



# Computational Optimization

Steepest Descent 2/8



# Most Obvious Algorithm

- The negative gradient always points downhill.
  - Head in the direction of the negative gradient.
  - Use linesearch to decided how far to go.
  - Repeat.
- 




# Steepest Descent Algorithm

- Start with  $x_0$
- For  $k = 1, \dots, K$ 
  - If  $x_k$  is optimal then stop

$$p_k = -\nabla f(x_k)$$


- Perform exact or backtracking linesearch to determine

$$x_{k+1} = x_k + \alpha_k p_k$$

# How far should we step?

Linesearch:  $x_{k+1} = x_k + \alpha_k p_k$

- fixed  $\alpha_k := c > 0$  a constant
  - Decreasing  $\alpha_k \rightarrow 0, \sum_{i=1}^{\infty} \alpha_k = \infty$
  - Approximate conditions
    - Armijo
    - Wolfe
    - Goldstein
- 



# Key Idea: Sufficient Decrease

The step can't go to 0 unless the gradient goes to 0.

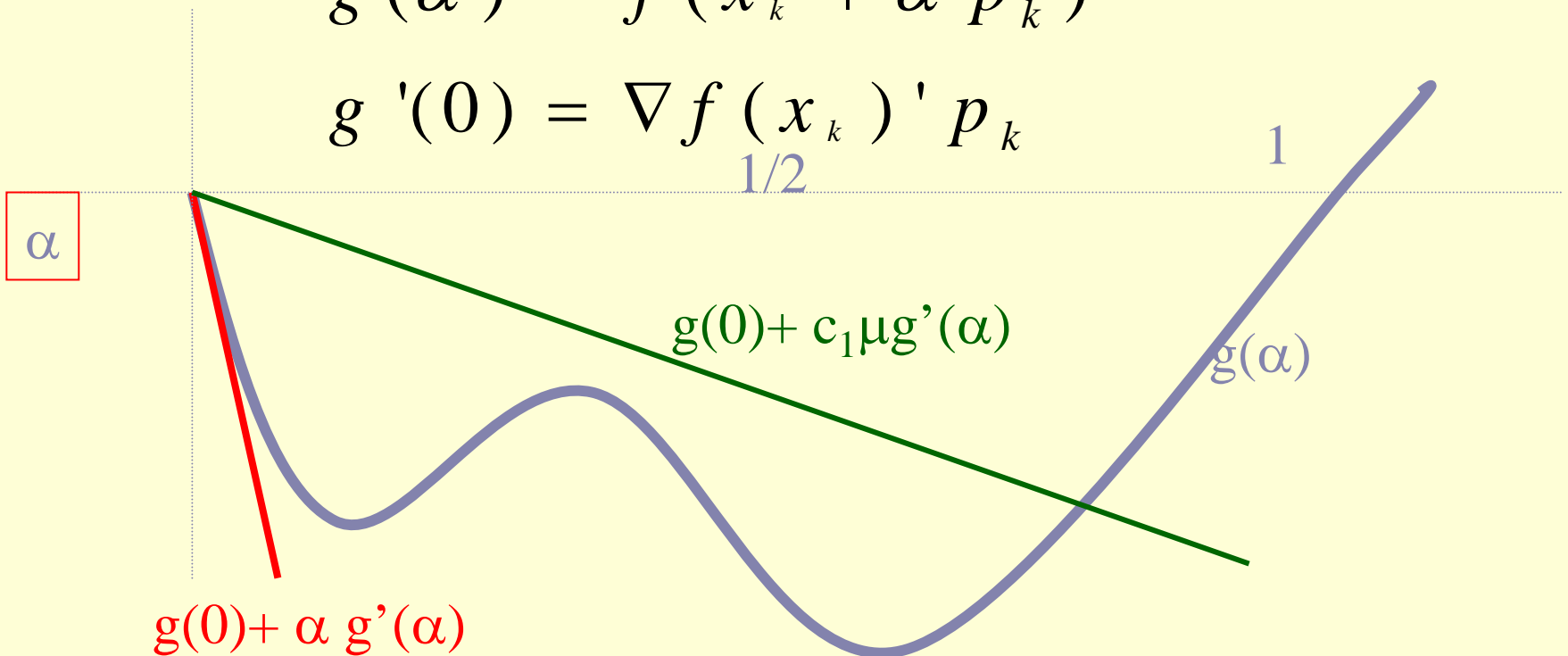


# Armijo Condition

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f(x_k)' p_k$$

