

# TIẾP CẬN BÀI TOÁN TÌM KIẾM VÀ LOẠI BỎ CÁC SAI SỐ THÔ TRONG DỮ LIỆU DỊ THƯỜNG TRỌNG LỰC KHI XÂY DỰNG CƠ SỞ DỮ LIỆU DỊ THƯỜNG TRỌNG LỰC THEO PHƯƠNG PHÁP KRIGING TỔNG QUÁT

PGS. TSKH. HÀ MINH HOÀ  
Viện Khoa học Đo đạc và Bản đồ

## Tóm tắt

Bài toán nội suy xác định các giá trị dị thường trọng lực tại các đỉnh của các ô chuẩn trong CSDL dị thường trọng lực quốc gia từ các giá trị dị thường trọng lực trên các điểm trọng lực là bài toán khoa học - kỹ thuật rất phức tạp. Hiện nay phương pháp kriging tổng quát đang được sử dụng để giải quyết bài toán nêu trên. Tuy nhiên, việc xác định một cách tin cậy các hệ số  $\beta$  của mô hình trên đòi hỏi phải tiến hành phát hiện và tìm kiếm các trị đo thô trong các dữ liệu dị thường trọng lực. Bài báo khoa học này đề xuất sử dụng thuật toán truy hồi T-T để giải quyết hiệu quả vấn đề phát hiện và tìm kiếm các trị đo thô trong các dữ liệu dị thường trọng lực.

## 1. Đặt vấn đề

Việc xây dựng cơ sở dữ liệu (CSDL) trọng lực quốc gia dưới dạng các ô chuẩn (grid) phủ trùm lãnh thổ quốc gia là bài toán khoa học - kỹ thuật rất phức tạp. Sự phức tạp của bài toán này nằm ở chỗ cần phải xác định các giá trị dị thường trọng lực trên các đỉnh của các ô chuẩn bằng phương pháp nội suy từ tập hợp rất lớn các điểm trọng lực có các giá trị dị thường không khí tự do ở vùng đồng bằng hoặc các giá trị dị thường Faye ở khu vực rừng núi. Các phương pháp nội suy được nghiên cứu để giải quyết bài toán này được đặt trên cơ sở phương pháp địa thống kê (Geostatistical Method) hay còn được gọi là *Dự báo bình phương nhỏ nhất* (Least Squares Prediction) hoặc *Dự báo không gian tối ưu* (Optimal Spatial Prediction), theo đó chúng ta phải xác định các hàm tương quan của tập hợp các biến ngẫu nhiên  $L(x)$  tạo nên trường ngẫu nhiên trong miền  $D$ . Trường

ngẫu nhiên  $L(x)$  với các tọa độ không gian  $x$  được định nghĩa như sau (Cressie N.A.C. (1993); Schabenger O., Gotway C.A. (2005)):

$$\{L(x) : x \in R^n\} \in D.$$

Giả thiết rằng trên tập hợp  $Q$  gồm  $n$  điểm trọng lực chúng ta đã xác định được  $n$  giá trị biến ngẫu nhiên (các giá trị dị thường trọng lực)  $L[x_i]$  ( $i=1,2,\dots,n$ ). Bài toán được đặt ra là cần xác định biến ngẫu nhiên  $\tilde{Z}(x_p)$  tại điểm  $P$  thuộc tập hợp  $P$  sao cho thỏa mãn các điều kiện:

- Không xê dịch

$$E[\tilde{L}(x_p) - L(x_p)] = 0, \quad (1)$$

- Sai số trung phương cực tiểu

$$E[\tilde{L}(x_p) - L(x_p)]^2 = \min, \quad (2)$$

ở đây  $E[.]$  - kỳ vọng toán học.

Các điều kiện (1) và (2) tương ứng với trường ngẫu nhiên tính tại theo cách tiếp

cận Wiener - Kolgomorov, theo đó Wiener N. đã xác định các điều kiện trên theo nguyên tắc tất định (deterministic principle) (xem Wiener N. (1949)), còn Kolgomorov A. N. theo nguyên tắc ngẫu nhiên (xem Kolmogorov A.N. (1933)). Khi đó, từ tập hợp các biến ngẫu nhiên (các giá trị dị thường trọng lực) tại n điểm trọng lực thuộc tập hợp Q, chúng ta xây dựng các hàm nội suy để xác định các giá trị của biến ngẫu nhiên (các giá trị dị thường trọng lực) tại các điểm p thuộc tập hợp P. Tập hợp Q thường là tập hợp các điểm đo nằm trên khu vực tính toán, còn P là tập hợp các đỉnh của các ô chuẩn (grid) trong CSDL dị thường trọng lực đang được xây dựng. Trong thực tiễn tính toán trong trắc địa vật lý, người ta thường dùng phương pháp collocation trung phương hoặc phương pháp kriging đơn giản để giải quyết bài toán được đặt ra. Tuy nhiên, các phương pháp nêu trên chỉ được áp dụng khi các biến ngẫu nhiên  $L[x_i]$  ( $i=1,2,\dots,n$ ) thỏa mãn điều kiện:

$$E[L(x)] = \mu(x_1) = \mu(x_2) = \dots = \mu(x_n) = \mu = const, \quad (3)$$

thêm vào đó giá trị trung bình của các biến ngẫu nhiên  $E[L(x)] = \mu$  có thể bằng 0 có thể khác 0 và luôn là đại lượng không đổi trong trường ngẫu nhiên tĩnh tại (xem các tài liệu Moritz H. (1980) ; Anberto Molteni, Lisa Pertusini, Mirko Reguzzoni (2009); Marcin Ligas, Marek Kulczycki (2010)).

