MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

Detection of auroral breakups using the correlation dimension

A. Kawaguchi, K. Kitamura
K. Yonemoto, T. Yanagawa
K. Yumoto

MHF 2004-5

(Received March 8, 2004)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Detection of auroral breakups using the correlation dimension

Atsushi Kawaguchi[*], Kentarou Kitamura[†], Koji Yonemoto[‡], Takashi Yanagawa[§], and Kiyofumi Yumoto[¶]

Abstract

In this paper, we develop the method for estimating the correlation dimension proposed by Kawaguchi (2002) and apply it to geomagnetic pulsations. It is found that the estimated values of the correlation dimension tend to drop at auroral breakups. This is developed into the detection of auroral breakups. At first, using the data that is easy to detect visually auroral breakups by the waveform, the training data is expressed with the pair of the estimates of the correlation dimension and the label which codes as +1 for auroral breakups and −1 for non-auroral breakups. Next, we construct the discriminant function by the support vector machine using the training data. Using the training data, the error for the discrimination is computed. The error is found to be 11.3%. This suggests the usefulness of the estimates of the correlation dimension in the discrimination.

Key Words: Correlation Dimension, Pi2 Pulsation, Support Vector Machine.

1 Introduction

Geomagnetic pulsations which have irregular and continuous waveform are called Pi-type and Pc-type, respectively. The Pi-type and Pc-type pulsations are classified into some classes according to their period. The Pi-type pulsations with period ranges of 40-150 seconds are denoted as Pi2 pulsations, and the Pc-type pulsations with period ranges of 45-150 seconds are denoted as Pc4 pulsations (Saito, 1969). The Pi2 pulsations appear to occur in one-to-one correspondence with auroral breakups (see for example, Yumoto, 2001).

As seen in Figure 1.1, the frequency decomposition by Fourier analysis make it easy to detect visually Pi2 pulsation on 12-24 UT (Universal Time), say nightside. On the other hand, The magnetosphere on 0-12UT, say dayside, is generally an active region for

[*]Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Fukuoka 812-8581, JAPAN
[†]Space Environment Research Center, Kyushu University, 6-10-1 Hakozaki, Fukuoka 812-8581, JAPAN
[‡]Graduate School of Medical Sciences, Kyushu University, Health C&C Center, Kubara 1822-1, Hisayama Town, Kasuya-gun, Fukuoka 811-2501, JAPAN
[§]Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Fukuoka 812-8581, JAPAN
[¶]Space Environment Research Center, Kyushu University, 6-10-1 Hakozaki, Fukuoka 812-8581, JAPAN

Pc4 pulsations. These Pc4 pulsations mask Pi2 activities and make it difficult to find visually the Pi2 pulsations on the dayside (Nosè, et al., 2003). Our goal is to detect the Pi2 pulsations on the dayside.

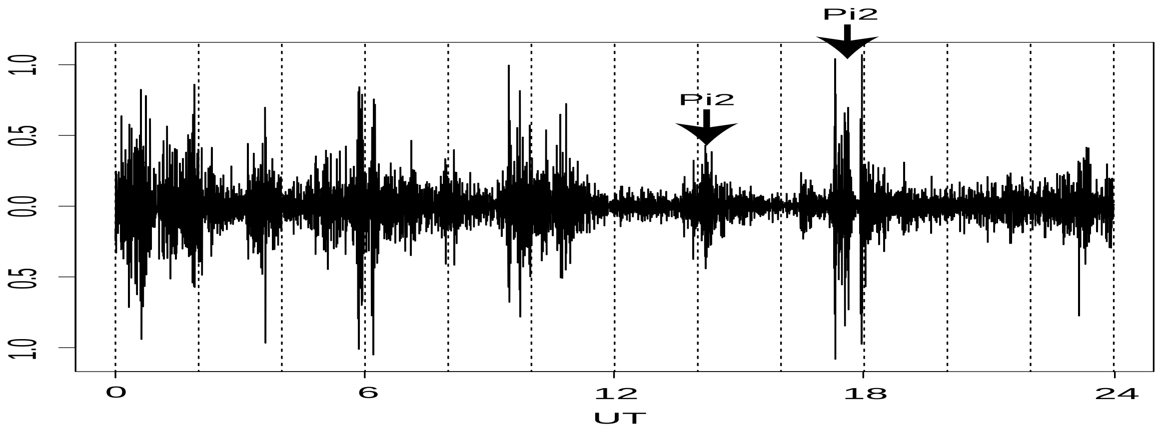

Figure 1.1: The magnetic field component measured at Kuju in Japan on 0-24UT March 23, 2003

This paper is organized as follows. In Section 2, it is shown the relation between the correlation dimension and the Pi2 pulsations. We introduce support vector machine and describe methods of discrimination in Section 3. In Section 4, we apply the procedure to data acquired from the Circum-pan Pacific Magnetometer Network (CPMN) (Yumoto, 2001) and show the results.

