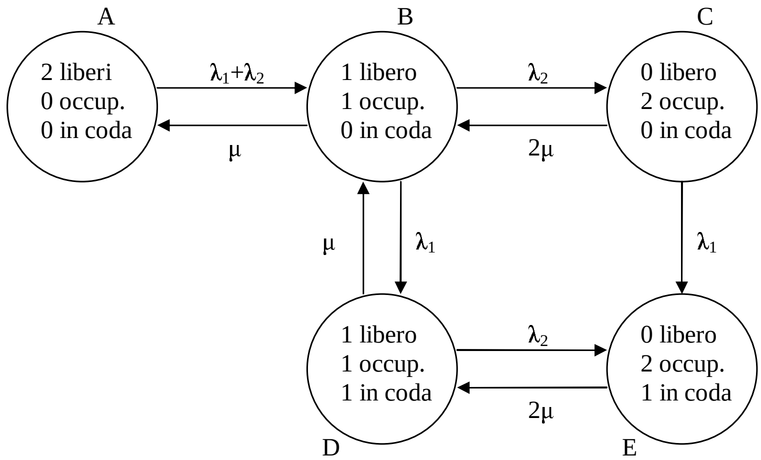
# Soluzioni degli esercizi della gara AIRO 2009-2010, fase locale

## 1 Il pronto soccorso

Il problema richiede anzitutto di rappresentare il funzionamento del sistema tramite un grafo, in cui ogni nodo corrisponde ad un possibile *stato* del sistema ed ogni arco corrisponde ad una possibile *transizione di stato*. E' poi possibile calcolare le probabilità associate ai possibili stati (indichiamo con $P_i$ la probabilità dello stato $i$), conoscendo la frequenza delle transizioni tra di essi (indichiamo con $f_{i \to j}$ la frequenza di transizione dallo stato $i$ allo stato $j$).

Nel nostro caso il sistema è rappresentato dal grafo in figura.




In figura le frequenze di transizione sono indicate da:

- Arrivo pazienti urgenti: $\lambda_1 = \frac{4}{3}$ pazienti/ora
- Arrivo pazienti non urgenti: $\lambda_2 = 4$ pazienti/ora
- Completamento servizio: $\mu = 3$ pazienti/ora.

La frequenza delle transizioni può essere riportata anche in una matrice con tante righe e colonne quanti gli stati, cioè 5.

Per ricavare le probabilità associate ad ogni stato, occorre impostare un sistema di 5 equazioni di bilanciamento tra la frequenza media di ingresso e la frequenza media di uscita da ciascuno stato (bastano quattro equazioni, perché la quinta è linearmente dipendente dalle altre). Ad esempio per lo stato $B$ si ha:

$$P_B * (f_{B \to A} + f_{B \to C} + f_{B \to D}) = P_A * f_{A \to B} + P_C * f_{C \to B} + P_D * f_{D \to B}$$

| $f$ | $A$ | $B$ | $C$ | $D$ | $E$ |
|---|---|---|---|---|---|
| $A$ | - | $\frac{16}{3}$ | 0 | 0 | 0 |
| $B$ | 3 | - | $\frac{4}{3}$ | 4 | 0 |
| $C$ | 0 | 6 | - | 0 | 4 |
| $D$ | 0 | 3 | 0 | - | $\frac{4}{3}$ |
| $E$ | 0 | 0 | 0 | 6 | - |

Tabella 1: Frequenze di transizione tra gli stati.

dove i valori di $f$ sono riportati nella tabella delle frequenze di transizione e i valori di $P$ sono le incognite del problema.

Il problema ha quindi 5 incognite e 5 vincoli di uguaglianza (uno dei quali ridondante). Bisogna infine imporre che la somma delle 5 probabilità sia pari a 1 (condizione di normalizzazione).

Il risultante sistema di 5 equazioni lineari in 5 incognite si può risolvere facilmente anche a mano o con un solutore software. Il modello Excel è nel file "1 - Pronto Soccorso.xls".

Le probabilità dei 5 stati risultano essere le seguenti:

- $P_A \cong 0,15$
- $P_B \cong 0,28$
- $P_C \cong 0,04$
- $P_D \cong 0,42$
- $P_E \cong 0,12$.

**Domanda 1.** Il sistema è saturo per i pazienti urgenti (cioè non accetta ulteriori pazienti urgenti) quando si trova negli stati C ed E, il che avviene con probabilità $P_C + P_E \cong 0,15$. La frazione di tempo è quindi pari al 15% circa.

**Domanda 2.** Il sistema è saturo per i pazienti non urgenti (cioè non accetta ulteriori pazienti non urgenti) quando si trova negli stati D ed E, il che avviene con probabilità $P_D + P_E \cong 0,53$. La frazione di tempo è quindi pari al 53% circa.

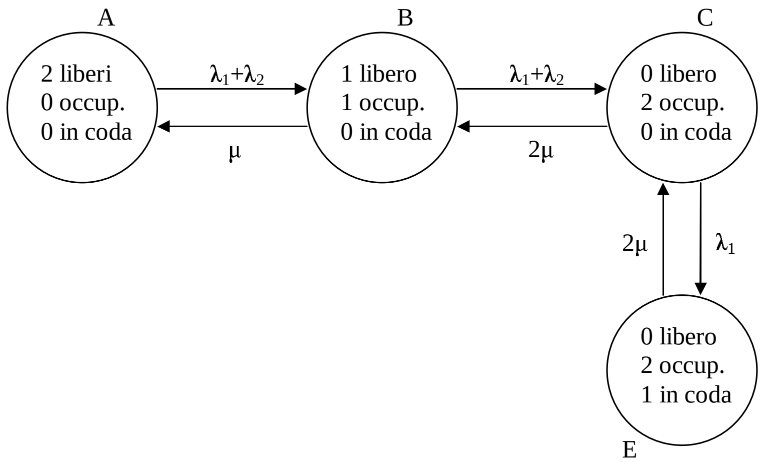
**Domanda 3.** La forza-lavoro disponibile non è impiegata quando il sistema si trova nello stato A, è impiegata al 50% (un medico su due) quando il sistema si trova negli stati B e D ed è impiegata al 100% quando il sistema si trova negli stati C ed E. Quindi il fattore di utilizzo dei medici è dato da

$$\rho = 0 * P_A + 0,5 * (P_B + P_D) + 1 * (P_C + P_E) \cong 0,50.$$

**Domanda 4.** Un paziente non urgente viene messo in coda quando al suo arrivo il sistema si trova negli stati B o C, il che accade con probabilità $P_B + P_C \cong 0,31$.

**Domanda 5.** Per rispondere a questa domanda occorre modificare leggermente la descrizione del sistema e ripetere poi lo stesso procedimento seguito per rispondere alle domande 1-4. Se i medici servono sempre i pazienti in arrivo, lo stato D non viene più raggiunto: all'arrivo di un paziente non urgente, se il sistema è nello stato B, passa allo stato C. Inoltre se il sistema si trova nello stato E, al completamento del trattamento di un paziente il sistema passa nello stato C, cioè il medico che si è liberato prende subito in carico il paziente che si trovava in coda.