Như đã đánh giá trong tài liệu (Chauvet P and Galli A. (1982) ; Reguzzi M., Sansò F. and Venuti G. (2005)), các đánh giá theo phương pháp collocation trung phương hoặc phương pháp kriging đơn giản được đặt trên cách tiếp cận Wiener - Kolgomorov trong trường ngẫu nhiên tĩnh tại D mà trong đó giá trị trung bình  $\mu$  của các biến ngẫu nhiên là không đổi (hoặc bằng 0) và thỏa mãn điều kiện (3). Tuy nhiên, trong thực tế, các biến ngẫu nhiên  $L(x)$  (các giá trị dị

thường trọng lực trên các điểm trọng lực) được phân bố tại các vị trí khác nhau trong trường vật lý không đồng nhất. Do đó điều kiện (3) không được thỏa mãn và mỗi biến ngẫu nhiên  $L(x)$  có thành phần trend  $A(x) = \mu(x)$  riêng rẽ và các thành phần trend của các biến ngẫu nhiên  $L(x)$  luôn khác nhau (Library A.L.T. (1998)). Phương pháp kriging đơn giản đã được trình bày trong tài liệu (Hà Minh Hòa (2014a)).

Bây giờ chúng ta ký hiệu vector của các thành phần trend đối với n biến ngẫu nhiên  $L(x)$  trên n điểm thuộc tập hợp Q ở dạng sau:

$$\Omega(x) = \begin{bmatrix} \mu(x_1) \\ \mu(x_2) \\ \vdots \\ \mu(x_n) \end{bmatrix}, \quad (4)$$

thêm vào đó  $\mu(x_i) = E[L(x_i)]$  và các thành phần của vector  $\Omega$  không thỏa mãn điều kiện (3).

Bây giờ chúng ta biểu diễn công thức đánh giá giá trị tin cậy nhất của biến ngẫu nhiên  $\tilde{L}(x_p)$  tại điểm p thuộc tập hợp P dưới dạng sau:

$$\tilde{L}(x_p) = \lambda^T L(x) = \lambda^T \Omega + \tilde{Z}(x_p) = \mu(x_p) + \lambda^T Z(x), \quad (5)$$

ở đây giá trị trung bình (trend) tin cậy nhất tại điểm p được xác định theo công thức:

$$\mu(x_p) = \lambda^T \Omega, \quad (6)$$

còn vector  $Z(x)$  có dạng:

$$Z(x) = \begin{bmatrix} Z(x_1) \\ Z(x_2) \\ \vdots \\ Z(x_n) \end{bmatrix},$$

thêm vào đó thành phần thứ i ( $i=1,2,\dots,n$ ) của vector  $Z(x)$  được xác định theo công

$$\text{thức: } Z(\chi_i) = L(\chi_i) - \mu(\chi_i). \quad (7)$$

Từ (7) không khó khăn để nhận thấy rằng

$$E[Z(\chi_i)] = E[L(\chi_i)] - \mu(\chi_i) = 0. \quad (8)$$

Từ điều kiện không xê dịch (1) đối với đánh giá  $\tilde{L}(\chi_p)$  và lưu ý (5), (7), (8) chúng ta có:

$$\begin{aligned} E[\tilde{L}(\chi_p) - L(\chi_p)] &= E[\lambda^T \cdot L(\chi) - L(\chi_p)] = \\ &= E[\lambda^T \cdot \Omega + \lambda^T \cdot Z(\chi) - L(\chi_p)] = \\ &= \lambda^T \cdot \Omega - \mu(\chi_p) = \mu(\chi_p) \cdot (\lambda^T e - 1) = 0 \end{aligned}$$

và từ đây suy ra điều kiện:

$$\lambda^T e = 1 \quad (9)$$

với vectơ

$$e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}$$

tương tự như đối với phương pháp kriging đơn giản.

Trong các tài liệu (Cressie N.A.C. (1993), Olea R. A. (1999), Jay D. Martin, Timothy W. Simpson (2004), Schabenger O., Gotway C.A. (2005), Marcin Ligas, Marek Kulczycki (2014)) đã đề xuất các phương pháp khác nhau để xác định vectơ trend  $\Omega(\chi)$  (4) và biến ngẫu nhiên  $\tilde{L}(\chi_p)$  dựa trên phương pháp kriging tổng quát. Tuy nhiên trong các phương pháp được đề xuất còn tồn tại hai vấn đề chưa được giải quyết:

- Kiểm tra và tìm kiếm các sai số thô trong các dữ liệu dị thường trọng lực thuộc tập hợp Q trong quá trình xác định vectơ trend  $\Omega(\chi)$  (4);

- Sự không hiệu quả của việc giải hệ phương trình chuẩn với ma trận chuẩn không xác định dương trong quá trình xác định vectơ  $\lambda^T$  trong công thức (5).

