

Gli schemi di stima

La stima di relazioni algebriche lineari da dati affetti da rumore additivo

Si consideri una sequenza di N campioni delle n variabili

$$x_1 \ x_2 \ \dots \ x_n \quad (1)$$

con $N > n$. Eventuali relazioni lineari tra tali grandezze sono descritte da equazioni del tipo

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = 0 \quad (2)$$

e, detta X la matrice ($N \times n$) avente colonne formate dalle osservazioni di tali variabili, le relazioni precedenti possono essere rappresentate dalle colonne, linearmente indipendenti, di una matrice A ($n \times q$) tale che

$$X A = 0. \quad (3)$$

Il numero di relazioni lineari che possono venire associate ai dati è $q = n - \text{rank } X$. Quando i dati a disposizione sono corrotti da rumore, il rango di X risulta sempre uguale ad n e X non risulta quindi compatibile con legami lineari tra le osservazioni.

Le considerazioni precedenti possono essere svolte anche sostituendo alla matrice ($N \times n$) X , la matrice ($n \times n$) Σ data da

$$\Sigma = \frac{X^T X}{N} \quad (4)$$

coincidente, nell'ipotesi di variabili a valor medio nullo, con la matrice di covarianza campionaria delle variabili considerate. In assenza di rumore e per variabili legate da relazioni lineari risulterà

$$\Sigma \geq 0 \quad (5)$$

e ogni soluzione, A , di rango massimo, dell'equazione

$$\Sigma A = 0 \quad (6)$$

sarà una base di $\ker \Sigma$. In assenza di relazioni lineari tra le variabili e/o in presenza di rumore risulterà

$$\Sigma > 0 \quad (7)$$

e sarà possibile estrarre relazioni lineari da Σ solo modificandola.

Definizione 1 – Un procedimento sistematico per estrarre relazioni lineari da dati affetti da rumore viene definito *schema*.

Per analizzare le caratteristiche dei diversi schemi di stima è necessario introdurre alcune ipotesi sul rumore presente sui dati; si assumerà che:

1) Il rumore sia di tipo additivo; ogni variabile x_i ($i = 1, 2, \dots, n$), risulta quindi definita come

$$x_i = \hat{x}_i + \tilde{x}_i \quad (8)$$

dove i termini incogniti, a destra del segno di uguaglianza, sono rispettivamente il valore esatto dell' i -esima variabile, \hat{x}_i , ed il corrispondente rumore, \tilde{x}_i .

2) Il valor medio delle sequenze di campioni \hat{x}_i e \tilde{x}_i sia nullo; risulti cioè

$$\frac{1}{N} \sum_{t=1}^N \hat{x}_{it} = 0 \quad (9)$$

$$\frac{1}{N} \sum_{t=1}^N \tilde{x}_{it} = 0. \quad (10)$$

3) Le sequenze di campioni del rumore siano ortogonali alle sequenze di campioni delle variabili prive di rumore; vale quindi la relazione

$$\sum_{t=1}^N \tilde{x}_{it} \hat{x}_{jt} = 0 \quad \text{per ogni } i, j. \quad (11)$$

Definendo in maniera ovvia le matrici \hat{X} , \tilde{X} , $\hat{\Sigma}$ e $\tilde{\Sigma}$, le ipotesi precedenti portano immediatamente alle seguenti proprietà

$$X = \hat{X} + \tilde{X} \quad (12)$$

$$\hat{X}^T \tilde{X} = 0 \quad (13)$$

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} \quad (14)$$

$$\Sigma > 0 \quad (15)$$

$$\tilde{\Sigma} \geq 0 \quad \text{o} \quad \tilde{\Sigma} > 0 \quad (16)$$

$$\hat{\Sigma} \geq 0. \quad (17)$$

Il problema della determinazione delle relazioni lineari compatibili con i dati affetti da rumore può essere formulato in termini matematici come segue.

Problema 1 – Data la matrice di covarianza campionaria dei dati affetti da rumore, Σ , determinare le matrici di covarianza dei rumori, $\tilde{\Sigma}$, definite o semidefinite positive che soddisfano la condizione

$$\hat{\Sigma} = \Sigma - \tilde{\Sigma} \geq 0. \quad (18)$$

Le corrispondenti soluzioni saranno quindi descritte da una qualunque base del sottospazio $\ker \hat{\Sigma}$.

Si noti come la formulazione del Problema 1 sia del tutto generale, non introduca cioè ulteriori ipotesi rispetto a quelle, relativamente blande, sulla additività e sulla indipendenza dei rumori. Tale formulazione

non implica l'unicità della soluzione che deriva, in genere, dalla introduzione, nei singoli schemi, di ipotesi particolari; non introduce, peraltro, neppure vincoli sul numero di relazioni lineari che possono legare le variabili \hat{x}_i . Si osservi inoltre come la formulazione data tratti tutte le variabili in modo simmetrico: non vengono cioè introdotte, a priori, ipotesi particolari sulla ripartizione del rumore sulle diverse variabili.

Lo schema dei minimi quadrati ordinari

Lo schema dei minimi quadrati ordinari (OLS, Ordinary Least Squares o, più semplicemente, LS, Least Squares) è basato sulla ipotesi addizionale che una sola delle variabili sia affetta da rumore cioè che la matrice di covarianza del rumore sia del tipo

$$\tilde{\Sigma} = \text{diag}[0 \dots 0 \tilde{\sigma}_i^2 0 \dots 0]. \quad (19)$$

In genere viene considerata affetta da rumore la variabile assunta come dipendente, il cui coefficiente nella relazione di lineare dipendenza viene normalizzato ad 1, e questa costituisce una ulteriore ipotesi. In tali condizioni la soluzione del Problema 1 è unica e data, considerando come affetta da rumore la variabile i -esima, dalla nota formula dei minimi quadrati

$$a = (X_i^T X_i)^{-1} X_i^T x_i^\circ \quad (20)$$

ove X_i è la matrice che si ottiene da X eliminando la i -esima colonna, x_i° indica il vettore dei coefficienti di dipendenza della proiezione ortogonale di x_i° sul sottospazio generato dalle colonne di X_i da tali vettori.