Pertanto il grafo degli stati si modifica come segue.




La matrice delle frequenze di transizione diventa la seguente.

| $f$ | $A$ | $B$ | $C$ | $E$ |
|---|---|---|---|---|
| $A$ | - | $\frac{16}{3}$ | 0 | 0 |
| $B$ | 3 | - | $\frac{16}{3}$ | 0 |
| $C$ | 0 | 6 | - | 4 |
| $E$ | 0 | 0 | 6 | - |

Tabella 2: Domanda 5: Frequenze di transizione tra gli stati.

Il modello Excel è nel file "1 - Pronto Soccorso.xls" (cartella "Risposta 5"). Le probabilità degli stati risultano essere le seguenti:

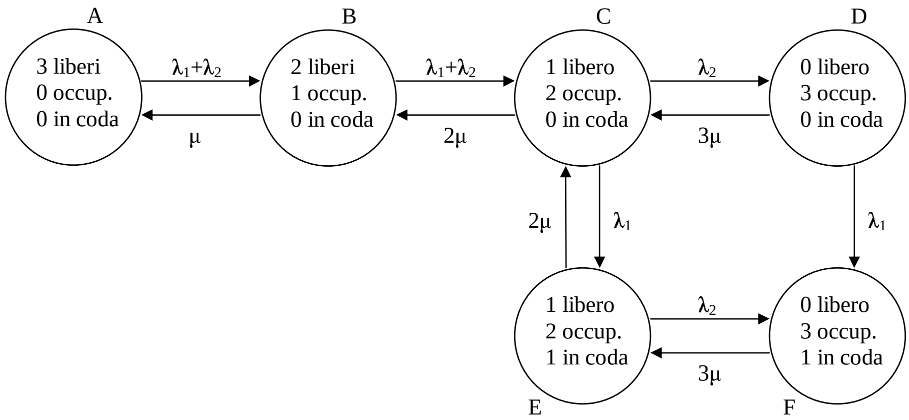
- $P_A \cong 0,18$
- $P_B \cong 0,33$

- $P_C \cong 0,29$
- $P_E \cong 0,19$.

Pertanto le quattro risposte alle quattro domande precedenti sono in questo caso:

- Domanda 1: $P_C + P_E = 48,7\%$ circa
- Domanda 2: $P_D + P_E = 19,5\%$ circa
- Domanda 3: $\rho \cong 65,1\%$
- Domanda 4: $P_C = 0,29$ circa.

**Domanda 6.** Se i medici fossero tre, il grafo degli stati verrebbe modificato come segue.




La matrice delle frequenza di transizione diventerebbe la seguente.

| $f$ | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ |
|---|---|---|---|---|---|---|
| $A$ | - | $\frac{16}{3}$ | 0 | 0 | 0 | 0 |
| $B$ | 3 | - | $\frac{16}{3}$ | 0 | 0 | 0 |
| $C$ | 0 | 6 | - | $\frac{4}{3}$ | 4 | 0 |
| $D$ | 0 | 0 | 9 | - | 0 | 4 |
| $E$ | 0 | 0 | 6 | 0 | - | $\frac{4}{3}$ |
| $F$ | 0 | 0 | 0 | 0 | 9 | - |

Tabella 3: Domanda 6: Frequenze di transizione tra gli stati.

Il modello Excel è nel file "1 - Pronto Soccorso.xls" (cartella "Risposta 6"). Le probabilità degli stati risultano essere le seguenti:

- $P_A \cong 0,17$

- $P_B \cong 0,30$
- $P_C \cong 0,27$
- $P_D \cong 0,03$
- $P_E \cong 0,20$
- $P_F \cong 0,04$.

Pertanto le quattro risposte alle quattro domande precedenti sono in questo caso:

- Domanda 1: $P_D + P_F = 7\%$ circa
- Domanda 2: $P_E + P_F = 23,7\%$ circa
- Domanda 3: $\rho \cong 47,7\%$
- Domanda 4: $P_C + P_D = 0,29$ circa.

**Commenti.** Questo esercizio richiede di applicare alcune nozioni di base sulle *catene di Markov*, uno strumento modellistico molto utile nell'ambito della *teoria delle code* (e non solo).

## 2 La legge di Ohm

Il modello matematico di questo problema si può formulare come segue.

**Dati.**

- un insieme $\mathcal{N}$ di studenti
- due vettori, indicati da $v$ e $I$, che contengono i valori di tensione e di intensità di corrente misurati da ogni studente.

**Variabili.** L'incognita del problema è una sola, la resistenza $R$.

**Vincoli.** Poiché il voltmetro è affetto da errore, la misura della tensione per ogni studente $j \in \mathcal{N}$ risulta essere pari a $v_j = RI_j + e_j$, dove $v_j$ ed $I_j$ sono rispettivamente la tensione e l'intensità di corrente misurate dallo studente, mentre $e_j$ è l'errore di misura.

**Obiettivi.** L'obiettivo è di trovare il valore di $R$ che rende minima una funzione degli errori e a questo scopo è richiesto di considerare tre diverse funzioni obiettivo.

**Prima funzione obiettivo.** La prima funzione obiettivo richiede di minimizzare il massimo errore in valore assoluto, ossia

$$\text{minimize } f_1 = \max_{j \in \mathcal{N}} \{|e_j|\}.$$

Per esprimere questa funzione obiettivo con funzioni lineari si può introdurre una variabile ausiliaria, che indichiamo con $z$, e richiedere

$$\min z$$

imponendo i vincoli

$$z \geq |e_j| \quad \forall j \in \mathcal{N}.$$

In questo modo $z$ assume il valore minimo tra tutti quelli maggiori o uguali agli errori (in valore assoluto), cioè il valore del massimo errore (in valore assoluto).

Per eliminare poi il valore assoluto, basta sdoppiare i vincoli e imporre

$$z \geq e_j \quad \text{e} \quad z \geq -e_j \quad \forall j \in \mathcal{N}.$$

**Seconda funzione obiettivo.** La seconda funzione obiettivo richiede di minimizzare il valor medio degli errori in valore assoluto, ossia

$$\text{minimize } f_2 = \frac{1}{|\mathcal{N}|} \sum_{j \in \mathcal{N}} |e_j|.$$

Per esprimere questa funzione obiettivo senza i valori assoluti si può introdurre una variabile ausiliaria $w_j$ per ogni studente ed imporre i vincoli

$$w_j \geq e_j \quad \text{e} \quad w_j \geq -e_j \quad \forall j \in \mathcal{N}.$$

In tal modo la funzione obiettivo viene riformulata come

$$\text{minimize } f_2 = \frac{1}{|\mathcal{N}|} \sum_{j \in \mathcal{N}} w_j.$$

**Terza funzione obiettivo.** La terza funzione obiettivo richiede di minimizzare il valor medio degli errori quadratici, ossia

$$\text{minimize } f_3 = \frac{1}{|\mathcal{N}|} \sum_{j \in \mathcal{N}} e_j^2.$$

