Задача распределенной оптимизации

Oleg Vasilev

April 2, 2020

1 Постановка задачи

Задан неориентированный граф коммуникации: в нем n агентов. Задана матрица смежности A и диагональная матрица степеней вершин Deg. Опеределена матрица Кирхгофа (Laplacian matrix) $L = Deg - A$ - т. е. сумма по каждой строчке и столбцу равна нулю.

Для каждого агента задана функция $f_i(x)$, и все агенты кооперативно действуют, чтобы оптимизировать сумму этих функций.

2 Сведение

Пусть каждый агент хранит свою копию переменных x_i. Введем функцию, в которой каждый агент считает свою функцию на своих переменных: $f'(x_1 \dots x_n) = \sum_i f_i(x_i)$. Тогда можно сформулировать исходную задачу в виде задачи с ограничениями:

$$\min f' \tag{1}$$

$$Lx = 0 \iff \sqrt{L}x = 0 \tag{2}$$

Т. е. на сумме функций достигается минимум при том, что очередной раунд коммуникации (домножение на L) не может ничего изменить.

Теперь полученную задачу можно свести к неограниченной при помощи аугментированного метода Лагранжа

$$A(x, u) = f'(x_1 \dots x_n) + \frac{\alpha}{2} \sum_{i,j} x_i \cdot x_j L_{ij} + \beta u^T \sqrt{L}x \tag{3}$$

Таким образом, переменные u можно рассматривать как двойственные - они показывают насколько то или иное ограничение выполннено.

Теперь можно решить полученную задачу прямо-двойственным алгоритмом, выполняя шаги одновременно по x и u. По x это будет шаг градиентного спуска, а по v - подъема в сторону увеличения важности насыщенных ограничений в выражении $A(x, u)$.

Можно выписать формулы пересчета переменных для i-того агента:

$$v_i = \sqrt{L}u_i \tag{4}$$

$$x_i \leftarrow x_i - \eta(\alpha \sum_j L_{ij}x_j + \beta v_i + \nabla f_i(x_i)) \tag{5}$$

$$v_i \leftarrow v_i + \eta\beta(\sum_j L_{ij}x_j) \tag{6}$$

Удивительным образом выражение для пересчета переменных зависит только от переменных, которые принадлежат "соседним" агентам - то есть тем, для которых в соответствующей строке L стоит не-ноль.

Далее, если специальным образом выбрать α и β, то можно добиться линейной скорости сходимости (т. е. очередное значение переменной в константу раз ближе к оптимуму). Константа будет зависить от матричных свойств L, т. е. чем сложнее добраться от одной части графа до другой, тем хуже будет сходимость.

3 Метод штрафов

Можно использовать альтернативный способ сведения задачи с ограничениями к задаче без ограничений

$$\arg\min f(x) + \frac{R_y^2}{\varepsilon}||Ax||_2^2 \tag{7}$$

$$\arg\min \sum_i f(x_i) + \frac{R_y^2}{\varepsilon}\sum_i\sum_k\left(\sum_j L_{ij}x_j^{(k)}\right)^2 \tag{8}$$

$$x_i^{(k)} \leftarrow \nabla^{(k)} f(x_i) + \frac{2R_y^2}{\varepsilon}\left(\sum_j L_{ij}x_j^{(k)}\right)L_{ii} \tag{9}$$

$$x_i \leftarrow \nabla f(x_i) + \frac{2R_y^2}{\varepsilon}\left(\sum_j L_{ij}x_j\right)L_{ii} \tag{10}$$

4 Шаг градиентного слайдинга

$$u_t = \arg\min\{g(u) + l_h(u_{t-1}, u) + \beta V(x, u) + \beta p_t V(u_{t-1}, u) + (u)\} \tag{11}$$

$$l_f(x, y) = f(x) + < \nabla f(x), y - x > \tag{12}$$

$$g(y) = l_f(x, y) \tag{13}$$

$$V(x, y) = \frac{1}{2}||x - y||_2^2 \tag{14}$$

$$\nabla u_t(u) = \nabla f(x) + \nabla h(u_{t-1}) - \beta(x - u) - \beta p_t(u_{t-1} - u) = 0 \tag{15}$$

$$\beta(1 + p_t)u = \beta x + \beta p_t u_{t-1} + \nabla f(x) + \nabla h(u_{t-1}) \tag{16}$$

$$u = \frac{1}{1 + p_t}\left(x + \frac{1}{\beta}\nabla f(x)\right) + \left(u_{t-1} + \frac{1}{(1 + p_t)\beta}\nabla h(u_{t-1})\right) \tag{17}$$