Tema 3
Global Convergence of Algorithms

Definition 1.1 Given two sets, X and Y , a set-valued mapping defined on X with range in Y is a map, Φ , which assigns to each x in X a subset $\Phi(x)$ of Y
.

Definition 1.2 *Let X be a set and $x_o \in X$ a given point. Then an* **iterative algorithm**, $\mathcal{A}$, *with initial point x_o is a set-valued mapping $\mathcal{A} : X \to X$ which generates a sequence $\{x_n\}_{n=1}^{\infty}$ according to*

$$x_{n+1} \in \mathcal{A}(x_n)\,, n = 0, 1, \ldots\ .$$

Amongst all the iterative algorithms, we are interested, in particular, in various descent algorithms. In order to define what we mean, we introduce the notion of a solution set Γ in X.

Whichever set is chosen as the solution set, we introduce the corresponding notion of a descent function.

Definition 1.3 *Given $\Gamma \subset X$ and an iterative algorithm $\mathcal{A}$ on X, a continuous real-valued function $Z : X \to \mathbb{R}$ is called a* **descent function** *provided*

1. *If $x \notin \Gamma$ and $y \in \mathcal{A}(x)$, $Z(y) < Z(x)$.*
2. *If $x \in \Gamma$ and $y \in \mathcal{A}(x)$, $Z(y) \leq Z(x)$.*

For the general non-linear programming problem

$$\min f(x), \;\; \text{subject to } \; x \in \Omega,$$

if we let Γ be the set of minimizing points (assuming that they exist) and if $\mathcal{A}$ is an algorithm defined on Ω for which, at each step, $f(x_{k+1}) < f(x_k)$, then we can use f itself as the descent function. This is often the case in practice. On the other hand, for unconstrained problems $\Omega = \mathbb{R}^n$, we often define $\Gamma = \{\boldsymbol{x} \in \mathbb{R}^n | \nabla f(\boldsymbol{x}) = 0\}$. Then, we may design the algorithm $\mathcal{A}$ such that $|\nabla f(\boldsymbol{x})|$ is the descent function.

By an iterative descent algorithm we simply mean an iterative algorithm with an associated solution set and descent function, i.e., a triple $\{\mathcal{A}, \Gamma, Z\}$. What we are interested in is clearly those algorithms whose iterates eventually end up in the solution set Γ.
For iterative descent algorithms there is a general notion of *global convergence* which describes the property that the algorithm converges to the solution set.

Definition 1.4 Let $\{A,\Gamma,Z\}$ be an iterative descent algorithm on a set X. It is **globally convergent** if for any starting point x_0 in X, any accumulation point of the sequence generated by A is in Γ.

Definition 1.5 *Given two metric spaces X and Y, a function $f: X \to Y$ is said to be* **continuous** *on X provided, given $x_o \in X$ and a sequence $\{x_n\}_{n=1}^{\infty}$ such that $x_n \to x_o$ as $n \to \infty$, then the sequence $\{y_n\}_{n=0}^{\infty} = \{f(x_n)\}_{n=1}^{\infty}$ converges to $y_o = f(x_o)$.*

We can think of this in a slightly different way. We consider the graph of f, denoted $Gr(f) := \{(x, y) \in X \times Y \,|\, y = f(x)\}$. Consider a sequence $\{x_n\}$ such that $x_n \to x_o$ and the corresponding sequence $y_n = f(x_n)\,, n = 1, 2, \ldots$. Then, for all n, $(x_n, y_n) \in Gr(f)$. Continuity of f means that if $y_n \to y_o$ then $y_o = f(x_o)$. In other words $(x_o, y_o) \in Gr(f)$. Indeed, we can see that we have the simple proposition,

Proposition 1.6 *Given two metric spaces X and Y, a function $f : X \to Y$ is continuous on X provided $Gr(f) \subset X \times Y$ is closed in $X \times Y$.*

For this reason, we think of continuity of a set-valued mapping in terms of its graph.

Definition 1.7 *Given two metric spaces X and Y and a set valued function Φ from X to Y, we define the graph of Φ,*

$$Gr(\Phi) := \{(x, y) \in X \times Y \,|\, y \in \Phi(x)\}.$$

Then the generalization of continuity to this case is most naturally defined in terms of this graph.

Definition 1.8 *A set-valued mapping $\Phi : X \to Y$ is said to* **closed** *at $x_o \in X$ provided*

(i) $x_k \to x_o$ *as* $k \to \infty$, $x_k \in X$,

(ii) $y_k \to y_o$ *as* $k \to \infty$, $y_k, y_o \in Y$,

implies $y_o \in \Phi(x_o)$. The map Φ is called **closed** *on $S \subset X$ provided it is closed at each $x \in S$.*

Clearly a set-valued map, Φ, that is closed on a set X is exactly one whose graph $\{(x, y) \in S \times Y \mid y \in \Phi(x)\}$ is closed. We remark that closed set-valued mappings are sometimes call **upper-semicontinuous** set-valued mappings.

