РЕШЕНИЕ БОЛЬШИХ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ НА МНОГОПРОЦЕССОРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ С РАСПРЕДЕЛЕННОЙ И ОБЩЕЙ ПАМЯТЬЮ

Л. Д. Попов, А. В. Некрасов

В различных алгоритмических подходах и вычислительных схемах линейного программирования часто встречаются операции положительной срезки векторов и конструкции, основанные на обычных или расширенных штрафных функциях, в состав которых эти операции входят существенным образом. Простейшим примером таких конструкций может служить функция квадратичной невязки $F(x) = \frac{1}{2}||(Ax-b)^+||^2$, поставленная в соответствие (не обязательно совместной) системе линейных неравенств $Ax \leq b$, где $A_{m\times n}$ — некоторая заданная матрица, $b = (b_1, \ldots, b_m)$ — вектор правых частей и $x = (x_1, \ldots, x_n)$ — вектор ее неизвестных. Здесь и далее используется обозначение $x^+ = \max(0, x)$ для числа x и $p^+ = (p_1^+, \ldots, p_n^+)$ для вектора $p = (p_1, \ldots, p_n)$.

Функция квадратичной невязки лежит в основе многих методов линейного программирования [1–9]. Это, в частности, классический метод штрафных функций (с добавлением к обычной квадратичной невязке дополнительного линейного слагаемого), метод фейеровских приближений (при переходе от задачи линейного программирования к симметрической системе линейных неравенств), метод модифицированных функций Лагранжа (со сдвигом вектора правых частей системы на вектор двойственных оценок), метод нагруженного функционала (с добавлением к системе ограничений задачи еще одного неравенства-ограничения с переменной правой частью, лимитирующего искомое значение целевой функции) и другие. В той или иной своей части все они нуждаются в блоке безусловной минимизации функции обычной или расширенной квадратичной невязки, либо в блоке ее минимизации при наличии ограничений простейшего вида.

Функция квадратичной невязки $F(x)$ дифференцируема, а ее градиент удовлетворяет условию Липшица. Вместе с тем ее гессиан, т. е. матрица ее вторых производных $H(x) = (\partial^2 F/\partial x_i \partial x_j)_{n\times n}$ разрывна, что не позволяет в прямом виде применять к ней методы минимизации второго порядка. Тем не менее недавние работы О.Мангасарьяна и Р.Канзоу [10, 11] показали эффективность применения для этих целей ее квадратичных аппроксимаций вида

$$\hat{F}(x, x_0) = \frac{1}{2}\,(Ax-b)^T D_0 (Ax-b) + \frac{\delta}{2}\,||x-x_0||^2,$$

где x_0 — точка, в окрестности которой происходит аппроксимация, $\delta > 0$ — параметр регуляризации, D_0 — диагональная $m\times m$-матрица знаков невязок системы $Ax \leq b$ в точке x_0, т. е. такая матрица, i-й диагональный элемент которой равен 1, если i-е неравенство этой системы нарушается в точке x_0, и 0 в противном случае. Очередной шаг минимизации функции $F(x)$ состоит в пересчете текущего приближения по классической формуле $x_{k+1} = x_k + \alpha_k \Delta x_k$ с использованием направления спуска

$$\Delta x_k = -\big(A^T D_k A + \delta E\big)^{-1} \nabla F(x_k) = \hat{x}_k - x_k,$$

где E — единичная $m\times m$-матрица, D_k — диагональная $m\times m$-матрица знаков невязок системы $Ax \leq b$ в точке x_k, $\hat{x}_k$ — точка минимума гладкой функции $\hat{F}(\cdot, x_k)$ по первому аргументу, величина шага α_k либо постоянна (равна 1), либо лежит в интервале $(0, 1)$ и в ходе вычислений дробится в духе Армийо, либо определяется обычным для таких алгоритмов линейным поиском (заметим, что $\nabla F(x_k) = \nabla_1 \hat{F}(x_k, x_k)$).

Впрочем, при определении направления спуска можно обойтись и без явного вычисления и последующего обращения матрицы $H(x_k) = A^T D_k A + \delta E$. Систему линейных уравнений $H(x_k)\Delta x = -\nabla F(x_k)$, которой удовлетворяет искомое направление Δx_k, можно решить, например, методом сопряженных градиентов, где достаточно уметь находить произведение обращаемой матрицы на серию промежуточных векторов. Для этих целей можно вполне ограничиться тем мультипликативным представлением матрицы $H(x_k)$, которое нам доступно. Более того, для разреженных матриц, т. е. матриц, большая часть элементов которых равна нулю, такой подход может оказаться даже менее затратным в вычислительном отношении.

