

УДК: 519.85

Взаимосвязь и реализация квазиньютоновских и ньютоновских методов безусловной оптимизации

А. Б. Свириденко¹, Г. А. Зеленков²

¹ ФГБОУ ВО «Кубанский государственный университет» филиал в г. Новороссийске,
Россия, 353922, г. Новороссийск, ул. Героев Десантников, д. 87

² ФГБОУ ВПО «Морской государственный университет им. адм. Ф. Ф. Ушакова»,
Россия, 353922, г. Новороссийск, пр. Ленина, д. 93

E-mail: roshechka@gmail.com

Получено 9 октября 2015 г.,
после доработки 16 февраля 2016 г.

Рассмотрены ньютоновские и квазиньютоновские методы безусловной оптимизации, основанные на факторизации Холецкого, с регулировкой шага и с конечно-разностной аппроксимацией первых и вторых производных. Для увеличения эффективности квазиньютоновских методов предложено модифицированное разложение Холецкого квазиньютоновской матрицы, определяющее и решение проблемы масштабирования шагов при спуске, и аппроксимацию неквадратичными функциями, и интеграцию с методом доверительной окрестности. Предложен подход к увеличению эффективности ньютоновских методов с конечно-разностной аппроксимацией первых и вторых производных. Приведены результаты численного исследования эффективности алгоритмов.

Ключевые слова: ньютоновские методы, квазиньютоновские методы, факторизация Холецкого, масштабирование шагов, метод доверительной окрестности, конечно-разностная аппроксимация, алгоритм, численные исследования, безусловная оптимизация

Correlation and realization of quasi-Newton methods of absolute optimization

A. B. Sviridenko¹, G. A. Zelenkov²

¹ FSEI of HPE "Kuban State University" branch in Novorossiysk, 87 Geroev-Desantnikov st., 353922, Russia

² Admiral Ushakov State Maritime University, Novorossisk, 93 Lenin's av., 353922, Russia

Abstract. — Newton and quasi-Newton methods of absolute optimization based on Cholesky factorization with adaptive step and finite difference approximation of the first and the second derivatives. In order to raise effectiveness of the quasi-Newton methods a modified version of Cholesky decomposition of quasi-Newton matrix is suggested. It solves the problem of step scaling while descending, allows approximation by non-quadratic functions, and integration with confidence neighborhood method. An approach to raise Newton methods effectiveness with finite difference approximation of the first and second derivatives is offered. The results of numerical research of algorithm effectiveness are shown.

Keywords: Newton methods, quasi-Newton methods, Cholesky factorization, step scaling, method of confidence neighborhoods, finite difference approximation, algorithm, numerical research, absolute optimization

Citation: *Computer Research and Modeling*, 2016, vol. 8, no. 1, pp. 55–78 (Russian).

1. Предисловие

Рассматривается задача безусловной минимизации:

$$\min_{x \in R^n} F(x), \quad (1.1)$$

где $F: R^n \rightarrow R^1$ — гладкая функция, а R^n обозначает n -мерное евклидово пространство.

Пусть h^k , H^k — градиент и гессиан, вычисленные на итерации k в точке x^k процесса решения задачи (1), тогда, в приведенных обозначениях, общий принцип построения большинства ньютоновских методов безусловной оптимизации с регулировкой шага состоит в следующем. На каждой итерации сначала строится некоторая «связанная» с H^k существенно положительно-определенная матрица \bar{H}^k , а затем направление спуска p^k вычисляется как решение системы:

$$\bar{H}^k p^k = -h^k \quad (1.2)$$

и определяется новая точка

$$x^{k+1} = x^k + \alpha_k p^k, \quad (1.3)$$

где α_k — длина шага, для которого $F^{k+1} < F^k$. При этом процедуру построения \bar{H}^k организуют так, чтобы \bar{H}^k совпадала с исходной матрицей H^k , если последняя сама является положительно-определенной, причем выяснение определенности H^k и построение \bar{H}^k осуществляется параллельно в рамках одной процедуры на основе некоторых матричных разложений, которые позволяют выявить знаки собственных чисел H^k и приспособиться для генерации \bar{H}^k .

На практике условие гладкости целевой функции является излишне строгим — как правило, или первые и (или) вторые производные не существуют, или процесс их вычисления слишком трудоемок. В этом случае быстро сходящийся алгоритм можно построить на основе их конечно-разностной аппроксимации. Подобно своему классическому прототипу он порождает целый класс модифицированных методов, причем модификации, предлагаемые для обобщения метода Ньютона на случаи со знаконеопределенными H^k , применимы и к его конечно-разностному аналогу, и наоборот [Зеленков и др., 2013].

В квазиньютоновских, как и в ньютоновских, методах направление спуска p^k определяется в результате решения систем линейных уравнений. В ньютоновских методах это системы с матрицами вторых производных H^k , которые на разных итерациях вычисляются независимо друг от друга, в то время как в квазиньютоновских методах это системы с квазиньютоновскими матрицами H^k , причем матрица очередной системы для построения p^k отличается от предыдущей лишь поправкой малого ранга, а выбор формулы пересчета квазиньютоновских матриц определяет вариант квазиньютоновского метода. Подробное описание формул пересчета квазиньютоновских матриц, таких как симметричная формула ранга один, DFP (Дэвидона–Флетчера–Пауэлла), PSB (Пауэлла–Бройдена), BFGS (Бройдена–Флетчера–Гольдфарба–Шанно), можно найти, например, в [Гилл и др., 1985].

В квазиньютоновских, как и в ньютоновских, методах выяснение определенности H^k и построение \bar{H}^k осуществляются параллельно в рамках одной процедуры на основе некоторых матричных разложений, которые позволяют выявить знаки собственных чисел H^k и приспособиться для генерации \bar{H}^k [Gill et al., 1974].

В данной работе в основе выяснения определенности H^k и построения \bar{H}^k лежит факторизация (разложение) Холецкого. Разложение Холецкого для симметричной положительно-

определенной матрицы H^k имеет вид:

$$H^k = L^k D^k (L^k)^T = (U^k)^T D^k U^k, \quad (1.4)$$

где $U^k = (L^k)^T$ — верхняя треугольная матрица с диагональными элементами, равными единице, а D^k — диагональная матрица с положительными ведущими элементами по диагонали. Коль скоро столбцы матрицы L^k с номерами от 1-го по $(j-1)$ -ый известны, ее j -ый столбец определяется по формулам:

$$d_{jj}^k = h_{jj}^k - \sum_{s=1}^{j-1} d_{ss}^k (l_{js}^k)^2, \quad l_{ij}^k = \frac{1}{d_{jj}^k} \left(h_{ij}^k - \sum_{s=1}^{j-1} d_{ss}^k l_{js}^k l_{is}^k \right). \quad (1.5)$$

Аналогичные формулы существуют и для построчной организации вычислений. В отличие от гауссовых исключений алгоритм Холесского численно устойчив без каких-либо перестановок [Гилл и др., 1985]. Это свойство определяется соотношением

$$\sum_{i=1}^n (l_{ij}^k)^2 d_{jj}^k = h_{jj}^k, \quad j = 1, \dots, n \quad (1.6)$$

между элементами H^k и L^k . В силу него существует априорное ограничение сверху на элементы $r_{ij}^k = l_{ij}^k \sqrt{d_{jj}^k}$: каждый из них не превосходит максимальной величины h_{jj}^k . Соответственно, «лавинообразный рост» элементов r_{ij}^k невозможен независимо от того, будут ведущие элементы малыми или нет.

Построение направлений ньютоновского поиска. Факторизация (1.4) неприменима для любой симметричной матрицы по следующим причинам. Во-первых, для знаконеопределенной матрицы H^k факторизация Холесского может не существовать. Во-вторых, даже если она и существует, то гарантировать численную устойчивость алгоритма уже нельзя, поскольку никаких априорных ограничений на субдиагональные элементы L^k в рассматриваемом случае не будет. Итак, на основе обычной факторизации Холесского построения \bar{H}^k не получить. Для этого нужно воспользоваться модифицированной факторизацией. Наиболее эффективная вычислительная схема предложена Гиллом и Мюрреем [Gill et al., 1974; Гилл и др., 1985]. Данный подход состоит в том, чтобы строить факторы Холесского D^k и L^k , подчиняющиеся двум требованиям: все диагональные элементы D^k должны быть существенно положительными, модули всех элементов треугольного фактора L^k должны быть равномерно ограничены сверху. Точнее говоря, требуется, чтобы для всех $j = 1, 2, \dots, n$ и некоторых заданных положительных δ и β выполнялись неравенства

$$d_{jj}^k \geq \delta, \quad |r_{ij}^k| \leq \beta, \quad i > j, \quad (1.7)$$

где r_{ij}^k — введенные для удобства изложения вспомогательные величины, по определению равные $u_{ij}^k \sqrt{d_{jj}^k}$. Гринштадт [Гилл и др., 1977] предложил при реализации алгоритма на конкретной ЭВМ, в которой под запись мантиссы отводится t битов, величину δ вычислять по формуле

$$\delta = \max \{ 2^{-t} \|H^k\|_{\infty}, 2^{-t} \}. \quad (1.8)$$

Для сохранения численной устойчивости процедуры построения H^k , а также для совпадения H^k и \overline{H}^k в случае положительно-определенной H^k целесообразно величину β вычислять по формуле

$$\beta^2 = \max \left\{ \gamma, \xi / \sqrt{n^2 - 1}, \varepsilon_M \right\}. \quad (1.9)$$

Здесь ξ — максимальный модуль недиагонального элемента H^k , γ — значение максимального из диагональных элементов H^k , ε_M — машинная точность. ε_M вводится в формулу расчета β , чтобы обеспечить устойчивость вычислений, когда норма H^k очень мала [Гилл и др., 1977]. При этом процедура расчета модифицированных факторов L^k , D^k фактически представляет собой обычный алгоритм факторизации Холецкого с попутным увеличением (по мере необходимости) диагонали исходной матрицы с целью добиться выполнения неравенств (1.7). Матрицы L^k и D^k , полученные по окончании описанной процедуры, будут факторами Холецкого для положительно определенной матрицы \overline{H}^k , связанной с H^k следующим образом:

$$L^k D^k (L^k)^T = H^k + E^k = \overline{H}^k, \quad (1.10)$$

где E^k — неотрицательная диагональная матрица, j -ый элемент которой равен e_{jj}^k . Таким образом, положительно-определенная матрица \overline{H}^k может отличаться от исходной матрицы H^k только диагональными элементами.

Направление спуска p^k вычисляется как решение системы $\overline{H}^k p^k = -h^k$, причем $p^k = -(\overline{H}^k)^{-1} h^k$ определяют последовательным решением двух систем линейных уравнений с треугольными матрицами:

$$L^k y^k = -h^k, \quad D^k (L^k)^T p^k = y^k. \quad (1.11)$$

Для построения модифицированного разложения Холецкого требуется выполнить около $\frac{1}{6}n^3$ арифметических операций, примерно столько же, сколько требуется для обычного разложения для положительно-определенной матрицы.

