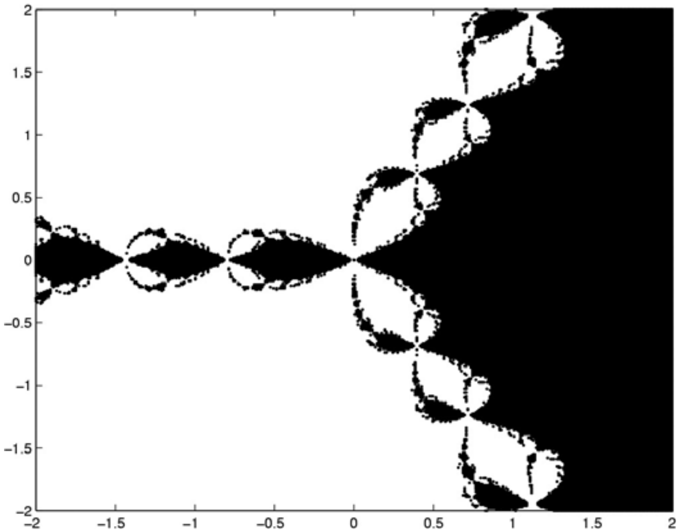

Рис. 2. Область притяжения для $z^* = 1$

изучены Жюлиа [24] и теперь называются *множествами Жюлиа*. Разнообразные примеры, порожденные методом Ньютона, могут быть найдены в книгах по фракталам [2, 39], статьях [12, 49] и в Интернете [14, 23] (мы упоминаем лишь часть из доступных источников). Некоторые из этих примеров демонстрируют гораздо более сложное поведение итераций, чем в примере Кэли. Например, можно встретить периодическое или хаотическое поведение траекторий итераций.

5. Преодоление локальности метода

В примере Кэли метод Ньютона сходится для почти всех начальных точек (исключением является счетное множество S_0); сложная природа областей притяжения вызывается существованием нескольких корней. Однако если (1) имеет единственный корень, то метод обычно сходится лишь локально, т. е. для хороших начальных приближений. Например, возьмем $F(x) = \operatorname{arctg} x$, эта функция гладкая, монотонная и имеет единственный корень $x^* = 0$. Легко проверить, что метод (2) сходится если и только если $|x_0| < x^* (x^* > 0$ — это корень уравнения $2x = (1 + x^2) \operatorname{arctg} x)$, для $|x_0| = x^*$ итерации являются периодическими $x_0 = -x_1 = x_2 = -x_3 \ldots$, а для $|x_0| > x^*$ итерации расходятся: $|x_k| \to \infty$.

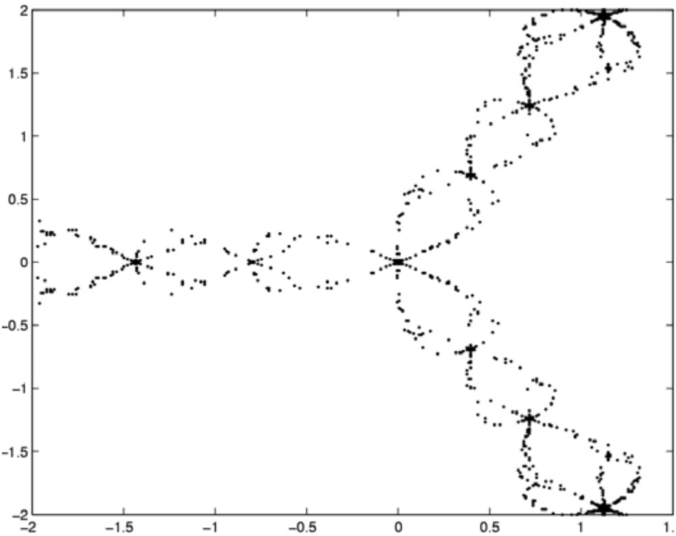

Рис. 3. Точки, из которых метод не сходится

Есть несколько способов модифицировать метод Ньютона чтобы добиться глобальной сходимости. Первый заключается в регулировке длины шага, предотвращающей большие шаги; это так называемый *демпфированный метод Ньютона*:

$$x_{k+1} = x_k - \alpha_k F'(x_k)^{-1} F(x_k), \quad k = 0, 1, \dots, \tag{10}$$

где шаг $0 < \alpha_k \leqslant 1$ выбирается так, чтобы добиться монотонного убывания $||F(x)||$, иначе говоря, шаг дробится, чтобы обеспечить условие $||F(x_{k+1})|| < ||F(x_k)||$. Мы обсудим конкретные алгоритмы выбора α_k позже для задач минимизации. Главная цель при построении таких алгоритмов — добиться баланса между скоростью сходимости и глобальной сходимостью, т. е. нужно выбирать $\alpha_k < 1$ когда x_k вне области притяжения «чистого» метода Ньютона и переключаться на $\alpha_k = 1$ внутри этой области.

