Forme Quadratiche Una forma quadratica é un polinomio omogeneo di grado 2 in n variabili.

In generale $x = (x_1, x_2, \dots, x_n)$

$$q(x) = \sum_{i,j=1}^{n} a_{i,j} x_i x_j = \sum_{i=1}^{n} a_{i,i} x_i^2 + \sum_{i \neq j}^{n} a_{i,j} x_i x_j$$

Scritta così, quanti sono i coefficienti di una forma quadratica in n variabili ? Sono n^2.

$$A = (a_{i,j})$$

In forma di prodotto scalare

$$q(x) = Ax \cdot x$$

In forma matriciale

$$q(x) = x^T A x$$

Esercizi di riscrittura dalla forma alla matrice associata e viceversa.

$$q(x_1, x_2, x_3) = x_1^2 + 3x_2^2 + x_3^2 - 24x_1x_2 - 6x_1x_3 + 2x_2x_3$$

La matrice associata è

$$\begin{pmatrix} 1 & -12 & -3 \\ -12 & 3 & 1 \\ -3 & 1 & 1 \end{pmatrix}$$

A una forma quadratica resta associata una matrice simmetrica.

Come riconoscere se la forma quadratica (o la matrice associata) è definita positiva ossia

$$q(x_1, x_2, x_3, \dots x_n) = \sum_{i,j=1}^{n} a_{i,j} x_i x_j \geq 0, \qquad \forall x \neq 0$$

soluzione n. 1 (calcolo degli autovalori)

$$D(\lambda) = \begin{vmatrix} \lambda - 1 & -12 & -3 \\ -12 & \lambda - 3 & 1 \\ -3 & 1 & \lambda - 1 \end{vmatrix}$$

$$D(\lambda) = 0,$$

$$(\lambda - 1) \begin{vmatrix} \lambda - 3 & 1 \\ 1 & \lambda - 1 \end{vmatrix} + 12 \begin{vmatrix} -12 & 1 \\ -3 & \lambda - 1 \end{vmatrix} - 3 \begin{vmatrix} -12 & \lambda - 3 \\ -3 & 1 \end{vmatrix} = 0$$

$$(\lambda - 1)((\lambda - 3)(\lambda - 1) - 1) - 144(\lambda - 1) + 36 + 36 - 9(\lambda - 3) =$$
$$(\lambda - 1)(\lambda^2 - 4\lambda + 3) - \lambda + 1 - 144\lambda + 144 + 72 - 9\lambda + 27 =$$
$$\lambda^3 - 4\lambda^2 + 3\lambda - \lambda^2 + 4\lambda - 3 - \lambda + 1 - 144\lambda + 144 + 72 - 9\lambda + 27$$
$$\lambda^3 - 5\lambda^2 - 147\lambda + 241 = 0$$

Osserviamo che

Tutte le soluzioni sono reali (due soluzioni sono positive e una negativa).

L'equazione di terzo grado per il Teorema Fondamentale dell'Algebra, ha tre soluzioni. L'equazione di terzo grado ha sempre almeno una radice reale, nel caso in esame sappiamo che tutte le soluzioni sono reali. Possiamo avere informazioni sul segno senza calcolarle?

Se r, s e t sono le sue tre radici, l'equazione si annulla se $x = r$ o se $x = s$ o se$x = t$, cioè dovrà essere:

(x – r) (x – s) (x – t) = 0

Si ottiene:

$$x^3 – (r + s + t)x^2 + (rs + rt + st)x – rst = 0$$

Da cui si trova $r + s + t = 5$ e $rst = -241$. Almeno una dovrà essere negativa, non non tutte e tre perché la somma è positiva.

Data l'equazione di terzo grado (Tartaglia, Cardano).

$$x^3 + bx^2 + cx + d = 0$$

la trasformiamo in un'equazione priva del termine di secondo grado

$$x = y – b/3$$

$$(y – b/3)^3 + b(y – b/3)^2 + c(y – b/3) + d = 0$$

$$(y^3 – 3(b/3)y^2 + 3(b^2/9)y – b^3/27) + b(y^2 – 2(b/3)y + b^2/9) + cy – bc/3 + d = 0$$

$$y^3 – by^2 + (b^2/3)y – b^3/27 + by^2 – (2b^2/3)y + b^3/9 + cy – bc/3 + d = 0$$

$$y^3 + (-b^2/3 + c)y + 2b^3/27 – bc/3 + d = 0$$

$$p = -b^2/3 + c \qquad q = 2b^3/27 – bc/3 + d$$

$$y^3 + py + q = 0$$

Introduciamo u e v

$$y = u + v$$

$$p = -3uv$$

$$(u + v)^3 - 3uv(u + v) + q = 0$$

Da cui

$$u^3 + v^3 = -q \qquad u^3v^3 = -p^3/27$$

Risolviamo

$$z^3 + qz - p^3/27 = 0$$

$$z_{1,2} = \frac{-q \pm \sqrt{p^2 + 4p^3/27}}{2}$$

Consideriamo un caso particolare : assumiamo

$$p^2 + 4p^3/27 \geq 0$$

Si ottiene una radice reale

$$y = \sqrt[3]{z}_1 + \sqrt[3]{z}_2.$$

Dividiamo poi il polinomio per $x - y$.

Fare per esercizio il caso

$$p^2 + 4p^3/27 < 0,$$

e mostrare che anche in questo caso una radice é reale.

0.1. **Forme quadratiche in $\mathbb{R}^2$ e la matrice Hessiana.** Una matrice Q si dice *non negativa* (rispettivamente, *non positiva*) se la forma $z^T Qz$ ($z = (x, y)$ is semidefinita positiva (rispettivamente negativa) cioè se

$$z^T Qz = \sum_{i,j=1}^{2} q_{i,j} x_i y_j \geq 0$$

$$(\text{rispettivamente, } z^T Qz \leq 0,)\, \forall (x, y) \in \mathbb{R}^2.$$

Una matrice Q si dice *positiva* (rispettivamente, *negativa*) se la forma $z^T Qz$ è *positiva* (rispettivamente, *negativa*)

$$z^T Qz = \sum_{i,j=1}^{2} q_{i,j} x_i y_j \; 0 \;\; (z^T Qz < 0) \forall (x, y) \in \mathbb{R}^2, x, y \neq 0.$$

Un esempio di matrice positiva è data da

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{0.1}$$

poichè $x^2 + y^2 > 0$. Un esempio di matrice non negativa è data da

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \tag{0.2}$$

mentre una matrice indefinita è data da

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \tag{0.3}$$

Se restringiamo l'analisi alle matrici 2×2 allora abbiamo il seguente

Teorema 0.1. *Sia*

$$Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \tag{0.4}$$

una matrice simmetrica.

$$|Q| = \det Q = q_{11} q_{22} - (q_{12})^2.$$

Allora

$$|Q| > 0 \quad \text{e } q_{11} > 0, \implies Q > 0$$

$$|Q| > 0 \quad \text{e } q_{11} < 0, \implies Q < 0$$

Se $\det \mathrm{Q} < 0$, *allora* Q *è indefinita.*

Proof. Data la forma quadratica

$$ax_1^2 + 2bx_1x_2 + cx_2^2,$$

essa può essere equivalentemente scritta

$$a\left(x_1 + \frac{b}{a}x_2\right)^2 + \frac{ac - b^2}{a}x_2^2,$$

da questa formula si evince chiaramente il risultato. □

Consideriamo le forme quadratiche nel caso di dimensione due associate alla matrice hessiana. Ricordiamo che la matrice hessiana, o matrice di Hesse, nel caso $n = 2$ è la matrice quadrata 2×2 delle derivate parziali seconde della funzione.

$$(Hf)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

ove il simbolo $\partial x_i \partial x_j$ indica che prima deriviamo rispetto a x_i e poi rispetto a x_j

1. Regressione lineare Minimi quadrati

Dati $n > 2$ di $\mathbb{R}^2$ con ascisse distinte tra loro si vuole determinare la retta che minimizza l'errore quadratico totale

$$F(a_0, a_1) = \sum_{j=1}^{n}(a_1x_j + a_0 - y_j)^2$$

Funzione di due variabili di cui possiamo calcolare i punti critici

$$\begin{cases} \frac{\partial F}{\partial a_0} = 2\sum_{j=1}^{n}(a_1x_j + a_0 - y_j) = 0 \\ \frac{\partial F}{\partial a_1} = 2\sum_{j=1}^{n} x_j(a_1x_j + a_0 - y_j) = 0 \end{cases}$$

Scritto sotto forma di sistema

$$\begin{cases} a_0 n + a_1\big(\sum_{j=1}^{n} x_j\big) = \sum_{j=1}^{n} y_j \\ a_0\big(\sum_{j=1}^{n} x_j\big) + a_1\big(\sum_{j=1}^{n} x_j^2\big) = \sum_{j=1}^{n} x_j y_j \end{cases}$$

$$D = \begin{vmatrix} n & \sum_{j=1}^{n} x_j \\ \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j^2 \end{vmatrix} = n\big(\sum_{j=1}^{n} x_j^2\big) - \big(\sum_{j=1}^{n} x_j\big)^2$$

Esercizio 1.1. *Dimostrare, assumendo* x_j *distinti tra loro al variare di* $j = 1, \dots,$

$$\big(\sum_{j=1}^{n} x_j\big)^2 < n\sum_{j=1}^{n} x_j^2, \; n \in \mathbb{N}, n \geq 2$$

Proof. La disuguaglianza è verificata per $n = 2$. **Assumendo vera la disuguaglianza** *al passo n* dobbiamo dimostare

$$\Big(\sum_{j=1}^{n+1} x_j\Big)^2 < (n+1)\sum_{j=1}^{n+1} x_j^2.$$

$$\Big(\sum_{j=1}^{n+1} x_j\Big)^2 = \Big(\sum_{j=1}^{n} x_j + x_{n+1}\Big)^2$$

$$\Big(\sum_{j=1}^{n} x_j\Big)^2 + x_{n+1}^2 + 2x_{n+1}\sum_{j=1}^{n} x_j <$$

$$n\sum_{j=1}^{n} x_j^2 + x_{n+1}^2 + 2x_{n+1}\sum_{j=1}^{n} x_j =$$

$$(n+1)\sum_{j=1}^{n} x_j^2 + nx_{n+1}^2 + x_{n+1}^2 - \underbrace{x_{n+1}^2 + \dots x_{n+1}^2}_{n \text{ volte}} - \sum_{j=1}^{n} x_j^2 + 2x_{n+1}\sum_{j=1}^{n} x_j =$$

$$(n+1)\sum_{j=1}^{n+1} x_j^2 - \sum_{j=1}^{n}\big(x_j - x_{n+1}\big)^2 < (n+1)\sum_{j=1}^{n+1} x_j^2.$$

□

Calcolo delle soluzioni

Un sistema con 2 equazioni e 2 incognite ha un'unica soluzione se e solo se il determinante D è diverso da zero. In questo caso, la soluzione è data da:

$$a_0 = \frac{\begin{vmatrix} \sum_{j=1}^n y_j & \sum_{j=1}^n x_j \\ \sum_{j=1}^n x_j y_j & \sum_{j=1}^n x_j^2 \end{vmatrix}}{\begin{vmatrix} n & \sum_{j=1}^n x_j \\ \sum_{j=1}^n x_j & \sum_{j=1}^n x_j^2 \end{vmatrix}}$$

$$a_1 = \frac{\begin{vmatrix} n & \sum_{j=1}^n y_j \\ \sum_{j=1}^n x_j & \sum_{j=1}^n x_j y_j \end{vmatrix}}{\begin{vmatrix} n & \sum_{j=1}^n x_j \\ \sum_{j=1}^n x_j & \sum_{j=1}^n x_j^2 \end{vmatrix}}$$

Gruppi di lavoro per impostazione problema. La dieta peso altezza, tariffa bus percorso prezzo, planning: consumi/ aumento salariale.

