Lecture 14: Globalization strategies

- Two basic globalization strategies: line search (Ch. 3) and trust-region (Ch. 4, not syllabus)
  - Note: “globalization” does not imply that we search for global optimum, but we make the algorithm work far from a (local or global) optimum!
- Step-length, Wolfe conditions
- Step-length computation
- Hessian modifications

Reference: N&W Ch.3-3.1, 3.4, 3.5

A general algorithm for unconstrained optimization

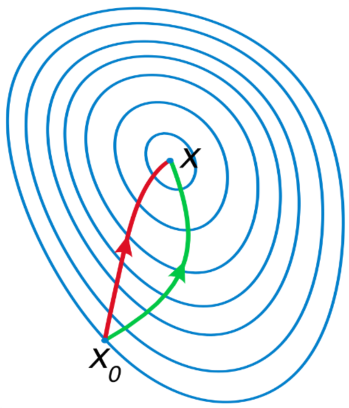
1. Initial guess x_0
2. While termination criteria not fulfilled
   a) Find descent direction p_k from x_k
   b) Walk along p_k to x_{k+1} (how long? – line search!)
   c) $k = k+1$
3. $x_M = x^*$? (possibly check sufficient conditions for optimality)

Termination criteria:
Stop when first of these become true:

- $\|\nabla f(x_k)\| \leq \epsilon$ (necessary condition)
- $\|x_k - x_{k-1}\| \leq \epsilon$ (no progress)
- $\|f(x_k) - f(x_{k-1})\| \leq \epsilon$ (no progress)
- $k \leq k_{\max}$ (kept on too long)

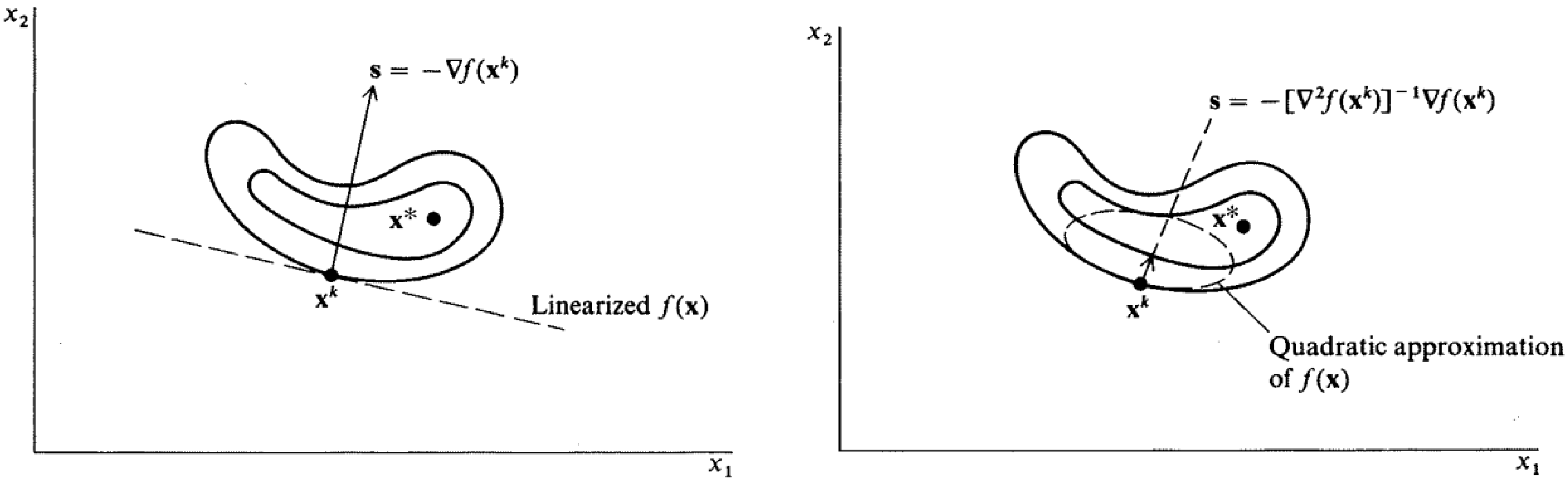
Descent directions:

- Steepest descent
  $p_k = -\nabla f(x_k)$
- Newton
  $p_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$
- Quasi-Newton
  $p_k = -B_k^{-1} \nabla f(x_k)$
  $B_k \approx \nabla^2 f(x_k)$