Ниже приводится детальное описание всех операций, выполняемых по ходу построения модифицированного разложения Холецкого с перестановками [Гилл и др., 1977]. Дана наиболее эффективная схема организации расчетов. В процессе вычисления j -го столбца матрицы L^k участвуют вспомогательные величины $c_{is}^k = l_{is}^k d_{ss}^k$, $s = 1, \dots, j$, $i = j, \dots, n$.

Вычислительная схема 1.1 модифицированного разложения Холецкого.

Шаг 0 (расчет порога для элементов). Вычислить $\beta^2 = \max \{ \gamma^0, \xi / \nu, \varepsilon_M \}$, где $\nu = \max \{ 1, \sqrt{n^2 - 1} \}$, а числа γ и ξ суть максимальные значения модулей диагонального и недиагонального элементов H^k .

Шаг 1 (инициализация). Присвоить индексу столбца j значение 1. Положить $c_{ii}^k = g_{ii}^k, i = 1, \dots, n$.

Шаг 2 (перестановка строк и столбцов). Найти индекс q такой, что $|c_{qq}^k| = \max_{j \leq i \leq n} |c_{ii}^k|$. Поменять местами все данные, отвечающие столбцам матрицы H^k с номерами q и j , а затем проделать то же самое с данными, отвечающими ее q -ой и j -ой строкам.

Шаг 3 (поиск максимальной по модулю величины $l_{ij}^k d_{ii}^k$). Вычислить $c_{ij}^k = g_{ij}^k - \sum_{s=1}^{j-1} l_{js}^k c_{is}^k$ для $i = j+1, \dots, n$ и найти $\theta_j = \max_{j+1 \leq i \leq n} |c_{ij}^k|$ (если $j = n$, взять $\theta_j = 0$).

Шаг 4 (расчет j -го диагонального элемента фактора D^k). Вычислить $d_{jj}^k = \max \{ \delta, |c_{jj}^k|, \theta_j^2 / \beta^2 \}$ и поправку $e_{jj}^k = d_{jj}^k - c_{jj}^k$. Если $j = n$, вычисления прекратить.

Шаг 5 (расчет j -ой строки L^k). Вычислить $l_{js}^k = c_{js}^k / d_{ss}^k$ для $s = 1, \dots, j-1$. Для $i = j+1, \dots, n$ пересчитать диагональные элементы c_{ii}^k по формуле $c_{ii}^k = c_{ii}^k - l_{ij}^k c_{ij}^k$. Присвоить j значение $j+1$ и вернуться к шагу 2.

Модифицированная факторизация Холесского, помимо прочего, позволяет определить и направление отрицательной кривизны, решая уравнение $(L^k)^T p^k = e_s$, где индекс s выбирается из условия $d_{ss}^k - e_{ss}^k \leq d_{jj}^k - e_{jj}^k, j = 1, \dots, n$.

Замечание. Применение вычислительной схемы 1.1 модифицированного разложения Холесского в подходе Городецкого и Гришагина к построению квазиньютоновских направлений поиска [Городецкий и др., 2003] обсуждается ниже.

Подход к увеличению эффективности алгоритма Гилла и Мюррея предложен в [Зеленков и др., 2013]. Доказано, что подход к построению направления спуска p^k определяет решение проблемы масштабирования шагов при спуске и интеграцию с методом доверительной окрестности следующим образом. В алгоритме Гилла и Мюррея направление спуска p^k в конечном счете вычисляется как решение системы

$$(L^k)^T p^k = U^k p^k = u^k, \tag{1.12}$$

где $u^k = (D^k)^{-1} y^k$, а y^k есть решение системы $L^k y^k = -h^k$. Величина элементов $|u_j^k|$ зависит от способа задания направления спуска. Следуя подходу Гилла и Мюррея, для численной устойчивости расчета элементов D^k, L^k, u^k достаточно потребовать изменения способа задания p^k так, чтобы для всех $j = 1, 2, \dots, n$ и некоторых заданных положительных δ, β и ω выполнялись неравенства:

$$d_{jj}^k \geq \delta, |r_{ij}^k| \leq \beta, i > j, |u_j^k| \leq \omega. \tag{1.13}$$

Это может стать ключом к решению проблемы масштабирования шагов при спуске, но такое предположение ошибочно — подход линейной алгебры к вычислению направления спуска p^k исключает расчет элементов вектора u^k по ходу построения модифицированного разложения Холесского [Парлетт, 1983; Свириденко, КиМ, 2015]. Нужен альтернативный подход, который опишем следующим образом. Равенства

$$H^k = (U^k)^T D^k U^k \tag{1.14}$$

достаточно для определения элементов матриц D^k, U^k . В скалярной форме это равенство выглядит следующим образом:

$$h_{ij}^k = \sum_{\mu=1}^i u_{\mu i}^k u_{\mu j}^k d_{\mu\mu}^k, \tag{1.15}$$

отсюда, полагая $u_{ii}^k = 1$, построим соотношения

$$d_{ii}^k = h_{ii}^k - \sum_{s=1}^{i-1} (u_{si}^k)^2 d_{ss}^k, \quad u_{ij}^k = \left(h_{ij}^k - \sum_{s=1}^{i-1} u_{si}^k u_{sj}^k d_{ss}^k \right) / d_{ii}^k, \quad j > i, \quad (1.16)$$

для расчета элементов D^k , U^k . Построение факторов Холецкого математически эквивалентно применению метода исключения Гаусса к системе уравнений $H^k p^k = -h^k$ в прямом порядке, при этом техника исключения Гаусса позволяет получить соотношение

$$u_i^k = - \left(h_i^k - \sum_{s=1}^{i-1} u_s^k u_{si}^k d_{ss}^k \right) / d_{ii}^k \quad (1.17)$$

для расчета элементов вектора u^k и систему уравнений (1.12) для вычисления направления спуска p^k . Интеграция техники исключения Гаусса и факторизации Холецкого приводит к соотношениям (1.16), (1.17), которые с помощью элементарных арифметических операций позволяют вычислить элементы u^k по ходу построения факторов Холецкого.

Интеграция техники исключения Гаусса и факторизации Холецкого порождает множество численно устойчивых способов задания направления спуска p^k , в основе которых лежит требование, чтобы на очередном шаге вычислений коэффициентов факторов Холецкого соответствующий диагональный элемент матрицы D^k и соответствующий элемент вектора u^k сначала рассчитывались по вычисленным ранее значениям этих коэффициентов. Затем диагональный элемент D^k увеличивается настолько, насколько необходимо, чтобы все диагональные элементы D^k были существенно положительными, модули всех элементов U^k , u^k были равномерно ограничены сверху. Например, достаточно потребовать, чтобы для всех $i = 1, 2, \dots, n$ выполнялись неравенства:

$$d_{ii}^k \geq \delta, \quad |u_{ij}^k| \leq 1, \quad j > i, \quad |u_i^k| \leq 1. \quad (1.18)$$

Положив $d_{ii}^k = \max \left\{ \delta, |c_{ii}^k|, \theta_i^2 / \beta^2, |c_i^k|, \theta_i \right\}$, поскольку элементы d_{ii}^k , c_{ij}^k / d_{ii}^k и c_i^k / d_{ii}^k определяют i -ые строки матриц D^k , U^k , u^k и удовлетворяют неравенствам (1.18). Здесь c_{ii}^k , c_{ij}^k , c_i^k — вспомогательные величины, определяемые следующим образом:

$$c_{ii}^k = h_{ii}^k - \sum_{s=1}^{i-1} (u_{si}^k)^2 d_{ss}^k, \quad c_{ij}^k = h_{ij}^k - \sum_{s=1}^{i-1} u_{si}^k u_{sj}^k d_{ss}^k, \quad j > i, \quad c_i^k = - \left(h_i^k - \sum_{s=1}^{i-1} u_s^k u_{si}^k d_{ss}^k \right). \quad (1.19)$$

Интеграция техники исключения Гаусса и факторизации Холецкого порождает множество способов интеграции с методом доверительной окрестности, например, следующим образом. В процессе построения D^k , U^k , u^k вычислить элементы $g_i^k = d_{ii}^k / l_i^k$ (если $d_{ii}^k > l_i^k \gamma$, взять $g_i^k = l_i^k / \gamma$), где $l_i^k = \max \left\{ \delta, |d_{ii}^k| \right\}$, вычислить $\gamma^k = \max_{1 \leq i \leq n} g_i^k$ и определить направление спуска p^k из системы

$$\frac{1}{\gamma^k} \overline{H}^k p^k = -h^k. \quad (1.20)$$

Здесь параметр γ задается пользователем, число γ^k ($\gamma^k \geq 1$ по построению) характеризует степень однородности, при этом (1.20) есть простейшая форма аппроксимации, отличная от квадратичной. Пусть

$$\overline{H}^k p^k = -\gamma^k h^k = -h^k - (\gamma^k - 1) h^k = -h^k - G^k p^k, \quad (1.21)$$

где G^k — диагональная матрица с элементами

$$g_{ii}^k = \frac{(\gamma^k - 1)h_i^k}{p_i^k} \quad (1.22)$$

на диагонали, или

$$\left(\bar{H}^k + G^k\right)p^k = -h^k, \quad (1.23)$$

что означает сдвиг на g_{ii}^k всех i -х собственных значений матрицы \bar{H}^k . В методах с регуляризацией шага матрицу H^k модифицируют так, чтобы изменения не затрагивали подпространства, натянутого на ее собственные векторы с положительными собственными значениями [Гилл и др., 1985]. Если же замена осуществляется в методе доверительной окрестности, то она отражается на всех векторах, так как результатом замены H^k на матрицу $\bar{H}^k = H^k + \lambda^k I$ будет сдвиг на λ^k всех собственных значений матрицы H^k [Гилл и др., 1985]. Из вышесказанного следует, что стратегия выбора направления спуска определяет и интеграцию с методом доверительной окрестности.

Ниже дана наиболее эффективная схема организации расчетов. Все фигурирующие в ней величины при реализации на ЭВМ могут размещаться в памяти, первоначально выделяемой для записи матриц H^k , h^k . При этом коэффициенты рассчитываемых факторов занимают места ее использованных элементов. В процессе вычисления i -й строки фактора U^k участвуют вспомогательные величины (1.19). В процессе построения системы уравнений $U^k p^k = u^k$ на каждой итерации k выявляются знаки собственных чисел и вычисляется количество отрицательных n_o и нулевых n_0 собственных значений, число переходов к направлению отрицательной кривизны n_s .

Вычислительная схема 1.2 модифицированного разложения Холецкого

Шаг 0. Вычислить $\beta^2 = \max\{\gamma^0, \xi/v, \varepsilon_M\}$, где $v = \max\{1, \sqrt{n^2 - 1}\}$, а числа γ^0 и ξ — максимальные значения модулей диагонального и недиагонального элементов H^k . Положить $\varepsilon_0 = 2^{-\frac{\tau_F}{2}}$, $\varepsilon_s = 2^{-\frac{\tau_F}{3}} \left(1 + \|F(x^{k-1}) - F(x^k)\|\right)$. Здесь τ_F — число правильных разрядов $F(x^k)$, которые хотелось бы получить.

Шаг 1. Положить $i = 1$, положить $c_{jj}^k = h_{jj}^k, j = 1, \dots, n$, $c_j^k = -h_j^k, j = 1, \dots, n$.