Riflessioni sulla validità del metodo. Regressione polinomiale

2. Elementi di topologia di $\mathbb{R}^m$

Ricordiamo che per $x, y \in \mathbb{R}$ valgono le seguenti proprietá

- $|x| \geq 0$
- $x \neq 0$ se e solo se $|x| > 0$
- $|x| = |-x|$
- $|xy| = |x||y|$
- $|x+y| \leq |x| + |y|$
- $||x| - |y|| \leq |x-y|$

2.1. **Norma.** Dato un punto $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$ si definisce la norma euclidea di x

$$\|x\|_2 = \left(x_1^2 + \cdots + x_m^2\right)^{1/2}.$$

$\forall x, y, z \in \mathbb{R}^m$ e $\lambda \in \mathbb{R}$, valgono le proprietà:

- $\|x\| \geq 0$,
- $\|x\| = 0 \iff x = 0$,
- $\|\lambda x\| = |\lambda| \cdot \|x\|$,
- $\|x + y\| \leq \|x\| + \|y\|$.

2.2. **Il prodotto scalare.** Il prodotto scalare in $\mathbb{R}^m$ è un numero reale dato da

$$x \cdot y = x_1 y_1 + \cdots + x_m y_m \qquad \text{per ogni x,y,z} \in \mathbb{R}^N\ \lambda \in \mathbb{R}$$

$$x \cdot y = y \cdot x, \qquad (x+y) \cdot z = x \cdot z + y \cdot z, \qquad \lambda(x \cdot y) = \lambda x \cdot y.$$

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

Dato $p > 1$, $p \in \mathbb{R}$ diciamo coniugato di p il numero reale q tale che

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Teorema 2.1. *Disuguaglianza di Young (esponenti reali): dati due numeri reali positivi a e b, e dati p, q numeri reali coniugati, si ha*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Proof. Fissiamo $b > 0$

$$f : [0, +\infty) \to \mathbb{R}$$

$$f(t) = \frac{t^p}{p} + \frac{b^q}{q} b^q - tb$$

$$\lim_{t \to +\infty} \frac{t^p}{p} p + \frac{b^q}{q} - tb = +\infty$$

$$f(0) = \frac{b^q}{q} > 0$$

$$f'(t) = t^{p-1} - b$$

$$t^{p-1} = b \iff t = b^{\frac{1}{p-1}}$$

$$f''(b^{\frac{1}{p-1}}) > 0$$

Punto di minimo relativo che è anche punto di minimo assoluto

$$f(b^{\frac{1}{p-1}}) = \frac{b^{\frac{p}{p-1}}}{p} + \frac{b^q}{q}b^q - b^{\frac{1}{p-1}}b = \left(\frac{1}{p} + \frac{1}{q} - 1\right)b^q = 0$$

Allora risulta per ogni numero reale $a \geq 0$

$$f(a) \geq 0,$$

ossia

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q$$

□

Teorema 2.2 (disuguaglianza di Hölder)**.** *Per* p, q *esponenti coniugati e* $p, q \in [1, +\infty)$ *e* $\forall x, y \in \mathbb{R}^m$ *abbiamo*

$$|x \cdot y| \leq \|x\|_p \|y\|_q. \tag{2.1}$$

Si fissi

$$a_i = \frac{|x_i|}{\|x\|_p}, \qquad b_i = \frac{|y_i|}{\|y\|_q}$$

Dalla disuguaglianza di Young

$$a_i b_i \leq \frac{1}{p}\frac{|x_i|^p}{\|x\|_p^p} + \frac{1}{q}\frac{|y_i|^q}{\|y\|_q^q}$$

Sommando

$$\sum_{i=1}^m a_i b_i \leq \frac{1}{p}\frac{\sum_{i=1}^m |x_i|^p}{\|x\|_p^p} + \frac{1}{q}\frac{\sum_{i=1}^m |y_i|^q}{\|y\|_q^q} = 1$$

Da cui si evince

$$\sum_{i=1}^m a_i b_i = \sum_{i=1}^m \frac{|x_i|}{\|x\|_p}\frac{|y_i|}{\|y\|_q} \leq 1$$

e la disuguaglianza di Hölder segue

$$|x \cdot y| \leq \|x\|_p \|y\|_q. \tag{2.2}$$

Teorema 2.3 (Disuguaglianza Minkowski)**.** *Per* $p \in [1, +\infty)$ *e* $\forall\ x, y \in \mathbb{R}^m$ *abbiamo*

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p. \tag{2.3}$$

$$|x_i + y_i|^p = |x_i + y_i|^{p-1}|x_i + y_i| \leq$$
$$|x_i + y_i|^{p-1}(|x_i| + |y_i|)$$

$$\sum_{i=1}^m |x_i + y_i|^p \leq \sum_{i=1}^m |x_i + y_i|^{p-1}|x_i| + \sum_{i=1}^m |x_i + y_i|^{p-1}|y_i|$$

Si ha

$$\sum_{i=1}^{m} |x_i + y_i|^{p-1}|x_i| \leq \|x\|_p \|(x+y)^{(p-1)}\|_q = \|x\|_p \left(\sum_{i=1}^{m} |x_i + y_i|^{(p-1)q} \right)^{\frac{1}{q}}$$

$$\sum_{i=1}^{m} |x_i + y_i|^{p-1}|y_i| \leq \|y\|_p \|(x+y)^{(p-1)}\|_q = \|y\|_p \left(\sum_{i=1}^{m} |x_i + y_i|^{(p-1)q} \right)^{\frac{1}{q}}$$

Quindi da $(p-1)q = p$

$$\|x+y\|_p^p \leq \|x+y\|_p^{p-1}(\|x\|_p + \|y\|_p)$$

e dividendo per $\|x+y\|_p^{p-1}$ (assumendo non nullo) si ottiene la Disuguaglianza di Minkowski

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p.$$

Esempio 2.4.

- *La formula*

$$\|x\|_1 = |x_1| + \cdots + |x_m|, \quad x = (x_1, \ldots, x_m) \in \mathbb{R}^m$$

definisce una norma su $\mathbb{R}^m$.

- *La formula*

$$\|x\|_\infty = \max\{|x_1|, \ldots, |x_m|\}$$

definisce una norma su $\mathbb{R}^m$.

Due norme $\|x\|_a$ $\|x\|_b$ si dicono equivalenti se esistono due costanti m e M tali che

$$m\|x\|_b \leq \|x\|_a \leq M\|x\|_b$$

Vale la relazione

$$\lim_{p\to+\infty} \|x\|_p = \|x\|_\infty$$

Sfruttando la relazione tra le norme, valida per ogni $p \geq 1$

$$\|x\|_\infty \leq \|x\|_p \leq m^{\frac{1}{p}} \|x\|_\infty$$

si ha che le norme p per $p \geq 1$ sono tra loro equivalenti.

Definizione 2.5. *Definiamo la* distanza *tra due punti di* $\mathbb{R}^m$ *tramite la formula*

$$d(x,y) := \|x-y\|$$

$$d(x,y) := \|x-y\| = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}$$

- $d(x,y) \geq 0$
- $d(x,y) = 0 \iff x = y$
- $d(x,y) = d(y,x)$
- $d(x,y) \leq d(x,z) + d(z,y)$

La base canonica in $\mathbb{R}^N$ è data da $e^1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, $e^m = (0, 0, \ldots, 1)$.

$$e^j = (0, \ldots 1, 0 \ldots 0)$$
$$e^k = (0, \ldots 0, 1 \ldots 0).$$
$$d(e^j, e^k) = \sqrt{2} \qquad j \neq k$$

2.3. **Interno, Esterno, Frontiera di un insieme.**

Definizione 2.6. *Sia $a \in \mathbb{R}^m$ e $r > 0$ un numero reale. $B_r(a)$ di centro a e raggio r per*

$$B_r(a) := \{x \in \mathbb{R}^m \; : \; \|x - a\| < r\}.$$

- Per $m = 1$ troviamo gli intervalli $]a - r, a + r[$.
- Per $m = 2$ troviamo il cerchio privato dei suoi punti frontiera
- Per $m = 3$ troviamo la palla privata dei suoi punti frontiera

Sia $A \subset \mathbb{R}^m$ e $x \in \mathbb{R}^m$.

- x è un *punto interno* di A se esiste $r > 0$ tale $B_r(x) \subset A$.
- x è un *punto esterno* di A se esiste $r > 0$ tale che $B_r(x) \subset \mathbb{R}^m \setminus A$.
- x è un *punto frontiera* di A se $B_r(x)$ incontra A e $\mathbb{R}^m \setminus A$ per ogni $r > 0$.

L'insieme dei punti interni esterni e di frontiera si chiama *interno*, *esterno* e la *frontiera* dia A, e si denota con $int(A)$, $ext(A)$ e ∂A. Si utilizza anche la notazione A° invece di $int(A)$.

Sia $A \subset \mathbb{R}^m$.

- Gli insiemi $int(A)$, $ext(A)$, ∂A formano una *partizione* di $\mathbb{R}^m$: sono disgiunti e la loro riunione fornisce $\mathbb{R}^m$.

3. Insiemi aperti, chiusi, compatti

Definizione 3.1. *Sia $X \subset \mathbb{R}^m$ e $x \in \mathbb{R}^m$. Si dice che il punto x è di aderenza per X se la palla $B_r(x)$ incontra X per ogni $r > 0$. L'insieme dei punti x con questa proprietà si chiama aderenza di X e si denota con $\overline{X}$.*

Proposizione 1. *Sia $X \subset \mathbb{R}^m$ e $x \in \mathbb{R}^m$, allora*

$$x \in \overline{X} \iff \exists\, (x_n) \subset X \, \mathrm{e}\, \mathrm{x_n} \to \mathrm{x}$$

Definizione 3.2. *Sia $X \subset \mathbb{R}^m$. Si dice che X è un insieme aperto se per ogni $x \in X$ se esiste $r > 0$ tale che la palla $B_r(x)$ è interamente contenuta in X.*

Dati due punti distinti di $\mathbb{R}^m$ x e y esistono due aperti X e Y tali che $x \in X$ $y \in X$ e $X \cap Y = \emptyset$.

Proposizione 2. $\emptyset$ *e $\mathbb{R}^m$ sono aperti. L' unione qualsiasi di insiemi aperti è un insieme aperto. L'intersezione di un numero finito di insiemi aperti è un insieme aperto.*

Definizione 3.3. *Un insieme X è chiuso se l'insieme complementare in $\mathbb{R}^m$ è aperto*

$\overline{X}$ è il più piccolo insieme chiuso che contiene X.

Proposizione 3. *$\emptyset$ e $\mathbb{R}^m$ sono chiusi. L'unione di un numero finito di insiemi chiusi è un insieme chiuso. L'intersezione qualunque di insiemi chiusi è un insieme chiuso .*

Definizione 3.4. *K è limitato $\iff$ esiste una costante L tale che $\|x\| < L$ per ogni $x \in K$ Il diametro di K è definito come*

$$diam(K) = \sup\{d(x, y), x, y \in K\}.$$

Se $diam(K) = +\infty$ diremo che K è illimitato.

Definizione 3.5. *K è compatto se $\forall (x_n) \subset X$ esiste una sottosuccessione (x_{n_k}) con $\lim x_{n_k} \in X$*

Teorema 3.6. *(***Teorema di Heine-Borel)** *K compatto $\iff$ K è chiuso e limitato*

L'intersezione infinita di una famiglia infinita di aperti può non essere aperta. Esempio $(-\frac{1}{n}, \frac{1}{n})$. L'intersezione $\{0\}$: un insieme chiuso.

$\emptyset$ e $\mathbb{R}^m$ sono gli unici insiemi che sono sia aperti che chiusi.

Ci sono insiemi che non sono né aperti né chiusi, ad esempio gli intervalli di $\mathbb{R}$ a cui appartiene un solo estremo.

