1 Постановка задачи

Рассмотрим линейное векторное пространство $E = \mathbb{R}^n$, с евклидовой нормой $\|\cdot\| = \|\cdot\|_2$.

Рассмотрим задачу выпуклой оптимизации:

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

где $f(x)$ выпуклая функция с липшицевой производной порядка p, т.е.

$$\|D^p f(x) - D^p f(y)\| \le L_p \|x - y\|. \tag{2}$$

Полином Тейлора функции $f(x)$ тогда запишется в виде

$$\Omega_p(f,x;y) = f(x) + \sum_{k=1}^{p} \frac{1}{k!} D^k f(x) \left[y - x\right]^k,\ y \in \mathbb{R}^n \tag{3}$$

Используя (2) и операцию интегрирования, получаем следующие неравенства

$$|f(y) - \Omega_p(f,x;y)| \le \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \tag{4}$$

$$\|\nabla f(y) - \nabla \Omega_p(f,x;y)\| \le \frac{L_p}{p!} \|y - x\|^{p}, \tag{5}$$

2 Неточный почти оптимальный ускоренный тензорный метод

Задача (1) может быть решена с помощью тензорных методов или их ускоренных версий. У этих методов присутствует следующий базовый шаг:

$$T_{H_p}(x) = \operatorname*{argmin}_y \left\{ \tilde{\Omega}_{p,H_p}(f,x;y) \right\},$$

где

$$\tilde{\Omega}_{p,H_p}(f,x;y) = \Omega_p(f,x;y) + \frac{H_p}{p!} \|y - x\|^{p+1} \tag{6}$$

Для $H_p \ge L_p$ эта подзадача выпуклая, а значит, осуществимая.

Рассмотрим случаи, когда невозможно решить именно эту подзадачу. Существуют методы Inexact pth-Order Basic Tensor Method (BTMI$_p$) и Inexact pth-Order Accelerated Tensor Method (ATMI$_p$). Их соотвествующие скорости сходимости - $O(k^{-p})$ и $O(k^{-(p+1)})$. В этом разделе представлен Inexact pth-Order Near-optimal Accelerated Tensor Method (NATMI$_p$) с улучшенной скоростью сходимости $\tilde{O}(k^{-\frac{3p+1}{2}})$, где $\tilde{O}(\cdot)$ означает с точностью до логарифмического множителя. Это улучшение алгоритма Accelerated Taylor Descent и обобщение Inexact Accelerated Taylor Descent.

Решение неточной подзадачи определяется следующим образом. Любая точка из множества

$$\mathcal{N}^{\gamma}_{p,H_p}(x) = \left\{ T \in \mathbb{R}^n \,:\, \|\nabla \tilde{\Omega}_{p,H_p}(f,x;T)\| \le \gamma \|\nabla f(T)\| \right\} \tag{7}$$

является решением неточной подзадачи, где $\gamma \in [0;1]$ является параметром точности. N^0_{p,H_p} является точным решением подзадачи.

Далее приведен Алогоритм 1.

Algorithm 1 Inexact pth-Order Near-optimal Accelerated Tensor Method (NATMI)

1: **Input:** convex function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\nabla^p f$ is L_p-Lipschitz, $H_p = \xi L_p$ where ξ is a scaling parameter, γ is a desired accuracy of the subproblem solution.

2: Set $A_0 = 0, x_0 = y_0$

3: **for** $k = 0$ **to** $k = K - 1$ **do**

4: Compute a pair $\lambda_{k+1} > 0$ and $y_{k+1} \in \mathbb{R}^n$ such that

$$\frac{1}{2} \le \lambda_{k+1} \frac{H_p \cdot \|y_{k+1} - \tilde{x}_k\|^{p-1}}{(p-1)!} \le \frac{p}{p+1}\,,$$

where

$$y_{k+1} \in \mathcal{N}^{\gamma}_{p,H_p}(\tilde{x}_k) \tag{8}$$

and

$$a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1}A_k}}{2}\,,\ A_{k+1} = A_k + a_{k+1}\,,\ \text{and}\ \tilde{x}_k = \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x_k\,.$$

5: Update $x_{k+1} := x_k - a_{k+1} \nabla f(y_{k+1})$

6: **return** y_K

Чтобы получить скорость сходимости Алгоритма 1, используются дополнительные леммы. Первая лемма говорит о промежуточном неравенстве и объединяет теорию неравенства и теорию метода.

Лемма 1. *Если* $y_{k+1} \in \mathcal{N}^{\gamma}_{p,H_p}(\tilde{x}_k)$, *то*

$$\|\nabla \tilde{\Omega}_{p,H_p}(f,\tilde{x}_k;y_{k+1})\| \le \frac{\gamma}{1-\gamma} \cdot \frac{(p+1)H_p + L_p}{p!} \|y_{k+1} - \tilde{x}_k\|^p. \tag{9}$$

Следующая лемма играет ключевую роль в доказательстве сходимости Алгоритма 1.