Bài báo khoa học này sẽ đề xuất phương pháp giải quyết vấn đề thứ nhất. Việc giải

quyết vấn đề thứ hai sẽ được đề xuất trong bài báo khoa học tiếp theo.

## 2. Giải quyết vấn đề

Trong tài liệu (Marcin Ligas, Marek Kulczycki (2014)) đã biểu diễn mô hình trend dưới dạng đa thức sau:

$$\mu(\chi) = \beta_0 + x \cdot \beta_1 + y \cdot \beta_2 + x \cdot y \cdot \beta_3 + x^2 \cdot \beta_4 + y^2 \cdot \beta_5 + \dots \quad (10)$$

ở đây x, y - các tọa độ của điểm.

Giả thiết đa thức (10) có q bậc, chúng ta ký hiệu:

$$f_0(\chi) = 1, f_1(\chi) = x, f_2(\chi) = y,$$

$$f_3(\chi) = x \cdot y, f_4(\chi) = x^2, f_5(\chi) = y^2 \dots$$

Khi đó công thức (10) có dạng:

$$\begin{aligned} \mu(\chi) &= f_0(\chi) \cdot \beta_0 + f_1(\chi) \cdot \beta_1 + f_2(\chi) \cdot \beta_2 + \\ &+ f_3(\chi) \cdot \beta_3 + f_4(\chi) \cdot \beta_4 + f_5(\chi) \cdot \beta_5 + \dots \quad (11) \end{aligned}$$

Lưu ý (11), vế trái của công thức (6) có dạng:

$$\begin{aligned} \mu(\chi_p) &= f_0(\chi_p) \cdot \beta_0 + f_1(\chi_p) \cdot \beta_1 + f_2(\chi_p) \cdot \beta_2 + \\ &+ f_3(\chi_p) \cdot \beta_3 + f_4(\chi_p) \cdot \beta_4 + f_5(\chi_p) \cdot \beta_5 + \dots \quad (12) \end{aligned}$$

còn vế phải của công thức (6), khi lưu ý (4) và  $f_0(\chi) = f_0(\chi_p) = 1$ , có dạng:

$$\begin{aligned} \lambda_1 \cdot \mu(\chi_1) + \lambda_2 \cdot \mu(\chi_2) + \dots + \lambda_n \cdot \mu(\chi_n) &= \\ &= \beta_0 \cdot \sum_{i=1}^n \lambda_i + \beta_1 \cdot \sum_{i=1}^n f_1(\chi_i) \cdot \lambda_i + \\ &+ \beta_2 \cdot \sum_{i=1}^n f_2(\chi_i) \cdot \lambda_i + \dots + \beta_q \cdot \sum_{i=1}^n f_q(\chi_i) \cdot \lambda_i. \quad (13) \end{aligned}$$

Với mục đích đánh giá được giá trị tin cậy nhất của biến ngẫu nhiên biến  $\tilde{L}(\chi_p)$  theo thức công thức (5), chúng ta phải xác định được một cách các hệ số  $\beta_0, \beta_1, \dots, \beta_q$  của đa thức (11). Khi ký hiệu

$$F^T(\chi) = \begin{bmatrix} f_0(\chi) \\ f_1(\chi) \\ \vdots \\ f_q(\chi) \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{bmatrix} \quad (14)$$

biểu thức (11) có dạng:  $\mu(\mathbf{x}) = F(\mathbf{x})\cdot\beta$ . (15)

Đối với tập hợp Q bao gồm n điểm trọng lực, ma trận hệ số  $F(\mathbf{x})$  kích thước n x q được biểu diễn dưới dạng :

$$\bar{F}^T(\mathbf{x}) = \begin{bmatrix} F(\mathbf{x}_1) \\ F(\mathbf{x}_2) \\ \vdots \\ F(\mathbf{x}_n) \end{bmatrix}, \quad (16)$$

ở đây  $F(\mathbf{x}_i)$  ( $i=1,2,\dots,n$ ) có dạng (14).

Mặt khác, vectơ các biến ngẫu nhiên (các giá trị dị thường trọng lực thuộc tập hợp Q bao gồm n điểm trọng lực) được biểu diễn dưới dạng (Library A.L.T. (1998)):

$$L(\mathbf{x}) = \mu(\mathbf{x}) + \nabla(\mathbf{x}), \quad (17)$$

ở đây  $\nabla(\mathbf{x})$  là vectơ các sai số ngẫu nhiên với kỳ vọng toán học  $E[\nabla(\mathbf{x})] = \mathbf{0}$  và ma trận hiệp phương sai  $C_{LL}$

