Ottimizzazione

Marco Caliari
Dipartimento di Informatica
Università di Verona

Simone Zuccher
Liceo Scientifico Statale "E. Medi"
Villafranca di Verona

Piano Lauree Scientifiche, a.s. 2014–2015

Capitolo 1

Ottimizzazione unidimensionale

1.1 Massimi e minimi di funzione

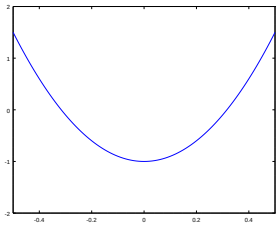
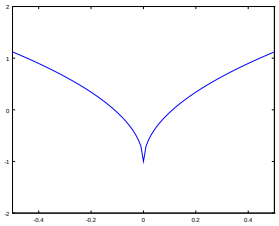
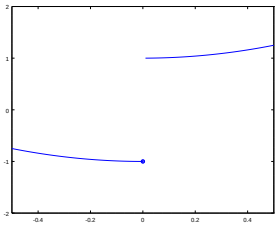
1.2 Funzioni unimodali

Una funzione $f\colon [a,b] \to \mathbb{R}$ è detta *strettamente unimodale* se esiste $t^* \in [a,b]$ tale che

$$f(t^*) = \min\{f(t)\colon t \in [a,b]\}$$

e se per ogni $a \leq t_1 < t_2 \leq b$ si ha

$$t_2 \leq t^* \Rightarrow f(t_1) > f(t_2)$$
$$t^* \leq t_1 \Rightarrow f(t_2) > f(t_1)$$

Vale il seguente risultato: se f è una funzione strettamente unimodale e $a \leq t_1 < t_2 \leq b$, allora

$$f(t_1) > f(t_2) \Rightarrow t^* \in (t_1, b) \tag{1.1a}$$
$$f(t_1) = f(t_2) \Rightarrow t^* \in (t_1, t_2) \tag{1.1b}$$
$$f(t_1) < f(t_2) \Rightarrow t^* \in (a, t_2) \tag{1.1c}$$

Un intervallo che contiene il minimo di f è detto *intervallo di incertezza*. Il risultato sopra dice che a partire dall'intervallo di incertezza $[a, b]$, è possibile ridurlo in $[t_1, b]$, $[t_1, t_2]$ oppure $[a, t_2]$ mediante *due* valutazioni della funzione f (in t_1 e in t_2). Con successive valutazione della funzione f si ridurrà ulteriormente l'intervallo di incertezza. I metodi di minimizzazione che consideriamo riescono a trovare un intervallo di incertezza di ampiezza minore di δ, δ fissato e piccolo a piacere, dentro il quale si trova il minimo della funzione.

1.3 Metodo della bisezione

Nel metodo della bisezione si cercano t_1 e t_2 in modo che il successivo intervallo di incertezza, qualunque ipotesi sia soddisfatta delle tre (1.1) abbia ampiezza non più di metà di quello precedente. Bisognerebbe allora scegliere $t_1 = t_2$, ma allora non avremmo a disposizione le due valutazioni di funzione. Se δ è l'ampiezza massima dell'intervallo di certezza, si potrebbe scegliere allora $t_1 = (a+b)/2 - \delta/2$ e $t_2 = (a+b)/2 + \delta/2$. In tal modo, l'intervallo di incertezza sarebbe "quasi" dimezzato (se $f(t_1) > f(t_2)$ o se $f(t_1) < f(t_2)$) o addirittura ridotto all'intervallo di incertezza voluto (se $f(t_1) = f(t_2)$). Nella pratica, si considerano solo i due casi

$$f(t_1) > f(t_2) \Rightarrow t^* \in (t_1, b)$$
$$f(t_1) \leq f(t_2) \Rightarrow t^* \in (a, t_2)$$

e si usa, per sicurezza, $\delta/3$ invece di $\delta/2$. In tal modo, l'intervallo di incertezza iniziale $[a_1, b_1] = [a, b]$ viene ridotto all'intervallo $[a_2, b_2] = [t_1, b_1]$ oppure $[a_2, b_2] = [a_1, t_2]$.

