Complementi di Algoritmi e Strutture Dati

Il teorema di dualità forte

Docente: *Nicolò Cesa-Bianchi* versione 1 giugno 2017

Ricordiamo la formulazione del problema di programmazione lineare nella sua forma primale e duale,

$$\begin{array}{llcll} \max & \boldsymbol{c}^\top \boldsymbol{x} & \qquad & \min & \boldsymbol{b}^\top \boldsymbol{y} \\ & A\boldsymbol{x} \le \boldsymbol{b} & & & A^\top \boldsymbol{y} \ge \boldsymbol{c} \\ & \boldsymbol{x} \ge \boldsymbol{0} & & & \boldsymbol{y} \ge \boldsymbol{0} \end{array} \tag{1}$$

dove A è una matrice reale $m \times n$, $\boldsymbol{b} \in \mathbb{R}^m$ e $\boldsymbol{c} \in \mathbb{R}^n$.

Formuliamo ora delle condizioni sufficienti per l'ottimalità note come condizioni di complementarietà. Data una coppia di soluzioni ammissibili $\boldsymbol{x}, \boldsymbol{y}$ di (1) possiamo scrivere

$$\sum_{j=1}^{n} c_j x_j \le \sum_{j=1}^{n} \left(\sum_{i=1}^{m} A_{i,j} y_i \right) x_j \tag{2}$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{i,j} x_j \right) y_i \tag{3}$$

$$\le \sum_{i=1}^{m} b_i y_i \tag{4}$$

Corollario 1 (Condizioni di complementarietà) *Sia (P) un programma lineare con duale (D), dove (P) e (D) hanno forma (1). Se $\boldsymbol{x}$ e $\boldsymbol{y}$ sono soluzioni ammissibili di (P) e (D) e inoltre valgono le condizioni*

1. *ogni volta che $x_j \neq 0$, $\boldsymbol{y}$ soddisfa il j-esimo vincolo con uguaglianza,*
2. *ogni volta che $y_i \neq 0$, $\boldsymbol{x}$ soddisfa l'i-esimo vincolo con uguaglianza,*

allora $\boldsymbol{x}$ e $\boldsymbol{y}$ sono entrambe ottime.

Dimostrazione. La condizione 1 può essere riscritta in forma compatta come

$$c_j x_j = \left(\sum_{i=1}^{m} A_{i,j} y_i \right) x_j$$

e implica che la (2) vale con l'eguaglianza. Analogamente, la condizione 2 può essere riscritta come

$$\left(\sum_{j=1}^{n} A_{i,j} x_j \right) y_i = b_i y_i$$

e implica che pure la (4) vale con l'eguaglianza. Quindi concludiamo che $\boldsymbol{c}^\top \boldsymbol{x} = \boldsymbol{b}^\top \boldsymbol{y}$. Ma dato che $\boldsymbol{c}^\top \boldsymbol{x} \le \boldsymbol{b}^\top \boldsymbol{y}$ vale per ogni coppia di soluzioni ammissibili, le due soluzioni devono essere entrambe ottime. $\square$

Possiamo usare le condizioni di complementarietà operativamente sviluppando algoritmi che cercano di soddisfare simultaneamente le seguenti tre proprietà:

1. $\boldsymbol{x}$ è ammissibile per (P),
2. $\boldsymbol{y}$ è ammissibile per (D),
3. $\boldsymbol{x}$ e $\boldsymbol{y}$ soddisfano le condizioni di complementarietà.

Enunciamo ora il teorema di dualità forte. Per semplicità, useremo una forma particolare di LP ricordando però che ogni LP può essere riscritto nella forma (1).