Несмотря на сделанное замечание о методе сопряженных градиентов и разреженных матрицах, наибольший успех методов квадратичной аппроксимации все же связывают с тем, насколько эффективно срабатывают классические методы линейной алгебры при вычислении и факторизации обобщенной матрицы вторых производных $H = A^T DA + \delta E$ (см., например, работы [12–15]). В этом отношении особенный интерес представляют разреженные матрицы, имеющие специальную структуру расположения своих ненулевых элементов и сохраняющие эту структуру целиком или частично при своем преобразовании в матрицу H. В качестве иллюстрирующего примера рассмотрим блочно-диагональную матрицу с вертикальным окаймлением

$$A = \left(\begin{array}{c|c|c|c|c} A_{11} & & & & A_{1,n} \\ \hline & A_{22} & & & A_{2,n} \\ \hline & & \ddots & & \vdots \\ \hline & & & A_{n-1,n-1} & A_{n-1,n} \end{array}\right).$$

Непосредственно проверяется, что матрица H также имеет блочно-диагональную структуру (правда, уже с двойным окаймлением)

$$H = \left(\begin{array}{c|c|c|c|c} H_{11} & & & & H_{1,n} \\ \hline & H_{22} & & & H_{2,n} \\ \hline & & \ddots & & \vdots \\ \hline & & & H_{n-1,n-1} & H_{n-1,n} \\ \hline H_{n,1} & H_{n,2} & \ldots & H_{n,n-1} & H_{n,n} \end{array}\right),$$

причем последняя не меняется при ее факторизации по Холецкому. Более того, факторизация всей матрицы в этом случае сводится к последовательной факторизации ее диагональных блоков [15], размерность которых несопоставима с размерностью всей матрицы. Прочие блоки результирующей матрицы получаются из исходных и факторизованных диагональных блоков простыми операциями.

В противовес предыдущему рассмотрим блочно-диагональную матрицу с горизонтальным окаймлением

$$A = \left(\begin{array}{c|c|c|c} A_{11} & & & \\ \hline & A_{22} & & \\ \hline & & \ddots & \\ \hline & & & A_{n-1,n-1} \\ \hline A_{n,1} & A_{n,2} & \ldots & A_{n,n-1} \end{array}\right).$$

Для нее матрица $H = A^T DA+\delta E$, скорее всего, окажется полностью заполнена ненулевыми элементами, т. е. потеряет важное для нас свойство разреженности. Напротив, матрица вида $\bar{H} = A\bar{D}^T A^T + \delta E$, в которой основные множители переставлены местами, очевидно сохраняет разреженную структуру исходного образца. Назовем эту форму гессиана (вообще говоря некоторой другой штрафной функции) двойственной к форме $H = A^T DA + \delta E$.

Очевидно, структура первой матрицы переходит в структуру второй матрицы при транспонировании последней. Это наталкивает нас на простую идею замены исходной задачи линейного программирования двойственной задачей и наоборот с тем, чтобы получить наиболее предпочтительную структуру расположения ненулевых элементов матрицы ее ограничений. Впрочем, это не единственный прием, позволяющий достичь поставленных выше целей, а именно — обеспечить условия, при которых применение стандартного матричного аппарата и многопроцессорной вычислительной техники для решения задач линейного программирования большой размерности было бы эффективно.

Доклад организован следующим образом. Вначале обсуждаются варианты квадратичной аппроксимации функции квадратичной невязки. Затем описываются различные алгоритмы и вычислительные схемы, в том числе параллельные, предназначенные для задач линейного программирования канонического формата с двусторонними границами на переменные. При этом, ориентируясь на блочно-диагональные матрицы с горизонтальным окаймлением, как наиболее характерные для прямых задач, будет делаться упор на алгоритмы, в которых основное ядро вычислений состоит в факторизации матриц типа $H = ADA^T$, хотя противоположный тип $H = A^T DA$ также будет рассмотрен. Последующее обсуждение связано с аналогичными алгоритмами и вычислительными схемами для стандартного формата записи задачи линейного программирования. Кроме того обсуждаются близкие схемы для симметрических систем линейных неравенств и уравнений, возникающих в линейном программировании. По всем алгоритмам приводятся результаты численных экспериментов. В заключение делается попытка выработать некоторые общие выводы и рекомендации по теме доклада.

