Che cos'è un modello MIP

Indice

Dati, variabili, obiettivo, vincoli

Perché l'interezza conta

Il rilassamento LP e il sandwich dei bound

Tre pattern che tornano ovunque

Che cosa fa il solver con i due bound

Quello che questo capitolo ha messo in mano

Obiettivi (1/2)

Alla fine di questo capitolo sarai in grado di:

- scrivere un modello con dati, variabili, funzione obiettivo, vincoli e domini;
- distinguere le classi LP, ILP, BIP e MILP dai domini delle variabili;
- spiegare con un esempio perché arrotondare l'ottimo dell'LP non risolve un ILP;

Obiettivi (2/2)

Alla fine di questo capitolo sarai in grado di:

- dire da quale lato di $z(MILP)$ stanno una soluzione ammissibile e un rilassamento, e calcolare il gap;
- riconoscere i vincoli di set partitioning, set packing e set covering.

Prerequisiti

- Dal modulo *Programmazione lineare intera*, Capitolo 1 «La programmazione lineare intera»: i problemi di programmazione lineare intera, i loro punti interi ammissibili e il rilassamento continuo
- Dal modulo *Programmazione lineare*, Capitolo 4 «Problemi di programmazione lineare»: la forma di un problema di programmazione lineare: funzione obiettivo, vincoli lineari e domini delle variabili

Dati, variabili, obiettivo, vincoli

Introduzione: Dati, variabili, obiettivo, vincoli

- Vediamo di quali ingredienti è fatto un modello di **programmazione matematica**
- Distinguiamo le quattro classi di modelli: **LP**, **ILP**, **BIP** e **MILP**
- Fissiamo la notazione per valori ottimi, soluzioni e **bound**, che useremo in tutto il corso

I quattro ingredienti di un modello

Un modello traduce una decisione in quattro ingredienti, sempre nello stesso ordine:

1. i **dati**: i numeri che descrivono il problema, noti prima di decidere (costi, capacità, domande, tempi. . .)
2. le **variabili**: le quantità che il decisore controlla
3. l'**obiettivo**: una funzione delle variabili da minimizzare o massimizzare
4. i **vincoli**: equazioni e disequazioni che le variabili devono soddisfare perché la soluzione sia realizzabile

Un problema di *ottimizzazione* chiede la soluzione ammissibile (che soddisfa tutti i vincoli) di valore migliore per l'obiettivo.

Le classi di modelli

- **Le sigle**: LP dritto è la *classe* di problemi, LP in corsivo è *un* problema di quella classe, quello di cui si scrivono $\tilde{x}(LP)$ e $z(LP)$

  Lo stesso per ILP, BIP e MILP

- **La taglia e i dati**:
  - n variabili, con $j \in \{1, 2, \ldots, n\}$
  - m vincoli, con $i \in \{1, 2, \ldots, m\}$
  - i costi c_j, i coefficienti a_{ij} e i termini noti b_i

Che cosa può contenere un modello

- una **funzione obiettivo lineare** nelle variabili, da minimizzare o da massimizzare
- **vincoli lineari** di tre tipi:
  - di verso $\geq$, per una richiesta da soddisfare
  - di verso $\leq$, per una capacità da rispettare
  - di uguaglianza, per un bilancio o un'assegnazione esatta
- **variabili**, in una o più famiglie, ciascuna con il proprio dominio, dichiarato in fondo al modello

Nessuna delle tre cose ha una forma obbligata: lo decide il problema.

Che cosa distingue le quattro classi: i domini

- **LP**: tutte le variabili sono continue, $x_j \in \mathbb{R}$, di solito non negative

  In questo corso un LP arriva sempre come *rilassamento* di un modello MIP, togliendo i vincoli di interezza

- **ILP**: tutte le variabili sono intere, $x_j \in \mathbb{Z}$
- **BIP**: tutte le variabili sono binarie, $x_j \in \{0, 1\}$
- **MILP**: alcune variabili intere o binarie, altre continue

È lì che vive la differenza di difficoltà: un LP si risolve in tempo polinomiale, un ILP e un MILP in generale no.

