

Capitolo 12

Ottimizzazione vincolata I: vincoli di uguaglianza

Completiamo lo studio dell'ottimizzazione con i problemi di *ottimo vincolato*. Distingueremo due casi, a cui dedicheremo questo e il prossimo capitolo rispettivamente: problemi con *vincoli di uguaglianza* e problemi con *vincoli di disuguaglianza*. Per quanto riguarda i primi introdurremo il metodo del Lagrangiano, che, sfruttando proprietà delle funzioni implicite, fornisce condizioni del primo ordine che, quando il vincolo è compatto, consentiranno di trovare gli estremi assoluti applicando una tecnica simile a quella esposta nel paragrafo 6.5 per problemi in una variabile.

12.1 Funzioni implicite

In questo paragrafo concentreremo la nostra attenzione soprattutto sulle *funzioni di due variabili*; solo alla fine faremo un breve cenno al caso generale di n variabili.

Consideriamo il problema (Pn) introdotto all'inizio del capitolo 11 nel caso di due variabili:

$$\begin{array}{ll} \max & f(x, y) \\ \text{sub} & (x, y) \in X. \end{array}$$

Supporremo ora che l'insieme ammissibile $X \subseteq \mathbb{R}^2$ sia una restrizione del dominio naturale della funzione obiettivo f : è giunto il momento di interpretare X come *vincolo* vero proprio. Il primo tipo di vincolo che affrontiamo richiede un tipo di analisi che può essere pensata come 'complementare' rispetto all'ottimizzazione libera: essa presuppone di trascurare completamente i punti interni e di occuparsi *esclusivamente dei punti di frontiera*. In questa prima analisi prenderemo dunque in considerazione vincoli che sono costituiti da punti di frontiera e non possiedono punti interni.

Prima ancora di provare a immaginare che forma possano assumere tali insiemi, sorge spontanea un'osservazione: insiemi costituiti esclusivamente da punti di frontiera sono necessariamente insiemi chiusi; poiché le nostre funzioni obiettivo f sono sempre continue, è sufficiente che tali insiemi siano limitati affinché siano compatti e quindi, per il Teorema di Weierstrass (Teorema 11.1), garantiscano l'esistenza del massimo e del minimo assoluti. Il tipo di approccio per affrontare questo tipo di problemi vincolati si baserà dunque sulla stessa filosofia della regola 6.1 per problemi in una variabile, nella quale si sfrutta la compattezza del vincolo e le condizioni del primo ordine per la funzione obiettivo. Poiché raramente tali insiemi sono convessi, l'altro approccio a nostra disposizione, quello basato sulla concavità di f , è invece quasi sempre inapplicabile in questo contesto.

‘Insiemi di frontiera’ per eccellenza sono le *curve nel piano* \mathbb{R}^2 , poiché ‘separano’ due superfici distinte in \mathbb{R}^2 , come mostra la figura 8.11(d); la prima tipologia di vincolo che consideriamo è dunque un vincolo “magro e sottile” (senza pancia). La definizione 1.6 asserisce che una curva in \mathbb{R}^2 è definita analiticamente da un’equazione in *forma implicita* del tipo $g(x, y) = b$, dove $b \in \mathbb{R}$ è un parametro.¹ Formalmente, il nostro vincolo è l’insieme

$$V = \{(x, y) \in \mathbb{R}^2 : g(x, y) = b, b \in \mathbb{R}\},$$

che ci permette di riformulare il problema (Pn) nella forma standard

$$\begin{aligned} &\text{ott } f(x, y) \\ &\text{sub } g(x, y) = b, \end{aligned} \tag{12.1}$$

dove ‘ott’ indica che cercheremo tutti i punti *estremi assoluti*, sia di massimo che di minimo.

Con a disposizione gli strumenti introdotti nel capitolo 10, possiamo specificare le caratteristiche della curva $g(x, y) = b$ a partire dalle caratteristiche della funzione di due variabili $g(x, y)$.

12.1.1 Curve di livello

Fissato il parametro b , la curva definita implicitamente dall’equazione $g(x, y) = b$ descrive l’andamento del grafico di g ‘al livello b ’, immaginando cioè di “camminare” sul grafico di g seguendo un percorso orizzontale, senza salire o scendere. La prossima definizione vale in generale per qualsiasi funzione di due variabili f , non soltanto per le funzioni g che individuano vincoli; manteniamo pertanto una notazione generica in termini di funzione f e livello c (‘costante’).

DEFINIZIONE 12.1 Si dicono **curve di livello** di $f : X \rightarrow \mathbb{R}$, $X \subseteq \mathbb{R}^2$, gli insiemi

$$CL = \{(x, y) \in X : f(x, y) = c, c \in \mathbb{R}\}.$$

Più specificamente chiameremo *curva di livello c* l’insieme CL caratterizzato dal livello $c \in \mathbb{R}$.

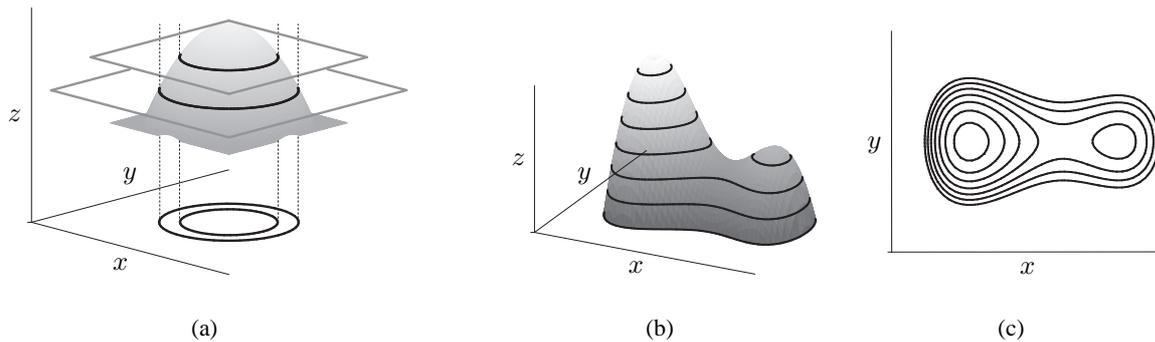


FIGURA 12.1: (a) le curve di livello sono la proiezione sul piano \mathbb{R}^2 delle curve ottenute intersecando il grafico di f con piani orizzontali; (b) altro esempio di f con in (c) le sue curve di livello.

La figura 12.1(a) mostra graficamente la costruzione delle curve di livello: esse sono la proiezione sul piano \mathbb{R}^2 delle curve ottenute intersecando il grafico di f con piani orizzontali a diversi

¹Qui usiamo la notazione ‘ b ’ per il parametro anziché ‘ c ’ come nel paragrafo 1.4.2 dove il parametro era riferito a una generica ‘costante’. Nel contesto dell’ottimizzazione vincolata applicata a problemi economici di solito il parametro b rappresenta il livello complessivo delle risorse (scarse) disponibili, ovvero il ‘budget’ di spesa.

‘livelli’ c . Immaginando che il grafico di una funzione rappresenti una “collina” o una “montagna” su una cartina geografica, possiamo interpretare le curve di livello come ‘isoipse’, le curve che rappresentano una quota di altitudine costante sui lati della collina/montagna stessa. La figura 12.1(b) mostra il grafico di una funzione le cui curve di livello sono riportate in figura 12.1(c).

Dagli esempi 8.6 e 8.7 visti nel paragrafo 8.5.3, intuimmo che le forme quadratiche $f(x, y) = x^2 + y^2$ e $f(x, y) = -x^2 - y^2$ hanno curve di livello che sono circonferenze concentriche con centro l’origine. La figura 12.2 mostra la costruzione delle curve di livello di $f(x, y) = x^2 + y^2$ a partire dal suo grafico come riportato in figura 8.18(a). Si noti che l’equazione in forma implicita $x^2 + y^2 = c$ definisce una circonferenza di raggio \sqrt{c} , che è indiscutibilmente una curva. Poiché la funzione può avere solo valori non negativi, si tratta di circonferenze che hanno senso per valori positivi di c , e che si “allargano” al crescere del livello c ; in altre parole esse rappresentano livelli via via crescenti, $0 < c_1 < c_2 < c_3 < c_4$. Osservando la figura 8.19(a) è facile intuire che per $f(x, y) = -x^2 - y^2$ vale una costruzione analoga in cui le circonferenze rappresentano livelli negativi e si “restringono” al crescere del livello c .

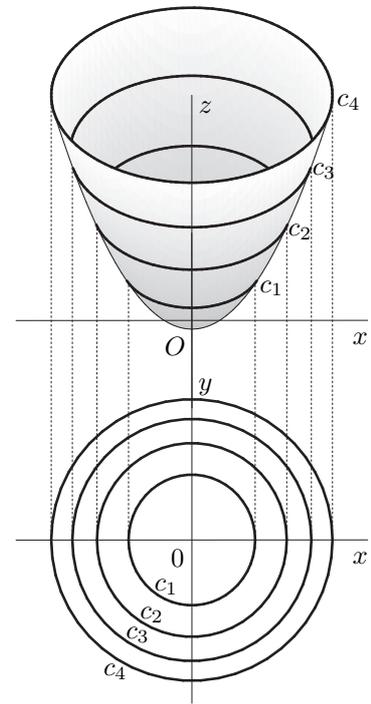


FIGURA 12.2: costruzione delle curve di livello di $f(x, y) = x^2 + y^2$.

OSSERVAZIONE 12.1 **Attenzione!** Secondo la definizione 12.1, le curve che compaiono sui grafici non sono le curve di livello!!! Sono curve di livello la loro proiezione sul piano \mathbb{R}^2 .

Quando è data la direzione di crescita dei diversi livelli c , le curve di livello sul piano \mathbb{R}^2 , pur essendo oggetti in due dimensioni, aiutano a farsi un’idea dell’andamento del grafico di f anche in mancanza di una rappresentazione tridimensionale. La figura 12.3 dovrebbe evidenziare questo aspetto. Quest’osservazione non è affatto secondaria, dietro di essa si cela lo scopo finale del nostro percorso: trasformare oggetti tridimensionali in oggetti a due dimensioni in modo da poterli gestire con gli strumenti che già conosciamo per funzioni di una sola variabile.

ESEMPIO 12.1

1. Ricaviamo le curve di livello di $f(x, y) = xy$, il cui grafico è riportato in figura 12.3(a) esplicitandole come funzioni di una variabile, $y = y(x)$, dalla forma implicita:

$$f(x, y) = xy = c \iff y(x) = c/x,$$

almeno per tutti i punti del dominio di f tali che $x \neq 0$. Deduciamo che le curve di livello sono tutte le iperboli che hanno come asintoti gli assi cartesiani. Dalla figura 12.3(d) si vede che le curve di livello per $c > 0$ sono iperboli nel primo e terzo ortante (l’iperbole equilatera per $c = 1$), mentre per $c < 0$ sono iperboli nel secondo e quarto ortante. Le frecce indicano la ‘direzione di crescita’ delle curve di livello secondo cui f assume livelli c via via più alti.

2. Studiamo le curve di livello di $f(x, y) = \sqrt{x} + \sqrt{y}$ il cui grafico è riportato in figura 12.3(b). Il suo dominio naturale è \mathbb{R}_+^2 e la funzione assume valori non negativi, quindi dovrà essere $c > 0$ (il caso $c = 0$ è ammissibile ma poco interessante). Esplicitiamo le curve di livello come funzioni di una variabile dall’equazione implicita:

$$\sqrt{x} + \sqrt{y} = c \iff \sqrt{y} = c - \sqrt{x} \iff y(x) = (c - \sqrt{x})^2$$

avendo cura di scegliere $c \geq \sqrt{x}$ affinché la seconda equazione abbia senso. La figura 12.3(e) mostra alcune di queste funzioni. La freccia indica la 'direzione di crescita' delle curve di livello secondo cui f assume livelli c via via più alti.

3. Le curve di livello della funzione affine $f(x, y) = ax + by + w$ con almeno uno dei parametri, a o b , diverso da zero, sono delle rette. La figura 12.3(c) riporta il grafico di una funzione affine con $a, b < 0$ e $w > 0$, mentre la figura 12.3(f) mostra alcune curve di livello, dove, come sempre, la freccia indica la 'direzione di crescita' secondo cui f assume livelli c via via più alti. Supponendo $b \neq 0$ esplicitiamo $y = y(x)$ dalla forma implicita:

$$f(x, y) = ax + by + w = c \quad \Leftrightarrow \quad y(x) = -(a/b)x + (c - w)/b,$$

a conferma che si tratta di funzioni affini (lineari se $c = w$) di una variabile.

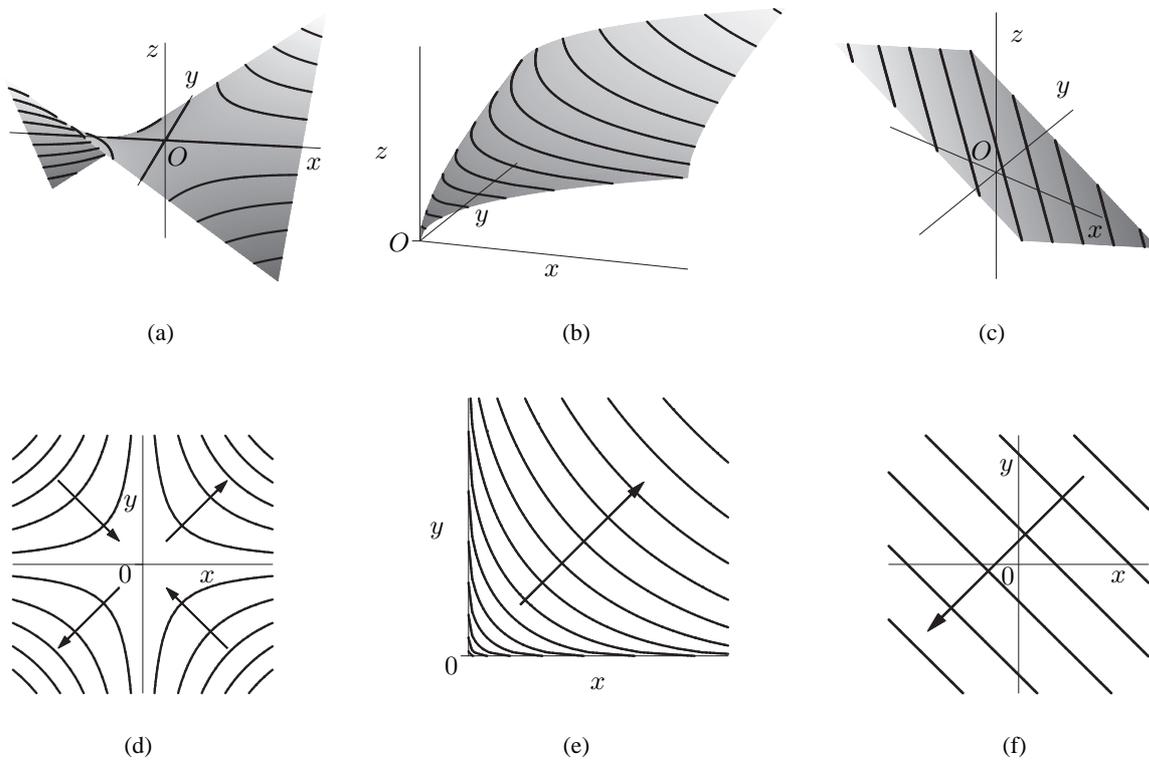


FIGURA 12.3: (a) grafico e (d) curve di livello di $f(x, y) = xy$; (b) grafico e (e) curve di livello di $f(x, y) = \sqrt{x} + \sqrt{y}$; (c) grafico e (f) curve di livello di $f(x, y) = ax + by + w$, con $a, b < 0$ e $w > 0$.

Terminiamo questo paragrafo con una nota dolente. Già ci eravamo accorti che *non tutte le forme implicite* $f(x, y) = c$ sono globalmente esplicitabili come funzioni di una sola variabile. Nell'esempio 1.3 avevamo scoperto che *le circonferenze non sono curve esplicitabili* come funzioni di una variabile. E questo non è trascurabile se pensiamo che le curve di livello delle forme quadratiche definite positive e negative [figure 8.18(a) e 8.19(a)], molto attraenti dal punto di vista dell'ottimizzazione, sono proprio delle circonferenze! Non solo, è evidente che nessuna forma quadratica definita possiede curve di livello esplicitabili, trattandosi di ellissi, anch'esse non esplicitabili globalmente. Non possiamo immaginare di costruire una teoria dell'ottimizzazione vincolata che escluda tali funzioni.

A rincarare la dose, esiste anche la possibilità che le curve di livello siano sì esplicitabili globalmente, ma solamente *in astratto*, essendo comunque impossibile ottenere una forma esplicita *in forma chiusa*.² Un esempio di questo tipo è la curva definita dalla forma implicita $y + \ln y + x = c$; se da un lato è immediato esplicitare $x(y) = c - y - \ln y$, ci rendiamo conto che ‘non siamo capaci’ di esplicitare $y = y(x)$! Eppure, la derivata $x'(y) = -1 - 1/y$ è negativa per tutti i valori di y ammissibili [$\text{Dom}(f) = \{(x, y) \in \mathbb{R}^2 : y > 0\}$], ovvero $x(y)$ è strettamente monotona e quindi invertibile, e il Teorema 1.5 ci assicura che *deve necessariamente esistere anche l'altra funzione esplicita*, $y = y(x)$; in effetti è proprio così, $y = y(x)$ esiste, ma non è esprimibile in forma chiusa.

Fortunatamente, ancora una volta un angelo custode corre in nostro aiuto a dipanare la matassa. L'angelo assume le sembianze delle due considerazioni seguenti.

Abbiamo imparato dall'analisi univariata che, se la funzione obiettivo non è concava, selezionare i punti di ottimo richiede un'analisi *locale*, incentrata sullo studio della funzione obiettivo in un intorno dei punti critici, stazionari ecc.. Pertanto, ciò di cui abbiamo bisogno non è la possibilità di esplicitare globalmente (su tutto il dominio di f) una curva di livello, bensì è sufficiente che ciò sia possibile solamente in un intorno (piccolo) di alcuni punti ‘incriminati’ che stanno sulla curva stessa. Ci rendiamo facilmente conto che, se ci concentriamo su un punto di una circonferenza e consideriamo il suo ingrandimento mediante una lente, ciò che vediamo è solo un ‘pezzetto’ di circonferenza che, in scala ridotta, ‘appare’ come il grafico di una funzione. Se pratichiamo quest'operazione sui punti $(-\sqrt{c}, 0)$ e $(0, \sqrt{c})$ di una circonferenza di raggio \sqrt{c} , cioè sui punti in cui la curva che rappresenta la circonferenza è ‘verticale’ e pertanto è precluso qualsiasi tentativo di ‘estrarre’ una funzione del tipo $y = y(x)$, ci accorgiamo che in un intorno (piccolo) di questi punti la curva è comunque esplicitabile, se non come $y = y(x)$, almeno come $x = x(y)$.

Vedremo che non è necessario esprimere la funzione esplicita in forma chiusa; una volta appurata la sua esistenza, sarà sufficiente conoscerne la derivata.

12.1.2 Il Teorema della funzione implicita

Il prossimo teorema soddisfa entrambe le esigenze di cui sopra: data una curva definita dalla forma implicita $f(x, y) = c$, sotto opportune ipotesi per f in un intorno di un punto (x_0, y_0) appartenente alla curva (ipotesi valide *localmente*, nelle vicinanze del punto stesso), esso stabilisce 1) l'*esistenza* (locale) di almeno una delle funzioni esplicite, $y = y(x)$ e/o $x = x(y)$, anche nei casi in cui non sia possibile esplicitarla/e analiticamente, e 2) fornisce il *valore della sua derivata* nel punto considerato, cioè $y'(x_0)$ e/o $x'(y_0)$, come funzione delle derivate parziali di f .

TEOREMA 12.1 (DELLA FUNZIONE IMPLICITA IN DUE VARIABILI) Dati una funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ e un ‘livello’ $c \in \mathbb{R}$, consideriamo un punto (x_0, y_0) che soddisfi l'equazione $f(x_0, y_0) = c$ [(x_0, y_0) appartiene alla curva definita da $f(x, y) = c$], e supponiamo che f sia *differenziabile* in un intorno di (x_0, y_0) , cioè che esistano le derivate parziali prime f_x e f_y continue in un intorno di (x_0, y_0) .

1. Se $f_y(x_0, y_0) \neq 0$, allora *esiste la funzione esplicita* $y = y(x)$ in un intorno $I_{x_0}^\delta$ di x_0 tale che
 - (a) $f[x, y(x)] = c$ per ogni $x \in I_{x_0}^\delta$
 - (b) $y_0 = y(x_0)$
 - (c) $y(x)$ è derivabile in x_0 con derivata continua, valendo

$$y'(x_0) = -\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}. \tag{12.2}$$

²Scrivere una funzione *in forma chiusa*, o, equivalentemente, *in forma analitica*, significa esprimerla come composizione di funzioni elementari, come, ad esempio, $y(x) = (e^{\sqrt{x} - \ln(1-1/x)}) / \ln(x^{27} + 34x^{1/52})$.

2. Se $f_x(x_0, y_0) \neq 0$, allora esiste la funzione esplicita $x = x(y)$, in un intorno $I_{y_0}^\delta$ di y_0 tale che
- $f[x(y), y] = c$ per ogni $x \in I_{y_0}^\delta$
 - $x_0 = x(y_0)$
 - $x(y)$ è derivabile in y_0 con derivata continua, valendo

$$x'(y_0) = -\frac{f_y(x_0, y_0)}{f_x(x_0, y_0)}. \quad (12.3)$$

Al di là di stabilire l'esistenza (locale) della funzione esplicita, l'importanza operativa del Teorema 12.1 risiede nelle formule (12.2) e (12.3) che forniscono il valore della sua derivata [la pendenza della curva di livello nel punto (x_0, y_0)] in termini di derivate parziali di f .

La parte difficile della dimostrazione di questo teorema è stabilire l'esistenza e la derivabilità (localmente) della funzione esplicita $y = y(x)$ [o $x = x(y)$], parte che ovviamente tralasciamo. Sapendo che la funzione esplicita esiste ed è derivabile, il calcolo della sua derivata, invece, si ricava facilmente utilizzando la regola 10.1. Supponiamo, ad esempio, che valga $f_y(x_0, y_0) \neq 0$, ovvero che esista la funzione esplicita $y = y(x)$ e che sia derivabile in un intorno di x_0 , come stabilisce la prima parte del teorema. La *restrizione della funzione f alla curva di livello* definita dalla funzione implicita $f(x, y) = c$ in un intorno di (x_0, y_0) non è altro che la funzione composta $h(x) = f[x, y(x)]$, cioè una funzione di una sola variabile.³ Poiché stiamo studiando una proprietà della curva di livello, dobbiamo imporre un vincolo ben preciso a $h(x)$: *il livello $h(x) = f[x, y(x)] = c$ deve rimanere costante al variare di x* , il "movimento" indotto dal variare della variabile x a partire dal punto (x_0, y_0) deve farci restare al livello $f(x, y) = c$, dobbiamo cioè rimanere 'in quota', senza salire o scendere. Ma questo non significa altro che la derivata $h'(x_0)$ nel punto x_0 dev'essere nulla. Calcoliamo questa derivata con la regola 10.1 e poniamola uguale a zero:

$$h'(x_0) = f_x[x_0, y(x_0)] + f_y[x_0, y(x_0)]y'(x_0) = 0,$$

dove nel primo addendo vale $x'(x_0) = 1$ perché si tratta della funzione identica $x(t) = t$. Dividendo per $f_y[x_0, y(x_0)] \neq 0$, otteniamo immediatamente la formula (12.2). Una procedura analoga porta alla (12.3) sotto l'ipotesi $f_x(x_0, y_0) \neq 0$.

OSSERVAZIONE 12.2 Le formule (12.2) e (12.3) implicano che quando esistono entrambe le funzioni esplicite, $y = y(x)$ e $x = x(y)$, in (x_0, y_0) *le loro derivate in (x_0, y_0) sono una il reciproco dell'altra*. Questa non è una semplice coincidenza. Il Teorema 1.5, infatti, afferma che le due funzioni esplicite sono una l'inversa dell'altra, e nel paragrafo 1.5 abbiamo imparato che il grafico della funzione inversa si ottiene ruotando di 180° il grafico di f attorno alla bisettrice, a conferma che una funzione e la sua inversa hanno derivate che sono una il reciproco dell'altra.

Applichiamo il Teorema 12.1 al primo punto dell'esempio 12.1 del paragrafo precedente.

ESEMPIO 12.2 La funzione $f(x, y) = xy$ è differenziabile sul suo dominio naturale \mathbb{R}^2 ; vale $f_x = y$ e $f_y = x$ e pertanto, per il Teorema 12.1, se $x \neq 0$, cioè su tutti i punti (x_0, y_0) che non stanno sull'asse delle ordinate, esiste la funzione esplicita $y = y(x)$ con derivata $y'(x_0) = -f_x(x_0, y_0)/f_y(x_0, y_0) = -y_0/x_0$. Ad esempio, il punto $(x_0, y_0) = (1, 1)$ appartiene alla curva di livello $c = 1$, dal momento che $f(1, 1) = 1$, che in $x_0 = 1$ ha pendenza $y'(1) = -1$; il punto $(x_0, y_0) = (1/3, -2)$ appartiene alla curva di livello $c = -2/3$ che in $x_0 = 1/3$ ha pendenza $y'(1/3) = 6$; mentre il punto $(x_0, y_0) = (-1, 0)$ appartiene alla curva di livello $c = 0$ che in $x_0 = -1$ ha pendenza $y'(-1) = 0$. Si noti dalle figure 12.3(a) e 12.3(d) che la curva di livello zero coincide con gli assi cartesiani, pertanto, poiché $x \neq 0$, in

³Si pensi alle funzioni studiate nel paragrafo 10.4.1 con la variabile x al posto di t .

questo caso stiamo esplicitando l'asse delle ascisse, cioè la funzione esplicita $y(x) \equiv 0$ (la costante nulla), che ha evidentemente ovunque derivata nulla.

