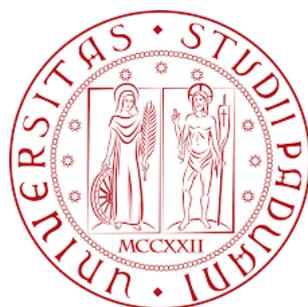


Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in

STATISTICA PER L'ECONOMIA E L'IMPRESA



RELAZIONE FINALE

Ottimizzazione vincolata e teorema di Karush–Kuhn–Tucker

Relatore: Prof.ssa Giulia Treu

Dipartimento di Matematica

Laureando: Mattia Bricolo

Matricola: 1149902

Anno Accademico 2018/2019

*«Capisci quello che stai leggendo?»
(At 8, 30-31)*

Sommario

In ambito scientifico, e non solo, problemi come l'ottimizzazione si trovano in ogni dove: dalla matematica alla statistica, dall'ingegneria all'economia. Molte sono le possibili soluzioni ai suddetti problemi ma in questo scritto ci si soffermerà in particolar modo nei principali teoremi. Uno di questi è rappresentato dal cosiddetto *teorema di Karush–Kuhn–Tucker* (KKT), una generalizzazione del teorema dei moltiplicatori di Lagrange. Obiettivo di tale tesi è quello di introdurre il teorema di KKT delineandone le principali definizioni e proponendo, infine, una semplice applicazione.

Abstract

In science, and not only, you'll find everywhere problems like the optimization: from mathematics to statistics, from engineering to economy. There are many solutions to those problems but in this paper we'll focus on the main theorems. One of these is represented by the so-called *Karush-Kuhn-Tucker theorem* (KKT), a generalization of the Lagrange multiplier theorem. The goal of this thesis is to introduce the KKT theorem illustrating the main definitions and proposing an easy application.

Indice

Introduzione	xi
1 Ottimizzazione non vincolata e vincolata	13
1.1 Il problema in \mathbb{R}	13
1.2 Il problema in \mathbb{R}^p , $p \in \mathbb{N}$, $p \geq 2$	16
1.2.1 Definizione dei vincoli di uguaglianza $g(\cdot)$	20
2 Il Teorema di Karush–Kuhn–Tucker	25
2.1 Programmazione lineare con vincoli non attivi	25
2.2 Programmazione lineare con vincoli non attivi e attivi	29
2.2.1 Semplice interpretazione geometrica del teorema di KKT	31
3 Caso pratico del teorema di Karush–Kuhn–Tucker	33
Conclusioni	41
Codici R	43
Bibliografia	47

Elenco delle figure

1.1	Paraboloido ellittico: In (a) le curve di livello, in (b) il grafico tridimensionale in prospettiva.	19
1.2	Paraboloido iperbolico: (a) curve di livello, (b) grafico tridimensionale in prospettiva.	22
2.1	Curve di livello, vincolo di uguaglianza e grafico in prospettiva.	28
2.2	Interpretazione geometrica del teorema di KKT.	32
3.1	In (a) le curve di livello, in (b) il grafico tridimensionale in prospettiva.	39
3.2	In rosso l'equazione $-\frac{1}{2}x_1 + 2 = 0$, in blu l'equazione $\frac{1}{2}x_1 = 0$.	40

Introduzione

Molto spesso in campo scientifico si ha a che fare con problemi di ottimizzazione, ossia determinare quali sono quei punti che massimizzano, o minimizzano, una certa funzione.

Ad esempio un'ottimizzazione frequente in statistica riguarda la massimizzazione della funzione di verosimiglianza, in economia può essere d'interesse determinare quale combinazione dei prezzi e dei beni massimizza l'utilità di un individuo o ancora, in ambito di *risk management*, minimizzare il rischio derivante da una certa attività rischiosa. Si pensi anche alla recente (2018) assegnazione della medaglia Fields al matematico italiano Alessio Figalli come riconoscimento "per i contributi alla teoria del trasporto ottimale [...]" ([13]). Sono alcuni esempi i quali per essere risolti hanno bisogno che si massimizzi/minimizzi una ben definita funzione.

Questa tesi ha come scopo quello di introdurre i più conosciuti concetti dell'analisi matematica con l'intento di spiegare semplicemente cosa sta dietro l'ottimizzazione. Lo studio di funzioni è alla base dell'ottimizzazione che assieme ad altri strumenti più o meno sofisticati, come ad esempio il *teorema di Karush–Kuhn–Tucker* (KKT), consentiranno di capire ed affrontare al meglio i problemi esposti nei capitoli seguenti.

Nel 1950 un matematico dell'università di Princeton di nome Albert Tucker tenne una conferenza dal titolo "Programmazione Nonlineare". Essa si basa-

va su di un lavoro congiunto con un giovane matematico, Harold Kuhn della medesima università: con tale scritto i due matematici dimostrarono i principali teoremi ed è da qui che deriva il cosiddetto "teorema di Kuhn–Tucker". Essi però non furono i primi matematici ad intraprendere la strada della programmazione nonlineare. Già nel 1939 il matematico William Karush dimostrò il suddetto teorema nella sua tesi di laurea magistrale presso l'università di Chicago. Anche se Karush provò prima le condizioni della programmazione lineare, oggi si sente parlare del teorema di Kuhn–Tucker, anche se non è raro trovare riferimenti al teorema di Karush–Kuhn–Tucker (KKT) con l'obiettivo di citare e ringraziare Karush per il lavoro svolto ([7], pagg. 331–332).

Ma cos'è la programmazione nonlineare?

In semplicità si ipotizzi il seguente problema (si veda anche [8] e [1])

$$\text{Ottimizzare } f(\mathbf{x}) \quad \text{sotto i vincoli } g(\mathbf{x}) = \mathbf{b}, \quad (1)$$

in cui $f : X \subseteq \mathbb{R}^p \rightarrow \mathbb{R}$ e $g : X \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^m$ sono funzioni note, $\mathbf{x} \in X$ e $\mathbf{b} \in Y$; si vuole quindi trovare quel vettore $\tilde{\mathbf{x}}$ tale per cui la funzione f risulta ottimizzata sotto le condizioni (meglio dette *vincoli*) g .

Due casi particolari della (1) che verranno analizzati in questo scritto, almeno in parte, sono:

1. posto $\mathbf{b} = \mathbf{0}$, ove $\mathbf{0} = (0, \dots, 0)^\top$, si ottiene il caso in cui $g(\mathbf{x}) = \mathbf{0}$ e le soluzioni al problema vengono risolte dal noto teorema dei moltiplicatori di Lagrange,
2. posto $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}$ e $g(\mathbf{x}) = V\mathbf{x}$ due funzioni lineari nella variabile \mathbf{x} , con $\mathbf{w} \in X$ e V una matrice $m \times p$, il problema si riconduce ad ottimizzare $f(\mathbf{x})$ sotto il vincolo $V\mathbf{x} \leq \mathbf{b}$ oppure $V\mathbf{x} = \mathbf{b}$.

A fine tesi verrà poi condotta una semplice applicazione per mostrare e dimostrare in pratica ciò che la teoria afferma.

Capitolo 1

Ottimizzazione non vincolata e vincolata

1.1 Il problema in \mathbb{R}

Il problema sopra definito può essere quindi tradotto in forma matematica come segue: trovare quel valore $\tilde{x} \in \mathcal{D}$ tale per cui $f(\tilde{x}) < f(x)$ (oppure $f(\tilde{x}) > f(x)$) per ogni $x \in \mathcal{D}$, ove $f : \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R}$ definisce una certa funzione.

Nel seguito ciò verrà scritto come

$$\underset{x \in \mathcal{D}}{\text{opt}} f(x), \quad (1.1)$$

ove *opt* sta per *optimization* e quindi ottimizzazione.