Questa funzione obiettivo è non-lineare (quadratica) e richiede quindi di essere minimizzata con un solutore per modelli di programmazione non-lineare. Tuttavia, poiché il problema è *convesso*, e quindi ha un solo punto di minimo locale, è possibile riformulare il problema di ottimizzazione quadratico come problema di esistenza lineare. Infatti il punto di minimo cercato è l'unico che soddisfa le condizioni analitiche del primo ordine, che impongono che la derivata prima della funzione obiettivo rispetto alla variabile sia nulla. Si ha quindi:

$$\frac{d(\frac{1}{|\mathcal{N}|} \sum_j e_j^2)}{d(R)} = 0$$

$$\sum_j \frac{d(e_j^2)}{d(R)} = 0$$

$$\sum_j 2e_j \frac{d(e_j)}{d(R)} = 0$$

$$\sum_j (v_j - RI_j) \frac{d(v_j - RI_j)}{d(R)} = 0$$

$$\sum_j (v_j - RI_j)(-I_j) = 0$$

$$\sum_j (-v_j I_j + RI_j^2) = 0$$

$$\sum_j (-v_j I_j) + \sum_j (RI_j^2) = 0$$

$$\sum_j (v_j I_j) = R \sum_j I_j^2$$

Poiché la variabile è una sola, il problema lineare risultante ha una sola incognita ($R$) ed una sola equazione (quella sopra riportata) ed è quindi risolubile direttamente:

$$R = \frac{\sum_j (v_j I_j)}{\sum_j I_j^2}.$$

**Commenti.** Questo esercizio richiede alcune semplici nozioni di programmazione matematica e si presta ad illustrare i concetti di *convessità* e di *linearità*, nonché ad applicare le condizioni analitiche del primo ordine per la ricerca di punti di minimo locale.

# 3 Il tracciato dell'elettrodotto

**Dati.** I dati del problema sono:

- un insieme $\mathcal{N}$ di paesi da raggiungere;
- le coordinate $(x_i, y_i)$ per ogni paese $i \in \mathcal{N}$;
- la struttura della rete elettrica.

**Variabili.** Il problema richiede di decidere dove localizzare la cabina di trasformazione per ciascuno dei paesi. Le variabili decisionali del problema sono quindi le coordinate cartesiane $(\overline{x}_i, \overline{y}_i)$ delle cabine di trasformazione per ogni paese $i \in \mathcal{N}$ tranne il primo, dove non è localizzato un paese bensì l'impianto di produzione.

**Funzione obiettivo.** Tutte le funzioni obiettivo da considerare richiedono di minimizzare costi. Per esprimere i costi è necessario conoscere le lunghezze dei tratti di linea ad alta tensione e le lunghezze degli scavi per i collegamenti a bassa tensione. I secondi sono facilmente esprimibili come distanze euclidee tra la cabina di trasformazione e la centralina per ogni paese. I primi invece devono rispecchiare la struttura ad albero della rete elettrica. La rete è un'*arborescenza* con radice nel punto A. Quindi ogni punto, eccetto A, ha uno e un solo *predecessore*. Ogni segmento dell'arborescenza è il collegamento tra un punto ed il suo predecessore. E' quindi consigliabile per scrivere il modello matematico del problema definire un vettore di predecessori, che contiene gli indici dei punti predecessori. Il primo valore nel vettore, quello che corrisponde al predecessore di A, non è usato nel modello e può contenere qualsiasi valore. Una volta definito il vettore dei predecessori, le distanze dei tratti ad alta tensione sono quelle da ogni punto al suo predecessore.

Una volta espresse le distanze dei tratti ad alta tensione e a bassa tensione relativi ad ogni nodo della rete, è immediato esprimere le funzioni obiettivo di Tizio, Caio e Sempronio.

**Vincoli.** I vincoli limitano superiormente le lunghezze dei tratti a bassa tensione.

Il modello risultante è non-lineare, a causa delle formule che esprimono la distanza tra due punti nel piano. La soluzione calcolata dal solutore perciò non viene garantita essere ottima globalmente, ma solo localmente.

Quando viene ottimizzata la funzione obiettivo di Tizio ovviamente la soluzione ottima ha costo nullo e consiste nel localizzare le cabine di trasformazione sempre in coincidenza delle centraline.

**Commenti.** Questo esercizio richiede di minimizzare funzioni di distanze euclidee calcolate in due dimensioni (programmazione non-lineare). Inoltre richiede di minimizzare la lunghezza complessiva degli archi di un'*arborescenza* e si presta quindi ad introdurre concetti di base di ottimizzazione su grafo.

# 4 Il cassiere

Tramite un modello matematico, il problema si può formalizzare come segue.

**Dati.** Sono dati:

- un insieme $\mathcal{N}$ di tagli di banconote (nel nostro caso 5 tagli);
- un valore $v_i$ per ogni taglio $i \in \mathcal{N}$;
- un numero $b_i$ di banconote disponibili per ogni taglio $i \in \mathcal{N}$;
- un importo $T$ da raggiungere.

**Variabili decisionali.** Le decisioni si possono rappresentare indicando il numero $x_i$ di banconote che vengono usate per ciascun taglio $i \in \mathcal{N}$.

**Obiettivo.** Si vuole minimizzare il numero totale di banconote usate:

$$minimize\ f = \sum_{i \in \mathcal{N}} x_i.$$

**Vincoli.**

- In tutti e tre i casi considerati esiste il vincolo che non si possono usare più banconote di quelle disponibili per ogni taglio:

$$x_i \leq b_i \quad \forall i \in \mathcal{N}.$$

- Inoltre si vuole ottenere esattamente l'importo $T$:

$$\sum_{i \in \mathcal{N}} v_i x_i = T.$$

Nel caso della domanda 1 non ci sono altri vincoli.

Nel caso della domanda 2 è necessario introdurre un vincolo che impone l'uso di almeno una banconota da 10 Euro (il taglio numero 4 dei 5 possibili) quando viene usata almeno una banconota da 20 Euro (il taglio numero 3). A questo scopo si possono introdurre variabili binarie $z_i$, che assumono valore 1 se il taglio $i \in \mathcal{N}$ viene usato e 0 altrimenti, attraverso i vincoli

$$x_i \leq b_i z_i$$

che sostituiscono i precedenti vincoli

$$x_i \leq b_i.$$

L'uso della banconota da 10 Euro viene quindi imposta dal vincolo

$$x_4 \geq z_3.$$

Nel caso della domanda 3, usando le stesse variabili $z$ di cui sopra, si può imporre

$$1 - z_2 \geq z_3.$$

**Commenti.** L'esercizio si può risolvere sia per via logica, sia scrivendo un modello di programmazione lineare intera e usando poi un solutore. L'esercizio richiede l'uso di variabili binarie per esprimere condizioni logiche.

