Modelli e Algoritmi di Ottimizzazione

2024-2025

Introduzione e problema del simplesso

Definiamo *algoritmo* la struttura matematica della soluzione di un determinato problema. Affrontiamo in questo corso gli algoritmi legati all'ottimizzazione. Un *problema di ottimizzazione* consiste nella ricerca del valore ideale di una *variabile obiettivo* ignota, in base a vincoli e relazioni tra *variabili decisionali* date, allo scopo di raggiungere un determinato fine. Per poter risolvere il problema, è necessario conoscere l'obiettivo dell'ottimizzazione. Può capitare in alcuni casi che possano coesistere diversi criteri secondo i quali cercare di ottimizzare. Ad esempio, in un'azienda, gli interessi dei dirigenti e gli interessi dei lavoratori possono entrare in contrasto. Si parla in questo caso di *programmazione multiobiettivo*. Nello scenario della programmazione multiobiettivo rientrano anche i modelli che affrontano l'incertezza futura. In questi problemi, infatti, si può decidere di massimizzare l'obiettivo principale, oppure di minimizzare il rischio dovuto all'incertezza futura. Sono inoltre distinguibili fin da subito due categorie di problemi di ottimizzazione, in base alla complessità matematica degli algoritmi risolutivi. Il primo gruppo, di più semplice risoluzione, coinvolge tutte e sole variabili continue, ed è detto *programmazione lineare* (LP). La seconda categoria, detta MIP (*mixed-integer programming*) contiene invece una o più variabili discrete, o addirittura binarie. La risoluzione è resa notevolmente più complessa dalla necessità di garantire un risultato intero per la variabile obiettivo.

Affrontiamo innanzitutto l'*algoritmo del simplesso* per la risoluzione di problemi di programamzione lineare. Esso fu sviluppato al termine della Seconda Guerra Mondiale da George Dantzig, addetto di logistica militare presso l'esercito americano, ai fini di generalizzare la soluzione ai problemi da lui affrontati nella propria attività lavorativa. Il primo passaggio consiste nella determinazione della variabile obiettivo. Si definiscono, in secondo luogo, le relazioni tra le variabili decisionali. Sono infine imposti vincoli sulle singole variabili decisionali, lasciando libera la variabile obiettivo. L'algoritmo non consente l'uso diretto delle disequazioni nonostante esse siano una parte fondamentale dei problemi di ottimizzazione. Per questa ragione, le disequazioni si riconducono ad equazioni mediante l'utilizzo di variabili scarto positive o nulle che bilanciano i due membri. Convenzionalmente esse sono positive o nulle e si aggiungono o sottraggono al termine sinistro. Le equazioni ottenute sono messe a sistema, ordinate per variabile. I coefficienti si possono raccogliere in forme matriciali o vettoriali. Le equazioni ottenute sono messe a sistema, ordinate per variabile. I coefficienti si possono raccogliere in forme matriciali o vettoriali. Questo tipo di rappresentazione permette di avere modelli compatti e facilmente riutilizzabili al variare dei valori delle variabili.

$$\begin{cases} 3x + 2y - 4z < 3 \;\rightarrow\; 3x + 2y + v_1 = 3 \\ 2x - y + z > 4 \;\rightarrow\; 2x - y + z - v_2 = 4 \\ x - 4y + 9z < 11 \;\rightarrow\; x - 4y + 9z + v_3 = 11 \end{cases}$$

$$\begin{bmatrix} 3 & 2 & -4 \\ 2 & -1 & 1 \\ 1 & -4 & 9 \end{bmatrix} \quad \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \quad \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$$

Figura 1: Trasformazione di una disequazione in equazione usando una variabile scarto