2 The correlation dimension and Pi2 pulsation

We recall the correlation dimension estimator proposed by Kawaguchi(2002). For one-dimensional time series $\{X_t\}_{t=1}^N$, putting $Y_t = (X_t, X_{t+1}, \ldots, X_{t+(d-1)})$ and

$$C_N(r) = \binom{N}{2}^{-1} \sum_{i<j}^{N} I(\|Y_i - Y_j\| \leq r),$$

where d is a positive integer called embedding dimension, I denotes an indicator function, and $\|\cdot\|$ is a norm, Grassberger and Procaccia (1983a,b) called $C(r) = \lim_{N\to\infty} C_N(r)$ the correlation integral and introduced the correlation dimension as

$$p = \lim_{r\to 0} \frac{\log C(r)}{\log r}$$

if the limit exists.

Let $i_1 = [M/2]$, $i_2 = [M/2]+1$, ... , $i_L = M$, where $M = \max\{m\ ; C_{N2}(r_m) \neq 0\}$,

$$C_{N2}(r) = \binom{N}{3}^{-1} \sum_{i\neq j, i\neq k, j\neq k}^{N} I(\|Y_i - Y_j\| \leq r, \|Y_k - Y_j\| \leq r)$$

for some $r > 0$, and $L = M - [M/2] + 1$. Putting $u_m = \log r_m$, $v_m = \log C_N(r_m)$, $r_m = r_0\phi^m$, $\bar{u} = L^{-1}\sum_{m=i_1}^{i_L} u_m$, and $\bar{v} = L^{-1}\sum_{m=i_1}^{i_L} v_m$, the estimator of p is given by

$$\hat{p} = \left(\sum_{m=i_1}^{i_L}(u_m - \bar{u})(v_m - \bar{v})\right) \Bigg/ \left(\sum_{m=i_1}^{i_L}(u_m - \bar{u})^2\right). \quad (2.1)$$

In the top panel of Figure 2.1, the bandpass-filtered (40-150s) geomagnetic field data for the period 12-24UT (nightside) on 8 March 2003, are plotted. The data set consists of 43200 data points. Our procedure for estimating the correlation dimension from the same data is as follows. Denoting the data by $\{X_1, X_2, \ldots, X_{43200}\}$, we constructed the data set i of size $N = 1800$ as $\{X_{300(i-1)+1}, X_{300(i-1)+2}, \ldots, X_{300(i-1)+1800}\}$, $(i = 1, 2, \ldots, 139)$. Setting embedding dimension $d = 5$, we estimated the correlation dimension using estimator (2.1) from each data set i.

The bottom panel of Figure 2.1 exhibits the result of computation, where UT is shown in the horizontal line and value of the estimated correlation dimension in the vertical line.