Esempio: un modello di ILP

Quante unità comprare da due fornitori, x_1 e x_2, a lotti interi.

$$\begin{array}{rrcrcl} \min & 4x_1 & + & 7x_2 & & \\ \text{soggetto a} & 2x_1 & + & 3x_2 & \geq & 12 \\ & x_1 & + & x_2 & \leq & 5 \\ & x_1, & & x_2 & \in & \mathbb{Z}_{\geq 0} \end{array}$$

- L'obiettivo minimizza la spesa
- Il primo vincolo ($\geq$) impone di coprire la domanda di 12 unità
- Il secondo ($\leq$) limita a cinque i lotti totali
- Ultima riga: il dominio. Tutte le variabili intere: è un **ILP**

Esempio: un modello di BIP

Quali progetti finanziare: $y_k = 1$ se il progetto k si finanzia, 0 altrimenti.

$$
\begin{array}{rrcrcrcl}
\max & 5y_1 & + & 4y_2 & + & 6y_3 & & \\
\text{soggetto a} & y_1 & + & y_2 & + & y_3 & \leq & 2 \\
 & y_1 & & & + & y_3 & \geq & 1 \\
 & y_1, & & y_2, & & y_3 & \in & \{0,1\}
\end{array}
$$

- L'obiettivo massimizza il valore dei progetti scelti
- Primo vincolo ($\leq$): se ne finanziano al più due
- Secondo ($\geq$): almeno uno fra il primo e il terzo
- Ultima riga: tutte le variabili binarie, è un **BIP**

Esempio: un modello di MILP (1/2)

Due prodotti, con x_1 e x_2 le quantità prodotte, e un impianto da attivare o no: $y = 1$ se lo si attiva, $y = 0$ altrimenti.

$$
\begin{array}{rrcrcrcl}
\max & 3x_1 & + & 8x_2 & - & 10y & & \\
\text{soggetto a} & x_1 & + & x_2 & & & = & 6 \\
 & & & x_2 & - & 4y & \leq & 0 \\
 & x_1 & & & & & \geq & 1 \\
 & x_1, & & x_2 & & & \geq & 0 \\
 & & & & & y & \in & \{0,1\}
\end{array}
$$

Esempio: un modello di MILP (2/2)

- L'obiettivo massimizza il ricavo dei due prodotti meno il costo dell'impianto
- Primo vincolo ($=$): produrre esattamente le sei unità richieste
- Secondo ($\leq$): del secondo prodotto al più quattro unità, e soltanto a impianto attivo, perché con $y = 0$ resta $x_2 \leq 0$
- Terzo ($\geq$): almeno un'unità del primo prodotto
- Ultime due righe: due variabili continue e una binaria, è un **MILP**

Come è scritto ogni modello del corso

Ogni dato, ogni insieme e ogni simbolo è *definito prima di essere usato*. In ogni modello:

1. le variabili sono introdotte esplicitamente prima della formulazione
2. obiettivo e ciascuna famiglia di vincoli hanno un numero d'equazione
3. i vincoli di dominio, che dichiarano le variabili, chiudono il modello
4. dopo il modello un elenco spiega obiettivo e vincoli famiglia per famiglia

Notazione dei valori ottimi e dei bound (1/2)

- I **valori ottimi** indicano sempre il modello: $z(MILP)$ è l'ottimo di un modello di MILP; lo stesso per rilassamenti e duali
- Le soluzioni **ammissibili** portano la barra, quelle **ottime** la tilde:

$$(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n) \qquad \text{e} \qquad (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$$

- I **bound** si chiamano LB e UB:

$$LB \leq z(MILP) \leq UB$$

Notazione dei valori ottimi e dei bound (2/2)

