Programmazione Matematica (A-L)

Ingegneria Informatica e Automatica - Sapienza Università di Roma

Dispense del corso: argomenti di teoria *

Simone Sagratella

1 Elementi di analisi matematica in $\mathbb{R}^n$

1.1 Norme

Dato un punto $x \in \mathbb{R}^n$ una **norma** è una funzione che prende in input x e ne restituisce una sua distanza dall'origine. Ogni norma $\|\cdot\|$ gode delle seguenti proprietà:

(i) $\|x\| \geqslant 0, \forall\, x \in \mathbb{R}^n$

(ii) $\|x\| = 0$ se e solo se $x = 0$

(iii) $\|ax\| = |a|\|x\|, \forall\, a \in \mathbb{R}, x \in \mathbb{R}^n$

(iv) $\|x + y\| \leqslant \|x\| + \|y\|, \forall\, x, y \in \mathbb{R}^n$.

Si elencano le norme più usate:

- norma euclidea: $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$

*Font ad alta leggibilità Biancoenero© di biancoenero edizioni srl, disegnata da Umberto Mischi, con la consulenza di Alessandra Finzi, Daniele Zanoni e Luciano Perondi (Brevetto n. RM2011O000128). Disponibile gratuitamente per tutte le istituzioni e i privati che la utilizzino per scopi non commerciali.

- norma 1: $\|x\|_1 = \sum_{i=1}^n |x_i|$
- norma infinito: $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$

Dato $x \in \mathbb{R}^n$ risulta che

$$\|x\|_\infty \leqslant \|x\| \leqslant \|x\|_1 \leqslant \sqrt{n}\|x\| \leqslant n\|x\|_\infty.$$

Proposizione 1.1.1. *Siano $x, y \in \mathbb{R}^n$, si ha che*

$$x^T x = \|x\|^2, \quad |x^T y| \leqslant \|x\|\|y\|.$$

Dimostrazione. Si ha

$$x^T x = \sum_{i=1}^n x_i^2 = \|x\|^2.$$

Si ha

$$\begin{aligned}
0 &\leqslant \sum_{i=1}^n \sum_{j=1}^n (x_i y_j - x_j y_i)^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n x_i^2 y_j^2 + \sum_{i=1}^n \sum_{j=1}^n x_j^2 y_i^2 - 2 \sum_{i=1}^n \sum_{j=1}^n x_i y_j x_j y_i \\
&= \sum_{i=1}^n x_i^2 \sum_{j=1}^n y_j^2 + \sum_{i=1}^n y_i^2 \sum_{j=1}^n x_j^2 - 2 \sum_{i=1}^n x_i y_i \sum_{j=1}^n x_j y_j \\
&= \|x\|^2\|y\|^2 + \|y\|^2\|x\|^2 - 2(x^T y)(x^T y) \\
&= 2\|x\|^2\|y\|^2 - 2(x^T y)^2.
\end{aligned}$$

Quindi si ha $(x^T y)^2 \leqslant \|x\|^2\|y\|^2$ che implica la disuguaglianza voluta. □

1.2 Insiemi

Sia $X \subseteq \mathbb{R}$, esistono sempre $\inf(X) \in \mathbb{R} \cup \{-\infty, +\infty\}$ e $\sup(X) \in \mathbb{R} \cup \{-\infty, +\infty\}$, gli **estremi inferiore e superiore di** X per cui valgono le seguenti proprietà:

(i) $\inf(X) \leqslant x \leqslant \sup(X)$, $\forall\, x \in X$

(ii) $\nexists\, z > \inf(X)$ tale che $x \geqslant z, \forall\, x \in X$

(iii) $\nexists\, w < \sup(X)$ tale che $x \leqslant w, \forall\, x \in X$.

Se $\inf(X) = -\infty$ si dice che X è illimitato inferiormente. Se $\sup(X) = +\infty$ si dice che X è illimitato superiormente. Se X non contiene punti allora $\inf(X) = +\infty$ e $\sup(X) = -\infty$.

Sia $X \subseteq \mathbb{R}^n$. Si indica con ∂X la **frontiera** di X, definita come

$$\partial X \triangleq \{z \in \mathbb{R}^n \,:\, \mathcal{B}(z,\rho) \cap X \neq \varnothing,\, \mathcal{B}(z,\rho) \cap (\mathbb{R}^n \backslash X) \neq \varnothing,\, \forall\, \rho > 0\},$$

dove con $\mathcal{B}(y,\rho)$ si intende la sfera aperta di raggio $\rho > 0$ e centro $y \in \mathbb{R}^n$:

$$\mathcal{B}(y,\rho) \triangleq \{z \in \mathbb{R}^n \,:\, \|z - y\| < \rho\}.$$

Si indica con $\mathring{X}$ l'**interno** di X, definito come

$$\mathring{X} \triangleq X \backslash \partial X.$$

L'insieme X si dice **aperto** se $\partial X \cap X = \varnothing$. L'insieme X si dice **chiuso** se $\partial X \subseteq X$.

L'insieme X si dice **limitato** se esiste $D > 0$ tale che $\|x\| \leqslant D/2$ per ogni $x \in X$. Altrimenti l'insieme X si dice **illimitato**.

L'insieme X si dice **compatto** se è sia chiuso che limitato.

L'insieme X si dice **non vuoto** se esiste $x \in X$.

L'insieme X si dice **convesso** se per ogni coppia di punti $x \in X$ e $y \in X$ risulta $[x,y] \subseteq X$, dove $[x,y]$ indica il segmento chiuso tra x e y:

$$[x,y] \triangleq \{z \in \mathbb{R}^n \,:\, z = \lambda x + (1-\lambda)y,\, 0 \leqslant \lambda \leqslant 1\}.$$

Sia X un insieme convesso. Il punto v è un **vertice** di X se:

(i) $v \in X$

(ii) per ogni $x, y \in X$ si ha $v \notin (x,y)$, dove (x,y) indica il segmento aperto tra x e y:

$$(x,y) \triangleq \{z \in \mathbb{R}^n \,:\, z = \lambda x + (1-\lambda)y,\, 0 < \lambda < 1\}.$$

Sia $X \subseteq \mathbb{R}^n$, per ogni punto $x \in \mathbb{R}^n$ si definisce la **proiezione** di x su X come

$$P_X(x) \triangleq \arg\min \left\{ \|z - x\|^2 \, : \, z \in X \right\}.$$

Proposizione 1.2.1. *Siano $x \in \mathbb{R}^n$ e $X \subseteq \mathbb{R}^n$. Se X è chiuso e non vuoto allora $P_X(x)$ esiste sempre, se inoltre X è convesso allora $P_X(x)$ è unico.*

Dimostrazione. Sia $f(z) = \|z - x\|^2 = \sum_{i=1}^n (z_i - x_i)^2$.

La funzione $f(z)$ è continua (vedi Proposizione 1.4.19) e coerciva, quindi valgono le ipotesi del Corollario 1.4.5, e quindi $P_X(x)$ esiste.

La funzione $f(z)$ è strettamente convessa (vedi Proposizione 1.4.19), quindi valgono le ipotesi del Teorema 1.4.15, e quindi $P_X(x)$ è unico. □

Proposizione 1.2.2. *Siano $x, y \in \mathbb{R}^n$ e $X \subseteq \mathbb{R}^n$. Se $\|y - x\| < \|P_X(x) - x\|$ allora $y \notin X$.*

Dimostrazione. Se per assurdo $y \in X$, allora si ha una contraddizione con $\|P_X(x) - x\|^2 \leqslant \|z - x\|^2, \, \forall \, z \in X$. □

Proposizione 1.2.3. *Siano $x, y \in \mathbb{R}^n$ e $X \subseteq \mathbb{R}^n$ chiuso, non vuoto e convesso, valgono le seguenti proprietà:*

(i) *$P_X(x)$ è l'unico punto tale che*

$$(P_X(x) - x)^T (z - P_X(x)) \geqslant 0, \, \forall \, z \in X$$

(ii) *$\|P_X(x) - P_X(y)\| \leqslant \|x - y\|$.*

Dimostrazione. Sia $f(z) = \|z - x\|^2 = \sum_{i=1}^n (z_i - x_i)^2$.

Dal Teorema 1.4.6 si ha che

$$2(P_X(x) - x)^T (z - P_X(x)) = \nabla f(P_X(x))^T (z - P_X(x)) \geqslant 0, \, \forall \, z \in X,$$

quindi vale la proprietà *(i)*. L'unicità è conseguenza dei Teoremi 1.4.14 e 1.4.15.

Utilizzando la proprietà *(i)* si ottiene

$$(P_X(x)-x)^T(P_X(y)-P_X(x)) \geqslant 0$$

quando si pone $z = P_X(y)$, e si ottiene

$$(P_X(y)-y)^T(P_X(x)-P_X(y)) \geqslant 0$$

quando si scrive la *(i)* per y e si pone $z = P_X(x)$. Sommando queste due disuguaglianze si ottiene

$$\begin{aligned}
0 &\leqslant (P_X(x)-x)^T(P_X(y)-P_X(x)) + (P_X(y)-y)^T(P_X(x)-P_X(y)) \\
&= (x-P_X(x))^T(P_X(x)-P_X(y)) + (P_X(y)-y)^T(P_X(x)-P_X(y)) \\
&= (x-P_X(x)+P_X(y)-y)^T(P_X(x)-P_X(y)) \\
&= (x-y)^T(P_X(x)-P_X(y)) - (P_X(x)-P_X(y))^T(P_X(x)-P_X(y)) \\
&= (x-y)^T(P_X(x)-P_X(y)) - \|P_X(x)-P_X(y)\|^2 \\
&\leqslant \|x-y\|\|P_X(x)-P_X(y)\| - \|P_X(x)-P_X(y)\|^2,
\end{aligned}$$

dove nelle ultime relazioni si è usata la Proposizione 1.1.1. Dividendo per $\|P_X(x)-P_X(y)\|$ si ottiene la proprietà *(ii)* (se $\|P_X(x)-P_X(y)\| = 0$ la proprietà *(ii)* è ovvia). □

1.3 Successioni

Sia $\{x^k\}$ una **successione** infinita di punti appartenenti a un insieme $X \subseteq \mathbb{R}^n$, ovvero $x^k \in X$, per ogni $k \in \mathbb{N}$.

Si dice che la successione **converge** se esiste un punto $\overline{x} \in \mathbb{R}^n$ tale che

$$\lim_{k\to\infty} \|x^k - \overline{x}\| = 0,$$

ovvero se $\forall\, \varepsilon > 0, \exists\, \overline{k}_\varepsilon \in \mathbb{N} : \|x^k - \overline{x}\| < \varepsilon, \forall\, k \geqslant \overline{k}_\varepsilon$. In questo caso si può scrivere $\lim_{k\to\infty} x^k = \overline{x}$, oppure $x^k \to \overline{x}$, e si dice che $\overline{x}$ è il **punto limite** della successione.