Analogamente, se $y \neq 0$, cioè su tutti i punti (x_0, y_0) che non stanno sull'asse delle ascisse, esiste la funzione esplicita $x = x(y)$ con derivata $x'(y_0) = -x_0/y_0$, il reciproco di $y'(x_0) = -y_0/x_0$. Ad esempio, il punto $(x_0, y_0) = (2, 1/2)$ appartiene alla curva di livello $c = 1$ che in $y_0 = 1/2$ ha pendenza $x'(1/2) = -4$; mentre il punto $(x_0, y_0) = (0, -3)$ appartiene alla curva di livello $c = 0$ che in $y_0 = -3$ ha pendenza $x'(-3) = 0$. Volendo cercare conferma grafica nella figura 12.3(d), dobbiamo leggere le curve di livello come funzioni $x = x(y)$, ovvero dobbiamo invertire gli assi cartesiani; pertanto, nel punto $(x_0, y_0) = (2, 1/2)$ la funzione $x = x(1/2)$ ha pendenza maggiore di 1, mentre se leggiamo la stessa curva di livello $c = 1$ come funzione $y = y(x)$ nello stesso punto la pendenza sarà minore di 1; analogamente, pendenza nulla della curva di livello zero nel punto $(x_0, y_0) = (0, -3)$ significa che stiamo esplicitando l'asse delle ordinate con la funzione esplicita $x(y) \equiv 0$ (la costante nulla), che ha ovunque derivata nulla se pensata come funzione di y , mentre coincide con una retta verticale (non esplicitabile) se pensata come funzione di x .

OSSERVAZIONE 12.3 I risultati dell'ultimo esempio sono confermati dall'approccio utilizzato nell'esempio 12.1, dove avevamo ricavato direttamente la funzione esplicita $y(x) = c/x$ per i punti $(x, y) \in \mathbb{R}^2$ tali che $x \neq 0$. Osserviamo però un fatto interessante: derivando direttamente $y(x) = c/x$, otteniamo l'espressione $y'(x) = -c/x^2$, piuttosto diversa dall'espressione $y'(x_0) = -y_0/x_0$ fornita dal Teorema 12.1. La spiegazione di questa apparente incongruenza è che i due approcci sono radicalmente distinti: esplicitare $y(x) = c/x$ dalla forma implicita $xy = c$ è una tecnica *globale* (almeno per $x \neq 0$) per ottenere una funzione di una variabile con tanto di *variabile indipendente* x e *variabile dipendente* y . Tutt'altra cosa è l'analisi *locale* del Teorema 12.1: esso stabilisce sì l'esistenza (locale) della funzione implicita e la sua derivabilità, ma nulla ci dice sull'espressione analitica della sua derivata, si limita a fornire il *valore della derivata nel punto* x_0 . Non a caso, quando applichiamo il Teorema 12.1 utilizziamo la *notazione puntuale* $y'(x_0) = -y_0/x_0$ riferita al punto (x_0, y_0) , da non confondersi con una notazione del tipo $y'(x) = -y/x$ che si riferirebbe, erroneamente, a un'ipotetica funzione della *variabile indipendente* x !

Calcoliamo la derivata $y'(x) = -c/x^2$ sui tre punti dell'esempio 12.2: $(x_0, y_0) = (1, 1)$ appartiene alla curva di livello $c = 1$, pertanto $y'(1) = -1/(1)^2 = -1$; $(x_0, y_0) = (1/3, -2)$ appartiene alla curva di livello $c = -2/3$, pertanto $y'(1/3) = -(-2/3)/(1/3)^2 = 6$; infine, $(x_0, y_0) = (-1, 0)$ appartiene alla curva di livello $c = 0$ e quindi $y'(-1) = 0$. Tutti i *valori puntuali* trovati nell'esempio 12.2 sono confermati dall'analisi diretta: il Teorema 12.1 funziona!

ESEMPIO 12.3 Studiamo la curva di livello $c = 1$ di $f(x, y) = x^2 + y^2$ nei quattro punti "delicati" $(1, 0)$, $(0, 1)$, $(-1, 0)$ e $(0, -1)$. Questa forma implicita rappresenta la circonferenza con centro l'origine e raggio 1, curva non esplicitabile globalmente. Vale $f_x = 2x$ e $f_y = 2y$, pertanto, per il Teorema 12.1, le curve di livello della funzione sono localmente esplicitabili rispetto a entrambe le variabili nei punti con coordinate $x_0 \neq 0$ e $y_0 \neq 0$ e su tali punti vale $y'(x_0) = -x_0/y_0$ e $x'(y_0) = -y_0/x_0$. In $(1, 0)$ e $(-1, 0)$ $f_x \neq 0$ (vale 2 e -2 rispettivamente), mentre $f_y = 0$, dunque esiste $x(y)$ con derivata nulla in $y_0 = 0$, $x'(0) = 0$; si tratta dei punti in cui la retta tangente alla curva è verticale, ovvero è orizzontale rispetto alla funzione esplicita $x(y)$, come mostra la figura 12.4. Viceversa, in $(0, 1)$ e $(0, -1)$ $f_x = 0$ e $f_y \neq 0$ (vale 2 e -2 rispettivamente), pertanto esiste $y(x)$ con derivata nulla in $x_0 = 0$, $y'(0) = 0$; si tratta dei punti in cui la retta tangente alla curva è orizzontale, come mostra la figura 12.4.

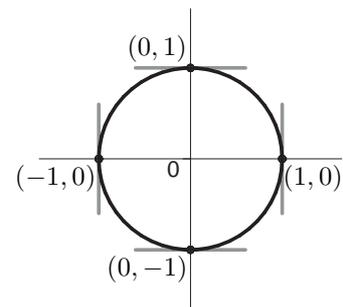


FIGURA 12.4: rette tangenti alla curva di livello $x^2 + y^2 = 1$ nei punti "delicati".

ESERCIZIO 12.1 Verificare che il punto $(x_0, y_0) = (3, 4)$ appartiene alla curva di livello $c = 7$ della funzione $f(x, y) = x^3 - 3xy + y^2$. Stabilire se esiste la funzione esplicita $y = y(x)$ per la curva di

livello $c = 7$ in un intorno di $(x_0, y_0) = (3, 4)$ e, nel caso, calcolarne la pendenza. Ripetere l'esercizio, se possibile, per la funzione esplicita $x = x(y)$.

ESERCIZIO 12.2 Verificare che esistono entrambe le funzioni esplicite, $y = y(x)$ e $x = x(y)$, di tutte le curve di livello per $f(x, y) = y + \ln y + x$. Fissata la coordinata $y_0 = 1$, ricavare la coordinata corrispondente x_0 del punto $(x_0, y_0) = (x_0, 1)$ appartenente alla curva di livello $c = 0$ e calcolare la pendenza di tale curva di livello in $(x_0, 1)$ pensandola sia come $y = y(x)$ che come $x = x(y)$.

12.1.3 Punti regolari e punti singolari

Vediamo ora un esempio patologico di una curva che contiene un punto sul quale essa non è esplicitabile neanche localmente, ovvero un caso in cui non valgono le ipotesi del Teorema 12.1.

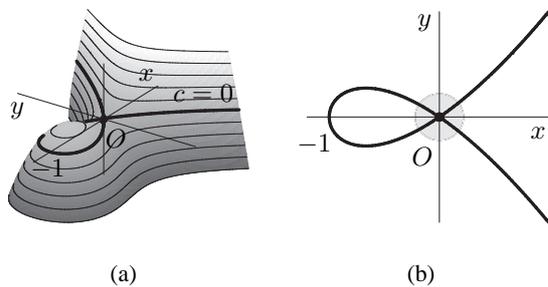


FIGURA 12.5: (a) grafico di $f(x, y) = x^2 + x^3 - y^2$ e (b) la sua curva di livello $c = 0$.

ESEMPIO 12.4 Studiamo la curva di livello $c = 0$ di $f(x, y) = x^2 + x^3 - y^2$. L'origine $(x_0, y_0) = (0, 0)$ appartiene a tale curva. Poiché $f_x = 2x + 3x^2$ e $f_y = -2y$, vale $\nabla f(0, 0) = (0, 0)$, *entrambe le derivate parziali si annullano* nel punto indagato e il Teorema 12.1 non è applicabile su tale punto. Perché $x^2 + x^3 - y^2 = 0$ non è esplicitabile né come $y(x)$ né come $x(y)$ sull'origine? La figura 12.5(a) mostra il grafico di f con evidenziata in grassetto la curva di livello $c = 0$, curva riportata in dettaglio in figura 12.5(b). La peculiarità di questa curva è che nell'origine *interseca se stessa*. *Qualsiasi intorno del punto* $(0, 0)$ (come quello evidenziato in grigio) *contiene in realtà due curve distinte che rappre-*

sentano lo stesso livello $c = 0$; in $(0, 0)$ è dunque impossibile esplicitare come una sola funzione un oggetto che è composto da due funzioni diverse e per questo il Teorema 12.1 fallisce.

L'origine è un *punto stazionario* per la funzione dell'esempio 12.4; è facile verificare che si tratta di un *punto di sella*. Ciò non è una coincidenza: la curva di livello che passa per un punto di sella ha la proprietà di intersecare se stessa in quel punto, come mostra la figura 12.5(a). I casi problematici, comunque, non si limitano ai punti di sella: in generale si tratta di punti sui quali la curva di livello *non è invertibile*. La nostra scarsa propensione al masochismo ci invita a non approfondire oltre.

Poiché il Teorema 12.1 presuppone che almeno una delle derivate parziali di f sia diversa da zero sul punto (x_0, y_0) , la costruzione di casi patologici, di qualsiasi tipo, richiede necessariamente che (x_0, y_0) sia un punto stazionario per f , tale cioè che $\nabla f(x_0, y_0) = (0, 0)$. Questa proprietà permette di circoscrivere l'insieme dei punti "malati" in modo rigoroso.

DEFINIZIONE 12.2 I punti (x_0, y_0) appartenenti alla curva definita dalla forma implicita $f(x, y) = c$, con f differenziabile, tali che $\nabla f(x_0, y_0) \neq (0, 0)$ si dicono **punti regolari**, mentre i punti (x_0, y_0) tali che $\nabla f(x_0, y_0) = (0, 0)$ si chiamano **punti singolari (non regolari)**.

Un punto stazionario per f , affinché sia singolare, deve *stare su una sua curva di livello*. Negli esempi del capitolo 11 abbiamo visto che le funzioni alla nostra portata hanno un numero esiguo di punti stazionari; ci vuole molta sfortuna per incontrare curve di livello passanti proprio per uno di essi! L'esempio 12.4 ammette solamente due punti stazionari, come si vede dalla figura 12.5, di cui solamente l'origine è irregolare per la curva di livello $c = 0$; l'altro, $(-2/3, 0)$, è un punto di massimo relativo e quindi non ci sono curve di livello passanti per quel punto.

ESERCIZIO 12.3 Verificare che il punto $(x_0, y_0) = (-1, 0)$ appartiene alla curva di livello $c = 0$ di $f(x, y) = x^2 + x^3 - y^2$. È applicabile il Teorema 12.1 su tale punto? Se sì, quale delle due funzioni, $y = y(x)$ o $x = x(y)$, è esplicitabile? Qual è la pendenza della curva in quel punto?

12.1.4 Curve di livello e gradiente

Menzioniamo una proprietà geometrica che lega la pendenza di una curva di livello in un punto (regolare) (x_0, y_0) al gradiente di f calcolato nello stesso punto, $\nabla f(x_0, y_0)$. Questa proprietà fornirà la chiave di lettura per interpretare i risultati sull'ottimizzazione vincolata.

TEOREMA 12.2 Si consideri la curva definita dalla forma implicita $f(x, y) = c$, con f differenziabile, e sia (x_0, y_0) un *punto regolare* di tale curva, cioè tale che $\nabla f(x_0, y_0) \neq (0, 0)$. Allora, il *vettore* che rappresenta il *gradiente* $\nabla f(x_0, y_0)$ è *perpendicolare alla curva* $f(x, y) = c$ nel punto (x_0, y_0) .

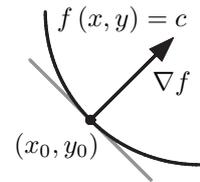


FIGURA 12.6: curva di livello e gradiente.

La figura 12.6 illustra il Teorema 12.2: il *vettore gradiente* $\nabla f(x_0, y_0) = (f_x(x_0, y_0), f_y(x_0, y_0))$, le cui coordinate sono le derivate parziali di f in (x_0, y_0) , è perpendicolare alla retta tangente (in grigio) alla curva definita da $f(x, y) = c$ (in nero); poiché tale retta rappresenta la pendenza della curva in (x_0, y_0) , ha senso dire che ∇f è *perpendicolare alla curva* $f(x, y) = c$ in (x_0, y_0) .

Essendo (x_0, y_0) regolare per la curva $f(x, y) = c$, il Teorema 12.1 vale su tale punto assicurando così l'esistenza della retta tangente. Se $f_y(x_0, y_0) \neq 0$, la pendenza della retta tangente in (x_0, y_0) è $y' = -f_x/f_y$ [omettiamo gli argomenti (x_0, y_0) per semplicità]; pertanto, un vettore rappresentante la *direzione parallela alla retta stessa* è $v = (1, -f_x/f_y)$. Poiché il prodotto scalare $v \cdot \nabla f$ è nullo, $v \cdot \nabla f = (1, -f_x/f_y) \cdot (f_x, f_y) = f_x - f_x = 0$, per il Teorema 8.1 i vettori v e ∇f sono fra loro perpendicolari, e il Teorema 12.2 è dimostrato.

Se (x_0, y_0) non è regolare, la retta tangente alla curva in quel punto non esiste e non ha dunque senso parlare di 'perpendicolarità rispetto alla curva'. Nel punto singolare dell'esempio 12.4 si intersecano due curve di livello distinte: non è possibile determinare la retta tangente (che dev'essere unica). D'altra parte, in un punto singolare $\nabla f = (0, 0)$, e quindi non esiste nemmeno la rappresentazione vettoriale del gradiente (il segmento orientato collassa a un punto).

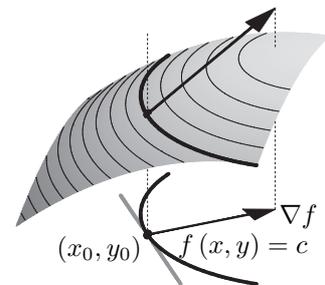


FIGURA 12.7: direzione di massima pendenza e perpendicolarità alla curva di livello c .

La figura 12.7 completa la figura 10.11 con le informazioni del Teorema 12.2: la direzione di massima pendenza di f in uscita da (x_0, y_0) è perpendicolare alla curva di livello passante per (x_0, y_0) . Camminando su di un sentiero in quota lungo il fianco di una montagna, se volessimo arrampicarci direttamente verso la cima, la direzione da prendere sarebbe per forza perpendicolare al sentiero.

12.1.5 Funzioni implicite nel caso di n variabili

Per una funzione di n variabili, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, la forma implicita $f(x) = f(x_1, \dots, x_n) = c$ individua ancora la **curva di livello c** . Ad esempio, la curva di livello 1 di $f(x, y, z) = x^2 + y^2 + z^2$, definita da $x^2 + y^2 + z^2 = 1$, è la sfera di raggio 1 con centro l'origine di \mathbb{R}^3 ; come per l'analoga in due variabili, essa non è globalmente esplicitabile. L'equazione $x + 2y - 3z = 4$ definisce la curva di livello 4 della funzione lineare $f(x, y, z) = x + 2y - 3z$; come in \mathbb{R}^2 , le funzioni lineari

sono globalmente esplicitabili: esplicitando $z = z(x, y) = x/3 + (2/3)y - 4/3$, che è una funzione affine di due variabili, ci rendiamo immediatamente conto che la curva di livello è un piano in \mathbb{R}^3 .

Le funzioni di n variabili, con $n > 3$, hanno curve di livello che chiameremo in astratto **ipersuperfici**, poiché sono oggetti invisibili ai nostri occhi. Si tratta di sottoinsiemi dello spazio \mathbb{R}^n di ‘dimensione’ $n - 1$ che, localmente, si comportano come \mathbb{R}^{n-1} , nel senso che i loro punti possono essere identificati mediante un sistema di $n - 1$ coordinate e su di essi ci si può muovere rispetto a $n - 1$ direzioni indipendenti (si riveda il paragrafo 8.1). Non approfondiamo questi aspetti delicati limitandoci ad osservare che le curve di livello delle funzioni di due variabili, $f(x, y)$, sono ipersuperfici di dimensione $2 - 1 = 1$, ovvero ‘curve’ vere e proprie in \mathbb{R}^2 , e le curve di livello delle funzioni di tre variabili, $f(x, y, z)$, sono ipersuperfici di dimensione $3 - 2 = 1$, ovvero ‘superfici’ ordinarie in \mathbb{R}^3 ; queste ultime si vedono bene quando si può esplicitare una variabile in funzione delle altre due, ad esempio $z = z(x, y)$, essendo il grafico di $z(x, y)$ una superficie in \mathbb{R}^3 .

Quando dalla forma implicita

$$f(x) = f(x_1, \dots, x_n) = c \quad (12.4)$$

si può esplicitare una variabile in funzione delle altre $n - 1$, $x_i = x_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, la (12.4) definisce x_i come **funzione implicita** delle $n - 1$ variabili $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$. Sappiamo che spesso ciò non è possibile globalmente; come nei paragrafi precedenti ci accontentiamo di poterlo fare localmente, in un intorno di un punto x^0 appartenente alla curva definita dalla (12.4). Il prossimo teorema estende il Teorema 12.1 a n variabili.

TEOREMA 12.3 (DELLA FUNZIONE IMPLICITA IN n VARIABILI) Data $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e un ‘livello’ $c \in \mathbb{R}$, consideriamo un punto $x^0 = (x_1^0, \dots, x_n^0)$ che soddisfi l’equazione $f(x^0) = c$ [x^0 appartiene alla curva definita da $f(x) = c$]. Supponiamo inoltre che f sia *differenziabile* in un intorno di x^0 . Se $f_{x_i}(x^0) \neq 0$, allora esiste la funzione esplicita $x_i : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$, $x_i = x_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, in un intorno I di $(x_1^0, \dots, x_{i-1}^0, x_{i+1}^0, \dots, x_n^0)$ tale che

1. $f[x_1, \dots, x_{i-1}, x_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n), x_{i+1}, \dots, x_n] = c$ per ogni $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in I$
2. $x_i^0 = x_i(x_1^0, \dots, x_{i-1}^0, x_{i+1}^0, \dots, x_n^0)$
3. $x_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ è differenziabile in $(x_1^0, \dots, x_{i-1}^0, x_{i+1}^0, \dots, x_n^0)$ e per ogni $j \neq i$

$$\frac{\partial}{\partial x_j} x_i(x_1^0, \dots, x_{i-1}^0, x_{i+1}^0, \dots, x_n^0) = -\frac{f_{x_j}(x^0)}{f_{x_i}(x^0)}.$$

I punti x^0 appartenenti alla curva definita dalla forma implicita $f(x) = c$, con f differenziabile, tali che $\nabla f(x^0) \neq (0, \dots, 0)$ si dicono **punti regolari**, mentre i punti x^0 tali che $\nabla f(x^0) = (0, \dots, 0)$ si chiamano **punti singolari (non regolari)**. In altre parole, i punti singolari sono tutti e soli i punti che sono stazionari per f e stanno su una curva di livello di f . Se tutti i punti di una curva di livello sono regolari $f(x) = c$, si dice che tale curva è un’**ipersuperficie regolare**, o una **varietà $n - 1$ -dimensionale** di \mathbb{R}^n ; ciò significa sostanzialmente che si tratta di un’ipersuperficie che “si comporta bene”: non interseca se stessa, non presenta escrescenze appuntite, ecc. Non approfondiamo le “patologie gravi” delle ipersuperfici.

Infine, poiché i vettori di qualsiasi dimensione n sono rappresentabili geometricamente mediante un segmento orientato, il Teorema 12.2 assume significato anche in spazi \mathbb{R}^n di dimensione $n > 2$.

TEOREMA 12.4 Se x^0 è un punto regolare della curva (ipersuperficie) definita dalla forma implicita $f(x) = c$, con f differenziabile, cioè se $\nabla f(x^0) \neq (0, \dots, 0)$, allora, il vettore che rappresenta il gradiente $\nabla f(x^0)$ è perpendicolare all’ipersuperficie nel punto x^0 .

In questo caso con ‘ $\nabla f(x^0)$ perpendicolare all’ipersuperficie’ si intende $\nabla f(x^0)$ perpendicolare all’iperpiano tangente all’ipersuperficie nel punto x^0 .

ESERCIZIO 12.4 Si consideri $f(x, y, z) = y + \ln y + xyz$.

1. Dopo aver verificato che $(x_0, y_0, z_0) = (0, 1, 0)$ appartiene alla curva di livello $c = 1$, stabilire quali delle funzioni esplicite, $z = z(x, y)$, $y = y(x, z)$ o $x = x(y, z)$, esistono per la curva di livello $c = 1$ in un intorno di $(x_0, y_0, z_0) = (0, 1, 0)$ e, nel caso, calcolarne il gradiente.
2. Dati $x_0 = 1$ e $y_0 = 1$, ricavare la coordinata corrispondente z_0 del punto $(x_0, y_0, z_0) = (1, 1, z_0)$ appartenente alla curva (superficie) di livello $c = 2$ e calcolare il gradiente del piano tangente su tale punto utilizzando tutte le funzioni esplicite possibili.

12.2 Il problema vincolato in due variabili

Affrontiamo il problema vincolato *in due variabili* (12.1) che qui riportiamo:

$$\begin{aligned} &\text{ott } f(x, y) \\ &\text{sub } g(x, y) = b, \end{aligned} \tag{12.5}$$

dove il vincolo V è rappresentato dall’uguaglianza che definisce la curva di livello b di $g(x, y)$,

$$V = \{(x, y) \in \mathbb{R}^2 : g(x, y) = b, b \in \mathbb{R}\}. \tag{12.6}$$

DEFINIZIONE 12.3 Diremo che (x_0, y_0) è un **punto di massimo (minimo) assoluto vincolato stretto (proprio)** per il problema (12.5) se

$$\begin{aligned} f(x, y) &< f(x_0, y_0) \\ &(>) \end{aligned} \tag{12.7}$$

per ogni $(x, y) \in X \cap V$, dove X è il dominio di f e V è il vincolo definito in (12.6). In altre parole, la disuguaglianza (12.7) deve valere solamente per i punti x che stanno contemporaneamente nel dominio di f e sulla curva definita implicitamente dall’equazione $g(x, y) = b$.

DEFINIZIONE 12.4 (x_0, y_0) si dice **punto di massimo (minimo) relativo vincolato stretto (proprio)** per il problema (12.5) se esiste un intorno $I_{(x_0, y_0)}^\delta$ del punto (x_0, y_0) tale che la (12.7) vale per ogni $(x, y) \in I_{(x_0, y_0)}^\delta \cap V$; in questo caso la disuguaglianza (12.7) deve valere solamente per i punti (x, y) che stanno contemporaneamente nell’intorno $I_{(x_0, y_0)}^\delta$ e sulla curva definita da $g(x, y) = b$.

Le definizioni di punto estremo debole si ottengono dalle precedenti sostituendo la disuguaglianza stretta in (12.7) con la disuguaglianza debole.

12.2.1 Interpretazione geometrica dei punti di ottimo vincolato

La figura 12.8 mostra un esempio di vincolo costituito da una curva definita dalla forma implicita $g(x, y) = b$. Il tratto in grassetto della curva è l’insieme $I_{(x_0, y_0)}^\delta \cap V$ usato nella definizione 12.4 per caratterizzare il punto estremo (x_0, y_0) , cioè l’insieme dei punti (x, y) che stanno nell’intorno $I_{(x_0, y_0)}^\delta$ e contemporaneamente soddisfano il vincolo V definito da $g(x, y) = b$.

Il fatto che il vincolo sia formato da una curva suggerisce l’idea chiave su cui si basa la tecnica di risoluzione del problema vincolato (12.5): sfruttando la costruzione di funzione composta sviluppata nel paragrafo 10.4.1 siamo in grado di trasformare il problema in due variabili (12.5) in un problema equivalente in una sola variabile; si tratta di *restringere la funzione obiettivo f alla curva $g(x, y) = b$* in modo da ottenere una funzione di una sola variabile (del tipo illustrato in figura 10.6).

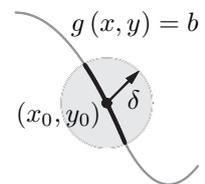


FIGURA 12.8: $I_{(x_0, y_0)}^\delta \cap V$.

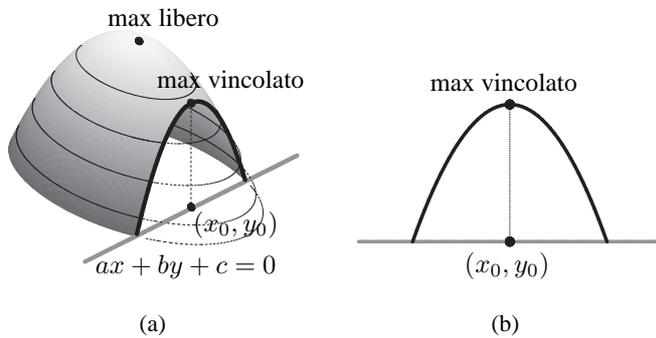


FIGURA 12.9: (a) max vincolato e max libero; (b) dettaglio bidimensionale della curva evidenziata in (a), ovvero riduzione a un problema univariato.

massimo vincolato distinto dal punto di massimo libero. La figura 12.9(b) mostra il dettaglio della restrizione del grafico di f al vincolo, ovvero la ‘riduzione’ del problema di massimo vincolato in due variabili a un problema di massimo in una sola variabile.

Il vincolo lineare in figura 12.9(a) è un caso del tutto particolare in cui è possibile trasformare il problema (12.5) direttamente nel problema in una sola variabile rappresentato in figura 12.9(b), dove è il vincolo ad assumere il ruolo di asse delle ascisse nella rappresentazione grafica bidimensionale [l’intersezione fra il piano verticale individuato dalla retta $ax + by + c = 0$ e il grafico di $f(x, y)$]. Se il vincolo è una curva vera e propria quest’operazione non è possibile; il principio però rimane valido: anche se si tratta di una curva in tre dimensioni (il “serpente” di figura 10.6), nessuno ci vieta di interpretarla come il grafico di una funzione di una sola variabile, anche se in questo caso, diversamente dalla sobrietà di comportamento a cui la variabile indipendente ci aveva abituato muovendosi su \mathbb{R} , la variabile indipendente ha bevuto un po’ e non riesce a mantenere una traiettoria rettilinea...