Второй подход — это *метод Левенберга—Марквардта* [36, 40]:

$$x_{k+1} = x_k - (\alpha_k I + F'(x_k))^{-1} F(x_k), \quad k = 0, 1, \dots . \tag{11}$$

Для $\alpha_k = 0$ метод превращается в ньютоновский, а для $\alpha_k \gg 1$ он близок к градиентному с малым шагом, последний обладает свойством глобальной сходимости. Существуют различные стратегии настройки α_k, они описаны, например, в [48]. Метод (11) работает даже тогда, когда метод Ньютона не определен — при вырождении оператора $F'(x_k)$. Как мы увидим, в задачах минимизации метод (11) очень перспективен, так как в нем не предполагается, что матрица $F'(x_k)$ положительно определена.

Еще один способ модифицировать метод Ньютона для предотвращения больших шагов — ввести так называемые *доверительные области*, где линейная аппроксимация обоснована. Такой подход (восходящий к [19] и подробно развитый в [11]) будет обсуждаться ниже, применительно к задачам минимизации.

6. Недоопределенные системы

В вышеприведенном анализе метода Ньютона предполагалось, что оператор $F'(x_k)$ обратим. Однако есть ситуации, когда это заведомо не так. Например, предположим что $F : \mathbf{R}^n \to \mathbf{R}^m$, где $m < n$. Иначе говоря, мы решаем недоопределенную систему m уравнений с n>m переменными; конечно, прямоугольная матрица $F'(x_k)$ не имеет обратной. Однако обобщение метода Ньютона на этот случай может быть получено, это было сделано Грейвсом [21], который использовал этот метод для получения теорем существования. Мы приводим результат из [51], где даны более точные оценки.

Теорема 1. *Пусть $F : X \to Y$ определена и дифференцируема на шаре $B = \{x : ||x - x_0|| \leqslant r\}$, ее производная удовлетворяет условию Липшица на B*

$$||F'(x) - F'(z)|| \leqslant L||x - z||, \quad x, z \in B,$$

$F'(x)$ отображает X на Y и имеет место оценка:

$$||F'(x)^* y|| \geqslant \mu||y|| \quad \forall\, x \in B, \quad y \in Y^*, \tag{12}$$

с $\mu > 0$ (звездочка означает сопряжение). Введем функцию

$$H_n(t) = \sum_{k=n}^{\infty} t^{2^k}$$

и предположим что

$$h = \frac{L\mu^2||F(x_0)||}{2} < 1, \quad \rho = \frac{2H_0(h)}{L\mu} \leqslant r. \tag{13}$$

Тогда метод

$$x_{k+1} = x_k - y_k, \quad F'(x_k)y_k = F(x_k), \quad ||y_k|| \leqslant (1/\mu)||F(x_k)||, \\ k = 0, 1, \ldots \tag{14}$$

определен и сходится к решению x^* уравнения $F(x) = 0, ||x^* - x_0|| \leqslant \rho$ со скоростью

$$||x_k - x^*|| \leqslant \frac{2H_k(h)}{L\mu}. \tag{15}$$

Таким образом, на каждой итерации нужно решать линейное уравнение $F'(x_k)y = F(x_k)$, где линейный оператор $F'(x_k)$ вообще говоря не имеет обратного; однако он отображает X на Y и это уравнение имеет решение (может быть не единственное). Среди решений существует такое y_k что $||y_k|| \leqslant (1/\mu)||F(x_k)||$; именно оно и используется в методе. В конечномерном случае ($X = \mathbf{R}^n$, $Y = \mathbf{R}^m$, $n > m$) такое решение дается формулой $y_k = F'(x_k)^+F(x_k)$, где A^+ обозначает псевдообратную к матрице A.

Рассмотрим приложение теоремы 1 к выпуклому анализу. Справедлив следующий результат о выпуклости образа малого шара в гильбертовом пространстве при нелинейном отображении [52].

