



Regressione non lineare

Regressione non lineare - Introduzione

- Sino ad ora si sono considerati casi con modelli lineari nei parametri:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

- Tale scrittura risulta più generica di quanto sembri dato che la singola variabile regressore può essere una funzione non lineare generica delle condizioni sperimentali.

- Esempi:

$$x_1 = x^2 z, x_2 = \log(z)$$

Regressione non lineare - Introduzione

- Ci possono essere dei casi in cui i modelli lineari non sono adeguati, e il valore osservato Y **NON È** funzione lineare dei parametri θ .

$$y = \exp(\theta_1 + \theta_2 x^2)$$

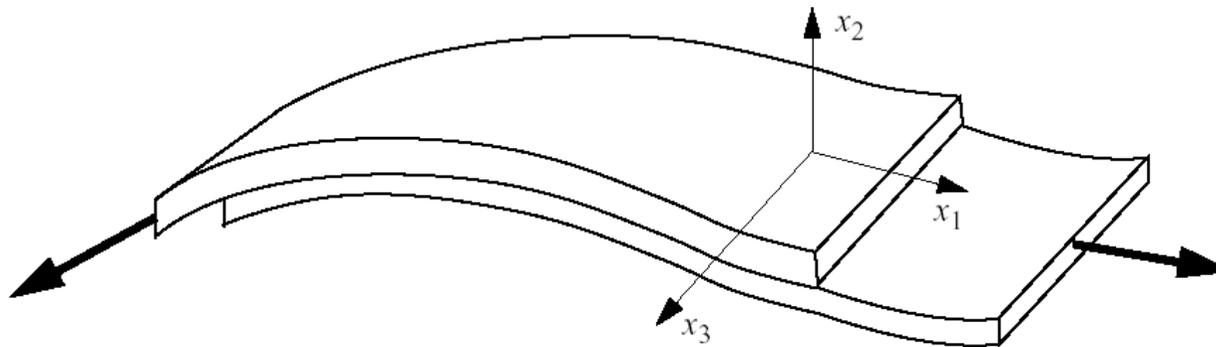
- Il modello può essere “linearizzato”, ma tale procedura porta comunque a delle stime dei parametri non ottimali

$$y = \theta_0 \exp[(-\theta_1 x) - \exp(-\theta_2 x)]$$

- Il modello non permette invece alcun tipo di linearizzazione.

Modelli non lineari

- Lo studio reologico di un materiale viene affrontato scegliendo condizioni di flusso semplici.
- Flusso di Scorrimento:
 - Un flusso è uno scorrimento se è possibile individuare una famiglia di superfici materiali (cioè fatte di particelle di fluido) le quali si mantengono inalterate durante il moto.
 - Le superfici sono in moto relativo l'una rispetto all'altra. Il moto di ciascuna di esse è però rigido (traslatorio e/o rotatorio)



Reometria

- Per effetto dello scorrimento si genera nel liquido uno sforzo tangenziale.
 - Se il liquido è Newtoniano allora: $\tau = \eta \dot{\gamma}$
 - Se il liquido è non Newtoniano: $\tau = f(\dot{\gamma})$
 - f è una funzione non lineare
 - Nel caso non-Newtoniano possono generarsi anche sforzi normali (ma qui questo aspetto non ci interessa).
- In generale $\dot{\gamma}$ dipende dal punto e dal tempo. Se esso non cambia nel tempo il flusso di scorrimento si dice viscometrico.
- Flusso viscometrico uniforme: $\dot{\gamma}$ è costante ovunque.
- Uno strumento capace di generare un flusso viscometrico uniforme può essere usato per caratterizzare la

$$f(\dot{\gamma})$$

Viscosimetri piatto e cono

- Viscosimetro rotazionale a piatto e cono
 - R è in genere circa 2cm.
 - α è pochi gradi.
- Campi di applicazione
 - E' una delle apparecchiature più utilizzate
 - shear rate da 10^{-6} a 1000 s^{-1} .
- Modello del processo
 - angolo del cono piccolo: $\sin\alpha \approx \tan\alpha \approx \alpha$
 - Legame tra coppia M e velocità angolare Ω :
 - La shear rate è data da:
 - Lo sforzo tangenziale:



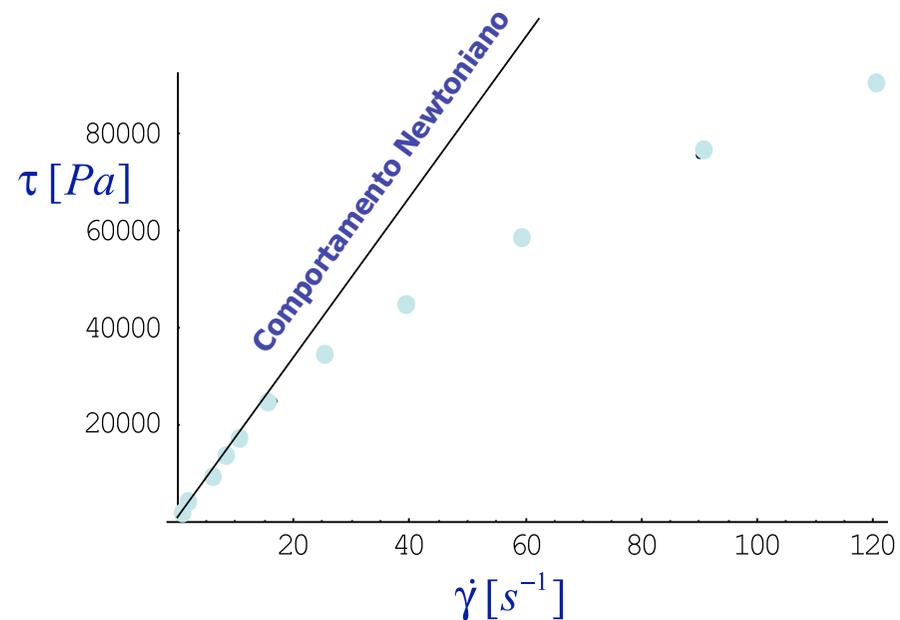
Parametri geometrici:
R, α

$$M = \eta \frac{2\pi R^3}{3\alpha} \Omega$$

$$\dot{\gamma} = \frac{\Omega}{\alpha}$$
$$\tau = \frac{3M}{2\pi R^3}$$

Equazione costitutiva reologica

- Modello dell'esperimento: $\tau_i = f(\dot{\gamma}_i)\dot{\gamma}_i + \varepsilon_i$
 - Variabile indipendente: velocità angolare $\Rightarrow \dot{\gamma}$ (reometro a deformazione imposta)
 - Variabile dipendente misurata: Coppia $\Rightarrow \tau$.
- Materiale: Polipropilene a 200°C

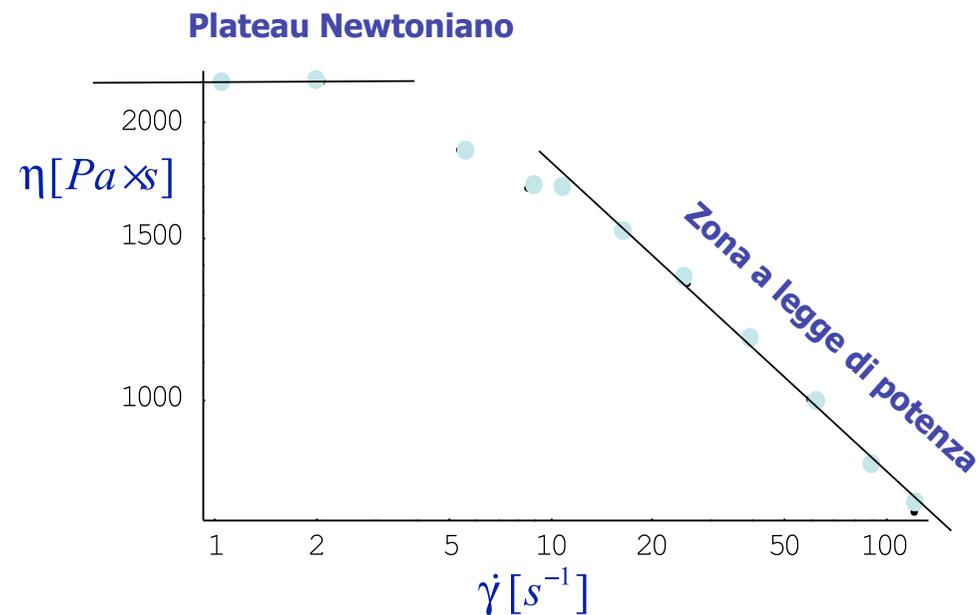


Equazione costitutiva reologica

- Se dividiamo tutta la relazione per $\dot{\gamma}$ che è una grandezza deterministica abbiamo:

$$\eta_i = f(\dot{\gamma}_i) + \varepsilon_i$$

- Come usualmente fanno i reologi, diagrammiamo questi dati su scale logaritmiche:



Equazione costitutiva reologica

- Una dipendenza della viscosità dalla shear rate di questo tipo è ben descritta dal modello di Carreau-Yasuda:

$$f(\dot{\gamma}) = \eta_0 \left(1 + |\lambda \dot{\gamma}|^a\right)^{\frac{n-1}{a}}$$

– Parametri:

- Valore di viscosità del plateau Newtoniano η_0 ;
 - Tempo caratteristico del materiale λ ;
 - Pendenza della legge di potenza $n-1$;
 - Ampiezza della zona di transizione tra i due comportamenti a .
- **Il modello è non lineare nei parametri.**
 - Obiettivo: Stimare il valore dei parametri sulla base dei dati sperimentali ottenuti.
 - Eventualmente suggerire lo svolgimento di nuovi esperimenti.

