

РАНДОМИЗИРОВАННЫЙ РАСПРЕДЕЛЕННЫЙ АДАПТИВНЫЙ АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ О МАКСИМАЛЬНОМ ПОТОКЕ*

Мальковский Н.В.¹

¹СПбГУ, Санкт-Петербург, Россия.

Аннотация

В этой статье предлагается метод решения одной из классических задач оптимизации — задачи о максимальном потоке. Алгоритм основан на общей процедуре балансирования дуг и не дает выигрыша по времени работы относительно существующих методов, однако обладает другими полезными свойствами: простую реализацию на распределенных вычислительных системах, сохранение близости к оптимальному решению при изменении параметров сети во времени. Рассматривается два варианта алгоритма: рандомизированный и синхронный. Рандомизированная версия использует идеи покоординатного градиентного спуска и рандомизированных алгоритмов стохастической аппроксимации. Рандомизированная версия оказывается предпочтительней, так как имеет такой же порядок сходимости (экспоненциальный), но при этом устойчива к помехам и допускает более простую распределенную реализацию по сравнению с синхронной версией.

Ключевые слова: оптимизация с ограничениями, градиентный спуск, распределенные алгоритмы, рандомизированные алгоритмы, стохастическая аппроксимация, задача о максимальном потоке.

Цитирование: Мальковский Н.В. Рандомизированный распределенный адаптивный алгоритм решения задачи о максимальном потоке // Компьютерные инструменты в образовании, 2016. № 5. С. 46–61 .

1. ВВЕДЕНИЕ

Задача о максимальном потоке является одной из классических задач оптимизации. Она имеет многие приложения в сетевых системах, а также часто возникает как техническая подзадача в других задачах оптимизации. Задача изучается уже более полувека, наиболее свежий обзор алгоритмов её решения можно найти в [1]. В этой работе представлен метод решения задачи о максимальном потоке, не являющийся самым эффективным по времени работы, однако обладающий некоторыми прочими полезными свойствами: возможностью распределенной / мультиагентной реализации (то есть если задача возникает в распределенной сети, при этом вычислительные устройства этой сети образуют вершины графа, а каналы связи образуют дуги графа, то алгоритм имеет

*Работа выполнена при поддержке гранта СПбГУ 6.37.181.2014.

«естественную» реализацию на узлах этой сети) и адаптивности (если пропускная способность дуг меняется по мере вычислений, то алгоритм сохраняет близость к максимальному потоку при условии, что эти изменения ограничены и достаточно малы). За основу взят метод балансирования дуг, описанный в [2]: оказывается, этот метод можно представить как проективный градиентный спуск [3]. В рамках этой работы будут исследованы две вариации метода: синхронный (обычный проективный градиентный спуск) и рандомизированный асинхронный (специальный вид рандомизированного алгоритма стохастической аппроксимации [4]). Второй вариант больше походит на метод, описанный в [2], с той лишь разницей, что используется случайный выбор дуг.

Основной целью этой статьи является разработка специфических алгоритмов решения задачи о максимальном потоке. Стоит отметить, что результаты работы также обобщают идеи проективного градиентного спуска [3], рандомизированного алгоритма стохастической аппроксимации [4, 5], алгоритмов «сплетен» [6] и протоколов консенсуса [7]. По сравнению с классическими результатами для проективного градиентного спуска, в этой работе получена экспоненциальная скорость сходимости для квадратичной функции, не являющейся сильно выпуклой. По сравнению с рандомизированными алгоритмами стохастической аппроксимации, рассматриваемыми в [4], в этой работе решается задача с ограничениями в виде неравенств. Алгоритмы консенсуса по общему стилю схожи с синхронной версией разработанного алгоритма, а алгоритмы сплетен — с рандомизированной версией. Связь между разработанными методами и сплетнями / консенсусом состоит в следующем: консенсус и сплетни минимизируют функционал $\mathbf{y}^T B B^T \mathbf{y}$, в то время как разработанный алгоритм минимизирует функционал $\mathbf{x}^T B^T B \mathbf{x}$ (в обоих случаях B — матрица инцидентности некоторого графа). Ключевым различием является наличие дополнительных ограничений в задаче, заданных неравенствами. Лемма 3 является ключевой для получения оценок сходимости, аналогичных [6, 7].

2. ЗАДАЧА О МАКСИМАЛЬНОМ ПОТОКЕ

Пусть задан некоторый ориентированный граф $G = \langle V, E \rangle$ с выделенными вершинами истока s и стока t . Традиционно задача о максимальном потоке ставится в следующем виде:

$$\begin{array}{ll} \text{максимизировать} & \text{val}(\mathbf{x}) = \sum_{e \in \text{in}(t)} x_e - \sum_{e \in \text{out}(t)} x_e, \\ \text{при условии} & \begin{cases} \sum_{e \in \text{in}(i)} x_e - \sum_{e \in \text{out}(i)} x_e = 0, & i \in V, i \neq s, t, \\ 0 \leq x_e \leq c_e, \end{cases} \end{array} \quad (1)$$

где переменными являются величины потока по дугам x_1, \dots, x_m (здесь и далее $\mathbf{x} = (x_1, \dots, x_m)^T$), $\text{in}(i)$ — множество входящих в i дуг, $\text{out}(i)$ — множество дуг, выходящих из i . Первое условие означает, что поток не задерживается в промежуточных вершинах. Единственные две вершины с неотрицательной разностью входящих / исходящих потоков — s и t . Более того,

$$\sum_{e \in \text{in}(t)} x_e - \sum_{e \in \text{out}(t)} x_e = - \left(\sum_{e \in \text{in}(s)} x_e - \sum_{e \in \text{out}(s)} x_e \right).$$

Задачу (1) можно записать в более компактной форме, используя матрицу инцидентности B :

$$B_{ie} = \begin{cases} 1, & e \text{ выходит из } i, \\ -1, & e \text{ входит в } i, \\ 0 & \text{в остальных случаях} \end{cases}$$

Обозначив вектор χ^{st} : $\chi_s^{st} = -1$, $\chi_t^{st} = 1$, $\chi_i^{st} = 0$, $i \neq s, t$ задачу (1) перепишем в виде:

$$\begin{array}{ll} \text{максимизировать} & v, \\ \text{при условии} & \begin{cases} B\mathbf{x} = v\chi^{st}, \\ 0 \leq x_e \leq c_e, \end{cases} \end{array}$$

3. ОБЩАЯ СХЕМА МЕТОДА БАЛАНСИРОВАНИЯ ДУГ

Метод балансирования дуг состоит из инициализации и последовательном применении некоторых локальных изменений потока:

Function Initialize(G, c)

```

for  $(i, j) \in E$  do
  if  $i = s$  или  $j = t$  then
     $x_{ij} \leftarrow c_{ij}$ ;
   $x_{ij} \leftarrow 0$ ;
return  $x$ ;

```

Function Move(G, c, x, e)

```

 $x_e \leftarrow x_e - \frac{1}{2}[B^T Bx]_e$ ;
if  $x_e > c_e$  then
   $x_e \leftarrow c_e$ ;
if  $x_e < 0$  then
   $x_e \leftarrow 0$ ;

```

Function ArcBalancing($G, c, \text{stopping criterion}$)

```

 $\mathbf{x} \leftarrow \text{Initialize}(G, c)$ ;
while stopping criterion is not satisfied do
  pick  $e = (i, j) \in E$ :  $i, j \neq s, t$ ;
   $\text{Move}(G, c, \mathbf{x}, e)$ ;
return  $\mathbf{x}$ ;

```

Величину $[Bx]_i = \sum_{e \in \text{in}(i)} x_e - \sum_{e \in \text{out}(i)} x_e$ обычно принято называть *избытком* вершины i . Величина $[B^T Bx]_{(i,j)} = [Bx]_j - [Bx]_i$ есть разница избытков вершин i, j . Отсюда название алгоритма: функция *Move* пытается изменить значение потока на дуге e так, чтобы разница между избытками на концах дуги была минимальной при условии, что $0 \leq x_e \leq c_e$. При этом от способа выбора дуги зависит сходимость функции *ArcBalancing*. Обычно выбираются только те дуги, для которых функция *Move* изменит значение потока, однако если выбирать дуги таким образом, а останавливаться в случае, когда таких дуг нет, то нет никакой гарантии, что *ArcBalancing* закончит работу.