Теорема 2. *Пусть X,Y гильбертовы пространства, $F\colon X \to Y$ определена и дифференцируема на шаре $B = \{x : ||x - a|| \leqslant r\}$, ее производная удовлетворяет условию Липшица на B*

$$||F'(x) - F'(z)|| \leqslant L||x - z||, \quad x, z \in B,$$

$F'(a)$ отображает X на Y и

$$||F'(a)^*y|| \geqslant \mu||y|| \quad \forall\, y \in Y, \tag{16}$$

где $\mu > 0$ и $r < \mu/(2L)$. Тогда образ шара B при отображении F выпуклый, т. е. множество $S = \{F(x) : x \in B\}$ выпукло в Y.

Эта теорема нашла многочисленные применения в оптимизации, [52], линейной алгебре [53], оптимальном управлении [54]. Например, псевдоспектр $n \times n$ матрицы (множество всех собственных чисел возмущенной матрицы с возмущениями, ограниченными во фробениусовой норме) оказывается объединением n выпуклых множеств на комплексной плоскости, при условии что все собственные числа исходной матрицы различны, а возмущения достаточно малы. Другой результат гарантирует выпуклость достижимого множества нелинейной системы при управлении, ограниченном в L_2 норме [54].

7. Безусловная оптимизация

Рассмотрим простейшую задачу безусловной минимизации в гильбертовом пространстве H:

$$\min f(x), \quad x \in H. \tag{17}$$

Предполагая что f дважды дифференцируема, мы можем придти к методу Ньютона двумя способами.

Во-первых, необходимое (а при выпуклости f и достаточное) условие экстремума имеет вид

$$\nabla f(x) = 0,$$

т. е. мы должны решить уравнение (3) с $F(x) = \nabla f(x)$. Применяя метод Ньютона для этого уравнения мы приходим к методу Ньютона для минимизации:

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k), \quad k = 0, 1, \dots, \tag{18}$$

где $\nabla^2 f$ означает вторую производную (матрицу Гессе в конечномерном случае).

Во-вторых, мы можем аппроксимировать $f(x)$ в окрестности точки x_k тремя членами ряда Тейлора:

$$f(x_k + h) \approx f_k(h) = f(x_k) + (\nabla f(x_k), h) + \frac{1}{2}(\nabla^2 f(x_k)h, h).$$

Тогда, минимизируя квадратичную функцию $f_k(h)$ мы получаем тот же метод (18). Обе интерпретации будут использоваться при построении более общих методов оптимизации.

Теоремы о сходимости метода Ньютона для уравнений могут быть немедленно перенесены на задачу безусловной минимизации (заменой $F(x)$ на $\nabla f(x)$ и $F'(x)$ на $\nabla^2 f(x)$). Главная черта метода — быстрая локальная сходимость — сохраняется без изменений. Однако, есть и некоторые особенности. Наиболее существенная: метод Ньютона в исходном виде не различает точки минимума, максимума и седловые. Если начать итерации из окрестности невырожденной критической точки (т. е. точки x^* с $\nabla f(x^*) = 0, \nabla^2 f(x^*)$ обратимой) метод сходится к ней независимо от знакоопределенности $\nabla^2 f(x^*)$.

Есть несколько способов превратить метод в глобально сходящийся, они аналогичны модификациям, описанным в разделе 5 (демпфированный метод Ньютона, методы Левенберга—Марквардта и доверительной области). Мы рассмотрим важный способ, предложенный Нестеровым и Немировским [44], для него может быть оценена *сложность* (число итераций для достижения заданной точности). Нестеров и Немировский ввели класс *самосогласованных* функций. Это три раза дифференцируемые выпуклые функции определенные на выпуклом множестве $D \subset \mathbf{R}^n$, удовлетворяющие свойству

$$|\nabla^3 f(x)(h, h, h)| \leqslant 2(\nabla^2 f(x)h, h)^{3/2} \quad \forall\, x \in D, \quad h \in \mathbf{R}^n.$$

Приведенная формула включает вторые и третьи производные f и их действие на вектор $h \in \mathbf{R}^n$. В более простом виде она может быть записана с помощью производных скалярной функции $\varphi(t) = f(x + th)$:

$$|\varphi'''(0)| \leqslant 2(\varphi''(0))^{3/2}$$

для всех $x \in D$, $h \in \mathbf{R}^n$. Например, $f(x) = -\ln x$, $x > 0$, $x \in \mathbf{R}^1$ — самосогласованная функция, а $f(x) = 1/x$, $x > 0$ — нет. Для $x_k \in D$ определим *ньютоновский декремент*

$$\delta_k = \sqrt{\nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k)}.$$

Тогда *вариант Нестерова—Немировского* для демпфированного метода Ньютона выглядит так:

$$x_{k+1} = x_k - \alpha_k (\nabla^2 f(x_k))^{-1} \nabla f(x_k), \quad k = 0, 1, \ldots, \tag{19}$$

$$\alpha_k = 1, \qquad \text{если} \quad \delta_k \leqslant \frac{1}{4}, \tag{20}$$

$$\alpha_k = \frac{1}{1+\delta_k}, \quad \text{если} \quad \delta_k > \frac{1}{4}. \tag{21}$$