$$g(\alpha) = f(x_k + \alpha p_k)$$

$$g'(0) = \nabla f(x_k)' p_k$$





# Curvature Condition

$$\nabla f(x_k + \alpha p_k)' p \geq c_2 \nabla f(x_k)' p_k$$

Make sure the step uses up most of the available decrease





# Wolfe Condition

For  $0 < c_1 < c_2 < 1$

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f(x_k)' p_k$$

$$\nabla f(x_k + \alpha p_k)' p_k \geq c_2 \nabla f(x_k)' p_k$$

Solution exists for any descent direction if  $f$  is bounded below on the linesearch.

(Lemma 3.1)



# Backtracking Search

- Key point: Stepsize cannot be allowed to go to 0 unless gradient is going to 0. Must have sufficient decrease.
- Fix  $\delta > 0$   $\beta \in (0, 1)$   $\mu \in (0, 1)$

$$\alpha_k = \delta \max(\beta^0, \beta^1, \beta^2, \dots)$$

which satisfies

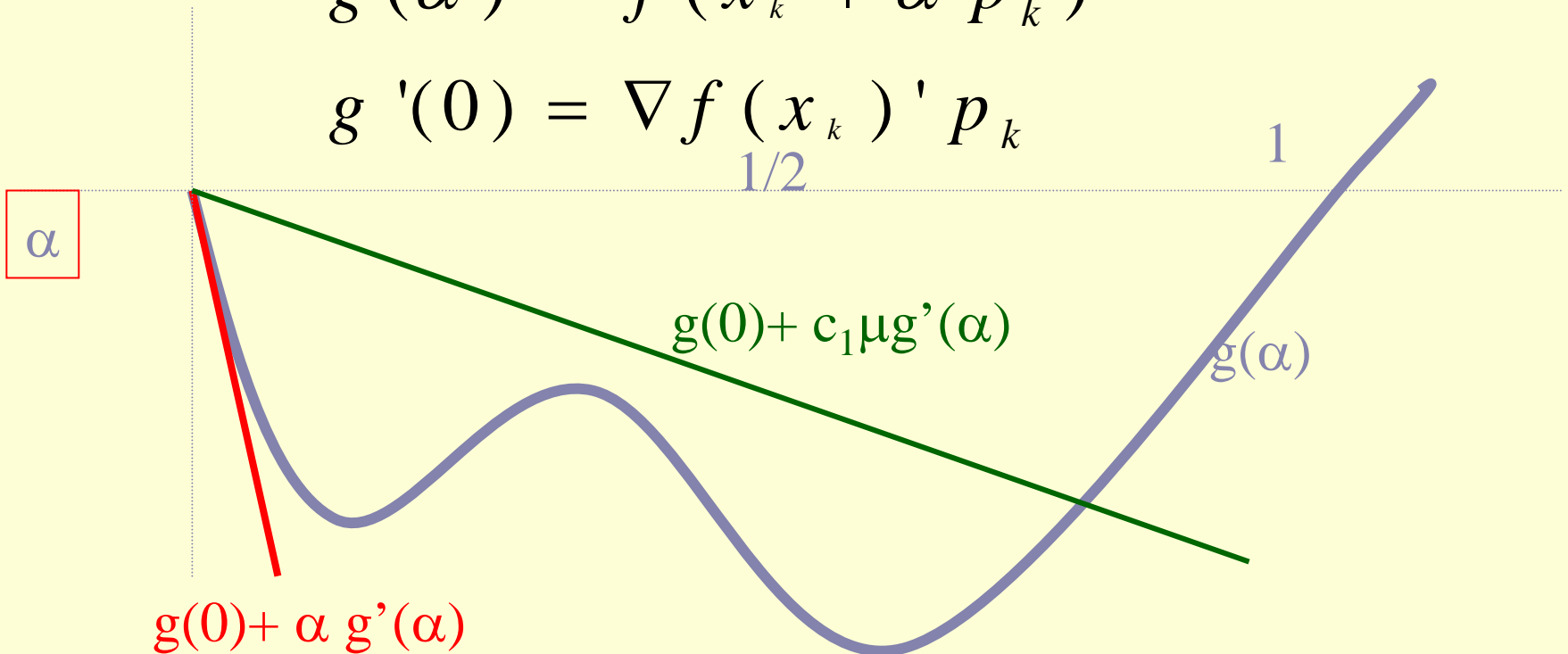
$$f(x_k + \alpha_k p_k) \leq f(x_k) + \mu \alpha_k \nabla f(x_k)' p_k$$

