УДК 519.853.62

Нетривиальность ускоренных покомпонентных методов выпуклой оптимизации

И.Н. Усманова , А.В. Гасников

Московский физико-технический институт (государственный университет)

Рассматривается метод получения ускоренного рандомизированного покомпонентного метода с помощью каплинга (комбинации) неускоренных прямого градиентного метода(ПГМ) и метода зеркального спуска(МЗС). Будем называть такой метод ACRCD(Accelerated by Coupling Randomized Coordinate Descent)

$$ACRCD(\alpha, \tau; x_0, f(x_0) - f_* \le d)$$

1. $x_{k+1} = \tau z_k + (1-\tau) y_k$
2. $i_{k+1} \in [1,...,n]$
3. $y_{k+1} = Grad_{i_{k+1}}(x_{k+1})$
4. $z_{k+1} = Mirr_{z_k}\left(\alpha n \nabla_{i_{k+1}} f(x_{k+1})\right)$

Где Grad() и Mirr() - стандартные шаги алгоритмов методов градиентного и зеркального спуска. Показано, что метод действительно является ускоренным(в смысле зависимости скорости сходимости от ε требуемой точности).

Доказана скорость сходимости метода при значении параметра $\alpha n^2 = \frac{1-\tau}{\tau}$

в среднем:
$$E\left[f\left(\bar{x}_k\right)\right] - f_* \le \frac{2n\sqrt{\Theta d}}{K} \le \frac{d}{2}$$

И по вероятности,

то есть число шагов, за которое гарантировано можно

достичь точности, ε с вероятностью, не меньшей $1-\sigma$

$$N = O\left(n\sqrt{\frac{\| x_0 - x_* \|^2}{\varepsilon}} \ln\left(\frac{\ln(d/\varepsilon)}{\sigma}\right)\right)$$

Главное преимущество покомпонентный методов - способность учитывать разреженность задачи, то есть вместо зависимости от размерности O(n) получать скорость O(s), где s -

среднее число ненулевых элементов в матрице(квадратичной целевой) функции. У метода ACRCD для произвольных функций остается зависимость от размерности пространства как O(n). Однако, учет разреженности работает для специального вида функций, таких что f(x) = \sum f(x_i), таких, как например

$$f(x) = \| Ax - b \|^2$$

Также, преимущество покомпонентных методов заключается в том, что константа липшица, появляющаяся в оценке есть максимальная константа по всем координатам(ограничение 1 нормы градиента), что меньше, а в случае с плохой обусловленностью сильно меньше, чем просто константа Липшица (ограничение 2 нормы градиента).

Также показано, в какие ограничения нужно наложить на множество Q, чтобы покомпонентные методы действительно заканчивали работу в точке минимума, так как это не всегда так.

Литература

1. *Bubeck S.* [*at al.*] Regret analysis of stochastic and nonstochastic multi-armed bandit problems. // Foundations and Trends in Machine learning. - 2012. - V. 5. № 1. P. 1-122
2. *Nesterov Y.* [*at al.]* Efficiency of coordinate descent methods on large scale optimization problem. // SIAM Journal on Optimization. 2012. V. 22. No 2. P. 341–362.
3. *Allen-Zhu Z., Orecchia L.* [*at al.*] Linear coupling: An ultimate unification of gradient and mirror descent // e-print, 2015.