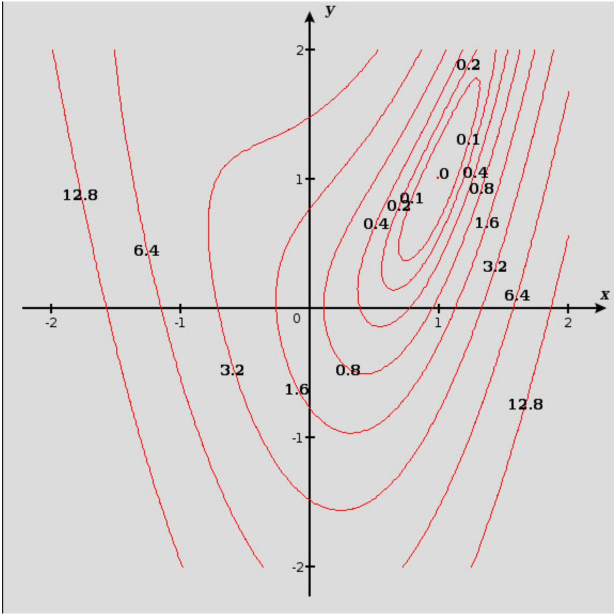
MATH529 – Fundamentals of Optimization

Line Search Algorithms

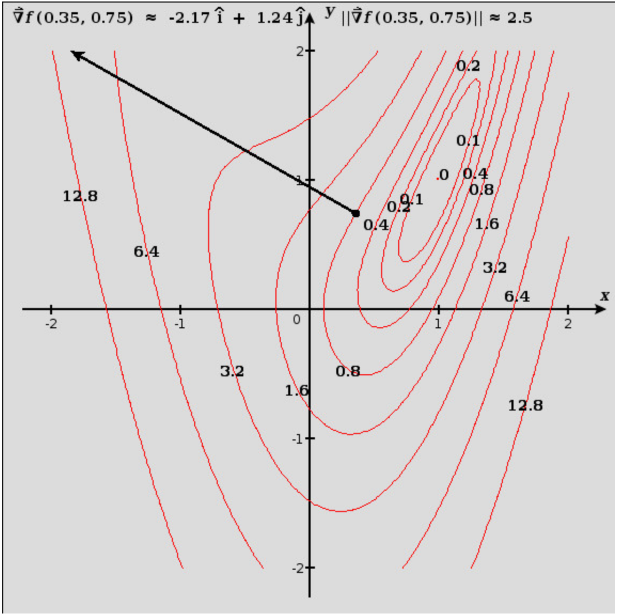
Marco A. Montes de Oca

Mathematical Sciences, University of Delaware, USA

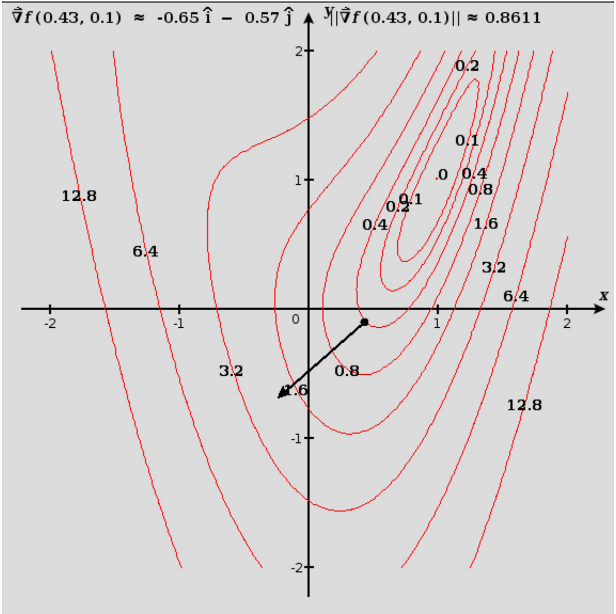
Contour plot of $f(\mathbf{x}) = (x_2 - x_1^2)^2 + (1 - x_1)^2$