# Armijo Condition

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f(x_k)' p_k$$

$$g(\alpha) = f(x_k + \alpha p_k)$$


$$g'(0) = \nabla f(x_k)' p_k$$





# Steepest Descent Algorithm

- Start with  $x_0$
- For  $k = 1, \dots, K$ 
  - If  $x_k$  is optimal then stop
  - $$p_k = -\nabla f(x_k)$$
  - Perform backtracking linesearch to determine

$$x_{k+1} = x_k + \alpha_k p_k$$


# Convergence Analysis

- We usually analyze quadratic problem

$$\min f(y) = \frac{1}{2}y'Qy - c \quad \text{where } \min f(y) = 0$$

- Equivalent to more general problems since let  $x^*$  solve  $\min f(x) = \frac{1}{2}x'Qx - bx$   
Let  $y = x - x^*$  so  $x = y + x^*$  then

$$\begin{aligned} f(x) &= \frac{1}{2}(y+x^*)'Q(y+x^*) - b'(y+x^*) \\ &= \frac{1}{2}y'Qy + \frac{1}{2}x^{*'}Qx^* - y'Qx^* - b'y - b'x^* \\ &= \frac{1}{2}y'Qy + -y'(Qx^* - b) + \frac{1}{2}x^{*'}Qx^* - b'x^* \\ &= \frac{1}{2}y'Qy - f(x^*) \end{aligned}$$

# Convergence Rate

- Consider case of constant stepsize

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) = [I - \alpha_k Q] x_k \text{ for our quadratic function}$$

- Square both sides

$$\begin{aligned} \|x_{k+1}\|^2 &= x_k' [I - \alpha_k Q]^2 x_k \\ &\leq \max \text{ eigenvalues} \left( [I - \alpha_k Q]^2 \right) \|x_k\|^2 \end{aligned}$$

and we know the eigenvalues are

$$(1 - \alpha_k \lambda_k)^2 \text{ where } \lambda_1 \leq \dots \leq \lambda_n \text{ are the eigs of } Q$$




# Convergence Rate...

Convergence rate based on

$$\frac{\mathit{error}_{k+1}}{\mathit{error}_k} = \frac{\|x_{k+1} - 0\|}{\|x_k - 0\|} \leq \max(|1 - \alpha_k \lambda_1|, |1 - \alpha_k \lambda_n|)$$