- Quando serve dire da quale soluzione arriva un bound, lo si scrive fra parentesi: una soluzione ammissibile $(\bar{x}_1, \ldots, \bar{x}_n)$ di un modello di minimo dà

$$UB(\bar{x}) = \sum_{j=1}^{n} c_j \, \bar{x}_j$$

- In un modello di massimo la stessa somma è $LB(\bar{x})$
- In un minimo: soluzione ammissibile $\Rightarrow$ UB, rilassamento $\Rightarrow$ LB
- In un massimo è il contrario

Che cosa usa questo corso

Questo corso lavora quasi solo con **MILP**:

variabili binarie o intere per le decisioni «sì/no» o di conteggio,

variabili continue per quantità che si misurano (tempo, denaro, flusso).

Riepilogo

Ingredienti di un modello:

- Dati, variabili, obiettivo, vincoli; il dominio delle variabili chiude il modello
- Vincoli lineari di tipo $\geq$, $\leq$, $=$

Classi:

- LP (continue), ILP (intere), BIP (binarie), MILP (miste)
- LP polinomiale, ILP e MILP in generale no

Notazione:

- $\bar{x}$ ammissibile, $\tilde{x}$ ottima; $LB \leq z(MILP) \leq UB$

Perché l'interezza conta

Introduzione: Perché l'interezza conta

- Un ILP o un MILP **non** si risolve «arrotondando» la soluzione del suo rilassamento continuo
- Lo mostriamo con un controesempio minimo, con due sole variabili binarie
- Ne ricaveremo due lezioni distinte sul rilassamento e sull'arrotondamento

Esempio: l'arrotondamento fallisce (1/5)

$$\begin{array}{lrcrcl} \max & x_1 & + & x_2 & & \\ \text{soggetto a} & 2x_1 & + & 2x_2 & \leq & 3 \\ & x_1, & & x_2 & \in & \{0,1\} \end{array}$$

- Il rilassamento LP sostituisce il dominio binario con

$$0 \leq x_1 \leq 1, \qquad 0 \leq x_2 \leq 1$$

- e ha valore ottimo

$$z(LP^+) = \frac{3}{2}$$

Esempio: l'arrotondamento fallisce (2/5)

- Il valore 3/2 è raggiunto da *infinite* soluzioni ottime: tutti i punti del segmento

$$x_1 + x_2 = \frac{3}{2}, \qquad 0 \leq x_1 \leq 1, \qquad 0 \leq x_2 \leq 1$$

  fra cui (3/4, 3/4), (1, 1/2) e (1/2, 1)

- Quale di queste restituisca il solver dipende dall'algoritmo, non dal modello: nessuna è «la» soluzione del rilassamento

- Sulla nostra installazione Gurobi restituisce il vertice (1/2, 1)

Esempio: l'arrotondamento fallisce (3/5)

Arrotondando si ottengono esiti diversi a seconda del punto di partenza *e* del verso:

- da $(3/4, 3/4)$:
  - per eccesso $(1, 1)$: viola il vincolo, $2 + 2 = 4 > 3$
  - per difetto $(0, 0)$: ammissibile ma di valore 0
  - *nessuno* dei due versi trova l'ottimo
- da $(1, 1/2)$:
  - per eccesso ancora $(1, 1)$, non ammissibile
  - per difetto $(1, 0)$, ammissibile e di valore 1: è proprio l'ottimo intero $z(MILP) = 1$

Esempio: l'arrotondamento fallisce (4/5)

- Non c'è dunque un verso «giusto»: qui funziona il difetto, e solo a partire da uno dei due punti
- Nulla, prima di risolvere, diceva quale punto il solver avrebbe restituito

