Lecture Notes in Electrical Engineering 323

Rongjun Shen
Weiping Qian
Editors

Proceedings of the 27th Conference of Spacecraft TT&C Technology in China

Wider Space for TT&C

清華大學出版社
TSINGHUA UNIVERSITY PRESS

Springer

Lecture Notes in Electrical Engineering

Volume 323

About this Series

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

More information about this series at http://www.springer.com/series/7818

Rongjun Shen · Weiping Qian
Editors

Proceedings of the 27th Conference of Spacecraft TT&C Technology in China

Wider Space for TT&C

清華大学出版社
TSINGHUA UNIVERSITY PRESS

Springer

Editors
Rongjun Shen
Chinese Academy of Engineering
Beijing
China

Weiping Qian
Beijing Institute of Tracking
and Telecommunications Technology
Beijing
China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
ISBN 978-3-662-44686-7 ISBN 978-3-662-44687-4 (eBook)
DOI 10.1007/978-3-662-44687-4

Jointly published with Tsinghua University Press, Beijing
ISBN: 978-7-302-38190-7 Tsinghua University Press, Beijing

Library of Congress Control Number: 2014949294

Springer Heidelberg New York Dordrecht London

© Tsinghua University Press, Beijing and Springer-Verlag Berlin Heidelberg 2015
This work is subject to copyright. All rights are reserved by the Publishers, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publishers' locations, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publishers can accept any legal responsibility for any errors or omissions that may be made. The publishers makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

As a result of over 50 years of development, China has already built state-of-the-art full-fledged spacecraft TT&C systems. TT&C means are more diversified, their capabilities are greatly uplifted, and the scope of work of TT&C systems has been expanding. TT&C systems play an indispensable role in fields including space launch missions and routine operation management of spacecraft.

With the prosperous development of China's space endeavors, TT&C systems are facing new challenges and opportunities. They are required to support deep space missions farther into the universe. They are required to provide more complicated and higher accuracy on-orbit management of spacecraft. They are required to maintain formations and constellations of more satellites and to play a bigger role in detection and cataloguing of smaller space objects. To meet the new requirements, TT&C systems have to enhance their long-distance space information transmission capability, weak signal extraction and processing capability, high accuracy space instrumentation and navigation capability, and long delay operation and control capability. Moreover, the accuracy of TT&C systems has to be further increased to provide more reliable, secure, flexible, and efficient support to spacecraft.

Taking "wider space for TT&C" as its theme, the 27th Conference of Spacecraft TT&C Technology of China highlights more utilization of modern technologies to lift the effectiveness of spacecraft TT&C systems to meet the demands of long-term development of space activities in the backdrop of China's growing deep space missions, maturing Beidou satellite navigation system and explosive workload on China's spacecraft TT&C systems.

From over 330 papers authored by scholars and specialists from different fields, 55 are selected for publication by Springer. The objective is to further increase the influence of the Spacecraft TT&C Committee of the Chinese Society of Astronautics and to promote international academic exchanges by sharing China's latest research achievements and engineering experiences in the field of spacecraft TT&C systems with the global space-faring community.

November 2014 Rongjun Shen

Contents

Part I Spacecraft Instrumentation Science and Technology

Part II Space Mission and Operations

Part III Space Object Exploration and Identification

Part IV Communication and Information Systems

Part I

Spacecraft Instrumentation Science and Technology

Chapter 1
The Application of the GNSS Receiver in the Third Stage of China Lunar Exploration Program

Changyuan Wen, Meng Wang, Lin Qi, Dongjun Li and Yuehua Qiu

Abstract Lunar exploration puts forward higher requirements in terms of performances for the subsystems. As for the navigation system, the GPS receiver provides autonomous navigation service on board in Low Earth Orbit environment, so it calls for a possible extension to Earth-to-Moon missions. This paper, referring to the Third Stage of China Lunar Exploration Program mission which is “returning” in the three stages “circling landing and returning”, investigates such a possibility, considering present and foreseen available GNSS signals. Carrier-to-noise levels achievable during the mission are evaluated. Moreover, the high-sensitivity GNSS receiver system design is discussed.

Keywords Lunar exploration · GNSS receiver · High sensitivity · Earth orbit determination

1.1 Introduction

With the successful task completion of Chang’E-1, Chang’E-2 and Chang’E-3, marking China’s lunar exploration project into the Third Stage of China Lunar Exploration Program mission. According to three step strategy, which is “circling landing and returning”, the Third Stage of China Lunar Exploration Program mission will be mainly probing on the moon landing and return. How to successfully landing on the moon, and accurately returns to the earth, asks for higher request to the track measurement and control detector.

At present, China existing deep space exploration spacecraft mainly use ground navigation and celestial navigation as the main means of navigation for orbit

---

C. Wen (✉) · M. Wang · D. Li · Y. Qiu
Space Star Technology Co., Ltd, Beijing 100086, China
e-mail: achang11241124@126.com

L. Qi
DFH Satellite Co., Ltd, Beijing 100094, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_1

parameters measurement. Ground navigation takes the VLBI as the main means, which has the limitation of the ground station, data transmission, and the real-time and the measurement error, so it is difficult to meet the requirements of high precision measurement of orbit. The GNSS receiver is mainly applied in the low orbit satellite, which is used for navigation and positioning. With the deep research of GNSS receiver, the use of GNSS receiver in the above GNSS constellation orbit, which received the GNSS signal from the earth opposite especially side-lobe, got more development, such as America AMSAT OSCAR-40 satellite during apogee 59,000 km has successfully received the GPS signal [1, 2].

So using the GNSS receiver in the lunar mission was considered for real-time receiving GNSS satellite signal, through which the three-dimensional velocity and three-dimensional position, UTC time was calculated, those results were put together with other inertial navigation system to constitute the integrated navigation system, which can be used to correct the orbit errors, realize the transfer phase of the orbit control, with real-time, low power consumption, small size, high ratio of performance to price advantage.

This paper is based on the GNSS receiver in the application of the Third Stage of China Lunar Exploration Program mission as the background. The high sensitivity receiver and GNSS autonomous orbit determination of Kalman filter, which is to provide the navigation service for the lunar spacecraft, was designed through the analysis of the visibility of GNSS receiver for the Earth-Moon-Earth orbit transfer task and the characteristics of the weak GNSS signal power.

1.2 Lunar Mission Availability for GNSS

Take the Chang'E-1 transfer orbit parameters as an example to analyze the availability of GNSS signal in the Earth-Moon-Earth transfer orbit. The transfer instantaneous orbit elements are numbered in the Table 1.1 [3]:

Earth-Moon-Earth transfer orbit is highly elliptical orbit (HEO), with perigee 200 km and apogee 380,000 km. Figure 1.1 shows the relative position of the HEO satellites and GNSS geometric diagram where ε is unilateral GNSS satellite antenna coverage angle, Φ is the angle between the GNSS satellites unilateral main-lobe, θ is the angle between the GNSS satellites and Earth unilateral tangent. α is for the receiver and GNSS satellite receiver connection point geocentric angle between the

Table 1.1 Earth–Moon transfer orbit elements for simulation

Epoch	2007/4/23 22:41:41
Semi-major axis (km)	212,857.337
Eccentricity	0.967199 336
Inclination angle (°)	30.983
Argument of perigee (°)	179.983
RAAN (°)	180.485
True anomaly (°)	0

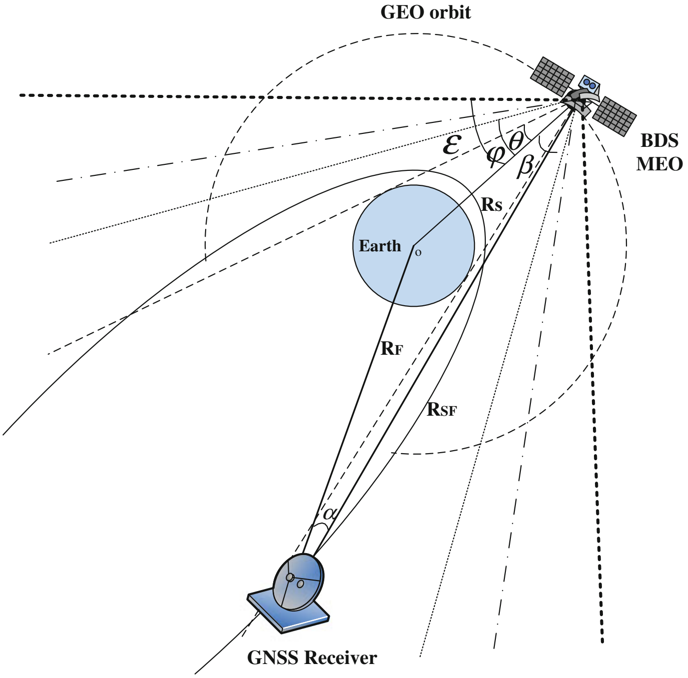


Fig. 1.1 Earth—Moon—Earth orbit GNSS signal receiver geometry schematic

direction, β is the angle between a link of the GNSS satellite to receiver and the direction of GNSS satellite pointing to center of the earth.

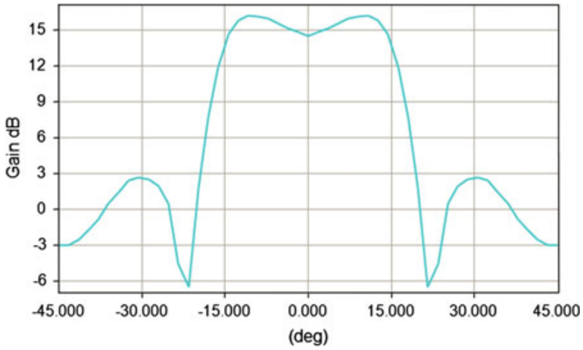
Seen from the Fig. 1.1, When the GNSS receiver orbit height less than GNSS satellite orbit height, it is required to receive GNSS signal from the zenith; When the GNSS receiver is higher than the height of the orbit of GNSS satellite orbit height, it is required to receive GNSS signal from the earth opposite for positioning.

1.2.1 Simulation Setting

According to the official GPS website information, into the March 14, 2013 satellite TLE file, the GPS constellation normal in operation is 31 satellites, not available is PRN 27. According to the published ICD, BDS design the constellation composed of 5 geostationary orbit (GEO) satellites, 27 circular earth orbit (MEO) satellites and 3 inclined geosynchronous orbit (IGSO) satellites [4].

GPS transmitting power is set to 14.28 dBW, GPS satellite antenna gain pattern (see Fig. 1.2) is saddle shaped gain, maximum gain value is 16.22 dB. Transmitting power and antenna gain of BDS is made by GPS.

Fig. 1.2 Transmitting antenna gain pattern

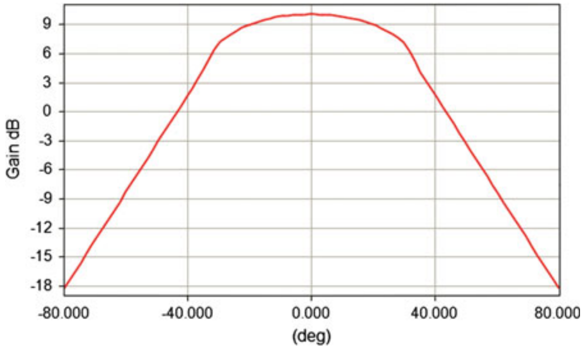


Consider in the ascent of Earth-Moon transfer orbit and the decent of Moon-Earth transfer orbit, when the orbit height less than GNSS satellite orbit height, it needs to receive the zenith signal; and the track height is higher than that of GNSS satellite orbit height, it needs to receive signals from the earth opposite, so GNSS receiver for lunar phase three use double antennas, antenna pointing to the earth and from the earth. In order to adapt to the different stages of task demands, the direction point to the zenith will use the high gain antenna, another direction will be the wide field receiving antenna.

The highest gain of the high gain antenna (see Fig. 1.3) is 10.1 dB, ensure the ±30° gain greater than 7 dBi. Wide field receiving antenna (see Fig. 1.4) installation point to the zenith, the gain is more than 0 dBi for the ±60° beam width.

According to the analysis of GNSS constellations and lunar exploration satellite geometry, and the GNSS signal link budget equation, and taking into account the current level of receiver design, it is known that in the Earth-Moon transfer orbit on the position of 100,000 km the received power (including high-gain receiving antenna reception gain) is about –175 dBW maximum [2], so the simulation is set to receive power threshold –175 dBW, the minimum elevation angle of cut in 5°.

Fig. 1.3 Received antenna gain patterns towards the down-side



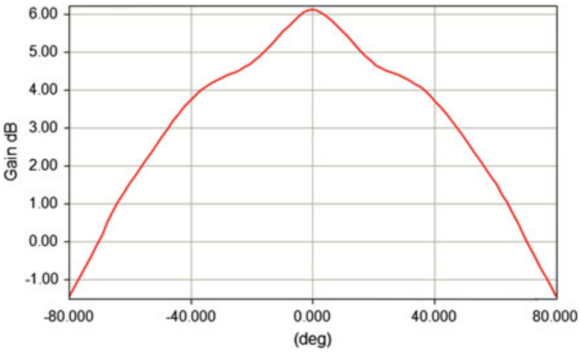


Fig. 1.4 Receive antenna gain patterns towards the up-side

1.2.2 *Earth–Moon–Earth Transfer Process GNSS Availability*

1.2.2.1 Earth to Moon Transition Phase

In view of earth-moon transfer stage, in the stage of altitude 100,000 km below, analysis the number of visible GNSS satellites (see Fig. 1.5), Doppler range (see Fig. 1.6) and signal duration (see Fig. 1.7) and other information.

1.2.2.2 Month to Earth Stage

In view of moon-earth transfer stage, in the stage of altitude 100,000 km below, analysis the number of visible GNSS satellites (see Fig. 1.8), Doppler range (see Fig. 1.9) and signal duration (see Fig. 1.10) and other information.

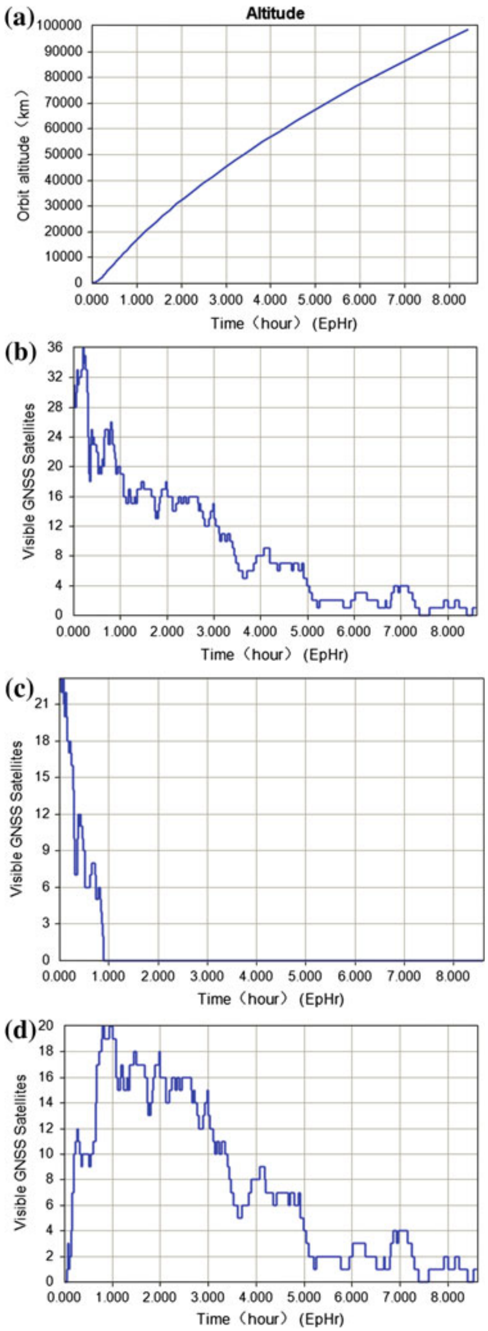
According to the simulation results, on the earth-moon-earth stage when the track height below 68,000 km, the number of GNSS satellites can be seen more than four, the up-side receiver will not be able to receive GNSS signal above 18,000 km. Signal Doppler is large when in the low height, close to ±50 kHz, decreases as the altitude increasing; as for the duration of the signal, the GPS signal has close to the maximum 2 h duration, some of the GEO satellite for BDS has nearly 4 h duration.

1.3 High Sensitivity GNSS Receiver Design

1.3.1 *High Sensitivity GNSS Receiver Scheme Design*

The sensitivity of conventional onboard GNSS receiver (see Fig. 1.11) for −160 dBW that can meet the application, when applied to the exploration of deep

Fig. 1.5 Visible GNSS satellites **a** orbital altitude **b** visible GNSS satellites **c** visible GNSS satellites to the up-side **d** visible GNSS satellites to the down-side



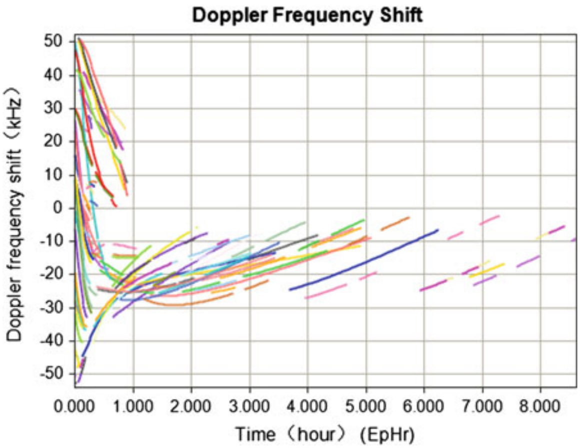


Fig. 1.6 Doppler shift range

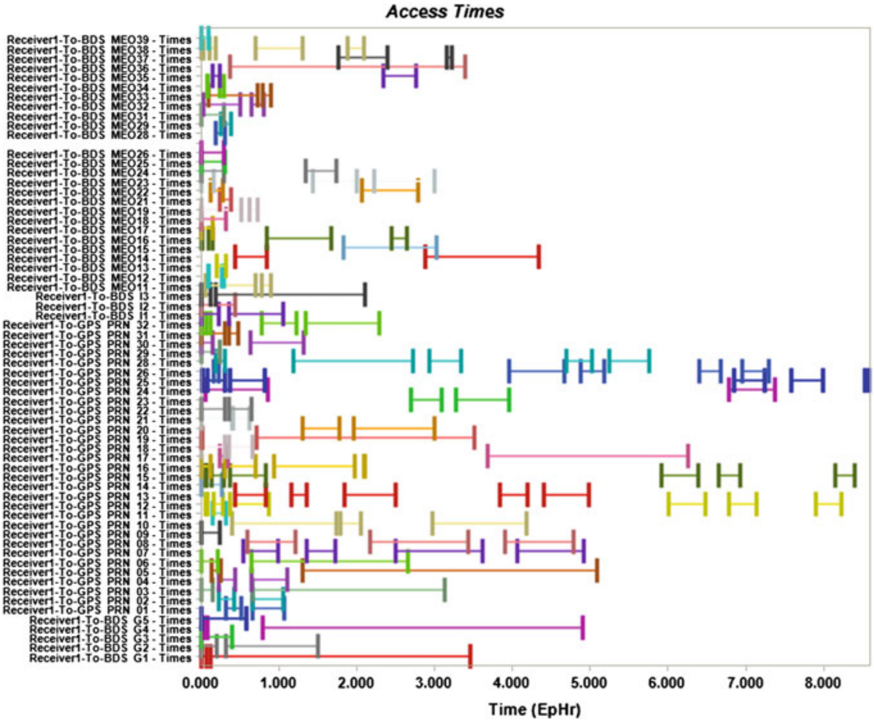


Fig. 1.7 GNSS signal visibility duration

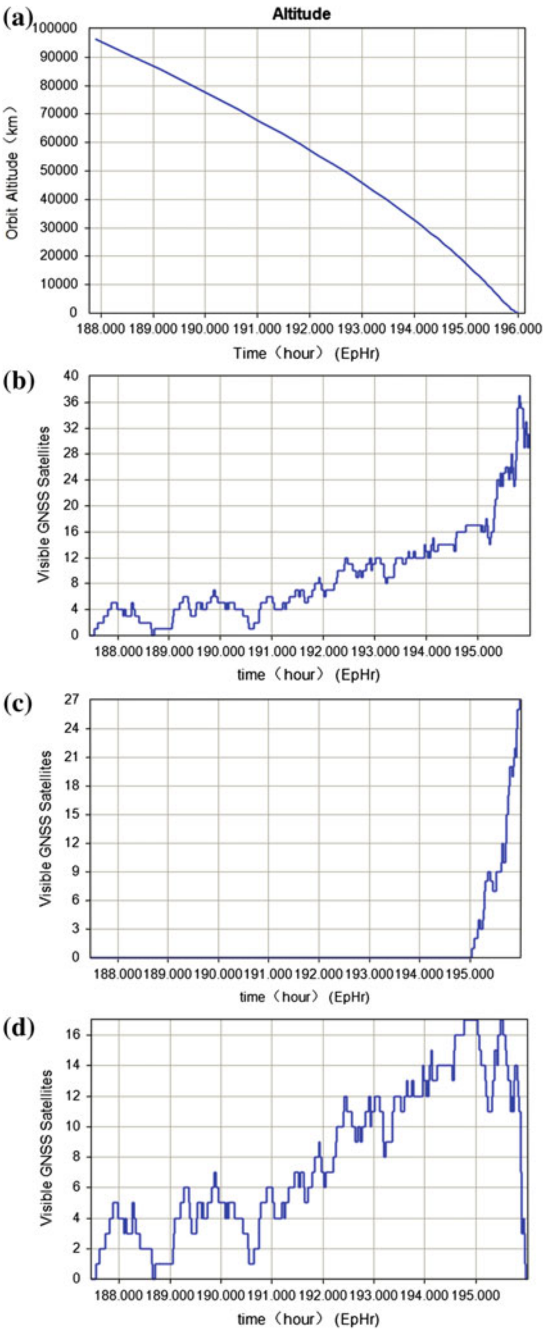


Fig. 1.8 Visible GNSS satellites **a** orbital altitude **b** visible GNSS satellites **c** visible GNSS satellites to the up-side **d** visible GNSS satellites to the down-side

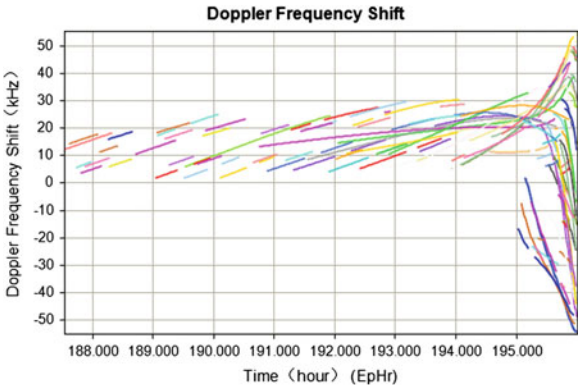


Fig. 1.9 Doppler shift range

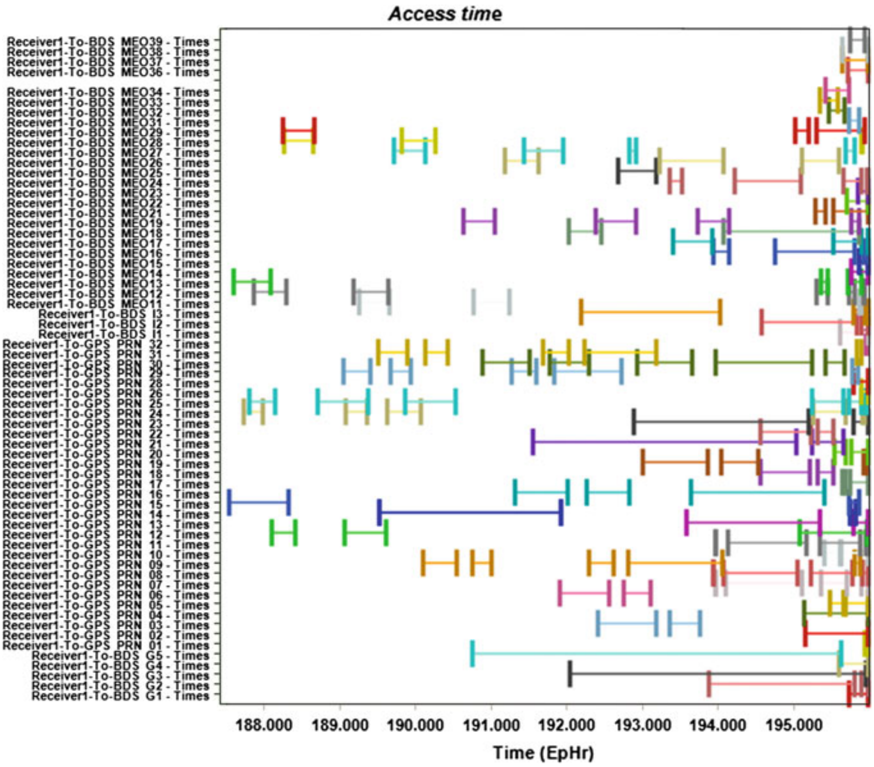


Fig. 1.10 GNSS signal visibility duration

space, the signal propagation path increases, need to receive the GNSS satellite side-lobe signal, the received signal power will become low, because the received power of −175 dBW than traditional signal lower above 15 dB, using traditional

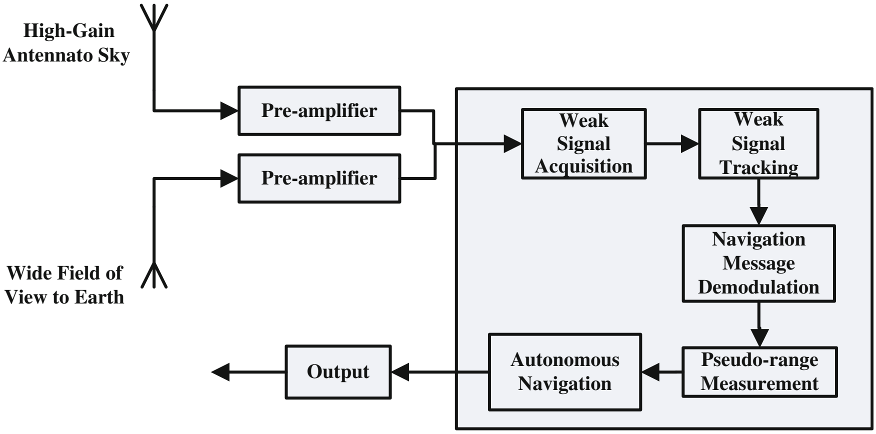


Fig. 1.11 Block diagram of the high sensitivity GNSS receivers

methods to capture the value of 1 ms data the signal peak detection will be difficult, result in the failure of demodulating navigation message. Also because the signal submerged in noise in the tracking process, resulting in the tracking loop cannot reach the pseudo range measurement accuracy of less than 10 m. So we need to solve the problem of high sensitivity receiver design.

The basic idea to solve these problems is the accumulation of the energy signal, through the accumulation of multiple 1 ms correlation values (coherent, non-coherent accumulation) summation, bring the SNR improvement. The GPS signal can be accumulated by the energy signal, but the BDS B1I with NH code, will result in greater different with GPS L1 C/A. The capture process, due to the unknown NH code phase, the coherent length is limited; if only rely on the gain of non-coherent processing to enhance the cost will be huge. In the tracking process, the same reason as the NH code and navigation data bit jump will result in not long pre-integration time tracking, with low precision of carrier phase and pseudo range measurement accuracy of GPS and BDS for code loop and carrier loop. Therefore, this paper use weak signal processing process as shown in Fig. 1.12, complete the acquisition and tracking weak signal of the BDS B1I and GPS L1 C/A.

The biggest difference between high sensitivity GNSS receiver and conventional baseband signal processing is during the acquisition phase and bit synchronization. For the acquisition of −175 dBW BDS and GPS weak signal, the acquisition process BDS and GPS signals are placed in different channels, BDS acquisition uses 20 ms data overlay code auto correlation method [5], this method can avoid the unknown NH code phase which causes the coherent length restriction; GPS signal can use the coherent accumulation with non-coherent accumulation, enhance the gain [6]. During the processing, bit synchronization is after acquisition and before tracking, with navigation information removed, it can be long time accumulations, to complete −175 dBW BDS/GPS weak signal tracking, and improve the measurement accuracy.

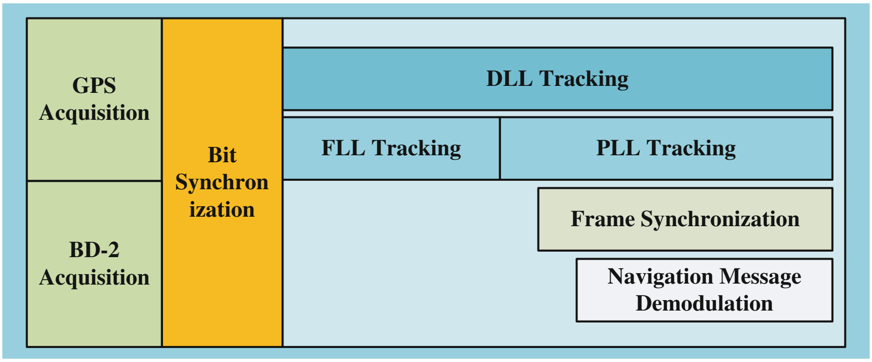


Fig. 1.12 High orbit BDS/GPS receiver weak signal processing

1.3.2 Weak Signal Acquisition Algorithm Design

1.3.2.1 Weak Signal Acquisition Processing

In this paper, the biggest difference between BDS and GPS weak signal acquisition schemelies in whether the need for NH code processing. BDS coherent accumulation is longer than the 1 ms time, should consider the removal effect of NH codes. After the rough acquisition, in order to get more accurate parameters of signal frequency, this also requires fine acquisition for the frequency.

BDS and GPS acquisition are used in parallel code phase search method based on FFT/IFFT. The difference lies in the BDS capture, the length of 10 ms signal are stored, and 0–20 ms after the FFT transform, then local 20 ms modulation of NH code data FFT transform, the conjugate and through the input data by FFT transform, and then IFFT transform. Then use the half bit method, followed by the odd sequence and a sequence of non-coherent accumulation, and finally through the peak detection, judge whether the signal acquisition is successfully accomplished. GPS is the 10 ms coherent accumulation of the 1 ms data, and eliminate the effect of data bit jump with half bit method, then the final peak detection will be done. In the scheme of BDS non-coherent accumulation while using the 20 ms data length, the actual loss due to zero by 3 dB, equivalent to the processing of 10 ms data.

1.3.2.2 Design Acquisition Processing Parameters

The power of the input signal −175 dBW, considering the frequency for loss, code without synchronization loss and pre-amplifier noise coefficient and A/D quantization loss and other factors, the power energy loss total close to 5 dB, it means that the normal signal relative to the −160 dBW, weak signal acquisition processing needs at least 20 dB gain in order to meet the requirement.

10 ms coherent provides 10 dB gain, non-coherent cumulative 20 provides 10.3 dB gain (false alarm probability 10^{-7}). So the −175 dBW ($C/N_0 = 29$ dB-Hz) of weak signals can be selected for the 10 ms coherent accumulation data of 20 non-coherent accumulation, then detect the signal peak value to judge the acquisition whether successful. The probability of false alarm is set to $P_{fa} = 1e^{-7}$, when the carrier to noise ratio is 29 dB-Hz, 20 times of non-coherent accumulation after 10 ms coherent accumulation, can achieve more than 99 % of the signal detection probability.

1.3.3 Weak Signal Tracking Algorithm Design

1.3.3.1 Weak Signal Tracking Processing

In this paper, the weak signal tracking using phase-locked loop (PLL), frequency locked loop (FLL) and delay locked loop (DLL). Take the accumulation data, according to the state transition corresponding function calls to achieve ranging code and carrier acquisition, tracking, bit synchronization and losing lock processing; according to the process design first complete the bit synchronization, and then complete 20 ms pre-integration time of the carrier and code tracking [7]·

In the low SNR case, in order to improve the tracking processing gain, by improving the pre-integration time, multi cycle energy accumulation, signal tracking processing technology needs to focus on the design of frequency, phase detector, loop filter.

1.3.3.2 Design Tracking Processing Parameters

In the FLL pre-integration time 10 ms, PLL and DLL pre-integration time 20 ms conditions, from Table 1.2 for loop tracking error analysis, the loop tracking error will be less than a threshold value, can guarantee the normal work of the loop. Therefore, to extend the pre-integration time can effectively solve the tracking problem under low SNR.

This paper uses the differential coherent accumulation method of bit synchronization in the process of acquisition transfer to tracking [8]; this method can

Table 1.2 Weak signal long pre-integration time tracking loop errors

Loop	Input power (dBW)	B_n (Hz)	T_{coh} (ms)	Theory errors	Simulation errors
PLL	−175	1	20	2.06°	2.36°
FLL	−175	1	10	1.19 Hz	1.59 Hz
DLL	−175	1(Chip interval in 0.5 chips)	20	7.79 m	8.46 m

eliminate the effect of initial frequency offset on the bit synchronization. The cumulative length of capture in 10 ms will cause the frequency search step in ±50 Hz, according to the analysis of 25 bit data length can be solved better bit synchronization, during this period, Doppler frequency offset to the code phase accumulation effect for the 0.017 chips, the error can be accepted as the code loop, so the bit synchronization before entering the loop tracking is feasible.

1.3.4 GNSS Orbit Determination Algorithm Design

High earth orbit satellite GNSS autonomous orbit determination algorithm using Kalman filtering method is based on dynamic orbit model, combined the orbit extrapolation without the use of visible satellite and Kalman filtering under positioning to carry out autonomous orbit determination. The main principle is to use dynamic orbit model to forecast the orbital elements, and then use the real-time data about the pseudo range and Doppler measurements to modify the orbital elements, in less than 4 visible satellites can filter, with orbit extrapolation ability.

Space bore GNSS autonomous orbit determination uses the GNSS measurement data with noise and gross error, Combined with the highly nonlinear dynamics model, without manual intervention, to obtain optimal estimation of satellite motion. Space borne GNSS autonomous orbit determination algorithm is to design a high precision, without manual intervention and long-term stable operation of the Kalman filter. In general, the model error mainly affected by three types of optimal estimation of satellite state: one is the error caused by nonlinear state model and observation model linearization; two is the orbit estimation process due to the calculation error caused by the word length of the computer and other restrictions; three is the satellite motion equation of perturbation force model selected does not possible and actual satellite mechanical model consistent, no model or error model is little perturbation to orbit estimation errors, that is to say, autonomous orbit determination filtering system state equation of noise does not satisfy the Gauss white noise characteristics, and belongs to the field of non-ferrous dynamic noise white noise driven.

Kalman filtering system of the satellite autonomous orbit determination software must consider the impact of these three kinds of model error, to design a stable and efficient filter. Three types of errors, the first kind of error can be extended with the Kalman filter (EKF) method to solve, in the filter state estimation, not used prenominal trajectory linearization, but a moment ago orbit around the estimated value of linearization, reduce the error caused by the linearization. The calculation error caused by computer word length, similar to U-D factorization to solve, not only avoid the state error covariance matrix for inverse matrix singular matrix operation possible situation, but also solve the problem of state estimation for satellite observation data in less than 4 GNSS satellites, enhance the autonomous orbit determination ability. For the third types of errors, need to use particular stochastic process to describe the perturbation without models or error model, which is the dynamic model compensation (DMC).

Space borne GNSS autonomous orbit determination Kalman filtering algorithm is to take into account the influence of the above three errors, Kalman filter design of a high precision, high stability, can use less than the observed data of 4 GNSS satellites for autonomous orbit determination. Simulation results are illustrated in Fig. 1.13.

For the Earth-Moon transfer orbit stage simulation of the autonomous orbit determination algorithm, below 60,000 km due to the ideal number of visible satellites, the position error can be guaranteed within 100 m. But with the orbit height increases, fewer visible satellites, position accuracy will become worse. Follow up for GNSS receiver in deep space applications visible satellites number less than 4, combining other navigation systems to improve the navigation accuracy for deep space exploration will be further research.

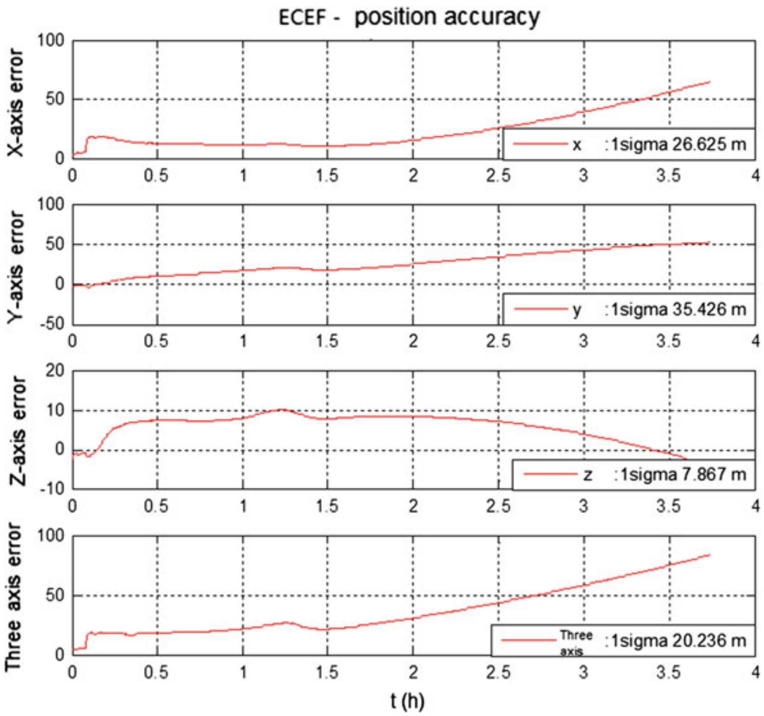


Fig. 1.13 Autonomous orbit determination of Kalman filter algorithm in the Earth–Moon transfer orbit simulation of position accuracy

1.4 Conclusion

According to the Third Stage of China Lunar Exploration Program mission requirements, the feasibility of GNSS under high below 100,000 km in the moon transfer orbit process was analyzed. The available number of stars under GPS + BDS constellation in the orbit transfer process, the duration and the Doppler characteristics of GNSS signal were obtained by STK simulation. A high-sensitivity GNSS receiver design method was proposed, in which the capture and tracking of BDS and GPS weak signal was completed respectively through the relevant processing method. Finally, according to the characteristics of the task, Kalman filtering method is employed to track dynamics model of orbit determination algorithm for navigation service.

Through the analysis, with the usage of high-sensitivity GNSS receiver in the Third Stage of China Lunar Exploration Program mission, a navigation service with the position accuracy better than 100 m will be provided for the lunar spacecraft, which is in great application demands and good prospects.

References

1. Winternitz LMB, Bamford WA, Heckler GW (2009) A gps receiver for high-altitude satellite navigation. IEEE J Sel Top Sig Process 3(4):541–556
2. Palmerini GB, Sabatini M, Perrottac G (2009) En route to the moon using GNSS signals. Acta Astronaut 64:467–483
3. Yang WL (2010) Phasing orbit design for Chinese lunar satellite CE-1. Chin Space Sci Technol 1:18–24
4. Beidou satellite navigation system—document of interface control for space signal (public service signal B1I 1.0 version.) http://www.beidou.gov.cn//2012/12/27. Cited 27 Dec 2012
5. Xu DY, Kou YH (2011) The dual component combined tracking of GPS L5 signal. In: Special journal for the second China satellite navigation conference. pp 182–190
6. Psiaki ML (2001) Block acquisition of weak GPS signals in a software receiver. In: Proceedings of Institute of Navigation GPS. pp 2838–2850
7. Wen CY, Fu ZY, Qiu YH et al (2013) High orbit environment BDS weak signal tracking technology research. J Spacecraft TT&C Technol 32(6):43–51
8. Li XS, Guo W, Xie XB (2011) A high dynamic, weak signal in GPS bit synchronization method. J Electron Inf Technol 33(10):87–95

Chapter 2
Chang'E-3 DOR Signal Processing and Analysis

Ke Xu, Gongyou Wu, Kun Jiang, Yiwen Jiao and Xin Lian

Abstract This paper researches the local correlation algorithm which is aimed at the characteristics of spacecraft DOR signal. The raw data is preprocessed to acquire the Doppler dynamic information of the received signal firstly. Then a model signal is generated which has the same frequency of the transmitted signal locally, and the model signal is utilized to do correlation processing with the raw data. Because of the error of initial delay model and estimated transmitted frequency, we propose to utilize the correctness of the frequency and phase based on polynomial fitting to improve the local correlation algorithm and adopt iteration to correct the model signal to solve the problems above. Finally, the differential phases of DOR signals are processed with Bandwidth Synthesis to calculate the DOR delay. This method has been utilized in Chang'E-3 interferometry test data processing, the obtained residual delay is relatively stable and verifies the correctness of the improved method.

Keywords DOR signal · Local correlation · Doppler dynamic · Polynomial fitting

2.1 Introduction

Chang'E-3 (CE-3) mission has utilized various new type TT&C techniques including Delta Differential One-Way Ranging (ΔDOR). ΔDOR is a new type TT&C technique based on conventional interferometry measurement technique developed by NASA's JPL [1, 2], which is different with conventional interferometry measurement technique in the definition of observable, the signal characteristics of spacecraft and the signal correlation processing method [3].

K. Xu (✉) · G. Wu
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: xk361@163.com

K. Jiang
China Xichang Satellite Launch Centre, Xichang 615000, China

Y. Jiao · X. Lian
Academy of Equipment, Beijing 101416, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_2

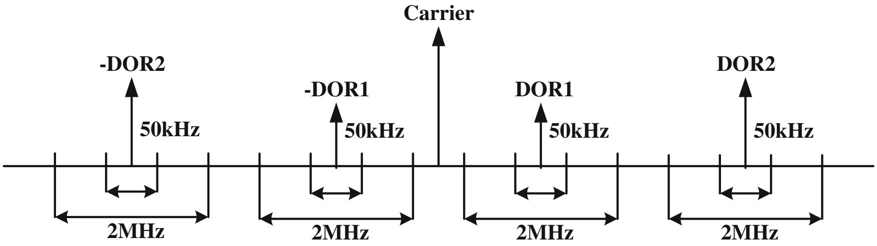


Fig. 2.1 The beacon structure schematic of DOR signal

Spacecraft DOR tone is point-frequency signal, its spectral characteristics is quite different with the radio source signal. The beacon structure schematic of spacecraft DOR tone is shown in Fig. 2.1 [4].

Multiple DOR signals with different bandwidths are generated across the spacecraft in the vicinity of the main carrier (each DOR signal generates two harmonics, such as DOR1 and –DOR1). The radio source signal and the spacecraft DOR signal are recorded in the centre of the tones in multiple frequency channels. For the spacecraft signal, the output bandwidth is 50 kHz, and the output bandwidth is 2 MHz for the radio source signal. The wider spanned bandwidth DOR signals are utilized to calculate the group delay, and make difference with the group delay generated by the radio source signal which has the same centre frequency, and obtain the ΔDOR measurement.

The conventional interferometry data processing algorithm for each observation station is making integer bit delay compensation, fringe stop, Fourier transform, fraction bit delay compensation and integral and so on [5, 6]. This method is quite practical for radio source broadband continuous spectrum signal. The spacecraft DOR signals are point-frequency signals, which have a few MHz interval with each other. Processing the DOR signals with conventional FX correlation processing method requires a very high spectral resolution to achieve the sufficient SNR and accuracy, and most of output data are noise data, so the accuracy is relative low [7].

This paper studies the DOR signal processing method, proposes to utilize the polynomial fitting method to improve the original algorithm, and processes the CE-3 test data and obtains an ideal result in the end.

2.2 DOR Signal Processing Method

2.2.1 Local Correlation Algorithm

The basic idea of local correlation algorithm is to build a none-noise model signal locally which has the same frequency of spacecraft DOR tone according to the relative motion between the spacecraft and the station. The model signal and recorded signal are made cross-correlation to eliminate the high dynamic variation of the DOR frequency and phase on account of the relative motion. The phases of corresponding channels are made difference after cross-correlation to calculate the final DOR delay observable by utilizing Bandwidth Synthesis.

References [7, 8] have made a complete description of the DOR signal local correlation algorithm.

Assume the one-way light time from the spacecraft signal to the Station 1 is p_1, which is p_2 to the Station 2. So the relation between the transmission-time t_1 and the reception-time t is:

$$t_1 = t - p_1 \tag{2.1}$$

$$t_2 = t - p_2 \tag{2.2}$$

Assume the spacecraft DOR signal expression at transmission-time is:

$$s_i(t) = e^{j(2\pi f_i t + \Phi_{oi})} \tag{2.3}$$

f_i is the DOR signal frequency, N is the number of DOR signals. Φ_{oi} is the original phase of DOR signals, so the received signals to the two stations are expressed respectively:

$$s_{rec1}(t) = e^{j[2\pi f_i (t-p_1) + \Phi_{o1}]} \tag{2.4}$$

$$s_{rec2}(t) = e^{j[2\pi f_i (t-p_2) + \Phi_{o2}]} \tag{2.5}$$

The spacecraft transmits a radio frequency signal, but the stations receive a baseband signal after down-conversion, add the equipment phase delay and the clock delay, the two stations recorded signals are expressed as:

$$s_{rec1-f_i}(t) = e^{j[2\pi(f_i - f_{0i})t - 2\pi f_i p_1 + \Phi_{o1} + \Phi_{1i}]} \tag{2.6}$$

$$s_{rec2-f_i}(t) = e^{j[2\pi(f_i - f_{0i})(t-\Delta\tau_c) - 2\pi f_i p_2 + \Phi_{o2} + \Phi_{2i}]} \tag{2.7}$$

f_{0i} is the sky-frequency, $\Delta\tau_c$ is the clock error of two stations. Φ_{1i} and Φ_{2i} are the signal phase delays through the two stations.

The none-noise local signal models of the two stations are expressed as:

$$s_{\text{mod}1-f_i}(t) = e^{j[2\pi(f_i^m - f_{0i})t - 2\pi f_i^m p_1^m]} \tag{2.8}$$

$$s_{\text{mod}2-f_i}(t) = e^{j[2\pi(f_i^m - f_{0i})(t-\Delta\tau_c) - 2\pi f_i^m p_2^m]} \tag{2.9}$$

f_i^m is the estimated transmitted frequency of spacecraft signal. p_1^m and p_2^m are the delay models of the received signal.

The DOR signal real frequency of Station 1 can be obtained from the test data, then obtain the estimated transmitted frequency f_i^m of spacecraft signal.

The playback data of the station is made cross-correlation with the local none-noise model signal. The two point-frequency signal correlation phase of Station 1 is expressed as:

$$\begin{aligned}\varphi_{cor1}^{i} &= 2\pi(f_i - f_{0i})t - 2\pi f_i p_1 + \varphi_{o1} + \varphi_{1i} - [2\pi(f_i^m - f_{0i})t - 2\pi f_i p_1^m] \\ &= 2\pi(f_i - f_i^m)t - 2\pi f_i(p_1 - p_1^m) + \varphi_{o1} + \varphi_{1i}\end{aligned} \tag{2.10}$$

Without considering the impact of the initial phase noise introduced by phase noise and instrument, so:

$$\varphi_{cor1}^{i} = 2\pi(f_i - f_i^m)t - 2\pi f_i(p_1 - p_1^m) \tag{2.11}$$

Similarly, the correlation phase of the two point-frequency of Station 2 is expressed as:

$$\varphi_{cor2}^{i} = 2\pi(f_i - f_i^m)t - 2\pi f_i(p_2 - p_2^m) \tag{2.12}$$

The differential phase of the two stations corresponding to the transmitted frequency f_1 is expressed as:

$$\varphi_{dif}^{f_1} = 2\pi f_1(p_2 - p_1) + 2\pi f_1(p_1^m - p_2^m) \tag{2.13}$$

The differential phase of the two stations corresponding to the transmitted frequency f_2 is expressed as:

$$\varphi_{dif}^{f_2} = 2\pi f_2(p_2 - p_1) + 2\pi f_2(p_1^m - p_2^m) \tag{2.14}$$

(2.13) and (2.14) make difference to obtain the DOR local correlation model:

$$p_2 - p_1 = \frac{\varphi_{dif}^{f_1} - \varphi_{dif}^{f_2}}{2\pi(f_1 - f_2)} - (p_1^m - p_2^m) \tag{2.15}$$

2.2.2 *Frequency and Phase Correction Based on Polynomial Fitting*

During the data processing, the residual frequency has much Doppler influence because of the error between the delay model and the real propagation delay and the error between the estimated transmitted frequency and the real transmitted frequency. To obtain the final DOR delay must correct the frequency and phase of the dynamic variation.

The correlation residual frequency of real signal and local signal is $f_{res}(t)$, residual phase is $\phi_{res}(t)$, the $f_{res}(t)$ in N order polynomial fitting can be expressed as:

$$f_{res}(t) = \sum_{i=0}^{N} a_i \cdot t^i \tag{2.16}$$

So the residual phase is:

$$\phi_{res}(t) = 2\pi \cdot \int f_{res}(t)dt = 2\pi \cdot \int \sum_{i=0}^{N} a_i \cdot t^i dt = 2\pi \cdot (\sum_{i=0}^{N} \frac{a_i \cdot t^{i+1}}{i+1} + \phi_0) \tag{2.17}$$

ϕ_0 is a constant.

The local model signal is corrected based on polynomial fitting of residual phase, there are two steps:

1. Single station residual phase correction term is obtained via the polynomial fitting of the residual frequency, and the phase of local signal is corrected to let the residual frequency in several Hz after the correlation.
2. Based on Step (1), the raw data is made cross-correlation with the local signal, the phase of the tone is extracted, the phase variation is made polynomial fitting, and the phase model of Station 2 is corrected to make the residual frequency equal to the two stations.

We can eliminate the single-station residual frequency dynamic variation and the difference of the corresponding channel between the two stations through the steps above, and obtain the final results with using Bandwidth Synthesis.

2.3 Test Data Processing and Analysis

The Station A and B test data of CE3 mission is processed. The test data is in 6 channels, quantized in 8bit and the bandwidth is 200 kHz. The duration is 300 s. The main carrier signal is in 3rd channel, the DOR signals are in the 1st, 2nd, 5th, 6th, channels, which have the intervals to the main carrier signal are −19.2, −3.85, 3.85 and 19.2 MHz respectively. The 4th channel is the ranging tone which has 500 kHz interval to the main carrier. The frequency spectrograms are shown in Fig. 2.2.

The clear tone is the DOR signal.

The raw data and the local model signal are made cross-correlation to obtain the DOR signal residual frequency variation expressed in Fig. 2.3.

During 300 s, the residual frequency of the carrier signal of Station A varies from −8,112 to −19.796 kHz, the same of Station B varies from −8,108 Hz to

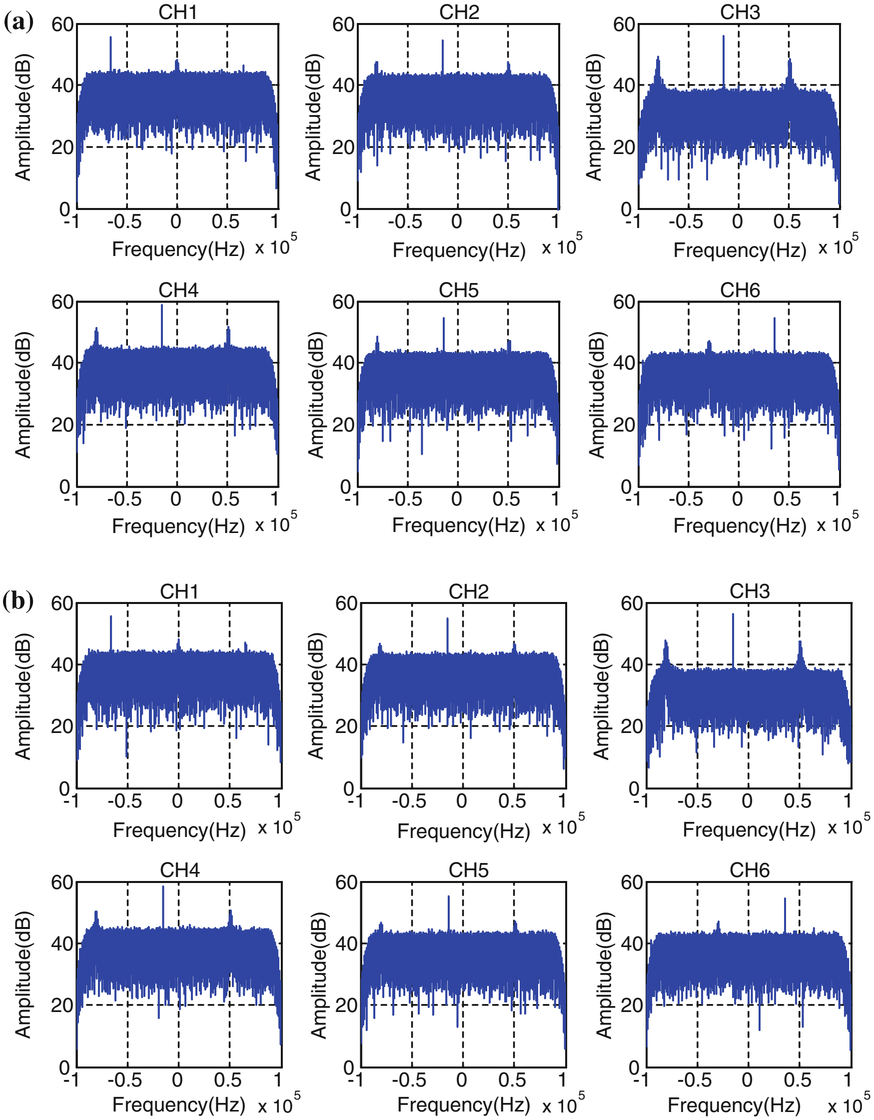


Fig. 2.2 Frequency spectrums of two stations. **a** Frequency spectrum of station A. **b** Frequency spectrum of station B

−19.792 kHz, which illustrates the local signal model is not accurate enough and the dynamic variation of the residual frequency is high.

The residual frequency is made polynomial fitting to correct the residual phase of the local signal model, the corrected residual frequency of each channel is expressed in Fig. 2.4.

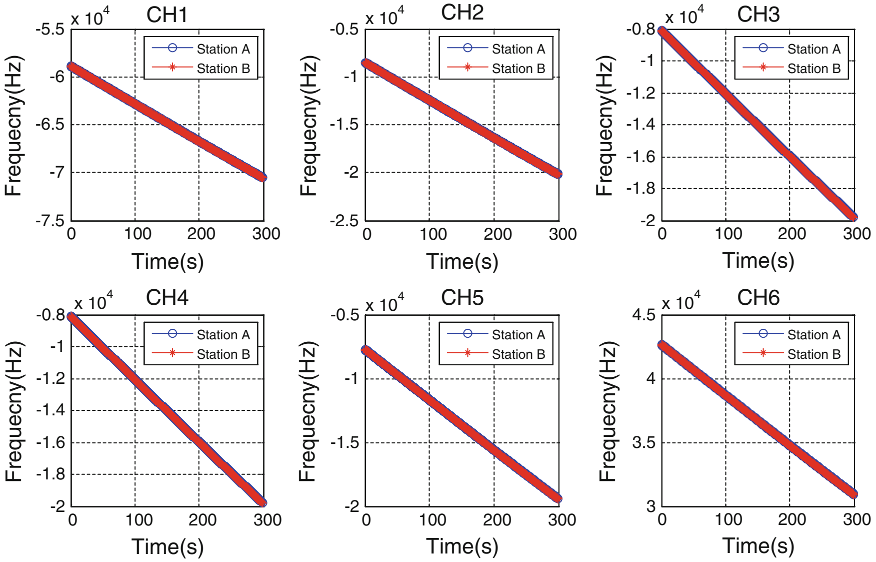


Fig. 2.3 Residual frequency dynamic variation of two stations

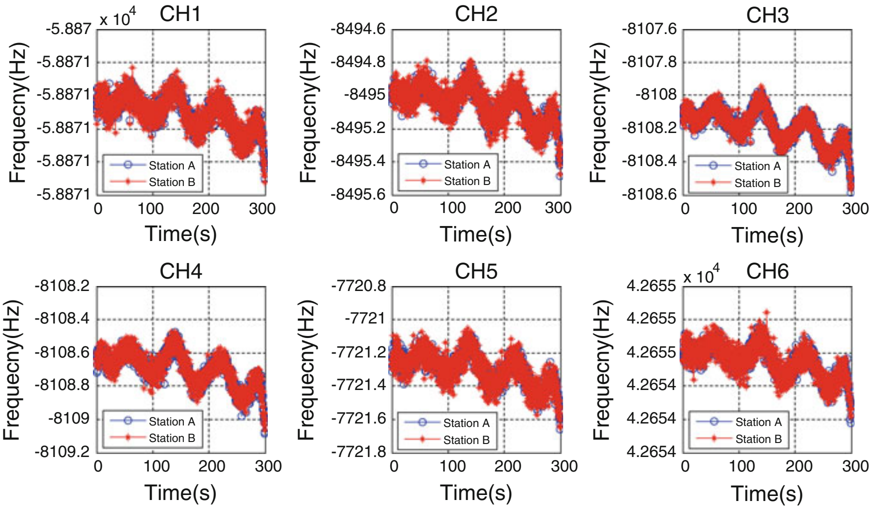


Fig. 2.4 Residual frequency dynamic variation of two stations via polynomial fitting correction

The residual frequency variation decreases obviously in 300 s, the maximum variation is about 0.69 Hz. The tiny fluctuation may be caused by the instability of the spaceborne oscillator. The variations of Station A and B are similar, this is consistent with the expected result. The differential residual phase of corresponding channel is shown in Fig. 2.5.

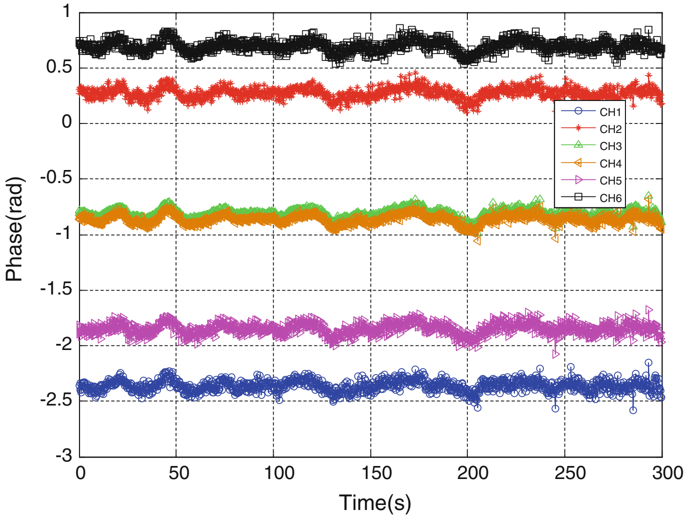


Fig. 2.5 Differential residual phase of two stations

The differential residual phase variation tendency of each channel is steady and similar, the differential residual phase of 20 s integral is shown in Fig. 2.6.

The variation tendency of differential residual phase decreases via integral. The residual delay is obtained by Bandwidth Synthesis which is shown in Fig. 2.7.

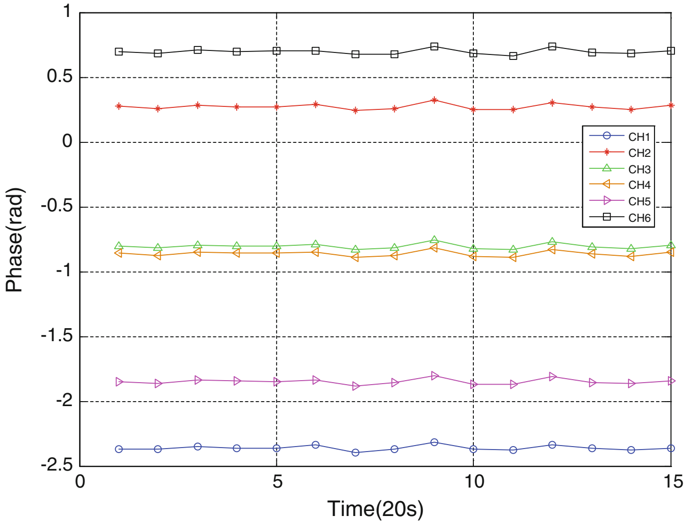


Fig. 2.6 Differential residual phase via 20 s integral

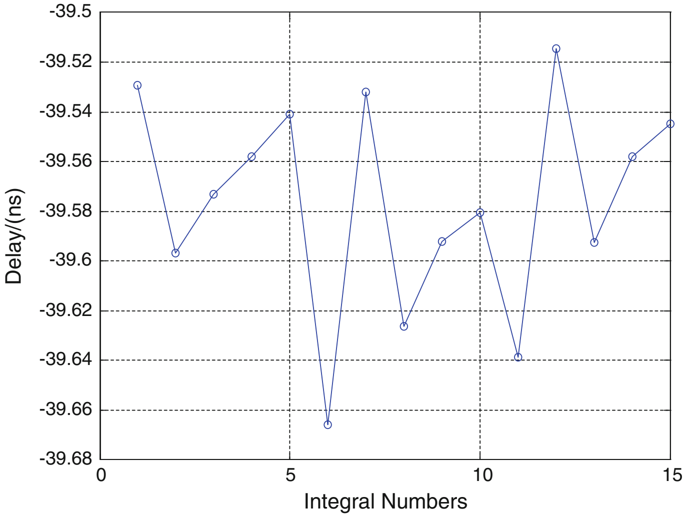


Fig. 2.7 Residual delay via bandwidth synthesis

The average residual delay is −39.576 ns, the mean square error is 43.3 ps. The stability and precision of the residual delay verify the correctness of the improved algorithm.

2.4 Conclusion

This paper studies the local correlation algorithm. Polynomial fitting method is utilized to improve the algorithm. The CE-3 test data is processed and obtain an ideal result. We can compare the result with the spacecraft subsequent precise obit for further analysis.

References

1. James SB, John AK (1993) Technical characteristics and accuracy capabilities of delta differential one-way ranging (δdor) as a spacecraft navigation tool. In: Report of the proceedings of the RF and modulation subpanel 1E meeting. German Space Operations Center, Munich, pp 1–2
2. Delta-differential one way ranging (Delta-DOR) operations (2011) Recommendation for space data system standards. CCSDS 506.0-M-2. Magenta Book. Washington, D.C.
3. Huang YD, Zheng WM, Ma ML (2013) Study on DOR delay calculating. In: The 9th China deep space exploration conference, pp 672–673
4. James N, Ricard A, Marco L et al (2009) Implementation of an ESA delta-DOR capability. Acta Astronaut 64:1041–1042

5. Li YF (2004) Study on VLBI data software correlation processing method. Doctoral thesis of Shanghai astronomical observatory, Chinese Academy of Sciences, Shanghai, pp 21–37
6. Deller AT, Tingay SJ, Bailes M et al (2007) DiFX: a software correlator for very long baseline interferometry using multiprocessor computing environment. Publ Astron Soc Pac 119:318–336
7. Shu FC, Ping JS, Zheng WM et al (2010) Local correlation model of DOR tones for tracking deep space probes. In: The 7th China deep space exploration conference, pp 79–83
8. Chen L, Tang GS, Han ST et al (2012) Modeling and simulation of local correlation processing of spacecraft DOR signals. J Spacecraft TT&C Technol 31(4):15–17

Chapter 3
Research on Technique of VLBI Data Correlation Post Processing in Deep Space Navigation

Jia Wang, Shilu Shen, Gongyou Wu, Dong Zhang and Ke Xu

Abstract Correlation post handling includes accomplishing error corrections and outputting delay and delay rate in VLBI data processing. For multicenter VLBI observation data, there is interchannel delay which should be calibrated with quasar phase, which is phase calibration. Correlation post handling makes bandwidth synthesis to phase after calibration and gets high precision delay consequently. We focus on interchannel data in data processing of Chang'E-3 Project, adopt difference methods to make correlation post handling and simulating analysis. The results show correctness and validity of method which is put forward in this paper.

Keywords Correlation post processing · Bandwidth synthesis · Phase delay · Group delay · Channel delay · Phase calibration

3.1 Introduction

Deep space exploration is a new domain in space technique after satellite application and manned space. Deep space exploration, especially lunar exploration has higher request in precision and real time performance of orbit measurement, which guarantee lunar Lander landed in destining position accurately and return equipment returned earth in scheduled orbit. Mastery of high precision measurement of deep space explorer orbit will support and promote the advancement of high tech and quicken up the development of deep space exploration science in our country [1].

Very Long Baseline Interferometry (VLBI) is one of a finite list of methods which can be depended in deep space exploration. VLBI has been widely used in fields of astrophysical radio astronomy and mother earth survey since its coming in 1965. Furthermore, VLBI has been widely used in tracking and navigation of satellite, lunar explorer, even interplanetary explorer from 1970s because of its high precision ability of measuring angle [2, 3]. We have launched Chang'E-1, Chang'E-2 and

---

J. Wang (✉) · S. Shen · G. Wu · D. Zhang · K. Xu
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: wj_ssl@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_3

Chang’E-3 exploration moon satellites in succession in our country since 2007. In order to make high precision tracking and navigation for exploration moon satellite, Chinese Academy of Sciences VLBI measuring system has been orbit measuring subsystem of Chang’E-1 TT&C system and joined in orbit measuring task [4].

It is to accomplish error correction and get correlation phase and delay in VLBI data correlation post process [5]. We make comparative analysis between two different handling methods in correlation post process of Chang’E-3 task.

3.2 Questions and Phenomenon Analysis

In correlation post process of Chang’E-3 task, for multicenter datum, we should make use of bandwidth synthesis [6] to get group delay. Due to the accretion of frequency bandwidth, bandwidth synthesis group delay measure errors are less than single channel group delay ones. Furthermore, we should carry out phase calibration between different channels and bandwidth synthesis.

In Ref. [7], Chen Guan lei et al. introduced method of calculating bandwidth synthesis group delay which named ‘calibration method’ for short hereinafter particularly. Firstly, it needs channel initial phase and single channel initial delay of quasar to calibrate satellite’s initial channel phase, which called interchannel delay correction. Then, it makes use of satellite’s single channel initial group delay and each channel phase after correction to put up bandwidth synthesis and get bandwidth synthesis group delay finally.

We come up with a method to calculating bandwidth synthesis group delay aiming at initial channel phase in Chang’E-3 task, which called ‘subtraction method’ for short hereinafter. Firstly, we use satellite’s and quasar’s initial group delay and each channel phase without interchannel delay correction for bandwidth synthesis directly and get satellite’s and quasar’s bandwidth synthesis group delay with deviation. The difference between satellite’s and quasar’s bandwidth synthesis group delay is bandwidth synthesis group delay after correction.

Aiming at calibration method and calibration method, we make calculating and analysis to get result, that is two methods result delay difference is 10^{-16}–10^{-14} ns, which show subtraction method and calibration method are equivalent.

3.3 Principle of Correlation Post Process

Generally, each channel phase can be expressed in Eq. (3.1)

$$\varphi_i = 2\pi f_i(\tau_i + \tau_{err}) + 2\pi N_i + \varphi_{0i} \quad i = 1, 2\ldots \tag{3.1}$$

There into, f_i is signal frequency in corresponding point in channel i, τ_i is signal residual phase delay in channel i, τ_{err} is delay which caused by troposphere, ionosphere or observation instruments, N_i is phase integer ambiguity in channel i, φ_{0i} is initial phase which caused by receiver in channel i.

3.3.1 Principle of Bandwidth Synthesis Method

Bandwidth synthesis method [8] is to make use of carrier wave and DOR side tones phases and their corresponding frequencies to calculate phases after eliminating ambiguity according to special steps and make linear fitting phases to get bandwidth synthesis delay. Figure 3.1 shows carrier wave, DOR side tones phases and corresponding frequencies.

Firstly, we make use of φ_{obsC}, φ_{obs1} and τ_{ini} to get predicted phase value φ_{pre1} and correction phase value after eliminating ambiguity φ_{corr1} at the frequency point f_1 and group delay $\tau_{groupC1}$ between f_C and f_1, which shows in Eqs. (3.2)–(3.4). INT means taking the whole.

$$\varphi_{pre1} = \varphi_{obsC} + 2\pi(f_1 - f_C) \cdot \tau_{ini} \tag{3.2}$$

$$\varphi_{corr1} = \varphi_{obs1} - \mathrm{INT}\big((\varphi_{obs1} - \varphi_{pre1})/2\pi\big) \cdot 2\pi \tag{3.3}$$

$$\tau_{groupC1} = (\varphi_{corr1} - \varphi_{obsC})/(2\pi \cdot (f_1 - f_C)) \tag{3.4}$$

Secondly, we make use of φ_{corr1}, φ_{obs2} and $\tau_{groupC1}$ to get predicted phase value φ_{pre2} and correction phase value after eliminating ambiguity φ_{corr2} at the frequency point f_2 and group delay $\tau_{group12}$ between f_1 and f_2, which shows in Eqs. (3.5)–(3.7).

$$\varphi_{pre2} = \varphi_{corr1} + 2\pi(f_2 - f_1) \cdot \tau_{groupC1} \tag{3.5}$$

$$\varphi_{corr2} = \varphi_{obs2} - \mathrm{INT}\big((\varphi_{obs2} - \varphi_{pre2})/2\pi\big) \cdot 2\pi \tag{3.6}$$

$$\tau_{group12} = (\varphi_{corr2} - \varphi_{corr1})/(2\pi \cdot (f_2 - f_1)) \tag{3.7}$$

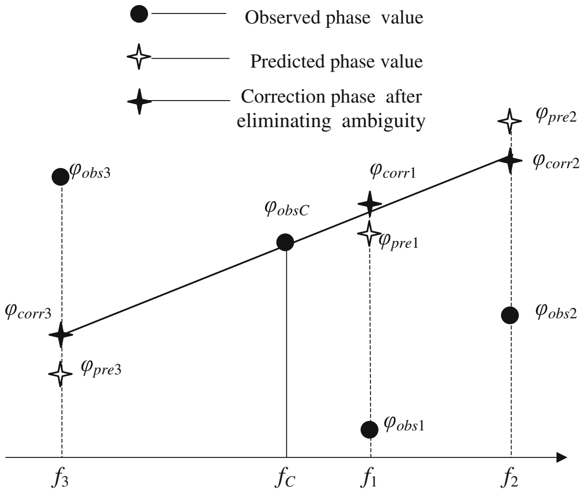


Fig. 3.1 DOR signal bandwidth synthesis handling sketch map

Similarly, we make use of φ_{corr2}, φ_{obs3} and $\tau_{group12}$ to get predicted phase value φ_{pre3} and correction phase value after eliminating ambiguity φ_{corr3} at the frequency point f_3, which shows in Eqs. (3.8)–(3.9).

$$\varphi_{pre3} = \varphi_{corr2} + 2\pi(f_3 - f_2) \cdot \tau_{group12} \tag{3.8}$$

$$\varphi_{corr3} = \varphi_{obs3} - \mathrm{INT}\left(\left(\varphi_{obs3} - \varphi_{pre3}\right)/2\pi\right) \cdot 2\pi \tag{3.9}$$

Finally, we make use of φ_{obsC} and correction phase value after eliminating ambiguity φ_{corr1}, φ_{corr2}, φ_{corr3} at the frequency point f_1, f_2 and f_3 to get bandwidth synthesis group delay adopting least square fitting method. The flow is shown in Fig. 3.2.

3.3.2 Principle of Calibration Method

Via the measurement of calibration quasar, delay in measurement mainly reflects the interchannel difference. So we make use of calibration quasar signal to calibrate system delay and initial phase in incepting channel. Furthermore, we get correction phase φ_{obsC}, φ_{obs1}, φ_{obs2} and φ_{obs3} and calculate satellite bandwidth synthesis group delay τ_{group} by using Eqs. (3.2)–(3.9).

3.3.3 Principle of Subtraction Method

By using Eqs. (3.2)–(3.9), we get calibration quasar bandwidth synthesis group delay τ_{groupQ} and make use of satellite's phase with initial phase to get bandwidth synthesis group delay $\tau_{groupIni}$ directly. Then, we use calibration quasar bandwidth

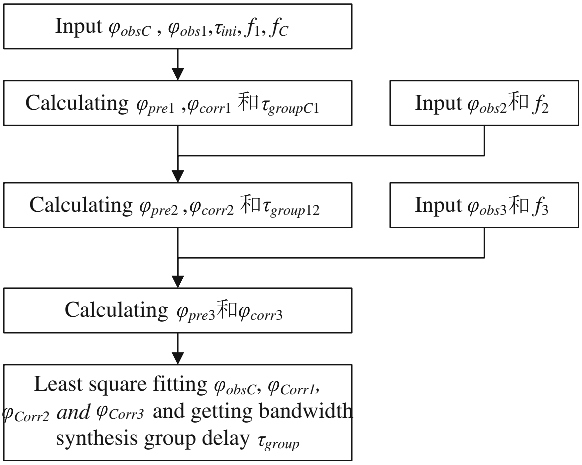


Fig. 3.2 The handling flow of DOR signal bandwidth synthesis

synthesis group delay τ_{groupQ} to calibrate $\tau_{groupIni}$ which with initial phase. Finally, we get satellite bandwidth synthesis group delay after correction, which is τ_{group}.

3.4 Calculating Example and Analysis

Using Chang'E-3 task observation data of station Tianma and Kunming of Shanghai astronomical observatory (SHAO) in Dec. 8th, 2013, we calculate and analyze calibration method and subtraction method.

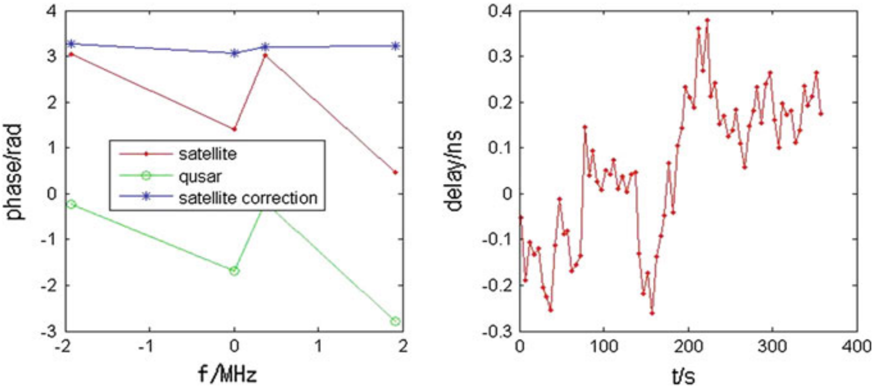


Fig. 3.3 Each channel phase of satellite, quasar and satellite correction with quasar calibration and satellite bandwidth synthesis group delay using calibration method

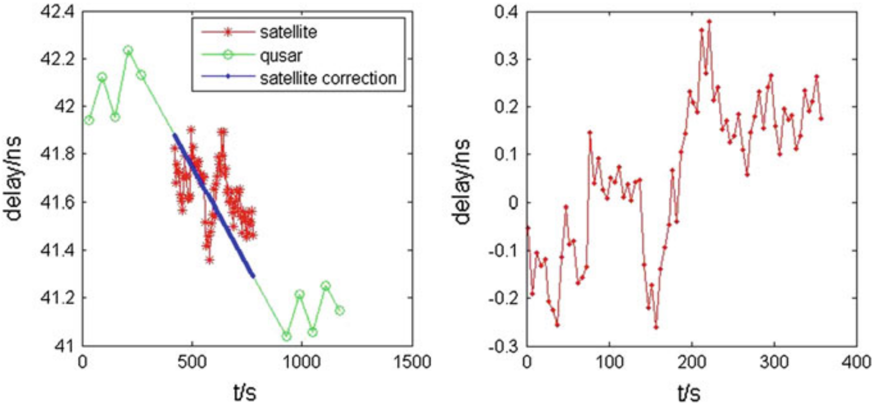


Fig. 3.4 Satellite, quasar and satellite after calibration bandwidth synthesis group delay using subtraction method

Table 3.1 Results and differences between calibration method and subtraction method

Time/s	Satellite bandwidth synthesis group delay/ns	Quasar bandwidth synthesis group delay/ns	Calibration method results/ns	Subtraction method results/ns	Differences/ns
2.480934	41.82543	41.87756	−0.052131849812227	−0.052131849812232	−4.5E−15
7.478055	41.6793	41.86929	−0.189992173390230	−0.189992173390237	−7.0E−15
12.47518	41.75566	41.86102	−0.105358863100179	−0.105358863100190	−1.1E−14
17.4723	41.72086	41.85274	−0.131888537045327	−0.131888537045334	−7.0E−15
22.46942	41.72601	41.84447	−0.118464199540007	−0.118464199540013	−6.0E−15
27.46654	41.63046	41.8362	−0.205734551370391	−0.205734551370393	−2.0E−15
32.46366	41.60353	41.82793	−0.224391711271167	−0.224391711271168	−1.0E−15
37.46078	41.56476	41.81965	−0.254893048391182	−0.254893048391191	−9.0E−15
42.4579	41.6981	41.81138	−0.113281257274238	−0.113281257274239	−1.0E−15
47.45502	41.79225	41.80311	−0.010852031225191	−0.010852031225192	−3.0E−16
52.45215	41.70711	41.79483	−0.087720347524446	−0.087720347524454	−8.1E−15
57.44927	41.70588	41.78656	−0.080682498241153	−0.080682498241161	−8.5E−15
62.44639	41.60988	41.77829	−0.168407377949130	−0.168407377949140	−1.0E−14
67.44351	41.61432	41.77002	−0.155691568805295	−0.155691568805294	1.0E−15
72.44063	41.62771	41.76174	−0.134034788520161	−0.134034788520162	−1.0E−15
77.43775	41.8992	41.75347	0.145732941048500	0.145732941048491	−9.0E−15
82.43487	41.78413	41.7452	0.038933430970893	0.038933430970883	−9.8E−15
87.43199	41.82957	41.73692	0.092642752528732	0.092642752528725	−7.4E−15
92.42912	41.75592	41.72865	0.027272701155169	0.027272701155168	−1.9E−15
97.42624	41.72831	41.72038	0.007927457031521	0.007927457031525	4.4E−15

Firstly, we calculate each channel phase of satellite, quasar and satellite correction with quasar calibration, which shown in Fig. 3.3. Bandwidth synthesis group delay which using calibration method or subtraction method is shown in Figs. 3.3 and 3.4 separately.

From Figs. 3.3, 3.4 and Table 3.1 we can see that bandwidth synthesis group delay difference between calibration method and calibration method is 10^{-16}–10^{-14} ns and can be neglected, which show that calibration method and subtraction method is equivalent.

Compare to calibration method, subtraction method use quasar and satellite bandwidth synthesis delay to subtract directly and needn't consider interchannel variance, which flow is so easy that can be realized in engineering easily.

3.5 Conclusions

Comparing and contrasting subtraction method with calibration method, we can educe that there is coherence and equivalence in results of two methods. Correlation post handling method which we put forward in the paper, that is subtraction method, needn't consider interchannel initial phase variance, which would be corrected with else in quasar calibration. Compare to calibration method, subtraction method has brief flow and steps and can be realized easier in VLBI engineering which is high sampling ratio and magnitude data size.

References

1. Ouyang ZY (2014) Scientific objective of Chinese lunar exploration and development strategy. Adv Earth Sci 19(3):351–358
2. Lichten S (2004) White paper on use of very long baseline interferometry for deep space tracking and navigation of NASA missions
3. Thornton CL, Border JS (2000) Radiometric tracking techniques for deep space navigation system center of excellence (DESCANSO). Jet Propulsion Laboratory, California Institute of Technology, Pasadena
4. Jiang DR, Hong XY (2008) VLBI for deep-space navigation. Science 60(1):10–14
5. Liu QH, Chen M, Xiong WM et al (2010) Relative position determination of a lunar rover using high-accuracy multi-frequency same-beam VLBI. Sci China Phys Mech Astron 53(3):571–578
6. Rogers AEE (1970) Very long baseline interferometry with large effective bandwidth for phase delay measurements radio science. Radio Sci 5(10):1239–1247
7. Chen GL, Zheng X, Chen M et al (2012) Method of calculating VLBI phase delay based on DOR signal of a satellite. Prog Astron 30(4):518–526
8. Chen M, Liu QH, Tang GS et al (2012) Method of calculating delay based on DOR signal in the Chang'E-2 project. Deep Space Explor 10(4):38–42

Chapter 4
Research on Simulation of Inter-satellite Link Based on the Navigating Constellation

Bo Jiang, Jing Li and Bo Guo

Abstract Establishing the inter-satellite link which connects the navigating constellation can optimize the resource of TT&C, and enhance management capacity of the ground station. For researching and validating topology, protocol and TT&C capacity of the ground station, an integrated simulating platform has been designed based on STK and OPNET. This platform includes the topology of inter-satellite link, the model of satellite node and ground station. Furthermore, the data format of telemetry between satellites and stations complies with Space Communication Protocol Specification. Finally, with constellation and stations combined, telemetry transmission is simulated in both normal mode and another mode with one satellite invalidated. Through analyzing indices like telemetry time delay and throughput, correctness and effectiveness is validated in the topology of constellation and routing algorithm. This platform has achieved a certain degree of universal property, and can provide technical method for researching other satellite communication networks.

Keywords Navigating constellation · Inter-satellite link · OPNET · Telemetry simulation · Telemetry time delay

4.1 Introduction

Inter-satellite link is an important part of the navigating constellation, which can be used for communication and ranging. There is necessary requirement in the field of TT&C using inter-satellite link, which is one of the crucial technique for improving the performance of navigating system. Recently, domestic research for inter-satellite

---

B. Jiang (✉) · J. Li
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: superb522@163.com

B. Guo
Weinan Station, Xi'an Satellite Control Center, Weinan 714000, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_4

link has been studied, including topology, routing protocol and so on. Based on these researches, the topology and routing protocol of inter-satellite link should be simulated to validate whether it could support TT&C through the use of computer simulation.

Network simulation technology can set up virtual network on the computer, which simulates all kinds of action to evaluate and forecast performance aiming to enhance the designing accuracy of network. It includes the simulation of both constellation and network protocol. SaVi and STK are commonly used to simulate designing constellation. OPNET, NS-2 and GloMoSim are used to simulate protocol of network. OPNET is a useful tool in the field of network simulation, which can perfectly model and construct wireless circumstance. The model of OPNET has been used broadly in the communication of satellites. SHEN [1] proposed the protocol and algorithm suiting LEO satellite network, and built simulating platform based on the IP. However, SCPS existing in navigating constellation in actual use differs from IP considerably [2]. The model of network, node and process for simulating performance of AOS protocol has been implemented in paper [3], which could simulate the curve illustrating relation of throughput and packet length on the condition of diverse BER. The function of SCPS-NP protocol based on OPNET was implemented in paper [4], which could analyze the performance of SCPS-NP. However, it didn't focus on concrete satellite traffic and application.

The simulating platform in correlative research rarely aims at combining the navigating constellation and ground stations using inter-satellite link. Consequently, this paper builds simulating platform of navigating constellation based on OPNET, which can analyze the performance of inter-satellite link, and research topology and routing algorithm of SCPS. The concrete telemetry data are simulated to relay through inter-satellite link and reach ground station on this platform. Finally, indices including telemetry time delay and throughput are counted and analyzed for statistics. In conclusion, this platform can validate the performance of network that includes navigating constellation and ground station, and the data originated from the correlative simulation can be used to analyze and research.

4.2 Design and Implementation of a Simulating Platform

According to the characteristics of the inter-satellite link, simulating model of navigating constellation based on STK, and that of network protocol based on OPNET are established, which introduce real orbit parameters. Consequently, an integrated simulating platform of navigating constellation is set up. Since there is no standard satellite node and ground station model in OPNET, and wireless models in OPNET are not suitable for satellite network, so a special navigating network is proposed based on wireless mechanism of OPNET in this section. This model includes three layers from the bottom up: process field, node field and network field. Process field realizes the algorithm protocol and queuing strategy, which has been implemented by Proto-C language. The basic data packet, detailed protocol

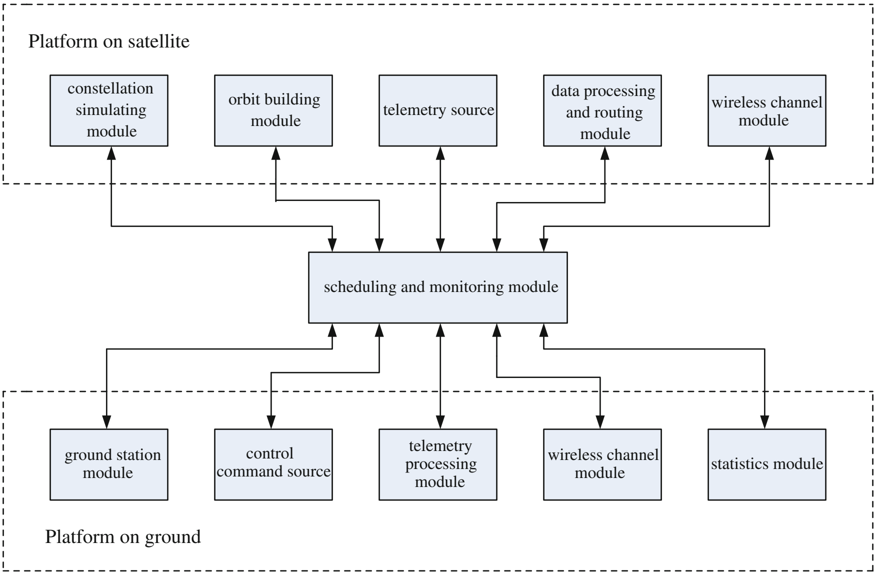


Fig. 4.1 Overall block diagram of a simulating platform

and queuing state are described by finite state machine (FSM). Forwards, FSM and packet stream constitutes node field, which simulates hardware or software resource inside the satellite node, like transmitter machine, data process unit and so on. Network field has implemented the function of satellite network, in order to reflect the performance of topology. The whole structure is described in Fig. 4.1 that is divided into two parts: satellite platform and ground platform. The satellite platform creates telemetry data, and sends them to ground station through inter-satellite link or direct link. The ground station is responsible for processing the telemetry data and completing the statistics. At the same time, the control command originated from ground station can reach the target satellite through the same route.

4.2.1 Constellation and Orbit Determination

This paper is based on the Walker constellation and its parameters is named 24/3/1 (it denotes the satellites number of 24, orbits number of 3 and phase factor of 1). The network is formed through inter-satellite link in the inner constellation, and each satellite is in procession of some links within one orbit plane and some ones within different orbit planes. Of course, the permanent link has priority.

The navigating constellation has been established based on STK [5]. The STK orbit files in accordance with the orbital parameters are imported into OPNET, so that the moving of satellites will comply with these orbits in the simulating process.

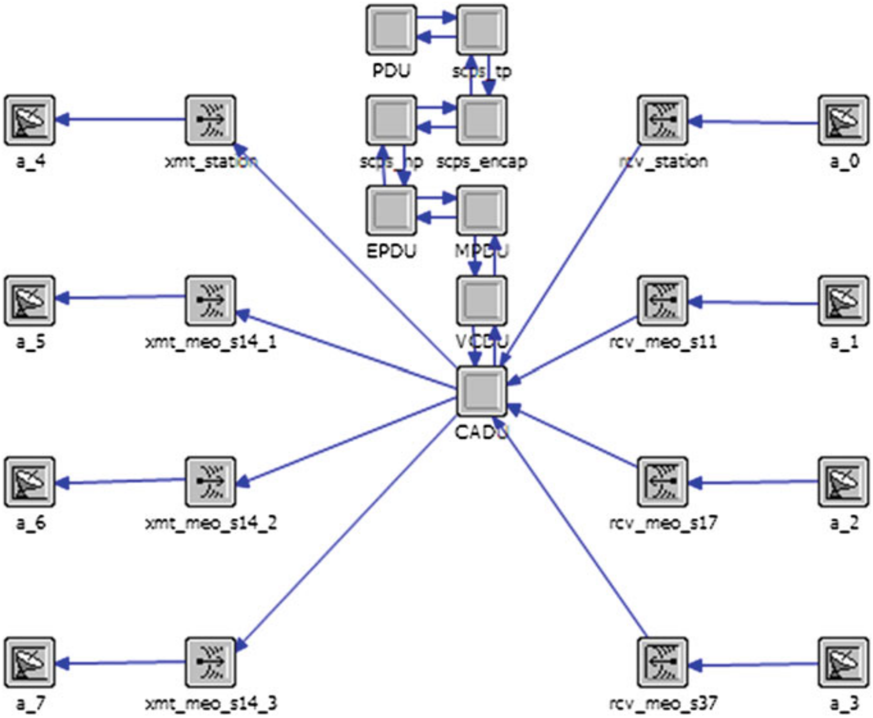


Fig. 4.2 Satellite simulating node model

Constellation is constituted by inter-satellite links and satellite nodes that include transmitter, receiver, antenna, telemetry source unit, data processing unit and so on, as shown in Fig. 4.2. The transmitter, receiver, and antenna unit are based on OPNET and wireless transmission is described by 14 pipeline stages, which are performed successively to simulate radio channel, wireless gain calculation, BER statistics and other processes.

4.2.2 The Source Module of Telemetry Data

It demonstrates the entire process of the telemetry package from generation to the physical transmitter layer, which is denoted as finite stat machine in Fig. 4.3. As shown in Fig. 4.3, the internal structure comprises three kinds of state modules, named the initial state of INIT, telemetry packet generation state of GENERATE, and end state of STOP. The broken line represents the transfer diagram, which can be transferred between two states according to the state transition condition when interrupt request is received.

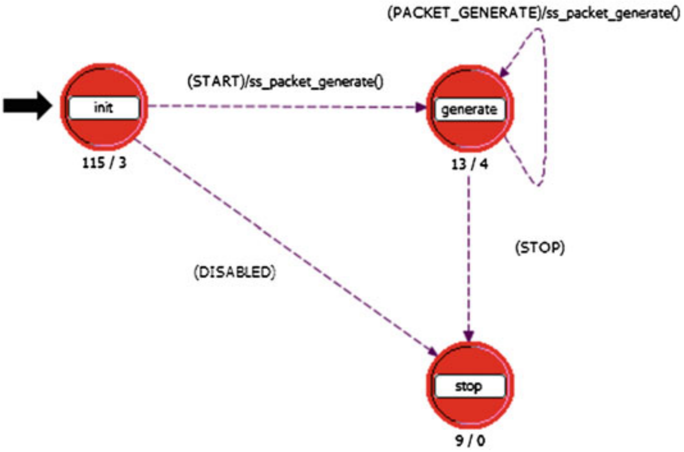


Fig. 4.3 State transition diagram of telemetry data source

The internal processes of finite state machine are implemented in Proto-C language, which includes core function and standard C/C++ language can efficiently describe the discrete event of system. Meanwhile, the module can be customized for a plurality of telemetry frame types, size and period to simulate the real telemetry state of the navigating satellite.

4.2.3 Network Routing Module

After the establishment of topology for inter-satellite link, the network routing problem should be considered firstly. For implementing routing of network, you can use a dynamic routing protocol, or plan-based routing scheme [6]. Different from dynamic routing schemes, plan-based routing is determined in advance, rather than relying on dynamic routing. It does not mean that routing tables always keep the same but in a period of time. As the satellite network has the characteristics of low data processing capability, limited bandwidth resources and long time delay, plan-based routing becomes the scheme of priority. Therefore, this paper adopts static routing scheme based on plan, which makes use of determined routing tables. The specific program is selected as follows: the system cycle is divided into several time slots by 15 min, and the satellite network route is fixed within a time slot. The routing table of next time slot will be enabled at the end of this time slot. The satellite does not require computing in real time, and you only need to switch the routing table at the end of the time slot. The real-time calculation of static routing tables by ground will be implemented and updated by remote control command after failure of a satellite or ground station.

The steps generating static routing tables are: (1) the report of visibility among satellite nodes and ground stations has been implemented using STK and some constraints [7]. The adjacent matrix of visibility has been obtained through STK/Matlab interface using the powerful computational function of Matlab, and the corresponding position in this matrix is 1 if the source node is constantly visual to the destination node or else 0. (2) These routing tables can be established through the adjacent matrix of visibility and ant colony algorithm [8] depending on independent search and mutual cooperation of ant colony. Consequently, the static routing tables can be calculated within the cycle of static routing slot. Starting with the first 15 min of a period, for example, the ground-based monitoring scheme generated from the static routing tables is shown as Fig. 4.4.

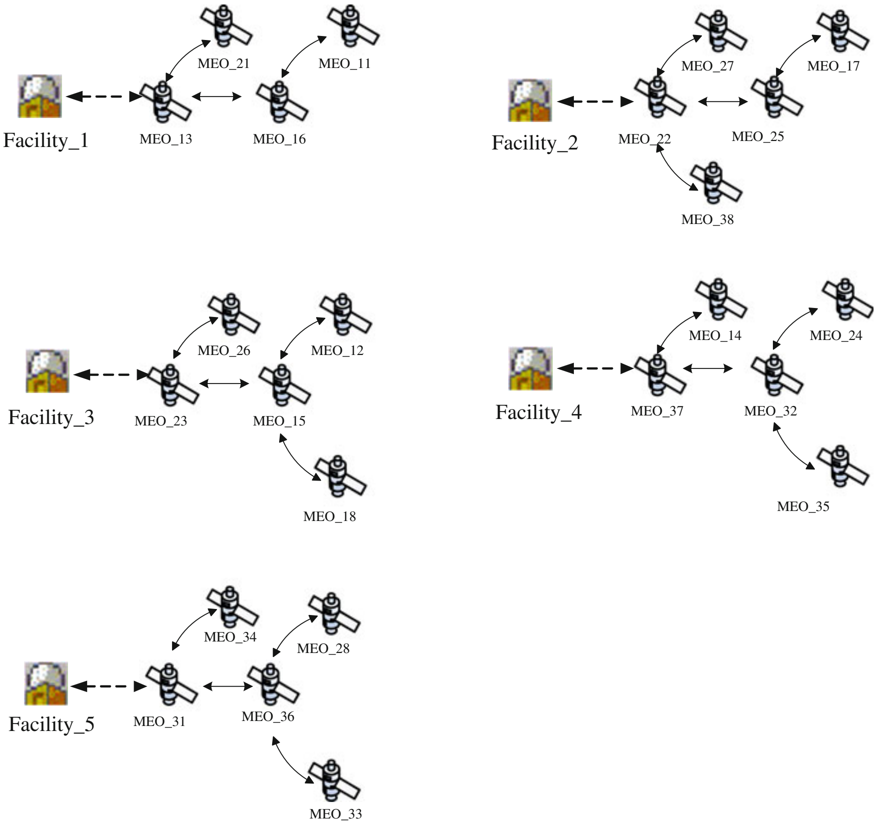


Fig. 4.4 Routing strategy between navigating satellites and ground stations

4.2.4 Data Transmitting and Processing Module

Advanced orbiting system (AOS) replaces the traditional PCM telemetry at this module. A series of protocol by CCSDS can be assorted to the physical layer, data link layer, network layer, transport layer and application layer like TCP/IP. AOS is adopted at data link layer, which suits large spacecrafts and scenarios for types of traffic. While at the network layer, transport layer and application layer, CCSDS has developed space communication protocol specification (SCPS) including SCPS-NP and SCPS-TP, which aims at satisfying the space as much as possible to achieve compatibility and interoperability with the Internet protocol.

Due to space limitation, the detailed designing process of other modules in Fig. 4.1 will not be elaborated here.

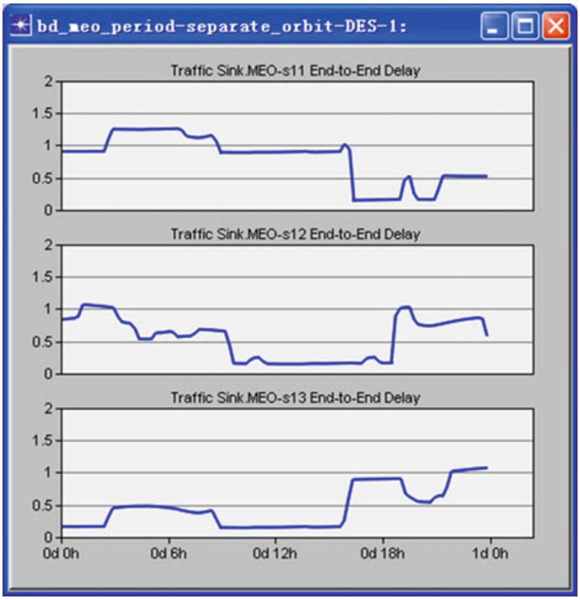
4.3 Simulating Experiment and Analysis

The simulating scenario includes: (1) a constellation comprises 24 MEO satellites; (2) 5 USB facilities located on the territory of China; (3) each single MEO satellite has 4 antennas, which are used to communicate with other satellite nodes. The operating model is: 5 facilities track 5 satellites as main nodes and communicate with other MEO node through inter-satellite links, so all of telemetry data from 24 satellites can be received by these 5 facilities. Next based on this scenario, the telemetry transmission is simulated both on the scene of normal mode and another mode with one satellite invalidated. Through analyzing telemetry time delay and throughput, correctness is validated about the topology of constellation and routing algorithm.

4.3.1 Telemetry Time Delay

Telemetry time delay is a crucial parameter that could measure performance of topology and routing. For the satellite network formed of moving nodes, the stability of network has a considerable significance, because re-constructed routing will seriously affect the delay of communication. Therefore, the changing of time delay should be test within 1 day on the simulating platform firstly. This paper sets the simulating time of 24 h, the frame period of 1 s, and the frame length of 2,048 bits during operating simulation. The ground stations receive telemetry data of 24 satellites in real time, and make the correlative statistics. The telemetry time delay and other types of indicators can be obtained after the completion of simulation. The graph curve of telemetry time delay obtained is shown in Fig. 4.5, which belong to MEO-s11, MEO-s12 and MEO-s13.

Fig. 4.5 The statistics *curve* of telemetry time delay (units s)



The variation of telemetry time delay is mainly affected by two factors: the first is the changing distance between satellite and the ground station, and switched static routing table on the other hand. As seen from Fig. 4.5, the impact introduced by changing distance is smooth and continuous, while the great jumping point of graph curve indicates the impact from switching of routing table. In order to analyze the specific value coming from this impact, this paper has made a correlative statistics within 24 h as shown in Table 4.1.

As seen from Table 4.1, the average time delay value of all satellites is 634 ms, at the same time minimum of 147 ms and maximum of 1,420 ms, respectively. It

Table 4.1 The statistics of telemetry time delay from navigating satellite within 24 h (units s)

	Maximum	Minimum	Mean value	Standard deviation
MEO-s11	1.261	0.148	0.794	0.366
MEO-s12	1.066	0.148	0.547	0.322
MEO-s13	1.071	0.147	0.454	0.310
MEO-s14	1.052	0.156	0.555	0.265
…	…	…	…	…
MEO-s35	1.367	0.536	0.945	0.272
MEO-s36	1.271	0.148	0.511	0.319
MEO-s37	0.687	0.150	0.413	0.204
MEO-s38	0.908	0.371	0.745	0.175
Mean value	1.077	0.196	0.634	0.275

indicates that switching the routing table causes some impact, however the particularly serious delay and packet loss does not occur. The standard deviation of time delay is 275 ms, which reflects the fluctuations of telemetry delay value within an entire period.

4.3.2 Scenario with the Failure of Single Satellite

This paper has considered the condition that one satellite within the constellation fails and could not participate in the network. The scenario like that has been simulated and counted about telemetry throughput and time delay. Now supposing the failure of MEO-s11 satellite, the re-counted routing table should be generated and updated to the remaining satellites, in order to keep the constant communications. The comparison with the normal mode is simulated and shown as Fig. 4.6.

As seen from Fig. 4.6, after the failure of MEO-s11 satellite occurs, the average value and standard deviation are as much as the value of normal state in addition to improve the maximum delay value.

In order to measure the impact of throughput on the ground station under rerouting, this paper validates the capability of telemetry throughput on ground station under the normal mode and single satellite failure mode. The simulation results include the average value of throughput and standard deviation, and the latter represents the value deviating from the average in Figs. 4.7 and 4.8 respectively.

As seen from Fig. 4.7, after the failure of MEO-s11 satellite and re-routing occurs, ground stations can normally receive all telemetry data of 23 satellites, and the average value has little changes. Further, the statistical value of standard deviation is shown as Fig. 4.8.

As seen from Fig. 4.8, after the failure of MEO-s11 satellite and re-routing occurs, the value of standard throughput deviation is greater than the prior one. This job indicates that the telemetry throughput of ground station from the average fluctuation becomes great when one satellite within the constellation fails and can not participate in the network, but the receiving of the ground station meets the overall requirements without affecting the normal receipt and demodulation.

Fig. 4.6 The comparison about telemetry time delay (units s)

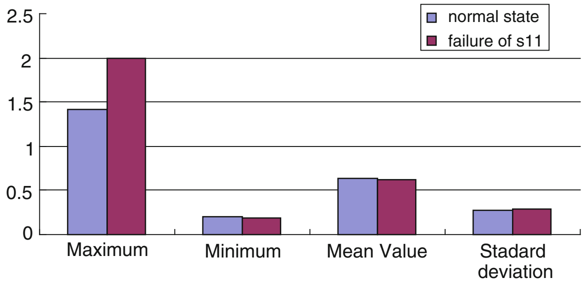


Fig. 4.7 The comparison from average of telemetry throughput on ground station (units bps)

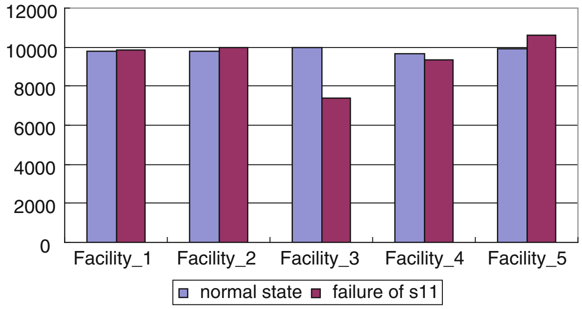
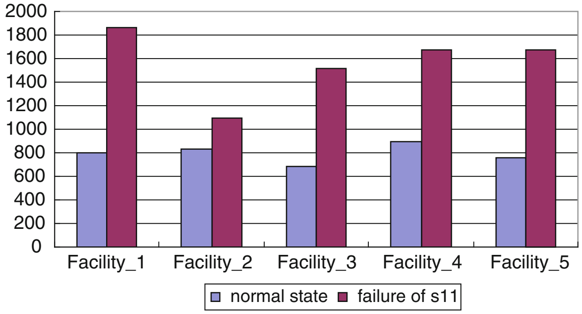


Fig. 4.8 The comparison from standard deviation of telemetry on ground station (units bps)



4.4 Conclusion

A simulating platform of navigating constellation has been established, which models the satellite node, inter-satellite link and ground station, and simulates the specific scenarios. It simulates telemetry reception for the constellation and ground station in normal mode and another mode with one satellite invalidated. This platform has achieved a certain degree of universal property, and can provide a simple and effective technical means for analyzing and researching inter-satellite links and satellite network.

Acknowledgments Due to the orbit parameter of navigating satellite needed, the author would like to extend sincere gratitude to iGMAS for the correlative data downloaded from its website www.igmas.org.

References

1. Shen JP, Yu HF (2009) OPNET based low earth orbit satellite network simulation platform. Comput Eng 35(18):237–239
2. Xu F, You Z (2007) CCSDS space communication protocol and its comparison with the internet communication protocol. Aerosp China 05:26–29

3. Li SF (2005) Throughput performance analysis of AOS protocol and OPNET modeling and simulation. Doctoral thesis of national university of defense technology, Changsha
4. Zheng WM, Fan YH, Wang YW (2011) Simulation design of SCPS-NP based on OPNET. Commun Technol 2(44):48–52
5. Zhang XG, Liu Y, Tang S (2011) Design and realization of STK-based visualization simulation software for satellites. J Spacecraft TT&C Technol 30(4):73–76
6. Sun LM, Lu ZX, Wu ZM (2004) Routing technology for LEO satellite network. Chin J Comput 27(5):659–667
7. Li J, Hu HJ, Yu PJ (2012) Analysis of requirement and constrained model of inter-satellite-link TT&C scheduling problem on navigation constellation. In: Proceedings of artificial intelligence and computational intelligence. Springer Berlin Heidelberg, Berlin, pp 160–170
8. Ye GQ, Li J, Zhang YJ (2013) Polymorphic ant colony algorithm for continuous monitoring of navigation constellation. J Astrodynamics 3(4):89–94

Chapter 5
The Co-simulation of Airborne Stable Platform Based on Adams/Simulink

Qiang Dong, Xiaoxu Yang, Junfeng Han and Yan Li

Abstract This paper proposes an approach to design a simulation method based on Adams and Simulink to improve performance of the stable platform. First, this paper gives an Adams/Simulink co-simulation method of airborne stable platform. Second, it designs mechanic model in Adams and control model in Simulink. Finally, simulate mechanic and control model of airborne stable platform, the simulation result shows that the co-simulation of Adams and Simulink is feasible. In the Adams model, it designs different error factors as follows: friction disturbance, motor disturbance, aircraft vibration disturbance and gear backlash. The simulation result is further close to the actual situation. Therefore, the co-simulation method of this paper is better than traditional simulink method. The design of co-simulation has guiding significance for improving the mechanic and control module of airborne stable platform.

Keywords Adams · Simulink · Co-simulation · Stable platform

5.1 Introduction

The airborne stabilized platform is a carrier in aviation field. It separates the load from aircraft's effects and provides a simple working environment for the target load. In an airborne stabilized platform project, control system frequency bandwidth and mechanical system frequency bandwidth cannot be effectively combined. System bandwidth greatly limits the platform performance. The reasonable distribution of control system frequency bandwidth and mechanical system frequency

---

Q. Dong (✉) · X. Yang · J. Han · Y. Li
Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Science,
Xi'an 710119, China
e-mail: dongqiang@opt.cn

Q. Dong
Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Science,
Beijing 100049, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_5

bandwidth will reduce development costs and improve system performance. Knowing the bandwidth of stabilized platform control system and mechanical system has great significance on airborne platform bandwidth allocation.

According to the experience of practical engineering, the principle of frequency allocation is improving control system frequency and reducing the mechanical system frequency as much as possible. The principle will inevitably lead to the rising of design and manufacturing costs, and bring resonance effect. This paper designs a co-simulation system based on Adams and simulink, the system analyze airborne stabilized platform control system and mechanical system in time-domain and frequency-domain, provides the simulation data. The simulation date provides the basis for the subsequent work frequency allocation on airborne platform and a reference for the improvement of airborne platform control system and mechanical system.

5.2 The Co-simulation of Stable Platform Based on Adams and Simulink

Adams has a very strong kinematics and dynamics analysis. The users have modeling, simulation and analysis to all kinds of mechanical system with Adams's powerful modeling and simulation environment. Matlab is indispensable in control system simulation because of powerful calculation capabilities. The co-simulation based on Adams/Simulink puts together mechanical system and control design simulation. It makes the combination analysis of mechatronics to be realized [1]. The flow chart to co-simulation of airborne stabilized platform's Mechanical structure and control system, as shown in Fig. 5.1 [2].

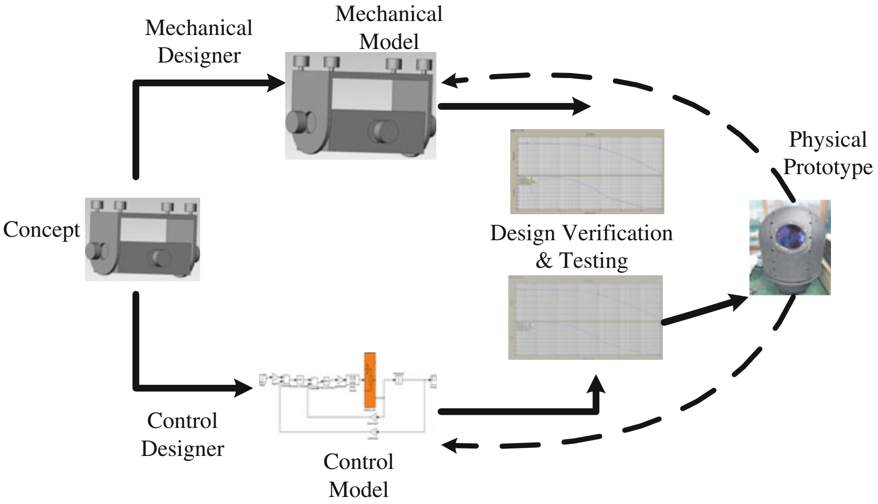


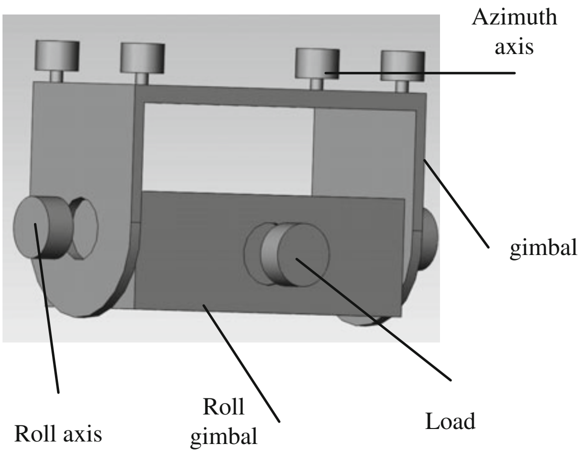
Fig. 5.1 Design process before ADAMS/controls

The flow chart of co-simulation is as follows. Firstly, it designs dynamics analysis model of airborne stabilized platform, and loads the different motion parameters of dynamics in Adams, as shown in Fig. 5.1, Mechanical Model. Secondly, it designs control system model of airborne stabilized platform, and load the different control parameters in Simulink, as shown in Fig. 5.1, Control Model. Thirdly, it turns airborne stable platform's nonlinear mechanical model into S-Function which Simulink identifiable function by Plant Export module outputs of Adams. Fourthly, it loads S-function into the load part of Simulink control system, realizes the co-simulation of Adams/Simulink [3].

5.3 The Design of Airborne Stabilized Platform Mechanical Module

The Adams model of airborne stabilized platform designed by Solidworks. Adams model includes airborne stabilized platform components. It imports module into Adams/View after each component module being assembled. In Adams/View, material of parts of each module will be specified, azimuth axis and horizontal roll axis will be constrained, load actual work load to simulate part [4]. Automatic Dynamic Analysis of Mechanical Systems model of airborne stabilized platform including azimuth gimbal, Horizontal roll gimbal, azimuth axis, Horizontal roll axis, simulated load, motor, reducing mechanism, angular velocity trans, IMU (Inertial Measurement Unit) etc. The Adams simulation mode of airborne stabilized platform, as shown in Fig. 5.2.

Fig. 5.2 Adams model of stable-platform



5.4 The Design of Airborne Stabilized Platform Control Module

The airborne stabilized platform uses azimuth—roll designs. Stabilized platform uses traditional rate—position control method. The control principle block diagram, as shown in Fig. 5.3.

The circuit for power conversion of Servo driver adopts H bipolar four-phase full bridge circuit which composed of IGBT power modules. It works in PWM (Pulse-Width Modulation) mode. In general, the driving cut-off frequency of PWM modulation circuit is far great than the cut-off frequency of torque motor in Airborne stabilized platform drive system. The transfer function of power amplifier simplifies to an amplifier [5], as shown in formula (5.1).

$$G_{PWM}(s) = K_{PWM} \tag{5.1}$$

The gyroscope is angular velocity measuring device in inertial space. Its output is an analog DC voltage which is proportional to the input angular velocity.

In Fig. 5.3, positional corrector and rate corrector uses PID correction. Therefore, the transfer function of angular rate gyro can approximate as a proportional cycles, as shown in formula (5.2).

$$G_{\mathrm{g}}(s) = K_{g0} \tag{5.2}$$

The proportional coefficient $G_{\mathrm{g}}(s)$ is determined by gyro prior-conditioning circuit amplification factor and scale factors [6].

According to the requirements of azimuth and elevation driving torque of airborne stabilized platform, drive motor adopts J90LYX01C1 permanent magnet DC torque motor which produces by Chengdu precision motors company. The reduction ratio both of azimuth and elevation motor are all 74:1. The gear transmission ratio of azimuth axis is 8:1. The gear transmission ratio of elevation axis is 6:1. According to the relevant technical indexes of the motor, calculation of motor parameters, are as shown in formula (5.3).

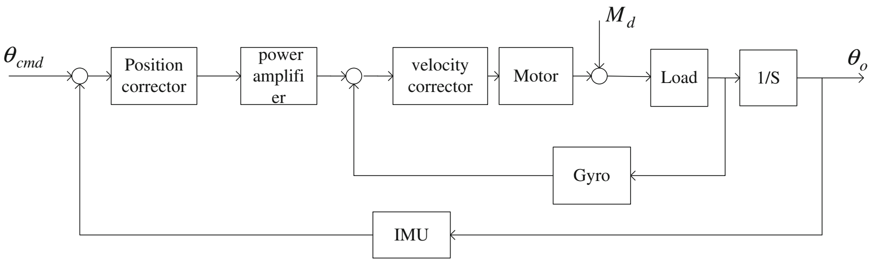


Fig. 5.3 Control block diagram of stable platform

$$\begin{cases} J_A = 0.1\,\text{kgm}^2\; J_E = 0.1\,\text{kgm}^2 \\ J_{\text{motor}} = 4 \times 10^{-4}\ \text{kgm}^2 \\ J_{\text{total}} = 0.1\,\text{kgm}^2 \\ \mathrm{R_a} = 7.45\,\Omega \\ L_a = 4.4\,\text{mH} \\ C_{\text{m}} = 0.36\,\text{N} \cdot \text{m/A} \\ C_{\text{e}} = 0.038\,\text{v} \cdot \text{s/rad} \end{cases} \tag{5.3}$$

J_{A} is gimbal inertia, J_{E} is load inertia, J_{motor} is motor moment of inertia, J_{total} is total moment of inertia which includes motor and load, R_a is resistance of armature winding, L_a is total inductance of the armature loop, C_m is motor torque coefficient, C_e is motor back-EMF coefficient.

Bringing the above parameters to formula (5.4), machine-electro time constant and electromagnetic time constant are as follows:

$$\begin{cases} T_e = \frac{L_a}{R_a} \\ T_m = \frac{R_a J}{C_e C_m} \end{cases} \tag{5.4}$$

T_m is machine-electro time constant, T_e is electromagnetic time constant.

S-function of motor drive module will be obtained by combining with motor back-EMF coefficient, machine-electro time and electro-magnetic time constant, and it is shown in formula (5.5).

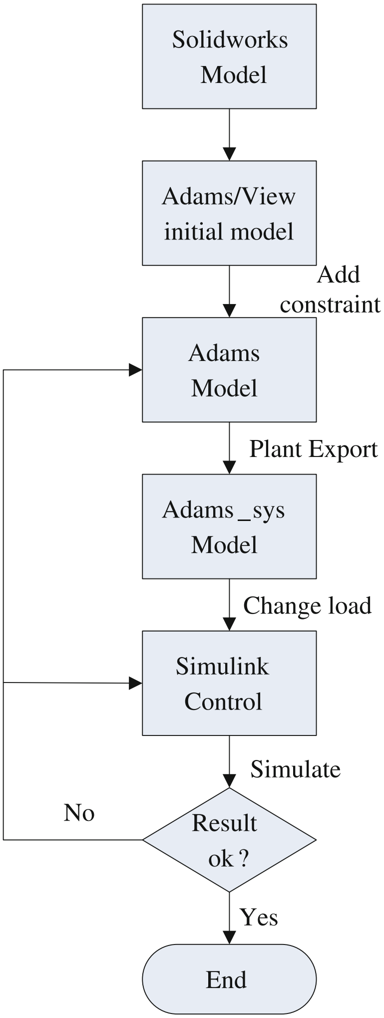
$$G_M(s) = \frac{\omega(\mathrm{s})}{U_a(s)} = \frac{1/C_e}{T_m T_e S^2 + T_m S + 1} \tag{5.5}$$

5.5 The Experiment of Co-simulation Based on Adams/Simulink

In Solidworks, designs airborne stabilized platform mechanical model, generates the IGS format file, imports rigid body parts into Adams. In Ansys, finite element analysis the flexible body parts, generating the MNF format file and importing into Adams. In Adams, the parts assembly, connection and all kinds of constraints will be added, and then Adams dynamic model will be achieved [7]. In Adams, the Plant Export outputs Adams model which can be read by matlab into matlab. In matlab, it generates Adams_sub module by running the plant module. Using Adams_sub module replace Simulink control module.

Operate simulation, modifying the mechanical design and control parts to perfection respectively when simulation results can not satisfy the demand. Modifying mechanical module specified materials, gimbal installation, Shafting arrangement, load installation to optimize system performance. It optimizes system performance by modifying control module PID parameters, Angular rate sensor and encoder. The flow chart of co-simulation based on Adams and simulink, as shown in Fig. 5.4.

Fig. 5.4 The flow diagram of Adams/simulink co-simulation



According to the flow chart of co-simulation based on Adams and simulink, it generates Adams_sub module, adds Adams_sub to control system in simulink, and generates stable platform co-simulation of control module. The co-simulation of control module based on Adams and simulink, as shown in Fig. 5.5 [8].

In the Adams model, it designs different error factors as follows: friction disturbance, motor disturbance, aircraft vibration disturbance and gear backlash. The simulation waveform accords to the method in Fig. 5.5, as shown in Fig. 5.6. In the simulation, the input is set to a step signal, the position loop PID parameter is set to 10, 2, 0, the speed loop PID parameter is set to 50, 10, 0.

The co-simulation of airborne stabilized platform improves system performance by adjusting the parameters of position of PID and rate of PID. In Fig. 5.6, the

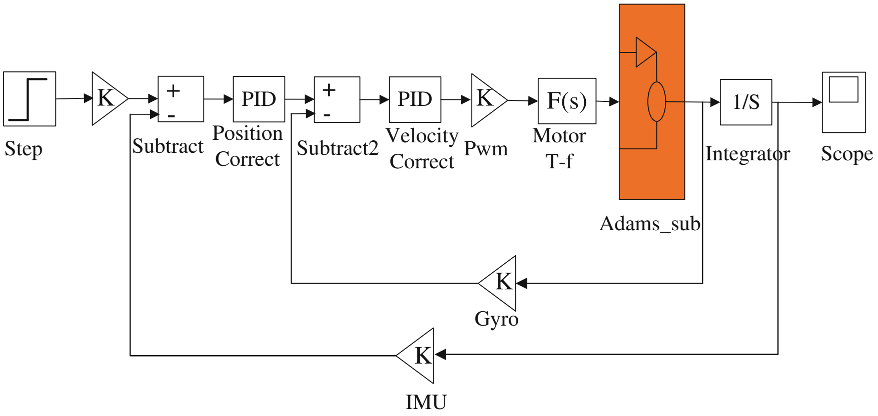


Fig. 5.5 The co-simulation of control module based on Adams and simulink

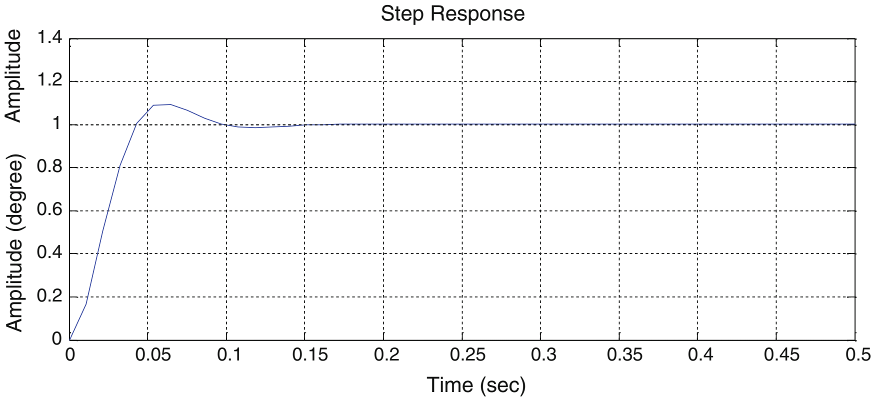


Fig. 5.6 Step response of azimuth axis in Adams/simulink simulation

position steady-state error is control under 0.02°, it satisfies system's demand, the step response is perfect. The frequency response of azimuth axis in Adams/simulink simulation is shown in Fig. 5.7. In Fig. 5.7, it can be observed that the cut-off frequency of system is 75.6 rad/s (12 Hz). The cut-off frequency is slightly higher than the actual system cut-off frequency because of the friction and vibration of Adams mechanical model does not be completely consistent with actual system.

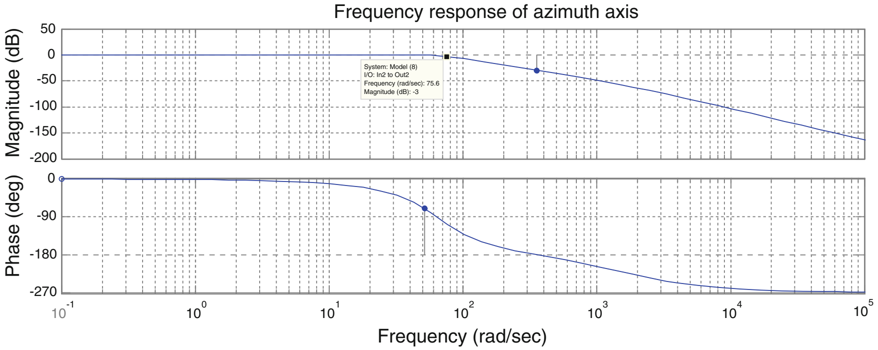


Fig. 5.7 Frequency response of azimuth axis in Adams/simulink simulation

5.6 Conclusions

This paper gives an optimal frequency both on the cut-frequency of mechanical system and control system based on airborne stabilized platform frequency assignment problem. The assignment of frequency reduces the system cost. The paper designs a co-simulation of airborne stabilized platform based on various areas co-modeling and co-simulation. It gives a design method of mechanical module and control module and running simulate [9]. The simulation result shows that the co-simulation system is feasible and effective. The co-simulation is better than the traditional simulation to reflect airborne stabilized platform actual performance. It will improve airborne stabilized platform performance by optimizing the mechanical system and control system design based on simulation results. The co-simulation design has guidance significance to the improvement of airborne stabilized platform performance. It can reduce the experimental cost and development cost of airborne stabilized platform design.

References

1. Tian FL, Sun RL (2006) The flexible robot arm simulation system based on ADAMS and Simulink. Mech Electron 10:62–64
2. Li H, Fan BH, Liu Q (2007) The co-simulation of ADAMS and MATLAB. Sci Technol Inf 9:126
3. Li SZ, Liu P, Wang YM (2012) Analysis of pantograph and centenary vibration control based on co-simulation of ADAMS and Simulink. Electric Locomot Mass Transit Vehicles 34 (6):35–37
4. Brezina T, Hadas Z, Vetiska J (2011) Using of co-simulation ADAMS-Simulink for development of mechatronic systems. In: IEEE Mechatronika 2011 14th international symposium

5. Affi Z, Romdhane L (2005) ADAMS/Simulink interface for dynamic modeling and control of closed loop mechanisms. In: Proceedings of the 7th WSEAS international conference on automatic control, modeling and simulation
6. Li HQ, Guo BW, Xu F (2009) Virtual prototype modeling and simulation analysis of servo based on ADAMS and Simulink. J Syst Simul 21(21):6886–6888
7. Xu ZM, Hu KB, Zhang ZF et al (2011) Modeling and control strategy simulation for automobile EPS system based on Adams and Matlab. J Chongqing Univ Technol (Nat Sci) 24 (10):7–12
8. Rao SJ (2009) Vehicle modeling and Adams-Simulink co-simulation with integrated continuously controlled electronic suspension (CES) and electronic stability control (ESC) models. Doctoral thesis of the Ohio state university
9. Rahmfeld R (2003) Modeling of hydraulic mechanical systems co-simulation between ADAMS and Matlab/Simulink. Lectures for design of mechatronic system I + II using CAD

Chapter 6
A Study on the Alignment of Platform INS Based on the Rotation Around Azimuth Axis

Chan Li, Yuan Cao and Shifeng Zhang

Abstract Platform INS is a type of high-precision navigation equipment and widely used in launch vehicles, missiles, aircrafts, ships and other types of carrier. The traditional gyrocompass alignment method used for the platform INS on stationary base is affected by the gyro's random drift, installation error and the scale factor errors and gimbal locking errors. This paper presents a new gyrocompass alignment method based on the continuous rotation around the platform's azimuth axis. Through locking the horizontal gimbal control loop and rotating around the azimuth axis, the alignment is carried out by estimating the misalignments based on horizontal gyros torque currents information. Simulation results show that the initial alignment can be quickly carried out with high precision and the proposed method has more advantages than traditional gyrocompass alignment method.

Keywords Platform INS · Initial alignment · Stationary base · Gyro torque current

6.1 Introduction

Platform inertial navigation system (PINS) is a type of high-precision navigation equipment and widely used in launch vehicles, missiles, aircrafts, ships and other types of carrier. The initial misalignments of PINS affect the navigation performance seriously and initial alignment has to be employed to provide adequate navigation precision for the carriers. The alignment methods used for PINS on stationary mainly include optical method, gyrocompass alignment and accelerometer-based multi-position alignment method. As the PINS with self-leveling and self-locking feature, so the self-align method like gyrocompass alignment has

---

C. Li (✉) · Y. Cao · S. Zhang
College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China
e-mail: lcpku@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_6

attracted widespread attention [1, 2]. However, the gyrocompass alignment method is affected by the gyro's random drift, installation error and the scale factor errors and gimbal locking errors. Traditional continuous calibration and alignment methods with multi-axes rotation in reference [3] can avoid the gimbal locking errors, but utilizing accelerometer outputs not only introduces the accelerometer error, but also many gyro error coefficients, which are very difficult to be estimated within one hour [4, 5]. Thus it cannot satisfy the fast response requirement. Besides, position selection for the alignment is another important problem to be considered [6, 7]. Observability analysis based on the observability matrix can help to analyze the alignment scheme, so a useful method for observability analysis have to be adopted for alignment scheme design [8].

Aiming at solving the above problems, we proposed a new gyrocompass alignment method based on the continuous rotation around the platform's azimuth axis. In this method, the angular velocities are sensed through the horizontal gyro's torque currents which can be obtained in the gyro torque servo loop, so it can be used to estimate the PINS's initial misalignments. Besides, the locking errors are eliminated by continuous rotation and the gyro torque errors are also considered and estimated to reduce the alignment errors. Simulation results show that the proposed method can provide higher precision azimuth information for the PINS than the traditional gyrocompass alignment method.

6.2 Alignment Scheme Design

In the inertial platform, there are two control loops to maintain the high precision navigation. One is used for gimbal control and the other is used to control the gyro torque. In the gimbal control loop, the angular velocities sensed by gyros are mainly used to control the rotation of the gimbals. In the gyro torque control loop, the commands of angular velocities are realized by control the precession angle rate of the gyro's input axis through the gyro torque. The proposed alignment scheme is developed based on these two control loops. Firstly, through the gimbal control loop and horizontal accelerators, we can level the platform and lock the horizontal axes. Then the gyro torque loop is employed to make the platform rotating around the azimuth axis and the torque current outputs of horizontal gyros are used to estimate the misalignments of PINS. When the estimation results are obtained, the control for PINS is switched to the mode of navigation and the misalignment compensation will be implemented. The designed alignment scheme can also be described as the following flow chart (Fig. 6.1).

Compared to the conventional multi-position gyrocompass alignment scheme, the proposed scheme has the following advantages: 1. Do not require locking the gimbal and leveling the platform many times, which can reduce locking errors and time; 2. Can take advantage of the gyro current information during the whole process of platform rotation. In the next section we will discuss the dynamics model and observation model used in this alignment method.

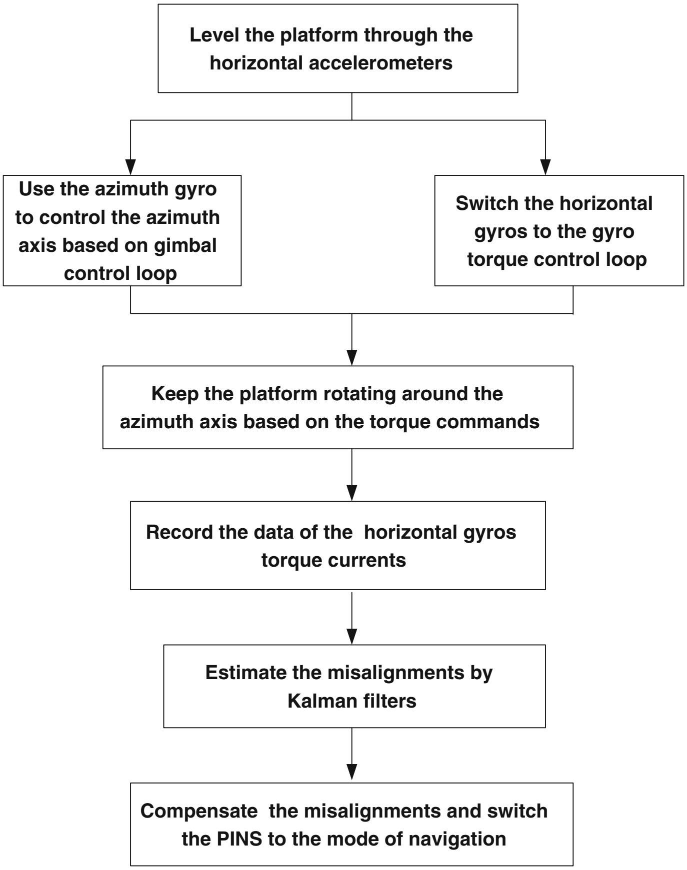


Fig. 6.1 The flow chart of alignment

6.3 Dynamic Model and Observation Model of the Alignment

6.3.1 Dynamic Model and Observation Model

When the platform rotating around the azimuth axis, the azimuth angle between computation frame of the inertial platform and the local geographical frame (navigation frame) is changing continuously. Denoting A_t as the azimuth angle, relationship between the computation frame and the local geographical frame can be expressed as follows.

$$\mathbf{C}_n^c = \begin{bmatrix} \cos A_t & 0 & \sin A_t \\ 0 & 1 & 0 \\ -\sin A_t & 0 & \cos A_t \end{bmatrix} \tag{6.1}$$

In the navigation frame, the rotation angular velocity of the platform can be expressed as

$$\boldsymbol{\omega}_r = \begin{bmatrix} 0 \\ \omega(t) \\ 0 \end{bmatrix} \tag{6.2}$$

Earth's rotation angular velocity vector in the navigation frame components is

$$\boldsymbol{\omega}_e = \begin{pmatrix} \Omega \cos L \\ \Omega \sin L \\ 0 \end{pmatrix} \tag{6.3}$$

where Ω is the angular velocity of the Earth's rotation, L is the local geographical latitude. When the horizontal axes of the platform are locked and rotating around the azimuth axis, the command angular velocity can be expressed as

$$\begin{aligned} \boldsymbol{\omega}_c &= \mathbf{C}_n^c(\boldsymbol{\omega}_r + \boldsymbol{\omega}_e) \\ &= \begin{bmatrix} \cos A_t & 0 & \sin A_t \\ 0 & 1 & 0 \\ -\sin A_t & 0 & \cos A_t \end{bmatrix} \cdot \begin{bmatrix} \Omega \cos L \\ \omega(t) + \Omega \sin L \\ 0 \end{bmatrix} = \begin{bmatrix} \cos A_t \Omega \cos L \\ \omega(t) + \Omega \sin L \\ -\sin A_t \Omega \cos L \end{bmatrix} \end{aligned} \tag{6.4}$$

In the actual frame of the platform (follow-up with the platform), this angular velocity can be expressed as

$$\boldsymbol{\omega}_p = \begin{bmatrix} \omega_{px} \\ \omega_{py} \\ \omega_{pz} \end{bmatrix} = \mathbf{C}_c^p \boldsymbol{\omega}_c = \begin{bmatrix} 1 & \varphi_z & -\varphi_y \\ -\varphi_z & 1 & \varphi_x \\ \varphi_y & -\varphi_x & 1 \end{bmatrix} \begin{bmatrix} \cos A_t \Omega \cos L \\ \omega(t) + \Omega \sin L \\ -\sin A_t \Omega \cos L \end{bmatrix} \tag{6.5}$$

where $\varphi_i (i = x, y, z)$ are the misalignment angles between the computation frame and the actual frame. Considering gyro's fixed drifts, torque scale errors and misalignment angles, the gyro error model can be expressed as

$$\begin{bmatrix} \varepsilon_{gx} \\ \varepsilon_{gy} \\ \varepsilon_{gz} \end{bmatrix} = \begin{bmatrix} k_{gx0} \\ k_{gy0} \\ k_{gz0} \end{bmatrix} + \begin{bmatrix} k_{Gx} & \Delta_{sx} & -\Delta_{ox} \\ \Delta_{oy} & k_{Gy} & -\Delta_{py} \\ \Delta_{sz} & \Delta_{oz} & k_{Gz} \end{bmatrix} \begin{bmatrix} \omega_{px} \\ \omega_{py} \\ \omega_{pz} \end{bmatrix} \tag{6.6}$$

where $k_{gi0} (i = 1, 2, 3)$ are the fixed drifts. $k_{Gi} (i = x, y, z)$ are torque scale errors. Δ_{ox}, Δ_{sx}, Δ_{sy}, Δ_{oy}, Δ_{sz} and Δ_{oz} are misalignment angles for X gyro, Y gyro and Z gyro, respectively. $\omega_{pi}(i = x, y, z)$ are the angular velocities of the platform relative to the inertial frame. When the platform is locked with the horizontal axes, the azimuth errors can be expressed as

$$\begin{bmatrix} 0 \\ \dot{\varphi}_y \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & \varphi_z & -\varphi_y \\ -\varphi_z & 0 & \varphi_x \\ \varphi_y & -\varphi_x & 0 \end{bmatrix} \begin{bmatrix} \omega_{cx} \\ \omega_{cy} \\ \omega_{cz} \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_{gy} \\ 0 \end{bmatrix} \tag{6.7}$$

Substituting Eq. (6.5) into Eq. (6.7), the dynamic model for the azimuth misalignment angle can be expressed as

$$\begin{aligned} \dot{\varphi}_y = {} & k_{g0y} - \Delta_{oy}\cos A_t\Omega\cos L + \Delta_{oy}\varphi_y(-\sin A_t\Omega\cos L) + \varphi_z\cos A_t\Omega\cos L \\ & + \varphi_x(-\sin A_t\Omega\cos L) \\ & + \Delta_{sy}\varphi_y\cos A_t\Omega\cos L + \Delta_{sy}(-\sin A_t\Omega\cos L) \\ & + k_{Gy}\left(-\Delta_{oy}\cos A_t\Omega\cos L + \omega(t) + \Omega\sin L + \Delta_{sy}(-\sin A_t\Omega\cos L)\right) \end{aligned} \tag{6.8}$$

The above equation is also the dynamic model for the alignment. As the horizontal axes are locked, the horizontal gyros torque current can reflect the platform's horizontal angular velocities, which can be expressed by

$$\begin{cases} I_X = \omega_{px} + \varepsilon_{gx} + \varepsilon_x \\ I_Z = \omega_{pz} + \varepsilon_{gz} + \varepsilon_z \end{cases} \tag{6.9}$$

where ε_x and ε_z are the random drifts for X gyro and Z gyro, respectively. Substituting Eqs. (6.5) and (6.6) into Eq. (6.9), the observation model based on gyro current can be expressed as

$$\begin{aligned} I_x = {} & (1 + k_{Gx})\big(\cos A_t\Omega\cos L + \varphi_z(\omega_0 + \Omega\sin L) - \varphi_y(-\sin A_t\Omega\cos L) \\ & + \Delta_{sx}(\omega_0 + \Omega\sin L) \\ & - \Delta_{ox}\left(\varphi_y\cos A_t\Omega\cos L - \sin A_t\Omega\cos L\right)\big) + k_{g0x} \end{aligned} \tag{6.10}$$

$$\begin{aligned} I_z = {} & (1 + k_{Gz})(\Delta_{sz}\cos A_t\Omega\cos L - \Delta_{sz}\varphi_y(-\sin A_t\Omega\cos L) - \Delta_{oz}(\omega_0 + \Omega\sin L) \\ & + \varphi_y\cos A_t\Omega\cos L - \varphi_x(\omega_0 + \Omega\sin L)\big) + k_{g0z} \end{aligned} \tag{6.11}$$

Deducting the ideal outputs, the observation models can be rewritten by

$$\begin{aligned} \Delta I_x \approx {} & \varphi_z\left(\omega_0 + \Omega\sin L\right) - \varphi_y\left(-\sin A_t\Omega\cos L\right) \\ & + \Delta_{sx}\left(\omega_0 + \Omega\sin L\right) - \Delta_{ox}\left(\varphi_y\cos A_t\Omega\cos L - \sin A_t\Omega\cos L\right) \\ & + k_{Gx}\big(\cos A_t\Omega\cos L - \varphi_y\left(-\sin A_t\Omega\cos L\right) + \Delta_{sx}\left(\omega_0 + \Omega\sin L\right) \\ & \quad - \Delta_{ox}\left(\varphi_y\cos A_t\Omega\cos L - \sin A_t\Omega\cos L\right)\big) + k_{g0x} \end{aligned} \tag{6.12}$$

$$\begin{aligned}\Delta I_z = {} & \Delta_{sz}\cos A_t\Omega\cos L - \Delta_{sz}\varphi_y(-\sin A_t\Omega\cos L) - \Delta_{oz}(\omega(t) + \Omega\sin L) \\ & + \varphi_y\cos A_t\Omega\cos L - \varphi_x(\omega(t) + \Omega\sin L) \\ & + k_{Gz}\big(\Delta_{sz}\cos A_t\Omega\cos L - \Delta_{oz}(\omega(t) + \Omega\sin L) + \varphi_y\cos A_t\Omega\cos L - \sin A_t\Omega\cos L\big) \\ & + k_{g0z}\end{aligned} \tag{6.13}$$

Equations (6.8), (6.12) and (6.13) form the model for the alignment of the platform. To facilitate the system observability analysis and misalignment angle estimation, derivative matrixes for dynamic and observation equations are needed. The state vector can be expressed as

$$\boldsymbol{X} = \begin{bmatrix} \varphi_x & \varphi_y & \varphi_z & k_{Gx} & k_{Gy} & k_{Gz} & k_{g0x} & k_{g0y} & k_{g0z} & \Delta_{sx} & \Delta_{sy} & \Delta_{sz} & \Delta_{ox} & \Delta_{oy} & \Delta_{oz} \end{bmatrix}^{\mathrm{T}} \tag{6.14}$$

The dynamic and observation can be described by

$$\begin{cases} \boldsymbol{f}(\boldsymbol{X}) = \begin{bmatrix} \dot{\varphi}_x & \dot{\varphi}_y & \dot{\varphi}_z & 0_{1\times 12} \end{bmatrix}^T \\ \boldsymbol{h}(\boldsymbol{X}) = \begin{bmatrix} \Delta I_x & \Delta I_y \end{bmatrix}^{\mathrm{T}} \end{cases} \tag{6.15}$$

Derivative matrix for $\boldsymbol{f}(\boldsymbol{X})$ can be expressed by

$$\boldsymbol{F}(\boldsymbol{X})_{15\times 15} = \frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}} \tag{6.16}$$

where

$$F_{2,1} = (-\sin A_t\Omega\cos L) \tag{6.17}$$

$$F_{2,2} = \Delta_{oy}(-\sin A_t\Omega\cos L) + \Delta_{sy}\cos A_t\Omega\cos L \tag{6.18}$$

$$F_{2,3} = \cos A_t\Omega\cos L \tag{6.19}$$

$$F_{2,5} = -\Delta_{oy}\cos A_t\Omega\cos L + \omega(t) + \Omega\sin L + \Delta_{sy}(-\sin A_t\Omega\cos L) \tag{6.20}$$

$$F_{2,8} = 1 \tag{6.21}$$

$$F_{2,12} = \varphi_y\cos A_t\Omega\cos L + \big(1 + k_{Gy}\big)(-\sin A_t\Omega\cos L) \tag{6.22}$$

$$F_{2,13} = -\big(1 + k_{Gy}\big)\cos A_t\Omega\cos L - \varphi_y(\sin A_t\Omega\cos L) \tag{6.23}$$

Other elements of the matrix F are zero. Derivative matrix for $\boldsymbol{h}(\boldsymbol{X})$ can be expressed by

$$\boldsymbol{H}(\boldsymbol{X})_{2\times 15} = \frac{\partial \boldsymbol{h}(\boldsymbol{X})}{\partial \boldsymbol{X}} \tag{6.24}$$

where

$$H_{1,2} = (1 + k_{Gx})(\sin A_t \Omega \cos L - \Delta_{ox} \cos A_t \Omega \cos L) \tag{6.25}$$

$$H_{1,3} = \omega(t) + \Omega \sin L \tag{6.26}$$

$$\begin{aligned} H_{1,4} = {} & \cos A_t \Omega \cos L - \varphi_y(-\sin A_t \Omega \cos L) + \Delta_{sx}(\omega(t) + \Omega \sin L) \\ & - \Delta_{ox}\left(\varphi_y \cos A_t \Omega \cos L - \sin A_t \Omega \cos L\right) \end{aligned} \tag{6.27}$$

$$H_{17} = 1 \tag{6.28}$$

$$H_{1,10} = (1 + k_{Gx})(\omega(t) + \Omega \sin L) \tag{6.29}$$

$$H_{1,11} = -(1 + k_{Gx})\left(\varphi_y \cos A_t \Omega \cos L - \sin A_t \Omega \cos L\right) \tag{6.30}$$

$$H_{2,1} = -(\omega(t) + \Omega \sin L) \tag{6.31}$$

$$H_{2,2} = (1 + k_{Gz}) \cos A_t \Omega \cos L - \Delta_{sz}(-\sin A_t \Omega \cos L) \tag{6.32}$$

$$\begin{aligned} H_{2,6} = {} & \Delta_{sz} \cos A_t \Omega \cos L - \Delta_{oz}(\omega(t) + \Omega \sin L) \\ & + \varphi_y \cos A_t \Omega \cos L - \sin A_t \Omega \cos L \end{aligned} \tag{6.33}$$

$$H_{2,9} = 1 \tag{6.34}$$

$$H_{2,14} = (1 + k_{Gz})(\cos A_t \Omega \cos L) - \varphi_y(-\sin A_t \Omega \cos L) \tag{6.35}$$

$$H_{2,15} = -(1 + k_{Gz})(\omega(t) + \Omega \sin L) \tag{6.36}$$

Other elements of the matrix H are zero.

6.3.2 Observability Analysis

Firstly, we will analyze the observability of the platform's misalignments φ_x, φ_y and φ_z based on the derivative matrix $\boldsymbol{H}(\boldsymbol{X})$.

1. When the platform rotating around the azimuth axis with uniform angular velocity, the coefficients for φ_x and φ_z are correlated with that of X gyro and Y gyro's fixed drifts. So we adopt a changing angular velocity during the rotation process.
2. In the matrix $\boldsymbol{H}(\boldsymbol{X})$, the coefficients for φ_y are correlated with that of Δ_{ox} and Δ_{sz}, the first-order components are $(1 + k_x) \sin A_t \Omega \cos L$ and $(1 + k_z) \cos A_t \Omega \cos L$, respectively.

Table 6.1 The singular values of the observability matrix

Rotation scheme	Rotation around the azimuth axis with an angular velocity of $(0.5 + t)^{\circ}/s$
Singular values of the observability matrix	13.38, 13.38, 0.0953, 0.0953, 0.0035, 0.0035, 0.0007, 0.00069, 0.000476, 0.00035, 0.000308, 8.35e-6, 3.03e-6, 1.70e-8, 1.32e-8

3. Because the gyro's torque scale error is very small, the coefficients for Δ_{sx} and Δ_{oz} are approximately equal to that of φ_x and φ_z, respectively.

The singular values of the observability matrix for this alignment scheme are listed in Table 6.1.

From the above table we can see that there are four singular values below 1e-5, which means there are at least four parameters' observability are low and this analysis is consistent with that of the derivative matrix analysis.

Additionally, we have to point out that the low observability of φ_x and φ_z will not affect the alignment for the horizontal axes seriously because the gimbal is locked based on the horizontal accelerometers.

6.4 Simulation and Analysis

6.4.1 Simulation

The simulation conditions are set as follows:

1. The gyros fixed drifts are $0.01^{\circ}/h(1\sigma)$; The misalignment angles of gyros are $60''(1\sigma)$; The gyro torque scale errors are $1e3\,\text{ppm}(1\sigma)$; The misalignments of platform are $\varphi_x = 20''(1\sigma), \varphi_y = 0.3^{\circ}(1\sigma), \varphi_z = 20''(1\sigma)$.
2. The measurement noise of the horizontal gyros are $0.01^{\circ}/h(1\sigma)$. The rotation angular velocity of the platform is $(0.5 + t)^{\circ}/s$ and the rotation time is 6 min.

Table 6.2 Estimation results of the parameters for the designed schemes

Parameters	Estimation errors (1σ)	Parameters	Estimation errors (1σ)
φ_x	$16.89''$	k_{g0z}	$0.0012^{\circ}/h$
φ_y	$17.23''$	Δ_{sx}	$17.72''$
φ_z	$17.73''$	Δ_{sy}	$16.45''$
k_{Gx}	13.2 ppm	Δ_{sz}	$18.11''$
k_{Gy}	12.6 ppm	Δ_{ox}	$18.70''$
k_{Gz}	23.9 ppm	Δ_{oy}	$16.68''$
k_{g0x}	$0.0046^{\circ}/h$	Δ_{oz}	$16.87''$
k_{g0y}	$0.0086^{\circ}/h$		

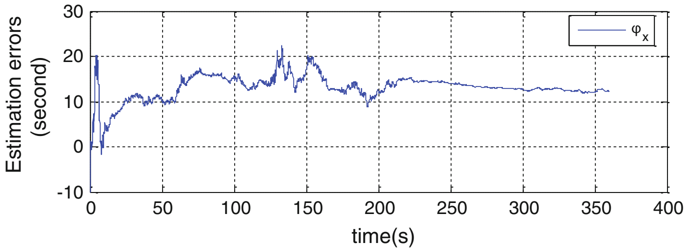


Fig. 6.2 Estimation of φ_x

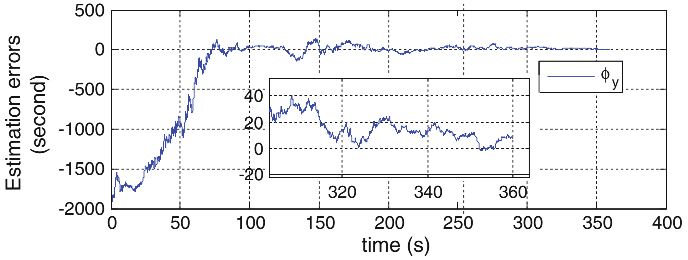


Fig. 6.3 Estimation of φ_y

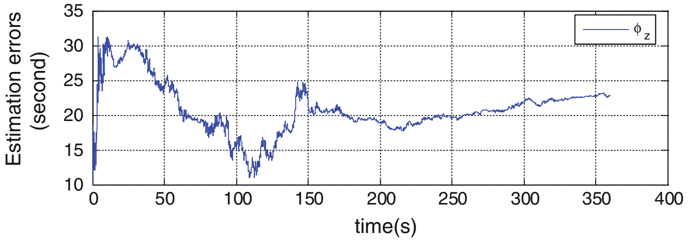


Fig. 6.4 Estimation of φ_z

The parameter estimation results are given in Table 6.2 and Figs. 6.2, 6.3 and 6.4 respectively.

Figures 6.2, 6.3 and 6.4 are the estimation results for the misalignments of the platform (the unit of estimation in the figures is arc second). From the figures we can see that the convergence times for φ_x, φ_x and φ_z are all less than 200 s. From Table 6.2 we can see that the estimation error for φ_y is less than 20 s, which can satisfy the requirement of the azimuth orientation for the carriers. The estimation errors for φ_x and φ_z are larger than 15 s because the coefficients of gyro misalignments Δ_{sx} and Δ_{oz} are correlated with that of φ_x and φ_z. Besides, we can see that the estimation errors of gyro's fixed drifts are less than $0.01^\circ/h$ through the time varying rotation angular velocity. To analyze the designed scheme more thoroughly, the two-position gyrocompass alignment method will be analyzed in the next section.

6.4.2 Compared with the Two-Position Gyrocompass Alignment Method

Denoting the outputs of X-gyro on the two positions as $\tilde{\omega}_{x1}$ and $\tilde{\omega}_{x2}$. Similarly, the outputs of Z-gyro are $\tilde{\omega}_{z1}$ and $\tilde{\omega}_{z2}$. Based on the two-position gyrocompass alignment method, the estimation for the misalignment of azimuth axis of the platform can be expressed by

$$\hat{\varphi}_y = \Delta_{ox} + \frac{\cos(A + \Delta A) \cdot (1 + k_{Gx})}{\sin(A + \Delta A)} - \frac{\tilde{\omega}_{x1} - \tilde{\omega}_{x2}}{2\sin(A + \Delta A) \cdot \Omega} \tag{6.37}$$

or

$$\hat{\varphi}_y = \Delta_{sz} - \frac{\tilde{\omega}_{z1} - \tilde{\omega}_{z2}}{2\cos(A + \Delta A) \cdot \Omega} - \frac{\sin(A + \Delta A) \cdot (1 + k_{Gz})}{\cos(A + \Delta A)} \tag{6.38}$$

where ΔA is the gimbal locking error.

From the above equations we can see obviously that gyro's misalignment angles, gyro torque scale errors and locking errors affect the alignment precision. Besides we can see that the initial azimuth angle also affect the alignment precision because it is coupled with other error sources. Compared with this traditional gyrocompass alignment method, the proposed alignment method has the following advantages:

1. Locking errors and the effect of initial azimuth angle are eliminated because the continuous rotation around the azimuth axis is adopted;
2. The gyro torque scale errors are estimated in the proposed method, so they will not affect the alignment precision.
3. Because two gyro's torque currents are both used in the proposed method, the error caused by Δ_{ox} and Δ_{sz} are not more than $\min(|\Delta_{sz}|, |\Delta_{ox}|)$.

Based on the analysis above, we can conclude that the proposed method is able to provide higher precision alignment results than the traditional gyrocompass alignment method.

6.5 Conclusions

In the paper, a study on the scheme design and analysis of the initial alignment of PINS is presented. A new gyrocompass alignment method based on the continuous rotation around the platform's azimuth axis is proposed. Based on the developed dynamic model and observation model, the observability of the platform's misalignment angles is analyzed. At last, the simulations are conducted to analyze the performance of the proposed method. It is shown that the initial alignment can be quickly carried out with high precision and the proposed method has more advantages than traditional gyrocompass alignment method.

References

1. Gong NY, Feng ZS (1980) Missile inertial platform self-alignment. J Natl Univ Def Technol 2 (1):173–184
2. Ren HW, Tang ZW (2010) Fast alignment method for platform inertial navigation system. J Chin Inert Technol 18(3):279–282
3. Zhao Z (1995) New scheme for reducing alignment time of inertial platform of Chinese aircraft. J Northwes Polytech Univ 13(3):432–435
4. Yang HB, Zhang SF, Cai H (2006) Continuous calibration and alignment techniques for a high precision inertial platform. J Astronaut 27(4):600–604
5. Jackson AD (1973) Continuous calibration and alignment techniques for an all-attitude inertial platform. In: AIAA guidance and control conference, Key Biscayne, Florida. AIAA Paper No. 73-865
6. Meskin G, Bar-Itzhack IY (1990) Observability analysis of piece-wise constant systems with application to inertial navigation. In: Proceedings of the 29th conference on decision and control, Honolulu, Hawaii, pp 821–826
7. Fang JC, Zhou R, Zhu SP (1999) Observability analysis of strapdown inertial navigation systemon moving base. J Beijing Univ Aeronaut Astronaut 25(6):714–719
8. Liu BQ, Fang JC (2008) A new adaptive feedback Kalman filter based on observability analysis for SINS/GPS. Acta aeronouticaet astronautica sinica 29(2):430–436

Chapter 7
Earth Polar Motion Parameters High Accuracy Differential Prediction

Lue Chen, Geshi Tang, Songjie Hu, Ming Chen, Yifei Li, Mei Wang, Jing Sun, Ming Shi and Li Li

Abstract This paper proposes a method of Earth polar motion parameters prediction by dual differential least-squares (LS) and autoregressive (AR) model. Firstly, polar motion parameters are processed by dual differential method, the stationarity of polar motion parameters is improved. Then, LS+AR method is utilized to analyze the dual differential polar motion parameters to obtain the preliminary prediction results. Finally, the preliminary prediction results are processed by inverse dual differential method to obtain high accuracy polar motion prediction results. The prediction results are compared with EOP prediction comparison campaign (EOP_PCC) results, it shows that the short-term polar motion parameters prediction error is at the same level of EOP_PCC. The one day prediction accuracy of PMX is at the level of 0.25 mas, PMY is 0.2 mas, they are better than EOP_PCC one day polar motion prediction accuracy.

Keywords Earth orientation parameters · Polar motion parameters prediction · Dual differential · Least-squares · AR model

7.1 Introduction

High accuracy Earth orientation parameters (EOP) is the basic parameters for conversion between the International Celestial Reference Frame (ICRF) and the International Terrestrial Reference Frame (ITRF) [1].

EOP has been monitored with increasing accuracy by advanced space-geodetic techniques, including lunar and satellite laser ranging (SLR), very long baseline

---

L. Chen (✉) · G. Tang · S. Hu · M. Chen · Y. Li · M. Wang · J. Sun · M. Shi · L. Li
National Key Laboratory of Science and Technology on Aerospace Flight Dynamic, Beijing 100094, China
e-mail: luechen0912@163.com

L. Chen · G. Tang · S. Hu · M. Chen · Y. Li · M. Wang · J. Sun · M. Shi · L. Li
Beijing Aerospace Control Center, Beijing 100094, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_7

interferometry (VLBI) and Global Navigation Satellite System(GNSS), etc. [2, 3]. The high accuracy EOP is essential for manned space flight and deep space exploration mission, especially for the real time and high accuracy navigation mission. EOP is usually available with a delay of hours to days, thus, EOP prediction is adapt to the growing demands for spacecraft navigation and physical geography science research.

Polar motion prediction, UT1-UTC prediction, and length of day (LOD) prediction are the hot research on EOP prediction. Many methods have been developed and applied to EOP predictions, such as LS extrapolation [4], LS extrapolation and AR prediction (LS+AR) [5], networks prediction [6], spectral analysis and least-squares extrapolation [7], wavelet decomposition and auto-covariance prediction [8], etc.

The Earth orientation parameters prediction comparison campaign (EOP_PCC) that started in 2005 was organized for the purpose of assessing the accuracy of EOP predictions. By contrast, LS+AR prediction method is one of the highest accuracy prediction methods. However, when LS+AR prediction method is utilized, the key problems are the selection of parameters in LS extrapolation, the best order determination in AR model, AR prediction of non-stationary EOP series, etc.

At present, the existing LS+AR prediction method in the study of data stationary requirements is lack. To a certain extent, this will affect the EOP prediction accuracy more or less.

This paper researches polar motion parameters prediction starting with data's stationary analysis, and proposes dual differential LS+AR prediction method to obtain high accuracy polar motion parameters short-term prediction results.

7.2 Theoretical Method

7.2.1 Least-squares

Least-squares model of polar motion prediction is shown in Formula (7.1), it contains linear term and periodic term. The periodic term contains Chandler wobbles, annual, half of a year, third of a year, etc.

$$\begin{aligned} X(t) = A + Bt + Ct^2 + D_1 \cos\left(\frac{2\pi t}{p_1}\right) + D_2 \sin\left(\frac{2\pi t}{p_1}\right) \\ + E_1 \cos\left(\frac{2\pi t}{p_2}\right) + E_2 \sin\left(\frac{2\pi t}{p_2}\right) + \cdots \end{aligned} \tag{7.1}$$

where, t is UTC time (unit is year). $A, B, C, D_1, D_2, E_1, E_2, \ldots$are the fitting parameters, $p_1, p_2, \ldots$ are the fitting periods, which could be determined by prior experience.

7.2.2 AR Model

For a stationary sequence $x_t(t = 1, 2, \ldots, N)$, the AR model is expressed as follows

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + a_t = \sum_{i=1}^{P} \varphi_i x_{t-i} + a_t \tag{7.2}$$

where $\varphi_1, \varphi_2, \ldots, \varphi_p$ are model parameters, a_t is white noise, p is model order. Formula (7.2) is called p order AR model, denoted by AR(p). $a_t \sim N\left(0, \sigma_n^2\right)$, σ_n^2 is the variance of the white noise.

The key technique of AR model is determining model order parameter p. There are many criterions which can be utilized for determining the model order parameter p, such as Final Prediction Error Criterion (FPE), Akake Information Criterion (AIC), Singular Value Decomposition (SVD) criterion, etc.

This study utilizes FPE for determining AR model order. FPE criterion function is as follows.

$$FPE(p) = \frac{N+p}{N-p} \sigma_n^2 \tag{7.3}$$

7.2.3 Prediction Error Estimates

In order to evaluate prediction error, Mean absolute error (MAE) is utilized as the prediction accuracy index shown as follows.

$$MAE_i = \frac{1}{n} \sum_{j=1}^{n} \left(\left| p_j^i - o_j^i \right| \right) \tag{7.4}$$

where o is the real observation, p is prediction value, i is prediction day, n is prediction number.

7.3 Dual Differential LS+AR Prediction Process

The process of the dual differential LS+AR polar motion parameters prediction is shown in Fig. 7.1.

7.4 Calculation and Analysis

Polar motion parameters come from IERS website for prediction and accuracy verification. EOP 05C04 data is selected for comparing with EOP_PCC results. The analyzed data is from January 1, 2000 to December 31, 2009. One day has one

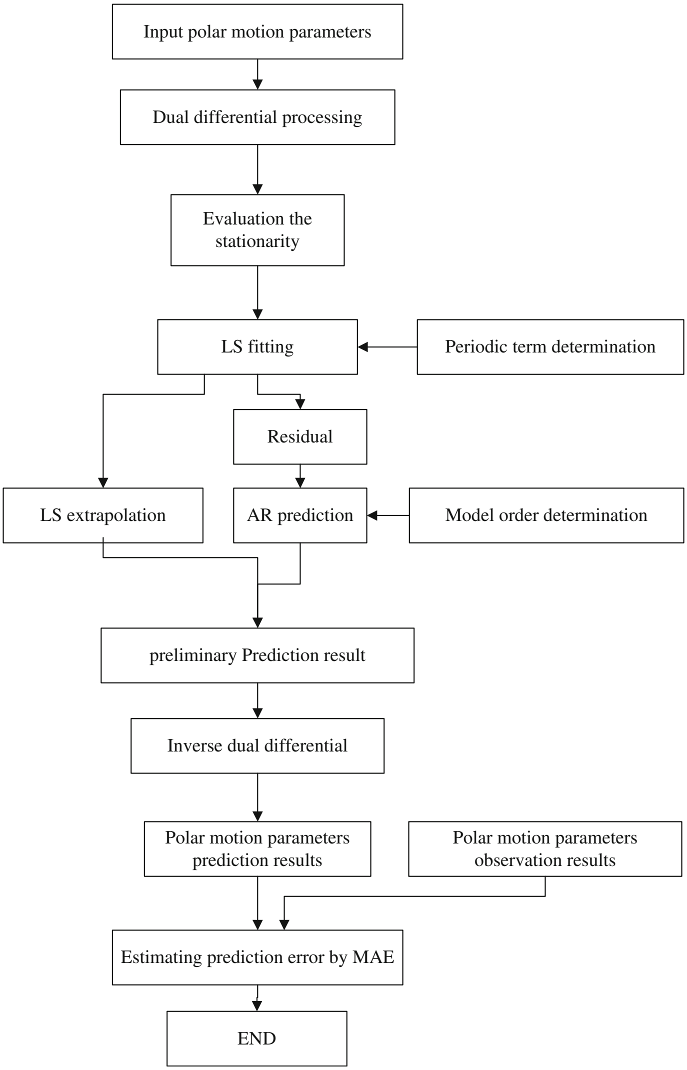


Fig. 7.1 Dual differential LS+AR prediction process

polar motion parameter. The basic polar motion sequences are from January 1, 2000 to December 31, 2007. Dual differential LS+AR method is utilized to predict polar motion parameters from January 1, 2008 to December 31, 2009. And the polar motion parameters prediction values are compared with the real observation values. The prediction days are from 1 to 30 days, the prediction number is 731 (corresponding to two years). The prediction results are shown as follows.

