Ha solo cinquanta anni: la programmazione non lineare

Giorgio Giorgi, Angelo Guerraggio

Introduzione

Sono numerosissimi i problemi, sia teorici che pratici, che si traducono nella massimizzazione o minimizzazione di una determinata espressione. Sono questi i cosiddetti *problemi di ottimizzazione*, al cui interno è subito possibile distinguere tra problemi di ottimizzazione *statica* e di ottimizzazione *dinamica*, caratterizzati – questi ultimi – da un'evoluzione temporale delle quantità considerate che viene a far parte integrante della struttura del problema. Noi qui ci occuperemo esclusivamente dei primi. È all'ottimizzazione statica che si riferisce il termine *programmazione matematica* (introdotto da Robert Dorfman nel 1949) per individuare più precisamente quei problemi in cui occorre trovare i valori ottimali di una funzione obiettivo, mentre le sue variabili indipendenti (o decisionali) sono soggette a soddisfare determinati vincoli espressi (classicamente) da uguaglianze o scritti, in termini più moderni e più generali, sotto forma di disuguaglianze o di appartenenza ad un insieme assegnato. Si parla, più in particolare, di programmazione (matematica) *non lineare* quando le funzioni obiettivo e di vincolo sono qualsiasi e non è detto che appartengano tutte alla classe delle funzioni lineari o affini. Il primo articolo che utilizza esplicitamente il termine di *non linear programming* è di H.W. Kuhn e A.W. Tucker nel 1951.

Situata, come disciplina, tra l'analisi e la ricerca operativa, la programmazione matematica ha avuto per sua natura – quando non ha favorito o addirittura generato – forti connessioni ed interscambi con altre teorie o campi di ricerca quali l'analisi convessa, l'analisi non-smooth, l'algebra lineare, il calcolo numerico, la teoria dei giochi, la teoria delle decisioni, ecc.; le sue caratteristiche rendono persino ovvio sottolinearne l'importanza per l'analisi economica. Analoga osservazione può essere brevemente sviluppata per coglierne l'interesse sul versante didattico: l'ottimizzazione nasce, da un punto di vista logico, con la ricerca dei punti di massimo o di minimo per una funzione reale di una variabile reale, dando subito luogo ad un argomento tipico dei corsi di analisi del primo biennio universitario, quali la ricerca degli estremanti liberi o condizionati per una funzione reale di n variabili reali.

Può allora apparire sorprendente che una teoria così centrale e dotata di un interscambio così ricco dia luogo ad una disciplina relativamente giovane, che festeggia in questi anni solo il primo mezzo secolo di vita. La ricostruzione storica della programmazione non lineare, che proponiamo in questo articolo, 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 della guerra e in quelli immediatamente successivi. Cominceremo ricordando le definizioni, i concetti e i risultati fondamentali riferiti per semplicità a spazi di dimensione finita. Alla "preistoria" della programmazione non lineare è dedicato il terzo paragrafo, mentre il quarto si avvia decisamente verso gli anni della "nascita" con i primi sviluppi della programmazione lineare. Il paragrafo centrale sarà allora il quinto, in cui analizzeremo i fondamentali lavori di W. Karush, di F. John e di H.W. Kuhn e A.W. Tucker.

Alcuni risultati notevoli

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:

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

La regione A viene chiamata *insieme ammissibile* del problema (o anche *campo di scelta*); la funzione f viene detta *funzione obiettivo* e le variabili $x_1, x_2, ..., x_n$ vengono talvolte 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 $f[tx^1 + (1-t)x^2] \geq tf(x^1) + (1-t)\, f(x^2)$, 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 $x^\circ \in A$ soluzione di (P_1) e sia f differenziabile sull'aperto $A \subseteq R^n$. Allora necessariamente è $\nabla f(x^\circ) = 0$ ossia x° è *punto stazionario* (o *punto critico*) per f.

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

a) se è $y^T\, Hf(x^\circ)\, y < 0$, $\forall\, y \in R^n \setminus \{0\}$, ossia se la forma quadratica rappresentata dal differenziale secondo di f è definita negativa, allora x° è punto di massimo locale stretto per f su A. Se è $y^T\, Hf(x^\circ)\, y > 0$, $\forall\, y \in R^n \setminus \{0\}$, allora x° è punto di minimo locale stretto per f su A;
b) se la forma quadratica $y^T\, Hf(x^\circ)\, y$ è indefinita, allora x° non è né punto di massimo né di minimo per f;
c) *se $y^T\, Hf(x^\circ)\, y$ è una forma quadratica semidefinita (negativa o positiva)*, occorrono invece ulteriori indagini per decidere la natura del punto stazionario x°.

5) Se in (P_1) f è differenziabile e concava sull'aperto e convesso $A \subseteq R^n$, allora ogni suo punto stazionario è 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*[1] 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, $\forall\ x,y \in A$, 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:

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

Se f è differenziabile sull'aperto e convesso A, la quasiconcavità di f viene equivalentemente caratterizzata dall'implicazione[2]:

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

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:

$$(\mathrm{P}_2) \qquad \underset{x \in S}{Max}\, f(x) \qquad \text{oppure} \qquad \underset{x \in S}{Min}\, f(x)$$

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

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 (P_2) 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 $\lambda_j \in R, j = 1, \ldots, r$, sono detti *moltiplicatori di Lagrange*, sussistono i seguenti fondamentali risultati.

---

[1] Il riferimento alla concavità generalizzata prevede la citazione ormai "classica" dell'articolo di Avriel M., Diewert W. E., Schaible S., Ziemba W. T. (1981) Introduction to concave and generalized concave functions, in *Generalized Concavity in Optimization and Economics* (S. Schaible, W.T. Ziemba eds.), New York, Academic Press.

[2] Un'ulteriore significatica estensione del concetto di convessità si ha con la classe delle funzioni *invesse*. A questo proposito si può vedere G. Giorgi, E. Molho (1992) Generalized Invexity: relationships with generalized convexity and applications to optimality and duality conditions, in *Generalized Concavity for Economic Application*, (P. Mazzoleni ed.), Bologna, Tecnoprint, pp. 53-70.