Одна из основных идей, предложенных в этой работе, заключается в том, что при равномерном выборе дуг для операции *Move* возможно доказать асимптотическую сходимость, при этом можно достаточно легко извлечь максимальный поток / минимальный разрез из предельного состояния. Далее в работе будут проанализированы следующие варианты метода *ArcBalancing*: *RandomizedArcBalancing* и *ConcurrentArcBalancing*.

Algorithm 1: RandomizedArcBalancing(G, \mathbf{c} , stopping criterion)

```

 $\mathbf{x} \leftarrow \text{Initialize}(G, \mathbf{c});$ 
while stopping criterion is not satisfied do
    pick random  $e = (i, j) \in E' : i, j \neq s, t$  w.p.  $p_i$ ;
    Move( $G, \mathbf{c}, \mathbf{x}, e$ );
return  $\mathbf{x}$ ;

```

Algorithm 2: ConcurrentArcBalancing(G, \mathbf{c} , stopping criterion)

```

 $\mathbf{x} \leftarrow 0_m;$ 
while stopping criterion is not satisfied do
     $\mathbf{x} \leftarrow \mathbf{x} - \frac{1}{2d} (B'^T B' \mathbf{x} + B'^T \mathbf{b});$ 
    for  $e \in E'$  do
        if  $x_e > c_e$  then
             $x_e \leftarrow c_e;$ 
        if  $x_e < 0$  then
             $x_e \leftarrow 0;$ 
return  $\mathbf{x}$ ;

```

В описании этих методов используются обозначения $E' = \{(i, j) \in E \mid i \neq s, j \neq t\}$, $d_i^- = |\{(i, j) \in E'\}|$, $d_i^+ = |\{(j, i) \in E'\}|$, $d > 0$ — некоторый скалярный параметр, B' — матрица инцидентности графа $G' = \langle V, E' \rangle$. Положим, что $\mathbf{c} = \begin{bmatrix} \mathbf{c}' \\ \mathbf{c}'' \end{bmatrix}$, \mathbf{c}' соответствует дугам из E' , \mathbf{c}'' соответствует дугам из $E \setminus E'$, $\mathbf{b} = B'' \mathbf{c}''$.

4. СХОДИМОСТЬ МЕТОДОВ

Для начала отметим, что оба метода очень схожи с *проективным градиентным спуском*, более того, если обозначить $f(\mathbf{x}) = \frac{1}{d} (\frac{1}{2} \mathbf{x}^T B'^T B' \mathbf{x} + \mathbf{x}^T B'^T \mathbf{b})$, $\nabla f(\mathbf{x}) = \frac{1}{d} (B'^T B' \mathbf{x} + B'^T \mathbf{b})$, то алгоритм *ConcurrentArcBalancing* в точности является проективным градиентным спуском для функции f . Общий метод проективного градиентного спуска был впервые сформулирован в работе [3]. Метод *RandomizedArcBalancing* обновляет поток похожим образом: на каждом шаге случайным образом выбирается координата и делается шаг градиентного спуска только по этой координате. Прежде чем переходить к анализу сходимости, выпишем некоторые ключевые свойства задачи.

Лемма 1. Если D — диагональная матрица с неотрицательными элементами, а $B \in \mathbb{R}^{n \times m}$ — произвольная вещественная матрица, то матрица $DB^T B$ имеет только вещественные неотрицательные собственные числа. Более того, если все диагональные элементы D строго положительны, то

$$DB^T B \mathbf{x} = 0_m \Leftrightarrow B \mathbf{x} = 0_n.$$

Доказательство. Так как D — диагональная матрица с неотрицательными элементами, то $DB^T B$ имеет только вещественные собственные числа, для матрицы D допустимо разложение

$$D = \sqrt{D}^T \sqrt{D},$$

где $\sqrt{D} = \text{diag}\{\sqrt{D_{11}}, \dots, \sqrt{D_{mm}}\}$. С другой стороны, у матриц $DB^T B$ и BDB^T одинаковый набор ненулевых собственных чисел:

$$\lambda \mathbf{x} = DB^T B \mathbf{x} \Rightarrow BDB^T (B \mathbf{x}) = B(DB^T B \mathbf{x}) = B(\lambda \mathbf{x}) = \lambda (B \mathbf{x}),$$

а значит, при $B \mathbf{x} \neq 0_n$ λ является собственным числом BDB^T . Если же $B \mathbf{x} = 0_n$, то $DB^T B \mathbf{x} = 0_m$, а значит, $\lambda = 0$. Аналогично в обратную сторону:

$$\lambda \mathbf{x}^T = \mathbf{x}^T BDB^T \Rightarrow (\mathbf{x}^T B)DB^T B = (\mathbf{x}^T BDB^T)B = (\lambda \mathbf{x}^T)B = \lambda (\mathbf{x}^T B),$$

при $\mathbf{x}^T B \neq 0_n$ λ является собственным числом $DB^T B$, при $\mathbf{x}^T B = 0_n$ $\mathbf{x}^T BDB^T = 0_m$, а значит, $\lambda = 0$. В силу

$$\mathbf{x}^T BDB^T \mathbf{x} = \|\sqrt{D}B^T \mathbf{x}\|^2 \geq 0,$$

получаем, что $DB^T B$ имеет только неотрицательные собственные числа. Наконец, если все диагональные элементы D строго положительны, то D обратима, а значит,

$$DB^T B \mathbf{x} = 0_m \Rightarrow 0 = \mathbf{x}D^{-1}DB^T B \mathbf{x} = \|\mathbf{Bx}\|^2 \Leftrightarrow B \mathbf{x} = 0_n.$$

□

Лемма 2. Если G' — граф без петель и кратных дуг, $D_{ee} \leq \max\{1/2(d_i^- + d_j^+), 1/(nd_i^-)\}$ при $e = (i, j)$, то $\sigma(DB'^T B') \leq 2$.

Доказательство. Для $e = (i, j)$ обозначим $d_e = d_{ij} = D_{ee}$, $d_{ij} = 0$ при $(i, j) \notin E'$. Так как у матриц $DB'^T B'$ и $B'DB'^T$ одинаковый набор ненулевых собственных чисел, то у них одинаковый спектральный радиус. Непосредственным вычислением получаем, что

$$[B'DB'^T]_{ij} = \begin{cases} -(d_{ij} + d_{ji}), & i \neq j, \\ \sum_{k \neq i} d_{ik} + d_{ki}, & i = j. \end{cases}$$

Если $d_{ij} \leq 1/2(d_i^- + d_j^+)$, то

$$\begin{aligned} \sum_{j \neq i} (d_{ij} + d_{ji}) &\leq \sum_{(i,j) \in E'} \frac{1}{2(d_i^- + d_j^+)} + \sum_{(j,i) \in E'} \frac{1}{2(d_j^- + d_i^+)} \leq \\ &\sum_{(i,j) \in E'} \frac{1}{2d_i^-} + \sum_{(j,i) \in E'} \frac{1}{2d_i^+} \leq 1. \end{aligned}$$

Последнее неравенство обращается в равенство, если обе суммы содержат хотя бы одно слагаемое. Если же $d_e \leq 1/(nd_i^-)$, то

$$\sum_{j \neq i} (d_{ij} + d_{ji}) \leq \sum_{(i,j) \in E'} \frac{1}{nd_i^-} + \sum_{(j,i) \in E'} \frac{1}{nd_j^-} \leq \frac{1}{n} + \frac{d_i^+}{n} \leq 1.$$

Как и в предыдущем случае, предпоследнее неравенство переходит в равенство, если соответствующие суммы содержат хотя бы одно слагаемое. В последнем неравенстве мы воспользовались тем, что в графе без петель и кратных дуг $\forall i : d_i^+ \leq n - 1$. Наконец, в силу критерия Гершгорина, мы получаем, что все собственные числа матрицы $B'DB'^T$ удовлетворяют соотношению $|\lambda - \max_i \sum_{j \neq i} (d_{ij} + d_{ji})| \leq \max_i \sum_{j \neq i} (d_{ij} + d_{ji})$, из чего следует, что $|\lambda - 1| \leq 1$, а значит, $\sigma(B'^T B' D) \leq 2$. □

Заметим, что в лемме 2 такая же оценка для спектрального радиуса получается, если выполняется неравенство $d_e \leq 1/(nd_j^+)$. Далее, для вывода основных результатов нам понадобится следующая лемма о евклидовой проекции на многогранники.