## 5 I libri scolastici

**Dati.** Sono dati:

- un insieme di origini $\mathcal{O}$;
- un insieme di destinazioni $\mathcal{D}$;
- un'offerta $o_i$ per ogni origine $i \in \mathcal{O}$;
- una domanda $d_j$ per ogni destinazione $j \in \mathcal{D}$;
- una capacità massima $q_{ij}$ per ogni tratta origine-destinazione;
- dei costi unitari di spedizione $c_{ij}$ su ciascuna tratta origine-destinazione.

**Varaibili.** Le variabili decisionali indicano il numero $x_{ij}$ di libri che vengono trasportati da ogni origine $i \in \mathcal{O}$ ad ogni destinazione $j \in \mathcal{D}$.

**Vincoli.** I vincoli del problema impongono che:

- la quantità complessiva in partenza da ogni origine $i \in \mathcal{O}$ non sia superiore all'offerta $o_i$:
$$\sum_{j \in \mathcal{D}} x_{ij} \leq o_i \ \forall i \in \mathcal{O}$$
- la quantità totale in arrivo in ogni destinazione $j \in \mathcal{D}$ sia almeno pari alla domanda $d_j$:
$$\sum_{i \in \mathcal{O}} x_{ij} \geq d_j \ \forall j \in \mathcal{D}$$
- La quantità trasportata su ogni tratta $(i, j)$ non ecceda il limite $q_{ij}$:
$$x_{ij} \leq q_{ij} \quad \forall i \in \mathcal{O}, \forall j \in \mathcal{D}.$$

**Obiettivo.** L'obiettivo è quello di minimizzare i costi, che sono dati dall'espressione
$$f_A = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} c_{ij} x_{ij}.$$

In questo primo caso, relativo all'azienda "TuttoTrasporti A" si ottiene quindi un problema di programmazione lineare. Poiché tutti i dati sono interi, grazie alla particolare struttura dei vincoli del problema, si ottiene una soluzione ottima con valori interi anche senza bisogno di imporre esplicitamente le condizioni di integralità sulle variabili $x$.

Dalla soluzione (v. foglio di lavoro "Modello PL" nel file Excel) si ricava che è possibile soddisfare tutta la domanda ed il costo minimo è pari a 10670 Euro.

In un secondo scenario, relativo all'azienda "TuttoTrasporti B", la formula per il calcolo del costo dà luogo ad un modello di programmazione non-lineare, poiché le quantità trasportate figurano al denominatore. In questo caso se non vengono imposte le condizioni di integralità sulle quantità trasportate si ottengono in generale soluzioni (localmente) ottime con valori frazionari delle variabili. Imponendo che le variabili $x$ assumano valori interi si ottiene una soluzione di costo 11353,07 Euro (v. foglio di lavoro "Modello PNL" nel file Excel).

**Commenti.** L'esercizio si presta al confronto tra due diversi sistemi logistici, descritti dagli stessi vincoli dal punto di vista tecnologico, ma caratterizzati da diverse funzioni costo, che danno luogo a modelli matematici diversi (PL e PNL 0-1), che richiedono algoritmi diversi. Poiché nel primo dei due casi il problema è un problema di trasporto, l'esercizio si presta ad introdurre i modelli e gli algoritmi di ottimizzazione su grafo (problemi di flusso) ed il concetto di integralità della soluzione ottima e a discutere di conseguenza la necessità di esplicitare le condizioni di integralità delle variabili nei modelli. Infine si presta ad illustrare la differenza tra le garanzie fornite da algoritmi per modelli di PL e di PNL (ottimalità globale o locale).

## 6 Noleggio auto

**Dati.** I dati del problema sono i seguenti:

- un insieme $\mathcal{N}$ di ordini;
- per ogni ordine $i \in \mathcal{N}$ un giorno iniziale $s_i$ ed un giorno finale $e_i$;
- per ogni ordine $i \in \mathcal{N}$ un profitto $p_i$;
- un numero massimo $V$ di veicoli disponibili (costante in tutti i giorni del periodo considerato).

**Variabili.** Le decisioni sono rappresentate da variabili binarie $x$ associate agli ordini: $x_i = 1$ indica che l'ordine viene soddisfatto, $x_i = 0$ indica che l'ordine viene rifiutato.

**Obiettivo.** La funzione obiettivo da ottimizzare, ossia il profitto del noleggiatore, è data da $f = \sum_{i \in \mathcal{N}} p_i x_i$.

**Vincoli.** I vincoli devono imporre che il numero di veicoli impegnati in ogni giorno non sia superiore al valore $V$. A questo scopo una possibilità è quella di usare una matrice con tante righe quanti gli ordini e tante colonne quanti i giorni del periodo considerato, dove ogni cella sulla riga $i$ e sulla colonna $g$ indica con un valore $a_{ig} = 1$ o $a_{ig} = 0$ se l'ordine $i \in \mathcal{N}$ richiede un veicolo nel giorno $g$ o no. L'insieme $\mathcal{G}$ dei giorni da considerare si ricava dai dati dell'esempio ed è pari a 35. Per esprimere i vincoli basta quindi contare per ogni giorno $g \in \mathcal{G}$ quanti sono gli ordini soddisfatti che richiedono un veicolo in quel giorno ed imporre che tale numero sia non superiore a $V$:

$$\sum_{i \in \mathcal{N}} a_{ig} x_i \leq V \quad \forall g \in \mathcal{G}.$$

Il modello risultante è lineare, con variabili binarie.

Come risulta dal file Excel, la soluzione ottima ha valore pari a 5200 Euro. Risolvendo lo stesso modello con 5 veicoli anziché 4, si ottiene una soluzione ottima di valore pari a 6400 Euro. Quindi noleggiare un veicolo aggiuntivo per 1000 Euro è conveniente, poiché il profitto aumenta di 1200 Euro.

**Commenti.** L'esercizio si presta sia alla formulazione matematica e all'uso di un solutore, sia alla realizzazione di un semplice *algoritmo di programmazione dinamica*. Si presta inoltre ad eseguire una semplice analisi post-ottimale sugli effetti che una variazione dei dati ha sul valore ottimo.

# 7 Energia eolica

**Dati.** I dati del problema sono i seguenti:

- un insieme $\mathcal{S}$ di settimane (numerate da 1 a 52);
- un fabbisogno $f_s$ di energia per ogni settimana $s \in \mathcal{S}$;
- una produzione $p_s$ di energia per ogni settimana $s \in \mathcal{S}$;
- un insieme $\mathcal{T}$ di tipi di batterie;
- per ogni tipo $t \in \mathcal{T}$ sono dati una tensione $v_t$, un amperaggio $a_t$ e un prezzo $c_t$.

Dall'analisi dei dati si ricava subito che la produzione totale eccede il fabbisogno totale di $\Delta = 10,750$ Kilowattora, che devono quindi andare dispersi.

**Variabili (1).** Per poter esprimere i vincoli e la funzione obiettivo è necessario sapere quanta energia è immagazzinata degli accumulatori al termine di ogni setitmana. A questo scopo si possono usare variabili non-negative $x_s \;\; \forall s \in \mathcal{S}$.

