Корзун В.

Сходимость квазиньютоновских матриц порожденных методом SR1.

**Данная статья является переводом оригинальной статьи «Convergence of quasi-Newton matrices generated by the symmetric rank one update», A.R. Conn* IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA N.I.M. Gould Rutherford Appleton Laboratory, Chilton, UK Ph.L. Toint Department of Mathematics, Faeultds Universitaires ND de la Paix, B-5000, Namur, Belgium*

**Некоторые названия могут отличаться от принятых, а в некоторых местах перевод может быть неверен, хоть я и старался перевести как надо.*

**Оригинал можно найти по ссылке*
ftp://ftp.numerical.rl.ac.uk/pub/nimg/pubs_old/ConnGoulToin91_mp.pdf

**Перевод сделал Корзун Владислав, МФТИ ФИВТ 397*

Квазиньютоновские алгоритмы безусловной нелинейной минимизации генерируют последовательность матриц, которую можно рассматривать как приближения целевой функции вторых производных. Эта статья дает условия, при которых эти приближения могут быть доказаны для глобально сходящихся к настоящим матрицам Гессе (квадратная матрица, состоящая из вторых частных производных скалярной функции или скалярного поля), в случае использования формулы SR1. Также рассмотрено и доказано, что скорость сходимости улучшается со скоростью сходимости базовых итераций. Теория подтверждается некоторыми численными экспериментами, которые также показывают, что сходимость приближений Гессе существенно медленнее других известных квазиньютоновских формул.

1. <u>Введение</u>

Квазиньютоновские методы признаются сегодня одними из самых эффективных способов решать нелинейные несвязные или связные ограниченные задачи оптимизации. Эти методы в основном используются, когда вторая производная матрицы целевой функции либо отсутствует, либо слишком сложна для вычисления. Эти методы похожи на ньютоновские, но избегают необходимость вычисления матриц Гессе рекурсивно, от итерации к итерации, в симметричную матрицу, которую можно рассматривать как приближение Гессиана. Следовательно, они позволяют эксплуатировать кривизну задачи в численном алгоритме, несмотря на то, что требуются только первые производные (градиенты) и значения функции. Мы отсылаем читателя к [3, 4, 6, 8] для дальнейшего анализа этих ставших уже классическими алгоритмами.

Рассматриваемая задача заключается в нахождении локального решения x' в R^n

$$min\, f(x), \tag{1}$$

где $f(x)$ — гладкая функция из R^n в R , использующая квазиньютоновский метод. Обозначим последовательность итераций { x_k }, генерируемым рассматриваемым методом. Важно отметить, что большая часть теории и практики этой области основывается на “линейном поиске”, модель итераций алгоритма которого имеет вид

$$x_{(k+1)}=x_k-a_k B_k^{-1}\nabla f(x), \tag{2}$$

где B_k - симметричное приближение Гессиана, которое удовлетворяет “уравнению секущей”

$$B_{(k+1)}(X_{(k+1)}-X_k)=\nabla f(X_{(k+1)})-\nabla f(x_k), \tag{3}$$

и где a_k - подходящий параметр релаксации, который вычисляется путем решения (иногда довольно неточно) одномерной задачи

$$\min_{a>0} f(x_k-a d_k), \tag{4}$$

где искомое направление d_k получается из

$$d_k=-B_k^{-1}\nabla f(x_k). \tag{5}$$

Эта последняя процедура часто называют ”линейным поиском” , отсюда и название алгоритмической модели. В этом контексте, сохраняя определенность B_k положительным имеет отличительное преимущество, которое гарантирует, что направление спуска d_k будет сонаправленно с $f(x)$. Эта положительная определенность матриц B_k может быть обеспечена путем введения подходящих условий на значение a_k и используя адекватные соотношения для обновления B_k. В этом контексте, обновление матриц B_k с помощью формулы Бройдена-Флетчера-Голдфарба-Шэнно (BFGS) принимает вид

$$B_{(k+1)}=B_k-\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}+\frac{y_k y_k^T}{y_k^T s_k}, \tag{6}$$

где

$$s_k:=x_{(k+1)}-x_k \tag{7}$$

и

$$y_k:=\nabla f(x_{(k+1)})-\nabla f(x_k), \tag{8}$$

Эта формула единогласно рекомендовалась в течение последних 15 лет. Поддержание положительной определенности приближений к матрице Гессе было, однако, были и некоторые важные недостатки. Первый заключался в том, что даже если истинный Гессиан обычно положительно определенный является решением задачи, может быть случай, когда текущая итерация алгоритма далеко и концепция приближения Гессиана на этих итерациях должна быть пересмотрена. Также отметим, что если решение задачи минимизации должно лежать в пределах области, определенной ограничениями на переменные, или на область наложены более общие ограничения, то Гессиан может быть неопределенным, даже в точке, являющейся решением. Наконец, положительная определенность несовместима на практике с симметрией, уравнением секущей (3) и разреженности B_k (см. [14]) . Все эти наблюдения оправданы, по мнению авторов, продолжением исследований неопределенных квазиньютоновских методов, особенно в контексте крупномасштабных задач, для которых сохранение структуры и разряженность в B_k имеют первостепенное значение.

С другой стороны, развитие «доверительной области» методов, в последнее время, позволило создать эффективные алгоритмы, которые способны обрабатывать неопределенные приближения Гессиана. В основном используя в сочетании с точной информацией о Гессиане (то есть в пределах метода Ньютона), эти алгоритмы стали важным предметом исследования в более общем контексте (см. [10] для лучшего понимания этих методов). Их успех вероятно связан с их замечательной численной надежностью и их комплексным теоретическим обоснованием. Среди основных отличий между доверительной области и методом линейного поиска тот факт, что шаг s_k больше не кратен направлению d_k, и что линейный поиск не выполняется.

В двух последних работах [1, 2], авторы описывают и анализируют класс методов доверительной области для неограниченной и связной условной минимизаций. Вторая из этих работ, в частности, дает численное сравнение нескольких квазиньютоновских формул обновления, используемых в доверительной области и применяемых на задачах небольших размерностей. Формула (6) рассматривается вместе с формулой Давидона-Флетчера-Пауэлла (DFP)

$$B_{(k+1)}=B_k+\frac{r_k y_k^T+y_k r_k^T}{y_k^T s_k}-\frac{(r_k^T s_k)\, y_k y_k^T}{(y_k^T s_k)^2}, \tag{9}$$

Симметричная формула ранга 1 (SR1)

$$B_{(k+1)}=B_k+\frac{r_k r_k^T}{r_k^T s_k} \tag{10}$$

Формула Пауэлла-симметрично-Бройдена (PSB)

$$B_{(k+1)}=B_k+\frac{r_k s_k^T+s_k r_k^T}{s_k^T s_k}-\frac{(r_k^T s_k)\, s_k s_k^T}{(s_k^T s_k)^2}, \tag{11}$$

