Applicazione della Forward Search in problemi di trasformazioni piane

Alessandro Carosio(*), Marco Piras(**), Dante Salvini(*)

(*) ETH, Politecnico federale di Zurigo, IGP, HIL Hönggerberg, 8093 Zurigo, Svizzera
(**) Politecnico di Torino, DITAG, C.so Duca degli Abruzzi 24, 10129 Torino, Italia

Riassunto
Il problema delle trasformazioni piane è solitamente affrontato risolvendo il sistema di equazioni che descrivono le relazioni tra i due sistemi di riferimento, mediante l'utilizzo del metodo dei minimi quadrati. Generalmente questo è un approccio efficiente che porta alla stima dei coefficienti in maniera corretta. Le complicazioni nascono quanto all'interno delle osservazioni ci sono uno o più outliers che, se non vengono individuati, conducono ad una stima dei coefficienti errata.
Soluzioni moderne a questo problema sono fornite dalla statistica robusta che permette di calcolare i parametri corretti anche in presenza di outliers. Tra le più recenti è da considerare il metodo *Forward Search*. La particolarità di questo metodo risiede nel passaggio graduale da una soluzione iniziale robusta, calcolata con *Least-Median Squares*, ad una soluzione ai minimi quadrati.

Abstract
The planar transformation problem is tackled solving the equations system, which describe the relationships between the two reference systems applying the *Least Squares* method. Generally this is an efficient approach that estimates the unknown coefficients in a correct manner. Some complications arise when one or more outliers are present in the original dataset. If these can not be identified they bring to the estimation of incorrect coefficients.
Modern solutions to this kind of problems are proposed by the robust statistic, by mean of which it is possible to calculate the correct parameters also if outliers are included in the data. One of the more recent proposals to be considered is the *Forward Search* technique. The characteristic of this method is a gradual moving from an initial robust solution computed with *Least-Median Squares* to a *Least Squares* solution.

Introduzione
I progressi tecnologici degli strumenti geodetici e topografici, hanno contribuito in modo considerevole all'incremento della quantità di dati acquisibili sul terreno. Con l'aumento della mole di dati reperibili aumentano anche il lavoro di analisi e inevitabilmente la presenza di outliers. Si rendono quindi necessarie nuove tecniche per l'analisi e il trattamento dei dati.
Nella maggior parte dei problemi trattati in geodesia, la ridondanza dei dati (mai troppo elevata) viene impiegata per migliorare la precisione della soluzione, ad esempio con il metodo dei minimi quadrati (LS), previa linearizzazione del problema stesso. La tecnica LS è un valido strumento per la compensazione di diverse tipologie di misure e per la ricerca ottimale della soluzione.
La presenza di un singolo outlier conduce però alla stima errata dei parametri incogniti. Nella soluzione classica, si identificano gli outliers mediante un'analisi dei residui per poi estrometterli dalle iterazioni successive. Il problema risulta di complessa soluzione quando il numero di outliers cresce e l'identificazione non è più immediata, in quanto, come è noto, l'errore non si localizza solo in corrispondenza dell'osservazione sbagliata, ma viene ripartito anche sulle altre misure. Si genera

in altre parole una combinazione di effetti che difficilmente può essere interpretata. Per il metodo LS occorre però giungere all'assenza di osservazioni affette da outliers.
La statistica robusta rappresenta una valida soluzione alternativa nei casi in cui possono esistere errori grossolani all'interno dei dati che porterebbero a risultati non corretti. Esistono stimatori robusti con diversi livelli del punto di rottura (*breakdown point*). Un punto di rottura più basso permette di raggiungere una migliore precisione, mentre un punto di rottura elevato resiste a un maggior numero di outliers.
In base a quanto detto, abbiamo cercato di capire se il metodo robusto della *Forward Search* possa essere uno strumento applicabile ai problemi in geodesia. Nei paragrafi che seguono descriveremo questo metodo e come l'abbiamo applicato alle trasformazioni piane, quali test abbiamo svolto e i risultati ottenuti.

