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 $\boldsymbol{x}^0$, set $k := 0$.

(2) **while** *stopping condition is not satisfied at* $\boldsymbol{x}^k$

  (a) Find $\boldsymbol{x}^{k+1}$ such that $f(\boldsymbol{x}^{k+1}) < f(\boldsymbol{x}^k)$.

  (b) $k := k + 1$

endwhile

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

- How to find $\boldsymbol{x}^{k+1}$ 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 $\boldsymbol{x}^0$?

Stopping Conditions for a minimization problem:

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

Practical Stopping conditions

Assumption: There are no *stationary* points

- $$\|\boldsymbol{g}(\boldsymbol{x}^k)\| \leq \epsilon$$
- $$\|\boldsymbol{g}(\boldsymbol{x}^k)\| \leq \epsilon(1 + |f(\boldsymbol{x}^k)|)$$
- $$\frac{f(\boldsymbol{x}^k) - f(\boldsymbol{x}^{k+1})}{|f(\boldsymbol{x}^k)|} \leq \epsilon$$

Speed of Convergence

- Assume that an optimization algorithm generates a sequence $\{\boldsymbol{x}^k\}$, which converges to $\boldsymbol{x}^*$.
- How *fast* does the sequence converge to $\boldsymbol{x}^*$?

Definition

The sequence $\{\boldsymbol{x}^k\}$ converges to $\boldsymbol{x}^*$ with order p if

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

- Asymptotically, $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\| = \beta\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^{p}$
- Higher the value of p, faster is the convergence.
- β : Convergence rate

(1) $p = 1, 0 < \beta < 1$ (Linear Convergence)

Some Examples:

- $\beta = .1, \|\boldsymbol{x}^0 - \boldsymbol{x}^*\| = .1$
  Norms of $\|\boldsymbol{x}^k - \boldsymbol{x}^*\| : 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, \ldots$
- $\beta = .9, \|\boldsymbol{x}^0 - \boldsymbol{x}^*\| = .1$
  Norms of $\|\boldsymbol{x}^k - \boldsymbol{x}^*\| : 10^{-1}, .09, .081, .0729, \ldots$

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

Example:

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

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

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

then this convergence is called superlinear convergence

Examples:

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

Use of Error Functions

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

Unconstrained Minimization Algorithm

(1) Initialize $\boldsymbol{x}^0$ and ϵ, set $k := 0$.

(2) **while** $\|\boldsymbol{g}(\boldsymbol{x}^k)\| > \epsilon$

  (a) Find $\boldsymbol{x}^{k+1}$ such that $f(\boldsymbol{x}^{k+1}) < f(\boldsymbol{x}^k)$.

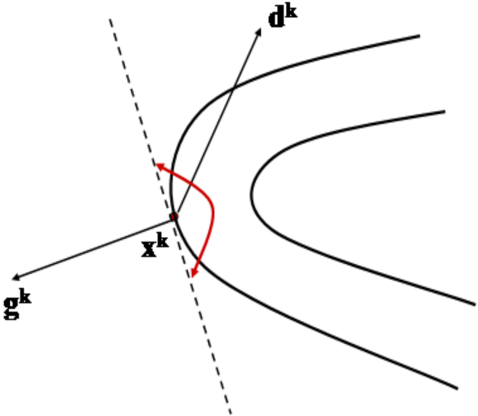
  (b) $k := k + 1$

endwhile

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

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

- Find a *descent direction* $\boldsymbol{d}^k$ for f at $\boldsymbol{x}^k$
- Take a step $\alpha^k (> 0)$ along $\boldsymbol{d}^k$ such that
  - $f(\boldsymbol{x}^{k+1}) < f(\boldsymbol{x}^k)$
  - $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$



- Descent direction set: $\{\boldsymbol{d} \in \mathbb{R}^n : \boldsymbol{g}^{k^T}\boldsymbol{d} < 0\}$ where $\boldsymbol{g}^k = \boldsymbol{g}(\boldsymbol{x}^k)$

Unconstrained Minimization Algorithm

(1) Initialize $\boldsymbol{x}^0$ and ϵ, set $k := 0$.

(2) **while** $\|\boldsymbol{g}(\boldsymbol{x}^k)\| > \epsilon$

  (a) Find a descent direction $\boldsymbol{d}^k$ for f at $\boldsymbol{x}^k$

  (b) Find $\alpha^k (> 0)$ along $\boldsymbol{d}^k$ such that $f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) < f(\boldsymbol{x}^k)$

  (c) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$

  (d) $k := k + 1$

endwhile

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

- How to determine α^k in Step 2(b)?