6) Sia $x^\circ \in S$ soluzione locale di (P_2) e siano soddisfatte le seguenti ipotesi:

i) f è differenziabile in x°;

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

Allora esiste un (unico) vettore $\lambda^\circ = (\lambda^\circ_1, ..., \lambda^\circ_r)$ per cui (x°, λ°) è punto stazionario per la funzione lagrangiana, ossia:

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

Generalmente, sotto determinate ipotesi, non particolarmente restrittive, λ° viene a rivestire un preciso significato in quanto le sue componenti λ°_j 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 λ°_j è possibile una valutazione economica del peso che ciascun vincolo assume nella definizione dell'ottimo del problema (P_2).

7) Sia $x^\circ \in S$; la coppia (x°, λ°) soddisfi la relazione (1) e sia $L(x,\lambda)$ funzione pseudoconcava rispetto al vettore x. Allora x° 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:

$$(P_3) \qquad \begin{array}{c} Max\, f(x) \\ x \in K \end{array}$$

$K = \{x \text{ e } A,\ g_i(x) \leq 0,\ i = 1, ..., m\}$, ove $A \subseteq R^n$ è aperto; $f, g_i: R^n \to R$. Per tale problema l'insieme degli *indici dei vincoli attivi* in $x^\circ \in K$ è dato da:

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

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

8) *Teorema di F. John.* Siano f e g_i $(i = 1, ..., m)$ differenziabili in $x^\circ \in K$, soluzione locale di (P_3). Esiste allora un vettore $(y^\circ_0, y^\circ_1, ..., y^\circ_m) \in R^{m+1}$ a componenti non negative e non tutte nulle tale che:

i) $y^\circ_0 \nabla f(x^\circ) - \sum_{i=1}^{m} y^\circ_i \nabla g_i(x^\circ) = 0$;

ii) $y^\circ_i g_i(x^\circ) = 0,\ i = 1, ..., m.$

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 $y^\circ_{(i=1,...,m)}$.

Confrontando i risultati 3), 6) e il teorema di John (relativi, rispettivamente, ai problemi di ottimizzazione libera, P_2) e P_3)), 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:

$$y^\circ \nabla f(x^\circ) - \sum_{i \in I(x^\circ)} y^\circ_i \nabla g_i(x^\circ) = 0$$

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 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 (P_2), richiede che i vettori $\nabla g_i(x^\circ)$, $i \in I(x^\circ)$, siano linearmente indipendenti.

9) *Teorema di Kuhn - Tucker.* Sia $x^\circ \in K$ soluzione locale di (P_3) nelle ipotesi che f e g_i $(i = 1, ..., m)$ siano differenziabili in x°. Se è soddisfatta una condizione di qualificazione dei vincoli, allora esiste un vettore λ° tale che:

i) $\nabla f(x^\circ) - \sum_{i=1}^{m} \lambda^\circ_i \nabla g_i(x^\circ) = 0$;

ii) $\lambda^\circ_i g_i(x^\circ) = 0,\ i = 1, ..., m$;

iii) $\lambda^\circ_i \geq 0,\ i = 1, ..., m.$

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

10) Sia $x^\circ \in K$ 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, $i \in I(x^\circ)$, funzioni quasiconvesse, ovvero $- g_i$ sia quasiconcava. Allora x° è soluzione di (P_3).

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 nei paragrafi successivi 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:

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

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

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

Al problema:

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

corrisponde invece il duale:

$$\begin{cases} \text{Max yb} \\ \text{yA} \geqq \text{c}. \end{cases}$$

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* $f: R^n \to R^m$, 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 $f(x) \leq f(x^\circ)$ che caratterizzava i punti di massimo nel caso scalare.

La preistoria

È chiaro, dal titolo e da qualche accenno che abbiamo già avuto modo di fare nel corso del paragrafo introduttivo, che stiamo parlando di una teoria il cui battesimo ufficiale risale a questo secolo, agli anni immediatamente successivi alla seconda guerra mondiale. Ciononostante, chiunque abbia un ricordo, sia pure vago, di un secondo (o primo) corso di Analisi troverà familiare l'espressione e il riferimento ai *moltiplicatori di Lagrange*. E il nome è sufficiente per comprendere come la storia – o meglio la preistoria – della programmazione non lineare risalga alla fine del '700, anche se ora questo termine viene sostanzialmente messo in relazione a (P_3) o a problemi ancora più generali, che contengono come caso particolare quello classico scritto con vincoli sotto forma di uguaglianza[3].

Lagrange introduce i "suoi" moltiplicatori nel 1778, nella quarta sezione della prima parte della *Mécanique Analytique*, come strumento per determinare la configurazione di equilibrio stabile di un sistema meccanico. Si

---

[3] Fra gli oramai numerosi *textbooks* di programmazione non lineare ci limitiamo a segnalare, per un primo punto di riferimento, M.S. Bazaraa, C.M. Shetty (1976) *Foundations of Optimization*, Berlin-Heidelberg, Springer-Verlag.

tratta di dedurre, dai principi generali della Statica, «des formules analytiques qui renferment la solution de tous les problèmes sur l'équilibre des corps [...], en reduisant en quelque manière tous les cas à celui d'un système entièrement libre».

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 λ_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 λ_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 l'«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 métaphysique 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».

Negli stessi anni di Lagrange, sempre con vincoli scritti nella forma di uguaglianza, abbiamo anche i primi problemi lineari. Una ricostruzione accurata, quale quella operata da I. Grattan-Guinness[4], registra gli interventi di Laplace (1793 e 1799), che riprende la teoria seguita nel 1755 da R.J. Boscovich in un lavoro di carattere geodetico, e di G.C.F.M. de Prony (1804) che adatta il metodo di Laplace ad un problema di idrodinamica.

Nel periodo a cavallo tra Settecento e Ottocento, l'altro contributo importante è comunque quello di Fourier, con una significativa apertura nella direzione della programmazione "moderna", se è vero che questa troverà una sua caratterizzazione, rispetto ai problemi di ottimo vincolato classico, nell'uso delle disuguaglianze (anziché delle uguaglianze) per definire i vincoli, venendo così portata su un terreno matematico completamente nuovo. Sempre nel contesto lagrangiano dei principi variazionali della meccanica, Fourier pubblica anzitutto nel 1798 la *Mémoire sur la Statique* dove intende provare quello che oggi è chiamato "principio dei lavori virtuali" e che veniva utilizzato da Lagrange come assioma. Dalla dimostrazione ottiene però una disuguaglianza debole (*principio di Fourier*) che ugualmente interpreta come condizione di minimo della funzione potenziale (sempre che questa esista). La ricerca delle condizioni di equilibrio condurrebbe così ancora alla minimizzazione della funzione potenziale, con la significativa novità che la discussione del problema fisico prevede ora che le variabili del sistema possano essere condizionate da vincoli rappresentati da disuguaglianze: «il arrive souvent que les points du système s'appuient seulement sur les obstacles fixes, sans y être attachés».