Sia $K \subseteq \mathbb{N}$ un insieme infinito di indici. Si indica con $\{x^k\}_K$ la **sottosuccessione** formata dai punti di $\{x^k\}$ i cui indici appartengono a K. Se esiste un insieme infinito di indici $\widetilde{K} \subseteq \mathbb{N}$ tale che la sottosuccessione $\{x^k\}_{\widetilde{K}}$ converge a un punto $\widetilde{x} \in \mathbb{R}^n$, allora si dice che $\widetilde{x}$ è un **punto di accumulazione** della successione. In questo caso si può scrivere $\lim_{k\to\infty, k\in\widetilde{K}} x^k = \widetilde{x}$, oppure $x^k \underset{\widetilde{K}}{\to} \widetilde{x}$.

Proposizione 1.3.1. *Sia $\overline{x} \in \mathbb{R}^n$. La successione $\{x^k\}$ converge a $\overline{x}$ se e solo se per ogni insieme infinito di indici $K \subseteq \mathbb{N}$ la sottosuccessione $\{x^k\}_K$ converge a $\overline{x}$.*

Dimostrazione. (Sufficienza) Per assurdo esiste $\widetilde{K}$ tale che $x^k \underset{\widetilde{K}}{\to} \widetilde{x} \neq \overline{x}$. Sia $\varepsilon = \|\overline{x} - \widetilde{x}\|/2$. Allora $\exists\, \widetilde{k}_\varepsilon \in \mathbb{N} : \|x^k - \widetilde{x}\| < \varepsilon, \forall\, k \geqslant \widetilde{k}_\varepsilon$ con $k \in \widetilde{K}$. Quindi

$$\|x^k - \overline{x}\| \geqslant \|\overline{x} - \widetilde{x}\| - \|x^k - \widetilde{x}\| > 2\varepsilon - \varepsilon = \varepsilon$$

per qualunque $k \geqslant \widetilde{k}_\varepsilon$ tale che $k \in \widetilde{K}$. Questo contraddice $x^k \to \overline{x}$.

(Necessità) Basta considerare $K = \mathbb{N}$ per ottenere $x^k \to \overline{x}$. □

Proposizione 1.3.2. *Sia $\{x^k\} \subseteq X \subseteq \mathbb{R}^n$. Se X è chiuso allora ogni punto di accumulazione di $\{x^k\}$ appartiene a X.*

Dimostrazione. Per assurdo esiste $\widetilde{K}$ tale che $x^k \underset{\widetilde{K}}{\to} \widetilde{x} \notin X$. Sia $\varepsilon = \|\widetilde{x} - P_X(\widetilde{x})\|$, si ha che $\exists\, \widetilde{k}_\varepsilon \in \mathbb{N} : \|x^k - \widetilde{x}\| < \varepsilon, \forall\, k \in \widetilde{K}$ tale che $k \geqslant \widetilde{k}_\varepsilon$. Quindi, per la Proposizione 1.2.2, $\forall\, k \in \widetilde{K}$ tale che $k \geqslant \widetilde{k}_\varepsilon$ si ottiene $x^k \notin X$, che contraddice $\{x^k\} \subseteq X$. □

Proposizione 1.3.3. *Se tutti i punti di accumulazione di ogni $\{x^k\} \subseteq X \subseteq \mathbb{R}^n$ appartengono a X, allora X è chiuso.*

Dimostrazione. Si può definire una successione convergente ad ogni punto di ∂X per mostrare che $\partial X \subseteq X$. □

Lemma 1.3.4. *Sia $\{x^k\} \subseteq X \subseteq \mathbb{R}$. Se X è limitato allora esiste almeno un punto di accumulazione di $\{x^k\}$.*

Dimostrazione. È sempre possibile definire un insieme infinito di indici $\widetilde{K}$ per cui la sottosuccessione $\{x^k\}_{\widetilde{K}}$ risulta essere monotona non decrescente o non crescente. Ipotizziamo senza perdita di generalità che sia monotona non decrescente, ovvero $x^{k_1} \leqslant x^{k_2}$ per ogni $k_1 < k_2$.

L'insieme X è limitato quindi $\sup(X)$ è finito, ed essendo $\{x^k\}_{\widetilde{K}} \subseteq X$, di conseguenza anche $\sup\left(\{x^k\}_{\widetilde{K}}\right)$ è finito.

Per ogni $\varepsilon > 0$ esiste $k_\varepsilon \in \mathbb{N}$ tale che $x^{k_\varepsilon} > \sup\left(\{x^k\}_{\widetilde{K}}\right) - \varepsilon$. Siccome la sottosuccessione è monotona non decrescente, e ricordando la definizione di estremo superiore, otteniamo che $\left\|x^k - \sup\left(\{x^k\}_{\widetilde{K}}\right)\right\| < \varepsilon, \forall\, k \geqslant k_\varepsilon$ tale che $k \in \widetilde{K}$, ovvero $\sup\left(\{x^k\}_{\widetilde{K}}\right)$ è il punto limite di $\{x^k\}_{\widetilde{K}}$. □

Proposizione 1.3.5. *Sia $\{x^k\} \subseteq X \subseteq \mathbb{R}^n$. Se X è limitato allora esiste almeno un punto di accumulazione di $\{x^k\}$.*

Dimostrazione. La successione delle prime coordinate è una successione reale limitata e perciò dal Lemma 1.3.4 essa ammette sottosuccessione convergente. Si ottiene quindi un insieme infinito di indici $\widetilde{K}_1 \subseteq \mathbb{N}$ per cui $x_1^k \underset{\widetilde{K}_1}{\to} \widetilde{x}_1$.

Da questa sottosuccessione possiamo estrarre una sottosottosuccessione (convergente) per la quale anche la seconda coordinata converge. Si ottiene quindi un sottoinsieme infinito di indici $\widetilde{K}_2 \subseteq \widetilde{K}_1$ per cui $x_2^k \underset{\widetilde{K}_2}{\to} \widetilde{x}_2$.

Iterando questo procedimento per tutte le n coordinate si definisce una n volte sottosuccessione (convergente) della successione di partenza per la quale anche l'n-esima coordinata converge. Si ottiene quindi un sottoinsieme infinito di indici $\widetilde{K}_n \subseteq \widetilde{K}_{n-1}$ per cui $x_n^k \underset{\widetilde{K}_n}{\to} \widetilde{x}_n$.

Si è quindi ottenuta una sottosuccessione convergente: $x^k \underset{\widetilde{K}_n}{\to} \widetilde{x}$. □

Se $X \subseteq \mathbb{R}^n$ è illimitato, allora esiste una successione $\{x^k\} \subseteq X$ tale che $\|x^k\| \to +\infty$.

Corollario 1.3.6. *Sia* $\{x^k\} \subseteq X \subseteq \mathbb{R}^n$*. Se* X *è compatto allora esiste almeno un punto di accumulazione di* $\{x^k\}$*, ed esso appartiene a* X*.*

Proposizione 1.3.7. *La successione* $\{x^k\} \subseteq X \subseteq \mathbb{R}^n$ *converge se e solo se è una* ***successione di Cauchy****, ovvero* $\forall\, \varepsilon > 0, \exists\, k_\varepsilon : \|x^{k_1} - x^{k_2}\| < \varepsilon, \forall\, k_1 \geqslant k_\varepsilon,\, k_2 \geqslant k_\varepsilon$*.*

Dimostrazione. (Sufficienza) Per ogni $\varepsilon > 0$, esiste $\overline{k}_{\varepsilon/2} \in \mathbb{N}$ tale che $\|x^k - \overline{x}\| < \varepsilon/2$ per ogni $k \geqslant \overline{k}_{\varepsilon/2}$, dove $\overline{x}$ è il punto limite della successione. Quindi per ogni $k_1 \geqslant \overline{k}_{\varepsilon/2}$ e ogni $k_2 \geqslant \overline{k}_{\varepsilon/2}$ si ha

$$\|x^{k_1} - x^{k_2}\| = \|(x^{k_1} - \overline{x}) + (\overline{x} - x^{k_2})\| \leqslant \|x^{k_1} - \overline{x}\| + \|x^{k_2} - \overline{x}\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Quindi la successione risulta essere di Cauchy.

(Necessità) Per prima cosa si dimostra che X può essere ipotizzato limitato. Per $\varepsilon = 1$ esiste $k_\varepsilon \in \mathbb{N}$ tale che $\|x^{k_1} - x^{k_2}\| < 1, \forall\, k_1 \geqslant k_\varepsilon, k_2 \geqslant k_\varepsilon$. Prendendo $k_2 = k_\varepsilon$ si ottiene $\|x^{k_1} - x^{k_\varepsilon}\| < 1, \forall\, k_1 \geqslant k_\varepsilon$. Essendo vero che $\|x^{k_1}\| \leqslant \|x^{k_1} - x^{k_\varepsilon}\| + \|x^{k_\varepsilon}\|$, si ottiene $\|x^{k_1}\| < 1 + \|x^{k_\varepsilon}\|, \forall\, k_1 \geqslant k_\varepsilon$. Usando questo limite per gli infiniti punti con indici dal k_ε-esimo in poi, e considerando i precedenti in modo esplicito, si ottiene infine un limite per l'intera successione: $\|x^k\| \leqslant \max\{\|x^0\|, \ldots, \|x^{k_\varepsilon - 1}\|, 1 + \|x^{k_\varepsilon}\|\}$, $\forall k \in \mathbb{N}$.

Siccome X è limitato, dalla Proposizione 1.3.5 si conclude che esiste un insieme infinito di indici $\widetilde{K} \subseteq \mathbb{N}$ tale che $x^k \underset{\widetilde{K}}{\to} \widetilde{x}$.

Sia $\varepsilon > 0$. Siccome la successione è di Cauchy, si ottiene che esiste $k_{\varepsilon/2} \in \mathbb{N}$ tale che $\|x^{k_1} - x^{k_2}\| < \varepsilon/2, \forall\, k_1 \geqslant k_{\varepsilon/2}, k_2 \geqslant k_{\varepsilon/2}$. Siccome la sottosuccessione definita da $\widetilde{K}$ converge a $\widetilde{x}$, si ottiene che esiste $\widetilde{k}_{\varepsilon/2} \in \widetilde{K}$ tale che $\widetilde{k}_{\varepsilon/2} \geqslant k_{\varepsilon/2}$ e $\|x^{\widetilde{k}_{\varepsilon/2}} - \widetilde{x}\| < \varepsilon/2$. Utilizzando queste relazioni, si ottiene per ogni $k \geqslant k_{\varepsilon/2}$:

$$\|x^k - \widetilde{x}\| = \|(x^k - x^{\widetilde{k}_{\varepsilon/2}}) + (x^{\widetilde{k}_{\varepsilon/2}} - \widetilde{x})\| \leqslant \|x^k - x^{\widetilde{k}_{\varepsilon/2}}\| + \|x^{\widetilde{k}_{\varepsilon/2}} - \widetilde{x}\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Quindi si può concludere che $x^k \to \widetilde{x}$. □

1.4 Funzioni

Continuità e differenziabilità

Una funzione $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ si dice **continua** se per ogni punto $\overline{x} \in Y$ e ogni successione di punti $\{x^k\} \subseteq Y$ tale che $x^k \to \overline{x}$, si ha

$$|f(\overline{x}) - f(x^k)| \to 0.$$

Ovvero se per ogni punto $\overline{x} \in Y$ si ha che $\forall\, \varepsilon > 0, \exists\, \delta > 0 : \|x - \overline{x}\| < \delta \Rightarrow |f(x) - f(\overline{x})| < \varepsilon$. In questo caso si scrive $f \in C^0$.

