UNIVERSITÁ DEGLI STUDI DI PAVIA
Dip. Ricerche Aziendali
QUADERNI DI RICERCA

# Interazioni tra economia e matematica: la nascita della programmazione matematica

**Giorgio Giorgi**

Ordinario di Matematica Generale - Facoltà di Economia – Università degli Studi di Pavia
Via San Felice,5 27100 Pavia, Italia - Phone: 0382.986238 - Fax (office): 0382.986228
eco.unipv.it - Email: ggiorgi@eco.unipv.it



## 1 – Introduzione

Da sempre, tra matematica e scienze della natura, fisica in particolare, c'è un rapporto di feconda interazione che sta alla base di alcune delle più profonde idee e concettualizzazioni della nostra conoscenza del mondo esterno. Più recente è invece un analogo rapporto tra matematica ed altre discipline, la cui dinamica essenziale è data dal processo di matematizzazione di diverse scienze nel corso del secolo appena concluso. Particolarmente significativo è stato il processo di matematizzazione che ha avuto luogo in Economia. Parecchi problemi dell'analisi economica hanno a che fare naturalmente con questioni di ottimizzazione che tradizionalmente sono affrontabili e risolvibili in termini matematici.

Difatti, nel corso degli ultimi due secoli gli studiosi hanno proposto numerose definizioni di Economia (o meglio di Teoria Economica, ovvero Analisi economica), parecchie delle quali mettono in evidenza un forte legame tra problemi economici e problemi matematici di ricerca del massimo o del minimo. Una delle definizioni più note è quella proposta da Lionel Robbins (1932).

Nella seconda edizione del suo saggio [47] leggiamo (pag. 16): "Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses".

Questa definizione pone dunque l'accento particolarmente sulla relazione tra fini e mezzi *limitati* suscettibili di impieghi alternativi.

Secondo Paul Anthony Samuelson (1948) e precisamente a pag. 5 dell'ottava edizione del suo manuale [48], datata 1970, si può leggere: "Economics is the study of how men and society end up choosing, with or without the use of money, to employ scarse productive resources which could have alternative uses, to produce various commodities and distribute them for consumption, now or in the future, among various people and groups in society".

Come si vede, anche se assai più articolata di quella di Robbins, la definizione di Samuelson rispecchia quella dell'economista inglese.

Per altre definizioni può essere utile consultare il capitolo 2 di Schumpeter [51] (1954). Va comunque notato che quasi tutte le definizioni pongono l'accento sulla (ovvia) scarsità dei beni economici e sulla conseguente necessità del loro impiego ottimale, secondo i criteri di volta in volta stabiliti.

Ribadiamo quindi che capita di frequente, nell'Analisi Economica, di volere stabilire quando un'assegnata espressione raggiunge il valore massimo o il minimo; tipicamente si è interessati al massimo quando la suddetta espressione (o meglio: funzione) ha il significato di utilità o di profitto, al minimo quando la medesima rappresenta un costo.

L'insieme degli strumenti matematici che permettono di trattare e spesso di risolvere, questi problemi va sotto il nome di *Teoria dell'Ottimizzazione*. È opportuno distinguere subito tra *problemi di ottimizzazione statica* e *dinamica*: i primi riguardano situazioni considerate in un dato istante ed in cui quindi si prescinde da qualunque variazione nel tempo, nei secondi invece l'evoluzione temporale delle variabili considerate assume un rilievo determinante.

Qui ci occuperemo, dal punto di vista essenzialmente storico, dei problemi di ottimizzazione statica e soprattutto di quei problemi che costituiscono l'oggetto della cosiddetta *programmazione matematica*, ossia di quei problemi in cui occorre trovare i valori ottimali di una funzione ("funzione obiettivo"), le cui variabili indipendenti (o decisionali) sono soggette a soddisfare determinati vincoli espressi, classicamente, da uguaglianze o, in termini più moderni e più generali, sotto forma di disuguaglianze e uguaglianze, con l'eventuale presenza di un vincolo insiemistico.

Lo stesso termine "programmazione matematica" è dovuto all'economista Robert Dorfman (si veda [5]), autore, assieme ai premi Nobel per l'Economia R. Solow e P.A. Samuelson, di un testo, divenuto ormai un "classico", sulla Programmazione Lineare ed Analisi Economica [10]. I termini "ottimizzazione", "programmazione matematica" e sinonimi entrano oramai nei titoli di parecchi libri ed articoli dedicati all'Analisi Economica: si vedano i riferimenti bibliografici [2, 9, 25, 26, 29, 39, 52].

La lezione tenuta da Samuelson in occasione del conferimento del premio Nobel (1970) aveva il significativo titolo "Maximum principles in analytical economics" [49]. Non dobbiamo poi dimenticare le altre applicazioni della programmazione matematica, soprattutto nell'ambito della Ricerca Operativa, delle Scienze Manageriali, della Statistica, delle Scienze Informatiche e Computazionali, della Fisica, dell'Ingegneria e perfino della Biologia e delle Scienze Medico-Chirurgiche!

Vale la pena di ricordare che una delle prime bibliografie di libri dedicati alla programmazione matematica è apparsa su Management Science [26] (si veda anche la rassegna [11], pubblicata sulla medesima rivista) e che, ad esempio, il primo volume di "Handbooks in Operations Research and Management Science" (a cura di G.L. Nemhauser e A.H.G. Rinnooy Kan [40]) si intitola semplicemente "Optimization".

Scendendo un po' di più nei dettagli, si parla, in particolare, di *programmazione non lineare* quando le funzioni obiettivo e di vincolo sono qualsiasi (e quindi non necessariamente lineari o affini), di *programmazione lineare* quando le suddette funzioni appartengono alla classe delle funzioni lineari o affini. Il primo articolo che utilizza esplicitamente il termine "non linear programming" è di H.W. Kuhn e A.W. Tucker nel 1951 [36]. Il termine "linear programming" è stato suggerito a G.B. Dantzig dall'economista (e premio Nobel) T.C. Koopmans (si veda [5]), curatore di uno dei volumi più importanti sulle interazioni tra Economia e Programmazione Matematica: quell'*Activity Analysis of Production and Allocation* [32], pubblicato nella prestigiosa collana della Cowles Foundation for Research in Economics e che conteneva, tra l'altro, il celebre lavoro di G.B. Dantzig sul metodo del simplesso ("Maximization of a linear function of variables subject to linear inequalities").

Altre classificazioni riguardano gli speciali problemi oggetto della programmazione matematica; avremo allora, ad esempio, la *programmazione intera*, se le variabili in giuoco sono discrete, la *programmazione stocastica* se i problemi contengono variabili aleatorie, la *programmazione vettoriale* o *Paretiana*, se la funzione obiettivo è a valori vettoriali, ecc.

Può allora apparire sorprendente che una teoria così centrale, così ricca e profonda per quanto concerne la strumentazione matematica usata, così dotata di un notevole interscambio con altre discipline, dia luogo ad una storia relativamente recente: poco più di 50 anni.

La ricostruzione storica della programmazione matematica, soprattutto non lineare, che qui vogliamo proporre, metterà in evidenza sia le difficoltà "tecniche", che possono averne ritardato lo sviluppo, sia il contesto economico e sociale (e militare) che ha favorito il suo radicamento negli anni dell'ultimo conflitto mondiale e soprattutto in quelli successivi.

Per quanto riguarda le “difficoltà” di natura matematica, anticipiamo subito che sfortunatamente le tecniche dell'Analisi Matematica classica non sono usualmente applicabili ai problemi di programmazione matematica, poiché le loro soluzioni di solito sono sulla frontiera dell'insieme ove le variabili indipendenti possono assumere i loro valori ("insieme ammissibile").

Nonostante una storia di poco più di mezzo secolo, la programmazione matematica ha avuto uno sviluppo veramente impressionante, di tipo esponenziale e non solo con riferimento a questioni teoriche ma anche e soprattutto con riferimento ad aspetti computazionali ed ad innumerevoli applicazioni. Basti osservare che la sezione "90" di Mathematical Reviews che fino a qualche anno fa si intitolava "Economics, Operations Research, Programming, Games", ora è interamente dedicata a "Operations Research, Mathematical Programming", mentre la sezione 91 raccoglie i lavori su "Game Theory, Economics, Social and Behavioral Sciences".

Vediamo ora, nel paragrafo successivo, alcune definizioni, concetti e risultati fondamentali riferiti per semplicità a spazi reali euclidei ad n dimensioni.

## 2 – Alcuni risultati fondamentali

Il più semplice schema di un problema di programmazione matematica può essere descritto tramite la massimizzazione o minimizzazione di una funzione $f\colon R^n \to R$ in un certo insieme A, sottoinsieme proprio o improprio del dominio di $f$ :

$$(P_1) \qquad \underset{x \in A}{\text{Max}}\ f(x) \quad \text{oppure} \quad \underset{x \in A}{\text{Min}}\ f(x)$$

La regione A viene chiamata insieme ammissibile del problema (o anche campo di scelta); la funzione $f$ viene detta funzione obiettivo e le variabili $x_1, x_2, \dots, x_n$ vengono talvolta chiamate variabili di decisione. In realtà, ci si può limitare a considerare uno solo dei due problemi indicati (ad esempio quello di massimo) in quanto è $\max f(x) = -\min\{-f(x)\}$.

Alcuni risultati generali sono ben noti.

1) *Teorema di Weierstrass.* Sia $f\colon R^n \to R$ definita su $X \subseteq R^n$. Se $X$ è compatto e $f$ è continua su $X$, allora $f$ ammette massimo e minimo su $X$.

Di tale fondamentale risultato (che comunque fornisce una condizione solo sufficiente) è possibile dare anche una versione più generale, particolarmente utile quando si sia interessati alla ricerca dei soli valori di massimo (minimo): se $f$ è superiormente (inferiormente) semicontinua sul compatto $X$, allora $f$ ammette massimo (minimo) su $X$.

2) Sia $f\colon R^n \to R$ definita su $X \subseteq R^n$. Se $X$ è un insieme convesso e $f$ è concava su $X$, ossia $\forall x^1, x^2 \in X$ e $\forall t \in [0,1]$ soddisfa la disuguaglianza $f\left[tx^1 + (1-t)x^2\right] \geq tf\left(x^1\right) + (1-t)f\left(x^2\right)$ allora:

a) i punti di massimo locale di $f$ sono anche di massimo globale;

b) l'insieme dei punti di massimo globale di $f$ è un sottoinsieme convesso di X;

c) se in particolare $f$ è strettamente concava su X, l'eventuale punto di massimo è unico.

Se il problema $(P_1)$ consiste nella ricerca dei punti e dei valori di massimo della funzione obiettivo $f$, allorché la regione ammissibile A è un *insieme aperto,* ovvero il vettore delle variabili decisionali assume valori solo in punti interni ad A, parleremo di problemi di *programmazione matematica libera* o di *problemi di ottimo libero.* In caso contrario, parleremo di *problemi di ottimo vincolato* (o di problemi di programmazione matematica *tout-court).*

Se $(P_1)$ è un problema di ottimo libero, i seguenti risultati sono familiari.

3) Sia $x^o \in \mathrm{A}$ soluzione di $(\mathrm{P}_1)$ e sia $f$ differenziabile sull'aperto $A \subseteq R^{\mathrm{n}}$. Allora necessariamente è $\nabla f(x^o) = 0$ ossia $x^o$ è *punto stazionario* (o *punto critico*) per $f$ .