L'*exploit* del 1798 non rimane un caso isolato. Dal materiale e dagli appunti inediti[5] emerge un Fourier sufficientemente consapevole degli spazi che potrebbe occupare quella che viene indicata come *analisi delle disu-*

[4] Cfr. Grattan-Guinness 1989.

[5] Cfr. Grattan-Guinness 1976. È a Grattan-Guinness, in particolare, che si deve la particolare considerazione di Fourier nei primissimi sviluppi della programmazione lineare.

guaglianze. In particolare, due articoli pubblicati nel 1826 e nel 1827 presentano quel metodo di soluzione di un sistema di disuguaglianze lineari che, basato sulla successiva eliminazione delle variabili, viene ora indicato come *metodo di eliminazione* di Fourier-Motzkin. Gli stessi articoli, con la discussione di alcuni problemi di programmazione lineare in due e in tre variabili (affrontati geometricamente, dando comunque una prima rudimentale versione del metodo del simplesso) giustificano ampiamente lo spazio dedicato a Fourier in una qualunque storia della programmazione lineare, ovviamente prima che questa esista e nasca come disciplina.

Nei lavori a stampa e nei manoscritti, dove Fourier anticipa in qualche misura i contenuti della programamzione moderna, non ci sono approfondimenti teorici né sistemazioni organiche, ma alcune prime acquisizioni – quali la convessità della regione ammissibile nei problemi lineari – e soprattutto l'indicazione e lo studio di molte situazioni "concrete" che documentano l'utilità dell'"analisi delle disuguaglianze" e ne consigliano l'approfondimento in diversi contesti, dall'aritmetica sociale all'astronomia, a quelle scienze in generale dove è opportuno minimizzare (con il metodo dei minimi quadrati) gli eventuali errori di misurazione. Tali suggerimenti non suscitano, nell'immediato e nei decenni subito successivi, interessi smodati nonostante la notorietà e l'autorevolezza di Fourier. Vengono ripresi da due suoi allievi, quali C.L.M.H. Navier (1825) e A.A. Cournot (1827) che assegna la condizione necessaria, per il problema di minimo con cui Fourier aveva formalizzato la ricerca della posizione di equilibrio, con argomentazioni *ad hoc*, che fanno specifico riferimento all'interpretazione meccanica dei termini in questione. Lo stesso teorema compare in termini più generali, ma sempre privo di una dimostrazione completa, in un articolo del 1838 di M.V. Ostrogradsky – da studente, a Parigi, prima di tornare a S. Pietroburgo, aveva seguito i corsi di Fourier – che asserisce che nel punto di minimo il gradiente della funzione obiettivo può essere espresso come combinazione lineare, con coefficienti non negativi, dei gradienti delle funzioni di vincolo; i moltiplicatori hanno invece segno qualsiasi quando i vincoli vengono espressi tramite uguaglianze.

Per il resto, per tutto l'Ottocento non ci sono significativi "ritorni di fiamma" verso le problematiche sollevate da Fourier (prime formalizzazioni di un problema (P_3), metodi risolutivi per problemi lineari, studio dei sistemi di disequazioni lineari). Vengono al più citati i nomi di Gauss (1829, sempre in relazione a problemi di Meccanica), di Boole (1854), 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, appunto, sotto forma di disuguaglianza.

L'interesse riprende, sul finire del secolo, con gli studi di P. Gordan (1873), che per primo assegna condizioni necessarie e sufficienti a proposito delle soluzioni di un sistema di disequazioni lineari, e soprattutto con la pubblicazione di alcuni articoli del matematico ungherese Julius Farkas, il cui ruolo verrà ulteriormente enfatizzato quando Kuhn e Tucker dimostreranno il loro teorema servendosi proprio di quello che adesso viene comunemente chiamato *lemma di Farkas* (o di Farkas-Minkowski). Di questa proposizione vengono subito date, a partire dal 1894, varie dimostrazioni (in ungherese o in tedesco) via via più complete e convincenti fino a quella che può considerarsi la versione definitiva, contenuta nell'articolo del 1902: "Theorie der einfachen Ungleichungen"[6]. Farkas (1847-1930) – professore di Fisica teorica all'Università di Kolozsvár e membro autorevole dell'Accademia delle Scienze ungherese, noto in particolare per i suoi contributi in Meccanica e termodinamica – dimostra che ogni soluzione della disuguaglianza matriciale $Ax \geqq 0$ (A matrice di tipo (m, n), le cui righe verranno indicate con a_i; $x \in R^n$) è anche soluzione della disequazione $bx \geq 0$ se e solo se esistono m numeri non negativi λ_i tali che $b = \sum_{i=1}^{m} \lambda_i a_i$. Un enunciato simile compare pressoché contemporaneamente nel libro di H. Minkowski, *Geometrie der Zahlen* (1ª edizione: 1896), all'interno della ricerca – per un sistema di disequazioni lineari omogenee – delle disequazioni ridondanti, tali cioè da poter essere cancellate senza alterare l'insieme delle soluzioni. L'applicazione del "lemma" di Farkas alla corretta soluzione del problema di equilibrio studiato da Fourier appare ora pressoché immediata. Sia $x°$ il punto di minimo della funzione potenziale V. Se $x° + dx$ rappresenta una diversa posizione nella regione ammissibile (tale cioè da soddisfare i vincoli $g_i\ (x° + dx) \geq 0$, in ipotesi di differenziabilità avremo $dg_i(x°) = \nabla g_i\ (x°)\ dx \geq 0$. Il viceversa non vale, ma se imponiamo una opportuna condizione di qualificazione dei vincoli (che ovviamente manca in Farkas), avremo che ogni dx tale che $\nabla g_i(x°)\ dx \geq 0$ appartiene anche alla regione ammissibile ed è tale, cioè, che $dV(x°) = \nabla V(x°)\ dx \geq 0$. Una volta "linearizzato" il problema, il "lemma" di Farkas garantisce l'esistenza di moltiplicatori non negativi λ_i per cui risulta $\nabla V(x°) = \sum \lambda_i\ \nabla gi(x°)$.

