Convergence Rate of Stochastic Gradient with Constant Step Size

Mark Schmidt
University of British Columbia

July 10, 2014

We want to minimize $f(x) = \mathbb{E}[f_i(x)]$, where the expectation is taken with respect to i. The most common case is minimizing a finite sum,

$$\min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} f_i(x), \tag{0.1}$$

as in problems like least squares and logistic regression. We use the iteraiton

$$x^{k+1} = x^k - \alpha f'_{i_k}(x^k),$$

where at each iteration i_k is sampled uniformly. We will assume that f' is L-Lipschitz, f is μ-strongly convex, $\|f'_i(x)\| \leq C$ for all x and i, that the minimizer is x^*, and $0 < \alpha < 1/2\mu$. We will show that

$$\mathbb{E}[f(x^k) - f(x^*)] \leq (1 - 2\alpha\mu)^k (f(x^0) - f(x^*)) + O(\alpha),$$

$$\mathbb{E}[\|x^k - x^*\|^2] \leq (1 - 2\alpha\mu)^k \|x^0 - x^*\|^2 + O(\alpha),$$

meaning that the function values and iterates converge linearly up to some error level proportional to α. For the special case of (0.1), Proposition 3.4 in the paper of Nedic and Bertsekas ('Convergence Rates of Incremental Subgradient Algorithms', 2000) gives a similar argument/result but here we also consider the function value and we work with the expectation to get rid of the dependence on n.

1 Useful inequalitites

By L-Lipschitz of f', for all x and y we have

$$f(y) \leq f(x) + f'(x)^T(y - x) + \frac{L}{2}\|y - x\|^2.$$

By μ-strong-convexity of f, for all x and y we have

$$f(y) \geq f(x) + f'(x)^T(y - x) + \frac{\mu}{2}\|y - x\|^2.$$

Minimizing both sides in terms of y, by setting $y = x - \frac{1}{\mu}f'(x)$ on the right hand side and using the definition of x^* on the left hand side,

$$f(x^*) \geq f(x) - \frac{1}{\mu}\|f'(x)\|^2 + \frac{1}{2\mu}\|f'(x)\|^2 = f(x) - \frac{1}{2\mu}\|f'(x)\|^2.$$

Also by strong-convexity,

$$f'(x)^T(x - x^*) = (f'(x) - f'(x^*))^T(x - x^*) \geq \mu\|x - x^*\|^2.$$

By definition of i_k and f,

$$\mathbb{E}[f'_{i_k}(x^k)] = f'(x^k).$$

Recall the limit of the geometric series,

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}, \text{ for } |r| < 1.$$

2 Function Value

$$\begin{aligned}
f(x^{k+1}) &\le f(x^k) + f'(x^k)^T(x^{k+1}-x^k) + \frac{L}{2}\left\|x^{k+1}-x^k\right\|^2 && (x = x^k,\ y = x^{k+1} \text{ in } L\text{-Lipshitz inequality}) \\
&= f(x^k) - \alpha f'(x^k)^T f_{i_k}(x^k) + \frac{L\alpha^2}{2}\left\|f'_{i_k}(x^k)\right\|^2 && (\text{eliminate } (x^{k+1}-x^k) \text{ using definition of } x^{k+1}) \\
&\le f(x^k) - \alpha f'(x^k)^T f_{i_k}(x^k) + \frac{L\alpha^2 C^2}{2}. && (\text{use } \left\|f'_i(x^k)\right\| \le C)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[f(x^{k+1}) - f(x^*)] &\le f(x^k) - f(x^*) - \alpha f'(x^k)\mathbb{E}[f_{i_k}(x^k)] + \frac{L\alpha^2 C^2}{2} && (\text{take expectation WRT } i_k, \text{ subtract } f(x^*)) \\
&\le f(x^k) - f(x^*) - \alpha\left\|f'(x^k)\right\|^2 + \frac{L\alpha^2 C^2}{2} && (\text{use } \mathbb{E}[f'_{i_k}(x^k)] = f'(x^k))) \\
&\le f(x^k) - f(x^*) - 2\alpha\mu(f(x^k) - f(x^*)) + \frac{L\alpha^2 C^2}{2} && (\text{use } \frac{1}{2\mu}\left\|f'(x^k)\right\|^2 \ge f(x^k) - f(x^*)) \\
&= (1-2\alpha\mu)(f(x^k) - f(x^*)) + \frac{L\alpha^2 C^2}{2}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[f(x^k) - f(x^*)] &\le (1-2\alpha\mu)^k(f(x^0) - f(x^*)) + \sum_{i=0}^{k}(1-2\alpha\mu)^i\frac{L\alpha^2 C^2}{2} && (\text{apply recursively, take total expectation}) \\
&\le (1-2\alpha\mu)^k(f(x^0) - f(x^*)) + \sum_{i=0}^{\infty}(1-2\alpha\mu)^i\frac{L\alpha^2 C^2}{2} && (\text{extra terms are positive because } \alpha < 1/2\mu) \\
&= (1-2\alpha\mu)^k(f(x^0) - f(x^*)) + \frac{L\alpha C^2}{4\mu}. && (\text{use that } \sum_{i=0}^{\infty}(1-2\alpha\mu)^i = 1/2\alpha\mu)
\end{aligned}$$

3 Iterates

$$\begin{aligned}
\left\|x^{k+1}-x^*\right\|^2 &= \left\|(x^k - \alpha f'_{i_k}(x^k)) - x^*\right\|^2 && (\text{definition of } x^{k+1}) \\
&= \left\|x^k - x^*\right\|^2 - 2\alpha f'_{i_k}(x^k)^T(x - x^*) + \alpha^2\left\|f'_{i_k}(x^k)\right\|^2 && (\text{group } (x^k - x^*), \text{ expand}) \\
&\le \left\|x^k - x^*\right\|^2 - 2\alpha f'_{i_k}(x^k)^T(x^k - x^*) + \alpha^2 C^2. && (\text{use } \left\|f'_i(x^k)\right\| \le C)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\left\|x^{k+1}-x^*\right\|^2] &\le \left\|x^k - x^*\right\|^2 - 2\alpha f'(x^k)^T(x^k - x^*) + \alpha^2 C^2 && (\text{take expectation WRT } i_k) \\
&\le \|x - x^*\|^2 - 2\alpha\mu\|x - x^*\| + \alpha^2 C^2 && (\text{use } f'(x)^T(x - x^*) \ge \mu\|x - x^*\|^2\ ) \\
&= (1-2\alpha\mu)\left\|x^k - x^*\right\|^2 + \alpha^2 C^2
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\left\|x^k - x^*\right\|^2] &\le (1-2\alpha\mu)^k\left\|x^0 - x^*\right\|^2 + \sum_{i=0}^{k}(1-2\alpha\mu)^i\alpha^2 C^2 && (\text{apply recursively, take total expectation}) \\
&\le (1-2\alpha\mu)^k\left\|x^0 - x^*\right\|^2 + \frac{\alpha C^2}{2\mu}. && (\text{as before, use that } \sum_{i=0}^{k}(1-2\alpha\mu)^i \le 1/2\alpha\mu).
\end{aligned}$$