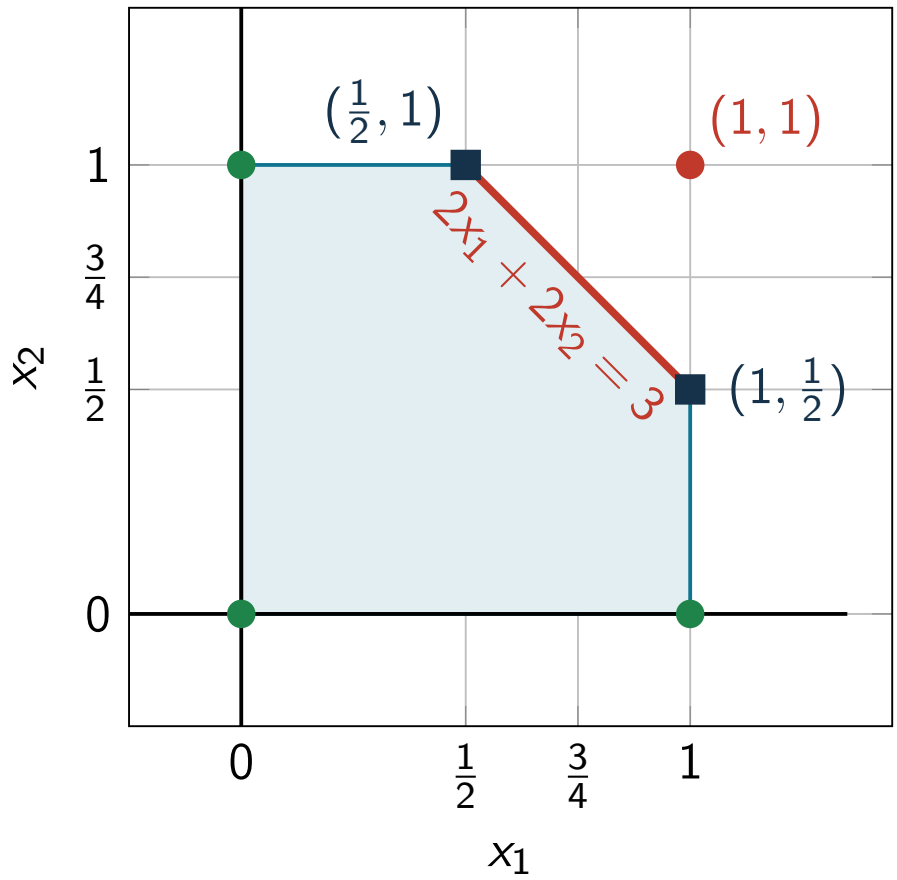
Le lezioni sono due, distinte:

- L'arrotondamento può produrre punti *non ammissibili*, e quando ne produce di ammissibili non c'è alcuna garanzia sul loro valore
- Il divario

$$\frac{3}{2} - 1 = \frac{1}{2}$$

non è colpa dell'arrotondamento: *nessun* punto intero ammissibile vale più di 1

Esempio: l'arrotondamento fallisce (5/5)



Azzurro: regione del rilassamento LP^+. Verde: punti interi ammissibili; rosso: $(1, 1)$ non ammissibile. Quadrati blu: due delle infinite soluzioni ottime del rilassamento, di valore 3/2.

Riepilogo

La lezione:

- Il rilassamento continuo dà un'indicazione (un **bound**), non una soluzione
- Arrotondare può dare punti non ammissibili, o ammissibili senza garanzie sul valore
- Il divario fra rilassamento e ottimo intero può esistere anche senza colpa dell'arrotondamento

Quindi:

- I MILP si risolvono con algoritmi dedicati, non con euristiche ingenue sul continuo

Il rilassamento LP e il sandwich dei bound

Introduzione: Il rilassamento LP e il sandwich dei bound

- Definiamo le due versioni del **rilassamento LP** usate nel corso
- Vediamo da che parte dell'ottimo intero sta ciascun **bound**, primale e duale
- Distinguiamo i tre significati della parola **gap**
- Capiamo a che cosa servono davvero i bound quando il solver non termina