Definiamo ora il concetto di *problemi sparsi*. Essi sono contraddistinti da una matrice dei coefficienti *sparsa*, contenente cioè molti elementi nulli. Il sistema che ne deriva è detto *sottodeterminato*. Si definisce un *coefficiente di sparsità* che rappresenta la quota di coefficienti nulli rispetto al numero totale di elementi nella matrice. Tanto più è alto il coefficiente di sparsità, tanto più inefficiente risulta il calcolo della soluzione. Molto tempo di elaborazione, infatti, risulterà sprecato in operazioni dal risultato nullo. Al fine di mitigare questo problema sono definite strutture dati apposite, che permettono di compattare le matrici sparse sfruttando l'algebra dei puntatori. Così facendo, sono risparmiati spazio in memoria e tempo di calcolo.

$$\begin{pmatrix} 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 7 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 5 \\ 6 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figura 2: Esempio di matrice sparsa

Questo corso impiega il *software* GAMS per la risoluzione dei problemi di ottimizzazione. Nel caso del problema del simplesso, è in grado di trasformare automaticamente le disequazioni in equazioni, mostrando anche le variabili scarto generate a fine elaborazione. L'input ha una sintassi molto leggibile, che permette all'utente di definire variabili ed equazioni, e dare direttive, in linguaggio quasi naturale. Il programma è in grado di gestire input vettoriale e matriciale fino a dimensione 10. La sezione dell'*output* è divisa in cinque sottosezioni. Esse sono *compilation*, *equation listing*, *column listing*, *model statistics* e *solution report*. *Compilation* esprime le informazioni relative alla compilazione dello script fornito dall'utente. *Equation Listing* mostra le disequazioni generate a partire da variabili e relazioni specificate dall'utente nello script. Queste ultime sono ancora mostrate come disequazioni, prima della trasformazione in equazioni con la tecnica precedentemente descritta. In *Column Listing* sono mostrati, sotto forma di vettori colonna, i coefficienti del sistema di equazioni generato a partire dalle equazioni. *Model Statistics* esprime le statistiche relative al modello generato. *Solution Report* contiene infine la soluzione vera e propria al problema di ottimizzazione, divisa in due sezioni. La prima sezione, EQU, indica come si colloca la soluzione rispetto ai vincoli, ed elenca le variabili scarto generate nella trasformazione delle disequazioni in equazioni. La seconda sezione, VAR, mostra nel medesimo formato le variabili decisionali. Il software, nell'enumerare le variabili e le equazioni, distingue tra *blocks* (numero di vettori) e *single* (cardinalità totale).

Le fasi della produzione si possono schematizzare utilizzando un diagramma a nodi. Ogni nodo rappresenta un prodotto o un semilavorato. Le lavorazioni sono indicate inserendo direttamente le loro tabelle d'impiego. Successivamente si tracciano le linee di produzione che, partendo dalla fase iniziale e terminando nel prodotto finito, congiungono nodi e tabelle. Esse sono disegnate facendo riferimento alle informazioni di compatibilità tra prodotti e impianti. Una singola linea di processo può ramificarsi in più archi quando la produzione da essa rappresentata può passare indifferentemente attraverso diversi impianti. Tutti i rami devono ricongiungersi in un'unica linea entro l'arrivo al prodotto finito. Al termine della schematizzazione si impone la garanzia dei bilanci, tramite opportuni vincoli, tenendo conto di ramificazioni e ricongiungimenti delle linee di produzione. Si impone inoltre che l'impiego di tempo sia minore o uguale del tempo totale disponibile. Si costruisce poi una funzione di profitto tenendo conto dei vincoli funzionali. Il profitto si definisce come differenza tra i ricavi e la somma dei costi (sia delle materie prime, che della produzione). Il passaggio finale dell'analisi consiste nella massimizzazione della funzione di profitto.