где определим остаток r_k как

$$r_k:=y_k-B_k s_k. \tag{12}$$

Несколько удивительно как традиционное превосходство обновления Бройдена-Флетчера-Голдфарба-Шэнно ставится под сомнение этими численными экспериментами, и формула SR1 кажется, значительно более эффективной в рамках доверительной области, чем другие квазиньютоновские методы, протестированные в данном контексте. Было предложено в [2], что это интересное отношение может быть связано с тем, что лучшая сходимость матриц B_k к истинному Гессиану цели решения наблюдается при использовании обновления SR1, по сравнению с другими формулами обновлений. Цель данной статьи прояснить это предложение, рассматривая сходимость матриц B_k к истинному Гессиану в предельной точке последовательности итераций, произведенных с помощью алгоритма минимизации, при условии, что направления поиска s_k остаются равномерно линейно независимы, это будет определено в следующем разделе. Скорость сходимости матриц также проверена, и оказалось, что сильно связана со скоростью сходимости последовательности { x_k }. Следующий раздел содержит краткий обзор того, что известно о сходимости квазиньютоновских матриц, а также обсуждаются сделанные предположения. Раздел 3 посвящен доказательству сходимости обновления SR1, а в разделе 4 представлены некоторые численные результаты, подтверждающие эту теорию, вместе с остальными формулами, рассмотренных выше.

2. <u>Сходимость приближений Гессиана в квазиньютоновских методах.</u>

Мы рассматриваем решение задачи безусловной оптимизации

$$\min_{x\in R^n} f(x), \tag{13}$$

поиска локального минимума, используя следующую алгоритмическую модель.

Алгоритм 1. Пусть x_0 - начальная точка, а также начальное состояние симметрической матрицы n на n B_0 .

Шаг 1. Обозначим k = 0 и вычислим $g_0:=\nabla f(x_0)$.

Шаг 2. найдем вектор s_k .

Шаг 3. Обозначим $x_{(k+1)}=x_k+s_k$, (14)

вычислим $g_{(k+1)}=\nabla f(x_{(k+1)})$ и

$$y_k=g_{(k+1)}-g_k \tag{15}$$

Шаг 4. Обновим матрицу B_k и получим $B_{(k+1)}$, использую одну из

квазиньютоновских формул, описанных выше.
Шаг 5. Увеличим k на единицу и перейдем к шагу 2.

Это все в общих чертах и , естественно, требует некоторых пояснений.

1. Мы не указали, каким образом должен быть вычислен s_k. Заметим, однако, что мы задали вычисление в классическом линейном поиске

$$s_k = -a_k B_k^{-1} g_k \tag{16}$$

из (2), где мы предположили, что B_k невырожденная.

Эта свобода выбора s_k ,тем не менее, весьма полезна. В частности, мы имеем в виду определение шага по схеме доверительной области (например, используя технику сопряженных градиентов), который не даст (16). Другой случай, когда (16) не будет проводить это в важном контексте больших скалярных оптимизаций с использованием частично разделяемых функций и распределенных обновлений [9], где шаг вычисляется путем принятия всех элементов функции во внимание и, следовательно имеет лишь косвенное отношение к какому-либо конкретному элементу приближения Гессиана.

2. Не были даны критерии остановки. Мы, действительно, заинтересованы в асимптотике алгоритма, и будем считать, что он производит бесконечную последовательность итераций { x_k } и бесконечную последовательность квазиньютоновских матриц { B_k }.
3. В случае, когда формула SR1 (10) используется для обновления B_k , мы должны убедиться, что они хорошо определены. Таким образом, формула (10) может быть использована только если

$$|r_k^T s_k| \geq c_1 \|r_k\| \|s_k\|, \tag{17}$$

где $c_1 \in (0,1)$ - константа. (Здесь и ниже, норма есть обычная $l_2 - норма$ на векторах соответствующая индуцированная норма матриц). Если условие нарушается, то зададим $B_{(k+1)} = B_k$.
4. Предполагается, что для всех итераций используется одна и та же квазиньютоновская формула: например, мы не используем формулу Бройдена-Флетчера-Голдфарба-Шэнно для одних итераций и формулу Давидона-Флетчера-Пауэлла для других.

Теперь мы представим наши предположения о проблеме и последовательности итераций, полученных нашей алгоритмической моделью.

(Пр.1) $f(x)$ - дважды всюду непрерывно дифференцируемая функция.
(Пр.2) $\nabla^2 f(x) := H(x)$ - постоянная Липшица, то есть существует константа $c_2 > 0,$ такая что для всех $x, y \in R^n$,

$$\|H(x) - H(y)\| \leq c_2 \|x - y\| \tag{18}$$

(Пр.3) Последовательность { x_k } сходится к некоторой конечной точке x'
(Пр.4) Последовательность { s_k } равномерно линейно независимы, то есть существуют $c_3 > 0, k_0 и\ m \geq n,$ такие что для любых $k \geq k_0$, можно выбрать n различных

индексов

$$k \le k_1 < \ldots < k_n \le k+m$$

с

$$\sigma_{min}(S_k) \ge c_3\,, \tag{19}$$

где $\sigma_{min}(S_k)$ - минимальное сингулярное значение матрицы

$$S_k := \left(\frac{s_{(k_1)}}{\|s_{(k_1)}\|}, \ldots, \frac{s_{(k_n)}}{\|s_{(k_n)}\|}\right). \tag{20}$$

(Мы отсылаем читателя к [11] для определения эквивалентной равномерно независимой последовательности.)

Мотивация для последнего из этих предположений (Пр.4) происходит от тщательного анализа численных результатов, представленных в [2]: в этих тестах, наименьшее сингулярное значение матриц S_k, содержащих n последовательных нормированных направлений всегда больше 10^{-4}. Превосходство обновления SR1 над BFGS, наблюдаемое в настоящей статье, может быть проанализировано в контексте этого предположения.

Однако, ясно, что (Пр.4) является достаточно сильным: есть, на самом деле, известные обстоятельства, когда (Пр.4) не выполняется. Во-первых, знаменатель в (10) может быть сколь угодно малым, и даже полностью исчезнуть при решении хорошо определенных и хорошо обусловленных задач минимизации. Это происходит, в частности, когда (2) используется с $a_k = 1$, в этом случает (10) принимает вид

$$B_{k+1} = B_k + \frac{g_{k+1} g_{k+1}^T}{g_{k+1}^T s_k}. \tag{21}$$

Обновление (10) становится непригодным, если $g_{k+1}^T s_k = 0$, хотя в большинстве квазиньютоновских методах предпочтительно, если длина шага минимизации целевой функции лежит вдоль направления поиска. Предполагая, что в этой формуле знаменатель не точно равен нулю, а лишь очень мал, сделается большая коррекция в B_k для получения B_{k+1}, возможно удалив полезную информацию, уже содержащуюся в B_k. Кроме того, теперь этот факт

$$B_{k+1} \approx \frac{g_{k+1} g_{k+1}^T}{g_{k+1}^T s_k}. \tag{22}$$

может привести к тому, что последующие направления s_{k+j} будут ортогональны к g_{k+1}, что может привести к тому, что эти направления станут линейно зависимы, что противоречит (Пр.4). Другие случаи, когда это предположение не подходит, возникают при решении разделяющихся и частично разделяющихся задач, когда минимум $f(x)$ ищется на различных итераций в независимых подпространствах. Компоненты направлений поиска вдоль подпространств, где найдется минимум первым, может затем стать очень малым и линейно зависимым.