Лемма 2. *Если* $y_{k+1} \in \mathcal{N}^{\gamma}_{p,H_p}(\tilde{x}_k)$, $H_p = \xi L_p$ *такие, что* $1 \ge 2\gamma + \frac{1}{\xi(p+1)}$ *и*

$$\frac{1}{2} \le \lambda_{k+1} \frac{H_p \cdot \|y_{k+1} - \tilde{x}_k\|^{p-1}}{(p-1)!} \le \frac{p}{p+1}\,, \tag{10}$$

то

$$\|y_{k+1} - (\tilde{x}_k - \lambda_{k+1}\nabla f(y_{k+1}))\| \leq \sigma \cdot \|y_{k+1} - \tilde{x}_k\| \tag{11}$$

и

$$\sigma \geq \frac{p\xi + 1 - \xi + 2\gamma\xi}{(1-\gamma)2p\xi}, \tag{12}$$

где $\sigma \leq 1$.

В результате получается следующая теорема.

Теорема 1. *Пусть* f *- выпуклая функция, производная* p^{th} *которой* L_p*-липшицева и* x_* *означает точку минимума* f*. Тогда Алгоритм 1 сходится со скоростью*

$$f(y_k) - f(x_*) \leq \tilde{O}\left(\frac{H_p R^{p+1}}{k^{\frac{3p+1}{2}}}\right), \tag{13}$$

где

$$R = \|x_0 - x^*\| \tag{14}$$

максимальный радиус исходного множества.

3 Супербыстрый метод второго порядка

Фиксируем параметры Алгоритма 1

$$p = 3, \quad \gamma = \frac{1}{2p} = \frac{1}{6}, \quad \xi = \frac{2p}{p+1} = \frac{3}{2}. \tag{15}$$

Из (12) получаем $\sigma = 0.6$, что близко к точному значению $\sigma_0 = 0.5$. Скорость сходимости с данными параметрами до достижения точности ε:

$$N_{out} = \tilde{O}\left(\left(\frac{L_3 R^4}{\varepsilon}\right)^{\frac{1}{5}}\right). \tag{16}$$

Очевидно, что на каждом шаге Алгоритма 1, требуется решить следующую задачу с точностью $\gamma = 1/6$

$$\underset{y}{\operatorname{argmin}}\left\{\langle\nabla f(x_i), y - x_i\rangle + \frac{1}{2}\nabla^2 f(x_i)[y-x_i]^2 + \frac{1}{6}D^3 f(x_i)[y-x_i]^3 + \frac{L_3}{4}\|y-x_i\|^4\right\}. \tag{17}$$

BGDM используется для решения подзадачи Алгоритма 1. В результате, это дает Hyperfast Second-Order method, как совмещение NATMI и BGDM.

Algorithm 2 Hyperfast Second-Order Method

1: **Input:** convex function $f : \mathbb{R}^n \to \mathbb{R}$ with L_3-Lipschitz 3rd-order derivative.
2: Set $A_0 = 0, x_0 = y_0$
3: **for** $k = 0$ **to** $k = K - 1$ **do**
4: Compute a pair $\lambda_{k+1} > 0$ and $y_{k+1} \in \mathbb{R}^n$ such that

$$\frac{1}{2} \leq \lambda_{k+1} \frac{3L_p \cdot \|y_{k+1} - \tilde{x}_k\|^2}{4} \leq \frac{3}{4}\,,$$

where $y_{k+1} \in \mathcal{N}^{1/6}_{3,3L_p/2}(\tilde{x}_k)$ solved by Algorithm 3 and

$$a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1}A_k}}{2}\ ,\ A_{k+1} = A_k + a_{k+1}\ ,\ \text{and } \tilde{x}_k = \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x_k\,.$$

5: Update $x_{k+1} := x_k - a_{k+1} \nabla f(y_{k+1})$
6: **return** y_K

Algorithm 3 Bregman-Distance Gradient Method

1: Set $z_0 = \tilde{x}_k$ and $\tau = \frac{3\delta}{8(2+\sqrt{2})\|\nabla f(\tilde{x}_k)\|}$
2: Set objective function

$$\varphi_k(z) = \langle \nabla f(\tilde{x}_k), z - \tilde{x}_k \rangle + \nabla^2 f(\tilde{x}_k)[z - \tilde{x}_k]^2 + D^3 f(\tilde{x}_k)[z - \tilde{x}_k]^3 + \frac{L_3}{4}\|z - \tilde{x}_k\|^4$$