1.4 Metodo della sezione aurea

Siano dati $a_1 = a$, $b_1 = b$ e $x_1 \in (a, b)$. Cerchiamo un punto $y_1 > x_1$, in modo che, all'iterazione successiva, l'intervallo di incertezza sia $[a_2, b_2] = [x_1, b_1]$ (se $f(x_1) > f(y_1)$) oppure $[a_2, b_2] = [a_1, y_1]$ (se $f(x_1) \leq f(y_1)$). Cerchiamo tale punto in modo che si abbia la stessa riduzione relativa dell'intervallo nei due casi, cioè

$$\frac{y_1 - a_1}{b_1 - a_1} = \frac{b_1 - x_1}{b_1 - a_1} = r \tag{1.2}$$

Supponiamo di essere nell'intervallo $[a_2, b_2] = [x_1, b_1]$, chiamiamo $x_2 = y_1$ e cerchiamo $y_2 > x_2$ con lo stesso criterio di prima (stessa riduzione dell'inter-

vallo) ed esattamente con lo stesso fattore di riduzione

$$\frac{y_2 - a_2}{b_2 - a_2} = \frac{b_2 - x_2}{b_2 - a_2} = r \tag{1.3}$$

Da (1.2) si ha

$$\frac{b_1 - y_1}{b_1 - x_1} = \frac{x_1 - a_1}{b_1 - x_1} = \frac{b_1 - a_1}{b_1 - x_1} - 1 = \frac{1}{r} - 1$$

e quindi da (1.3)

$$r = \frac{b_2 - x_2}{b_2 - a_2} = \frac{b_1 - y_1}{b_1 - x_1} = \frac{1}{r} - 1$$

Pertanto $r = (\sqrt{5} - 1)/2$. Se invece siamo nell'intervallo $[a_2, b_2] = [a_1, y_1]$, chiamiamo $y_2 = x_1$. Cerchiamo $x_2 < y_2$ con lo stesso criterio (1.3). Da (1.2) si ha

$$\frac{x_1 - a_1}{y_1 - a_1} = \frac{b_1 - y_1}{y_1 - a_1} = \frac{b_1 - a_1}{y_1 - a_1} - 1 = \frac{1}{r} - 1$$

e quindi da (1.3)

$$r = \frac{y_2 - a_2}{b_2 - a_2} = \frac{x_1 - a_1}{y_1 - a_1} = \frac{1}{r} - 1$$

e dunque r è lo stesso. Da (1.2) si ricavano x_1 e y_1 e da (1.3) x_2 oppure y_2.

La riduzione dell'intervallo di incertezza è dunque pari ad $r \approx 0.62$, dunque minore della riduzione che si ottiene con il metodo di bisezione. Ma ogni riduzione avviene con *una sola* valutazione della funzione f.

1.5 Approssimazione mediante parabole

Una tecnica diversa dalle precedenti consiste nell'approssimare la funzione di cui trovare il minimo con una funzione più semplice, per esempio una parabola. Si può procedere in questo modo: dati i punti a, b e $x_1 = (a+b)/2$, si trova l'ascissa del vertice, diciamo y_1, della parabola passante per i tre punti. Poniamo adesso $t_1 = \min\{x_1, y_1\}$ e $t_2 = \max\{x_1, y_1\}$. Se $f(t_1) > f(t_2)$, allora $t^* \in (t_1, b)$ e dunque si può considerare la parabola passante per t_1, t_2 e b, altrimenti si può considerare la parabola passante per a, t_1 e t_2 e continuare iterativamente.

Questo metodo è in generale più efficiente dei due visti, ma genera anche situazioni più difficili da controllare:

1. y_1 potrebbe coincidere con x_1. Potrebbe voler dire che abbiamo trovato il minimo, ma anche no (vedi Figura 1.1). Come capire di fermarsi o come proseguire?

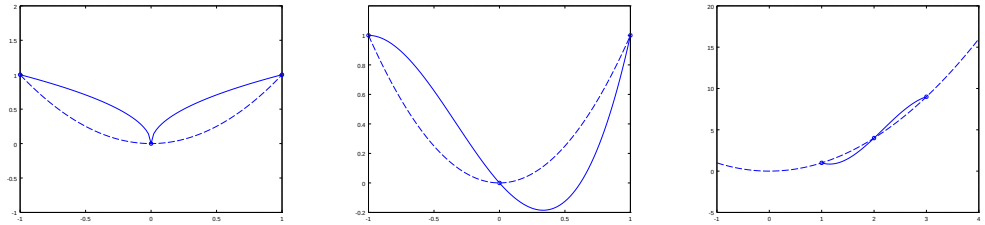


Figura 1.1: Approssimazione con parabola.

2. y_1 potrebbe cadere al di fuori dell'intervallo $[a, b]$ (vedi Figura 1.1). Come proseguire?

1.6 Minimizzazione bidimensionale

Le funzioni possono dipendere anche da due variabili indipendenti e la ricerca del minimo, in questo caso, è molto più difficile in generale. Per fortuna non sempre: data la funzione $z = g(x, y) = x^2 + y^2$ è facile scoprire che il minimo si ha nel punto $(0, 0)$, poiché solo in quel punto la funzione vale 0 e in tutti gli altri assume un valore maggiore.