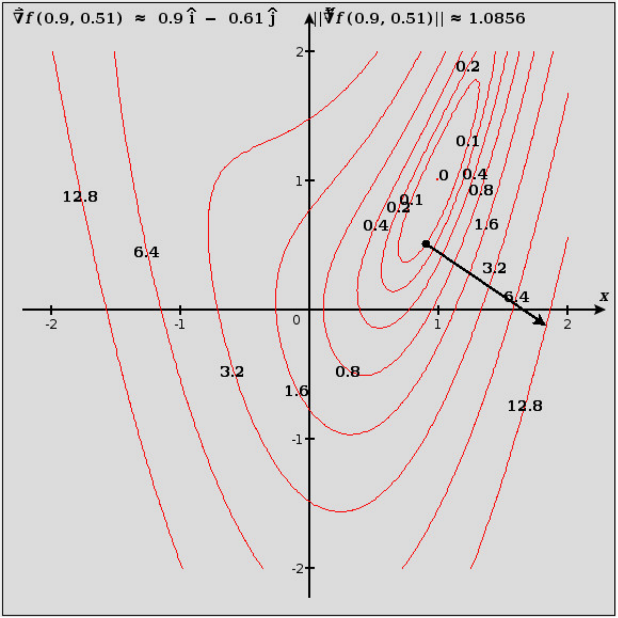
Contour plot of $f(\mathbf{x}) = (x_2 - x_1^2)^2 + (1 - x_1)^2$



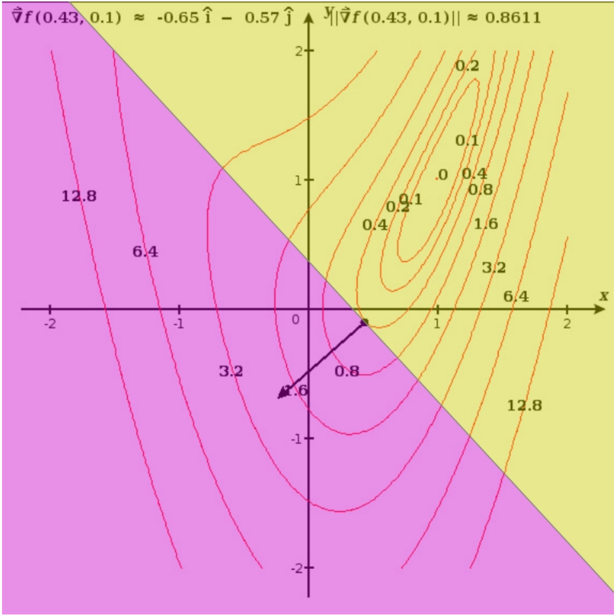
Contour plot of $f(\mathbf{x}) = (x_2 - x_1^2)^2 + (1 - x_1)^2$



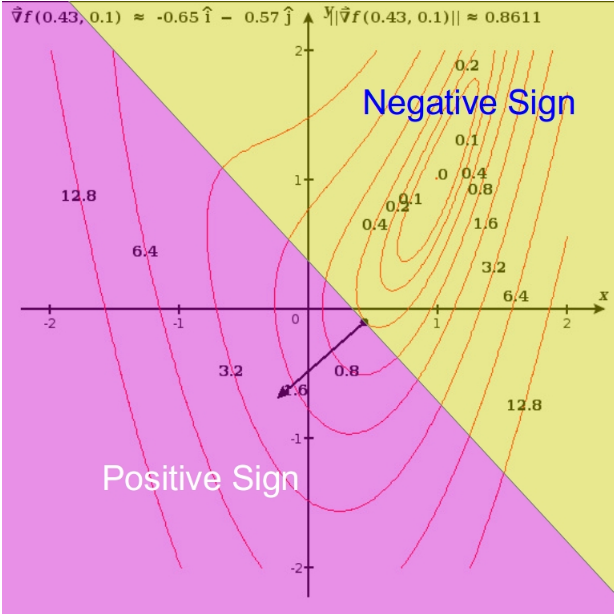
Contour plot of $f(\mathbf{x}) = (x_2 - x_1^2)^2 + (1 - x_1)^2$