Шаг 2. Найти индекс q такой, что $|c_{qq}^k| = \max_{i \leq j \leq n} |c_{ij}^k|$, и поменять местами все данные, отвечающие строкам матрицы H^k с номерами q и i , а затем проделать то же самое с данными, отвечающими ее q -му и i -му столбцам.

Шаг 3. Если $c_{ii}^k < \varepsilon_0$, то положить $l_i^k = \delta$; иначе положить $l_i^k = c_{ii}^k$.

Шаг 4. Если $|c_{ii}^k| \leq \varepsilon_0$, то положить $n_o = n_o + 1$ и перейти к шагу 8; иначе перейти к шагу 5.

Шаг 5. Если $c_{ii}^k > 0$, то перейти к шагу 8; иначе — перейти к шагу 6.

Шаг 6. Положить $n_o = n_o + 1$. Если $\|h^k\| > \varepsilon_s$, то перейти к шагу 8; иначе перейти к шагу 7.

Шаг 7. Построить систему уравнений $U^k p^k = e_i$ для поиска направления отрицательной кривизны по формулам $u_j^k = 0, j = 1, \dots, n$, $u_i^k = 1$, положить $n_s = n_s + 1$ и остановиться.

Шаг 8. Если $i = n$, то положить $\theta_i = 0$; иначе вычислить вспомогательные величины $c_{ij}^k, j = i + 1, \dots, n$, по формуле $c_{ij}^k = h_{ij}^k - \sum_{s=1}^{i-1} u_{si}^k u_{sj}^k d_{ss}^k$ и найти $\theta_i = \max_{i+1 \leq j \leq n} |c_{ij}^k|$.

Шаг 9. Вычислить диагональный элемент фактора D^k и элемент вектора u^k по формулам $d_{ii}^k = \max\{\delta, |c_{ii}^k|, \theta_i^2 / \beta^2, |c_i^k|, \theta_i\}$, $u_i^k = c_i^k / d_{ii}^k$. Вычислить $g_i^k = d_{ii}^k / l_i^k$ (если $d_{ii}^k > l_i^k \gamma$, взять $g_i^k = l_i^k / \gamma$). Если $i = n$, то перейти к шагу 11.

Шаг 10. Вычислить строку $u_{ij}^k = c_{ij}^k / d_{ii}^k$, $j = i+1, \dots, n$, фактора U^k , пересчитать вспомогательные величины по формулам $c_{jj}^k = c_{jj}^k - u_{ij}^k c_{ij}^k$, $j = i+1, \dots, n$, $c_j^k = c_j^k - u_{ij}^k c_i^k$, $j = i+1, \dots, n$, положить $i = i+1$ и перейти к шагу 2.

Шаг 11. Найти $\gamma^k = \max_{1 \leq i \leq n} g_i^k$, положить $u^k = \gamma^k u^k$ и остановиться.

Развитие подхода к построению вычислительной схемы 1.2 модифицированного разложения Холесского обсуждается ниже.

Построение направлений квазиньютоновского поиска. В данной работе рассматриваются способы задания квазиньютоновского поиска, позволяющие ослабить или снять недостатки квазиньютоновских методов, которые опишем следующим образом. Квазиньютоновские методы очень близки к методам сопряженных градиентов — большинство вариантов квазиньютоновских методов (например, одна из наиболее эффективных схем Бройдена–Флетчера–Гольдфарба–Шенно) при минимизации сильно выпуклых квадратичных функционалов приводят к одной и той же траектории спуска, вырождаясь в хорошо изученные методы сопряженных градиентов [Nazareth, 1979; Черноруцкий, 2013]. Предположение о невыпуклости вносит дополнительные трудности: в этих условиях метод сопряженных градиентов по характеристикам сходимости эквивалентен градиентному методу наискорейшего спуска со всеми вытекающими отсюда последствиями [Черноруцкий, 2011].

Кроме отмеченных дефектов, общих для методов сопряженных градиентов и квазиньютоновских методов, последние имеют дополнительные недостатки, связанные с проблемой потери положительной определенности квазиньютоновских матриц из-за накопления вычислительных погрешностей в рекуррентных процедурах аппроксимации матриц Гессе [Гилл и др., 1985].

Городецкий и Гришагин в [Городецкий и др., 2003] предложили подход к частичному или полному снятию отмеченных выше недостатков, в основе которого лежит модифицированное разложение Холесского квазиньютоновской матрицы H^k (вычислительная схема 1.1). Это связано с тем, что процедура модифицированной факторизации Холесского — численно устойчивый алгоритм, генерирующий положительно-определенную матрицу, которая может отличаться от исходной матрицы только диагональными элементами. Это оптимизированный алгоритм в том смысле, что параметр β подбирается в нем путем минимизации априорной оценки нормы поправки E^k при условии сохранения существенно положительно-определенной матрицы неизменной [Gill et al., 1974]. Следует также отметить, что реальная величина нормы почти всегда оказывается меньше априорной оценки. Развитие подхода Городецкого и Гришагина обсуждается ниже.

1.1. Тестовые задачи и результаты вычислений

Трудности сравнения между собой нескольких методов оптимизации по результатам численных экспериментов состоят в следующем.

Во-первых, сравнению подвергаются не методы, а машинные реализации соответствующих алгоритмов. Результат сравнения зависит от качества программирования, выбора параметров алгоритма, числовых типов данных и так далее.

Во-вторых, неясно, каким образом соизмерять трудоемкость различных методов. Естественный на первый взгляд критерий — затраченное машинное время — в действительности не подходит: сложно сравнивать быстродействие различных машин, не всегда имеются данные о затратах времени при работе в мультипроцессорном режиме и так далее. Более надежным по-

казателем является число вычислений минимизируемой функции или другая «внутренняя характеристика метода». Но и здесь возникает ряд проблем. Например, неясно, как соизмерять трудоемкость вычисления функции и решения различных вспомогательных задач. Затруднительно сопоставлять вычисления функции и ее производных. В тех задачах, где применяются конечно-разностные аппроксимации производных, все намного проще — одно вычисление градиента эквивалентно n вычислениям функции, вычисление гессиана — эквивалентно $n(n+1)/2$ вычислениям функции. Однако для дискретной задачи оптимального управления градиент лишь примерно вдвое «дороже» функции [Поляк, 1983]. Для квадратичной функции градиент вычисляется даже проще, чем ее значения. Так же дело обстоит и с субградиентом в минимаксной задаче и так далее.

В-третьих, методы могут по-разному вести себя на разных этапах процесса минимизации. Более того, поведение разных величин, характеризующих точность решения, может быть также различным.

Никакого удовлетворительного способа преодоления перечисленных трудностей не существует. Единственное, что можно сделать в подобной ситуации, — приводить данные о результатах вычислений в развернутой форме, чтобы иметь возможность сравнивать методы по разным критериям.

Рекомендуется при публикации результатов проверки методов придерживаться следующих правил [Поляк, 1983]:

- приводить точную формулировку задачи, для которой проводился счет, включая все ее параметры и начальное приближение;
- указывать тип ЭВМ и данных, язык и среду программирования, сведения о программе;
- давать подробное описание применяемого алгоритма или отсылать к его публикации, если она имеется;
- сообщать различные характеристики точности приближения, невязку в ограничениях и выполнении условий экстремума. В задачах малой размерности приводить и сами приближения;
- указывать подробные сведения о трудоемкости вычислений.

При выполнении этих условий можно воспроизводить полученные данные и сравнивать их с другими по разным показателям.

Общие требования к тестовым задачам. Сравнение методов целесообразно проводить на стандартных специально подобранных задачах-тестах. Желательно, чтобы тесты удовлетворяли следующим требованиям [Поляк, 1983]:

- тесты должны быть унифицированы и быть общепринятыми; обычно такие задачи формируются в результате своеобразного естественного отбора; основным достоинством тема является его популярность;
- тесты должны моделировать типовые трудности для данного класса задач, например, для задач безусловной минимизации нужны тесты с разной обусловленностью, разной размерностью, разной кривизной линий уровня, одно- и многоэкстремальные и так далее;
- решение в тестовой задаче должно быть известным;
- задачи должны быть достаточно компактными — не требовать больших массивов информации или сложных правил вычисления функций;
- в качестве тестовых не годятся задачи, обладающие специфическими особенностями, дающие преимущества тому или иному методу.

Безусловная минимизация гладких функций. Всюду далее x^* — точка глобального минимума $F(x)$ на R^n , $F^* = F(x^*)$, k — число итераций, $\delta F = |F(x^*) - F(x^k)|$ — точность решения по функции, $\delta x = \max |x_i^* - x_i^k|$ — точность решения по аргументам, k_0 — число вычислений функции, $\nabla F(x)$ — градиент $F(x)$, $\nabla^2 F(x)$ — матрица вторых производных (гессиан).

Тестовые задачи с известным решением можно построить следующим образом. Выбираются точка x^* и функции $\phi_i(x)$, $i = 1, 2, \dots, m$. Тогда

$$F(x) = \sum_{i=1}^m (\phi_i(x) - \phi_i(x^*))^2 \quad (1.1.1)$$

достигает минимума в точке x^* . При этом $F^* = 0$, $\nabla^2 F(x^*) = \sum_{i=1}^m \nabla \phi_i(x^*) \nabla^T \phi_i(x^*)$, так что можно регулировать обусловленность задачи за счет выбора $\phi_i(x)$. В частности, если $m < n$ или $m = n$, но $\nabla \phi_i(x^*) = 0$ для некоторых i , то получаем вырожденную точку минимума. Функции вида (1.1.1), вообще говоря, невыпуклы и могут иметь локальные и даже глобальные минимумы, отличные от x^* .

Другой способ связан с выбором функций вида

$$F(x) = \sum_{i=1}^{n+1} f_i(x_i, x_{i-1}), x_0 = a, x_{n+1} = b. \quad (1.1.2)$$

Тогда, задавшись $x_0^* = a$ и решив последовательно одномерные уравнения

$$\partial F / \partial x_i = \partial / \partial x_i [f_i(x_i, x_{i-1}^*) + f_{i+1}(x_{i+1}, x_i^*)] = 0 \quad (1.1.3)$$

(решение обозначим x_{i+1}^*) для $i = 0, 1, \dots, n$, выберем $b = x_{n+1}^*$. Очевидно, что тогда $\nabla F(x^*) = 0$. Если $f_i(x_i, x_{i-1})$ выпуклы по x_i, x_{i-1} , то $F(x)$ выпукла и x^* — точка минимума $F(x)$.

Наконец, можно взять произвольную гладкую выпуклую функцию $\phi(x)$ и произвольную точку x^* и построить

$$F(x) = \phi(x) - (\nabla \phi(x^*), x). \quad (1.1.4)$$

Тогда $\nabla F(x^*) = 0$, $F(x)$ выпукла, и потому x^* — точка глобального минимума $F(x)$. Существуют и другие приемы построения функций с известной точкой минимума.

Приводимые ниже примеры для численного исследования как ньютоновских, так и квази-ньютоновских методов оптимизации построены по одному из описанных способов. Следует отметить, что для функций вида (1.1.1), (1.1.2) существуют специальные методы, обладающие повышенной эффективностью. Например, для задач (1.1.1) эффективен метод Гаусса–Ньютона [Гилл и др., 1985], для (1.1.2) — метод динамического программирования и покоординатного спуска [Поляк, 1983]. Аналогичным образом специфика задач (1.1.1), (1.1.2) может сказаться на поведении общих методов оптимизации.