Definizione 1.1 (Massimo assoluto). $M \in \mathbb{R}$ è detto *massimo assoluto* di $f : \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R}$ se esiste $\tilde{x} \in \mathcal{D}$ detto *punto di massimo assoluto* di f se

$$f(x) \leq M \quad \forall x \in \mathcal{D},$$

$$f(\tilde{x}) = M.$$

Definizione 1.2 (Minimo assoluto). $m \in \mathbb{R}$ è detto *minimo assoluto* di $f : \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R}$ se esiste $\tilde{x} \in \mathcal{D}$ detto *punto di minimo assoluto* di f se

$$f(x) \geq m \quad \forall x \in \mathcal{D},$$

$$f(\tilde{x}) = m.$$

Definizione 1.3 (Massimo relativo). $M_0 \in \mathbb{R}$ è detto *massimo relativo* di $f : \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R}$ se esiste un intorno I di \tilde{x} , e si scriverà $I(\tilde{x})$, tale per cui

$$f(\tilde{x}) \geq f(x) \quad \forall x \in I,$$

$$f(\tilde{x}) = M_0.$$

Definizione 1.4 (Minimo relativo). $m_0 \in \mathbb{R}$ è detto *minimo relativo* di $f : \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R}$ se esiste un intorno $I(\tilde{x})$ tale per cui

$$f(\tilde{x}) \leq f(x) \quad \forall x \in I,$$

$$f(\tilde{x}) = m_0.$$

Il tipo di ottimizzazione vista finora è detta *ottimizzazione non vincolata* (o *libera*) in quanto non si tiene conto di eventuali vincoli. E se invece ce ne fossero?

Si supponga ad esempio di voler minimizzare la funzione $f(x) = x^2$ con $x \in \mathcal{D} \subseteq \mathbb{R}$; utilizzando i simboli introdotti poc'anzi la soluzione è data da

$$\tilde{x} = 0,$$

$$m = 0.$$

Banalmente se viene ristretto il dominio della funzione passando ad esempio da $\mathcal{D} \subseteq \mathbb{R}$ a $\mathcal{D} = [5, 10]$, i nuovi valori saranno

$$\tilde{x} = 5,$$

$$m = 25.$$

Si è quindi introdotto un vincolo, ossia si è "obbligato" la variabile x a muoversi nell'intervallo $[5, 10]$ anziché in tutto l'asse reale.

È necessario introdurre e delimitare ulteriori concetti che verranno poi ripresi in maniera più approfondita nei prossimi capitoli.

Considerando l'obiettivo (1.1)

Definizione 1.5 (Punto stazionario). Sia $f(x)$ una funzione derivabile in \tilde{x} . Allora $\tilde{x} \in \mathcal{D}$ è detto *punto stazionario* se

$$f'(x)|_{x=\tilde{x}} = \left. \frac{d}{dx} f(x) \right|_{x=\tilde{x}} = 0. \quad (1.2)$$

Ogni punto che soddisfa la (1.2) è un possibile "candidato" ad essere punto di massimo (o minimo) relativo (o assoluto).

È necessario quindi determinare dove la derivata prima risulta positiva (oppure negativa). Se essa è negativa prima di \tilde{x} e positiva dopo, si può affermare che tale punto è di minimo; viceversa se risulta positiva prima \tilde{x} e negativa dopo esso è di massimo.

Una ulteriore analisi riguarda lo studio della derivata seconda, definita da

$$f''(x) = \frac{d^2}{dx^2} f(x). \quad (1.3)$$

Utilizzando definizioni più rigorose diremo che

Definizione 1.6 (Funzione convessa). Sia $I \in \mathbb{R}$ un intervallo e siano $x_1, x_2 \in I$ con $x_1 \neq x_2$. Si dice allora che una funzione è *convessa* in I se il segmento che congiunge i punti $(x_1, f(x_1))$ e $(x_2, f(x_2))$ risulta sopra la curva. Se f è derivabile la derivata prima di f è una funzione crescente e se f è derivabile due volte $f''(x) \geq 0$.

Analogamente

Definizione 1.7 (Funzione concava). Una funzione si definisce *concava* se il segmento che congiunge $(x_1, f(x_1))$ e $(x_2, f(x_2))$ risulta essere sotto la curva. Se f è derivabile la derivata prima di f è decrescente e se f è due volte derivabile $f''(x) \leq 0$.

1.2 Il problema in \mathbb{R}^p , $p \in \mathbb{N}$, $p \geq 2$

Finora si è parlato dell'ottimizzazione (vincolata e non) per funzioni in una variabile a valori reali.

Si supponga ora di avere il seguente problema

$$\underset{\mathbf{x} \in \mathbb{R}^p}{\text{opt}} f(\mathbf{x}), \quad (1.4)$$

In questo caso la variabile \mathbf{x} appartiene ad uno spazio p -dimensionale (con $p \in \mathbb{N}$, $p \geq 2$) ed è quindi descritta dal vettore $\mathbf{x} = (x_1, \dots, x_p)^\top$. I concetti in generale però rimangono gli stessi solamente che la derivata prima della funzione è sostituita dal differenziale primo, rappresentato dal gradiente, e la derivata seconda della funzione è rappresentata dal differenziale secondo, rappresentato dalla matrice Hessiana.

Si pensi ad esempio alla funzione $z = f(x, y) = x^2 + y^2$ che descrive una parabola (detta *paraboloide ellittico*) nello spazio (x, y, z) mostrata in Figura 1.1.

Per poter ottimizzare tale funzione è necessario introdurre dei concetti più avanzati rispetto quelli della sezione 1.1.

Definizione 1.8 (Gradiente). Sia $f : \mathcal{D} \subseteq \mathbb{R}^p \rightarrow \mathbb{R}$ una funzione differenziabile in \mathcal{D} e $\mathbf{x} = (x_1, \dots, x_p)^\top$ il vettore di variabili. Si definisce *gradiente* di

f il vettore

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_p} f(\mathbf{x}) \end{bmatrix}. \quad (1.5)$$

Allora, per determinare se un punto $\tilde{\mathbf{x}}$ è un punto critico per f si procede come nel caso in una variabile ponendo $\nabla f(\mathbf{x}) = \mathbf{0}$, in cui $\mathbf{0} = (0, \dots, 0)^\top$ rappresenta il vettore nullo.

Definizione 1.9 (Matrice hessiana). Sia $f : \mathcal{D} \subseteq \mathbb{R}^p \rightarrow \mathbb{R}$ una funzione differenziabile due volte in \mathcal{D} . Si indichi con $\mathbf{x} = (x_1, \dots, x_p)^\top$ il vettore di variabili. Si definisce *matrice hessiana* di f la matrice

$$H_f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1^2} & \frac{\partial f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial f(x)}{\partial x_1 \partial x_p} \\ \frac{\partial f(x)}{\partial x_2 \partial x_1} & \frac{\partial f(x)}{\partial x_2^2} & \dots & \frac{\partial f(x)}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(x)}{\partial x_p \partial x_1} & \frac{\partial f(x)}{\partial x_p \partial x_2} & \dots & \frac{\partial f(x)}{\partial x_p^2} \end{bmatrix}. \quad (1.6)$$

Tale matrice calcola $p \times p$ derivate parziali. Un particolare teorema fornisce una semplificazione per quanto riguarda le derivate seconde miste (si veda [3], pag. 342).

Teorema 1.1 (Teorema di Schwarz). *Siano $f : \mathcal{D} \subseteq \mathbb{R}^p \rightarrow \mathbb{R}$, con \mathcal{D} insieme aperto, e $x \in \mathcal{D}$. Se f è continua con derivate prime e seconde continue, allora $H_f(x)$ è simmetrica.*

Nel caso in cui $\mathbf{x} \in \mathbb{R}^p$ non è immediato verificare che un punto sia di massimo/minimo. Le cose si semplificano se $p = 2$ mentre bisogna calcolare particolari quantità se $p > 2$.

Definizione 1.10 (Punto di sella). Un punto \tilde{x} è detto *punto di sella* se è un punto stazionario per f e non esiste nessun intorno di \tilde{x} in cui tale punto è di massimo o minimo.

Per farsi un'idea di tale definizione si pensi al grafico del cosiddetto *paraboloide iperbolico* descritto dalla funzione $z = f(x, y) = x^2 - y^2$ in Figura 1.2. Esso ha un punto di sella esattamente in $(x, y) = (0, 0)$: ciò significa che la sezione di piano parallela a (x, z) ha un minimo in $(0, 0)$ (in quanto, posto $y = 0$ si ottiene $z = x^2$) mentre quella parallela a (y, z) ha un massimo in $(0, 0)$ (in quanto ponendo $x = 0$ si ottiene $z = -y^2$).