Una funzione continua $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ si dice **continuamente differenziabile** se esiste $\nabla f : \mathring{Y} \to \mathbb{R}^n$ continuo tale che per ogni punto $\overline{x} \in \mathring{Y}$ e ogni successione di punti $\{x^k\} \subseteq \mathring{Y}$ tale che $x^k \to \overline{x}$, si ha

$$\frac{|f(\overline{x}) + \nabla f(\overline{x})^T (x^k - \overline{x}) - f(x^k)|}{\|x^k - \overline{x}\|} \to 0.$$

In questo caso si scrive $f \in C^1$ e si dice che ∇f è il gradiente di f.

Una funzione continuamente differenziabile $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ si dice **due volte continuamente differenziabile** se esiste $\nabla^2 f : \mathring{Y} \to \mathbb{R}^{n\times n}$ continua tale che per ogni punto $\overline{x} \in \mathring{Y}$ e ogni successione di punti $\{x^k\} \subseteq \mathring{Y}$ tale che $x^k \to \overline{x}$, si ha

$$\frac{|f(\overline{x}) + \nabla f(\overline{x})^T (x^k - \overline{x}) + \frac{1}{2}(x^k - \overline{x})^T \nabla^2 f(\overline{x})(x^k - \overline{x}) - f(x^k)|}{\|x^k - \overline{x}\|^2} \to 0.$$

In questo caso si scrive $f \in C^2$ e si dice che $\nabla^2 f$ è l'Hessiana di f. La matrice $\nabla^2 f(\overline{x})$ è sempre simmetrica per ogni $\overline{x} \in \mathring{Y}$.

Dati $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, con $f \in C^1$, e $\overline{x} \in \mathring{Y}$, si indica con $\mathcal{L}_{f,\overline{x}} : \mathbb{R}^n \to \mathbb{R}$ l'**approssimazione lineare** di f calcolata in $\overline{x}$:

$$\mathcal{L}_{f,\overline{x}}(y) \triangleq f(\overline{x}) + \nabla f(\overline{x})^T (y - \overline{x}),\ \forall\, y \in \mathbb{R}^n.$$

Proposizione 1.4.1. *Siano $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, con $f \in C^1$, e $\overline{x} \in \mathring{Y}$, valgono le seguenti proprietà:*

(i) $\mathcal{L}_{f,\overline{x}}(\overline{x}) = f(\overline{x})$

(ii) $\nabla \mathcal{L}_{f,\overline{x}}(y) = \nabla f(\overline{x}),\ \forall\, y \in \mathbb{R}^n$

(iii) $\lim_{y \to \overline{x}} \frac{|\mathcal{L}_{f,\overline{x}}(y) - f(y)|}{\|y - \overline{x}\|} = 0$

(iv) $\nabla f(\overline{x})^T (y - \overline{x}) = \lim_{\lambda \downarrow 0} \frac{f(\overline{x} + \lambda(y - \overline{x})) - f(\overline{x})}{\lambda},\ \forall\, y \in \mathbb{R}^n.$

Dimostrazione. Le proprietà *(i)* e *(ii)* valgono per verifica diretta, la proprietà *(iii)* è conseguenza della definizione di gradiente.

La proprietà *(iv)* si ottiene dalla definizione di gradiente osservando che deve essere vero il seguente limite

$$
\begin{aligned}
0 &= \lim_{\lambda \downarrow 0} \frac{f(\overline{x} + \lambda(y - \overline{x})) - f(\overline{x}) - \nabla f(\overline{x})^T (\overline{x} + \lambda(y - \overline{x}) - \overline{x})}{\lambda} \\
&= \lim_{\lambda \downarrow 0} \frac{f(\overline{x} + \lambda(y - \overline{x})) - f(\overline{x}) - \lambda \nabla f(\overline{x})^T (y - \overline{x})}{\lambda} \\
&= \lim_{\lambda \downarrow 0} \frac{f(\overline{x} + \lambda(y - \overline{x})) - f(\overline{x})}{\lambda} - \nabla f(\overline{x})^T (y - \overline{x}).
\end{aligned}
$$

□

Dati $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, con $f \in C^2$, e $\overline{x} \in \mathring{Y}$, si indica con $\mathcal{Q}_{f,\overline{x}} : \mathbb{R}^n \to \mathbb{R}$ l'**approssimazione quadratica** di f calcolata in $\overline{x}$:

$$\mathcal{Q}_{f,\overline{x}}(y) \triangleq f(\overline{x}) + \nabla f(\overline{x})^T (y - \overline{x}) + \frac{1}{2}(y - \overline{x})^T \nabla^2 f(\overline{x})(y - \overline{x}),\ \forall\, y \in \mathbb{R}^n.$$

Proposizione 1.4.2. *Siano* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$*, con* $f \in C^2$*, e* $\overline{x} \in \mathring{Y}$*, valgono le seguenti proprietà:*

(i) $\mathcal{Q}_{f,\overline{x}}(\overline{x}) = f(\overline{x})$

(ii) $\nabla \mathcal{Q}_{f,\overline{x}}(\overline{x}) = \nabla f(\overline{x})$

(iii) $\nabla^2 \mathcal{Q}_{f,\overline{x}}(y) = \nabla^2 f(\overline{x}),\ \forall\, y \in \mathbb{R}^n$

(iv) $\lim_{y \to \overline{x}} \frac{|\mathcal{Q}_{f,\overline{x}}(y) - f(y)|}{\|y - \overline{x}\|^2} = 0$

(v) $\nabla^2 f(\overline{x})^T (y - \overline{x}) = \lim_{\lambda \downarrow 0} \frac{\nabla f(\overline{x} + \lambda(y - \overline{x})) - \nabla f(\overline{x})}{\lambda},\ \forall\, y \in \mathbb{R}^n.$

Dimostrazione. Le proprietà *(i)*, *(ii)* e *(iii)* valgono per verifica diretta essendo

$$\nabla \mathcal{Q}_{f,\overline{x}}(y) = \nabla f(\overline{x}) + \nabla^2 f(\overline{x})(y - \overline{x}),\ \forall\, y \in \mathbb{R}^n.$$

La proprietà *(iv)* è conseguenza della definizione di Hessiana. La proprietà *(v)* si dimostra in parte sfruttando passaggi simili a quelli della Proposizione 1.4.1. In ogni caso questo punto della dimostrazione non fa parte del programma del corso. □

Proposizione 1.4.3. *siano $f : T \subseteq \mathbb{R}^m \to \mathbb{R}$ e $g : Y \subseteq \mathbb{R}^n \to T$ tali che $f, g \in C^1$. La funzione $f \circ g : Y \to \mathbb{R}$ è tale che $f \circ g \in C^1$ e*

$$\nabla f \circ g(x) = \nabla g(x) \nabla f(g(x)).$$

Valori estremali

Siano $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ e $X \subseteq Y$, esistono sempre $\inf(f, X) \in \mathbb{R} \cup \{-\infty, +\infty\}$ e $\sup(f, X) \in \mathbb{R} \cup \{-\infty, +\infty\}$, gli **estremi inferiore e superiore di** f **su** X per cui valgono le seguenti proprietà:

(i) $\inf(f, X) \leqslant f(x) \leqslant \sup(f, X),\ \forall\, x \in X$

(ii) $\nexists\, \overline{m} > \inf(f, X)$ tale che $f(x) \geqslant \overline{m},\ \forall\, x \in X$

(iii) $\nexists\, \overline{M} < \sup(f, X)$ tale che $f(x) \leqslant \overline{M},\ \forall\, x \in X$.

Se $\inf(f, X) = -\infty$ si dice che f è illimitata inferiormente su X. Se $\sup(f, X) = +\infty$ si dice che f è illimitata superiormente su X. Se X non contiene punti (è vuoto) allora $\inf(f, X) = +\infty$ e $\sup(f, X) = -\infty$.

Il seguente risultato è noto come **Teorema di Weierstrass**.

Teorema 1.4.4. *Sia $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, con $f \in C^0$, e sia $X \subseteq Y$ un insieme compatto e non vuoto. Esistono $\check{x}, \hat{x} \in X$ tali che*

$$-\infty < \inf(f, X) = f(\check{x}) \leqslant f(x) \leqslant f(\hat{x}) = \sup(f, X) < \infty,\ \forall\, x \in X.$$

Dimostrazione. Per prima cosa si mostra che $\text{inf}(f, X) > -\infty$. Infatti se per assurdo questo non è vero, allora per ogni $k \in \mathbb{N}$ esiste $x^k \in X$ tale che $f(x^k) < -k$. Per la Proposizione 1.3.5 esiste un insieme infinito di indici $\widetilde{K} \subseteq \mathbb{N}$ tale che $x^k \underset{\widetilde{K}}{\rightarrow} \widetilde{x}$. Per la Proposizione 1.3.2 si ha $\widetilde{x} \in X$. Per la continuità di f si ha che

$$-\infty = \lim_{k\to\infty, k\in\widetilde{K}} -k \geqslant \lim_{k\to\infty, k\in\widetilde{K}} f(x^k) = f(\widetilde{x}),$$

ovvero $f(\widetilde{x}) = -\infty$ che è in contraddizione con $\widetilde{x} \in X \subseteq Y$.

Essendo $\text{inf}(f, X) > -\infty$, per ogni $k \in \mathbb{N}$ esiste $x^k \in X$ tale che $f(x^k) < \text{inf}(f, X) + 1/k$. Per la Proposizione 1.3.5 esiste un insieme infinito di indici $\overline{K} \subseteq \mathbb{N}$ tale che $x^k \underset{\overline{K}}{\rightarrow} \breve{x}$. Per la Proposizione 1.3.2 si ha $\breve{x} \in X$. Per la continuità di f si ha che

$$\text{inf}(f, X) = \lim_{k\to\infty, k\in\overline{K}} \left(\text{inf}(f, X) + \frac{1}{k}\right) \geqslant \lim_{k\to\infty, k\in\overline{K}} f(x^k) = f(\breve{x}) \geqslant \text{inf}(f, X),$$

ovvero $f(\breve{x}) = \text{inf}(f, X)$ e quindi $\breve{x}$ è un punto di minimo di f su X.