4. Insiemi convessi

Definizione 4.1. *$\Omega \subset \mathbb{R}^N$ si dice* convesso *se per ogni x e $y \in \Omega$,*

$$\lambda x + (1-\lambda)y \in \Omega \qquad \textit{per ogni } \lambda \in [0, 1].$$

Se Ω contiene x e y, allora contiene tutti i punti del segmento di estremi x e y che indicheremo con $[x, y]$. Nel piano il cerchio $B_R(x)$ è un insieme convesso, mentre una corona circolare non è un insieme convesso. In $\mathbb{R}^m$ l'insieme $B_r(a) := \{x \in \mathbb{R}^m \ : \ \|x - a\| < r\}$ è convesso. Siano x, $y \in B_r(a)$ allora per $\lambda \in [0, 1]$

$$\|\lambda x + (1-\lambda)y - a\| = \|\lambda(x-a) + (1-\lambda)(y-a)\| \leq \lambda \|(x-a)\| + (1-\lambda) \|y-a)\| \leq r$$

L'intersezione di due insieme convessi è un insieme convesso.

Esempio l'intersezione non vuota di un cerchio con un quadrato.

L'intersezione di un numero finito di insiemi convessi è un insieme convesso. Esempi di insieme convessi

- p non nullo.

  Iperpiano (insieme chiuso)

  $$H = \{x \in \mathbb{R}^m : p^T x = \alpha\},$$

  ossia l'insieme dei vettori di $\mathbb{R}^m$ ortogonali a p

- p non nullo.

  Semispazi (insieme chiuso)

$$H^+ = \{x \in \mathbb{R}^m : p^T x \geq \alpha\},$$
$$H_- = \{x \in \mathbb{R}^m : p^T x \leq \alpha\},$$

Un insieme X è un poliedro se è intersezione di un numero finito di semispazi chiusi e iperpiani. In $\mathbb{R}^3$ un esempio di poliedro è il cubo. In $\mathbb{R}^2$ i poligoni piani.

Verificare che l'insieme

$$\{(x, y) : y + x \leq 1, x \geq 0, y \geq 0\}$$

è un sottoinsieme convesso di R^2.

Dato un insieme Ω, l'involucro convesso di Ω, $\mathrm{co}(\Omega)$ è il più piccolo insieme convesso che contiene Ω.

Teorema 4.2 (Teorema di separazione)**.** *Siano C_1 e C_2 due sottoinsiemi convessi di $\subset \mathbb{R}^N$ tali $\mathrm{int}(C_1) \cap C_2 = \emptyset$. allora esiste $p \in \mathbb{R}^N$, $p \neq 0$, tale che*

$$py \geq px, \qquad \forall x \in C_1, \forall y \in C_2 \tag{4.1}$$

5. Convergenza

5.1. Successioni.

Definizione 5.1. *Una successione (x_n) $x_n \in \mathbb{R}^m$ è* convergente, *se esiste un punto $a \in \mathbb{R}^m$, detto* limite *della successione tale che $\|x_n - a\| \to 0$ per $n \to \infty$.*

Diremo anche che (x_n) converge *ad a, e scriviamo*

$$x_n \to a \quad \text{anche} \ \lim x_n = a$$

.

Esempio 5.2.

- *Per $m = 1$ si ha la nozione di convergenza per successioni reali*

Definizione 5.3. *Una successione (x_n) $x_n \in \mathbb{R}^m$ è una successione di Cauchy se $\forall \varepsilon > 0$ $\exists \nu > 0$ tale che $\|x_n - x_m\| < \varepsilon$, $\forall n, m > \nu$*

Equivalentemente

Definizione 5.4. *Una successione (x_n) $x_n \in \mathbb{R}^m$ è una successione di Cauchy se $\forall \varepsilon > 0$ $\exists \nu > 0$ tale $\|x_{n+p} - x_n\| < \varepsilon$, $\forall n > \nu$, $\forall p \in \mathbb{N}$*

(Caratterizzazione della convergenza). Sia (x_n) $x_n \in \mathbb{R}^m$, $a \in \mathbb{R}^m$ scriviamo

$$x_n = (x_{n1}, \dots, x_{nm}) \quad \text{e} \quad a = (a_1, \dots, a_m).$$

Allora $x_n \to a$ in $\mathbb{R}^m \iff x_{nk} \to a_k$ in $\mathbb{R}$, per ciascuna componente k. $k = 1, \dots, m$.

In $\mathbb{R}^m$ non valgono più i risultati che fanno uso della monotonia.

Proposizione 4. *Siano* $(x_n)(,y_n)$ *due successioni con* $x_n, y_n \in \mathbb{R}^m$*e* $(\lambda_n) \subset \mathbb{R}$.

- *Il limite di una successione convergente è unico : se* $x_n \to a$ *e* $x_n \to b$*, allora* $a = b$.
- *Se* $x_n \to a$*, allora* $x_{n_k} \to a$ *per ogni sottosuccessione* (x_{n_k}) *della successione* (x_n).
- *Se* $x_n \to a$ *e* $y_n \to b$*, allora* $x_n + y_n \to a + b$.
- *Se* $\lambda_n \to \lambda$ *(in* $\mathbb{R}$*) e* $x_n \to a$ *(in* $\mathbb{R}^m$*), allora* $\lambda_n x_n \to \lambda a$ *(in* $\mathbb{R}^m$*).*
- *Se* $x_n \to a$ *(in* $\mathbb{R}^m$*), allora* $\|x_n\| \to \|a\|$ *(in* $\mathbb{R}$*).*

Definizione 5.5. *Una successione* (x_n) $x_n \in \mathbb{R}^m$ è limitata, *se esiste* $L \in \mathbb{R}$ *tale che* $\|x_n\| < L$ *per ogni* n.

Tutte le successioni convergenti sono limitate e vale

Teorema 5.6. *(Bolzano–Weierstrass)Tutte le successioni limitate di* $\mathbb{R}^m$ *ammettono una sottosuccessione convergente*

Denotiamo con $B(x,r)$ la *palla* di centro $x \in \mathbb{R}^N$ e raggio $r > 0$, ossia

$$B(x,r) = \{y \in \mathbb{R}^N : \|y - x\| \leq r\},$$

Definizione 5.7. *Dato* $\Omega \subset \mathbb{R}^m$*, diciamo che* $x \in \mathbb{R}^m$ *è un punto di accumulazione per* Ω *se esiste una successione* (x_n) *tale che* $x_n \in \Omega$*,* $x_n \neq x$ *e* $\lim_{n\to\infty} \|x_n - x\| = 0$.

Lemma 5.8. *Sia* $f : A \to \mathbb{R}^n$ *(*$A \subset \mathbb{R}^m$*) e* $a \in A$ *di accumulazione per* A

Le due proprietà seguenti sono equivalenti

(a) $\forall \varepsilon > 0\ \exists \delta > 0$ *tale che se* $x \in A$ *e* $\|x - a\| < \delta$*, allora* $\|f(x) - f(a)\| < \varepsilon$*;*

(b) (x_n) $x_n \in D(f)$ *e* $x_n \to a$*, allora* $f(x_n) \to f(a)$.

f è continua in A se é continua in ogni punto $a \in A$, ossia
$\forall \varepsilon > 0\ \exists \delta > 0$ tale che se $x \in A$ e $\|x - a\| < \delta$, allora $\|f(x) - f(a)\| < \varepsilon$

5.2. **Teorema di Weierstrass.** Una funzione continua in un compatto ammette minimo e massimo assoluto.

6. Funzione Distanza

Consideriamo la funzione distanza di $x \in \mathbb{R}^N$ non appartenente all' insieme compatto e convesso D

$$d_D(x) = \inf_{y \in D} |x - y| \,. \tag{6.1}$$

$|x - y|$: si tratta di una funzione continua, D è un insieme compatto, come applicazione del teorema di Weierstrass, abbiamo che il minimo è assunto. Allora per qualche $y^* \in D$.

$$d_D(x) = |x - y^*|$$

$d_D : \mathbb{R}^N \to \mathbb{R}$ la funzione distanza di $x \in \mathbb{R}^N$ dall'insieme chiuso D è allora definita

$$d_D(x) = \min_{y \in D} |x - y| \,.$$

Proposizione 5. *Sia D un insieme compatto. Allora d_D è Lipschitz continua*

$$\left| d_D(x) - d_D(x') \right| \leq \left| x - x' \right| \ \forall x, x' \in \mathbb{R}^N.$$

Proof. $d_D(x') = x' - \hat{y}$ per qualche $\hat{y} \in D$ allora

$$d_D(x) - d_D(x') \leq |x - \hat{y}| - \left| x' - \hat{y} \right| \leq \left| x - x' \right| .$$

La dimostrazione segue scambiando il ruolo di x and x'. □

7. Test degli autovalori

Gli autovalori di una matrice simmetrica $n \times n$ sono numeri reali che possiamo ordinare dal più piccolo al più grande

$$\lambda_1 \leq \lambda_2 \leq \dots \lambda_n$$

Teorema 7.1. *Se λ_1 e λ_n sono il più piccolo e il più grande autovalore di A allora*

$$\lambda_1 \|h\|^2 \leq Ah \cdot h \leq \lambda_n \|h\|^2 , \qquad \forall h \in \mathbb{R}^n \tag{7.1}$$

Introduciamo la forma quadratica

$$f(h) = Ah \cdot h = \sum_{i,j=1}^{n} a_{i,j} h_i h_j$$

in

$$K = \{h \in \mathbb{R}^n : \|h\| = 1\}$$

$f(h)$ è una funzione continua su K (insieme chiuso e limitato). Per il teorema di Weierstrass ammette minimo e massimo in K, siano h_1 e h_2 rispettivamente con $f(h_1) = m$ e $f(h_2) = M$. Se $h = 0$ la (7.1) è ovviamente verificata. Supponiamo $h \in \mathbb{R}^n$ con $h \neq 0$. Poniamo

$$\mu = \frac{h}{\|h\|}$$

e consideriamo

$$A\mu \cdot \mu = \sum_{i,j=1}^{n} a_{i,j} \mu_i \mu_j .$$

Si ha $\mu \in K$, segue allora

$$m \leq \sum_{i,j=1}^{n} a_{i,j} \mu_i \mu_j \leq \sum_{i,j=1}^{n} a_{i,j} \frac{h_i}{\|h\|} \frac{h_j}{\|h\|} = \frac{1}{\|h\|^2} \sum_{i,j=1}^{n} a_{i,j} h_i h_j \leq M$$

La funzione
$$g(h)=\frac{1}{\|h\|^2}\sum_{i,j=1}^{n}a_{i,j}h_ih_j$$
in $\mathbb{R}^n\setminus 0$ ammette in h_1 un punto di minimo e in h_2 un punto di massimo.

Rimane da dimostrare che m e M sono autovalori di A $Ah_1=mh_1$, $Ah_2=Mh_2$, più esattamente sono il più piccolo e il più grande autovalore di A.

Calcoliamo le derivate parziali della funzione g
$$\frac{\partial g}{\partial h_i}=\left(\sum_{i,j}^{n}a_{i,j}h_ih_j\right)\frac{\partial}{\partial h_i}\left(\frac{1}{\|h\|^2}\right)+\frac{1}{\|h\|^2}\left(\frac{\partial}{\partial h_i}\sum_{i,j=1}^{n}a_{i,j}h_ih_j\right)$$
Calcoliamo
$$\frac{\partial}{\partial h_i}\left(\frac{1}{\|h\|^2}\right)=\frac{\partial}{\partial h_i}\left(\frac{1}{h_1^2+h_2^2+\ldots h_n^2}\right)=-\frac{2h_i}{(h_1^2+h_2^2+\ldots h_n^2)^2}==-\frac{2h_i}{\|h\|^4}$$
e utilizzando $a_{i,j}=a_{j,i}$
$$\left(\frac{\partial}{\partial h_i}\sum_{i,j=1}^{n}a_{i,j}h_ih_j\right)=2\sum_{i,j=1}^{n}a_{j,i}h_j.$$
Da cui si deduce
$$\frac{\partial g}{\partial h_i}=\frac{2}{\|h\|^2}\left(\sum_{i,j=1}^{n}a_{j,i}h_j-\frac{Ah\cdot h}{\|h\|^2}h_i\right)$$
$Dg(h_1)=0 \iff Ah_1-g(h_1)h_1=0 \qquad Dg(h_2)=0 \iff Ah_2-g(h_2)h_2=0$
ossia $m=g(h_1)$ e $M=g(h_2)$ autovalori per A. Inoltre sono il più piccolo e il più grande autovalore come segue facilmente.