12.2.2 Metodo per sostituzione

L’idea di trasformare la funzione obiettivo di due variabili in una funzione di una sola variabile è semplicemente geniale. La costruzione della funzione composta di una sola variabile $h(t) = f[x(t), y(t)]$ presuppone però che la curva definita dalle due funzioni $x(t)$ e $y(t)$ sia ‘ben definita’. Nel problema (12.5) la curva in questione è definita implicitamente dall’equazione $g(x, y) = b$, e abbiamo già speso abbastanza parole per avvertire che le funzioni implicite spesso risultano “poco maneggevoli”. In questo breve paragrafo fingeremo di vivere nel migliore dei mondi possibile, quello in cui dalla forma implicita $g(x, y) = b$ è possibile esplicitare ‘globalmente’ una delle due funzioni, $y = y(x)$ e/o $x = x(y)$. Questo al solo fine di illustrare la portata dell’idea di fondo: la trasformazione di (12.5) in un problema equivalente in una variabile sola. Il *metodo per sostituzione* si basa sulla sostituzione del vincolo direttamente nella funzione obiettivo $f(x, y)$, rimpiazzando una delle due variabili x o y con la rispettiva funzione esplicita $x(y)$ o $y(x)$ e ottenendo così una nuova funzione obiettivo (composta) che dipende da una variabile sola.

Sono evidenti i limiti di questo approccio: esso è inapplicabile tutte le volte che il *vincolo* $g(x, y) = b$ non è *globalmente esplicitabile* come funzione di una variabile *in forma chiusa* (o analitica), casi peraltro, come abbiamo visto, tutt’altro che infrequenti. Inoltre, pur consentendo l’effettiva

Prima di studiare questa trasformazione in dettaglio, illustriamo graficamente il concetto di punto di massimo vincolato. In figura 12.9(a) è disegnato il grafico di una funzione di due variabili $f(x, y)$ e, sul piano orizzontale a livello zero, il vincolo rappresentato da una retta (in grigio) definita implicitamente dalla curva di livello $b = 0$ di una funzione affine g , cioè dall’equazione $g(x, y) = ax + by + c = 0$. La restrizione del grafico di f al vincolo è la curva evidenziata in nero che si trova sul grafico di f sopra la retta $ax + by + c = 0$. Dalla figura 12.9(a) si vede che (x_0, y_0) è un *punto di*

risoluzione di un problema vincolato quando ciò è possibile, si tratta di un metodo “grossolano” in quanto non racconta nulla sulle implicazioni economiche del modello sottostante.

Torniamo al nostro mondo ideale e assumiamo che la curva $g(x, y) = b$ rappresentante il vincolo in (12.5) sia globalmente (!) esplicitabile, ovvero supponiamo che esista (in forma chiusa!) la funzione di una variabile $y = y(x)$ [oppure $x = x(y)$]. Allora basta sostituire tale funzione nel secondo (o primo) argomento di $f(x, y)$ e ottenere automaticamente un problema di massimo in una sola variabile con una nuova funzione obiettivo (composta) $h : \mathbb{R} \rightarrow \mathbb{R}$ definita da $h(x) = f[x, y(x)]$ – oppure $\hat{h}(y) = f[x(y), y]$ – vincolata al dominio della funzione $y = y(x)$ [oppure $x = x(y)$]. Se, ad esempio, supponiamo che esista (globalmente) $y = y(x)$, il problema (12.5) viene trasformato nel problema equivalente in una variabile

$$\begin{aligned} \text{ott } h(x) &= f[x, y(x)] \\ \text{sub } x &\in \text{Dom}[y(x)]. \end{aligned}$$

che sappiamo risolvere con gli strumenti dell’analisi univariata.

ESEMPIO 12.5 Risolviamo il seguente problema vincolato:

$$\begin{aligned} \text{ott } (4 - x^2 - y^2) \\ \text{sub } x + y = 2. \end{aligned}$$

La funzione obiettivo è definita su tutto \mathbb{R}^2 , e il vincolo è una retta (globalmente esplicitabile: $y = 2 - x$) che è un insieme non compatto. Quindi il Teorema di Weierstrass (Teorema 11.1) non è applicabile. Con il metodo descritto sopra trasformiamo il problema dato in un problema di massimo in una variabile; poiché il vincolo è una retta, cioè un insieme illimitato, il nuovo problema in una variabile sarà un problema di *massimo libero*, cioè ‘vincolato’ al dominio della retta $y = 2 - x$, che è tutto \mathbb{R} ! Sostituendo $y = 2 - x$ nel secondo argomento della funzione obiettivo otteniamo il nuovo problema libero (non vincolato) nella sola variabile x :

$$\text{ott } [4 - x^2 - (2 - x)^2] = \text{ott } (-2x^2 + 4x).$$

La derivata prima è $h'(x) = -4x + 4$ che si annulla in $x^* = 1$, l’unico punto stazionario per questo problema (non ci sono punti di frontiera). Calcolando la derivata seconda, $h''(x) = -4 < 0$, scopriamo che h è *strettamente concava* e quindi deduciamo che $x^* = 1$ è l’unico punto di massimo (assoluto) per il problema in una variabile e, essendo V non compatto, non c’è nient’altro.

Per trovare la soluzione del problema originale calcoliamo la seconda coordinata y^* del punto di massimo tramite l’equazione del vincolo originale, che dev’essere soddisfatta nel punto di massimo: da $y = 2 - x = 2 - 1 = 1$ ricaviamo l’unico punto di massimo assoluto $(x^*, y^*) = (1, 1)$.

Se il vincolo non è globalmente esplicitabile ci troviamo in un vicolo cieco. Il metodo descritto a partire dal prossimo paragrafo ha portata generale, funziona sotto ipotesi molto deboli – valide in tutte le applicazioni economiche – e costituisce il cuore della teoria economica.

12.2.3 Il Teorema di Lagrange in due variabili

Il propulsore del prossimo risultato è il teorema della funzione implicita. Ricordiamo che il teorema non funziona quando la curva $g(x, y) = b$ contiene punti singolari. Nella sfortunata e rara eventualità di incontrare punti singolari, essi andranno aggiunti ad una lista di *punti critici* per analizzarli separatamente, alla stregua dei punti di frontiera per i problemi univariati (regola 6.1).

Forniremo *condizioni necessarie* e non sufficienti per punti estremi di (12.5). Essendo il vincolo sempre un insieme chiuso di \mathbb{R}^2 , basta che esso sia limitato per applicare un metodo analogo a quello del paragrafo 6.5 per problemi univariati. In questo contesto, condizioni necessarie del primo ordine sono tutto ciò che ci serve per identificare i punti estremi assoluti, non serve nient'altro.

Il calcolo si basa sullo stesso principio illustrato nel paragrafo precedente, però senza l'ambizione di affrontare il problema in modo 'globale', bensì sfruttando il valore numerico della pendenza della curva $g(x, y) = b$ nel punto (x_0, y_0) fornito dalla formula (12.2) [o dalla (12.3)]. In questo senso, seguiremo un approccio *locale*. Poiché stiamo cercando condizioni necessarie, ovvero vogliamo stabilire un risultato analogo al Teorema di Fermat (Teorema 11.2) per l'ottimizzazione libera, il punto (x_0, y_0) in questione sarà un punto estremo (che soddisfa la definizione 12.4).

Se valgono le ipotesi del Teorema 12.1 nel punto (x_0, y_0) , $g(x, y) = b$ è esplicitabile localmente mediante una (o entrambe) delle due funzioni $y = y(x)$ e $x = x(y)$, e tale funzione è derivabile in (x_0, y_0) . Senza perdere generalità, supponiamo che $g_y(x_0, y_0) \neq 0$; allora esiste $y = y(x)$ in un intorno di x_0 e $y'(x_0) = -g_x(x_0, y_0)/g_y(x_0, y_0)$. L'esistenza di $y = y(x)$ ci permette di sostituire tale funzione nel secondo argomento della funzione obiettivo $f(x, y)$ e trasformare (12.5) in un problema univariato con funzione obiettivo (composta) $h: \mathbb{R} \rightarrow \mathbb{R}$ definita da $h(x) = f[x, y(x)]$, esattamente come avevamo fatto nell'esempio 12.5. La differenza fondamentale rispetto a tale esempio è che ora si tratta di un *problema libero* definito in un intorno (che magari è molto piccolo ma sicuramente è un insieme aperto) di x_0 . Inoltre sappiamo che x_0 è la soluzione di tale problema: x_0 è un punto estremo per la funzione di una variabile h , e quindi vale il Teorema di Fermat in una variabile (Teorema 6.2) che stabilisce la condizione del primo ordine $h'(x_0) = 0$. Sotto l'ipotesi che anche la funzione obiettivo f sia differenziabile in un intorno di (x_0, y_0) , possiamo applicare la regola 10.1 alla funzione composta $h(x) = f[x, y(x)]$ nel punto stazionario x_0 :

$$h'(x_0) = f_x[x_0, y(x_0)] + f_y[x_0, y(x_0)]y'(x_0) = 0,$$

che, poiché $y(x_0) = y_0$ (punto 1b del Teorema 12.1), riscriviamo come

$$f_x(x_0, y_0) + f_y(x_0, y_0)y'(x_0) = 0,$$

dalla quale, sostituendo $y'(x_0) = -g_x(x_0, y_0)/g_y(x_0, y_0)$ [formula 12.2], otteniamo

$$f_x(x_0, y_0) - f_y(x_0, y_0) \frac{g_x(x_0, y_0)}{g_y(x_0, y_0)} = 0. \quad (12.8)$$

Se indichiamo con λ il rapporto $f_y(x_0, y_0)/g_y(x_0, y_0)$, se poniamo cioè

$$\lambda = \frac{f_y(x_0, y_0)}{g_y(x_0, y_0)}, \quad (12.9)$$

e sostituiamo in (12.8) otteniamo $f_x(x_0, y_0) = \lambda g_x(x_0, y_0)$, che associata alla (12.9), opportunamente riscritta, porta alla coppia di condizioni del primo ordine per il problema (12.5)

$$\begin{cases} f_x(x_0, y_0) = \lambda g_x(x_0, y_0) \\ f_y(x_0, y_0) = \lambda g_y(x_0, y_0). \end{cases} \quad (12.10)$$

Quasi senza accorgercene, abbiamo appena dimostrato il prossimo teorema. Prima di enunciarlo formalmente, alcune osservazioni sono d'obbligo.

1. Ripercorrendo gli stessi passaggi sotto l'ipotesi $g_x(x_0, y_0) \neq 0$ e ponendo $\lambda = f_x(x_0, y_0) / g_x(x_0, y_0)$ nell'ultimo passaggio, ciò che otteniamo (sorpresa!) è ancora una volta il sistema (12.10). Questo significa che le condizioni (12.10) valgono in tutti i casi in cui è possibile applicare il Teorema 12.1 al vincolo $g(x, y) = b$, a prescindere da quale delle due funzioni, $y = y(x)$ oppure $x = x(y)$, sia esplicitabile; ovvero è sufficiente che esista *soltanto una* delle funzioni esplicite, non sono necessarie entrambe. In particolare, se la curva che esprime il vincolo è verticale in (x_0, y_0) , $y(x)$ non esiste, ma $x(y)$ sì e ha pendenza $x'(y_0) = -g_y(x_0, y_0) / g_x(x_0, y_0) = 0$, da cui segue $g_y(x_0, y_0) = 0$, che, sostituito nella seconda equazione in (12.10), porta a $f_y(x_0, y_0) = 0$ [mentre $f_x(x_0, y_0) = \lambda g_x(x_0, y_0) \neq 0$], situazione del tutto plausibile. In effetti, anche nei punti in cui la curva ha pendenza verticale, il vincolo si "comporta bene" in quanto rimane comunque una curva derivabile.
2. In nessuno dei passaggi svolti abbiamo utilizzato informazioni sulla funzione obiettivo f : tutte le ipotesi fatte riguardano il vincolo $g(x, y) = b$. Per quanto riguarda f , l'unica richiesta è che f sia differenziabile, per poter applicare la regola della catena e calcolare la derivata della funzione composta $h(x) = f[x, y(x)]$ in x_0 . La tentazione di dividere entrambi i termini della (12.8) per $f_y(x_0, y_0)$ era forte, ma ci siamo guardati bene dal farlo perché questo avrebbe richiesto, oltre all'ipotesi $g_y(x_0, y_0) \neq 0$, l'ipotesi aggiuntiva $f_y(x_0, y_0) \neq 0$. Quest'ultima non è necessaria per pervenire alle (12.10), che, come abbiamo visto al punto precedente, funzionano benissimo anche quando $f_y(x_0, y_0) = 0$.
3. Se, oltre a $g_y(x_0, y_0) \neq 0$, ipotizziamo anche $f_y(x_0, y_0) \neq 0$, siamo in grado di comprendere appieno l'intuizione geometrica sottostante al prossimo Teorema 12.5. In questo caso possiamo effettivamente dividere entrambi i termini della (12.8) per $f_y(x_0, y_0)$ ottenendo

$$\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)} = \frac{g_x(x_0, y_0)}{g_y(x_0, y_0)}. \quad (12.11)$$

Se moltiplichiamo entrambi i termini della (12.11) per -1 ci rendiamo conto che la pendenza della curva rappresentante il vincolo nel punto (x_0, y_0) è la stessa della curva di livello di f passante per il punto (x_0, y_0) . Se indichiamo con c il livello definito dal valore assunto da f in (x_0, y_0) , $c = f(x_0, y_0)$, la curva di livello di f passante per (x_0, y_0) è definita dall'equazione $f(x, y) = c$. Riassumendo: se $g_y(x_0, y_0) \neq 0$ e $f_y(x_0, y_0) \neq 0$, le curve $g(x, y) = b$ (il vincolo) e $f(x, y) = c$ hanno la stessa pendenza nel punto (x_0, y_0) , ovvero, sono *parallele* in (x_0, y_0) , ossia sono *tangenti nel punto* (x_0, y_0) .⁴ La (12.11) comprende anche il caso in cui $g_x(x_0, y_0) = 0$, da cui segue $f_x(x_0, y_0) = 0$: entrambe le curve $g(x, y) = b$ e $f(x, y) = c$ sono orizzontali e quindi tangenti in (x_0, y_0) .

Verifichiamo infine che anche quando $g_y(x_0, y_0) = 0$ la condizione di tangenza è rispettata. Poiché vale il Teorema 12.1, in questo caso $g_x(x_0, y_0) \neq 0$ e le condizioni (12.10) valgono ancora; da esse deduciamo che anche $f_y(x_0, y_0) = 0$, ovvero entrambe le curve $g(x, y) = b$ e $f(x, y) = c$ sono verticali e, ancora una volta, tangenti in (x_0, y_0) .⁵

⁴Essendo g e f differenziabili e (x_0, y_0) un punto regolare per entrambe [perché $g_y(x_0, y_0) \neq 0$ e $f_y(x_0, y_0) \neq 0$], si tratta di curve lisce (senza spigoli) che non intersecano se stesse. Ha pertanto senso parlare di 'tangenza'.

⁵Se ripercorriamo i passaggi precedenti con $g_x(x_0, y_0) \neq 0$, sotto l'ipotesi (restrittiva, ma utile ai fini interpretativi) che anche $f_x(x_0, y_0) \neq 0$, perveniamo a

$$\frac{f_y(x_0, y_0)}{f_x(x_0, y_0)} = \frac{g_y(x_0, y_0)}{g_x(x_0, y_0)},$$

che è equivalente alla (12.11) in quanto considera i reciproci dei rapporti.

Concludendo, le condizioni del primo ordine (12.10) per il problema vincolato (12.5) caratterizzano nel modo più generale possibile il punto estremo (x_0, y_0) come **punto di tangenza fra il vincolo $g(x, y) = b$ e la curva di livello $c = f(x_0, y_0)$ della funzione obiettivo f** , includendo i punti di tangenza orizzontale e verticale. Formalizziamo questa proprietà fondamentale.

TEOREMA 12.5 (LAGRANGE I) È dato il problema (12.5) dove entrambe le funzioni f (funzione obiettivo) e g (vincolo) sono differenziabili. Sia (x_0, y_0) un *punto di massimo o minimo* relativo per (12.5) che sia un *punto regolare* per il vincolo $g(x, y) = b$, cioè tale che $\nabla g(x_0, y_0) \neq (0, 0)$. Allora esiste un *moltiplicatore* $\lambda \in \mathbb{R}$ che soddisfa le condizioni (12.10):

$$\begin{cases} f_x(x_0, y_0) = \lambda g_x(x_0, y_0) \\ f_y(x_0, y_0) = \lambda g_y(x_0, y_0) \end{cases} \quad (12.12)$$

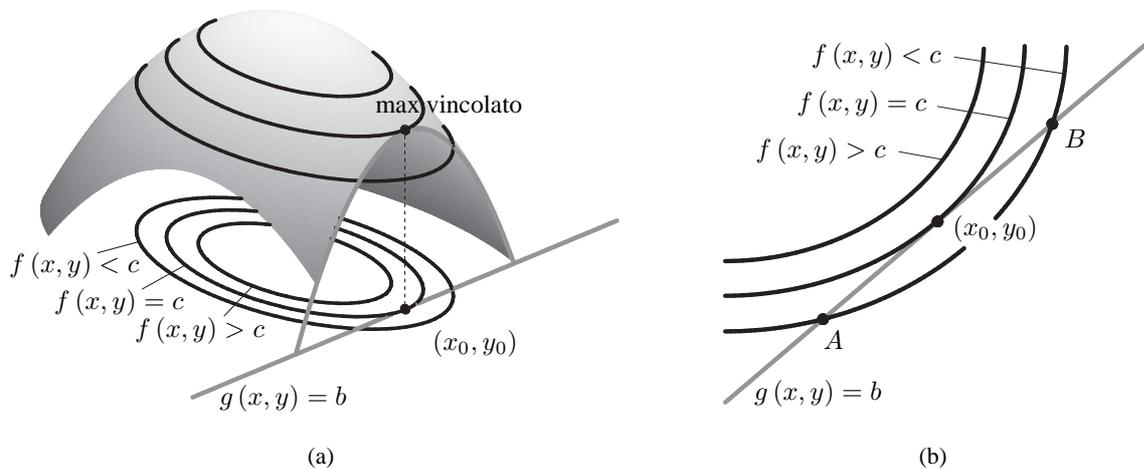


FIGURA 12.10: (a) il punto di max vincolato (x_0, y_0) è il punto di tangenza tra la curva di livello $f(x, y) = c$ e il vincolo $g(x, y) = b$; (b) dettaglio di un'intorno del punto di tangenza (x_0, y_0) .

La figura 12.10 illustra il punto di massimo vincolato (x_0, y_0) di un esempio simile a quello riportato in figura 12.9(a) in cui si evidenzia la condizione (necessaria) di tangenza fra le curve di livello. Nella figura 12.10(a) sono rappresentate (in nero) tre curve di livello della funzione obiettivo f e la loro proiezione 'in quota' sul grafico di f . Quella centrale è la curva definita dal valore raggiunto da f in (x_0, y_0) , il livello $c = f(x_0, y_0)$, ed è tangente al vincolo $g(x, y) = b$ (la retta grigia) nel punto (x_0, y_0) ; le altre due curve rappresentano rispettivamente un livello inferiore e un livello superiore a c e vengono indicate sinteticamente con le disuguaglianze $f(x, y) < c$ e $f(x, y) > c$. La figura 12.10(b) riporta un ingrandimento dell'area intorno al punto di ottimo vincolato (e di tangenza) (x_0, y_0) sul piano \mathbb{R}^2 . In essa si vede chiaramente che la curva di livello superiore, $f(x, y) > c$, indica un livello troppo elevato, incompatibile con il vincolo; nessun punto su tale curva interseca il vincolo, ovvero nessun punto della curva è ammissibile per il problema vincolato (12.5). La curva di livello inferiore $f(x, y) < c$, diversamente, interseca il vincolo in due punti, A e B ; cionondimeno nessuno di questi punti può essere soluzione del problema vincolato (12.5) perché a sinistra (e in alto) rispetto a tale curva esistono altre curve di livello che intersecano anch'esse il vincolo (contengono cioè punti ammissibili per il problema) ma rappresentano valori (livelli) di f maggiori. Esistono altri punti ammissibili per il problema sui quali f assume valori

superiori a quelli in A e B : A e B non soddisfano la definizione 12.4. La figura 12.10 illustra geometricamente la massimizzazione vincolata come *individuazione della curva di livello più alta della funzione obiettivo f che “tocchi” la curva definita dal vincolo $g(x, y) = b$* . Tale curva è necessariamente tangente alla curva $g(x, y) = b$ perché, se la intersecasse, esisterebbero curve riferite a livelli superiori per la funzione obiettivo anch’esse intersecanti il vincolo.

12.2.4 Il moltiplicatore di Lagrange

Il numero ‘ λ ’ che compare nella condizione (12.12) si chiama **moltiplicatore di Lagrange** e, se $g_y(x_0, y_0) \neq 0$, è pari al rapporto $f_y(x_0, y_0) / g_y(x_0, y_0)$, come abbiamo stabilito in (12.9), mentre, se $g_x(x_0, y_0) \neq 0$, è pari al rapporto $f_x(x_0, y_0) / g_x(x_0, y_0)$. Se $g_y(x_0, y_0) \neq 0$ e $g_x(x_0, y_0) \neq 0$ contemporaneamente, entrambi i rapporti fra le corrispondenti derivate parziali di f e di g sono uguali fra loro e il loro valore comune è proprio λ :

$$\frac{f_y(x_0, y_0)}{g_y(x_0, y_0)} = \frac{f_x(x_0, y_0)}{g_x(x_0, y_0)} = \lambda. \tag{12.13}$$

Il moltiplicatore di Lagrange può essere pensato come a un parametro (o variabile, come vedremo tra breve) accessorio introdotto per ottenere la condizione di tangenza più generale possibile: la (12.12).

La condizione (12.12) può essere scritta in *forma vettoriale* come

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0), \tag{12.14}$$

da cui si vede che sul punto di ottimo vincolato, (x_0, y_0) , il gradiente della funzione obiettivo, $\nabla f(x_0, y_0)$, è uguale alla *moltiplicazione scalare* fra il gradiente del vincolo, $\nabla g(x_0, y_0)$, e lo ‘scalare’ λ . Questo significa che i vettori ∇f e ∇g sono *paralleli* (sono proporzionali, giacciono sulla stessa retta) nel punto (x_0, y_0) . Sappiamo dal paragrafo 12.1.4 che il gradiente di una funzione è un vettore perpendicolare alla curva di livello che passa per il punto sul quale il gradiente è calcolato (Teorema 12.2); se ∇f e ∇g sono paralleli nel punto (x_0, y_0) deduciamo che necessariamente anche le curve $f(x, y) = c$ e $g(x, y) = b$ – entrambe perpendicolari a ∇f e ∇g – devono essere parallele (tangenti) in (x_0, y_0) in quel punto, a conferma di quanto già dimostrato.

La figura 12.11 illustra i due casi possibili: in (a) i vettori ∇f e ∇g giacciono sulla stessa retta e ‘puntano’ nella medesima direzione, in (b) essi giacciono ancora sulla stessa retta ma ‘puntano’ in direzioni opposte. In base alla (12.14) deduciamo che i vettori ∇f e ∇g puntano nella medesima direzione se $\lambda > 0$, mentre puntano in direzioni opposte se $\lambda < 0$. Il moltiplicatore, dunque, può assumere valori sia positivi che negativi (in circostanze rarissime può anche essere nullo).

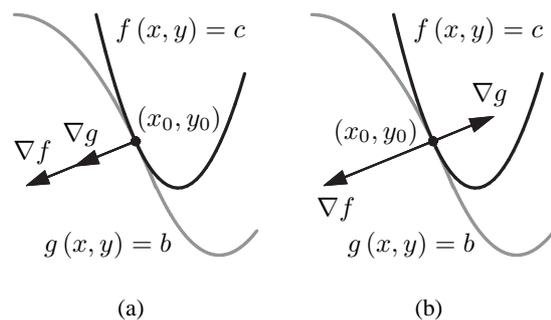


FIGURA 12.11: gradienti paralleli, (a) stesso verso, (b) verso opposto.

12.2.5 Metodo del Lagrangiano per problemi con vincolo compatto

Il Teorema 12.5 stabilisce che un punto di ottimo (x_0, y_0) per il problema vincolato (12.5) deve soddisfare due condizioni: 1) deve appartenere alla curva che definisce il vincolo, cioè deve valere

$g(x_0, y_0) = b$, e 2) la curva di livello $f(x, y) = c$ definita da $c = f(x_0, y_0)$ dev'essere tangente alla curva $g(x, y) = b$ (il vincolo) nel punto (x_0, y_0) . La regola generale che ci apprestiamo a sviluppare permette di verificare entrambe queste condizioni contemporaneamente risolvendo le condizioni del primo ordine di un problema libero (fittizio) in tre variabili.

Definiamo la **funzione Lagrangiana**, $L : \mathbb{R}^3 \rightarrow \mathbb{R}$, come

$$L(x, y, \lambda) = f(x, y) + \lambda [b - g(x, y)], \quad (12.15)$$

nella quale la terza variabile aggiunta è il moltiplicatore di Lagrange. Poiché tutti i punti (x, y) del vincolo sono tali che $b - g(x, y) = 0$, su di essi il secondo addendo in (12.15) è nullo: sul vincolo la funzione Lagrangiana L coincide con la funzione obiettivo f che vogliamo massimizzare (minimizzare). Se (x, y) non appartiene al vincolo possiamo interpretare il prodotto $\lambda [b - g(x, y)] \neq 0$ come una 'sanzione' da pagare per il fatto di "trasgredire le regole" (non rispettare il vincolo). In questo modo si introduce una penalità che è nulla sul vincolo e cresce proporzionalmente alla distanza da esso. Il moltiplicatore λ assume il significato di prezzo (in economia viene chiamato *prezzo ombra*) di questa sanzione: il costo complessivo della "trasgressione" è dato dal prodotto fra sanzione unitaria λ e distanza dal vincolo, misurata, quest'ultima, dalla differenza $b - g(x, y)$.

Se imponiamo le condizioni del primo ordine (Teorema 11.2) alla funzione Lagrangiana (12.15), miracolosamente otteniamo proprio ciò che cercavamo: l'appartenenza al vincolo e le condizioni (12.12) contemporaneamente. Suggelliamo questo risultato nel prossimo teorema.