Si dimostrano analogamente $\text{sup}(f, X) < \infty$ e l'esistenza di $\widehat{x}$. $\square$

Una funzione $f : \mathbb{R}^n \to \mathbb{R}$ si dice **coerciva** se per ogni successione $\{x^k\}$ tale che $\|x^k\| \to +\infty$ si ha

$$\lim_{k\to\infty} f(x^k) = +\infty.$$

Il seguente risultato è un corollario del Teorema di Weierstrass che può essere usato se la funzione f è continua e coerciva e l'insieme X è chiuso e non vuoto, ma non necessariamente limitato. Quindi la richiesta di limitatezza dell'insieme X viene scambiata con la richiesta di coercività della funzione f.

Corollario 1.4.5. *Sia $f : \mathbb{R}^n \to \mathbb{R}$, con $f \in C^0$ e coerciva, e sia X un insieme chiuso e non vuoto. Esiste $\breve{x} \in X$ tale che*

$$-\infty < \text{inf}(f, X) = f(\breve{x}) \leqslant f(x),\ \forall\, x \in X.$$

Dimostrazione. Per ogni $\overline{x} \in X$ l'insieme $\overline{X} = \{x \in X : f(x) \leqslant f(\overline{x})\}$ è compatto e contiene tutti i possibili minimi di f su X. Infatti se $\overline{X}$ non è chiuso allora per la Proposizione 1.3.3 esiste $\{x^k\} \subseteq \overline{X}$ tale che $x^k \to \widetilde{x} \notin \overline{X}$. Per la chiusura di X dalla Proposizione 1.3.2 si ha $\widetilde{x} \in X$, e quindi necessariamente si ottiene $f(\widetilde{x}) > f(\overline{x})$. Ma questo contraddice la continuità di f, infatti

$$\lim_{k\to\infty} f(x^k) \leqslant f(\overline{x}) < f(\widetilde{x}).$$

Quindi $\overline{X}$ è chiuso. L'insieme $\overline{X}$ risulta anche limitato perché per ogni successione $\{x^k\} \subseteq X$ tale che $\|x^k\| \to +\infty$, dalla coercività di f si ha che esiste un $\overline{k} \in \mathbb{N}$ tale che per ogni $k \geqslant \overline{k}$ si ha $f(x^k) > f(\overline{x})$, e quindi $\{x^k\} \nsubseteq \overline{X}$.

Per concludere la dimostrazione basta quindi usare il Teorema 1.4.4 sostituendo X con l'insieme compatto $\overline{X}$. □

Il seguente risultato è noto come **principio del minimo (o del massimo)**.

Teorema 1.4.6. *Sia* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$*, con* $f \in C^1$*, e sia* $X \subseteq \mathring{Y}$ *un insieme convesso. Siano* $\breve{x} \in X$ *e* $\rho > 0$ *tali che* $f(\breve{x}) \leqslant f(x), \forall x \in X \cap \mathcal{B}(\breve{x}, \rho)$*. Si ha*

$$\nabla f(\breve{x})^T (x - \breve{x}) \geqslant 0, \forall x \in X.$$

Siano $\widehat{x} \in X$ *e* $\rho > 0$ *tali che* $f(\widehat{x}) \geqslant f(x), \forall x \in X \cap \mathcal{B}(\widehat{x}, \rho)$*. Si ha*

$$\nabla f(\widehat{x})^T (x - \widehat{x}) \leqslant 0, \forall x \in X.$$

Dimostrazione. Se per assurdo esiste $\widetilde{x} \in X$ tale che $\nabla f(\breve{x})^T(\widetilde{x}-\breve{x}) < 0$, si ottiene per la Proposizione 1.4.1

$$0 > \nabla f(\breve{x})^T (\widetilde{x} - \breve{x}) = \lim_{\lambda \downarrow 0} \frac{f(\breve{x} + \lambda(\widetilde{x} - \breve{x})) - f(\breve{x})}{\lambda}.$$

Quindi per ogni ε tale che $0 < \varepsilon < -\nabla f(\breve{x})^T(\widetilde{x}-\breve{x})$ esiste $\lambda_\varepsilon > 0$ tale che $(f(\breve{x}+\lambda_\varepsilon(\widetilde{x}-\breve{x}))-f(\breve{x}))/\lambda_\varepsilon < -\varepsilon$, che implica $f(\breve{x}+\lambda_\varepsilon(\widetilde{x}-\breve{x}))-f(\breve{x}) <$

0. Sia senza perdita di generalità $\lambda_\varepsilon \leqslant 1$ e abbastanza piccolo da fare in modo che il punto $(\breve{x}+\lambda_\varepsilon(\widetilde{x}-\breve{x})) \in \mathcal{B}(\breve{x},\rho)$, si ottiene che il punto $(\breve{x}+\lambda_\varepsilon(\widetilde{x}-\breve{x})) \in [\widetilde{x},\breve{x}]$ che, per la convessità di X, implica che $(\breve{x}+\lambda_\varepsilon(\widetilde{x}-\breve{x})) \in X$. Quindi il punto $(\breve{x}+\lambda_\varepsilon(\widetilde{x}-\breve{x}))$ nega l'ottimalità di $\breve{x}$.

Analogamente si dimostra la seconda proprietà con $\widehat{x}$. □

Corollario 1.4.7. *Sia* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, *con* $f \in C^1$, *sia* $X \subseteq \mathring{Y}$ *un insieme convesso, e siano* $\overline{x} \in \mathring{X}$ *e* $\rho > 0$ *tali che* $f(\overline{x}) \leqslant f(x), \forall\, x \in X \cap \mathcal{B}(\overline{x},\rho)$, *oppure* $f(\overline{x}) \geqslant f(x), \forall\, x \in X \cap \mathcal{B}(\overline{x},\rho)$. *Si ha*

$$\nabla f(\overline{x}) = 0.$$

Dimostrazione. In entrambi i casi, per ogni $i = 1,\ldots,n$ esistono $x, y \in X$ e $\lambda > 0$ tali che $(x-\overline{x}) = \lambda e^i$ e $(y-\overline{x}) = -\lambda e^i$, dove $\{e^1,\ldots,e^n\}$ sono i vettori della base canonica. Quindi per ogni $i = 1,\ldots,n$ per il Teorema 1.4.6 si ottiene $\lambda \nabla_{x^i} f(\overline{x}) = \nabla f(\overline{x})^T(x-\overline{x}) \geqslant 0$ e $-\lambda \nabla_{x^i} f(\overline{x}) = \nabla f(\overline{x})^T(y-\overline{x}) \geqslant 0$. Questo implica $\nabla_{x^i} f(\overline{x}) = 0$, essendo $\lambda \neq 0$. □

Teoremi della media

Lemma 1.4.8. *Siano* $f : Y \subseteq \mathbb{R} \to \mathbb{R}$, *con* $f \in C^1$, *e* $\overline{x}, y \in \mathring{Y}$ *con* $[\overline{x}, y] \subseteq \mathring{Y}$. *Esiste* $z \in [\overline{x}, y]$ *tale che*

$$f(\overline{x}) + \nabla f(z)(y-\overline{x}) = f(y).$$

Dimostrazione. Se $\overline{x} = y$ il risultato è banalmente vero, quindi si ipotizza $\overline{x} \neq y$.

Si considera la funzione

$$g(t) = f(t) + \frac{f(y)-f(\overline{x})}{y-\overline{x}}(y-t) - f(y).$$

Si ha che $g : Y \subseteq \mathbb{R} \to \mathbb{R}$, con $g \in C^1$, $g(y) = g(\overline{x}) = 0$, e

$$\nabla g(t) = \nabla f(t) - \frac{f(y)-f(\overline{x})}{y-\overline{x}}.$$

Per il Teorema 1.4.4, essendo $[\overline{x}, y]$ compatto e non vuoto, esistono $\breve{x}, \widehat{x} \in [\overline{x}, y]$ tali che $-\infty < \mathtt{inf}(g, [\overline{x}, y]) = g(\breve{x}) \leqslant g(x) \leqslant g(\widehat{x}) = \mathsf{sup}(g, [\overline{x}, y]) < \infty$, $\forall\, x \in [\overline{x}, y]$.

Ci sono 3 casi:

1. se $\mathtt{inf}(g, [\overline{x}, y]) < 0$, allora $\breve{x} \in (\overline{x}, y)$
2. se $\mathsf{sup}(g, [\overline{x}, y]) > 0$, allora $\widehat{x} \in (\overline{x}, y)$
3. se $\mathtt{inf}(g, [\overline{x}, y]) = \mathsf{sup}(g, [\overline{x}, y]) = 0$, allora si può ipotizzare $\breve{x} \in (\overline{x}, y)$ e $\widehat{x} \in (\overline{x}, y)$.

In ognuno dei 3 casi esiste $z \in (\overline{x}, y)$ che verifica le ipotesi del Corollario 1.4.7, e si ottiene $\nabla g(z) = 0$. E quindi si ha

$$0 = \nabla g(z) = \nabla f(z) - \frac{f(y) - f(\overline{x})}{y - \overline{x}},$$

che implica il risultato cercato. □

Teorema 1.4.9. *Siano* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, *con* $f \in C^1$, *e* $\overline{x}, y \in \mathring{Y}$ *con* $[\overline{x}, y] \subseteq \mathring{Y}$. *Esiste* $z \in [\overline{x}, y]$ *tale che*

$$f(\overline{x}) + \nabla f(z)^T (y - \overline{x}) = f(y).$$

Dimostrazione. Si considera la funzione $g(t) = f(\overline{x} + t(y - \overline{x}))$. Si ha che $g : T \subseteq \mathbb{R} \to \mathbb{R}$, con $[0, 1] \subseteq \mathring{T}$ e $g \in C^1$. Per il Lemma 1.4.8 esiste $\lambda \in [0, 1]$ tale che $g(0) + \nabla g(\lambda)(1 - 0) = g(1)$, ovvero, per la Proposizione 1.4.3,

$$f(y) - f(\overline{x}) = g(1) - g(0) = \nabla g(\lambda) = \nabla f(\overline{x} + \lambda(y - \overline{x}))^T (y - \overline{x}),$$

che implica il risultato cercato essendo $(\overline{x} + \lambda(y - \overline{x})) \in [\overline{x}, y]$. □

Lemma 1.4.10. *Siano* $f : Y \subseteq \mathbb{R} \to \mathbb{R}$, *con* $f \in C^2$, *e* $\overline{x}, y \in \mathring{Y}$ *con* $[\overline{x}, y] \subseteq \mathring{Y}$. *Esiste* $z \in [\overline{x}, y]$ *tale che*

$$f(\overline{x}) + \nabla f(\overline{x})(y - \overline{x}) + \frac{1}{2}\nabla^2 f(z)(y - \overline{x})^2 = f(y).$$

Dimostrazione. Se $\overline{x} = y$ il risultato è banalmente vero, quindi si ipotizza $\overline{x} \neq y$.