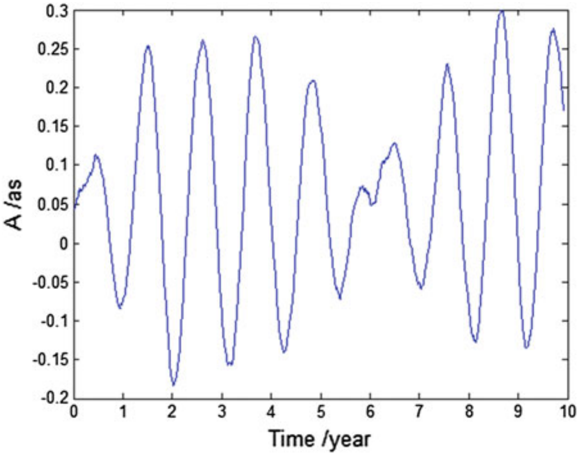


Fig. 7.2 PMX of from 2000 to 2009

Figure 7.2 shows the original polar motion component X orientation (PMX) from 2000 to 2009. Figure 7.3 shows the direct LS fitting result of PMX from 2000 to 2009. In Fig. 7.3, it can be found that the residual after LS fitting shows obvious fluctuation, and it is non-stationary. Thus, this study utilizes dual differential method for preliminarily processing polar motion parameters, then LS method is utilized to fit the differential polar motion parameters, the results are shown in Fig. 7.4. It shows that the differential results are more stationary.

In the same way, PMY from 2000 to 2009 is shown in Fig. 7.5. The LS fitting dual differential PMY results are shown in Fig. 7.6.

Then, utilizing LS extrapolation, residual AR prediction and inverse dual differential, the one day polar motion prediction value can be obtained. In the same way, different prediction days prediction value can be obtained. The prediction

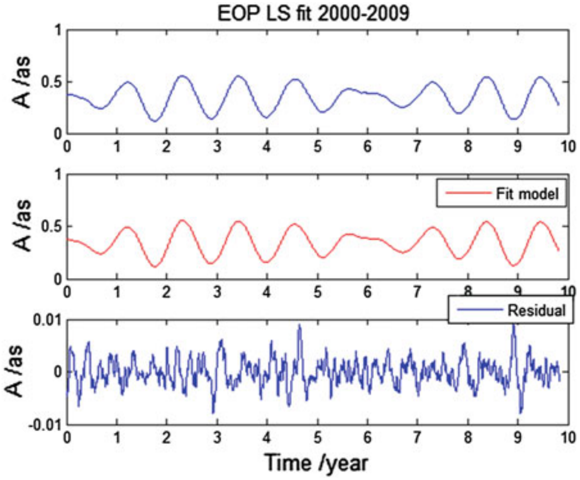


Fig. 7.3 LS fitting result of PMX from 2000 to 2009

Fig. 7.4 LS fitting result of PMX dual differential from 2000 to 2007

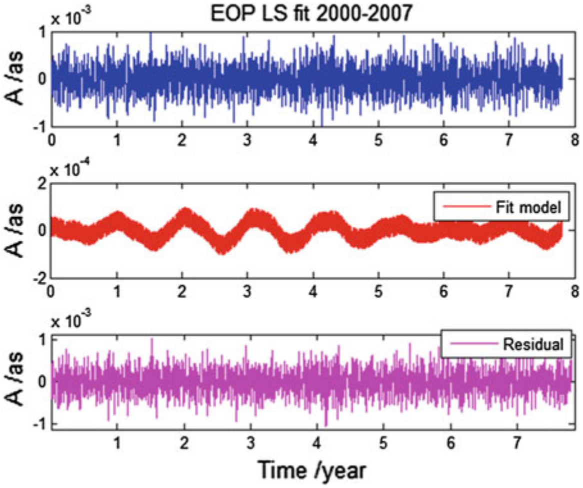


Fig. 7.5 PMY from 2000 to 2009

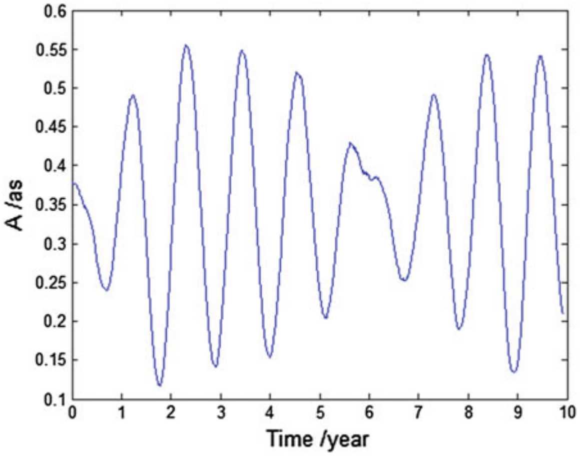


Fig. 7.6 LS fitting result of PMY dual differential from 2000 to 2007

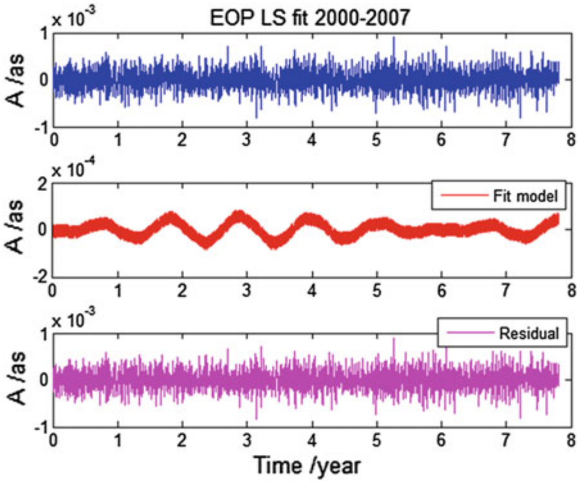


Table 7.1 MAE of polar motion short term prediction

Prediction day	PMX error(as)	PMY error(as)
1	0.000255753	0.000201034
2	0.000602844	0.000481510
3	0.000976849	0.000782508
4	0.001384357	0.001105586
5	0.001796136	0.001428774
6	0.002199790	0.001749022
7	0.002581433	0.002057742
8	0.002952354	0.002358710
9	0.003327655	0.002667279
10	0.003713629	0.002989965
11	0.004112756	0.003325439
12	0.004519159	0.003658391
13	0.004943598	0.003992372
14	0.005385053	0.004340725
15	0.005828831	0.004699438
16	0.006274278	0.005058480
17	0.006719786	0.005407787
18	0.007157375	0.005750121
19	0.007594186	0.006092389
20	0.008034826	0.006432489
21	0.008475994	0.006783072
22	0.008926775	0.007140763
23	0.009389878	0.007513585
24	0.009866693	0.007900196
25	0.010361002	0.008295547
26	0.010858821	0.008695167
27	0.011348414	0.009085973
28	0.011835469	0.009478140
29	0.012335856	0.009890083
30	0.012851078	0.010306225

number is 731 (corresponding to two years). MAE is utilized to evaluate the prediction accuracy, and the result is shown in Table 7.1.

Earth Orientation Parameters Prediction Comparison Campaign attracted 12 participants coming from 8 countries, who are the top professors or scholars in the time sequence analyzing filed. EOP_PCC referenced to more than 20 prediction methods. EOP_PCC contains the ultra short term (predictions to 10 days into the future), short term (30 days), and medium term (500 days) predictions.

This paper compared our results (BACC results) with EOP_PCC results in ultra short term prediction and short term prediction. Figure 7.7 shows the EOP_PCC results, Fig. 7.8 shows the BACC results.

By comparison, BACC prediction results of polar motion are at the same level as EOP_PCC. According to one day prediction accuracy, BACC one day polar motion prediction accuracy is better than EOP_PCC. In BACC prediction, one day prediction error of PMX is at the level of 0.25 mas, PMY is at the level of 0.2 mas. EOP_PCC polar motion minimum one day predictions error is at the level of 0.5 mas for PMX, 0.35 mas for PMY [1].

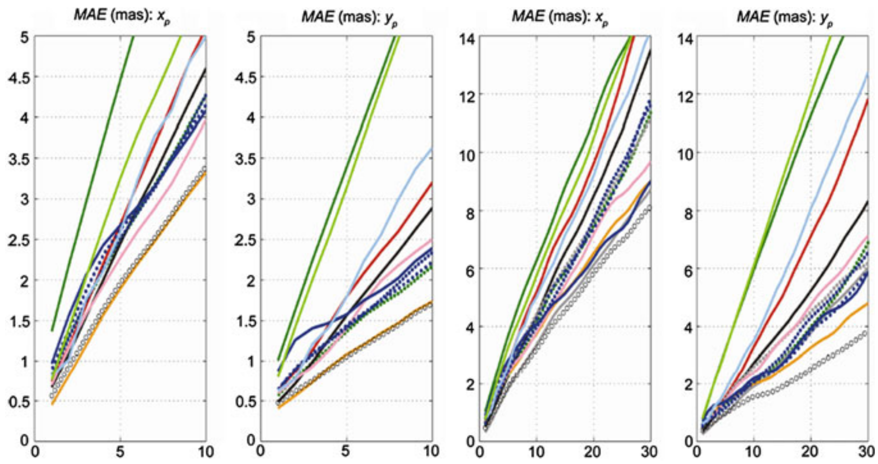


Fig. 7.7 EOP_PCC prediction results of polar motion

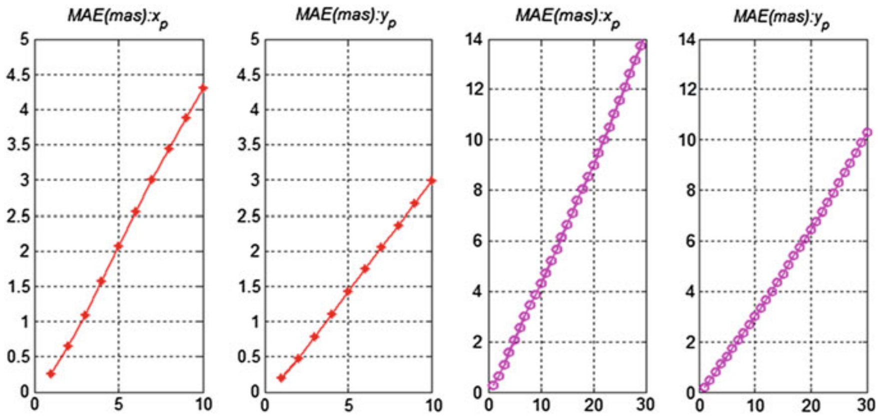


Fig. 7.8 BACC prediction results of polar motion

7.5 Conclusion

The paper proposes dual differential LS+AR prediction method for polar motion prediction. The Earth polar motion parameters from IERS are utilized for prediction by dual differential LS+AR method. By comparison, this paper's prediction accuracy is at the same level as EOP_PCC in ultra short term prediction and short term prediction. The one day prediction error of PMX is at the level of 0.25 mas, PMY is at the level of 0.2 mas in our study, this is better than EOP_PCC result. In the following work, the proposed prediction method can be used to UT1-UTC and LOD prediction.

Acknowledgments We thank the International Earth Rotation and Reference Systems Service (IERS) for providing the EOP data.

References

1. Kalarus M, Schuh H, Kosek W et al (2010) Achievements of the earth orientation parameters prediction comparison campaign. J Geodyn 62:587–596
2. Karbon M, Nilsson T, Schuh H (2013) Determination and prediction of EOP using Kalman filtering. Geophys res 15
3. Xu XQ, Zhou YH (2010) High precision prediction method of earth orientation parameters. J spacecraft TT&C technol 29(2):70–76
4. Xu X, Zhou Y, Liao X (2012) Short-term earth orientation parameters predictions by combination of the least-squares, AR model and Kalman filter. J Geodyn 62:83–86
5. Guo JY, Li YB, Dai CL et al (2013) A technique to improve the accuracy of earth orientation prediction algorithms based on least squares extrapolation. J Geodyn 70:36–48
6. Wang QJ, Liao DC, Zhou YH (2008) Real-time rapid prediction of variations of earth's rotational rate. Chin Sci Bull 53:969–973
7. Akulenko LD, Kumakshev SA, Rykhlova LV (2002) A model for the polar motion of the deformable earth adequate for astrometric data. Astron Rep 46:74–82
8. Kosek W, Kalarus M, Johnson TJ et al (2005) A comparison of LOD and UT1-UTC forecasts by different combination prediction techniques. Artif Satell 40:119–125

Chapter 8
Real-Time Magnetometer-Bias Calibration of Micro-satellite Without Attitude Information

Zhen Zhang, Jianping Xiong and Jin Jin

Abstract Due to the disadvantages in present magnetometer bias determination methods such as complex calculation, low accuracy of estimation and non real-time, a differential value method is developed for magnetometer bias estimation of micro-satellite without attitude information. This method avoids the impact of quartic nature and use the iterative method to satisfy real-time applications. Simulation results indicate that the accuracy of estimation using the differential value method can be less than 50 nT, which is more accurate compared with other methods and has no need to set an initial state. This method will reduce the demand for computing power effectively and improve the accuracy, while the magnetometer bias calibration of micro-satellite is being carried out on-orbit.

Keywords Magnetometer bias · Calibration · Satellite · Attitude-independent

8.1 Introduction

Micro-satellites work mainly in low-Earth orbit. To detect the ambient magnetic field strength, it is often equipped with a three-axis magnetometer, for the following four main reasons: first, geomagnetic field model which possesses a rather high degree of accuracy such as IGRF11 [1] has already been established in low earth orbit space. Hence the use of magnetometer alone or in combination with other attitude sensors will suffice to determine the attitude of the satellite; second, for micro-satellite equipped with magnetorquers as its attitude control actuator, magnetometer is indispensable for the determination of geomagnetic field strength; third, in conjunction with other attitude sensors, it can determine the remanence moment of the entire satellite; forth, magnetometer also has the merits of being light

---

Z. Zhang (✉) · J. Xiong
Department of Precision Instrument, Tsinghua University, Beijing 100084, China
e-mail: zhangzhen05@tsinghua.org.cn

J. Jin
Space Center, Tsinghua University, Beijing 100084, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_8

in weight, small in size, easy to install, low in cost and power consumption etc. In light of the reasons mentioned above, magnetometer has gained wide application in attitude determination and control system of micro-satellite.

However, influenced by the micro-satellite remanence and other factors such as the installation error, instrument error and measurement noise, magnetometer will inevitably have some errors when used for attitude determination. For micro-satellite which does not employ special technology to reduce the remanence of the whole satellite, errors caused by remanence will be much larger than other factors [2, 3]. When working in steady state, the internal equipment unit of micro-satellite will generate a stable magnetic field in conjunction with permanent magnet, which could be viewed as the approximately constant bias of the magnetometer measured value within a certain time. This bias and the constant bias of the magnetometer jointly constitute the magnetometer constant bias, often referred to as the constant bias for short [3]. Numerous studies have been carried out by related scholars about the on-orbit estimation of the constant bias, and different methods have also been developed by them.

The joint use of magnetometer and other attitude sensors could obtain the real-time estimation of the constant bias through extended Kalman filter (EKF). Reference [4] provides a method to determine the constant bias through unscented Kalman filter (UKF), using star sensor, fiber optical gyro, sun sensor and magnetometer as sensors. According to its study, the accuracy of estimation could reach up to 50 nT in 660 km sun-synchronous orbit. Reference [3] also presents a method to estimate the constant bias through EKF with sun sensor, earth sensor, gyroscope and magnetometer as its sensors. This kind of methods need additional attitude sensor to provide the attitude information of the satellite, which makes it impossible to determine the constant bias by using the magnetometer alone, and the accuracy of estimation is also related to the attitude information acquired. Besides, there is a close correlation between the initial state and the performance of EKF, thus the setting of an appropriate initial state is required.

The use of magnetometer alone can also determine the constant bias with the help of the maximum likelihood estimation method. Roberto Alonso and Malcolm D. Shuster have introduced a fast and stable estimation algorithm named TWO-STEP and also proved that this method is superior to other statistical algorithms [5]. However, as a statistical algorithm, it usually involves dealing with a large number of sample data in order to ensure the quality of estimation, which has increased the burden of the satellite attitude determination and control system. Reference [4] talks about an offline working mode which downloads data collected by micro-satellite to the ground and then uploads the results back to micro-satellite after being processed on the ground. This will not only take up the communication resources, but also impair the autonomous operation of the micro-satellite. In addition, statistical algorithm also has the drawback of non real-time.

In order to solve the problems of large computing and non real-time associated with the statistical algorithm, Ref. [6] introduces a recursive algorithm based on Centered statistical algorithm and Ref. [7] presents another one based on Kalman filter.

Based on the analysis of the principles in the estimation of constant magnetometer bias by statistical algorithm, this paper develops a differential value method for on-orbit real-time estimation, and realizes the real-time recursive calculation of the algorithm as well. From the comparison of the differential value recursive algorithm, the Centered recursive algorithm and the extended Kalman estimation (EKF), we can find that the differential value recursive algorithm discussed in the paper is a rather satisfactory one, with such advantages as high accuracy of estimation, real-time computing ability and needlessness of initial conditions setting.

8.2 Measurement Model and Base of Estimation Algorithms

8.2.1 Measurement Model

With strict calibration, the installation errors and instrument errors are not the master part compared to the remanence. So, we mainly focus on the magnetometer bias caused by steady remanence. The measurement errors of magnetometer are discussed in somewhat more detail in Ref. [2]. The measurement model is [5]

$$\boldsymbol{B}_{b_k} = \boldsymbol{A}_k \boldsymbol{B}_{o_k} + \boldsymbol{b} + \boldsymbol{\varepsilon}_k, \quad k = 1, \ldots, N \tag{8.1}$$

where $\boldsymbol{B}_b$ is the measurement of the magnetic field by the magnetometer; $\boldsymbol{B}_o$ is the corresponding value of the geomagnetic field with respect to an orbit coordinate system; $\boldsymbol{A}$ is the attitude matrix of the magnetometer with respect to the orbit coordinates; $\boldsymbol{b}$ is the magnetometer bias and $\boldsymbol{\varepsilon}$ is the measurement noise. The k denotes time t_k. So that at each time

$$|\boldsymbol{B}_{o_k}|^2 = |\boldsymbol{A}_k \boldsymbol{B}_{o_k}|^2 = |\boldsymbol{B}_{b_k} - \boldsymbol{b} - \boldsymbol{\varepsilon}_k|^2 \tag{8.2}$$

We can define

$$z_k \equiv |\boldsymbol{B}_{b_k}|^2 - |\boldsymbol{B}_{o_k}|^2, \quad v_k \equiv 2(\boldsymbol{B}_{b_k} - \boldsymbol{b}) \cdot \boldsymbol{\varepsilon}_k - |\boldsymbol{\varepsilon}_k|^2 \tag{8.3}$$

then

$$z_k = 2\boldsymbol{B}_{b_k} \cdot \boldsymbol{b} - |\boldsymbol{b}|^2 + v_k, \quad k = 1, \ldots, N \tag{8.4}$$

For the assumption that the magnetometer measurement noise is white and Gaussian with covariance matrix $\boldsymbol{C}_k$, we can get

$$\boldsymbol{\varepsilon}_k \sim N(0, \boldsymbol{C}_k), \quad \mathrm{E}\{\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_l^{\mathrm{T}}\} = 0, \quad k \neq l \tag{8.5}$$

where $\mathrm{E}\{\cdot\}$ denotes the expectation. It follows that

$$u_k \equiv \mathrm{E}\{v_k\} = -\mathrm{tr}(\boldsymbol{C}_k) \tag{8.6}$$

$$\sigma_k^2 \equiv \mathrm{E}\{v_k^2\} - u_k^2 = 4(\boldsymbol{B}_{b_k} - \boldsymbol{b})^{\mathrm{T}} \boldsymbol{C}_k (\boldsymbol{B}_{b_k} - \boldsymbol{b}) + 2(\mathrm{tr}\boldsymbol{C}_k^2) \tag{8.7}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace operation and v_k satisfies [5]

$$v_k \sim N(u_k, \sigma_k^2), \quad \mathrm{E}\{v_k v_l\} = u_k u_l,\ k \neq l \tag{8.8}$$

8.2.2 Maximum-Likelihood Estimation

With the samples at different moments, we can get the estimation of magnetometer bias via maximum-likelihood estimation. The negative-log-likelihood function is given by

$$J(\boldsymbol{b}) = \frac{1}{2}\sum_{k=1}^{N}\left[\frac{1}{\sigma_k^2}\left(z_k - 2\boldsymbol{B}_{b_k}\cdot\boldsymbol{b} + |\boldsymbol{b}|^2 - u_k\right)^2\right] + \frac{1}{2}\sum_{k=1}^{N}\left[\log\sigma_k^2 + \log 2\pi\right] \tag{8.9}$$

The value which minimizes the function $J(\boldsymbol{b})$ is the maximum-likelihood estimate of magnetometer bias (noted by $\boldsymbol{b}^*$). So $\boldsymbol{b}^*$ must satisfy

$$\left.\frac{\partial J}{\partial \boldsymbol{b}}\right|_{\boldsymbol{b}^*} = 0 \tag{8.10}$$

To solve the Eq. (8.10), we can use Newton-Raphson approximation. The sequence is

$$\boldsymbol{b}_{i+1}^{NT} = \boldsymbol{b}_i^{NT} - \left[\frac{\partial^2 J}{\partial\boldsymbol{b}\partial\boldsymbol{b}^T}\left(\boldsymbol{b}_i^{NT}\right)\right]^{-1}\frac{\partial J}{\partial\boldsymbol{b}}\left(\boldsymbol{b}_i^{NT}\right), \quad i = 0, 1, \ldots \tag{8.11}$$

$$\boldsymbol{b}_0^{NT} = 0 \tag{8.12}$$

Since the quartic nature of $J(\boldsymbol{b})$ leads to several minima and maxima and it would take a multistage iterative process to achieve a satisfactory result, this method is not the best choice.

8.2.3 Centered Estimation

Reference [8] presents an estimation algorithm named Centered to avoid the impact of quartic nature in $J(\boldsymbol{b})$. The main steps are introduced as follows.

Firstly, It defines four weighted averages ($\bar{z}$, $\bar{\boldsymbol{B}}_b$, $\bar{v}$, $\bar{u}$) and four centered variables ($\tilde{z}_k$, $\tilde{\boldsymbol{B}}_{b_k}$, $\tilde{v}_k$, $\tilde{u}_k$) (see Ref. [8] for more details).

Then we can get the negative-log-likelihood function as

$$\begin{aligned}\tilde{J}(\boldsymbol{b}) = & \frac{1}{2}\sum_{k=1}^{N}\left[\frac{1}{\sigma_k^2}\left(\tilde{z}_k - 2\tilde{\boldsymbol{B}}_{b_k}\cdot\boldsymbol{b} - \tilde{u}_k\right)^2\right] \\ & + \frac{1}{2}\left\{\sum_{k=1}^{N}\left[\log\sigma_k^2 + \log 2\pi\right] - \left[\log\bar{\sigma}^2 + \log 2\pi\right]\right\}\end{aligned} \tag{8.13}$$

Equation (8.13) is no longer quartic and the estimate from Centered algorithm is

$$\tilde{\boldsymbol{b}}^* = \tilde{\boldsymbol{P}}_{bb}\sum_{k=1}^{N}\left[\frac{1}{\sigma_k^2}(\tilde{z}_k - \tilde{u}_k)2\tilde{\boldsymbol{B}}_{b_k}\right] \tag{8.14}$$

where

$$\tilde{\boldsymbol{P}}_{bb} = \tilde{\boldsymbol{F}}_{bb}^{-1} = \left[\sum_{k=1}^{N}\frac{1}{\sigma_k^2}4\tilde{\boldsymbol{B}}_{b_k}\tilde{\boldsymbol{B}}_{b_k}^{\mathrm{T}}\right]^{-1} \tag{8.15}$$

$\tilde{\boldsymbol{P}}_{bb}$ is the covariance matrix of $\tilde{\boldsymbol{b}}^*$.

Centered algorithm converts the negative-log-likelihood function to quadratic and this leads to a simple batch solution. But the disadvantage of this method is that the accuracy of estimation would become poor when the magnetometer bias is much smaller compared to the geomagnetic field [5].

8.2.4 TWOSTEP Estimation

TWOSTEP is a fast robust algorithm. It uses the centered approximation to provide the initial estimate, before Gauss-Newton method is being introduced to continue the estimation as the second step. The estimate of TWOSTEP is

$$\hat{\boldsymbol{b}}^* = \tilde{\boldsymbol{b}}^* - \boldsymbol{P}_{bb}\frac{\partial J}{\partial\boldsymbol{b}}\left(\tilde{\boldsymbol{b}}^*\right) \tag{8.16}$$

where $\boldsymbol{P}_{bb}$ is the covariance matrix and the values of these parameters are

$$\boldsymbol{P}_{bb} = \boldsymbol{F}_{bb}^{-1} = \left(\tilde{\boldsymbol{F}}_{bb} + \bar{\boldsymbol{F}}_{bb}\right)^{-1} \tag{8.17}$$

$$\frac{\partial J}{\partial \boldsymbol{b}}(\boldsymbol{b}) = \frac{\partial \tilde{J}}{\partial \boldsymbol{b}}(\boldsymbol{b}) + \frac{\partial \bar{J}}{\partial \boldsymbol{b}}(\boldsymbol{b}) \tag{8.18}$$

$$\bar{\boldsymbol{F}}_{bb} = \frac{4}{\bar{\sigma}^2}(\bar{\boldsymbol{B}}_b - \boldsymbol{b})(\bar{\boldsymbol{B}}_b - \boldsymbol{b})^{\mathrm{T}} \tag{8.19}$$

$$\frac{\partial \tilde{J}}{\partial \boldsymbol{b}}(\boldsymbol{b}) = -\sum_{k=1}^{N}\left[\frac{1}{\sigma_k^2}\left(\tilde{z}_k - 2\tilde{\boldsymbol{B}}_{b_k} \cdot \boldsymbol{b} - \tilde{u}_k\right)2\tilde{\boldsymbol{B}}_{b_k}\right] \tag{8.20}$$

$$\frac{\partial \bar{J}}{\partial \boldsymbol{b}}(\boldsymbol{b}) = -\frac{1}{\bar{\sigma}^2}\left(\bar{z} - 2\bar{\boldsymbol{B}}_b \cdot \boldsymbol{b} + |\boldsymbol{b}|^2 - \bar{u}\right)2(\bar{\boldsymbol{B}}_b - \boldsymbol{b}) \tag{8.21}$$

TWOSTEP is proved to be a fast and robust algorithm [5]. But to reach a sufficient accuracy of estimation, it needs to deal with a large amount of data at a time. Such as sampling once per 0.25 with a period of 24,000 s, the number of data points can be 96,000. It would be a large load for micro-satellites. In addition, as a batch algorithm, it needs to store the measured data for computing. Unfortunately, we always want to get the estimate in real time for in-flight computing. So it is necessary to develop an effective and robust sequential process.

8.2.5 Centered Iterative Algorithm

Reference [6] presents a iterative algorithm based on Centered algorithm to solve the disadvantages of non real-time. The main process is as follows.

First, the iteration of weighted averages can be expressed as

$$\bar{z}_{N+1} = \frac{1}{\sigma_{N+1}^2 + \bar{\sigma}_N^2}\left(\sigma_{N+1}^2 \bar{z}_N + \bar{\sigma}_{N+1}^2 z_{N+1}\right) \tag{8.22}$$

$$\bar{\boldsymbol{B}}_{b_{N+1}} = \frac{1}{\sigma_{N+1}^2 + \bar{\sigma}_N^2}\left(\sigma_{N+1}^2 \bar{\boldsymbol{B}}_{b_N} + \bar{\sigma}_{N+1}^2 \boldsymbol{B}_{b_{N+1}}\right) \tag{8.23}$$

$$\bar{u}_{N+1} = \frac{1}{\sigma_{N+1}^2 + \bar{\sigma}_N^2}\left(\sigma_{N+1}^2 \bar{u}_N + \bar{\sigma}_{N+1}^2 u_{N+1}\right) \tag{8.24}$$

where

$$\frac{1}{\bar{\sigma}_{N+1}^2} = \frac{1}{\bar{\sigma}_N^2} + \frac{1}{\sigma_{N+1}^2} \tag{8.25}$$

From Eq. (8.15), it can be represented as

$$\begin{aligned}\tilde{\boldsymbol{P}}^{-1}_{bb_{N+1}} &= \tilde{\boldsymbol{F}}_{bb_{N+1}} = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} 4\tilde{\boldsymbol{B}}_{b_k}\tilde{\boldsymbol{B}}^{\mathrm{T}}_{b_k} + \frac{1}{\sigma_{N+1}^2} 4\tilde{\boldsymbol{B}}_{b_{N+1}}\tilde{\boldsymbol{B}}^{\mathrm{T}}_{b_{N+1}} = \tilde{\boldsymbol{F}}_{bb_N} + \frac{1}{\sigma_{N+1}^2} 4\tilde{\boldsymbol{B}}_{b_{N+1}}\tilde{\boldsymbol{B}}^{\mathrm{T}}_{b_{N+1}} \\ &= \tilde{\boldsymbol{P}}^{-1}_{bb_N} + \frac{1}{\sigma_{N+1}^2} 4\tilde{\boldsymbol{B}}_{b_{N+1}}\tilde{\boldsymbol{B}}^{\mathrm{T}}_{b_{N+1}}\end{aligned} \tag{8.26}$$

Then, from the matrix inversion lemma, the following sequential formulation is developed

$$\tilde{\boldsymbol{b}}^*_{N+1} = K_N\tilde{\boldsymbol{b}}^*_N + \frac{1}{\sigma_{N+1}^2}(\tilde{z}_{N+1} - \tilde{u}_{N+1})2\tilde{\boldsymbol{P}}_{bb_{N+1}}\tilde{\boldsymbol{B}}_{b_{N+1}} \tag{8.27}$$

$$\tilde{\boldsymbol{P}}_{bb_{N+1}} = K_N\tilde{\boldsymbol{P}}_{bb_N} \tag{8.28}$$

where

$$K_N \equiv \boldsymbol{E}_{3\times3} - \tilde{\boldsymbol{P}}_{bb_N}\tilde{\boldsymbol{B}}_{b_{N+1}}\left(\tilde{\boldsymbol{B}}^{\mathrm{T}}_{b_{N+1}}\tilde{\boldsymbol{P}}_{bb_N}\tilde{\boldsymbol{B}}_{b_{N+1}} + \frac{1}{4}\sigma_{N+1}^2\right)^{-1}\tilde{\boldsymbol{B}}^{\mathrm{T}}_{b_{N+1}} \tag{8.29}$$

$\boldsymbol{E}_{3\times3}$ is the identity matrix.

8.3 Differential Value Iterative Algorithm

8.3.1 Basic Theory

The differential value iterative algorithm is an expansion of Acuña's algorithm mentioned in Ref. [8]. Its core idea is carrying out the difference calculation of magnetometer measurement at different times, which can avoid the impact of quartic nature, and further repeating the computing to meet the requirement of on-orbit real time application.

We can define

$$z_{d_k} \equiv z_{2k} - z_{2k-2l+1} = 2(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}}) \cdot \boldsymbol{b} + \Delta_{d_k} \tag{8.30}$$

where

$$\varDelta_{d_k} = v_{2k} - v_{2k-2l+1} \tag{8.31}$$

If the measurement noise $\boldsymbol{\varepsilon}_k$ at different time meet the same distribution characteristics, we can get

$$\varDelta_{d_k} \sim N(0, \sigma_{d_k}^2) \tag{8.32}$$

$$\sigma_{d_k}^2 = 4(\boldsymbol{B}_{b_k} - \boldsymbol{b})^{\mathrm{T}} \boldsymbol{C} (\boldsymbol{B}_{b_k} - \boldsymbol{b}) + 4\left(\boldsymbol{B}_{b_{2k-2l+1}} - \boldsymbol{b}\right)^{\mathrm{T}} \boldsymbol{C} \left(\boldsymbol{B}_{b_{2k-2l+1}} - \boldsymbol{b}\right) + 4(\mathrm{tr}\boldsymbol{C}^2) \tag{8.33}$$

The negative-log-likelihood function is given by

$$J(\boldsymbol{b}) = \frac{1}{2}\sum_{k=1}^{N}\left[\frac{1}{\sigma_{d_k}^2}\left(z_{d_k} - 2(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}}) \cdot \boldsymbol{b}\right)^2\right] + \frac{1}{2}\sum_{k=1}^{N}\left[\log \sigma_{d_k}^2 + \log 2\pi\right] \tag{8.34}$$

Since this equation isn't quartic of $\boldsymbol{b}$, it is easy to get the maximum-likelihood estimate via one-step calculation as

$$\widehat{\boldsymbol{b}}^* = \widehat{\boldsymbol{P}}_{bb} \sum_{k=1}^{N}\left[\frac{1}{\sigma_{d_k}^2} 2(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}}) z_{d_k}\right] \tag{8.35}$$

$$\widehat{\boldsymbol{P}}_{bb} = \widehat{\boldsymbol{F}}_{bb}^{-1} = \left[\sum_{k=1}^{N} \frac{1}{\sigma_{d_k}^2} 4(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}})(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}})^{\mathrm{T}}\right]^{-1} \tag{8.36}$$

8.3.2 The Convergence of Differential Value Iterative Algorithm

From Eqs. (8.35, 8.36) and the definition of z_{d_k}, we can write

$$\widehat{\boldsymbol{b}}^* = \boldsymbol{b} + \widehat{\boldsymbol{P}}_{bb} \sum_{k=1}^{N}\left[\frac{1}{\sigma_{d_k}^2} 2(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}}) \varDelta_{d_k}\right] = \boldsymbol{b} + \boldsymbol{\delta} \tag{8.37}$$

where

$$\boldsymbol{\delta} \sim N(0, \widehat{\boldsymbol{P}}_{bb}) \tag{8.38}$$

It is easy to see that $\widehat{\boldsymbol{F}}_{bb}$ is a real symmetric positive definite matrix from Eq. (8.36).

Considering the characteristics of sun-synchronous orbit, the geomagnetic field measured on board changes periodically. That is

$$\boldsymbol{B}_{b_k} = \boldsymbol{B}_{b_{k+T}} \tag{8.39}$$

where k denotes time t_k and T denotes the orbit period. It follows that

$$\sigma^2_{d_k} = \sigma^2_{d_{k+T}} \tag{8.40}$$

Further,

$$\begin{aligned}\widehat{\boldsymbol{F}}_{bb} &= \sum_{k=1}^{N} \frac{1}{\sigma^2_{d_k}} 4(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}})(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}})^{\mathrm{T}} \\ &= n\sum_{k=1}^{T} \frac{1}{\sigma^2_{d_k}} 4(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}})(\boldsymbol{B}_{b_{2k}} - \boldsymbol{B}_{b_{2k-2l+1}})^{\mathrm{T}}\end{aligned} \tag{8.41}$$

$\widehat{\boldsymbol{F}}_{bb}$ can be represented as

$$\widehat{\boldsymbol{F}}_{bb} = n\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\lambda}\boldsymbol{Q} = n\begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \tag{8.42}$$

where $\boldsymbol{Q}$ is an orthogonal matrix. And there are real numbers U and V which satisfy $U \geq \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq V > 0$. It follows that

$$\widehat{\boldsymbol{P}}_{bb} = \widehat{\boldsymbol{F}}_{bb}^{-1} = \frac{1}{n}\begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \frac{1}{\lambda_1} & 0 & 0 \\ 0 & \frac{1}{\lambda_2} & 0 \\ 0 & 0 & \frac{1}{\lambda_3} \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \tag{8.43}$$

where the diagonal elements of $\widehat{\boldsymbol{P}}_{bb_{ii}}$ can be written as

$$\widehat{\boldsymbol{P}}_{bb_{ii}} = \frac{1}{n\lambda_1}q^2_{1i} + \frac{1}{n\lambda_2}q^2_{2i} + \frac{1}{n\lambda_3}q^2_{3i} \leq \frac{1}{nV} \tag{8.44}$$

Obviously, when $N \to \infty$ $(n \to \infty)$, $\widehat{\boldsymbol{P}}_{bb_{ii}} \to 0$. Then we can write

$$\lim_{N\to\infty} \widehat{\boldsymbol{b}}^{*} = \boldsymbol{b} \tag{8.45}$$

That is the differential value iterative algorithm can converge to magnetometer bias.

8.4 Simulation and Experimental Results

8.4.1 Satellite Model

A model of the satellite's attitude dynamics can be expressed as

$$\dot{\boldsymbol{q}} = \frac{1}{2}(q_0\boldsymbol{E} + [\boldsymbol{q}\times])\boldsymbol{\omega} \tag{8.46}$$

$$\boldsymbol{I}\underline{\boldsymbol{\omega}} = -\boldsymbol{\omega} \times (\boldsymbol{I}\boldsymbol{\omega} + \boldsymbol{h}) - \dot{\boldsymbol{h}} + \boldsymbol{T}_{mt} + \boldsymbol{T}_d \tag{8.47}$$

where $\boldsymbol{q}$ and q_0 are the vector and scalar representations of quaternion, $\boldsymbol{\omega}$ is the relative angular velocity between the body fixed frame and the orbital reference frame expressed in body coordinates, $\boldsymbol{I}$ is the moment of inertia, $\boldsymbol{h}$ is the angular momentum of reaction wheels, $\boldsymbol{T}_{mt}$ is the actuator torque, $\boldsymbol{T}_d$ is the disturbance torque.

8.4.2 Comparison of Estimation Algorithms from Simulation

In this section, the software simulation results are presented to show the accuracy of estimation of the differential value, Centered and EKF methods.

The simulated satellite is an Earth-pointing spacecraft in an 800 km low-Earth sun-synchronous orbit. The geomagnetic field is simulated using an 11th-order International Geomagnetic Reference Field model. The satellite works in three axis stabilized mode. The precision and stability of attitude determination and control are required to be within 2° and 0.1°/s respectively. The initial states of attitude angle and angular velocity are set to zero and the magnetometer measurement noise is assumed to be white and Gaussian and the covariance is set to be isotropic with a standard deviation of 350 nT. The magnetometer bias is not added in initial state and then changes to $[114, -250, 1{,}642]^{\mathrm{T}}$nT after the satellite enters working order. The measurements are sampled every 250 ms over 24,500 s. The difference calculation time between two samples is set to 500 s.

The EKF method is presented in Ref. [7]. The state model is given by

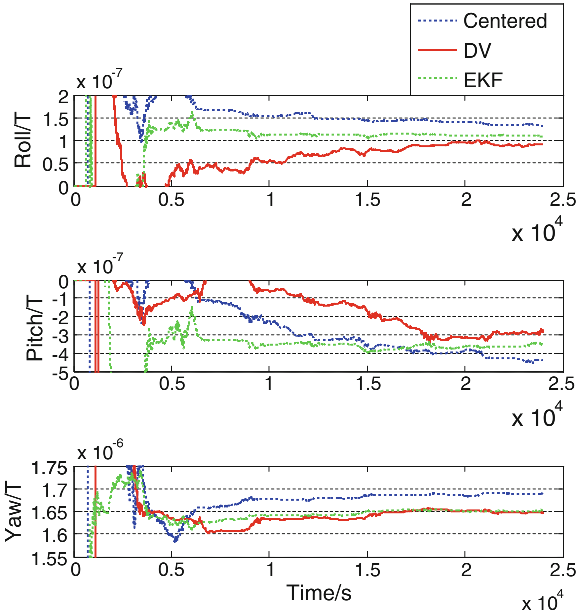
$$\dot{\boldsymbol{b}}(t) = 0 \tag{8.48}$$

And the measurement model is given by

$$z = h(\boldsymbol{b}) + v \tag{8.49}$$

where $h(\boldsymbol{b}) = 2\boldsymbol{B}_b \cdot \boldsymbol{b} - |\boldsymbol{b}|^2$. Then the EKF equations are presented as

Fig. 8.1 Estimation of magnetometer-bias



$$\hat{\boldsymbol{b}}_{k+1/k} = \hat{\boldsymbol{b}}_k, \quad \hat{\boldsymbol{b}}_{k+1} = \hat{\boldsymbol{b}}_{k+1/k} + \boldsymbol{K}_{k+1}\left[z_{k+1} - h\left(\hat{\boldsymbol{b}}_{k+1/k}\right)\right] \tag{8.50}$$

$$\boldsymbol{P}_{k+1} = [\boldsymbol{E}_{3\times3} - \boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1}]\boldsymbol{P}_k, \quad \boldsymbol{H}_{k+1} = 2\left(\boldsymbol{B}_{b_{k+1}} - \hat{\boldsymbol{b}}_{k+1/k}\right)^{\mathrm{T}} \tag{8.51}$$

$$\boldsymbol{K}_{k+1} = \boldsymbol{P}_k\boldsymbol{H}_{k+1}^{\mathrm{T}}\left(\boldsymbol{H}_{k+1}\boldsymbol{P}_k\boldsymbol{H}_{k+1}^{\mathrm{T}} + \sigma_{k+1}^2\right)^{-1} \tag{8.52}$$

The initial state of magnetometer bias is set to $[0, 0, 0]^{\mathrm{T}}$nT with a standard deviation of 1,000 nT.

The comparison of the three algorithms is shown as Fig. 8.1.

It isn't hard to see from Fig. 8.1 that the performance of differential value method is better than the other two and its accuracy of estimation is less than 50 nT.

The estimates of magnetometer bias via three different algorithms are shown as Table 8.1.

Table 8.1 Simulation results of the estimation of magnetometer-bias

	DV			Centered		EKF	
	Truth (nT)	Estimate (nT)	Variance (T^2)	Estimate (nT)	Variance (T^2)	Estimate (nT)	Variance (T^2)
b11	114	91.19	7.56×10^{-16}	130.30	4.87×10^{-17}	111.85	4.91×10^{-17}
b12	−250	−265.05	8.98×10^{-15}	−454.88	3.04×10^{-15}	−326.50	2.89×10^{-15}
b13	1,642	1644.77	2.06×10^{-16}	1692.01	7.15×10^{-17}	1648.75	6.67×10^{-17}

8.5 Conclusion

In this paper, we analyzes the basic principles of attitude independent magnetometer bias estimation and develops the differential value algorithm from the Acuña's algorithm. This method avoids the impact of quartic nature successfully and meets the requirement of on-orbit real-time application. Simulation results indicate that the accuracy of estimation would be less than 50 nT with the differential value method, whose performance is better than Centered and EKF algorithms and has no need to set an initial state in advance. The differential value algorithm can reduce the amount of calculations effectively and improves the accuracy of estimation.

References

1. Finlay C, Maus S, Beggan C et al (2010) International magnetic reference field: the eleventh generation. Geophys J Int 183(3):1216–1230
2. Yang ZH, Yu YJ, Qi ZQ (2012) Error analysis and on-board calibration of magnetometer in space environment exploration satellite. J Astronaut 33(8):1104–1111
3. Huang L, Jing WX (2006) Effects of remanence on attitude determination and control of satellites. Acta Aeronautica et Astronautica Sinica 3(27):390–394
4. Inamori T, Sako N, Nakasuka S (2010) Strategy of magnetometer calibration for nano-satellite missions and in-orbit performance. AIAA guidance, navigation and control conference. Canada, Toronto, pp 2–5
5. Alonso R, Shuster MD (2002) TWOSTEP: a fast robust algorithm for attitude-independent magnetometer-bias determination. J Astronaut Sci 50(4):433–451
6. Crassidis JL, Markley FL, Lightsey EG (1999) Global positioning system integer ambiguity resolution without attitude knowledge. J Guid Control Dyn 22(2):212–218
7. Crassidis JL, Lai KL, Harman RR (2005) Real-time attitude-independent three-axis magnetometer calibration. J Guid Control Dyn 28(1):115–120
8. Alonso R, Shuster MD (2002) Attitude-independent magnetometer-bias determination: a survey [J]. J Astronaut Sci 50(4):453–475

Chapter 9
Unified Fit Algorithm for Satellite Clock Parameters of BDS

Lan Du, Peiyuan Zhou, Yu Lu, Li Liu and Lingfeng Zhu

Abstract Satellite clock offset correction of GNSS is very important to positioning, navigation and timing users. In this paper, a unified two-step fit to solve the clock parameters is proposed using both long- and short-term data sets and can be applied for all satellite clocks under the various durations of out of view. Firstly, a standard quadric polynomial model is fitted with longer dataset. Then a linear fit is done to correct the frequency bias with local short-term data. Finally, a synthetic solution is combined to precisely predict clock offsets. The computation procedures are presented in details. Experiment results with a GEO on-board rubidium clock have proved the superiority of the proposed Algorithm in 2-h prediction and fast recovery compared with stand-alone linear and quadric polynomial models.

Keywords Navigation satellite · Clock offset prediction · Clock offset · Frequency bias · Frequency drift

9.1 Introduction

In order to provide real-time GNSS navigation services, precise satellite clock correction have to be estimated and predicted, and then broadcast to users through navigation message [1–3]. In view of the ultra-stable characteristics of atomic clock on-board GNSS satellites, usually a low-order polynomial is a simple and effective choice as the short-term clock offset predicting model. Actually, 3-parameter quadric polynomial model is commonly adopted by GNSS such as GPS, GLONASS, Galileo, BDS, etc.

L. Du (✉) · P. Zhou · Y. Lu
College of Navigation and Aerospace Engineering, Information Engineering University, Zhengzhou 450052, China
e-mail: lan.du09@qq.com

L. Liu · L. Zhu
Beijing Global Information Application and Development Center, Beijing 100094, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_9

Precise clock parameters are estimated from the satellite clock offset series measured at time-comparison stations [4]. Unfortunately, due to the regional distribution of these stations for BDS, the measurements cannot be achieved continuously 24 h per day for the Beidou hybrid constellation. For example, the IGSO satellites have about 5 h per day being out of view, while the unreachable time period for the MEO satellites may last as long as up to 20 h in extreme conditions.

Therefore, BDS employs a considerably sophisticated algorithm to solve the clock parameters to deal with the frequent in- and out of view of its navigation satellites. In this algorithm, the fitting polynomial can be varied with the quantity and quality of the historical data available for a specific satellite clock. Normally, linear model is fitted with 2 h of data and updated hourly. Whenever longer periods of out of view or data missing take place, quadric model is preferred and the length of fitting data is accordingly extended to 24–72 h. Once the tracking data recover, at such cases, dozens of minutes of new data are accumulated to re-fit a linear model which is so-called fast recovery.

It is noted that three parameters of quadric model are presently assigned to clock offset correction in broadcast ephemeris for GNSS. Therefore, the linear model does not make full use of the precious resources of navigation message. Moreover, the conversion threshold between linear and quadric models should be optimized for each of satellite clocks according to clock behaviors, instead of a fixed value commonly set for all clocks.

The purpose of this study is to develop a uniform two-step algorithm to unify satellite clock models for GEOs, IGSOs and MEOs of BDS. This paper is organized as follows. First, the currently used piecewise clock fitting model is introduced. Next, two-step algorithm is investigated in details. Finally, a numerical example is employed to validate the developed algorithm in term of predicting accuracy and clock jump of update epoch.

9.2 Satellite Clock Offset Model and Algorithm in BDS

The satellite-ground time comparison has two tasks: satellite clock offset computation and clock offset prediction [5–7]. Usually, the former is the process of mathematical fit and the fitted model is then used to predict. Hence, the selected model is vital to ensure the accuracy of predicting clock offset. In most GNSS operations, the prediction model of quadric polynomial is just the fit model. Within the valid period, the clock offset, say a_t, can be obtained by

$$a_t = \hat{a}_0 + \hat{a}_1 \cdot (t - t_{oc}) + \hat{a}_2 \cdot (t - t_{oc})^2 \tag{9.1}$$

where t_{oc} is reference epoch of clock offset prediction, $\hat{a}_0$, $\hat{a}_1$ and $\hat{a}_2$ are the fitted coefficients of quadric polynomial, which represent the clock offset of the epoch, frequency bias and frequency drift, respectively.

BDS, however, employs a considerably sophisticated algorithm to solve the clock parameters. In this algorithm, the fitting polynomial can be varied with the quantity and quality of the historical data available for a specific satellite clock.

I. If the data are continuously sampled, like all GEOs and parts of orbital arc of IGSOs and MEOs, the linear model is adopted with 2 h of nearest historical data. Herein, the frequency drift $\hat{a}_2$ held as zero.
II. If the data gap is larger than several hours, such as IGSOs and MEOs' out of view from the ground tracking network, at least two changes need to be considered. Firstly, the quadric polynomial is employed with much longer historical data, from 24 up to 72 h, to precisely estimate the clock parameters. Next, the predicted result may be too worse to be acceptable after long period of out-of-view. So, a new linear model needs to be fitted rapidly after the satellite appears in view again. Unlike the normal condition, there might only several or dozens of minutes' data to fit at these circumstances.

It is noted that the clock parameters is hourly updated in order to control the error propagation of the predicting clock offset. So, the valid period of the clock parameters is merely 2 h, which shows that the performance of BDS's on-board clocks still lower than that of GPS satellite clocks in terms of stability and accuracy.

9.3 Two-Step Fit Algorithm of Uniformed Model

It should be noticed that quadric polynomial model plays dual roles in clock offset modeling. Firstly, it is a mathematical approximation function in fitting the 3 coefficients of polynomial. More importantly, it is a prediction model which is required to follow the characteristic variation of clock as closely as possible in order to decrease prediction errors.

Existing data fitting algorithms of clock parameters have proved different merits of low order polynomials. On the one hand, the linear model fitted with 2 h of short-term data can reflect the frequency bias characteristics, and thus is suitable for short-term prediction. On the other hand, quadratic polynomial model resulting from more than 24 h of data can give frequency bias and frequency drift values at the same time, and is more applicable for medium- and long-term prediction. Therefore, different error propagation laws over time make it difficult to decide when it should be transferred from one model to the other, even for a given clock. As for multiple on-board clocks with different performances and irregular data disruptions, it is unreasonable to give a unified transfer criterion between two models.

In fact, the broadcast ephemeris has already assigned 3 parameters for the clock offset prediction, which should be taken full advantage of in the fitting strategy. Apart from frequency drift $\hat{a}_2$ fitted with quadric model, there are two frequency bias estimates $\hat{a}_1$, separately derived from linear and quadric polynomial fits. For a long time series, it is noted that some considerable compromises between the coefficients of the first and second order item may be unavoided in order to best fit a quadric polynomial as a whole. In other words, the fitted first order coefficient has more or less mathematical adjustment attached on the real linear trend. So, it is worthwhile combining the two fit models to ensure the prediction accuracy.

We propose a two-step algorithm to obtain the 3 coefficients of the quadric polynomial model. Firstly, a standard quadric polynomial model is fitted with longer dataset. Then a linear fit is done near the prediction epoch to recover the local variation of frequency bias. Finally, we can combine the long-term estimates with the locally revised linear trend to obtain a synthetic solution, which reflects simultaneously the long-term frequency drift and rapidly varied frequency bias.

The detailed computation procedures are as follows:

1. Intercept a time series of clock offsets of 24–72 h. Notice that several data gaps may appear in the fit data.
2. Take the last epoch as the fit reference epoch t_0, and fit a quadratic polynomial model with the whole data set. The fitted parameters are denoted by $\left(\hat{a}'_0, \hat{a}'_1, \hat{a}_2\right)$.
3. Re-fit a linear model with 2 h of fitting residuals nearest to t_0, and get the locally linear corrections $(\delta\hat{a}_0, \delta\hat{a}_1)$ about t_0.
4. Augment the quadratic polynomial model with the locally linear corrections, and the final fitting parameters are denoted by $(\hat{a}_0, \hat{a}_1, \hat{a}_2)$.
5. Propagate the clock offset $\hat{a}_0$ from the fit reference epoch t_0 to the prediction reference epoch t_{oc} with the final fitting quadric polynomial model.

In the last step, a fast clock offset recovery would be needed if the prediction reference epoch is too far from the fit reference epoch, such as out of view or data breakdown. In these cases, the computation will go back to step 3, but this time linear fit can only be done with the newly obtained data of less than 20 min after the satellite is back in view.

9.4 Experiments and Analysis

A GEO on-board rubidium clock is used to validate the proposed unified fit algorithm. GEO navigation satellites can be tracked 24 h a day with a regional ground network. So, the real clock data are employed to carry out experiments in all kinds of data gap conditions. The tested data start from October 2, 2013, and last for 7 days. The nominal accuracy is 0.1 ns and the sampling interval is 1 s.

These clock offset data are separated into 120 data groups, and follow the fitting procedure of hourly updated. Each group has 48 h of data. The data of the first 24 h are treated as inputs for the parameter estimation and the rest are taken as the true values for the statistical analysis of predict accuracy.

Based on the regional tracking network for GEO/IGSO/MEO satellites, the tested conditions of prediction gap are simplified as follows: (a) no gap; (b) 5 h of out-of-view; (c) 18 h of out-of-view.

The presented two-step fit algorithm is compared with two classical one-step algorithms: (I) 24-h data fit for quadric polynomial; (II) 2-h data fit for linear model. And herein the 2-h data is from the last part of 24-h fit data.

The performances of the three algorithms are evaluated by the prediction accuracy within the 2-h validation period and the predicting clock jump of update epoch between two consecutive groups. Finally, the clock offset fast recovery is also tested for the two prediction gap conditions.

9.4.1 Fitted Frequency Bias and Frequency Drift

The fitted frequency drift with 24-h data is shown in Fig. 9.1, and frequency bias from three fit algorithms are shown in Fig. 9.2.

- The frequency drift sequence has some large corner points at which the values can reach as large as 40 ns/d^2 (Fig. 9.1). These abrupt changes, accordingly, induce the evident fluctuations in the frequency bias estimates, such as group 40,

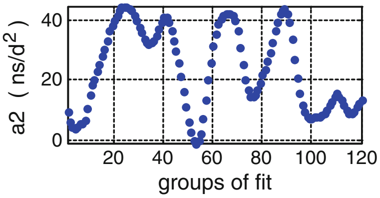


Fig. 9.1 Frequency drift series of 24-h quadratic fit

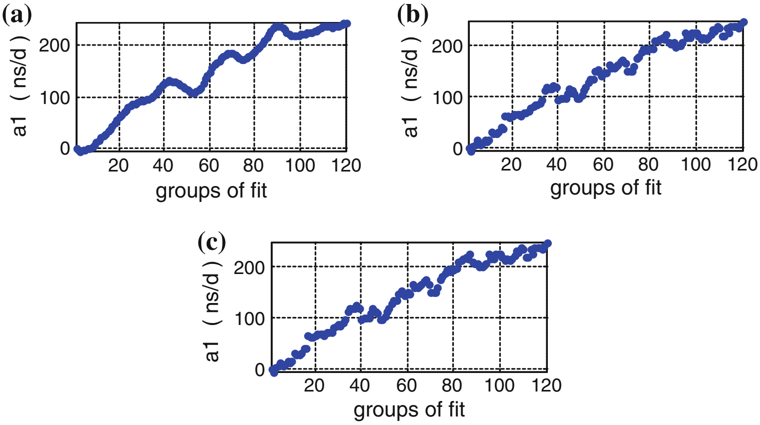


Fig. 9.2 Frequency bias series of three fitting algorithms **a** 24-h quadric polynomial fitting **b** 2-h linear fitting **c** two-step fitting

70 and 90, especially in 24-h quadric polynomial fit in Fig. 9.2a, which shows that the clock acceleration term should be taken into account in the prediction model.

- From Fig. 9.2a, b, the frequency biases from 2-h linear fit can demonstrate local variations in details, while those from 24-h quadric polynomial fit are far more smooth.
- With the frequency drift unchanged, the two-step algorithm uses localized short-term data to improve the frequency bias. Therefore, the local details of Fig. 9.2c follow very well with those of 2-h linear fit; see Fig. 9.2b.

To take a closer look at the details, the frequency biases of 2-h linear fit and 24-h quadric polynomial fit are differenced respectively with those of the two-step fit, and then illustrated as PN1 and PN2 in Fig. 9.3. It can be seen that the fitted clock rates of two-step algorithm are sufficiently consistent with the results of 2-h linear fit, which means they are capable of reflecting the short-term characteristics of the atomic clocks. In contrast, the discrepancy of PN2 from 24-h quadric polynomial fit, are significantly large or even up to nearly 30 ns/d. Hence, it is doubtable that the first order coefficient of quadric polynomial fit can properly illustrate the detailed short-term clock behaviors.

9.4.2 Accuracy of Predicted Clock Offset

The clock offsets within 2 h of validation period are predicted using the above-mentioned three models and the statistics of predicting RMS are listed in Table 9.1. Note that the prediction time period would be prolonged by extra 5 and 18 h respectively to take the out-of-view of IGSOs and MEOs into account.

It has been admitted that linear model is suitable to short-term prediction, while quadric polynomial model is preferred to predict long-term variations. In Table 9.1, we can see that neither of the two one-step models is competent simultaneously to short- and long-term prediction. When the time period of prediction is no more than

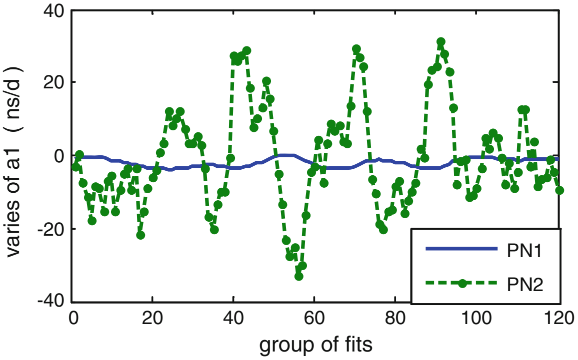


Fig. 9.3 Differences of frequency bias series of two classical methods against two-step method

Table 9.1 RMS statistics of predicting residuals using three models

Gap	RMS (ns)	Quadric polynomial prediction	Linear prediction	Two-step prediction
No	<0.5	17 (14.2 %)	70 (58.3 %)	75 (62.5 %)
	<1	40 (33.3 %)	108 (90 %)	111 (92.5 %)
	<2	84 (70 %)	120 (100 %)	120 (100 %)
5 h	<2	35 (29.2 %)	42 (35 %)	42 (35 %)
	<5	69 (57.5 %)	99 (82.5 %)	103 (86 %)
	<10	109 (90.8 %)	120 (100 %)	120 (100 %)
18 h	<2	6 (5 %)	7 (5.8 %)	12 (10 %)
	<5	14 (11.7 %)	14 (11.7 %)	30 (25 %)
	<10	42 (35 %)	27 (22.5 %)	60 (50 %)

5 h, the linear prediction maintains overall superiority. While the out-of-view period approaches 18 h, however, the quadric polynomial becomes to exceed as a whole. Therefore, the traditional algorithms require coexistence of short-term and long-term prediction models, and thus frequent transfers between them are also needed.

Two-step fit algorithm gives an unified model to predict clock offset. On the one hand, it suits to long-term prediction with the form of quadric polynomial. On the other hand, it improves the frequency bias estimate in order to precisely predict short-term variations. It can be shown in Table 9.1 that the prediction accuracies of the two-step prediction are higher than those of the two traditional algorithms in each of three prediction cases.

Moreover, the predicting clock jumps at update epochs are listed statistically as in Table 9.2 for no-gap circumstances. The smaller jumps (less than 0.5 ns) from two-step algorithm are 76 %, which is still a little higher than those from the two traditional algorithms. That exhibits from another aspect that the two-step model is a more stable model to predict the clock offset.

9.4.3 Fast Recovery of Clock Offset

Because all the three models are low order polynomials, they have common drawbacks such as rapid degradation of prediction accuracy with time. Take two-step algorithm for example, the percentage of predicting RMS of less than 2 ns

Table 9.2 Statistics of clock jumps of updating epoch without out-of-view

Clock jump of update epoch (ns)	Quadric polynomial prediction	Linear prediction	Two-step prediction
<0.5	86 (71.7 %)	84 (70 %)	91 (76 %)
<1	112 (93.3 %)	107 (89.2 %)	109 (91 %)
<2	117 (97.5 %)	117 (97.5 %)	117 (97.5 %)

decreases from 100 % for normal 2-h prediction period down to 35 and 10 % for extension of 5 and 18 h respectively, see Table 9.3.

Therefore, if out-of-view duration is more than 5 h, the fast recovery must be implemented once the satellite is back in view and new data, namely recovering data, are obtained.

RMSs of 2-h prediction residuals of 120 groups are calculated using three fast recovery algorithms respectively, in term of accuracy and rapid-response time. The fit portions of these algorithms are: (a) linear fit with 20-min recovering data; (b) two-step fit without recovering data; (c) modified two-step fit with 20-min recovering data. Again, the statistics of the clock jumps from the hourly update predictions are shown in Table 9.4.

Compared with the traditional strategy of using 2-h data, the fast recovery strategy can use merely 20-min data to achieve almost same results, which indicates the high quality of clock offset data as well as the high short-term stability of the on-board clocks.

The two-step fit without re-cumulating data, however, become to exhibit its weakness since prediction accuracy degrades rapidly with time. When the out-of-view period lasted for 18 h, the percentage of prediction RMS of less than 0.5 ns decreases from 52.5 to 48.3 %.

Finally, the modified two-step fit can retain the consistency of prediction accuracy regardless of the out-of-view duration. Furthermore, the two-step fit algorithms with- and without recovering data are able to demonstrate better prediction stability than linear mode because they both takes into account the effect of frequency drift term. In Table 9.4, the smaller jumps of less than 0.5 ns from both two-step modes (b) and (c) are 67.5 %, about 5 % higher than mode (a).

Table 9.3 RMS statistics of prediction residuals after recovery

Gap (h)	RMS (ns)	20-min linear fit	Two-step fit	Modified two-step fit
5	<0.5	62 (51.7 %)	63 (52.5 %)	65 (54.2 %)
	<1	103 (85.8 %)	103 (85.8 %)	103 (85.8 %)
	<2	116 (96.7 %)	117 (97.5 %)	117 (97.5 %)
18	<0.5	64 (53.3 %)	58 (48.3 %)	63 (52.5 %)
	<1	104 (86.7 %)	103 (85.8 %)	103 (85.8 %)
	<2	116 (96.7 %)	116 (96.7 %)	116 (96.7 %)

Table 9.4 Statistics of clock jumps of recovery prediction

Clock jump of update epoch (ns)	2 h linear model	Two-step prediction	Modified two-step prediction
<0.5	75 (62.5 %)	81 (67.5 %)	81 (67.5 %)
<1	97 (80.8 %)	101 (84.2 %)	101 (84.2 %)
<2	111 (92.5 %)	111 (92.5 %)	111 (92.5 %)

9.5 Conclusions

A uniformed two-step fit algorithm has been presented, and can be an alternative of the currently available multi-model fit algorithm to solve the satellite clock parameters for broadcast ephemeris in BDS. The new fit algorithm makes use of long-term data to estimate the frequency drift and local short-term data to correct the frequency bias [8].

Furthermore, the two-step fit is introduced into the fast recovery after the navigation satellite is back in view from the regional tracking network. The difference is that there is only much fewer data to be employed to revise locally the frequency bias.

The satellite clock parameters can be solved in a unified way by the two-step model for all on-board clocks. Firstly, the model transfer is avoided and thus the likely unreasonable transfer threshold between models is removed. Secondly, the prediction accuracy and smoothness are enhanced with the two-step model even in the fast recovery.

References

1. He F, Wang G, Liu L et al (2011) Ephemeris fitting and experiments analysis of GEO satellites. Acta Geodaetica Cartogr Sin 40(Suppl):52–58
2. Wang JG, Hu YH, He ZM et al (2012) Combination models for navigation satellite clock prediction. J Geodesy Geodyn 32(1):84–88
3. Huang GW, Zhang Q, Wang JG et al (2009) Research on estimation and prediction of GPS satellite clock error. J Geodesy Geodyn 29(6):118–122
4. Zhu LF, Li C, Liu L et al (2009) Research on methods for predicting clock error based on domestic hydrogen atomic clock. J Geodesy Geodyn 29(1):148–151
5. Zheng ZY, Lu XS, Yang F (2009) Analysis on GPS satellite clock error linear prediction from IGS ultra-rapid products. Sci Survey Map 34(6):48–50
6. Zheng ZY, Dang YM, Lu XS et al (2010) Study on GPS satellite clock error prediction with IGS ultra rapid ephemeris. Sci Survey Mapp 35(2):8–12
7. Huang GW, Yang YX, Zhang Q (2011) Estimate and predict satellite clock error using adaptively robust sequential adjustment with classified adaptive factors based on opening windows. Acta Geodaetica Cartogr Sin 40(1):15–21
8. Du L, Lin L, Wang RP et al (2013) Satellite clock parameter two-step fit algorithm for period of out-of-view. J Geodesy Geodyn 33(5):120–123

Chapter 10

Microwave Holographic Metrology of the Surface Accuracy of Reflector Antenna—Simulation Method

Yuhu Duan

Abstract The basic principle of microwave holographic metrology is simply described. The mathematic relationships between the phase error of aperture field and the optical path-length difference of antenna reflector, between normal error and surface accuracy were deduced. Take the gravity deformation of a large reflector antenna for deep space observation for example, the simulation method in the far-field used for microwave holographic metrology of surface accuracy of reflector antenna was presented. The holographic results of the pattern simulation with low-resolution medium accuracy and high-resolution high accuracy were simulated respectively. The simulation results show that reflector accuracy was increased from 1.77 to 0.241 mm and the antenna gain was increased by 1.75 dB at X band in the period of low-resolution medium accuracy; The gain at Ka band was increased by 0.95 dB and the side-lobe level was decreased by 0.8 dB in the period of low-resolution using high-frequency high-resolution simulation.

Keywords Surface accuracy · Holographic metrology · Resolution · Radiation pattern · Fourier transform

10.1 Introduction

The surface accuracy of reflector antenna is the most important factor in affecting antenna gain, beam width, beam direction and the side-lobe performance of the antenna radiation pattern [1]. For large reflector antenna, the error caused by gravity deformation is the biggest error source for antenna accuracy [2], whose effect on the RF performance of the antenna operating in high frequency is unacceptable. So the major concern of antenna designer is how to accurately measure the surface error of reflector and how to reduce its effect.

Y. Duan (✉)
The North West China Research Institute of Electronic Equipment, Xi'an 710065, China
e-mail: duanyuhu678@gmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_10

In recent years, many new metrology methods such as photogrametry method [3–5] and microwave holographic metrology method [6] have been developed. Using them, the antenna deformation at any elevation angle can be quickly and accurately measured. Microwave holographic metrology method refers to the method of measuring the antenna far-field and near-field amplitude and phase using the geostationary satellite beacon, artificially erected beacon or other space radiation source as measurement radiation source. Using the method the amplitude and phase distribution of antenna aperture field can be obtained through the inverse Fourier Transform of measured value of amplitude and phase. And the error distribution of antenna reflector can be obtained after a series of mathematic calculation of phase distribution, guiding the reflector adjustment. The Jet Propulsion Laboratory of USA Aeronautics and Space Administration designed a measuring system MAHST (Microwave Antenna Holography System) for antenna holographic metrology [7] in order to measure the surface accuracy of deep space station antenna. Its measurement accuracy can reach 0.25 mm. The antenna highest operating frequency after measurement and adjustment can reach 95 GHz. A microwave holographic metrology method used in ground near-field of measuring and data processing millimeter-wave antennas was introduced in reference [8]. Using this method the millimeter-wave antenna with 12 m diameter operating at 104 GHz frequency can be measure within 315 m limited distance, with the 16–17 μm measured rms error.

Microwave holographic metrology must meet several requirements. Firstly, the elevation angle and the source with proper radiation intensity are necessary to offer appropriate SNR for measurement system to meet the requirement of measurement accuracy. Secondly, fast measurement function must be realizable to finish once measurement of low-resolution and medium accuracy in little time (about 45 min/a measurement) in order to calibrate the position of sub-reflector. Then, the measurement ability of high resolution and accuracy must be provided to control the measuring time within 12 h in order to reduce the influence of temperature change.

10.2 The Mathematic Processing Method of Holographic Metrology

10.2.1 The Radiation Field of Antenna

The reciprocity theorem describes the equivalent relation of antenna radiation characteristics at the T_x and R_x state. So the equivalent radiation characteristic at the T_x state can be found through the analysis of the radiation characteristic at the T_x state. As shown in Fig. 10.1, the relationship between the radiation field $f(x, y, z)$ in space point $P(x, y, z)$ and the aperture field distribution function $F(x', y')$ can be formulated as:

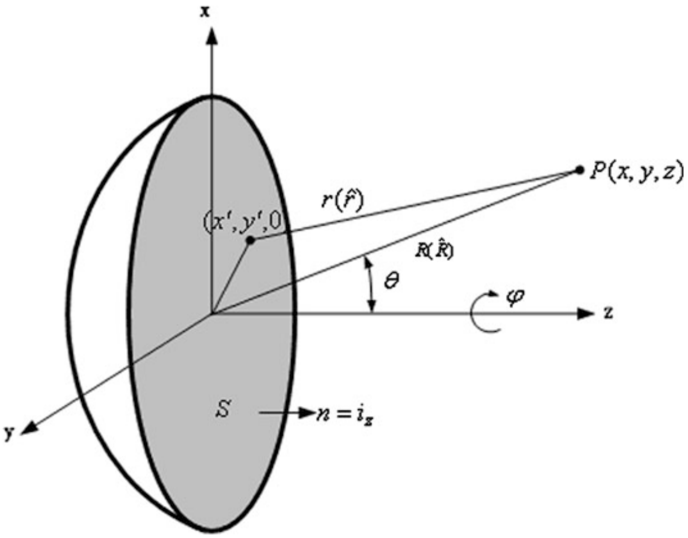


Fig. 10.1 The integral geometry of field point aperture with limited distance

$$f(x,y,z) = \iint\limits_S \frac{F(x',y') \cdot e^{-jkr}}{4\pi r}\left[\left(jk + \frac{1}{r}\right) \cdot i_z \cdot \hat{r} + jki_z S\right] \mathrm{d}x'\mathrm{d}y' \tag{10.1}$$

Where, the integral region is on the whole antenna aperture surface, $k = \frac{2\pi}{\lambda}$, and the unit vector is shown in Fig. 10.1. supposing the antenna aperture is far larger than operating wave length, and the field point P is far enough from antenna aperture surface, the formula (10.1) can be simplified to the following formula:

- The $1/r$ at the inside of parenthesis can be neglected;
- The $1/r$ at the outside of parenthesis can be replaced by the distance from the aperture central to the field point $1/R$;
- $i_z \cdot \hat{r}$ may be approximate to $i_z \cdot \hat{R} = \cos\theta$;
- $i_z \cdot S$ is the deviation of aperture surface phase from the uniform distributed phase, which is very small and supposed to be 1.

According to the above supposition, the formula (10.1) can be simplified to

$$f(x,y,z) = \frac{j}{2\lambda R}\iint\limits_S F(x',y') \cdot (1 + \cos\vartheta) \cdot e^{-jkr}\mathrm{d}x'\mathrm{d}y' \tag{10.2}$$

The distance from any point in antenna aperture to the field point P is r:

$$r = \left[(x - x')^2 + (y - y')^2 + z^2\right]^{0.5} \tag{10.3}$$

The coordinate of field point $P(x, y, z)$ may be revised to the following spherical coordinate:

$$\begin{aligned} x &= R \cdot \sin\vartheta \cdot \cos\phi \equiv R \cdot u \\ y &= R \cdot \sin\theta \cdot \sin\phi \equiv R \cdot v \\ z &= R \cdot \cos\theta = R\sqrt{1 - u^2 - v^2} \end{aligned} \tag{10.4}$$

Introduce the direction cosine of field point:

$$P(u, v) = (\sin\theta \cdot \cos\phi, \sin\theta \cdot \sin\phi) \tag{10.5}$$

Then, formula (10.3) can be revised to:

$$\begin{aligned} r &= \left\{(Ru - x')^2 + (Rv - y')^2 + R^2\left(1 - u^2 - v^2\right)\right\}^{0.5} \\ &= R\left\{1 - 2\frac{ux' + vy'}{R} + \frac{x'^2 + y'^2}{R^2}\right\}^{0.5} \end{aligned} \tag{10.6}$$

The formula (10.6) can be expanded by use of series:

$$\begin{aligned} r \approx R - (ux' + vy') + \frac{x'^2 + y'^2}{2R} - \frac{\left(x'^2 + y'^2\right)^2}{8R^3} \\ - \frac{(ux' + vy')^2}{2R} + \frac{(x'^2 + y'^2)(ux' + vy')}{2R^2} - \cdots \end{aligned} \tag{10.7}$$

10.2.2 Far-Field Region Approximation

In far-field position, R tends to be infinite. $\hat{R}$can be regarded to be parallel with $\hat{r}$. r in the series position of Eq. (10.2) can be replaced by the linear part of Eq. (10.7):

$$r = R - (ux' + vy') \tag{10.8}$$

In addition, for the antenna with high gain, the angle region θ in which we are interested is very small. In Eq. (10.2), $\cos\theta \approx 1$, whose introduced error is less than 0.1 % within 3° deviation of antenna beam from axial. So Eq. (10.2) can be revised to:

$$f(u, v) = \frac{je^{-jkR}}{\lambda R} \iint\limits_S F(x', y') e^{\{-jk(ux' + vy')\}} \mathrm{d}x' \mathrm{d}y' \tag{10.9}$$

The integral region mentioned above is on the whole antenna aperture surface S. In Eq. (10.9), $f(u, v)$ can be changed into $F(x', y')$ by Fourier Transform. Neglect $\frac{j}{\lambda}$, the following inverse Fourier Transform formula can be gained:

$$F(x', y') = \frac{e^{-jkR}}{4\pi R} \iint_S f(u, v) e^{\{jk(ux' + vy')\}} \mathrm{d}u\mathrm{d}v \tag{10.10}$$

The integral region of formula (10.10) is in the closed area of antenna aperture surface. So if the amplitude and phase of antenna radiation pattern of the whole spherical surface are known, the plurality field distribution in the antenna aperture $F(x', y')$, namely the amplitude value and phase value of aperture field, can be calculated. This method is known as microwave holographic metrology of reflector antenna. The deviation of aperture phase from the uniform aperture phase can reflect the error between actual reflector and theoretical reflector.

In real work, it is impossible to measure the antenna far-field radiation pattern on the whole spherical surface.

Nyquist sampling method can show the relationship between the deviation angle range of measured pattern from the antenna axial $\Theta = n \cdot \Theta_A$and the spatial resolution of antenna aperture field distribution, namely, $\delta = \frac{D}{n}$. Where, $\Theta_A \approx \frac{\lambda}{D}$ is the half-power beam width of antenna, D is antenna diameter, and λ is wavelength.

10.2.3 Surface Error Calculation

The geometry of distorted reflector is shown in Fig. 10.2.

In order to gain the residual surface error, the ray tracing method of geometrical optics is used. Then the Following relationship between normal error, axial error and aperture surface phase can be gained:

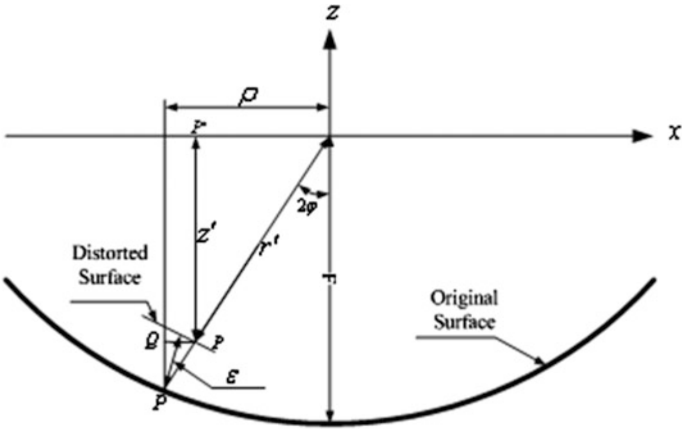


Fig. 10.2 The geometric relation of distorted reflector

$$\begin{aligned}\frac{1}{2}\Delta \boldsymbol{PL}_i &= \frac{1}{2}(\boldsymbol{P'P} + \boldsymbol{PQ}) \\ &= \frac{1}{2}\left(\frac{\varepsilon_i}{\mathbf{cos}\,\boldsymbol{\phi}_i} + \frac{\varepsilon_i \mathbf{cos}\,2\boldsymbol{\phi}_i}{\mathbf{cos}\,\boldsymbol{\phi}_i}\right) \\ &= \varepsilon_i \mathbf{cos}\,\boldsymbol{\phi}_i\end{aligned} \tag{10.11}$$

$$\boldsymbol{phase}(\Delta \boldsymbol{PL}_i) = \frac{4\pi}{\lambda}\varepsilon_i \mathbf{cos}\,\boldsymbol{\phi}_i \tag{10.12}$$

Where F is focal length, $\frac{1}{2}\boldsymbol{\Delta PL}$ is half optical path-length difference, and ε_i is normal error. substitute (10.4) into (10.2), the following formula is gained:

$$\Delta \boldsymbol{u} = \Delta \boldsymbol{v} = \frac{\boldsymbol{\kappa}}{\boldsymbol{D}/\lambda} \quad 0.5 \le \boldsymbol{\kappa} \le 1 \tag{10.13}$$

$$\varepsilon_i(x', y') = \frac{1}{2k\cos\phi_i} phase\{\exp(j2k\mathrm{F})IFFT[f(u, v)]\} \tag{10.14}$$

ε_i is the normal error of every adjustment point.

10.3 Simulation Methods

10.3.1 Radiation Patterns Simulation for Distorted Antenna Reflector

According to the distortion data of antenna reflector surface given by structural mechanics simulation, establish GRASP [9] model for antenna RF simulation which can be used to calculate antenna radiation patterns in u and v coordinate system, providing primitive data for microwave holography simulation. MATLAB language was used to write simulation software.

10.3.2 Low Resolution Simulation with Medium Accuracy

Firstly, perform low resolution simulation with medium accuracy to determine the adjustment amount of vertex of antenna subreflector. Low resolution simulation was performed at X band, frequency of 8.5 GHz. Sampling factor is $\boldsymbol{\kappa} = 0.65$, the number of sampling points of pattern is $25 \times 25 = 625$, calculation scope is $u = v = \pm 0.0082$, x and y coordinate resolution of corresponding antenna aperture is $\Delta x = \Delta y = 2153.846$ mm, the number of effective sampling points on the

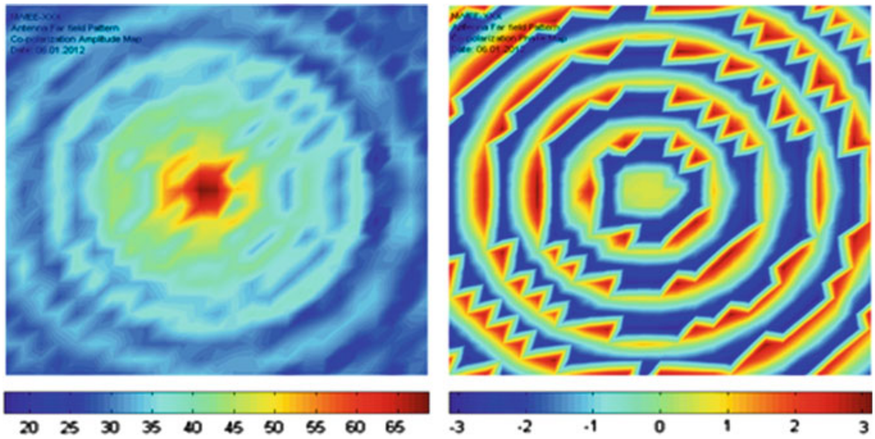


Fig. 10.3 Amplitude and phase patterns of distorted reflector (*left* is amplitude, *right* is phase)

reflector surface is 212. Calculated amplitude and phase pattern of antenna are shown in Fig. 10.3. Secondly, perform inverse Fourier transforms for the pattern, 1.777 mm rms error of antenna reflector surface is achieved. After performing the best fitting for surface error distribution, fitting parameter of main reflector and the amount of vertex adjustment on the subreflector respectively are as follows:

$$\begin{array}{ll} x_0 = 0.0\ \text{mm} & x_s = 0.0\ \text{mm} \\ y_0 = -0.731\ \text{mm} & y_s = -4.77\ \text{mm} \\ z_0 = -0.946\ \text{mm} & z_s = -1.65\ \text{mm} \\ \alpha = 0.0113^\circ & \varphi_x = 0.0113^\circ \\ \beta = 0.0^\circ & \varphi_y = 0.0^\circ \end{array}$$

The rms error of surface after best fitting is 0.241 mm.

The result shows that antenna directivity factor increases by 0.31 dB and sidelobe level decreases by 1.75 dB.

10.3.3 High Resolution Simulation with High Accuracy

The high resolution simulation with high accuracy includes two steps: firstly, use data given by low resolution simulation to calculate antenna radiation pattern at Ka band. The number of sampling points on the main reflector is 212, frequency is 32 GHz, sampling points of pattern is $101 \times 101 = 10201$, calculation scope is $u = v = \pm 0.0088$, calculated antenna pattern is shown in Fig. 10.4. Secondly, after performing inverse Fourier transforms, x, y coordinate resolution of corresponding antenna aperture is $\Delta x = \Delta y = 533.13$ mm, error distribution on the antenna

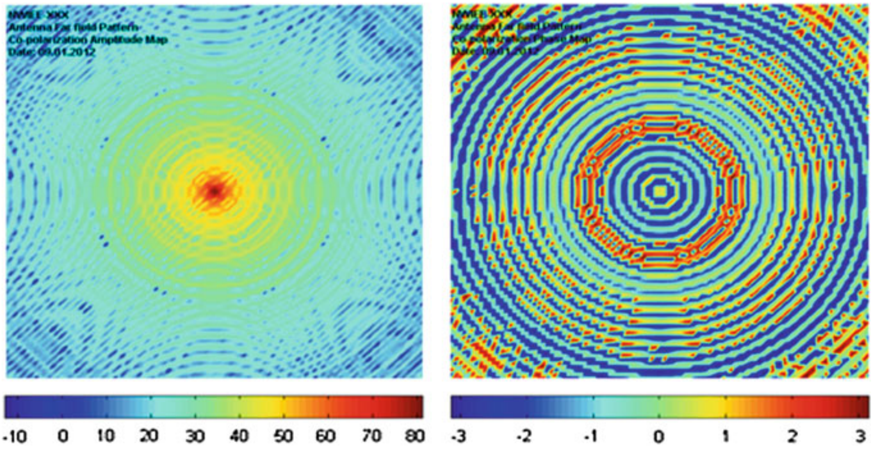


Fig. 10.4 Amplitude and phase patterns of distorted reflector after best fitting with high resolution

surface is achieved, and rms error is 0.348 mm(with 3644 sampling points). After best fitting of surface error distribution, the amount of subreflector vertex adjustment is achieved as follows:

$$\begin{array}{ll} x_0 = 0.0\ \text{mm} & x_s = 0.0\ \text{mm} \\ y_0 = -0.0045\ \text{mm} & y_s = 0.0\ \text{mm} \\ z_0 = -0.6828\ \text{mm} & z_s = 1.37\ \text{mm} \\ \alpha = 0.0^\circ & \varphi_x = 0.0^\circ \\ \beta = 0.0^\circ & \varphi_y = 0.0^\circ \end{array}$$

The surface rms error after best fitting is 0.072 mm.

Use data obtained after best fitting to calculate antenna radiation pattern at X and Ka band. Compared with low resolution simulation, Ka band antenna gain is increased by about 0.95 dB, sidelobe level decreases by about 0.9 dB. X band antenna gain increases by about 0.06 dB, sidelobe level almost has no change.

10.4 Conclusions

Based on the gravity distortion of a large reflector antenna for deep space observation, the paper calculates radiation pattern after antenna reflector surface is deformed by gravity when antenna points at 10°of elevation angle. Microwave holography is used to simulate relationship between phase error of antenna aperture field and surface distortion. The least square method is used to obtain best fitting of distorted reflector surface, achieving optimal position of antenna subreflector and the amount of adjustment corresponding to antenna main reflector adjustment points, simulating antenna radiation pattern. The result shows that microwave holography method is able to effectively improve antenna surface accuracy and antenna radiation characteristics.

References

1. Ruze J (1966) Antenna tolerance theory—a review. Proc IEEE 54(4):633–640
2. Imbriale WA, Britcliffe MJ, Brenner M (2001) Gravity deformation measurements of NASA's deep space network 70-meter reflector antenna. The interplanetary network progress report 42–147, pp. 1–15. http://ipnpr.jpl.nasa.gov/progress_report/42-147/147I.pdf
3. Rochblatt DJ, Serdel BL (1992) Microwave antenna holography. IEEE Trans Microwave Theor Tech 40(6):1294–1300
4. Rochblatt DJ (1991) A microwave holography methodology for diagnostics and improvement for large reflector antenna. The telecommunications and data acquisition progress report 42–108, pp 235–252. http://ipnpr.jpl.nasa.gov/progress_report/42-108/108R.pdf
5. Bennet JC, Anderson AP, McInnes PA et al (1976) Microwave holographic metrology of large reflector antennas. IEEE Trans Antennas Propag AP-24:295–303
6. Jones TW, Pappa RS (2002) Dot projection photogrammetric technique for shape measurements of aerospace test articles. AIAA Paper 2002-0532
7. Rochblatt DJ, Withington PM, Jackson HJ (1995) DSS-24 microwave holography measurements. The telecommunications and data acquisition progress report 42–121, pp 252–270. http://ipnpr.jpl.nasa.gov/progress_report/42-121/121A.pdf
8. Baars J W M, Lucas R (2007) Near-field radio holography of large reflector antennas. IEEE Antennas Propag Mag 49(5):24
9. GRASP 9, Technical description editor: Knud Pontoppidan, TICRA, Denmark.© 2005 TICRA engineering consultants, Copenhagen, Denmark

Chapter 11
Analysis of Space Probe Performance of Different Array Manifold Spectrum Estimation

Ting Li and Jianping Hu

Abstract Different kinds of spectrum estimation algorithm based on the Uniform Linear Array (ULA) may not be directly applied to the Uniform Circular Array (UCA), due to the complexity of the expression which describes the UCA manifold. The UCA occupies a huge area, so in practice where space is limited, the direct of arrival (DOA) may not be properly estimated. On the basis of the above, an L-shaped array of 9-elements is proposed to satisfy the requirement of the space probe. Its calculation is simplified, and the error of the angle estimation is reduced. Thus, in the same simulation condition, the performance of the L-shaped array is better than that of the UCA, and the L-shaped array have been successfully applied in the non-cooperative target detection.

Keywords Space probe · DOA estimation · Array manifold · Spectrum estimation · Non-cooperative target detection

11.1 Introduction

The spectrum estimation is one of DOA estimation ways; the DOA estimation is a key point in the spatial spectrum estimation research. It is widely applied in such areas as radar, communication, sonar, navigation, earthquake, biomedical engineering and radio astronomy to detect and locate the target. Through the DOA estimation, the azimuth and elevation angle parameters may be obtained to locate the target by the single station. This technology may be applied to detect and locate the space target passively. However, the incident wave may be the signal from the target which emits it, or the signal reflected from the third party, like the communication signal from the mobile base station.

T. Li (✉) · J. Hu
Southwest China Institute of Electronic Technology, Chengdu 610036, China
e-mail: liting142@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_11

The development history of the DOA estimation is as follows:

In 1899, Schuster firstly proposed the periodogram [1, 2], and then in 1958 Blackman and Turkey proposed BT (Blackman-Turkey) algorithm. The two methods are the initial methods of DOA estimation. However, due to the restriction of the Rayleigh limit, they did not have high resolution, and were rather sensitive to noise, which may not satisfy a higher requirement.

In 1967, Burg proposed maximum entropy method [3, 4], which established the modern spatial spectrum estimation research. It had a high resolution, but required a huge calculation and its robustness is poor. Hence, it might not be applied in practice because of these limits.

In 1969, Capon applied maximum likelihood spectral estimation method [5, 6] when he conducted the wave number research in earth-physics. It attracted people's attention due to its high-resolution performance and better robustness. However, this method needs to search the high dimensional parameter space, which produces a huge calculation. Therefore, it is hard to apply it in practical detection.

After 1970s, Subspace decomposition algorithm was proposed, like MUltiple SIgnal Classification (MUSIC) algorithm [7] proposed by Schmidt and ESPRIT algorithm [8, 9] proposed by Roy. They had high-resolution performance. The MUSIC algorithm requires a great calculation due to its spectrum peak search, while the ESPRIT algorithm requires special array structure to conduct DOA estimation, which limits its application area.

In 1973, Professor Brennan from the U.S proposed space-time adaptive [10] signal processing technology. It might reduce the limitation of the array structure, and had great Noise immunity. The typical DOA estimation method ignored the signal's time response, which might not fully utilize the signal's useful information. Hence, it is necessary to bring in the proper time-domain information processing when the space-domain processing by typical methods is conducted.

In 1975, Gardner and Franks proposed cyclic stable signal DOA estimation method-cyclic MUSIC and Cyclic ESPRIT [11, 12]. Considering that in certain condition where the communication signal and radar signal might tend to be cyclic stable, researchers combined the technology of cyclic stable signal processing and the conventional spectrum estimation method to propose a series of cyclic stable signal DOA estimation method.

In 1991, Sanjay and Tariq [13] proposed the neural-network method in the DOA estimation. Considering that vocal and radar etc. except a few special cases, signal mainly shows unstable characteristics (limited duration and time variable). Therefore, researchers began to apply neural-network method in DOA estimation. It has been a hotspot in recent years.

At present, the majority of the mentioned spectrum estimation methods are still in the stage of the theory research and lab simulation, but not applied in practice. Among them, the MUSIC method has a better resolution, moderate calculation, better stability and widely application in different array structures. They are always adopted prior in engineering practical experiment. Moreover, a lot of corresponding equipment and devices are developed. In addition, since the ULA can mostly provide the azimuth information of 180° none-fuzzy and cannot provide elevation angle information, it is necessary to design such arrays as L-shaped array, rectangle array, UCA and conformal array etc., which can obtain the azimuth and elevation angle information of 2-D DOA estimation.

Based on the former analysis, the UCA and L-shaped array space probe performance will be discussed and compared in the method of 2-D MUSIC algorithm. The azimuth and elevation angle of the detection target are estimated properly to achieve the single-station 2-D detection location of the non-cooperative target. The corresponding technologies are analyzed and compared, thus providing a theory and simulation data to space target passive detection.

11.2 Signal Model and Array Manifold

Assume the receiving antenna array is arranged in a certain way. The characteristics of the receiving array elements are related to their location but independent of their size (assume that the element is a point. Each element is an omnidirectional array element with an equal gain. The coupling between each other is negligible). The spatial source signal received by the array may be regarded as a bunch of parallel plane wave. The different time delay when the spatial source signal is received by each element may be decided by the array geometry and incident direction of space wave. The incident direction of space wave is estimated by the DOA.

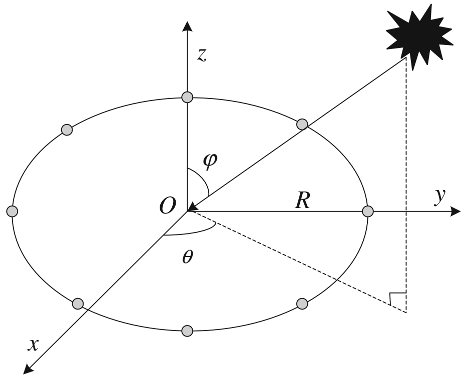
The space source signals analyzed in this thesis are regarded as narrow-band signals (compared with the signal of several MHz and even wider bandwidth in TT&C system). The narrow-band signal has a rather narrow bandwidth envelope (slowly changing) compared with the signal (complex signal) of the carrier frequency. Therefore, at the same moment the different effects of the received signal of each array element vary in different phase, which is only decided by the wave journey of the signal received by each array element.

The typical array types applied in 2-D or more dimensional direction finding system are such as: UCA, non-uniform circle array, L-shaped array, Y-shaped array etc. These direction finding system may obtain both azimuth and elevation angle. The following is to introduce array manifold of the UCA and L-shaped array and compare them in detail.

11.2.1 UCA

The UCA is a typical surface array. Its array elements are distributed uniformly around the circle. It may provide azimuth and elevation angle information which is 360° omnidirectional and none-fuzzy at the same time. Assume that there are M array elements distributed uniformly around the circle whose radius is R on the plane XOY. They are same and omnidirectional seen in Fig. 11.1. Taking Spherical coordinate system into consideration, its origin O is located in the center of the

Fig. 11.1 UCA schematic



circle array. Azimuth and the elevation angle of the incident signal is θ and φ. The angle φ is the intersection angle between the shadow which the incident signal ray makes on the plane XOY and axis X. The angle φ is the intersection angle between the incident signal ray and axis Z. The arrival of azimuth is θ and the arrival of elevation angle is φ.

If there are N incident sources of narrow-band signals, the time delay between the reference array element and the $k(1,2,\ldots,M)$ element caused by a certain incident signal $i(1,2,\ldots,N)$.

$$\tau_{k,i} = \frac{R}{c}\left(\cos\left(\frac{2\pi(k-1)}{M} - \theta_i\right)\sin\phi_i\right) \tag{11.1}$$

If A is the steering vector of the assuming array above, it may be a $M \times N$ direction matrix, as follows:

$$A = [A_1, A_2, \ldots, A_N] \tag{11.2}$$

$$A_i = a(\theta_i, \phi_i), \quad i = 1, \ldots, N \tag{11.3}$$

Where $a(\theta_N, \phi_N)$ denotes the steering vector of the Nth signal source given by:

$$a(\theta_N, \phi_N) = \begin{bmatrix} 1 & \mathrm{e}^{-j2\pi c\tau_{2,N}/\lambda} & \ldots & \mathrm{e}^{-j2\pi c\tau_{M,N}/\lambda} \end{bmatrix} \tag{11.4}$$

Taking Eq. (11.4) into the above steering vector:

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathrm{e}^{-j2\pi c\tau_{2,1}/\lambda} & \mathrm{e}^{-j2\pi c\tau_{2,2}/\lambda} & \cdots & \mathrm{e}^{-j2\pi c\tau_{2,N}/\lambda} \\ \vdots & \vdots & & \vdots \\ \mathrm{e}^{-j2\pi c\tau_{M,1}/\lambda} & \mathrm{e}^{-j2\pi c\tau_{M,2}/\lambda} & \cdots & \mathrm{e}^{-j2\pi c\tau_{M,N}/\lambda} \end{bmatrix} \tag{11.5}$$

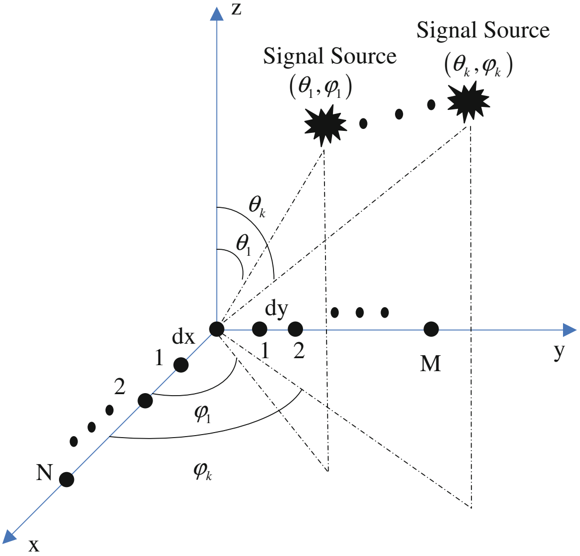


Fig. 11.2 L-shaped array schematic

11.2.2 L-Shaped Array

The L-shaped antenna array is a form of URA (Uniformly Rectangular Array) with a certain number of elements in the plane. It consists of N identical isotropous antennas placed on X-axis and M identical isotropous antennas placed on Y-axis, as shown in Fig. 11.2. The spacing between neighbor columns and neighbor rows is dx and dy, respectively. Suppose that there are k narrow-band sources, $s_i(t)(i = 1, \ldots, k)$, with the same wavelength λ impinging on the array from directions $(\theta_1, \varphi_1), (\theta_2, \varphi_2), \ldots, (\theta_k, \varphi_k)$, where θ_i and φ_i denote the elevation angle and the azimuth angle of arrival of the ith signal, respectively.

The corresponding array steering vector of N identical isotropous antennas placed on X-axis is given by:

$$A_x = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathrm{e}^{-j2\pi d \sin\varphi_1 \cos\theta_1/\lambda} & \mathrm{e}^{-j2\pi d \sin\varphi_2 \cos\theta_2/\lambda} & \cdots & \mathrm{e}^{-j2\pi d \sin\varphi_k \cos\theta_k/\lambda} \\ \vdots & \vdots & & \vdots \\ \mathrm{e}^{-j2\pi d(N-1) \sin\varphi_1 \cos\theta_1/\lambda} & \mathrm{e}^{-j2\pi d(N-1) \sin\varphi_2 \cos\theta_2/\lambda} & \cdots & \mathrm{e}^{-j2\pi d(N-1) \sin\varphi_k \cos\theta_k/\lambda} \end{bmatrix} \tag{11.6}$$

The corresponding array steering vector of M identical isotropous antennas placed on Y-axis is given by:

$$A_y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-j2\pi d \sin\theta_1 \sin\varphi_1/\lambda} & e^{-j2\pi d \sin\theta_2 \sin\varphi_2/\lambda} & \cdots & e^{-j2\pi d \sin\theta_k \sin\varphi_k/\lambda} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi d(M-1) \sin\theta_1 \sin\varphi_1/\lambda} & e^{-j2\pi d(M-1) \sin\theta_2 \sin\varphi_2/\lambda} & \cdots & e^{-j2\pi d(M-1) \sin\theta_k \sin\varphi_k/\lambda} \end{bmatrix} \tag{11.7}$$

11.3 Performance of Different Elements in the Same Array Manifold

The same array manifold and the same spatial spectrum algorithm are applied to analyze the law of the Root Mean Square Error (RMSE) of the 2-D angle (azimuth angle and elevation angle), which is decided by the number of array elements.

Take the UCA for example to be analyzed, and the spatial spectrum estimation adopts the MUSIC algorithm. The UCA applies three arrays with different number of elements, like 11, 9, 7; the biggest aperture of array is 50 m; The number of snapshots is 1,000; Assume the signal source is uncorrelated narrowband complex sinusoidal signal from the direction of arrival ($\theta = 300^\circ, \varphi = 30^\circ$), the sounding frequency is 11.8 MHz, the azimuth is θ and the elevation angle is φ.

The azimuth and elevation angle are obtained by the UCA. When signal to noise ratio (SNR) grows from -10 to 10 dB, the corresponding simulation azimuth and elevation angle might be obtained.

With the same SNR, the RMSE of azimuth and elevation angle estimation are obtained from 300 Monte-Carlo simulations. Assume the azimuth $\theta_i (i = 1, 2, \ldots, n)$ and the elevation angle $\varphi_i (i = 1, 2, \ldots, n)$, then, the corresponding RMSE of azimuth and elevation angle are respectively θ_{RMSE} and φ_{RMSE}.

$$\theta_{\mathrm{RMSE}} = \sqrt{\sum_{i=1}^{n} (\theta_i - \theta) \Big/ n} \tag{11.8}$$

$$\varphi_{\mathrm{RMSE}} = \sqrt{\sum_{i=1}^{n} (\varphi_i - \varphi) \Big/ n} \tag{11.9}$$

In Fig. 11.3, the red star represents the output of 11 array elements; the blue rhombus represents the output of 9 array elements; the black pentagram represents the output of 7 array elements.

Make the number of array element of the UCA respectively to be 7, 9 and 11. The relationship between the RMSE of the corresponding azimuth and SNR is shown in Fig. 11.3a as follows: if the number of array elements is constant, the

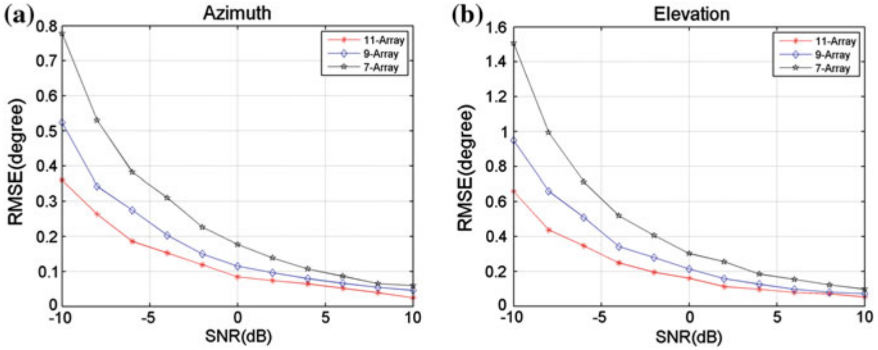


Fig. 11.3 Comparison of angle RMSE of the same array manifold with different elements. **a** Comparison of azimuth RMSE; **b** Comparison of elevation angle RMSE

RMSE of the azimuth reduces as the SNR grows from −10 to 10 dB; if the SNR is constant, the RMSE of the azimuth reduces as the number of array elements grows.

The relationship between the RMSE of the corresponding elevation angle and SNR is shown in Fig. 11.3b as follows: if the number of array elements is constant, the RMSE of the elevation angle reduces as the SNR grows from −10 to 10 dB; if the SNR is constant, the RMSE of the elevation angle reduces as the number of array elements grows.

According to the former analysis, the performance of angle estimation is the best when the number of array elements is 11; the performance of angle estimation is the worst when the number of array elements is 7; the performance of angle estimation is moderate when the number of array elements is 9, while as SNR grows, the performance of 9 elements is close to the performance of 11 elements. Thus, under the ideal condition, the number of array elements is more, the performance of angle estimation is better. However, in practice, considering the area, the cost of antenna and the demanding of location precision, the number of array elements is selected to be 9, which will be simulated in the next step.

11.4 Analysis of the Performance of L-Shaped Array and UCA

11.4.1 Comparison of the Two Array Manifold with Different SNR

Compare and analyze the L-shaped array and the UCA with different SNR. The relationship between the RMSE of the corresponding azimuth and elevation angle and SNR is obtained. The spatial spectrum estimation adopts the MUSIC algorithm; the array manifold applies 9 array elements; the biggest aperture of array is 50 m; The number of snapshots is 1,000; Assume the signal source is uncorrelated narrowband

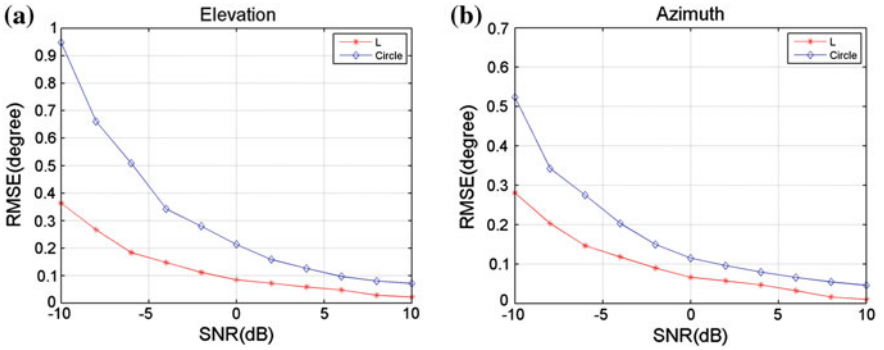


Fig. 11.4 Comparison of angle RMSE of the two array manifold with different SNR. **a** Comparison of azimuth RMSE; **b** Comparison of elevation angle RMSE

complex sinusoidal signal from the direction of arrival ($\theta = 300°, \varphi = 30°$), the sounding frequency is 11.8 MHz, the azimuth is θ and the elevation angle is φ.

The azimuth and elevation angle are obtained by the UCA and the L-shaped array. When SNR grows from −10 to 10 dB, the corresponding simulation azimuth and elevation angle might be obtained. With the same SNR, the azimuth and elevation angle are obtained by 300 Monte-carlo simulations. According to Eqs. (11.8) and (11.9), the corresponding RMSE of azimuth and elevation angle are obtained.

In Fig. 11.4, the red star represents the output of the L-shaped array; the blue rhombus represents the output of the UCA.

With the L-shaped array and the UCA, the corresponding azimuth and elevation angle are obtained respectively. The relationship between the RMSE of the corresponding azimuth and SNR is shown in Fig. 11.4a as follows: if the array manifold is fixed, the RMSE of the azimuth reduces as the SNR grows from −10 to 10 dB; if SNR is constant, the RMSE of the azimuth of the L-shaped array is smaller than that of the UCA.

The relationship between the RMSE of the corresponding elevation angle and SNR is shown in Fig. 11.4b as follows: if the array manifold is fixed, the RMSE of the elevation angle reduces as the SNR grows from −10 to 10 dB; if SNR is constant, the RMSE of the elevation angle of the L-shaped array is smaller than that of the UCA.

In a conclusion, the performance of the L-shaped array is better than that of the UCA with the different SNR.

11.4.2 Comparison of the Two Array Manifold with Different Snapshots

Compare and analyze the L-shaped array and the UCA with different snapshots. The relationship between the RMSE of the corresponding azimuth and elevation angle and the number of snapshots is obtained. The spatial spectrum estimation adopts the MUSIC algorithm; the array manifold applies 9 array elements; the biggest aperture of array is 50 m; The SNR is 5 dB; Assume the signal source is uncorrelated narrowband complex sinusoidal signal from the direction of arrival ($\theta = 300^\circ, \varphi = 30^\circ$), the sounding frequency is 11.8 MHz, the azimuth is θ and the elevation angle is φ.

The azimuth and elevation angle are respectively obtained by the UCA and the L-shaped array. When the number of snapshots varies from 100 to 1,000, the corresponding azimuth and elevation angle might be obtained. With the same number of snapshots, the azimuth and elevation angle are obtained by 300 Monte-Carlo simulations. According to Eqs. (11.8) and (11.9), the corresponding RMSE of azimuth and elevation angle are obtained.

In Fig. 11.5, the red star represents the output of the L-shaped array; the blue rhombus represents the output of the UCA.

With the L-shaped array and the UCA, the corresponding azimuth and elevation angle are obtained respectively. The relationship between the RMSE of the corresponding azimuth and the snapshots is shown in Fig. 11.5a as follows: if the array manifold is fixed, the RMSE of the azimuth reduces as the snapshots grows from 100 to 1.000; if the number of snapshots is constant, the RMSE of the azimuth of the L-shaped array is smaller than that of the UCA.

The relationship between the RMSE of the corresponding elevation angle and the snapshots is shown in Fig. 11.5b as follows: if the array manifold is fixed, the RMSE of the elevation angle reduces as the snapshots grows from 100 to 1,000; if the number of snapshots is constant, the RMSE of the elevation angle of the L-shaped array is smaller than that of the UCA.

In a conclusion, the performance of the L-shaped array is better than that of the UCA with the different number of snapshots.

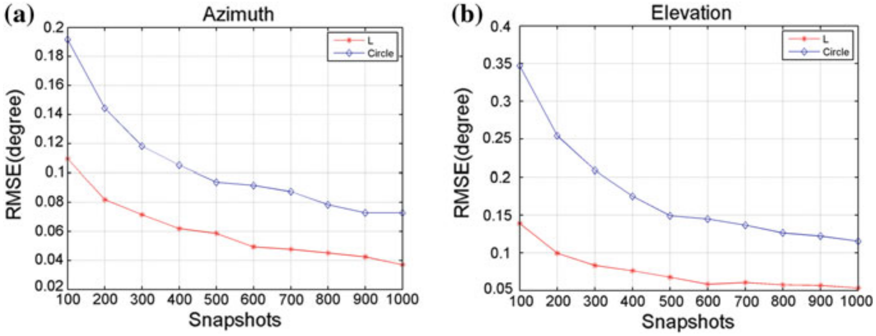


Fig. 11.5 Comparison of angle RMSE of the two array manifold with different number of snapshots. **a** Comparison of azimuth RMSE; **b** Comparison of elevation angle RMSE

11.5 Conclusion

Based on the simulation analysis of the 2-D array spectrum estimation, it is concluded that: (1) The RMSE of the corresponding angle of both the L-shaped array and UCA reduces as SNR grows; (2) The RMSE of the corresponding angle of both the L-shape array and UCA reduces as Snapshots grows; (3) If SNR is constant, the RMSE of angle of the L-shaped array is smaller than that of the UCA; (4) If the number of snapshots is constant, the RMSE of azimuth of the L-shaped array is smaller than that of the UCA. Hence, under the same simulation condition, the performance of the L-shaped array is better than that of the UCA.

The relevant algorithm and technology may be applied in the passive detection system which estimates the incident wave with the combination of the known target signal and the mobile communication reflection signal. Consequently, the application area of the traditional TT&C system is expanded, since the non-cooperative target may be detected and located. If the signal of the mobile communication network is applied to passively detect and locate the space target, its antenna may be properly designed and distributed according to the space location and detection range of the detection target.

References

1. Blackman RB, Tukey JW, Tukey JW et al (1959) The measurement of power spectra: from the point of view of communications engineering. Dover, New York
2. Blackman RB, Tukey JW, Teichmann T (1960) The measurement of power spectra. Phys Today 13:52
3. Burg JP (1968) A new analysis technique for time series data. NATO advanced study institute on signal processing with emphasis on underwater acoustics, pp 12–23
4. Burg JP (1972) The relationship between maximum entropy spectra and maximum likelihood spectra. Geophysics 37(2):375–376
5. Capon J, Greenfield RJ, Kolker RJ (1967) Multidimensional maximum-likelihood processing of a large aperture seismic array. Proc IEEE 55(2):192–211
6. Capon J (1983) Maximum-likelihood spectral estimation. Nonlinear methods of spectral analysis. Springer, Berlin, pp 155–179
7. Schmidt R (1986) Multiple emitter location and signal parameter estimation. IEEE Trans Antennas Propag 34(3):276–280
8. Roy R, Paulraj A, Kailath T (1986) ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise. IEEE Trans Acoust Speech Signal Process 34(5):1340–1342
9. Roy R, Kailath T (1989) ESPRIT—estimation of signal parameters via rotational invariance techniques. IEEE Trans Acoust Speech Signal Process 37(7):984–995
10. Brennan LE, Reed LS (1973) Theory of adaptive radar. IEEE Trans Aerosp Electron Syst 2:237–252
11. Gardner W, Franks L (1975) Characterization of cyclostationary random signal processes. IEEE Trans Inf Theory 21(1):4–14
12. Gardner WA, Napolitano A, Paura L (2006) Cyclostationarity: half a century of research. Sig Process 86(4):639–697
13. Jha Sanjay, Durrani Tariq (1991) Direction of arrival estimation using artificial neural networks. IEEE Trans Syst Man Cybern 21(5):1192–1201

Chapter 12
Research on Analysis and Numerical Simulation of Wind-Load on Observatory Antenna

Demin Qiu, Diyu Zhou, Qi Zhang and Qing Sun

Abstract For large observatory antenna working in the open air, wind-load is the primary disturbance which should be considered because it not only influences the structure strength of the antenna, but also causes distorts of emitting area and the back-frame, which influences the antenna's precision seriously and creates focal diverge. In order to analyze the effects of the wind-load on the antenna, a dynamic model is presented based on the computational wind engineering, which is deduced with the relation between wind speed and wind-load torques acting on the azimuth and elevation axis, through coordinate transformation. Finally, a numerical simulation analyzing of the effects of wind-load on the antenna is carried out using MATLAB, which reveals the variety law of wind-load under a typical trajectory of antenna.

Keywords Wind-Load · Coordinate transformation · Numerical simulation · Dynamic emulation

12.1 Introduction

Antenna is interfered by load mainly includes wind-load, snow/ice-load, inertia-load while turning round, self-weight-load and so on [1]. Now we can make self-weight-load become a subordination influence factor by conformal design or using new materials. But environment-load, especially wind-load, influences the control quality of antenna a lot, which turns into a prominent problem to be solved. The wind is one of the most common phenomenon of nature, so all kinds of outdoor-working equipment need to consider the effects of wind-load while operating [2]. Limited by conditions of engineering, large observatory antenna mostly work in outdoor environment without protection of shield, and wind-load become the main

---

D. Qiu (✉) · D. Zhou · Q. Zhang · Q. Sun
Qingdao Station, Xi'an Satellite Control Center, Qingdao 266114, China
e-mail: murray001@sina.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_12

external load. That will reduce the strength of the antenna structure strength, causing focal diverge and deviation while sending or receiving radio waves. That also make the load torque of driving motor fluctuating, and affect the tracking-pointing precision of the antenna. Therefore, analyzing wind-load characteristics and variety law of wind-load and setting up the model of wind disturbance, guarantee high performance and reliability of antenna control system while designing and verifying.

12.2 Analysis of Antenna Wind-Load

The wind-load's influence of observatory antenna is quit complex. Its adverse effects will cause change and vibration of antenna structure, even affecting tracking-pointing precision of the antenna. So analysis of the composition and function is the premise of accurate simulation observatory antenna wind-load.

12.2.1 Composition of Wind-Load

Antennas that we research in this paper work in the low altitude atmosphere, where flow belongs to the turbulent flow, which means viscous Newtonian fluid with low speed and no compression [3]. Wind is caused by different pressure of air, and the relative motion between air and objects causes wind-load. The surface force of object by wind is mainly made up of friction and pressure drag. For streamline body, mainly consider friction rather than pressure drag. As for non-streamline body, pressure drag should be mainly considered, because of the vortex that formed on the underside of antenna by wind [4].

After antenna structure bearing wind-load, wind generates forward wind and lateral wind, and wind torque which acting on the antenna azimuth and elevation axis. According to the speed record of forward wind, velocity-time curve of wind consists of two components. The period of one of them is more than 10 min, and the other one's period is short to a few seconds or dozens of seconds. They become average wind speed and pulsating wind speed respectively [5]. Relative to the resonant frequency of antenna's mechanical structure, the effect of average wind can be considered as a static force, and due to high frequency, energy is very small, the pulsating wind doesn't affect antenna much. Figure 12.1 shows the forward wind speed in some time, we divide the wind into average wind and pulsating wind according to the period. And in Fig. 12.1, $\bar{V}$ represents the average wind speed, while V_f is on behalf of the pulsating wind speed.

Average wind means that we consider the direction, speed and other quantity of wind that acting on antenna as constants, which do not change over time. Because the wind's period is longer than the natural vibration period of large observatory antenna structure, its effect is similar to the static effect. And when analyzing what

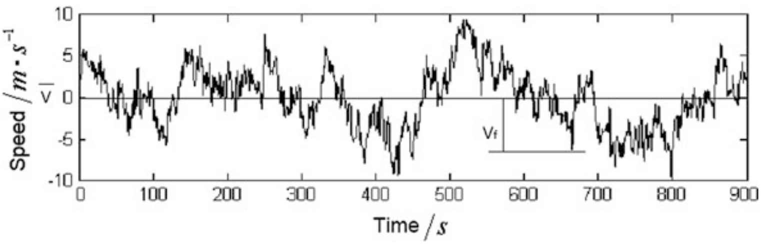


Fig. 12.1 Record of forward wind speed

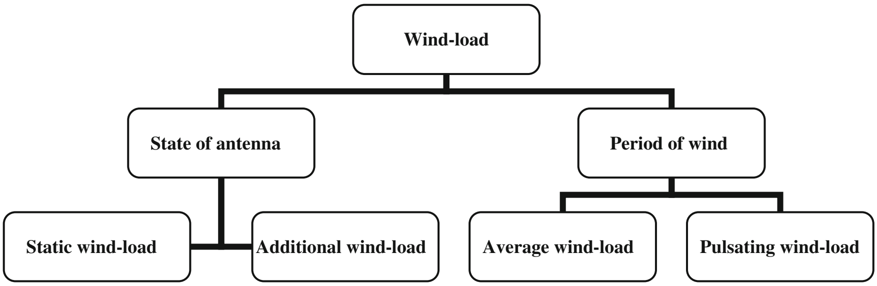


Fig. 12.2 Classification of wind-load

effect does the average wind act on antenna, it can be treated as static effect. While the strength of pulsating wind is basically a stationary stochastic process, and its effect on the antenna has the characteristics of short and irregular period. So it shall be carried out in accordance with the properties of dynamic effect analysis [6].

In addition, antenna wind-load caused by average wind can be divided into static wind-load and additional wind-load when the antenna is stationary and rotating respectively. The classification of wind-load is shown in Fig. 12.2 according to the effect.

12.2.2 Force Analysis of the Rotation Axis of Antennas

Large observatory antennas usually use circular paraboloid (rotating paraboloid) as the main reflector. In 2D (two-dimensional) coordinate system, the wind-load components is shown in Fig. 12.3.

The three components of wind-load can be expressed as:

$$F_C = C_C q A \tag{12.1}$$

$$F_A = C_A q A \tag{12.2}$$

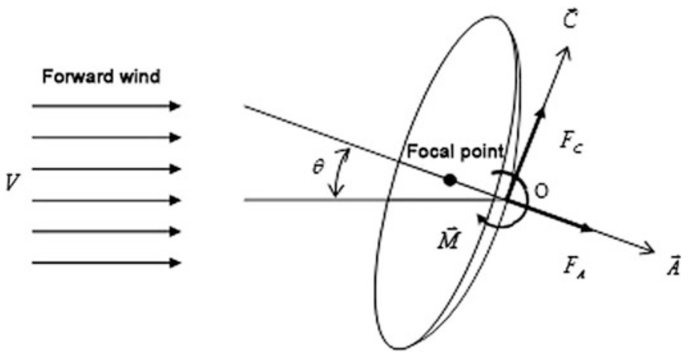


Fig. 12.3 Wind-load of circular paraboloid reflector

$$M = C_M q A \tag{12.3}$$

In expression (12.1) to expression (12.3), F_C, F_A, M, C_C, C_A and C_M represent axial force, lateral force, wind torque, axial wind-load coefficient, lateral wind-load coefficient and wind torque coefficient respectively, q is dynamic pressure, and A is Antenna windward area. Because the circular paraboloid has symmetry, in 2D coordinate system, wind-load coefficient variety law is the same when antenna rotates around azimuth axis and elevation axis. When calculating the actual wind-load on antenna, the 2D coordinate system should be converted to 3D (three-dimensional) coordinate system first [7]. The relationship between these two kinds of coordinate system is shown in Fig. 12.4.

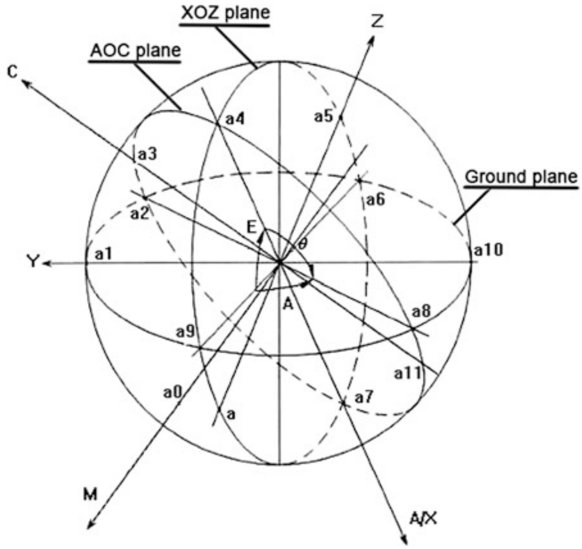


Fig. 12.4 Relationship between 2D coordinate system and 3D coordinate system

In Fig. 12.4, A-C-M coordinate system stands for 2D coordinate system, while x-y-z coordinate system represents 3D coordinate system. When doing coordinate transformation, the structure of antenna can be regarded as rigid body, so the wind-load coefficient can convert A-C-M coordinate system to x-y-z coordinate system through force system translation. After conversion of 3D coordinate system, the expression of wind-load coefficient is:

$$\begin{cases} C_x = C_A \\ C_y = C_C \sin A/\sqrt{1-\cos^2 E\cdot\cos^2 A} \\ C_{MX} = 0 \\ C_{MY} = -C_M \sin E\cos A/\sqrt{1-\cos^2 E\cdot\cos^2 A} \\ C_{MZ} = C_M \sin A/\sqrt{1-\cos^2 E\cdot\cos^2 A} \end{cases} \tag{12.4}$$

In expression (12.4), C_x, C_y and C_z represent the wind-load coefficient of 3D coordinate system; C_{MX}, C_{MY} and C_{MZ} represent the wind-load torque coefficient of 3D coordinate system; C_C, C_A and C_M represent the axial wind-load coefficient, lateral wind-load coefficient and wind torque coefficient of 2D coordinate system.

In practical engineering, the peak point of observatory antenna reflector is usually not on the rotating axis. So when calculating the wind-load on antenna rotating axis, we need to shift the wind-load effect of reflector peak point into rotating axis by force system translation. The translation process is shown in Fig. 12.5.

From Figs. 12.4 and 12.5, we can see the corresponding relations between peak point of circular paraboloid antenna and wind-load. Thus, we can get the expressions of wind-load coefficient of antenna rotating axis:

$$\begin{cases} {C_x}' = C_x = C_A(A,E) \\ {C_y}' = C_y = C_C(A,E)\sin A/\sqrt{1-\cos^2 E\cdot\cos^2 A} \\ {C_z}' = C_z = C_C(A,E)\sin E\cos A/\sqrt{1-\cos^2 E\cdot\cos^2 A} \\ {C_{MX}}' = C_{MX} = 0 \\ {C_{MY}}' = C_{MY} = -C_M(A,E)\sin E\cos A/\sqrt{1-\cos^2 E\cdot\cos^2 A} \\ {C_{MZ}}' = C_{MZ} = C_M(A,E)\sin A/\sqrt{1-\cos^2 E\cdot\cos^2 A} \end{cases} \tag{12.5}$$

From expression (12.5), the expressions of static wind torque coefficient of antenna azimuth axis is available:

$$C'_{MZA} = \left\{ \begin{array}{l} (a/D)\sin A\cdot C_C(A,E) + \\ {[(b/D)C_C(A,E) + C_M(A,E)]}\sin A\cdot\cos E \end{array} \right\} / \sqrt{1-\cos^2 E\cdot\cos^2 A} \tag{12.6}$$

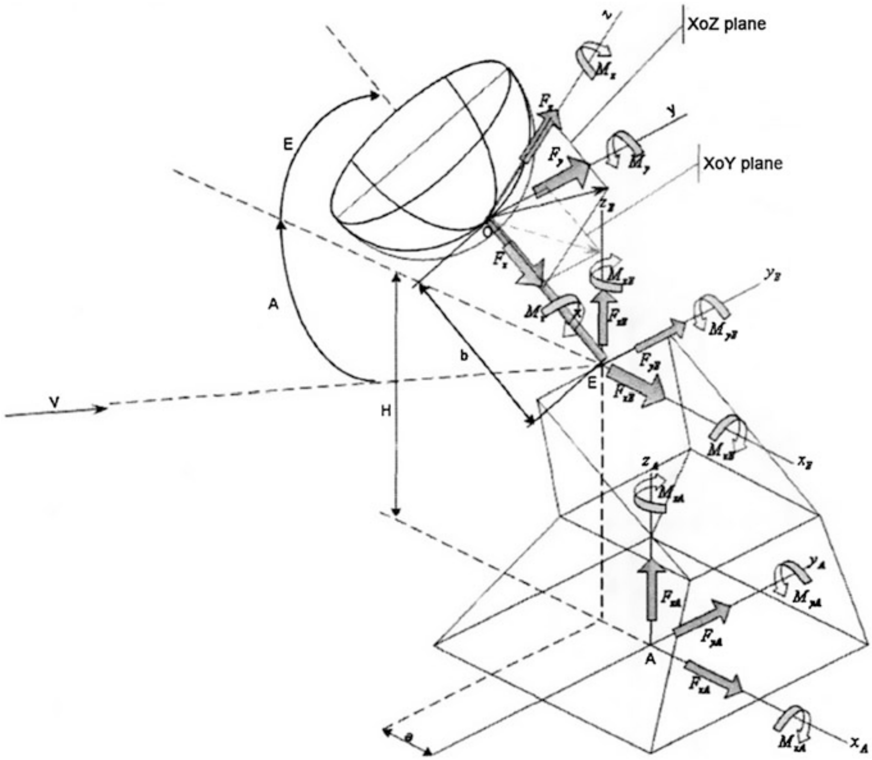


Fig. 12.5 Force system translation in 3D coordinate system

And from expression (12.6), the expression of static wind torque coefficient of antenna elevation axis is gained:

$$C'_{MYE} = -\sin E \cdot \cos A[(b/D)C_C(A,E) + C_M(A,E)]/\sqrt{1-\cos^2 E \cdot \cos^2 A} \quad (12.7)$$

In expression (12.5)–(12.7), azimuth angle $A \in [0°, 360°)$, elevation angle $E \in [-90°, 90°], \theta = \cos^{-1}[\cos A \cdot \cos E], (\theta \in [0°, 180°])$ and $C_C(A,E)$ is the wind-load coefficient of 3D coordinate system.

12.3 Numerical Simulation of Antenna Wind-Load

Observatory antenna will always be under the interference of wind when operating, so wind-load is changing with the variation of antenna attitude and wind power. The design of traditional antenna control system only think of average wind, and wind-load is compensated as a constant interference. But if the antenna is quite large, the additional wind-load and effect of pulsating wind caused by antenna rotating will not be ignored.

12.3.1 Dynamic Wind-Load Model of Antenna

The interference of antenna control system by wind-load is that, the force in antenna reflector formed by wind-load will eventually be passed to the antenna drive motor axis. That will reduce the dynamic performance of servo system, and even will make control system unable to get closed loop normally. The high frequency pulsation component of the wind may also cause self-oscillation of the antenna structure. Thus, when establishing dynamic wind-load model, we should consider the effects of both average and pulsating wind. Figure 12.6 shows pulsating wind power spectrum of different elevation.

As we can see in Fig. 12.6, most energy of the pulsating wind spectrum is concentrated in low frequency band [8]. The energy of pulsating wind rapidly reduce with the increase of frequency, and normally peak frequency of pulsating wind concentrate between 0 and 1 Hz. While the mechanical resonance frequency of large observatory antenna is generally greater than 4 Hz, which is far away from the peak frequency of pulsating wind spectrum. Therefore, when calculating the pulsating wind effects on the antenna, we can only consider the wind vibration effect of low frequency pulsating wind.

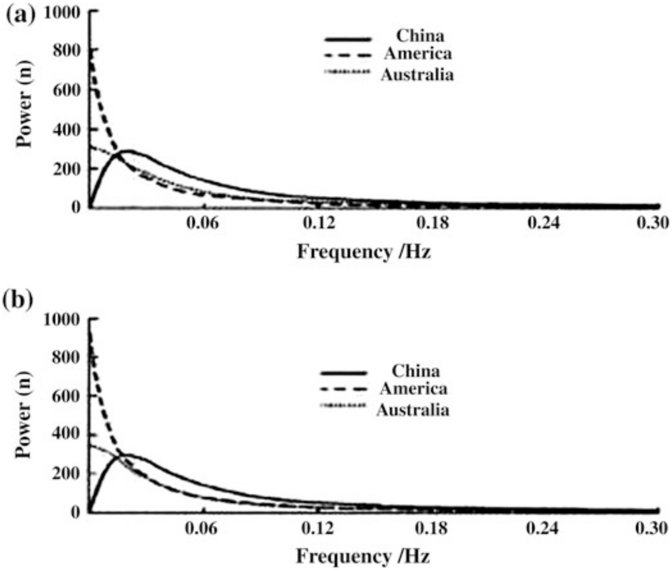


Fig. 12.6 Pulsating wind power spectrum

It may be required that antenna should be work precisely when wind scale is 8, and should not be damaged when wind scale is 12. According to the standard elevation of 10 m and the standard time of 10 min, the average wind speed is 20.7 m/s when wind scale is 8, and 32.7 m/s when 12.

The function expression of dynamic wind-load is:

$$M_{WA} = 0.5C_{MA}\rho(V + V_f \cdot \sin(2\pi ft))^2 AD \tag{12.8}$$

$$M_{WE} = 0.5C_{ME}\rho(V + V_f \cdot \sin(2\pi ft))^2 AD \tag{12.9}$$

In expression (12.8) and (12.9), V is the average wind speed, V_f is the peak speed of pulsating wind, f represents the frequency of pulsating wind, A represents the windward area of antenna, D represents the diameter of antenna, C_{MA} stands for the wind torque coefficient of antenna azimuth axis, and C_{ME} stands for the wind torque coefficient of antenna elevation axis.

12.3.2 Wind-Load Dynamic Simulation

Figure 12.7 shows that azimuth and elevation angle of antenna is get by the actual angle's horizontal projection when tracking satellite. And when building the simulation model of wind-load, we set the wind blows along the shortcut direction of satellite trajectory to the antenna reflector in front.

By the conversion relationship between 2D coordinate system and 3D coordinate system shown in Fig. 12.5, we can get the variation law of static wind torque coefficient of azimuth and elevation axis in 3D coordinate system. Figures 12.8 and 12.9

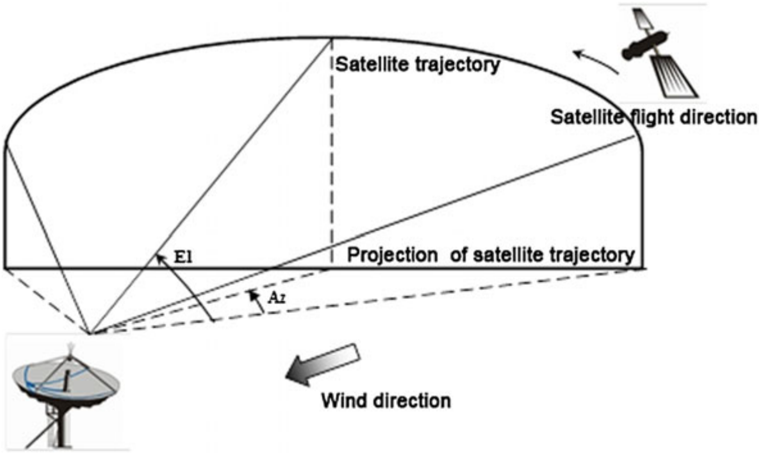


Fig. 12.7 Trajectory of antenna and satellite

Fig. 12.8 3D diagram of static wind torque coefficient of azimuth axis

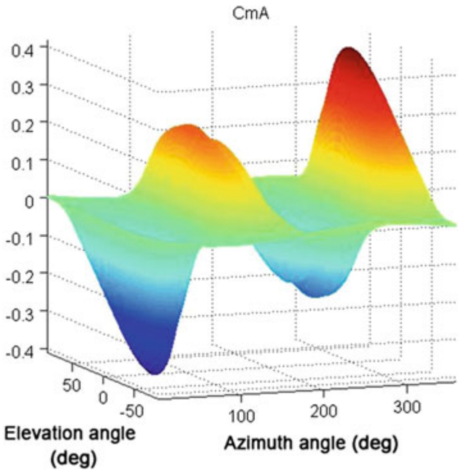
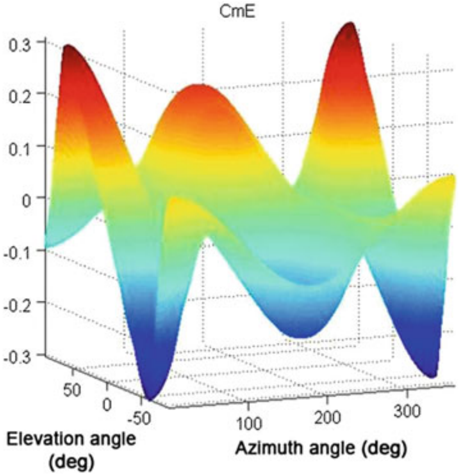


Fig. 12.9 3D diagram of static wind torque coefficient of elevation axis



shows 3D diagrams of static wind torque coefficient of azimuth and elevation axis when antenna has different attitude angle.

In Figs. 12.8 and 12.9, we select elevation angle 15°, 30°, 45°, 60°, 75° and 80°, and do the profile respectively. Then we get the diagrams of static wind torque coefficient of azimuth and elevation axis under the typical condition of elevation angle while azimuth angle changes from 0° to 180°, as shown in Figs. 12.10 and 12.11.

From Figs. 12.8, 12.9, 12.10 and 12.11, it can be seen that the static wind torque coefficient of antenna azimuth axis is about azimuth angle 180° odd symmetry, and also about elevation angle 0° even symmetry; the static wind torque coefficient of

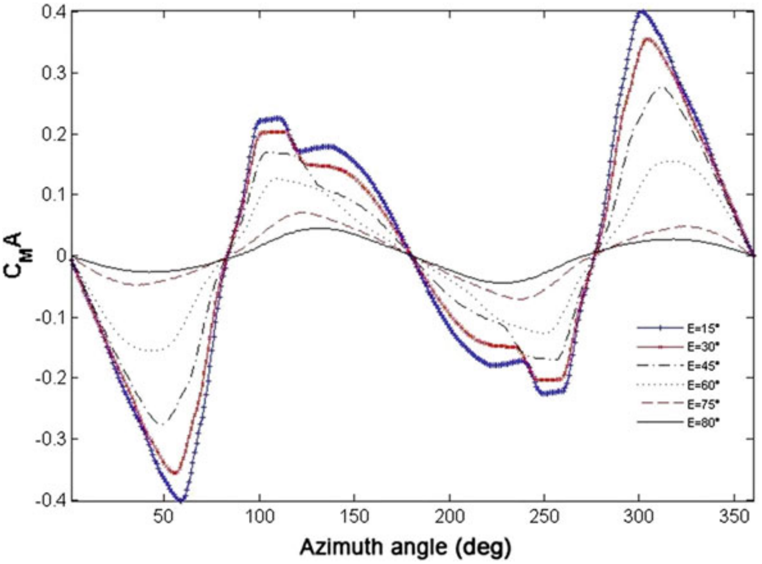


Fig. 12.10 Diagram of static wind torque coefficient of azimuth axis

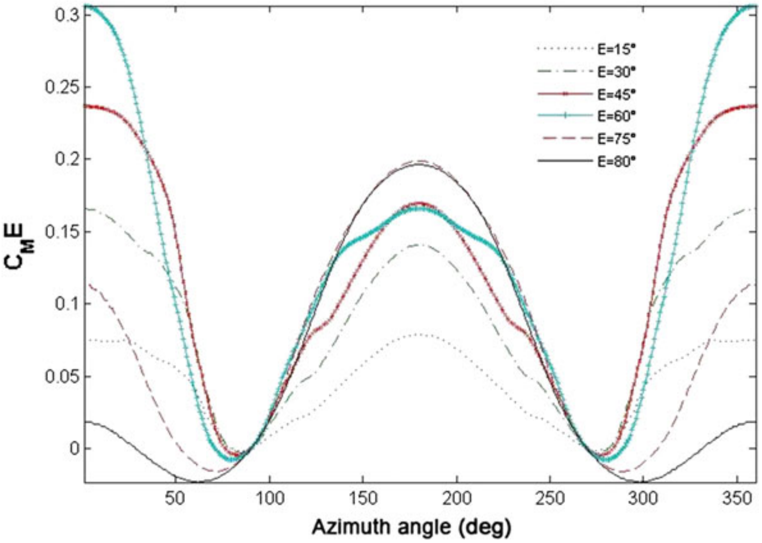


Fig. 12.11 Diagram of static wind torque coefficient of elevation axis

Fig. 12.12 3D diagram of wind torque of azimuth axis

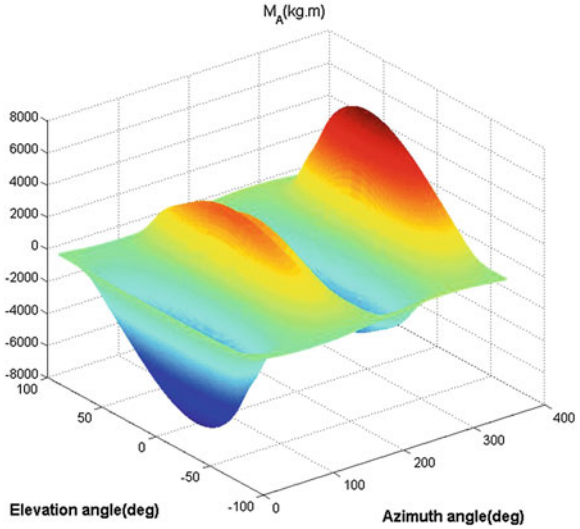
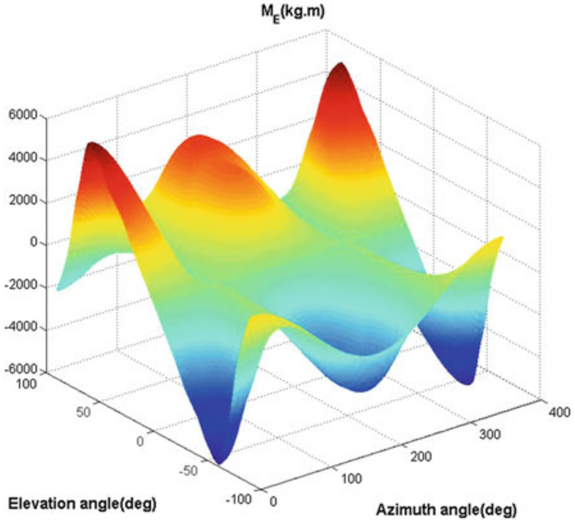


Fig. 12.13 3D diagram of wind torque of elevation axis



antenna elevation axis is about azimuth angle 180° even symmetry, and also about elevation angle 0° odd symmetry. The results of simulation calculation accord with the law of practical experiment data.

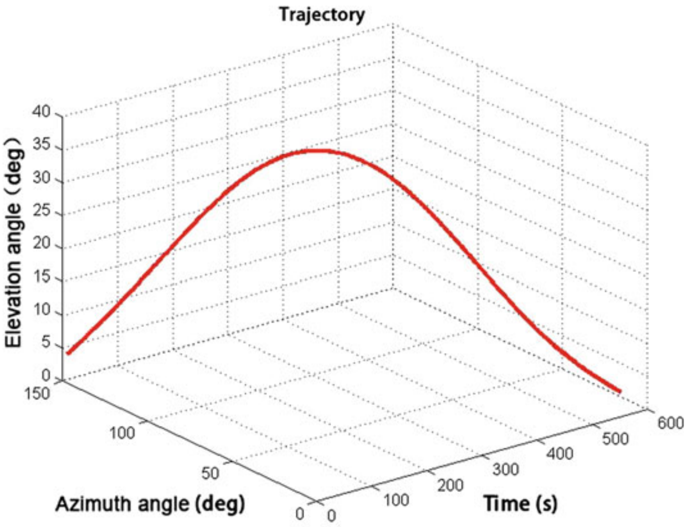


Fig. 12.14 Typical trajectory of antenna

Using expression (12.8) and (12.9), combined with Figs. 12.8 and 12.9, by simulation program, dynamic wind torque model of antenna azimuth and elevation axis is available, as shown in Figs. 12.12 and 12.13.

12.3.3 Analysis of Wind-Load under the Typical Antenna Trajectory

Most of the near-earth spacecraft use recursive orbit, and the principle of ground antenna tracking spacecraft is shown in Fig. 12.7. When antenna tracking satellite, its posture also changes real-time. The projection path of azimuth and elevation angles in 3D space is shown in Fig. 12.14. The variety law of wind torque of azimuth and elevation axis is shown in Figs. 12.15 and 12.16. Thus, the wind-load interference model will be close to the real value, and will add the wind torque of azimuth and elevation axis to the azimuth and elevation branch in antenna control system simulation model.

From Figs. 12.15 and 12.16, we can see that, when practical tracking satellites, the variety law of wind-load torque on antenna azimuth and elevation axis conform to the trend shown in Figs. 12.10 and 12.11.

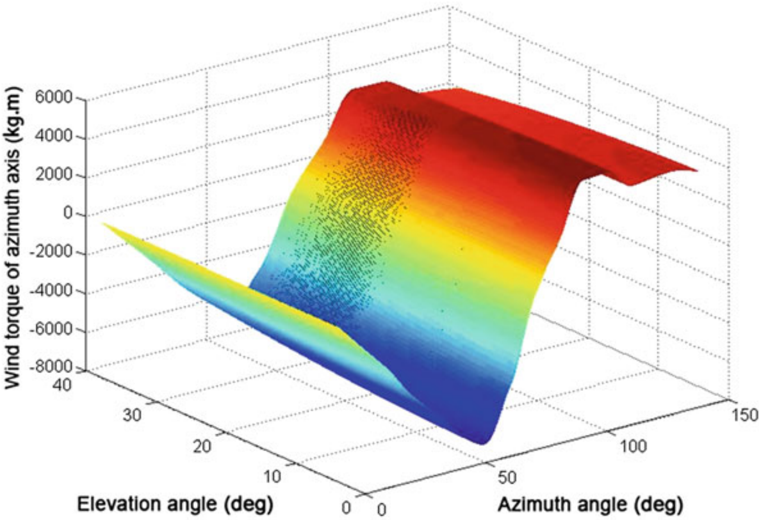


Fig. 12.15 3D diagram of wind torque of azimuth axis in typical trajectory of antenna

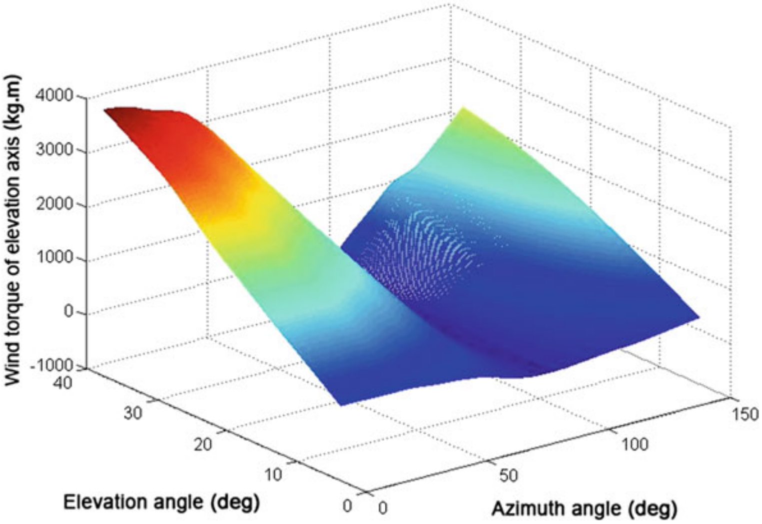


Fig. 12.16 3D diagram of wind torque of elevation axis in typical trajectory of antenna

12.4 Conclusion

For characteristics of wind-load on large observatory antenna, according to the fundamental relation between wind speed and wind-load, based on the coordinate system transformation, dynamic wind-load function expression is deduced. In order to verify the feasibility and correctness of the algorithm, we use MATLAB language to code a computer simulation program. Then we execute numerical simulation of the wind-load in typical trajectory of antenna when tracking satellites. The results of simulation show that wind-load on antenna azimuth axis achieve maximum when antenna is in low elevation angle and broadside on the wind, while wind-load on antenna elevation axis achieve maximum when antenna is in high elevation angle and broadside on the wind. Dynamic wind-load value and its variety law generated by the simulation are consistent with the actual situation. This study has practical reference value to improve the ability of resistance to wind-load disturbance in observatory antenna designing.

References

1. Xiao WX (1998) Determination on load of the antenna of shipborne radar. Mod radar 20 (5):69–77
2. Zhai JH, Wang HX, Chen J et al (2007) Active disturbance rejection controller to compensate the wind disturbance for the optoelectronic telescope with large aperture. Opto-electronic Eng 34(12):13–16
3. Pottebaum Tait S, MacMynowski Douglas G (2004) Wind tunnel testing of a generic telescope enclosure. SPIE 5495:270–281
4. Cho Myung K, Stepp Larry M, Angeli George Z (2003) Wind loading of large telescopes. SPIE 4837:352–367
5. Davenport AG (1961) The spectrum of horizontal gustiness in high winds. Q J Roy Meteorol Soc 87:194–211
6. Zhang ZT (1997) Consideration of wind load in design of radar antenna structure. Mod Electron 4
7. Wan FL (1992) Relationship between two-dimension paraboloidal plane wind load and three-dimension space wind load. Communication and TT&C: 62–65
8. Gawronski W, Bienkiewez B, Hill R (1994) Wind induced dynamics of a deep space network antenna. Sound Vibr 178(1):66–77

Chapter 13
Research on Angle Tracking Technology of DS/FH Communication System

Changqing Li, Yong Li, Manhong Lu and Ming Chen

Abstract The direct sequence spread spectrum communication system adopts the coherent demodulation method to acquire the angular error signal at present. However, the carrier frequencies hop continuously in DS/FH communication system, and this system works in the interference environment. It will bring some difficulties to use the coherent demodulation to attain the angular error signal. For this purpose, a new dual-channel monopulse angle tracking algorithm is proposed. This algorithm can apply to dynamic interference environment. The algorithm can realized the angular error signal demodulation and automatic phase calibration under the rapid hopping in carrier frequencies condition. The test results show that this algorithm can effectively demodulate angular error signal under the interference environment, and realize the angle tracking and phase calibration function.

Keywords DS/FH communication system · Interference suppression · Angular error signal demodulation · Automatic phase calibration

13.1 Introduction

For direct sequence spread spectrum communication system, making use of traditional dual-channel monopulse angle tracking algorithm would obtain the angular error signal via correlation processing in sum and difference channels [1]. However, the traditional algorithm will bring some new problem about angle tracking in DS/FH system [2]. In particular, this system works in the complex electromagnetic environment. The jamming signal is more serious effects on different signal. In the

---

C. Li (✉) · Y. Li · M. Lu · M. Chen
Beijing Research Institute of Telemetry, Beijing 100076, China
e-mail: changqingli@sohu.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_13

mean time, DS/FH communication system occupies wide bandwidth which leads to the different delay in sum and difference channels. So it's necessary to calibrate the amplitude and phase at all frequencies within the whole bandwidth. If calibration at every frequency is utilized, it would consume much time and it is a disadvantage of improving the equipment efficiency. Aim at the mentioned issue, a new dual-channel monopulse angle tracking algorithm is proposed. This algorithm can realized the angular error signal demodulation and automatic phase calibration in an interference condition.

13.2 Angle Tracking Technology of DS/FH System in Interference Condition

There are series of problems while applying the traditional angle tracking algorithm to DS/FH system. Especially in an interference environment, the correlation peak about the interference signal is stronger and there is no way to acquire the angle error information in traditional cross-correlation method. Hereby, this paper proposes an algorithm which adapts to an interference environment. The algorithm can automatically detect the narrowband interference and suppress it so that the demodulation can be achieved. It also makes baseband equipments track the target quickly while the target appears and accomplish acquiring the angle error signal. The flow-process diagram of the algorithm is shown in Fig. 13.1.

DS/FH communication system uses dual-channel monopulse system to receive the IF signal which two signals in sum and difference channel go through the LNA and DC. After sampling the IF signal, the identification and suppression module begin to detect this signal. If any interference occurs, suppress it with related algorithm; or else do nothing.

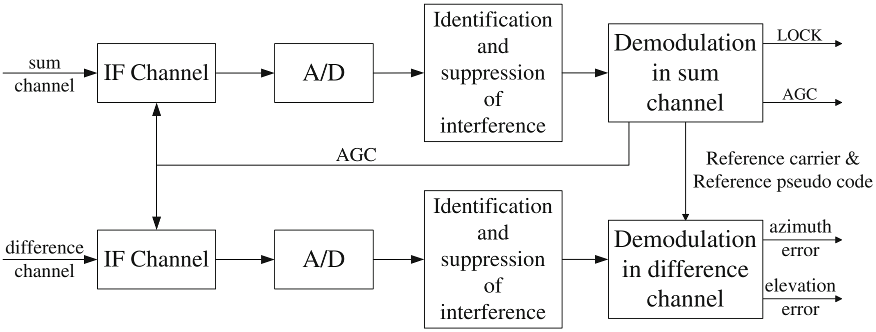


Fig. 13.1 Angular error signal flow process diagram

13.2.1 Extraction of Angle Error Signal in DS/FH Communication System

After the interference suppression, the output signal can be express as:

$$\begin{aligned} E_{\Sigma} &= d \times P_n \times E_1 \sin(\omega t) \\ E_{\Delta} &= d \times P_n \times E_2 \mu\theta \sin(\omega t + \beta) \\ &= d \times P_n \times E_2 \mu\theta (\cos\beta \sin\omega t + \sin\beta \cos\omega t) \end{aligned} \tag{13.1}$$

In this formula, θ is the deflection angle about the flying target deviate from the antenna axis. β is the angle formed the antenna azimuth axis and θ mapping in the rectangular coordinate system. μ is the antenna slope coefficient. ω is the carrier frequency. E_1 is the amplitude of the sum signal. E_2 is the amplitude of the different signal. d is the data symbol. P_n is the pseudo code.

The different signal E_{Δ} includes the azimuth angle information and the elevation angle information, besides the two carriers are direct cross. So we must adopt the method of cross-correlation demodulation to attain the error voltage [3]. The flow diagram is shown in Fig. 13.2.

The sum signal and different signal are sent into the tracking receiver to make amplification, frequency conversion and A/D conversion. Among them, the sum signal would generate the orthogonal I signal and Q signal after the sum signal go through digital IF receiver. The I signal and the Q signals can be describe as:

$$\begin{aligned} u_I &= E_3 \sin(\omega t + \varphi_1(t)) \\ u_Q &= E_3 \cos(\omega t + \varphi_1(t)) \end{aligned} \tag{13.2}$$

After the normalization processing, the differential signal can be express as:

$$u_{\Delta 1} = E_4 \mu\theta \sin(\omega t + \beta + \varphi_2(t)) \tag{13.3}$$

In the formula, E_3 and E_4 are the signals' amplitude. $\varphi_1(t)$ is the phase shift which generate by different carrier frequency pass through the sum channel. $\varphi_2(t)$ is

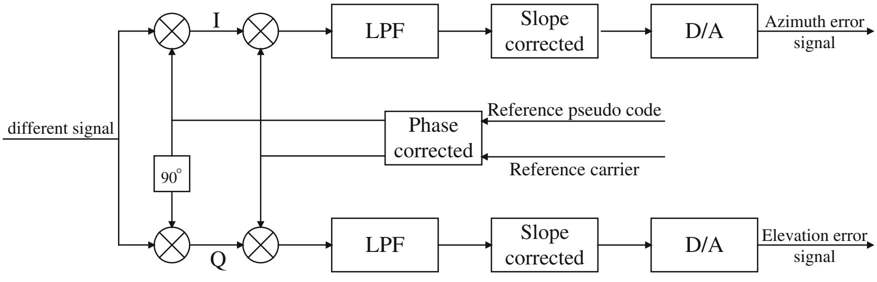


Fig. 13.2 The flow diagram of angle error demodulation in difference channel

the phase shift which generate by different carrier frequency pass through the difference channel.

If the DS/FH signal carrier frequency is f_1 right now, then the angle error signal demodulation processing can be describe as:

$$\begin{aligned} u_I \times u_{\Delta 1} &= \frac{1}{2} E_3 E_4 \mu\theta[\cos(\beta - (\varphi_{11} - \varphi_{12})) - \cos(2\omega_1 t + \beta + \varphi_{11} + \varphi_{12})] \\ u_Q \times u_{\Delta 1} &= \frac{1}{2} E_3 E_4 \mu\theta[\sin(\beta - (\varphi_{11} - \varphi_{12})) - \sin(2\omega_1 t + \beta + \varphi_{11} + \varphi_{12})] \end{aligned} \tag{13.4}$$

After the low pass filter, the receiver output the azimuth and elevation angle error signal, and its expression is:

$$\begin{aligned} \Delta u_A &= \mu\theta E_5[\cos\beta\cos(\varphi_{11} - \varphi_{12}) + \sin\beta\sin(\varphi_{11} - \varphi_{12})] \\ \Delta u_E &= \mu\theta E_5[\sin\beta\cos(\varphi_{11} - \varphi_{12}) + \cos\beta\sin(\varphi_{11} - \varphi_{12})] \end{aligned} \tag{13.5}$$

As can be seen from the formula, the part of the signal component Δu_A is contained in Δu_E, the part of the signal component Δu_E is contained in Δu_A. This signal component is called cross-coupling signal. Its value is approximately equal to $\sin(\varphi_{11} - \varphi_{12})$.

After the tracking equipment calibration, make the difference between φ_{11} and φ_{12} equal to zero. Then you can gain the real angle error signal.

$$\begin{aligned} \Delta u_A &= \mu\theta E_5 \cos\beta \\ \Delta u_E &= \mu\theta E_5 \sin\beta \end{aligned} \tag{13.6}$$

The angle error signal feed into the servo device. The device drives the antenna to move until Δu_A and Δu_E is zero, so as to realize the target tracking.

If the carrier frequency of the DS/FH system hop to f_2, the reference signal's phase shift value is still the phase shift value of f_1. It must be bring in the cross-coupling signal about the azimuth error voltage and elevation error voltage, its expression is as follows:

$$\begin{aligned} \Delta u_A &= \mu\theta E_5[\cos\beta\cos(\varphi_{21} - \varphi_{22} + \varphi_{12} - \varphi_{11}) + \sin\beta\sin(\varphi_{21} - \varphi_{22} + \varphi_{12} - \varphi_{11})] \\ \Delta u_E &= \mu\theta E_5[\sin\beta\cos(\varphi_{21} - \varphi_{22} + \varphi_{12} - \varphi_{11}) + \cos\beta\sin(\varphi_{21} - \varphi_{22} + \varphi_{12} - \varphi_{11})] \end{aligned} \tag{13.7}$$

where the $\sin(\varphi_{21} - \varphi_{22} + \varphi_{12} - \varphi_{11})$ is the cross-coupling signal. When the cross-coupling signal comes to 1, the antenna can't track the target. Therefore, the system need calibrate the phase and the directional sensitivity after the carrier frequency hop to another frequency [4].

13.2.2 The Automatic Phase Calibration in DS/FH System

Since the DS/FH communication system occupied huge bandwidth, the characteristic of the group delay of the wideband system will affect the angle measuring accuracy. The dissimilar results of the phase calibration and the directional sensitivity are obtained at the different carrier frequency [5]. It will take much time if using the phase calibration at every carrier frequency. In order to realize quickly phase calibration for this system, this paper proposes an automatic phase calibration based on the normalization carrier frequency. This algorithm breaks through the bottleneck of the phase calibration of the multi-frequency points and implement quickly phase calibration based on the dual-channel monopulse system. Its core theory is to normalizing the multi-frequency points to the center frequency point, and then does the calibration at the center frequency point. Finally, the results mapped to each frequency according to the frequency corresponding criterion so that the phase calibration of DS/FH system can be achieved. Specific phase calibration flow chart is shown in Fig. 13.3.

In cooperation with the standard signal tower, the system automatically adjusts the phase shifter of the difference channel to make sure the servo system can track the target normally [6]. The monitor system schedules the baseband equipment and the ACU equipment together to complete the phase calibration.

Step 1. Start the automatic phase calibration program and set the carrier frequency as the center frequency of the DS/FH signal.

Step 2. The antenna aim at the beacon machine, then adjust the tracking receiver to make the angle error voltage minimum or to be zero. Record the azimuth A_0 and the elevation E_0.

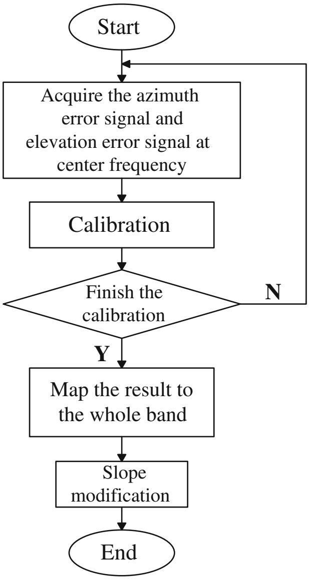


Fig. 13.3 The automatic phase calibration algorithm flow chart

Step 3. Keep the value of the azimuth and alter the elevation by θ_E. When the receiver tracks the target, record the voltage value of the angle error.

Step 4. Keep the value of the elevation and alter the azimuth by θ_A. When the receiver tracks the target, record the voltage value of the angle error.

Step 5. Calculate the accurate phase value by φ_0 according to the angle error voltage. It is the result of the phase calibration.

Step 6. Map the result of the phase calibration to the whole band. Assuming the center frequency of the receiver is f_0, the each carrier frequency of the DS/FH system is that the interval frequency of the FH system times the sequence number of carrier frequency plus the initial frequency. The formula can be describe as: $\varphi = \varphi_0 \times \frac{f_1 + \Delta f \times n}{f_0}$. If the result of the phase calibration greater than 2π, then the integer multiple of 2π subtracted from the result gives the final phase calibration result. It is a completion of the automatic phase calibration in DS/FH system.

Much time can be saved by using this method to complete the calibration in DS/FH system. Also it improves the efficiency of tracking receiver and ensures the accuracy of the system.

13.2.3 The Influence of the Angle Tracking Performance by Using Interference Suppression Algorithm

Because the system works in an interference environment, use of the interference suppression algorithm will bring new problem to angle measurement and tracking. This paper adopts the interference suppression algorithm based on transform domain [7] and the flow chart is shown in Fig. 13.4. To avoid the dissemination of the interferer's energy over a wide frequency range, a windowing function can be used prior to the transformation processes. After the windowing, the input signal is the Fourier transformed to get the spectral line of signal. Based on the amplitude of the DS/FH signal and the interference signal to detect the position of the interference signal, and then suppress the interference at frequency domain. After the interference is suppressed, convert signal to time domain by IFFT algorithm.

When suppressed the interference to input data, the data will generate the sequence overlapping $x'(n)$, the conventional diagram is shown in Fig. 13.5.

After the windowing and the Fourier transform, the data stream can be express as:

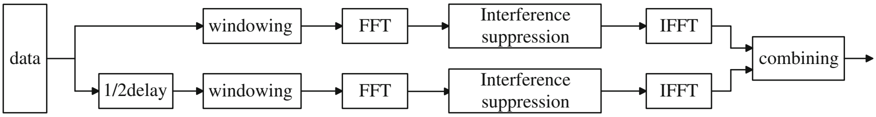


Fig. 13.4 The interference suppression algorithm flow diagram

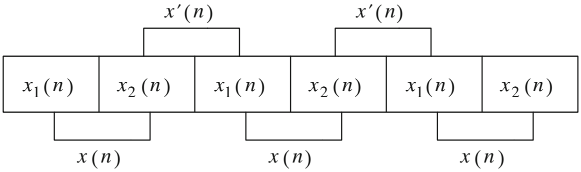


Fig. 13.5 Data stream of the input signal

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n)w(n)e^{-j\frac{2\pi}{N}kn} \quad 0 \le k \le N-1 \\ X'(k) &= \sum_{n=0}^{N-1} x'(n)w(n)e^{-j\frac{2\pi}{N}kn} \quad 0 \le k \le N-1 \end{aligned} \tag{13.8}$$

If detect the interference at $k_i\,N-k_i$, in $X(k)$ and $k_s, N-k_s$ in $X'(k)$, and then set it to zero, namely, $X(k_i)=0, X(N-k_i)=0, X'(k_s)=0, X'(N-k_s)=0$. After the interference is suppressed, convert signal to time domain to get the signal by $z(n)$[8].

$$\begin{aligned} z(n) = {} & 2\alpha x(n) - \frac{1}{N}X(k_i)e^{j\frac{2\pi}{N}k_i n} - \frac{1}{N}X^*(k_i)e^{-j\frac{2\pi}{N}k_i n} \\ & - \frac{1}{N}X_1(k_s)e^{j\frac{2\pi}{N}k_s\left(n+\frac{N}{2}\right)} - \frac{1}{N}X_1^*(k_s)e^{-j\frac{2\pi}{N}k_s\left(n+\frac{N}{2}\right)} \end{aligned} \tag{13.9}$$

where α is the window function. When $\alpha = 0.5$, it is the Hanning window. When $\alpha = 0.54$, it is the Hamming window.