A comparison of steepest descent (green) and Newton's method (red) for minimizing a function (with small step sizes). Newton's method uses curvature information to take a more direct route. (wikipedia.org)

Steepest descent direction vs Newton direction from objective function approximation

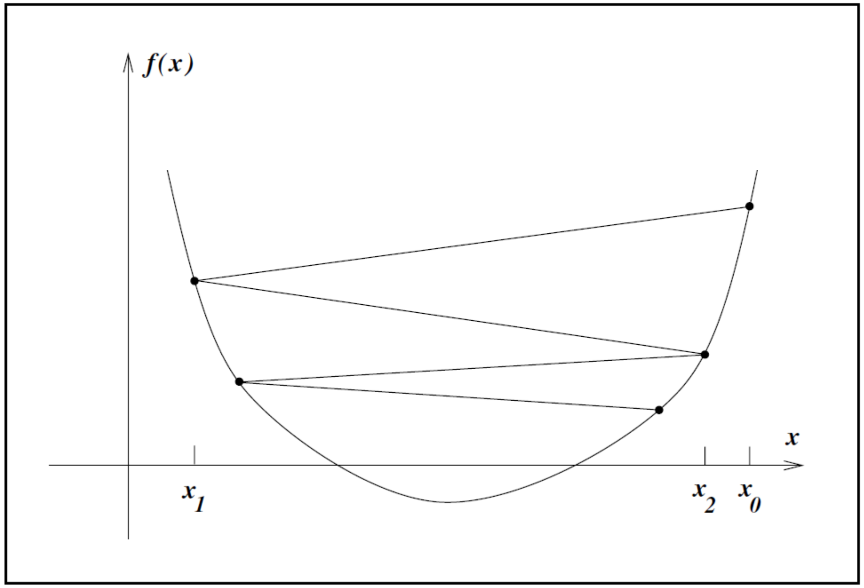


(a) Steepest descent: first-order approximation (linearization) of $f(\mathbf{x})$ at $\mathbf{x}^k$

(b) Newton's method: second-order (quadratic) approximation of $f(\mathbf{x})$ at $\mathbf{x}^k$

From Edgar, Himmelblau, Lasdon: “Optimization of Chemical Processes”

Why sufficient decrease?



- Decrease not enough, need <u>sufficient decrease</u> (1st Wolfe condition)