**Vincoli.** Il profilo dell'energia accumulata nella batteria in funzione del tempo si ricava imponendo che l'energia accumulata al termine della settimana $s$, indicata con $x_s$, sia pari a quella accumulata al termine della settimana precedente più quella prodotta nella settimana $s$ meno quella consumata nella settimana $s$:

$$x_s = x_{s-1} + p_s - f_s.$$

Gli indici in questa formula sono intesi modulo 52, cioè si ha $x_1 = x_{52} + p_1 - f_1$ in modo da definire un andamento ciclico che si ripete ogni anno. Le 52 variabili $x$, introdotte nel modello per indicare l'andamento dell'energia accumulata, sono definite a meno di una costante: è possibile infatti aggiungere una qualsiasi costante a tutte, mantenendo tuttavia soddisfatti i vincoli. Per definire questa costante è possibile ad esempio mettere in evidenza la quantità di energia che si trova accumulata all'inizio dell'anno, ossia $x_0 = x_{52}$.

I vincoli del problema impongono che tutte le variabili $x_s$ siano comprese tra 0 e la capacità, indicata con $y$, della batteria.

**Obiettivo.** La capacità $y$ è a sua volta una variabile decisionale ed è anche la funzione obiettivo da minimizzare per rispondere alla prima domanda.

Per minimizzare la capacità necessaria, è conveniente che la dispersione dell'energia $\Delta$ prodotta in eccesso avvenga tra il punto di minimo e il punto di massimo del profilo dell'energia. In tal modo è sempre possibile abbassare il picco di una quantità pari a $\Delta$, riducendo quindi la capacità massima richiesta per l'accumulatore.

**Variabili (2).** Alternativamente si può introdurre un altro insieme di variabili, indicate con $z_s$, che indicano la dispersione di energia in ogni settimana $s \in \mathcal{S}$. In questo caso l'equazione di conservazione dell'energia si modifica come segue:

$$x_s = x_{s-1} + p_s - f_s - z_s.$$

E' sempre conveniente scegliere $x_0$ in modo che il punto di minimo del profilo dell'energia accumulata corrisponda ad un valore nullo; diversamente infatti sarebbe possibile traslare tutto il grafico delle variabili $x$ verso il basso senza violare alcun vincolo ma riducendo il valore richiesto per la capacità.

Riassumendo, le variabili del problema sono:

- le variabili $x_s \quad \forall s \in \mathcal{S}$;
- le variabili $z_s \quad \forall s \in \mathcal{S}$;
- il picco di accumulo $y$.

I vincoli del problema invece sono:

- le equazioni di conservazione dell'energia in ogni settimana, inclusa l'equazione che collega l'ultima settimana alla prima;
- il vincolo di non-negatività delle variabili $x$ e $z$;
- il vincolo sul massimo accumulo, che impone che tutti i valori di $x_s$ siano sempre non superiori al valore di $y$.

La funzione obiettivo richiede semplicemente di minimizzare $y$.

Il modello risultante è di programmazione lineare. Il valore minimo del picco di accumulo (v. foglio di lavoro "Domanda 1" nel file Excel) è $y^* = 21,125$ Kilowattora.

La seconda domanda richiede di trovare la combinazione di batterie di minimo costo che può soddisfare la domanda di capacità di accumulo ricavata dalla risposta alla domanda precedente.

Entrano in gioco solo a questo punto i dati relativi ai vari tipi di batterie. Le variabili decisionali, che indichiamo con $w$, indicano il numero $w_t$ di batterie di ogni tipo $t \in \mathcal{T}$ e sono variabili *intere*.

C'è un solo vincolo, espresso dalla disuguaglianza $\sum_{t \in \mathcal{T}} c_t a_t w_t \geq y^*$.

La funzione obiettivo richiede di minimizzare il costo complessivo, dato da $\sum_{t \in \mathcal{T}} c_t w_t$.

Il modello che ne risulta è lineare con variabili intere e la soluzione ottima (v. foglio di lavoro "Domanda 2" nel file Excel) ha costo 3425 Euro e consiste nel comporre una batteria di tipo 2 con sei batterie di tipo 4.

Alla domanda numero 3 si può rispondere senza calcoli, osservando che ai fini della determinazione del picco di energia accumulata e del costo delle batterie è ininfluente la scelta del periodo in cui effettuare la dispersione purché sia compreso tra la settimana in cui si ha il valore minimo (nullo) e la settimana in cui si ha il valore massimo ($y^*$) tra tutti i valori delle variabili $x$.

La domanda 4 richiede di modificare ulteriormente l'equazione di conservazione dell'energia, tenendo conto di una dispersione che avviene ogni settimana, proporzionalmente alla quantità di energia accumulata. Si ha perciò:

$$x_s = \alpha x_{s-1} + p_s - f_s - z_s.$$

Il valore di $\alpha$ deve essere determinato e quindi il problema è non-lineare a causa del prodotto tra la variabile $\alpha$ e le variabili $x$. Si chiede qual è il minimo valore che $\alpha$ può assumere garantendo tuttavia l'esistenza di una soluzione ammissibile. In assenza di vincoli sulla capacità dell'accumulatore è sicuramente una scelta ottima quella di azzerare i valori delle variabili $z$, che quindi non è più necessario considerare. Di conseguenza le uniche variabili del problema sono $\alpha$ e $x_0$.

Il valore cercato è circa $98,45\%$ (v. foglio di lavoro "Domanda 4" nel file Excel). E' interessante notare che in questo caso il picco massimo di energia accumulata raggiunge il valore di 26,354 Kilowattora, ben superiore a quello del caso precedente.

Per rispondere alla domanda 5 occorre risolvere lo stesso problema già considerato per rispondere alla domanda 2, ma con un valore di capacità richiesta superiore (26,354 Kilowattora anziché 21,125). Se è possibile ri-progettare l'impianto di accumulazione da capo, allora la soluzione di minimo costo è data da una batteria di tipo 1 e otto di tipo 4, con un costo complessivo di 4385 Euro. Se invece il sistema è stato progettato inizialmente per batterie senza autoscarica e a causa dell'autoscarica si vuole aumentare la capacità del sistema di accumulazione pre-esistente, allora è necessario imporre che il numero di batterie di ciascun tipo nella soluzione del problema con autoscarica sia non inferiore al numero di batterie dello stesso tipo nella soluzione del problema senza autoscarica. In tal caso la soluzione ottima prevede l'aggiunta di una batteria di tipo 3 e una batteria di tipo 4 alla precedente soluzione ottima del caso senza autoscarica (una batteria di tipo 2 e sei di tipo 4) per un costo complessivo di 4415 Euro (di cui 3425 Euro per l'impianto senza autoscarica e 990 Euro per la sua espansione).