Farkas non pubblicherà nessun altro lavoro sulle disequazioni lineari fino agli anni Venti. Nel periodo tra le due guerre mondiali un altro matematico ungherese, A. Haar, generalizzerà il suo risultato ai sistemi non omogenei (1924). Ma il primo effettivo riconoscimento dell'importanza del lavoro di Farkas per i problemi lineari verrà qualche anno dopo con la tesi di dottorato di T. Motzkin (1933). Siamo ormai alla vigilia degli atti ufficiali di nascita della programmazione, lineare e non.

---

[6] L'articolo è stato pubblicato in *Journal für die reine und angewandte Mathematik*, pp. 1-27.

U.R.S.S e U.S.A.

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'origine 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, 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 URSS e USA i Paesi che registrano i primi fondamentali sviluppi.

Per l'Unione Sovietica il padre della programmazione lineare è Leonid Vitalevic Kantorovich (1912-1996)[7]. Giovane professore di Leningrado,

---

[7] Kantorovich ha insegnato all'Università di Leningrado dal 1934 al 1960. All'interno della comunità matematica è ricordato soprattutto per i suoi notevoli contributi all'analisi funzionale.

viene contattato nella primavera del '36, per una consulenza, da un'azienda che produce legno compensato e intende rendere più efficiente l'utilizzo del proprio macchinario; si tratta di aumentare il livello della produzione di cinque diversi tipi di compensato, realizzati in otto impianti dotati di diversa capacità produttiva, considerando come dati i rapporti tra le quantità prodotte dei tipi di compensato. Kantorovich si accorge che la questione propostagli ha una struttura matematica che la accomuna ad altre situazioni (apparentemente diverse) che si presentano «in the organization and planning in the fields of industry, construction, transportation and agricolture» e a questo tipo di problemi, alla loro comune formulazione matematica e alla loro risoluzione – anche numerica – dedica un lungo articolo pubblicato nello stesso anno[8] e che sarà seguito da *Ekonomicheskii raschet nailuchshego ispol'zovanic resursov*[9] e da tutta una serie di ulteriori contributi. La programmazione lineare nasce così, in Unione Sovietica, strettamente collegata alle esigenze della sfera produttiva, all'interno del terzo piano quinquennale, nella speranza che «will play a very useful role in the development of our socialist industry».

Nell'articolo del '39 Kantorovich presenta una serie di problemi microeconomici (che traggono spunto dalla pianificazione della produzione e dalla gestione di determinate imprese), che inquadra in tre schemi generali. Il primo di questi - gli altri ne rappresentano una variante più generale - considera l'allocazione di *m* macchine, che possono produrre *n* prodotti. Se a_{ij} indica il numero di unità del j-esimo output prodotte nell'unità di tempo dalla i-esima macchina; x_{ij} il numero di macchine dell'i-esimo tipo allocate per produrre l'j-esimo output e z_j il numero totale di unità prodotte sempre dell'j-esimo output, avremo:

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

dove b_i rappresenta il numero di macchine disponibili del tipo i-esimo. I vincoli sono chiari. Per fissare invece la funzione obiettivo, supponiamo (con un'ipotesi che verrà successivamente rimossa) 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 $z_0 = \min_j z_j$.

Una volta formalizzato il problema, Kantorovich dimostra (sia analiticamente che geometricamente) l'esistenza delle variabili duali, associate a ciascun vincolo e chiamate *moltiplicatori risolventi*, che più tardi verranno

[8] (Kantorovich 1960) è la traduzione in lingua inglese di questo saggio.

[9] La sua traduzione inglese è *The best use of economic resources*, Oxford, Oxford University Press, 1965.

precisate come «objectively determined evaluations» dato che i loro valori non dipendono da variabili esogene (quali prezzi o costi) ma dalla stessa natura fisica del problema[10]. La sua principale attenzione è comunque rivolta alle procedure di calcolo. La dimostrazione dell'esistenza dei moltiplicatori – anche se cruciale – è pertanto "relegata" nell'ultima appendice, con la precisazione che «the ignorance of the proof [...] in no way interferes with mastering the method of its practical application». Il metodo dei "moltiplicatori risolventi" deve soprattutto essere «sufficiently simple and effective» perché i problemi in questione, al di là del loro interesse teorico, hanno una rilevanza pratica che richiede «the solution of ten of thousands or even millions of systems of equations for completion». Kantorovich individua così un algoritmo risolutivo che, nel lavoro del '39, è comunque ancora applicato ai diversi problemi, piuttosto che descritto accuratamente nella sua generalità; formalmente diverso da quello successivo di Dantzig, è nella sostanza equivalente a questo come è stato poi dimostrato[11].

Abbiamo presentato Kantorovich, negli anni Trenta, come giovane professore di Leningrado. È un fatto conosciuto come, con il passare degli anni, la sua autorevolezza all'interno della comunità scientifica sovietica sia aumentata considerevolmente. Eppure, la diffusione delle sue idee in URSS si è rivelata estremamente difficoltosa. Attraverso gli ostacoli frapposti allo sviluppo della programmazione lineare possiamo, anzi, ricostruire una buona parte del dibattito[12] che nel secondo dopoguerra ha contrapposto i "pianificatori ottimali" al potere politico e all'ideologia più ortodossa. È noto come Stalin, già ai suoi esordi, avesse classificato l'economia matematica tra i semplici "giochi con i numeri". L'espressione, al di là della sua volgarità, voleva specificare come la pianificazione e l'organizzazione ottimale delle forze produttive non fossero problemi di economia politica bensì di politica economica, appartenenti dunque alla precisa sfera di competenza dei politici; il ruolo degli economisti era un altro: apologeti del sistema, a loro toccava il compito di elaborare un quadro teorico che giustificasse *a posteriori* le scelte politiche. La situazione comincia a cambiare solo negli anni '60 ed è in questo periodo che le idee di Kantorovich hanno una prima diffusione non clandestina. Nel '58 viene costituito il Laboratorio matematico-economico dell'Accademia delle Scienze, da cui

[10] È evidente la preoccupazione di Kantorovich di non uscire dai confini dell'ortodossia, a proposito della marxiana teoria del valore.

[11] Cfr. van de Panne C., Rahnama F. (1985) The First Algorithm for Linear Programming: An Analysis of Kantorovich's Method, *Economics of Planning*, 19, 2, pp. 76-91.

