Norwegian University of Science and Technology

TTK4135 – Lecture 14
Line search

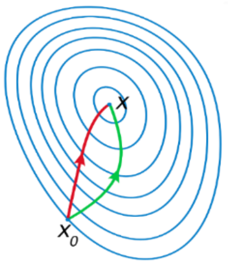
Lecturer: Lars Imsland

Learning goal Ch. 2, 3 and 6: Understand this slide

Line-search unconstrained optimization

$$\min_x f(x)$$

1. Initial guess x_0
2. While termination criteria not fulfilled
   a) Find descent direction p_k from x_k
   b) Find appropriate step length α_k ; set $x_{k+1} = x_k + \alpha_k p_k$
   c) $k = k+1$
3. $x_M = x^*$? (possibly check sufficient conditions for optimality)

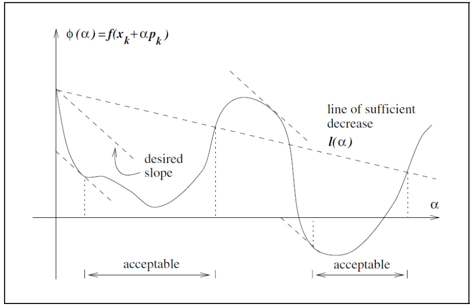


A comparison of steepest descent and Newton's method. Newton's method uses curvature information to take a more direct route. (wikipedia.org)

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

Step length line search (Wolfe):



How to calculate derivatives – Ch. 8

How many iterations? (Convergence rates)

Outline today

Objective of line search: make gradient algorithms work when you start far away from optimum

- These algorithms are sometimes called globalization strategies
- Two basic globalization strategies: line search (Ch. 3) and trust-region (Ch. 4, not syllabus)
  - Note again: “globalization” does not imply that we search for global optimum, but we make the algorithm work far from a (local or global) optimum!

Line search elements:

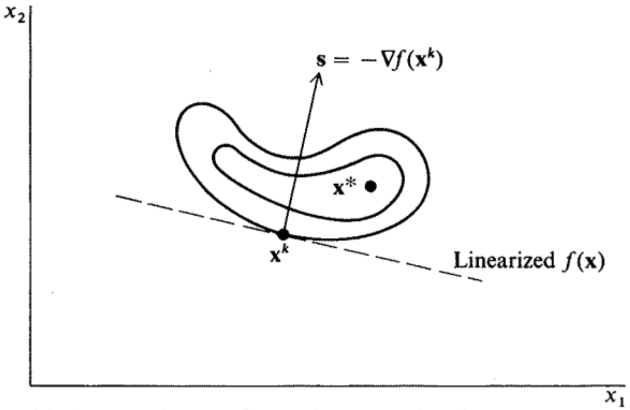
- Conditions on step-length: Wolfe conditions
- Step-length computation

Hessian modification for Newton

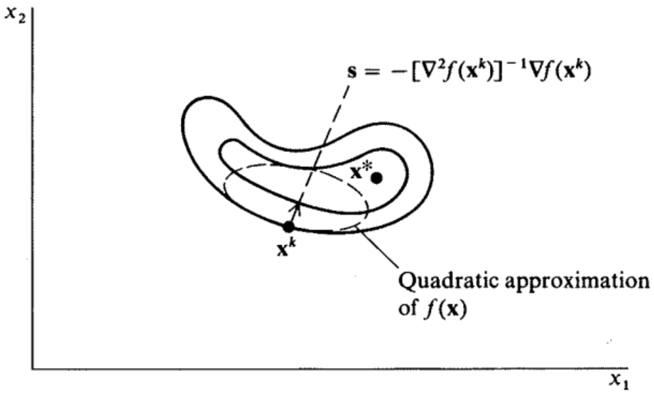
Reference: N&W Ch.3-3.1, 3.4, 3.5

Steepest descent direction vs Newton direction from objective function approximation

