A Richiami sulle norme

Una funzione $\|\cdot\| : R^n \to R$ si definisce *norma vettoriale*,se soddisfa le seguenti proprietà:

- $\|v\| \geq 0$ per ogni $v \in R^n$;
- $\|v\| = 0$ se e solo se $v = 0$;
- $\|\alpha v\| = |\alpha| \|v\|$ per ogni $v \in R^n$ e $\alpha \in R$;
- $\|v + w\| \leq \|v\| + \|w\|$ per ogni $v, w \in R^n$ (disguaglianza triangolare).

Le norme più frequentemente utilizzate sono le norme ℓ_p (o norme di Hölder) indicate con $\|\cdot\|_p$ e definite come

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}}$$

con $1 \leq p < \infty$. Per $p = 2$ si ottiene la *norma euclidea*

$$\|v\|_2 = \left(\sum_{i=1}^{n} |v_i|^2 \right)^{\frac{1}{2}}$$

che è quella maggiormente usata. Si può anche definire una norma ℓ_∞ come

$$\|v\|_\infty = \max_{1 \leq i \leq n} |v_i|.$$

Vale sempre la seguente relazione:

$$\|v\|_2 \leq \|v\|_\infty.$$

Nel caso di norma euclidea vale la *disuguaglianza di Schwartz,* ovvero

$$|v^T w| \leq \|v\|_2 \|w\|_2.$$

Consideriamo ora il caso di norme di matrici. Indichiamo con M_n l'insieme di tutte le matrici $n \times n$[4].

Assegnata una matrice

$$A = (a_{ij})$$

in M_n, una norma di A può essere introdotta sia considerando A come un insieme di n^2 elementi a_{ij}, sia considerando A come un operatore lineare da R^n in R^n.

Nel primo caso possiamo ovviamente definire come norma di A una qualsiasi norma vettoriale relativa agli elementi di A. Un esempio di norma interpretabile come norma del vettore degli elementi è costituito dalla *norma di Frobenius* di A, data da:

$$\|A\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$$

[4]Ci riferiamo, per semplicità al caso di matrici quadrate, anche se le definizioni esposte nel seguito si estendono in modo immediato al caso di matrici rettangolari.

Se $A, B \in M_n$ si ha:

$$\|A\ B\|_F \leq \|A\|_F\ \|B\|_F.$$

Indicando con $\mathrm{Tr}(A)$ la *traccia* della matrice A, ossia:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii},$$

la norma di Frobenius si può anche esprimere nella forma:

$$\|A\|_F = \left(\mathrm{Tr}(A^T A)\right)^{1/2}.$$

Se A è pensata come un operatore lineare, si può definire la norma di A ponendo

$$\|A\| = \sup_{x \in R^n, x \neq 0} \frac{\|Ax\|}{\|x\|}, \tag{120}$$

o, equivalentemente:

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

Supponendo, per semplicità, che venga usata la stesssa norma vettoriale $\|\cdot\|$ sia per x che per Ax, la norma matriciale definita dalla (120) si dice *norma indotta* dalla norma vettoriale considerata.

Ponendo:

$$\|A\|_p = \sup_{x \in R^n, x \neq 0} \frac{\|Ax\|_p}{\|x\|_p},$$

si ha, in particolare:

$$\|A\|_1 = \max_{1 \leq i \leq n} \sum_{i=1}^{n} |a_{ij}|,$$

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|,$$

$$\|A\|_2 = \left(\lambda_M(A^T A)\right)^{1/2},$$

essendo $\lambda_M(A^T A)$ il massimo autovalore di $A^T A$. Se A è una matrice simmetrica risulta ovviamente

$$\|A\|_2 = \max_{1 \leq i \leq n} |\lambda_i(A)|,$$

essendo $\lambda_i(A)$ gli autovalori di A.
Se A è simmetrica semidefinita positiva, per cui tutti gli autovalori sono non negativi si può porre:

$$\|A\|_2 = \lambda_M(A).$$

B Richiami sulla differenziazione in R^n

Richiamiamo alcuni concetti essenziali sulla differenziazione delle funzioni definite su R^n, la cui conoscenza è richiesta nello studio dei problemi di ottimizzazione di tipo differenziabile. Nel seguito supponiamo, per semplicità, che f sia una funzione definita su tutto R^n; è immediato tuttavia estendere le definizioni qui riportate al caso in cui x appartiene ad un insieme aperto $\mathcal{D} \subseteq R^n$.

