LECTURE 6: CONVERGENCE OF AN ALGORITHM

1. Concept of convergence
2. Rate of convergence

General solution method

- General iterative descent algorithm

Step 1: (Initialization)

Start from a solution point x^0.

Set $k = 0$.

Step 2: (Optimality Check)

Check if x^k is optimal (or near optimal).

If Yes, stop and output x^k.

Step 3: (Movement)

Move to an improved solution point x^{k+1}.

(Possibly, $x^{k+1} = x^k + \alpha_k d_k$.)

Set $k \leftarrow k + 1$ and go to Step 2.

Basic terminologies

- Proceeding from k to $k+1$ (a cycle from Steps $2 \to 3 \to 2$) means an iteration.
- An algorithm always terminates (at Step 2) with a desired solution point in a finite number of iterations is a finite algorithm.
- For an infinite sequence of solution points generated by an algorithm, $\{x^k \mid k = 0, 1, \cdots\}$, if $\{x^k\}$ (or a subsequence $\{x^{k_i}\}$) converges to a point x^* that is a desirable solution point, then the algorithm is convergent.

Convergence proof

- Given an iterative algorithm, we need to show that, under certain conditions, the sequence of solutions generated by the algorithm indeed converges to a desired solution.

  - How to come up with such a proof of convergence for the algorithm you designed?

Basic terminologies

- An algorithm converges to a desired solution from any given starting point is said to be <u>globally convergent</u>.

- Let X be a "space" of interests. An algorithm "A" initiated at $x^0 \in X$ would generate a sequence $\{x^k\}$ defined by

$$x^{k+1} = A(x^k) \ ,$$

when A is a point-to-point mapping,

or $$x^{k+1} \in A(x^k) \ ,$$

when A is a point-to-set mapping.

Definitions

1. An algorithm "A" is a mapping defined on a space X that assigns to every point $x \in X$ a subset of X.
2. Let $\Gamma \subset X$ be a solution set of interests and A is an algorithm on X. A continuous real-valued function z on X is a <u>descent function</u> for Γ and A, if

$$\text{(i)}\ z(y) < z(x),\ \forall x \notin \Gamma \text{ and } y \in A(x).$$

$$\text{(ii)}\ z(y) \leq z(x),\ \forall x \in \Gamma \text{ and } y \in A(x).$$

Definitions

3. A point-to-set mapping $A : X \longrightarrow Y$ is closed at $x \in X$, if the conditions

   " $x_k \longrightarrow x,\ \ x_k \in X$ " and

   " $y_k \longrightarrow y,\ \ y_k \in A(x_k)$ "

   imply "$y \in A(x)$".

   Moreover, A is closed on X if it is closed at each point of X.

Observations

Let X be closed. Then

(i) A is closed on X if and only if graph(A)=$\{(x, y) \mid x \in X,\ y \in A(x)\}$ is closed.

(ii) If A is a point-to-point mapping, then continuity implies closedness.

Global convergence theorem

Let X be a space of interests, $\Gamma \subset X$ be a solution set of interests, A be an algorithm on X, and $\{x^k\}_{k=0}^{\infty}$ be a sequence of solutions generated by A from a given x^0 such that $x^{k+1} \in A(x^k)$.

If

(i) $\{x^k\} \subset S$ (a compact subset of X);

(ii) $\exists$ a descent function z for Γ and A;

(iii) A is closed at points outside Γ;

then the limit of any convergent subsequence of $\{x^k\}$ is a solution point in Γ.

Rate of convergence

- Basic concept:

  Let $\{r_k\}_{k=0}^{\infty}$ be a decreasing sequence of "errors" between a current solution and the desired solution. If the sequence converges to zero, we would like to know "how fast" it converges.

- Example

  $\{\frac{1}{k}\}$? $\{(\frac{1}{2})^k\}$? $\{(\frac{1}{k})^k\}$? $\{(\frac{1}{2})^{2^k}\}$?

- Which one converges fastest?

Terminologies

- Definition

Let $\{r_k\}_{k=0}^{\infty}$ be a bounded sequence of real numbers and $s_k = \sup\{r_i \mid i \geq k\}$.

The <u>limit superior</u> of $\{r_k\}$ is

$$\overline{\lim_{k \to \infty}} r_k \triangleq \lim_{k \to \infty} s_k \ .$$

- Definition

Let $\{r_k\}_{k=0}^{\infty}$ be a convergent sequence of real numbers with $\lim_{k \to \infty} r_k = r^*$.

The <u>order of convergence</u> of $\{r_k\}$ is defined as the supremum of the nonnegative numbers p satisfying

$$0 \leq \overline{\lim_{k \to \infty}} \frac{|\ r_{k+1} - r^*\ |}{|\ r_k - r^*\ |^p} < \infty \ .$$

Observations

(i) If the situation of $\frac{0}{0}$ is not involved, we usually consider the definition as

$$\lim_{k\to\infty} \frac{|r_{k+1}-r^*|}{|r_k-r^*|^p} = \beta \ .$$

In this case,

$$| \ r_{k+1} - r^* |= \beta \mid r_k - r^* \mid^p \ .$$

Observations

(ii) $\{r_k\}_{k=1}^{\infty}$ converges faster for larger p.

(a) $r_k = a^k$ (with $0 < a < 1$) converges to 0 with $p = 1$, since $\frac{r_{k+1}}{r_k} = a$;

(b) $r_k = a^{2^k}$ (with $0 < a < 1$) converges to 0 with $p = 2$, since $\frac{r_{k+1}}{r_k^2} = 1$.

Linear rate of convergence

- **Definition**:

  If the sequence r_k converges to r^* such that

  $$\lim_{k\to\infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|} = \beta < 1 \ ,$$

  we say $\{r_k\}$ converges linearly to r^* with convergence ratio β.

Observations

(i) The tail of $\{r_k\}$ converges at least as fast as the geometric sequence $c\beta^k$ for some constant c (geometric convergence).

(ii) A linearly convergent sequence with smaller β converges faster.

(iii) $\beta = 0$ is referred to as superlinear convergence.

(iv) Any convergent sequence with $p > 1$ is superlinear.

Examples

(i) $\{r_k = \frac{1}{k}\} \to 0.$

$$\lim_{k\to\infty} \frac{r_{k+1}}{r_k} = 1.$$

The convergence is of order 1, but it is not linear!

(ii) $\{r_k = (\frac{1}{k})^k\} \to 0.$

(a) $\lim_{k\to\infty} \frac{r_{k+1}}{r_k} = 0.$

The convergence is superlinear.

(b) $\lim_{k\to\infty} \frac{r_{k+1}}{r_k^p} = \infty$ for any $p > 1$. The convergence is of order 1 only !

Examples

- Order 1 convergence: $\{\frac{1}{k}\}$.

  (Arithmetic convergence)

- Linear convergence: $\{(\frac{1}{2})^k\}$.

  (Geometric convergence)

- Superlinear convergence: $\{(\frac{1}{\log(k+1)})^k\}$; $\{(\frac{1}{k})^k\}$.

- Quadratic convergence: $\{(\frac{1}{2})^{2^k}\}$.