Step Length Determination

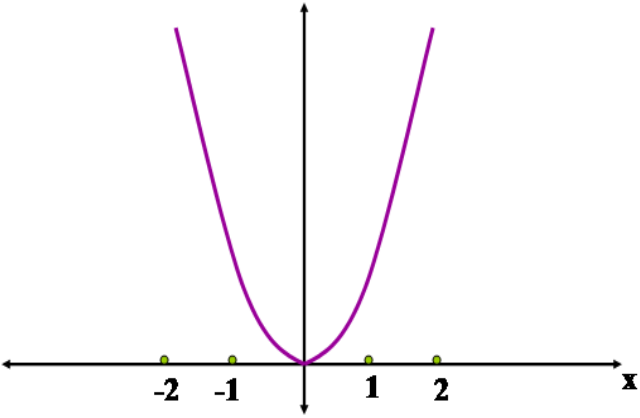
- Exact Line Search : Given a descent direction $\boldsymbol{d}^k$, determine α^k by solving the optimization problem:

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

- Inexact Line Search :
    - Choice of α^k is crucial

Consider the problem,

$$\min \quad x^2$$



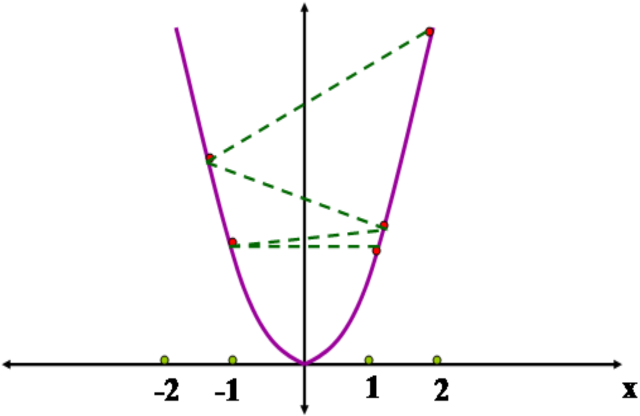
Example: Consider the problem,

$$\min \quad x^2$$

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

$$\{x\} : \{2, -\frac{3}{2}, \frac{5}{4}, -\frac{9}{8}, \ldots\}$$
$$\{f\} : \{4, \frac{9}{4}, \frac{25}{16}, \frac{81}{64}, \ldots\}$$

- $f(x^{k+1}) < f(x^k) \; \forall \; k = 0, 1, 2, \ldots$
- The sequence x^k *does not* converge.



- Small decrease in function values relative to the step length

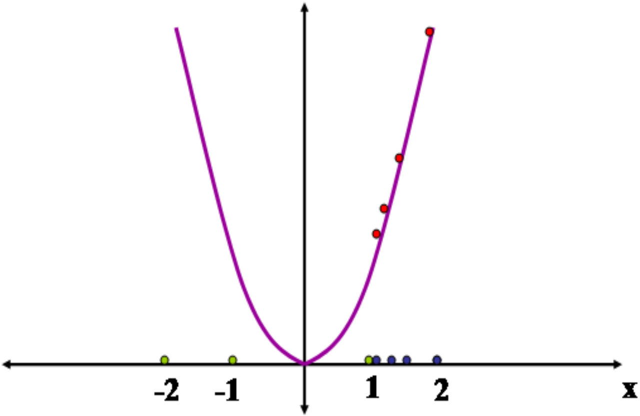
Example: Consider the problem,

$$\min \quad x^2$$

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

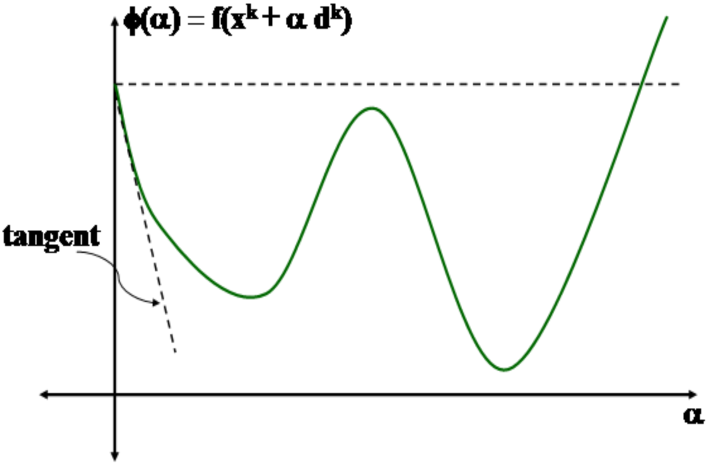
$$\{x\} : \{2, \frac{3}{2}, \frac{5}{4}, \frac{9}{8}, \ldots\}$$
$$\{f\} : \{4, \frac{9}{4}, \frac{25}{16}, \frac{81}{64}, \ldots\}$$