Modelli non lineari nei parametri

- La stima che abbiamo imparato ad eseguire era basata sullo stimatore della MV e si limitava al caso di un modello del processo lineare nei parametri.
- Come si modificano le cose nel caso non lineare?
 - Il modello del processo è NON lineare $g(\theta, x)$

Modello del processo vero

- IPOTESI

$$Y_i = g(x_i, \underline{\theta}) + \varepsilon_i \quad \text{Modello dell'esperimento}$$

$$\varepsilon_i := N(0, \sigma^2) \quad \text{Tipo di esperimento}$$

$$\varepsilon_i \text{ indipendente da } \varepsilon_j \quad \text{Tipo di esperimento}$$

- La funzione Verosimiglianza: non cambia

$$f_Y = \prod_{i=1}^N f_{Y_i} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(y_i - g(x_i, \underline{\theta}))^2}{\sigma^2}\right]$$

STIMA PUNTUALE DI PARAMETRI ESEMPIO CINETICA NON ISOTERMA

- Il modello non lineare potrebbe essere linearizzabile con trasformazioni non lineari
- Consideriamo la reazione chimica $A+B \rightarrow C$ in fase liquida con cinetica che dipende dalla sola T ($r=k_0e^{-E/RT}$).

$$\phi(\theta) = \sum_{i=1}^n (r_i - k_0 e^{-\frac{E}{RT}})^2$$

- Il problema è diventato non lineare nei parametri. Si perverrà ad un sistema di 2 equazioni non lineari in 2 incognite.
- C'è però da notare che in questo caso la non linearità è linearizzabile con una trasformazione non lineare.

STIMA PUNTUALE DI PARAMETRI ESEMPIO CINETICA NON ISOTERMA

- Se infatti facciamo il logaritmo della velocità di reazione otteniamo un **modello di processo** lineare nei parametri.

$$\log_e r = \log_e k_0 - \frac{E}{RT}$$

$$y' = \log_e r, \quad x' = \frac{1}{RT}, \quad \theta = [\log_e k_0, E]$$

$$y' = \theta_0 + \theta_1 x'$$

- **ATTENZIONE:** cosa succede al modello dell'esperimento

$$r = k_0 e^{\frac{E}{RT}} + \varepsilon, \quad \log(r) = \log\left(k_0 e^{\frac{E}{RT}} + \varepsilon\right)$$

- Il risultato dell'esperimento trasformato con il logaritmo non ha più lo stesso tipo di VA dell'errore sperimentale.

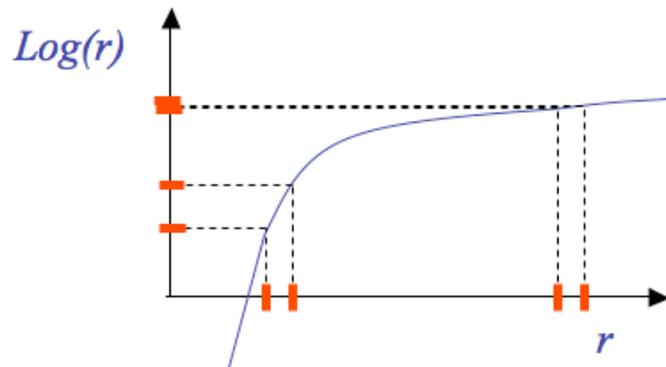
- Il suo tipo deriva dalla trasformazione $\log\left(N\left(k_0 e^{\frac{E}{RT}}, \sigma^2\right)\right)$

STIMA PUNTUALE DI PARAMETRI ESEMPIO CINETICA NON ISOTERMA

- La MV si complica notevolmente data la maggiore complicazione della congiunta. Possiamo procedere con i MQ ma sappiamo che le incertezze sulla stima cresceranno
- Se procediamo con i MQ senza riflettere richiamo di non avere buoni risultati

$$\phi(\theta) = \sum_{i=1}^N \left(\log r_i - \log k_0 + \frac{E}{RT_i} \right)^2$$

- La trasformazione di r altera l'errore



L'errore si amplifica per r piccoli e si riduce per r grandi. Se ricordiamo la definizione di Gauss del metodo dei minimi quadrati ci rendiamo conto di dover introdurre dei pesi per compensare le diverse accuratezze.

STIMA PUNTUALE DI PARAMETRI ESEMPIO CINETICA NON ISOTERMA

- I termini affetti da minor errore devono pesare di più nella minimizzazione:

$$\phi(\theta) = \sum_{i=1}^N w_i \left(\log r_i - \log k_0 + \frac{E}{RT_i} \right)^2$$

- In questo caso è facile scegliere i pesi, basta usare l'inverso della derivata della trasformazione. Dato che i pesi non possono essere grandezze negative si eleva al quadrato l'inverso:

$$w_i = \left. \frac{d\Gamma}{dy} \right|_i^{-2}$$

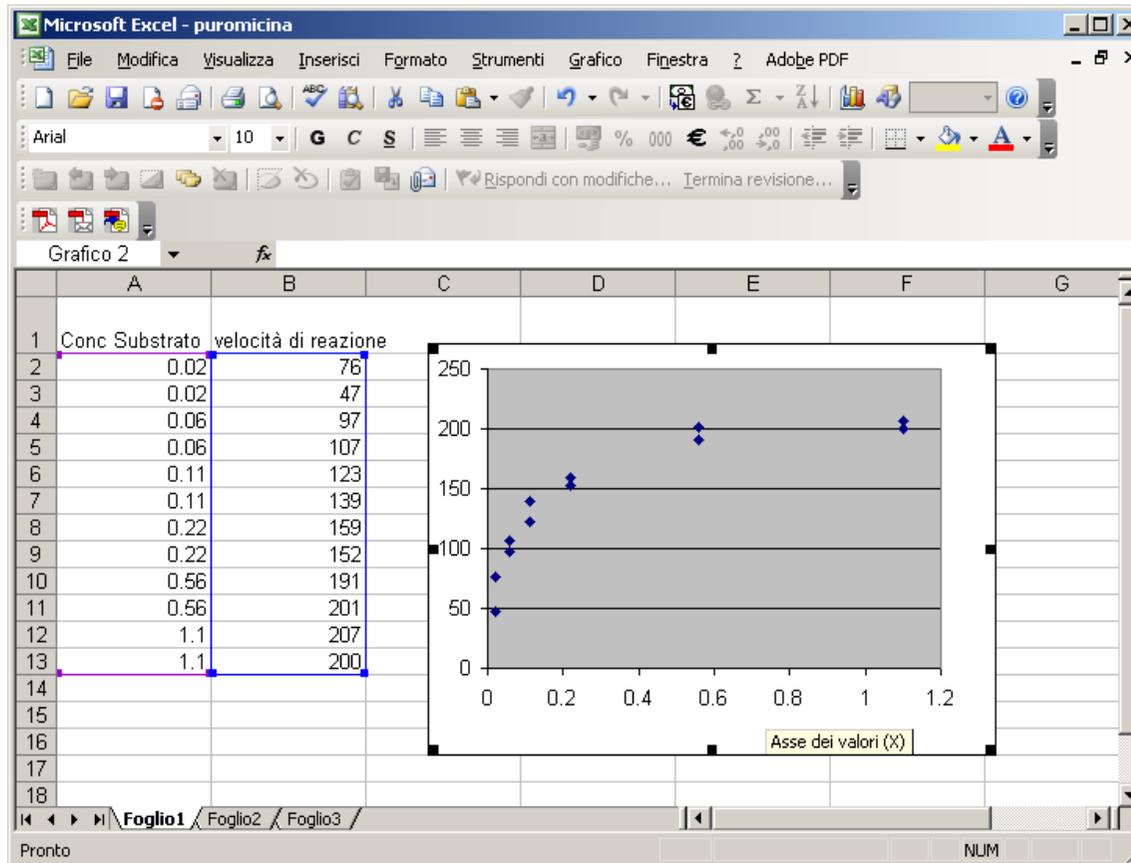
- Questo è un esempio di stima ottenuta con i **Minimi Quadrati Pesati**

Regressione non lineare - Introduzione

- Nel seguito saranno introdotti alcuni concetti elementari sulla regressione non lineare
- Il modello in esame è un modello linearizzabile
- Si effettuerà prima la regressione linearizzata, e in seguito la procedura non lineare rigorosa,
 - Si intende mostrare i limiti della linearizzazione e mettere in evidenza le differenze significative tra i risultati dei due approcci
- Saranno infine dati dei cenni teorici sugli algoritmi matematici sfruttati nella regressione non lineare

Regressione linearizzata - Esempio

- Reazione di interesse biologico



Il modello usato per descrivere la dipendenza della velocità di reazione dalla concentrazione della specie chimica è di tipo Monod

$$r = \frac{\mu S}{K + S}$$

- I parametri da stimare nel modello sono μ e K
- È chiaramente un modello non lineare nei parametri