B.1 Derivate del primo ordine di una funzione reale

Un qualsiasi vettore assegnato $d \in R^n$ non nullo definisce una *direzione* in R^n. Una prima nozione di derivata che si può introdurre è quella di *derivata direzionale.*

Definizione 9 (Derivata direzionale)

Sia $f : R^n \to R$. Si dice che f ammette derivata direzionale $Df(x, d)$ *nel punto $x \in R^n$ lungo la direzione $d \in R^n$ se esiste finito il limite*

$$\lim_{t \to 0^+} \frac{f(x + td) - f(x)}{t} := Df(x, d). \qquad \square$$

Se consideriamo f come funzione di una sola variabile x_j, supponendo fissate tutte le altre componenti possiamo introdurre il concetto di *derivata parziale* rispetto a x_j.

Definizione 10 (Derivata parziale)

Sia $f : R^n \to R$. Si dice che f ammette derivata parziale $\partial f(x)/\partial x_j$ *nel punto $x \in R^n$ rispetto alla variabile x_j se esiste finito il limite*

$$\lim_{t \to 0} \frac{f(x_1, \ldots, x_j + t, \ldots, x_n) - f(x_1, \ldots, x_j, \ldots, x_n)}{t} := \frac{\partial f(x)}{\partial x_j}. \qquad \square$$

L'esistenza della derivata parziale $\partial f(x)/\partial x_j$ nel punto x implica, ovviamente, che esistano e coincidano le derivate direzionali lungo le direzioni e_j e $-e_j$.

Se f ammette derivate parziali rispetto a tutte le componenti possiamo definire il *gradiente* di f nel punto x.

Definizione 11 (Gradiente)

Sia $f : R^n \to R$ ed $x \in R^n$. Se esistono le derivate parziali prime di f in x definiamo gradiente di f in x il vettore $\nabla f(x) \in R^n$

$$\nabla f(x) := \begin{pmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \cdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{pmatrix}.$$

A differenza di quanto avviene sulla retta reale, nel caso di R^n l' esistenza del gradiente non consente, in generale, di poter approssimare, con la precisione voluta, il valore di f nell'intorno di x con una funzione lineare dell'incremento. Una tale possibilità è legata ad una nozione più forte di differenziabilità, che è quella riportata nella definizione successiva.

Definizione 12 (Funzione differenziabile)

*Sia $f : R^n \to R$. Si dice che f è differenziabile (*secondo Frèchet*, o* in senso forte*) nel punto $x \in R^n$ se esiste $g(x) \in R^n$ tale che, per ogni $d \in R^n$ si abbia*

$$\lim_{\|d\| \to 0} \frac{|f(x+d) - f(x) - g(x)^T d|}{\|d\|} = 0,$$

o, equivalentemente, se per ogni $d \in R^n$ si può porre

$$f(x+d) = f(x) + g(x)^T d + \alpha(x,d),$$

dove $\alpha(x,d)$ soddisfa:

$$\lim_{\|d\| \to 0} \frac{\alpha(x,d)}{\|d\|} = 0.$$

L'operatore $g(x)$: $R^n \to R$ si dice derivata (di Frèchet) di f in x. □

È da notare che la sola esistenza di ∇f non implica, in generale, la differenziabilità nel senso prima definito. Si dimostra, tuttavia, che se $\nabla f(x)$ esiste ed è continuo rispetto ad x, allora f è differenziabile in x. (La continuità del gradiente è una condizione *sufficiente* per la differenziabilità in senso forte.) Vale il risultato seguente

Teorema 45 *Sia $f : R^n \to R$ e sia $x \in R^n$. Si ha:*

(i) *se f è differenziabile in x, allora f è continua in x, esiste il gradiente $\nabla f(x)$ e $\nabla f(x)^T$ coincide con la derivata di Frèchet di f in x*

(ii) *se esiste il gradiente $\nabla f(x)$ e $\nabla f(x)$ è continuo rispetto a x, allora f è differenziabile in x, e la derivata di Frèchet di f in x coincide con $\nabla f(x)^T$.* □

Dal teorema precedente segue che se $\nabla f(x)$ è continuo si può scrivere, per ogni $d \in R^n$:

$$f(x+d) = f(x) + \nabla f(x)^T d + \alpha(x,d),$$

dove $\alpha(x,d)$ soddisfa:

$$\lim_{\|d\| \to 0} \frac{\alpha(x,d)}{\|d\|} = 0.$$

Se f è differenziabile, è immediato verificare che esiste anche la derivata direzionale di f lungo una qualsiasi direzione $d \in R^n$ e risulta:

$$\lim_{t \to 0^+} \frac{f(x+td) - f(x)}{t} = \nabla f(x)^T d.$$

B.2 Differenziazione di un vettore di funzioni

Sia $g : R^n \to R^m$ un vettore a m componenti di funzioni reali. Possiamo introdurre la definizione seguente, che estende la nozione di gradiente.

