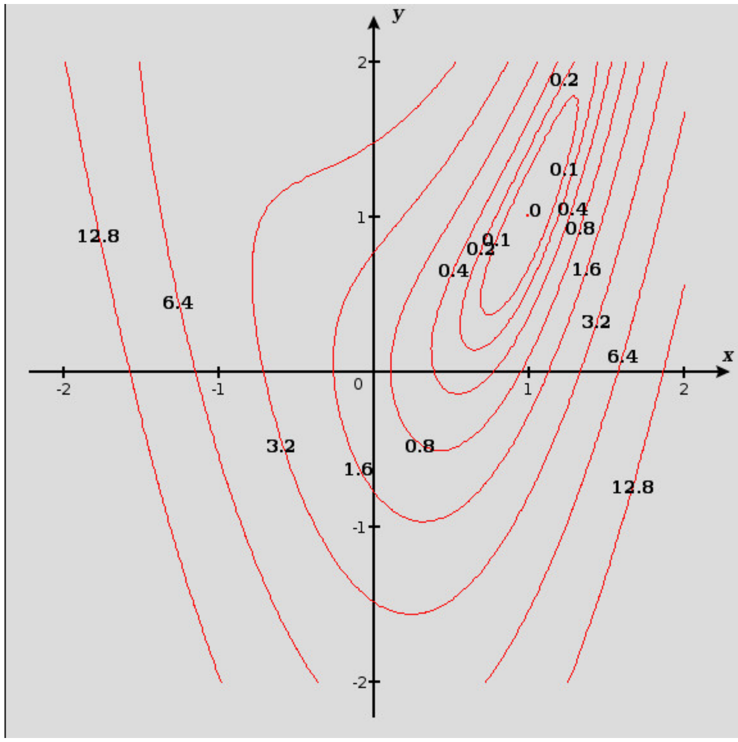
MATH529 – Fundamentals of Optimization
Unconstrained Optimization VI

Marco A. Montes de Oca

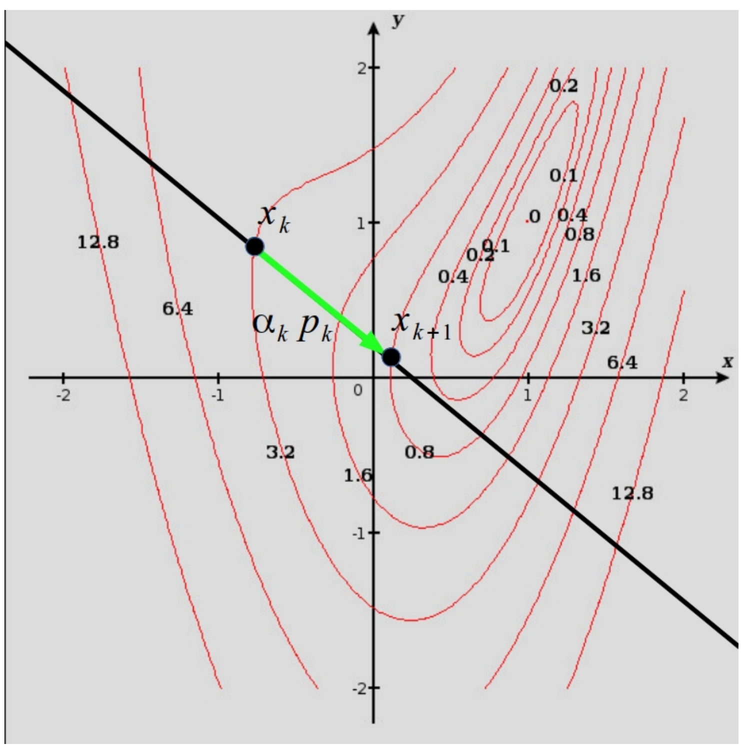
Mathematical Sciences, University of Delaware, USA

Recap

Line Search



Line Search



Line Search

Two questions:

- What should be the direction $\mathbf{p}_k$? (Q1)
- What is the best value α_k in $\mathbf{x}_k + \alpha_k \mathbf{p}_k$? (Q2)

Newton's Method Approach

Newton's method approach to answer questions Q1 and Q2:

- Create a local quadratic model m_k of f at $\mathbf{x}_k$.
- Solve the problem $\min_p m_k(\mathbf{p})$. Call the solution $\mathbf{p}_k$.
- Generate the next estimate as follows $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$, with $\alpha_k = 1$.