Лемма 3. Пусть $Q = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{d}\}$, $P_Q(\mathbf{x}) = \operatorname{argmin}_{\mathbf{y} \in Q} \|\mathbf{x} - \mathbf{y}\|$, $\mathbf{b}_1, \dots, \mathbf{b}_n$ — некоторый ортогональный базис, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\alpha_1, \dots, \alpha_n$ и β_1, \dots, β_n — разложения \mathbf{x}, \mathbf{y} по базису \mathbf{b} , $\alpha'_1, \dots, \alpha'_n$ и $\beta'_1, \dots, \beta'_n$ — разложения $P_Q(\mathbf{x}), P_Q(\mathbf{y})$ по базису \mathbf{b} , тогда

$$|\alpha'_i - \beta'_i| \leq |\alpha_i - \beta_i|, \quad 1 \leq i \leq n.$$

Доказательство. Во-первых, отметим, что P_Q представляется в виде композиции проекций $P_{Q_1} \circ \dots \circ P_{Q_m}$, где $Q_i = \{\mathbf{x} \in \mathbb{R}^n \mid A_i \mathbf{x} \leq d_i\}$, поэтому достаточно доказать указанное свойство для полупространства. Пусть $Q = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} \leq d\}$. Рассмотрим некоторый ортогональный базис, содержащий вектор \mathbf{a} : $\mathbf{a} = \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, g_1, \dots, g_n и h_1, \dots, h_n — разложения \mathbf{x}, \mathbf{y} по этому базису. Очевидно, что разложение по этому базису для $P_Q(\mathbf{x}), P_Q(\mathbf{y})$ задается формулами:

$$\begin{cases} g'_1 = \min\left(g_1, \frac{d}{\|\mathbf{a}\|^2}\right) \\ g'_i = g_i, \quad i > 1 \end{cases}, \quad \begin{cases} h'_1 = \min\left(h_1, \frac{d}{\|\mathbf{a}\|^2}\right) \\ h'_i = h_i, \quad i > 1 \end{cases}$$

из чего следует

$$\begin{cases} |g'_1 - h'_1| \leq |g_1 - h_1| \\ g'_i - h'_i = g_i - h_i, \quad i > 1. \end{cases}$$

Так как \mathbf{b} — ортогональный базис, то

$$\alpha_i \|\mathbf{b}_i\| = \mathbf{x}^T \mathbf{b}_i = \sum_{j=1}^n g_j \mathbf{a}_j^T \mathbf{b}_i.$$

Остальные коэффициенты получаются аналогичным образом. Наконец,

$$\begin{aligned} |\alpha'_i - \beta'_i| &= \frac{1}{\|\mathbf{b}_i\|} \left| \sum_{j=1}^n g'_j \mathbf{a}_j^T \mathbf{b}_i - \sum_{j=1}^n h'_j \mathbf{a}_j^T \mathbf{b}_i \right| = \frac{|g'_1 - h'_1| \|\mathbf{a}_1^T \mathbf{b}_i\|}{\|\mathbf{b}_i\|} \leq \\ &= \frac{|g_1 - h_1| \|\mathbf{a}_1^T \mathbf{b}_i\|}{\|\mathbf{b}_i\|} = \frac{1}{\|\mathbf{b}_i\|} \left| \sum_{j=1}^n g_j \mathbf{a}_j^T \mathbf{b}_i - \sum_{j=1}^n h_j \mathbf{a}_j^T \mathbf{b}_i \right| = |\alpha_i - \beta_i|. \end{aligned}$$

□

Далее мы воспользуемся следующим очевидным следствием этой леммы: если M — линейное подпространство \mathbb{R}^n , $\mathbf{x} = \mathbf{u}_x + \mathbf{v}_x$, $\mathbf{y} = \mathbf{u}_y + \mathbf{v}_y$, $P_Q(\mathbf{x}) = \mathbf{u}'_x + \mathbf{v}'_x$, $P_Q(\mathbf{y}) = \mathbf{u}'_y + \mathbf{v}'_y$, $\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}'_x, \mathbf{u}'_y \in M$, $\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}'_x, \mathbf{v}'_y \in M^\perp$, то

$$\|\mathbf{u}'_x - \mathbf{u}'_y\| \leq \|\mathbf{u}_x - \mathbf{u}_y\|,$$

$$\|\mathbf{v}'_x - \mathbf{v}'_y\| \leq \|\mathbf{v}_x - \mathbf{v}_y\|.$$

4.1. Сходимость синхронного алгоритма

Во всей этой секции мы будем считать, что матрица $\|\frac{1}{d} B'^T B\| \leq 2$ (параметр d можно выбирать как угодно, поэтому достаточно легко гарантировать это свойство, выбрав d , используя, например, лемму (2)). Как и раньше, обозначим $f(\mathbf{x}) = \frac{1}{d} (\frac{1}{2} \mathbf{x}^T B'^T B' \mathbf{x} + \mathbf{x}^T B'^T \mathbf{b})$,

$\nabla f(\mathbf{x}) = \frac{1}{d}(B'^T B' \mathbf{x} + B'^T \mathbf{b})$. Метод *ConcurrentArcBalancing* генерирует последовательность $\{\mathbf{x}^k\}_{k=0}^\infty$ по правилу

$$\mathbf{x}^{k+1} = P_Q \left(\mathbf{x}^k - \frac{1}{2} \nabla f(\mathbf{x}^k) \right), \quad (2)$$

при этом начальное приближение выбирается как $\mathbf{x}^0 = \mathbf{0}_m$. Если бы матрица $B'^T B'$ была строго положительно определенной, то можно было бы вывести экспоненциальную сходимость из результатов [3]. К сожалению, это происходит только в случае, если G' не содержит циклов (без учета ориентации). Тем не менее, для квадратичных функций также можно доказать экспоненциальную сходимость. Обозначим $A = \frac{1}{d} B'^T B'$, $\mathbf{w} = \frac{1}{d} B'^T \mathbf{b}$.

Теорема 1. Пусть \mathbf{x}^k — последовательность, генерируемая по правилу (2) с любым начальным приближением $\mathbf{x}^0 \in Q$, $\|A\| \leq 2$, $\gamma = \max_{\mathbf{x} \in Q} \|A\mathbf{x} + \mathbf{w}\|$, $\lambda_2(A)$ — минимальное отличное от нуля собственное число матрицы A , $q = \left(1 - \frac{\lambda_2(A)}{2}\right)$, тогда

1. Последовательность \mathbf{x}^k сходится к некоторой точке минимума \mathbf{x}^* f на Q , причем

$$\|\mathbf{x}^k - \mathbf{x}^*\| \leq \frac{q^{k/2}}{1 - \sqrt{q}} \sqrt{\gamma \|\mathbf{x}^0 - \mathbf{x}^*\|}.$$

2. $f(\mathbf{x}^k)$ убывает, причем

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq \gamma q^k \|\mathbf{x}^0 - \mathbf{x}^*\|.$$

3. Если Ω — ядро A , то есть $\Omega = \{\mathbf{y} \in \mathbb{R}^m \mid A\mathbf{y} = \mathbf{0}_m\}$, то

$$\delta(\mathbf{x}^k - \mathbf{x}^*, \Omega) \leq q^k \|\mathbf{x}^0 - \mathbf{x}^*\|,$$

где $\delta(\cdot, \cdot)$ — евклидова метрика.