3: Set feasible set

$$S_k = \left\{ z : \|z - \tilde{x}_k\| \leq 2 \left(\frac{2+\sqrt{2}}{L_3} \|\nabla f(\tilde{x}_k)\| \right)^{\frac{1}{3}} \right\} \tag{18}$$

4: Set scaling function

$$\rho_k(z) = \frac{1}{2} \left\langle \nabla^2 f(\tilde{x}_k)(z - \tilde{x}_k), z - \tilde{x}_k \right\rangle + L_3 d_4(z - \tilde{x}_k) \tag{19}$$

5: **for** $k \geq 0$ **do**
6: Compute the approximate gradient $g_{\varphi_k,\tau}(z_i)$ by (20).
7: **IF** $\|g_{\varphi_k,\tau}(z_i)\| \leq \frac{1}{6}\|\nabla f(z_i)\| - \delta$, then **STOP**
8: **ELSE** $z_{i+1} = \operatorname*{argmin}_z \left\{ \langle g_{\varphi_k,\tau}(z_i), z - z_i \rangle + 2\left(1 + \frac{1}{\sqrt{2}}\right) \beta_{\rho_k}(z_i, z) \right\},$
9: **return** y_K

Здесь $\beta_{\rho_k}(z_i,z)$ - это расстояние Бергмана, вычисленное с помощью $\rho_k(z)$

$$\beta_{\rho_k}(z_i,z) = \rho_k(z) - \rho_k(z_i) - \langle \nabla \rho_k(z_i), z - z_i \rangle .$$

С помощью $g_{\varphi_k,\tau}(z)$ обозначаем неточный градиент подзадачи (17)

$$g_{\varphi_k,\tau}(z) = \nabla f(\tilde{x}_k) + \nabla^2 f(\tilde{x}_k)[z - \tilde{x}_k] + g^{\tau}_{\tilde{x}_k}(z) + L_3 \|z - \tilde{x}_k\|^2 (z - \tilde{x}_k) \tag{20}$$

и $g^{\tau}_{\tilde{x}_k}(z)$ обозначает неточное приближение $D^3 f(\tilde{x}_k)[y - \tilde{x}_k]^2$

$$g^{\tau}_{\tilde{x}_k}(z) = \frac{1}{\tau^2} \left(\nabla f(\tilde{x}_k + \tau(z - \tilde{x}_k)) + \nabla f(\tilde{x}_k - \tau(z - \tilde{x}_k)) - 2\nabla f(\tilde{x}_k) \right). \tag{21}$$

Если

$$\delta = O\left(\frac{\varepsilon^{\frac{3}{2}}}{\|\nabla f(\tilde{x}_k)\|_*^{\frac{1}{2}} + \|\nabla^2 f(\tilde{x}_k)\|^{\frac{3}{2}} / L_3^{\frac{1}{2}}} \right),$$

то точное количество внутренних итераций будет следующим

$$T_k(\delta) = O\left(\ln \frac{G + H}{\varepsilon} \right), \tag{22}$$

где G и H верхние границы норм градиентов и Гессианов, вычисленные в точках, полученных основным алгоритмом.

Теорема 2. *Пусть f - выпуклая функция, такая что прозводная 3 порядка L_3 является липшецевой и пусть x_* есть минимум f. Тогда чтобы достичь точности ε Алгоритм 2, использующий Алгоритм 3, для решения подзадачи, вычисляет*

$$N_1 = \tilde{O}\left(\left(\frac{L_3 R^4}{\varepsilon} \right)^{\frac{1}{5}} \right), \tag{23}$$

Гессианов и

$$N_2 = \tilde{O}\left(\left(\frac{L_3 R^4}{\varepsilon} \right)^{\frac{1}{5}} \log \left(\frac{G + H}{\varepsilon} \right) \right), \tag{24}$$

градиентов, где G и H равномерные верхние границы норм градиентов и Гессианов, вычисленные в точках, полученных основным алгоритмом.

4 Супербыстрый слайдинг второго порядка

В этом разделе рассматривается задача

$$\min_x f(x) = g(x) + h(x), \tag{25}$$

где $g(x)$ и $h(x)$ являются такими выпуклыми функциями, что $\nabla^3 g$ $L_{3,g}$-липшицева функция и $\nabla^3 h$ $L_{3,h}$-липщицева функция, также $L_{3,g} \leq L_{3,h}$.

Слайдинг - схема, разделяющая обращения к оракулам для $g(x)$ и $h(x)$, вследствие чего уменьшается количество вызовов для $g(x)$. В этом разделе комбинируется слайдинг и Hyperfast Second-Order Method 2, из чего получается Hyperfast Second-Order Sliding.

Во-первых, потребуется вариант агоритма NATMI с гладкой составной частью в качестве внешнего базового метода.