Teorema 2 (Dualità forte) *Per ogni programma lineare (P) con duale (D) della forma*

$$\begin{array}{llcll} \min & \boldsymbol{c}^\top \boldsymbol{x} & \qquad & \max & \boldsymbol{b}^\top \boldsymbol{y} \\ & A\boldsymbol{x} \geq \boldsymbol{b} & & & A^\top \boldsymbol{y} = \boldsymbol{c} \\ & \boldsymbol{x} \in \mathbb{R}^n & & & \boldsymbol{y} \geq \boldsymbol{0} \end{array} \tag{5}$$

Se (P) e (D) sono entrambi limitati con soluzioni ottime $\boldsymbol{x}^*$ e $\boldsymbol{y}^*$, allora $\boldsymbol{c}^\top \boldsymbol{x}^* = \boldsymbol{b}^\top \boldsymbol{y}^*$.

La dimostrazione del teorema di dualità forte utilizza il lemma di Farkas. A sua volta, la dimostrazione del lemma di Farkas si basa sul seguente risultato geometrico.

Lemma 3 (Iperpiano separatore) *Sia $C \subset \mathbb{R}^d$ chiuso e convesso e $\boldsymbol{z} \notin C$. Allora esiste un iperpiano che separa $\boldsymbol{z}$ da C. Ovvero, esiste $\boldsymbol{w} \in \mathbb{R}^d$ tali che*

$$\boldsymbol{w}^\top \boldsymbol{z} > \max_{\boldsymbol{x} \in C} \boldsymbol{w}^\top \boldsymbol{x} \ .$$

L'intuizione dietro la dimostrazione di questo teorema è semplice. Dato che C è chiuso e convesso, la proiezione ortogonale $\boldsymbol{z}_0$ di $\boldsymbol{z}$ su C esiste ed è unica. Si consideri il punto medio del segmento che unisce $\boldsymbol{z}$ e $\boldsymbol{z}_0$. L'iperpiano passante per questo punto e perpendicolare al vettore $\boldsymbol{z} - \boldsymbol{z}_0$ soddisfa le condizioni del teorema.

Dimostrazione del lemma dell'iperpiano separatore. Prima di tutto, si noti che l'enunciato del lemma può essere riscritto come:

$$\exists \gamma > 0 \quad \text{tale che} \quad \boldsymbol{w}^\top \boldsymbol{z} > \gamma \quad \text{e} \quad \forall \boldsymbol{x} \in C \quad \boldsymbol{w}^\top \boldsymbol{x} \leq \gamma \ .$$

Sia

$$\boldsymbol{z}_0 = \operatorname*{argmin}_{\boldsymbol{x} \in C} \|\boldsymbol{x} - \boldsymbol{z}\|$$

la proiezione di $\boldsymbol{z}$ su C. Questa esiste ed è unica perché C è chiuso e convesso. Poniamo $\boldsymbol{w} = \boldsymbol{z} - \boldsymbol{z}_0$ e $\gamma = \boldsymbol{w}^\top \boldsymbol{z}_0$. La prima disuguaglianza dell'enunciato si dimostra facilmente,

$$\boldsymbol{w}^\top \boldsymbol{z} = \boldsymbol{w}^\top \boldsymbol{z} + \boldsymbol{w}^\top \boldsymbol{z}_0 - \boldsymbol{w}^\top \boldsymbol{z}_0 = \boldsymbol{w}^\top \big(\boldsymbol{z} - \boldsymbol{z}_0\big) + \boldsymbol{w}^\top \boldsymbol{z}_0 = \|\boldsymbol{w}\|^2 + \gamma > \gamma \ .$$

Per dimostrare la seconda disuguaglianza dell'enunciato, scegliamo un $\boldsymbol{x} \in C$ arbitrario. Dato che C è convesso tutti i punti del segmento che unisce $\boldsymbol{x}$ con $\boldsymbol{z}_0$ appartengono a C. Ovvero, $\boldsymbol{x}_\alpha \in C$

per ogni $0 < \alpha \leq 1$, dove $\boldsymbol{x}_\alpha = (1-\alpha)\boldsymbol{z}_0 + \alpha\boldsymbol{x}$. Inoltre, dato che $\boldsymbol{z}_0$ è il punto di C più vicino a $\boldsymbol{z}$, vale $\|\boldsymbol{z} - \boldsymbol{z}_0\| \leq \|\boldsymbol{z} - \boldsymbol{x}_\alpha\|$. Quindi,