Теорема 1. *Если $f(x)$ самосогласованная и $f(x) \geqslant f^* \ \forall\, x \in D$ то для метода* (19)–(21) *с $x_0 \in D, \varepsilon > 0$ имеет место оценка*

$$f(x_k) - f^* \leqslant \varepsilon$$

для

$$k = c_1 + c_2 \log\, \log\left(\frac{1}{\varepsilon}\right) + c_3(f(x_0) - f^*). \tag{22}$$

Здесь c_1, c_2, c_3 — некоторые абсолютные константы.

Идея доказательства достаточно прозрачна. Процедура минимизации состоит из двух этапов. На этапе 1 применяется метод (19), (21), и при этом функция убывает монотонно: $f(x_{k+1}) \leqslant f(x_k) - \gamma$, где γ положительная константа. Очевидным образом этот этап заканчивается за конечное число итераций (поскольку $f(x)$ ограничена снизу). На этапе 2 работает метод Ньютона (19), (20) и он сходится с квадратичной скоростью: $2\delta_{k+1} \leqslant (2\delta_k)^2$. Заметим что ньютоновский декремент является удобным инструментом для настройки метода: в приведенной теореме нет констант типа L, K, η, фигурирующих в теоремах 1–3.

Для оценки (22) можно выписать более явный вид. Заметим что $\log(\log(1/\varepsilon))$ невелико если ε мало; для всех разумных ε справедлива оценка

$$k = 5 + 11(f(x_0) - f^*).$$

Более того, результаты вычислений для многих примеров [8] подтверждают что число шагов в методе (19)–(21) не превосходит:

$$k = 5 + 0{,}6(f(x_0) - f^*).$$

Серьезным ограничением в вышеприведенном анализе сложности является предположение о выпуклости. Недавно [45] была предложена другая модификация метода Ньютона, для которой удается добиться глобальной сходимости и получить оценки числа шагов для общих гладких функций. Предположим что $f \in C^{2,1}$, т. е. что f дважды дифференцируема на $\mathbf{R}^n$ и ее вторая производная удовлетворяет условию Липшица с константой L на множестве $\{x : f(x) \leqslant f(x_0)\}$. В [45] рассматривается метод:

$$x_{k+1} = \arg\min_x f_k(x), \quad h = x - x_k,$$
$$f_k(x) = f(x_k) + (\nabla f(x_k), h) + \frac{1}{2}(\nabla^2 f(x_k)h, h) + \frac{L}{6}||h||^3. \tag{23}$$

Таким образом, на каждой итерации мы решаем задачу минимизации с тем же квадратичным членом, что и в методе Ньютона, но с регуляризацией с помощью кубического члена. Такая задача выглядит трудной и невыпуклой, однако она может быть сведена к одномерной выпуклой проблеме (см. подробности в [45]). Удивительным образом, такой подход имеет преимущества перед «чистым» методом Ньютона. Во первых, он сходится глобально, из любой начальной точки. Во вторых, он не сходится к точкам максимума или точкам перегиба, а лишь к точкам минимума (возможно локальным). В третьих, можно оценить его сложность для различных классов функций. Наконец, возможен реализуемый вариант метода (23), где знание константы L не требуется.

8. Оптимизация при наличии ограничений

Мы начнем с некоторых частных случаев задач оптимизации с ограничениями.

Первый — оптимизация при *простых ограничениях*:

$$\min f(x), \quad x \in Q, \tag{24}$$

где множество Q в гильбертовом пространстве H является «простым» в том смысле, что задача (24) с квадратичной функцией f может быть решена явно. Например, Q может быть шаром, линейным подпространством и т. д. Обобщение метода Ньютона на этот случай основывается на второй интерпретации этого метода для задач безусловной минимизации:

$$x_{k+1} = \arg\min_{x \in Q} f_k(x),$$
$$f_k(x) = (\nabla f(x_k), x - x_k) + \frac{1}{2}\big(\nabla^2 f(x_k)(x - x_k), (x - x_k)\big). \tag{25}$$

Метод сходится при тех же условиях, что и в задаче без ограничений: если f выпукла и дважды дифференцируема с липшицевой второй производной на Q, множество Q выпукло и замкнуто, f достигает минимума на Q в точке x^*, $\nabla^2 f(x^*) > 0$, тогда последовательность (25) локально сходится к x^* с квадратичной скоростью. Этот результат был получен в [37], дальнейшие подробности и примеры можно найти в [6, 55].