TEOREMA 12.6 Nelle stesse ipotesi del Teorema 12.5 per il punto di ottimo (x_0, y_0) del problema (12.5), esiste un *moltiplicatore* $\lambda \in \mathbb{R}$ tale che (x_0, y_0, λ) è un *punto stazionario* della funzione Lagrangiana (12.15), cioè tale che

$$\begin{cases} L_x(x_0, y_0, \lambda) = f_x(x_0, y_0) - \lambda g_x(x_0, y_0) = 0 \\ L_y(x_0, y_0, \lambda) = f_y(x_0, y_0) - \lambda g_y(x_0, y_0) = 0 \\ L_\lambda(x_0, y_0, \lambda) = b - g(x_0, y_0) = 0. \end{cases} \quad (12.16)$$

Riconosciamo nelle prime due equazioni in (12.16) la condizione (12.12) del Teorema 12.5, mentre l'ultima equazione stabilisce semplicemente che il vincolo dev'essere rispettato.

Il Teorema 12.6 esprime le condizioni già stabilite dal Teorema 12.5 inequivocabilmente come *condizioni necessarie del primo ordine*; esse, pertanto, si rivelano del tutto inutili per distinguere massimi da minimi. Se però la curva che definisce il vincolo è limitata, trattandosi di un insieme chiuso per definizione essa è un *insieme compatto*, e questo ci permette di utilizzare il Teorema 12.6 con modalità analoghe a quelle impiegate nella regola 6.1 per problemi in una variabile, consentendo l'individuazione dei punti di massimo (minimo) assoluti, ma non gli estremi relativi.

REGOLA 12.1 (DEL LAGRANGIANO)

1. Verificare le ipotesi del Teorema 12.5: f e g devono essere differenziabili e il vincolo non deve contenere punti singolari: le derivate parziali g_x e g_y non devono mai annullarsi contemporaneamente sul vincolo. **Vincoli** che soddisfano questa condizione si dicono **regolari**. Se esistono punti singolari (\hat{x}, \hat{y}) , cioè tali che $\nabla g(\hat{x}, \hat{y}) = (0, 0)$ e $g(\hat{x}, \hat{y}) = b$, aggiungerli alla lista dei punti critici sui quali poi si confronteranno i valori della funzione obiettivo f .
2. Appurare che il *vincolo sia compatto*, che si tratti cioè di una curva limitata.
3. Scrivere la funzione Lagrangiana $L(x, y, \lambda)$ definita in (12.15) e calcolarne i punti stazionari (x^*, y^*, λ^*) risolvendo il sistema (12.16).
4. Calcolare i valori della funzione obiettivo f su tutte le coppie (x^*, y^*) che costituiscono i punti stazionari (x^*, y^*, λ^*) della funzione Lagrangiana ricavati al punto precedente e su eventuali punti non regolari del vincolo. Il *massimo assoluto* è il punto su cui f assume il valore più elevato, mentre il *minimo assoluto* è il punto su cui f assume il valore più basso.

Le condizioni sufficienti del secondo ordine per la ricerca di estremi relativi in problemi vincolati del tipo (12.5) coinvolgono lo studio della matrice Hessiana della funzione Lagrangiana. Poiché, almeno ad un livello elementare, tali condizioni vengono scarsamente utilizzate nelle applicazioni economiche, ne omettiamo la trattazione, rinunciando così allo studio degli estremi relativi.

Applichiamo la regola 12.1 ad alcuni problemi con vincolo compatto.

ESEMPIO 12.6 Risolviamo il seguente problema vincolato:

$$\begin{aligned} & \text{ott } (4 - x) \\ & \text{sub } x^2 + y^2 = 1. \end{aligned}$$

Entrambe le funzioni $f(x, y) = 4 - x$ e $g(x, y) = x^2 + y^2$ sono differenziabili. La *funzione obiettivo* ha come grafico un piano inclinato parallelo all'asse y (è una funzione costante rispetto a y) e con pendenza -1 rispetto alla direzione individuata dall'asse delle ascisse. Vale $f_x = -1$ e $f_y = 0$.

Il *vincolo* è la circonferenza di raggio 1 con centro l'origine, quindi è *compatto* e, poiché la funzione obiettivo è continua, esistono il massimo e minimo assoluti (Teorema 11.1). Inoltre il vincolo è *regolare* perché $\nabla g(x, y) = (2x, 2y) \neq (0, 0)$ per ogni punto (x, y) che sta sulla circonferenza.

Scriviamo la funzione Lagrangiana:

$$L(x, y, \lambda) = 4 - x + \lambda(1 - x^2 - y^2)$$

e calcoliamo i punti stazionari risolvendo il seguente sistema nelle tre variabili x, y e λ :

$$\begin{cases} L_x = -1 - 2\lambda x = 0 \\ L_y = -2\lambda y = 0 \\ L_\lambda = 1 - x^2 - y^2 = 0. \end{cases}$$

La seconda equazione è sempre soddisfatta se $\lambda = 0$ oppure $y = 0$; $\lambda = 0$ è impossibile perché non soddisfa la prima equazione e deduciamo che l'unico valore ammissibile per y è $y^* = 0$. Sostituendo $y^* = 0$ nell'ultima equazione (il vincolo) otteniamo $x^2 = 1$ che ha per soluzione due valori distinti, $x_1^* = -1$ e $x_2^* = 1$, che, sostituiti nella prima equazione, restituiscono i due valori corrispondenti per il moltiplicatore λ : $\lambda_1^* = 1/2$ e $\lambda_2^* = -1/2$. Otteniamo pertanto due punti stazionari per la funzione Lagrangiana: $(x_1^*, y_1^*, \lambda_1^*) = (-1, 0, 1/2)$ e $(x_2^*, y_2^*, \lambda_2^*) = (1, 0, -1/2)$.

Calcolando f su questi due punti stazionari [$f(x, y)$ dipende solo da x e da y e non da λ , quindi utilizziamo solo i primi due valori x^* e y^* della terna (x^*, y^*, λ^*)] ricaviamo

$$f(x_1^*, y_1^*) = f(-1, 0) = 4 + 1 = 5 \quad \text{e} \quad f(x_2^*, y_2^*) = f(1, 0) = 4 - 1 = 3.$$

Poiché non ci sono punti singolari, $(x_1^*, y_1^*) = (-1, 0)$ è l'unico punto di *massimo assoluto* mentre $(x_2^*, y_2^*) = (1, 0)$ è l'unico punto di *minimo assoluto* e non ci sono massimi o minimi relativi.

Nella figura 12.12(a) l'ellisse evidenziata in grigio sul grafico della funzione obiettivo $f(x, y) = 4 - x$ è la proiezione verticale del vincolo [la circonferenza $g(x, y) = x^2 + y^2 = 1$] sul grafico di f ; tale ellisse costituisce il grafico della funzione obiettivo ristretta al vincolo. I punti di massimo e di minimo per la f vincolata, $(-1, 0)$ e $(1, 0)$ rispettivamente, sono i punti di tangenza fra la circonferenza stessa e le curve di livello $f(x, y) = 5$ e $f(x, y) = 3$ (le rette in nero sul grafico di f).

ESEMPIO 12.7 Nel problema vincolato

$$\begin{aligned} & \text{ott } (x^2 + 2y^2 + 3) \\ & \text{sub } 2x^2 + y^2 = 4 \end{aligned}$$

il vincolo è regolare poiché $\nabla g(x, y) = (4x, 2y) \neq (0, 0)$ su tutti i punti tali che $2x^2 + y^2 = 4$. Trattandosi di un'ellisse, il vincolo è un insieme compatto di \mathbb{R}^2 e quindi esistono il massimo e minimo

assoluti. La funzione Lagrangiana è $L(x, y, \lambda) = x^2 + 2y^2 + 3 + \lambda(4 - 2x^2 - y^2)$ e i suoi punti stazionari risolvono il sistema

$$\begin{cases} L_x = 2x - 4\lambda x = 0 \\ L_y = 4y - 2\lambda y = 0 \\ L_\lambda = 4 - 2x^2 - y^2 = 0. \end{cases}$$

Dalla prima equazione otteniamo $2x(1 - 2\lambda) = 0$ che ci obbliga a studiare separatamente i due casi:

1. se $x = 0$, dalla terza equazione otteniamo $y_1 = 2$ e $y_2 = -2$ che, sostituiti nella seconda equazione, forniscono lo stesso valore del moltiplicatore, $\lambda_1 = \lambda_2 = 2$;
2. se $1 - 2\lambda = 0$, ovvero $\lambda = 1/2$, sostituendo nella seconda equazione otteniamo $y_3 = y_4 = 0$, che, a sua volta sostituito nella terza equazione, determina $x_3 = \sqrt{2}$ e $x_4 = -\sqrt{2}$.

Abbiamo quindi quattro punti stazionari per L : $(0, 2, 2)$, $(0, -2, 2)$, $(\sqrt{2}, 0, 1/2)$ e $(-\sqrt{2}, 0, 1/2)$. Il valore di f su ciascun punto (sulle prime due coordinate, x^* e y^*) è: $f(0, 2) = f(0, -2) = 11$, $f(\sqrt{2}, 0) = f(-\sqrt{2}, 0) = 5$; deduciamo che i primi due sono entrambi punti (distinti!) di *massimo assoluto* mentre gli ultimi due sono entrambi punti (distinti!) di *minimo assoluto*, e non c'è nient'altro.

Nella figura 12.12(b) l'ellisse distorta in grigio sul grafico della funzione obiettivo $f(x, y) = x^2 + 2y^2 + 3$ è la proiezione verticale del vincolo, l'ellisse definita da $g(x, y) = 2x^2 + y^2 = 4$; essa è il grafico di f ristretta al vincolo. I punti di ottimo per f vincolata, $P_1 = (0, 2)$, $P_2 = (0, -2)$, $P_3 = (\sqrt{2}, 0)$ e $P_4 = (-\sqrt{2}, 0)$, sono i punti di tangenza fra l'ellisse stessa e le curve di livello $f(x, y) = 5$ e $f(x, y) = 11$ (le altre due ellissi in nero sul grafico di f).

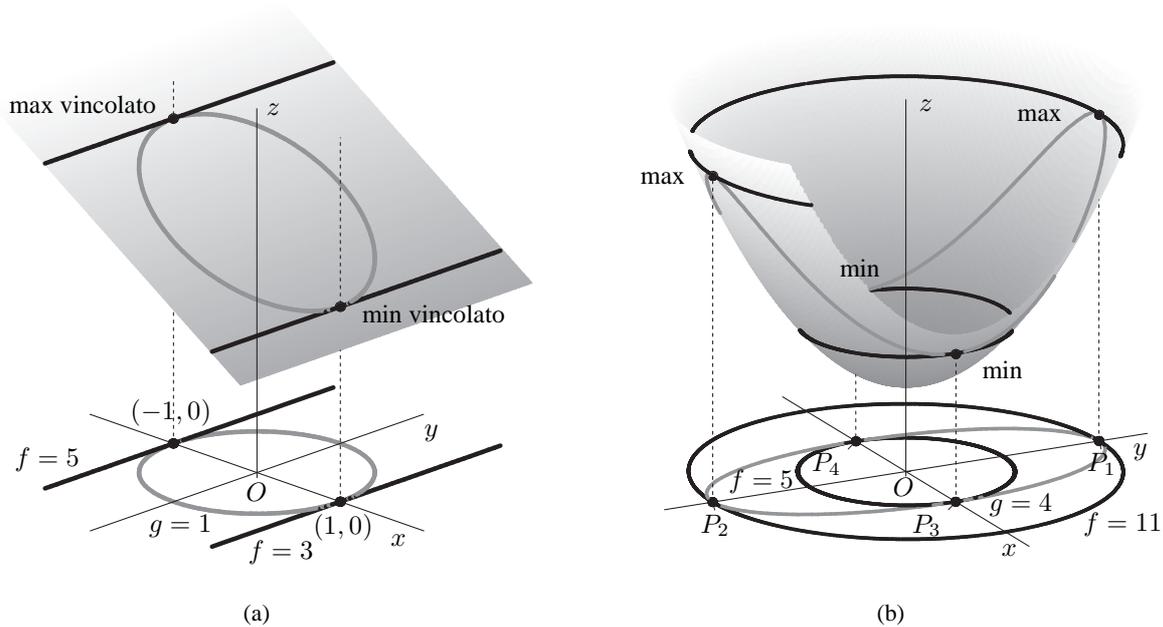


FIGURA 12.12: illustrazione della soluzione di problemi vincolati negli esempi (a) 12.6 e (b) 12.7.

ESEMPIO 12.8 (VINCOLO NON REGOLARE) Nel problema vincolato

$$\begin{aligned} & \text{ott } y \\ & \text{sub } y^3 - y^4 - x^2 = 0 \end{aligned}$$

il vincolo non è regolare, dal momento che $\nabla g(x, y) = (-2x, 3y^2 - 4y^3)$ si annulla nell'origine $(0, 0)$, punto appartenente al vincolo. Inseriamo, dunque, il *punto singolare* $(0, 0)$, nella lista dei punti critici.

La compattezza del vincolo si evince osservando che, poiché $y^3 - y^4 = y^3(1 - y) < 0$ per $y < 0$ e $y > 1$, la funzione $g(x, y) = y^3 - y^4 - x^2$ è negativa su tutti i punti (x, y) che distano più (hanno norma maggiore) di 1 dall'origine. Pertanto, la curva di livello 0 è tutta contenuta nell'intorno di raggio 1 di $(0, 0)$ ed è, dunque, limitata; deduciamo che il massimo e il minimo assoluti esistono e, poiché la funzione obiettivo non è costante, sono raggiunti in (almeno) due punti distinti. Lo studio della funzione Lagrangiana, $L(x, y, \lambda) = y + \lambda(x^2 + y^4 - y^3)$, fornisce le seguenti condizioni:

$$\begin{cases} L_x = 2\lambda x = 0 \\ L_y = 1 + 4\lambda y^3 - 3\lambda y^2 = 1 + \lambda y^2(4y - 3) = 0 \\ L_\lambda = x^2 + y^4 - y^3 = 0. \end{cases}$$

La prima equazione propone due scenari: 1) se $x = 0$ dalla terza equazione otteniamo $y^3(y - 1) = 0$ che fornisce due valori per y , $y_1 = 0$ e $y_2 = 1$, di cui, sostituendo nella seconda equazione, il primo porta ad un assurdo, $1 = 0$, e il secondo produce $1 + \lambda(4 - 3) = 0$ che ha come soluzione $\lambda = -1$; 2) se $\lambda = 0$ la seconda equazione porta nuovamente all'assurdo $1 = 0$. Il metodo del Lagrangiano è dunque capace di fornire *un solo punto stazionario*, $(0, 1, -1)$, da aggiungere al punto singolare $(0, 0)$ nella lista dei punti critici. Poiché $f(0, 0) = 0$ e $f(0, 1) = 1$, concludiamo che il punto singolare $(0, 0)$ è l'unico punto di *minimo assoluto* e $(0, 1)$ è l'unico punto di *massimo assoluto*.

La figura 12.13 mostra (in grigio) la curva definita dal vincolo $g(x, y) = y^3 - y^4 - x^2 = 0$ dell'esempio 12.8: sono evidenziati il punto di tangenza $(0, 1)$ tra il vincolo e la curva di livello $c = 1$ di f (la retta nera orizzontale superiore) e il punto singolare $(0, 0)$, che *non è un punto di tangenza* fra il vincolo e la curva di livello $c = 0$ di f (la retta nera orizzontale inferiore). La curva $g(x, y) = 0$ non è localmente esplicitabile nel punto $(0, 0)$ perché lì due tratti della curva sono tra loro tangenti (si tratta di una **cuspid**); per questo motivo il Teorema 12.1 non è applicabile e dunque il metodo del Lagrangiano fallisce.

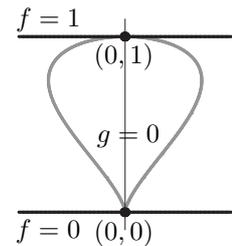


FIGURA 12.13: vincolo dell'esempio 12.8.

ESERCIZIO 12.5 Risolvere i seguenti problemi vincolati:

- | | | |
|--|---|---|
| 1. ott $(x - y)$
sub $x^2 + y^2 = 1$ | 2. ott $(x^2 - y^2)$
sub $x^2 + y^2 = 1$ | 3. ott $(x + y^2)$
sub $x^2 + (1/2)y^2 = 1$ |
| 4. ott e^{x+y}
sub $x^2 + 2y^2 = 1$ | 5. ott $(y - x^2)$
sub $x^4 + y^4 - y^2 = 0$ | 6. ott $(x^2 + y^2 - 2x - 4y + 7)$
sub $x^2 + y^2 = 3$. |

ESERCIZIO 12.6 Si dimostri il Teorema 10.5 nel caso di due variabili; si dimostri, cioè, che il vettore gradiente $\nabla f = (f_x, f_y)$ individua la direzione di massima pendenza in uscita da un punto prefissato (x_0, y_0) , sotto l'ipotesi che $\nabla f = (f_x, f_y) \neq (0, 0)$ nel punto (x_0, y_0) ; per semplicità, si assuma che entrambe le derivate parziali siano positive, $f_x > 0$ e $f_y > 0$.

12.3 Il problema vincolato in n variabili

Abbiamo visto che, quando le variabili sono n , la forma implicita $f(x) = f(x_1, \dots, x_n) = c$ determina un'ipersuperficie di dimensione $n - 1$ in \mathbb{R}^n . Ciò significa che un vincolo determinato da una sola uguaglianza, $g(x) = g(x_1, \dots, x_n) = b$, dove g è una funzione di $n > 2$ variabili, è un oggetto ben più "corposo" di una semplice curva; se, ad esempio, $n = 4$ il vincolo è un'ipersuperficie di dimensione 3, ovvero a tutti gli effetti un "solido" in \mathbb{R}^4 . Pertanto, quando $n > 2$, una sola uguaglianza definisce un vincolo che lascia ancora sufficiente "spazio di movimento" all'argomento di una funzione obiettivo, non essendo in grado di "ridurre" il problema di partenza ad un problema univariato secondo le modalità discusse nel paragrafo 12.2.3.

La struttura teorica del problema vincolato rimane la stessa. In particolare, un'ipersuperficie di dimensione $n - 1$ è un insieme *di frontiera* in \mathbb{R}^n : come qualsiasi intorno di un punto su una curva (ipersuperficie di dimensione 1) in \mathbb{R}^2 , contiene necessariamente punti sia della curva che del suo complemento [figura 8.11(d)], qualsiasi intorno di un punto appartenente a una superficie in \mathbb{R}^3 , che è una palla [figura 8.9(b)], contiene necessariamente punti sia della superficie che del suo complemento; lo stesso argomento vale per dimensioni n superiori a 3. La novità rispetto ai problemi in due variabili, è che in n variabili abbiamo la possibilità di considerare *vincoli determinati da più uguaglianze*, ovvero da più forme implicite $g_j(x) = b_j$.

12.3.1 Vincoli definiti da più uguaglianze

Se il vincolo è definito da una sola uguaglianza, che ora indicheremo con $g_1(x) = b_1$, una costruzione analoga a quella discussa nel paragrafo precedente presuppone, nelle ipotesi del Teorema 12.3, la sostituzione (localmente) di una delle variabili argomento della funzione obiettivo, ad esempio x_i , con la funzione esplicita di $n - 1$ variabili $x_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ 'estratta' dalla forma implicita $g_1(x) = b$. *In presenza di una sola uguaglianza*, quest'operazione non è altro che la *trasformazione (localmente) di un problema vincolato di n variabili in un problema libero equivalente in $n - 1$ variabili*. Se $n > 2$, la nuova funzione obiettivo è dunque ancora una funzione di più variabili ($n - 1$, per l'esattezza); ciò prefigura due scenari possibili: da un lato possiamo affrontare direttamente il problema 'trasformato' trattandolo come un problema libero in $n - 1$ variabili (sfruttando gli strumenti del capitolo 11), dall'altro abbiamo l'opportunità di specificare meglio il vincolo aggiungendo un'altra uguaglianza che indicheremo con $g_2(x) = b_2$. Nel secondo scenario, sempre se vale il Teorema 12.3, seguendo la solita procedura di sostituzione (localmente) di una variabile con una funzione esplicita 'estratta' da $g_2(x) = b_2$ nell'argomento della nuova funzione obiettivo, è possibile 'ridurre' ulteriormente il problema in un altro equivalente in $n - 2$ variabili; se $n > 3$, ciò che otteniamo è una nuova funzione obiettivo che dipende ancora da più di una variabile ($n - 2$, per l'esattezza). Sono possibili diverse 'riduzioni' del problema di partenza in problemi equivalenti: ciascuna riduzione richiede un'uguaglianza $g_j(x) = b_j$ che caratterizzi il vincolo e serve a trasformare un problema in $n - i$ variabili in un problema equivalente in $n - i - 1$ variabili.

Quante trasformazioni di questo tipo sono possibili? Ovvero, quante forme implicite $g_j(x) = b_j$ sono ammesse per definire il vincolo? Ovviamente bisogna fermarsi quando l'ultima trasformazione produce una funzione obiettivo di una variabile sola, togliere di mezzo anche quella renderebbe il problema di partenza privo di significato. Deduciamo che, quando le variabili sono $n > 2$, *il vincolo V può essere rappresentato da un numero di uguaglianze minore o uguale a $n - 1$* . Possiamo concludere che un problema vincolato in n variabili il cui vincolo è definito da $m \leq n - 1$ equazioni è (localmente) equivalente ad un problema libero in $n - m$ variabili.

Il problema vincolato in n variabili si scrive nel modo seguente:

$$\text{sub } \begin{cases} \text{ott } f(x) \\ g_1(x) = b_1 \\ g_2(x) = b_2 \\ \vdots \\ g_m(x) = b_m, \end{cases} \quad (12.17)$$

dove sia la funzione obiettivo f che tutte le funzioni g_j dipendono da n variabili, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$, per $j = 1, 2, \dots, m$, e il numero delle *uguaglianze* che definiscono il vincolo V è inferiore al numero delle variabili, $m \leq n - 1$.

Il sistema di equazioni che definisce il vincolo V in (12.17) individua l'insieme dei vettori $x \in \mathbb{R}^n$ che soddisfano 'contemporaneamente' tutte le m equazioni; in altre parole, il vincolo V è l'intersezione delle curve di livello (ipersuperfici) definite da ciascuna forma implicita $g_j(x) = b_j$. Più equazioni ci sono e più "sottile" è il vincolo. In \mathbb{R}^3 , ad esempio, una sola equazione $g(x) = b$ determina un vincolo che è una superficie (ipersuperficie di dimensione 2) in \mathbb{R}^3 , mentre due equazioni $g_j(x) = b_j, j = 1, 2$, determinano l'intersezione di due superfici, cioè una curva vera e propria (un'ipersuperficie di dimensione 1) in \mathbb{R}^3 . Se considerassimo un vincolo formato da tre equazioni in \mathbb{R}^3 , o, più in generale, un vincolo formato da n equazioni in \mathbb{R}^n , la soluzione del sistema in (12.17) sarebbe, salvo casi rarissimi, un numero esiguo di punti isolati (magari uno solo o addirittura nessuno!), perché questo è il numero delle soluzioni di un sistema di n equazioni in n incognite, rendendo quindi il problema privo di significato. Anche da questo punto di vista rimane confermata la necessità che le equazioni non siano più di $n - 1$.

12.3.2 Vincoli regolari

Quando è definito da una sola uguaglianza $g(x) = b$, affinché il vincolo sia regolare è sufficiente che $\nabla g(x) \neq (0, 0, \dots, 0)$ su tutti i punti x appartenenti alla curva definita da $g(x) = b$. Nel caso di vincoli determinati da più uguaglianze, $g_j(x) = b_j, j = 1, 2, \dots, m$, non basta più la semplice condizione $\nabla g_j(x) \neq (0, 0, \dots, 0)$ per ogni $j = 1, 2, \dots, m$. Poiché in presenza di più uguaglianze il vincolo è definito dai punti di intersezione delle curve di livello definite da ciascuna uguaglianza, tali punti di intersezione devono "comportarsi bene".

ESEMPIO 12.9 Studiamo la curva definita in \mathbb{R}^3 dal sistema di equazioni

$$\begin{cases} x^2 + y^2 + z^2 = 1 \\ 2(x - 1/2)^2 + 2y^2 = 1/2 \end{cases} \quad (12.18)$$

che descrive un vincolo in \mathbb{R}^3 del tipo (12.17) con $g_1(x, y, z) = x^2 + y^2 + z^2$, $g_2(x, y, z) = 2(x - 1/2)^2 + 2y^2$, $b_1 = 1$ e $b_2 = 1/2$. La prima equazione definisce la sfera di centro l'origine e raggio 1 in \mathbb{R}^3 , mentre la seconda definisce il cilindro verticale con centro passante nel punto $(1/2, 0, 0)$ e raggio⁶ $1/2$. La figura 12.14 mostra il vincolo come *intersezione* delle due superfici: si tratta di una curva unidimensionale immersa nello spazio tridimensionale \mathbb{R}^3 (la curva in nero). Il punto $(x_0, y_0, z_0) = (1, 0, 0)$ è palesemente problematico: in esso la curva *interseca se stessa*, esattamente come accadeva nell'esempio 12.4, l'unica differenza è che ora la curva è un sottoinsieme di \mathbb{R}^3 anziché di \mathbb{R}^2 . Siamo costretti a concludere, per le stesse ragioni illustrate nel paragrafo 12.1.3, che $(x_0, y_0, z_0) = (1, 0, 0)$ è un *punto singolare* e dunque la costruzione del paragrafo precedente non può funzionare su tale punto. Cionondimeno, i gradienti delle g_j sono entrambi non nulli sul punto: $\nabla g_1(1, 0, 0) = \nabla g_2(1, 0, 0) = (2, 0, 0)$.

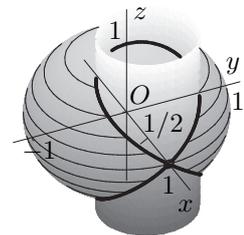


FIGURA 12.14: punto non regolare in \mathbb{R}^3 .

Dalla figura 12.14 evinciamo che $(1, 0, 0)$ è un punto di intersezione particolare: è un *punto di tangenza* fra le due superfici definite in (12.18); infatti, in $(1, 0, 0)$ i gradienti sono uguali e dunque paralleli.⁷ Deduciamo che un criterio per individuare i punti singolari dovrà per forza mettere i gradienti ∇g_j in relazione fra loro. Abbiamo appena visto che, oltre alla condizione $\nabla g_j \neq (0, 0, \dots, 0)$ per ogni j , dovrà valere $\nabla g_i \neq \nabla g_j$ per ogni $i \neq j$. In realtà serve qualcosa di più.