В качестве иллюстрации приведем авторский вариант классического метода нагруженного функционала в применении к канонической задаче линейного программирования с блочно-диагональной структурой ограничений вида

$$\min\,(c,x):\quad B_i x_i = b_i\ (i=1,\ldots,r),\quad Ax = b_0,\quad x = [x_1,\ldots,x_r] \geq 0.$$

Здесь разбиение вектора неизвестных на подвекторы инициирует соответствующее разбиение на блоки вектора $c = [c_1,\ldots,c_r]$ и матрицы $A = [A_1,\ldots,A_r]$, где $c_i \in R^{n_i}$, B_i и A_i — $m_i \times n_i$- и $m_0 \times n_i$-матрицы соответственно, $m = \sum_{i=0}^r m_i$, $n = \sum_{i=1}^r n_i$. Ограничения $Ax = b_0$ назовем связывающими.

Соотнесем с выписанной задачей штрафную функцию

$$\Phi(x) = \frac{1}{2}\sum_{i=1}^{r}(||B_i x_i - b_i||^2 + ||x_i^-||^2)$$

и рассмотрим двухуровневый итерационный процесс

$$\begin{cases} x^k \ \in \ \mathrm{Arg}\ \min\{\Phi(x):\ Ax = b_0,\ (c,x) = \gamma_k\}, \\ \gamma_{k+1} \ = \ \gamma_k + 2\lambda_k^{-1}\,\Phi(x^k), \quad k = 0,1,2,\ldots. \end{cases}$$

Здесь на каждом шаге вначале ищется минимум штрафной функции при связывающих ограничениях и единственном дополнительном линейном ограничении на значение целевой функции, а затем это значение уточняется при помощи соответствующего множителя Лагранжа λ_k ($\lambda_k > 0$). Для запуска процесса необходимо знать нижнюю границу оптимального значения исходной задачи γ_0. В предположении, что все вычисления производятся точно, метод является конечным, т. е. при некотором K имеет место включение вектора x_K в оптимальное множество прямой задачи.

Для минимизации функции $\Phi(x)$, организуем стандартный внутренний итерационный процесс вида $x^{s+1} = x^s + \alpha_s \Delta x^s \quad (s = 0, 1, \dots)$, где $\alpha_s = 1$ (для повышения надежности можно подключать процедуру дробления шагового параметра по Армийо или линейный поиск), а направление корректировки текущего приближения определяется как $\Delta x^s = \hat{x}^s - x^s$. Вектор $\hat{x}^s$ представляет собой точку минимума гладкой функции

$$\hat{\Phi}(x, x^s) = \frac{1}{2} \sum_{i=1}^{r} \left(||B_i x_i - b_i||^2 + x_i^T D_s^{(i)} x_i \right) + \frac{\delta}{2} ||x - x^s||^2$$

при линейных ограничениях-равенствах $Ax = b_0, \quad (c, x) = \gamma_k$ и может быть получен из классической системы ККТ-условий

$$\begin{cases} (B_i^T B_i + D_s^{(i)} + \delta E_{n_i \times n_i})\, \hat{x}_i^s + \ A_i^T \hat{y} + \ \lambda_s c = B_i^T b_i + \delta x_i^s \quad (i = 1, \dots, r), \\ A_1 \ \hat{x}_1^s \ + \ \dots \ + \ A_r \ \hat{x}_r^s = b_0, \\ (c_1, \hat{x}_1^s) + \dots + (c_r, \hat{x}_r^s) = \gamma_k, \end{cases}$$

где $D_s^{(i)} = \mathrm{diag}(\ \mathrm{sign}((-x_i^s)_1^+), \dots, \mathrm{sign}((-x_i^s)_{n_i}^- +)\ )$, E — единичные матрицы соответствующих размеров, λ_s — уже упоминавшийся множитель Лагранжа, $\delta > 0$ — малый параметр (может не стремиться к нулю). Матрица коэффициентов этой системы имеет блочно-диагональный вид с симметричным горизонтальным и вертикальным окаймлением. Применяя матричный вариант метода Гаусса, получаем

Утверждение. Формулы решения выписанной ККТ-системы имеют вид

$$\hat{y} = H_0^{-1} \Big[\sum_{i=1}^{r} A_i H_i^{-1} \left(p_i - \hat{\lambda}^s c_i \right) - b_0 \Big], \quad \hat{x}_i^s = H_i^{-1}(p_i - A_i^T \hat{y} - \hat{\lambda}^s c_i), \quad i = 1, \dots, r,$$