Ma trận hiệp phương sai  $C_{LL}$  được xác định trên cơ sở xác định dạng và các tham số của hàm bán phương sai lý thuyết. Trong phương pháp kriging người ta thường sử dụng các hàm bán phương sai lý thuyết như hàm số mũ

$$\gamma(d) = C_0 + C_1 \cdot \left(1 - e^{-\frac{d}{a}}\right),$$

hàm Gauss

$$\gamma(d) = C_0 + C_1 \cdot \left(1 - e^{-\frac{d^2}{a^2}}\right),$$

hàm cầu với dạng:

$$\gamma(d) = C_0 + C_1 \cdot \left(\frac{3d}{2a} - \frac{1}{2} \left(\frac{d}{a}\right)^3\right),$$

ở đây  $0 < d \leq a$  và  $\gamma(d) = C_0 + C_1$  khi  $d > a$  và hàm tuyến tính:

$$\gamma(d) = C_0 + C_1 \cdot \left(\frac{d}{a}\right),$$

ở đây  $0 < d \leq a$  và  $\gamma(d) = C_0 + C_1$  khi  $d > a$ .

Việc lựa chọn dạng và các tham số của hàm bán phương sai lý thuyết được thực

hiện dựa trên việc xây dựng hàm bán phương sai thực nghiệm. Trong tài liệu (Hà Minh Hoà, Nguyễn Tuấn Anh (2014b)) đã trình bày kết quả thực nghiệm xây dựng hàm bán phương sai thực nghiệm và xác định các tham số của hàm số mũ để nội suy các giá trị dị thường trọng lực trên các điểm trọng lực vào các đỉnh của các ô chuẩn dị thường trọng lực 5' x 5' ở khu vực Đông Bắc Việt Nam.

Lưu ý (15), (16), từ (17) chúng ta có hệ phương trình sai số:

$$\nabla(\mathbf{x}) = L(\mathbf{x}) - \bar{F}(\mathbf{x})\cdot\beta, \quad (18)$$

với ma trận trọng số  $C_{LL}^{-1}$ .

Do  $\mu(\mathbf{x}) = F(\mathbf{x})\cdot\beta$  là các giá trị trung bình tin cậy nhất của các biến ngẫu nhiên  $L(\mathbf{x})$ , nên trong hàng loạt công trình, ví dụ Cressie N.A.C. (1993), Olea R. A. (1999), Jay D. Martin, Timothy W. Simpson (2004), Schabenger O., Gotway C.A. (2005), Marcin Ligas, Marek Kulczycki (2014) đã đề xuất sử dụng phương pháp hợp lý cực đại (maximum likelihood estimation) dựa trên phân bố Gauss để đánh giá các giá trị tin cậy nhất của vectơ  $\beta$  và phương sai của biến ngẫu nhiên  $L(\mathbf{x})$ . Phương pháp này được áp dụng trên cơ sở xác định được rằng các biến ngẫu nhiên  $L(\mathbf{x})$  là các đại lượng ngẫu nhiên tuân theo phân bố chuẩn. Phương pháp này không cho đánh giá hiệu quả trong trường hợp tồn tại các sai số thô trong các dữ liệu dị thường trọng lực thuộc tập hợp Q.

Trong tài liệu (Jay D. Martin, Timothy W. Simpson (2004)), phương pháp đưa ra - tương hỗ (cross - validation) thường được sử dụng trong trường hợp không xác định được quy luật phân bố của các biến ngẫu nhiên  $L(\mathbf{x})$ . Trong thực tế thường sử dụng rộng rãi phương pháp thống kê đưa ra - tương hỗ bỏ - một - ra (Leave - one - out cross - validation). Bản chất của phương

pháp này là lần lượt loại bỏ 1 trị đo và đưa n - 1 trị đo còn lại vào tính toán. Như vậy chúng ta có n phương án tính toán. Phương án được chấp nhận là phương án cho sai số trung phương sau bình sai nhỏ nhất. Tuy nhiên, trong trường hợp tồn tại nhiều trị đo thô (lớn hơn 1 trị đo thô), chúng ta không thể nhận được đánh giá tin cậy được.

Việc phát hiện, tìm kiếm và sửa chữa các trị đo thô khi giải hệ phương trình (19) dưới điều kiện  $\nabla^T(\chi) \cdot \nabla(\chi) = \min$  chỉ hiệu quả khi sử dụng phương pháp bình sai truy hồi. Trong bài báo này sẽ sử dụng phương pháp hồi với phép biến đổi xoay (thuật toán  $T^{-T}$ -thuận) (xem tài liệu Hà Minh Hòa (2013a)) khi tính đến các khả năng của nó trong việc phát hiện, tìm kiếm các trị đo thô, giảm sự tích lũy của các sai số làm tròn trong quá trình tính toán và giải quyết hiệu quả hệ phương trình chuẩn với ma trận chuẩn xác định không dương. Khi coi biến thứ i là  $L(\chi_i)$ , ( $i=1,2,\dots,n$ ) như trị đo, chúng ta biểu diễn lại phương trình (18) dưới dạng:

$$V(\chi_i) = \bar{F}(\chi_i) \cdot \beta - L(\chi_i), \quad (19)$$

thêm vào đó coi các biến ngẫu nhiên có cùng độ chính xác và có trọng số  $P_i$ .