È possibile estendere il metodo della sezione aurea al caso bisimensionale? Sì, procedendo per direzioni. Si deve scegliere un "ragionevole" punto x_0 e poi trovare il minimo della funzione di una variabile

$$f_1(y) = g(x_0, y)$$

usando il metodo della sezione aurea e quindi, in particolare, fissare un intervallo di incertezza $[y_{\min}, y_{\max}]$. Una volta trovato, diciamo y_0, possiamo considerare la funzione di una variabile

$$f_2(x) = g(x, y_0)$$

e trovarne il minimo, diciamo x_1, in un intervallo di incertezza $[x_{\min}, x_{\max}]$. E continuare così. Questo metodo si chiama *metodo di discesa per direzioni coordinate*. Poiché gli intervalli di incertezza per ogni variabile sono sempre gli stessi, bisogna trovare il modo di terminare l'algoritmo. Un *test di arresto* comune prevede di valutare la differenza tra $[x_n, y_n]$ e $[x_{n+1}, y_{n+1}]$ per esempio tramite la loro distanza Euclidea

$$\delta_{n+1} = \sqrt{(x_{n+1} - x_n)^2 + (y_{n+1} - y_n)^2}$$

Quando δ_{n+1} è minore di una tolleranza prefissata δ, si termina l'algoritmo.

Capitolo 2

Best fit

In questo capitolo ci occupiamo di determinare la forma analitica di una funzione $f : \mathbb{R} \to \mathbb{R}$ che rappresenti al meglio una nuvola di N coppie di dati $(x_i; y_i)$ con $i \in \mathbb{N},\ 1 \leq i \leq N$ derivanti, per esempio, da degli esperimenti.

Come può essere "misurata" la bontà o meno di una certa approsimazione $y = f(x)$? Evidentemente l'approssimazione è tanto migliore quanto più la differenza tra y_i e $f(x_i)$ è piccola. Al limite, se l'approsimazione fosse "perfetta", la funzione $y = f(x)$ passerebbe per tutti i punti $(x_i; y_i)$ per cui per ogni punto si avrebbe $y_i = f(x_i)$. Questo tipo di approsimazione viene detto *interpolazione*.

Qui non ci interessa l'interpolazione dei dati $(x_i; y_i)$, quanto piuttosto di determinare una funzione che li rappresenti adeguatamente. Consideriamo come misura della bontà dell'approssimazione la quantità

$$E_f = \sum_{i=1}^{N} \left[y_i - f(x_i) \right]^2 .$$

Perché proprio questa scelta? Certamente la somma degli scarti non può andare bene in quanto, data la nuvola di punti qualunque retta passante per $\bar{x} = (\sum_{i=1}^{N} x_i)/N$ e per $\bar{y} = (\sum_{i=1}^{N} y_i)/N$ rende nulla la sommma degli scarti. La somma dei valori assoluti degli scarti è più plausibile, cosí come la somma di qualunque funzione pari degli scarti. Si sceglie la somma dei quadrati degli scarti come funzione da minimizzare perché la retta che ne risulta, che è una sorta di "retta media" ha la proprietà di cui gode anche la media aritmetica. Infatti, supponiamo di fare N misure ripetute di una lunghezza di un banco. Detti y_i i valori misurati, un indicatore della misura cercata è la media aritmetica

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

Tale valore minimizza la somma dei quadrati degli scarti. Infatti la funzione di y

$$\sum_{i=1}^{N}(y_i - y)^2 = \sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} 2y_i\right) y + Ny^2$$

è una parabola il cui vertice corrisponde a $y = -b/(2a) = (\sum_{i=1}^{N} y_i)/N = \bar{y}$. Interpretando questo solo come il caso particolare della ricerca della miglior retta orizzontale, è naturale minimizzare la somma dei quadrati degli scarti.

2.1 La retta dei minimi quadrati

La funzione più semplice cui si possa pensare, diversa da una costante, è un polinomio di primo grado, ossia una retta del tipo $f(x) = mx + q$. La bontà dell'approssimazione si misura, quindi, con la funzione

$$E(m, q) = \sum_{i=1}^{N} \left[y_i - (mx_i + q)\right]^2, \tag{2.1}$$

dove si è messo in evidenza il fatto che E dipende da due variabili reali, m e q.

L'obiettivo è minimizzare $E(m, q)$, ossia determinare i valori m e q che rendono minima E. Si osservi che, essendo E la somma di quadrati, il suo minimo non può che essere una quantità positiva.