$$\begin{aligned}
\|\boldsymbol{z}_0 - \boldsymbol{z}\|^2 &\leq \|(1-\alpha)\boldsymbol{z}_0 + \alpha\boldsymbol{x} - \boldsymbol{z}\|^2 = \|(1-\alpha)(\boldsymbol{z}_0 - \boldsymbol{z}) + \alpha(\boldsymbol{x} - \boldsymbol{z})\|^2 \\
&= (1-\alpha)^2 \|\boldsymbol{z}_0 - \boldsymbol{z}\|^2 + 2\alpha(1-\alpha)(\boldsymbol{z}_0 - \boldsymbol{z})^\top(\boldsymbol{x} - \boldsymbol{z}) + \alpha^2 \|\boldsymbol{x} - \boldsymbol{z}\|^2 \\
&\leq (1-\alpha)^2 \|\boldsymbol{z}_0 - \boldsymbol{z}\|^2 + 2\alpha(1-\alpha)(\boldsymbol{z}_0 - \boldsymbol{z})^\top(\boldsymbol{x} - \boldsymbol{z}) + \alpha^2 \|\boldsymbol{x} - \boldsymbol{z}\|^2
\end{aligned}$$

da cui, osservando che $(1-\alpha)^2 = 1 - 2\alpha + \alpha^2$, otteniamo

$$\alpha(2-\alpha)\|\boldsymbol{z}_0 - \boldsymbol{z}\|^2 \leq 2\alpha(1-\alpha)(\boldsymbol{z}_0 - \boldsymbol{z})^\top(\boldsymbol{x} - \boldsymbol{z}) + \alpha^2 \|\boldsymbol{x} - \boldsymbol{z}\|^2 \ .$$

Dividendo entrambi i membri per $2\alpha > 0$ e scegliendo α arbitrariamente vicino a zero otteniamo

$$\|\boldsymbol{z}_0 - \boldsymbol{z}\|^2 \leq (\boldsymbol{z}_0 - \boldsymbol{z})^\top(\boldsymbol{x} - \boldsymbol{z}) \ . \tag{6}$$

Dato che $\|\boldsymbol{z}_0 - \boldsymbol{z}\|^2 = (\boldsymbol{z}_0 - \boldsymbol{z})^\top(\boldsymbol{z}_0 - \boldsymbol{z})$, la disuguaglianza (6) è equivalente a

$$(\boldsymbol{z} - \boldsymbol{z}_0)(\boldsymbol{z}_0 - \boldsymbol{x}) \geq 0 \ .$$

Ricordando che $\boldsymbol{w} = \boldsymbol{z} - \boldsymbol{z}_0$ e $\gamma = \boldsymbol{w}^\top\boldsymbol{z}_0$ concludiamo che $\gamma \geq \boldsymbol{w}^\top\boldsymbol{x}$ e la dimostrazione è conclusa.
$\square$

Siamo ora pronti per enunciare e dimostrare il lemma di Farkas.

Lemma 4 (Farkas) *Sia A una matrice $m \times n$ e $\boldsymbol{b} \in \mathbb{R}^m$ un vettore. Allora il sistema*

$$\begin{aligned}
A\boldsymbol{x} &= \boldsymbol{b} \\
\boldsymbol{x} &\geq \boldsymbol{0}
\end{aligned}$$

ha soluzione se e solo se il sistema

$$\begin{aligned}
A^\top\boldsymbol{y} &\leq \boldsymbol{0} \\
\boldsymbol{b}^\top\boldsymbol{y} &> 0
\end{aligned}$$

non ha soluzioni.