4) Sia $x^o \in \mathrm{A}$ punto stazionario per $f$ e sia $f$ due volte differenziabile con continuità sull'aperto A. Allora, denotata con $Hf(x^o)$ *la matrice Hessiana* di $f$ valutata in $x^o$,

a) se è $y^T Hf(x^o) y < 0, \forall y \in R^n \setminus \{0\},$ ossia se la forma quadratica rappresentata dal differenziale secondo di $f$ è definita negativa, allora $x^o$ e punto di massimo locale stretto per $f$ su A. Se è $y^T Hf(x^o) y > 0, \forall y \in R^n \setminus \{0\},$ allora $x^o$ è punto di minimo locale stretto per $f$ su A;

b) se la forma quadratica $y^T Hf(x^o) y$ è indefinita, allora $x^o$ non è né punto di massimo né di minimo per $f$ ;

c) se $y^T Hf(x^o) y$ è una forma quadratica semidefinita (negativa o positiva), occorrono invece ulteriori indagini per decidere la natura del punto stazionario $x^o$.

5) Se in $(\mathrm{P}_1)$ $f$ è differenziabile e concava sull'aperto e convesso $A \subseteq R^{\mathrm{n}}$, allora ogni suo punto stazionario e anche punto di massimo globale di $f$ su A.

Già i risultati 2) e 5) mettono in evidenza l'importanza della concavità nei problemi di ottimo. La richiesta che la funzione obiettivo sia concava (o convessa, nei problemi di minimo) è quanto mai usuale quando si vuole assegnare un carattere globale ad una proprietà ipotizzata o dimostrata solo localmente o si intende trasformare in sufficiente una condizione per ora solo necessaria. Proprio l'ottimizzazione è risultata una delle principali motivazioni, per la cosiddetta *concavità generalizzata* dove, tramite opportune estensioni della definizione di funzione concava (o convessa), si studiano classi funzionali più vaste. In particolare, il risultato 5) può essere indebolito ipotizzando che la funzione (differenziabile) sia *pseudoconcava* sull'aperto e convesso A ovvero, $\forall x, y \in \mathrm{A}$, soddisfi l'implicazione:

$$\nabla f(x)(y - x) \leq 0 \Rightarrow f(y) \leq f(x).$$

Una ulteriore generalizzazione, particolarmente significativa nei problemi di programmazione matematica, è costituita dalle funzioni *quasiconcave*, definite dalla disuguaglianza:

$$f[tx + (1-t)y] \geq \min\{f(x), f(y)\}; \ (\forall \mathrm{t} \in [0,1], \forall x, y \in \mathrm{A}, \mathrm{A} \text{ convesso}).$$

Se $f$ è differenziabile sull'aperto e convesso A, la quasiconcavità di $f$ viene equivalentemente caratterizzata dall'implicazione:

$$x, y \in \mathrm{A}, f(y) \geq f(x) \Rightarrow \nabla f(x)(y - x) \geq 0.$$

Si veda, ad esempio [1, 20].

Un primo tipo di problema di ottimo vincolato è quello ove l'insieme ammissibile A è individuato dalle soluzioni di un'equazione o di un sistema di equazioni:

(P2) $\underset{x \in S}{\text{Max}}\ f(x)$ oppure $\underset{x \in S}{\text{Min}}\ f(x)$

ove $S = \left\{ x \in A,\ h_j(x) = 0, j = 1,\dots, r < n \right\}$, $A \subseteq R^n$ è aperto; $f$, $h_j : R^n \to R$.

Tali problemi sono noti come problemi *classici di ottimo vincolato* e, anche nella terminologia, fanno riferimento all'opera di Lagrange (cui accenneremo nel prossimo paragrafo).

Se per (P2) introduciamo la cosiddetta *funzione lagrangiana:*

$$L(x, \lambda) = f - \lambda h = f(x) - \sum_{j=1}^{r} \lambda_j h_j(x)$$

ove i numeri $\lambda_j \in R, j = 1, \dots, r,$ sono detti *moltiplicatori di Lagrange,* sussistono i seguenti fondamentali risultati.

6) Sia $x^o \in S$ soluzione locale di (P2) e siano soddisfatte le seguenti ipotesi:

i) $f$ è differenziabile in $x^o$;

ii) le funzioni $h_j\ (j = 1, \dots, r)$ sono in $x^o$ differenziabili con continuità e la matrice jacobiana $\nabla h\left(x^o\right)$ ha rango pieno, ovvero i vettori $\nabla h_j\left(x^o\right)$ sono linearmente indipendenti (i vincoli sono cioè regolari).

Allora esiste un (unico) vettore $\lambda^o = \left(\lambda_1^o, \dots, \lambda_r^o\right)$ per cui $(x^o, \lambda^o)$ è punto stazionario per la funzione lagrangiana, ossia:

$$\nabla f\left(x^o\right) - \lambda^o \nabla h\left(x^o\right) = 0. \tag{1}$$

Generalmente, sotto determinate ipotesi, non particolarmente restrittive, $\lambda^o$ viene a rivestire un preciso significato, in quanto le sue componenti $\lambda_j^o$ indicano la misura dell'effetto che una variazione marginale del j-esimo vincolo esercita sul valore ottimale della funzione obiettivo. Nei problemi economici la variazione marginale di un vincolo rappresenta spesso la variazione di quantità disponibile di una data risorsa, mentre la funzione obiettivo esprime il profitto o il costo; per queste ragioni, i moltiplicatori di Lagrange sono chiamati *prezzi ombra* (unitari) di quella data risorsa. Attraverso i valori $\lambda_j^o$ è possibile una valutazione economica del peso che ciascun vincolo assume nella definizione dell'ottimo del problema (P2).

7) Sia $x^o \in S$; la coppia $\left(x^o, \lambda^o\right)$ soddisfi la relazione (1) e sia $L(x,\lambda)$ funzione pseudoconcava rispetto al vettore $x$. Allora $x^o$ è punto di massimo globale di $f$ su $S$.

Il problema ($P_2$) è stato il primo problema di ottimo vincolato a ricevere attenzione. Quando, in tempi molto più vicini a noi (come si è già avuto modo di accennare), si è cominciato ad affrontare il caso in cui la regione ammissibile è determinata da vincoli anche espressi da disuguaglianze, lo si è fatto in due direzioni. Da un lato ci si è occupati di *programmazione lineare,* ossia di problemi in cui funzione obiettivo e vincoli sono dati da funzioni lineari o affini; dall'altro si è sviluppato il problema classico di ottimo vincolato con la stessa generalità:

$$(P_3) \qquad \underset{x \in K}{Max} f(x)$$

$$K = \{x \in A,\ g_i(x) \leq 0,\ i = 1, \dots m\},\ \text{ove } A \subseteq R^n \text{ è aperto; } f,\ g_i: R^n \to R.$$

Per tale problema l'insieme degli *indici dei vincoli attivi* in $x^o \in K$ è dato da:

$$I(x^o) = \{i : g_i(x^o) = 0\}$$

I due fondamentali risultati per $(P_3)$ sono tradizionalmente attribuiti a F. John e a H.W. Kuhn e A.W. Tucker.

*8) Teorema di F. John.* Siano $f$ e $g_i (i = 1, \dots m)$ differenziabili in $x^o \in K$, soluzione locale di $(P_3)$. Esiste allora un vettore $(y_0^o, y_1^o, \dots, y_m^o) \in R^{m+1}$ *a componenti non negative e non tutte nulle,* tale che:

$$\text{i)} \quad y_0^o \nabla f(x^o) - \sum_{i=1}^{m} y_0^o \nabla g_i(x^o) = 0;$$

$$\text{ii)} \quad y_i^o g_i(x^o) = 0,\ i = 1, \dots, m.$$

La dimostrazione del teorema si basa sostanzialmente sulla constatazione che l'esistenza di un punto ottimale implica che certi insiemi hanno intersezione vuota, risultando così impossibile un sistema di disuguaglianze lineari; "scatta" a questo punto un teorema dell'alternativa, che può essere interpretato geometricamente come esistenza di un iperpiano separatore nella cui equazione i parametri sono proprio i moltiplicatori $y_i^o, i = 0, \dots, m$.

Confrontando i risultati 3), 6) e il teorema di John (relativi, rispettivamente, ai problemi di ottimizzazione libera, P2) e P3)), emerge un filo conduttore sufficientemente stabile nelle condizioni necessarie del primo ordine. Le "novità" comunque non mancano. Ora otteniamo precise informazioni sul segno dei moltiplicatori. Inoltre le relazioni i), ii) possono essere compendiate nella seguente:

$$y_0^o \nabla f(x^o) - \sum_{i \in I(x^o)}^{m} y_i^o \nabla g_i(x^o) = 0$$

in cui si considerano solo i vincoli attivi e che, rispetto all'analoga (1) valida per il problema vincolato classico, presenta un moltiplicatore associato anche alla funzione obiettivo.

Lo scalare $y_0^o$ può essere nullo. Per evitare questo caso degenere in cui la funzione obiettivo non giocherebbe alcun ruolo, occorre imporre una condizione di regolarità dei vincoli, detta anche di qualificazione (dei vincoli). Una tale condizione (sufficiente) assume diverse forme, non tutte equivalenti e dotate di diversi gradi di generalità. Quella più affine alla condizione di regolarità dei vincoli, già vista a proposito del problema (P2), richiede che i vettori $\nabla g_i(x^o)$ $i \in I(x^o)$, siano linearmente indipendenti.

*9) Teorema di Kuhn-Tucker.* Sia $x^o \in K$ soluzione locale di (P3) nelle ipotesi che $f$ e $g_i (i=1,\dots,m)$ siano differenziabili in $x^o$. Se è soddisfatta una condizione di qualificazione dei vincoli, allora esiste un vettore $\lambda^o$ tale che:

$$\text{i)} \quad \nabla f(x^o) - \sum_{i=1}^{m} \lambda_i^o \nabla g_i(x^o) = 0;$$

$$\text{ii)} \quad \lambda_i^o g_i(x^o) = 0, \quad i = 1,\dots, m;$$

$$\text{iii)} \quad \lambda_i^o \geq 0, i = 1,\dots,m.$$

Il successivo teorema fornisce una condizione sufficiente per l'ottimalità di un vettore $x^o \in K$ che verifica le precedenti condizioni di Kuhn-Tucker.

10) Sia $x^o \in K$ un punto che soddisfa le condizioni i), ii), iii) del teorema 9). Sia $f$ pseudoconcava sull'insieme convesso $A \subseteq R^n$ e siano $g_i$, $i \in I(x^o)$, funzioni quasiconvesse, ovvero ogni $-g_i$ sia quasiconcava. Allora $x^o$ è soluzione di (P3).

Un'introduzione così schematica alla programmazione matematica non può naturalmente dare conto di tutte le teorie, anche classiche, che si sono sviluppate al suo interno. Faremo, in conclusione di questo paragrafo, due eccezioni.

La prima riguarda la *teoria della dualità,* che avremo spesso modo di richiamare successivamente e che permette di dedurre alcune interessanti caratteristiche di un problema (P) di programmazione matematica attraverso l'analisi di un altro problema, per così dire "speculare" a (P), chiamato problema *duale* e costruito a partire da (P) con certe regole.

La teoria della dualità è nata parallelamente agli studi sui problemi di programmazione lineare e a quelli sulla teoria dei giochi matriciali. Se (P) rappresenta il problema lineare:

$$\begin{cases} \text{Min } cx \\ \text{A}x \leqq b \\ x \geqq 0 \end{cases}$$

(dove $A$ è una matrice di ordine (m, n); $c$, $x \in \mathrm{R}^n$; $b \in \mathrm{R}^m$), il suo duale (P') è il problema:

$$\begin{cases} \text{Max } yb \\ y\text{A} \geqq c \\ y \geqq 0. \end{cases}$$

Al problema:

$$\begin{cases} \text{Min } \ cx \\ \text{A}x = b \\ x \geqq 0 \end{cases}$$

corrisponde invece il duale:

$$\begin{cases} \text{Max } \ yb \\ y\text{A} \geqq \text{c} \end{cases}$$

nel quale sono assenti i vincoli di non-negatività delle variabili.

