LECTURE 4:

PREREQUISITES:

LINEAR ALGEBRA:

- NORMS $\|x\|$, $\|A\|$
- MATRIX MULTIPLICATIONS, Ax, AB, $A^T y$, $x^T y$
- LINEAR INDEPENDENCE
- $\mathrm{rank}(A)$

CALCULUS: (MULTI-VAR):

- ∇f (GRADIENT), $\nabla^2 f$ (HESSIAN)
- TAYLOR EXPANSIONS IN $\mathbb{R}^n$
- $\{x_k\}$ SEQUENCES $\lim_{k \to \infty} x_k$

LECTURE 4: UNCONSTRAINED OPT. ALGORITHMS

$$\boxed{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{array}}$$

- SOLVED USING ITERATIVE ALGORITHMS
  - START IN x_0
  - ITERATE x_k BY SOME RULE
  - TERMINATE

DEF: **LINE SEARCH TYPE ALGORITHM**

STEP 0: CHOOSE STARTING POINT $x_0 \in \mathbb{R}^n$, LET $k=0$

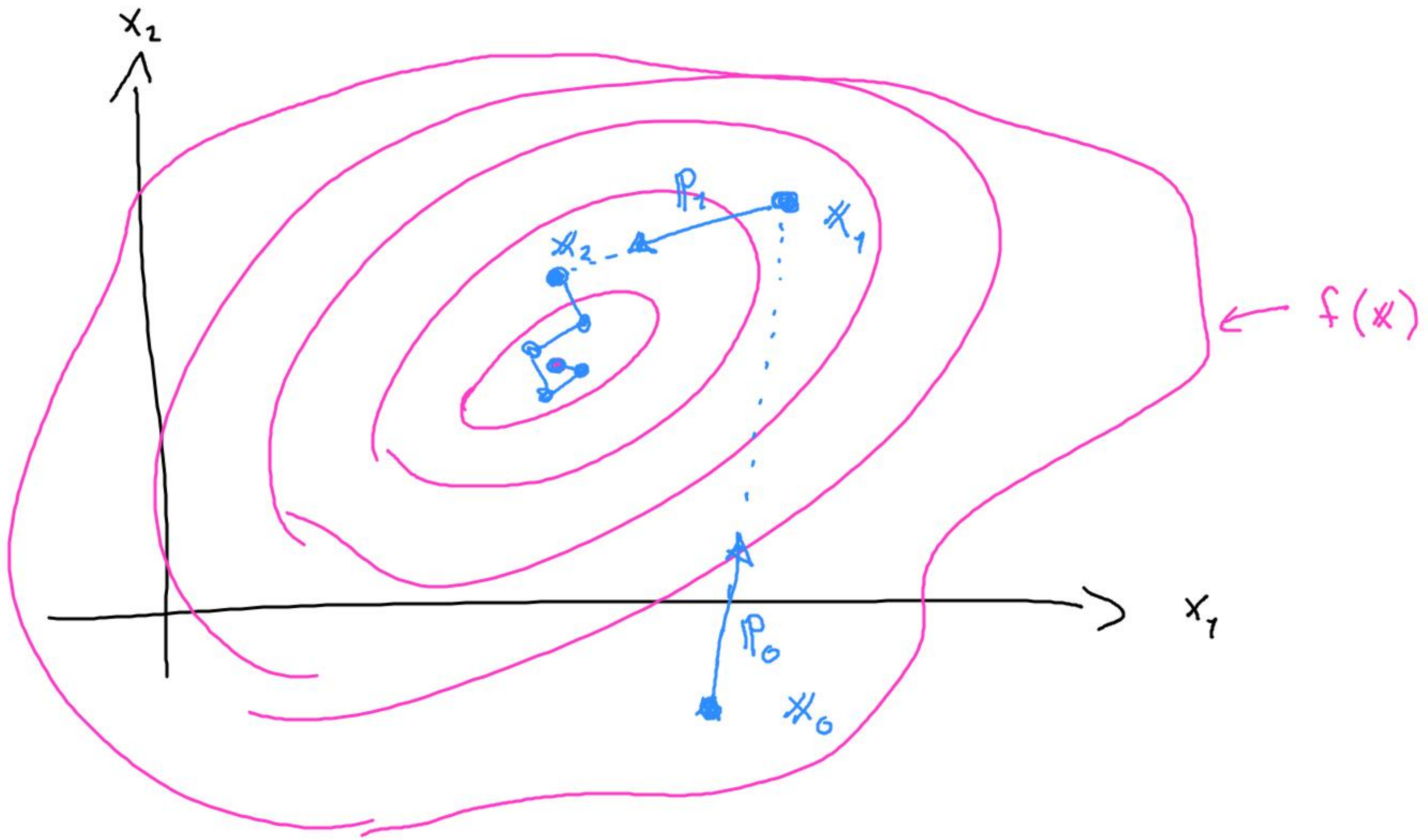
STEP 1: FIND A SEARCH DIRECTION $p_k \in \mathbb{R}^n$

