



SAPIENZA
UNIVERSITÀ DI ROMA

Programmazione lineare intera

Volume 1 — Modelli, bound e formulazioni

Che cos'è un problema di programmazione lineare intera (MIP, ILP, BIP), le sue soluzioni e i suoi esiti; bound primali e duali e gap; il rilassamento lineare; formulazioni ideali, disuguaglianze valide e disuguaglianze di cover. Esempi svolti ed esercizi con la soluzione completa.

Fabio Furini

fabiofurini.github.io

Dipartimento di Ingegneria informatica, automatica e gestionale — Sapienza Università di Roma

Versione online e materiale scaricabile:

fabiofurini.github.io/programmazione-lineare-intera/
github.com/fabiofurini/programmazione-lineare-intera

Anno accademico 2026–2027

Indice

I	Modelli e bound	3
1	La programmazione lineare intera	4
1.1	Forma canonica di massimizzazione	5
1.2	Esiti possibili	9
1.3	Enumerazione completa	10
1.4	Errori tipici	11
1.5	In sintesi	12
2	Bound primali e duali	13
2.1	Bound primali e bound duali	14
2.2	Bound primali: soluzioni euristiche	15
2.3	Il rilassamento lineare	16
2.4	Arrotondare la soluzione del rilassamento lineare	18
2.5	Upper bound sulle variabili	20
2.6	Bound duali	22
2.7	Errori tipici	25
2.8	In sintesi	26
II	Rafforzare la formulazione	27
3	Formulazioni ideali	28
3.1	L'involuppo convesso	29
3.2	Formulazioni e formulazioni ideali	30
3.3	Confronto tra formulazioni	32
3.4	Esistenza delle formulazioni ideali	33
3.5	Errori tipici	33
3.6	In sintesi	33
4	Disuguaglianze valide	35
4.1	Disuguaglianze valide	36
4.2	Ricavare disuguaglianze valide per arrotondamento	37

4.3	Migliorare una disuguaglianza valida	38
4.4	Famiglie di disuguaglianze valide e tagli	40
4.5	Errori tipici	41
4.6	In sintesi	41
5	Disuguaglianze di cover	43
5.1	Un vincolo con coefficienti positivi e il suo rilassamento lineare	44
5.2	Coperture e disuguaglianze di copertura	45
5.3	Interpretazione geometrica	48
5.4	Separazione delle disuguaglianze di copertura	54
5.5	Dal CSP a un problema complementare	58
5.6	Disuguaglianze di copertura estese	60
5.7	Violazione e distanza da una disuguaglianza di copertura	61
5.8	Errori tipici	63
5.9	In sintesi	64

Parte I

Modelli e bound



La programmazione lineare intera

Prerequisiti

- Dal modulo *Programmazione Lineare*: [Capitolo 2 «Semispazi e poliedri»](#) e [Capitolo 4 «Problemi di programmazione lineare»](#) (la forma canonica di massimizzazione, la regione ammissibile, le soluzioni ottime e il valore ottimo), e [Capitolo 5 «Esiti possibili e formulazioni equivalenti»](#) (problemi inammissibili, illimitati e con soluzione ottima).

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- scrivere un problema PLI in forma canonica di massimizzazione e distinguere i problemi PLI, BIP e MIP;
- disegnare la regione ammissibile di un piccolo problema PLI e trovare graficamente una soluzione ottima;
- spiegare perché la regione ammissibile non è convessa e quali sono i tre esiti possibili di un problema PLI;
- spiegare perché l'enumerazione completa fallisce e quanto costa;
- spiegare perché servono bound sul valore ottimo per evitare l'enumerazione.

1.1 Forma canonica di massimizzazione

Definizione 1.1: Problemi di Programmazione Lineare Intera in forma canonica di massimizzazione

Dati:

1. un vettore dei costi $\mathbf{c} \in \mathbb{Z}^n$
2. una matrice dei vincoli $\mathbf{A} \in \mathbb{Z}^{m \times n}$
3. un vettore dei termini noti (RHS) $\mathbf{b} \in \mathbb{Z}^m$ con m righe

$$\underbrace{\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}}_{\mathbf{c}} \in \mathbb{Z}^n \quad \underbrace{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}}_{\mathbf{A}} \in \mathbb{Z}^{m \times n} \quad \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{\mathbf{b}} \in \mathbb{Z}^m$$

un **problema di Programmazione Lineare Intera (PLI)** in **forma canonica di massimizzazione** con n variabili intere non negative

$$\underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} \in \mathbb{Z}_{\geq 0}^n$$

e m vincoli lineari di minore o uguale è il seguente problema di ottimizzazione:

$$\max \underbrace{\sum_{j=1}^n c_j x_j}_{\mathbf{c}'\mathbf{x}} \quad (1.1a)$$

$$\text{soggetto a } \underbrace{\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad \forall i \in \{1, 2, \dots, m\}}_{\mathbf{A}\mathbf{x} \leq \mathbf{b}}, \quad (1.1b)$$

$$\underbrace{x_j \in \mathbb{Z}_{\geq 0}, \quad \forall j \in \{1, 2, \dots, n\}}_{\mathbf{x} \in \mathbb{Z}_{\geq 0}^n}. \quad (1.1c)$$

- I dati $\mathbf{c}$, $\mathbf{A}$ e $\mathbf{b}$ sono **interi**, come in tutto il modulo.
- Con un leggero abuso di notazione, scriviamo il problema PLI (1.1) in forma compatta come

$$\max \{ \mathbf{c}'\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \in \mathbb{Z}_{\geq 0}^n \}.$$

- L'insieme $\mathbb{Z}^n$ dei punti di $\mathbb{R}^n$ con componenti intere si chiama **reticolo intero**.

Un problema PLI (1.1) è un problema LP in forma canonica di massimizzazione con il requisito aggiuntivo $\mathbf{x} \in \mathbb{Z}^n$: la funzione lineare $\mathbf{c}'\mathbf{x}$ viene massimizzata sui punti del **reticolo intero** che soddisfano i vincoli.

- Se tutte le variabili sono ristrette a $\{0, 1\}$, il problema è un problema di **programmazione lineare intera binaria** (BIP). Se solo alcune delle variabili sono intere, il problema è un problema di **programmazione lineare intera mista** (MIP).
- I vincoli $\geq$ e $=$, e i problemi di minimizzazione, si riportano alla forma canonica di massimizzazione come nella programmazione lineare.

Esempio 1.1: Problema PLI in forma canonica di massimizzazione

Consideriamo i seguenti dati di un problema PLI con $n = 2$ e $m = 2$:

$$\mathbf{c} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} 10 & 4 \\ 4 & 14 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 45 \\ 49 \end{pmatrix}$$

Le componenti numeriche dei due vettori colonna e della matrice sono

$$c_1 = 3, c_2 = 2, a_{11} = 10, a_{12} = 4, a_{21} = 4, a_{22} = 14, b_1 = 45, b_2 = 49.$$

Il problema PLI associato in forma canonica di massimizzazione, con $n = 2$ variabili intere non negative e $m = 2$ vincoli lineari di minore o uguale, è il seguente:

$$\begin{array}{llllll} \max & 3x_1 & + & 2x_2 & & \\ \text{soggetto a} & 10x_1 & + & 4x_2 & \leq & 45 \\ & 4x_1 & + & 14x_2 & \leq & 49 \\ & x_1, & & x_2 & \in & \mathbb{Z}_{\geq 0} \end{array}$$

Le variabili x_1 e x_2 si possono leggere come il numero di auto di due tipi da produrre: questo problema PLI è l'**esempio dell'auto**, l'esempio guida del modulo.

1.1.1 Regione ammissibile e funzione obiettivo

Definizione 1.2: regione ammissibile e funzione obiettivo

- La **regione ammissibile** di un problema PLI (1.1) è l'insieme F così definito:

$$F = \left\{ \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} \in \mathbb{Z}^n : \underbrace{\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad \forall i \in \{1, 2, \dots, m\}}_{\mathbf{A} \mathbf{x} \leq \mathbf{b}}, \right. \quad (1.2)$$

$$\left. \underbrace{x_j \geq 0, \quad \forall j \in \{1, 2, \dots, n\}}_{\mathbf{x} \geq \mathbf{0}} \right\}.$$

- La **funzione obiettivo** (1.1a) di un problema PLI (1.1) da massimizzare è la **funzione lineare** (scalare) f così definita:

$$f : \mathbb{Z}^n \rightarrow \mathbb{R}, \quad f : \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} \mapsto \sum_{j=1}^n c_j x_j \quad (1.3)$$

$$\underbrace{\hspace{1.5cm}}_{\mathbf{c}'\mathbf{x}}$$

- Sia $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A} \mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ il poliedro dei vincoli, cioè la regione ammissibile del problema LP con gli stessi dati. La regione ammissibile del problema PLI è

$$F = P \cap \mathbb{Z}^n.$$

- L'insieme F è fatto di punti isolati del reticolo, quindi **non è convesso**.

Esempio 1.2: regione ammissibile

Consideriamo il problema PLI dell'Esempio 1.1. La regione ammissibile F è l'insieme dei 16 punti neri della figura seguente, i punti del reticolo $\mathbb{Z}^2$ nel poliedro P (rosso).

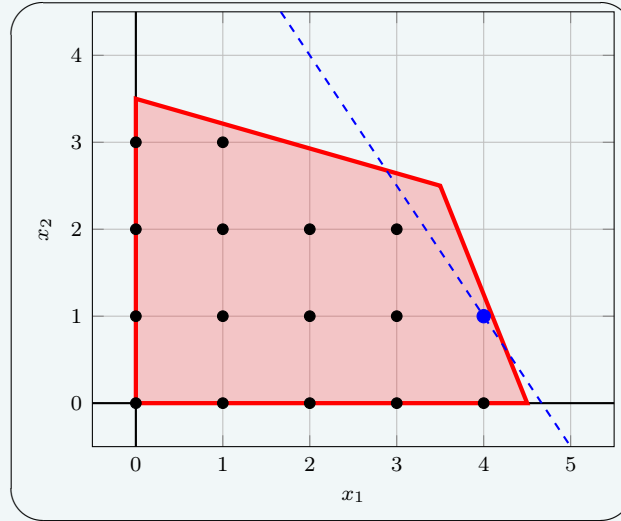


Figura 1.1: L'esempio dell'auto: la regione ammissibile del problema PLI è l'insieme dei suoi 16 punti interi (neri) dentro il poliedro (rosso). La soluzione ottima $(4,1)$ è in blu, con la retta obiettivo $3x_1 + 2x_2 = 14$ (tratteggiata).

Il punto $\bar{x} = (2,2)$ appartiene a F , perché $10 \cdot 2 + 4 \cdot 2 = 28 \leq 45$ e $4 \cdot 2 + 14 \cdot 2 = 36 \leq 49$. Il punto $(4,2)$ non appartiene a F , perché $10 \cdot 4 + 4 \cdot 2 = 48 > 45$. Il punto $(\frac{7}{2}, \frac{5}{2})$ appartiene a P ma non a F , perché non è intero.

1.1.2 Soluzione ottima e valore ottimo

Definizione 1.3: soluzione ottima e valore ottimo

- Un punto $\tilde{x}(\text{ILP}) \in F$ è una **soluzione ottima** del problema PLI (1.1) se

$$\tilde{x}(\text{ILP}) \in \arg \max \{ c'x : x \in F \}.$$

- Il valore

$$z(\text{ILP}) = \underbrace{\max \{ c'x : x \in F \}}_{=c'\tilde{x}(\text{ILP})}$$

è il **valore ottimo** del problema PLI (1.1).

- Un punto $\bar{x} \in F$ è una **soluzione ammissibile** del problema PLI. Una soluzione ottima è una soluzione ammissibile con il valore più alto della funzione obiettivo.

L'obiettivo di un problema PLI (1.1) è **trovare una soluzione ottima** $\tilde{x}(\text{ILP})$, quando esiste, e non soltanto determinare il valore ottimo $z(\text{ILP})$.

1.1.3 Metodo grafico

Una soluzione ottima di un problema PLI (1.1) con due variabili si può determinare graficamente spostando la linea di livello della funzione obiettivo nella direzione del suo gradiente, fino a raggiungere l'ultimo punto del reticolo nel poliedro P .

Esempio 1.3: Metodo grafico

Consideriamo il problema PLI dell'Esempio 1.1. Il gradiente della funzione obiettivo è $\nabla f = (3, 2)$. Spostando la linea di livello $3x_1 + 2x_2 = z$ nella direzione di ∇f , l'ultimo punto di F raggiunto è $(4, 1)$, sulla linea tratteggiata $3x_1 + 2x_2 = 14$ della figura dell'Esempio 1.2. Nessun altro punto di F giace su questa linea. L'esempio dell'auto ha l'**unica** soluzione ottima $\tilde{x}(\text{ILP}) = (4, 1)$, con valore ottimo $z(\text{ILP}) = 14$.

1.2 Esiti possibili

- Come nella programmazione lineare, un problema PLI ha esattamente uno di tre **esiti**: è **inammissibile** ($F = \emptyset$), **illimitato** (la funzione obiettivo non è limitata superiormente su F) oppure ha una **soluzione ottima**.
- Se il poliedro P è limitato, allora F è un **insieme finito**, e una soluzione ottima esiste non appena $F \neq \emptyset$.

Esempio 1.4: un problema PLI inammissibile con un problema LP ammissibile

Il problema PLI

$$\max \{ x_1 : 2x_1 \leq 1, -2x_1 \leq -1, x_1 \in \mathbb{Z}_{\geq 0} \}$$

è inammissibile, perché i vincoli richiedono $x_1 = \frac{1}{2}$. Il problema LP con gli stessi dati è ammissibile, con l'unica soluzione ammissibile $x_1 = \frac{1}{2}$.

Esempio 1.5: un problema PLI illimitato

Il problema PLI

$$\max \{ x_1 + x_2 : x_1 - x_2 \leq 1, x_1, x_2 \in \mathbb{Z}_{\geq 0} \}$$

è illimitato: i punti (k, k) , $k \in \mathbb{Z}_{\geq 0}$, sono ammissibili, con valore $2k$.

1.3 Enumerazione completa

- Quando la regione ammissibile è un insieme finito, possiamo trovare una soluzione ottima per **enumerazione completa**: calcoliamo la funzione obiettivo per ogni soluzione ammissibile e teniamo la migliore. Per l'esempio dell'auto ci sono 16 punti:

x_1	x_2	z	x_1	x_2	z
0	0	0	3	1	11
1	0	3	4	1	14
2	0	6	0	2	4
3	0	9	1	2	7
4	0	12	2	2	10
0	1	2	3	2	13
1	1	5	0	3	6
2	1	8	1	3	9

- Il valore migliore è 14, in (4, 1) (in grassetto). Qui il metodo funziona, perché il modello è minuscolo.
- Consideriamo ora il caso più semplice: n **variabili binarie** e nessun altro vincolo. Ogni variabile assume 2 valori, quindi ci sono 2^n soluzioni: i vertici di un ipercubo n -dimensionale.

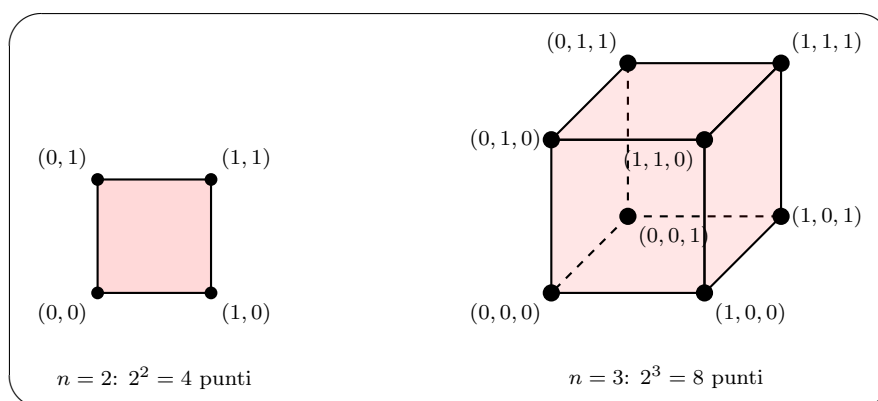


Figura 1.2: Le soluzioni di n variabili binarie senza altri vincoli: i 2^n vertici di un ipercubo n -dimensionale, per $n = 2$ (un quadrato) e $n = 3$ (un cubo).

- Il numero 2^n esplode. Con $n = 80$ variabili ci sono circa $1.2 \cdot 10^{24}$ soluzioni, un numero paragonabile a quello stimato delle stelle dell'universo, e con $n = 266$ variabili circa $1.2 \cdot 10^{80}$, paragonabile al numero stimato di atomi dell'universo osservabile.
- Supponiamo che un computer esamini 10^9 soluzioni al secondo, e poi un computer 1000 volte più veloce, 10^{12} soluzioni al secondo. Il tempo necessario per enumerare 2^n soluzioni è:

n	2^n	a 10^9 soluzioni/s	a 10^{12} soluzioni/s
20	$\approx 1.05 \cdot 10^6$	≈ 1.05 ms	≈ 1.05 μ s
30	$\approx 1.07 \cdot 10^9$	≈ 1.07 s	≈ 1.07 ms
40	$\approx 1.10 \cdot 10^{12}$	≈ 18.3 min	≈ 1.10 s
50	$\approx 1.13 \cdot 10^{15}$	≈ 13.0 days	≈ 18.8 min
60	$\approx 1.15 \cdot 10^{18}$	≈ 36.5 years	≈ 13.3 days
70	$\approx 1.18 \cdot 10^{21}$	≈ 374 centuries	≈ 37.4 years
80	$\approx 1.21 \cdot 10^{24}$	$\approx 3.8 \cdot 10^5$ centuries	≈ 383 centuries

L'enumerazione completa è impraticabile già per modelli di dimensione moderata. Un computer 1000 volte più veloce guadagna solo circa 10 variabili, perché $2^{10} = 1024 \approx 1000$.

- La via d'uscita è **evitare di enumerare**. Se conosciamo un lower bound e un upper bound sul valore ottimo, possiamo escludere intere regioni di soluzioni ammissibili senza visitarle. I due bound vengono da due fonti diverse, le soluzioni ammissibili e i rilassamenti: sono l'argomento del capitolo successivo, «Bound primali e bound duali».