Другой простой случай — оптимизация при *ограничениях типа равенств*:

$$\min f(x), \quad g(x) = 0, \tag{26}$$

где $f : X \to \mathbf{R}^1$, $g : X \to Y$; X, Y — гильбертовы пространства. Если решение x^* задачи существует и является регулярной точкой ($g'(x^*)$ отображает X на Y), тогда найдутся множители Лагранжа y^* такие, что пара x^*, y^* является стационарной точкой функции Лагранжа

$$L(x, y) = f(x) + (y, g(x)),$$

т. е. корнем уравнения

$$L'_x(x, y) = 0, \quad L'_y(x, y) = 0. \tag{27}$$

Поэтому мы можем применить метод Ньютона для его решения. При естественных предположениях он сходится локально к x^*, y^*, см. точные формулировки в [5, 6, 56].Там же могут быть найдены и различные модификации метода.

Есть и другие способы решения (26), которые не используют двойственных переменных y. Исторически первый метод для отыскания наибольшего собственного значения матрицы $A = A^T$ путем сведения к задаче условной минимизации:

$$\max(Ax, x), \quad ||x||^2 = 1$$

был предложен Рэлеем (1899). Позже Канторович [27] обосновал «чистый» метод Ньютона для этой же задачи. Другая ситуация, допускающая упрощение метода, это случай линейных ограничений $g(x) = Cx - d$. Тогда метод такой же, как и при безусловной минимизации f на аффинном подпространстве $Cx = d$.

Перейдем к применениям метода Ньютона для выпуклых задач оптимизации — той области, где этот метод играет ключевую роль при конструировании эффективных алгоритмов. Основная схема *метода внутренних точек* выглядит следующим образом [44, 66]. Для *задачи выпуклой оптимизации*

$$\min f(x), \quad x \in Q \tag{28}$$

с выпуклой самосогласованной f и выпуклой $Q \in \mathbf{R}^n$ строится самосогласованный *барьер* $F(x)$, определенный на внутренности Q и растущий к бесконечности, когда точка приближается к границе Q:

$$F : \mathrm{int}Q \to \mathbf{R}^1, \quad F(x) \to \infty \quad \text{для} \quad x \to \partial Q.$$

Такие барьеры существуют для многих ограничений, например, если $Q = \{x : x \geqslant 0\}$, тогда годится логарифмический барьер:

$$F(x) = \sum_{i=1}^{n} -\ln x_i.$$

Используя такие барьеры, можно построить штрафную функцию

$$f_k(x) = t_k f(x) + F(x),$$

зависящую от параметра $t_k > 0$. При естественных предположениях можно доказать, что $f_k(x)$ имеет точку минимума x_k^* на Q и $f(x_k^*) \to f^*$ (минимальное значение в (28)) когда $t_k \to \infty$ (это так называемый *центральный путь*). Однако нет необходимости находить x_k^*, точно, достаточно сделать один шаг метода (19)–(21) и затем изменить t_k. Существуют способы настройки параметров α_k, t_k, при которых метод имеет полиномиальную сложность, см. [4, 8, 44, 46]. Теоретический результат о полиномиальной сложности очень важен, однако результаты численных экспериментов не менее существенны. Они демонстрируют очень высокую эффективность метода для широкого круга оптимизационных задач — от *линейного программирования* до *полуопределенного программирования*. В частности, метод является успешным конкурентом симплекс-метода для линейного программирования.

Близкие идеи используются в невыпуклых задачах оптимизации, где так называемое *последовательное квадратичное программирование* (*SQP*) является основой для эффективных алгоритмов минимизации, см. [7, 11, 34, 61, 66].

9. Некоторые обобщения

Метод Ньютона допускает многочисленные обобщения и приложения, мы рассмотрим лишь некоторые из них.

- **Ослабленные предположения о гладкости.** Метод Ньютона может быть применен в ситуации, когда уравнения (или минимизируемая функция) являются недостаточно гладкими. Простейший пример — решение уравнения (1) при кусочно-линейной F. Тогда, если решение существует, метод (2) (естественно обобщенный) находит его за конечное число итераций.