STEP 2: PERFORM LINE SEARCH, i.e. FIND A $\alpha_k > 0$ SUCH THAT

$$f(x_k + \alpha_k p_k) < f(x_k)$$

STEP 3: LET $x_{k+1} = x_k + \alpha_k p_k$

STEP 4: IF TERMINATION CRITERIA IS FULFILLED, THEN STOP!
OTHERWISE LET $k = k+1$ AN GO TO STEP 1.



x_2
x_1
p_1
x_1
x_2
$f(x)$
p_0
x_0

STEP 1: FIND SEARCH DIRECTION.

- VECTOR p_k IS A DESCENT DIRECTION AT x_k IF

$$f(x_k + \alpha p_k) < f(x_k) \quad \text{FOR ALL } \alpha \in (0, \delta] \text{ FOR SOME } \delta > 0$$

- IF $f \in C^1$ (CONTINUOUS ∇f). WE KNOW THAT IF $\nabla f(x_k)^T p_k < 0$, THEN p_k IS A DESCENT DIRECTION

- $p_k = -\nabla f(x_k)$ IS A DESCENT DIRECTION (WHY? $-\nabla f(x_k)^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0$ AS LONG AS $\nabla f(x_k) \neq 0$)

- THIS DIRECTION $(p_k = -\nabla f(x_k))$ IS CALLED THE STEEPEST DESCENT DIRECTION SINCE IT SOLVES

$$\begin{aligned} \min \quad & \nabla f(x_k)^T p_k \\ \text{s.t.} \quad & \|p_k\| = 1, \; p_k \in \mathbb{R}^n \end{aligned}$$

- STEEPEST DESCENT DIRECTION: $p_k = -\nabla f(x_k)$
- LET $Q \in \mathbb{R}^{n \times n}$ BE A POSITIVE DEF. MATRIX. THEN

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

IS DESCENT DIRECTION TO f AT x_k. $\left(\text{WHY? } -\nabla f(x_k)^{T} Q \nabla f(x_k) < 0 \text{ SINCE } Q \succ 0 \right)$

- NEWTON'S SEARCH DIRECTION: $Q = [\nabla^2 f(x_k)]^{-1}$

STEEPEST DESCENT DIRECTION: $Q = I$

NEWTON'S METHOD

ASSUME $f \in C^2$. AND TAYLOR EXPAND AROUND x_k IN THE DIRECTION p:

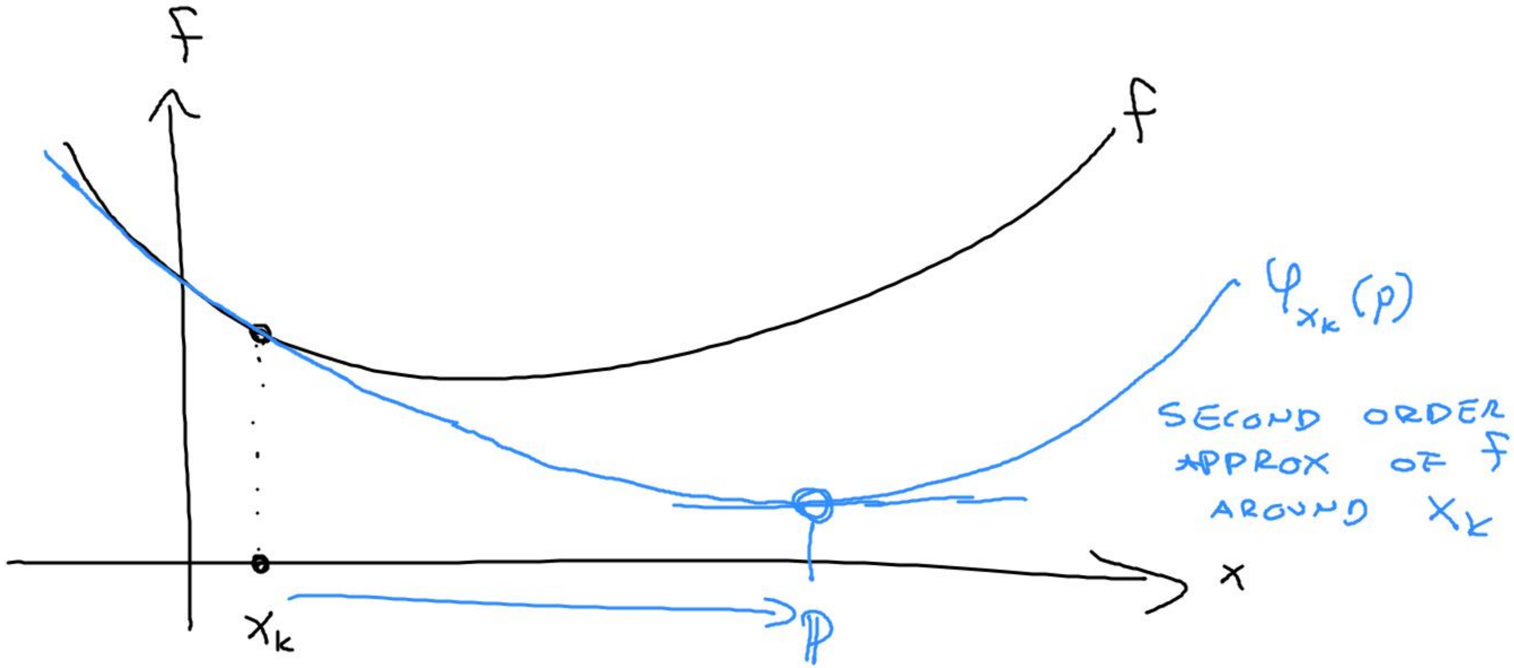
$$f(x_k + p) \approx \underbrace{f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p}_{\varphi_{x_k}(p)} = \varphi_{x_k}(p)$$

LETS TRY TO MINIMIZE $\varphi_{x_k}(p)$ W.R.T p. ASSUME $\nabla^2 f(x_k) > 0$ THEN WE CAN USE "$\nabla_p \varphi_{x_k}(p) = 0$" TRICK.