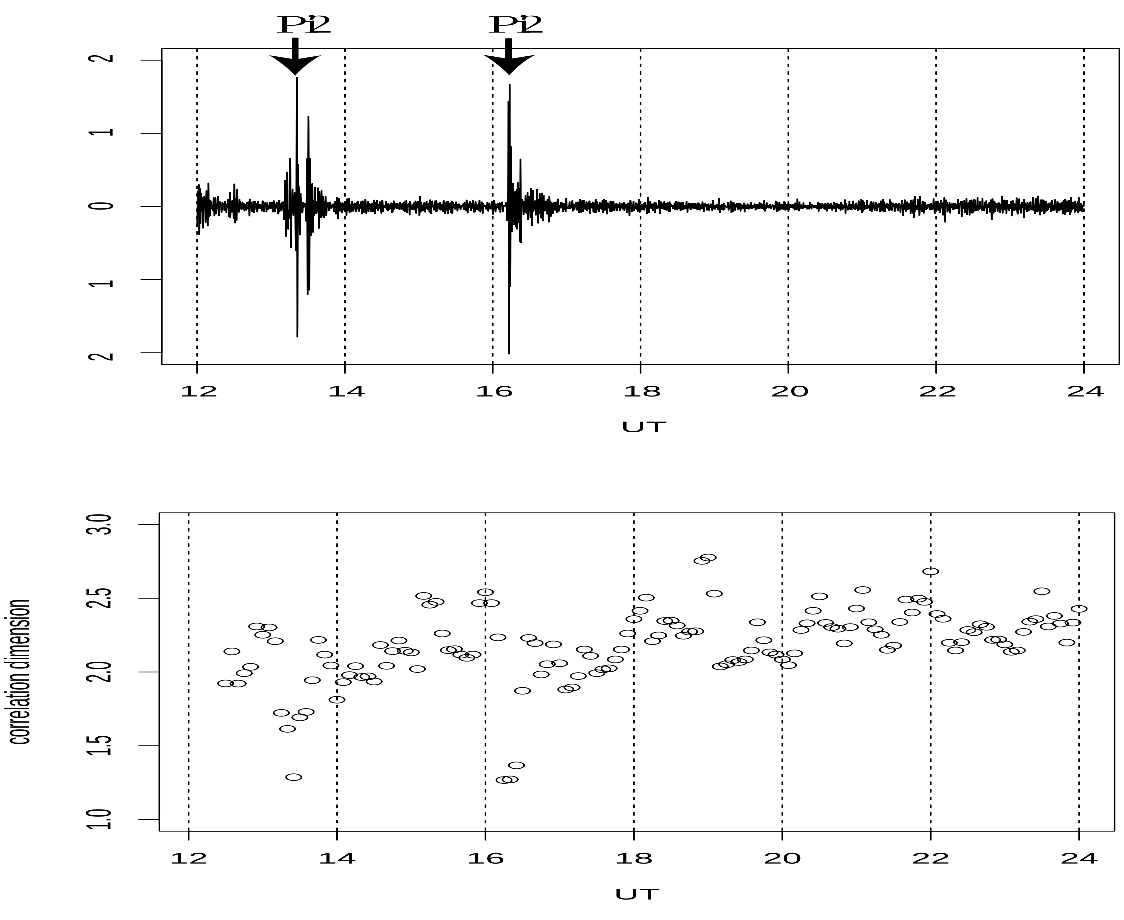

Figure 2.1: The magnetic field component measured at Kuju in Japan on 12-24UT March 8, 2003 (Top) The estimates of the correlation dimension (Bottom)

Figure 2.1 shows that the Pi2 pulsation corresponds with the drop of the correlation dimension estimates. Using this finding, a procedure for the discrimination of Pi2 pulsations is given as follows. At first, we construct the discriminant function based on the correlation dimension estimated from the nightside data. Next, using the constructed discriminant function, we detect the Pi2 pulsations from the dayside data.

3 Methods for the discrimination

3.1 Support vector machine

We use support vector machine (SVM) (Vapnick, 1998) as a method of discrimination. Suppose that we are given two classes Class 1 and Class 2, and l pairs $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)$, called the training data, with $\mathbf{x}_i \in \mathbf{R}^q$, y_i codes as 1 for Class 1 and -1 for Class 2. Discrimination by SVM is based on the function given as

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{l} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right), \tag{3.1}$$

where $\alpha_i \in \mathbf{R}$, $b \in \mathbf{R}$, and $K(\mathbf{x}, \mathbf{y})$ is a kernel. We use Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = \exp(-\sigma\|\mathbf{x} - \mathbf{y}\|^2)$ where $\|\cdot\|$ is a norm.

It is well known that $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l)^T$ is given by solving the quadratic optimization problem as follows:

$$\begin{aligned} &\min_{\boldsymbol{\alpha}} \left\{\frac{1}{2}\boldsymbol{\alpha}^T Q \boldsymbol{\alpha} + \mathbf{e}^T \boldsymbol{\alpha}\right\} && (3.2) \\ &\text{subject to} \quad 0 \leq \alpha_i \leq C, \quad (i = 1, 2, \ldots, l) \\ &\qquad\qquad\quad \mathbf{y}^T \boldsymbol{\alpha} = 0, \end{aligned}$$

where $\mathbf{e} = (1, 1, \ldots, 1)^T$, $C > 0$ is a constant, $\mathbf{y} = (y_1, y_2, \ldots, y_l)^T$, and Q is a $l \times l$ positive semidefinite matrix, $Q_{i,j} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$. Denote by $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_l)^T$ the solution of (3.2). Once $\hat{\boldsymbol{\alpha}}$ is obtained, the solution for b in (3.1) is given as follows,

$$\hat{b} = \frac{1}{l} \sum_{i=1}^{l} \left\{ y_i - \sum_{j=1}^{l} \hat{\alpha}_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \right\}.$$