Dimostrazione. Sia $\boldsymbol{x}_0$ una soluzione del primo sistema. Ovvero, $A\boldsymbol{x}_0 = \boldsymbol{b}$ e $\boldsymbol{x}_0 \geq \boldsymbol{0}$. Allora il secondo sistema non ha soluzioni. Infatti, se $\boldsymbol{y}_0$ fosse una soluzione del secondo sistema, allora

$$\boldsymbol{b}^\top\boldsymbol{y}_0 = \big(A\boldsymbol{x}_0\big)^\top\boldsymbol{y}_0 = \boldsymbol{x}_0^\top\big(A^\top\boldsymbol{y}_0\big) \leq 0$$

che contraddice $\boldsymbol{b}^\top\boldsymbol{y}_0 > 0$.

Dimostriamo ora l'implicazione nell'altro senso. Siano $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n \in \mathbb{R}^m$ le colonne della matrice A e sia C l'insieme dei punti $x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$ con $\boldsymbol{x} \geq \boldsymbol{0}$. In altre parole, C contiene tutti i vettori della forma $A\boldsymbol{x}$ per ogni $\boldsymbol{x} \geq \boldsymbol{0}$. Geometricamente, C è un cono chiuso e convesso; ovvero, un prisma centrato nell'origine, senza base e con spigoli ottenuti prolungando i vettori $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ nel verso positivo. Ora assumiamo che il primo sistema non abbia soluzione. Questo è equivalente a $\boldsymbol{b} \notin C$. Dato che C è chiuso e convesso, per il lemma dell'iperpiano separatore esiste un $\boldsymbol{y}$ tale che

$$\max_{\boldsymbol{x} \in C} \boldsymbol{y}^\top \boldsymbol{x} < \boldsymbol{y}^\top \boldsymbol{b} \; . \tag{7}$$

Dimostriamo ora che $\boldsymbol{y}$ è soluzione del secondo sistema. Per prima cosa abbiamo che $\boldsymbol{y}^\top \boldsymbol{b} > 0$ perché $\boldsymbol{0} \in C$. Dimostriamo ora che vale $A^\top \boldsymbol{y} \leq \boldsymbol{0}$. Se infatti fosse, per esempio, $\left(A^\top \boldsymbol{y}\right)_1 > 0$, allora esisterebbe un vettore $\boldsymbol{x}' = (\alpha, 0, \ldots, 0)$ e un $\alpha > 0$ così grande che

$$\left(A^\top \boldsymbol{y}\right)^\top \boldsymbol{x}' = \left(A^\top \boldsymbol{y}\right)_1 \alpha > \boldsymbol{b}^\top \boldsymbol{y} > 0 \; .$$

Ma ciò non è possibile. Infatti

$$\left(A^\top \boldsymbol{y}\right)^\top \boldsymbol{x}' = \boldsymbol{y}^\top \left(A \boldsymbol{x}'\right) < \boldsymbol{y}^\top \boldsymbol{b}$$

per (7) dato che $A\boldsymbol{x}' \in C$. Questo conclude la dimostrazione. □

Per dimostrare il teorema di dualità forte useremo una variante del Lemma di Farkas la cui dimostrazione (che omettiamo) è simile a quella del Lemma 4.

Lemma 5 (Farkas, seconda variante) *Sia A una matrice $m \times n$ e $\boldsymbol{b} \in \mathbb{R}^m$ un vettore. Allora il sistema $A\boldsymbol{x} \geq \boldsymbol{b}$ ha soluzione in $\mathbb{R}^n$ se e solo se il sistema*

$$\begin{aligned} A^\top \boldsymbol{y} &= \boldsymbol{0} \\ \boldsymbol{b}^\top \boldsymbol{y} &> 0 \\ \boldsymbol{y} &\geq \boldsymbol{0} \end{aligned}$$

non ha soluzioni.

Siamo ora pronti a dimostrare il teorema di dualità forte.