In generale, dato un problema (P) di programmazione lineare, si possono notare le seguenti proprietà:

a) il duale di (P) è unico;

b) se (P') è il duale di (P), il duale di (P') è ancora (P);

c) se (P) è un problema di minimo, (P') è un problema di massimo (e viceversa) nel quale funzione obiettivo e di vincolo scambiano in qualche modo il proprio ruolo.

Tra un problema di programmazione lineare e il suo duale esistono legami profondi che sono alla base sia di sviluppi teorici che computazionali. Un primo risultato fondamentale afferma che, se (P) e (P') hanno entrambi regione ammissibile non vuota, allora entrambi i problemi ammettono ottimo e i rispettivi valori ottimali coincidono.

La teoria della dualità si è poi sviluppata anche con riferimento a problemi non lineari. Preferiamo però dedicare la conclusione del paragrafo accennando ad un modo necessariamente diverso di intendere l'ottimizzazione, già a partire dalle prime definizioni. Se la funzione obiettivo è rappresentata da una funzione *vettoriale* $f\colon \mathrm{R}^{\mathrm{n}} \to \mathrm{R}^{\mathrm{m}}$, abbiamo la classe dei *problemi di ottimo paretiano* (dal nome di Vilfredo Pareto che per primo li considerò) e più in generale di ottimo *multi-criteria.* In questo caso, prima ancora di considerare le condizioni necessarie e/o sufficienti di ottimalità, sarà fondamentale specificare la nozione di punto di ottimo, dato che lo spazio immagine è solo parzialmente ordinato e non è dunque possibile trasferire immediatamente al caso vettoriale la disuguaglianza $f(x) \le f(x^\circ)$ che caratterizzava i punti di massimo nel caso scalare. Si veda, ad esempio, [20, 50].

# 3 – La programmazione matematica "classica": il metodo dei moltiplicatori di Lagrange

Abbiamo già osservato che la moderna programmazione matematica è nata verso la metà del secolo scorso; tuttavia i problemi consistenti nel massimizzare o minimizzare una data funzione, soggetta ad un sistema di vincoli espressi da *uguaglianze* (e non quindi da *disuguaglianze*, come avviene nella programmazione matematica vera e propria) furono studiati da G.L. Lagrange nella seconda metà del 18° secolo.

Chiunque abbia un ricordo, sia pur vago, di un corso di Analisi Matematica ove si trattano le funzioni di più variabili reali, troverà non nuova l'espressione "moltiplicatori di Lagrange", oppure "metodo dei moltiplicatori di Lagrange".

E il nome è sufficiente per comprendere come in realtà la storia - o meglio la preistoria - della programmazione matematica, sia pure limitata al caso di vincoli espressi da uguaglianze, risalga alla fine del 1700. Qualcuno parla infatti anche di programmazione matematica "classica", di problemi di ottimo vincolato "classici", in opposizione alla moderna teoria dei problemi di ottimo i cui vincoli sono costituiti anche da disuguaglianze o addirittura da insiemi astratti. Con riferimento ai problemi di ottimo "classici" (cioè senza vincoli o con vincoli espressi da solo uguaglianze), segnaliamo il bel volume di Hancock [23], apparso nel 1917 e successivamente ristampato da Dover nel 1960.

Lagrange introduce i suoi "moltiplicatori" nel 1778, nella quarta sezione della prima parte del suo famoso libro *Mécanique Analytique*, quale strumento per determinare la configurazione di equilibrio stabile in un sistema meccanico. Si tratta, in particolare, di minimizzare la funzione potenziale, supponendo che essa esista e tenendo conto che il sistema - individuato nelle sue posizioni da $n + r$ coordinate - è soggetto ai vincoli $h_1 = h_2 = ... = h_r = 0$. È proprio “en différentiant ces équations” che si avrà subito $dh_1 = ... = dh_r = 0$ e, “comme ces équations ne doivent servir qu'à éliminer un pareil nombre de différentielles dans la formule générale de l'équilibre, après quoi les coefficients des différentielles restantes doivent être égalés chacun à zéro, il n'est pas difficile de prouver, par la théorie de l'élimination des équations linéaires, qu'on aura les mêmes résultats si l'on ajoute simplement à la formule dont il s'agit les différentes équations [...] multipliées chacune par un coefficient indéterminé $\lambda_j$”. È da questa osservazione, di carattere algebrico, che segue la regola "extrêmement simple" che permette di determinare le configurazioni di equilibrio. Si considera l'uguaglianza $df-\lambda_1 dh_1-...-\lambda_r dh_r=0$ dove $df$ rappresenta, nel linguaggio di Lagrange, “la somme des *moments* de toutes les puissances qui doivent être en equilibre”. Se a questo punto si scelgono i moltiplicatori $\lambda_j$ in modo che siano nulli i coefficienti di $dx_{n+j}$, rimane un'equazione in $dx_1,...,dx_n$ i cui coefficienti devono tutti annullarsi. Si sono così ottenute le $n$+$r$ condizioni (cui vanno aggiunte le $r$ uguaglianze $h_1=0,...,h_r=0$ nelle $n + 2r$ variabili $x_1,...,x_{n+r}$, $\lambda_1,...\lambda_r$. Introdotti come strumento algebrico, per avere che “le nombre de ces équations sera égal à celui de toutes les cordonnées des corps”, i moltiplicatori hanno anche un significato fisico. Lagrange, infatti, non discute il sistema delle $n + 2r$ equazioni nelle $n + 2r$ incognite ottenute tramite le “équations particulières de l'équilibre” - si limita a osservare che il valore dei moltiplicatori “pourra toujours exécuter par les moyens connus, mais il conviendra, dans chaque cas, de choisir ceux qui pourront conduire aux résultats les plus simples” - ma preferisce sottolineare che i vari $\lambda_j \; dh_j$ rappresentano “les moments de différentes forces appliquées au même systeme”. È la considerazione di queste forze, espresse dai vincoli, che permette di passare - diremmo oggi - ad un problema di ottimizzazione libera: “et de là on voit la raison metaphysique pourquoi l'introdution des termes” $\lambda_1 dh_1+...+\lambda_r dh_r$ nella condizione del primo ordine “peut ensuite traiter cette équation comme si tout les corps du système étaient entièrement libres”.

Nella *Théorie de fonctions analytiques (*1ª ed.: 1797) il metodo dei moltiplicatori viene presentato in tutta la sua generalità, non riferito ad alcuna specifica questione di Meccanica ma introdotto per i problemi di ottimizzazione, quando tra le variabili sussiste “une ou plusieurs équations” e si vuole semplificarne la soluzione senza ogni volta dover ricorrere all'eliminazione di alcune variabili utilizzando le equazioni di vincolo. Nel paragrafo 58 del capitolo XI della seconda parte, ritroviamo la dimostrazione della condizione necessaria, condotta sempre seguendo quello *standard* di rigore e quei principi che erano stati dichiarati nell'introduzione. Il paragrafo si chiude con l'enunciato della regola ("principe") generale: “il suffira d'ajouter à la fonction proposée les fonctions qui doivent être nulles, multipliées chacune par une quantité indéterminée, et de chercher ensuite le *maximum* ou *minimum* comme si les variables étaient indépendantes”.

## 4 – La nascita della programmazione matematica moderna

Dopo l'opera di Lagrange, nel periodo a cavallo tra Settecento ed Ottocento, un altro contributo importante è quello del matematico francese J. Fourier, che introduce nel problema del potenziale di Lagrange l'uso di disuguaglianze, anziché uguaglianze.

Siamo nel 1798 ed il lavoro di Fourier, nel quale avviene questa "apertura" nella direzione della programmazione matematica "moderna" è la *Mémoire sur la Statique*.

I risultati non sono però dimostrati, né in maniera rigorosa né in maniera completa. Lo stesso dicasi con riferimento ai lavori, sulle medesime questioni, di due allievi dello stesso Fourier, ossia C.L.M.H. Navier (1825) ed il famoso economista (e matematico!) A.A. Cournot (1827) che assegna la condizione necessaria per un caso speciale del problema di Fourier, ma senza fornire la dimostrazione rigorosa del risultato. Lo stesso teorema compare in termini generali, ma sempre privo di dimostrazione completa in un articolo del 1838 (ma scritto nel 1834) di Ostrogradsky [41] che, prima di fare ritorno a San Pietroburgo, aveva seguito a Parigi i corsi di Fourier.

Per il resto, per tutto l'Ottocento non ci sono significativi "ritorni di fiamma" verso le problematiche sollevate da Fourier, implicanti disuguaglianze anziché uguaglianze. Vengono al più citati i nomi di Gauss (1829), sempre in relazione a problemi di Meccanica, di Boole (1864), per qualche spunto che si trova nei suoi lavori di Logica e di Probabilità, e di Gibbs (1879) che considera problemi di dinamica con vincoli sotto forma di disuguaglianza.

L'interesse verso lo studio di sistemi di disuguaglianze *lineari* riprende nella seconda metà dell'Ottocento con un celebre lavoro di P. Gordan (1873) che è l'antesignano di tutti i risultati sui cosiddetti "teoremi dell'alternativa" per sistemi lineari e soprattutto con la pubblicazione di alcuni articoli del matematico ungherese Gyula Farkas. Il lavoro di Farkas verrà in seguito riconosciuto fondamentale proprio dai "padri" della programmazione non lineare H.W. Kulm e A.W. Tucker che dimostreranno, nel 1951, il loro celebre teorema, servendosi proprio di quello che nella nostra epoca è comunemente chiamato *teorema* (o lemma) *di Farkas* o *di Farkas-Minkowski*. Farkas pubblicò diverse dimostrazioni del suo teorema ma quella completa ed esatta apparve in ungherese nel 1898 e in tedesco nel 1899. Tale dimostrazione è ripresa nel successivo e più noto lavoro [14] del 1901. Il risultato di cui stiamo parlando è il seguente:

- Ogni soluzione $x \in \mathrm{R}^n$ del sistema lineare di disequazioni $\mathrm{A}x \geqq 0$, con A matrice di tipo (m, n), è anche soluzione della disequazione $bx \geq 0$ se e solo se esiste un vettore $y \in \mathrm{R}^m$, $y \geqq 0$, tale che $\mathrm{A}^T y = b$.

Tale risultato è equivalente al seguente, forse più utilizzato e che meglio mette in luce la natura di "teorema dell'alternativa" del risultato di Farkas:

- Il sistema $\mathrm{A}x \geqq 0$, $bx < 0$ ammette soluzione $x \in \mathrm{R}^n$ se e solo se il sistema $\mathrm{A}^T y = b$, $y \geqq 0$ non ammette soluzione $y \in \mathrm{R}^m$.

È quindi palese che mai i sistemi appena scritti possono entrambi ammettere soluzione o essere entrambi impossibili.

Un enunciato simile compare pressoché contemporaneamente nel libro di H. Minkowski "Geometrie der Zahlen" (1ª edizione: 1896).

Farkas non pubblicherà nessun altro lavoro sulle disequazioni lineari fino agli anni venti del secolo successivo.

Nel periodo tra le due guerre mondiali un altro matematico ungherese, A. Haar, generalizzerà il risultato di Farkas ai sistemi non omogenei (1924). Con la tesi di dottorato di Motzkin [38] nel 1933 si darà il primo effettivo riconoscimento dell'importanza delle ricerche di Farkas nell'ambito della teoria dei sistemi lineari di disuguaglianze.

Siamo oramai alla vigilia degli atti ufficiali di nascita della programmazione matematica, lineare e non.

Cogliamo l'occasione per fare osservare che in realtà il teorema di Farkas è il generatore di tutti i teoremi dell'alternativa per sistemi lineari, scoperti successivamente. Ciò è stato dimostrata dall'autore della presente nota in [17], lavoro ripreso ed ampliato in [7]. Non solo il teorema di Farkas è in grado di generare ben 225 diversi teoremi dell'alternativa per sistemi lineari, ma è addirittura possibile dimostrare che il primo teorema dell'alternativa pubblicato, ossia quello di Gordan (già richiamato e pubblicato nel 1873, quindi 25 anni prima della versione corretta del teorema di Farkas) implica a sua volta (ovviamente in modo non banale) il risultato di Farkas. Si veda [18]. Il teorema dell'alternativa di Gordan ha la seguente semplice formulazione.