Poiché lo scopo di avere due equazioni per definire il vincolo di un problema in tre variabili è ridurre il problema da 3 dimensioni a 1 dimensione, la presenza di un punto di tangenza nell'insieme

⁶Ciò appare più evidente se riscriviamo il secondo vincolo nella forma equivalente $(x - 1/2)^2 + y^2 = 1/4$.

⁷Si rammenti, inoltre, che vale il Teorema 12.4.

intersezione fra le superfici definite da ciascuna equazione crea un'ambiguità: non è più certo, infatti, che l'intersezione sia una semplice curva perché la proprietà di 'tangenza' tende ad "allargare" la curva stessa trasformandola in una superficie vicino al punto stesso. Per quanto piccola possa essere, si tratta sempre di un insieme di dimensione maggiore di 1. Ciò è sufficiente a compromettere la strategia di fondo: sul punto di tangenza le due equazioni si equivalgono (individuano la stessa curva); una delle due è ridondante e quindi inutile al fine di ridurre la dimensione del problema.

La situazione è più complessa quando le variabili sono $n > 3$, anche se l'interpretazione rimane sostanzialmente la stessa: il vincolo è regolare se ciascuna equazione che lo definisce porta un contributo indipendente dalle altre; in altri termini, il vincolo dev'essere una curva che rappresenta, in dimensione $n - m$, una *funzione invertibile* su ciascun punto del vincolo. Non approfondiamo ulteriormente questi aspetti e forniamo direttamente il criterio operativo per verificare la regolarità o meno di un vincolo – o, più precisamente, di un punto appartenente ad esso.

Scriviamo le m funzioni che individuano il vincolo V in (12.17) sotto forma di *funzione vettoriale* $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (definizione 9.6) definita dalle m funzioni g_1, g_2, \dots, g_m :

$$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{bmatrix}.$$

In questo modo possiamo riscrivere il problema (12.17) nella forma più succinta:

$$\begin{array}{l} \text{ott } f(x) \\ \text{sub } g(x) = b, \end{array} \quad (12.19)$$

dove $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, b è il vettore di coefficienti $b^T = (b_1, b_2, \dots, b_m) \in \mathbb{R}^m$, e $m \leq n - 1$. Definiamo la matrice J ($m \times n$) le cui righe sono formate dagli m gradienti $\nabla g_j(x)$:

$$J(x) = \begin{bmatrix} (\partial/\partial x_1) g_1(x) & (\partial/\partial x_2) g_1(x) & \cdots & (\partial/\partial x_n) g_1(x) \\ (\partial/\partial x_1) g_2(x) & (\partial/\partial x_2) g_2(x) & & (\partial/\partial x_n) g_2(x) \\ \vdots & & \ddots & \vdots \\ (\partial/\partial x_1) g_m(x) & (\partial/\partial x_2) g_m(x) & \cdots & (\partial/\partial x_n) g_m(x) \end{bmatrix}. \quad (12.20)$$

La matrice J costituisce a tutti gli effetti la *derivata prima di g* e si chiama **matrice Jacobiana**.

Ricordiamo che il *rango di una matrice* A ($m \times n$), $\rho(A)$, è la massima dimensione delle sottomatrici quadrate estraibili da A con determinante non nullo (definizione 9.12).

DEFINIZIONE 12.5 Il punto x^0 appartenente al vincolo V del problema (12.19), ovvero che soddisfa l'equazione $g(x) = b$ [o il sistema di equazioni $g_j(x) = b_j$, per $j = 1, 2, \dots, m$], dove tutte le funzioni componenti, g_j , per $j = 1, 2, \dots, m$, sono differenziabili, si dice **regolare** se la *matrice Jacobiana* $J(x^0)$ definita in (12.20) ha *rango pieno* nel punto x^0 . Poiché necessariamente $m < n$, ciò equivale alla condizione $\rho(J) = m$.

Diremo che il **vincolo è regolare** se tutti i punti appartenenti ad esso sono regolari.

Il punto di tangenza $(x_0, y_0, z_0) = (1, 0, 0)$ dell'esempio 12.9 viola la definizione 12.5: la matrice Jacobiana è $J = \begin{bmatrix} 2x & 2y & 2z \\ 4(x - 1/2) & 4y & 0 \end{bmatrix}$, che in $(x_0, y_0, z_0) = (1, 0, 0)$ vale $J(1, 0, 0) = \begin{bmatrix} 2 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$; chiaramente J non ha rango pieno poiché vale $\rho[J(1, 0, 0)] = 1 < 2 = m$.

Osserviamo che la definizione 12.5 implica entrambe le condizioni $\nabla g_j \neq (0, 0, \dots, 0)$ per ogni $j = 1, 2, \dots, m$, e $\nabla g_i \neq \nabla g_j$ per ogni $i \neq j$; infatti, la matrice J non può avere rango pieno né se contiene una riga tutta composta da zeri, né se contiene due righe uguali (proprietà 9.1).

12.3.3 Il Teorema di Lagrange in n variabili

Il prossimo teorema estende il Teorema 12.6 stabilendo *condizioni necessarie del primo ordine* per problemi vincolati con n variabili e $m \leq n - 1$ equazioni che definiscono il vincolo. Al fine di riscrivere la funzione Lagrangiana (12.15) per il problema (12.17) associamo a ciascuna delle m equazioni che definiscono il vincolo, $g_j(x) = b_j$, per $j = 1, 2, \dots, m$, un *moltiplicatore* $\lambda_j \in \mathbb{R}$. La *funzione Lagrangiana*, $L : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, diventa quindi:

$$\begin{aligned} L(x, \lambda) &= f(x) + \lambda \cdot [b - g(x)] \\ &= f(x) + \lambda_1 [b_1 - g_1(x)] + \lambda_2 [b_2 - g_2(x)] + \dots + \lambda_m [b_m - g_m(x)], \end{aligned} \tag{12.21}$$

dove $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $b^T = (b_1, b_2, \dots, b_m) \in \mathbb{R}^m$ e $\lambda^T = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^m$ è il vettore dei moltiplicatori di Lagrange.

TEOREMA 12.7 (LAGRANGE II) È dato il problema (12.19) [ovvero il problema (12.17)] dove la funzione obiettivo f e tutte le funzioni g_j , per $j = 1, 2, \dots, m$, sono differenziabili. Sia x^0 un *punto di massimo o minimo* relativo per (12.19) che sia un *punto regolare* per il vincolo, cioè tale che $J(x^0)$ abbia rango pieno. Allora esistono m *moltiplicatori* $\lambda_j \in \mathbb{R}$, $j = 1, 2, \dots, m$ tali che $(x^0, \lambda) \in \mathbb{R}^{n+m}$ è un *punto stazionario* della funzione Lagrangiana (12.21), cioè tale che $\nabla L(x^0, \lambda) = (0, 0, \dots, 0)$. Specificamente, valgono le $n + m$ condizioni:

$$\left\{ \begin{array}{l} L_{x_1}(x^0, \lambda) = f_{x_1}(x^0) - \lambda_1 (\partial/\partial x_1) g_1(x^0) - \lambda_2 (\partial/\partial x_1) g_2(x^0) - \dots - \lambda_m (\partial/\partial x_1) g_m(x^0) = 0 \\ \vdots \\ L_{x_n}(x^0, \lambda) = f_{x_n}(x^0) - \lambda_1 (\partial/\partial x_n) g_1(x^0) - \lambda_2 (\partial/\partial x_n) g_2(x^0) - \dots - \lambda_m (\partial/\partial x_n) g_m(x^0) = 0 \\ L_{\lambda_1}(x^0, \lambda) = b_1 - g_1(x^0) = 0 \\ \vdots \\ L_{\lambda_m}(x^0, \lambda) = b_m - g_m(x^0) = 0. \end{array} \right.$$

Possiamo riscrivere il precedente sistema di $n + m$ equazioni nella seguente forma più succinta:

$$\left\{ \begin{array}{l} \nabla_x L(x^0, \lambda) = \nabla f(x^0) - \lambda^T J(x^0) = 0 \\ \nabla_\lambda L(x^0, \lambda) = b - g(x^0) = 0, \end{array} \right. \tag{12.22}$$

dove ∇_x e ∇_λ indicano i gradienti di L rispetto alle prime n variabili, x_1, x_2, \dots, x_n e rispetto alle ultime m variabili, $\lambda_1, \lambda_2, \dots, \lambda_m$, rispettivamente, e $\lambda^T J(x^0)$ è il prodotto matriciale fra la matrice $1 \times m$ (vettore riga) λ^T e la matrice Jacobiana $J(m \times n)$ calcolata in x^0 . Si tenga presente che lo zero della prima equazione in (12.22) è il vettore nullo $0 \in \mathbb{R}^n$ e lo zero della seconda equazione è il vettore nullo $0 \in \mathbb{R}^m$.

ESEMPIO 12.10 Risolviamo il seguente problema vincolato:

$$\begin{array}{l} \text{ott} (x^2 + y^2) \\ \text{sub} \left\{ \begin{array}{l} x^2 + z^2 = 1 \\ y - x = 0. \end{array} \right. \end{array}$$

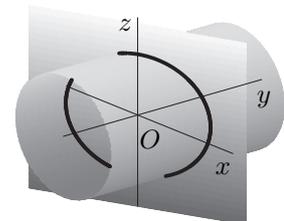


FIGURA 12.15: vincolo dell'esempio 12.10.

Il vincolo è l'intersezione delle curve (superfici) definite dalle equazioni $x^2 + z^2 = 1$ e $x - y = 0$ in \mathbb{R}^3 ; si tratta dell'intersezione fra il cilindro orizzontale centrato sull'asse y di raggio 1 e il piano verticale individuato dalla bisettrice fra il primo e il terzo ortante del piano orizzontale individuato dagli assi x e y , ovvero l'ellisse evidenziata in grassetto in figura 12.15. Chiaramente è un *insieme compatto*.

La matrice Jacobiana, $J = \begin{bmatrix} 2x & 0 & 2z \\ -1 & 1 & 0 \end{bmatrix}$ ha rango pieno su tutti i punti del vincolo, $\rho(J) = 2$,

per ogni (x, y, z) ammissibile; infatti, le due sottomatrici $\begin{bmatrix} 2x & 0 \\ -1 & 1 \end{bmatrix}$ e $\begin{bmatrix} 0 & 2z \\ 1 & 0 \end{bmatrix}$ non possono avere contemporaneamente determinante nullo perché la prima equazione del vincolo, $x^2 + z^2 = 1$, non ammette che x e z siano contemporaneamente nulli. Pertanto il *vincolo è regolare*.

La funzione Lagrangiana è

$$L(x, y, z, \lambda_1, \lambda_2) = x^2 + y^2 + \lambda_1(1 - x^2 - z^2) + \lambda_2(x - y),$$

e i suoi punti stazionari risolvono le condizioni (12.22):

$$\begin{cases} L_x = 2x - 2\lambda_1 x + \lambda_2 = 0 \\ L_y = 2y - \lambda_2 = 0 \\ L_z = -2\lambda_1 z = 0 \\ L_{\lambda_1} = 1 - x^2 - z^2 = 0 \\ L_{\lambda_2} = x - y = 0 \end{cases}$$

Dall'ultima equazione ricaviamo $y = x$, che, assieme alla seconda, stabilisce che $\lambda_2 = 2x = 2y$. Sostituendo nelle altre equazioni, otteniamo il sistema equivalente in tre sole incognite:

$$\begin{cases} 2x(2 - \lambda_1) = 0 \\ 2\lambda_1 z = 0 \\ x^2 + z^2 = 1. \end{cases}$$

La seconda equazione prefigura due scenari possibili: deve valere $\lambda_1 = 0$ oppure $z = 0$. Se $\lambda_1 = 0$ la prima equazione implica $x = 0$ e quindi anche $y = 0$ e $\lambda_2 = 0$, mentre dalla terza equazione ricaviamo $z^2 = 1$, ovvero $z = \pm 1$; deduciamo che i primi due punti stazionari sono $(0, 0, -1, 0, 0)$ e $(0, 0, 1, 0, 0)$. Se $z = 0$ la terza equazione diventa $x^2 = 1$, da cui ricaviamo $x = \pm 1$ e dalla prima equazione $\lambda_1 = 2$. Deduciamo che ci sono altri due punti stazionari: $(-1, -1, 0, 2, -2)$ e $(1, 1, 0, 2, 2)$.

Poiché $f(0, 0, -1) = f(0, 0, 1) = 0$ e $f(-1, -1, 0) = f(1, 1, 0) = 2$, deduciamo che $(0, 0, -1)$ e $(0, 0, 1)$ sono entrambi punti di *minimo assoluto* mentre $(-1, -1, 0)$ e $(1, 1, 0)$ sono entrambi punti di *massimo assoluto*; non ci sono massimi o minimi relativi.

ESERCIZIO 12.7 Risolvere i seguenti problemi vincolati:

1. $\text{ott } [x - (1/3)y^3]$
sub $x^2 + y^2 + z^2 = 2$
2. $\text{ott } (xyz)$
sub $x^2 + y^2 + z^2 = 1$
3. $\text{ott } (x + y^2)$
sub $\begin{cases} x^2 + z^2 = 1 \\ y - x = 0. \end{cases}$
4. $\text{ott } (x + y + z)$
sub $\begin{cases} x^2 + y^2 = 1 \\ x - y + z = 1. \end{cases}$
5. $\text{ott } (x + y + 2z)$
sub $\begin{cases} x^2 + z^2 = 8 \\ x - y = 1. \end{cases}$
6. $\text{ott } (xyz)$
sub $\begin{cases} x^2 + z^2 = 1 \\ x + y + z = 0. \end{cases}$

Capitolo 13

Ottimizzazione vincolata II: vincoli di disuguaglianza

I problemi con vincoli di disuguaglianza costituiscono il caso più generale possibile poiché presuppongono l'analisi *dei punti interni e di frontiera contemporaneamente*. È necessaria pertanto una sintesi di entrambe le tecniche viste nei capitoli 11 e 12 rispettivamente, quella per problemi di ottimo libero e quella per problemi con vincoli di uguaglianza, che sfocerà nel metodo unificato delle condizioni di Kuhn-Tucker. A seconda delle caratteristiche della funzione obiettivo e/o del vincolo sarà possibile scegliere uno dei due approcci principali trattati finora: quello basato sulla concavità della funzione obiettivo oppure quello basato sulla compattezza del vincolo.

13.1 Definizione del problema

Dobbiamo dunque occuparci di vincoli “grassi” (con la “pancia”): insiemi di \mathbb{R}^2 formati da superfici e dal loro contorno, solidi in \mathbb{R}^3 compresa la loro “buccia”, ecc.. La tecnica che costruiremo ci permetterà di studiare i *punti di massimo (o minimo) assoluti*, sia che si trovino sulla frontiera oppure siano punti interni; in questo senso si tratta di una sintesi dei due approcci visti in precedenza che implica il loro utilizzo congiuntamente. Avremo a disposizione entrambi gli schemi adottati finora: 1) quello basato sulle ipotesi di convessità del vincolo e concavità della funzione obiettivo e 2) quello basato sull'ipotesi di compattezza del vincolo e continuità della funzione obiettivo (Teorema di Weierstrass). Data la generalità del tipo di vincolo considerato, questi schemi non si escludono a vicenda: un vincolo convesso può benissimo essere anche compatto e viceversa, mentre tutte le funzioni obiettivo considerate saranno differenziabili (e quindi continue) e spesso anche concave.

I vincoli formati sia da punti interni che da punti di frontiera vengono ancora espressi da funzioni di n variabili $g : \mathbb{R}^n \rightarrow \mathbb{R}$, come nel caso dei vincoli di uguaglianza, però in questo caso le funzioni g , anziché essere coinvolte nei termini di un'equazione, compaiono come termini di una disequazione. Osserviamo fin da subito una novità: diversamente dai problemi con vincoli di uguaglianza, per i quali il numero di equazioni doveva essere minore o uguale a $n - 1$, ora *non ci sono limiti sul numero di disuguaglianze da impiegare*: poiché ciascuna disequazione definisce un insieme (con relativa frontiera) di \mathbb{R}^n e il vincolo è *l'intersezione di tutti questi insiemi*, non c'è limite al numero di insiemi da intersecare.

La formulazione standard dei problemi oggetto di questo paragrafo è la seguente:

$$\text{sub} \begin{cases} \max f(x) \\ g_1(x) \leq b_1 \\ g_2(x) \leq b_2 \\ \vdots \\ g_m(x) \leq b_m \end{cases} \quad \text{sub} \begin{cases} \min f(x) \\ g_1(x) \geq b_1 \\ g_2(x) \geq b_2 \\ \vdots \\ g_m(x) \geq b_m, \end{cases} \quad (13.1)$$

dove la funzione obiettivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e tutte le funzioni $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ per $j = 1, 2, \dots, m$ dipendono da n variabili, e il numero delle disuguaglianze, m , è arbitrario.

In questo caso è necessario distinguere se si tratta di un problema di massimo oppure di minimo. La differenza è nel verso delle disuguaglianze: per i problemi di massimo tutte le disuguaglianze sono del tipo ' \leq ', mentre per i problemi di minimo le disuguaglianze sono del tipo ' \geq '. L'adozione di disuguaglianze deboli, ' \leq ' e ' \geq ', anziché disuguaglianze strette, ' $<$ ' e ' $>$ ', lascia intendere che il vincolo identificato da m disequazioni contiene tutti i suoi punti di frontiera.

DEFINIZIONE 13.1 Diremo che x^0 è un **punto di massimo (minimo) assoluto vincolato stretto** per il problema (13.1) se $f(x) < f(x^0)$ [$f(x) > f(x^0)$] per ogni $x \in X \cap V$, $x \neq x^0$, dove X è il dominio di f e il vincolo, V , è la soluzione del sistema di disequazioni in (13.1).

Per la definizione di punto estremo debole basta sostituire $f(x) \leq f(x^0)$ [$f(x) \geq f(x^0)$] al posto di $f(x) < f(x^0)$ [$f(x) > f(x^0)$] e contemplare anche il caso $x = x^0$.

13.2 Vincoli espressi da disuguaglianze

Vediamo come si individuano sottoinsiemi "corpulenti" di \mathbb{R}^n mediante disequazioni.

DEFINIZIONE 13.2 Data $f : X \rightarrow \mathbb{R}$, $X \subseteq \mathbb{R}^n$, si dicono **insiemi di livello inferiore** gli insiemi

$$LI = \{x \in X : f(x) \leq c, c \in \mathbb{R}\},$$

e si dicono **insiemi di livello superiore** gli insiemi

$$LS = \{x \in X : f(x) \geq c, c \in \mathbb{R}\}.$$

Più specificamente parleremo di *insieme di livello inferiore (superiore)* c l'insieme LS (LI) caratterizzato dal livello $c \in \mathbb{R}$.

Gli insiemi di livello, a differenza delle curve di livello, sono insiemi "grassi" di \mathbb{R}^n : sono *insiemi di dimensione n* (la stessa dello spazio di appartenenza) *delimitati dalle curve di livello c* . L'esempio più eclatante è costituito dalle disequazioni del tipo $f(x) \leq c$ [$f(x) \geq c$] che coinvolgono una funzione di una sola variabile, $f : \mathbb{R} \rightarrow \mathbb{R}$: tali disequazioni hanno come soluzione un *intervallo* chiuso oppure l'*unione di intervalli* chiusi; ciò significa che gli insiemi di livello delle funzioni di una variabile hanno la stessa dimensione della retta reale \mathbb{R} , $n = 1$, ovvero hanno la stessa "consistenza" dello spazio, \mathbb{R} , di appartenenza. Si noti che un'equazione $f(x) = c$ con f di una variabile ha come soluzione al più un numero finito di punti isolati, che sono le 'curve di livello' in \mathbb{R} ; si tratta di insiemi di dimensione $1 - 1 = 0$, inferiore a quella di \mathbb{R} .

In \mathbb{R}^2 un insieme di livello è una delle due superfici separate dalla curva di livello c . Le figure 13.1(a) e 13.1(b) mostrano, rispettivamente, gli insiemi LI e LS (le aree grigie sul piano \mathbb{R}^2) della

funzione di due variabili $f(x, y) = x^2 + y^2$ per un certo livello c ; gli stessi insiemi sono evidenziati come superfici nel piano \mathbb{R}^2 nelle figure 13.1(c) e 13.1(d). Entrambi gli insiemi contengono la frontiera, la curva di livello c , che in questo caso è una circonferenza concentrica con centro l'origine (figura 12.2). Pertanto, i livelli inferiori, LI , sono i cerchi (superfici) contenuti in ciascuna circonferenza (inclusa la circonferenza), mentre i livelli superiori LS sono tutta la superficie che sta fuori dal cerchio (di nuovo inclusa la circonferenza), come si vede nelle figure 13.1(c) e 13.1(d). Si noti che gli insiemi LI e LS sono insiemi "quasi" complementari: hanno sempre in comune la frontiera, l'insieme $LI \cap LS$ che coincide con la curva di livello c (l'insieme costituito da tutti i punti che soddisfano la definizione 13.2 con l'uguaglianza, $f(x) = c$).

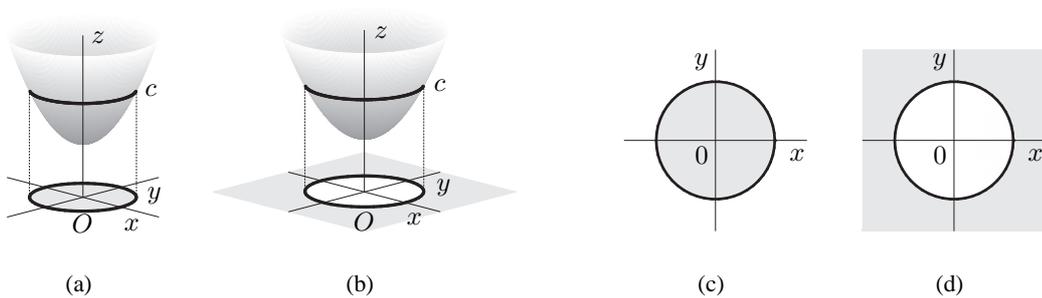


FIGURA 13.1: (a) insiemi di livello inferiore e (b) insiemi di livello superiore di $f(x, y) = x^2 + y^2$ (superfici in grigio sul piano \mathbb{R}^2) con i relativi dettagli in (c) e (d).

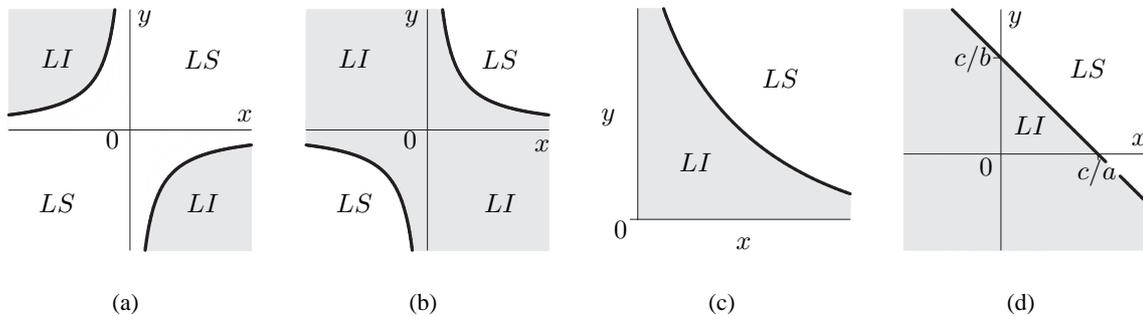


FIGURA 13.2: insiemi di livello inferiore LI (superfici grigie) e superiore LS (superfici bianche) delle funzioni (a) xy per $c < 0$, (b) xy per $c > 0$, (c) $\sqrt{x} + \sqrt{y}$ e (d) $ax + by$ con $a, b > 0$ per $c > 0$.

Per ricavare gli insiemi di livello inferiore della funzione $f(x, y) = xy$ studiata al primo punto dell'esempio 12.1 osserviamo la figura 12.3(d): se $c < 0$, poiché le iperboli si allontanano dall'origine al decrescere di c , un generico insieme di livello inferiore LI è costituito dalle aree grigie in figura 13.2(a), mentre il corrispondente insieme di livello superiore LS ne è l'insieme complementare (con la frontiera in comune a LI); per valori positivi di c , invece, gli insiemi LI assumono la forma della figura 13.2(b). Analogamente, dalla figura 12.3(e) è immediato rendersi conto che gli insiemi di livello inferiore (e superiore) della funzione $f(x, y) = \sqrt{x} + \sqrt{y}$ studiata al secondo punto dell'esempio 12.1 sono del tipo rappresentato in figura 13.2(c). La figura 13.2(d) mostra infine l'insieme di livello inferiore (e superiore) per $c > 0$ di una funzione lineare $f(x, y) = ax + by$ con $a, b > 0$;

la curva di livello c , esplicitabile con la retta definita da $y = h(x) = -(a/b)x + c/b$, interseca gli assi x e y nei punti $x = c/a$ e $y = c/b$ rispettivamente.

Il vincolo nei problemi di massimo in (13.1) è costituito dall'**intersezione degli insiemi di livello inferiore LI definiti dai livelli b_i** per ciascuna funzione g_i , $i = 1, \dots, m$; mentre il vincolo nei problemi di minimo in (13.1) è costituito dall'**intersezione degli insiemi di livello superiore LS definiti dai livelli b_i** per ciascuna funzione g_i , $i = 1, \dots, m$. Poiché l'intersezione di insiemi n -dimensionali di \mathbb{R}^n in generale è ancora un insieme n -dimensionale, a differenza dei vincoli definiti da curve di livello (uguaglianze), non c'è alcun motivo per limitare il numero degli insiemi di livello (delle disuguaglianze) in (13.1). In presenza di una disequazione unica, $g(x) \leq b$ [o $g(x) \geq b$], il vincolo corrisponde all'intero insieme di livello b (LI o LS) della g .