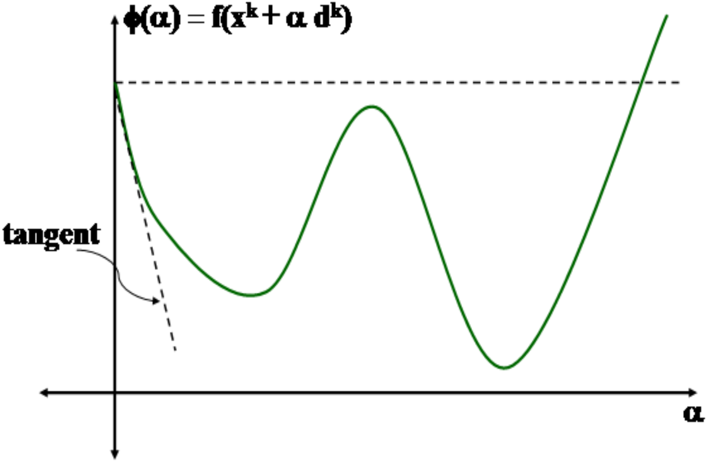
- $f(x^{k+1}) < f(x^k) \; \forall \, k = 0, 1, 2, \ldots$
- $\lim_{k\to\infty} x^k = 1 \neq \boldsymbol{x}^*$



- Step sizes are too small relative to the initial rate of decrease of f

Inexact Line Search

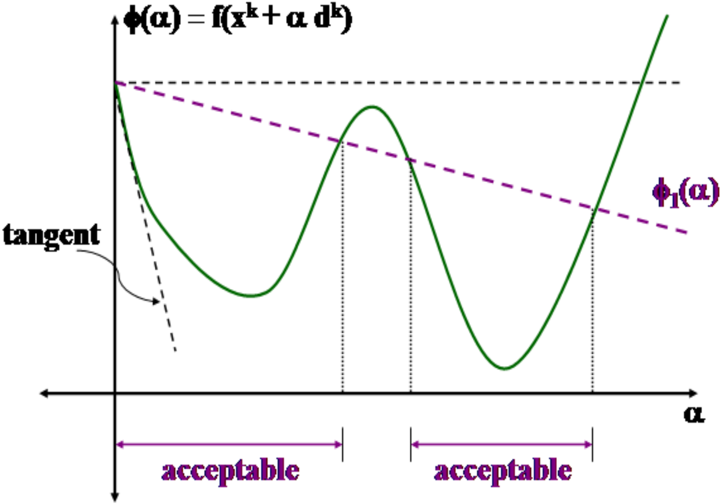


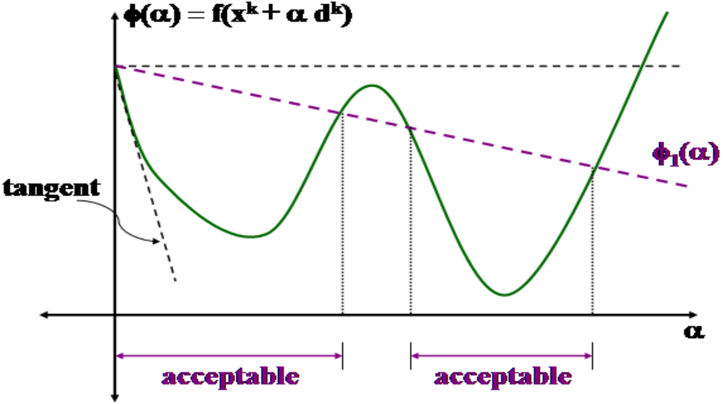


Need to avoid

- Small decrease in function values relative to the step length
- Small step sizes

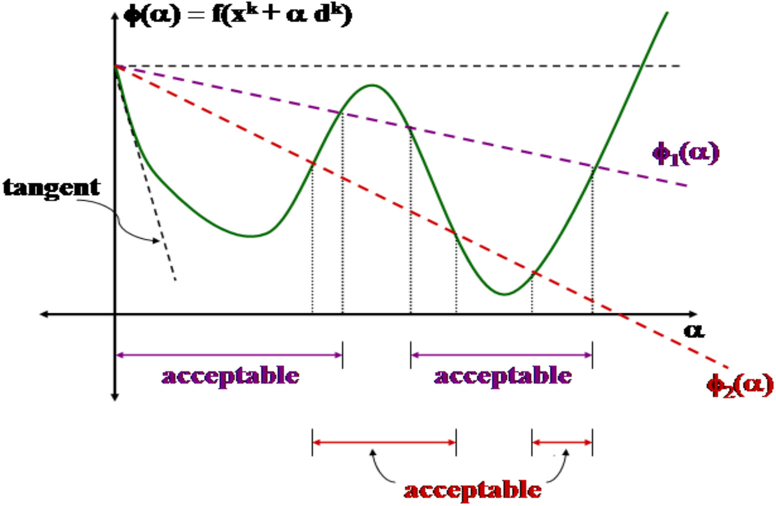
Armijo's condition ensures sufficient decrease in the function value