Rilassamento LP: due versioni da non confondere (1/2)

Il **rilassamento LP** di un MILP si ottiene togliendo i vincoli di interezza e binarietà e tenendo tutto il resto. Due modi:

- **rilassamento con i bound**, $z(LP^+)$:
  - $x \in \{0,1\}$ diventa $0 \leq x \leq 1$, $x \in \mathbb{Z}_{\geq 0}$ diventa $x \geq 0$
  - è quello che risolve Gurobi con `relax()`, il primo che il solver risolve
- **rilassamento senza i bound**, $z(LP)$:
  - $x \in \{0,1\}$ diventa il solo $x \geq 0$
  - più debole (o uguale), ma ha un duale più compatto: è quello di cui si scrive il duale a mano

Rilassamento LP: due versioni da non confondere (2/2)

- Il rilassamento senza i bound è il più debole dei due, cioè il bound *più lontano* dall'ottimo intero
- I due coincidono quando gli altri vincoli implicano già $x \leq 1$
- Succede, per esempio, con un vincolo di assegnamento

$$\sum_m x_{jm} = 1$$

Il rilassamento è un bound: da che parte, e perché

- Il rilassamento *toglie* vincoli: ogni soluzione ammissibile del MILP resta ammissibile per il rilassamento, non viceversa
- Ottimizzare su un insieme più grande non può dare un valore peggiore:

$$\text{massimo:} \quad z(LP) \geq z(LP^+) \geq z(MILP)$$

$$\text{minimo:} \quad z(LP) \leq z(LP^+) \leq z(MILP)$$

- In un massimo è un *upper* bound, in un minimo un *lower* bound
- È sempre **ottimistico**: mai una soluzione, e il valore promesso può essere irraggiungibile

Da quale lato arriva ciascun bound (1/3)

Osservazione 1: Da quale lato arriva ciascun bound

- *I **bound primali** vengono dalle soluzioni intere ammissibili, costruite a mano o trovate dal solver: il lato **pessimistico***
- *I **bound duali** vengono dalla soluzione ottima del rilassamento LP o da una qualunque soluzione ammissibile del suo duale: il lato **ottimistico***

Da quale lato arriva ciascun bound (2/3)

- Il duale del rilassamento *non* dà un bound «dall'altro lato»
- Per **dualità debole**, in un minimo ogni soluzione duale ammissibile vale al più $z(LP)$, quindi al più $z(MILP)$
- Sta dalla *stessa* parte del rilassamento, in generale più debole
- Siano $\bar{x}$ ammissibile per il MILP, di valore $\sum_{j=1}^{n} c_j \bar{x}_j$, e $\bar{\pi}$ ammissibile per il duale del rilassamento, di valore $\sum_{i=1}^{m} b_i \bar{\pi}_i$

Da quale lato arriva ciascun bound (3/3)

Minimo: $\bar{x}$ dà $UB(\bar{x})$, $\bar{\pi}$ dà $LB(\bar{\pi})$

$$LB(\bar{\pi}) \leq z(D(LP)) = z(LP) \leq z(LP^+) \leq z(MILP) \leq UB(\bar{x})$$

Massimo: i ruoli si scambiano, $\bar{x}$ dà $LB(\bar{x})$, $\bar{\pi}$ dà $UB(\bar{\pi})$

$$LB(\bar{x}) \leq z(MILP) \leq z(LP^+) \leq z(LP) = z(D(LP)) \leq UB(\bar{\pi})$$

- L'uguaglianza in mezzo è la **dualità forte**: il duale del rilassamento vale quanto il rilassamento
- La disuguaglianza vera è fra il *valore di una soluzione duale ammissibile* e quel valore comune

Il sandwich dei bound

I due bound si incontrano nel *sandwich*

$$LB \leq z(MILP) \leq UB$$

che percorre tutto il corso.