Конечно, анализ сходимости матриц, полученных с помощью обновления SR1 без (Пр.4) является желательным, но авторы верят, что такой анализ, вероятно, будет очень трудным.

Может быть, самый ранний результат о сходимости квазиньютоновских матриц B_k, без предположения о точном линейном поиске, является то, что с учетом квадратичной целевой функции и последовательности n линейно независимых шагов $\{s_k\}_{k=1}^n$, формула

SR1 определяет истинный Гессиан $H=\nabla^2 f$ (то есть $B_{n+1}=H$), при условии формулы (10) хорошо определенной для этих шагов. Этот результат, похоже, впервые был доказан Фиакко и Маккормикои [5], и довольно хорошо известен (см [6], например). Этот результат имеет очень интересную особенность, что он никак не зависит от метода, используемого для расчета шагов s_k , таким образом, имеет место в нашем основном алгоритмическом режиме.

Сходимость матриц, полученных с помощью обновления PSB, была также изучена. В [12], Пауэлл доказал, что, предполагая (Пр.l) и (Пр.2), сходимость последовательности $\{ x_k \}$ к точке x' и равномерную линейную независимость шагов $\{ s_k \}$(См. [11]), последовательность $\{ B_k \}$, полученная с помощью обновления PSB, сходится к *H(x')*. Доказательство основывается на том факте, что обновление PSB определяет B_{k+1} как симметричную матрицу, близкую к B_k по норме Фробениуса, что удовлетворяет условию (3). (Интересно заметить, что [12] также была основополагающей работой в изучении методов доверительных областей). Аналогичная теорема доказана для разреженного обновления PSB в [15].

Обновления DFP и BFGS были рассмотрены Пауэллом и Ге в более поздней статье [7]. Они рассматривали случай, когда шаг s_k получался из (16) с выбором $a_k=1,$ и доказали, что последовательность $\{ B_k \}$ сходится, когда целевая функция строго выпуклая квадратичная. Тем не менее, предел последовательности $\{ B_k \}$ не обязан быть, в основном, эквивалентным матрице Гессе квадратичной функции. Они, действительно, представили пример, где наблюдается это поведение (другие примеры можно найти в [4]). Ге и Пауэлл также показывают, что их результат может быть продлен до неквадратичных целевых функций при условии, что выполняются (Пр.l) и (Пр.2), и что последовательность $\{ x_k \}$ сходится к пределу x' , где *H(x')* – положительно определена. Доказательства в их работах достаточно громоздкие и зависят в значительной степени от их выбора шагов s_k . Довольно удивительно, что они не зависят от предположения о типе в (Пр.4).

Сходимость квазиньютоновских матриц, полученных с помощью формулы BFGS, была также рассмотрена Шуллером в [13]. Он показывает интересную связь

$$\|B_{k+n+1}-H(x')\|\leq c_4\|x_k-x'\| \tag{23}$$

для некоторой константы $c_4>0,$ но его предположения достаточно сильны. Действительно, кроме (Пр.3) и (Пр.4), он также предполагает, что задача линейного поиска (4) решается асимптотически точно, что B_0 и *H(x')* положительно определены, что нормы B_k *и* его обратной ограничены, что x_0 достаточно близко к x' и что последовательность $\{ x_k \}$ сходится к x' Q-линейно. Первое из этих предположений, в частности, весьма нереально в нашем контексте, потому что мы хотим покрыть случай доверительной области, где вообще нет линейного поиска.

3. Симметричное обновление первого ранга для неквадратичных функций

Теперь обратим наше внимание снова на формулу SR1, но на общую неквадратичную целевую функцию, которая удовлетворяет (Пр.1) и (Пр.2). Сначала докажем следующую важную лемму, удивительно простой вариант квадратичного случая.

Лемма 1. Предположим, что (Пр.1) и (Пр.2) выполняются , а также что $\{ x_k \}$ - последовательность итераций, генерируемых с помощью алгоритма, описанного выше, используя формулу SR1(10). Кроме того, предположим, что (17) выполняется при каждой итерации. Тогда

$$\|y_j-B_{(j+1)}s_j\|=0 \tag{24}$$

для всех j и

$$\|y_j - B_i s_j\| \le \frac{c_2}{c_1}\left(\frac{2}{c_1}+1\right)^{(i-j-2)} \eta_{ij}\|s_j\| \qquad (25)$$

для любых $j\ u\ i \ge j+1, где$

$$\eta_{ij} := max[\|x_p - x_s\| : j \le s \le p \le i]. \qquad (26)$$

Доказательство. Заметим сначала, что (24) и (25) с i = j + 1 немедленно следует из (3). Докажем (25) по индукции. Выберем $k \ge j+1$ и предположим что выполняется предположение индукции для всех $i = j+1, \ldots, k.$ Теперь заметим, что

$$|r_k^T s_j| = |y_k^T s_j - s_k^T B_k s_j| \le |y_k^T s_j - s_k^T y_j| + \frac{c_2}{c_1}\left(\frac{2}{c_1}+1\right)^{(k-j-2)} \eta_{ij}\|s_j\|\|s_k\|, \qquad (27)$$

где мы используем (12), наше индуктивное предположение и неравенство Коши-Шварца. Теперь, используя теорему о среднем, получаем, что для всех l,

$$y_l = H_l s_l, \qquad (28)$$

где

$$H_l = \int_0^1 H(x_l + ts_l)\, dt. \qquad (29)$$

Отсюда (27) вытекает, что

$$|r_k^T s_j| \le |s_k^T (H_k - H_j) s_j| + \frac{c_2}{c_1}\left(\frac{2}{c_1}+1\right)^{(k-j-2)} \eta_{ij}\|s_j\|\|s_k\| \le .$$
$$c_2 \eta_{(k+1\,j)}\|s_j\|\|s_k\| + \frac{c_2}{c_1}\left(\frac{2}{c_1}+1\right)^{(k-j-2)} \eta_{ij}\|s_j\|\|s_k\|, \qquad (30)$$

где мы используем (29), (Пр.2) и определение (26). Но из (10), (17) и неравенства треугольника вытекает

$$\|y_j - B_{(j+1)} s_j\| = \|y_j - B_j s_j - \frac{r_k r_k^T s_j}{r_k^T s_k}\| \le \|y_j - B_j s_j\| + \frac{|r_k^T s_j|}{c_1 \|s_k\|} \qquad (31)$$

и следовательно

$$\|y_j - B_{(j+1)} s_j\| \le \left(\frac{c_2}{c_1}\left(\frac{2}{c_1}+1\right)^{(k-j-2)} + \frac{c_2}{c_1^2}\left(\frac{2}{c_1}+1\right)^{(k-j-2)}\right)\eta_{kj}\|s_j\| + \frac{c_2}{c_1}\eta_{(k+1\,j)}\|s_j\| \qquad (32)$$

используя (30), предположение индукции и (31). Последнее неравенство затем дает (25) для $i = k+1$, если принять во внимание тот факт, что $c_1 \in (0,1)$ и используя простое неравенство

$$\eta_{kj} \leq \eta_{(k+1\, j)} \tag{33}$$

Лемма доказана.