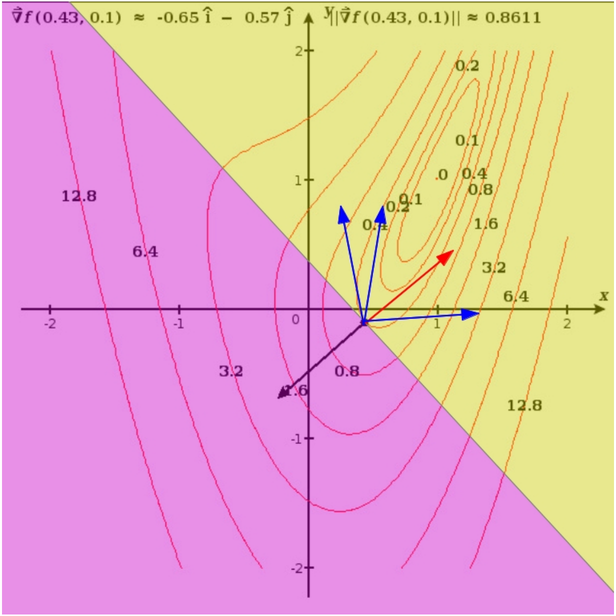
Sign of Directional Derivatives



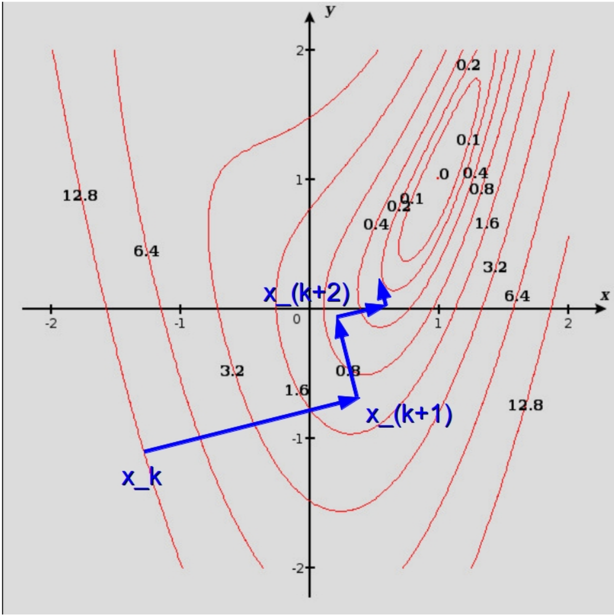
Sign of Directional Derivatives



Descent Search Directions



Line Search Algorithms



Steepest Descent Direction

Let's $\mathbf{p}_k$ denote the search direction chosen at step k, that is, $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$. Then, according to Taylor's formula, we have:

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) =$$
$$f(\mathbf{x}_k) + \alpha_k \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k + \frac{\alpha_k^2}{2} \mathbf{p} \cdot Hf(\mathbf{x}_k + t\mathbf{p}_k)\mathbf{p}_k$$

for some $t \in (0, \alpha_k)$.

Then,

$$\frac{\mathrm{d} f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)}{\mathrm{d}\alpha_k} = \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k + \alpha_k \mathbf{p} \cdot Hf(\mathbf{x}_k + t\mathbf{p}_k)\mathbf{p}_k.$$

When $\alpha_k = 0$, that is, the rate of change of f at $\mathbf{x}_k$ in the direction of $\mathbf{p}_k$ is: $\nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k = ||\nabla f(\mathbf{x}_k)||||\mathbf{p}_k|| \cos\theta$, where θ is the angle between $\mathbf{p}_k$ and $\nabla f(\mathbf{x}_k)$.

Thus, the direction of most rapid decrease at $\mathbf{x}_k$ is given by:$-\frac{\nabla f(\mathbf{x}_k)}{||\nabla f(\mathbf{x}_k)||}$.