Перейдем к описанию конкретных тестовых задач, предложенных Поляком [Поляк, 1983].

Задача 1 (тест-функция Розенброка, $n = 2$):

$$F(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2, \quad x^0 = (-1.2 \ 1)^T,$$

$x^* = (1 \ 1)^T$, $F(x^*) = 0$. Функция плохо обусловленная, невыпуклая, с параболическим оврагом, точка x^0 далека от точки x^* .

Задача 2 (тест-функция Пауэлла, $n = 4$):

$$F(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4,$$

$x^0 = (3 \ -1 \ 0 \ 1)^T$, $x^* = (0 \ 0 \ 0 \ 0)^T$, $F(x^*) = 0$. Функция не выпуклая, точка минимума — вырожденная.

Задача 3 (тест-функция Поляка, среднеквадратичная аппроксимация экспонентами, $n = 4$):

$$F(x) = \sum_{j=1}^{10} [\exp(-0.2j) + 2\exp(-0.4j) - x_1 \exp(-0.2jx_2) - x_3 \exp(-0.2jx_4)]^2,$$

$x^0 = (0.5 \ 0 \ 2.5 \ 3)^T$, $x^* = (1 \ 1 \ 2 \ 2)^T$. Функция не выпуклая, с искривленным оврагом, обусловленность велика.

Задача 4 (тест-функция Вуда, $n = 4$):

$$F(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1),$$

$x^0 = (-3 \ -1 \ -3 \ -1)^T$, $x^* = (1 \ 1 \ 1 \ 1)^T$, $F(x^*) = 0$. Функция имеет несколько локальных минимумов, это обстоятельство может вызвать преждевременное окончание процесса.

Задача 5 (тест-функция Степенная, $n = 2$):

$$F(x) = (10(x_1 - x_2)^2 + (x_1 - 1)^2)^4,$$

$$x^0 = (-1.2 \ 0)^T, \quad x^* = (1 \ 1)^T, \quad F(x^*) = 0.$$

Завершение итераций производилось с использованием рекомендаций из [Свириденко, КиМ, 2015]:

$$|F(x^{k-1}) - F(x^k)| < 2^{-\tau_F} (1 + |F(x^{k-1}) - F(x^k)|), \quad (1.1.5)$$

$$\|h^k\| \leq 2^{-\frac{\tau_F}{3}} (1 + |F(x^{k-1}) - F(x^k)|), \quad (1.1.6)$$

$$\|h^k\| \leq 2^{-\frac{\tau_F}{3}} (1 + |F(x^k)|), \quad (1.1.7)$$

что является ужесточением условий останова из [Гилл и др., 1985], позволяющим повысить точность решения, и хорошо согласуется с численными экспериментами [Зеленков и др., 2013].

Программные реализации ньютоновских и квазиньютоновских методов. Приложения Nmbm (Свириденко А. Б. Свидетельство № 2015610399 от 12.01.2015) и NmbmApp (Свириденко А. Б., Зеленков Г. А. Свидетельство № 2015610347 от 12.01.2015) — программная реализация ньютоновских методов оптимизации с регулировкой шага, основанных на факторизации Холесского.

NmbmApp отличается от своего классического прототипа Nmbm конечно-разностной аппроксимацией первых и вторых производных. Все версии алгоритмов реализованы на языке Visual Basic.NET, среда разработки — Microsoft Visual Studio 2010. Исходные тексты программ доступны на сайте журнала «Компьютерные исследования и моделирование» (<http://cfm.ics.org.ru>) и могут быть использованы по универсальной общедоступной лицензии GNU.

В верхней части окна располагаются описания всех задач, сохраненных пользователями в информационной базе данных. Запустить выбранную задачу на решение можно нажатием кнопки «Решить задачу». Ниже дан пример результатов решения одной из тестовых задач, сохраненных в информационной базе данных (рис. 1.1.2).

Ниже представлено окно настроек алгоритма (рис. 1.1.3). Помимо указания параметров алгоритма осуществляется и ввод данных о задаче при переходе по следующим кнопкам:

F — функция (рис. 1.1.4);

- x — стартовая точка (рис. 1.1.5);
 h — градиент (рис. 1.1.6);
 H — матрица вторых производных (гессиан) (рис. 1.1.7).

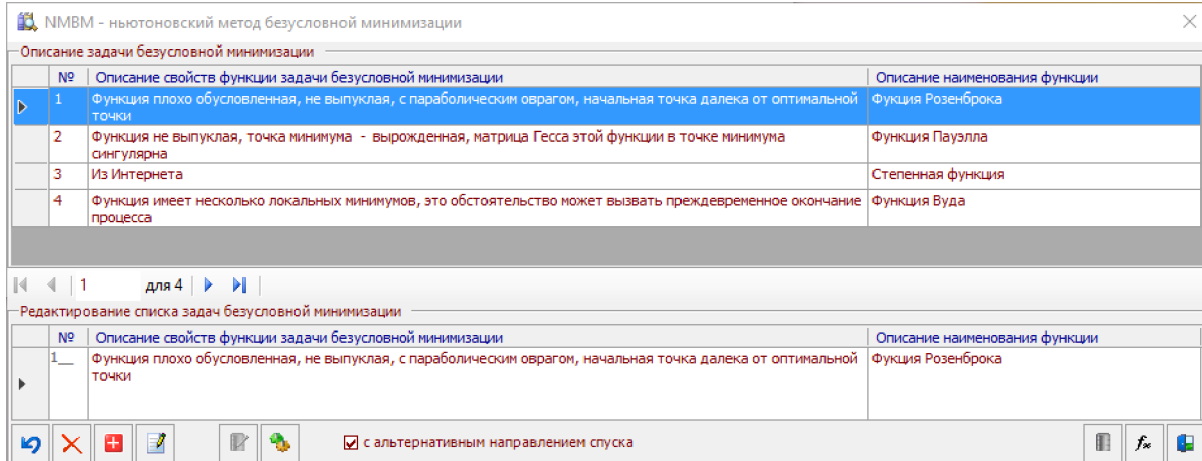


Рис. 1.1.1. Главное окно приложения Nmbm

Назначение кнопок приложения:

- отменить
- удалить
- добавить
- редактировать
- сохранить в локальной базе данных
- настройки алгоритма
- сохранить в информационной базе данных
- решить задачу
- закрыть

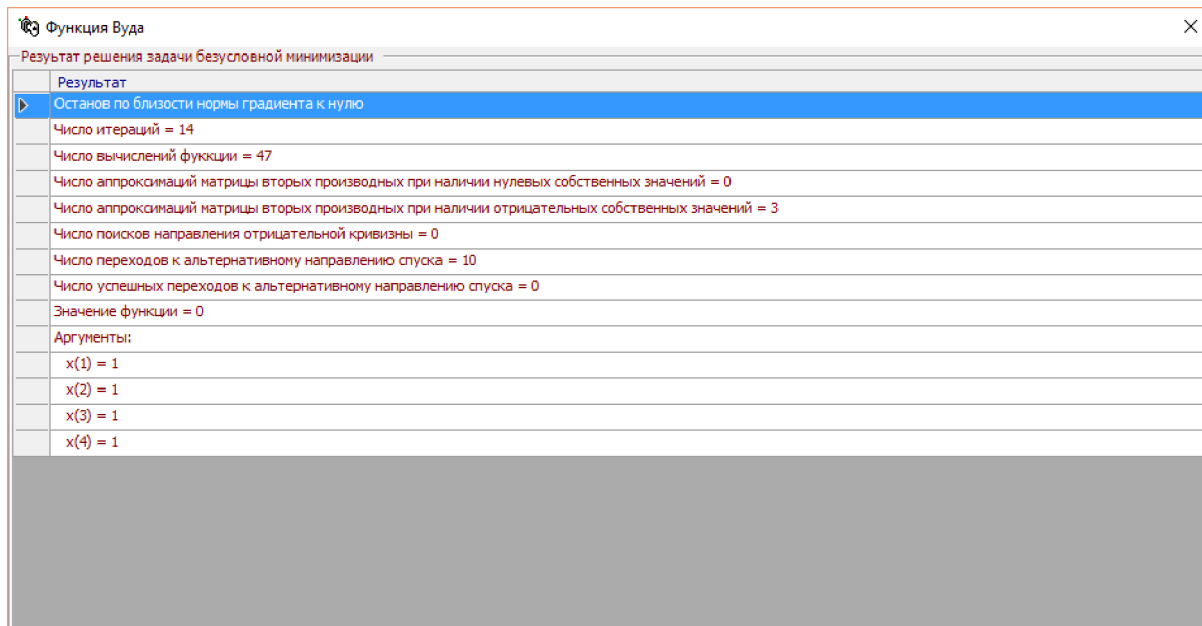


Рис. 1.1.2. Результаты решения задачи 4 (тест-функция Вуда)