8. Funzioni convesse e concave

Definizione 8.1. *Sia C un insieme aperto e convesso. $f:C\to\mathbb{R}$ si dice*

i) convessa *se*

(8.1) $$f(\lambda x+(1-\lambda)y)\le,\lambda f(x)+(1-\lambda)f(y)\quad\forall x,y\in C\lambda\in[0,1].$$

Si dice strettamente convessa *se in* (8.2) *si ha la disuguaglianza stretta per $\lambda\in(0,1)$*

ii) concava *se $-f$ è convessa ossia*

(8.2) $$\lambda f(x)+(1-\lambda)f(y)\le f(\lambda x+(1-\lambda)y)\quad\forall x,y\in C,\lambda\in[0,1].$$

Le funzioni affini definite in $\mathbb{R}^N$ sono convesse e concave, un esempio di funzione strettamente convessa è $f(x)=\|x\|^2$. La funzione $f:\mathbb{R}\to\mathbb{R}$ definita da

(8.3) $$f(x)=\begin{cases}|x|^2, & x\ge 0,\\ |x| & x<0\end{cases}$$

è convessa, ma non strettamente convessa.

Per funzioni regolari si ha la seguente caratterizzazione della convessità .

Proposizione 6. *Sia $f : C \to \mathbb{R}$, con C sottoinsieme aperto e convesso di $\mathbb{R}^N$. Allora*

i) *Se $f \in C^1(C)$, allora f è convessa in C se e solo se*

$$f(y) - f(x) \geq Df(x)(y - x)\forall x, y \in C. \tag{8.4}$$

ii) *Se $f \in C^2(C)$, allora f è convessa in C se e solo se $D^2 f(x) \geq 0$ $\forall x \in C$.*

Proof. □

8.1. **Funzione distanza.** Assumiamo che D sia un insieme convesso. Segue che la funzione $d_D(x)$ è convessa. In ipotesi di convessità vediamo come dimostrare l'unicità del punto di minimo. Prima osservazione è che il punto di minimo si trova sulla frontiera, sia μ tale che

$$\mu = |x - y^*|$$

e se supponiamo ne esista un altro punto y^-

$$\mu = \left|x - y^-\right|$$

I due punti non possono trovarsi sullo stesso segmento, pertanto $x - y^-$ $x - y^*$ sono linearmente indipendenti. In tal caso, prendendo y^- e y^* si ha

$$f(\alpha y^- + (1 - \alpha)y^*) < \alpha f(y^-) + (1 - \alpha)f(y^*) = \mu,$$

una contraddizione. Resta univocamente determinata

$$\mathrm{proj}_D(x) = \{y \in D \ \text{ tale che } d_D(x) = |x - y|\} \quad x \in \mathbb{R}^N.$$

8.2. **Convessità delle forme quadratiche.** Esercizio. Riprendere i conti sul paragrafo test degli autovalori $q(h) = h^T Q h$ convessa $\iff$ Q è semidefinita positiva.

8.3. **Disuguaglianza discreta di Jensen.** Si $f : C \to \mathbb{R}$ una funzione convessa in C insieme convesso. Dati k punti con $k \geq 2$

$$x_1, x_2, \ldots, x_k$$

si ha

$$\frac{x_1 + x_2 + \cdots + x_k}{k} \in C$$

e

$$f\left(\frac{x_1 + x_2 + \cdots + x_k}{k}\right) \leq \frac{1}{k}\Bigg(f(x_1) + f(x_2) + \ldots f(x_k)\Bigg)$$

se $k = 2$ allora $\frac{x_1+x_2}{2} \in C$ segue dalla combinazione convessa di due punti. e $f\left(\frac{x_1+x_2}{2}\right) \leq \frac{1}{2}\Bigg(f(x_1) + f(x_2)),$ segue dalla convessità di f. Assumiamo che al passo k sussista (ipotesi induttiva) si ha

$$\frac{x_1 + x_2 + \cdots + x_k}{k} \in C$$

e

$$f\big(\frac{x_1 + x_2 + \cdots + x_k}{k}\big) \leq \frac{1}{k}\bigg(f(x_1) + f(x_2) + \dots f(x_k)\bigg)$$

e dimostriamo

$$\frac{x_1 + x_2 + \cdots + x_k + x_{k+1}}{k+1} \in C$$

e

$$f\big(\frac{x_1 + x_2 + \cdots + x_k + x_{k+1}}{k+1}\big) \leq \frac{1}{k+1}\bigg(f(x_1)+f(x_2)+\dots f(x_k)+f(x_{k+1})\bigg).$$

Poniamo

$$\lambda = \frac{k}{k+1} \qquad 1 - \lambda = 1 - \frac{k}{k+1} = \frac{1}{k+1},$$

allora

$$\frac{x_1 + x_2 + \cdots + x_k + x_{k+1}}{k+1} = \lambda\frac{x_1 + x_2 + \cdots + x_k}{k} + (1-\lambda)x_{k+1} \in C$$

Si ha

$$f\big(\frac{x_1 + x_2 + \cdots + x_k + x_{k+1}}{k+1}\big) = f\big(\frac{x_1 + x_2 + \cdots + x_k}{k+1} + \frac{1}{k+1}x_{k+1}\big)$$

e

$$f\big(\frac{x_1 + x_2 + \cdots + x_k + x_{k+1}}{k+1}\big) = f\big(\frac{x_1 + x_2 + \cdots + x_k}{k+1} + \frac{1}{k+1}x_{k+1}\big) =$$

$$f\big(\lambda\frac{x_1 + x_2 + \cdots + x_k}{k} + \frac{1}{k+1}x_{k+1}\big) \leq$$

$$\lambda f\big(\frac{x_1 + x_2 + \cdots + x_k}{k}\big) + (1-\lambda)f(x_{k+1}) \leq$$

$$\lambda\big(\frac{1}{k}\big(f(x_1) + f(x_2) + \dots f(x_k)\big)\big) + (1-\lambda)f(x_{k+1}) =$$

$$\frac{1}{k+1}\big(f(x_1) + f(x_2) + \dots f(x_k) + f(x_{k+1})\big)$$

8.4. **Trasformata di Legendre-Fenchel.** La trasformata di Legendre è un procedimento che trasforma una funzione convessa a valori reali di variabile reale in un' altra funzione convessa.

Sia $f : \mathbb{R}^N \to \mathbb{R}$, $f \in C^2(\mathbb{R}^N)$ e f tale che $D^2 f > 0$ in C. Allora f è strettamente convessa e la trasformata di Legendre-Fenchel di f è definita come

$$f^*(x) = \sup_{y\in\mathbb{R}^N} \big[x \cdot y - f(y)\big] \qquad x \in \mathbb{R}^N \tag{8.5}$$

Esempio 8.2. $1-$*dimensione*

$$f(x) = \frac{1}{2}x^2$$

$$f^*(x) = \frac{1}{2}x^2.$$

In dimensione n siano p.q esponenti coniugati.

$$f(x) = \frac{1}{p}\,\|x\|^p$$

con

$$\|x\|^p = (x_1^2 + x_2^2 + \cdots + x_n^2)^{\frac{p}{2}}$$

Allora

$$f^*(x) = \frac{1}{q}\,\|x\|^q\,.$$

Infatti annullando il gradiente si ha

$$x_j - \|\hat{y}\|^{p-1}\,\frac{\hat{y}_j}{\|\hat{y}\|} = 0 \iff x_j - \|\hat{y}\|^{p-2}\,\hat{y}_j = 0$$

Si ricava

$$\|\hat{y}\|^{p-1} = \|x\| \iff \|\hat{y}\| = \|x\|^{\frac{1}{p-1}}\,.$$

Perciò

$$\hat{y}_j = x_j\,\|x\|^{-\frac{p-2}{p-1}}$$

Sostituendo il valore si ottiene

$$f^*(x) = \sum_j x_j\hat{y}_j - \frac{1}{p}\,\|\hat{y}\|^p =$$

$$\sum_j x_j x_j\,\|x\|^{-\frac{p-2}{p-1}} - \frac{1}{p}\,\|x\|^{\frac{p}{p-1}} =$$

$$\|x\|^2\,\|x\|^{-\frac{p-2}{p-1}} - \frac{1}{p}\,\|x\|^{\frac{p}{p-1}} =$$

$$\|x\|^{\frac{p}{p-1}} - \frac{1}{p}\,\|x\|^{\frac{p}{p-1}} = \frac{1}{q}\,\|x\|^q$$

Esempio 8.3. *Sia $f(x) = x^2$ con $x \in C = [2,3]$. Per ogni y fissato $yx - x^2$ è continua in un intervallo compatto. Ne segue cha f^* è definite in $\mathbb{R}$. Si ha un punto stazionario $2x = y$ con $x \in [2,3] \iff y = \frac{x}{2} \qquad 4 \le y \le 6$, altrimenti ($4 > y$ oppure $y > 6$) il massimo è assunto agli estremi ($x = 2$ oppure $x = 3$). In conclusione la trasformata è definita in tutto $\mathbb{R}$ e $f^*(y)$ vale*

i) $2y - 4 \qquad y < 4$
ii) $\frac{y^4}{4} \qquad 4 \le y \le 6$
iii) $3y - 9 \qquad y > 6$

Teorema 8.4. *Sia C un insieme aperto e convesso. Sia $f : C \to \mathbb{R}$ $f \in C^2(C)$ e $D^2 f > 0$ allora*

i) *Per ogni x, esiste $y = y(x)$ tale che il* sup *in* (8.5) *è assunto.*
ii) *f^* è convessa.*
iii) *$f^{**} = f$*

Nel caso unidimensionale: osserviamo che per le ipotesi di regolarità

$$D_x\big[xy - f(x)\big] = 0 \iff y = f'(x),$$

sia x la soluzione ossia $x = g(y)$ con $g = (f')^{-1}(y)$ allora

$$D_y g(y) = \frac{1}{f''(g(y))}$$

$$D_y f^*(y) = g(y) + yg'(y) - f'(g(y))g'(y) = g(y).$$

Risulta (la derivata della funzione inversa è il reciproco della derivata della funzione calcolata nella controimmagine del punto)

$$D_y^2 f^*(y) = \frac{1}{f''(g(y))}$$

Esercizio 8.5. *Se f è omogenea di grado $p > 1$ alllora f^* è omogenea di grado q con p e q coniugati tra loro. Sia $\lambda > 0$*

$$f^*(\lambda x) = \sup_{y\in\mathbb{R}^N}\big[\lambda x \cdot y - f(y)\big] = \sup_{y\in\mathbb{R}^N}\big[\lambda^{q+1-q} x \cdot y - f(y)\big] = \tag{8.6}$$

$$\lambda^q \sup_{y\in\mathbb{R}^N}\big[x \cdot (\lambda^{1-q})y - \lambda^{-q} f(y)\big] = \lambda^q \sup_{y\in\mathbb{R}^N}\big[x \cdot (\lambda^{1-q} y) - f(\lambda^{-\frac{q}{p}} y)\big]$$

Osserviamo

$$\frac{q}{p} = q - 1,$$

pertanto ponendo $\xi = \lambda^{1-q} y$ si ottiene

$$f^*(\lambda x) = \lambda^q \sup_{\xi\in\mathbb{R}^N}\big[x \cdot \xi - f\xi)\big] = \lambda^q f^*(x)$$

8.5. **Esempio: Minimi e Massimi per funzioni di 2 variabili.** $f(x, y) = x^2 + y^2$, generalizzare a $x = (x_1, x_2)$ $\quad f(x) = x^T Ax$ con A matrice definita positiva.

8.6. **Esempio: Minimo e Massimo assoluti per funzioni di 2 variabili su un insieme chiuso e limitato.** $f(x, y) = x - y$ in un rettangolo (esempio $-1 \le x \le 1, \quad -2 \le y \le 2$)

Esercizi su mimino e massimo assoluti in due variabili.