Доказательство.

$$\begin{aligned} f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) &= \frac{1}{2} \mathbf{x}^{k+1T} A \mathbf{x}^{k+1} + \mathbf{w}^T \mathbf{x}^{k+1} - \frac{1}{2} \mathbf{x}^{kT} A \mathbf{x}^k - \mathbf{w}^T \mathbf{x}^k = \\ &= \frac{1}{2} (\mathbf{x}^{k+1} - \mathbf{x}^k)^T A (\mathbf{x}^{k+1} - \mathbf{x}^k) + (A \mathbf{x}^k + \mathbf{w})^T (\mathbf{x}^{k+1} - \mathbf{x}^k) \leq \\ &= (A \mathbf{x}^k + \mathbf{w})^T (\mathbf{x}^{k+1} - \mathbf{x}^k) + \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 = \\ \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - 2(\mathbf{x}^k - \frac{1}{2}(A \mathbf{x}^k + \mathbf{w}) - \mathbf{x}^{k+1})^T (\mathbf{x}^{k+1} - \mathbf{x}^k) - 2(\mathbf{x}^{k+1} - \mathbf{x}^k)^T (\mathbf{x}^{k+1} - \mathbf{x}^k) &\leq \\ -\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \end{aligned}$$

В последнем равенстве использовалось свойство проекции $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in Q: (\mathbf{x} - P_Q(\mathbf{x}))^T (\mathbf{y} - P_Q(\mathbf{x})) \leq 0$. Таким образом, $f(\mathbf{x}^k)$ убывает. Так как f ограничена снизу, то существует предел $f(\mathbf{x}^k)$, а значит, $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \rightarrow 0$. Пусть \mathbf{x}^* — любая точка минимума f на Q , тогда, повторив аналогичное рассуждение для \mathbf{x}^* и $\mathbf{x}' = P_Q(\mathbf{x} - \frac{1}{2} \nabla f(\mathbf{x}^*))$, получаем

$$0 \leq f(\mathbf{x}') - f(\mathbf{x}^*) \leq -\|\mathbf{x}^* - \mathbf{x}'\|^2 \Rightarrow \mathbf{x}^* = \mathbf{x}'.$$

Далее, рассмотрим последовательности $\mathbf{u}^k \in \Omega$, $\mathbf{v}^k \in \Omega^\perp: \mathbf{x}^k = \mathbf{u}^k + \mathbf{v}^k$ (эти последовательности задаются однозначно по \mathbf{x}^k). Отметим, что $f(\mathbf{x}^k) = f(\mathbf{v}^k)$:

$$f(\mathbf{x}^k) - f(\mathbf{v}^k) = \frac{1}{2} \underbrace{(\mathbf{x}^k - \mathbf{v}^k)^T A (\mathbf{x}^k + \mathbf{v}^k)}_{=\mathbf{u}^k A = \mathbf{0}_m} + \underbrace{\mathbf{w}^T (\mathbf{x}^k - \mathbf{v}^k)}_{=0 \text{ по лемме 1}} = 0.$$

Пусть $\mathbf{u}^* \in \Omega, \mathbf{v}^* \in \Omega^\perp, \mathbf{u}^* + \mathbf{v}^* = \mathbf{x}^*$, тогда

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) = f(\mathbf{v}^k) - f(\mathbf{v}^*) = \frac{1}{2}(\mathbf{v}^k - \mathbf{v}^*)^T A(\mathbf{v}^k + \mathbf{v}^*) + \mathbf{w}^T(\mathbf{v}^k - \mathbf{v}^*) \leq \left\| \frac{1}{2}A(\mathbf{v}^k + \mathbf{v}^*) + \mathbf{w} \right\| \|\mathbf{v}^k - \mathbf{v}^*\|.$$

В последнем выражении левый множитель ограничен на Q , в силу выбора γ , выполняется неравенство

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq \gamma \|\mathbf{v}^k - \mathbf{v}^*\|.$$

Далее, применяя лемму 3 к точкам $\mathbf{x}^k - \frac{1}{2}\nabla f(\mathbf{x}^k)$ и $\mathbf{x}^* - \frac{1}{2}\nabla f(\mathbf{x}^*)$ и учитывая $\mathbf{u}^k, \mathbf{u}^* \in \Omega, \mathbf{v}^k, \mathbf{v}^* \in \Omega^\perp$ получаем

$$\|\mathbf{v}^{k+1} - \mathbf{v}^*\| \leq \|(I - \frac{1}{2}A)(\mathbf{v}^k - \mathbf{v}^*)\| \leq \left(1 - \frac{\lambda_2(A)}{2}\right) \|\mathbf{v}^k - \mathbf{v}^*\|.$$

Таким образом,

$$\delta(\mathbf{x}^k - \mathbf{x}^*, \Omega) \leq q^k \|\mathbf{x}^0 - \mathbf{x}^*\|$$

и

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq \gamma q^k \|\mathbf{x}^0 - \mathbf{x}^*\|.$$

Теперь заметим следующее: раз $f(\mathbf{x}^k) \rightarrow f(\mathbf{x}^*)$, то $\sum_{i=0}^{\infty} \|\mathbf{x}^{i+1} - \mathbf{x}^i\|^2 \leq \sum_{i=0}^{\infty} f(\mathbf{x}^i) - f(\mathbf{x}^{i+1}) = f(\mathbf{x}^0) - f(\mathbf{x}^*)$. Для сходимости \mathbf{x}^k достаточно сходимости ряда $\sum_{i=0}^{\infty} \|\mathbf{x}^{i+1} - \mathbf{x}^i\|$, что в общем случае не следует из сходимости $\sum_{i=0}^{\infty} \|\mathbf{x}^{i+1} - \mathbf{x}^i\|^2$. Однако в нашем случае сходимост экспоненциальная, что позволяет вывести сходимост \mathbf{x}^k :

$$\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \leq f(\mathbf{x}^k) - f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k) - f(\mathbf{x}^*),$$

следовательно,

$$\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \leq q^{k/2} \sqrt{\gamma \|\mathbf{x}^0 - \mathbf{x}^*\|},$$

а значит, \mathbf{x}^k сходится. Не умаляя общности, можно считать, что $\mathbf{x}^k \rightarrow \mathbf{x}^*$. Наконец, выведем скорость сходимости для \mathbf{x}^k :

$$\|\mathbf{x}^k - \mathbf{x}^*\| \leq \sum_{i=k}^{\infty} \|\mathbf{x}^i - \mathbf{x}^{i+1}\| \leq \left(\sum_{i=0}^{\infty} q^{i/2} \right) q^{k/2} \sqrt{\gamma \|\mathbf{x}^0 - \mathbf{x}^*\|} = \frac{q^{k/2}}{1 - \sqrt{q}} \sqrt{\gamma \|\mathbf{x}^0 - \mathbf{x}^*\|}.$$

□

Замечания. Для процесса $\mathbf{x}^{k+1} = \mathbf{x}^k - \frac{1}{2}(A\mathbf{x}^k + \mathbf{w})$ легко показать, что $\mathbf{u}^* = \mathbf{u}^0 = \mathbf{u}^1 = \dots$, что облегчает доказательство, однако при наличии проекции это свойство не выполняется, для чего и понадобилась лемма 3. Величину $\lambda_2(A)$ принято называть *числом Фидлера* или алгебраической связностью, оно всегда фигурирует в качестве параметра скорости сходимости лапласиановых систем. Пункт 3 дает оценку расстояния до ближайшей точки минимума, хотя нет гарантии, что последовательность сходится именно к ней.

Так как удалось установить экспоненциальную сходимост для $\|\mathbf{x}^k - \mathbf{x}^*\|$ и при этом размер шага постоянен, то без особого труда можно доказать, что в случае если на каждой итерации пропускные способности немного изменяются, то $\|\mathbf{x}^k - \mathbf{x}^*\|$ будет экспоненциально сходиться к некоторому шару с центром в \mathbf{x}^* . Пусть \mathbf{c}^k — вектор пропускных

способностей на итерации k , $Q_k = \{\mathbf{x} \in \mathbb{R}^m \mid 0 \leq x_e \leq c_e^k\}$, $f_k(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A \mathbf{x} + \mathbf{x}^T \mathbf{w}^k$, $\mathbf{w}^k = B'^T \mathbf{b}^k$, $\mathbf{b}^k = B'' \mathbf{c}^k$. В этих обозначениях мы полагаем, что в случае изменения пропускных способностей последовательность оценок генерируется по правилу:

$$\mathbf{x}^{k+1} = P_{Q_k} \left(\mathbf{x} - \frac{1}{2} \nabla f_k(\mathbf{x}^k) \right), \quad (3)$$

Обозначим также $\mathcal{Q} = \cup_{k=1}^{\infty} Q_k$.

