Advanced Mathematical Programming
IE417

Lecture 13

Dr. Ted Ralphs

Reading for This Lecture

- Chapter 7

Iterative Algorithms

- In previous courses, we have discussed algorithms that were guaranteed to terminate in a finite number of steps, usually with an optimal solution.
- For non-linear optimization, things are not so nice.
- We will be dealing with *iterative algorithms* that produce an infinite sequence of points.
- These algorithms may or may not converge to the optimal solution.

Properties of Iterative Algorithms

- We will be interested in the following properties of an algorithm:
  - Does the algorithm converge?
    - Under what conditions does the algorithm converge?
    - Does it converge to a global optimal solution?
    - Does it converge to a local optimal solution?
  - How quickly does it converge?
  - How much computational effort is involved in each iteration?
  - How robust is the algorithm?
- We will also be interested in the *termination criteria* and the accuracy of the solution.

The Algorithmic Map

- An algorithm is defined by its *algorithmic map*.
- Given our current location, where do we go next?
- This is determined by a mapping $A : X \to 2^X$ which maps each point in the *domain* X to a set of possible "next iterates."
- In other words, if the current iterate is x_k, then $x_{k+1} \in A(x_k)$.
- After terminating the algorithm, the final iterate x^* will be called a *solution*.

The Solution Set

- Our goal is to find a global optimal solution.
- This is usually an unattainable goal.
- A solution x^* may be considered *acceptable* if
  - x^* is a local optimal solution
  - $f(x^*) < b$, an "acceptable value"
  - $f(x^*) < LB + \epsilon$.
  - $f(x^*) < OPT + \epsilon$.
  - x^* is a KKT point.
- The set Ω of acceptable solutions is the *solution set*.
- An algorithmic map is said to be *convergent* over $Y \subseteq X$, if, starting from any point $x_1 \in Y$, the limit of any convergent subsequence of $x_1, x_2, \ldots$, is in Ω.

Closed Maps

- An algorithmic map A is said to be *closed* at $x \in X$ if

  - $x_k \in X$ and $\{x_k\} \to x$
  - $y_k \in A(x_k)$ and $\{y_k\} \to y$

  implies that $y \in A(x)$.

- The map A is said to be closed on $Z \subseteq X$ if it is closed at each point in Z.

A Typical Algorithm

- Let X be a nonempty closed set in $\mathbb{R}^n$, $\Omega \subseteq X$ the solution set, and $A : X \to 2^X$ an algorithmic map on X (for the remainder of the lecture)
- A typical algorithm generates a sequence as follows:
  - Start at x_1 (given)
  - If $x_k \in \Omega$, STOP.
  - Otherwise, choose $x_{k+1} \in A(x_k)$, set $k \to k + 1$ and iterate.
- A function $\alpha : X \to \mathbb{R}$ is called a *descent function* if $\alpha(y) < \alpha(x)$ when $x \notin \Omega$ and $y \in A(x)$.

A Convergence Theorem

Theorem 1. *If the map A is closed over the complement of Ω and α is a continuous descent function, then either the algorithm stops in a finite number of steps or it generates an infinite sequence $\{x_k\}$ such that*

- *Every convergent subsequence of $\{x_k\}$ has a limit in Ω.*
- $\alpha(x_k) \rightarrow \alpha(x)$ *for some* $x \in \Omega$.

- Note this implies that if Ω is a singleton, then the sequence must converge to that single value.
- Typically, the descent function is the objective function.

Termination Criteria

- It's not always easy to know when to stop.
- We may not even know if our current iterate is in Ω.
- Typical termination criteria
  - $\|x_{k+N} - x_k\| < \epsilon$
  - $\|x_{k+1} - x_k\|/\|x_k\| < \epsilon$
  - $\alpha(x_k) - \alpha(x_{k+N}) < \epsilon$
  - $\alpha(x_k) - \alpha(x_{k+N})/|\alpha(x_k)| < \epsilon$

Composition of Mappings

- Most nonlinear algorithms involve the *composition* of two or more mappings.
- Typical example
  - The first mapping determines a descent direction d.
  - The second mapping determines the step size λ.
- The *composite map* $A = CB$ is defined by

$$A(x) = \cup\{C(y) : y \in B(x)\}$$

Closedness of Composite Maps

Theorem 2. *Let X, Y, and Z be nonempty closed sets in $\mathbb{R}^n$, $\mathbb{R}^p$, and $\mathbb{R}^q$ respectively. Let $B : X \to 2^Y$, $C : Y \to 2^Z$ be algorithmic maps and consider $A = CB$. Suppose*

- *B is closed at x and C is closed at $B(x)$, and*
- *if $\{x_k\} \to x$ and $y_k \in B(x_k)$, then there is a convergent subsequence of $\{y_k\}$.*

Then A is closed at x.

- Note that if Y is compact, then the closedness of B and C implies the closedness of A.

Convergence of Composite Maps

- Although both maps must be closed to ensure the composite is closed, we can get convergence with fewer assumptions.
- Let $\alpha : X \rightarrow \mathbb{R}$ be a continuous function. Consider the algorithmic map $C : X \rightarrow 2^X$ such that $\alpha(y) \le \alpha(x)$ for $y \in C(x)$. Let $B : X \rightarrow 2^X$ be an algorithmic map closed over the complement of Ω such that $\alpha(y) < \alpha(x)$ for $y \in B(x)$, $x \notin \Omega$. Consider the typical algorithm defined by $A = CB$ and suppose that $\{x : \alpha(x) \le \alpha(x_1)\}$ is compact. Then, either the algorithm stops in a finite number of steps or all accumulation points of x_k are in Ω.

Minimizing Along Independent Directions

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and consider the problem to minimize $f(x)$ over $x \in \mathbb{R}^n$.
- Suppose we use the following algorithm
  - Find n linearly independent search directions $d_1, \ldots, d_n$.
  - Minimize f along each search direction to get the next iterate.
- Suppose the minimum of f along each search direction is unique.
- If the sequence $\{x_k\}$ is contained in a compact set, then each accumulation point x satisfies $\nabla f(x) = 0$.

Rate of Convergence

- We want not only convergence, but "fast" convergence.
- Let the sequence $\{x_k\}$ converge to limit point x.
- Intuitively, the *order of convergence* of the sequence $\{x_k\}$ is the largest power p such that

$$|x_{k+1} - x| \approx \beta |x_k - x|^p$$

- If $p = 1$ and $\beta < 1$ then we have *linear convergence*.
- If $\beta > 1$ then we have *superlinear convergence*.
- If $p = 2$ then we have *quadratic convergence*.