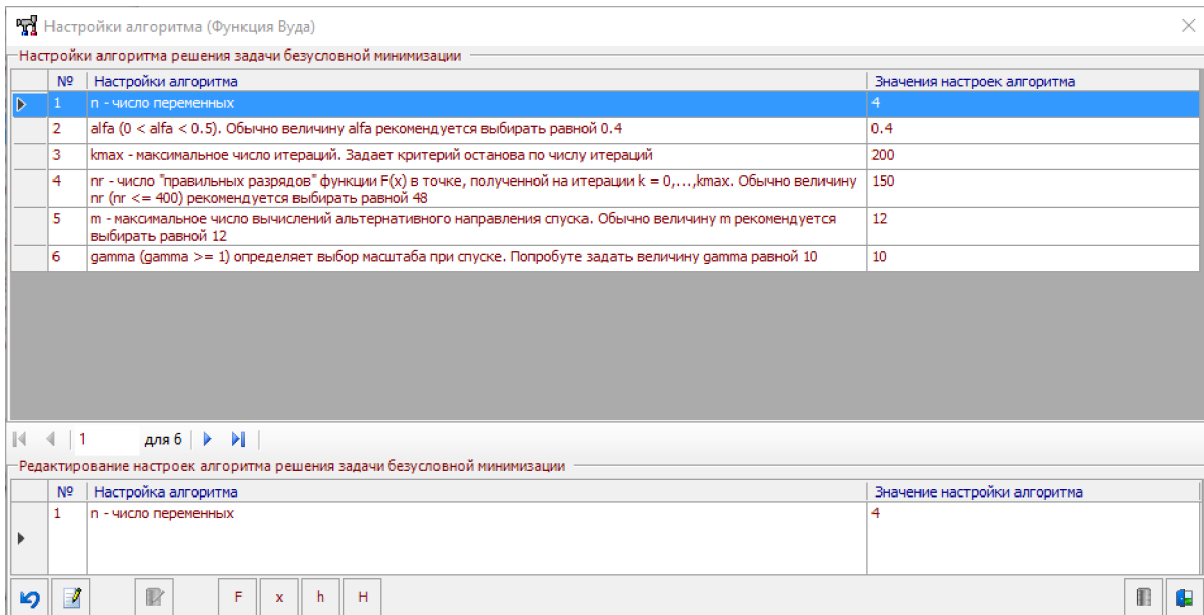


Рис. 1.1.3. Настройки алгоритма

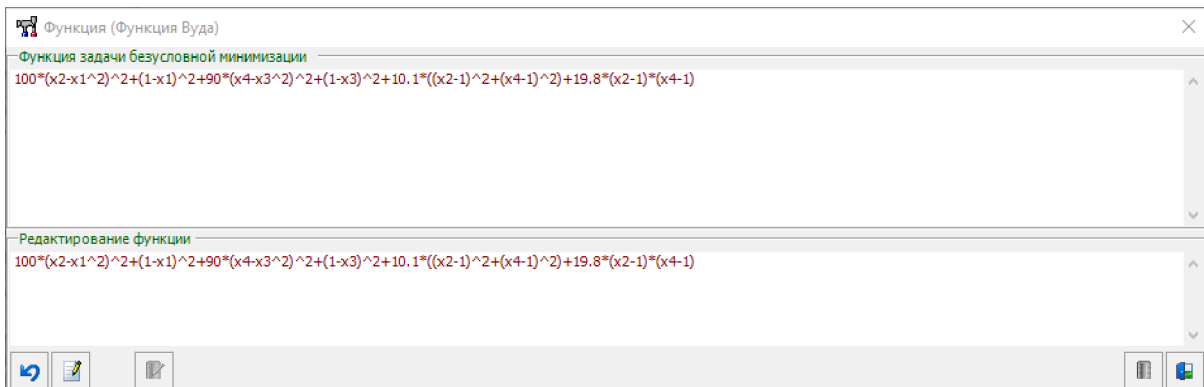


Рис. 1.1.4. Задание функции

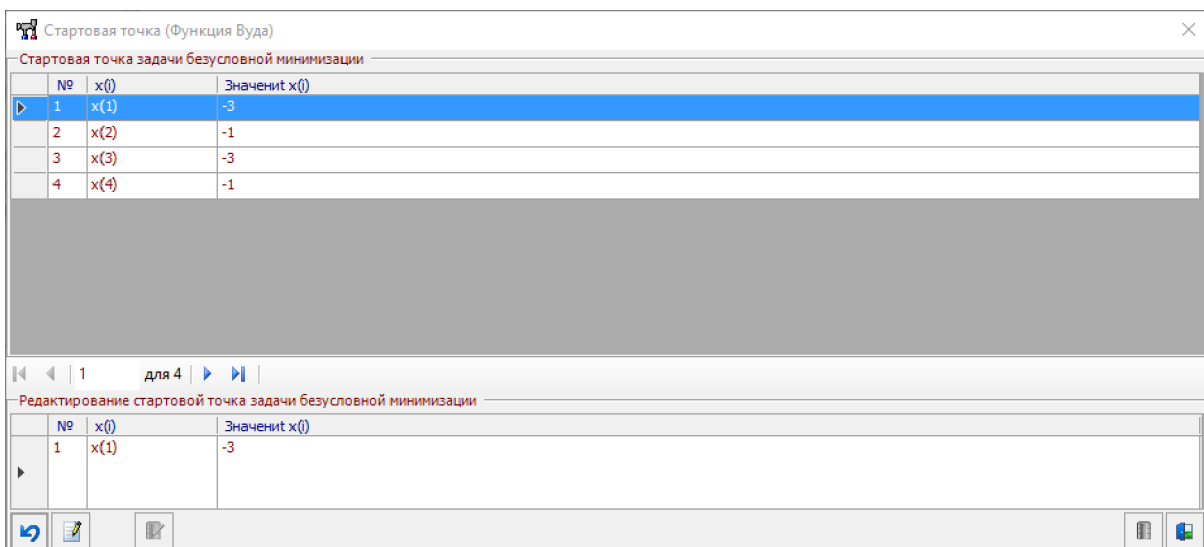


Рис. 1.1.5. Задание начальной точки

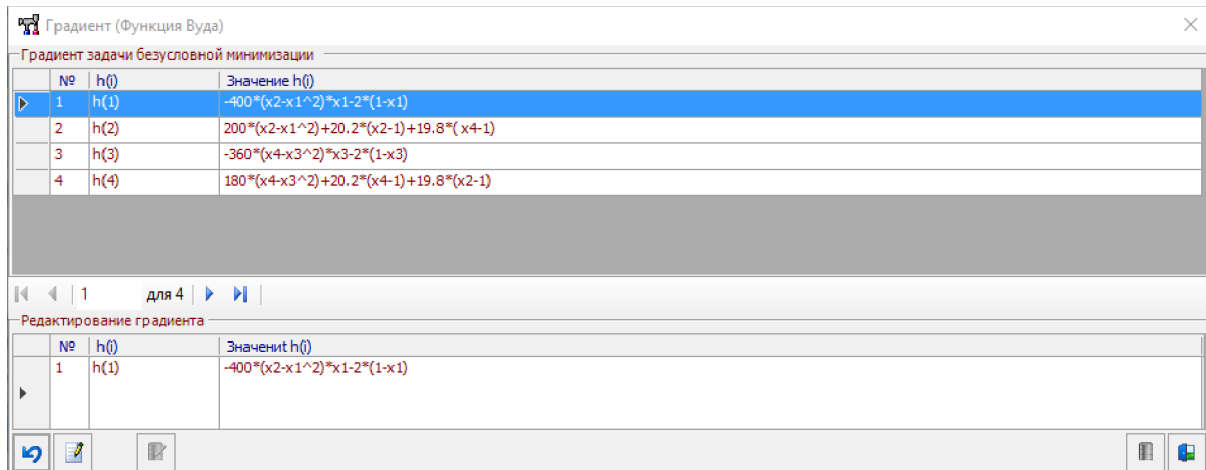


Рис. 1.1.6. Задание градиента функции

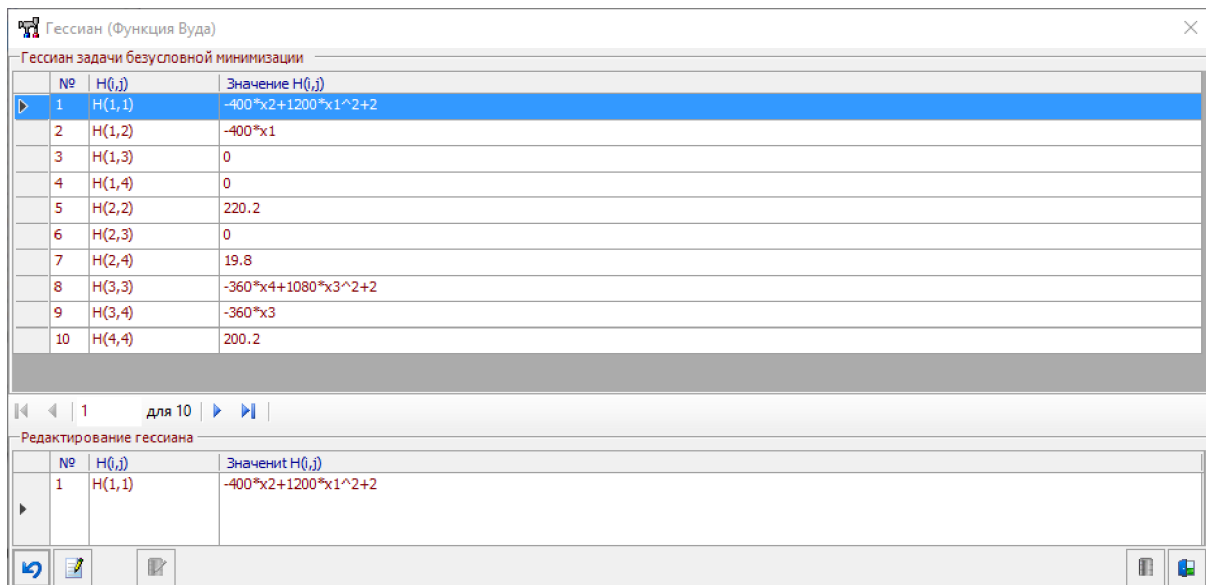


Рис. 1.1.7. Задание матрицы вторых производных

Приложения KNmbm и KNmbmApp — программная реализация квазиньютоновских методов оптимизации с регулировкой шага, основанных на факторизации Холецкого. Интерфейс аналогичен рассмотренному выше.

KNmbmApp отличается от своего классического прототипа KNmbm конечно-разностной аппроксимацией первых производных. Все версии алгоритмов реализованы на языке Visual Basic.NET, среда разработки — Microsoft Visual Studio 2010. Исходные тексты программ доступны на сайте журнала «Компьютерные исследования и моделирование» (<http://crm.ics.org.ru>) и могут быть использованы по универсальной общедоступной лицензии GNU.

Варианты квазиньютоновского метода:

- симметричная формула ранга один;
- BFGS (Бройдена–Флетчера–Гольдфарба–Шанно);
- DFP (Дэвидона–Флетчера–Пауэлла);
- PSB (Пауэлла–Бройдена).

Замечания к разделу 1. Предложенный подход к вычислению длины шага α_k можно найти в [Хакимова и др., 2010; Хакимова, Дикусар и др., 2010]. В основе лежит интеграция

подходов Пшеничного [Пшеничный и др., 1975], Полака [Полак, 1974], Карманова [Карманов, 1975], Гилла, Мюррея и Райт [Гилл и др., 1985].

Предложенный подход к модификации критериев останова, разработанных Гиллом, Мюрреем и Райт [Гилл и др., 1985], можно найти в [Свириденко, КиМ, 2015].

Предложенный подход к оцениванию конечно-разностных интервалов и подход к конечно-разностной аппроксимации первых и вторых производных можно найти в [Зеленков и др., 2013]. В основе лежит интеграция подходов Полака [Полак, 1974], Гилла, Мюррея и Райт [Гилл и др., 1985]. Что же касается способов выбора конечно-разностных интервалов, то они практически ограничены работами Полака [Полак, 1974], Гилла, Мюррея и Райт [Гилл и др., 1985]. Это связано с тем, что стандартные программы численного дифференцирования нацелены на поиск конечно-разностного интервала, обеспечивающего минимальную суммарную ошибку вычисляемого приближения. Алгоритмы выбора интервала для центральной конечно-разностной формулы можно найти, например, у Стиплмена и Винарски [Stempleman et al., 1979]. Однако для оптимизационных приложений такие программы не подходят, так как, обеспечивая избыточную точность вычисления оценок производных, они обычно требуют слишком большого числа обращений к процедуре расчета значений функции.

Замена градиентов их конечно-разностными оценками лучше всего проходит в квазиньютоновских методах. Для методов ньютоновского типа результаты получаются менее удовлетворительными.

2. Развитие подходов и численные исследования

Гилл и Мюррей в [Гилл и др., 1985] доказали, что значение нормы поправки E^k можно дополнительно уменьшить, если использовать симметричные перестановки строк и столбцов H^k на шаге 2 вычислительной схемы 1.1 модифицированного разложения Холесского следующим образом. Найти индекс q такой, что $|c_{qq}^k| = \max_{i \leq j \leq n} |c_{ij}^k|$, и поменять местами все данные, отвечающие строкам матрицы H^k с номерами q и i , а затем проделать то же самое с данными, отвечающими ее q -му и i -му столбцам.