**Commenti.** La prima parte dell'esercizio richiede di lavorare su una quantità di dati che preclude la possibilità di ricavare soluzioni con carta e penna in tempi ragionevoli e si presta quindi ad illustrare la necessità di saper lavorare con strumenti come il foglio elettronico. Consente inoltre ragionamenti di tipo qualitativo, ad esempio sulla scelta ottima del periodo in cui disperdere l'energia in eccesso. L'esercizio richiede anche di usare indici ciclici e di risolvere modelli diversi in cui il dato di uno è il valore ottimo dell'altro. La seconda parte richiede la soluzione di un problema non-lineare con sole due variabili decisionali (che si può anche affrontare per tentativi, oltre che con un solutore) e la nuova risoluzione di un modello già risolto nella prima parte che dà luogo a due diverse possibili interpretazioni, con conseguenti diverse soluzioni e diversi costi.

# 8 Rete sociale

**Dati.** I dati del problema sono descritti da:

- un grafo $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ di relazioni sociali, rappresentabile con una matrice simmetrica i cui elementi $a_{ij}$ hanno valori binari: $a_{ij} = 1$ se e solo se l'insieme degli spigoli $\mathcal{E}$ contiene l'elemento $[i, j]$;
- un parametro intero $k$.

**Variabili.** Il problema ammette diverse formulazioni, ma richiede comunque l'uso di variabili binarie. In tutti i casi tuttavia non si riesce a formulare il modello matematico senza ricorrere a variabili con tre indici o ad insiemi di vincoli su triplette di nodi del grafo. Quindi alla soluzione con il foglio elettronico, dove i dati si rappresentano agevolmente in forma di vettori o matrici ma non oltre, è da preferirsi l'uso di un linguaggio di modellizzazione matematica come MathProg o AMPL.

Una formulazione possibile è quella che fa uso di variabili binarie $x_{ij}$ per ogni coppia di nodi $i$ e $j$ del grafo: $x_{ij} = 1$ se e solo se $i$ e $j$ vengono inclusi nella stessa partizione del grafo.

**Vincoli.** Con tale scelta delle variabili, i vincoli del problema devono imporre:

- la proprietà simmetrica:

$$x_{ij} = x_{ji} \quad \forall i, j \in \mathcal{V}$$

- la proprietà transitiva:

$$x_{ij} + x_{jk} \leq 1 + x_{ik} \quad \forall i, j, k \in \mathcal{V}.$$

Il secondo insieme di vincoli impone che se $i$ e $j$ sono inclusi nella stessa partizione e $j$ e $k$ lo sono pure, allora anche $i$ e $k$ devono esserlo.

**Funzikone obiettivo.** La funzione obiettivo da massimizzare è data da tre termini:

- premio per vertici adiacenti e inclusi nella stessa partizione:

$$+k \sum_{i,j \in \mathcal{V}} a_{ij} x_{ij}$$

- penalità per vertici non adiacenti inclusi nella stessa partizione:

$$-\sum_{i,j \in \mathcal{V}} (1 - a_{ij}) x_{ij}$$

- penalità per vertici adiacenti non inclusi nella stessa partizione:

$$-\sum_{i,j\in\mathcal{V}} a_{ij}(1-x_{ij})$$

ed è lineare.

Il modello risultante è di programmazione lineare con variabili binarie.

Risolvendolo per i cinque diversi valori del parametro $k = 1, \ldots, 5$ si ottengono le soluzioni ottime riportate nei files "rete1.out",...,"rete5.out".

All'ultima domanda si può rispondere considerando la configurazione con una sola partizione e chiedendosi a quale condizione può essere più conveniente una configurazione in cui i vertici sono divisi in due partizioni, indicate con $\mathcal{S}$ e $\mathcal{V}\backslash\mathcal{S}$. Per ogni coppia di vertici $[i, j]$ adiacenti con $i \in \mathcal{S}$ e $j \in \mathcal{V}\backslash\mathcal{S}$, la configurazione con una sola partizione è preferibile per un termine $k + 1$ a causa del primo e terzo addendo della funzione obiettivo; per ogni coppia di vertici $[i, j]$ non adiacenti con $i \in \mathcal{S}$ e $j \in \mathcal{V}\backslash\mathcal{S}$, la configurazione con due partizioni è preferibile per un termine pari a 1 a causa del secondo addendo della funzione obiettivo. Pertanto, per rendere più conveniente la configurazione con una sola componente, il valore di $k$ deve essere abbastanza grande da soddisfare la condizione

$$(k+1)|\mathcal{T}(\mathcal{S})| \geq |\overline{\mathcal{T}}(\mathcal{S})| \quad \forall\mathcal{S}\subset\mathcal{V} : \mathcal{S}\neq\emptyset$$

dove $\mathcal{T}(\mathcal{S})$ indica il taglio che separa $\mathcal{S}$ da $\mathcal{V}\backslash\mathcal{S}$ e $\overline{\mathcal{T}}(\mathcal{S})$ indica il suo complemento. La cardinalità di $\overline{\mathcal{T}}(\mathcal{S})$ è data da $|\mathcal{S}||\mathcal{V}\backslash\mathcal{S}| - |\mathcal{T}(\mathcal{S})|$. Usando variabili binarie $y$ per identificare quali vertici appartengono ad $\mathcal{S}$ si può scrivere un modello di programmazione matematica come segue:

$$\text{maximize } k = \frac{(\sum_{i\in\mathcal{V}} y_i)(\sum_{i\in\mathcal{V}}(1-y_i)) - \sum_{i,j\in\mathcal{V}} a_{ij}y_i(1-y_j)}{\sum_{i,j\in\mathcal{V}} a_{ij}y_i(1-y_j)} - 1 \tag{1}$$

$$\text{subject to } \sum_{i\in\mathcal{V}} y_i \geq 1 \qquad \forall i\in\mathcal{V} \tag{2}$$

$$y_i \in \{0,1\} \qquad \forall i\in\mathcal{V}. \tag{3}$$

Il modello è non lineare e può essere risolto con un apposito solutore.

Volendo seguire un approccio meno generale, ma valido solo per l'esempio in esame, si possono considerare le componenti ottenute ad esempio con $k = 5$ e calcolare per ciascuna di essere il valore di $k$ per cui diventerebbe conveniente fonderle con il resto del grafo. La componente che genera il valore più alto è data da $\mathcal{S} = \{5, 6, 7, 11, 17\}$ per la quale si ha $|\mathcal{S}| = 5$, $|\mathcal{V}\backslash\mathcal{S}| = 29$, $|\mathcal{T}(S)| = 4$ e quindi $k \geq \frac{29*5-4}{4} - 1$ cioè $k \geq 34,25$. Per $k = 35$ si verifica che la soluzione ottima risulta composta da un'unica componente, mentre con $k \leq 34$ la componente $\{5, 6, 7, 11, 17\}$ rimane separata nella soluzione ottima.

