Shashank Singh[1]

4 Convergence rate of subgradient method [25 points] (Adona)

(a) (4 pts) Since the 2-norm is induced by an inner product $\langle \cdot, \cdot \rangle$,

$$\begin{aligned}
\|x^{(k)} - x^\star\|_2^2 &= \|x^{(k-1)} - x^\star - t_k g^{(k-1)}\|_2^2 && \text{(def. of } x^{(k)}) \\
&= \langle x^{(k-1)} - x^\star - t_k g^{(k-1)}, x^{(k-1)} - x^\star - t_k g^{(k-1)} \rangle \\
&= \|x^{(k-1)} - x^\star\|_2^2 - 2t_k \langle x^{(k-1)} - x^\star, g^{(k-1)} \rangle + t_k^2 \|g^{(k-1)}\|_2^2 && \text{(bilinearity of } \langle \cdot, \cdot \rangle) \\
&\leq \|x^{(k-1)} - x^\star\|_2^2 - 2t_k \left(f(x^{(k-1)}) - f(x^\star) \right) + t_k^2 \|g^{(k-1)}\|_2^2,
\end{aligned}$$

where the inequality follows from the definition of a subgradient. ■

(b) (5 pts) If g is a subgradient of f at x, then by the Lipschitz condition on f,

$$\|g\|_2^2 = g^T(x + g - x) \leq f(x + g) - f(x) \leq G\|x + g - x\|_2 = G\|g\|_2, \tag{3}$$

and so $\|g\| \leq G$. Thus, applying the recursive bound from (a) k times then gives

$$\begin{aligned}
0 \leq \|x^{(k)} - x^\star\|_2^2 &\leq \|x^{(0)} - x^\star\|_2^2 + \sum_{i=1}^{k} (-2t_i) \left(f(x^{(i-1)}) - f(x^\star) \right) + t_i^2 \|g^{(i-1)}\|_2^2 \\
&\leq R^2 - 2\sum_{i=1}^{k} t_i \left(f(x^{(i-1)}) - f(x^\star) \right) + G^2 \sum_{i=1}^{k} t_i^2. \quad \blacksquare
\end{aligned}$$

(c) (4 pts) Since $x_{\text{best}}^{(k)}$ is chosen so as to minimize $f(x_{\text{best}}^{(k)})$ over $\{x^{(0)}, \dots, x^{(k)}\}$,

$$2\sum_{i=1}^{k} t_i \left(f(x_{\text{best}}^{(k)}) - f(x^\star) \right) \leq 2\sum_{i=1}^{k} t_i \left(f(x^{(i-1)}) - f(x^\star) \right) \leq R^2 + G^2 \sum_{i=1}^{k} t_i^2,$$

using a rearrangement of the result of part (b). Thus, further rearranging, we have

$$f(x_{\text{best}}^{(k)}) - f(x^\star) \leq \frac{R^2 + G^2 \sum_{i=1}^{k} t_i^2}{2\sum_{i=1}^{k} t_i}. \quad \blacksquare \tag{4}$$

(d) (4 pts) Plugging $t_1 = \cdots = t_k = t$ into (4) and taking the desired limit gives

$$\lim_{k \to \infty} f(x_{\text{best}}^{(k)}) - f(x^\star) \leq \lim_{k \to \infty} \frac{R^2 + G^2 k t^2}{2kt} = \boxed{\frac{G^2 t}{2}.}$$

Thus, the subgradient method with a constant step size t converges to a point at which the objective function exceeds its minimum by no more than $G^2 t/2$.

[1] sss1@andrew.cmu.edu

(e) (4 pts) Taking the desired limit in (4) gives

$$\lim_{k\to\infty} f(x_{\text{best}}^{(k)}) - f(x^\star) \le \lim_{k\to\infty} \frac{R^2 + G^2 \sum_{i=1}^k t_i^2}{2\sum_{i=1}^k t_i} \le \frac{R^2 + G^2 \lim_{k\to\infty} \sum_{i=1}^k t_i^2}{2\lim_{k\to\infty} \sum_{i=1}^k t_i} = \boxed{0.}$$

Thus the subgradient method with step sizes as specified converges to a minimum of f.

(f) (4 pts) Plugging $t_i = R/(G\sqrt{k})$ into (4) gives

$$f(x_{\text{best}}^{(k)}) - f(x^\star) \le \frac{R^2 + R^2 k/k}{2k(R/G)\sqrt{k}} = RGk^{-3/2}. \tag{5}$$

Since the t_i was chosen to minimize (4), this is the best bound we can derive from (4).