Теорема 2. Пусть \mathbf{x}^k — последовательность, генерируемая по правилу (3) с любым начальным приближением $\mathbf{x}^0 \in Q_0$, при этом $\|\mathbf{c}^k\| \leq \sigma$, $\|\mathbf{c}^k - \mathbf{c}^{k-1}\| \leq \beta$, $\gamma = \sup_{\mathbf{x} \in \mathcal{Q}, k} \|A\mathbf{x} + \mathbf{w}^k\|$, $q = 1 - \lambda_2(A)/2$, $\nu = \beta(1 + \frac{\|B'^T B''\|}{d\lambda_2(A)})$, тогда

1. Ближайшая к \mathbf{x}^k точка минимума \mathbf{x}^{*k} функции f_k находится на расстоянии не более

$$q^k \left(\|\mathbf{x}^0 - \mathbf{x}^{*0}\| - \frac{\nu}{1-q} \right) + \frac{\nu}{1-q}.$$

2. Близость значения функции f_k к оптимальному значению оценивается следующим неравенством

$$f_k(\mathbf{x}^k) - f_k(\mathbf{x}^{*k}) \leq \gamma \left(q^k \left(\|\mathbf{x}^0 - \mathbf{x}^{*0}\| - \frac{\nu}{1-q} \right) + \frac{\nu}{1-q} \right).$$

Доказательство. Доказательство практически повторяет теорему 1, если обозначить \mathbf{x}^{*k} , \mathbf{u}^{*k} , \mathbf{v}^{*k} — точка минимума f_k и её разложение по Ω, Ω^\perp соответственно, то все выкладки вплоть до

$$f_k(\mathbf{x}^k) - f_k(\mathbf{x}^{*k}) \leq \gamma \|\mathbf{v}^k - \mathbf{v}^{*k}\|$$

идентичны. Далее, из-за изменения пропускных способностей точка минимума сдвигается не более, чем на β из-за изменения множества Q_k , и не более чем на $\|B'^T B''\| \beta / \lambda_2(B'^T B')$ из-за изменения функции f_k . Таким образом,

$$\|\mathbf{v}^{k+1} - \mathbf{v}^{*k+1}\| \leq \|(I - \frac{1}{2}A)(\mathbf{v}^k - \mathbf{v}^{*k})\| + \nu \leq$$

$$q \|\mathbf{v}^k - \mathbf{v}^{*k}\| + \nu.$$

Вычтем из обеих частей $\frac{\nu}{(1-q)}$:

$$\|\mathbf{v}^{k+1} - \mathbf{v}^{*k+1}\| - \frac{\nu}{(1-q)} \leq q \|\mathbf{v}^k - \mathbf{v}^{*k}\| + \nu - \frac{\nu}{(1-q)} = q \left(\|\mathbf{v}^k - \mathbf{v}^{*k}\| - \frac{\nu}{(1-q)} \right).$$

Таким образом,

$$\|\mathbf{v}^k - \mathbf{v}^{*k}\| \leq q^k \left(\|\mathbf{v}^0 - \mathbf{v}^{*0}\| - \frac{\nu}{1-q} \right) + \frac{\nu}{1-q},$$

$$\delta \left(\mathbf{v}^k, B \left(\mathbf{v}^{*k}, \frac{\nu}{1-q} \right) \right) \leq q^k \left(\|\mathbf{x}^0 - \mathbf{x}^{*0}\| - \frac{\nu}{1-q} \right).$$

Так как множество $\Omega_k = \{\mathbf{x}^{*k} + \mathbf{y} \mid \mathbf{y} \in \Omega\}$ состоит только из точек минимума f_k , то расстояние до ближайшей точки минимума и есть $\|\mathbf{v}^k - \mathbf{v}^{*k}\|$. □

Замечание. Константа σ нигде в явной форме не использовалась в теореме, однако от нее зависит размер множества \mathcal{Q} , а значит, и константа γ .

4.2. Сходимость рандомизированного алгоритма

Как уже было сказано, рандомизированная версия алгоритма представляет специальный вид рандомизированной стохастической аппроксимации. Обозначим $f_k(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T B'^T B' \mathbf{x} + \mathbf{x}^T B'^T \mathbf{b}^k$, Δ_k — рандомизированное пробное возмущение на итерации k , определяемое по правилу

$$\Delta_k = \pm e_m^i \text{ с. в. } \frac{P_i}{2}, \quad (4)$$

где e_m^i — m -мерный i -ый орт. Далее, $A^k = \Delta_k \Delta_k^T B'^T B'$, $A = EA^i$, $\mathbf{w}^{ik} = e_m^i e_m^{iT} B'^T \mathbf{b}^k$, $\mathbf{w}^k = E\mathbf{w}^{ik}$, $y_k^+ = f(\mathbf{x}^k + \alpha\Delta_k) + \xi_{2k}$ и $y_k^- = f(\mathbf{x}^k - \alpha\Delta_k) + \xi_{2k+1}$ — зашумленные измерения на итерации k , при этом ξ_k — погрешность, возникающая при измерении \mathbf{b}^k (который, в свою очередь, зависит от пропускных способностей), ξ_k — случайная величина, распределение которой неизвестно, $Q_k = \{\mathbf{x} \in \mathbf{R}^m \mid 0 \leq x_e \leq c_e^k\}$. В этих обозначениях *RandomizedArcBalancing* генерирует последовательность оценок по правилу

$$\mathbf{x}^{k+1} = P_{Q_k} \left(\mathbf{x}^k - \Delta_k \frac{y_k^+ - y_k^-}{4\alpha} \right). \quad (5)$$

Теорема 3. Пусть \mathbf{x}^k — последовательность, генерируемая по правилу (5) с любым начальным приближением $\mathbf{x}^0 \in Q_0$, $E\xi_k^2 \leq \sigma^2$, множество $\cup_k Q_k$ ограничено и $\gamma = \sup_{\mathbf{x} \in \mathcal{Q}, k} \|A\mathbf{x} + \mathbf{w}^k\|, \|\mathbf{c}^k - \mathbf{c}^{k+1}\| \leq \beta$, $\lambda_2(A)$ — минимальное отличное от нуля собственное число матрицы A , $q = \left(1 - \frac{\lambda_2(A)}{2}\right)$, $\alpha > 0$, $\nu = \beta \left(1 + \frac{\|B'^T B''\|}{\lambda_2(B'^T B')}, $R = \nu + \frac{\sigma}{2\sqrt{\alpha}}$, тогда$

1. Если \mathbf{x}^{*k} — любая точка минимума f_k на Q_k , то

$$E\delta(\mathbf{x}^k - \mathbf{x}^{*k}, \Omega) \leq q^{k/2} (\|\mathbf{x}^0 - \mathbf{x}^{*0}\| - R) + R,$$

где δ — евклидова метрика.