Cenni sulle tecniche classiche

In geodesia la maggior parte dei parametri è determinata tramite procedure statistiche di stima. Queste tecniche calcolano i valori richiesti sulla base di osservazioni multivariate e dei corrispettivi modelli funzionali. Il metodo comunemente applicato è il principio dei minimi quadrati (LS). Questo stimatore è ottimale per osservazioni seguenti la distribuzione normale, in quanto fornisce le migliori stime dei parametri incogniti (in un modello lineare). In forma matriciale possiamo esprimerla nel modo seguente:

$$V^T Q_{ll}^{-1} V = \min \tag{1}$$

dove v = vettore dei residui
Q_{ll}^{-1} = matrice dei pesi

Il metodo LS consente di considerare le diverse precisioni delle osservazioni mediante l'impiego della matrice dei pesi e permette di unire diverse tipologie di misura (angoli, distanze, ecc.) grazie alla linearizzazione del problema. L'inconveniente del metodo LS risiede nella forte sensibilità alla presenza di errori grossolani o comunque a variazioni non consoni alla distribuzione normale. Come è ben visibile nella più elementare applicazione di questo principio, la media aritmetica, un solo elemento errato della serie, falsa il valore medio cercato. Si parla in questo caso di un metodo con un punto di rottura (*breakdown point* BP) molto basso, praticamente pari a 0. In geodesia, come in tanti altri settori dove è richiesta la stima statistica di parametri incogniti, spesso non può essere garantita l'assenza di errori nei dati da elaborare. In questi casi si presentato due possibilità: l'applicazione del metodo descritto, eliminando in modo iterativo i dati considerati errati, fino ad ottenere una soluzione pulita, oppure fare capo a metodi di calcolo meno sensibili agli errori grossolani e agli outliers, in altre parole metodi con un punto di rottura più alto. In questa categoria troviamo per esempio gli stimatori M (derivati dal metodo *Maximum-Likelihood*). Tra questi figura lo stimatore robusto di *Huber*, che ben si presta alle problematiche geodetiche per l'affinità al metodo LS. La funzione da minimizzare in questo caso è la somma della seguente funzione ρ di tutti i residui:

$$\rho(v_i) = \begin{cases} \frac{1}{2} v_i^2 & per\,|v_i| < c \\ c|v_i| - \frac{1}{2} c^2 & per\,|v_i| \geq c \end{cases} \tag{2}$$

dove c = costante
v = residuo

Nell'intervallo [-*c*,*c*] la funzione è identica a quella del metodo LS, mentre all'esterno è lineare. La costante *c* è da scegliersi in base alla precisione delle osservazioni (p.es. $c = 3\sigma$). Uno sviluppo di questo metodo prevede una definizione variabile di *c*, dipendente dalla ridondanza delle osservazioni (proposte di *Mallows* o *Schweppe*).

Un metodo efficace pensato espressamente per le problematiche geodetiche è lo stimatore BIBER, (*bounded influence by standardized residuals*) che fa pure parte della classe degli stimatori M (Wicki, 1999). La particolarità di questo metodo è quella di ridurre l'influsso degli outlier sui parametri stimati, analizzando i residui standardizzati. La funzione da minimizzare in questo caso è la somma della seguente funzione ρ di tutti i residui:

$$\rho(\frac{v_i}{\sigma_{v_i}}) = \begin{cases} \frac{1}{2\sigma_{v_i}} v_i^2 & per \left| \frac{v_i}{\sigma_{v_i}} \right| < c \\ \frac{c}{\sigma_{v_i}} |v_i| - \frac{1}{2} c^2 & per \left| \frac{v_i}{\sigma_{v_i}} \right| \geq c \end{cases} \tag{3}$$

dove v = residuo
c = costante
σ_{v_i} = scarto quadratico medio del residuo

Nella categoria degli stimatori ad alto punto di rottura (BP = 0.5) troviamo il metodo *Least-Median Squares* (LMS). Questo metodo assume quale stima ottimale dei parametri incogniti la combinazione di valori per la quale il residuo quadrato mediano risulta essere minore. La condizione da soddisfare è quindi la seguente:

$$med(v_i^2) \rightarrow \min \tag{4}$$

dove v = residuo

I parametri stimati che soddisfano l'equazione (4) hanno un punto di rottura del 50%.
La definizione non riporta informazioni su come trovare i parametri da stimare. In pratica si ricerca il minimo utilizzando metodi approssimati. In sintesi si tratta di individuare tra i possibili sottoinsiemi di osservazioni quello che soddisfa la condizione posta. È evidente che appena il sistema assume una certa ampiezza il numero di combinazione da calcolare aumenta a dismisura. Gli algoritmi sviluppati a questo proposito contemplano dei meccanismi per la ricerca efficiente della soluzione cercata. Per le caratteristiche accennate, gli stimatori robusti sono un valido strumento per l'analisi di dati affetti da outliers.