**Commenti.** Il problema si formula su un grafo simmetrico e richiede quindi la rappresentazione del grafo in termini di matrice di adiacenza. Il problema ammette formulazioni diverse, lineari e quadratiche, con insiemi di variabili diversi, e richiede comunque l'uso di variabili binarie per rappresentare condizioni logiche (if...then...). La risoluzione è richiesta per 5 valori diversi di $k$ e porta quindi all'introduzione del concetto di *parametro* come grandezza data dal punto di vista dell'algoritmo ma variabile dal punto di vista del modello. Il problema si presta ad illustrare i limiti del foglio elettronico nella rappresentazione dei problemi e ad introdurre quindi i linguaggi di modellizzazione algebrici come MathProg (gratuito) e i relativi solutori come GLPK (gratuito). Per rispondere all'ultima domanda si richiede di introdurre la nozione di componente di un grafo e di *taglio*. La risposta può essere data in particolare analizzando il grafo dato oppure in generale formulando e risolvendo un modello di programmazione non-lineare con variabili binarie.

# 9 Piano d'emergenza

**Dati.** I dati del problema sono i seguenti:

- un insieme $\mathcal{Z}$ di zone;
- un numero di abitanti $a_i$ per ogni zona $i \in \mathcal{Z}$;
- un costo di installazione $f_i$ per ogni zona $i \in \mathcal{Z}$;
- un coefficiente di rischio $r_i$ per ogni zona $i \in \mathcal{Z}$;
- una matrice di distanze $d_{ij}$ per ogni coppia di zone $i, j \in \mathcal{Z}$;
- una capacità massima $Q$ per ogni sito di emergenza (500 persone);
- una frazione stimata $\eta$ di popolazione da alloggiare in caso di emergenza (25%);
- un costo unitario di trasporto $c$ per ogni chilometro e per ogni persona (3 Euro).

**Variabili.** Le variabili del problema rappresentano le scelte di localizzazione dei siti di emergenza e di assegnamento degli sfollati ai siti. Ci sono quindi due distinti insiemi di variabili:

- Variabili $y$ di localizzazione (binarie): $y_i = 1$ se e solo se nella zona $i \in \mathcal{Z}$ viene installato un sito di emergenza;
- Variabili $x$ di assegnamento (intere): $x_{ij}$ è il numero di persone provenienti dalla zona $i \in \mathcal{Z}$ che vengono assegnate al sito della zona $j \in \mathcal{Z}$.

**Vincoli.** I vincoli del problema impongono che:

- siano assegnati ai siti di emergenza tutti gli sfollati provenienti da ciascuna zona, pari ad una frazione $\eta$ degli abitanti:

$$\sum_{j \in \mathcal{Z}} x_{ij} = \eta a_i \quad \forall i \in \mathcal{Z}$$

- a nessuna zona vengano assegnate più persone di quante ne può ospitare. La capacità massima di ogni zona $j \in \mathcal{Z}$ è pari a $Q$ se la zona $j$ ospita un sito di emergenza ($y_j = 1$) e nulla altrimenti ($y_j = 0$):

$$\sum_{i \in \mathcal{Z}} x_{ij} \leq Q y_j \quad \forall j \in \mathcal{Z}$$

**Obiettivo.** L'obiettivo è di minimizzare i costi, che sono dati da due termini distinti: i costi di installazione dei siti e i costi di trasporto degli sfollati. I primi dipendono dalle scelte di localizzazione, cioè dalle variabili $y$, i secondi dipendono dalle scelte di assegnamento, cioè dalle variabili $x$. I costi di localizzazione sono dati da

$$\sum_{j \in \mathcal{Z}} f_j y_j$$

mentre i costi di trasporto sono dati da

$$\sum_{i \in \mathcal{Z}} \sum_{j \in \mathcal{Z}} c\ d_{ij}\ x_{ij}$$

Per il fatto che i termini noti dei vincoli sono interi e per la struttura del problema (problema di trasporto a costo minimo), non è necessario imporre esplicitamente che le variabili $x$ siano intere: la soluzione di costo minimo ha sicuramente valori interi per le variabili $x$. Basta imporre che esse siano non-negative.

Il solutore di Excel non è in grado di gestire correttamente un esempio di queste dimensioni, che invece viene risolto facilmente da un solutore come GLPK. La soluzione ottima prevede l'installazione di 8 siti nelle zone 1, 2, 9, 12, 13, 15, 18 e 20 e ha un costo minimo complessivo pari a 236960.30 Euro.

La domanda 2 richiede di valutare l'indice medio di rischio associato alla soluzione ed imporre che non superi un dato limite $\theta = 0,05$. Si richiede quindi di introdurre il vincolo

$$\frac{\sum_{j \in \mathcal{Z}} (r_j \sum_{i \in \mathcal{Z}} x_{ij})}{\sum_{i \in \mathcal{Z}} \eta a_i} \leq \theta.$$

Questo vincolo, che è violato dalla soluzione ottima calcolata precedentemente, fa venir meno la garanzia che la soluzione ottima abbia valori interi per tutte le variabili $x$. Tuttavia la soluzione ottima è ancora intera anche in questo secondo caso, anche senza imporre il vincolo di integralità. Il costo sale a 259829,00 Euro e la scelta delle zone in cui localizzare i siti di emergenza è diversa rispetto allo scenario precedente.

La terza domanda richiede di scegliere alcuni siti da mettere in sicurezza e poi di risolvere ancora il problema con un vincolo sul costo complessivo del piano di emergenza e con l'obiettivo di minimizzare l'indice di rischio. A tal fine vengono introdotti ulteriori dati:

- la messa in sicurezza di un sito costa $s = 1000000$ Euro
- il budget complessivo disponibile è $b = 3000000$ Euro

Alle variabili decisionali del problema occorre quindi aggiungere ulteriori variabili binarie $z$ che indicano la scelta di quali siti mettere in sicurezza: $z_i = 1$ se e solo se viene messo in sicurezza il sito $i \in \mathcal{Z}$.

La funzione obiettivo dei casi precedenti diviene ora il primo membro del vincolo sul budget, nel quale occorre tener conto anche delle spese per la messa in sicurezza dei siti scelti:

$$\sum_{j \in \mathcal{Z}} f_j y_j + \sum_{i \in \mathcal{Z}} \sum_{j \in \mathcal{Z}} c x_{ij} + \sum_{j in \mathcal{Z}} s z_j \leq b.$$

La funzione obiettivo è ora l'indice di rischio, come definito nel vincolo introdotto nel caso precedente

$$\min \sum_{j \in \mathcal{Z}} (r_j * (1 - z_j) * \sum_{i \in \mathcal{Z}} x_{ij})$$

dove il rischio delle zone messe in sicurezza è nullo. Con questa formulazione della funzione obiettivo il problema diventa non-lineare, poiché nel modello compare il prodotto tra le variabili $z$ e $x$. Di conseguenza non è possibile risolverlo con solutori come GLPK che richeidono in ingresso modelli lineari.