Để tiện sử dụng tiếp theo, chúng ta ký hiệu  $v_i = V(\chi_i)$ ,  $a_i = \bar{F}(\chi_i)$ ,  $l_i^{(0)} = -L(\chi_i)$ . Khi đó phương trình (19) có dạng:

$$v_i = a_i \cdot \beta_{i-1} + l_i^{(0)}, \quad (20)$$

với trọng số  $P_i$ , ở đây  $\beta_{i-1}$  là giá trị của vector  $\beta$  sau khi đưa vào tính toán truy hồi i - 1 trị đo  $L(\chi)$  đầu tiên.

Việc giải hệ phương trình (20), về nguyên tắc, được thực hiện theo ba bước. Bước 1 được thực hiện để phát hiện sự có mặt của các trị đo thô. Nếu trong các trị đo không có các trị đo thô, thì sau bước 1 việc giải hệ phương trình (20) sẽ kết thúc và chúng ta xác định được một cách tin cậy

vectơ nghiệm  $\beta$ . Trong trường hợp ngược lại, chúng ta phải chuyển qua bước 2 để tìm kiếm và sửa chữa các trị đo thô. Sau khi sửa chữa xong các trị đo thô, chúng ta phải chuyển về bước 1 để giải lại hệ phương trình (20) (bước này được gọi là bước thứ ba). Chúng ta lần lượt xem các bước giải.

**Bước 1:** Giải hệ phương trình (20) kết hợp với việc phát hiện sự có mặt của các trị đo thô. Ma trận chuẩn  $R = \bar{F}^T(\chi) \cdot \bar{F}(\chi)$  bậc  $q \times q$  được khai triển tam giác ở dạng  $R = T^T T$ . Khi đó, ma trận nghịch đảo  $R^{-1} = T^{-1} T^{-T}$ . Trong phương pháp bình sai truy hồi với phép biến đổi xoay, thuật toán  $T^{-T}$  thuận làm việc với ma trận tam giác dưới  $T^{-T}$ . Tổng  $\Phi = [P V V]$  phục vụ việc xác định sai số trung phương đơn vị trọng số  $\mu_0$  sẽ được xác định trong quá trình tính toán truy hồi. Để chuẩn bị tính toán truy hồi, các ma trận ban đầu được nhận ma trận ban đầu  $T_0^{-T} = 10^6 \cdot E_{qq}$ , ở đây  $E_{qq}$  - ma trận đơn vị bậc q; vectơ - cột nghiệm  $\beta_0 = 0$  với kích thước  $q \times 1$  và tổng  $\Phi_0 = 0$ .

Giả thiết rằng sau khi đưa vào tính toán truy hồi i - 1 trị đo  $L(\chi)$  đầu tiên, chúng ta nhận được các ma trận  $T_{i-1}^{-T}$ , vectơ nghiệm  $\beta_{i-1}$  và tổng  $\Phi = \Phi_{i-1}$ . Khi đưa trị đo  $y_i = L(\chi_i)$  với phương trình số cải chính dạng (20) vào tính toán truy hồi, chúng ta đầu tiên tính vectơ

$$t_i = T_{i-1}^{-T} \cdot a_i^T. \quad (21)$$

Trước khi đưa trị đo  $y_i$  vào quá trình tính toán bình sai truy hồi, chúng ta cần tiến hành kiểm tra sự có mặt của các trị đo thô. Trước tiên, chúng ta kiểm tra xem trị đo i là trị đo cần thiết hay trị đo dư. Chúng ta tính trọng số  $g_i$  của số hạng tự do

$$l_i = a_i \cdot \beta_{i-1} + l_i^{(0)} \quad (22)$$

theo công thức sau:

$$g_i = p_i^{-1} + t_i^T t_i. \quad (23)$$

Nếu  $g_i > \frac{100}{p_i}$ , thì trị đo  $y_j$  là trị đo cần thiết. Trong trường hợp này chúng ta không kiểm tra sự có mặt của trị đo thô. Chúng ta chuyển sang quy trình tính toán bình sai truy hồi đối với trị đo  $y_j$  theo thuật toán biến đổi xoay  $T^{-T}$  theo quy trình được trình bày ở dưới đây.

Nếu  $g_i \leq \frac{100}{p_i}$ , thì trị đo  $y_j$  là trị đo dư. Trong trường hợp này, để kiểm tra sự có mặt của các trị đo thô, chúng ta xác định số hạng tự do  $l_j$  theo công thức (22) và so sánh nó với giá trị cho phép

$$(l_i)_{cp} = 3 \cdot m_0 \cdot \sqrt{g_i},$$

ở đây  $m_0$  là sai số trung phương được xác định một cách tiên nghiệm của trị đo  $y_j$  trọng số đảo  $g_j$  của số hạng tự do  $l_j$  được xác định theo công thức (23).