La soluzione dei minimi quadrati può venire ricavata anche imponendo la condizione di singolarità (18) su $\hat{\Sigma}$. Si consideri, per semplicità di notazione, come affetta da rumore la variabile n -esima; tale ipotesi non limita in alcun modo le considerazioni che seguono essendo sempre possibile ottenerla mediante una permutazione delle variabili. Si consideri poi la matrice $\hat{\Sigma}$ (18) partizionata come segue

$$\hat{\Sigma} = \begin{bmatrix} \Sigma^\# & \Sigma^{\#\#} \\ \Sigma^{*\#} & \sigma_n^2 - \tilde{\sigma}_n^2 \end{bmatrix}. \quad (21)$$

Utilizzando il Lemma sul determinante delle matrici partizionate (vedi Appendice) si ottiene

$$\det \hat{\Sigma} = \det \Sigma^\# [\sigma_n^2 - \tilde{\sigma}_n^2 - \Sigma^{*\#} \Sigma^{\#-1} \Sigma^{\#\#}]; \quad (22)$$

dalla relazione precedente si ricava il valore di $\tilde{\sigma}_n^2$ che rende semidefinita $\hat{\Sigma}$

$$\tilde{\sigma}_n^2 = \sigma_n^2 - \Sigma^{*\#} \Sigma^{\#-1} \Sigma^{\#\#} = \frac{\det \Sigma}{\det \Sigma^\#}. \quad (23)$$

In generale quindi la varianza del rumore additivo sulla variabile i -esima che si ottiene considerando le variabili rimanenti prive di rumore è data da

$$\tilde{\sigma}_i^2 = \frac{\det \Sigma}{\det \Sigma_i} \quad (24)$$

ove Σ_i indica la matrice che si ottiene da Σ eliminando la i -esima riga e la i -esima colonna. Si noti come il valore fornito dalla (24) descriva l'entità massima del rumore che può essere presente sulla i -esima variabile; un valore maggiore non sarebbe infatti compatibile con la condizione $\hat{\Sigma} \geq 0$. Nello schema dei minimi quadrati l'unicità della soluzione corrispondente alla ipotesi di rumore presente su una sola variabile deriva dalla unicità della soluzione del problema algebrico di rendere semidefinita $\hat{\Sigma}$ operando su un solo elemento della diagonale principale; sottraendo infatti quantità minori del valore fornito dalla (24) $\hat{\Sigma}$ resta definita positiva, sottraendo quantità maggiori alcuni autovalori di $\hat{\Sigma}$ diventano negativi. In entrambi i casi non si ha alcuna soluzione.

Osservazione 1 – È di interesse valutare le ragioni che portano usualmente, nella applicazione dello schema dei minimi quadrati, ad assumere tutto il rumore sulla variabile che si esprime come combinazione lineare delle rimanenti; questa operazione porta a calcolare coefficienti che minimizzano la somma dei quadrati degli scarti tra le misure (affette da rumore o meno) di tale variabile e la combinazione lineare delle misure (affette da rumore o meno) delle variabili rimanenti. In questo contesto l'obiettivo non è quindi quello di cercare di stimare i parametri della relazione lineare che si suppone legghi le variabili prive di rumore \hat{x}_i ma di ottenere una relazione che legghi le misure affette da rumore, x_i , minimizzando un criterio quadratico.

Se si considera, al di fuori del contesto precedente, lo schema dei minimi quadrati come legato alla ipotesi di una sola variabile affetta da rumore, ne segue che questo schema porta ad un insieme formato dalle n diverse soluzioni che si ottengono attribuendo, a turno, tutto il rumore alle singole variabili.

Esempio 1

Si consideri la matrice di covarianza ottenuta da dati affetti da rumore

$$\Sigma = \begin{bmatrix} 14 & -6 & -4 \\ -6 & 6.5 & 1 \\ -4 & 1 & 5 \end{bmatrix} \quad (25)$$

con spettro dato da $\lambda_1 = 2.4245$, $\lambda_2 = 4.5688$ e $\lambda_3 = 18.5067$. Le varianze dei rumori massimi su x_1 , x_2 e x_3 sono

$$\tilde{\sigma}_1^2 = \det[\Sigma] / \det[\Sigma_1] = 205/31.5 = 6.5079$$

$$\tilde{\sigma}_2^2 = \det[\Sigma] / \det[\Sigma_2] = 205/54 = 3.7963$$

$$\tilde{\sigma}_3^2 = \det[\Sigma] / \det[\Sigma_3] = 205/55 = 3.7273.$$

I coefficienti a_1 , a_2 ed a_3 dei corrispondenti modelli OLS si possono ottenere come basi di $\ker[\Sigma - \tilde{\Sigma}]$ considerando le matrici di covarianza del rumore $\tilde{\Sigma} = \text{diag}[\tilde{\sigma}_1^2 \ 0 \ 0]$, $\tilde{\Sigma} = \text{diag}[0 \ \tilde{\sigma}_2^2 \ 0]$ e $\tilde{\Sigma} = \text{diag}[0 \ 0 \ \tilde{\sigma}_3^2]$. Normalizzando a_3 ad 1, considerando cioè modelli del tipo $a_1x_1 + a_2x_2 + x_3 = 0$, si ottengono i modelli

$$\begin{array}{lll} a_1 = 1.5750 & a_1 = 2.6000 & a_1 = 0.3636 \\ a_2 = 1.3000 & a_2 = 5.4000 & a_2 = 0.1818. \end{array}$$

Lo schema dei minimi quadrati classici

Lo schema dei minimi quadrati classici (CLS, Classic Least Squares) può venire considerato come una generalizzazione di quello dei minimi quadrati ordinari e viene usato per estrarre q relazioni di lineare dipendenza dalle misure delle n variabili disponibili. Anche nello schema CLS vengono assunte affette da rumore le q variabili considerate come dipendenti dalle rimanenti $n - q$. Supponendo, senza con questo limitare la generalità della trattazione, che le variabili prive di rumore siano le prime $n - q$, si partizioni la matrice X come segue

$$X = [X^\# \ X^*], \quad (26)$$

ove $X^\#$ ($N \times n - q$) contiene le misure prive di rumore e X^* ($N \times q$) quelle affette da rumore (dipendenti). La soluzione (unica) fornita dallo schema è ancora data dalla formula dei minimi quadrati

$$A = (X^{\#T} X^\#)^{-1} X^{\#T} X^*. \quad (27)$$

Le colonne di A descrivono i coefficienti di lineare dipendenza delle proiezioni ortogonali delle colonne di X^* sul sottospazio generato dalle colonne di $X^\#$ da tali vettori. Dalla partizione (26) di X segue

$$\tilde{\Sigma} = \begin{bmatrix} 0 & 0 \\ 0 & \tilde{\Sigma}_q \end{bmatrix}, \quad (28)$$