Per determinare simultaneamente la coppia (m, q) che minimizza $E(m, q)$ basta osservare che, fissato $q = \bar{q}$, la (2.1) si riduce a

$$E_m = E(m, \bar{q}) = \sum_{i=1}^{N} \left[y_i - (mx_i + \bar{q})\right]^2,$$

che è una parabola nell'incognita m, mentre fissato $m = \bar{m}$, la (2.1) si riduce a

$$E_q = E(\bar{m}, q) = \sum_{i=1}^{N} \left[y_i - (\bar{m}x_i + q)\right]^2,$$

che è una parabola nell'incognita $\bar{m}$. Ricordando che una parabola del tipo $y = ax^2+bx+c$, con $a > 0$, assume il proprio valore minimo in corrispondenza di $x = -b/(2a)$, si possono ottenere il valore di m in corrispondenza del quale E_m è minima ed il valore q in corrispondenza del quale E_q è minima. Mettendo a sistema queste due condizioni, nel caso il sistema sia determinato, si ottengono i valori $(\bar{m}; \bar{q})$ che garantiscono il minimo di $E(m, q)$.

Fissando $q = \bar{q}$ si ha

$$\begin{aligned}
E_m &= \sum_{i=1}^{N} \left[y_i - (mx_i + \bar{q})\right]^2 \\
&= \sum_{i=1}^{N} \left[y_i^2 + m^2 x_i^2 + \bar{q}^2 - 2mx_i y_i - 2\bar{q}y_i + 2m\bar{q}x_i\right] \\
&= \sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} m^2 x_i^2 + \sum_{i=1}^{N} \bar{q}^2 - \sum_{i=1}^{N} 2mx_i y_i - \sum_{i=1}^{N} 2\bar{q}y_i + \sum_{i=1}^{N} 2m\bar{q}x_i \\
&= \sum_{i=1}^{N} y_i^2 + m^2 \sum_{i=1}^{N} x_i^2 + N\bar{q}^2 - 2m \sum_{i=1}^{N} x_i y_i - 2\bar{q} \sum_{i=1}^{N} y_i + 2m\bar{q} \sum_{i=1}^{N} x_i \\
&= m^2 \sum_{i=1}^{N} x_i^2 - 2m \left(\sum_{i=1}^{N} x_i y_i - \bar{q} \sum_{i=1}^{N} x_i \right) + \sum_{i=1}^{N} y_i^2 + N\bar{q}^2 - 2\bar{q} \sum_{i=1}^{N} y_i,
\end{aligned} \tag{2.2}$$

da cui il valore di m che assicura il minimo di $E_m(m)$:

$$m = -\frac{-2 \left(\sum_{i=1}^{N} x_i y_i - \bar{q} \sum_{i=1}^{N} x_i \right)}{2 \sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N} x_i y_i - \bar{q} \sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2} \tag{2.3}$$

Svolgendo calcoli analoghi su $E_q = E(\bar{m}, q)$ (che lasciamo allo studente diligente), si ottiene

$$E_q = Nq^2 - 2q \left(\sum_{i=1}^{N} y_i - \bar{m} \sum_{i=1}^{N} x_i \right) + \sum_{i=1}^{N} y_i^2 + \bar{m}^2 \sum_{i=1}^{N} x_i^2 - 2\bar{m} \sum_{i=1}^{N} x_i y_i, \tag{2.4}$$

da cui il valore di q che assicura il minimo di $E_q(q)$:

$$q = -\frac{-2 \left(\sum_{i=1}^{N} y_i - \bar{m} \sum_{i=1}^{N} x_i \right)}{2N} = \frac{\sum_{i=1}^{N} y_i - \bar{m} \sum_{i=1}^{N} x_i}{N}. \tag{2.5}$$

In definitiva, la coppia $(\bar{m}, \bar{q})$ che minimizza $E(m, q)$ si determina risolvendo

il sistema

$$\begin{cases} \bar{m} = \dfrac{\displaystyle\sum_{i=1}^{N} x_i y_i - \bar{q} \sum_{i=1}^{N} x_i}{\displaystyle\sum_{i=1}^{N} x_i^2} \\ \bar{q} = \dfrac{\displaystyle\sum_{i=1}^{N} y_i - \bar{m} \sum_{i=1}^{N} x_i}{N} \end{cases}$$

che ha come unica soluzione

$$\begin{cases} \bar{m} = \dfrac{N \displaystyle\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \displaystyle\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2} \\ \bar{q} = \dfrac{\displaystyle\sum_{i=1}^{N} y_i \sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i}{N \displaystyle\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2} \end{cases}$$