2. Если \mathbf{x}^{*k} — любая точка минимума f_k на Q_k , то

$$E\{f(\mathbf{x}^k) - f(\mathbf{x}^{*k})\} \leq \gamma(q^{k/2} (\|\mathbf{x}^0 - \mathbf{x}^{*0}\| - R) + R).$$

3. Если помехи в измерениях отсутствуют, а пропускные способности не изменяются (то есть $R = 0$, $Q_k = Q_{k+1}$, $f_k \equiv f_{k+1}$) и \mathbf{x}^k сходится к \mathbf{x}^* , то

$$E\|\mathbf{x}^k - \mathbf{x}^*\| \leq \frac{q^{k/4}}{1 - \sqrt[4]{q}} \sqrt{\gamma\|\mathbf{x}^0 - \mathbf{x}^{*0}\|}.$$

Доказательство. Во-первых, отметим, что для любого возможного Δ_k имеет место $\|A^k\| \leq 2$. Непосредственным вычислением получаем, что при выборе дуги $l = (i, j)$ выполняется

$$e_m^l e_m^{lT} B'^T B' \mathbf{x} = \sum_{(h,j) \in E'} x_h - \sum_{(j,h) \in E'} x_h - \sum_{(h,i) \in E'} x_h + \sum_{(i,h) \in E'} x_h.$$

Каждая компонента вектора \mathbf{x} может входить в любую из этих сумм, а значит любая компонента входит с коэффициентом $-2, -1, 0, 1$ или 2 , из чего и следует $\|A^k\| \leq 2$. Из выбора γ следует

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq \gamma\|\mathbf{v}^k - \mathbf{v}^*\|,$$

где \mathbf{v} и \mathbf{u} — разложения \mathbf{x} по Ω^\perp , Ω . Из леммы 1 следует, что ядро A и $B'^T B'$ совпадают. Так как $0 \leq A^k \leq 2I$, то

$$E\left\{\left(I - \frac{A^k}{2}\right)^T \left(I - \frac{A^k}{2}\right)\right\} \leq E\left\{I - \frac{A^k}{2}\right\} = I - \frac{A}{2}.$$

Из определения Δ_k следует, что $E\Delta_k = 0_m$, $\Delta_k \Delta_k^T \Delta_k = \Delta_k$, а значит, $E\Delta_k \Delta_k^T \Delta_k = 0_m$ и, следовательно, $E\Delta_k A^k = 0$. Таким образом, применяя лемму 3 для точек $\mathbf{x}^k - \Delta^k(y_k^+ - y_k^-)/4\alpha$ и \mathbf{x}^{*k} , получаем

$$\begin{aligned} E\{\|\mathbf{v}^{k+1} - \mathbf{v}^{*k}\|^2 \mid \mathbf{x}^k\} &\leq E\left\{\left\|\left(I - \frac{A^k}{2}\right)(\mathbf{v}^k - \mathbf{v}^*) + \Delta_k \frac{\xi_{2k+1} - \xi_{2k}}{4\alpha}\right\|^2\right\} = \\ &(\mathbf{v}^k - \mathbf{v}^*)^T E\left\{\left(I - \frac{A^k}{2}\right)^T \left(I - \frac{A^k}{2}\right)\right\}(\mathbf{v}^k - \mathbf{v}^*) + \\ &\frac{1}{2\alpha} E\{(\xi_{2k+1} - \xi_{2k}) \Delta_k^T \left(I - \frac{A^k}{2}\right)\}(\mathbf{v}^k - \mathbf{v}^*) + \frac{1}{4\alpha} E\{\Delta_k^T \Delta_k (\xi_{2k+1} - \xi_{2k})\} \leq \\ &\left(1 - \frac{\lambda_2(A)}{2}\right) \|\mathbf{v}^k - \mathbf{v}^*\|^2 + \frac{\sigma^2}{4\alpha}, \end{aligned}$$

а значит,

$$\begin{aligned} E\{\|\mathbf{v}^{k+1} - \mathbf{v}^{*k+1}\| \mid \mathbf{x}^k\} &\leq E\{\|\mathbf{v}^{k+1} - \mathbf{v}^{*k}\| \mid \mathbf{x}^k\} + \nu \leq \\ (E\{\|\mathbf{v}^{k+1} - \mathbf{v}^{*k}\|^2 \mid \mathbf{x}^k\})^{1/2} + \nu &\leq \left(1 - \frac{\lambda_2(A)}{2}\right)^{\frac{1}{2}} \|\mathbf{v}^k - \mathbf{v}^*\| + \nu + \frac{\sigma}{2\sqrt{\alpha}}. \end{aligned}$$

Отсюда выводим

$$E\{\|\mathbf{v}^k - \mathbf{v}^{*k}\|\} - R \leq q^{k/2} (\|\mathbf{x}^0 - \mathbf{x}^{*0}\| - R).$$

По выбору γ

$$E f_k(\mathbf{x}^k) - f_k(\mathbf{x}^{*k}) \leq \gamma E \|\mathbf{v}^k - \mathbf{v}^{*k}\| \leq \gamma (q^{k/2} (\|\mathbf{x}^0 - \mathbf{x}^{*0}\| - R) + R).$$

Если же $\sigma = 0$, $f_k \equiv f_{k+1}$, $Q_k = Q_{k+1}$, то $R = 0$, множество точек минимума не изменяется, и расстояние до ближайшей точки минимума в среднем сходится с экспоненциальной скоростью.

Так как Q_k — прямоугольный параллелепипед, а на одном шаге изменения выполняются только по одной координате, то существует такой набор скаляров $0 \leq \varphi_k \leq 1$, что

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \varphi_k \left(\mathbf{x}^k - \frac{1}{2}(A^k \mathbf{x}^k + \mathbf{w}^k)\right) - \mathbf{x}^k = \mathbf{x}^k - \varphi_k \frac{1}{2}(A^k \mathbf{x}^k + \mathbf{w}^k). \quad (6)$$

Из выпуклости f_k следует

$$\begin{aligned} f_k(\mathbf{x}^{k+1}) - f_k(\mathbf{x}^k) &\leq \\ \left(f_k(\mathbf{x}^k) + \varphi_k \left(f_k\left(\mathbf{x}^k - \frac{1}{2}(A^k \mathbf{x}^k + \mathbf{w}^k)\right) - f_k(\mathbf{x}^k)\right)\right) - f_k(\mathbf{x}^k) &= \\ \varphi_k \left(f_k\left(\mathbf{x}^k - \frac{1}{2}(A^k \mathbf{x}^k + \mathbf{w}^k)\right) - f_k(\mathbf{x}^k)\right). & \quad (7) \end{aligned}$$

Далее, заметим, что $f_k(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T B'^T B' \mathbf{x} + B'^T \mathbf{b}^k = \frac{1}{2} (\|B' \mathbf{x} + \mathbf{b}^k\|^2 - \|\mathbf{b}^k\|^2)$. В случае отсутствия помех (5) делает оптимальный шаг в выбранном направлении: выбирается дуга $e = (i, j)$

и уравниваются значения $[B'\mathbf{x}^k + \mathbf{b}^k]_i$ и $[B'\mathbf{x}^k + \mathbf{b}^k]_j$, при этом прочие координаты вектора $B'\mathbf{x}^k + \mathbf{b}^k$ не изменяются. Используя этот факт, можно вычислить $f_k(\mathbf{x}^{k+1}) - f_k(\mathbf{x}^k)$. Для удобства обозначим $\mathbf{z} = B'\mathbf{x}^k + \mathbf{b}^k$, тогда

$$f_k\left(\mathbf{x}^k - \frac{1}{2}(A^k\mathbf{x}^k + \mathbf{w}^k)\right) - f_k(\mathbf{x}^k) = \left(\frac{z_i + z_j}{2}\right)^2 - \left(\frac{z_i^2 + z_j^2}{2}\right) = -\frac{1}{2}(z_i - z_j)^2.$$

С другой стороны,

$$\left\|\frac{1}{2}(A^k\mathbf{x}^k + \mathbf{w}^k)\right\|^2 = \frac{1}{2}|[B'\mathbf{x}^k + \mathbf{b}^k]_i - [B'\mathbf{x}^k + \mathbf{b}^k]_j|^2 = \frac{1}{2}(z_i - z_j)^2,$$

а значит,

$$f_k\left(\mathbf{x}^k - \frac{1}{2}(A^k\mathbf{x}^k + \mathbf{w}^k)\right) - f_k(\mathbf{x}^k) = -\left\|\frac{1}{2}(A^k\mathbf{x}^k + \mathbf{w}^k)\right\|^2.$$

Следовательно, в силу (6) и (7), получаем

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \leq f_k(\mathbf{x}^k) - f_k(\mathbf{x}^{k+1}).$$

Если $R = 0$ и \mathbf{x}^k сходится к \mathbf{x}^* , то

$$\begin{aligned} E\|\mathbf{x}^k - \mathbf{x}^*\| &\leq \sum_{i=k}^{\infty} E\|\mathbf{x}^i - \mathbf{x}^{i+1}\| \leq \sum_{i=k}^{\infty} E(f_k(\mathbf{x}^i) - f_k(\mathbf{x}^{*k}))^{1/2} \leq \\ &\sum_{i=k}^{\infty} q^{i/4} \sqrt{\gamma\|\mathbf{x}^0 - \mathbf{x}^{*0}\|} = \frac{q^{k/4}}{1 - \sqrt[4]{q}} \sqrt{\gamma\|\mathbf{x}^0 - \mathbf{x}^{*0}\|} \end{aligned}$$

□

Замечания. $A = DB'^T B'$, где $D = \text{diag}\{p_1, \dots, p_m\}$. Для того чтобы выполнялось $\|A\| \leq 2$, годится выбор $p_e = 1/(nd_i^-)$, $e = (i, j)$ или же просто $p_e = 1/|E'|$. По сравнению с синхронной версией алгоритма, сходимость немного более чем в два раза медленней, но при этом на одну итерацию гораздо меньше действий.