8.7. **Esempio: Minimi e Massimi relativi per funzioni di 3 variabili.** $f(x, y) = x^2 + y^2 + z^2$, generalizzare a $x = (x_1, x_2, x_3)$ $\quad f(x) = x^T Ax$ con A matrice definita positiva.

8.8. **Esempio: Minimo e Massimo assoluti per funzioni di 3 variabili su un insieme chiuso e limitato.** $f(x,y) = x - y - z$ in un parallelepipedo (esempio $-1 \leq x \leq 1, \quad -2 \leq y \leq 2, \quad -3 \leq z \leq 3$).

Tra tutti i parallelepipedi retti a base rettangolare inscritti in un ellissoide di equazione

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

determinare quello di volume massimo. Se x, y, z indicano i semispigoli si tratta di massimizzare la funzione

$$f(x,y,z) = 8xyz,$$

soggetta al vincolo $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$, $x > 0, y > 0, z > 0$. Il problema può essere scritto nella forma

$$\begin{cases} \max_{x,y,z} 8xyz \\ \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \\ x > 0, y > 0, z > 0 \end{cases}$$

8.9. **Metodo di Lagrange.**

$$\begin{cases} \max_{x,y,z} 8xyz \\ \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \end{cases}$$

$$L(x,y,z,\lambda) = \max_{x,y,z} 8xyz + \lambda(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1)$$

Si calcolano i punti stazionari della funzione L

$$\begin{cases} 8yz + \frac{2\lambda x}{a^2} = 0 \\ 8xz + \frac{2\lambda y}{b^2} = 0 \\ 8xy + \frac{2\lambda z}{c^2} = 0 \\ \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \end{cases}$$

Si ricava dalla prima equazione

$$\lambda = -\frac{4a^2yz}{x}$$

$$\begin{cases} 8x^2zb^2 - 8a^2y^2z = 0 \\ 8x^2yc^2 + 8yz^2a^2 = 0 \\ \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \end{cases}$$

Semplificando

$$\begin{cases} \frac{x^2}{a^2} = \frac{y^2}{b^2} \\ \frac{x^2}{a^2} = \frac{z^2}{c^2} \\ \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \end{cases}$$

$$x^2 = \frac{a^2}{3},$$

la cui soluzione positiva è

$$x = \frac{a}{\sqrt{3}}.$$

Dunque

$$x = \frac{a}{\sqrt{3}} \qquad y = \frac{b}{\sqrt{3}} \qquad z = \frac{c}{\sqrt{3}}.$$

In generale il problema di minimizzazione per funzioni a valori reali soggetta a m vincoli

$$\min f(x), \qquad x \in \Omega = \{x \in \mathbb{R}^N : \phi_1(x) = \phi_2(x) = \cdots = \phi_m(x) = 0\}$$

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i \phi_i(x) = f(x) + \lambda^T \phi(x)$$

In ipotesi di regolarità i punti stazionari sono dati dalla soluzione nelle $N+m$ variabili

$$DL(x, \lambda) = Df(x) + \sum_{i=1}^{m} \lambda_i D\phi_i(x) = Df(x) + \lambda^T D\phi(x) = 0$$

Vediamo le condizioni per il problema di minimizzazione per la forma quadratica

$$\min x^T Ax, \qquad x \in \Omega = \{x \in \mathbb{R}^N : Cx = d\},$$

con A definita positiva, C matrice $m \times n$, $d \in \mathbb{R}^m$ $m < n$ di rango m

$$L(x, \lambda) = x^T Ax + \lambda(Cx - d),$$

$$D_x L(x, \lambda) = Ax + C^T \lambda = 0 \quad D_\lambda L(x, \lambda) = Cx - d = 0$$

$\lambda \in \mathbb{R}^m$. Esempio.

8.10. **Vincoli di disuguaglianza.** Condizioni necessarie (KKT).

$$\min x^2 + y, \qquad x \in \Omega = \{x \in \mathbb{R}^N : x^2 + y^2 \leq 4, x + y \leq 1\},$$
$$L(x, y, \lambda_1, \lambda_2) = x^2 + y + \lambda_1(x^2 + y^2 - 4) + \lambda_2(x + y - 1).$$

Minimizzazione sotto la condizione $x^2 + y^2 \leq 4, x + y \leq 1,\ \lambda_1, \lambda_2 \geq 0$ $\lambda_1(x^2 + y^2 - 4) = 0$ $\lambda_2(x + y - 1)$=0

- $x + y < 1,\ x^2 + y^2 < 4$ il minimo non è influenzato dai vincoli $\lambda_1 = \lambda_2 = 0$.

  $$D_x L(x, y, \lambda) = 2x \qquad D_y L(x, y, \lambda) = 1.$$

  Non vi sono punti interni in cui si annulla il gradiente
- $x + y = 1,\ x^2 + y^2 < 4$ il minimo non è influenzato da un vincolo $\lambda_1 = 0$.

  $$D_x L(x, y, \lambda) = 2x + \lambda_2 \qquad D_y L(x, y, \lambda) = 1 + \lambda_2 \qquad .$$

  $1 + \lambda_2 = 0$ non è possibile.

- $x^2 + y^2 = 4$, $x + y < 1$ il minimo non è influenzato da un vincolo $\lambda_2 = 0$.

  $D_x L(x, y, \lambda) = 2x + 2x\lambda_1 \qquad D_y L(x, y, \lambda) = 1 + 2y\lambda_1 \qquad .$

  Pertanto
  $$\lambda_1 = -1,$$
  non è possibile oppure
  $$x = 0.$$
  In tal caso
  $$y = \pm 2.$$
  Se $y = 2$ allora $\lambda_1 = -\frac{1}{2}$, non accettabile. Pertanto l'unica soluzione è
  $$(0, -2, \frac{1}{4}, 0) \qquad f(0, -2, \frac{1}{4}, 0) = -2$$
- Entrambi i vincoli sono attivi $x^2 + y^2 = 4$, $x + y = 1$
  $$2x + 2x\lambda_1 + \lambda_2 = 0,$$
  $$1 + 2y\lambda_1 + \lambda_2 = 0$$
  Completare e concludere il calcolo.

9. Convoluzione Inf-Sup per funzioni non regolari

Data una funzione $f : \mathbb{R}^N \to \mathbb{R}$, $f \in C(\mathbb{R}^N)$ la convoluzione Inf-Sup f indicate con f_ϵ e con f^ϵ

$$f_\epsilon(x) = \inf_{y \in \mathbb{R}^N} \left(f(y) + \frac{\|x - y\|^2}{2\epsilon} \right) \tag{9.1}$$

e

$$f^\epsilon(x) = \sup_{y \in \mathbb{R}^N} \left(f(y) - \frac{\|x - y\|^2}{2\epsilon} \right) \tag{9.2}$$

Per ϵ che tende a 0 $f^\epsilon \searrow f$ e $f_\epsilon \nearrow f$ localmente uniformemente.

Ref. Fundamentals of Convex Analysis, Di Jean-Baptiste Hiriart-Urruty,Claude Lemaréchal, Springer

Ref. A.Avantaggiati, P.Loreti, An approximation of Hopf-Lax type formula via idempotent analysis Tropical and Idempotent Mathematics and Applications, Proceedings of the International Workshop on Tropical and Idempotent Mathematics, Independent University of Moscow, Russia, August 26–31, 2012. Contemporary Mathematics, Publication Year 2014: Volume 616, ISBNs: 978-0-8218-9496-5 (print); 978-1-4704-1684-3 (online) (pag 47-58) DOI: http://dx.doi.org/10.1090/conm/616) Consideriamo il seguente esempio. Sia

$$f = \|x\|.$$

$$f_\epsilon(x) = \inf_{y \in \mathbb{R}^N} \left[\left(\sum_{k=1}^N y_k^2 \right)^{\frac{1}{2}} + \frac{1}{2\epsilon} \sum_{k=1}^N (x_k - y_k)^2 \right]. \tag{9.3}$$

Assumiamo $y \neq 0$, calcoliamo il gradiente e poniamolo uguale a 0. Troviamo

$$\frac{y_k}{\|y\|} - \frac{1}{\epsilon}(x_k - y_k) = 0 \qquad \forall k = 1 \dots N. \tag{9.4}$$

Allora abbiamo, quadrando

$$\epsilon^2 \frac{y_k^2}{|y|^2} = (x_k - y_k)^2,$$

e sommando su k

$$\|x - y\|^2 = \epsilon^2.$$

Questo implica

$$\|x - y\| = \epsilon.$$

Moltiplichiamo (9.4) per y_k e sommiamo allora si ottiene

$$\|y\| - \frac{1}{\epsilon}(yx - \ \|y\|^2) = 0.$$

Dalla formula (9.4) otteniamo anche

$$y_k(\|y\| + \epsilon) = \|y\| x_k \qquad \forall k = 1 \dots N.$$

Facendo il quadrato e addizionando

$$\|y\|^2(\|y\| + \epsilon)^2 = \|y\|^2 \|x\|^2.$$

Ossia

$$\|y\| + \epsilon = \|x\|$$

e quindi

$$\|x\| > \epsilon \qquad y = \|x\| - \epsilon.$$

per questo valore di y si ha

$$f_\epsilon(x) = \|x\| - \frac{1}{2\epsilon}.$$

Assumiamo $|y| \neq 0$ e $|x| \leq \epsilon$, allora

$$\|y\| + \frac{1}{2\epsilon}\|y - x\|^2 \geq$$

$$\|y\| + \frac{1}{2\epsilon}(\|y\|^2 + |x|^2 - 2\|x\|\|y\| = |y|(1 - \frac{\|x\|}{\epsilon}) + \frac{\|y\|^2}{2\epsilon} + \frac{|x|^2}{2\epsilon} \geq \frac{|x|^2}{2\epsilon}.$$

Mentre f_ϵ in 0 fornisce

$$f_\epsilon(x) = \frac{\|x\|^2}{2\epsilon}.$$

Allora si ottiene

$$f_\epsilon(x) = \begin{cases} \frac{\|x\|^2}{2\epsilon} & \|x\| \leq \epsilon \\ \|x\| - \frac{\epsilon}{2} & \|x\| > \epsilon. \end{cases}$$

Esercizio: fare il grafico nel caso unidimensionale per alcuni valori di ϵ.

10. Ottimizzazione

$$\min f(x) \qquad\qquad x \in \mathcal{F}$$

- $\mathcal{F} = \mathbb{R}^n$, non vincolato
- $\mathcal{F} \subset \mathbb{R}^n$ problema vincolato

$\mathcal{F}$ insieme chiuso. Caso particolare espresso come

$$\mathcal{F} = \{x \in \mathbb{R}^n : g(x) \leq 0 \; h(x) = 0\},$$

con $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$ un vincolo g_k si dice attivo in x^* se $g_k(x^*) = 0$.

Nella teoria che svilupperemo f, g e h sono assunte C^1, abbiamo come primo punto il problema dell'esistenza del minimo. Assumeremo che $\mathcal{F} \neq \emptyset$. Osserviamo che se g è una funzione convessa in $\mathbb{R}^n$ allora l'insieme

$$\mathcal{F} = \{x \in \mathbb{R}^n : g(x) \leq 0\},$$

è un insieme convesso perché intersezione di un numero finito di insieme convessi.
Il problema di minimizzazione si dice convesso che la funzione da minimizzare è una funzione convessa e l'insieme dei vincoli è un insieme convesso.

