Lecture 13: Unconstrained optimization

- Optimality conditions for unconstrained optimization
- Ingredients in a general algorithm for unconstrained optimization
  - Descent directions (steepest descent, Newton, Quasi-Newton)
  - How far to walk in descent direction (line search, trust region)
  - Termination criteria
- Scaling

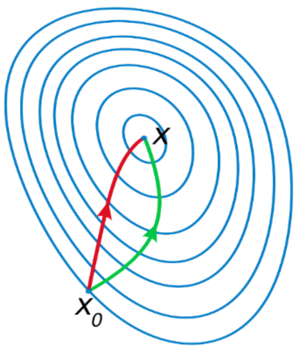
Reference: N&W Ch.2.1-2.2

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

$$\min_x f(x)$$

Line-search unconstrained optimization

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)

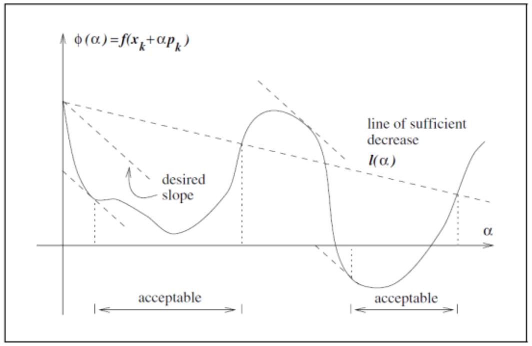


A comparison of steepest descent and Newton's method. Newton's method uses curvature information to take a more direct route. (wikipedia.org)

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

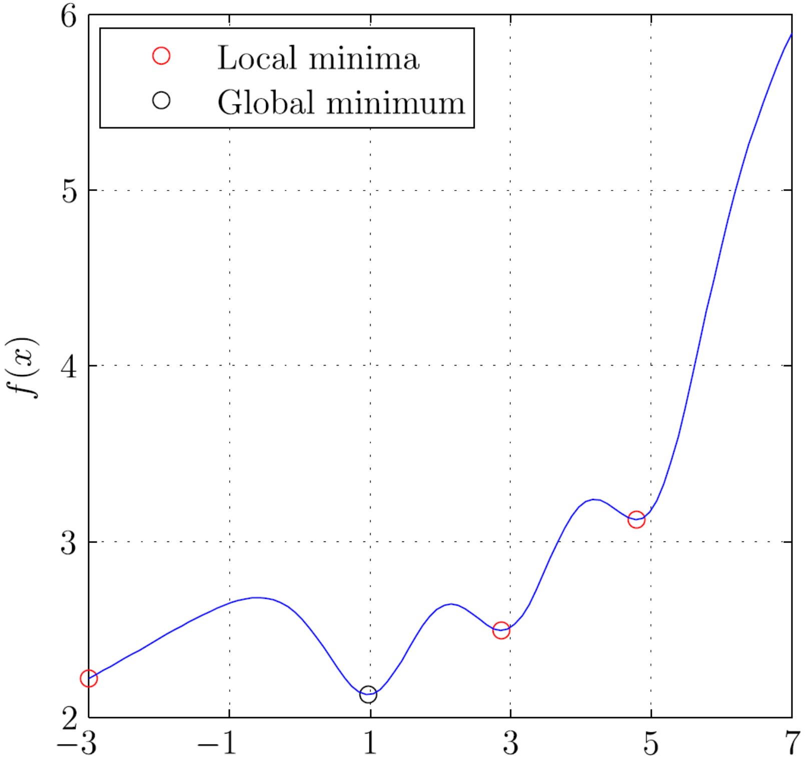
Step length (Wolfe):