Una soluzione ammissibile dà il lato pessimistico,
un rilassamento o una soluzione del suo duale
danno quello ottimistico.

Tre «gap» da non confondere (1/3)

La parola *gap* indica tre quantità diverse. Convenzioni:

- denominatore sempre $|z(MILP)|$ oppure $|UB|$, come indicato
- numeratore in valore assoluto
- se il denominatore è nullo, il gap relativo non si scrive (si riporta la differenza assoluta)

1. Gap dell'euristica. Se $\bar{x}$ è la soluzione dell'euristica, il suo valore è $UB(\bar{x})$ in un minimo ($LB(\bar{x})$ in un massimo), e quando l'ottimo è noto

$$\text{gap} = \frac{\left|UB(\bar{x}) - z(MILP)\right|}{|z(MILP)|}$$

Tre «gap» da non confondere (2/3)

- Il gap dell'euristica è quello delle tabelle dei bound: le istanze del corso si risolvono all'ottimo
- Quando l'ottimo *non* è noto lo si calcola rispetto a un bound duale, che prende il posto di $z(MILP)$: il numero è più grande, ed è l'unico che si possa dichiarare

2. Divario certificato fra due bound noti, senza conoscere $z(MILP)$:

$$\frac{UB - LB}{|UB|} \qquad \text{(minimo, } UB > 0\text{)}$$

- Garantisce che $z(MILP) \in [LB, UB]$, non che sia vicino a uno dei due estremi

Tre «gap» da non confondere (3/3)

3. `MIPGap` **del solver**:
   - la stessa idea, calcolata da Gurobi con $UB =$ `ObjVal` dell'incumbent e $LB =$ `ObjBound`
   - convenzione $|UB - LB|/|UB|$, con le sue tolleranze
   - vale 0 quando l'ottimalità è dimostrata

Le formule dei punti 2 e 3 richiedono che l'incumbent esista e che il denominatore non sia nullo: sono **convenzioni**, non identità matematiche, e vanno dichiarate ogni volta che si riporta un numero.

A che cosa servono davvero i due bound (1/2)

- Negli esercizi del corso le istanze sono piccole apposta: Gurobi trova $z(MILP)$ in una frazione di secondo
- LB e UB servono a *capire* da dove viene un bound e a verificare il ragionamento sull'euristica e sul duale
- Nella pratica un MILP reale può avere migliaia di variabili e il solver può non terminare in tempo utile
- Lo si interrompe: restano la migliore soluzione trovata e il miglior bound dimostrato al momento dello stop

A che cosa servono davvero i due bound (2/2)

- Il divario certificato fra i due è tutto quello che si può dire con sicurezza sulla distanza dall'ottimo, *senza* conoscere $z(MILP)$
- Saper costruire in fretta una soluzione ammissibile, e a mano un bound ottimistico, sono gli strumenti che restano quando l'ottimo esatto non è raggiungibile
- Sono anche quelli che permettono di giudicare i numeri che il solver riporta

Riepilogo (1/2)

Rilassamento LP:

- Con i bound ($z(LP^+)$) e senza i bound ($z(LP)$, più debole, duale più compatto)
- In un massimo è un upper bound, in un minimo un lower bound: sempre ottimistico

Bound:

- Primali da soluzioni ammissibili (pessimistici)
- Duali dal rilassamento o da soluzioni duali ammissibili (ottimistici)
- Sandwich: $LB \leq z(MILP) \leq UB$

Riepilogo (2/2)

Tre gap:

- Gap dell'euristica rispetto a $z(MILP)$ (o a un bound duale)
- Divario certificato $(UB - LB)/|UB|$
- `MIPGap` di Gurobi

In pratica:

- Quando il solver non termina, il divario certificato è tutto ciò che si può dichiarare

Tre pattern che tornano ovunque

Introduzione: Tre pattern che tornano ovunque

- Molti vincoli MILP con variabili binarie di selezione, indicizzate da un insieme I, appartengono a una di tre famiglie
- La famiglia dipende solo dal verso del vincolo
- Riconoscere il pattern dice subito come si comporta il vincolo