Example 1.9 As we have seen above, a continuous single valued function is one whose graph is closed. Let us look at a simple example of a *discontinuous* function, namely the Heaviside function, considered as as set-valued function, and defined by

$$H(x) = \begin{cases} \{0\}, x \leq 0 \\ \{1\}, x > 0 \end{cases}$$

This set-valued function does not have a closed graph at $x_o = 0$ since, if $\{x_n\}$ is any sequence of points converging to 0 such that $x_n > 0$ and if, for each n, $y_n = 1 \in H(x_n) = \{1\}$ then $y_n \to 1 = y_o$ as $n \to \infty$, but $y_o = 1 \notin H(0) = \{0\}$. Hence this set-valued function is not closed at $x_o = 0$.

Example 1.10 Let $\Phi : \mathbb{R} \to \mathbb{R}$ be given by

$$\Phi(x) := \left[-\frac{|x|}{2}, \frac{|x|}{2}\right].$$

Now suppose that for some $x_o \in \mathbb{R}$, $x_n \to x_o$ and $y_n \to y_o$ as $n \to \infty$ are sequences such that $y_n \in \Phi(x_n)$. This means that we have

$$-\frac{|x_n|}{2} \leq y_n \leq \frac{|x_n|}{2},$$

and, taking $n \to \infty$, we clearly have

$$-\frac{|x_o|}{2} \leq y_o \leq \frac{|x_o|}{2}.$$

Hence Φ is closed.

Example 1.11 Let us consider the algorithm

$$\mathcal{A}(x) = \begin{cases} \{\frac{1}{2}(x-1)1\}, & x > 1 \\ \{\frac{1}{2}x\}, & 0 \leq x \leq 1\,. \end{cases}$$

and let $\Gamma = \{0\}$. Here, we may take $Z(x) = x$ as the descent function. To see this, suppose $x \neq 0$ which means just that $x \notin \Gamma$. Then we have two cases:

1. If $x > 1$ then $(1/2)(x-1)+1 = (x+1)/2$ which, since $x > 1$ implies that $(x+1)/2 < x$.

2. If $0 < x \leq 1$ then $(1/2)x < x$.

If we start the algorithm with $x_o > 1$ then the sequence x_n generated by the algorithm converges to $x = 1$ which is not in the solution set Γ. This algorithm is not closed at $x = 1$.

Example 1.12 One common algorithm that is imbedded as a sub-algorithm in many descent algorithms is a *line search* algorithm in which we minimize the function $\varphi(\alpha) = f(x_k - \alpha\nabla f(x_k))$ along the line $x_k - \alpha\nabla f(x_k)$, $0 \leq \alpha < \infty$, which we denote by S. Note that, since f may have several minima along the line, the algorithm defined by S is indeed set-valued.

To be more precise, we define the set-valued function $S : \mathbb{R}^{2n} \to \mathbb{R}^n$ by

$$S(\boldsymbol{x}, d) = \{\boldsymbol{y} \in \mathbb{R}^n \,|\, \boldsymbol{y} = \boldsymbol{x} + \alpha\,\boldsymbol{d}, \alpha \geq 0, \text{ and } f(\boldsymbol{y}) = \min_{0 \leq \alpha < \infty} f(\boldsymbol{x} + \alpha\,\boldsymbol{d})\}.$$

It is important later that S, so defined, be closed provided $\boldsymbol{d} \neq 0$. We also *assume* that the set-valued function has non-empty values, i.e., that the function f does indeed have a minimum along lines. Mild conditions, e.g., that f is both continuous and coercive suffice.

Proposition 1.13 *Let f be continuous on $\mathbb{R}^n$. Then the algorithm S is closed at any point $(\boldsymbol{x}, \boldsymbol{d}) \in \mathbb{R}^{2n}$ at which $\boldsymbol{d} \neq 0$.*

Proof: Suppose the $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ and $\{\boldsymbol{d}_k\}_{k=1}^{\infty}$ are sequences and that $\boldsymbol{x}_k \to \boldsymbol{x}_o$ and $\boldsymbol{d}_k \to \boldsymbol{d}_o$ as $k \to \infty$ where $\boldsymbol{d}_o \neq 0$. Suppose $\{\boldsymbol{y}_k\}_{k=1}^{\infty}$ is a sequence such that $\boldsymbol{y}_k \in S(\boldsymbol{x}_k, \boldsymbol{d}_k)$ for all k and that $\boldsymbol{y}_k \to \boldsymbol{y}_o$ as $k \to \infty$. We wish to show that $\boldsymbol{y}_o \in S(\boldsymbol{x}_o, \boldsymbol{d}_o)$.