It can be seen that this method will lose part of useful signal while suppressing the interference signal. However, after the DS/FH signal demodulation, the useful signal will compressed to a narrow band. Loss signal energy has weak effect on the signal demodulation and the angle tracking, it can be ignored.

13.3 Simulation

The model of angle error demodulation of DS/FH system is shown in Fig. 13.1, which is built in the simulink of Matlab. The simulation parameter is set as follows:

Data rate: 10 kbps
Chip rate: 10 Mbps
Frequency hopping rate: 10 khop/s
Amount of frequencies: 256
Interval frequency of the FH system: 400 kHz
Carrier frequency: 50–150 MHz
Modulation mode: BPSK

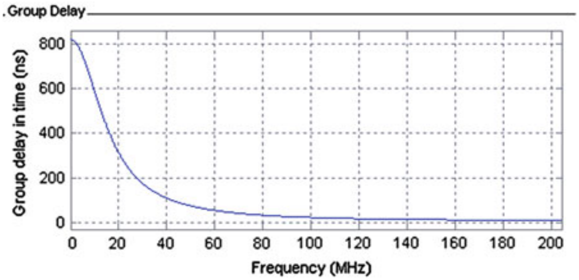


Fig. 13.6 The characteristic of the nonlinear channel

Due to the DS/FH system occupies the wide bandwidth, the nonlinear characteristics of the channel will bring some influence of the angle tracking. The module of the channel consists of the white Gaussian noise module and the nonlinear channel module. Among them, the characteristic of the nonlinear channel is shown in Fig. 13.6.

Figure 13.7 shows the value of normalization error voltage under the dissimilar channel condition. Set the signal noise ratio of the output signal as $E_b/N_0 = 13\text{dB}$. Figure 13.7a shows that the mean of the normalization error voltage is 0.9995 V and the variance of the normalization error voltage is 3.8589e−6 on the linear channel condition. This voltage result can be used for antenna tracking. Figure 13.7b shows that the mean of the normalization error voltage is 0.9658 V and the variance of the normalization error voltage is 0.1478 on the non-linear channel condition. This voltage result varies obviously, it can't be used for antenna tracking.

Then add 40 dB wideband interference and 60 dB narrowband interference at the center frequency, and the narrowband interference bandwidth would be 5 % of the system bandwidth. The angle tracking system adopts the dual-channel monopulse system. Demodulate the angle error signal coherently by making use of the carrier phase and the pseudo code phase. Observe the angle tracking performance while adopting the overlapping and windowing interference suppression algorithm and not adopting this algorithm.

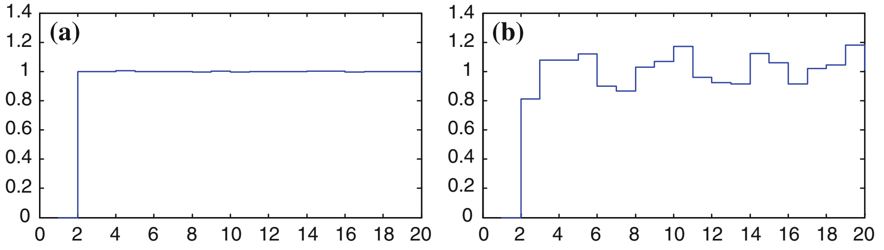


Fig. 13.7 The value of the normalization error voltage under the dissimilar channel condition. **a** The value of the normalization error voltage on the non-linear channel condition. **b** The value of the normalization error voltage on the linear channel condition

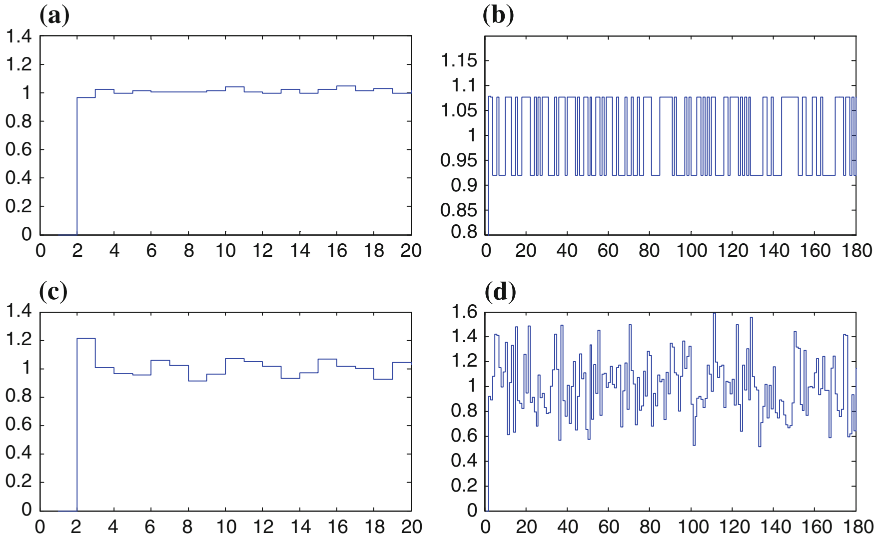


Fig. 13.8 The value of the normalization error voltage. **a** Join the narrowband interference with the interference suppression module. **b** Join the narrowband interference without the interference suppression module. **c** Join the narrowband interference and wideband interference with the interference suppression module. **d** Join the narrowband interference and wideband interference without the interference suppression module

Join the 60 dB narrowband interference in DS/FH system and observe the value of the normalization error voltage. Figure 13.8a is the simulation result which adding the interference suppression module in receiver. By calculating, we can conclude that the mean of the normalization error voltage is 1.0005 V and the variance of the normalization error voltage is 1.1636e−4. This result can be used for antenna tracking. Figure 13.8b is the simulation result without the interference suppression module in receiver. After calculating, we can get that the mean of the normalization error voltage is 0.9914 V and the variance of the normalization error voltage is 0.8013. This voltage result varies obviously, it can't be used for antenna tracking.

Join the 60 dB narrowband interference and 40 dB wideband interference in DS/FH system and observe the value of the normalization error voltage. Figure 13.8c is the simulation result which adding the interference suppression module in receiver. By calculating, we can conclude that the mean of the normalization error voltage is 0.9442 V and the variance of the normalization error voltage is 0.0312. This result can't be used for antenna tracking because the angle error output bandwidth is wide (10 kHz) during simulation. If using the narrow bandwidth output, the result can be used for antenna tracking. Figure 13.8d is the simulation result without the interference suppression module in receiver. It is obvious that the variance of the normalization error voltage is huge. Even if adopt a narrow bandwidth output, it can't be used for antenna tracking.

13.4 Outfield Test Results Analysis

In DS/FH communication system, angle tracking and automatic phase calibration need a special test to verify these functions in the outfield. In December 2013, Beijing Research Institute of Telemetry successfully completed the verification test in Xi'an. In this experiment the device using wireless access, to test the automatic phase calibration and angle tracking performance.

13.4.1 The Automatic Phase Calibration Testing

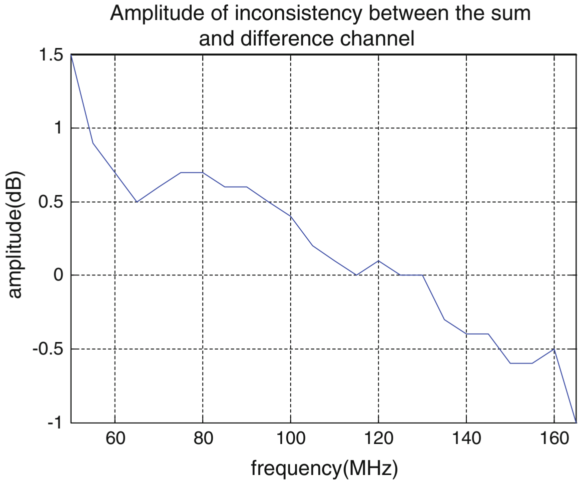
The automatic phase calibration in conjunction with the standard signal tower to automatic adjusts the phase shifter in the difference channel, in order to ensure that this system can track the target successfully. After the phase calibration, this system needs to be an antenna deviation test to verify whether the measured value of directional sensitivity is consistent with the setting value of directional sensitivity.

Firstly, the test signal is amplified by LNA, and then transmits via a long cable. After the long cable transmission, it will lead to the amplitude difference and phase difference between the sum channel and difference channel. Measuring the difference value is the precondition of normalized processing. The system connection graph as shown in (Fig. 13.9).

With 5 MHz for step by step, set the frequency output range of the signal source from 50–165 MHz. In the signal transfer board, the output signal amplitude and phase is measured respectively, and then check the amplitude and phase inconsistency at the receiving frequency range. Test results are as follows.

Fig. 13.9 The system connection graph

Fig. 13.10 Amplitude inconsistency curve



It can be seen from the Figs. 13.10 and 13.11, the amplitude inconsistency and the phase inconsistency consistent with the theoretical analysis. And then put the above data into the phase corrector to accomplish the phase calibration.

13.4.2 The Automatic Tracking Testing

When the carrier locked and the AGC voltage satisfied the given threshold, the system starts on the closed-loop tracking testing. The test time for 5 min, the results are shown in Table 13.1.

Fig. 13.11 Phase inconsistency curve

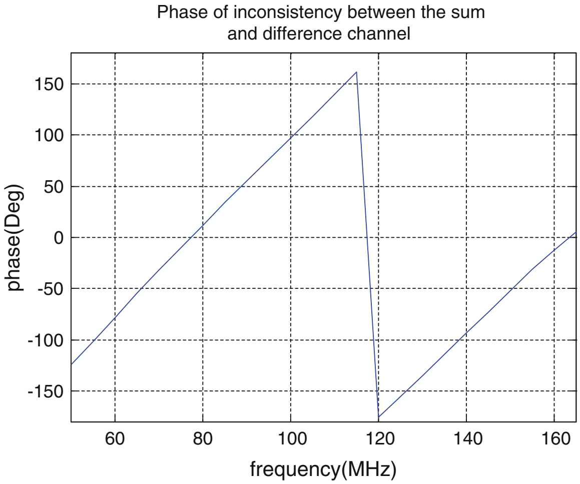


Table 13.1 Automatic tracking testing

Serial Num	Mean (Deg)		Variance (Deg)	
	Azimuth	Elevation	Azimuth	Elevation
1	272.980	6.684	0.00045	0.0012
2	272.987	6.621	0.00033	0.0011
3	272.986	6.633	0.00028	0.0007
4	272.985	6.641	0.00070	0.0007
5	272.985	6.634	0.00044	0.0011

It can be seen from above table, the angle tracking module works well. The angle measuring accuracy satisfies the system requirement.

13.5 Conclusion

Focused on the problem that the angle tracking function of the DS/FH system can't work successfully under an interference condition, this paper propose an angle tracking algorithm which adapts the dynamic interference environment. Firstly, analyze the received signal in the frequency domain, and then automatically detect the narrowband interference and suppress it. Second of all, attain the angle error signal. This paper adopts the cross-correlation method, its theory is simple and the system no longer needs dispreading. However, the inconsistency of amplitude and phase will eventually affect the angle measuring accuracy, and then the phase calibration is needed. As the carrier frequencies hop continuously in DS/FH system, it's hard for this system to calibrate automatically. Therefore, this paper proposed an automatic phase calibration algorithm based on the normalization carrier frequency. This algorithm breaks through the bottle neck of the phase calibration of the multi-frequency points and implement quickly phase calibration based on the dual-channel monopulse system. Simulation results and outfield test results show that using this algorithm can acquire the angle error signal in an interference environment, and the automatic phase calibration can be implemented.

References

1. Guo SL, Lou ZG (2008) Study on tracking technology of DSSS TT&C system. J spacecra TT&C technol 27(1):49–52
2. Cherubini G, Milstein LB (1989) Performance analysis of both hybrid and frequency-hopped phase-coherent spread spectrum systems part I: a hybrid DS/FH system. IEEE Trans Commun 37:600–611
3. Hao JH, Xu B, Liu R (2008) Study of tracking methods using spread spectrum signal. Key Mag China Technol 27(4):10–13

4. Liu B, Lu RJ (2010) Study on automatic phase calibration method of on-orbit target tracking in unified TT&C system. J Telemetry Track Command 31(1):37–40
5. Su X, Xi WJ (2012) Phase calibration technology of dual-channel tracking receiver with ground beacon. Telecommun Eng 52(3):268–272
6. Su X (2010) S/X band test report for deep space TT&C system phase calibration/distance calibration. The 10th Institute of China Electronics Technology Group Corporation, Chengdu
7. Kumpumaki TJ, Isohookana MA, Juntti JK (1997) Narrow-band interference rejection using transform domain signal processing in a hybrid DS/FH spread-spectrum system. IEEE, New Jesey, pp 89–93
8. Li Y, Zhang LX, Meng YS (2012) Effect of frequency-domain overlapping and windowing interference suppression on intermediate frequency signal measuring-range accuracy. Sp Electr Technol 1:23–29

Chapter 14
Optimizing Deep-Space Antennas Array Configuration by Multi-objective Genetic Algorithm

Xueshu Shi, Hong Ma and Yiwen Jiao

Abstract In order to increase the capacity of communication link between ground and planetary spacecraft, antenna arraying technology is adopted in deep space communication. For antenna arrays construction, array elements geometric configuration optimization is a key technology to improve array performance. Firstly, calculation method of instantaneous synthesized beams is presented, and an optimization model of array for a given tracking direction is built. Secondly, for limitation of standard genetic algorithms, an improved genetic algorithm (IGA) for array configuration optimizing is given. Finally, based on IGA optimization algorithms and Pareto multi-objective genetic algorithms, multi-objective array configurations optimization is carried out between minimum sidelobe level and cable length, and between minimum sidelobe level and main lobe width.

Keywords Deep-space communication · Array antennas · Configuration optimizing · Genetic algorithm · Multi-objective optimizing

14.1 Introduction

Spacecraft tracking and telemetry reception using large arrays composed of numerous small inexpensive reflector antennas is becoming a new choice in future deep-space network development [1]. The problem of deciding the relative positions of antennas in array has many degrees of freedom. Finding methods to solve this problem and optimizing array performance relevance to spacecraft tracking applications is a key technology. In array design process, the array performance parameters include minimum sidelobe level, shortest possible cable length and minimum main lobe width, all should be optimized. Multi-objective genetic algorithm provides solution to optimization problems, with its intelligence, universality,

---

X. Shi (✉) · H. Ma · Y. Jiao
Equipment Academy, Beijing 101416, China
e-mail: altel@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_14

robustness and global search ability [2]. In this paper based on GA operators and Pareto strategies [3], a multi-objective optimizing algorithm is developed to optimize array parameters.

In Sect. 14.2 of this paper, a calculation model of array instantaneous synthesized beams for a given tracking direction is built. In Sect. 14.3, based on calculation model, an improved genetic algorithm (IGA) for arrays geometric configurations optimization is proposed to improve array performance. For practical optimizing needs, designer only care about available solution, so in Sect. 14.4 based on IGA, some multi-objective array configurations optimization are carried out between minimum sidelobe level and cable length, and between minimum sidelobe level and main lobe width. Then MOPGA (Multi-Objective Pareto Genetic Algorithm) is proposed to solve the multi-objective optimization of deep space array configuration, finally satisfied Pareto front and Pareto solutions database are obtained for configuration result selection.

14.2 Calculation Methods of Array Instantaneous Synthesized Beams

According to array signal process theory, the beam pattern of the array can be verified by the following equation [4]:

$$\underset{u\in[-1,1],v\in[-1,1]}{E(u,v)} = \sum_{m=1}^{M}\sum_{n=1}^{N} a_k e^{-j[\pi(ux_m+vy_n)]} \tag{14.1}$$

where a_k denotes antenna amplitude weights, suppose antenna elements have same size and have similar receiving capacity, uniform amplitude weights is adopted in following analysis, that is $a_k = 1$. In above equation, u and v stand for orientation parameters of synthesized beam, and the vector $\vec{r}_k = (x_m, y_n)$ determines the position of the kth array element at the aperture of the array, measured at wave lengths. The instantaneous synthesized power radiated by the aperture can be written as $P(\vec{e})$, which denotes projection aggregate of baseline decided by any two array antenna elements onto beam vector direction, presented as:

$$\begin{aligned} P(\vec{e}) &= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{n=1}^{N} e^{\{-j2\pi[(\vec{r}_k-\vec{r}_n)\cdot\vec{e}]\}} \\ &= \frac{1}{N}\sum_{k=1}^{N} e^{(-j2\pi\vec{r}_k\cdot\vec{e})}\cdot\frac{1}{N}\sum_{n=1}^{N} e^{(j2\pi\vec{r}_n\cdot\vec{e})} \\ &= \frac{1}{N^2}\left|E(\vec{e})\right|^2 \end{aligned} \tag{14.2}$$

Above mathematical derivation are on the assumption that array located on the equator and antenna beam pointed on zenith position, in practical application, wherever latitude array located on and which direction beam pointed on, instantaneous synthesized power of array can equivalent to Eq. (14.1).

14.3 Improved Genetic Algorithms (IGA) for Array Configuration Optimizing

Genetic algorithm (GA) is an optimization technique with parallel, robust, and probabilistic search characteristic [5], and is suitable for deep space array optimization problems that with many design variables and nonlinear processing. But standard genetic algorithm (SGA)cannot be adopted directly in optimization, so in this paper we proposed an improved genetic algorithm (IGA) for array configuration optimizing. Figure 14.1 shows flowchart of IGA for arrays optimizing.

Compared with SGA, improved optimization approach of IGA can be summarized by follows.

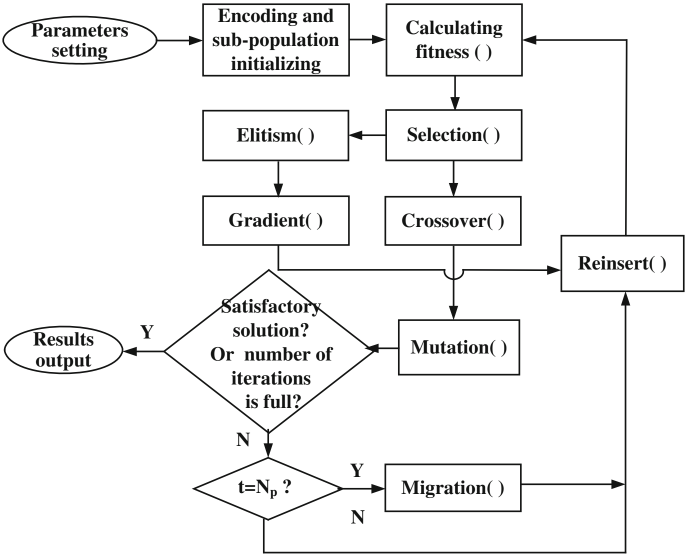


Fig. 14.1 Flowchart of IGA for deep space arrays optimizing

14.3.1 Chromosome with Hybrid Coding

If using standard real-coded GA, position vector of antenna elements is directly appointed as gene on chromosome, which is denoted by $gene_i = \{x_i, y_i\}$, and chromosome of array geometric configuration can be given:

$$Chrom(k) = gene_1(k), gene_2(k), \ldots gene_N(k) \tag{14.3}$$

whereas in array optimization process, using real-coding can't deal with constraints of geometric distribution such as those antennas must be fitted within the available land, avoid antenna one another shadowing and so on. In order to reduce geometric constraints operating and make gene to denote array topology identity, an antenna position coding parameter s_i is added in chromosome, which denotes the ith antenna's relative coded position in planar array. Position coding can use Cartesian coordinates or Polar coordinates, if choose Cartesian, and then the ith antenna's coded position vector is expressed as follows:

$$\{x_i, y_i\} = \{[f(s_i) + \Delta x_i],\ [g(s_i) + \Delta y_i]\} \tag{14.4}$$

where $(\Delta x_i,\ \Delta y_i)$ is relative coordinate position of the ith antenna, and uses real-coding. In formula (4) $f(s_i)$ and $g(s_i)$ is coding coordinate translation function. Finally gene string of the ith antenna position is deduced as $gene(i) = \{s(i),\ \Delta x(i),\ \Delta y(i)\}$. Through the improvement of coding, gene parameters presents certain dominance characteristic, at the same time maintained the advantage of real-coding that can approach the optimal solution by arbitrarily high precision.

14.3.2 Improving the Convergence Speed of Algorithm

With the increase of antennas numbers, and due to ultra-sparse distribution characteristic of antenna array, the solution space will get exponential growth, even if the initial population is composed of thousands of chromosomes, the population size is too small relative to the solution space. Therefore, when generating the initial population, must let individuals different from each other as far as possible, let algorithm has relative abundance of genetic material from beginning of iteration. Meanwhile, in order to further improve the convergence rate of small populations of evolution, the use of parallel genetic population is divided into several sub-populations, and inter-generational migration operation is added between sub-populations in the evolutionary process. Also in the early iterations, using niche technology in the offspring group, and making gene of antenna location s_i as the parameter of the fitness sharing function, so that GA optimization maintain a certain distance between the individual gene, and enhance the ability of the optimal solution search.

In practice due to limits of calculation accuracy, terrain, antenna position coordinates error, antenna calibration accuracy, therefore geometric optimization only requires engineering satisfactory solution [6]. Based on the above analysis, the gradient search operator is added in Elitism operation to accelerate the rate of convergence of the algorithm, expressed as follows, if individual's fitness in a generation groups is particularly high, then put the individual store for gradient search, and after several generations of gradient calculation, insert these individuals into the population and continue genetic optimization.

14.4 Pareto Multi-objective Genetic Algorithms Based on IGA

When the number of antennas is fixed, the minimum element spacing and maximum configuration area constraints directly determine the border of the instantaneous synthesized beam sidelobe level and main lobe width. In order to reduce the mutual shadow between the antenna element, spacing usually set at a fixed minimum distance, while configuration of the array area has a lot of flexibility. In engineering design, determination of the maximum configuration area needs to consider the number of antennas, the angular resolution requirements, the phase shift constraints, construction cost, terrain and other conditions. Although there are no optimal solution of the antenna array configuration for each design goal, but near-optimal solution between several conflict design goals need to be selected. This near-optimal trade-off solution is called Pareto optimal solution [7].

In contrast to a traditional GA, a Pareto GA spreads its population over the objective space in order to simultaneously explore different performance trade-offs. Through iteration of the optimizer, the population converges toward a Pareto front, which describes a near-optimal trade-off between different objectives over a wide range of objective values. This essentially provides a design curve by which one can select the desired trade-off.

When solving Pareto GA optimal solution, the real coding and binary encoding all shown an endpoint effect, that cause the obtained solution clustering unexpectedly. While paper adopts hybrid encoding and niche technology of IGA in Pareto multi-objective optimizing (MOPGA) [8], not only can overcome endpoint effect to some extent, but also can improve the diversity of genetic material. In addition, coding and cross, mutation operator of MOPGA all copy from the corresponding operation of IGA, that will not be discussed in detail. Below several multi-objective optimizations instance is given.

14.4.1 Sidelobe Level and Cable Length Optimizing Based on MOPGA

During large-scale antenna array configuration design, we always want to save costs while maintaining the track detection sensitivity performance of the antenna array.

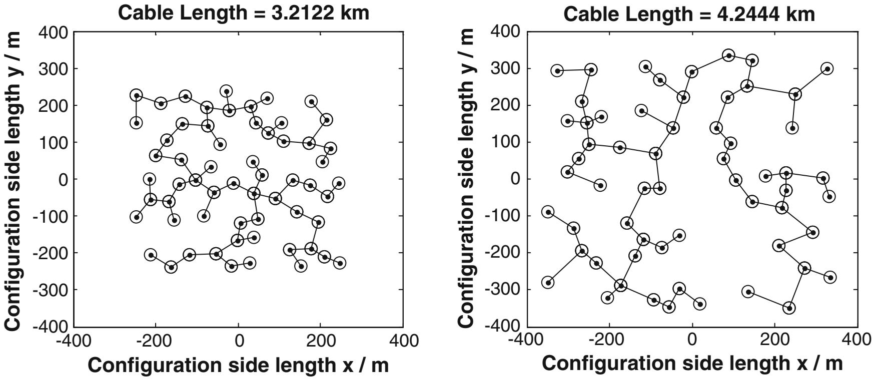


Fig. 14.2 Cable length of optimized configuration in different upper limit of optimization area based on IGA

Configuration area is smaller, the lower the potential laying cables costs. Figure 14.2 shows when the parameters of configuration area are 500 m × 500 m and 700 m × 700 m constraints, the corresponding optimal configuration of the cable length (cable length based on the minimum spanning tree algorithm). The simulation set the initial population size of 8 × 50, and the minimum spacing between antenna elements is 40 m, the antenna array element number 60, RF frequency 32 GHz, antenna diameter 12 m with 3 dB main beam width of 0.058°.

According to Fig. 14.2, configuration optimizations of the two constraints have the cable length difference of about 1 km, but with the increase in the number of antenna elements, the cable length difference will be greater. In addition, Fig. 14.3 shows two configuration optimization results have similar side-lobe level (<0.09), so the 500 m × 500 m configuration result better than 700 m × 700 m result based on cost considerations.

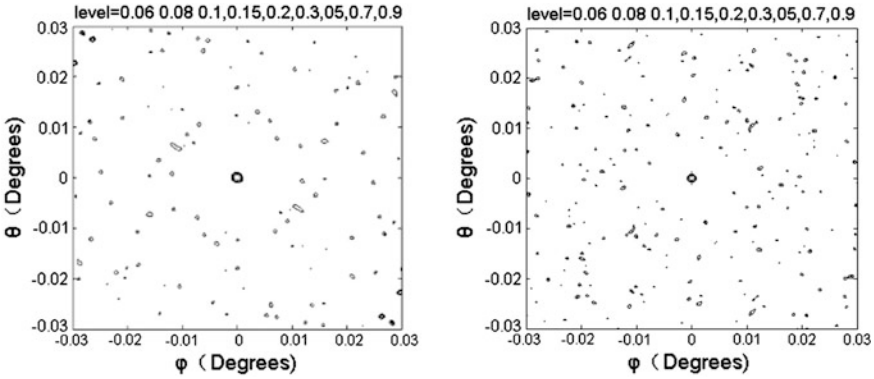


Fig. 14.3 Beam pattern of optimal configuration of two area constraints

Below, we analyze the antenna array configuration optimization between the sidelobe level and the cable length. Configuration optimization mathematical model as follows:

$$\begin{cases} Min \quad f_1(x_i, y_i) = \max\left|\dfrac{P(\alpha,\delta)}{P_{\max}}\right| & s.t. \quad \begin{array}{l} -L \le x_i \le L, -L \le y_i \le L \\ \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \ge L_{\min} \\ i,j \in [1, \ldots N], i \ne j \end{array} \\ Min \quad f_2(x_i, y_i) = MST(x_i, y_i) & \end{cases} \quad (14.5)$$

Where $f_1(x_i,y_i)$ is sidelobe level optimization objective function, and $f_2(x_i,y_i)$ is the cable length optimization objective function, *MST()* is the minimum spanning tree algorithm used to calculate the minimum cable laying length. By analyzing Eq. (14.5), potential conflict can be found between these two objectives: the smaller configuration area, the shorter cable laying length. But if the configuration area is too small, the solution space of optimization will get relatively small, sidelobe optimization results will be worse.

Based on MOPGA operations, optimization of 60 antenna elements corresponding to the minimum sidelobe level and cable length was simulated. In Fig. 14.4, the configuration area constraint is 500 m × 500 m, and the minimum antenna spacing constraints 40 m. Figure 14.4 shows Pareto front of 200 generations evolution, configuration topology results on the Pareto front. The MOPGA result shows the wave front is ideal, Pareto optimal solution with high credibility, which provides an optional solution database for engineering applications.

In Fig. 14.4, the Pareto front is not continuous, but can be improved by increasing the number of population individuals and the evolution iteration

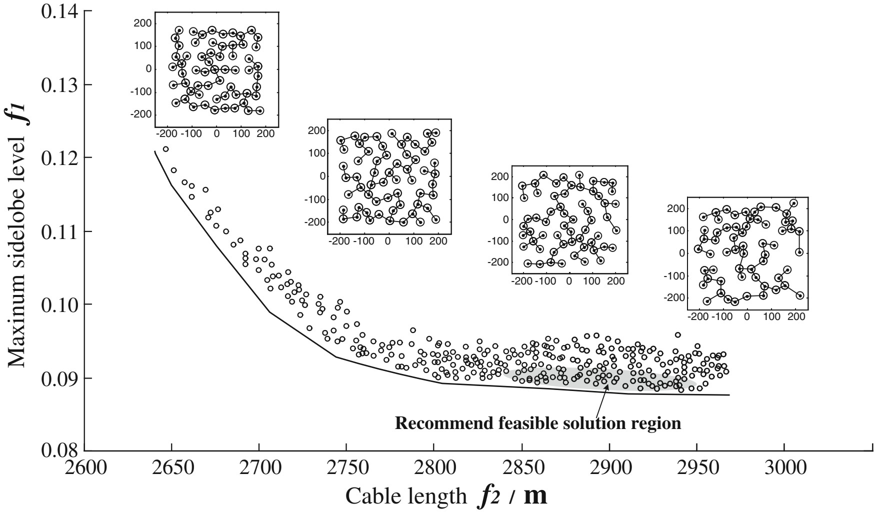


Fig. 14.4 Pareto wave front of the bi-objective optimization between sidelobe level and cable length

numbers. In addition, according to engineering application, additional 200 meters of cable laying length will not significantly increase the cost of the system. So if reasonable constraints of the configuration area are adopted, only minimize sidelobe level problem is needed to consider and optimize.

14.4.2 Sidelobe Level and Main Lobe Width Optimizing Based on MOPGA

During configuration optimization design of the antenna array, cable laying length and configuration area are conflicting, while increasing the configuration area and reduce the width of the synthesized beam, increase the beam angular resolution is consistent, so shorten cable length and reduce the synthesized beam width is conflicting in Multi-objective optimizing. But in the actual deep space communication situation, we are more concerned about whether the side lobe level is the minimum, so the simulation of array sidelobe level and main lobe width is given below.

Figure 14.5 shows optimization results (the sidelobe amplitude and the corresponding width of the main lobe) using single-objective IGA algorithms, under the conditions of the different configuration area. The simulation parameters set same as the above. With the configuration of the area increases, the solution space will increase rapidly. Figure 14.5 shows that configuration area is larger, converges to the low side lobe is more difficult, and configuration area is smaller, the more likely convergence at a lower sidelobe level. But the width of the main lobe is significantly reduced with the increase of configuration area. So if the requirement of the main lobe beam width is met, optimization should be within a smaller configuration to find the smallest sidelobe.

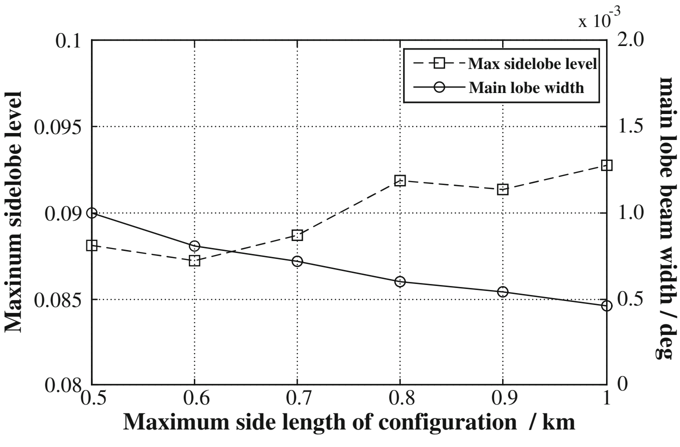


Fig. 14.5 Maximum sidelobe level and 3-dB beam width of main lobe of 60 antennas with different upper limit of optimization area based on IGA

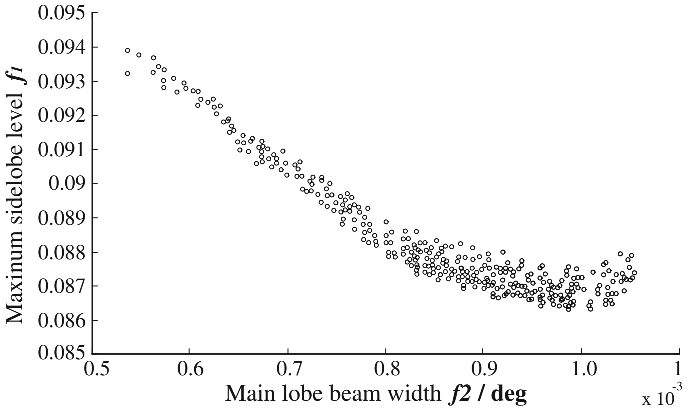


Fig. 14.6 Pareto wave front of the bi-objective optimization between sidelobe level and main lobe beam width

Based on proposed IGA genetic algorithm, with MOPGA operations, optimization of 60 array elements corresponding to the minimum sidelobe level and main lobe width was simulated. Optimization mathematical model as follows:

$$\begin{cases} \min \quad f_1(x_i, y_i) = \max\left|\dfrac{P(\alpha,\delta)}{P_{\max}}\right| \\ \min \quad f_2(x_i, y_i) = W_{\text{beam}}(P(\alpha,\delta)) \end{cases} \quad s.t. \quad \begin{array}{l} -L \le x_i \le L, -L \le y_i \le L \\ \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \ge L_{\min} \\ i, j \in [1, \ldots N], i \ne j \end{array} \tag{14.6}$$

Where $f_1(x_i, y_i)$ is sidelobe level optimization objective function, $f_2(x_i, y_i)$ is beam width optimization objective function, $W_{\text{beam}}()$ is used to calculate the synthesized beam width. Figure 14.6 shows the evolved Pareto front after 200 generations; Obtained Pareto optimal solutions provide an optional solution database for engineering applications.

14.5 Conclusion

With the increase of the number of deep space array antennas, as well as by the limitations of the terrain and antenna configuration area, in practice the configuration optimization will encounter many difficulties. Multi-objective optimization of antenna array configuration can provide designers with an alternative solution database, and according to certain criteria such as area constraints etc., based on solution database designers can select a better solution from the Pareto optimal solution.

Acknowledgment This work is supported by the National Natural Science Foundation of China (Grant No. 61271265).

References

1. Jamnejad V, Huang J, Cesarone RJ (2001) Array antennas for JPL/NASA deep space network. IEEE Aerospace Conf Proc 10:219–230
2. Eckart Z, Kalyanmoy Deb (2000) Comparison of multi-objective evolutionary algorithms empirical results. Evol Comput 8:173–195
3. Kay CT, Tong HL (2002) Evolutionary algorithms for multi-objective optimization performance assessments and comparisons. Artif Intell 17:251–290
4. Kumar BP, Branner GR (1999) Design of unequally spaced arrays for performance improvement. IEEE Trans. Antennas Propag 47:511–524
5. Goldberg DE (1989) Genetic algorithm. In: Science Amen. Addison-Wesley, New York
6. Kogan L (2000) Optimizing an large array configuration to minimize the sidelobes. IEEE Trans Antennas Propag 48:1075–1078
7. Rahmat-Samii Y, Michielssen E (1999) Electromagnetic optimization by genetic algorithms. Wiley, New York
8. Shi XS, Wang YQ (2010) Research of optimizing algorithm for deep space large arrays geometric configuration. J Astronaut 31:478–484

Chapter 15
Research on Inter-Satellite Capturing and Tracking Simulation Model

Lanjuan Tong, Hui Guan, Zhilin Wang and Shaolin Yu

Abstract The running rule of inter-satellite capturing and tracking is studied and the simulation model is established, which consists of orbit prediction, attitude setting and antenna pointing simulation. Based on the model, both algorithms and software were designed, realized, and verified by carrying out a certain data-relay task drilling combined with actual data. The result shows that the model is capable of simulating user's relay terminal tracking TDRS continuously at the precision of 0.400684 degree for 8 h and supporting 8-circle task planning with the efficiency improved by 7 times, compared with 1-circle data preparation at present. High precise and efficient simulation supports vigorously for validation of capturing and tracking schemes.

Keywords Relay terminal · Capturing and tracking · Orbit prediction · Antenna pointing · Simulation modeling

15.1 Introduction

With the construction of TDRSS improving, the number of the relay terminal used in LEO spacecrafts is increasing sharply. In order to fulfill data-relay task, it is necessary to carry out TDRSS task planning and to analyze the process of spacecraft capturing and tracking TDRS. During the prophase, lots of task drills are carried out, supported by relay terminal simulation to verify capturing and tracking schemes, especially for vital tasks. Now more user relay terminal access verification needs more simulation support, shorter planning time and higher efficiency and accuracy, so the relay terminal simulation model must be studied to meet TDRSS development requirements.

L. Tong (✉) · H. Guan · Z. Wang · S. Yu
Beijing Space Information Relay and Transmission Technology Research Center,
Beijing 100094, China
e-mail: tonglanjuan@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_15

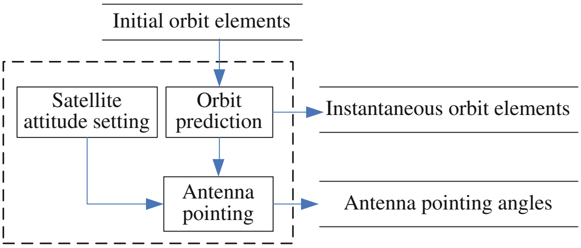


Fig. 15.1 Relay-terminal capturing and tracking process

15.2 Principle Analysis of Relay Terminal Capturing and Tracking

When the user spacecraft captures and tracks TDRS, its relay terminal servo control unit calculates antenna pointing angles and then drives the antenna pointing to TDRS, based on the two objects' orbital elements and satellite attitude, as shown in Fig. 15.1.

In Fig. 15.1, the core is shown in the dashed box, which consists of orbit prediction part, satellite attitude setting part and antenna pointing part. Inputs are satellite initial orbit elements and outputs are instantaneous orbit elements and antenna pointing angles.

15.3 Simulation Model Design of Relay Terminal Capturing and Tracking

Derived from the principle of the relay terminal, the simulation model consists of orbit prediction module, satellite attitude setting module and antenna pointing module, as shown in Fig. 15.2.

Among three modules, satellite attitude setting module can be simplified. For TDRS is a geostationary satellite, whose east-west and north-south station-keeping range can be negligible, TDRS is assumed to be stationary. While user satellites are

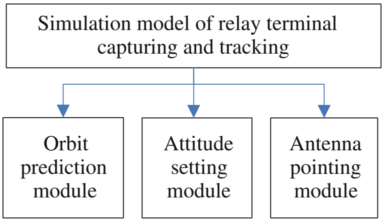


Fig. 15.2 Simulation model architecture

basically ground orientation during data-relay tasks, and its attitude control deviation can be ignored under simulation conditions, therefore attitude deviation can be set as constants in simulation conditions (for example, set flying yaw angle as zero, inverted yaw angle as 180°).

Orbit prediction module and antenna pointing module are mainly studied, which involve J2000 orbital coordinate system, J2000 inertial coordinate system, user satellite orbit coordinate system, user satellite body coordinate system, antenna body coordinate system [1].

15.3.1 Module Design of Orbit Prediction

15.3.1.1 Establishment of Orbital Perturbation Equation

LEO satellites in space are mostly affected by atmospheric drag and additional force generated by non-spherical Earth and uneven quality, and the solar radiation and the sun's and moon's gravitation can be ignored, no influencing simulation accuracy [2].

Orbit prediction method mainly contains numerical method and analytical method [3], of which the former is more simple and accurate, so is it adopted. In order to simplify the right function and trigonometric functions of the orbit perturbation equation, J2000 coordinate system is adopted. For non-spherical gravitation perturbation, J_2 is mainly considered with the magnitude of 10^{-3}, while J_3 and J_4 is neglected with the magnitude of 10^{-6}. In the perturbation equation, the position vector is set as $r = (x, y, z)^T$, the speed vector as $v_r = (v_x, v_y, v_z)^T$, and Eqs. (15.1)–(15.6) are shown as follows [4].

$$\frac{dx}{dt} = v_x \tag{15.1}$$

$$\frac{dy}{dt} = v_y \tag{15.2}$$

$$\frac{dz}{dt} = v_z \tag{15.3}$$

$$\begin{aligned} \frac{dv_x}{dt} = & - U_e x/c_r^3 + u_a(J_2(7.5z\_r^2 - 1.5) \\ & - c\_CdSpV(v_x + W_e y) \end{aligned} \tag{15.4}$$

$$\begin{aligned} \frac{dv_y}{dt} = & - U_e y/c_r^3 + u_a(J_2(7.5z\_r^2 - 1.5) \\ & - c\_CdSpV(v_y - W_e x) \end{aligned} \tag{15.5}$$

$$\begin{aligned}\frac{dv_z}{dt} = & - U_e z/cr^3 + u_a(J_2(7.5z\_r^2 - 4.5) \\ & - c\_CdSpV \cdot v_z\end{aligned} \tag{15.6}$$

In the equations, U_e is gravity constant, J_2 is perturbation term, c_CdSpV is atmospheric drag coefficient, Re is the earth equator radius, W_e is the earth angular velocity, and other parameters are defined in Fomulas (15.7)–(15.9).

$$c_r = \sqrt{x^2 + y^2 + z^2} \tag{15.7}$$

$$u_a = \frac{U_e^3}{c_r^5} \tag{15.8}$$

$$z\_r = \frac{z}{c_r} \tag{15.9}$$

15.3.1.2 Runge-Kutta Integration Method

Common numerical methods contain Runge-Kutta method and Adams-Cowell method [5], of which the former is easier, more simple and stabile, so is it chosen. It is a single-step integral method, and formulas are simple in form, shown as Fomulas (15.10)–(15.13).

Suppose the problem is described as follows.

$$\begin{cases} \dot{y}(t) = f(t, y) \\ y(t_n) = y_n \end{cases} \tag{15.10}$$

Then Runge-Kutta expressions are deduced as Formulas (15.11) and (15.12).

$$y_{n+1} = y_n + h\sum_{i=0}^{k} C_i f_i \tag{15.11}$$

$$\begin{cases} f_0 = f(t_n, y_n) \\ f_1 = f(t_n + a_1 h, y_n + a_1 b_{10} h f_0) \\ \quad \vdots \\ f_k = f(t_n + a_k h, y_n + a_k h \sum_{i=0}^{k} b_{ki} f_i) \end{cases} \tag{15.12}$$

In the Formulas, C_i, a_i and b_{ij} are known, k is the rank, and h is integral step, shown as Formula (15.13).

$$h = t_{n+1} - t_n \tag{15.13}$$

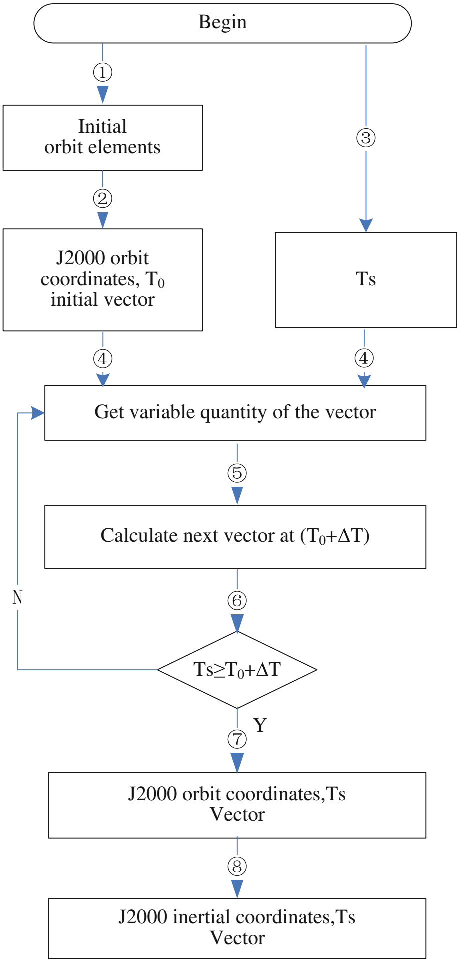
15.3.1.3 Algorithm Design of Orbit Prediction

Based on orbit perturbation equations and solution algorithms, orbit prediction algorithm is designed [6] as shown in Fig. 15.3.

It is described step by step.

Step 1. Input initial orbit elements of the user satellite.
Step 2. Convert orbit elements into the vector of position and velocity in J2000 orbit coordinates at T_0 moment.
Step 3. Input prediction time T_S ($\geq T_0$).
Step 4. Solve equations and got variable quantity of the position and speed.

Fig. 15.3 Algorithm flow of orbit prediction



Step 5. Calculate the vector of position and speed at $(T_0 + \Delta T)$ moment by four-rank Runge-Kutta and ΔT is the integral step.

Step 6. If time condition is met, go to the next step, or else to the fourth step.

Step 7. Output the vector of position and velocity at T_S moment.

Step 8. Convert the vector of position and velocity at T_S moment from J2000 orbital coordinate system to J2000 inertial coordinate system.

So is the orbit prediction algorithm of the user satellite, while TDRS orbit prediction is realized based on the present platform in existence, with no details.

15.3.2 Module Design of Orbit Antenna Pointing

15.3.2.1 Principle of Antenna Pointing

When the user satellite captures and tracks TDRS, their relay terminal antennas are pointing at each other, as shown in Fig. 15.4.

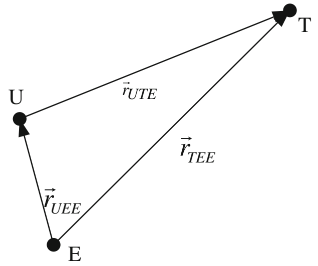
Let U denote user satellite, E denote the Earth and T denote TDRS. All of them are seen as particles. Also set the position vectors of the user satellite and TDRS as $\vec{r}_{UTE}$ and $\vec{r}_{TEE}$ in J2000 inertial coordinate system, and then the pointing vector from the user satellite to TDRS is shown as Formula (15.14).

$$\vec{r}_{UTE} = \vec{r}_{TEE} - \vec{r}_{UEE} \tag{15.14}$$

The pointing vector is converted from J2000 inertial coordinate system to antenna coordinate system and antenna orientation angles are defined as A and B, which describe the direction of the center line of the antenna beam. A is called azimuth angle between x-axis and the projection of the center line of the antenna beam in the xoy plane, and B is called pitch angle between the z-axis and the center line of the antenna beam, shown in Fig. 15.5. T denotes pointing point, and $\overrightarrow{OT}$ denote antenna pointing vector.

Antenna pointing expressions are shown as Formulas (15.15)–(15.16).

Fig. 15.4 Antenna pointing principle



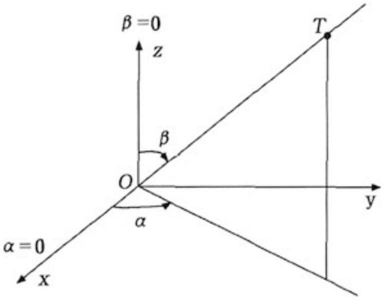


Fig. 15.5 Antenna orientation

$$\alpha = -arctan(\frac{y}{z}) \tag{15.15}$$

$$\beta = a\sin(\frac{x}{r}) \tag{15.16}$$

15.3.2.2 Algorithm Design of Antenna Pointing

According to the principle of antenna pointing, the algorithm is designed [7] as shown in Fig. 15.6.

It is described step by step as follows.

Step 1. Using orbit elements resulted from orbit prediction simulation, calculate position vector of the user satellite and TDRS in J2000 inertial coordinate system.

Step 2. Calculate antenna pointing vector $\vec{r}_{UTE}$ in J2000 inertial coordinate system by using formulas in Sect. 15.3.2.1.

Step 3. Convert $\vec{r}_{UTE}$ from J2000 inertial coordinate system to the user satellite orbit coordinate system, expressed as Formula (15.17).

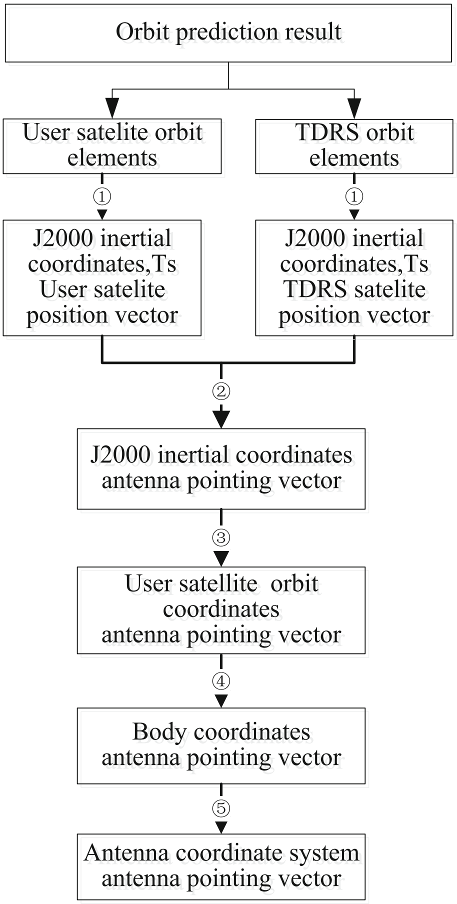
$$\vec{r}_{yh} = R_{YH} \cdot \vec{r}_{UTE} \tag{15.17}$$

R_{YH} is transition matrix.

$$R_{YH} = \begin{bmatrix} \frac{y^2v_x + xzv_2 - z^2v_x + xyv_y}{r^2} & \frac{x^2v_y + xyv_z - yzv_z + z^2v_y}{r^2} & \frac{y^2v_z + yzv_y - x^2v_z + xzv_x}{r^2} \\ \frac{yv_z - zv_y}{r} & \frac{xv_z - zv_x}{r} & \frac{xv_y - yv_x}{r} \\ \frac{x}{r} & \frac{y}{r} & \frac{z}{r} \end{bmatrix} \tag{15.18}$$

Step 4. Convert $\vec{r}_{yh}$ from the user satellite orbit coordinate system to the body coordinate system, expressed as Formula (15.19).

Fig. 15.6 Algorithm flow of antenna pointing



$$\vec{r}_b = R_Y(\theta) \cdot R_X(\varphi) \cdot R_Z(\psi) \cdot \vec{r}_{UTE} \tag{15.19}$$

$R_Z(\psi)$ denotes the rotation ψ around Z axis, $R_X(\phi)$ denotes the rotation ϕ around X axis, and $R_Y(\theta)$ denotes the rotation θ around Z axis. ψ, ϕ and θ represent respectively yaw angle, roll angle and pitch angle.

Step 5. Convert $\vec{r}_b$ from the body coordinate system to antenna coordinate system, expressed as Formula (15.20).

$$\vec{r}_a = R_Y(\alpha) \cdot R_X(\beta) \cdot R_Z(\gamma) \cdot \vec{r}_b \tag{15.20}$$

$R_Z(\gamma)$ denotes the rotation γ around Z axis, $R_X(\beta)$ denotes the rotation β around X axis, and $R_Y(\alpha)$ denotes the rotation α around Y axis. α, β and γ represents respectively the angle between *Y* axis, *X*-axis, *Z*-axis in antenna coordinates and corresponding axis in body coordinates.

Step 6. According to Formulas (15.15) and (15.16), calculate antenna elevation angle and azimuth angle.

15.4 Simulation Results

Based on the algorithms above, software is designed and realized on VC2008 platform. In order to validate both algorithms and software, a typical task scene is designed and performed, with simulation results obtained as follows.

15.4.1 Simulation Results of Orbit Prediction

The key of orbit prediction algorithm is to confirm four-rank Runge-Kutta step. In experiments, four groups of steps are choosen as 1, 10, 20 and 30 s to conduct 12-h orbit prediction, and furthermore position error is calculated based on Formula (15.21), with results as shown in Table 15.1.

$$\sigma_i = \sqrt{(x_i - x_i)^2 + (y_i - y_i)^2 + (z_i - z_i)^2} \tag{15.21}$$

σ_i represents the position error, (x_i, y_i, z_i) and (x_t, y_t, z_t) represent simulation data and actual data respectively.

As seen in Table 15.1, under 1-s step position error is minimum and time consuming is the longest, while error reaches the maximum under 30-s step, and time consuming shortest. As the model is mainly used for task planning, accuracy is prior to real-time requirements, so 1-s step is adopted. Orbit prediction error curve is shown in Fig. 15.7.

In Fig. 15.7, as time goes up, the curve is oscillating and rising. When time is 41,290 s (11:28:10), the error reaches the maximum of 3,732.3 m within the simulation range. There are two reasons for the error: one is accumulative error of Runge-Kutta integral step, and the other is ignorance of solar radiation, compared with actual orbit prediction.

Table 15.1 Prediction result comparation of different steps

Step (s)	Position error (m)	Time (s)
1	3,732.312	4.125
10	3,732.506	0.406
20	3,736.360	0.203
30	3,757.836	0.141

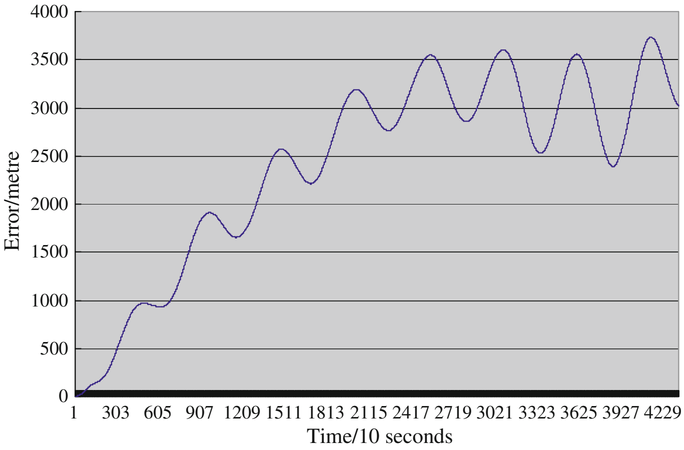


Fig. 15.7 Orbit prediction error for 12 h 7

15.4.2 Simulation Results of Antenna Pointing

Based on orbit prediction results and satellite attitude, antenna pointing module covers all visible arcs during the prediction period (for 12 h) and outputs pointing angles. Compared with theoretical data (no considering real-time attitude) and actual data (considering real-time attitude), antenna pointing error is calculated, which is reflected by three values, *X*-axis pointing error, *Y*-axis pointing error and antenna pointing deviation [8], shown as Formulas (15.22)–(15.24).

$$\sigma_{xi} = x_i - x_t \tag{15.22}$$

$$\sigma_{yi} = y_i - y_t \tag{15.23}$$

$$\sigma_i = \cos^{-1}(\sin y_i \sin y_t + \cos y_i \cos y_t \cos \sigma_{xi}) \tag{15.24}$$

x_i, y_i represent simulation data, and x_t, y_t represent theoretical data or actual data, and calculation results are shown in Table 15.2.

In Table 15.2, as time goes up, antenna pointing error rises. As known the pointing error index of 0.41°, when visible arc is from 7 to 8 h, pointing error reaches the maximum of 0.400684°, less than 0.41°, which meets the accuracy requirement.

Table 15.2 Antenna pointing error during the prediction period

Prediction period (h)	Theoretical antenna pointing deviation (Deg)	Actual antenna pointing deviation (Deg)
1–2	0.182694	0.358670
2–3	0.183235	0.374748
3–4	0.187736	0.383101
4–5	0.194206	0.384500
5–6	0.199267	0.393080
6–7	0.199835	0.389977
7–8	0.204205	0.400684
8–9	0.247405	0.418282

15.5 Conclusion

The principle of spacecrafts capturing and tracking TDRS is analyzed, and relay terminal simulation is modeled, which consists of orbit prediction module, satellite attitude module and antenna pointing module. Based on the model, a set of software algorithms are designed, realized, and validated with antenna pointing accuracy of 0.400684°. The result reveals that if the relay terminal simulation model performs once, it will support 8-h capturing and tracking simulation. In actual, 1-circle task costs less than 1 h, so it is estimated that the model can sustain 8-circle data-relay task at least. Compared with the current mode of 1-circle task regulation, the accuracy and efficiency is greatly improved.

References

1. Guo WG, Li YJ (2009) User inter-satellite antenna pointing algorithm. J Spacecraft TT&C Technol 28(1):74–76
2. Liu L (2006) Orbit theory of spacecraft. National Defense Industry Press, Beijing
3. Liu L, Wang YR (2005) An analytical method for satellite orbit forecasting. Acta Astraonmica Sinica 46(3): 307
4. Zhang RW (1992) Satellite orbit and attitude control. Beihang University Press, Beijing
5. Tong KW, Zhou JP, He LS (2009) A reduced orbit propagation model of low-earth orbit satellite. J Astronaut 30(4):1327–1333
6. Meng B, Han C (2008) Development of simulation software for high precision spacecraft orbit prediction. Comput Simul 25(1):62–73
7. Li XC, Wang SS, Yu XC (2007) Azimuth angle and pitch angle calculation of TDRS antenna pointing at user's satellite under procedure control. Space Electronic Technol 4:17–22
8. Chen H, Li BJ, Jia YP (2011) Study on the inter-satellite antenna pointing error spread quotient of tracking and data relay satellites. J Spacecraft TT&C Technol 30(3):26–31

Chapter 16
Dynamic Matrix Control for ACU Position Loop

Hainan Gao, Hongguang Pan, Quanjie Zang and Gangfeng Liu

Abstract For a closed-loop control problem of position loop in servo system with an integral process in TT&C (Telemetry, Track and Command) ground station, this paper studies the optimal control strategy, which combines the improved dynamic matrix control (DMC) algorithm with feed-forward compensation. The predictive truncation error of the above system can be avoided by modifying the shift matrix of original DMC; the lag of system response can be effectively overcome via adding feed-forward compensation. The validity of the algorithm is verified by a simulation example, which shows that the set point and the target curve can be well tracked using this control strategy and the strategy can be applied to the position loop control of antenna control unit (ACU) in servo system.

Keywords Dynamic matrix control · Position loop · ACU · Feed-forward compensation

16.1 Introduction

A servo system of TT&C ground station antenna can be used to measure the angle of tracking object continuously by controlling the antenna on target aircraft, meanwhile, if the speed and distance data of aircraft are supplemented, the target orbit can be determined. The antenna servo system contains three nested loops, which are current loop, speed loop and position loop, separately. The current loop and the speed loop are included in an antenna drive unit (ADU), and the position loop is realized in antenna control unit. The current loop guarantees the start and brake of constant acceleration; The speed loop satisfies the requirement of the

---

H. Gao · Q. Zang · G. Liu
Movable TT&C Station, Xi'an Saterlite Control Center, Weinan 714000, China

H. Pan (✉)
Department of Automation, Xi'an Jiaotong University, Xi'an 710049, China
e-mail: hongguangpan@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_16

antenna stable speed; The position loop is used to track target accurately and stably based on the angular deviation information [1]. Therefore, the performance of a ACU position loop controller affects the system tracking accuracy directly.

At present, the ACU position loop controller of servo system adopts the digital PID control algorithm, and there are many literatures on the PID algorithm and the improved PID algorithm, such as PID control algorithm with feed-forward compensation [2, 3]. Some typically intelligent control algorithms are also used in position loop control, for instance, fuzzy control, adaptive control, neural network control, etc. [4–6].

Up to now, dynamic matrix control (DMC) is one of the most influentially advanced control algorithm accepted in process industry. In DMC, system step response information is adopted to constitute prediction model, traditional auto-correction of single step prediction is expanded to multiple steps prediction, and the traditionally optimal control is displaced by online rolling optimization on the basis of feedback correction. Because of the characteristic of the step model, the basic DMC algorithm is only used in gradual stability system, however, for the unstable object, the PID controller is usually used to stabilize system. Then, DMC can be applied to optimization control.

Based on mathematical model of ACU position loop control object and the utilizing of the integral property, the system predictive truncation error can be avoided by modifying the shift matrix, and then the improved DMC algorithm can be directly applied to control unstable systems. In order to overcome the track lag, feed-forward compensation control is introduced in the improved DMC algorithm. The effectiveness of the control strategies is illustrated by comparing the control performance between the proposed algorithm and the traditional PID algorithm in the simulation example.

16.2 Servo System Modeling

In the azimuth and pitching branches of the servo system, dual-motor driving solution is adopted to enhance the system reliability and eliminate the backlash, hence, the system can run more smoothly and get a favorable dynamic stiffness characteristics and ideally favorable dynamic property. Regardless of gearbox backlash and the friction between motor and gearbox, a dual-motor driving antenna can be expressed as physical model shown in Fig. 16.1 [1].

The symbols shown in Fig. 16.1 are defined as follows:

U_a: motor armature terminal voltage; I: motor armature current; R_a: equivalent resistance of motor armature circuit; L_a: equivalent inductance of motor armature circuit; e_a: EMF; C_e: motor electrical potential constant; C_m: motor torque constant; J_m: motor rotating inertia; M_a: motor electromagnetic torque; θ_m, ω_m, ω $\dot{\omega}_m$: motor rotation axis angular position, angular velocity and angular acceleration; N: gearbox ratios; K_g :equivalent stiffness of single-chain gearbox output end; J_a: antenna

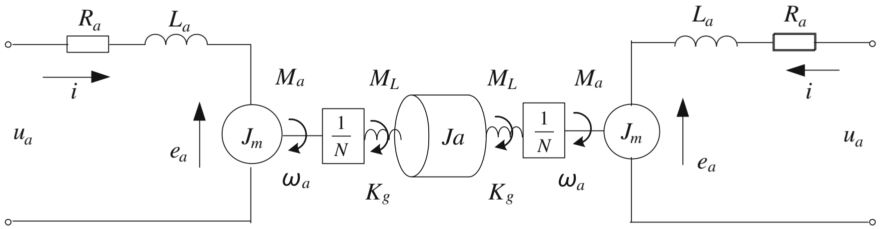


Fig. 16.1 Physical model of dual-motor driving system

rotating inertia;: torque from gearbox output end to antenna; θ_a, ω_a, $\dot{\omega}_a$: antenna shaft angular position, angular velocity and angular acceleration.

Considering the electromagnetic and mechanical laws of motion, differential equations can be got as follows:

$$\begin{aligned} U_a &= e_a + R_a i + L_a \frac{di}{dt} \\ e_a &= C_e \omega_m \end{aligned} \tag{16.1}$$

$$\begin{aligned} 2M_a &= 2C_m i \\ 2M_a &= \frac{1}{N} M_L + 2J_m \dot{\omega}_m \\ M_L &= 2K_g \left(\frac{1}{N}\theta_m - \theta_a\right) \\ M_L &= J_a \dot{\omega}_a \end{aligned} \tag{16.2}$$

Taking into account current loop having impact on the control object of speed loop, EMF influence is negligible. By formula (16.1), transfer function of current loop object is found

$$G_{ui}(s) = \frac{I(s)}{U_a(s)} = \frac{1}{R_a + L_a s} \tag{16.3}$$

K_g is generally large in actual system, so $\frac{1}{N}\theta_m \approx \theta_a$ and $\frac{1}{N}\dot{\omega}_m \approx \dot{\omega}_a$. With formula (16.2), transfer function of current i and motor rotation axis angular velocity ω_m is

$$G_{u\omega}(s) = \frac{\omega_m(s)}{U_a(s)} = \frac{C_m}{J_{\sum}(R_a + L_a s)s} \tag{16.4}$$

This transfer function is an integral equation. With the current loop in series, we get the speed loop control object. By defining $J_{\sum} = J_m + \frac{J_a}{2N^2}$, the transfer function of motor armature terminal voltage U_a and ω_m is

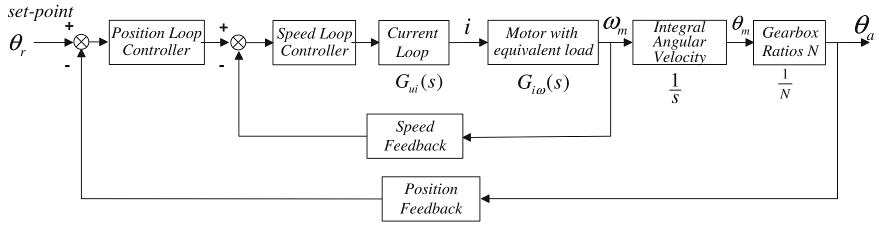


Fig. 16.2 Three-loop control block diagram of the servo system

$$G_{u\omega}(s) = \frac{\omega_m(s)}{U_a(s)} = \frac{C_m}{J_{\sum}(R_a + L_a s)s} \tag{16.5}$$

The system block diagram of the servo system is shown as Fig. 16.2.

In engineering applications, the speed loop controller can be simplified to a proportional component with K_v as its proportionality factor. Assuming velocity feedback coefficient is α_v, we can get the speed loop transfer function:

$$G_v(s) = \frac{K_v G_{u\omega}}{1 + \alpha_v K_v G_{u\omega}} = \frac{K_v C_m}{J_{\sum} L_a s^2 + J_{\sum} R_a s + \alpha_v K_v C_m} \tag{16.6}$$

Furthermore, the position loop transfer function of the control object is

$$G_p(s) = \frac{G_v(s)}{Ns} = \frac{K_v C_m}{Ns\left[J_{\sum} L_a s^2 + J_{\sum} R_a s + \alpha_v K_v C_m\right]} \tag{16.7}$$

The system is an integral process with a second-order system in series. Based on the above model, the position loop controller of the servo system can be designed. For convenience, the formula (16.7) is abbreviated as follows

$$G_p(s) = \frac{K}{s(T_\zeta^2 s^2 + 2\zeta T_\zeta s + \alpha_v)} \tag{16.8}$$

Considering PID algorithm is simple and the parameters are easy to adjust, we choose PID algorithm to design the ACU position loop controller. However, PID algorithm is based on the control deviation, and therefore, once there is interference effect, the corrective action would be adopted until the input impacts on the output and forming deviation. Obviously, the output always lags the input. Generally, in order to improve the response speed and the dynamic tracking performance, feed-forward compensation is added into PID controller. When the interference occurs, the corrective action works according to interference, directly. For continuous-time system, if the feed-forward unit multiplied by system close-loop transfer function is equal to 1, the output fully reproduces the input, and the bias caused by the

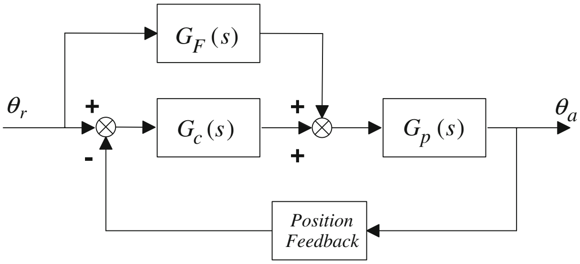


Fig. 16.3 Block diagram of position loop feed-forward and feedback control

disturbance can be completely eliminated in theory. $G_c(s)$ is the position loop controller and $G_F(s) = 1/G_p(s)$ is feed-forward controller. $G_c(s)$ and $G_F(s)$ are shown in Fig. 16.3.

16.3 Algorithm of DMC for ACU Position Loop

16.3.1 Principle of DMC

DMC is a model predictive control algorithm based on the step response model. It makes use of system predictive information and implements a rolling optimization strategy to optimize cost function in limited time domain, which can overcome the interference brought by the noises and the nondeterminancy of the system model. DMC includes the following three parts, i.e., prediction model, rolling optimization and feedback correction [7].

16.3.1.1 Predictive Model

Considering the sample values of the system step response $a_i = a(iT)$, where $i = 1, 2, \cdots$ and T is the sampling period. For a certain time there exists $a_i \approx a_N (i > N)$, then the system dynamic information can be described approximately by the vector $\boldsymbol{a} = [a_1 \quad \cdots \quad a_N]^T$, where N is modeling time-domain. At time k, suppose that the control input remains unchanged, and the future N predictive outputs are $\tilde{y}_0(k+i|k)$. Giving M increments$\Delta u(k); \cdots, \Delta u(k+M-1)$, the future outputs $\tilde{y}_M(k+i|k)$ at time k+i is:

$$\tilde{y}_M(k+i|k) = \tilde{y}_0(k+i|k) + \sum_{j=1}^{\min(M,i)} a_{i-j+1} \Delta u(k+j-1), i \in [1, N] \tag{16.9}$$

16.3.1.2 Rolling Optimization

The cost function adopted in DMC as follows:

$$\min_{\triangle u_M(k)} J(k) = \|w_P(k) - \tilde{y}_{PM}(k)\|_Q^2 + \|\triangle u_M(k)\|_R^2 \tag{16.10}$$

where $\tilde{y}_{PM}(k) = \tilde{y}_{P0}(k) + \boldsymbol{A}\triangle \boldsymbol{u}(k)$, $w_P(k) = \begin{bmatrix} w(k) & \cdots & w(k+P) \end{bmatrix}^T$, $\tilde{y}_{P0}(k) = \begin{bmatrix} \tilde{y}_0(k+1|k) \\ \vdots \\ \tilde{y}_0(k+P|k) \end{bmatrix}$, $\tilde{y}_{PM}(k) = \begin{bmatrix} \tilde{y}_M(k+1|k) \\ \vdots \\ \tilde{y}_M(k+P|k) \end{bmatrix}$, $\boldsymbol{A} = \begin{bmatrix} a_1 & \cdots & 0 \\ a_M & \ddots & a_1 \\ \vdots & \vdots & \vdots \\ a_P & \cdots & a_{P-M+1} \end{bmatrix}$ $\boldsymbol{A}$ is dynamic matrix; $w_P(k)$ is the set-point of the future output; Q and R are weighting matrices; M is control domain and P is optimization domain. By applying formula (16.9), the minimization of formula (16.10) is

$$\triangle u_M(k) = (AQA + R)^{-1} A^T Q[w_P(k) - \tilde{y}_{P0}(k)] \tag{16.11}$$

M optimal control moves $\triangle u_M(k)$ are calculated each time, but only the first control move $\triangle u(k)$ is implemented. At the next time, $\triangle u(k+1)$ can be solved. This measure is so called "rolling optimization" strategy.

16.3.1.3 Rolling Optimization

At time k, the control move $\triangle u(k)$ is implemented and the output prediction is

$$\tilde{y}_{N1}(k) = \tilde{y}_{N0}(k) + \boldsymbol{a}\triangle u(k) \tag{16.12}$$

$\tilde{y}_{N1}(k)$ is the output prediction when the control moves at time $k+1$ and keeps invariant in the future time instants, so it is the basis for constructing initial output prediction after shifting $\tilde{y}_{N1}(k)$ next time. In order to overcome the effects of model uncertainties and interference, at time $k+1$, the system output error is denoted as $e(k+1) = y(k+1) - \tilde{y}_1(k+1|k)$, which is used for feedback correction, and the modified future output prediction is described by

$$\tilde{y}_{cor}(k+1) = \tilde{y}_{N1}(k) + \boldsymbol{h}e(k+1) \tag{16.13}$$

where $\boldsymbol{h} = [h_1 \cdots h_N]^T$ is the correction vector.

With time changing, $\tilde{y}_{cor}(k+1)$ should be shifted to be the initial predictive value of time $k+1$. Denote

$$\tilde{y}_0(k+1+i|k+1) = \tilde{y}_{cor}(k+1+i|k+1), i \in [1, N-1] \tag{16.14}$$

For a stable system, $\tilde{y}_0(k+1+N|k+1)$ equals $\tilde{y}_{cor}(k+N|k+1)$ by approximation, and the shifting process can be represented as fallows

$$\tilde{y}_{N0}(k+1) \approx S\tilde{y}_{cor}(k+1) \tag{16.15}$$

where, shift matrix

$$S_{uns} = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix}$$

.

16.3.2 Improved DMC Algorithm for Integral Process

The above DMC algorithm can be used only for asymptotically stable system in view of selecting principle of the step response coefficients. For the position loop control object described by formula (16.8) which contains an integrator process, the step response coefficients $\tilde{y}_0(k+1+N|k+1)$can't be approximated at $\tilde{y}_{cor}(k+N|k+1)$ with the sampling time increasing. However, the integral process has the characteristics: the difference between two adjacent step response coefficients is constant after a sampling interval, therefore, we can still use the finite step response coefficients for model prediction [8]. The infinite-dimensional step response coefficient of the system with integral process can be described as

$$\boldsymbol{a}_{uns} = [\, a_1 \quad \cdots \quad a_{N-1} \quad a_{N-1} + \triangle a \quad a_{N-1} + 2\triangle a \quad \cdots \,]^T$$

Utilizing the integral process characteristics above motioned yields

$$\tilde{y}_{cor}(k+N+1) - \tilde{y}_{cor}(k+N) = \tilde{y}_{cor}(k+N) - \tilde{y}_{cor}(k+N-1) \tag{16.16}$$

That is

$$\tilde{y}_{cor}(k+N+1) = 2\tilde{y}_{cor}(k+N) - \tilde{y}_{cor}(k+N-1) \tag{16.17}$$

By applying formula (16.17) we can obtain the vector form:

$$\tilde{y}_{N0}(k+1) = S_{uns}\,\tilde{y}_{cor}(k++1) \tag{16.18}$$

where

$$S_{uns} = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & -1 & 2 \end{bmatrix}$$

.

The N step response coefficients $\boldsymbol{a} = [\,a_1 \quad \cdots \quad a_{N-1} \quad a_{N-1} + \triangle a\,]^T$ can be selected to construct dynamic matrix $\boldsymbol{A}$. For the position loop control system with integral process, the only change in improved DMC algorithm is the shift matrix.

16.3.3 Improved DMC Algorithm with Feed-Forward Compensation

Considering velocity feed-forward and acceleration feed-forward of a controller, $G_F(s)$ is given by

$$G_F(s) = \frac{s(2\zeta T_\zeta s + \alpha_v)}{K} = \frac{\alpha_v}{K}s + \frac{2\zeta T_\zeta}{K}s^2 \tag{16.19}$$

Suppose the system set value is $w(k)$ at time k, rewrite formula (16.19) by replacing differential operational with differential operation, the digital feed-forward controller output $\triangle u_f$ can be derived as fellows

$$\triangle u_f = K_{fv}(w(k) - w(k-1)) + K_{fa}[w(k) + w(k-2) - 2w(k-1)] \tag{16.20}$$

where $K_{fv} = \frac{\alpha_v}{KT}$ and $K_{fa} = \frac{2\zeta T_\zeta}{KT^2}$. The system input is

$$u(k) = u(k-1) + \triangle u(k) + \triangle u_f(k) \tag{16.21}$$

It is shown that in Fig. 16.4, under the combined effect of the DMC control and feed-forward control, at time k, the formula (16.12) needs to be modified as

$$\tilde{y}_{N1}(k) = \tilde{y}_{N0}(k) + \boldsymbol{a}\left[\triangle u(k) + \triangle u_f(k)\right] \tag{16.22}$$

The overall algorithm can be divided into off-line and on-line stage.

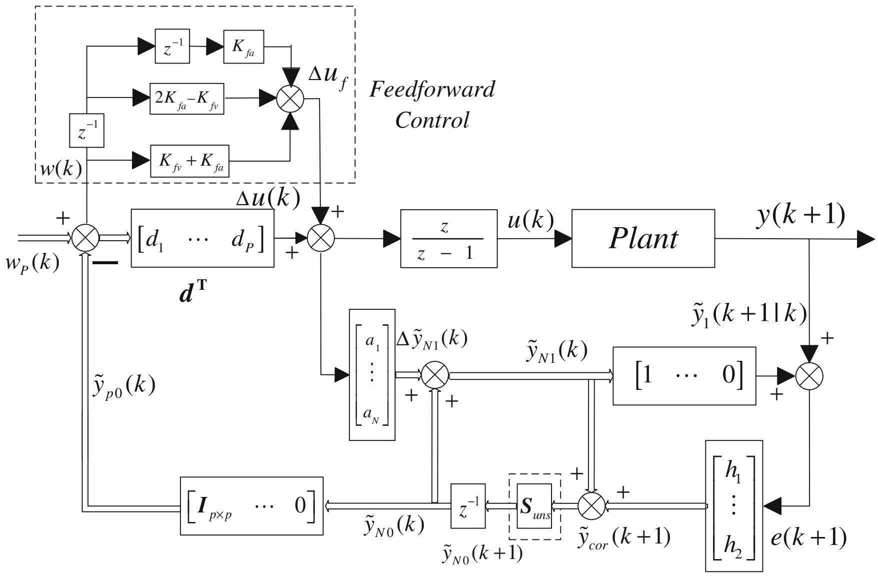


Fig. 16.4 Block diagram of improved dynamic matrix control with feed-forward compensation

Off-line stage: Determine system sampling and control time T and obtain step response coefficients $\boldsymbol{a} = [a_1 \quad \cdots \quad a_{N-1} \quad a_{N-1} + \Delta a\ ]^T$ via theoretical calculation or system identification, then calculate $(AQA + R)^{-1}A^TQ$.

On-line stage: at each $k \geq 0$,

Step 1. measure actual output to calculate the output error $e(k)$;
Step 2. correct the predicted value via formula (16.13);
Step 3. set the initial shift value via formula (16.18);
Step 4. calculate the DMC control move and feed-forward control move via formulas (16.11) and (16.20), respectively;
Step 5. implement the control inputs calculated via formula (16.21);
Step 6. calculate predictive output via formula (16.22).

16.4 Simulation Example

Consider the azimuth branch position loop of the servo control system with the transfer function

$$G_v(s) = \frac{2}{s(0.0012s^2 + 0.067s + 1)}$$

We adopt the PI algorithm and DMC algorithm to solve the set-point regulation problem and set trajectory tracking problem in simulation, then we compare the control performance of the two algorithms. Firstly, we build control model of the position loop via Simulink in Matlab R2008a, and choose the system sampling time T as $0.01s$. The parameters of PI algorithm are the suboptimal values obtained by the PID controller tuning parameters method. The DMC parameters are modeling time-domain $N = 30$, deviation weight matrix $Q = 1000\boldsymbol{I}_{P\times P}$ and control weight matrix $R = \boldsymbol{I}_{M\times M}$.

16.4.1 Set-Point Regulation Problem

The tuning PI control parameters are $K_p = 3.5$ and $T_i = 200$; For the DMC algorithm, the prediction-horizon $P = 7$ and the control time-domain $M = 1$. Due to the set-point adjustment problem, the feed-forward compensation control has no effect on the output result.

The input and output simulation results are shown in Fig. 16.5 (the left part is the input curve and the right output curve). We can see that DMC output reaches the set-point quickly and the regulation time is less than the PI algorithm. Therefore, the control performance of DMC algorithm is superior to PID algorithm. In addition, the simulation results also show that the two algorithms for set-point regulation problems can track without steady-state error.

16.4.2 Set Trajectory Tracking Problem

Select the target trajectory $r(t) = 0.1\sin(2\pi t)$ to test the tracking performance of the two algorithms. Figure 16.6 shows the tracking curves of the two algorithms without feed-forward compensation. Obviously, both of the two algorithms outputs

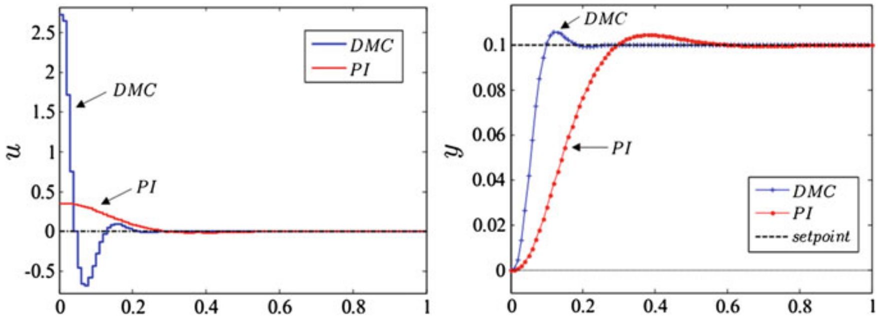


Fig. 16.5 Set-point control results of the PI algorithm and DMC algorithm

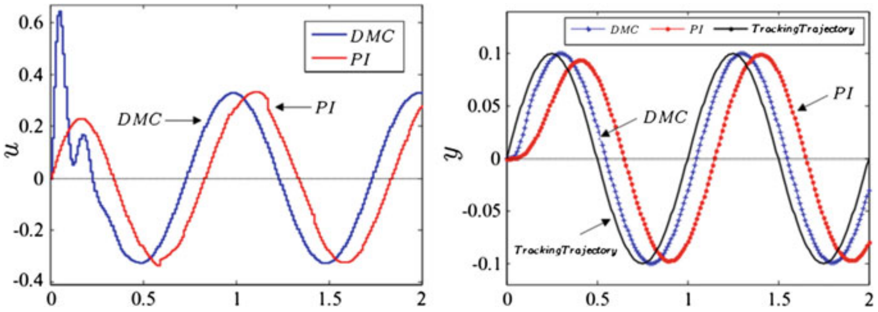


Fig. 16.6 Set trajectory tracking results of the PI algorithm and DMC algorithm without feed-forward compensation

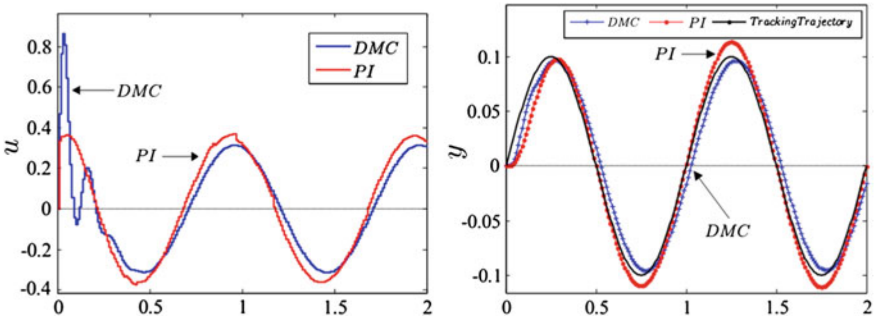


Fig. 16.7 Set trajectory tracking results of the PI algorithm and DMC algorithm with feed-forward compensation

significantly lag behind the tracing trajectory, but the lag time of DMC algorithm is less than that of the PI algorithm.

When the feed-forward compensation is joined, the tracking curve of the two algorithms is shown in Fig. 16.7. The feed-forward compensation significantly decreases the tracking lag, but the overshoot of DMC algorithm is less than that of the PI algorithm.

16.5 Conclusion

This paper focuses on the DMC algorithm for ACU position loop. Firstly, the ACU position loop control object is obtained by analyzing the current loop and speed loop mathematical models of the TT&C servo system. Then, the DMC method is introduced and an improved DMC algorithm with feed-forward compensation is given to deal with position loop with integral process. Simulation results indicate that the proposed DMC algorithm is superior to the classic PID algorithm. Therefore, the algorithm can be used to replace the current PID algorithm for ACU position loop control to improve the dynamic response and tracking performance in servo system.

References

1. Li QF (2007) An improved design for shipboard TT&C radar antenna control system. Dotoral thesis of Nanjing University of Science And Technology, Nanjing
2. Hou BJ, Li XQ et al (2009) The development of feed-forward plus PID controller for linear motor. Mach Tool Hydraulics 37(2):56–58
3. Hu QB, Lv ZY (2005) Feed-forward position controller design of full-digitalized servo system. Electric Drive 35(5):24–27
4. Li Q, Liang L (2003) Direct-current brushless servo control system based on fuzzy control. Control Theory Appl 20(4):547–550
5. Wang YM, Liu YY (2006) Adaptive control research for electro-hydraulic position servo system. Trans Chin Soc Agric Mach 37(12):160–163
6. Liu JK (2011) Advanced PID control and MATLAB simulation. Publishing House of Electronics Industry, Beijing
7. Xi YG (1993) Predictive control. National Defense Industry Press, Beijing
8. Zou T, Liu HB, Li SY (2004) Dynamic matrix control algorithm on the boiler level integral process. Control Theory Appl 21(3):386–390

Chapter 17
Research on Open-Loop Measurement Technique for Spacecraft

Wenquan Chen and Lei Huang

Abstract In order to solve the problem that Phase-Locked loop tracking measurement technique losses lock when the SNR is low or dynamic range is high, we do the research of spacecraft open-loop measurement technique without Phase-Locked loop. Open-Loop measurement technique includes Open-Loop data reception and data processing technique. Spacecraft Open-Loop distance measurement technique mainly adopts preset frequency band, FFT analysis, high-intermediate frequency or RF collection technique. Spacecraft Open-Loop data processing mainly includes Open-Loop Doppler data processing and Open-Loop distance measurement data processing. With high accuracy estimation and compensation for Doppler and Doppler rate, the random error of Open-Loop Doppler and distance measurement is about an order of magnitudes better than Closed-Loop. Open-Loop measurement technique has the disadvantage of complicated algorithms and huge amount of calculation, so it can't be implemented on the present FPGA. As a result, the paper presents a method to implement Open-Loop measurement technique.

Keywords Open-loop Doppler · Open-loop distance measurement · Phase-locked loop · Doppler compensation · Frequency spectrum calibration · Unified compute device architecture

17.1 Introduction

The difficulty for high accurate measurement of spacecraft mainly includes high dynamic range of the target and low SNR (signal-to-noise ratio). High dynamic range of the target refers to that due to the rapid movement of the target relative to

---

W. Chen (✉)
The 27th Research Institute of China Electronics Technology Group Corporation,
Zhengzhou 450047, China
e-mail: chwq722@126.com

L. Huang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_17

measure station, there is a large relatively velocity and relative acceleration and its higher order term between the transmitter and receiver, and the received signal will contain large Doppler and Doppler rate and its higher order term. Low SNR refers to that the target signal is so weak that it would be submerged in the thermal noise. The higher the dynamic state is, especially the higher Doppler rate and its higher order is, the more difficult design of the phase-locked loop is. So bandwidth of the loop must be wider so that it will not be easily unlocked. However, in this case, the performance of filtering noise will be poor and the measure accuracy will be poor, too. Under the condition of low SNR, the loop will be unlocked and measurement will not go on [1, 2].

Open-Loop data processing technique adopts several kinds of technologies such as Doppler, Doppler rate compensate search, Fourier transform, frequency and phase fitting and forward-estimating, rather than PLL(phase-locked loop) to measure the target. After compensating for Doppler and Doppler rate, integration time of signal will increase to improve the SNR. After the search, frequency and phase fitting and forward-estimating technologies are adopted to predict frequency and phase of the next moment so as to reduce computational complexity. Thanks to Open-Loop data processing technique, which can do accurate compensation for Doppler and its higher order term, frequency and phase can be measured in high-accuracy.

17.2 Open-Loop Data Reception Technique

Open-loop reception doesn't adopt feedback such as loop track and AGC (Automatic gain control) so that the signal will not be influenced by fluctuating of adjustment of the loop and feedback and it can truly reflect the movement of the target relative to measuring station. Open-loop receiver does not have the ability to make its bandwidth aligned (locked or tracked) with received signal, but depends on the forecast to set the bandwidth. This will lead a risk that forecast error may lead to faulty part on the receiving spectrum to be processed. In order to discern this problem, every sub-channel of the open-loop receiver analyzes the data and provides indication of the signal detected of the sub-channel. Except for indication of the signal detected, by using FFT analysis, every sub-channel provides the frequency domain band-pass characteristics of the signal recorded in this band. Users can control the characteristics parameter of FFT such as points, number of smooth and update rate [1].

As open-loop receiver has no AGC function, AD converter devices are acquired to adapt to large dynamic range. Under normal circumstances AD can hardly adapt to multiple kinds of signal situations of deep space signal, high-orbit spacecraft and near-earth spacecraft at the same time. The program controls channel design. Situation of signal amplification differs according to different measurement tasks. Channel control state remains unchanged in every measurement task so that the signal can reflect the movement of the target accurately as far as possible.

In order to reduce the influence of analog channel, AD sampling should be moved toward the front end so that it can be sampled at RF. Sampling at RF part, pass band of analog path can be as wide as possible. In this condition, a channel can receive various kinds of information and the signal to process occupies only a small part of the whole bandwidth. Besides, it has wonderful flatness, low dispersion and high signal fidelity in the bandwidth of signal.

For in-orbit spacecraft at the present time, the downlink signal has been confirmed and it is hard to change the form the signal to measure. Downlink signal usually refers to carrier, sub-carrier, pseudo random code, lateral sound signal and other general data signals. If open-loop receiving equipment can meet the collection of all kinds signal, analog channel pass-band should be as wide as possible. In order to reduce the amount of data collected, recorded and processed, the signal part of interest can be extracted according to the requirements and can be able to do multi-sub-band collection for the devices.

Figure 17.1 illustrates the relationship between the extracting flow and processing bandwidth of every sub-band of RSR (Radio Science Radio Science Receivers). Intermediate frequency signal selected by intermediate frequency switch feeds to digital converter (DIG), where the bandwidth range from 265 to 375 MHz limited by filtering and the center frequency is 320 MHz. Accordingly receiving frequency ranges from 2,265 to 2,375 MHz at S frequency band, 8,365–8,475 MHz at X frequency band and 31,965–32,075 MHz at Ka frequency band. However, the actual receiving frequency range depends on the characteristics of antenna equipment selected. The center frequency of filter's output signal will be down-converted to 64 MHz and the signal will be digitalized at 8-bit resolution at frequency of 256 MHz. Digitalized data will be sent to digital down-converter, which chooses any one 16 MHz bandwidth from the original pass bands with a resolution of 1 MHz and down-convert to baseband to form 16 Ms/s and 8-bit plural data-flow. Baseband processing provides as many as 4 sub-channel filters to choose the frequency band of interest to record. The number of usable filters depends on the selectable bandwidth and can be roughly divided into narrow-band, medium bandwidth and wide-band [3].

17.3 Open-Loop Doppler Measurement Technique

17.3.1 Calculation Method of Open-Loop Doppler

Because of movement of the target, signal received by receiving station contains Doppler message. No matter velocity or distance measurement, the more accurate Doppler compensation is, the more accurate velocity and distance measurement will be. The curve of the movement of target is continuous and step will not occur even with the rapid mobility. Besides, the velocity is continuous and because of being proportional to the velocity, Doppler is also continuous. Continuous function can be expanded in a Taylor series so that instantaneous Doppler $f_d(t)$ can be denoted as

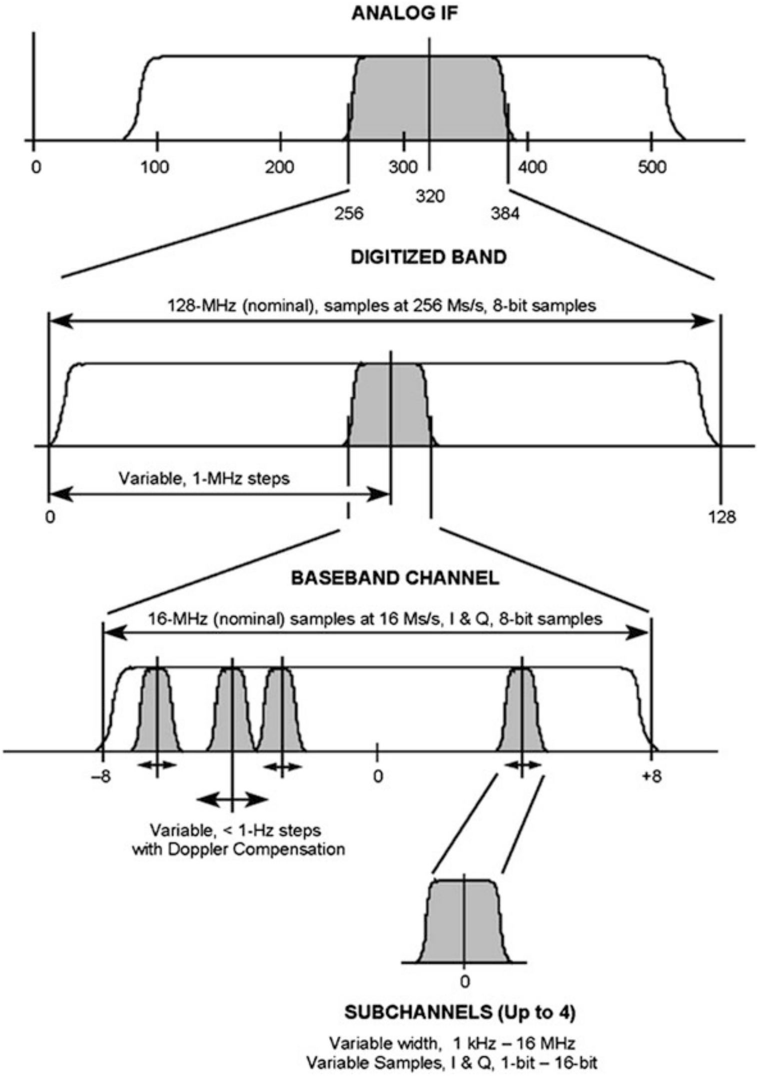


Fig. 17.1 Relationships between RSR processing bands

$$f_d(t) = f_d(t_0) + f_d^{(1)}(t_0)(t - t_0) + \frac{f_d^{(2)}(t_0)}{2!}(t - t_0)^2 + \cdots + \frac{f_d^{(n)}(t_0)}{n!}(t - t_0)^n + R_n$$

In the equation above, t_0 stands for reference time and time of polynomial expanded from instantaneous Doppler. $f_d(t_0)$ Denotes the Doppler of reference time and $f_d^n(t_0)$denotes n derivative of $f_d(t)$ at t_0. R_n is the high order infinitesimal of $(t - t_0)^n$ of $f_d(t)$. The greater the velocity of target on connection direction between

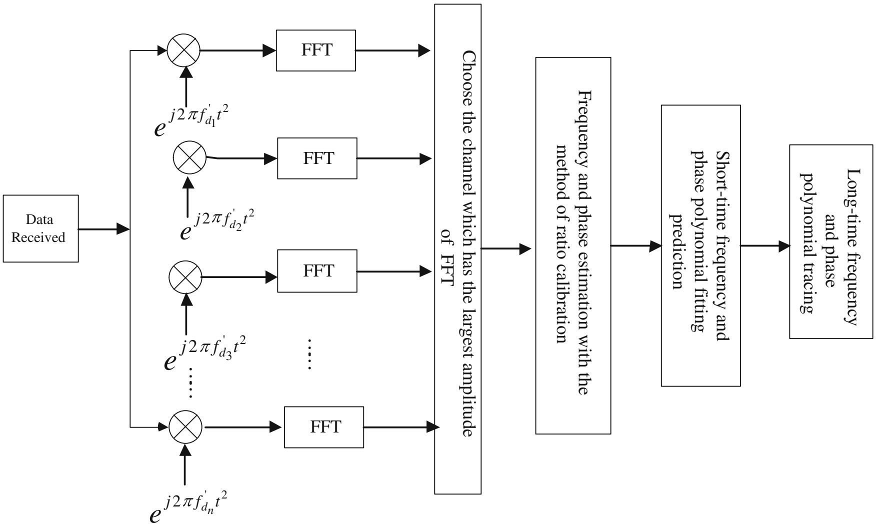


Fig. 17.2 Open-loop Doppler processing

the station and target is, the larger absolute value of $f_d(t)$, the greater the velocity of target maneuver or in the other direction is, and the greater the high-order item of Doppler is. In the estimation of frequency and phase, we always adopt FFT, which in fact searches for constant frequency and is accurate for the frequency and phase of fixed frequency. When higher order term of Doppler exists, the error of FFT transformation will be great if integration time is long. Integration time has to be increased to suppress the noise when SNR is low relatively. In order to increase integration time, 2nd higher order of Doppler should be compensated and other higher order will have little influence in a short time. Besides, one dimensional search will be added if higher order term is added, what will increase the number of calculation of multidimensional search. 2-dimensional can basically meet the need of estimation of low SNR and high dynamic range. The whole process is shown in Fig. 17.2.

Received data will do multichannel parallel one-order Doppler rate compensation and then FFT transformation to find greatest one of the amplitude of the spectrum. The corresponding Doppler rate is the coarse value of Doppler rate estimated. Then relatively accurate frequency and phase value can be obtained by method of ratio calibration. The search for Doppler and Doppler rate will carry on only in the absence of a priori information. After a piece of data searched, there is no need to make search of large amount of calculation if short-term frequency and phase polynomial fitting is adopted. The search only uses the estimation by the first short data segment to do polynomial fitting, gets Doppler and Doppler rate and estimates and compensates for the signal's frequency to be calculated. Higher accuracy will be obtained if long time track using calculated Doppler and phase occurs.

Ratio of spectrum correction method uses the window function ratio between two spectrum lines which have the difference value of main-lope peak nearby of 1 after frequency(window-length) is normalized to establish an equation whose normalized corrected frequency is treated as a variable. Then solve the normalized corrected frequency values and correct frequency, amplitude and phase. The relative error of ratio correction method is below 5 % [4].

Doing FFT for the truncated data is in fact to do FFT after the signal multiply by window function. The truncated spectrum of sine signal is the frequency where center frequency of the gate function translates from 0 to sine wave. The maximum and minimum value respectively correspond to the $(k+1)^{\mathrm{st}}$ and k^{st} spectrum lines. ∇f^1 stands for the frequency difference between the estimated value and real one. y_k denotes the amplitude of the k^{st} spectrum line and y_{k+1} the $(\mathrm{k}+1)^{\mathrm{st}}$ spectrum line.

For the rectangular window,

$$\nabla f^1 = -\frac{y_{k+1}}{y_{k+1} + y_k}$$

For Hanning window

$$\nabla f^1 = -\frac{y_k - 2y_{k+1}}{y_{k+1} + y_k}$$

The real frequency after correction is

$$f_0 = (k - \nabla f^1)\frac{f_s}{N}$$

In the equation, N denotes analysis point number of FFT and f_s stands for sampling frequency. Make sure that and respectively correspond to the front and back spectrum line of the maximum and sub-maximum values.

After correction, the phase becomes to

$$\theta = \phi_k + \pi(k - f_0^1) = \arctan(\frac{I_k}{R_k}) + \pi\nabla f^1$$

17.3.2 *Open Doppler Procession for Deep-Space Signal*

Measuring station is Kashgar deep-space station. Target carrier frequency is 8,419,226,562.5 Hz and 1-order subcarrier is 262,144 Hz. Sampling frequency of the two stations is 50,000 Hz. Four sub-channel respectively sample carrier message, the second sub-carrier, -14th sub-carrier and 20th sub-carrier of the target signal. The length of the data is 600s. Here take the carrier with the strongest signal and the 20th sub-carrier of with the weakest signal as an example.

From Fig. 17.3 we can see that without Doppler rate compensation, the spectrum lines of carrier Doppler have very wide bandwidth and the 20th sub-carrier have been completely submerged in noise.

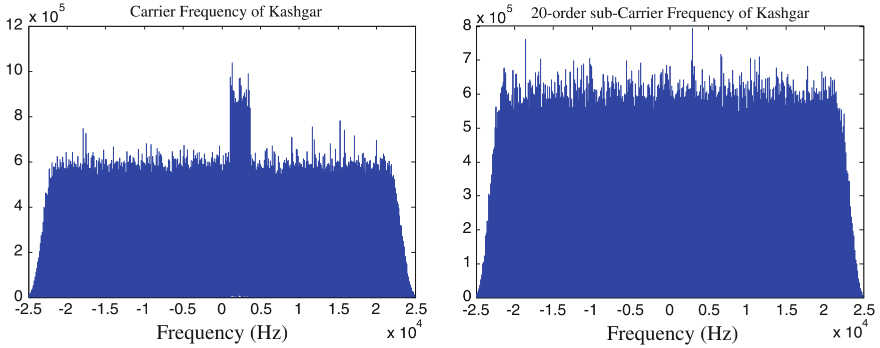


Fig. 17.3 Spectrum of carrier and the 20th sub-carrier without Doppler rate compensation in Kashgar station

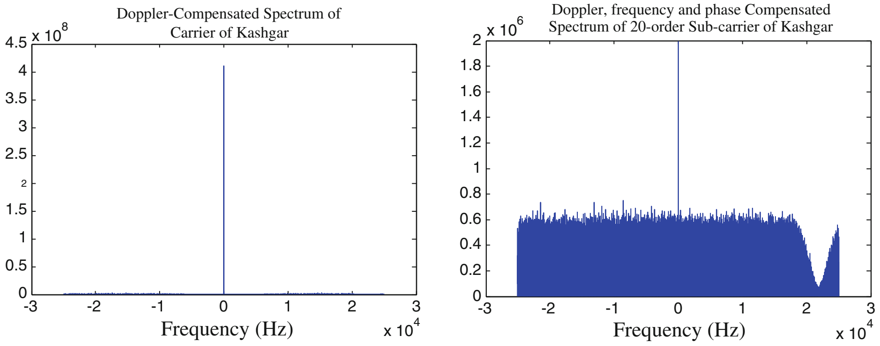


Fig. 17.4 Spectrum of carrier with Doppler compensation and the 20th sub-carrier with interpolation phase compensation in Kashgar station

From Fig. 17.4 we can see that after accurate Doppler rate compensation, spectrum of carrier doppler and the 20th sub-carrier have only one spectrum line.

Figure 17.5 shows the residual error of frequency after frequency compensation (in this condition the length of integration is 10s and the output gap is 1s). The frequency residual has an order of 10^{-4} Hz and the phase residual has an order of 10^{-3} circumference.

17.4 Open-Loop Distance Measurement Technique

17.4.1 Theory of Open-Loop Distance Measurement

Open-loop distance measurement requires the antenna to be able to receive and transmit. Acquisition equipment needs to record received target signal and

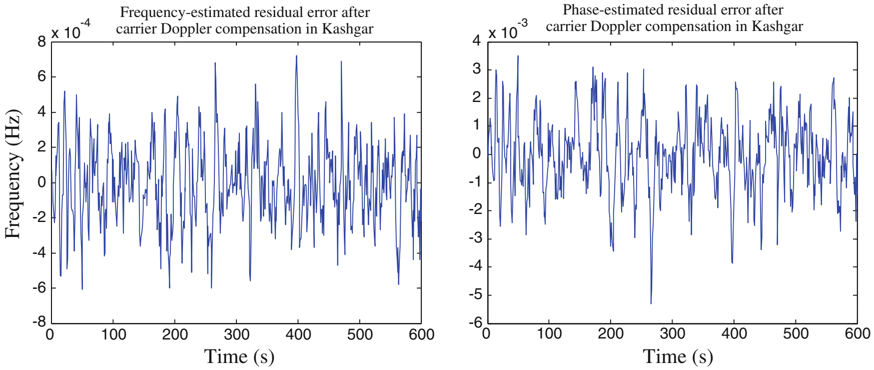


Fig. 17.5 Residual error of frequency and phase of the carrier in Kashgar station after Doppler compensation

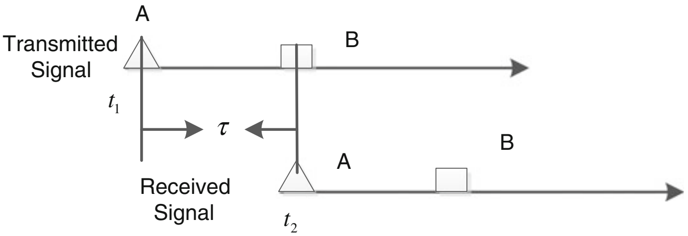


Fig. 17.6 Block diagram of the basic theory of distance measurement

baseband signal transmitted by transmitter at the same time. After comparison of two signals, we can obtain the distance from the target to the receiver.

As long as we know the time delay relationship between messages, we can get the time delay between receiver and transmitter according to the relationship between the messages of the receiver and transmitter at the same time to measure distance. From the Fig. 17.6, the delay between message A and B is τ and the transmitted signal at t_2 is B. The distance between the target and the measurement station when the received signal is A can be denoted as

$$R = \frac{C\tau}{2}$$

C stands for the speed of light. When measuring distance, we usually use a certain regularity of message such as distance measurement for side sound and pseudo code [5]. Even if without regular message, we can measure distance in open-loop using relevant processing technique. With the side sound distance measurement, the calculation method is similar to Doppler estimation method previously. After estimating the frequency and phase received and transmitted, we

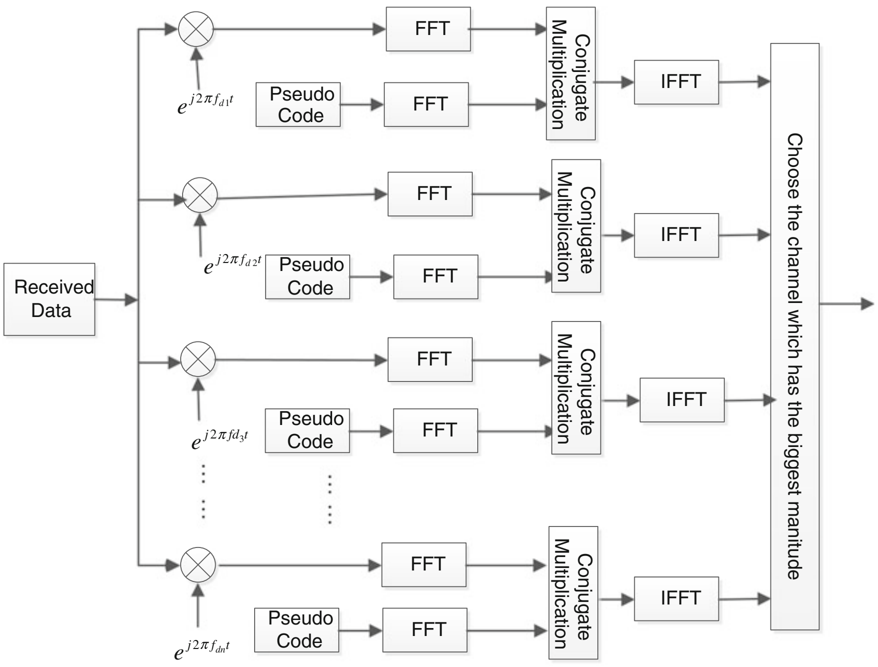


Fig. 17.7 Parallel search for Doppler pseudo code

can calculate the phase difference, low frequency solution of fuzzy, high fidelity of the same time.

Pseudo code distance measurement containing Doppler need parallel search for pseudo and Doppler. Search flow shows as Fig. 17.7. In FPGA, in order to save source, we usually adopt time-domain correlation, open-loop receive technique and general computer to implement and FFT/IFFT for search. After the search we extract the data de-spread spectrum filter directly, take the data's square and estimate the frequency and phase by piecewise FFT transformation. By polynomial fitting and forward predicting, estimate Doppler and Doppler rate. According to Doppler estimation we obtain the correlation result of the new pseudo code and estimate accurate phase of pseudo code gradually. Finally we get highly accurate measurement result by successive approximation of track of pseudo code and carrier.

17.4.2 Simulation of Open-Loop Distance Measurement

Simulation condition is as follows:

Carrier-frequency $f_c = 2.03\,\text{GHz}$, Doppler $f_d = 18{,}000\,\text{Hz}$, Pseudo code rate $f_{code} = 10.23\,\text{MHz}$, Code Doppler is $f_d/f_c * f_{code}$; Sampling rate is 32 MHz, Fragment the 500,000-points long data and the length of fragment is 1,024.

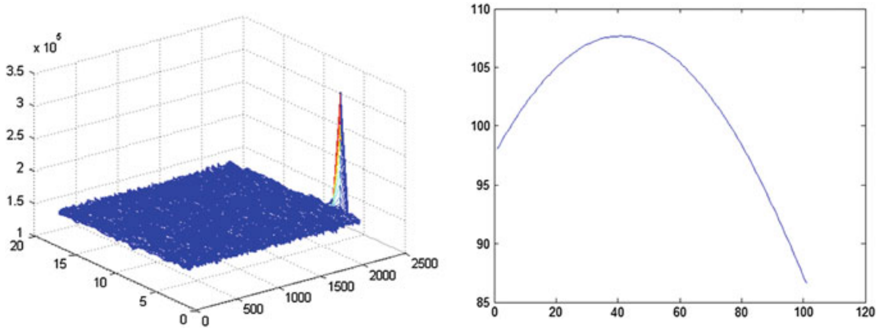


Fig. 17.8 Simulation diagram of pseudo code distance measurement

The beginning-time difference of couple of correlation-processed data is 512.4 samples of local pseudo code behind received one. One channel is supposed to be local signal which has no Doppler. Step length of Doppler search is 31,250 Hz. Doppler search has 17 channels in total, from 8 × 31,250 to 8 × 31,250 Hz.

Figure 17.8 is the simulation diagram of pseudo code distance measurement. The simulation condition is integer bit and fractional bit search without tracing received pseudo code. The left diagram is the result of integer bit search and the maximum value corresponds to searched frequency and place of integer bit. The right diagram illustrates the fractional bit search after integer bit and Doppler compensation. The accuracy of search can be as accurate as 0.001 sampling.

17.5 Implement of Open-Loop Measurement Technique

17.5.1 Implement of Data Collection Equipment for Open-Loop Receiving

In order to make the open-loop data collection equipment to meet the need of high accuracy open-loop velocity measurement, distance measurement and collection of various kinds of signal, we require dual-AD to sample two channels signal. AD can adapt to a wide range of input signal, has a high acquisition rate and bandwidth and can do multi sub-band collection. As the variety of the signal of spacecraft, especially in the form of wideband signal, if the rate of data collection is too high, the pressure of record and processing to signal will be too high. As a result, it can hardly meet the need of real-time or quasi real-time processing. In order to meet the need of real-time and quasi real-time processing, we demand that the sub-band width and dimension of truncation can be set. In order to make the collection equipment meet the need of difference of center frequency of RF or intermediate frequency at the front end, we require the center frequency of collection equipment can be set. In order to meet the above need of collection equipment, we put forward high request to the DA converter devices, FPGA processing devices and the record ability.

AD device chooses high speed AD—ADC12D1800, which has two AD channels with an output of standard LVDS signal conforming to IEEE 1596.3-1996 and meet the need of FPGA/ASIC input electrical level. It supports dual channel 1.8 GSPS or single channel 3.6 GSPS, 12 bit sampling rate and 2.15 GHz input bandwidth.

FPGA processing device chooses XILINX Virtex6 LX240T, whose frequency lock can support to 500 MHz. It has rich resources, for example 301440 registers, 150720 LUTs, 460 RAM/FIFO 36E1s and 768 DSP48E1 s. It adopts ISE 13.1 to program and implementation program, which has various IP core, such as FFT, FIR, RAM/ROM, PCI, PCIE and so on and can help programmers achieving functional requirements quickly.

The block diagram of Open-loop data reception data collection shows in Fig. 17.9. Two AD channels can collect two-channel reception signal at the same time and can collect a channel of reception and local-producing signal. With frequency conversion and low-pass filtering, the sub-band data will be formed. After buffered, sub-band data add frame-head and frame frame-tail to form sub-band data

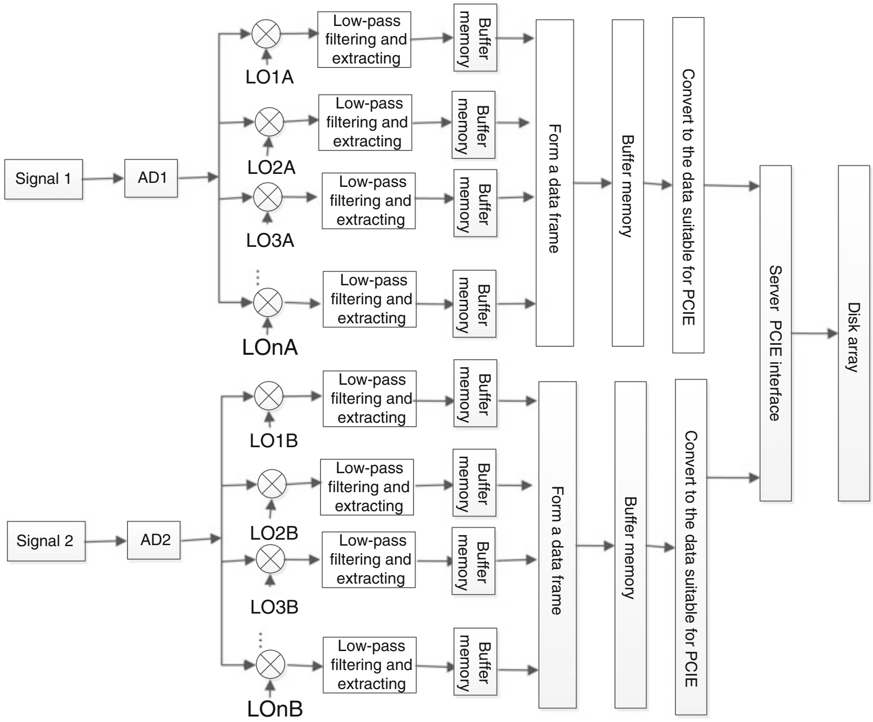


Fig. 17.9 The block diagram of open-loop data reception data collection

frame. The data frame forms the data suitable for PCIE transmission after buffer's changing rate. The collected data will be recorded by the disk array from PCIE on the server.

17.5.2 Implementation of Open-Loop Information Data Processing

Open-loop data processing algorithms are usually complicated and long data FFT, fitting and high accuracy phase estimation will occupy great resources, so it's not suitable to implement on FPGA. We consider open-loop received data processing using general computer's CPU + GPU (Graphic Processing Unit) to process. Real-time data processing by using general computer is always the dream of science and technology personnel. General computer can implement complicated algorithms, but the velocity is always bottleneck of real-time data processing. With the development of computer technology, especially with the emergence of the development environment of GPU calculation of CUDA (Compute Unified Device Architecture), we can treat GPU as equipment which can do parallel data calculation and distribute and manage for the operating computer. CUDA architecture builds around an extensible multi-threaded flow and multi-processor (SMs) array. When the CUDA program on the hose calls the kernel grid, the inside block of the grid enumerates and distributes to multiprocessor which has usable execution resources. Threads in block execute concurrently in more than one processor and multiple blocks can execute concurrently in one flow multiprocessor. When the threads terminate, it will transmit new blocks in the free multiprocessor. Flow multiprocessor is designed to perform hundreds of threads concurrently at the same time. In order to manage so many threads, flow multiprocessor adopts an architecture called SIMT (single instruction, multi-threads). Command flow uses instruction level within single thread to parallel and the flow multiprocessor uses multi-threads as well as hardware to parallel. Different from the core of CPU, the instructions transmit orderly and have no branch prediction and forecast executing [6].

To deal with the deep-space four-channel Doppler and difference Doppler above-mentioned, where integration length of the data is 100s and data will be calculated repeatedly for 100 times, it will take as long as 5 h if we use MATLAB. However, we only spend more than 10 min if we use a Tesla M2070 GPU card on NVIDIA. At present, the ability to calculate of latest TESLA K40 is nearly 3 times Tesla M2070, which is shown in Table 17.1 [7, 8]. The new TESLA K40 can meet the need of processing ability of deep-space Doppler estimation. We can use multiple chips GPU to implement a great amount of real-time calculation such as wideband data processing and search. In addition, for aerospace telemetering and command, the amount of calculation is great in the beginning, for example when we do frequency search and pseudo code search. When turning to tracking period, we can predict the information behind using the one in the front so that the amount of calculation will decrease. Relative to FPGA, GPU has great advantage that it can

Table 17.1 Device performance of NVIDIA

	Tesla M2070/M2075	TESLA K40
Double-precision floating-point performance (Peak)	515 Gflops	1.43 Tflops
Single-precision floating-point performance(Peak)	1.03 Tflops	4.29 Tflops
Special storage capacity	6 GB GDDR5	12 GB GDDR5
Memory bandwidth	150 GB/s	288 GB/s
CUDA core number	448	2,880

use nearly all the resources to search rapidly. Because the amount of data to process become smaller and smaller, it will quickly get into real-time measuring phase even though it can't totally process the data in real time.

17.6 Conclusion

As Doppler compensation to the received signal is accurate, open-loop technique can increase integration, improve the calculation accuracy and solve the problem that closed-loop receiving method can't measure in case of high dynamic range of the target or low SNR. Open-loop receiving technique is more complicated than closed-loop receiving. However, compared with the closed-loop, computation dependency between the forward and behind is reduced greatly, so parallel computation can be used to increase the speed. Adopting CPU + GPU in which CPU is used to do process control and GPU is used to do high-speed parallel process, we can implement aerospace measurement.

References

1. Li SC, Li JH, Sun JH et al (2011) A modified carrier tracking loop using an extended Kalman filter in high dynamic conditions. J Harbin eng univ 32(11):1509–1513
2. Lu HB, Ang WW, Gu QC (2010) Carrier synchronization technique for low SNR and high dynamic condition. Inf Control 39(4):451–454
3. 810-005, Rev. E DSMS (2009) Telecommunications link design handbook. 209 Open-loop radio science
4. Ding K, Xie M, Yang ZJ (2008) The theory and technology of discrete spectrum correction. Science Press, Beijing
5. Xie HW, Zhang M (2000) Space TT&C system. Press of National University of Defense Technology, Changsha
6. (2012) CUDA C programming guide. PG-02829-001_v5.0
7. TESLA™ M-class GPU computing modules accelerating science. http://www.nvidia.com/tesla/ DS-Tesla-M-Class-Aug11
8. NVIDIA® TESLA® GPU accelerators. http://www.nvidia.com/tesla/Tesla-KSeries- Overview - LR

Chapter 18
A Multi-parameter Joint Demodulation Algorithm Based on Alternating Projection

Ge Wang, Peijie Zhang and Wei Liu

Abstract For unknown synchronization parameters symbol detection problem, this paper gives the joint estimation processing structure of symbol detection and carrier phase based on alternating projection. For the traditional hierarchical approach, carrier phase estimation is required before symbol detection. the traditional optimum reception algorithm, parameter estimation and symbol detection will be optimized separately, that may not get a global joint optimal solution. This paper studies this problem, and proposes a multi-parameter joint processing algorithm based on alternating projection. Simulation results show that the algorithm outperforms the traditional optimum reception algorithm, which estimates the carrier phase using the non-data-aided method before symbol detection. Simulation results show that the algorithm can effectively enhance the performance of joint detection.

Keywords Signal detection · Alternating projection · SER · Joint process

18.1 Introduction

Satellite measurement and control communication system is to achieve access to information, technical support and functional security information transmission and information control and other functions, is the only channel satellite to keep in touch with the ground, with the transmission and read and control signals to the satellite control station can be completed in tracking test track, telemetry, remote control and many other features and world communication.

G. Wang (✉) · P. Zhang · W. Liu
Movable TT&C Station, Xi'an Satellite Control Center, Weinan 714000, China
e-mail: superwangge@sohu.com

P. Zhang
e-mail: 124145636@qq.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_18

Unified S-band satellite (USB) measurement and control system is characterized by the use of sub-carrier frequency division multiplexing, through the carrier and subcarrier modulation to achieve measurement and control. Sub-carrier is the directly carrier of measurement and control, research satellite telemetry signal demodulation algorithm for satellite USB system has important academic significance and application value.

In signal demodulation, for the problem of symbol detection in the presence of unknown synchronization parameter, the traditional optimum receive method is estimated the synchronization parameter first, compensated the received signal by estimation results later. In the non-cooperative reception environment due to the non-data-aided parameter estimation performance is poor. Therefore, has a great impact for data detection performance [1, 2].

In joint process different levels of single data stream, the part of signal demodulation mainly to complete the synchronization parameter estimation and symbol detection. Synchronization performance has a direct impact on the symbol detection results. The estimation of the carrier phase and symbol timing may be accomplished separately as described above or jointly. Joint ML estimation of two or more signal parameters yields estimated that are as good and usually better than the estimates obtained from separate optimization of the likelihood function. In other words, the variances of the signal parameters obtained from joint optimization are less than or equal to the variance of parameter estimates obtained from separately optimizing the likelihood function. Reference [2] proposed that the carrier frequency, carrier phase and symbol timing can be estimated together. Reference [3] mention that one can combine the parameter estimation problem with the demodulation of the information sequence. Thus one can consider the joint maximum-likelihood estimation of information sequence, the carrier phase, and the symbol timing parameter. Hisashi Kobayashi 1971 explanation the problem of sequence decision, sample timing and carrier phase recovery in a class of linear modulation data transmission systems are treated from the viewpoint of multi-parameter estimation theory in reference [4]. The structure of the maximum likelihood estimator is first obtained, and a decision-directed receiver is then derived. Reference [5] proposed the idea of multi-parameter joint processing, but the optimal solution did not get for these parameters, and still deal with the problems by dimension reduction and decision feedback methods to get the maximum likelihood estimate of information sequence, carrier phase and symbol timing. Reference [6] proposed a joint sequence detection and phase estimation algorithm based on EM, the algorithm directly calculated the carrier phase by the previous symbol decision value. Reference [7] proposed that taking the average of the symbol by the EM algorithm, estimated the carrier phase while the symbol soft information is obtained. The maximum value of the posterior probability of the symbol as a symbol hard decision results, the algorithm is also based on the NDA carrier phase estimation algorithm. Therefore, the bit error rate performance is still affected.

In the present paper we will focus on the problem of multi-parameter joint demodulation. We proposed a multi-parameter joint processing algorithm based on alternating projection. The rest of the paper is organized as follows; Sect. 18.2

introduced the traditional optimum receives algorithm. The proposed algorithm is developed in Sect. 18.3. In Sect. 18.4, simulation results and analysis are given. Finally, conclusions are drawn in Sect. 18.5.

18.2 The Traditional Optimum Receives Algorithm

18.2.1 Algorithm Describes

We wish to design a signal detector that makes a decision on the transmitted signal in each signal interval based on the observation of the vector **r** in each interval such that the probability of a correct decision is maximized. With this goal in mind, we consider a decision rule based on the computation of the posterior probabilities defined as $p(signal\,\mathbf{s}_m\,was\,transmitted\,|\mathbf{r})\quad(m = 1, 2, \ldots, M)$ which we abbreviate as $p(\mathbf{s}_m|\mathbf{r})$. The decision criterion is based on selecting the signal corresponding to the maximum of the set of posterior probabilities $\{p(\mathbf{s}_m|\mathbf{r})\}$. Later, we show that this criterion maximizes the probability of a correct decision and, hence, minimizes the probability of error. This decision criterion is called the maximum a posterior probability (MAP) criterion.

Using Bayes' rules, the posterior probability may be expressed as

$$p(\mathbf{s}_m|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s}_m)p(\mathbf{s}_m)}{p(\mathbf{r})} \tag{18.1}$$

where $p(\mathbf{r}|\mathbf{s}_m)$ is the conditional pdf of the observed vector given $\mathbf{s}_m$, and $p(\mathbf{s}_m)$ is the a priori probability of the mth signal being transmitted. The denominator of (18.1) may be expressed as

$$p(\mathbf{r}) = \sum_{m=1}^{M} p(\mathbf{r}|\mathbf{s}_m)p(\mathbf{s}_m) \tag{18.2}$$

Some simplification occurs in the MAP criterion when the M signals are equally probable a priori, i.e. $p(\mathbf{s}_m) = 1/M$ for all M. Furthermore, we mote that the denominator in (18.1) is independent of which signal is transmitted. Consequently, the decision rule based on finding the signal that maximizes $p(\mathbf{s}_m|\mathbf{r})$.

The conditional pdf $p(\mathbf{r}|\mathbf{s}_m)$ or any monotonic function of it is usually called the likelihood function. The decision criterion based on the maximum of $p(\mathbf{r}|\mathbf{s}_m)$ over the M signals is called the maximum-likelihood (ML) criterion. We observe that a detector based on the MAP criterion and one that is based on the ML criterion make the same decision as long as the a priori probabilities $p(\mathbf{s}_m)$ are all equal, i.e., the signal $\mathbf{s}_m$ equiprobable.

In the case of an AWGN channel, the likelihood function is

$$p(\mathbf{r}|\mathbf{s}_m) = \frac{1}{(\pi N_0)^{N/2}} \exp\left[-\sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0}\right] \quad (m = 1, 2, \ldots, M) \tag{18.3}$$

To simplify the computations, we may work with the natural logarithm of $p(\mathbf{r}|\mathbf{s}_m)$, which is a monotonic function. Thus,

$$\ln p(\mathbf{r}|\mathbf{s}_m) = -\frac{1}{2} N \ln(\pi N_0) - \frac{1}{N_0} \sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0} \tag{18.4}$$

The maximum of $\ln p(\mathbf{r}|\mathbf{s}_m)$ over $\mathbf{s}_m$ is equivalent to finding the signal $\mathbf{s}_m$ that minimizes the Euclidean distance

$$D(\mathbf{r}, \mathbf{s}_m) = \sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0} \tag{18.5}$$

We call $D(\mathbf{r}, \mathbf{s}_m) \quad (m = 1, 2, \ldots, M)$ is the distance metrics. Hence, for the AWGN channel, the decision rule based on the ML criterion to finding the signal $\mathbf{s}_m$ that is closest in distance to the receive signal vector $\mathbf{r}$. We shall refer to this decision rule as minimum distance detection.

In summary, we have demonstrated that the optimum ML detector computes a set of M distance $D(\mathbf{r}, \mathbf{s}_m)$, and selects the signal corresponding to the smallest (distance) metric. The above development for the optimum detector treated the important case in which all signal are equally probable. In this case, the MAP criterion is equivalent to the ML criterion. However, when the signals are not equally probable, the optimum MAP detector bases its decision on the probabilities $p(\mathbf{s}_m|\mathbf{r}) \quad (m = 1, 2, \ldots, M)$, given by (18.1) or, on the metrics

$$PM(\mathbf{r}, \mathbf{s}_m) = p(\mathbf{r}|\mathbf{s}_m) p(\mathbf{s}_m) \tag{18.6}$$

18.2.2 The Problem of Traditional Optimum Receives Algorithm

The fundamental purpose of communication is reliable transmission. The receiver should recover the original signal from the error signal as much as possible. In the traditional optimum receive structure, cascaded subsystem to optimize the design of each subsystem separately according to their respective capabilities. The decision value of each module is transmitted between each subsystem. The process gain cannot access from pre-system to the stage. From the perspective of information processing, performance loss always exist when information processing. Therefore, this does not make full use of the information carried by the received signal. So the receiver sensitivity is lower and the receiver capability is limited.

We have shown in last section, the likelihood function $p(\mathbf{r}|\mathbf{s}_m)$ is important to the optimum decision rule achieve. But, for the complex structures of several subsystems cascade, described the input-output relationship with probabilistic is difficult. We considered the impact of the actual channel on the transmitted signal, so using ML criterion or MAP criterion to achieve the overall optimal detection almost infeasible.

18.3 Joint Multi-parameter Demodulation Algorithm Based on Alternating Projection

18.3.1 Signal Model

When the target signal generation and construction, there is some correlation between the different information levers, for example channel code, estimation and detection and so on. The optimum receive is joint multi-level information processing, and obtain the global optimum.

As the global joint search computationally intensive, the alternating projection algorithm performance was significantly better than the traditional hierarchical optimization approach. The symbol detection and carrier phase estimation between multi-level processes is studied in this paper.

Suppose that timing information is completely known, the matched filter outputs sampled per symbol signal is

$$r_k = a_k e^{j\theta} + n_k, \quad k = 0, 1, 2, \ldots, N-1 \tag{18.7}$$

Let $\mathbf{r} = [r_0, r_1, r_2, \ldots, r_{N-1}]^T$ is the received signal vector, $\mathbf{a} = [a_0, a_1, a_2, \ldots, a_{N-1}]^T$ is the transmitted symbol vector. The conditional probability density of $\mathbf{r}$, $\mathbf{a}$ and θ is:

$$p(\mathbf{r}|\mathbf{a},\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}\left|r_k - a_k e^{j\theta}\right|^2\right) \tag{18.8}$$

18.3.1.1 The Traditional ML Method

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathrm{ML}} &= \arg\max_{\boldsymbol{\theta}}\{p(\mathbf{r}|\boldsymbol{\theta})\} = \arg\max_{\boldsymbol{\theta}}\left\{\int p(\mathbf{r},\mathbf{a}|\boldsymbol{\theta})d\mathbf{a}\right\} \\ &= \arg\max_{\boldsymbol{\theta}}\left\{\int p(\mathbf{r}|\mathbf{a},\boldsymbol{\theta})p(\mathbf{a}|\boldsymbol{\theta})d\mathbf{a}\right\} = \arg\max_{\boldsymbol{\theta}}\{E_{\mathbf{a}}[p(\mathbf{r}|\mathbf{a},\boldsymbol{\theta})]\}\end{aligned} \tag{18.9}$$

When the signal carries the information sequence, we can adopt one of two approaches: either we assume that information sequence is known or we treat as a random sequence and average over its statistics. In decision-directed parameter estimation, we assume that the information sequence over the observation interval has been estimated and, in the absence of demodulation errors. In this case, information sequence is completely known except for the carrier phase, decision-directed phase estimation was obtained. Instead of using a decision-directed scheme to obtain the phase estimate, we may treat the data as random variables and simply average the likelihood function over these random variables prior to maximization.

18.3.1.2 Global Joint ML Method

Joint maximum likelihood symbol detection and phase estimation can be obtained by maximize the joint maximum likelihood function. Thus,

$$\begin{aligned}\left[\hat{\mathbf{a}}_{\mathrm{GML}} \quad \hat{\boldsymbol{\theta}}_{\mathrm{GML}}\right] &= \arg\max_{\mathbf{a},\boldsymbol{\theta}}\{p(\mathbf{r}|\mathbf{a},\boldsymbol{\theta})\} \\ &= \max_{\mathbf{a},\boldsymbol{\theta}}\sum_{k=0}^{N-1}\mathrm{Re}\left\{r_k a_k^* \mathrm{e}^{-j\theta}\right\} - \frac{1}{2}|a_k|^2 \qquad (18.10)\end{aligned}$$

Clearly, its global optimal solution can be obtained according to two-dimensional search of the likelihood function, but its computational complexity grows exponentially with the number of symbols.

18.3.2 The Proposed Algorithm

For practical signal processing problems, we often encounter multi dimensional parameter for extreme value problems. There are some effective methods of initialization for dimensional optimization problem, alternating projection algorithm (Alternating Projection, AP) is one of them, the algorithm is proposed by Ziskind and Wax for multiple signal DOA estimation. The basic idea is this: First, assume that the received signal contains only one multi path echoes, using a single signal parameter estimation method to estimate the parameters, and then increments one way the multi path echoes and estimate its parameters, until all L way multi path parameters of the signal was estimated.

The idea of alternating projection is conditions maximize iteration, the results of $i + 1$ can be expressed as:

$$\hat{\theta}^{(i+1)} = \arg\max_{\theta} p\left(\mathbf{r}|\hat{\mathbf{a}}^{(i)}, \theta\right) = \max_{\theta} \mathrm{Re}\left\{ \mathrm{e}^{j\theta} \sum_{k=0}^{N-1} r_k \left(\hat{a}_k^{(i)}\right)^* \right\} = \arg\left\{ \left(\hat{\mathbf{a}}^{(i)}\right)^H \mathbf{r} \right\} \tag{18.11}$$

$$\hat{a}_k^{(i+1)} = \arg\max_{a_k} p\left(\mathbf{r}|a_k, \hat{\theta}^{(i+1)}\right) = \max_{a_k} \mathrm{Re}\left\{ \mathrm{e}^{j\theta} \sum_{k=0}^{N-1} r_k a_k^* \mathrm{e}^{-j\hat{\theta}(i+1)} \right\} - \frac{1}{2}|a_k|^2 \tag{18.12}$$

Like other iterative algorithms, the iteration initial value of algorithm need to give before the first iteration, the initial value selection is very important to algorithm ultimately performance. Appropriate initial value not only to ensure a greater probability of iteration scan converge to the global optimal solution, and can accelerate the speed of convergence of the iteration. Both to ensure the accuracy of the estimate, but also reduces the computational of the algorithm. If the initial value of the algorithm is not sufficiently close to the ML solution, the algorithm may converge to a local minima rather than the global extreme point.

So we use the traditional V&V algorithm to get a rough estimate as the initial phase of alternating projection algorithm. The initial value is close to the ML estimation of the carrier phase. The same can also be used to solve the problem that initial symbol.

So the multi-parameter (assuming two parameters is θ and φ) joint process algorithm based on alternating projection can be described as follow,

Step (1) : Set initial value of the estimate parameter $\varphi_{GML}^{(0)}$;

Step (2) : With the $\varphi_{GML}^{(i)}$, maximize the objective function $p(\mathbf{r}\,|\varphi, \boldsymbol{\theta})$,

$$\hat{\boldsymbol{\theta}}_{GML}^{(i)} = \arg\max_{\boldsymbol{\theta}} \left\{ p\left(\mathbf{r}\middle|\hat{\varphi}_{GML}^{(i)}, \boldsymbol{\theta}\right) \right\} \tag{18.13}$$

Step (3) : With the $\theta_{GML}^{(i)}$, maximize the objective function $p(\mathbf{r}\,|\varphi, \boldsymbol{\theta})$,

$$\hat{\varphi}_{GML}^{(i+1)} = \arg\max_{\varphi} \left\{ p\left(\mathbf{r}\middle|\varphi, \boldsymbol{\theta}_{GML}^{(i)}\right) \right\} \tag{18.14}$$

Step (4) : Repeat step (2) and step (3), until algorithm convergence.

Conventional demodulation processing methods: first, estimate the carrier phase, compensate the received signal phase according the estimation result to, and last detect symbol based on the minimum Euclidean distance criterion.

18.4 Algorithm Performance Simulations

The performance of the proposed joint global multi-parameter processing algorithm based on alternating projection is evaluated by computer simulations. In simulations, the transmitted signal was selected satellite monitoring and control system in a unified S-band telemetry subcarriers MPSK signals with $M = 2, 4$. In particular, the symbol error rate (SER) and the carrier phase estimate performance are measured with simulations.

18.4.1 Ser

The Sect. 18.3.2 simulate the symbol error rate performance for BPSK and QPSK signal, compared with the EM-based joint symbol detection and parameter estimation in the literature [8, 9] given by numerical integration methods and the analytical solutions. From Figs. 18.1 and 18.2 we can see that Multi-parameter joint process algorithm based on alternating projection compared with the numerical integration methods. The EM algorithm to calculate the soft information of data approximated by numerical integration, because of the integral scope and the size of rectangle in rectangular numerical integration method, the performance of symbol error rate is worse than the algorithm based on alternating projection. And compared with the joint detection and estimation algorithm based on EM, The simulation result shows that the two algorithms is equivalent on the treatment methods, therefore no change detection performance.

Fig. 18.1 SER of alternating projection algorithm (BPSK)

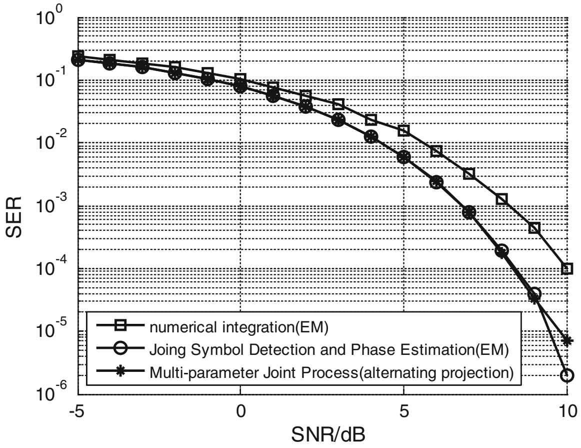
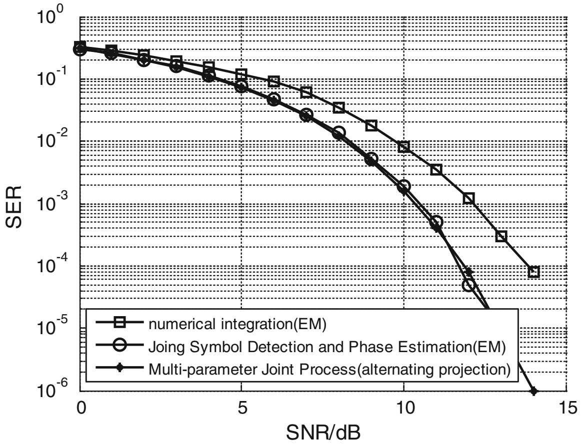


Fig. 18.2 SER of alternating projection algorithm (QPSK)



18.4.2 Carrier Phase Estimation

For Sect. 18.3.2 algorithm described, Figs. 18.3 and 18.4 respectively indicate that the standard deviation of phase estimation performance curves when the carrier phase offset $\theta = \frac{4}{9}\pi$, the BPSK modulated signal and the carrier phase offset $\theta = \frac{2}{9}\pi$, the QPSK modulated signal. Compared with the joint detection and estimation algorithm based on EM and the MCRB, The simulation result shows that the two algorithms is equivalent on the treatment methods, therefore the carrier phase estimation performance no obviously change.

Fig. 18.3 Carrier phase estimation variance of alternating projection algorithm (BPSK)

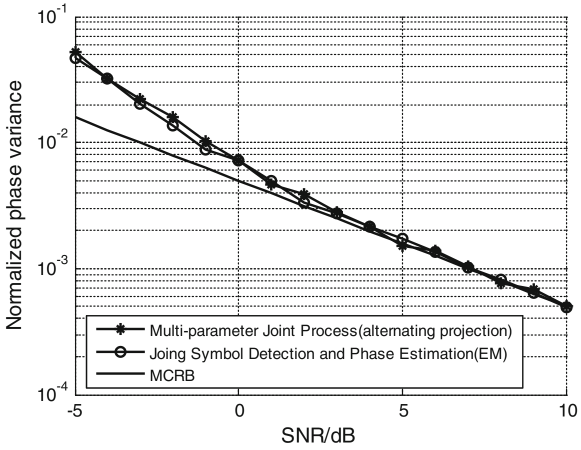
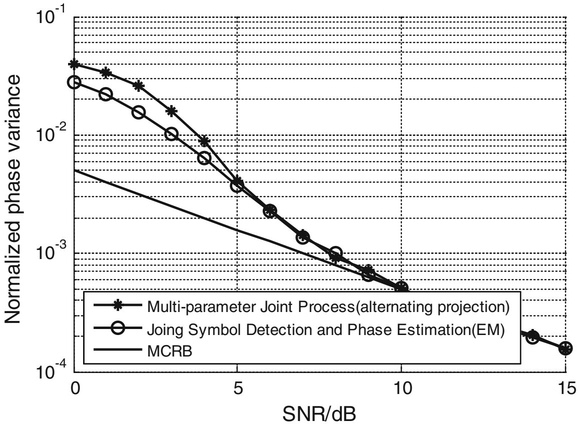


Fig. 18.4 Carrier phase estimation variance of alternating projection algorithm (QPSK)



18.5 Conclusions

In this paper, the multi-parameter signal demodulation processing problems is studied. For the traditional hierarchical approach, carrier phase estimation is required before symbol detection. the traditional optimum reception algorithm, parameter estimation and symbol detection will be optimized separately, that may not get a global joint optimal solution. This paper proposes a joint global multi-parameter processing algorithm based on alternating projection. Simulation results show that the algorithm outperforms the traditional optimum reception algorithm, which estimates the carrier phase using the non-data-aided method before symbol detection. Simulation results show that the algorithm can effectively enhance the performance of joint detection.

References

1. Moeneclaey M, Jonghe G (1994) ML-oriented NDA carrier synchronization for general rotationally symmetric signal constellations. IEEE Trans Commun 42(8):2531–2533
2. Gworghiades CN (1997) Blind carrier phase acquisition for QAM communication. IEEE Trans Commun 45(11):1477–1486
3. Proakis JG (1995) Digital communications. McGraw-Hill, New York
4. Kobayashi H (1971) Simultaneous adaptive estimation and decision algorithm for carrier modulated data transmission systems. IEEE Trans Commun Technol 19(3):268–280
5. Zamiri-Jafarian H, Pasupathy S (1997) Adaptive MLSDE using the EM algorithm. IEEE Trans Commun 47(8):1181–1193
6. Nassar CR, Soleymani MR (1994) Joint sequence detection and phase estimation using the EM algorithm. In: Conference on electrical and computer engineering, vol 1. pp 296–299
7. Cacciamani ER, Wolejsza CJ (1971) Phase-ambiguity resolution in a four-phase PSK communications system. IEEE Trans Commun Technol 19(6):1200–1210

8. Wang G, Shen ZX, Yu HY (2014) Joint symbol detection and phase estimation algorithm based on expectation-maximum. J Jilin Univ (Eng Sci) 44(2)
9. Wang G, Yu HY, Shen ZX (2012) An improved symbol detection algorithm based on expectation-maximum. In: Lecture note in electrical engineering

Chapter 19
Compressive Sensing for DS TT&C Signals Based on Basic Dictionary Building

Yanhe Cheng, Wenge Yang and Jiang Zhao

Abstract A notable trend of the TT&C system is the broadband, which would bring high-speed sampling pressure and big data problem. In this paper, the compressive sensing for DS TT&C signals based on basic dictionary building is presented in response to big data problem. The sparsity of the DS TT&C signal is analyzed by building basic dictionary firstly. Then based on delay-doppler basic dictionary which is built by theoretical analysis, the performance of the sparse representation and compressive sensing for the DS TT&C signal is studied by the simulation experiment. The results of simulation show that the DS TT&C signal gets a strong sparsity on the delay-doppler basic dictionary, the compressive sensing for the DS TT&C signal is feasible which can effectively bring down the data rate, and has some noise reduction performance.

Keywords DS TT&C signal · Compressive sampling · Sparsity · Delay-Doppler basic dictionary

19.1 Introduction

Spread spectrum communication system have lots of advantages, such as better anti-jamming, concealment, CDMA multiplexing, so that it get widespread concern in the field of aerospace tracking, telemetry and command (TT&C). Direct sequence spread spectrum (DS) TT&C technique has firstly been applied in the Tracking and Data Relay Satellite System (TDRSS). At present, the DS TT&C system has been achieved, the next step of development is to improve the system abilities, such as anti-jamming, anti-interception and anti-multipath interference, which respond to security threats from the increasingly complex environment of the space

---

Y. Cheng (✉) · W. Yang · J. Zhao
Department of Optical and Electrical Equipment, Equipment Academy,
Beijing 101416, China
e-mail: cheng20130810@foxmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_19

information confrontation. Overall, the develop trend of spread spectrum TT&C system has a notable feature, that is the system bandwidth getting broader [1]. In particular the bandwidth of DS TT&C system will exceed 200 MHz [2]. According to the Shannon sampling theorem, in order to acquire signals without distortion the receiver must work at a sampling frequency no less than twice of the signal bandwidth. Along with the development of spread spectrum TT&C systems, the signal bandwidth is much wider. Moreover, the increasing in bandwidth means increasing in the amount of data which is from the Nyquist sampling, so as to TT&C system will be exposed to the massive data problem, and there will be a heavy burden for the data store, transfer and demodulation processing. Therefore, it is necessary to get a novel method which can resolve the massive data problem.

Compressive sensing (CS) theory states that if the signal is sparse in a transform domain, we can project the transformed high-dimensional signal onto a certain lower-dimensional space with a measurement matrix non-correlated with the transform basic dictionary, and it is proved that the random measurements contain all of information for successfully reconstructing the original signal. CS exploits the sparsity of the signal to reduce the measurements, and the original signal can be reconstructed with high probability by solving the convex optimization problem. CS theoretical frame consists of three main areas, such as signal sparsity, non-correlated measurement and non-linear optimization reconstruction, and the signal sparsity is the essential prerequisite and theoretical foundation of the CS theory [3].

In the following sections, on basis of the signal analysis, the sparsity basic dictionary will be built firstly. Then, the DS TT&C signal sparsity is explored by sparse representation. Thirdly, the DS TT&C signals are compressive sensed and reconstructed on the basis of the built basic dictionary. Lastly, the reconstruction performance of CS is deeply analyzed to study the feasibility of the compressive sensing for DS TT&C signals.

19.2 Compressive Sensing Framework

19.2.1 Signal Sparsity

Signal sparsity states that a vector has few non-zero elements, which is measured by the 0-norm, and the specific definitions is given below [4].

Definition 1 (*self-sparsity*) A signal $\boldsymbol{x} \in \mathrm{R}^N$ or C^N is called K-sparse, if $\|\boldsymbol{x}\|_0 \leq K$.

The above definition is very strict, so that there are few strict sparse signals in real life. Most of signals are compressible, which can be approximately represented by sparse signals called as approximately sparse signal. Therefore the pure sparse signal is a special case of compressible signals, generally exists on theory.

Typically, the signals that exhibit self-sparsity are very few, but we can sparsify them by choosing an appropriate transformation domain, which can be represented by sparsity basic dictionary.

Definition 2 (*dictionary-sparsity*) Let $\boldsymbol{\Psi}$ is the basic dictionary of a transformation domain, if a signal $\boldsymbol{x} \in \mathrm{R}^N$ or C^N is called K-sparse in the basic dictionary $\boldsymbol{\Psi}$, there exists $\boldsymbol{x} = \boldsymbol{\Psi s}$, and the vector $\boldsymbol{s}$ is K-sparse, such that $\|\boldsymbol{s}\|_0 \leq K$.

Let $\boldsymbol{\Psi} = \boldsymbol{I}^N$, and $\boldsymbol{I}^N$ is N degree identity matrix, we can find that the self-sparsity is a special case of dictionary-sparsity. And the strict dictionary-sparse signal also exists on theory. In real life, most signals are the compressible signal in some transformation domain, which fade by the power function law.

If the signal x is dictionary-sparse, there exists the vector s, let $\boldsymbol{x} = \boldsymbol{\Psi s}$. The absolute value of coefficient vector elements are arranged from largest to smallest, as $|s_{(1)}| \geq |s_{(2)}| \geq \cdots \geq |s_{(N)}|$, and the power function relation is as follows

$$|s_{(k)}| = \alpha \times j^{-\beta} \tag{19.1}$$

where α and β are constants. Let K main coefficients to approximate to s

$$\|\hat{\boldsymbol{s}}_K - \boldsymbol{s}\|_2 \leq \alpha' \times K^{-\beta'} \tag{19.2}$$

where α' and β' are constants that are determined by α and β. Note that the compressible signal can been approximated by the sparse signal, and the approximation error can been controlled. In following analysis, we don't distinguish the compressible signal from sparse signal, all been called sparse signal.

19.2.2 Compressive Sensing Theory

To simply subsequent analysis, the discretization of the signal is given firstly. The vector x is a discrete model of the signal $x(t)$, as follow

$$x[n] = x(nT_s), \quad T_s = 1/f_s, \quad f_s > f_N \tag{19.3}$$

where f_s and T_s represent the discretization frequency and time interval, f_N is the Nyquist frequency.

To enable CS, the first condition which must be fulfilled is that the sampled signal is sparse. The signal $x(t)$ is sparse if its discrete model x can be approximated in a given domain $\boldsymbol{\Psi}$, as follow

$$\boldsymbol{x} = \boldsymbol{\Psi s}, \quad \|s\|_0 < K \tag{19.4}$$

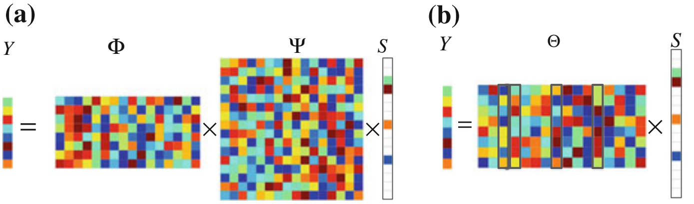


Fig. 19.1 Compressive sensing schematic diagram: **a** Measurement; **b** Reconstruction

where $\boldsymbol{s}$ is a K-sparse vector, and the number of the non-zero elements in the vector is no more than K.

In the standard CS framework, as shown in Fig. 19.1, we acquire the signal x via the linear measurements [5],

$$\boldsymbol{y} = \boldsymbol{\Phi x} = \boldsymbol{\Phi \Psi s} = \boldsymbol{\Theta s} \tag{19.5}$$

where $\boldsymbol{\Phi}$ is an $M \times N$ measurement matrix representing the sampling system, $\Theta = \Phi\Psi \in C^{M\times N}$ is called CS matrix for the reconstruction, $\boldsymbol{y}$ is an $M \times 1$ vector of measurements, and N is the signal length.

Classical sampling theory states that, in order to ensure there is no loss of information, the number of samples M should be as large as the signal length N. On the other hand, CS theory allows for $M \ll N$ as long as the signal is sparse or compressible in some domain [6]. When the CS matrix $\boldsymbol{\Theta}$ satisfies the restricted isometry property (RIP), the signal x can be accurately reconstructed from the measurement vector $\boldsymbol{y}$ by solving 0-norm optimization problem

$$\min \|\boldsymbol{s}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \boldsymbol{\Theta s} \tag{19.6}$$

where $\|\cdot\|_p$ denotes *p-norm.* The RIP is the important foundation of the reconstruction, and we say a matrix $\boldsymbol{\Theta}$ satisfies the RIP of order K, only if there exists a constant $\delta_K \in (0,1)$, such that

$$(1-\delta_K)\|\boldsymbol{s}\|_2^2 \le \boldsymbol{\Theta}\|\boldsymbol{s}\|_2^2 \le (1-\delta_K)\|\boldsymbol{s}\|_2^2 \tag{19.7}$$

where δ_K is monotone increasing, and $\delta_{K1} \ge \delta_{K2}$ for any pair K_1, K_2 with $K_1 \ge K_2$. If the length of measurement vector satisfies

$$M = cK\log(N/K) \ll N \tag{19.8}$$

We can believe that the matrix $\boldsymbol{\Theta}$ probably meets the RIP condition. In practice, there generally is noise in the signal. Furthermore, the RIP can also ensure the compressible signal is recovered from the noisy measurements by bounding the recovery error

$$\min\|\boldsymbol{s}\|_0 \quad \text{s.t.} \quad \|\boldsymbol{y}-\boldsymbol{\Theta}\boldsymbol{s}\|_2 \le \delta \tag{19.9}$$

where δ is the noise power.

Note that the CS framework consists of three main areas, such as sparsity basic dictionary $\boldsymbol{\Psi}$, non-correlated measurement matrix $\boldsymbol{\Phi}$ and reconstruction algorithms, and the basic dictionary $\boldsymbol{\Psi}$ is the essential prerequisite and theoretical foundation of CS. And in the application research of CS, the sparsity content is embodied in the design of a suitable basic dictionary, and lets the signal become sparse in the transformation domain that is represented by the basic dictionary. Basic dictionary construction based on theoretical analysis can be carried out mainly from two aspects: time-frequency analysis, and the signal mathematical model.

19.3 Basic Dictionary Building by Theoretical Analysis

19.3.1 Signal Model

The paper just considers the PCM-CDMA-BPSK system signal for the BPSK odulated DSSS is the foundation of the other type of modulation DSSS. The signal at the receiver can be modeled as [1]

$$r(t) = AD(t-\tau)P(t-\tau)\exp(j(2\pi(f_c+f_d)t+\theta)) \tag{19.10}$$

where A, τ, f_d, f_c, θ are the amplitude, code phase, Doppler frequency, carrier frequency and carrier phase respectively, let $\theta = 0$ to simply the following analysis. $P(t)$, $D(t)$ and $n(t)$ represent spread spectrum waveform at R_c cps, data waveform at R_b bps $(R_b = 1/T_b \ll R_c = 1/T_c)$.

As shown in (19.10), the DS TT&C signals don't belong to the self-sparse signal. On the next step, we should study of the dictionary-sparsity of the DS TT&C signal.

19.3.2 Time-Frequency Analysis

By time-frequency analysis theory, energy distribution of the signal in time-frequency plane can show the signal energy character [7]. For DS TT&C signals, we can let Wigner-Ville distribution represent energy distribution characteristics of the signal in time-frequency plane, and in order to eliminate cross terms, the paper exploits the smooth pseudo Wigner-Ville distribution, the result is shown in Fig. 19.2.

The figure shows that the energy of the DS TT&C signal is scattered on almost all the time-frequency plane. Therefore, the DS TT&C signals don't get obvious

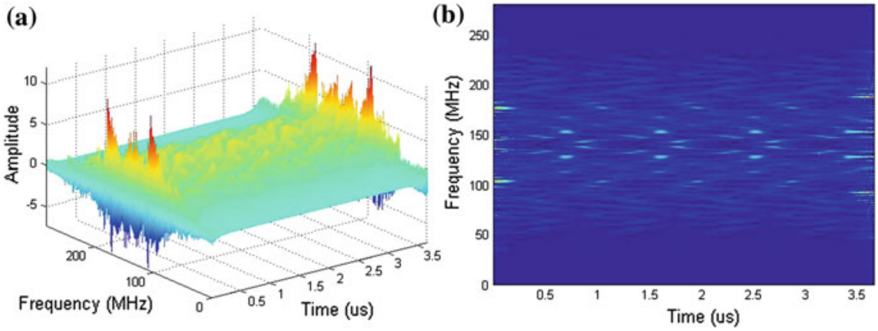


Fig. 19.2 Time-frequency energy distribution

sparsity in the time-frequency plane, such that it isn't suitable to construct the basic dictionary of the signal by time-frequency analysis.

19.3.3 Signal Mathematical-Model Analysis

Spread spectrum waveform of the DS TT&C signal can be expressed as

$$P(t) = \sum_{i=1}^{I-1} P[i]g(t - iT_c) \tag{19.11}$$

where I is spreading gain (namely length of PN code), $P[i]$ is chip of PN code, T_c is the chip period, and $g(t) = \begin{cases} 1 & 0 \le t \le T_c \\ 0 & \text{else} \end{cases}$ is a wideband short pulse. The signal $r(t)$ can be represented as

$$\begin{aligned} r(t) &= \sum_{k=K_{data}} AD(k)P(t - \tau - kT_b)\exp(j(2\pi(f_c + f_d)t)) \\ &= \sum_{k=K_{data}} AD(k)\psi_{k,\tau,f_d}(t) \end{aligned} \tag{19.12}$$

where $D(k)$ is the data bit, K_{data} is the total number of data bits included in the processing length of the signal, T_b is the data period. To simply subsequent analysis, we get Nyquist sampling frequency f_s to discretize (19.12), as follows

$$r(n) = \sum_{k=K_{data}} AD(k)\psi_{k,\tau,f_d}(nT_s) \tag{19.13}$$

where T_s is the sample interval, N is the number of samples in the processing length of the signal, and $n \in [1, N]$. According to the Shannon sampling theorem, the discretize-time signal in (19.13) is equivalent to the continuous-time signal in (19.12). The function ψ_{k,τ,f_d} is determined by the code delay phase τ, Doppler frequency f_d, and corresponding data bit k, and we let that ψ_{k,τ,f_d} denotes the basic dictionary atom. Specifically, by discretizing τ and f_d to expand ψ_{k,τ,f_d}, we can get the delay-Doppler basic dictionary $\boldsymbol{\Psi}_{\mathrm{S}}$ as

$$\boldsymbol{\Psi}_{\mathrm{S}} = \left\{ \psi_{k,\tau^i,f_d^j} | \tau^i = i\Delta\tau, f_d^j = j\Delta f_d, i = [1, N_\tau], j = \left[1, N_{f_d}\right], k = [1, K_{data}] \right\} \tag{19.14}$$

where $\Delta\tau$ and Δf_d are the discretization accuracy of the code delay phase τ and Doppler frequency f_d, respectively. While we assume that phase range and Doppler range are $[0, \tau^{\max}]$ and $[f_d^{\min}, f_d^{\max}]$, we can get the phase number $N_\tau = \tau^{\max} / \Delta\tau$ and the Doppler number $N_{f_d} = (f_d^{\max} - f_d^{\min}) / \Delta f_d$, hence the number of delay-Doppler basic dictionary atoms is $N_\tau \times N_{f_d} \times K_{data}$.

19.4 Numerical Simulations

In this section, we run numerical simulations to demonstrate the DS TT&C signal sparsity and the compressive sensing based on the delay-Doppler basic dictionary. In the simulation, the DS TT&C signal sparse representation and compressive reconstruction are carried out by orthogonal matching pursuit (OMP) algorithm, and the compressive measurement matrix exploits the Gaussian matrix [8]. The data bit per a segment is $K_{data} = 5$, which is at the rate of $R_b = 1$ Mbps. The length of PN code is 127 at the rate of $R_c = 127$ MHz. The intermediate frequency is $f_c = 140$ MHz, while the sampling frequency is $f_s = 3R_c$, hence the number of samples is $N = 1{,}905$. We assume that the discretization accuracy of the code phase and Doppler frequency are $\Delta\tau = T_s$ and $\Delta f_d = 5$Hz, and phase range and Doppler range are $[0, 1/R_b]$ and [−1,000, 1,000] Hz.

Firstly, according to (19.14) to construct the delay-Doppler basic dictionary $\boldsymbol{\Psi}_{\mathrm{S}}$; Secondly, while the SNR and the atom number are 0 dB and 20 respectively, based on $\boldsymbol{\Psi}_{\mathrm{S}}$ to sparsely decompose and represent the DS TT&C signal, which is noiseless and additive white Gaussian noise (AWGN), the results shown in Figs. 19.3 and 19.4; Lastly, in the same conditions as above, and the number of compressive samples is 0.5 N, we compressively sense and reconstruct the DS TT&C signal, the results shown in Figs. 19.5 and 19.6.