GAMS permette di definire non solo insiemi di vincoli, ma anche sottoinsiemi di tali insiemi. È richiesto invece di definire variabili nulle per rappresentare le decisioni non attuabili. Si tratta delle combinazioni incompatibili di prodotti e impianti, che vanno a generare vincoli nulli. Questo perché è conveniente creare un vettore dei prodotti e un vettore degli impianti, e moltiplicarli per ottenere un'unica matrice decisionale. All'interno di essa le coppie non attuabili vengono semplicemente rappresentate da un valore nullo, andando a generare un certo grado di sparsità nella matrice. La forma matriciale permette di esprimere vincoli generici usando le sommatorie. Procedendo in questo modo, non risulta necessario fare considerazioni particolari per le coppie incompatibili, oltre all'annullamento delle variabili ad esse associate.

$$\begin{bmatrix} A & B & C \end{bmatrix} \begin{bmatrix} D \\ E \\ F \end{bmatrix} \begin{bmatrix} & A & B & C \\ D & A,D & B,D & C,D \\ E & A,E & B,E & C,E \\ F & 0 & B,F & C,F \end{bmatrix}$$

Figura 3: A, B, C: prodotti; D, E, F: linee; A e F sono incompatibili

Possiamo osservare i vincoli di bilancio come primissimo elemento della sezione EQU nei risultati di GAMS. Le risorse non completamente impiegate sono riconoscibili dal valore nullo del loro marginal. Dove esso non è nullo, si può invece verificare con un semplice procedimento se la struttura della soluzione cambierebbe in seguito all'aumento della risorsa scarsa. Si aggiunge un'unità e si riavvia l'ottimizzazione. Se l'aumento di produttività è pari al *marginal* del caso precedente, allora la struttura della soluzione non cambia. Se invece il valore è diverso, la struttura della soluzione è cambiata. Il marginal è detto anche prezzo ombra. Si può interpretare come la valorizzazione interna della risorsa, ovvero il suo valore ai fini della specifica produzione analizzata. Il concetto è diverso dalla valorizzazione esterna, corrispondente al valore di mercato della risorsa analizzata.

Un'altra distinzione derivabile dai risultati dell'analisi riguarda la divisione dei vincoli in attivi e inattivi. Sono definiti inattivi i vincoli che, se rimossi, non cambierebbero la soluzione. In altre parole, si tratta dei vincoli legati alle risorse non completamente sfruttate. Il punto ottimale non si trova sul confine definito da tali vincoli. I vincoli attivi, al contrario, sono legati alle risorse limitanti. Il punto ottimale si trova esattamente sul confine da essi definito. Segue dunque che i vincoli di uguaglianza siano sempre attivi. L'attività dei vincoli di disuguaglianza, invece, dipende da che si tratti di disuguaglianza stretta o non stretta. Il concetto di vincoli attivi è fondamentale nella casistica, non affrontata dal corso, dei modelli non lineari. Esiste una classe di algoritmi di ottimizzazione non lineare chiamata active set methods, basata sull'identificazione iterativa di un sottoinsieme di vincoli che si ritiene saranno attivi nel punto ottimale.

Esiste una categoria di problemi di ottimizzazione, chiamata dei *problemi di livello*, molto studiata in tempi recenti. In questo tipo di problemi esistono più decisiori indipendenti. Questo si contrappone all'ottimizzazione classica, in cui un singolo decisore ha a disposizione informazioni sull'intero problema. Decisori differenti gestiscono aree diverse, ma le decisioni di ognuno influenzano indirettamente anche le aree di competenza altrui. I decisori possono essere allo stesso livello, distribuirsi gerarchicamente, oppure assumere una configurazione ibrida. Un esempio di problema di livello è la gestione del mercato libero dell'energia. Il gestore di rete sceglie giorno per giorno le offerte di produzione di varie aziende energetiche in base a una classifica di merito basata sul seguente criterio. I produttori dichiarano ogni giorno quanto saranno disposti a produrre il giorno successivo, e a che prezzo. Nonostante la grande quantità di variabili coinvolte, si tratta di un problema di programmazione lineare. Ogni giorno si osservano scostamenti, e risulta necessario introdurre piccole variazioni nei piani di produzione, sempre ottimizzandole matematicamente. Esistono *mercati di aggiustamento* dedicati al sopperimento di tali bisogni dinamici di energia. I modelli energetici devono essere evolutivi per poter raggiungere obiettivi futuri imposti a livello politico, come ad esempio il piano Energia e Clima dell'Unione Europea.