Nếu  $l_i \leq (l_i)_{cp}$ , thì không tồn tại trị đo thô trong  $i$  trị đo đầu tiên được đưa vào tính toán bình sai truy hồi. Trong trường hợp ngược lại, trong  $i$  trị đo đầu tiên có chứa trị đo thô. Chúng ta chuyển sang quy trình tính toán bình sai truy hồi đối với trị đo  $y_j$  theo thuật toán biến đổi xoay  $T^{-T}$  theo quy trình được trình bày ở dưới đây. Sở dĩ đã phát hiện sự có mặt của các trị đo thô, chúng ta không tiến hành tìm kiếm chúng ngay, mà vẫn đưa vào tất cả các trị đo vào tính toán truy hồi rồi mới tiến hành tìm kiếm sau, bởi vì chúng ta cần xác định vector các số cải chính  $V$  của tất cả các trị đo. Điều này mới cho phép áp dụng nguyên tắc modul cực tiểu để tìm kiếm các trị đo thô khi các số cải chính  $V$  của tất cả các trị đo được sử dụng làm trọng số của tất cả các trị đo trong lần lặp đầu tiên.

Quy trình triển khai thuật toán truy hồi  $T^{-T}$  cụ thể như sau. Giả thiết rằng sau khi đưa vào tính toán truy hồi  $i - 1$  trị đo đầu tiên, chúng ta nhận được ma trận  $T_{i-1}^{-T}$ , vector tham số ẩn  $\beta_{i-1}$  và  $[PVV]_{i-1}$ . Bây giờ để đưa vào trị đo  $i$  với phương trình số cải chính (21) vào tính toán truy hồi, chúng ta xác định vector - cột  $t_j$  theo công thức (21), cho vector - hàng  $\{\eta_i\}_0 = 0$  và xác định số

$$\delta^{(0)} = \frac{1}{\sqrt{p_i}}.$$

Chúng ta biến đổi ma trận  $T_{i-1}^{-T}$  lần lượt theo các hàng  $j = 1, 2, \dots, q$ .

(\*) Biến đổi hàng  $J$ :

$$F = \left[ (\delta^{(j-1)})^2 + \{t_i\}_j^2 \right]^{1/2},$$

$$C = \frac{\delta^{(j-1)}}{F}, \quad S = -\frac{\{t_i\}_j}{F},$$

ở đây  $\{t_i\}_j$  - thành phần thứ  $j$  của vector - cột  $t_j$ ,  $C$  và  $S$  - các thành phần của ma trận xoay.

Tiếp theo, đối với hàng  $J$  chúng ta biến đổi các cột  $j_1$  lần lượt từ 1 đến  $J$ . Đối với cột  $j_1$ :

$$(T_i^{-T})_{j,j_1} = C(T_{i-1}^{-T})_{j,j_1} + S\{\eta_i\}_{j_1};$$

$$\{\tilde{\eta}_i\}_{j_1} = C\{\eta_i\}_{j_1} - S(T_{i-1}^{-T})_{j,j_1}.$$

ở đây  $\{\eta_i\}_{j_1}$  - thành phần thứ  $j_1$  của vector - hàng  $\eta_i$  chưa được biến đổi,  $\{\tilde{\eta}_i\}_{j_1}$  - thành phần thứ  $j_1$  của vector - hàng  $\eta_j$  đã được biến đổi.

Sau khi biến đổi xong hàng  $j$ , chúng ta tiến hành xác định số:

$$\delta^{(j)} = C\delta^{(j-1)} - S\{t_i\}_j.$$

Tiếp theo chúng ta chuyển về (\*) để tính toán đối với hàng  $J + 1$  tiếp theo và cứ thế cho đến khi biến đổi xong hàng  $q$ .

Kết quả chúng ta nhận được ma trận  $T_i^T$ , vector - hàng  $\eta_i$  và số  $\gamma_i = \delta^{(q)}$ . Vector

nghiệm  $\beta_i$  và tổng  $[PVV]_i$  được xác định theo các công thức:

$$\beta_i = \beta_{i-1} - \frac{1}{\gamma_i} \cdot \eta_i^T \cdot l_i;$$

$$[PVV]_i = [PVV]_{i-1} + \varepsilon_i^2,$$

ở đây số hạng tự do  $l_i$  được xác định theo công thức (22), còn số  $\varepsilon_i = \frac{l_i}{\gamma_i}$ .

Chúng ta tiếp tục thực hiện quy trình toán toán truy hồi nêu trên đối với các trị đo tiếp theo  $i+1, \dots, n$ .

Vectơ nghiệm cần tìm  $\beta = \beta_n$ . Sai số trung phương đơn vị trọng số sau bình sai được xác định theo công thức:

$$\mu = \pm \sqrt{\frac{\Phi_n}{n-q}}.$$

**Bước 2:** Tìm kiếm và sửa chữa các trị đo thô

Trong quá trình đưa  $n$  trị đo vào tính toán truy hồi theo thuật toán truy hồi  $T^{-T}$  ở bước 1, khi đã phát hiện được sự có mặt của các trị đo thô, chúng ta sẽ tiến hành tìm kiếm chúng. Trong tài liệu (Hà Minh Hòa (2013a)) đã đề xuất phương pháp tìm kiếm các trị đo thô theo nguyên tắc mô đun cực tiểu dưới điều kiện:

$$\sum_{i=1}^n |\bar{v}_i| = \min, \quad (26)$$

ở đây  $\bar{v}_i = p_i^{1/2} \cdot v_i$ , còn  $p_i$  là trọng số của trị đo  $y_i$ .