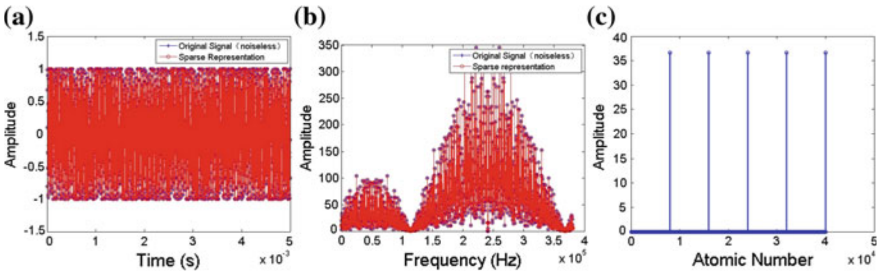


Fig. 19.3 Comparison between sparsere presentation and the original signal (noiseless): **a** in time-domain; **b** in frequency-domain; **c** sparsity coefficient

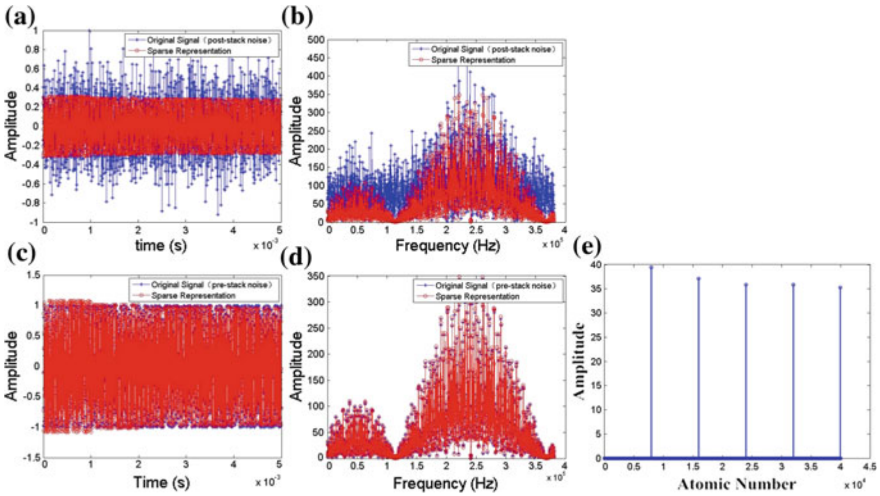


Fig. 19.4 Comparison between sparsere presentation and the original signal (SNR = 0 dB). Comparison with post-stack noise original signal in time-domain (**a**) and frequency-domain (**b**), and with pre-stack noise original signal in time-domain (**c**) and frequency-domain (**d**) respectively **e** Sparsity coefficient

19.4.1 Sparse Representation

In the case of noiseless, Fig. 19.3 shows that the relative error between the sparsere presentation and the original in both time and frequency domain are both not obvious, and the sparsity coefficients are in correspondence with the data bits. In the case of AWGN, it can be seen from Fig. 19.4 while there is a big relative error between the sparsere presentation and the post-stack noise original signal, the relative error between the sparsere presentation and the corresponding pre-stack noise original signal is very small, and the sparsity coefficients are also correspond

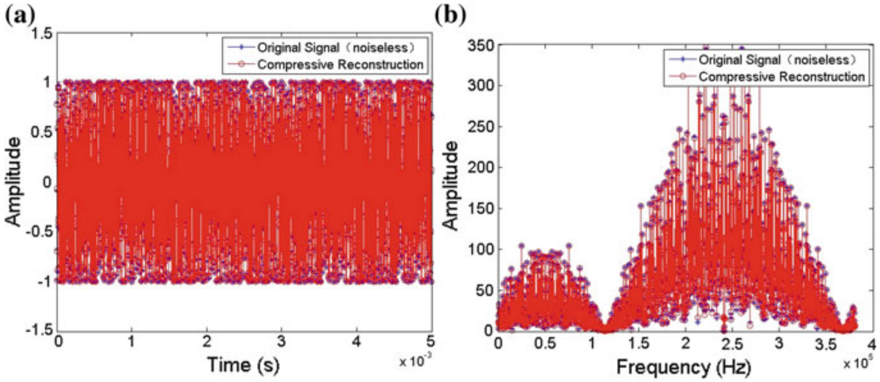


Fig. 19.5 Comparison between the compressive reconstruction and original signal (noiseless). **a** In time-domain. **b** In frequency-domain

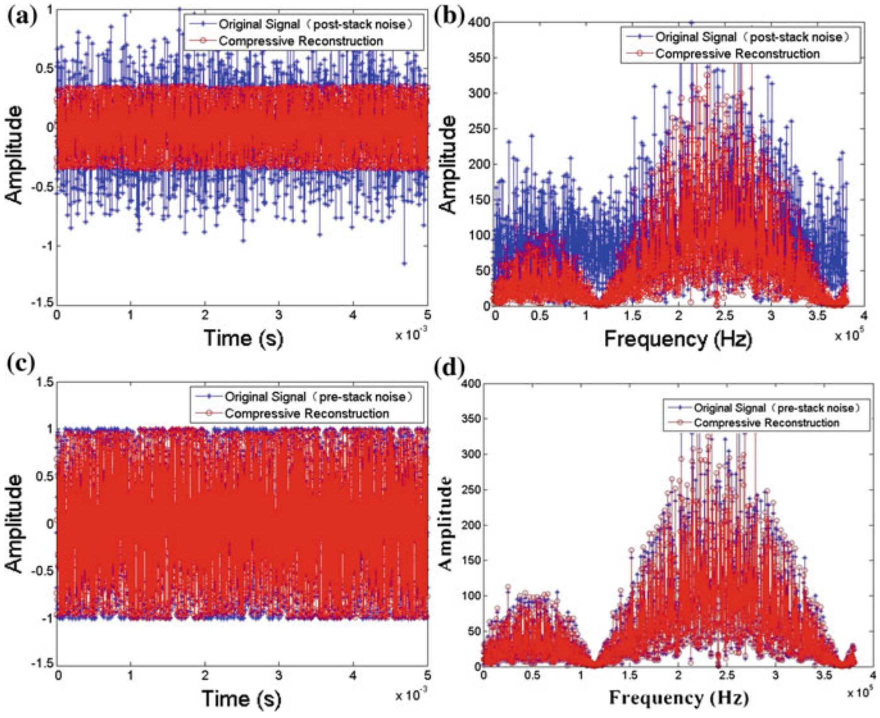


Fig. 19.6 Comparison between the compressive reconstruction and original signal (SNR = 0 dB). comparison with post-stack noise original signal in time-domain (**a**) and frequency-domain (**b**), and with pre-stack noise original signal in time-domain (**c**) and frequency-domain (**d**) respectively

with the data bits. Hence, we can assume that the DS TT&C signal can be very sparse in the delay-Doppler basic dictionary, specifically the sparsity is equal to the data bits, and the dictionary can get some noise reduction performance.

19.4.2 Compressive Sensing and Reconstruction

As shown in Figs. 19.5 and 19.6, in the case of noiseless, the relative error between the reconstruction signal and the original in both time and frequency domain are both very small; In the case of AWGN, the error between the reconstruction signal and the post-stack noise original is big, but the error between reconstruction signal the corresponding pre-stack noise original and is small. Hence, we can conclude that compressive sensing for DS TT&C signals based on delay-Doppler basic dictionary is feasible, which can get some noise reduction performance.

19.5 Conclusions

In this paper, we have analyzed the sparsity of DS TT&C signals by the theoretical analysis. On the basis of the in-depth sparse analysis for the DS TT&C signal, the delay-Doppler basic dictionary has been built, and then the sparse representation and compressive sensing for the DS TT&C signal has been studied by simulations. The results of simulation shows that the DS TT&C signal is sparse in the delay-Doppler basic dictionary, specifically the sparsity is equal to the data bits, and compressive sensing for the signal based on the dictionary is feasible which can effectively bring down the data rate, furthermore it has some noise reduction performance, which has established the theoretical foundation for the CS application to the DS or DS/FH TT&C system.

References

1. Liu JX (2011) Spacecraft TT&C and information transmission technology. National defense industry press, Beijing
2. Huang ZJ, Zhang XL, Chang XM (2012) Design of the structure of a high frequency hopping rate broadband carrier frequency synthesizer. J Spacecraft TT&C Technol 31(1):42–46
3. Jiao LC, Yang SY, Liu F (2011) Development and prospect of compressive sensing. Acta Electronica Sin 38(7):1651–1662
4. Gao L (2011) Study on application of compressive sensing theory in chirp echoes processing for broadband imaging radar. Doctoral thesis of national university of defense technology, Changsha
5. Li SD, Yang J, Ma XY (2013) High resolution ISAR imaging algorithm based on compressive sensing. J Commun 34(9):150–157
6. Candes E, Tao T (2004) Decoding by linear programming. Trans Inf Theory 51(12):4203–4215

7. Mallat SG, Zhang ZF (1993) Matching pursuits with time-frequency dictionaries. Trans Sig Process 41(12):3397–3415
8. Tropp JA, Gilberta C (2007) Signal recovery from random measurements via orthogonal matching pursuit. Trans Inf Theory 53(12):4655–4666

Chapter 20
Research and Implementation of FFT-Based High Dynamic Spread Spectrum Signal with Parallel Acquisition Method by GPU

Bingyin Han, Haixin Zheng, Xianglong Gu and Zibo Li

Abstract In order to acquire DSSS signal based on the PC platform, a method accomplished by GPU has been presented. In this method, one-dimensional search is used to replace the traditional two-dimensional search for PN code phase and Doppler frequency. For this algorithm, we study the influencing factors which are focused on initial phase of the PN code, Doppler frequency, symbol modulation and noise. Due to large number of FFT calculation in the algorithms, parallel processing in the GPU significantly improves the efficiency of operations, as compared with the CPU.

Keywords GPU · High-Performance computers · Parallel acquire · Spread spectrum signal · High dynamic · FFT

20.1 Introduction

Direct sequence spread spectrum (DSSS) technology has been widely used in the fields of communication, control and navigation and positioning, but in high dynamic environment, the received signal with Doppler frequency and large changes in the rate of. Synchronous tracking for High Dynamic DSSS signal, we must first accurately, quickly complete the signal acquisition, the estimated Doppler frequency and code phase [1].

Acquire DSSS signal, essence is the time domain (i.e. PN code phase) of two-dimensional and frequency acquisition process. Parallel fast acquisition is the use of the time domain or the frequency domain parallel method for high dynamic signal, frequency domain parallel is the carrier of parallel, serial PN code search method,

---

B. Han (✉) · H. Zheng · X. Gu · Z. Li
The Academy of Equipment, 3380-172, Beijing 101416, China
e-mail: hanby1002@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_20

time domain parallel or parallel PN code, carrier serial search method. Fast capture all of them can achieve high dynamic spread spectrum signal, but in the spread spectrum signal acquisition of high dynamic and high PN code rate [2]. The hardware resources are limited, the capture time is longer, the signal processing in high performance computer can adopt more effective methods, improve the acquisition speed and accuracy.

With the rapid development of programmable graphics processors (graphic processing unit, GPU), current GPU have parallel computing ability is very strong. Floating-point processing capacity can be achieved even with the generation of more than 10 times that of CPU [3]. At the same time as the NVIDIA Compute Unified Device Architecture Compute Unified Device architecture (CUDA) of the launch, the GPU has good programmability. How to make full use of the parallel fast solving complex operational characteristics of auxiliary CPU calculation of the GPU, has become one of the hot issues of today's. In 2009 Timo Balz and Uwe Stilla in synthetic aperture radar (synthetic aperture radar, SAR) image processing on the effective attempt to use Microsoft GPU center; Asian Studies at the Tsinghua University, the Beijing Jiaotong University, joint design of four phase shift keying (quaternary phase shift keying (QPSK) software demodulation of Sora system in the corresponding position using the GPU to realize the data processing speed, fine result is achieved [4, 5].

This paper combines of parallel hardware features and GPU algorithm for the proposed method, so that the acquisition of DSSS signal accelerated on GPU. And it effectively improves the real-time processing capability of the system.

20.2 Parallel Acquisition Method Based on FFT

In DSSS communication system, signal acquisition essentially means to determine PN code phase and carrier Doppler frequency shift, which are used to produce local recurrence of PN code and carrier signal for demodulation. So it is generally described as a two-dimensional search processing. Main index to judge the performance of DSSS signal acquisition is accuracy of Doppler frequency shift, resolution of code phase and acquisition time.

The acquisition methods of time domain before frequency domain or frequency domain before time domain are usually taken. The fast algorithm contains FFT-based PN code parallel acquisition and FFT-based PN code parallel acquisition. And the acquisition time and accuracy is often affected by the influence of Doppler frequency range.

In this paper, a fast acquisition method based on the FFT parallel, to realize the parallel searching for PN code phase and Doppler frequency, is proposed. Supposed that calculation ability is enough, this method can achieve the purpose of fast computation.

In DSSS communication system, considered with the channel Doppler, baseband signal can be expressed as

$$s(t) = A(t)d[(1+\frac{f_{\mathrm{d}}}{f_t})t]c[(1+\frac{f_{\mathrm{d}}}{f_t}t)t]\cdot \\ \cos[2\pi(f_{\mathrm{IF}}+f_{\mathrm{d}})t]+N(t) \quad (20.1)$$

In the formula (20.1), $A(t)$ for the signal amplitude, $d(t)$ for information symbol modulation waveform, $c(t)$ is the spread spectrum code modulation waveform, f_{IF} is baseband frequency, $N(t)$ baseband signal noise, f_{d} is the instantaneous Doppler frequency.

Without considering the noise and symbol modulation, the ideal signal representation for

$$s(t) = c[(1+\frac{f_{\mathrm{d}}}{f_t}t)t]\cos[2\pi(f_{\mathrm{IF}}+f_{\mathrm{d}})t] \quad (20.2)$$

As can be seen from formula (20.2), PN code modulation directly confuses the frequency of carrier include Doppler, while the carrier also confuses the accumulation of PN code. It is essence that each parallel acquisition method previous of FFT-based PN code and FFT-based PN code is a process of two-dimensional.

The essence of the acquisition process is to estimate the instantaneous state of accurate Doppler frequency and PN code phase. Considering the existence of such a fact that in a certain precision range, the number of initial phase of local PN code is limited, and even it is the same as the length of PN code.

Without considering the influence of information symbol modulation of DSSS signal, Various kinds of PN code with probable initial phase are used to be multiplied with input signal, and then FFT operation is performed on them. So the whole result constitutes a two-dimensional distribution of power with the horizontal axis of frequency and the vertical axis of PN code phase. Through threshold judgment and feature detection with the power of result, we can acquire the Doppler frequency and PN code phase. It is the principle and the sketch map of result in Fig. 20.1.

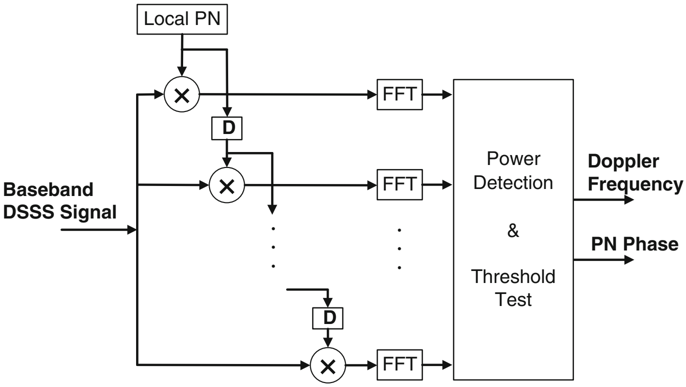


Fig. 20.1 Parallel acquisition method based on FFT

20.3 Performance Analysis and Simulation

In the parallel fast acquisition method, the main analysis method is FFT treatment with the probable recovery carrier. According to the formula (20.1) analysis, the main factors influencing calculation results are PN code phase deviation, instantaneous Doppler frequency, noise and information symbol modulation.

20.3.1 Influence of PN Code Phase Deviation

The condition of simulation is as follows: the signal sampling rate is 56 MHz, the rate of PN code is 10.23 MHz, and the length of PN code is 1023. The simulation result is shown as below. Figure 20.2 shows spectral power, in 3D graph, of PN code phase offset with respect to the baseband signal range of −1.5 to 1.5. Further, the result of spectral distribution with the PN code phase offset at 0.5 chips is shown in Fig. 20.3.

As can be seen from Fig. 20.2, when the PN code phase of simulation signal is the same as local PN code phase, the peak of result is at zero and all of others are zero. At this time, the maximum of the result convoluted with the carrier frequency it is value of parallel acquisition peak.

When deviation of phase between the PN code of simulation signal and local PN code is less than one chip, the correlation spectrum PN code is peak at zero. And the harmonic peak is at the overtones of PN code rate 10.23 MHz; when the phase deviation is 0.5 chips, the harmonic peak at 10.23 MHz comes to be maximal value.

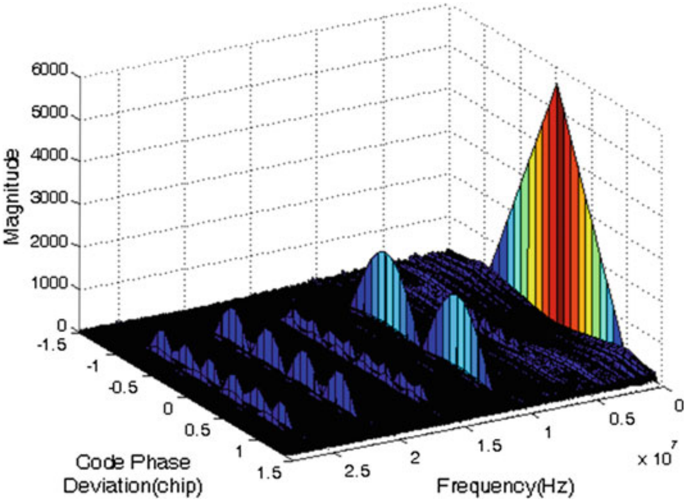


Fig. 20.2 Spectral power with PN code phase offset of −1.5 to 1.5 chips

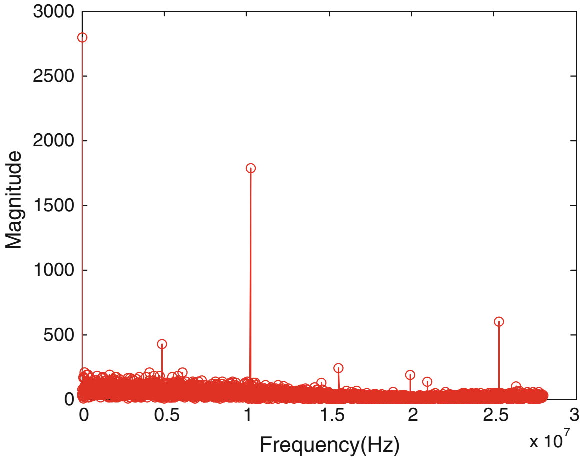


Fig. 20.3 Spectral power with PN code phase offset of 0.5 chips

20.3.2 *Influence of Doppler Frequency*

As to the scheme of parallel acquisition, it is the premise for two-dimensional search being simplified to one-dimensional search that influence on despreading process by Doppler PN code clock is in certain range.

Without considering the influence of information symbol modulation and noise, the signal is expressed as

$$s(t) = A(t)c[(1+\frac{f_{\mathrm{d}}}{f_t})t + \varphi_c(t)]\cos[2\pi(f_0+f_{\mathrm{d}})t] \tag{20.3}$$

The influence of the deviation phase has been discussed in Sect. 20.3.1. And then the influence of Doppler frequency would be studied. Assumed that $\varphi_c(t) = 0$, then the DSSS signal is shown as below:

$$s(t) = A(t)c[(1+\frac{f_{\mathrm{d}}}{f_t})t]\cos[2\pi(f_0+f_{\mathrm{d}})t] \tag{20.4}$$

The despread signal is shown as below:

$$s(t) = A(t)c[(1+\frac{f_{\mathrm{d}}}{f_t})f_c t]c(f_c t)\cos[2\pi(f_0+f_{\mathrm{d}})t] \tag{20.5}$$

Analyzed with the formula of Ref. [5], we can see that the main influencing factor is $c[(1+\frac{f_{\mathrm{d}}}{f_t})f_c t]c(f_c t)$. In order to study the effect of Doppler frequency on parallel acquisition it needs to be considered with both RF carrier frequency and Doppler frequency. The Doppler phenomenon of PN code frequency caused by Doppler frequency is smaller when the RF carrier frequency is higher. The simulation result is shown in Figs. 20.4 and 20.5.

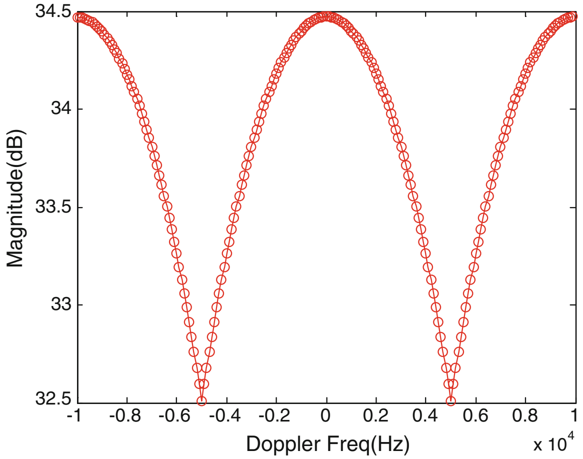


Fig. 20.4 Fluctuation caused by Doppler frequency

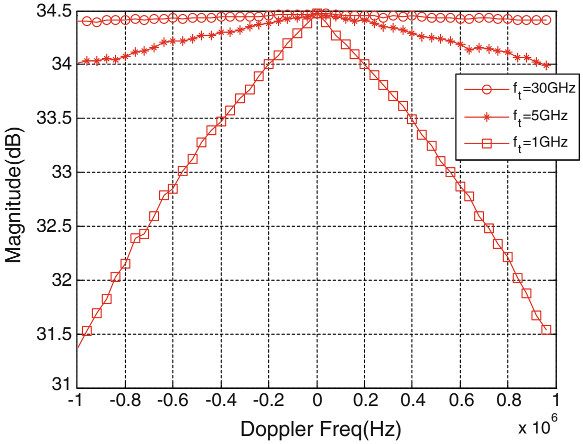


Fig. 20.5 Fluctuation caused by RF carrier frequency

As we can see from Fig. 20.4, with the same RF carrier frequency, fluctuation caused by the small frequency offset is approximated to 2 dB. Figure 20.5 shows that variation of the spectral peak caused by the Doppler frequency under the condition with different RF carrier frequency. It verifies the above conclusions.

20.3.3 Influence of Noise

According to the related theory on signal processing [6], the spectrum peak would effectively increase with the increasing number of data points. In actual environment including noise, when the signal spectrum peak is close to or less than the substrate noise in the received signal, it can be higher with increased FFT points. And then the spectrum peak of signal could be distinguished from noise.

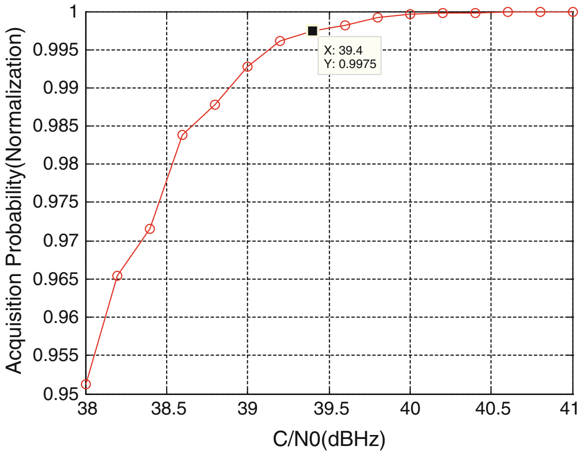


Fig. 20.6 Acquisition probability of DSSS signal with noise

$$\mathrm{X}(k) = \sum_{n=0}^{N-1} x(n) e^{\left(-j\frac{2\pi}{N}nk\right)} \tag{20.6}$$

From formula (20.6) we know that for the signal with single frequency, when the number of calculation points is doubled, the spectrum peak is expected to increase by 3 dB.

DSSS signal is acquired with 262144 (8192 × 32) points under the conditions of different carrier to noise ratios with the 105 times simulation for each, the average of acquisition probability is shown in Fig. 20.6.

The simulation result of Fig. 20.6 shows that when C/N_0 is greater than 39.4 dBHz, parallel acquisition probability is higher than 99.97 %.

20.3.4 Influence of Symbol Modulation

For DSSS signal with information symbol modulation the despreaded signal by PN code is equivalent to the BPSK modulation signal [7]. If the signal data that used to be acquired with is not in a same symbol unit, the spectrum power will scatter. It will lead to the expansion of spectrum, but the main power is concentrated in the center frequency. Supposing that the rate of information symbol is 1 kHz, and then the distinguishing frequency of acquisition is ±1 kHz. Under these conditions the simulation of acquisition probability, the result is shown in Fig. 20.7.

From the simulation result of Fig. 20.7 we can see that as to the rate of information symbol being 1 kHz, when $\mathrm{C/N_0}$ is greater than 43.6 dBHz, parallel acquisition probability is higher than 99.97 %. As can be seen, in these conditions, modulation data information of 1 kHz caused a loss of 4 dB for the acquisition probability.

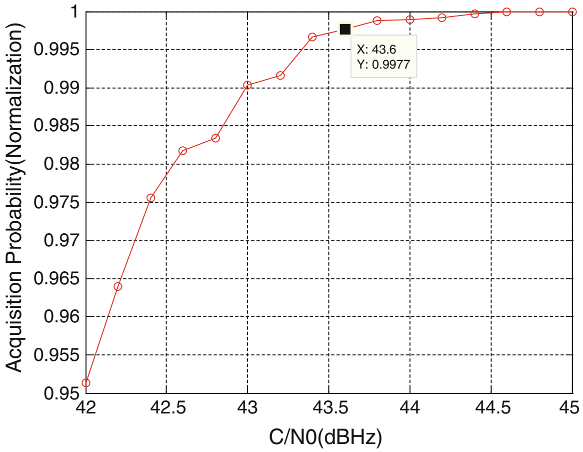


Fig. 20.7 Acquisition probability of DSSS signal with information symbol

20.3.5 Simulation Results

The simulation conditions: $f_s = 56$ MHz, $C/N_0 = 63$ dBHz, $f_{IF} = 14$ MHz, $f_d = 1$ MHz, $\varphi_{c(t)} = 11$. The simulation results as following figures show.

From Fig. 20.8 we can see that the estimations of Doppler frequency and PN code phase are accurate.

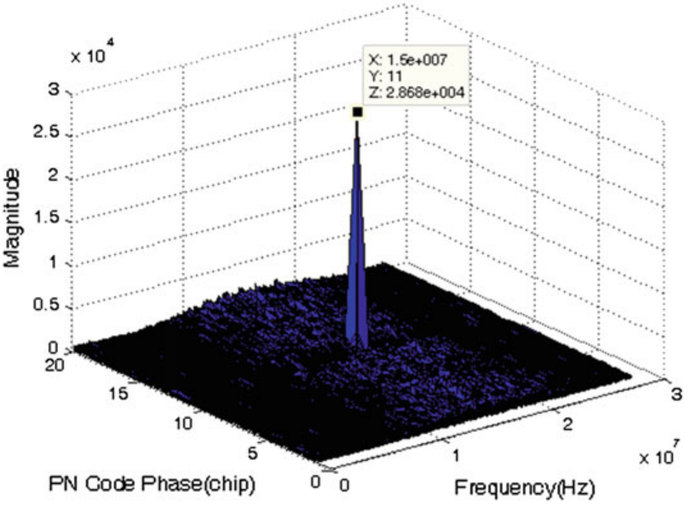
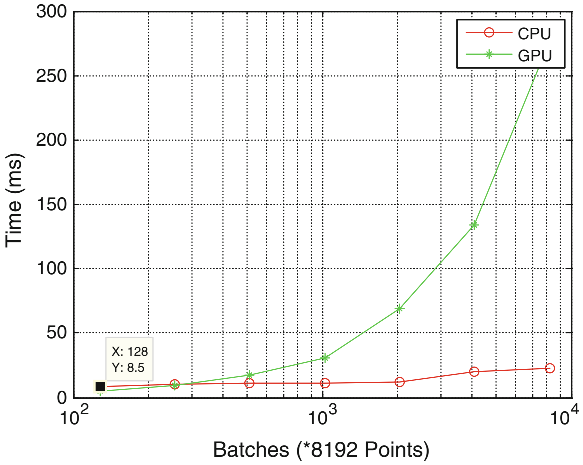


Fig. 20.8 Acquisition results of $C/N_0 = 63$ dBHz

Fig. 20.9 The comparison between CPU and GPU



20.4 Implementation of GPU Parallel Acquisition Method

In the process of this algorithm which introduced in front, there is so much operation of FFT that with the larger number of data the volume of data operation would be increased sharply [8]. As we know that all the operation processes of each branch are similar and simple logic, so the GPU is used to realize the algorithm in this paper. In GPU, hundreds of threads can be launched simultaneously to compute more efficiently, so that DSSS signal can be acquisition fast.

In order to illustrate the full description of the operational efficiency of GPU, this paper is compared between CPU and GPU implementation of the algorithm on the platform of the Z800 high performance computer. The configuration of CPU is as follows: 16 nuclear Intel® Xeon® E5530@2.4 GHz, 12 GB RAM; The configuration of GPU is as follows: NVIDIA Tesla C2070, 448 CUDA core, 6 GB GDDR5 RAM. Test data points for 8192, parallel computing PN code phase number is 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, the experimental result is shown in Fig. 20.9.

20.5 Conclusions

In this paper, it is discussed that the fast parallel acquisition method of DSSS signal based on FFT. In this algorithm, calculation of frequency spectrum would be done on all signals despreaded with probable PN code phase at the same time. It means that a parallel search is made in two-dimensional of Doppler frequency and PN code phase. So this algorithm improves the acquisition efficiency. Especially for the acquisition of high dynamic DSSS signal carried in high frequency, it is more significant to improve the efficiency with this parallel acquisition method.

It is the main disadvantage that this algorithm contains large quantity of calculation and high repetitive operation. But those characteristics of large, simple and high repetitive operation, is applicable to GPU. Implementation of GPU parallel acquisition method based on high performance computer, not only meets the precision requirements of the situation, but also has high flexibility.

References

1. Zhang X (2004) Digital spread spectrum communication baseband signal processing algorithm and VLSI implementation. Science Press, Beijing, pp 136–140
2. Grant PM, Spangenberg SM (1998) Doppler estimation for fast acquisition in spread spectrum communication systems. IEEE J Spread Spectr T Appl
3. Akopian D (2005) Fast FFT based GPS satellite acquisition methods radar. IEEE proceedings of sonar and navigation
4. TOP500 Supercomputer Site (2011) TOP500 Supercomputer Nov 2010 List. http://www.top500.org/lists/2010/11
5. Lee V, Kim C et al (2010) Debunking the 100X GPU versus CPU MYTH: an evaluation of throughput computing on CPU and GPU. ACM SIGARCH computer architecture news
6. Hu GS (2003) Digital signal processing-theory, algorithm and implementation. Tsinghua university press, Beijing, pp 121–129
7. Huang B, Gao J, Li X (2009) An empirically optimized radix sort for GPU. In: IEEE international symposium on parallel and distributed processing with applications 2009, Chengdu, China
8. Woh M, Lin Y, Seo S et al (2011) Analyzing the next generation software defined radio for future architectures. J Sig Process Syst 63(1):83–94

Chapter 21
An Analysis of Minimum Delay Based on Fixed-Hop Routing Strategy

Yi Liu, Bin Wu, Bo Wang and Guangyao Wu

Abstract As for a structure of 24/3/1 Walker-δ constellation with TDMA inter-satellite links (ISL) system, this article establishes the links via permanent ones and divides a time period into 8 time slots which is the basic unit of time delay in this paper. The delays between arbitrary two satellites are measured based on fixed one and two hops routing strategy, the minimum of which are analyzed statistically. The research may make a difference in conducting the routing strategy schematization.

Keywords Inter-satellite links (ISL) · Time slot · Routing strategy · Time delay

21.1 Introduction

As space missions are getting more and more complex and diverse, to accomplish the tasks only by one satellite cannot meet the demands. Multiple satellites' jointly completing assignments has become a mainstream in the development of satellite application and technology [1]. If the geometry structure and space relations between multiple satellites remain stable, they are called satellite constellation. The great advantages of satellite constellation system are irreplaceable. What's more, it has been widely used in communication, reconnaissance, meteorological, navigation and many other fields of space.

The progress of satellite constellation technology requires the ability of auto-navigation and anti-.ISL has been widely researched worldwide and practically applied in GPS and GLONASS [2]. How to transmit the information from a source

---

Y. Liu (✉) · B. Wu · B. Wang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: liuyibittt@163.com

G. Wu
School of Mechanical Engineering and Automation, National University of Defense Technology, Changsha 410073, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_21

satellite to a destination one taking proper routing cost is essential to the constellations with ISL, which is becoming a priority trend. Many satellite routing algorithms have been proposed such as Snapshot-based Routing Algorithm [3] and the improved one [4], Dynamic Virtual Topology Routing Algorithm [5] and Priority-based Adaptive Routing Algorithm [6] and so on. This article proposes a routing method design and assessment, which is mainly based on a walker constellation with TDMA ISL.

21.2 Constellation Properties

The Walker-δ constellation has an orbit height h = 21,000 km, an inclined angle i = 55° and a structure of 24/3/1, where 24 represents 24 satellites, 3 represents 3 orbit planes, and 1 represents the phase factor. It is illustrated in Fig. 21.1. According to the properties above, we can calculate that the eastern orbit is 15° ahead than the western one. In order to make the analysis simple and the statistics cooler, 24 satellites are numbered M_{11}–M_{18}, M_{21}–M_{28}, M_{31}–M_{38}, first subscript of which represents the identifier of orbit plane, while the other subscript represents the identifier of satellites in this plane.

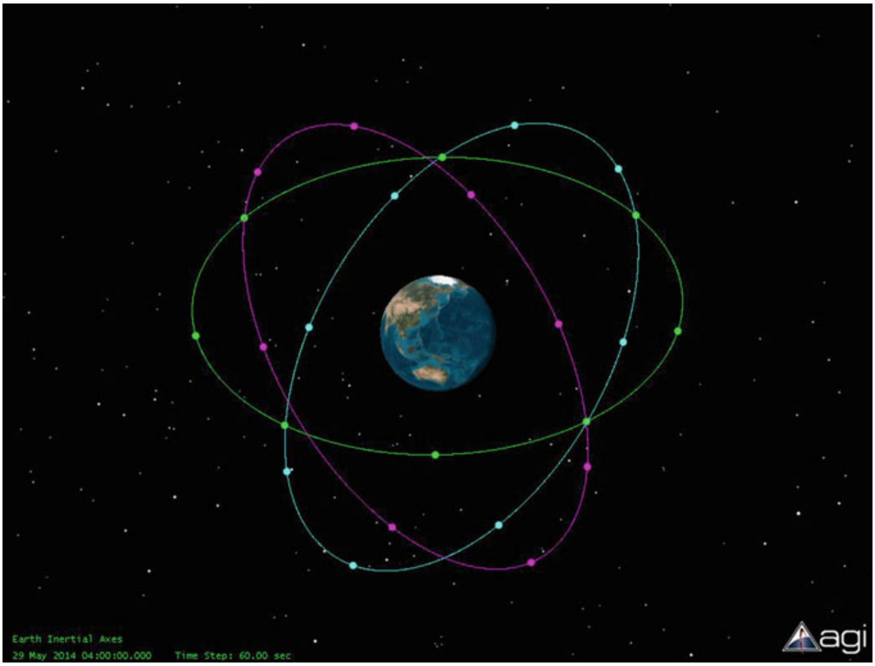


Fig. 21.1 24/3/1 Walker constellation

When on-orbit satellites are capable of communicating and measuring,the constellation is with ISL. Theoretically, a satellite can establish link with the other 23 ones with no constraint conditions. However due to earth block, relative motion and antenna ability, some links are always there while some are interrupted in the whole period. The former links are defined as Permanent ISL, and the latter temporary ISL. Because of time variation and instability of temporary ISL, only permanent ISL is taken into consideration when establishing links. According to visibility analysis, any satellite can make a permanent ISL with another 8 satellites during the period. This paper selects the ISL is built on TDMA system basis, i.e. a whole data frame transmission cycle is divided into 8 time slots of the same (every time slot as a unit of time, a time slot is defined as 3 s). In different time slots, the link between any two satellites is unique. According to the construction of chain planning table, any one of 24 satellites has one and only one satellite to build the link in the each different time slot. For instance, satellite M_{11} can establish link with M_{14}, M_{16}, M_{24}, M_{35}, M_{13}, M_{17}, M_{21}, M_{38} from time slot 1 to 8.

21.3 Routing Strategy

Satellite routing can be classified into Border routing, Up/Down link routing, routing (UDL, Up/Down Link) and space segment routing (ISL, Intel—Satellite Link). Border routing operates between border gateway and ground station.UDL is the link between satellite and ground station or other terrestrial users.ISL is the optimal path to meet certain price between Internal sources of satellite constellation and target satellite. In this paper, we mainly study the routing problem about data packet transmitting from the ground to a satellite constellation, namely, ISL. In order to analyze problems more clearly, a few concepts is defined below.

Definition 1: Source satellite: the satellite giving message or forwarding the ground command data

Definition 2: Target satellite: the final satellite receiving data

Definition 3: Relay satellite:data passes through some satellites between a source satellite and a target satellite. These satellites are called relay satellite

Definition 4: Hop count: The number of relay satellite.

Definition 5: Time delay: The waiting time for data transmitting from the source satellite to a target satellite after a time slot. Its unit is “time slot unit”. The time of the data i transmitting to the satellite directly connected is not included. For example, M_{11} starts from the first time slot and then passes through the relay satellite. It will transmit to the satellite M_{12} after 4 time slot. The time delay is 4 time slot. The time delay of satellite transmitting to itself is defined as 0.

Routing algorithm is the methods or strategies chosen by routing. Due to the limited storage space and processing power on satellite, the long distance between stars, the short time of information should be stored in the satellite, and the need of reducing the transmission error rate, we should not use too much hop counts to transmit data. Based on the above reasons, we select a fixed-one-hop or fixed-two-hop routing strategy and analyze the minimum time delay of data transmitting from an arbitrary satellite to the target satellite after a fixed-one-hop or fixed-two-hop routing, with the minimum delay being the target.

21.4 Simulation

The strategy of fixed-one-hop (fixed-two-hop) means the data package is transmitted from a (two) relay satellite to the target satellite. We have stated and evaluated the minimum time delay of two different strategies via Matlab.

Under the circumstance of fixed-one-hop and fixed-two-hop strategy, we stated the time delay that the data transmission from a satellite to others, and defined the time delay as infinity when cannot reach under the strategy. Meanwhile this article takes minimum time delay as statistical data. The results shows that the minimum time delay is one, two, three, four, seven when fixed one hop, and the probability distribution is shown in Fig. 21.2.; While it is two, three, four, five, six when fixed two hops, and the

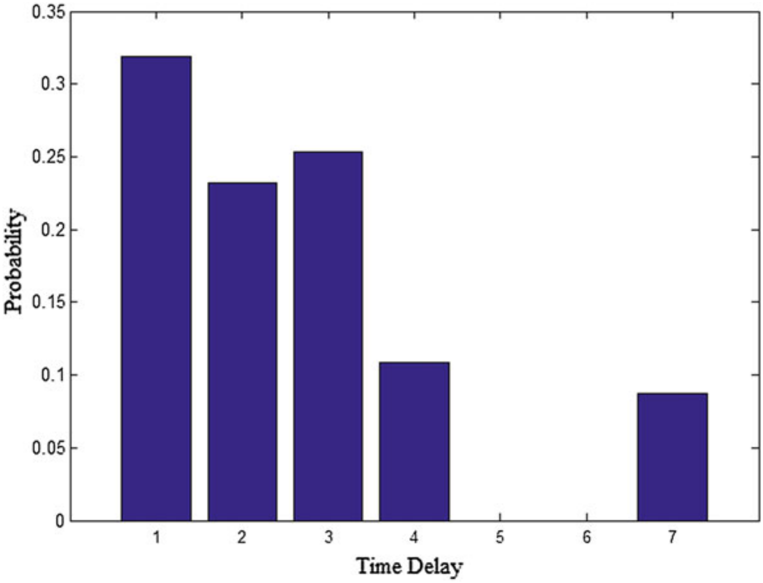


Fig. 21.2 Minimum time delay probability distribution under fixed-one-hop strategy

probability distribution is shown in Fig. 21.3. Also the minimum time delay, routing, start time slot of M_{11} under fixed-one-hop strategy are listed in Table 21.1.

As it is shown in Table 21.1, we can conclude that different start time slot brings different relay satellite (routing); destination variation makes the minimum time delay varies; different start time slot and different routing may make the minimum time delay equal. The conclusion can also applied to fixed-two-hop strategy.

As shown in the Fig. 21.2, when fixed one-hop, the minimum time delay is mainly 1, 2, 3 unit of time slot, which occupies more than eighty percent. The average time delay $E_{\text{delay}} = 2.5870$ unit of time slot, the time delay varies fiercely, the variance $D_{\text{delay}} = 3.1674$. When focus on Fig. 21.3, we may find that under fixed-two-hop strategy, the minimum time delay is mainly 2, 3, 4 unit of time slot, occupying more than eighty percent. The average time delay $E_{\text{delay}} = 3.1449$ unit of time slot, the time delay varies fiercely, the variance $D_{\text{delay}} = 2.0898$. The minimum average time slot under fixed-one-hop is less than the other one, while the delay jitter is obviously greater than the latter one.

The above results tell that information can be transferred through different routes via choosing different fixed-hop strategy according to the start time. At the same time, when the ground station needs to transfer data to target satellite on a certain time slot, the satellite would be selected as the source satellite whose time delay is least according to the above results.

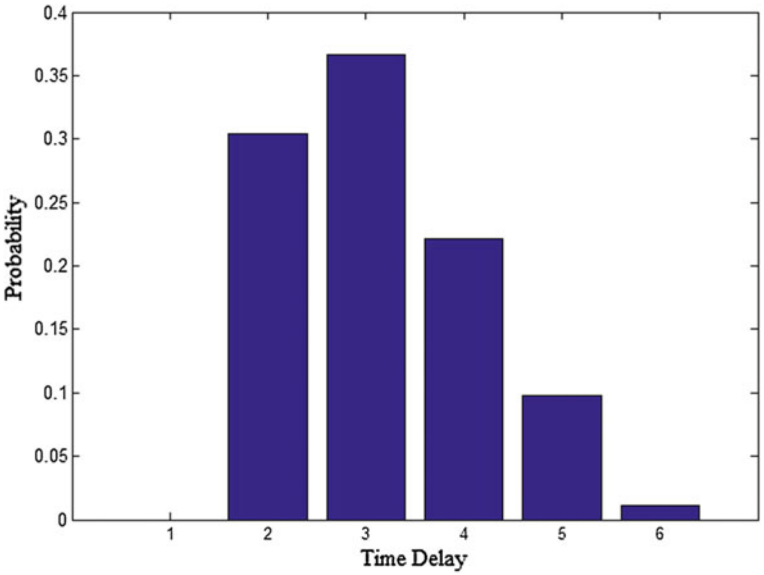


Fig. 21.3 Minimum time delay probability distribution under fixed-two-hop strategy

Table 21.1 Time delay of the source satellite M_{11} under fixed-one-hop strategy

Destination ID	Minimum delay	Routing	Starting time
M_{12}	3	M_{17}	6
M_{13}	7	M_{16}	2
M_{14}	3	M_{38}	8
M_{15}	1	M_{13}	5
M_{16}	3	M_{35}	4
M_{17}	1	M_{14}	1
M_{18}	3	M_{16}	2
M_{21}	1	M_{16}	2
M_{21}	2	M_{24}	3
M_{23}	2	M_{13}	5
M_{24}	2	M_{21}	7
M_{25}	4	$M_{35}(M_{38})$	4(8)
M_{26}	3	$M_{24}(M_{21})$	3(7)
M_{27}	1	M_{17}	6
M_{28}	7	M_{38}	8
M_{31}	1	M_{21}	7
M_{31}	2	M_{16}	2
M_{33}	2	$M_{35}(M_{38})$	4(8)
M_{34}	4	M_{24}	3
M_{35}	1	M_{38}	8
M_{36}	2	M_{17}	6
M_{37}	1	$M_{24}(M_{35})$	3(4)
M_{38}	2	M_{14}	1

21.5 Conclusion

This paper chooses fixed-one-hop and fixed-two-hop strategy to carry on minimum time delay analysis based on permanent ISL through link establishment. According to the simulation, we can select a proper cost routing under these two strategies.

The next research will concentrate on selecting the satellite linked to ground as a source in constellation, more flexible hops routing, and taking the temporary ISL into consideration.

References

1. Zhang YL, Fan L, Zhang Y (2008) The theory and design of satellite constellation. Science Press, Beijing
2. Lin YM, He SB, Zheng JJ (2010) The suggestion to ISL technology development of global navigation satellites. Spacecraft Eng 19(6):1–7
3. Gounder VV, Prakash R, Abu Amara H (1999) Routing in LEO-based satellite networks. In: Proceedings of IEEE emerging technologies symposium: wireless communications and systems, vol 22, pp, 1–22
4. Wang JL, Yan J, Cao ZG (2009) LEO satellite network routing algorithm to optimize the snapshot sequence. J Astronaut 30(5):2003–2007
5. Werner M (1997) A dynamic routing concept for ATM-based satellite personal communication networks. IEEE J Sel Areas Commun 15(8):1636–1648
6. Uzunalioglu H (1998) Probabilistic routing protocol for low earth orbit satellite. In: IEEE international conference on communications, pp 89–93

Chapter 22
Private-Key Error Correcting Cryptography Based on QC-LDPC Codes for Telemetry System

Jinrong Zhang, Ling Wu, Shengli Liu and Junfeng Wang

Abstract Aiming at the security and reliability of data transmission in a wireless telemetry system, a new error correcting cryptography algorithm is proposed in this paper. The novel algorithm employs quasi-cyclic low-density parity-check (QC-LDPC) codes for simultaneous encryption and channel encoding, which is different from the traditional concatenated systems where encryption and channel encoding perform serially. Thus the proposed cryptosystem exhibits attracting characteristics when it is applied to a telemetry system. The parameters are obtained according to the cryptanalysis and the corresponding key size is estimated from the compact description of the quasi-cyclic matrices. The analysis shows that the proposed error correcting cryptosystem with lower-complexity is more secure than the traditional concatenated system at a price of larger key size which is still acceptable for implementation.

Keywords Error correcting cryptography · Telemetry · QC-LDPC code · Private-key · Security · Key size

22.1 Introduction

Security and reliability are two crucial aspects of a wireless telemetry system. In the conventional way, they are achieved respectively by two techniques. On the one hand, in order to avoid eavesdroppers getting the orbital target's capabilities and parameters from telemetry signals, encryption techniques are generally employed to guarantee the secure transmission of telemetry data. On the other hand, at the receiver side some data may be demodulated in error or even be lost since telemetry

---

J. Zhang (✉) · L. Wu · S. Liu · J. Wang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: jinrong@mail.ustc.edu.cn

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_22

signals are likely corrupted during transmission, which will lead to errors propagation after decryption or even lead to decryption failure because of synchronization loss of encrypted data block, therefore, error correcting codes are adopted to provide reliability of telemetry signal transmission.

In this paper, an innovative technique of an error correcting cryptosystem based on linear codes is developed to combine security and reliability. This is an attractive idea since it can reduce the implementation complexity of the system and the delay of telemetry data.

The cryptography schemes based on algebraic linear codes have being developed since several decades. The first public-key cryptosystem was proposed by McEliece in 1978 [1]. The private-key cryptosystem was designed by Rao in 1984 [2]. These original cryptosystems were both based on Goppa codes and exhibited two major drawbacks: large key size and low transmission rate (about 0.5), which restricted the applications to practical systems. In order to overcome these drawbacks, other linear codes were adopted to substitute Goppa codes. Monico suggested using low density parity-check (LDPC) codes in place of Goppa codes [3]. The cryptosystem based on LDPC codes was more feasible than the original one: firstly, the key size was smaller since an LDPC code could be described by a very sparse matrix; secondly, LDPC codes with higher rate could be designed with no difficulty. Marco Baldi [4] presented a cryptosystem based on quasi-cyclic (QC) LDPC code. It reduced the key size in a further step since a QC-LDPC code could be completely represented by a row of the parity check matrix due to its quasi-cyclic structure. Otherwise, punctured LDPC codes [5], nonhomogeneous LDPC Codes [6] and extended difference family QC-LDPC codes [7] were also introduced to the cryptosystem.

In consideration of key size, implementation complexity, security level and error correcting performance together, this paper focuses on a private-key cryptosystem based on QC-LDPC codes. The parity check matrix of the QC-LDPC code, a scrambling matrix, a permutation matrix and an error vector are treated as the private key and provided only for legal users for encryption and decryption. The private-key cryptosystem based on QC-LDPC codes has smaller key size and lower complexity compared with the one based on Goppa code or LDPC codes, while it does not degrade error correction capability compared with a public-key cryptosystem based on QC-LDPC codes. Otherwise, it possesses higher security level and lower complexity than the traditional concatenated system. Therefore, the private-key cryptosystem based on QC-LDPC codes is very promising for a wireless telemetry system.

The rest of this paper is organized as follows. In Sect. 22.2, the framework of a telemetry system based on the new error correcting cryptography algorithm is described and the merit is discussed. In Sect. 22.3, the algorithm of key design, encryption and decryption is presented. Then, the security is discussed in Sect. 22.4 and the key size is estimated in Sect. 22.5. Finally, Sect. 22.6 concludes this paper.

22.2 Framework of Error Correcting Cryptosystem

In the existed wireless telemetry system, encryption and error correcting code are both employed to provide security and reliability for telemetry data transmission. As illustrated in Fig. 22.1, at the transmitter side, the telemetry data are encrypted first, the encrypted data are then encoded by a channel encoder and the encoded symbols are modulated to produce signals for transmission; at the receiver side, after telemetry signals are demodulated, the demodulated data are decoded first by a channel decoder and then decrypted to recover the original telemetry data.

A new error correcting cryptosystem is designed in this paper to combine security and reliability. As illustrated in Fig. 22.2, the telemetry data are encrypted and encoded simultaneously by one module at the transmitter side and the demodulated data are error correcting decoded and decrypted simultaneously by one module at the receiver side.

Compared with the traditional system as shown in Fig. 22.1, the proposed error correcting cryptosystem presented in Fig. 22.2 exhibits four merits as follows. First, since encryption and encoding are performed by one module while decryption and decoding are also accomplished by one module, the interface between encryption and encoder and the interface between decryption and decoder are both needless in the new cryptosystem whereas they have to be elaborately designed in the traditional system. Thereby, the implementation complexity of the new cryptosystem is lower and the system architecture is simplified. Second, there needs only one synchronization module for decryption and decoding in the new cryptosystem while both code block synchronization and cipher block synchronization are necessary in the

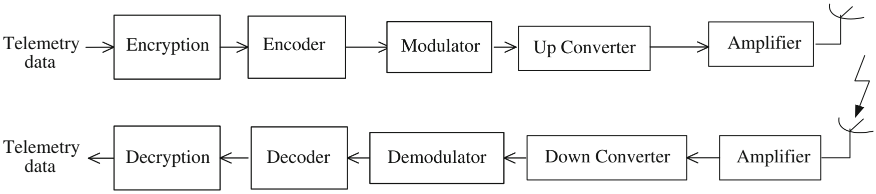


Fig. 22.1 The framework of the existed telemetry system

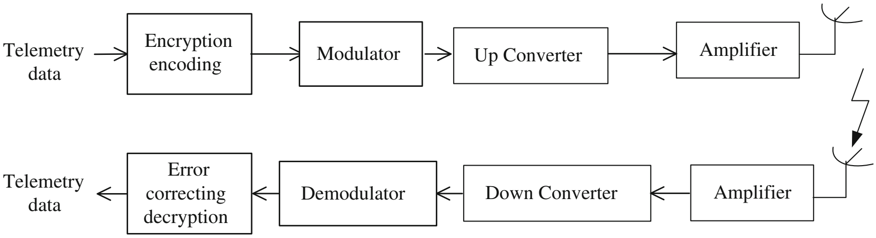


Fig. 22.2 The framework of the error correcting cryptosystem

traditional system. This will reduce data loss caused by synchronization loss. Third, the data delay of the new cryptosystem is smaller since encryption and encoding are performed simultaneously at the transmitter side and decryption and decoding are accomplished simultaneously at the transmitter side. Finally, the security level of the new error correcting cryptosystem is higher than the traditional one.

Generally, there are public-key cryptosystems and private-key cryptosystems. In a public-key cryptosystem, the public key which is used for encryption can be published while the private key used for decryption is only allocated for authorized receivers. This is beneficial for key administration. However, there is a trade off between security and reliability in the public-key error correcting cryptosystem and the parameters have to be elaborately chosen [4]. In a private-key cryptosystem, the key for encryption have to be strictly kept secret, which increases the difficulty in key administration. However, there is no error correcting performance loss in the private-key error correcting cryptosystem and the key size is smaller. Therefore, a private-key error correcting cryptosystem is more appropriate for the wireless telemetry system where error correcting performance is the prior factor and there are mature strategies of key administration.

The error correcting cryptosystems based on linear codes have a common characteristic that the parity check matrix of the error correcting code is treated as the private key which is only supplied for authorized receivers for decryption. It is simple for authorized receivers who know the private key to implement decryption by performing a fast decoding algorithm of error correcting codes while it is impossible for eavesdroppers who do not know the private key to recover the plaintext from the cipher text in polynomial time. The private-key cryptosystem is also called symmetric-key cryptosystem since the key for encryption and the key for decryption are just the same.

22.3 Algorithm of Error Correcting Cryptography

The algorithm of a private-key error correcting cryptography is illustrated in Fig. 22.3, where the insecure noisy channel includes the modulator, up converter, amplifier, antenna at the transmitter side and the antenna, amplifier, down converter, demodulator at the receiver side.

22.3.1 Key

The secret key is composed of the set $\{\mathbf{H}, \mathbf{S}, \mathbf{P}, \mathbf{e}\}$ which is constructed as follows.

1. Design a QC-LDPC code based on difference families [8]. The code length n and information block length k of the QC-LDPC code are both multiple of a positive integer p, i.e. $n = n_0 \times p$ and $k = k_0 \times p$. Denote $r_0 = n_0 - k_0$.

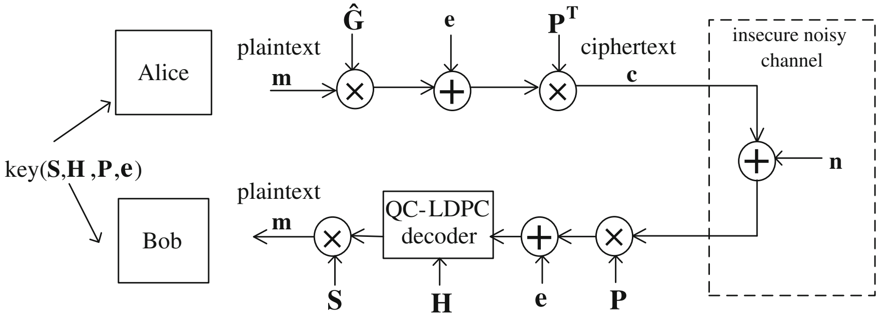


Fig. 22.3 Algorithm of private-key error correcting cryptography

The QC-LDPC code can be described by a parity check matrix **H** that is formed by $r_0 \times n_0$ array of circulant submatrices, as is presented by formula (22.1).

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}^c_{0,0} & \mathbf{H}^c_{0,1} & \cdots & \mathbf{H}^c_{0,n_0-1} \\ \mathbf{H}^c_{1,0} & \mathbf{H}^c_{1,1} & \cdots & \mathbf{H}^c_{1,n_0-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}^c_{r_0-1,0} & \mathbf{H}^c_{r_0-1,1} & \cdots & \mathbf{H}^c_{r_0-1,n_0-1} \end{bmatrix} \tag{22.1}$$

Where $\mathbf{H}^c_{i,j}$ is a $p \times p$ circulant matrix with the following form:

$$\mathbf{H}^c_{i,j} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{p-1} \\ a_{p-1} & a_0 & a_1 & \cdots & a_{p-2} \\ a_{p-2} & a_{p-1} & a_0 & \cdots & a_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix} \tag{22.2}$$

2. Choose a $k \times k$ nonsingular scrambling matrix **S** that is formed by $k_0 \times k_0$ array of circulant submatrices, as is presented by formula (22.3).

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{0,0} & \mathbf{S}_{0,1} & \cdots & \mathbf{S}_{0,k_0-1} \\ \mathbf{S}_{1,0} & \mathbf{S}_{1,1} & \cdots & \mathbf{S}_{1,k_0-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{k_0-1,0} & \mathbf{S}_{k_0-1,1} & \cdots & \mathbf{S}_{k_0-1,k_0-1} \end{bmatrix} \tag{22.3}$$

Where $\mathbf{S}_{\mathrm{ij}}$ is a $p \times p$ circulant matrix with the form described by formula (22.2).

3. Choose a $n \times n$ random permutation matrix $\mathbf{P}$ with block-diagonal form, as is presented by formula (22.4).

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{P}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{P}_{n_0-1} \end{bmatrix} \tag{22.4}$$

Where $\mathbf{P}_\mathrm{i}$ is a $p \times p$ circulant matrix.

4. Construct a predetermined set of error vectors with length n. The Hamming weight of error vectors (represented by t) equals approximately to $n/2$.

22.3.2 Encryption

Alice who wants to send messages to Bob divides the messages into k-bit vectors $\mathbf{m}$. She obtains the encryption matrix $\hat{\mathbf{G}} = \mathbf{S}^{-1}\mathbf{G}$**,** where $\mathbf{G}$ is the generator matrix of the QC-LDPC code.

The ciphertext $\mathbf{c}$ of the plaintext $\mathbf{m}$ is produced in the following way:

1. Send the plaintext $\mathbf{m}$ to an encoder defined by the generator matrix $\hat{\mathbf{G}}$ to produce codeword $\mathbf{c}' = \mathbf{m}\hat{\mathbf{G}}$.
2. Select an error vector $\mathbf{e}$ randomly from the predetermined set of error vectors.
3. Compute the ciphertext $\mathbf{c} = (\mathbf{c}' + \mathbf{e})\mathbf{P}^{\mathbf{T}}$.

22.3.3 Decryption

Bob obtains the ciphertext $\mathbf{r}$ by demodulating the received signals. Denote $\mathbf{r} = \mathbf{c} + \mathbf{e}_{\mathrm{ch}}$, where $\mathbf{e}_{\mathrm{ch}}$ is an error vector with t_{ch} errors that are introduced by the channel noise. The plaintext $\mathbf{m}$ is recovered by decryption operations as follows:

1. Calculate $\mathbf{r}' = \mathbf{rP} = (\mathbf{c} + \mathbf{e}_{\mathrm{ch}})\mathbf{P} = \mathbf{c}' + \mathbf{e} + \mathbf{e}_{\mathrm{ch}}\mathbf{P} = \mathbf{mS}^{-1}\mathbf{G} + \mathbf{e} + \mathbf{e}_{\mathrm{ch}}\mathbf{P}$.
2. Compute $\mathbf{r}'' = \mathbf{r}' + \mathbf{e} = \mathbf{mS}^{-1}\mathbf{G} + \mathbf{e}_{\mathrm{ch}}\mathbf{P}$ to eliminate the intentional errors imposed by the transmitter.
3. Send the codeword $\mathbf{r}''$ to the QC-LDPC decoder to obtain $\mathbf{m}' = \mathbf{mS}^{-1}$ by belief propagation decoding.
4. Recover the plaintext $\mathbf{m} = \mathbf{m}'\mathbf{S}$.

It can be seen from the decryption process that errors introduced by the channel noise turns to be $\mathbf{e}_{\mathrm{ch}}\mathbf{P}$ after the first step. Since the Hamming weight of the permutation matrix **P** is 1, the vector $\mathbf{e}_{\mathrm{ch}}\mathbf{P}$ has the same Hamming weight with $\mathbf{e}_{\mathrm{ch}}$, i.e., the number of errors does not vary. All errors can be corrected with high probability by the third step as long as t_{ch} is smaller than the correction capability of the QC-LDPC code. Obviously, there is no reliability loss in the private-key error correcting cryptosystem.

22.4 Security

Many potential attacks can be considered for the cryptanalysis of the cryptosystem based on QC-LDPC codes. In this paper, we limit to discuss four types of attacks.

22.4.1 Brute Force Attack

Brute force attacks could be tempted by enumerating all possible secret key {**H**,**S**,**P**,**e**}. If there are a huge number of possible values of the key, it is impossible to find the exact key in acceptable time by enumerating, that is, brute force attacks will fail.

22.4.1.1 Attack Against H

The cardinality of a set of equivalent codes based on random difference families with fixed parameters can be evaluated through probabilistic arguments and is approximately [9]:

$$C(n_0, d_v, p) = \binom{p}{d_v}^{n0} \prod_{l=0}^{n0-1} \prod_{j=1}^{dv-1} \frac{p-j\left[2-p(\mathrm{mod}2)+(j-1)^2/2+ld_v(d_v-1)\right]}{p-j} \tag{22.5}$$

where d_v is the column weight of **H**.

As an example, by assuming $p = 500$, $n_0 = 5$, $d_v = 7$, we have $C(5, 7, 500) \approx 2^{263}$, which is high enough to discourage brute force attack against **H**.

22.4.1.2 Attack Against S

Given that the row weight of the submatrix $\mathbf{S}_{\mathrm{ij}}$ is μ_{S} for all i and j, the number of all possible **S** equals to

$$N_{\mathrm{S}} = \prod_{i=0}^{k_0 \times k_0 - 1} \left(\binom{p}{\mu_{\mathrm{S}}} - i \right) \tag{22.6}$$

when $p = 500$, $k_0 = 4$, $\mu_{\mathbf{S}} = 2$, we have $N_s = 2^{270}$ which is large enough to resist against brute force attack. When **S** is a denser matrix, a bigger value of $\mu_{\mathbf{S}}$ will result in larger N_s so that the brute force attack against **S** is more infeasible.

22.4.1.3 Attack Against P

The number of possible perturbation matrices **P** with row weight $m = 1$ equals to

$$N_{\mathbf{P}} = (p!)^{n_0} \tag{22.7}$$

When $p = 500$, $n_0 = 5$, we have $N_{\mathbf{P}} = (500!)^5 \gg 2^{500}$ which is obviously too huge to break **S**.

22.4.1.4 Attack Against E

The number of possible error vectors **e** with Hamming weight $t \approx n/2$ equals to

$$N_{\mathbf{e}} = \binom{n}{t} \tag{22.8}$$

When $p = 500$, $n_0 = 5$, $n = 2{,}500$, we have $N_e \gg 2^{500}$ which is large enough to resist against brute force attack.

22.4.2 Majority Voting Attack

The goal of majority voting attack is to obtain the encryption matrix $\mathbf{G}' = \mathbf{S}^{-1}\mathbf{G}$ in an efficient way [7]. The encryption matrix $\mathbf{G}'$ can be recovered as follows:

1. Choose an arbitrary k-bit vector **x**, and produce $N_e = 2^{n-k}$ distinct ciphertext of **x**, denoted by $\boldsymbol{c}_i = \mathbf{x}\mathbf{G}' + \boldsymbol{e}'_i$. Let $C = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_2^{n-k}\}$ be a set of ciphertext, $\mathbf{M}(\mathbf{x}\mathbf{G}')$ be a $2^{n-k} \times n$ matrix where the n-bit vector $\mathbf{x}\mathbf{G}'$ iterates 2^{n-k} times in its rows, $\mathbf{M}(\mathbf{e})$ be a matrix whose ith row is $\boldsymbol{e}'_i$.
2. Perform majority voting on each column of the matrix $\mathbf{M}(C) = \mathbf{M}(\mathbf{x}\mathbf{G}') + \mathbf{M}(\mathbf{e})$ to obtain an estimate $\mathbf{x}\mathbf{G}'$ of $\mathbf{x}\mathbf{S}^{-1}\mathbf{G}$.
3. Repeat the above steps for k linearly independent vectors **x**. Let $\mathbf{X}_{k\times k}$ be a matrix whose rows are these k vectors, $(\mathbf{X}\mathbf{G}')_{k\times n}$ be a matrix whose rows are the corresponding k estimates. Then we obtain an estimate of the encryption matrix by calculating $\mathbf{G}' = \mathbf{X}^{-1}(\mathbf{X}\mathbf{G}')$.

These steps require $k \times 2^{n-k}$ majority votes over n columns of $\mathbf{M}(C)$. Therefore, the work factor of this attack is $O(k \times n \times 2^{n-k})$. When $p = 500$, $n_0 = 5$, $r = 4/5$, the work factor is 2^{522} which is infeasible.

22.4.3 Rao-Nam Attack

Rao-Nam attack is a chosen plaintext attack. It attempts to recover the encryption matrix $\mathbf{G}'$ from ciphertext as follows: choose two plaintext vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ that are differing only in the ith position, and get the difference of the corresponding ciphertext $\mathbf{c_1} - \mathbf{c_2} = (\mathbf{m_1} - \mathbf{m_2})\mathbf{G}' + (\mathbf{e_1} - \mathbf{e_2}) = \mathbf{g_i}' + (\mathbf{e_1} - \mathbf{e_2})$, where $\mathbf{g_i}'$ is the ith row vector of $\mathbf{G}'$. The weight of $\mathbf{e_1} - \mathbf{e_2}$ is not larger than $2t$. If t is much smaller than n, we get $\mathbf{g_i}'$ by majority voting on the vector $\mathbf{c_1} - \mathbf{c_2}$ and obtain $\mathbf{G}'$ by repeating this process for all $i = 1,2, \ldots, k$.

As can be seen from the process of Rao-Nam attack, the precondition of obtaining $\mathbf{G}'$ from ciphertext is that t/n is a small value. In the proposed cryptosytem we set $t \approx n/2$, thus, Rao-Nam attack is infeasible.

22.4.4 Struik-Tilburg Attack

Struik-Tilburg attack is a chosen plaintext attack proposed by Struik and Tilburg [10]. It works as follows:

1. The plaintext $\mathbf{m}$ is encrypted $O(N_e \log N_e)$ times to produce different ciphertext $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \ldots, \mathbf{c}^{(Ne)}$, where, $N_e = 2^{n-k}$ is the number of different error vectors.
2. Repeat step 1) using $\mathbf{m}_i = \mathbf{m} + \mathbf{u}_i$ in place of $\mathbf{m}$ and obtain ciphertext $\mathbf{c}_i^{(1)}, \mathbf{c}_i^{(2)}, \ldots, \mathbf{c}_i^{(N)}$.
3. Construct the encryption matrix $\mathbf{G}'$ by regarding the vector $\mathbf{e}_i = \mathbf{c}_i - \mathbf{c}$ as the ith row.

The work factor of Struik-Tilburg attack is $O(N_e \log N_e)$. When $p = 500$, $n_0 = 5$, $r = 4/5$, the work factor is far greater than 2^{500} which is large enough to resist against Struik-Tilburg attack.

22.4.5 Parameters Choice

Nowadays, complexity of order 2^{80} is considered as the lower bound of security. According to the above cryptanalysis, we choose a set of parameters $\{p = 500, n_0 = 5, k_0 = 4, d_v = 7, t \approx n/2\}$ to provide high security. Whichever eavesdroppers employ brute force attack, majority voting attack, Rao-Nam attack or Struik-Tilburg attack, the work factor is larger than 2^{500}. As is known that the work factor of a cryptosystem based on advanced encryption standard (AES) is not larger than 2^{256} [11]. Obviously, the proposed private-key error correcting cryptosystem is far more secure than the traditional cryptosystem based on AES.

22.5 Key Size

Large key size is one of the major drawbacks of the cryptosystem based on linear codes. The proposed error correcting cryptosystem based on QC-LDPC code has a smaller key size than the original cryptosystem since the matrices $\{\mathbf{H}, \mathbf{S}, \mathbf{P}\}$ have a quasi-cyclic structure.

$\mathbf{H}$ is formed by $(n_0 - k_0) \times n_0$ array of submatrices $\mathbf{H_{ij}}$ that is a $p \times p$ sparse circulant matrix. $\mathbf{H_{ij}}$ can be completely described by its 1th row which only has d_v nonzero values that can be represented by $\log_2 p$ bits. Thus, the required memory for $\mathbf{H}$ is $M_{\mathbf{H}} = (n_0 - k_0) \times n_0 \times d_v \times \log_2 p$.

$\mathbf{S}$ is formed by $k_0 \times k_0$ array of circulant submatrices which can be represented by one row. Since $\mathbf{S}$ is a dense matrix, the required memory is $M_{\mathbf{S}} = k_0 \times k_0 \times p$.

$\mathbf{P}$ is a permutation matrix with block-diagonal form and can be defined by n_0 circulant matrices. Each circulant matrix can be represented by the nonzero positions of its1^{th} row. Thus, the required memory for $\mathbf{P}$ is $M_{\mathbf{P}} = n_0 \times \log_2 p$.

$\mathbf{e}$ is a n-bit error vector which can be generated by a linear feedback shift register (LFSR). When a LFSR with 10 states is used, the required memory for $\mathbf{e}$ is $M_{\mathbf{e}} = 10$.

When a set of parameters $\{p = 500, n_0 = 5, k_0 = 4, d_v = 7\}$ is chosen, the key size of the proposed cryptosystem is $M_{\mathbf{H}} + M_{\mathbf{S}} + M_{\mathbf{p}} + M_{\mathbf{e}} = (n_0 - k_0) \times n_0 \times d_v \times \log_2 p + k_0 \times k_0 \times p + n_0 \times \log_2 p + 10 = 8{,}370$ bits = 1,047 bytes. Although the key size is larger than that of the traditional cryptosystem based on AES, it is much smaller than that of a cryptosystem based on Goppa codes or random LDPC codes [2, 4]. It is not a problem to store and administrate the key of 1047 bytes for the current hardware storage techniques.

22.6 Conclusion

This paper presents a private-key error correcting cryptosystem based on QC-LDPC codes which combine encryption and error correcting code to provide security and reliability simultaneously. When applied to a wireless telemetry system, the new error correcting cryptosystem exhibits better performance including lower complexity, lower delay and higher security compared with the traditional concatenated system. The proposed cryptosystem can resist against four typical attacks including brute force attack, majority voting attack, Rao-Nam attack and Struik-Tilburg attack as long as the parameters are appropriately chosen. Compared with the cryptosystem based on AES, the private-key error correcting cryptosystem based on QC-LDPC codes can provide significantly higher level of security whereas needs a larger memory to store its key. It is worth noticing that an AES cryptosystem is unable to provide reliability and it has to be concatenated with an error correcting code. Compared with the concatenated system, the proposed cryptosystem has a simplified structure and can be implemented with lower complexity.

References

1. McEliece RJ (1978) A public-key cryptosystem based on algebraic coding theory. DSN progress report 42–44:114–116
2. Rao TRN (1984) Crypotsystems using algebraic codes. In: International conference on computer systems and signal processing, Bangalore, India
3. Monico C, Rosenthal J, Shokrollahi A (2000) Using low density parity check codes in the McEliece cryptosystem. In: Proceeding of IEEE ISIT2000, Sorrento, Italy, p 215
4. Baldi M, Chiaraluce F, Garello R (2006) On the usage of quasi-cyclic low-density parity-check codes in the McEliece cryptosystem. In: Proceedings of first international conference on communication and electronics (ICCE'06), Hanoi, Vietnam, pp 305–310
5. Adshar AAA, Sharifi A, Eghlidos R (2009) On the security and efficiency of secure channel coding schemes based on punctured codes. In: SAM-STACNS '09 conference
6. Lay KT, Hou CH, Peng LC (2011) Nonhomogeneous LDPC codes and their application to encrypted communication. In: 2011 third international conference on communications and mobile computing, pp 353–357
7. Hooshmand R, Eghlidos T, Reza Aref M (2012) Improving the Rao-Nam secret key cryptosystem using regular EDF-QC-LDPC codes. ISC Int J Inf Secur 4(1):3–14
8. Baldi M, Chiaraluce F (2005) New quasi cyclic low density parity check codes based on difference families. In: Proceedings of international symposium communication theory and application, ISCTA 05, Ambleside, UK, pp 244–249
9. Baldi M, Chiaraluce F, Garello R et al (2007) Quasi-cyclic low-density parity-check codes in the McEliece cryptosystem. In: Proceedings of IEEE international conference on communications (ICC2007), Glasgow, Scotland, pp 951–956
10. Struik R, Tilburg J (1988) The Rao-Nam scheme is insecure against a chosen-plaintext attack. In: Advances in cryptology-CRYPTO'87, lecture notes in computer science, Springer, New York, pp 445–457
11. FIPS-197 (2001) Advanced encryption standard (AES). http://csrc.nist.gov/encryption/aes

Chapter 23
Error Analysis of Angle Measurement for Ground Based Photoelectric Tracking and Measuring Devices

Sheng Cai, Yanfeng Qiao and Yi Yu

Abstract Ground based photoelectric tracking and measuring devices are important in range tests, and angle measurement accuracy is one of the key technical indicators. Therefore, appropriate modeling and analyzing of the angle measurement error of the photoelectric tracking and measuring devices are important for the overall design of such equipments. Traditionally, the angle measurement errors are divided into two categories: angle measurement error of the tracker and detector error. They are usually calculated respectively and then combined using root mean square theory. However, this analysis method does not take into account the interaction between different error terms, and assumes that all the error terms are independent. A theoretical (true value) angle measurement equation and a valuation equation are established, based on the coordinate transformation theory and working process of the ground based photoelectric tracking and measuring devices, in this paper. Then the mean squares of angle measurement errors are obtained using Monte Carlo statistical analysis method. The results indicate that the simulation results and that of traditional methods are almost the same.

Keywords Ground based photoelectric tracking and measuring devices · Error analysis of angle measurement error · Coordinate transformation · Monte Carlo method

23.1 Introduction

Ground based photoelectric tracking and measuring devices, yielding several advantages such as high measurement accuracy, able to obtain the image of flying target, intuitive, and the ability to avoid effect of blackout and noise jamming from

---

S. Cai (✉) · Y. Qiao · Y. Yu
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130033, China
e-mail: cais_ciomp@hotmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_23

the ground, are important equipments in modern range. With the development of technology, the performance of observing targets and new weapons is improved, which put forward higher requirements of photoelectric tracking and measuring devices, especially in maneuverability and accuracy. Angle measurement accuracy is an important manifestation of the measurement accuracy [1].

Traditionally, the angle measurement errors are divided into two categories: angle measurement error of the tracker and detector error. They are usually calculated respectively and then combined using root mean square theory [2–4]. However, this analysis method does not take into account the interaction between different error terms, and assumes that all the error terms are independent. A theoretical (true value) angle measurement equation and a valuation equation are established, based on the coordinate transformation theory and working process of the ground based photoelectric tracking and measuring devices, in this paper. Then the mean squares of angle measurement errors are obtained using Monte Carlo statistical analysis method. This analysis method can provide a theoretical basis for the foundation of the overall design, especially error distribution, of photoelectric tracking and measuring devices.

23.2 Traditional Error Analysis Method for Angle Measurement

The error sources, in traditional error analysis of angle measurement, are divided into two components which are shown in Tables 23.1 and 23.2.

According to the data in Tables 23.1 and 23.2, the composite angle measurement error of the system can be calculated by

$$\begin{aligned}\text{angle measurement error} = {} & (\text{image processing error}^2 \\ & + \text{angle measurement error of the seeker}^2)^{1/2}\end{aligned}$$

When the azimuth of the seeker is $E = 65°$, the measured azimuth and pitch angle are respectively $6.32''$ and $5.74''$.

23.3 Angle Measurement Error Analysis Based on Monte Carlo Method

The basic idea of angle measurement error analysis using Monte Carlo method is described as follows. The target star is regarded as the measurement reference. A theoretical (true value) angle measurement equation under ideal condition and a valuation equation under actual working conditions (considering various error factors) are established, based on the coordinate transformation theory and working

Table 23.1 Angle measurement error of the seeker (the pitch angel $E = 65°$ in units of ″)

Error sources		Error		Azimuth angle error			Pitch angle error		
		Systematic	Stochastic	Transformational relation	Systematic	Stochastic	Transformational relation	Systematic	Transformational relation
Vertical axis error		2.0	0.25	$1/\sqrt{2}\tan E$	3.03	0.38	$1/\sqrt{2}$	1.41	0.18
Horizontal axis error		3.0	0.3	tan$E$	6.34	0.64	0	0	0
Zero position error		2.0	0.3	0	0	0	1	2.0	0.3
Aiming error		2.0	0.3	sec$E$	4.73	0.71	0	0	0
Encoder error		1.5	0.5	1	1.5	0.5	1	1.5	0.5
Angle measurement error of the seeker	Static error	5.05	0.83		8.67	1.14		2.87	0.61
	Dynamic increment (60 % static)				5.2	0.69		1.72	0.37
	Dynamic error				10.12	1.33		3.35	0.71
	Dynamic horizontal projection error			10.2			3.42		
	Dynamic point error			4.32			3.42		

Table 23.2 Sensor error

Items	Visible light television	Remarks
Pixel number	1,920 × 1,080	
Pixel center distance (μm)	10 μm	
Focal length (mm)	1,000 mm	
Pixel resolution $\delta('')$	2.1	
Pixel resolution error $\sigma_F('')$	1.19	Pixel resolution error: $\sigma_F = \delta/\sqrt{3}$
Signal processing error $\sigma_C('')$	1.19	$\sigma_C = N_c \cdot \sigma_F$
		$(N_c = 1)$
Static error $\sigma_J('')$	1.68	$\sigma_J = (\sigma_F^2 + \sigma_C^2)^{1/2}$
Error caused by moving of the target $\sigma_Y('')$	1.68	$\sigma_Y = N_Y \cdot \sigma_F$
		$(N_y = 1)$
Shaking error of optical axis $\sigma_H('')$	4.13	Correction remnant/focal length
Dynamic error $\sigma_D('')$	4.61	$\sigma_D = (\sigma_J^2 + \sigma_Y^2 + \sigma_H^2)^{1/2}$

process of the ground based photoelectric tracking and measuring devices. Then the statistical property (RMS) of the angle measurement error is obtained using Monte Carlo statistical analysis method [5–8].

23.3.1 Definition of the Coordinate System

Six coordinate systems, such as earth horizontal coordinate system, are introduced to meet the requirement of error analysis. The coordinate systems (right-handed coordinate systems) are defined as follows.

1. Earth horizontal coordinate system $E_g(g_1, g_2, g_3)$

The origin is the station of the ground based photoelectric tracking and measuring device, axis g_1 points to the north, and axis g_3 points to the zenith.

2. Earth geocentric coordinate system $E_c(c_1, c_2, c_3)$

The origin is the geocenter, axis c_1 is the rotation axis of the earth pointing to the North Pole, and axis c_3 is in the equatorial plane and it is also in the same longitude with the origin of the earth horizontal coordinate system.

3. Base coordinate system $B(b_1, b_2, b_3)$

The base coordinate system namely B is fixed on the base of the ground based photoelectric tracking and measuring device, it represents spatial position of the base, and the origin is located at the central symmetry point in the level surface of the base. When the base coordinate system is completely leveled, the base coordinate system coincides with the earth horizontal coordinate system. Axis b_1 and b_2 are in the earth horizontal plane, b_1 points to the North, and b_2 points to the zenith. The leveling error between the two coordinate systems can be decomposed into the rotation angle φ_{BC} and θ_{BC}.

4. Photoelectric tracking and measuring device alidade coordinate system (vertical axis coordinate system) $V(v_1, v_2, v_3)$

The vertical axis coordinate system originally coincides with the earth horizontal coordinate system. The bottom of the alidade coordinate system is fixed together with the rotation component of vertical axis, and the top end is fixed with the stationary of horizontal axis. The alidade rotates with the horizontal axis and the telescope axis around the vertical axis, which is defined as azimuth angle (A) rotation.

5. Photoelectric tracking and measuring device horizontal coordinate system $H(h_1, h_2, h_3)$

Horizontal coordinate system, which is fixed on the rotation component of the horizontal shafting, represents the space position of the seeker's horizontal axis. The origin is the intersection of horizontal axis and telescope optical axis. Mutually orthogonal axis h_1 and h_2 are in the horizontal plane, and h_1 points to the North, h_3 points to the zenith. The horizontal coordinate system, with the telescope axis, rotates around the horizontal axis, namely pitch angle (E) rotation.

6. Telescope coordinate system $T(t_1, t_2, t_3)$

The telescope coordinate system represents the space position of the telescope, and its origin is the intersection of telescope optical axis and the seeker's horizontal axis. Axis t_1 is the telescope's optical axis, which points to the North. Axis t_3 points to the zenith.

7. Barycentric coordinate system of the nth target star $O_{cn}(o_{cn1}, o_{cn2}, o_{cn3})$

The origin is the target star, axis o_{cn1} points to the North, and axis o_{cn3} points to the zenith crossing the geocenter.

23.3.2 The Error Terms

The error terms are generated in the process of coordinate transformation from target barycentric coordinate system to the geographic horizontal coordinate system. Due to the design, processing, manufacturing, and other factors which introduce deviations from the theoretical definition, the error of angle measurement occurs. The error terms are shown in Table 23.3.

Table 23.3 Statistical table of the error terms

No.	Description of error terms	Error distribution
1	$\Delta\varphi_{b1}$—base coordinate system, the level deviation rotating around axis g_1 in the earth horizontal coordinate system	Uniform distribution
2	$\Delta\vartheta_{b2}$—base coordinate system, the level deviation rotating around axis g_1 in the earth horizontal coordinate system	Uniform distribution
3	$\Delta\varphi_{v1}$—the shaking error of alidade coordinate system rotating axis b_1 in the base coordinate system	Uniform distribution
4	$\Delta\vartheta_{v2}$—the shaking error of alidade coordinate system rotating axis b_2 in the base coordinate system	Uniform distribution
5	ΔA—reading error of vertical axis optical-electricity encoder	Normal distribution
6	i—orthogonal error of horizontal and vertical axis	Uniform distribution
7	$\Delta\varphi_{h1}$—the shaking error of horizontal coordinate system rotating axis h_1 in the alidade coordinate system	Uniform distribution
8	$\Delta\psi_{h3}$—the shaking error of horizontal coordinate system rotating axis h_3 in the alidade coordinate system	Uniform distribution
9	ΔE—reading error of horizontal axis optical-electricity encoder	Normal distribution
10	C—orthogonal error of telescope optical axis and vertical axis	Uniform distribution
11	$\Delta\vartheta_{t2}$—shaking error of pitch of telescope optical axis	Uniform distribution
15	$\Delta\psi_{t3}$—shaking error of sway of telescope optical axis	Uniform distribution
16	$\Delta f'$—measurement error of focal length	Uniform distribution
17	Δx—position measurement error of the star image point in x direction	Normal distribution
18	Δy—position measurement error of the star image point in y direction	Normal distribution

23.3.3 *Establish the Measurement Equation*

23.3.3.1 Theoretical (True Value) Measurement Equation

Position of the nth target star in its barycentric coordinate system can be expressed as

$$O_{cn0} = [\,0 \quad 0 \quad 0 \quad 1\,]^{\mathrm{T}} \tag{23.1}$$

The coordinate transformation from barycentric coordinate system of the the nth target star to the earth horizontal coordinate system is shown in Fig. 23.1.

According to the coordinate transformation process expressed in Fig. 23.1, the position of the nth target star is

$$E_{gn} = [\,E_{gn1} \quad E_{gn2} \quad E_{gn3} \quad 1\,]^{\mathrm{T}} = M_5 M_4 M_3 M_2 M_1 O_{cn0} \tag{23.2}$$

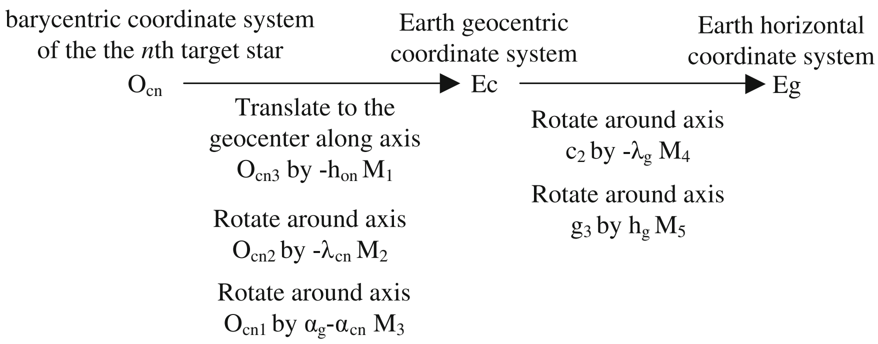


Fig. 23.1 Coordinate transformation process of the theoretical measurement equation

Hence the azimuth angle A_n and the pith angel E_n of the *n*th target star in the earth horizontal coordinate system can be calculated by

$$E_n = \arcsin\left(E_{gn3}/\left(E_{gn1}^2 + E_{gn2}^2 + E_{gn3}^2\right)^{1/2}\right) \tag{23.3}$$

$$A_n = \arccos\left[E_{gn1}/\left(E_{gn2}^2 + E_{gn3}^2\right)^{1/2}\right] \tag{23.4}$$

23.3.3.2 Valuation Measurement Equation

The valuation measurement equation is established by the following processes. The target star is firstly tracked using the telescope. Then the telescope coordinate system is transformed to the earth horizontal coordinate system through transformation of horizontal coordinate, vertical coordinate and the base coordinate. The overall coordinate transformation process is shown as Fig. 23.2.

According to the coordinate transformation process expressed in Fig. 23.2, the position of the *n*th target star in the earth horizontal coordinate system is

$$E_{gn} = \begin{bmatrix} E_{gn1} & E_{gn2} & E_{gn3} & 1 \end{bmatrix}^{\mathrm{T}} = \prod_{i=11}^{i=25} M_i O_{cn0} \tag{23.5}$$

The valuation of the azimuth angle $\hat{A}_n$ and the pith angle $\hat{E}_n$ can be calculated by Eqs. (23.2) and (23.5), thus the angle measurement errors can be represented as

$$\Delta A_n = \hat{A}_n - A_n \tag{23.6}$$

$$\Delta E_n = \hat{E}_n - E_n \tag{23.7}$$

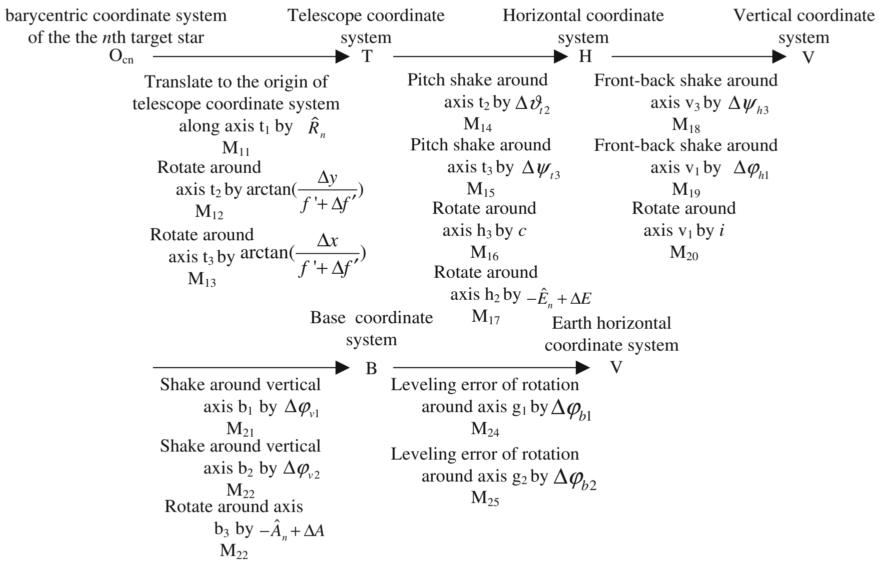


Fig. 23.2 Coordinate transformation process of the valuation measurement equation

23.3.4 Analysis of Angle Measurement Error Using Monte Carlo Method

According to the established in Sect. 23.3.3 and the distribution of error terms described in Sect. 23.3.2, the valuation errors of the azimuth angle and the pith angle can be calculated by a MATLAB program based on Mote Carlo method. The statistical histogram is shown in Figs. 23.3 and 23.4, the RMS are respectively 5.95″ and 2.70″.

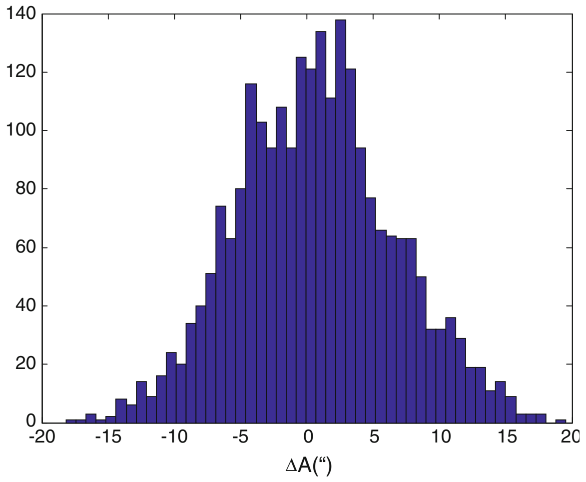
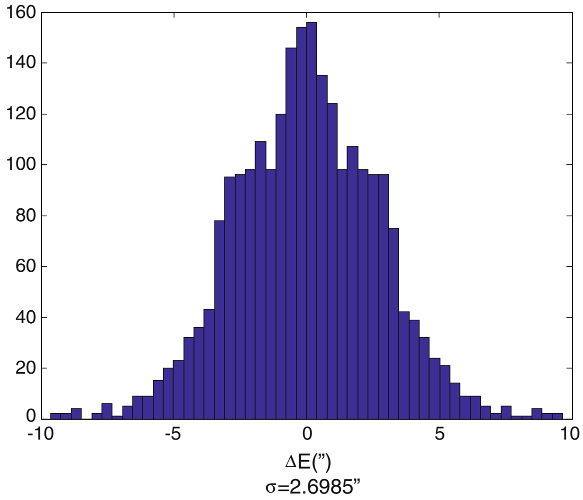


Fig. 23.3 Statistical histogram of azimuth valuation error

ΔA(“)
σ=5.9464”

Fig. 23.4 Statistical histogram of pitch valuation error



23.4 Conclusions

Angle measurement error analysis, as an important foundation for the overall design of the photoelectric tracking and measuring devices, is of great significance for systematic error allocation. In this paper, the angle measurement error analysis results of the Monte Carlo method and that of the traditional one are almost the same. Therefore, the analysis method is can be used as a reference for analyzing the precision of angle measurement using the ground based photoelectric tracking and measuring devices. Besides, establishing a measurement equation using coordinate transformation methods is a strict modeling process for the error analysis of complex photoelectric measuring devices, which indicates that the error analysis results calculated using the Monte Carlo method is reliable. In general, the proposed angle measurement error analyzing method is valid for planning the indexes of a system in the argumentation process. Further studies should focus on considering whether the error distributions of the error terms in this paper are reasonable and how to determine the value of error, which provides basis for more precise error analysis models. Besides, high precision error test methods, especially the ones in field trials, are required in order to provide the actual measurement foundations for error analysis of angle measurement.

References

1. Wang JQ (2003) Integrated design of optical instruments. Master thesis of Changchun institute of optics fine mechanics and physics, Changchun
2. Du JF, Li ZZ (2002) Analysis of the axial accuracy of the GD-220 photoelectric theodolite. Optics Precis Eng 10(4):416–419

3. Li H, Shen XH (2008) New shafting error simulating method of photoelectric theodolite. Infrared Laser Eng 37(2):334–337
4. Du JF, Zhang MW, Zhang XM (2012) Angle measurement accuracy of photoelectric theodolite. J Appl Optics 33(3):466–473
5. Wang JQ, Jin G, Yan CX (2005) Orientation error analysis of airborne opto-electric tracking and measuring device. Optics Precis Eng 13(2):105–116
6. Wang F, Jia T, Zhang CL (2009) Dynamic correction of target deviations for photoelectric theodolites by coordinate transform. Optics Precis Eng 17(12):2939–2944
7. Wang T, Tang J, Song LW (2010) Correction of the measuring error of vehicular photoelectric theodolite. Infrared and laser engineering 41(5):1335–1338
8. Li WC (2010) Analysis of angle measurement accuracy based on vehicle photoelectric theodolites. Master thesis of Changchun institute of optics fine mechanics and physics, Changchun

Chapter 24
Autocollimator for Small Angle Measurement over Long Distance

Jun Luo and Wei Zhao

Abstract In order to send the load to the right track, it is necessary to aim the initial emission angle before rocket launching. We designed an autocollimator which can accurately measure small angle over long distance to meet the requirements of rocket launching. By adding a new optical imaging system to the common one, the focal length and resolution both largely increase. Cubic polynomial fitting was used to calculate the deflection angle which reduces the influence of laser beam drifts and other undesirable changes of autocollimator. Experiment results demonstrated that the system has an accuracy of ±2″ over a measurement distance of 150 m and remain such precision during an hour. The autocollimator achieves high accuracy, stability and anti-interference of unwanted lights.

Keywords Auto-collimation · Small-angle measurement · Optical system combination · Cubic polynomial fitting

24.1 Introduction

Optical levers have been used for contactless measurement of angular deflection since 1826 [1]. Optical apparatuses for precision small angle measurement have wide applications in many different fields such as calibration of machine tools,

---

J. Luo (✉)
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science, Changchun 130033, China
e-mail: ciomp_cekoong@sina.com

J. Luo
University of Chinese Academy of Sciences, Beijing 100049, China

W. Zhao
PLA Military Representative Office, No. 206 Institute of the 2nd Academy of CASIC, Beijing 100854, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_24

alignment of precision instruments during assembly, measurement of large scale structures deformations, etc. [2, 3]. Several optical methods based on interference [4, 5], auto-collimation [6–9], and internal reflection effect [10] have been proposed for measurement of small angles. Optical apparatus based on these methods are compact and could achieve high accuracy, but few could conduct outdoors or measure deflection angle over a long distance. Nonetheless, such cases are inevitable in the initial attitude determination before missile launches or rocket launches.

Laser beam drifts due to laser generation mechanism and environmental interferences. Therefore, an autocollimator using laser datum for small angle measurement can hardly achieve high accuracy and robustness over a long measurement distance [11, 12]. The light generated by laser is quasi-parallel, the beam width increases with the increasing of distance. Thus we can hardly do anything to advance the laser generation mechanism. However, varies methods can be applied to lessen the influence of beam transmission medium. The pulse width modulation technology and a Fourier phase shift method for sub-pixel estimation are adopted to autocollimator used outdoors, and achieve $\pm 1''$ measurement accuracy at 0.5 m measuring distance [13]. A common-path compensation principle was proposed to improve the measuring beam spot's angular drift [14]. Signal processing both in hardware and software are also used to reduce environmental interferences. However, these experimental results were still unsuitable for measuring deflection angle over a measuring distance of 150 m.

We proposed a simple and novel method based on the widely used autocollimator. Our system is easily set up by adding a new optical imaging system to the common one; the new system contains two convex lenses and a right angle prism. The combination of two optical systems largely increases the focal length and attains higher resolution. Moreover, a neutral filter is placed in front of CCD (charge coupled device) to eliminate the interference of unwanted lights. In order to reduce the influence of the laser beam drift and other undesirable changes of autocollimator, cubic polynomial fitting was used to calculate the deflection angle. Experimental results have shown that our system can achieve an accuracy of $\pm 2''$ over a measurement distance of 150 m and remain such precision during an hour.

24.2 General Principle

The principle of auto-collimation is shown in Fig. 24.1. When the parallel light is projected onto a mirror plane which is square to the beam, the beam will be reflected back along the same path (see Fig. 24.1a). When the mirror plane has small deflection angle α with the optical axis, the rays reflected from the mirror are rotated by twice that angle 2α relative to the optical axis, and the image in the plane of the CCD array changes its position accordingly (see Fig. 24.1b) [9]. The relationship between the deviation distance d and the deflection angle α is given by

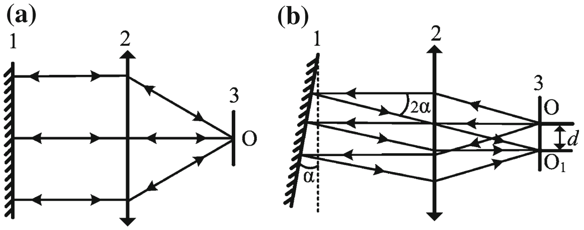


Fig. 24.1 The principle of small angle measurement based on auto-collimation. *1* Mirror; *2* objective lens; *3* focal plane. **a** the mirror plane is square to the beam; **b** the mirror plane has small deflection angle *a*

$$d = f\ \tan 2\alpha \tag{24.1}$$

f is the focal length of the objective lens. When α is very small, we can get

$$\alpha = \frac{d}{2f} \tag{24.2}$$

24.3 System Configuration

The autocollimator we developed consists of two optical imaging systems, laser, CCD and an embedded control processing system based on the digital signal processing unit.

Generally, common autocollimator only includes one optical imaging system; we add a new one composed of two convex lenses and a right angle prism to it to increase the focal length. As shown in Fig. 24.2, the laser at a wavelength of 658 nm with output power of 5 mw was chosen as the light source which can be seen, therefore, easily handled and aligned. Rays from the laser after being reflected twice issue from objective as a quasi-parallel beam enter the right angle prism, make an angle of 90° with the incident beam. The reflected laser beam goes through the objective lens and converges to the beam splitter; beam splitter separates the light whose wavelength is at 658 nm from the reflected light. The light at a wavelength of 658 nm reflected by the beam splitter reaches the first imaging plane where no CCD detector is placed on, passes through the two convex lenses, then is reflected by a right angle prism and finally reaches the second imaging plane where CCD is placed on. Neutral filter in front of CCD can effectively reduce the power of laser beam and reject other useless lights. The remaining reflected light passes through the beam splitter and reaches the plane of reticle which is exactly located in the focal plane of the objective lens and then we can see it via the eyepiece.

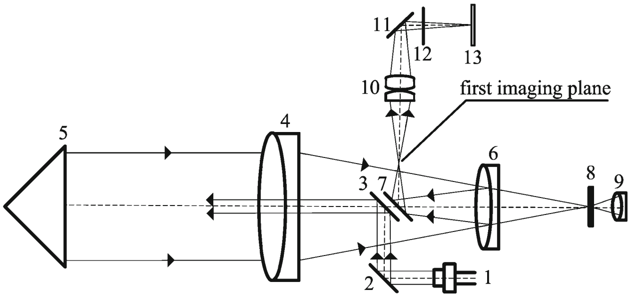


Fig. 24.2 Schematic diagram of the angle measuring system. *1* Laser; *2* reflector; *3* reflector; *4* objective lens; *5* right angle prism; *6* beam splitter; *7* reflector; *8* rectile; *9* eyepiece; *10* two convex lenses; *11* right angle prism; *12* neutral filter; *13* CCD

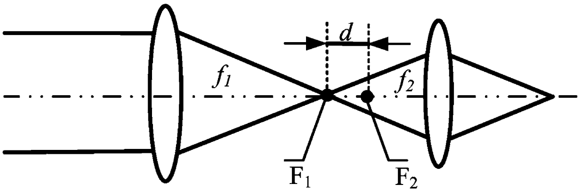


Fig. 24.3 Schematic diagram of the optical system's combination

The combination of two optical systems significantly increases focal length. As shown in Fig. 24.3, we assume the optical distance between two optical systems is d, the focal length of optical system 1 and 2 are f_1 and f_2 respectively. Hence the total focal length of the combined optical system is given by

$$f = \frac{f_1 f_2}{d} \tag{24.3}$$

In our system, $f_1 = 140$ mm, $f_2 = 10$ mm, $d = 2.5$ mm, thus $f = 560$ mm. Meanwhile, two convex lenses can largely decrease the spherical aberration to improve imaging quality.

One can infer from Eq. (24.2) that the resolution will be improved with the focal length increased. When the laser gets back into the tube through the objective lens, laser image will be accurately adjusted at the second imaging plane instead of the first one, therefore the focal length will be increased by 4 times. It is obvious that the resolution of optical system is 1.10″ according to Eq. (24.2) when the focal length of optical system f is 560 mm and image size p is 6 μm. The sub-pixel positioning method can further improve the resolution tremendously.

The chosen CCD chip (MTV9V032C12STM, Aptina) has 752H × 480 V pixels, whose unit cell size is 6 μm × 6 μm. The control and processing unit was constructed based on digital signal processing chip (TMS320VC33, Texas instrument, USA) with a clock rate of 60 MHz, which carried out the drive signals of the CCD, image capturing and sub-pixel locating. The sub-pixel subdivision technology is one of the effective solutions to improve the fixed and limited resolution of the CCD; the algorithm of barycenter is adopted to compute the coordinates of the image for it is easy to be implemented on digital signal processor. The barycenter algorithm effectively reduces noise and achieves a positioning accuracy of 0.1 pixels.

24.4 Device Calibration

The calibration device is consist of a precise angle dividing table whose accuracy is 0.1″, a right angle prism, a corner cube. Given the laser is quasi-collimated light beam; Eq. (24.2) is no longer suitable for calculating the deflection angle especially over long distance. Cubic polynomial fitting was used for calibration and calculation, which makes it possible to attain high accuracy.

The calibration procedure is as follows.

Firstly, level up the autocollimator precisely; place the precise angle dividing Table 150 m away from it; put the right angle prism on the precise angle dividing table and the corner cube on the top of right angle prism.

Secondly, the autocollimator emits laser beam and the beam projects onto the corner cube. Now the back scattered laser is uniformly distributed on optical lens of autocollimator; the imaging location on CCD is zero point x_0 and the deviation angle is 0″.

Thirdly, adjust the autocollimator pitch knob till the laser reaches the center of right angle prism and the reflected light is uniformly distributed on optical lens too.

Finally, rotate the precise angle dividing table from 35″ to −35″ with an interval of 5″, and record the deflection angle θ of the precise angle dividing table at each time. The image signals are captured by CCD and the image location x is calculated by digital signal processor using centroid estimation algorithm simultaneously.

According to cubic polynomial equation, we have

$$\theta = k_3 d^3 + k_2 d^2 + k_1 d^1 + k_0 \tag{24.4}$$

where θ is the deflection angle of the precise angle dividing table which is equivalent to the deflection angle of the right angle prism, d is the deviation distance on CCD and equals to x minus x_0, the aim of calibration is to acquire the calibration coefficients k_3, k_2, k_1, k_0.

Table 24.1 shows the results of the reflected beam imaging position x and deviation distance d on CCD as the right angle prism mounted on precise angle dividing table rotates from 35″ to −35″ with an interval of 5″.

Table 24.1 Deflection angle θ of right angle prism and corresponding deviation distance d on CCD

$\theta('')$	x(pixel)	d(pixel)	$\theta('')$	x(pixel)	d(pixel)
0	257.1	0	0	257.1	0
5	258	0.9	−5	256.1	−1
10	259.2	2.1	−10	255.1	−2
15	260.3	3.2	−15	253.8	−3.3
20	261.6	4.5	−20	252.5	−4.6
25	262.6	5.5	−25	251.3	−5.8
30	264.2	7.1	−30	249.8	−7.3
35	266.2	9.1	−35	248.1	−9

Using θ and d based on the data in Table 24.1, a cubic polynomial fitting curve can be attained. Figure 24.4 shows the fitting result; the x axis indicates the deviation distance d on CCD, the y axis means the deflection angle θ.

Calibration coefficients k_3, k_2, k_1, k_0 can easily be attained. $k_3 = -0.0106$, $k_2 = -0.0002$, $k_1 = 4.7284$, $k_0 = 0.1633$.

Figure 24.4 indicates that cubic polynomial curves have good linearity between $-25''$ to $25''$ and high non-linearity when deviation angle θ is bigger than $30''$ or lower than $-30''$. A large number of test results demonstrate that measurement accuracy and stability both increase as linearity of cubic polynomial curve improves, therefore, our autocollimator can achieve high accuracy within $\pm 30''$. This range can hardly be improved over a long distance for few reflected light can reach CCD detector with measuring distance increased.

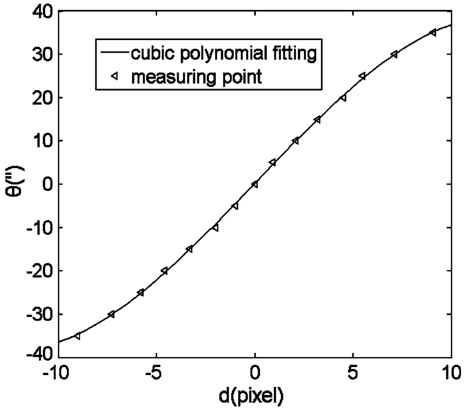


Fig. 24.4 Cubic polynomial fitting between the deviation distance d and the deflection angle θ

24.5 Experiment

Have known the calibration coefficients, the deflection angle of right angle prism can be calculated using Eq. (24.4). A large amount of experiments were conducted to test the accuracy, stability and anti-interference performance of the autocollimator.

24.5.1 *Accuracy*

A precise angle dividing table with a right angle prism placed on it is located 150 m away from the autocollimator to test its accuracy. Rotate the precise angle dividing table from 30″ to −30″,step by step (5″ per step), and record the image signals captured by CCD in the autocollimator simultaneously. Table 24.2 shows the relationship between real deflection angle of right angle prism and its measuring angle with measuring distance of 150 m. The real deflection angle of right angle prism obtained from precise angle dividing table is denoted as θ, the measuring angle obtained from autocollimator with measuring distance of 150 m are denoted as α. Δ indicates the measurement error and equals to the difference between θ and α.

The Standard deviation of Δ is 1.03″(1σ). The autocollimator achieves an accuracy of ±2″ over a distance of 150 m. Such accuracy can perfectly meet the requirement of initial attitude determination before missile launches or rocket launches.

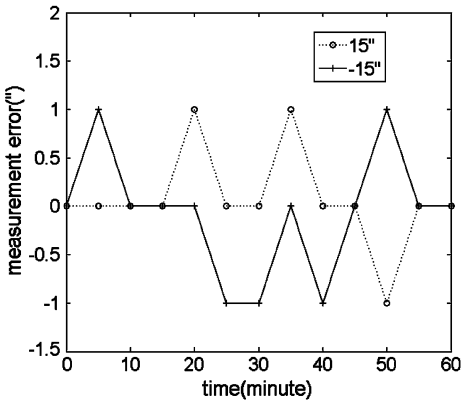
24.5.2 *Stability*

In order to monitor the stability of the autocollimator, an experiment lasted for a whole hour was conducted. The autocollimator was put 150 m away from the precise angle dividing table, deflection angle of right angle prism which is fixed on the table was then set at 15″ and the measuring angle on the autocollimator was

Table 24.2 The relationship between real deflection angle θ and measurement error Δ

θ(″)	α(″)	Δ(″)	θ(″)	α(″)	Δ(″)
0	0	0	0	0	0
5	6	1	−5	−4	1
10	11	1	−10	−9	1
15	15	0	−15	−16	−1
20	20	0	−20	−19	1
25	26	1	−25	−23	2
30	32	2	−30	−28	2

Fig. 24.5 Measurement error during an hour



recorded for every 5 min. Then another test was executed with the deflection angle set at −15″. Figure 24.5 shows the measurement errors, it can be seen that the measurement errors remain below ±1″ at 15″ and at −15″ during the whole process. In conclusion, the tests demonstrated that the stability of the autocollimator is quite high.

We also conducted repetitive experiment to test our autocollimator. Rotate the precise angle dividing table from 30″ to −30″ for every 15 min in 1 h,step by step (5″ per step), and record the measuring angle α by the autocollimator. The measurement error Δ, which is equal to real deflection angle θ minus to α, is less than ±2″. The test of repetition shows the autocollimator attains high robustness.

24.5.3 Anti-interference

Generally speaking, autocollimator performance is heavily influenced by illumination conditions. A standard or laboratory designed illumination level is often required to guarantee satisfied measuring accuracy. However, missile launches and rocket launches are operating over a long distance. To validate the anti-interference performance of the autocollimator, comprehensive tests spanning sunlight, flashlight, and CD reflection were performed.

We conducted an experiment in the sunlight, the deviation angle display on autocollimator changed slightly with a measurement error less than ±2″. Flashlight and CD reflection were used as interference sources; the measured deviation angles remain stable. It must be said that all the anti-interference experiments were conducted with a measuring distance of 150 m. It is as clear as crystal that the beam splitter and neutral filter in autocollimator are capable of filtering nearly all of the interfering lights.

24.6 Summary

The designed system can perform an azimuth measurement with a measuring distance of 150 m. An accuracy of ±2″ can readily be achieved over a measurement distance of 150 m during an hour; It can be used in cases which needs small angle measurement over long distance, such as rocket launches. The proposed method could be further enhanced to measure deflection angles in both azimuth and pitch using two-dimensional CCD arrays and appropriate reflector. Ranges of angle measurements could be improved as the optical lens which receives reflected light has bigger diameter.

References

1. Arp TB, Hagedorn CA, Schlamminger S et al (2013) A reference-beam autocollimator with nanoradian sensitivity from mHz to kHz and dynamic range of 107. Rev Sci Instrum 84:095007
2. Martinelli P, Musazzi S, Perini U (1994) An autocollimator based optical system for precise angular alignment control over large exploring areas. Rev Sci Instrum 65:1012
3. Yoder JPR, Schlesinger ER, Chickvary JL (1975) Active annular-beam laser autocollimator system. Appl Opt 14(8):1890–1895
4. Pedro MB, Girão S, Postolache OA et al (2001) An overview and a contribution to the optical measurement of linear displacement. IEEE Sens J 1(4):322–331
5. Ikram M, Hussain G (1999) Michelson interferometer for precision angle measurement. Appl Opt 38(1):113–120
6. Yuan J, Long XW (2003) CCD-area-based autocollimator for precision small-angle measurement. Rev Sci Instrum 74:1362
7. Luther GG, Deslattes RD, Towler WR (1984) Single axis photoelectronic autocollimator. Rev Sci Instrum 55:747
8. Korolev AN, Gartsuev AI, Polishchuk GS et al (2009) A digital autocollimator. J Opt Technol 76(10):624–628
9. Yuan J, Long XW, Yang KY (2005) Temperature-controlled autocollimator with ultrahigh angular measuring precision. Rev Sci Instrum 76:125106
10. Huang PS, Kiyono S, Kamada O (1992) Angle measurement based on the internal-reflection effect: a new method. Appl Opt 31(28):6047–6055
11. Zhao WQ, Tan JB, Qiu LR et al (2005) Enhancing laser beam directional stability by single-mode optical fiber and feedback control of drifts. Rev Sci Instrum 76:036101
12. Zhu F, Tan JB, Cui JW (2013) Common-path design criteria for laser datum based measurement of small angle deviations and laser autocollimation method in compliance with the criteria with high accuracy and stability. Opt Lett 21(9):11391–11403
13. Gao M, Dong ZR, Bian ZL et al (2011) Robust CCD photoelectric autocollimator for outdoor use. Chinese opt lett 9(9):091201
14. Li K, Kuang CF, Liu X (2013) Small angular displacement measurement based on an autocollimator and a common path compensation principle. Rev Sci Instrum 84:015108

Chapter 25
Study of Optical Environment Effect for Target Detect Algorithm Based on the Template Match

Di Yang, Yonghong Zhan, Honggang Wang, Chang'e Zeng and Qing Liu

Abstract Focus on the optical environment effect for target detect, the model of optical environment effect was modeled. The analysis tools were digital optical environment simulation platform based on the HLA. The models of target, optical environment, complexity evaluation, image match algorithm, detect probability, and false-alarm probability was molded. Based on the simulation platform, the image quality degradation caused by the optical environment and the performance of the optical detect were analyzed. Several conclusions were provided at last.

Keywords Complex optical environment · Image effect · Image quality degradation · Image detect

25.1 Introduction

Evaluating the performance of the infrared image system in the complex optical environment approached to the real usage condition is an importance work component in the field of test and measurement [1]. If simulating the all elements and boundary conditions for the infrared image system, the simulation and test equipments should be numerous and jumbled the test should be organized difficultly, and the consistency of the test result should be bad. So, the effect of optical environment could be separated to obscuration effect and decoying effect, and the test should simulate the environment effect not entity environment. For obscuration effect, the smoke generator could be utilized. The decoying effect could be equivalent effect simulated. In the equivalent effect simulation, the optical environment element entity should be neglected, and the decoying characteristic should be simulated. If so, the difficulty in the optical simulated equipments, the test organization, and the test result analysis should be reduced.

D. Yang (✉) · Y. Zhan · H. Wang · C. Zeng · Q. Liu
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: guangxuesheji@sina.com.cn

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_25

Two difficulties coming from the environment and detect system respectively would be existed in the equivalent effect simulating for decoy. Firstly, lack of the index to evaluate the complex optical environment, namely, changed quantity value between the optical environment included and the origin optical scene could not be acquired. Secondly, lack of the authorized index to evaluate the saliency of the target, namely, the relationship between the effect of the target saliency and detection performance of infrared image system could be acquired [2]. For the first problem, the model of image quality degradation should be used. Through compare the image matrix of the optical environment included image and origin image, the value of the difference should be calculated, and the difference were considered as the reason which cause the failure of acquiring the target. For the second problem, the relationship model should be build up, and the detection probability or false-alarm probability were considered as the characteristic effect specification of the saliency of the target.

To confirm and verify the viewpoint, a digital simulation platform was built up. The origin scene, complex optical environment included scene, detect system model, and analysis model were composed together in the platform [3]. All the models and units were netted by the HLA software. Through the simulation, the result should be provided for analyzing the relationship between the optical environment and the detect performance.

25.2 Architecture of PIP and Its Features

The digital optical simulation platform was composed with scene generation computer, optical environment computer, detection system computer, analysis computer and RTI network server [4]. The scheme of the simulation platform was shown as Fig. 25.1. Four computers were linked together as open loop, and represent the optical signal transmit and process coursed respectively from the target and recognize performance. The RTI server was used for the data interact between the computers. The computer function is shown as Fig. 25.1.

Scene generation unit: load and delete the scene and target; generate the terrain, target, background, weather image; and set the target parameters, such as size, position, temperature, and etc.

Optical environment unit: load the optical environment; set the parameters of the radiation environment (burst, flame, flash) and obscuration environment (smoke, dust), such as position, scale, number, power, and etc.; load and calculate the atmosphere model.

Detection unit: provide and chose the type of detect system; set the parameters of the optical lens and electric device for simulating different optical detection system; and provide the model of visible and infrared detection model.

Analysis unit: load the auto target recognize (ATR) algorithm; show the tracking effect and detect performance; evaluate the image degradation; show the effect of the degraded image to ATR algorithm and detect performance, and

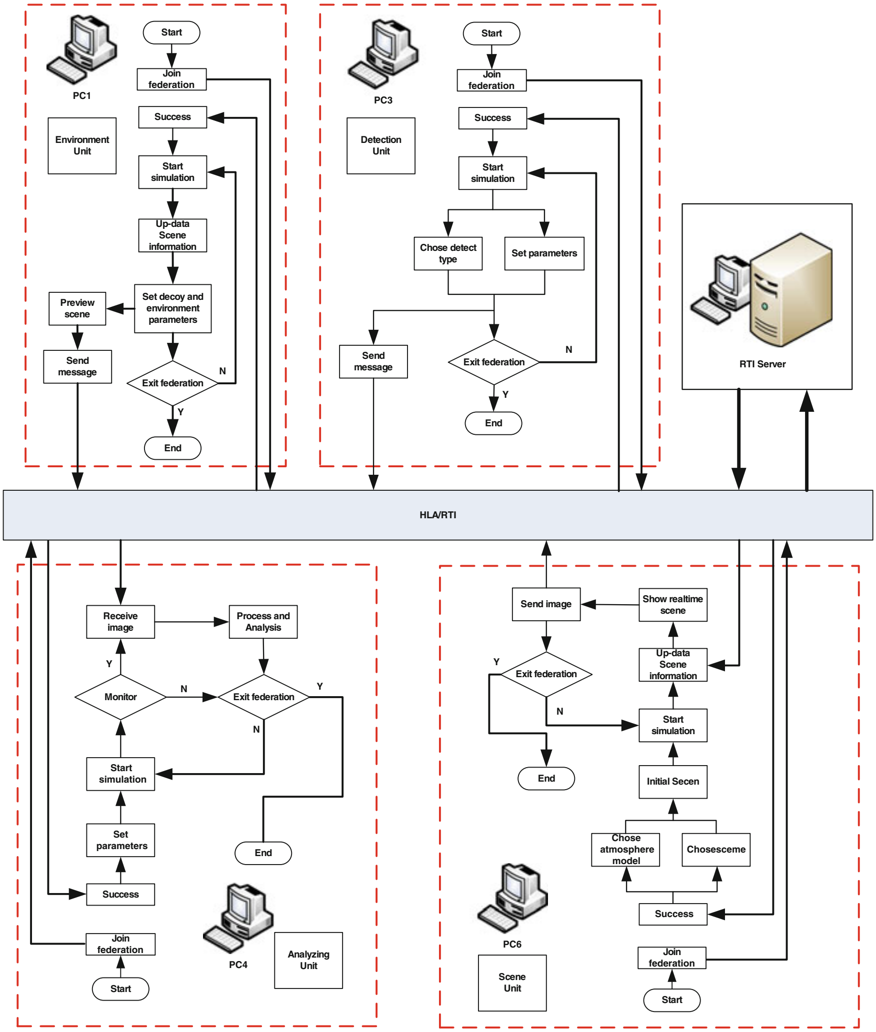


Fig. 25.1 The digital optical environment simulation platform

compare the tracking performance in origin and optical environment included scene.

RTI sever: manage the federation member, such as join or exit the federation; reorganize and reconfigure the simulation resource. Single federation which supports the interconnection and interoperation among the federal members was built up in the local net.

25.3 Quantity Value of Complex Optical Environment

When the optical environment was included to the origin scene, the saliency of the target should change, and the whole scene should change too. The mean average error (MAE), mean square error (MSE), normalized mean average error (NMSE), signal to noise rate (SNR), peak signal to noise rate (PSNR) and improvement factor of the signal to noise rate (ISNR) were used to weight the variation of the complex optical environment included and not scene [5], which is shown as function (25.1–25.6).

$$MAE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}|g_1(i,j) - g_0(i,j)|}{M \cdot N} \tag{25.1}$$

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[g_1(i,j) - g_0(i,j)]^2}{M \cdot N} \tag{25.2}$$

$$NMSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[g_1(i,j) - g_0(i,j)]^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}[g_0(i,j)]^2} \tag{25.3}$$

$$SNR = 10\log_{10}[\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}g_1(i,j)^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}[g_1(i,j) - g_0(i,j)]^2}] \tag{25.4}$$

$$PSNR = 10\log_{10}[\frac{255^2 M \cdot N}{\sum_{i=1}^{M}\sum_{j=1}^{N}[g_1(i,j) - g_0(i,j)]^2}] \tag{25.5}$$

$$ISNR = 10\log_{10}[\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[g_1(i,j) - g_0(i,j)]^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}[f_1(i,j) - g_0(i,j)]^2} \tag{25.6}$$

where, M and N are the pixel number of rows and columns of the image; $g_0(i, j)$ and $g_1(i, j)$ are the gray scale of the pixel of the point (i, j) in the optical environment include and not included image; and $f_1(i, j)$ is gray scale of the point (i, j) in standard recover image.

Low of MAE, MSE, NMSE mean the quality of the image is fine, while high of SNR and PSNR means mean the quality of the image is fine. ISNR is used to evaluate the image recover algorithm. The models above could be used to measure the quantity value of variation of the complex optical environment included and not included image. SNR and PSNR are the parameters used to calculate the recognize performance, and always used to weight the degree of image degraded. Note that the texture is an important and fine parameter to evaluate the ATR performance which is not considered in the paper.

25.4 Target Detect Algorithm Based on the Template Match

The template math and knowledge judge are two important ATR algorithms. The template mate algorithm is process the acquired image of the sensor and search the image template (target) in the processed image, and return the position information to the control system. Knowledge judge algorithm is process the acquired image and provides the characteristic of the image, then based on the knowledge to judge to track or attack the target.

The template match is a normal algorithm used in fire control and guidance. The template match algorithm could be classified to image template and characteristic template. The image template match is that a known image is fabricated to a matched template before, the algorithm shall search the most possible zone which is familiar with the template in the acquired image, and return the center index of the image which is recognized as the target position to the guidance system, and the direction should be the attack axis.

T is the template image, and S is acquired image of the sensor, the function (25.7) is used to describe the familiar between the target and searched image [6, 7].

$$\begin{aligned} D(i,j) &= \sum_{m=1}^{M}\sum_{n=1}^{N}\left[S^{ij}(m,n) - T(m,n)\right]^2 \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N}\left[S^{ij}(m,n)\right]^2 - 2\sum_{m=1}^{M}\sum_{n=1}^{N} S^{ij}(m,n) \times T(m,n) + \sum_{m=1}^{M}\sum_{n=1}^{N}[T(m,n)]^2 \end{aligned} \tag{25.7}$$

where, the first item is image energy, the third item is template energy, and the two items are non-relevant to match. So, the decisive item is correlation calculate item. The normalized correlation correlativity R is shown as function (25.8).

$$R(i,j) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} S^{ij}(m,n) \times T(m,n)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N} [S^{ij}(m,n)]^2}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N} [T(m,n)]^2}} \tag{25.8}$$

If the template is same as some zone of the image, the R is 0; if not, the R is a value between 0–1. In the working course, the algorithm is always searching the most possible position in the acquired image for each frame.

25.5 Effect of Image Degradation to ATR Algorithm

The simulation course is as follow: firstly, choose the interesting zone as the template ("+" target board), which is shown as the left image of Fig. 25.2; then set the flight route, range and speed, the frame speed is 10 fps which is shown as the middle image of Fig. 25.2. Then set the fire size, scale, and power, and comparing the image degradation value and recognition error which is shown as right image of the Fig. 25.2.

The right image and middle image shown the different of the fire included and not include. As the simulation show: when the fire environment is existed, the algorithm should provide the error position of the target, and the aircraft might lose the target [8].

In the interface of the analysis of the image degradation, the MAE, MSE, NMSE, SNR, PSN and ISNR could be selected to evaluate the image degradation after the fire included. In the Fig. 25.3, only the SNR, PSNR and MSE were selected.

As shown in the Fig. 25.3: when the aircraft flight to target board, SNR and PSNR get lower and lower, and MSE get higher and higher. Namely, the effects of the fire to the image degradation get higher and higher (according to the origin image).

Based on the template match algorithm, the max correlation point should be provided in the acquired image, and error should be existed between the correlation

Fig. 25.2 The simulation and the flight course

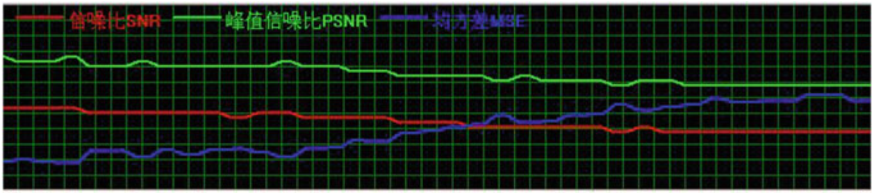


Fig. 25.3 The value of the image degradation

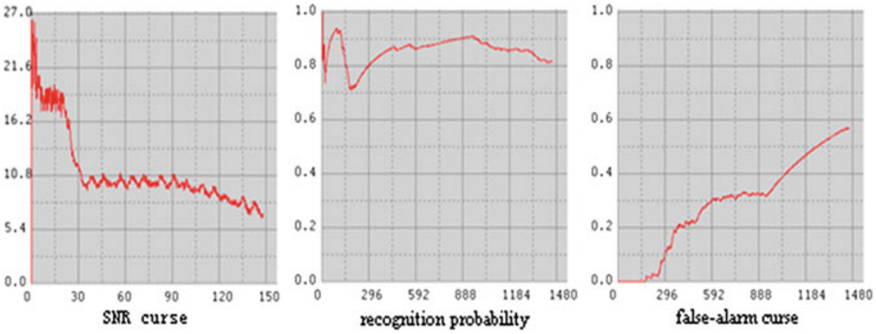


Fig. 25.4 Effect to recognition probability and false-alarm

point and target center. A fixed value could be set for ATR permitted, such as the 1/10 of the target board size. If the error is above the value, false-alarm was considered as existed, which is shown as Fig. 25.4.

From the Fig. 25.4, we can see: when the aircraft flight to the target board, the SNR gets lower and lower. Based on the template match algorithm, the maximum point could be acquired in each frame image, and the recognition probability was keep above 80 %. However, the error of departure gets higher and higher, which means ATR algorithm could not recognition the target efficiently.

25.6 Conclusion and Discuss

Based on the HLA, ATR algorithm performance was evaluated in the complex optical environments through the simulation platform. The result provided by the platform is familiar with the anticipation. It proved that the digital simulated scene, model of the image degradation, and ATR algorithm could be used for evaluating the performance of the image sensor in the complex optical environment.

Only the fire and image template match algorithm were utilized for evaluating the effect of the complex optical environments. The next, the other image degradation evaluation parameters and recognition algorithm should be involved in the simulation; the robust of different evaluation parameters should be test based platform; and the performance of the control system should be involved.

References

1. Yang D, Su JG, Zeng CE et al (2009) Study of optical target imaging quality evaluation. J Spacecraft TT&C Technology 28(3):34–38
2. Luo XM, Min HQ, Yang D (2008) Research on the impact of complex battlefield electromagnetic environment on the capability of missile combat system. J Acad Equip Command Technol 19(6):6–10
3. Zhang YN, Tang XY (2007) Research on performance evaluation approaches of infrared automatic target recognition algorithms. Infrared 28(6):15–20
4. Sadjadi FA, Bazakos M (2001) A perspective on automatic target recognition Evaluation technology. Opt Eng 30(2):2–5
5. Clark LG, Velten VJ (1990) Imaging characterization for automatic target recognition algorithm evaluations. In: Proceedings of SPIE 1310, Orlando, FL. doi:10.1117/12.21795
6. Peter RA, Strichland RN (1992) Imaging complexity metrics for automatic target recognizers. In: Automatic target recognizers systems and technology conference, London
7. Schmieder DE, Weathersby MR (1983) Detection performance in clutter with variable resolution. IEEE Trans Aerospace Electrons Sys AES-19(4):622–630
8. Shirvakar MV, Trivedi MM (1992) Studies in robust approaches to object detection in high-clutter background. Proc SPIE 1468:52–59

Chapter 26
The New Numerical Differentiation for Ballistic Determination

Min Chai, Min Wang, Tiening Nie and Si Shen

Abstract In the measuring system of modern shooting ranges, numerical differentiation, which is key to the process of external ballistic measurement determination of carrier rockets, must be adopted for calculating kinematic velocity and acceleration parameters of carrier rockets by means of optical and radar measurement data. As existing numerical differentiation is not steady enough, the loss function is improved by fitting differential algorithm and this paper establishes a new steady fault-tolerant numerical differentiation with relatively high accuracy and fault-tolerant ability, so as to improve and enrich high-accuracy external ballistic measurement computing technology. The simulation results show that new numerical differentiation technique has high accuracy and practical engineering value.

Keywords Numerical differentiation · Ballistic determination · Steady fitting differential · Fault-tolerant

26.1 Introduction

Numerical differentiation technology is the critical technology to determine velocity, acceleration, trajectory inclination angle and deflection angle of carrier rockets by making use of measurement data of theodolites, radar, Doppler velocity-measuring system and GPS system [1].

As is known to all, the metrical information of a theodolite is as follows: range R, angle of pitch E, and azimuth A and so on; information for pulse radar is range R, angle of pitch E, azimuth A, range rate $\dot{R}$, etc.; GPS system: range R, range rate $\dot{R}$, and carrier phase ϕ; Doppler velocity-measuring system: range rate $(\dot{R}_1, \ldots, \dot{R}_m)$ [2, 3]. In

---

M. Chai (✉) · M. Wang
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: 13991177352@163.com

M. Chai · M. Wang · T. Nie · S. Shen
Xi'an Satellite Control Center, Xi'an 710043, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_26

consideration of the complexity of kinetic model for carrier rockets and the afterwards treatment concerning external ballistic measurement data, orientation information may calculate the three-dimensional coordinates $(x(t), y(t), z(t))$ of a target by appropriately choosing reference system, and velocity information may calculate the three-dimensional velocity component $(\dot{x}(t), \dot{y}(t), \dot{z}(t))$ of a target in reference system above independently or in a way of combination with orientation information.

However, in order to make analysis of the flight conditions of carrier rockets and separate guiding system errors, in the high-accuracy treating process of external ballistic measurement data, we have to calculate the following for a target: position coordinates $(x(t), y(t), z(t))$, flight speed parameters $(\dot{x}(t), \dot{y}(t), \dot{z}(t))$, trajectory inclination angle $\theta(t)$, deflection angle $\sigma(t)$, component of acceleration $(\dot{x}(t), \dot{y}(t), \dot{z}(t))$, tangential acceleration $\dot{v}(t)$, lateral acceleration $\dot{v}_\theta(t)$ and normal acceleration $\dot{v}_\sigma(t)$. Therefore, numerical differentiation must be adopted. Orientation information or coordinate component is used to calculate kinematic velocity and acceleration parameters of a target; or speed information and parameters are used to calculate acceleration parameters. In the treating process of external ballistic measurement data on space flight measurement and control engineering, the common numerical differentiation is based on multinomial-fitting multinomial smoothing differential or orthogonal polynomial smoothing differential technology [1]. As existing numerical differentiation is not steady enough and short of fault-tolerant ability, this paper establishes a new steady fault-tolerant numerical differentiation with relatively high accuracy and fault-tolerant ability, so as to improve and enrich high-accuracy external ballistic measurement computing technology; and utilizes actual numerical value for simulation to verify the reliability and effectiveness of the new numerical differentiation technology.

26.2 Steady Fitting Differential Algorithm Design Model

In statistics, the loss function is a measure of the loss and the degree of error function. The common multinomial smoothing differential algorithm is not steady and short of fault-tolerant ability, basically because it adopts the loss function $\rho(x) = \frac{1}{2}x^2$ whose derived function is $\phi(x) = \dot{\rho}(x) = x$; it equally treats each measurement data in calculation, rather than differentiate them to check whether information acquired is normal or not, and thus, outliers cause adverse effect in the calculation process that they are obviously higher than normal data points [4–6].

26.2.1 *Loss Function and ϕ-Function*

In order to improve the fault-tolerant ability of algorithm for data, based on the existing fitting smoothing differential algorithm, this paper improves loss function as follows:

$$\rho_{FT}(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq 3\sigma \\ 1.5|x|\sigma, & 3\sigma < |x| \leq 4.5\sigma \\ -x^2 + 12|x|\sigma - 27\sigma^2, & 4.5\sigma < |x| \leq 6\sigma \\ 9\sigma^2, & |x| > 6\sigma \end{cases} \quad (26.1)$$

The threshold value index ϕ in the formula can be in Error variance estimates of root of meassured data, or can be empirically chosen.

Correspondingly, ϕ- function is the mode of redescending:

$$\phi_{FT1}(x) = \begin{cases} x & |x| \leq 3\sigma \\ 3\sigma & 3\sigma < x \leq 4.5\sigma \\ -3\sigma & -4.5\sigma \leq x < -3\sigma \\ -2x + 12\sigma, & 4.5\sigma < x \leq 6\sigma \\ -2x - 12\sigma, & -6\sigma \leq x < -4.5\sigma \\ 0 & |x| > 6\sigma \end{cases} \quad (26.2)$$

Compare Formulae (26.1) and (26.2) and the loss function based on least square theory and corresponding ϕ- function, Formulae (26.1) and (26.2) may effectively overcome the disadvantage "that differential algorithm's lack in fault-tolerant ability is as a result of measurement data used in common quadratic loss function that fails to be distinguished", as shown in Fig. 26.1.

In Comparative Analysis Chart 26.1a or b, we may learn from the two curves that adopting Formula (26.2) to replace LS estimated loss function has the following advantages [4]:

1. For normal measurement data, $\rho_{LS}(x) = \rho_{FT1}(x)$ and $\phi_{LS}(x) = \phi_{FT1}(x)$ $(|x| \leq 3\sigma)$ may guarantee steady fitting differential algorithm can obtain common fitting smoothing differential result, and make sure of high-accuracy in calculation.
2. For abnormal measurement data, $\rho_{FT1}(x) = c$ (a constant) and $\phi_{FT1}(x) = 0$ $(|x| > 6\sigma)$ may effectively reduce the adverse effect on abnormal data and ensure reliability of differential computing result and avoid ballistic parameter distortion.
3. The dedicate design of $\rho_{FT1}(x)$ and $\phi_{FT1}(x)$, enables data information to have a quantitative transition from normal data to abnormal data in the using process of data information, which differs from conventional "outlier picking out" processing. It not only avoids the computing trouble caused by data with unequal intervals, but also makes full use of effective information in a scientific way, which is in favor of improving ballistic computation quality.

26.2.1.1 Steady Fault-Tolerant Fitting Differential Algorithm Design

Numerical differential algorithm is simplify described in order to facilitate the research and establish, and numerical differential will be abstractively simplified

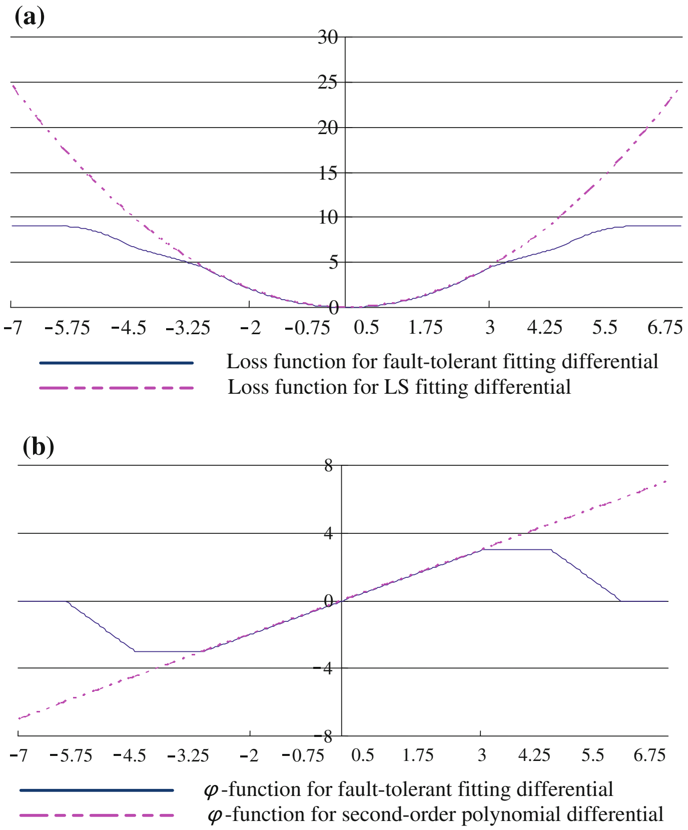


Fig. 26.1 Comparison of loss function and ϕ- function in LS fitting differential and fault-tolerant differential. **a** Loss function **b** ϕ- function

representation for: Assume that the data sequence $\{\kappa(t_i)|i = 1, 2, 3, \ldots\}$ is the direct measurement of objects or after data conversion calculation and generation of discrete time series data of $\kappa(t)(t \geq 0)$, the core problem of numerical differential is how to struct operator T appropriate:

$$T\{\kappa(t_i)|i = i_a, i_a + 1, \ldots, i_b - 1, i_b\} \to \kappa(t)(t_{ia} \leq t_{ib}) \tag{26.3}$$

$\hat{\kappa} = T\{\kappa(t_i)|i = i_a, i_a + 1, \ldots, i_b - 1, i_b\}$ is called numerical differential algorithm in short. The performance of the algorithm, and the degree of agreement depends directly on the. Algorithm directly depends on the degree of agreement of $\hat{\dot{\kappa}}(t)$ and $\dot{\kappa}(t)$.

Suppose a measuring object $y(t)$ is measured by a type of external-ballistic measurement system, we may acquire the total points N, smoothing differential semi-points n, and the measurement data sequence$\{y(t_i)|t_i = t_0 + ih, i = 1, 2, 3, \ldots N\}$ with sample intervals h [7, 8].

1. In the window n as the point of the sliding sampling interval, suppose $\{y(t_j)|j = i-n,\ldots,i+n\}$ is the sampling data fragment, use formula (26.1) to construct the objective function $S(a,b,c) = \sum_{j=i-n}^{i+n} \rho_{FT}(\kappa(t_j) - a - bt_j - ct_j^2)$, and the minimum point:

$$(\hat{a},\hat{b},\hat{c}) = \arg\min \sum_{j=i-n}^{i+n} \rho_{FT}(\kappa(t_j) - a - bt_j - ct_j^2)$$

2. Calculate steady fitting multinomial coefficient:

For each point of $j = n+1,\ldots,N-n$, take $2n+1$ points of $y_{j-n},\ldots,y_j,\ldots y_{j+n}$ as coefficients of multinomial fitting calculation.

$$a_0^{(0)} = \frac{-c\sum_{i=-n}^{n} y_i}{b^2 - ac} + \frac{bh^2\sum_{i=-n}^{n} i^2 y_i}{b^2 - ac} \tag{26.4}$$

$$\begin{aligned} a_1^{(0)} &= \frac{h}{b}\sum_{i=-n}^{n} iy_i \\ a_2^{(0)} &= \frac{b}{b^2 - ac}\sum_{i=-n}^{n} y_i + \frac{ah^2}{b^2 - ac}\sum_{i=-n}^{n} i^2 y_i \end{aligned} \tag{26.5}$$

$$a = 2n+1, \quad b = \frac{h^2}{3}n(n+1)(2n+1)$$

$$c = \frac{h^4}{15}n(n+1)(2n+1)(3n^2+3n-1)$$

3. Residual examination:

$$\bar{y}_i = a_0^{(0)} + a_1^{(0)} ih + a_2^{(0)}(ih)^2 \tag{26.6}$$

$$v_i' = y_i - \bar{y}_i \tag{26.7}$$

$$v_i = \begin{cases} -\varepsilon & v_i' < -\sigma \\ v_i' & |v_i'| \leq -\sigma \\ \varepsilon & v_i' > -\sigma \end{cases} \tag{26.8}$$

$$\phi_i = \begin{cases} 0 & \left|v_i'\right| > \sigma \\ 1 & \left|v_i'\right| \le \sigma \end{cases} \tag{26.9}$$

4. If φ_i is totally equal to 1, then we may obtain:

$$\bar{y}_j = \bar{y}_0 = a_0^{(0)} \tag{26.10}$$

$$\bar{y}_j' = \bar{y}_0' = a_1^{(0)} h \tag{26.11}$$

5. If φ_i is not totally equal to 1, then make:

$$J = \begin{bmatrix} 1 & -nh & n^2h^2 \\ \vdots & \vdots & \vdots \\ 1 & nh & n^2h^2 \end{bmatrix}_{(2n+1,3)}$$

$$\Phi = \begin{bmatrix} \varphi_{-n} & & 0 \\ & \ddots & \\ 0 & & \varphi_n \end{bmatrix}_{(2n+1,2n+1)}$$

$$V = \begin{bmatrix} v_{-n} \\ \vdots \\ v_n \end{bmatrix}_{(2n+1)}$$

Calculate Δa and $a^{(1)}$, make

$$a^{(0)} = \begin{pmatrix} a_0^{(0)} \\ a_1^{(0)} \\ a_2^0 \end{pmatrix}, \ a^{(1)} = \begin{pmatrix} a_0^{(1)} \\ a_1^{(1)} \\ a_2^{(1)} \end{pmatrix}$$

We may obtain:

$$\left.\begin{array}{l} \Delta a = (J'\Phi J)^{-1} J'V \\ a^{(1)} = a^{(0)} + \Delta a \end{array}\right\} \tag{26.12}$$

From this

$$\bar{y}_j = a_0^{(1)} \tag{26.13}$$

$$\bar{y}'_j = a_1^{(1)} h \tag{26.14}$$

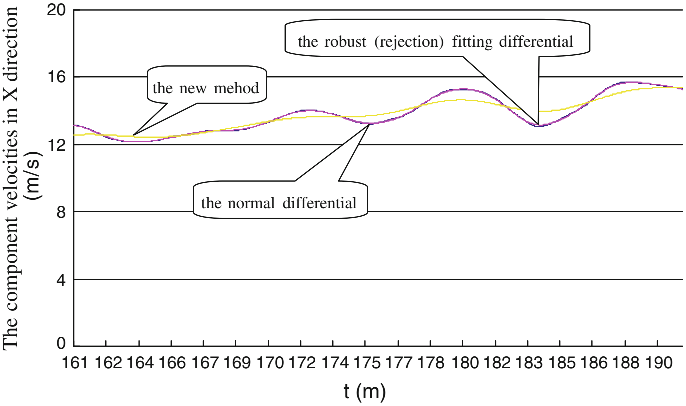


Fig. 26.2 The component velocities in *X* direction for differential results achieved by adopting the new and former methods

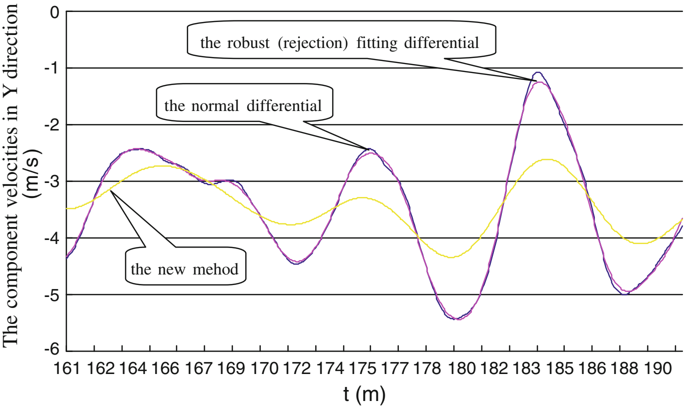


Fig. 26.3 The component velocities in *Y* direction for differential results achieved by adopting the new and former methods

26.3 Simulation Analysis Based on Actually Measured Data

In order to analyze and verify the effect of steady fault-tolerant fitting differential, this paper takes one experiment for example, the normal differential (no rejection), robust (rejection) fitting differential and fault tolerant robust fitting differential calculation of ballistic target parameters of the simulation analysis, making analysis of simulation. Figures 26.2, 26.3, 26.4, 26.5, 26.6, and 26.7: The result curve charts obtained by adopting the two methods to determine targeted ballistic parameters

Via simulation analysis, we learn that the steady fault-tolerant fitting differential algorithm may well avoid the fitting curve deformation caused by abnormal data in

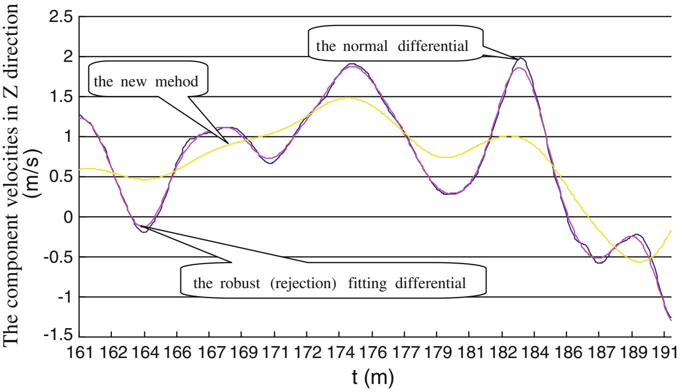


Fig. 26.4 The component velocities in *Z* direction for differential results achieved by adopting the new and former methods

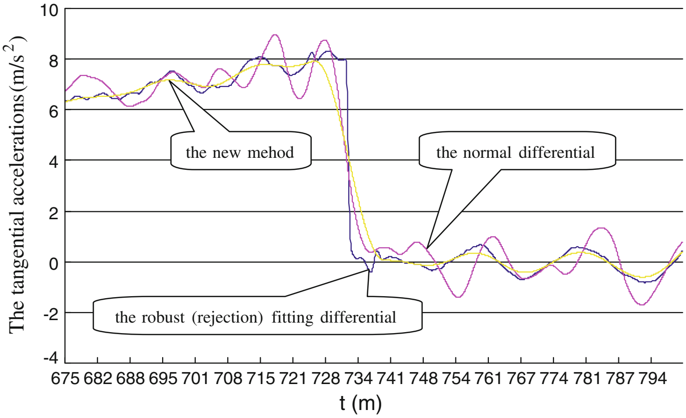


Fig. 26.5 The tangential accelerations for differential results achieved by adopting the new and former methods

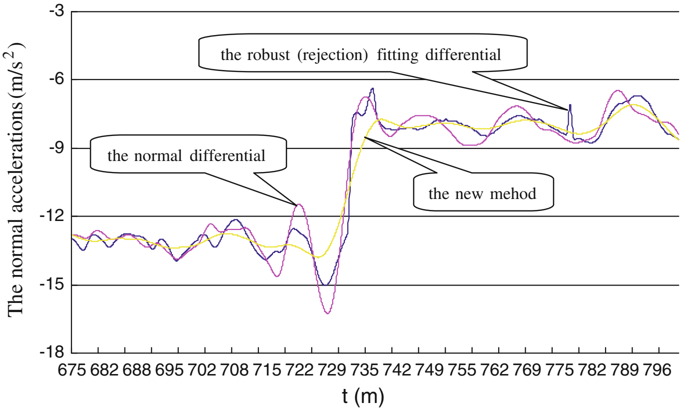


Fig. 26.6 The normal accelerations for differential results achieved by adopting the new and former methods

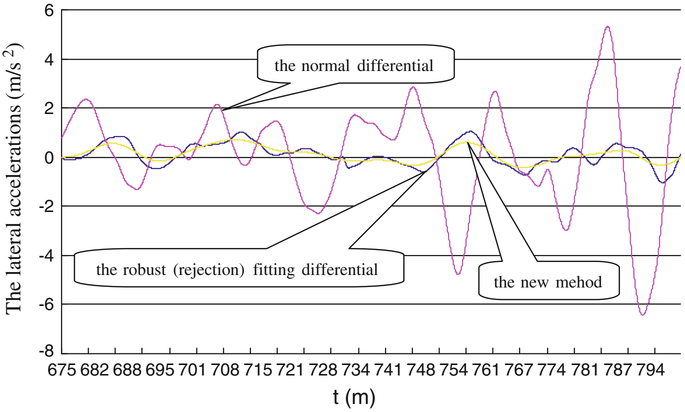


Fig. 26.7 The lateral accelerations for differential results achieved by adopting the new and former methods

fitting process; and that the algorithm is less sensitive to random error, avoiding getting unsafe data. As shown in Figs. 26.2 and 26.7, obviously, in the normal arc, despite the sliding interval, differential (no rejection), robust (rejection) fitting differential and fault tolerant robust fitting differential calculation results are basically identical, but in the feature segment, robust (rejection) fitting differential causes distortion of ballistic parameters. In the overall trend analysis, fault tolerance and robust fitting differential the treatment results were better than the normal differential (no rejection) and robust(rejection) processing the results of fitting differential, and also it is proved that the new method has good robustness and fault tolerance.

26.4 Conclusions

Actual result proves that the steady fault-tolerant fitting differential algorithm in this paper is obviously superior to classical numerical differentiation, for it may effectively eliminate the effect of outliers on computing results. Moreover, the steady fault-tolerant fitting differential algorithm is able to better reduce the effect of random error on computing results, and satisfy the need for high-accuracy data processing to a higher extent.

References

1. Liu LS (2002) External-ballistic measurement data processing. National Defense Industry Press, Beijing, pp 270–366
2. Wang ZM (1999) Calibration and assessment of ballistic tracking data. The Press of National University of Defense Technology, Changsha, pp 186–271
3. Wang ZM, Yi DY (1996) Measurement data modeling and parameter estimation. The Press of National University of Defense Technology, Changsha, pp 310–322
4. Hu F, Sun GJ (2000) The fault-tolerant smoothing and fault-tolerant differential smoothing based on multinomial fitting. Chinese J Eng Math 17(2):53–57
5. Hu SL, Sun GJ (1999) The statistical diagnosis technology regarding outliers of external-ballistic measurement data in shooting ranges. J Astronaut 20(2):68–74
6. Matassini L, Kantz H (2002) Optimizing of recurrence plots for noise reduction. Phys Rev E 65 (2):1–6
7. Gao C, Qiao YF (2007) Real time error correction method for photoelectrical theodolite. Opt Precis Eng 15(6):846–851
8. Hu SL, Xu AH, Guo XH (2007) Technic of pulse radar tracking and measurement data processing. National Defense Industry Press, Beijing, pp 168–171

Chapter 27
Methods of Increasing Velocity Measurement Precision for Sky Screen

Zhiyuan Feng, Haibin Hu, Shunhua Zhu, Zhen Wang and Liucheng Miao

Abstract The velocity measurement precision of sky screen was analyzed in this paper. Problems involving measuring distance between two sky screens and placement of sky screen will be discussed. Using tapeline to measure the distance between two sky screens would cause 5 kinds of error; using normal counter to measuring time interval would cause a "±1" pulse error. A few ways were brought forward to increase precision for measuring distance and time interval, which could eliminate the influence of human error. Those methods are controllable, simple and convenient. Applying those methods would increase the precision of velocity measurement of sky screen to a degree that would satisfy the requirement of testing.

Keywords Sky screen · Time interval measurement · Distance measurement · Velocity measurement

27.1 Introduction

The velocity measurement is a regular testing item. Velocity measurement device has a great variety, such as velocity measurement radar, coil target, light screen, sky screen and tinfoil target. Owning to its convenient operation, small measurement error, high reliability, sky screen has become the mainstream measurement device. This paper analyzed velocity measurement precision of sky screen, discussed interfering factor on using sky screen, and brought up some methods for improving measurement precision. After use those methods, velocity measurement data produced by sky screen can be more reliable.

Z. Feng (✉) · H. Hu · S. Zhu · Z. Wang · L. Miao
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: fengzhiyuanmail@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_27

27.2 Analysis of the Velocity Measurement Precision of Sky Screen

The sky screen is mainly used for detecting time that a flying object arrives at a certain place. It utilizes the sky as background, and uses phototube to detect the change of luminous flux. Two sky screens and one time measurement device can form a velocity measurement system, which can measure velocity of flying object. As the flying object flies over two sky screens, a pulse signal denoting starting will be sent from one screen to the time measurement device, and another signal denoting stopping will be sent from the other. The time measurement device computes the time interval (t) of the object takes to fly over the two sky screens. Given the distance (S) between two sky screens, the average velocity of the flying object between two sky screens can be computed by the following expressions.

$$V = \frac{S}{t} \tag{27.1}$$

According to the theory of error transferring [1], the precision of average velocity is expressed as:

$$\Delta v = \frac{\Delta S}{S} + \frac{\Delta t}{t} \tag{27.2}$$

where ΔS is the error of distance measurement between two sky screens, $\Delta \mathrm{t}$ is the error of measured time interval.

The above expression shows that the precision of computed velocity is caused by error of measured distance between two sky screens and error of measured time interval for the object to fly over the two sky screens.

27.2.1 Error of Measured Time Interval

Recently, time interval measurement mainly uses counter. Waveform mechanism of the counter is shown in Fig. 27.1.

In reality, starting time and stopping time are random, thus their positions at time axis are random. As a result, time interval from counter has measurement quantization error. If counter acts under front edge of pulse, Δt_1 is time interval between strobe starting and front edge of following pulse, and Δt_2 is time interval between strobe stopping and front edge of following pulse, Δt_1 and Δt_2 are quantization error of time measurement. Both Δt_1 and Δt_2 are less than T, so counter has "±1" pulse error. Normally, the quantization error of time measurement of counter is 1 μs [2].

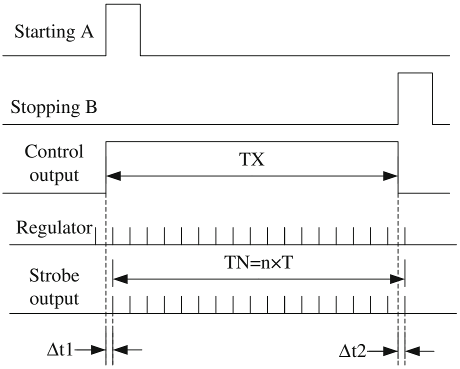


Fig. 27.1 Waveform demonstrating mechanism of counter

27.2.2 Error of Measured Distance

In reality, the distance between two sky screens can be measured by steel tapeline. The sign of sky screen is used to pinpoint the location of screen. Human operator measures the two signs on the two sky screens. Due to characteristic of steel tapeline, there are 5 kinds of error in measured distance between two sky screens [3]:

Error of indication: this error is a half of the minimum scale. Now, the minimum scale of steel tapeline is 1 mm, so the error of indication (ΔS_1) is 0.5 mm.

Observational error: when human operator observes the sign of sky screen and the scale line of steel tapeline, the sign may not be on a straight line with the scale line observational error can be computed by the following expression:

$$\Delta S_2 = \frac{L\psi}{e} \tag{27.3}$$

where L is distance of distinct vision for human eyes, which is normally 250 mm; ψ is pointing accuracy of the eyes of operator, which is normally 20″; e is conversion coefficient between radian and second, $e = 2.06 \times 105$.

Temperature error: the length and error shown in calibration certificate of steel tapeline are obtained in 20 °C. However the temperature of operating environment may be different from standard temperature. Length of tapeline will change as temperature changes. The change of length of tapeline can be computed by following expression:

$$\Delta S_3 = S\alpha\Delta T \tag{27.4}$$

where S is measured distance; α is expansion coefficient for the steel tapeline, which is 11.5×10^{-6}; ΔT is difference between the temperature of operating environment and standard temperature.

Pull change error: when measurement range exceeds 5 m, human operator shall exert a force to tapeline for it to be plain and straight. According to the calibration certificate, the standard force exerted should be 49 N. When the force is different from standard force, there will be pull change error. This error will be computed by the following expression:

$$\Delta S_4 = \frac{\Delta PS}{EA} \tag{27.5}$$

where ΔP is difference between the applied force and standard force; S is measured distance; E is coefficient of elasticity for steel tapeline, which is 205,800 N/mm^2; A is cross-sectional area of tapeline.

Measurement status error: tapeline is placed in testing platform when calibrating, it is suspended between two sky screens when actually used. Because of gravity, the length of tapeline will change, this error can be computed by following expression:

$$\Delta S_5 = \frac{(W_g \times S)^2 S}{24P^2} \tag{27.6}$$

where W_g is weight of a unit length of tapeline; S is measured distance; P is the force that exerts the tapeline.

27.2.3 Analysis of Precision of Velocity Measurement

27.2.3.1 Tapeline Factor

Assume that the actual pull force is 100 N. The distance between two sky screens varies from 2 to 10 m, and the velocity of the flying object is 100, 500 or 900 m/s. According to the velocity measurement precision expression, the precision will be computed, and shown in Fig. 27.2.

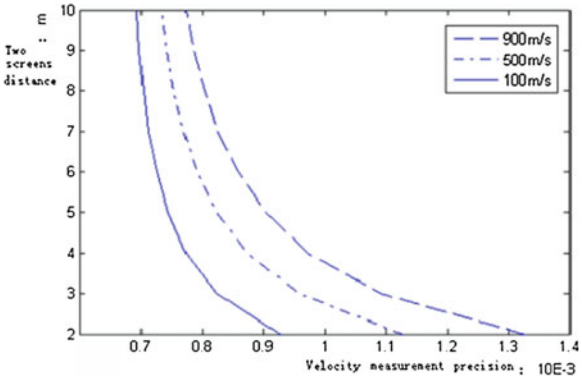


Fig. 27.2 Precision curve considering tapeline only

From the curve of precision in Fig. 27.2, the higher the velocity is, the bigger the precision will be. That is because as velocity becomes higher, the time interval the object takes to fly over the two sky screens become shorter, so error of measured time interval will influence precision of measured velocity more badly. Similarly, as the distance between two sky screens becomes shorter, the time interval the object takes to fly over them also gets shorter, causing the error of measured time interval to influence the precision of measured velocity more badly.

27.2.3.2 Placement of Sky Screen

1. Two screens are not parallel.

Normal placement method of sky screen is to use plain and straight tapeline to close to sign of sky screen and make side of sky screen parallel to the tapeline. When distance from two side of sky screen to the sign is equal for both sky screens, the two screens are parallel. According to analysis of error of steel tapeline, error exits in distance measured by tapeline. If this error is 1 mm and length of side of sky screen is 140 mm, the error will cause the two screens to be unparallel, with an angle about 0.409°. If object flies over screen form a position that deviates from the center of the screen by 0.1 m, the distance measurement error (ΔS_6) will be 0.7 mm [4].

2. The height of two sky screens is not the same.

Because two sky screens are not designed for one bed, the height of two sky screens may not be the same, especially when two sky screens are far away from each other. The distance of two sky screens is measured by tapeline, thus an error may be introduced due to difference in height of two sky screens. This error can be computed by the following expression:

$$\Delta S_7 = \sqrt{S^2 + \Delta H^2} - S \tag{27.7}$$

where S is measured distance; ΔH is height difference between two sky screens, which is assumed to be 0.1 m.

In Fig. 27.2, when the distance of two sky screens is longer than 4 m, velocity of object is 100, 500 or 900 m/s, the precision of velocity measurement is less than 0.1 %. When error caused by placement of sky screen is considered, even if the distance of two sky screens is longer than 10 m, the precision of velocity measurement does not satisfy the requirement of testing.

In Fig. 27.3, these parameters were selected to be very small. If human operator does not try to avoid error caused by placement method, the precision of velocity measurement is very badly. This paper will study methods to control precision of velocity measurement to provide credible testing data.

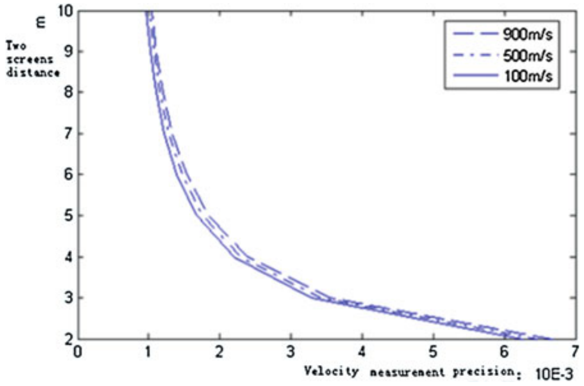


Fig. 27.3 Error curve add placement factor

From Fig. 27.2, method of measuring time interval shall be studied and improved, decreasing influence of error of time interval to such a small scale that it can be ignored, and measurement method of distance between two sky screens shall also be studied, in order to increase velocity measurement precision.

27.3 Improvement on Time Interval Measurement Methods

Previous analysis shows error of time interval influences the precision of velocity measurement. Set distance of two sky screens is 10 m, error of time interval as 10, 100 and 1,000 ns, compute the precision of velocity measurement result is shown in Fig. 27.4.

From Fig. 27.4, when the error of time interval is less than 100 ns, the precision of velocity measurement satisfies the requirement of testing. So we can improve the error of time interval in order to improve the precision of velocity measurement.

Recently, there are some methods for measuring time interval, such as counter, vernier method, analog interpolation method and delay method.

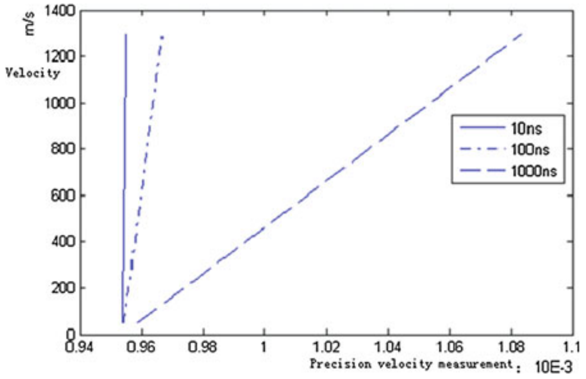


Fig. 27.4 Error curve in time interval influence

Mechanism of counter is starting and stopping. There are two independent channels: one channel is channel A, the other is channel B. Channel A will be input starting signal, and start trigger for starting counting. Channel B will be input stopping signal, and start trigger for stopping counting. Width of pulse of time strobe is time interval. In the period of time strobe, time standard (frequency is f) counts the number of cycle, measures time interval between signal A and signal B. To obtain more precise time interval, the frequency of time standard should be increased.

Vernier method is based on counter, it adopts vernier theory to measure oddment time (from strobe start to front edge of following pulse) and mantissa time (from strobe stopping to front edge of following pulse). This method uses two counters with close time standards to measure oddment time and mantissa time [5, 6].

Analog interpolation method is also based on counter. In the period of oddment time and mantissa time, it uses large current to charge the capacitor, then uses small current to discharge the capacitor. It then uses time of discharge and the ratio of large current and small current to compute oddment time and mantissa time [2].

Delay method is a measurement device. The device has a lot of delay units. Delay of those units is the same. Time interval can be measured by detecting signals that spread in these units [2].

The error of time interval measurement of counter may be ns. It requires the frequency of time standard of counter is very high, so the operation environment of counter is very rigorous. The error of time interval measurement of vernier method is 20 ps. the frequency of time standard of counter is low, but it requires stability and veracity of the frequency of time standard is very high. Ideally, the error of time interval measurement of analog interpolation method is ps, but there are many disadvantages as well, such as long discharge time, non-linear discharge ratio, instability under different temperatures. The error of time interval measurement of delay method is very high. Because the device is complex, the range of measurement is very small [7].

From above analysis, vernier method may be used in field of velocity measurement using sky screen.

27.4 Improvement on Distance Measurement Method of Two Sky Screens

27.4.1 Improvement on Distance Measurement Method

According to previous analysis, tapeline-measured distance has obvious error. The paper suggests that distance measurement should be conducted using laser telemeter.

GLE laser telemeter from IVO corp, has a range of measurement from 0.1 to 500 m. This laser telemeter measures distance by comparing phase. However long the distance is, this laser telemeter can provide exact position data. Error of distance is 0.125 mm. Repetition is also feasible. According to the range of measured distance, error of distance measurement is 0.5–2 mm. Height of telemeter is 105 mm. Width of telemeter is 54 mm. Depth of telemeter is 140 mm. Weight of telemeter is about 1 kg. Installation of the device is very simple. Working environment temperature is −25 to 45°C. This temperature range can meet testing environment [8].

Laser telemeter is installed on one sky screen, baffle-board is installed on another. Irradiation position should be pinpointed in the laboratory. In using, aim laser beam of meter at irradiation position. Laser telemeter can provide exact distance data. In this method, error of operator and method will be avoided in distance measurement by using tapeline. The error of measured distance comes from laser telemeter only.