Теперь мы можем доказать теорему о сходимости.

Теорема 2. Пусть выполняются (Пр.1) — (Пр.4), где { x_k } - последовательность итераций, генерируемых с помощью алгоритма, описанного выше, используя формулу SR1(10). Кроме того, предположим, что (17) выполняется при каждой итерации. Тогда существует константа c_5 , такая что для $k \geq k_{0,}$

$$\| B_{k+m+1} - H(x') \| < -c_5 \varepsilon_k , \tag{34}$$

где

$$\varepsilon_k := max[\| x_s - x' \| : k \leq s \leq k+m+1], \tag{35}$$

и

$$\lim_{k \to \infty} \| B_k - H(x') \| = 0. \tag{36}$$

Доказательство. Для доказательства теоремы заметим, что для любых s и p,

$$\| x_s - x_p \| \leq \| x_s - x' \| + \| x_p - x' \|, \tag{37}$$

что означает

$$\eta_{k+m+1,k} \leq 2\varepsilon_k , \tag{38}$$

в силу определений (26) и (35). Также заметим, что в силу (Пр.2),

$$\| y_j - A(x') s_j \| = \| (H_j - H(x')) s_j \| \leq c_2 \varepsilon_k \| s_j \| \tag{39}$$

для любых $k \leq j \leq k+m, где H_j$ определена как (29). Кроме того, для любого такого j мы можем вывести из Леммы 1 и (38), что

$$\| y_j - B_{k+m+1} s_j \| \leq \frac{2c_2}{c_1} \left(\frac{2}{c_1} + 1 \right)^m \varepsilon_k \| s_j \|. \tag{40}$$

Собирая (39) и (40) и используя неравенство треугольника, мы получим, что для любых $k \leq j \leq k+m$

$$\| (B_{k+m+1} - H(X')) \frac{s_j}{\| s_j \|} \| \leq \left(\frac{2c_2}{c_1} \left(\frac{2}{c_1} + 1 \right)^m + c_2 \right) \varepsilon_{k.} \tag{41}$$

Это неравенство выполняется также для $j = k_{1,} \ldots, k_n .$ Мы также имеем

$$\| B_{k+m+1} - H(x') \| \leq \frac{1}{c_3} \| (B_{k+m+1} - H(x')) S_k \|, \tag{42}$$

используя (Пр.4). Это последнее неравенство, вместе с (41), подразумевает (34) с

$$c_5 := \frac{c_2}{c_3}\left(\frac{2}{c_1}\left(\frac{2}{c_1}+1\right)^m+1\right)\sqrt{n}. \qquad (43)$$

Теперь (Пр.3) подразумевает в свою очередь, что ε_k стремится к нулю, и , следовательно, (36) следует из (34).

Теорема доказана.

Заметим, что эта теорема достаточно мощная. Она не только гарантирует глобальную сходимость последовательности $\{ B_k \}$ к истинному Гессиану, но и также предоставляет некоторые представления о скорости сходимости в силу (34). В самом деле, можно получить простое следствие.

Следствие 3. Пусть выполняются (Пр.1) — (Пр.4), где $\{ x_k \}$ - последовательность итераций, генерируемых с помощью алгоритма, описанного выше, используя формулу SR1(10). Кроме того, предположим, что (17) выполняется при каждой итерации и существует константа $c_6 \geq 0$, такая что для достаточно больших k выполняется

$$\|x_{k+1}-x'\| \leq c_6 \|x_k-x'\|. \qquad (44)$$

Тогда существует константа c_7 , такая что

$$\|B_{k+m+1}-H(x')\| \leq c_7 \|x_k-x'\| \qquad (45)$$

для достаточно больших k.

Доказательство. Доказательство очевидно, если заметить, что из (44) и определения (35) следует, что

$$\varepsilon_k \leq max[1, c_6^m] \|x_k-x'\| \qquad (46)$$

для достаточно больших k. Следовательно (45) следует из (34) с

$$c_7 := max[1, c_6^m] c_5. \qquad (47)$$

Следствие доказано.

Таким образом, мы получаем довольно интересный результат. Чем быстрее сходится $\{ x_k \}$ к x' , тем быстрее сходятся $\{ B_k \}$ к *H(x')!*

Этот результат аналогичен результату Шуллера, но если наше доказательство проводится только для обновления SR1, а сделанные предположения, по существу, слабее. В частности, не потребовался ни линейный поиск, ни положительная определенность матриц.

Мы также можем сосредоточить наше внимание на последних n итерациях и заменить (Пр.4) следующим предположением.

(Пр.4б). Существует k_0 такое,что для любых $k \geq k_0$ матрица

$$S'_k := \left(\frac{s_k}{\|s_k\|}, \ldots, \frac{s_{k+n-1}}{\|s_{k+n-1}\|}\right) \qquad (48)$$

невырожденная.

Теперь мы получаем следующую оценку ошибки.

Следствие 4. Пусть выполняются (Пр.1) — (Пр.3) и (Пр.4б), где $\{ x_k \}$ - последовательность итераций, генерируемых с помощью алгоритма, описанного выше, используя формулу SR1(10). Кроме того, предположим, что (17) выполняется при каждой итерации. Тогда существует константа $c_8 \geq 0$, такая что для любых $k \geq k_0$ выполняется

$$\|B_{k+m+1} - H(x')\| \leq c_8 \kappa(S'_k) \varepsilon_k, \tag{49}$$

где S'_k и ε_k определены в (48) и (35) соответственно, и где *κ(X)* — номер состояния матрицы *X*.

Доказательство. Этот результат может быть непосредственно выведен из неравенства (41), определения (48), ограничения

$$\kappa(S'_k) \geq \frac{\|S'_k\|}{\sigma_{min}(S'_k)} \geq \frac{1}{\sigma_{min}(S'_k)}, \tag{50}$$

и определения

$$c_8 := c_2 \sqrt{n} \left(\frac{2}{c_1} \left(\frac{2}{c_1} + 1 \right)^n + 1 \right). \tag{51}$$

Следствие доказано.