Sufficient decrease

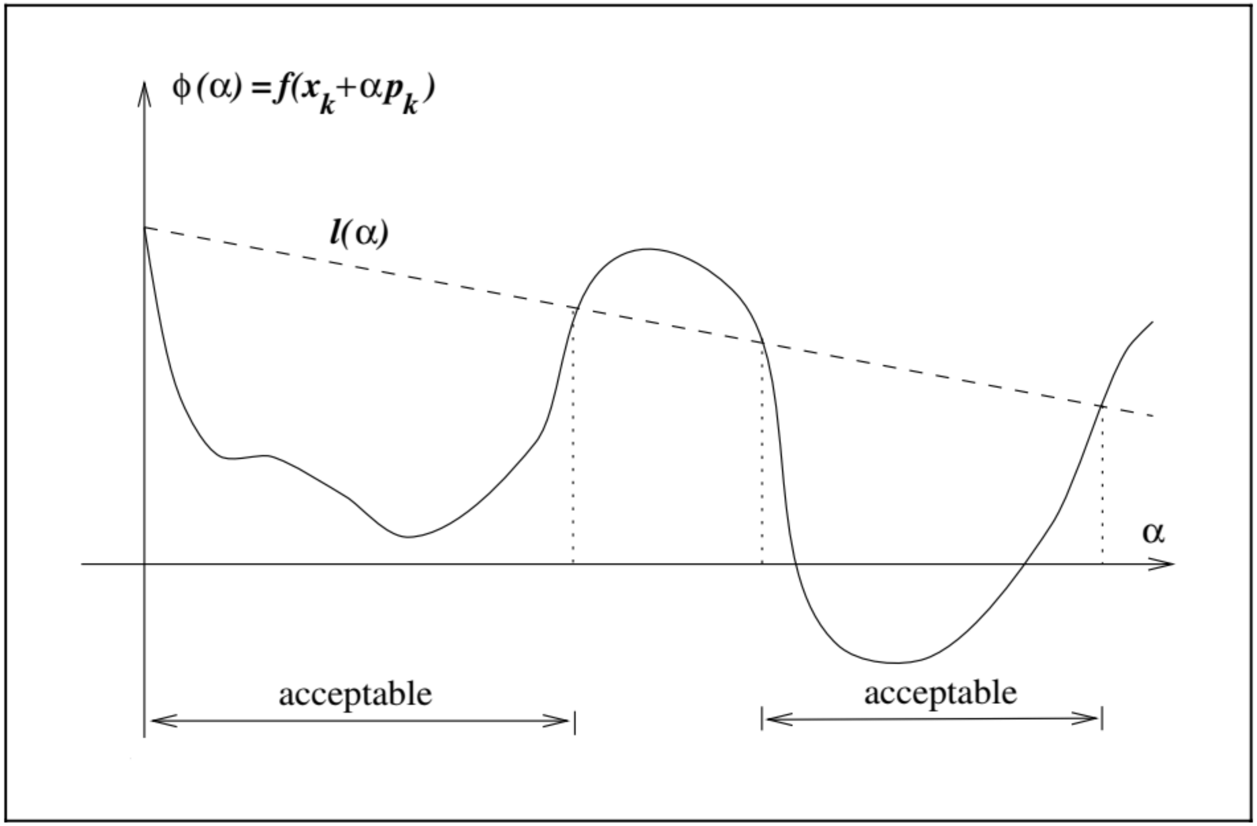


Figure 3.3 Sufficient decrease condition.

Curvature condition

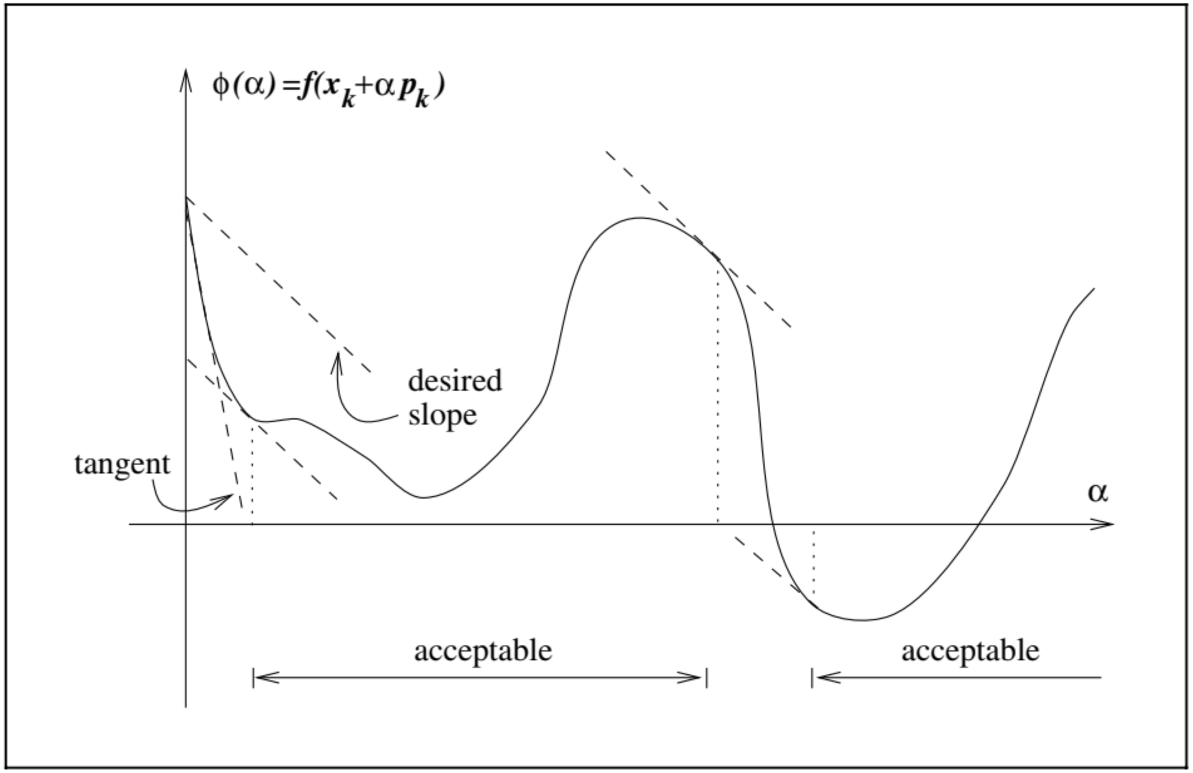


Figure 3.4 The curvature condition.