Newton's Method Approach

The quadratic model of f at $\mathbf{x}_k$ is defined by

$$m_k(\mathbf{p}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{p} + \frac{1}{2}\mathbf{p}^T Hf(\mathbf{x}_k)\mathbf{p}$$

and the solution to $\min_p m_k(\mathbf{p})$, $\mathbf{p}_k$, is given by

$$\mathbf{p}_k = -(Hf(\mathbf{x}_k))^{-1}\nabla f(\mathbf{x}_k)$$

$\mathbf{p}_k$ is a descent direction only if $Hf(\mathbf{x})$ is positive definite. (Proof: Homework exercise!)

Steepest Descent Approach

The approach taken by the steepest descent method to answer questions Q1 and Q2:

- At $\mathbf{x}_k$, choose $\mathbf{p}_k = -\nabla f(\mathbf{x}_k)$. (Homework exercise: Show that $\mathbf{p}_k$ is the direction in which f decreases most rapidly at $\mathbf{x}_k$.)
- Calculate a sensible value for α_k using certain principles (**Today's class**).

Steepest Descent Approach

The most natural answer to Q2 using the steepest descent method is to choose α_k such that $\mathbf{x}_{k+1}$ is a minimizer (a local minimizer at least) along the line $\mathbf{x}_k + \alpha_k \mathbf{p}_k$.

However, we know how hard it is to find minimizers, even of 1-dimensional functions.

Calculating the Step Length

An approach:

Let $\phi(\alpha_k) = f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)$. Using the chain rule:

$\phi'(\alpha) = \nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)^T \mathbf{p}_k$. If we want to find the critical points of α, we need to solve the equation

$$\nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)^T \mathbf{p}_k = 0$$

for α.
Since this equation represents a dot product, the optimal value for α_k is the one that makes $\nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)$ perpendicular to $\mathbf{p}_k$.

To ensure that α_k is a global minimizer, we need that $\phi''(\alpha_k) = \mathbf{p}_k^T Hf(\mathbf{x}_k)\mathbf{p}_k > 0$, which means that we need that the Hessian of f be positive definite. (**Make sure you verify this!**)

Calculating the Step Length

The exact steepest descent method is difficult to implement, and inefficient. Thus, we are content with decreasing the function at each step *sufficiently*.

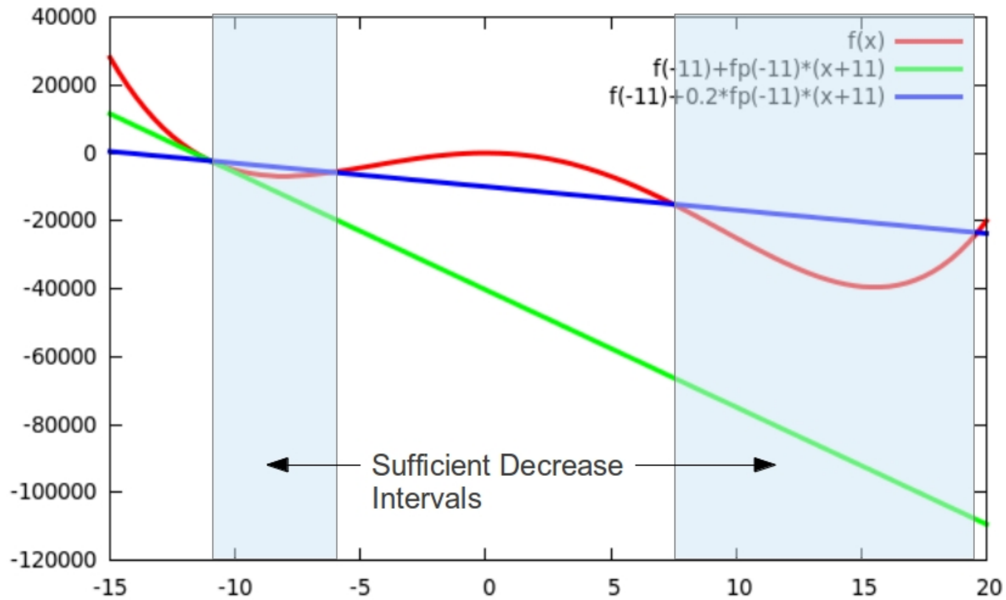
Desired conditions for step length selection

- Decrease: $f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) < f(\mathbf{x}_k)$
- Wolfe conditions:
  - Sufficient Decrease (Armijo condition): $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k$

Typically, c_1 is small. For example $c_1 = 10^{-4}$.