Regressione linearizzata - Esempio



- Una strada percorribile è una linearizzazione del modello

$$\frac{1}{r} = \frac{1}{\mu} + \frac{K}{\mu} \frac{1}{S}$$

- In questo modo è possibile ricondurre il modello ad una forma lineare

$$z = \beta_0 + \beta_1 f$$

dove

$$z = \frac{1}{r}$$

$$\beta_0 = \frac{1}{\mu}$$

$$f = \frac{1}{S}$$

$$\beta_1 = \frac{K}{\mu}$$



$$\mu = \frac{1}{\beta_0}$$

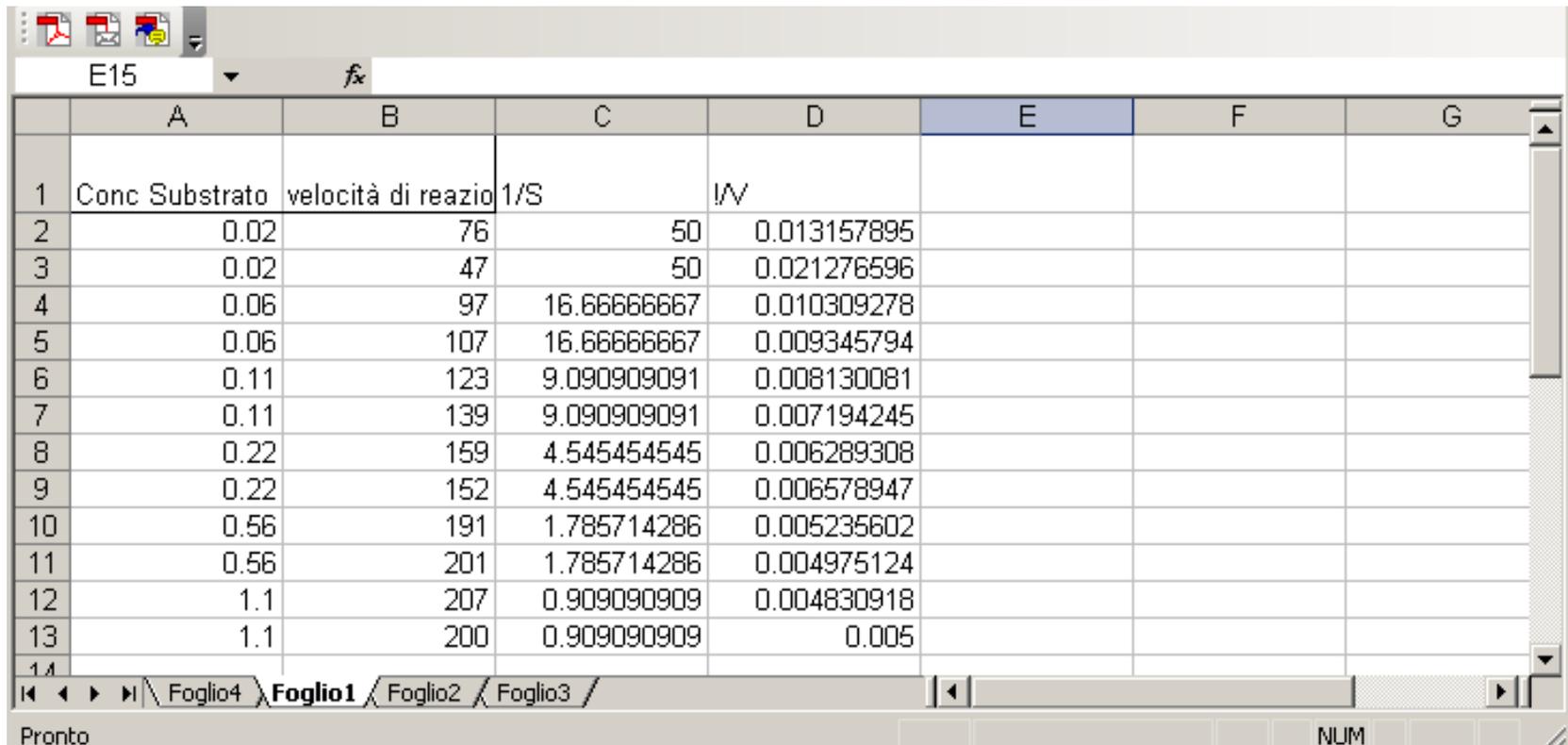
$$K = \frac{\beta_1}{\beta_0}$$

Regressione linearizzata con Excel

- È possibile eseguire la trasformazione delle variabili in maniera molto semplice con Excel
 1. Selezionare con il mouse la cella C2
 2. Nella finestra di input di Excel digitare: “=1/” e selezionare la cella A2 (dovrebbe automaticamente apparire nella finestra di dialogo). Premere enter. In questo modo dovrebbe essere eseguita l’operazione $C2=1/A2$
 3. Una volta eseguita questa operazione per la cella C2 è possibile ripeterla per tutte le altre celle: copiare il contenuto della cella C2 e “incollarla” nelle celle in basso: Excel automaticamente esegue l’operazione per la riga corrispondente
 4. Ripetere a questo punto l’operazione per calcolare l’inverso della velocità di reazione (da mettere in colonna D)

Regressione linearizzata con Excel

- Lo schermo di Excel si dovrebbe presentarsi nella seguente forma:



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G
1	Conc Substrato	velocità di reazio	1/S	1/V			
2	0.02	76	50	0.013157895			
3	0.02	47	50	0.021276596			
4	0.06	97	16.66666667	0.010309278			
5	0.06	107	16.66666667	0.009345794			
6	0.11	123	9.090909091	0.008130081			
7	0.11	139	9.090909091	0.007194245			
8	0.22	159	4.545454545	0.006289308			
9	0.22	152	4.545454545	0.006578947			
10	0.56	191	1.785714286	0.005235602			
11	0.56	201	1.785714286	0.004975124			
12	1.1	207	0.909090909	0.004830918			
13	1.1	200	0.909090909	0.005			

Regressione linearizzata con Excel

- Possiamo ora eseguire la regressione lineare sulla colonna C e la colonna D
- Eseguendo i soliti comandi

Attenzione:
Si riporta anche il grafico dei
residui

Regressione

Input

Intervallo di input Y:

Intervallo di input X:

Etichette Passa per l'origine

Livello di confidenza %

Opzioni di output

Intervallo di output:

Nuovo foglio di lavoro:

Nuova cartella di lavoro

Residui

Residui

Residui standardizzati

Tracciati dei residui

Tracciati delle approssimazioni

Probabilità normale

Tracciati delle probabilità normali

OK

Annulla

?

Regressione linearizzata con Excel

- Come appaiono i risultati della regressione
- Dai coefficienti della regressione è possibile valutare i parametri del modello:

$$\mu = \frac{1}{\beta_0} = 195.8$$

$$K = \frac{\beta_1}{\beta_0} = 0.0484$$

Microsoft Excel - puomicinawork

File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ? Adobe PDF

Arial 10 G C S

N29 fx

	E	F	G	H	I	J	K	L	M	N	O
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35											
36											
37											
38											

OUTPUT RIEPILOGO

Statistica della regressione

R multiplo	0.925037699
R al quadrato	0.855694744
R al quadrato cor	0.841264218
Errore standard	0.00189226
Osservazioni	12

ANALISI VARIANZA

	ddl	SQ	MQ	F	Significatività F
Regressione	1	0.000212324	0.000212324	59.29754512	1.64217E-05
Residuo	10	3.58065E-05	3.58065E-06		
Totale	11	0.00024813			

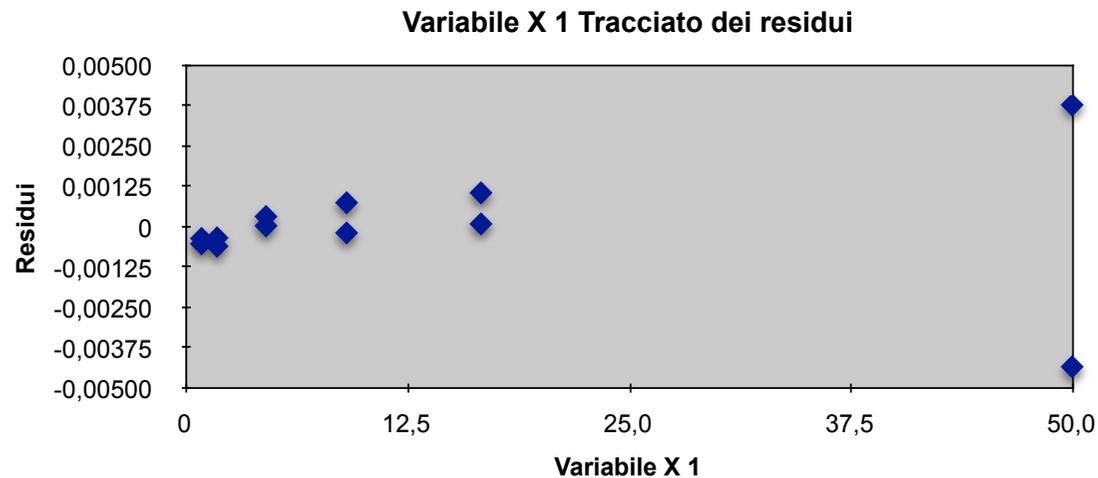
	Coefficienti	Errore standard	Stat t	Valore di significatività	Inferiore 95%	Superiore 95%	Inferiore 95.0%	Superiore 95.0%
Interoetta	0.005107182	0.000703998	7.2545362	2.74385E-05	0.003538576	0.006675798	0.003538576	0.006675798
Variable X1	0.000247221	3.21046E-05	7.700489928	1.64217E-05	0.000175688	0.000318754	0.000175688	0.000318754

OUTPUT RESIDUI

Osservazione	Y prevista	Residui	Residui standard
1	0.01746823	-0.004310335	-2.389057062
2	0.01746823	0.003808366	2.110834515
3	0.009227531	0.001081747	0.599572003
4	0.009227531	0.000118263	0.065548968
5	0.007354645	0.000775436	0.42979531
6	0.007354645	-0.0001604	-0.088903884
7	0.006230913	5.83949E-05	0.0323661
8	0.006230913	0.000348034	0.192902245
9	0.005548648	-0.000313046	-0.173509411
10	0.005548648	-0.000573523	-0.317882443
11	0.005331928	-0.00050101	-0.277891112
12	0.005331928	-0.000331928	-0.183975229

Regressione linearizzata con Excel

- Il grafico dei residui da delle prime informazioni abbastanza interessanti: **la dispersione dei residui cresce all'aumentare di $1/S$**



- Ricordiamo che:

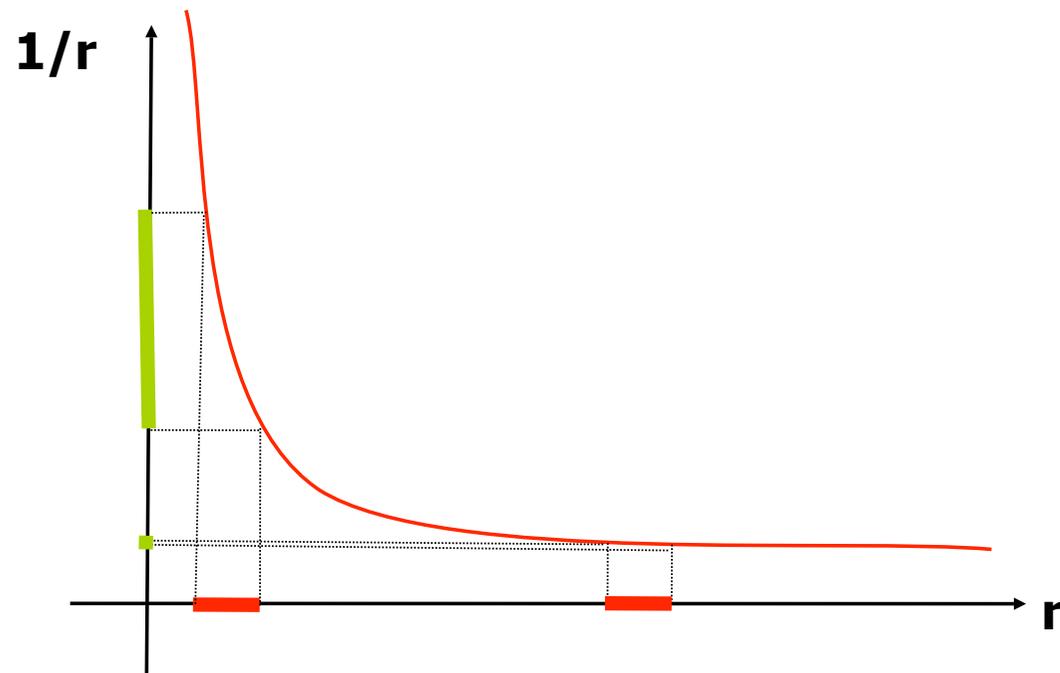
$\frac{1}{S}$ alti \Rightarrow concentrazioni S basse \Rightarrow velocità V basse

Regressione linearizzata con Excel

- Ricordiamo che la trasformazione non lineare implica una variazione dell'incertezza al variare della variabile regressore

Le incertezze a bassi valori di r sono amplificate

Sono invece attenuate ad alti valori di r



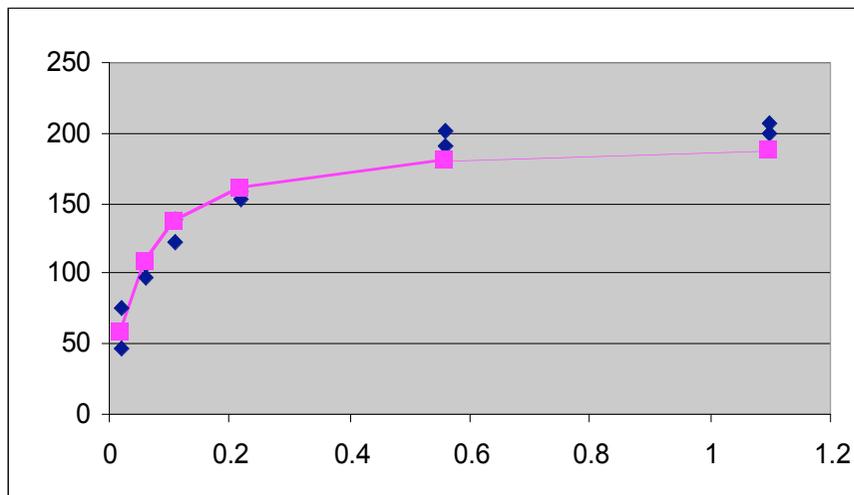
Regressione linearizzata con Excel

- Un'ulteriore conferma dei limiti della linearizzazione può essere ottenuta confrontando i dati sperimentali con i valori previsti dal modello
- A questo scopo eseguiamo i seguenti passi:
 1. Selezionare la colonna delle Y previste dall'output di regressione:

OUTPUT RESIDUI			
<i>Osservazione</i>	<i>Y prevista</i>	<i>Residui</i>	<i>Residui standard</i>
1	0.01746823	-0.004310335	-2.389057062
2	0.01746823	0.003808366	2.110834515
3	0.009227531	0.001081747	0.599572003
4	0.009227531	0.000118263	0.065548968
5	0.007354645	0.000775436	0.42979531
6	0.007354645	-0.0001604	-0.088903884
7	0.006230913	5.83949E-05	0.0323661
8	0.006230913	0.000348034	0.192902245
9	0.005548648	-0.000313046	-0.173509411
10	0.005548648	-0.000573523	-0.317882443
11	0.005331928	-0.00050101	-0.277691112
12	0.005331928	-0.000331928	-0.183975229

Regressione linearizzata con Excel

2. Copiare la colonna in un altro punto della cartella (per esempio, da P2 in giù)
3. Calcolare in un'altra colonna i reciproci delle Y predette (tali valori rappresentano le predizioni corrette delle velocità di reazione)
4. Selezionare il grafico dove sono rappresentati i punti sperimentali
5. Dal menù selezionare: Grafico Aggiungi dati
6. Si aprirà una nuova finestra in cui definire i nuovi punti da aggiungere al grafico. Selezionare la colonna delle predizioni della velocità di reazione



La linearizzazione fallisce in maniera evidente la predizione del comportamento ad alti valori di s

Regressione non lineare

- Per avere una stima più affidabile dei parametri del modello si dovrebbe rinunciare alla linearizzazione e affrontare la regressione non lineare
- Dal punto di vista matematico il problema è analogo al caso lineare:
 - È necessario determinare i valori dei parametri che rendano minima la distanza tra osservazioni e predizioni del modello
 - Nel caso della regressione lineare la ricerca del minimo della funzione obiettivo nei parametri ammette soluzione analitica e dipende linearmente dalle osservazioni
- Nel caso di un modello non lineare (nei parametri) la procedura è assolutamente analoga: si può definire la funzione somma quadratica degli errori

- Nel caso in esame
$$\phi(\theta) = \sum_{i=1}^n (y_i - f(x_i, \underline{\theta}))^2$$

$$\phi(\theta) = \sum_{i=1}^n \left(r_i - \frac{\mu S_i}{K + S_i} \right)^2$$

Modelli non lineari nei parametri

- In generale con applicando il procedimento appena visto si ottiene una stima di primo tentativo.
- Per applicare la MV dovremo procedere sul modello non linearizzato.
- Ciò che dovrà cambiare è la tecnica con cui si perviene alla stima in quanto non è più possibile operare analiticamente la minimizzazione.
- **Siamo costretti ad eseguire la minimizzazione numericamente.**



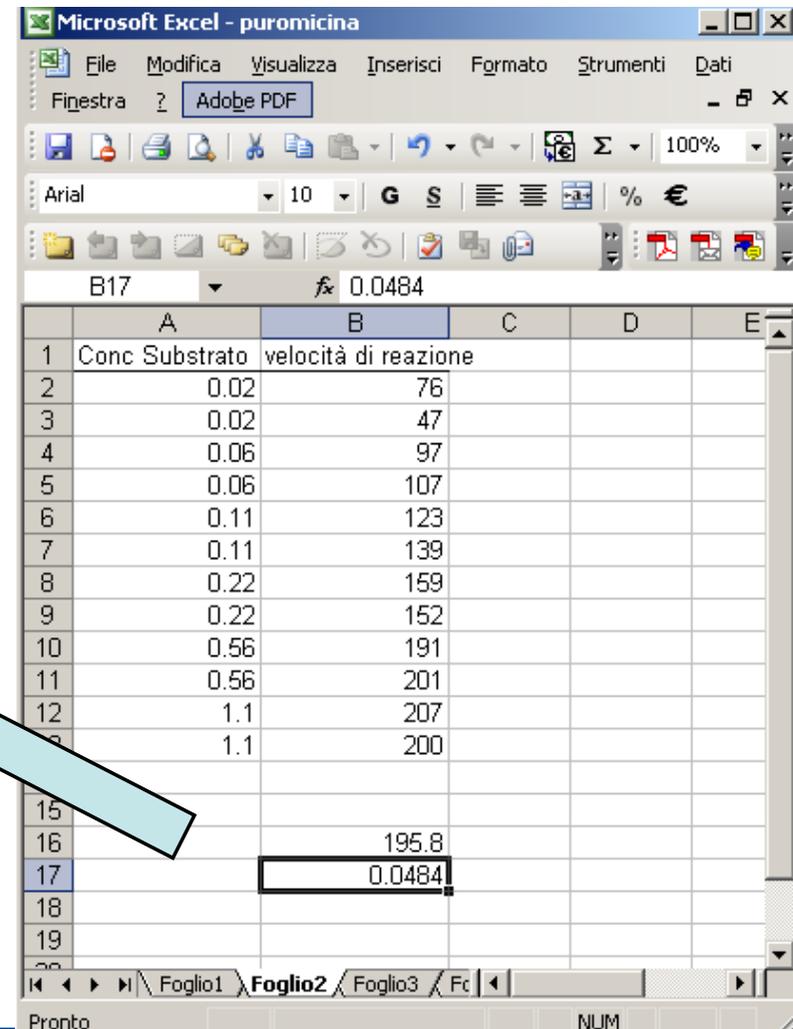
- Scopo: determinare il minimo di una funzione non lineare di più variabili.
- Il tipo di funzione da studiare è dettato dalla funzione obiettivo ridotta (quella che contiene solo i parametri del modello del processo) ed in genere è del tipo somma di scarti quadratici.
- Metodi Numerici:
 - Metodi diretti (è necessaria solo la conoscenza di Φ)
 - Metodi del gradiente (serve la conoscenza di Φ e del suo gradiente)
 - Metodi che richiedono la conoscenza del gradiente e dell'Hessiano della funzione obiettivo.

Regressione non lineare con Excel

- Dobbiamo cercare quindi di minimizzare la funzione obiettivo rispetto ai parametri ignoti.
- Per tale scopo è necessario usare un altro strumento di Excel: il menu Risolutore
- Di seguito saranno riportati i passi per eseguire una regressione non lineare con Excel come ricerca del minimo della funzione obiettivo

Regressione non lineare con Excel

1. I dati da cui partire sono foglio del file puromicina.xls
4. Selezionare alcune celle su cui scrivere i parametri μ e K . È necessario assegnare dei valori di primo tentativo. Si possono inserire dei valori a caso, la cosa migliore è inserire i valori ottenuti dalla regressione linearizzata



The screenshot shows the Microsoft Excel interface with the file 'puromicina.xls' open. The active cell is B17, containing the formula $\mu = 0.0484$. The data table is as follows:

	A	B	C	D	E
1	Conc Substrato	velocità di reazione			
2	0.02	76			
3	0.02	47			
4	0.06	97			
5	0.06	107			
6	0.11	123			
7	0.11	139			
8	0.22	159			
9	0.22	152			
10	0.56	191			
11	0.56	201			
12	1.1	207			
13	1.1	200			
15					
16		195.8			
17		0.0484			
18					
19					

A light blue arrow points from the text 'valori di primo tentativo' in the instructions to the cell B17 containing the value 0.0484.