Dimostrazione del Teorema 2. Il teorema di dualità debole ci dice che se $\boldsymbol{x}, \boldsymbol{y}$ sono soluzioni ammissibili di (P) e (D) della forma (5), allora vale sempre $\boldsymbol{c}^\top \boldsymbol{x} \geq \boldsymbol{b}^\top \boldsymbol{y}$. Sia $\boldsymbol{y}^*$ la soluzione ottima di (D) con valore $\boldsymbol{b}^\top \boldsymbol{y}^* = \delta$. Si consideri il sistema

$$\begin{aligned} A\boldsymbol{x} &\geq \boldsymbol{b} \\ \boldsymbol{c}^\top \boldsymbol{x} &\leq \delta \end{aligned} \tag{8}$$

Se (8) ha soluzione, allora (P) deve avere una soluzione $\boldsymbol{x}^*$ tale che $\boldsymbol{c}^\top \boldsymbol{x}^* \leq \delta$ e il teorema è dimostrato.

Per dimostrare che (8) ha soluzione, notiamo che il sistema (8) può essere equivalentemente riscritto come

$$\begin{bmatrix} A \\ -\boldsymbol{c}^\top \end{bmatrix} \boldsymbol{x} \geq \begin{bmatrix} \boldsymbol{b} \\ -\delta \end{bmatrix} \tag{9}$$

Supponiamo per assurdo che (9) non abbia soluzioni. Allora per il Lemma 5 esistono $\boldsymbol{y} \in \mathbb{R}^m$ e $z \in \mathbb{R}$ tali che

$$\left[A^\top, \; -\boldsymbol{c}\right] \begin{bmatrix} \boldsymbol{y} \\ z \end{bmatrix} = \boldsymbol{0} \qquad \begin{bmatrix} \boldsymbol{b} \\ -\delta \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{y} \\ z \end{bmatrix} > 0 \qquad \begin{bmatrix} \boldsymbol{y} \\ z \end{bmatrix} \geq \boldsymbol{0} \tag{10}$$

Ciò implica $A^\top \boldsymbol{y} = z\boldsymbol{c}$ e $\boldsymbol{b}^\top \boldsymbol{y} > z\delta$. Consideriamo due casi.

Caso 1: $z \neq 0$. Dato che $(\boldsymbol{y}, z) \geq \boldsymbol{0}$ per (10), $\boldsymbol{y}' = \boldsymbol{y}/z$ soddisfa $\boldsymbol{y}' \geq \boldsymbol{0}$. Inoltre $A^\top \boldsymbol{y} = z\boldsymbol{c}$ implica $A^\top \boldsymbol{y}' = \boldsymbol{c}$ mentre $\boldsymbol{b}^\top \boldsymbol{y} > z\delta$ implica $\boldsymbol{b}^\top \boldsymbol{y}' > \delta$. Quindi $\boldsymbol{y}'$ è una soluzione ammissibile di (D) con un valore maggiore di δ e abbiamo una contraddizione.

Caso 2: $z = 0$. Allora $A^\top \boldsymbol{y} = \boldsymbol{0}$ e $\boldsymbol{b}^\top \boldsymbol{y} > 0$. Dato che $A^\top \boldsymbol{y}^* = \boldsymbol{c}$ e $\boldsymbol{b}^\top \boldsymbol{y}^* = \delta$, vale $A^\top \big(\boldsymbol{y}^* + \varepsilon \boldsymbol{y}\big) = \boldsymbol{c}$ per ogni $\varepsilon > 0$. Inoltre, dato che $\boldsymbol{y} \geq \boldsymbol{0}$ per (10) e $\boldsymbol{y}^* \geq \boldsymbol{0}$, abbiamo che $\big(\boldsymbol{y}^* + \varepsilon \boldsymbol{y}\big) \geq \boldsymbol{0}$. Infine, $\boldsymbol{b}^\top \big(\boldsymbol{y}^* + \varepsilon \boldsymbol{y}\big) = \delta + \varepsilon \boldsymbol{b}^\top \boldsymbol{y} > \delta$. Quindi $\boldsymbol{y}^* + \varepsilon \boldsymbol{y}$ è una soluzione ammissibile di (D) con un valore maggiore di δ e abbiamo ancora una contraddizione.

Ma allora (8) ha soluzioni e il teorema è dimostrato. $\square$