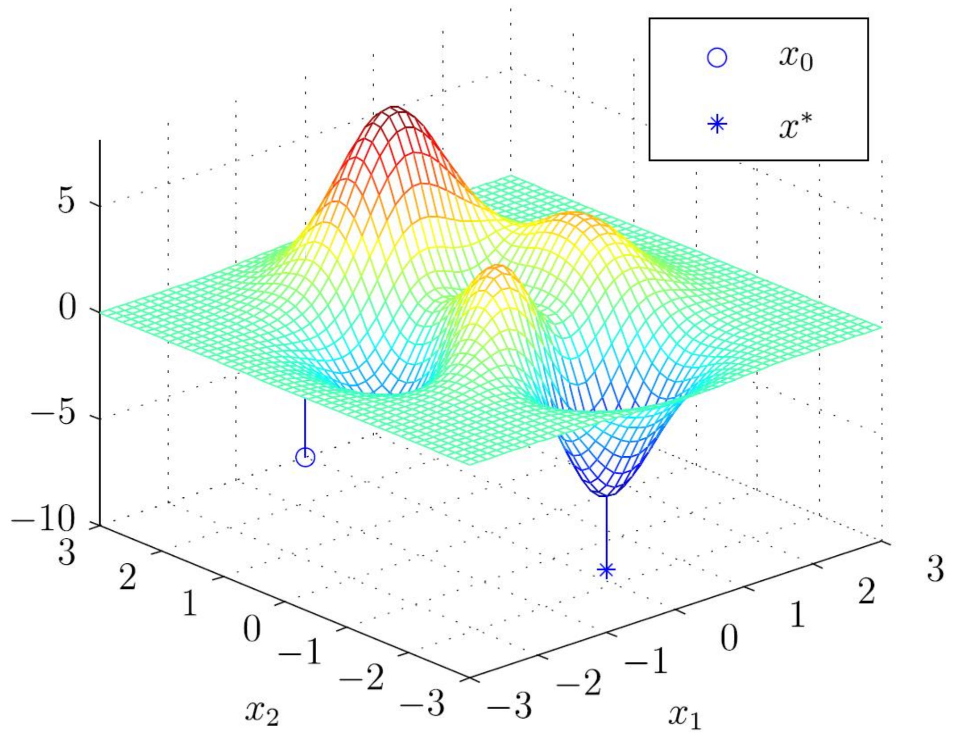


How to calculate derivatives (Ch. 8)?

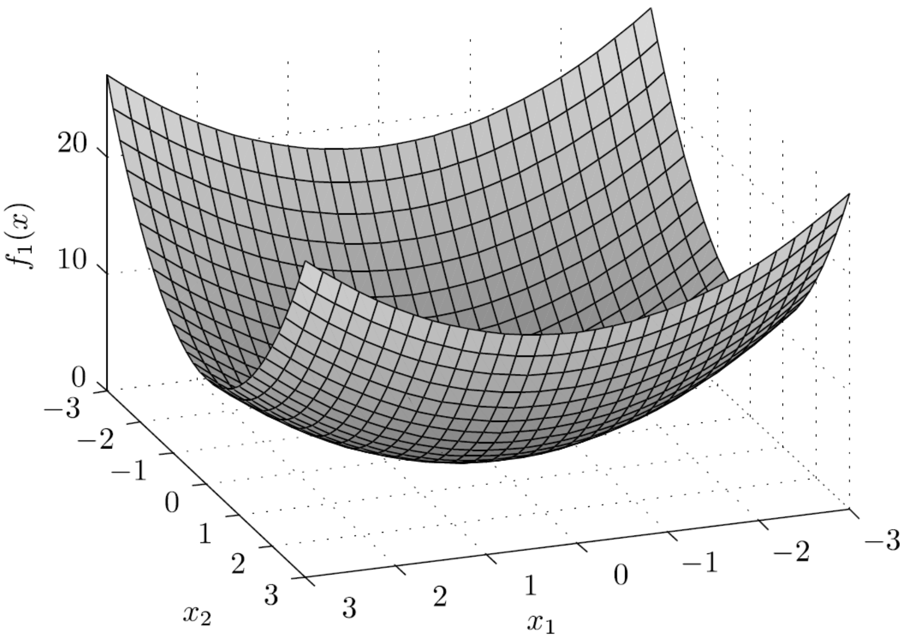
How many iterations? (Convergence rates)