Definizione 13 (Matrice Jacobiana)

Sia $g : R^n \to R^m$ e $x \in R^n$. Se esistono le derivate parziali prime $\partial g_i(x)/\partial x_j$, per $i = 1\dots, m$ e $j = 1 \dots n$ in x definiamo matrice Jacobiana *di g in x la matrice $m \times n$*

$$J(x) := \begin{pmatrix} \dfrac{\partial g_1(x)}{\partial x_1} & \dots & \dfrac{\partial g_1(x)}{\partial x_n} \\ \dots & \dots & \dots \\ \dfrac{\partial g_m(x)}{\partial x_1} & \dots & \dfrac{\partial g_m(x)}{\partial x_n} \end{pmatrix}.$$

Possiamo estendere in modo ovvio la nozione di differenziabilità al caso di un vettore di funzioni.

Definizione 14 (Derivata prima di un vettore di funzioni)

*Sia $g : R^n \to R^m$ e $x \in R^n$. Si dice che g è differenziabile (*secondo Frèchet*, o* in senso forte*) nel punto $x \in R^n$ se esiste una matrice $G(x)$ tale che, per ogni $d \in R^n$ si abbia*

$$\lim_{\|d\|\to 0} \frac{\|g(x+d) - g(x) - G(x)d\|}{\|d\|} = 0,$$

o, equivalentemente, se per ogni $d \in R^n$ si può porre

$$g(x+d) = g(x) + G(x)d + \gamma(x, d),$$

dove $\gamma(x, d)$ soddisfa:

$$\lim_{\|d\|\to 0} \frac{\gamma(x, d)}{\|d\|} = 0.$$

L'operatore lineare $G(x) : R^n \to R^m$ si dice derivata (di Frèchet) di g in x. □

Anche in questo caso, ovviamente, la sola esistenza della matrice Jacobiana in x non implica la differenziabilità e si ha il teorema seguente.

Teorema 46 *Sia $g : R^n \to R^m$ e $x \in R^n$. Si ha:*

(i) *se g è differenziabile in x, allora g è continua in x, esiste la matrice Jacobiana $J(x)$ e $J(x)$ coincide con la derivata di Frèchet di g in x*

(ii) *se esiste la matrice Jacobiana $J(x)$ di g in x e $J(x)$ è continua rispetto a x, allora g è differenziabile in x, e la derivata di Frèchet di g in x coincide con $J(x)$.* □

Dal teorema precedente segue che se $J(x)$ è continua si può scrivere, per ogni $d \in R^n$:

$$g(x+d) = g(x) + J(x)d + \gamma(x, d),$$

dove $\gamma(x, d)$ soddisfa:

$$\lim_{\|d\|\to 0} \frac{\gamma(x, d)}{\|d\|} = 0.$$

Se $m = 1$ si ha ovviamente $J(x) = \nabla g(x)^T$.
Per analogia con la notazione usata per il gradiente si può usare anche la notazione $\nabla g(x)^T$ per indicare la derivata prima di g, ossia

$$\nabla g(x) = J(x)^T = (\nabla g_1(x), \dots, \nabla g_m(x)).$$

B.3 Derivate del secondo ordine di una funzione reale

Sia $f : R^n \to R$ una funzione reale. Con riferimento alle derivate del secondo ordine, possiamo introdurre innanzitutto la definizione seguente.

Definizione 15 (Matrice Hessiana)

Sia $f : R^n \to R$ e $x \in R^n$. Se esistono le derivate parziali seconde $\partial^2 f(x)/\partial x_i \partial x_j$, per $i = 1 \dots, n$ e $j = 1 \dots n$ in x definiamo matrice Hessiana *di f in x la matrice $n \times n$*

$$\nabla^2 f(x) := \begin{pmatrix} \dfrac{\partial^2 f(x)}{\partial {x_1}^2} & \dots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dots & \dots & \dots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \dots & \dfrac{\partial^2 f(x)}{\partial {x_n}^2} \end{pmatrix}.$$

Possiamo introdurre la nozione di derivata seconda mediante la definizione seguente.

Definizione 16 (Derivata seconda)

*Sia $f : R^n \to R$ e $x \in R^n$. Si dice che f è due volte differenziabile (*secondo Frèchet*, o* in senso forte*) nel punto $x \in R^n$ se la derivata prima di f, $\nabla f(x)^T$ è differenziabile in x. La derivata prima di $\nabla f(x)^T$ si dice derivata seconda (di Frèchet) di f.* □

Anche in questo caso, la sola esistenza della matrice Hessiana (che possiamo interpretare come matrice Jacobiana di $\nabla f(x)$) non implica la differenziabilità e si ha il teorema seguente.