$$\hat{\Sigma} = \begin{bmatrix} \Sigma^\# & \Sigma^{\#\#} \\ \Sigma^{*\#} & \Sigma^* - \tilde{\Sigma}_q \end{bmatrix} \quad (29)$$

e le q soluzioni sono descritte da qualunque base di $\ker \hat{\Sigma}$. Applicando il lemma sul determinante delle matrici partizionate (vedi Appendice) a $\hat{\Sigma}$ si ottiene

$$\det \hat{\Sigma} = \det \Sigma^\# \det [\Sigma^* - \tilde{\Sigma}_q - \Sigma^{*\#} \Sigma^{\#-1} \Sigma^{\#\#}]; \quad (30)$$

la condizione $\text{rank } \hat{\Sigma} = n - q$ può venire soddisfatta imponendo la condizione

$$\Sigma^* - \tilde{\Sigma}_q - \Sigma^{*\#} \Sigma^{\#-1} \Sigma^{\#\#} = 0 \quad (31)$$

che, essendo $\Sigma^\# > 0$ e $\Sigma^* - \tilde{\Sigma}_q - \Sigma^{*\#} \Sigma^{\#-1} \Sigma^{\#\#}$ simmetrica, risulta necessaria e sufficiente per la presenza di $n - q$ autovalori nulli in $\hat{\Sigma}$. Dalla (31), infine, si ricava l'espressione di $\tilde{\Sigma}_q$

$$\tilde{\Sigma}_q = \Sigma^* - \Sigma^{*\#} \Sigma^{\#-1} \Sigma^{\#\#}. \quad (32)$$

La matrice $\tilde{\Sigma}_q$ (32) può anche venire data in una forma alternativa osservando che, in base al lemma sull'inversa delle matrici partizionate (vedi Appendice), risulta

$$\Sigma^{-1} = \begin{bmatrix} \Delta & \Delta \\ \Delta & (\Sigma^* - \Sigma^{*\#} \Sigma^{\#-1} \Sigma^{\#\#})^{-1} \end{bmatrix} \quad (33)$$

ove si sono indicate con il simbolo Δ sottomatrici che non interessano in questo contesto. In base alla (33) $\tilde{\Sigma}_q$ può venire definita come l'inversa della matrice che si ottiene da Σ^{-1} eliminando le prime $n - q$ righe e colonne.

Osservazione 2 – Si è già osservato come qualunque base di $\ker \hat{\Sigma}$ definisca univocamente la soluzione; in generale tuttavia le soluzioni corrispondenti alle diverse basi includono, nelle relazioni di lineare dipendenza da loro definite, tutte le colonne di X . Indicando con

$$B = \begin{bmatrix} B^\# \\ B^* \end{bmatrix} \quad (34)$$

una arbitraria base di $\ker \hat{\Sigma}$, nella quale $B^\#$ denota le prime $n - q$ righe e B^* le rimanenti q , è immediato verificare come la soluzione, A , fornita dalla (27) sia data da

$$A = -B^\# B^{*-1}. \quad (35)$$

Osservazione 3 – La soluzione CLS può anche venire ricavata applicando q volte lo schema OLS includendo in X la sottomatrice $X^\#$ e, a turno, le colonne di X^* .

Osservazione 4 – Si noti come nella soluzione CLS la matrice $\tilde{\Sigma}_q$ non risulti, in generale, diagonale. Non viene quindi introdotta alcuna ipotesi di ortogonalità tra le colonne (non nulle) di \tilde{X} , ed i rumori presenti sulle q variabili considerate dipendenti possono risultare correlati.

Esempio 2

Si consideri ancora la matrice di covarianza (25); si applicherà lo schema CLS per determinare due relazioni di lineare dipendenza tra x_2 e x_3 , supposte affette da rumore, e x_1 , supposto privo di rumore; nel caso in esame risulta pertanto $n = 3$ e $q = 2$. La matrice $\tilde{\Sigma}_2$ può essere calcolata mediante la (32) o, indifferentemente, invertendo la matrice che si ottiene da Σ^{-1} cancellando la prima riga e la prima colonna; si ottiene

$$\tilde{\Sigma}_2 = \begin{bmatrix} 3.9286 & -0.7143 \\ -0.7143 & 3.8571 \end{bmatrix}.$$

Una base ortogonale di $\ker \hat{\Sigma}$ è data da

$$B = \begin{bmatrix} -0.4469 & 0.0996 \\ -0.8427 & -0.3805 \\ -0.3003 & 0.9194 \end{bmatrix};$$

gli elementi delle colonne di B descrivono due relazioni lineari tra le colonne di X . Utilizzando la (35) si ottiene la matrice

$$A = -\begin{bmatrix} -0.4469 & 0.0996 \end{bmatrix} \begin{bmatrix} -0.8427 & -0.3805 \\ -0.3003 & 0.9194 \end{bmatrix}^{-1} = \begin{bmatrix} -0.4286 & -0.2857 \end{bmatrix}$$

che descrive le relazioni lineari

$$\begin{aligned} x_2 &= -0.4286 x_1 \\ x_3 &= -0.2857 x_1. \end{aligned}$$

Lo schema dei minimi quadrati totali

Lo schema dei minimi quadrati totali (TLS, Total Least Squares) estrae una singola relazione di lineare dipendenza dalle misure affette da rumore, X , determinando la matrice \hat{X} di rango $n - 1$ che minimizza la norma di Frobenius di $X - \hat{X}$; la soluzione è cioè determinata da

$$\min_{\hat{X}} \|X - \hat{X}\|_F. \quad (36)$$

L'obiettivo di tale schema è quello di ottenere una soluzione modificando il meno possibile i dati affetti da rumore. Ricordando che la norma di Frobenius di una matrice M ($m \times n$) è data da

$$\|M\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n m_{ij}^2} = \sqrt{\text{tr}(M^T M)}, \quad (37)$$

ne segue che la soluzione dello schema TLS è caratterizzata da

$$\min_{\tilde{\Sigma} \in \mathcal{S}_{\Sigma}} \text{tr} \tilde{\Sigma} \quad (38)$$

ove \mathcal{S}_{Σ} indica l'insieme delle matrici $\tilde{\Sigma}$ che soddisfano la condizione $\hat{\Sigma} = \Sigma - \tilde{\Sigma} \geq 0$. Per meglio comprendere le interessanti proprietà geometriche della soluzione TLS si consideri il caso $n = 2$ nel quale la soluzione, $a_1 x_1 + a_2 x_2 = 0$, individua una retta nel piano delle variabili x_1 e x_2 . Se si utilizza lo schema OLS e si suppone tutto il rumore su x_1 si determina la retta che minimizza la somma dei quadrati delle distanze dai punti misurate secondo l'asse x_1 (Figura 1); analogamente, supponendo tutto il rumore su x_2 , si determina la retta che minimizza la somma dei quadrati delle distanze dai punti misurate secondo l'asse x_2 .