Il metodo Forward Search: principi generali

Il *Forward Search* (FS) può essere visto come un metodo robusto e dinamico che consente di passare da una soluzione ottenuta con il metodo LMS (alto punto di rottura) ad una soluzione LS (alta precisione) in base ai bisogni.
La rappresentazione schematica della struttura del metodo mediante un flowchart è la seguente:

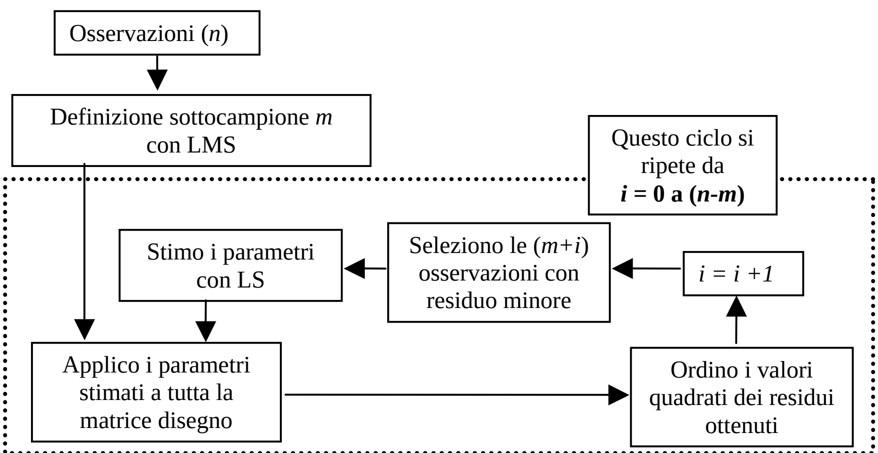


Figura 1 - Flowchart Forward Search

Il concetto di fondo di questa tecnica è di utilizzare un sottoinsieme iniziale *m* di dati privi di outliers, estratti dal campione originale. Questo sottocampione è selezionato con lo stimatore LMS. Solitamente si definisce un numero *m* di elementi del sottoinsieme pari a *p*, numero dei parametri incogniti, lasciando in questo modo *n-p* osservazioni da testare. I coefficienti, così stimati, vengono applicati a tutte le altre osservazioni per definire gli scarti di ogni osservazione. Si analizzano i valori degli scarti quadrati, identificando le *m+1* osservazioni con il residuo quadrato più piccolo. Questo gruppo di osservazioni viene quindi usato per ricalcolare con LS dei nuovi parametri. Il ciclo prosegue unità per unità, aggiungendo una osservazione alla volta sino ad ottenere $m = n$, cioè giungendo alla fine del processo ad una soluzione LS comune.
La novità del metodo consiste, oltre nella sua variabilità, in quanto consente di passare dal LMS a LS, nell'effettuare un monitoraggio continuo, epoca per epoca, di alcuni parametri statistici (residui, distanza di Cook, stima dei coefficienti, t-stastistica, ecc.). Durante i *n–p* passaggi, si controllano le variazioni di questi parametri individuando quale nuova osservazione introdotta porta ad una brusca alterazione, permettendo così di distinguere eventuali "classi" all'interno dei dati. Nel nostro caso siamo in grado di distinguere i dati in "puliti", e quindi utilizzabili per la stima dei coefficienti incogniti, e gli "outliers".
Sino a quando si utilizzano osservazioni non affette da outliers i parametri osservati si mantengono stabili. La soluzione finale corrisponde alla stima dei parametri incogniti ottenuta al passo precedente la prima variazione significativa dei parametri stessi.
In sintesi il metodo FS si compone di tre passi:

- *Ricerca di un sottoinsieme iniziale;*
- *Fase di avanzamento, con aggiunta di singole osservazioni ad ogni passo;*
- *Analisi dei parametri statistici durante la fase di avanzamento del calcolo.*

Applicazioni in geodesia

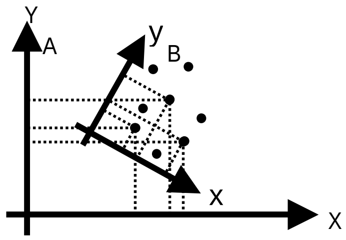
Con il fine di meglio descrivere le caratteristiche del metodo FS e di evidenziarne i vantaggi rispetto ad altri metodi robusti, in questa sezione si intende discuterne l'applicazione ad un problema classico in geodesia: le trasformazioni piane. I risultati ottenuti saranno discussi e confrontati con quelli calcolati applicando altri metodi.
Si definiscono trasformazioni piane, tutte quelle trasformazioni che fanno corrispondere biunivocamente ad un dato insieme piano di punti un altro insieme piano di punti.
Una trasformazione molto usata è quella detta conforme, dove i parametri che legano i due sistemi di riferimento sono essenzialmente una rotazione rispetto all'origine, due traslazioni dell'origine e un fattore di scala isotropo. Le equazioni che descrivono la trasformazione dal sistema locale (O', X', Y') al sistema globale (O, X, Y) saranno le seguenti:

$$\begin{aligned} X_P &= T_X + \mu\left(X'_P \cos\alpha + Y'_P \sin\alpha\right) \\ Y_P &= T_Y + \mu\left(-X'_P \sin\alpha + Y'_P \cos\alpha\right) \end{aligned} \qquad (5)$$

Il problema, comune a tutti i tipi di trasformazione piana, è quello della determinazione dei parametri incogniti, che nel nostro caso saranno: μ, α, T_X, T_Y.
Il calcolo può essere eseguito partendo da un numero sufficiente di punti noti nei due sistemi di riferimento. Ogni punto noto nei due sistemi di riferimento, permetterà la scrittura delle equazioni (5). Per ricavare i quattro parametri incogniti, saranno necessarie almeno quattro equazioni derivanti da due punti (P_1 e P_2) noti nei due sistemi di riferimento. Il sistema di equazioni così ottenuto può essere poi linearizzato semplicemente ponendo: $a = \mu\cos\alpha$ e $b = \mu\sin\alpha$.
Una volta definiti i parametri incogniti, possiamo applicare la relazione (5) ai punti noti solo in uno dei due sistemi di riferimento, per effettuare la trasformazione.
È possibile anche determinare i parametri incogniti usando un numero di punti noti eccedente rispetto al minimo indispensabile, giungendo alla soluzione del sistema di equazioni mediante l'applicazione di un metodo di compensazione. Di seguito presentiamo un riassunto del calcolo

delle incognite applicando il metodo comune LS, i metodi robusti Huber e BIBER, nonché LMS e FS incrementando il numero di outlier nei dati.
L’implementazione usata del metodo FS è una libreria per S-Plus sviluppata dai ricercatori che hanno proposto questo metodo (A.C. Atkinson, M. Riani). Gli altri algoritmi applicati fanno parte di un programma di calcolo standard Svizzero per le trasformazioni piane di coordinate (Transint).

Situazione



Dati iniziali

Punti comuni (in A e B).:	10 (x,y)
Precisione……………...:	centimetrica
Parametri incogniti.……:	4
Ridondanza…………… :	16 ($2\cdot 10-4$)
Parametri cercati…….T_x:	-0.186856919
T_y:	29.51869566
a:	0.9999938289
b:	0.0000357561

Figura 2 – Schema problema trasformazioni piane

Per avere un parametro di confronto, si è calcolata la soluzione ai minimi quadrati, ottenendo così una soluzione comparabile con l’ultimo passo del metodo FS. Paragonando i risultati possiamo confermare la corretta definizione del problema, e analizzare nel dettaglio l’avanzamento del processo nel FS. In un primo test abbiamo introdotto un errore grossolano nelle osservazioni (una coordinata del sistema A falsata di 10 metri) e ricalcolato i parametri analizzando i residui. Il metodo FS fornisce i parametri corretti al passo 19 identificando l’outlier. Anche gli altri metodi robusti applicati forniscono la soluzione corretta e individuano l’outlier. I parametri ottenuti con LS sono falsati dall’outlier; un’analisi dei residui permette però di individuare chiaramente la fonte dell’errore. Aumentando il numero di outliers a 3, la situazione è simile a quella presentata.
Per dare una descrizione qualitativa e quantitativa dell’efficacia dei vari metodi, elenchiamo di seguito i residui ottenuti dal calcolo dei dati contenenti 5 outliers (un quarto delle osservazioni):

Metodo LS		**Metodo FS (*m=15*)**		**Metodo LMS**		**Metodo BIBER**	
V_y [mm]	V_x [mm]	V_y [mm]	V_x [mm]	V_y [mm]	V_x [mm]	V_y [mm]	V_x [mm]
2138.9	-227.0	1.0	1.8	0.9	0.3	-8.8	-12.4
-5460.4	-1543.7	-7999.3	-3.2	-7998.6	-4.4	-8011.6	-6.7
2583.7	9201.7	0.1	12000.0	1.7	11999.7	-11.7	120007.3
1293.5	-2932.8	-3.5	0.7	-1.7	-1.2	-4.8	7.5
-9362.2	-2181.4	-10003.8	2.6	-10002.4	0.2	-10000.4	3.0
74.8	-1087.2	-0.6	1.2	0.1	-1.7	6.6	-7.5
-240.7	4856.0	1.5	4997.5	1.5	4994.3	10.6	4981.0
5964.0	-626.1	7002.7	-1.2	7003.1	-4.9	7018.4	-14.3
2361.9	-3521.0	3.2	2.4	5.3	1.3	-6.2	14.6
656.6	-1938.4	-1.8	-4.4	-0.6	-6.8	1.4	-6.0