1.4 Errori tipici

Attenzione

- **Credere che la regione ammissibile sia convessa.** È l'insieme dei punti interi di un poliedro, non il poliedro. Il punto medio di due soluzioni ammissibili può essere inammissibile.
- **Credere che la regione ammissibile sia sempre finita.** È finita solo se il poliedro è limitato; altrimenti il problema PLI può essere illimitato.
- **Credere che un problema LP ammissibile dia un problema PLI ammissibile.** Il problema LP può avere soluzioni mentre nessun punto intero soddisfa i vincoli.
- **Prendere una buona soluzione ammissibile per ottima.** Senza un bound dall'altro lato (capitolo successivo) non c'è alcun certificato di ottimalità.

1.5 In sintesi

In sintesi

- Un problema PLI massimizza una funzione lineare sui punti interi di un poliedro; la sua regione ammissibile non è convessa, ed è finita quando il poliedro è limitato.
- Gli esiti sono inammissibile, illimitato oppure con soluzione ottima, e un problema LP ammissibile può avere una versione PLI inammissibile.
- L'enumerazione completa costa 2^n valutazioni per n variabili binarie: non è praticabile oltre i modelli molto piccoli.
- Per evitare l'enumerazione servono bound su $z(\text{ILP})$ da entrambi i lati: soluzioni ammissibili e rilassamenti, l'argomento del capitolo successivo.

Cruscotto dell'esempio dell'auto. Ancora nessun bound: il capitolo successivo fornisce i primi lower bound e upper bound.

Bound primali e duali

Prerequisiti

- Da questo modulo, il capitolo «Programmazione lineare intera»: problemi PLI, soluzioni ammissibili e ottime, enumerazione completa e l'esempio dell'auto.
- Dal modulo *Programmazione Lineare*, [Capitolo 9 «Teoria della dualità»](#): il problema duale, la dualità debole e la dualità forte.
- Dallo stesso modulo, [Capitolo 6 «Algoritmo del simplesso»](#): un problema LP che ha una soluzione ottima ha un vertice ottimo.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- ottenere un bound primale (un lower bound) sul valore ottimo da una soluzione ammissibile, anche con un semplice algoritmo euristico;
- costruire il rilassamento lineare di un problema PLI e usarne il valore ottimo come bound duale (un upper bound), arrotondato per difetto quando è lecito;
- calcolare il gap di ottimalità da un bound primale e un bound duale;
- spiegare perché arrotondare una soluzione ottima del rilassamento lineare dà, nel migliore dei casi, una soluzione ammissibile;
- ricavare dai vincoli upper bound sulle variabili, e dire quando servono;
- ottenere un bound duale da moltiplicatori non negativi dei vincoli, e riconoscere nel problema duale il problema dei moltiplicatori migliori.

Dove eravamo

Nel capitolo precedente abbiamo visto che l'enumerazione completa non è praticabile e che servono bound sul valore ottimo da entrambi i lati. Ora li otteniamo: **bound primali** dalle soluzioni ammissibili e **bound duali** dai rilassamenti.

2.1 Bound primali e bound duali

Definizione 2.1: bound primale, bound duale e gap di ottimalità

Consideriamo un problema PLI in forma canonica di massimizzazione, con regione ammissibile F e valore ottimo $z(\text{ILP})$.

- Data una soluzione ammissibile $\bar{x} \in F$, il suo valore $\text{LB}(\bar{x}) = \sum_{j=1}^n c_j \bar{x}_j$ è un **bound primale** su $z(\text{ILP})$.
- Un valore UB tale che $\sum_{j=1}^n c_j x_j \leq \text{UB}$ per ogni $x \in F$ è un **bound duale** su $z(\text{ILP})$.
- Dati un bound primale $\text{LB} \neq 0$ e un bound duale UB , il **gap di ottimalità** è

$$\text{gap di ottimalità} = \frac{\text{UB} - \text{LB}}{|\text{LB}|}. \quad (2.1)$$

- In un problema di massimizzazione i bound primali sono **lower bound** e i bound duali sono **upper bound**. Un bound primale viene da una sola soluzione ammissibile; un bound duale è un'affermazione su tutte le soluzioni ammissibili insieme.

Per ogni soluzione ammissibile $\bar{x}$ e ogni bound duale UB :

$$\text{LB}(\bar{x}) \leq z(\text{ILP}) \leq \text{UB}.$$

Una soluzione ammissibile è **certificata ottima** quando il gap di ottimalità è zero.

Definizione 2.2: rilassamento

Dato un problema PLI $\max\{c'x : x \in F\}$, il problema $\max\{c'x : x \in R\}$ è un **rilassamento** del problema PLI se $F \subseteq R$. Indichiamo il suo valore ottimo con $z(R)$.

- Massimizzare su una regione più grande non può dare un valore più piccolo: il valore ottimo di un rilassamento è un bound duale, $z(\text{ILP}) \leq z(R)$.

2.2 Bound primali: soluzioni euristiche

Definizione 2.3: algoritmo euristico

Un **algoritmo euristico** è un metodo che trova rapidamente soluzioni ammissibili, senza garanzia di ottimalità. Il valore di ogni soluzione che trova è un bound primale.

Una buona soluzione ammissibile è un buon bound primale.

Esempio 2.1: bound primale

Per la soluzione ammissibile $(2, 2)$ dell'esempio dell'auto otteniamo $LB = 3 \cdot 2 + 2 \cdot 2 = 10$, quindi $z(ILP) \geq 10$.

Definizione 2.4: parte intera di un numero

Dato $y \in \mathbb{R}$, la **parte intera** di y è $\lfloor y \rfloor = \max\{q \in \mathbb{Z} : q \leq y\}$.

Esempio 2.2: due algoritmi euristici per l'esempio dell'auto

- L'**euristica a una variabile** porta una sola variabile al suo valore massimo, con l'altra uguale a 0, e tiene la migliore delle due soluzioni.

1. Solo x_1 ($x_2 = 0$): $\bar{x}_1 = \min\{\lfloor \frac{45}{10} \rfloor, \lfloor \frac{49}{4} \rfloor\} = \min\{4, 12\} = 4$, quindi $\bar{x} = (4, 0)$ e $LB = 12$.
2. Solo x_2 ($x_1 = 0$): $\bar{x}_2 = \min\{\lfloor \frac{45}{4} \rfloor, \lfloor \frac{49}{14} \rfloor\} = \min\{11, 3\} = 3$, quindi $\bar{x} = (0, 3)$ e $LB = 6$.

- L'**euristica a due fasi** porta prima una variabile al suo valore massimo, e poi l'altra al suo valore massimo con i **termini noti residui**. Tiene la migliore delle due soluzioni.

1. Prima x_1 : $\bar{x}_1 = 4$ usa 40 e 16, e lascia 5 e 33. Poi $\bar{x}_2 = \min\{\lfloor \frac{5}{4} \rfloor, \lfloor \frac{33}{14} \rfloor\} = \min\{1, 2\} = 1$, quindi $\bar{x} = (4, 1)$ e $LB = 14$.
2. Prima x_2 : $\bar{x}_2 = 3$ usa 12 e 42, e lascia 33 e 7. Poi $\bar{x}_1 = \min\{\lfloor \frac{33}{10} \rfloor, \lfloor \frac{7}{4} \rfloor\} = \min\{3, 1\} = 1$, quindi $\bar{x} = (1, 3)$ e $LB = 9$.

Euristica	Ordine	$\bar{x}_1$	$\bar{x}_2$	LB
a una variabile	x_1	4	0	12
a una variabile	x_2	0	3	6
a due fasi	prima x_1	4	1	14
a due fasi	prima x_2	1	3	9

- L'euristica a due fasi dà un bound primale migliore, ma richiede più calcoli. Questo compromesso tra la qualità della soluzione e lo sforzo di calcolo è tipico degli algoritmi euristici.

- Qui il bound primale migliore è $LB = 14$, in $(4, 1)$. Un bound primale da solo non certifica l'ottimalità: serve un bound duale.

2.3 Il rilassamento lineare

Definizione 2.5: rilassamento lineare

Il **rilassamento lineare** (rilassamento lineare) di un problema PLI in forma canonica di massimizzazione è il problema LP che si ottiene sostituendo i vincoli di integralità $x_j \in \mathbb{Z}_{\geq 0}$ con i vincoli di non negatività $x_j \geq 0$, per ogni $j \in \{1, 2, \dots, n\}$:

$$\max \sum_{j=1}^n c_j x_j \quad (2.2a)$$

$$\text{soggetto a } \sum_{j=1}^n a_{ij} x_j \leq b_i, \quad \forall i \in \{1, 2, \dots, m\}, \quad (2.2b)$$

$$x_j \geq 0, \quad \forall j \in \{1, 2, \dots, n\}. \quad (2.2c)$$

- Indichiamo con $z(\text{LP})$ il valore ottimo del rilassamento lineare, e con $\tilde{x}(\text{LP})$ una sua soluzione ottima. Scriviamo sempre l'argomento, perché nello stesso esempio $\tilde{x}(\text{ILP})$ e $\tilde{x}(\text{LP})$ sono punti diversi.

Proposizione 2.6: upper bound dal rilassamento lineare

Ogni soluzione ammissibile del problema PLI è ammissibile per il suo rilassamento lineare, quindi

$$z(\text{ILP}) \leq z(\text{LP}). \quad (2.3)$$

Inoltre, se una soluzione ottima del rilassamento lineare è intera, allora è una soluzione ottima del problema PLI.

Dimostrazione. La regione ammissibile F del problema PLI è contenuta nel poliedro del rilassamento lineare, e la funzione obiettivo è la stessa. Il massimo su un insieme più grande non è minore del massimo su un suo sottoinsieme. Se $\tilde{x}(\text{LP})$ è intera, appartiene a F e ha valore $z(\text{LP}) \geq z(\text{ILP})$: è ottima per il problema PLI. $\square$

- Il valore di una soluzione ammissibile può essere un intero, ed è al più $z(\text{LP})$: allora è al più il più grande intero che non supera $z(\text{LP})$.

Osservazione 2.7: arrotondare l'upper bound

Se tutte le variabili sono intere e tutti i costi c_j sono interi, allora

$$z(\text{ILP}) \leq \lfloor z(\text{LP}) \rfloor. \quad (2.4)$$

In tutti gli altri casi il bound $\lfloor z(\text{LP}) \rfloor$ non è valido.

- In questo modulo tutti i problemi PLI hanno dati interi, quindi il bound (2.4) vale sempre. In un problema MIP con una variabile continua il valore di una soluzione ammissibile non è necessariamente intero, e il bound non si può arrotondare.

Esempio 2.3: il rilassamento lineare dell'esempio dell'auto

- Il rilassamento lineare dell'esempio dell'auto ha la soluzione ottima

$$\tilde{x}_1 = \frac{7}{2}, \quad \tilde{x}_2 = \frac{5}{2}, \quad z(\text{LP}) = \frac{31}{2},$$

che è frazionaria: non è ammissibile per il problema PLI.

- Poiché i costi sono interi, abbiamo l'upper bound $\lfloor z(\text{LP}) \rfloor = 15$. Insieme al bound primale 14 della soluzione ammissibile (4, 1) dell'Esempio 2.2:

$$14 \leq z(\text{ILP}) \leq 15.$$

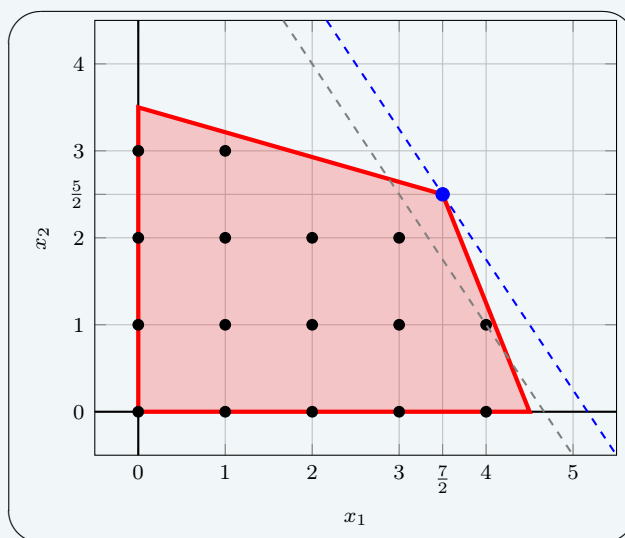


Figura 2.1: Il rilassamento lineare dell'esempio dell'auto: la regione ammissibile (rossa), i suoi 16 punti interi (neri) e la soluzione ottima $(\frac{7}{2}, \frac{5}{2})$ del rilassamento lineare (blu), con la retta obiettivo $3x_1 + 2x_2 = \frac{31}{2}$ che vi passa (tratteggiata). La linea grigia tratteggiata è la retta obiettivo della soluzione ottima del problema PLI.

- Il valore ottimo è 14 o 15. Con $\text{UB} = 15$ e $\text{LB} = 14$ il gap di ottimalità (2.1) è $\frac{1}{14} \approx 7.1\%$: la soluzione (4, 1) non è ancora certificata ottima.

2.4 Arrotondare la soluzione del rilassamento lineare

Di solito si ottengono soluzioni ammissibili del problema PLI arrotondando una soluzione ottima del rilassamento lineare. Ma nessuna tecnica di arrotondamento garantisce di trovare una soluzione ottima del problema PLI.

Esempio 2.4: arrotondamento nell'esempio dell'auto

- La soluzione ottima del rilassamento lineare è $\tilde{\mathbf{x}}(\text{LP}) = (\frac{7}{2}, \frac{5}{2})$. Arrotondando all'intero più vicino si ottiene $(4, 3)$, che è inammissibile: $10 \cdot 4 + 4 \cdot 3 = 52 > 45$.
- Arrotondando per difetto entrambe le componenti si ottiene la soluzione ammissibile $\bar{\mathbf{x}} = (3, 2)$, con $\text{LB} = 3 \cdot 3 + 2 \cdot 2 = 13$. La soluzione ottima del problema PLI è $\tilde{\mathbf{x}}(\text{ILP}) = (4, 1)$, con $z(\text{ILP}) = 14$: l'arrotondamento non dà l'ottimo, e i bound sono $13 \leq z(\text{ILP}) \leq 15$.

Esempio 2.5: l'arrotondamento funziona, ma il bound non è stretto

Consideriamo il seguente problema PLI; il suo rilassamento lineare ha $x_1, x_2 \geq 0$ al posto dei vincoli di integralità:

$$\begin{array}{llll} \max & 2x_1 & + & 4x_2 \\ \text{soggetto a} & -x_1 & + & x_2 \leq 2 \\ & 8x_1 & + & 2x_2 \leq 19 \\ & x_1, & & x_2 \in \mathbb{Z}_{\geq 0} \end{array}$$

- La soluzione ottima del rilassamento lineare è $\tilde{\mathbf{x}}(\text{LP}) = (\frac{3}{2}, \frac{7}{2})$, con $z(\text{LP}) = 17$ e $\lfloor z(\text{LP}) \rfloor = 17$, quindi $z(\text{ILP}) \leq 17$.
- Arrotondando per difetto entrambe le componenti otteniamo la soluzione ammissibile $\bar{\mathbf{x}} = (\lfloor \frac{3}{2} \rfloor, \lfloor \frac{7}{2} \rfloor) = (1, 3)$, con $\text{LB} = 2 \cdot 1 + 4 \cdot 3 = 14$.
- La soluzione ottima del problema PLI è $\tilde{\mathbf{x}}(\text{ILP}) = (1, 3)$, con $z(\text{ILP}) = 14$: la soluzione arrotondata è ottima. Ma i bound $14 \leq z(\text{ILP}) \leq 17$ non lo dimostrano.

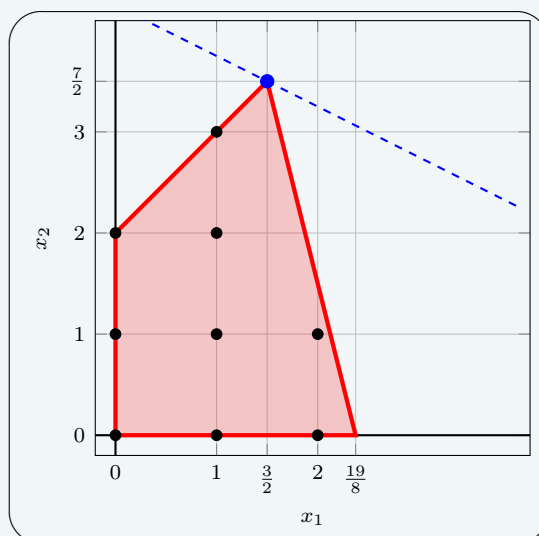


Figura 2.2: Il secondo esempio: la regione ammissibile del rilassamento lineare (rossa), i punti interi (neri) e la soluzione ottima del rilassamento lineare (blu), con la retta obiettivo $2x_1 + 4x_2 = 17$ (tratteggiata).

Esempio 2.6: l'arrotondamento fallisce

Consideriamo il seguente problema PLI:

$$\begin{aligned} \max \quad & 25x_1 + 16x_2 \\ \text{soggetto a} \quad & 50x_1 + 31x_2 \leq 250 \\ & -3x_1 + 2x_2 \leq 4 \\ & x_1, x_2 \in \mathbb{Z}_{\geq 0} \end{aligned}$$

- La soluzione ottima del rilassamento lineare è $\tilde{\mathbf{x}}(\text{LP}) = (\frac{376}{193}, \frac{950}{193}) \approx (1.94, 4.92)$, con $z(\text{LP}) = \frac{24600}{193} \approx 127.46$. Quindi $z(\text{ILP}) \leq \lfloor z(\text{LP}) \rfloor = 127$.
- Arrotondando all'intero più vicino si ottiene $(2, 5)$, che è inammissibile: $50 \cdot 2 + 31 \cdot 5 = 255 > 250$. Arrotondando $\tilde{x}_1$ per eccesso e $\tilde{x}_2$ per difetto si ottiene la soluzione ammissibile $\bar{\mathbf{x}} = (2, 4)$, con $\text{LB} = 25 \cdot 2 + 16 \cdot 4 = 114$.
- La soluzione ottima del problema PLI è $\tilde{\mathbf{x}}(\text{ILP}) = (5, 0)$, con $z(\text{ILP}) = 125$. È lontana da $\tilde{\mathbf{x}}(\text{LP})$, e nessun arrotondamento di $\tilde{\mathbf{x}}(\text{LP})$ la trova.