OSSERVAZIONE 13.1 Gli insiemi LI di una funzione f corrispondono agli insiemi LS della funzione opposta, $-f$, e viceversa. Ad esempio, $f(x, y) = -x^2 - y^2$ ha insiemi LS come quelli in figura 13.1(c) e insiemi LI come in figura 13.1(d). Questa proprietà permette di scrivere i vincoli in modo equivalente con disuguaglianze del tipo ' \geq ' o con disuguaglianze di tipo ' \leq ': lo stesso vincolo è ottenibile invertendo la disuguaglianza e cambiando il segno delle funzioni g_i e delle costanti b_i .

ESEMPIO 13.1 Date le funzioni $g_1(x, y) = (x-1)^2 + (y-1)^2$, $g_2(x, y) = -\sqrt{x} - \sqrt{y}$ e le due costanti $b_1 = 1/2$ e $b_2 = -2$, disegniamo il vincolo V definito dalla coppia di disequazioni

$$\begin{cases} (x-1)^2 + (y-1)^2 \leq 1/2 \\ -\sqrt{x} - \sqrt{y} \leq -2, \end{cases}$$

espresse in modo da rappresentare il vincolo per un problema di massimo in (13.1). g_1 ha come grafico un paraboloide con centro in $(1, 1)$ (è la forma quadratica in figura 8.18(a) traslata di 1 su ciascun asse cartesiano), pertanto la prima disequazione individua il cerchio di centro $(1, 1)$ e raggio $1/\sqrt{2}$ (inclusa la circonferenza). g_2 ha come insieme di livello inferiore LI per $b_2 = -2$ l'insieme LS in figura 13.2(c), si tratta infatti dell'opposto della funzione $f(x, y) = \sqrt{x} + \sqrt{y}$ lì considerata e pertanto i suoi insiemi LI coincidono con gli insiemi LS di f . La figura 13.3(a) mostra il vincolo V (la superficie in grigio scuro) ottenuto dall'intersezione tra gli insiemi LI $g_1(x, y) \leq 1/2$ e $g_2(x, y) \leq -2$: è un insieme compatto, convesso ed ha la stessa dimensione dello spazio che lo contiene.

ESEMPIO 13.2 Consideriamo le funzioni $g_1(x, y) = -y$, $g_2(x, y) = -x^2 - (y-1)^2$ e $g_3(x, y) = y - e^{-x}$ con relative costanti $b_1 = -1$, $b_2 = -1$ e $b_3 = 0$. Il vincolo V che ci interessa è definito da

$$\begin{cases} -y \leq -1 \\ -x^2 - (y-1)^2 \geq -1 \\ y - e^{-x} \geq 0. \end{cases}$$

Supponiamo di voler usare questo vincolo per definire un problema di minimo; le ultime due disequazioni si presentano già con le disuguaglianze giuste, dobbiamo solamente trasformare la prima disequazione in una disuguaglianza di tipo ' \geq '. Il nostro vincolo 'operativo' sarà pertanto:

$$\begin{cases} y \geq 1 \\ -x^2 - (y-1)^2 \geq -1 \\ y - e^{-x} \geq 0. \end{cases}$$

L'insieme LS di livello $b_1 = 1$ della prima funzione, $g_1(x, y) = y$, è il semipiano che sta sopra la retta $y \equiv 1$, inclusa la retta medesima; l'insieme LS di livello $b_2 = -1$ di $g_2(x, y) = -x^2 - (y-1)^2$ è il cerchio di centro $(0, 1)$ e raggio 1, inclusa la circonferenza; mentre l'insieme LS di livello $b_3 = 0$ di $g_3(x, y) = y - e^{-x}$ è la regione che sta sopra il grafico della funzione $y = e^{-x}$ ($y - e^{-x} = 0$ è una forma implicita esplicitabile), incluso il grafico stesso. La figura 13.3(b) mostra il vincolo V ottenuto dall'intersezione fra i tre insiemi di livello appena descritti. Ancora una volta si tratta di una superficie – insieme di dimensione 2 – ed è un insieme compatto e convesso.

ESEMPIO 13.3 (VINCOLO DI BILANCIO) Vediamo come si rappresenta il vincolo di bilancio già discusso al punto 4 dell'esempio 8.4 mediante un sistema di tre disequazioni. Innanzitutto, affinché qualsiasi problema economico soggetto a un vincolo di bilancio abbia significato, *le variabili* x e y , che esprimono quantità di beni, fattori ecc., *devono essere non negative*, ovvero devono senz'altro valere le disuguaglianze $x \geq 0$ e $y \geq 0$. La funzione $g_3(x, y) = p_1x + p_2y$ definisce la *spesa complessiva* per l'acquisto di due beni in quantità x e y dati i prezzi, p_1 e p_2 , dei due beni (ovviamente positivi!). Se la ricchezza disponibile è $w > 0$, la spesa non può superare questo valore (escludiamo, per semplicità, la possibilità di potersi rivolgere a uno strozzino), pertanto, la terza disequazione è $p_1x + p_2y \leq w$, e il nostro vincolo è completo. Poiché i vincoli di bilancio sono legati in modo naturale a *problemi di massimo* (massimo profitto, utilità massima ecc.), dobbiamo riscrivere le prime due disequazioni in modo appropriato, ottenendo il sistema di disequazioni in forma standard

$$\begin{cases} -x \leq 0 \\ -y \leq 0 \\ p_1x + p_2y \leq w. \end{cases}$$

Questo sistema di condizioni definisce il *vincolo di bilancio*, che indicheremo con B : l'insieme delle combinazioni possibili di quantità acquistate x e y dati i prezzi p_1 e p_2 e la ricchezza w .

Il vincolo B è rappresentato dal triangolo rettangolo grigio scuro in figura 13.3(c): è, cioè, l'intersezione fra l'ortante positivo \mathbb{R}_+^2 (a sua volta intersezione dei due semipiani $x \geq 0$ e $y \geq 0$) e il semipiano inferiore (l'area grigio chiaro) individuato dalla retta di bilancio $p_1x + p_2y = w$ [esplicitabile come $y = w/p_2 - (p_1/p_2)x$], che interseca gli assi cartesiani nei punti w/p_1 e w/p_2 rispettivamente. Anche questo è un insieme di dimensione 2 compatto e convesso. I punti sui cateti del triangolo presuppongono l'acquisto di uno solo dei due beni (in particolare, nell'origine non si acquista nulla e si risparmia tutta la ricchezza, eventualità sempre ammissibile), mentre i punti sull'ipotenusa equivalgono a spendere tutta la ricchezza disponibile senza risparmiare nulla; infine, i punti interni al triangolo rappresentano quantità acquistate senza impiegare tutta la ricchezza.

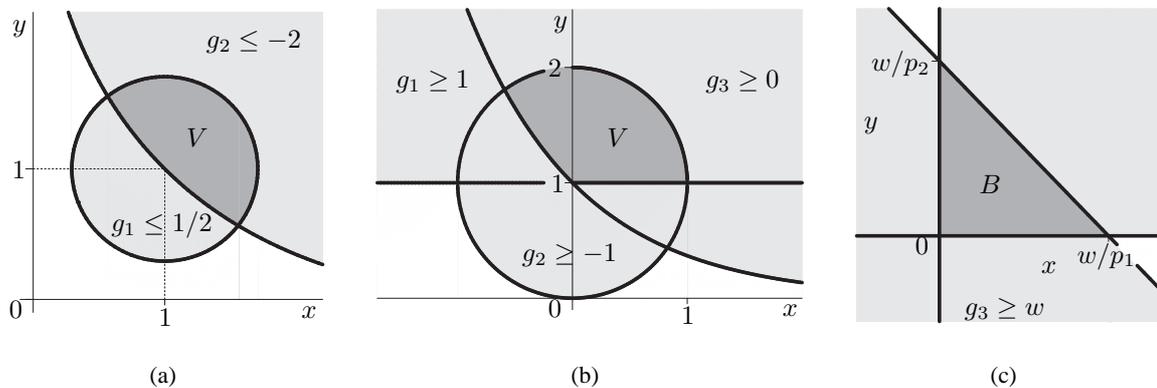


FIGURA 13.3: vincoli degli esempi 13.1, 13.2 e 13.3.

La figura 13.4 mostra il *vincolo di bilancio nel caso di tre beni*, in quantità x , y e z , dati i rispettivi prezzi, p_1 , p_2 e p_3 , e la ricchezza, w , definito da

$$\begin{cases} -x \leq 0 \\ -y \leq 0 \\ -z \leq 0 \\ p_1x + p_2y + p_3z \leq w. \end{cases}$$

Si noti la somiglianza con la figura 13.3(c): si tratta di una “piramide” in \mathbb{R}^3

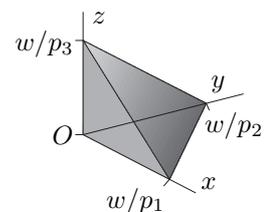


FIGURA 13.4: vincolo di bilancio in \mathbb{R}^3 .

con vertici $(0, 0, 0)$, $(w/p_1, 0, 0)$, $(0, w/p_2, 0)$ e $(0, 0, w/p_3)$; è ancora un insieme compatto e convesso ma questa volta ha dimensione 3.

ESERCIZIO 13.1 Scrivere in forma standard e disegnare il vincolo espresso dalle quattro disequazioni, $x \geq -1$, $x \leq 1$, $y \geq -1$ e $y \leq 1$, in \mathbb{R}^2 .

13.2.1 Vincoli convessi

Gli esempi visti mostrano che è facile costruire vincoli contemporaneamente compatti e convessi. La convessità, in particolare, è un requisito essenziale per poter applicare la teoria del paragrafo 11.6. I prossimi due risultati forniscono condizioni sufficienti per la convessità del vincolo in (13.1).

TEOREMA 13.1 Data $f : X \rightarrow \mathbb{R}$, dove $X \subseteq \mathbb{R}^n$ è un insieme convesso,

1. se f è convessa i suoi insiemi di livello inferiore LI sono convessi per ogni livello c , mentre
2. se f è concava i suoi insiemi di livello superiore LS sono convessi per ogni livello c .

Presi due punti $x, y \in LI$, se f è (debolmente) convessa per ogni $0 \leq \alpha \leq 1$ vale

$$f[\alpha x + (1 - \alpha)y] \leq \alpha f(x) + (1 - \alpha)f(y) \leq \alpha c + (1 - \alpha)c = c,$$

dove la prima disuguaglianza è la definizione 11.7 e la seconda sfrutta la definizione 13.2 per i punti x e y ; deduciamo che anche $\alpha x + (1 - \alpha)y \in LI$ e la prima parte del Teorema 13.1 è dimostrata. La seconda parte si prova in modo analogo. La figura 13.1(a) illustra il teorema: una funzione convessa ha il grafico come una “scodella” verso il basso, poiché gli insiemi LI sono la proiezione sul piano orizzontale di questa scodella si tratta per forza di insiemi convessi.

TEOREMA 13.2 L'intersezione di un qualsiasi numero di insiemi convessi è un insieme convesso.

Osservando la figura 13.3 notiamo che continuando a intersecare insiemi convessi ciò che otteniamo o è l'insieme vuoto \emptyset oppure è un insieme a sua volta convesso. La dimostrazione del Teorema 13.2 si basa sull'idea che due punti qualsiasi nell'intersezione di insiemi convessi stanno in ciascun insieme preso singolarmente; essendo questi insiemi tutti convessi, tutti contengono il segmento congiungente i due punti che pertanto è contenuto anche nell'intersezione.

I Teoremi 13.1 e 13.2 presi congiuntamente ci dicono che per verificare la convessità del vincolo è sufficiente verificare che tutte le funzioni g_i che lo definiscono siano convesse nel caso di un problema di massimo, ovvero che siano concave nel caso di un problema di minimo; si tratta cioè di applicare gli strumenti del paragrafo 11.6 a ciascuna funzione g_i .

OSSERVAZIONE 13.2 Poiché le funzioni lineari sono sia concave che convesse (debolmente), i vincoli costruiti a partire da funzioni g_i lineari sono sempre convessi, sia nei problemi di massimo che nei problemi di minimo. Il caso eclatante è il vincolo di bilancio B descritto nell'esempio 13.3.

13.2.2 Frontiera del vincolo e gradiente

È il momento di chiarire il motivo della distinzione fra le disuguaglianze del tipo ‘ \leq ’ e le disuguaglianze del tipo ‘ \geq ’ usate per definire il vincolo in problemi di massimo e minimo rispettivamente in (13.1). Pur avendo appreso che per definire lo stesso insieme entrambe le disuguaglianze vanno bene, esiste una differenza importante nell'usare una disuguaglianza piuttosto che l'altra.

Soffermiamo la nostra attenzione sugli insiemi di livello inferiore LI delle g_i , cioè gli insiemi che, riferiti alle funzioni $g_i, i = 1, \dots, m$, definiscono il vincolo per i *problemi di massimo* in (13.1) mediante disuguaglianze del tipo ‘ \leq ’. Supponiamo, per semplicità, che il vincolo sia rappresentato da un’unica disequazione $g(x) \leq b$, ovvero che coincida con l’insieme LI corrispondente al livello b di g . La curva di livello $g(x) = b$ contiene tutti i punti di frontiera dell’insieme LI . Sappiamo che in ciascun punto x di tale curva il gradiente, $\nabla g(x)$, è il vettore perpendicolare alla curva stessa il cui verso indica la direzione di massima pendenza di g in uscita da x (Teoremi 12.4 e 10.5). Poiché la curva di livello $g(x) = b$ in questo caso delimita un *insieme di livello inferiore* per g , siamo in grado di stabilire fin da subito qual è il verso del gradiente ∇g su ciascun punto della curva: esso indica sempre una *direzione di uscita dal vincolo*, ovvero dall’insieme LI . Infatti, immaginando di camminare sul grafico di g in prossimità del livello b , se “entriamo” nell’insieme LI (nel vincolo) a partire da un punto della frontiera “scendiamo” verso valori inferiori per g , mentre se “usciamo” dall’insieme LI (dal vincolo), ovvero entriamo nell’insieme LS , a partire da un punto della frontiera, “saliamo” verso valori più alti di g ; poiché ∇g indica la direzione di massima pendenza, è evidente che esso deve necessariamente ‘puntare fuori’ da LI , ovvero ‘dentro’ LS .

Viceversa, per analogia è facile vedere che se consideriamo *problemi di minimo* in (13.1), poiché in questo caso, sempre nell’ipotesi di vincolo unico, il vincolo è definito da un *insieme di livello superiore* LS , il gradiente ∇g calcolato sui punti di frontiera del vincolo punterà invece verso l’interno del vincolo stesso; in altre parole, esso indica sempre una *direzione di entrata nel vincolo*, (‘dentro’ l’insieme LS). La figura 13.5 rappresenta i due casi distinti: in (a) tutti i gradienti calcolati sul bordo di un vincolo espresso da $g(x) \leq b$ sono orientati verso l’esterno e in (b) tutti i gradienti calcolati sul bordo di un vincolo espresso da $g(x) \geq b$ sono orientati verso l’interno.

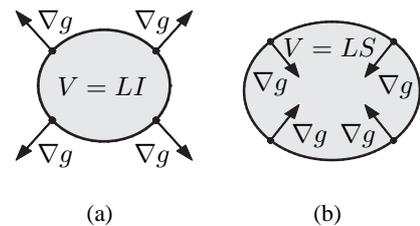


FIGURA 13.5: gradienti sul bordo di vincoli del tipo (a) $g(x) \leq b$, (b) $g(x) \geq b$.

Lo stesso argomento si estende facilmente a vincoli definiti da più disequazioni: poiché la frontiera del vincolo è l’unione di alcuni tratti di ciascun bordo dell’insieme di livello (inferiore o superiore) delle funzioni g_i , il gradiente ∇g_i indicherà una ‘direzione di uscita dal vincolo’ nel caso di disuguaglianze ‘ \leq ’, o una direzione di ‘entrata nel vincolo’ nel caso di disuguaglianze ‘ \geq ’.

13.2.3 Qualificazione dei vincoli attivi

Nei paragrafi 12.1.3 e 12.3.2 abbiamo discusso le condizioni di regolarità dei vincoli formati da punti di frontiera che sono curve (ipersuperfici) in \mathbb{R}^n . Ora ci troviamo di fronte vincoli che hanno la *frontiera composta da più spezzoni di curve*, ciascuna definita implicitamente da una funzione g_i : oltre a richiedere che ciascuna di queste curve si comporti bene, dovremo prestare attenzione ai punti di incontro (intersezione) tra tali curve. La questione, però, è rilevante solo se il punto estremo che risolve il problema (13.1) è anche un punto di frontiera; se si tratta di un punto interno, il comportamento delle curve che costituiscono la frontiera ci lascia del tutto indifferenti.

DEFINIZIONE 13.3 Data una **soluzione** x^0 di un problema del tipo (13.1), un punto cioè che soddisfa la definizione 13.1, diremo che il **vincolo è attivo** in x^0 se il punto medesimo soddisfa almeno una delle disuguaglianze $g_i(x) \leq b_i$ [o $g_i(x) \geq b_i$] con l’uguaglianza, cioè se vale $g_i(x^0) = b_i$ per almeno un i . Viceversa, se la soluzione x^0 è tale che $g_i(x^0) < b_i$ [o $g_i(x^0) > b_i$] per ogni $i = 1, \dots, m$, allora diremo che il vincolo è **inattivo** in x^0 .

Il vincolo è dunque attivo se la funzione obiettivo in (13.1) raggiunge il suo valore massimo (minimo) su un punto di frontiera del vincolo, mentre è inattivo se il punto estremo è interno.

OSSERVAZIONE 13.3 Nel capitolo 12 abbiamo imparato che l'intersezione di m curve di livello in \mathbb{R}^n è un insieme di dimensione $n - m$. Ciò significa che in un vincolo definito da m disuguaglianze ci possono essere al più n disuguaglianze che valgono con uguaglianza; se $m > n$, un numero di uguaglianze maggiore di n definirebbe un insieme di dimensione negativa, che non esiste! A differenza dei problemi con vincoli di sola uguaglianza per i quali doveva essere $m \leq n - 1$ perché l'insieme ammissibile aveva dimensione $n - m$, ora che il vincolo ha dimensione n ci interessano anche i punti isolati (di dimensione 0) determinati dall'eventuale intersezione di n curve di livello perché l'ottimo potrebbe benissimo essere raggiunto in uno di questi punti di frontiera.

La prossima definizione tiene conto delle definizioni 12.2 e 12.5 nel caso di vincolo attivo.

DEFINIZIONE 13.4 (QUALIFICAZIONE DEL VINCOLO) Se il vincolo del problema in (13.1) è definito da una sola disequazione, $g(x) \leq b$ [o $g(x) \geq b$], diremo che si tratta di un **vincolo regolare** se è attivo e se la soluzione di (13.1) non contiene punti singolari (definizione 12.2).

In un problema del tipo (13.1) il cui vincolo è definito da m disequazioni, $g_i(x) \leq b_i$ [o $g_i(x) \geq b_i$], $i = 1, \dots, m$, diremo che tale **vincolo è regolare** se è attivo e se per ciascun punto x^0 che risolve (13.1) la matrice Jacobiana $J(x^0)$ ($k \times n$) formata dai k gradienti $\nabla g_j(x^0)$ corrispondenti alle disequazioni che valgono con uguaglianza, $g_j(x^0) = b_j$ per gli indici $j = 1, \dots, k$, ha rango pieno, cioè $\rho[J(x^0)] = k \leq n$, nel punto x^0 .

Se il vincolo è espresso mediante $m > 1$ disuguaglianze, la seconda parte della definizione 13.4 dice che i punti x^0 di intersezione fra le curve definite implicitamente dalle k equazioni $g_j(x^0) = b_j$, qualora x^0 sia un punto di ottimo, devono soddisfare la definizione 12.5. Ciò implica, in particolare, $\nabla g_j(x^0) \neq (0, 0, \dots, 0)$ per ogni $j = 1, \dots, k$, e $\nabla g_i(x^0) \neq \nabla g_j(x^0)$ per coppie di indici $i \neq j$ che individuano uguaglianze in x^0 , escludendo così che x^0 possa essere un punto stazionario di qualche funzione g_j coinvolta nelle uguaglianze, ovvero che tale punto possa essere un punto di tangenza fra due curve di livello distinte $g_i(x) = b_i$ e $g_j(x) = b_j$.

13.3 Punti di Kuhn-Tucker

Siamo finalmente in grado di studiare in dettaglio le soluzioni dei problemi vincolati in (13.1). Come abbiamo accennato all'inizio, il metodo per risolvere problemi con un vincolo che contiene sia punti interni che di frontiera dovrà presupporre l'utilizzo contemporaneo delle tecniche per problemi liberi (da usarsi per i punti estremi interni) e di quelle per problemi vincolati a curve (da usarsi sui punti estremi di frontiera). Sostanzialmente, ciò di cui abbiamo bisogno è un meccanismo che riesca a discriminare a priori queste due tipologie di punti estremi e che ci incanali in modo automatico verso lo studio dei punti interni, ovvero verso l'analisi dei punti di frontiera.

13.3.1 Vincoli definiti da una sola disuguaglianza

Supponiamo che il vincolo sia definito da un'unica disequazione. Se c'è un punto estremo di frontiera, la tentazione è quella di individuarlo come punto di tangenza fra la curva che esprime la frontiera del vincolo e una curva di livello della funzione obiettivo, ovvero definire la funzione Lagrangiana e applicare il Teorema 12.7. L'intuizione è giusta, però ora c'è una differenza fondamentale rispetto al caso con vincolo di uguaglianza: la regione ammissibile non è circoscritta ai soli

punti della curva (il bordo del vincolo) bensì una delle due superfici individuate (e separate) dalla curva è essa stessa ammissibile; pertanto, non siamo costretti a “muoverci” esclusivamente lungo la curva ma possiamo “abbandonare” la curva ed “entrare” in una delle due superfici rimanendo sempre nella regione ammissibile (il vincolo). Ciò introduce la possibilità che esistano punti di tangenza fra il bordo del vincolo e curve di livello della funzione obiettivo che non sono punti estremi.

Concentriamo la nostra attenzione su un *problema di massimo* del tipo (13.1) con un solo vincolo espresso da $g(x) \leq b$. Supponiamo che il massimo (assoluto) per la funzione obiettivo f sia raggiunto nel punto x^0 e che x^0 sia un punto di frontiera per il vincolo, cioè tale che $g(x^0) = b$. Queste informazioni sono sufficienti per garantire che x^0 è un punto di tangenza fra la curva di livello $c = f(x^0)$ della funzione obiettivo e la curva definita implicitamente da $g(x) = b$; è evidente infatti che, sotto l'ipotesi di qualificazione del vincolo (definizione 13.4), x^0 soddisfa le ipotesi del Teorema 12.7. Ma osserviamo più a fondo le caratteristiche di questo punto. In figura 13.5(a) si vede che il gradiente $\nabla g(x^0)$ punta verso l'esterno del vincolo, perché, e non è una coincidenza, il vincolo in un problema di massimo è rappresentato da insiemi di livello inferiore per le g . Trattandosi di un punto di tangenza, $\nabla f(x^0)$ e $\nabla g(x^0)$ sono paralleli, ovvero giacciono sulla stessa retta; rimane da stabilire il verso di $\nabla f(x^0)$. In linea di principio potrebbe coincidere con quello di ∇g oppure potrebbe essere il suo opposto. Ma ora abbiamo due informazioni in più: 1) sappiamo che l'area interna al vincolo è ammissibile e 2) x^0 è un punto di massimo per f ; ciò è sufficiente per escludere la possibilità che ∇f abbia verso opposto a ∇g . Se così fosse, infatti, dal momento che $\nabla f(x^0)$ individua la direzione di massima pendenza per f in uscita da x^0 , esisterebbero punti x interni al vincolo (e quindi ammissibili) vicini a x^0 tali che $f(x) > f(x^0)$, contraddicendo la definizione 13.1 valida per il punto x^0 . Dunque, $\nabla f(x^0)$ deve avere lo stesso verso di $\nabla g(x^0)$.

Gli eventuali punti di tangenza fra una curva di livello di f e bordo del vincolo sui quali i gradienti ∇f e ∇g puntano in direzioni opposte (∇g , come sempre, verso l'esterno del vincolo e ∇f verso l'interno del vincolo) non possono dunque essere punti di massimo per il problema (13.1). Poiché vogliamo escludere questi punti, alla usuale condizione di tangenza caratteristica di un punto di massimo di frontiera x^0 dobbiamo aggiungere la condizione che entrambi i gradienti $\nabla f(x^0)$ e $\nabla g(x^0)$ abbiano la medesima direzione di uscita da x^0 : quella verso l'“esterno” del vincolo. Rileggendo il Teorema 12.5 capiamo come imporre analiticamente questa condizione: dobbiamo richiedere che il moltiplicatore λ sia positivo; questo non solo assicura che il rapporto fra le derivate parziali di f e di g in x^0 sia lo stesso, ma anche che i due gradienti abbiano lo stesso verso.

Appare chiaro, finalmente, il motivo per cui i vincoli dei problemi di massimo debbano essere espressi con disuguaglianze di tipo ‘ \leq ’. Per lo stesso motivo i vincoli dei problemi di minimo devono essere definiti da disuguaglianze di tipo ‘ \geq ’: così facendo, anche in questo caso i punti di minimo di frontiera hanno i gradienti ∇f e ∇g che puntano nella stessa direzione, questa volta verso l'“interno” del vincolo, e dunque il moltiplicatore λ è ancora positivo. Punti di tangenza sui quali i gradienti ∇f e ∇g puntano in direzioni opposte – caratterizzati da un moltiplicatore λ negativo – non possono essere né di massimo né di minimo.

Abbiamo capito che non possiamo esimerci dallo studiare la funzione Lagrangiana introdotta nel paragrafo 12.2.5 come strumento deputato all'analisi dei punti di tangenza:

$$L(x, \lambda) = f(x) + \lambda [b - g(x)]. \quad (13.2)$$

C'è un'importante novità rispetto ai problemi con vincoli di uguaglianza: ora il moltiplicatore λ dev'essere positivo (o meglio, non negativo, come vedremo tra breve), sia che si tratti di problemi di massimo che di minimo, mentre prima era libero di assumere valori sia positivi che negativi poiché il verso dei gradienti ∇f e ∇g non forniva alcuna informazione rilevante.

