

# Diagnosis of the Ill-condition of the RFM Based on Condition Index and Variance Decomposition Proportion (CIVDP)

ZHOU QING<sup>1,2</sup>, JIAO WEILI<sup>1</sup>, LONG TENGFEI<sup>1</sup>

<sup>1</sup>Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China.

<sup>2</sup>Graduate University of Chinese Academy of Sciences, Beijing, China.

wljiao@ceode.ac.cn

**Abstract.** The Rational Function Model (RFM) is a new generalized sensor model. It does not need the physical parameters of sensors to achieve a high accuracy that is compatible to the rigorous sensor models. At present, the main method to solve RPCs is the Least Squares Estimation. But when coefficients has a large number or the distribution of the control points is not even, the classical least square method loses its superiority due to the ill-conditioning problem of design matrix. Condition Index and Variance Decomposition Proportion (CIVDP) is a reliable method for diagnosing the multicollinearity among the design matrix. It can not only detect the multicollinearity, but also can locate the parameters and show the corresponding columns in the design matrix. In this paper, the CIVDP method is used to diagnose the ill-condition problem of the RFM and to find the multicollinearity in the normal matrix.

## 1. Introduction

The Rational Function Model (RFM) is a new generalized sensor model. It can achieve a high accuracy that compatible to the rigorous sensor models [1]. Some high resolution satellite imagery providers keep the sensor parameters confidential for commercial reasons, and it becomes difficult to develop a physical sensor model. For such reason, more and more researchers divert their attention to the RFM. The key to use RFM to process the satellite image is to find the precise rational polynomial coefficients(RFCs). At present, the main method to solve RPCs is the Least Squares Estimation. But when coefficients has a large number or the distribution of the control points is not even, the classical least square method loses its superiority due to the ill-conditioning problem of design matrix. The ill-conditioning problem is usually caused by the multicollinearity that exists among the columns of the design matrix. A lot of methods have been proposed to diagnose the ill-condition of the normal equation including the condition number method and the CIVDP method in data processing field [2,3]. Some researchers have proposed new methods to overcome the ill-condition of RFM, such as ridge estimation and Artificial Intelligence [4,5,6]. But the research on the diagnosis of the ill-condition of RFM is limited.

CIVDP is a reliable method for diagnosing the multicollinearity among the design matrix. It can not only detect the multicollinearity, but also can locate the parameters and show the corresponding



columns in the design matrix [7]. In this paper, the CIVDP method is used to diagnose the ill-condition problem of the RFM and to find out the multicollinearity among normal matrix. Based on the diagnosis, the model parameters can be divided into two categories, affected or not affected by the multicollinearity.

The remainder of this paper is organized as follows. The RFM is briefly introduced in Section2. The diagnosing method of CIVDP and the main steps of implementation is described in Section3. Experiments and analysis are provided in Section4. Finally, conclusions are drawn in Section5.

## 2. Rational function model

In the rational function model, the pixel coordinates  $(r, c)$  are expressed as the ratio of polynomials containing the ground coordinates  $(X, Y, Z)$ . It can be expressed as the following equation [8]:

$$r_n = \frac{p_1(X_n, Y_n, Z_n)}{p_2(X_n, Y_n, Z_n)} \quad (2.1)$$

$$c_n = \frac{p_3(X_n, Y_n, Z_n)}{p_4(X_n, Y_n, Z_n)} \quad (2.2)$$

The polynomials can be described as the following equation:

$$\begin{aligned} p = \sum_{i=0}^{m1} \sum_{j=0}^{m2} \sum_{k=0}^{m3} a_{ijk} X^i Y^j Z^k = & a_1 + a_2 Z + a_3 Y + a_4 X + a_5 ZY \\ & + a_6 ZX + a_7 YX + a_8 Z^2 + a_9 Y^2 + a_{10} X^2 + a_{11} ZYX \\ & + a_{12} Z^2 Y + a_{13} Z^2 X + a_{14} Y^2 Z + a_{15} Y^2 X + a_{16} ZX^2 \\ & + a_{17} YX^2 + a_{18} Z^3 + a_{19} Y^3 + a_{20} Z^3 \end{aligned} \quad (2.3)$$

There are totally 80 parameters for the complete 3-order form of RFM, and except for  $b_1 = d_1 = 1$  there remains 78 parameters including  $a_1 \sim a_{20}, b_2 \sim b_{20}, c_1 \sim c_{20}, d_2 \sim d_{20}$ .