Tabella 1 – Riassunto dei residui (V) calcolati con vari metodi di compensazione

I residui calcolati dai metodi robusti lasciano chiaramente riconoscere quali osservazioni sono state falsate e l’entità dell’errore introdotto. Per il metodo FS, i residui sono quelli relativi al passo precedente l’introduzione del primo outlier (*m=15*). I residui forniti dal metodo LS mostrano, invece, come questo ripartisce gli errori su tutte le osservazioni, rendendone difficile l’individuazione.

Tra i vari parametri di monitoraggio forniti dal metodo FS presentiamo di seguito l'andamento della varianza (s^2) per il caso discusso:

Avanzamento	s^2	
m=15	$0.71 \cdot 10^{-5}$	Ultima osservazione "pulita"
m=16	1.50	Primo outlier integrato nel calcolo
m=17	4.36	Secondo outlier integrato nel calcolo
m=20	19.61	Tutte le osservazioni sono state integrate

Tabella 2 – Descrizione andamenti varianza

Abbiamo messo alla prova i metodi aumentando di volta in volta il numero di outliers. Riassumendo possiamo dire che fintanto il numero di outliers è inferiore alla metà del numero di osservazioni i metodi robusti qui presentati sono in grado di giungere alla soluzione corretta. Se gli outliers superano il numero di osservazioni corrette il risultato non è scontato. Nel caso del metodo FS molto dipende dalla scelta del subset iniziale. Se questo comprende solo un numero minore di outliers (o nessuno), il calcolo riesce, negli altri casi è quanto meno possibile fare una classificazione dei dati sulla basi dei parametri statistici forniti da questo metodo.
Concludendo possiamo dire che i tre metodi robusti dimostrano la loro efficienza nella compensazione di dati affetti da outliers. I vantaggi del metodo FS risiedono nel fatto di basarsi sul metodo LS per il calcolo dei coefficienti.

Conclusioni

Il metodo *Forward Search* si è dimostrato, relativamente ai casi studiati, valido per l'individuazione di outliers e eventuali sistematismi all'interno di problemi di trasformazioni piane.
La struttura del metodo FS e il suo procedimento "graduale", che consente di passare dal metodo LMS a LS, si dimostra particolarmente utile per la "classificazione" dei dati in "puliti" e "outliers". Si nota, quindi, come questo strumento oltre ad essere un mezzo di calcolo potrebbe fungere anche come valido strumento di filtraggio dati. Il metodo risulta valido anche nei casi in cui esistono fenomeni di mascheramento, nei quali sia LMS che LS risultano essere inadatti al trattamento dei dati. Una particolarità che si è riscontrata, utilizzando il metodo FS, è la tendenza a distinguere all'interno dei dati, le classi di appartenenza delle osservazioni, vale a dire a separare la coordinata X dalla coordinata Y. In questo modo si perde la condizione di appartenenza delle singole coppie di coordinate all'entità punto. Il problema potrebbe essere risolto applicando un FS modificato, che prevede l'inserimento di blocchi di dati invece che singole osservazioni. Per il futuro, si prospetta di applicare il metodo a problemi di trasformazione tridimensionali, dove per la complessità del problema l'individuazione degli outlier multipli risulta essere molto importante.

Bibliografia

Comoglio G. (2004), *Elementi di Cartografia*, CELID, Torino.
Huber J. P. (1981), *Robust statistics*, Wiley, New York.
Huber J. P., Ronchetti M. E., Rousseeuw P. J., Stahel W. A. (1986), *Robust statistics: the approach based on influence functions*, Wiley, New York.
Atkinson A., Riani M. (2000), *Robust Diagnostic Regression Analysis*, Springer, New York.
Atkinson A., Cerioli A., Riani M. (2003), *Exploring Multivariate Data with the Forward Search*, Springer, New York.
Wicki, F. (1999), *Robuste Schätzverfahren für die Parameterschätzung in geodätischen Netzen*, Dissertation ETHZ Nr. 12894, Zürich