[12] Anche in questo caso la letteratura comincia ad essere sufficientemente ampia. Si può vedere, tra gli altri, tradotto in italiano, (Ellman 1973).

succesivamente nascerà il TSEMI. Nello stesso anno Kantorovich è eletto membro corrispondente dell'Accademia delle Scienze (che nel '65 gli conferirà il premio Lenin). Nel '60 si tiene a Mosca un convegno sull'impiego dei metodi matematici nella scienza economica e nella pianificazione, che consolida quel percorso di legittimazione che idealmente si chiuderà dieci anni dopo quando, sempre a Mosca, verrà organizzato un incontro sulla pianificazione ottimale (1971).

Anche in Occidente – per altre ragioni – la conoscenza dei primi elementi della teoria e dei metodi della programmazione lineare è lenta. La guerra, i problemi di lingua, e poi la fase della "guerra fredda", sono tutti elementi che spiegano un ritardo che, anche in questo caso, si aggira sui venti anni. È T.C. Koopmans[13] che fa conoscere agli studiosi occidentali le idee di Kantorovich, con la traduzione su "Management Science", nel 1960, dei contributi presentati a Leningrado nel '39.

Lo stesso Koopmans – premio Nobel nel '75 per l'Economia, assieme a Kantorovich – una decina d'anni prima aveva contribuito in modo significativo alla diffusione (soprattutto tra gli economisti) dei risultati di George Bernard Dantzig, il "padre occidentale" di una programmazione lineare maggiormente legata – all'atto della sua nascita – a ricerche svolte in ambiente militare.

Durante la guerra, dal '42 al '44, Koopmans 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»[14] 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

---

[13] Koopmans (1910-1985) è nato in Olanda e ha compiuto i suoi studi nelle Università di Utrecht e di Leiden. Negli U.S.A. ha insegnato a Chicago e poi (dal '55) a Yale. Dal '61 al '67 è stato direttore della "Cowles Foundation".

[14] Queste parole di Dantzig, così come le sue successive citazioni, sono tratte da (Dantzig 1991).

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[15], 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[16]. 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

---

[15] Wassily Leontief è nato a S. Pietroburgo nel 1906. Si è laureato a Berlino nel '28. Subito dopo è stato consulente economico del governo cinese. Dal '31 si è trasferito negli USA; durante la seconda guerra mondiale, in particolare, è stato consulente del *U.S. Department of Labor* per lo studio dei possibili effetti del disarmo.

[16] Il riferimento "classico" è qui a von Neuman J., Morgenstern O. (1953) *Theory of Games and Economic Behavior*, Princeton, Princeton Univ. Press.

advent or rather *the promise* that the electronic computer would exist soon, the exposure of theoretical mathematicians and economist 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». Gli Atti verranno pubblicati dallo stesso Koopmans[17], 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".

La nascita

La nascita della programmazione *non* lineare avviene in anni e in ambienti molto prossimi a quelli che abbiamo avuto modo di ricordare a proposito di

---

[17] Koopmans T.C. (ed.) (1951) *Activity Analysis of Production and Allocation*, New York, Wiles.

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[18], 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[19]. 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 sotto

---

[18] Tra gli scritti cui Kuhn ha affidato i suoi ricordi faremo riferimento a Kuhn 1976 e all'articolo Non linear Programming. A historical Note, pubblicato in *History of Mathematical Programming* (Lenstra J.K., Rinnaykan A.N.G., Schrijver A. (eds) (1991) Amsterdam, North Holland, pp. 82-96.

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

forma di disugualianza. 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.

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 vengono in particolare ricordati tre risultati. Il primo è la cosiddetta condizione di Carathéodory[20], 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° è soluzione locale del problema $Min\, f(x)$, con $x \in S = \{x \in A,\ h_j(x) = 0,\ j = 1, ..., r < n\}$, esiste un vettore non nullo $\lambda^\circ = (\lambda_0^\circ, \lambda_1^\circ, ..., \lambda_r^\circ)$ per cui le derivate della funzione lagrangiana $L = \lambda_0^\circ f + \sum_{j=1}^{r} \lambda_j^\circ h_j$ si annullano in x°.

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 $\lambda_0^\circ = 1$): se x° è soluzione del problema in questione e la matrice jacobiana $\nabla h(x^\circ)$ ha rango pieno, allora il differenziale secondo in x° della funzione lagrangiana:

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

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

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

soddisfa le condizioni $L_{x_i}(x^\circ) = 0$ e $d^2 L(x^\circ) > 0$ per ogni $dx \neq 0$ tale che $\nabla h_j(x^\circ)\, dx = 0$, allora x° 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 $g_i(x) \geq 0$ e vengono considerati tutti attivi in x°, dato che per ragioni di continuità un eventuale vincolo del tipo $g_i(x^\circ) > 0$ non impone localmente alcuna restrizione.

La prima condizione necessaria, se letta frettolosamente, può dare l'impressione – erronea – di anticipare quello che nel secondo paragrafo abbia-

[20] Il teorema di Carathéodory è un "teorema à la John" (con un moltiplicatore non necessariamente unitario associato alla funzione obiettivo), relativo però al problema (P_2).

mo indicato come teorema di John: se x° 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 $\lambda^\circ = (\lambda_0{}^\circ, \lambda_1{}^\circ, \ldots, \lambda_m{}^\circ)$ tale che x° è stazionario per la funzione lagrangiana:

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

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° (che chiama delle direzioni ammissibili) delle direzioni $dx \neq 0$ tali che $\nabla g_i(x^\circ)\, dx \geq 0$ e dimostra la seguente condizione necessaria: «suppose that for each admissible direction there is an admissible arc issuing from x° in the direction dx[21]. Then a first necessary condition for f(x°) to be a minimum is that there exist multipliers $\lambda_i \leq 0$ such the derivatives L_{x_i} of the function $L = f + \sum \lambda_j g_j$ all vanish at x°». 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 $x(0) = x^\circ$ e $x'(0) = dx$. Da $f[x(t)] \geq f(x^\circ) = f[x(0)]$, $\forall t \in [0,t_0]$, segue immediatamente $f'[x(0)] \geq 0$ ovvero $\nabla f(x^\circ) \cdot x'(0) = \nabla f(x^\circ)\, dx \geq 0$. 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 functions $g_i(x)$ will satisfy property Q» che, interpretata geometricamente, ipotizza che per ogni direzione del cono linearizzante – e quindi, in sostanza,

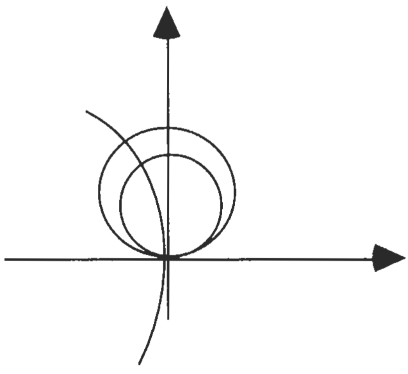
---

[21] Questa condizione di qualificazione dei vincoli viene solitamente citata ora come "condizione di Kuhn-Tucker". Quella che Karush introdurrà successivamente è riportata come "condizione di Cottle"; R.W. Cottle l'ha introdotta, nel '63, nell'articolo: A Theorem of Fritz John in *Mathematical Programming*, RAND Corporation Memorandum. La condizione di Cottle implica quella di Kuhn-Tucker. 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*.