In equation (2.1) and (2.2),  $r_n$  and  $c_n$  are the normalized pixel coordinates,  $X_n, Y_n$  and  $Z_n$  are the normalized ground point coordinates. They can be transferred from the normalization equations:

$$\begin{aligned} r_n &= \frac{r - r_0}{r_s} \\ c_n &= \frac{c - c_0}{c_s} \\ X_n &= \frac{X - X_0}{X_s} \\ Y_n &= \frac{Y - Y_0}{Y_s} \\ Z_n &= \frac{Z - Z_0}{Z_s} \end{aligned} \quad (2.4)$$

In equation (2.4),  $(r_0, c_0, X_0, Y_0, Z_0)$  are the standardized translation parameters,  $(r_s, c_s, X_s, Y_s, Z_s)$  are the standardized scaling parameters. The multinomial coefficients are the RPCs needed to be solved.

The number of polynomial coefficients needed to be solved and the minimum GCPs needed under the nine different RPC models are shown in table 1. When the denominators of the RPC model are the same and equal to 1 (cases 7~9), the RPC model becomes the conventional polynomial model. In the case of  $p_2 = p_4$  with the first order (case 4 in table 1), the model degrades into the direct linear transformation model.

**Table 1.** Nine different forms of the RFM.

Case	Denominator	Orders of polynomials	Number of RFCs	Min number of GCPs
1	P2=P4=1	1	8	4
2		2	20	10
3		3	40	20
4	P2=P4	1	11	6
5		2	29	15
6		3	59	30
7	P2≠P4	1	14	7
8		2	38	19
9		3	78	39

### 3. Principles of CIVDP method

The ill-conditioning problem of the RFM is usually caused by the multicollinearity in the design matrix [9]. When the multicollinearity reaches a certain extent, the normal matrix will get very small eigenvalue, and the Least Square estimation will become significantly bad. CIVDP method is generalized and improved from the method of traditional condition number. It can diagnose the existence, number and composition of the multicollinearity in the columns of the design matrix.

Considering the Gauss-Markov model of the survey adjustment in equation (3.1):

$$\begin{cases} L = AX + \Delta \\ E(\Delta) = 0, Cov(\Delta) = \sigma_0^2 P^{-1} \end{cases} \quad (3.1)$$

Where,  $L$  is the  $n \times 1$  observation vector,  $P$  is the weight matrix,  $A$  is the  $n \times t$  design matrix and  $\Delta$  is the  $n \times 1$  observation error vector.

The Least Squares estimation of the parameter  $X$  can be expressed as below:

$$\hat{X}_{LS} = (A^T P A)^{-1} A^T P L \quad (3.2)$$

The principles and the implementing steps of the method of CIVDP are as follows [10]:

- First, normalize the design matrix  $A$ ;
- Spectral decomposing the normal matrix  $A^T P A$ ;

$$A^T P A = Q \Lambda Q^T \quad (3.3)$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_t), Q = (q_{ij}) = (q_1, \dots, q_t)$$

Where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t > 0$  are the eigenvalue of normal matrix  $A^T P A$  and  $q_1, \dots, q_t$  are the eigenvector to the corresponding eigenvalue.

- Calculating the value of Condition Index(CI)  $\eta_k$ :

$$\eta_k = \frac{\lambda_1}{\lambda_k}, (k = 1, \dots, t) \quad (3.4)$$

Check the value of  $\eta_k$  : if it is larger than the given threshold value, there is multicollinearity in the design matrix.