Regressione non lineare con Excel

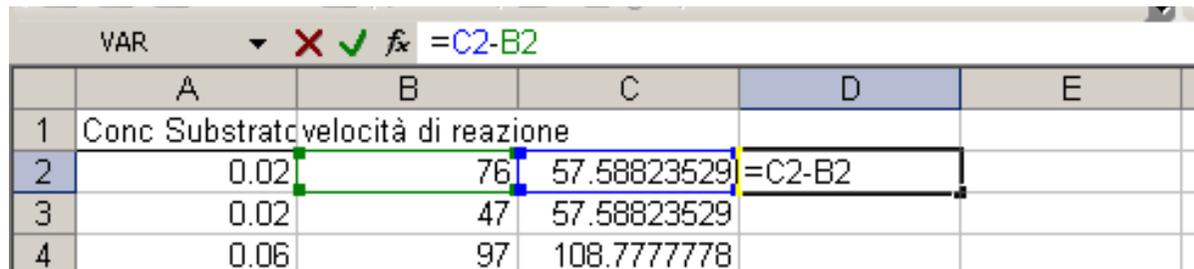
3. Nella colonna C calcolare i valori della velocità usando i parametri appena introdotti nelle celle, i dati nella colonna A secondo il modello non lineare a disposizione

- **Attenzione:**
- La formula è meno banale da implementare dei casi precedenti: i termini B15:B16 devono rimanere costanti e l'indice di A deve variare lungo la colonna.
- Non si può usare un semplice copia ed incolla

	A	B	C	D
1				
2				
3	0.02	76	57.58823529	
4	0.02	47	57.58823529	
5	0.06	97	108.7777778	
6	0.06	107	108.7777778	
7	0.11	123	136.3164557	
8	0.11	139	136.3164557	
9	0.22	159	160.7313433	
10	0.22	152	160.7313433	
11	0.56	191	180.3421053	
12	0.56	201	180.3421053	
13	1.1	207	187.6132404	
14				
15		195.8		
16		0.048		

Regressione non lineare con Excel

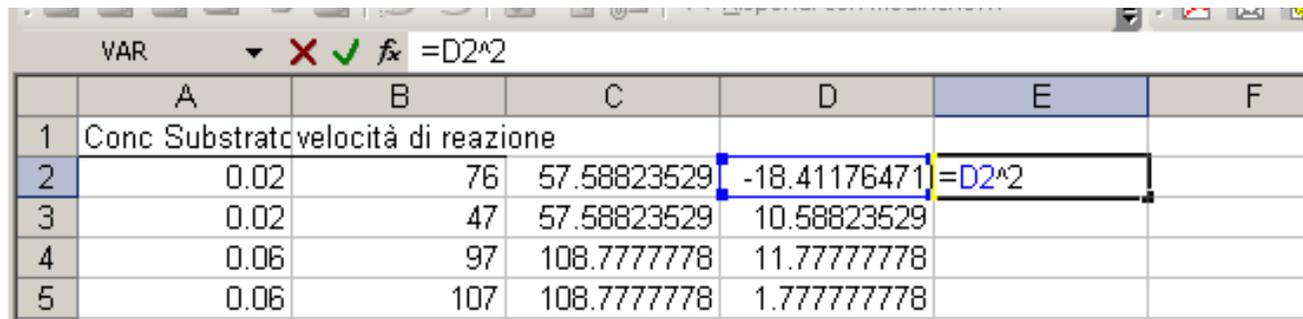
4. Calcolare la colonna D come differenza tra le colonne C e B.



VAR X ✓ fx =C2-B2

	A	B	C	D	E
1	Conc Substrato	velocità di reazione			
2	0.02	76	57.58823529	=C2-B2	
3	0.02	47	57.58823529		
4	0.06	97	108.7777778		

Valutare poi i quadrati dei risultati e metterli nella colonna F



VAR X ✓ fx =D2^2

	A	B	C	D	E	F
1	Conc Substrato	velocità di reazione				
2	0.02	76	57.58823529	-18.41176471	=D2^2	
3	0.02	47	57.58823529	10.58823529		
4	0.06	97	108.7777778	11.77777778		
5	0.06	107	108.7777778	1.77777778		

Regressione non lineare con Excel

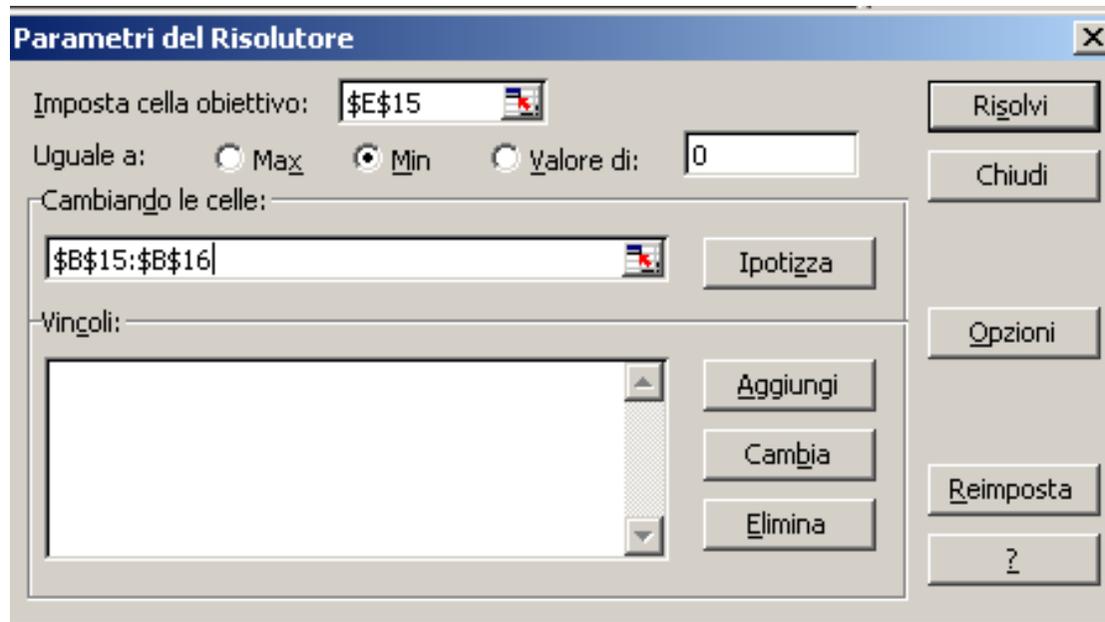
5. Inserire una cella (per esempio la cella E15) in cui sommare tutti gli elementi della colonna E

Tale cella rappresenta la nostra funzione obiettivo

	A	B	C	D	E	F
1	Conc Substrato	velocità di reazione				
2	0.02	76	57.58823529	-18.41176471	338.9930796	
3	0.02	47	57.58823529	10.58823529	112.1107266	
4	0.06	97	108.7777778	11.77777778	138.7160494	
5	0.06	107	108.7777778	1.777777778	3.160493827	
6	0.11	123	136.3164557	13.3164557	177.3279923	
7	0.11	139	136.3164557	-2.683544304	7.20141003	
8	0.22	159	160.7313433	1.731343284	2.997549566	
9	0.22	152	160.7313433	8.731343284	76.23635554	
10	0.56	191	180.3421053	-10.65789474	113.5907202	
11	0.56	201	180.3421053	-20.65789474	426.748615	
12	1.1	207	187.6132404	-19.38675958	375.8464471	
13	1.1	200	187.6132404	-12.38675958	153.4318129	
14						
15		195.8			=ma(E2:E13)	
16		0.048				
17						

Regressione non lineare con Excel

- Lo scopo è minimizzare la cella E15 variando i valori B15 e B16
- A tale scopo aprire la componente: **Strumenti Risolutore**
- Si apre una finestra in cui si deve selezionare la cella obiettivo (E15) le celle che devono essere variate (B15 e B16).
- Ricordarsi di scegliere l'opzione "Min"