27.4.2 Improvement on Placement Method

Laser collimator will be installed in sky screen to ensure that two screens are parallel, and the height of two screens is the same. The laser beam plumb up the screen. This step should be finished in the laboratory. In using, human operator sets screen plumb up the earth, then adjusts two laser collimators aim at each other. When two laser beams become one beam, two screens should plumb up earth, and height of two screens is the same.

27.5 Precision Analyzing After Improvement

From Fig. 27.4, if error of measured time interval less than 10 ns, it can be ignored. Using improved distance measurement method and placement method, error of measured distance is very exact. Select 2 m as distance between two sky screens.

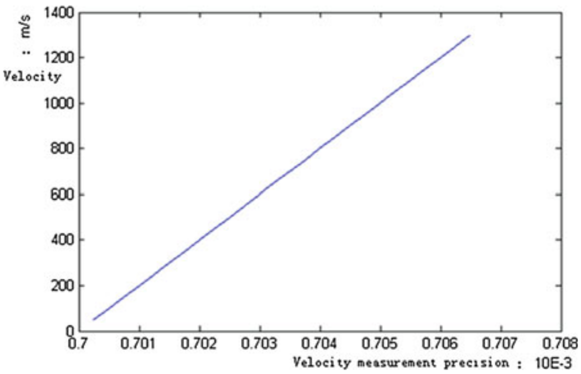
Laser telemeter measures 2 m distance. Error can be controlled in 1 mm.

Diameter of laser bean is 1 mm. Distance of two sky screens is 2 m. The angle between two screens should be 0.03°. Assume object flies over screen from position that deviates from the center of screen at 0.1 m, after computing, the error of measured distance is 0.05 mm. The error of measured distance coming from height difference of sky screen is 2.5×10^{-4} mm.

When human operator place sky screen, the screen do not plumb strictly up direction of object flying. There is an angle between screen and direction of object flying in azimuth. This angle is less than 1°. At the same time, the screen do not plumb strictly up earth. A water level will be installed the shell of sky screen. The error of water level is 30″. If water level departs from center 1/2 scale, the angle between screen and earth is 0.0042°. After computing, the error of distance measurement is 0.3 mm [9].

The velocity of flying object is from 50 to 1300 m/s, the precision of velocity measurement is shown in Fig. 27.5.

Fig. 27.5 Velocity measurement precision after improving



From Fig. 27.5, after using new methods of time interval measurement, distance measurement and placement, the precision of velocity measurement satisfies the requirement of testing at the range from 50 to 1,300 m/s.

27.6 Conclusions

Time interval measurement methods were discussed in this paper. Enhancing error of time interval measurement diminishes precision of velocity measurement. Using laser telemeter for distance measurement, error of operator and method caused by the use of tapeline will be avoided in distance measurement, such as error of indication, temperature error, pull change error and measurement status error. By installing laser collimator on sky screen, the error coming from unparallel and unlevel screen is decreased. After computing, the precision of velocity measurement is less than 0.1 %. Those methods are controllable, simple, convenient and independent of human operator. Those methods can be popularized in other velocity measurement field.

References

1. Compiling Group of Handbook of Mathematics (1979) Handbook of mathematics. Higher Education Press, Beijing
2. Liu HY (2006) Design and achieved for education system of time interval measurement. Doctoral thesis of university of electronic science and technology of China, Chengdu, pp 6–9
3. Hu JH, Gao X, Yin DS (2005) Correction of errors caused by temperature, pull and measurement state in actual measurement with steel tapeline. Tianjin Sci Technol 2:50–51
4. Dong T, Li H, Ma SL (2008) Improved distance measurement of wide-angle sky screen. J Xi'an Inst Technol 28(3):212–223
5. Liu L, Li SJ, Shao DR (2005) Time interval measuring instrument based on FPGA with vernier method. Telemetry Telecontrol 26(3):63–67

6. Zhu HB (1997) Time interval measuring instrument based on vernier method. J East China Shipbuilding Inst 11(2):90–97
7. Yang J (2012) Research and application of the precise time interval measurement techniques based on the TDCs and FPGA. Doctoral thesis of Chinese academy of sciences, pp 2–5
8. Kortmann T, Reiff EC (2003) Exactitude position—a widely used laser telemeter. Maschinen Markt 25:46–47
9. An Y, Wang TL (1997) The wide angle sky screen accuracy analysis. J Xi'an Inst Technol 17 (1):46–49

Chapter 28
The Integration Algorithm of Space-Based and Ground-Based Tracking Data for Measurement Element

Yan Ma and Jingwen Xie

Abstract In this article, an orbit determination method is introduced which is based on the integration of space-based and ground-based tracking data and the spline representation of trajectory. The feasibility of spline representation of trajectory is introduced and proved in the first; Subsequently, the problem of measurement element optimization and data weighting are solved, and the effect of measurement element on parameter estimation in linear model is analyzed and evaluated in detail; Finally theoretically analysis proves that the measuring element weighting can improve the parameters estimation precision in consideration of measuring element linearity and correlation.

Keywords Measurement data · Orbit determination via velocity measurement · Data integration

28.1 Introduction

With the rapid development of space navigation technology especially satellite navigation technology applied in the field of space tracking and control, data integration technology which makes full use of the space-based and ground-based tracking data, is one of the most effective ways to realize trajectory measurement with high precision. This article utilizes spline function as mathematical tool to parameterize the estimated functions, and subsequently convert the problem of measurement data processing into the problem of parameter estimation of linear or nonlinear regression analysis model, finally settle the problem of data integration in the processing of measurement data using the research achievements of modern regression analysis.

Y. Ma (✉) · J. Xie
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: mayan888@sina.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_28

From the viewpoint of mathematics, the measurement value of the different sensors construct different measuring subspace, and data integration is a projection from the measurement subspace to the integration data space according to certain rules. The process of data integration is as follows: Unify the references time and space of radar and GPS data by time alignment and coordinate transformation; Identify and eliminate abnormal measurement element and then assign proper weight according to the random error and correlation of measurement element; Estimate the trajectory parameters based on the principle of velocity measurement and the method of nonlinear parameter estimation, realizing the orbit determination of target.

28.2 The Spline Representation of Trajectory

Weierstrass theorem of polynomial approximation [1]: Bounded continuous functions on closed interval can use algebraic polynomial approximation to arbitrary precision. Spline function is an effective tool of function approximation. Compared with polynomial, spline function approximation reduces the continuity conditions, and shows better ability to adapt to variation in data and functions. Because flight trajectory is regular and time-relevant, the method that represent flight trajectory with temporal correlation, can save a lot of estimated parameters, compared with the traditional calculation method. Generally B-spline function is used in practice, especially for the least squares problem of a large experimental data.

28.2.1 The Basic Concept of B-Spline Function

In order to avoid the ill-conditioned problem of coefficient matrix, the cubic B spline is usually applied to represented trajectory [2]. Therefore, it mainly introduces the B spline with equidistant nodes applied to the representation of smooth trajectory. The following equation is $k-1$th B spline function with equidistant nodes.

$$B_k(t) = \frac{\sum_{j=0}^{k} (-1)^j C_k^j (t + \frac{k}{2} - j)_+^{k-1}}{(k-1)!} \tag{28.1}$$

where: $(t + \dfrac{k+1}{2} - j)_+^k$ is $k-1$th truncation monomial. When $k-1=0$, we define :

$$t_+^0 = \begin{cases} 0 & t<0 \\ 0.5 & t=0 \\ 1 & t>0 \end{cases} \tag{28.2}$$

28.2.2 The Spline Representation of Trajectory

The cubic B spline is usually applied to represented smooth trajectory [3], which is shown as follows:

$$B_4(t) = \begin{cases} 0 & |t| \geq 2 \\ |t|^3/2 - t^2 + 2/3 & |t| < 1 \\ -|t|^3/6 + t^2 - 2|t| + 4/3 & else \end{cases} \tag{28.3}$$

Suppose that the parameter of trajectory on time interval [a, b] is $X(t) = (x(t), y(t), z(t), \dot{x}(t), \dot{y}(t), \dot{z}(t))$, which can be represented by cubic B spline for smooth trajectory:

$$\begin{cases} x(t) = \sum\limits_{j=-1}^{N_x+1} \beta_{x,j} B_4\left(\frac{t-\tau_{x,j}}{h_x}\right) & \dot{x}(t) = \sum\limits_{j=-1}^{N_x+1} \beta_{x,j} \dot{B}_4\left(\frac{t-\tau_{x,j}}{h_x}\right) \\ y(t) = \sum\limits_{j=-1}^{N_y+1} \beta_{y,j} B_4\left(\frac{t-\tau_{y,j}}{h_y}\right) & \dot{y}(t) = \sum\limits_{j=-1}^{N_y+1} \beta_{y,j} \dot{B}_4\left(\frac{t-\tau_{y,j}}{h_y}\right) \\ z(t) = \sum\limits_{j=-1}^{N_z+1} \beta_{z,j} B_4\left(\frac{t-\tau_{z,j}}{h_z}\right) & \dot{z}(t) = \sum\limits_{j=-1}^{N_z+1} \beta_{z,j} \dot{B}_4\left(\frac{t-\tau_{z,j}}{h_z}\right) \end{cases} \tag{28.4}$$

Equation (28.4) means a linear regression model [1], in which the trajectory parameters can be estimated after calculating the spline coefficients by least square method.

28.2.3 The Error of the Spline Representation

In order to illustrate the feasibility of trajectory expressed by spline function, XX–XX missile is taken for example. Without loss of generality, the approximation precision of smooth trajectory expressed by cubic B-spline is analyzed. In the example, time interval is on 3–61 s, the trajectory is represented by cubic B-spline with equidistant nodes, the node spacing is 2 s, then the trajectory in X direction is simulated in Fig. 28.1 as follows.

As is shown in Fig. 28.1, the error is no more than 10^{-10} when the trajectory is expressed by cubic B-spline with node spacing 2 s, and meets the demand of trajectory precision. Consequently, it is feasible to represent trajectory with spline function in theory [4].

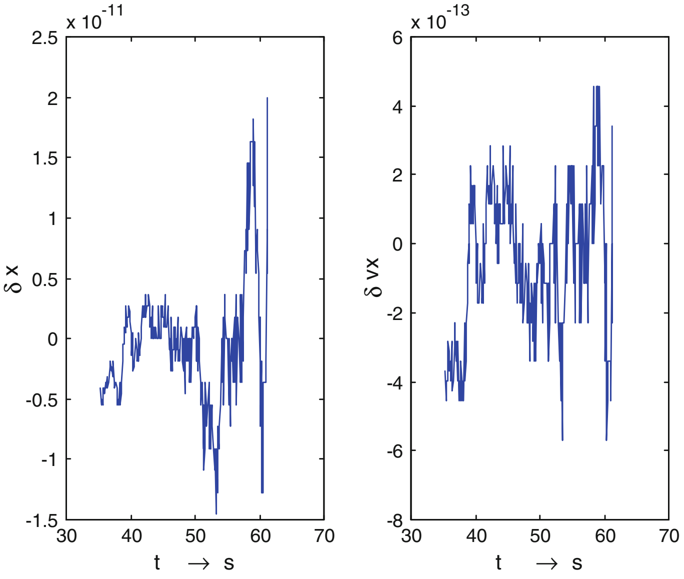


Fig. 28.1 The position and velocity errors of trajectory represented by spline in *X* direction

28.3 The Model of Data Integration Based on Measurement Element

Data integration can be simply divided into state vector integration and measurement integration (i.e., data integration based on measurement element). State vector integration combines the state vector associated to a linear estimator in order to export a trajectory estimation; Measurement integration is a more accurate estimation that calculates target state using sensor information instead of sensor estimation.

28.3.1 Observation Model

The integration of space-based and ground-based tracking data mainly indicate the acquisition of missile trajectory by the integration data of DGPS velocity measurement and ground radar velocity measurement [5].

The observation equation [6] of velocity-measuring radar is shown by Eq. (28.5):

$$\dot{S}_i = F_i(X) + v_i + \varepsilon_i \quad i = 1, 2, \ldots, M \tag{28.5}$$

where $\dot{S}_i$ is measurement data of the *i*th information source, $F_i(\cdot)$ is the function of *i*th information source on ballistic parameter X, v_i is the systematic error of the *i*th information source, ε_i is the random error of the *i*th information source.

The observation equation of navigation satellites is shown by Eq. (28.6):

$$\begin{aligned} R_j &= G_j(X) + v_j + \varepsilon_j \\ \dot{R}_j &= \dot{G}_j(X) + \widehat{v}_j + \widehat{\varepsilon}_j \end{aligned} \quad j = 1, 2, \ldots N \tag{28.6}$$

where the meanings of parameters $G, v, \varepsilon, \widehat{v}, \widehat{\varepsilon}$ are similar to Eq. (28.5).

Therefore the integration observation equation of radar and GPS is

$$Y(t) = F(X, t) + v(t) + \varepsilon(t) \tag{28.7}$$

where:

$$\begin{aligned} F(X,t) &= (F_1(X,t), F_2(X,t), \ldots, F_M(X,t), G_1(X,t), \ldots, G_N(X,t), \dot{G}_1(X,t), \\ &\qquad \ldots, \dot{G}_N(X,t))^T; \\ \varepsilon(t) &= (\varepsilon_1(t), \ldots, \varepsilon_M(t), \varepsilon_{M+1}(t), \ldots, \varepsilon_{M+N}(t), \varepsilon_{M+1}(t), \ldots, \varepsilon_{M+N}(t))^T; \\ Y(t) &= (\dot{S}_1(t), \dot{S}_2(t), \ldots, \dot{S}_M(t), R_1(t), \ldots, R_N(t), \dot{R}_1(t), \ldots, \dot{R}_N(t))^T; \\ v(t) &= (v_1(t), \ldots, v_M(t), v_{M+1}(t), \ldots, v_{M+N}(t), v_{M+1}(t), \ldots, v_{M+N}(t))^T; \end{aligned}$$

After taking Eq. (28.4) into Eq. (28.7), an observation equation simplified (28.8) is obtained.

$$Y = F(\beta, v) + \varepsilon \tag{28.8}$$

where $Y = (Y(t_1)^T, Y(t_2)^T, \ldots, Y(t_m)^T)^T$, m is sampling number;

$$\begin{aligned} v &= (v(t_1)^T, v(t_2)^T, \ldots, v(t_m)^T)^T; \\ F(\beta) &= (F(\beta, t_1)^T, F(\beta, t_2)^T, \ldots, F(\beta, t_m)^T)^T; \\ F(\beta, t_i) &= (F_1(\beta, t_i), F_2(\beta, t_i), \ldots, F_M(\beta, t_i), G_1(\beta, t_i), \ldots, G_N(\beta, t_i), \dot{G}_1(\beta, t_i), \\ &\qquad \ldots, \dot{G}_N(\beta, t_i))^T \; \varepsilon = (\varepsilon(t_1)^T, \varepsilon(t_2)^T, \ldots, \varepsilon(t_m)^T)^T \end{aligned}$$

28.3.2 Integration Method

Multiple-velocity orbit determination which is based on the spline function representation, estimate the spline coefficient using nonlinear iterative algorithm, and then to realize the location and velocity estimation of target, thus accomplishes the orbit determination.

The model Eq. (28.8) mentioned above and the spline representation of trajectory (28.4) indicate: the data measuring velocity of radar and GPS contain common unknown parameters $X = (x(t), y(t), z(t), \dot{x}(t), \dot{y}(t), \dot{z}(t))$, that's to say, there are common parameters to be estimated in different measurement data. Therefore, the joint estimation of spline coefficients using GPS and radar data can realize the integration process of space-based and ground-based measurement data. The integration process is illustrated in the Fig. 28.2.

Consequently it is impossible to solve the nonlinear regression problem of model Eq. (28.8). In practice, nonlinear iterative algorithms, such as Gauss-Newton

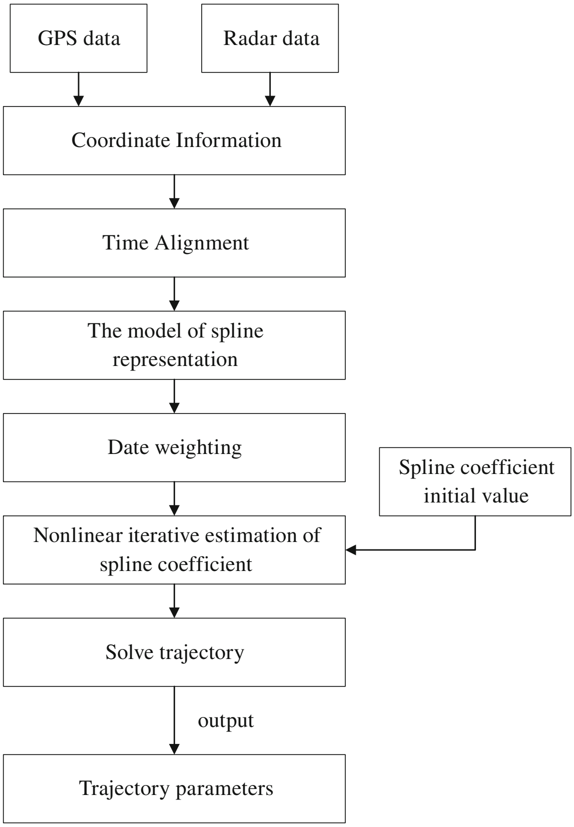


Fig. 28.2 The integration process of measurement data

algorithm, Marquardt algorithm, Hartley algorithm, can be applied to estimate trajectory parameters to realize the orbit determination of space target.

28.4 Measurement Element Optimization

The data integration of off line processing is to decide how to make use of data from different information sources in the condition of a mass of redundancy. The first question is element optimization in data integration. In this passage, we recognize the gross error of certain information sources mainly by mutually checking among the measurement elements [7].

Generally the model of measurement element information is presented as

$$S_k = G_k(\beta) + v_k + \varepsilon_k, \quad k = 1, 2, \ldots, n \tag{28.9}$$

Suppose $G_k(\beta)$ is linear function, that is $G_k(\beta) = G_k\beta$, G_k is matrix, Eq. (28.9) is the linear model of measurement element. Integrate all the information

$$S = G\beta + v + \varepsilon \tag{28.10}$$

where $S = (S_1^T, \ldots, S_K^T)^T$, $G = (G_1^T, \ldots, G_K^T)^T$, $v = (\delta_1 e_1^T, \ldots, \delta_K e_K^T)^T$, $\varepsilon = (\varepsilon_1^T, \ldots, \varepsilon_K^T)^T$.

δ_k is an independent random variable as well as independent of ε_k. $P(\delta_k = 1) = P_k, P(\delta_k = 0) = 1 - P_k$, Usually P_k is a relative small positive number and, $\varepsilon_k$is error variable with mean zero.

From Eq. (28.10) the least square estimation of β is shown by

$$\hat{\beta} = (G^T G)^{-1} G^T S = \beta + (G^T G)^{-1} G^T (v + \varepsilon) \tag{28.11}$$

Suppose that there is abiogeny system error in measurement element K, $c_K I^T$, in which $I^T = (1, \ldots, 1)$. Denote as $G(K) = (G_1^T, \ldots, G_{K-1}^T)^T$, that's $G = \begin{bmatrix} G(K) \\ G_K \end{bmatrix}$, S (K), v(K), $\varepsilon(K)$ and $G(K)$ own similar meaning. From Eq. (28.9), we could get the model when not using measurement element K:

$$S(K) = G(K)\beta + v(K) + \varepsilon(K) \tag{28.12}$$

$\hat{\beta}(K)$ is the least square estimation of β, thus:

$$\hat{\beta}(K) = \hat{\beta} - AG_K^T S_K + AG_K^T (I - G_K A G_K^T) G_K A G^T(K) S(K) \tag{28.13}$$

In which $A = (G^T G)^{-1}$, $\hat{\beta}$ is Eq. (28.11).

Formula Eq. (28.13) indicates, the different estimations $\hat{\beta}$ and $\hat{\beta}(K)$ of β can be presented as the last term of Eq. (28.13). Suppose v is not random,

$$v = (v^T(K), v_K^T)^T = (0^T, c_K I^T)^T \tag{28.14}$$

where $I^T = (1, \ldots, 1)$, thus:

If the system error in model is formula (28.14), The average mean square error $E\|e(K)\|^2/(n(K-1) - 3m)$ when element K isn't used is $c_K^2 \sum_{i,j=1}^n d_{i,j}^{(K)}/(nK - 3m)$ smaller than $E\|e\|^2/(nK - 3m)$ when K is used, in which $d_{i,j}^{(K)}$ is the submatrix of $(I - GAG^T)$.

Without using measurement element K, the residual error vector is:

$$e(K) = S(K) - G(K)\hat{\beta}(K) = (I - G(K)(G^T(K)G(K))^{-1} G^T(K))S(K)$$

where residual error is $e_i = S_i - G_i\beta$, Residual Sum of Squares:

$$RSS = \sum_{i=1}^{m} e_i^2$$

From the analysis above,the abiogeny system error in measurement element can reduce error of parameter estimation. Model (28.10) [or (28.12)] can descript the measurement data when abiogeny system error is not clear.

The abiogeny system error in measurement element can reduce the precision of trajectory parameters estimation. Model (28.10) [or (28.12)] fits well when it's not clear for abiogeny system error. The problem of judging for abiogeny system error converts to the following hypothesis testing [3, 7].

H_0: the system error of cell K is $c_K\delta_K$ (system error occurs at little P_K);
H_1: the system error of cell K is $c_K 1$ (system error c_K occurs at 1).

Compare $\overset{\wedge}{\beta}$ and $\overset{\wedge}{\beta}(K)$ as follows:

$$E(\hat{\beta}) = \beta + (G^T G)^{-1} G^T \cdot E(v) = \beta + (G^T G)^{-1} G^T (P_1 e_1, \ldots, P_K e_K)^T \tag{28.15}$$

Similar:

$$\begin{aligned} E(\hat{\beta}(K)) &= \beta + (G^T(K)G(K))^{-1} G^T(K)(e_1 P_1, \ldots, e_{K-1} P_{K-1})^T \\ &= \beta + [(G^T G)^{-1} + (G^T G)^{-1} G_K^T (I - G_K (G^T G)^{-1} G_K^T)^{-1} G_K (G^T G)^{-1} \\ &\qquad G^T (e_1 P_1, \ldots, e_{K-1} P_{K-1}, 0)^T \end{aligned} \tag{28.16}$$

Thus:

$$\begin{aligned} E(\hat{\beta} - \hat{\beta}(K)) = {} & (G^T G)^{-1} G^T (0, \ldots, e_K P_K)^T \\ & - (G^T G)^{-1} G_K^T (I - G_K (G^T G)^{-1} G_K^T)^{-1} G_K (G^T G)^{-1} \\ & \qquad G^T (e_1 P_1, \ldots, e_{K-1} P_{K-1}, 0)^T \end{aligned} \tag{28.17}$$

when H_0 is true, and $P_i (i = 1, 2, \ldots, K)$ is small, formula (28.17) converses at 0. The difference between $\hat{\beta}$ and $\hat{\beta}(K)$ is small. When H_1 is true, there is constant system error c_K for K, and the gain of the first term in (28.17) increases with c_K, so the modulus of $E(\hat{\beta} - \hat{\beta}(K))$ is. By this method can we easily decide which element is not going to be used.

When H_0 is true, the mean difference of trajectory parameters estimation between using K and not is $E[B(t)(\hat{\beta} - \hat{\beta}(K))]$, where $B(t)$ is spline base matrix. Thus, if $E[B(t)(\hat{\beta} - \hat{\beta}(K))]$ is large enough, the corresponding element K should be removed.

28.5 The Weight of Measurement Element Coefficient

Measurement data contains valuable signal, system error and random error. As different element has different random value, they make different contribution to the precision of orbit determination in trajectory calculation. We can determine the weight allocation of each elements in data integration by their random error characteristics.

In the institution of multiple velocity measurement, we normally use one transmitting station with several receiving station, so at the same time, the gradient of the sum distance among one transmitting station and related receiving stations is correlative. Besides, correlative matrix of random error vector is not diagonal matrix. The random error characteristics of GPS data is similar to that of radar data, but they are not correlative. The correlative matrix of random error of the sun of distance can be calculated as following:

$$D_{\dot{S}} = \operatorname{cov}[\varepsilon_1(t_i), \varepsilon_2(t_i), \ldots, \varepsilon_M(t_i)]$$

Similar:

$$D_{\dot{R}} = \operatorname{cov}[\varepsilon_{M+1}(t_i), \varepsilon_{M+2}(t_i) \ldots, \varepsilon_{M+N}(t_i)]$$

Therefore, the correlative matrix of random error in integration model at time t_i is $K_i = \operatorname{cov}[\varepsilon(t_i)] = diag(D_{\dot{S}}, D_{\dot{R}})$, $i = 1, 2, \ldots, n$ is sampling number.

$$D_{\dot{S}} = diag((1 + b_1 t_i)^2 \sigma_1^2, \ldots, (1 + b_m t_i)^2 \sigma_m^2) + (1 + b_T t_i)^2 \sigma_T^2 \begin{bmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \ldots & 1 \end{bmatrix} \tag{28.18}$$

The first and second terms in the right hand are positive definite, the last term is positive semidefinite. Consequently $D_{\dot{S}}$ is positive definite. Suppose the error of measurement element is Gaussian noise, when solve the parameters D_R and $D_{\dot{R}}$. Therefore, D_R and $D_{\dot{R}}$ are both diagonal matrix.

So, we get K_i, set

$$\tilde{Y}_i = K_i^{-\frac{1}{2}} Y_i,\ \tilde{F}_i = K_i^{-\frac{1}{2}} F_i,\ \tilde{\varepsilon}(t_i) = K_i^{-\frac{1}{2}} \varepsilon(t_i)$$

From formula (28.8), we get

$$\tilde{Y} = \tilde{F}(\beta, v) + \tilde{e}(t_i),\ \tilde{e}(t_i) \sim (0, I_{(m+2n)N}) \tag{28.19}$$

Obviously, formula (28.19) is the observation equation of system after the measurement element is weighted, indicating that the random error of every element is not effective on finally result because of their variation.

From Guass-Markoff theorem, for every constant vector C, $C^T\beta$ is the estimation with least error among all the linear unbiased estimation class of $C^T\beta$, where $\hat{\beta} = (F^T K^{-1} F)^{-1} F^T K^{-1} S$, $\hat{\beta}$ is the estimation with weight of $\beta . \tilde{\beta} = (F^T F)^{-1} F^T S$ is the estimation without weight of β.

From above formula:

$$\begin{aligned} E\hat{\beta} &= \left(\sum_{i=1}^{m} K_i^{-1} \right)^{-1} \sum_{i=1}^{m} (K_i^{-1} S_i) = \beta \\ E\left| \hat{\beta} - \beta \right|^2 &= \left(\sum_{i=1}^{m} K_i^{-1} \right)^{-1} \end{aligned} \tag{28.20}$$

The same as:

$$\begin{aligned} E\tilde{\beta} &= \beta \\ E\left|\tilde{\beta}-\beta\right|^2 &= \frac{K_1+K_2+\cdots+K_m}{m^2} \end{aligned} \tag{28.21}$$

From Holder's inequality:

$$m^2 = \left[\sum_{i=1}^{m}\sqrt{K_i}\frac{1}{\sqrt{K_i}}\right]^2 \le \left(\sum_{i=1}^{m}K_i\right)\left(\sum_{i=1}^{m}\frac{1}{K_i}\right)$$

That is

$$\left(\sum_{i=1}^{m}K_i\right)\frac{1}{m^2} \ge \left(\sum_{i=1}^{m}\frac{1}{K_i}\right)^{-1}$$

Thus

$$E\left|\tilde{\beta}-\beta\right|^2 \ge E\left|\hat{\beta}-\beta\right|^2 \tag{28.22}$$

Therefore, the precision of estimated parameters is better when we using weighted average of measurement elements than not. Besides, the trajectory parameters are linear function of spline coefficients, the precision of trajectory parameters estimated are getting better.

28.6 Simulation Analysis

In order to verify the feasibility and validity of obit determination based on data integration of GPS and radar, the trajectory of a tested missile is simulated. In simulation 1, there are 6 distances with gradients of ground-based data and 2 velocity of GPS, and the system error is modeled to solve trajectory, in which the system error of R4# contains accidental error that cannot be modeled, and measurement element weighting is not introduced into obit determination. In simulation 2, the measurement elements containing accidental error are removed only. In simulation 3, measurement element weighting is introduced as well as removing the measurement elements containing accidental error. All the system and random errors are set equally. The precision of location and velocity of target in X direction are illustrated in Figs. 28.3 and 28.4.

The simulation indicates that: (a) it is feasible to fuse the radar and GPS tracking data to realize orbit determination. (b) the measurement element with the system

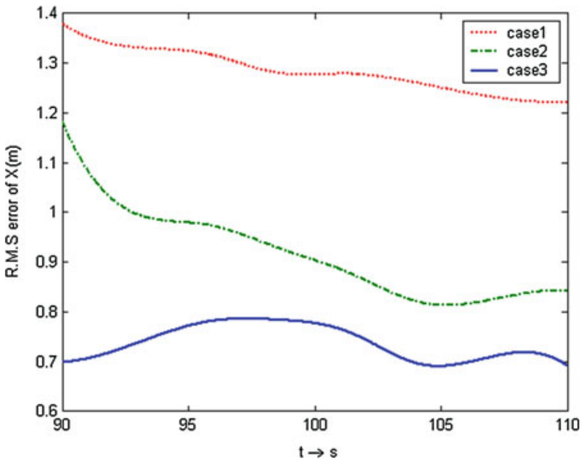


Fig. 28.3 Location precision of trajectory in *X* direction

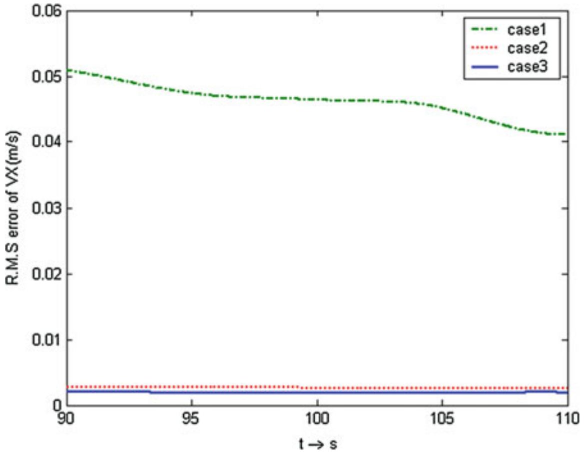


Fig. 28.4 Speed precision of trajectory in *X* direction

error that cannot be modeled can reduce the trajectory error and it is necessary to optimize the measurement element. (c) On the basis of optimization in measurement element, measurement element weighting according to their statistical property can improve the precision of obit determination.

28.7 Conclusion

There are two advantages of the integration of GPS data and radar data in trajectory calculation. The one is the full usage of space-based and ground-based measurement information, which increases the redundancy of measurement data and ensures the high reliability and high precision of measurement and control system.

The other is that, the data integration of space-based measurement improves the gradient matrix $(\nabla F(\beta))$ structure of nonlinear iterative algorithm in parameter estimation.

References

1. Wang ZM, Yi DY (1999) The calibration and evaluation of trajectory tracking data. National University of Defense Technology Press, Changsha
2. Guo JH (2000) The smoothing method of observed data based on optimal nodes spline approximation. Space Science and Technology of China (3)
3. Wang ZM, Zhou HY (1996) Mathematical processing of range and range rate tracking data, AD-A310796
4. Brown DC (1964) The error model best estimation of trajectory. AD602799
5. Li YS, Qi DX (1979) Spline function methods. Science Press, Beijing
6. Wei BC (1989) Modern linear regression analysis. Southeast University Press, Nanjing
7. Wang ZM, Yi DY (1996) The model of observed data and parameter estimation. National University of Defense Technology Press, Changsha

Chapter 29
Comparative Research on Models of Refractive Index in the Troposphere over Xiamen

Xiaodong Ding, Lusha Wang, Hailong Yan, Hongxing Zhu and Yanhua Lu

Abstract Considering the fact that refraction effect of radio wave in atmosphere becomes a main error source of data quality from ground stations. Using the previous ground data and actual meteorological sounding data of Xiamen region, it builds a vertical distribution model of atmospheric parameters and calculates height vertical distribution data of atmospheric refraction in this region. Then it analyzes the characteristics of changes of meteorological factors in troposphere over this region and their effect on the refractive index. To be specific, it compares the forecast accuracy of three models including ITR-U Model, segmented model, and Hopfield Model with the measured atmospheric refractive index data in this region. The forecast accuracy of different refractive index models applied in the same region is comparatively analyzed. Finally, the effect of the models is measured by using the root-mean-square error so as to denote the useable ranges of three models, which shed a certain light on studying measure-control precision of TT&C systems in China.

Keywords Atmospheric parameters · Troposphere · Refraction index model

29.1 Introduction

With the development of aerospace experiment techniques, requirements for satellite launching and high precision orbit tracking are increasingly demanding. Therefore, it is necessary to correct atmospheric refractive error which is one of the main sources of error affecting the accuracy of ground-based monitoring system.

X. Ding (✉)
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: dingxd04@163.com

X. Ding · L. Wang · H. Yan · H. Zhu · Y. Lu
Xi'an Satellite Control Center, Xi'an 710043, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_29

Currently, the main methods of atmospheric refraction error correction can be divided into the following categories [1, 2].

1. Ray tracing method [3, 4]. This method is based on the assumption of spherical atmosphere, starting from geometrical optics formula for calculating each refractive error, which not only adapts to a variety of elevation but also has higher precision, so this method has been widely used in practice. However, due to its large amount of calculation and the need to provide real-time measurements of the refractive index of the data path, this method is mainly used for post-processing and data validation.
2. Radiometer method [5]. Such methods can achieve high-precision real-time correction of refractive error, but it has limited application because of its high cost.
3. Simplified correction method [1, 6–8]. It is characterized by fast calculation, but lower precision, and generally used for real-time correction, such as the equivalent radius of the Earth, linear regression method, empirical correction method, structural modeling, and so on.

Regardless of any correction method, the precondition is obtaining atmospheric refractive index distribution as most accurately as possible. At present, the vertical structure of the refractive index is obtained from the refractive index model of atmospheric structure, which is accomplished generally through the use of sounding instruments to measure atmospheric parameters. The actual atmospheric structure is complex and changeable, and it changes in a certain order, such as seasonal variations, it also changes randomly such as irregular diurnal variation, and change significantly with regional variations. So it is difficult to obtain an accurate model of the atmospheric structure, which largely limits the accuracy of error correction of refractive index. Based on the characteristics of the coastal atmosphere in China, empirical model of the atmosphere is established in Xiamen field area by using sounding environmental data in recent years. This research studied the changes of meteorological factors in the field area troposphere and their impact on the refractive index, and made comparative analysis of the forecast accuracy of different refractive index model, then provided the applicable conditions for different models, which provides a theoretical basis for refractive error correction method.

29.2 Atmospheric Refractive Index

Radio wave propagation velocity in the medium is affected by the magnetic permeability and permittivity of the medium, and it is generally less than velocity of light. c represents the speed of light in vacuum, and define the ratio of c to v (radio wave propagation velocity in the medium) as refractive index of air, denoted by n. Atmospheric refraction index n is close to 1, for ease of analysis and calculation, usually the refractive index N is defined to represent the refraction index

$$N = (n - 1) \times 10^6 \tag{29.1}$$

From molecular physics knowledge, the refractive index N can be expressed as the function atmospheric state parameters (dry air pressure P_d, vapor pressure P_e and temperature T)

$$N = a\frac{P_d}{T} + b\frac{P_e}{T} + c\frac{P_e}{T^2} \tag{29.2}$$

Currently, international use of coefficient a, b, c generally adopts the measurement results by Smith and Weintraub [9]: $a = 77.6$, $b = 72$, $c = 3.75 \times 10^5$, put into formula (29.2):

$$N = 77.6\frac{P_d}{T} + 72\frac{P_e}{T} + 3.75 \times 10^5\frac{P_e}{T^2} \approx 77.6\frac{P}{T} + 3.73 \times 10^5\frac{P_e}{T^2} \tag{29.3}$$

where $P = P_d + P_e$.

29.3 Atmospheric Environmental Characteristics of Xiamen

Usually, only surface meteorological data is used by the monitoring of atmospheric refraction correction. However, due to the effects of climate and seasonal changes and other factors, there is a large difference between this model and the actual three-dimensional distribution, resulting in large deviation on atmospheric refraction error correction, which is difficult to meet the increasing demand for precision orbit determination. Accurate atmospheric refractive index profile is obtained by converting the meteorological data from radiosonde observation. It takes about 1 h for radiosonde observation to detect one profile data. For one day, 2 detections are done at 0:00 (Universal Time, the same below) and 12:00 [2].

29.3.1 Average Height Distribution of Atmospheric Parameters

By analyzing atmospheric parameters from original data of sounding test in Xiamen in recent years, excluding outliers from these tens of thousands of sets of data, an average of 12 months of meteorological data is obtained using the monthly average method. For each month, data are selected at 0:00 and 12:00 respectively, finally getting a total of 24 sets of data. Then monthly polynomial fitting formula of atmospheric pressure and temperature is obtained by using polynomial fitting method. For example, at 12:00 in January, the formulas (29.4) and (29.5) are the

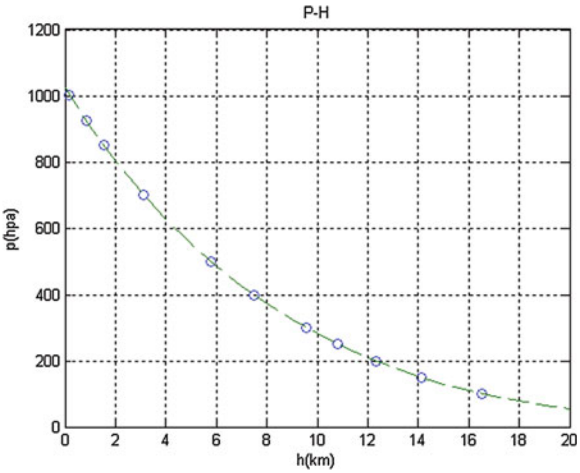


Fig. 29.1 Trend of pressure change with height at 12:00 UTC, January

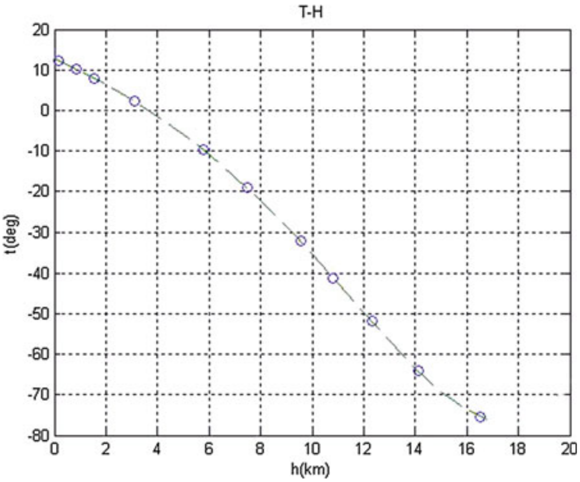


Fig. 29.2 Trend of temperature change with height at 12:00 UTC, January

atmospheric parameters model of January for Xiamen, Figs. 29.1 and 29.2 shows trend of pressure and temperature change with height in Xiamen at 12:00 in January.

$$\begin{aligned} p = {} & -6.164 \times 10^{-5} h^5 - 5.582 \times 10^{-3} h^4 \\ & - 0.234 h^3 + 6.609 h^2 - 121.6 h + 1021.5 \end{aligned} \tag{29.4}$$

$$\begin{aligned} t = {} & 6.654 \times 10^{-5} h^5 - 1.034 \times 10^{-3} h^4 \\ & - 1.863 \times 10^{-3} h^3 - 0.132 h^2 \\ & - 2.947 h + 12.828 \end{aligned} \tag{29.5}$$

29.3.2 Vertical Distribution of Atmospheric Refractive Index

Using atmospheric parameters measured data to obtain polynomial fitting formula in Xiamen, the vertical structure of atmospheric features of the measured refractive index profile is obtained according to the Eq. (29.3).

As can be seen from Fig. 29.3, with the change of different atmospheric environment characteristics, atmospheric vertical refractive index profile shows a large fluctuation in the lower level 3 km or less, especially most significant about 1 km. When analyzing this variation, it is found that in the main meteorological factors affecting the change in refractive index, the temperature variation with height fluctuations is relatively stable over time (see Fig. 29.4), but the magnitude of the vapor pressure shows big changes (see Fig. 29.5), especially in lower level the

Fig. 29.3 Years measured vertical atmospheric refractive index distributions in Xiamen

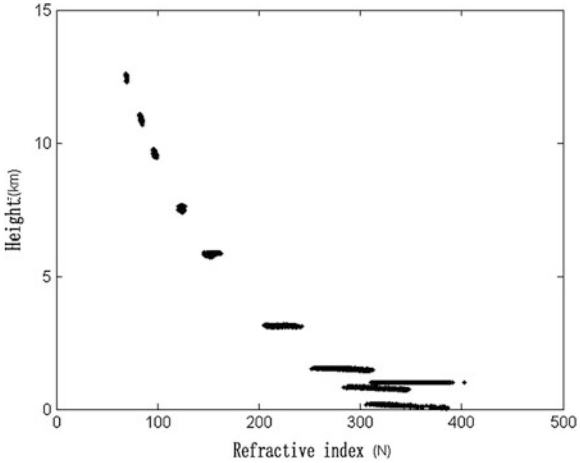
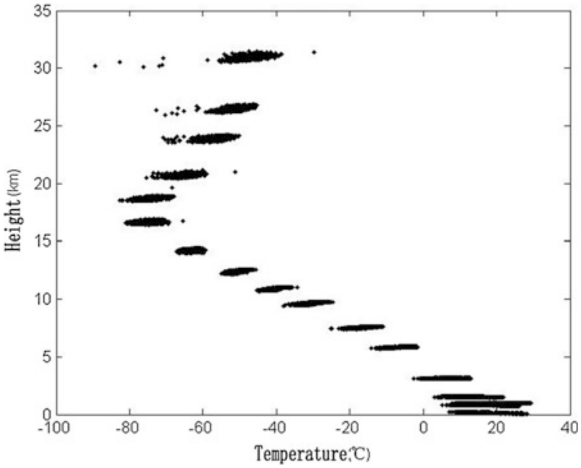


Fig. 29.4 Years measured vertical distribution of temperature in Xiamen



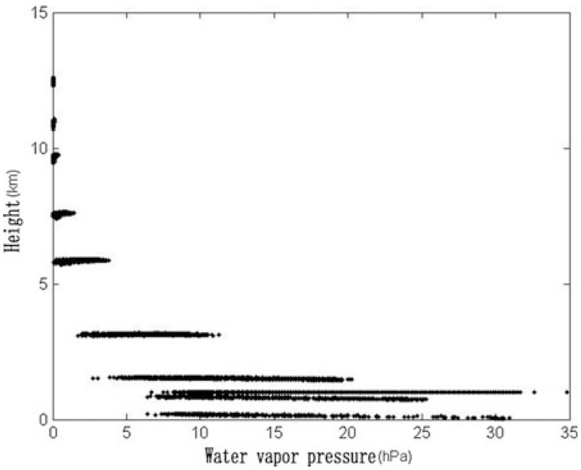


Fig. 29.5 Years measured vertical distribution of water vapor pressure in Xiamen

amplitude of variation varies from less than 10 hPa to over 30 hPa. With height increasing, the fluctuation characteristics of amplitude of variation significantly reduce, namely the higher it is, the smaller of the amplitude of variation of water vapor pressure. It is thus that changes of water vapor are the most difficult part for atmospheric refraction correction. Figure 29.5 also shows that the vapor pressure is almost zero when the height is more than 10 km. Figure 29.3 shows that the refractive index has a maximum area in the vicinity of 1 km, at the same time, there is a inversion layer in vertical distribution of temperature, therefore the change of refractive index is not only affected by water vapor, but also by temperature.

29.4 Comparison of Refractive Index Profile Models

The above vertical distributions of meteorological parameters facilitate understanding of atmospheric inversion precision refractive index correction. Here 3 high-precision refractive index models which currently commonly used are analyzed.

29.4.1 Refractive Index Model

29.4.1.1 ITU-R Exponential Model

ITU-R P. 834-6 suggested [11] a general exponential model (hereinafter referred to as ITU-R model), in which the rate of exponential decay is described as reciprocal of the feature height H which is correlated to zenith delay.

$$N(h) = N_0 \exp\left(-\frac{h - h_0}{H}\right) \tag{29.6}$$

where H is feature height; ΔS_v is zenith delay, $\Delta S_v = 0.00227 \times P_0 + f(t_0)$ RH0; P_0 for ground pressure; RH_0 for ground relative humidity. The Relationship between $f(t_0)$ with the ground temperature t_0 is, $f(t_0) = a \times 10^{bt_0}$, in coastal areas a takes value 5.5×10^{-4}, b takes value 2.91×10^{-2}.

29.4.1.2 Segmentation Model

Through statistical analysis on atmospheric refraction index by a large number of measured data, it is found that the atmospheric refractive index nearly 1 km below the ground level changes linearly approximately, while the refractive index at an altitude of 9 km is relatively stable throughout the year, and for 1–9 km and 9–60 km, it can be estimated with negative exponential model so as to get segmentation model.

$$N(h) = \begin{cases} \mathrm{N_0} - \Delta\mathrm{N_1}(h - \mathrm{h_0}) & \mathrm{h_0} \le h \le \mathrm{h_0} + 1 \\ \mathrm{N_1}\exp[-\mathrm{c_{a1}}(h - \mathrm{h_0} - 1)] & \mathrm{h_0} + 1 < h \le 9 \\ \mathrm{N_9}\exp[-\mathrm{c_{a9}}(h - 9)] & 9 < h \le 60 \end{cases} \tag{29.7}$$

where ΔN_1 stands for refractive index gradient at 1 km from the ground; N_1 stands for the refractive index at 1 km from the ground; c_{a1} represents decaying exponential rate at an altitude from 1 to 9 km from the earth; N_9 represents refractive index at an altitude of 9 km; c_{a9} denotes decaying exponential rate at an altitude between 9 and 60 km from the ground.

To ensure the continuity of the obtained cross-sectional view at the segment points, in actual calculation, the profile parameters of segment above 9 km N_9 and c_{a9} take the lead in statistics. Then N_1 c_{a1} N_0 and ΔN_1 can be obtained under the conditions of the continuousness of 9 and 1 km from the ground. Because the refractive index profile at 9 km has been relatively stable, this has been proved to improve forecast accuracy of refractive index profile [12].

29.4.1.3 Hopfield Model

Hopfield model [13], also known as double refraction index of the fourth power model. Different from the several previous models, its static parts are based on the theory of atmospheric statics. Hopfield static entry model is based on the characteristic that the temperature of the troposphere change with height at a certain gradient attenuation. A theoretical model, derived from the atmospheric statics equations, can be used to describe the refractive index distribution throughout the neutral atmosphere as following:

$$N_i(h) = \begin{cases} N(h) = \mathrm{N_d} + \mathrm{N_w}(h) & \\ \mathrm{N}_{i0}\left(\frac{H_j - h}{H_i - \mathrm{h}_0}\right)^4 & i = \mathrm{d,w}\, h \leq H_i \\ \mathrm{N}_i = 0 & i = \mathrm{d,w}\, h \geq H_i \end{cases} \quad (29.8)$$

where H_d is the equivalent height of static items, $H_\mathrm{d} = 40.136 + 0.148\,72 \times t_0$; H_w is the equivalent height of the wet items, take value of 11 km.

29.4.2 Error Assessment Method

In error analysis, air sounding data as a real profile of the atmosphere is used to compare with profiles which are calculated through a variety of cross-sectional model. Then root-mean-square error used to measure of the prediction effect of the model [10], computational formula is:

$$RMS = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(N^i_{Model} - N^i_{Radio}\right)^2} \quad (29.9)$$

where m is the point number of the profile; N^i_{Model} and N^i_{Radio} represent the refractive index values of model profile and radiosonde observation profile.

29.4.3 Comparison of the Refractive Index Models

To compare the prediction accuracy of each model, historical sounding meteorological parameters in Xiamen in recent years were put into to different refractive index model to obtain the vertical distribution of the refractive index. Due to large amount of data and long-time span, there are some missing values and unreasonable singular value weather raw data. So, before the calculation, all observational data were under quality control, judgment and exclusion of defaults and unreasonable data. As the meteorological sounding data provided meteorological data parameters on the surface of the fixed pressure, before using, the difference value should be converted to a fixed height segment. In order to facilitate the calculation and more fully reflect the forecasting e effect of refractive index of various models, the radiosonde data interpolated by the altitude, take 500 m as a step-length for calculation. Since the atmospheric structure above the troposphere has been relatively stable and f forecast accuracy gap of each model in the troposphere has been very evident, here vertical altitude just takes the value up to 12 km.

In this article, ITU-R model, Hopfield model and segmentation model were compared with measured atmospheric refractive index profile. Take the measured date in May 2007 as an example, as shown in Fig. 29.6:

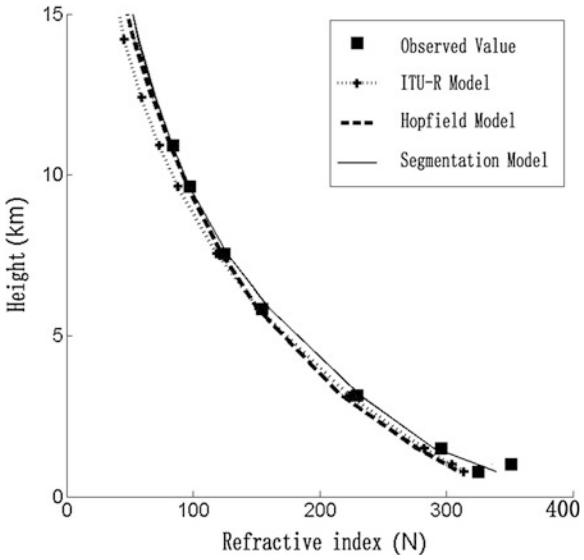


Fig. 29.6 Vertical distributions of predicted values of different Models in Xiamen

1. In the overall trend, each statistical model can fit and predict the vertical distribution of atmospheric refraction quite well, but Hopfield model and segmentation model are relatively more effective than ITU-R model, especially at the height of above 5 km; compare Eq. (29.6), it can be seen that the vertical distribution of ITU-R model is only associated with atmospheric parameters on the ground, namely when the ground atmospheric parameters input cannot better represent the average state of the ground in Xiamen, error of model is larger. In Fig. 29.6, ITU-R model is generally smaller than the actual measured value at the same height, which denotes that it is difficult to select appropriate ground atmospheric parameters for the model.
2. There are varying differences of deviation between different refractive index models at different height ranges. Errors of Hopfield model and Segmentation model are relatively small at the area of 5 km or above, with Hopfield model better. Comparative analysis of Figs. 29.3, 29.4, and 29.5, it can be clearly found that in the upper atmosphere, the troposphere has a relatively high degree of stability without significant change of vapor pressure, temperature, etc., which makes the refractive index changes less and regularly. Hopfield model, as a theoretical model based on the equations of atmospheric statics, is more suitable for upper atmosphere.
3. For the lower atmosphere, there are large changes of moisture and temperature, the complex change of refractive index, but the segmentation model can better fit the actual vertical distribution of the refractive index with smaller error.

In assessing the overall prediction accuracy of three models, this paper makes statistics of the vertical distribution of root mean square error of prediction and measured values of different models in Xiamen, and the results are shown in

Table 29.1 Vertical distribution of refractive index of the root mean square error (RMS) of different Models

Height (base ground) (km)	ITU-R model	Hopfield model	Segmentation model
1	2.193	2.397	1.288
2	1.135	0.870	0.454
3	0.374	0.276	0.488
4	0.640	0.830	0.499
5	0.420	0.542	0.433
6	0.135	0.198	0.335
7	0.158	0.130	0.228
8	0.305	0.079	0.127
9	0.390	0.068	0.031
10	0.389	0.087	0.038
11	0.478	0.411	0.121
12	1.043	1.092	0.298

Table 29.1. As can be seen from Table 29.1, RMSE of different forecasting models and measured values are relatively small, in which Hopfield model and segmentation model can better forecast the vertical distribution of the refractive index in Xiamen. The only difference is that the accuracy of these two models is inconsistent with height changes. Hopfield model can better represent the atmospheric characteristics of measured refractive index at the middle level (6–8 km), while segmentation model can better represent the atmospheric characteristics of measured refractive index at low-level and high-level. Different models are based on different theories. Hopfield model is mainly is derived from the atmosphere static equations, so it can be more effective at middle-level of the atmosphere where the atmosphere is relatively stable and better meet the statics equation. However, segmentation model is based on the accuracy needs of different levels of height to do segmentation fitting, which has wider applicability. Above conclusion is consistent with the example of Fig. 29.6.

29.5 Summary

China has vast territory, complex topography, and large differences in weather and other environmental factors, resulting in large error changes when the same model used in different regions, the error changes between various models is more dramatic. This paper, through a case study in Xiamen, analyzes variation characteristics of refractive index with height. By comparing three typical models' prediction accuracy on vertical distribution of atmospheric refractive index including ITU-R model, Hopfield model and Segmentation model, the following conclusions are made:

1. Except for water vapor, the characteristics of pressure and temperature parameters distribution with height are stable in Xiamen.
2. Vertical distribution of atmospheric refractive index has a large fluctuation at lower level below 3 km; in corresponding height range, the vapor pressure shows the most dramatic changes, temperature changes followed.
3. Segmentation model has higher forecast accuracy at lower atmosphere, Hopfield model adapts better to the upper atmosphere in the troposphere, ITU-R model uses the global average distribution model of atmospheric refractive index to do fitting and shows relatively poor accuracy.
4. In practical applications, it is not recommended to use general parameters of the model. There must be some improvements when using these models with accuracy ensured, namely the statistics of the actual region's air data should be analyzed to obtain adaptable model coefficients.

In this paper, the analysis is more emphasized on the characteristics of the vertical distribution of the refractive index on the average state of the atmosphere. However, changes in the atmosphere refractive index are affected by the meteorological elements, especially sensitive by moisture and temperature changes, future work will further study impact of seasonal variation and diurnal variation of meteorological parameters on vertical distribution of refractive index. At the same time, the study on the atmospheric refraction correction model with uneven level of multidimensional data is also an important research orientation to high precision measurement and task security.

References

1. Huang J, Zhang RF, Hu DZ (1999) Wave refraction error correction. National Defence Industry Press, Beijing, pp 12–13
2. Yan HJ (2006) Progresses in research of atmospheric refraction. World Sci-tech R&D 28 (1):48–58
3. Zhang Y, Yuan QL (2006) The method research of radiowave radial track description from radar to objective. J Henan Normal Univ (Nat Sci) 34(2):50–53
4. Jiang CY, Wang BD (2001) Atmospheric refraction corrections of radiowave propagation for airborne and satellite-borne radars. Sci China (Series E) 44(3):280–290
5. Zhang Y (2001) Compensation method of radiowave refraction correction by microwave radiometer at low-angle. J Microwaves 17(1):96–98
6. Yang L, Sui LF (2001) The correction of the tropospheric propagation delay error in GPS surveying. J Inst Surveying Mapp 18(3):182–185
7. Bean BR, Dutton EJ (1968) Radio meteorology. Dover Publications Inc, New York, pp 231–232
8. Zhang Y, Gao X (1996) Affect latency and low-altitude atmospheric bending of radio positioning. Electron Technol Inf Technol 3:87–89
9. Smith EK, Weintraub S (1953) The constants in the equation for atmospheric refractive index at radio frequencies. Proc IRE 41(8):1035–1037

10. Chen XM (2008) Studied on atmospheric refractivity profile model and radio wave refractive error correction method. Doctoral thesis of ocean university of China, Qingdao
11. ITU-R P.834-6 (2009) Effects of tropospheric refraction on radiowave propagation
12. Liu YM, Chen JS, Zhao ZW et al (2007) Regional distribution of atmospheric refraction error In: 2007 space control technology symposium proceedings in Beijing, pp 875–877
13. Hopfield HS (1969) Two-quartic tropospheric refractivity profile for correcting satellite data. J Geophys Res 74(18):4487–4499

Part II
Space Mission and Operations

Chapter 30
Study on Emergency Scheduling Model of Space and Ground TT&C Resources

Ying Shen and Jian Bai

Abstract The scheduling of space and ground TT&C resources is an optimization problem with complex constraints such as multi-time windows and long-time windows.In the actual TT&C resource management activities, the arranged TT&C network plan needs to be adjusted in emergency, such as the spacecraft in abnormal or the TT&C devices in malfunction. Based on the practical applications, the workflow of emergency scheduling of space and ground TT&C resources is designed. Taking the adjustment as scheduling decision variable, the emergency scheduling model and algorithm of TT&C resources is established corresponding to our TT&C network management mode, which minimize the change range of network plan and maximize the task satisfaction rate, and then the evaluate function is also provided. Simulation result demonstrates that the model and algorithm are feasible for emergency scheduling of TT&C resources to ensure the task completion of the whole TT&C network.

Keywords Space and ground TT&C resources · Scheduling model · In emergency · Heuristic algorithm · Greedy policy

30.1 Introduction

According to the requirements of task in emergency, the emergency scheduling of space and ground TT&C resources realizes the dynamic arrangement to the allocated TT&C resources to ensure the satisfaction rate of tasks. The study on space and ground TT&C resources scheduling is shown as follows, the 0–1 integer programming model based on the possible begin time of task and the lagrangian relaxation algorithm written by Kang [1], and model based on the integrated

---

Y. Shen (✉) · J. Bai
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: shenying_163@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_30

scheduling of space and ground TT&C resources and genetic algorithm written by Chen [2]. Till now, there has few references about the space and ground TT&C resources scheduling in emergency.

The limitation exists in the emergency scheduling by only using the ground resource. The tracking coverage for spacecrafts in middle or low earth orbit can reach 85–100 % by tracking and data relay satellite system (TDRSS). With their own special characteristics, using these two TT&C resources to realize the emergency scheduling is significant. According to management model of our space and ground TT&C network mode, the preliminary exploration of the emergency scheduling model and its workflow is provided in this paper.

30.2 Classification of Emergency Requirements

30.2.1 Emergency Requirements of Spacecrafts

The emergency requirements of spacecrafts in low, middle and high earth orbit in long time can be divided into two situations as follows:

1. Space mission emergency: Supporting in disaster or quick response by several spacecrafts joint TT&C.
2. Abnormal state of spacecrafts: Instability or unlocking state of spacecrafts caused by its own problems.

More TT&C resources need to be added to ensure the task in emergency. The ground station or TDRSS can support low earth orbit spacecraft, but middle or high earth orbit spacecrafts can only be supported by ground station. The visible arcs of middle or high earth orbit spacecrafts supported by ground station and low orbit spacecraft supported by TDRSS are long. Without special requirements, the emergency scheduling can be maximized by sliding time window in this situation.

30.2.2 Emergency Requirements of TT&C Devices

The ground TT&C resource contains land-based devices (fixed, moving or oversea devices) and ocean-based devices (tracking ship). The TT&C network plan needs to be rearranged when the devices suffer the thunderstorms or its own problems in emergency. If there is no arc to substitute by the other devices, the space TT&C resource needs to be used.

Currently our ground and space TT&C resources are controlled by different resource management center. There are special operating mode and system. How to perform the work and scheduling algorithm in an integrated ground and space TT&C is a key point to solve the emergency scheduling problem.

30.3 Workflow of Emergency Scheduling

30.3.1 Workflow of Spacecrafts in Emergency

The spacecraft emergency support is provided by ground and space TT&C resource network. After the emergency requirements offered by users, to choose space or ground TT&C resource is decided by the management center of ground TT&C resource, according to the emergency requirement and the arranged network plan. Integrated scheduling of TT&C resources is the most complicated mode compared with only using ground or space TT&C resource. Emergency scheduling function is started after receiving the requirement of emergency by the management center of ground TT&C resource. The workflow is shown in Fig. 30.1.

There are two situations in the emergency adjustment result, satisfied or not. Based on the confirmation of adjustment by users, the TT&C network plan will be updated by management center of ground TT&C resource.

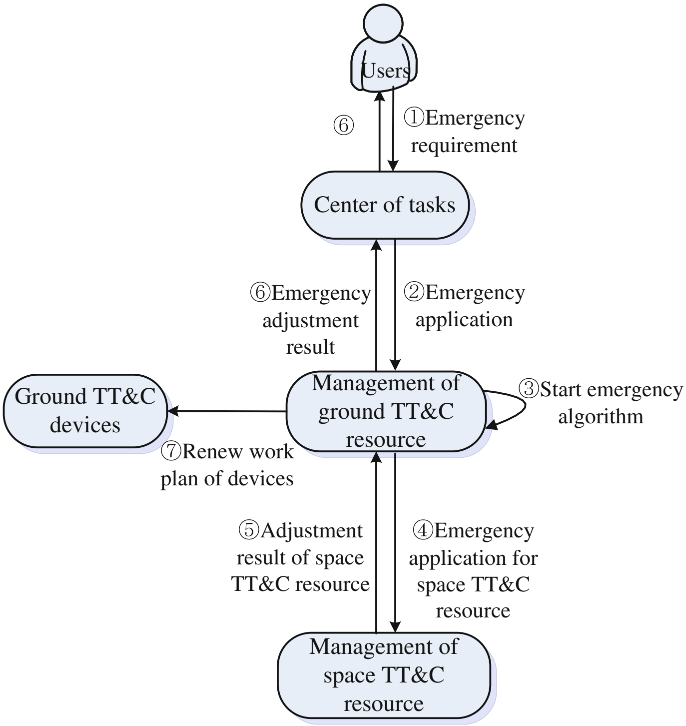


Fig. 30.1 Workflow of spacecrafts in emergency

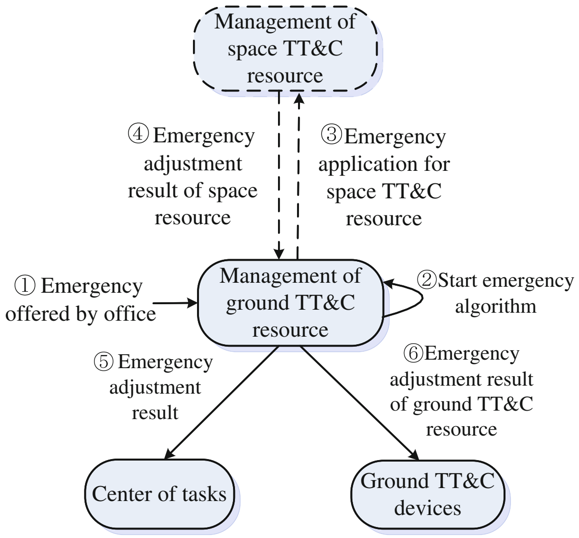


Fig. 30.2 Workflow of TT&C devices in emergency

30.3.2 Workflow of TT&C Devices in Emergency

The TT&C devices emergency is offered by office, which contains the interval that the device could not work. The emergency algorithm of ground TT&C resource is taken by management center of ground TT&C resource. If there is a set of unsatisfied tasks, the emergency algorithm [3] of space TT&C resource will be taken. We offer the requirement to the management center of space TT&C resource, and feedback the adjustment result. The workflow is shown in Fig. 30.2.

30.4 Emergency Scheduling Model

30.4.1 Principle for Emergency Scheduling

The principle for emergency scheduling design is to satisfy the emergency requirements, and maximize the normal requirements of spacecrafts and also minimize the adjustment of the arranged resources. Minimizing the adjustment of the arranged resources means the adjustment should not affect other spacecrafts as possible.

30.4.2 Scheduling Model and Algorithm

If the spacecrafts emergency require both ground and space TT&C resources, the emergency algorithm will be carried out respectively. The device emergency requirement is satisfied mainly by ground TT&C resource using the heuristic adjustment algorithm as follows.

30.4.2.1 Spacecraft Emergency

The spacecrafts in emergency needs p arcs based on its normal q arcs, and the $p + q \leq$ all the visible passes. The priority of emergency spacecrafts will be promoted. The redundant resource in the required pass is the first choice, and the transferable resource will be taken if there is no redundancy according to their different priority. The set of unsatisfied task arcs contains all the spacecrafts whose resources are occupied by the spacecraft in emergency.

30.4.2.2 TT&C Device Emergency

The redundant resource in the same pass with the malfunction device will be taken at first. If there is no redundancy, the resource in neighbor passes will be chosen. Transferable resource based on priority should be taken at last. The set of unsatisfied task arcs contains all the spacecrafts whose resources are occupied by the spacecraft in emergency.

If there remains a set of unsatisfied task arcs after using the heuristic adjustment algorithm of ground TT&C resource, the space TT&C resource emergency scheduling algorithm will be taken. The policy of greedy algorithm [4, 5] is accepted in emergency scheduling of the space TT&C resource. Based on the normal plan and the newest spare window of the relay satellite, the emergency task will be added in the scheduling queue by rearranging the plan among other tasks. Since a part of time has been occupied by the tasks in the normal task plan, the required tracking arcs will be inserted in the excess time window according to the priority of tasks. If there is no excess time window, the arranged task arcs can be moved a segment to achieve a new spare time. If not, this requirement arc will be quitted. The set of unsatisfied task arcs need to be treated in this way successively. At last, a new normal task plan and the time for emergency tasks can be received. Based on principle of value density, the relay satellite emergency task scheduling model [6, 7] is built as follows,

$$\mathrm{Max}\sum_{i=1}^{n} p(y_i) \Big/ d_i \cdot \gamma_i$$

$$\left.\begin{array}{l} Y = \{y_1, y_2 \ldots, y_n\} \\ \gamma_i \in \{0, 1\} \\ p(y_i) \in \{1, 2, \ldots, p\}, i = 1, 2, \ldots n \\ T_i \geq STW_i^{\tau}, if\ \gamma_i = 1 \\ T_i + d_i \leq ETW_i^{\tau}, if\ \gamma_i = 1 \\ T_s \leq T_i \leq T_e, \quad T_s \leq T_i + d_i \leq T_e \end{array}\right\} \tag{30.1}$$

The constraints are instructed as follows.

The number of the required arrangement tasks is described by $Y = \{y_1, y_2 \ldots, y_n\}$.

In the set of task variable $\{\gamma_1, \gamma_2, \ldots \gamma_n\}, \gamma_i = 1$ means the task y_i in emergency can be arranged in the time $\left[STW_i^\tau, ETW_i^\tau\right]$, and $\gamma_i = 0$ means not, in which $\left[STW_i^\tau, ETW_i^\tau\right]$ means a set of executed window of task y_i in emergency. The mark of corresponding time window is τ.

The priority of the task y_i in emergency is $p(y_i)$. The higher of the priority, the easier to be scheduled for this task.

Constraint of the visible part The start executed time for task y_i in emergency is T_i, and the corresponding required relay satellite service is d_i. The inequality $T_i \geq STW_i^\tau (1 \leq \tau \leq |TW_i|)$ means that the task y_i needs to be executed after the start time of visible time window $\left[STW_i^\tau, ETW_i^\tau\right]$, and $T_i + d_i \leq ETW_i^\tau$ means that the task y_i needs to be executed before the end time of visible time window $\left[STW_i^\tau, ETW_i^\tau\right]$.

Constraint of the task time The inequalities $T_s \leq T_i \leq T_e$ and $T_s \leq T_i + d_i \leq T_e$ $(i = 1, 2, \ldots n)$ mean all the tasks in emergency need to be executed in time $[T_s, T_e]$. The start and end time of the scheduling segment are described as T_s and T_e respectively.

Constraint of the conflict task The arcs arranged in emergency cannot be conflict. The objective function of task schedule $\mathrm{Max} \sum_{i=1}^{n} p(y_i) \Big/ d_i \cdot \gamma_i$ definitudes the purpose of scheduling is to ensure more emergency tasks in higher priority could be scheduled.

The flow chart of emergency scheduling of space and ground TT&C resources is designed in Fig. 30.3.

30.4.3 Evaluation of Results

The following factors will be taken into consideration for the problem of emergency scheduling.

1. Completion rate of the tasks in emergency. This is an evaluation for the completion of the tasks in emergency.
2. Satisfaction rate of the normal tasks. This is an evaluation for the completion of the normal tasks which affected by the tasks in emergency.
3. Changing degree of the whole tasks. Mainly taking the tracking arcs into consideration, the ratio of the sum of original arcs and new arcs, except the added arcs of the emergency tasks, can be defined as follows,

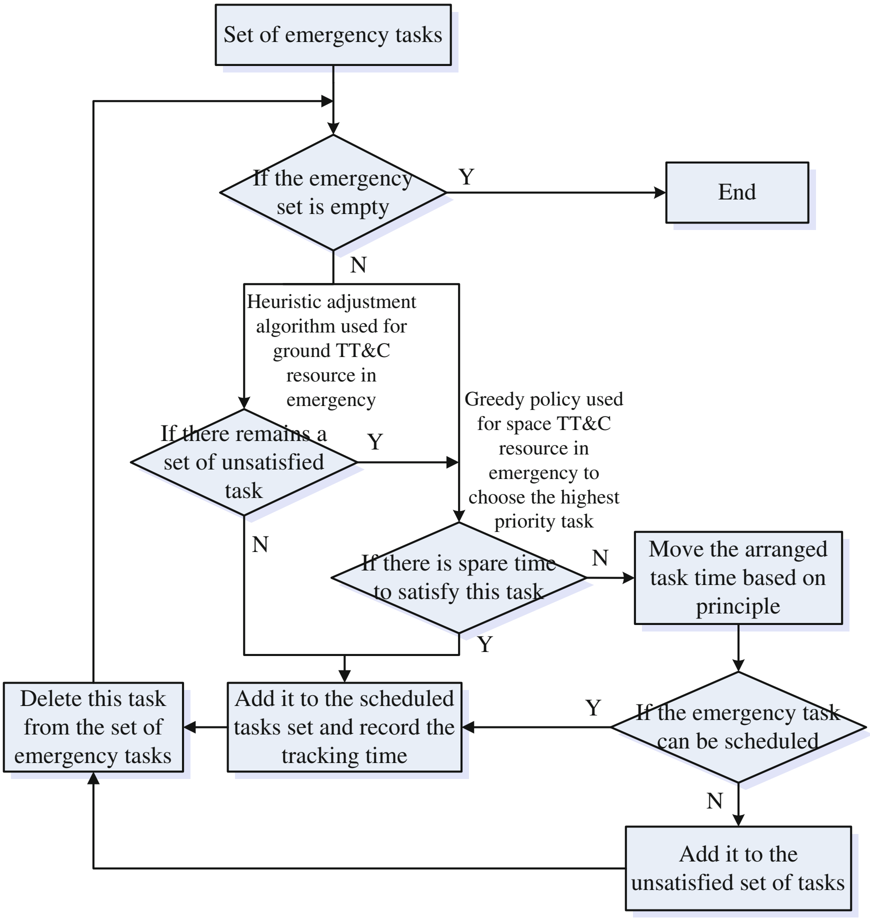


Fig. 30.3 Flow chart of emergency scheduling of space and ground TT&C resources

$$L_c = \sum_{h=1}^{H'} A'_h \Big/ \sum_{h=1}^{H} A_h \tag{30.2}$$

$$T_c = \sum_{m=1}^{M'} A'_m \Big/ \sum_{m=1}^{M} A_m \tag{30.3}$$

The arc tracked by ground station is h, and the new arc is H'. The arc tracked by relay satellites is m, and the new arc is M'. The changing degree of ground resource and space resource is described as L_c and T_c respectively. The whole changing degree is defined as follows,

$$C = \omega_1 L_c + \omega_2 T_c,\ \omega_1 = \omega_2 = 0.5 \tag{30.4}$$

According to the principle of minimizing the changing, the nearer of the sum of new tracking arcs with the sum of original tracking arcs, the better it is, which means $C \to 1$.

30.5 Simulations

The tracking arcs can be defined in special time by the definition position of the spacecrafts and relay satellites orbits and the ground station. The ground and space resources can be scheduled according to the visible tracking predicts and the tracking requirements [8] of each satellite. The tool Satellite Tool Kits (STK) is used to simulate four LEO satellites and one MEO satellite tracking predicts to three devices (D1, D2, and D3), and the visible window for the satellites of relay satellite TDRS1 on 6th Nov. 2013. Matlab is used to realize our emergency algorithms. Based on the normal tracking plan, an arc of the emergency task is simulated in the concentrated time of D3, and the requirement is ground and space TT&C resources (Table 30.1). The conflicts for the resource are shown in Table 30.2, and the rearranging result is shown in Table 30.3.

The results above have shown that:

1. The completion rate of normal and emergency tasks are both 100 %, and the changing degree of the plan is $C = 1.1$.
2. The tracking arcs of middle and high earth orbit satellites can be moved to avoid the conflicts with the emergency task.
3. There is no ground resource to be adjusted for Sat4, so its requirement is also offered to the management of space TT&C resource.
4. The arranged arcs of relay satellite for Sat2 and Sat4 need to be moved to form new requirements offered to the management of space TT&C resource.

The algorithm of greedy policy is used for the relay satellite scheduling in emergency. Making use of the spare time by inserting and moving tracking time, the requirements of space TT&C resource can be fully satisfied.

Table 30.1 Requirement of the emergency task

Task ID	Requirement	Priority	Emergency arc	Work
Sat-Y	D3	5	14:30–14:40	TT&C
	TDRS1	5	15:07–15:20	Telemetry

Table 30.2 Conflicts for the resource

Task ID	Arranged resource	Priority	Start	End	During (min)
MEO1	D3	3	14:15	14:26	11
Sat4		1	14:31	14:38	7
Sat5		4	14:43	14:52	9
Sat2	TDRS1	2	15:01	15:07	6
Sat4		1	15:27	15:34	7
The switching time of ground device is 5 min, and the relay satellite is 17 min					

Table 30.3 Rearranging result

Task ID	Resource	Priority	Start	End	During (min)
MEO1	D3	3	14:14	14:25	11
Sat-Y	D3	5	14:30	14:40	10
Sat5	D2	4	11:40	11:51	11
Sat4	TDRS1	1	14:23	14:30	7
Sat-Y		5	15:07	15:15	8
Sat4		1	15:32	15:39	7
Sat2		2	16:00	16:05	5

30.6 Conclusions

The emergency scheduling model and work flow of space and ground TT&C resources in this paper are applicable for the spacecraft or device in emergency. The satisfaction rate of the tasks can be maximized and the changing is minimized in whole. It is an initial attempt to the integrated scheduling in emergency, and the father study may be taken on solving efficiency and its whole evaluation etc.

References

1. Kang N, Wu XY (2011) The scheduling model of TT&C resources of TDRSS and ground stations based on task begin time. J Acad Equip Command Technol 22(6):97–101
2. Chen F, Wu XY (2010) Space and ground TT&C resource integrated scheduling model. J Astronaut 31(5):1405–1412
3. Bai J, Shen Y et al (2013) One Heuristic adjusting algorithm for space TT&C resources under emergencies. J Astronaut Dyn 3(3):102–108
4. Cormen TH, Leiserson CE, Rivest RL et al (2008) Introduction to algorithm, 2nd edn. China Machine Press, Beijing
5. Cui ZX (2009) Study on the relay satellites scheduling. Engineering Master thesis of Institute of University of Defense Technology, Changsha

6. Preindl B, Seidl M, Mehnen L et al (2010) A performance comparison of different satellite range scheduling algorithms for global ground station networks. In: 61st international astronautical congress, Prague, Czech Republic
7. Barbulescu L, Howe AE, Whitley LD (2006) Understanding algorithm performance on an oversubscribed scheduling application. Artif Intell Res 27:577–615
8. Ling XD, Wu XY, Liu Q (2009) Requirement-oriented TT&C scheduling algorithm. Syst Eng Electron 31(7):1661–1666

Chapter 31
State Trend Prediction of Spacecraft Using PSO-SVR

Yu Gao, Tianshe Yang, Weiping Li and Hailong Zhang

Abstract Fault prediction is the core content and crucial technology for health monitoring of the in-orbit spacecraft, and predicting the future trend of telemetry data is the prerequisite and basis for fault prediction. This paper presents a state trend prediction method for spacecraft based on particle swarm optimization (PSO) and support vector regression (SVR). The method applies SVR to construct a regression prediction model of telemetry data. SVR is a learning procedure based on statistical learning theory, which employs the training data to build an excellent prediction model in the situations of small sample. The complexity and generalization performance of the SVR model is influenced by its training parameters. In this paper, PSO is applied to optimize the parameters of SVR model. The results show that the method is efficient and practical for telemetry data prediction of the in-orbit spacecraft.

Keywords Spacecraft · Trend prediction · Support vector regression · Particle swarm optimization

31.1 Introduction

With the development of the space industry, modern spacecraft are becoming more advanced and complex. Safety and reliability requirements are also increasing. Because spacecrafts are running in the harsh space environment, it is practically impossible to completely eliminate the possibility of anomalies or faults, even if we increase the reliability of the system components to the limit. Once a fault occurs, it may cause huge economic losses and tremendous impact. Therefore, predicting a fault of spacecraft before it occurs is significantly important to avoid disastrous accidents.

Y. Gao (✉) · T. Yang · W. Li · H. Zhang
Key Laboratory of Fault Diagnosis and Maintenance for in-Orbit Spacecraft, Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: gy615@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_31

Spacecraft is a highly complex nonlinear system. Most of the changes in telemetry parameters of spacecraft are nonlinear. Artificial intelligence as a typical non-linear modeling method has been widely used in nonlinear system fault prediction [1, 2]. Artificial neural network (ANN) has a strong adaptive learning and nonlinear approximation ability and has been widely applied in fault prediction [3–5]. However, the model of ANN is vulnerable to the local minimum value and slow convergence. Support vector regression (SVR) effectively overcomes the shortcomings of ANN. SVR is based on the structural risk minimization principle. At present, SVR has been emerging as an alternative and powerful technique to solve the nonlinear regression problem [6–8]. When using SVR, the main problem is how to set the best kernel parameters. Proper parameters setting can improve the SVR regression accuracy. However, inappropriate parameters in SVR lead to over-fitting or under-fitting. Different parameter settings can cause significant differences in performance. Therefore, selecting optimal parameter is an important step in SVR.

In this paper, we propose a state trend prediction method for spacecraft using support vector regression based on particle swarm optimization (PSO) algorithm. We apply SVR to construct the regression prediction model of telemetry data and apply PSO to determine three parameters of SVR. The experimental results show that the method achieves high prediction accuracy.

31.2 Support Vector Regression

Support vector machine (SVM) is proposed to solve the classification issues [9]. However, with the introduction of loss function, SVM has been developed to solve nonlinear regression estimation problems, known as support vector regression (SVR). In SVR, the basic idea is to map x from the input space into a higher dimensional feature space F via a nonlinear mapping function $\varphi(x)$, and then to conduct linear regression in F space. Given the sample data (x_i, y_i), $i = 1, 2, \ldots, l$, SVR is to find the linear relation:

$$f(x) = \omega \cdot \varphi(x) + b \quad \varphi : R^n \to F, \quad \omega \in F \tag{31.1}$$

where ω is a vector for regression coefficients, b is a bias. They are estimated by minimizing the regularized risk function, namely:

$$R(\omega) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} L_\varepsilon(y, f(x)) \tag{31.2}$$

where C is a penalty factor, $L_\varepsilon(y, f(x))$ is ε-insensitive loss function, which is defined as follows:

$$L_{\varepsilon}(y,f(x)) = \begin{cases} 0 & |y-f(x)| \leq \varepsilon \\ |y-f(x)| - \varepsilon & |y-f(x)| > \varepsilon \end{cases} \tag{31.3}$$

Considering those regression errors beyond precision, slack variables $\xi_i \geq 0$ and $\xi_i^* \geq 0$ are introduced to deal with the data points that do not satisfy Eq. (31.3). So SVR can be expressed as follows minimal optimization problem:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ s.t \quad & \begin{cases} wx_i + b_i - y_i \leq \varepsilon + \xi_i^* \\ -wx_i - b_i + y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{31.4}$$

Equation (31.4) is a typical quadratic programming problem and can be solved by Lagrange multiplier method. According to the saddle point conditions and Wolfe duality theory, the original problem can be transform to its dual problem, as following:

$$\begin{aligned} \max \quad & \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) - \tfrac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\varphi(x_i)^T\varphi(x_j) \\ s.t \quad & \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \tag{31.5}$$

According to Karush-Kuhn-Tucker theorem, we can obtain the optimal solution α_i, α_i^*, $i = 1, 2, \ldots, l$. Based on the relationship between ω, b and α_i, α_i^*, we can get the optimal estimation function:

$$\begin{aligned} \omega &= \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\varphi(x_i) \\ b &= y_i - (\omega \cdot \varphi(x_i)) - \varepsilon, \alpha_i \in (0, C) \quad or \\ b &= y_i - (\omega \cdot \varphi(x_i)) + \varepsilon, \alpha_i^* \in (0, C) \end{aligned} \tag{31.6}$$

From the Eq. (31.5), we can see that the optimal estimate value of ω is only related to sample vector x_i $(\alpha_i \neq \alpha_i^*,\ i = 1, 2, \ldots, l)$. Such samples are called support vectors. Support vector is only a small portion of the total sample and can be used to simply the training process.

Using kernel function $K(x_i, x_j) = (\varphi(x_i) \cdot \varphi(x_j))$ to replace inner product, we can get the ε-insensitive support vector regression function:

$$f(x) = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) K(x_i, x) + b \tag{31.7}$$

Common kernel functions include polynomial kernel function, Gaussian radial basis function (RBF) and Sigmoid kernel function and so on. Considering the RBF kernel function is a general kernel function and has less restrictive conditions. So we use RBF kernel function $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{\sigma^2})$ in this paper.

31.3 Prediction Model Based on SVR

Telemetry data of spacecraft is typical time-series data. For a given nonlinear time series $X_N = \{x_1, x_2, \ldots, x_n\}$, Assuming that data before time t are known, in order to predict data value in time $t + 1$, we can construct the mapping $f : R^m \to R$, so that:

$$\underset{t+1}{\overset{*}{x}} = f(x_{t-m+1}, x_{t-m+2}, \ldots, x_t) \tag{31.8}$$

$\underset{t+1}{\overset{*}{x}}$ is the prediction value of time $t + 1$. m is the input dimension.

Based on the support vector regression algorithm, the regression function for training support vector machine is:

$$f(\bar{t}) = \sum_{i=1}^{n-m} (\alpha_i - \alpha_i^*) K(\bar{x}_i, \bar{x}_t) + b, \quad t = m + 1, \ldots, n \tag{31.9}$$

Then one-step SVR prediction model is:

$$\underset{n+1}{\overset{*}{x}} = \sum_{i=1}^{n-m} (\alpha_i - \alpha_i^*) K(\bar{x}_i, \bar{x}_{n+1}) + b \tag{31.10}$$

where $\bar{x}_{n+1} = \{x_{n-m+1}, x_{n-m+2}, \ldots, x_n\}$ and $\underset{n+1}{\overset{*}{x}}$ is the prediction value of time $n + 1$.

Similarly, the prediction model of Step l is:

$$\underset{n+l}{\overset{*}{x}} = \sum_{i=1}^{n-m} \left(\alpha_i - \alpha_i^*\right) K(\bar{x}_i, \bar{x}_{n+l}) + b \tag{31.11}$$

where $\bar{x}_{n+l} = \{x_{n-m+l}, x_{n-m+l+1}, \cdots, \underset{n+1}{\overset{*}{x}}, \cdots, \underset{n+l-1}{\overset{*}{x}}\}$ and $\underset{n+l}{\overset{*}{x}}$ is the prediction value of Step l.

31.4 Parameter Optimization of SVR Based on PSO

In our support vector regression prediction model, there are three parameters that need to been determined. The three parameters are insensitive loss function parameter ε, penalty factor C and kernel function parameter σ. The ε-insensitive loss function controls the width of ε-insensitive zone, is used to fit the training date. The value of ε affects the number of support vectors used for prediction. Therefore, a larger value of ε will lead to a smaller number of support vectors and a less complex model, vice versa. Penalty factor C determines the tradeoff cost between minimizing the training error and minimizing model complexity. A small value for C will increase the number of training errors, while a large C will lead to a behavior similar to that of a hard-margin SVR. The kernel function parameter σ implicitly defines the nonlinear mapping from input space to some high-dimensional feature space and will affect the complex of data distribution in high-dimensional feature space. Big value of σ indicates the good fitting performance of the nuclear function and poor pan-ability variation. However, it was noticed by researchers that each of parameters reaching an optimal value point sometimes does not achieve a good performance for SVR, just when the combination of their optimal values, one can get a nice performance.

In this study, PSO was introduced to search the optimal subset (ε, C, σ) for SVR. The process is shown in Fig. 31.1.

The parameters optimization process consists of two modules: PSO algorithm module and SVR module. Firstly, PSO algorithm module passes the initial parameters (ε, C, σ) to the SVR module. SVR module computes the prediction error and passes the prediction error back to the PSO algorithm module. PSO algorithm module compares the current prediction error with the last minimum error. If the

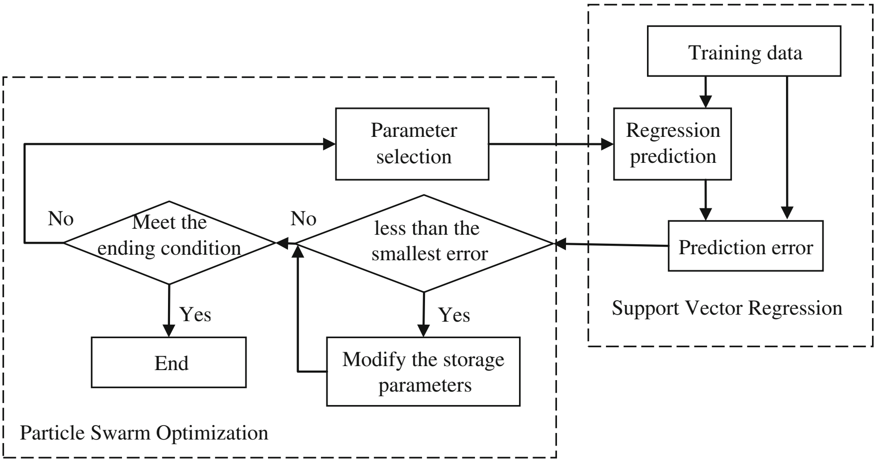


Fig. 31.1 Parameters optimization process of SVR based on PSO

current prediction error is smaller than the last minimum error, we store the current prediction error and the current parameters. After that, the algorithm determines whether the ending condition is satisfied. If not, the parameters are re-selected and the above process is repeated. If the ending condition is satisfied, the stored parameters are the optimal parameters.

Each particle in our PSO algorithm is made up of a parameter vector (ε, C, σ). The ith particle is looked as a point in the 3D space and represented as $\bullet x_i = (x_{i1}, x_{i2}, x_{i3})^T$, its velocity is represented as $v_i = (v_{i1}, v_{i2}, v_{i3})^T$. The position of each particle with its best-fit value that is its local best, is remembered and denoted as p_i, its global best, which is the position with the best-fit value of all particles, is also recorded as p_g. At each iterative process, the velocity and position of each particle are adjusted by tracking its local best value, global best value and its present velocity, their iterative equations are as follows:

$$\begin{aligned} v_{id} &= \omega v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \\ x_{id} &= x_{id} + v_{id} \end{aligned} \tag{31.12}$$

where v_{id} and x_{id} are respectively the speed and position of the ith particle; r_1 and r_2 are random values between 0 and 1; c_1 and c_2 are both learning factors, normally was set as 2; ω is a weighting factor to accelerate the convergence rate, its value should be automatically regulated with the iterative time of algorithm extending, defined generally as:

$$\omega = \omega_{\min} + (iter_{\max} - iter)(\omega_{\max} - \omega_{\min})/iter_{\max} \tag{31.13}$$

where $\omega_{\max}$ and $\omega_{\min}$ are the biggest and smallest weighting factors respectively, commonly set as 0.9 and 0.4, $iter$ is the number of current iteration. $iter_{\max}$ is the total number of iterations. In this study, mean square error (MSE), which directly reflects the regression performance of SVR, is selected as the fitness function:

$$MSE = \frac{1}{n}\sqrt{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2} \tag{31.14}$$

31.5 Results Analysis

We have implemented the approach described above and verified it using actual satellite telemetry data. The prediction results of the satellite telemetry parameter TG1, IN2 and VN10 are shown in Figs. 31.2, 31.3 and 31.4 respectively. For each parameter, figure (a) shows the measured values and the prediction values. Figure (b) shows the prediction error. Table 31.1 shows the average absolute error, the mean square deviation, the average relative error and the maximum absolute deviation of all experimental telemetry data.

Table 31.2 shows the comparison between our PSO algorithm and the cross validation method in SVM tool LIBSVM [10]. From Table 31.2 we can see that the

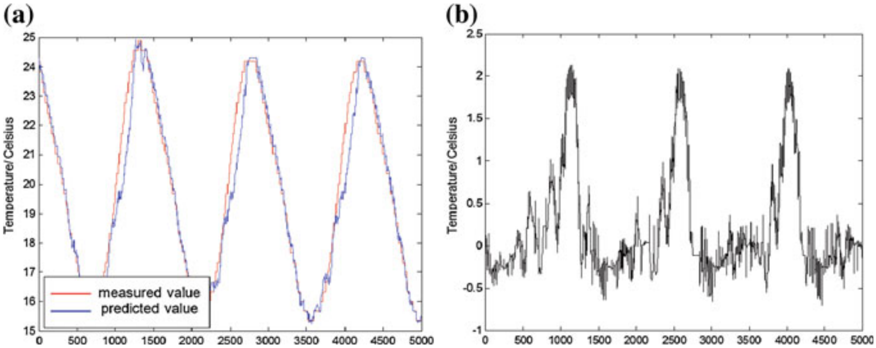


Fig. 31.2 Prediction result and error of parameter TG1. **a** Prediction result. **b** Prediction error

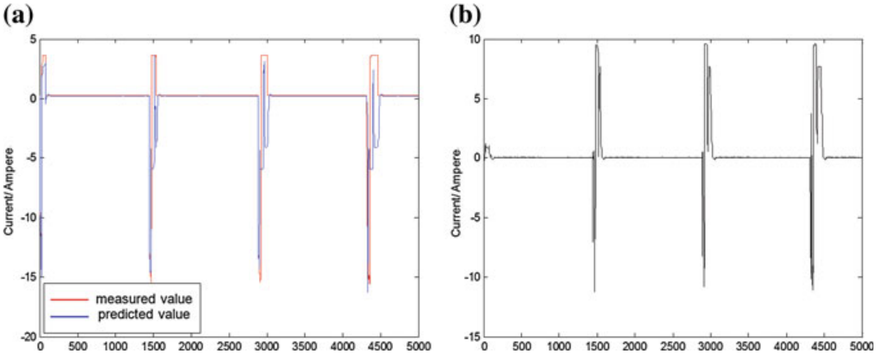


Fig. 31.3 Prediction result and error of parameter IN2. **a** Prediction result. **b** Prediction error

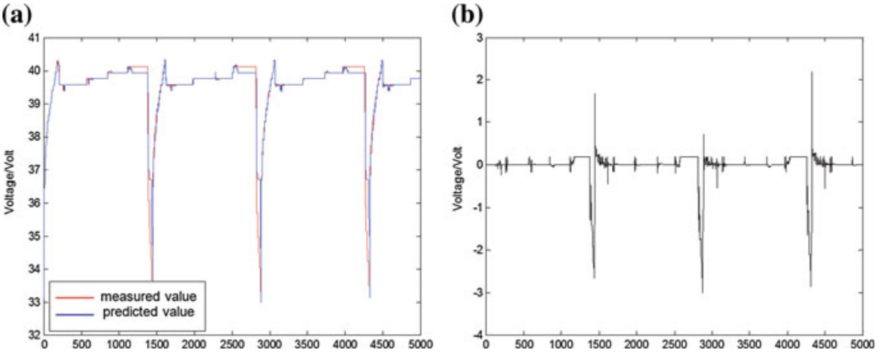


Fig. 31.4 Prediction result and error of parameter VN10. **a** Prediction result. **b** Prediction error

Table 31.1 Prediction precision of different parameters

Parameter	Average absolute error	Mean square deviation	Average relative error	Maximum absolute deviation
TG1 (Temperature)	0.4042	0.6564	0.0194	2.1339
TK2 (Temperature)	0.0161	0.0608	0.0311	0.6876
TK8 (Temperature)	0.3693	0.5456	0.0246	1.9553
TN1 (Temperature)	0.1938	0.2688	0.0594	1.7932
IN2(Current)	0.0214	0.0332	0.5409	0.2331
IN6(Current)	0.8459	0.9385	2.4451	2.7512
VN2 (Voltage)	2.2845	2.3205	0.0574	2.8522
VN10 (Voltage)	0.1075	0.3726	0.0029	3.0170

Table 31.2 Comparison between the PSO algorithm and the cross validation method

Item	PSO algorithm	Cross validation method
ε	0.0015	0.0032
C	632	389
σ	3.25	6.53
Prediction error	0.4042	0.4859
Optimization time (s)	62.8	75.9

prediction accuracy of the PSO algorithm and the cross validation method are very close. But the PSO algorithm is better than the cross validation method in optimization time. When the training data becomes large, this advantage will become more apparent.

31.6 Conclusion

This paper proposed a telemetry data trend prediction method for in-orbit spacecraft based on particle swarm optimization and support vector regression. The method uses support vector regression to construct the prediction model of the telemetry data. The parameters of the support vector regression model are optimized by particle swarm algorithm. The result shows that the method is efficient and practical for telemetry data prediction of the spacecraft system.

References

1. Huang YL et al (2013) Research on method of electronic equipment fault prediction. In: Proceedings of IEEE fourth international conference on digital manufacturing and automation (ICDMA), pp 1080–1085
2. Yang TS, Chen B, Gao Y et al (2013) Data mining-based fault detection and prediction methods for in-orbit satellite. In: Proceedings of IEEE international conference on measurement, information and control (ICMIC), pp 805–808
3. Gao Y, Yang TS, Xu MQ et al (2012) A neural network approach for satellite telemetry data prediction. In: Proceedings of IEEE international conference on electronics, communications and control (ICECC), pp 150–153
4. Yang TS, Chen B, Zhang HL et al (2013) State trend prediction of spacecraft based on BP neural network. In: Proceedings of IEEE international conference on measurement, information and control (ICMIC), pp 809–812
5. Cui JG, Zhang L, Wang GH et al (2014) Fault prediction method of the marine gas turbine based on neural network-Markov. Appl Mech Mater 538:171–174
6. Deng S, Jing B, Zhou HL et al (2012) Fault prediction method based on improved AdaBoost-SVR algorithm. Acta Armament 33(8):991–996
7. Fan J, Tang Y (2013) An EMD-SVR method for non-stationary time series prediction. In: Proceedings of IEEE international conference on quality, reliability, risk, maintenance, and safety engineering (QR2MSE), pp 1765–1770
8. Liu L, Shen J, Zhao H (2012) Fault forecast of electronic equipment based on ε-SVR. Web information systems and mining. Springer, Berlin, pp 521–527
9. Poyhonen S, Arkkio A, Jover P et al (2005) Coupling pairwise support vector machines for fault classification. Control Eng Pract 13(6):759–769
10. Lin CJ. LIBSVM. http://www.csie.ntu.edu.tw/~cjlin/

Chapter 32
Application of Differential Evolution Algorithm in Multi-satellite Monitoring Scheduling

Jianguang Wu, Shuo Wang, Yang Li, ChaoPing Dou and Jin Hu

Abstract This paper analyzes the situation of multi-satellite monitor scheduling problem, according to the constraints in which, two models were established respectively, one is of total successful scheduling tasks in same weight, the other one is in different weights. Then a codec-based differential evolution algorithm was designed to solve the scheduling problem. Firstly, each evolution individual was encoded into real-coding for the use of mutation and crossover, and the optimum individual was singled out for the next iteration of the loop to get the best result by greedy selection method. After the process of the algorithm, the sequences of tasks which to assign ground station resources and the executing time were listed by decoding result code. Finally we get the results of general task scheduling and weighted task scheduling with the scheduling model. The simulating shows that the algorithm could achieve satisfactory scheduling results in satellite monitoring scheduling problem.

Keywords Real code · Monitoring schedule · Differential evolution algorithm (DE algorithm) · Scheduling model

32.1 Introduction

With the development of the aerospace industry, the quantity of spacecraft increases and the number of monitoring tasks will grow rapidly. In aerospace TT&C network system, how to assign the limited resources and equipment reasonably and to improve the task executing efficacy becomes an important issue. In terms of the monitoring section of tasks, among different monitoring tasks exist some problems, such as back and forth dependences, resource contention, the task executing priority

---

J. Wu (✉) · S. Wang · Y. Li · C. Dou · J. Hu
School of Information and Electronics, Beijing Institute of Technology,
Beijing 100081, China
e-mail: younacoplus@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_32

and specific parameter index requirements of different business sectors, which are usually different. The traditional model's study on monitoring resources scheduling mainly focus on two aspects of monitoring scheduling model and monitoring scheduling algorithm [1–5]. They are considered insufficient for such difference in the situation. Schalck's [1] scheduling model proposed that all tasks will be seen as the same priority to let the maximum number of executed tasks as the goal, it established a mixed integer programming model. Jin et al. [2] proposed a Petri net model of monitoring station and get the indicator of two owners, which are both satellite users and ground station users aimed to system evaluation model. But when the scale of monitoring tasks is large, this model's complexity will be exponential increased. In order to solve large-scale scheduling problem, Chen et al. [3] proposed a two-stage progressive evolution model of space and earth monitoring resources, which is decomposing the optimal solution of multiple objective into stages of progressive search optimal solution. Jin [4] proposed an expended CSP model to describe and solve the problem of monitoring resources scheduling, but it has the limits of common constraints to consider not comprehensive and lacks of studies to objective function. On this basis, Lin et al. [5] proposed further specific studies of CSP model and established the CSP model of monitoring resources scheduling for the specific problems of multiple satellites monitoring scheduling, and used the heuristic algorithm to solve them.

The current monitoring scheduling algorithms mainly include local search algorithm, mathematical programming algorithm, heuristic algorithm, intelligent algorithm, etc. The local search algorithm [6, 7] was studied earlier, which is used in the problem of smaller scale. Mathematical programming algorithm [1] achieved better results of early monitoring resources, but was only suitable for the situation of smaller scale. Heuristic algorithm relies on specific issues form and thus has lower applicability. Genetic algorithm [6, 8, 9] as the representative algorithm of intelligent algorithm was widely used in the aerospace monitoring scheduling. Due to the time constraints of specific monitoring tasks, each mutation or crossover of genetic algorithm would generate lots of unreasonable solutions, and the lack of efficiency and computational speed of genetic algorithm particularly appear when the monitoring constraints or the task increased. Particle swarm optimization algorithm [10] improved the computing speed and efficiency compared to the genetic algorithm in monitoring scheduling aspect [11], but it is easy to make the monitoring problem fall into local optimal solution [12].

The differential evolution algorithm [13] applied in this paper is also an intelligent algorithm. This paper firstly applied DE algorithm into the scheduling of monitoring tasks. It is conducive to a better solution of the problem in different scheduling tasks.

This paper is structured as follows: Sect 32.2 briefly introduces the DE algorithm, Sect. 32.3 analyses and describes the model of multi-satellite monitoring scheduling problem, Sect. 32.4 designs the model of monitoring tasks scheduling and the implements of codec of DE algorithm in the monitoring scheduling problem, Sect. 32.5 gets the simulation results by calculating the AFIT benchmarks, and obtaining the feasibility and advantages of this algorithm by comparing with the classical algorithms, finally, Sect. 32.6 gives summary and future view.

32.2 Differential Evolution Algorithm Profile

DE algorithm is a population-based evolutionary algorithm, which has the characters of remembering the optimal values of individuals and sharing information within the population, namely through cooperation, evolution, competition among individuals within population to achieve the solution of the optimization problem. It is essentially a special genetic algorithm based on real-coded and greedy strategy [13].

For the optimization problem:

$$G = \max f(\mathrm{x}), \quad \mathrm{x} = [x_1, x_2, \ldots x_d], \quad l_k \le x_k \le u_k, \quad k = 1, 2, \ldots d \tag{32.1}$$

where G is the objective function, l_k and u_k respectively denote the down-bound and up-bound of the k-dimensional variable. d is the dimension of the variable.

In DE algorithm, each individual is corresponded to a solution $\mathrm{x_i} = [x_{i1}, x_{i2}, \ldots x_{id}]$, where x_i represents first ith individual of the population. During DE initialization, N random individuals are generated to constitute initial population, as in formula (32.2)

$$x_{ik}(0) = l_k + rand() \cdot (u_k - l_k), \quad k = 1, 2, \ldots d, \quad i = 1, 2, \ldots N \tag{32.2}$$

where rand() means an uniformly distributed random number between 0 and 1. Standard DE algorithm consists of three core operations: mutation, crossover, selection.

32.2.1 Mutation

In each generational search, DE algorithm generates a target individual $t_i(g)$ for the current individual $x_i(g)$ of population through mutation operation, where g represents the generation of evolution. Currently, DE algorithm has many different variation mechanisms which can be expressed as DE/*a*/*b*, where *a* represents the type of mutation operation object, generally taking the value of rand and best. Rand denotes randomly selecting ***a*** individual as the mutation operation object, while best denoted selecting the best individual as the mutation operation object; ***b*** indicates the number of the differential item during mutation, generally rounded. Some common mutation mechanisms are as follows:

DE/rand/1 $t_i(g) = x_{r_1}(g) + F \cdot [x_{r_2}(g) - x_{r_3}(g)]$
DE/best/1 $t_i(g) = x_{best}(g) + F \cdot [x_{r_1}(g) - x_{r_2}(g)]$

DE/rand/2 $t_i(g) = x_{r_1}(g) + F \cdot [x_{r_2}(g) - x_{r_3}(g)] + F \cdot [x_{r_4}(g) - x_{r_5}(g)]$
DE/best/2 $t_i(g) = x_{best}(g) + F \cdot [x_{r_1}(g) - x_{r_2}(g)] + F \cdot [x_{r_3}(g) - x_{r_4}(g)]$
DE/target-to-best/1 $t_i(g) = x_i(g) + F \cdot [x_{best}(g) - x_i(g)] + F \cdot [x_{r_1}(g) - x_{r_2}(g)]$

where $r_1, r_2, r_3, r_4, r_5 \in \{1, 2, \ldots, N\}$ are integers which are different from each other except for i. $x_{best}(g)$ represents the best individual of first g generational population. $F \in (0, 1)$ is the scaling factor. This paper adopts the variation mechanism of DE/target-to-best.

32.2.2 Crossover

The crossover operation of DE algorithm executes after mutation operation by replacing partial variables of current individual with corresponding variables of target individual, thereby generating test individual which could retain better variable of individual and enhance the ability of local exploration.

During the crossover operation in this paper:

1. An integer k is randomly selected for each individual between 1 and d as the starting point of the cross.
2. A random number r_k is generated corresponding to this genetic bit k. If $r_k < Cr$ (Cr is the crossover probability for controlling whether to receive the variable of target individual), then the genetic bit k of current individual would be replaced with the corresponding variable of target individual. Then the next bit $(k+1)_{d+1}$ would be examined, and similarly a random number $r_{(k+1)_{d+1}}$ corresponding to this genetic bit $(k+1)_{d+1}$ would be generated. If $r_{(k+1)_{d+1}} < Cr$, the corresponding component of the target individual would be accepted until $r_{(k+p)_{d+1}} \geq Cr$ or $(k+p)_{d+1} = k - 1$, where $()_{d+1}$ represents the modulo operation of an integer relative to $d + 1$, $0 \leq p \leq d$ and p represents the total number of crossing genetic bits.

32.2.3 Selection

Standard DE algorithm adopts greedy selection method, for the test individual $v_i(g)$ and the current individual $x_i(g)$, the better one is selected into the next generational search. Namely:

$$x_i(g+1) = \begin{cases} v_i(g), & f[v_i(g)] > f[x_i(g)] \\ x_i(g), & otherwise \end{cases} \tag{32.3}$$

By greedy selection mechanism the quality of the next generational individuals is ensured in the population, so that the average performance of the population gradually reaches the optimal solution or satisfactory solution.

32.3 Multi-satellite Monitoring Problem and Symbol Definitions

32.3.1 Selection

This paper studies the problem of multiple resources monitoring scheduling, which exists multiple monitoring resources (ground stations and relay satellites) and multiple monitoring satellites. As the constraints of the satellite's orbit, the ground station's location and the ground station monitoring equipment must be matched, between satellite and monitoring station or relay satellite can be visible and meet the requirements of communication only within a specific time. This time interval is defined as the visible monitoring arcs. Satellite and ground station or relay satellite can establish a link to complete the appropriate monitoring tasks only in the visible arcs. Considering the limit of monitoring stations and satellites resources, the essence of the problem is solving the conflicts and matching competitions between multiple satellites and multiple stations in the visible fixed monitoring arcs, in order to make full use of the monitoring resources to meet the needs of monitoring tasks.

32.3.2 Model and Symbol Definitions

Assuming the set of monitoring tasks is $Task = \{task_1, task_2, \ldots, task_z\}$, n ground stations were $\mathrm{g}_1, \mathrm{g}_2, \ldots, g_n$, among them, the ground station g_i has k_i antennas, respectively, a_i^1, $a_i^2, \ldots, a_i^{k_i}$; let m satellites were $s_1, s_2, \ldots s_m$. There are h visible-windows between g_i and s_j within a scheduling period. The first r visible-time-window is $tw_{ij}^r[t_s, t_e]$,$r = [1, 2, \ldots, h]$, t_s represents the start time of visible-window, t_s represents the end time of visible-window. The monitoring between satellites and ground stations needs to meet the following two conditions: first, satellite's antenna and ground station's antenna must be matched; second, the satellite and ground station must be in the visible arc paragraph. Thus, we can use a six elements array to describe the satellite monitoring and control tasks, that is named as $\{T, S, G, A, TW, P\}$, in which, T represents the task; S indicates the satellite which is needed for satellite monitoring mission; G means the ground station set which contains the ground station qualified the ability of receiving satellite monitoring data; A means the antenna collection which includes the antenna capable of establishing a connection with satellites. Each of the antennas among A belong to a particular ground station of G; TW said the set of the visible window between satellite S and all ground stations of G during a scheduling cycle; P is the task priority.

In addition, each monitoring tasks are required a resource preparation time before executing. The preparation time is mainly associated with the ground station, which will be considered in the algorithm.

32.3.3 Problem Assumption

In this paper, making the following assumptions of monitoring process:

1. Tasks are independent.
2. The mission can't be preempted when it is executing.
3. Equipment failure is not considered.
4. Resource preparation time of ground stations is constant.
5. The antenna can't be used for task during the preparation and release time.
6. Multiple monitoring devices belong to the same ground control station are seemed as different devices.

Assuming constraints:

- Constraint 1: At a moment, an antenna is served to a satellite only, and a satellite only requires a single antenna for its service.
- Constraint 2: Only considering the visible time windows which are greater than a certain minimum time, it is because before performing monitoring tasks, earth station antennas and satellite dishes need to go through acquisition, tracking, targeting and other stages, each stage will take some time, so the visible time window must be greater than the minimum threshold time $TW_{\min}$.
- Constraint 3: All data monitoring tasks must be achieved within the visible monitoring arcs.

32.4 DE Algorithm of Satellite Monitoring Tasks Scheduling Problem

32.4.1 Satellite Monitoring Scheduling Model

Let space monitoring task set as $Task = \{task_1, task_2, \ldots, task_z\}$, to maximize the success rate of weighted scheduling task as the scheduling objective function G, the scheduling objective function G is

$$G = f(task) = Max\left\{ \sum_{i=1}^{z} p_i c_i \middle/ \sum_{i=1}^{z} p_i \right\} \tag{32.4}$$

$$c_i = \begin{cases} 1 & \text{task}_i \ sucessed \\ 0 & otherwise \end{cases} \tag{32.5}$$

where p_i is the task priority of $task_i$.

When the time length of the visible window is not only less than the minimum threshold time $TW_{\min}$, but ground station resources are also allocated for the task successfully, it indicates the task scheduling successfully.

32.4.2 Codec-Based Differential Evolution Algorithm Design

The core of using DE algorithm to solve the satellite monitoring scheduling problem is to determine the order of ground resources and execution time allocation for the tasks, so the form of solution is the task sequence. For conflict resolution, real coding is used for the unassigned tasks there, coding method is random number coding [14], and then the order of allocation of ground resources and execution time for the tasks is determined by the result of decoding.

32.4.2.1 Encoding

Encoding: First, assuming $Task' = \{task^1, task^2, \ldots, task^z\}$ is a set of tasks sequence which is sorted ascended by the ceiling of required executing time, then randomly generating z random numbers between 0 and 1 corresponding to each of these tasks, it is called a random number sequence. The size of random number sequence which is corresponding to each task determines the order of allocation of ground station resources and execution time (Fig. 32.1). In the algorithm, the task sequence remains unchanged, but the corresponding random number sequence will change in the algorithm. Each random number sequence is called an individual (or chromosome). Each number of the sequence corresponds to a location which is called a genetic bit. The random number is the value of that genetic bit.

32.4.2.2 Decoding

The main process of decoding is mapping between task sequence and random number sequence, the order of allocation of ground resources and task execution

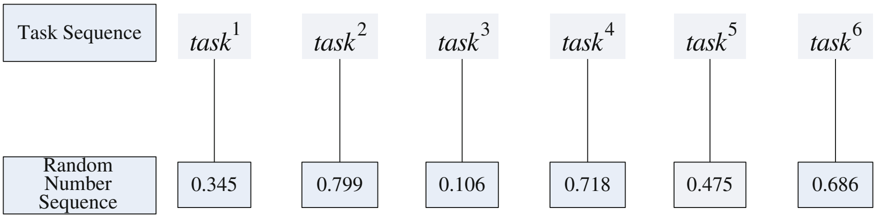


Fig. 32.1 Relation between task sequence and random number sequence

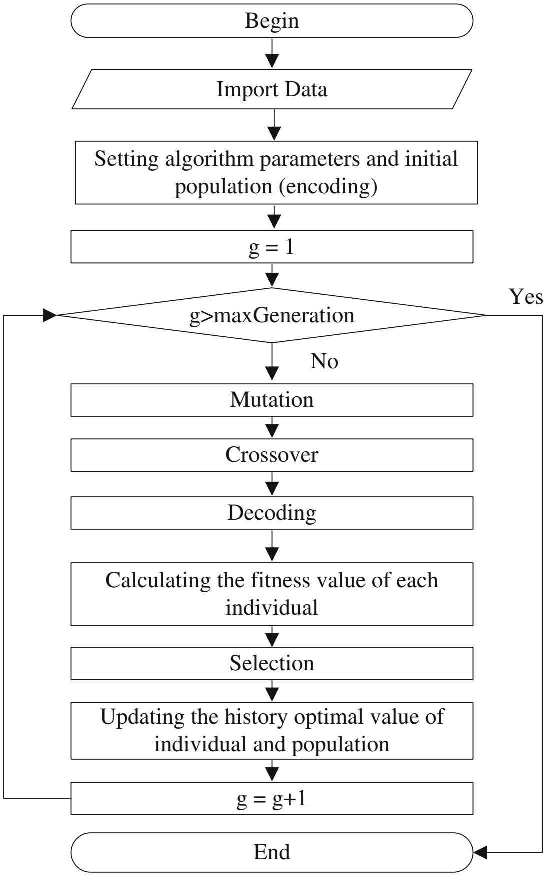


Fig. 32.2 DE algorithm's flowchart

time of task sequence is determined by random number sequence. There we convention: the descending order of random numbers is the order of allocation of ground resources and task execution time. The top surface of task is priority assignment when two tasks corresponding to the same random number. Decoding the chromosomes of Fig. 32.1, we can get the executing order of tasks:

$$task^2 - task^4 - task^6 - task^5 - task^1 - task^3$$

The main DE algorithm is illustrated by the Fig. 32.2.

32.5 Simulation Results

This paper aims AFIT LEO satellite test data[2] which is used to verify the satellite scheduling algorithm to conduct simulation. This problems is developed in DEV-C ++ 5, computer environment is configured as Intel (R) Core (TM) 2 Duo CPU

E8400 @ 3.00 GHz, RAM 4.00 GB. This paper makes some assumptions for AFIT data as follows:

1. Task priority is the random integer between 1 and 10; a larger number indicates a higher priority.
2. Assuming the time of resource preparation as a basic attribute of ground stations. Before the scheduling of each task, the ground stations need to begin preparing 20 min earlier than the start time of visible window.
3. The minimum threshold time $TW_{\min} = 3\text{min}$.
4. Let the task corresponding to multiple antennas in data could be as multiple tasks to encode.

In this paper, the maximum number of successful scheduling tasks(same priority) and the maximum weighted percentage of successful scheduling tasks(different priority) are used as the objective function simulation respectively, these results are shown in Tables 32.1 and 32.2.

Table 32.1 The maximum number of successful scheduling tasks

Evaluation	Data file						
	Reqlf12	Reqlf13	Reqlf14	Reqlf15	Reqlf16	Reqlf17	Reqlf18
Total tasks	153	137	146	142	142	144	142
Successful scheduling tasks	144	133	144	139	141	138	140
Execution time (ms)	234	203	250	219	234	219	203

Note mutation mechanism: DE/target-to-best, scaling factor: 0.6, crossover probability: 0.9, individual number of population: 10, evolution generations: 10

Table 32.2 The maximum weighted percentage of successful scheduling tasks

Evaluation	Data file						
	Reqlf 12	Reqlf 13	Reqlf 14	Reqlf 15	Reqlf 16	Reqlf 17	Reqlf 18
Total tasks	153	137	146	142	142	144	142
Weighted percentage of successful scheduling tasks (%)	96.42	97.66	99.05	98.90	99.77	97.74	99.06
Average weighted percentage of successful scheduling tasks (%)	98.37						
Execution time (ms)	4,369	3,514	4,037	3,881	4,028	3,924	3,670

Note mutation mechanism: DE/target-to-best, scaling factor: 0.6, crossover probability: 0.9, individual number of population: 20, evolution generations: 100

Table 32.3 Gooley and Parish's scheduling results of algorithm

Algorithm	Scheduling results (Indicator: successful scheduling percentage of tasks)
Gooley (%)	91–95
Parish (%)	96

Further, Gooley [15] proposed the mixed integer programming algorithm based on insertion and commutation and Parish [16] proposed Genitor genetic algorithm also tested the performance of algorithm aiming to AFIT data. According to document the scheduling results shown in Table 32.3.

The results show that the DE algorithm in this paper can achieve better scheduling results aimed to the scheduling of LEO satellite data transmission. Its average weighted percentage of successful scheduling tasks is 98.37 %, which is superior to the scheduling results of Gooley's and Parish's algorithms. In addition, because the algorithm studied in this paper adopts random crossover length, mutation of tending to the optimal solution of population, greedy selection strategy, the algorithm has some characters of low computation cost and short run time, so it is easier to meet the requirement of calculation timeliness in that field, all above are provided to illustrate the feasibility of the algorithm.

32.6 Conclusion

For satellite scheduling problem, we studied a scheduling algorithm based on DE algorithm. The algorithm, firstly, codes for each individual, then gets the sequence of allocation of ground resources and execution time for tasks according to the decoding result, finally, adopts greedy selection method to select better individual into the next generation. Through simulation verified the DE algorithm has the advantages of fast speed and high-quality scheduling result on the premise of ensuring corresponding successful percentage in the process of solving the scheduling problem of satellite data transmission tasks, indicates the effectiveness of the algorithm for the tasks scheduling in the case of complicated differences. The further research of the algorithm in theoretical aspect includes the convergence of algorithm, the convergence speed, limited-time performance, parameter selection, parameter robustness et al., in the aspect of algorithm itself includes presenting reasonable core update formula, the strategy of effectively balance the global search and local improvement.

References

1. Schalck S (1993) Automating satellite range scheduling. Master thesis of Air Force Institute of Technology, Wright-Patterson AFB, OH
2. Jin G, Wu XY, Gao WB (2004) Simulation-based study on resource deployment of satellite ground station. J Syst Simul 16(11):2401–2403

3. Chen F, Wu XY (2010) Two-stage successive genetic algorithm for space and ground TT&C scheduling. J Nat Univ Defense Technol 32(2):17–22
4. Jin G (2007) CSP model for satellite and ground station TT&C resource scheduling problems. Syst Eng Electron 7:023
5. Lin XD, Wu XY, Liu B (2012) Study on the CSP model of satellite TT&C resource scheduling. Syst Eng Electron 34(11):2275–2279
6. Barbulescu L, Watson JP, Whitley LD et al (2004) Scheduling space-ground communications for the air force satellite control network. J Sched 7(1):7–34
7. Barbulescu L, Howe AE, Watson JP et al (2002) Satellite range scheduling: a comparison of genetic, heuristic and local search. In: Parallel problem solving from nature-PPSN VII. Springer, Berlin, pp 611–620
8. Barbulescu L, Howe AE, Whitley D (2006) AFSCN scheduling: how the problem and solution have evolved. Math Comput Model 43(9):1023–1037
9. Li YQ, Wang RX, Xu MQ et al (2012) An improved genetic algorithm for a class of multi-resource range scheduling problem. J Astronaut 33(1):85–90
10. Kennedy J, Eberhart R (1995) Particle swarm optimization. Proc IEEE Int Conf Neural Networks 4(2):1942–1948
11. Chang F, Wu XY (2009) Satellite data transmission task scheduling based on advanced particle swarm optimization. Syst Eng Electron 31(10):2404–2408
12. Feng JZ, Chen X, Zheng SL (2014) Improved MOPSO algorithm and its application. Appl Res Comput 31(3):675–678
13. Storn R, Price K (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 11(4):341–359
14. Li YF, Wu XY (2008) Application of genetic algorithm in satellite data transmission scheduling problem. Syst Eng Theory Pract 28(1):124–131
15. Gooley TD (1993) Automating the satellite range scheduling process. Air Force Institute of Technology, Wright-Patterson AFB
16. Parish SA (1994) A genetic algorithm approach to automating satellite range scheduling. Air Force Institute of Technology, Wright-Patterson AFB

Chapter 33
Research of AI Planning for Space Flight Control Based on PDDL

Jianping Liu, Fan Yang and Jing Li

Abstract The application of Artificial Intelligence (AI) theories and methods are required to enable more efficient space flight control center system. Applying AI planning theories, an AI planning method is proposed for space flight control based on planning domain definition language (PDDL) in this paper. Beginning from analyzing characteristics of space flight control planning problem in terms of AI planning theories, field model and problem model for space flight control planning are established. Then a solving architecture of space flight control planning problem based on PDDL is presented. Finally taking example from east-west station-keeping control, the feasibility of this method is proved.

Keywords Artificial intelligence · Intelligent planning · PDDL · Space flight control · Station-keeping control

33.1 Introduction

The main object of space flight control planning is the controlling event. According to some experts' knowledge and experiences, its purpose is to formulate the proper and available satellite controlling plan which meets some given timing constraints, resources constraints and logical constraints. Its key problem is to answer "what time and order are all the controlling events assigned?" Along with high dense space control mission, the increasing sophisticated spacecraft operations and more heavy work in multi-satellites long-term management, the flight control planning certainly will be corn component in the high efficient space flight control center system [1, 2].

J. Liu (✉) · J. Li
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: ljpnudt@sina.com

F. Yang
Xi'an Satellite Control Centre, Xi'an 710043, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_33

The space flight control planning is a typical knowledge-dealing process, which involves complicated logical reasoning and a large numerous of constraints, and this problem is well suited to AI approaches. Seen from current open literatures [3–7], AI planning based rules is a general method to solve this problem. The basic process of the method is: the rule library based on all kinds of judging rules and constraints for planning process is firstly formulated, and then the reasoning machine base on a specific searching algorithm is completed. During the running period, in terms of current mission state, the reasoning machine carries out searching, judging and reasoning to generate finally the flight controlling plan.

There are the two aspects concerning about AI planning methods. On one hand, though AI planning method based rules can efficiently generate the flight controlling plan especially for missions with high real-time requirements, it is dedicated to the flight mission, and the reasoning machine is closely connected with rules. On the other hand, AI planning formulates a specific planning problem into domain and problem description, and problem formulations is separated from problem-solving [8], which helps to plan various problem with same domain. So, AI planning method based on domain and problem description is chosen.

In this paper, An AI planning method is proposed for space flight control based on planning domain definition language (PDDL) in this paper. Beginning from analyzing characteristics of space flight control planning problem in terms of AI planning theories, domain model and problem model for space flight control planning are established. Then a solving architecture of space flight control planning problem based on PDDL is presented. Finally taking example from east-west station-keeping control, the feasibility of this method is proved.

33.2 Characteristics of Space Flight Control Planning Problem

The following summarized the seven characteristics of space flight control planning problem in terms of AI planning theories.

1. The nature of space flight control planning problem is a temporal planning problem.
   There are five input elements for this problem: the first one is the satellite control network scheduling result; the second is the predictive results of all flight control computing which can be predicted; the third is the uplink requirements of payload commands; the fourth is the requirements of satellite specific controlling events which have given beginning time and ending time; the fifth is the current state of spacecraft. For the above elements, there are many temporal constraints within single elements or among them, such as minimum time interval and maximum time interval. A time sequence output made of all controlling events is provided ground operators to carry out on schedule. So, space flight control planning is not state-transferring process in classical planning

problem, but more likely is a temporal planning problem with lots of various running time events.

2. The basic object is the controlling event.
   Mentioned above, the space flight control plan is a temporal sequence which is used by ground engineers to carry out all involved controlling events, and really is a timeline of controlling events. A controlling event may be a spacecraft commanding sequence such as gyro calibration procedure, or a single command such as Stationkeeping firing startup, or a tracking procedure of ground device such as a tracking orbit event, or some ground operation procedure such as Stationkeeping parameters computing. Each controlling event has specific running time, and various events have various times, from seconds to hours. For example, a single command needs the least 2 s, and a tracking orbit event may go on for 6 h.
3. Space flight control planning aims at spacecraft and ground integrated system.
   Seen from the above description of a controlling event, it is not only a spacecraft operation, but also an operation of space flight control center system. For example, a tracking orbit event involves the operation of ground station and data processing of flight control center.
4. Some controlling events are determined by some predictable exogenous event constraints that happen at known times.
   For space flight control planning problem, there are two main kinds of predictable exogenous event constraints: one is time window constraints on resources, which means that occurrence of some controlling events must be within time windows. The other is the predictive times constraints on all predictable control computing events, such as Earth Sensor probe prohibition time and Stationkeeping firing time.
5. Predictable exogenous event constraints are beneficial to problem-solving.
   A planning problem is made up of initial states and goal states, and its resolution is an action sequence from initial states to goal states. Problem-solving algorithms generally begin searching from initial states to goal states or from goal states to initial states. Space flight control plan is a timeline of controlling events, where many events have predictable exogenous event constraints. So these predictable exogenous event constraints should be used to improve problem-solving efficiency.
6. This planning problem has soft constraints.
   Soft constraints are relative to hard constraints. The latter must be met when solving the problems, and the former need be met as far as possible. For example, for station keeping control, GYRO calibration procedure is planned according to some time constraints, which include time windows, engine firing time and the visible time of Sun Sensor. When the time constraints are met, this procedure is assigned; when the time constraints are not met, this procedure is not assigned.

7. The representations of this planning problem have few distinct goal states. The generation of space flight control plan is a off-ling knowledge dealing process, and the order and time of all events is based on constraints and initial states. So, compared with classical planning problem, space flight control planning problem has few distinct goal states.

33.3 The PDDL Domain Representation of Space Flight Control Planning Problem

In the planning model based on PDDL [9], space flight control planning domain can be described as a four tuples: (object, predicate, function, controlling event), and each element is described as follows: Object is used to represent different objects of space flight control planning problem, such as sensors etc. Predicate is used to represent all the attribute state of this planning problem, such as sensor switch state etc. Function is used to represent the duration time of some controlling events. Controlling event is used to represent action that can be taken, and change the state attribute of an object through the implementation of action.

In this paper an east-west station-keeping control of geostationary orbit satellite is taken as an example to introduce space flight control planning domain representation in detail. The representation for other control process planning can be formalized according to this example.

33.3.1 Definition of Objects

In the planning model of east-west station-keeping control, the object is the basis of predicate, function, controlling event. For this planning domain, eight types of object are defined, such as satellite, sensor, mode, submode, Gyro Assembly Electronics (GAE), Power Voltage Converter (PVC), Thrusts and Momentum wheel. Table 33.1 shows these objects and examples.

33.3.2 Definition of Predicate

Table 33.2 shows the definition and description of various predicates in the planning domain model of east-west station-keeping control.

Table 33.1 Objects and instances

Object	Instance
Satellite	Satellite0
Sensor	Sun_sensor,Earth_sensor
Mode	SKM(stationkeeping mode), NM(normal mode)
Submode	Submode0, submode1
GAE	Gyro GAE_A, GAE_B
PVC	PVC_A, PVC_B
Thrusts	Thrusts_A, Thrusts_B
Mwheel	Mwheel0

Table 33.2 Definition and description of predicates

Predicates	Descriptions
on_board?p-PVC?s-satellite	PVC? p is on Satellite? s
power_avail?p-PVC	The power of PVC? p is available
power_on?p-PVC	The power of PVC? p is on
on_board?ss-SSE?s-satellite	Sun sensor? ss is on Satellite? s
power_avail? ss-SSE	The power of Sun sensor? ss is available
power_on? ss-SSE	The power of Sun sensor? ss is on
Coefficients_for_attitude_ready? s-satellite	Coefficients for Satellite? s attitude is ready
Coefficients_for_attitude_on_sat? s-satellite	Attitude coefficients is uploaded on Satellite? s
attitude_caculate_working? s-satellite	Satellite? s uses coefficients to calculate attitude and to adjust attitude
attitude_caculate_ending? s-satellite	Satellite? s uses coefficients to calculate attitude, and attitude adjusting is over
new_threshold_ready? s-satellite	New threshold of attitude is ready for Satellite? s
new_threshold_working? s-satellite	New threshold of Satellite? s attitude is working
new_threshold_cancel? s-satellite	New threshold of Satellite? s attitude is cancelled
fire_parameters_ready? s-satellite	Firing parameters is ready for Satellite? s
fire_parameters _on_sat? s-satellite	Firing parameters is already uploaded on Satellite? s
Compensation_torque_ready? mw_Mwheel	Compensation torque of Mwheel? mw is ready
Compensation_torque_working? mw_Mwheel	Compensation torque of Mwheel? mw is working
Changing_NM_back? s-satellite	Satellite? s has restored normal mode
GAE_tempreture_ok? g-GAE	Gyro? g temperature is ok
EWSK_SUCCESS? s-satellite	East-west station-keeping control of Satellite? s is successful
…	…

Table 33.3 Definition and description of functions

Functions	Descriptions
upload_yaw_time? s-satellite	Duration time of Satellite? s uploading yaw Coefficients
slew_time?a?b_Mode?c? d_SubMode	Duration time of Mode? a SubMode? c transferred to Mode? b SubMode? d
Fire_time?t_Thrusts	Duration time of thrusts firing
waiting_GAE_time? s-satellite	Duration time of gyro waiting for normal temperature on Satellite? s.
GAE_calibration_time? s-satellite	Duration time of gyro calibration

33.3.3 Definition of Function

Functions are divided into three types, involving duration function of action, resource function and utility function. In the space flight control planning domain, the main function applied is duration function of action. Table 33.3 shows the definition and description of various functions in the planning domain model of east-west station-keeping control.

33.3.4 Definition of Controlling Event

The definition of controlling event is the most fundamental part in the planning domain model. The controlling events can be executed by integrated system, such as turning on some sensor, and changing mode etc. Table 33.4 shows the definition and description of controlling events in the planning domain model of east-west station-keeping control.

33.4 The PDDL Problem Representation of Space Flight Control Planning Problem

Generally, the planning problem representation based on PDDL [9] can be described as a six-tuples: (planning problem name, planning domain name, object instance, initial state, goal state, planning time span). For space flight control planning problem, the initial states of space flight control planning problem include the initial state set of satellite platform, the initial state set of ground flight control center system, the initial state set of predictable exogenous events and the internal state of planning domain. Planning domain name is the name which is defined in the planning domain representation. Object instances include all instances of the objects which are defined in the planning domain representation. Planning time

Table 33.4 Definition and description of controlling events

Controlling events	Descriptions
Turn_on_PVC	PVC power is turned on
Turn_on_GAE	GAE power is turned on
Waiting_for_temperature_ok	To wait for normal temperature
GAE_startup	GAE begins working.
Turn_on_SSE	Sun sensor power is turned on
GAE_calibration	GAE is calibrated
Uploading_Yaw_Coefficients	Yaw coefficients is uploaded
Executing_Yaw_Coefficients	Yaw coefficients is working
Changing_Threshold	Threshold of three-axis attitudes are changed
Changing_mode_Submode	The satellite's mode and sub mode is changed
Reset_GAE_integration	GAE integration is reset
Uploading_Fire_Parameters	Firing parameters are uploaded
Mwheel_startup	Momentum wheel begins working
Uploading_Compensation_Torque	Compensation torque for momentum wheel is uploaded
firing_startup	Thrusts begin firing
firing_ending	Thrusts firing is turned off
memory_downloading	Satellite memory begins downloading
Changing_mode_submode_back	Satellite controlling mode is transferred into Normal
waiting_mode_back	To wait for being normal mode
Modifying_pulse_width	The pulse wide of momentum wheel upload is modified
uploading_Mwheel_uploading_threshold	The threshold of momentum wheel upload is uploaded
GAE_closing	The Gyro is turned off
GAE_turnoff	GAE is turned off
Cancel_yaw_coefficient	The yaw coefficient of satellite attitude is cancel
SSE_turnoff	SSE is turned off
PVC_turnoff	PVCis turned off

span indicates the planning period. Figure 33.1 shows the planning problem representation of east-west station-keeping control planning in NM mode and visible Sun sensor.

33.5 Problem-Solving Framework

The problem-solving framework of general intelligent planning problem is divided into planning representation and the planner [10], as shown in Fig. 33.2. A planning problem can be described as a binary group D = (Dom,Prob): which Dom is the

```
(define (problem stationkeep-east_west-skm-sun_avail) ::problem name
(:domain stationkeep-east_west)::domain name
(:objects
   ::object instances
    satellite0 – satellite
    pvc_a – PVC
    sse_a – SSE
    mw_a – Mwheel
    thrusts_a – Thrusts
    normal – Mode
    skm – Mode
    sm_0 – Submode
    sm_3 – SubMode
    gae_a – GAE
)
(:init
    ::spacecraft initial states
    (working normal sm_0 satellite0)
    (=(slew_time sm_0 sm_3 normal skm satellite0) 16)
    (=(slew_time sm_3 sm_0 skm normal satellite0) 16)
    (=(waiting_GAE_time satellite0) 3600)
    (on_board1 pvc_a satellite0)
    (on_board2 sse_a satellite0)
    (on_board3 mw_a satellite0)
    (on_board4 thrusts_a satellite0)
    (on_board5 gae_a satellite0)
    (power_avail1 pvc_a satellite0)
    (power_avail2 sse_a satellite0)
    (power_avail3 gae_a satellite0)
      ::initial states of ground system
    (GAE_galibration_ready satellite0)
    (=(fire_time satellite0) 600)
    ::initial states of predictable exogenous events
    (at 5640.00 (TL2 satellite0))
    (at 6000.00 (TL3 satellite0))
    (at 6600.00 (TL4 satellite0))
    ::internal states of planning domain
    (controlling_model_submodelskm sm_3 satellite0)
    (controlbace_mode_submodel normal sm_0 satellite0)
    (attitude_caculate_working satellite0)
    (new_threshold_cancel satellite0)
    (at 5000.00 (transfer_submode_ready satellite0))
)
(:goal  (and
(EWSK_SUCCESS satellite0) ::goal states
))
(:metric minimize(total-time))::time span
)
```

Fig. 33.1 The problem representation of east-west station-keeping control in NM mode and visible Sun sensor

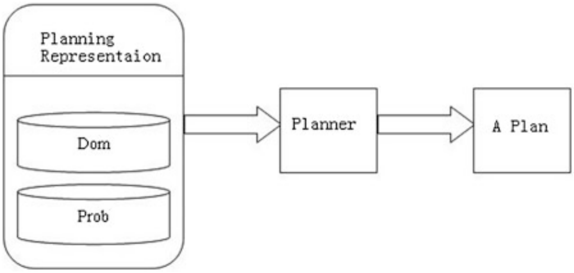


Fig. 33.2 General problem-solving framework of AI planning problem

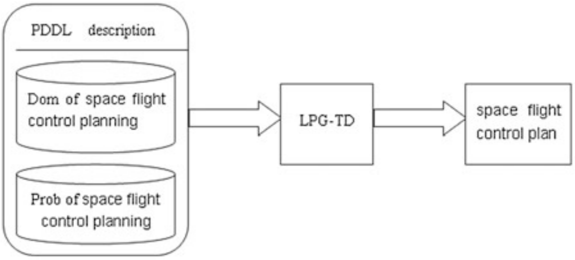
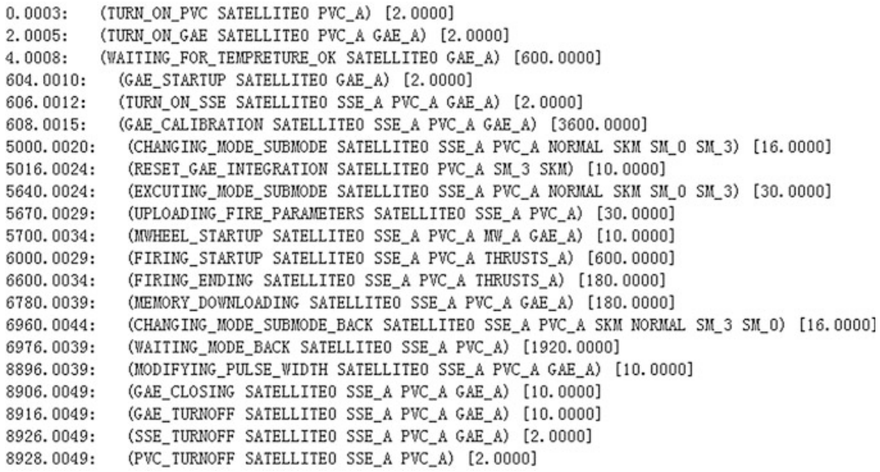


Fig. 33.3 The problem-solving framework of space flight control planning based on PDDL

domain representation file, Prob is problem representation file. This modular description method can repeatedly use the domain representation file, and different planning problem instances only need to define different problem representation files. Planner is a important component of the planning problem solving architecture based on PDDL. The input of the planner is the domain representation file and problem representation file. The output of planner is a plan, which is made up of start time, parameters instance, action duration), and the start time can be relative time or absolute time.

For space flight control planning problem, planning problem representation is described by the PPDL method given in Sects. 33.3 and 33.4, and planner applies local search algorithm based on constraint inference incorporating graph planning framework with predictable exogenous event constraints(LPG-TD) [11]. This paper gives the problem-solving framework of space flight control planning based on PDDL in Fig. 33.3.

Taking east-west station-keeping control for example, the controlling plan is generated by using the problem-solving framework of space flight control planning based on PDDL. The name of domain representation file is stationkeep-east_west.pddl, and the name of problem representation file is stationkeep-east_west-skm-sun_notaval.pddl. Figure 33.4 gives the east-west station-keeping control plan.Seen from the solving result, the action sequence of solution plan is consistent with east-west station-keeping control, and it further proves that the solution algorithm of LPG-TD is feasible for space flight control planning problem.



```
0.0003:    (TURN_ON_PVC SATELLITE0 PVC_A) [2.0000]
2.0005:    (TURN_ON_GAE SATELLITE0 PVC_A GAE_A) [2.0000]
4.0008:    (WAITING_FOR_TEMPRETURE_OK SATELLITE0 GAE_A) [600.0000]
604.0010:    (GAE_STARTUP SATELLITE0 GAE_A) [2.0000]
606.0012:    (TURN_ON_SSE SATELLITE0 SSE_A PVC_A GAE_A) [2.0000]
608.0015:    (GAE_CALIBRATION SATELLITE0 SSE_A PVC_A GAE_A) [3600.0000]
5000.0020:    (CHANGING_MODE_SUBMODE SATELLITE0 SSE_A PVC_A NORMAL SKM SM_0 SM_3) [16.0000]
5016.0024:    (RESET_GAE_INTEGRATION SATELLITE0 PVC_A SM_3 SKM) [10.0000]
5640.0024:    (EXCUTING_MODE_SUBMODE SATELLITE0 SSE_A PVC_A NORMAL SKM SM_0 SM_3) [30.0000]
5670.0029:    (UPLOADING_FIRE_PARAMETERS SATELLITE0 SSE_A PVC_A) [30.0000]
5700.0034:    (MWHEEL_STARTUP SATELLITE0 SSE_A PVC_A MW_A GAE_A) [10.0000]
6000.0029:    (FIRING_STARTUP SATELLITE0 SSE_A PVC_A THRUSTS_A) [600.0000]
6600.0034:    (FIRING_ENDING SATELLITE0 SSE_A PVC_A THRUSTS_A) [180.0000]
6780.0039:    (MEMORY_DOWNLOADING SATELLITE0 SSE_A PVC_A GAE_A) [180.0000]
6960.0044:    (CHANGING_MODE_SUBMODE_BACK SATELLITE0 SSE_A PVC_A SKM NORMAL SM_3 SM_0) [16.0000]
6976.0039:    (WAITING_MODE_BACK SATELLITE0 SSE_A PVC_A) [1920.0000]
8896.0039:    (MODIFYING_PULSE_WIDTH SATELLITE0 SSE_A PVC_A GAE_A) [10.0000]
8906.0049:    (GAE_CLOSING SATELLITE0 SSE_A PVC_A GAE_A) [10.0000]
8916.0049:    (GAE_TURNOFF SATELLITE0 SSE_A PVC_A GAE_A) [10.0000]
8926.0049:    (SSE_TURNOFF SATELLITE0 SSE_A PVC_A GAE_A) [2.0000]
8928.0049:    (PVC_TURNOFF SATELLITE0 SSE_A PVC_A) [2.0000]
```

Fig. 33.4 The produced east-west station-keeping control plan

33.6 Conclusion

The application of AI theories and approaches has became the developing trend of high efficient space flight control center system. Space flight control AI planning is a good application [12]. An AI planning method based on PDDL is proposed for space flight control planning problem in this paper. Different from rule-based AI planning method, this method is to separate domain knowledge from searching and control information of this planning problem, and space flight control planning problem is described based on PDDL and solved by LPG-TD. The example of east-west station-keeping control given in the paper shows that this method is feasibility, and next we will demonstrate its adaptability by wide examples in space flight control planning field.

Acknowledgments The planner LPG-td was developed by Alfonso Gerevini, etc, and its code is available in http://prometeo.ing.unibs.it/lpg. I would like to appreciate you and your code to provide contributions for my research.

References

1. Truszkowski W (2009) Autonomous and autonomic operations and exploration systems-with applications to NASA intelligent spacecraft systems. Springer, Berlin
2. Hao Y (2004) Space flight TT&C network. National Defense Industry Press, Beijing
3. Gutiérrez LG, Tejo JA, Pereda M et al (2006) Flexplan: deployment of powerful comprehensive mission planning systems. In: SpaceOps conference, AIAA 2006-5679
4. Haddow CR, Adamson K, Sousa B et al (2010) Mission planning—Establishing a common concept for ESOC's missions. In: SpaceOps conference, AIAA 2010-1969

5. Lee BS, Kim JH (2003) Design and implementation of the mission planning functions for the KOMPSAT-2 mission control element. J Astron Space Sci 20(3):227–238
6. Hall S, Moreira F, Franco TG (2008) Operations planning for the Galileo constellation. In: SpaceOps conference, AIAA 2008-3542
7. Zheng X (2007) Study on mission planning of spaceflight applying artificial intelligence. Acta Aeronautica Astronaut Sin 28(4):791–795
8. Gu WX, Yan MH, Xu L (2010) Intelligent planning and recognition. Science Press, Beijing
9. Edelkamp S, Hoffmann J (2004) PDDL 2.2: the language for the classical part of the 4th international planning competition. Technical report, IPC
10. Ghallab M, Nau D, Traverso P (2004) Automated planning: theory and practice. Morgan Kaufmann, Los Altos
11. Gerevini A, Saetti A, Serina I (2006) An approach to temporal planning and scheduling in domains with predictable exogenous events. J Artif Intell Res 25(5):187–231
12. LaVallee DB, Jacobsohn J, Olsen C et al (2006) Intelligent control for spacecraft autonomy - an industry survey. In: Space conference, AIAA 2006-7384

Chapter 34
Operation Orbit Options for Asteroid Explorations

Xiaosheng Xin, Lin Liu and Shengxian Yu

Abstract For the exploration of asteroids, especially those with small masses, there are several options for the probe: flybying the asteroid, becoming a satellite of the asteroid, or formation flying with the asteroid. This study focuses on the last two kinds of orbits. For the circulate orbit, the model of the asteroid's gravity field is an important issue. Difficulties will be encountered when constructing the analytical solutions to the circulate orbits. Two kinds of formation flying orbits are considered in this study: (1) around the collinear libration points (CLPs) of the Sun-asteroid system, and (2) directly around the asteroid itself. The asteroid Apophis is used as an example throughout the paper. For the first formation flying strategy, the operation orbits (Lissajous orbits and halo orbits) around the CLP L_1 are firstly constructed in the circular restricted three-body problem (CRTBP), and then generalized to the real force model. For the second formation flying strategy, the operation orbits are firstly constructed using the well-known C-W equation, and then also generalized to the real force model, where the asteroid's true orbit and the solar radiation pressure (SRP) are considered. Studies in this work may be used as references for future asteroid missions.

Keywords Asteroid · Circulate orbit · Formation flying orbit · Libration point · C-W equation

34.1 Introduction

When exploring asteroids, especially those with small masses, one problem is whether the weak gravity field can capture the spacecraft to a satellite orbit around the asteroids. For asteroids with considerable masses, this problem usually does not

---

X. Xin · L. Liu (✉)
School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China
e-mail: lliu@nju.edu.cn

S. Yu
Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_34

exist, but another problem is the shape and mass distribution of the asteroids. Generally, asteroids have irregular shapes quite different from the spheres. One example is, the Amor type asteroid, 433 Eros. It has a size of 34.4 × 11.2 × 11.2 km [1]. Another example is the asteroid 25143 Itokawa, which was visited by the Japanese spacecraft Hayabusa launched in May 2003. The asteroid's orthogonal axes are 533, 294 and 209 m [2]. Due to the shape irregularity and possible inhomogeneous mass distribution, the asteroids' gravity fields may be quite different from those of the planets, especially for regions close to the bodies. Many new problems appear, such as modelling of the asteroid gravity field, orbit dynamics (shape, equilibrium points, stability, etc.) in the irregular shaped gravity field with other possible external perturbations (solar and planetary tidal perturbation, SRP). These problems should be considered in designing the orbit, and the guidance, navigation and control (GNC) of the spacecraft, and the ground TT&C operations.

Many works have been done about these problems [3–6]. In this contribution, however, we only briefly discuss the difficulties associated with the construction of analytical solutions to the circulate orbits and then focus on the formation flying orbits. The formation flying orbits are especially suitable and practical for asteroids with weak gravity fields, because usually it is difficult for these asteroids to host a spacecraft as a stable satellite.

Two kinds of formation flying orbits exist. One strategy is that the spacecraft orbit around the asteroid's collinear libration points (L_1 or L_2). In this case the mass of the asteroid is considered and the periodic or quasi-periodic orbits are computed in the Sun-asteroid circular restricted three-body problem (CRTBP). The other strategy is that the spacecraft directly fly around the asteroid itself. The mass of the asteroid is neglected and together with the spacecraft they make formation flying around the Sun. In fact, since the asteroid's mass is extremely small compared with that of the Sun, formation flying orbits around the asteroid itself and around its libration points do not differ much with each other in geometry. For example, according to the data obtained from the Minor Planet Center (MPC) [1] about the Eros gravitational constant $GM = 4.463 \times 10^{-4}\ \text{km}^3\ \text{s}^{-2}$ and other related orbit parameters, it can be calculated that its libration point L_1/L_2 only deviates about 1,600 km from its center. This distance tends to decrease for asteroids with smaller masses.

These two kinds of formation flying orbits have no fundamental difference. Mathematically, they can be computed with the same method. The only difference lies in whether or not the mass of the asteroid is considered. The formation flying orbits directly around the asteroid neglect the asteroid's mass while the formation flying orbits around the collinear libration points consider it. We still treat them as distinct cases in this paper for clarity.

34.2 Difficulties of Constructing Analytical Solutions to Circulate Orbits

The circulate orbits for proximity exploration of asteroids are mainly affected by the non-spherical and third-body (the Sun and major planets) gravitation. The treatment of the latter force is relatively easy. But problems would be encountered when computing the former one. The first problem is the modelling of the gravity field. We will not elaborate on the different methods here but examine the main physical parameters of the gravity field of an example asteroid 433 Eros [1], which are shown in Tables 34.1 and 34.2.

For comparison, the gravitational constants of the Sun and the Earth are listed as follows:

$$GS = 1.32712442076 \times 10^{11}\,\mathrm{km}^3\,\mathrm{s}^{-2}, \quad GE = .986004418 \times 10^{5}\,\mathrm{km}^3\,\mathrm{s}^{-2}$$

Table 34.1 The gravity field model for asteroid 433 Eros ($R_0 = 16.0\,\mathrm{km}$)

l, m	$\overline{C}_{l,m}$	$\overline{S}_{l,m}$
2, 0	−0.052478	0
2, 1	0	0
2, 2	0.082538	−0.027745
3, 0	−0.001400	0
3, 1	0.004055	0.003379
3, 2	0.001792	−0.000686
3, 3	−0.010337	−0.012134
4, 0	0.012900	0
4, 1	−0.000106	0.000136
4, 2	−0.017495	0.004542
4, 3	−0.000319	−0.000141
4, 4	0.017587	−0.008939

Table 34.2 Main physical parameters of asteroid 433 Eros

Parameters	Values
Volume	$2503 \pm 25\,\mathrm{km}^3$
Bulk density	$2.67 \pm 0.03\,\mathrm{g\,cm}^{-3}$
Mass	$(6.6904 \pm 0.0030) \times 10^{15}\,\mathrm{kg}$
GM (optical radiometric)	$(4.4631 \pm 0.0003) \times 10^{-4}\,\mathrm{km}^3\,\mathrm{s}^{-2}$
GM (radiometric)	$(4.4584 \pm 0.0030) \times 10^{-4}\,\mathrm{km}^3\,\mathrm{s}^{-2}$
GM (radiometric and optical pole)	$(4.4621 \pm 0.0015) \times 10^{-4}\,\mathrm{km}^3\,\mathrm{s}^{-2}$
Rotation rate	$1639.38922 \pm 0.00020\,^\circ/\mathrm{day}$

From the values of the main terms of the gravity field $C_{2,0}$ and $C_{2,2}$, it is apparent that the characteristics like that of the Earth gravity field no longer exists. Though difficulties can be overcome to construct the perturbation solutions under certain circumstances, the large values of the oblateness $(C_{2,0})$ and equatorial ellipticity ($C_{2,2}$ and $S_{2,2}$) would cause trouble when determining the number of terms needed for the small parameter power series solutions. In addition, the nonautonomous tesseral terms $C_{2,2}$ and $S_{2,2}$ also pose great challenges for the construction of the analytical orbit solutions. In the worst case, when the orbit eccentricity of the spacecraft is too large, they would cause the construction process to fail [7].

The above discussion is based on the assumption that the asteroid's gravity field can still be modeled with spherical harmonics. For those asteroids with extremely irregular shape and mass distribution that spherical harmonics are almost impossible to approximate, other approaches are needed.

34.3 Formation Flying Orbit Option—Orbit Around the Libration Points

For the formation flying of two earth satellites in close proximity, both their masses can be neglected which is the basic theoretical priority of solving this problem. However, when both satellites are heavy enough, such as those large GEO satellites (several tons) positioned in the 'same' location (about 100 m apart) above the earth equator, obtaining their precise orbits entail the careful consideration of their mutual gravitation [8]. This can be solved from two different approaches. One method is to add their mutual attraction as one additional perturbation force in their respective force model. The other is a whole new mathematical treatment that involves the model of the CRTBP, which we elaborate in the following.

For asteroids with small masses, such as Eros, orbits around its libration point L_1/L_2, which is close to its center, can serve the purpose of formation flying around the asteroid.

34.3.1 Nominal Orbits Around the Libration Points

In this section, we take the L_1 libration point as an example. We adopt the multiple shooting method to construct its Lissajous and halo orbits in the real force model. The high-order analytical solutions of the Lissajous and halo orbits are taken as initial values for the iteration process. The real force model is based on the Sun-Earth-asteroid CRTBP with additional perturbations by the eight major planets as well as Pluto and the Moon.

34.3.1.1 High-Order Solutions of the Lissajous and Halo Orbits

In the synodic coordinate system where the origin coincides with the barycentre of the two primaries, the equation of motion of the CRTBP [9] is

$$\begin{cases} \ddot{X} - 2\dot{Y} = \Omega_X \\ \ddot{Y} + 2\dot{X} = \Omega_Y \\ \ddot{Z} + Z = \Omega_Z \end{cases} \tag{34.1}$$

$$\Omega = \frac{1}{2}\left(X^2 + Y^2\right) + (1-\mu)r_1^{-1} + \mu r_2^{-1}, \quad \begin{cases} r_1^2 = (X-\mu)^2 + Y^2 + Z^2 \\ r_2^2 = (X-\mu+1)^2 + Y^2 + Z^2 \end{cases} \tag{34.2}$$

where $\mu = M/(S+M)$ is the non-dimensional mass of the asteroid.

We change the origin from the barycentre to the L_1 point and magnify the size of the coordinate by a factor of $1/\gamma$, then

$$x = -\frac{1}{\gamma}(X - \mu + 1 - \gamma), \quad y = -\frac{1}{\gamma}Y, \quad z = \frac{1}{\gamma}Z \tag{34.3}$$

in which γ is the distance between the libration point L_1 and the asteroid. In the new coordinate system, the equation of motion becomes

$$\begin{aligned} \ddot{x} - 2\dot{y} - (1+2c_2)x &= \frac{\partial}{\partial x}\sum_{n \geq 3} c_n r^n P_n\left(\frac{x}{r}\right) \\ \ddot{y} + 2\dot{x} + (c_2 - 1)y &= \frac{\partial}{\partial y}\sum_{n \geq 3} c_n r^n P_n\left(\frac{x}{r}\right) \\ \ddot{z} + c_2 z &= \frac{\partial}{\partial z}\sum_{n \geq 3} c_n r^n P_n\left(\frac{x}{r}\right) \end{aligned} \tag{34.4}$$

where P_n is the Legendre polynomials, $r^2 = x^2 + y^2 + z^2$, and the coefficients of the expansion is given by

$$c_n = \gamma^{-3}\left[\mu + (-1)^n(1-\mu)\left(\frac{\gamma}{1-\gamma}\right)^{n+1}\right] \tag{34.5}$$

The solution of Eq. (34.4) yields the Lissajous orbit around the L_1 libration point, which is represented by the following trigonometric series:

$$\begin{cases} x(t) = \sum\limits_{i,j=1}^{\infty} \left(\sum\limits_{|k| \le i, |m| \le j} x_{ijkm} \cos(k\theta_1 + m\theta_2) \right) \alpha^i \beta^j \\ y(t) = \sum\limits_{i,j=1}^{\infty} \left(\sum\limits_{|k| \le i, |m| \le j} y_{ijkm} \sin(k\theta_1 + m\theta_2) \right) \alpha^i \beta^j \\ z(t) = \sum\limits_{i,j=1}^{\infty} \left(\sum\limits_{|k| \le i, |m| \le j} z_{ijkm} \cos(k\theta_1 + m\theta_2) \right) \alpha^i \beta^j \end{cases} \tag{34.6}$$

where $\theta_1 = \omega t + \phi_1$, $\theta_2 = vt + \phi_2$, ϕ_1, ϕ_2 are constants, $N = i + j$ is the order of the solution. In the summation symbol, $i, j = 1$ signifies $i + j = 1, i \geq 0, j \geq 0$. ω, v can be represented by the series in the form

$$\omega = \sum_{i,j=0}^{\infty} \omega_{ij} \alpha^i \beta^j, \quad v = \sum_{i,j=0}^{\infty} v_{ij} \alpha^i \beta^j \tag{34.7}$$

The higher-order solution can be derived from the lower-order solution. Therefore, the solution obtained from the linearized model can serve as the initial value for the recurrence.

Normally, ω and v are not commensurate with each other. But when the amplitude of in-plane α and out-of-plane β are large enough, some of the combinations of the two amplitudes would yield $\omega = v$. This is the case of periodic orbit in the 3D space, which is also named as halo orbit. The halo orbit can be represented as expansion of the trigonometric series in the form:

$$\begin{cases} x(t) = \sum\limits_{i,j=1}^{\infty} \left(\sum\limits_{|k| \le i} x_{ijk} \cos(k\theta) \right) \alpha^i \beta^j \\ y(t) = \sum\limits_{i,j=1}^{\infty} \left(\sum\limits_{|k| \le i} y_{ijk} \sin(k\theta) \right) \alpha^i \beta^j \\ z(t) = \sum\limits_{i,j=1}^{\infty} \left(\sum\limits_{|k| \le i} z_{ijk} \cos(k\theta) \right) \alpha^i \beta^j \end{cases} \tag{34.8}$$

The detailed derivation already exists in the literature [10–12].

We take asteroid Apophis as an example. On JD = 2455800.5, its orbital elements are

$$\begin{cases} a = 0.9223002432\,AU, & e = 0.1910762290, & i = 3.3319600435^\circ \\ \Omega = 204.4304100445^\circ, & \omega = 126.4244766663^\circ, & M = 287.5823055950^\circ \end{cases},$$

and its mass is $M_a = 2.7 \times 10^{10}$ kg. The non-dimensional units are taken as

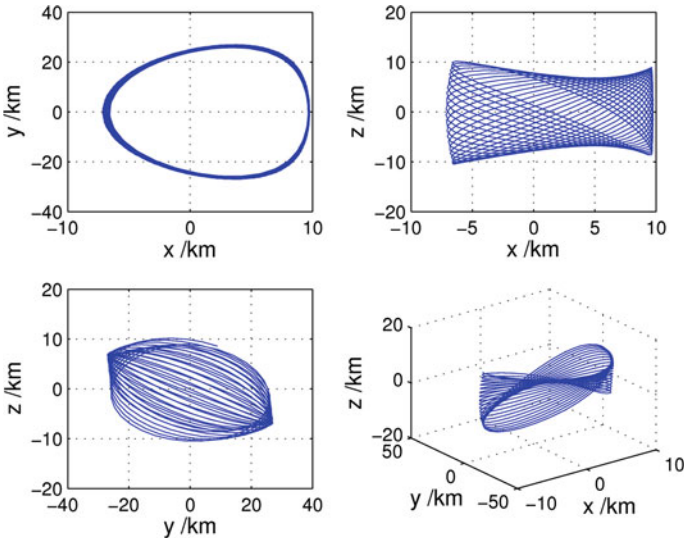


Fig. 34.1 Lissajous orbits around the Apophis L$_1$ libration point ($\alpha = \beta = 2 \times 10^{-8} R_0$)

$$\begin{cases} [L] = R_0 = 0.9223002432 \text{ AU}, \\ [M] = M_{\text{sun}} + M_{\text{a}}, \\ [T] = \left[R_0^3/G(M_{\text{sun}} + M_{\text{a}})\right]^{1/2} \approx 51.490492199842 \text{ d} \end{cases} \tag{34.9}$$

Without special notice, all quantities below are expressed in the non-dimensional units. The orbit insertion time is 00:00:00 March 14, 2018, which corresponds to Julian Date JD = 2458191.5. The orbits around the L$_1$ point are shown here as an example. The Lissajous and halo orbits are shown in Figs. 34.1 and 34.2, respectively, where $\gamma = 1.653975024 \times 10^{-7}$.

34.3.1.2 Numerical Computation of the Lissajous and Halo Orbits in Real Force Model

Taking the third-order analytical solutions as initial values, we find the Lissajous and halo orbits in the real force model via multiple shooting method (i.e. differential correction). The real force model is based on the Sun-Earth-asteroid CRTBP with additional perturbations by the eight major planets as well as Pluto and the Moon, whose ephemerides are given by DE 405. The ephemeris of the asteroid is obtained by integration in this real force model. The Lissajous and halo orbits computed are depicted in Figs. 34.3 and 34.4, respectively.

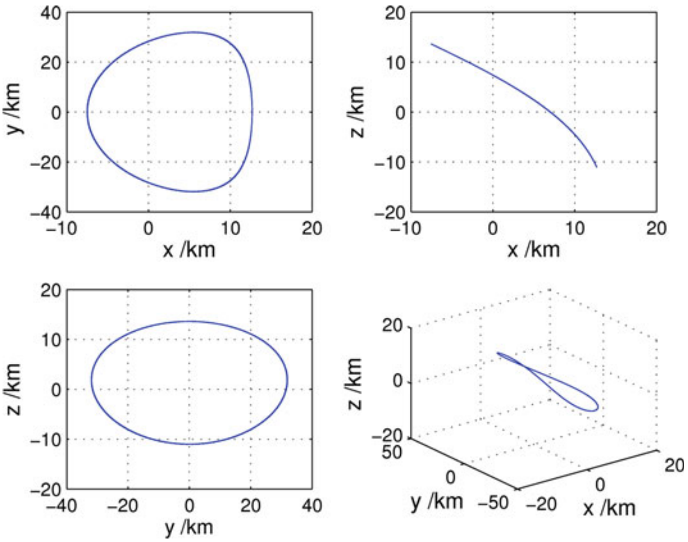


Fig. 34.2 Halo orbits around the Apophis L$_1$ libration point $(\alpha = 2.48 \times 10^{-8} R_0)$

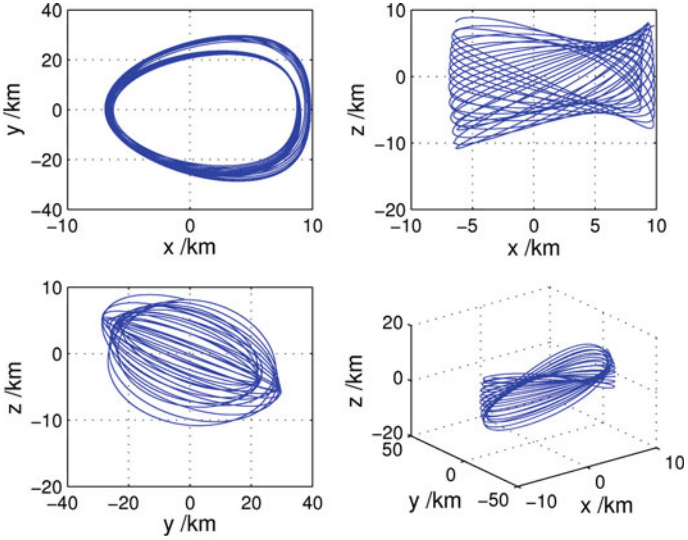


Fig. 34.3 Lissajous orbits around the Apophis L$_1$ libration point in the real force model $(\alpha = \beta = 2 \times 10^{-8} R_0)$

Comparing the orbits obtained by the third-order analytical solutions and those computed in the real force model, we can find that the two kinds of orbits, especially the halo orbits, have larger drifting regions in the real force model scenario

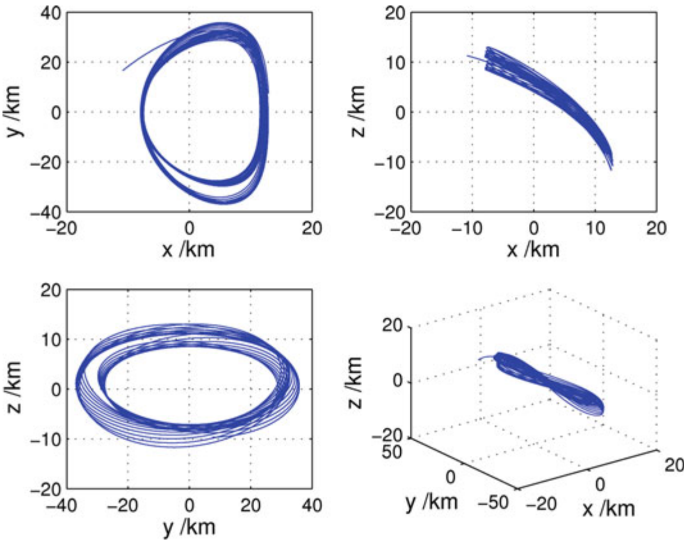


Fig. 34.4 Halo orbits around the Apophis L$_1$ libration point in the real force model ($\alpha = 2.48 \times 10^{-8} R_0$)

due to the perturbations of other celestial bodies. In spite of this, the orbits in the real force model still stay around those produced by the analytical solutions; in other words, they still retain their respective orbit characteristics after certain amount of time.