Per determinare se un punto stazionario è un punto di massimo, minimo o sella nel caso in cui $p \geq 2$ è necessario studiare la matrice hessiana nel punto candidato.

In particolare, se $p = 2$ ed M un matrice quadrata del tipo

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

con determinante $\det(M) = m_{11}m_{22} - m_{12}m_{21}$ allora

- se $\det(M) > 0$ e $m_{11} > 0$, M è *definita positiva*,
- se $\det(M) > 0$ e $m_{11} < 0$, M è *definita negativa*,
- se $\det(M) = 0$, M è *semidefinita*,
- se $\det(M) < 0$, M è *indefinita*.

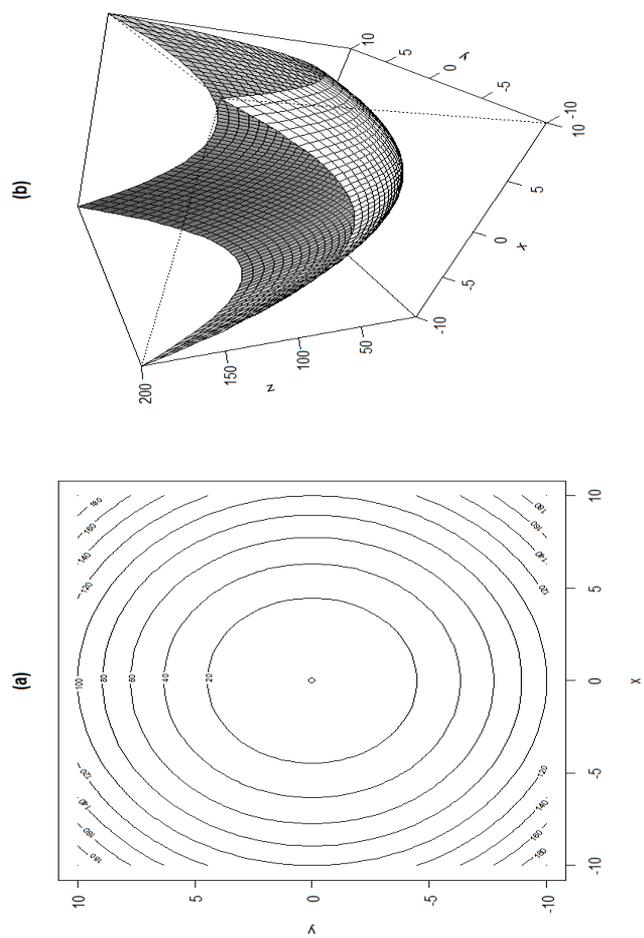


Figura 1.1: Paraboloide ellittico: In (a) le curve di livello, in (b) il grafico tridimensionale in prospettiva.

Posto $M = H_f(\tilde{\mathbf{x}})$ con $\tilde{\mathbf{x}} \in \mathbb{R}^2$ punto stazionario per f ; se

- $H_f(\tilde{\mathbf{x}})$ è definita positiva il punto $\tilde{\mathbf{x}}$ è di minimo locale,
- $H_f(\tilde{\mathbf{x}})$ è definita negativa il punto $\tilde{\mathbf{x}}$ è di massimo locale,
- $H_f(\tilde{\mathbf{x}})$ è indefinita il punto $\tilde{\mathbf{x}}$ è di sella,
- $H_f(\tilde{\mathbf{x}})$ è semidefinita non si giunge a nessuna conclusione immediata.

Nel caso, invece, in cui $p > 2$ si studia il segno degli autovalori della matrice hessiana valutata nei punti stazionari (non essendo l'obiettivo di questo scritto trattate argomenti di algebra lineare, per approfondimenti sugli autovalori si consigliano [14] e [11]). Detti δ_i , $i = 1, \dots, p$, gli autovalori di $H_f(\tilde{\mathbf{x}})$,

- se $\delta_i > 0 \forall i$, allora $H_f(\tilde{\mathbf{x}})$ è definita positiva e $\tilde{\mathbf{x}}$ è di minimo locale,
- se $\delta_i < 0 \forall i$, allora $H_f(\tilde{\mathbf{x}})$ è definita negativa e $\tilde{\mathbf{x}}$ è di massimo locale,
- se $\exists \delta_i > 0$ e $\delta_j < 0$ allora $H_f(\tilde{\mathbf{x}})$ è indefinita e $\tilde{\mathbf{x}}$ è di sella,
- se $\delta_i \leq 0 \forall i$ (o $\delta_i \geq 0$) non si giunge a nessuna conclusione circa la natura di $\tilde{\mathbf{x}}$ in quanto $H_f(\tilde{\mathbf{x}})$ è semidefinita negativa (o positiva).

1.2.1 Definizione dei vincoli di uguaglianza $g(\cdot)$

Si è fin qui parlato dei vincoli solamente in maniera 'intuitiva' senza entrare troppo nello specifico.

Sia ora l'obiettivo il seguente:

$$\underset{x \in \mathbb{R}^p}{\text{opt}} f(x) \quad \text{sotto i vincoli} \quad g_i(x) = c, \quad i = 1, \dots, m, \quad (1.7)$$

con $\mathcal{D} \subseteq \mathbb{R}^p$, $f : \mathcal{D} \rightarrow \mathbb{R}$ e $g : \mathcal{D} \rightarrow \mathbb{R}^m$, $1 \leq m \leq p - 1$.

L'obiettivo è dunque determinare quel (o quei) valore \tilde{x} tale per cui si ottimizzi¹ la funzione $f(x)$ sottostando però ai vincoli imposti dalle equazioni $g_i(x) = c$ per $i = 1, \dots, m$. Il termine noto c verrà posto solitamente pari a zero senza perdita di generalità. Il problema (1.7) può essere visto nella sua versione vettoriale

$$\underset{\mathbf{x} \in \mathbb{R}^p}{\text{opt}} f(\mathbf{x}) \quad \text{sotto i vincoli} \quad g(\mathbf{x}) = \mathbf{0}, \quad (1.8)$$

con $\mathbf{0} = (0, \dots, 0)^\top$ vettore nullo.

Un altro concetto che è bene conoscere è la differenza tra *vincolo esplicitabile* e *vincolo non esplicitabile* in quanto d'ora in avanti si terrà conto dei vincoli nelle analisi.

Definizione 1.11 (Vincolo esplicitabile). Un vincolo g è detto *esplicitabile* se è possibile scrivere (esplicitare) una variabile (o alcune variabili) in funzione delle rimanenti.

Ad esempio, posti $f(x, y)$ una funzione e $g(x, y) = x^2 - y = 0$ il vincolo, quest'ultimo può essere scritto come $\gamma(x) = y = x^2$ e quindi sostituito nella funzione come $f(x, \gamma(x))$.

Se non fosse possibile esplicitarlo si parla di *vincolo non esplicitabile*. É questo il caso più interessante da analizzare in quanto se il vincolo risultasse esplicitabile sarebbe banale la determinazione dei punti critici.

Ci si mette quindi nella condizione in cui non si riesca ad esplicitare il vincolo. Per trovare i punti stazionari viene in aiuto il cosiddetto *teorema dei moltiplicatori di Lagrange* (si veda [9], [12] e [3]).

¹Viene utilizzato il termine generico "ottimizzare" il quale andrà sostituito dai termini "massimizzare" o "minimizzare" in base alle situazioni.

