УДК 518.12

Исследование быстрого градиентного метода при изменении длины шага

***А.В. Чернов*[1]**

[1]Московский физико-технический институт (государственный университет)

В данной работе вводится в рассмотрение однопараметрическое семейство методов на основе быстрого градиентного метода и численно исследуется вопрос сходимости таких методов для задачи энтропийно-линейного программирования с ограничениями типа равенства и неравенства (1)

$$f(x)=\sum_{i=1}^{n} x_i \ln\frac{x_i}{\xi_i} \to \min_{x\in G}; \quad G=\left\{x\in S_n(1): C_1x=b_1; C_2x\le b_2\right\}. \tag{1}$$

Здесь и далее $S_n(1)=\left\{x\in R_{++}^n : \sum_{i=1}^{n} x_i=1\right\}$ – вероятностный симплекс, $b_1\in R^{m_1}$ и $C_1\in R^{m_1\times n}$, $b_2\in R^{m_2}$ и $C_2\in R^{m_2\times n}$, $C=\left[C_1,C_2\right]\in R^{(m_1+m_2)\times n}$ и $b=\left[b_1;b_2\right]\in R^{(m_1+m_2)}$.

Для задачи (1), полагая $S_n(1)$ простым множеством, нетрудно выписать двойственную задачу (2) в форме задачи минимизации функции $\varphi(y)$.

$$\varphi(y)=\langle y,b\rangle+\ln\left(\sum_{i=1}^{n}\xi_i \exp\left(-\left[C^T y\right]_i\right)\right)\to \max_{y\in Q}; \quad Q=R^{m_1}\times R_+^{m_2} \tag{2}$$

При этом функция $x(y)$, полученная при построении двойственной задачи, вычисляется по формуле

$$x_i(y)=\frac{\xi_i \exp\left(-\left[C^T y\right]_i\right)}{\sum_{j=1}^{n}\xi_j \exp\left(-\left[C^T y\right]_j\right)}$$

В работах [1], [2] обоснована возможность решения такой задачи с помощью быстрого градиентного метода в форме (3) без рестартов при значении параметра $m=1$.

$$\tilde{y}_k^j=\begin{cases}\left(y_k-\dfrac{m}{L}\nabla\varphi(y_k)\right)_+^j & j>m_1;\\ \left(y_k-\dfrac{m}{L}\nabla\varphi(y_k)\right)^j & j\le m_1;\end{cases} \qquad \breve{y}_k^j=\begin{cases}\left(y_k-\dfrac{m}{L}\sum_{i=0}^{k}\alpha_i\nabla\varphi(y_i)\right)_+^j & j>m_1;\\ \left(y_k-\dfrac{m}{L}\sum_{i=0}^{k}\alpha_i\nabla\varphi(y_i)\right)^j & j\le m_1;\end{cases}$$

$$y_{k+1}=\tau_k\tilde{y}_k+(1-\tau_k)\breve{y}_k; \quad x_k=\sum_{i=0}^{k}\frac{\alpha_i}{A_k}x(y_i). \tag{3}$$

Здесь $x_+=\max(x,0)$ – функция-срезка, $\{\alpha_i\}_{i=0}^{\infty}$, $\{A_i\}_{i=0}^{\infty}$, $\{\tau_i\}_{i=0}^{\infty}$ – произвольные последовательности, удовлетворяющие следующим свойствам:

$$\alpha_0\in(0,1]; \quad A_k=\sum_{i=0}^{k}\alpha_i; \quad \alpha_k^2\le A_k; \quad \tau_k=\frac{\alpha_{k+1}}{A_{k+1}}.$$

Численные эксперименты были проведены на ЭВМ с процессором Intel Core i5, 2.5 ГГц и оперативной памятью 6 Гб в среде Matlab 2012® (8.0) под управлением операционной системой Microsoft Windows 7 (64 разрядная). Поиск решения задачи (1) проводился с точностью ε_f по функции и ε_g по ограничениям. Решение задачи (x,y) будем искать по следующим критериям

1) $f(x)+\varphi(y)\le\varepsilon_f$ (проверка зазора двойственности)
2) $\left\|C_1x-b_1\right\|+\left\|(C_2x-b_2)_+\right\|\le\varepsilon_g$