- Il sistema $\mathrm{A}\ x > 0$ ammette soluzione $x \in \mathrm{R}^n$ se e solo se il sistema $\mathrm{A}^T y = 0$, $y \geqq 0$, $y \neq 0$, non ammette soluzione $y \in \mathrm{R}^m$.

Perché una disciplina e una teoria si sviluppino in un dato periodo storico o in un certo Paese, piuttosto che in un altro, è proprio una "bella" domanda. Quasi sempre troppo generale per ricevere risposte (anche parzialmente) soddisfacenti; opportuna e utile comunque, per porre il problema dell'ordine delle teorie scientifiche e non credere sempre e in ogni caso ai "miracoli" di una generazione che risulterebbe incontrollabile nella sua dinamica e insensibile ad ogni direttiva programmatica, rispondendo solo all'imprevedibile genialità umana. Nel caso della programmazione matematica, perché il lemma di Farkas è rimasto nel cassetto, inutilizzato per mezzo secolo? Perché fino agli anni '50 non registriamo alcun interesse consistente verso l'ottimizzazione, anche se la massimizzazione o la minimizzazione di una data quantità possono essere facilmente incluse tra le "curiosità naturali" dell'uomo (e l'economia matematica aveva già raggiunto da alcuni decenni un suo statuto scientifico)? Perché la programmazione lineare e non lineare, riceve il suo battesimo solo dopo la seconda guerra mondiale?

Parlavamo di risposte solo parziali. Nel nostro caso, la nascita della disciplina appare correlata allo sviluppo di determinati strumenti matematici e di altre teorie e alla maturazione di alcune situazioni socio-economiche. Relativamente ai primi è ovvio constatare come il ricorso al calcolo differenziale fosse possibile da tempo, anche per funzioni di più variabili reali e nella sua versione astratta, ma la stessa cosa non può dirsi per gli strumenti dell'analisi convessa, resi opportuni dall'introduzione delle disuguaglianze e per gli sviluppi e il radicamento dell'economia matematica. Quasi negli stessi anni, a cavallo della seconda guerra mondiale, diventa fattore propulsivo anche la percezione della complessità delle società industriali e delle interdipendenze in esse presenti. L'esperienza bellica, con la naturale presenza di precisi obiettivi da raggiungere, la necessità di considerare simultaneamente un gran numero di variabili e l'incombenza drammatica del fattore tempo, torna estremamente utile nella fase di riconversione, per andare oltre a tecniche escogitate per un singolo problema e stabilire invece metodi sufficientemente generali. Questioni economiche e militari costituiscono il naturale terreno di cultura, inizialmente, per la programmazione *lineare* e nessuno si sorprenderà che siano la ex URSS e gli Stati Uniti d'America i due Paesi che registrano i primi fondamentali sviluppi.

Per l'Unione Sovietica il padre della programmazione lineare è Leonid Vitalevic Kantorovich (1912-1996).

Leonid Vitalevic era una persona eccezionale, nel senso più completo del termine: si iscrisse alla Facoltà di Matematica di San Pietroburgo (allora Leningrado) all'età di quattordici anni, a diciotto era laureato e a ventidue aveva già una cattedra presso la stessa Facoltà. Successivamente: premio Stalin, premio Lenin, premio Nobel (si veda [43]). Premio Nobel, nel 1975 assieme al già ricordato T.C. Koopmans.

Il problema, di natura economico- manageriale, posto e risolto da Kantorovich nel 1939, è un caso particolare di problema di programmazione lineare che va sotto il nome di "problema del trasporto" o "problema dell'allocazione ottima".

Nella fattispecie, nel più semplice tra i problemi che Kantorovich considera, si tratta di allocare in modo ottimale m macchine che possono produrre n prodotti. Se $a_{ij}$ indica il numero di unità del j-esimo output prodotte nell'unità di tempo dall'i-esima macchina; $x_{ij}$ il numero di macchine dell'n-esimo tipo allocate per produrre il j-esimo output e $z_j$ il numero totale di unità prodotte, sempre del j-esimo output, avremo

$$z_j = \sum_{i=1}^{m} a_{ij} x_{ij}; \ \sum_{j=1}^{n} x_{ij} \leq b_i; \ x_{ij} \geq 0$$

dove $b_i$ rappresenta il numero di macchine disponibili del tipo i-esimo. Per fissare poi la funzione obiettivo, supponiamo che ciascun prodotto finale richieda esattamente un'unità di ogni output.

Se $z_0$ è il numero di articoli finiti, in questo primo schema la funzione da massimizzare sarà allora data da $z_0 = \min_j z_j$.

L'articolo (in russo) di Kantorovich sarà ignorato nel mondo occidentale (e non solo per problemi di lingua) fino al 1960, allorché T.C. Koopmans lo farà conoscere agli studiosi occidentali con la traduzione in inglese su "Management Science" [28].

Indipendentemente da Kantorovish, negli Stati Uniti il problema di programmazione lineare, detto "del trasporto", era stato affrontato nel 1941 da F.L. Hitchcock [24] e nel 1947 da T.C. Koopmans (ancora lui!) [31].

Abbiamo già avuto modo di accennare che Koopmans, all'inizio degli anni '50 del secolo scorso, aveva contribuito in modo significativo alla diffusione (soprattutto nell'ambito degli studi economici e di Ricerca Operativa) dei risultati di George Bernard Dantzig, il "padre occidentale" della Programmazione Lineare, disciplina che al suo nascere era per la verità maggiormente legata a ricerche svolte per fini militari e logistici.

Durante la guerra, dal '42 al '44, Koopmans (1910-1985) aveva lavorato come statistico per "The Allied Shipping Adjustement Board", occupandosi in particolare di un modello di trasporti. Dantzig (nato nel 1914) collaborava con il Pentagono come esperto di metodi di programmazione, sviluppati con l'uso di "desk calculators". Con la fine della guerra può terminare gli studi, ottenendo il dottorato di ricerca; le possibilità di "carriera" gli provengono da un'offerta dell'Università di Berkeley e dal prolungamento del proprio rapporto di lavoro con il Pentagono, come consulente matematico per l'aviazione. È nel tentativo di "non lasciarselo scappare" che due colleghi del Pentagono - D. Hitchcock e M. Wood - gli propongono di "mechanize the planning process" [5] che Dantzig aveva formalizzato, trovando i sistemi di disequazioni lineari come denominatore comune formale delle interdipendenze che legavano le quantità variabili in molte situazioni. Naturalmente la formalizzazione di un processo decisionale e la costruzione di un algoritmo vanno riferite alla tecnologia di un'epoca che deve aspettare ancora qualche anno per ospitare gli esordi della rivoluzione micro-elettronica e la comparsa dei computers: "in those days mechanizing planning meant using analog devices or punch-card equipment". Se il punto di riferimento iniziale è costituito per Dantzig dal modello *input-output* di Leontief, le esigenze militari e le richieste del committente ne sollecitano una generalizzazione che consideri anche il fattore tempo, che preveda inoltre la possibilità di scelta tra diverse alternative e soprattutto porti ad una modellizzazione computabile: "once the model was formulated then had to be a practical way to compute what quantities of these activities to engage in that was consistent with their respective input-output characteristics and with given resources. This would be no mean task since the military application had to be *large scale,* with hundreds and hundreds of items and activities".

Il *metodo del simplesso* - ideato da Dantzig per risolvere numericamente un problema di programmazione lineare - nasce all' interno di questa storia e viene presentato per la prima volta nell'estate del '47. Poco prima, nel giugno dello stesso anno, Dantzig aveva rivisto Koopmans, che subito intuisce le potenzialità del nuovo algoritmo e si impegna per la sua diffusione all'interno del gruppo di giovani economisti con i quali collabora; i loro nomi - K. Arrow, P. Samuelson, M. Simon, R. Dorfman, L. Hurwicz, H. Scarf - danno immediatamente un'idea dell'eccezionale contesto in cui la programmazione lineare muove i suoi primi passi. In autunno

Dantzig decide di consultare anche John von Neumann, a Princeton, e viene così "introdotto" alla teoria della dualità e alla problematica dei rapporti della programmazione lineare con la teoria dei giochi. Per questa teoria si veda il fondamentale testo [60]. Per essere più precisi, la teoria della programmazione lineare è equivalente alla teoria dei giochi discreti e a somma nulla, tra due avversari. Qualche mese ancora - per cominciare a sviluppare le potenzialità anche teoriche del metodo del simplesso, per altri incontri che non rimarranno senza conseguenze nella storia della ottimizzazione (ad esempio, con A.W. Tucker e i suoi allievi H.W. Kuhn e D. Gale) o per qualche presentazione in gruppi ancora ristretti di ricercatori - e poi ecco il "debutto in società". L'occasione è fornita da un primo convegno sulla programmazione matematica, organizzato da Koopmans a Chicago nel '49. Dantzig nei suoi ricordi enfatizza l'evento e il periodo che il convegno quasi viene a rappresentare, ma un semplice sguardo ai temi trattati e ai nomi dei principali relatori è sufficiente per avere conferma di quanto era maturato rapidamente, dopo la guerra: "the advent or rather *the promise* that the electronic computer would exist soon, the exposure of theoretical mathematicians and economists to real problems during the war, the interest in mechanizing the planning process, and last but not least the availability of money for such applied research all converged during the period 1947-1949. The time was ripe. The research accomplished in exactly two years is, in my opinion, one of the remarkable events of history". Come abbiamo già ricordato, gli atti verranno pubblicati dallo stesso Koopmans, nel '51, con una introduzione che sottolinea il ruolo svolto nella nascita della programmazione lineare da quattro distinte linee di ricerca: il dibattito e le generalizzazioni successivamente intervenute sul tema dell'equilibrio economico generale, la nuova economia del benessere, lo studio delle interdipendenze sollecitato dal modello di Leontief e, infine, proprio lo specifico lavoro di Dantzig (e Wood) inizialmente motivato da "the organization of defense, the conduct of the war, and other specifically war-related allocation problems".

Sempre nei suoi ricordi, Dantzig attribuirà il nome di "metodo del simplesso" ad uno scambio di idee avuto con T. Motzkin, che rileggeva il procedimento alla luce di una sua interpretazione geometrica. Al di là del nome, il metodo trae origine dall'osservazione che, in un problema di programmazione lineare, la regione ammissibile è descritta da una disuguaglianza matriciale del tipo $Ax \leqq b$ che individua, quale intersezione di un numero finito di semispazi, un poliedro. Qui il massimo (o il minimo) di una funzione lineare, se esiste, è raggiunto in corrispondenza di uno dei vertici. È facile, nel caso di una funzione lineare di due o tre variabili, trovare geometricamente la regione ammissibile ed elencarne tutti i vertici (e quindi trovare il valore massimo o il valore minimo) ma in generale, nei problemi "concreti", il numero delle variabili decisionali è elevato e quello dei vertici raggiunge livelli elevatissimi. È qui che si inserisce il metodo di Dantzig, che dà una procedura effettiva per la ricerca dei punti di massimo o di minimo, con l'individuazione di un percorso "ragionevole" che evita il transito per tutti i vertici del poliedro, tramite la scelta di una sorta di "percorso ottimale".

Questo, per quanto riguarda la nascita della Programmazione Lineare (termine dovuto non a Dantzig, ma al solito Koopmans; si veda [5]). La nascita della Programmazione Non Lineare avviene in anni ed in ambienti molto prossimi a quelli che abbiamo avuto modo di ricordare a proposito dell'opera di Dantzig.

