Optimization Theory and Algorithms

On Linear Convergence Rates

Linear Rates

Let $\{e_k\} \subset \mathbb{R}$ be a positive error sequence that converges to 0. Typically, if $\{x^k\}$ converges to x^*, then $e_k = \|x^k - x^*\|$ in some norm. The so-called q_1-*factor* of the sequence is defined to be

$$q_1 = \limsup_{k\to\infty} \frac{e_{k+1}}{e_k} \in [0, 1]. \tag{1}$$

The q_1-factor must be less than or equal to 1, otherwise $\{e_k\}$ will diverge to infinity. The smaller the q_1-factor is, the faster the convergence will be.

Whenever $q_1 \in (0, 1)$, we say that the sequence $\{e_k\}$ converges to 0 linearly. Otherwise, we say that $\{e_k\}$ converges to 0 sublinearly if $q_1 = 1$, and superlinearly if $q_1 = 0$.

When $\{e_k\}$ converges to 0 linearly, then there exists a constant $\beta \in [q_1, 1)$ such that for all k sufficiently large,

$$e_{k+1} \le \beta\, e_k. \tag{2}$$

In particular, the sequence

$$\{\beta^k\} = \{\beta, \beta^2, \beta^3, \cdots\}$$

converges to 0 linearly for $\beta \in (0, 1)$. In this case, the q_1-factor is β. If we plot the logarithm sequence $\{\log \beta^k\} = \{(\log \beta)k\}$ against the index k, we get a straight line with a negative slope equal to $\log \beta$.

On the other hand, for $p > 0$ (for example $p = 1$ or 2), the sequence

$$\left\{\frac{1}{k^p}\right\} = \left\{1, \frac{1}{2^p}, \frac{1}{3^p}, \cdots\right\}$$

converges to 0 at a sublinear rate, since

$$\lim_{k\to\infty} \frac{e_{k+1}}{e_k} = \lim_{k\to\infty} \left(\frac{k}{k+1}\right)^p = 1.$$

When is $e_k \le \epsilon$?

First consider the sequence with $e_k = 1/k^p$. Set $1/k^p = \epsilon$ and solve for k. We immediate see that $e_k \le \epsilon$ only when

$$k \ge \left(\frac{1}{\epsilon}\right)^{1/p}.$$

For $\epsilon = 10^{-8}$, we need $k \ge 10^8$ for $p = 1$, and $k \ge 10^4$ for $p = 2$.

Now consider the sequence with $e_k = \beta^k$ ($\beta < 1$). Again set $\beta^k = \epsilon$ and solve for k. We obtain $k = \ln \epsilon / \ln \beta$. Hence, $e_k \le \epsilon$ if and only if

$$k \ge \frac{\ln \epsilon}{\ln \beta} = \frac{-1}{\ln \beta} \ln \frac{1}{\epsilon}$$

Let $\beta = 1 - 1/\gamma$ for some $\gamma > 1$. Then by Taylor expansion at 1,

$$\ln \beta = \ln(1 - 1/\gamma) = 0 - (1/\gamma) + O(1/\gamma^2) \ge -1/\gamma.$$

Hence, $-1/\ln \beta \le \gamma$ and $e_k \le \epsilon$ when

$$k \ge \gamma \ln \frac{1}{\epsilon}.$$

For very small ϵ, the term $\ln \frac{1}{\epsilon}$ is much smaller than $\frac{1}{\epsilon}$ or even $\left(\frac{1}{\epsilon}\right)^{1/2}$. For example, if $\frac{1}{\epsilon} = 10^8$, then $\ln \frac{1}{\epsilon} \approx 18$.

In general and in practice, a linear convergence rate is faster than a sublinear one as long as the number γ (i.e., the condition number) is not excessively large.