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

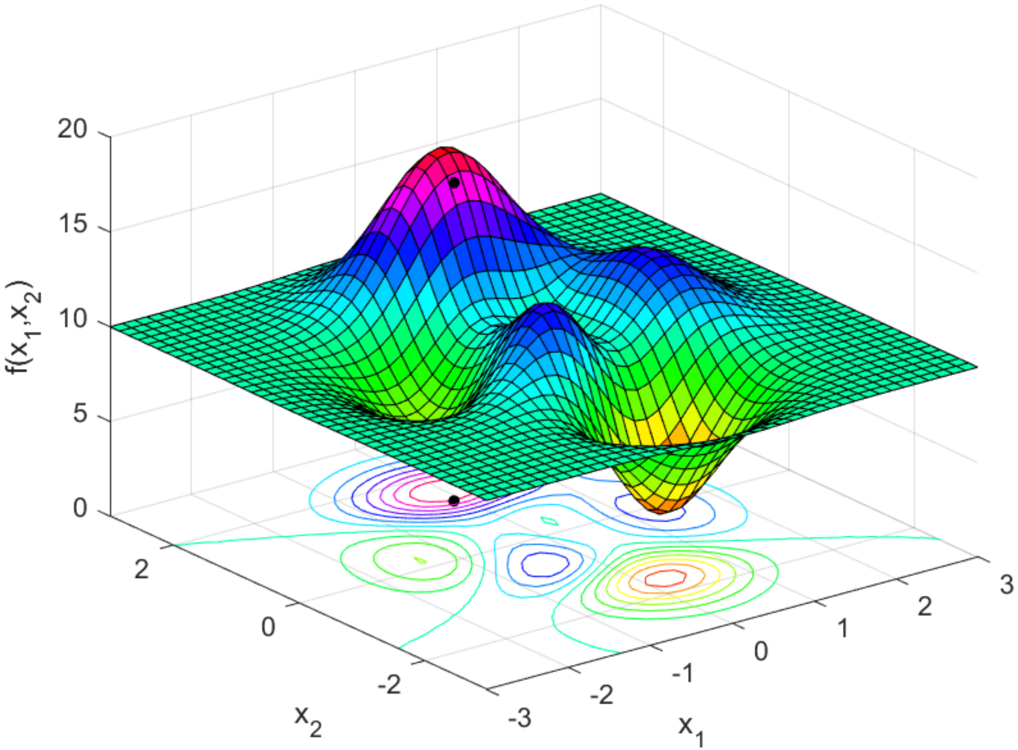


(a) Steepest descent: first-order approximation (linearization) of $f(\mathbf{x})$ at $\mathbf{x}^k$

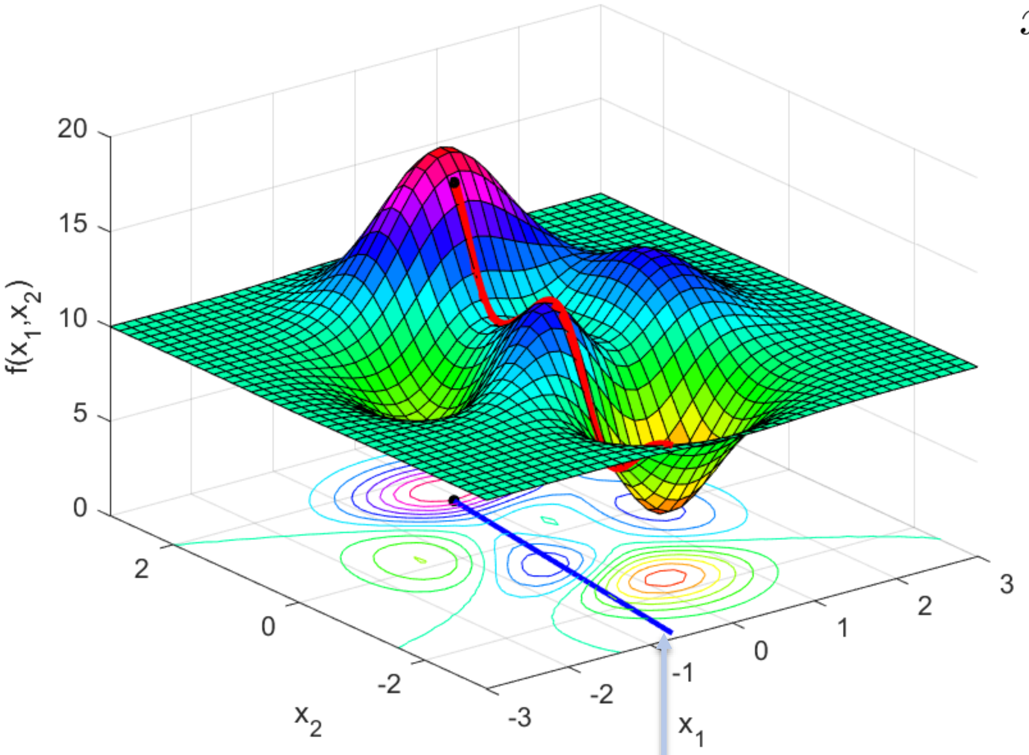


(b) Newton's method: second-order (quadratic) approximation of $f(\mathbf{x})$ at $\mathbf{x}^k$