Regressione non lineare con Excel

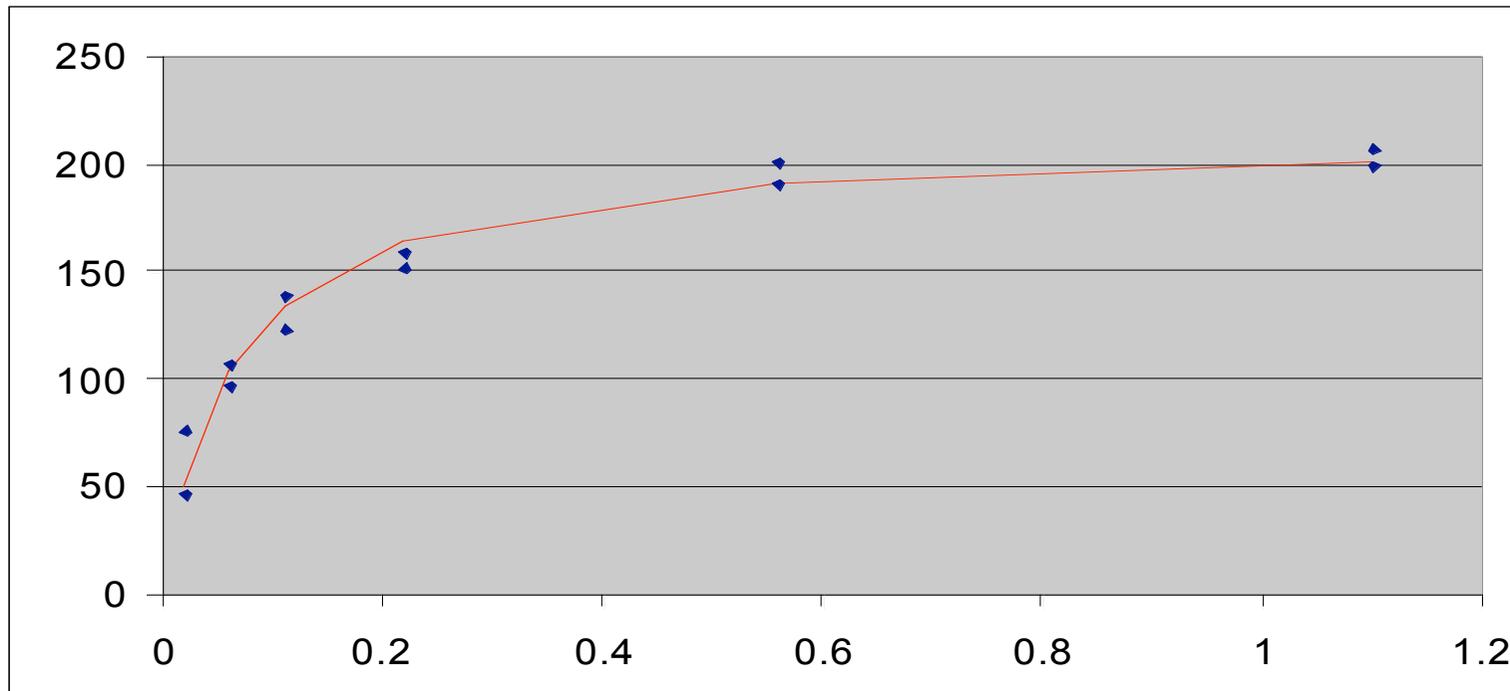
- Una volta eseguito il risolutore i risultati della finestra di Excel cambiano

Da notare che i
valori dei parametri
sono cambiati
rispetto al caso
linearizzato

	A	B	C	D	E
1	Conc Substrato	velocità di reazione			
2	0.02	76	50.56615338	-25.43384662	646.8805537
3	0.02	47	50.56615338	3.566153384	12.71744996
4	0.06	97	102.8111271	5.811127053	33.76919762
5	0.06	107	102.8111271	-4.188872947	17.54665657
6	0.11	123	134.3617152	11.36171521	129.0885726
7	0.11	139	134.3617152	-4.638284786	21.51368576
8	0.22	159	164.684691	5.684691014	32.31571192
9	0.22	152	164.684691	12.68469101	160.9013861
10	0.56	191	190.8327935	-0.167206544	0.027958028
11	0.56	201	190.8327935	-10.16720654	103.3720889
12	1.1	207	200.9686413	-6.03135868	36.37728753
13	1.1	208	200.9686413	0.96864132	0.938266007
14					
15			212.6834454		1195.448815
16			0.064120872		
17					

Regressione non lineare con Excel

- Qualche riflessione
 1. Se abbiamo la pazienza di confrontare i risultati della regressione con i dati sperimentali possiamo osservare che i nuovi parametri si adeguano molto meglio ai punti sperimentali rispetto alla linearizzazione

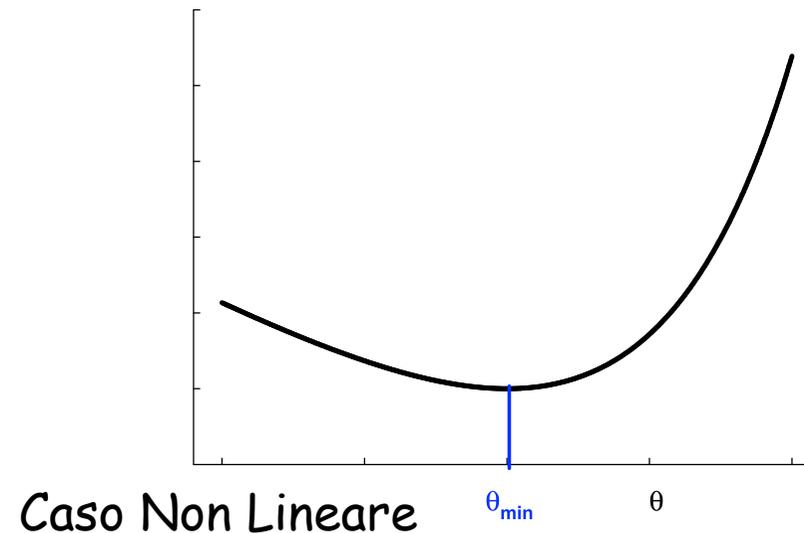
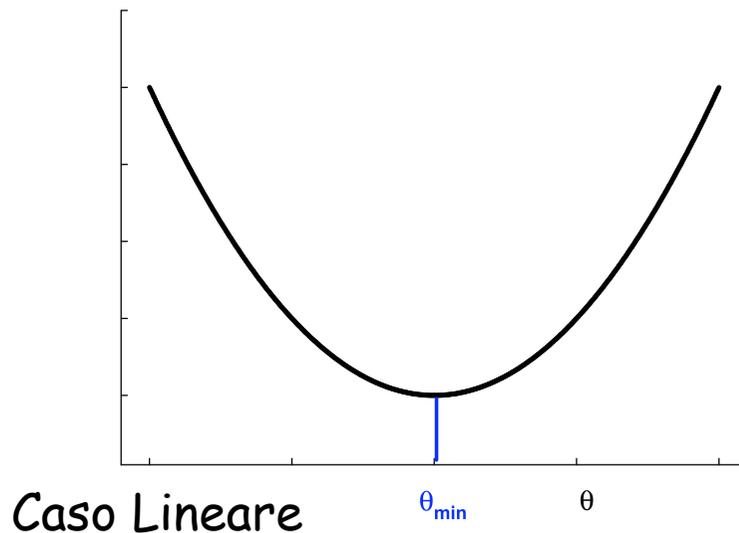


Regressione non lineare con Excel

- Provare a vedere cosa succede se si assegna come valore di primo tentativo i valori $\mu = -100$ e $L = 200$.
- I risultati che ottengo dal risolutore sono plausibili oppure no?
- Il ruolo del valore di **primo tentativo** è **importante**: i valori forniti dalla linearizzazione rappresentano una buona scelta
- Nei prossimi lucidi cercheremo di introdurre qualche concetto di teoria che può aiutare a farci capire i risultati che si sono ottenuti

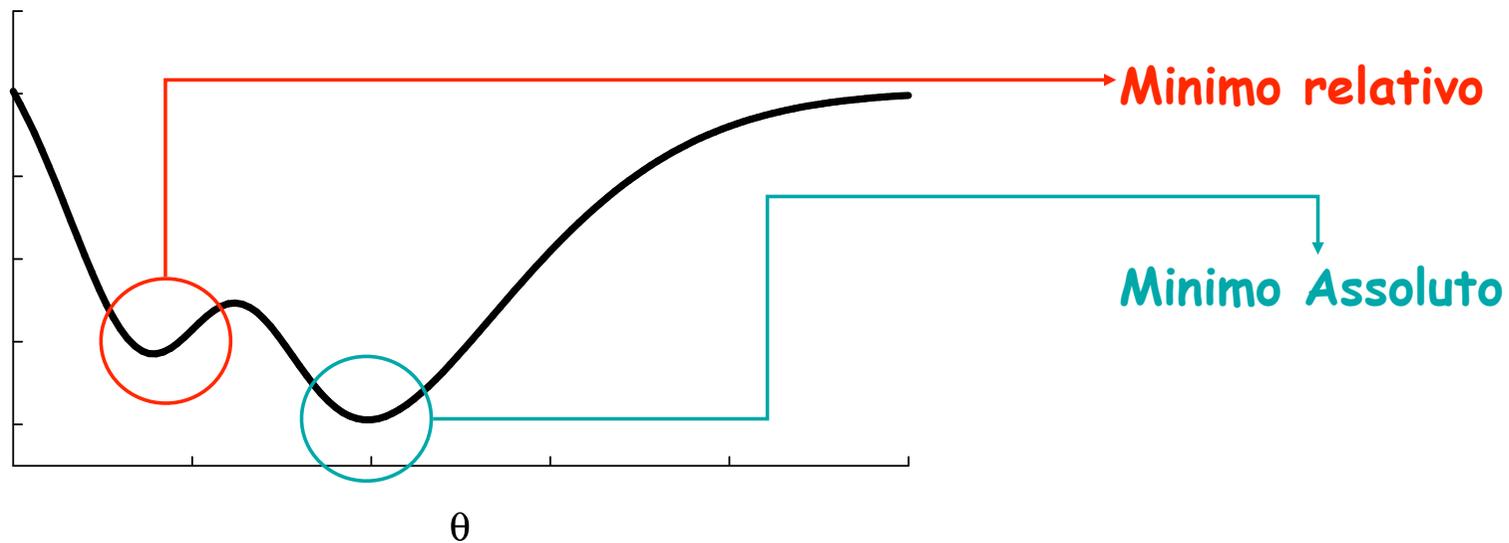
Algoritmi di ottimizzazione

- Si è già visto come la stima dei parametri per il modello non lineare consista nella ricerca del minimo della funzione obiettivo
- Non abbiamo più la possibilità di una soluzione analitica: il diagramma della funzione $\Phi(\theta)$ non sarà più un paraboloide come nel caso della regressione lineare.
- Per esempio, in un caso scalare ($\theta \in \mathbb{R}$) si può avere:



Algoritmi di ottimizzazione

- Si possono avere casi più complicati con la presenza di più minimi nella funzione obiettivo:



- In questo caso è necessario individuare il minimo **assoluto**

Algoritmi di ottimizzazione

- La ricerca del minimo di una funzione obiettivo è effettuata per via **numerica** utilizzando metodi **iterativi**:
- Si parte da un **valore di primo tentativo** θ_0 e si genera una successione di valori

$$\theta_0, \theta_1, \theta_2, \dots, \theta_n$$

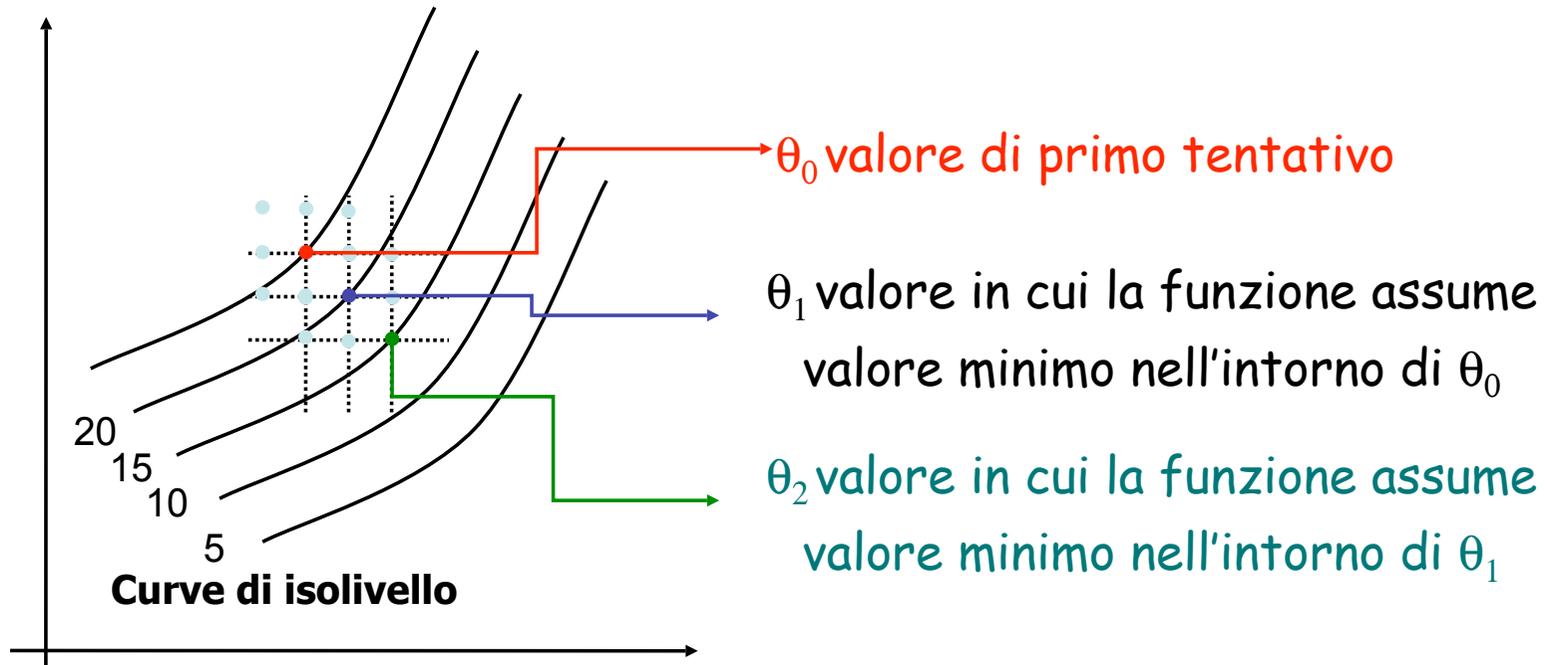
- Che si **spera** converga verso il minimo
- Quando i valori della successione θ_i non cambiano più in modo significativo (ovvero sono uguali a meno di una fissata **tolleranza**) il metodo è giunto a convergenza
- Non è garantito che si sia giunti al minimo assoluto!!! (può essere che si è in presenza di un minimo relativo)

Le stime θ ottenute per un modello non lineare non sono più variabili aleatorie di tipo Gaussiano

Algoritmi di ottimizzazione – Metodi diretti

- I metodi diretti sono i più semplici da implementare.
- Si fissa una griglia nello spazio dei parametri e si cerca nell'intorno del valore i -esimo della successione quale è il punto dell'intorno che abbia valore minimo.

Esempio: funzione obiettivo a due parametri



Algoritmi di ottimizzazione – Metodi diretti

- Caratteristiche dei metodi diretti:
- **Vantaggi**
 - È necessaria la conoscenza della sola funzione obiettivo.
 - Risulta molto semplice da implementare
- **Svantaggi**
 - Può rivelarsi molto oneroso dal punto di vista computazionale: per ogni iterazione è necessaria la valutazione della funzione in (3^n-1) punti, dove n è la dimensione dello spazio delle variabili di stato (nel caso in esame $n = 2$).

ALGORITMI DI MINIMIZZAZIONE

RIPETIZIONE DI ANALISI

- Che cosa è il gradiente di una funzione?
- Il gradiente di una funzione $f(\theta)$ è il vettore \mathbf{g} delle sue derivate prime.

$$\underline{\mathbf{g}} = \begin{pmatrix} \frac{\partial f}{\partial \theta_1} \\ \vdots \\ \frac{\partial f}{\partial \theta_n} \end{pmatrix}$$

- Il gradiente \mathbf{g} di una funzione f indica la direzione lungo la quale la funzione cresce più rapidamente. Perciò $-\mathbf{g}$ è la direzione lungo cui la funzione **diminuisce** più velocemente.

ALGORITMI DI MINIMIZZAZIONE

RIPETIZIONE DI ANALISI

- Che cosa è l'Hessiano di una funzione?
- L'Hessiano di una funzione $f(\theta)$ è la matrice **H** delle sue derivate seconde.

$$\underline{\underline{H}} = \begin{pmatrix} \frac{\partial^2 f}{\partial \theta_1^2} & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_n} \\ \frac{\partial^2 f}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 f}{\partial \theta_2^2} & \dots & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \theta_n \partial \theta_1} & \frac{\partial^2 f}{\partial \theta_n \partial \theta_2} & \dots & \frac{\partial^2 f}{\partial \theta_n^2} \end{pmatrix}$$

- L'Hessiano è una matrice simmetrica.

ALGORITMI DI MINIMIZZAZIONE

RIPETIZIONE DI ANALISI

- Condizioni necessaria e sufficiente affinché θ^* sia un minimo:
- Condizione **NECESSARIA**: $\mathbf{g}(\theta^*) = \mathbf{0}$
- Condizione **SUFFICIENTE**: $\mathbf{H}(\theta^*)$ deve essere positiva definita.

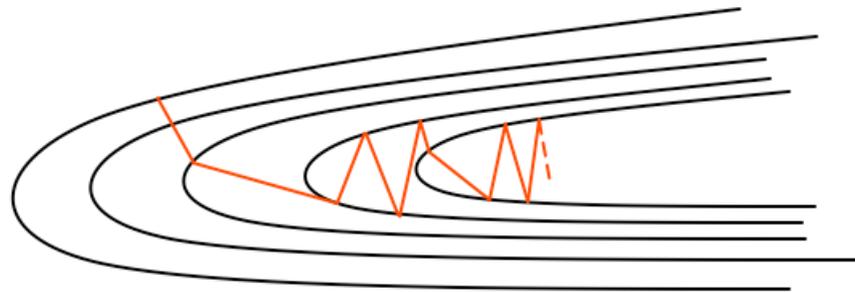
METODI NUMERICI ITERATIVI

- Si parte da una stima di primo tentativo e si itera in modo da avvicinarsi al minimo assoluto.

ALGORITMI DI MINIMIZZAZIONE

Problemi storici:

- Scelta della direzione: il gradiente.
 - In vicinanza del minimo si usa la curvatura della funzione (Hessiano) per accelerare la convergenza
 - Problemi di occupazione di memoria.
 - **MOLTI METODI NON CONVERGEVANO** (i problemi nascono quando si è lontani dalla soluzione)
- I metodi falliscono quando si hanno valli strette



ALGORITMI DI MINIMIZZAZIONE

- Non esiste il metodo migliore per ogni problema di minimizzazione.
- Un buon programma deve sfruttare più metodi selezionati tra quelli adatti a risolvere un particolare problema.
- La scelta dell'algoritmo da utilizzare è legata a:
 - Robustezza:
 - deve risolvere problemi con valli molto strette.
 - Deve trovare il minimo assoluto
 - Accuratezza
 - Efficienza
 - Deve convergere velocemente al minimo assoluto
 - Scarsa occupazione di memoria

ALGORITMI DI MINIMIZZAZIONE

- Nei problemi di regressione non lineare
 - Il numero delle variabili è ridotto
 - Si è in presenza di valli strette
 - Il minimo assoluto deve essere localizzato accuratamente
 - La funzione da minimizzare è una somma di quadrati
 - Si può essere in presenza di più minimi.