Now for each integer k, $\boldsymbol{y}_k = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$ for some number $\alpha_k > 0$. Hence

$$\alpha_k = \frac{\|\boldsymbol{y}_k - \boldsymbol{x}_k\|}{\|\boldsymbol{d}_k\|} \underset{k\to\infty}{\longrightarrow} \overline{\alpha} := \frac{\|\boldsymbol{y}_o - \boldsymbol{x}_o\|}{\|\boldsymbol{d}_o\|}$$

which implies that $\boldsymbol{y}_o = \boldsymbol{x}_o + \overline{\alpha}\, \boldsymbol{d}_o$.

It remains to prove that this $\boldsymbol{y}_o$ minimizes f along the line $\boldsymbol{x}_o + \alpha\, \boldsymbol{d}_o$. Observe that, for each k and each α, $0 \leq \alpha < \infty$, we have, by definition of $S(\boldsymbol{x}_k, \boldsymbol{d}_k)$:

$$f(\boldsymbol{y}_k) \leq f(\boldsymbol{x}_k + \alpha\, \boldsymbol{d}_k)\,.$$

By continuity of f, taking $k \to \infty$ leads to $f(\boldsymbol{y}_o) \leq f(\boldsymbol{x}_o + \alpha\, \boldsymbol{d}_o)$ for all α which implies that

$$f(\boldsymbol{y}_o) \leq \min_{0\leq\alpha<\infty} f(\boldsymbol{x}_o + \alpha\, \boldsymbol{d}_o)\,,$$

which implies, by definition, that $\boldsymbol{y}_o \in S(\boldsymbol{x}_o, \boldsymbol{d}_o)$. $\square$

Example 1.14 Consider the scalar-valued function $f(x) = (x-1)^2$. For any $d \neq 0$ we have

$$\min_{0 \leq \alpha < \infty} f(\alpha\, d) = \min_{0 \leq \alpha < \infty} (\alpha\, d - 1)^2 = f(1) = 0\,.$$

Hence $S(0,d) = \{1\}$. On the other hand, for $d = 0$

$$\min_{0 \leq \alpha < \infty} f(\alpha\, d) = \min_{0 \leq \alpha < \infty} f(\alpha\, 0) = f(0) = 1\,.$$

So $S(0,0) = \{0\}$. We see then that if $y_k \in S(0, d_k)\,, d_k \neq 0$ that $y_k = 1$ for all k and certainly $y_k \to 1$ as $k \to \infty$. But $1 \notin S(0,0)$ so S is not closed whenever $d = 0$.

It is often possible to decompose and algorithm $\mathcal{A}$ into two well-defined algorithms $\mathcal{B}$ and $\mathcal{C}$ in the sense that the results of one become the input of the next. We speak of the *composition* of algorithms and write, for example $\mathcal{A} = \mathcal{B} \circ \mathcal{C}$. For example, if we consider the algorithm of steepest descent, the sequence is generated first by the map $G : \mathbb{R}^n \to \mathbb{R}^{2n}$ given by $G(x) = (x, \nabla f(x))$ which gives the initial point and direction for the next step of the overall algorithm and is followed by a line search.

Now we will define what we mean by the composition of two set-valued functions.

Definition 1.15 *Let $\mathcal{A} : X \to Y$ and $\mathcal{B} : Y \to Z$ be two point to set mappings. The composite map $\mathcal{C} = \mathcal{B} \circ \mathcal{A}$ which takes points $x \in X$ to sets $\mathcal{C}(x) \subset Z$ is defined by*

$$\mathcal{C}(x) := \bigcup_{y \in \mathcal{A}(x)} \mathcal{B}(y).$$

Of course, it is of interest to know when a composite map of this type is closed.

Proposition 1.16 *Let $\mathcal{A} : X \to Y$ and $\mathcal{B} : Y \to Z$ be two set-valued mappings. Suppose*

(i) $\mathcal{A}$ is closed at x_o,

(ii) $\mathcal{B}$ is closed on $\mathcal{A}(x_o)$,

(iii) If $x_k \to x_o$ and $y_k \in \mathcal{A}(x_k)$ then there exists a y such that, for some subsequence $\{y_{k_j}\}$, $y_{k_j} \to y$ as $j \to \infty$.

Then the composite map $\mathcal{C} = \mathcal{B} \circ \mathcal{A}$ is closed at x.