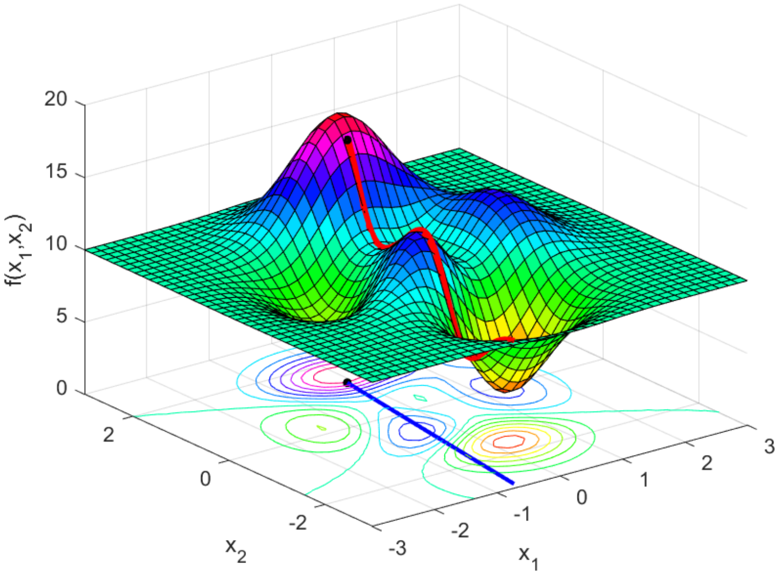
$$x_0 = (-0.1, 1.3)^\top$$



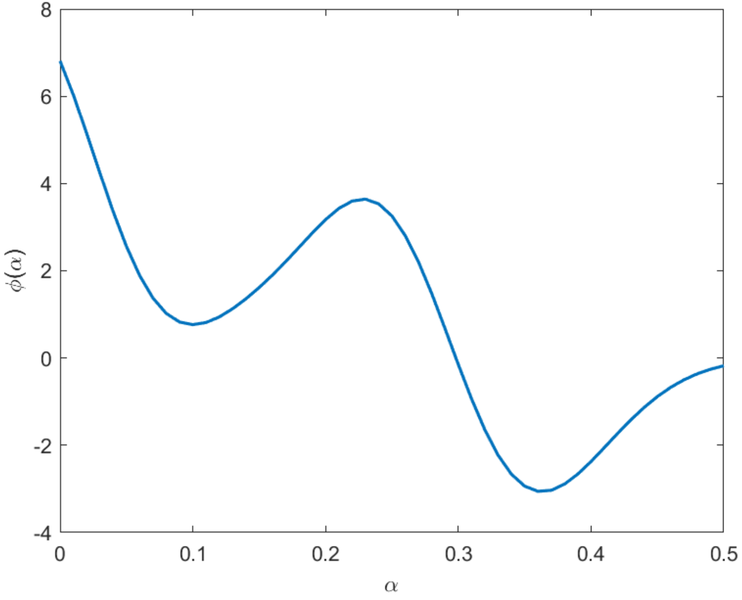
$$x_0 = (-0.1, 1.3)^\top$$



Steepest descent direction from x_0

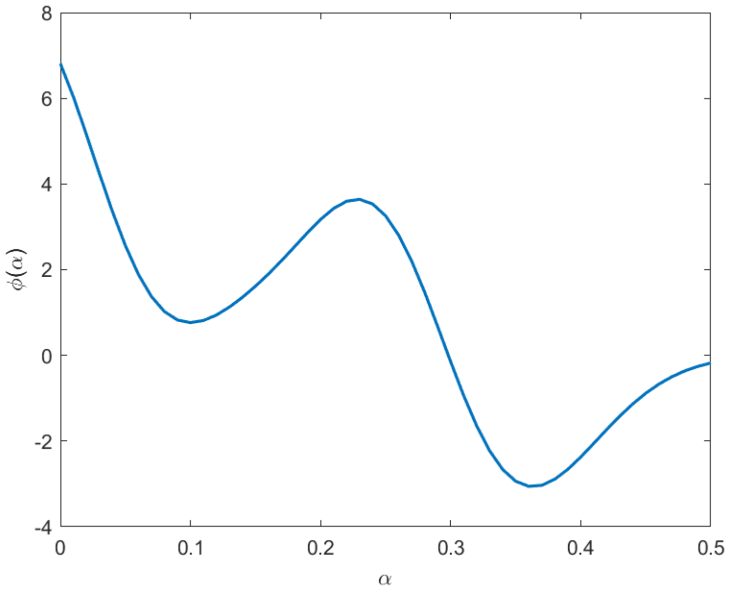


20
15
10
5
0
f(x1,x2)
2
0
-2
x2
-3
-2
-1
0
1
2
3
x1