METODI

- Metodi che richiedono l'uso del gradiente
- Metodi che richiedono gradiente ed Hessiano

Algoritmi di ottimizzazione – Metodi del gradiente

- La ricerca del minimo risulterebbe meno onerosa se si avessero informazioni sulla “direzione” lungo cui muoversi nello spazio dei parametri.

dove

$$\underline{\theta}_{i+1} = \underline{\theta}_i + \underline{c}_i$$

$$\underline{c}_i = \lambda_i \underline{v}_i$$

Modulo del
passo i-esimo

Direzione del
passo i-esimo

**Lungo la direzione
fissata il problema
diventa di natura
scalare**

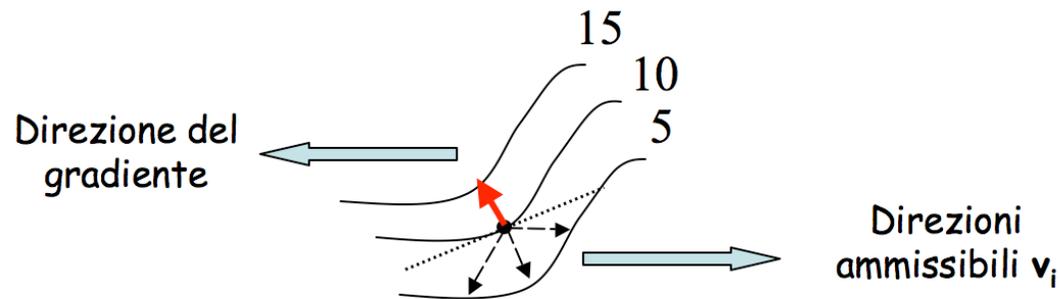
**Il valore λ può quindi
essere determinato con
molta accuratezza**

Algoritmi di ottimizzazione – Metodi del gradiente

- Il gradiente della funzione obiettivo

$$\frac{\partial \phi}{\partial \theta} \Big|_{\theta_i}$$

- Rappresenta la direzione di massima crescita della funzione nello spazio delle variabili θ .
- Quindi, per procedere verso il minimo si deve comunque scegliere una qualunque direzione opposta alla direzione del gradiente



Algoritmi di ottimizzazione – Metodi del gradiente

- La direzione \mathbf{v}_i può essere scritta nella seguente forma:

$$\underline{v}_i = -\mathbf{R}_i \frac{\partial \phi}{\partial \underline{\theta}} \Big|_{\underline{\theta}_i}$$

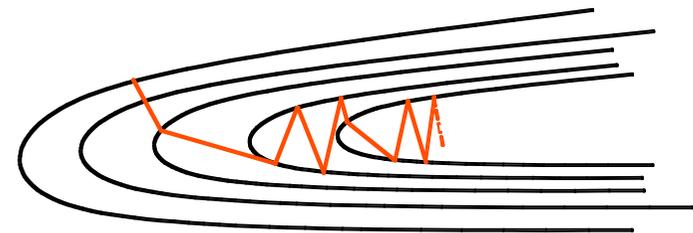
- Il tensore \mathbf{R} è un tensore **definito positivo** che “modula” opportunamente la direzione di $\underline{\theta}_i$.
- Il vettore \mathbf{v} ha un ruolo più importante di λ . Infatti, una volta scelto \mathbf{v} si può scegliere λ con grandissima precisione

Metodi del Gradiente – Metodo Massima Pendenza

- In inglese: **Steepest Descent Method**
- Nel metodo della massima pendenza si adotta $\mathbf{R} = \mathbf{I}$ (matrice identità) e λ fissato (da notare che sarà spesso necessario adottare valori piccoli di λ per problemi di stabilità). Il metodo si traduce quindi nella formula:

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \lambda \frac{\partial \phi}{\partial \underline{\theta}} \Big|_{\underline{\theta}_i}$$

- **Vantaggi:**
 - Molto semplice da implementare.
- **Svantaggi:**
 - E' estremamente lento nella convergenza ed instabile.



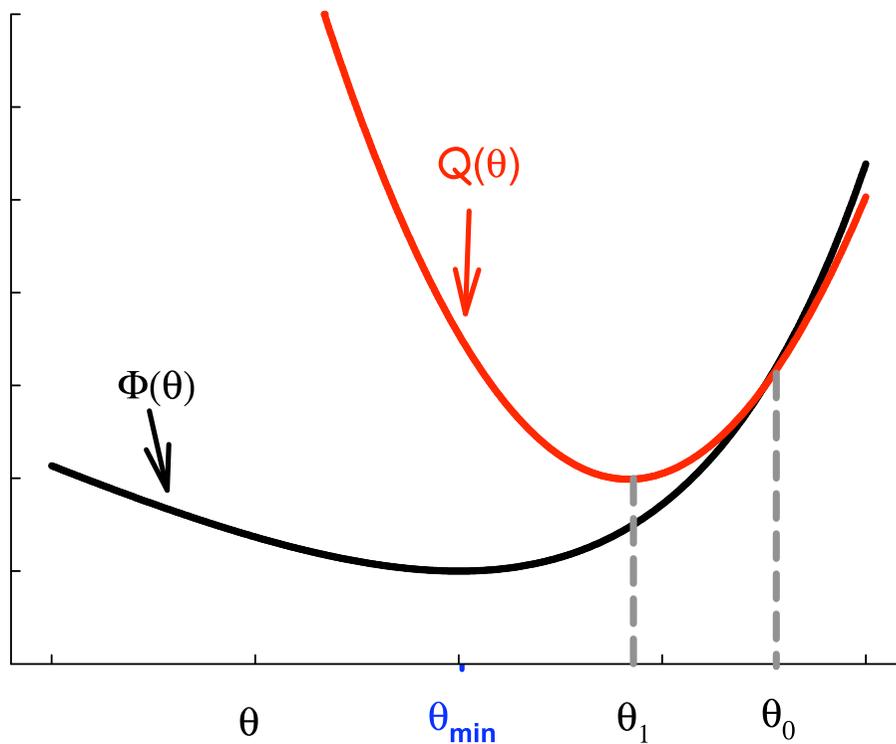
Metodi del Gradiente – Metodo di Newton-Raphson

- Il metodo della massima pendenza può essere migliorato con una opportuna scelta della matrice **R**.
- È possibile approssimare la funzione obiettivo con un'altra funzione il cui minimo è molto più semplice da calcolare.
- Per semplicità si incominci a considerare un caso scalare e si sviluppi la funzione obiettivo in serie di Taylor intorno al punto generico θ_i delle iterazioni:

$$\phi(\underline{\theta}) \approx Q(\underline{\theta}) = \phi(\underline{\theta}_i) + \frac{\partial \phi}{\partial \underline{\theta}} \Big|_{\underline{\theta}_i} (\underline{\theta} - \underline{\theta}_i) + \frac{1}{2} \frac{\partial^2 \phi}{\partial \underline{\theta}^2} \Big|_{\underline{\theta}_i} (\underline{\theta} - \underline{\theta}_i)^2$$

Metodi del Gradiente – Metodo di Newton-Raphson

- Dal punto di vista grafico, si approssima la funzione con una curva quadratica passante per θ_i :



La curva $Q(\theta)$ è la quadratica che approssima meglio la funzione in θ_0

Il nuovo valore di tentativo θ_1 è il valore in cui $Q(\theta)$ assume valore minimo

Metodi del Gradiente – Metodo di Newton-Raphson

- Per trovare il minimo, si può derivare la $Q(\theta)$ rispetto a θ :

$$\frac{\partial Q}{\partial \theta} = \frac{\partial \phi}{\partial \theta} \Big|_{\theta_i} + \frac{\partial^2 \phi}{\partial \theta^2} \Big|_{\theta_i} (\theta - \theta_i) = 0$$

- Da cui, con semplici passaggi, si perviene al valore θ_{i+1} della successione

$$\theta_{i+1} = \theta_i - \frac{1}{\frac{\partial^2 \phi}{\partial \theta^2} \Big|_{\theta_i}} \cdot \frac{\partial \phi}{\partial \theta} \Big|_{\theta_i}$$

Metodi del Gradiente – Metodo di Newton-Raphson

- Il caso precedente può essere concettualmente esteso senza problemi al caso vettoriale:

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \mathbf{H}_{\theta_i}^{-1} \cdot \frac{\partial \phi^T}{\partial \theta_{\theta_i}}$$

- con H indichiamo l'**Hessiano** della funzione Φ

ALGORITMI DI MINIMIZZAZIONE

Metodo di Newton

- **Vantaggi** del metodo di Newton
 - Quando converge lo fa molto rapidamente
 - Se l'Hessiano è positivo definito la funzione diminuisce ad ogni iterazione.
- **Svantaggi** del metodo di Newton
 - Il metodo va in crisi quando l'Hessiano non è positivo definito
 - Il metodo può non convergere
 - Il metodo non è robusto
 - Ad ogni iterazione deve essere calcolato l'Hessiano
 - $N(N+1)/2$ valutazioni di funzioni
 - Ad ogni iterazione deve essere risolto un sistema di equazioni lineari

- **Varianti del metodo di Newton vettoriale**
- **Metodi di Newton modificato:**
 - L'Hessiano viene calcolato (numericamente o analiticamente) ad ogni iterazione
 - Viene risolto un sistema lineare ad ogni iterazione
 - L'Hessiano viene modificato in modo da evitare i problemi
- **Metodi di quasi Newton:**
 - L'Hessiano viene aggiornato ad ogni iterazione sommandogli una opportuna matrice.
 - Viene evitata la soluzione di un sistema di equazioni lineari
 - L'Hessiano viene modificato in modo da evitare eventuali problemi.

Metodi del Gradiente – Metodo LM

- Per ovviare al problema della gestione dell'eventualità di matrici malcondizionate la matrice \mathbf{H} può essere modificata, con l'introduzione di un opportuno fattore di condizionamento k :
- Metodo di Levenberg (1944)

$$\mathbf{Q} = \mathbf{H} + k\mathbf{I} \quad \mathbf{R} = \mathbf{Q}^{-1}$$

- Metodo di Marquardt (1963)

$$\mathbf{Q} = \mathbf{H} + k\mathbf{D} \quad \mathbf{R} = \mathbf{Q}^{-1}$$

– dove \mathbf{D} è una matrice diagonale

Riferimenti bibliografici di interesse

- L'implementazione della regressione non lineare con Excel è ispirata dal libro
 - **Introduction to Chemical Engineering Computing, B. Finlayson, Wiley & Sons, New Jersey, 2006**
- Ottimi libri di riferimento per la regressione non lineare sono:
- Un testo un po' datato ma tuttora valido e di lettura relativamente semplice è:
 - **Nonlinear parameter Estimation, Bard Y., Academic Press, 1974**
- Un ottimo testo, anche se non di facile lettura è il seguente testo:
 - **Nonlinear Regression Analysis and Its Applications, Bates D. and Watts D., Wiley & Sons, New York, 1988**

Sommario

- Modelli non lineari linearizzabili
 - Attenzione ad usare i MQ (i risultati sono buoni come primo tentativo)
 - Meglio MQP
- Ricerca del minimo (o massimo) per via iterativa
 - Codici numerici
 - Problemi di convergenza
 - Problema del primo tentativo
 - Necessità di tentare molti primi tentativi alla ricerca del minimo assoluto