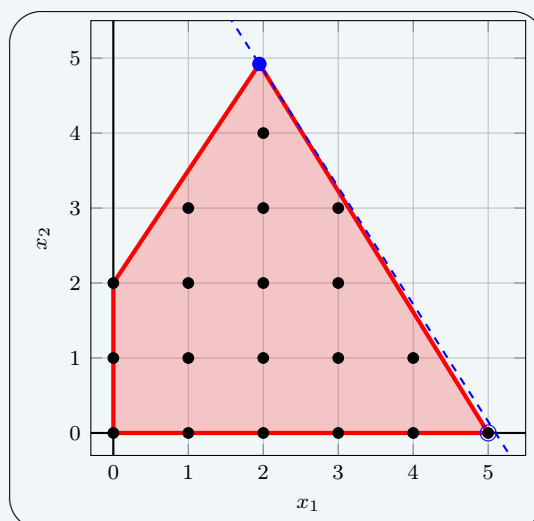


Figura 2.3: Il terzo esempio: i punti interi (neri) sono lontani dalla soluzione ottima $(\frac{376}{193}, \frac{950}{193})$ del rilassamento lineare (blu). La soluzione ottima del problema PLI è $(5,0)$ (cerchio blu). La linea tratteggiata è la retta obiettivo $25x_1 + 16x_2 = \frac{24600}{193}$.

Le soluzioni ottime del problema PLI possono essere molto lontane dalle soluzioni ottime del rilassamento lineare. Il valore del rilassamento lineare, invece, è spesso un buon bound: qui $125 \leq z(\text{LP}) \approx 127.46$.

2.5 Upper bound sulle variabili

Proposizione 2.8: upper bound sulle variabili

Sia $\sum_{j=1}^n a_{ij} x_j \leq b_i$ un vincolo con $a_{ij} \geq 0$ per ogni j . Allora ogni soluzione ammissibile del problema PLI soddisfa

$$x_j \leq \left\lfloor \frac{b_i}{a_{ij}} \right\rfloor, \quad \forall j \text{ tale che } a_{ij} > 0. \quad (2.5)$$

Dimostrazione. Poiché tutte le variabili e tutti i coefficienti sono non negativi, $a_{ij} x_j \leq \sum_{k=1}^n a_{ik} x_k \leq b_i$, quindi $x_j \leq \frac{b_i}{a_{ij}}$. La variabile è intera, quindi non supera la parte intera di $\frac{b_i}{a_{ij}}$. $\square$

- Se più vincoli danno un bound su x_j , teniamo il più piccolo. I bound tagliano parti del poliedro del rilassamento lineare, e nessun punto intero.

Gli upper bound sulle variabili migliorano il rilassamento lineare solo se ne tagliano la soluzione ottima.

Esempio 2.7: upper bound sulle variabili dell'esempio dell'auto

Entrambi i vincoli dell'esempio dell'auto hanno coefficienti non negativi:

$$x_1 \leq \min \left\{ \left\lfloor \frac{45}{10} \right\rfloor, \left\lfloor \frac{49}{4} \right\rfloor \right\} = \min\{4, 12\} = 4,$$

$$x_2 \leq \min \left\{ \left\lfloor \frac{45}{4} \right\rfloor, \left\lfloor \frac{49}{14} \right\rfloor \right\} = \min\{11, 3\} = 3.$$

- Le disuguaglianze $x_1 \leq 4$ e $x_2 \leq 3$ sono soddisfatte da tutti i punti interi dell'esempio dell'auto, e si possono aggiungere al modello. Tagliano due regioni del poliedro, e nessun punto intero.

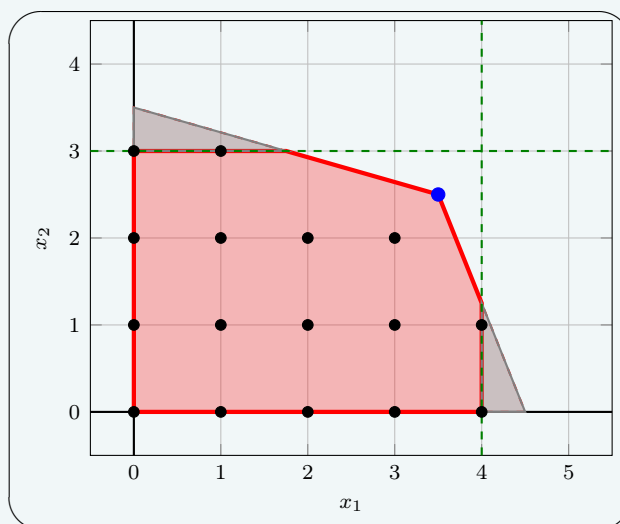


Figura 2.4: Gli upper bound sulle variabili $x_1 \leq 4$ e $x_2 \leq 3$ (verdi) tagliano le due regioni grigie del rilassamento lineare (rosso) e nessun punto intero. La soluzione ottima del rilassamento lineare (blu) non viene tagliata.

- La soluzione ottima del rilassamento lineare, $(\frac{7}{2}, \frac{5}{2})$, soddisfa entrambe le disuguaglianze. Non viene tagliata, quindi il valore ottimo del rilassamento lineare non cambia, e l'upper bound resta 15.

Esempio 2.8: upper bound che tagliano l'ottimo del rilassamento lineare

- Consideriamo la funzione obiettivo $x_1 - x_2$ sui vincoli dell'esempio dell'auto, la cui soluzione ottima del problema PLI è $(4, 0)$ con $z(\text{ILP}) = 4$. Prima dei bound, la soluzione ottima del rilassamento lineare è $(\frac{9}{2}, 0)$ con $z(\text{LP}) = \frac{9}{2}$; dopo $x_1 \leq 4$ è $(4, 0)$ con $z(\text{LP}) = 4$, che è intera e quindi ottima.

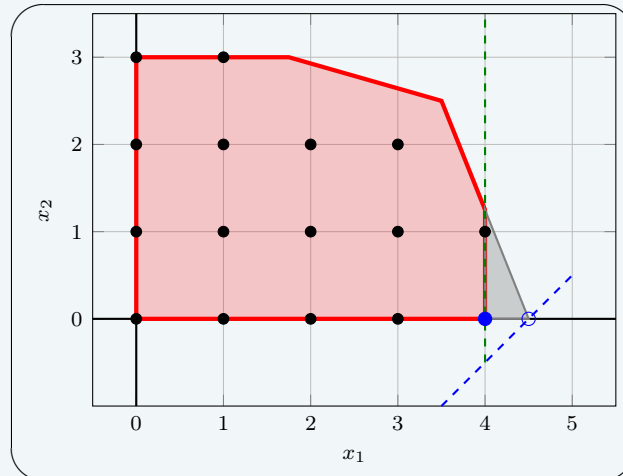


Figura 2.5: L'obiettivo $x_1 - x_2$: l'upper bound sulla variabile $x_1 \leq 4$ (verde) taglia il triangolo grigio e la soluzione ottima $(\frac{9}{2}, 0)$ del rilassamento lineare (cerchio blu). La nuova soluzione ottima $(4, 0)$ (blu) è intera. La linea tratteggiata è $x_1 - x_2 = \frac{9}{2}$.

- Il valore del rilassamento lineare migliora da $\frac{9}{2}$ a 4. Il bound arrotondato $\lfloor \frac{9}{2} \rfloor = 4$ valeva già 4: una formulazione migliore non migliora sempre il bound arrotondato.

2.6 Bound duali

- Un bound duale si può ottenere anche senza risolvere il rilassamento lineare, combinando i vincoli con **moltiplicatori non negativi** π_i , uno per ciascun vincolo.

Definizione 2.9: problema duale del rilassamento lineare

Il **problema duale** D(LP) del rilassamento lineare (2.2) è il seguente problema LP, con valore ottimo $z(D(LP))$:

$$\min \sum_{i=1}^m b_i \pi_i \quad (2.6a)$$

$$\text{soggetto a } \sum_{i=1}^m a_{ij} \pi_i \geq c_j, \quad \forall j \in \{1, 2, \dots, n\}, \quad (2.6b)$$

$$\pi_i \geq 0, \quad \forall i \in \{1, 2, \dots, m\}. \quad (2.6c)$$

- Moltiplicando il vincolo i per $\pi_i \geq 0$ e sommando, otteniamo una disuguaglianza i cui coefficienti sono almeno pari ai costi, per (2.6b). I moltiplicatori devono essere non negativi: un moltiplicatore negativo invertirebbe la disuguaglianza.

Proposizione 2.10: bound duali

Per ogni soluzione ammissibile $\bar{\pi}$ del problema duale (2.6),

$$z(\text{ILP}) \leq z(\text{LP}) \leq \sum_{i=1}^m b_i \bar{\pi}_i.$$

Per la dualità forte, $z(\text{D}(\text{LP})) = z(\text{LP})$: i moltiplicatori migliori danno esattamente il valore del rilassamento lineare.

Dimostrazione. Sia $\mathbf{x}$ ammissibile per il rilassamento lineare. Moltiplicando il vincolo i per $\bar{\pi}_i \geq 0$ e sommando, $\sum_{j=1}^n (\sum_{i=1}^m a_{ij} \bar{\pi}_i) x_j \leq \sum_{i=1}^m b_i \bar{\pi}_i$. Per (2.6b) e $x_j \geq 0$, il membro sinistro è almeno $\sum_{j=1}^n c_j x_j$. Quindi $\sum_{j=1}^n c_j x_j \leq \sum_{i=1}^m b_i \bar{\pi}_i$ per ogni $\mathbf{x}$ ammissibile. $\square$

Ogni soluzione ammissibile del problema duale è il certificato di un upper bound, trovato senza risolvere il rilassamento lineare. Il certificato migliore è una soluzione duale ottima, e ha lo stesso valore del rilassamento lineare.

Esempio 2.9: moltiplicatori per l'esempio dell'auto

- Una soluzione ammissibile del rilassamento lineare non dà un bound duale: il suo valore non è né un lower bound né un upper bound su $z(\text{ILP})$.

(x_1, x_2)	ammissibile per il rilassamento lineare	intera	$3x_1 + 2x_2$
$(4, \frac{5}{4})$	sì	no	$\frac{29}{2} > 14$
$(2, \frac{5}{2})$	sì	no	$11 < 14$

- Moltiplicando il primo vincolo dell'esempio dell'auto per $\frac{3}{4}$ otteniamo

$$\frac{15}{2} x_1 + 3x_2 \leq \frac{135}{4}.$$

Poiché $x_1, x_2 \geq 0$ e i coefficienti $\frac{15}{2}$ e 3 sono almeno pari ai costi 3 e 2, abbiamo $3x_1 + 2x_2 \leq \frac{15}{2}x_1 + 3x_2 \leq \frac{135}{4}$: un upper bound pari a $\frac{135}{4}$.

- Con i moltiplicatori $\pi_1, \pi_2 \geq 0$ dei due vincoli, i coefficienti della somma devono essere almeno pari ai costi:

$$\begin{aligned} (10\pi_1 + 4\pi_2)x_1 + (4\pi_1 + 14\pi_2)x_2 &\leq 45\pi_1 + 49\pi_2, \\ 10\pi_1 + 4\pi_2 &\geq 3, \quad 4\pi_1 + 14\pi_2 \geq 2. \end{aligned}$$

(π_1, π_2)	moltiplicatori validi	upper bound $45\pi_1 + 49\pi_2$
$(\frac{3}{4}, 0)$	sì	$\frac{135}{4}$
$(0, \frac{3}{4})$	sì	$\frac{147}{4}$
$(\frac{1}{5}, \frac{1}{4})$	sì	$\frac{85}{4}$
$(\frac{1}{2}, 0)$	sì	$\frac{45}{2}$
$(\frac{1}{5}, 0)$	no: $10 \cdot \frac{1}{5} = 2 < 3$	—

- I moltiplicatori migliori, con il valore più piccolo di $45\pi_1 + 49\pi_2$, risolvono il problema duale (2.6) dell'esempio dell'auto:

$$\begin{array}{llll}
 \min & 45\pi_1 & + & 49\pi_2 \\
 \text{soggetto a} & 10\pi_1 & + & 4\pi_2 \geq 3 \\
 & 4\pi_1 & + & 14\pi_2 \geq 2 \\
 & \pi_1, & & \pi_2 \geq 0
 \end{array}$$

Esempio 2.10: soluzione duale ottima dell'esempio dell'auto

La soluzione ottima del problema duale dell'Esempio 2.9 è $\tilde{\pi}(D(LP)) = (\frac{17}{62}, \frac{2}{31})$, e

$$z(D(LP)) = 45 \cdot \frac{17}{62} + 49 \cdot \frac{2}{31} = \frac{765}{62} + \frac{98}{31} = \frac{31}{2} = z(LP).$$

Gli altri due vertici della regione ammissibile duale danno i bound $\frac{45}{2}$ e $\frac{147}{4}$.

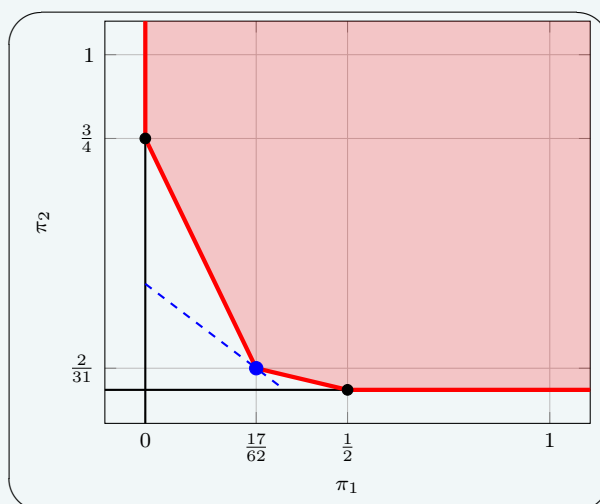


Figura 2.6: La regione ammissibile del problema LP duale (rossa, illimitata) e la sua soluzione ottima $(\frac{17}{62}, \frac{2}{31})$ (blu), con la retta obiettivo $45\pi_1 + 49\pi_2 = \frac{31}{2}$ (tratteggiata). Gli altri vertici (neri) danno i bound $\frac{45}{2}$ e $\frac{147}{4}$.

2.6.1 Riepilogo della notazione

- **Problema PLI:** soluzione ottima $\tilde{x}(\text{ILP})$, valore ottimo $z(\text{ILP})$.
- **Bound primale:** soluzione ammissibile $\bar{x}$ con $\text{LB}(\bar{x}) = \sum_j c_j \bar{x}_j \leq z(\text{ILP})$.
- **Rilassamento lineare:** soluzione ottima $\tilde{x}(\text{LP})$, valore ottimo $z(\text{LP})$, upper bound $z(\text{ILP}) \leq z(\text{LP})$.
- **Problema duale:** soluzione ottima $\tilde{\pi}(\text{D}(\text{LP}))$, valore ottimo $z(\text{D}(\text{LP})) = z(\text{LP})$, e $z(\text{ILP}) \leq \sum_i b_i \tilde{\pi}_i$ per ogni soluzione duale ammissibile $\tilde{\pi}$.

2.7 Errori tipici

Attenzione

- **Prendere una soluzione euristica per ottima.** Una soluzione ammissibile dà solo un bound primale. Senza un bound duale dello stesso valore non c'è alcun certificato di ottimalità.
- **Arrotondare il bound quando non è lecito.** Il bound $\lfloor z(\text{LP}) \rfloor$ richiede che tutte le variabili e tutti i costi siano interi. In un problema MIP il bound è $z(\text{LP})$.
- **Arrotondare la soluzione e chiamarla ottima.** Una soluzione arrotondata è, nel migliore dei casi, ammissibile, e in generale non ottima. Può anche essere inammissibile, come il punto intero più vicino $(2, 5)$ dell'[Esempio 2.6](#).
- **Usare come bound il valore di una soluzione ammissibile del rilassamento lineare.** Solo il valore ottimo del rilassamento lineare, o il valore di una soluzione duale ammissibile, è un upper bound.
- **Moltiplicatori con segno negativo.** I moltiplicatori dei vincoli $\leq$ devono essere non negativi, e i coefficienti combinati devono essere almeno pari ai costi.
- **Ricavare un bound su una variabile da un vincolo con coefficienti negativi.** La dimostrazione di (2.5) richiede che tutti i coefficienti siano non negativi.
- **Credere che una formulazione più forte migliori sempre il bound.** I bound $x_1 \leq 4$ e $x_2 \leq 3$ lasciano invariato il bound 15 dell'esempio dell'auto.

2.8 In sintesi

In sintesi

- Ogni soluzione ammissibile è un bound primale (un lower bound) su $z(\text{ILP})$, e gli algoritmi euristici trovano rapidamente buone soluzioni ammissibili.
- Il rilassamento lineare elimina l'integralità, quindi $z(\text{ILP}) \leq z(\text{LP})$; con variabili intere e costi interi questo bound duale (un upper bound) si può arrotondare per difetto a $\lfloor z(\text{LP}) \rfloor$.
- Bound primale e bound duale definiscono il gap di ottimalità; l'ottimalità è certificata quando il gap è zero.
- Arrotondare una soluzione ottima del rilassamento lineare dà, nel migliore dei casi, una soluzione ammissibile, mai un certificato.
- Gli upper bound sulle variabili $x_j \leq \lfloor b_i/a_{ij} \rfloor$ vengono da vincoli con coefficienti non negativi, e servono solo se tagliano l'ottimo del rilassamento lineare.
- I moltiplicatori non negativi dei vincoli danno bound duali; i migliori risolvono il problema duale, il cui valore ottimo è $z(\text{LP})$.