В пункте 1 оценивается расстояние до ближайшей точки минимума. В случае отсутствия помех и изменений функций f_k (пропускных способностей) $R = 0$, имеет место экспоненциальная сходимость к множеству точек минимума f_k . В пункте 3 можно вывести сходимость ряда $\sum_{i=0}^{\infty} E\|\mathbf{x}^i - \mathbf{x}^{i+1}\|$, а также сходимость ряда $\sum_{i=0}^{\infty} \|\mathbf{x}^i - \mathbf{x}^{i+1}\|^2$ для любой возможной последовательности \mathbf{x}^k , что, тем не менее, ничего не говорит о сходимости \mathbf{x}^k .

5. ИЗВЛЕЧЕНИЕ РЕШЕНИЯ ЗАДАЧИ О МАКСИМАЛЬНОМ ПОТОКЕ

Из теорем 1 и 3 следует, что оба рассматриваемых алгоритма решают следующую задачу оптимизации:

$$\begin{aligned} &\text{минимизировать} && \frac{1}{2}\mathbf{x}^T B'^T B' \mathbf{x} + \mathbf{x}^T B'^T \mathbf{b}, \\ &\text{при условии} && 0 \leq x_e \leq c_e, \quad e \in E'. \end{aligned} \quad (8)$$

При этом на исходном графе G мы получаем псевдопоток, удовлетворяющий пропускным способностям, в котором насыщены все дуги, инцидентные истоку и стоку. Обозначим $V_0(\mathbf{x}) = \{i \in V \setminus \{s, t\} \mid [B'^T \mathbf{x} + B'^T \mathbf{b}]_i = 0\}$, $V_-(\mathbf{x}) = \{i \in V \setminus \{s, t\} \mid [B'^T \mathbf{x} + B'^T \mathbf{b}]_i < 0\}$, $V_+(\mathbf{x}) = \{i \in V \setminus \{s, t\} \mid [B'^T \mathbf{x} + B'^T \mathbf{b}]_i > 0\}$. Ключевым моментом в сведении задачи о максимальном потоке к задаче (8) является следующая теорема:

Теорема 4. Если \mathbf{x}^* — решение задачи (8), то

1. Множества $A = \{s\} \cup V_+(\mathbf{x}^*) \cup V_0(\mathbf{x}^*)$ и $V \setminus A$ образуют минимальный st -разрез графа G .
2. Для получения максимального потока достаточно провести частичную декомпозицию потока \mathbf{x}^* .

Доказательство. Во-первых, применяя условия Каруша-Куна-Такера к задаче (8), получаем

$$\begin{cases} [B'\mathbf{x}^* + \mathbf{b}]_j - [B'\mathbf{x}^* + \mathbf{b}]_i + \mu_e - s_e = 0, \\ \mu, s \geq 0_m, \\ 0 = \mu_e(c_e - x_e) = s_e x_e, \end{cases}$$

где $e = (i, j)$, μ_e — множитель Лагранжа, соответствующий ограничению $x_e \leq c_e$, а s_e — множитель Лагранжа, соответствующий $0 \leq x_e$. В данном случае как s_e , так и μ_e играют вспомогательную роль, и от них можно избавиться следующим образом: пусть $E_x = \{e = (i, j) \in E' \mid c_e - x_e > 0\} \cup \{(j, i) \mid e = (i, j) \in E', x_e > 0\}$, тогда условия ККТ для (8) можно переписать в виде:

$$[B'\mathbf{x}^* + \mathbf{b}]_i \leq [B'\mathbf{x}^* + \mathbf{b}]_j, \quad (i, j) \in E_{x^*}. \quad (9)$$

Проще говоря, если избыток вершины i (величина $[B'\mathbf{x} + \mathbf{b}]_i$) больше избытка вершины j , то $(i, j) \notin E_{x^*}$. Теперь вернемся к исходному графу G . Избыток потока для вершин в графе G задается вектором $B \begin{bmatrix} \mathbf{x}^* \\ \mathbf{c}'' \end{bmatrix}$, по построению для вершин, отличных от s, t , он совпадает с $B'\mathbf{x} + \mathbf{b}$, для истока s — это величина $(-\sum_{(s,i) \in E} c_e) \leq 0$, для стока — величина $\sum_{(i,t) \in E} c_e \geq 0$. По теореме о декомпозиции потока существует конечный набор путей, каждый из которых ведет из вершины с отрицательным избытком в вершину с положительным избытком и проходит только по дугам с положительным потоком. В силу (9), для вершин с положительным избытком кроме t такой путь полностью проходит только по вершинам из A , а значит, начинается в s , для вершин с отрицательным избытком кроме s такой путь полностью проходит только по вершинам из $V \setminus A$, а значит, заканчивается в t . Пусть, \mathbf{y} — поток, получившийся в результате удаления потока по соответствующим путям из s в $V_+(\mathbf{x}^*)$ и путям из $V_-(\mathbf{x}^*)$ в t . По построению \mathbf{y} является допустимой точкой задачи о максимальном потоке (1), при этом не существует дуг $(i, j) \in E_y$, $i \in A, j \in V \setminus A$, а значит, из теоремы Форда-Фалкерсона \mathbf{y} — максимальный поток, а $A, V \setminus A$ — минимальный st -разрез. \square

Замечание. В этой теореме также можно было взять $A = \{s\} \cup V_+(\mathbf{x}^*)$. В работе [2] также отмечено, что, после того как в общей процедуре балансирования дуг ни одна из операций *Move* не дает изменения потока больше, чем на $1/n^2$, возможно точное извлечение максимального потока / минимального разреза, что позволяет использовать приближенное решение (8), а не точное.

6. РАСПРЕДЕЛЕННАЯ РЕАЛИЗАЦИЯ АЛГОРИТМОВ

Оба алгоритма *ConcurrentArcBalancing* и *RandomizedArcBalancing* допускают распределенную реализацию в сети, если топология этой сети задана тем же графом, что и задача о максимальном потоке. Также можно считать, что вычислительные узлы соответствуют не всем вершинам, а только $V \setminus \{s, t\}$. Для ускорения вычислений имеет смысл непосредственным образом вычислять избыток вершины. Для обоих алгоритмов мы будем предполагать следующее:

- Алгоритмы оперируют с величинами $x_e, e \in E', y_i, i \in V \setminus \{s, t\}$.
- Узел i хранит в локальной памяти величины x_e и c_e , где $e = (i, j) \in E'$.
- В синхронной версии используются дополнительные промежуточные переменные y_1, \dots, y_n , в рандомизированной версии каждый узел хранит дополнительно четыре переменных для промежуточных вычислений $z_i^+, z_j^+, z_i^-, z_j^-$.

Далее представлены распределенные версии алгоритмов *ConcurrentArcBalancing* и *RandomizedArcBalancing*.