Lo schema TLS porta invece a determinare la retta che minimizza la somma dei quadrati delle distanze dei punti misurate ortogonalmente alla retta stessa.

Se il minimo autovalore di Σ , λ_m , ha molteplicità uno, la matrice $\tilde{\Sigma}$ che soddisfa la (38) è data da

$$\tilde{\Sigma} = \lambda_m v_m v_m^T \quad (39)$$

ove v_m denota l'autovettore associato a λ_m con norma Euclidea normalizzata ad 1. Risulta inoltre

$$\|\tilde{\Sigma}\|_F = \lambda_m \quad (40)$$

$$\text{rank} \tilde{\Sigma} = 1. \quad (41)$$

Esempio 3

Si faccia ancora riferimento alla matrice di covarianza (25). Per tale matrice risulta

$$\lambda_m = 2.4245,$$

$$v_m = \begin{bmatrix} 0.5248 \\ 0.6329 \\ 0.5693 \end{bmatrix},$$

$$\tilde{\Sigma} = \begin{bmatrix} 0.6677 & 0.8052 & 0.7243 \\ 0.8052 & 0.9711 & 0.8735 \\ 0.7243 & 0.8735 & 0.7857 \end{bmatrix}.$$

La corrispondente relazione lineare che si ottiene da $\hat{\Sigma} = \Sigma - \tilde{\Sigma}$ normalizzando a_3 ad 1, del tipo cioè $a_1x_1 + a_2x_2 + x_3 = 0$, è descritta dai parametri

$$a_1 = 0.9218$$

$$a_2 = 1.1117.$$

Il metodo dell'autovettore

Lo schema dell'autovettore (EM, Eigenvector Method) è basato sulla ipotesi addizionale che tutte le variabili considerate siano affette dalla stessa quantità di rumore e che le diverse sequenze di rumore siano tra loro ortogonali, valga cioè la condizione

$$\sum_{t=1}^N \tilde{x}_{it} \tilde{x}_{jt} = 0 \quad \text{per } i \neq j. \quad (42)$$

In tale situazione la matrice di covarianza del rumore è diagonale

$$\tilde{\Sigma} = \text{diag}[\sigma^2 \dots \sigma^2] \quad (43)$$

e, dovendo valere la (17), σ^2 è dato dal minimo autovalore, λ_m , di Σ ,

$$\sigma^2 = \lambda_m. \quad (44)$$

Nel caso in cui la molteplicità di λ_m sia pari ad uno, la corrispondente stima dei coefficienti della relazione lineare tra le variabili coincide, a meno di una normalizzazione, con gli elementi dell'autovettore, v_m , di Σ associato a λ_m .

Osservazione 5 – È interessante notare come lo schema EM porti, pur essendo basato su considerazioni essenzialmente diverse ed utilizzando matrici di covarianza del rumore diverse, alla stessa stima dello schema TLS. Nel caso in cui la molteplicità di λ_m sia pari ad uno, dalla relazione

$$(\Sigma - \lambda_m I) v_m = 0 \quad (45)$$

si ricava infatti, se $\|v_m\|_2^2 = 1$, la relazione

$$\Sigma v_m = \lambda_m v_m = \lambda_m v_m v_m^T v_m \quad (46)$$

dalla quale, infine,

$$(\Sigma - \lambda_m v_m v_m^T) v_m = 0. \quad (47)$$

Il confronto delle (45) e (47) evidenzia come v_m costituisca la soluzione di entrambi gli schemi.

Lo schema EM tratta le variabili in maniera simmetrica ma, diversamente dallo schema dei minimi quadrati, non fornisce risultati consistenti qualora si effettui una diversa messa in scala delle variabili x_i ; lo stesso vale, ovviamente, per lo schema TLS.

Esempio 4

Si consideri ancora la matrice di covarianza (25); il minimo autovalore ed il corrispondente autovettore (normalizzato) sono dati, come già visto, da

$$\lambda_m = 2.4245, \quad v_m = \begin{bmatrix} 0.5248 \\ 0.6329 \\ 0.5693 \end{bmatrix}.$$

Normalizzando a_3 ad 1, come negli esempi precedenti, si ottiene la stima EM

$$a_1 = 0.9218$$

$$a_2 = 1.1117$$

coincidente, come si è visto, con la stima TLS ottenuta nell'Esempio 3. Si utilizzi ora, per la variabile x_1 , una unità di misura pari ad un decimo della precedente lasciando invariate quelle relative a x_2 e x_3 ; la matrice di covarianza dei dati diventa

$$\Sigma' = \begin{bmatrix} 1400 & -60 & -40 \\ -60 & 6.5 & 1 \\ -40 & 1 & 5 \end{bmatrix} \quad (48)$$

con spettro dato da $\lambda_1 = 3.1697$, $\lambda_2 = 4.6074$ e $\lambda_3 = 1403.72$. Le varianze dei rumori massimi su x_1 , x_2 e x_3 sono

$$\tilde{\sigma}_1^2 = \det[\Sigma'] / \det[\Sigma'_1] = 20500/31.5 = 650.79$$

$$\tilde{\sigma}_2^2 = \det[\Sigma'] / \det[\Sigma'_2] = 20500/5400 = 3.7963$$

$$\tilde{\sigma}_3^2 = \det[\Sigma'] / \det[\Sigma'_3] = 20500/5500 = 3.7273;$$

come si vede le varianze dei rumori additivi su x_2 e x_3 non sono variate rispetto a quelle trovate nell'Esempio 1, mentre quella del rumore su x_1 risulta, correttamente, 100 volte superiore. I corrispondenti modelli OLS, con a_3 normalizzato ad 1, sono

$$a_1 = 0.1575 \quad a_1 = 0.2600 \quad a_1 = 0.03636$$

$$a_2 = 1.3000 \quad a_2 = 5.4000 \quad a_2 = 0.1818,$$

e coincidono, considerando la diversa scala di x_1 , con quelli trovati in precedenza. L'autovettore di norma Euclidea unitaria associato al minimo autovalore di Σ' è

$$v' = \begin{bmatrix} 0.0503 \\ 0.6893 \\ 0.7228 \end{bmatrix};$$

normalizzando a_3 ad 1 si ottiene la stima EM/TLS

$$\begin{aligned} a_1 &= 0.0696 \\ a_2 &= 0.9537, \end{aligned}$$

diversa da quella ottenuta con la messa in scala iniziale.