Teorema 47 *Sia $f : R^n \to R$ e $x \in R^n$. Si ha:*

(i) *se f è due volte differenziabile in x, allora il gradiente $\nabla f(x)$ esiste ed è continuo in x, la matrice Hessiana $\nabla^2 f(x)$ esiste ed è una matrice* simmetrica *e $\nabla^2 f(x)$ coincide con la derivata seconda di Frèchet di f in x*

(ii) *se esiste la matrice Hessiana $\nabla^2 f(x)$ in x e $\nabla^2 f(x)$ è continua rispetto a x, allora f è due volte differenziabile in x, $\nabla^2 f(x)$ è necessariamente simmetrica e la derivata seconda di Frèchet di f in x coincide con $\nabla^2 f(x)$.* □

Dal teorema precedente segue che se $\nabla^2 f(x)$ è continua si può scrivere, per ogni $d \in R^n$:

$$f(x + d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d + \beta(x, d),$$

dove $\beta(x, d)$ soddisfa:

$$\lim_{\|d\| \to 0} \frac{\beta(x, d)}{\|d\|^2} = 0.$$

Nelle stesse ipotesi si ha anche, come si è detto, che la matrice Hessiana $\nabla^2 f(x)$ è una **matrice simmetrica**, ossia si ha, per $i, j = 1 \dots, n$:

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.$$

Osserviamo anche che se $\nabla^2 f(x)$ è continua si può scrivere

$$\nabla f(x+d) = \nabla f(x) + \nabla^2 f(x)d + \gamma(x,d),$$

con

$$\lim_{\|d\|\to 0} \frac{\gamma(x,d)}{\|d\|} = 0.$$

B.4 Teorema della media e formula di Taylor

Nel caso di funzioni differenziabili valgono anche i risultati seguenti (che si possono tuttavia stabilire sotto ipotesi più deboli).

Teorema 48 (Teorema della Media)

Sia $f : R^n \to R$ una funzione differenziabile. Allora, per ogni $h \in R^n$, si può scrivere

$$f(x+h) = f(x) + \nabla f(z)^T h,$$

in cui $z \in R^n$ è un punto opportuno (dipendente da x e h) tale che $z = x + \zeta h$, con $\zeta \in (0,1)$.$\Box$

Possiamo dare di tale risultato anche una formulazione integrale.

Teorema 49 *Sia $f : R^n \to R$ una funzione differenziabile. Allora, per ogni $h \in R^n$, si può scrivere*

$$f(x+h) = f(x) + \int_0^1 \nabla f(x+th)^T h dt, \qquad .\Box$$

Utlizzando le derivate seconde si ha il risultato seguente.

Teorema 50 Teorema di Taylor

Sia $f : R^n \to R$ una funzione due volte differenziabile. Allora, per ogni $h \in R^n$ si può scrivere:

$$f(x+h) = f(x) + h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(w) h$$

in cui $w \in R^n$ è un punto opportuno (dipendente da x e h) tale che $w = x + \xi h$, con $\xi \in (0,1)$. $\Box$

Anche in questo caso possiamo considerare una formulazione di tipo integrale.

Teorema 51 *Sia $f : R^n \to R$ una funzione due volte differenziabile. Allora, per ogni $h \in R^n$ si può scrivere:*

$$f(x+h) = f(x) + h^T \nabla f(x) + \int_0^1 (1-t) h^T \nabla^2 f(x+th) h dt.$$

$\Box$

Nel caso di funzioni vettoriali $g : R^n \to R^m$ non è possibile stabilire un teorema della media. È tuttavia possibile considerare un'espressione di tipo integrale.

Teorema 52 *Sia $g : R^n \to R^m$ una funzione differenziabile. Allora, per ogni $h \in R^n$, si può scrivere*

$$g(x+h) = g(x) + \int_0^1 J(x+th)h dt,$$

in cui J è la matrice Jacobiana di g. $\square$

Come caso particolare del risultato precedente, se g è il gradiente ∇f di una funzione due volte differenziabile $f : R^n \to R$, si ha:

$$\nabla f(x+h) = \nabla f(x) + \int_0^1 \nabla^2 f(x+th)h dt.$$

Esempi

Alcuni esempi di interesse di funzioni differenziabili sono:

$f(x) = c^T x$; si ha: $\nabla f(x) = c, \quad \nabla^2 f(x) = 0$;

$f(x) = \frac{1}{2}x^T Q x + c^T x$, (con Q simmetrica); si ha: $\nabla f(x) = Qx + c; \quad \nabla^2 f(x) = Q$;

$f(x) = \|Ax - b\|^2$; si ha: $\nabla f(x) = 2A^T(Ax - b); \quad \nabla^2 f(x) = 2A^T A$

References

[1] L.E. Blume and Carl P. Simon "Matematica 2 per l'economia e le Scienza sociali", a cura di A. Zaffaroni, Università Bocconi Editore, EGEA s.p.a, Ed. italiana 2002.

[2] D. Bertsekas. "Nonlinear Programming", Athena Scientific, Belmont, Massachusetts, 1995.

[3] L. Grippo. "Metodi di ottimizzazione non vincolata", Rap.Tec. IASI, N. 64, 1988.