Si considera la funzione

$$g(t) = f(t) + \nabla f(t)(y-t) + \frac{f(y) - f(\overline{x}) - \nabla f(\overline{x})(y-\overline{x})}{(y-\overline{x})^2}(y-t)^2 - f(y).$$

Si ha che $g : Y \subseteq \mathbb{R} \to \mathbb{R}$, con $g \in C^2$, $g(y) = g(\overline{x}) = 0$, e

$$\nabla g(t) = \left(\nabla^2 f(t) - 2\frac{f(y) - f(\overline{x}) - \nabla f(\overline{x})(y-\overline{x})}{(y-\overline{x})^2}\right)(y-t).$$

Analogamente alla dimostrazione del Lemma 1.4.8, si ottiene che esiste $z \in (\overline{x}, y)$ tale che $\nabla g(z) = 0$. E quindi si ha

$$0 = \nabla g(z) = \left(\nabla^2 f(z) - 2\frac{f(y) - f(\overline{x}) - \nabla f(\overline{x})(y-\overline{x})}{(y-\overline{x})^2}\right)(y-z),$$

che implica

$$\nabla^2 f(z) = 2\frac{f(y) - f(\overline{x}) - \nabla f(\overline{x})(y-\overline{x})}{(y-\overline{x})^2}$$

essendo $y \neq z$, ovvero il risultato cercato. □

Teorema 1.4.11. *Siano* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, *con* $f \in C^2$, *e* $\overline{x}, y \in \mathring{Y}$ *con* $[\overline{x}, y] \subseteq \mathring{Y}$. *Esiste* $z \in [\overline{x}, y]$ *tale che*

$$f(\overline{x}) + \nabla f(\overline{x})^T(y-\overline{x}) + \frac{1}{2}(y-\overline{x})^T \nabla^2 f(z)(y-\overline{x}) = f(y).$$

Dimostrazione. Si considera la funzione $g(t) = f(\overline{x} + t(y-\overline{x}))$. Si ha che $g : T \subseteq \mathbb{R} \to \mathbb{R}$, con $[0,1] \subseteq \mathring{T}$ e $g \in C^2$. Per il Lemma 1.4.10 esiste $\lambda \in [0,1]$ tale che $g(0) + \nabla g(0)(1-0) + \frac{1}{2}\nabla^2 g(\lambda)(1-0)^2 = g(1)$, ovvero, per la Proposizione 1.4.3,

$$\begin{aligned} f(y) - f(\overline{x}) - \nabla f(\overline{x})^T(y-\overline{x}) &= g(1) - g(0) - \nabla g(0) = \frac{1}{2}\nabla^2 g(\lambda) \\ &= \frac{1}{2}(y-\overline{x})^T \nabla^2 f(\overline{x} + \lambda(y-\overline{x}))^T(y-\overline{x}), \end{aligned}$$

che implica il risultato cercato essendo $(\overline{x} + \lambda(y-\overline{x})) \in [\overline{x}, y]$. □

Convessità

Una funzione $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, con Y convesso, si dice:

- **convessa** se per ogni $\overline{x}, y \in Y$ si ha
$$f(\lambda\overline{x} + (1-\lambda)y) \leqslant \lambda f(\overline{x}) + (1-\lambda)f(y),\ \forall\, \lambda \in [0,1],$$
- **strettamente convessa** se per ogni $\overline{x}, y \in Y$ con $\overline{x} \neq y$ si ha
$$f(\lambda\overline{x} + (1-\lambda)y) < \lambda f(\overline{x}) + (1-\lambda)f(y),\ \forall\, \lambda \in (0,1),$$
- **fortemente convessa** se esiste $\mu > 0$ per cui per ogni $\overline{x}, y \in Y$ si ha
$$f(\lambda\overline{x}+(1-\lambda)y) \leqslant \lambda f(\overline{x})+(1-\lambda)f(y)-\lambda(1-\lambda)\frac{\mu}{2}\|\overline{x}-y\|^2,\ \forall\, \lambda \in [0,1],$$
ovvero se la funzione $g(x) = f(x) - \frac{\mu}{2}x^T x$ è convessa. In questo caso μ si dice modulo di forte convessità.

La forte convessità implica la stretta convessità. La stretta convessità implica la convessità.

Proposizione 1.4.12. *Sia $f : \mathring{Y} \subseteq \mathbb{R}^n \to \mathbb{R}$, con Y convesso e $f \in C^1$. I seguenti fatti sono equivalenti:*

(i) f è convessa,

(ii) per ogni $\overline{x}, y \in \mathring{Y}$ si ha
$$f(y) \geqslant f(\overline{x}) + \nabla f(\overline{x})^T(y-\overline{x}) = \mathcal{L}_{f,\overline{x}}(y),$$

(iii) per ogni $\overline{x}, y \in \mathring{Y}$ si ha
$$(\nabla f(y) - \nabla f(\overline{x}))^T(y-\overline{x}) \geqslant 0,$$

(iv) se $f \in C^2$, per ogni $\overline{x} \in \mathring{Y}$ si ha
$$\nabla^2 f(\overline{x}) \succeq 0,$$
ovvero
$$v^T \nabla^2 f(\overline{x}) v \geqslant 0,\ \forall\, v \in \mathbb{R}^n.$$

Dimostrazione. (*(i)* implica *(ii)*) Per ogni $\lambda \in (0,1]$ si ha

$$f(\overline{x}+\lambda(y-\overline{x})) \leqslant f(\overline{x})+\lambda(f(y)-f(\overline{x}))$$

e quindi

$$f(y)-f(\overline{x}) \geqslant \frac{f(\overline{x}+\lambda(y-\overline{x}))-f(\overline{x})}{\lambda}.$$

Questo implica, per la Proposizione 1.4.1,

$$f(y)-f(\overline{x}) \geqslant \lim_{\lambda \downarrow 0} \frac{f(\overline{x}+\lambda(y-\overline{x}))-f(\overline{x})}{\lambda} = \nabla f(\overline{x})^T(y-\overline{x}).$$

(*(ii)* implica *(i)*) Per ogni $\lambda \in [0,1]$, si indica $z=\lambda\overline{x}+(1-\lambda)y$, si ha

$$f(\overline{x}) \geqslant f(z)+\nabla f(z)^T(\overline{x}-z), \quad f(y) \geqslant f(z)+\nabla f(z)^T(y-z).$$

Moltiplicando la prima per λ e la seconda per $(1-\lambda)$ e sommando si ottiene

$$\lambda f(\overline{x})+(1-\lambda)f(y) \geqslant f(z)+\nabla f(z)^T(\lambda\overline{x}+(1-\lambda)y-z) = f(z) = f(\lambda\overline{x}+(1-\lambda)y).$$

(*(ii)* implica *(iii)*) Si ha

$$f(y) \geqslant f(\overline{x})+\nabla f(\overline{x})^T(y-\overline{x}), \quad f(\overline{x}) \geqslant f(y)+\nabla f(y)^T(\overline{x}-y),$$

e quindi

$$\nabla f(\overline{x})^T(\overline{x}-y) \geqslant f(\overline{x})-f(y), \quad -\nabla f(y)^T(\overline{x}-y) \geqslant -f(\overline{x})+f(y).$$

Sommando le due disuguaglianze si ottiene il risultato.

(*(iii)* implica *(ii)*) Per il Teorema 1.4.9 esiste $\lambda \in [0,1]$ tale che

$$f(y) = f(\overline{x})+\nabla f(\overline{x}+\lambda(y-\overline{x}))^T(y-\overline{x}) \geqslant f(\overline{x})+\nabla f(\overline{x})^T(y-\overline{x}),$$

dove l'ultima disuguaglianza deriva dalla *(iii)* in quanto si ha

$$\nabla f(\overline{x}+\lambda(y-\overline{x}))^T\lambda(y-\overline{x}) \geqslant \nabla f(\overline{x})^T\lambda(y-\overline{x}).$$

(*(iii)* implica *(iv)*) Per ogni $v \in \mathbb{R}^n$ si ha

$$(\nabla f(\overline{x}+tv)-\nabla f(\overline{x}))^T tv \geqslant 0,$$

per ogni $t > 0$ tale che $\overline{x} + tv \in \mathring{Y}$. Quindi si ottiene

$$0 \leqslant \left(\lim_{t \downarrow 0} \frac{\nabla f(\overline{x} + tv) - \nabla f(\overline{x})}{t} \right)^T v = v^T \nabla^2 f(\overline{x}) v,$$

dove nell'ultima relazione si è applicata la Proposizione 1.4.2.

(*(iv)* implica *(ii)*) Utilizzando il Teorema 1.4.11 si ottiene che esiste $z \in [\overline{x}, y]$ tale che

$$f(y) = f(\overline{x}) + \nabla f(\overline{x})^T (y - \overline{x}) + \frac{1}{2}(y - \overline{x})^T \nabla^2 f(z)(y - \overline{x}) \geqslant f(\overline{x}) + \nabla f(\overline{x})^T (y - \overline{x}).$$

□

Applicando gli stessi risultati ottenuti nella Proposizione 1.4.12 sulla funzione $g(x) = f(x) - \frac{\mu}{2} x^T x$, si ottengono analoghe proprietà per la funzione f quando è fortemente convessa con modulo $\mu > 0$.