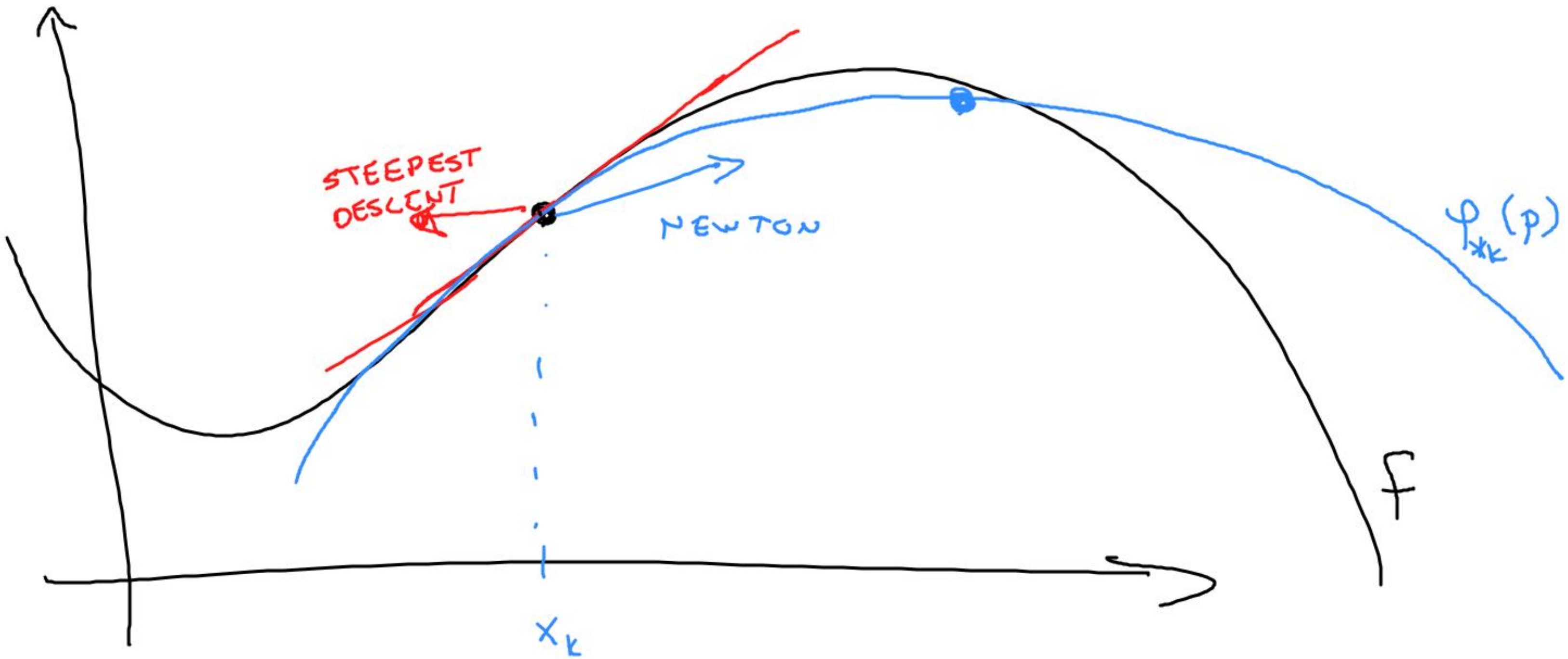
$$\nabla_p \varphi_{x_k}(p) = \nabla f(x_k) + \nabla^2 f(x_k) p = 0 \implies \nabla^2 f(x_k) p = -\nabla f(x_k)$$

$$\implies \boxed{p = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)}$$

NEWTON



NEWTON



- NEWTON'S METHOD ONLY GIVES DESCENT DIRECTIONS IF $\nabla^2 f(x_k) \succeq 0$

- LEWENBERG-MARQUARDT MODIFICATION TO NEWTON'S METHOD

  - NEWTON METHOD: $p_k = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$

    STEEPEST DESC: $p_k = -\nabla f(x_k)$

    LEWENBERG-MARQUARDT: $p_k = -\left[\nabla^2 f(x_k) + \lambda I\right]^{-1} \nabla f(x_k)$