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 $\nabla g_i(x^\circ)\, dx > 0$, $\forall i$. In questo caso, infatti, chiamata con $x(t)$ una curva regolare definita su un intervallo $[0,t_0]$ e tale che $x(0) = x^\circ$ e $x'(0) = dx$, si ha $g_i'(x^\circ) = \nabla g_i(x^\circ)\, x'(0) > 0$; le funzioni $g_i[x(t)]$ risultano crescenti e da $g_i[x(t)] \geq g[x(0)] = 0$ segue che $x(t)$ è un arco ammissibile. Di conseguenza, è $f[x(0)] \leq f[x(t)]$ e quindi $f'[x(0)] = \nabla f(x^\circ)\, dx \geq 0$. 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 $\nabla g_i(x^\circ)\, dz > 0$ e quindi (ripetendo quanto dimostrato per la direzione dx) $\nabla f(x^\circ) dz \geq 0$. La linearità di quest'ultima funzione permette di estendere la disuguaglianza al caso $t = 1$: $\nabla f(x^\circ)\, dy \geq 0$, che è dunque una "conseguenza" delle disuguaglianze $\nabla g_i(x^\circ)\, dy \geq 0$. 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 $x^\circ = (x_0,y_0) = (0,0)$; $g_1(x,y) = x^2+(y-1)^2-1$, $g_2(x,y) = 4 - [x^2+(y-2)^2]$, $g_3(x,y) = x+y^2$.



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° è di minimo se $m \geq n$, la matrice jacobiana $\nabla g_i(x^\circ)$ ha rango n ed esiste un vettore di moltiplicatori $\lambda^\circ=(\lambda_1^\circ, ..., \lambda_m^\circ)$ a componenti negative, tale che la funzione $L(x) = f(x) + \sum_{i=1}^{m} \lambda_i^\circ gi(x)$ annulla in x° 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° per cui la matrice $\nabla g_i(x^\circ)$ ha rango m, e che la provenienza dal calcolo delle variazioni suggerisce di chiamare *normale*, afferma che condizione necessaria perché x° sia di minimo è che $L_{x_i}(x^\circ) = 0$ (con $L = f + \sum_{i=1}^{m} \lambda_i\, g_i$ e $\lambda_i \leq 0$) e che risulti $d^2 L(x^\circ) \geq 0$ in corrispondenza ad ogni direzione ammissibile dx che soddisfi le condizioni $\nabla g_i(x^\circ)dx = 0$. Il teorema 6.1 assicura che x° è soluzione del problema in questione se è possibile associargli un vettore $\lambda^\circ = (\lambda_1^\circ, ..., \lambda_m^\circ)$ a componenti negative, per il quale risulta $L_{x_i}(x^\circ) = 0$ e $d^2L(x^\circ) > 0$ per ogni direzione ammissibile dx soddisfacente le condizioni $\nabla g_i(x^\circ)\, dx = 0$ (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 (John 1948),[22] nel volume che festeggiava i 60 anni di Richard Courant.

---

[22] John (1910-1994), allievo di Courant a Göttingen, fu costretto a lasciare la Germania per motivi razziali. Dopo un breve soggiorno in Inghilterra, si trasferì negli Stati Uniti

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° è un punto di minimo, nelle ipotesi dette, esistono un numero finito di punti $y_1, \ldots, y_s \in S$ (con $s<n$) e un vettore $\lambda^\circ = (\lambda_0^\circ, \lambda_1^\circ, \ldots, \lambda_s^\circ) \neq 0$ con $\lambda_0^\circ \geq 0$ e $\lambda_i^\circ > 0$ ($j = 1, \ldots, s$) per cui la funzione $L(x) = \lambda_0^\circ f(x) - \sum_{i=1}^{s} \lambda_i^\circ g(x,y_i)$ ha un punto critico in x°. 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:

$$\begin{cases} \nabla f(x^\circ)\, dx < 0 \\ \nabla g_i(x^\circ)\, dx > 0 \end{cases}$$

dove $g_i(x) = g(x,y_i)$ sono le funzioni che intervengono nei vincoli attivi in x°. 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 e da R.W. Stokes[23]. John lo ricostruisce geometricamente: l'impossibilità del precedente sistema implica che lo spazio immagine del punto x°, tramite le funzioni $(-\nabla f,$

dove collaborò con l'U.S. War Department (dal '43 al '45), insegnando poi nelle Università del Kentuchy e di New York e al Courant Institute. La sua produzione scientifica riguarda principalmente la geometria convessa, le equazioni alle derivate parziali, la teoria dell'elasticità, l'analisi numerica. I nuovi lavori sono raccolti in due volumi editi da J. Moser (Birkäuser, Basel, 1985).

[23] Dines L.L. (1936) Linear Inequalities, *Bulletin American Mathematical Society*, 42, pp. 353-365. Stokes R.W. (1931) A geometric theorem of linear inequalities, *Transactions American Mathematical Society*, 33, pp. 782-805.