La programmazione lineare, non appena si configura come disciplina autonoma, funge da punto di riferimento per ulteriori generalizzazioni e il passaggio al caso non lineare è pressoché immediato e naturale.

Harold Kuhn, nei suoi ricordi [33, 34] risale ad un viaggio di Dantzig a Princeton, nel maggio '48, successivo a quello che abbiamo già citato, sempre per incontrare von Neumann e discutere con lui delle possibili connessioni tra la programmazione lineare e la teoria dei giochi: "Tucker happened to give Dantzig a lift to the train station for his return trip to Washington. On the way, Dantzig gave a short exposition of what linear programming was, using the transportation problem as a simple illustrative example. [...] Dantzig's visit to Princeton resulted in the initiation of a research project which had as its original object the study of the relation between linear programming and matrix games". Di questa ricerca, del resto, rimane una traccia precisa attraverso il contributo di Gale, Kuhn e Tucker - nel '48, a Princeton, David Gale e Harold Kuhn erano giovani studenti, già laureati, di Albert Tucker - al già citato volume di *Atti* curato da Koopmans. È Tucker che, verso la fine del '49, invita Gale e Kuhn a generalizzare, inizialmente al caso quadratico, i risultati ottenuti in tema di dualità per la programmazione lineare. "Gale declined. Kuhn accepted". Nasce così la programmazione non lineare, che a lungo ha identificato il suo esordio con il celebre articolo di Kuhn e Tucker: *Non Linear Programming* del 1951.

In realtà, come gli stessi Kuhn e Tucker hanno progressivamente riconosciuto, giungendo a parlare di "a theorem which has been incorrectly attributed to Kuhn and Tucker", il primo organico contributo alla teoria risale ad una decina di anni prima con la tesi di dottorato discussa da William Karush nel '39, all'Università di Chicago, mai pubblicata[1]. Bastano questi riferimenti - Chicago, 1939 - per intuire come sia il *calcolo delle variazioni* il contesto matematico nel quale collocare la tesi di Karush. Stiamo parlando di una scuola che, ancora in quegli anni, annoverava matematici del calibro di G.A. Bliss, L.M. Graves, F.A. Valentine, M.R. Hestenes. Si trattava in particolare, nel lavoro di Karush, di sviluppare una versione finito-dimensionale di quanto Valentine andava elaborando per il calcolo delle variazioni, considerando il problema di Lagrange con vincoli scritti sottoforma di disuguaglianza. E anche nel titolo - *Minima of functions of several variables with inequalities as side conditions* - la tesi di Karush ricorda un "corrispondente" articolo di Valentine da cui trae ispirazione [58].

La sua struttura è essenziale: la presentazione del problema viene seguita da condizioni necessarie e da condizioni sufficienti che coinvolgono, prima, solo le derivate prime e poi anche quelle seconde. La tesi ha così inizio ricordando come il problema al centro dell'ottimizzazione vincolata classica abbia ormai raggiunto un trattamento soddisfacente per funzioni di classe $C^2$. Di una recente pubblicazione di Bliss [3] vengono in particolare ricordati tre risultati. Il primo è la cosiddetta condizione di Carathéodory [4], più generale del classico teorema da noi riportato nel punto 6) del paragrafo 2, in quanto non richiede alcuna condizione di regolarità dei vincoli: se $x^o$ è soluzione locale del problema Min $f(x)$, con $x \in S = \{x \in A,\ h_j(x) = 0, j = 1,\dots, r < n\}$

[1]Dobbiamo alla cortesia del Prof. M.El-Hodiri (dell'Università del Kansas) la trasmissione del testo tratto dall'unico microfilm disponibile.

esiste un vettore non nullo $\lambda^o = \left(\lambda_1^o, \lambda_1^o, \dots \lambda_r^o\right)$ per cui le derivate della funzione lagrangiana $L = \lambda_0^o f + \sum_{j=1}^{r} \lambda_j^o h_j$ si annullano in $x^o$.

L'ulteriore condizione necessaria (del secondo ordine) reintroduce una di quelle condizioni che, dopo il lavoro di Kuhn e Tucker, siamo abituati a chiamare "di qualificazione dei vincoli" (ottenendo così $\lambda_0^o = 1$): se $x^o$ è soluzione del problema in questione e la matrice jacobiana $\nabla h\left(x^o\right)$ ha rango pieno, allora il differenziale secondo in $x^o$ della funzione lagrangiana:

$$L(x) = f(x) + \sum_{j=1}^{r} \lambda_j^o h_j(x)$$

è una forma quadratica vincolata semidefinita positiva, per tutte le direzioni $dx$ soddisfacenti le condizioni $\nabla h_j\left(x^o\right)dx = 0$. Il terzo teorema di Bliss esprime una condizione sufficiente: se ad un punto $x^o$ è possibile associare un vettore di moltiplicatori $\lambda^o$ tale che la funzione lagrangiana:

$$L = f + \sum_{j=1}^{r} \lambda_j^{\circ} h_j$$

soddisfa le condizioni $L_{x_i}\left(x^o\right) = 0$ e $d^2 L\left(x^o\right) > 0$ per ogni $dx \neq 0$ tale che $\nabla h_j\left(x^o\right)dx = 0$, allora $x^o$ è un punto di minimo (locale stretto).

Il problema affrontato da Karush è invece quello che ormai conosciamo come tipico della programmazione non lineare. I vincoli sono scritti nella forma $g_j(x) \geq 0$ e vengono considerati tutti attivi in $x^o$, dato che per ragioni di continuità un eventuale vincolo del tipo $g_j\left(x^o\right) > 0$ non impone localmente alcuna restrizione.

La prima condizione necessaria, se letta frettolosamente, può dare l'impressione - erronea - di anticipare quello che nel secondo paragrafo abbiamo indicato come teorema di John: se $x^o$ è soluzione locale del problema Min $f(x)$ con $x \in S = \{x \in A,\ g_i(x) \geq 0\}$ allora esiste un vettore non nullo $\lambda^o = \left(\lambda_0^o, \lambda_1^o, \dots \lambda_m^o\right)$ tale che $x^o$ è stazionario per la funzione lagrangiana:

$$L(x) = \lambda_0^o f + \sum_{i=1}^{m} \lambda_i^o g_i(x).$$

In realtà, al teorema di John non ci siamo ancora. La proposizione 3.1 di Karush non dà alcuna informazione sul segno dei moltiplicatori e, in accordo con ambizioni così limitate, la sua dimostrazione è immediata. Per m < n è addirittura "unnecessary", in quanto il teorema è diretta conseguenza della precedente condizione necessaria di Carathéodory; a questo teorema d'altra parte ci si può ricondurre anche nel caso generale, introducendo delle variabili ausiliarie e trasformando i vincoli nelle uguaglianze $g_i(x) - z_i^2 = 0$.

Quello che sempre nel paragrafo 2 abbiamo invece ricordato come teorema di Kuhn e Tucker compare effettivamente (nel 1939!) come teorema 3.2. Karush introduce il *cono linearizzante* in $x^o$ che chiama "cono delle direzioni ammissibili") delle direzioni $dx \neq 0$ tali che $\nabla g_i(x°)dx \geq 0$ e dimostra la seguente condizione necessaria: “suppose that for each admissible direction there is an admissible arc issuing from $x^o$ in the direction *dx*. Then a first necessary condition for $f(x^o)$ to be a minimum is that there exist multipliers $\lambda_i \leq 0$ such the derivatives $L_{x_i}$ of the function $L = f + \sum \lambda_j g_i$, all vanish at $x^o$“. Se si nota nella citazione di Karush, qualche diversità (a proposito del segno dei moltiplicatori) con quanto affermato nel paragrafo 2, si ricordi che il problema qui trattato è di *minimo*. Anche in questo caso, grazie ora al lemma di Farkas la dimostrazione è sufficientemente semplice. Sia *dx* una direzione del cono linearizzante e *x(t)* una curva tale che $x(0) = x^o$ e $x'(0) = dx$.

Da $f[x(t)] \geq f(x^o) = f[x(0)]$, $\forall t \in [0, t_0]$, segue immediatamente $f'[x(0)] \geq 0$ ovvero $\nabla f(x°) \cdot x'(0) = \nabla f(x°)\ dx \geq 0$. Questa disuguaglianza è dunque una "conseguenza” dell'"ammissibilità" di *dx*. Abbiamo così la tesi (comprensiva dell'osservazione sul segno dei moltiplicatori) applicando il lemma di Farkas.

Karush naturalmente non si nasconde il ruolo esercitato, nel corso della dimostrazione, dalla condizione di qualificazione dei vincoli (che chiama *"property* Q") senza la quale non è in grado di provare neppure un teorema "à la John". Si chiede così “what the probability is, roughly, that the funtions $g_i(x)$ will satisfy property Q”, che, interpretata geometricamente, ipotizza che per ogni direzione del cono linearizzante - e quindi, in sostanza, per ogni direzione "tangente" - esista una curva regolare della regione ammissibile che la approssimi. La conclusione è sufficientemente rassicurante: “if the functions $g_i$ are regular enough, it seems that the satisfaction of property Q is not a great restriction”.

La proprietà Q può comunque essere sostituita da altre ipotesi, che ugualmente assicurano la condizione necessaria Si può ad esempio chiedere, in alternativa, che esista una particolare direzione ammissibile *dx* per la quale risulta $\nabla g_i(x^o)dx > 0, \forall i$. In questo caso, infatti, chiamata con *x(t)* una curva regolare definita su un intervallo $[0, t_0]$ e tale che $x(0) = x^o$ e $x'(0) = dx$, si ha $g'_i(x^o) = \nabla g_i(x^o)x'(0) > 0$; le funzioni $g_i[x(t)]$ risultano crescenti e da $g_i[x(t)] \geq g[x(t)] = 0$ segue che *x(t)* è un arco ammissibile. Di conseguenza, è $f[x(0)] \leq f[x(t)]$ e quindi $f'[x(0)] = \nabla f(x^o)dx \geq 0$. A questo punto si può considerare un'altra direzione ammissibile *dy* e una combinazione lineare convessa *dz=t dy + (1-t) dx*. È chiaro che, per ogni t $\neq$1, risulta $\nabla g_i(x^o)dz > 0$ e quindi (ripetendo quanto dimostrato per la direzione *dx*) $\nabla f(x^o)dz \geq 0$. La linearità di quest'ultima funzione permette di estendere la disuguaglianza al caso $t = 1: \nabla f(x^o)dy \geq 0$, che è dunque una "conseguenza" delle disuguaglianze $\nabla g_i(x^o)dy \geq 0$. A questo punto, come prima, interviene il lemma di Farkas.

Karush non confronta le due condizioni di qualificazione dei vincoli utilizzate, ma illustra significativamente un esempio bidimensionale che soddisfa la proprietà Q e non quella dell'ultimo teorema. Siano

$x^o = (x_0, y_0) = (0,0); g_1(x,y) = x^2 + (y-1)^2 - 1, g_2(x,y) = 4 - \left[x^2 + (y-2)^2\right],$
$g_3(x,y) = x + y^2.$

Risultando $\nabla g_1(0,0) = (0,-2)$, $\nabla g_2(0,0) = (0,4)$, $\nabla g_3(0,0) = (1,0)$, il cono linearizzante è costituito dall'unica direzione ($a$,0), con $a > 0$, per la quale è sempre possibile trovare un arco ammissibile $x$(t) con $x(0) = (0,0)$ e $x'(0) = (a,0)$. Non è invece soddisfatta la richiesta che sia (per j = 1, 2, 3) $\nabla g_j(0,0) \cdot (a,0) > 0$.