Этот результат вероятно лучше отражает поведение, которое можно увидеть на практике, как показано в следующем разделе.

Разумеется, границы приведенные в Теореме 2 и Следствиях 3 и 4 могут быть достаточно грубыми, в основном из-за того, что они сильно зависят от c_1 в (17), а также от константы Липшица c_2.

Также ясно, что если условие (17) не выполняется для некоторых итераций, то весь процесс сходимости и соответствующие оценки только замедлятся. Если (17) не будет выполнятся для бесконечного числа итераций, то будет нарушена сходимость квазиньютовских матриц. Тем не менее, мы можем ожидать, что число итераций, где условие (17) не будет выполняться, будет мало на практике.

Наконец, интересно отметить, что сходимость матриц $\{ B_k \}$ к *H(x')* не требует, чтобы *x'* была стационарной точкой. Это полезно в контексте доверительных областей методов минимизации, потому что это гарантирует, что любое отрицательное собственное значение присутствующее в *H(x),* в конечном итоге будет присутствовать в B_k , и шаги отрицательной кривизны будут предотвращать сходимость итераций к *x'.*

4. <u>Некоторые численные эксперименты</u>

В этом разделе мы приводим некоторые численные эксперименты, которые проводились с двойной целью: проверка теоретических оценок погрешности квазиньютоновских матриц, полученных с помощью обновления SR1, а также сравнение этих свойств сходимости с другими квазиньютоновскими формулами, таких как BFGS, DFP и PSB. Как уже упоминалось выше, все эти обновления были уже рассмотрены в [2].

Все эксперименты проводились на языке Фортран 77 на Cray X-MP / 24 в лаборатории Харуэлле (Harwell Laboratory), в основном, потому что эта машина имеет большую длину слова (64 бита) и ,следовательно, позволяет проводить более подробный анализ асимптотического поведения квазиньютоновских матриц вблизи к минимуму целевой функции. Все тесты проводились в режиме повышенной точности под компилятором CFT77. Точность соответствующей машины вычисления ε_M порядка 10^{-29}.

В качестве основного использования квазиньютоновских формул для нелинейной оптимизации алгоритмов, наши эксперименты были смоделированы для анализа сходимости квазиньтоновских матриц для последовательности { x_k }, сходящейся к минимуму некоторой целевой функции. Мы сосредоточим наше внимание на безусловной минимизации квартичных тестовых функций вида

$$f(x)=\frac{1}{2}x^T Hx+\frac{1}{3}\sum_{i=1}^{n} t_i x_i^3+\frac{1}{4}\sum_{i=1}^{n} q_i x_i^4. \quad (52)$$

Теперь подробно рассмотрим выбор, сделанный для *H* и коэффициентов t_i и q_i . Они генерируются с помощью обычной FA01 подпрограммы библиотеки Харуэлл для генерации равномерно распределенных псевдослучайных чисел в заданном диапазоне от "точки инициализации" θ . Чтобы воспроизвести наши результаты, рекурента, использууемая этой программой, приведена в приложении.

- Даны целое число $\nu \geq 1$, коэффициенты u_i , t_i и q_i , полученные для $i=1,\ldots,n$ в порядке u_1 , t_1 , q_1 , u_2 , t_2 , q_2 , …, используя FA01 с точкой инициализации $\theta=\nu+\nu*16^4$ и диапазонами [0,1], [0,1] и $[0,10*2^{\nu}]$ соответственно.
- Матрица *n на n H* — симметричная, положительно определенная и получена с помощью следующей процедуры.

Используя вектор $u=(u_{1,}\ldots,u_n)^T,$ определим матрицу отражения

$$R=I-2\mathrm{uu}^T/\|u\|^2 \quad (53)$$

тогда матрица H будет построена как

$$H=RDR^T, \quad (54)$$

где *D* — диагональная матрица, элементы которой распределены интервалами между 1 и $2^{-\nu}$.

Как диагональные элементы *D* являются собственными для *H* , так параметр ν позволяет обрабатывать вариационные задачи. Минимум для этих задач (для всех значений ν) явно лежит в начале координат.

Были получены три различных типа последовательностей, сходящихся к этому минимуму:

- искусственные последовательности, имеющие хорошо определенную скорость сходимости,
- последовательности итераций производимые алгоритмами из доверительной области, описанные в [2],
- последовательности итераций генерируемые алгоритмом минимизации, использующий линейный поиск, как описано в Разделе 1, где метод определения размера шага a_k описан ниже.

Все эти последовательности начинаются как

$$x_0^T = (1, \dots, 1)$$

и начальное квазиньютоновское приближение всегда выбирается как еденичная матрица.

Формула SR1 пропускалась, как это было предложено в теории выше, когда тест (17) не проходил, где c_1 была выбрана как 10^{12}. Обновления BFGS и DFP пропускались, когда

$$y_k^T < 10^{-8} \| y_k \| \| s_k \| \quad . \tag{55}$$

Стоит отметить что эти условия не помешали обновлению второй производной приближений в любых тестах, приводящихся здесь.

4.1 <u>Искусственно сгенерируемые последовательности</u>

Последовательность { x_k } сходящаяся к началу, в этих тестах, была сгенерирована с помощью следующей процедуры

Шаг 1. Определим норму следующей итерации x_{k+1}. Это делается с помощью следующего отношения

$$\| x_{k+1} \| = \begin{cases} \beta_k \| x_k \| & (\textit{линейная сходимость}), \\ \beta_k \| x_k \| / k & (\textit{суперлинейная сходимость}), \\ \beta_k \| x_k \|^2 & (\textit{квадратичная сходимость}), \end{cases} \tag{56}$$

где β_k - случайное число, равномерно распределенное в одном из диапазонах [0.7,0.9], [0.4, 0.5] и [0.1, 0.3].

Шаг 2. Сгенерируем x_{k+1} путем вычисления случайного вектора z_{k+1} и зададим

$$x'_{k+1} = \| x_{k+1} \| z_{k+1}. \tag{57}$$

Вектор z_{k+1} вычисляется путем случайного определения его углов Эйлера в соответствующих диапазонах.

Шаг 3. Убедимся, что новая точка дается строгий спуск по целевой функции. Если $f(x'_{k+1}) < f(x_k)$, тогда примем $x_{k+1} = x'_{k+1}$ на следующем шаге. Иначе, вернемся к Шагу 2.

Заметим, что Шаг 2 можно повторить несколько раз, прежде чем найдется точка, удовлетворяющая условию. Как показано в (56), были рассмотрены три различных скорости сходимости: линейные, суперлинейные и квадратичные.