Ammissibilità

I vincoli lineari dei problemi LP sono rappresentabili come rette in un piano. La loro convergenza, quando esiste, va a generare una *regione di ammissibilità.* I vincoli di disuguaglianza limitano la soluzione ad un semipiano, mentre quelli di uguaglianza limitano la soluzione ai soli punti della retta stessa. L'aspetto della regione di ammissibilità è determinato dai vincoli di non negatività delle variabili. Può prendere la forma di un semipiano, di una retta, di un poligono, di un segmento o di un punto. Le intersezioni tra le frontiere dei vincoli sono chiamate *vertici.* I vertici della regione di ammissibilità sono *soluzioni di base.* Per determinare l'appartenenza di punti del piano alla regione di ammissibilità si valuta il segno della variabile scarto. Essa si annulla in corrispondenza delle soluzioni di base. È fondamentale individuare tali soluzioni perché esistono algoritmi in grado di determinare sequenze di soluzioni ammissibili a partire da una singola soluzione base nota.

$$x_1 + 2x_2 \leq 0$$
$$s = 10 - (x_1 + 2x_2)$$
$$x_1 + 2x_2 + s = 10 \quad \text{con} \quad s_1 \geq 0$$
$$\begin{cases} s > 0\text{: dentro la regione di ammissibilità} \\ s < 0\text{: fuori dalla regione di ammissibilità} \\ s = 0\text{: sulla frontiera} \end{cases}$$

Figura 4: Esempio di determinazione dei punti base

Possiamo dare una rappresentazione matriciale al sistema:

$$\begin{cases} Fw = b \\ w \geq 0 \end{cases}$$

dove:

- F è la matrice dei coefficienti
- w è il vettore delle incognite
- b è il vettore dei termini noti

Sappiamo che $F \in \mathbb{R}^{m,n+m}$ dove n è il numero di variabili decisionali (x) e m è il numero di vincoli. Il rango di F è $\text{rk}(F) = m$. Possiamo separare il il vettore w in due blocchi, uno di variabili decisionali e uno di variabili scarto:

$$w = \begin{pmatrix} x_1 \\ x_2 \\ s_1 \\ s_2 \end{pmatrix} \Rightarrow \begin{pmatrix} x \\ s \end{pmatrix}$$

Anche la matrice dei coefficienti si può spezzare in una matrice di coefficienti delle variabili decisionali, e una matrice identità di coefficienti delle variabili scarto:

$$F = [AI] \,.$$

Riesprimendo in questo modo la precedente formula, abbiamo:

$$[AI] \begin{bmatrix} x \\ s \end{bmatrix} = Ax + s = b.$$

Supponiamo, continuando il corrente esempio, di voler calcolare uno dei punti base. Questo punto è dato dall'intersezione tra le due rette base dei vincoli, con l'annullamento delle rispettive variabili scarto.

$$\begin{array}{cc} \begin{array}{cc} x_1 & x_2 \end{array} & \\ \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right] & \left(\begin{array}{c} x_1 \\ x_2 \\ \hline s_1 \\ s_2 \end{array}\right) \\ B \quad N & \end{array}$$

Annullando le due variabili scarto s_1 e s_2, la parte N della matrice diventa in pratica irrilevante e il calcolo si restringe a

$$\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \end{pmatrix}$$

calcolando solo il prodotto tra la base B e le variabili di base x_i da cui dobbiamo determinare w_B:

$$w_B = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

Invece w_N è

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Da questo possiamo dare un'ulteriore riscrittura a $Fw = B$:

$$[BN] \begin{bmatrix} w_B \\ w_N \end{bmatrix} = Bw_B + Nw_N = b$$

$$B^{-1}Bw_B + B^{-1}Nw_N = B^{-1}b \;\rightarrow\; \text{moltiplico tutto per l'inversa di } B$$

e quindi

$$\begin{cases} w_B = B^{-1}b - B^{-1}Nw_n \\ w_N = 0 \end{cases}$$

Quest'ultimo è l'insieme di tutte le soluzioni, anche quelle non ammissibili, del sistema sottodeterminato $Fw = b$ espresso rispetto alla base B.