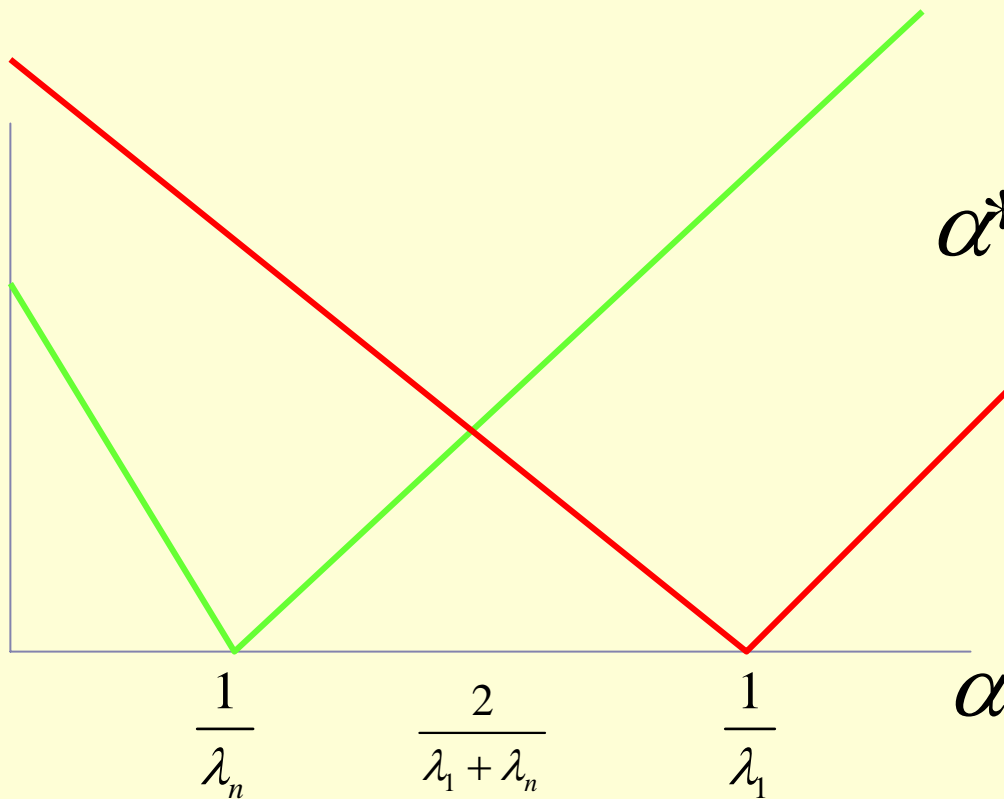
$$\text{since } \frac{\|x_{k+1}\|^2}{\|x_k\|^2} \leq \max\left((1 - \alpha_k \lambda_1)^2, (1 - \alpha_k \lambda_n)^2\right)$$

So we want this ratio to be small as possible



# Best Step

Minimizes  $\max(|1 - \alpha_k \lambda_1|, |1 - \alpha_k \lambda_n|) = \frac{\lambda_n - \lambda_1}{\lambda_1 + \lambda_n}$



$$\alpha^* = \frac{2}{\lambda_1 + \lambda_n}$$

# Condition number determines convergence rate

- Linear convergence since

$$\frac{\|x_{k+1}\|}{\|x_k\|} = \frac{\lambda_n - \lambda_1}{\lambda_1 + \lambda_n} < 1$$

- Condition number:**  $\rho = \frac{\lambda_n}{\lambda_1} \geq 1$

$$\frac{\lambda_n - \lambda_1}{\lambda_1 + \lambda_n} = \frac{\rho - 1}{\rho + 1} \leq 1$$

