METODO DEL GRADIENTE

(P) $\min_{x \in \mathbb{R}^n} f(x)$

Sia $x^0 \in \mathbb{R}^n$ NON STAZIONARIO

allora $D_s(x^0) \neq \emptyset$ e $D_s(x^0) \underset{\downarrow}{\subseteq} D(x^0)$

è cont. nell'insieme di direzioni di discesa

Quindi posso considerare $d \in D_s(x^0)$
e muovermi lungo questa direzione e determinare un nuovo punto x^1 con valore di funzione obiettivo minore del valore di funzione obiettivo in x^0

$$x^1 = x^0 + \alpha d \quad \text{con } \alpha > 0$$

$$f(x^1) < f(x^0)$$

Consideriamo come particolare direzione $d \in D_s(x^0)$
l'antigradiente $\underbrace{d = -\nabla f(x^0)}$

[f è continuamente differenziabile]

DIREZIONE DI MASSIMA DISCESA

PROP. ANTIGRADIENTE:

L'antigradiente è:

1) direzione di massima discesa

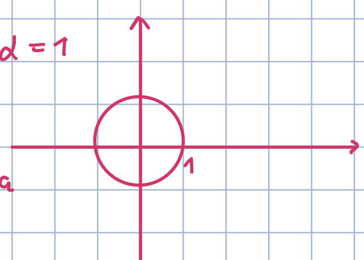
2) perpendicolare alle curve di livello della funzione

$\gamma = \{x \in \mathbb{R}^n \mid f(x) = \alpha\}$
(l'insieme dei punti per cui $f(x) = \alpha$)

es: $f(x) = \|x\|^2$

$\alpha = 1$

γ sarà l'insieme dei punti che $\in$ alla circonferenza



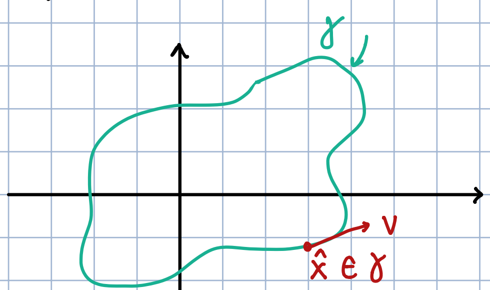
OSS (2):

Consideriamo una linea di livello

$$\{x \in \mathbb{R}^n \mid f(x) = \alpha\}$$

e sia $v \in \mathbb{R}^n$ tangente a γ in un punto $\hat{x} \in \gamma$

es. $f: \mathbb{R}^2 \to \mathbb{R}$



La derivata direzionale di f lungo v sarà 0

$$D_v f(\hat{x}) = \nabla f(\hat{x})^T \cdot v =$$

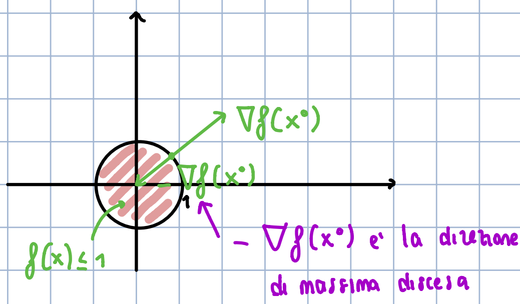
$$= \|\nabla f(\hat{x})\| \cdot \|v\| \cdot \cos\theta$$

con θ angolo tra $\nabla f(\hat{x})$ e v
quindi $\theta = \pi/2$ ovvero $\nabla f(\hat{x}) \perp v$

es:

$f(x) = \|x\|^2$

$\gamma = \{x \in \mathbb{R}^2 \mid f(x) = 1\}$



direzione $\nabla f(x)$?

↓

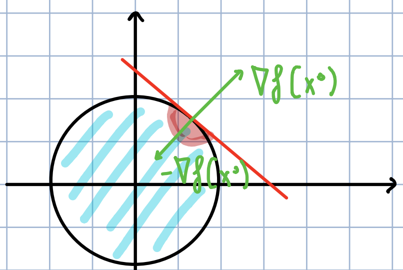
Verso l'esterno perchè i punti nella circ. sono ≤ 1

direzione $-\nabla f(x)$?