- Calculating the Variance Decomposition Proportion(VDP):

The variance of the  $k$ – $th$  unknown parameter can be described as the following formula:

$$\text{Var}(X_{LS}^{(k)}) = \sigma_0^2 \sum_{j=1}^t \frac{q_{kj}^2}{\lambda_j} = \sigma_0^2 \left( \frac{q_{k1}^2}{\lambda_1} + \dots + \frac{q_{kj}^2}{\lambda_j} + \dots + \frac{q_{kt}^2}{\lambda_t} \right) \quad (3.5)$$

And the ratio of the  $j$ – $th$  term of the  $k$ – $th$  unknown parameter variance formula to the  $k$ – $th$  unknown parameter variance is the Variance Decomposition Proportion of the term  $(k, j)$ , and marked as  $\pi_{jk}$ :

$$\pi_{jk} = \frac{\phi_{kj}}{\phi_k}, k, j = 1, \dots, t \quad (3.6)$$

$$\phi_{kj} = \frac{q_{kj}^2}{\lambda_j}, \phi_k = \sum_{j=1}^t \phi_{kj}$$

So a  $t \times t$  Variance Decomposition Proportion matrix  $\Pi = (\pi_{jk})$  can be obtained.

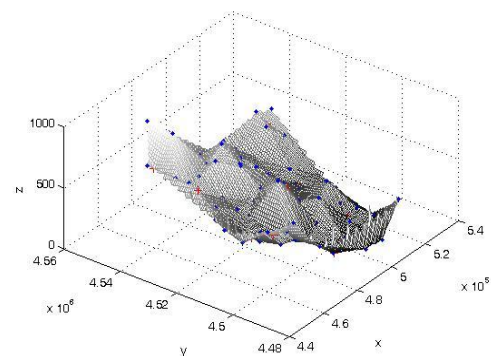
- Find the VDP value that  $\pi_{jk} > 0.5$ , and then sort out the columns that affected by the multicollinearity;

#### 4. Experiment and analysis

In order to verify the validation of the CIVDP method in diagnosing the morbid of the RFM, an imagery of SPOT5 was selected as the experimental data. The data were collected in Beijing, China on 2nd June, 2004. The area is 60kilometers×60 kilometers and the image size is24000×24000 pixels. The test region includes both mountain areas and flat resident areas as shown in figure.1. The elevation of ground control points ranges from 150 meters to 879.74meters. In this paper, we use terrain-dependent method to solve the RFM. Eighty GCPs were collected from an ortho-rectified SPOT5 image with 2.5 meters resolution by using the PCI software, and another 12 checkpoints were also collected from it. The 3-D view of the distribution of these control points (marked by blue dot “•”) and checkpoints (marked by red cross “+”) is given in figure.2.



**Figure 1.** The 2.5 meters resolution SPOT5 imagery of Beijing area (part of the image).



**Figure 2.** The 3-D view of the terrain surface and the distribution of GCPs(•) and CPs(+).

##### 4.1. Morbid diagnosis

Using the CIVDP method to diagnose the morbid of the RFM with the principles and main steps discussed in section 3, and finally we can find out the parameters (in RFM we call it RPCs) that affected by the multicollinearity exist in the design matrix. The parameters are shown in table 2.

**Table 2.** RPCs that affected by multicollinearity diagnosed with CIVDP method.

RPC type	RPCs influenced by morbid
a	$a_7$ 、 $a_{11}$ 、 $a_{13}$ 、 $a_{17}$
b	$b_3$ 、 $b_6$ 、 $b_8$ 、 $b_9$ 、 $b_{13}$ 、 $b_{18}$ 、 $b_{20}$
c	$c_1$ 、 $c_5$ 、 $c_6$ 、 $c_{11}$ 、 $c_{14}$ 、 $c_{16}$ 、 $c_{19}$ 、 $c_{20}$
d	$d_2$ 、 $d_5$ 、 $d_6$ 、 $d_{11}$ 、 $d_{12}$ 、 $d_{15}$ 、 $d_{16}$ 、 $d_{19}$ 、 $d_{20}$

The corresponding columns of the influenced parameters by morbid in the design matrix are list below:  
7、11、13、17、23、26、28、29、33、38、40、44、45、50、53、55、58、59、61、64、65、70、71、74、75、78.