Possiamo trovare altri punti base annullando altre variabili. Questo può essere ottenuto anche cambiando l'ordine delle colonne di F. Ad esempio, annullando x_2 e s_2, il sistema

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \end{pmatrix}$$

diventa (per scambio, riportando a sinistra le due colonne di base):

$$\begin{bmatrix} 1 & 1 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ s_1 \\ x_2 \\ s_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \end{pmatrix} \quad \Rightarrow \quad \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ s_1 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \end{pmatrix}$$

Trovo l'inversa di B:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \Rightarrow B^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

Per ottenere infine:

$$\overset{B^{-1}}{\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}} \begin{array}{c} B \qquad\quad N \qquad\quad b \\ \begin{bmatrix} 1 & 1 & 2 & 0 & | & 10 \\ 1 & 0 & 0 & 1 & | & 4 \end{bmatrix} \end{array} = \begin{array}{c} \begin{bmatrix} 1 & 0 & | & 0 & 1 & | & 4 \\ 0 & 1 & | & 2 & -1 & | & 5 \end{bmatrix} \\ B^{-1}B = I \quad B^{-1}N \quad B^{-1}B \end{array}$$

$$\text{quindi} \quad \begin{array}{ccccccc} \begin{pmatrix} x_1 \\ s_1 \end{pmatrix} & = & \begin{pmatrix} 6 \\ 4 \end{pmatrix} & - & \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix} & \begin{pmatrix} x_2 \\ s_2 \end{pmatrix} \\ w_B & = & B^{-1}b & - & B^{-1}N & w_N \end{array}$$

Terminiamo dunque la risoluzione numerica:

$$\begin{cases} x_1 = 4 - s_2 \\ s_1 = 6 - 2x_2 + s_2 \end{cases} \qquad \text{con } x_2 \text{ e } s_2 \text{ arbitrari in } \mathbb{R}^2.$$

La soluzione di base è

$$x_1 = 4 \quad x_2 = 0$$

$$s_1 = 6 \quad s_2 = 0.$$

Se trovo degli s negativi, non sono in zona ammissibile. Ricordiamo inoltre che la base BB per definizione deve essere singolare, cioè le sue colonne devono essere linearmente indipendenti.

La risoluzione dei problemi LP si divide dunque in due fasi, una chiamata *indagine di ammissibilità*, e l'altra coincidente con l'ottimizzazione vera e propria. L'indagine di ammissibilità consiste nel determinare se la regione di ammissibilità esista, trovando almeno un punto che ne faccia parte. I problemi i cui vincoli siano esclusivamente di minore o uguale relativamente a termini noti non necessitano di una vera indagine di ammissibilità, perché si può verificare immediatamente che l'origine del piano sia una soluzione ammissibile. Questo non vale invece per problemi nei quali coesistono vincoli di uguaglianza e disuguaglianza. In tal caso è necessario costruire un problema ausiliario, immediatamente ammissibile, la cui soluzione ottimale corrisponda ad un punto base della regione ammissibile relativa al problema originario. Il problema è irrisolvibile se la regione di ammissibilità è delimitata da un vincolo limitato da una variabile che necessiti di essere massimizzata nella funzione obiettivo. Esiste infine un caso limite di ammissibilità, in cui esiste una sola soluzione ammissibile puntuale. Tale soluzione è chiamata punto degenere e causa problematiche, seppur risolvibili, nella soluzione algebrica del problema.