где $\quad H_0 = \sum_{i=1}^{r} A_i H_i^{-1} A_i^T, \quad H_i = B_i^T B_i + D_s^{(i)} + \delta E_{n_i \times n_i}, \quad p_i = B_i^T b_i + \delta x_i^s, \quad i = 1, \dots, r,$

$$\hat{\lambda}^s = \frac{\gamma_k + (\sum_{i=1}^{r} c_i^T H_i^{-1} A_i^T) H_0^{-1} (\sum_{i=1}^{r} A_i H_i^{-1} p_i - b_0) - \sum_{i=1}^{r} c_i^T H_i^{-1} p_i}{(\sum_{i=1}^{r} c_i^T H_i^{-1} A_i^T) H_0^{-1} (\sum_{i=1}^{r} A_i H_i^{-1} c_i) - \sum_{i=1}^{r} c_i^T H_i^{-1} c_i}.$$

Таким образом, наиболее трудоемкая часть алгоритма заключается в обращении r матриц H_i размера $n_i \times n_i$ и одной матрицы H_0 размера $m_0 \times m_0$, причем обращение матриц H_i, $i = 1, \dots, r$, можно выполнять параллельно.

Более подробно тема освещена в работе автора [16].

Работа поддержана РФФИ, проект 07-01-00399.

ЛИТЕРАТУРА

1. Eremin I.I. Theory of Linear Optimization. Inverse and Ill-Posed Problems Series. VSP. Utrecht, Boston, Keln, Tokyo, 2002.
2. Еремин И.И., Мазуров Вл.Д. Нестационарные процессы математического программирования. –М.: Наука, 1979.
3. Евтушенко Ю.Г. Методы решения экстремальных задач и их применение в системах оптимизации. – М.: Наука, 1982.
4. Васильев Ф.П. Численные методы решения экстремальных задач. –М.: Наука. Гл. ред. физ.-мат.лит., 1988.
5. Васильев Ф.П., Иваницкий А.Ю. Линейное программирование. М.: Факториал, 2003.
6. Антипин А.С. Методы нелинейного программирования, основанные на прямой и двойственной модификации функции Лагранжа.- М.: ВНИИСИ (препринт), 1979.
7. Лебедев В.Ю. Приближенный алгоритм решения задачи ЛП // ЖФМиМФ, 1974, т.14, №4, 1052—1058.
8. Разумихин Б.С. Метод касательных для стохастических и динамических задач оптимизации // Автоматика и телемеханика. 1977, N 1, с.5-15.
9. Поляк Б.Т., Третьяков Н.В. Об одном итерационном методе линейного программирования и его экономической интерпретации // Экономика и матем. методы. 1972. т.8. вып.5. с.740–751.
10. Mangasarian O.L. A finite Newton method for classification // Optimizat. Meth. Software. 2002. Vol.17, p.913—930.
11. Kanzow C., Qi H., Qi L. On the minimum norm solution of linear program // J. Optimizat. Theory and Appl. 2003. Vol.116. p. 333–345.
12. Голиков А.И., Евтушенко Ю.Г., Моллаверди Н. Применение метода Ньютона к решению задач линейного программирования большой размерности // Журн. вычисл. мат. и матем. физики, 2004, том 44, №9, с. 1564–1573.
13. Голуб Дж., Ван Лоун Ч. Матричные вычисления: Пер. с англ. —М.: Мир, 1999.
14. Ортега Дж. Введение в параллельные и векторные методы решения линейных систем / Пер. с англ. – М.: Мир, 1991.
15. Gondzio J., Sarkissian R. Parallel interior-point solver for structured linear programs // Math. Progam., Ser. A. 2003, Vol. 96, p. 561–584.
16. Попов Л.Д. Квадратичная аппроксимация штрафных функций при решении задач линейного программирования большой размерности // Журн. вычислит. математ. и мат. физики. 2007. том 47, N2, с.206-221.

Попов Леонид Денисович, Институт математики и механики УрО РАН,
ул. С. Ковалевской, 16, Екатеринбург, 6200219, Россия,
тел. (8-343-3)75-34-23, факс (8-343-3)74-25-81, e-mail:popld@imm.uran.ru

Некрасов Антон Владимирович, Уральский государственный университет