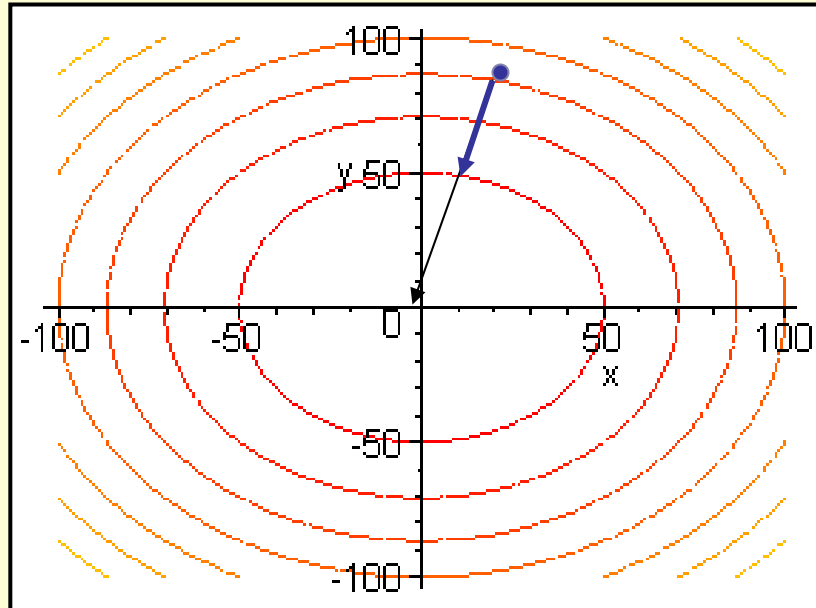
1 is best

# Well Conditioned Problem

$$x^2 + y^2$$

$$\text{Cond num} = 1/1 = 1$$

**Steepest Descent  
does one step like Newton**

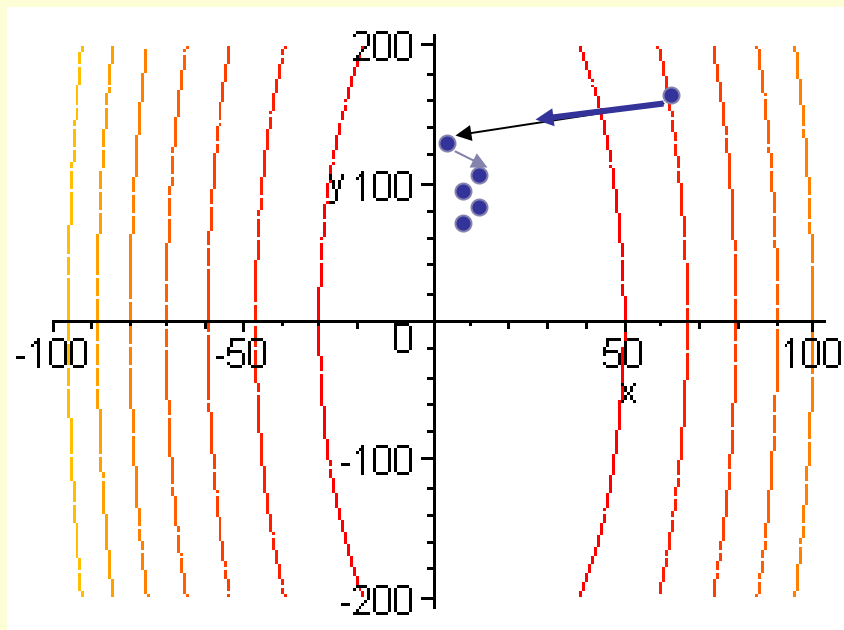


# ill-Conditioned Problem

$$50(x-10)^2 + y^2$$

$$\text{Cond num} = 50/1 = 50$$

**Steepest Descent**  
**ZIGZAGS!!!**





# Linear Convergence for our fixed stepsize

$$\frac{\|x_{k+1}\|}{\|x_k\|} = \frac{\lambda_n - \lambda_1}{\lambda_1 + \lambda_n} < 1$$



# Theorem 3.3 MW

Let  $\{x_k\}$  be the sequence generated by steepest descent with exact linesearch applied to function  $\min 1/2x'Qx-b'x$  where  $Q$  is p.d.

Then for any  $x_0$ , the sequence converges to the unique minimizer  $x^*$  and

$$\|x_{k+1} - x^*\|_Q^2 \leq \left[ \frac{\text{cond}(Q) - 1}{\text{cond}(Q) + 1} \right]^2 \|x_k - x^*\|_Q^2$$

That is the method converges linearly

# Condition Number

Officially:

$$\text{cond}(A) = \|A\| \cdot \|A^{-1}\|$$

But this as ratio of max and min eigenvalues if  $A$  is pd and symmetric

$$\begin{aligned} \|A\| &= \sqrt{\lambda_{\max}(A' A)} \\ &= \sqrt{\lambda_{\max}(A)^2} = \lambda_{\max}(A) \text{ if } A \text{ is pd and symmetric} \end{aligned}$$

$$\begin{aligned} \|A^{-1}\| &= \sqrt{\lambda_{\max}(A^{-1} ' A^{-1})} \\ &= \sqrt{\lambda_{\max}(A^{-1})^2} = 1 / \lambda_{\min}(A) \end{aligned}$$

# Theorem 3.4

- Suppose  $f$  is twice cont diff and the sequence of steepest descent converge to  $x^*$  satisfying SOSC.