Local and global minimizers

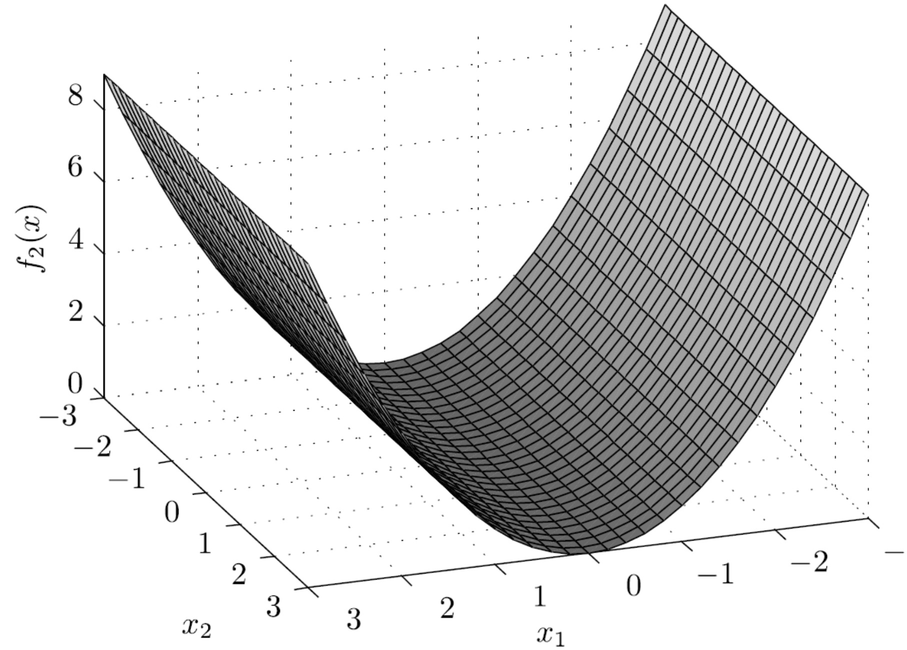




(Strict and non-strict optimizers)



$x^* = 0$ is a *strict minimizer.*



$x_1^* = 0$ is a *non-strict minimizer.*

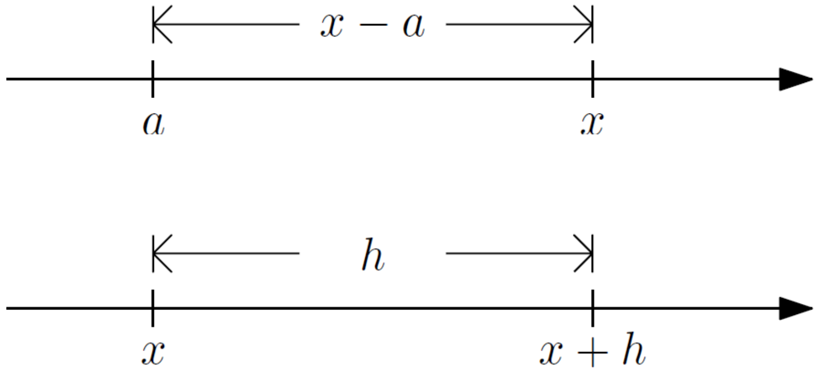
Taylor expansions

- From Calculus?

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2}f''(a) + \cdots$$

- In this course:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \cdots$$



Taylor's theorem

$$f : \mathbb{R}^n \to \mathbb{R},\ p \in \mathbb{R}^n$$

- First order: If *f* is continuously differentiable,

$$f(x+p) = f(x) + \nabla f(x+tp)^\top p, \quad \text{for some } t \in (0,1)$$

- Second order: If *f* is twice continuously differentiable

$$f(x+p) = f(x) + \nabla f(x)^\top p + \frac{1}{2} p^\top \nabla^2 f(x+tp)^\top p, \quad \text{for some } t \in (0,1)$$

Quadratic approximation to objective function

$$f(x_k + p) \approx m_k(p) = f(x_k) + p^\top \nabla f(x_k) + \frac{1}{2} p^\top \nabla^2 f(x_k) p$$

Minimize approximation: $\nabla_p m_k(p) = 0 \Rightarrow p_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$

"Newton step": $x_{k+1} = x_k + p_k = x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$