∇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: se esiste una funzione della forma $L(x) = \lambda_0^\circ f(x) - \sum_{i=1}^{s} \lambda_i^\circ g(x,y_i)$, con $g(x^\circ,y_i) = 0$, $\lambda_0^\circ \geq 0$ e $\lambda_i^\circ > 0$, per cui x° è punto critico e inoltre la matrice:

$$\begin{bmatrix} \lambda_0^\circ f_{x1}(x^\circ) & g_{x1}(x^\circ,y_1) & \ldots\ldots & g_{x1}(x^\circ,y_s) \\ \lambda_0^\circ f_{x2}(x^\circ) & g_{x2}(x^\circ,y_1) & \ldots\ldots & g_{x2}(x^\circ,y_s) \\ \ldots\ldots\ldots\ldots & & & \\ \lambda_0^\circ f_{xn}(x^\circ) & g_{xn}(x^\circ,y_1) & \ldots\ldots & g_{xn}(x^\circ,y_s) \end{bmatrix}$$

ha rango n, allora x° è un punto di minimo rispetto all'insieme individuato in R dalle disuguaglianze $g(x,y_i) \geq 0$ ($i = 1, \ldots, s$) e quindi a maggior ragione rispetto a R'. Né, comunque, da quest'ultima conclusione si può dedurre in generale l'equivalenza tra infinite disuguaglianze e un suo sottoinsieme finito, al fine di determinare i punti di minimo. Ad esempio, se si tratta di minimizzare la funzione $f(x) = -x^2$ con i vincoli $g(x,y) = y-x^2 \geq 0$, dove y varia nel compatto $[0,1]$, si ottiene quale soluzione $x^\circ = 0$ in quanto la regione ammissibile è costituita da questo unico punto; se invece i precedenti infiniti vincoli sono sostituiti da un numero finito di disuguaglianze del tipo $y_i - x^2 \geq 0$, la regione ammissibile viene a coincidere con un opportuno intorno dell'origine e f non ammette minimo.

È a questo punto (paragrafo 3) che si apre lo spazio dedicato alle applicazioni geometriche. Se è corretta la ricostruzione di Kuhn, per cui sarebbe stato proprio un simile interesse a portare John verso la programmazione non lineare, la collocazione delle considerazioni di carattere geometrico nella parte terminale dell'articolo, come "semplice" esemplificazione della parte analitica, è una ulteriore testimonianza dello *stile* "moderno", ormai impostosi verso la metà del secolo: «the main impulse came from trying to prove the theorem [...] that asserts that the boundary of a compact convex set S in R^n lies between two homotetic ellipsoid of ratio $\leq$ n, and that the outer ellipsoid can be taken to be the ellipsoid of least volume containing S. The case $n = 2$ had been settled by F. Behrend with whom John had become acquainted in 1934 in Cambridge, England. A student of John's, O.B. Ader dealt with the case $n = 3$ in 1938. By that time, John had become deeply interested in convex sets and in the inequalities connected with them». Il punto di partenza, per questa che rimane la principale applicazione geometrica nel lavoro del '48, è comunque costituito da una generalizzazione a R^n del problema che J.J. Sylvester aveva posto nel 1857 relativamente al caso bidimensionale. Si tratta ora di trovare la sfera di minimo raggio contenente un assegnato insieme limitato $S \subset R^n$, adottando quindi come variabili decisionali le coordinate del centro della sfera $x = (x_1, ..., x_n)$ e la misura x_{n+1} del raggio. John "usa" questo problema come applicazione della sua condizione necessaria, dopo averlo formalizzato come ricerca del minimo della funzione $f(x,x_{n+1}) = x_{n+1}$, tenendo conto dei vincoli $x_{n+1} - d(x,y) \geq 0$, $\forall\, y \in S$.

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 apparire 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 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: $x°$ è un punto di minimo per una funzione lineare f, considerando i vincoli $g_i(x) = b_i - \sum_{j=1}^{n} a_{ij} x_j \geq 0$ e $x_j \geq 0$, se e solo se esiste un moltiplicatore $\lambda°$ a componenti non negative tale che $(x°,\lambda°)$ è punto di sella per la funzione lagrangiana $L = f + \lambda\, g$ ovvero soddisfa le disuguaglianze $L(x,\lambda°) \leq L(x°,\lambda°) \leq L(x°,\lambda)$, per ogni x e λ 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 $\lambda_i \geq 0$. La condizione necessaria richiede che nel punto $(x°,\lambda°)$ siano soddisfatte le condizioni:

$$\varphi_{x_j} \leq 0; \qquad x_j\, \varphi_{x_j} = 0; \qquad x_j \geq 0; \tag{1}$$
$$\varphi_{\lambda_i} \geq 0; \qquad \lambda_i\, \varphi_{\lambda_i} = 0; \qquad \lambda_i \geq 0. \tag{2}$$

Le stesse condizioni sono sufficienti a garantire che (x°,λ°) è un punto di sella per φ, se ad esse si aggiungono le disuguaglianze:

$$\varphi(x,\lambda^\circ) \leq \varphi(x^\circ,\lambda^\circ) + \nabla_x\, \varphi(x^\circ,\lambda^\circ) \cdot (x-x^\circ) \tag{3}$$
$$\varphi(x^\circ,\lambda) \geq \varphi(x^\circ,\lambda^\circ) + \nabla_\lambda\, \varphi(x^\circ,\lambda^\circ) \cdot (\lambda-\lambda^\circ) \tag{4}$$

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 φ rispetto a x (rispetto a λ).

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 $g_1 = x \geq 0$, $g_2 = y \geq 0$, $g_3 = (1-x)^3 - y \geq 0$ 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 λ° 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° 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 $\nabla f(x^\circ)dx \leq 0$; 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 defіnzione 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 $x°$ è punto di massimo se e solo se per qualche $\lambda°$, con $\lambda_i° \geq 0$, $(x°,\lambda°)$ è di sella per $L = f + \lambda\, g$ dove f e g_i ($i = 1, 2, ...$) 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 (1) si riduce alle uguaglianze $L_{x_s} = 0$. 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.

Questa, invero, ha un ultimo momento di grande interesse quando applica il consueto schema di una condizione necessaria che diventa anche sufficiente, aggiungendo un'ulteriore ipotesi, al problema di ottimo *vettoriale*.