  Для общего случая существуют различные обобщения метода Ньютона, начиная с концепции полугладкой функции Мифлина [42]. Многие исследователи работали в этом направлении, см. [9, 34, 58, 62] и приведенные там ссылки; книга [34] содержит много идей и результатов. Другой подход связан со сглаживанием уравнений, примером служат работы [9, 18].

- **Кратные корни.** Мы исследовали метод Ньютона в окрестности простого корня x^*. В случае кратного корня метод Ньютона либо сохраняет сходимость, но теряет высокую скорость, либо расходится. Существует модификация метода, восходящая к Шредеру (1870), которая обладает квадратичной скоростью сходимости. В одномерном случае для (1) она имеет вид:

$$x_{k+1} = x_k - pF'(x_k)^{-1}F(x_k), \quad k = 0, 1, \dots, \tag{29}$$

где p — кратность корня. К сожалению, здесь нужно заранее знать кратность корня, кроме того в многомерном случае ситуация более сложна.

- **Методы более высокого порядка**. Скорость сходимости метод Ньютона достаточно высока; однако иногда предлагают методы с еще более высокой скоростью сходимости. Такие попытки не кажутся актуальными (в действительности лишь несколько итераций метода Ньютона требуются для достижения высокой точности, если мы находимся в области притяжения метода), поэтому нет нужды вводить усложнения (обычно за счет использования производных более высоких порядков) для ускорения сходимости.

- **Непрерывный вариант**. Вместо дискретных итераций метода Ньютона (4) можно использовать его непрерывный аналог

$$\dot{x}(t) = -F'(x(t))^{-1}F(x(t)). \tag{30}$$

Простейшая разностная аппроксимация (30) ведет к демпфированному методу Ньютона (10) с постоянным шагом $\alpha_k = \alpha$. Последний может обладать глобальной сходимостью, поэтому можно ожидать этого же свойства для непрерывного аналога. Действительно, глобальная сходимость (30) была доказана Смейлом [63], а скорость сходимости исследована в [59]. Конечно, практическая ценность непрерывного метода спорна, так как при его практической реализации приходится прибегать к дискретизации.

- **Условия в одной точке.** Все результаты о сходимости метода Ньютона получены в предположении о выполнении тех или иных условий на некотором шаре. В противоположность этому Смейл [64] предложил теорему о сходимости, в которой используются условия только в начальной точке. Однако эти условия включают оценки для всех производных.

- **Решение задач дополнительности и равновесных задач**. Есть много работ, где метод Ньютона применяется к проблемам, которые не сводятся к решению уравнений или оптимизации. Примерами могут служить задачи дополнительности и равновесные задачи. Результаты в этом направлении можно найти в [16, 34].
- **Проблемы реализации**. Здесь нет возможности обсудить проблемы численной реализации различных вариантов метода Ньютона; между тем эти вопросы очень важны. Много информации в этом направлении можно почерпнуть из книги [13], а сами программы можно загрузить с [15]. Для выпуклой оптимизации эти проблемы обсуждаются в [8].
- **Сложность**. Существуют глубокие результаты о сложности основных задач вычислительной математики (таких как отыскание всех корней полинома). Многие из них тесно связаны с методом Ньютона (он часто оказывается оптимальным в некотором смысле методом, пригодным для решения этих задач). Мы рекомендуем читателю, интересующемуся данной проблематикой, обратиться к [65].