Partitioning, packing, covering

Nome	Vincolo
set partitioning	$\sum_{i \in I} x_i = 1$
set packing	$\sum_{i \in I} x_i \leq 1$
set covering	$\sum_{i \in I} x_i \geq 1$

- **partitioning**: esattamente un elemento di I è scelto
- **packing**: al più un elemento di I è scelto
- **covering**: almeno un elemento di I è scelto

Riconoscere il pattern

- «Ogni lavoro a esattamente una macchina» è un **set partitioning**
- «Al più un lavoro per fascia oraria» è un **set packing**
- «Ogni zona coperta da almeno un presidio» è un **set covering**, traduzione diretta di una clausola OR
- Un *covering* ammette che più elementi siano scelti insieme, un *packing* no, un *partitioning* impone l'uno e l'altro insieme

Riepilogo

Tre pattern:

- $\sum_{i \in I} x_i = 1$: set partitioning (esattamente uno)
- $\sum_{i \in I} x_i \leq 1$: set packing (al più uno)
- $\sum_{i \in I} x_i \geq 1$: set covering (almeno uno, clausola OR)

Perché servono:

- Il verso del vincolo dice subito quanti elementi possono essere scelti insieme

Che cosa fa il solver con i due bound

Introduzione: Che cosa fa il solver con i due bound

- Un MILP con insieme ammissibile limitato si risolve con algoritmi dedicati
- Come funzionano (esplorazione, tagli, ramificazione) è materia di un corso a parte
- Qui interessa una cosa sola: i due bound che costruiamo a mano sono i due lati dello stesso sandwich che il solver cerca di chiudere
- Non glieli si passa: il solver se li costruisce da sé, con le sue euristiche e i suoi rilassamenti

Il ruolo dei due bound nel solver

- Il **bound primale** è una soluzione ammissibile, e dà un valore da battere: ogni parte dello spazio che non può fare meglio viene scartata senza esplorarla
- Il **bound duale** viene dal rilassamento, e dice quanto si può sperare al massimo: più è vicino all'ottimo intero, cioè più la formulazione è **stretta**, meno lavoro resta da fare
- Il divario fra i due è il `MIPGap`, ed è l'unica cosa che si può affermare con sicurezza quando l'ottimo non si raggiunge

Riepilogo

Il solver e i bound:

- Il bound primale permette di scartare parti dello spazio
- Il bound duale misura quanto si può ancora sperare
- Il `MIPGap` è il divario fra i due

Che cosa serve di questo corso:

- Scrivere il modello con una formulazione stretta
- Saper produrre a mano i due bound

Quello che questo capitolo ha messo in mano

Introduzione: Quello che questo capitolo ha messo in mano

- Rimettiamo in fila gli oggetti con cui si scrive ogni modello MIP
- e i concetti di soluzione, rilassamento, bound e gap che useremo nel resto del corso

Riepilogo finale (1/2)

Un modello MIP si scrive sempre con gli stessi cinque oggetti:

- i **dati**, noti e non decisi
- le **variabili**, una per decisione, con il loro dominio
- la **funzione obiettivo**, che dice che cosa si massimizza o si minimizza
- i **vincoli**, uno per ogni frase dell'enunciato
- il **dominio** delle variabili, un vincolo anche lui, il più facile da dimenticare

Riepilogo finale (2/2)

- **Soluzione ammissibile**: rispetta tutti i vincoli; **soluzione ottima**: la migliore fra le ammissibili
- Rilassare l'interezza dà il **rilassamento lineare**, più facile, il cui ottimo è un **bound**
- Bound pessimistico da una soluzione ammissibile, ottimistico da un rilassamento
- La distanza fra i due è il **gap**: il resto del corso serve a costruire i due lati e a stringerli