Non è detto che una funzione convessa in un insieme convesso ammetta minimo. Sussiste il seguente risultato

Proposizione 7. *Sia $C \subset \mathbb{R}^N$ un insieme convesso e $f : C \to \mathbb{R}$ una funzione strettamente convessa. Allora ogni punto di mínimo locale è punto di mínimo globale.*

Proof. Assumiamo che ci sia un punto x^* di minimo locale non globale. Allora esiste $\delta > 0$ tale che

$$f(x^*) \leq f(x), \qquad \forall x \in C \cap B_\delta(x^*)$$

e, per qualche $\hat{x}$, $f(\hat{x}) < f(x^*)$. Poichè $\hat{x} \neq x^*$, prendiamo $\lambda \in (0, \min\{1, \frac{1}{|\hat{x}-x^*|}\}$ e $x_\lambda = \lambda\hat{x} + (1-\lambda)x^*$. Allora

$$f(x_\lambda) \leq \lambda f(\hat{x}) + (1-\lambda)f(x^*) < \lambda f(x^*) + (1-\lambda)f(x^*) = f(x^*),$$

contraddicendo l'ipotesi che x^* sia un punto di minimo locale. □

10.1. Ottimizzazione con vincoli unilaterali e bilaterali.

Il problema è il seguente:

Data $f : \mathbb{R}^N \to \mathbb{R}$ e $g : \mathbb{R}^N \to \mathbb{R}^M$, $h : \mathbb{R}^N \to \mathbb{R}^P$, trovare

$$\min \quad \{f(x) : x \in \mathbb{R}^N \text{s.t. } g_i(x) \geq 0,\ i = 1, \dots, m, \\ h_i(x) = 0,\ i = 1, \dots, P\} \tag{10.1}$$

11. Condizioni necessarie: il teorema di Fritz-John

Teorema 11.1. *Sia I un sottoinsieme di $\mathbb{R}^N$, $f : I \to \mathbb{R}$, $g : I \to \mathbb{R}^M$, $h : I \to \mathbb{R}^P$, funzioni $C^1(I)$ e $x_0 \in I$. Se esiste un intorno aperto U di x_0 in $\mathbb{R}^N$ tale che*

$$f(x_0) \le f(x) \qquad \forall x \in U \cap \{x \in I : g(x) \le 0,\, h(x) = 0\}$$

allora esistono λ_0, $\lambda = (\lambda_1, \dots, \lambda_M)$ e $\mu = (\mu_1, \dots, \mu_P)$ tali che

i)
(11.1)
$$\begin{cases} \lambda_0 \frac{\partial f}{\partial x_i}(x_0) + \sum_{j=1}^M \lambda_j \frac{\partial g_j}{\partial x_i}(x_0) + \sum_{j=1}^P \mu_j \frac{\partial h_j}{\partial x_i}(x_0) = 0,\ i = 1, \dots, N \\ \lambda_i g_i(x_0) = 0,\ i = 1, \dots, M,\ (\lambda_0, \lambda) \ge 0,\ (\lambda_0, \lambda, \mu) \ne 0 \\ g(x_0) \le 0,\ h(x_0) = 0 \end{cases}$$

Proof. Dalla Definitione di minimo vincolato e dalla continuità di f, g e h possiamo considerare $\delta > 0$ tale che per ogni $x \in B(x_0, \delta) \cap \{x \in I : g(x) \le 0,\, h(x) = 0\}$

$$\begin{aligned} &f(x_0) \le f(x) \\ &g_i(x) < 0 \qquad \text{se } g_i(x_0) < 0 \end{aligned}$$

Introduzione il funzionale penalizzato

$$\mathcal{F}_k(x) = f(x) + \frac{1}{2}\|x - x_0\|^2 + \frac{k}{2}\left(\sum_{i_1}^M g_i^+(x)^2 + \sum_{i_1}^P h_i(x)^2\right)$$

dove $g_i(x)^+ = \max\{g_i(x), 0\}$ il cui quadrato è una funzione C^1 con gradiente $2g_i^+(x)Dg_i(x)$. Consideriamo il problema di ottimizzazione

$$\min_{x \in B(x_0,\delta)} \mathcal{F}_k(x) \tag{11.2}$$

Dal teorema di Weierstrass, esiste x_k punto di minimo per $\mathcal{F}_k$ in $\overline{B(x_0, \delta)}$. In particolare abbiamo

$$\mathcal{F}_k(x_k) \le \mathcal{F}_k(x_0) = f(x_0) \tag{11.3}$$

(poiché $g_i(x_0) \le 0$ e $h_i(x_0) = 0$). Inoltre, per il teorema di Bolzano-Weierstrass (ogni successione reale limitata ammette almeno una sottosuccessione convergente), la successione $\{x_k\}_{k\in\mathbb{N}}$ ammette una sottosuccessione che denotiamo ancora con $\{x_k\}$ convergente a $x^* \in \overline{B(x_0, \delta)}$. Dalla (11.3)

$$\sum_{i=1}^M g_i^+(x_k)^2 + \sum_{i=1}^P h_i(x_k)^2 \le \frac{2}{k}\left(f(x_0) - f(x_k) - \frac{1}{2}\|x_k - x_0\|^2\right)$$

e per la continuità di g_i, h_i abbiamo per $k \to \infty$

$$\sum_{i=1}^M g_i^+(x^*)^2 + \sum_{i=1}^P h_i(x^*)^2 \le 0$$

e quindi poichè

$$g_i(x)^= \begin{cases} g_i(x) & g_i(x) > 0 \\ 0 & g_i(x) \le 0 \end{cases}$$

Esempio $\max(0, \sin x)$ e $\max(0, \sin x))^2$

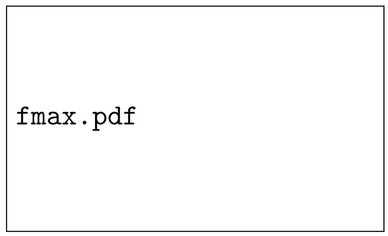

Figure 1. max(0, sin x) e $(\max(0, \sin x))^2$

$$g_i(x^*) \le 0,\ i = 1, \dots, M, \text{ e } h_i(x^*) = 0,\ i = 1, \dots, P. \tag{11.4}$$

In più dalla (11.3), abbiamo

$$f(x_k) + \frac{1}{2}\|x_k - x_0\|^2 \le \mathcal{F}_k(x_k) \le f(x_0)$$

e passando al limite per $k \to \infty$ abbiamo

$$f(x^*) + \frac{1}{2}\|x^* - x_0\|^2 \le f(x_0). \tag{11.5}$$

Dalla (11.4), $x^* \in \{x \in I : g(x) \le 0, h(x) = 0\}$ e quindi $f(x^*) \ge f(x_0)$. Allora (11.6) concludiamo che

$$\|x^* - x_0\|^2 = 0$$

e quindi

$$x^* = x_0.$$

Poiché $x_k \to x_0$, possiamo concludere che per k sufficientemente grande $x_k \in B(x_0, \delta)$ e allora, per il Teorema di Fermat, si ottiene

$$\begin{aligned} \frac{\partial \mathcal{F}_k}{\partial x_i}(x_k) = \frac{\partial f}{\partial x_i}(x_k) + (x_{k,i} - x_{0,i}) + \sum_{j=1}^{M} k g_j^+(x_k) \frac{\partial g_j}{\partial x_i}(x_k) \\ + \sum_{j=1}^{P} k h_j(x_k) \frac{\partial h_j}{\partial x_i}(x_k) = 0, \qquad i = 1, \dots, N \end{aligned} \tag{11.6}$$

Definiamo L^k, $\lambda_0^k \in \mathbb{R}$, $\lambda^k \in \mathbb{R}^M$, $\mu^k \in \mathbb{R}^P$

$$L^k = \left(1 + \sum_{j=1}^{M}(kg_j^+(x_k))^2 + \sum_{j=1}^{P}(kh_j(x_k))^2\right)^{\frac{1}{2}},$$

$$\lambda_0^k = \frac{1}{L^k}, \qquad \lambda_i^k = \frac{kg_i^+(x_k)}{L^k}, \qquad \mu_i^k = \frac{kh_i(x_k)}{L^k}$$

allora

$$|(\lambda_0^k, \lambda^k, \mu^k)|^2 = \left(\frac{1}{L^k}\right)^2 + \sum_{j=1}^{M}\left(\frac{kg_j^+(x_k)}{L^k}\right)^2 + \sum_{j=1}^{M}\left(\frac{kh_j(x_k)}{L^k}\right)^2 =$$

$$= \left(\frac{1}{L^k}\right)^2 \left(1 + \sum_{j=1}^{M}\left(kg_j^+(x_k)\right)^2 + \sum_{j=1}^{M}(kh_j(x_k))^2\right) = 1$$

Per il teorema di Bolzano-Weierstrass la successione

$$(\lambda_0^k, \lambda^k, \mu^k)_{k\in\mathbb{N}}$$

ammette una sottosuccessione che denotiamo ancora con $(\lambda_0^k, \lambda^k, \mu^k)$ convergente a $(\lambda_0, \lambda, \mu)$, tale che $|(\lambda_0, \lambda, \mu)| = 1$. Dividendo (11.6) per L^k, si ha

$$\lambda_0^k \frac{\partial f}{\partial x_i}(x_k) + \frac{(x_{k,i} - x_{0,i})}{L^k} + \sum_{j=1}^{M} \lambda_j^k \frac{\partial g_j}{\partial x_i}(x_k) + \sum_{j=1}^{P} \mu_j^k \frac{\partial h_j}{\partial x_i}(x_k) = 0 \tag{11.7}$$

e ricordando che $x_k \to x_0$ e $(\lambda_0^k, \lambda^k, \mu^k) \to (\lambda_0, \lambda, \mu)$ otteniamo la prima condizione in (11.1). Poiché $\lambda_0^k, \lambda^k \geq 0$ abbiamo al limite $\lambda_0, \lambda \geq 0$.

Sia i tale che $g_i(x_0) < 0$, allora $g_i(x_k) < 0$ e quindi

$$\lambda_i^k = \max\{g_i(x_k), 0\} = 0 \tag{11.8}$$

Quindi se $g_i(x_0) < 0$, abbiamo

$$\lambda_i g(_i(x_0) = 0.$$

Con gli stessi argomenti per gli altri indici i i, $g_i(x_0) = 0$, possiamo scrivere $\lambda_i g_i(x_0) = 0$ per ogni $i = 1, \ldots, M$ ottenendo in questo modo (11.1).

□

12. Penalizzazione

Problema vincolato

$$\begin{cases} \min f(x) & x \in \mathbb{R}^n \\ x \in K \end{cases}$$

Problema libero con penalizzazione $\gamma > 0$

$$\begin{cases} \min f(x) + \gamma P(x) & x \in \mathbb{R}^n \end{cases}$$

ove $P(x)$ verifica

$$\begin{cases} P(x) \in C(\mathbb{R}^n) \\ P(x) \geq 0 \qquad \forall x \in \mathbb{R}^n \\ P(x) = 0 \iff x \in K \end{cases}$$

Esercizio 12.1. *Disegnare l'insieme ammissibile definito tramite le funzioni di penalizzazione Siano g_1 e g_2 due funzioni da $\mathbb{R}$ in $\mathbb{R}$ definite da*

$$g_1(x) = x - 3 \qquad g_2(x) = -(x+2)^3$$

Prendiamo

$$p^+(x) = (\max(p(x), 0))^2$$

$$g_1^+(x) \begin{cases} x - 3 & x \geq 3 \\ 0 & x < 3 \end{cases}$$

$$g_2^+(x) \begin{cases} -(x+2)^3 & x \leq -2 \\ 0 & x > -2 \end{cases}$$

Esercizio 12.2. *Costruire la soluzione penalizzata del problema con la funzione di penalizzazione di Courant-Beltrami*

$$p^+(x) = (\max(p(x), 0))^2$$

$$\begin{cases} \min x \\ x \geq a \end{cases}$$

$$p^+(x) = (\max(a - x, 0))^2$$

Per $\gamma > 0$ consideriamo

$$F_\gamma(x) = x + \gamma(\max(a - x, 0))^2.$$

Completare confrontando il risultato.