Let  $r \in \left( \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)$  where  $\lambda_1 \leq \dots \leq \lambda_n$  eigenvalues of  $\nabla^2 f(x^*)$

The for all  $k$  sufficient large

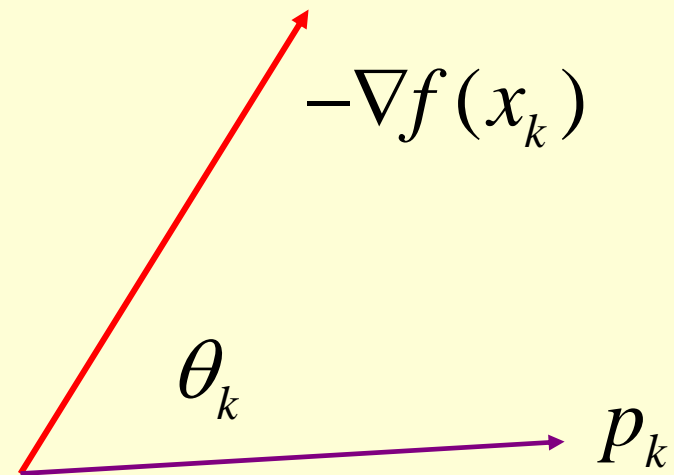
$$f(x_{k+1}) - f(x^*) \leq r^2 [f(x_k) - f(x^*)]$$

# Other possible directions

- Any direction  $\nabla f(x_k)' p_k < 0$  satisfying is a descent direction.

Which ones will work?

$$\cos \theta_k = \frac{-\nabla f(x_k)' p_k}{\|\nabla f(x_k)\| \|p_k\|}$$





# Zoutendijk's Theorem

For any descent direction and step size satisfying the Wolfe conditions,  $f$  bounded, below, Lipschitz, and differentiable

$$\sum_{k>0} \cos^2 \theta_k \|\nabla f(x_k)\|^2 < \infty$$

Corollary:  $\cos^2 \theta_k \|\nabla f(x_k)\|^2 \rightarrow 0$





# Convergence Theorem

If  $\cos \theta_k \geq \delta > 0$  for all iteration  $k$

Then

$$\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$$

So steepest descent converges!

Many other variations.





# Theorem Convergence of Gradient Related Methods

(i) Assume the set  $S$  is bounded where

$$S = \{x : f(x) \leq f(x_0)\}$$

(ii) Let  $\nabla f(x)$  be Lipschitz cont on  $S$  i.e.

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

for all  $x, y \in S$  and some fixed finite  $L$

For the sequence  $\{x_k\} \subset S$  generated by

$$x_{k+1} = x_k + p_k$$


# Theorem 10.2 Nash and Sofer

(iii) Let the directions  $p_k$  satisfy:

$$\frac{p_k^T \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} \geq \varepsilon > 0$$

(iv) be gradient related:

$$\|p_k\| \geq m \|\nabla f(x_k)\| \text{ for all } k \text{ (with } m > 0)$$

and bounded in form:

$$\|p_k\| \leq M \text{ for all } k$$




# Theorem 10.2

- (v) Let the stepsize be the first element of the sequence  $\{1, 1/2, 1/4, 1/8, \dots\}$  satisfying

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \mu \alpha_k \nabla f(x_k)' p_k$$

with  $0 < \mu < 1$ .

Then

$$\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0.$$


# Steepest Descent

- An obvious choice of direction is

$$p_k = -\nabla f(x_k)$$

- Called steepest descent because

$$\min \frac{p_k' \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} = -\frac{\nabla f(x_k)' \nabla f(x_k)}{\|\nabla f(x_k)\| \|\nabla f(x_k)\|} = -1$$

- Clearly satisfies gradient related requirement of Theorem 10.2



# Steepest Descent Summary

- Simple algorithm
  - Inexpensive per iteration
  - Only requires first derivative information
  - Global convergence to local minimum
  - Linear convergence – may be slow
  - Convergence rate depends on condition number
  - May zigzag
- 