Algorithm 4 3rd-Order NATMI with smooth composite part

1: **Input:** convex functions $h(x)$ and $g(x)$ such that $\nabla^3 h$ is $L_{3,h}$-Lipschitz and $\nabla^3 g$ is $L_{3,g}$-Lipschitz.
2: Set $A_0 = 0, x_0 = y_0$
3: **for** $k = 0$ **to** $k = K - 1$ **do**
4: Compute a pair $\lambda_{k+1} > 0$ and $y_{k+1} \in \mathbb{R}^n$ such that

$$\frac{1}{2} \leq \lambda_{k+1} \frac{3L_{3,g} \cdot \|y_{k+1} - \tilde{x}_k\|^2}{4} \leq \frac{3}{4},$$

where $y_{k+1} \in \left\{ T \in E \,:\, \|\nabla \tilde{\Omega}_{3,3L_{3,g}/2}(g,\tilde{x}_k; T) + \nabla h(T)\| \leq \frac{1}{6} \|\nabla f(T)\| \right\}$

$$a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1} A_k}}{2}\,,\; A_{k+1} = A_k + a_{k+1}\,,\text{ and } \tilde{x}_k = \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x_k\,.$$

5: Update $x_{k+1} := x_k - a_{k+1} \nabla f(y_{k+1})$
6: **return** y_K

Сходимость этого метода доказывается с помощью объединенных доказательств NATMI с CATD; получается та же самая скорость сходимости.

Теперь приводится Hyperfast Second-Order Sliding Method. Он содержит три уровня методов: внешний Алгоритм 4 для $g(x)$ и $h(x)$ как составной части, на следующем уровне Алгоритм (промежуточный) 4 решает подзадачу внешнего с моделью $g(x)$ и $h(x)$, затем внутренний Алгоритм 3 решает внутреннюю подзадачу суммы двух моделей $g(x)$ и $h(x)$.

Algorithm 5 Hyperfast Second-Order Sliding

Input: convex functions $g(x)$ and $h(x)$ such that $\nabla^3 h$ is $L_{3,h}$-Lipschitz and $\nabla^3 g$ is $L_{3,g}$-Lipschitz.
Set $z_0 = y_0 = x_0$
for $k = 0$, **to** $K-1$ **do**
  Run Algorithm 4 for problem $g(x) + h(x)$, where $h(x)$ is a composite part.
  for $m = 0$, **to** $M-1$ **do**
    Run Algorithm 4 up to desired accuracy for subproblem

$$\min_y \left(\tilde{\Omega}_{3,3L_{3,g}/2}(g,\tilde{x}_k;y) + h(y)\right)$$

    for $l = 0$, **to** $L-1$ **do**
      Run Algorithm 3 up to desired accuracy for subproblem

$$\min_y \left(\tilde{\Omega}_{3,3L_{3,g}/2}(g,\tilde{x}_k;y) + \tilde{\Omega}_{3,3L_{3,h}/2}(h,\tilde{x}_k;y)\right)$$

return y_K

Общая сложность - это произведение сложностей каждого внутреннего метода в Алгоритме 5. Можно получить скорости сходимости для внешнего и промежуточного метода. Чтобы достичь $f(x_N) - f(x^*) \leq \varepsilon$, необходимы N_g вычислений производных $g(x)$ и N_h вычислений производных $h(x)$, где

$$N_g = \tilde{O}\left[\left(\frac{L_{3,g}R^4}{\varepsilon}\right)^{\frac{1}{5}}\right], \tag{26}$$

$$N_h = \tilde{O}\left[\left(\frac{L_{3,h}R^4}{\varepsilon}\right)^{\frac{1}{5}}\right]. \tag{27}$$

Также из предыдущего раздела получена скорость сходимости внутреннего Алгоритма (3), которая равна $O\left(\ln\frac{G+H}{\varepsilon}\right)$. Из этого вытекает следующая теорема.

Теорема 3. *Пусть $f(x)$ и $g(x)$ - выпуклые функции с липшицевой 3-й производной и $L_{3,g} < L_{3,h}$. Тогда Метод 5 сходится к $f(x_N) - f(x^*) \leq \varepsilon$ за N_g (26) вычислений Гессиана $f(x)$, N_h (27) вычислений Гессиана $h(x)$, $O\left(N_g \ln\frac{G+H}{\varepsilon}\right)$ вычислений градиентов $g(x)$ и $O\left(N_h \ln\frac{G+H}{\varepsilon}\right)$ вычислений градиентов $h(x)$, где G и H - равномерные верхние границы для норм градиентов и гессианов $f(x)$, вычисленных в точках, генерируемых главным алгоритмом.*

Можно обобщить этот результат для суммы n функций, отсортированных по L_{3,f_i} и примененных последовательно, также возможно разделение на батчи.