Развитие подхода к построению вычислительной схемы 1.2

Подход Гилла и Мюррея к дополнительному уменьшению нормы поправки E^k развит в [Свириденко, КиМ, 2015]. Доказано что фактическое значение нормы поправки E^k можно дополнительно уменьшить, если использовать симметричные перестановки строк и столбцов H^k . На очередном i -м шаге факторизации в качестве i -й строки и i -го столбца надо брать ту из нетронутых пар, для которой величина $|c_{ii}^k| + |c_i^k|$ максимальна. Такая стратегия дает возможность увеличить численную устойчивость расчета элементов D^k , U^k , u^k и остается работоспособной и в том случае, когда $\|h^k\|$ не равна нулю, но очень мала.

Там же отмечено: «Параметр β подбирается путем минимизации априорной оценки нормы поправки E^k и с очевидностью зависит от способа задания направления спуска и, как следствие, от величин c_j^k , g_i^k . Рассмотрение предлагаемой формулы вычисления величины β , учитывающей такую зависимость, не является целью данной работы и находится в стадии экспериментального сравнения эффективности».

В данной работе величину β предлагается вычислять по формуле

$$\beta^2 = \max \{ \gamma^0, \rho, \xi/v, \varepsilon_M \}, \quad (2.1)$$

где $\rho = \max_{1 \leq i \leq n} |h_i^k|$, ε_M — машинная точность. ε_M вводится в формулу расчета β , чтобы обеспечить устойчивость вычислений, когда норма H^k очень мала. Применение формулы (2.1) хорошо согласуется с численными экспериментами, приведенными в данной работе.

Ниже приводится предлагаемая модификация вычислительной схемы 1.2, дана наиболее эффективная схема организации расчетов. Все фигурирующие в ней величины при реализации на ЭВМ могут размещаться в памяти, первоначально выделяемой для записи матриц H^k , h^k .

Вычислительная схема 2.1 модифицированного разложения Холецкого

Шаг 0. Вычислить $\beta^2 = \max\{\gamma^0, \rho, \xi/v, \varepsilon_M\}$, где $\rho = \max_{1 \leq i \leq n} |h_i^k|$, $v = \max\{1, \sqrt{n^2 - 1}\}$, а числа γ^0 и ξ суть максимальные значения модулей диагонального и недиагонального элементов H^k , ε_M — машинная точность. Положить $\varepsilon_0 = 2^{-\frac{\tau_F}{2}}$, $\varepsilon_s = 2^{-\frac{\tau_F}{3}} (1 + \|F(x^{k-1}) - F(x^k)\|)$. Здесь τ_F — число правильных разрядов $F(x^k)$, которые хотелось бы получить.

Шаг 1. Положить $i = 1$, положить $c_{jj}^k = h_{jj}^k, j = 1, \dots, n$, $c_j^k = -h_j^k, j = 1, \dots, n$.

Шаг 2. Найти индекс q такой, что $|c_{qq}^k| = \max_{i \leq j \leq n} (|c_{ij}^k| + |c_j^k|)$, и поменять местами все данные, отвечающие строкам матрицы H^k с номерами q и i , а затем проделать то же самое с данными, отвечающими ее q -му и i -му столбцам.

Шаг 3. Если $c_{ii}^k < \varepsilon_0$, то положить $l_i^k = \delta$; иначе положить $l_i^k = c_{ii}^k$.

Шаг 4. Если $|c_{ii}^k| \leq \varepsilon_0$, то положить $n_0 = n_0 + 1$ и перейти к шагу 8; иначе перейти к шагу 5.

Шаг 5. Если $c_{ii}^k > 0$, то перейти к шагу 8; иначе перейти к шагу 6.

Шаг 6. Положить $n_o = n_o + 1$. Если $\|h^k\| > \varepsilon_s$, то перейти к шагу 8; иначе перейти к шагу 7.

Шаг 7. Построить систему уравнений $U^k p^k = e_i$ для поиска направления отрицательной кривизны по формулам $u_j^k = 0, j = 1, \dots, n$, $u_i^k = 1$, положить $n_s = n_s + 1$ и остановиться.

Шаг 8. Если $i = n$, то положить $\theta_i = 0$; иначе вычислить вспомогательные величины $c_{ij}^k, j = i + 1, \dots, n$, по формуле $c_{ij}^k = h_{ij}^k - \sum_{s=1}^{i-1} u_{si}^k u_{sj}^k d_{ss}^k$ и найти $\theta_i = \max_{i+1 \leq j \leq n} |c_{ij}^k|$.

Шаг 9. Вычислить диагональный элемент фактора D^k и элемент вектора u^k по формулам $d_{ii}^k = \max\{\delta, |c_{ii}^k|, \theta_i^2 / \beta^2, |c_i^k|, \theta_i\}$, $u_i^k = c_i^k / d_{ii}^k$. Вычислить $g_i^k = d_{ii}^k / l_i^k$ (если $d_{ii}^k > l_i^k \gamma$, взять $g_i^k = l_i^k / \gamma$). Если $i = n$, то перейти к шагу 11.

Шаг 10. Вычислить строку $u_{ij}^k = c_{ij}^k / d_{ii}^k, j = i + 1, \dots, n$, фактора U^k , пересчитать вспомогательные величины по формулам $c_{jj}^k = c_{jj}^k - u_{ij}^k c_{ij}^k, j = i + 1, \dots, n$, $c_j^k = c_j^k - u_{ij}^k c_i^k, j = i + 1, \dots, n$, положить $i = i + 1$ и перейти к шагу 2.

Шаг 11. Найти $\gamma^k = \max_{1 \leq i \leq n} g_i^k$, положить $u^k = \gamma^k u^k$ и остановиться.

В таблице 2.1 приводятся результаты численных исследований, выполненных в двух версиях предложенного ньютоновского метода безусловной минимизации: Nmbm и NmbmApp [Свириденко, 2015; Свириденко и др., 2015; Зеленков и др., 2013].

Достоинством подхода, основанного на применении формул численного дифференцирования, кроме его универсальности является низкая стоимость подготовки задачи к компьютерному моделированию [Черноруцкий, 2011]. От пользователя требуется лишь написание програм-

мы для вычисления значения $F(x)$ при заданном x , поэтому реализованные на основе численных производных методы оптимизации оказываются, по существу, прямыми методами, то есть методами, не использующими в своей схеме производные от $F(x)$, в то время как результаты численных исследований, приведенные в таблице 2.1, показывают, что конечно-разностные аналоги проигрывают в точности решения задач оптимизации. Предлагаемый подход к увеличению точности, в основе которого лежит интеграция ньютоновских и квазиньютоновских направлений поиска, обсуждается ниже.

Таблица 2.1

Метод	Функция	k	δF	δx	k_0
Nmbm	Розенброка	11	0	0	17
NmbmApp	Розенброка	20	0	0	313
Nmbm	Пауэлла	4	0	0	6
NmbmApp	Пауэлла	4	0	0	108
NmbmApp	Поляка	40	$4.23 \cdot 10^{-33}$	0	1341
Nmbm	Вуда	13	0	0	36
NmbmApp	Вуда	14	$4.4 \cdot 10^{-27}$	$2 \cdot 10^{-14}$	470
Nmbm	Степенная	12	0	0	163
NmbmApp	Степенная	13	$4.3 \cdot 10^{-61}$	$8 \cdot 10^{-9}$	323

Развитие подхода Городецкого и Гришагина. В основе подхода к дальнейшему ослаблению отмеченных выше недостатков квазиньютоновских методов лежит предлагаемое модифицированное разложение Холесского квазиньютоновской матрицы H^k (вычислительная схема 2.1). Такой подход к построению квазиньютоновского поиска определяет и решение проблемы масштабирования шагов при спуске, и аппроксимацию неквадратичными функциями, и интеграцию с методом доверительной окрестности. Кроме этого, фактическое значение нормы поправки E^k можно дополнительно уменьшить, если использовать симметричные перестановки строк и столбцов H^k .

В таблице 2.2 приводятся результаты численных исследований, выполненных в двух версиях предлагаемого квазиньютоновского метода безусловной минимизации: KNmbm и KNmbmApp.

Таблица 2.2

Метод	Функция	k	δF	δx	k_0	Вариант квазиньютоновского метода
KNmbm	Розенброка	13	0	0	35	PSB
KNmbmApp	Розенброка	29	0	0	525	PSB
KNmbm	Пауэлла	4	0	0	6	Симметричная формула ранга один
KNmbmApp	Пауэлла	4	0	0	113	Симметричная формула ранга один
KNmbmApp	Поляка	45	$4.23 \cdot 10^{-33}$	0	1457	DFP
KNmbm	Вуда	16	0	0	50	PSB
KNmbmApp	Вуда	14	0	0	470	PSB
KNmbm	Степенная	17	0	0	222	DFP
KNmbmApp	Степенная	79	0	0	1719	DFP

Сравнительные численные исследования KNmbm и KnmbmApp с системой квазиньютоновской оптимизации ProfMiniHP. Программная система квазиньютоновской оптимизации ProfMiniHP, разработанная Яновским, предназначена для решения задач оптимизации с высокой точностью [Яновский, 2005]. Система ProfMiniHP включает 77 программных модулей, расположенных на 7 уровнях вложенности и представленных 3455 операторными строками языка Borland C++ v5.02.

Подсистемы программных модулей ProfMiniHP:

- расчет вектор-градиентов на основе адаптивных правой и центральной конечно-разностной аппроксимаций, а также экстраполяции Ричардсона [Мэтьюз и др., 2001];
- начального задания обратной матрицы вторых производных на основе единичной матрицы или модификаций Денниса–Шнабеля–Фрэнка и Шанно–Фуа [Деннис и др., 1988];
- рекуррентной оценки обратных квазиньютоновских матриц методами DFP, проективного Ньютона–Рафсона, Бройдена, Пирсона 2 и 3 [Fletcher et al., 1969];
- численного анализа обратной квазиньютоновской матрицы на основе сингулярного разложения и оценки ее обусловленности, используемой для выбора уровня регуляризации обратной квазиньютоновской матрицы и ее реализации на основе подхода [Яновский, 2005], опирающегося на результаты Гершгорина [Форсайт и др., 1980; Хорн и др., 1989];
- одномерной оптимизации на основе методов модифицированного золотого сечения, параболической аппроксимации Мэтьюза–Финка, а также регуляризованных методов Брента–Форсайта–Малькольма–Моулера и Дэвиса–Свенна–Кемпи–Пауэлла [Мэтьюз и др., 2001; Форсайт и др., 1980; Brent, 1973; Гилл и др., 1985];
- многоаспектной проверки условий останова на основе регуляризованных критериальных функций, опирающихся на относительную машинную точность вычислений;
- масштабирования пространства независимых переменных и значений функции.

Для численного исследования системы квазиньютоновской оптимизации ProfMiniHP и подсистем квазиньютоновской оптимизации MatLab, Mathematica, MathCAD на тест-функциях Вуда и Пауэлла Яновским было выполнено более 300 численных экспериментов с позиций точности и надежности сходимости численного решения к теоретическому, что позволило ему сделать следующие выводы [Яновский, 2005].