Литература

1. *Арнольд В. И.* Малые знаменатели и проблема устойчивости в классической и небесной механике // Успехи математических наук. 1963. 18 (6). С. 91–192.
2. *Barnsley M.* Fractals Everywhere. London: Academic Press, 1993.
3. *Bennet A. A.* Newton's method in general analysis // Proceedings of National Academy of Sciences USA. 1916. 2 (10). С. 592–598.
4. *Ben-Tal A.*, *Nemirovski A.* Lectures on Modern Convex Optimization: Analysis, Algorithms and Engineering Applications. SIAM, Philadelphia, 2001.
5. *Bertsekas D. P.* Constrained Optimization and Lagrange Multiplier Method. N. Y.: Academic Press, 1982.
6. *Bertsekas D. P.* Nonlinear Programming. Athena Scientific, Belmont, 1999.
7. *Biegler L. T.*, *Grossmann I. E.* Part I: Retrospective on Optimization, Part II: Future Perspective on Optimization, preprints. Chem. Eng. Dept., Carnegie-Mellon Univ., 2004.
8. *Boyd S.*, *Vandenberghe L.* Convex Optimization. Cambridge: Cambridge Univ. Press, 2004.
9. *Chen X.*, *Nashed Z.*, *Qi L.* Smoothing methods and semismooth methods for nondifferentiable operator equations // SIAM Journal on Numerical Analysis. 2000. № 38. P. 1200–1216.
10. *Коллатц Л.* Функциональный анализ и вычислительная математика. М.: Мир, 1969.
11. *Conn A. B.*, *Gould N. I. M.*, *Toint Ph. L.* Trust Region Methods. SIAM, Philadelphia, 2000.
12. *Curry J. H. Garnett L. Sullivan D.* On the iteration of a rational function: computer experiments with Newton's method // Communications on Mathematical Physics. 1983. №. 91. P. 267–277.
13. *Deulfhard P.* Newton Methods for Nonlinear Problems: Affine Invariant and Adaptive Algorithms. Berlin: Springer, 2004.
14. *Dickau R. M.* Newton's method (http//mathforum.org/advanced/robertd/newnewton.html).
15. Electronic library ZIB (http://www.zib.de/SciSoft/NewtonLib).
16. *Facchinei F. Pang J. S.* Finite-Dimensional Variational Inequalities and Complementarity Problems. New-York: Springer, 2003.

17. *Fine H.* On Newton's method of approximation // Proceedings of National Academy of Sciences USA. 1916. 2 (9). P. 546–552.
18. *Fukushima M.*, *Qi L.* Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods. Dordrecht: Kluwer, 1999.
19. *Goldfeld S.*, *Quandt R.*, *Trotter H.* Maximization by quadratic hill climbing // Econometrica. 1966. № 34. P. 541–551.
20. *Gill P. E.*, *Murray W.*, *Saunders M. A.* SNOPT: an SQP algorithm for large-scale constrained optimization // SIAM Journal on Optimization. 2002. № 12. P. 979–1006.
21. *Graves L. M.* Some mapping theorems // Duke Mathematical Journal. 1950. № 17. P. 111–114.
22. *Ioffe A. D.* On the local surjection property // Nonlinear Analysis: Theory, Methods and Applications. 1987. 11 (5). P. 565–592.
23. *Joyce D. E.* Newton basins (http//aleph0.clarku.edu/ djoyce/newton/newton.html).
24. *Julia G.* Sur l'iteration des fonctions rationelles // Journal de Mathematiques Pure et Applique. 1918. № 8. P. 47–245.
25. *Канторович Л. В.* О методе Ньютона для функциональных уравнений // Доклады АН СССР. 1948. 59 (7). С. 1237–1240.
26. *Канторович Л. В.* Функциональный анализ и прикладная математика // Успехи математических наук. 1948. 3 (1). С. 89–185.
27. *Канторович Л. В.* О методе Ньютона // Труды МИАН. 1949. № 28. С. 104–144.
28. *Канторович Л. В.* Принцип мажорант и метод Ньютона // Доклады АН СССР. 1951. 76 (1). С. 17–20.
29. *Канторович Л. В.* О дальнейшем применении метода мажорант // Доклады АН СССР. 1951. 80 (6). С. 849–852.
30. *Канторович Л. В.* О приближенном решении функциональных уравнений // Успехи математических наук. С. 1956. 11 (6). С. 99–116.
31. *Канторович Л. В.* Некоторые дальнейшие применения метода Ньютона // Вестник ЛГУ. Сер.: мат.-мех. 1957. № 7. С. 68–103.
32. *Канторович Л. В.*, *Акилов Г. П.* Функциональный анализ в нормированных пространствах. М: Физматгиз, 1959.
33. *Канторович Л. В.*, *Акилов Г. П.* Функциональный анализ. М: Наука, 1977.
34. *Klatte D.*, *Kummer B.* Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications. Dordrecht: Kluwer, 2002.
35. *Красносельский М. А.*, *Вайникко Г. М.*, *Забрейко П. П.*, *Рутицкий Я. Б.*, *Стеценко В. Я.* Приближенное решение операторных уравнений. М: Наука, 1969.
36. *Levenberg K.* A method for the solution of certain nonlinear problems in least squares // Quarterly of Applied Mathematics. 1944. № 2. P. 164–168.
37. *Левитин Е. С.*, *Поляк Б. Т.* Методы минимизации при наличии ограничений // ЖВМ и МФ. 1966. 6 (5). С. 787–823.
38. *Люстерник Л. А.* Об условном экстремуме функционалов // Мат. сборник. 1934. 41 (3). С. 390–401.
39. *Mandelbrot B.* The Fractal Geometry of Nature. New York: Freeman W. H., 1983.
40. *Marquardt D.* An algorithm for least squares estimation of nonlinear parameters // SIAM Journal of Applied Mathematics. 1963. 11. P. 431–441.
41. *Mathews J. H.* Bibliography for Newton's method (math.fullerton.edu/mathews/newtons-method/Newton'sMethodBib/Links/Newton'sMethodBib_lnk_3.html).
42. *Mifflin R.* Semismooth and semiconvex functions in constrained optimization // SIAM Journal on Control and Optimization. 1977. № 15. P. 959–972.

