Numerical Optimization
Unconstrained Optimization

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization
Unconstrained Minimization Algorithm

(1) Initialize $x^0$, set $k := 0$.

(2) while stopping condition is not satisfied at $x^k$
   
   (a) Find $x^{k+1}$ such that $f(x^{k+1}) < f(x^k)$.
   
   (b) $k := k + 1$

endwhile

Output: $x^* = x^k$, a local minimum of $f(x)$.

- How to find $x^{k+1}$ in Step 2(a) of the algorithm?
- Which stopping condition can be used?
- Does the algorithm converge? If yes, how fast does it converge?
- Does the convergence and its speed depend on $x^0$?
Stopping Conditions for a minimization problem:

- \( \|g(x^k)\| = 0 \) and \( H(x^k) \) is positive semi-definite

Practical Stopping conditions

Assumption: There are no stationary points

- \( \|g(x^k)\| \leq \epsilon \)
- \( \|g(x^k)\| \leq \epsilon(1 + |f(x^k)|) \)
- \( \frac{f(x^k) - f(x^{k+1})}{|f(x^k)|} \leq \epsilon \)
Assume that an optimization algorithm generates a sequence \( \{x^k\} \), which converges to \( x^* \).

How fast does the sequence converge to \( x^* \)?

**Definition**

The sequence \( \{x^k\} \) converges to \( x^* \) with order \( p \) if

\[
\lim_{k \to \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^p} = \beta, \quad \beta < \infty
\]

- Asymptotically, \( \|x^{k+1} - x^*\| = \beta \|x^k - x^*\|^p \)
- Higher the value of \( p \), faster is the convergence.
- \( \beta \) : Convergence rate
(1) $p = 1, 0 < \beta < 1$ (Linear Convergence)
Some Examples:

- $\beta = .1, \|x^0 - x^*\| = .1$
  Norms of $\|x^k - x^*\| : 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, \ldots$

- $\beta = .9, \|x^0 - x^*\| = .1$
  Norms of $\|x^k - x^*\| : 10^{-1}, .09, .081, .0729, \ldots$

(2) $p = 2, \beta > 0$ (Quadratic Convergence)
Example:

- $\beta = 1, \|x^0 - x^*\| = .1$
  Norms of $\|x^k - x^*\| : 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}, \ldots$

(3) Suppose an algorithm generates a convergent sequence $\{x^k\}$ such that

$$\lim_{k \to \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0 \text{ and } \lim_{k \to \infty} \frac{\|x^{k+1} - x^*\|^2}{\|x^k - x^*\|^2} = \infty$$

then this convergence is called superlinear convergence
Examples:

- The sequence with $x^k = 1 + a^k$ where $0 < a < 1$ converges to 1 \textit{linearly}, with convergence rate, $\beta = a$.
- The sequence $x^k = a^{(2^k)}$ where $0 < a < 1$ converges to zero \textit{quadratically}, with convergence rate, $\beta = 1$.
- The sequence $1 + k^{-k}$ converges \textit{superlinearly} to 1.
Use of Error Functions

- Suppose the sequence $x^k$ converges to $x^*$. 
- Let $E : \mathbb{R}^n \to \mathbb{R}$, $E \in C^0$ 
- Convergence properties of $x^k$ can be studied by analyzing the convergence of $E(x^k)$ to $E(x^*)$. 
- In general, the order of convergences of a sequence is \textit{insensitive} to the choice of error function.
Unconstrained Minimization Algorithm

(1) Initialize $x^0$ and $\epsilon$, set $k := 0$. 
(2) while $\|g(x^k)\| > \epsilon$
   (a) Find $x^{k+1}$ such that $f(x^{k+1}) < f(x^k)$.
   (b) $k := k + 1$
endwhile
Output: $x^* = x^k$, a stationary point of $f(x)$.

How to find $x^{k+1}$ in Step 2(a)?

- Find a descent direction $d^k$ for $f$ at $x^k$
- Take a step $\alpha^k (> 0)$ along $d^k$ such that
  - $f(x^{k+1}) < f(x^k)$
  - $x^{k+1} = x^k + \alpha^k d^k$
Descent direction set: \( \{ d \in \mathbb{R}^n : g_k^T d < 0 \} \) where \( g^k = g(x^k) \)
Unconstrained Minimization Algorithm

(1) Initialize $x^0$ and $\epsilon$, set $k := 0$.