Lo schema delle componenti principali

Lo schema delle componenti principali (PC, Principal Components) è basato sulla scomposizione ai valori singolari della matrice di covarianza dei dati, Σ ,

$$\Sigma = U \Lambda U^T \quad (49)$$

ove

$$\Lambda = \text{diag} [\lambda_1 \dots \lambda_n] \quad (50)$$

è formata dagli autovalori di Σ e U è una matrice ortogonale; scambiando le colonne di U è possibile ordinare in maniera arbitraria gli autovalori di Σ sulla diagonale di Λ . Nello schema delle componenti principali si ottengono q ($1 \leq q \leq n - 1$) relazioni costruendo la matrice di covarianza del rumore

$$\tilde{\Sigma} = U \text{diag} [\lambda_1 \dots \lambda_q 0 \dots 0] U^T \quad (51)$$

che, sottratta a Σ , porta ad una matrice $\hat{\Sigma} \geq 0$ di rango $n - q$ con spettro dato da $0, \dots, 0, \lambda_{q+1}, \dots, \lambda_n$. La scelta (51) di $\tilde{\Sigma}$ consente quindi di annullare q arbitrari autovalori di Σ senza modificare i rimanenti; le q soluzioni vengono poi estratte da $\ker \hat{\Sigma}$.

Osservazione 6 – Lo schema delle componenti principali consente, come i minimi quadrati classici, di ottenere il numero di relazioni desiderato e, una volta fissato q , lascia ulteriori gradi di libertà nella scelta del sottoinsieme di q autovalori da utilizzare per costruire $\tilde{\Sigma}$. La logica usualmente seguita nella applicazione di tale schema consiste tuttavia nel selezionare i q autovalori più piccoli di Σ in modo da ottenere la soluzione con la minima perturbazione dei dati. Anche il numero di relazioni da estrarre viene determinato valutando quanti autovalori di Σ risultino sensibilmente inferiori ai rimanenti.

Osservazione 7 – Essendo Σ simmetrica, la matrice U ha come colonne gli autovettori normalizzati di Σ . Ne segue che per $q = 1$ lo schema PC porta, qualora si costruisca $\tilde{\Sigma}$ con il minimo autovalore di Σ , alla stessa stima degli schemi TLS ed EM.

Esempio 5

La scomposizione ai valori singolari della matrice di covarianza (25) porta alle matrici

$$U = \begin{bmatrix} -0.5248 & -0.0748 & 0.8480 \\ -0.6329 & -0.6319 & -0.4474 \\ -0.5693 & 0.7714 & -0.2842 \end{bmatrix}$$

$$\Lambda = \text{diag} [2.4245, 4.5688, 18.5067].$$

Si determineranno due relazioni lineari annullando i due autovalori più piccoli di Σ mediante la matrice di covarianza del rumore

$$\tilde{\Sigma} = U \text{diag} [2.4245, 4.5688, 0] U^T = \begin{bmatrix} 13.3323 & -6.8052 & -4.7243 \\ -6.8052 & 5.5289 & 0.1265 \\ -4.7243 & 0.1265 & 4.2143 \end{bmatrix}.$$

Una base ortogonale di $\ker \hat{\Sigma}$ è data da

$$B = \begin{bmatrix} -0.3629 & -0.3864 \\ -0.0991 & -0.8888 \\ -0.9266 & 0.2464 \end{bmatrix};$$

utilizzando la (35) per esprimere x_2 e x_3 in funzione di x_1 si ottengono le relazioni lineari

$$x_2 = -0.5276 x_1$$

$$x_3 = -0.3352 x_1.$$

Lo schema di Frisch

Tale schema, proposto da Frisch nel 1934, introduce, rispetto alle (12)–(17), l'ulteriore ipotesi (42) di ortogonalità delle sequenze di rumore presenti sulle diverse variabili, già vista per il metodo dell'autovettore. Lo schema di Frisch non introduce, tuttavia, alcuna ipotesi sull'ammontare dei rumori presenti sulle singole variabili che vengono trattate in maniera simmetrica. Nell'ambito dello schema di Frisch ogni matrice di covarianza del rumore, diagonale,

$$\tilde{\Sigma} = \text{diag} [\tilde{\sigma}_1^2 \dots \tilde{\sigma}_n^2] \quad (52)$$

che soddisfi la condizione (18) porta ad una soluzione.

Osservazione 8 – Le matrici di covarianza del rumore considerate dagli schemi OLS ed EM sono del tipo (52); ne segue che lo schema di Frisch è compatibile con una famiglia di soluzioni cui appartengono sia le stime OLS che la stima EM.

Osservazione 9 – Lo schema di Frisch non introduce ipotesi sul numero di relazioni lineari che possono venire estratte dai dati. L'Osservazione 8 evidenzia come il sottoinsieme delle soluzioni costituite da una singola relazione lineare non risulti mai vuoto e contenga più elementi. Nella famiglia delle soluzioni possono essere presenti anche sottoinsiemi di soluzioni costituite da più relazioni lineari; tali sottoinsiemi non conterranno, in generale, le soluzioni fornite dagli schemi CLS e PC poiché le matrici di covarianza previste da tali schemi non soddisfano la (52).

Definizione 2 – Si considerino le matrici di covarianza del rumore $\tilde{\Sigma}$ definite o semidefinite positive tali che

$$\text{rank} (\Sigma - \tilde{\Sigma}) = n - q_i \quad (53)$$

con $q_i \geq 1$; il massimo corank di Σ nell'ambito dello schema di Frisch viene definito come

$$\text{maxcor}_F(\Sigma) = \max_i \{ q_i \}. \quad (54)$$

In base alla definizione precedente, per ogni $\Sigma > 0$, $\text{maxcor}_F(\Sigma) \geq 1$; il massimo numero di relazioni lineari compatibili con un insieme di dati nello schema di Frisch è quindi uguale al massimo corank della corrispondente matrice di covarianza.

Osservazione 10 – La condizione $\text{maxcor}_F(\Sigma) = q > 1$ non implica che i dati privi di rumore siano legati proprio da q relazioni lineari ma solo che tale situazione va considerata come possibile ossia che soluzioni costituite da più relazioni lineari sono presenti nell'insieme delle soluzioni dello schema di Frisch.