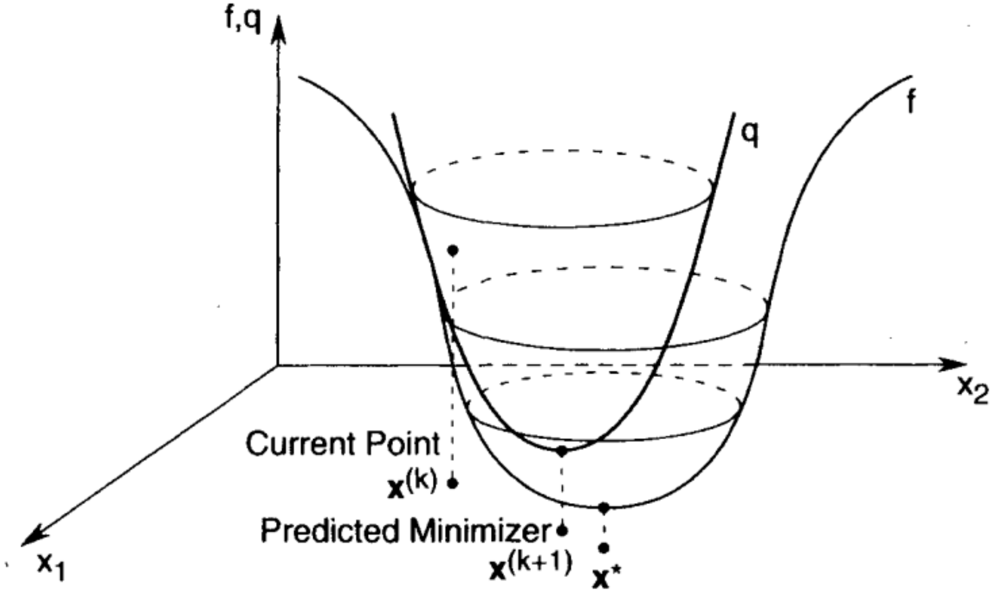
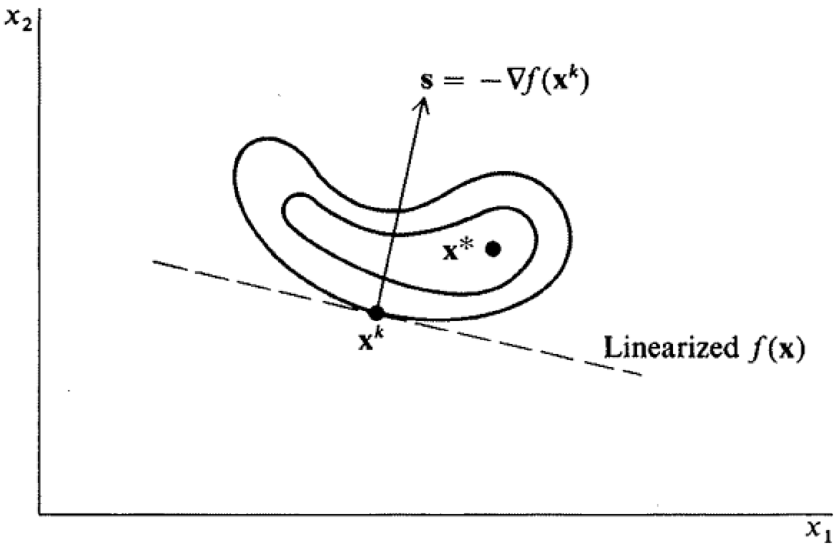


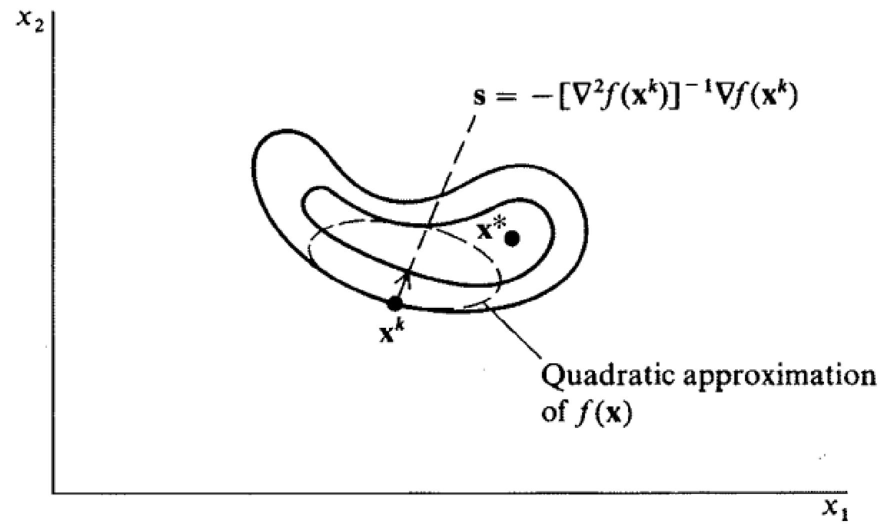
Figure 9.1 Quadratic approximation to the objective function using first and second derivatives.