La parte centrale dell'inedito di Karush sarà quella che maggiormente si confronta con il successivo lavoro di Kuhn e Tucker. Per offrire comunque un'idea anche dei paragrafi successivi, riprendiamo l'osservazione sulla struttura "lineare" della tesi. I teoremi che abbiamo presentato sono seguiti da alcune condizioni sufficienti, sempre del primo ordine, che sviluppano con uno stretto parallelismo le precedenti condizioni necessarie, opportunamente rafforzate. Ad esempio; il teorema 4.1 garantisce che il punto $x^o$ è di minimo se $m \geq n$, la matrice jacobiana $\nabla g_i(x^o)$ ha rango $n$ ed esiste un vettore di moltiplicatori $\lambda^o = (\lambda_1^o, \ldots, \lambda_m^o)$ a componenti negative, tale che la funzione $L(x) = f(x) + \sum_{i=1}^{m} \lambda_i^o g_i(x)$ annulla in $x^o$ tutte le sue derivate prime.

Sempre per dare un'idea, ora, delle condizioni necessarie e delle condizioni sufficienti del secondo ordine citiamo, dalla tesi di Karush, il corollario del teorema 5.1 e il teorema 6.1. Il primo, per un punto $x^o$ per cui la matrice $\nabla g_i(x^o)$ ha rango $m$, e che la provenienza dal calcolo delle variazioni suggerisce di chiamare *normale,* afferma che condizione necessaria perché $x^o$ sia di minimo è che $L_{x_i}(x^o) = 0$ (con $L = f + \sum_{i=1}^{m} \lambda_i g_i$ e $\lambda_i \leq 0$) e che risulti $d^2 L(x^o) \geq 0$ in corrispondenza ad ogni direzione ammissibile $dx$ che soddisfi le condizioni $\nabla g_i(x^o) dx = 0$. Il teorema 6.1 assicura che $x^o$ è soluzione del problema in questione se è possibile associargli un vettore $\lambda^o = (\lambda_1^o, \ldots, \lambda_m^o)$ a componenti negative, per il quale risulta $L_{x_i}(x^o) = 0$ e $d^2 L(x^o) > 0$ per ogni direzione ammissibile $dx$ soddisfacente le condizioni $\nabla g_i(x^o) \mathrm{dx} = 0$ con $(\text{con } L = f + \sum \lambda_i g_i)$.

Questi riferimenti all'opera di Karush sono sufficienti a comprendere come, con il 1939, entriamo nella storia vera e propria della programmazione non lineare, nel momento in cui il calcolo delle variazioni le fornisce motivazioni e metodi per un loro approfondimento originale. È giusto ricordare questa origine e questa diversità (rispetto ai successivi e ugualmente fondanti lavori di John e di Kuhn e Tucker), per mettere in evidenza la ricchezza delle differenti anime che stanno alla base della programmazione non lineare, ma l'eccessiva enfasi - quale a volte si legge nelle ricostruzioni operate da Kuhn ("the motivation for Karush's work was different from the spirit of mathematical programming that prevailed at the end of the 1940's") - appare piuttosto funzionale a preservare una sufficiente originalità anche ai momenti successivi. Torneremo comunque più avanti sul confronto tra il contributo di Karush e quello di Kuhn e Tucker. Il loro è solo il terzo atto della nostra "storia", preceduto dalla pubblicazione del 1948 di un articolo del matematico (di origine tedesca) Fritz John, nel volume che festeggiava i 60 anni di Richard Courant [27].

Anche se Kuhn ha messo in giusta evidenza le motivazioni geometriche presenti in questo lavoro, in realtà l'approccio di John è molto generale, quasi compiaciuto dell'astrattezza del contesto nel quale il problema è formalizzato. Si tratta di minimizzare una funzione $f$, definita su un insieme $R$ di uno spazio $E$, considerando il sottoinsieme $R' \subseteq R$ descritto dalle disuguaglianze $g(x, y) \geq 0$, al variare del parametro $y$ in un insieme $S$ di uno spazio $H$. La generalità viene poi attenuata, assumendo $E = R^n$ e $S$ sottoinsieme compatto di uno spazio metrico $H$; si suppone inoltre che $f$ e $h$ siano funzioni di classe $C^1$, sia pure con una significativa apertura verso quelle che qualche decennio dopo saranno conosciute come tipiche problematiche dell'ottimizzazione non-smooth: "from the point of view of applications it would seem desiderable to extend the methods used here to cases, where the functions involved are not necessarily differentiable".

Quello che ancora oggi citiamo come *Teorema di John* viene presentato subito: se $x^o$ è un punto di minimo, nelle ipotesi dette, esistono un numero finito di punti $y_1, \ldots, y_s \in S$ (con $s \leq n$) e un vettore $\lambda^o = \left(\lambda_0^o,\ \lambda_1^o, \ldots,\ \lambda_s^o\right) \neq 0$ con $\lambda_0^o \geq 0$ e $\lambda_j^o \geq 0$ $(j = 1, \ldots, s)$ per cui la funzione $L(x) = \lambda_0^o f(x) - \sum_{i=1}^{s} \lambda_i^o g(x, y_i)$ ha un punto critico in $x^o$. Anche la dimostrazione è una di quelle che attualmente si trovano nei *textbooks* di ottimizzazione. John dimostra, anzitutto, che non vi può essere alcun vettore $dx$ soluzione del sistema:

$$\begin{cases} \nabla f\left(x^o\right) dx < 0 \\ \nabla g_i\left(x^o\right) dx > 0 \end{cases}$$

dove $g_i(x) = g(x, y_i)$ sono le funzioni che intervengono nei vincoli attivi in $x^o$. La verifica di questa affermazione è in John più laboriosa del consueto, coinvolgendo anche la compattezza dell'insieme S che qui non è necessariamente finito. A questo punto scatta, anziché il "solito" lemma di Farkas, un teorema dell'alternativa dimostrato da L.L. Dines [8] e da R.W. Stokes [55]. John lo ricostruisce geometricamente: l'impossibilità del precedente sistema implica che lo spazio immagine del punto $x^o$, tramite le funzioni $(-\nabla f, \nabla g_i)$ non appartenga al primo ottante e pertanto possa essere da questo separato tramite un iperpiano passante per l'origine. È tale iperpiano che fornisce i moltiplicatori che figurano nella tesi.

La parte più propriamente analitica dell'articolo è completata da una condizione sufficiente, qui riportata nella versione data da Stoer e Witzgall [54]: se esiste una funzione della forma

$\mathrm{L}(x)=\lambda_0^{\mathrm{o}} f(x)-\sum_{\mathrm{i}=1}^{\mathrm{s}} \lambda_{\mathrm{i}}^{\mathrm{o}}\ g(x, y_i)$, con $g(x^o, y_i)=0, \lambda_0^{\mathrm{o}} \geq 0$ e $\lambda_{\mathrm{i}}^{\mathrm{o}} \geq 0$, per cui $x^o$ è punto critico e inoltre la matrice:

$$\begin{bmatrix} \lambda_0^o f_{x_1}(x^o) & \lambda_1^o g_{x_1}(x^o, y_1) & \ldots\ldots & \lambda_s^o g_{x_1}(x^o, y_s) \\ \lambda_0^o f_{x_2}(x^o) & \lambda_1^o g_{x_2}(x^o, y_1) & \ldots\ldots & \lambda_s^o g_{x_2}(x^o, y_s) \\ \ldots\ldots\ldots\ldots & & & \\ \lambda_0^o f_{x_n}(x^o) & \lambda_1^o g_{x_n}(x^o, y_1) & \ldots\ldots & \lambda_s^o g_{x_n}(x^o, y_s) \end{bmatrix}$$

ha rango n, allora $x^o$ è un punto di minimo rispetto all'insieme individuato in $R$ dalle disuguaglianze $g(x, y_i) \geq 0 \ \ (\mathrm{i}=1, \ldots, \mathrm{s})$ e quindi a maggior ragione rispetto a $R'$.

Nel secondo paragrafo abbiamo visto come una prima presentazione dei teoremi di John e di Kuhn e Tucker possa agevolmente insistere sulla loro analogia, appena scalfita da un'unica variante: se alle ipotesi del teorema di John si aggiunge una condizione di qualificazione dei vincoli, si ha la certezza (con il teorema di Kuhn e Tucker) che il moltiplicatore associato alla funzione obiettivo, nella lagrangiana, sarà diverso da zero. Può allora appare sorprendente - alla luce di questa ricostruzione logica - che nessuno dei protagonisti dei primi sviluppi della programmazione non lineare abbia fatto un passo in questa direzione. Karush anticipa - nettamente - il teorema di Kuhn e Tucker, ma non svolge alcuna osservazione sulla situazione che si verrebbe a creare in assenza della "property Q"; viceversa, John non si pone neppure il problema della opportunità, di avere una funzione lagrangiana che veda sempre la funzione obiettivo tra le sue costituenti. La stessa osservazione vale per Kuhn e Tucker che, come Karush, non spendono nessuna parola per valutare le conseguenze della mancata validità della condizione di qualificazione dei vincoli.

Il loro articolo [36] richiama subito alla memoria l'incontro tra Dantzig e Tucker, a Princeton. Inizia, infatti, ricordando la formalizzazione di un problema di programmazione lineare e la sua equivalenza con un "problema di sella": $x^o$ è un punto di minimo per una funzione lineare $f$ considerando i vincoli $g_i(x)=b_i-\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{a}_{\mathrm{ij}} x_j \geq 0$ e $x_j \geq 0$, se e solo se esiste un vettore di moltiplicatori $\lambda^{\mathrm{o}}$ a componenti non negative tale che $(x^o, \lambda^{\mathrm{o}})$ è punto di sella per la funzione lagrangiana $L=f+\lambda \mathrm{g}$, ovvero soddisfa le disuguaglianze $\mathrm{L}(x, \lambda^{\mathrm{o}}) \leq \mathrm{L}(x^o, \lambda^{\mathrm{o}}) \leq \mathrm{L}(x^o, \lambda)$, per ogni $x$ e $\lambda$ non negativi, fornendo così una soluzione per il corrispondente gioco, a somma zero, fra due persone. Obiettivo dichiarato dell'articolo è l'estensione di questo risultato ad un problema di programmazione non lineare, nel quale vengono esplicitamente considerate le condizioni di non-negatività nelle variabili decisionali.

In questa direzione, il primo passo è costituito dalla dimostrazione di una condizione necessaria e di una condizione sufficiente per i punti di sella di una generica funzione $\varphi(x, \lambda)$

differenziabile, unicamente considerata per $x_j \geq 0$ e $\lambda_j \geq 0$. La condizione necessaria richiede che nel punto $\left(x^o, \lambda^o\right)$ siano soddisfatte le condizioni:

$$\varphi_{x_j} \leq 0; \qquad x_j \ \varphi_{x_j} = 0; \qquad x_j \ \geq 0; \tag{1}$$

$$\varphi_{\lambda_j} \geq 0; \qquad \lambda_j \ \varphi_{\lambda_i} = 0; \qquad \lambda_j \ \geq 0. \tag{2}$$

Le stesse condizioni sono sufficienti a garantire che $\left(x^o, \lambda^o\right)$ è un punto di sella per $\varphi$, se ad esse si aggiungono le disuguaglianze:

$$\varphi\left(x, \lambda^o\right) \leq \varphi\left(x^o, \lambda^o\right) + \nabla_x \varphi\left(x^o, \lambda^o\right) \cdot \left(x - x^o\right) \tag{3}$$

$$\varphi\left(x^o, \lambda\right) \geq \varphi\left(x^o, \lambda^o\right) + \nabla_\lambda \varphi\left(x^o, \lambda^o\right) \left(\lambda - \lambda^o\right) \tag{4}$$

che Kuhn e Tucker si sentono in obbligo di giustificare, perché non sembrino “as artificial as may appear at first sight”, ma che oggi - a distanza di mezzo secolo - è immediato leggere come ipotesi di concavità (convessità) di $\varphi$ rispetto a $x$ (rispetto a $\lambda$ ).