Newton's Direction

Let's instead choose $\alpha_k \mathbf{p}_k$ such that $f(\mathbf{x}_k + \alpha \mathbf{p}_k)$ is a minimum. We may do this by approximating $f(\mathbf{x}_k)$ with

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

Then, $\mathbf{x}_k + \alpha_k \mathbf{p}_k$ must satisfy:

$$\nabla m_k(\mathbf{x}_k + \alpha_k \mathbf{p}_k) = \alpha_k \nabla f(\mathbf{x}_k) + \alpha_k^2 Hf(\mathbf{x}_k) \mathbf{p}_k = \mathbf{0}.$$

So, $\alpha_k^2 Hf(\mathbf{x}_k) \mathbf{p}_k = -\alpha_k \nabla f(\mathbf{x}_k)$. Therefore:

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

Basic search direction selection

Steepest descent direction:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \frac{\nabla f(\mathbf{x}_k)}{||\nabla f(\mathbf{x}_k)||}, \text{ or simply } \mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)$$

Newton's direction:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (Hf(\mathbf{x}_k))^{-1} \nabla f(\mathbf{x}_k),$$

Micro-Lab

To Note

Steepest Descent Direction:

- We need a smart choice of the step length because the length of the gradient can change very rapidly.
- It can be slow in comparison with other methods.

Newton's Direction:

- The Hessian must be positive definite at every step in order to obtain a descent direction.
- The Hessian may be singular.

Modifications

Steepest Descent Direction:

- Strategies for modifying the step length
- It may still be slow.

Newton's Direction:

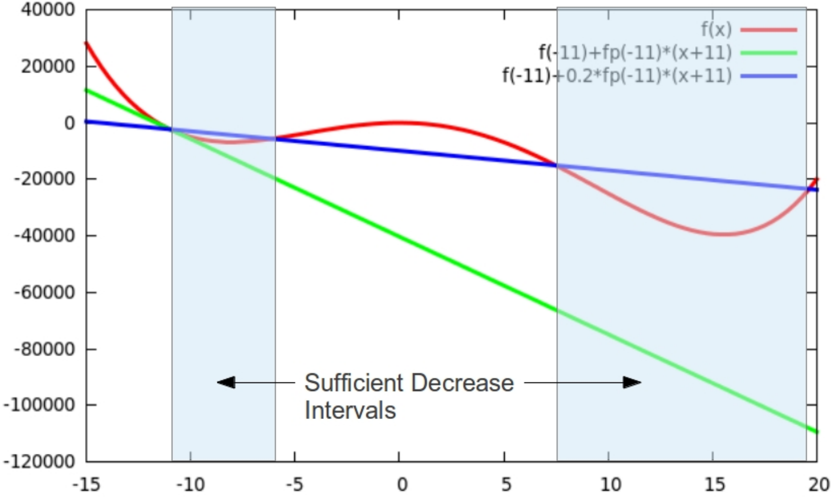
- Quasi-Newton methods: Instead of using the Hessian, a symmetric, non-singular, positive definite matrix is used.

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):
    $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k$

Desired conditions for step length selection

Armijo Condition:

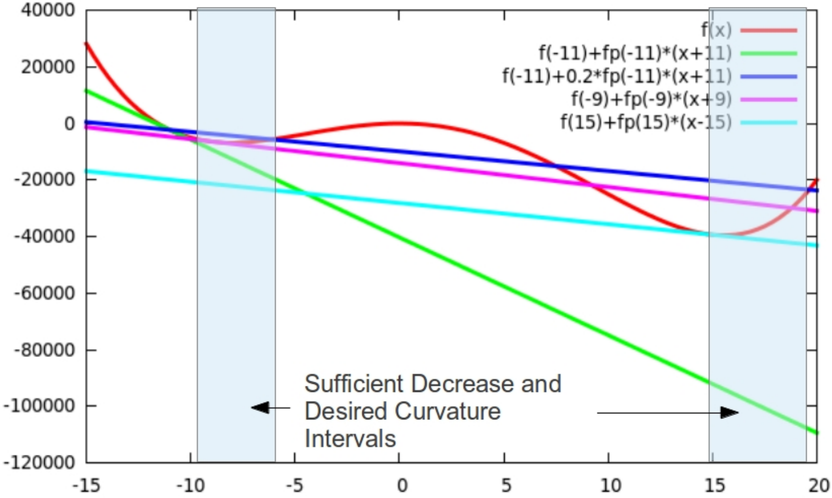


Desired conditions for step length selection

- 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: $\nabla f(\mathbf{x}_{k+1}) \cdot \mathbf{p}_k \geq c_2 \nabla f(\mathbf{x}_k) \cdot \mathbf{p}_k$, $c_2 \in (c_1, 1)$. Typically, $c_2 \approx 0.9$ for Newton or quasi-Newton methods.