Si noti inoltre che la condizione $\text{maxcor}_F(\Sigma) = q > 1$ non implica l'esistenza di q relazioni lineari indipendenti tra le colonne di \hat{X} , come viene messo in evidenza dal seguente esempio.

Esempio 6

Si consideri la matrice di covarianza

$$\Sigma = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 4 & 1 \\ 1 & 1 & 1 & 5 \end{bmatrix}; \quad (55)$$

è immediato verificare che $\text{maxcor}_F(\Sigma) = 3$ ($\text{rank}(\Sigma - \tilde{\Sigma}) = 1$ per $\tilde{\Sigma} = \text{diag}[1, 2, 3, 4]$). Tuttavia x_2, x_3 e x_4 sono legate a x_1 dalla stessa relazione lineare.

La dimensione dell'insieme delle soluzioni di Frisch per Σ è

$$p_i = \dim_F(\Sigma, q_i) = n - \frac{q_i(q_i + 1)}{2} \quad (56)$$

ove $q_i \leq q = \text{maxcor}_F(\Sigma)$. La famiglia delle soluzioni formate da una singola relazione lineare contiene pertanto ∞^{n-1} elementi, . . . , la famiglia delle soluzioni date da q relazioni lineari indipendenti contiene ∞^p elementi ove $p = \dim_F(\Sigma, q)$.

Osservazione 11 – Si noti come la relazione (56) faccia riferimento alle sole soluzioni indipendenti. Così nel caso della matrice di covarianza considerata nell'Esempio 6, si è trovato che $\text{maxcor}_F(\Sigma) = 3$ mentre la (56) indica come tale matrice non sia compatibile con soluzioni formate da più di due relazioni lineari indipendenti.

Di notevole rilevanza il caso $\text{maxcor}_F(\Sigma) = 1$ nel quale l'insieme delle soluzioni è formato da singole relazioni lineari; una condizione necessaria e sufficiente perché ciò avvenga è data dal seguente teorema.

Teorema 1 – $\text{maxcor}_F(\Sigma) = 1$ se e solo se tutti gli elementi di Σ^{-1} sono positivi o se questa condizione può essere ottenuta mediante opportuni cambiamenti di segno nella definizione delle variabili x_i .

Osservazione 12 – Una matrice ad elementi tutti positivi è anche detta di Frobenius. La verifica se Σ^{-1} è riconducibile ad una matrice di Frobenius cambiando il segno di alcune variabili può essere facilmente effettuata cambiando il segno alle variabili corrispondenti ad elementi negativi nella prima riga (colonna) di Σ^{-1} .

Quando è soddisfatta la condizione fornita dal Teorema 1 vale il seguente importante risultato che descrive l'intera famiglia delle soluzioni dello schema di Frisch in funzione delle n soluzioni fornite dallo schema OLS.

Teorema 2 – Se $\max_{cor_F}(\Sigma) = 1$, i coefficienti a_1, \dots, a_n di tutte le possibili relazioni lineari (2) compatibili con i dati sono contenute, normalizzando uno dei coefficienti (ad esempio a_n) ad 1, nel semplice con vertici individuati, nel primo ortante di \mathcal{R}^{n-1} , dai coefficienti a_1, \dots, a_{n-1} delle n soluzioni OLS.

Il Teorema 2 descrive, per il caso $\max_{cor_F}(\Sigma) = 1$, il luogo delle soluzioni nello spazio dei parametri; di rilevante interesse anche il corrispondente luogo delle soluzioni nello spazio del rumore descritto dal teorema che segue.

Teorema 3 – Nel caso $\max_{cor_F}(\Sigma) = 1$, le n -ple $(\tilde{\sigma}_1^2, \dots, \tilde{\sigma}_n^2)$ che definiscono matrici di covarianza $\tilde{\Sigma}$ (52) ammissibili, appartengono ad una (iper)superficie convessa $\mathcal{S}(\Sigma)$, situata nel primo ortante di \mathcal{R}^n , che interseca i piani paralleli ai piani coordinati secondo segmenti di iperbole.

Dimostrazione – Si consideri una matrice di covarianza del rumore $\tilde{\Sigma} = \text{diag}[\tilde{\sigma}_1^{2*} \dots \tilde{\sigma}_n^{2*}]$, compatibile con lo schema di Frisch, nella quale risulti $\tilde{\sigma}_{n-1}^{2*} \neq 0$ e $\tilde{\sigma}_n^{2*} \neq 0$. Si considerino poi le matrici di covarianza

$$\tilde{\Sigma} = \text{diag} [\tilde{\sigma}_1^{2*} \dots \tilde{\sigma}_{n-2}^{2*}, \tilde{\sigma}_{n-1}^2, \tilde{\sigma}_n^2] \quad (57)$$

relative a punti dello spazio del rumore appartenenti ad un piano parallelo al piano coordinato contenente i punti $(0, \dots, 0, \tilde{\sigma}_{n-1}^2, \tilde{\sigma}_n^2)$. Risultando $\tilde{\sigma}_{n-1}^{2*} \neq 0$ e $\tilde{\sigma}_n^{2*} \neq 0$, $\Sigma - \text{diag}[\tilde{\sigma}_1^{2*}, \dots, \tilde{\sigma}_{n-2}^{2*}, 0, 0]$ sarà definita positiva. Si partizioni ora la matrice

$$\hat{\Sigma} = \Sigma - \text{diag} [\tilde{\sigma}_1^{2*} \dots \tilde{\sigma}_{n-2}^{2*}, \tilde{\sigma}_{n-1}^2, \tilde{\sigma}_n^2] \quad (58)$$

come segue

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{11} & \Sigma_{12} \\ \Sigma_{21} & \hat{\Sigma}_{22} \end{bmatrix}$$

ove

$$\hat{\Sigma}_{11} = \Sigma_{11} - \text{diag} [\tilde{\sigma}_1^{2*} \dots \tilde{\sigma}_{n-2}^{2*}]$$

$$\hat{\Sigma}_{22} = \Sigma_{22} - \text{diag} [\tilde{\sigma}_{n-1}^2, \tilde{\sigma}_n^2].$$

$\hat{\Sigma}_{11}$ è una sottomatrice sulla diagonale principale della matrice definita positiva $\Sigma - \text{diag}[\tilde{\sigma}_1^{2*}, \dots, \tilde{\sigma}_{n-2}^{2*}, 0, 0]$ ed è quindi a sua volta definita positiva. L'applicazione del lemma sul determinante delle matrici partizionate (vedi Appendice) a $\hat{\Sigma}$ (58) fornisce

$$\det \hat{\Sigma} = \det \hat{\Sigma}_{11} \det [\hat{\Sigma}_{22} - \Sigma_{21} \hat{\Sigma}_{11}^{-1} \Sigma_{12}]. \quad (59)$$

Dalla (59) segue che $\det \hat{\Sigma} = 0$ se e solo se $\det [\hat{\Sigma}_{22} - \Sigma_{21} \hat{\Sigma}_{11}^{-1} \Sigma_{12}] = 0$; sviluppando tale espressione si trova una espressione del tipo

$$(\alpha - \tilde{\sigma}_{n-1}^2)(\beta - \tilde{\sigma}_n^2) - \gamma^2 = 0$$

che descrive una iperbole con concavità rivolta verso l'origine. Tale iperbole degenera in una coppia di rette quando $\gamma^2 = 0$.