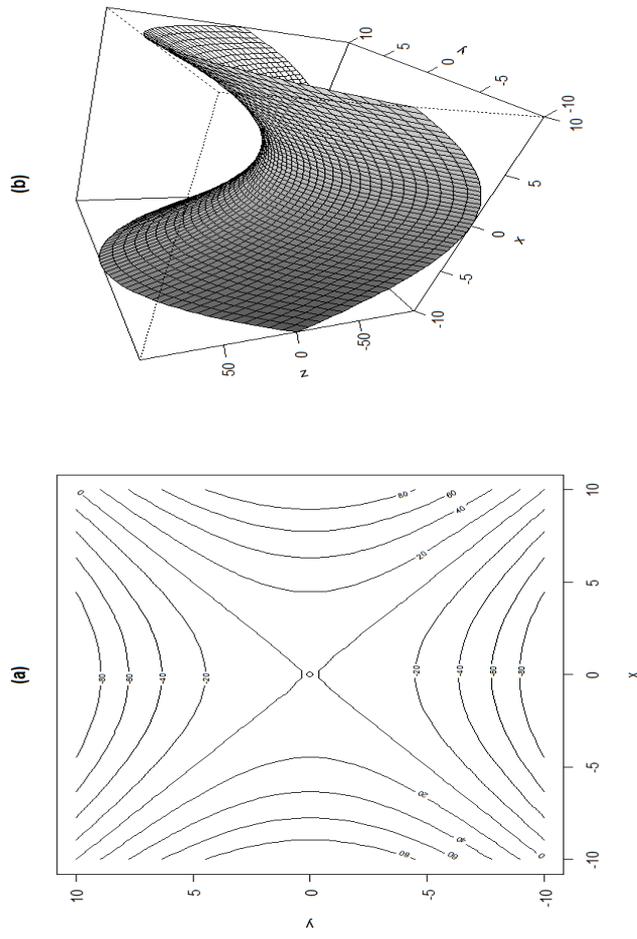


Figura 1.2: Paraboloide iperbolico: (a) curve di livello, (b) grafico tridimensionale in prospettiva.

Introducendo quindi i vincoli si parla, in questo caso, di 'estremi vincolati'; ossia determinare quei punti di massimo (o minimo) vincolati a giacere, ad esempio, in una certa porzione del piano o su di una retta o curva in generale.

Definizione 1.12 (Rango). Sia A una matrice di dimensione $m \times n$. Si definisce *rango* di A , e lo si identifica con $rank(A)$, il massimo numero di colonne (o righe) linearmente indipendenti di A .

Definizione 1.13 (Punto regolare). Il punto $\tilde{\mathbf{x}} \in \mathcal{D} \subseteq \mathbb{R}^p$ è detto *punto regolare* per il vincolo g se $rank(Jg(\tilde{\mathbf{x}})) = m$, in cui J rappresenta la matrice jacobiana.

Teorema 1.2 (Dei moltiplicatori di Lagrange). Siano $f \in \mathcal{C}^1(\mathcal{D}, \mathbb{R})$, e $g \in \mathcal{C}^1(\mathcal{D}, \mathbb{R}^m)$ con $\mathcal{D} \subseteq \mathbb{R}^p$ e sia $\tilde{\mathbf{x}} \in \mathcal{D}$ un punto regolare per g . Allora se $\tilde{\mathbf{x}}$ è estremo vincolato per la funzione f esistono m moltiplicatori $\tilde{\lambda}_j \in \mathbb{R}$ tali per cui, posto $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_m)^\top$

$$\mathcal{L}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}) = \nabla f(\tilde{\mathbf{x}}) + \tilde{\boldsymbol{\lambda}}^\top Jg(\tilde{\mathbf{x}}) = \mathbf{0}.$$

In (1.7) e (1.8) si è usato, nel vincolo, il simbolo "=" (uguale) ma in molti problemi applicativi alcuni vincoli sono rappresentati da equazioni (in questo caso viene detto che il vincolo è *attivo*) ed altri da disequazioni (il vincolo qui è detto *non attivo*).

Il teorema dei moltiplicatori di Lagrange riesce a risolvere il problema dell'ottimizzazione vincolata solamente se il vincolo è espresso da m equazione e quindi se è attivo; negli altri casi è necessario introdurre ulteriori concetti e in particolare quello che viene chiamato *teorema di Karush-Kuhn-Tucker*.

Capitolo 2

Il Teorema di Karush–Kuhn–Tucker

2.1 Programmazione lineare con vincoli non attivi

L'ultima parte di questo scritto analizza particolari condizioni tali per cui un punto si possa definire *punto di Kuhn–Tucker* (come riferimento bibliografico si veda [4], [5] e [2]).

In particolare sia dato il seguente problema di ottimizzazione

$$\begin{aligned} \min_{x \in \mathbb{R}^p} f(x) \quad \text{sotto i vincoli} \quad & g_i(x) \leq 0 \quad i = 1, \dots, n, \\ & h_j(x) = 0 \quad j = 1, \dots, m \quad 1 \leq m \leq p - 1. \end{aligned} \tag{2.1}$$

In questa prima parte i vincoli $h_j(x) = 0$ non verranno utilizzati nelle analisi ma ci si soffermerà solamente sui vincoli di disuguglianza.

Teorema 2.1 (Teorema di Karush–Kuhn–Tucker). *Sia \tilde{x} di minimo locale per f tale per cui soddisfi la condizione 2.1. Siano f e g_i , per $i =$*

$1, \dots, n$, funzioni differenziabili. Esiste allora un vettore $\tilde{\lambda} \geq \mathbf{0}$, ove $\tilde{\lambda} = (\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_n)^\top$, tale per cui

$$\begin{aligned} \mathcal{L}(\tilde{\mathbf{x}}, \tilde{\lambda}) &= \nabla f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \tilde{\lambda}_i \nabla g_i(\tilde{\mathbf{x}}) = \mathbf{0} \\ \tilde{\lambda}_i g_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \dots, n. \end{aligned} \tag{2.2}$$

L'equazione $\mathcal{L}(\tilde{\mathbf{x}}, \tilde{\lambda}) = \mathbf{0}$ prende il nome di condizione di KKT e i $\tilde{\lambda}_i$ sono detti moltiplicatori di KKT.

Per capire e semplificare la dimostrazione del teorema di KKT è necessario introdurre le cosiddette *condizioni di Fritz-John*.

Teorema 2.2 (Condizioni necessarie di Fritz-John). *Sia $\tilde{\mathbf{x}}$ di minimo locale per f tale per cui soddisfi la condizione 2.1. Esistono allora $\lambda_0, \lambda_1, \dots, \lambda_n \geq 0$, non tutti nulli, tali per cui*

$$\begin{aligned} \lambda_0 \nabla f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \lambda_i \nabla g_i(\tilde{\mathbf{x}}) &= \mathbf{0} \\ \lambda_i g_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \dots, n. \end{aligned}$$

Dimostrazione. Il teorema di KKT viene dimostrato dando per noto il teorema sulle condizioni necessarie di Fritz-John.

Stante alle condizioni di Fritz-John esistono $\tilde{\lambda}_0, \tilde{\lambda}_1, \dots, \tilde{\lambda}_n$, non tutti nulli, tali per cui

$$\begin{aligned} \tilde{\lambda}_0 \nabla f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \tilde{\lambda}_i \nabla g_i(\tilde{\mathbf{x}}) &= \mathbf{0} \\ \tilde{\lambda}_i g_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \dots, n. \end{aligned}$$

É necessario inoltre che $\tilde{\lambda}_0 \neq 0$ in quanto se ciò non fosse, $\sum_{i \in \mathcal{I}} \tilde{\lambda}_i \nabla g_i(\tilde{\mathbf{x}}) = \mathbf{0}$. Non potendo essere tutti i $\tilde{\lambda}_i = 0$ si avrebbe una contraddizione. Allora $\tilde{\lambda}_0 > 0$. Se si definisce $\lambda_i = \tilde{\lambda}_i / \tilde{\lambda}_0$, si ottiene la condizione del teorema di KKT. □

Osservazioni.

Un'interpretazione geometrica del teorema di KKT, in un caso speciale, può essere la seguente. In Figura 2.1 (a sinistra) sono rappresentate le curve di livello della funzione il cui grafico in 3D è rappresentato nella medesima Figura 2.1 (a destra). La retta in rosso è il vincolo che stiamo considerando. Nel caso rappresentato in figura il vincolo è tangente ad un insieme di livello. Se si considera il valore assunto da f nel punto e ci si muove sul vincolo verso destra o sinistra si osserva che i valori crescono sempre in quanto il vincolo è contenuto completamente nell'insieme di sopralivello¹ per f . Quindi il punto in cui il vincolo è tangente all'insieme di livello è di minimo per f . Se il vincolo, invece, intersecasse l'insieme di livello si avrà che in un intorno del punto d'intersezione ci sarebbero anche valori di f più piccoli e il punto non potrebbe essere di minimo.