(2) while $\|g(x^k)\| > \epsilon$
   (a) Find a descent direction $d^k$ for $f$ at $x^k$
   (b) Find $\alpha^k (> 0)$ along $d^k$ such that $f(x^k + \alpha^k d^k) < f(x^k)$
   (c) $x^{k+1} = x^k + \alpha^k d^k$
   (d) $k := k + 1$

endwhile

Output: $x^* = x^k$, a stationary point of $f(x)$.

- How to determine $\alpha^k$ in Step 2(b)?
Step Length Determination

- **Exact Line Search**: Given a descent direction $d^k$, determine $\alpha^k$ by solving the optimization problem:

$$
\alpha^k = \arg \min_{\alpha > 0} \phi(\alpha) \equiv f(x^k + \alpha d^k)
$$

- **Inexact Line Search**: Choice of $\alpha^k$ is crucial
Consider the problem,

$$\min \ x^2$$
Example: Consider the problem,

$$\min x^2$$

- Local and global minimum at $$x^* = 0$$
- Let $$x^k = (-1)^k(1 + 2^{-k})$$ and $$d^k = (-1)^k$$, $$k = 0, 1, 2, \ldots$$

$$\{x\} : \{2, -\frac{3}{2}, \frac{5}{4}, -\frac{9}{8}, \ldots\}$$

$$\{f\} : \{4, \frac{9}{4}, \frac{25}{16}, \frac{81}{64}, \ldots\}$$

- $$f(x^{k+1}) < f(x^k) \forall k = 0, 1, 2, \ldots$$
- The sequence $$x^k$$ does not converge.
• Small decrease in function values relative to the step length
Example: Consider the problem,

$$\min \ x^2$$

- Local and global minimum at $$x^* = 0$$
- Let $$x^k = (1 + 2^{-k})$$ and $$d^k = -1, \ k = 0, 1, 2, \ldots$$

$$\{x\} : \{2, \frac{3}{2}, \frac{5}{4}, \frac{9}{8}, \ldots\}$$

$$\{f\} : \{4, \frac{9}{4}, \frac{25}{16}, \frac{81}{64}, \ldots\}$$

- $$f(x^{k+1}) < f(x^k) \ \forall \ k = 0, 1, 2, \ldots$$
- $$\lim_{k \to \infty} x^k = 1 \neq x^*$$
Step sizes are too small relative to the initial rate of decrease of $f$
Inexact Line Search

\[ \phi(\alpha) = f(x^k + \alpha d^k) \]
Need to avoid

- Small decrease in function values relative to the step length
- Small step sizes
Armijo’s condition ensures sufficient decrease in the function value

\[ \phi(\alpha) = f(x^k + \alpha d^k) \]
Define $\phi_1(\alpha) = f(x^k) + c_1 \alpha g^k d^k$, $c_1 \in (0, 1)$
Choose $\alpha^k$ such that $f(x^k + \alpha^k d^k) \leq \phi_1(\alpha^k)$ (Armijo’s condition)
Goldstein’s condition ensures that step lengths are not too small

\[ \phi(\alpha) = f(x^k + \alpha d^k) \]
Define $\phi_2(\alpha) = f(x^k) + c_2 \alpha g_k^T d^k$, $c_2 \in (c_1, 1)$

Choose $\alpha^k$ such that $f(x^k + \alpha^k d^k) \geq \phi_2(\alpha^k)$ (Goldstein’s condition)
Armijo-Goldstein Conditions: Choose $\alpha^k$ such that

$$\phi_2(\alpha^k) \leq f(x^k + \alpha^k d^k) \leq \phi_1(\alpha^k)$$
Wolfe’s condition ensures sufficient rate of decrease of function value in the given direction

Choose $\alpha_k$ such that

$$\phi'(\alpha^k) \geq c_2 \phi'(0), \quad c_2 \in (c_1, 1)$$

Wolfe’s Condition