    FOR SOME $\lambda > 0$ SUCH THAT

    $$\left[\nabla^2 f(x_k) + \lambda I\right] \succeq 0$$

<u>NOTE</u>: WHEN $\lambda = \infty$, THEN LEVENBERG-MARQUARDT IS STEEPEST DESC

S.D

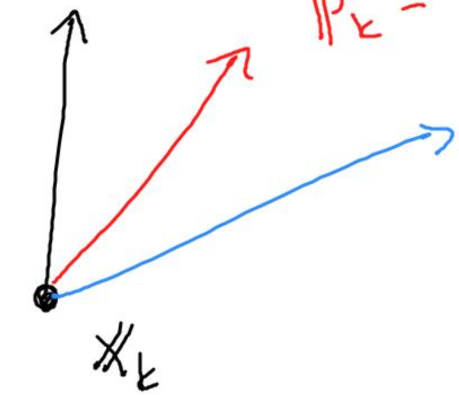
$-\nabla f(x_k)$

L-M

$$p_k = -\left[\nabla^2 f(x_k) + \lambda I\right]^{-1} \nabla f(x_k)$$

NEWTON

$$p_k = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$$



- QUASI-NEWTON DIRECTIONS ARE USED WHEN $\nabla^2 f(x_k)$ IS EXPENSIVE TO COMPUTE. TRY TO APPROXIMATE HESSIAN WITH B_k.

- SUMMARY

  - STEEPEST DESCENT: $p_k = -\nabla f(x_k)$
  - NEWTON'S $\nabla^2 f(x_k)\, p_k = -\nabla f(x_k)$
  - LEVENBERG-MARQUARDT: $[\nabla^2 f(x_k) + \lambda I]\, p_k = -\nabla f(x_k)$
  - QUASI-NEWTON: $B_k\, p_k = -\nabla f(x_k)$

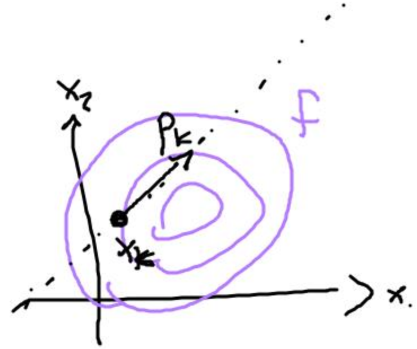
THIS IS HOW YOU COMPUTE IT

NOTE: $\boxed{\nabla^2 f(x_k)\, p_k = -\nabla f(x_k)} \iff p_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$

STEP 2: STEP LENGTH α_k

ASSUME WE ARE STANDING IN x_k AND HAVE A SEARCH DIRECTION p_k

WE WOULD LIKE TO SOLVE
$$\min f(x_k + \alpha p_k) \quad \text{s.t. } \alpha \geq 0$$



THIS IS UNNECESSARY! SINCE THE SOLUTION TO THE ORIGINAL PROBLEM $\min f(x)$ s.t. $x \in \mathbb{R}^n$ IS NOT GOING TO LIE ON THAT LINE.

WAYS OF CHOOSING STEP LENGTHS:

- EXACT STEP LENGTH: SOLVING $\min f(x_k + \alpha p_k)$ s.t. $\alpha \geq 0$

  ONLY USED WHEN f IS SIMPLE.

- INTERPOLATION: USE $f(x_k)$, $\nabla f(x_k)$, $\nabla^2 f(x_k)$ TO APPROXIMATE f ALONG THE DIRECTION OF p_k. AND THEN FIND EXACT STEP LENGTH FOR THAT APPROX.

- NEWTON'S METHOD: $\alpha := \alpha - \varphi'(\alpha) / \varphi''(\alpha)$ WHERE φ IS SECOND ORDER APPROX ALONG p_k.

- GOLDEN SECTION: DERIVATIVE-FREE METHOD WHERE YOU SHRINK AN INTERVAL ITERATIVELY

- ARMIJO RULE: NEXT.

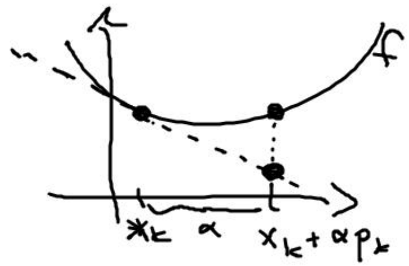
ARMIJO STEP LENGTH RULE

THE IDEA IS TO FIND A STEP LENGTH α_k WHICH PROVIDES A <u>SUFFICIENT DECREASE</u> OF f. WE HAVE THAT

$$f(x_k + \alpha p_k) \approx f(x_k) + \underbrace{\alpha \nabla f(x_k)^T p_k}_{}$$

(1ST ORDER TAYLOR)

PREDICTED CHANGE OF f IF WE MOVE FROM x_k TO $x_k + \alpha p_k$



- THE PREDICTED DECREASE IN f IS IN MANY CASES OPTIMISTIC.