Prima di avventarci a calcolare le condizioni del primo ordine per L con i metodi del capitolo 12, però, dobbiamo ricordarci che esiste la possibilità che la soluzione di (13.1) sia un punto interno, per il quale la pomposità della funzione L risulta senz'altro ridondante. Osserviamo che sui punti estremi interni il vincolo è *inattivo*, il che significa che possiamo tranquillamente fare a meno della curva definita implicitamente da $g(x) = b$, essa non racconta nulla di utile. Ancora una volta possiamo manovrare l'esclusione di suddetta curva attraverso la 'leva' fornita dal moltiplicatore λ : se il vincolo è inattivo, basta porre $\lambda = 0$ in (13.2), e le condizioni del primo ordine per L si tramutano miracolosamente nelle condizioni del primo ordine per la sola funzione obiettivo f , ovvero ci troviamo di fronte un problema di ottimo libero.

Come riformulare il Teorema 12.7 in questa nuova situazione? Da un lato, annullare le prime n derivate parziali della funzione Lagrangiana, $L_{x_i} = f_{x_i} - \lambda g_{x_i} = 0$ per $i = 1, 2, \dots, n$, ha ancora senso: queste condizioni stabiliscono, correttamente, che il punto estremo è un punto di tangenza se il vincolo è attivo, mentre si riducono alle condizioni del primo ordine per ottimo libero, $f_x = 0$ e $f_y = 0$, se il vincolo è inattivo ($\lambda = 0$). Più problematica è la $n + 1$ -esima condizione: derivando L rispetto a λ nel caso di vincolo di uguaglianza si otteneva semplicemente il vincolo stesso, $b - g(x) = 0$, che certamente doveva essere soddisfatto. Ora il vincolo è definito da una disuguaglianza, $g(x) \leq b$, anziché da un'uguaglianza, e pertanto non possiamo imporre a priori la condizione $b - g(x) = 0$ che escluderebbe automaticamente tutti i casi in cui il punto di ottimo x^0 è interno e soddisfa il vincolo con disuguaglianza stretta, $g(x^0) < b$. Non possiamo più limitarci ad annullare la derivata parziale L_λ .

Sostanzialmente ci sono due possibilità: o il vincolo è attivo e vale la condizione $b - g(x^0) = 0$ associata a un moltiplicatore $\lambda \geq 0$, oppure il vincolo è inattivo e vale $b - g(x^0) > 0$, nel qual caso dev'essere $\lambda = 0$. In altre parole, abbiamo due termini, $b - g(x^0)$ e λ , dei quali almeno uno (o entrambi, anche se si tratta di una circostanza piuttosto rara) deve sempre annullarsi; deve valere, cioè, $b - g(x^0) = 0$ oppure $\lambda = 0$. Un modo naturale per imporre che almeno uno di due termini sia nullo è richiedere che il loro prodotto sia sempre nullo:

$$\lambda [b - g(x^0)] = 0. \quad (13.3)$$

La (13.3) si chiama **condizione di complementarità**; essa presuppone almeno una delle due uguaglianze: $\lambda = 0$ (quando x^0 è interno) oppure $b - g(x^0) = 0$ (quando x^0 è di frontiera).

La figura 13.6 presenta un ricco panorama di situazioni possibili per un problema di *massimo vincolato in due variabili*. La figura 13.6(a) riproduce il grafico di una funzione obiettivo $f(x, y)$ vincolata a un rettangolo¹ V sul quale sono riportate alcune curve di livello per f , prolungate anche fuori dal vincolo. Sul vincolo sono evidenti cinque punti interessanti² contrassegnati dalle lettere A, B, C, D e E . In corrispondenza dei quattro punti che stanno sulla frontiera del vincolo, A, C, D e E , sono evidenziate in grassetto le curve di livello di f tangenti alla frontiera del vincolo in quei punti, sia sul piano \mathbb{R}^2 che 'in quota', sul grafico di f . Inoltre, agli stessi punti vengono "applicati" i gradienti del vincolo (le frecce in grigio indicate con ∇g) e i gradienti di f (le frecce in nero indicate con ∇f che indicano la direzione di massima pendenza di f in uscita da ciascun punto e sono riportate anche sul grafico di f). La figura 13.6(b) mostra il dettaglio dei cinque punti sul piano \mathbb{R}^2 , sul quale, oltre al vincolo V , sono riportate anche alcune curve di livello di f , in particolare quelle tangenti ai punti critici evidenziate in grassetto.

¹Dall'esercizio 13.1 sappiamo che un vincolo a forma di rettangolo necessita di quattro disequazioni del tipo $-x \leq b_1$, $x \leq b_2$, $-y \leq b_3$, $y \leq b_4$, anziché una sola disequazione, ma la figura ha uno scopo puramente illustrativo e quindi non sottilizziamo, dal momento che l'idea descritta in precedenza si applica a qualsiasi tipo di vincolo.

²In verità ce ne sono altri due non evidenziati nella figura; si invita il lettore a individuarli.

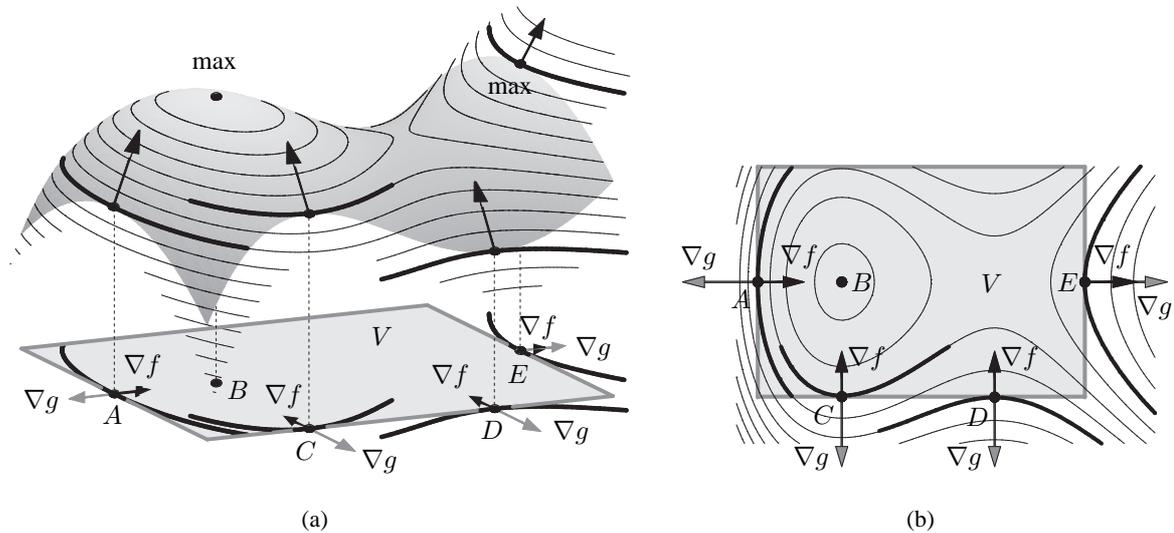


FIGURA 13.6: (a) i punti B ed E sono le due soluzioni, interna e di frontiera rispettivamente, del problema vincolato mentre A, C e D sono punti di tangenza fra curve di livello di f e bordo del vincolo che non sono punti estremi; (b) dettaglio sul piano \mathbb{R}^2 di tutti i punti critici in (a).

I punti B ed E costituiscono la soluzione del problema di massimo vincolato, il primo è un punto interno sul quale la f vincolata a V assume il valore massimo, e il secondo è un punto di frontiera sul quale nuovamente la f raggiunge il valore massimo. Per quanto riguarda il punto B , valgono semplicemente le condizioni del primo ordine per la f , poiché lì il vincolo è inattivo. Il punto E , diversamente, è un punto sul quale il vincolo è attivo: si tratta di un punto di tangenza fra il bordo del vincolo e una curva di livello di f ; in esso il gradiente ∇g del vincolo e il gradiente ∇f della funzione obiettivo non solo sono paralleli, ma puntano entrambi nella stessa direzione verso l'esterno del vincolo, come si vede nella parte destra della figura 13.6(b). Gli altri tre punti di tangenza, A, C e D , viceversa, pur essendo tutti punti di tangenza fra il bordo del vincolo e curve di livello di f , non sono affatto punti di massimo vincolato per f ; infatti si tratta di punti caratterizzati da gradienti ∇g e ∇f che puntano in direzioni opposte. In particolare, come sempre i ∇g puntano verso l'esterno del vincolo, mentre i ∇f puntano verso l'interno del vincolo, direzione verso la quale f cresce più rapidamente a partire da ciascun punto.

TEOREMA 13.3 È dato il *problema di massimo* (13.1) con il vincolo espresso da un'unica *disequazione*, $g(x) \leq b$. Supponiamo che entrambe la funzione obiettivo f e la funzione g siano differenziabili e x^0 sia un *punto di massimo* assoluto per (13.1) che, nell'ipotesi che il vincolo sia attivo, soddisfi la *qualificazione del vincolo* (definizione 13.4), valendo $\nabla g(x^0) \neq (0, 0, \dots, 0)$. Allora esiste un *moltiplicatore* $\lambda \in \mathbb{R}$ per la funzione Lagrangiana $L(x, \lambda) = f(x) + \lambda[b - g(x)]$ che soddisfa le seguenti condizioni:

$$\begin{cases} L_{x_1}(x^0, \lambda) = f_{x_1}(x^0) - \lambda g_{x_1}(x^0) = 0 \\ \vdots \\ L_{x_n}(x^0, \lambda) = f_{x_n}(x^0) - \lambda g_{x_n}(x^0) = 0 \\ \lambda L_\lambda(x^0, \lambda) = \lambda [b - g(x^0)] = 0 \\ L_\lambda(x^0, \lambda) = b - g(x^0) \geq 0 \\ \lambda \geq 0. \end{cases} \quad (13.4)$$

Le prime n condizioni in (13.4) coincidono con le prime n condizioni del Teorema di Lagrange per vincoli sottili; si noti che esse si riducono alle condizioni del primo ordine per ottimo libero se il vincolo è inattivo ($\lambda = 0$), mentre stabiliscono la nota proprietà di tangenza se il vincolo è attivo ($\lambda > 0$). Riconosciamo la condizione di complementarietà (13.3) nella $n + 1$ -esima condizione, scritta come *prodotto fra il moltiplicatore λ e la derivata parziale della funzione Lagrangiana rispetto a λ uguale a zero*, $\lambda L_\lambda = 0$, per facilitarne la memorizzazione. La penultima condizione esprime il vincolo, $g(x) \leq b$, scritto come derivata parziale della funzione Lagrangiana rispetto a λ maggiore o uguale a zero, $L_\lambda \geq 0$, per facilitarne la memorizzazione. Infine, l'ultima condizione serve a escludere automaticamente punti di tangenza che non sono punti di massimo per f . Si noti che la qualificazione del vincolo deve valere solamente nel caso in cui il vincolo sia attivo.

Il teorema equivalente per problemi di minimo con il vincolo definito da una disuguaglianza di tipo $g(x) \geq b$, come nel problema di destra in (13.1) è sostanzialmente identico al Teorema 13.3 con la disuguaglianza $b - g(x^0) \leq 0$ al posto della disuguaglianza $b - g(x^0) \geq 0$ nella penultima condizione (attenzione: l'ultima condizione rimane sempre la stessa, $\lambda \geq 0$).

NOTAZIONE 13.1 I punti x^0 che soddisfano il Teorema 13.3 si dicono **punti di Kuhn-Tucker**, in breve, punti di K.T..

Poiché il Teorema 13.3 stabilisce *condizioni solo necessarie*, possiamo pensare ai punti di Kuhn-Tucker come alla generalizzazione a problemi con vincoli di disuguaglianza dei *punti stazionari* nei problemi di ottimo libero e con vincolo di uguaglianza. Tali condizioni, però, anche senza essere sufficienti, sono comunque molto potenti perché sono in grado di escludere tutti i punti di minimo di frontiera, sia relativi che assoluti, e alcuni punti di tangenza che non sono né massimi né minimi. Ad esempio, l'insieme dei punti di K.T. per un problema di massimo contiene tutti i massimi, assoluti e relativi, interni e di frontiera, i minimi interni, assoluti e relativi, e le selle, oltre ad alcuni punti di tangenza che non sono né massimi né minimi. Questo teorema risolve completamente i *problemi con vincolo compatto*, selezionando, come sempre, il valore più alto della funzione obiettivo calcolato su tutti i punti di K.T., e i *problemi concavi*, quelli cioè con vincolo convesso e funzione obiettivo concava, per i quali se esiste un punto di K.T. esso è necessariamente la soluzione.

13.3.2 Vincoli definiti da m disuguaglianze

Enunciamo il teorema generale per problemi con vincolo definito da più disequazioni. Come per i vincoli di uguaglianza, attribuiamo un moltiplicatore λ_i distinto a ciascuna delle m disequazioni $g_i(x, y) \leq b_i$ [o $g_i(x, y) \geq b_i$], per $i = 1, \dots, m$, che definiscono il vincolo e consideriamo la funzione Lagrangiana generalizzata (12.21), $L : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, che qui riportiamo:

$$L(x, \lambda) = f(x) + \lambda_1 [b_1 - g_1(x)] + \lambda_2 [b_2 - g_2(x)] + \dots + \lambda_m [b_m - g_m(x)]. \quad (13.5)$$

TEOREMA 13.4 È dato il problema di massimo (13.1) con vincolo espresso da m disequazioni, $g_i(x, y) \leq b_i$, per $i = 1, \dots, m$. Supponiamo che la funzione obiettivo f e tutte le funzioni g_i siano differenziabili e x^0 sia un *punto di massimo* assoluto per (13.1). Supponiamo inoltre che, nell'ipotesi che il vincolo sia attivo, x^0 soddisfi la *qualificazione del vincolo* (definizione 13.4), ovvero che la matrice Jacobiana $J(x^0)$ ($k \times n$) formata dai k gradienti $\nabla g_j(x^0)$ corrispondenti alle disequazioni che valgono con uguaglianza, $g_j(x^0) = b_j$ per gli indici $j = 1, \dots, k$, abbia *rango pieno*, $\rho[J(x^0)] = k \leq n$, nel punto x^0 . Allora esistono m *moltiplicatori* $\lambda_1, \dots, \lambda_m$ per la funzione Lagrangiana (13.5) che soddisfano le seguenti $n + 3m$ condizioni:

1. $L_{x_i}(x^0, \lambda) = f_{x_i}(x^0) - \lambda_1 (\partial/\partial x_i) g_1(x^0) - \lambda_2 (\partial/\partial x_i) g_2(x^0) - \dots - \lambda_m (\partial/\partial x_i) g_m(x^0) = 0$,
per $i = 1, 2, \dots, n$,
2. $\lambda_j L_{\lambda_j}(x^0, \lambda) = \lambda_j [b_j - g_j(x^0)] = 0$, per $j = 1, 2, \dots, m$,
3. $L_{\lambda_j}(x^0, \lambda) = b_j - g_j(x^0) \geq 0$, per $j = 1, 2, \dots, m$,
4. $\lambda_j \geq 0$, per $j = 1, 2, \dots, m$.

Il Teorema 13.4 si limita a estendere le condizioni del Teorema 13.3 a tutti i moltiplicatori relativi a ciascuna disequazione che definisce il vincolo. Anche in questo caso la qualificazione del vincolo deve valere solamente nel caso in cui il vincolo sia attivo.

Il teorema equivalente per problemi di minimo con vincolo definito da più disuguaglianze di tipo $g_j(x) \leq b_j$, per $j = 1, 2, \dots, m$, come nel problema di destra in (13.1), è identico al Teorema 13.4 tranne per le disuguaglianze $b_j - g_j(x^0) \geq 0$ nella terza serie di condizioni che vanno invertite: $b_j - g_j(x^0) \leq 0$, per $j = 1, 2, \dots, m$ (attenzione: le ultime m condizioni rimangono sempre le stesse, $\lambda_j \geq 0$, per $j = 1, 2, \dots, m$).

13.3.3 Un esempio in una sola variabile

È istruttivo illustrare il ruolo giocato dalla non negatività del moltiplicatore λ nel distinguere i punti di massimo sulla frontiera da altri punti ‘critici’ di frontiera in un problema univariato. Anche se la struttura del Teorema 13.3 appare ridondante per risolvere un problema così semplice, questo esercizio fornisce l’intuizione elementare su cui si basa il teorema stesso.

Nel paragrafo 6.5 avevamo definito il vincolo come intervallo o un’unione di intervalli. Ad esempio, il problema

$$\begin{aligned} & \max f(x) \\ & \text{sub } x \in [a, b], \end{aligned} \tag{13.6}$$

ha come vincolo l’intervallo chiuso $[a, b]$. Per applicare il Teorema 13.3 a (13.6) dobbiamo esprimere il vincolo $[a, b]$ mediante una disequazione $g(x) \leq 0$, con $g : \mathbb{R} \rightarrow \mathbb{R}$. Innanzitutto poniamo $b = 0$ in modo da non creare confusione con l’estremo destro dell’intervallo $[a, b]$, poi scegliamo una funzione $g(x)$ tale che $g(x) \leq 0$ per ogni $x \in [a, b]$ e $g(x) > 0$ se $x \notin [a, b]$. Possiamo sempre scegliere una funzione g strettamente convessa, cosicché è sufficiente che valga $g(a) = g(b) = 0$. La figura 13.7 mostra una funzione strettamente convessa g che interseca l’asse delle ascisse nei due punti a e b : l’insieme definito da $g(x) \leq 0$ è proprio l’intervallo chiuso $[a, b]$, in grigio sull’asse delle ascisse. Pertanto il problema (13.6) è equivalente a

$$\begin{aligned} & \max f(x) \\ & \text{sub } g(x) \leq 0, \end{aligned} \tag{13.7}$$

scritto in forma standard come in (13.1). La figura 13.7 riporta anche la funzione obiettivo f di (13.6) e quindi illustra completamente il problema impostato come in (13.7).

La frontiera del vincolo $[a, b]$ in (13.7) è costituita da due soli punti, a e b , che costituiscono un insieme di dimensione 0 di \mathbb{R} . Per risolvere il problema è sufficiente confrontare i valori di f su tali punti con quelli assunti sugli altri punti critici interni all’intervallo, senza scomodare le condizioni previste dal Teorema 13.3. Cionondimeno, il suddetto teorema vale senz’altro anche in questo caso, anche se la condizione di ‘parallelismo fra gradienti’

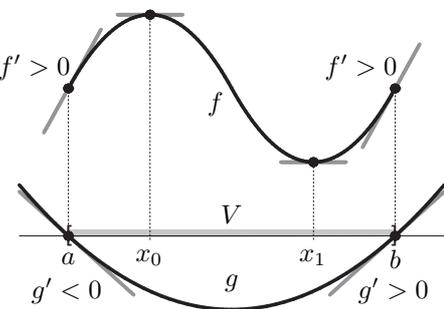


FIGURA 13.7: punti di K.T. in problemi con una sola variabile.

espressa dalle prime n condizioni del teorema perde significato (non ha senso parlare di “punti paralleli”). La funzione Lagrangiana è $L(x, \lambda) = f(x) - \lambda g(x)$, da cui, applicando il Teorema 13.3, otteniamo le seguenti quattro condizioni: $f'(x_0) = \lambda g'(x_0)$, $\lambda g(x_0) = 0$, $-g(x_0) \geq 0$ e $\lambda \geq 0$ valide per tutti i punti di K.T. x_0 . La prima dice che in x_0 la derivata della funzione obiettivo f è pari alla derivata della funzione g (che rappresenta il vincolo) moltiplicata per λ ; essa esprime il fatto che in un punto di K.T. i gradienti di f e di g sono paralleli e sono l'uno un multiplo (o una frazione) dell'altro. Questo significa che essi giacciono sulla stessa retta, che in questo caso è l'unica retta possibile: la retta reale. Le altre condizioni sono le solite.

La figura 13.7 mette in evidenza quattro punti critici di un problema in una variabile: i due punti di frontiera a e b e i due punti stazionari (interni) x_0 e x_1 . Il Teorema 13.3 ci permette di escludere a priori il punto a , dal momento che in a le derivate $f'(a)$ e $g'(a)$ hanno segno opposto e la ‘condizione di parallelismo’ $f'(x_0) = \lambda g'(x_0)$ implica $\lambda < 0$, contraddicendo l'ultima condizione del teorema. I rimanenti punti, b , x_0 e x_1 , sono senz'altro punti di K.T. poiché soddisfano tutte le condizioni (13.4) (il punto b si comporta come il punto E nella figura 13.6, mentre x_0 e x_1 presuppongono $\lambda = 0$, trattandosi di punti interni stazionari). Di questi punti, b è un punto di massimo relativo di frontiera, x_0 è il punto di massimo assoluto interno e x_1 è il punto di minimo assoluto interno, che non ci interessa. Il vantaggio di usare il Teorema 13.3 nell'esempio illustrato dalla figura 13.7 consiste nell'avere un punto critico in meno da confrontare con gli altri punti critici.

Anche nei problemi univariati, dunque, lo studio dei punti di K.T. riduce il numero di punti critici da analizzare, anche se certamente non vale la pena scomodare il Teorema 13.3 in questi casi.

ESEMPIO 13.4 Risolviamo l'esempio 6.11 del paragrafo 6.5 utilizzando il Teorema 13.3:

$$\begin{aligned} \max & (x^2 + x - 1) \\ \text{sub } & x \in [-1, 1]. \end{aligned}$$

Per studiare i punti di K.T. dobbiamo riscrivere il vincolo in forma standard. Una funzione g utile a questo scopo si ottiene trasformando la parabola x^2 in modo da farle intersecare l'asse delle ascisse nei punti -1 e 1 : la funzione $g(x) = x^2 - 1$, ad esempio, fa al caso nostro. Si noti che si tratta di una funzione strettamente convessa. Possiamo quindi riscrivere il problema come

$$\begin{aligned} \max & (x^2 + x - 1) \\ \text{sub } & x^2 - 1 \leq 0. \end{aligned}$$

La funzione Lagrangiana è $L(x, \lambda) = x^2 + x - 1 + \lambda(1 - x^2)$ e le condizioni (13.4) sono

$$\begin{cases} L_x = 2x + 1 - 2\lambda x = 0 \\ \lambda L_\lambda = \lambda(1 - x^2) = 0 \\ L_\lambda = 1 - x^2 \geq 0 \\ \lambda \geq 0. \end{cases}$$

Consideriamo separatamente i due casi: 1) $\lambda = 0$, che corrisponde alla ricerca di punti estremi interni al vincolo, cioè appartenenti all'intervallo aperto $(-1, 1)$, e 2) $\lambda > 0$, che riguarda lo studio dei due punti di frontiera -1 e 1 .

1. Sostituendo $\lambda = 0$ nella prima equazione otteniamo il punto interno $x_0 = -1/2 \in (-1, 1)$, che è l'unico punto stazionario per $f(x) = x^2 + x - 1$; poiché la coppia $x_0 = -1/2$ e $\lambda = 0$ soddisfa le altre condizioni, si tratta senz'altro di un punto (interno) di K.T..
2. Se $\lambda > 0$, dalla seconda equazione otteniamo i due punti $x_1 = -1$ e $x_2 = 1$ (la frontiera del vincolo). Per $x_1 = -1$, dalla prima equazione otteniamo $\lambda = 1/2 > 0$, e deduciamo che la

coppia $x_1 = -1$ e $\lambda = 1/2$ è un altro punto di K.T. poiché essa soddisfa tutte le condizioni. Infine, per $x_2 = 1$, dalla prima equazione ricaviamo $\lambda = 3/2 > 0$ e deduciamo quindi che anche la coppia $x_2 = 1$ e $\lambda = 3/2$ è un punto di K.T.

Calcolando i valori di f in corrispondenza dei tre punti di K.T. trovati vediamo facilmente che il massimo assoluto è raggiunto in, $x_2 = 1$, a conferma di quanto avevamo già trovato nell'esempio 6.11. Si invita il lettore ad aggiungere il grafico di $g(x) = x^2 - 1$ alla figura 6.15 e di trarre le conclusioni appropriate analoghe a quelle della figura 13.7.

13.3.4 Il metodo e alcuni esempi

REGOLA 13.1 Consideriamo un *problema di massimo* vincolato del tipo (13.1).

1. Verificare le ipotesi del Teorema 13.3 o del Teorema 13.4: f e g devono essere funzioni di n variabili *differenziabili* e deve valere la *qualificazione del vincolo* (definizione 13.4).

- (a) Se il vincolo è definito da una sola disequazione, $g(x) \leq b$, accertarsi che la sua frontiera sia regolare, cioè che $\nabla g(x) \neq (0, 0, \dots, 0)$ su tutti gli x tali che $g(x) = b$.
- (b) Se il vincolo è formato da m disequazioni, $g_i(x) \leq b_i, i = 1, \dots, m$, con $m \geq 2$, una condizione sufficiente per la qualificazione del vincolo è che siano *regolari sia i singoli tratti di curve di livello* $g_i(x) = b_i$ appartenenti alla frontiera del vincolo, *che le curve (o i punti) formate(i) dall'intersezione di più curve di livello* $g_i(x) = b_i$, tenendo presente che al più $\min(m, n)$ curve possono intersecarsi sulla frontiera.³

Operativamente, conviene scrivere la *matrice Jacobiana* $J(m \times n)$ formata dagli m gradienti ∇g_i e, per $1 \leq k \leq \min(m, n)$, verificare che tutte le *sottomatrici* J_k di J di dimensione $k \times n$ *abbiano rango pieno*, $\rho[J_k(x)] = k$, *su tutti i punti di frontiera x appartenenti all'intersezione delle k curve di livello*⁴ $g_j(x) = b_j, j = 1, \dots, k$.⁵

I punti singolari che non soddisfano queste condizioni vanno aggiunti alla lista di punti critici.

2. Scrivere la funzione Lagrangiana di $n + m$ variabili:

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, \dots, x_n) + \lambda_1 [b_1 - g_1(x_1, \dots, x_n)] + \lambda_2 [b_2 - g_2(x_1, \dots, x_n)] + \dots + \lambda_m [b_m - g_m(x_1, \dots, x_n)]$$

e risolvere il sistema di $n + 3m$ condizioni previste dal Teorema 13.4,

$$\left\{ \begin{array}{l} L_{x_1} = f_{x_1} - \lambda_1 (\partial/\partial x_1) g_1 - \lambda_2 (\partial/\partial x_1) g_2 - \dots - \lambda_m (\partial/\partial x_1) g_m = 0 \\ \vdots \\ L_{x_n} = f_{x_n} - \lambda_1 (\partial/\partial x_n) g_1 - \lambda_2 (\partial/\partial x_n) g_2 - \dots - \lambda_m (\partial/\partial x_n) g_m = 0 \\ \lambda_1 L_{\lambda_1} = \lambda_1 (b_1 - g_1) = 0 \\ \vdots \\ \lambda_m L_{\lambda_m} = \lambda_m (b_m - g_m) = 0 \\ L_{\lambda_1} = b_1 - g_1 \geq 0 \\ \vdots \\ L_{\lambda_m} = b_m - g_m \geq 0 \\ \lambda_1 \geq 0, \lambda_2 \geq 0, \dots, \lambda_m \geq 0, \end{array} \right. \quad (13.8)$$

rispetto alle $m + n$ variabili $x_1, \dots, x_n, \lambda_1, \dots, \lambda_m$, omesse nel sistema (13.8) per semplificarne la scrittura. Il (13.8) è un sistema misto di equazioni e disequazioni. La tecnica

³Si ricordi che al più n disuguaglianze $g_i(x) \leq b_i$ possono valere con uguaglianza (osservazione 13.3).