Corollary 1.17 *If $\mathcal{A}$ is closed at x_o and $\mathcal{B}$ is closed on $\mathcal{A}(x_o)$, then, if Y is compact, the composite map is closed.*

Corollary 1.18 *If f is a scalar-valued function and $\mathcal{B}$ is a set-valued mapping, then if f is continuous at x and $\mathcal{B}$ is closed at $f(x_o)$ then $\mathcal{C} = \mathcal{B} \circ f(x)$ is closed at x_o.*

Theorem 1.19 Let $\boldsymbol{A}$ be an algorithm on X, and suppose that, given $\boldsymbol{x}_o \in X$, the sequence $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ is generated and satisfies

$$\boldsymbol{x}_{k+1} \in \boldsymbol{A}(\boldsymbol{x}_k).$$

Let a solution set $\Gamma \subset X$ be given, and suppose that

(i) the sequence $\{\boldsymbol{x}_k\}_{k=0}^{\infty} \subset S$ for $S \subset X$ a compact set.

(ii) there is a continuous function Z on X such that

  (a) if $\boldsymbol{x} \notin \Gamma$, then $Z(\boldsymbol{y}) < Z(\boldsymbol{x})$ for all $\boldsymbol{y} \in \boldsymbol{A}(\boldsymbol{x})$.

  (b) if $\boldsymbol{x} \in \Gamma$, then $Z(\boldsymbol{y}) \leq Z(\boldsymbol{x})$ for all $\boldsymbol{y} \in \boldsymbol{A}(\boldsymbol{x})$.

(iii) the mapping $\boldsymbol{A}$ is closed at all points of $X \setminus \Gamma$.

Then the limit of any convergent subsequence of $\{\boldsymbol{x}_k\}_{k=0}^{\infty}$ is a solution.

Proof: Suppose that x^* is a limit point of the sequence $\{x_k\}_{k=0}^{\infty}$. Then there is a subsequence $\{x_{k_j}\}_{j=0}^{\infty}$ such that $x_{k_j} \to x^*$ as $j \to \infty$. Since the descent function Z is continuous, we have $Z(x_{k_j}) \to Z(x^*)$ as $j \to \infty$.

We show, first, that in fact $Z(x_k) \to Z(x^*)$ as $k \to \infty$. To this end, observe first that Z is monotonically decreasing on the sequence $\{x_k\}_{k=0}^{\infty}$ as follows from the property that $x_{k+1} \in A(x_k)$ and from (a) and (b) of (ii). Hence we must have $Z(x_k) - Z(x^*) \geq 0$ for all k.

Now, since $Z(x_{k_j}) \to Z(x^*)$ as $j \to \infty$, given $\varepsilon > 0$ there is a j_0 such that, for $j \geq j_0$, we have

$$Z(x_{k_j}) - Z(x^*) < \varepsilon, \quad \text{for all } j \geq j_0.$$

Hence, for all $k > j_0$

$$Z(x_k) - Z(x^*) = Z(x_k) - Z(x_{k_{j_0}}) + Z(x_{k_{j_0}}) - Z(x^*) < \varepsilon,$$

which shows that $Z(x_k) \to Z(x^*)$ as $k \to \infty$.

Now we want to show that the limit point x^* is a solution. We prove this by contradiction. Suppose that x^* is *not* a solution. We consider the sequence $\{x_{k_j+1}\}_{j=1}^{\infty}$ which has the property that, for each j, $x_{k_j+1} \in A(x_{k_j})$. This new sequence lies in the compact set S and hence contains a convergent subsequence $x_{(k_j+1)_\ell} \to \overline{x}$ as $\ell \to \infty$. Since A is closed on $X \setminus \Gamma$ and, by assumption $x^* \notin \Gamma$, we see that

$$\overline{x} \in A(x^*).$$

On the other hand, the fact that, along the *original* sequence, $Z(x_k) \to Z(x^*)$ implies that we must have $Z(\overline{x}) = Z(x^*)$ and this contradicts property (ii) (a) of the theorem. □

Theorem 1.20 (Convergence with composite maps)

Let X be a nonempty closed set in $\mathbf{R}^n$ and let Ω included in X a nonempty solution set. Let α be a continuous function from $\mathbf{R}^n$ into R and C a point-to-set map from X into X satisfying $\alpha(y)<=\alpha(x)$ for y in C(x). Let B be another point-to-set map that is closed over the complement of Ω and satisfies $\alpha(y)<\alpha(x)$ for y in C(x) if x is not in C(x).

Consider the algorithm defined by the composite map A=CB. Given the initial solution x1 ε X, the sequence is generated as follows:

1. If x_k in Ω then STOP
2. Otherwise, let x_{k+1} in $A(x_k)$, replace k by k+1 and repeat.

Suppose that the set $\Lambda=\{x: \alpha(y)<=\alpha(x_1)\}$ is compact. Then, the algorithm stops in a finite number of steps with a point in Ω or all accumulation points of $\{x_k\}$ belong to Ω.