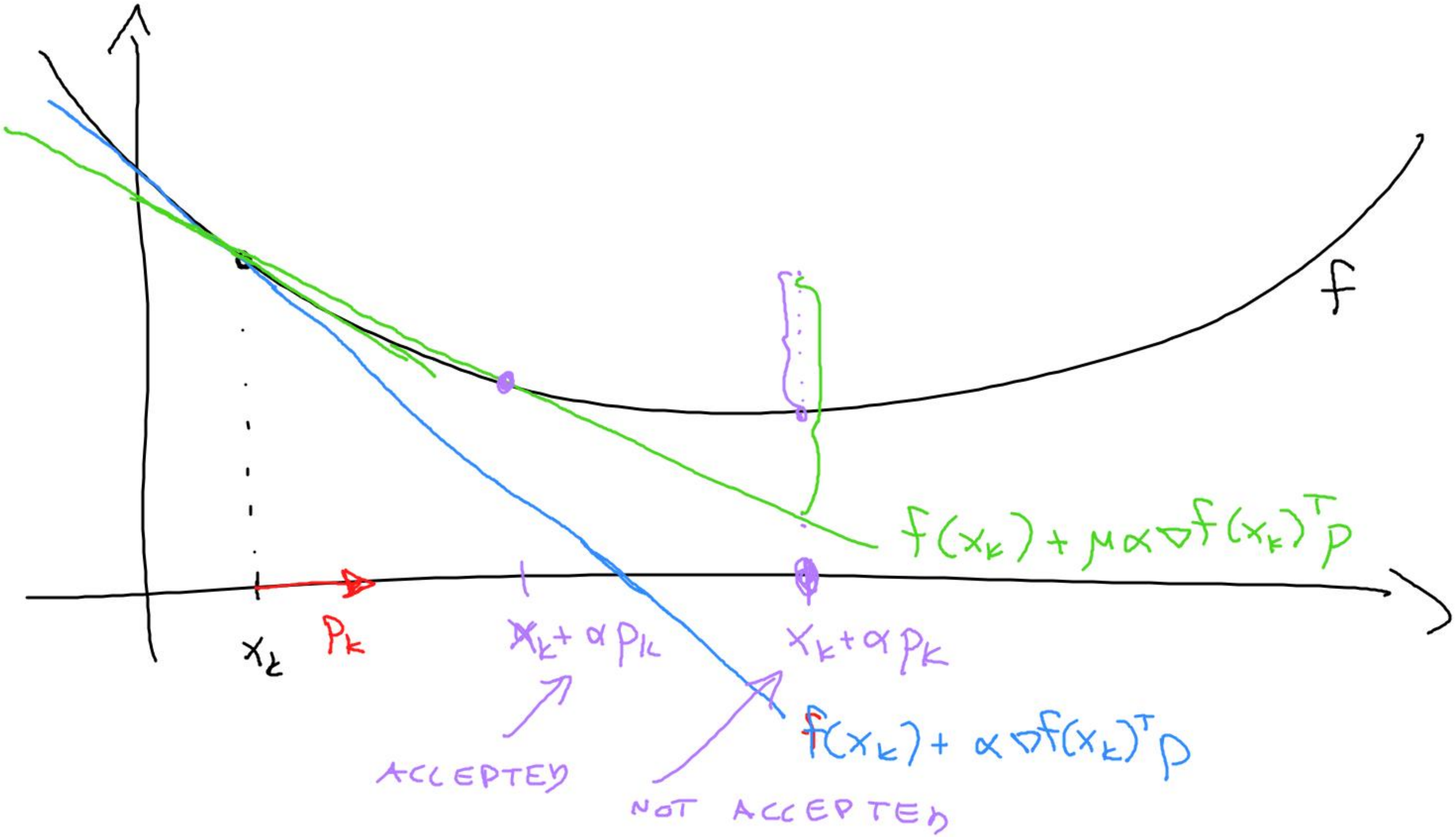
- ACCEPT A STEP LENGTH α IF THE DECREASE OF f IS A FRACTION μ OF THE PREDICTED DECREASE