Результаты работы пакетов оптимизации MatLab и Mathematica показали, что с их помощью не всегда можно добиться корректного решения задачи оптимизации. Так, при минимизации тест-функции Вуда из начальной точки $(-3 \ -1 \ -3 \ -1)^T$ пакет оптимизации системы MatLab не смог найти точку глобального минимума, указав точку, близкую к псевдорешению. Пакет оптимизации системы Mathematica не достиг верного решения в задаче с той же функцией из начальной точки $(2 \ -1 \ -3 \ -1)^T$ и, более того, не смог даже начать решать задачу минимизации тест-функции Пауэлла для начальной точки $(1 \ 1 \ 1 \ 1)^T$.

В целом лучшим пакетом оптимизации среди протестированных коммерческих систем обладает система MathCAD — все корректные результаты. Однако при минимизации тест-функции Вуда из точки $(-3 \ -1 \ -3 \ -1)^T$ установлена низкая точность оценки минимума целевой функции, отклоняющаяся от истинного значения уже в третьем знаке после запятой.

Таким образом, результаты данного сравнительного исследования позволяют оценить систему оптимизации ProfMiniHP как предпочтительную по сравнению с пакетами оптимизации систем MatLab, Mathematica, MathCAD.

В таблицах 2.3–2.6 приведены сравнения с результатами численного исследования системы квазиньютоновской оптимизации ProfMiniHP и подсистем квазиньютоновской оптимизации MatLab, Mathematica, MathCAD, проведенные Яновским [Яновский, 2005] на тест-функциях Вуда и Пауэлла с позиций точности и надежности сходимости численного решения к теоретическому.

Для тест-функции Вуда фрагменты результатов исследований приведены в таблицах 2.3–2.4, причем в таблице 2.3 отражены результаты оптимизации из начальной точки $(-3 \ -1 \ -3 \ -1)^T$, а в таблице 2.4 — из точки $(2 \ -1 \ -3 \ -1)^T$.

Для тест-функции Пауэлла фрагменты результатов исследований приведены в таблицах 2.5–2.6, причем в таблице 2.5 отражены результаты оптимизации из начальной точки $(2 \ -1 \ -3 \ -1)^T$, а в таблице 2.6 — из точки $(1 \ 1 \ 1 \ 1)^T$.

Таблица 2.3

Метод	δF	δx
MatLab	7,3825	2,364
Mathematica	$4,6756 \cdot 10^{-2}$	0,19584
MathCAD	$1,154 \cdot 10^{-3}$	$1,6 \cdot 10^{-2}$
ProfMiniHP	$4,95 \cdot 10^{-8}$	$2,35 \cdot 10^{-4}$
KNmbm	0	0
KNmbmApp	0	0

Таблица 2.4

Метод	δF	δx
MatLab	$3 \cdot 10^{-9}$	0,000
Mathematica	1,71832	1,0646
MathCAD	$2 \cdot 10^{-9}$	0,000
ProfMiniHP	$9 \cdot 10^{-8}$	$3,13 \cdot 10^{-4}$
KNmbm	0	0
KNmbmApp	0	0

Таблица 2.5

Метод	δF	δx
MatLab	$1 \cdot 10^{-6}$	$2,51 \cdot 10^{-2}$
Mathematica	0,000000	0,000000
MathCAD	$2,1 \cdot 10^{-5}$	$5,7 \cdot 10^{-2}$
ProfMiniHP	$1,8 \cdot 10^{-8}$	$8,96 \cdot 10^{-3}$
KNmbm	0	0
KNmbmApp	0	0

Таблица 2.6

MatLab	$6 \cdot 10^{-6}$	$3,5 \cdot 10^{-2}$
Mathematica	122	1
MathCAD	$3,17 \cdot 10^{-4}$	$5,9 \cdot 10^{-2}$
ProfMiniHP	$9 \cdot 10^{-10}$	$1,93 \cdot 10^{-3}$
KNmbm	0	0
KNmbmApp	0	0

Таким образом, результаты данного сравнительного исследования позволяют оценить описанные выше программные реализации квазиньютоновских методов оптимизации с регулируемой шага, основанных на факторизации Холесского, как более предпочтительные по сравнению с системой квазиньютоновской оптимизации ProfMiniHP-2004 и с подсистемами квазиньютоновской оптимизации MatLab, Mathematica, MathCAD.

Замечания к разделу 2. Результаты численных исследований, приведенные в таблице 2.2, показали, что предлагаемый квазиньютоновский метод с конечно-разностной аппроксимацией первых производных не уступает в точности решения тестовых задач ньютоновскому методу с аналитическим вычислением первых и вторых производных.

В настоящее время на основе экспериментальных данных принята точка зрения, согласно которой ньютоновские методы с конечно-разностной аппроксимацией H^k на основе аналитических выражений для h^k надежнее, чем квазиньютоновские методы, и сходятся в более широком классе случаев, в частности в задачах с очень плохой обусловленностью H^k . С другой стороны, согласно тем же источникам, если аналитические выражения для h^k отсутствуют и применяются версии нулевого порядка соответствующих алгоритмов, в большинстве случаев

более эффективными оказываются квазиньютоновские методы с конечно-разностной аппроксимацией [Черноруцкий, 2013]. Данные замечания не только отражают сравнительные оценки, основанные на опыте проведения реальных вычислений, но и являются основанием для построения вычислительных схем квазиньютоновского типа.

Описание вариантов квазиньютоновских методов, не укладывающихся в классическую схему, задаваемую семейством методов Бройдена, которые дают результаты расчетов в среднем не хуже, а для отдельных целевых функций лучше, чем BFGS, можно найти в [Ветошкин, 2010].

3. Интеграция направлений поиска

Результаты численных исследований, приведенные в таблице 2.1, показывают, что алгоритм с конечно-разностной аппроксимацией первых и вторых производных проигрывает в точности решения задач оптимизации методу с аналитическим вычислением первых и вторых производных, однако точность можно увеличить проведением на итерациях $k+1$ «альтернативной процедуры спуска» в тех случаях, когда регулярная процедура поиска направления спуска не дает существенных результатов.

Если аналитические выражения для h^k отсутствуют и применяются версии нулевого порядка соответствующих алгоритмов, в большинстве случаев более эффективными оказываются квазиньютоновские методы с конечно-разностной аппроксимацией h^k [Черноруцкий, 2013]. Данные замечания не только отражают сравнительные оценки, основанные на опыте проведения реальных вычислений, но и являются основанием для выбора квазиньютоновского направления в качестве альтернативного направления спуска.

Достоинства ньютоновских методов с конечно-разностной аппроксимацией первых и вторых производных те же, что и у обычных методов ньютоновского типа, — высокая скорость сходимости и способность «ощущать» приближение седловой точки и уходить от нее. К недостатку следует отнести необходимость вычисления на каждой итерации n значений градиента, которые требуются для построения n столбцов матрицы вторых производных, когда в качестве конечно-разностных направлений используются столбцы единичной матрицы.

Для существенного ослабления отмеченного выше недостатка предлагается переход от проведения «альтернативной процедуры спуска» в тех случаях, когда регулярная процедура поиска направления спуска не дает существенных результатов, к построению на итерациях k ньютоновского, а на итерациях $k+1$ квазиньютоновского направления спуска p^{k+1} :

$$H^{k+1} p^{k+1} = -h^{k+1}. \quad (3.1)$$

В данной работе выбор варианта квазиньютоновского метода определяет симметричная формула ранга один:

$$H^{k+1} = H^k + \frac{1}{(y^k - B^k s^k)^T s^k} (y^k - B^k s^k)(y^k - B^k s^k)^T \quad (3.2)$$

для пересчета матрицы H^k , где $y^k = h^{k+1} - h^k$, $s^k = x^{k+1} - x^k = \alpha_k p^k$ и вычисляется в результате шага α^k вдоль направления p^k , поэтому формулу можно упростить, используя равенство $B^k s^k = -\alpha_k h^k$ [Гилл и др., 1985]. В этом случае (3.2) преобразуется к виду

$$H^{k+1} = H^k + c_k z^k (z^k)^T, \quad (3.3)$$

где $z^k = y^k + \alpha_k h^k$, $c_k = 1/\alpha_k (z^k)^T p^k$. На итерациях $k+1$ $(U^k)^T D^k U^k$ -разложение матрицы H^k известно, поэтому

$$H^{k+1} = (U^k)^T D^k U^k + c_k z^k (z^k)^T = (U^k)^T (D^k + c_k v^k (v^k)^T) U^k, \quad (3.4)$$

где v^k — решение треугольной системы $(U^k)^T v^k = z^k$. При этом структура матрицы

$$D^k + c_k v^k (v^k)^T \tag{3.5}$$

настолько проста, что для элементов ее факторов Холецкого, которые обозначим через \bar{U}^{k+1} и \bar{D}^{k+1} , можно выписать нерекуррентные формулы. Таким образом,

$$H^{k+1} = (U^k)^T (\bar{U}^{k+1})^T \bar{D}^{k+1} \bar{U}^{k+1} U^k = (U^{k+1})^T D^{k+1} U^{k+1}, \tag{3.6}$$

где через U^{k+1} обозначим произведение $\bar{U}^{k+1} U^k$, а $D^{k+1} = \bar{D}^{k+1}$. Поскольку результатом перемножения двух верхних треугольных матриц с диагональными элементами, равными единице, является матрица того же вида, полученное равенство и есть искомое разложение. Вычисления здесь можно организовать таким образом, что расчет компонент вектора v^k и умножение \bar{U}^{k+1} на U^k будут вестись параллельно, причем потребуют порядка n^2 операций. В данной работе для пересчета матричных разложений использовался экономный алгоритм вычисления факторов Холецкого, предложенный в [Gill et al., 1974].

Ниже в таблице 3.1 приведены результаты исследования предлагаемой схемы нахождения направления спуска.

Таблица 3.1

Функция	k	δF	δx	k_0
Розенброка	26	0	0	567
Пауэлла	4	0	0	117
Поляка	66	$4.23 \cdot 10^{-33}$	0	1773
Вуда	23	0	0	730
Степенная	68	0	0	959

Выбор способа задания направления спуска тем актуальнее, чем выше размерность задачи. Поэтому возьмем тест-функцию переменной размерности аргумента, например расширенную функцию Розенброка [Деннис и др., 1988], и исследуем зависимость сходимости от размерности тестовой задачи:

$$\min_{x \in R^n} F(x) = \min_{x \in R^n} \sum_{i=1}^n f_i^2(x). \tag{3.7}$$

Расширенная функция Розенброка:

- n — любое положительное целое, кратное 2;
- для $i = 1, \dots, n/2$

$$f_{2i-1}(x) = 10(x_{2i} - x_{2i-1}^2),$$

$$f_{2i}(x) = 1 - x_{2i-1};$$

- $x^0 = (-1, 2 \ 1 \ \dots \ -1, 2 \ 1)^T$;
- $x^* = (1 \ 1 \ \dots \ 1 \ 1)^T$.

Ниже, в таблице 3.2, приведены результаты исследования зависимости сходимости от размерности n расширенной функции Розенброка, выбор варианта квазиньютоновского метода определяет формула Дэвидона–Флетчера–Пауэлла.