Proposizione 1.4.13. *Sia* $f : \mathring{Y} \subseteq \mathbb{R}^n \to \mathbb{R}$, *con* Y *convesso e* $f \in C^1$. *I seguenti fatti sono equivalenti:*

(i) f *è fortemente convessa con modulo* $\mu > 0$,

(ii) per ogni $\overline{x}, y \in \mathring{Y}$ *si ha*

$$f(y) \geqslant f(\overline{x}) + \nabla f(\overline{x})^T (y - \overline{x}) + \frac{\mu}{2} \|y - \overline{x}\|^2 = \mathcal{L}_{f,\overline{x}}(y) + \frac{\mu}{2} \|y - \overline{x}\|^2,$$

(iii) per ogni $\overline{x}, y \in \mathring{Y}$ *si ha*

$$(\nabla f(y) - \nabla f(\overline{x}))^T (y - \overline{x}) \geqslant \mu \|y - \overline{x}\|^2,$$

(iv) se $f \in C^2$, *per ogni* $\overline{x} \in \mathring{Y}$ *si ha*

$$(\nabla^2 f(\overline{x}) - \mu I) \succeq 0,$$

ovvero

$$v^T \nabla^2 f(\overline{x}) v \geqslant \mu v^T v, \; \forall\, v \in \mathbb{R}^n.$$

Teorema 1.4.14. *Sia* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ *convessa, con* $f \in C^1$*, e sia* $X \subseteq \mathring{Y}$ *un insieme convesso. Sia* $\breve{x} \in X$ *tale che*

$$\nabla f(\breve{x})^T(x - \breve{x}) \geqslant 0, \, \forall \, x \in X,$$

allora si ha $f(\breve{x}) \leqslant f(x), \, \forall \, x \in X$*.*

Dimostrazione. Dalla Proposizione 1.4.12 si ha

$$f(x) - f(\breve{x}) \geqslant \nabla f(\breve{x})^T(x - \breve{x}) \geqslant 0, \, \forall \, x \in X.$$

$\square$

Teorema 1.4.15. *Sia* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ *strettamente convessa e sia* $X \subseteq Y$ *un insieme convesso. Sia* $\breve{x} \in X$ *tale che* $f(\breve{x}) \leqslant f(x), \, \forall \, x \in X$*. Allora non può esistere* $\widetilde{x} \in X$ *tale che* $\widetilde{x} \neq \breve{x}$ *e* $f(\widetilde{x}) \leqslant f(x), \, \forall \, x \in X$*.*

Dimostrazione. Se per assurdo esiste $\widetilde{x}$, allora necessariamente si ha $f(\breve{x}) = f(\widetilde{x}) = \inf(f, X)$, e per ogni $\lambda \in (0, 1)$ si ha

$$\inf(f, X) = f(\breve{x}) = \lambda f(\breve{x}) + (1-\lambda) f(\breve{x}) = \lambda f(\breve{x}) + (1-\lambda) f(\widetilde{x}) > f(\lambda \breve{x} + (1-\lambda)\widetilde{x}),$$

con $(\lambda \breve{x} + (1-\lambda)\widetilde{x}) \in X$ perché X è convesso. Quindi impossibile. $\square$

Condizioni di Lipschitz

Una funzione differenziabile $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ ha il **gradiente Lipschitz continuo** se esiste $L > 0$ tale che

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|, \, \forall \, x, y \in \mathring{Y}.$$

In questo caso si scrive $f \in C^{1,1}$ e si dice che L è la costante di Lipschitz di ∇f.

Il seguente risultato è noto come **lemma di discesa**.

Proposizione 1.4.16. *Siano* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$*, con* $f \in C^{1,1}$ *con costante* L*, e* $\overline{x}, y \in \mathring{Y}$*, si ha*

$$f(\overline{x}) + \nabla f(\overline{x})^T(y - \overline{x}) + \frac{1}{2} L \|y - \overline{x}\|^2 \geqslant f(y).$$

Dimostrazione. Si considera la funzione

$$g(z) = \frac{1}{2} L z^T z - f(z).$$

Per ogni $z, w \in \mathring{Y}$ si ha

$$\begin{aligned}(\nabla g(z) - \nabla g(w))^T (z - w) &= (Lz - \nabla f(z) - Lw + \nabla f(w))^T (z - w) \\ &= L\|z - w\|^2 - (\nabla f(z) - \nabla f(w))^T (z - w) \\ &\geqslant L\|z - w\|^2 - \|\nabla f(z) - \nabla f(w)\| \|z - w\| \\ &\geqslant L\|z - w\|^2 - L\|z - w\|^2 = 0,\end{aligned}$$

Quindi g è convessa per la Proposizione 1.4.12. Per la Proposizione 1.4.12 si ha

$$\begin{aligned}\frac{1}{2} L y^T y - f(y) = g(y) &\geqslant g(\overline{x}) + \nabla g(\overline{x})^T (y - \overline{x}) \\ &= \frac{1}{2} L \overline{x}^T \overline{x} - f(\overline{x}) + (L\overline{x} - \nabla f(\overline{x}))^T (y - \overline{x}),\end{aligned}$$

che implica

$$\begin{aligned}f(y) &\leqslant f(\overline{x}) + \nabla f(\overline{x})^T (y - \overline{x}) + \frac{1}{2} L y^T y - \frac{1}{2} L \overline{x}^T \overline{x} - L \overline{x}^T (y - \overline{x}) \\ &= f(\overline{x}) + \nabla f(\overline{x})^T (y - \overline{x}) + \frac{1}{2} L \|y - \overline{x}\|^2.\end{aligned}$$

□

Proposizione 1.4.17. *Sia* $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ *con* $f \in C^{1,1}$*, allora* $f \in C^1$*.*

Dimostrazione. Sia L la costante di Lipschitz. Per ogni punto $\overline{x} \in \mathring{Y}$ e ogni $\varepsilon > 0$ si ha che $\|x - \overline{x}\| < \varepsilon / L \Rightarrow \|\nabla f(x) - \nabla f(\overline{x})\| < \varepsilon$. □

Lineari e quadratiche

Una funzione $f : \mathbb{R}^n \to \mathbb{R}$ è **lineare** se ha la seguente struttura:

$$f(x) = c^T x + d = \sum_{i=1}^{n} c_i x_i + d,$$

con $c \in \mathbb{R}^n$ e $d \in \mathbb{R}$.

Proposizione 1.4.18. *La funzione* $f : \mathbb{R}^n \to \mathbb{R}$ *è lineare se e solo se vale la seguente proprietà:*

$$f\left(\sum_{j=1}^{p} \alpha_j x^j\right) = \sum_{j=1}^{p} \alpha_j f(x^j) + \left(1 - \sum_{j=1}^{p} \alpha_j\right) f(0),\ \forall\, x^j \in \mathbb{R}^n,\ \alpha_j \in \mathbb{R},\ j = 1, \ldots, p.$$

Dimostrazione. (Sufficienza) Per ogni $x^j \in \mathbb{R}^n$, $\alpha_j \in \mathbb{R}$, $j = 1, \ldots, p$ si ha

$$\begin{aligned}
f\left(\sum_{j=1}^{p} \alpha_j x^j\right) &= c^T \left(\sum_{j=1}^{p} \alpha_j x^j\right) + d \\
&= \sum_{j=1}^{p} \alpha_j c^T x^j + d \\
&= \sum_{j=1}^{p} \alpha_j c^T x^j + \sum_{j=1}^{p} \alpha_j d - \sum_{j=1}^{p} \alpha_j d + d \\
&= \sum_{j=1}^{p} \alpha_j (c^T x^j + d) + \left(1 - \sum_{j=1}^{p} \alpha_j\right) d \\
&= \sum_{j=1}^{p} \alpha_j f(x^j) + \left(1 - \sum_{j=1}^{p} \alpha_j\right) f(0).
\end{aligned}$$

(Necessità) Per ogni $x \in \mathbb{R}^n$, indicando con $c_i = f(e^i) - f(0)$, $i = 1, \ldots, n$, $d = f(0)$, si ha

$$\begin{aligned}
f(x) &= f\left(\sum_{i=1}^{n} x_i e^i\right) \\
&= \sum_{i=1}^{n} x_i f(e^i) + \left(1 - \sum_{i=1}^{n} x_i\right) f(0) \\
&= \sum_{i=1}^{n} x_i (f(e^i) - f(0)) + f(0) \\
&= \sum_{i=1}^{n} c_i x_i + d = c^T x + d.
\end{aligned}$$

□

Una funzione $f : \mathbb{R}^n \to \mathbb{R}$ è **quadratica** se ha la seguente struttura:

$$f(x) = \frac{1}{2}x^T Qx + c^T x + d = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij}x_i x_j + \sum_{i=1}^{n} c_i x_i + d,$$

con $Q \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$ e $d \in \mathbb{R}$. Una funzione lineare è una funzione quadratica con $Q = 0$.

Proposizione 1.4.19. *Sia $f : \mathbb{R}^n \to \mathbb{R}$ quadratica, allora valgono le seguenti proprietà:*

(i) $\nabla f(x) = Qx + c$, $\forall\, x \in \mathbb{R}^n$

(ii) $\nabla^2 f(x) = Q$, $\forall\, x \in \mathbb{R}^n$

(iii) se $Q \succeq 0$ allora f è convessa

(iv) se $Q \succ 0$, ovvero se per ogni $v \in \mathbb{R}^n$ tale che $v \neq 0$ si ha $v^T Qv > 0$, allora f è fortemente convessa con modulo μ pari all'autovalore minimo di Q

(v) $f \in C^{1,1}$ con costante $L = \|Q\|$, ovvero con L pari all'autovalore di Q di modulo massimo

(vi) $f(y) - f(\overline{x}) = (Q\overline{x} + c)^T (y - \overline{x}) + \frac{1}{2}(y - \overline{x})^T Q(y - \overline{x})$, $\forall\, \overline{x}, y \in \mathbb{R}^n$.

2 Programmazione matematica

La Programmazione Matematica è la disciplina che ha per oggetto lo studio dei problemi di ottimizzazione in cui si vuole minimizzare o massimizzare una funzione $f : Y \subseteq \mathbb{R}^n \to \mathbb{R}$ le cui n variabili rappresentano misure che sono vincolate a soddisfare alcuni requisiti, e quindi che devono appartenere ad un insieme prefissato $X \subseteq \mathring{Y}$:

$$\begin{aligned} &\min f(x) \\ &\quad x \in X. \end{aligned}$$

La funzione f viene chiamata **funzione obiettivo** e l'insieme X **insieme ammissibile** del problema. Ogni $x \in X$ si dice **punto ammissibile** per il problema.

Si parla di **problema inammissibile** quando $X = \varnothing$, cioè se non esistono punti ammissibili per il problema. In questo caso si ha $\inf(f, X) = +\infty$.

Si parla di **problema illimitato** quando comunque scelto un valore $M \in \mathbb{R}$ esiste un punto $x \in X$ tale che $f(x) < M$, cioè se per ogni valore della funzione obiettivo esiste almeno un punto per cui f restituisce un valore migliore. In questo caso si ha $\inf(f, X) = -\infty$.

Si parla di **problema che ammette soluzione** quando esiste $\breve{x} \in X$ tale che

$$f(\breve{x}) \leqslant f(x), \; \forall\, x \in X.$$

In questo caso si ha $\|\breve{x}\| < \infty$ e $-\infty < \inf(f, X) = f(\breve{x}) < +\infty$. Il punto $\breve{x}$ è una **soluzione del problema** e $f(\breve{x}) = \inf(f, X)$ è il **valore ottimo del problema**.

Si parla indifferentemente di problemi di massimo o di minimo in quanto vale il seguente risultato.