Per avere una condizione di tangenza fra il vincolo e l'insieme di livello si deve avere per prima cosa che l'insieme di livello e il vincolo devono avere la stessa retta normale (o ortogonale alla tangente) e questo, per la proprietà geometrica del gradiente, significa che il gradiente della funzione f e il gradiente della funzione g che descrive il vincolo devono essere paralleli. In termini matematici si scriverà

$$\nabla f(\cdot) = \lambda \nabla g(\cdot).$$

Invece, nel caso in cui i vincoli siano di disuguaglianza si può supporre che il vincolo sia uno dei due semipiani in cui resta diviso il piano dalla retta rossa nella Figura 2.1. Se è il semipiano che sta sopra la retta, il gradiente di g è ortogonale alla retta e punta verso il basso, mentre il gradiente di f è ortogonale alla medesima retta ma punta verso l'alto, cioè il moltiplicatore λ è positivo ed effettivamente il punto di tangenza è di minimo. Se, invece, il

¹Un insieme F_ξ è detto di *sopralivello di quota* ξ se $F_\xi = \{x \in X : f(x) \geq \xi\}$.

vincolo descritto dalla disuguaglianza fosse quello sotto la retta si avrebbe un moltiplicatore λ negativo e, dall'analisi del grafico il punto non è di minimo.

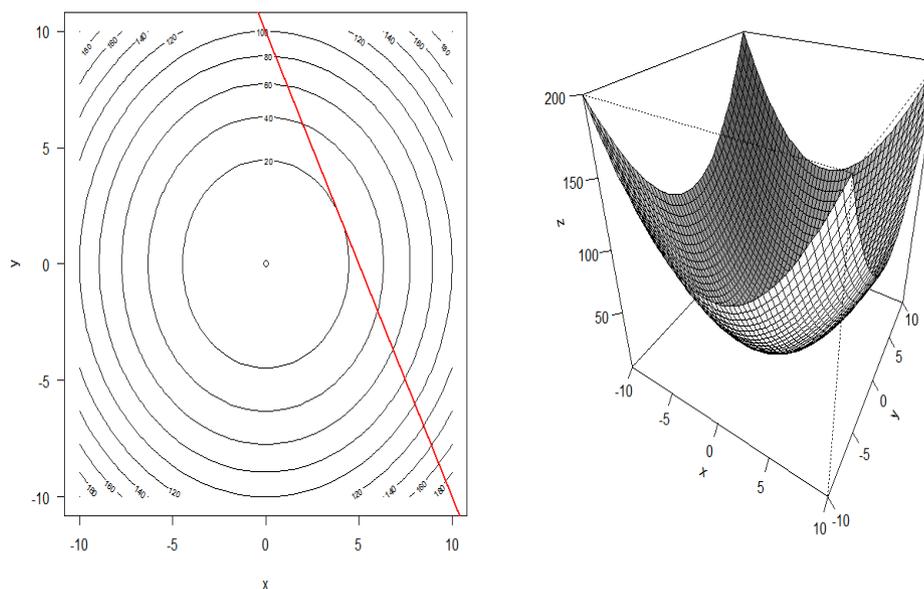


Figura 2.1: Curve di livello, vincolo di uguaglianza e grafico in prospettiva.

La coppia $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}})$ è detta punto di Kuhn–Tucker e quindi soddisfa le condizioni precedenti.

Teorema 2.3 (Condizioni necessarie e sufficienti per il minimo assoluto).
Sia f una funzione convessa con gradiente ∇f continuo in \mathbb{R}^n . Il punto $\tilde{\mathbf{x}}$ è di minimo assoluto per f se esiste un vettore $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^n$ tale per cui siano soddisfatte le seguenti condizioni

$$\mathcal{L}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}) = 0,$$

$$\tilde{\boldsymbol{\lambda}}^\top g(\tilde{\mathbf{x}}) = 0,$$

$$\tilde{\lambda}_i \geq 0, \quad i = 1, \dots, n.$$

Definizione 2.1 (Funzione affine). Sia $f : \mathcal{D} \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^m$ una funzione e siano $\mathbf{x}, \mathbf{y} \in \mathcal{D}$ e $\delta \in \mathbb{R}$. f prende il nome di *funzione affine* (o *funzione lineare*) se: $f(\mathbf{x} \pm \mathbf{y}) = f(\mathbf{x}) \pm f(\mathbf{y})$ e $f(\delta \mathbf{x}) = \delta f(\mathbf{x})$.

Definizione 2.2 (Combinazione lineare). Siano f_1, f_2, \dots, f_n funzioni e $\delta_1, \delta_2, \dots, \delta_n$ dei valori reali. La forma $\sum_{i=1}^n \delta_i f_i = \delta_1 f_1 + \delta_2 f_2 + \dots + \delta_n f_n$ è detta *combinazione lineare* di f_i di coefficienti $\delta_i, i = 1, \dots, n$.

Definizione 2.3 (Combinazione linearmente indipendente). Una combinazione lineare è detta *linearmente indipendente* se alcuna funzione può essere espressa come combinazione lineare delle rimanenti.

Affinché il punto $\tilde{\mathbf{x}}$ soddisfi le condizioni del Teorema 2.1 i vincoli devono "sottostare" ad alcune restrizioni chiamate *qualificazione dei vincoli*. Le principali sono

1. la linearità dei vincoli (*linearity constraint qualification* (LCQ)): se i vincoli g_i , per $i = 1, \dots, n$, e h_j , per $j = 1, \dots, m$, con $1 \leq m \leq p - 1$, sono funzioni affini allora le condizioni di KKT sono soddisfatte;
2. l'indipendenza lineare dei vincoli (*linear independence constraint qualification* (LICQ)): se i gradienti ∇g_i , per $i = 1, \dots, n$, e ∇h_j , per $j = 1, \dots, m$, con $1 \leq m \leq p - 1$, sono tra loro linearmente indipendenti le condizioni di KKT risultano soddisfatte.

2.2 Programmazione lineare con vincoli non attivi e attivi

Teorema 2.4 (Teorema di Karush–Kuhn–Tucker). *Sia $\tilde{\mathbf{x}}$ minimo locale per f tale per cui soddisfi la condizione 2.1. Siano f, g_i, h_j , per $i = 1, \dots, n$ e*

$j = 1, \dots, m$, funzioni differenziabili. Esistono allora $\tilde{\lambda} \geq \mathbf{0}$ e $\tilde{\mu} \in \mathbb{R}^m$, ove $\tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_m)^\top$, tali che

$$\mathcal{L}(\tilde{\mathbf{x}}, \tilde{\lambda}, \tilde{\mu}) = \nabla f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \tilde{\lambda}_i \nabla g_i(\tilde{\mathbf{x}}) + \sum_{j=1}^m \tilde{\mu}_j \nabla h_j(\tilde{\mathbf{x}}) = \mathbf{0} \quad (2.3)$$

$$\tilde{\lambda}_i g_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, \dots, n.$$

Nel caso in cui le funzioni siano convesse le condizioni del teorema di KKT risultano essere sufficienti.

Teorema 2.5 (Teorema di Karush–Kuhn–Tucker nel caso di funzioni convesse). *Sia $\tilde{\mathbf{x}}$ una possibile soluzione al problema 2.1. Siano f, g_i, h_j , per $i = 1, \dots, n$ e $j = 1, \dots, m$, funzioni differenziabili e convesse e siano h_j funzioni affini. Si supponga che esistano $\tilde{\lambda} \geq \mathbf{0}$ e $\tilde{\mu} \in \mathbb{R}^m$ tali per cui*

$$\mathcal{L}(\tilde{\mathbf{x}}, \tilde{\lambda}, \tilde{\mu}) = \nabla f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \tilde{\lambda}_i \nabla g_i(\tilde{\mathbf{x}}) + \sum_{j=1}^m \tilde{\mu}_j \nabla h_j(\tilde{\mathbf{x}}) = \mathbf{0} \quad (2.4)$$

$$\tilde{\lambda}_i g_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, \dots, n.$$

Allora il punto $\tilde{\mathbf{x}}$ risolve la 2.1.