Cruscotto dell'esempio dell'auto. I bound su $z(\text{ILP})$ trovati finora in questo modulo (bound primale: lower bound; bound duale: upper bound):

Capitolo	Lower bound	Upper bound	Da
Bound primali e bound duali	14	15	soluzione euristica $(4, 1)$, e $\lfloor z(\text{LP}) \rfloor$ con $z(\text{LP}) = \frac{31}{2}$

Parte II

Rafforzare la formulazione

Formulazioni ideali

Prerequisiti

- Da questo modulo, i capitoli «Programmazione lineare intera» e «Bound primali e bound duali»: la regione ammissibile di un problema PLI, il rilassamento lineare e il suo bound, e l'esempio dell'auto.
- Dal modulo *Programmazione Lineare*, [Capitolo 15 «La geometria della Programmazione Lineare»](#): i vertici di un poliedro, e il fatto che un problema LP con una soluzione ottima ha un vertice ottimo.
- Dal modulo *Ottimizzazione convessa*, [Capitolo 1 «Insiemi convessi»](#): combinazioni convesse e inviluppi convessi. In questo capitolo ne diamo un breve richiamo.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- calcolare l'inviluppo convesso delle soluzioni ammissibili di un piccolo problema PLI, e classificarne i punti in vertici, punti del bordo e punti interni;
- definire una formulazione di un problema PLI, e decidere se è ideale;
- spiegare perché un problema LP su una formulazione ideale risolve il problema PLI;
- confrontare due formulazioni dello stesso problema attraverso i loro rilassamenti lineari;
- spiegare perché le formulazioni ideali esistono ma sono difficili da trovare.

Dove eravamo

Nell'esempio dell'auto abbiamo il lower bound 14 e l'upper bound $\lfloor z(\text{LP}) \rfloor = 15$. Il gap di $\frac{1}{14}$ è dovuto al poliedro dei vincoli, che contiene punti frazionari fuori dall'inviluppo convesso dei punti interi. Questo capitolo definisce il poliedro che non lascia alcun gap.

3.1 L'involuppo convesso

Definizione 3.1: involuppo convesso

Dato un insieme $S \subseteq \mathbb{R}^n$, una **combinazione convessa** dei punti $s^1, s^2, \dots, s^k \in S$ è un punto $\sum_{h=1}^k \lambda_h s^h$ con $\lambda_h \geq 0$ e $\sum_{h=1}^k \lambda_h = 1$. L'**involuppo convesso** di S , indicato con $\text{conv}(S)$, è l'insieme di tutte le combinazioni convesse di punti di S .

- L'involuppo convesso è il più piccolo insieme convesso che contiene S . Se S è un insieme finito di punti, l'involuppo convesso è un poliedro limitato (un **politopo**), e i suoi **vertici** sono punti di S . Gli altri punti di S stanno sul bordo del politopo oppure al suo interno: non servono per descriverlo.
- Per un problema PLI prendiamo $S = F$, l'insieme delle soluzioni ammissibili. I vertici di $\text{conv}(F)$ sono soluzioni intere ammissibili.

Esempio 3.1: involuppo convesso dell'esempio dell'auto

L'involuppo convesso dei 16 punti interi dell'esempio dell'auto ha 6 vertici:

$$(0, 0), \quad (4, 0), \quad (4, 1), \quad (3, 2), \quad (1, 3), \quad (0, 3).$$

Degli altri 10 punti, cinque stanno sul bordo, cioè $(0, 1)$, $(0, 2)$, $(1, 0)$, $(2, 0)$ e $(3, 0)$, e cinque all'interno, cioè $(1, 1)$, $(2, 1)$, $(3, 1)$, $(1, 2)$ e $(2, 2)$.

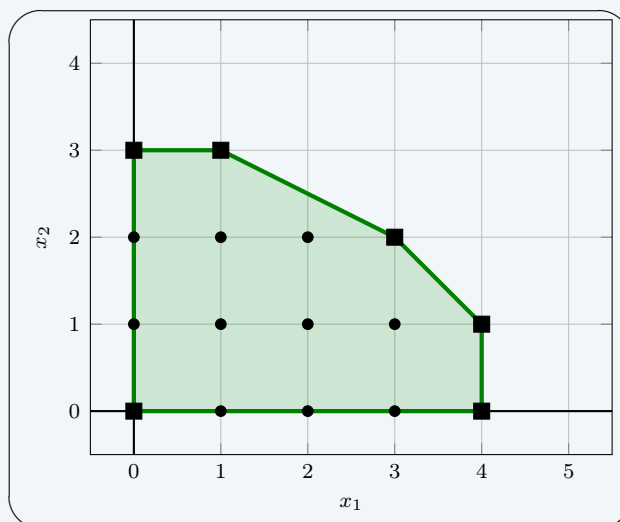


Figura 3.1: L'involuppo convesso dei 16 punti interi dell'esempio dell'auto (verde). I suoi 6 vertici (quadrati neri) sono punti interi; altri cinque punti stanno sul bordo e cinque all'interno.

3.2 Formulazioni e formulazioni ideali

Definizione 3.2: formulazione

Una **formulazione** di un problema PLI con regione ammissibile F è un poliedro $P = \{x \in \mathbb{R}^n : \sum_j a_{ij} x_j \leq b_i, \forall i, x_j \geq 0, \forall j\}$ tale che $P \cap \mathbb{Z}^n = F$. Il **rilassamento lineare della formulazione** è il problema di massimizzare $\sum_j c_j x_j$ su P .

- Un problema PLI ha molte formulazioni: hanno in comune i punti interi e differiscono nei punti frazionari. Poiché P è convesso e contiene F , ogni formulazione contiene l'involuppo convesso:

$$\text{conv}(F) \subseteq P.$$

Definizione 3.3: formulazione ideale

Una formulazione P di un problema PLI è **ideale** se $P = \text{conv}(F)$, cioè se la regione ammissibile del suo rilassamento lineare è esattamente l'involuppo convesso delle soluzioni intere ammissibili.

Teorema 3.4: LP su una formulazione ideale

Sia P una formulazione ideale di un problema PLI con $F \neq \emptyset$ e P limitato. Allora il rilassamento lineare ha un vertice ottimo, ogni vertice ottimo è una soluzione ottima del problema PLI, e $z(\text{LP}) = z(\text{ILP})$.

Dimostrazione. Il poliedro $P = \text{conv}(F)$ è non vuoto e limitato, quindi il problema LP ha un vertice ottimo di P . I vertici di $\text{conv}(F)$ appartengono a F , quindi questo vertice è una soluzione ammissibile del problema PLI, di valore $z(\text{LP})$. Poiché $z(\text{ILP}) \leq z(\text{LP})$, il vertice è ottimo per il problema PLI. $\square$

- L'algoritmo del simplesso restituisce un vertice ottimo, quindi con una formulazione ideale restituisce una soluzione intera. Se il problema LP ha più soluzioni ottime, un metodo che non restituisce vertici può dare un punto frazionario sulla faccia ottima: per questo il teorema parla di un vertice ottimo.

Con una formulazione ideale, un problema PLI si risolve come un problema LP: un vertice ottimo del rilassamento lineare è una soluzione ottima del problema PLI.

Esempio 3.2: la formulazione dell'esempio dell'auto non è ideale

Il punto $(\frac{7}{2}, \frac{5}{2})$ appartiene al poliedro P dell'esempio dell'auto ma non a $\text{conv}(F)$, perché viola $x_1 + x_2 \leq 5$, che tutti i vertici dell'involuppo dell'Esempio 3.1 soddisfano: $\text{conv}(F) \subsetneq P$.

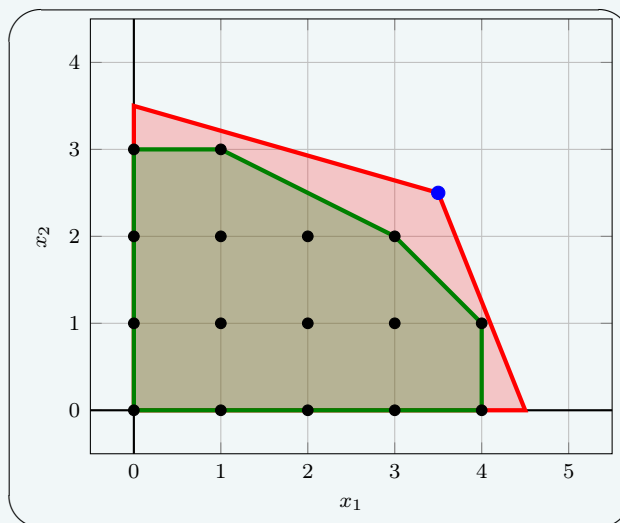


Figura 3.2: L'esempio dell'auto: il poliedro del rilassamento lineare (rosso) contiene l'involuppo convesso dei punti interi (verde), e la soluzione ottima del rilassamento lineare (blu) sta fuori dall'involuppo.

Esempio 3.3: una formulazione ideale dell'esempio dell'auto

L'involuppo convesso dell'esempio dell'auto è descritto dalle quattro disuguaglianze seguenti e dai vincoli di non negatività:

$$\begin{array}{llll}
 \max & 3x_1 & + & 2x_2 \\
 \text{soggetto a} & x_1 & + & x_2 \leq 5 \\
 & x_1 & + & 2x_2 \leq 7 \\
 & x_1 & & \leq 4 \\
 & & x_2 & \leq 3 \\
 & x_1, & x_2 & \geq 0
 \end{array}$$

Queste disuguaglianze definiscono esattamente i 6 vertici dell'involuppo, e gli stessi 16 punti interi. Il rilassamento lineare di questa **formulazione ideale** ha la soluzione ottima $(4, 1)$, che è intera, con $z(\text{LP}) = 14 = z(\text{ILP})$.

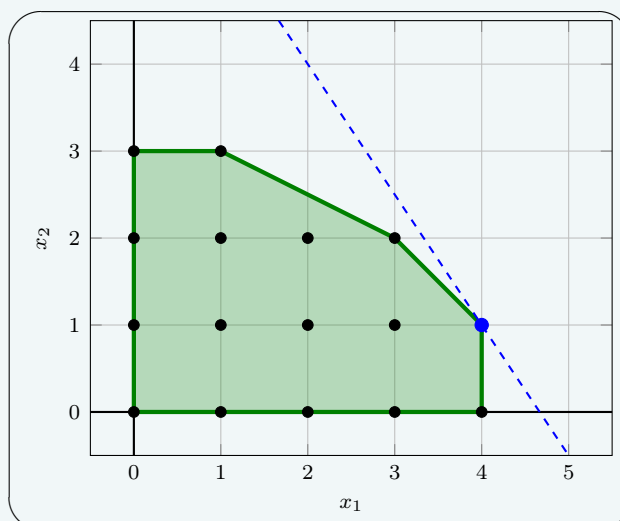


Figura 3.3: La formulazione ideale dell'esempio dell'auto: la sua regione ammissibile (verde) è l'involuppo convesso dei punti interi. La soluzione ottima del problema LP è il punto intero (4, 1) (blu).

3.3 Confronto tra formulazioni

Definizione 3.5: formulazione più forte

Date due formulazioni P_1 e P_2 dello stesso problema PLI, P_1 è **più forte** di P_2 se $P_1 \subseteq P_2$.

- Una formulazione più forte non dà mai un upper bound peggiore: il massimo su un insieme più piccolo non è più grande, quindi

$$P_1 \subseteq P_2 \implies z(\text{ILP}) \leq z(\text{LP}_1) \leq z(\text{LP}_2).$$

- La formulazione ideale è la più forte di tutte, perché è contenuta in ogni formulazione. Nell'esempio dell'auto la formulazione originale dà il bound $\lfloor \frac{31}{2} \rfloor = 15$, e quella ideale dell'Esempio 3.3 dà il bound 14.

Tra le formulazioni dello stesso problema PLI preferiamo quelle con il poliedro più piccolo: i loro rilassamenti lineari danno gli upper bound migliori, e la formulazione ideale dà il valore ottimo.

3.4 Esistenza delle formulazioni ideali

- Una formulazione ideale esiste per ogni problema PLI con regione ammissibile limitata o, più in generale, con dati razionali (teorema di Meyer; in questo modulo i dati sono interi). L'involuppo convesso dei punti interi è un poliedro, quindi è descritto da un numero finito di disuguaglianze lineari.
- Trovare la formulazione ideale è in generale molto difficile. Può servire un numero esponenziale di disuguaglianze, e trovare quelle necessarie è spesso difficile quanto risolvere il problema PLI stesso.
- In pratica teniamo la formulazione che abbiamo e aggiungiamo, passo dopo passo, disuguaglianze soddisfatte da tutti i punti interi che tagliano punti frazionari: le **disuguaglianze valide** del prossimo capitolo.

3.5 Errori tipici

Attenzione

- **Credere che i punti che non sono vertici stiano strettamente all'interno dell'involuppo.** Possono stare sul bordo, come $(0, 1)$ e $(3, 0)$ nell'esempio dell'auto.
- **Controllare solo la soluzione ottima.** Una formulazione può restituire una soluzione ottima intera senza essere ideale: l'ottimo del rilassamento lineare può essere intero per una funzione obiettivo e frazionario per un'altra.
- **Credere che ogni formulazione sia ideale.** La formulazione dell'esempio dell'auto ha i punti interi giusti, ma contiene punti frazionari fuori dall'involuppo.
- **Aggiungere disuguaglianze che eliminano punti interi.** Una formulazione deve conservare tutti i punti di F : il poliedro può restringersi solo fino all'involuppo convesso, mai oltre.

3.6 In sintesi

In sintesi

- L'involuppo convesso delle soluzioni ammissibili è il più piccolo poliedro che le contiene, e i suoi vertici sono soluzioni intere ammissibili.
- Una formulazione è ideale se il suo poliedro è l'involuppo convesso; allora un vertice ottimo del rilassamento lineare è una soluzione ottima del problema PLI.
- Una formulazione più forte, con un poliedro più piccolo, non dà mai un bound peggiore.
- L'esempio dell'auto ha una formulazione ideale con 4 disuguaglianze, che dà $z(LP) = 14 = z(ILP)$.

- Le formulazioni ideali esistono per dati razionali, ma in generale richiedono un numero esponenziale di disuguaglianze e sono difficili da trovare.

Cruscotto dell'esempio dell'auto. I bound su $z(\text{ILP})$ trovati finora in questo modulo:

Capitolo	Lower bound	Upper bound	Da
Bound primali e bound duali	14	15	soluzione euristica $(4, 1)$, e $\lfloor z(\text{LP}) \rfloor$ con $z(\text{LP}) = \frac{31}{2}$
Formulazioni ideali	14	14	formulazione ideale, vertice ottimo $(4, 1)$

Disuguaglianze valide

Prerequisiti

- Da questo modulo, il capitolo «Bound primali e bound duali»: il rilassamento lineare, il bound arrotondato $\lfloor z(\text{LP}) \rfloor$, e i bound ottenuti con moltiplicatori non negativi dei vincoli.
- Da questo modulo, il capitolo «Formulazioni ideali»: le formulazioni, l'involuppo convesso e il confronto tra formulazioni attraverso i loro poliedri.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- definire una disuguaglianza valida, e dire quando taglia un punto frazionario;
- ricavare una disuguaglianza valida da moltiplicatori non negativi dei vincoli, arrotondando per difetto;
- migliorare una disuguaglianza valida finché non tocca un punto intero ammissibile;
- misurare l'effetto di una disuguaglianza valida sul valore ottimo del rilassamento lineare e sull'upper bound;
- spiegare perché le disuguaglianze valide sono il modo pratico per avvicinarsi a una formulazione ideale.

Dove eravamo

Nell'esempio dell'auto la formulazione ideale dà l'upper bound 14, uguale al lower bound, e quindi certifica che $(4, 1)$ è ottima. Ma abbiamo trovato le sue quattro disuguaglianze guardando l'involuppo convesso, cosa che in generale non è possibile. Ora impariamo a **ricavare** queste disuguaglianze dai vincoli.

4.1 Disuguaglianze valide

Definizione 4.1: disuguaglianza valida

Una disuguaglianza $\sum_{j=1}^n \alpha_j x_j \leq \beta$ è una **disuguaglianza valida** per un problema PLI con regione ammissibile F se è soddisfatta da tutti i punti di F .

Definizione 4.2: punto tagliato e disuguaglianza valida saturata

Sia $\sum_{j=1}^n \alpha_j x_j \leq \beta$ una disuguaglianza valida per un problema PLI con regione ammissibile F .

- La disuguaglianza **taglia** un punto $\mathbf{x}^* \in \mathbb{R}^n$ se $\sum_{j=1}^n \alpha_j x_j^* > \beta$.
 - La disuguaglianza è **saturata** se almeno un punto di F la soddisfa all'uguaglianza.
- Una disuguaglianza valida non taglia nessun punto di F , ma può tagliare punti frazionari del rilassamento lineare. Il termine noto di una disuguaglianza saturata non si può diminuire senza perdere una soluzione ammissibile.
 - Aggiungere una disuguaglianza valida alla formulazione non cambia F , e restringe il poliedro P . La nuova formulazione è più forte, quindi il suo rilassamento lineare dà un upper bound non peggiore.
 - Una disuguaglianza valida è soddisfatta dai punti di F se e solo se è soddisfatta dal loro inviluppo convesso, quindi è valida per $\text{conv}(F)$. Le disuguaglianze della formulazione ideale sono disuguaglianze valide.

Una disuguaglianza valida elimina punti frazionari e mai una soluzione intera ammissibile. Aggiungere disuguaglianze valide rende la formulazione più forte e l'upper bound migliore.

Esempio 4.1: disuguaglianze valide dell'esempio dell'auto

Le disuguaglianze $x_1 \leq 4$ e $x_2 \leq 3$ dei capitoli precedenti sono disuguaglianze valide dell'esempio dell'auto. Non tagliano la soluzione ottima $(\frac{7}{2}, \frac{5}{2})$ del rilassamento lineare, quindi non migliorano l'upper bound 15.

4.2 Ricavare disuguaglianze valide per arrotondamento

- Combiniamo i vincoli con moltiplicatori non negativi, come per i bound duali; poiché le variabili sono intere, il termine noto della combinazione si può arrotondare per difetto.