Proposizione 2.0.1. *Ogni soluzione del problema* $\min_{x \in X} f(x)$ *è anche soluzione del problema* $\max_{x \in X} (-f(x))$ *e viceversa. Inoltre* *si ha che*

$$\inf(f, X) = -\sup(-f, X), \quad \sup(f, X) = -\inf(-f, X).$$

Dimostrazione. Sia $\check{x} \in X$ tale che

$$f(\check{x}) \leqslant f(x), \forall\, x \in X,$$

ovvero $\inf(f, X) = f(\check{x})$. Allora si ha

$$-f(\check{x}) \geqslant -f(x), \forall\, x \in X,$$

ovvero $\sup(-f, X) = -f(\check{x})$. Quindi $\inf(f, X) = f(\check{x}) = -\sup(-f, X)$.

Viceversa, sia $\hat{x} \in X$ tale che

$$f(\hat{x}) \geqslant f(x), \forall\, x \in X,$$

ovvero $\sup(f, X) = f(\hat{x})$. Allora si ha

$$-f(\hat{x}) \leqslant -f(x), \forall\, x \in X,$$

ovvero $\inf(-f, X) = -f(\hat{x})$. Quindi $\sup(f, X) = f(\hat{x}) = -\inf(-f, X)$.

$\square$

Nei problemi pratici l'insieme ammissibile X viene descritto da un numero finito m di vincoli dati dalle funzioni $g_j : Y \subseteq \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, m$, e dai valori $b_j \in \mathbb{R}$, $j = 1, \ldots, m$:

$$X = \{x \in \mathbb{R}^n \,:\, g_j(x) \leqslant b_j,\, j = 1, \ldots, m\}.$$

Diversi Vincoli (requisiti)

[min $g(x)$ s.t. $x \in X$, $X \subseteq \mathbb{R}^n$]

In questa formulazione dell'insieme X si sono utilizzati solo vincoli di disuguaglianza nella forma di minore o uguale, ma questa notazione risulta comunque generale in quanto include i casi in cui i vincoli sono espressi con vincoli di disuguaglianza nella forma di maggiore o uguale e con vincoli di uguaglianza. Infatti si può sempre trasformare in modo equivalente $g_j(x) \geqslant b_j$ in $-g_j(x) \leqslant -b_j$, e $g_j(x) = b_j$ può essere riscritto nella forma equivalente delle due disuguaglianze $g_j(x) \leqslant b_j$ e $-g_j(x) \leqslant -b_j$.

$g_j(x) = b_j$

$\underbrace{\begin{cases} g_j(x) \leqslant b_j \\ g_j(x) \geqslant b_j \end{cases}}$ 2 disuguaglianze

$-g_j(x) \leqslant -b_j$

Usiamo prevalentemente $\underline{g(x) \leqslant b_j}$, $j=1,\ldots,m$

Dato $g_j(x) \leqslant b_j$, questo si dice:

PROBLEMA DI MINIMO

- **vincolo violato** in $\overline{x}$ se $g_j(\overline{x}) > b_j$
- **vincolo verificato** in $\overline{x}$ se $g_j(\overline{x}) \leqslant b_j$
- **vincolo attivo** in $\overline{x}$ se $g_j(\overline{x}) = b_j$
- **vincolo ridondante** se con la sua eliminazione l'insieme ammissibile X rimane immutato.

Dato un punto $x \in \mathbb{R}^n$, si indica con $A(x) \subseteq \{1, \ldots, m\}$ l'insieme degli indici dei vincoli attivi in x:

$$A(x) \triangleq \{j \in \{1, \ldots, m\} \,:\, g_j(x) = b_j\}\,.$$

Proposizione 2.0.2. *Se per ogni $j = 1, \ldots, m$ si ha che $g_j \in C^0$, allora X è chiuso.*

Dimostrazione. Se per assurdo X non è chiuso, allora per la Proposizione 1.3.3 esiste $\{x^k\} \subseteq X$ tale che $x^k \to \overline{x} \notin X$. Quindi esiste $\overline{j} \in \{1, \ldots, m\}$ tale che $g_{\overline{j}}(\overline{x}) > b_{\overline{j}}$ per ogni $k \in \mathbb{N}$. Quindi

$$\lim_{k \to \infty} g_{\overline{j}}(x^k) \leqslant b_{\overline{j}} < g_{\overline{j}}(\overline{x}),$$

che nega la continuità di $g_{\overline{j}}$. □

Proposizione 2.0.3. *Se per ogni $j = 1, \ldots, m$ si ha che g_j è convessa, allora X è convesso.*

Dimostrazione. Siano $y, w \in X$, quindi per ogni $j = 1, \ldots, m$ si ha $g_j(y) \leqslant b_j$ e $g_j(w) \leqslant b_j$. Si dimostra che per ogni $\lambda \in [0, 1]$ si ha $z = \lambda y + (1 - \lambda)w \in X$. Infatti per ogni $j = 1, \ldots, m$ si ha

$$\begin{aligned} g_j(z) &= g_j(\lambda y + (1 - \lambda)w) \\ &\leqslant \lambda g_j(y) + (1 - \lambda) g_j(w) \\ &\leqslant \lambda b_j + (1 - \lambda) b_j = b_j, \end{aligned}$$

ovvero z verifica tutti gli m vincoli e quindi $z \in X$. Ne consegue che $[y, w] \subseteq X$, ovvero X è convesso. □

2.1 Programmazione lineare

I problemi di Programmazione Lineare (PL) sono caratterizzati dall'essere descritti esclusivamente da funzioni lineari. Quindi sia f che ogni g_j sono funzioni lineari:

$$f(x) = c^T x = \sum_{i=1}^{n} c_i x_i, \quad g_j(x) = (a^j)^T x = \sum_{i=1}^{n} a_i^j x_i,\ j = 1, \ldots, m,$$

con $c \in \mathbb{R}^n$, $a^j \in \mathbb{R}^n$, $j = 1, \ldots, m$.

L'insieme ammissibile di un problema di PL si chiama **poliedro** e si indica con

$$P = \left\{ x \in \mathbb{R}^n \, : \, (a^j)^T x \leqslant b_j,\ j = 1, \ldots, m \right\}.$$

Il poliedro P è un insieme chiuso e convesso. Infatti le funzioni g_j sono continue e convesse (vedi Proposizione 1.4.19) e quindi si possono richiamare le Proposizioni 2.0.2 e 2.0.3.

Teorema 2.1.1. *Un punto v è un vertice di P se e solo se:*

(i) $v \in P$

(ii) esiste $J \subseteq A(v)$ tale che $|J| = n$

(iii) i vettori a^j con $j \in J$ sono linearmente indipendenti.

Dimostrazione. (Sufficienza) Si suppone per assurdo che il numero di vettori linearmente indipendenti a^j corrispondenti ai vincoli attivi in v è minore di n. Allora esiste $d \in \mathbb{R}^n$ tale che $(a^j)^T d = 0, \forall\, j \in A(v)$. Sia $\varepsilon > 0$ tale che $x = v + \varepsilon d$ e $y = v - \varepsilon d$ verificano tutti i vincoli non attivi in v:

$$(a^j)^T x = (a^j)^T (v + \varepsilon d) \leqslant b_j, \quad (a^j)^T y = (a^j)^T (v - \varepsilon d) \leqslant b_j, \quad \forall\, j \notin A(v).$$

Per la definizione di d risulta che

$$\begin{aligned}
&(a^j)^T x = (a^j)^T (v + \varepsilon d) = (a^j)^T v + \varepsilon (a^j)^T d = (a^j)^T v = b_j, \\
&(a^j)^T y = (a^j)^T (v - \varepsilon d) = (a^j)^T v - \varepsilon (a^j)^T d = (a^j)^T v = b_j, \quad \forall\, j \in A(v),
\end{aligned}$$

e quindi $x, y \in P$. Inoltre risulta $\frac{1}{2}x + \frac{1}{2}y = v$, ovvero $v \in (x, y)$, e quindi v non è un vertice di P.

(Necessità) Si suppone per assurdo che v non è un vertice di P e quindi che esistono $x, y \in P$ tali che $v = \lambda x + (1-\lambda)y$ per un certo $\lambda \in (0,1)$. Quindi $x \neq v$. Inoltre per ogni $j \in A(v)$ si ha

$$\begin{aligned} b_j &= (a^j)^T v = (a^j)^T(\lambda x + (1-\lambda)y) \\ &= \lambda(a^j)^T x + (1-\lambda)(a^j)^T y \leqslant \lambda(a^j)^T x + (1-\lambda)b_j, \end{aligned}$$

e quindi $\lambda b_j \leqslant \lambda(a^j)^T x$, che, insieme a $(a^j)^T x \leqslant b_j$, implica che $j \in A(x)$. Ovvero $A(v) \subseteq A(x)$.

Quindi, per ogni $J \subseteq A(v)$ tale che $|J| = n$ si ha

$$\left[(a^j)_J\right]^T v = b_J, \quad \left[(a^j)_J\right]^T x = b_J$$

con $v \neq x$, che implica $\left[(a^j)_J\right]$ singolare, e quindi i vettori a^j non possono essere linearmente indipendenti. □

Corollario 2.1.2. *Il poliedro P ha al più*

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}$$

vertici.

Dimostrazione. Il numero $\binom{m}{n}$ corrisponde ai possibili modi di prendere gli m vincoli che compongono P a gruppi di n in modo da verificare almeno la condizione *(ii)* del Teorema 2.1.1. □

Il seguente risultato è noto come **teorema fondamentale della PL**.

Teorema 2.1.3. *Si considera il problema di PL*

$$\begin{gathered} \min f(x) = c^T x \\ x \in P, \end{gathered}$$

una e una sola delle seguenti affermazioni è vera:

(i) il problema è inammissibile

(ii) il problema è illimitato

(iii) il problema ammette soluzione e se l'insieme V dei vertici di P è non vuoto allora V contiene almeno una soluzione del problema.

Dimostrazione. La *(i)* implica che $\inf(f, P) = +\infty$, la *(ii)* implica che $\inf(f, P) = -\infty$, e la *(iii)* implica che $-\infty < \inf(f, P) < +\infty$. Quindi le tre affermazioni sono autoescludenti. Per dimostrare il teorema basta far vedere che se *(i)* e *(ii)* sono false, allora *(iii)* è vera.

Caso generale. Se P è limitato, allora P è compatto e la *(iii)* è vera per il Teorema 1.4.4. Sia quindi P illimitato, allora per la convessità di P esistono $x \in P$ e $d \in \mathbb{R}^n$ tali che $\|x\| < \infty$, $0 < \|d\| < \infty$ e $(x + \lambda d) \in P$ per ogni $\lambda \geqslant 0$. La *(iii)* non è immediatamente vera solo se

$$\inf(f, P) = \lim_{\lambda \to \infty} f(x + \lambda d) = c^T x + c^T d \lim_{\lambda \to \infty} \lambda,$$

che implica

$$c^T d = \frac{\inf(f, P) - c^T x}{\lim_{\lambda \to \infty} \lambda} = 0.$$

Quindi $f(x) = \inf(f, P)$ e x è soluzione del problema.