Điều kiện (26) tương ứng với trường hợp phân bố Laplace (xem chi tiết trong tài liệu Hà Minh Hòa (2013b)). Để triển khai điều kiện (26), trong tài liệu (Hà Minh Hòa (2013a)) đã đề xuất sử dụng phương pháp bình phương nhỏ nhất lặp và còn được gọi là phương pháp Fletcher – Grand - Hebden (Fletcher R., Grand J.A, Hebden M.D.

(1971)), theo đó điều kiện (26) được biến đổi dưới dạng:

$$\sum_{i=1}^n |\bar{v}_i| = \sum_{i=1}^n \bar{p}_i \cdot \bar{v}_i^2 = \min, \quad (27)$$

ở đây  $\bar{p}_i = \frac{1}{|\bar{v}_i|}$ .

Dưới dạng xử lý tính toán lặp, tại lần lặp thứ  $m$  ( $m = 1, 2, \dots$ ), điều kiện (27) có dạng sau:

$$\sum_{i=1}^n \bar{p}_i^{(m)} \cdot (\bar{v}_i^{(m)})^2 = \min, \quad (28)$$

ở đây  $\bar{p}_i^{(m)} = \frac{1}{|\bar{v}_i^{(m-1)}|}$ , thêm vào đó khi lưu ý phương trình (II.23), số cải chính của trị đo  $y_i$  ( $i = 1, 2, \dots, n$ ) được xác định theo công thức:

$$\bar{v}_i^{(m-1)} = p_i^{1/2} \left( a_i \cdot \beta_{i-1}^{(m-1)} - L(x_i) \right).$$

Sự giảm dần của  $\sum_{i=1}^n |\bar{v}_i^{(m)}|$  sau mỗi lần lặp  $m$  ( $m = 1, 2, \dots$ ) và sự hội tụ của quá trình giải lặp theo điều kiện (28) đã được chứng minh trong tài liệu (Hà Minh Hòa (2013a)). Như vậy, sau mỗi lần lặp  $m$ , chúng ta có quan hệ:

$$|\bar{v}_i^{(m)}| \leq |\bar{v}_i^{(m-1)}|.$$

Dựa trên bất đẳng thức Trébussep, đối với các trị đo ngẫu nhiên tuân theo phân bố chuẩn, điều kiện (24) dẫn đến việc các số cải chính  $v$  sẽ hội tụ về 0 (về kỳ vọng toán học của chúng), còn đối với các trị đo thô với các sai số thô tuân theo phân bố Laplace, các số cải chính sẽ tiến đến các giá trị của các sai số thô (xem tài liệu Hà Minh Hòa (2013b)). Quá trình tính toán lặp được dừng lại (hội tụ) khi đối với tất cả các trị đo  $L(x_i)$  ( $i=1, 2, \dots, n$ ) thỏa mãn điều kiện  $\left| \left[ \bar{v}_i^{(m)} \right] - \bar{v}_i^{(m-1)} \right| \leq 0,0001$ . Bằng cách như vậy, sau khi quá trình tính toán lặp hội tụ, dựa trên độ lớn của các số cải chính sau lần lặp cuối cùng, chúng ta sẽ xác định được các trị đo thô.

Để bắt đầu quá trình tính toán lặp chúng ta phải xác định các số cải chính ban đầu

$v_i^{(0)} = v_i$  ( $i=1,2,\dots,n$ ) dựa trên các kết quả xác định vector  $\beta_n$  đã được xác định ở bước 1, theo đó

$$v_i^{(0)} = a_i \cdot \beta_n + l_i^{(0)}.$$

Sau khi phát hiện và sửa chữa các trị đo thô, chúng ta quay lại bước 1 để tiến hành lại từ đầu quá trình tính toán với n trị đo theo thuật toán  $T^{-T}$  thuận dựa trên quy trình đã được trình bày ở trên để xác định một cách tin cậy vector nghiệm  $\beta$ .

### 3. Kết luận

Bài toán nội suy xác định các giá trị dị thường trọng lực tại các đỉnh của các ô chuẩn trong CSDL dị thường trọng lực quốc gia từ các giá trị dị thường trọng lực trên các điểm trọng lực là bài toán khoa học - kỹ thuật rất phức tạp. Do các điểm trọng lực phân bố trên bề mặt địa hình, nên các giá trị trung bình xác suất (trend) của các giá trị dị thường trọng lực không bằng nhau, nên trong thực tế ở các nước phát triển đã sử dụng phổ biến phương pháp kriging tổng quát để giải quyết bài toán nêu trên. Tuy nhiên, với số lượng các điểm trọng lực rất lớn phân bố trên phạm vi lớn của lãnh thổ quốc gia và được xây dựng trong giai đoạn dài, việc phát hiện, tìm kiếm và sửa chữa các giá trị dị thường trọng lực có chứa các sai số thô là một nhiệm vụ bắt buộc. Nhưng các đề xuất hiện nay trong các tài liệu quốc tế chưa có các phương pháp tin cậy để thực hiện nhiệm vụ nêu trên. Bài báo khoa học này đã đề xuất sử dụng triển khai thuật toán truy hồi  $T^{-T}$  để thực hiện hiệu quả nhiệm vụ này với mục đích xác định một cách tin cậy các hệ số  $\beta$  của mô hình trend (10). Nhiệm vụ tiếp theo xác định một cách tin cậy vector  $\lambda^T$  theo thuật toán truy hồi  $T^{-T}$  sẽ được nghiên cứu trong bài báo khoa học tiếp theo. ○