Proposizione 4.3: disuguaglianze valide per arrotondamento

Siano $\pi_1, \pi_2, \dots, \pi_m \geq 0$ moltiplicatori tali che $\sum_{i=1}^m \pi_i a_{ij} \geq \alpha_j$, con $\alpha_j \in \mathbb{Z}$, per ogni $j \in \{1, 2, \dots, n\}$. Allora la disuguaglianza

$$\sum_{j=1}^n \alpha_j x_j \leq \left\lfloor \sum_{i=1}^m \pi_i b_i \right\rfloor \quad (4.1)$$

è valida per il problema PLI.

Dimostrazione. Per ogni soluzione ammissibile, $\sum_j \alpha_j x_j \leq \sum_j (\sum_i \pi_i a_{ij}) x_j \leq \sum_i \pi_i b_i$, per $x_j \geq 0$ e per i vincoli moltiplicati per $\pi_i \geq 0$. Il membro sinistro è intero, perché α_j e x_j sono interi, quindi è al più la parte intera del membro destro. $\square$

- La proposizione richiede che tutte le variabili siano intere e non negative. Con una variabile continua il membro sinistro non è intero, e il termine noto non si può arrotondare.

Esempio 4.2: arrotondamento con moltiplicatore $\frac{1}{2}$

Nell'esempio dell'auto moltiplichiamo il primo vincolo per $\frac{1}{2}$: otteniamo $5x_1 + 2x_2 \leq \frac{45}{2}$. Poiché il membro sinistro è intero, otteniamo la disuguaglianza valida

$$5x_1 + 2x_2 \leq 22.$$

Moltiplicando il secondo vincolo per $\frac{1}{2}$ otteniamo $2x_1 + 7x_2 \leq \frac{49}{2}$, quindi

$$2x_1 + 7x_2 \leq 24.$$

Entrambe le disuguaglianze tagliano la soluzione ottima $(\frac{7}{2}, \frac{5}{2})$ del rilassamento lineare: $5 \cdot \frac{7}{2} + 2 \cdot \frac{5}{2} = \frac{45}{2} > 22$ e $2 \cdot \frac{7}{2} + 7 \cdot \frac{5}{2} = \frac{49}{2} > 24$. I nuovi rilassamenti lineari danno:

Disuguaglianza aggiunta	$\tilde{x}(\text{LP})$	$z(\text{LP})$	$\lfloor z(\text{LP}) \rfloor$
nessuna	$(\frac{7}{2}, \frac{5}{2})$	$\frac{31}{2}$	15
$5x_1 + 2x_2 \leq 22$	$(\frac{105}{31}, \frac{157}{62})$	$\frac{472}{31}$	15
$2x_1 + 7x_2 \leq 24$	$(\frac{219}{62}, \frac{75}{31})$	$\frac{957}{62}$	15

L'ottimo del rilassamento lineare si sposta, ma l'upper bound non migliora: $\lfloor \frac{472}{31} \rfloor = \lfloor \frac{957}{62} \rfloor = 15$.

Una disuguaglianza valida che taglia la soluzione ottima del rilassamento lineare non sempre migliora l'upper bound. La scelta dei moltiplicatori conta: cerchiamo disuguaglianze che tocchino i punti interi.

4.3 Migliorare una disuguaglianza valida

- Dati i coefficienti α_j , la disuguaglianza valida più forte ha il termine noto più piccolo possibile, che è il massimo di $\sum_j \alpha_j x_j$ su F . Lo troviamo abbassando il termine noto β della (4.1) finché un punto intero ammissibile non lo raggiunge.

Algoritmo

Input: Un problema PLI in forma canonica di massimizzazione con $F \neq \emptyset$, e coefficienti interi $\alpha_1, \alpha_2, \dots, \alpha_n$.

Output: Una disuguaglianza valida saturata $\sum_{j=1}^n \alpha_j x_j \leq \beta$.

Passo 0: Bound grossolano

Trovare moltiplicatori $\pi_i \geq 0$ con $\sum_i \pi_i a_{ij} \geq \alpha_j$ per ogni j , e porre $\beta \leftarrow \lfloor \sum_i \pi_i b_i \rfloor$.

Passo 1: Verifica del termine noto

Se nessun punto intero ammissibile soddisfa $\sum_j \alpha_j x_j = \beta$, porre $\beta \leftarrow \beta - 1$ e ripetere il Passo 1.

Passo 2: Disuguaglianza saturata

Un punto intero ammissibile soddisfa $\sum_j \alpha_j x_j = \beta$, quindi $\sum_j \alpha_j x_j \leq \beta$ è valida e saturata. **STOP**.

- Nel Passo 0 i moltiplicatori si possono scegliere liberamente: migliori sono, meno ripetizioni del Passo 1 servono. Nel Passo 1, con due variabili, sostituiamo una variabile usando l'equazione e limitiamo l'altra con i vincoli: se i limiti si contraddicono, il punto non esiste.

Esempio 4.3: la disuguaglianza $x_1 + x_2 \leq 5$

Cerchiamo una disuguaglianza valida con $\alpha = (1, 1)$.

- **Passo 0.** Con i moltiplicatori $(\frac{1}{14}, \frac{1}{14})$ la somma dei vincoli, divisa per 14, è $x_1 + \frac{9}{7}x_2 \leq \frac{47}{7}$. I coefficienti sono almeno $(1, 1)$, quindi $x_1 + x_2 \leq \lfloor \frac{47}{7} \rfloor = 6$, e poniamo $\beta = 6$.
- **Passo 1, $\beta = 6$.** Sia $x_1 + x_2 = 6$, cioè $x_2 = 6 - x_1$. Il primo vincolo diventa $6x_1 + 24 \leq 45$, cioè $x_1 \leq \frac{7}{2}$, quindi $x_1 \leq 3$. Il secondo diventa $84 - 10x_1 \leq 49$, cioè $x_1 \geq \frac{7}{2}$, quindi $x_1 \geq 4$. I limiti si contraddicono: nessun punto ammissibile ha $x_1 + x_2 = 6$, e poniamo $\beta = 5$.
- **Passo 1, $\beta = 5$.** Sia $x_2 = 5 - x_1$. I vincoli diventano $x_1 \leq \frac{25}{6}$, quindi $x_1 \leq 4$, e $x_1 \geq \frac{21}{10}$, quindi $x_1 \geq 3$. Non c'è contraddizione, e i punti $(3, 2)$ e $(4, 1)$ sono ammissibili con $x_1 + x_2 = 5$.
- **Passo 2.** La disuguaglianza $x_1 + x_2 \leq 5$ è valida e saturata.

Taglia $(\frac{7}{2}, \frac{5}{2})$ (la somma è 6). Il nuovo rilassamento lineare ha la soluzione ottima $(\frac{25}{6}, \frac{5}{6})$ con $z(\text{LP}) = \frac{85}{6}$, quindi l'upper bound è $\lfloor \frac{85}{6} \rfloor = 14$.

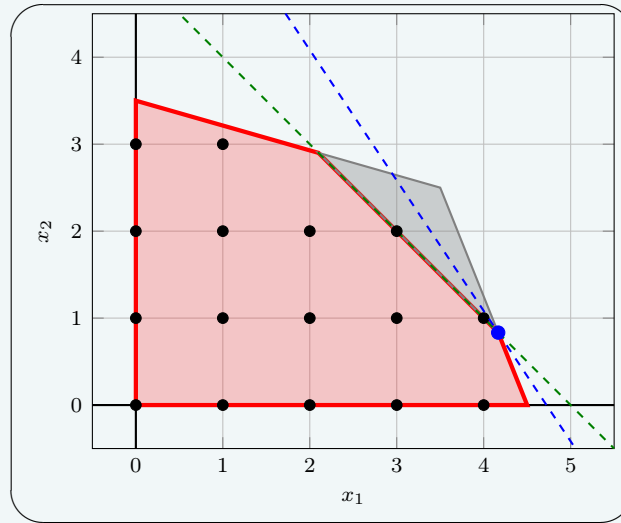


Figura 4.1: La disuguaglianza valida $x_1 + x_2 \leq 5$ (verde) taglia il triangolo grigio, che non contiene punti interi, e la soluzione ottima del rilassamento lineare precedente. La nuova soluzione ottima è $(\frac{25}{6}, \frac{5}{6})$ (blu), con $z = \frac{85}{6}$ (linea tratteggiata).

Esempio 4.4: la disuguaglianza $x_1 + 2x_2 \leq 7$

Cerchiamo una disuguaglianza valida con $\alpha = (1, 2)$.

- **Passo 0.** Con i moltiplicatori $(\frac{3}{62}, \frac{4}{31})$ la combinazione dei vincoli è $x_1 + 2x_2 \leq \frac{17}{2}$, perché $\frac{3 \cdot 10}{62} + \frac{4 \cdot 4}{31} = 1$, $\frac{3 \cdot 4}{62} + \frac{4 \cdot 14}{31} = 2$ e $\frac{3 \cdot 45}{62} + \frac{4 \cdot 49}{31} = \frac{17}{2}$. Quindi $x_1 + 2x_2 \leq \lfloor \frac{17}{2} \rfloor = 8$, e poniamo $\beta = 8$.
- **Passo 1, $\beta = 8$.** Sia $x_1 = 8 - 2x_2$. Il primo vincolo diventa $80 - 16x_2 \leq 45$, quindi $x_2 \geq 3$; il secondo diventa $32 + 6x_2 \leq 49$, quindi $x_2 \leq 2$. I due limiti si contraddicono, e poniamo $\beta = 7$.
- **Passo 1, $\beta = 7$.** Sia $x_1 = 7 - 2x_2$: i vincoli diventano $70 - 16x_2 \leq 45$, quindi $x_2 \geq 2$, e $28 + 6x_2 \leq 49$, quindi $x_2 \leq 3$. I punti $(3, 2)$ e $(1, 3)$ sono ammissibili con $x_1 + 2x_2 = 7$.
- **Passo 2.** La disuguaglianza $x_1 + 2x_2 \leq 7$ è valida e saturata.

Il nuovo rilassamento lineare ha la soluzione ottima $(\frac{31}{8}, \frac{25}{16})$ con $z(\text{LP}) = \frac{59}{4}$, quindi l'upper bound è $\lfloor \frac{59}{4} \rfloor = 14$.

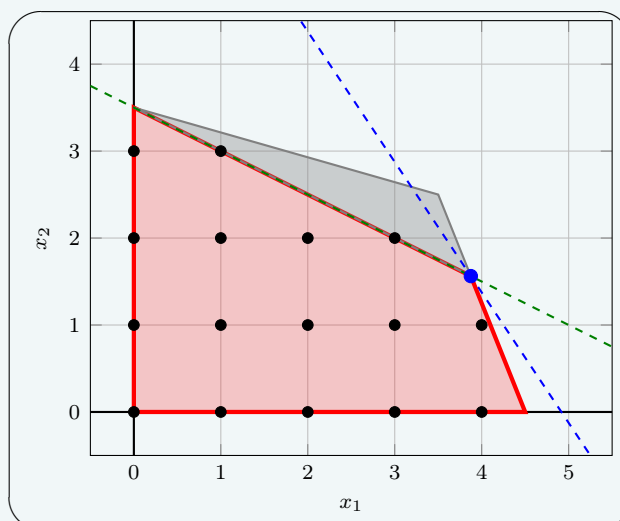


Figura 4.2: La disuguaglianza valida $x_1 + 2x_2 \leq 7$ (verde) taglia il triangolo grigio, che non contiene punti interi. La nuova soluzione ottima è $(\frac{31}{8}, \frac{25}{16})$ (blu), con $z = \frac{59}{4}$ (linea tratteggiata).

Con la disuguaglianza valida $x_1 + x_2 \leq 5$, oppure con $x_1 + 2x_2 \leq 7$, l'upper bound scende da 15 a 14. È uguale al lower bound dato da $(4, 1)$: **la soluzione $(4, 1)$ è certificata ottima**, e il gap è zero.

- Le due disuguaglianze, insieme a $x_1 \leq 4$ e $x_2 \leq 3$, sono esattamente le quattro disuguaglianze della formulazione ideale del capitolo precedente. Per l'esempio dell'auto abbiamo raggiunto l'involuppo convesso con disuguaglianze valide ricavate dai vincoli.

4.4 Famiglie di disuguaglianze valide e tagli

- Con due variabili le direzioni α si scelgono guardando la figura. In generale, le disuguaglianze valide vengono da due fonti.
- Famiglie di disuguaglianze per strutture specifiche dei vincoli, come le **disuguaglianze di copertura** di un vincolo a coefficienti positivi di un problema binario (prossimo capitolo).
- L'arrotondamento di questo capitolo, applicato in modo sistematico alle righe del dizionario ottimo del rilassamento lineare: i **tagli frazionari di Gomory** dell'algoritmo dei piani di taglio, e poi l'algoritmo **branch-and-cut**.

4.5 Errori tipici

Attenzione

- **Arrotondare quando una variabile è continua.** L'arrotondamento della (4.1) richiede variabili intere e coefficienti α_j interi.
- **Arrotondare per eccesso invece che per difetto.** Il termine noto si arrotonda per difetto. Arrotondare per eccesso dà una disuguaglianza che non è valida, e può eliminare soluzioni intere.
- **Credere che ogni disuguaglianza valida migliori il bound.** Le disuguaglianze $5x_1 + 2x_2 \leq 22$ e $2x_1 + 7x_2 \leq 24$ tagliano l'ottimo del rilassamento lineare, ma lasciano il bound a 15.
- **Usare moltiplicatori non abbastanza grandi.** I coefficienti combinati devono essere almeno α_j per ogni j , e i moltiplicatori devono essere non negativi.
- **Fermarsi a un termine noto grossolano.** La disuguaglianza $x_1 + x_2 \leq 6$ è valida, ma non taglia l'ottimo del rilassamento lineare dell'esempio dell'auto (la sua somma è 6); la disuguaglianza saturata $x_1 + x_2 \leq 5$ lo taglia.

4.6 In sintesi

In sintesi

- Una disuguaglianza valida è soddisfatta da tutte le soluzioni intere ammissibili; può tagliare punti frazionari del rilassamento lineare, e rende la formulazione più forte.
- Moltiplicatori non negativi dei vincoli, con il termine noto arrotondato per difetto, danno disuguaglianze valide per i problemi PLI.
- Abbassare il termine noto finché un punto intero non tocca la retta dà una disuguaglianza saturata.
- Una disuguaglianza valida serve solo se taglia l'ottimo del rilassamento lineare e migliora il bound arrotondato.
- Nell'esempio dell'auto, $x_1 + x_2 \leq 5$ oppure $x_1 + 2x_2 \leq 7$ abbassa l'upper bound da 15 a 14, e certifica che $(4, 1)$ è ottima.

Cruscotto dell'esempio dell'auto. I bound su $z(\text{ILP})$ trovati finora in questo modulo:

Capitolo	Lower bound	Upper bound	Da
Bound primali e bound duali	14	15	soluzione euristica $(4, 1)$, e $\lfloor z(\text{LP}) \rfloor$ con $z(\text{LP}) = \frac{31}{2}$
Formulazioni ideali	14	14	formulazione ideale, vertice ottimo $(4, 1)$
Disuguaglianze valide	14	14	$x_1 + x_2 \leq 5$ dà $z(\text{LP}) = \frac{85}{6}$

Disuguaglianze di cover

Prerequisiti

- Da questo modulo, [Capitolo 1 «Programmazione lineare intera»](#) e [Capitolo 2 «Rilassamento lineare»](#): problemi di PLI, il rilassamento lineare e l'upper bound fornito dal rilassamento lineare.
- Da questo modulo, [Capitolo 3 «Formulazioni ideali»](#) e [Capitolo 4 «Disuguaglianze valide»](#): l'involuppo convesso delle soluzioni intere, le disuguaglianze valide e i tagli che rimuovono punti frazionari.
- Il rilassamento lineare di un problema binario con un vincolo, risolto inserendo le variabili in ordine non crescente del rapporto c_j/a_j tra i coefficienti dell'obiettivo e del vincolo (regola di Dantzig).

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- scrivere le disuguaglianze di copertura di un vincolo $\leq$ con coefficienti positivi di un MIP binario e dimostrare che sono valide;
- elencare le coperture e le coperture minimali di un'istanza, e descrivere la chiusura delle coperture;
- separare una disuguaglianza di copertura da un punto frazionario risolvendo un problema di separazione delle coperture, in modo esatto o con l'algoritmo euristico greedy;
- rafforzare una disuguaglianza di copertura in una disuguaglianza di copertura estesa;
- calcolare la violazione di una disuguaglianza di copertura e la distanza di un punto da essa.

5.1 Un vincolo con coefficienti positivi e il suo rilassamento lineare

Definizione 5.1: vincolo con coefficienti positivi di un MIP binario

Date le variabili binarie x_j , per $j \in \mathcal{J} = \{1, 2, \dots, n\}$, i coefficienti interi $a_j \in \mathbb{Z}_{\geq 1}$ e un termine noto $b \in \mathbb{Z}_{\geq 1}$, consideriamo il vincolo

$$\sum_{j \in \mathcal{J}} a_j x_j \leq b, \quad (5.1a)$$

$$x_j \in \{0, 1\}, \quad \forall j \in \mathcal{J}. \quad (5.1b)$$

Il suo insieme delle soluzioni ammissibili è

$$X = \left\{ \mathbf{x} \in \{0, 1\}^n : \sum_{j \in \mathcal{J}} a_j x_j \leq b \right\}, \quad (5.2)$$

e il suo rilassamento lineare è

$$P = \left\{ \mathbf{x} \in [0, 1]^n : \sum_{j \in \mathcal{J}} a_j x_j \leq b \right\}. \quad (5.3)$$

- I vincoli di questo tipo si chiamano **vincoli di zaino** o **vincoli di budget**.
- Supponiamo che

$$a_j \leq b \quad \text{per ogni } j \in \mathcal{J} \quad \text{e} \quad \sum_{j \in \mathcal{J}} a_j > b.$$

La prima ipotesi garantisce che ogni variabile possa essere posta a 1, la seconda che non possano essere poste tutte a 1 insieme.

- X è un insieme finito di punti binari, e una descrizione esatta del suo involucro convesso $\text{conv}(X)$ è in generale difficile da ottenere. L'obiettivo è rafforzare P rimuovendo punti frazionari senza rimuovere alcun punto di X .
- Un MIP può contenere più vincoli di questo tipo, e altri vincoli ancora. Le disuguaglianze valide per X restano valide quando si impongono altri vincoli.