In conclusione del secondo paragrafo, abbiamo accennato alla difficoltà che qui si presenta subito nella definizione di punto di massimo (chiamato *efficiente*, nell'ottimizzazione vettoriale). Un punto $x°$ della regione ammissibile K è detto Pareto-efficiente (di massimo) quando il decisore non trova "conveniente" spostarsi in un altro punto di K ovvero ogni altro $x \in K$, magari permette di migliorare qualche criterio f_i, ma sicuramente peggiora almeno una *performance*; in altre parole, $x^\circ \in K$ è Pareto-efficiente quando per ogni $x \in K$ esiste i, per cui $f_i(x) < f_i(x^\circ)$. L'articolo di Kuhn e Tucker è il primo in cui viene data una definizione più restrittiva di efficienza. Un punto $x^\circ \in K$ viene detto *propriamente efficiente* quando è efficiente e per nessuna direzione dx del cono linearizzante in x° si ha $\nabla f_i(x^\circ)\, dx \geq 0$ con almeno un indice j per cui risulta $\nabla f_j(x^\circ)\, dx > 0$. Nei decenni successivi si aggiungeranno diverse definizioni di efficienza propria, quasi tutte più restrittive di quella di Kuhn-Tucker, alcune basate sempre sulla nozione di *trade-off*, altre più geometriche e riferite allo spazio immagine[24]. Al di là dell'apparente ermetismo, le motivazioni alla base della considerazione dell'efficienza propria sono sufficientemente chiare. Si tratta di eliminare delle anomalie, rappresentate da punti che, pur essendo efficienti, non risultano completamente affidabili. Un punto x°, per esempio, è *impropriamente* efficiente secondo Kuhn-Tucker quando esiste una direzione dx del cono linearizzante per la quale si ha $f_j(x^\circ+dx) > f_j(x^\circ)$ mentre le perdite $f_i(x) - f_i(x^\circ) < 0$, che si registrano negli altri criteri, sono di ordine superiore al primo, risultando $df_i(x^\circ) = 0$. È chiara l'anomalia: x° è efficiente; eppure, valutando costi e ricavi e il loro *trade-off*, il decisore è egualmente incentivato a spostarsi, raggiungendo il punto $x^\circ + dx$. Il semplice esempio fornito da Kuhn eTucker chiarisce ulteriormente la situazione. Sia la funzione obiettivo data da:

$$f(x) = (f_1(x), f_2(x)) = (x,\, 2x - x^2), \quad x \geqq 0.$$

Il punto $x^\circ = 1$ è efficiente, perché qualunque spostamento peggiora (almeno) i valori assunti da f_2; lo stesso punto non è però propriamente efficiente, come si constata considerando la direzione $dx = 1$.

Come dicevamo, anche la struttura delle proposizioni che riguardano il problema vettoriale ricalca lo schema già visto. Condizione necessaria perché x° sia un punto propriamente efficiente (di massimo) per $f = (f_1, f_2, \ldots, f_p)$, quando si considerino i vincoli $g_i(x) \geq 0$ e $x_s \geq 0$, è che esistano un moltiplicatore $\theta^\circ \in R^p$ a componenti positive e un moltiplicatore λ°, a componenti non negative, per cui la funzione $L = \theta f + \lambda\, g$ soddisfa le condizioni (1) e (2) in $(x^\circ, \theta^\circ, \lambda^\circ)$. La stessa condizione diventa sufficiente se, alle precedenti condizioni, aggiungiamo la disuguaglianza (3). Infine, in ipotesi di concavità per le funzioni obiettivo e di vincolo, x° è punto di massimo proprio *se e solo se* (x°, λ°) è di sella per la funzione $L = \theta^\circ f + \lambda\, g$. L'unica osservazione che si può aggiungere riguarda la "constraint qualification":

[24] Per le diverse definizioni di efficienza propria si può vedere Guerraggio A., Molho E., Zaffaroni A. (1994) On the notion of proper efficiency in vector optimization, *Journal of Optimization. Theory and Applications*, 82, 1, pp. 1-21.

nel caso vettoriale, anche in assenza di una condizione di regolarità dei vincoli riusciamo ad avere $\theta^\circ \neq 0$ perché ora è più forte l'ipotesi di massimo (che richiede che x° sia *propriamente* efficiente).

Alcune considerazioni

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 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, 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 – considerando anche quella vettoriale, cui abbiamo dedicato qualche spunto – 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 non lineare 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 non lineare 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 – 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 di questo secolo 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.

Riferimenti bibliografici

Dantzig G.B. (1991) Linear Programming, in *History of Mathematical Programming* (Lenstra J.K., Rinnoykan A.H.G., Schrijver A. eds.) Amsterdam, North Holland, pp. 19-31.

Dantzig G.B. Thapa M.N. (1997) *Linear Programming*, 1: Introduction, Berlin-Heidelberg, Springer Verlag.

Dorn W.S. (1963) Non linear programming. A survey, *Management Science*, 9, 2, pp. 171-208.

Ellman M. (1973) *Planning Problems in the URSS*, Cambridge Univ. Press, Cambridge (trad. it. *I problemi della pianificazione in URSS*, Napoli, Liguori ed., 1979).

Grattan-Guinness I. (1976) Joseph's Fourier anticipation of Linear programming, *Operational Research Quarterly*, 27, pp. 361-364.

Grattan-Guinness I. (1989) On the Prehistory of Linear and Non-linear Programming, in *The History of Modern Mathematics*, III, (Knobloch E., Rowe D.E. eds.), Academic Press, London, pp. 43-89.

Ingrao B., Israel G. (1987) *La mano invisibile*, Roma-Bari, Laterza.

John F. (1948) Extremum problems with Inequalities as Subsidiary Conditions, *Studies and Essays Presented to R. Courant on his 60.th Birthday*, pp. 187-204, New York, Interscience.

Kantorovich L.V. (1960) Mathematical methods of organizing and planning production (trad. inglese di "Mathematicheskie metody organizatriya planirovanija proizvodska", 1939), *Management Science*, 6, 4, pp. 366-422.

Karush W. (1939) Minima of functions of several variables with inequalities as side conditions, MA Thesis, Univ. Chicago

Kuhn H.W. (1976) Non Linear Programming. A historical view, SIAM-AMS Proc., pp. 1-26.

Kuhn H.W., Tucker A.W. (1951) Non linear programming, Proc. of *II Berkeley Symposium on Mathematical Statistics and Probability*, Neyman J. ed., University of California Press, pp. 481-492.

Prékopa A. (1980) On the development of optimization theory, *American Mathematical Monthly*, 87, pp. 527-542.

Pourciaiu B.H. (1980) Modern Multipliers Rules, *American Mathematical Monthly*, 87, pp. 433-452

Tikhomirov V.M. (1996), The Evolution of Methods of Convex Optimization, *American Mathematical Monthly*, 103, pp. 65-71.

Weintraub E.R. (1985) *General Equilibrium Analysis*, Cambridge, Cambridge Univ. Press.