↓

Verso l'interno

es:



$$D_s(x^\circ) = \{d \in \mathbb{R}^n \mid \nabla f(x^\circ)^T d < 0\}$$

PASSI METODO DEL GRADIENTE

Passo 0: Dati $x^\circ \in \mathbb{R}^n$, $\gamma \in (0,1)$, si pone $k = 0$

entriamo in un ciclo → for $k = 0, 1, \dots$

Passo 1: Se $\nabla f(x^k) = 0$ STOP (perchè abbiamo trovato un punto stazionario)

Passo 2: Si pone $d^k = -\nabla f(x^k)$

Passo 3: Si calcola $\alpha^k > 0$ t.c.

i) $f(x^k + \alpha^k d^k) \leq f(x^k) + \alpha^k \gamma \nabla f(x^k)^T d$

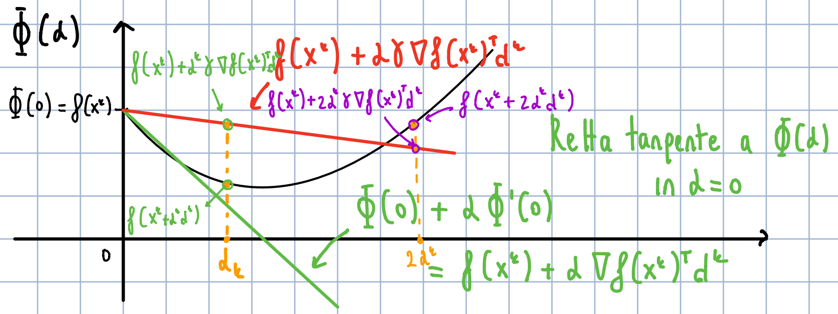
ii) $f(x^k + 2\alpha^k d^k) > f(x^k) + 2\alpha^k \gamma \nabla f(x^k)^T d^k$

Passo 4: Si pone $x^{k+1} = x^k + \alpha^k d^k$; $k = k+1$

End for.

OSS:

Nel passo 3 si determina $\alpha^k \in \mathbb{R}$; $\alpha^k > 0$, tentando di minimizzare (in maniera approssimata) la funzione $\Phi(\alpha) = f(x^k + \alpha d^k)$



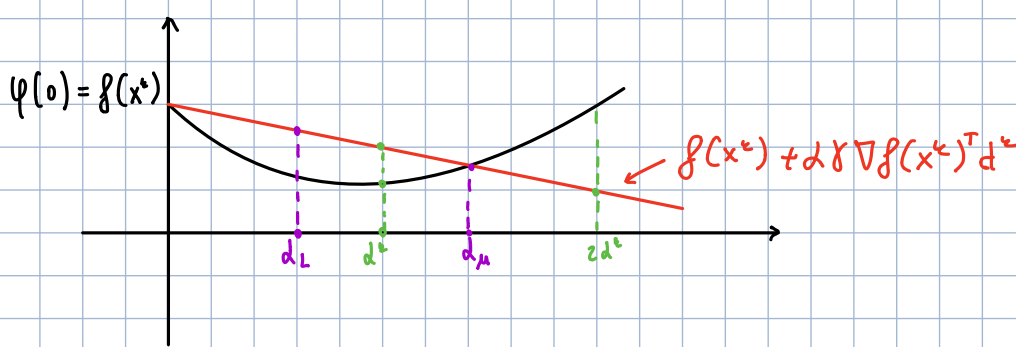
$\Phi(\alpha^k)$ deve essere al di sotto della retta tangente

$\Phi(2\alpha^k)$ deve essere al di sopra

Le condizioni al passo 3 impongono che lo scalare α^k appartenga all'intervallo $[\alpha_L, \alpha_u]$