Con i coefficienti dell'obiettivo c_j , massimizzare $\sum_{j \in \mathcal{J}} c_j x_j$ su P fornisce un upper bound per il problema intero di massimizzazione; le disuguaglianze valide per X possono migliorare questo upper bound.

5.2 Coperture e disuguaglianze di copertura

Definizione 5.2: Copertura e insieme di tutte le coperture

Una *copertura* (cover) è un sottoinsieme $C \subseteq \mathcal{J}$ il cui coefficiente totale $\sum_{j \in C} a_j$ supera il termine noto b . L'insieme di tutte le coperture è

$$\mathcal{C} = \left\{ C \subseteq \mathcal{J} : \sum_{j \in C} a_j > b \right\}. \quad (5.4)$$

Poiché i coefficienti e il termine noto sono interi, la condizione equivale a $\sum_{j \in C} a_j \geq b+1$.

Definizione 5.3: Disuguaglianza di copertura

Data $C \in \mathcal{C}$, la *disuguaglianza di copertura* associata è

$$\sum_{j \in C} x_j \leq |C| - 1. \quad (5.5)$$

Proposizione 5.4

Per ogni copertura $C \in \mathcal{C}$, la disuguaglianza di copertura (5.5) è valida per X .

Dimostrazione. Supponiamo che un punto $\mathbf{x} \in X$ violi (5.5). La somma di $|C|$ variabili binarie è allora maggiore di $|C| - 1$, quindi $x_j = 1$ per ogni $j \in C$. Poiché tutti i coefficienti sono positivi,

$$\sum_{j \in \mathcal{J}} a_j x_j \geq \sum_{j \in C} a_j > b.$$

Questo contraddice il vincolo (5.1a): al più $|C| - 1$ variabili di C possono valere 1 insieme. $\square$

5.2.1 Coperture minimali

Definizione 5.5: Copertura minimale

Una copertura C è *minimale* se nessun sottoinsieme proprio di C è una copertura. In modo equivalente,

$$\sum_{j \in C} a_j > b, \quad \sum_{j \in C} a_j - a_i \leq b \quad \forall i \in C. \quad (5.6)$$

- Poiché i coefficienti sono positivi, basta rimuovere una variabile alla volta: se ciascuno di questi $|C|$ sottoinsiemi soddisfa il vincolo, anche ogni sottoinsieme più piccolo lo soddisfa.

Proposizione 5.6

Le disuguaglianze delle coperture minimali, insieme agli upper bound $x_j \leq 1$ per ogni $j \in \mathcal{J}$, implicano ogni disuguaglianza di copertura.

Dimostrazione. Partendo da una copertura C , rimuoviamo variabili finché l'insieme rimanente è ancora una copertura: otteniamo una copertura minimale $D \subseteq C$. Sommando la disuguaglianza di copertura di D e gli upper bound delle variabili in $C \setminus D$ otteniamo

$$\sum_{j \in C} x_j = \sum_{j \in D} x_j + \sum_{j \in C \setminus D} x_j \leq |D| - 1 + |C \setminus D| = |C| - 1. \quad \square$$

5.2.2 La chiusura delle coperture

Definizione 5.7: Chiusura delle coperture

La *chiusura delle coperture* di P è il poliedro

$$P^{\text{cov}} = \left\{ \mathbf{x} \in P : \sum_{j \in C} x_j \leq |C| - 1 \quad \forall C \in \mathcal{C} \right\}. \quad (5.7)$$

- Poiché tutte le disuguaglianze di copertura sono valide,

$$\text{conv}(X) \subseteq P^{\text{cov}} \subseteq P,$$

e in generale la prima inclusione è stretta.

- Per la Proposizione 5.6, le disuguaglianze delle coperture minimali bastano a descrivere P^{cov} .
- Il numero di coperture può essere esponenziale in n . Per esempio, se $a_j = 1$ per ogni $j \in \mathcal{J}$ e $b = 1$, ogni sottoinsieme con almeno due indici è una copertura, cioè ci sono $2^n - n - 1$ coperture. Anche le coperture minimali possono essere in numero esponenziale: per $a_j = 1$ per ogni $j \in \mathcal{J}$, $n = 2m$ e $b = m$, sono i $\binom{2m}{m+1}$ sottoinsiemi di $m + 1$ indici.
- Per questo le disuguaglianze di copertura non si aggiungono tutte insieme: si aggiungono solo quando sono violate, cioè si *separano* (Sezione 5.4).

Esempio 5.1: Un problema binario con un vincolo di questa forma

Consideriamo $\mathcal{J} = \{1, 2, 3, 4\}$, il termine noto $b = 9$, e i seguenti coefficienti dell'obiettivo c_j e coefficienti del vincolo a_j .

j	1	2	3	4
c_j	6	5	8	9
a_j	2	3	6	7

Il modello LP associato è:

$$\begin{array}{llllll}
 \max & 6x_1 & + & 5x_2 & + & 8x_3 & + & 9x_4 \\
 \text{soggetto a} & 2x_1 & + & 3x_2 & + & 6x_3 & + & 7x_4 & \leq & 9 \\
 & x_1 & & & & & & & \leq & 1 \\
 & & & x_2 & & & & & \leq & 1 \\
 & & & & & x_3 & & & \leq & 1 \\
 & & & & & & & x_4 & \leq & 1 \\
 & x_1, & & x_2, & & x_3, & & x_4 & \geq & 0
 \end{array}$$

I rapporti c_j/a_j sono $3, 5/3, 4/3, 9/7$, in ordine non crescente. La regola di Dantzig pone a 1 le variabili in questo ordine, finché il vincolo è soddisfatto, e assegna alla prima variabile che lo violerebbe il massimo valore frazionario consentito: si ottiene così una soluzione ottima di LP. Una soluzione ottima $\tilde{x}(\text{LP})$ è:

$$\tilde{x}_1 = \tilde{x}_2 = 1, \quad \tilde{x}_3 = \frac{2}{3}, \quad \tilde{x}_4 = 0.$$

Il valore ottimo $z(\text{LP})$ è $6 + 5 + 8(2/3) = 49/3$.

Esempio 5.2: La chiusura delle coperture dell'istanza a quattro variabili

L'istanza dell'Esempio 5.1 ha le seguenti sette coperture.

C	$\sum_{j \in C} a_j$	$ C $	Disuguaglianza di copertura	Minimale
$\{2, 4\}$	10	2	$x_2 + x_4 \leq 1$	sì
$\{3, 4\}$	13	2	$x_3 + x_4 \leq 1$	sì
$\{1, 2, 3\}$	11	3	$x_1 + x_2 + x_3 \leq 2$	sì
$\{1, 2, 4\}$	12	3	$x_1 + x_2 + x_4 \leq 2$	no
$\{1, 3, 4\}$	15	3	$x_1 + x_3 + x_4 \leq 2$	no
$\{2, 3, 4\}$	16	3	$x_2 + x_3 + x_4 \leq 2$	no
$\{1, 2, 3, 4\}$	18	4	$x_1 + x_2 + x_3 + x_4 \leq 3$	no

Solo le prime tre coperture sono minimali; le loro disuguaglianze descrivono già la chiusura delle coperture. Indichiamo con LP^{cov} l'LP ottenuto aggiungendole a LP. Una soluzione ottima $\tilde{x}(\text{LP}^{\text{cov}})$ è:

$$\tilde{x}_1 = 1, \quad \tilde{x}_2 = \frac{3}{5}, \quad \tilde{x}_3 = \frac{2}{5}, \quad \tilde{x}_4 = \frac{2}{5}.$$

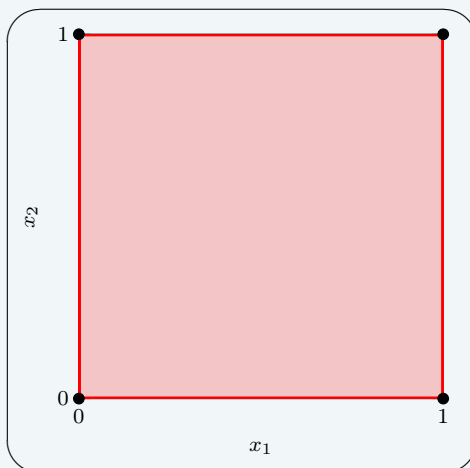
Il valore ottimo è $z(\text{LP}^{\text{cov}}) = 79/5 < z(\text{LP}) = 49/3$.

5.3 Interpretazione geometrica

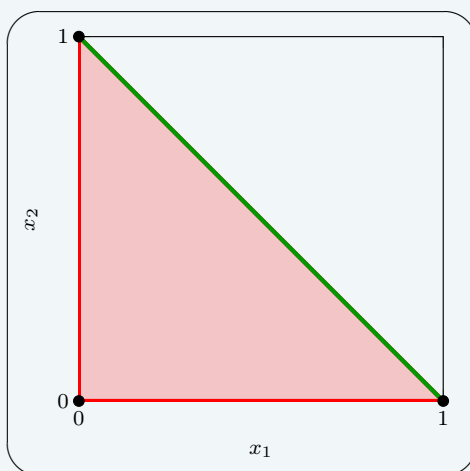
- Tutte le coordinate nelle figure sono i valori effettivi delle variabili. Le tacche indicano le coordinate del punto blu e dei vertici principali; alcuni vertici non hanno tacca, per lasciare leggibili gli assi. In 3D, gli spigoli tratteggiati sono nascosti dietro le facce visibili.
- Ogni passo è illustrato sia in due sia in tre dimensioni, con gli stessi colori in tutte le figure: la regione rossa è la regione continua considerata; il verde indica una disuguaglianza di copertura (un lato in 2D, una faccia in 3D); i punti neri sono i punti binari ammissibili; il punto blu è il punto frazionario discusso nel testo.
- Consideriamo prima solo il dominio, $[0, 1]^2$ o $[0, 1]^3$: nei due esempi che seguono non imponiamo alcun vincolo oltre ai limiti delle variabili.

Esempio 5.3: Una disuguaglianza di copertura nel quadrato unitario

In $[0, 1]^2$, la disuguaglianza $x_1 + x_2 \leq 1$ rimuove solo il punto binario $(1, 1)$ e conserva i punti binari $(0, 0)$, $(1, 0)$ e $(0, 1)$.



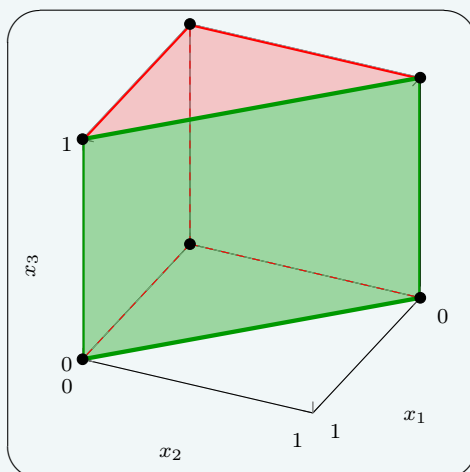
Quadrato unitario: solo il dominio.



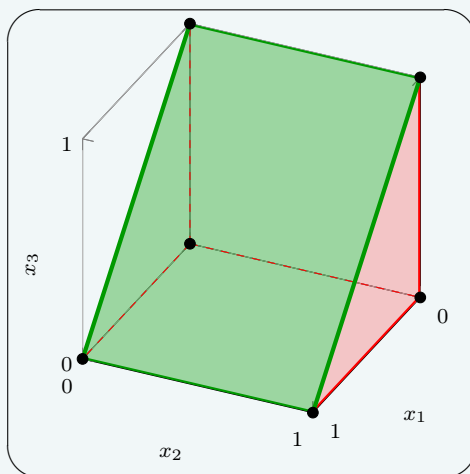
Dominio con $x_1 + x_2 \leq 1$.

Esempio 5.4: Disuguaglianze di copertura nel cubo unitario

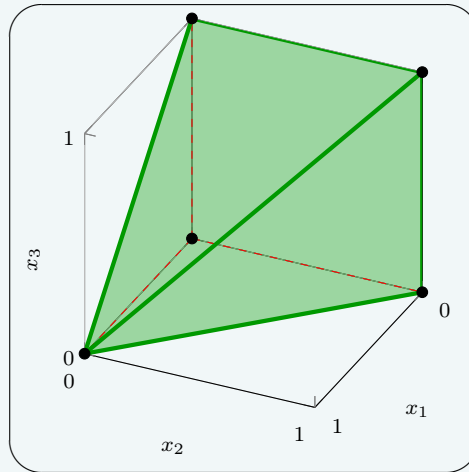
- In $[0, 1]^3$, la disuguaglianza $x_1 + x_2 \leq 1$ non dipende da x_3 ; la sua frontiera è parallela all'asse x_3 . Allo stesso modo, la frontiera di $x_1 + x_3 \leq 1$ è parallela all'asse x_2 .
- I due semispazi rimuovono porzioni diverse del cubo. La loro intersezione conserva esattamente i punti binari $(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (0, 1, 1)$.
- La disuguaglianza $x_1 + x_2 + x_3 \leq 2$ rimuove solo il punto binario $x_1 = x_2 = x_3 = 1$.



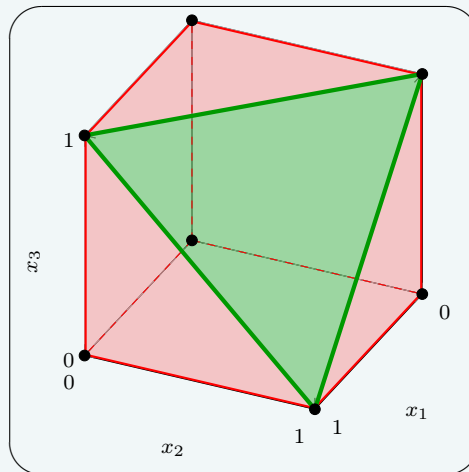
Dominio con $x_1 + x_2 \leq 1$.



Dominio con $x_1 + x_3 \leq 1$.



Dominio con entrambe le disuguaglianze.


 Dominio con $x_1 + x_2 + x_3 \leq 2$.

Esempio 5.5: Una copertura in due dimensioni

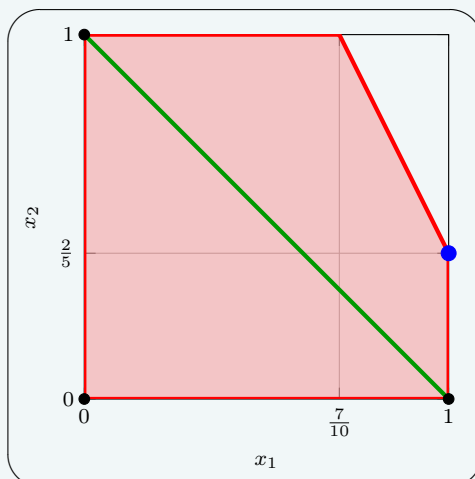
Per $\mathcal{J} = \{1, 2\}$, consideriamo

$$P_2 = \{x \in [0, 1]^2 : 10x_1 + 5x_2 \leq 12\}.$$

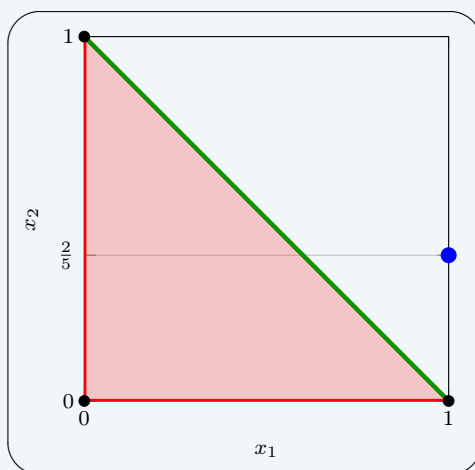
La copertura $C = \{1, 2\}$ dà $x_1 + x_2 \leq 1$. Il punto $\tilde{x}_1 = 1$, $\tilde{x}_2 = 2/5$ soddisfa il vincolo all'uguaglianza, ma viola la disuguaglianza di copertura di $2/5$. Il poligono originale ha vertici

$$(0, 0), (1, 0), (1, 2/5), (7/10, 1), (0, 1).$$

Dopo l'aggiunta del taglio, la regione è il triangolo di vertici $(0, 0), (1, 0), (0, 1)$. Quindi tutti e tre i punti binari ammissibili sono conservati, e il vincolo diventa ridondante, poiché $10x_1 + 5x_2 \leq 10 < 12$ sul triangolo.



P_2 originale e la frontiera $x_1 + x_2 = 1$.



$P_2 \cap \{x_1 + x_2 \leq 1\}$: il punto blu è rimosso.