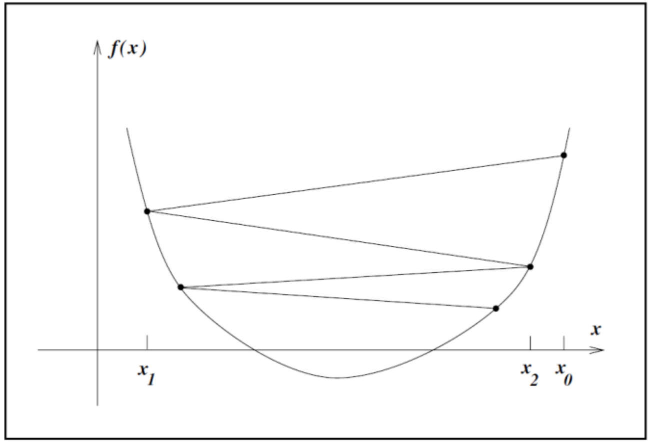


8
6
4
2
0
-2
-4
φ(α)
0
0.1
0.2
0.3
0.4
0.5
α

Exact linesearch



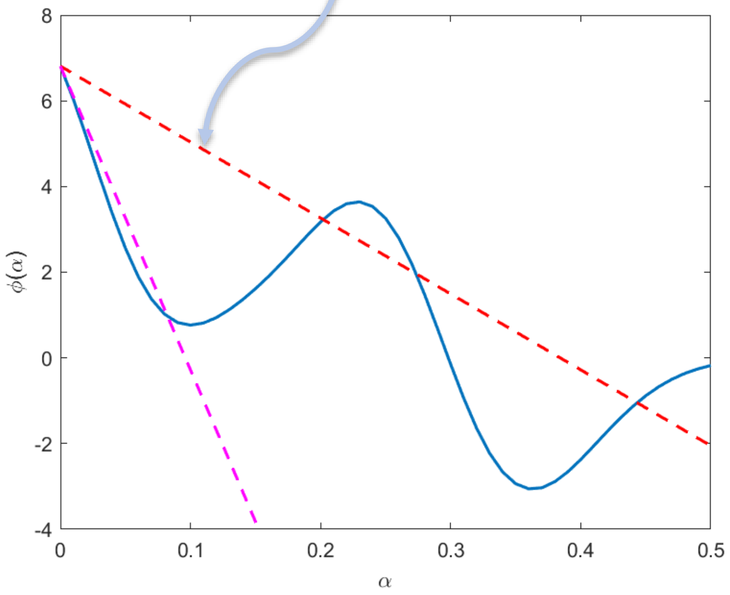
Why sufficient decrease?



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

Condition 1: Sufficient decrease (*Armijo*)