Osservazione 13 – Se si considera una matrice $\tilde{\Sigma}$ (57) ove $\tilde{\sigma}_{n-1}^{2*} = 0$ e $\tilde{\sigma}_n^{2*} = 0$, la corrispondente iperbole degenera in un punto. La dimostrazione del Teorema 3 mostra che le condizioni $\tilde{\sigma}_{n-1}^{2*} \neq 0$ e $\tilde{\sigma}_n^{2*} \neq 0$ possono venire sostituite, senza perdita di generalità, da $\tilde{\sigma}_{n-1}^{2*} \neq 0$ o $\tilde{\sigma}_n^{2*} \neq 0$.

Osservazione 14 – $\mathcal{J}(\Sigma)$ partiziona il primo ortante dello spazio del rumore in due regioni. I punti al di sopra di tale (iper)superficie corrispondono a matrici $\hat{\Sigma}$ non definite e definite negative, quelli al di sotto, a matrici $\hat{\Sigma}$ definite positive.

Osservazione 15 – $\mathcal{J}(\Sigma)$ interseca gli assi coordinati nei punti $(0 \dots 0 \tilde{\sigma}_i^2 0 \dots 0)$ corrispondenti agli n rumori OLS che definiscono il massimo rumore possibile sulle singole variabili.

Osservazione 16 – I punti di $\mathcal{J}(\Sigma)$ sono isomorfi ai punti del simpleso definito dal Teorema 2 (che definiscono univocamente punti di \mathcal{R}^n). Le intersezioni di $\mathcal{J}(\Sigma)$ con gli assi coordinati corrispondono, in tale isomorfismo, ai vertici del simpleso.

I punti di $\mathcal{J}(\Sigma)$ possono venire calcolati mediante il seguente algoritmo.

Algoritmo 1

1) Si calcoli la varianza del massimo rumore (OLS) ammissibile su x_1 , $\tilde{\sigma}_1^{2m}$, mediante la (24).

2) Si assuma su x_1 , un rumore con varianza $\tilde{\sigma}_1^{2*} = k_1 \tilde{\sigma}_1^{2m}$ pari ad una frazione ($0 \leq k_1 \leq 1$) di $\tilde{\sigma}_1^{2m}$. Si definisca ora $\hat{\Sigma}^1$ come segue

$$\hat{\Sigma}^1 = \Sigma - \text{diag} [\tilde{\sigma}_1^{2*} 0 \dots 0]. \quad (60)$$

Se $\hat{\Sigma}^1$ è singolare ($k_1 = 1$), si è determinato un punto di $\mathcal{J}(\Sigma)$ e la procedura prosegue con un nuovo valore di k_1 .

3) Per il rumore considerato su x_1 , con varianza $\tilde{\sigma}_1^{2*}$, si calcoli la varianza del massimo rumore ammissibile sulla seconda variabile, x_2 , data da

$$\tilde{\sigma}_2^{2m} = \frac{\det \hat{\Sigma}^1}{\det \hat{\Sigma}_2^1} \quad (61)$$

ove $\hat{\Sigma}_2^1$ si ottiene eliminando la seconda riga e la seconda colonna di $\hat{\Sigma}^1$. Si consideri ora una frazione del massimo rumore ammissibile su x_2 , $\tilde{\sigma}_2^{2*} = k_2 \tilde{\sigma}_2^{2m}$ ($0 \leq k_2 \leq 1$) e si definisca la corrispondente matrice

$$\hat{\Sigma}^2 = \Sigma - \text{diag} [\tilde{\sigma}_1^{2*}, \tilde{\sigma}_2^{2*}, 0 \dots 0]. \quad (62)$$

Se $\hat{\Sigma}^2$ è singolare ($k_2 = 1$), si è ottenuta una nuova soluzione e la procedura prosegue con un nuovo valore di k_2 finché la matrice di covarianza del rumore $\tilde{\Sigma} = \text{diag} [\tilde{\sigma}_1^{2*} \tilde{\sigma}_2^{2*} \dots \tilde{\sigma}_n^{2*}]$ non è completata. Si osservi che $\tilde{\sigma}_i^{2*}$ deve essere una frazione di $\tilde{\sigma}_i^{2m}$ per $i \leq n - 1$ mentre all'ultimo passo $\tilde{\sigma}_n^{2*} = \tilde{\sigma}_n^{2m} = \det [\hat{\Sigma}^{n-1}] / \det [\hat{\Sigma}_n^{n-1}]$.

Esempio 7

Si consideri nuovamente la matrice di covarianza (25) per la quale sono già state determinate, nell’Esempio 1, le tre soluzioni OLS che definiscono i vertici del semplice che contiene tutte le soluzioni dello schema di Frisch riportato in Figura 1.

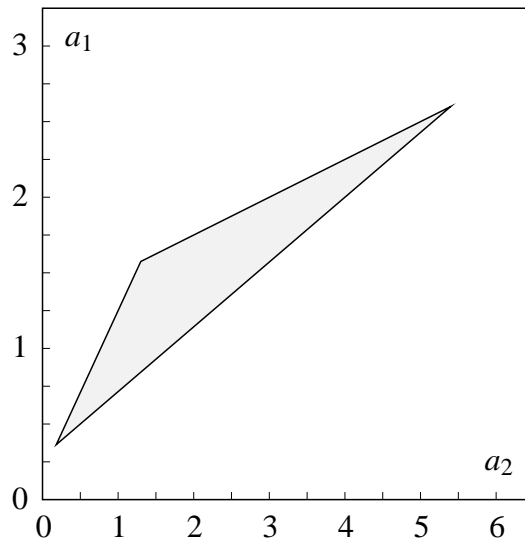


Figura 1 – Simpleso delle soluzioni per la matrice (25)

Nell’Esempio 1 sono state calcolate anche le varianze massime dei rumori sulle variabili, date da $\tilde{\sigma}_1^{2m} = 6.5079$, $\tilde{\sigma}_2^{2m} = 3.7963$ e $\tilde{\sigma}_3^{2m} = 3.7273$, che definiscono le intersezioni con gli assi di $\mathcal{S}(\Sigma)$. I rimanenti punti di tale superficie possono venire calcolati mediante l’Algoritmo 1; si ottiene la superficie riportata in Figura 2.