Backtracking Line Search

Algorithm 3.1 (Backtracking Line Search).

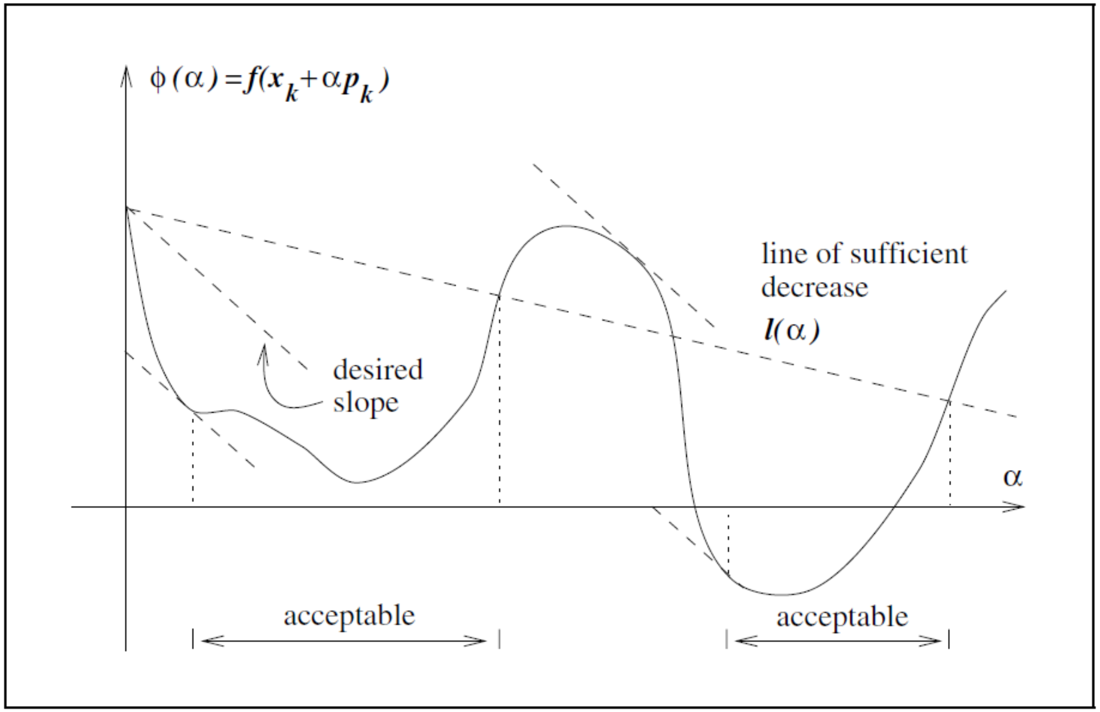
Choose $\bar{\alpha} > 0$, $\rho \in (0, 1)$, $c \in (0, 1)$; Set $\alpha \leftarrow \bar{\alpha}$;

repeat until $f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \nabla f_k^T p_k$

$\alpha \leftarrow \rho\alpha$;

end (repeat)

Terminate with $\alpha_k = \alpha$.



Line search Newton

Algorithm 3.2 (Line Search Newton with Modification).

Given initial point x_0;

for $k = 0, 1, 2, \ldots$

Factorize the matrix $B_k = \nabla^2 f(x_k) + E_k$, where $E_k = 0$ if $\nabla^2 f(x_k)$ is sufficiently positive definite; otherwise, E_k is chosen to ensure that B_k is sufficiently positive definite;

Solve $B_k p_k = -\nabla f(x_k)$;

Set $x_{k+1} \leftarrow x_k + \alpha_k p_k$, where α_k satisfies the Wolfe, Goldstein, or Armijo backtracking conditions;

end

Local convergence rates

Steepest descent:
Linear convergence

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq r \quad \text{for all } k \text{ sufficiently large, } r \in (0,1)$$

Newton:
Quadratic convergence

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq M \quad \text{for all } k \text{ sufficiently large, } M > 0$$

Quasi-Newton:
Superlinear convergence

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

$$\frac{\|x_{k+1} - x^*\|}{\|x_0\|}$$