⁴Poiché non tutte le sottomatrici J_k corrispondono a punti di intersezione fra curve di livello, non è necessario che $\rho(J_k) = k$ valga per ciascuna matrice J_k , come vedremo nell'esempio 13.7.

⁵Per $k = 1$ ciò significa verificare che *ciascuna riga di* $J(m \times n)$ *non sia identicamente nulla* sui punti di frontiera del vincolo, ovvero che $\nabla g_i(x) \neq (0, 0, \dots, 0)$ su tutti i punti x di frontiera tali che $g_i(x) = b_i$, per ogni $i = 1, \dots, m$.

per risolverlo si basa sullo studio separato di tutti i casi possibili individuati dalle ultime m disequazioni: ciascuna va considerata prima come uguaglianza e poi come disuguaglianza stretta; in ciascun caso si studiano tutte le implicazioni per le altre equazioni e disequazioni fino ad arrivare a una contraddizione oppure ottenere un punto di K.T.. Questa procedura, di fatto, significa analizzare separatamente i punti interni e i punti di frontiera del vincolo.

3. Applicare uno dei due approcci che abbiamo a disposizione.

- (a) Se il *vincolo* è *compatto*, calcolare il valore della funzione obiettivo f sulle prime n variabili, x_1^0, \dots, x_n^0 , dei punti di K.T. $(x^0, \lambda) = (x_1^0, \dots, x_n^0, \lambda_1, \dots, \lambda_m)$ trovati al punto precedente e sugli eventuali punti singolari. Il *massimo assoluto* è il punto su cui f assume il valore più elevato. Si noti che per verificare la compattezza del vincolo è sufficiente accertarsi che si tratta di un *insieme limitato*, dal momento che è sempre un insieme chiuso in quanto definito da disuguaglianze deboli.
- (b) Se il *vincolo* è *convesso*, verificare se la funzione obiettivo f è concava con gli strumenti del paragrafo 11.6; se f è concava ed esiste un punto di K.T. (necessariamente unico se f è strettamente concava) che risolve il sistema (13.8), esso è la soluzione cercata.

La regola per risolvere problemi di minimo vincolato è simile: basta invertire le disuguaglianze che definiscono il vincolo [il penultimo gruppo di m disuguaglianze in (13.8)]; le disuguaglianze per i moltiplicatori nell'ultima riga di (13.8) rimangono le stesse: $\lambda_1 \geq 0, \dots, \lambda_m \geq 0$.

ESEMPIO 13.5 Risolviamo il problema vincolato

$$\begin{aligned} & \max(xy + 4) \\ & \text{sub } (1/4)x^2 + y^2 \leq 1. \end{aligned}$$

Il *vincolo* è un'ellisse con centro l'origine, inclusa la frontiera, e quindi è un insieme sia *compatto* che *convesso*. La funzione obiettivo però non è concava poiché, come si verifica immediatamente, ha matrice Hessiana ovunque indefinita. Non valendo il Teorema 11.6, dobbiamo per forza seguire l'approccio basato sul Teorema di Weierstrass. Vale la *qualificazione del vincolo* (caso 1a della regola 13.1) perché $\nabla g(x, y) = ((1/2)x, 2y) \neq (0, 0)$ per ogni punto (x, y) che sta sul bordo dell'ellisse. La *funzione Lagrangiana* è $L(x, y, \lambda) = xy + 4 + \lambda [1 - (1/4)x^2 - y^2]$. Dobbiamo risolvere il sistema

$$\begin{cases} L_x = y - (1/2)\lambda x = 0 \\ L_y = x - 2\lambda y = 0 \\ \lambda L_\lambda = \lambda [1 - (1/4)x^2 - y^2] = 0 \\ L_\lambda = 1 - (1/4)x^2 - y^2 \geq 0 \\ \lambda \geq 0. \end{cases}$$

Poiché il vincolo è definito da un'unica disequazione, lo studio del sistema presuppone due casi distinti: 1) $\lambda = 0$, che corrisponde alla ricerca di punti estremi interni, per i quali vale $(1/4)x^2 + y^2 < 1$, e 2) $\lambda > 0$, che riguarda lo studio dei punti di frontiera, per i quali vale $(1/4)x^2 + y^2 = 1$.

1. Sostituendo $\lambda = 0$ nelle prime due equazioni otteniamo il punto interno al vincolo $(0, 0)$, che è un punto stazionario per la funzione obiettivo $f(x, y) = xy$. Il punto $P_1 = (0, 0, 0)$ soddisfa banalmente la terza equazione (perché $\lambda = 0$) e la quarta disequazione con disuguaglianza stretta; deduciamo, senza alcun dubbio, che si tratta di un punto (interno) di K.T..
2. Se $\lambda > 0$, dalla terza equazione deduciamo che necessariamente

$$1 - (1/4)x^2 - y^2 = 0, \tag{13.9}$$

ovvero stiamo studiando i punti di frontiera del vincolo. Dalle prime due equazioni ricaviamo $y = (1/2)\lambda x$ e $y = x/(2\lambda)$ rispettivamente (si ricordi che $\lambda \neq 0$), e quindi $(1/2)\lambda x = x/(2\lambda)$, ovvero $\lambda^2 x = x$. Osserviamo che se $x = 0$, la prima equazione implica che anche $y = 0$,

contraddicendo la (13.9), pertanto dev'essere necessariamente $x \neq 0$ e possiamo scrivere $\lambda^2 = 1$, da cui otteniamo due valori per il moltiplicatore, $\lambda = -1$ e $\lambda = 1$. Ovviamente scartiamo il primo, dal momento che dev'essere $\lambda > 0$. Sostituendo $\lambda = 1$ nella prima equazione otteniamo $y = (1/2)x$, che, a sua volta sostituito in (13.9), porta ai due valori $x = -\sqrt{2}$ e $x = \sqrt{2}$, ciascuno associato a $y = -1/\sqrt{2}$ e $y = 1/\sqrt{2}$ rispettivamente. In conclusione, abbiamo trovato due punti di K.T.: $P_2 = (-\sqrt{2}, -1/\sqrt{2}, 1)$ e $P_3 = (\sqrt{2}, 1/\sqrt{2}, 1)$.

I valori di f in corrispondenza dei tre punti di K.T. trovati sono $f(0, 0) = 4$ e $f(-\sqrt{2}, -1/\sqrt{2}) = f(\sqrt{2}, 1/\sqrt{2}) = 5$, da cui deduciamo che P_2 e P_3 sono due punti distinti di massimo assoluto; si tratta di due punti in cui il vincolo è attivo e non ci sono punti di massimo assoluto interni. La figura 13.8(a) illustra la soluzione; si noti che P_1 è un punto (interno) di sella, mentre $P_4 = (-\sqrt{2}, 1/\sqrt{2})$ e $P_5 = (\sqrt{2}, -1/\sqrt{2})$ sono due punti distinti di minimo assoluto che vengono automaticamente esclusi dal fatto che in corrispondenza di essi il moltiplicatore λ è negativo.

ESEMPIO 13.6 Massimizziamo $f(x, y) = \ln x + \ln y$ soggetta al vincolo definito da $xy \geq 1$ e $x^2 + y^2 \leq 4$. Scrivendo la prima disequazione come $-xy \leq -1$ il problema in forma standard è

$$\begin{aligned} & \max (\ln x + \ln y) \\ \text{sub } & \begin{cases} -xy \leq -1 \\ x^2 + y^2 \leq 4. \end{cases} \end{aligned}$$

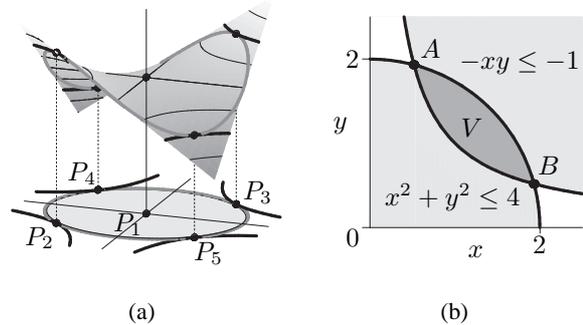


FIGURA 13.8: (a) soluzione dell'esempio 13.5; (b) vincolo dell'esempio 13.6.

Il vincolo è formato dall'intersezione fra l'insieme LI di livello -1 della funzione $-xy$ e il cerchio di raggio 2 con centro l'origine (l'insieme LI di livello 4 di $x^2 + y^2$). L'insieme ammissibile è dunque l'intersezione fra il dominio naturale della funzione obiettivo, \mathbb{R}_{++}^2 , e il vincolo; la figura 13.8(b) mostra che si tratta di un insieme sia compatto

che convesso. La funzione obiettivo è *concava* in quanto somma di due funzioni strettamente concave in una variabile (Proposizione 11.2). Poiché vale il Teorema di Weierstrass, ci sarà un unico punto di K.T. che sarà la soluzione cercata. Siamo nel caso 1b della regola 13.1 con $\min(m, n) = m = n = 2$. Per $k = 1$ entrambe le righe della matrice Jacobiana, $J = \begin{bmatrix} -y & -x \\ 2x & 2y \end{bmatrix}$, sono non nulle poiché $x \neq 0$ e $y \neq 0$ su tutto il vincolo. Per $k = 2$ J non ha rango pieno (è, cioè, singolare), $\rho(J) < 2$, solamente se $\det(J) = -2y^2 + 2x^2 = 0$, ovvero se $y = x$; ma sostituendo $y = x$ nel sistema di equazioni che definisce la frontiera del vincolo, $\begin{cases} -xy = -1 \\ x^2 + y^2 = 4 \end{cases}$, otterremmo $\begin{cases} x^2 = 1 \\ x^2 = 2 \end{cases}$, che è impossibile. Pertanto, $y \neq x$ sui due punti d'intersezione fra le curve di livello [A e B in figura 13.8(b)] e vale la *qualificazione del vincolo* poiché su di essi J ha rango pieno, $\rho(J) = 2$.

La funzione Lagrangiana è $L(x, y, \lambda_1, \lambda_2) = \ln x + \ln y + \lambda_1(-1 + xy) + \lambda_2(4 - x^2 - y^2)$. Dobbiamo risolvere il sistema

$$\begin{cases} L_x = 1/x + \lambda_1 y - 2\lambda_2 x = 0 \\ L_y = 1/y + \lambda_1 x - 2\lambda_2 y = 0 \\ \lambda_1 L_{\lambda_1} = \lambda_1(xy - 1) = 0 \\ \lambda_2 L_{\lambda_2} = \lambda_2(4 - x^2 - y^2) = 0 \\ L_{\lambda_1} = xy - 1 \geq 0 \\ L_{\lambda_2} = 4 - x^2 - y^2 \geq 0 \\ \lambda_1 \geq 0, \lambda_2 \geq 0. \end{cases} \quad (13.10)$$

Ci sono quattro possibilità: 1(a) $\lambda_1 = 0$ e $\lambda_2 = 0$, 1(b) $\lambda_1 = 0$ e $\lambda_2 > 0$, 2(a) $\lambda_1 > 0$ e $\lambda_2 = 0$, 2(b) $\lambda_1 > 0$ e $\lambda_2 > 0$.

1. Se $\lambda_1 = 0$, le prime due equazioni diventano

$$\begin{cases} 1/x - 2\lambda_2 x = 0 \\ 1/y - 2\lambda_2 y = 0. \end{cases} \quad (13.11)$$

- (a) Il sistema (13.11) implica $\lambda_2 \neq 0$ e pertanto passiamo subito al caso successivo.
 (b) Se $\lambda_2 > 0$ dalle (13.11) ricaviamo $\lambda_2 = 1/(2x^2) = 1/(2y^2)$, cioè $y = x$. Poiché $\lambda_2 > 0$ implica $4 - x^2 - y^2 = 0$ nella quarta equazione del sistema (13.10), sostituendo in essa $y = x$ otteniamo $4 - 2x^2 = 0$ che porta a due soluzioni: $x = -\sqrt{2}$ e $x = \sqrt{2}$, di cui scartiamo la prima che è negativa. Sostituendo $x = \sqrt{2}$ a ritroso otteniamo il punto $(\sqrt{2}, \sqrt{2}, 0, 1/4)$, che è senz'altro un punto di K.T. in quanto risolve tutte le condizioni in (13.10), in particolare, $\lambda_2 = 1/4 > 0$. Poiché la funzione obiettivo è concava e il vincolo è convesso e compatto, si tratta dell'unica soluzione.

2. Per scrupolo, verificiamo che le ultime due possibilità conducono a una contraddizione.

- (a) Se $\lambda_1 > 0$ e $\lambda_2 = 0$, le prime due equazioni diventano $1/x + \lambda_1 y = 0$ e $1/y + \lambda_1 x = 0$, da cui segue $\lambda_1 = -1/(xy)$, che è negativo essendo $x, y > 0$: una contraddizione.
 (b) Per $\lambda_1 > 0$ e $\lambda_2 > 0$, la terza e quarta equazione in (13.10) implicano il sistema

$$\begin{cases} xy = 1 \\ x^2 + y^2 = 4, \end{cases} \quad (13.12)$$

che ha due soluzioni positive, $(x, y) = (\sqrt{2 - \sqrt{3}}, 1/\sqrt{2 - \sqrt{3}})$ e $(x, y) = (\sqrt{2 + \sqrt{3}}, 1/\sqrt{2 + \sqrt{3}})$, corrispondenti ai punti di intersezione A e B fra le curve di livello $xy = 1$ e $x^2 + y^2 = 4$ in figura 13.8(b). Sostituendo entrambe le coppie di valori nelle prime due equazioni in (13.10) giungiamo a due contraddizioni: se, ad esempio, vale $x = \sqrt{2 - \sqrt{3}}$ e $y = 1/\sqrt{2 - \sqrt{3}}$, otteniamo

$$\begin{cases} 1/\sqrt{2 - \sqrt{3}} + \lambda_1/\sqrt{2 - \sqrt{3}} = 2\lambda_2\sqrt{2 - \sqrt{3}} \\ \sqrt{2 - \sqrt{3}} + \lambda_1\sqrt{2 - \sqrt{3}} = 2\lambda_2/\sqrt{2 - \sqrt{3}} \end{cases} \iff \frac{1 + \lambda_1}{2 - \sqrt{3}} = (1 + \lambda_1)(2 - \sqrt{3})$$

da cui segue $(2 - \sqrt{3})^2 = 1$, chiaramente impossibile. Con gli stessi passaggi, per $x = \sqrt{2 + \sqrt{3}}$ e $y = 1/\sqrt{2 + \sqrt{3}}$ giungiamo alla contraddizione $(2 + \sqrt{3})^2 = 1$.

In conclusione, la soluzione è $(\sqrt{2}, \sqrt{2}, 0, 1/4)$. Vale $f(\sqrt{2}, \sqrt{2}) = \ln 2$. Si tratta di un punto in cui il vincolo è attivo, in particolare è un punto della curva di livello definita implicitamente da $x^2 + y^2 = 4$, cioè un punto di frontiera per l'insieme LI definito dalla seconda disequazione. Naturalmente si perviene alla stessa conclusione seguendo un ordine diverso per le sostituzioni nel sistema (13.10).

ESEMPIO 13.7 Vogliamo minimizzare la funzione obiettivo $f(x, y) = (x - 1)^2 + (y - 1)^2$ sul quadrato $[0, 2] \times [0, 2]$ (definizione 11.8). Formuliamo il problema come

$$\begin{aligned} & \min [(x - 1)^2 + (y - 1)^2] \\ & \text{sub } \begin{cases} x \geq 0 \\ -x \geq -2 \\ y \geq 0 \\ -y \geq -2. \end{cases} \end{aligned}$$

Il vincolo è compatto e convesso, e la funzione obiettivo è *convessa*; esiste dunque un unico punto di K.T. che è la soluzione. Siamo di nuovo nel caso 1b della regola 13.1 con $\min(m, n) = n = 2 < m = 4$. La matrice Jacobiana,

$$J = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix},$$

per $k = 1$ ha tutte le righe non nulle e, per $k = 2 [= \min(m, n)]$, *soltanto le due sottomatrici* 2×2 composte dalle prime due righe e dalle ultime due righe hanno rango minore di 2 (sono *singolari*), ma esse non corrispondono a punti di intersezione fra curve di livello poiché $x = 0$ e $-x = -2$ (e $y = 0$ e $-y = -2$) non possono valere simultaneamente.⁶ Deduciamo che vale la *qualificazione del vincolo*. La funzione Lagrangiana è $L(x, y, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = (x - 1)^2 + (y - 1)^2 + (\lambda_2 - \lambda_1)x - 2\lambda_2 + (\lambda_4 - \lambda_3)y - 2\lambda_4$, funzione di sei variabili, di cui quattro moltiplicatori, $\lambda_1, \lambda_2, \lambda_3$ e λ_4 . Dobbiamo risolvere il sistema

$$\begin{cases} L_x = 2(x - 1) + \lambda_2 - \lambda_1 = 0 \\ L_y = 2(y - 1) + \lambda_4 - \lambda_3 = 0 \\ \lambda_1 L_{\lambda_1} = -\lambda_1 x = 0 \\ \lambda_2 L_{\lambda_2} = \lambda_2(x - 2) = 0 \\ \lambda_3 L_{\lambda_3} = -\lambda_3 y = 0 \\ \lambda_4 L_{\lambda_4} = \lambda_4(y - 2) = 0 \\ L_{\lambda_1} = -x \leq 0, L_{\lambda_2} = x - 2 \leq 0, L_{\lambda_3} = -y \leq 0, L_{\lambda_4} = y - 2 \leq 0, \\ \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0, \lambda_4 \geq 0. \end{cases} \quad (13.13)$$

Si noti che, diversamente dagli esempi precedenti, le disuguaglianze della penultima riga sono di tipo ‘ \leq ’ anziché ‘ \geq ’, dal momento che ora le disuguaglianze del vincolo sono di tipo ‘ \geq ’ anziché ‘ \leq ’. Ci sono sedici casi possibili, pari a tutte le combinazioni possibili di moltiplicatori nulli/positivi. Molti di questi, però, sono impossibili perché, essendoci solo due variabili, al più due disuguaglianze alla volta possono valere con uguaglianza (le curve di livello delle g_j si intersecano al più due alla volta); pertanto, non possono esserci più di due moltiplicatori strettamente positivi contemporaneamente.

Se $\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0$ e $\lambda_4 = 0$, le prime due equazioni forniscono il punto $(1, 1, 0, 0, 0, 0)$ che soddisfa tutte le condizioni in (13.13) e dunque è l'unico punto di K.T. cercato, ovvero l'unico punto di minimo assoluto. Si tratta di un punto interno, il vincolo è inattivo. Nonostante l'elevato numero di condizioni da studiare abbia generato cupe aspettative, l'esercizio è concluso. Lasciamo al lettore l'“entusiasmante” compito di verificare che tutte le altre possibilità conducono a delle contraddizioni.

ESEMPIO 13.8 Vogliamo massimizzare $f(x, y, z) = \ln(1 + x) - y^2 - z^2$ soggetta al vincolo definito da $-x + z \geq -1, y \geq 0$ e $z \geq 0$. Il problema in forma standard è

$$\begin{aligned} & \max [\ln(1 + x) - y^2 - z^2] \\ & \text{sub} \begin{cases} x - z \leq 1 \\ -y \leq 0 \\ -z \leq 0. \end{cases} \end{aligned}$$

Poiché le funzioni che definiscono il vincolo sono lineari, il vincolo è *convesso* (osservazione 13.2), ma non è compatto (perché?). Fortunatamente la funzione obiettivo f è strettamente concava sul suo dominio, $CE = \{(x, y, z) \in \mathbb{R}^3 : x > -1\}$. Siamo ancora nel caso 1b della regola 13.1, questa

volta con $\min(m, n) = m = n = 3$. La matrice Jacobiana, $J = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$, per $k = 1$ ha

tutte le righe non nulle, per $k = 2$ tutte le (tre) sottomatrici $J_2 (2 \times 3)$ formate da coppie di righe hanno rango pieno, $\rho(J_2) = 2$, e infine, per $k = 3$, la matrice stessa, J , ha rango pieno (è, cioè, non singolare), $\rho(J) = 3$. Poiché J non dipende dalle variabili, x, y, z , ciò è senz'altro vero su tutta la frontiera del vincolo, e quindi vale la *qualificazione del vincolo*. La funzione Lagrangiana è

⁶I punti di intersezione fra curve di livello sulla frontiera del vincolo individuati da $x = 0$ e $-x = -2$ (e da $y = 0$ e $-y = -2$) corrispondono ad angoli distinti del quadrato $[0, 2] \times [0, 2]$.

$L(x, y, z, \lambda_1, \lambda_2, \lambda_3) = \ln(1+x) - y^2 - z^2 + \lambda_1(1-x+z) + \lambda_2 y + \lambda_3 z$, e il sistema da risolvere è

$$\begin{cases} L_x = 1/(1+x) - \lambda_1 = 0 \\ L_y = -2y + \lambda_2 = 0 \\ L_z = -2z + \lambda_1 + \lambda_3 = 0 \\ \lambda_1 L_{\lambda_1} = \lambda_1(1-x+z) = 0 \\ \lambda_2 L_{\lambda_2} = \lambda_2 y = 0 \\ \lambda_3 L_{\lambda_3} = \lambda_3 z = 0 \\ L_{\lambda_1} = 1-x+z \geq 0, L_{\lambda_2} = y \geq 0, L_{\lambda_3} = z \geq 0 \\ \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0. \end{cases}$$

Dalla prima equazione appare evidente che $\lambda_1 \neq 0$, quindi dev'essere $\lambda_1 > 0$ e dalla quarta equazione segue $z = x - 1$. La seconda e la quinta equazione assieme implicano $y = \lambda_2 = 0$. Se $z = 0$ la terza equazione diventa $\lambda_1 + \lambda_3 = 0$, che, poiché $\lambda_1 > 0$, implicherebbe $\lambda_3 < 0$, ma ciò è impossibile e deduciamo che $z \neq 0$, che, dalla sesta equazione, implica $\lambda_3 = 0$. Unendo la prima e la terza equazione e tenendo conto che $z = x - 1$ otteniamo $x^2 = 3/2$, e quindi due valori per x , $x = \sqrt{3/2}$ e $x = -\sqrt{3/2}$, di cui scartiamo il secondo perché $-\sqrt{3/2} < -1$, mentre il *CE* presuppone $x > -1$; dunque $z = \sqrt{3/2} - 1$ e $\lambda_1 = 2z = 2(\sqrt{3/2} - 1) = \sqrt{6} - 2 > 0$. Abbiamo così trovato l'unico punto di K.T. $(\sqrt{3/2}, 0, \sqrt{3/2} - 1, \sqrt{6} - 2, 0, 0)$, che, essendo la funzione obiettivo strettamente concava, è la soluzione del problema. Il vincolo è attivo.

ESERCIZIO 13.2 Risolvere i seguenti problemi vincolati:

$$\begin{array}{llll} 1. & \max e^{x+y} & 2. & \max(-x^2 - y^2) \\ & \text{sub } x^2 + (1/2)y^2 \leq 1 & & \text{sub } y \geq x - 1 \\ 3. & \min(x^2 - y + e^z - z) & 4. & \min[x^2 - \ln(y+z)] \\ & \text{sub } x \geq y^2 & & \text{sub } y^2 + z^2 \leq x \end{array}$$

ESERCIZIO 13.3 Risolvere i seguenti problemi vincolati:

$$\begin{array}{lll} 1. & \max[\ln(1+x) - y^2] & 2. & \min(-\sqrt{x+3y}) \\ & \text{sub } \begin{cases} x \geq 0 \\ -x - y \geq -1 \end{cases} & & \text{sub } \begin{cases} x \geq 0 \\ y \geq 1 \\ y \leq 4 - x^2 \end{cases} \\ 3. & & & \max[x^2 + y^2 - 3(x+y) + 9/2] \\ & & & \text{sub } \begin{cases} x \geq 0 \\ y \geq 0 \\ 2x \leq 2 - y \\ (1/2)x \leq 1 - y \end{cases} \\ 4. & \max[(x-3/2)^2 + (y-3/2)^2] & 5. & \min(x-3)^2 \\ & \text{sub } \begin{cases} x \geq 0 \\ y \geq 0 \\ y \leq 2 - x \end{cases} & & \text{sub } \begin{cases} (x-1)^2 + (y-1)^2 \leq 1 \\ (1/9)(x-1/2)^2 + (y-1)^2 \leq 1/4 \end{cases} \\ 6. & & & \min(x+y) \\ & & & \text{sub } \begin{cases} y \leq x^3 \\ x \leq 1 \\ y \geq 0 \end{cases} \\ 7. & \max(-x^2 - y^2 - z^2) & 8. & \max(-x^2 + 2y - z^2) \\ & \text{sub } \begin{cases} 3x + y + 2z \geq 1 \\ x \geq 0 \end{cases} & & \text{sub } \begin{cases} x + y + z \leq 8 \\ y \leq 4 \end{cases} \\ 9. & & & \min[\ln(1-x^2-y^2-z^2) - 2z] \\ & & & \text{sub } \begin{cases} x^2 + y^2 + z^2 \leq 1/2 \\ x \leq 0 \\ y \leq 0 \end{cases} \\ 10. & \max[-(e^{-x} + e^{-z})^2 - y] & 11. & \min(x_1^2 + x_2^2 + x_3^2 + x_4^2) \\ & \text{sub } \begin{cases} x \leq 0 \\ y \geq 0 \\ z \leq 1 \end{cases} & & \text{sub } \begin{cases} -x_1 - x_2 - x_3 - x_4 \leq -2 \\ x_1 - x_3 \leq 0 \end{cases} \end{array}$$