Отметим, что использование случайных направлений может быть весьма полезным для сходимости матриц, генерируемых обновлением SR1: действительно, поведение рассматриваемое в обсуждении (Пр.4) выше, вряд ли произойдет в данном контексте, потому что шаги s_{k+j} больше не связаны с матрицей B_{k+1} и больше не должны быть почти ортогональными к g_{k+1}. В самом деле, предполагая, что s_{k+1} имеет значительную составляющую вдоль g_{k+1}, остаток r_{k+1} сам по себе будет доминировать в выражении

$g_{k+1}(g_{k+1}^T s_{k+1})/(g_{k+1}^T s_k)$ из (22) и заметим, что

$$B_{k+2}-B_{k-1}=\frac{r_{k+1}r_{k+1}^T}{r_{k+1}^T s_{k+1}}\approx\frac{-g_{k+1}g_{k+1}^T}{g_{k+1}^T s_k}, \tag{58}$$

коррекция большого терма ранга один, введенная обновлением в B_{k+1}. Случайные направления, тем не менее, появляются ,например, в связи обновлением SR1 в решении большого числа хорошо обусловленных, частично отделимых задач, или когда обновление используется в контексте неточной схеме доверительной области.

Сначала приведем в Таблице 1 четыре серии экспериментов, связанных с линейно сходящимися последовательностями в 10 и 3 переменные, соответственно. Для того чтобы получить результаты этих таблиц, линейно сходящиеся последовательности были получены, как описано выше, и процесс был остановлен как только $\|x_{k+n+1}\|\leq 10^{-8}$. Тогда ошибка вычисляется как

$$E:=\max_{i,j=1,\dots,n}|[B_{k+m+1}-H(x')]_{ij}|. \tag{59}$$

Эти ошибки появляются в таблицах под заголовками $E_{SR1}, E_{PSB}, E_{BFGS}$ и E_{DFP} для четырех рассматриваемых обновлений, соответственно. Значения ε_k и $\kappa(S'_k)$ также приведены для сравнения E_{SR1} с их результатами, как было предложено в Следствии 4.

Некоторые выводы можно сделать уже из этих результатов. Во-первых, предположения Следствия 4, кажется, проверены достаточно красиво со значение константы c_8 немного большей 0.1 для наших тестовых функций. Наблюдается действительно довольно гладкая сходимость квазиньютоновских матриц, полученных обновлением SR1, к истинным матрицам Гессе в точке решения, даже в не очень хорошо описанных примерах (значение ν большое). Что касается других формул обновления, то они, кажется, производят гораздо менее точное приближение. Качество приближений, полученных от BFGS и DFP, существенно ухудшается, когда размерность увеличивается от 3 до 10, особенно для DFP. Теоретическая сходимость матриц PSB едва показывает лучшие результаты на условных задачах с n = 3, но снова его производительность заметно снижается с качеством обработки, и когда размерность задачи увеличивается. Ухудшение связано с обработкой, чего и следовало ожидать, в PSB, в отличие от других трех обновлений, нет инварианта по отношению к линейному масштабированию переменных.

Таблица 1
Линейно сходящиеся последовательности.

n	β_k range	ν	E_{SR1}	ε_k	$\kappa(S'_k)$	E_{PSB}	E_{BFGS}	E_{DFP}
3	[0.1, 0.3]	2 4 6 8 10	$1.35*10^{-7}$ $6.53*10^{-8}$ $1.04*10^{-6}$ $9.52*10^{-7}$ $1.70*10^{-8}$	$5.51*10^{-6}$ $1.84*10^{-6}$ $7.28*10^{-6}$ $1.14*10^{-5}$ $4.80*10^{-6}$	3.06 1.47 3.65 3.05 8.89	1.33 3.70 2.68 2.46 $1.36*10^{1}$	$2.09*10^{-2}$ $1.38*10^{-1}$ $4.34*10^{-1}$ $1.50*10^{-1}$ $2.02*10^{-1}$	$9.75*10^{-1}$ $1.11*10^{2}$ $4.32*10^{1}$ $1.24*10^{3}$ $1.40*10^{1}$
3	[0.4, 0.5]	2 4 6 8 10	$1.72*10^{-7}$ $8.82*10^{-9}$ $4.85*10^{-8}$ $1.42*10^{-7}$ $4.87*10^{-8}$	$1.70*10^{-7}$ $2.06*10^{-7}$ $2.08*10^{-7}$ $1.55*10^{-7}$ $1.80*10^{-7}$	$9.49*10^{1}$ 2.16 8.71 6.96 4.93	$6.09*10^{-2}$ $2.89*10^{-4}$ $1.26*10^{-3}$ $2.87*10^{-3}$ $4.28*10^{-2}$	$7.86*10^{-4}$ $3.78*10^{-2}$ $3.65*10^{-3}$ $1.60*10^{-3}$ $3.84*10^{-2}$	$1.31*10^{-2}$ $3.68*10^{1}$ 4.66 $6.62*10^{1}$ $1.55*10^{2}$
3	[0.7, 0.9]	2	$9.06*10^{-7}$	$1.94*10^{-8}$	4.57	$7.13*10^{-10}$	$8.28*10^{-10}$	$8.28*10^{-10}$

		4 6 8 10	$7.59*10^{-9}$ $1.23*10^{-8}$ $1.56*10^{-8}$ $1.38*10^{-8}$	$2.37*10^{-8}$ $2.46*10^{-8}$ $2.52*10^{-8}$ $2.34*10^{-8}$	5.59 6.03 $8.88*10^{1}$ 7.61	$4.90*10^{-9}$ $4.12*10^{-7}$ $1.89*10^{-4}$ $1.24*10^{-1}$	$4.16*10^{-9}$ $7.20*10^{-9}$ $6.19*10^{-7}$ $3.06*10^{-8}$	$4.16*10^{-9}$ $7.20*10^{-9}$ $4.08*10^{-7}$ $2.61*10^{-8}$
10	[0.7, 0,9]	2 4 6 8 10	$1.06*10^{-6}$ $6.85*10^{-7}$ $6.77*10^{-6}$ $6.57*10^{-5}$ $1.63*10^{-5}$	$2.00*10^{-7}$ $1.05*10^{-7}$ $1.02*10^{-7}$ $1.06*10^{-7}$ $1.23*10^{-7}$	$1.72*10^{2}$ $4.05*10^{2}$ $2.38*10^{4}$ $4.60*10^{3}$ $4.75*10^{3}$	1.21 6.67 $1.23*10^{1}$ $5.90*10^{2}$ $8.36*10^{2}$	$3.82*10^{-1}$ $4.49*10^{-1}$ $8.22*10^{-1}$ $8.82*10^{-1}$ $6.09*10^{-1}$	2.70 5.77 $1.78*10^{1}$ $9.24*10^{2}$ $4.55*10^{2}$

Результаты суперлинейных и квадратичных последовательностей приведены в Таблице 2. Первая из этих последовательностей была остановлена на первом таком k ,что $\|x_k\| \le 10^{-30}$, в то время как вторая была остановлена, когда $\|x_k\| \le 10^{-50}$. Эти значения позволяют достичь отношения асимптотики пока не превысился уровень точности машины, несмотря на относительно высокую скорость сходимости. Оба этих теста были проведены только для *n = 3,* потому что задержка n итераций, участвующих в границах задаваемых Следствием 4, должна быть достаточно мала по отношению к скорости сходимости для достижения необходимого асимтотического поведения.