$$l(\alpha) = \phi(0) + c_1 \alpha \phi'(0), \quad c_1 = 0.25$$



Sufficient decrease

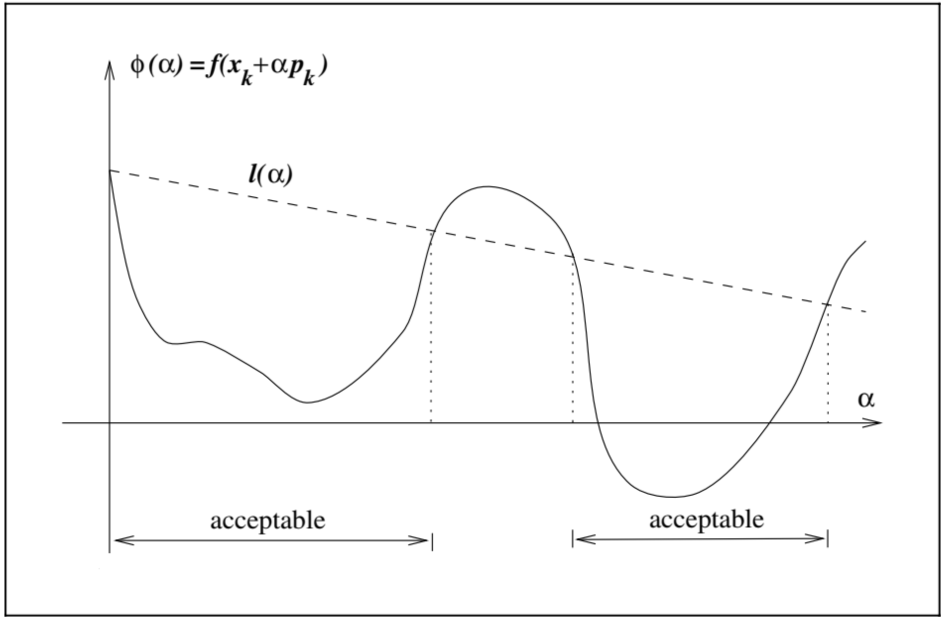
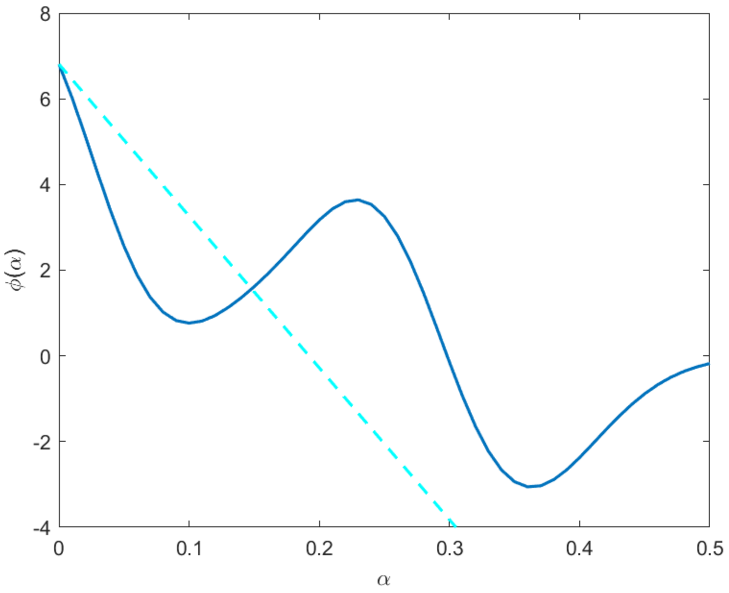


Figure 3.3 Sufficient decrease condition.

Condition 2: Curvature condition

$$l(\alpha) = \phi(0) + c_2 \alpha \phi'(0), \quad c_2 = 0.5$$



Curvature condition

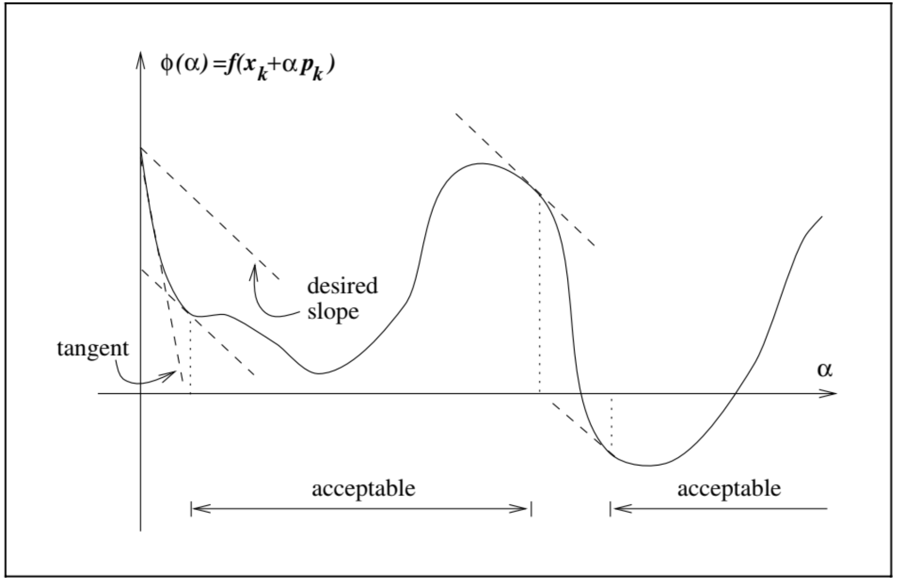


Figure 3.4 The curvature condition.