Valori di α^k più grandi di α_u non soddisfano i)

Valori più piccoli di α_L non soddisfano ii)



Idealmente vorrei cercare $\alpha^k > 0$, $\alpha^k \in \mathbb{R}$ t.c

$$f(x^k + \alpha^k d^k) = \min_{\alpha > 0} f(x^k + \alpha d^k)$$

Nella pratica si cerca di trovare una soluzione approssimata a questo problema e si definiscono le così dette

PROCEDURE DI LINE SEARCH

ovvero metodi numerici che mi garantiscono la determinazione di un passo α^k che soddisfi le condizioni i) e ii) del passo 3.

TEOREMA:

Sia $f \in C^1(\mathbb{R}^n)$ e sia $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n \mid f(x) \le f(x^0)\}$ un insieme non vuoto e compatto $\Rightarrow$ $\min f(x)$, $x \in \mathbb{R}^n$ ammette soluzione

Sia $\{x^k\} \subseteq \mathbb{R}^n$ la successione generata dal metodo del gradiente. Si ha che:

i) $\{x^k\}$ ammette almeno un punto di accumulazione

ii) $\lim_{k \to \infty} \|\nabla f(x^k)\| = 0 \iff$ dal punto di vista numerico fissato $\varepsilon > 0$ $\exists\, \bar{k}$ t.c $\|\nabla f(x^k)\| < \varepsilon$

iii) Ogni punto di accumulazione $\bar{x}$ di $\{x^k\}$ e' un punto stazionario di f e inoltre $f(\bar{x}) \le f(x^0)$

DIMOSTRAZIONE:

i) Dalle condizioni su d^k al passo 3 del metodo del gradiente e dato che $d^k = -\nabla f(x^k)$ si ha

$$f(x^{k+1}) = f(x^k + \alpha^k d^k) \le f(x^k) + \alpha^k \gamma \nabla f(x^k)^T d$$

$$= f(x^k) \underbrace{- \alpha^k \gamma \|\nabla f(x^k)\|^2}_{<0} \le f(x^k)$$

$d^k = -\nabla f(x^k)$

sommando una quantita' negativa ho sicuramente una quantita' $\le f(x^k)$

Pertanto $x^k \in \mathcal{L}(x^0) \quad \forall k \in \mathbb{N}$

$\mathcal{L}(x^0)$ e' compatto per ipotesi $\overset{\text{T. B-W}}{\Longrightarrow}$ $\{x^k\}$ ammette un punto di accumulazione.

Vediamo una possibile procedura per trovare α^k

PROCEDURA DI LINE SEARCH:

PASSO 0: Dati $x^k, d^k, \nabla f(x^k) \in \mathbb{R}^n$, $\gamma \in (0,1)$, $\alpha^0 > 0$
si pone $i = 0$

PASSO 1: si calcola $f(x^k + \alpha^i d^k)$

PASSO 2: se $f(x^k + \alpha^i d^k) > f(x^k) + \alpha^i \gamma \nabla f(x^k)^T d^k$
si pone $\alpha^{i+1} = \alpha^i / 2$ e si torna al passo 1

PASSO 3: si calcola $f(x^k + 2\alpha^i d^k)$

PASSO 4: se $f(x^k + 2\alpha^i d^k) \le f(x^k) + 2\alpha^i \gamma \nabla f(x^k)^T d^k$ si pone $\alpha^{i+1} = 2\alpha^i$, $i = i+1$ e si torna al passo 3

PASSO 5: si pone $\alpha^k = \alpha^i$ e STOP

TEOREMA:

Sia $f \in C^1(\mathbb{R})$ e sia $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\}$ non vuoto e compatto

Se $d^k = -\nabla f(x^k)$ e $\nabla f(x^k) \neq 0$

la procedura di linesearch termina dopo un numero finito di iterazioni producendo un α^k che soddisfa le condizioni al passo 3 del metodo del gradiente.