Come secondo momento, viene introdotta - ora proprio con il nome di *constraint qualification* - quella stessa ipotesi di regolarità dei vincoli, che abbiamo già incontrato in Karush. Serve per eliminare, sulla frontiera della regione ammissibile, situazioni patologiche create da punti “such as an outward pointing cusp”. Si consideri, ad esempio, in $R^2$, la regione ammissibile definita dai vincoli $g_1 = x \geq 0$, $g_2 = x \geq 0$, $g_3 = (1-x)^3 - y \geq 0$ e, in particolare, il punto (1,0). La direzione $(dx, dy) = (1,0)$ è "tangente" alla regione ammissibile in quanto soddisfa le condizioni $\nabla g_2(1,0) \cdot (1,0) \geq 0$ e $\nabla g_3(1,0) \cdot (1,0) \geq 0$, è cioè più precisamente una direzione del cono linearizzante. Eppure - questa è la patologia che la *constraint qualification* qui non verificata, si impegnerebbe a rimuovere - pur essendo "tangente", conduce da (1,0) in punti di $R^2$ che nulla hanno a che fare con la regione ammissibile.

Il terzo e ultimo passo, per raggiungere l'obiettivo indicato, è la dimostrazione di una condizione necessaria e di una condizione sufficiente, ora, per il problema $Max\, f(x)$, quando siano presenti i vincoli $g_i \geq 0$ e $x_s \geq 0$. La condizione necessaria assicura che esiste un moltiplicatore $\lambda^o$ a componenti non negative per cui la funzione $L = f + \lambda g$ soddisfa le precedenti condizioni (1) e (2); se a questo sistema aggiungiamo la disuguaglianza (3), allora $x^o$ è soluzione del problema di massimo considerato. Vale la pena di soffermarci brevemente sulla dimostrazione della condizione necessaria, perché è quella che ci permetterà di precisare il confronto con il lavoro di Karush e di trarre qualche conclusione in merito alla questione di priorità. È bene dire subito che i passaggi cruciali sono identici a quelli del teorema 3.2 di Karush: anzitutto, grazie alla condizione di qualificazione dei vincoli, ogni direzione $dx$ del cono linearizzante verifica la disuguaglianza $\nabla f\left(x^o\right) dx \leq 0$; a questo punto entra in scena il lemma che, nelle parole di Kuhn e di Tucker, è stato “indicated by H. Minkowski and proved by J. Farkas at the turn of the century”. I passaggi successivi trovano la loro motivazione unicamente in una tesi leggermente diversa (rispetto al teorema 3.2), in quanto è leggermente diverso il problema di ottimo qui preso in esame.

Confrontando le varie proposizioni finora enunciate, relative ai punti di sella e a quelli di massimo, è chiaro che si avrà l'equivalenza auspicata, anche nel caso non lineare, sistemando in un *loop* la definizione di sella, le condizioni (1)-(2)-(3) e, infine, la definizione di massimo. È altrettanto chiaro che sarà possibile percorrere questo *loop,* in entrambi i versi, aggiungendo la richiesta di concavità per la funzione lagrangiana rispetto alle variabili decisionali (quella di convessità rispetto a λ è ovviamente soddisfatta). È quanto effettivamente fanno Kuhn e Tucker quando dimostrano *("Theorem 3. (Equivalence Theorem)")* che $x^o$ è punto di massimo se e solo se per qualche $\lambda^o$ con $\lambda_i^o \geq 0$, $(x^o, \lambda^o)$ è di sella per $L = f + \lambda g$ dove $f$ e $g_i(i = 1, 2, \ldots, m)$ siano supposte concave.

A questo punto, è possibile trarre qualche conclusione sufficientemente fondata sull'originalità degli articoli di (Karush 1939) e di (Kuhn e Tucker 1951). Anche loro ottengono le "condizioni di Kuhn e Tucker" quando considerano la situazione particolare che si viene a presentare con vincoli unicamente scritti come $g_i(x) \geq 0$, essendo dunque assenti le condizioni di non negatività delle variabili; è facile infatti verificare, in questo caso, che la (l) si riduce alle uguaglianze $L_{x_j}$. Simile, dunque, l'enunciato e simili le dimostrazioni basate - come abbiamo osservato - su una stessa condizione di qualificazione dei vincoli e sul ricorso al lemma di Farkas. Quello che rimane diverso - e permette di parlare di contributi indipendenti, che pure raggiungono lo stesso risultato - è il contesto nel quale la condizione necessaria è collocata. Più "moderno", nel senso di maggiormente simile a molte delle attuali presentazioni, è il percorso di Karush che non rinuncia a sottolineare l'analogia che la programmazione non lineare conserva con l'elementare ricerca degli estremanti di una funzione di una variabile reale e sostanzialmente articola la sua esposizione nella formalizzazione del problema e nelle condizioni necessarie e sufficienti del primo e del secondo ordine. Diversamente, Kuhn e Tucker orientano la loro attenzione - e quella del lettore - sul problema di sella e sulla generalizzazione di quanto già era stato ottenuto in tema di programmazione lineare, ottenendo poi il "loro" teorema solo come caso particolare, quasi come un risultato secondario della loro analisi.

Se l'approccio di Karush (1939) è davvero "più moderno" rispetto a quello di Kuhn e Tucker (1951) e anche rispetto a quello di Fritz John (1948), c'è da chiedersi come mai la tesi di Karush rimase "oscura" per molti anni. Una prima risposta ovvia (addirittura banale) pone l'accento sul fatto che tale lavoro non fu mai pubblicato (e non lo è tutt'oggi, almeno a conoscenza dello scrivente). Però il lavoro di Fritz John fu pubblicato, eppure non ebbe risonanza che parecchi anni dopo la sua pubblicazione, quando oramai la programmazione matematica era diventata un'area autonoma di studi e ricerche. L'articolo di Kuhn e Tucker diede invece quasi subito la stura a numerosi articoli e note. Il motivo essenziale di tale disparità nel processo di diffusione dei risultati scientifici è dovuto, a mio parere, al diverso ambiente ove i medesimi furono concepiti e prodotti. Karush e Fritz John (quest'ultimo - non dimentichiamolo - matematico affermato, forse l'allievo più brillante di Richard Courant) operavano con intendimenti esclusivamente teorici in ambienti ove solo la matematica "pura" o "purissima" aveva una qualche dignità. Kuhn e Tucker operavano in un ambiente ove le motivazioni applicative (soprattutto di tipo economico e manageriale) erano la molla che spingeva a nuove ricerche e a trovare nuovi risultati.

Il lavoro di Karush fu, di fatto, posto all'attenzione degli studiosi da due economisti: M. El-Hodiri [12, 13] e A. Takayama [56]. Lo stesso Kuhn [34] dichiara candidamente ma molto onestamente, che apprese dell'esistenza della tesi di Karush dal libro di Takayama (1974!). Precedentemente, oltre a El-Hodiri che aveva a fondo studiato Karush, tale tesi appare citata solo in [15, 42].

Un altro contributo importante di quegli anni alla programmazione non lineare è il lavoro (anch'esso non pubblicato!) di M. Slater [53], del 1950 (in tale lavoro si cita il contributo di Kuhn e Tucker che quindi era già circolato negli ambienti scientifici l'anno precedente la sua pubblicazione).

Slater riprende la questione trattata da Kuhn e Tucker, sui legami tra un problema di programmazione non lineare e un opportuno problema di sella della relativa funzione lagrangiana. La "novità" sta nel fatto che Slater non impone nessuna ipotesi di differenziabilità sulle funzioni implicate, che sono richieste essere concave o convesse a seconda del tipo di problema; Slater impone poi una condizione di qualificazione dei vincoli, oggi universalmente nota come "condizione di Slater".

Le dimostrazioni sono per la verità inutilmente complesse e utilizzano il teorema del punto fisso di Kakutani.

Il risultato di Slater fu poi ridimostrato in modo più sintetico, con l'utilizzo di classici teoremi di separazione tra insiemi convessi, indipendentemente da Uzawa [57] e da Karlin [29]: di nuovo ci troviamo sotto l'egida delle interazioni con la teoria economica.

Abbiamo già menzionato che due grandi studiosi e pionieri della programmazione matematica, T.C. Koopmans e L.V. Kantorovich, ricevettero nel 1975 il premio Nobel per l'Economia. Ad essi possiamo senz'altro aggiungere il nome di K.J. Arrow, premio Nobel per l'Economia nel 1972, i cui studi su fondamentali aspetti teorici della programmazione matematica (soprattutto in collaborazione con L. Hurwicz e H. Uzawa) meritano di essere segnalati anche ai nostri giorni. Infine H.M. Markowitz fu insignito dello stesso premio Nobel nel 1990 per i suoi contributi alla programmazione quadratica nell'ambito del celebre modello di scelta di portafoglio finanziario.

## 5 – Alcune considerazioni finali

La storia della nascita della programmazione non lineare finisce qui. Con l'articolo di Kuhn e Tucker, la teoria acquista una sua definitiva autonomia, raggiungendo in breve tempo una tale tentacolarità e una tale quantità di contributi da non rendere più possibile una descrizione del suo stato in un'unica sede, con sufficiente analiticità. A quanto finora abbiamo documentato è forse opportuno aggiungere ancora qualche breve considerazione conclusiva.

È sicuramente appena il caso di ribadire come la nascita della programmazione (lineare e non) sia una "grande storia", vuoi per l'interesse e la rilevanza dei contenuti - trafficato crocevia di diversi settori di ricerca e stimolante osservatorio anche per la didattica - vuoi per i prestigiosi nomi che si incontrano - "grandi" matematici, premi Nobel - tra le prime espressioni di due scuole che eserciteranno una vera e propria *leadership* nella seconda metà del secolo.

La scelta, pressoché obbligata, di raccontare della programmazione matematica solo gli esordi, ha portato inevitabilmente a sottolinearne alcuni aspetti tipici delle "nascite": il debito verso alcuni strumenti - qui l'algebra lineare e l'analisi convessa - che si sono rivelati essenziali per il decollo e le convergenze e la temporanea sovrapposizione di diversi linguaggi - il calcolo delle variazioni, quello geometrico, quello più analitico - dai cui apporti e confronti nasce un *mix* che si pone come nuova e originale teoria. Sempre la "scelta degli esordi" porta altrettanto inevitabilmente a considerarli come punto terminale di diversi fili e a mettere maggiormente a fuoco il precedente periodo preparatorio. È il caso della programmazione matematica, che "nasce" a metà del secolo, ricevendo in eredità alcune caratteristiche che forse meglio aderiscono a fasi precedenti. Basti pensare allo stretto intreccio che abbiamo descritto tra matematici ed economisti, tra esigenze "pratiche" e approfondimenti teorici. Da questo punto di vista, la programmazione matematica può essere ancora considerata quale espressione di quella grande tradizione analitica che "offre" il concetto di funzione e il calcolo differenziale (e tutti i loro raffinamenti) come la chiave più adatta per supportare l'analisi nelle stesse scienze economiche e sociali.