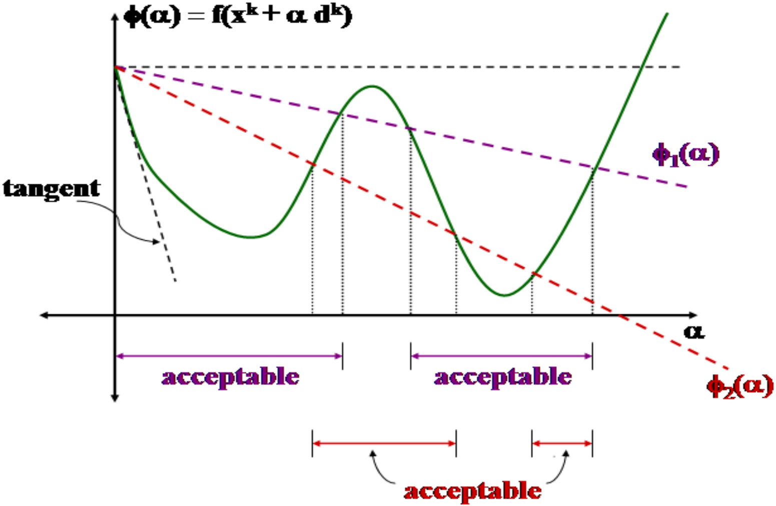




Define $\phi_1(\alpha) = f(\boldsymbol{x}^k) + c_1 \alpha \boldsymbol{g}^{k^T} \boldsymbol{d}^k, \quad c_1 \in (0, 1)$
Choose α^k such that $f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) \leq \phi_1(\alpha^k)$ (Armijo's condition)

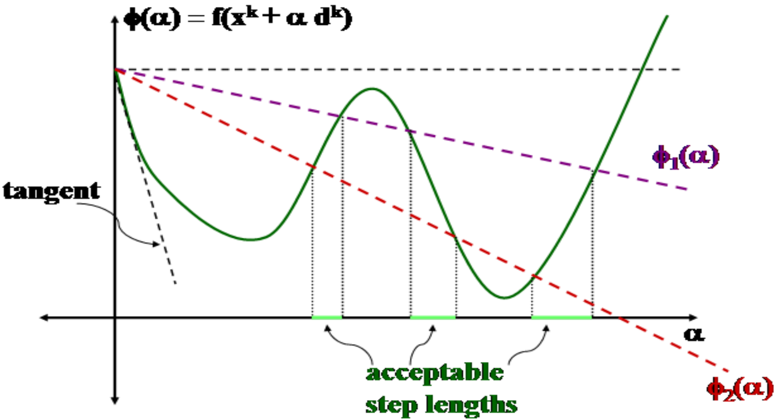
Goldstein's condition ensures that step lengths are not too small





Define $\phi_2(\alpha) = f(\boldsymbol{x}^k) + c_2 \alpha \boldsymbol{g}^{k^T} \boldsymbol{d}^k, \quad c_2 \in (c_1, 1)$

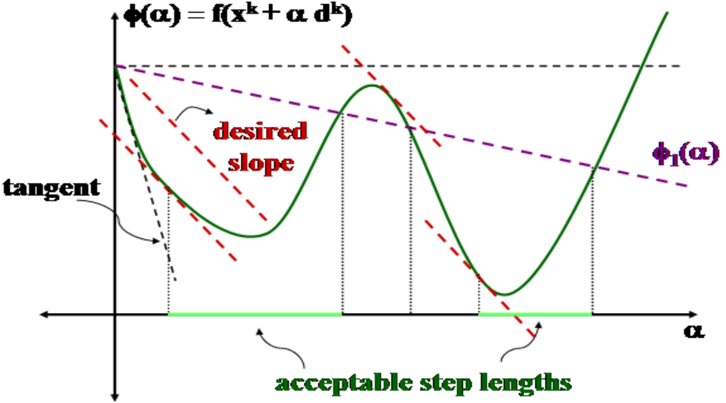
Choose α^k such that $f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) \geq \phi_2(\alpha^k)$ (Goldstein's condition)



Armijo-Goldstein Conditions: Choose α^k such that

$$\phi_2(\alpha^k) \leq f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) \leq \phi_1(\alpha^k)$$

Wolfe's condition ensures sufficient rate of decrease of function value in the given direction



Choose α^k such that

$$\phi'(\alpha^k) \geq c_2 \phi'(0), \quad c_2 \in (c_1, 1) \qquad \text{Wolfe's Condition}$$