Dimostrazione. Sia $\tilde{\mathbf{x}}$ una soluzione al problema 2.1. Essendo la funzione $\omega(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^n \tilde{\lambda}_i g_i(\mathbf{x}) + \sum_{j=1}^m \tilde{\mu}_j h_j(\mathbf{x})$ convessa, per determinare quel punto $\tilde{\mathbf{x}}$ tale per cui la funzione sia minima è necessario calcolarne il gradiente come $\nabla \omega(\mathbf{x}) = \nabla f(\mathbf{x}) + \sum_{i=1}^n \tilde{\lambda}_i \nabla g_i(\mathbf{x}) + \sum_{j=1}^m \tilde{\mu}_j \nabla h_j(\mathbf{x})$. Ne segue che quel $\tilde{\mathbf{x}}$ che risolve $\nabla \omega(\tilde{\mathbf{x}}) = \mathbf{0}$ è il punto di minimo e che $\omega(\tilde{\mathbf{x}}) \leq \omega(\mathbf{x})$.

Si giunge quindi alla conclusione nel seguente modo.

$$\begin{aligned}
 f(\tilde{\mathbf{x}}) &= f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \tilde{\lambda}_i g_i(\tilde{\mathbf{x}}) + \sum_{j=1}^m \tilde{\mu}_j h_j(\tilde{\mathbf{x}}) \\
 &= \omega(\tilde{\mathbf{x}}) \\
 &\leq \omega(\mathbf{x}) \\
 &= f(\mathbf{x}) + \sum_{i=1}^n \tilde{\lambda}_i g_i(\mathbf{x}) + \sum_{j=1}^m \tilde{\mu}_j h_j(\mathbf{x}) \\
 &\leq f(\mathbf{x}).
 \end{aligned}$$

Si è dimostrato che $f(\tilde{\mathbf{x}}) \leq f(\mathbf{x})$. □

Per concludere, le condizioni di Karush–Kuhn–Tucker sono condizioni necessarie per determinare il punto di ottimo ma non sono necessarie e sufficienti. Se la funzione obiettivo f e i vincoli $g_i(\cdot)$ risultano essere differenziabili e convesse e $h_j(\cdot)$ sono funzioni lineari, le condizioni di KKT sono necessarie e sufficienti.

2.2.1 Semplice interpretazione geometrica del teorema di KKT

Sia $\mathcal{I} = \{i : g_i(\tilde{\mathbf{x}}) = 0\}$ e sia la condizione di KKT $\nabla f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \lambda_i \nabla g_i(\tilde{\mathbf{x}}) + \sum_{j=1}^m \mu_j \nabla h_j(\tilde{\mathbf{x}}) = 0$ ([10]). Si può allora riscrivere tale condizione come

$$-\nabla f(\tilde{\mathbf{x}}) = \sum_{i \in \mathcal{I}} \lambda_i \nabla g_i(\tilde{\mathbf{x}}) + \sum_{j=1}^m \mu_j \nabla h_j(\tilde{\mathbf{x}}).$$

Allora, l'opposto del gradiente di f non è altro che una combinazione lineare del gradiente di g_i , con $i \in \mathcal{I}$, e del gradiente di h_j . (Figura tratta da [6].)

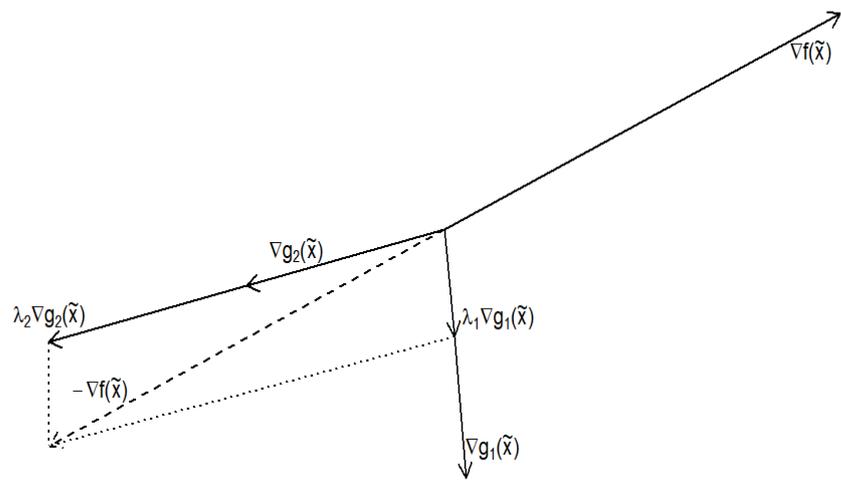


Figura 2.2: Interpretazione geometrica del teorema di KKT.

Capitolo 3

Caso pratico del teorema di Karush–Kuhn–Tucker

Fino a questo momento si sono date le basi matematiche per l'ottimizzazione di funzioni in generale. Un semplice esempio per testare quanto detto può essere dato dalla seguente formulazione:

Sia $f(x_1, x_2) = 9x_1^2 + 4x_2^2$ la funzione obiettivo e siano $x_2 \geq -\frac{1}{2}x_1 + 2$, $x_2 \leq \frac{1}{2}x_1$ i vincoli.

La funzione $f(x_1, x_2)$ è mostrata in Figura 3.1.

Si noti innanzitutto che la funzione obiettivo è una funzione convessa (si veda la definizione 2.3), con derivate parziali continue, inoltre i vincoli risultano essere lineari (si veda il punto 1. della qualificazione dei vincoli) e con derivate parziali continue. Per poter utilizzare il teorema di Karush–Kuhn–Tucker è necessario che tutti i vincoli siano del tipo $g \leq 0$. Allora

$$g_1(x_1, x_2) = -\frac{1}{2}x_1 + 2 - x_2 \leq 0,$$

$$g_2(x_1, x_2) = -\frac{1}{2}x_1 + x_2 \leq 0.$$

La forma precedente ricorda molto quella vista nel punto 2. del capitolo introduttivo in cui

$$V = \begin{bmatrix} -\frac{1}{2} & -1 \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -2 \\ 0 \end{bmatrix},$$

che in questo caso viene scritta come disequaglianza $V\mathbf{x} \leq \mathbf{b}$.

Quando si ha di fronte un problema di questo tipo conviene procedere per casi. Innanzitutto i vincoli sono rappresentati da disequazioni ed entrambe rappresentano un piano in \mathbb{R}^2 , come mostrato in Figura 3.2. La regione evidenziata rappresenta l'insieme dei punti che soddisfano i due vincoli.

- Si supponga che i vincoli g_1 e g_2 siano entrambi attivi e quindi che $g_1(x_1, x_2) = 0$ e $g_2(x_1, x_2) = 0$. *In primis* si valuta se i gradienti dei vincoli sono o meno linearmente indipendenti:

$$\nabla g_1 = \begin{bmatrix} -\frac{1}{2} \\ -1 \end{bmatrix}, \quad \nabla g_2 = \begin{bmatrix} -\frac{1}{2} \\ 1 \end{bmatrix}.$$

Viene dunque creata la matrice

$$A = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ -1 & 1 \end{bmatrix}$$

e si procede con l'algoritmo di Gauss come segue

$$\begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ -1 & 1 \end{bmatrix} \xrightarrow{E_1(-2)} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \xrightarrow{E_{21}(1)} \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \xrightarrow{E_2(\frac{1}{2})} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = B.$$

Avendo la matrice B due *pivot*, pari a $b_{11} = 1$ e $b_{22} = 1$, il suo rango è pari a 2 e quindi i due vettori di partenza ∇g_1 e ∇g_2 sono linearmente indipendenti.