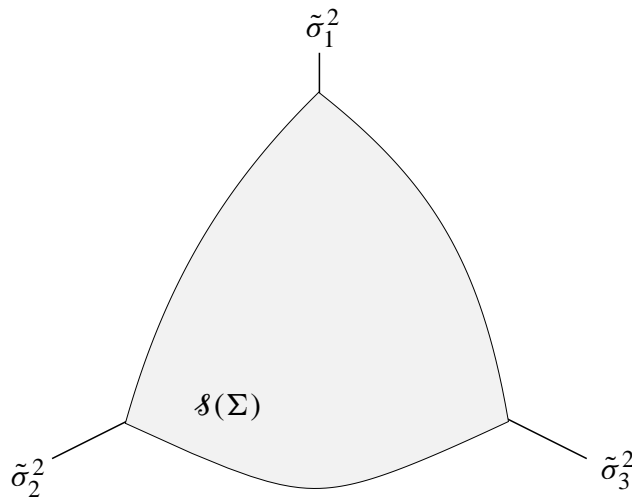


Figura 2 – Superficie di singolarità per la matrice (25)

Osservazione 17 – La superficie del simpleso delle soluzioni nello spazio dei parametri può venire

considerata come una misura dell'incertezza sui coefficienti della relazione lineare che lega i dati privi di rumore. Al tendere del rumore a zero il simpleso collassa sul punto che descrive tale relazione. Se alla stessa sequenza vengono sommate quantità minori di rumore, si ottiene un nuovo simpleso contenuto nel precedente. Considerazioni analoghe possono venire fatte sulle superfici di singolarità $\mathcal{S}(\Sigma)$ che si ottengono sommando ad una stessa sequenza priva di rumore quantità decrescenti di rumore; le corrispondenti superfici di singolarità si sposteranno verso l'origine (restando sempre al di sotto delle precedenti), sino a collassare sull'origine nella situazione di rumore nullo.

Esempio 8

Si consideri la matrice di covarianza (63) relativa agli stessi dati utilizzati per costruire la matrice di covarianza (25) ma con rumore additivo dimezzato,

$$\Sigma' = \begin{bmatrix} 12 & -6 & -4 \\ -6 & 5.75 & 1 \\ -4 & 1 & 4 \end{bmatrix}. \quad (63)$$

Lo spettro è ora dato da $\lambda_1 = 1.2309$, $\lambda_2 = 3.6629$ e $\lambda_3 = 16.8562$. Le varianze dei rumori massimi su x_1 , x_2 e x_3 sono

$$\tilde{\sigma}_1^{2m} = \det[\Sigma'] / \det[\Sigma'_1] = 76/22 = 3.4545$$

$$\tilde{\sigma}_2^{2m} = \det[\Sigma'] / \det[\Sigma'_2] = 76/32 = 2.3750$$

$$\tilde{\sigma}_3^{2m} = \det[\Sigma'] / \det[\Sigma'_3] = 76/33 = 2.3030.$$

I coefficienti a_1 , a_2 ed a_3 dei corrispondenti modelli OLS (con a_3 normalizzato ad 1) sono dati da

$$a_1 = 1.2941 \quad a_1 = 1.6667 \quad a_1 = 0.5151$$

$$a_2 = 1.1765 \quad a_2 = 2.6667 \quad a_2 = 0.3636.$$

La Figura 3 riporta il simpleso associato alla matrice (25) (grigio chiaro) e quello (grigio scuro) associato alla matrice (63) che risulta contenuto nel precedente.

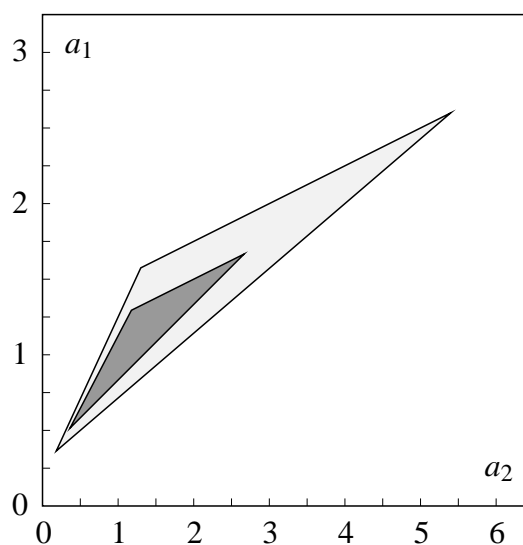


Figura 3 – Simpleso delle soluzioni per le matrici (25) e (63)

Nella Figura 4 sono poi riportate le superfici di singolarità per la matrice di covarianza (25) ($\mathcal{J}(\Sigma)$, in grigio chiaro) e per la matrice (63) ($\mathcal{J}(\Sigma')$, in grigio scuro) che risulta più vicina all'origine della precedente.

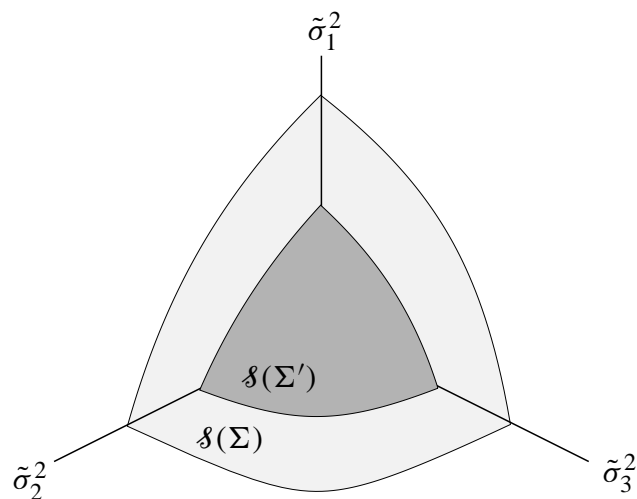


Figura 4 – Superfici di singolarità per le matrici (25) e (63)