Chong & Zak, "An introduction to optimization"

Steepest descent directions vs Newton directions from objective function approximations



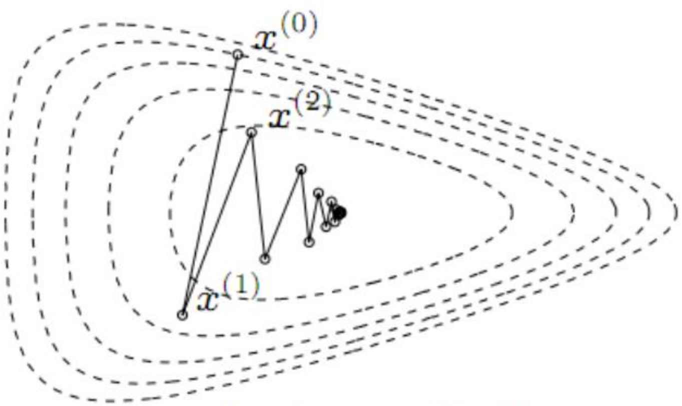
(a) Steepest descent: first-order approximation (linearization) of $f(\mathbf{x})$ at $\mathbf{x}^k$



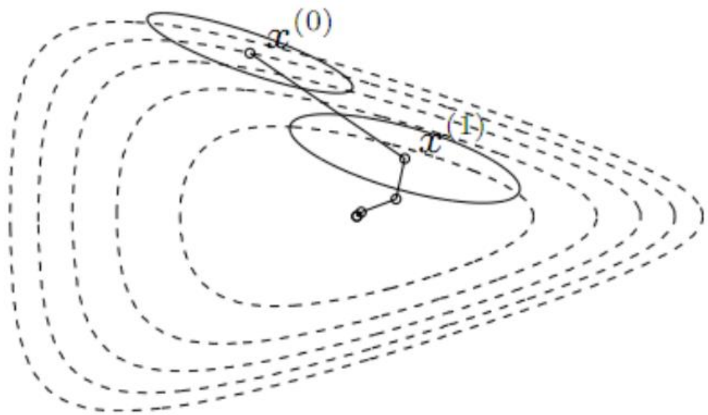
(b) Newton's method: second-order (quadratic) approximation of $f(\mathbf{x})$ at $\mathbf{x}^k$

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

Steepest descent vs Newton



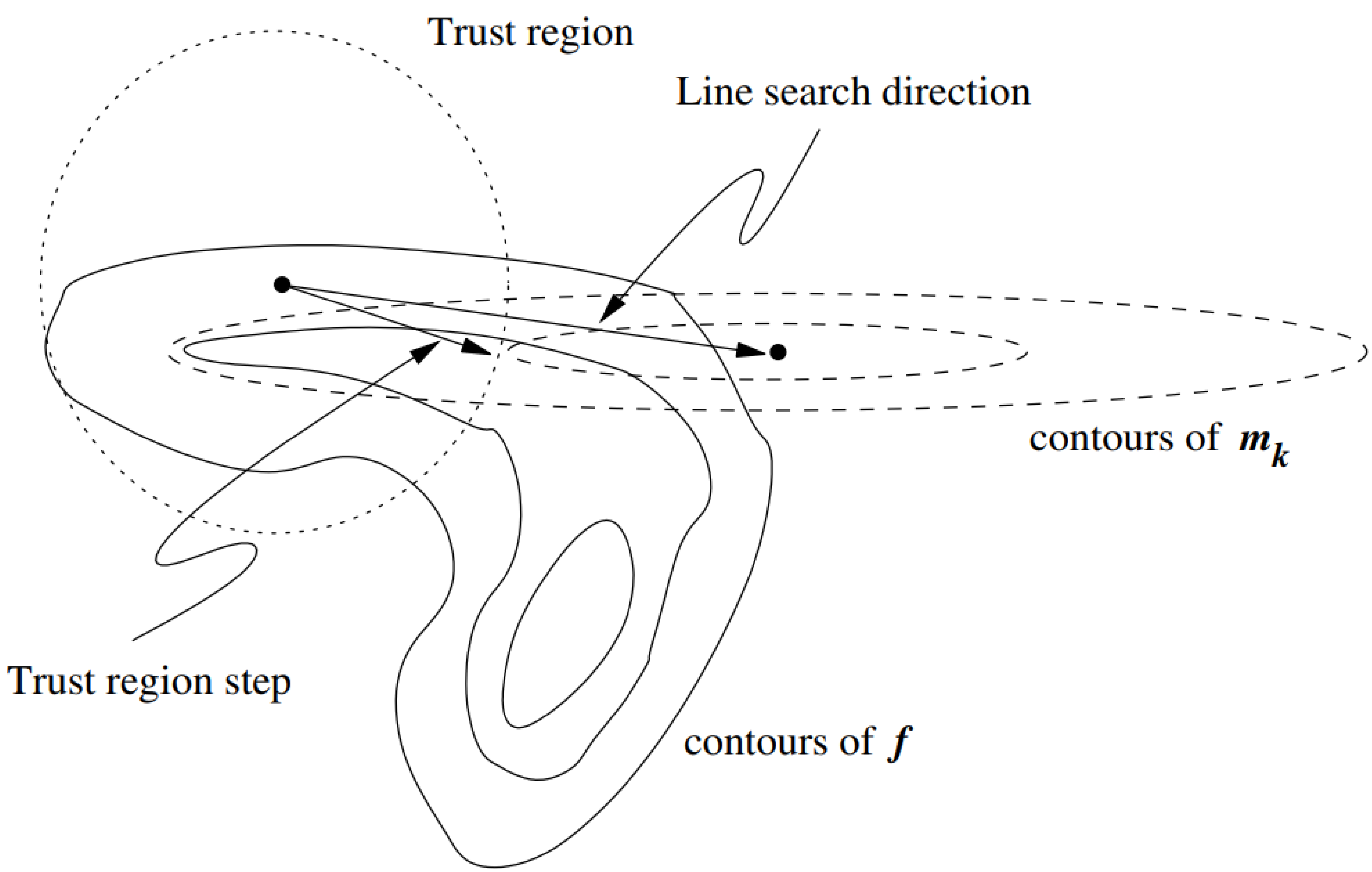
gradient descent with
backtracking line search