Результаты численных экспериментов приведены на рис. 1, 2. При этом точности ε_f по функции и ε_g по ограничениям выбиралась как 1% от минимального значения функции $f(x)$ и от отклонения начальной точки соответственно

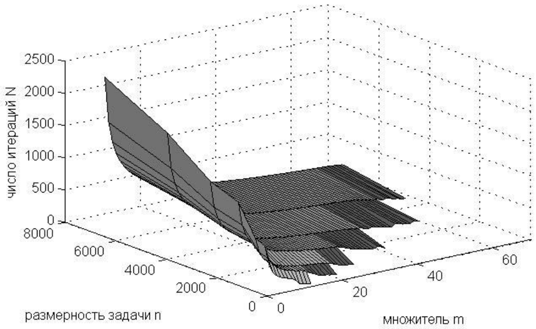

Рис. 1. График зависимости числа итераций от размерности задачи и множителя m

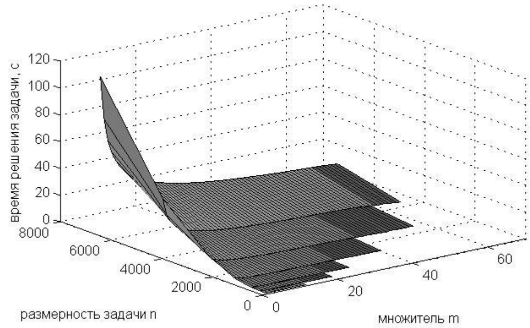

Рис. 2. График зависимости времени поиска решения от размерности задачи и множителя m

Из рис. 1 и рис. 2 видно, что увеличение шага в быстром градиентном методе позволяет на порядок ускорить решение исследуемой задачи. Такое поведение объясняется тем, что константа L, определяющая величину шага, выбирается из соображений гарантии сходимости метода с запасом, а такое значение не всегда оптимально для величины шага. Эксперименты также показали, что дальнейшее увеличение длины шага (даже небольшое, по сравнению со значением, указанным на графике) значительно увеличивает время сходимости метода, а также приводит к тому, что последний начинает расходиться.

На рис. 3 отражена экспериментальная зависимость оптимального значения параметра m, дающего максимальный эффект от размерности задачи n.

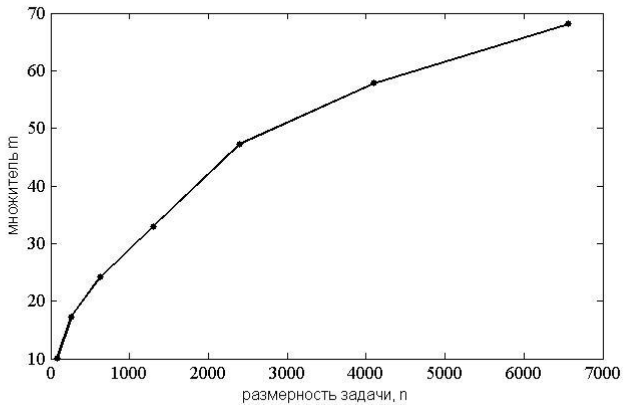

Рис. 3. График изменения зависимости значения m от размерности задачи n

Из рис. 3 видно, что оптимальное значение параметра m растет вместе с размерностью задачи.

Сказанное выше позволяет сделать вывод, что с помощью грамотного выбора параметра m можно добиться существенного, до 10 раз, ускорения работы быстрого градиентного метода. Причем величина оптимального значения параметра m растет вместе с размерностью задачи. Дополнительно можно, что даже небольшие значения параметра m, например 5, 10, для задач большой размерности уже дают кратное уменьшение времени решения задачи и числа итераций необходимого для такого решения.

Автор выражает благодарность Бирюкову А.Г. и Гасникову А.В. за ряд ценных замечаний

Исследование выполнено при поддержке гранта РФФИ, проект номер 14-01-00722.

Литература

1. *Чернов А.В.* Прямо-двойственный метод решения задачи энтропийно-линейного программирования // Интеллектуальные системы. Теория и приложения. 2016, Т.20, В.1, С. 39-57.
2. *Chernov A., Dvurechensky P., Gasnikov A.* Fast Primal-Dual Gradient Method for Strongly Convex Minimization Problems with Linear Constraints // 9th International Conference «Discrete Optimization and Operations Research». 2016. P. 391-403