- IF $$\underbrace{f(x_k + \alpha p_k) - f(x_k)}_{\text{DECREASE}} \leq \underbrace{\mu}_{\text{FRACTION}} \alpha \underbrace{\nabla f(x_k)^T p_k}_{\text{PREDICTED DECREAS}}$$

  THEN WE ACCEPT α AS STEP LENGTH

- IF NOT LET $\alpha := \frac{\alpha}{2}$ AND TRY AGAIN
- FINALLY SOME α WILL BE ACCEPTED (HOMEWORK INTUITION TO UNDERSTAND WHY)

- USUALLY $\mu \in [0.001, 0.1]$



f
$f(x_k) + \mu\alpha\nabla f(x_k)^T p$
x_k
p_k
$x_k + \alpha p_k$
$x_k + \alpha p_k$
$f(x_k) + \alpha\nabla f(x_k)^T p$
ACCEPTED
NOT ACCEPTED

STEP 4: TERMINATION CRITERIA [NOTE: $\nabla f(x) = 0$ CAN NEVER BE FOUND]

a) $\|\nabla f(x_k)\| \leq \varepsilon_1 (1 + |f(x_k)|)$ [STOP WHEN GRADIENT IS SMALL]

FOR SOME SMALL ε_1.

b) $f(x_{k-1}) - f(x_k) \leq \varepsilon_2 (1 + |f(x_k)|)$ [STOP WHEN NO FURTHER IMPROVEMENT IS MADE]

c) $\|x_{k-1} - x_k\| \leq \varepsilon_3 (1 + \|x_k\|)$ [STOP WHEN ITERATES ARE NOT MOVING]

NOTE: USE AT LEAST 2.

CONVERGENCE

WE NEED SEARCH DIRECTIONS p_k TO FULFILL (★)

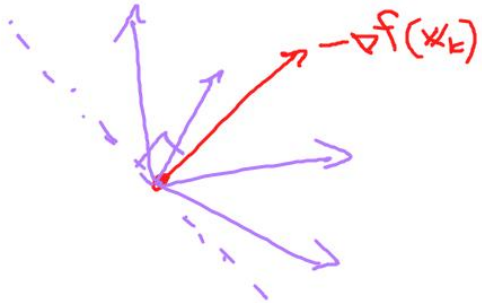
$$-\frac{\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \cdot \|p_k\|} \geq s_1 , \quad \|p_k\| \geq s_2 \|\nabla f(x_k)\| , \quad \|p_k\| \leq M$$

FOR SOME $s_1, s_2 > 0$.

ENSURES THAT ONLY WHEN $\nabla f(x_k) = 0$ CAN $p_k = 0$. SO p_k CANNOT BE 0 UNLESS $\nabla f(x_k) = 0$

SEARCH DIRECTIONS SHOULD BE BOUNDED

ENSURES THAT THE ANGLE BETWEEN p_k AND $\nabla f(x_k)$ IS BETWEEN 0 AN $\pi/2$ AND NOT TOO CLOSE TO $\pi/2$



NOTE: $p_k = -\nabla f(x_k)$ FULFILLS REQ.

THEOREM: SUPPOSE $f \in C^1$. AND FOR STARTING POINT $x_0 \in \mathbb{R}^n$ THE LEVEL SET $\{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ IS BOUNDED. ASSUME p_k FULFILLS REQUIREMENTS (*) AND α_k ARE CHOSEN ACCORDING TO ARMIJO'S RULE. THEN

a) THE SEQUENCE $\{x_k\}$ IS BOUNDED

b) THE SEQUENCE $\{f(x_k)\}$ IS DESCENDING AND LOWER BOUNDED

c) EVERY LIMIT POINT OF $\{x_k\}$ IS A STATIONARY POINT $(\nabla f(\bar{x}) = 0)$.