Esempio 5.6: Tagli di copertura successivi

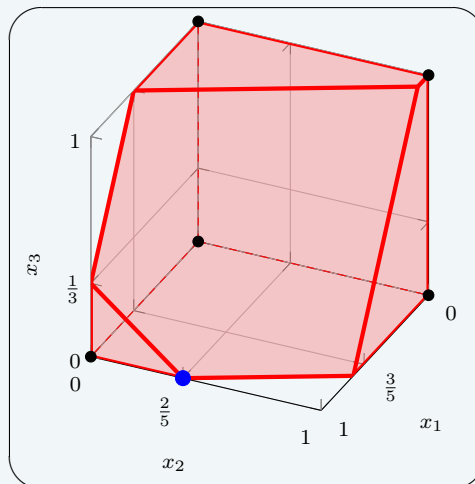
Per $\mathcal{J} = \{1, 2, 3\}$, sia

$$P_A = \{\mathbf{x} \in [0, 1]^3 : 10x_1 + 5x_2 + 6x_3 \leq 12\}.$$

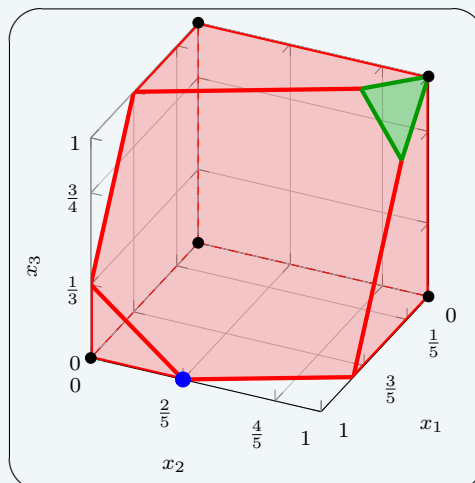
Ci sono esattamente tre coperture:

$$\begin{aligned} \{1, 2, 3\} &: x_1 + x_2 + x_3 \leq 2, \\ \{1, 3\} &: x_1 + x_3 \leq 1, \\ \{1, 2\} &: x_1 + x_2 \leq 1. \end{aligned}$$

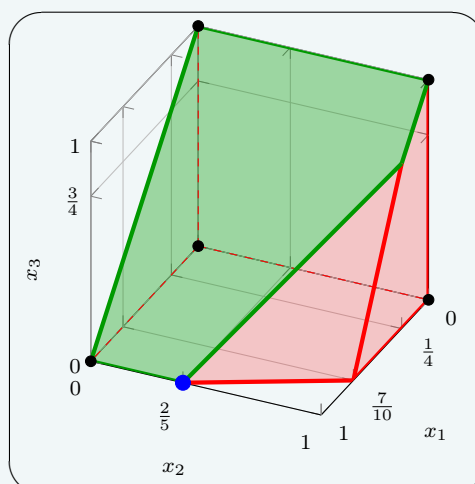
Solo le ultime due sono minimali. Il punto blu $\tilde{x}_1 = 1$, $\tilde{x}_2 = 2/5$, $\tilde{x}_3 = 0$ appartiene a P_A . Soddisfa la prima disuguaglianza ed è sulla frontiera della seconda; la terza è violata di $2/5$, esattamente come nell'Esempio 5.5. A differenza dei riquadri precedenti, che mostravano solo il dominio, ogni riquadro che segue include il vincolo $10x_1 + 5x_2 + 6x_3 \leq 12$.



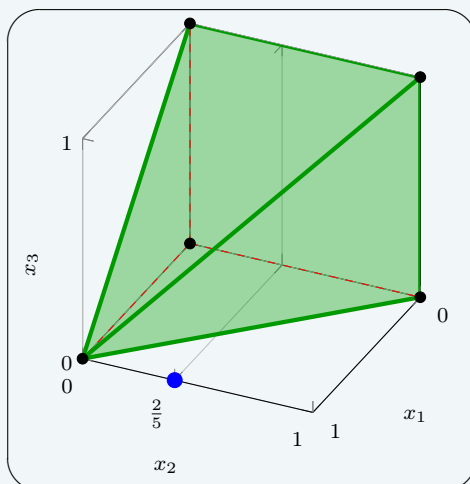
P_A originale.



Aggiungiamo $x_1 + x_2 + x_3 \leq 2$.



Aggiungiamo anche $x_1 + x_3 \leq 1$.



Aggiungiamo anche $x_1 + x_2 \leq 1$: la chiusura delle coperture.

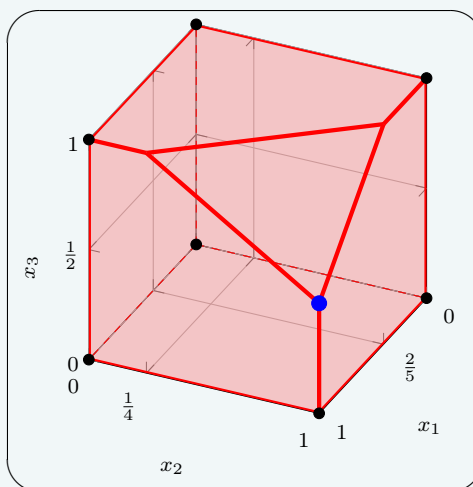
- Per l'obiettivo $\max 40x_1 + 15x_2 + 12x_3$, la soluzione ottima su P_A è $\tilde{x} = (1, 2/5, 0)$, di valore 46. Sulla chiusura delle coperture, la soluzione ottima è $(1, 0, 0)$, di valore 40, che è ottima anche per il problema binario.
- In questa istanza la chiusura delle coperture coincide con $\text{conv}(X)$: i suoi vertici, mostrati nell'ultimo riquadro, sono tutti binari. È una proprietà di questa istanza, non di ogni chiusura delle coperture.

Esempio 5.7: Un vincolo diverso

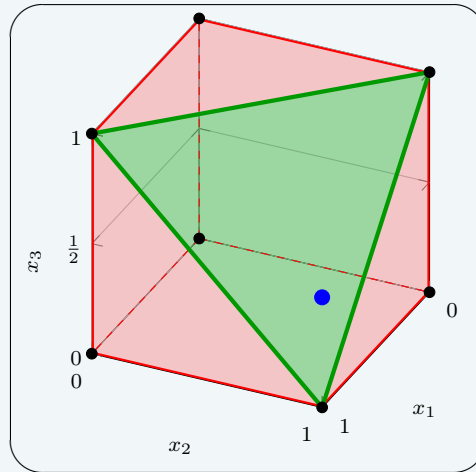
Per $\mathcal{I} = \{1, 2, 3\}$, consideriamo ora l'istanza

$$P_B = \{x \in [0, 1]^3 : 5x_1 + 4x_2 + 6x_3 \leq 12\}.$$

Ogni coppia di variabili soddisfa il vincolo, mentre i tre coefficienti hanno somma $15 > 12$. Quindi l'unica copertura è $\{1, 2, 3\}$, ed è minimale. La sua disuguaglianza è $x_1 + x_2 + x_3 \leq 2$. Il punto blu $\tilde{x}_1 = \tilde{x}_2 = 1, \tilde{x}_3 = 1/2$ ha membro sinistro 12 e viola questo taglio di $1/2$.



P_B originale.



$$P_B \cap \{x_1 + x_2 + x_3 \leq 2\}.$$

- Il piano del vincolo interseca gli spigoli del cubo in $(1, 1, 1/2)$, $(1, 1/4, 1)$ e $(2/5, 1, 1)$.
- La disuguaglianza di copertura rimuove questi vertici frazionari e conserva tutti e sette i punti binari ammissibili. Come nell'esempio in 2D, dopo il taglio il vincolo è ridondante: la somma dei due coefficienti più grandi è $6 + 5 = 11 < 12$.
- Il poliedro $P_B \cap \{x_1 + x_2 + x_3 \leq 2\}$ ha solo vertici binari (ultimo riquadro): qui una sola disuguaglianza di copertura dà $\text{conv}(X)$.

5.4 Separazione delle disuguaglianze di copertura

Dato un punto $\tilde{x} \in P$, il **problema di separazione delle coperture** (CSP, cover separation problem) consiste nel trovare una disuguaglianza di copertura violata da $\tilde{x}$, o nel dimostrare che $\tilde{x}$ soddisfa tutte le disuguaglianze di copertura.

Formalmente, il CSP chiede di trovare una copertura

$$C \in \mathcal{C} : \quad \sum_{j \in C} \tilde{x}_j > |C| - 1,$$

o di dimostrare che

$$\sum_{j \in C} \tilde{x}_j \leq |C| - 1 \quad \forall C \in \mathcal{C}.$$

- Di solito $\tilde{x}$ è una soluzione ottima del rilassamento lineare corrente, ma si può separare qualunque punto di P .
- Per una copertura C , la disuguaglianza di copertura è violata da $\tilde{x}$ se

$$\sum_{j \in C} \tilde{x}_j > |C| - 1. \quad (5.8)$$

Portando $\sum_{j \in C} \tilde{x}_j$ al membro destro e 1 al membro sinistro, e scrivendo $|C| = \sum_{j \in C} 1$, otteniamo le condizioni equivalenti:

$$\begin{aligned} |C| - \sum_{j \in C} \tilde{x}_j &< 1, \\ \sum_{j \in C} 1 - \sum_{j \in C} \tilde{x}_j &< 1, \\ \sum_{j \in C} (1 - \tilde{x}_j) &< 1. \end{aligned} \tag{5.9}$$

- Assegnando a ogni indice j il costo $q_j = 1 - \tilde{x}_j \geq 0$, il CSP consiste nel trovare una copertura di *costo minimo*: esiste una disuguaglianza di copertura violata se e solo se questo costo minimo è minore di 1.

5.4.1 Una formulazione di PLI per il CSP

- Introduciamo n variabili binarie z_j : $z_j = 1$ se j appartiene alla copertura e $z_j = 0$ altrimenti.
- L'obiettivo minimizza il costo totale della copertura; l'unico vincolo impone che l'insieme scelto sia una copertura.

$$(\text{ILP}_{\text{CSP}}) \quad \min \sum_{j \in \mathcal{J}} (1 - \tilde{x}_j) z_j \tag{5.10a}$$

$$\text{soggetto a} \quad \sum_{j \in \mathcal{J}} a_j z_j \geq b + 1, \tag{5.10b}$$

$$z_j \in \{0, 1\}, \quad \forall j \in \mathcal{J}. \tag{5.10c}$$

Proposizione 5.8

Esiste una disuguaglianza di copertura violata da $\tilde{\mathbf{x}}$ se e solo se $z(\text{ILP}_{\text{CSP}}) < 1$.

Dimostrazione. Ogni assegnamento ammissibile delle variabili binarie z_j rappresenta una copertura $C = \{j : z_j = 1\}$, e il suo valore dell'obiettivo è $\sum_{j \in C} (1 - \tilde{x}_j)$. Per la (5.9), questo valore è minore di 1 esattamente quando il taglio corrispondente è violato. Prendendo il minimo su tutte le coperture si dimostrano entrambe le implicazioni. Il CSP è ammissibile perché $\mathcal{J}$ stesso è una copertura. $\square$

Sia $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_n$ una soluzione ottima di ILP_{CSP} . Se $z(\text{ILP}_{\text{CSP}}) < 1$, la copertura

$$\tilde{C} = \{j \in \mathcal{J} : \tilde{z}_j = 1\}$$

è associata alla disuguaglianza di copertura $\sum_{j \in \tilde{C}} x_j \leq |\tilde{C}| - 1$ massimamente violata da $\tilde{\mathbf{x}}$. Altrimenti, nessuna disuguaglianza di copertura è violata da $\tilde{\mathbf{x}}$.

- Per trovare una disuguaglianza di copertura violata non è necessario risolvere ILP_{CSP} all'ottimo: qualunque soluzione ammissibile $\bar{z}_1, \dots, \bar{z}_n$ di costo minore di 1 fornisce la copertura $\bar{C} = \{j \in \mathcal{J} : \bar{z}_j = 1\}$, la cui disuguaglianza di copertura è violata da $\tilde{x}$.
- Al contrario, una soluzione ammissibile di costo almeno 1 *non* dimostra che non esistano disuguaglianze di copertura violate: per questo serve il valore ottimo (o un lower bound) di ILP_{CSP} .
- Se $z(\text{ILP}_{\text{CSP}}) < 1$, la massima violazione di una disuguaglianza di copertura in $\tilde{x}$ è $1 - z(\text{ILP}_{\text{CSP}})$.

Esempio 5.8: Separazione delle disuguaglianze di copertura

Riprendiamo l'Esempio 5.1. Le disuguaglianze di copertura si separano in modo iterativo: a ogni iterazione si risolve l'LP corrente, si risolve ILP_{CSP} per la sua soluzione ottima $\tilde{x}$ e si aggiunge la disuguaglianza di copertura massimamente violata, finché nessuna disuguaglianza di copertura è violata.

Iterazione 1. Una soluzione ottima di LP è

$$\tilde{x}_1 = 1, \quad \tilde{x}_2 = 1, \quad \tilde{x}_3 = \frac{2}{3}, \quad \tilde{x}_4 = 0, \quad z(\text{LP}) = \frac{49}{3}.$$

Il ILP_{CSP} associato è:

$$\begin{aligned} \min \quad & \frac{1}{3} z_3 + z_4 \\ \text{soggetto a} \quad & 2z_1 + 3z_2 + 6z_3 + 7z_4 \geq 10 \\ & z_1, z_2, z_3, z_4 \in \{0, 1\} \end{aligned}$$

Una soluzione ottima di ILP_{CSP} è $\tilde{z}_1 = \tilde{z}_2 = \tilde{z}_3 = 1, \tilde{z}_4 = 0$, di valore $z(\text{ILP}_{\text{CSP}}) = \frac{1}{3} < 1$. Di conseguenza, $\tilde{C} = \{j \in \mathcal{J} : \tilde{z}_j = 1\} = \{1, 2, 3\}$ è la copertura associata alla disuguaglianza di copertura massimamente violata da $\tilde{x}$:

$$x_1 + x_2 + x_3 \leq 2, \quad \text{poiché} \quad \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 = 1 + 1 + \frac{2}{3} = \frac{8}{3} > 2.$$

Iterazione 2. Aggiungendo questa disuguaglianza di copertura a LP, otteniamo il seguente LP, indicato con LP_1 :

$$\begin{aligned} \max \quad & 6x_1 + 5x_2 + 8x_3 + 9x_4 \\ \text{soggetto a} \quad & 2x_1 + 3x_2 + 6x_3 + 7x_4 \leq 9 \\ & x_1 + x_2 + x_3 \leq 2 \\ & x_1 \leq 1 \\ & x_2 \leq 1 \\ & x_3 \leq 1 \\ & x_4 \leq 1 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

Una soluzione ottima di LP_1 è

$$\tilde{x}_1 = 1, \quad \tilde{x}_2 = 1, \quad \tilde{x}_3 = 0, \quad \tilde{x}_4 = \frac{4}{7}, \quad z(\text{LP}_1) = \frac{113}{7}.$$

Il ILP_{CSP} associato è:

$$\begin{array}{ll} \min & z_3 + \frac{3}{7} z_4 \\ \text{soggetto a} & 2 z_1 + 3 z_2 + 6 z_3 + 7 z_4 \geq 10 \\ & z_1, z_2, z_3, z_4 \in \{0, 1\} \end{array}$$

Una soluzione ottima di ILP_{CSP} è $\tilde{z}_1 = 0, \tilde{z}_2 = 1, \tilde{z}_3 = 0, \tilde{z}_4 = 1$, di valore $z(\text{ILP}_{\text{CSP}}) = \frac{3}{7} < 1$ (la copertura $\{1, 2, 4\}$ ha lo stesso costo; scegliamo la copertura minimale). Di conseguenza, $\tilde{C} = \{j \in \mathcal{J} : \tilde{z}_j = 1\} = \{2, 4\}$ e la disuguaglianza di copertura massimamente violata è

$$x_2 + x_4 \leq 1, \quad \text{poiché} \quad \tilde{x}_2 + \tilde{x}_4 = 1 + \frac{4}{7} = \frac{11}{7} > 1.$$

Iterazione 3. Aggiungendo questa disuguaglianza di copertura a LP_1 , otteniamo il seguente LP, indicato con LP_2 :

$$\begin{array}{ll} \max & 6 x_1 + 5 x_2 + 8 x_3 + 9 x_4 \\ \text{soggetto a} & 2 x_1 + 3 x_2 + 6 x_3 + 7 x_4 \leq 9 \\ & x_1 + x_2 + x_3 \leq 2 \\ & x_2 + x_4 \leq 1 \\ & x_1 \leq 1 \\ & x_2 \leq 1 \\ & x_3 \leq 1 \\ & x_4 \leq 1 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{array}$$

Una soluzione ottima di LP_2 è

$$\tilde{x}_1 = 1, \quad \tilde{x}_2 = \frac{3}{5}, \quad \tilde{x}_3 = \frac{2}{5}, \quad \tilde{x}_4 = \frac{2}{5}, \quad z(\text{LP}_2) = \frac{79}{5}.$$

Il ILP_{CSP} associato è:

$$\begin{array}{ll} \min & \frac{2}{5} z_2 + \frac{3}{5} z_3 + \frac{3}{5} z_4 \\ \text{soggetto a} & 2 z_1 + 3 z_2 + 6 z_3 + 7 z_4 \geq 10 \\ & z_1, z_2, z_3, z_4 \in \{0, 1\} \end{array}$$

Una soluzione ottima di ILP_{CSP} è $\tilde{z}_1 = 0, \tilde{z}_2 = 1, \tilde{z}_3 = 0, \tilde{z}_4 = 1$, di valore $z(\text{ILP}_{\text{CSP}}) = 1$ (le coperture $\{1, 2, 3\}$ e $\{1, 2, 4\}$ hanno lo stesso costo): **nessuna** disuguaglianza di copertura è violata da $\tilde{x}$, e la separazione si ferma.

Il valore finale $z(\text{LP}_2) = \frac{79}{5}$ coincide con il valore $z(\text{LP}^{\text{cov}})$ della chiusura delle coperture (Esempio 5.2): la separazione raggiunge il bound della chiusura delle coperture aggiungendo solo 2 delle 7 disuguaglianze di copertura.

La soluzione $\tilde{x}$ è ancora **frazionaria**, anche se soddisfa tutte le disuguaglianze di copertura: la chiusura delle coperture non coincide, in generale, con l'involuppo convesso delle soluzioni intere.

5.5 Dal CSP a un problema complementare

- Consideriamo un assegnamento ammissibile delle variabili z_j del CSP. Definiamo le variabili binarie complementari y_j ponendo

$$y_j = 1 - z_j \quad \forall j \in \mathcal{J}.$$

Quindi $y_j = 1$ se j è fuori dalla copertura, mentre $y_j = 0$ se j appartiene alla copertura.

- Il vincolo diventa

$$\sum_{j \in \mathcal{J}} a_j(1 - y_j) \geq b + 1 \iff \sum_{j \in \mathcal{J}} a_j y_j \leq \sum_{j \in \mathcal{J}} a_j - (b + 1).$$

Per l'ipotesi $\sum_{j \in \mathcal{J}} a_j > b$, questo termine noto è non negativo.