Un'alternativa per formulare il problema in modo da ottenere un modello lineare è quella di sdoppiare tutte le varibili $x$, in variabili $x'$ e $x''$, rappresentanti rispettivamente i flussi di persone verso siti a rischio e verso siti messi in sicurezza. Il modello del problema è dunque il seguente:

$$\begin{aligned}
\text{minimize } \rho = & \frac{\sum_{j \in \mathcal{Z}} (r_j \sum_{i \in \mathcal{Z}} x'_{ij})}{\sum_{i \in \mathcal{Z}} \eta a_i} && && (4) \\
\text{subject to } & \sum_{j \in \mathcal{Z}} f_j y_j + \sum_{i \in \mathcal{Z}} \sum_{j \in \mathcal{Z}} c d_{ij} (x'_{ij} + x''_{ij}) + && \sum_{j in \mathcal{Z}} (s + f_j) z_j \leq b && (5) \\
& \sum_{i \in \mathcal{Z}} x'_{ij} \leq Q y_j && \forall j \in \mathcal{Z} && (6) \\
& \sum_{i \in \mathcal{Z}} x''_{ij} \leq Q z_j && \forall j \in \mathcal{Z} && (7) \\
& y_j + z_j \leq 1 && \forall j \in \mathcal{Z} && (8) \\
& \sum_{j \in \mathcal{Z}} (x'_{ij} + x''_{ij}) = \eta a_i && \forall i \in \mathcal{Z} && (9) \\
& y_j \in \{0, 1\} && \forall j \in \mathcal{Z} && (10) \\
& z_j \in \{0, 1\} && \forall j \in \mathcal{Z} && (11) \\
& x'_{ij} \geq 0 && \forall i \in \mathcal{Z} \forall j \in \mathcal{Z} && (12) \\
& x''_{ij} \geq 0 && \forall i \in \mathcal{Z} \forall j \in \mathcal{Z} && (13)
\end{aligned}$$

Risolvendo questo modello con GLPK si ottiene un valore minimo di rischio pari a circa 0.0175, mettendo in sicurezza le zone 7 e 9 e attivando siti di emergenza anche in quasi tutte le altre, poiché il vincolo sul budget lo consente.

**Commenti.** Questo esercizio si svolge in tre parti distinte, corrispondenti alle tre domande. Nella prima parte si richiede di formulare un problema con due insiemi di variabili, uno dei quali con vincolo di integralità (variabili binarie) e di combinare due voci di costo nella funzione obiettivo. Nella seconda parte si richiede di aggiungere un ulteriore vincolo, relativo all'indice di rischio, e risolvere nuovamente il problema. Nella terza parte l'indice di rischio viene assunto come funzione obiettivo mentre i costi vengono assoggettati ad un vincolo di budget. In questa terza parte inoltre si introduce il concetto di programmazione bi-livello, poiché esistono due livelli decisionali, il primo dei quali è relativo alla scelta di quali siti mettere in sicurezza. Il terzo modello si può formulare sia in modo non-lineare, sia - meno intuitivamente - in modo lineare. In tutti e tre i casi la dimensione dell'esempio richiede l'uso di un solutore di PLI come GLPK.

# 10 Test diagnostico

**Dati.** I dati del problema sono i seguenti:

- un insieme $\mathcal{S}$ di campioni sani;
- un insieme $\mathcal{M}$ di campioni malati;
- una coppia di coordinate $(x_i, y_i)$ per ogni campione $i \in \mathcal{S} \cup \mathcal{M}$.

**Variabili (1).** Si vuole determinare una retta in un piano cartesiano. Essa è definita da tre coefficienti $a$, $b$ e $c$ che determinano la sua equazione $ax+by+c = 0$. E' necessario usare l'equazione della retta in forma generale poiché non ci sono garanzie che la retta non sia orizzontale o verticale. I coefficienti sono legati tra di loro da una condizione di normalizzazione: ad esempio $a^2 + b^2 = 1$.

**Vincoli.** Per imporre che i campioni sani giacciano tutti da una parte della retta e quelli malati dall'altra parte bisogna imporre i vincoli

$$ax_i + by_i + c \leq 0 \quad \forall i \in \mathcal{S} \qquad \text{e} \qquad ax_i + by_i + c \geq 0 \quad \forall i \in \mathcal{M}.$$

**Obiettivo (1).** Il problema non ha funzione obiettivo; è un problema di *esistenza*, non di *ottimizzazione*. Risolvendo questo esempio non si trova alcuna soluzione ammissibile, poiché i due insiemi di punti dati nel piano cartesiano non sono *linearmente separabili*. Per questo motivo è data nel testo dell'esercizio la possibilità di trascurare alcuni dei campioni e si vuole trascurarne il minor numero possibile.

**Varaibili (2).** Si introduce quindi un insieme di variabili binarie $z_i$ che indica se il campione $i \in \mathcal{S} \cup \mathcal{M}$ va trascurato ($z_i = 1$) o no ($z_i = 0$). I vincoli vengono quindi modificati come segue:

$$ax_i + by_i + c - Hz_i \leq 0 \quad \forall i \in \mathcal{S} \qquad \text{e} \qquad ax_i + by_i + c + Hz_i \geq 0 \quad \forall i \in \mathcal{M}$$

dove $H$ è un numero molto grande. In questo modo ogni volta che $z_i = 1$ il termine $Hz_i$ fa sì che il vincolo sia sempre soddisfatto, indipendentemente da dove cade il punto $i$-esimo rispetto alla retta.

**Obiettivo (2).** Inoltre si ha in questo caso una funzione obiettivo:

$$\text{minimize} \sum_{i \in \mathcal{S} \cup \mathcal{M}} z_i.$$

Ne risulta un problema di ottimizzazione non-lineare (a causa della condizione di normalizzazione) con variabili binarie. La soluzione ottima si ha trascurando 7 campioni.

Per rispondere all'ultima domanda è necessario introdurre un'ulteriore variabile $\delta$, indicante la minima distanza tra un campione (non trascurato) e la retta. Si vuole quindi massimizzare il valore di $\delta$, imponendo tuttavia che la distanza tra punto e retta sia sempre maggiore o uguale a $\delta$ per tutti i campioni non trascurati. I vincoli del problema divengono pertanto:

$$ax_i + by_i + c - Hz_i \leq -\delta \quad \forall i \in \mathcal{S} \qquad \text{e} \qquad ax_i + by_i + c + Hz_i \geq \delta \quad \forall i \in \mathcal{M}.$$

Il numero di campioni trascurati, che era la funzione obiettivo nel caso precedente, ora diventa un vincolo:

$$\sum_{i \in \mathcal{S} \cup \mathcal{M}} z_i \leq z^*$$

dove il valore ottimo $z^*$ è stato ottenuto in precedenza ed è pari a 7.

La soluzione ottima si ottiene trascurando 4 campioni sani e 3 malati e dà un valore di $\delta$ pari a $0,213$.

**Commenti.** Questo esercizio richiede di applicare alcuni concetti di base di geometria del piano: in particolare l'equazione di un retta nel piano, l'appartenenza di un punto ad un semipiano e la distanza di un punto da una retta. Consente inoltre di introdurre il concetto di normalizzazione per garantire l'unicità di una soluzione. Esso richiede dapprima la formulazione di un problema di esistenza che non ha soluzione e successivamente la formulazione di due problemi di ottimizzazione non-lineari, dove le variabili binarie vengono usate per annullare alcuni vincoli e dove l'obiettivo nel primo modello diventa un vincolo nel secondo.