Caso $V \neq \varnothing$. Per ogni $x \in P$ esiste $v_x \in V$ tale che $f(v_x) \leqslant f(x)$. Infatti se $x \notin V$ allora esistono $d \in \mathbb{R}^n$ e $\lambda > 0$ tali che $(a^j)^T d = 0, \forall\, j \in A(x)$, $(x + \lambda d) \in P$, $|A(x + \lambda d)| > |A(x)|$, e $f(x + \lambda d) \leqslant f(x)$. Reiterando questo procedimento si arriva a v_x che verifica le ipotesi del Teorema 2.1.1.

Per il Corollario 2.1.2 si ha $|V| < \infty$, quindi esiste $\breve{v} \in V$ tale che $f(\breve{v}) \leqslant f(v), \forall\, v \in V$. E quindi si ottiene

$$f(\breve{v}) \leqslant f(v_x) \leqslant f(x), \forall\, x \in P,$$

ovvero $\breve{v}$ è soluzione del problema. □

Come conseguenza del Teorema 2.1.3, gli algoritmi di soluzione per i problemi di PL possono concentrarsi sulla ricerca di un vertice ottimo di P, invece che cercare tra tutti gli infiniti punti che compongono l'insieme ammissibile P.

Proposizione 2.1.4. *L'insieme delle soluzioni di un problema di PL*

$$\begin{aligned} \min f(x) = c^T x \\ x \in P, \end{aligned}$$

è un poliedro contenuto in P.

Dimostrazione. Se il problema non ha soluzione, allora l'insieme delle soluzioni è vuoto ed è quindi un poliedro contenuto in P.

Sia $\breve{x}$ una soluzione, allora l'insieme delle soluzioni è definito da

$$\begin{aligned} &c^T x \leqslant c^T \breve{x} \\ &x \in P, \end{aligned}$$

che è un poliedro contenuto in P. □

2.2 Programmazione non lineare

Per quanto riguarda l'esistenza di soluzioni nei problemi di Programmazione Non Lineare (PNL) non si può definire qualcosa di simile al Teorema 2.1.3 per la PL. Un esempio di problema di PNL che non è inammissibile o illimitato, ma non ha soluzione è $f(x) = e^x$ e $X = \mathbb{R}$. Tuttavia si possono utilizzare il Teorema 1.4.4 e il Corollario 1.4.5 per ottenere delle condizioni sufficienti per l'esistenza di soluzioni.

Nell'ambito della PNL si può definire un concetto di ottimalità più debole. Un punto $\overline{x} \in X$ è una **soluzione locale** se esiste $\rho > 0$ tale che

$$f(\overline{x}) \leqslant f(x), \forall x \in X \cap \mathcal{B}(\overline{x}, \rho).$$

Condizioni necessarie di ottimalità locale possono essere prese dal Teorema 1.4.6 e dal Corollario 1.4.7.

Un **problema convesso** è un problema di ottimizzazione in cui la funzione obiettivo f è convessa e l'insieme ammissibile X è convesso.

Teorema 2.2.1. *Sia il problema convesso, allora i seguenti fatti sono equivalenti*

(i) $\breve{x} \in X, \quad f(\breve{x}) \leqslant f(x), \forall\, x \in X$

(ii) $\breve{x} \in X, \rho > 0, \quad f(\breve{x}) \leqslant f(x), \forall\, x \in X \cap \mathcal{B}(\breve{x}, \rho)$

(iii) se $f \in C^1$, $\breve{x} \in X, \quad \nabla f(\breve{x})^T(x - \breve{x}) \geqslant 0, \forall\, x \in X$*.*

Dimostrazione. (*(i)* implica *(ii)*) Ovvio.

(*(ii)* implica *(i)*) Si suppone per assurdo che esiste $\widetilde{x} \in X$ tale che $f(\widetilde{x}) < f(\breve{x})$. Per la convessià di X si ha $[\widetilde{x}, \breve{x}] \subseteq X$, e per la convessità di f si ha per ogni $\lambda \in (0, 1]$

$$f(\lambda\widetilde{x} + (1-\lambda)\breve{x}) \leqslant \lambda f(\widetilde{x}) + (1-\lambda) f(\breve{x}) = f(\breve{x}) + \lambda(f(\widetilde{x}) - f(\breve{x})) < f(\breve{x}).$$

Quindi si può definire la successione $\{x^k\} = \{((1/k)\widetilde{x} + (1-(1/k))\breve{x})\} \subseteq X$ tale che $f(x^k) < f(\breve{x})$ e $x^k \to \breve{x}$. Questo nega la *(ii)*.

(*(ii)* implica *(iii)*) Teorema 1.4.6.

(*(iii)* implica *(i)*) Teorema 1.4.14. □

Un **problema strettamente convesso** è un problema di ottimizzazione in cui la funzione obiettivo f è strettamente convessa e l'insieme ammissibile X è convesso.

I problemi di PL sono convessi, ma non strettamente convessi.

Teorema 2.2.2. *Sia il problema strettamente convesso, allora o non ammette soluzione, oppure questa è unica.*

Dimostrazione. Dal Teorema 1.4.15. □

Sia $f \in C^1$, con l'obiettivo di calcolare un punto che verifica le condizioni necessarie di ottimalità descritte nel Teorema 1.4.6, si può fare riferimento al seguente schema algoritmico:

$$x^{k+1} \leftarrow P_X(x^k - \alpha \nabla f(x^k)),$$

con $\alpha > 0$. Questo metodo numerico prende il nome di **algoritmo del gradiente proiettato**.

Teorema 2.2.3. *Sia $f \in C^{1,1}$ con costante L e sia X chiuso, non vuoto e convesso, allora ogni punto di accumulazione $\overline{x}$ della successione $\{x^k\}$ prodotta con l'algoritmo del gradiente proiettato con $\alpha \in (0, 2/L)$ verifica*

$$\overline{x} \in X, \quad \nabla f(\overline{x})^T (x - \overline{x}) \geqslant 0, \, \forall \, x \in X.$$

Dimostrazione. La successione $\{x^k\}$ è ben definita per la Proposizione 1.2.1. Si ipotizza che esiste il punto di accumulazione $\overline{x}$.

La successione $\{x^k\} \subseteq X$, quindi dalla chiusura di X si ha $\overline{x} \in X$, vedi Proposizione 1.3.2.

Si considera un generico $k \in \mathbb{N}$. Usando la *(i)* della Proposizione 1.2.3 e considerando $z = x^k$, si ha

$$(x^{k+1} - x^k + \alpha \nabla f(x^k))^T (x^k - x^{k+1}) \geqslant 0,$$

che implica

$$\nabla f(x^k)^T (x^{k+1} - x^k) \leqslant -\frac{1}{\alpha} \|x^{k+1} - x^k\|^2.$$

Combinandola con la disuguaglianza della Proposizione 1.4.16 si ottiene

$$\begin{aligned} f(x^{k+1}) &\leqslant f(x^k) + \nabla f(x^k)^T (x^{k+1} - x^k) + \frac{L}{2} \|x^{k+1} - x^k\|^2 \\ &\leqslant f(x^k) + \left(\frac{L}{2} - \frac{1}{\alpha} \right) \|x^{k+1} - x^k\|^2. \end{aligned}$$

Siccome $\left(\frac{L}{2} - \frac{1}{\alpha}\right) < 0$ si ha la successione $\{f(x^k)\}$ monotona non crescente. Esistendo il punto di accumulazione $\overline{x}$, dalla continuità di f si ottiene $f(x^k) \to f(\overline{x})$. Che implica $\|x^{k+1} - x^k\| \to 0$. Dalla continuità della funzione $g(x) = P_X(x - \alpha\nabla f(x))$ (l'operatore di proiezione è continuo dalla *(ii)* della Proposizione 1.2.1 e ∇f è continuo per ipotesi) si ottiene che esiste un insieme infinito di indici $\overline{K} \subseteq \mathbb{N}$ tale che

$$0 = \lim_{\substack{k \to \infty \\ \overline{K}}} \|g(x^k) - x^k\| = \|g(\overline{x}) - \overline{x}\|,$$

e quindi $\overline{x} = P_X(\overline{x} - \alpha\nabla f(\overline{x}))$. Usando la *(i)* della Proposizione 1.2.3 si ottiene

$$(\overline{x} - \overline{x} + \alpha\nabla f(\overline{x}))^T(x - \overline{x}) \geqslant 0, \; \forall\, x \in X,$$

che prova il risultato cercato, essendo $\alpha > 0$. □

Teorema 2.2.4. *Sia $f \in C^{1,1}$ con costante L e fortemente convessa con modulo μ, e sia X chiuso, non vuoto e convesso. Se il problema di ottimizzazione (che è strettamente convesso) ha una soluzione $\breve{x}$, la successione $\{x^k\}$ prodotta con l'algoritmo del gradiente proiettato con $\alpha \in (0, 2\mu/L^2)$ converge a $\breve{x}$.*

Dimostrazione. Per ogni $k \in \mathbb{N}$ si ha

$$\begin{aligned}
\|x^{k+1} - \breve{x}\|^2 &= \|P_X(x^k - \alpha\nabla f(x^k)) - P_X(\breve{x} - \alpha\nabla f(\breve{x}))\|^2 \\
&\leqslant \|x^k - \alpha\nabla f(x^k) - \breve{x} + \alpha\nabla f(\breve{x})\|^2 \\
&= \|x^k - \breve{x}\|^2 - 2\alpha(x^k - \breve{x})^T(\nabla f(x^k) - \nabla f(\breve{x})) + \alpha^2\|\nabla f(x^k) - \nabla f(\breve{x})\|^2 \\
&\leqslant (1 - 2\alpha\mu + \alpha^2 L^2)\|x^k - \breve{x}\|^2,
\end{aligned}$$

dove: nella prima uguaglianza $\breve{x} = P_X(\breve{x} - \alpha\nabla f(\breve{x}))$ deriva dalla *(i)* della Proposizione 1.2.3; la prima disuguaglianza è la *(ii)* della Proposizione 1.2.3; l'ultima disuguaglianza è vera per il fatto che $f \in C^{1,1}$ ed è fortemente convessa.

Quindi si ottiene

$$\|x^k - \breve{x}\|^2 \leqslant \tau^k \|x^0 - \breve{x}\|^2,$$

dove $\tau = (1 - 2\alpha\mu + \alpha^2 L^2)$.

Nell'intervallo considerato per α si ha $\tau < 1$, e quindi

$$\lim_{k \to \infty} \|x^k - \breve{x}\|^2 \leqslant \|x^0 - \breve{x}\|^2 \lim_{k \to \infty} \tau^k = 0,$$

che implica il risultato cercato. $\square$

Se si richiede la limitetezza di X nel Teorema 2.2.3 o 2.2.4, si ottiene l'esistenza di soluzioni dal Teorema 1.4.4, e l'esistenza di punti di accumulazione per la successione dal Corollario 1.3.6.