- L'obiettivo si può scrivere come

$$\sum_{j \in \mathcal{J}} (1 - \tilde{x}_j)(1 - y_j) = n - \sum_{j \in \mathcal{J}} \tilde{x}_j - \sum_{j \in \mathcal{J}} (1 - \tilde{x}_j)y_j.$$

Poiché i primi due termini sono costanti, minimizzarlo equivale al problema seguente, detto *problema complementare*, con coefficienti dell'obiettivo non negativi (eventualmente nulli) $1 - \tilde{x}_j$.

$$(\text{ILP}_{\text{CSP}}^2) \quad \max \quad \sum_{j \in \mathcal{J}} (1 - \tilde{x}_j)y_j \quad (5.11a)$$

$$\text{soggetto a} \quad \sum_{j \in \mathcal{J}} a_j y_j \leq \sum_{j \in \mathcal{J}} a_j - (b + 1), \quad (5.11b)$$

$$y_j \in \{0, 1\}, \quad \forall j \in \mathcal{J}. \quad (5.11c)$$

- La complementazione è una corrispondenza biunivoca tra le soluzioni ammissibili, quindi

$$z(\text{ILP}_{\text{CSP}}) = n - \sum_{j \in \mathcal{J}} \tilde{x}_j - z(\text{ILP}_{\text{CSP}}^2). \quad (5.12)$$

Sia $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n$ una soluzione ottima di $\text{ILP}_{\text{CSP}}^2$. Se $n - \sum_{j \in \mathcal{J}} \tilde{x}_j - z(\text{ILP}_{\text{CSP}}^2) < 1$, la copertura

$$\tilde{C} = \{j \in \mathcal{J} : \tilde{y}_j = 0\}$$

è associata alla disuguaglianza di copertura $\sum_{j \in \tilde{C}} x_j \leq |\tilde{C}| - 1$ massimamente violata da $\tilde{\mathbf{x}}$. Altrimenti, nessuna disuguaglianza di copertura è violata da $\tilde{\mathbf{x}}$.

- Qualunque soluzione ammissibile $\bar{y}_1, \dots, \bar{y}_n$ di $\text{ILP}_{\text{CSP}}^2$ fornisce la copertura $\bar{C} = \{j \in \mathcal{J} : \bar{y}_j = 0\}$, di costo $n - \sum_{j \in \mathcal{J}} \tilde{x}_j - \sum_{j \in \mathcal{J}} (1 - \tilde{x}_j)\bar{y}_j$: se questo costo è minore di 1, la disuguaglianza di copertura di $\bar{C}$ è violata da $\tilde{\mathbf{x}}$.

5.5.1 Un'euristica greedy di separazione

La seguente procedura greedy costruisce valori ammissibili $\bar{y}_j$ per il problema complementare.

1. Porre $\bar{y}_j = 0$ per ogni $j \in \mathcal{J}$ e il termine noto residuo $r = \sum_{j \in \mathcal{J}} a_j - (b + 1)$. Ordinare gli indici per rapporto $q_j/a_j = (1 - \tilde{x}_j)/a_j$ non crescente, risolvendo i pareggi per indice crescente.
 2. Scorrere gli indici in questo ordine. Se $q_j > 0$ e $a_j \leq r$, porre $\bar{y}_j = 1$ e $r := r - a_j$; altrimenti saltare l'indice.
 3. Ricostruire la copertura $C = \{j : \bar{y}_j = 0\}$ e valutarne il costo. Se lo si desidera, ridurla a una copertura minimale rimuovendo indici finché $\sum_{j \in C} a_j$ supera b .
- Gli indici con $q_j = 0$ ($\tilde{x}_j = 1$) non vengono mai selezionati: restano nella copertura, che quindi può non essere minimale.
 - Rimuovere un indice da una copertura ne diminuisce il costo di $q_j \geq 0$: ridurre una copertura violata a una minimale non può diminuirne la violazione.
 - L'algoritmo greedy è un'euristica per il problema complementare: se non trova una disuguaglianza di copertura violata, non possiamo concludere che non ne esistano.

Esempio 5.9: Problema complementare e scansione greedy

Per il punto $\tilde{x} = (1, 1, \frac{2}{3}, 0)$ dell'Iterazione 1 dell'Esempio 5.8,

$$4 - \sum_{j \in \mathcal{J}} \tilde{x}_j = 4 - 1 - 1 - \frac{2}{3} = \frac{4}{3},$$

$$\sum_{j \in \mathcal{J}} a_j - (b + 1) = 2 + 3 + 6 + 7 - (9 + 1) = 8.$$

Il problema complementare è

$$\begin{array}{llllll} \max & & & \frac{1}{3} y_3 & + & y_4 \\ \text{soggetto a} & 2 y_1 & + & 3 y_2 & + & 6 y_3 & + & 7 y_4 & \leq & 8 \\ & y_1, & & y_2, & & y_3, & & y_4 & \in & \{0, 1\} \end{array}$$

j	q_j	a_j	q_j/a_j	Decisione	Residuo
4	1	7	$\frac{1}{7}$	$\bar{y}_j = 1$	1
3	$\frac{1}{3}$	6	$\frac{1}{18}$	$\bar{y}_j = 0$	1
1	0	2	0	$q_j = 0, \bar{y}_j = 0$	1
2	0	3	0	$q_j = 0, \bar{y}_j = 0$	1

La scansione dà $\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = 0$ e $\bar{y}_4 = 1$, di valore 1. Gli indici 3 e 4 non possono essere presi insieme; le uniche alternative con coefficiente dell'obiettivo positivo hanno valori $1/3$ e 1. Quindi questa soluzione greedy è anche ottima, e

$$z(\text{ILP}_{\text{CSP}}) = 4 - \sum_{j \in \mathcal{J}} \tilde{x}_j - z(\text{ILP}_{\text{CSP}}^2) = \frac{4}{3} - 1 = \frac{1}{3}.$$

La copertura è $\tilde{C} = \{j \in \mathcal{J} : \bar{y}_j = 0\} = \{1, 2, 3\}$, come nell'Iterazione 1 dell'Esempio 5.8.

5.6 Disuguaglianze di copertura estese

- Una disuguaglianza di copertura si può rafforzare considerando anche le variabili fuori dalla copertura i cui coefficienti sono grandi almeno quanto ogni coefficiente della copertura. Sia

$$\hat{a}(C) = \max_{j \in C} a_j$$

il massimo coefficiente delle variabili in C .

Definizione 5.9: Disuguaglianza di copertura estesa

Per una copertura C , sia

$$E(C) = C \cup \{j \in \mathcal{J} \setminus C : a_j \geq \hat{a}(C)\}.$$

La *disuguaglianza di copertura estesa* associata è

$$\sum_{j \in E(C)} x_j \leq |C| - 1. \quad (5.13)$$

Il termine noto dipende da $|C|$, non da $|E(C)|$.

Proposizione 5.10

Per ogni copertura $C \in \mathcal{C}$, la disuguaglianza di copertura estesa (5.13) è valida per X .

Dimostrazione. Sia $k = |C|$ e prendiamo un qualunque insieme $S \subseteq E(C)$ con $|S| = k$. Gli insiemi $S \setminus C$ e $C \setminus S$ hanno lo stesso numero di variabili, e ogni coefficiente a_j con $j \in S \setminus C$ è almeno $\hat{a}(C)$, cioè grande almeno quanto ogni coefficiente con $j \in C \setminus S$. Sostituendo gli indici di $C \setminus S$ con quelli di $S \setminus C$ otteniamo

$$\sum_{j \in S} a_j \geq \sum_{j \in C} a_j > b.$$

Quindi nessun punto $x \in X$ può avere k variabili di $E(C)$ uguali a 1: al più $k - 1$ di esse possono valere 1 insieme. $\square$

- Poiché $C \subseteq E(C)$ e il termine noto è lo stesso, la disuguaglianza di copertura estesa domina la disuguaglianza di copertura di C .
- Si può vedere come un *lifting* (sollevamento) dei coefficienti delle variabili in $E(C) \setminus C$ da 0 a 1. Non vieta a una variabile con un coefficiente grande di valere 1 insieme a variabili di C : limita a $|C| - 1$ il numero di variabili dell'intero insieme $E(C)$ uguali a 1.

Esempio 5.10: Rafforzare la copertura dell'istanza a quattro variabili

Nell'Esempio 5.1, consideriamo la copertura $C = \{1, 2, 3\}$. Allora $\hat{a}(C) = 6$, e il coefficiente di x_4 è $7 \geq 6$, quindi

$$E(C) = \{1, 2, 3, 4\}, \quad x_1 + x_2 + x_3 + x_4 \leq 2.$$

L'LP rafforzato, indicato con LP_{ext} , è

$$\begin{array}{llllll}
 \max & 6x_1 & + & 5x_2 & + & 8x_3 & + & 9x_4 \\
 \text{soggetto a} & 2x_1 & + & 3x_2 & + & 6x_3 & + & 7x_4 & \leq & 9 \\
 & x_1 & + & x_2 & + & x_3 & + & x_4 & \leq & 2 \\
 & x_1 & & & & & & & \leq & 1 \\
 & & & x_2 & & & & & \leq & 1 \\
 & & & & & x_3 & & & \leq & 1 \\
 & & & & & & & x_4 & \leq & 1 \\
 & x_1, & & x_2, & & x_3, & & x_4 & \geq & 0
 \end{array}$$

Una soluzione ottima di LP_{ext} è $\tilde{\mathbf{x}} = (1, 0, 0, 1)$, di valore $z(LP_{\text{ext}}) = 6 + 9 = 15$. Poiché è binaria, $\tilde{\mathbf{x}}$ è ottima anche per il problema binario. Abbiamo $z(LP_{\text{ext}}) = 15 < z(LP^{\text{cov}}) = 79/5$: in questa istanza, una sola disuguaglianza di copertura estesa è più forte di tutte le disuguaglianze di copertura insieme.

5.7 Violazione e distanza da una disuguaglianza di copertura

- Sia $C \subseteq \mathcal{J}$ una copertura. La sua disuguaglianza di copertura è

$$\sum_{j \in C} x_j \leq |C| - 1,$$

e il suo iperpiano di frontiera è

$$H_C = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{j \in C} x_j = |C| - 1 \right\}.$$

- La violazione della disuguaglianza di copertura in un punto $\tilde{\mathbf{x}}$ è

$$v_C(\tilde{\mathbf{x}}) = \max \left\{ 0, \sum_{j \in C} \tilde{x}_j - (|C| - 1) \right\}. \quad (5.14)$$

Il punto viola la disuguaglianza di copertura esattamente quando $v_C(\tilde{\mathbf{x}}) > 0$.

- I coefficienti della disuguaglianza di copertura sono 1 per $j \in C$ e 0 altrimenti, quindi la loro norma euclidea è $\sqrt{|C|}$. Di conseguenza, la distanza euclidea di $\tilde{\mathbf{x}}$ da H_C è

$$d(\tilde{\mathbf{x}}, H_C) = \frac{\left| \sum_{j \in C} \tilde{x}_j - (|C| - 1) \right|}{\sqrt{|C|}}. \quad (5.15)$$

Se la disuguaglianza di copertura è violata,

$$d(\tilde{\mathbf{x}}, H_C) = \frac{v_C(\tilde{\mathbf{x}})}{\sqrt{|C|}}.$$

- La proiezione ortogonale $\mathbf{x}^\perp$ di $\tilde{\mathbf{x}}$ su H_C ha componenti

$$x_j^\perp = \begin{cases} \tilde{x}_j - \frac{\sum_{k \in C} \tilde{x}_k - (|C| - 1)}{|C|}, & j \in C, \\ \tilde{x}_j, & j \in \mathcal{J} \setminus C. \end{cases} \quad (5.16)$$

Infatti, sostituendo si ottiene $\sum_{j \in C} x_j^\perp = |C| - 1$.

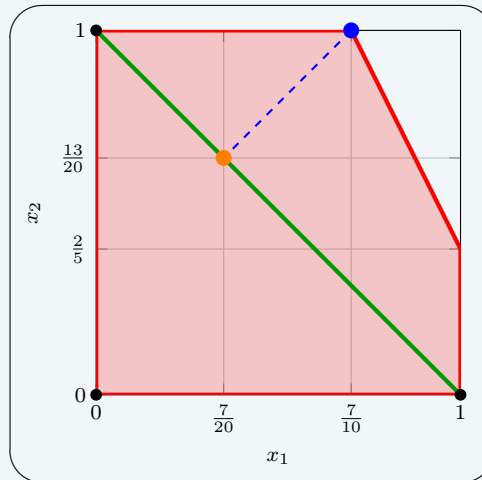
- Il CSP massimizza la violazione, non la distanza: quando le coperture hanno cardinalità diverse, i due criteri possono ordinarle in modo diverso. Per esempio, nel punto $\tilde{\mathbf{x}} = (1/10, 1, 1)$ di P_A (Esempio 5.6), le coperture $\{1, 2\}$, $\{1, 3\}$ e $\{1, 2, 3\}$ hanno tutte violazione $1/10$, ma le loro distanze sono rispettivamente $1/(10\sqrt{2})$, $1/(10\sqrt{2})$ e $1/(10\sqrt{3})$.

Esempio 5.11: Proiezione ortogonale in due dimensioni

Prendiamo il vertice $\tilde{x}_1 = 7/10$, $\tilde{x}_2 = 1$ di P_2 nell'Esempio 5.5 e la copertura $C = \{1, 2\}$. La sua disuguaglianza di copertura è $x_1 + x_2 \leq 1$ e $|C| = 2$. Allora

$$v_C(\tilde{\mathbf{x}}) = \frac{7}{10}, \quad x_1^\perp = \frac{7}{20}, \quad x_2^\perp = \frac{13}{20}, \quad d(\tilde{\mathbf{x}}, H_C) = \frac{7}{10\sqrt{2}}.$$

La regione rossa è P_2 e il lato verde è la frontiera H_C . Il segmento blu tratteggiato è perpendicolare a H_C e il punto arancione è la proiezione ortogonale.

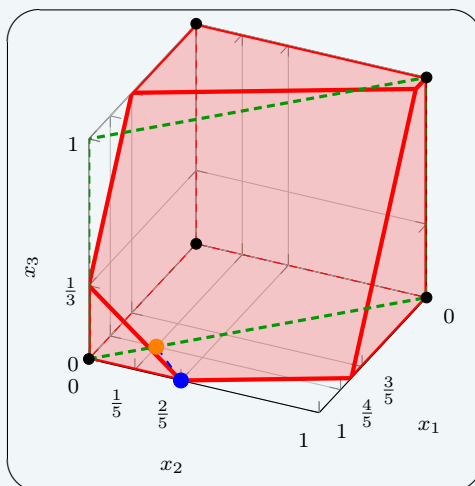


Esempio 5.12: Proiezione ortogonale in tre dimensioni

Prendiamo il punto $\tilde{x}_1 = 1, \tilde{x}_2 = 2/5, \tilde{x}_3 = 0$ di P_A nell'Esempio 5.6 e la copertura $C = \{1, 2\}$. La sua disuguaglianza di copertura è $x_1 + x_2 \leq 1$ e $|C| = 2$. Allora

$$v_C(\tilde{x}) = \frac{2}{5}, \quad x_1^\perp = \frac{4}{5}, \quad x_2^\perp = \frac{1}{5}, \quad x_3^\perp = 0, \quad d(\tilde{x}, H_C) = \frac{2}{5\sqrt{2}}.$$

La regione rossa è P_A e il rettangolo verde tratteggiato è H_C ristretto al cubo unitario. Poiché $3 \notin C$, la proiezione non modifica $\tilde{x}_3$.



5.8 Errori tipici

Attenzione

- **Usare un insieme che non è una copertura.** La disuguaglianza $\sum_{j \in C} x_j \leq |C| - 1$ è valida solo se il coefficiente totale $\sum_{j \in C} a_j$ supera il termine noto b ; controllare i coefficienti prima di scriverla.
- **Usare $|E(C)| - 1$ per una copertura estesa.** Il termine noto della disuguaglianza di copertura estesa è $|C| - 1$, dove C è la copertura, non $|E(C)| - 1$.
- **Rimuovere da una copertura tutti gli indici con $q_j = 0$.** Per ottenere una copertura minimale, si rimuovono gli indici uno alla volta finché l'insieme rimanente è ancora una copertura: rimuoverli tutti può lasciare un insieme che non è una copertura.
- **Fermare la separazione a un fallimento dell'euristica.** Se l'algoritmo euristico greedy non trova disuguaglianze di copertura violate, non possiamo concludere che non ne esistano: lo dimostra solo il valore ottimo del problema di separazione.
- **Confondere violazione e distanza.** Il problema di separazione delle coperture massimizza la violazione; coperture con la stessa violazione possono avere distanze diverse dal punto.

- **Credere che la chiusura delle coperture sia l'involuppo convesso.** Un punto può soddisfare tutte le disuguaglianze di copertura ed essere comunque frazionario (Esempio 5.8).

5.9 In sintesi

In sintesi

- Una copertura è un insieme di variabili il cui coefficiente totale supera il termine noto; la sua disuguaglianza di copertura consente che al più $|C| - 1$ delle sue variabili valgano 1, ed è valida per X .
- Le coperture minimali bastano: le loro disuguaglianze implicano tutte le altre, ma possono essere in numero esponenziale, quindi si separano solo quando sono violate.
- Separare una disuguaglianza di copertura significa trovare una copertura di costo minimo $\sum_{j \in C} (1 - \tilde{x}_j)$: esiste una disuguaglianza violata se e solo se il costo minimo è minore di 1; è un problema 0–1 con un solo vincolo $\leq$, definito nel capitolo.
- Una disuguaglianza di copertura estesa domina la disuguaglianza di copertura della stessa copertura.
- La chiusura delle coperture è più vicina a $\text{conv}(X)$ del rilassamento lineare, ma in generale non coincide con esso.

Cruscotto dell'istanza a quattro variabili. I lower bound e gli upper bound sul valore ottimo del problema binario dell'Esempio 5.1. I coefficienti dell'obiettivo sono interi, quindi l'upper bound è il valore del rilassamento lineare arrotondato per difetto.

Fonte	Lower bound	Upper bound	Da
Rilassamento lineare	11	16	soluzione greedy $(1, 1, 0, 0)$, $\lfloor z(\text{LP}) \rfloor$ con $z(\text{LP}) = \frac{49}{3}$
Chiusura delle coperture	11	15	$\lfloor z(\text{LP}^{\text{cov}}) \rfloor$ con $z(\text{LP}^{\text{cov}}) = \frac{79}{5}$, 2 coperture separate
Copertura estesa	15	15	$z(\text{LP}_{\text{ext}}) = 15$, soluzione ottima $(1, 0, 0, 1)$