Penalizzazione con vincolo di uguaglianza

$$n = 2,\ \min \frac{1}{2}(x_1^2 + x_2^2), Ax = b,\ A = [1\ 1],\ b = 2$$

$$n = 2,\ \min \frac{1}{2}(x_1^2 + x_2^2), \quad x_1 + x_2 = 2$$

Trovare la soluzione. Trovare la soluzione con penalizzazione

$$\min \frac{1}{2}(x_1^2 + x_2^2) + \gamma(x_1 + x_2 - 2)^2$$

e confrontare con la soluzione. Generalizzare al caso

$$F(x) = \min \frac{1}{2}\|x\|^2 + \gamma\|Ax - b\|^2$$

13. Programmazione lineare (PL)

Un problema di PL in forma standard è espresso come

$$\begin{cases} \min cx \qquad c \in \mathbb{R}^n \\ x \in K = \{x \in \mathbb{R}^n : \ Ax = b, \ x \geq 0\}, \\ A = (a_{i,j}) \ i = 1, \dots, m \qquad j = 1, \dots, n \qquad b \in \mathbb{R}^m \end{cases}$$

Infatti vediamo come le altre condizioni possono essere

$$\begin{cases} \min cx \qquad c \in \mathbb{R}^n \\ x \in K = \{x \in \mathbb{R}^n : \ Ax \geq b, \ x \geq 0\} \end{cases}$$

e

$$\begin{cases} \min cx \qquad c \in \mathbb{R}^n \\ x \in K = \{x \in \mathbb{R}^n : \ Ax \leq b, \ x \geq 0\} \end{cases}$$

possono essere ricondotte al caso standard mediante aggiunta di variabile.

Caso I $\qquad x \in \mathbb{R}^n : \ Ax \geq b, \ x \geq 0.$

Vediamo come: la condizione

$$Ax = b \iff a_{i,1}x_1 + a_{i,2}x_2 \cdots + a_{i,n}x_n = b_i \qquad i = 1 \quad , m$$

Introduciamo

$$\begin{cases} \min cx \qquad c \in \mathbb{R}^n \\ a_{i,1}x_1 + a_{i,2}x_2 \cdots + a_{i,n}x_n - y_i = b_i \qquad i = 1 \quad , \ x \geq 0, \ y \geq 0\} \end{cases}$$

che possiamo riscrivere

$$Ax - I_m y = [A, -I_m][x, y]^T = b \qquad x \geq 0, \ y \geq 0$$

Caso II $\qquad x \in \mathbb{R}^n : \ Ax \leq b, \ x \geq 0.$

Intrduciamo

$$\begin{cases} \min cx \qquad c \in \mathbb{R}^n \\ a_{i,1}x_1 + a_{i,2}x_2 \cdots + a_{i,n}x_n + y_i = b_i \qquad i = 1 \quad , \ x \geq 0, \ y \geq 0\} \end{cases}$$

che possiamo riscrivere

$$Ax + I_m y = [A, I_m][x, y]^T = b \qquad x \geq 0, \ y \geq 0$$

A prima vista i problemi possono apparire differenti. In effetti geometricamente la condizione legata alle disuguaglianze da luogo a intersezioni di semispazi nello spazio n-dimensionale mentre il problema con variabili aggiunte da luogo a intersezioni di semispazi e iperpiani nello spazio $n + m$ dimensionale.

Vediamo su un esempio

$$x_1 \leq 3,$$

diventa

$$\begin{cases} x_1 + x_2 = 3, \\ x_2 \geq 0 \end{cases}$$

e

$$x_1 \geq 3,$$

diventa

$$\begin{cases} x_1 - x_2 = 3 \\ x_2 \geq 0 \end{cases}$$

Disegnare gli insiemi e verificare che possiamo associare al primo problema le soluzioni $(x_1, 0)$.

Osserviamo che in un problema di programmazione lineare può essere $K = \emptyset$ oppure il problema può essere inferiormente illimitato. Vediamo in un semplice esempio il secondo caso

$$\begin{cases} \min x \\ x \leq 2 \end{cases}$$

Trattatiamo un problema di esistenza quando $K \neq \emptyset$ e il problema si assume essere inferiormente limitato. Dapprima ricordiamo

13.1. **Cono convesso.** Un cono è un sottoinsieme C di $\mathbb{R}^m$ chiuso rispetto alla moltiplicazione per scalari positivi, cioè

$$x \in C, \quad \lambda > 0 \implies \lambda x \in C$$

I coni appaiono naturalmente come rappresentazione dei vincoli. Consideriamo

$$C = \{y \in \mathbb{R}^m : y = Ax, \quad x \geq 0\}.$$

C è un cono. Inoltre C è un cono convesso ossia se v e w appartengono a C allora se $0 \leq \alpha \leq 1$, risulta $\alpha v + (1 - \alpha)w \in C$. Si può dimostrare che C è un cono convesso chiuso (un insieme C è chiuso se e solo se comunque presa una successione convergente $\{x_k\} \subset C$ si ha $\lim_{k \in +\infty} x_k = x \in C$).

Teorema 13.1. *Se $K \neq \emptyset$, e la funzione cx è inferiormente limitata in K, allora il problema di minimizzazione*

$$\begin{cases} \min cx \qquad c \in \mathbb{R}^n \\ x \in K = \{x \in \mathbb{R}^n : \ Ax \geq b, \ x \geq 0\} \end{cases}$$

ammette soluzione ottimale

La dimostrazione utilizza principalmente la riduzione del problema in forma standard e le proprietà del problema dei vincoli in forma standard viste precedentemente.

Esercizi.

Poliedro convesso: Si ottiene come intersezione di un numero finito di semispazi. Politopi: Un politopo n-dimensionale o n-politopo è l'analogo di un poligono nel piano (n=2). Politopo corrispone al caso poliedro limitato.

esempio di ottenimento di politopo e problema di minimizzazione con una funzione concava su un politopo. Ricordiamo che un problema con funzione obiettivo lineare e insieme ammissibile convesso è sia concavo che convesso.

Per le funzioni concave si ha il seguente

Teorema 13.2. *Sia $C \subset \mathbb{R}^n$ un insieme aperto convesso e f concava (non costante) su $\overline{C}$. Se esiste un punto di minimo globale del problema di minimizzazione* $\min f(x), \quad x \in \overline{C}$*, allora ogni punto di minimo globale appartiene alla frontiera di C.*

Proof. Assumiamo che ci sia un punto x^* di minimo globale soluzione del problema di ottimizzazione. Poiché f non è costante, allora esiste almeno un punto $\hat{x} \in \overline{C} : f(\hat{x}) > f(x*)$. Consideriamo Sia z un punto interno a C e $B_\delta(z)$, $\delta > 0$ un intorno di raggio δ. Sul segmento $[\hat{x}z]$ determiniamo $y \in B_\delta(z)$ diverso da z tale che

$$z = \lambda y + (1 - \lambda)\hat{x},$$

per $0 \leq \lambda < 1$. Allora

$$f(z) \geq (1 - \lambda)f(\hat{x}) + \lambda f(y) > (1 - \lambda)f(x^*) + \lambda f(x^*).$$

Se z fosse un punto di mínimo globale si avrebbe una contraddizione, in quanto $f(z) > f(x^*)$, contro l'ipotesi che x^* sia un punto di minimo globale soluzione del problema di ottimizzazione. □

Nel caso di PL (Programmazione Lineare) allora se il problema di minimizzazione ammette soluzione allora il punto di minimo globale si trova nella frontiera del poliedro ammissibile.

14. CONDIZIONI KKT ($\lambda_0 \neq 0$): ESEMPI

Siamo interessati a trovare condizioni per cui (11.1) $\lambda_0 > 0$.

Corollario 1. *Sotto le stesse ipotesi del teorema 11.1, definiamo l'insieme degli indici $I^*(x_0) = \{i \in \{1, \dots, M\} : g(x_0) = 0\}$ (vincoli attivi) e assumiamo che i $\#(I^*(x_0) + p)$ vettori $\{Dg_i(x), i \in I^*(x_0)\}$, $\{Dh_i(x_0), i = 1, \dots, p\}$ siano linearmente indipendenti. Allora esiste $\lambda = (\lambda_1, \dots, \lambda_M)$ e $\mu = (\mu_1, \dots, \mu_P)$ tali che*

$$\begin{cases} \frac{\partial f}{\partial x_i}(x_0) + \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i}(x_0) + \sum_{j=1}^p \mu_j \frac{\partial h_j}{\partial x_i}(x_0) = 0, \; i = 1, \dots, N \\ \lambda_i g_i(x_0) = 0, \; i = 1, \dots, M, \\ g(x_0) \leq 0, \; h(x_0) = 0, \; \lambda \geq 0 \end{cases} \tag{14.1}$$

Proof. Dal teorema 11.1 sappiamo che esiste λ_0, λ e μ, non tutti nulli, tali che le condizioni (11.1) sono soddisfatte. Vogliamo dimostrare che $\lambda_0 \neq 0$. Assumiamo per assurdo che $\lambda_0 = 0$, allora ricordando che $\lambda_i = 0$ se $g_i(x_0) < 0$, abbiamo

$$\sum_{j \in I^*(x_0)} \lambda_j \frac{\partial g_j}{\partial x_i}(x_0) + \sum_{j=1}^p \mu_j \frac{\partial h_j}{\partial x_i}(x_0) = 0 \qquad i = 1, \dots, N.$$

Per l'indipendenza lineare dei vettori, abbiamo $\lambda = 0$ e $\mu = 0$. Non è possibile. Allora $\lambda_0 \neq 0$ e possiamo dividere per λ_0 nella prima condizione (11.1) e ottenere (14.1). □

Un altro caso importante di regolarità è il seguente

Corollario 2. *Sotto le stesse ipotesi del Teorema 11.1, assumiamo che la funzione h sia lineare e le funzioni g_i, $i = 1, \ldots, M$ concave in x_0. Allora esistono $\lambda = (\lambda_1, \ldots, \lambda_M)$ e $\mu = (\mu_1, \ldots, \mu_P)$ tali che sussista la* (14.1)

Proof. In un intorno di x_0 possiamo scrivere

$$\begin{aligned} &h_i(x) = h_i(x_0) + Dh_i(x_0)(x - x_0) \\ &g_i(x) \le g_i(x_0) + Dg_i(x_0)(x - x_0) \qquad i \in I^*(x_0) \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^{p} \mu_i h_i(x) + \sum_{i=1}^{M} \lambda_i g_i(x) \le \sum_{i=1}^{p} \mu_i h_i(x_0) + \sum_{i \in I^*(x_0)} \lambda_i g_i(x_0) \\ + \left(\sum_{i=1}^{P} \mu_i Dh_i(x_0) + \sum_{i \in I^*(x_0)} \lambda_i Dg_i(x_0) \right) (x - x_0) \end{aligned}$$

Se in (11.1) assumiamo per assurdo $\lambda_0 = 0$ e ricordiamo $\lambda_i = 0$ per $i \notin I^*(x_0)$, dalla precedente disuguglianza otteniamo

$$\sum_{i=1}^{p} \mu_i h_i(x) + \sum_{i=1}^{M} \lambda_i g_i(x) \le 0.$$

Vale (con le stesse notazioni F-J) (da dimostrare) *ii)* in un qualsiasi intorno x_0, esiste x tale

$$\begin{aligned} &\lambda_i g_i(x) > 0 \qquad \forall\ i \text{ s.t. } \lambda_i > 0 \\ &\mu_i h_i(x) > 0 \qquad \forall\ i \text{ s.t. } \mu_i \neq 0 \end{aligned} \tag{14.2}$$

Se $\lambda \neq 0$ e $\lambda_0 = 0$, esiste un punto x nell'intorno di x_0 tale che

$$\sum_{i=1}^{p} \mu_i h_i(x) + \sum_{i=1}^{M} \lambda_i g_i(x) > 0$$

e quindi una contraddizione per $\lambda_0 = 0$. □

Esempio 1. *Le condition* (14.1) *sono le condizioni* Kurush-Kuhn-Tucker (KKT) *: sono condizioni necessarie*

Esempio 2. *La Lagrangiana $L : \mathbb{R}^N \times \mathbb{R}^M_+ \times \mathbb{R}^p$ associata al problema di ottimizzazione* (10.1) *è data da*

$$L(x, \lambda, \nu) = f(x) + \lambda g(x) + \nu h(x), \tag{14.3}$$

con $\lambda, \nu \in \mathbb{R}^M_+ \times \mathbb{R}^P$. Le condizioni KKT conditions possono essere riformulate

$$\begin{cases} \frac{\partial L}{\partial x_i}(x_0, \lambda, \mu) = 0,\ i = 1, \ldots, N \\ \lambda_i g_i(x_0) = 0,\ i = 1, \ldots, M, \\ g(x_0) \le 0,\ h(x_0) = 0,\ \lambda \ge 0 \end{cases} \tag{14.4}$$

15. Caso Convesso

Sotto alcune ipotesi in pù le condizioni KKT (14.4) diventano sufficienti.