Procedendo con il calcolo della lagrangiana,

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2) = 9x_1^2 + 4x_2^2 + \lambda_1 \left(-\frac{1}{2}x_1 + 2 - x_2 \right) + \lambda_2 \left(-\frac{1}{2}x_1 + x_2 \right)$$

e di conseguenza le derivate parziali risultano

$$\frac{\partial \mathcal{L}}{\partial x_1} = 18x_1 - \frac{1}{2}\lambda_1 - \frac{1}{2}\lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 8x_2 - \lambda_1 + \lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = -\frac{1}{2}x_1 + 2 - x_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = x_2 - \frac{1}{2}x_1.$$

In termini matriciali tali quantità possono essere viste come

$$A = \begin{bmatrix} 18 & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 8 & -1 & 1 \\ -\frac{1}{2} & -1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} x_1 \\ x_2 \\ \lambda_1 \\ \lambda_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ -2 \\ 0 \end{bmatrix}$$

che portano alle soluzioni

$$A\boldsymbol{\beta} = \mathbf{b} \implies \boldsymbol{\beta} = A^{-1}\mathbf{b} = \begin{bmatrix} x_1 \\ x_2 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 40 \\ 32 \end{bmatrix}.$$

Andando a sostituire x_1 e x_2 in g_1 e g_2 si ottiene

$$g_1(2, 1) = 0,$$

$$g_2(2, 1) = 0.$$

Quindi le condizioni tali per cui $g_1 \leq 0$ e $g_2 \leq 0$ sono soddisfatte, i moltiplicatori risultano entrambi strettamente positivi e infine $\lambda_1 g_1(2, 1) = 0$ e $\lambda_2 g_2(2, 1) = 0$ come da ipotesi. Tutte le ipotesi del teorema di KKT sono soddisfatte. Il punto ammissibile trovato è allora $(2, 1)$.

- Si ipotizzi ora che nessuno dei due vincoli sia attivo; la lagrangiana dunque sarà

$$\mathcal{L}(x_1, x_2) = 9x_1^2 + 4x_2^2$$

e le derivate parziali

$$\frac{\partial \mathcal{L}}{\partial x_1} = 18x_1 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 8x_2 = 0,$$

da cui si ricava che $x_1 = 0$ e $x_2 = 0$ e quindi, di conseguenza, $\lambda_1 = 0$ e $\lambda_2 = 0$ non essendo attivo nessun vincolo e quindi la regione ammissibile è vuota.

- Sia ora solamente il vincolo g_1 attivo. La funzione di Lagrange è

$$\mathcal{L}(x_1, x_2, \lambda_1) = 9x_1^2 + 4x_2^2 + \lambda_1 \left(-\frac{1}{2}x_1 + 2 - x_2 \right)$$

e le derivate parziali

$$\frac{\partial \mathcal{L}}{\partial x_1} = 18x_1 - \frac{1}{2}\lambda_1 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 8x_2 - x_1 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = -\frac{1}{2}x_1 + 2 - x_2 = 0$$

la cui forma matriciale è

$$\begin{bmatrix} 18 & 0 & -\frac{1}{2} \\ 0 & 8 & -1 \\ -\frac{1}{2} & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}$$

dalla quale si ricavano le quantità $x_1 = \frac{2}{5}$, $x_2 = \frac{9}{5}$, $\lambda_1 = \frac{72}{5}$ e $\lambda_2 = 0$.

Valutandole come nei punti precedenti si ha che

$$g_1 \left(\frac{2}{5}, \frac{9}{5} \right) = 0,$$

$$\lambda_1 g_1 \left(\frac{2}{5}, \frac{9}{5} \right) = 0.$$

Il punto $\left(\frac{2}{5}, \frac{9}{5} \right)$ è un possibile punto di minimo relativo.

- Infine utilizzando solamente il vincolo g_2 come vincolo attivo si ottiene la seguente lagrangiana

$$\mathcal{L}(x_1, x_2, \lambda_2) = 9x_1^2 + 4x_2^2 + \lambda_2 \left(x_2 - \frac{1}{2}x_1 \right)$$

le cui derivate parziali sono

$$\frac{\partial \mathcal{L}}{\partial x_1} = 18x_1 - \frac{1}{2}\lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 8x_2 + \lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = x_2 - \frac{1}{2}x_1 = 0.$$

Ottenendo sempre la forma matriciale si ricava

$$\begin{bmatrix} 18 & 0 & -\frac{1}{2} \\ 0 & 8 & 1 \\ -\frac{1}{2} & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

38CAPITOLO 3. CASO PRATICO DEL TEOREMA DI KARUSH-KUHN-TUCKER

da cui $x_1 = 0$, $x_2 = 0$, $\lambda_2 = 0$ e $\lambda_1 = 0$, che sostituiti portano a

$$g_2(0, 0) = 0,$$

$$\lambda_2 g_2(0, 0) = 0$$

Anche il punto $(0, 0)$ è un possibile candidato.

Vengono quindi valutati i punti ottenuti nella funzione f ottenendo

$$f(2, 1) = 40,$$

$$f\left(\frac{2}{2}, \frac{9}{5}\right) = \frac{72}{5},$$

$$f(0, 0) = 0.$$

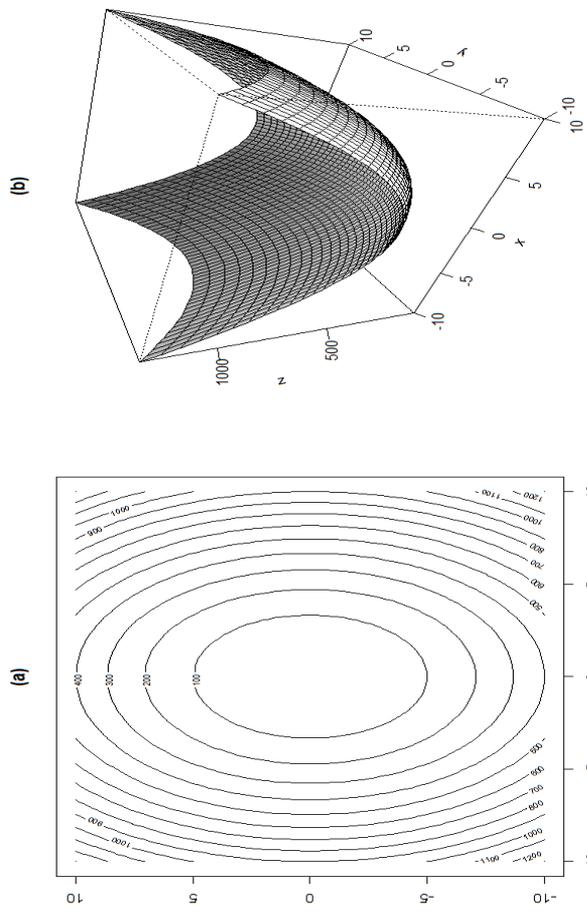


Figura 3.1: In (a) le curve di livello, in (b) il grafico tridimensionale in prospettiva.

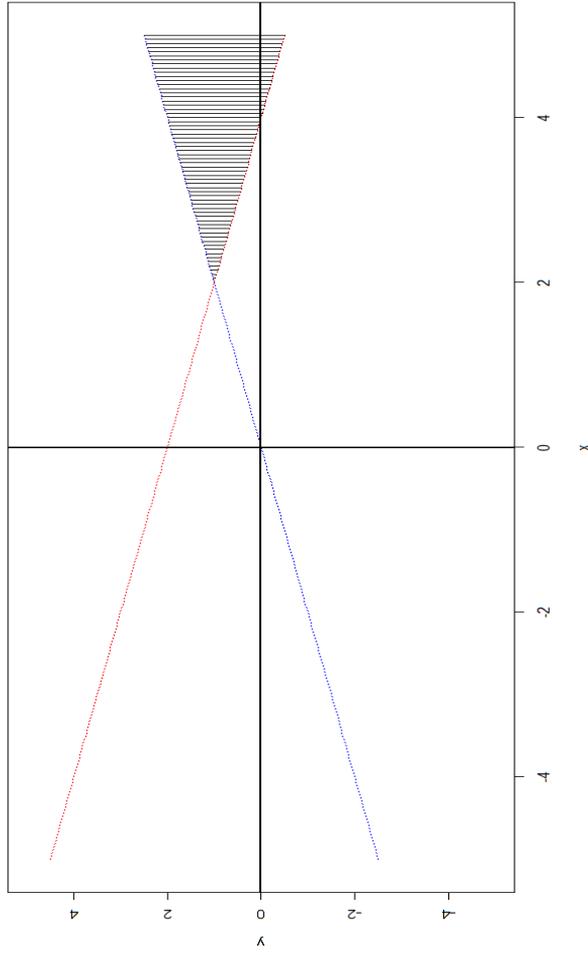


Figura 3.2: In rosso l'equazione $-\frac{1}{2}x_1 + 2 = 0$, in blu l'equazione $\frac{1}{2}x_1 = 0$.