34.4 Formation Flying Orbit Option—Orbit Around the Asteroid Itself

If the asteroid's mass is so small that it can be neglected, i.e. the mass parameter $\mu = M/(S + M) = 0$, then the above case of orbits around the libration points is degenerated to the case of spacecraft formation flying with the asteroid itself. In this case, the L$_1$ and L$_2$ libration points merge together with the center of the asteroid $(x = 1, y = 0, z = 0)$, and the corresponding equation of motion can be obtained from Eq. (34.4) with $c_2 = 1$ [13]. The linearized form is just the C-W equation [14] well-known in the satellite formation flying research. The coordinate symbols are denoted as (ξ, η, ζ) instead of (x, y, z) for discrimination. The linearized equation reads

$$\ddot{\xi} - 2\dot{\eta} = 3\xi, \quad \ddot{\eta} + 2\dot{\xi} = 0, \quad \ddot{\varsigma} + \varsigma = 0 \tag{34.10}$$

34.4.1 Solutions of Conditional Periodic Orbits in the Linearized Circumstance

The solution of Eq. (34.10) is

$$\begin{cases} \xi = -\dfrac{2}{3}C_2 + \dfrac{1}{2}C_3 \sin t - \dfrac{1}{2}C_4 \cos t \\ \dot{\xi} = \dfrac{1}{2}C_3 \cos t + \dfrac{1}{2}C_4 \sin t \\ \eta = C_1 + C_2 t + C_3 \cos t + C_4 \sin t \\ \dot{\eta} = C_2 - C_3 \sin t + C_4 \cos t \end{cases} \tag{34.11}$$

The motion in the direction of ς is a simple harmonic oscillation. With proper initial values, by setting $C_1 = C_2 = 0$, the conditional periodic solution can be obtained. The initial conditions at time t_0 should satisfy

$$\xi_0,\ \eta_0,\ \dot{\xi}_0 = \eta_0/2,\ \dot{\eta}_0 = -2\xi_0,\ \varsigma_0,\ \dot{\varsigma}_0 \tag{34.12}$$

and the solution is

$$\begin{cases} \xi = \xi_0 \cos t + (\eta_0/2) \sin t, & \dot{\xi} = -\xi_0 \sin t + (\eta_0/2) \cos t \\ \eta = -2\xi_0 \sin t + \eta_0 \cos t, & \dot{\eta} = -2\xi_0 \cos t - \eta_0 \sin t \\ \varsigma = \varsigma_0 \cos t + \dot{\varsigma}_0 \sin t, & \dot{\varsigma} = -\varsigma_0 \sin t + \dot{\varsigma}_0 \cos t \end{cases} \tag{34.13}$$

Its projection on the $\xi - \eta$ plane is an ellipse. In addition, if $\dot{\varsigma}_0$ also satisfies $\dot{\varsigma}_0 = \pm(\eta_0/2\xi_0)\varsigma_0$, the projection on the $\eta - \varsigma$ is also an ellipse. Taking $\xi_0 = \eta_0 = \varsigma_0 = 1 \times 10^{-4} R_0$, we can get the orbit shown in Fig. 34.5. Various formation flying orbits with different configurations can be computed by modifying the initial conditions. The conditional periodic solution (34.13) is later taken as the initial orbit for iteration.

34.4.2 Numerical Result of Formation Flying Around Asteroid Itself in the Real Force Model

In the rotating coordinate system centered on the asteroid, the equation of motion of the spacecraft is

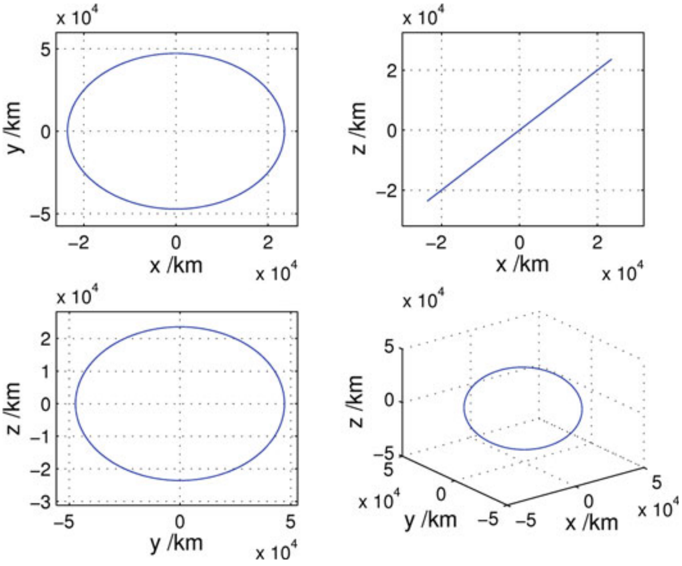


Fig. 34.5 Conditional periodic orbits around Apophis ($\xi_0 = \eta_0 = \varsigma_0 = 1 \times 10^{-4} R_0$)

$$
\begin{aligned}
&\ddot{\boldsymbol{\rho}} = \mathbf{F}_1 + \mathbf{F}_2 \\
&\mathbf{F}_1 = -2C^T\dot{C}\dot{\boldsymbol{\rho}} - C^T\ddot{C}\boldsymbol{\rho} - \mu_{11}\mathbf{r}/r^3 + \mu_{11}\mathbf{r}_{12}/r_{12}^3 - \sum_{i\in S, i\neq 11,12} \mu_i\left(\boldsymbol{\delta}_i/\delta_i^3 - \boldsymbol{\delta}_i^{12}/\left(\delta_i^{12}\right)^3\right) \\
&\mathbf{F}_2 = -\mu_{11}\mathbf{r}_{12}/r_{12}^3 - \sum_{i\in S, i\neq 11,12} \mu_i\left(\boldsymbol{\delta}_i^{12}/\left(\delta_i^{12}\right)^3 + \mathbf{r}_i/r_i^3\right) - \ddot{\mathbf{r}}_{12} - 2C^T\dot{C}\dot{\mathbf{r}}_{12} - C^T\ddot{C}\mathbf{r}_{12}
\end{aligned}
\tag{34.14}
$$

where $\mathbf{r}, \mathbf{r}_i$ is the position vector in the Sun-centered rotating system, $\mathbf{r}$ is the position vector of the spacecraft, and $\mathbf{r}_i$ is the position vector of the major planet μ_i. $\boldsymbol{\rho}(\xi, \eta, \zeta) = \mathbf{r} - \mathbf{r}_{12}$ is the position vector of the spacecraft relative to the asteroid, and $\boldsymbol{\delta}_{12}^i = \mathbf{r}_{12} - \mathbf{r}_i$ is the position vector of the asteroid relative to the major planet μ_i. C is the transformation matrix between the Sun-centered rotating system and the Sun-centered inertial coordinate system.

Using the orbit obtained in Sect. 34.4.1 as initial value, we have computed the corrected conditional quasi-periodic orbit in the real force model via multiple shooting method, which is shown in Fig. 34.6.

Figure 34.6 shows the change of the orbit over 8.458 years. It can be seen that although perturbations of major planets and orbit evolution of the asteroid are considered in the real force model the corrected orbit still stays in the vicinity of the orbit calculated from the solution of linearized model, only with minor drift relative to the initial orbit.

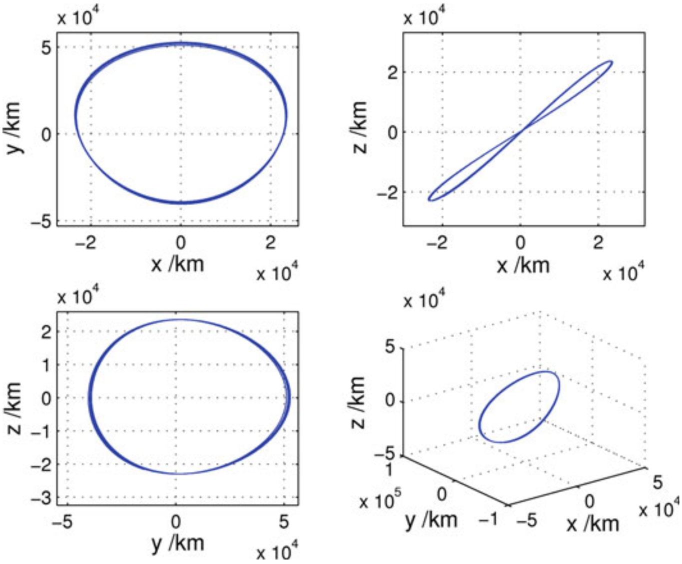
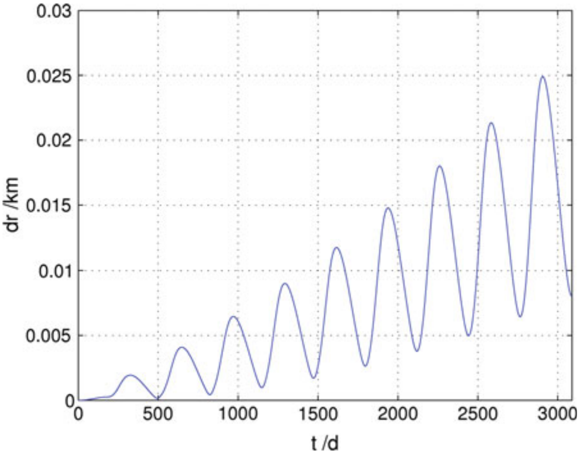


Fig. 34.6 Conditional quasi-periodic orbits around Apophis in the real force model ($\xi_0 = \eta_0 = \varsigma_0 = 1 \times 10^{-4} R_0$)

The above computation does not take into account the gravity effect of the asteroid on the spacecraft. Taking this factor into consideration as a perturbation, with the same initial orbit for ephemeris integration, the difference between the resultant 'true' orbit and the nominal orbit in Fig. 34.6 is shown in Fig. 34.7.

Figure 34.7 tells us that the effect of neglecting Apophis' mass on the orbit with altitude of $O(10^4)$ km over 8.458 years is on the order of $O(10^2)$ m. Therefore, the nominal orbit is not influenced much without considering the asteroid's mass.

Fig. 34.7 Difference between the 'true' orbit considering the Apophis mass and the nominal orbit



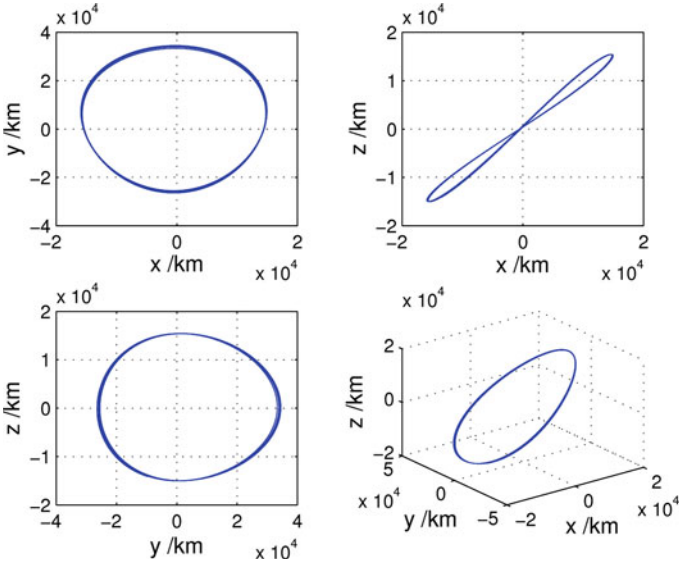


Fig. 34.8 Conditional periodic orbit around Apophis in the real force model with SRP perturbation. $(\xi_0 = \eta_0 = \varsigma_0 = 1 \times 10^{-4} R_0)$

34.4.3 Effect of the SRP Perturbation on Formation Flying Orbit

The origin of the formation flying orbit is still the asteroid's center after taking the perturbation of SRP into account. As a result, the orbit calculated from the analytical solution is the same as in Fig. 34.6. For the real force model, the perturbation of SRP should be added to the model to numerically correct the initial orbit. The final orbit is shown in Fig. 34.8.

34.5 Conclusion

In this paper, we proposed two kinds of orbits for asteroid explorations, the formation flying orbits around the collinear libration points of the asteroid and the formation flying orbits around the asteroid itself. Which kind of orbits should be used depends on the mass of the asteroid. Moreover, under certain circumstances, the SRP perturbation should also be considered, but no differences appear in the numerical method described in the paper.

Orbit control to the two kinds of formation flying orbits is necessary, because both of them are unstable. We also did some work on this topic, with two possible approaches: continuous low-thrust and solar sail. We leave this topic for future discussions.

References

1. Miller JK et al (2002) Determination of shape, gravity, and rotational state of asteroid 433 Eros. Icarus 155(1):3–17
2. Saito J et al (2006) Detailed images of asteroid 25143 Itokawa from Hayabusa. Science 312 (5778):1341–1344
3. Werner RA, Scheeres DJ. Exterior gravitation of a polyhedron derived and compared with harmonic and mascon gravitation representations of asteroid 4769 Castalia. Celestial Mech Dyn Astron 65(3):313–344
4. Scheeres DJ (1994) Dynamics about uniformly rotating triaxial ellipsoids: applications to asteroids. Icarus 110(2):225–238
5. Scheeres DJ et al (1996) Orbits close to asteroid 4769 Castalia. Icarus 121(1):67–87
6. Scheeres DJ et al (1996) Dynamics of orbits close to asteroid 4179 Toutatis. Icarus 132 (1):53–79
7. Liu L (2000) Orbit theory of spacecraft. Defense industry press, Beijing
8. Zhang Q, Liu L (1999) On the binary star system of geostationary satellites. J Nanjing Univ 35:7–13
9. Liu L, Hou XY (2012) Deep space spacecraft orbit dynamics. Publishing House of Electronics Industry, Beijing
10. Jorba A, Masdemont J (1999) Dynamics in the center manifold of the collinear points of the restricted three body problem. Phys D: Nonlinear Phenom 132(1):189–213
11. Richardson DL (1980) Analytic construction of periodic orbits about the collinear points. Celest Mech 22(3):241–253
12. Kim M (2001) Periodic spacecraft orbits for future space-based deep space observations. Technical report, Vienna Institute of Technology, Vienna
13. Liu L, Wang HH, Ma JB (2004) On the formation flying of satellite constellations. Chin Astron Astrophys 28(2):188–199
14. Clohessy WH (1960) Terminal guidance system for satellite rendezvous. J Aerosp Sci 27:653–658

Chapter 35
Real-Time Estimation of Maneuvering Orbit by Compensative Modeling of "Current" Acceleration

Ye Liu, Jianfeng Cao and Lei Liu

Abstract The unknown maneuver is a key restriction for the real-time spacecraft tracking. For the modeling residual by dynamic of orbits, a compensative modeling method is developed according to the idea of "current" statistic. Then an orbit filter is established based on the new model and the unscented filter. By utilizing a non-mean value and time correlated stochastic process for the maneuvering acceleration, the new model can achieve an online auto-adaptivity for orbit acceleration with different maneuvers mode. Therefore, it can obtain a continuous and accurately estimated orbit for orbit with unknown maneuver. Simulation results for space target with series of random pulsar and continuous maneuvers validated the feasibility and efficiency of the new tracking filter.

Keywords Orbit maneuvering · "Current" statistic · Acceleration · Real-time

35.1 Introduction

It is not an easy task for tracking spacecraft with high maneuverability or wide range of dynamics [1]. This is will be more difficult for a non-cooperative target since the maneuver character cannot be obtained in time. For the discontinuous or rapidly changed acceleration, there is a mismatch for the ordinary ballistic model [2] or dynamic model [3]. This modeling error will lead to a tracking accuracy decline or even tracking loss [4].

The model-decision method is originated in the maneuvering target tracking problem. The basis of the method is that the dynamic mode changing of the system state can be reflected in the measurement or estimated state. Maneuver detection

---

Y. Liu (✉) · J. Cao · L. Liu
Science and Technology on Aerospace Flight Dynamic Laboratory, Beijing 100094, China
e-mail: liuye_new@sina.com

Y. Liu · J. Cao · L. Liu
Beijing Aerospace Control Center, Beijing 100094, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_35

and maneuver identification are two common used model-decision techniques. Both of them can be applied in nonlinear filters which are widely applied for spacecraft tracking. The maneuver detection can give a reference to the model switch by detecting the acceleration variation. This depends on the exact modeling for different flying phases as well the construction of test statistic [5]. Improper statistic will cause false alarm or missing alarm. The maneuver identification aims to make up an auto-decision model by fusion the detecting and estimating step, where adaptive model parameters or model structure is utilized [6]. Such as the Singer model, the Noval statistic model, the Semi-Markov Jump Process model, the "current" statistic model (CS) [3]. As one of the widely applied kinetic model, the CS model utilized a Markov Model with adaptive mean acceleration and can rapidly identical the currently maneuver [3, 6]. It is also applied in the boost phase of a rocket-powered BT target [5]. According to the smoothness of the flight, some semi- parametric or non-parametric modeling algorithms are also used for orbit estimation problem [7]. However, most of the identical methods focus on the smooth transition while maneuvers occur. They may not give an accurate depiction as the dynamic modeling dose for target without maneuver, and thus take a decline of tracking accuracy during the un-maneuvering flight phase.

Therefore, the paper focuses on the modeling method by integrated utilizing the accuracy of the dynamic modeling and the adaptability of the model-decision technique. A new tracking filter with compensative modeling for the maneuvering acceleration is discussed referring to the idea of "current" statistic in Sects. 35.2 and 35.3. This may obtain a continuous and accurate tracking for spacecraft with complicated maneuver mode, which is validated by simulations in Sect. 35.4.

35.2 Maneuver Orbit Modeling

35.2.1 Review of the "Current" Statistical Modeling for Maneuvering Target

The CS model was first suggested by Zhou [6] for benchmark tracking problem. He believes that the acceleration can just be in the neighborhood of the current acceleration, and the neighborhood is limited according to the maneuver ability. Then, the maneuver acceleration is supposed as a time correlated stochastic process with nonzero adaptive mean and conditional Rayleigh probability density.

Let $\boldsymbol{X} = (x, y, z, \dot{x}, \dot{y}, \dot{z}, \ddot{x}, \ddot{y}, \ddot{z})^T$ be the state vector at time step k. The acceleration is assumed decoupled along the three dimensions in the CS model. Take x-orientation as example, the acceleration is expressed as a first-order stationary Markov process which can be written as

$$\ddot{x}(t) = \overline{\lambda}(t) + \lambda(t) \tag{35.1}$$

where $\lambda(t)$ is a zero-mean colored noise and satisfies $\dot{\lambda}(t) = -\alpha\lambda(t) + w(t)$. The mean value $\overline{\lambda}(t)$ is considered to be constant during one sample period, and can be taken as the predicted acceleration $\overline{x}(t)$ referred to [3]. α is the maneuver frequency, and $w(t)$ is a zero-mean Gaussian white noise with covariance as $\sigma_w^2 = 2\alpha\sigma_a^2$. Then, we can obtain the state-space formula for CS as

$$\begin{bmatrix} \dot{x}(t) \\ \ddot{x}(t) \\ \dddot{x}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\alpha \end{bmatrix} \begin{bmatrix} x(t) \\ \dot{x}(t) \\ \ddot{x}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \alpha \end{bmatrix} \overline{\lambda}(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(t) \tag{35.2}$$

To carry more accurate information, the probability density of the acceleration is replaced by a conditional density $f(\ddot{x}|\widehat{\ddot{x}}_k)$, where $\widehat{\ddot{x}}_k$ is the estimated acceleration at time step k. Conditional Rayleigh density is often proposed for the maneuver acceleration in the CS model [3, 6]. But in extreme case when estimated accelerations beyond prior range, the calculated covariance may be too big to applied. Therefore, we give a modified covariance equation as

$$\sigma_k^2 = \begin{cases} (4-\pi)(\ddot{x}_{\max} - \hat{\ddot{x}}_k)^2/\pi, & \ddot{x}_{\max} \geq \hat{\ddot{x}}_k > 0 \\ (4-\pi)(\ddot{x}_{-\max} - \hat{\ddot{x}}_k)^2/\pi, & \ddot{x}_{-\max} \leq \hat{\ddot{x}}_k < 0 \\ (4-\pi)\hat{\ddot{x}}_k^2/\pi, & \hat{\ddot{x}}_k > \ddot{x}_{\max} \; or \; \hat{\ddot{x}}_k < \ddot{x}_{-\max} \end{cases} \tag{35.3}$$

where $\ddot{x}_{-\max}$ is the negative acceleration limit, not necessarily equal to the acceleration limit $\ddot{x}_{\max}$. Both of them could be given prior or by an adaptive strategy.

35.2.2 *Maneuvering Orbit Compensative Modeling of "Current" Acceleration*

For earth orbit flying spacecraft, there is accurate dynamic model can be referred for the acceleration calculation

$$\boldsymbol{a}(t) = [a_x(t), a_y(t), a_z(t)]^T = f(x(t), y(t), z(t), \dot{x}(t), \dot{y}(t), \dot{z}(t)) \tag{35.4}$$

Here $f(\cdot)$ is often derived referring to the accuracy requirement and the computational complexity. Take x-orientation as example, the dynamic state equation can be reformulated as

$$\begin{cases} x_{k+1} = x_k + \int\limits_{t_k}^{t_{k+1}} \dot{x}(t)\mathrm{d}t = x_k + \dot{x}_k \cdot (t_{k+1} - t_k) + \int\limits_{t_k}^{t_{k+1}} \int\limits_{t_k}^{t_{k+1}} \ddot{x}(t)\mathrm{d}t\mathrm{d}t \\ \dot{x}_{k+1} = \dot{x}_k + \int\limits_{t_k}^{t_{k+1}} \ddot{x}(t)\mathrm{d}t \\ \ddot{x}_{k+1} = a_{x,k+1} \end{cases} \tag{35.5}$$

If a maneuver is occurring, an unknown component $\boldsymbol{b}_k$ should be added into the acceleration. The maneuver is usually arisen by pulsar and continuous thrust which is a limited value. Therefore, we can modeling $\boldsymbol{b}_k$ according to the "current" statistic technique. Assume that $\boldsymbol{b}_k = [b_{x,k}, b_{y,k}, b_{z,k}]^T$ is decoupled along the three dimensions, then it can be formula as

$$b_{s,k} = \bar{b}_{s,k} + \lambda_{s,k} \quad s = x, y, z \tag{35.6}$$

where λ_s is a zero-mean colored noise and satisfies $\dot{\lambda}_s(t) = -\alpha_s \lambda_s(t) + w_s(t)$. Take $\bar{b}_{s,k}$ as the predicted maneuver acceleration $\ddot{x}_{s,k} - a_{s,k}$. Then we can finally obtain the compensative dynamic model by carrying an integral operation in Eq. (35.2). The general formula of the model in x-orientation is

$$\begin{cases} x_{k+1} = x_k + \int\limits_{t_k}^{t_{k+1}} \int\limits_{t_k}^{t_{k+1}} \big(a_{x,k}(t) + w_x(t)\big)\mathrm{d}t\mathrm{d}t + 0.5 \cdot (\ddot{x}_{s,k} - a_{s,k})(t_{k+1} - t_k)^2 \\ \dot{x}_{k+1} = \dot{x}_k + \int\limits_{t_k}^{t_{k+1}} \big(a_{x,k}(t) + w_x(t)\big)\mathrm{d}t + (\ddot{x}_{s,k} - a_{s,k})(t_{k+1} - t_k) \\ \ddot{x}_{k+1} = a_{x,k+1} + \underset{s,k}{\ddot{x}} - a_{s,k} \end{cases} \tag{35.7}$$

The noise $w_x(t)$ possesses a modified Rayleigh density as in (35.3), where the "current" acceleration is the compensative one, i.e. $\ddot{x}_{s,k} - a_{s,k}$.

35.3 Maneuver Orbit Filter

Let Δt be the sample time, and the state-space equation can be described as by the compensative dynamic model

$$\begin{cases} \boldsymbol{X}_k = f(\boldsymbol{X}_{k-1}) + \mathbf{C} \cdot \boldsymbol{b}_{k-1} + \boldsymbol{w}_{k-1} \\ \boldsymbol{Y}_k = h(\boldsymbol{X}_k) + \boldsymbol{v}_k \end{cases} \tag{35.8}$$

where $\boldsymbol{Y}_k$ is the n-dimension observation vector, $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are zero-mean white noise vector with $\mathbf{P}_{\boldsymbol{w}}$ and $\mathbf{P}_{\boldsymbol{v}}$ as their covariance respectively. The nonlinear function $f(\cdot)$ is the first item in the right of first equation in (35.7). $\boldsymbol{b}_k$ denotes the maneuvering acceleration as the third item in the right of first equation in (35.7), and C is the coefficient matrix. For commonly applied instruments, the observe equation $h(\cdot)$ is nonlinear too.

The system (35.8) is usually solved by a nonlinear filter, such as the extended Kalman filter (EKF), the unscented filter (UKF) and the particle filter (PF) et al. [5]. The following algorithm is derived referring to the UKF [8] for real-time as well as high accuracy purpose.

Let $\boldsymbol{X}_{k-1|k-1}$ be the state estimation with covariance $\boldsymbol{P}_{\boldsymbol{X},k-1}$ at time step $k-1$. Take sigma points X_{k-1}^i with coefficients W_i^M and W_i^C from the state vector ($i = 1, \ldots, L$) and L is the number of sigma points. The sampling method can be referred to [8]. Then the filter has the expression as follows.

1. State Prediction

$$\begin{cases} \overline{\boldsymbol{X}}_{k|k-1} = \sum\limits_{i=0}^{L} W_i^M \overline{\mathrm{X}}_{k|k-1}^i \\ \overline{\mathbf{P}}_{\boldsymbol{X},k} = \sum\limits_{i=0}^{L} W_i^C (\overline{\mathrm{X}}_{k|k-1}^i - \overline{\boldsymbol{X}}_{k|k-1})(\overline{\mathrm{X}}_{k|k-1}^i - \overline{\boldsymbol{X}}_{k|k-1})^T + \mathbf{P}_{\boldsymbol{w},k} \end{cases} \tag{35.9}$$

where $\boldsymbol{b}_{k-1} = \hat{\ddot{x}}_{k-1} - \boldsymbol{a}_{k-1} = \hat{\ddot{x}}_{k-1} - f(x_{k-1}, y_{k-1}, z_{k-1}, \dot{x}_{k-1}, \dot{y}_{k-1}, \dot{z}_{k-1})$, $\overline{\mathrm{X}}_{k|k-1}^i = f(\mathrm{X}_{k-1}^i) + \mathbf{C} \cdot \boldsymbol{b}_{k-1}$, $\overline{\boldsymbol{X}}$ and $\overline{\mathbf{P}}_{\boldsymbol{X},k}$ are the mean and covariance of the predicted state. $\mathbf{P}_{\boldsymbol{w}}$ is the covariance of the modified conditional Rayleigh density which can be calculated as

$$\mathbf{P}_{\boldsymbol{w}} = \begin{bmatrix} \sigma_{x,k} \cdot \int\limits_{t_k}^{t_{k+1}} \int\limits_{t_k}^{t_{k+1}} \mathbf{F}(t)\mathrm{d}t\mathrm{d}t & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \sigma_{y,k} \cdot \int\limits_{t_k}^{t_{k+1}} \int\limits_{t_k}^{t_{k+1}} \mathbf{F}(t)\mathrm{d}t\mathrm{d}t & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \sigma_{z,k} \cdot \int\limits_{t_k}^{t_{k+1}} \int\limits_{t_k}^{t_{k+1}} \mathbf{F}(t)\mathrm{d}t\mathrm{d}t \end{bmatrix} \tag{35.10}$$

where $\sigma_{x,k}$, $\sigma_{y,k}$ and $\sigma_{z,k}$ can be calculated as in (35.3). The transfer matrix $\mathbf{F}(t)$ is formulated as $\mathbf{F}(t) = \begin{bmatrix} 1 & t & (\alpha t - 1 + \mathrm{e}^{\alpha t})/\alpha^2 \\ 0 & 1 & (1 - \mathrm{e}^{-\alpha t})/\alpha \\ 0 & 0 & \mathrm{e}^{-\alpha t} \end{bmatrix}$.

2. Observation Prediction

$$\begin{cases} \overline{\boldsymbol{Y}}_{k|k-1} = \sum_{i=0}^{L} W_i^M \overline{\mathrm{Y}}_{k|k-1}^i \\ \overline{\mathbf{P}}_{\boldsymbol{Y},k} = \sum_{i=0}^{L} W_i^C (\overline{\mathrm{Y}}_{k|k-1}^i - \overline{\boldsymbol{Y}}_{k|k-1})(\overline{\mathrm{Y}}_{k|k-1}^i - \overline{\boldsymbol{Y}}_{k|k-1})^T + \mathbf{P}_{\boldsymbol{v},k} \\ \overline{\mathbf{P}}_{\boldsymbol{XY},k} = \sum_{i=0}^{L} W_i^C (\overline{\mathrm{X}}_{k|k-1}^i - \overline{\boldsymbol{X}}_{k|k-1})(\overline{\mathrm{Y}}_{k|k-1}^i - \overline{\boldsymbol{Y}}_{k|k-1})^T \end{cases} \tag{35.11}$$

where $\overline{\mathrm{Y}}_{k|k-1}^i = h(\overline{\mathrm{X}}_{k|k-1}^i)$,$\overline{\boldsymbol{Y}}_{k|k-1}$ and $\overline{\mathbf{P}}_{\boldsymbol{Y},k}$ are the mean and covariance of the predicted observation, $\overline{\mathbf{P}}_{\boldsymbol{XY},k}$ is the covariance between the predicted state and observation.

3. State filtering

$$\begin{cases} \mathbf{K}_k = \overline{\mathbf{P}}_{\boldsymbol{XY},k} \cdot \overline{\mathbf{P}}_{\boldsymbol{Y},k}^{-1} \\ \hat{\boldsymbol{X}}_{k|k} = \overline{\boldsymbol{x}}_{k|k-1} + \mathbf{K}_k \cdot (\overline{\boldsymbol{Y}}_k - \overline{\boldsymbol{Y}}_{k|k-1}) \\ \mathbf{P}_{\boldsymbol{X},k} = \overline{\mathbf{P}}_{\boldsymbol{X},k} - \mathbf{K}_k \cdot \overline{\mathbf{P}}_{\boldsymbol{Y},k} \cdot \mathbf{K}_k^T \end{cases} \tag{35.12}$$

where $\mathbf{K}_k$ is the filter gain, $\hat{\boldsymbol{X}}_{k|k}$ and $\mathbf{P}_{\boldsymbol{X},k}$ are the estimate state and covariance.

35.4 Simulation

In this section, the proposed modeling and filter are evaluated via numerical simulations. The reference orbit come from a spacecraft with orbit elements as

$$(a, e, i, \Omega, \omega, M) = (22031.384\,\mathrm{km},\ 0.892,\ 29.521^\circ,\ 19.287^\circ,\ 357.819^\circ,\ 122.883^\circ)$$

The total orbit time is 15,000 s and seven maneuvers are involved as seen in Table 35.1. The motion is measured by three radar with range R, rate of range v,

Table 35.1 Maneuvers in the simulated orbit

No.	Maneuver time (s)	Maneuver mode (along the orbit)
1	3,001–3,006	Velocity increment (T-orientation): 2 m/s each second
2	3,501–3,502	Velocity increment (T-orientation): 10 m/s each second
3	4,001–4,004	Velocity increment (T-orientation): 2 m/s each second
4	4,501–4,503	Velocity increment (T-orientation): 4 m/s each second
5	6,001–8,001	Acceleration (T-orientation): 0.02 m/s^2
6	10,001–10,031	Acceleration (T-orientation): 1 m/s^2
7	13,001–13,051	Acceleration (T-orientation): 0.2 m/s^2

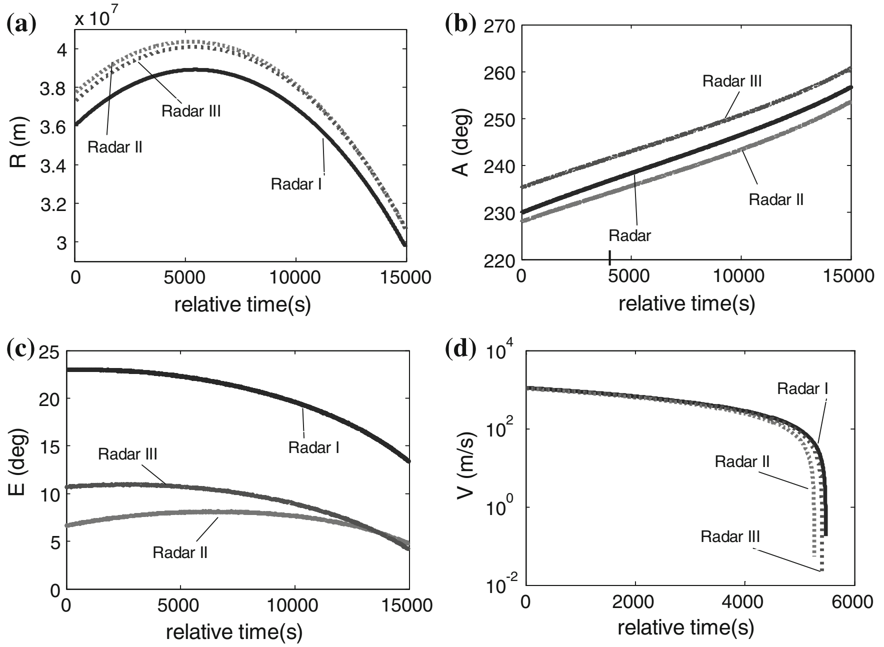


Fig. 35.1 Simulated observations of the orbit. **a** Range. **b** Azimuth. **c** Pitching. **d** Rate of range

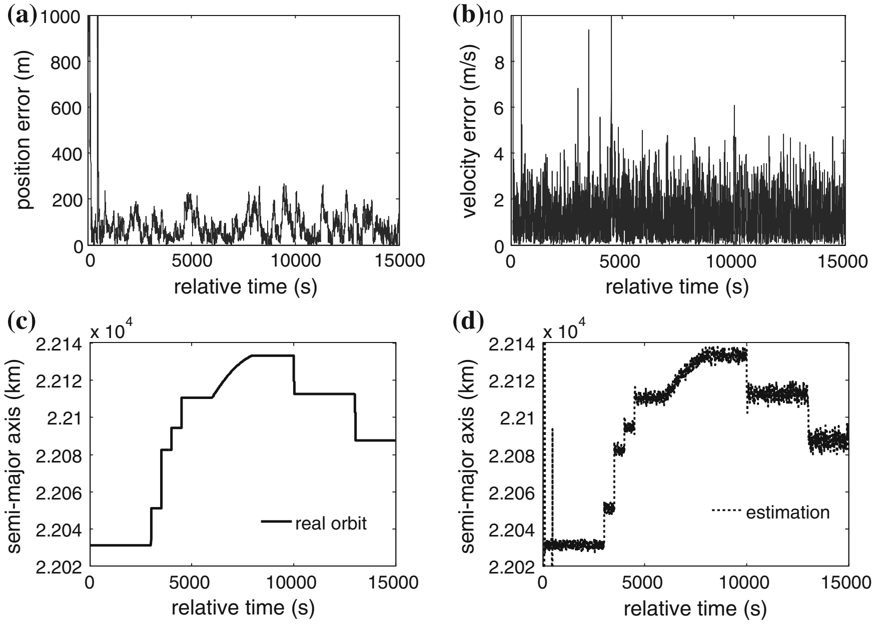


Fig. 35.2 Orbit estimation by the "current" statistical modeling (CS). **a** Position error. **b** Velocity error. **c** Real a. **d** Estimated a

pitching E, and azimuth A at a sample frequency of 1 Hz. The radar are disposed $(-30^\circ,\ 220^\circ,\ 27\,\mathrm{m})$,$(-35^\circ,\ 240^\circ,\ 1,273\,\mathrm{m})$和$(-40^\circ,\ 230^\circ,\ 129\,\mathrm{m})$ at respectively. The observations are shown in Fig. 35.1.

Figures 35.2, 35.3 and 35.4 are the position and velocity filter error as well as estimated semi-major axis by different modeling methods. The CS model have obtained continuous orbit estimation and reflected large orbit maneuvers by the estimated semi-major axis which can be seen in Fig. 35.2. But the accuracy is not satisfactory. The curve of dynamic model in Fig. 35.3 has a very slow convergence

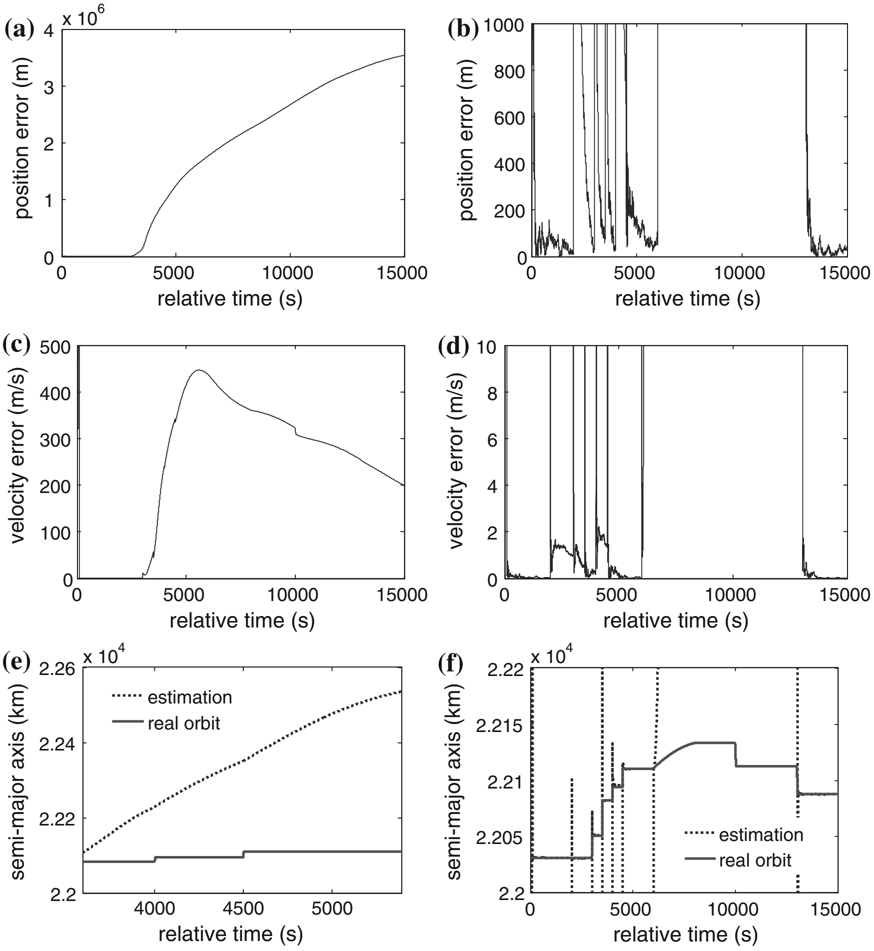


Fig. 35.3 Orbit estimation by the dynamic modeling. **a** Position error (unknown maneuver). **b** Position error (known maneuver). **c** Velocity error (unknown maneuver). **d** Velocity error (known maneuver). **e** Estimated a (unknown maneuver). **f** Estimated a (known maneuver)

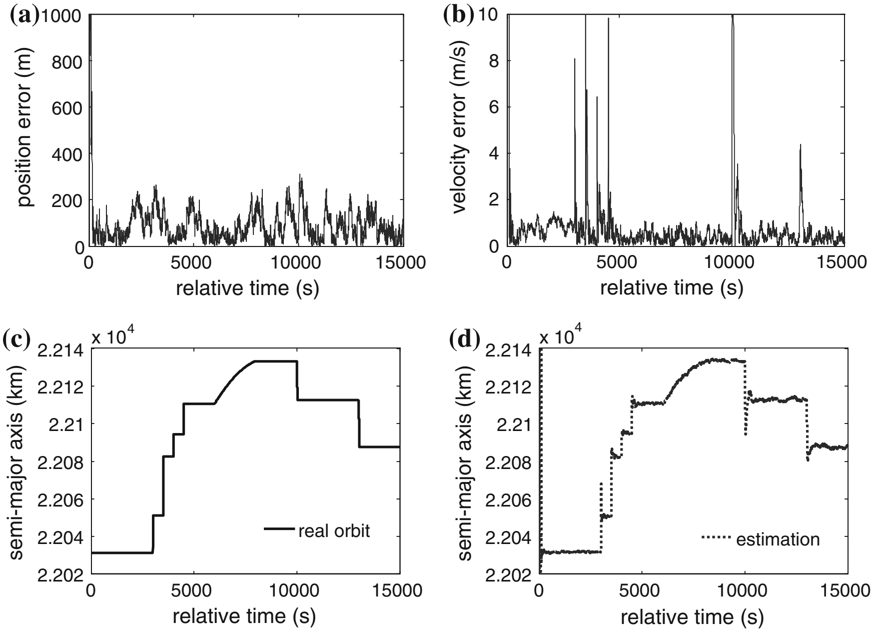


Fig. 35.4 Orbit estimation by the compensative modeling of "current" acceleration. **a** Position error. **b** Velocity error. **c** Real *a*. **d** Estimated *a*

Table 35.2 Detected result of the maneuvering time

No.	1	2	3	4	5	6	7
Maneuver time (s)	3,001	3,500	4,000	4,501	10,004	13,006	13,001

and became divergent when large maneuvers occur. Even taking a filter restart operation at every maneuvering time, the filter could not give a satisfied convergence for frequent pulsar maneuver or small-continuous maneuver. Figure 35.4 shows the tracking result of the compensative model, which is continuous and more accurate than the CS model.

Furthermore, all maneuvers are reflected by the estimated semi-major axis as shown in Table 35.2 and Fig. 35.5. This is obtained by maneuver detection according to the continuity and stability of the estimated orbit which will be discussed in another paper. As we can see the detection delay is no more than five sample period except the small-continuous maneuver during 6001–8001s.

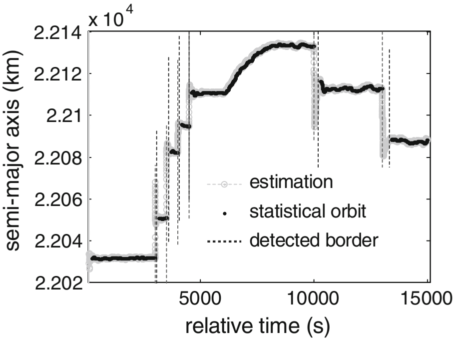


Fig. 35.5 Sketch of the maneuver detection by estimated semi-major axis of the compensative dynamic model

35.5 Conclusion

High accurate tracking for spacecraft depends on the accurate modeling, especial the adaptability of model for the potential maneuver. By combining the tracking accuracy of the dynamic model and the maneuver adaptability of the CS model, a maneuvering acceleration compensation modeling method is proposed. Then a real-time tracking filter is developed utilized the new model and the UKF algorithm. Numerical experiments show that the method can be adapted to the mixed maneuvering tracking mode and accesses continuous and accreted orbit estimation. To further improve the tracking performance for spacecraft with complicated maneuver, the future research will be focused on the high-order or coupled model according to the idea of compensative modeling. In addition, maneuver detection based on compensation model will be studied too, especially for the small-continuous maneuver case. This may provide a support for other space tasks, e.g. the target recognition task.

References

1. Xu T, Zhu XP, Yang J (2012) Research on maneuvering target tracking algorithm with bearing-only measurements. In: The 3rd international conference on mechanic automation control engineering, pp 2432–2436
2. Li XR, Jilkov VP (2001) A survey of maneuvering target tracking—part II: ballistic target models. In: Proceedings of 2001 SPIE conference on signal and data processing of small targets, San Diego, vol 4473, pp 1–23
3. Li XR, Jilkov VP (2003) A survey of maneuvering target tracking part I: dynamic models. IEEE Trans Aerosp Electron Syst 39(4):1333–1364
4. Li XR, Jilkov VP (2002) A survey of maneuvering target tracking—part IV: decision-based methods. In: Proceedings of SPIE conference on signal data processinf small targets, Orlando, pp 1–24
5. Liu Y, Yu AX, Zhu JB et al (2013) Survey of the filters for ballistic target real-time tracking. J Astronaut 34(11):1–9
6. Zhou HR (1991) Maneuvering target tracking. National Defense Industry Press, Beijing

7. Pang WH, Wen YL (1954) Liao Y (2008) Orbit determination method with batch processor based on semi-parametric regression model. J Astronaut 29(6):1917–1921
8. Julier SJ, Uhlmann JK (1997) A new extension of the Kalman filter to nonlinear systems. In: Proceedings of aerosense: 11th international symposium aerospace/defense sensing, simulation and controls, Orlando, pp 54–65

Part III

Space Object Exploration and Identification

Chapter 36
A Robotic Identification Method for Space Object Tracking

Zhanwei Xu, Lianda Wu and Xin Wang

Abstract A practical method is proposed for the robotic identification and capture of space object in tracking mode. Method chooses proper tracking speed of the telescope and exposure time of the sensor to achieve the image which generates significant difference between the static and moving objects. Algorithm, which recognizes moving object with single frame and correlate moving path on multiple frames, drastically reduces the computational resources compared to the traditional image-differencing method. Path correlation methods are given separately for both low and high altitude object. Practical experiments show that the method is effective and convenient, realizes the fully robotic tracking for electro-optical facilities well as the key technology for the network observation system.

Keywords Space object · Path correlation · Automatic identification · Moving-target identification

36.1 Introduction

Electro-optical telescopes play very important role in the space surveillance of space object (SO) by collecting angular measurements for orbit determination [1]. The method of optical positioning uses the telescope with CCD camera to achieve the image of target object and surrounding stars as well as the corresponding time and pointing of the telescope and the measurements of the target is obtained according the relative positions of the object to the stars [2, 3] or the absolute position according to the encoding disk [4]. For telescope with large field of view

---

Z. Xu · L. Wu · X. Wang (✉)
Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008, China
e-mail: wangxin@pmo.ac.cn

Z. Xu · L. Wu · X. Wang
Key Laboratory of Space Object and Debris Observation, Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_36

(FOV), auto recognition usually requires multiple frames of the same object in the same FOV [5].While for the smaller FOV, research focus mainly on the tracking process [6, 7], recognition and identification of the target is still done manually by the human observer, in other words, the initial information which is required by the following tracking depends on human. This method is limited with the ability of human eye and cannot take the advantage of current technology in the detection of faint object and is also affected by the operators' level of proficiency and fatigue.

In this paper, differ from the previous work, we focus on the recognition and identification of the target, a practical method is proposed for the robotic identification and capture of SO in tracking mode for telescope with small FOV. The method realizes the moving object recognition with a single frame by choosing suitable tracking speed and exposure time. For the SOs with different orbiting characteristic, different path correlation algorithm is used for low earth orbit (LEO) object and high-altitude earth orbit (HEO) object separately. Robotic identification for all kinds of SOs is implemented. With this method, the tracking of SOs no longer relies on manual operation and is fully robotic. The observation efficiency is improved, especially for large network.

36.2 Acquisition of Characteristic Image

Short exposure, which brings about the similar morphology for both stars and moving objects, is widely used in the optical observation of SO to improve the accuracy of positioning. The similar morphology is the key reason for manual recognition. For large FOV, multiple frames of image are captured with fixed pointing and the moving object can be detected by compare these frames with the image-differencing method. This way consumes much computing power and does not meet the requirement of tracking which must be real-time. To solve this problem, firstly an image with distinguishable morphology between static and moving object should be achieved, then the recognition can be done with a single frame while path correlation is the only link between multi-frames, and the computing power decreases rapidly.

Though different objects have different orbiting speed, the stars, which have the same sidereal speed, are the main part in the background. The way of tracking the moving object, not the starts, is chosen, CCD images are captured during the telescope moving. To enchase degree of distinction of stars and moving SO in the CCD image, the exposure time is calculated according to the speed of SO to keep the aspect ratio of SO in the image to one, which means the circular shape, while the shape of star is elongated. Exposure time t is:

$$t = S/(V_o - V_T) \tag{36.1}$$

where V_o is the apparent velocity of SO, V_T is the velocity of telescope. S is the ideal length of SO, 5–6 pixel is the typical value.

36.3 Choice of Tracking Speed

In the practice of SO tracking, because of predicting error, to improve the success rate of capture, the telescope is moving along the direction of SO ahead of predicted time, a smaller speed is chosen $V_T < V_o$, which deals both situations of delay and ahead of schedule. During the tracking, the system processes every image to detecting moving object; close-loop tracking is triggered after the identification of moving target. As the low relative speed between the telescope and moving object, SO will transit the whole FOV slowly. In this mode, the length of star R_s is:

$$R_s = R_o(V_s - V_T)/(V_o - V_T) \tag{36.2}$$

where V_s is the apparent velocity of stars; R_o is the length of SO, with the exposure time as Eq. (36.1). Normalized unit is often used and $R_o = 1$. Compared with the method of elongating SO, an obvious advantage of this method is the increase of the detection ability, magnitude is increased by:

$$\Delta M = 2.5\,\log(V_o/|V_o - V_T|)(\mathrm{MAG}) \tag{36.3}$$

With this method, star and SO have significant difference in the image. As shown in Fig. 36.1, point-like object is SO and streak objects are stars.

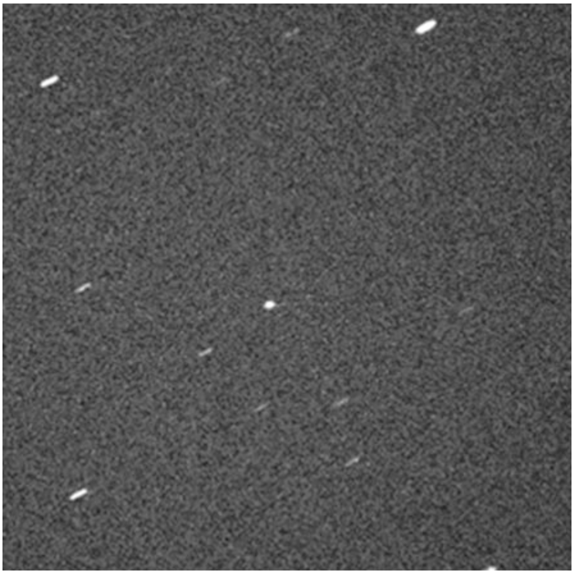


Fig. 36.1 Space object as *point* and stars as *streaks*

36.4 Identification of Moving Object

All the objects, both starts and moving object, can be obtained with full frame segment algorithm [8] from the acquired image. To distinguish the star and SO, 4 parameters, coordinates in X and Y axis, aspect ratio and area objects are extracted. According the previous section, the most significant difference between moving object and stars is the aspect ratio, $R < R_m$ is used to find the candidate moving target. The lower bound of area is also employed to reduce the false alarm. Multiple candidates will be selected according to these two thresholds. Follow-up frames are used to identify the unique target.

For every frame, candidates can be extracted with same algorithm and filtering is carried out with the prior information of the orbit of the target. LEO and HEO are considered separately because of different apparent motion.

Let the XY-coordinate of candidates are $O_{ij} \in R^2$, where i is the sequence number of the images, j is the number of candidates, $j = 1, 2, \ldots, n$, n_i is the total number of candidates in frame i. For tracking problem, the prediction coordinate is known and let C_i is the predicted position of the target on frame i. Three frames are often used in the identification, for LEO objects, there is no need for the continuity. Firstly, the candidates on first two frames are compared one by one. The velocity is chosen as the universal criterion, because the largest error occurs in the along-track direction which is expressed as a shift of the streak on the image. Let three frames are (i, $i + 1$, $i + 2$), condition is:

$$\left|O_{aj} - O_{bk}\right| - |C_a - C_b| < G_v \tag{36.4}$$

where G_v is the threshold. Pair (j, k), which meets the condition, is constructed as path l_{jk}^{ab}. The value of (a, b) takes $(i+1, i)$, $(i+2, i)$, and $(i+2, i+1)$ sequentially until the construction of the path. After this process, if there's no path constructed, the process will repeated with $i = i + 2$.

After path l_{jk}^{ab} constructed, follow-up frame $(c = b + 1)$ will be compared with the path to check whether it is the predicted target or not with the condition:

$$\left|O_{cl} - O_{bj}\right| - |C_c - C_b| < G_v \tag{36.5}$$

And the objects on three frames must be the same object is the last condition, as the SO is moving linearly in a very short span [9], we have:

$$\arccos \frac{(O_{cl} - O_{bj}) \cdot (O_{bj} - O_{ak})}{\left|O_{cl} - O_{bj}\right| \left|O_{bj} - O_{ak}\right|} < G_l \tag{36.6}$$

where G_l is the threshold of the degree of linearity. The object which meets conditions of Eqs. (36.5) and (36.6) are identified as the target. If no target is identified, let $c = c + 1$ and repeat the comparing with newly acquired image. The nearest object to the center of FOV is identified as the target, if multiple objects meet the conditions. The whole flowchart is shown as Fig. 36.2.

The moving characteristic of HEO objects is very similar with the stars and they are often confused by the people. Considering that the moving speed is very slow and the prediction accuracy is very high, a strict criterion of same object on continues three frames is used. As the HEO object moves very slowly, so Eq. (36.4) is the only condition used for the path construction. Let two frames $(i, i+1)$ are obtained, path l_{jk} is constructed if the following condition is met.

$$\left|O_{i+1,j} - O_{i,k}\right| - \left|C_{i+1} - C_i\right| < G_v \tag{36.7}$$

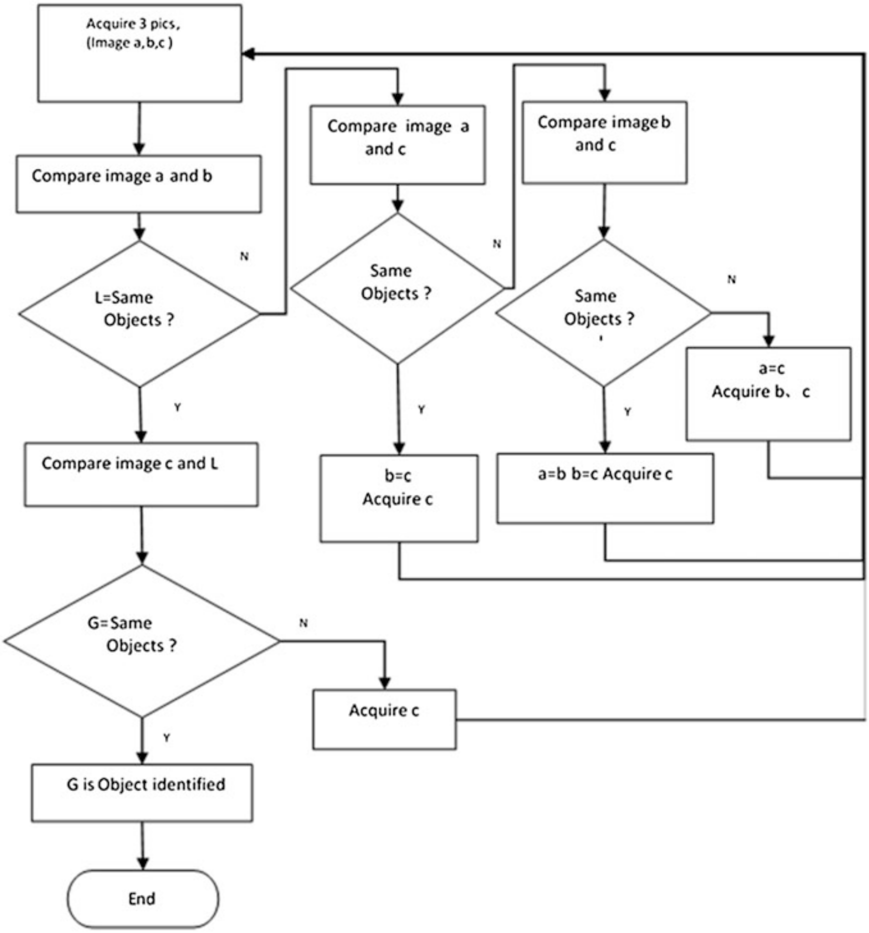


Fig. 36.2 Flow chart of identification of low-altitude object

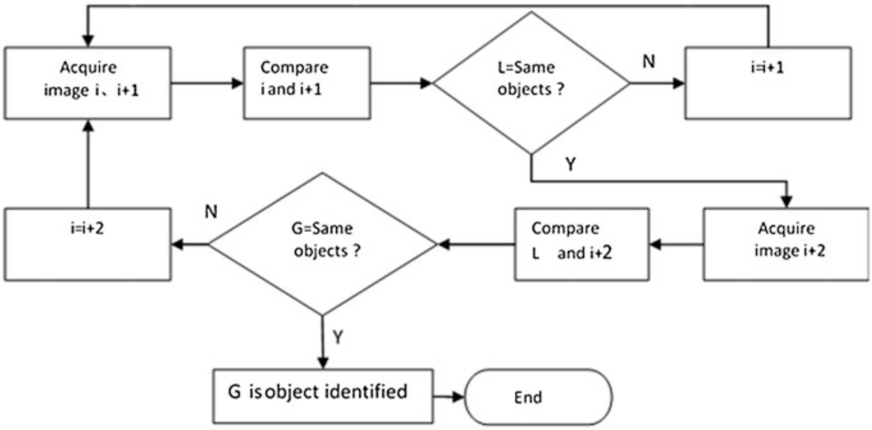


Fig. 36.3 Flow chart of identification algorithm of high-altitude object

The process is repeating with $i = i + 1$ until the path is constructed. After the construction, follow-up frame is compared with path l_{jk} by:

$$\left|O_{i+2,l} - O_{i+1,k}\right| - \left|C_{i+2} - C_{i+1}\right| < G_v \tag{36.8}$$

The tracking speed is equal to the apparent speed of the object, so $V_T = V_o$ which means the SO is fixed to the telescope and the XY-coordinate is kept as constants, we have

$$\left|O_{i+2,l} - O_{i+1,k}\right| < G_p \tag{36.9}$$

where G_p is the threshold of target-missing quantity. The object which meets the condition is identified as the target. Let $i = i + 2$ and repeat the path construction algorithm, while no target is identified, as shown in Fig. 36.3.

36.5 Experiments

To check the effectiveness of the method proposed in this paper, experiments with a 40 cm-aperture telescope of the Space Object Optical Observation Network of Chinese Academy of Sciences are carried out for both LEO and HEO. The entire image is 512 × 512 pixel with 0.7° × 0.7° FOV.

36.5.1 Experiment of LEO Object

In the experiment, telescope is waiting according prediction and start moving before the predict time with 80 % velocity of the target $V_T = 0.8\ V_o$. Full image extraction is taken to every image. Because the telescope moves very fast, the sidereal speed is ignored. According to Eq. (36.1), the streak length of the star is $4R_o$ as the width of the star is not exactly as the same as the SO, $R_m = 1.5$ is chosen which is a little more than the theoretical value and the threshold of area is 10 pixels. Candidates are listed in Table 36.1. Tagged object in Fig. 36.4 are the candidates. As there's only one candidate object on each image, the path is unique and confirmed with thresholds $G_v = 5$ pixel and $G_l = 5°$. The result of correlation is shown in Fig. 36.5. The tracking is successfully according to the post processing of the collected measurements.

36.5.2 Experiment of HEO Object

Experiment of HEO is carried out with same telescope, telescope starts at the predicted time with same velocity and direction of the target. There are multiple candidates on each image and the total number of candidates is also different between images (see Fig. 36.6 and Table 36.2).

Take $G_p = 5$ pixel, we obtained the identified object as in Table 36.3 and the tracking is successfully.

Table 36.1 Result of candidate low-altitude objects

Frame #	X (axis)	Y (axis)	Aspect ratio	Area (pixel)
1	10.778	426.650	1.333	60
2	20.109	419.026	1.203	69
3	28.404	412.276	1.199	66

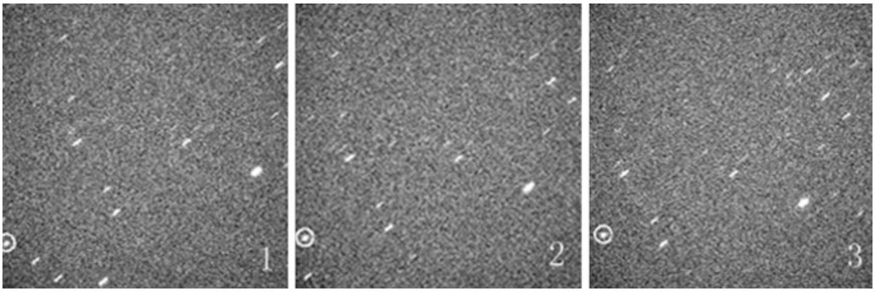


Fig. 36.4 Candidate low-altitude objects in image frames

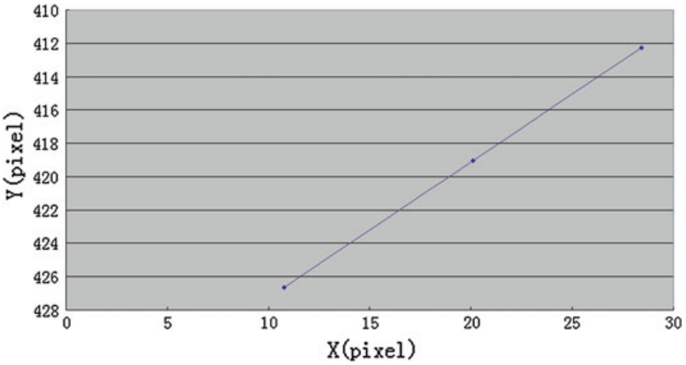


Fig. 36.5 Associate path of the LEO target

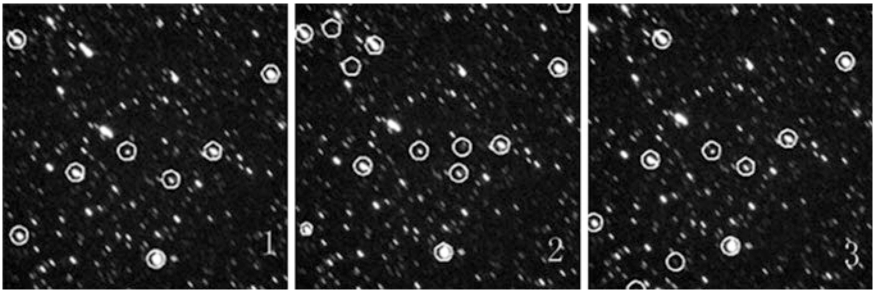


Fig. 36.6 Candidate high-altitude objects in image frames

Table 36.2 Result of candidate high-altitude objects

Frame #	Object #	*X* (pixel)	*Y* (pixel)	Aspect ratio	Area
1	1	26.886	65.159	1.4720	429
	2	274.871	454.599	1.2780	585
	3	31.485	412.512	1.3030	298
	4	301.560	313.458	1.2200	223
	5	131.752	301.020	1.4980	470
	6	375.980	263.121	1.3520	459
	7	222.898	263.242	1.0860	117
	8	478.446	125.878	1.4460	486
2	1	18.083	54.317	1.4910	412
	2	479.590	1.843	1.4240	11
	3	266.024	443.753	1.2880	563

(continued)

Table 36.2 (continued)

Frame #	Object #	*X* (pixel)	*Y* (pixel)	Aspect ratio	Area
	4	22.692	401.643	1.2990	302
	5	292.825	302.686	1.2400	224
	6	122.974	290.160	1.4990	477
	7	222.877	263.200	1.1540	127
	8	367.199	252.429	1.3370	449
	9	299.092	256.342	1.4930	52
	10	469.678	115.102	1.4530	493
	11	102.673	113.631	1.3400	25
	12	143.640	75.718	1.4970	423
	13	67.687	45.361	1.2440	11
3	1	134.733	64.540	1.4900	435
	2	88.253	510.676	1.4060	24
	3	158.326	460.640	1.4960	13
	4	257.083	432.604	1.2680	565
	5	13.605	390.535	1.3220	308
	6	283.933	291.673	1.2570	217
	7	114.066	279.035	1.4850	468
	8	223.170	263.149	1.0810	110
	9	358.371	241.334	1.3810	436
	10	460.796	104.053	1.4310	494

Table 36.3 Result of identified objects

Frame #	Object #	*X* (pixel)	*Y* (pixel)	Aspect ratio	Area
1	7	222.898	263.242	1.086	117
2	7	222.877	263.200	1.154	127
3	8	223.170	263.149	1.081	110

36.6 Conclusion

The method proposed in this paper realized the automatic recognition and capture of space target with three steps. Firstly, according to the prediction of space target, appropriate parameters are employed to obtain the image with distinction between background stars and SOs; Secondly after full image extraction, candidate targets are selected according to parameters of the aspect ratio, area parameters and etc. Finally, considering the along track error and path slope, path correlation is carried out to identify the tracking target. The method, which is simple, effective and feasible, has been applied to the daily operation of some devices equipped in the Space Object Optical Observation Network of Chinese Academy of Sciences. Ordinary PC machine supplies enough computational requirements for CCD image of 512 by 512 and 1,024 by 1,024 and achieved very good results.

This method realizes the unmanned observation with electro-optical telescope, and also provides technical basis for the network observation which needs fast response. And there's very good application prospect in the field of space surveillance.

References

1. Wu LD (2010) Orbit and detection of artificial earth satellite and space debris. Chin Science and Technology Press, Beijing
2. Shen Z, Luo C, Hua WH (2010) Methods accurately locating optical positions of space object based on ccd stellar images. Astron Res Technol 72(3):231–236
3. Lu CL, Zhang XX, Gu GD (2005) Real time astronomical calibration method for space object. China patents CN1710377
4. Chen JQ (2003) Principles of astronomical telescope design. China Science and Technology Press, Beijing
5. Han JT, Zhang Y, Chen ZP (2005) Real time detection algorithm of weak targets in astronomical images. Opto-electron Eng 32(12):1–4
6. Zhang XX, Wu LD, Xiong JN (2004) Circular orbit tracking method of space objects. Chin Astron Astrophy 28(1):94–104
7. Wang X, Xu ZW (2012) A new approach to the electro-optical tracking of a space object. Chin Astron Astrophy 36(4):426–434
8. Lutz RK (1980) An algorithm for the real time analysis of digitized images. Comput J 23 (3):262–269
9. Liu L (1998) Method of celestial mechanics. Nanjing University Press, Nanjing

Chapter 37
Calculation of Optical Scale of Space Debris

Jianfeng Wang, Xiaomeng Lu and Yongna Mao

Abstract This paper proposes the concept of the optical scale of space debris and presents a preliminary study of the acquisition and calculation method of the optical scale of space debris. Based on the large data obtained by ground-based observing facilities, the multi-object, multi-circle and multi-data-point luminosity information of space debris can be measured by utilizing high-precision technology of flux extraction and calibration to eliminate instrumental and atmospheric effects. Furthermore, reliable optical scale of space debris can be obtained by correcting the effect on luminosity caused by the position change of space debris and eliminating the changes of projected area on the observer's line of sight due to the rotation of space debris through statistic analysis of large sample data. The experimental results show that the statistical error of optical scale of space debris between calculated and real values is 50 % (RMS). With the advantages of easy acquisition, broad applications and high precision, optical scale can be regarded as one of physical characteristic to characterize the size of space debris well.

Keywords Space debris · Optical size scale · Slant range · Phase angle · Flux

37.1 Introduction

The research of physical characteristics of space debris is an international new field in the recent years. Since each piece of space debris has its own geometry shape, material composition, surface paint coat, orbit, attitude, etc., it has particular

---

J. Wang (✉) · X. Lu · Y. Mao
Key Laboratory of Optical Astronomy, National Astronomical Observatories,
Chinese Academy of Sciences, Beijing 100012, China
e-mail: wjf@nao.cas.cn

J. Wang
University of Chinese Academy of Sciences, Beijing 100049, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_37

irradiation and visible relationships with the sun and the ground base observing sites as well as particular energy spectral dynamical variation rules. By obtaining and analyzing of these rules, we can master the physical characteristics of space debris and provide important information for mastering and understanding the origin and application of space debris [1, 2].

Size is an important physical characteristic of space debris, now the Radar Cross Section (RCS) is usually used to characterize the size of the space debris. Contrast with optical band, the radar wavelength is longer, and the RCS related to many factors, such as instrument's frequency and form of polarization, the structure and attitude of the space debris. The RCS is very sensitive to edge where it always gets a high value, so it cannot represent the real size of space debris well. In addition, the radar's signal strength, which is proportional to the fourth power of distance, will be depleted from ground clutter and atmosphere, and also be confined to its power and wavelength. Thus it is difficult to use radar to detect high apogee space debris. Compared with radar, optical observation has characteristics of shorter-wavelength, longer detection distance and passively receiving. This paper is based on these characteristics to study the calculation of optical size scale of space debris, to precisely obtain the characteristic parameters of characterization of size scales of space debris.

37.2 Principle

Space debris's luminosity is mainly due to the reflection of sunlight, The density of solar radiation near the Earth has been accurately measured with value 1.367×10^3 w/m^2, so the brightness relates to the cross-sectional area and surface materials albedo, the function can be expressed as:

$$\mathrm{Flux_C} = \rho \times \mathrm{S} \times v \tag{37.1}$$

When material's albedo can be obtained by ground testing, so the optical scale is calculated as follows:

$$\mathrm{S} = \mathrm{Flux_C}/\rho/v \tag{37.2}$$

37.2.1 Acquisition of High-Precision Luminosity Information

When observing space debris with optical telescopes, we can obtain not only the position information, which is used to determine the orbit of space debris, but also the flux information (luminosity). The accessing of the flux information of debris greatly enrich the data types of observing targets, extend the location data, which can reveal the orbit information, to the data which characterizes the physical properties itself and provide a solid data foundation for carrying out the physical characteristics research of space debris. Similarly, studies of optical scale which is one of physical properties of space debris should be based on its flux information, and the accuracy of flux information directly affect the accuracy of the acquired optical scale [3–5]. So, how to get the high-precision flux information of the space debris is the first problem that we need to solve.

In addition to high-precision data requirements, a large sample of observations is also necessary basis of our research. There are many kinds of space debris, such as failure of the satellite, rocket debris, satellite debris, etc., and the difference between individuals is also large. Therefore, factors such as the regularity of space debris shape, stability of orbit control, and the different operating attitude, etc., directly result the ever-changing characteristics of its luminosity. We have to pick the right space debris and perform intensive observations to provide a large enough sample data to support the follow-up study.

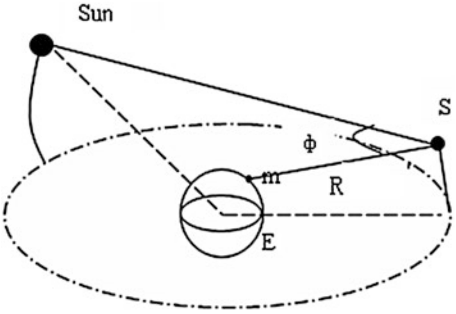
When the radiation of space debris outside atmospheric is acquired by the telescope, it has been affected by the instrument (telescope optics, filters, detectors, etc.) and the atmosphere [6, 7], so that observations of the same object at the same time by using different instruments, in different locations, the radiation flux obtained is not the same. So we need to correct these effects to reduce the real flux information of space debris itself.

37.2.2 Correction of Effect of Position Change

When the space debris running in space, it does not emit light itself, but illuminated by reflecting sunlight. Since the debris is always in motion, and the spatial geometry position of the sun, which is the irradiation source, is also changing constantly. Relative to the ground observer, the moving space debris and changing lighting conditions bring uncertainty factors to observations of space debris, so that the photometric properties derived from observations of space debris in different times and different locations are different. Relative to the observer, the movement of space debris consists of two parts: the attitude of rotation (rotation) and spatial position movement (revolution around the Earth). Rotation is running attitude of the space debris mentioned above, while revolution is the changes of the slant range and phase angle mentioned above. Through correcting of the luminosity variations due to rotation and revolution, we can eliminate these effects and eventually get the information for characterizing the shape scales. Thus, how to establish the corrected relationship of the observed luminosity and the slant range, the phase angle to eliminate the effects of the revolution is the key technology of this work.

As shown in Fig. 37.1, m is ground based observing station, R is the distance between observing station and space debris, i.e. slant range, Φ is the angle of station —debris—the sun, i.e., the phase angle (PA).

Fig. 37.1 Schematic of space debris observation



Over time, the sun's position and the spatial geometry of space debris has also changed due to the movement of spatial location, slant range and phase angle are also changing, which leads to the energy of sunlight reflected by space debris change, a direct manifestation is the change of luminosity of space debris. For the optical telescope, the observations of space debris can only obtain the angle and photometric information, neither distance information, nor phase angle derived by transforming the stellar angle commonly used in astronomy, so when getting the slant range and phase angle of space debris, we mainly use orbit prediction techniques of space debris. By this calculation, we will get the slant range and phase angle of space debris, which would eliminate the effects of the two factors by the model algorithm in the calculation process of extracting the optical size scale, which can be used to represent the scale size, from the photometric information of space debris, so that the optical size scale can be used to describe the size of space debris more accurately.

The correction function of the slant range and phase angle is as follows:

$$Flux_C = f(R,\phi) = Flux \times \frac{R^2}{(R-R_C)^2} \times \frac{180^\circ - \phi_C}{180^\circ - \phi} \tag{37.3}$$

In this formula, the Flux is corresponding to flux value of the slant range of R and the phase angle Φ; Rc is normalized distance; Φc is normalized phase angle; $Flux_C$ is the corresponding flux.

By modifying the function (37.3) we can make a correction of irradiation angle, which are due to the location and movement, and the distance relationship of space debris, and eliminate these two effects to obtain the corrected flux information of space debris.

37.2.3 Optical Scale Computing

Optical observations are always carried out at night when the brightness of the space debris mainly comes from the sun's radiation and the flux is directly related to the solar radiation. Astronomical observation shows that solar radiation density near the earth is approximately a constant. If the scale of space debris is known, the total flux of space debris illuminated by solar radiation can be obtained.

Space debris scatters solar radiation at different rate depending on its surface material [8]. Space debris scattered flux over the incoming flux is called albedo. We can roughly determine the upper limit and lower limit of albedo through simulation studies on the surface material of space debris.

The relationship between optical scale and flux of space debris can be obtained with the corrected flux and the density of the solar radiation.

$$OCS = \frac{Flux_C}{\rho_{\text{sun}} \times \upsilon} \tag{37.4}$$

After slope distance and phase angle corrected, optical scale still contains shape and attitude information of space debris. We can only observe the projection of the true scale at a certain viewing angle due to attitude rotation of space debris and rotation of the earth which changes with time. To obtain the scale closer to the true size of space debris, we need to do statistical processing to remove the influence of attitude change with large amounts of data.

37.3 Results and Analysis

In order to verify the calculation method, a group of space debris with known shape and size have been chosen and observed intensively. The observations are shown in Table 37.1.

Obtaining information on the brightness of space debris, slant range and phase angle from the phase correction and calculation while using a large sample of data for statistical processing, when the rectangular parallelepiped by the albedo mean value 0.1, 0.02 cylinder is calculated to obtain the optical scale value. Meanwhile according to its dimensions information, calculate the scope of its cross-sectional area, and take the mid-value as a typical value, the result is shown in Table 37.2 and Fig. 37.2.

Preliminary study on the statistical analysis of a large sample data showed that optical scale coincide well with the real dimensions of space debris with a statistical error of 40.66 % (RMS), less than 50 % (RMS). Since we cannot obtain all the points of view of the objects, sometimes the deviation between the optical scale and the real size would be very large, e.g. a point of 140 % in Fig. 37.2.

Table 37.1 Situation on photometric observations of space debris

No.	Shape	Size (m)	Observation times
1	Cuboid	2.16 × 2.46 × 3.15	25
2	Cuboid	2.10 × 2.27 × 3.40	20
3	Cuboid	2.80 × 4.90 × 3.80	31
4	Cuboid	3.80 × 2.80 × 4.30	24
5	Cuboid	3.40 × 3.50 × 5.80	4
6	Cuboid	2.20 × 1.72 × 2.00	7
7	Cuboid	3.10 × 2.50 × 6.10	22
8	Cylinder	Diameter 2.16, Length 6.59	22
9	Cylinder	Diameter 2.16, Length 6.59	25
10	Cylinder	Diameter 2.10, Length 4.10	5
11	Cuboid	3.60 × 2.70 × 4.30	8
12	Cuboid	7.30 × 3.62 × 3.62	10
13	Cuboid	3.40 × 3.50 × 5.80	25
14	Cuboid	2.20 × 1.72 × 2.00	8
Total			236

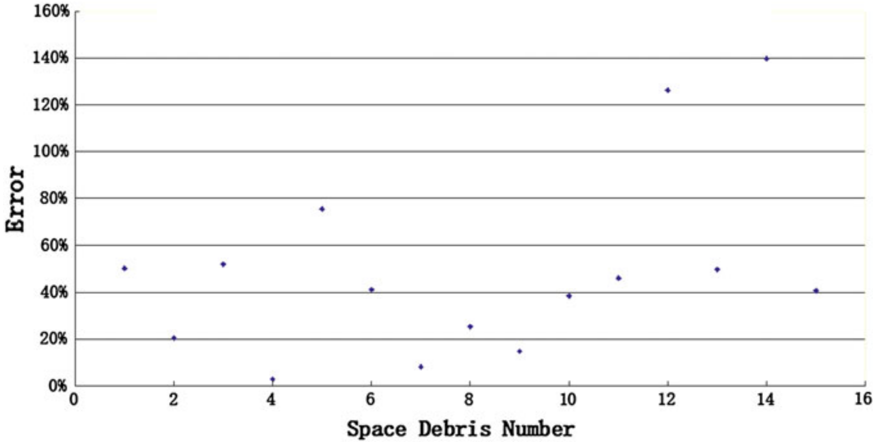


Fig. 37.2 Relative error of optical scale to real size

Table 37.2 Comparison between optical scales and body scales of space debris

No.	Size (m)	Observation times	Range of cross section (m^2)	Albedo	Mid-value (m^2)	Optical scale (m^2)	Relative error (%)
1	2.16 × 2.46 × 3.15	25	5.31–10.31	0.10	7.81	11.74	50.32
2	2.10 × 2.27 × 3.40	20	4.77–10.51	0.10	7.64	6.08	20.42
3	2.80 × 4.90 × 3.80	31	10.64–23.13	0.10	16.88	8.11	51.95
4	3.80 × 2.80 × 4.30	24	10.64–20.30	0.10	15.47	15.89	2.71
5	3.40 × 3.50 × 5.80	4	11.90–28.30	0.10	20.10	4.94	75.42
6	2.20 × 1.72 × 2.00	7	3.44–5.80	0.10	4.62	6.52	41.13
7	3.10 × 2.50 × 6.10	22	7.75–24.29	0.10	16.02	17.31	8.05
8	Diameter 2.16, Length 6.59	22	3.66–14.23	0.02	8.95	6.69	25.25
9	Diameter 2.16, Length 6.59	25	3.66–14.23	0.02	8.95	7.63	14.75
10	Diameter 2.10, Length 4.10	5	3.46–8.61	0.02	6.04	8.36	38.41
11	3.60 × 2.70 × 4.30	8	9.72–19.35	0.10	14.54	7.84	46.08
12	7.30 × 3.62 × 3.62	10	13.10–37.37	0.10	25.24	57.13	126.35
13	3.40 × 3.50 × 5.80	25	11.90–28.30	0.10	20.10	10.10	49.75
14	2.20 × 1.72 × 2.00	8	3.44–5.80	0.10	4.62	11.08	139.83
RMS							40.66

37.4 Conclusions

The study of physical characteristics of space debris is a newly emerging field with hot topics. In this new situation, despite the traditional track-based measurements in space debris observation, several observational methods can be used to obtain the shape, size, surface properties of space debris and get the information of their distribution and their origin from the analysis of these parameters.

Based on the observation on brightness of space debris and data analysis, the result shows that as one of the physical characteristic information, Optical scale can characterize space debris size better.

With short wavelengths, passively received and long detection distance of optical observation, Optical scale is a more accurate parameter to characterize the size of space debris. The method can be easily applied to the existing space debris observing equipments.

References

1. Cowardin H, Seitzer P, Abercromby K et al (2010) Characterization of orbital debris photometric properties derived from laboratory-based measurements. In: Proceedings of AMOS
2. Mao YN, Lu XM, Wang JF (2012) Simultaneous multicolor photometry of fast-moving objects using the 1-meter telescope at Xinglong observatory. Science China physics, mechanics & astronomy 55
3. Wellems D, Bowers D (2007) Laboratory imaging of satellites and orbital appearance estimation. In: Proceedings of AMOS
4. Lukesh G, Chandler S, Barnard C (2001) Estimation of satellite laser optical cross section: a comparison of simulations and field results. In: Proceedings of SPIE 4167, 53
5. Lukesh G, Chandler S, Voelz D (2002) Analysis of satellite laser optical cross sections from the active imaging testbed. In: Proceedings of SPIE 4538, 24
6. Tang YJ, Jiang XJ, Wei JY et al (2009) Diffuse reflection optical characteristics of three-axis stabilization GEO satellite. Acta Photonica Sinica 38(11):2863–2866
7. Tang YJ, Jiang XJ, Lu XM (2010) Analysis of photometric characteristics of medium and high apogee satellites based on light-reflection model. Acta Photonica Sinica 39(3):763–767
8. Li SJ, Gao XD, Zhu QX (2004) Analysis for luminosity of a satellite with solar battery panels. Opto-electronic engineering 31(4):1–4, 8

Chapter 38
Infrared Radiation Measurement Based on Simplified Calibration Method

Songtao Chang, Min Li and Yaoyu Zhang

Abstract In order to improve the efficiency of infrared radiometric calibration, a simplified calibration method is put forward. Firstly, a radiometric calibration model comprising integral time is proposed, and the correctness is verified by curves of gray value versus integral time and gray value versus radiance. Then a simplified method for radiometric calibration is deduced. Therefore calibration at all integral times can be obtained by using only three infrared images collected at two typical integral times. Finally, the experiments are designed to illustrate the accuracy of calibration and radiation measurement based on the proposed simplified method. Experimental results illustrate the effectiveness of the proposed simplified calibration method,it improves the efficiency of radiometric calibration greatly, and can be directly applied to target radiation measurement tasks.

Keywords Infrared detector · Radiance calibration · Integration time · Radiometry

38.1 Introduction

Radiometric calibration of infrared imaging system is to establish a quantitative relationship between the input and infrared radiation detector digital output [1–4]. After absolute radiometric calibration, the relationship between output gray value and parameters such as the input radiance and the integration time can be determined. Therefore, radiance of the target can be measured.

S. Chang (✉) · M. Li · Y. Zhang
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
e-mail: stchang2010@sina.com

S. Chang
University of Chinese Academy of Sciences, Beijing 100049, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_38

Radiometric calibration methods of ground based radiometry devices can be divided into two categories: laboratory calibration and field calibration. Among them, the laboratory radiometry calibration is the one that performed under stable laboratory conditions, its advantages are stability and high precision. However, field calibration is the one that carried out in the field with complex environmental conditions. The effect of ambient temperature fluctuations therefore is included in the on radiometric calibration results [5]. In order to maintain the accuracy of radiometry, field radiometric calibrations are usually required. As is known, several different integration times are usually selected achieving high dynamic range radiometry. Hence the calibration results under all integration times shall be obtained. In a consequence, the levels of the calibration references should cover a high temperature range, which naturally makes the calibration more complex.

In this paper, the laboratory calibrations of cooled midwave infrared detectors are performed, a radiometry formula considering the integration time is then deduced. According to the model proposed, a simplified radiometric calibration method, which only needs to collect three images under two integration times and two irradiance levels, is developed. The radiometric calibration model of the infrared imaging system is obtained based on the former three images, which is then verified by comparing the calculated radiometric calibration to with the experimental ones. Finally, a field radiometry experiment is performed using the infrared calibration model obtained by our simplified method, and the radiometry accuracy is calculated for comparison.

38.2 Radiometric Calibration and Measurement Model

38.2.1 Mathematical Model of Radiometric Calibration

An extended area reference source such as a high precision blackbody, which covers the entrance pupil of the imaging system, is essential to radiometric calibration of infrared imaging systems. In order to avoid the influence of atmospheric attenuation and path radiation, the blackbody is usually placed directly in front of the optical system, as is shown in Fig. 38.1.

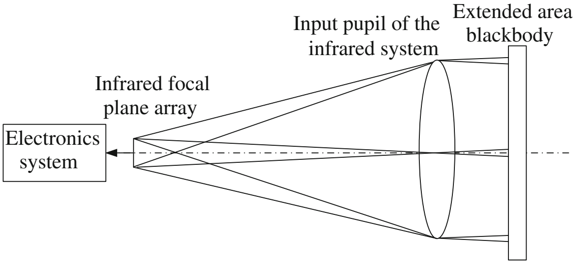


Fig. 38.1 Schematic diagram of radiometric calibration using extended area blackbody

If the temperature of the reference blackbody is denoted as T_b, and the emissivity is ε_{BB}, the radiance is then can be expressed as

$$L_B(T_b) = \varepsilon_B \cdot L(T_b) = \frac{\varepsilon_B}{\pi} \int_{\lambda_1}^{\lambda_2} Planck(\lambda,\ T_b) d\lambda \tag{38.1}$$

where $\lambda_1 \sim \lambda_2$ is the working waveband of the infrared detector, $L(T_b)$ is the radiance of an ideal blackbody at temperature $T_b, Planck(\lambda,\ T_b)$ is the Planck's blackbody radiation formula. Within the linear response range of the infrared imaging system, gray value of a pixel of the focal plane arrays is given by [6–8]

$$h = K_0 \times L_B(T_b) + B_0 \tag{38.2}$$

where K_0 is the responsivity and B_0 denotes the offset which can be represented as

$$B_0 = h_{stray} + h_{detector} \tag{38.3}$$

where h_{stray} is the output gray value caused by stray radiation, let $h_{stray} = K_0 \times L_{stray}$, thus L_{stray} is the incident radiance, consisting of radiation of lens and other components of the system, that reaches the detector element. $h_{detector}$ is the output gray value that caused by factors such as dark current of the detector.

Considering the integration time, the calibration formula is expressed as

$$h = t \times \left(G \times L_B(T_b) + h_{stray}\right) + h_{detector} \tag{38.4}$$

where t is the integration time in units of ms, G and h_{stray} are the responsivity and output gray value result from stray radiation under an integration time of 1 ms, respectively.

The calibrations, under integration times $t_1, t_2, \ldots, t_n$ that required, are usually performed one by one, which leads to heavy workload. Moreover, some infrared imaging systems work at continuously adjustable integration times, such that it is impossible to cover all the integration times by direct radiometric calibration.

38.2.2 Simplified Method for Radiometric Calibration

There are only three unknown parameters, namely G, h_{stray} and $h_{detector}$ in Eq. (38.4). Therefore, if we obtain the above parameters, calibration result at arbitrary integration time can be determined. The blackbody's temperature is set T_1, and the integration times are t_1 and t_2, thus the output gray values are

$$h(t_1,\ T_1) = t_1 \times (G \times L_B(T_1) + h_{stray}) + h_{detector} \tag{38.5}$$

$$h(t_2,\ T_1) = t_2 \times (G \times L_B(T_1) + h_{stray}) + h_{detector} \tag{38.6}$$

Similarly, set the integration time t_2 and the temperature T_2, we obtain

$$h(t_2,\ T_2) = t_2 \times (G \times L_B(T_2) + h_{stray}) + h_{detector} \tag{38.7}$$

Equations (38.4)–(38.6) can be expressed as a linear equation form:

$$\begin{pmatrix} t_1 L_B(T_1) & t_1 & 1 \\ t_2 L_B(T_1) & t_2 & 1 \\ t_2 L_B(T_2) & t_2 & 1 \end{pmatrix} \begin{pmatrix} G \\ h_{stray} \\ h_{detector} \end{pmatrix} = \begin{pmatrix} h(t_1, T_1) \\ h(t_2, T_1) \\ h(t_2, T_2) \end{pmatrix} \tag{38.8}$$

Three parameters namely G, h_{stray} and $h_{detector}$ can be obtained by solving Eqs. (38.8), then the calibration result at arbitrary integration time is determined.

The principles of integration time and blackbody temperature selection are: Under the selected integration time, the output gray value shall be constrained in the range of 30–70 %. The process of selecting the appropriate integration time and blackbody temperature for radiometric calibration is as follows:

1. Set the temperature of the blackbody T_1 then collect the infrared images at integration times namely t_1 and t_2.
2. Set the temperature of the blackbody T_2 then collect the infrared images at integration time namely t_2.
3. Obtain G, h_{stray} and $h_{detector}$ of every pixels of the arrays by solving Eq. (38.8) using the three images above.

The above is a simplified radiometric calibration process, which has the ability to improve the efficiency of the radiometric calibration.

38.2.3 Radiometry Model

Measuring the radiation characteristics of the target, the output gray value of the detector pixel can be expressed as:

$$\begin{aligned} h = {} & \varepsilon_{target} \times \tau_{atm} \times K \times L(T_{target}) \\ & + (1 - \varepsilon_{target}) \times \tau_{atm} \times K \times L(T_{background}) + K \times L_{path} + B \end{aligned} \tag{38.9}$$

where ε_{target} is emissivity of the target, τ_{atm} is the atmospheric transmittance, T_{target} is the surface temperature of the target, K and B are respectively the gain and offset, and L_{path} denotes the path radiation. $(1 - \varepsilon_{target}) \times K \times L(T_{background})$ is the reflected background radiation, and $T_{background}$ is the background temperature. The radiance of the target can be calculated by

$$L(T_{target}) = \frac{h - B - (1 - \varepsilon_{target}) \times \tau_{atm} \times K \times L(T_{background}) - K \times L_{path}}{\varepsilon_{target} \times \tau_{atm} \times K} \tag{38.10}$$

where K and B can be determined by direct calibration or calculated by the simplified calibration method propose in this paper. ε_{target} and $T_{background}$ are directly measured, τ_{atm} and L_{path} can be obtained by using atmospheric transmittance calculation software such as MODTRAN. The radiance of the target can be calculated by Eq. (38.10), and the temperature can be determined by Planck's formula, thus completing the radiation characteristics measurement of the target.

38.3 Verifying the Radiometric Calibration Model

38.3.1 Output Gray Value as a Function of the Integration Time

The experimental set-up used for radiometric calibration is shown in Fig. 38.2.

The experiments were performed with a MWIR camera of FLIR systems having a large-scale MCT FPA (640 × 512 pixels). The camera operates in 3.7–4.8 μm waveband, with a 14-bit digital output. The extended surface blackbody, selected as the reference source, has a 100 × 100 mm size and exhibits high effective emissivity (0.97 in 3.7–4.8 μm waveband). Its temperature accuracy is 0.01 °C over an operating temperature range of 0–125°C.

Figure 38.3 shows the gray value of the center pixel (320, 256) as a function of the integration time.

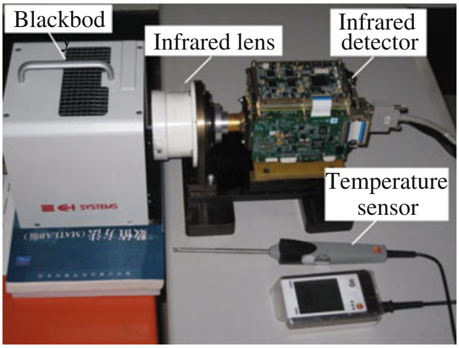
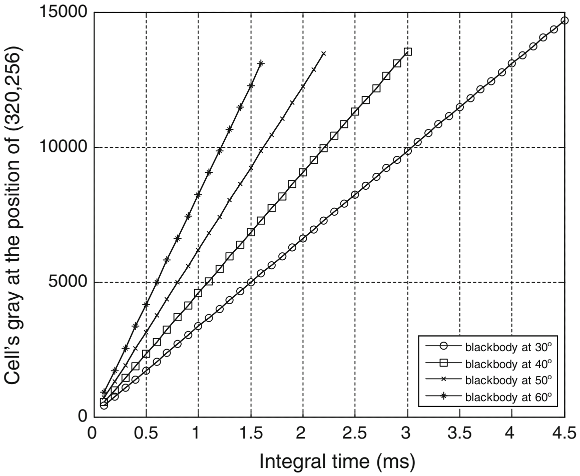


Fig. 38.2 The calibration using an area blackbody

Fig. 38.3 Output gray value as a function of the integration time



As illustrated in Fig. 38.3, when the temperature of the reference blackbody is fixed, the gray value is linear to the integration time. Additionally, the responsivity increases with the rising of blackbody temperature.

38.3.2 Relationship Between the Output Gray Value and the Blackbody Radiance

Figure 38.4 is the curve of output gray value versus the blackbody radiance at a fixed integration time, and the ambient temperature is kept stable.

Fig. 38.4 Output gray value as a function of the integration time

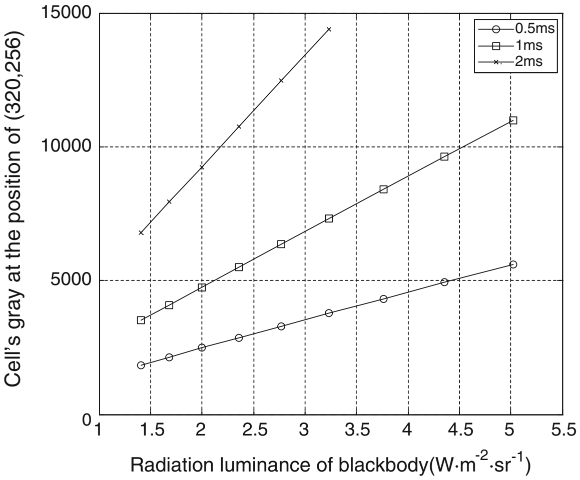


Table 38.1 The calibration results of pixel (160, 128)

Temperature of blackbody/°C	Integral time/ms	Radiance/ ($W \cdot m^{-2} \cdot sr^{-1}$)	Gray value/DN
40	5.5	1.9365	7,186
40	5	1.9365	6,607
60	5	3.6495	9,962

As is shown in Fig. 38.4, the gray value is linear to the radiance. Additionally, the responsivity increases when the integration time rises.

38.3.3 Verifying the Calibration Accuracy of the Simplified Method

A cooled infrared imaging system is used for calibration, and the parameters are as follows: the waveband is 3.7–4.8 μm and the pixel size is 30 × 30 μm, the F number is 2. The reference source is a high precision extended blackbody, namely SR800-20A of CI system, with the size 500 × 500 mm. The blackbody works in a range of 5–150 °C with the precision 0.01 °C and the emissivity is 0.97. According to the simplified calibration method, we select the blackbody temperatures $T_1 = 40$ °C and $T_2 = 60$°C, the integration times are $t_1 = 5.5\,\text{ms}$ and $t_2 = 5\,\text{ms}$. The calibration results of pixel (160, 128) are shown in Table 38.1.

The radiometric radiation model is determined by adopting Eq. (38.8).

$$h = t \times (391.7104 \times L_B(T_b) + 399.4528) + 817.0000 \qquad (38.11)$$

The calibration results at integration times 0.8, 2.5, 3, 3.5, 4 and 4.5 ms are calculated using Eq. (38.11), and then compared with the real ones, as is shown in Fig. 38.5. The temperatures of the blackbody vary from 30 to 90 °C with the interval 5 °C.

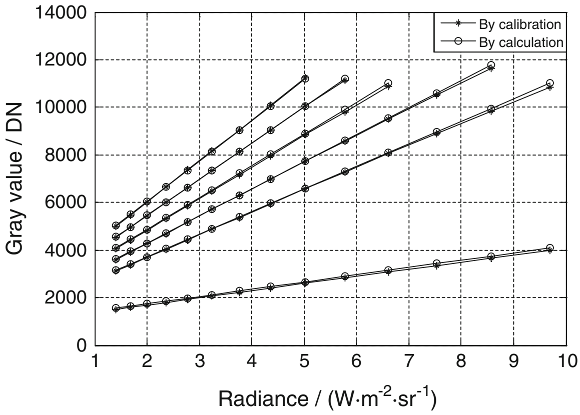


Fig. 38.5 Errors between actual calibration results and calculated ones

The maximum error between the actual calibration results and calculated ones is 2.75 %, which demonstrates that the simplified calibration method ensures the accuracy of radiometric calibration. Therefore the simplified calibration method is valid for infrared imaging systems to improve the efficiency of radiometric calibration.

38.4 Radiometry Adopting the Simplified Calibration Method

38.4.1 Experiment Introduction

In order to verify the precision of radiometry using the proposed simplified calibration method, radiometry experiments in the field are performed. The optical system has a diameter of 400 mm, and a focal length of 800 mm. The target, a 100 mm × 100 mm size blackbody, is placed at a distance of 500 m. The temperature of the target varies from 0 to 125 °C.

38.4.2 Radiometric Calibration

The extended area blackbody SR800-20A of CI systems is used for radiometric calibration of the infrared imaging system. The integration times are selected as 5.5 and 5 ms, Fig. 38.6 is the experimental set-up of radiometric calibration.

38.4.3 Calculation of Atmospheric Transmittance

During the experiments, the ambient temperature, measured by the meteorological station, is −3.4 °C. The atmospheric pressure is 983.5 hPa, the relative humidity is

Fig. 38.6 Equipment drawing of radiometric calibration

Fig. 38.7 Mid-wave infrared image at a distance of 500 m

58.9 %, and the visibility is 33 km measured by a Visibility Meter. The average atmospheric transmittance in waveband 3.7–4.8 μm is $\tau_{atm} = 0.7222$, which is calculated by MODTRAN using the above measured parameters, and the path radiance $L_{path} = 0.1175\ \mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mathrm{sr}^{-1}$.

Table 38.2 Radiation measurement results of infrared targets

Integral time/ms	Temperature/°C	Radiance/ $\mathrm{W{\cdot}m^{-2}{\cdot}sr^{-1}}$	Gray value/DN	Measured radiance/ $\mathrm{W{\cdot}m^{-2}{\cdot}sr^{-1}}$	Measured temperature/°C	Error of radiance/%
2.5	40	1.9964	3,423	2.1639	42.41	8.39
	50	2.7670	3,956	2.9409	51.94	6.28
	60	3.7624	4,663	3.9714	61.82	5.55
	70	5.0273	5,707	5.4933	73.18	9.27
	80	6.6107	6,938	7.2877	83.70	10.24
	90	8.5658	8,331	9.3183	93.37	8.78
3.5	40	1.9964	4,469	2.1677	42.46	8.58
	50	2.7670	5,300	3.0329	52.92	9.61
	60	3.7624	6,246	4.0179	62.22	6.79
	70	5.0273	7,605	5.4329	72.78	8.07
	80	6.6107	9,366	7.2665	83.59	9.92
	90	8.5658	11,427	9.4124	93.78	9.88
4.5	40	1.9964	5,589	2.2297	43.32	11.69
	50	2.7670	6,553	3.0103	52.68	8.79
	60	3.7624	7,840	4.0526	62.51	7.71
	70	5.0273	9,501	5.3977	72.54	7.37
	80	6.6107	11,719	7.1939	83.20	8.82
	90	8.5658	14,560	9.4946	94.13	10.84

38.4.4 Radiometry of an Infrared Target

Images of the targets at temperatures including 40, 50, 60, 70, 80 and 90 °C are collected, and the ambient temperatures of the infrared imaging system is set at 2.5, 3.5 and 4.5 ms. The gray image of the target at a distance of 500 m is shown in Fig. 38.7.

By using the calibration results calculated by the simplified calibration model, we obtain the radiometry results in Table 38.2.

It is illustrated in Table 38.2 that the maximum error of radiometry is 11.69 % using the proposed simplified calibration method. In addition, the error is quite stable, and it does not change with the variation of the integration time and the temperature of the target. In conclusion, the simplified calibration method ensures the accuracy of calibration and radiometry while improving the calibration efficiency.

38.5 Conclusion

A radiometric calibration formula considering the integration time, as the base for the simplified calibration method we developed, is proposed in this paper and verified by experiments. Calibration and Radiometry experiments are performed in the field, and the results indicate that the maximum calibration and radiometry using the simplified method are 2.75 and 11.69 %, respectively.

In summary, the proposed method ensures the accuracy of radiometric calibration and measurement, while improving the efficiency of calibration, by avoiding the complex operation of traditional calibration methods in which the calibration steps are required to be carried out once at each selected integration time. Therefore, the simplified calibration method is of great significance for researches on radiometry calibration theories, besides it is valid for radiometry missions in the field.

References

1. Tang J, Gao X (2006) Calibration method for radiation characteristics measuring system of test range. Infrared Laser Eng 35(10):266–270
2. Li XS, Ren JW, Zhang LG et al (2006) Research on a radiometric calibration device for a large aperture infrared opto-electric on spot. J Optoelectron Laser 17(2):175–178
3. Cao LH, Li N, Yang CY et al (2012) Radiance calibration for 3–5 μm infrared detector. Infrared Laser Eng 41(4):858–864
4. A study of the radiometric calibration of spectral bands in the mid wave infrared (MWIR) spectral range, 1.5–5.5 μm. In: Proceedings of SPIE 7300, 73000Z
5. Sun ZY, Zhu W, Qiao YF (2010) Amendment of gray drift in infrared temperature measurement. Chinese J Opt Appl Opt 3(4):391–396

6. Luo MJ, Zhou J, Fu J et al (2013) Integration time as variable for radiometric calibration of infrared system. Infrared Laser Eng 42(1):36–40
7. Holst G C (1933) Testing and evaluation of infrared imaging systems. JCD Publishing Co
8. Yang CY, Zhang JP, Cao LH (2012) Infrared radiation measurement based on proportional corrected atmospheric transmittance. Opt Precis Eng 20(7):1626–1635

Chapter 39
Analysis and Simulation of Intensity Correlation Imaging Noise Towards High-Orbit Satellite

Xiyu Li, Xin Gao, Changming Lu and Jia Tang

Abstract Intensity correlation imaging method can obtain high-resolution image by low-precision equipment on ground, which is a developing high-orbit satellite imaging method, but the method has some disadvantages such as low signal to noise ratio. In order to make clear the main noise source when observing high-orbit satellite by this method, analysis and simulations for the noise of intensity correlation imaging has been done. In this paper the theory and history of intensity correlation is described; the deviation and noise of intensity correlation imaging is analyzed according to the characteristic of high-orbit satellite; simulation about shot noise and partial coherence deviation has been done on semi-classical model. Analysis shows that the area of light-collecting mirrors affected the noise and deviation in intensity correlation imaging, the area should be optimized according to the diameter and brightness of satellite.

Keywords Intensity correlation · Partial coherence · Shot noise · Stray light noise · Dark current noise

39.1 Introduction

In the process of imaging satellites by ground-based optical telescope, the resolution is limited by atmosphere turbulence and the aperture of telescope, the image of low-orbit satellite can be obtained by large adaptive telescope. But due to the diffraction limit of telescope, 20 m aperture telescope is needed to obtain the image of high-orbit satellite, more the ten thousands adaptive cell is required to correct the influence of atmosphere turbulence. It is impossible to realized because of the difficulty in science and economy, new high-resolution imaging method is need to obtain the image of high-orbit satellite.

X. Li · X. Gao (✉) · C. Lu · J. Tang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: gaoxin526@sina.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_39

Intensity correlation imaging (ICI) is a developing optical synthetic aperture imaging method which can obtain high resolution with intensity interferometry array. In the imaging process, spatial frequency modulus of objects can be obtained by calculated the correlation of intensity fluctuations and the intensity distribution image can be reconstituted by measured data [1]. The imaging resolution of ICI is in proportion to the length of baselines, the image of high-orbit satellite can be obtain without using large adaptive telescope. The image quality of ICI is immune to atmosphere turbulence and lens aberration, high quality image can be obtained without using high-precision mirrors and adaptive system and the complexity and cost of imaging system can be reduced [2]. Intensity correlation imaging is firstly used in measuring the stellarangular diameter, with the development of electronics manufacturing technology and phase retrieval algorithm, the research imaging star and high-orbit satellite by intensity interferometry array is ongoing [3].

The major minus of intensity correlation imaging the low signal to noise ratio and large deviations of measured spatial frequency modulus when the observing target is dim, it decrease the quality of image which is reconstituted by intensity interferometry array. In this paper, we summarized the theory of intensity correlation imaging and analyzed the main factors of noise and deviations when observing high-orbit satellite. We simulated the noise of ICI on the basis of semi-classical model of light detection and imaging result of high-orbit satellite by ICI method. These analysis and simulations will provide a reference to the later research of improving the imaging quality of ICI.

39.2 Theory of Intensity Correlation Imaging

Intensity Correlation Imaging is a kind of synthetic aperture imaging technology. Compared to traditional amplitude interference imaging method, ICI method can obtain higher imaging resolution with low-precision equipment. Intensity correlation imaging method is firstly used in the measurement of stellar angular diameter, the resolution can be up to 0.5 milli-arcsecond [4]. In the imaging process, large mirrors in different position converge the light from object to high-frequency intensity detector in focal point of mirrors, the detectors convert the light intensity fluctuations to alternating current with the same frequency. According to Hanbury Brown and Twiss effect, we know that the correlation of intensity fluctuations in different position is in proportion to the square of light's spatial coherence γ:

$$|\gamma|^2 = \frac{\langle \delta i_1 \delta i_2 \rangle}{\langle i_1 \rangle \langle i_2 \rangle} \tag{39.1}$$

where $\langle\rangle$ represents the average in time domain, i_1 is the light current form detector. According to Van Citter-Zernike Theory, the spatial coherence of object's light field is in direct proportion to the spatial frequency of object's intensity distribution image, so we can reconstituted the image by ICI method [1, 5].

If the stellar image is simplified as a uniform disc, the stellar angular diameter can be calculated by the measured spatial frequency modulus. But due to the loss of phase information of spatial frequency, only the image's autocorrelation can be obtained by directly Fourier transform of measured data. Combine with phase retrieval algorithm and prior knowledge about object's space constrain, the image of object can be reconstituted with low noise and aberrations. In intensity correlation imaging process, the intensity interferometry array on the ground should be designed to acquire more non-repeating baselines to cover the spatial frequency plane uniformly. The imaging resolution is in direct proportion to the longest baselines and the view field is in reciprocal to the shortest baselines of the observing array, so the observing array of ICI should be arranged by the size of satellite [6].

Intensity correlation imaging method obtained the spatial frequency modulus of high-orbit satellite by calculated the correlation of intensity fluctuations from satellite in different position instead of amplitude interference of object's light field. In ICI imaging process, the optical path between satellite and different detectors needn't maintain the same strictly, the measure result is not sensitive to optical path difference caused by atmosphere turbulence and equipment errors, the difficulty of building long-baseline interferometry array can be reduced tremendously and shorter wavelength can be used in observing process to obtain higher resolution image. The baselines of Narrabri Intensity Interferometry which is built in 1960s is more than 170 m, the detection wavelength is less than 450 nm, the resolution in theory is even higher than that of the amplitude interferometer now we have.

When we observing object by intensity interferometry array on ground, only the intensity fluctuations of light from object is collected, the change of light field's phase information didn't influence measured spatial frequency modulus, atmosphere turbulence and mirror aberrations can't introduce significant measurement errors, so we can obtain low deviation result by low precision imaging equipment. On condition that the sample rate of intensity detectors which is used to collect object's intensity fluctuations is 1 GHz and optical path difference caused by mirrors is less than 0.1 m, the deviation of measured data caused by atmosphere and equipment is less than 10 %.

According to classical wave theory of light, all natural sources do not emit at an ideal monochromatic wavelength; instead the finite bandwidth of the light causes the light intensity to vary with unique beat frequencies as the multiple frequency components interfere with each other. These intensity modulations vary in a random and uncorrelated manner so that each element of the scene will vary its intensity. The frequency of intensity fluctuations is in direct proportion to the spectral width of satellite and much higher than the sampling rate of all detectors we have, the information contained in higher frequency intensity fluctuations can't be obtained by intensity correlation imaging method and it cause its signal to noise ratio is much lower than that of other interference method. Without considering eliminable noise, the SNR of measured spatial frequency modulus by ICI can be written as:

$$SNR = A\alpha n|\gamma|^2\sqrt{\Delta fT/2} \tag{39.2}$$

where A the area of is reflect mirrors, α is the quantum efficiency of intensity detector, n is the spectral density of object on the ground. Δf is the sampling frequency of detectors of intensity interferometry array. T is the observing time. In 1970s, high SNR measured result can be obtained on condition that observing star which is brighter than 2.5 Mv as long as 3 h by two large mirrors whose diameter is larger than 6.5 m [7]. With the improvement of sampling frequency and quantum efficiency of intensity detectors and the decrease of detector's dark current, the noise in intensity correlation imaging is much lower than before. The research about reconstituted stellar image by intensity correlation method using Cherenkov telescope array is undergoing. This array has more than one hundred detectors and light collectors, which can obtain more than then thousands pairs spatial frequency modulus of stellar intensity distribution image, with the help of high order correlation and phase retrieval algorithm, high resolution image of star can be obtain by Cherenkov telescope array [8].

39.3 Analyze of Noise and Deviation

Compared with stars which is far from earth, high-orbit satellites have the characteristics of smaller angular diameter and lower brightness. In the process of measuring high-orbit satellites' spatial frequency modulus by intensity correlation imaging method, there will be overmuch noise and deviation in the measured result. In this section we will analyze the source of noise and deviation of intensity correlation imaging when observing high-orbit satellite.

39.3.1 Partial Coherence Deviation

When observing incoherent radiation source on measuring plane, we assume the light field received in coherent area is complete coherence approximately [9], the coherent area of incoherent in distance can be expressed as follows:

$$A_C = (\lambda z)^2\Big/A_s = \lambda^2/\Omega_S \tag{39.3}$$

where A_S is the area of incoherent radiation source, Ω_S is the solid angle of radiation source to measuring plane.

From Formula (39.2) we know that the SNR of intensity correlation imaging can be promoted by enlarge the mirrors of intensity interferometers arrays which is used to collect intensity fluctuations form object. But when the area of light-collecting mirrors is larger than the coherence area of satellite, the light focused on intensity detectors is not completely coherence, in this condition, the spatial frequency of satellite calculated by intensity interferometry array will deviate from theory value, so the area of light-collecting mirrors in intensity interferometry array should be limited according to the size of target. Usually stellar angular diameter is less than 5 milli-arcsecond, the mirrors whose diameter is more than 20 m can be used to collect the intensity fluctuations without introducing overmuch deviation. At the same time the angular diameter of high-orbit satellite is more than 100 milli-arcsecond, the diameter of light-collecting mirrors should be limited less than 1 m to guarantee that the deviation caused by partial coherence is less than 10 % of real value.

When the light-collecting mirrors of intensity interferometry array is larger than coherent area of satellite, the spatial frequency modulus measured by intensity interferometry will deviate real value, the degree of deviation of measurement is related to object's real spatial frequency modulus and the shape of light-collecting mirrors. For example, when we measure the spatial frequency modulus of circle star whose brightness is uniform by intensity interferometry, the measured result $\Delta|\gamma|^2$ can be expressed as follows:

$$\Delta|\gamma|^2 = \frac{1}{A_1 A_2} \iint\limits_{A_1} \iint\limits_{A_2} \left[\frac{2J(\xi)}{\xi}\right]^2 dx_1 dx_2 dy_1 dy_2 \tag{39.4}$$

where A_1 and A_2 is the region of light-collecting mirrors, ξ is the coordinate of spatial frequency, it can be expressed as $\xi = (\pi\theta v_0/c)\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. When the size and shape of light-collecting mirrors are the same with each other, the measured result is approximately equal to the convolution of real spatial frequency and region of mirrors.

39.3.2 Shoot Noise

Because of partial coherence deviation of intensity correlation imaging, the size of light-collecting mirrors of intensity interferometry array should be limited less than 1 m which measuring the spatial frequency modulus of high-orbit satellite. However, most high-orbit satellites' brightness is 16 Mv approximately, the intensity fluctuations from satellite collected by small aperture mirrors is much lower than the intensity fluctuations of stars which are collected by large light-collecting mirrors. In a sampling period of high-frequency detectors, only a few photons from satellites are received by detectors. So the fluctuations of light current include the classical fluctuations caused by non-monochromatic light and shot noise caused by faint light received by high-frequency detectors. In the intensity correlation imaging process towards high-orbit satellite, shot noise is the main source of measure noise.

On condition that the light intensity received by high-frequency detectors is dim, the intensity fluctuations of satellite can be seen as the photon density in time domain, the correlation of intensity fluctuations in different position can be equivalent to the probability of photons reach different detectors in the same time. Here we will analyze the influence of shot noise in intensity correlation imaging process on the basis of semi-classical model of light detection, in which only the photoelectric translation process of light detection is calculated by quantum mechanics theory, other process such as the light propagation and refraction is calculated by classical wave theory about lightfiled, it can simplified the computing process vastly.

According to semi-classical model, in a ultra-short time horizon, the probability of generate one photoelectron by detectors is in direct proportion to the intensity of light, the length of sampling period and the area of photosensitive surface, the generation of photoelectron is mutually independent between different detectors and different time period. When the light intensity received by detectors is constantly, the statistical distribution of generating K photoelectron in time period $(t, t+\tau)$ obey Poisson distribution, the probably can be expressed as follows:

$$P(K)_{con} \equiv P(K; t, t+\tau)_{con} = \frac{(\bar{K})^{K}}{K!} e^{-\bar{K}} \tag{39.5}$$

But in process of intensity correlation imaging, the light intensity on detectors fluctuate in high frequency. The statistical distribution function of the intensity received by detector can be expressed as $P_W(W)$, in which time the probability of generating K photoelectron in sampling period can be written as follows:

$$P(K) = \int_{T}^{T+\tau} P(K|W) P_W(W) dW = \int_{T}^{T+\tau} \frac{(\alpha W)^{K}}{K!} e^{-\alpha W} P_W(W) dW \tag{39.6}$$

In this expression, the photoelectric conversion coefficient α can be expressed as $\alpha = \eta / \hbar v$, η is the quantum efficiency of detectors. From above equation, the variance of photocurrent export by intensity detectors can be expressed as $\sigma_K^2 = \alpha \bar{W} + \alpha^2 \sigma_W^2$, in which $\alpha \bar{W}$ is the shot noise caused by quantum effects of light in light detection process, $\alpha^2 \sigma_W^2$ is the fluctuation of photocurrent caused by the variation of light intensity itself. In the process of calculating spatial frequency modulus by ICI method,the shot noise in different detectors is not coherent, so the shot noise can only introduce noise rather than deviation in the measurement, the measurement noise obey zero-mean-value Gaussian distribution approximately.

39.3.3 Other Noise

In the process of intensity correlation imaging, the light from satellite should be collected by low-precision mirrors and detected by high-frequency detectors on the focal point. Due to the large aberrations of light-collecting mirrors, the stray light from other object will be also detected by high-frequency detectors. At the same time, the dark current noise of detectors will also disturb the measurement result. These extra noise from different detectors have no coherence with each other, but they can improve the average of photocurrent and lead the mathematical expectation of measured spatial frequency modulus lower than real value. In addition, if the correlation of photocurrent is calculated by electronic analog multiplier, there will be vast noise caused by the distortion of photocurrent. The brightness of high-orbit satellite is much lower than that of star, dark current noise and stray light noise will introduce much noise in intensity correlation imaging, it should be considered when observing high-orbit satellite by ICI method.

39.4 Simulations

In order to describe these noise in intensity correlation imaging towards high-orbit satellite intuitively, in this section we simulated the influence of partial coherence deviations and shot noise in ICI process when observing high-orbit satellite.

Firstly we simulated the partial coherence deviation caused by the extension of light-collection mirrors in intensity correlation imaging process, a uniform brightness disk whose angular diameter is 20 milli-arcsecond is observed by ground-based intensity interferometry. When the light-collecting mirrors are point, 5 and 20 m respectively, the measured spatial frequency modulus is shown in Fig. 39.1: the blue line represent the real value of object's spatial frequency modules; black lines represents the result obtained by 5 m mirrors and the red line represents the result obtained by 20 m mirrors.

Next we simulated the shot noise in the intensity detection of ICI process, the detecting wavelength is 500 nm and the spectral width is limited to 0.1 nm, the

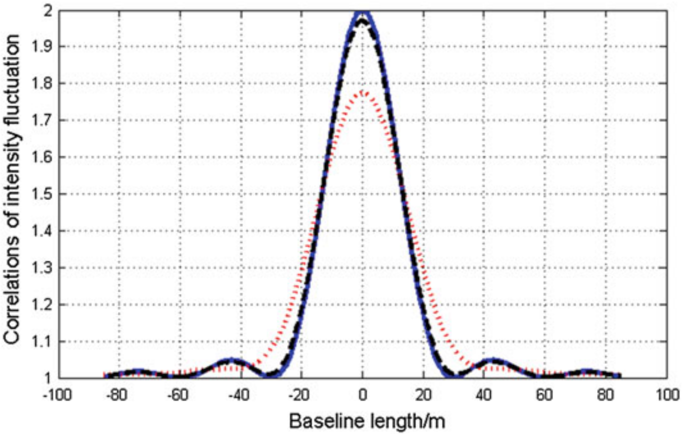


Fig. 39.1 Deviation caused by partial coherence effect

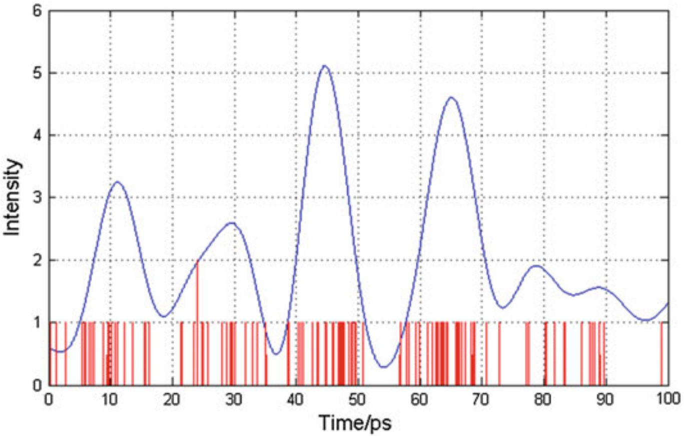


Fig. 39.2 Shot noise in light detection

luminous power received by detectors is 400 nW, the photon time sequence of intensity detector is shown in Fig. 39.2.

In this paper, the blue line represents for the classical fluctuations of light, red pulse represents the photon received by detectors, in this reason the fluctuations of photocurrent exported from detectors include a part caused classical light fluctuations and others caused by shot noise.

39.5 Conclusion

Due to the low-brightness and large angular diameter of high-orbit satellite, we can't calculated high SNR spatial frequency modulus of satellite by ground-based intensity interferometry array. The shot noise and partial coherence deviation is the main noise source of ICI. In order to reduce the partial coherence deviation, smaller light-collecting mirrors should be used to reduce measurement deviation; at the same time smaller mirrors of intensity interferometry array will reduce the light intensity received by high-frequency, it will lead to more shot noise and decrease the quality of image obtained by ICI method. The area of light-collecting mirror should be balanced by the brightness and angular diameter of high-orbit satellite.

From the analyze and simulations above we know that the main constrain of intensity correlation imaging towards high-orbit satellite is the low SNR of measured spatial frequency modulus, some research has been done to decrease the noise in observing process. Redundancy baselines method, laser illumination and coincidence counting method is the new trend of intensity correlation imaging towards high-orbit object.

References

1. Ofir A, Ribak EN (2006) Offline, multidetector intensity interferometers—I. theory. Mon Not R Astron Soc 368(4):1646–1651
2. Nuñez PD, Holmes R, Kieda D et al (2012) High angular resolution imaging with stellar intensity interferometry using air Cherenkov telescope arrays. Mon Not R Astron Soc 419 (1):172–183
3. Rou J, Nu N, Ez PD, Kieda D et al (2013) Monte Carlo simulation of stellar intensity interferometry. Mon Not R Astron Soc 430(4):3187–3195
4. Brown RH, Twiss RQ (1956) A test of a new type of stellar interferometer on Sirius. Nature 178 (1046):19
5. Fontanta PR (1983) Multidetector intensity interferometers. J Appl Phys 54:473–480
6. Dravins D, LeBohec S, Jensen H et al (2013) Optical intensity interferometry with the Cherenkov Telescope Array. Astropart Phys 43:331–347
7. Hyland D (2005) Exo-planet detection via stellar intensity correlation interferometry: 590511
8. Trahan III RE (2012) Phase retrieval with application to intensity correlation interferometers. Texas A&M University, Texas
9. Goodman JW (2004) Statistical optics. Courier Dover Publications, New York

Chapter 40
Research on Space-Based Missile Trajectory Characteristic Extraction Methods

Ying Ci, Shuang Wang and Yandong Li

Abstract It is very hard to estimate missile trajectory based on space observation because of multi-error sources and complicated estimation methods. In this paper, we derive the relationship of wavelet coefficient and function derivative. Wavelet scales selection rule and characteristic extraction methods based on white noise are also introduced. Simulation results show that the ratio derived in this paper is correct and the rule of selecting wavelet scales is practical. It has great application worth on extracting characteristic of missile trajectory quickly and exactly by using wavelet transform.

Keywords Missile trajectory · Wavelet transform · Characteristic extraction

40.1 Introduction

Ballistic missile is one of the main used weapons in the world because of its fast flight velocity, long shooting distance, powerful damage efficiency and high hit accuracy.

In order to defense ballistic missile launched from enemy state and offer early warning message of target, most country in the world such as America and Russia are construct their own ballistic missile defense system. Space-Based warning technology as an effective method becomes the major development direction in missile defense system. In missile defense process, warning satellite uses its infrared detector to detect azimuth angle through tail fire of powered-flight missile. If target is found, angle information is obtained through passive measurement at first. Missile trajectory is calculated base on measurement by information processing system. Then trajectory is used to determine and recognize launching position, range, impact point position and other information. According to these

---

Y. Ci (✉) · S. Wang · Y. Li
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: ciyinghappy@hotmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_40

information, free flight phase and re-entry phase trajectory is estimated. Finally, intercept missile will be launched by missile defense system. Accuracy of trajectory characteristic extraction is one of the most important factors in defense process.

Space-based trajectory estimation process is complicated because of different error sources such as satellite attitude, camera accuracy, distance between satellite and missile, environmental interference and calculate method error based on single satellite observation. However, trajectory accuracy is low so characteristic may be covered by white noise which makes extraction and estimation harder. Based on the above background, this paper derives the relationship of wavelet coefficient and function derivative. Wavelet scales selection rule and characteristic extraction methods based on white noise are also introduced. Simulation results show that trajectory characteristic are extracted quickly and accurately using wavelet transform under the wavelet scale selection rule which is derived in this paper. And it also has great application worth in future [1, 2].

40.2 Analysis of Trajectory Characteristics

According to different operational missions, ballistic missile can be divided into tactical missile, strategic missile and battle missile. These three types of missile have different maximum range that covers different areas. Ballistic missile flight process including powered-flight phase and unpowered-flight phase. Powered-flight phase including a vertical section, a turning section (pitchover turning section), transition section (gravity turning section) and stable section (aiming section). Unpowered-flight phase including free flight phase and re-entry phase [3, 4]. Maximum range and impact point of missile depends on position and velocity of cut-off point because there is no power in unpowered-flight phase. Trajectory characteristics reflect operational mission of missile to some extent. Therefore, it has very important significance in recognizing missile type correctly, determining combat intention and supporting anti-missile early-warning system.

Through analyzing, there are four types trajectory characteristics influence recognition that are thrust acceleration, cut-off point, range and maneuvering characteristics(as shown in Fig. 40.1). Cut-off point and maneuvering points information can be obtain by characteristic extraction methods. Thrust acceleration, maneuvering acceleration and range can be estimate through characteristic information and trajectory parameters (as shown in Fig. 40.2). Therefore, characteristic extraction is the precondition for subsequent characteristic calculation. This paper focus on characteristic extraction methods, so other estimation methods will not be mentioned.

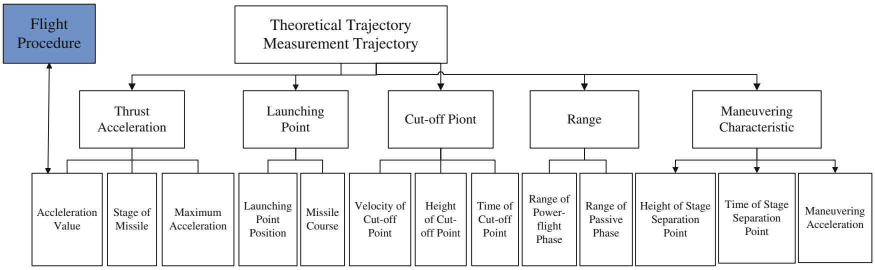


Fig. 40.1 The relationship of trajectory and trajectory characteristic

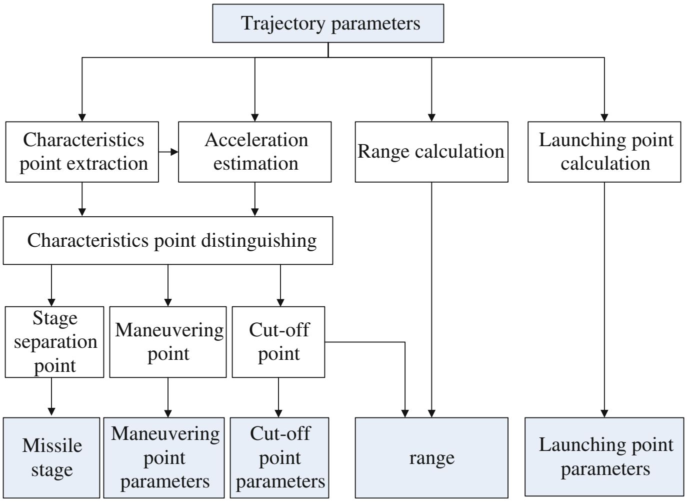


Fig. 40.2 Flow chart of trajectory characteristic extraction methods

40.3 Trajectory Characteristic Extraction Methods Based on Wavelet Transform

Missile force condition changes suddenly on stage separation, cut-off and maneuvering flight process. This sudden change causes singularity of a continuous measurement signal, so detection of these characteristic points are converted to detection of singularity in the time series. There two kinds of signal singularity. One is breakpoint that is amplitude changed at some time which cause non-continuous of signal. The other one is that it seems smooth and non-amplitude changed, but its first or second order differential changed. Missile force condition changed on stage separation, cut-off and maneuvering flight process which cause acceleration (second order differential) changed. So it belongs to second singularity point. It has good effect using frequency detection methods to detect this kind of singularity point. Well, wavelet transform is a good choice.

40.3.1 The Relationship of Wavelet Coefficient and Function Derivative

If trajectory is $\mathbf{x}(t)$, its wavelet coefficient is [5].

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \mathbf{x}(t) \boldsymbol{\psi}\left(\frac{t-b}{a}\right) dt \tag{40.1}$$

For easy to calculate and in order to improve time domain resolution, we usually choose short supported wavelet as basis function which has non-zero value on a finite interval. If it is assumed that support interval of wavelet is [0,1], then Eq. (40.1) can be expressed as Eq. (40.2).

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{b}^{aL+b} \mathbf{x}(t) \boldsymbol{\psi}\left(\frac{t-b}{a}\right) dt \tag{40.2}$$

If $t' = t - b$, Eq. (40.2) can be expressed as follows.

$$\begin{aligned} W(a,b) &= \frac{1}{\sqrt{a}} \int_{0}^{aL} \mathbf{x}(t'+b) \boldsymbol{\psi}\left(\frac{t'}{a}\right) dt' \\ &\triangleq \frac{1}{\sqrt{a}} \int_{0}^{aL} \mathbf{x}(t+b) \boldsymbol{\psi}\left(\frac{t}{a}\right) dt \end{aligned} \tag{40.3}$$

Using second-order Taylor expansion at point b, $\mathbf{x}(t+b) \approx \mathbf{x}(b) + t \cdot \mathbf{x}'(b) + \frac{1}{2}t^2 \cdot \mathbf{x}''(b)$. Take it into the above equation, then

$$\begin{aligned} W(a,b) &= \frac{1}{\sqrt{a}} \int_{0}^{aL} \mathbf{x}(t+b) \boldsymbol{\psi}\left(\frac{t}{a}\right) dt \\ &\approx \frac{1}{\sqrt{a}} \int_{0}^{aL} \left(\mathbf{x}(b) + t \cdot \mathbf{x}'(b) + \frac{1}{2}t^2 \cdot \mathbf{x}''(b)\right) \boldsymbol{\psi}\left(\frac{t}{a}\right) dt \\ &\overset{\tilde{t}=\frac{t}{a}}{=} \sqrt{a} \int_{0}^{L} \left(\mathbf{x}(b) + a\tilde{t} \cdot \mathbf{x}'(b) + \frac{1}{2}(a\tilde{t})^2 \cdot \mathbf{x}''(b)\right) \boldsymbol{\psi}(\tilde{t}) d\tilde{t} \end{aligned} \tag{40.4}$$

If we choose wavelet function whose vanishing moment is 2. We can convert Eqs. (40.4) to (40.5).

$$W(a,b) = \frac{1}{2}\sqrt{a} \cdot a^2 \cdot \mathbf{x}''(b) \int_0^L t^2 \psi(t)dt \tag{40.5}$$

If

$$\mathbf{Rat}(\psi, a) \triangleq \frac{W(a,b)}{\mathbf{x}''(b)} \tag{40.6}$$

Then

$$\mathbf{Rat}(\psi, a) = \frac{1}{2}\sqrt{a} \cdot a^2 \cdot \int_0^L t^2 \psi(t)dt \tag{40.7}$$

Form above formula, $\mathbf{Rat}(\psi, a)$ represents the relationship of function second order derivative and wavelet coefficient whose vanishing moment is 2. It only associated with wavelet scale a and wavelet function $\psi(t)$. Therefore, if a and $\psi(t)$ are selected, $\mathbf{Rat}(\psi, a)$ is a constant value. That means it has fixed proportion relation between wavelet coefficient and function derivation.

40.3.2 Characteristic of White Noise Under Wavelet Transform

White noise is one of the typical instrumentation errors in measurement. It is difficult to filtering by regulating scales because white noise has wide distribution in frequency domain. So it is necessary to analyze characteristics of white noise under different scales.

Theorem 1 If $n(x)$ is a real wide stationary white noise, its variance is σ^2, the expectation of white noise wavelet transform is $W_n(s,x)$ [5, 6]:

$$\mathbf{E}(|W_n(s,x)|^2) = \frac{\|\mathbf{\Psi}\|^2\sigma^2}{s} \tag{40.8}$$

That means attenuation of $\mathbf{E}(|W_n(s,x)|^2)$ is proportional to $1/s$. If scale increased, amplitude of wavelet transform deduced.

Theorem 2 If a white noise is a Gauss white noise, the average density of transform modulus in scale s is as Eq. (40.9) [5, 6].

$$ds = \frac{1}{s\pi}\left(\frac{\left\|\boldsymbol{\Psi}^{(2)}\right\|}{2\left\|\boldsymbol{\Psi}^{(1)}\right\|} + \frac{\left\|\boldsymbol{\Psi}^{(1)}\right\|}{\|\boldsymbol{\Psi}\|}\right) \tag{40.9}$$

where $\boldsymbol{\Psi}^1(x)$ and $\boldsymbol{\Psi}^2(x)$ are represent first and second order derivative of $\boldsymbol{\Psi}(x)$ respectively. This theorem shows that the average density of wavelet transform modulus is proportional to *1/s*. That means density will deduce during scale s increase.

According to these two theorems above, wavelet spectrum will disappear with increase of scale. It is easy to distinguish wavelet transform modulus from signal and white noise by observing the evolution of their wavelet transform modulus increased (or deduced) with scale.

40.3.3 Trajectory Extraction Methods Under White Noise Background

40.3.3.1 Choice of Wavelet Basis Function

According to derivation in Sect. 41.3.1, acceleration changed in result of missile force situation mutation. That means second order derivative $\mathbf{x}''(t)$ of $\mathbf{x}(t)$ has changed. If choose wavelet function whose vanishing moment is 2, expression of wavelet coefficient can be convert to Eq. (40.10).

$$W(a,b) = \frac{1}{2}\sqrt{a}\cdot a^2\cdot \mathbf{x}''(b)\int_0^L t^2\psi(t)dt \tag{40.10}$$

That means only in the region of acceleration mutation will produce wavelet coefficients correspondingly mutated. Wavelet function whose vanishing moment is 2 should be chose as basis function. Characteristics of wavelet function which used frequently are listed in Table 40.1 [7, 8] . According to db2 wavelet, sym2 wavelet and coif1 wavelet are all meet the requirements.

40.3.3.2 Wavelet Scale Selection

Wavelet transform scales affect the final accuracy. Usually, the smaller wavelet scale is, the higher detection accuracy of singularity is. However, due to the presence of measurement error, the choice of scale is not as smaller as better. So it

Table 40.1 Comparison of characteristic

Wavelet function	Orthogonality	Biorthogonality	Compactness	DWT	CWT	Support width	Symmetry	Vanishing moment
Haar	Yes	Yes	Yes	Yes	Yes	1	Yes	1
DBN	Yes	Yes	Yes	Yes	Yes	2N − 1	No	N
Boir	No	Yes	Yes	Can	Can	For rebuilding is 2Nr + 1,for resolving is 2Nd + 1	Yes	Nr
CoifN	Yes	Yes	Yes	Can	Can	6N − 1	Close to the symmetry	For wavelet function is 2N, for scale function is 2N − 1
SymN	Yes	Yes	Yes	Can	Can	2N − 1	Close to the symmetry	N
Morlet	No	No	No	Cannot	Can	[−4,4] Infinite	Yes	——
Mexican hat	No	No	No	Cannot	Can	[−5,5] Infinite	Yes	——
Meyer	Yes	Yes	No	Can	Can	[−8,8] Infinite	Yes	——
Gaussian derivatives	No	No	No	Cannot	Can	[−5,5] Infinite	Symmetrical if N is even; else is antisymmetric	——

is necessary to select an appropriate scale according to the actual situation. After maneuvering and stage separation, measure noise would increased due to attitude change of missile. More data in time is accumulated in order to inhibit random error. Thus large scale is selected correspondingly.

If measure signal denotes as Eq. (40.11)

$$\mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\varepsilon}(t) \tag{40.11}$$

where $\mathbf{x}(t)$ is a true signal, $\boldsymbol{\varepsilon}(t)$ is measure noise. Because the wavelet transfer WT is a lineal transformation, then

$$\mathrm{WT}\mathbf{y}(t) = \mathrm{WT}\mathbf{x}(t) + \mathrm{WT}\boldsymbol{\varepsilon}(t) \tag{40.12}$$

From above equation, wavelet transform result is the sum of true signal wavelet transform result and measure noise wavelet transform result. The wavelet transform scale is smaller, the detection precision is higher, but the result also can be affected by noise. If we choose a scale without analyzing, the true signal characteristics may be marked by noise. So it is important to choose scale according to true signal characteristics. Amplitude of true signal wavelet become larger by increasing of scale. But when scale increased, amplitude of white noise wavelet coefficient will not become larger. So it can easily extracting real signal from white noise.

In this paper, rule of selection the optimal wavelet transform scale under known trajectory noise is given. If we need select true signal from white noise in high probability, maximum of true signal wavelet transform coefficient should exceed three times of the average noise.

According to this rule, if we use wavelet function whose vanishing moment is 2 to transfer trajectory, we can select wavelet scale using maximum acceleration of target as a prior information which can be obtained in actual easily.

40.4 Simulation

In order to prove the relationship of wavelet coefficient and function derivative is correct, analyzing influence of white noise using wavelet transform and accuracy of trajectory characteristics extraction under different noise. This paper design a simulation trajectory under actual background. Suppose first stage separation time is 130 s, second stage separation time is 215 s, sample time is 0.01 s, maximum acceleration on Y direction is 40 m/s2, coif1 wavelet function is chose to calculate and the results are shown as follows.

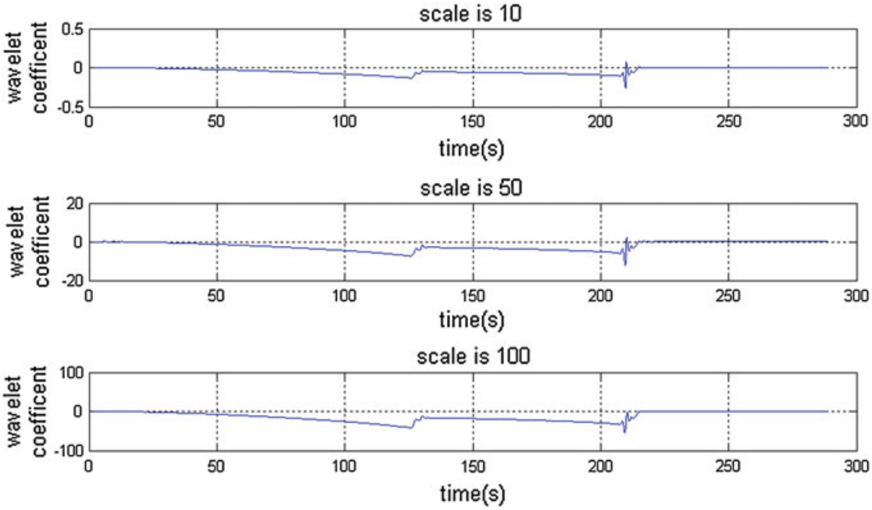


Fig. 40.3 The wavelet coefficient of Y direction data under different scales

40.4.1 *Simulation on Relationship of Wavelet Coefficient and Function Derivative*

40.4.1.1 Analysis of Wavelet Coefficient Under Different Scales

Using wavelet coefficient whose vanishing moment is 2 and scale are 10, 50, 100 respectively to transfer Y direction data. The result is shown on Fig. 40.3.

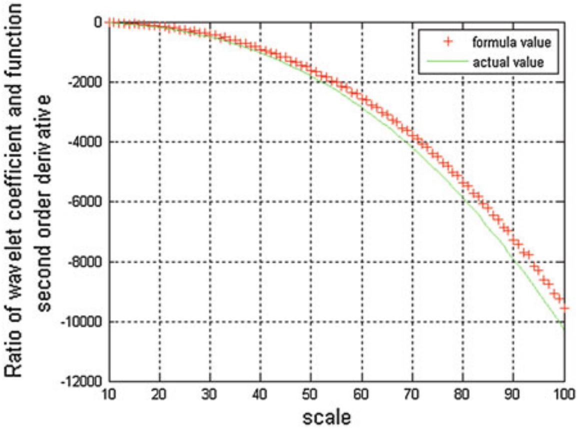
Simulation results show that it has difference in amplitude of wavelet coefficient under different scale. And the value of scale is not change the shape of wavelet coefficient.

40.4.1.2 Analysis Wavelet Coefficient and Function Derivative

In order to prove the relationship of wavelet coefficient and function derivative is correct. This paper wavelet transfer Y direction data using wavelet coefficient whose vanishing moment is 2. Red "+" is the result of Eq. (40.6) which is the ratio of wavelet coefficient and acceleration. Green "-" is the result of Eq. (40.7). The result is shown in Fig. 40.4.

As can be seen from simulation, for ratio of wavelet coefficient and second order derivative derive value and real value maintain a good consistency.

Fig. 40.4 Ratio of wavelet coefficient and function second order derivative



40.4.2 Analysis of White Noise Wavelet Transform

Design a white noise whose mean value is 0 and variance is 15. Using wavelet function whose vanishing moment is 2 to simulate. Wavelet coefficient of white noise under different scale is given in Fig. 40.5. Wavelet coefficient variance of white noise under different scale is given in Fig. 40.6.

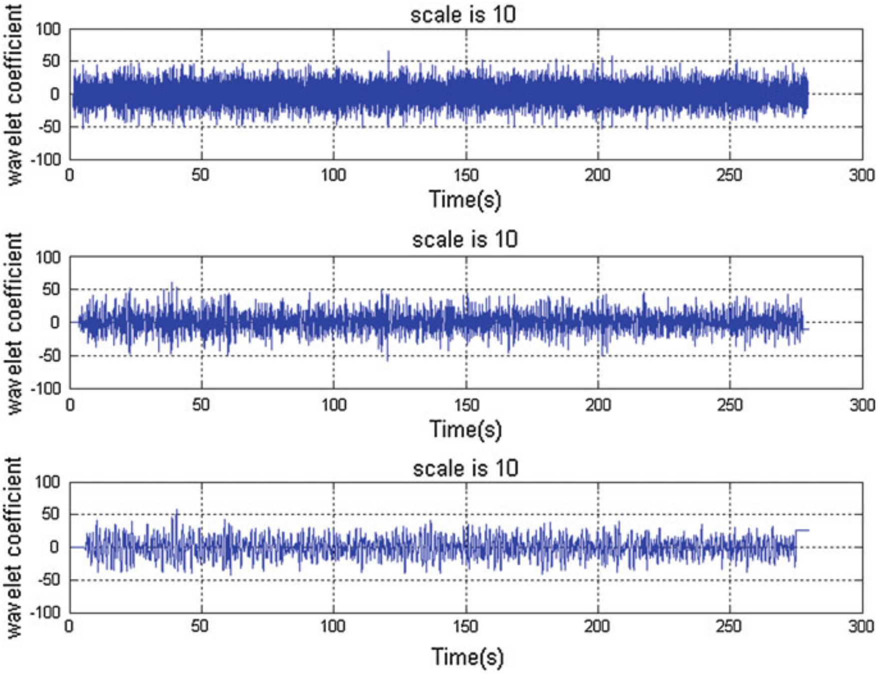


Fig. 40.5 Wavelet coefficient under different scales

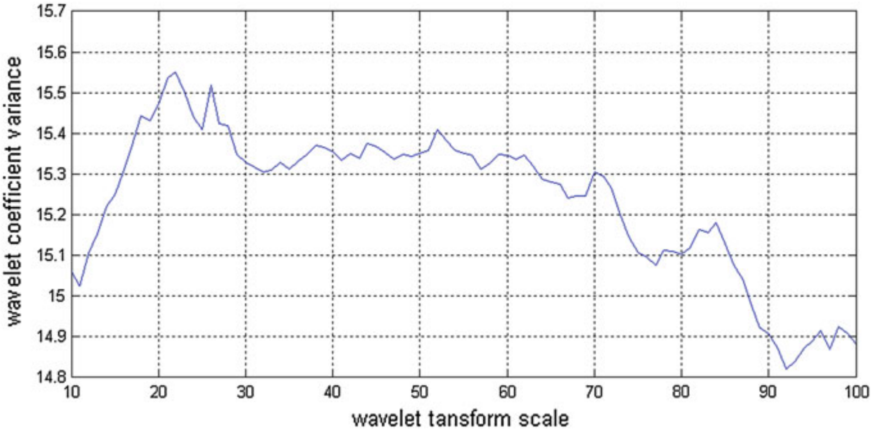


Fig. 40.6 Relationship of wavelet coefficient variance and scale

From simulation, the larger scale is, the smaller amplitude is. But variance contains 15 which has nothing to do with scale. That is to say wavelet transform cannot change the intensity of the noise.

40.4.3 Accuracy of Trajectory Characteristic Extraction

40.4.3.1 Simulation of Scale Selection Rule

In order to verify the accuracy of the selection rule, random noise of variance 10 is superimposed to simulation trajectory. Using coif wavelet function whose vanishing moment is 2 to transfer Y direction data. From prior information, the maximum acceleration on Y direction is 40 m/s^2. We should choose wavelet coefficient greater than 30 because variance of noise is 10. So we need the ratio of wavelet coefficient and second order derivative greater than 30/40 = 0.75. According to equations above, we can calculate the ratio of coif1 wavelet under different scales. Ratio of coif1 wavelet coefficient and second order derivative is shown in Fig. 40.7.

When scale is 87, the ratio is 0.75. To verify the accuracy of above rule, we transfer to Y direction data with wavelet scale 10,87,200 respectively. Results are shown in Fig. 40.8.

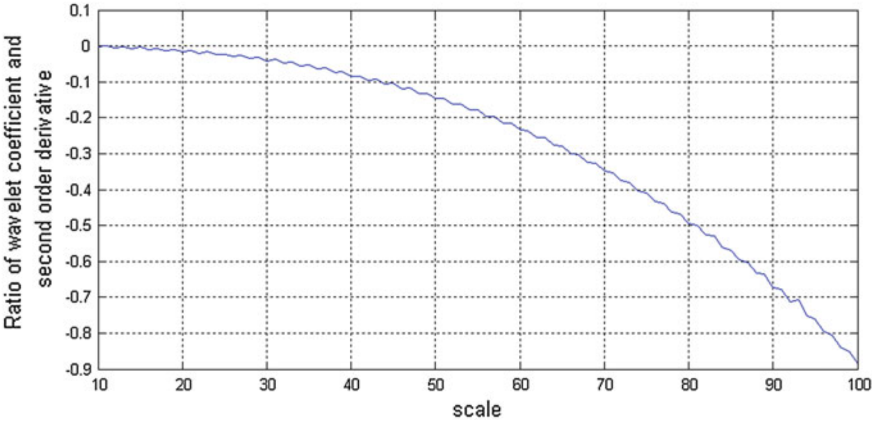


Fig. 40.7 Ratio of coif1 wavelet coefficient and second order derivative

Form simulation, when scale is 10, we cannot extract characteristic because characteristic of theory wavelet coefficient is totally marked with noise. when scale is 87, wavelet coefficients of signal and white noise have obvious difference, so we can extract characteristic of signal preliminarily. With the increasing scale, wavelet coefficients feature become increasingly clear from which we can clearly distinguish trajectory characteristics.

40.4.3.2 Simulation of Characteristic Extractive Accuracy

According to above rule, if we want to extract characteristic information, scale must greater than 87. In order to verify accuracy of characteristics extraction under white noise background, random noise of variance 10 is superimposed to simulation trajectory. Using coif wavelet function whose vanishing moment is 2 to transfer Y direction data. Mont Carlo simulation 100 times and the RMS of characteristics are given in Figs. 40.9 and 40.10.

The result show that extract error of first cut-off time is less than 0.96 s. When scale between 137 and 167 which extract error less than 0.65 s, it appears the highest extracting accuracy. Extracting error of second cut-off time is less than 1.6 s. When scale between 147 and 167 which extract error less than 1.35 s, it appears the highest extracting accuracy.

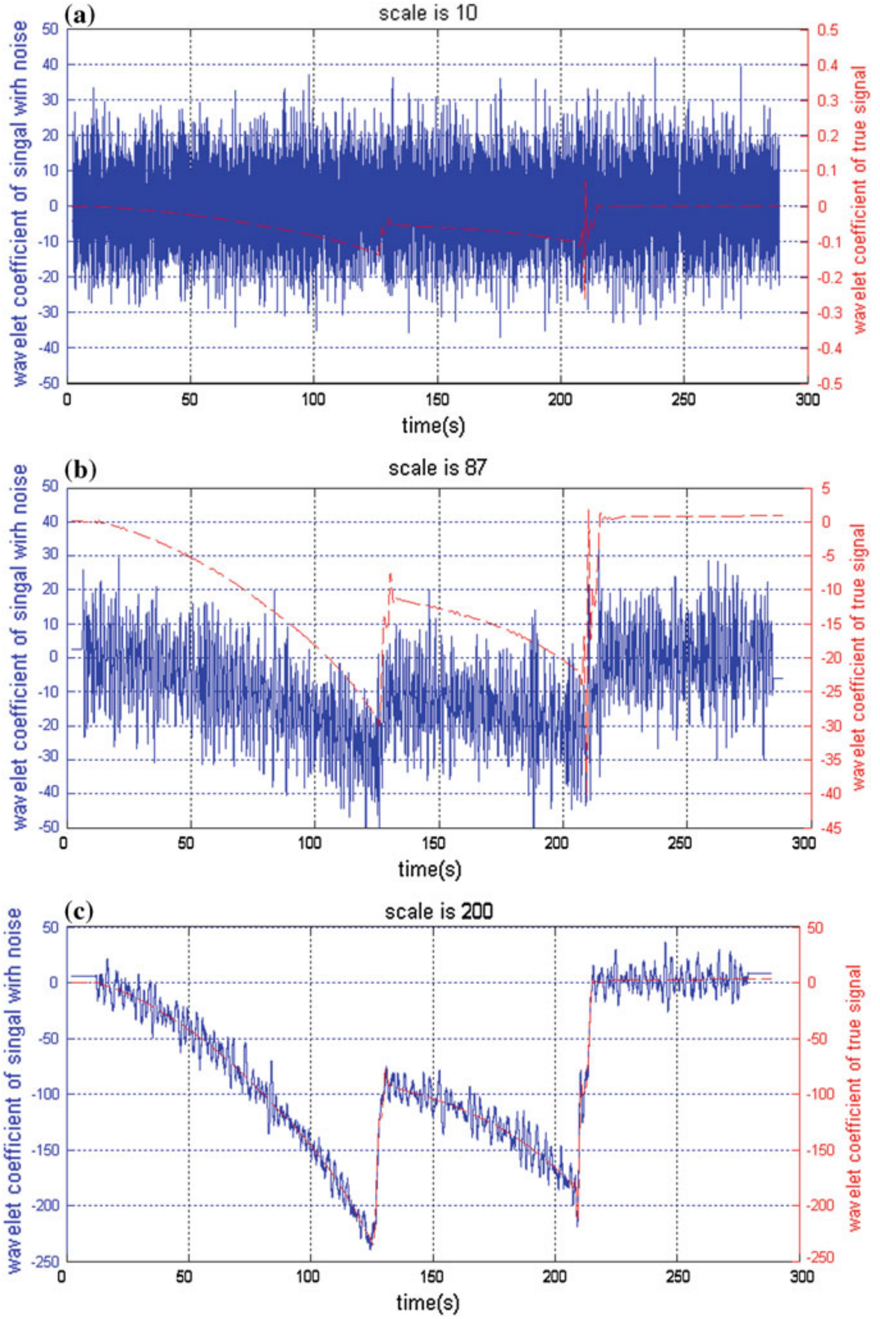


Fig. 40.8 Wavelet coefficient under different scale on Y direction **a** scale is 10 **b** scale is 87 **c** scale is 200

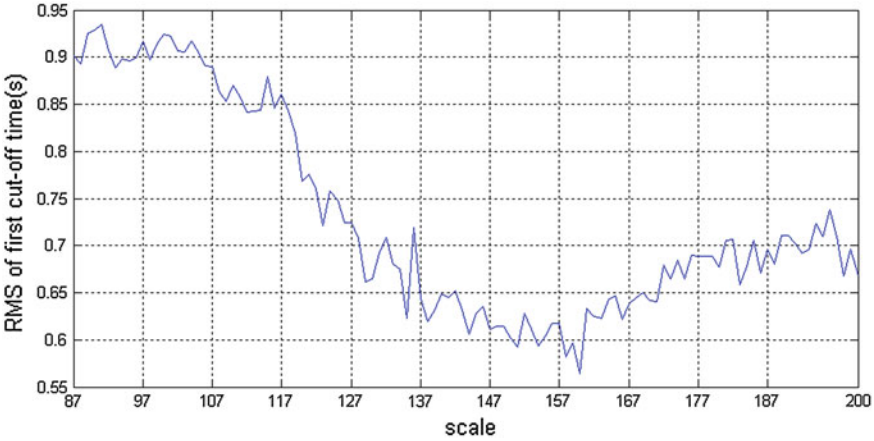


Fig. 40.9 RMS of first cut-off time

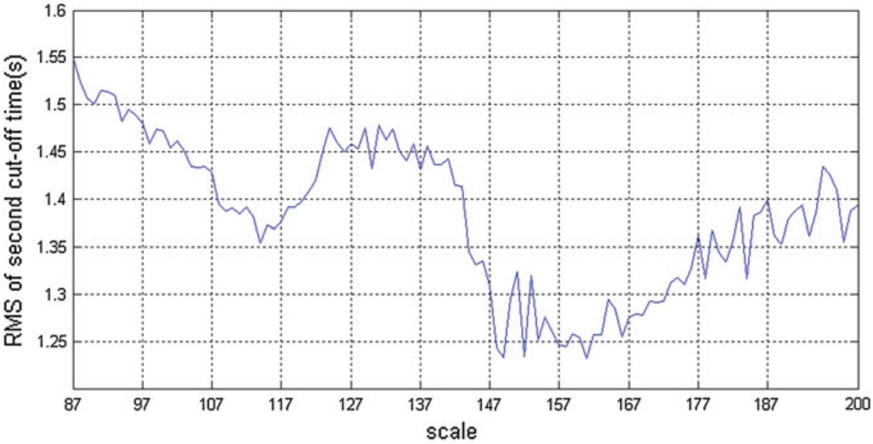


Fig. 40.10 RMS of second cut-off time

40.5 Conclusions

The research show that if we choose scale based on the derived ratio of wavelet coefficient and function derivative, the situation of failure extract characteristic information can be avoided when we transfer to a signal with noise. The speed and precision of trajectory feature extraction is guaranteed. If this method is used for missile early warning system, we can quickly extract missile characteristics and lay the foundation for the rapid identification of missile target and give active defense.

References

1. Qiang S, Shen Z, Yi DY (2011) Method of early warning for a single satellite based on missile dynamic characteristics. Syst Eng Electron 33(10):2234–2238
2. Guo JL, Li J, An W (2011) Performance analysis of boost-phase trajectory estimation in space early warning system. Electron opt and control 14(2):50–54
3. Shen Z (2010) The single satellite early warning ballistic missile parameter estimation research. Doctoral thesis of national university of defense technology, Changsha
4. Shen Z (2010) Research on estimating missile trajectory characteristic based on single satellite. J syst simul 22(2):295–297
5. Li JP (1997) Wavelet analysis and signal processing theory, applications and software implementations. Chongqing publishing house, Chongqing
6. Goswami JC (2007) Fundamentals of wavelets: theory, algorithms, and applications. National defence industry press, Beijing
7. Zhang DF et al (2012) Matlab wavelet analysis. China Machine Press, Beijing
8. Daubechies I (2011) Ten lectures on wavelets. National defence industry press, Beijing

Chapter 41
A Rapid Method of Extracting Unintentional Modulation Feature Using Gabor Atoms

Siji Quan, Weiping Qian and Junhai Guo

Abstract Radar emitter signal recognition is currently the hot field of radar countermeasures. The unintentional modulation feature of radar signal is unique and stable, thus becoming a good choice of radar emitter feature while analyzing. In this paper, we use matching pursuit (MP) algorithm with Gabor atoms to apply atomic decomposition to emitter signal added with unintentional modulation, and successfully extracted such modulation features. These features can be effectively used to restructure radar signal and distinguish radar emitters. Due to the compute-intensive feature of MP algorithm, we introduce a quantum genetic algorithm (QGA), which reduces the computational complexity of each step when signal decomposing.

Keywords Radar emitter · Unintentional modulation · Time-frequency atomic decomposition · Matching pursuit · Quantum genetic algorithm

41.1 Introduction

Radar confrontation is a major form of electronic warfare, including radiation detection as an important part. Radiation detection system is the core of integrated electronic warfare radar system. Radar emitter signal recognition is the key technique measuring whether radar confrontation techniques are advanced or not, and thus becoming a hot spot in study. We need to extract useful features from radar signals in order to recognize radar emitter.

Unintentional modulation is caused by the inherent properties of components and circuits used in radar transmitter, producing slight yet unique effect on radar emitters. The main reflection of unintentional modulation is on phase noise, and such feature cannot be affected artificially. Different radar radiation sources holds

---

S. Quan (✉) · W. Qian · J. Guo
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: quansiji@gmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_41

different, unique and stable features, which can be treated as radiation's "fingerprint". Since unintentional modulation holds such excellent features, it's attracting attention of researchers from home and abroad.

Currently, in order to extract unintentional modulation features from radar signals, there are ways extracting features from frequency stability, ways analyzing signal spectral features, ways using wavelet transform processing on pulse signal, and ways using wavelet packets analyzing. These methods analyze signal features in frequency domain, time domain and other transform domains separately to achieve unintentional modulation features extraction.

Time-frequency atomic decomposition algorithm (TFAD) is a new signal processing method that is gradually developing. This method decompose input signal into a series of linear atomic signals chosen from a redundant atomic set. The input signal is divided into a small amount of component atoms with obvious features, while meeting the accuracy requirements. Gabor atomic set and Chip atomic set is commonly used currently. This algorithm can do a good job in signal decomposition and building signal models, but its drawback is its excessive calculation. Every step obtaining a component atom requires a lot of inner product calculation.

To raise the speed of TFAD algorithm, we introduce quantum genetic algorithm (QGA). When selecting current best atom from the atomic set, QGA can reduce computational complexity and speed up TFAD, while still maintain a relatively good reconstruction capability, which is enough to extract unintentional modulation feature and distinguish radar emitter signals.

41.2 Analysis of Radar Emitter Phase Noise Features

In this paper, phase-locked loop (PLL) frequency synthesizer with signal loop is used for analysis. The phase noise of each part in frequency synthesizer can be considered as statistically independent stationary process, and therefore we can calculate total PLL phase noise by simply summing the phase noise power spectrum of components in PLL [1]. The structure of PLL frequency synthesizer is shown in Fig. 41.1.

The model of PLL frequency synthesizer is [1]

$$\begin{aligned} L_{\Phi,FS}(f) = 10\log\Bigg[&\left|N_{ref}(f)G_c(f)\frac{N}{R}\right|^2 + |N_{R_2}(f)|^2 \\ &+ \left|N_{VCO}(f)\frac{1}{1+G_o(f)}\right|^2 + |N_{R_1}(f)|^2 \\ &+ \left|\big(N_{pd}(f)+N_{DN}(f)+N_{DR}(f)\big)G_c(f)N\right|^2\Bigg] \end{aligned} \tag{41.1}$$

where $N_{part}(f)$ is the phase noise power spectrum of corresponding "*part*" component. N and R are parameters of dividers. $G(f)$ is transfer function. About how to

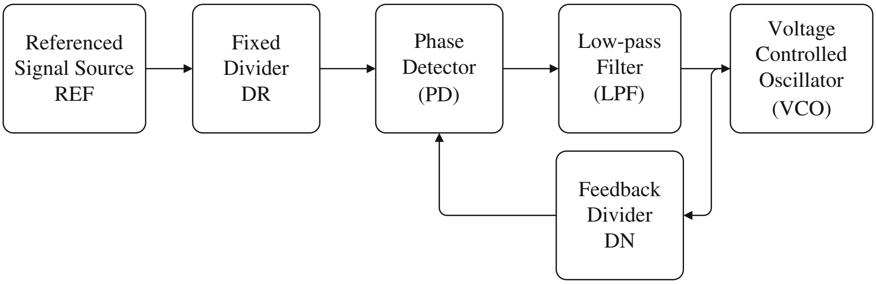


Fig. 41.1 The structure of PLL frequency synthesizer

select LPF, characteristics of various components noise power, and how to derive this transfer function, please refer [1] and [2].