43. *Мысовских И. П.* О сходимости метода Л. В. Канторовича решения функциональных уравнений и его применения // Доклады АН СССР. 1950. 70 (4). С. 565–568.
44. *Nesterov Yu.*, *Nemirovski A.* Interior-Point Polynomial Algorithms in Convex Programming // SIAM. Philadelphia, 1994.
45. *Nesterov Yu.*, *Polyak B.* Cubic regularization of Newton method and its global performance // Mathematical Programming. Ser. A. 2006. 108. P. 177–205.
46. *Nesterov Yu.* Introductory Lectures on Convex Programming. Boston: Kluwer, 2004.
47. *Островский А. М.* Решение уравнений и систем уравнений. М.: Изд-во иностр. лит., 1963.
48. *Ортега Дж.*, *Рейнболдт В.* Итерационные методы решения нелинейных систем уравнений со многими неизвестными. М.: Мир, 1975.
49. *Peitgen H. O. Saupe D. Haeseler F.* Cayley's problem and Julia sets // Mathematical Intelligencer. 1984. 6 (2). P. 11–20.
50. Newton's Method and Dynamical Systems / Ed. H. O. Peitgen. Berlin: Springer, 1989.
51. *Поляк Б. Т.* Градиентные методы решения уравнений и неравенств // ЖВМ и МФ. 1964. 4 (6). С. 995–1005.
52. *Polyak B. T.* Convexity of nonlinear image of a small ball with applications to optimization // Set-Valued Analysis. 2001. 9 (1–2). P. 159–168.
53. *Поляк Б. Т.* Локальное программирование // ЖВМ и МФ. 2001. 41 (9). С. 1324–1331.
54. *Polyak B. T.* Convexity of the reachable set of nonlinear systems under L_2 bounded controls, Dynamics of Continuous, Discrete and Impulsive Systems. 2004. 11 (2–3). P. 255–268.
55. *Поляк Б. Т.* Введение в оптимизацию. М: Наука, 1983.
56. *Поляк Б. Т.* Итерационные методы, использующие множители Лагранжа, для решения экстремальных задач с ограничениями типа равенств // ЖВМ и МФ. 1970. 10 (5). С 1098–1106.
57. *Поляк Б. Т.* Минимизация негладких функционалов // ЖВМ и МФ. 1969. 9 (3). С. 509–521.
58. *Qi L.*, *Sun J.* A nonsmooth version of Newton's method // Mathematical Programming. 1993. 58. P. 353–367.
59. *Ramm A. G.* Acceleration of convergence: a continuous analog of the Newton's method // Applicable Analysis. 2002. 81 (4). P. 1001–1004.
60. *Rheinboldt W. C.* Methods for Solving Systems of Nonlinear Equations // SIAM, Philadelphia, 1998.
61. *Robinson S. M.* Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms // Mathematical Programming. 1974. № 7. P. 1–16.
62. *Robinson S. M.* Newton's method for a class of nonsmooth functions // Set-Valued Analysis. 1994. 2. P. 291–305.
63. *Smale S.* A convergent process of price adjustment and global Newton's methods // Journal of Mathematical Economics. 1976. № 3. P. 107–120.
64. *Smale S.* Newton's method estimates from data at one point // Ewing R., Gross K., Martin C. (Eds.). The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics. Springer, 1986.
65. *Smale S.* Complexity theory and numerical analysis // Acta Numerica. 1997. № 6. P. 523–551.
66. *Wright S. J.* Primal-Dual Interior-Point Methods // SIAM. Philadelphia, 1997.
67. *Ypma T. J.* Historical development of the Newton—Raphson method // SIAM Review. 1995. 37 (4). P. 531–551.