Wolfe conditions

- Good step lengths should fulfill the Wolfe conditions:

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^\top p_k \qquad \text{Sufficient decrease (Armijo condition)}$$

$$\nabla f(x_k + \alpha_k p_k)^\top p_k \geq c_2 \nabla f_k^\top p_k \qquad \text{Desired slope (Curvature condition)}$$

- How do we compute such a step length?

Backtracking Line Search

Algorithm 3.1 (Backtracking Line Search).

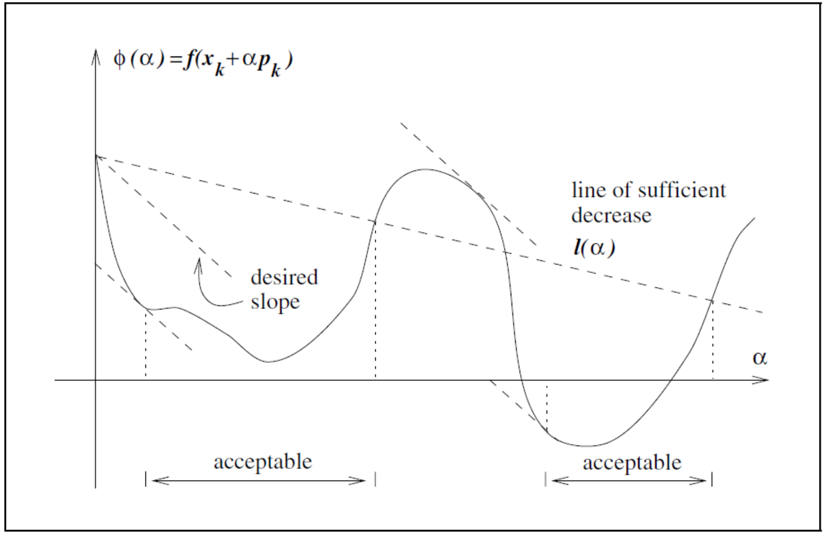
Choose $\bar{\alpha} > 0$, $\rho \in (0, 1)$, $c \in (0, 1)$; Set $\alpha \leftarrow \bar{\alpha}$;

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

$\alpha \leftarrow \rho\alpha$;

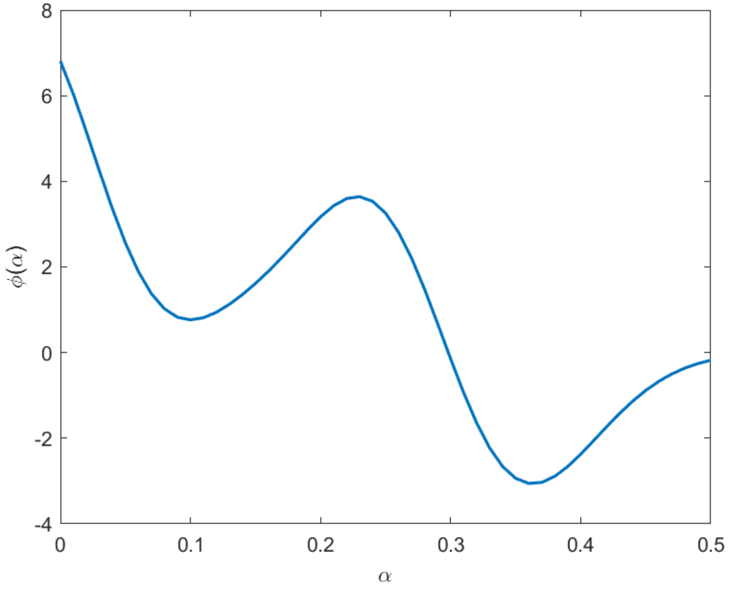
end (repeat)

Terminate with $\alpha_k = \alpha$.



Curvature condition (desired slope)
not needed since we start with long step length

Interpolation



Example: Line search for convex quadratic objective function

Newton: Hessian modification

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 (close to optimum)

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

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

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