41.3 TFAD Algorithm Based on Matching Pursuit (MP)

MP algorithm states as follows [3]:

The chosen time-frequency atomic set is $\boldsymbol{D} = \{g_r(t)\}$, and all atoms from this set build a Hilbert space $\boldsymbol{H} = R^N$. For any signal $s \in \boldsymbol{H}$, we can choose a subset $\boldsymbol{B} \subseteq \boldsymbol{D}$ that satisfies s linear expandable in $\boldsymbol{B}$:

$$s = \sum_{h=0}^{H-1} c_{\gamma_h} g_{\gamma_h} \tag{41.2}$$

where $g_{\gamma_h} \in \boldsymbol{B}$ and $H = card(\boldsymbol{B})$. c_{γ_h} is projection of signal s on atom g_{γ_h}, namely the coefficient of corresponding atom in linear decomposition.

In order to reach a sparse signal approximation, the expanding set $\boldsymbol{B}$ should be as small as possible. If all atoms are orthogonal in $\boldsymbol{D}$, as is in wavelet decomposition and in Fourier decomposition, subset $\boldsymbol{B}$ is composed of the H atoms having maximum absolute inner product with signal s. But in TFAD, $\boldsymbol{D}$ is over-complete atomic set, making this simple solution infeasible. To achieve feature extraction, limited times of signal decomposition is needed, which will surely produce residuals. As was proved in ref. [4], looking for an optimal approximation to a energy limited signal with given length is NP-hard. To reduce residuals left after given times of decomposition, an algorithm which can efficiently choose subset $\boldsymbol{B}$ is a must.

When carrying out subset $\boldsymbol{B}$, there exist a number of methods, including basis pursuit (BP), matching pursuit (MP), and orthogonal matching pursuit (OMP). MP is proposed by Mallat. This method is a greedy algorithm, and it joins the best matching atom into subset $\boldsymbol{B}$ in each iteration process. MP algorithm gives up finding a global optimal solution, but instead it finds local optimal solution at each decomposition, which is the feature of greedy algorithm. Finally we obtain a solution that is optimum at each step. The specific process of MP algorithm describes as follows:

For an atomic set $\boldsymbol{D} = \{g_r(t)\}$ that satisfies $\|g_r(t)\| = 1$, let iteration times count $h = 0$, and residuals left after h times of iteration is R_h, namely $R_0 = s$. Break down signal:

$$R_0 = \langle g_{\gamma_0}, R_0 \rangle g_{\gamma_0} + R_1 \tag{41.3}$$

getting the first decomposition atom, its coefficient and residuals. Choose atom as

$$\left|\langle g_{\gamma_0}, R_0 \rangle\right| = \sup_{g_\gamma \in \boldsymbol{D}} \left|\langle g_\gamma, R_0 \rangle\right| \tag{41.4}$$

Repeat decomposition for H times, and we get

$$s = \sum_{h=0}^{H-1} \langle g_{\gamma_h}, R_h \rangle g_{\gamma_h} + R_H \tag{41.5}$$

while choosing the *h*th atom, it satisfies

$$\left|\langle g_{\gamma_h}, R_h \rangle\right| = \sup_{g_\gamma \in \boldsymbol{D}} \left|\langle g_\gamma, R_h \rangle\right| \tag{41.6}$$

According to Eq. (41.3) and references [5] and [6], signal energy satisfies

$$\|s\|^2 = \sum_{h=0}^{H-1} \left|\langle g_{\gamma_h}, R_h \rangle\right|^2 + \|R_H\|^2 \tag{41.7}$$

The energy of signal s gradually disperses into each atom as iteration continues. So for an energy limited signal with given length, residuals energy is approaching zero as iteration continues, thus ensuring the validity of MP algorithm.

In each iteration, choosing the optimal atom needs to calculate the inner product of signal and each atom from the whole atomic set. There are 4 parameters for atoms in Gabor atomic set, and 5 for atoms in Chirp atomic set. As signal length increases, the valid range of every parameter will expand, causing an explosive increase on the size of atomic set. In order to speed up searching optimal atom, we introduce quantum genetic algorithm (QGA).

41.4 Quantum Genetic Algorithm (QGA)

QGA introduces the concept of quantum theory info genetic algorithms. Chromosomes no longer holds determined value, and instead they holds qubits, namely quantum bits, whose value is between 0 and 1. A qubit can be expressed as

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{41.8}$$

where $\alpha, \beta \in \mathbb{C}$ are the probability amplitudes of state 0 and state 1, satisfying the normalization condition

$$|\alpha|^2 + |\beta|^2 = 1 \tag{41.9}$$

Then $|\alpha|^2$ and $|\beta|^2$ represent the probability of 0 and 1 respectively. Using qubits, the format of individual chromosome is $[\alpha_1\alpha_2\ldots\alpha_M\beta_1\beta_2\ldots\beta_M]$, in which α_i and β_i are paired amplitude probability, and M is the total qubit number same as the binary length sum of all undetermined parameters. Then the procedures of QGA are [7]:

Population initialization. Set population size to n, and population $p = \{p_1, p_2, \ldots, p_n\}$. The chromosome of some individual $p_j(j = 1, 2, \ldots n)$ is $p_j = [\alpha_1\alpha_2\ldots\alpha_M|\beta_1\beta_2\ldots\beta_M]_j$, where M is the total qubit number. In order to meet the normalization condition, all the probability amplitude $\alpha_i, \beta_i\,(i = 1, 2, \ldots M)$ is initially set to $2^{-1/2}$. Note this population as generation 0.

Observed states construction. According to the probability amplitude in population p, observe all individuals' states $R = \{r_1, r_2, \ldots, r_n\}$, where $r_j\,(j = 1, 2, \ldots n)$ is the observed state of jth individual. r_j is a binary string with a length of M. Observing is comparing amplitude probability with a random number to decide whether the observed qubit is 0 or 1.

Fitness evaluation. Use the binary string obtained in the previous step to encode the parameters of atomic set. Then the atoms in the population is determined. Calculate the inner product of each individual atom and signal, and evaluate fitness according to Eq. (41.4). If there exists any individual that has reached the stop condition, stop iteration or continue to the next step.

Population adjustment. Use quantum revolving door G to update all individuals' probability amplitude, and then back to step two. Quantum revolving door expresses as

$$G = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{42.10}$$

where θ is the revolving angle. The revolving angle for ith qubit is

$$\theta_i = k \cdot f(\alpha_i, \beta_i) \tag{41.12}$$

where k is the convergence coefficient associated to evolution algebra, and $f(\cdot)$ is a function to adjust the revolving direction using the probability amplitudes of both current individual and the optical individual. The aim of quantum revolving door is to approach the optimal solution.

41.5 Simulation

41.5.1 Design of Simulation Parameters

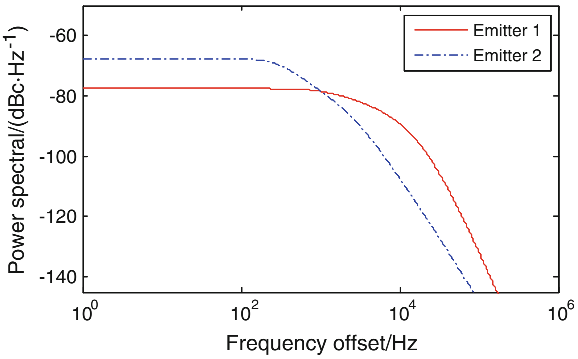
To achieve the purpose of distinguish radar emitter signal, we design the parameters of two radar emitters as Table 41.1, referring to [8]:

Using data in Table 41.1 and Eq. (41.1), we can draw single side band (SSB) phase noise power spectrum as shown in Fig. 41.2:

Table 41.1 The design of parameters of two radar emitters

Parameters	Divider R	Divider N	Signal frequency f_0	PD working frequency f_s
Emitter 1	1	200	10 MHz	10 MHz
Emitter 2	1	2,000	1 MHz	1 MHz
Parameters	R_1	R_2	PD optimization factor FOM	Voltage control sensitivity K_{VCO}
Emitter 1	750 Ω	2 kΩ	−220 dBc/Hz	50 MHz/V
Emitter 2	540 Ω	1.3 kΩ	−210 dBc/Hz	68.4 MHz/V
Parameters	C_1	C_2	C_3	Charge pump gain I_{cp}
Emitter 1	560 pF	10 nF	160 pF	3 mA
Emitter 2	6 nF	33 nF	37 pF	3 mA
Parameters	REF phase noise		VCO phase noise	
Emitter 1	−105 dBc/Hz @ 10 Hz		−50 dBc/Hz @ 10 Hz	
	−120 dBc/Hz @ 100 Hz		−83 dBc/Hz @ 100 Hz	
	−135 dBc/Hz @ 1 kHz		−108 dBc/Hz @ 1 kHz	
	−145 dBc/Hz @ 10 kHz		−126 dBc/Hz @ 10 kHz	
	−150 dBc/Hz @ 100 kHz		−142 dBc/Hz @ 100 kHz	
Emitter 2	−120 dBc/Hz @ 10 Hz		−8 dBc/Hz @ 10 Hz	
	−136 dBc/Hz @ 100 Hz		−38 dBc/Hz @ 100 Hz	
			−67 dBc/Hz @ 1 kHz	
	−146 dBc/Hz @ 1 kHz		−92 dBc/Hz @ 10 kHz	
	−151 dBc/Hz @ 10 kHz		−113 dBc/Hz @ 100 kHz	
	−152 dBc/Hz @ 100 kHz		−133 dBc/Hz @ 1 MHz	

Fig. 41.2 SSB phase noise power spectrum of emitter 1 and emitter 2



41.5.2 Adding Phase Noise

Using emitter 1 as example, we draw signal without and with unintentional modulation phase noise, as shown in Fig. 41.3. Phase noise is so weak that it's hard to find their differences. There is no significant effect on the signal.

41.5.3 Performing Time-Frequency Atomic Decomposition

We choose Gabor atomic set with 4 parameters $[s, u, \xi, \omega]$, and perform TFAD with emitter 1 signal and emitter 2 signal for 30 times. Still, we use emitter 1 as example. The original signal and reconstructed signal is shown in Fig. 41.4. After 30 times of decomposition, the RMS of residuals is only 0.0170, thus there is no significant difference between original signal and reconstructed signal.

After 30 times of decomposition, we get 30 sets of Gabor atom parameters. The distribution of these parameters changes as shown in Fig. 41.5. As of the first 12 atoms, the parameters from two emitters are all the same, but when it comes to the 13th atom and on, the parameters appear quite different. This proves that TFAD can extract features from emitter signals and distinguish different such features.

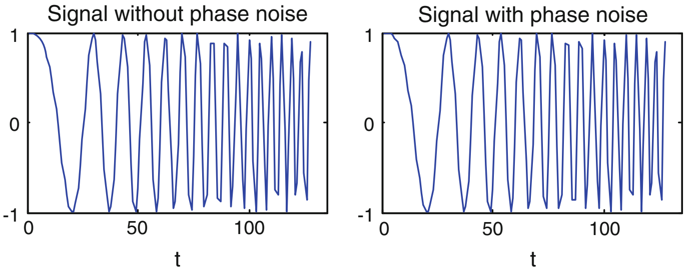


Fig. 41.3 Comparison on emitter 1 signal without and with unintentional modulation phase noise

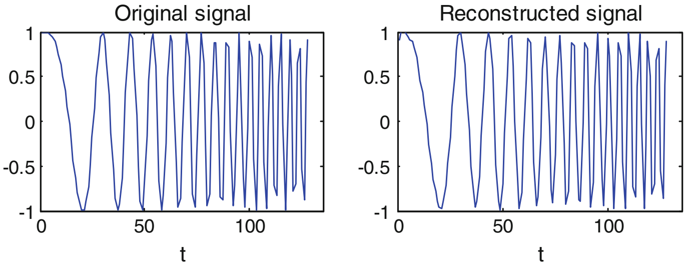


Fig. 41.4 Comparison on original signal and reconstructed signal for emitter 1

To get timing result, we use Monte-Carlo method to measure the time needed for each decomposition in average. We do the same thing with QGA applied, and here is the statistical results, as shown in Table 41.2:

Applying QGA algorithm do have some improvements on performance. While the four-parameter search space of Gabor atomic set is still not large enough for

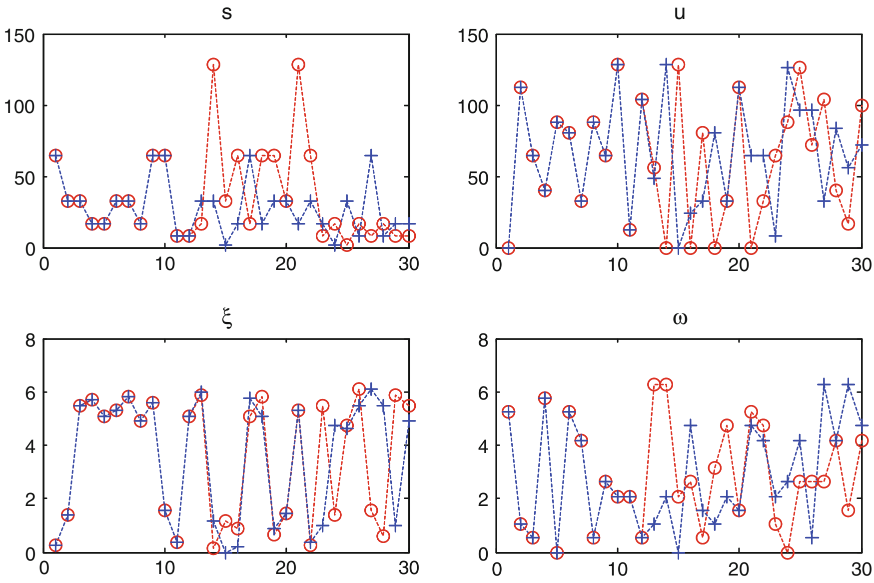


Fig. 41.5 Comparison on distribution of 30 Gabor atom parameters

Table 41.2 Comparison on time cost in each decomposition using MP and QGA algorithm

Algorithm	Emitter 1	Emitter 2
MP algorithm	1.3618 s	1.3613 s
QGA algorithm	1.1376 s	1.1381 s

QGA algorithm, it will work better and show a bigger advantage if we use Chirp atomic set with five parameters, which can do a better job decomposing radar signal and yet expands its search space into dozens of times.

41.6 Conclusion

Radar emitter unintentional modulation is an intrinsic property of radar transmitter, mainly in the form of phase noise. In this paper, we design specific parameters of two radar emitters, decompose radar emitter signal with phase noise using Gabor atomic set and then reconstruct it. The decomposed result, namely the parameters of Gabor atoms are treated as unintentional modulation features obtained. Such parameters are sufficient to represent unintentional modulation features, and also can be used to distinguish different emitters. The introduction of quantum genetic algorithm (QGA) solves the problem that MP algorithm is computationally intensive, and the amount of computation is reduced effectively. If we use Chirp atomic set with a larger search space, QGA will have better performance. This is set as the further work in the future.

References

1. Tian B (2010) Unintentional modulation of radar emitter signals. Doctoral thesis of Southwest Jiaotong University, Chengdu
2. Liu X (2011) Research on unintentional modulation analysis and recognition methodology of radar emitter. Doctoral thesis of National University of Defense Technology, Changsha
3. Wang L (2009) Radar emitter signal feature analysis based on time-frequency atom. Doctoral thesis of Southwest Jiaotong University, Chengdu
4. Davis G, Mallat S, Avellaneda M (1997) Adaptive greedy approximation. Constr Appro 13 (1):57–98
5. Cheng JX (2011) Radar emitter signal recognition based on time-frequency atom approach. Doctoral thesis of Southwest Jiaotong University, Chengdu
6. Mallat SG, Zhang ZF (1993) Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 41(12):3397–3415
7. Zhang GX, Li N, Jin WD et al (2004) A novel quantum genetic algorithm and its application. Acta Electronica Sinica 3:476–479
8. Tian B, Zhang GX, Long LJ et al (2010) Unintentional modulation feature extraction of radar emitter signals based on Gabor atoms. Foreign Electron Meas Technol 9:17–21

Chapter 42
An Improved Method on Associating Ascending and Descending Passes of Observations

Yuanxin Li and Jian Huang

Abstract Association of ascending and descending passes (A-D passes) in uncorrelated tracks (UCTs) is an efficient way to find out a "new" object and to generate its elaborate orbit sequence gradually. This paper proposes an improved method for A-D passes association problem. This method takes a three-level association structure; extensive filtering is performed in level 1 with time interval and orbital parameters a, i, Ω as decision variables; elaborate filtering is performed in level 2 with time offset and orbital parameters Ω, λ as decision variables; validating by orbit determination (OD) is performed in level 3 with OD residual as decision variable. The corresponding filtering thresholds depend mainly on sensor coverage in longitude, time interval of A-D passes, initial orbit precision, measurement error, model error, etc. Simulation analysis indicates that the method has satisfactory aspects in association accuracy, stability and efficiency.

Keywords Space object · UCTs · Association · Correlation

42.1 Introduction

In the process of detecting and cataloging space debris, the problem of how to establish the catalog orbit for a "newly incoming" space object rapidly stands on the first line to be resolved. For a type of sensor which can be used to searching and detecting space objects, such as the "dual-fence" radar [1], phased array radar, etc., when there is no a priori orbit for a newly detected space object, the uncorrelated tracks (UCTs) should be associated together for the purpose of transiting the new object into the catalogue stably.

This paper focuses on the association of an ascending pass and a descending pass observed by a single senor. First of all, compared to simply using ascending

---

Y. Li (✉) · J. Huang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: tryqtl1@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_42

(or descending) passes for orbit determination, the combined observations of ascending and descending passes (A-D passes) have the advantage of better space-time distribution, which is favorable for generating a satisfactory orbit to guide the following tracking. Secondly, it is realizable to obtain a combination of A-D passes in a short time, since there are at least two opportunities (with an ascending pass and a descending pass) for observing LEO objects when a ground station is continuously working for more than 12 h. In addition, the time interval between the adjacent ascending pass and descending pass is comparatively short (no more than 12 h for LEO objects), the orbit differences reflected by those two groups of observing data are not too large, which benefits data association.

Association of A-D passes is the principal step in using UCTs to find out a "new" object and to generate its elaborate orbit sequence gradually (see Fig. 42.1). At the beginning of catalog maintenance, or when space collision occurs, leading to a great increase of space debris, the amount of "new" objects found by a certain sensor is large, which means a higher requirement for accuracy and timeliness of A-D passes association.

A-D passes association is a kind of problem involved in observed orbital data association, for which some kind of methods based on orbital parameters association are mainly used currently, though the Naval Space Command uses a search and determine algorithm to construct orbits from uncorrelated observations [2, 3]. For example, an orbital element comparison method is introduced to use orbit parameters of inclination i, mean motion n, eccentricity e, etc. to associate two UCTs [4, 5]. Moreover, a covariance based association method is described based on the following assumptions: (a) the forces put on a space object are conservative; (b) the equations of motion can be linearized about the estimated trajectory. It is clear that this method is of certain limitation for those large area-mass ratio objects in lower orbit, for the reason that objects in this domain are seriously influenced by

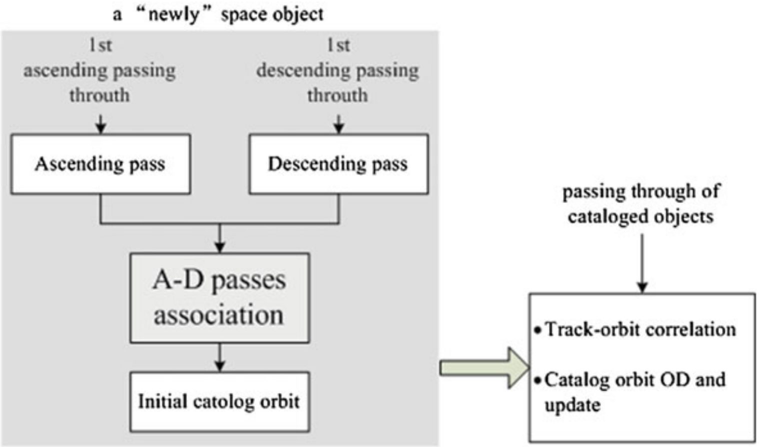


Fig. 42.1 Catalog process for a "newly" space object from its first discovering to daily update

atmospheric drag [6], which is non-conservative, and the deviation of the motion from the reference orbit predicted by an initial orbit will also be very large. It is also proposed that transforming inter-data association into association between observed data and orbit element, or association based on comparison of orbit elements [7]. While the first way needs to predict observation with all orbit elements in advance, leading to a low computational efficiency when the volume of space objects is great. The second way requires covariance value of orbit elements as input, and is sensitive to threshold values, a loose threshold leading to fuzzy association and a narrow one leading to miss association.

The purpose of this paper is to provide an improved method for the problem of A-D passes association with regard to a great amount of space objects.

42.2 Analysis of A-D Passes Association

A-D passes association is the process of associating a certain ascending (or descending) pass of observed track to a prior descending (or ascending) one, ensuring these two passes are obtained by observing a same space object, and then generating its initial catalog orbit with the associated two passes. It is a kind of decision making problem in nature, and a decision model should be created on the basis of correlation of A-D passes, some relevant decision variables and thresholds should be determined, to select the correct pass from a candidate set.

To associate an ascending pass and a descending pass distributed in a relatively short time interval (less than approximately 12 h), some common association characteristics should be integrated with its inherent specialties of A-D passes association. Here are:

Match in orbital parameters A-D passes correspond to the orbital motion characteristics of a same space object in different period of time. When the time interval is small, there exists a relationship between the two groups of orbit elements deriving from its ascending (or descending) pass (only taking into account the first order long-term perturbation):

$$\begin{cases} a(t_2) = a(t_1) \\ i(t_2) = i(t_1) \\ \Omega(t_2) = \Omega(t_1) + \Omega_1 \cdot (t_2 - t_1) \\ \lambda(t_2) = \lambda(t_1) + (n + \lambda_1) \cdot (t_2 - t_1) \end{cases} \tag{42.1}$$

where t_1 and t_2 are the epochs of orbit elements corresponding to two groups of observations respectively ($t_2 > t_1$), a denotes semi-major axis, i denotes inclination, Ω denotes right ascension of ascending node, λ denotes mean argument of latitude, n denotes mean motion; Ω_1 and λ_1 are the first order long terms of Ω and λ respectively (see Ref. [8] for detailed formulas).

Correlation of time interval The time interval of the adjacent A-D passes generated by a space object passing through certain detection area of coverage is relevant to specific orbital characteristics. If we define that φ represents the corresponding geocentric latitude of the coverage area forming by a certain detection sensor, u represents the argument of latitude when object passing through this latitude with a ascending (or descending) pass mode on the first time, then the value of the argument of latitude when object passing through this latitude with a descending (or ascending) pass mode on the nest time will be $\pi - u$, and the time interval between two pass can be broken down into two parts:

$$\Delta T_{A-D} = m \cdot T + \Delta u / n \tag{42.2}$$

If the first pass mode is ascending, then $u \in (\pi/2, \pi)$, $\Delta u = 2u - \pi$; otherwise $u \in (0, \pi/2)$, $\Delta u = 2u + \pi$. And for a single ground sensor, m may be value of $N - 1$, N or $N + 1$ generally, where $N = [\theta/\omega_e]$, in which "[]" denotes a round operation, ω_e denotes angular rate of earth rotation, and $\cos\theta = -(\sin 2\varphi + \cos 2u)/\cos^2\varphi$.

Self-accordance of residual Orbit determination (OD) is to seek the optimal fit (for the minimum of weighted residual sum of squares) between the calculated values and observed values. The resulted observation residual after OD processing is composed of two parts:

$$\Delta = y_o - y_c = (y_o - y_{true}) - (y_c - y_{true}) \tag{42.3}$$

in which the first part means inherent measurement error of detection sensor, and the second part means orbital dynamics model error. For the observations of one ascending pass and one descending pass, the corresponding OD residuals should be approximately within the scope determined by the specific measurement error and model error. If the two passes of observations are generated from two different space objects respectively, compared to the correct association, the OD residuals may be relatively larger.

42.3 Design of A-D Pass Association Method

42.3.1 Description of the Problem

Definition

1. $P = (f, t, \sigma, O)$ represents an individual pass of observation, in which f identifies pass mode (with 0 for ascending, 1 for descending), the combination of (t, σ) identifies the epoch and orbit elements of the initial orbit created by this single pass of observations, O represents a sequence of observations tagged with time in this pass.

2. $A = \{P_1, P_2, \ldots, P_n\}$ represents a set of un-associated passes, and P_0 denotes the pass to be associated.

A-D passes association is using a certain associating and matching method, to select an individual $P' \in A$, which satisfying:

- P_0 and P' are different in the value of pass mode;
- P_0 and P' originate from the same space object;
- An Initial catalog orbit can be derived by OD processing with P_0 and P' jointly.

Where restriction 1 is a precondition of A-D passes association, restriction 2 makes sure of the correctness of association results, and restriction 3 validates the association result and guarantees the availability of the result.

42.3.2 Association Model

According to the association characteristics of A-D passes, a three-level association model is applied to improve the accuracy and efficiency of candidate pass filtering and associating:

Level 1: extensive filtering Use the approximate time interval between ascending pass and descending pass, and some steady orbital parameters as decision variables; and select a candidate set I from a variety of un-associated passes, to primarily diminish filtering domain. The extensive filtering restrictions are set as below:

$$\begin{cases} f_0 \cdot f' = 0 \\ |t' - (t_0 - \Delta T)| \leq \Delta m_1' \cdot T_0 \\ |a' - a_0| \leq \Delta a' \\ |i' - i_0| \leq \Delta i' \\ |\Omega' - \Omega_0| \leq \Delta \Omega_1' \end{cases} \tag{42.4}$$

where subscript 0 denotes the pass to be associated, and superscript $'$ denotes a candidate pass; f denotes the pass mode, while t, a, i, Ω and T are the epoch time, semi-major axis, inclination, right ascension of the ascending node and period of mean motion of initial orbit respectively; ΔT is the approximate time interval of A-D passes calculated with P_0; $\Delta m_1{}'$, $\Delta a'$, $\Delta i'$, $\Delta\Omega_1{}'$ are extensive thresholds.

Level 2: elaborate filtering Use the time offset of A-D passes when they go through the same geocentric latitude, together with the orbital parameters Ω and λ as decision variables; and select a candidate set II from candidate set I. The elaborate filtering restrictions are set as below:

$$\begin{cases} \left|\Delta N_0' - round\left(\Delta N_0'\right)\right| \leq \Delta m_2' \\ \left|(\Omega_0 - \Omega') - 0.5 \cdot (t_0 - t') \cdot \left(\Omega_{10} + \Omega_1'\right)\right| \leq \Delta \Omega_2' \\ \left|(\lambda_0 - \lambda') - 0.5 \cdot (t_0 - t') \cdot \left(n_0 + \lambda_{10} + n' + \lambda_1'\right)\right| \leq \Delta \lambda' \end{cases} \tag{42.5}$$

where $\Delta N_0' = (t_0 - \Delta T - t')/T_0$ is the period difference of A-D passes; T_0 and n_0 are the period and angular rate of initial orbit mean motion of P_0; Ω_{10} and λ_{10} are the first order long terms of Ω and λ calculated with the initial orbit elements of P_0, while Ω_1' and λ_1' are calculated with the initial orbit elements of P'; $\Delta m_2'$, $\Delta\Omega_2'$, $\Delta\lambda'$ are elaborate thresholds.

Level 3: validating by OD processing Combine the pass to be associated P_0 and each pass in candidate set II to make an OD processing one by one; select the one of which the OD processing converges, the residual RMS is within a given threshold and is minimal as final association result. The corresponding filtering restrictions are set as below:

$$\begin{cases} \delta_0' \le \Delta V' \\ \delta_0' = \min(\delta_1', \delta_2', \ldots, \delta_k') \end{cases} \tag{42.6}$$

where δ_0' is the minimum residual RMS of OD processing with A-D passes, k is the number of pass in candidate set II; $\Delta V'$ is the given residual threshold.

42.3.3 Analysis of Filtering Thresholds

42.3.3.1 Thresholds of Extensive Filtering

The value of $\Delta m_1'$ is mainly determined by the longitude coverage of the detection sensor. If the longitude coverage is large, it will still leave one or more times of orbit period after eliminating ΔT from the time interval of adjacent A-D passes $\Delta t = t_0 - t'$. In order to reduce the probability of mistakenly deleting the correct result, the value of $\Delta m_1'$ can be set to:

$$\Delta m_1' = \frac{\Delta L}{\omega_e \cdot T_0} \tag{42.7}$$

The value of $\Delta a'$ and $\Delta i'$ depend primarily on the precision of initial orbital parameters a and i:

$$\begin{cases} \Delta a' = k_a \cdot \sqrt{2}\sigma_a \\ \Delta i' = k_i \cdot \sqrt{2}\sigma_i \end{cases} \tag{42.8}$$

where k_a and k_i are control factors, σ_a and σ_i are respectively the errors of initial orbital parameters a and i.

The value of $\Delta\Omega_1'$ depends on the precision of initial orbital parameter Ω, as well as the rate of orbit plane procession:

$$\Delta\Omega_1' = k_\Omega \cdot \sqrt{2}\sigma_\Omega + \dot{\Omega} \cdot \left(\Delta T + \Delta m_1' \cdot T_0\right) \tag{42.9}$$

where k_Ω is a control factor; the value of $\dot{\Omega}$is truncated to the first-order long term, and σ_Ω is the initial error of parameter Ω.

42.3.3.2 Thresholds of Elaborate Filtering

The value of $\Delta m_2'$ is mainly determined by the orbit period error, as well as the time interval of A-D passes, in which the orbit period error is caused by the semi-major axis error. The corresponding threshold is set to:

$$\Delta m_2' = k_{m_2} \cdot \frac{3}{2} \cdot \frac{\sigma_a}{a_0} \cdot \Delta N_0' \tag{42.10}$$

where k_{m2} is a control factor.

The value of $\Delta\Omega_2'$ depends primarily on the initial orbit accuracy of A-D passes:

$$\Delta\Omega_2' = k_\Omega \cdot \sqrt{2}\sigma_\Omega \tag{42.11}$$

The value of $\Delta\lambda'$ is relevant to the initial orbit parameter λ, as well as the time interval, of A-D passes:

$$\Delta\lambda' = k_\lambda \cdot \left(2\sigma_\lambda^2 + \left(\frac{3\pi \cdot \Delta N_0'}{2a_0}\right)^2 \cdot \sigma_a^2\right)^{1/2} \tag{42.12}$$

where k_λ is a control factor, and σ_λ is the initial error of parameter λ.

42.3.3.3 The Threshold of OD Validation

The value of $\Delta V'$ depends primarily on measurement error and model error. The measurement error is restricted by the accuracy of detection sensor; model error is related to the dynamics model used in OD processing, and space objects with large area-mass ration in low orbit is mainly influenced by model error of atmospheric drag.

$$\Delta V' = k_V \cdot \left(\sigma_R^2 + \left(\frac{3\pi}{2} \cdot \frac{\left(\Delta N_0'\right)^2}{4}\right)^2 \cdot \sigma_{da}^2\right)^{1/2} \tag{42.13}$$

where k_V is the control factor; σ_R is the trajectory error mapped by the measurement error, σ_{da} is the rate error of the semi-major axis caused by atmospheric drag model error (with unit of m/rev).

In order to reduce the probability of false filtering, when setting the values of filtering thresholds, it may be appropriate to enlarge the control factors (generally in the range of 5–8).

42.4 Simulation Analysis of Association Performance

A great deal of passes of simulated observations, which are supposedly "obtained" by a transformed "electronic fence" detection sensor for LEO space objects in large amount, are taken as input to analyze the performance of the three-level association method mentioned above.

Simulation conditions According to the space distribution of existed LEO objects, observations are simulated for those ones of which altitude are less than 1,000 km and totally 86,625 passes of 36,410 objects in 1 day are obtained. Detailed conditions are as below:

- The "Electronic fence" takes a dual-fence structure, with each fence inclining 10 degrees from zenith; when a space object passes through the dual-fence, at least 2 points of observations are created to form a valid pass.
- The simulated sensor locates near 40°N, and the longitude coverage is approximately 30° for objects of which the height $H = 1{,}000$ km.
- Observation types include range and angles (direction cosines), with values of corresponding measurement errors set in Table 42.1.

Settings of thresholds Under the above-mentioned simulation conditions, the initial orbits are calculated with each pass of observations; and the standard deviations of initial orbit parameter errors are counted, as shown in Table 42.2.

In accordance with the analysis in Sect. 42.3.3, three groups of different filtering thresholds are set to perform A-D passes association. See Table 42.3 for the values of thresholds.

Results of association The accuracy of association is analyzed first. The probability of A-D passes association correctness is defined as:

$$P_{correct} = 1 - P_{error} = 1 - \left(P_{fault1} + P_{fault2} + P_{miss}\right) \tag{42.14}$$

in which the mistakenly associated results are composed of 3 parts:

- virtual association, there does not actually exist a combination of A-D passes, while resulting one;

Table 42.1 Observation error settings of simulation

System errors		Random errors	
Range (m)	Angle (°)	Range (m)	Angle (°)
50.0	0.01	50.0	0.01

Table 42.2 Standard deviations of initial orbital parameter errors

σ_a (km)	σ_i (°)	σ_Ω (°)	σ_λ (°)
7.47	0.021	0.019	0.038

Table 42.3 Three groups of value settings of filtering thresholds. Notice that except $\Delta V'$, other thresholds shift to a degree of more than 20 % relatively

Group	$\Delta m_1'$	$\Delta a'$ (km)	$\Delta i'$ (°)	$\Delta\Omega_1'$(°)	$\Delta m_2'$ (°)	$\Delta\Omega_2'$ (°)	$\Delta\lambda'$ (°)	$\Delta V'$ (km)
1	2.0	48	0.16	8.0	0.16	0.16	12.0	4.5
2	2.5	60	0.20	10.0	0.20	0.20	15.0	4.5
3	3.0	72	0.24	12.0	0.24	0.24	18.0	4.5

- false association, there does actually exist a combination of A-D passes, and a false result is obtained for it, i.e. the two passes associated don't originate from the same space object;
- missing association, there does indeed exist a combination of A-D passes, but no result returns for it.

Simulated association results are shown in Table 42.4, where N_{all} is the number of actually existent combinations of A-D passer, N_{assoc} is the number of combinations obtained after association process, N_0 is the number of combinations which actual exist, and are correctly associated, N_1 is the number of virtual association, N_2 is the number of false association, N_3 is the number of missing association. It is show that with the jointly employment of a variety of filtering criterions, include time interval, orbital parameters, and OD residual, and so on, a high probability of more than 99 % of A-D passes association correctness can be obtained in this simulation. Moreover, even when some of the filtering thresholds shift to the degree of more than 20 % relatively, the probability of association correctness still varies quite small, which certificates a favorable stability of this method.

The following analysis is about association efficiency. An efficient association method excels in smallest computational spending to obtain the correct result and create an initial catalog orbit. For the A-D passes association problem, the workload of OD processing is of maximum, so it is expected the number of remaining

Table 42.4 Simulated association results for three groups of filtering thresholds settings

Group	N_{all}	N_{assoc}	N_0	N_1	N_2	N_3	$P_{correct}$ (%)
1	33,111	33,052	33,042	10	0	67	99.77
2	33,111	33,107	33,095	12	0	8	99.94
3	33,111	33,117	33,083	34	0	1	99.89

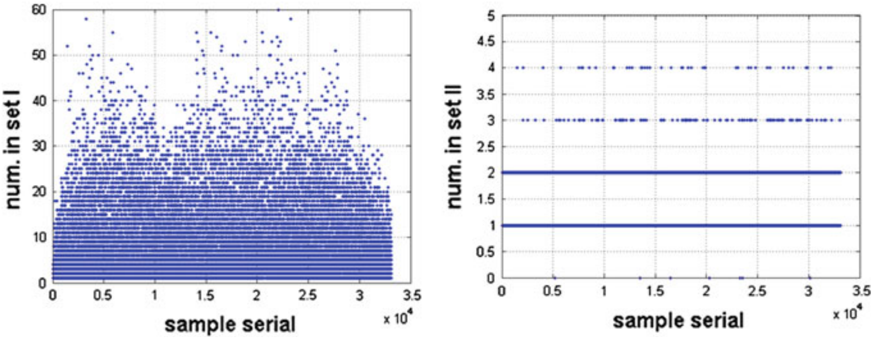


Fig. 42.2 Number distribution of candidate passes after extensive filtering and elaborate filtering (*left* candidate set I, *right* candidate set II)

candidate passes be a minimum after extensive filtering of level 1 and elaborate filtering of level 2. Simulation results show that the number of passes in candidate set I is less than 100 after filtering level 1, and about 90 % in all samples are less than 20 (as shown in the left side of Fig. 42.2). The number of passes in candidate set II is no more than 5 after filtering level 2, and about 99.5 % are less than 2 (see right part of Fig. 42.2). The total OD processing number is up to 38,456, with an average of less than 1.2 for each association.

42.5 Conclusions

Several association characteristics have been analyzed for the problem of A-D passes association. A three-level association model and the corresponding settings for filtering thresholds have been proposed. Simulation analysis has been performed to certificate the method proposed is of high accuracy, stability and efficiency for A-D passed association.

References

1. Huang J, Hu WD, Qi Xin et al (2012) Novel data association scheme for LEO space debris surveillance based on a double fence radar system. Adv Space Res 50:1451–1461
2. Coffey SL, Jenkinsi E, Neal HL et al (1996) Parallel processing of uncorrelated observations into satellite orbits. In: AAS/AIAA astrodynamics specialist conference
3. Coffey SL, Neal HL, Berry MM (2002) Uncorrelated observations processing at naval space command. J Guidance Control Dyn 25(4):676–684
4. Alfriend KT (1997) A dynamic algorithm for processing uncorrelated tracks, paper no. 97-607. In: AAS/AIAA astrodynamics conference, Sun Valley, ID
5. Alfriend KT Covariance based track association. http://aero.tamu.edu/sites/default/files/faculty/alfriend/s1.3%20AlfriendF.pdf
6. Montenbruck O, Gill E (2001) Satellite orbits: models, methods, and applications. Springer, Berlin
7. Wu LD (2011) Orbit and detection of satellite and space debris. China Science and Technology Press, Beijing
8. Liu L, Hu SJ, Wang X (2006) An introduction of astrodynamics. Nanjing University Press, Nanjing

Chapter 43
Analysis of False Alarm and Missing Alarm in Conjunction Assessment of Space Objects

Chaowei Ma and Xianzong Bai

Abstract The risks of collision between space objects are becoming more serious than ever. In order to ensure on-orbit safety, collision probability (Pc), which provides a quantitative measure of likelihood between space objects, is widely used to evaluate the risk of collision. But the conjunction assessment of space objects is substantially a discriminant analysis problem, and there is always a question about the reliability of Pc. In this paper, the probabilities of missing alarm and false alarm are studied based on the explicit expressions of Pc, the safety-region and danger-region of conjunction assessment are defined, and the definition, formula, and basic property of probabilities of missing alarm and false alarm are provided. This paper provides an idea for the reliability calculation of Pc, which is useful for us to directly know the reliability of Pc, balance the cost of taking action to maneuver the spacecraft or not, and make the final decision.

Keywords Space objects · Conjunction assessment · Collision probability · Missing alarm probability · False alarm probability

---

C. Ma (✉)
School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
e-mail: machaowei@bittt.cn

C. Ma
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China

X. Bai
College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_43

43.1 Introduction

In recent years, the risks of collision between space objects cause important attentions on spacecraft flight-safety all over the world [1, 2]. In order to ensure on-orbit safety, maneuver should be taken by operational spacecrafts to avoid dangerous conjunctions with space debris which are larger than 10 cm and can probably be tracked and cataloged.

Miss distance and collision probability (Pc) are widely used to evaluate the risk of collision. Collision probability is more often to use than miss distance now. There are many methods [3–23] used to calculate collision probability, in which state vector, state covariance, and combined size of two space objects are basically needed. Due to orbit tracking error, determination error and prediction error, it is inevitable for collision probability to be imprecise. In other words, the calculation value of collision probability is not the actual value. And what we suppose a dangerous conjunction event maybe not dangerous.

When two space objects getting closer, there are always two discriminant analysis problems. One is probability of missing alarm, which means the two space objects are actually collide but we suppose they are safety. The other is probability of false alarm, which means the two space objects are safety but we suppose they are dangerous to collide.

In this paper, we define the safety-region and danger-region of conjunction assessment by using explicit expression of Pc, and put forward the methods for calculating probabilities of missing alarm and false alarm. In this way, it is useful for us to directly know the reliability of Pc, balance the cost of taking action to maneuver the spacecraft or not, and make the final decision.

43.2 Safety-Region and Danger-Region

Conjunction risk assessment is essentially a discriminant problem. According to certain criteria, miss distance, collision probability, maximum probability of collision, and other conjunction risk metrics are used to determine whether the risk metrics in hazardous areas. For example, if miss distance is less than the distance threshold, or collision probability is greater than the probability threshold, the two space objects are in dangerous conjunction, and appropriate measures should be taken.

As a discrimination problem, there are always two types of judgement: missing alarm and false alarm. In a conjunction risk assessment, It is very important to control the probability of missing alarm (P_m) and the probability of false alarm (P_{fa}). The two parameters are controversial and need to be balanced. In general, the smaller P_m the greater P_{fa} is, and the greater P_m the smaller P_{fa} is.

For scenarios with conjunction risk analysis, calculation values of collision probability are not as same as actual values. In fact, the actual values of Pc are located in three-dimensional covariance ellipsoids centred by the calculation values with high probability at the time of closest approaches (TCA). For a specific conjunction event, the actual values of Pc are deterministic but unknown. The calculation values of Pc are around the actual values by some kind of probability distribution. Based on above assumption, safety-region and danger-region of conjunction assessment are determined as follows.

In a conjunction event, collision plane can be established by perpendicular to the relative velocity vector at the time of closest approach. Assuming the actual relative position vector is $(x_{\text{true}}, y_{\text{true}})$, the prediction relative position vector is $\left(x_{pred}, y_{pred}\right)$, whose uncertainties are represented by Gaussian distribution. Then,

$$x_{\text{pred}} \sim N\left(x_{\text{true}}, \sigma_x^2\right), \quad y_{\text{pred}} \sim N\left(y_{\text{true}}, \sigma_y^2\right) \tag{43.1}$$

Collision probability is calculated by the prediction relative position vector $\left(x_{\text{pred}}, y_{\text{pred}}\right)$, which can be explicit expressed as [19],

$$P_c = \exp\left[-\frac{1}{2}\left(\frac{x_{\text{pred}}^2}{\sigma_x^2} + \frac{y_{\text{pred}}^2}{\sigma_y^2}\right)\right]\left[1 - \exp\left(-\frac{r_{\text{A}}^2}{2\sigma_x\sigma_y}\right)\right] \tag{43.2}$$

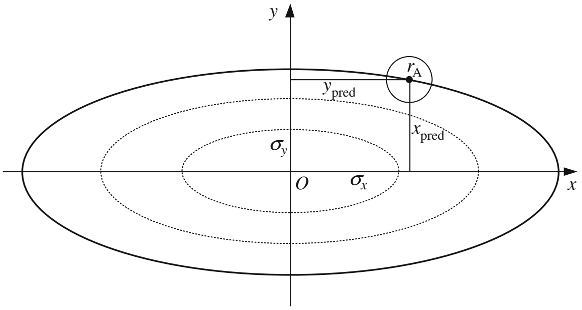
Since the prediction relative position vector $\left(x_{\text{pred}}, y_{\text{pred}}\right)$ is random, the collision probability Pc is also random. Equal values of collision probabilities form a ellipse (as shown in Fig. 43.1), means the collision probability ellipse, which can be expressed as,

$$\frac{x_{\text{pred}}^2}{\sigma_x^2} + \frac{y_{\text{pred}}^2}{\sigma_y^2} = C^2 \tag{43.3}$$

For those prediction relative position vectors on the collision probability ellipse, the values of collision probability are constant.

Let the collision probability threshold is P_T, safety-region S and danger-region D of conjunction assessment can be expressed as follows,

Fig. 43.1 Ellipse of equal collision probability



$$\begin{aligned} S&: \left\{ \left(x_{\text{pred}}, y_{\text{pred}}\right) \middle| P_c\left(x_{\text{pred}}, y_{\text{pred}}\right) < P_{\text{T}} \right\} \\ D&: \left\{ \left(x_{\text{pred}}, y_{\text{pred}}\right) \middle| P_c\left(x_{\text{pred}}, y_{\text{pred}}\right) \geq P_{\text{T}} \right\} \end{aligned} \tag{43.4}$$

From Eq. (43.2), safety-region S can be written as

$$P_c = \exp\left[-\frac{1}{2}\left(\frac{x_{\text{pred}}^2}{\sigma_x^2} + \frac{y_{\text{pred}}^2}{\sigma_y^2}\right)\right]\left[1 - \exp\left(-\frac{r_{\text{A}}^2}{2\sigma_x\sigma_y}\right)\right] < P_{\text{T}} \tag{43.5}$$

Let

$$C^2(P_{\text{T}}) = 2\left[\ln\left(1 - e^{-\frac{r_{\text{A}}^2}{2\sigma_x\sigma_y}}\right) - \ln P_{\text{T}}\right] \tag{43.6}$$

Then

$$\begin{aligned} S&: \left\{ \left(x_{\text{pred}}, y_{\text{pred}}\right) \middle| \frac{x_{\text{pred}}^2}{\sigma_x^2} + \frac{y_{\text{pred}}^2}{\sigma_y^2} > C^2(P_{\text{T}}) \right\} \\ D&: \left\{ \left(x_{\text{pred}}, y_{\text{pred}}\right) \middle| \frac{x_{\text{pred}}^2}{\sigma_x^2} + \frac{y_{\text{pred}}^2}{\sigma_y^2} \leq C^2(P_{\text{T}}) \right\} \end{aligned} \tag{43.7}$$

Equation (43.7) shows that safety-region S is outside the collision probability ellipse, and danger-region D is inside the collision probability ellipse. That means, the sizes of safety-region and danger-region are determined by the collision probability threshold P_{T}. When P_{T} is growing, the ellipse is becoming smaller and the safety-region S is becoming larger. When P_{T} is decreasing, the ellipse is becoming larger and the safety-region S is becoming smaller. As shown in Fig. 43.2.

Assume combined radii of two space objects is r_A, and relative position vector uncertainty in collision plane is $\left(\sigma_x, \sigma_y\right)$, and let $r_A = 20$, $\sigma_x = 1{,}000$, $\sigma_y = 100$ (distance unit). Figure 43.3 gives out the curve of the boundary of ellipse C changing with the collision probability threshold P_{T}. When P_{T} is greater than a

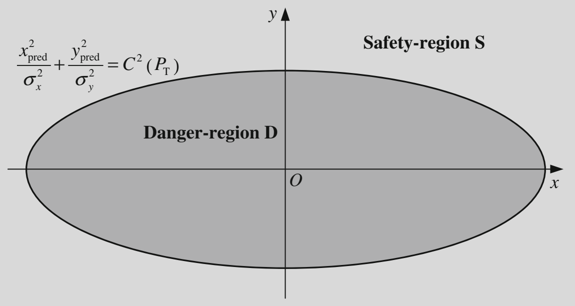


Fig. 43.2 Safety-region and danger-region of collision assessment

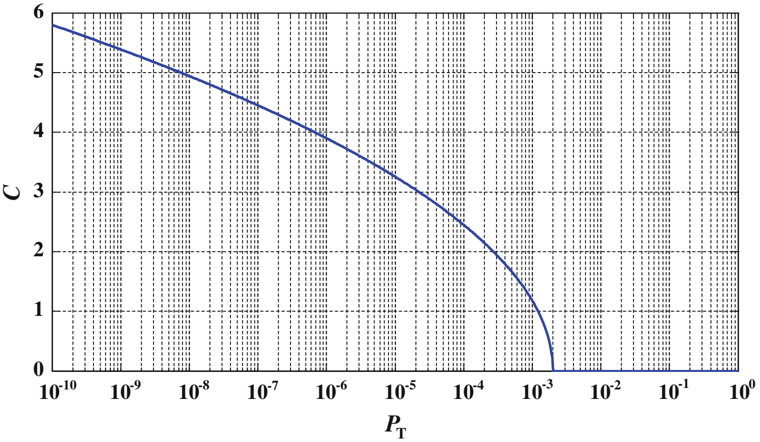


Fig. 43.3 Curve of size of threshold ellipse versus threshold value of Pc

certain threshold value, the boundary of the ellipse C is equal to 0, which indicates that the danger-region is degradation to the coordinate origin, and the entire two-dimensional collision plane is the safety-region. That means the collision probability threshold is greater than the maximum probability of collision. The collision probability in any case is not greater than the collision probability threshold. The conjunction risk assessment of the two space objects is always determined as safety.

43.3 Missing Alarm and False Alarm Model

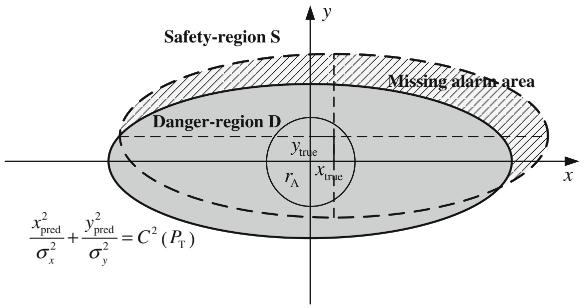
43.3.1 Probability of Missing Alarm

The probability of missing alarm P_m is defined as two space objects actually collide but supposed to be safety. That means the probability computed by predicted relative position vector $\left(x_{\text{pred}}, y_{\text{pred}}\right)$ is in the safety-region, but the actual relative position vector (x_{true}, y_{ture}) falls in the combined hardball of the two space objects, $x_{\text{true}}^2 + y_{\text{true}}^2 \le r_{\text{A}}^2$. Then, P_m can be described as

$$P_{\text{m}} = P\left(\left(x_{\text{pred}}, y_{\text{pred}}\right) \in S \middle| x_{\text{true}}^2 + y_{\text{true}}^2 \le r_{\text{A}}^2\right) \tag{43.8}$$

As shown in Fig. 43.4, the actual relative position vector $(x_{\text{true}}, y_{\text{true}})$ is less than combined radii of the two space objects. The predicted relative position vector $\left(x_{pred}, y_{pred}\right)$, which is around the actual relative position vector, satisfies Gaussian distribution. Notice that the dashed covariance ellipsoid in Fig. 43.4 is a 3σ ellipsoid, the area of missing alarm within safety-region is spreading the entire two-dimensional collision plane.

Fig. 43.4 Missing alarm region of collision assessment



From Eq. (43.1), The χ^2 distribution with two degrees of freedom and non-central argument δ can be obtained by $\left(x^2_{\text{pred}}/\sigma_x^2\right)+\left(y^2_{\text{pred}}/\sigma_y^2\right)$ [24]. That is

$$\frac{x^2_{\text{pred}}}{\sigma_x^2}+\frac{y^2_{\text{pred}}}{\sigma_y^2}\sim\chi^2(2,\delta),\quad \delta=\sqrt{\frac{x^2_{\text{true}}}{\sigma_x^2}+\frac{y^2_{\text{true}}}{\sigma_y^2}} \tag{43.9}$$

Let the probability density function (PDF) of the χ^2 distribution is $F(x|2,\delta)$. The missing alarm of probability can be expressed as

$$\begin{aligned} &P_{\text{m}}=1-F\left(C^2|2,\delta\right)\\ &\delta=\sqrt{\frac{x^2_{\text{true}}}{\sigma_x^2}+\frac{y^2_{\text{true}}}{\sigma_y^2}},\quad x^2_{\text{true}}+y^2_{\text{true}}\leq r_{\text{A}}^2 \end{aligned} \tag{43.10}$$

Equation (43.10) shows that missing alarm probability P_{m} is determined not only by the boundary of ellipse C, but also by the actual relative position vector (x_{true},y_{ture}). Let $r_A=20$, $\sigma_x=1{,}000$, $\sigma_y=100$ (distance unit), $P_{\text{T}}=10^{-4}$. The size of the boundary of ellipse C ($C=2.447$) can be yielded by Eq. (43.6). Figure 43.5 shows the curved surface of the missing alarm probability P_{m} is changing with the the actual relative position (x_{true},y_{ture}) in the collision plane.

As shown in Fig. 43.5, the maximum missing alarm probability can be yielded when the relative position $(x_{\text{true}},y_{\text{true}})$ is at the end of minor axis direction (y direction). In this case, the maximum missing alarm probability $P_{\text{m,max}}=0.0654$ when $(x_{\text{true}},y_{\text{true}})$ is at the two separate points of $(0,20)$ and $(0,-20)$.

$$P_{\text{m,max}}=1-F\left(C^2|2,\delta\right),\quad \delta=\frac{r_{\text{A}}}{\sigma_y} \tag{43.11}$$

Let the threshold value of collision probability $P_{\text{T}}=10^{-4}$. The boundary of ellipse $C=2.447$, which means the size of covariance ellipse is 2.447σ. Then the actual relative position is approximately at the coordinate origin, ($x_{\text{true}}=y_{\text{true}}=0$),

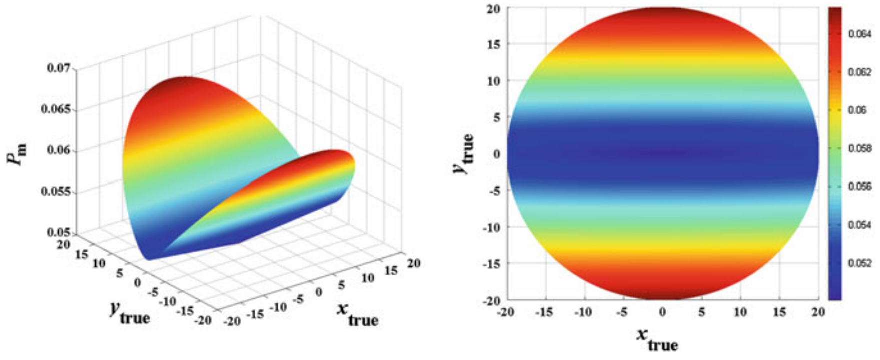
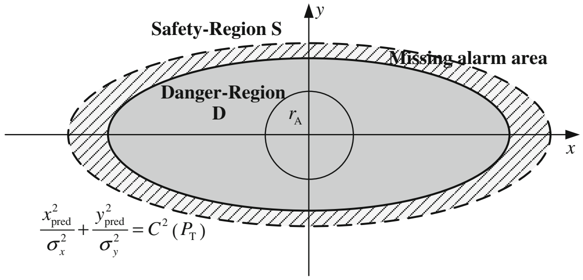


Fig. 43.5 Curved surface of $\mathrm{P_m}$ versus $(x_{\mathrm{true}}, y_{\mathrm{true}})$

Fig. 43.6 Missing alarm region when the true value is in the origin



and the area of missing alarm is a ring, as shown in Fig. 43.6. In this case, the missing alarm probability can be simplified as

$$P_{\mathrm{m}} = \frac{P_{\mathrm{T}}}{1 - \exp\left(-\frac{r_{\mathrm{A}}^2}{2\sigma_x\sigma_y}\right)} \tag{43.12}$$

The assumptions of $r_A = 20$, $\sigma_x = 1{,}000$, $\sigma_y = 100$, $P_{\mathrm{T}} = 10^{-4}$ are substituted into Eq. (43.12), then the missing alarm probability is equal to 0.05005 ($P_{\mathrm{m}} = 0.05005$). In many cases, Eq. (43.12) can be used to calculate the approximate value of missing alarm probability.

When the following inequality satisfies

$$P_m = \frac{P_{\mathrm{T}}}{1 - \exp\left(-\frac{r_{\mathrm{A}}^2}{2\sigma_x\sigma_y}\right)} > 1 \tag{43.13}$$

Then, the missing alarm probability is always equal to 1 ($P_{\mathrm{m}} = 1$).

$$P_{\mathrm{T}} > 1 - \exp\left(-\frac{r_{\mathrm{A}}^2}{2\sigma_x\sigma_y}\right) \tag{43.14}$$

$$\sigma_x\sigma_y > \frac{r_{\mathrm{A}}^2}{-2\ln(1 - P_{\mathrm{T}})} \tag{43.15}$$

Equations (43.14) and (43.15) means that it is always missing alarm when the threshold of collision probability P_{T} or the orbit propagation errors are overlarge. So, the maximum of collision probability is always smaller than the threshold of collision probability when Eqs. (43.14) and (43.15) are satisfied.

43.3.2 Probability of False Alarm

The probability of false alarm $\mathrm{P_{fa}}$ is defined as two space objects are safety but supposed be dangerous to collide. That means the probability computed by predicted relative position vector $(x_{\mathrm{pred}}, y_{\mathrm{pred}})$ is in the danger-region, but the actual relative position vector (x_{true}, y_{ture}) is out of the combined hardball of the two space objects, $x_{\mathrm{true}}^2 + y_{\mathrm{true}}^2 > r_{\mathrm{A}}^2$. Then, $\mathrm{P_{fa}}$ can be described as

$$P_{\mathrm{fa}} = P\left((x_{\mathrm{pred}}, y_{\mathrm{pred}}) \in D \middle| x_{\mathrm{true}}^2 + y_{\mathrm{true}}^2 > r_{\mathrm{A}}^2\right) \tag{43.16}$$

As shown in Fig. 43.7, the actual relative position vector $(x_{\mathrm{true}}, y_{\mathrm{true}})$ falls out of the combined radii of the two space objects. The predicted relative position vector (x_{pred}, y_{pred}), which is around the actual relative position vector, satisfies Gaussian distribution.

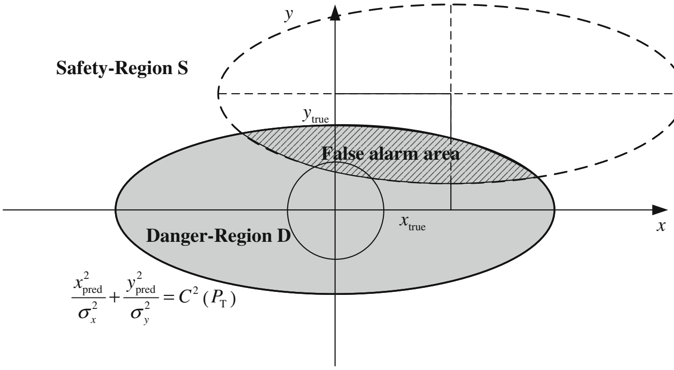


Fig. 43.7 False alarm region of collision assessment

Similar with the missing alarm probability, false alarm probability can also be described by the χ^2 distribution with two degrees of freedom and non-central argument δ [24].

$$\begin{aligned} P_{\text{fa}} &= F\left(C^2 \middle| 2, \delta\right) \\ \delta &= \sqrt{\frac{x_{\text{true}}^2}{\sigma_x^2} + \frac{y_{\text{true}}^2}{\sigma_y^2}}, \quad x_{\text{true}}^2 + y_{\text{true}}^2 > r_{\text{A}}^2 \end{aligned} \tag{43.17}$$

Equation (43.17) shows that false alarm probability P_{fa} is determined not only by the boundary of ellipse C, but also by the actual relative position vector (x_{true}, y_{ture}). Let $r_A = 20$, $\sigma_x = 1{,}000$, $\sigma_y = 100$ (distance unit), $P_{\text{T}} = 10^{-4}$. The size of the boundary of ellipse C ($C = 2.447$) can be yielded by Eq. (43.6). Figure 43.8 shows the curved surface of the false alarm probability P_{fa} is changing with the the actual relative position (x_{true}, y_{ture}) in the collision plane.

As shown in Fig. 43.8, the maximum false alarm probability can be yielded when the actual relative position $(x_{\text{true}}, y_{\text{true}})$ is at the end of major axis direction (y direction). In this case, the maximum false alarm probability $P_{\text{fa,max}} = 0.9484$ when $(x_{\text{true}}, y_{\text{true}})$ is at the two separate points of $(20, 0)$ and $(-20, 0)$.

$$P_{\text{fa,max}} = F\left(C^2 \middle| 2, \delta\right), \quad \delta = \frac{r_{\text{A}}}{\sigma_x} \tag{43.18}$$

Also can be seen from Fig. 43.8, the false alarm probability P_{fa} slowly becomes smaller, as the actual relative position in the collision plane changes along the major axis of the error ellipse (x direction), which is still equal to 0.77 when $(x_{\text{true}}, y_{\text{true}})$ is already at $(\pm 2{,}000, 0)$. But it quickly become smaller as $(x_{\text{true}}, y_{\text{true}})$ is changing along the minor axis (y direction), which is only equal to 0.0148 when $(x_{\text{true}}, y_{\text{true}})$ is at $(0, \pm 2{,}000)$.

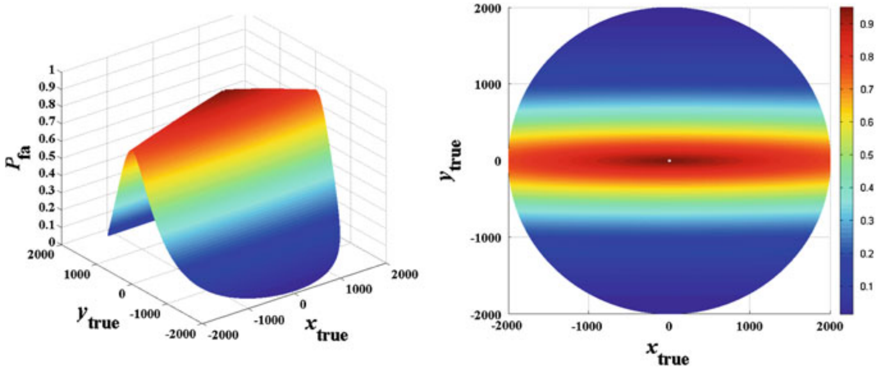


Fig. 43.8 Curve surface of P_{fa} versus $(x_{\text{true}}, y_{\text{true}})$

43.4 Conclusion

This paper gives out the concepts of probabilities of missing alarm and false alarm in a conjunction event between two space objects, which indicates that conjunction risk assessment of space objects is essentially a discriminant problem, and always exists two types of judgements: missing alarm and false alarm.

Based on the explicit expression of the collision probability, the safety-region and danger-region of conjunction assessment are defined, the methods for calculating probabilities of missing alarm and false alarm are put forward, and the basic property of probabilities of missing alarm and false alarm are provided.

Analysis of missing alarm and false alarm is still an exploratory research to solve conjunction assessment between space objects in the new issue, in which there are many issues that need to be resolved. In this paper, a preliminary analysis of this issue are given out, which provides ideas for further research.

References

1. Crowther R (2002) Space junk—protecting space for future generations. Science 296: 1241–1242
2. Liou JC, Johnson NL (2006) Risk in space from orbiting debris. Science 311:340–341
3. Foster JL, Estes HS (1992) A parametric analysis of orbital debris collision probability and maneuver rate for space vehicles. NASA/JSC-25898. NASA Johnson Space Flight Center, Houston
4. Patera RP (2001) General method for calculating satellite collision probability. J Guid Control Dyn 24(4):716–722
5. Patera RP (2002) Quick method to determine long-term orbital collision risk. In: AIAA 2002-1809. SatMax 2002—satellite performance workshop, Arlington
6. Patera RP (2004) Conventional form of the collision probability integral for arbitrary space vehicle shape. In: AIAA 2004-5218. AIAA/AAS astrodynamics specialist conference and exhibit, Providence, Rhode Island
7. Patera RP (2005) Calculating collision probability for arbitrary space-vehicle shapes via numerical quadrature. J Guid Control Dyn 28(6):1326–1328
8. Alfano S (2004) Accommodating rectangular objects in probability calculations. In: AIAA 2004-5217. AIAA/AAS astrodynamics specialist conference and exhibit, Providence, Rhode Island
9. Alfano S (2005) A numerical implementation of spherical object collision probability. J Astronaut Sci 53(1):103–109
10. Alfano S (2006) Satellite collision probability enhancements. J Guid Control Dyn 29 (3):512–588
11. Alfano S. Beta conjunction analysis tool (2007) AAS 07 393. In: AAS/AIAA space flight mechanics meeting, Sedona, Arizona
12. Chan FK (2008) Spacecraft collision probability. The Aerospace Press, El Segundo
13. Chan FK (1997) Collision probability analyses for Earth-orbiting satellites. Adv Astronaut Sci 96:1033–1048
14. Chan FK (2003) Improved analytical expressions for computing spacecraft collision probabilities. In: AAS 03-184. AAS/AIAA space flight mechanics meeting, Ponce, Puerto

15. Alfriend KT, Akella MR, Frisbee J et al (1999) Probability of collision error analysis. Space Debris 1:21–35
16. Akella MR, Alfriend KT (2000) The probability of collision between space objects. J Guid Control Dyn 23(5):769–772
17. Alfano S (2007) Review of conjunction probability methods for short-term encounters. In: AAS 07-148. AAS/AIAA space flight mechanics meeting, Sedona, Arizona
18. Bai XZ, Chen L (2008) Research on calculational method of collision probability between space objects. J Astronaut 29(4):1435–1442
19. Bai XZ, Chen L (2009) A rapid algorithm of space debris collision probability based on space compression and infinite series. Acta Mathematicae Applicatae Sinica 32(2):336–353
20. Bai XZ, Chen L (2009) Explicit expression and influencing factor analysis of collision probability between space objects. Chin J Space Sci 29(4):422–431
21. Bai XZ, Chen L, Tang GJ (2010) Explicit expression of collision probability for space objects in arbitrary-shape orbits. In: The 61st international astronautical congress, Prague, Czech
22. Bai XZ, Chen L, Tang GJ (2010) Space objects maximum collision probability analysis based on explicit expression. J Astronaut 31(3):880–887
23. Bai XZ, Chen L, Tang GJ (2013) Explicit expression of collision probability in terms of RSW and NTW components of relative position. Adv Space Res 52(6):1078–1096
24. Ma ZH (2000) Modern applied mathematics handbook: probability statistics and stochastic process. Tsinghua University Press, Beijing

Part IV
Communication and Information Systems

Chapter 44
The Design of Spacecraft TT&C Autonomous Management System

Xingyu Wang, Wenjie Liang, Huipeng Liang, Ailing Zhang and Ting Wang

Abstract Along with the quickening pace of mankind's space exploration, the number of space missions grows sharply and such space missions are becoming much harder; the task of space ground TT&C management is getting increasingly heavier and more complicated day by day. At the same time, space ground TT&C management system is lack of adequate autonomous technology support, which cannot meet the needs of space missions. In order to solve this problem, this paper adopts the method of increasing autonomous technology in space ground TT&C management system, applies Agent and MAS to space TT&C ground management system, designs the structure of space TT&C autonomous management system and states the working procedure of the system. The adoption of this system would greatly improve the efficiency and quality of ground TT&C system and solve the problem that ground TT&C system does not adapt to space missions. This system not only promotes the development of space TT&C technology, also meets the new requirements of space missions to TT&C system.

Keywords Agent · Spacecraft · MAS · TT&C · Autonomy

44.1 Introduction

With the development of the space exploration of mankind, space missions are becoming more and more complex. Meanwhile, the space technology develops faster and faster, especially autonomous operation technology of space is developing rapidly in recent years. In terms of its concept, significance or its great support to spatial system, space autonomous operation is unprecedented [1]. However, space missions are the jobs of integration of space and earth, the development of space autonomous operation technology cannot totally meet the all

---

X. Wang (✉) · W. Liang · H. Liang · A. Zhang · T. Wang
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: wxyyx_001@sina.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_44

needs of space mission; therefore, during the development of space autonomous operation technology, the research of ground autonomous operation technology can't be ignored. The perfect combination of ground and space autonomous operation can better solve the problem of slow emergency response speed, avoid serious accidents, improve working efficiency greatly and reduce the cost of construction and maintenance effectively.

In the 1990s, researchers in NASA has been working on the study of applying artificial intelligence technology to ground control system. The realization of automation of ground system software operation reduces the number of ground operators [2]. In recent years, domestic scholars have studied the satellite system based on the Agent, however, the research about the application of Agent technology to the system of aerospace TT&C ground operation (hereinafter referred to as the ground operation system) is less. Therefore, this paper analyzes the relevant concepts of Agent and MAS, designs the structure of ground autonomous operation system and provides reference for the realization of ground autonomous operation system from the perspective of the functional requirements of the ground operation system.

44.2 Basic Concept of the Agent

Since 1970s, the concept of Agent appeared in the field of artificial intelligence, along with the rapid development of computer and artificial intelligence technology, Agent technology has been widely used in many fields. Agent is often translated as "intelligent Agent" or "autonomous Agent" in artificial intelligence, which is an entity with active behaviors gained from the abstraction of decision in the process of operation or control task [3, 4]. Agent is a kind of method that describes the complex phenomena, studies the complex system and realizes the complex and adaptive calculation; it can use mathematical calculation or rules inference to complete certain operation task and information transmission and coordination through message mechanism, procedure object and the interaction of other Agents. In general, Agent has the basic features as Table 44.1 shows [5, 6].

Table 44.1 The basic characteristics of Agent

Characteristics	Meaning
Autonomy	The operation of Agent is not controlled directly by the people or other software, but to control his behavior and state
Social ability	Agent uses the agent communication language for information exchange
Response ability	Agent perception environment changes, and make the necessary response
Initiative	Agent has its own goals, can not only passively response to environmental changes, and can be proactive in organizational behavior

MAS is the multi-agent system, whose concept is derived from distributed artificial intelligence. MAS is a relatively loose consortium made up of multiple Agents, of which each Agent independently acts on itself and the surrounding environment and can coordinate with other Agents [7]. Multi-agent expresses the structure of the system by the communication, cooperation, mutual understanding, coordination, scheduling, management and control of different Agents [8]. Through task decomposition and task coordination, Multi-agent improves the overall ability of solving problems and overcomes the shortcomings that individual Agent is incomplete in knowledge and deals with information incorrectly. MAS can plan high level goal through coordination and communication and provide the behaviors of completing these goals on its own. In addition, MAS can decompose a large and complex problem into small and easy questions to simplify problems and achieve automation and intellectualization to certain degree, which just meet the needs of the ground TT&C management system.

44.3 System Design

44.3.1 Functional Requirements

In order to complete the task of spacecraft in orbit, ground TT&C management system should contain such functions as Table 44.2 shows.

Table 44.2 The ground control management system function

Function	Meaning
Telemetry data processing	Receiving and processing the spacecraft telemetry data, provide the basis for mastering the spacecraft state
Telecontrol data transmission	Processing and sending telecontrol command, control and management of spacecraft
Diagnosis and treatment of abnormity	Diagnosis of spacecraft telemetry data judgment, judgment and disposal of spacecraft anomalies
Orbit control and determination	According to requirements of the spacecraft orbit control and maintenance, and to measure and calculate the spacecraft orbital parameters
TT&C resources regulation	According to the spacecraft ground management needs, the distribution and adjustment of TT&C resources
Information management	When the spacecraft ground mass of information generated by the management of storage, easy retrieval of data analysis
Exception handling system	When the system autonomous operation is abnormal, it must be safe and effective implementation of the disposal

44.3.2 The System Structure

According to the basic concepts of Agent and MAS and the functional requests of the ground TT&C management system, the design about the structure model of space TT&C autonomous management system is shown in Fig. 44.1.

This structural model mainly consists of eight parts, including interaction Agent, communication Agent, control and decision Agent, information Agent, TT&C resources regulation Agent, different function Agent, knowledge base and database, etc.

The features of Agent in the system is fully displayed: each part of the system has its own individual functions and the ability of completing function on its own, what's more, system can complete the task of ground TT&C management. Every part, as one link of the system is to complete the duty of space TT&C management, and they have the abilities to solve the cooperative problems autonomously in the

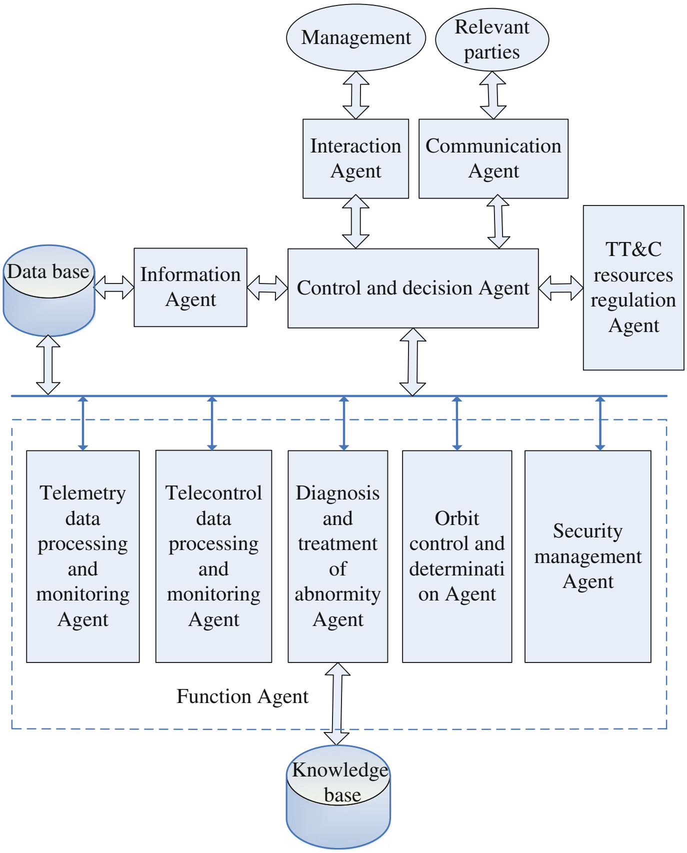


Fig. 44.1 The structure model of space TT&C autonomous management system

system during the management of the ground TT&C; each system has certain intelligence to handle the predictable abnormity of the system (such predictive ability can be increased) to assure the stability of system's operation.

44.3.2.1 Function Agent

It includes telemetry data processing and monitoring, telecontrol data processing and monitoring, diagnosis and treatment of abnormity, orbit control and determination and security management function Agent, and these agents complete the specific assignment through regulation and cooperation.

1. Telemetry data processing and monitoring Agent. It receives and processes the telemetry data of the spacecraft sent by communication Agent according to the track plan; monitors the accuracy of the data, and if the data has something wrong, it would give request to the control and decision Agent to send the telemetry data again and reprocess the new data. The processed telemetry data would be sent to communication Agent, to the other relevant Agents at the same time and finally is stored in the database.
2. Telecontrol data processing and monitoring Agent. It processes and handles telecontrol up-link data of satellite according to the track plan, does the comparisons and analysis based on certain rules to ensure the correctness, then gives the data to control and decision Agent, which sends it to TT&C equipment through communication Agent, and finally to satellite. At last, it judges the implementation of telecontrol data according to relevant telemetry data and gives feedback to control and decision Agent to decide whether follow-up plan of TT&C should be carried on as planned.
3. Diagnosis and treatment of abnormity Agent. It analyses and calculates the telemetry data, uses the parameter threshold, event rules and logic knowledge in the knowledge base to diagnose and to judge satellite running state on orbit. After it judges abnormity and sends the authorized disposal of abnormity to control and decision Agent directly to deal with it, TT&C resource transfer Agent should transfer TT&C equipment urgently and give feedback to control and decision Agent. After the TT&C plan of the disposal to meet the emergency of abnormity is formed, according to the predetermined proposed plan, telecontrol data processing and monitor Agent processes telecontroled up-link data of satellite and finally sends to satellite to finish the disposal of abnormity on its own. As for the unauthorized abnormity, it should be sent to satellite users and research department through communication Agent and when the plan for emergency handling is certain, such unauthorized abnormity can be returned to control and decision Agent to be dealt with.
4. Orbit control and determination Agent. Based on the definite result of the spacecraft orbit and input requirement of manager or relevant department, it analyses control requests of orbit. When orbit control is needed, it would send control request to control and decision Agent. Control and decision Agent

coordinates TT&C resource, arranges track plan and requires telecontrol data processing and monitor Agent to process telecontroled up-link data of satellite. Control and decision Agent arranges track plan to carry out orbit control order; diagnosis and treatment of abnormity Agent carries out real-time monitoring the condition of spacecraft during the control; after the control, ground station, according to the track plan, measures orbit and sends the data of orbital measurement. Orbit control and determination Agent receives such data to carry out obit determination, thus to evaluate the controlling effect of orbit.

5. Security management Agent. When the problem or abnormity that system cannot solve appears, the system has handling strategies. Security management Agent would provide solutions.

44.3.2.2 Control and Decision Agent

Control and decision Agent is the core of the whole system, assumes responsibility for duty management, coordinates the operation of other different Agent and monitors their work progress. According to the designed management requirements of the system, the demands of managers and relevant departments, control and decision Agent uses the rules of knowledge base to do task planning and coordination, chooses the suitable Agent to shoulder duty, gives the performance of doing task back to interaction Agent and communication Agent and finally gives feedback to managers and relevant department.

44.3.2.3 Interaction Agent

Interaction Agent is the window of man-machine communication, which provides interface of monitoring and managing TT&C equipment and central system to managers. The main functions of interaction Agent includes input of different management requirements, display function of managing task administration and receiving interaction order of the users to send to control and decision Agent.

44.3.2.4 Communication Agent

Communication Agent is the channel of exchanging data between system and external relevant department. Control and decision Agent gets different TT&C data needed by exterior relevant department from data base according to the rules and provides the result data to relevant department by communication Agent. Communication Agent sends the information transmitted by relevant department to control and decision Agent to decide or to data base to be stored.

44.3.2.5 Information Agent

Information Agent is mainly in charge of classification management and analysis and utilization of relevant information of system. It classifies and sorts out a lot of information given by system itself and relevant departments and then sends the information to data base to be stored. Besides, information Agent can search the needed information from data base to use; it can analyze and process the data from searched information and present the feedback to managers and diagnosis and treatment of abnormity Agent. Therefore, diagnosis and treatment of abnormity Agent make full use of the result of analysis to revise diagnosis and judgement, when managers make sure the revised contents, then they can be sent to knowledge base.

44.3.2.6 TT&C Resources Regulation Agent

TT&C resources regulation Agent is mainly to regulate and control TT&C resource adopted by the system and arrange use plan of equipment. Based on the relevant department's TT&C requirements sent by control and decision Agent, available condition of the ground station and predetermined rules of the knowledge base in system and the track prediction of ground station, TT&C resources regulation Agent arranges use plan of equipment.

44.3.2.7 Knowledge Base

Diagnostic threshold used by abnormal judgement, rules, logic knowledge and processing rules used by control and decision Agent while analyzing decision are stored in knowledge base. Relevant Agent uses knowledge and rules in the base to judge and decide.

44.3.2.8 Data Base

Data base contains various data produced by system including telemetry data, telecontrol data, orbit measurement data, task information and calculation result, etc. Relevant Agent uses data in the base to search information, inquire data and analyze tendency.

44.4 Overview of the System Work

The design about autonomous Management system of space TT&C reflects the features of Agent, which determines the autonomy, intelligence and cooperativity of the system in the course of working. The chart of the system working is given

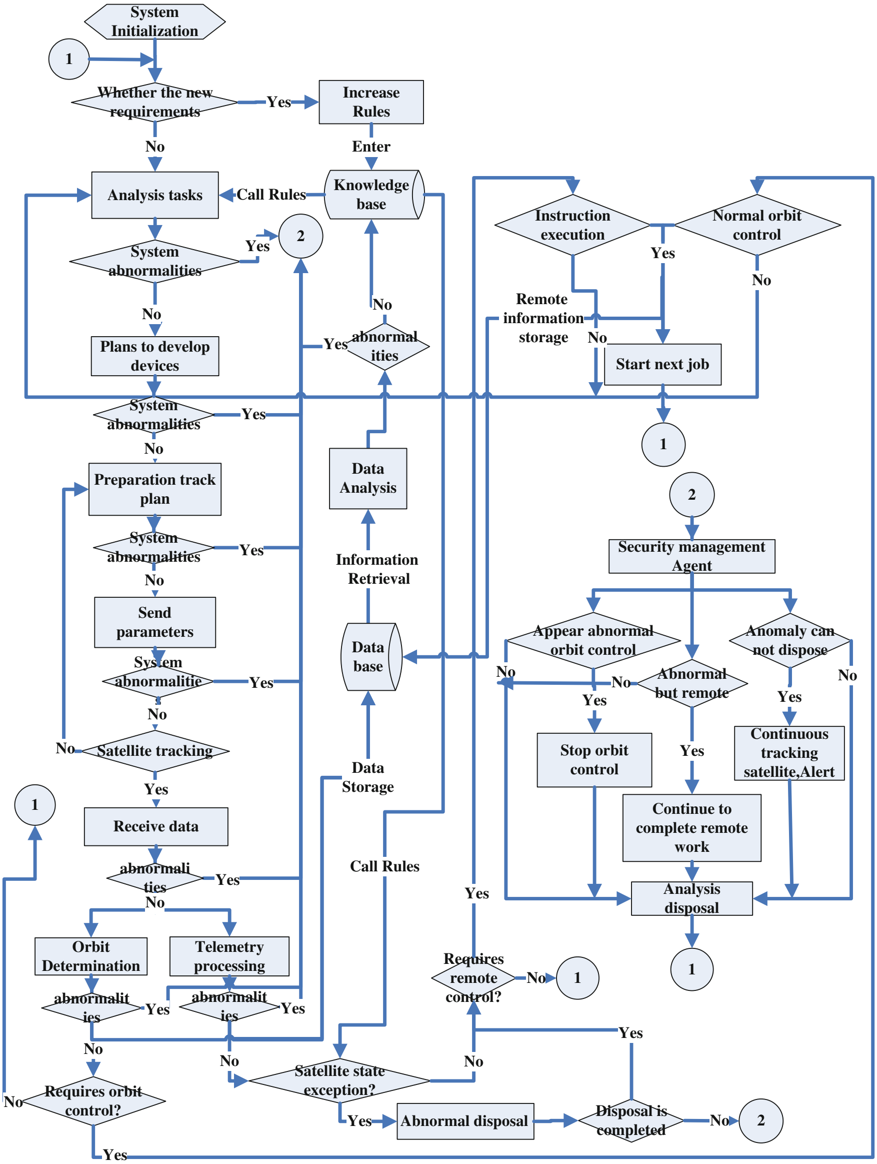


Fig. 44.2 The flowchart of space TT&C autonomous management system

below as Fig. 44.2 shows. In the chart, ① means the data interface after executing a task normally; ② suggests the data interface, which means that the system cannot handle the problems or the system is abnormal.

44.4.1 Decomposition and Allocation of Tasks

After the system initializes, in the case of receiving no new requirements from managers and relevant department, control and decision Agent decompose tasks as planed rules, distribute tasks to different Agent and coordinate the working connection and procedure between Agents. If new requirements from managers and relevant department come, control and decision Agent would decompose management task and distribute that to relevant Agent to implement based on the new management rules.

44.4.2 Spacecraft Tracking Control Equipment

According to requirements of control and decision Agent, TT&C resources regulation Agent considers the usages of ground station to form usage plan of equipment, meanwhile, control and decision Agent forms track plan. Control and decision Agent sends the control parameter of the ground station to communication Agent on the basis of track plan and control TT&C equipment to track spacecraft.

44.4.3 Telemetry Data Processing

TT&C equipment receives telemetry data and sends back to system by communication Agent, and control and decision Agent sends telemetry data to telemetry data processing and monitor Agent to process data and monitor task, then the processed result would be sent to information Agent to sort and analyze, next, the result would be sent to data base to be stored. What's more, the result would also be sent to telecontrol data processing and monitor Agent, diagnosis and treatment of abnormity Agent, Orbit control and determination Agent and optimizing management function Agent.

44.4.4 Orbit Control and Determination

TT&C equipment receives exterior tracking and measurement data of satellite and sends back to system by communication Agent, and control and decision Agent would transmit exterior tracking and measurement data to orbit control and determination Agent to determine the orbit; system calculate orbital element on its own and put it in storage after sending it to information Agent, and orbit control and determination Agent would do the orbit control analysis based on the new orbit to make sure the next opportunity of controlling satellite. If control is needed, orbit control and determination Agent would send control requirement to control and decision Agent, accordingly, control and decision Agent would apply for the usage plan of equipment to TT&C resources regulation Agent, give the requirement to telecontrol data processing and monitor Agent to create telecontrol up-link data of satellite. And based on the produced telecontrol order, control and decision Agent would begin to control according to track plan; during the control, diagnosis and treatment of abnormity Agent would monitor satellite.

44.4.5 Send Remote Data

Based on the content stated in the track plan, telecontrol data processing and monitor Agent processes telecontrol up-link data of satellite; after processing, it sends the processed data to control and decision Agent, then to communication Agent into TT&C equipment, finally to satellite. Control and decision Agent makes sure about the event of uplink according to the returned telemetry data.

44.4.6 Abnormality Diagnosis and Management

During the normal track, the result processed by telemetry data processing and monitoring Agent is transmitted to diagnosis and treatment of abnormity Agent directly, which utilizes diagnosis threshold, rules and logic knowledge in the knowledge base to judge the telemetry parameter and the state of satellite. If abnormity emerges, diagnosis and treatment of abnormity Agent would apply to control and decision Agent for handling it. As for the abnormity which is permitted to dispose, control and decision Agent invokes relevant agent to create telecontrol up-link data of satellite and invokes telecontrol equipment to implement the disposal of abnormity. As for the abnormity which is not permitted to dispose, this should be sent to the users and research department of satellite through communication Agent by control and decision Agent. After coordination and communication, managers add the negotiated plan of disposal to the scheme of disposal of abnormity in the management rules in control and decision Agent, next, control and decision Agent invokes relevant Agent to create telecontrol order, invoke telecontrol equipment and send telecontrol order to implement the disposal of abnormity.

44.4.7 Exception Handling

When system operates normally, security management Agent would not be involved in controlling; only when the regulations in the system fail to solve the practical problems or the system is abnormal, the system would turn to security management Agent to reduce the influence on the spacecraft management. Security Agent can preinstall multi predictable subfunction to deal with abnormity of system. For example, if system is abnormal, while the function of telecontrol order works, then the abnormity being disposed in the track plan, orbit control and other telecontrol order should be assured to work normally; if system is abnormal, meanwhile, telecontrol order cannot be given, then the system should stop all remote-control equipment and the task of telecontrol order, and the sound and light alarm should be given to managers to wait for their analysis and disposal; if system cannot solve practical problems during the time when spacecraft is abnormal, then control and decision Agent would invoke TT&C equipment to track spacecraft, report the abnormal situation to users and research department and give sound and light alarm to managers in order to warn them to dispose in time; if system cannot solve practical problems during the time when orbit is controlled, then control and decision Agent would maintain TT&C equipment and track spacecraft, besides, orbit control must be stopped immediately and sound and light alarm should be given to managers warning them to dispose in time.

44.5 Conclusion

This paper analyzes the advantages of applying Agent to space ground operation, designs the structure of autonomous management system of space TT&C and the functions of different Agent in system and researches the procedure of cooperation of the system from the point of the basic concepts and features of Agent and MAS. This idea breaks a new direction for the autonomy of the space TT&C ground operation, meets the requirements between the autonomous running of the space TT&C ground operation and management, complies with the tendency of the development of space technology and the integration of space and earth and provides reference for the realization of this system in the future. The follow-up would further study on the realization of engineering issue of this system.

References

1. Jing ZL, Yuan JP et al (2011) Measurement and control of spacecraft autonomous operation. China Aerospace Press, Beijing
2. Yu PJ (2012) Autonomous intelligent systems and their applications in NASA spacecraft operations and detection systems. Defense Industry Press, Beijing
3. Ni JJ (2011) Theory and application of complex systems modeling and control of multi-agent. Electronic Industry Press, Beijing
4. Xu YZ, Huang L (2009) Pumping station automation system based on integrated multi-agent. Adv Sci Technol Water Resour 29(3):64–67
5. Wooldr1dgel J, Jennings NR (1995) Intelligent agents: theory and practice. Knowl Eng Rev 10 (2):115–152

6. Dimltris NC (1998) Agent technology handbook. McGraw-Hill, New York
7. Wu YH, Cao XB et al (2007) Research on multi-agent architecture is based on formation flying satellites. J Syst Simul 19(18):4284–4287
8. Zou GP, Huang Z, Zou Q (2006) Highway monitoring system of multi-agent-based technology research. Commun Stand 9:64–67

Chapter 45
Web-Based Realization of Integrated Monitoring System to TT&C Equipments

Qi Dang, Dongxing Wu, Daming Bai and Guomin Li

Abstract With the increasing of TT&C equipments, C/S architecture is convenient for remote monitoring and control, but it is unsuitable for centralized state collection. This paper proposes a Web-based architecture of Integrated monitoring system for TT&C equipments which adopts the centralized monitoring protocol parsing, dynamic pages creation, data interaction based on HTTP protocol and other technologies, and designs an prototype software based on B/S. The result shows that this design can satisfy the requirement of Integrated monitoring on the current instruments and the requirement of upcoming expansion, easily realizes the design on demand, avoid the repetitive development of the software at lower level and improve work efficiency of the system.

Keywords TT&C equipments · Integrated monitoring · B/S · Web

45.1 Introduction

With the continuous development and progress of China's manned spaceflight project and lunar exploration project, a fully functional space TT&C network has been established by China [1]. In these TT&C equipments, client/server (C/S) architecture is often adopted for monitoring and controlling system [2], and centralized monitoring of equipments is primarily achieved through centralized remote monitoring client. Thus generally, if a new set of equipment is added, a corresponding remote client computer will be added to achieve the status of monitoring and control of the equipments. The centralized monitoring equipment was quickly deployed, however, some inconvenience has been introduced into the monitoring and decision-making process. Therefore, this study was designed to use Web

---

Q. Dang (✉) · D. Wu · D. Bai · G. Li
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: dangqifuping@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_45

services to provide integrated operational status of TT&C equipments to users, without changing the existing monitoring system, in order to improve the efficiency of equipment monitoring, while offering a more good user experience.

45.2 Monitoring Situation of TT&C Equipments

Definition 1 Sm_i-Subsystem monitoring, also known as primary monitoring, measured and controlled by various sub-systems, while controlled by secondary monitoring equipment. For example, subsystems such as baseband, antenna servo and feed, recording, automated testing and data transmission of unified S-band (USB) equipments, can be operated and managed not only via subsystem panel or operation console, but also be monitored and controlled by monitoring subsystem.

Definition 2 Em_i-Equipment monitoring, also known as secondary monitoring, measured and controlled locally by its own monitoring computer, meanwhile being monitored by remote control system.

Definition 3 Rm_i-Remote monitoring and Control, also known as the third level monitoring, monitored and controlled by TT&C network management center. Possibly there are several centers are monitoring the same equipment, but there is only one center able to control the equipment at the same time.

Currently, C/S architecture is often used in monitoring and control systems. Monitoring server is mainly responsible for collecting, reporting monitoring information, and executing commands sent from the client, monitoring and control will be implemented by the client via a good man-machine interface. Remote monitoring and control is realized by the use of information exchange between remote control client of network management center and TT&C equipments.

In this model, subsystem is responsible to achieve primary monitoring, local monitoring client and server responsible for secondary monitoring, and remote monitoring client and for monitoring, remote monitoring client and monitoring servers work together to achieve remote monitoring. For any set of monitoring and control equipment, Sm_i and Em_i are uniquely determined while monitoring network resource management center has a corresponding Rm_i. While the numbers of Sm_i and Em_i increases, the number of Rm_i will also continue to increase. The C/S architecture of the monitoring model to TT&C equipments is shown in Fig. 45.1.

Although there are a lot of advantages in the C/S architecture to meet the requirement of remote control, these two outstanding problems occurs with the increasing of equipment:

1. It is difficult to conduct integrated monitoring with continuous increasing numbers of remote monitoring computer in the monitoring network management center.
2. There are great environmental differences for various remote monitoring and control client, and the management and deployment of software is limited [3].

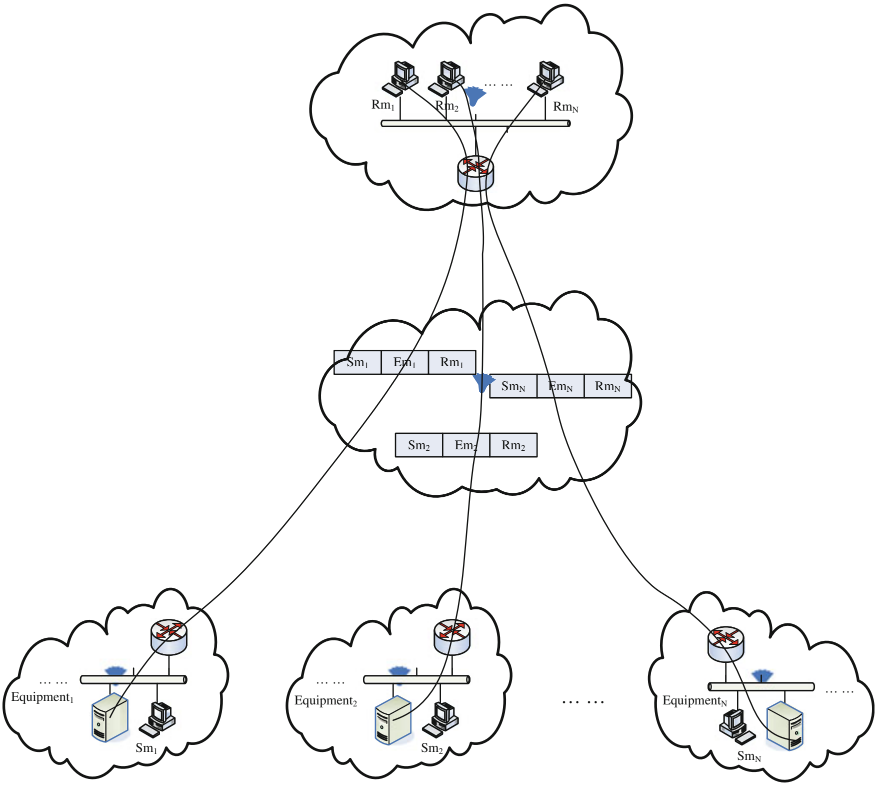


Fig. 45.1 TT&C equipments monitoring and control model

45.3 Integrated Monitoring Architecture

Integrated monitoring equipments architecture presented in this paper come into being on the basis of the current C/S monitoring mode by adding Protocol parsing layer and the presentation layer, and its hierarchical model is shown in Fig. 45.2.

The existing monitoring systems (including monitoring servers, local clients, remote clients) act as a equipment monitoring layer, and layer Protocol parsing has been added on the previous basis, then the equipment and operating parameters are extracted, unified and written into the database. Comprehensive presentation layer is responsible for Web-based information publishing [4]. This model is relatively traditional, for real-time monitoring of the C/S architecture has greater flexibility and openness with its convenience to deploy, to expand, to maintain and manage [5]. The system is developed in ASP.NET development environment, using static SQL Server database as the data source of basic information, information-based situation IP protocol network packet reception and transmission subsystems as real-time information and data sources, to achieve dynamic display control network integrated operational status, as shown in Fig. 45.3.

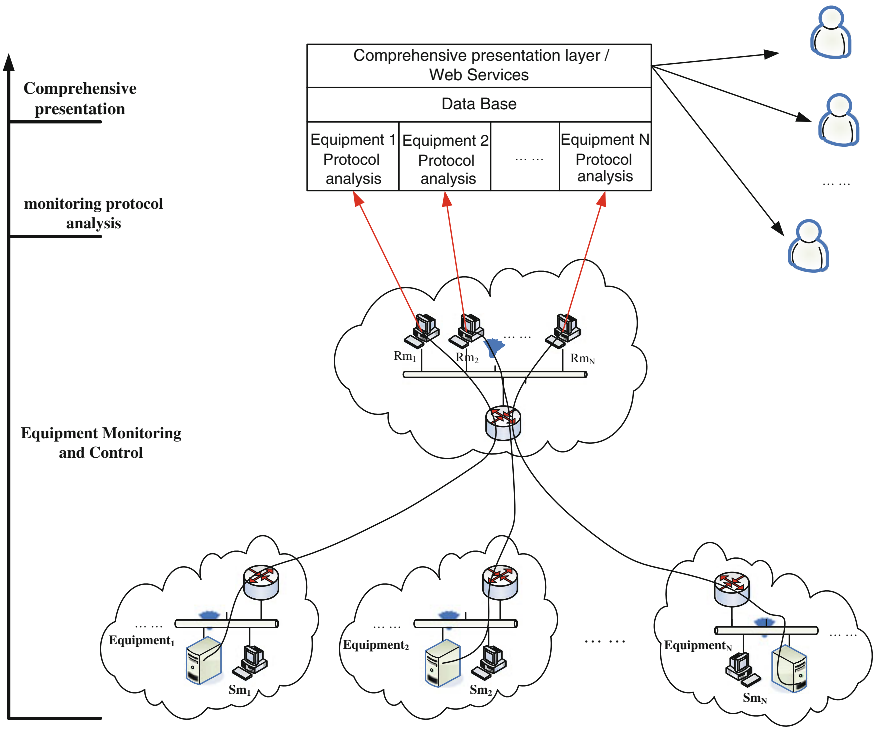


Fig. 45.2 Hierarchical model of integrated monitoring architecture

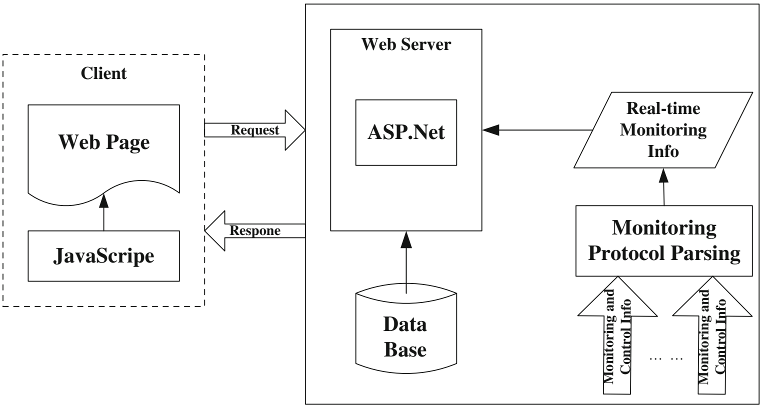


Fig. 45.3 Web-based running architecture

In order to meet all kinds of duties, management and operational needs of the user's browser, the system must meet the requirements of intuitive display, interactive operation, excellent running performance.

45.4 Realization of Technology

45.4.1 Framework and Strategic of Protocol Parsing

Protocol parsing is responsible for collecting equipment status information sent from each equipment monitoring server, and processing the monitoring parameters to extract and analysis the comprehensive operational status, fault information equipments and so on. To facilitate system expansion, the Protocol parsing capabilities are designed for each equipment individually replaceable components. Taking into account the data access redundancy, an independent monitoring data cache is established for each equipment by the use of standardized access interface.

In accordance with the function, protocol analysis is divided into the following sections, shown in Fig. 45.4.

1. Main control, to complete the various software modules for system management and scheduling, implementation of the interface between the operators, the process of creating and destroying various types of sub-windows; displaying monitoring information source and monitoring the status of the parsed information.
2. Network Communications, to create transmitting and receiving threads according to external configuration of the network, to receive monitoring information from the equipment monitoring network, to convert the information into the internal frame and to forward Protocol parsing object; data processing will be required to be sent out to meet the specified requirements in the agreement data frames sent out over the network.
3. Log management, to fully manage all the logs generated from the system, including the log display, record storage and browsing.

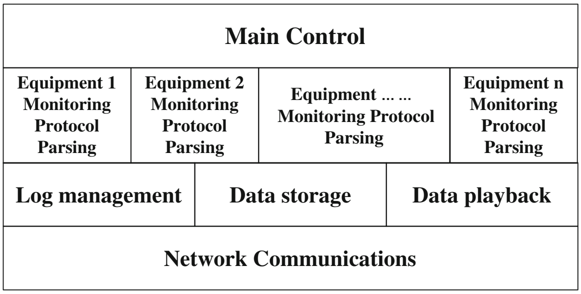


Fig. 45.4 Monitoring protocol parsing function diagram

4. Data storage, to store the real-time monitoring data received by the direction and the source.
5. Protocol parsing, to analyze remote monitoring information for each receiver after pretreatment control equipment, call the corresponding data parsing components.
6. Data playback, to replay the corresponding raw equipment files stored, according to the original transmission format for playback at a certain rate.

45.4.2 Ajax-Based Web Application Design

Firstly, Web-based Integrated monitoring program to TT&C equipments needs to be resolved is the real-time problem. The parameters of TT&C equipments change rapidly, data exchange quite a great deal, and it requires high efficiency of information exchange and the state of the real-time changes; Secondly, there are many types of TT&C equipments, the parameters of each equipment have various corresponding monitoring interface, the equipment monitoring interface needs to be dynamically created.

Asynchronous JavaScript and XML (Asynchronous JavaScript and XML, Ajax) is a way to create interactive web application development technology. Different from traditional Web development, Ajax use XMLHttpRequest object in the manner of communication with the server asynchronously, after obtaining the required data access file object model (Document Object Model, DOM) via JavaScript interface to update some elements of the page [6]. Relative to traditional Web applications, it has the following features:

1. Page rendering separated from data.
2. Reducing the response server redundancy, reducing the burden on the server.
3. Reducing the interactive data between server and client, reducing the network traffic load [7].

45.4.2.1 Real-Time Monitoring Data Interaction

Real-time data exchange between server and client can make use of the asynchronous data access mechanism of Ajax. Completed by the end of the monitoring service Protocol parsing, the analytical results will be sent into the cache server, the client queries sent by regular asynchronous request monitoring data acquisition equipment, and access the DOM interface to update the equipment display area by the use of JavaScript. Real-time monitoring interactive data flow diagram is shown in Fig. 45.5.

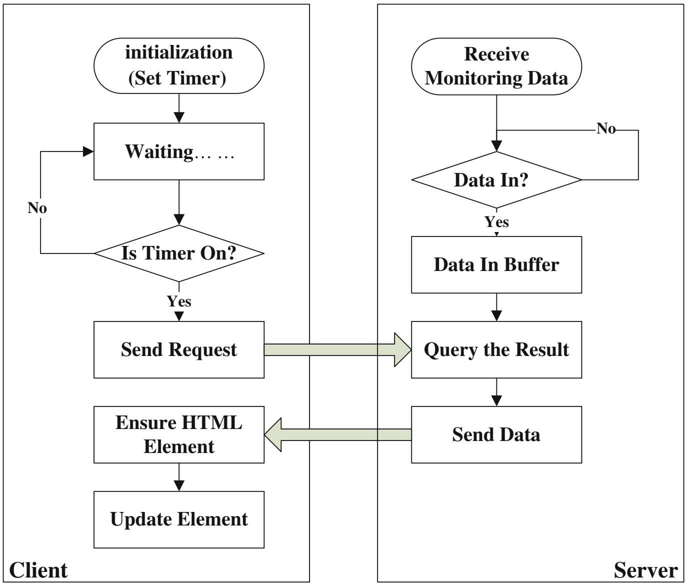


Fig. 45.5 Real-time monitoring interactive data flow diagram

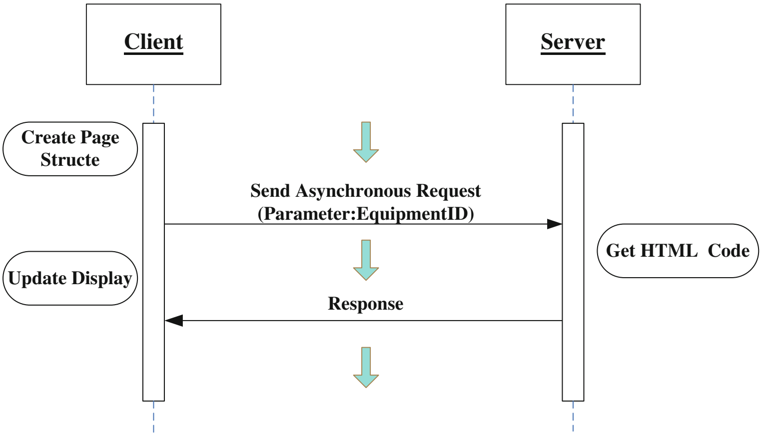


Fig. 45.6 Create web monitoring pages flow chart

45.4.2.2 Dynamic Creation of Web Monitoring Pages

At page initialization phase, the framework of monitoring page layout will be firstly created and the area will be allocated for the equipment to be monitored, then the client will send an asynchronous request to the server with the equipment identification parameters from the client, meanwhile change the corresponding region refresh icon into waiting status. After the server receives the request, allocate the equipment identification key into the database section of the HTML code sent to the client according to the equipment identification query. The client will update the corresponding region in asynchronous access callback function [8]. It is shown in Fig. 45.6 to create a equipment display page.

45.5 Conclusion

Although it is easy to realize remote monitoring based on C/S structure and control equipment monitoring system, the equipment status monitoring and control relies on the monitoring clients, independent of each other to form a large amount of information “islands”. In this paper, a Integrated information monitoring platform has been created based on Web services technology. Currently, protocol parsing have been realised in more than 20 TT&C sets in Kashi, Qingdao and other TT&C stations, which greatly facilitates the monitoring network management center to monitor and control the equipment status in time, shown in Fig. 45.7.

Proven by the practice, Web-based Integrated monitoring system has the following advantages:

距离 0.0Km
速度 0.0m/s
方位 0.00°
俯仰 0.00°
AGC 0.00V
功率 0W

Fig. 45.7 Interface of integrated status revelation to TT&C equipments

1. Simple deployment of maintenance with low cost. The client of B/S architecture is a Web browser which belongs to the thin client mode, users simply manage and maintain servers without maintaining clients [9].
2. Flexible display and easy realization. Based on the browser, B/S architecture brings much better user experience with richness, vividness, interactiveness, meanwhile easier to control development costs.
3. Reduction of data exchange and network transmission. Compared to the data exchange layer structure, the client need only one single time for data exchange. Multiple data interaction between the intermediate layer may be defined in the database, and the data traffic on the network is greatly reduced [10].

In summary, it provides a cross-platform, simple and consistent viewing environment for Integrated TT&C equipments on the basis of Web services platform. The system development environment is separated from application environment. Compared with C/S architecture, the B/S architecture greatly improves the intuitiveness and flexibility to well meet the diverse needs of TT&C equipments with favorable scalability and versatility.

References

1. Yu ZJ (2006) Status quo and development of spaceflight TT&C systems. Eng Sci 8(10):42–46
2. Wang XD, Yang XH (2010) Application of SNMP in C/S mode remote monitoring and control management system design. Telecommun Eng 50(1):42–46
3. Bai DM, Dang Q (2012) Design and implementation of an integrated monitor system for TT&C equipment. In: The 26th conference proceeding of spacecraft TT&C technology in China, Nanjing, 686–691
4. Banerjee A, Corera A (2002) C# web service—building web services with .Net remoting and ASP .Net (trans: by Kang B). Tsinghua University Press, Beijing
5. Wang JT (2005) The research and implementation of GI services based on web technology. Doctoral thesis of information engineering university, Zhenzhou
6. You LZ, Guo YC, Li CX (2006) The principle and application of Ajax engine. Control Autom 22(6):205–207
7. Asleson R, Schutta NT (2006) AJAX tutorial (trans: Jin L, etc). Posts and Telecom Press, Beijing
8. Liu YQ, Lei P, Han HL et al (2013) Visual collaborative flight simulation based on Google Earth. Flight dynamics 31(3)
9. Wang Y, Li JD, Yin XH et al (2013) Designing of general railway monitoring system base structure. In: Proceedings of the 32nd Chinese control conference, Xi'an, 8189–8194
10. Yao XR (2008) The research of system in remote fault diagnosis based on the web. Doctoral thesis of Xi'an Technology University, Xi'an

Chapter 46
Design and Implement of Service Bus in the Multi-mission System

Hongwei Xing and Wenli Zhai

Abstract In the future, with the development of spaceflight, aerospace and various of flight experiment missions, the frequency of technology, innovation and technical complexity is increased, as well, the mission complexity, diversity and randomicity become more and more outstanding. Therefore, adaptability, flexibility and the flexibility of multi-mission are demanded more. The design of the software architecture is demanded to provide the ability for the different type mission. Based on the Service-Oriented Architecture (SOA), service bus of the multi-mission system is analyzed and designed in this paper, the service bus framework for the multi-mission system is built, the implement and application procedure of the service bus is proposed.

Keywords Service bus · SOA · Multi-mission system · Server registration · Message exchange

46.1 Introduction

Along with the increasing development of the science and technology in Chinese spaceflight, there are more and more missions in the spaceflight, space and all kinds of flight-experiment, which also require the development of multi-mission parallel and cross-functional cooperation. Therefore, the design of the systemic organization must fully support various missions, and to some extent, fulfill the requirements from business modification and new business addition. A good systemic organization will play a very active role in supporting of the growth of multi-mission, cross-platform, interactive and real-time system, and also contribute to the efficient development of the software.

H. Xing (✉) · W. Zhai
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: xing_star@sina.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_46

The multi-mission system is adopted from the service oriented architecture (SOA) [1], which achieves the flexibility and generality for the software service platform through standardization and granularity of service functions. The functions offered by the service provider are generally utilized to fulfill the operation, use, display and control etc. It is convenient for the functional custom-built and extension and to improve the flexibility of the software through the development of systemic organization. The key element of SOA is the message exchange mechanism: the service bus. For the multi-mission system, there is high demanding for the real-time operation. However, there are many different types of business application software and the interface of the software is varied and complicated. Now, the Enterprise Service Bus (ESB) [2] products cannot fulfill the requirements from the real multi-mission system. Therefore, it is critical to design and construct the service bus for the multi-mission system to provide the data communication interface for the service and to exchange and process the data in real-time. The design of the multi-mission service bus is based on the.net framework technology and the message mechanism.

46.2 Architecture of SOA Service Bus

46.2.1 SOA Conception

SOA, acting as a software architecture model, is based on the service which is the fundamental structure and functional unit for the software design, development, dispose, invocation and maintenance [3]. Service is in a raw granularity module which can complete some business function independently. The communication between services is through the simple and accurately defined interface without detailed programming or communication module, which implement the distributed dispose, combination and utilization of the loose coupled granularity service. It can overcome different hardware platforms, software platforms and website. It can merge various services to a new application. It demonstrated standardized service interface externally, enhanced flexibility and comprehensive interactions between services [4]. There are 3 different roles in the service organization: service provider, service customer and service registration center [5]. Figure 46.1 shows the interaction among the 3 roles in the service organization.

46.2.2 Service Bus Introduction

The development of the SOA is aiming to provide full-scale service. The service bus framework has been considered as the core framework to centralize the service management [6]. The service bus is the combination of the traditional middleware

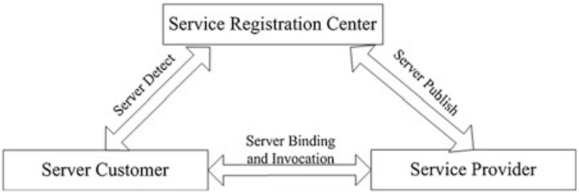


Fig. 46.1 Roles interaction of SOA

technology, xml, and web service [7]. It is a standard integration of the loose coupled service and implicational program, which can be used for message transmission and can be achieved through various middleware technology and program models from technical perspectives.

All the application systems are interconnected through the service bus framework. It is not necessary to aware the location, the standard and the platform of the communication system when any specific application system is exchanging message with other systems, but only to send the message to the service bus. Once the message gets into the service bus, the service bus will process the message, select the route and deliver the message to the destination [8].

46.3 Analysis and Design of Service Bus

In order to construct the multi-mission system based on SOA, it is important to classify, to break down and to abstract the systemic function between the public service and specific applications. It is necessary to standardize the design for the identical or similar functions among different experiment missions, and to take the service provider role and provide both public and specific service to the business applications. Making a comprehensive usage of the service offered by the service provider and focusing on the basic business application formed by all missions, these applications were assembled to the specific business applications with different missions to fulfill the requirements of multi-mission operation, application, display and control. It, therefore, provides direct support to the multi-mission system and takes the service bus role in the service registration center and completes the service registration, management, query, routing/mapping and the service invocation via business applications. Figure 46.2 shows the architecture of the multi-mission system based on SOA.

The service bus of the multi-mission system processes and exchanges the data between the applications and the services or among services, and it also processes the data in real-time and provides the data communication interface. Figure 46.3 shows the framework of the service bus for the multi-mission system.

Service register management center. The main functions of a service register management center include the management of service registration, the

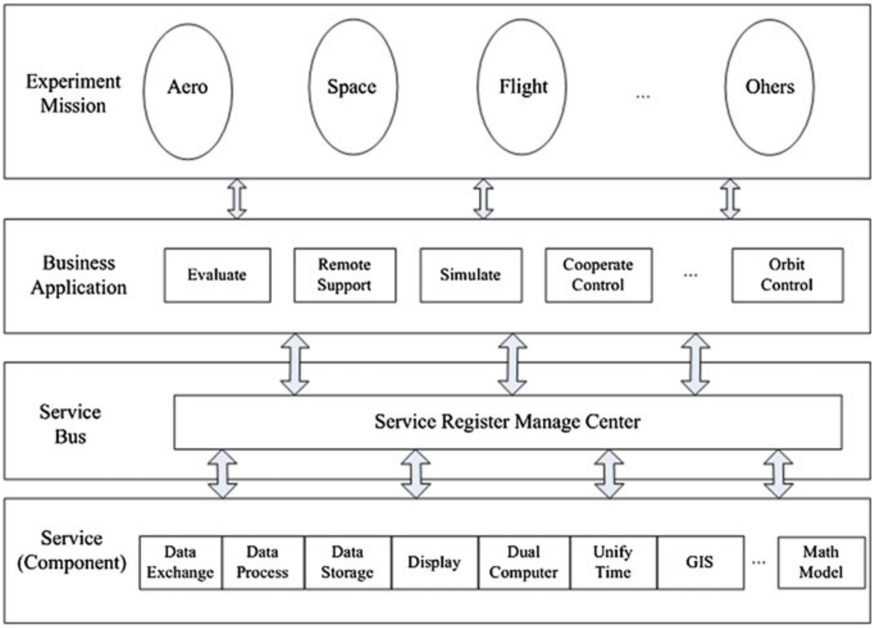


Fig. 46.2 Architecture of the multi-mission system based on SOA

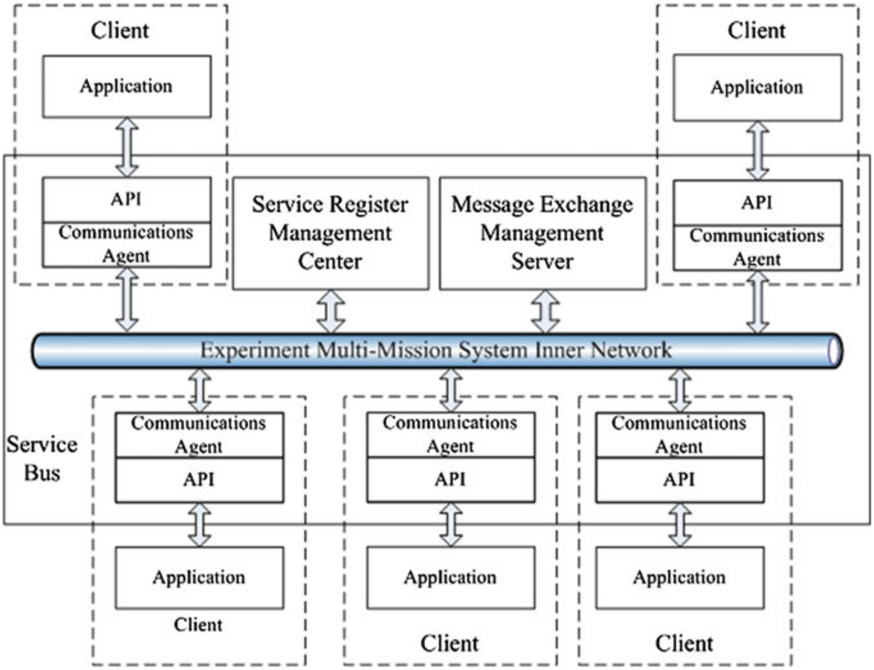


Fig. 46.3 Framework of the service bus for the multi-mission system

synchronization of service information, message publication/subscription, and the synchronization of the message publication/subscription, etc. The service registration is to register the recording, registration cancelling and synchronization registration information to the client of the service bus. The message publication/subscription system is responsible to record the maintenance information, to synchronize the subscription information for the message publisher, and to manage the information publishing address.

Message exchange management server. The message exchange server is responsible for the receiving of non-real time message, message queue management, message transmission, the permanent storage of message and the monitoring and controlling of the message exchange system. The message receiving function is to receive the non-real time message sent by the message publisher, the message queue management function is the message enqueue and dequeue operations. The message transmission function is responsible to send the message based on the subscription information. The monitoring function is to collect the working status of the message exchange server and the end user, and monitor the exceptionality occurred during the operation process.

Service bus client. The service bus client is responsible for service registration, service query, message registration, message subscription, service agent and message sending. Service registration is to send the software service registration information to the service registration center, and the service query is responsible to check the service registration information from the service registration center upon clients' requirements. Message registration is to register the software information to the service registration center, and the message subscription is to subscript or unsubscript the message. The service agent is responsible to establish the connection with the service provider and forward the service application and feedback. The message sending is to send the real-time data to the designation via the point to point or multicast mode.

46.4 Implement of Service Bus Software

The service bus software is public service software, and runs on the distributed computer network of the multi-mission system. It shares the service provided by the internal provider and fulfills the operation, utilization, display and control requirements from the users. It provided direct support to the missions from the system and complete various missions, such as service sharing, message sending and bus management.

The service bus software is responsible for the service listing updates, synchronization and message exchange among different subsystems. The service bus software mainly implemented the management of registration, message exchange, the service bus client and the service bus monitoring and controlling.

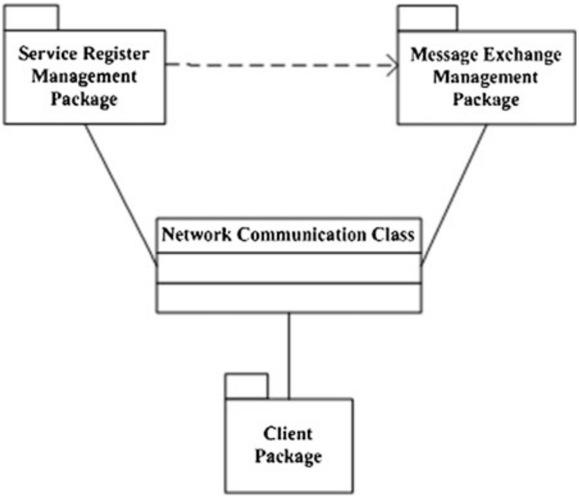


Fig. 46.4 Service bus logic structure

Registration management. It is installed on the registration management server and implement the service management registration and message management registration.

Message exchange. It is installed on the message exchange server and implement the sending/receiving of the non-real time message, message queue management, the permanent storage of the message/data and the monitoring management.

The service bus client. It is installed on the distributed computer network and implement the operations of the service bus agent, real-time/non-real time message exchange agency, and file exchange agency etc. The service bus client communicated with the software applications through the unified API.

The service bus monitoring and controlling management. It is installed on any terminal of the network and implement the remote monitoring and management of the service bus and message exchange etc.

46.4.1 Logic Structure

As shown in Fig. 46.4, service bus software consists of the service register management package, message exchange management package, client package and network communication class.

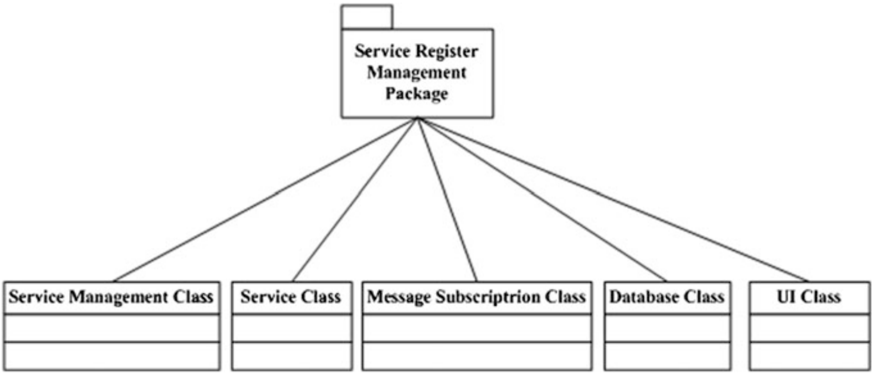


Fig. 46.5 Service register management package

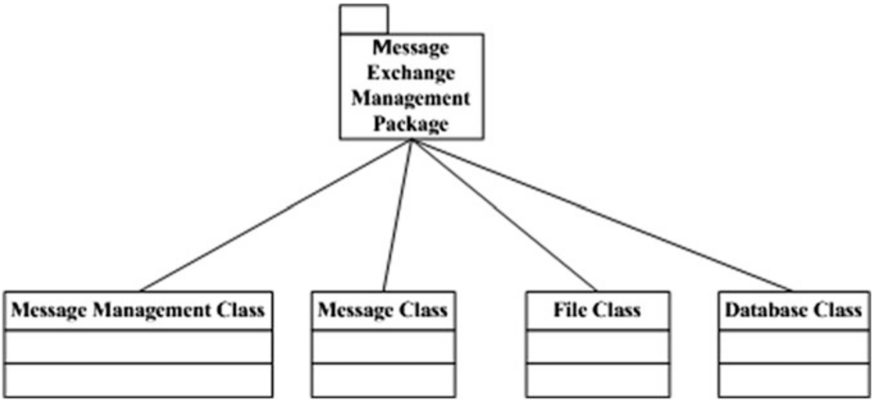


Fig. 46.6 Message exchange management package

46.4.2 Package Structure

Service register management package. As shown in Fig. 46.5, service register management package consists of service management class, service class, message subscription class 、database class and UI class. These classes are used to describe service management method, service definition and operation, message subscribing definition and operation, database management and user operation respectively.

Message exchange management package. As shown in Fig. 46.6, message exchange management package consists of message management class, message class, file class and database class. These classes are used to describe news type, structure and transmission, and file definition and transmission as well.

Message exchange management package. As shown in Fig. 46.7, service bus client package consists of user class, service management class, service class, file

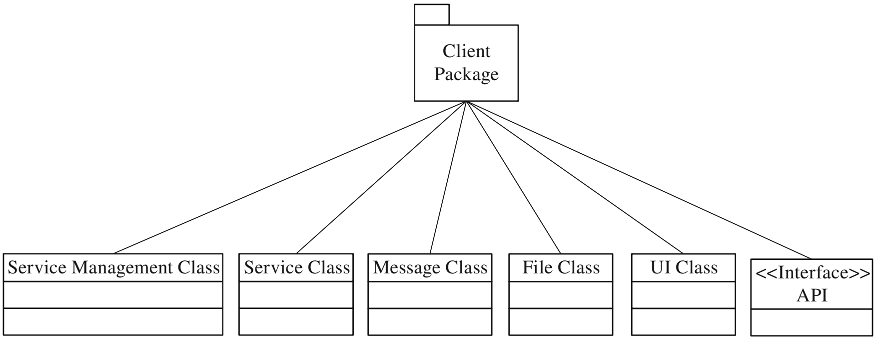


Fig. 46.7 Service bus client package

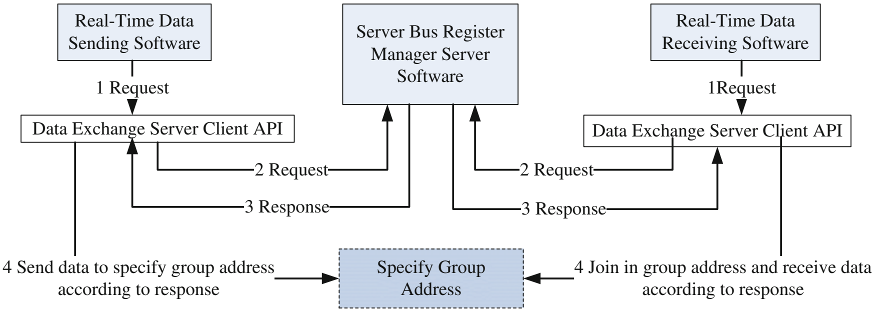


Fig. 46.8 Flow of real-time data exchange

class, message class and user API class. These classes are used to describe user identification, service management operation, service definition and operation, message subscription definition and operation, message structure and transmission, file definition and transmission, database management and user operation.

46.4.3 Data Exchange Service

To illustrate the principle that how these services utilize service bus, a example of data exchange service flow is presented as shown in Fig. 46.8. Data exchange service components provide data communication for those users who need real time data exchange, and all users will share data exchange service via unified API interface.

When real-time data sending software initiates, a real-time data sending request is sent, via data exchange service client API component that integrated in the software, to the service bus registration server software. Network address and port number are assigned based on real-time data exchange definition which designed in

Table 46.1 Data exchange service client API

Attribute/method		Description	Note
Attribute: DataService		Information ID	
Attribute: networkIP		IP Address	
Receive request	Method:	Initialize request	return Sting
	Sting ReceReg (IPaddress networkIP)		
	Sting ReceReg (IPaddress networkIP, string DataSer)		
	Method: Byte[] Receive ()	Receive data	return Byte[]
	Method: void UnSubscribe ()	Cancel request	
Send registration	Method:	Initialize registration	return Sting
	Sting SendReg (IPaddress networkIP)		
	Sting SendReg (IPaddress networkIP, string DataSer)		
	Method: Bool Send (Byte[] Sending)	Send data	return Bool
	Method: void RegisterCancel ()	Cancel registration	

advance, or allocated dynamically. After getting the feedback response, data exchange service client API component is added to the group address, and then, real-time data sending is achieved.

When real-time data receiving software initiates, a real-time data receiving request is sent, via data exchange service client API component that integrated in the software, to the service bus registration server software. Network address and port number are assigned based on real-time data exchange definition which designed in advance, or allocated dynamically. After getting the feedback response, data exchange service client API component is added to the group address, and then, real-time data reception is achieved.

The description of API interface for client data exchange service is given in Table 46.1.

46.5 Tag

The service bus is located at the center of the multi-mission system in the SOA framework. It provides stronger support to the multi-mission system by decreasing the number, the size and the complexity of interface. The service bus achieved the integration, the management and the running of upper layer application and the service components, as well as among service components. The newly added service components and applications can be installed in the “soft receptacle” of the service bus as the "soft plug-in" and provide support to the multi-mission management. So far, there is no standard process or method of technology for the service bus software. Therefore, the optimization of the service bus related technology is very important in the near future.

References

1. Xing SM, Zhou BS (2008) Progress of research on service-oriented architecture. Comput Sci 35 (9):13–20
2. Qi Q, Liao JX, Wang C et al (2008) Research and design of data service bus based on SOA. Comput Sci 35(6):126–127
3. Xu XM (2013) Research on range information system integration based on SOA. J Telemetry, Tracking Command 34(5):73–76
4. Deng LJ, Luo JH, Wei XJ (2008) Research on integration of equipment management information system based on SOA. Microcomput Inf 24(12–3):28–29
5. Zhao L, Shi WF (2009) On enterprise service bus based on SOA and its application. Comput Appl Softw 26(5):117–118
6. Ge FZ, Zhu CJ, Peng TL (2009) Intelligent transportation system framework based on SOA and multi-softman. Microcomput Inf 25(6–1):301–303
7. Wang J, Jiang T (2012) A application of enterprise service bus based on SOA in the banking system integration. Microcomput Appl 31(14):15–16
8. Wang SJ, Jiang S (2006) Enterprise service bus technology in enterprise integration. Comput Eng 32(13):251–250

Chapter 47
Multiparadigm Design of SLE API's Program Interface Based on Design Pattern

Xiaowei Li, Bin Tian, Wei Guo and Shengjun Luo

Abstract It's an important way to improve the degree of internationalization of data transferring between ground stations in deep space TT&C by applying the SLE protocol. Meanwhile, the design of SLE API is the most important part in SLE applications development. This paper introduces the encapsulation of SLE API's interface using of design patterns, describes how to use design patterns in multiparadigm. Moreover, it's also emphasized on the point that how the Class mechanism in object-oriented paradigm is adapted to procedure-oriented paradigm and call-back function is introduced to applied to Observer pattern in latter paradigm. Through multiparadigm encapsulation, the SLE API can acclimatize to most software development environments.

Keywords SLE application program interface · Design pattern · Multiparadigm · Deep space TT&C · Space link extension protocol · Software development · Callback function

47.1 Introduction

Space Link Extension (SLE) helps the deep space station and mission center to exchange the information of telemetry, telecontrol and digital transmission in global scope. As an international standard, SLE is a kind of space data communication protocol which is developed by Consultative Committee for Space Data System (CCSDS) [1], and could help the ground system providing cross-support in space flight missions [2, 3]. SLE API (Space Link Extension Application Program

---

X. Li (✉) · B. Tian · W. Guo · S. Luo
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: jinglelxw@163.com

X. Li · B. Tian · W. Guo · S. Luo
Xi'an Satellite Control Center, Xi'an 710043, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_47

Interface) is an important part of the application of SLE protocol, which encapsulates the SLE protocol and communication protocol [4–6]. CCSDS provides the complete SLE API specification and issues the versions of NASA and ESA, meanwhile it also encourages each country to develop their own SLE API, only requires the cooperation of both sides using SLE protocol to supply the cross-support. China attempt on the self-control SLE API development in the construction process of deep space TT&C. Interface is the difficulty of SLE API design, and according to the CCSDS recommendations, SLE API can only support the object-oriented paradigm in C++ language which would limit its usage. This paper introduces the interface encapsulation process of the multiparadigm, and emphasizes how the design pattern conception in Object-Oriented domain is used to switch between different paradigms.

47.2 SLE Protocol and API Synopsis

47.2.1 SLE Protocol Introduction

CCSDS SLE protocol is the extension of the Space Link's TC and TM services between the ground stations and spacecraft, provides a standard interface and control process for space organization using the global control of resources [1]. It's a integrated system including the design thought, core algorithm, interactive timing, and control flow.

The SLE Services include two major elements:

- data transfer services that move space link data units between ground stations, control centers and end-user facilities;
- management services that control the scheduling and provisioning of the transfer services.

47.2.2 SLE API Introduction

As a key section in the SLE application, SLE API realizes the SLE protocol and could provide the programming interface for SLE provider and SLE user application above it. By encapsulating the underlying details about data processing and network transmission, SLE API let the upper users achieve the required return and forward data processing function nothing more than invoking the interfaces according to the business logic. What's more, it also has the other functions defined in SLE standard, such as parameter setting, status monitoring and reporting. Internal structure of SLE API is shown in the Fig. 47.1.

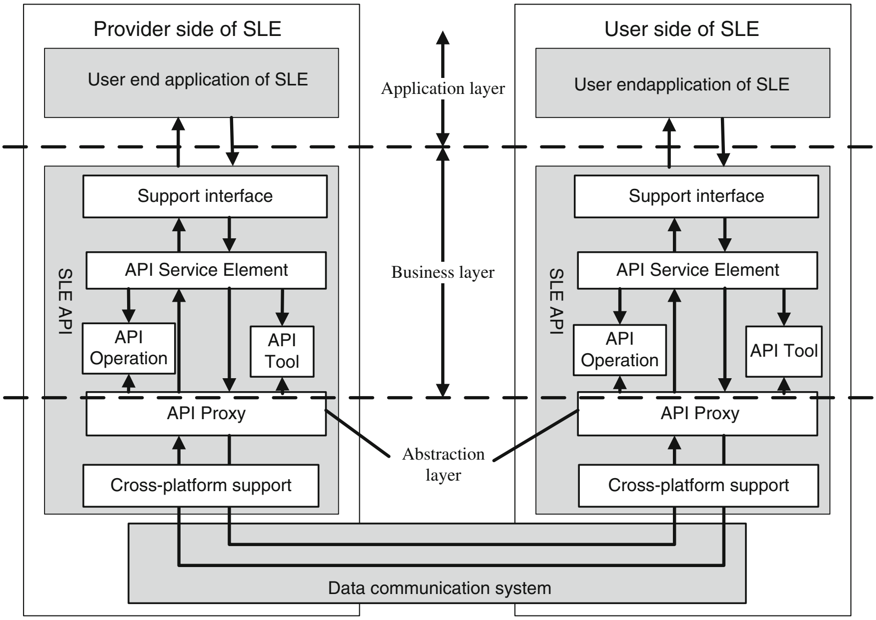


Fig. 47.1 Internal structure of SLE API

The main effect of SLE API consists of two parts:

- To interact the information with the SLE application software. It receives the call command from the SLE application software, and return the transmission state and results back to SLE application software;
- To interact the information with the peer SLE API software. Through ensuring provider and user follow the same transfer protocol, data unit format and information exchange rules, SLE API implements the transmission, control and certification of space link data.

Interface design is one of the most important parts of the API design, this paper focuses on how to design SLE interface by the design pattern.

47.2.3 SLE API Characteristics

CCSDS SLE API has the following characteristics [5]:

- Function structure is complex. SLE API encapsulates SLE protocol specified complex logic, including Return All Frames (RAF), Return Channel Frames (RCF), Communications Link Transmission Unit (CLTU) services and the underlying network transmission functions. It also provides the time, reports and other auxiliary functions. In conclusion the API's structure is very complex.

- Strong coupling. SLE API's internal function modules are tightly coupled, working together to provide the SLE service for SLE applications.
- Only supports the object-oriented paradigm. CCSDS recommends the SLE API developing by C++ language and only provide the access interface about object-oriented paradigm, which will limit the application scope.

In order to overcome the shortcomings of CCSDS SLE API, design pattern is used in the real development of SLE API to simplify the interface and support future multiparadigm application.

47.3 Using Design Pattern to Encapsulate the Interface of SLE API

Design pattern is an excellent design method making the framework elements independence, and then each element is visited through the message. It is helpful for maintaining the relatively stable application framework [7–9]. Design pattern is an important concept of object oriented domain. Facade, Adapter and Observer pattern are most commonly used in components and class library designs [9]. The thought of design pattern is also applicable to cross-supply programming multiparadigm, and then the overall design can be partitioned into three parts which will be introduced later.

47.3.1 Facade Pattern to Encapsulate the Overall Interface

Facade pattern provides a consistent interface for the system, which defines a high-level interface to make it easier to access the system.

47.3.1.1 Process of the Encapsulation

Facade can eliminate the dependencies of objects and establish the hierarchical structure system in practical application. The structure of facade is shown in Fig. 47.2a:

Facade is particularly suitable for the encapsulation of SLE API components, the external interface diagram of SLE AP is shown in Fig. 47.2b. According to CCSDS standard, SLE API includes four core modules and more than 60 core classes. After using the facade, API reduces the number of the interface, achieves the decoupling between API and SLE applications.

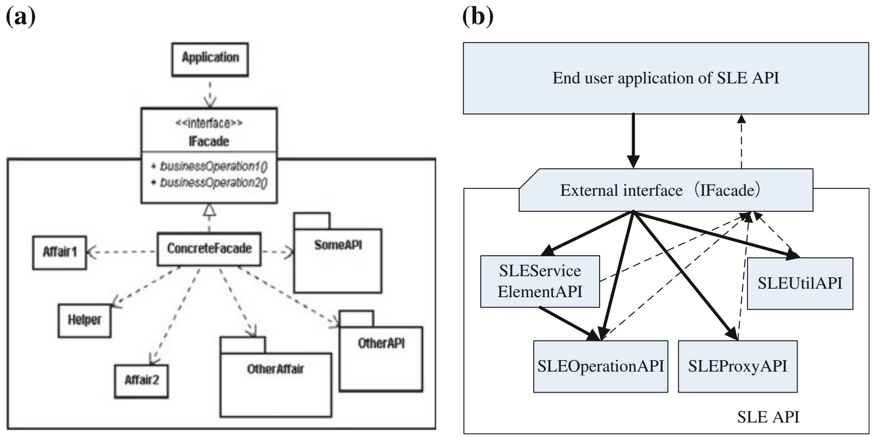


Fig. 47.2 **a** The structure schematic diagram of facade; **b** The external interface diagram of SLE API

47.3.1.2 Description of the Structure

IFacade interface contains initialization, service operations and other ten methods, and provides the methods for SLE end user to realize all the service operations of SLE API. The relationship between IFacade interfaces and other subsystems is shown in Fig. 47.3:

IFacade interface only plays the bridge role between SLE end user and each module of the API, when the SLE end user sends the request to IFacade interface to communicate with the subsystem, IFacade will forward the message to the appropriate objects to do the real work. A part of IFacade's initialize codes are shown as follows:

```
bool IFacade::Initialise(const char * proxyPath...){
        pg_ISLE_Proxy=new ISLE_ProxyAdmin();
        result=pg_ISLE_Proxy->Configure(proxyPath,..);
        ……
}
```

When the end user invokes the method, IFacade interface will call the Configure method of ISLE_ProxyAdmin module to complete the configuration.

47.3.1.3 Advantages of Using Facade Pattern

Facade pattern has the following advantages:

- Reduce the number of access interfaces. With Facade pattern the clients do not need to access subsystem object directly, and then reduce the number of fan-out in the process of access.

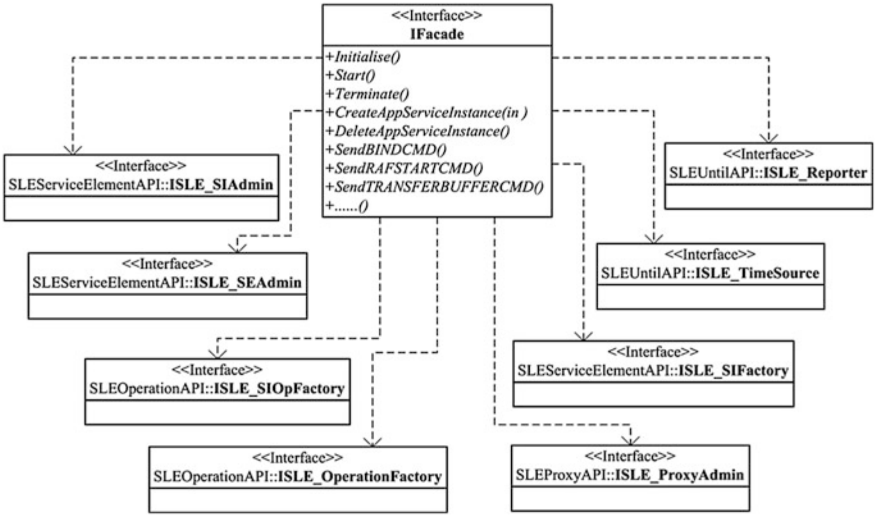


Fig. 47.3 Relationship between IFacade interfaces and other subsystems

- Reduce the costs of interaction. As SLE API uses the same language inside, it helps to reduce the overhead of context switching, thereby reducing the costs of the whole system interaction.

47.3.2 Adapter Pattern to Support Procedure-Oriented Paradigm

This section introduces the encapsulation process that suit to procedure-oriented paradigm by adapter pattern in further, which provide pure C interface of SLE API.

47.3.2.1 Problem

In the recommendation of CCSDS SLE, SLE API should be developed by C++ and only provide C++ foreign interface. This will limit the application of SLE API, the reasons are as follows:

- C++ interface can be accessed only by the C++ application. Essentially, C++ class is set up by the data structure and its operation, and both have adopted different compilation process. Reference name and calling conventions factors make the compiled C++ class almost impossible for a non-C++ language to access, even C# language which is also object-oriented paradigm language;

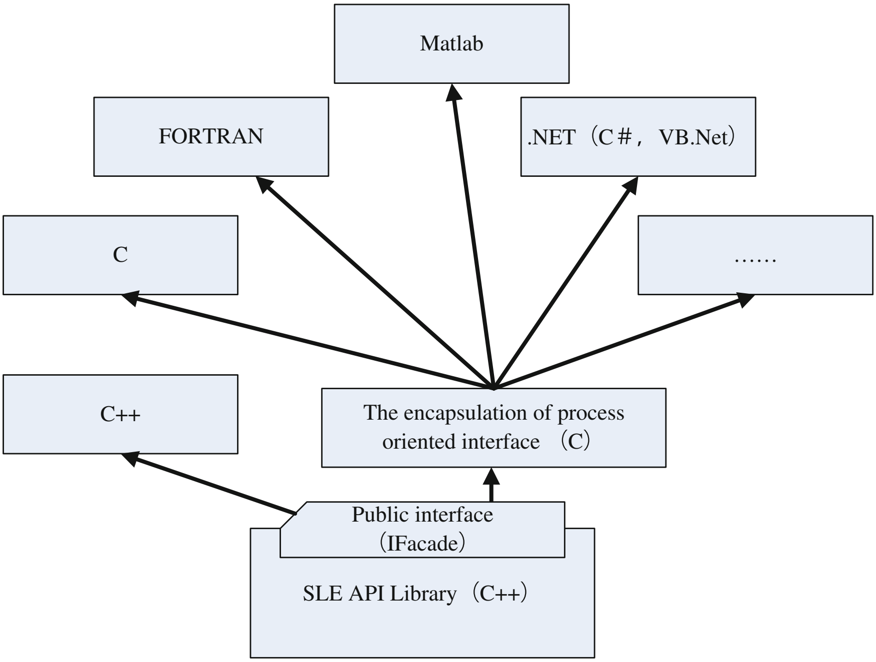


Fig. 47.4 SLE API's different programming languages supporting diagram

- C++ language is not suitable for the development of large-scale visualization applications. Because of the inherent complexity, the difficulty and cost of visual system development will be doubled and redoubled by using C++, compared with. Net framework. Therefore.Net frame and other languages do not support the C++ interface (but support pure C function).

As SLE applications deployed in deep space of TT&C are general requirements with friendly graphical interface, pure C interface encapsulation of SLE API libraries will has important application significance. After encapsulation, it can really support object-oriented and procedure-oriented paradigm and will be compatible with the current most programming languages such as C, C++, C #. The multi-paradigm extension of SLE API which supports different programming languages is shown in Fig. 47.4:

47.3.2.2 Adapter Pattern Cross Different Paradigms

Adapter pattern is defined as "to convert one class interface to another interface which customers want, and then make the irrelevant classes working together" [7]. Adapter pattern is completely suitable for encapsulating the C++ interface in pure C

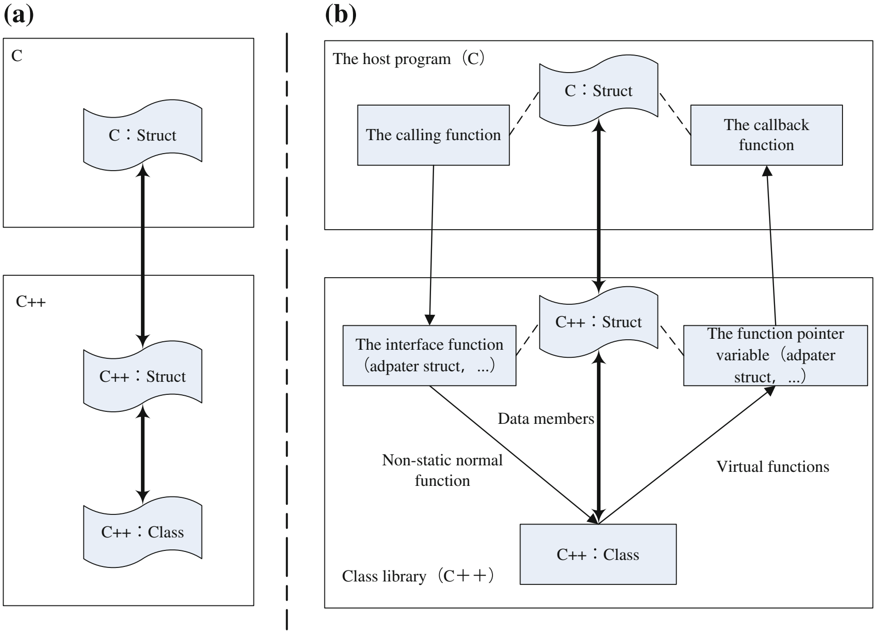


Fig. 47.5 **a** The mapping diagram of language mechanism between different paradigm; **b** The process of mapping class

language on the concept. Due to the interface concept in C++ language generally refers to the abstract class, which contains virtual functions, the crux of this problem is how to encapsulate the class C++ function by using pure C language.

After analysis of C and C++ language mechanism, we find that the "class" of C++ language mechanism is based on the "struct" of C language added with the operation. Simultaneity the C++ structure mechanism has been expanded greatly and can support member function and inheritance mechanism. The compilation process and results of struct and class in C++ are exactly the same and can be considered to equivalent on the language level [10]. Therefore, "struct" in C++ can be as a communication bridge between "struct" of pure C and "class" of C++, as shown in Fig. 47.5a:

Based on the above basis, the data and function members of C ++ "class" can be mapped to the elements of C language, as shown in Fig. 47.5b. The concrete content includes:

- Data members can be directly replaced by the struct;
- Non-static conventional function can add a matching interface function for adaptation; the function should include the class parameter;
- Virtual function is relatively complex and will be refilled by call-back function.

47.3.2.3 Adaptation Process

This section describes the adaptation process of non-static conventional function, the virtual function part is introduced in Sect. 47.3.3. When the pure C function access IFacade member function, and two part libraries using the "struct" keyword can be adapted to:

- in the C++ class library of SLE API, establish the adaptation function of C language to each IFacade interface;
- in the external access program, add a function with the same name to access the function definition.

The codes of "initialization" methods above are shown below:

```
//Codes Inside SLE API:
class IFacade {
bool Initialise(const char* proxy-Path...);
};
extern "C" bool call_IFacade_Initialise (IFacade * p,const char *
proxyConfigFi-lePath...){
  return p->Initialise (const char * prox-yConfigFilePath...);
}

//Codes Outside SLE API:
double call_IFacade_Initialise (struct IFacade * p, const char *
proxyPath...);
void ccc(struct IFacade * p, int i)
 {call_C_f(p,i);}
```

Some tests prove that the SLE end user could access the internal library method of SLE API by pure C interface function smoothly.

47.3.3 Observer Pattern to Feed Back the Information

Observer pattern defines a one-to-many dependency between objects, when an object changes state, all the objects that depend on it will be informed and updated automatically [7]. This interaction is known as publish-subscribe, its target is let the component inside call the method outside, or resolve the problem when a state of the component inside changes, how to make the external methods to respond. Notification does not need to know who is the observer. Procedure-oriented development of traditional can use the call-back function to achieve the same effect as the interface and the virtual functions that commonly used in object-oriented practice.

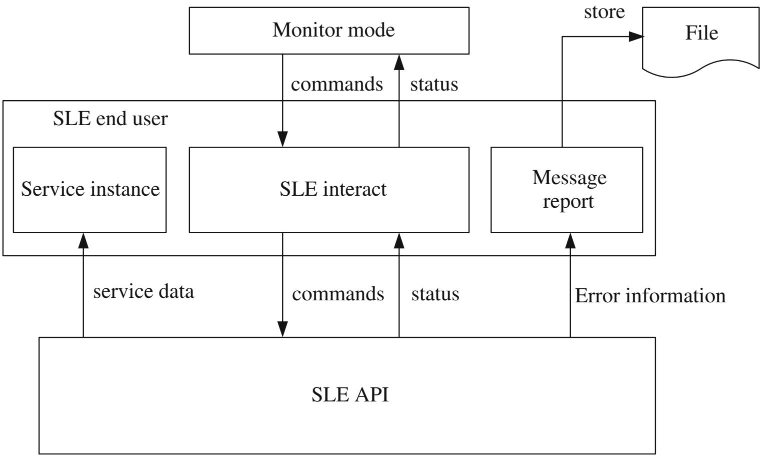


Fig. 47.6 Interaction relationship between SLE application and SLE API

Although observer pattern use different mechanism in procedure-oriented and object-oriented paradigms, both of them have the operation of registration and response. This section will give the example about the different paradigm of the observer pattern by log-reporting functions.

47.3.3.1 The Application Foundation of the Observer Pattern

A large number of information will be exchanged between SLE application and SLE API in the running time. Downward information includes the commands such as the initialization, start, shutdown. Upward information includes the status of SLE API, the service data, and so on. It is shown in Fig. 47.6:

As SLE API is used by upper application, if it want to return notice correctly, it must know the information of subscriber in SLE end user, which provides the application space to use observer pattern.

47.3.3.2 Observer Pattern in Object-Oriented Based on Interface Mechanism

In observer pattern of object-oriented, the key elements are the subject and the observer, a subject can have many observers dependent on it. When the state is changed, all observers could obtain the notified message. The structure diagram of observer pattern and its application of SLE API are shown in Fig. 47.7 respectively:

As shown above, observer pattern is used in SLE API log-class design. It consists of two steps:

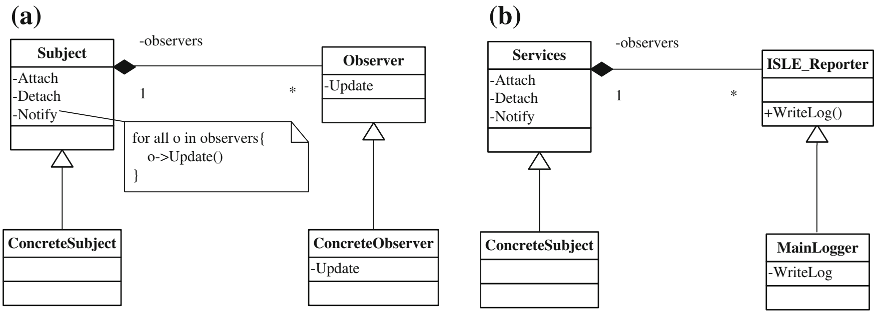


Fig. 47.7 **a** Structure diagram of observer pattern; **b** Application of observer pattern in SLE API

- SLE application registers the log interface object pointer in the observers list of service class in SLE API;
- In the run-time process, SLE API drive the WriteLog function of MainLogger class in SLE application through the registry pointer.

47.3.3.3 Observer Pattern in Procedure-Oriented Based on Call-Back Function

Call-back function is the one that call the other functions which has been registered of invoker reversely. This design allows the underlying code calls subroutines in high definition. Call-back function has the different forms in different development languages. For example, the pure C language directly uses the function pointer, yet.net framework provides a type of safely reference encapsulation, which named "Delegate" to define the function pointer. Call-back function is usually in the same abstract level with the original caller.

Observer pattern in procedure-oriented is based on call-back function, as shown as in Fig. 47.8a. A function pointer to reverse the operation is directly carried out and can fulfill Run-time Type Identification (RTTI) as virtual function in object-oriented language.

Application of call-back function in SLE API is shown in Fig. 47.8b. Corresponding log processing observer pattern steps include:

- During the process of initialization, SLE application registers the function pointer of the log call-back to SLE API;
- When need to fill the abnormal logs in the processing of the execution, SLE API calls the registered function pointer, the SLE application call-back the corresponding function of MainLogger.

By using a call-back function mechanism, SLE API can call the function of SLE application even though it knows nothing about the application, and thus breaking the cycle dependent.

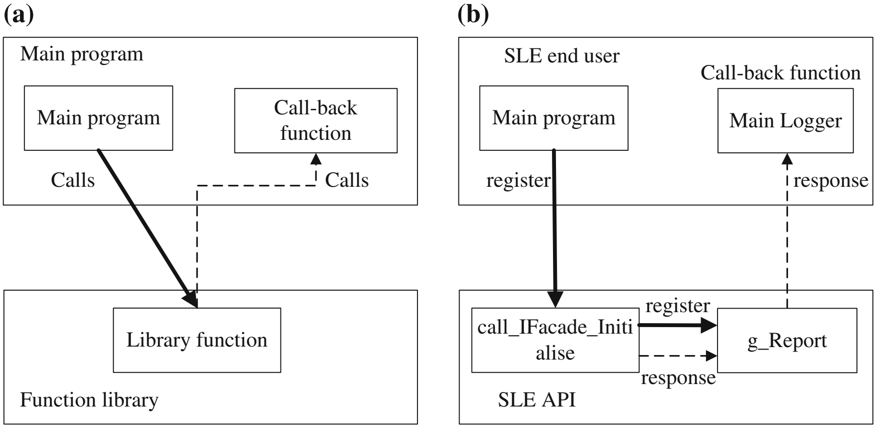


Fig. 47.8 **a** Diagram of call-back function; **b** Application of call-back function in SLE API

47.4 Conclusion

This paper introduces the whole process of using design patterns to encapsulate the external interface in the design of SLE API software, and apply several design pattern of object-oriented domain in mutilparadigm, and then gives the complete solution of "Class" adaptation to procedure-oriented. After the multiparadigm encapsulation of interface in SLE API, the SLE protocol can be applied in a variety of common development platform, and then greatly reduce the threshold and achieve remarkable effect. It's hoped that this research can help the readers understand the SLE protocol in further, deepen the understanding of design patterns and design more excellent software in deep space TT&C domain finally.

References

1. CCSDS (1996) Space cross support reference model. Part 1: space link extension services. 910.4-B-1, Washington, DC
2. Tai W, Maldai P, Hell W (2001) Key system architecture issues concerning space link extension services. In: 2nd ESA workshop on tracking, telemetry and command systems for space applications
3. Wang GH, Yang HJ (2008) Application of space link extension(SLE) services in interagency space TT&C networks. Telecommun Eng 48(1):20–25
4. Zhao S, Lu XL (2010) The application of SLE API based on SCM. J Spacecraft TT&C Technol 29(6):74–78
5. CCSDS (2006) Space link extension-application program interface for transfer services–summary of concept and rationaleccsds. 914.1-G-1, Washington D. C
6. CCSDS (2008) Space link extension-application program interface for transfer services—core specification. 914.0-M-1, Washington DC

7. Gamma E, Helm R, Johnson R et al (1994) Design patterns, elements of reusable object-oriented software. China Machine Press, Beijing
8. Li XW, Zhang YF, Wang HB (2012) Research of design patterns application in communication components of TT&C domain. J Spacecraft TT&C Technol 31(6):63–68
9. Reddy M (2003) API design for C++. Posts and Telecom Press, Beijing
10. Lippman SB (2001) Inside the C++ object model. Huazhong Science and Technology University Press, Wuhan

Chapter 48
Product Line Architecture for TT&C Software Support Framework

Zhilin Wang, Xinming Li and Dongdong Li

Abstract One of the most promising techniques to improve the quality and productivity of TT&C software is product line engineering, the core of which is the product line architecture for TT&C software support framework (PLATSSF) which is of great help to improve the reusability of architectural design. Since component models are cornerstones of support framework design, the characteristics of them including variety, common requirement and variant requirement were firstly analyzed. Then, mainly by analyzing the three stages of TT&C software, which include development deployment and runtime, the structure of PLATSSF is designed. Thirdly, the function of PLATSSF is instantiated by adding a new satellite to the framework. In this instance the software requirements are analyzed and the contents of software development are provided. In conclusion, PLATSSF implements the production of the TT&C software series, reduces the workload of software development and improves the software quality.

Keywords TT&C software · Software product line architecture · Software support framework · Component model

48.1 Introduction

Space TT&C Network is a special network by which orbited spacecrafts are tracked, measured and controlled [1]. With the expanse of the space enterprise, space TT&C center and the space application center need highly cooperate with

---

Z. Wang (✉) · D. Li
Beijing Space Information Relay and Transmission Technology Research Center, Beijing 100094, China
e-mail: zhilinwww@126.com

Z. Wang · X. Li
Science and Technology on Complex Electronic System Simulation Laboratory, Beijing 100092, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_48

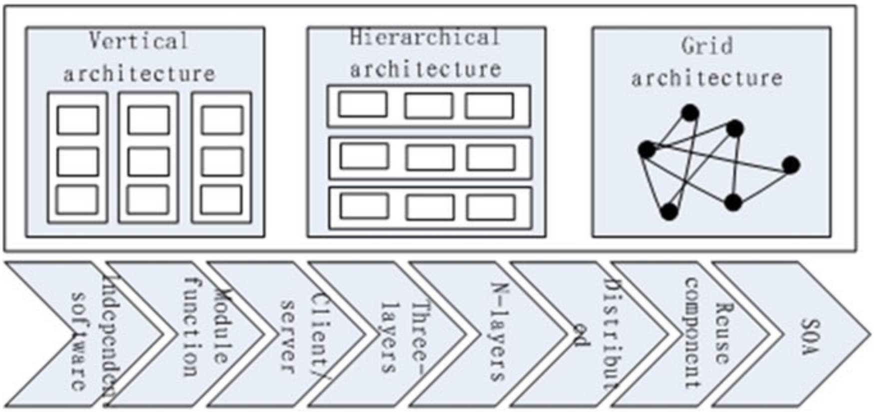


Fig. 48.1 Evolvement of TT&C center software architecture

each other to complete the complex tasks. Space TT&C center is the main control node of the TT&C network, in which TT&C software plays a key role to automate and intellectualize the system. Software support framework provides a deploying and running environment for the function components, and is foundation of space TT&C center software. With increase of mission requirements, TT&C center software's architecture is making the transit to the network one. All in all, software support framework is now facing with new challenge.

In the past 10 years, TT&C center software architecture went through three phases. The first one is vertical architecture, including many representations, such as independent software, module function and client/server. The second one is hierarchical architecture, including many representations, such as client/server, three-layers, N-layers and distributed structure. The Last one is grid architecture, including many representations, such as distributed structure, reused component and SOA. As Shown in Fig. 48.1.

By Far, hierarchical architecture is mainstream and popular for TT&C center software. With increase of requirements, some of TT&C center software is adopting grid architecture.

Component-Based Development (CBD) [2] builds software system by combining the developed components, which can reduce development cost, and improve reuse rate of developed components [3].The developed components are the base units of combination, describe the offered service by interface, define the demand for environment, and can be independently deployed by a third party.

The support framework provides the running and deployable environment for components [4], and is fundamental of the component-based distributed development. During the running state, support framework builds and manages instances of components, allocates system resources, maintains intercommunications of components, and transparently heads off request of the outer system. As support framework has achieved common design and implementation [5], developers only need to focus on certain parts, without considering common parts.

Product line engineering is a software engineering technology which develops special domain product series by using common kernel assets [6]. It involves three phase: domain analysis, domain design and domain completion. Domain analysis is different with common requirement analysis, adds description for commonness and variability of the special domain, and guides the manufacture of reused software assets. Product line architecture (PLA) [7] is a product of domain design. By supporting the domain requirement and implementing the common requirement, PLA realizes systematic reuse. Research on PLA can promote structural reuse of support framework, and improve productivity and quality.

48.2 Domain Analysis of Support Framework

48.2.1 Component Sort

Considering universal principle of function definition in the field of TT&C, TT&C system falls into two categories: task function component and system manage component. Task function component satisfies requirements of special tasks, including compute component, satellite control component, equipment control component, customer data relay component and customer data parse component. System manage component satisfies requirements of system support framework, including task layout component, task schedule component and status surveillance component. The following lists are the details of each component function:

1. Compute component. It receives and deals with the trajectory data, and computes the orbits and the control coefficients of satellites.
2. Satellite control component. It implements the TT&C operations of the satellites.
3. Equipment control component. It remotely monitors and controls of the ground stations.
4. Customer data relay component. It exchanges customer data customer data between the ground stations and the customer centers.
5. Customer data parse component. It deals with data of satellites.
6. Task layout component. It allocates the resources of Space TT&C Network, establishes and manages the task layout.
7. Task schedule component. According to the layout or human instruction, it schedules the all function components of the software system.
8. Status surveillance component. It shows the task progress and system status.

Relation among aforementioned components is described in Fig. 48.2.

Liu Guo-Liang [8] details the analysis framework of the component model. The author aims at the software system of space TT&C.