#### 4.2. Geometric rectification

In order to testify that the multicollinearity detected by the CIVDP method are really the reason to cause the morbid of RFM, we give up the parameters influenced by the morbid and remove the corresponding columns, and then use the remaining parameters to do image geometric rectification. A comparison was made with the traditional least squares estimation with no parameters given up. The results are shown in table 3.

**Table 3.** Image geometric rectification accuracy (unit: pixel).

Model	RMSE of control points			RMSE of check points			Condition number
	RMSE of row coordinate	RMSE of column coordinate	RMSE	RMSE of row coordinate	RMSE of column coordinate	RMSE	
Optimized by CIVDP	0.5137	0.4732	0.6984	0.6590	0.8185	1.0508	$9.8 \times 10^5$
Original model	5.108	22.2637	22.8422	1.8639	72.5126	72.5365	$3.7 \times 10^{13}$

The geometric rectification precision of the two different model are shown in table 3, and we can draw comprehensive conclusions from the experimental results.

- From table 3 we can see that the precision of the original model solved by traditional least square estimation is very low, the condition number of the design matrix is  $3.7 \times 10^{13}$ , which means the normal equation becomes severely ill-conditioned. The root mean square error (RMSE) of the check points is 72.5365 pixels and that is far from the precision that needed.
- The precision of the model optimized by the CIVDP morbid diagnosing method is significantly improved. The RMSE of control points is within one pixel and the precision of check points is also nearly one pixel. The precision can satisfy most applications of the imagery.
- It proves that CIVDP method is an effective method for diagnosing the morbid of RFM. It can precisely locate the parameters and the corresponding design matrix columns that influenced by the multicollinearity.

#### 5. Conclusion

When the number of the parameters is large or the distribution of the control points is not well, the RFM often becomes ill-conditioned and the least square estimation cannot get reliable solution. In this

paper we brought CIVDP method to diagnose the morbid of RFM and sort out the parameters and columns in design matrix that influenced by multicollinearity. Furthermore, an optimized model is generated by abandoning the parameters that influenced by the morbid. And by applying the simplified model to do SPOT5 image geometric correction, the results prove that the CIVDP method is effective in the morbid diagnosing of RFM.

### Acknowledgments

The research has been supported by the grants from the National Natural Science Foundation of China (61271013).

### References

- [1] Dowman I and Dolloff J.T 2000 *J. An Evaluation of Rational Functions for Photogrammetric Restitution. International Archives of Photogrammetry Engineering and Remote Sensing.* **58** 85-89
- [2] Liang B and Wen D Q 2007 *Research of Ill-conditioned Equation in GPS Data Processing* (Nanjing: Nanjing University)
- [3] Guo J F 2002 *Diagnostics and Processing of Ill-conditioning Adjustment System* (Zhenzhou: The PLA Information Engineering University)
- [4] Hoerl A E and Kennard R W 1970 *J. Ridge Regression: Biased Estimation for Non-orthogonal Problems. Technometrics.* **12** 55-57
- [5] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-posed Problems* (New York: Wiley)
- [6] Zhu W Q and Jiao W L 2008 *J. Solving Rational Function Model with Genetic Algorithm. Science Technology and Engineering* **8** 3530-3535
- [7] Zhang L, Gu Y W, Gui Q M and Ma C Z 2007 *J. Biased Estimator Based on Diagnosis and Measure of Multicollinearity. Journal of Geodesy and Geodynamics* **27** 99-102
- [8] Zhang J, Wen H Y and Li L 2006 *J. Discussion on the Ill-conditioned Problems in Normal Equation of Geodesy. Hydrographic Surveying and Charting* **26** 1-3
- [9] Belsey D A 1991 *Conditioning diagnostics: Collinearity and Weak Data in Regression* (New York: John Wiley& Sons)
- [10] GUI Q M, YAO S W and Gu Y W 2006 *J A New Method to Diagnose Multicollinearity Based on Condition Index and Variance Decomposition Proportion (CIVDP). Acta Geodaetica et Cartographia Sinica.* **35** 210-14