Teorema 15.1. *Sotto le ipotesi 11.1, asumiamo f e g_i, $i = 1\dots, M$ convesse e $h(x) = Ax - b$. Allora x_0 è soluzione di* (10.1) *se e solo se soddisfa* (14.4).

In più, se f è strettamente convessa, x_0 è l'unico punto soluzione di(10.1).

Proof. Poiché $\lambda \geq 0$ per ogni $x \in \{x \in I : \, g(x) \leq 0, \, h(x) = 0\}$,

$$f(x) \geq f(x) + \lambda g(x) + \mu h(x).$$

In più dalla linearità di h, la convessità di f and g_i e $\lambda_0 \geq 0$ abbiamo

$$\begin{aligned} h(x) &= h(x_0) + Jh(x_0)(x - x_0) \\ f(x) &\geq f(x_0) + Df(x_0)(x - x_0) \\ \lambda g(x) &\geq \lambda g(x_0) + \lambda Jg(x_0)(x - x_0) \end{aligned}$$

$$\begin{aligned} f(x) \geq f(x) + \lambda g(x) + \mu h(x) &\geq f(x_0) + Df(x_0)(x - x_0) \\ +\lambda g(x_0) + \lambda Jg(x_0)(x - x_0) &+ \mu h(x_0) + \mu Jh(x_0)(x - x_0) \\ \geq f(x_0) + \left(Df(x_0) + Jg(x_0)^T \lambda + Jh(x_0)^T \mu\right) &(x - x_0) = f(x_0) \end{aligned}$$

per x ammissibile, allora x_0 è un punto di minimo .

Se f è strettamente convessa un calcolo simile al precedente $f(x) > f(x_0)$ per ogni x ammissibile, mostrando che x_0 è l'unico puntodi minimo globale. □

15.1. **Vincoli di non negatività.** Consideriamo la seguente classi di problemi

$$\min\{f(x) : x \in \mathbb{R}^N \text{tali che} \, x \geq 0, \, i = 1, \dots, M\} \tag{15.1}$$

ossia i vincoli sono lineari con A la matrice identità e $b = 0$. Otteniamo

$$\begin{aligned} Df(x_0) - \lambda = 0 \\ x_0 \geq 0, \, \lambda \geq 0, \, \lambda x = 0 \end{aligned}$$

quindi $\lambda_i = \frac{\partial f}{\partial x_i}(x_0)$ e

$$\begin{aligned} \frac{\partial f}{\partial x_i}(x_0) \geq 0 \quad \text{se } x_{0,i} = 0 \\ \frac{\partial f}{\partial x_i}(x_0) = 0 \quad \text{se } x_{0,i} > 0 \end{aligned}$$

15.2. **Vincoli di tipo** ***box*****.** Consideriamo la seguente classe di problemi

$$\min\{f(x) : x \in \mathbb{R}^N \text{tali che}\, a_i \leq x_i \leq b_i,\ i = 1, \dots, N\} \tag{15.2}$$

dove $a, b \in \mathbb{R}^N$ con $a_i < b_i$. Consideriamo la Lagrangiana

$$L(x, \lambda) = f(x) + \lambda(a - x) + \nu(x - b)$$

Otteniamo

$$\begin{aligned}
&Df(x_0) - \lambda + \nu = 0\\
&(a - x_0)\lambda = 0,\ (x_0 - b)\nu = 0,\ (\lambda, \nu) \geq 0
\end{aligned}$$

Poniamo

$$J_a = \{j : x_{0,j} = a_j\},\ J_b = \{j : x_{0,j} = b_j\},\ J_a = \{j : a_j < x_{0,j} < b_j\}$$

Se $j \in J_a$, e $x_j < b_j$, allora $\nu_j = 0$. Ne segue

$$\frac{\partial f}{\partial x_j}(x_0) = \lambda_j \geq 0.$$

In modo simile, se $j \in J_b$, $\lambda_j = 0$ e

$$\frac{\partial f}{\partial x_j}(x_0) = -\nu_j \leq 0.$$

Se $j \in J_0$, allora $\lambda_j = \nu_j = 0$ e quindi

$$\frac{\partial f}{\partial x_j}(x_0) = 0$$

Le condizioni necessarie di ottimalità sono

$$\begin{aligned}
&\frac{\partial f}{\partial x_j}(x_0) \geq 0 \quad \text{se } x_{0,j} = a_j\\
&\frac{\partial f}{\partial x_j}(x_0) \leq 0 \quad \text{se } x_{0,j} = b_j\\
&\frac{\partial f}{\partial x_j}(x_0) = 0 \quad \text{se } a_j < x_{0,j} < b_j.
\end{aligned}$$

Queste conditioni sono anche sufficienti se f è convessa.

15.3. **Condizioni KKT per un problema in forma standard.** Consideriamo la seguente classe di problemi

$$\min\{f(x) : x \in \mathbb{R}^N \text{tali che} Ax = b,\ x \geq 0\} \tag{15.3}$$

Consideriamo la Lagrangiana

$$L(x, \lambda) = f(x) + \mu(Ax - b) - \lambda x$$

Otteniamo

$$\begin{aligned}
&Df(x_0) - \lambda + A^T\mu = 0\\
&\lambda x_0 = 0,\ \lambda \geq 0\\
&Ax_0 = b,\ x_0 \geq 0
\end{aligned}$$

Esempio di applicazione del risultato.

16. Il problema della dieta e il suo duale

Consideriamo n cibi diversi tra loro

$$C_1, C_2, \dots, C_n.$$

Consideriamo m sostanze che il paziente deve assumere ogni giorno

$$V_1, V_2, \dots, V_m.$$

Inoltre p_i denota il costo del cibo C_i, $a_{i,j}$ quantità di sostanza nutritiva V_i nel cibo C_j, e b_i quantità minima di sostanza nutritiva V_i che deve essere garantita ogni giorno.

Si vuole minimizzare il costo giornaliero della dieta

$$\min\{px : x \in \mathbb{R}^n \text{tali che} Ax \geq b,\ x \geq 0\} \tag{16.1}$$

Consideriamo la Lagrangiana

$$L(x, \lambda) = px + \lambda_1(b - Ax) - \lambda_2 x$$

Otteniamo

$$\begin{aligned}
&p = \lambda_2 + A^T \lambda_1 \\
&\lambda_1 \geq 0,\ \ \lambda_2 \geq 0 \\
&Ax_0 \geq b,\ x_0 \geq 0 \\
&\lambda_1(b - Ax_0) = 0\ \ \lambda_2 x_0 = 0
\end{aligned}$$

Problema duale.

16.1. **Funzione Lagrangiana e sua duale.**

Ritorniamo al problema

$$\min f(x) \qquad \text{sotto le condizioni } g(x) \leq 0 \qquad h(x) = 0,$$

$x \in \mathcal{F}$ insieme ammissibile p_{opt} il minimo,

$$\mathcal{F} = (\cap_{i=1}^m \mathrm{dom}(g_i)) \cap (\cap_{i=1}^p \mathrm{dom}(h_i))$$

$$p_{opt} = \inf \Big[f(x) \,\text{sotto le condizioni } g(x) \leq 0 \qquad h(x) = 0 \Big] = \inf \Big[f(x) : x \in \mathcal{F} \Big]$$

La funzione Lagrangiana è definita

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{i=1}^p \mu_i h_i(x).$$

Ricordiamo che la funzione Lagrangiana è definita nell'insieme $\mathcal{F}$. Poiché $g_i(x) \leq 0$ la funzione Lagrangiana assumerà valori inferiori o uguali alla funzione obiettivo f.

Si definisce la funzione duale della Lagrangiana L la funzione G $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}^p$

$$G(\lambda, \mu) = \inf_{x \in \mathcal{F}} L(x, \lambda, \mu) = \inf_{x \in \mathcal{F}} \big(f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{i=1}^p \mu_i h_i(x) \big)$$

La funzione duale fornisce un limite inferiore del valore ottimale

$$G(\lambda,\mu) \le p_{opt}$$

Infatti se $x \in \mathcal{F}$ e $\lambda > 0$ allora

$$\sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \le 0,$$

e

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \le f(x)$$

la disuguaglianza vale passando all'estremo inferiore su x.

La funzione duale è una funzione concava perché estremo inferiore puntuale di funzioni affini.

La funzione duale è definita nell'insieme

$$Dom(G) = \{(\lambda,\mu) : G(\lambda,\mu) > -\infty\}$$

Esempio.

$$\min cx \qquad \text{sotto le condizioni } Ax = b \qquad x \ge 0$$

$$L(x,\lambda,\mu) = cx - \lambda x + \mu(Ax - b) = -\mu b + (c - \lambda + A^T\mu)x$$

Allora

$$G(\lambda,\mu) = \begin{cases} -\mu b & c - \lambda + A^T\mu = 0 \\ -\infty & \text{altrimenti} \end{cases}$$

16.2. **Problema duale.**

$$\max G(\lambda,\mu) \qquad \text{sotto la condizione } \lambda \ge 0$$

$$d_{opt} = \max_{\lambda \ge 0} G(\lambda,\mu)$$

$$(\lambda,\mu) \in dom(G)$$

Esempio

Problema

$$\min cx \qquad \text{sotto le condizioni } Ax = b \qquad x \ge 0$$

$$L(x,\lambda,\mu) = cx - \lambda x + \mu(Ax - b) = -\mu b + (c - \lambda + A^T\mu)x$$

Scrivere il problema duale.

16.3. **Esercizio n.1.** Nel caso di assenza di vincoli di uguaglianza

$$\min f(x) \qquad \text{sotto le condizioni } g(x) \leq 0,$$

$\mathcal{F}$ insieme ammissibile p_{opt} il minimo. La funzione Lagrangiana è definita

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x))$$

Si definisce la funzione duale della Lagrangiana L la funzione G

$$G(\lambda) = \inf_{x \in \mathcal{F}} L(x, \lambda) = \inf_{x \in \mathcal{F}} \big(f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)\big)$$

16.4. **Esercizio n.2.** Nel caso di assenza di vincoli di disuguaglianza

$$\min f(x) \qquad \text{sotto le condizioni } h(x) = 0,$$

$\mathcal{F}$ insieme ammissibile p_{opt} il minimo. La funzione Lagrangiana è definita

$$L(x, \mu) = f(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

Si definisce la funzione duale della Lagrangiana L la funzione G

$$G(\mu) = \inf_{x \in \mathcal{F}} L(x, \mu) = \inf_{x \in \mathcal{F}} \big(f(x) + \sum_{i=1}^{p} \mu_i h_i(x)\big)$$

16.5. **Esercizio n.3.** Minimizzare

$$f(x) = \frac{1}{2} \|x\|^2 \qquad \text{sotto le condizioni } Ax = b$$

$$L(x, \mu) = \frac{1}{2} \|x\|^2 + \mu(Ax - b)$$

Per minimizzare L calcoliamo

$$DL(x, \mu) = x + A^T \mu = 0 \implies x = -A^T \mu$$

Allora

$$G(\mu) = -\frac{1}{2} \left\|A^T \mu\right\|^2 - \mu b.$$

La funzione G risulta concava. Vero in generale.In questo caso la funzione G è definita ovunque ma può risultare che la funzione G sia uguale a $-\infty$ in sottoinsiemi.

16.6. **Dualità debole e forte.** Dualità debole

$$p_{opt} \geq d_{opt}$$

Dualità forte

$$p_{opt} = d_{opt}$$

Esempio.

Condizione di Slater Condizione sufficiente per la dualità forte.