Desired conditions for step length selection

Armijo Condition:

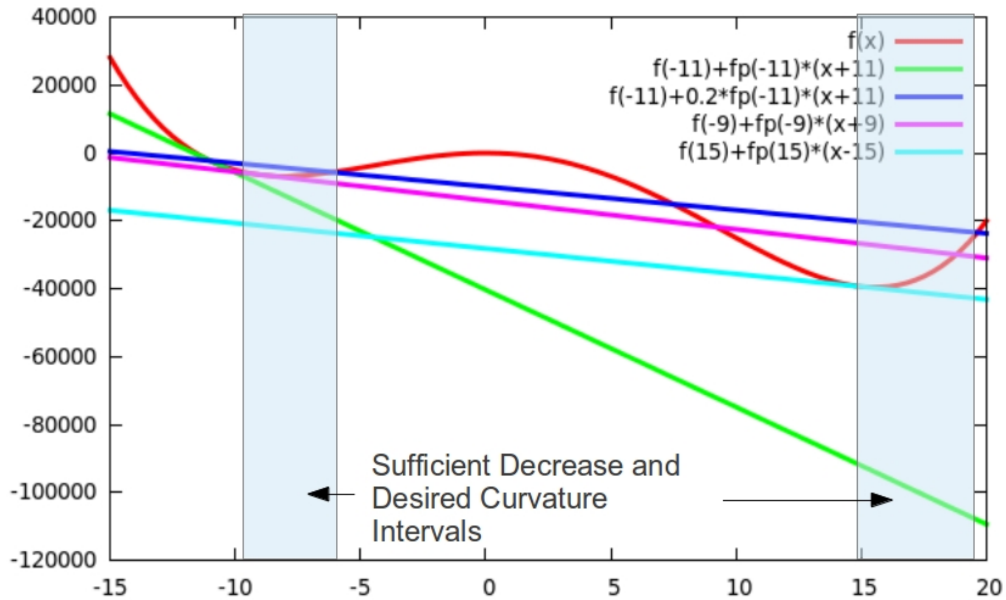


Desired conditions for step length selection

- Decrease: $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$
- Wolfe conditions:
  - Sufficient Decrease (Armijo condition): $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k$, $c_1 \in (0, 1)$ but $c_1 \approx 10^{-4}$.
  - Curvature condition: $\nabla f(\mathbf{x}_{k+1}) \cdot \mathbf{p}_k \geq c_2 \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k$, $c_2 \in (c_1, 1)$.

Desired conditions for step length selection

Desired Curvature Condition:



Desired conditions for step length selection

- Decrease: $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$
- Wolfe conditions:
  - Sufficient Decrease (Armijo condition): $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k$, $c_1 \in (0, 1)$ but $c_1 \approx 10^{-4}$.
  - Curvature condition: $\nabla f(\mathbf{x}_{k+1})^T \mathbf{p}_k \geq c_2 \nabla f(\mathbf{x}_k)^T \mathbf{p}_k$, $c_2 \in (c_1, 1)$. Typically, $c_2 \approx 0.9$ for Newton's method.

Desired conditions for step length selection

How do we implement Wolfe conditions?

Backtracking algorithm:

- Let $\alpha > 0$, $0 < \rho < 1$, $0 < c < 1$.
- Set $\alpha_k = \alpha$.
- While $f(\mathbf{x}_k + \alpha_i \mathbf{p}_k) > f(\mathbf{x}_k) + c\alpha_k \nabla f(\mathbf{x}_k)^T \mathbf{p}_k$
  - $\alpha_k = \rho\alpha_k$
- End While
- Return α_k.

Mini-Lab