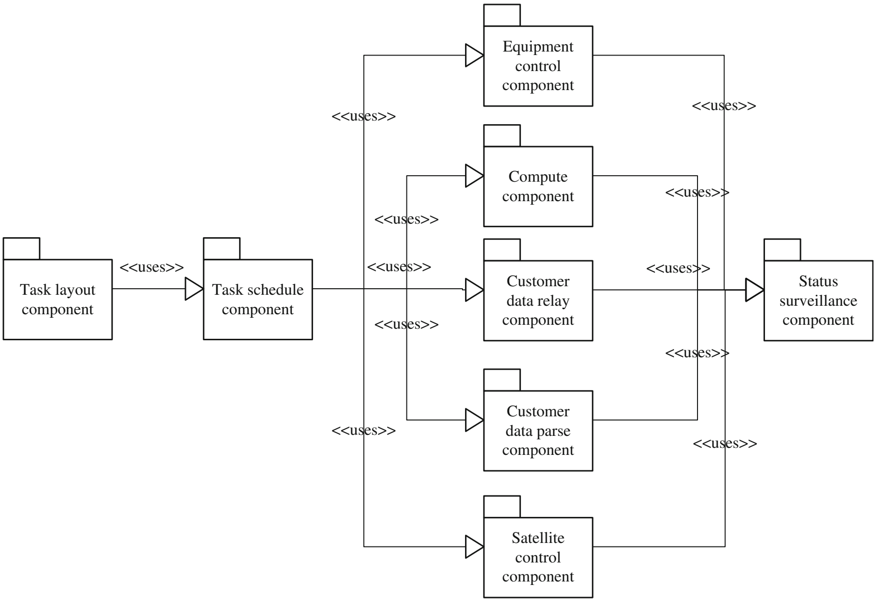


Fig. 48.2 Relation among function components

48.2.2 Component Common Requirement Analysis

The common requirements of component support framework include component deployment, component running, and component maintenance. There are five actors associated with component support framework, which are deployer, manager, operator, resources, and component instance. By analyzing relation among actors and the support framework, domain case model is educed as shown in Fig. 48.3.

By decomposing cases into a series of independent acts, it is achieved that the function requirements refine. In outer view, each act is indivisible. For example, the task cooperation case represents components' cooperates in task. It includes scheduling synchronization, data synchronization, and operation synchronization.

48.2.3 Component Variant Requirement Analysis

There are five sorts of variant requirements, which are symbol configuration, script template, network configuration, display configuration, and special algorithm. As shown in Table 48.1.

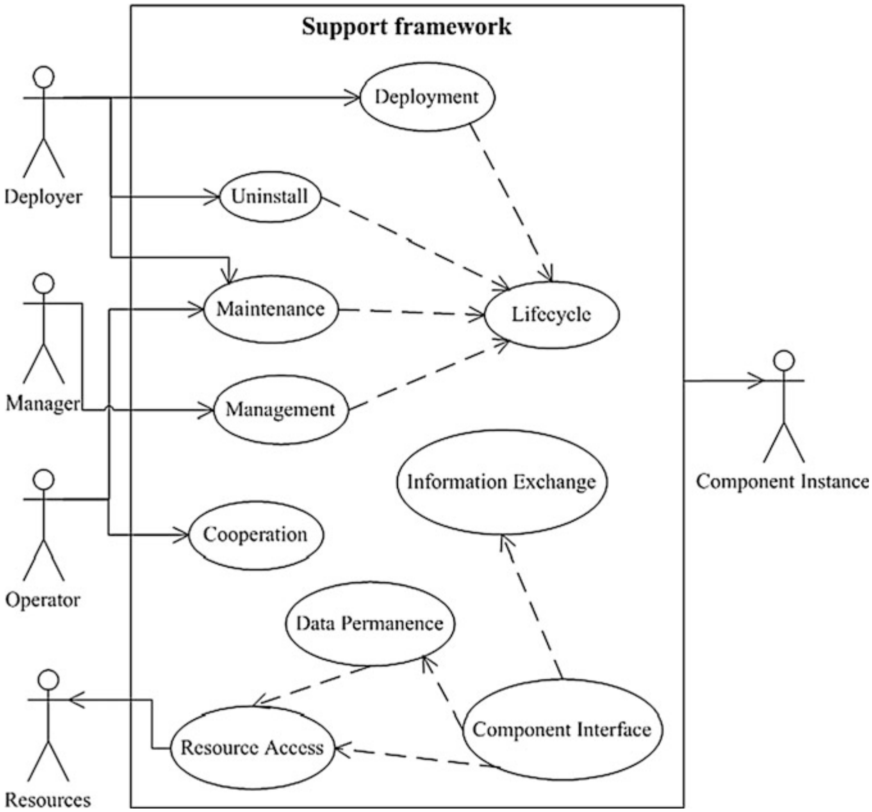


Fig. 48.3 Domain case model

48.3 Product Line Architecture for TT&C Software Support Framework

48.3.1 Development Structure

The development structure includes two parts. See Fig. 48.4.

1. Basic software layer. It contains support framework and component set. It mainly solves the problem of the system common function design, and builds basic platform for task expanse. The edition of basic software is independent with task.
2. Task software layer. Aiming at the special requirements of tasks, it is developed based on basic software layer. It mainly contains various configuration files and some programs for special algorithm. Each task software is independent.

Table 48.1 Analysis of component variant requirement

Type	Description	Software structure	Application
Symbol configuration	There are much similar algorithms in the different tasks. Only by adjusting several arguments, the same algorithm can be reused	The structure of argument set is adopted. By adding new configuration options, the new tasks' requirements are satisfied. By choosing the "file + directory" as the form of argument set, the chief file appoints the configuration directory. The configuration directory includes many files. Each file only contains the requirements of one task	Compute component Task layout component
Script template	The procedure is mostly consistent in different tasks. The main difference is the amount and values of arguments	The structure of "script + argument set" is adopted. The script is collection of procedure and argument symbol. The argument set is a list of argument symbols and values. In running time, script and argument set are dynamically combined. In order to satisfy the different types of tasks, different scripts are applied	Satellite control component Equipment control component Task schedule component
Network configuration	The basic function of data exchange for all customers is same. The differences include IP address, amount of link, and the exchange rule of application layer	The structure of "domain name service + route algorithm" is adopted. The domain name service achieves the mapping between IP addresses and customers. The route algorithm implements the exchange rule of application layer	Customer data relay component
Display configuration	The display style is much different for each task. But the basic display elements are same	The structure of "display framework + page deployment" is adopted. The display framework provides the basic display elements. The page deployment defines the combination of the display elements	Status surveillance component
Special algorithm	There is completely different in customer data. Special algorithm need to developed	The structure of "software algorithm framework + special program + compatible interface" is adopted. The software algorithm framework and special program insure that each special algorithm is developed independently. The compatible interface insures that the data format is consistent	Compute component Customer data parse component

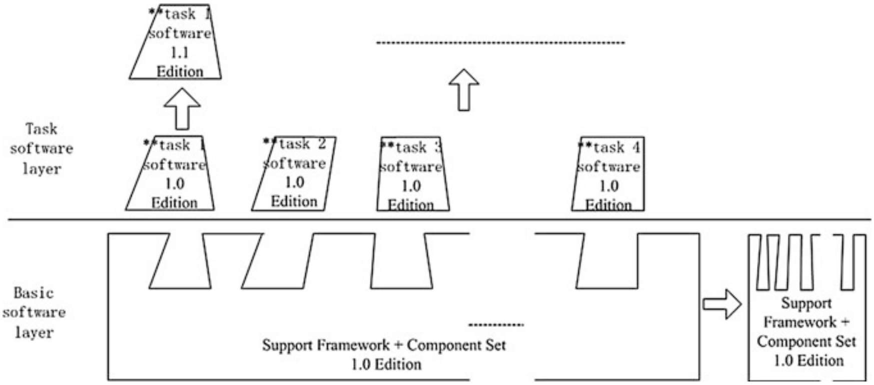


Fig. 48.4 Development structure

Under this development structure, maintenance of basic software and task software can be simultaneously implemented, and different editions of softwares are managed and upgraded respectively.

48.3.2 Layer Structure

The analysis of variant requirement is the base of the component reuse. There are four layers for the component reuse. See Fig. 48.5.

1. Task Procedure. A series of acts make up of one task procedure. Each act matches one component.
2. Component support framework. It provides the running environment for component.
3. Component module. It is component set, and in.
4. Component configuration set. It includes series of configuration files.

Every Task Procedure uses the command interface to send instruction to the component support framework. The component support framework parses the instruction and uses the module loading interface to initialize component. The component module uses the file configuring interface to get task arguments, so as to configure itself at initialization time.

48.3.3 Running Structure

Based on support framework, task software implements the unity among development, deployment, and maintenance. However, in running structure, each task software is independent. According to the task requirements, the editions of

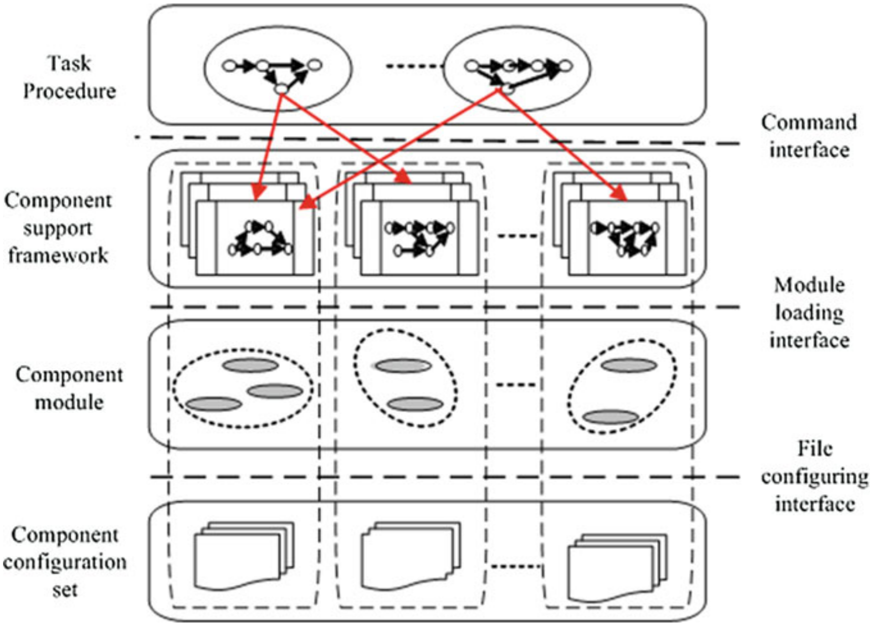


Fig. 48.5 Layer structure

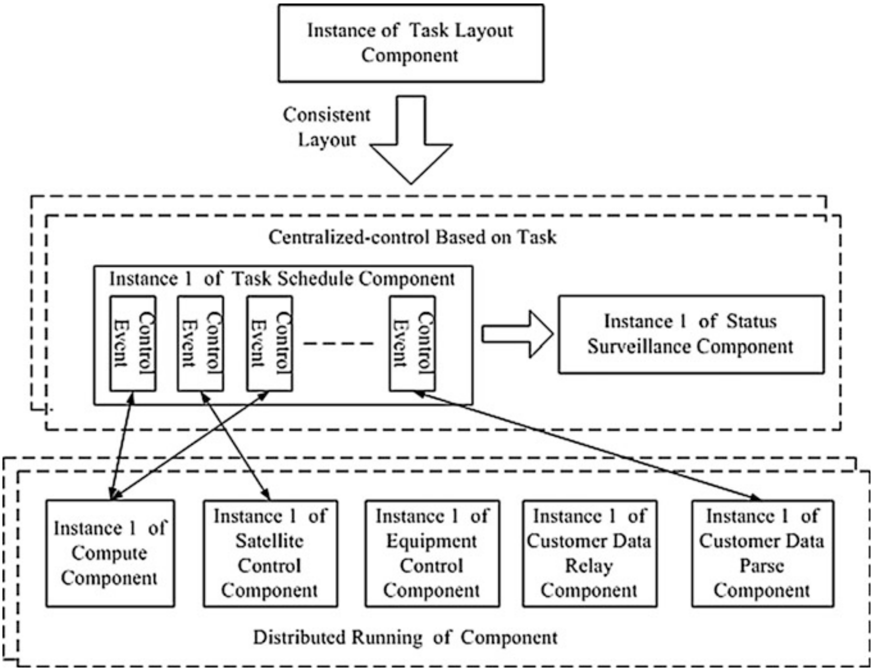


Fig. 48.6 Running structure

configuration files are finished. By configuration files, we initialize the instances of the requisite components. See Fig. 48.6.

In the system runtime, there are three kinds of initialization modes for all component instances. Firstly, it's task layout component which initializes only one instance in the whole system. This instance provides service for all customers and confirms the exclusive result of resource layout. Secondly, they're task schedule component and status surveillance component, which initialize one instance for each task. They build the independent platform for special task to be monitored. At last, all other components are initialized on demand. Every instance is independently running and focus on single function.

In a word, ultimate purposes are achieved, including consistent layout, centralized-control based on task and distributed running of component.

48.4 Software Product

By adding a new satellite to space TT&C Network, we validate the effect of PLATSSF. The task requirements of the satellite have two ways. The one is providing management of satellite platform from orbiting, which includes three stages: going into orbit, on-orbit, seceding orbit. The other one is providing service of data relay.

48.4.1 Task Analysis

48.4.1.1 Management of Satellite Platform

The task requirements include position keep, attitude control, power management, and orbit calculation etc. We take more complex position keep as case and analyze the task requirements. See Table 48.2.

48.4.1.2 Service of Data Relay

Through operating satellite and station equipment, we build a data link and implement application data exchange between customer and satellite. See Table 48.3.

Table 48.2 TT&C task analysis

Num	Require of task control	Phase	Operation instance
1	Orbit prediction	Task layout	Software process
2	TT&C task application		Software process
3	Station equipment prepared	Task running	Station equipment
4	Bidirectional capture control		Station equipment
5	Telecontrol channel examine		Satellite
6	Control variable compute		Software process
7	Instruction inject		Satellite
8	Task implement		Software process
9	Procedure surveillance		Software process
10	Orbit calculate	Task evaluation	Software process
11	Control effect analysis		Software process

Table 48.3 Data relay task analysis

Num	Require of task control	Phase	Operation instance
1	Orbit prediction	Task layout	Software process
2	Data relay task application		Software process
3	Station equipment prepared	Task running	Station equipment
4	Data relay equipment prepared		Station equipment
5	Bidirectional capture control		Satellite
6	Telecontrol channel examine		Satellite
7	Instruction inject		Software process
8	Customer data relay		Station equipment
9	Customer data parse		Software process
10	Procedure surveillance		Software process
11	Task implement		Software process
12	Information exchange	Other	Software process

48.4.2 Maintenance Requirement and Product Format

According to the function of all components, we divide task requirements into sixteen parts, get the maintenance requirement of the software, and design the product modes for every function. See Table 48.4.

Through combining the above software products, we get the special software product line of the TT&C center, and make the reuse rate of the software greatly improved. See Table 48.5.

Table 48.4 Maintenance requirement and product format

Num	Component	Require of task control	Maintenance requirement	Product mode
1	Compute component	Orbit prediction	Adding satellite symbol	Argument configuration file
2		Orbit calculate		
3		Control variable compute	Adding satellite constants and control requirement	
4		Control effect analysis		
5	Satellite control component	Telecontrol channel examine	Adding satellite instruction sequence and telemetering description	Satellite telecommand script
6		Instruction inject		
7	Equipment control component	Station equipment prepared	Adding control command set	Equipment control command
8		Bidirectional capture control		
9		Data relay equipment prepared		
10		Customer data relay		
11	Customer data relay component	Information exchange	Adding network configuration	Route configuration file
12	Customer data parse component	Customer data parse	Adding customer data format description	Data parse configuration file
13	Task layout component	TT&C Task application	Adding satellite symbol	Argument configuration file
14		Data relay task application		
15	Task schedule component	Task implement	Adding task procedure	Task schedule script
16	Status surveillance component	Procedure surveillance	Adding display pages	Page configuration files

Table 48.5 Summarization of software reuse

Task type	Maintenance type	Reuse rate (%)
The sequence kind one of current spaceflight	Configuration files	95
New spaceflight, but task requirements is similar	Configuration files + special algorithm	80
New spaceflight, but task requirements is different	Configuration files + special algorithm + new component	70

48.5 Conclusion

According to theories of software product line and support framework, software requirement of the TT&C domain is analyzed, including component sort, common requirement and variant requirement. Product line architecture is designed for TT&C Software support framework. From three views of development structure, layer structure, and running structure, Systematic structure is described in detail to achieve a software system based on the support framework. Aiming at adding a new satellite into TT&C network, task requirements are analyzed, maintenance requirement and product format of software system are demonstrated, and the reuse rate of software systems is simply summarized.

References

1. Hao Y (2004) TT&C network. China national defense industry press, Beijing
2. Wang YH (2009) Software component and architecture. China machine press, Beijing
3. Zhang W, Mei H (2003) A feature-oriented domain model and its modeling process. J Softw 14 (8):1345–1356
4. Bachmann F, Bass L, Buhman C et al (2000) Volume II: technical concepts of component-based software engineering. In: Technical report, 2nd edn CMU/SEI-2000-TR-008, Software Engineering Institute, Carnegie Mellon University, Pittsburgh
5. Mei H, Chen F et al (2003) ABC: an architecture based, component oriented approach to software development. J Softw 14(4):721–732
6. Clements P, Northrop L (2002) Software product line: practices and patterns. Addison-Wesley Professional, Boston
7. Bosch J (2000) Design and use of software architectures: adopting and evolving a product-line approach. Addison-Wesley Professional, Reading
8. Liu GL, Wei J (2010) Container product line architecture based on component model analysis. J Softw 21(1)

Chapter 49
Research on Cloud Computing Technology Serving Space TT&C Applications

Bowei Zhang, Li Su, Yanmeng Sun and Manhong Lu

Abstract Technologies used in space TT&C are becoming varied and complex, in which case it requires developing a novel service environment with scale-out ability, general access interfaces and multiple technologies adaption to serve processes of information manufacture and information analysis in better quality. Cloud computing utilizes a model based on shared resources composing of mass storage devices, computing equipments, software and networks, from which the service can be flexibly obtained according to the needs of tasks. In this research, we propose a method design of the space TT&C cloud service platform (STCSP) based on cloud computing technology. Up on the analysis of design purpose and service modes, the key technologies of space TT&C cloud service are highlighted. Then the system framework of STCSP is described, as well as proposed deployment strategies and contained functions. At last, three study cases are given to illustrate how STCSP serves space TT&C applications.

Keywords Cloud computing · Space TT&C cloud service · System framework of STCSP · Deployment strategies of STCSP · TT&C baseband cloud · TT&C analysis cloud · Space environment sensing cloud

49.1 Introduction

Operated for measuring spacecraft and delivered TT&C data processing, the space TT&C system plays one of the most important roles in space tasks [1]. Space TT&C has the characteristics of various tasks, mass acquisition information, multiple data types and real time reaction. Other than upgrading standalone machines,

---

B. Zhang (✉) · L. Su · Y. Sun · M. Lu
Beijing Research Institute of Telemetry, Beijing 100076, China
e-mail: antony01@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_49

building TT&C system in future with interconnected units, flexible configuration methods, as well as powerful processing ability against lower cost might be a reasonable solution. Cloud computing is the third revolution of IT after the inventions of PC and Internet. The cloud computing technology provides new solutions to the challenges in the 21st century: data intensity, computing intensity, concurrent access intensity and spatiotemporal intensity. The development of cloud computing also provides new opportunities for the popularization of TT&C and business development.

49.2 The Background of Cloud Computing and Development in Aerospace Field

49.2.1 The Background of Cloud Computing

Up to now there has been yet no unified standard to define cloud computing. The concept of cloud computing derives from the cluster, grid and effectiveness computing [2]. The cluster and grid computing solve large tasks by running an amount of computers in parallel, while the effectiveness computing and SaaS (Software as Service) serve users by supplying resource. As the new revolution of IT, cloud computing is always defined and explained as:

1. As a type of resource scheduling model: cloud computing gathers virtual resource (like virtualized hardware, development platform or I/O services) in quantity, then schedules them with reconfiguration dynamically according to loads in order to maximum the usage rate;
2. As a type of computing model: cloud computing distributes tasks on resource pool of large number of computers, from what applications can obtain computing ability, storage space and other info services according to their actual needs;
3. As capability of service: cloud computing supplies users with IT services, due to what users can be finely served via Internet and be free from the effort to know detail techniques and knowledge or to operate real devices.

Therefore, cloud computing is in essence a resource sharing service model, called cloud computing services or cloud services. Upon this model business applications can be served via the Internet while the computing devices or other info equipments are shared among different users. Moreover, without resource left unused the overall cost decreases [3]. By means of cloud computing technology, resources (like computing, storage, network, data or software) can be formed into a unified resource pool and packaged into measurable services similar to public facilities (like water supply, power supply, etc.).

49.2.2 Development in Aerospace Field

NASA is the first aerospace agency who used the cloud platform technology for serving the informatization work and scientific tasks. In year of 2008, the Ames Research Center of NASA firstly proposed an informatization cloud platform called Nebula [4]. The Nebula was built upon OpenStack, an open source technology or a platform, which has become a seamless and self-ruled service platform during recent decades via integrating serials of open source management software. This platform was provided with powerful capacities of computing, storage and network. The platform was constructed to ensure the security of IT framework, compensation with NASA's institutes and USA federal policies as a prerequisite. The Nebula was designed in bottom-up method while visualization and scalable framework are utilized for reaching economics and energy efficiency.

In year of 2010, Nebula helped NASA in tasks of exploring Luna and Mars by being charged with recording and processing large amount of high resolution pictures and videos [4]. In traditional solutions, the construction of the IT infrastructures with configuration is estimated to spend nearly 150–180 days. By using Nebula instead of expensive typical data center, it was no longer necessary for NASA to build more IT infrastructures for additional big data. Upon this, researchers in NASA could fix all tasks of building, configuring, monitoring and system upgrading in a few minutes. Overall, Nebula provided framework, source code library, interfaces inside outside and network, those supported all Nebula services. Moreover, data could be delivered into sub institutes of NASA from Nebula via safe channel that got rid of massive accesses to internal network. In NASA Mars projects of "Spirit MER-A" and "Opportunity MER-B", Nebula performed outstanding on its data processing efficiency with reduction of the cost.

Besides, NASA extended Nebula to many other scientific tasks for further metrics for establishing advanced task-level enterprise data center [5]. Some of successful cases (shown in Fig. 49.1) included:

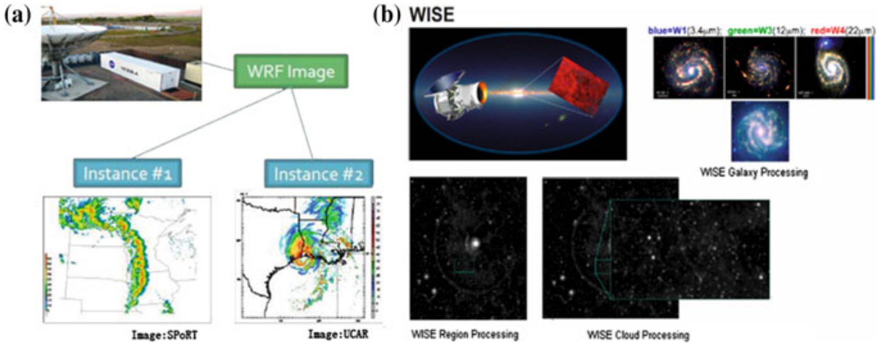


Fig. 49.1 Example: Nebula servers NASA projects. **a** SERVIR Project. **b** WISE Project [4]

1. serving SERVIR.net (for the earth observation science work and development) [4], SPoRT (Short-term Prediction Research and Transition) and serving WRF (Weather Research and Forecasting) [4] by rapid processing work of combining great numbers of the high-resolution pictures for weather forecasting without interrupting other business routine.
2. serving WISE (Wide Field infrared Survey Explorer) [4] project by improving image processing ability and providing massive data storage.
3. serving Ames center itself in the task of aerodynamic calculation data analysis instead of expensive Pleiades devices.
4. serving public organization customers by providing big data storage and management.
5. serving software programmer from all over the world by providing developing platforms with virtual workstation for compiling then testing Nebula's source code. In addition, Nebula continued to share its modules and libraries with these volunteer developers.

49.3 TT&C Cloud Service Technology

49.3.1 Concept

Normally, CORBA, DCOM, RMx and webservice based on UDDI are considered when we talk about service technologies. However, such technologies are of static implementation. The space TT&C cloud services are the application services provided through a network as a way of on-demand sharing of integrated TT&C devices resources, software and data center based on cloud computing. The space TT&C cloud services should reflect the advantages of cloud computing in aspects of scale-out, virtualization and HPC. One is to improve efficiency in building system by means of rapid configuration based on cloud computing; the other is to benefit TT&C service aggressively by means of its overall reliability, efficiency and flexibility due to cloud platform's natural distributed structure and virtual levels. Moreover, the cloud computing model is designed to manage mass data in terms of its powerful capability in high-parallel storage, real-time processing, long-term recording, scientific management and rapid indexing since the model can be easily combined with HPC clusters or interconnected desktops.

In delivered service hierarchy above, the TT&C devices, data center (general devices for storage or computation), system software as well as application software are key components. Way to design such components should differ from normal approaches, thus following new features should be considered:

TT&C devices. Originally, the TT&C platforms are made of special digital signal processing devices like DSPs or FPGAs. Such devices drive multiple TT&C systems or functions by reconfiguration. In proposed space TT&C cloud service hierarchy, original a standalone machine (like a TT&C baseband, with devices like DSPs and FPGAs) should play an additional role of the server machine, which is supposed to be equipped with general purpose processors or powerful embedded processors. Such a machine or called a main board enhances the ability to access open platforms freely by supplying more general I/Os, high-speed interconnect ports and general software. In this case, moreover, one can reduce cost per-unit by re-organizing the function modules on the proposed TT&C machine. For instance, a single main board can be issued to sever multiple TT&C baseband function sub-boards simultaneously. Furthermore, widely used APIs and open platform access protocols should be installed on such TT&C platform in order to make applications compatible with other virtual access.

Data center. Traditional data centers are always made of large numbers of interconnected nodes of servers. In mind of cloud service, computation tasks should be always moved to such data centers from desktops (including moves of data processing, storage position and application services). In our TT&C cloud service hierarchy, the data center is one of the most important components and it should be built in awareness of the specific performance requirements of the TT&C applications. To metric such data center, designers should pay attention to features on real-time performance, data throughput, bandwidth between nodes, capacity of storage, capacity of disaster recovery and especially the maintains of mass data task routines. Furthermore, the middle and small size data centers should be developed for custom use as well as the PnP feature.

System software. The DFS (Distributed File System) and open platform accessible interfaces should be employed.

Application software. Designing of data-intensive, storage-intensive and I/O-intensive software applications should be under the model of the data center in order to maximum the efficiency of computation resource. In summary, way to design and use the software should be authorized by the principle of quality assurance.

49.3.2 Service Modes

The TT&C cloud service should support service models as listed:

TT&C data analysis as a service (TDAaaS). TDAaaS provides data analysis results based on business requirements and modularized analysis models. Users can review and download the results, thus get meaningful information without producing it themselves.

TT&C data as a service (TDaaS). TDaaS provides users with TT&C data itself and related operating approaches. There is no need for users to purchase additional devices for producing and storing the data. Data operations as indexing, exploring, querying and reviewing are also involved in this kind of service.

TT&C processing software as a service (TPSaaS). TPSaaS provides configurable virtual terminals in terms of platforms like virtual machines. On such virtual machines, compiled software tools, the database and network kits are pre-installed. Users namely clients can finish all their online TT&C workflows while be free from the annoying steps of building the work environment.

TT&C platform as a service (TT&CPaaS). TT&CPaaS provides users who are willing to develop their own specific applications with development platforms. Such platforms are equipped with drivers, SDK and API libraries to support custom workflow's developments on delivered low-level backend TT&C tasks.

TT&C application infrastructure as a service (TT&CAIaaS). Virtualization is one of the most important features of cloud computing. TT&CAIaaS provides users with the access to rent or schedule virtualized elements (like TT&C hardware devices, resources of the data center) for custom usage or re-development within the limitation of administrations. Users can also build their own private cloud service platforms based on the open platform framework in this service model.

49.3.3 Key Technologies

49.3.3.1 Cloud Network Technology

In the TT&C cloud service driven by the data center, data processing is supposed to be run on clusters, either real or virtual. The key features of the cloud service interconnection technology, namely the cloud network technology, should involve low-delay data transportation, high bandwidth, MPI (Massage-Passing Interface) supported, fault tolerance supported, DFS accessible as well as P2P and group communication supported. Moreover, maintaining good flexibility is another important consideration in building cloud network since the network topology might be reconfigured according to needs of expanding clusters by adding more server nodes. Data should be able to be shifted between nodes for balancing overall payloads of the clusters. Approaches to build the data center with flexible network topology could be found in reference [6].

49.3.3.2 Cloud Storage Technology

Cloud storage organizes a large variety of storage devices in a storage resources pool to provide unified dynamic scaling storage service. By means of large file chunks, distributed storage, and multiple copies, the cloud storage can automatically schedule required data and storage resources on the user's demand and ensure the reliability of data and the efficiency of access by redundant storage.

As the amount of TT&C data is huge and is still being increased rapidly with the operation of satellites and service platforms, cloud storage with unparalleled scalability and devices reuse can meet scalability requirements of data growing, and reduce investment for new equipments. At present, the DFS of the cloud computing platform based on stream bytes access can meets most requirements of TT&C data tasks. However, specific storage strategies should be considered for multi-media data or non-structure data. Such strategies include the utilization of the non-structure database, for example the HBase for DFS special which is sponsored by the Apache corp.

49.3.3.3 Cloud Processing Technology

The TT&C cloud processing technology utilizes high performance, high scalability, and high availability cloud computing technology as well as large-scale multi-core programming, multi-threading technology and high throughput technology to realize the high-speed processing of massive TT&C task data based on distributed storage and parallel computing model. In the other hand, it is necessary to follow the uniform parallel framework or the distributed principle (like MapReduce method) in designing parallel programs. Such kind of architectures provides programmers with abstract interfaces by means of utilizing dataflow transparence technologies, in which situation the programming efficiency can be maximized against an even shorter TTM (Time to Market);

Particularly, other than original HPC clusters or grid computing, the cloud processing technology has the feature of a loosely-coupled relationship between the hardware and real tasks, since resources are all shared and scheduled in the pool dynamically. In this case, the technology of virtual clusters is necessary. In the other words, the virtual clusters are composed of many virtual client terminals/nodes which are located on one or several distributed servers of real clusters.

49.3.3.4 Security Technology

Information security is one of the key problems affecting wide adoption of cloud computing technology and this problem determines whether users are willing to store their own data on the TT&C cloud service platform. High-precision TT&C data and business information are highly confidential. Ensuring information security is crucial to the practical application of TT&C cloud service platform and the TT&C cloud services model [7].

The security strategies that the TT&C cloud service should employ includes: (a) physic isolation: for preventing leak of the electromagnetic wave carried business information, or EM/RF hijacking; (b) data encryption during the whole life-cycle:

limited authorization of information decoding and understanding; (c) utilization of new generation technologies as firewall, IDSee and PIK: preventing the risk of leaking by fault or manual invasion; (d) principle of authorization and certification: data and application services only certificated by central management unit; (e) building different security grades for keeping classified data safe in service routine.

49.4 Space TT&C Cloud Service Platform

With providing TT&C data service as the core function, the space TT&C cloud service platform can be established by integrating TT&C devices, computation infrastructures, TT&C data storage, TT&C data management, analysis software, algorithm modules as called library, business model, and dynamic expanded open interfaces.

49.4.1 Cloud Service Platform Architecture

As shown in Fig. 49.2, the TT&C cloud service platform has a four-tier framework. The resource tier is the foundation and integrated resources of the system; in the function tier, the basic functions of the system are implemented; in the service tier, the particular service model is realized with combining the system function following business requirements; in the transaction tier, the business application procedures are provided. The function tier and service tier are built up on the cloud

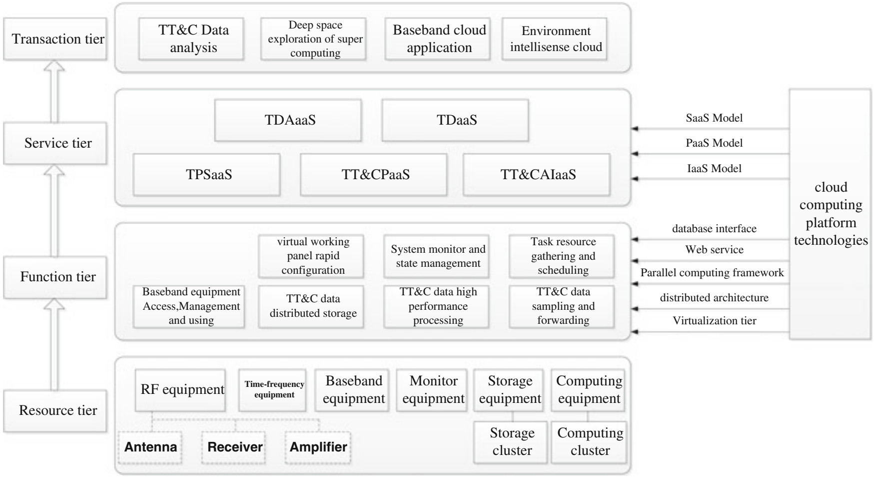


Fig. 49.2 Space TT&C cloud service platform (STCSP) structure

computing platform technologies including virtualization, distributed architecture, parallel computing framework, Web service and database interface. With above the platform can provides functions of device virtualized management, TT&C data distributed storage, TT&C data high performance processing, virtual working panel rapid configuration, library of algorithm modules, resources pool management, resources dynamic scheduling as well as user management with certification rules, data encryption, massive user access supported, payload balancing and failure recovering. Illustrated by Fig. 49.2, proposed service modes the platform runs in are derived from three of the modes described in Sect. 49.3.2.

The TT&C cloud service based on proposed platform should have features as listed:

1. The four-tier architecture of proposed platform plays either as a single system, or separated abstract layers. For the usage of upper layers does not depend on particular implementations of lower layers. The architecture has flexible interfaces and scale-out ability.
2. The service provides particular TT&C applications with integrating operation procedure, TT&C data, and software with device and computing resources it needs.
3. The service supports data sharing and multiple virtual terminal co-works. The virtual working panel can be rapidly configured with special resources and software. It should be able to dynamically create virtual machines, deploy and update of remote sensing data and software, realize use work point reservations, data sharing and collaboration of different users.
4. The platform allows TT&C service elements (sampling, storage, processing algorithm, application models, etc.) to be accessed in and be dynamically scheduled, including collaborated utilization of data, information and other technical resource as well as providing ability to organize the workflow, to implement cooperating services of information chain, technology chain and industry chain.

49.4.2 Cloud Service Platform Deployment Strategy

Figure 49.3 shows a typical deployment strategy of the STCSP which consists of three parts: regional cloud, function private cloud and hybrid cloud.

The regional cloud is an open platform that can be accessed in a controllable range, and it should be signed in by way of Web2.0 or the virtual working panel built and authorized by administrators.

The function private cloud is a platform that specifies services of configuring TT&C devices (like time-frequency device or baseband device), supporting real

TT&C tasks execution by providing high-level encrypted access authorization. Hence, the function private cloud should be kept away from the public region and be maintained within a LAN.

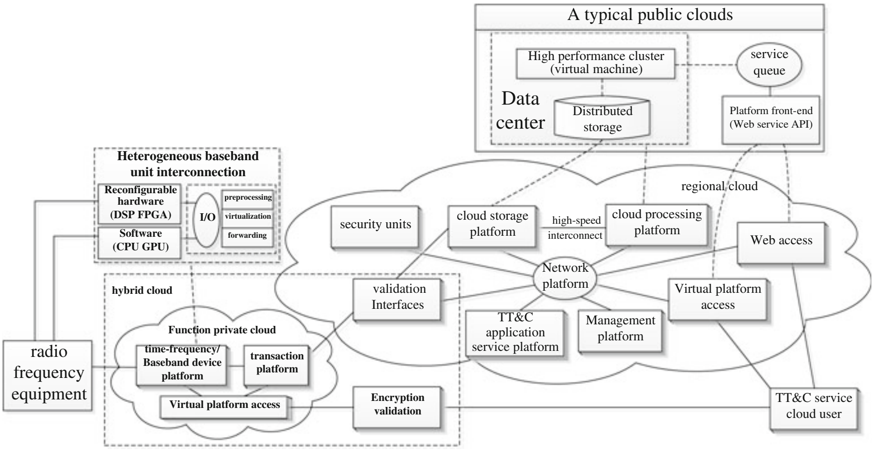


Fig. 49.3 STCSP deployment method

The hybrid cloud in fact is a conjunction of the above two. Interfaces with validation are utilized for communication between the regional and the private platform. The content of the vehicles on such interfaces can be files of raw data, intermediate data or the result information.

In the regional cloud, components are connected by the switching network. The key components should be the cloud storage platform, the cloud processing platform, the management platform, TT&C application service platform and security units. The union of the storage cloud and processing cloud presents a typical public cloud deployed on the data center. Data is transmitted and received on high-speed channel between computation nodes and storage nodes inside the data center clusters or virtual clusters.

The management platform should provide approaches that schedule tasks for real computing, data loading and storing operations while a service queue is controlled at the same time. The management platform also provides visual monitoring service of status report, behavior control and command control by way of Gb Ethernet network.

The TT&C application service platform are working on application software servers those should be pre-installed with commercial CAD tools, TT&C task special software equipped with typical TT&C algorithm libraries and user libraries.

The security components are provides in hardware and software. The hardware components includes: hardware firewall, secret-key and biological feature recognition sensors while the software includes the third generation network defense tools and other cloud security components. A novel strategy in assumption is that by utilizing many trusted terminals and the abilities of analysis those have, all data generated by users' behavior can be monitored, recorded and analyzed with sample comparison. The analysis result then is gathered and fed back to management platform that becomes objects for detecting Trojan virus, malicious programs and illegal intrusion.

The function private cloud specifies components of TT&C device platform, data forwarding platform and special virtual platform accessed via encrypted link. The TT&C device platform can be either standalone time-frequency/baseband equipment or the network that manages such equipments. According to the situation that the software radio technology and large scale multi-core processor array (like GPU for HPC or general purpose coprocessor based on multi-core micro-architecture) develops rapidly, issuing a software based TT&C device compatible model should have been taken in consideration before the platform I/O is proposed in future. Particularly, the deployment of the function private cloud should provide users with some option like shifting some of HPC processors into private domain form the regional cloud with reconfiguring flexible interfaces between the two. Such option will help in case that the user application has much greater data scale fan-in than fan-out. In other words, overloading can be avoided when transporting data between the private domain and the regional domain.

49.4.3 Cloud Service Platform Functions

According to the cloud service platform architecture and deployment strategy, the basic functions of the system are as follows:

1. TT&C devices access, management, virtualization and usage;
2. TT&C data sampling and forwarding;
3. High performance data processing on parallel architecture;
4. Distributed data storage;
5. Data indexing, context replaying and reproducing;
6. Called libraries of TT&C core algorithms;
7. Testing software developing, simulating, and debugging;
8. Rapid configuration of virtual working panel;
9. System monitoring, controlling, and status reporting;
10. Tasks and resources gathering and scheduling;
11. Security defense, data encryption and authorization;
12. IT information services.

49.4.4 Case Study

49.4.4.1 TT&C Baseband Cloud

It is a trend in future that the TT&C system will enhance the standardization currency and agility of the TT&C baseband instrument and processing platform as the space mission is become increasingly complicated and diversified. A GPP (General Purpose Processor) software baseband that runs on cloud platform has the

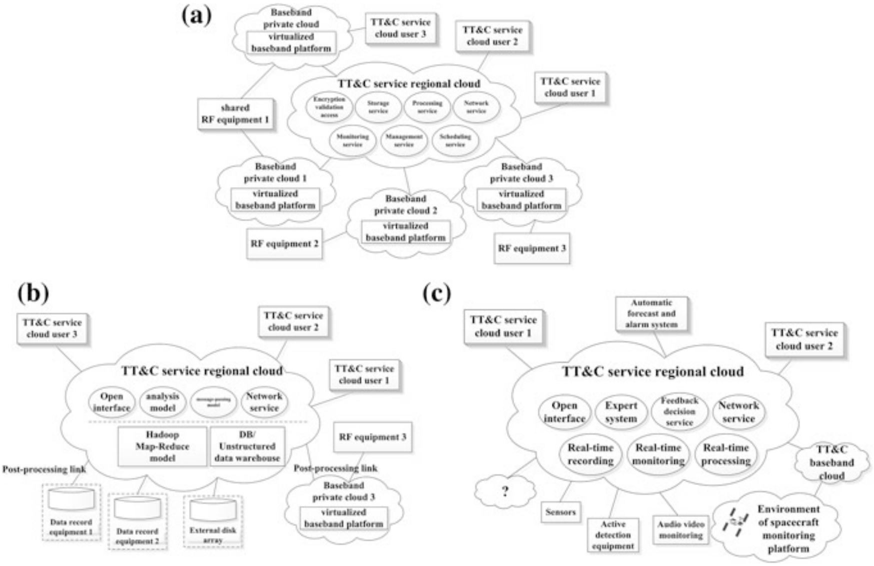


Fig. 49.4 Study cases of space TT&C cloud service. **a** TT&C Baseband Cloud. **b** TT&C Analysis Cloud. **c** Space Environment Intelligence Cloud

advantages of expandability, virtualization and high performance. Meanwhile cloud computing technology adapts for future TT&C system more easily in the aspects of comprehensive function, equipment generalization, interface standardization, equipment remote management and operation transparency.

TT&C baseband cloud (shown in Fig. 49.4a) consists of TT&C service regional cloud and one or more baseband private cloud. Different baseband private cloud can be in same level or in the cloud of local overlay to achieve load balancing. The cloud provides services for scheduling gathered time-frequency equipments at the bottom of the private cloud. Homogeneous or heterogeneous architecture GPPs are utilized to realize the baseband capabilities. Such GPPs are recognized as virtualized basebands via abstract interfaces. Share the input and output of the RF equipment through extending or forwarding between private clouds. This way need virtual baseband equipment has strong general processing ability to deal with the I/O with ultra high data rate transmission. The cloud provides TDaaS and TPSaaS to local or web users, who can obtain data and software data service by way of network access.

49.4.4.2 TT&C Data Analysis Cloud

Space missions are great businesses for human to explore space. Big data is always produced during the process of space missions. The cloud platform can be the location of data storage and the way to access. We can rely on this environment of cloud platform and its ability of cloud computing to mine an effective and meaningful data set, which is suit to specific task. Like the spatial data, the amount of space big data is more than normal social application result in higher processing request. In addition, for tasks of remote controlling spacecraft, the system should process the data very rapidly during some TT&C task. Hence, the TT&C data analysis platform should not only has the characteristics of effective big data analyzing, but also have high reliability.

TT&C data analysis cloud (shown in Fig. 49.4b) consists of links between the regional cloud and data terminals, including recording equipments and external data raid. Data terminals are intelligent equipments as well as extended resources disk storage tank. Such storage can be accessed by way of DAS, NAS or SAN [8]. Inside the regional cloud, data is stored in a distributed architecture based on a switched network. The regional cloud is equipped with distributed framework like Hadoop [9] MapReduce, or other distributed computing [10, 11] for large-scale data sets. The platform should support functions of loading modules with application analysis model automatically. The analysis cloud obtains real-time results of task data processing and responds rapidly via the real-time transmission pathway.

49.4.4.3 Space Environment Intelligence Cloud

The basis of modern information is technologies of information collection, transmission and processing. The intelligence cloud measures, monitors and records possible damage and failure of space environment effectively with sensors. It processes and manages the data efficiently, scientifically and reliably. These methods help to guarantee the quality of space environment detection and improvement of regional space environment. The space environment intelligence cloud generally consists of perception layer, network layer and transaction layer, which are based on event-driven architecture. The platform is built on network, also has friendly interface and is adaptive to sensing devices and applications.

The space environment intelligence cloud (shown in Fig. 49.4c) is built with regional cloud and many sensing platforms or equipments. Equipments of the sensing platform are set for long-term controlling. The task data is big and geographically sporadic. It provides significant advantage by utilizing distributed architecture to bearing large amount of task data in the aspect of efficiency and cost. The TT&C cloud provides the support of real-time recording, processing, monitoring, expert judgment and decision-making based on the real-time returned sensor information via virtual storage pool and computing pool. The cloud provides application-oriented analysis services and software toolbox. Such applications are specially focus on measuring environmental data. Moreover, the space environment intelligence cloud can be integrated with environmental monitoring platform of spacecraft though baseband cloud. It can be even connected with some Internet of Things to share information.

49.5 Conclusion

Cloud computing technology is supposed to change modes of the TT&C data storage, processing technology and service profoundly, so as to bring significant opportunities of technology innovation and industrialization development [12, 13]. The technology based on cloud computing can package the data, processing technology and the software hardware resources as controllable services. This solution will address the bottleneck among data, technology, equipment, cost and personnel in TT&C application to achieve efficient use of cloud computing services technology. The proposed cloud service platform is not only the integration between information technology and new service mode, but also providing technical innovation platform. It will derive more operation modes and application direction with the space characteristics and age characteristics on this platform.

At present, there is no unified standard about construction of cloud service platform which covers the theory, technology and service [14–16]. Moreover, cloud platform is not mature enough and the market is not a monopoly. Therefore it is conducive to the design and promotion of domestic information cloud platform solutions. Furthermore, the standard of cloud computing has not yet been issued. There is not technical monopoly, and it is a good opportunity for us to develop and utilize space cloud platform in real space TT&C tasks. The TT&C information application technique is an important part of space technology. Tracking application demand and the developing trend of space, developing TT&C application technology and service model based on cloud technology will not only meet the application need, which is brought by the age of the cloud, but also promote the construction of cloud platform standardization system based on service.

References

1. Shen RJ, Zhao J (2001) Development tendency and strategy of space TT&C technology of our country. J Astronaut 22(3):1–5
2. Hwang K, Fox GC, Dongarra JJ et al (2013) Distributed and cloud computing: from parallel processing to the internet of things. China Machine Press, Beijing, p 132
3. Chen K, Zheng WM (2009) Cloud computing: system instances and current research. J Softw 20(5):1337–1348
4. Williams J (2012) NASA's Nebula cloud computing initiative, cloud innovation at NASA, NASA Ames Research Center
5. Hu XY (2010) Cloud computing and space exploration discussion. Satell Netw 12:18–23
6. Al-Fares M, Loukissas A, Vahdat A (2008) A scalable, commodity datacenter network architecture. In: Proceedings of the ACM SIGCOMM 2008 conference on data communication, Seattle, WA
7. Feng DG, Zhang M, Zhang Y et al (2011) Study on cloud computing security. J Softw 22 (1):77–83
8. Zhang D (2011) Big storage II—analysis of storage system architecture and the underlying principle limitation. Tsinghua University Press, Beijing, p 253
9. White T (2012) Hadoop: the definitive guide, 3rd edn. O'Reilly, California

10. Reed D (2008) Clouds, clusters and many core: the revolution ahead. In: Proceedings of the 2008 IEEE international conference on cluster computing
11. Foster I, Zhao Y, Raicu Y et al (2008) Cloud computing and grid computing 360-degree compared. In: Grid computing environments workshop, pp 1–10
12. Yuan B, Zhang M (2010) Ideasfor applications of aerospace TT&C cloud computing. Sci Technol Innov Herald 17:13
13. Ren FH, Wang JN (2012) Turning remote sensing to cloud services: technical research and experiment. J Remote Sens 16(6)
14. Velte T, Velite A, Elsenpeter R (2010) Cloud computing, a practical approach. McGraw-Hill Osborne Media, New York
15. Cloud Security Alliance (2009) Security guidance for critical areas of focus in cloud computing
16. Gonzalez LMV, Merino LR, Caceres J et al (2009) A break in the clouds: towards a cloud definition. Comput Commun Rev (CCR) 39(1):50–55

Chapter 50
The Research on Servo DC Motor Fault Diagnosis with LVQ Neural Network Theory

Qing Sun, Guangping Liu, Demin Qiu, Qi Zhang and Jiaqing Xi

Abstract According to the analysis on the fault diagnosis technology of Servo DC motor in aerospace measurement and control, LVQ Neural Network is studied in this paper. The ability of feature extraction and pattern recognition is the characteristics of LVQ neural network, and it is applied to analyze current signals in different branches. We proposed a servo DC motor fault diagnosis model based on LVQ neural network theory, determined the input, output, network structure and inference of the model, and designed a DC motor fault diagnosis system to identify the failure mode in the fault diagnosis of DC motor. In order to demonstrate the superiority of LVQ neural network, a classical BP neural network has been developed to solve the same problem for comparison. The simulation results shows that the LVQ neural network is effective and superiority in fault location.

Keywords Fault diagnosis · LVQ neural network · Fault mode identify · Servo DC motor

50.1 Introduction

In the aero space measurement and control devices, servo DC motors play an important role as it is the actuators in the antenna control system. Once a failure occurs in the servo DC motors, it will seriously delay the implementation of the monitoring and control tasks. However, due to the implementation of the long tube heavier task and its automated operation, the continuous operation time of monitoring and control equipment is greatly increased. This means it will bring the heavier burden of the monitoring and control equipment, and reduces the time for equipment maintenance and inspection. Therefore, it is needed to design a servo DC

Q. Sun (✉) · G. Liu · D. Qiu · Q. Zhang · J. Xi
Qingdao Station, Xi'an Satellite Control Center, Qingdao 266114, China
e-mail: sq2006@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_50

motor fault diagnosis system, which could effectively monitoring and diagnosing the fault of servo DC motor on-line.

In fault diagnosis, the fault pattern recognition is the key point to achieve on-line fault monitoring and intelligent diagnosis, which is also the core of fault diagnosis system. The recognition efficiency and accuracy will directly affect the performance of the fault diagnosis system. Because the servo DC motor has complex structure, and its fault characteristic parameters have dispersion, randomness and fuzziness characteristics [1–3]. Moreover, there is no clear linear relationship between failure modes and failure characteristic parameters; therefore we cannot establish an accurate mathematical model. To this end, there is a bottleneck to get the complete knowledge for building an effective diagnostic model for servo DC motor pattern recognition [4]. To solve this problem, BP neural network is introduced to solve this problem [5]. However, the gradient descent method is used to minimize the error in BP neural network, which maybe convergency slowly and not meet the requirements of precision [6]. Besides the gradient descent method maybe converges to local optimal point [7, 8]. Therefore, we introduce the LVQ neural network to solve the fault pattern recognition problems here. Firstly, we train the LVQ neural network with collected data, and then realize the automatically elicitation of diagnostic knowledge. Finally, the trained LVQ neural network model are used for automatically fault diagnosis.

50.2 LVQ Neural Network

The full name of LVQ neural network is Learning Vector Quantization neural network, which is a supervised learning algorithm with training in competitive layer, the transfer function can be used in the hidden layer, depending on the target's class will be the input vector combined classification, so LVQ algorithm can be considered to be the self-organizing map algorithm to improve teachers learning algorithm [9–11].

The basic idea of this algorithm is to calculate the distance of the input vector layer neurons recent competition to find the linear output layer neurons connected If the input vector classes and linear output layer neurons consistent with each other, the corresponding competition layer neuron input weights moved along the direction of the vector; Conversely, the corresponding neuron weights of competitive layer moves in the opposite direction along the input vector [12].

LVQ neural network has strong robustness, memory capacity, nonlinear mapping and self-learning ability. Its learning rules are simple, and easy to realize in the computer. Therefore, it is the idea model to solve the fault diagnosis of servo DC motor fault pattern recognition problem.

50.3 Servo DC Motor Fault Diagnosis Model Based on LVQ Neural Network

50.3.1 Description of the Model

50.3.1.1 Input of the Model

In the actual operation of the servo DC motor, it has many features which may characterize the dynamic behavior of the system. However, there is no need to select all of them as the input feature vectors. To meet the condition of real-time fault diagnosis and lower computation complexity of the servo DC motor, and the results of the literature analysis, we only use the features: the current the pulse frequency f_w, the mean steady-state current i_{av}, the standard differential steady state current i_{std}, the starting current peak i_m, the rate of current peak change k. Therefore, the characteristic parameters for the fault diagnosis are $T_e = [f_w, i_{av}, i_{std}, i_m, k]$, where corresponds input vector machine fault diagnosis for the corresponding LVQ neural network $X = [x_1, x_2, x_3, x_4, x_5]^T$.

Here, the x_1 represents the current the pulse frequency f_w, x_2 represents the mean steady-state current i_{av}, x_3 represents the standard differential steady state current i_{std}, x_4 represents the starting current peak i_m, x_5 represents the rate of current peak change k_o.

50.3.1.2 Output of the Model

Servo DC motor fault types mainly include brush fault, element opens, and shorts between turns. Therefore, the corresponding servo motor failure modes are: no failure, brush failure, component open, and inter-turn short circuit. Four states are labeled as 1, 2, 3, 4 respectively. And the output vector is $C = [c_1, c_2, c_3, c_4]^T$. Therefore, the associated fault status code is no fault (1, 0, 0, 0), the wiper fault (0, 1, 0, 0), components open (0, 0, 1, 0), and inter-turn short circuit (0, 0, 0, 1).

To this end, the problem is converted into the problem that dividing the five-dimensional Euclidean space into a four-dimensional decision space.

50.3.1.3 Network Structure of the Model

The model structure is shown in Fig. 50.1, which contains the layer characteristic parameters, the competitive layer, and the failure mode layer. Here, the number of neurons in characteristic parameters layer is 5, which corresponding to five input variables. The number of neurons in failure mode layer is 4, which corresponding to the four output variables. The number of neurons in competitive layer is 8, each of them corresponding to one of two neurons fault state. There is a fully connected neural network layer and the characteristic parameters of the competition between

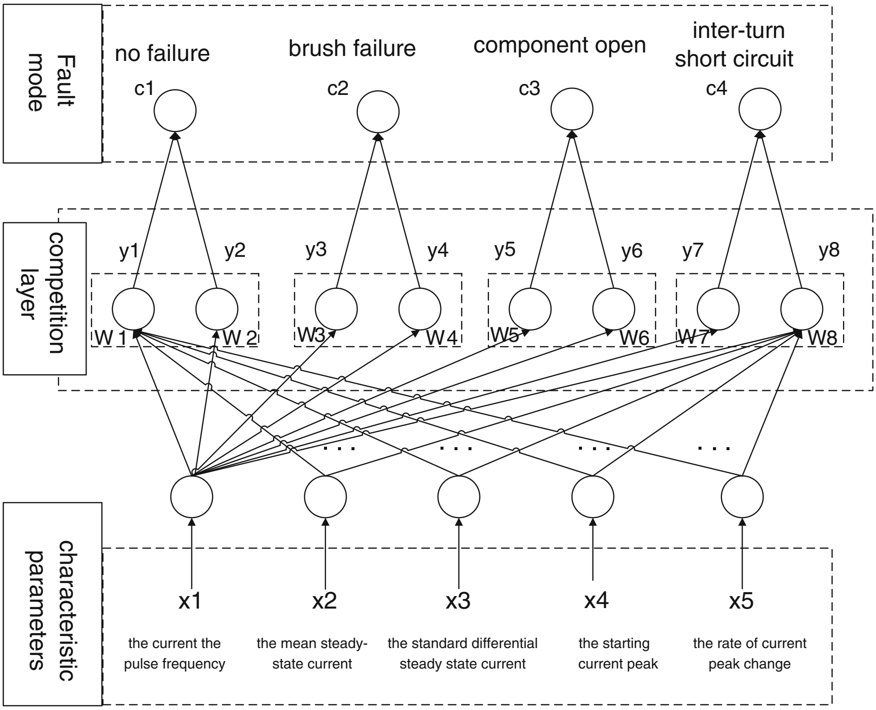


Fig. 50.1 Servo DC motor fault diagnosis model based on LVQ neural network

the layer in which the connection weights represents the *i*-th vector of neurons compete with the five characteristic parameters, and the failure mode in the competitive layer is partially connected, each output neuron element with different groups of neurons are connected to the competitive layer. Competitive layer connection weights and failure mode layer value is fixed at 1. LVQ network in the training process, the connection weights characteristic parameter layer and competitive layer is gradually adjusted to the cluster center. When a sample is sent to the network, the reference vector closest to the input neuron model competition due to get intimated to win the competition with the output 1, meanwhile the other neurons output is 0. Generating a given output neuron input mode category; each output neuron represents a different category.

50.3.2 Inference of the Model

The inference of LVQ neural network model for servo motor fault diagnosis has following steps:

1. Initialize the connection weights W_j in the parameters layer and competition layer (j = 1, 2,…, m, represents the j-th weights competitive layer neurons

characteristic parameter vector), and determine the initial learning rate $\eta(0)(\eta > 0)$ and the number of training t_m;

2. Put $X = [x_1, x_2, x_3, x_4, x_5]^T$ into parameter layer, and calculate the distance between vector of parameters characteristic and competitive layer neurons based on the Eq. (50.1):

$$d_j = \|X - W_j\| \tag{50.1}$$

3. Select the neuron which has the smallest vector distance in competitive layer. If a_i is the minimal value, then the connected failure mode neurons are labeled as c_i, where the real failure tags are c_x;
4. The actual output is compared with the target output, which is used for adjusting the weights of the winner neurons. Assuming in competitive layer, the winning neuron is the *i*-th neuron:

If that is the correct classification $c_i = c_x$, adjust its weights

$$W_i^{(t+1)} = W_i^t + \eta\left[X - W_i^t\right] \tag{50.2}$$

If the classification is not correct $c_i \neq c_x$, adjust the weights in the opposite direction of the input sample

$$W_i^{(t+1)} = W_i^t - \eta\left[X - W_i^t\right] \tag{50.3}$$

5. Update the learning rate

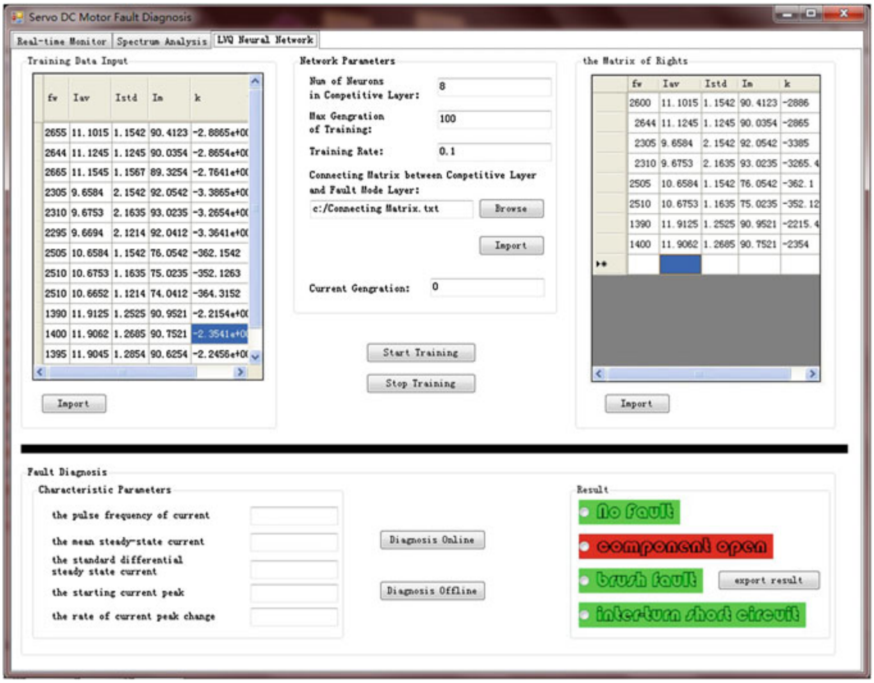
$$\eta(k) = \eta(0)(1 - t/t_m) \tag{50.4}$$

At the time $t < t_m, t = t + 1$, go to step 2, enter the next training samples, and repeat all the steps until $t = t_m$.

50.4 Design and Model Verification of Servo DC Motor Fault Diagnosis System

50.4.1 Interface of the System

The servo DC motor fault diagnosis system is based on LVQ neural network functions, which includes the fault detection and the fault pattern recognition. Moreover, the pattern recognition could be divided into parts: training LVQ neural network, and



Training Data Input

fw	Iav	Istd	Im	k
2655	11.1015	1.1542	90.4123	-2.8865e+0
2644	11.1245	1.1245	90.0354	-2.8654e+0
2665	11.1545	1.1567	89.3254	-2.7641e+0
2305	9.6584	2.1542	92.0542	-3.3865e+0
2310	9.6753	2.1635	93.0235	-3.2654e+0
2295	9.6694	2.1214	92.0412	-3.3641e+0
2505	10.6584	1.1542	76.0542	-362.1542
2510	10.6753	1.1635	75.0235	-352.1263
2510	10.6652	1.1214	74.0412	-364.3152
1390	11.9125	1.2525	90.9521	-2.2154e+0
1400	11.9062	1.2685	90.7521	-2.3541e+0
1395	11.9045	1.2854	90.6254	-2.2456e+0

Network Parameters

Num of Neurons in Competitive Layer: 8

Max Gengration of Training: 100

Training Rate: 0.1

Connecting Matrix between Competitive Layer and Fault Mode Layer: c:/Connecting Matrix.txt

Current Gengration: 0

the Matrix of Rights

fw	Iav	Istd	Im	k
2600	11.1015	1.1542	90.4123	-2886
2644	11.1245	1.1245	90.0354	-2865
2305	9.6584	2.1542	92.0542	-3385
2310	9.6753	2.1635	93.0235	-3265.4
2505	10.6584	1.1542	76.0542	-362.1
2510	10.6753	1.1635	75.0235	-352.12
1390	11.9125	1.2525	90.9521	-2215.4
1400	11.9062	1.2685	90.7521	-2354

Fig. 50.2 The interface of servo DC motor fault diagnosis system

make fully usage of trained neural network for fault diagnosis. The system has two modes: on-line and off-line. Its GUI is designed as shown in Fig. 50.2, which contains two parts: neural network training and fault pattern recognition.

In the neural network training part, there are functions support import the faults identified data, import and automatically generate competitive layer and failure mode layer connection weights, set the number of competing neurons, set the largest training algebra, set the learning rate and etc.

The fault pattern recognition part includes on-line and off-line diagnostic modes, which supports a real-time visual display of diagnostic results and findings exportation function.

50.4.2 Process of the System

1. Data acquisition, collect training data about the current the pulse frequency f_w, the mean steady-state current i_{av}, the standard differential steady state current i_{std}, the starting current peak i_m, the rate of current peak change k, and associated fault status.

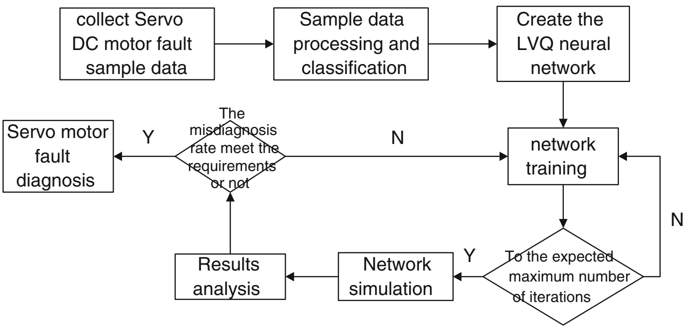


Fig. 50.3 The process of servo DC motor fault diagnosis system

2. Create the network, and determine the number of neurons in the competitive layer, the connection weights and the connection between competition and failure mode layer between the characteristic parameters and the competition layers.
3. Network training, and determine the maximum rate of training and learning algebra of LVQ neural networks, via collected training data from step 1 and neural network model in step 2.
4. Network simulation, when the network reaches a preset maximum number of training iterations, save the network, and input the test set of five quantify characteristics, to get diagnostic results output.
5. Results analysis, analyzing the network simulation results to get the rate of misdiagnosis, which includes misdiagnosis the no fault condition as the fault condition and the misdiagnosis between different no fault conditions. When the misdiagnosis rate higher than a threshold, this algorithm will go back to step 3 and repeat training the network until meet the requirement.
6. Fault diagnosis, using the trained neural network model for fault pattern recognition with both on-line and off-line diagnostics mode. The system operation flow diagram is shown in Fig. 50.3.

50.4.3 Verification of the Model

Here, we get the synthesized servo DC motor operation mode data and fault data from simulation, a total number of 500 sample data and 20 sets of test data are generated, part of the data are shown in Tables 50.1 and 50.2:

We set the network learning rate as 0.1, and the number of training set as 200. The initial right settings between the feature vector layer and the competition layer are shown in Table 50.3. The connection settings between the failure mode layer and the competitive layer are shown in Table 50.4.

Table 50.1 Part of simple data

Group	Characteristic parameters					Fault status
	f_w	i_{av}	i_{std}	i_m	k	
1	2655	11.1015	1.1542	90.4123	−2.8865e +003	No failure
2	2644	11.1245	1.1245	90.0354	−2.8654e +003	No failure
3	2665	11.1545	1.1567	89.3254	−2.7641e +003	No failure
4	2305	9.6584	2.1542	92.0542	−3.3865e +003	Brush failure
5	2310	9.6753	2.1635	93.0235	−3.2654e +003	Brush failure
6	2295	9.6652	2.1214	92.0412	−3.3641e +003	Brush failure
7	2505	10.6584	1.1542	76.0542	−362.1542	Component open
8	2510	10.6753	1.1635	75.0235	−352.1263	Component open
9	2510	10.6652	1.1214	74.0412	−364.3152	Component open
10	1390	11.9125	1.2525	90.9521	−2.2154e +003	Inter-turn short circuit
11	1400	11.9062	1.2685	90.7521	−2.3541e +003	Inter-turn short circuit
12	1395	11.9045	1.2854	90.6254	−2.2456e +003	Inter-turn short circuit

Table 50.2 Part of test data

Group	Characteristic parameters					Fault status
	f_w	i_{av}	i_{std}	i_m	k	
1	2655	11.1254	1.1235	90.5462	−2.8452e +003	No failure
2	2310	9.6541	2.1413	92.5422	–3.2645e +003	Brush failure
3	2505	10.6521	1.1245	74.0254	−352.1542	Component open
4	1390	11.9012	1.2435	90.5412	−2.2465e +003	Inter-turn short circuit

Using the model in Sect. 50.3.2 for LVQ neural network training, after 200 generations, we can get the trained LVQ neural network. After testing the trained neural network, part of the diagnostic results are shown in Table 50.5.

Here we found there a 1 out of 20 sets of test data has diagnosis error, the misdiagnosis rate is 5 %, which is acceptable. Based on the analysis, we found changes in the relationship between fault features and fault modes as shown in Table 50.6.

Table 50.3 The initial right settings table between the feature vector layer and the competition layer

Num	Characteristic parameters				
	f_w	i_{av}	i_{std}	i_m	k
1	2600	11.1015	1.1542	90.4123	−2886
2	2644	11.1245	1.1245	90.0354	−2865
3	2305	9.6584	2.1542	92.0542	−3386
4	2310	9.6753	2.1635	93.0235	−3265.4
5	2505	10.6584	1.1542	76.0542	−362.1
6	2510	10.6753	1.1635	75.0235	−352.12
7	1390	11.9125	1.2525	90.9521	−2215.4
8	1400	11.9062	1.2685	90.7521	−2354

Table 50.4 The connection setting table between the failure mode layer and the competitive layer

Num	Fault mode			
	No failure	Brush failure	Component open	Inter-turn short circuit
1	1	0	0	0
2	1	0	0	0
3	0	1	0	0
4	0	1	0	0
5	0	0	1	0
6	0	0	1	0
7	0	0	0	1
8	0	0	0	1

Table 50.5 Part of the diagnostic results

Group	Characteristic parameters					Fault status	Diagnostic results
	f_w	i_{av}	i_{std}	i_m	k		
1	2655	11.1254	1.1235	90.5462	−2.8452e +003	No failure	No failure
2	2310	9.6541	2.1413	92.5422	−3.2645e +003	Brush failure	Brush failure
3	2505	10.6521	1.1245	74.0254	−352.1542	Component open	Component open
4	1390	11.9012	1.2435	90.5412	−2.2465e +003	Brush failure	Brush failure

Table 50.6 Relationship between fault features and fault modes

Typical fault modes	Characteristic parameters				
	f_w	i_{av}	i_{std}	i_m	k
Brush failure	Decrease	Decrease	Increase	Increase	Obvious decrease
Component open	Decrease	Decrease	A little change	Obvious decrease	Obvious increase
Brush failure	Obvious decrease	Increase	Increase	A little change	Increase

50.5 Conclusion

This article applied the LVQ neural network theory to solve a servo DC motor diagnosis problem. The input are the current the pulse frequency f_w, the mean steady-state current i_{av}, the standard differential steady state current i_{std}, the starting current peak i_m, the rate of current peak change k. The four output modes are: no failure, brush failure, component open, and inter-turn short circuit. Based on these, we designed the system user interface of the mode and user guide, the final simulation results show the effectiveness and accuracy of our method, also it verify the correctness of the LVQ neural model in solving the DC motor diagnosis problem.

References

1. Filbert D (1995) Advanced fault diagnosis for the mass production of small power electric motros. Electromotion 11(2):159–166
2. Lu F, Wang XQ (2001) Advance in motor fault diagnosis. J Shanghai Marit Univ 22(3):1–4
3. Shen BZ (1999) Electric motor fault diagnosis technology. China Machine Press, Beijing
4. Liu ML, Cui SM, Guo B (2011) A method of failure recognition based on fuzzy C-means support vector machines for permanent magnetic DC motor. Micromotors 44(10):78–80
5. Cui LZ, Yu FS, Hu ZG (2005) Design and realization of BP neural network in the fault diagnosis of DC motor. Electric Switchgear 5:45–48
6. Liu WX, Yan F, Tian L et al (2012) LVQ neural network approach for fault location of distribution network. Power Syst Prot Control 40(5):90–95
7. Thukaram D, Khincha HP, Vijaynarasimha HP (2005) Artificial neural network and support vector machine approach for locating faults in radial distribution systems. IEEE Trans Power Delivery 20(2):710–721
8. Bi TS, Ni YX, Wu FL (2002) A novel neural network approach for fault section estimation. Proc CSEE 22(2):73–78
9. Bayindir NS, Ukrer OK, Yakup M (2003) DSP based PLL controlled 50–100 kHz 20 kW high-frequency induction heating system for surface hardening and welding applications electric power applications. IEEE Proc 150(3):365–371

10. Liu YN (2009) Image classification of remote sensing and its application based on LVQ neural network and gray level co-occurrence matrix. Master thesis of Chengdu University of Technology, Chengdu
11. Gao F, Sun SZ (2010) Application of AI in fault diagnosis. Sci Technol Inf 23:63–64
12. Yang LL (2012) Model recognition and classification of hand-written numbers based on artifical neural networks. Master thesis of North university of China, Taiyuan

Chapter 51
Studies on Robots Technology of Autonomous Decision Based on Self-adaptive Constraint Transmission

Hang Yin, Quan Zhou and Haiwei Wang

Abstract Aiming at the problem of autonomous decision in the robot intelligent control, state transition model is used to completely describe the action and observation of the robot system and the data model is for robots autonomous decision. And then a method of self-adaptive constraint transmission is put forward to solve the explosion problem of solution space in the process of robot autonomous decision, to reduce the solution search space and remarkably speed up the solution efficiency. It provides the reference of the research of autonomous and intelligent control for the future robot system.

Keywords Robot system · Autonomous decision · State transition model · Self-adaption constraint transmission

51.1 Introduction

The robot manipulation refers to the way of human being to control robot system, mainly divided into three ways of the remote control, semi-autonomous control and autonomous control [1]. Remote control means that the operators monitor and perceive the system state and environments of the robot through sensor data in the distance [2]. And then the operators control the actuator in robot; Semi-autonomous control, namely independent control monitored by the operator, based on the remote control, the part of control tasks are encapsulated as the automatically executable package. There are two purposes. One is that the autonomous system must rely on the existing knowledge in independently "thinking" on the questions, knowing

---

H. Yin (✉)
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: yinhangbbb@163.com

H. Yin · Q. Zhou · H. Wang
Xichang Satellite Launch Center, Xichang 615000, China

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_51

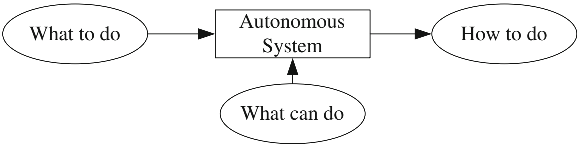


Fig. 51.1 The functions of the model of planning and scheduling in autonomous system

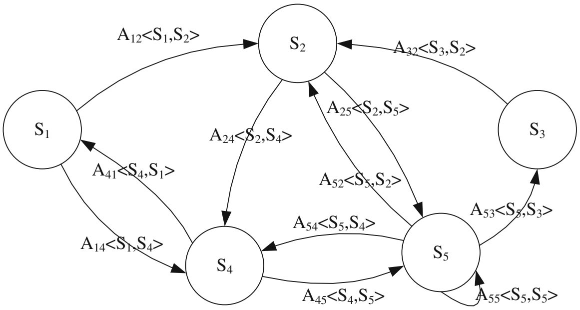
"what can we do" [3, 4], which requires abstraction of the behavior of the system. Another aspect is that the model of planning and scheduling provides the regulated interface people and systems can understand easy to pass over information "what to do" [5, 6] for autonomous system. It is necessary to describe the task target, as shown in Fig. 51.1.

As for the designer of decision control system of the robot, planning and scheduling model is the tool used to describe and define the robot system capability. As to manipulation personnel, planning and scheduling model is the data interface to control the robot system [7]. State transfer model is based on the model of planning and scheduling of state transition. Its basic description factors mainly include action and state, the action execution capable of dynamically changing the state. Every state transition model is equivalent to a woven mesh made up of a finite state and transfer action among states [8] as shown in Fig. 51.2.

The circle represents a state, the directed line among states represents to the action of state transfer between two states. The state transition model can describe all changes in a system in the form of matrix. Figure 51.2 the state transition matrix of the state transition model is as below:

$$A = \begin{bmatrix} 0 & A_{12} & 0 & A_{14} & 0 \\ 0 & 0 & 0 & A_{24} & A_{25} \\ 0 & A_{32} & 0 & 0 & 0 \\ A_{41} & 0 & 0 & 0 & A_{45} \\ 0 & 0 & A_{53} & A_{54} & A_{55} \end{bmatrix}$$

Fig. 51.2 The schematic diagram of a state transition model



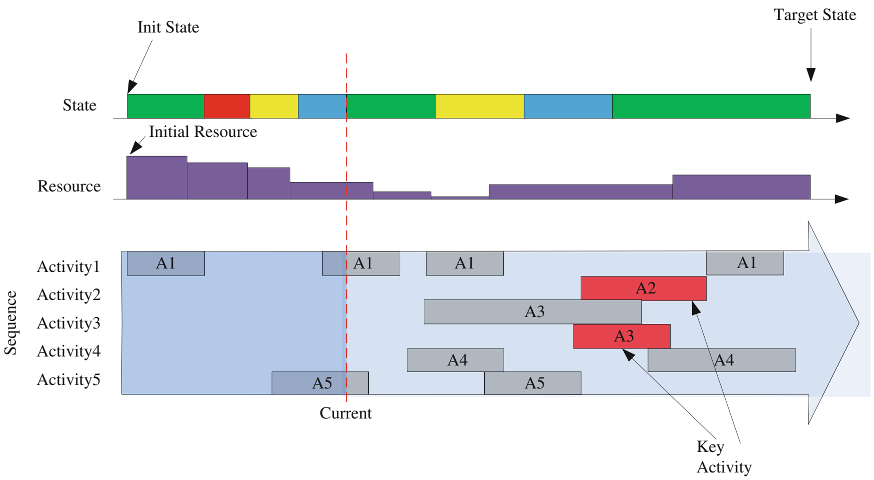


Fig. 51.3 Robot planning and scheduling model based on time axis

You can get the state transition model of the system: $X(n) = AX(n-1)$. $X(n-1)$ and $X(n)$ indicate the state transfer results in $n-1$ and n time. $X(0)$ indicates the initial state of the system. In the problems of planning and scheduling based on the model of state transition, from the beginning of the initial state, the execution actions are constantly chosen to make the state of the system continuously changed until it achieves the goal state. Finally the action sequences obtained is the planning results.

Autonomous decision model of the robot system is described through activities types, resources factors, state factors and constraint relationship and the 2 dimensions plane based on the time axis is used to show the planning model and results, as shown in Fig. 51.3.

Among them, transverse direction is for the sequence time, representing the dynamics of activities, resources and state changes with time, the longitudinal direction as state resources and activity types. The results of planning and scheduling are the timing sequence of activities of the robot system.

51.2 Self-adaptive Constraint Transmission Algorithms

51.2.1 The Analysis of Autonomous Decision Problems

The main purpose of planning and scheduling solution is to assemble the activity sequence capable of completing the target through the behavior model of the control object. Since planning and scheduling model of the above robot combines two types of features such as state transfer and constraint satisfaction. Among them, the problem of state transfer focuses on the activities selected to make the system obtain the goal, whereas the problem of constraint satisfaction does on the

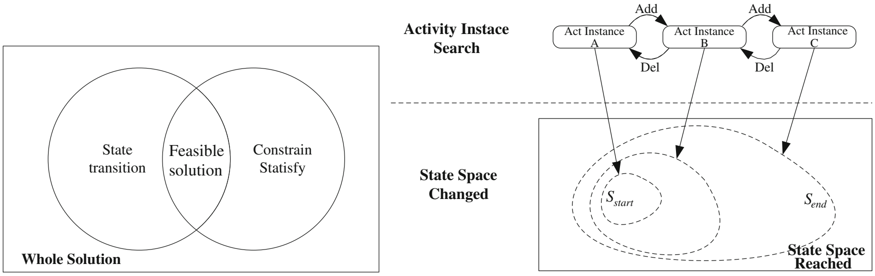


Fig. 51.4 The schematic diagram of the search solution procedure

conditions selected to meet the constraints. Therefore, the targets of decision solution is to search for the simultaneous state transfer rules and various constraints solutions from two angles under the premise of completing the task goals as shown in Fig. 51.4 (left) shows.

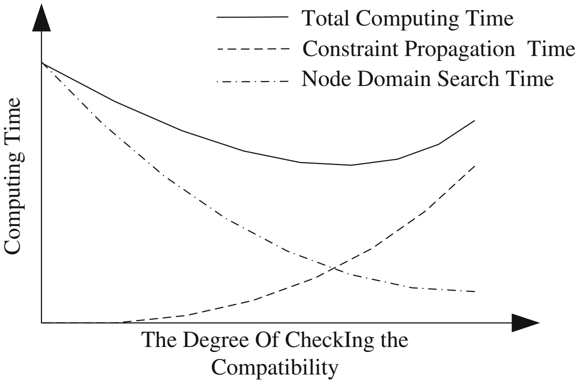
Therefore, the solution process of the single objective planning and scheduling includes two levels, as shown in Fig. 51.4 (right).

1. Activity case set is searched, through expanding or shrinking the case set in order to find can activity case collection capable of reaching the goal state. The way of activity set is to add a new activity case or delete the activity case from the Activity example set.
2. The state space of activity case set is searched. Different activities case set can achieve different states; the same activity case set is different in the reachable state space under different assignment of activity case variables circumstances.

51.2.2 The Design of Self-adaptive Constraint Transmission Algorithm

As for unary constraints, when all values in a variable domain of definition satisfies a unary constraint on the variables in constraint satisfaction problems, and then the variable is accorded with the nodes. As for the binary constraint problem, the binary constraint between variables v_1 and v_2 can be regarded as an arc $\langle v_1, v_2 \rangle, v_1 \in D_1, v_2 \in D_2$. If and only if each meets with value d_1 of a unary constraint in v_1, then a value d_2 averagely exist in $v_2$the domain of definition. If the binary constraint between v_1 and v_2 is satisfied, the two variables are or compatible or the arcs is in consistency. Since compatible detection and calculation of constraint transmission will cause substantial consumption of calculation in the process of reducing the search space. With the accumulation of this time, the consumption of the search node time may even be exceeded. In turn, the overall solution time is caused to increase and the search efficiency to decrease. As shown in Fig. 51.5.

Fig. 51.5 The relationship between the inspection level of compatibility and the solution time of the total calculation



From Fig. 51.5 the statistical solution time can be seen that if the incompatibility of search nodes cannot be effectively eliminated, then the number of search nodes must be increased. Thus solving efficiency will be reduced. So the key question is how to choose a balance point or make the consumption time controllable between constraint transmission calculation and node search calculation. The overall solution time will become better.

Here is the self-adaptive constraint transmission algorithm proposed (see Table 51.1). Its main idea is that the ratio remains at a moderate range between the computational time of node expansion and computing arc consistency transmission, to avoid excessive time consumption for constraint propagation calculation and also

Table 51.1 The self-adaptive constraint transmission algorithm

Input	Extended node points path of searching, the maximum $T_{\max}$ in a single transmission
Output	The threshold value D after constraint transmission
1	Solving time ratio K and tolerance ratio Δk Te = 0, Tc = 0 are set up
2	All nodes in the node points path existing in constraint arc are added in the queue P of constraint transmission solution
3	The non-transmission constraint arc $C < P_o, P_n >$ is taken out from the propagation queue
4	P_n is constrained and spread through the assignment of P_o
5	Setting the current solution state = constraint propagation solution
6	The current transmission queue is not empty
7	Judge the proportion in time between node expansion and constraint transmission for estimation (node expansion time, constraint transmission time)
8	If back to TRUE, and going on solution according to the current state of solution, cumulating corresponding solution time
9	If back to FALSE, executing another solution calculation, and cumulating corresponding solution time
10	If the current queue is empty, expanding the nodes, and recording the solution time
11	Return to step 8

Table 51.2 Estimation and judgment of the adaptive algorithm

Input	Node expansion time, constraint transmission time
Output	TRUE does not change the calculation method, FALSE switching the way of calculation
1	If the current state for solving constraint propagation, and, return TRUE
2	If the current state for solution is constraint transmission, and $K - \Delta k \leq T_c/T_e \leq K + \Delta k$, return TRUE
3	If the current state for solution is constraint transmission, and $T_c/T_e \geq K + \Delta k$, return FALSE
4	If the current state for solution is node expansion and $T_c/T_e \geq K + \Delta k$, return TRUE
5	If the current state for solution is node expansion and $K - \Delta k \leq T_c/T_e \leq K + \Delta k$, return TRUE
6	If the current state for solution is node expansion and $T_c/T_e \leq K - \Delta k$, return FALSE

to avoid the large search space caused by not fully constraint propagation. The time relationship is in self-adaptive constraint transmission.

$$K - \Delta k \leq \frac{T_c}{T_e} \leq K + \Delta k \, \Delta k \leq K$$

where K refers to the ratio threshold of the constraint propagation calculation and node expansion time; Δk means self-adaptive value ($\Delta k \leq K$); T_c indicates the consumption time of constraint transmission; T_e refers to the consumption time of the node expansion. On $\frac{T_c}{T_e} \geq K - \Delta k$, if the arc consistency is being currently calculated, then the calculation is stopped and node expansion calculation is start up. On $K - \Delta k \leq \frac{T_c}{T_e} \leq K + \Delta k$, the current process of solution is not intervened. On $\frac{T_c}{T_e} \leq K - \Delta k$, if the node expansion computation is done, and the constraint transmission queue is not empty currently, the expansion calculation of variable node is immediately stopped and the arc compatible calculation is done. Thus can avoid the excessive time of unnecessary constraint transmission calculation and the time required for solving problems by itself can be reduced (see Table 51.2).

51.3 Simulation Verification

The simulation example (part) of the robot system designed, as shown in Table 51.3.

Due to the characteristics of dynamic planning domain model of the robot, the activity type of any number may cause the solution space containing infinitely many instances of the activity. Thus the boundary conditions in the process of

Table 51.3 The simulation samples from the part of the robot system

Activity name	Duration time	Constraint of activity state	Activities cause the state of change
The unlock of the gripper	5	"The lock state of the gripper" = lock	"The lock state of the gripper" = unlock
The lock of the gripper	7	"on and off state of the gripper" = off	"The lock state of the gripper" = lock
The close gripper	5	"on and off state of the gripper" = on, envelopment	"on and off state of the gripper" = off
The envelopment of the gripper	5	"on and off state of the gripper" = on	"on and off state of the gripper" = envelopment
The opening of the gripper	5	"on and off state of the gripper" = envelopment, off	"on and off state of the gripper" = on
The target of capture	10	"The state of the target capture" = not be captured	"The state of the target capture" = be captured
Visually guided movement	10	"The relative position to get the target from the hand and eye camera of the manipulation arm" = obtain	"The control state of the manipulation arm" → Visually guided control

Table 51.4 The constraints of the search process

The maximum time of search	3,600 s
The drawn conditions of search	The improved time is more than 10 to search for optimal solving

solution are set to limit the process of search, as shown in Table 51.4. The parameters of self-adaptive constraint transmission algorithm are shown in Table 51.5.

The fitness of self-adaptive constraint transmission algorithm.

With no consistent constraint transmission, the search process of the solution space from the complete consistency transmission and self-adaptive constraint transmission is shown in Fig. 51.6.

Among them, "⊙" represents the nodes to add the activity case; "—" is on behalf of a movement of the activity case node added on the time axis; "×" indicates the backtracking after the nodes are moved. The abscissa represents the cycle of planning and scheduling and the ordinate means the types of activities in the planning domain. Each statistics of the search process is as shown in Tables 51.6, 51.7 and 51.8.

Table 51.5 The parameters of self-adaptive constraint transmission algorithm

The ratio value of transmission time and node time	0.2
The switch toleration of solving calculation	0.1

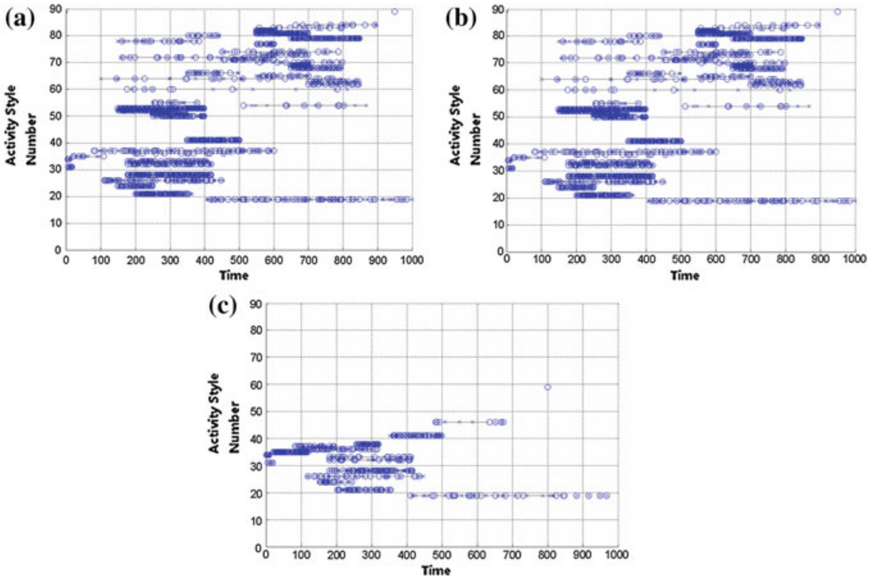


Fig. 51.6 The schematic diagram of the searching process of the solution space in three times

Table 51.6 The solution results of no consistent constraint transmission

The number of search nodes	253965
Whether the solution is obtained or not	Yes
The total number of the activity case of optimal solution	69
The total time of the activity case of optimal solution	1565
The quantity of electricity consumption of optimal solution	224
The fuel consumption of optimal solution	100

Table 51.7 The solution results of complete consistent constraint transmission

The number of search nodes	62383
Whether the solution is obtained or not	Yes
The total number of the activity case of optimal solution	38
The total time of the activity case of optimal solution	1085
The quantity of electricity consumption of optimal solution	174
The fuel consumption of optimal solution	100

Table 51.8 The solution results of self-adaptive constraint transmission

The number of search nodes	89243
Whether the solution is obtained or not	Yes
The total number of the activity case of optimal solution	32
The total time of the activity case of optimal solution	960
The quantity of electricity consumption of optimal solution	142
The fuel consumption of optimal solution	100

51.4 Conclusions

The solving space of the solution problems of robot system decision is very large, although the initial minimum upper bound limit is given for the activity resource consumption, the number of activity case and the total length of activity case and so on, in theory, the limits of the search space is to a limited range. Practically the large solving space is still remained. The part of calculation resources is required to be consumed to compute the constraint transmission in the consistency constraint transmission. In order to effectively reduce the solution space and more effectively search the solving space. Self-adaptive constraint transmission algorithm can avoid the excessive consumption of calculation resources to detect the consistency of search nodes. The “excessive transmission” problems are avoided in the late of spread. Relative to complete constraint transmission, the search efficiency is further improved.

References

1. Gao J (2008) Inference based on algorithms for solving constraint satisfaction problems. Master thesis of Jilin University, Changchun
2. Liu CH (2008) Research on the constraint propagation based constraint solving methods. Master thesis of Jilin University, Changchun
3. Wang H, Nian SR, Fan BF et al (2010) Multi-objective dynamic gait parameters optimization for biped robot based on constraint satisfaction. Appl Res Comput RFC 27(12):4507–4510
4. Wang SJ (2008) Intelligent scheduling technology in flexible manufacturing shop floor based on mechanism of autonomy and coordination. Master thesis of Shanghai Jiaotong University, Shanghai
5. Xu ZB, Gu TL, Chang L (2011) Symbolic ADD algorithm for weighted constraint satisfaction problem. Pattern Recog Aitif Intell RFC 24(1):14–21
6. Ghallab M, Nau D, Traverso P (2006) Automated planning theory and practice. Translated by Jiang YF, Yang Q, Ling YB. Tsinghua University Press, Beijing
7. Sather DG, Sullivan TJ, Nishinaga RG (2010) A compatible approach to satellite operations (SATOPS). In: SpaceOps 2010 conference, Huntsville, Alabama. AIAA RFC 2304
8. Zheng JQ, Wang YL, Li ZG (2004) Evaluating model of process plan for large complicated stampings. Forging Stamp Technol 29(5):12–15

Chapter 52
Test Suite Generation Based on Minimum Cut Set and Interaction Testing

Weixiang Zhang and Wenhong Liu

Abstract How to select a few representative test cases for efficient testing is one of the most significant subjects in software testing. It has becoming even more and more important due to the increasing software complexity. This paper gave an integrative method for test suite generation. The method mainly took advantage of fault tree analysis and interaction testing algorithm to generate and select test cases. It could be divided into three steps: firstly, brought forward an amended Fussell-Vesely algorithm and used it to get software minimal cut sets; secondly, used black-box testing methods to obtain typically discrete values of each element in minimal cut sets; finally, presented an interaction testing algorithm to generate the test suite. Compared with existing Fussell-Vesely algorithm and interaction testing algorithm, this method is more suitable in wide area. Practice has shown that the method can significantly reduce test cases count along with ensuring software testing effect.

Keywords Software testing · Test suite generation · Interaction testing · Fault tree analysis · Minimum cut set · Software engineering

52.1 Introduction

With the rapid development of information technology, software quality wins a growing concern.

Software testing is one of the most important means to ensure software quality. With the increasing software complexity, software testing is becoming more and more difficult and expensive. Data show that software testing has accounted for more than 50 % of the total cost of software development process [1]. How to select a set of software test methods and a few of test cases to test software effectively becomes an outstanding problem.

W. Zhang (✉) · W. Liu
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: wxchung@gmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_52

Test case generation has become a most important subject in software testing. So far, there are some results on application of fault tree analysis (FTA) to software testing [2–5], which aim to mainly FTA modeling and minimal cut set acquisition. Due to lack of test suite generation methods, these results often lead to a large count of test cases and so affect testing efficiency. Interaction testing is an effective software test technology, which can improve the chance of finding software bugs due to taking account of the relationship between software parameters [6–9]. But, there is rare research on how to get the required basic data for combination.

This paper proposes a new test suite generation method combine with interaction testing and FTA. First, gives an amended Fussell-Vesely algorithm to transform software testing requirements into minimum cut set. Then, in Sect. 52.3, gives our test suite generation algorithm using of interaction testing technique. Based on the minimum cut set and making using of interaction testing and black-box techniques, the method can not only improve test adequacy but also significantly reduce the number of test cases. At last, an example is given to illustrate the process of the method and validate its effect.

52.2 Fault Tree Analysis and Minimum Cut Set

52.2.1 Fault Tree Analysis

Fault tree (FT) is an inverted tree-like causal diagram. Fault tree analysis (FTA) method makes use of FT to indicate the logical relationship between faults with its causes by deductive means.

FTA method is helpful to analysis the trigger conditions and transition probabilities of the transformation from a correct or an incorrect state to an insecurity state. So, we can use FTA as a tool on test cases design to find the potential software flaws and weaknesses, then to improve software reliability ultimately.

The general process of FTA includes 4 steps [10]. The first step named system definition is to analysis the scope of software faults and collects their information. The second step named top event selection is to choose the fault to be analyzed. The third step named fault tree construction is to analysis the causes of every event step by step, from the top event to bottom events. The last step, named minimal cut sets generation, is to identify the minimal set of the elements that would lead to the given fault. In addition, if necessary, further analysis can be done to get the occurrence probability of the top event according to the bottom events.

52.2.2 Amended Fussell-Vesely Algorithm

The key of FTA is to obtain minimal cut sets, and the Fussell-Vesely algorithm proposed by Fussell is a standard algorithm to do this. But, the Fussell-Vesely algorithm cannot cover the “2 fix 3” relationship, which is often used in engineer, especially in the aerospace software field.

The “2 fix 3” relationship is an addition to the usual logic NAND gate, which means that a proposition becomes true if and only if there are at least two among all its three conditions are true. We amend the Fussell-Vesely algorithm as follow to catch the “2 fix 3” relationship.

Amended Fussell-Vesely algorithm:

1. Construct a set S contains only the root node T and drill down on all child nodes of T.
2. If the relationship between the current node and its child nodes is logical AND, put all the child nodes into and removed the current node from the current set; else if the relationship is logical OR, copy the current set into the same number of copies with children’s count, insert one different chide into and removed the current node from each copy; else if the relationship is “2 fix 3”, copy the current set into 4 copies, removed the current node from each copy and insert into one of the following combinations: 1st child node, 1st and 2nd child node, 1st and 3rd child node, 2nd and 3rd child node.
3. If the current node is not the root and has extended completely, set the parent node as the current node; else if the current node is the root node and has extended completely, go to Step (5).
4. From the collection of the child nodes of the current node, select one that is not expanded as the current node, perform Step (2). If all child nodes are completed, perform Step (3).
5. In all generated set, remove duplicate elements respectively.
6. Eliminate redundant sets using of Boolean algebra law $A + AB = A$ and $A + (B + C) = AB + AC$, obtain minimal cut sets.

52.3 Interaction Testing and Test Suite Generation Method

Although minimal cut sets provide a good foundation, it is still necessary to take advantage of effective testing cases design methods when generating test suite to avoid huge count of test cases, particularly when the minimal cut sets are large.

52.3.1 Interaction Testing

Software failure may be due to a default of one software parameter, but it is more often due to the interaction of multiple parameters. Interaction testing is an effective method to check the relationship between several parameters, and it considers with less count of test cases to achieve coverage of various combinations [9].

To use the interaction testing method, it is necessary firstly to find out the various software parameters in every given scenarios such as software status, configuration parameters, user input and external event parameters, etc. Every parameter has its own continuous or discrete range. It is impossible and unnecessary to traverse the entire value of all parameters. The essence of interaction testing is to select a suitable number of test cases to cover certain combination of these parameters.

In general, according to the extent of coverage, interaction testing can be divided into single factor covering, couple factors covering, triple factors covering, and so on.

With the increasing of coverage requirements, the probability of finding software defects would be higher. But on the other side, the cost of software testing would become bigger as the count of test cases would increase exponentially. In practice, couple factors covering is used most commonly [11].

52.3.2 Software Testing Suite Generation Method

Combing with FTA, black-box testing methods and interaction testing technique, we bring out our method here. It can be divided into four main steps (as shown in Fig. 52.1):

Step 1 According to the fault tree analysis method to get test software fault tree, then using of the amended Fussell-Vesely algorithm (shown above in Sect. 52.2) to obtain minimal cut sets.

Step 2 Using black-box testing methods such as equivalence partitioning and boundary value analysis, to discrete every element of each minimal cut set for their typical values set.

Without loss of generality, here are m elements $c_1, c_2, \ldots, c_m$ in any given cut set K, whose typical values are $T_1, T_2, \ldots, T_m$ respectively.

Lets $\alpha_j = |T_j|$ denoted the count of values in T_j, $\alpha_m \le \alpha_{m-1} \le \cdots \le \alpha_1 = \max_{1 \le j \le m} \alpha_j \equiv n$.

Step 3 Using interaction testing to generate test cases suite of minimal cut set K. This step is divided into three small steps.

(1) *To construct matrix $C = (c_{ij})_{n \times m}$, which is composed of the typical values of each elements in minimum cut set K.*

The specific method is as follow: beginning with c_1, put all typical values of c_1 into the 1st column of matrix C; then, put all typical values of c_2 into the 2nd column of matrix C; and so on, loops until the last element c_m is over.

If there are vacancies in any column of C, use any other value in the same column to fill them to ensure each position is occupied. The value to fill a vacancy is named Stopgap, for distinguishing with other values.

(2) *Expand the matrix C to obtain the initial test suite I_0 of the minimal cut set K. According to the magnitude relation of n and m, perform 2A or 2B respectively.*

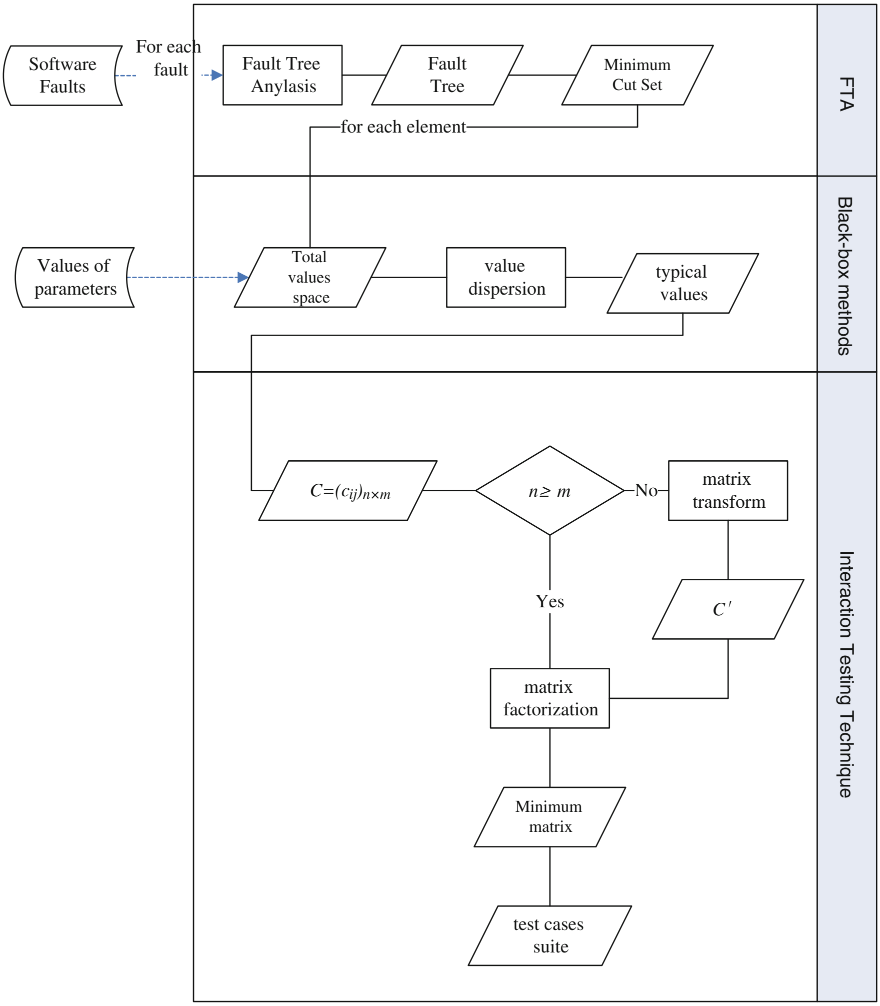


Fig. 52.1 Main process of the testing suite generation

(2A) For n > m, expand the matrix C by the manner similar to matrix factorization method.

(2A.1) Expand the matrix C to the minimum matrix D.

If $n > m$, D is $(n - m) \times 1$; else, D is 2×2.
The specific method is as follow: taking the first column of the matrix as beginning, sum up all ITEM by multiplying each value by its remained matrix, where remained matrix meaning the matrix eliminated the row and column located the value, recursive until the minimum matrix complete.

During this process, if there is Stopgap in any ITEM outside minimum matrix, exclude the ITEM.

(2A.2) Expand the minimum matrix D.

- For $n > m$, if there is more than one NORMAL value in D, multiply each NORMAL value by the rest of ITEM except D; if every value in D is Stopgap, multiply any a Stopgap by the rest of ITEM except D. Insert each product we obtained which is a test case into I_0.
- For $n = m$, expand D by matrix factorization manner, multiply its each part by the rest of ITEM except D. Insert each product we obtained which is a test case into I_0.

(2B) For $n<m$, transform C into C^T and expand C^T by matrix factorization manner. This step is divided into two small steps.

(2B.1) Expand the matrix C^T to the minimum matrix D.

The approach is similar to 2A.1. D is $(m-n) \times 1$ here.

(2B.2) Expand the minimum matrix D.

Multiply each value in D by the rest of ITEM except D. Insert each product we obtained which is a test case into I_0.

(3) *Adjust I_0 to I. This step is divided into two small steps.*

(3.1) *If there are more than two NORMAL values in any row of C, take up all values of the row and insert the resulting product into I_0*

(3.2) *If there are at least two test cases identical in I_0, retaining only one of them and eliminate the others. Until now, I_0 turns into the last test suite I of minimal cut sets K.*

Step 4 Loop STEP 3, until we get all test suite of each minimum cut set. The union of all test suite is just the test suite of targeted software.

Example Here is an example to further illustrate the use of the method. For the fault tree as shown in Fig. 52.2, apply our method as follow:

Firstly, get minimum cut set by amended Fussell-Vesely algorithm (as shown in Fig. 52.3):

(1) Create a collection S = {T}, set T as the current node;
(2) Search all child nodes of T, as M0 is logical AND operation, put all children of T (M1 and M2) into and delete T from the set, so the set {T} comes into {M1, M2};

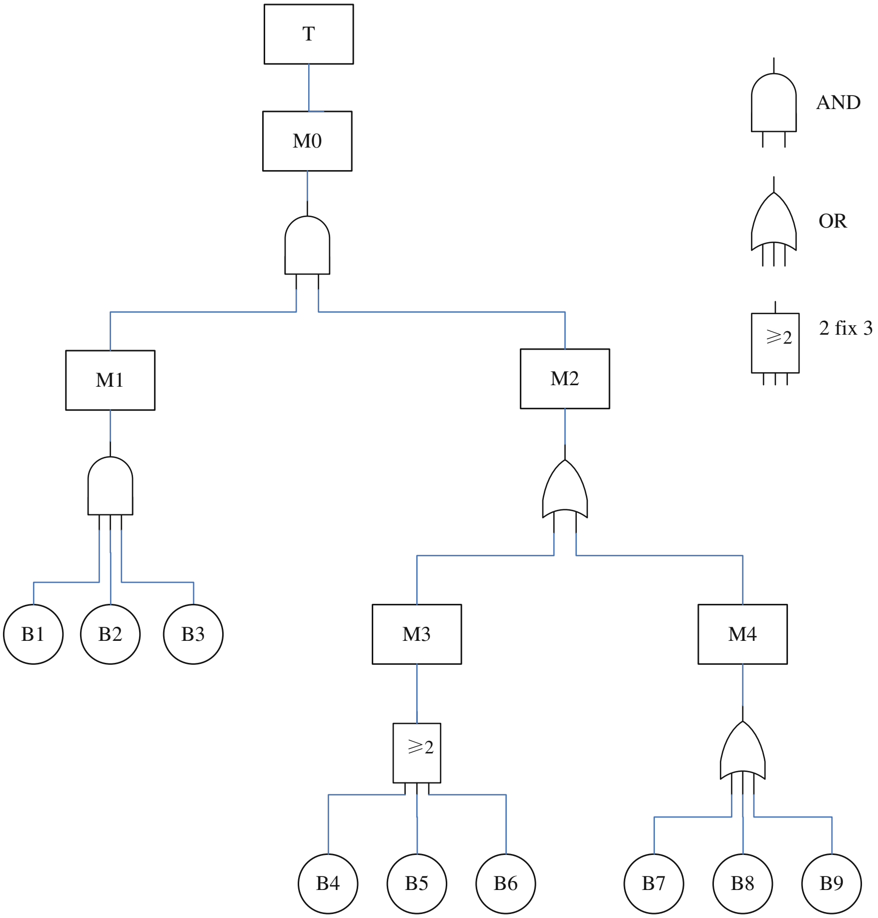


Fig. 52.2 An example of a fault tree

(3) Set M1 as the current node, as M1 is logical OR operation, {M1, M2} becomes {B1, B2, B3, M2}; because of each of B1, B2, and B3 is leaf node, M1 is completed;

(4) Since M1 is not root node, set its unexpanded sibling M2 as the current node; as M2 is logical OR operation, {B1, B2, B3, M2} becomes {B1, B2, B3, M3} and {B1, B2, B3, M4};

(5) Set M3 as the current node, as M3 is "2 fix 3" operation, {B1, B2, B3, M3} becomes {B1, B2, B3, B4}, {B1, B2, B3, B4, B5}, {B1, B2, B3, B4, B6} and {B1, B2, B3, B5, B6};

(6) Since M3 is not root and already completed, set M4 as the current node, as M4 is logic AND operation, {B1, B2, B3, M4} comes into {B1, B2, B3, B7}, {B1, B2, B3, B8} and {B1, B2, B3, B9};

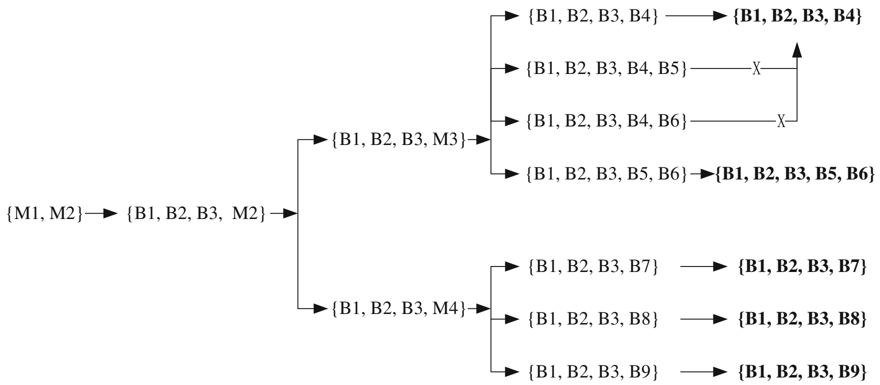


Fig. 52.3 An example of getting minimum cut sets

(7) Now, expansion process is completed. Check all sets by Boolean algebra rules, get five minimal cut sets: {B1, B2, B3, B4}, {B1, B2, B3, B5, B6}, {B1, B2, B3, B7}, {B1, B2, B3, B8} and {B1, B2, B3, B9}.

Secondly, make use of black-box methods and interaction testing technique to access to test suite of each minimal cut sets.

Lets $K = \{B_1, B_2, B_3, B_4\}$ as a example, suppose its values are: $B_1{:}a_1, a_2, a_3, a_4$; $B_2{:}b_1, b_2, b_3$; $B_3{:}c_1, c_2, c_3$; $B_4{:}d_1, d_2$.

According to our method, it can be obtained as follow:

$$\begin{aligned}
C = \begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & \\ a_4 & & & \end{bmatrix} &= \begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_1^* \\ a_4 & b_1^* & c_1^* & d_2^* \end{bmatrix} \\
&= \ldots a_1b_2\begin{bmatrix} c_3 & d_1^* \\ c_1^* & d_2^* \end{bmatrix} + a_1b_3\begin{bmatrix} c_2 & d_2 \\ c_1^* & d_2^* \end{bmatrix} + a_2b_1\begin{bmatrix} c_3 & d_1^* \\ c_1^* & d_2^* \end{bmatrix} + a_2b_3\begin{bmatrix} c_1 & d_1 \\ c_1^* & d_2^* \end{bmatrix} + a_4b_1\begin{bmatrix} c_2 & d_2 \\ c_3 & d_1^* \end{bmatrix} \\
&\quad + a_4b_2\begin{bmatrix} c_1 & d_1 \\ c_3 & d_1^* \end{bmatrix} + a_4b_3\begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \end{bmatrix} + a_3b_1\begin{bmatrix} c_2 & d_2 \\ c_1^* & d_2^* \end{bmatrix} + a_3b_2\begin{bmatrix} c_1 & d_1 \\ c_1^* & d_2^* \end{bmatrix} \\
&= a_1b_2\left(c_3d_2^* + c_1^*d_1^*\right) + a_1b_3\left(c_2d_2^* + c_1^*d_2\right) + a_2b_1\left(c_3d_2^* + c_1^*d_1^*\right) + a_2b_3\left(c_1d_2^* + c_1^*d_1\right) \\
&\quad + a_3b_1\left(c_2d_2^* + c_1^*d_2\right) + a_3b_2\left(c_1d_2^* + c_1^*d_1\right) + a_4b_1\left(c_2d_1^* + c_3d_2\right) \\
&\quad + a_4b_2\left(c_1d_1^* + c_3d_1\right) + a_4b_3\left(c_1d_2 + c_2d_1\right) \\
&= a_1b_2c_3d_2^* + a_1b_2c_1^*d_1^* + a_1b_3c_2d_2^* + a_1b_3c_1^*d_2 + a_2b_1c_3d_2^* + a_2b_1c_1^*d_1^* + a_2b_3c_1d_2^* \\
&\quad + a_3b_1c_2d_2^* + a_3b_1c_1^*d_2 + a_3b_2c_1d_2^* + a_3b_2c_1^*d_1 + a_4b_1c_2d_1^* + a_4b_1c_3d_2 \\
&\quad + a_4b_2c_3d_1 + a_4b_3c_1d_2 + a_4b_3c_2d_1 + a_2b_3c_1^*d_1 + a_4b_2c_1d_1^*
\end{aligned}$$

Now, I_0 is achieved:

$$\begin{aligned}
I_0 = \{&a_1b_2c_3d_2^*, a_1b_2c_1^*d_1^*, a_1b_3c_2d_2^*, a_1b_3c_1^*d_2, a_2b_1c_3d_2^*, a_2b_1c_1^*d_1^*, a_2b_3c_1d_2^*, a_2b_3c_1^*d_1, \\
&a_3b_1c_2d_2^*, a_3b_1c_1^*d_2, a_3b_2c_1d_2^*, a_3b_2c_1^*d_1, a_4b_1c_2d_1^*, a_4b_1c_3d_2, a_4b_2c_1d_1^*, a_4b_2c_3d_1, \\
&a_4b_3c_1d_2, a_4b_3c_2d_1\}
\end{aligned}$$

Table 52.1 Comparison of the count of test suite

	Our method	Full coverage	Rate (%)
$K_1 = \{B_1, B_2, B_3, B_4\}$ $B_1{:}a_1, a_2, a_3, a_4;\ B_2{:}b_1, b_2, b_3$ $B_3{:}c_1, c_2, c_3;\ B_4{:}d_1, d_2$	17	72	23.6
$K_2 = \{B_1, B_2, B_3, B_5, B_6\}$ $B_5{:}e_1, e_2, e_3, e_4, e_5$ $B_6{:}f_1, f_2, f_3, f_4, f_5, f_6$	89	1080	8.2
$K_3 = \{B_1, B_2, B_3, B_7\}$ $B_7{:}g_1, g_2, g_3, g_4, g_5, g_6, g_7$	58	252	23.0
$K_4 = \{B_1, B_2, B_3, B_8\}$ $B_8{:}h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8$	70	288	24.3
$K_5 = \{B_1, B_2, B_3, B_9\}$ $B_9{:}i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8, i_9$	82	324	25.3
Total	316	2016	15.7

Put into $a_1b_1c_1d_1, a_2b_2c_2d_2, a_3b_3c_3d_1^*$ and eliminate unnecessary items, we obtain test suite of K:

$$I_1 = \{a_1b_2c_3d_2,\ a_1b_2c_1d_1,\ a_1b_3c_2d_2,\ a_1b_3c_1d_2,\ a_2b_1c_3d_2,\ a_2b_3c_1d_1, a_3b_1c_2d_2, a_4b_1c_3d_2, a_4b_1c_2d_1,\ a_4b_2c_1d_1,\ a_4b_2c_3d_1, a_4b_3c_1d_2,\ a_4b_3c_2d_1,\ a_1b_1c_1d_1, a_3b_3c_3d_1,\ a_3b_2c_1d_1, a_2b_2c_2d_2\}$$

At last, after getting the test suite I_2, I_3, I_4, I_5 of remaining four minimal cut sets by the same manner, the total test suite is $I = I_1 \cup I_2 \cup I_3 \cup I_4 \cup I_5$.

Comparison of the testing cases count of each cut sets from our method to full-coverage is shown in Table 52.1. As shown in Table 52.1, the effect of our method is very obvious. In fact, the effect will become better and better along with increasing parameters count and more complexity.

52.4 Conclusion

In the field of software testing, there are many difficult issues, such as the selection of test methods, the reduction of test cases and the timing of test termination. The fundamental reason is the conflict between test adequacy and test cost. How to minimize the test cost to get the best test effect is the ultimate goal of study on software testing.

This paper presented a test suite generation method based on FTA and interaction testing. We gave its specific algorithm steps and illustrated its effect.

By transforming software testing requirements into minimum cut sets and making use of interaction testing to cover the interactions between multiple parameters, the method can better ensure testing adequacy and substantially reduce the count of test cases. In the future, we will aim to software test suite reduction and optimal generation technology for further study.

References

1. Ammann P, Offutt J (2008) Introduction to software testing. Cambridge University Press, Cambridge
2. Xiao YB, Xiang JW, Xu RZ (1999) The fault tree analysis of software reliability stable growth and safety testing. Mini-Micro Syst 20(9):668–671
3. Liu WH, Wang ZW, Wu X (2004) Application of the failure tree analysis in software test. Syst Eng Electron 26(7):985–988
4. Hu Z, Yin RK (2006) Dynamic generation of safety test case based on minimal cuts sets. Comput Eng Des 27(16):3018–3020
5. Zhu YP (2006) Research on technique of software testing based on fault tree analysis. Comput Eng Des 27(16):3018–3020
6. Zhang WX, Liu WH, Du HS (2012) A software quantitative assessment method based on software testing. Lecture notes in artificial intelligence (LNAI) 7390:300–307
7. Nie CH, Xu BW (2004) An algorithm for automatically generating black-box test cases based interface parameters. Chin J Comput 27(3):382–388
8. Kobayashi N, Tsuchiya T, Kikuno T (2002) A new method for constructing pair-wise covering designs for software testing. Inf Process Lett 81(2):85–91
9. Chen X, Gu Q, Wang XP et al (2010) Research advances in interaction testing. Comput Sci 37 (3):1–5
10. Zhu JiZ (1989) Fault tree theory and its application. Xi'an Jiaotong University Press, Xi'an
11. Zhang WX, Liu WH (2008) A method for automatically generating black-box test cases. J. Spacecraft TT&C Technol 27(5):53–56

Chapter 53
The Realization of High Speed Data Transmission and Processing of Space-Based TT&C Network

Tao Ji, Fan Yang and Luyang Pan

Abstract In terms of the characteristics of and demand for the safe and reliable high-speed data transmission in the space-based TT&C network, several approaches of the key technology are fully discussed. The main technical difficulties and problems about the network security control, high-performance data distributed processing and massive data real-time storage are analyzed. Based on the above, the development and exploration ways of high speed data transmission technology in the future are summarized. This can facilitate the engineering implementation of high speed data transmission in the field of space-based TT&C.

Keywords Space-based TT&C · High-speed data transmission · Security control · Parallel storage

53.1 Introduction

The space-based TT&C will become more reliable, secure, high cover, accurate, high frequency band, remote and networking in the future [1]. Because of its advantages of high orbit coverage, multi-target tracking and high speed data transmission, the space-based TT&C becomes widely deployed in the field of aerospace, and has showed considerable economic benefits. With the increasing number of users and application types, high speed data transmission becomes one of the core missions of space-based TT&C. How to achieve the high throughput, real-time, security, uninterrupted processing and distribution of large capacity data will become the technology difficulties and goals for building and utilizing space-based TT&C.

T. Ji (✉) · F. Yang · L. Pan
Beijing Space Information Relay and Transmission Technology Research Center,
Beijing 100094, China
e-mail: jt_htch@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_53

53.2 The Processing Flow of High Speed Data Transmission

In space-based TT&C, space information which include multiple channels and the transmission rate of each channel can reach several hundred Megabits, will be first downloaded to multiple remote ground sites. Through the network transmission equipment and security protection equipment, mass data of Gigabits will be transmitted to the central node of information system, on which the two layers of data processing, i.e., system layer and data distribution and application layer, can be performed to implement the real-time data receiving, processing, storing and forwarding. Finally, via the different business network, the data will be distributed to different user node. The transmission and processing flow of high speed data is illustrated in Fig. 53.1.

Combined with analysis of processing flow of high speed data transmission, there are three aspects that can influence transmission capacity of the whole ground links: security control, data distribution and processing, real-time data storage of mass data. These will be analyzed in the following.

53.3 Security Control of High Speed Transmission Network

During recent years, the high speed ethernet technology develops very fast. With the exploding of volume of IP-based business data, 100 Gbit/s optical transmission network is becoming commercial in large-scale, the routing and switching technology of 10 Gigabits LAN has been mature. Through testing, some middle or high-end routers and switches in the market can support 10 Gb line-card speed

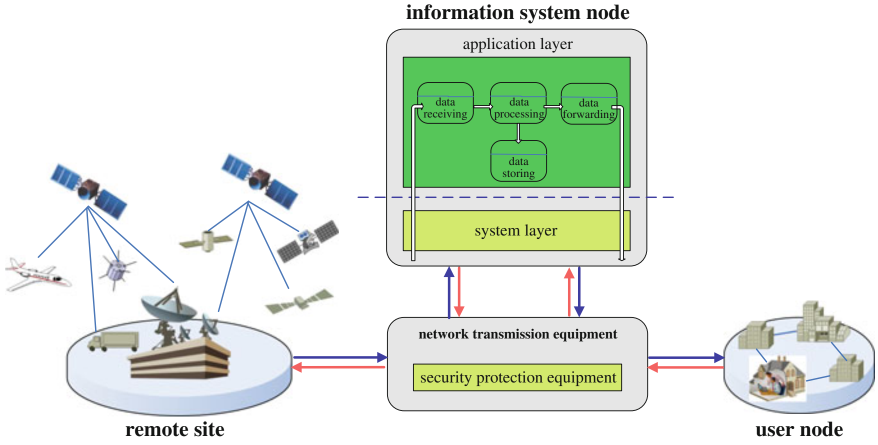


Fig. 53.1 The transmission and processing flow of high speed data of space-based TT&C

forwarding (when the capacity of back plane is enough). Thus, compared with the processing capacity of network devices like routers and switches, lagging developments of security protection, QoS, burst traffic control and so on are the main problems in network applications.

53.3.1 Security Protection of Network

There exists a certain gap between the real capacity of security protection equipment and the requirement of large capacity network transmission, this is mainly in the lack of bandwidth and processing capacity of security protection equipment. For example, firewall is an important security protection equipment, however, the firewalls need to be cascaded in the transmission network, their performances will directly influence the capacity of network transmission. Different with routing and switching equipment, the packet forwarding of firewalls is based on the state table that is constructed according to the data flows. In the state table, there exists more information rather than the normal five tuples, such as matching time, protocol state, application identification, attack state (Source routing attack, Land attack, Smurf attack, SYN/FIN scan, Ping-of-Death attack, etc.), and other related identifications [2]. Thus, the forwarding capacity will be more degraded if we enable more protection function, or add more security policies in the firewalls.

At the same time, because security protection equipments need to adapt to different characteristics of various attacks, thus the processing can not only depend on the simple hardware logic, but also the CPU capacity. Currently, multi-core based parallel processing technology, that can provide high-throughput, flexibility and scalability, has become the trend in high-capacity security equipment. However, multi-core based parallel processing will certainly bring problems in the operate mode and resource allocation among multiple processors. Although multi-core based technology can enhance the business capacity, it can cause packet disorder, because of different processing delay and load in different CPUs.

This case of disorder has very little effect on vast majority of internet applications, because a large number of internet applications using TCP at the transport layer can automatically adjust to the disorder packets in the system-level. Even applications use the UDP protocol, due to its transmission rate is not high, we can also use the software correcting for disorder packets. However, the situation of disorder using UDP protocol for high-speed data transmission in the space-based TT&C network is unacceptable, which requires the firewall to guarantee the order of packet flow. To this end, we propose the main logic of keeping packet order in Fig. 53.2. Receiving the packets, equipment need to mark every packet, and store them in the buffer. CPU needs to find the data to be processed in the receiving buffer according to the dispatcher. After confirming the order of these packets, CPU then will put the data in the forwarding buffer and forward them. If there exists disorder, then the related processing algorithm that can re-order the packets begins to re-queue the packets, after that, CPU can forward the packets.

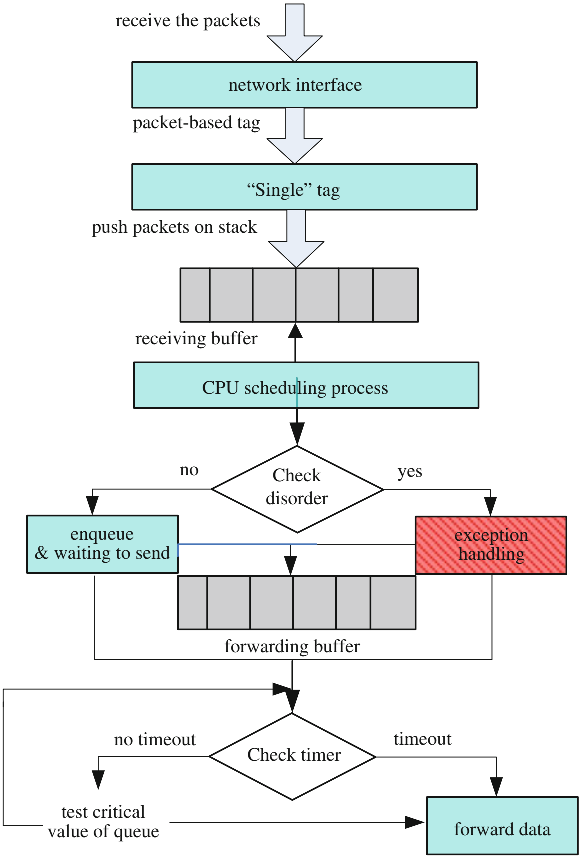


Fig. 53.2 The main logic of keeping packet order

Because of handling packets in order logic needs to consider a variety of unusual circumstances, such as packets loss, long time yet to come, mark overflow, etc., it will consume a lot of CPU processing resources of the firewall. Through optimized design performance of keeping packet order module, ultimately the capacity for processing packets of more than 512 bytes is close to the original state. Combined with a firewall or similar security devices handling scenarios, only to obtain optimization balance between high throughput and keeping packets in order, which can achieve high-speed data distribution performance in the actual application.

53.3.2 QoS and Burst Traffic Control

TT&C network carries a variety of transmission services, such as TT&C message, services, audio and video, documents, etc. Different business application in TT&C has different requirements for network environment. For example, The telegram applications need to guarantee real-time and few packet loss; Multi-media applications, like video and audio applications need to guarantee low transmission delay and stable jitter; Web and FTP applications do not have such requirements in delay, jitter and packet loss. In the high speed transmission network, we need to use QoS technology for guaranteeing the high availability and stability of networks, avoiding network overloading and congestion when there exists multiple business traffic, and ensuring the performance of the importance business traffic [3]. Combined with the practical situations of TT&C, the two key points for optimizing the QoS include:

The first one is for the key business and time-sensitive video/audio applications. Through configuring QoS in network environment, and assigning different transmission priority of different data flow, marking the relative importance of data flow and using the priority-based forwarding policies and congestion avoiding mechanisms provided by the equipments, we can prevent the low-priority flow from occupying more bandwidth and influencing the transmission of high-priority flow, and increase the predictability of the network performance.

The second one is for the link interface congestion problem when facing the burst traffic of high-speed applications. We need to analyze the characteristics of the data in different segment of the full link. For example, in the practical applications of data transmission equipment, one problem occured because the accuracy of the packet sending timer was not high enough, making the sending hardware module exceed the normal speed, sent the packets of a period in a shorter time, causing large volume of burst traffic. Thus the interface of smaller data cache communication transmission device became congested and packet-loss, which influenced the transmission of business data. So we need network analyze/management tools, to monitor and locate the problems for traffic delay, jitter and loss. We can also guarantee the high availability and security of the whole high speed TT&C network, by improving the technologies in both sending and receiving sides.

53.4 High-Performance Process of Data Distribution

The data process of the space-based TT&C network requires high throughput, real-time, and reliability, so the equipments of data process need high bandwidth networking ability, high throughput of I/O disks and high real-time schedule capability. However, the current high-speed data process is based on the conventional model of data distribution, which is implemented by distribution servers and the data need to be copied twice, i.e. from the buffer of the server NIC to the kernel cache of the system and then to the cache of users. Moreover, there will be many hardware interruptions in the process of high speed-data and need to switch in and out the data process thread frequently. As a result, the efficiency of data process is greatly degraded and the requirements of high-performance data distribution will not be met. Jointly considering how to reduce the number of data copies and the influence of hardware interruptions and increase the process ability of servers will be the key issue of implementing the high-performance data distribution.

53.4.1 Zero Copy Technology

In order to reduce the influence of high-speed data transmission caused by the data copies, we reduce the number of data copies and the number of operations sharing the bus, so the cost of the operating system and communication protocols is reduced and the performance of data transmission is improved. This is known as zero-copy technology [4] which is mainly used in the field of high-speed data acquisition records, as well as high-performance network traffic monitoring. Data distribution server which is the intermediate part of the whole links of ground, not only needs to receive data sent from the remote site, but also forwards the data to each user nodes, besides shakes hand with the above application data in both directions. So it must realize efficient data exchange from NIC to user layer, which can use zero-copy mode. Currently there are two main zero-copy technology implementation [5], One is named memory sharing and memory remapping which allocates a sharing storage in the space of network card, the space of system kernel and the space of user cache, it uses the kernel to manage the space of network card, and then transfers data between the network storage and the user cache using DMA. This method has the advantages of reducing the complexity and manageability of the network card, and no need of modifying the application software since the operation semantics above the Socket layer is remained. The other method is dumping the engine network card by the TOE protocol. The TOE network card can run a real-time operating system in its embedded processor, so it can dump the TCP/IP protocol and any protocol supporting the embedded operating system into the network card. As a result, the number of data copies and interruptions is greatly reduced and the load of CPU is mitigated by omitting the process of TCP/IP protocol in CPU and supporting the direct communication between the kernel space and the user space. Via a synthesis of technology upon, we solve the performance bottleneck of data distribution through the development and using of zero-copy card driver based on Linux kernel, and ultimately the server ability of send and receive data increases of more than 30 %.

53.4.2 NAPI Technology

In order to reduce the frequency of the network hardware interruptions, the technology of New Application Programming Interface (NAPI) [6], which optimizing the interruption process of operating system by integrating interruption and polling. When the frequency of receiving data is high, the interruptions will be processed in combination, i.e. the network card will produce an interruption for every N Ethernet frames or for a time interval of T. As a result, the frequency of interruptions is reduced, the amount of data transferred for each interruption is increased, and the throughput of the server is improved.

The working process of the system using NAPI technology is illustrated in Fig. 53.3. The hardware produces an interruption to wake the data receiving

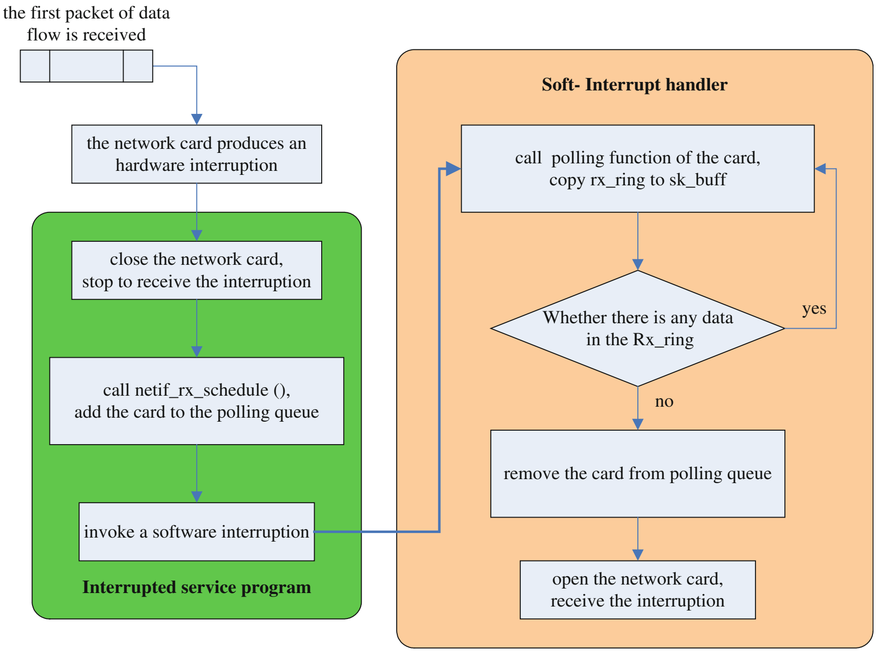


Fig. 53.3 The working process of NAPI

procedure, and then it will get the data by polling. In practice, when the first packet of a series of data is received, it will notify the system by an interruption signal. Then, the system will register the receiving equipment to an polling queue and close the response to the registered equipment. Next, the processing procedure of hardware interruption will invoke a software interruption, and polling the registered network equipment and reading the data. The interruption of receiving will remain in the state of closeness until all the data packets in the receiving caches of equipments in the list are processed. If there is no data in the receiving cache, the equipment will be deleted from the polling queue and its interruption response will be opened at the same time. This method processes the hardware interruptions by combination, so the efficiency of processing high-speed data flows will be improved, especially when most packets are small but the overall traffic is huge.

53.4.3 Optimizing the Selection of Communication Protocols and the Character of Packet Length

The previous analysis shows that the high-speed data distribution equipments have the issues of huge amount of interruptions and limited capability when small packets induce huge traffic. In practice, when transferring many small packets, it will be more rational to consider using the TCP protocol and fully employing the character of flow processing of TCP/IP. Moreover, the problem of low efficiency of data distribution equipments caused by many interruptions of small packets can be solved by selecting the appropriate packet length and the appropriate frequency of packets to transfer. The testing results show that the transmission capability of 1,024 byte packets of the data distribution equipments is larger by a factor of 4.9 compared with the 64 byte packets, and by a factor of 2.7 compared with the 128 byte packets. However, the quality of link is also an important issue when transferring data between remote site and the information system center or users. It needs to be cautious that the low quality of link and the high delay may induce the low efficiency of the TCP protocol and the high data error rate caused by the loss of big packets.

53.5 Optimization of Mass Data Real-Time Storage

In the field of storage technology, whatever storage technology being used, the basic architecture is composed of host I/O interface, connected data cables and storage equipment interface. It is related with I/O bus, transmission control protocol, storage equipment interface and embedded controller [7]. Considering that high speed data distribution need to occupy large network bandwidth, and the load of the TCP/IP protocol stack is very high in the current server operating system. So we mainly use SAN data storage to achieve storage access of multiple front-end processing equipment, data manipulation and backup, data sharing. This application mode do not need to occupy large network bandwidth, can enhance the network efficiency. More importantly, it can achieve centralization of data from different applications and servers physically, these data are all used in shared disk arrays, data management and copy operation can be completed in the same equipment, thus improving the utilization of storage resources.

For achieving high availability of the system storage, and satisfying high performance of synchronous storing and reading in multiple front-end processing equipment, we still need to optimize and adjust the storage system. The processing flow starts with writing operation of users in some application, and ends with the data being written into the physical storage. It includes multiple aspects such as application software, file system, server hardware and disk array. The basic flow of writing operation in storage is shown in Fig. 53.4, thus we can optimize the storage as following:

1. Optimizing the configuration of disk array, making the dual-controller work in parallel, increasing the number of optical fiber channel in the end of the array, and using scatter read/write to make use of the disk array more efficiently.
2. Optimize the file system, first we need to stripe the file system, change the serialize read/write mode to parallel read/write mode [8], reduce the overload of

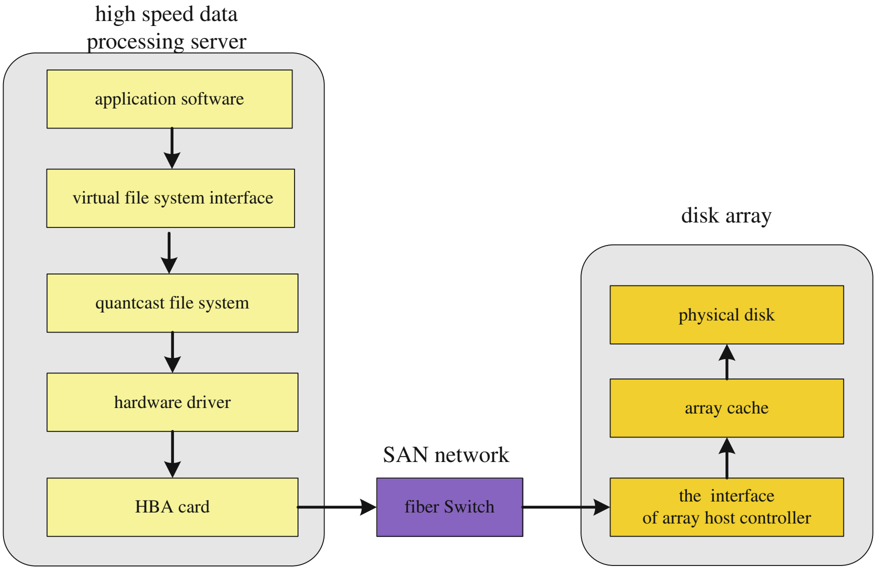


Fig. 53.4 The basic flow of writing data to the storage medium

a single logic disk, and allow multiple hosts or threads to read/write a file at the same time while reducing the waiting time.

3. Considering the high speed characteristics of most applications that need continuous and mass capacity data read/write operations, applications use relatively large single file, and adopt asynchronous mode to write data into physical disk, to make full use of the high speed buffer capacity of the disk array.

By such optimization, and evaluation of the storage interface in the hosts, we find that the read/write capacity of a single interface exceeds 6 Gbit/s, which is almost the upper limit of the interface of FC hosts.

53.6 Future Work

In the future, the business traffic of space-based TT&C will grow exponentially. It can be predicted that the real-time data processing capability will reach several Gigabits in multiple channels, which requires the whole system to have the higher transmission rate, security protection, and the more powerful continuous service capability, and thus requires that the efficiency of data transmission in each step of the data link should be improved. In the aspect of network transmission, the follow-on researches will focus on the related technology of network architecture, device compatibility and device of high security. To solve the problems associated with data forwarding and processing, we need to explore new technologies through either general computer system or dedicated device. For the general computer system, the capacity limit of single device in data processing must be overcome, we can use server cluster technology to enhance the overall throughput of the information system [9], while resolving the confliction between multi-task parallel processing and response time of multiple devices, and also need to design optimized load balancing policy to make full use of resources in the devices. The focus is on the high performance dedicated distributed device, which is based on FPGA or network flow processor. We need to resolve the technology difficulties, such as high performance and flexible loading of multi-task mode, adapting to different network protocols. For high performance data storage, some concurrent sharing and parallel storage systems can already satisfy the basic demands for real-time shared read/write operations of mass data. We need to study the compatibility and efficiency optimization problems for storage system and practical applications of TT&C.

With the development of our own controllable software and hardware computer system, the TT&C high speed data transmission system platform will still face new technology challenges. It is believed that the system will be more controllable, secure, reliable, efficient and usable.

53.7 Conclusion

To satisfy the demand for high throughput, real-time, large capacity, security and reliable transmission of space-based TT&C network in the future, the key technologies of high speed transmission and processing are needed to be researched urgently. In this paper, through analyzing the main technology and key problems of high speed data transmission, we obtain the solution and further exploration way for development, which can provide support for further system design, and also pave the way for the construction and application of the high speed data distribution and processing system of space-based TT&C.

References

1. Lei L, Hu JP, Zhu QZ (2011) Future spacecraft TT&C communication architecture and key technology. Telecommun Eng 7
2. Zhang XQ, Pan Q, Gong B (2013) A network service method based on the control flow and data flow separation mechanism. Digital Technol Appl 10
3. Wu ZA, Luo JZ, Song AB (2006) QoS-based grid resource management. J Softw 17 (11):2264–2276
4. Ke XM, Gong ZH, Xia JD (2000) Research of the zero-copy technique and its implementation. Comput Eng Sci 22(5)
5. Zhao YB (2012) Research on efficient network packet transmission mechanism based on Zero-Copy. Master thesis of Harbin Engineering University, Harbin
6. Tan YS, Jia YC et al (2007) A study methods to improve NAPI polling mechanism. Microcomput Appl 28(9)

7. Jin CY, Li XY, Bai YC (2011) Development and prospect of mass storage system. Comput Appl Softw 28(8)
8. Dai HD (2002) Studies on shared-memory management and optimization technologies in parallel and distributed operating systems. Doctoral thesis of National University of Defense Technology, Changsha
9. Li L (2013) The analysis of the server cluster technology. Digital Technol Appl 10

Chapter 54
Modeling and Synchronization for IEEE 1588 Clock Based on Kalman

Lei Chen, Tianlin Zhu, Feng Liu and Wei Wang

Abstract A modeling method and synchronization algorithm based on accelerated motion Kalman model of IEEE 1588 clock is proposed in this paper, against the deterioration from IEEE 1588 slave clock uncertainties. The clock offset calculated by timestamps is regarded as direct input. To strong the input reliability, the clock skew calculated by clock offset is regarded as indirect input. With the estimation of clock offset, clock skew and clock aging rate, the IEEE 1588 clock model could be modified adaptively and both of the clock offset and clock skew could be compensated with different statistical characters of measurement noise. Verified by some meaningful simulations, this algorithm could attenuate the synchronization performance deterioration effect from IEEE 1588 slave clock. The synchronization accuracy is almost double optimized, and the stability and the deterioration inflection point is optimized with almost one order.

Keywords IEEE 1588 · Kalman · Slave clock uncertainty · Time synchronization

54.1 Introduction

With the network and distribution development of aerospace measurement and control systems, the synchronization among the terminals in a system becomes more and more important [1]. Since that terminals in a distributed system are driven by independent clocks, the clock synchronization performance is just one of the most important indexes in a network distributed system. The hardware synchronization technology based on IRIG-B code which relies on dedicated line, could not fit for the networked and distributed character. And the traditional network timing synchronization based on Network Time Protocol (NTP) [2], could only achieve millisecond level, which not meet the needs of aerospace measurement and control

---

L. Chen (✉) · T. Zhu · F. Liu · W. Wang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: chenlei0831@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 27th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 323,
DOI 10.1007/978-3-662-44687-4_54

systems. IEEE 1588 standard, which was published in 2002 [3] and upgraded in 2008 [4], defines precision time protocol (PTP) fit for networked distributed systems. The synchronization performance based on PTP could achieve sub-millisecond level.

But IEEE 1588 only defines the rules to compensate the clock offset, ignored clock skew which is the root cause of clock offset. So in the field of compensating clock offset, kinds of algorithms have already been put forward, like asymmetric ratio estimation algorithm in Ref. [5] and packet delay estimation based on its distribution algorithm in Ref. [6]. While Kalman filtering also has been applied in optimizing IEEE 1588 performance for its good character in filtering both observation noise and system noise, for example, a 2-states IEEE 1588 clock model in Ref. [7] and a 3-states clock model based on 1-observation input in Ref. [8] which was upgraded from Ref. [7]. However, the 2-states clock model do not take the influence of clock skew variation caused by aging rate into consideration, and the 1-observation input do not consider much in the inter-influence between clock offset and clock skew. So based on the these situation, this paper established a 3-states IEEE 1588 clock model with more observation inputs and put forward a synchronization algorithm based on Kalman. This algorithm introduces clock skew as the new observation input together with traditional input clock offset, realizing 3-states estimation for clock offset, clock skew and clock aging rate, adaptive update of clock model and enhancement of IEEE 1588 synchronization performance.

54.2 Basic Clock Model

The clock is usually comprised of oscillator and counter, whose complex physical characters decide the complexity of clock. So to design clock servo based on observation input, we need model the basic clock. While the main difference between basic clock and IEEE 1588 clock is that basic clock's state could not be adjusted but IEEE 1588 clock's state could be.

Based on the traditional "simple skew model" (SKM) introduced in Ref. [7], the recursive state equations that account for clock behavior at successive synchronization instants can be written as (54.1)

$$\begin{cases} \theta(k+1) = \theta(k) + r(k)\Delta t + \omega_\theta(k) \\ r(k+1) = r(k) + \omega_r(k) \end{cases} \tag{54.1}$$

where $\theta(k)$ is called *offset* and represents the difference between the time reported by the local clock and the time reference at the time instant T_k, and $r(k)$, which is called *skew*, represents the normalized difference between the frequency of the local clock and the frequency of the time reference, at the time instant T_k. And this traditional clock model is parameterized by three constants, namely, the interval Δt between consecutive reference time instants and $\omega_\theta(k)$ and $\omega_r(k)$ are two uncorrelated white Gaussian random noise processes, with variances $\sigma_\theta^2\Delta t$ and $\sigma_r^2\Delta t$, respectively.

The traditional clock model could be good if the environment was constant temperature. However, most real application system could not achieve this, so the clock aging could change the clock skew and thus the clock aging rate should be taken into consideration. We defined the clock aging rate a:

$$a(t) = \frac{\mathrm{d}r(t)}{\mathrm{d}t} \tag{54.2}$$

where $r(t)$ is clock skew. So according to the comparable derivation in Ref. [7], we can make a new 3-states clock model shown as (54.3) which upgrade from (54.1):

$$\begin{cases} \theta(k+1) = \theta(k) + r(k)\Delta t + a\Delta t^2 + \omega_\theta(k) \\ r(k+1) = r(k) + a(k)\Delta t + \omega_r(k) \\ a(k+1) = a(k) + \omega_a(k) \end{cases} \tag{54.3}$$

where $\omega_a(k)$ is another white Gaussian random noise process, uncorrelated with $\omega_\theta(k)$ and $\omega_r(k)$, with variance $\sigma_a^2\Delta t$.

54.3 IEEE 1588 Clock Model

IEEE 1588 standard estimated the clock offset and skew with timestamps in the exchange of packets between master and slave clocks, and realizing synchronization with compensation to local clock. The basic IEEE 1588 synchronization process is shown as Fig. 54.1 [4].

As shown in Fig. 54.1, in the kth synchronization period, master clock firstly send synchronization packet and mark its leaving time as T_k. When this packet arrived at salve clock with the transmission delay D_{ms}, the slave clock would mark its arriving time as $C(T_k + D_{ms})$. Then the master clock would send follow up

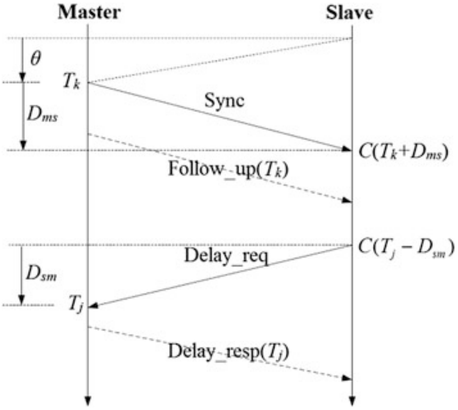


Fig. 54.1 Basic IEEE 1588 synchronization process

packet to send timestamp T_k to slave, finishing the first step in a process. The clock offset could be calculated as formula (54.4) [4]:

$$\theta_M(k) = C(T_k + D_{ms}) - T_k - D_{ms} \tag{54.4}$$

Since that the transmission delay D_{ms} is unknown, so to estimate clock offset there is need the second step in a process. In the second step, the slave clock firstly send delay requirement packet and mark its leaving time as $C(T_j - D_{sm})$. When this packet arrived at master clock with the transmission delay D_{sm}, the master clock would mark its arriving time as T_j. Then the master clock would send back delay response packet to send timestamp T_j to slave, finishing the second step. The transmission delay could be calculated as formula (54.5) [4]:

$$D_{sm} = \theta_M(k) + T_j - C(T_j - D_{sm}) \tag{54.5}$$

IEEE 1588 standard made an assumption that the transmission delay from master to slave is equal to the delay from slave to master, so from formulas (54.4) and (54.5), we can get the observed value of clock offset and clock skew:

$$\theta_M(k) = \frac{[C(T_k + D_{ms}) - T_k] - [T_j - C(T_j - D_{sm})]}{2} \tag{54.6}$$

$$r_M(k) = \frac{\theta_M(k) - \theta_M(k-1)}{T_k - T_{k-1}} \tag{54.7}$$

Both of the observed values would be impacted on by measurement uncertainty. Based on formula (54.6), the uncertainty could be divided into 3 parts: the uncertainty of master clock σ_T^2, the uncertainty of slave clock $\sigma_{C(t)}^2$ and the asymmetry of transmission delay σ_d^2. Since that the asymmetry of transmission delay affects synchronization performance much, the optimization in this field usually would be considered independently and not the key point in this paper. Meanwhile, the master clock usually has so good performance that its uncertainty could be ignored. Thus based on the symmetry assumption of transmission delay, we can deduce the measurement uncertainty of the observed values of clock offset and clock skew:

$$\sigma_{\theta_M}^2 = \frac{1}{2}(\sigma_{C(t)}^2 + \sigma_T^2 + \sigma_d^2) \approx \frac{1}{2}\sigma_{C(t)}^2 \tag{54.8}$$

$$\sigma_{r_M}^2 = 2 \cdot \frac{\sigma_{\theta_M}^2}{(T_k - T_{k-1})^2} = 2(\frac{\sigma_{\theta_M}}{T_{sync}})^2 \tag{54.9}$$

where T_{sync} is the synchronization period of IEEE 1588. Since that clock offset and clock skew are cross-correlation, there is need to consider their covariance:

$$\sigma_{\theta r}^2 = \sigma_{r\theta}^2 = \frac{\sigma_{\theta_M}^2}{T_{sync}} \tag{54.10}$$

IEEE 1588 clock could be compensated according to the estimation value of clock offset and clock skew. So the compensation input need to be introduced into IEEE 1588 clock model. Let $u_\theta(k)$, $u_r(k)$ and $u_a(k)$ be the compensation input for clock offset, clock skew and clock aging rate in the kth synchronization period, although actually the aging rate is the inherent attribute of clock and $u_a(k)$ would be 0. With the introduction of compensation input into basic clock model (54.3) and $\Delta t = T_{sync}$, we can get the IEEE 1588 clock model:

$$\begin{cases} \theta(k+1) = (\theta(k) - u_\theta(k)) + (r(k) - u_r(k))T_{sync} + (a(k) - u_a(k))T_{sync}^2 + \omega_\theta(k) \\ r(k+1) = (r(k) - u_r(k)) + (a(k) - u_a(k))T_{sync} + \omega_r(k) \\ a(k+1) = a(k) - u_a(k) + \omega_a(k) \end{cases} \tag{54.11}$$

The introduction of compensation input is based on assumption that the compensation was effective and immediate.

54.4 IEEE 1588 Clock Synchronization Based on Kalman

The measured offset and skew could be directly applied as corrections to the local clock in order to achieve synchronization. However, in practice, timing information is occasionally inaccurate and needs to be preprocessed by some form of filter. Once a recursive state-variable clock model has been developed, as shown in the previous sections, it becomes almost natural to consider the implementation of a recursive estimator based on Kalman filter equations. For this purpose, the clock state model (54.11) can be written in matrix form:

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k) + \boldsymbol{\omega}(k) \tag{54.12}$$

where $\boldsymbol{x}(k) = [\theta(k), r(k), a(k)]^T$ is the state vector, $\boldsymbol{u}(k) = [u_\theta(k), u_r(k), u_a(k)]^T$ is the input vector and $\boldsymbol{A}$ is state-transition matrix, $\boldsymbol{B}$ is input-control matrix:

$$\boldsymbol{A} = \begin{bmatrix} 1 & T_{sync} & T_{sync}^2 \\ 0 & 1 & T_{sync} \\ 0 & 0 & 1 \end{bmatrix} \tag{54.13}$$

$$\boldsymbol{B} = \begin{bmatrix} -1 & -T_{sync} & -T_{sync}^2 \\ 0 & -1 & -T_{sync} \\ 0 & 0 & -1 \end{bmatrix} \tag{54.14}$$

$\boldsymbol{\omega}(k) = [\omega_\theta(k), \omega_r(k), \omega_a(k)]^T$ is system noise vector with character:

$$\begin{cases} \mathrm{E}\{\boldsymbol{\omega}(k)\} = 0 \\ \mathrm{E}\left\{\boldsymbol{\omega}(k)\boldsymbol{\omega}(j)^T\right\} = \boldsymbol{Q}\delta_{kj} \end{cases} \tag{54.15}$$

where $\boldsymbol{Q}$ is the system noise covariance matrix:

$$\boldsymbol{Q} = \begin{bmatrix} \sigma_\theta^2 T_{sync} & 0 & 0 \\ 0 & \sigma_r^2 T_{sync} & 0 \\ 0 & 0 & \sigma_a^2 T_{sync} \end{bmatrix} \tag{54.16}$$

Likewise, the observation equation could be written with (3.3) and (3.4) as:

$$\boldsymbol{z}(k+1) = \boldsymbol{H}\boldsymbol{x}(k+1) + \boldsymbol{v}(k+1) \tag{54.17}$$

where $\boldsymbol{z}(k) = [\theta_M(k), r_M(k)]^T$ is the observation input vector, $\boldsymbol{H}$ is the observation matrix:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{54.18}$$

$\boldsymbol{v}(k) = [v_\theta(k), v_r(k)]^T$ is the observation noise vector with character:

$$\begin{cases} \mathrm{E}\{\boldsymbol{v}(k)\} = 0 \\ \mathrm{E}\left\{\boldsymbol{v}(k)\boldsymbol{v}(j)^T\right\} = \boldsymbol{R}\delta_{kj} \end{cases} \tag{54.19}$$

where $\boldsymbol{R}$ is the observation noise covariance matrix:

$$\boldsymbol{R} = \begin{bmatrix} \sigma_{\theta_M}^2 & \sigma_{\theta r}^2 \\ \sigma_{r\theta}^2 & \sigma_{r_M}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{\theta_M}^2 & \frac{\sigma_{\theta_M}^2}{T_{sync}} \\ \frac{\sigma_{\theta_M}^2}{T_{sync}} & 2\frac{\sigma_{\theta_M}^2}{T_{sync}^2} \end{bmatrix} \tag{54.20}$$

So finally based on Eqs. (54.12) and (54.17), we can get IEEE 1588 synchronization based on Kalman:

The prediction equations are:

$$\hat{\boldsymbol{x}}(k|k-1) = \boldsymbol{A}\hat{\boldsymbol{x}}(k-1) + \boldsymbol{B}\boldsymbol{u}(k-1) \tag{54.21}$$

$$\boldsymbol{P}(k|k-1) = \boldsymbol{A}\boldsymbol{P}(k-1)\boldsymbol{A}^T + \boldsymbol{Q} \tag{54.22}$$

where $\hat{\boldsymbol{x}}(k|k-1)$ is the one-step state prediction, $\boldsymbol{P}(k|k-1)$ is the apriori error prediction covariance matrix, and $\boldsymbol{P}(k)$ is the aposteriori estimation error covariance. If the $\boldsymbol{Q}$ is known apriori, $\boldsymbol{P}(k|k-1)$ could be initialized as $\boldsymbol{P}(1|0) = \boldsymbol{Q}$.

The correction equations are:

$$\boldsymbol{K}(k) = \boldsymbol{P}(k|k-1)\boldsymbol{H}^T[\boldsymbol{H}\boldsymbol{P}(k|k-1)\boldsymbol{H}^T + \boldsymbol{R}]^{-1} \tag{54.23}$$

$$\boldsymbol{P}(k) = (\boldsymbol{I} - \boldsymbol{K}(k)\boldsymbol{H})\boldsymbol{P}(k|k-1) \tag{54.24}$$

$$\hat{\boldsymbol{x}}(k) = \hat{\boldsymbol{x}}(k|k-1) + \boldsymbol{K}(k)(\boldsymbol{z}(k) - \hat{\boldsymbol{x}}(k|k-1)) \tag{54.25}$$

Finally, the aposteriori estimate $\hat{\boldsymbol{x}}(k)$ provided by the Kalman filter is used to update the local clock, which corresponds to setting $\boldsymbol{u}(k) = \hat{\boldsymbol{x}}(k)$.

54.5 Simulation and Verification

To verify the Kalman-based IEEE 1588 synchronization, this paper built a simulation system on OMNet++, with parameters shown as Table 54.1 decided by the real hardware platform based on STM32F107 [9].

Simulation 1: set slave clock uncertainty as 10^{-8} fixed. The simulation results are shown as Fig. 54.2.

Traditional IEEE 1588 synchronization corrected the local clock with the measured offset and skew directly, meanwhile the clock model could not be modified adaptively according to the clock skew variation. So there is overshoot in synchronization process.

Table 54.1 IEEE 1588 synchronization simulation parameters

Slave clock uncertainty	Slave clock aging rate	Slave clock skew	Synchronization period	Number of synchronization
10^{-8}–10^{-3}	10^{-12}	5×10^{-7} s	1 s	1,000

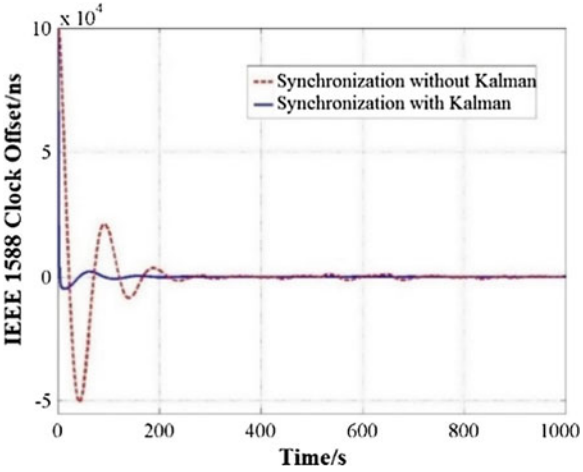


Fig. 54.2 IEEE 1588 synchronization simulation

Table 54.2 Statistical data of IEEE 1588 synchronization performance

Optimization state	Synchronization accuracy (ns)	Standard deviation of clock offset (s)	Number of synchronization
Without Kalman	119.8284	1.3069×10^{-5}	1,000
With Kalman	59.3287	3.4352×10^{-6}	1,000

While IEEE 1588 Kalman-based synchronization could estimate clock offset, skew and aging rate with the iterative running results of formulas (3.18)–(3.22) and correct the local clock with the final estimation value. Since that the clock model could be modified adaptively according to the clock skew variation, the synchronization performance would be better certainly. Table 54.2 just shows the statistical results of the simulation, and the data could verify the effectiveness of the proposed IEEE 1588 Kalman-based synchronization algorithm.

Traditional IEEE 1588 synchronization corrected the local clock with the measured offset and skew directly, meanwhile the clock model could not be modified adaptively according to the clock skew variation. So there is overshoot in synchronization process.

While IEEE 1588 Kalman-based synchronization could estimate clock offset, skew and aging rate with the iterative running results of formulas (3.18)–(3.22) and correct the local clock with the final estimation value. Since that the clock model could be modified adaptively according to the clock skew variation, the synchronization performance would be better certainly. Table 54.2 just shows the statistical results of the simulation, and the data could verify the effectiveness of the proposed IEEE 1588 Kalman-based synchronization algorithm.

Simulation 2: set slave clock uncertainty as 10^{-8}–10^{-3} to verify the effectiveness of attenuating slave clock uncertainty. The simulation results are shown as Fig. 54.3.

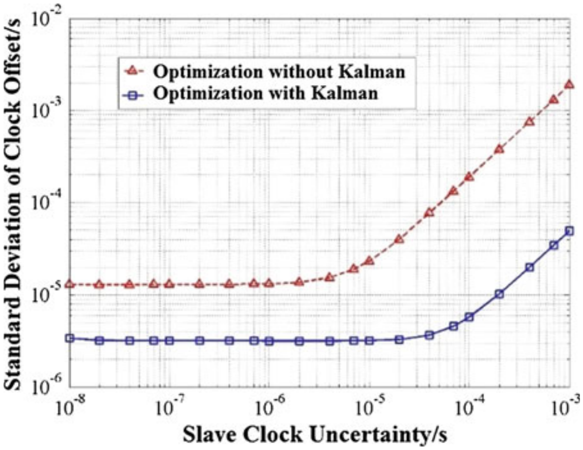


Fig. 54.3 IEEE 1588 synchronization simulation

The slave clock uncertainty decides the observation noise matrix ***R***. With the increase of observation noise, system synchronization performance decreased and standard deviation of clock offset increased. But Kalman-based synchronization algorithm could adjust the gain matrix ***K*** adaptively if observation noise increased. So the modification from input value would be controlled and the synchronization performance would be enhanced.

The slave clock uncertainty decides the observation noise matrix ***R***. With the increase of observation noise, system synchronization performance decreased and standard deviation of clock offset increased. But Kalman-based synchronization algorithm could adjust the gain matrix ***K*** adaptively if observation noise increased. So the modification from input value would be controlled and the synchronization performance would be enhanced.

Figure 54.3 also shows that the synchronization performance would be retained when the slave clock uncertainty was smaller than 1 μs. However, when the slave clock uncertainty was bigger than 1 μs, the synchronization performance with no optimization would be deteriorated visibly, on the contrary the deterioration process of synchronization performance with optimization was attenuated. The deterioration starts from uncertainty 2×10^{-5} s, compared with 2×10^{-6} s without optimization, optimized with almost one order.

The simulation results could verify that IEEE 1588 Kalman-based synchronization algorithm could attenuate clock uncertainty and enhance synchronization performance.

54.6 Conclusions

To attenuate IEEE 1588 synchronization performance deterioration caused by clock uncertainty, this paper modeled a 3-states IEEE 1588 clock model and put forward IEEE 1588 Kalman-based synchronization algorithm based on this 3-states model. This algorithm introduced clock skew as new input vector, with clock offset as traditional input, realizing the correction of both clock offset and skew. The introduction of clock skew as new input vector strengthened the stability of IEEE 1588 synchronization.

Verified by some meaningful simulations, the IEEE 1588 Kalman-based synchronization algorithm could enhance the synchronization performance, with accuracy almost double optimized, the stability and the deterioration inflection point optimized with almost one order. The model and algorithm could help realize good IEEE 1588 synchronization, even with big clock uncertainty.

References

1. Tong BR (2003) Timing system. National Defense Industry Press, Beijing
2. Deeths D (2001) Using NTP to control and synchronize system clocks, system clocks. Sun Microsystems, Inc.
3. IEEE 1588-2002 standard (2002) IEEE 1588 Standard for a precision clock synchronization protocol for networked measurement and control systems
4. IEEE 1588-2008 standard (2008) IEEE 1588 standard for a precision clock synchronization protocol for networked measurement and control systems
5. Sungwon L, Seunggwan L, Choongseon H (2012) An accuracy enhanced IEEE 1588 synchronization protocol for dynamically changing and asymmetric wireless links. IEEE Commun Lett 16(2):190–192
6. Takahide M, Yukio H (2009) Improvement of synchronization accuracy in IEEE 1588 using a queuing estimation method. In: IEEE symposium on precision clock synchronization for measurement, control and communication, Brescia, Italy, pp 12–16
7. Giada G, Claudio N (2011) Performance analysis of Kalman-filter-based clock synchronization in IEEE 1588 networks. IEEE Trans Instrum Meas 60(8):2902–2909
8. Zhuang XY, Wang HH (2012) IEEE 1588 clock synchronization algorithm based on Kalmanfilter. J Electron Meas Instrum 26(9):747–751
9. ST-AN3411 (2011) IEEE 1588 precision time protocol demonstration for STM32F107 connectivity line microcontroller. http://www.st.com. Cited July 11, 2011