Come ogni storia del Novecento, infine, anche quella della programmazione matematica si presta ad alcune riflessioni relative alla storiografia della matematica moderna. Occuparsi di una matematica quasi contemporanea è, per certi versi, più impegnativo - si pensi ai contenuti disciplinari - ma, per altri (ad esempio, la reperibilità delle fonti), più semplice. Il punto in questione è però un altro e viene avanti tramite le obiezioni di chi rimprovera alla storiografia moderna di non poter disporre di un distacco sufficiente e quindi di un orizzonte temporale abbastanza aperto; troppo schiacciate "sull'oggi", le ricostruzioni della matematica moderna non riuscirebbero a produrre una valutazione misurata né a superare, mancando consolidati riferimenti storici e filosofici, la soglia della divulgazione. La storia della programmazione matematica indica tutta la ragionevolezza di queste perplessità, ma anche la necessità di considerarle uno stimolo e un incentivo, e non un freno paralizzante. La distanza che ci separa dagli anni '50 del secolo scorso - forse troppo breve per vedere tutti i frutti di determinate ricerche o la loro caducità - non può costituire un alibi per rinviare ricostruzioni, sistemazioni, giudizi che appaiono utili (almeno) ad una migliore comprensione dell'ottimizzazione, per chi la usa o si aggira nei suoi dintorni dal punto di vista della ricerca e/o dell'insegnamento. E la stessa osservazione vale per il rischio di fermarsi al livello di una divulgazione, che tutti concordano nel ritenere propedeuticamente necessaria ma che non può ovviamente esaurire l'indagine storica. Sarebbe indubbiamente più rassicurante poter disporre di cornici concettuali già pronte, alle quali fare riferimento e nelle quali collocare la ricerca storico-matematica che vi attingerebbe anche per la propria "nobiltà". Sarebbe più facile, ma così non è. La storia delle idee del secolo appena concluso è un'opera *in fieri.* Eccitante, anche perché in corso e non ancora imbalsamata E se c'è un corrispettivo per gli storici delle singole discipline, per la loro maggiore fatica, questo è proprio nel poter contribuire alla costruzione senza limitarsi al ruolo di fruitori.

## Bibliografia

[1] Avriel M., Diewert W. E., Schaible S., Zang I., *Generalized Concavity*, Plenum Press, New York, 1988.

[2] Beavis B., Dobbs I., *Optimization And Stability Theory For Economic Analysis*, Cambridge Univ. Press, Cambridge, 1990.

[3] Bliss G. A., "Normality And Abnormality In The Calculus Of Variations", *Trans. Amer. Math. Soc.*, 43, 1938, 36s-376.

[4] Caratheodory C., *Calculus Of Variations And Partial Differential Equations Of The First Order, Part Ii, Calculus Of Variations*, Holden Day, San Francisco, 1967 (Tradotto Dall'originale In Tedesco Pubblicato Nel 1935).

[5] Dantzig G. B., "Linear Programming"; In [37], 19-31.

[6] Dantzig G. B., Thapa M. N., *Linear Programming, 1: Introduction*, Springer Verlag, Berlin, 1997.

[7] De Giuli M. E., Giorgi G., Magnani U., "A General Theorem Of The Alternative: How To Get Its Special Cases Quickly", *Pure Mathematics And Applications*, 8, 1997, 215-232.

[8] Dines L. L., "Linear Inequalities", *Bulletin Of The American Math. Society*, 42, 1936, 353-365.

[9] Dixit A. K., *Optimization In Economic Analysis*, Oxford Univ. Press, Oxford, 1990.

[10] Dorfman R., Samuelson P. A., Solow R., *Linear Programming And Economic Analysis*, Mc Graw-Hill, New York, 1958.

[11] Dorn W. S,. "Nonlinear Programming. A Survey", *Management Science*, 9, 1963, 171-208.

[12] El-Hodiri M.A., "The Karush Characterization Of Constrained Extrema Of Functions Of A Finite Number Of Variables", *Research Memorandum A.3*, Uar Ministry Of Treasury, 1967, Cairo.

[13] El-Hodiri M.A., *Constrained Extrema: Introduction To The Differentiable Case With Economic Applications*, Springer Verlag, Berlin, 1971 (Originalmente: M. A. El-Hodiri, Constrained Extrema Of Functions Of A Finite Number Of Variables: Review And Generalizations, Krannert Institute Paper, 141, 1966, Purdue Univ., Purdue).

[14] Farkas "Theorie Der Einfachen Ungleichungen", *Journal Fuer Die Reine Und Angewandte Mathematik*, 124, 1901, 1-27.

[15] Fiacco A.V., Mc Cormick G.P., *Nonlinear Programming: Sequential Uncostrained Minimization Techniques,* J. Wiley, New York, 1968.

[16] Gerber F., "Mathematical Programming Book Reviews: 1965-1972", *Management Science*, 20, 1973/74, 875-895.

[17] Giorgi G., *Un Approccio Unificante Ai Teoremi Dell'alternativa Per Sistemi Lineari*, Atti X Convegno A.M.A.S.E.S., Pitagora Editrice, Bologna, 1986, 103-136.

[18] Giorgi G., *Sui Teoremi Di Gordan, Stiemke E Tucker*, Atti Xxiv Convegno A.M.A.S.E.S., Università Degli Studi Di Bergamo E Brescia, 2000, 623-630.

[19] Giorgi G., Guerraggio A., "Ha Solo Cinquanta Anni: La Programmazione Non Lineare", *Pristem/Storia, 1*, Springer Verlag Italia, Milano, 1998, 1-32.

[20] Giorgi G., Guerraggio A., Thierfelder J., *Mathematics Of Optimization: Smooth And Nonsmooth Case,* American Elsevier, Amsterdam, 2004.

[21] Grattan-Guinness I., "Joseph's Fourier Anticipation Of Linear Programming", *Operational Research Quarterly*, 27, 1976, 361-364.

[22] Grattan-Guinness I., "On The Prehistory Of Linear And Nonlinear Programming", In E. Knobloch, D. E. Rowe (Eds.), *The History Of Modern Mathematics,* Academic Press, London, 1989, 43-89.

[23] Hancock H., *Theory Of Maxima And Minima*, Dover, New York, 1960 (Pubblicato Originalmente Nel 1917).

[24] Hitchcock F. L., "Distribution Of A Product From Several Sources To Numerous Localities", *Journal Of Mathematical Physics*, 20, 1941, 224-230.

[25] Intriligator M.D., *Mathematical Optimization And Economic Theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

[26] Intriligator M.D., "Mathematical Programming With Applications In Economics", Chapter 2 In K. J. ARROW, M. D. INTRILIGATOR (Eds.), *Handbook Of Mathematical Economics, Vol. I*, North Holland, Amsterdam, 1981.

[27] John F., "Extremum Problems With Inequalities As Subsidiary Conditions", In *Studies And Essays Presented To R. Courant On His 60th Birthday*, Interscience, New York, 1948, 187-204.

[28] Kantorovich L.V., "Mathematical Methods Of Organizing And Planning Production" (Traduzione Inglese Di "Mathematischeskie Metody Organizatriya Planirovanija Proizvodska, Univ. Di Leningrado, 1939), *Management Science*, 6, 1960, 366-422.

[29] Karlin S., *Mathematical Methods And Theory In Games, Programming And Economics*, 2 Volumi, Addison-Wesley, Reading, Mass., 1959.

[30] Karush W., Minim*a Of Functions Of Several Variables With Inequalities As Side Conditions,* M. A. Thesis, University Of Chicago, 1939.

[31] Koopmans T.C., "Optimum Utilization Of The Transportation System", Proceedings Of The International Statistical Conference, Washington, D. C., 1947. Ristampato Su *Econometrica*, 17, Supplement, 1949.

[32] Koopmans T.C. (Ed.), *Activity Analysis Of Production And Allocation*, J. Wiley, New York, 1951.

[33] Kuhn H.W., "Nonlinear Programming. A Historical View", *Siam-Ams Proceedings*, 9, 1976, 1-26.

[34] Kuhn H.W., "Nonlinear Programming: A Historical Note"; In [37] 82-96.

[35] Kuhn H.W., "The Discovery And Application Of Linear And Nonlinear Programming", In M. EMMER (Ed.), *Mathematics And Culture I,* Springer Verlag, Berlin, 2004, 13-18.

[36] Kuhn H.W., Tucker A.W., "Nonlinear Programming", In J. NEYMAN (Ed.), *Proceedings Of The Second Berkeley Symposium On Mathematical Statistics And Probability,* University Of California Press, Berkeley, 1951, 481-492.

[37] Lenstra J.K., Rinnooy Kan A.H.G., Schrijver A. (Eds.), *History Of Mathematical Programming*, Cwi-North Holland, Amsterdam, 1991.

[38] Motzkin T.S., *Beitraege Zur Theorie Der Linearen Ungleichungen*, Dissertation, Universitaet Basel, 1933 E Azriel, Jerusalem,1936.

[39] Y. Murata, *Mathematics For Stability And Optimization Of Economic Systems*, Academic Press, New York, 1977.

[40] Nemhauser G.L., Rinnooy Kan A.H.G. (Eds.), *Handbooks In Operations Research And Management Science - Volume 1 - Optimization*, North Holland, Amsterdam, 1989.

[41] Ostrogradsky M., "Considérations Générales Sur Les Moments Des Forces", *Mémoires De L'académie Impériale Des Sciences De Saint-Pétersbourg*, Sixième Série, 1, 1838, 129-150.

[42] Pennisi L., "An Indirect Sufficiency Proof For The Problem Of Lagrange With Differential Inequalities As Added Side Conditions", *Trans. Amer. Math. Soc*., 74, 1953, 177-198.

[43] Polyak B.T., "History Of Mathematical Programming In The Ussr: Analyzing The Phenomenon", *Math. Programming*, 91, Ser. B, 2002, 401-416.

[44] Pourciau B.H., "Modern Multipliers Rules", *American Math. Monthly*, 87, 1980; 433-452.

[45] Prekopa A., "On The Development Of Optimization Theory", *American Math. Monthly,* 87, 1980, 527-542.

[46] Ranzini L., "Il Contributo Di L. V. Kantorovich Ai Problemi Di Programmazione Lineare", *Fascicoli Della Sezione Di Matematica Generale Ed Applicata - Dipartimento Di Ricerche Aziendali*, Università Degli Studi Di Pavia, N. 63, A.A. L999/2000.

[47] Robbins L., *An Essay On The Nature And Significance Of Economic Science*, Macmillan, London, 1932.

[48] Samuelson P.A., *Foundations Of Economic Analysis*, Harvard Univ. Press, Cambridge, Mass.,1947.

[49] Samuelson P.A., "Maximum Principles In Analytical Economics", *American Economic Review*, 62, 1972, 249-262.

[50] Sawaragi Y., Nakayama H., Tanino T., *Theory Of Multiobjective Optimization*, Academic Press, Orlando, 1985.

[51] Schumpeter J.A., *History Of Economic Analysis*, Allen & Unwin, London, 1954.

[52] Simon C., "Scalar And Vector Maximization: Calculus Techniques With Economic Applications", In S. REITER (Ed.), *Studies In Mathematical Economics*, The Mathematical Association Of America, 1986, 62-159.

[53] Slater M., "Lagrange Multipliers Revisited - A Contribution To Non-Linear Programming", *Cowles Commission Discussion Paper: Mathematics*, 403, November 7, 1950.

[54] Stoer J., Witzgall C., *Convexity And Optimization In Finite Dimensions, I*, Springer Verlag, Berlin, 1970.

[55] Stokes R.W., "A Geometric Theorem Of Linear Inequalities", *Trans. Amer. Math. Soc*., 33, 1931, 782-805.

[56] Takayama A., *Mathematical Economics*, Dryden Press, Hinsdale, 1974. Ii Edition: Cambridge Univ. Press, Cambridge, 1985.

[57] Uzawa H., "The Kuhn-Tucker Theorem In Concave Programming", In K.J. ARROW, L. HURWICZ, H. UZAWA (Eds.), *Studies In Linear And Nonlinear Programming*, Stanford Univ. Press, Stanford, 1958, 32-37.

[58] Valentine F.A., "The Problem Of Lagrange With Differential Inequalities As Added Side Conditions", In *Contributions To The Calculus Of Variations* (1933-1937), Univ. Of Chicago Press, 1937, Chicago.

[59] Van De Panne C., Rahnama F., "The First Algorithm For Linear Programming: An Analysis Of Kantorovich's Method", *Economics Of Planning*, 19, 1985, 76-91.

[60] Von Neumann J., Morgenstern O., *Theory Of Games And Economic Behaviour,* Princeton Univ. Press, Princeton, 1944.