Опять же, можно наблюдать превосходную сходимость матриц, порожденных обновлением SR1. Лучшим из оставшихся, в среднем, является BFGS, PSB работает относительно хорошо в суперскалярном случае. DFP явно худший.

Также интересно численно показать, насколько сходимость матриц SR1 зависит от скорости сходимости базовой последовательности итераций минимизации. Чтобы проиллюстрировать это отношение, мы проводили тесты с *n = 3 и* $\nu=5$ для четырех обновлений, а затем подсчитывали количество итераций, которые необходимы для того, чтобы ошибка (59) была строго меньше, чем 10^{-6} . Эти подсчеты представлены в Таблице 3.

Таблица 2.
Суперскалярная и квадратичная сходимость последовательностей ($n=3, \beta_k \in [0.7, 0.9]$)

	ν	E_{SR1}	ε_k	$\kappa(S'_k)$	E_{PSB}	E_{BFGS}	E_{DFP}
Суперскалярная сходимость	2 4 6 8 10	$2.52*10^{-29}$ $2.75*10^{-27}$ $3.84*10^{-27}$ $4.66*10^{-27}$ $5.25*10^{-25}$	$3.12*10^{-26}$ $5.21*10^{-25}$ $5.33*10^{-25}$ $5.36*10^{-25}$ $5.68*10^{-25}$	2.31 3.73 1.37 2.54 $5.34*10^{1}$	$4.41*10^{-3}$ $1.73*10^{-2}$ $5.85*10^{-8}$ $5.36*10^{-5}$ $4.00*10^{-2}$	$1.03*10^{-3}$ $3.40*10^{-2}$ $2.50*10^{-2}$ $1.20*10^{-3}$ $4.83*10^{-2}$	$4.10*10^{-2}$ $4.12*10^{1}$ $2.34*10^{1}$ $3.38*10^{1}$ $3.28*10^{2}$
Квадратичная сходимость	2 4 6 8 10	$1.45*10^{-11}$ $1.18*10^{-4}$ $3.41*10^{-8}$ $1.71*10^{-9}$ $2.50*10^{-8}$	$4.00*10^{-6}$ $8.17*10^{-4}$ $2.95*10^{-4}$ $1.01*10^{-4}$ $7.27*10^{-5}$	7.90 $1.02*10^{2}$ 3.74 2.61 $2.20*10^{1}$	4.70 $4.02*10^{1}$ $9.96*10^{-1}$ $1.45*10^{2}$ $1.72*10^{2}$	$3,47*10^{-1}$ 1.56 $8.33*10^{-1}$ $1.49*10^{-1}$ $5.92*10^{-1}$	6.95 $1.06*10^{2}$ $8.77*10^{1}$ $1.19*10^{2}$ $4.07*10^{2}$

Таблица 3.
Количество итераций для достижения точности 10^{-6}

	β_k range	SR1	PSB	BFGS	DFP
Линейная	[0.7, 0.9] [0.4, 0.5]	70 23	79 40	72 56	73 82

Суперлинейная Квадратичная	[0.1,0.3] [0.7, 0.9] [0.7, 0.9]	14 12 10	29 29 >11(10^0)	>88(10^{-5}) >46(10^{-3}) >11(10^{-1})	>88(10^{-2}) >46(10^0) >11(10^{-1})

Если желаемая точность не была достигнута до этого нового критерия остановки, то в таблице обозначается следующий номер итерации со знаком «>» , в скобках обозначается порядок величины полученной точности квазиньютоновской матрицы.

Уменьшение количества итераций, необходимых для достижения такой точности , зависит от увеличения скорости сходимости последовательности { x_k }, что вполне очевидно для обновления SR1.

4.2 Последовательности, генерируемые с помощь алгоритмов из доверительной области

Следующий эксперимент использует последовательности { x_k }, которые были получены с помощью алгоритма их доверительной области. Используемый алгоритм полностью описан в [2], и был применен к квартичным тестовым примерам(52) с $n = 3$. Он останавливается как только

$$\|\nabla f(x_k)\| \le \varepsilon_M^{2/3}. \tag{60}$$

Окончательная ошибка в квазиньютоновских матрицах по сравнению с истинными Гессианами в точке минимума была вычислена как описано выше. Результаты представлены в Таблице 4, в то время как Таблица 5 дает более детальные результаты при использовании обновления SR1. В этих таблицах заголовок «ng» обозначает количество оценок градиента, которые были необходимы для достижения требуемой точности.

Таблица 4

Последовательность полученная с помощью метода доверительной области

ν	SR1		PSB		BFGS		DFP	
	ng	E_{SR1}	ng	E_{PSB}	ng	E_{BFGS}	ng	E_{DFP}
2	15	$9.74*10^{-10}$	63	$1.11*10^{-3}$	24	$1.39*10^{-3}$	45	$3.53*10^{-3}$
4	25	$3.67*10^{-13}$	59	$1.74*10^{-4}$	34	$2.02*10^{-3}$	83	$4.16*10^{-3}$
6	24	$4.96*10^{-9}$	116	$3.81*10^{-4}$	30	$1.37*10^{-3}$	42	$3.22*10^{-3}$
8	35	$8.55*10^{-10}$	169	$3.86*10^{-4}$	43	$2.38*10^{-3}$	50	$2.12*10^{-3}$
10	50	$8.63*10^{-13}$	111	$3.75*10^{-3}$	49	$1.89*10^{-3}$	93	$2.17*10^{-3}$

Таблица 5

Детали метода доверительной области для случая SR1

ν	ng	E_{SR1}	ε_k	$\kappa(S'_k)$
2	15	$9.74*10^{-10}$	$7.83*10^{-7}$	$5.40*10^1$
4	25	$3.67*10^{-13}$	$3.48*10^{-9}$	9.23
6	24	$4.96*10^{-9}$	$5.16*10^{-7}$	$3.63*10^1$

8	35	$8.55*10^{-10}$	$1.21*10^{-8}$	$3.57*10^{1}$
10	50	$8.63*10^{-13}$	$3.30*10^{-10}$	$7.30*10^{1}$

Опять же, поведение прогнозируемое в Следствии 4 наблюдается для SR1, в то время как остальные три формулы значительно менее эффективны.

4.3 <u>Последовательности, полученные с помощью алгоритма линейного поиска</u>

Наконец, рассмотрим последовательности { x_k }, полученные с помощью простого алгоритма линейного поиска. Эта традиционная структура традиционной, в которой используется формула BFGS, и была надежда, что эта формула покажет повышение эффективности в данном контексте. Поэтому мы провели тесты только для обновлений SR1 и BFGS.