Фактически, они ничем не отличаются от представленных ранее, но имеют более явную форму. В алгоритме *ConcurrentArcBalancing* предполагаем, что в сети есть синхронизация времени, все операции двух внутренних циклов *for* происходят одновременно по тактам системы. В этом случае, на каждой итерации достаточно хранить только значения $\mathbf{x}^k, \mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{y}^{k+1}$. В обоих алгоритмах предполагается, что узел i должен иметь возможность запросить информацию соседних узлов, то есть таких j , что $(i, j) \in E$, и при этом передать информацию на этот узел. Другими словами, условие двустороннего взаимодействия необходимо для реализации этого алгоритма (однако это еще не значит, что для соответствующей задачи о максимальном потоке пропускные способности должны быть двусторонними).

Algorithm 3: Distributed ConcurrentArcBalancing(G, \mathbf{c} , stopping criterion)

```

 $\mathbf{x}^0 \leftarrow 0_m, k \leftarrow 0;$ 
for  $e = (s, i) \in E$  do
   $x_e^0 \leftarrow c_e;$ 
for  $e = (i, t) \in E$  do
   $x_e^0 \leftarrow c_e;$ 
for  $i \in V \setminus \{s, t\}$  do
   $y_i^0 \leftarrow \sum_{(j,i) \in E} x_e^0 - \sum_{(i,j) \in E} x_e^0;$ 
/* End of initialization */
while stopping criterion is not satisfied do
  for  $e = (i, j) \in E'$  do
     $x_e^{k+1} \leftarrow x_e^k - \frac{y_j^k - y_i^k}{2};$ 
    if  $x_e^{k+1} > c_e$  then
       $x_e^{k+1} \leftarrow c_e.$ 
    if  $x_e < 0$  then
       $x_e^{k+1} \leftarrow 0.$ 
  for  $i \in V \setminus \{s, t\}$  do
     $y_i^{k+1} \leftarrow \sum_{(j,i) \in E} x_e^k - \sum_{(i,j) \in E} x_e^k;$ 
   $k \leftarrow k + 1;$ 
return  $\mathbf{x};$ 

```

В рандомизированном алгоритме предполагается, что в системе нет синхронизации, каждый из узлов совершает *Move* относительно случайной исходящей дуги. Таким образом, можно считать, что сначала случайным образом выбирается узел, который совершает действие, после чего случайно выбирается исходящая дуга, что дает вероятность выбора дуги $e = (i, j)$: $p_e = \frac{1}{nd_i}$. Из леммы 2 следует, что такой выбор вероятности га-

рантирует выполнение неравенства $\|A\| \leq 2$ в соответствующих теоремах сходимости. Результатом обоих алгоритмов является решение (8), извлечение решения задачи о максимальном потоке было обсуждено в предыдущем разделе, оно заключается в частичной декомпозиции потока и представляет собой гораздо менее сложную задачу, нежели сама задача о максимальном потоке. Извлечение минимального разреза возможно в распределенном виде: если $y_i > 0$, то узел i находится в одной части разреза, если же $y_i \leq 0$, то в другой.

Algorithm 4: Distributed RandomizedArcBalancing(G, c , stopping criterion)

```

x ← 0m; while stopping criterion is not satisfied do
  pick random node  $i$  uniformly;
  pick random outgoing arc  $e = (i, j) \in E'$  of  $i$  uniformly;
  let  $\Delta \leftarrow \pm 1$  w. p. 1/2;
   $z_i^+ \leftarrow (\sum_{(k,i) \in E'} x_e - \sum_{(i,k) \in E'} x_e + \sum_{(s,i) \in E} c_e - \sum_{(i,t) \in E'} x_e + \alpha \Delta)$ ;
   $z_j^+ \leftarrow (\sum_{(k,j) \in E'} x_e - \sum_{(j,k) \in E'} x_e + \sum_{(s,j) \in E} c_e - \sum_{(j,t) \in E'} x_e + \alpha \Delta)$ ;
   $z_i^- \leftarrow (\sum_{(k,i) \in E'} x_e - \sum_{(i,k) \in E'} x_e + \sum_{(s,i) \in E} c_e - \sum_{(i,t) \in E'} x_e - \alpha \Delta)$ ;
   $z_j^- \leftarrow (\sum_{(k,j) \in E'} x_e - \sum_{(j,k) \in E'} x_e + \sum_{(s,j) \in E} c_e - \sum_{(j,t) \in E'} x_e - \alpha \Delta)$ ;
   $x_e \leftarrow x_e - \Delta \frac{(z_j^+)^2 + (z_i^+)^2 - (z_i^-)^2 - (z_j^-)^2}{8\alpha}$ ;
  if  $x_e > c_e$  then
     $x_e \leftarrow c_e$ .
  if  $x_e < 0$  then
     $x_e \leftarrow 0$ .
  return x;

```

7. ЗАКЛЮЧЕНИЕ

В этой работе были предложены и проанализированы два распределенных алгоритма решения задачи о максимальном потоке. Было доказано свойство адаптивности для разработанных алгоритмов, то есть при ограниченном изменении пропускных способностей на каждой итерации алгоритмы сохраняют экспоненциальную сходимость, но при этом сходятся только в некоторую окрестность точки минимума и остаются в ней. Для рандомизированной версии алгоритма была показана устойчивость к помехам в измерениях пропускных способностей.

Список литературы

1. Goldberg A., Tarjan R. Efficient maximum flow algorithms // Communications of the ACM, 2014. Vol. 57, №. 8. P. 82–89.
2. Tarjan R., Ward J., Zhang B., Zhou Y., Mao J. Balancing applied to maximum network flow problems // European Symposium on Algorithms, 2006. Springer, P. 612–623.
3. Левитин Е.С., Поляк Б.Т. Методы минимизации при наличии ограничений // Журнал вычислительной математики и математической физики, 1966. Т. 6, №. 5. С. 787–823.
4. Граничин О.Н. Поисковые алгоритмы стохастической аппроксимации с рандомизацией на входе // Автоматика и телемеханика, 2015. №. 5. С. 43–59.
5. Granichin O., Volkovich V., Toledano-Kitai D. Randomized algorithms in automatic control and data mining. Springer, 2014.

6. Boyd S., Ghosh A., Prabhakar B., Shah D. Randomized gossip algorithms // Information Theory, IEEE Transactions on, 2006. Vol. 52, №. 6. P. 2508–2530.
7. Olfati-Saber R., Murray R.M. Consensus problems in networks of agents with switching topology and time-delays // Automatic Control, IEEE Transactions on, 2004. Vol. 49, №. 9. P. 1520–1533.

Поступила в редакцию 10.08.2016, окончательный вариант — 11.09.2016.

Computer tools in education, 2016

№ 5: 46–61

<http://ipo.spb.ru/journal>

RANDOMIZED DISTRIBUTED ADAPTIVE ALGORITHM FOR A MAXIMUM FLOW PROBLEM

Мальковский Н.В.¹

¹SPbSU, Saint-Petersburg, Russia

Abstract

This paper presents an algorithm for solving one of the fundamental optimization problems — maximum flow problem. The algorithm is based on the ideas of arc balancing procedure and projected gradient method. Although the algorithm does not improve asymptotic complexity over existing methods it still possess some usefull properties: natural implementation in distributed systems and convergence to an optimal solution even if network parameters are changing slightly over time. Two versions of the algorithm are considered: randomized and concurrent. Both versions mainain adaptability but randomized version is preferable due to better convergence rate and simpler implementation in distributed systems.

Keywords: *constraint optimization, gradient descent, distributed algorithms, randomized algorithms, network flow.*

Citation: Malkovskiy, N., 2016. "Randomizirovannyi raspredelennyi adaptivnyi algoritm resheniya zadachi o maksimal'nom potoke" ["Randomized Distributed Adaptive Algorithm for a Maximum Flow Problem"]. *Computer tools in education*, no. 5, pp. 46–61.

Received 10.08.2016, the final version — 11.09.2016.

Мальковский Николай Владимирович, магистрант математико-механического факультета СПбГУ; 198504 Санкт-Петербург, Старый Петергоф, Университетский пр. 28, математико-механический факультет, malkovskynv@gmail.com.

Мальковский Николай Владимирович, магистрант математико-механического факультета СПбГУ; 198504 Санкт-Петербург, Старый Петергоф, Университетский пр. 28, математико-механический факультет, malkovskynv@gmail.com.

©

Наши авторы, 2016.

Our authors, 2016.