Таблица 3.2

n	k	δF	δx	k_0
2	30	0	0	628
4	30	0	0	1065
6	34	0	0	1756
8	33	0	0	2256

Замечания к разделу 3. Разложение Холецкого матриц H^{k+1} позволяет получать оценки их чисел обусловленности. Например, если максимальный и минимальный элементы фактора D^{k+1} равны d_{\max}^{k+1} и d_{\min}^{k+1} , то $\text{cond}(H^{k+1}) \geq d_{\min}^{k+1}/d_{\max}^{k+1}$ [Гилл и др., 1985]. На промежуточных итерациях такие оценки требуются для выявления ситуаций, когда матрица H^{k+1} обусловлена настолько плохо, что в численном решении p^{k+1} системы (3.1), скорее всего, не окажется ни одной правильной цифры. При этом факторы разложения можно подправить так, что обусловленность результирующей матрицы станет приемлемой.

Факторизация Холецкого матриц H^{k+1} позволяет избежать потери положительной определенности. Здесь все зависит от значения диагональных элементов фактора D^{k+1} . Пока они больше нуля, положительная определенность квазиньютоновской матрицы H^{k+1} гарантирована. Контролируя эти значения при пересчете диагонального фактора, всегда можно выявить опасность потери положительной определенности в тот момент, когда она возникает, и принять своевременные меры.

Применение факторизации Холецкого для построения численно устойчивых реализаций квазиньютоновского поиска можно найти в [Gill et al., 1972; Fletcher et al., 1974; Gill et al., 1975; Brodlie, 1977].

Заключение

Доказано, что интеграция техники исключения Гаусса и факторизации Холецкого определяет решение проблемы масштабирования шагов при спуске, а следовательно, и аппроксимацию не квадратичными функциями, и интеграцию с методом доверительной окрестности [Зеленков и др., 2013]. Доказано, что интеграция техники исключения Гаусса и факторизации Холецкого определяет и подходы к дальнейшему уменьшению значения нормы поправки E^k , если использовать симметричные перестановки строк и столбцов H^k : на очередном, i -м шаге факторизации в качестве i -й строки и i -го столбца надо брать ту из нетронутых пар, для которой величина $|c_{ii}^k| + |c_i^k|$ максимальна [Свириденко, КиМ, 2015]. Доказано, что модифицированная факторизация Холецкого — оптимизированный алгоритм в том смысле, что параметр β подбирается в нем путем минимизации априорной оценки нормы поправки E^k при условии сохранения существенно положительно-определенной матрицы неизменной и с очевидностью зависит от способа задания направления спуска [Зеленков и др., 2013; Свириденко, КиМ, 2015]. В данной работе доказанные утверждения позволили модифицировать формулу вычисления величины β , предложенную в [Gill et al., 1974], и построить вычислительную схему 2.1 модифицированной факторизации Холецкого. Полученные результаты хорошо согласуются с численными экспериментами, приведенными в данной работе.

Рассмотрен способ задания квазиньютоновского поиска, позволяющий ослабить или снять недостатки квазиньютоновских методов:

- большинство вариантов квазиньютоновских методов при минимизации сильно выпуклых квадратичных функционалов приводят к одной и той же траектории спуска, вырождаясь в хорошо изученные методы сопряженных градиентов;

- предположение о невыпуклости вносит дополнительные трудности: в этих условиях метод сопряженных градиентов по характеристикам сходимости эквивалентен градиентному методу наискорейшего спуска со всеми вытекающими отсюда последствиями;
- кроме отмеченных дефектов, общих для методов сопряженных градиентов и квазиньютоновских методов, последние имеют дополнительные недостатки, связанные с проблемой потери положительной определенности квазиньютоновских матриц из-за накопления вычислительных погрешностей в рекуррентных процедурах аппроксимации матриц вторых производных.

В основе способа задания направления спуска лежит модифицированное разложение Холецкого квазиньютоновской матрицы. Такой подход определяет для квазиньютоновских методов и решение проблемы масштабирования шагов при спуске, и аппроксимацию неквадратичными функциями, и интеграцию с методом доверительной окрестности. Кроме этого, фактическое значение нормы поправки E^k можно дополнительно уменьшить, если использовать симметричные перестановки строк и столбцов квазиньютоновской матрицы H^k . Результаты численных исследований, приведенные в данной работе, показали, что предлагаемый квазиньютоновский метод с конечно-разностной аппроксимацией первых производных не уступает в точности решения тестовых задач ньютоновскому методу с аналитическим вычислением первых и вторых производных.

Изучен подход к увеличению эффективности ньютоновских методов с конечно-разностной аппроксимацией первых и вторых производных. Подход является основой для дальнейших исследований, результаты которых могут быть использованы для построения численных методов квазиньютоновского типа. Результаты численных исследований, приведенные в данной работе, показали, что предлагаемый метод квазиньютоновского типа с конечно-разностной аппроксимацией первых и вторых производных не уступает в точности решения тестовых задач ньютоновскому методу с аналитическим вычислением первых и вторых производных.

Список литературы

- Ветошкин А. М.* Структура квазиньютоновских методов оптимизации // Журнал вычислительной математики и математической физики. — 2010. — Т. 50, № 5. — С. 817–831.
- Гилл Ф., Мюррей У.* Численные методы условной оптимизации. — М.: Мир, 1977.
- Гилл Ф., Мюррей У., Райт М.* Практическая оптимизация. — М.: Мир, 1985.
- Городецкий С. Ю., Гришагин В. А.* Учебный курс «Модели и методы конечномерной оптимизации». Ч. 2. Нелинейное программирование и многоэкстремальная оптимизация. — Нижний Новгород: НГУ им. Лобачевского, 2003.
- Деннис Дж., Шнабель Р.* Численные методы безусловной оптимизации и решения нелинейных уравнений. — М.: Мир, 1988.
- Зеленков Г. А., Хакимова А. Б.* Подход к разработке алгоритмов ньютоновских методов оптимизации, программная реализация и сравнение эффективности // Компьютерные исследования и моделирование. — 2013. — Т. 5, № 3. — С. 367–377.
- Карманов В. Г.* Математическое программирование. — М.: Мир, 1975.
- Мэтьюз Дж. Г., Финк К. Д.* Численные методы. — М.: Изд. дом «Вильямс», 2001.
- Парлетт Б.* Симметричная проблема собственных значений. Численные методы. — М.: Мир, 1983.
- Полак Э.* Численные методы оптимизации. Единый подход. — М.: Мир, 1974.
- Поляк Б. Т.* Введение в оптимизацию. — М.: Наука. Главная редакция физико-математической литературы, 1983.
- Пиеничный Б. Н., Данилин Ю. М.* Численные методы в экстремальных задачах. — М.: Наука, 1975.
- Свириденко А. Б.* Априорная поправка в ньютоновских методах оптимизации // Компьютерные исследования и моделирование. — 2015. — Т. 7, № 4 — С. 835–863.

- Свириденко А. Б.* Программа для ЭВМ «MNB (Ньютоновский метод безусловной оптимизации)». Свидетельство № 2015610399 от 12.01.2015. Официальный бюллетень Федеральной службы по интеллектуальной собственности (РОСПАТЕНТ) «Программы для ЭВМ, Базы данных, Топологии интегральных микросхем». № 2(100). 20.02.2015.
- Свириденко А. Б., Зеленков Г. А.* Программа для ЭВМ «MNBApp (Ньютоновский метод безусловной оптимизации с численным вычислением первых и вторых производных)». Свидетельство № 2015610347 от 12.01.2015. Официальный бюллетень Федеральной службы по интеллектуальной собственности (РОСПАТЕНТ) «Программы для ЭВМ, Базы данных, Топологии интегральных микросхем» №2(100) 2015, 20.02.2015.
- Форсайт Дж., Малькольм М., Моулдер К.* Машинные методы математических вычислений. — М.: Мир, 1980. — 279 с.
- Хакимова А. Б., Дикусар В. В., Зеленков Г. А.* Увеличение эффективности ньютоновских методов оптимизации. Информодинамический подход // Труды ИСА РАН «Динамика неоднородных систем». — Вып. 14-А. — Т. 53. — М.: Книжный дом «ЛИБРОКОМ», 2010. — С. 97–114.
- Хакимова А. Б., Зеленков Г. А.* Увеличение эффективности ньютоновских методов минимизации. Вычисление длины шага // Труды ИСА РАН «Динамика неоднородных систем». — Вып. 14-А, Т. 53. — М.: Книжный дом «ЛИБРОКОМ», 2010. — С. 115–120.
- Хорн Р., Джонсон Ч.* Матричный анализ: Пер. с англ. — М.: Мир, 1989. — 655 с.
- Черноруцкий И. Г.* Методы оптимизации. Компьютерные технологии. — СПб.: БХВ-Петербург, 2011. — 384 с.
- Черноруцкий И. Г.* Практическая оптимизация и невыпуклые задачи // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. — СПб.: Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого». — 2013. — № 4(176). — С. 79–86.
- Яновский Т. А.* Сравнительные численные исследования системы квазиньютоновской оптимизации ProfMiniNP-2004 и подсистем оптимизации MatLab 6, Mathematica 4, MathCAD 8 / Т. А. Яновский // Вестник Волгоградского государственного университета. Серия 9, Исследования молодых ученых. — 2005. — Вып. 4 (Ч. 2. Физико-математические и экономические науки). — С. 23–27.
- Brent R. P.* Algorithms for Minimization without Derivatives, Prentice-Hall, Inc., Englewood Cliffs. New Jersey, 1973.
- Brodli K.W.* An assessment of two approaches to variable metric methods // Math. Prog. — 1977. — Vol. 12. — P. 344–355.
- Fletcher R. and Powell M. J. D.* On the modification of LDL^T factorization // Mathematics of Computation. — 1974. — Vol. 28. — P. 1067–1087.
- Fletcher R., Powell M. J. D.* A rapidly convergent descent method for minimization // Computer J. — 1963. — No. 6; Pearson J. D.; Computer J. — 1969. — No. 13.
- Gill P. E. and Murray W.* Newton-type methods for unconstrained and linearly constrained optimization // Math. Prog. — Vol. 28 — 1974. — P. 311–350.
- Gill P. E. and Murray W.* Quasi-Newton methods for unconstrained optimization // J. Inst. Maths. Applics. — 1972. — Vol. 9. — P. 91–108.
- Gill P. E., Golub G. H., Murray W., and Saunders M. A.* Methods for Modifying Matrix Factorization // Mathematics of computation. — 1974 — Vol. 28, No. 126, — P. 505–535.
- Gill P. E., Murray W., and Saunders M. A.* Methods for computing and modifying the LDV factors of a matrix // Mathematics of Computation. — 1975. — Vol. 29. — P. 1051–1077.
- Nazareth L.* A relationship between the BFGS and conjugate-gradient algorithms and its implications for new algorithms // SIAM J. Numer. Anal. — 1979. — Vol. 16. — P. 794–800.
- Stpleman R. S., Winarsky N. D.* Adaptive numerical differentiation // Mathematics of Computation. — 1979. — Vol. 33. — P. 1257–1264.