Направление поиска было определено как в (5) и размер шага a_k был вычислен путем последовательного деления пополам (начиная с 1) для того, чтобы удовлетворять условию

$$f(x_k)-f(x_k+1)\geq 0.1\, a_k \nabla f(x_k)^T d_k . \tag{61}$$

При использовании формулы обновления BFGS, добавлялись дополнительные условия

$$y_k^T s_k \geq 10^{-8}\|y_k\|\|s_k\| \tag{62}$$

для того, чтобы гарантировать положительную определенность приближений Гессиана.

Отметим, что эта стратегия на самом деле не рекомендуется для обновления SR1, так как эти формулы могут генерировать неопределенные или даже вырожденные матрицы B_k. Для решения этой потенциальной проблемы, направление поиска было обращено, когда d_k не было направлением спуска, что происходило несколько раз. Вырожденный случай никогда не происходил в наших тестах.

Тестовыми функциями опять были квартичные (52) с *n=3.* Алгоритм минимизации останавливался как только

$$\|\nabla f(x_k)\|\leq 10\,\varepsilon_M \tag{63}$$

(который требует чуть больше чем (60)). Соответствующие величины были вычислены в конечной точке, и приведены в Таблице 6. Более детально для случая SR1 разобраны в Таблице 7.

Даже в этом несоответствующем состоянии, SR1 остается весьма эффективным и очень согласованным с оценками ошибок, приведенных в Следствии 4.

Таблица 6
Последовательность полученная с помощью метода линейного поиска

ν	SR1		BFGS	
	ng	E_{SR1}	ng	E_{BFGS}
2	21	$2.97*10^{-14}$	33	$2.16*10^{-3}$
4	24	$5.99*10^{-13}$	39	$5.57*10^{-4}$

6	35	$4.01*10^{-10}$	47	$1.36*10^{-3}$
8	34	$1.98*10^{-17}$	56	$1.04*10^{-3}$
10	43	$5.76*10^{-11}$	61	$3.37*10^{-4}$

Таблица 7
Детали для случая SR1 в методе линейного поиска

ν	ng	E_{SR1}	ε_k	$\kappa(S'_k)$
2	21	$2.97*10^{-14}$	$2.06*10^{-9}$	$1.40*10^{1}$
4	24	$5.99*10^{-13}$	$1.97*10^{-9}$	4.55
6	35	$4.01*10^{-10}$	$6.37*10^{-8}$	$2.38*10^{1}$
8	34	$1.98*10^{-17}$	$5.38*10^{-12}$	1.79
10	43	$5.76*10^{-11}$	$7.60*10^{-10}$	$4.90*10^{1}$

5. Заключение

Эта статья сравнивает поведение квазиньютоновских формул обновления как средства генерирования симметричных приближений к матрицам Гессе. Известный результат сходимости матриц SR1 на квадратичных функциях был продлен до общего класса достаточно гладких нелинейных функций. Эта сходимость является глобальной, и показано. что скорость сходимости улучшается с повышением скорости сходимости последовательности итераций, полученных с помощью основного алгоритма минимизации.

Представленные численные эксперименты также достаточно хорошо поддерживают эту теорию. Кроме того, эти расчеты показывают, что, по сравнению с другими формулами, таких как BFGS, DFP или PSB, SR1 генерирует более точные приближения Гессиана в рядя обстоятельств.

Хотя нельзя полностью объяснить, почему метод доверительной области на основе обновления SR1 превысил более классический линейный поиск на основе алгоритма BFGS из работы [2], это, конечно, проливает некоторый свет на их относительные достоинства и отличия.

Приложение: рекурента для FA01

Как упоминалось выше, подпрограмма FA01 из библиотеки HSL (Harwell Subroutine Library) использовалась для получения коэффициентов квартичной функции из тестового примера (52), и других равномерно распределенных псевдослучайных чисел в заданном диапазоне значений. Для воспроизведения наших численных экспериментов или же использования (52) для других целей, рассмотрим детально рекуренту, используемую в FA01, для генерирования этих чисел.

На вход FA01 дается стартовая точка θ и первые переопределения

$$\theta=(9228907*\theta)\,mod\,16^{8}. \qquad (П.1)$$

Это новое значения затем используется в следующем вызове процедуры. Псевдослучайные величины в диапазоне [0, 1] затем вычисляются по формуле

$$«псевдослучайное число» = \theta*16^{-8}. \qquad (П.2)$$

Это число , наконец, масштабируется с помощью линейного преобразования, которое отображает [0, 1] на желаемый диапазон значений.

Благодарности

Авторы благодарят М.Дж.Д. Пауэлл за его вдумчивый и конструктивные замечания по первой версии этой статьи.

Литература

[1] A.R. Conn, N.I.M. Gould and Ph.L. Toint, "Global convergence of a class of trust region algorithms for optimization with simple bounds," SIAM Journal on Numerical Analysis 25 (1988) 433-460 (with a correction given in SlAM Journal on Numerical Analysis 26 (1989) 764-767).

[2] A.R. Conn, N.I.M. Gould and Ph.L. Toint, "Testing a class of methods for solving minimization problems with simple bounds on the variables," Mathematics of Computation 50 (1988) 399-430.

[3] J.E. Dennis and J.J. Mor6, "Quasi-Newton methods, motivation and theory," SIAM Review 19 (1977) 46-89.

[4] J.E. Dennis and R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[5] A.V. Fiacco and G.P. McCormick, Nonlinear Programming (Wiley, New York, 1968).

[6] R. Fletcher, Practical Methods of Optimization: Unconstrained Optimization (Wiley, Chichester, 1980).

[7] R.P. Ge and M.J.D. Powell, "The convergence of variable metric matrices in unconstrained optimization," Mathematical Programming 27 (1983) 123-143.

[8] P.E. Gill, W. Murray and M.H. Wright, Practical Optimization (Academic Press, New York, 1981).

[9] A. Griewank and Ph.L. Toint, "Partitioned variable metric updates for large structured optimization problems," Numerische Mathematik 39 (1982) 119-137.

[10] J.J. Mor6, "Recent developments in algorithms and software for trust region methods," in: A. Bachem, M. Gr6tschel and B. Korte, eds., Mathematical Programming: The State of the Art (Springer, Berlin, 1983).

[11] J.M. Ortega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables (Academic Press, New York, 1970).

[12] M.J.D. Powell, "A new algorithm for unconstrained optimization," in: J.B. Rosen, O.L. Mangasarian and K. Ritter, eds., Nonlinear Programming (Academic Press, New York, 1970).

[13] G. Schuller, "On the order of convergence of certain quasi-Newton methods," Numerische Mathematik 23 (1974) 181-192.

[14] D.C. Sorensen, "An example concerning quasi-Newton estimates of a sparse Hessian," SIGNUM Newsletter 16 (1981) 8-10.

[15] Ph.L. Toint, "On the superlinear convergence of an algorithm for solving a sparse minimization problem," SIAM Journal on Numerical Analysis 16 (1979) 1036-1045.