### Tài liệu tham khảo

- [1]. Anberto Molteni, Lisa Pertusini, Mirko Reguzzoni (2009). Collocation and Kriging. Statistical Analysis of Enviromental Data - Academic Year 2008 - 2009, 28 p. [www.geomatica.como.polimi.it/.../Trattamentol Es.5\\_ FILTER](http://www.geomatica.como.polimi.it/.../Trattamentol Es.5_ FILTER).
- [2]. Chauvet P and Galli A. (1982). Universal kriging Course. - C-96, Centre de Geostatistique, Ecole des Mines de Paris.
- [3]. Cressie N.A.C. (1993). Statistics for spatial data. John Wiley & Sons, New York. 900 p.
- [4]. Hà Minh Hòa (2013a). Phương pháp bình sai truy hồi với phép biến đổi xoay. Nhà Xuất bản Khoa học và Kỹ thuật, 244 trg. Hà Nội - 2013.
- [5]. Hà Minh Hòa (2013b). Phương pháp xử lý toán học các mạng lưới trắc địa quốc gia. Nhà Xuất bản Khoa học và Kỹ thuật, 244 trg. Hà Nội - 2013.
- [6]. Hà Minh Hòa (2014a). Lý thuyết và thực tiễn của Trọng lực trắc địa. Nhà Xuất bản Khoa học và Kỹ thuật, 592 trg. Hà Nội - 2014.
- [7]. Hà Minh Hòa, Nguyễn Tuấn Anh (2014b). Nghiên cứu khả năng hiệu chỉnh các hệ số khai triển điều hòa cầu của thế trọng trường Quả đất của mô hình EGM2008 dựa trên các dữ liệu đo trọng lực chi tiết ở Việt Nam. Báo cáo khoa học. Kỷ yếu Hội nghị Khoa học "Trắc địa và Bản đồ vì hội nhập quốc tế" ngày 08/07/2014. Viện Khoa học Đo đạc và Bản đồ, Hội Trắc địa, Bản đồ và Viễn thám Việt Nam, trg. 21 - 37.
- [8]. Kolmogorov A.N. (1933). Foundations of the Theory of Probability. Chelsea Publishing, New York.
- [9]. Jay D. Martin, Timothy W. Simpson (2004). On the use of kriging models to



approximate deterministic computer models. Proceedings of DTEC'04: ASME 2004 International Design Engineering Technical Conference and Computers and Information in Engineering Conference Salt Lake City, Utah USA, September 28 - October 2, 2004. pp. 1 - 12.

[10]. Library A.L.T. (1998). Principles of Geographical Information System. Oxford University Press . 1998. 333 p.

[11]. Marcin Ligas, Marek Kulczycki (2010). Simple spatial prediction - least squares prediction, simple kriging, and conditional expectation of normal vector. J. Geodesy And Cartography, Vol. 59, N02, pp. 69-81. Polish Academy of Sciences.

[12]. Marcin Ligas, Marek Kulczycki (2014). Kriging approach for local height transformations. J. Geodesy And Cartography, Vol. 63, N01, pp. 25-37. Polish Academy of Sciences. Doi: 10.2478/geo-cart-2014-0002.

[13]. Moritz H. (1980). Advanced Physical Geodesy. Herbert wichmann Verlag Karlsruhe, Abacus Press Tunbridge Wells Ken<sup>T</sup>, 512 p.

[14]. Olea R. A. (1999). Geostatistics for engineers and earth scientists: Norell, Mass., Boston, Kluwer Academic Publishers, 313 p.

[15]. Reguzzi M., Sansò F. and Venuti G. (2005). The theory of general kriging, with applications to the determination of a local geoid. Geophys. J. Int., 162, 303 - 314, doi: 10.1111/j.1365-246X.2005.02662.x

[16]. Schabenger O., Gotway C.A. (2005). Statistical methods for spatial data analysis. Chapman & Hall/CRC, New York.

[17]. Wiener N. (1949). Extrapolation, Interpolation and Smoothing of Stationary Time Series, MIT Press, Cambridge, MA.○

## Summary

### **Approach to detection of outliers in gravity anomaly data for construction of state gravity anomaly database by general kriging method**

*Assoc. Prof. Dr. Sc. Ha Minh Hoa*

*Vietnam Institute of Geodesy and Cartography*

A task of interpolation of gravity anomalies at points of gravity anomaly grid from gravity anomaly values on the gravimetric points is very complicated science - technical one. At present general kriging method is widely used for solving abovementioned task. However an accurate determination of coefficients  $\beta$  of the trend model requires detection of outliers in gravity anomaly data. This scientific article proposes using recurrent algorithm  $T^{-T}$  for effective solution of detection of outliers in gravity anomaly data.○

*Ngày nhận bài: 15/01/2015.*