Newton's method with
backtracking line search

Boyd & Vanderberghe, P. Abbeel

Line search and trust region steps



Scaling, scale invariance

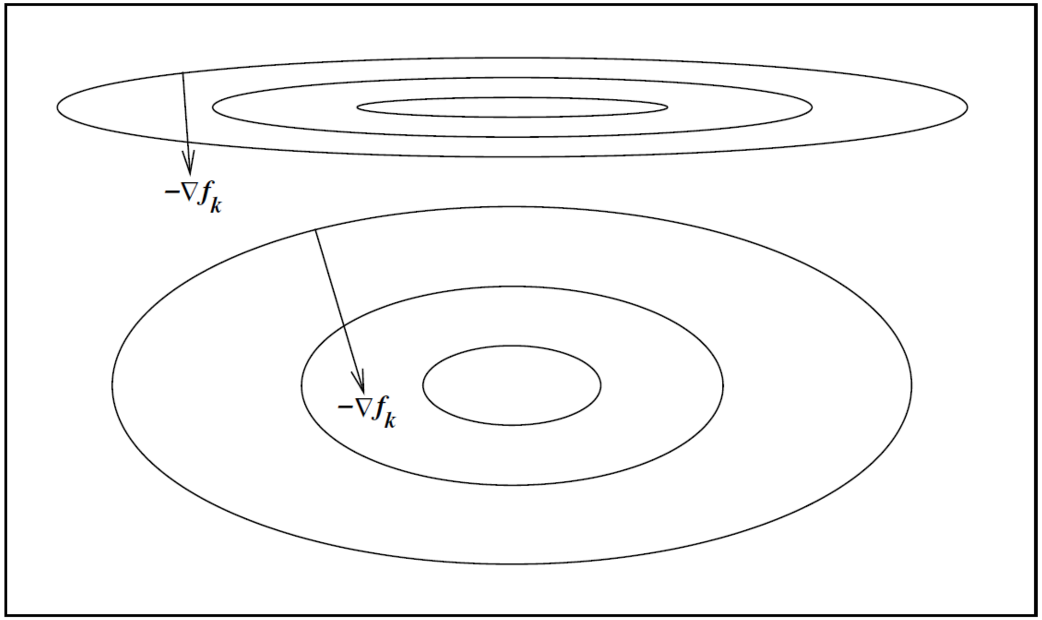


Figure 2.7 Poorly scaled and well scaled problems, and performance of the steepest descent direction.