Conclusioni

In tale scritto si sono volute dare nozioni di base e avanzate di cosa significhi ottimizzare una funzione, dalla più semplice alla più complessa. Si è partiti dalle funzioni reali, in una variabile, per giungere infine a quelle in più variabili ove gli accorgimenti risultano certamente molteplici e più intricati ma gli strumenti messi a disposizione, come il gradiente e l'hessiana *in primis* per poi passare ai moltiplicatori di Lagrange e le condizioni di Karush–Kuhn–Tucker, riescono a 'semplificare' il processo di ottimizzazione. È inoltre importante sapere perché le condizioni del teorema di KKT vengono definite in quel modo con quei vincoli specifici.

In particolare la differenza e la difficoltà risiedono nell'aggiunta nel problema di ottimo delle due tipologie di vincoli che possono risultare utili in certe situazioni. Infine il caso pratico ha dimostrato come non sia del tutto complesso ottimizzare una funzione.

Codici R

Funzione x^3

```
1 x11(6, 6, 12)
2 x3 <- function(x) x^3
3 curve(x3, -5, 5, las = 1, ylab = expression(f(x)),
4       lwd = 2)
5 points(x = 0, y = 0, pch = "X", col = 2, lwd = 2,
6        cex = 1.3)
7 legend("bottomright", legend = c("f '(x)=0"),
8        pch = "X", cex = 1.3, col = 2, bty = "n")
```

Paraboloide ellittico

```
1 x <- y <- seq(-10, 10, length = 50)
2 z <- outer(x, y, function(x, y) x^2 + y^2)
3
4 x11(6, 6, 12)
5 par(mfcol=c(1, 2))
6 contour(x, y, z, main = "(a)", xlab = "x", ylab = "y", las = 1)
7 points(x = 0, y = 0)
8 persp(x, y, z, theta = 30, phi = 30, main = "(b)", ticktype = "
  detailed", shade = .3)
```

Paraboloide iperbolico

```

1 z. <- outer(x, y, function(x, y) x^2 - y^2)
2
3 x11(6, 6, 12)
4 par(mfcol=c(1, 2))
5 contour(x, y, z., main = "(a)", xlab = "x",
6         ylab = "y", las = 1)
7 points(x = 0, y = 0)
8 persp(x, y, z., theta = 30, phi = 30,
9        main = "(b)", ticktype = "detailed",
10       shade = .3, zlab = "z")

```

Funzione $9x_1^2 + 4x_2^2$

```

1 x <- y <- seq(-10, 10, length = 50)
2 z.. <- outer(x, y, function(x, y) 9*x^2+4*y^2)
3
4 x11(6, 6, 12)
5 par(mfcol=c(1, 2))
6 contour(x, y, z.., main = "(a)")
7 persp(x, y, z.., theta = 30, phi = 30, shade = .3, main = "(b)",
8       zlab = "z", ticktype = "detailed")

```

Vincoli $-\frac{1}{2}x_1 + 2 = 0$ e $\frac{1}{2}x_1 = 0$

```

1 x11(6, 6, 12)
2 curve(-1/2*x + 2, xlim = c(-5,5), ylim = c(-5,5), lty = "dotted",
3       ylab = "y", col = "red")
3 curve(1/2*x, add = T, lty = "dotted", col = "blue")

```

```

4 abline(h = 0, lwd = 2)
5 abline(v = 0, lwd = 2)
6 a <- seq(2, 5, by = 0.05)
7 for(i in a){
8   segments(x0 = i, y0 = 1/2*i, x1 = i, y1 = -1/2*i+2, col = 1)
9 }

```

Interpretazione geometrica

```

1 library("latex2exp")
2 x1 <- y1 <- c(0, -1, -2)
3 x11(6, 6, 12)
4 {
5   plot(x1, y1, xlab = "", ylab = "", pch = ".", xlim = c(-2.2, 2)
6     , ylim = c(-2.3, 2), lwd = 2, axes = F)
7   arrows(0, 0, -2, -2, length = .18, lty = "dashed", lwd = 2)
8   text(1.85, 1.65, cex = 1.5, labels = latex2exp("$\\nabla f(\\
9     \\tilde{x})$"))
10  arrows(0, 0, 2, 2, length = .18, lwd = 2)
11  text(-1.75, -1.47, cex = 1.5, labels = latex2exp("$-\\nabla f
12    (\\tilde{x})$"))
13  arrows(0, 0, -1, -.52, length = .18, lwd = 2)
14  text(-.75, -.18, cex = 1.5, labels = latex2exp("$\\nabla g_2(\\
15    \\tilde{x})$"))
16  arrows(-1, -.52, -2, -1.04, length = .18, lwd = 2)
17  text(-2, -.78, cex = 1.5, labels = latex2exp("$\\lambda_2\\
18    \\nabla g_2(\\tilde{x})$"))
19  arrows(0, 0, .05, -.99, length = .18, lwd = 2)
20  text(.27, -.8, cex = 1.5, labels = latex2exp("$\\lambda_1\\
21    \\nabla g_1(\\tilde{x})$"))
22  arrows(.05, -.99, .11, -2.3, length = 0.18, lwd = 2)

```

```
17 text(.24, -2, cex = 1.5, labels = latex2exp("$\\nabla g_1(\\  
    tilde{x})$"))  
18 segments(-2, -1.04, -2, -2, lty = "dotted", lwd = 2)  
19 segments(.05, -.99, -2, -2, lty = "dotted", lwd = 2)  
20 }
```

Bibliografia

- [1] Mokhtar S Bazaraa, Hanif D Sherali e Chitharanjan M Shetty. *Nonlinear programming: theory and algorithms*. John Wiley & Sons, 2013.
- [2] Amir Beck. *Introduction to Nonlinear Optimization. The KKT Conditions*. URL: https://web.iem.technion.ac.il/images/user-files/becka/lec11_kkt_layers.pdf.
- [3] Michiel Bertsch, Roberta Dal Passo e Lorenzo Giacomelli. *Analisi matematica*. McGraw-Hill, 2007.
- [4] Dean Corbae, Maxwell B. Stinchcombe e Juraj Zeman. *An introduction to mathematical analysis for economic theory and econometrics*. Princeton University Press, 2009.
- [5] Morgan A. Hanson. «On sufficiency of the Kuhn-Tucker conditions». In: *Journal of Mathematical Analysis and Applications* 80.2 (1981), pp. 545–550.
- [6] R. Huňady e P. Pavelka. «Geometric Optimization of a Tuning Fork in NX Nastran». In: *American Journal of Mechanical Engineering* (2016), pp. 320–324.
- [7] Tinne Hoff Kjeldsen. «A contextualized historical analysis of the Kuhn–Tucker theorem in nonlinear programming: the impact of World War II». In: *Historia mathematica* 27.4 (2000), pp. 331–361.

- [8] Harold W Kuhn. «Nonlinear programming: a historical note». In: (1991).
- [9] *Lagrange multipliers, examples*. 2019. URL: <https://www.khanacademy.org/math/multivariable-calculus/applications-of-multivariable-derivatives/constrained-optimization/a/lagrange-multipliers-examples>.
- [10] Mahmut Parlar. *Interactive operations research with Maple: methods and models*. Springer Science & Business Media, 2012.
- [11] Michael Thomas e Sepideh Stewart. «Eigenvalues and eigenvectors: Embodied, symbolic and formal thinking». In: *Mathematics Education Research Journal* 23 (set. 2011), pp. 275–296.
- [12] William F Trench. «The Method of Lagrange Multipliers». In: *S. Trauductor, Trad.* (2013).
- [13] Wikipedia. *Alessio Figalli* — *Wikipedia, L'enciclopedia libera*. [Online; controllata il 2 luglio 2019]. 2019. URL: https://it.wikipedia.org/wiki/Alessio_Figalli.
- [14] James Hardy Wilkinson. *The algebraic eigenvalue problem*. Vol. 662. Oxford Clarendon, 1965.