With $\hat{\boldsymbol{\alpha}}$ and $\hat{b}$, a new data $\mathbf{x}' \in \mathbf{R}^q$ is discriminated as

$$\mathbf{x}' \in \begin{cases} \text{Class 1}, & \text{if } \hat{f}(\mathbf{x}') = 1 \\ \text{Class 2}, & \text{if } \hat{f}(\mathbf{x}') = -1 \end{cases},$$

where

$$\hat{f}(\mathbf{x}) = sign\left(\sum_{i=1}^{l} y_i \hat{\alpha}_i K(\mathbf{x}_i, \mathbf{x}) + \hat{b}\right). \tag{3.3}$$

The data $\mathbf{x}'$ is called the test data. The details of the training data and the test data is given as follows.

3.2 Training data

The training data is obtained by the estimates of the correlation dimension from nightside (12-24UT) data. For the entire data set, consisting of 43200 data points, say $\{X_{1,1}, X_{1,2}, \ldots, X_{1,43200}\}$, we constructed the data set i of size $N = 1800$ as $\{X_{1,300(i-1)+1}, X_{1,300(i-1)+2}, \ldots, X_{1,300(i-1)+1800}\}$ $(i = 1, 2, \ldots, 139)$.

Setting embedding dimension $d = 5$, we estimated the correlation dimension using the estimator (2.1) from each data set i. Denote by $\hat{p}_{1,i}$ the estimates of the correlation dimension.

Using the data that is easy to judge visually whether the data set contains Pi2 pulsations or not, we set our training data as follows. For some given integer η, put

$$\begin{aligned} \mathbf{x}_i &= (\hat{p}_{1,i-\eta},\ \hat{p}_{1,i},\ \hat{p}_{1,i+\eta}) \qquad \text{and} \\ y_i &= \begin{cases} 1 & \text{if the data set } i \text{ contains Pi2 pulsations} \\ -1 & \text{if not.} \end{cases} \end{aligned}$$

Then the training data consists of $l - 2\eta$ pairs $(\mathbf{x}_{1+\eta}, y_{1+\eta})$, $(\mathbf{x}_{2+\eta}, y_{2+\eta})$, $\ldots$, $(\mathbf{x}_{l-\eta}, y_{l-\eta})$.

Note that in order to obtain α from (3.2) based on the training data we must decide the values of C, σ, and η at first. We apply Cross-Validation criterion to decide the value of C, σ, and η. Namely, by putting

$$CV(C, \sigma, \eta) = \frac{1}{l - 2\eta} \sum_{j=1+\eta}^{l-\eta} (y_j - \hat{f}^{\backslash j}(\mathbf{x}_j, C, \sigma, \eta))^2,$$

where $\hat{f}^{\backslash j}(\mathbf{x}, C, \sigma, \eta)$ denote the discriminant function estimator computed with the j-th training data $(\mathbf{x}_j, y_j)$ removed with some given C, σ, and η, we selected parameters $(\hat{C}, \hat{\sigma}, \hat{\eta})$ in a class of $C \in \Theta_C$, $\sigma \in \Theta_\sigma$, and $\eta \in \Theta_\eta$ that minimizes $CV(C, \sigma, \eta)$ where Θ_C, Θ_σ, and Θ_η denote parameter sets.

3.3 Test data

The test data is obtained by the estimates of the correlation dimension from the dayside (0-12UT) data. For the entire data set, consisting of 43200 points, say $\{X_{2,1}, X_{2,2}, \ldots, X_{2,43200}\}$, we constructed the data set i of size $N = 1800$ as $\{X_{2,300(i-1)+1}, X_{2,300(i-1)+2}, \ldots, X_{2,300(i-1)+1800}\}$ $(i = 1, 2, \ldots, 139)$.

Setting embedding dimension $d = 5$, we estimated the correlation dimension using the estimator (2.1) from each data set i. Denote by $\hat{p}_{2,i}$ the estimates of the correlation dimension.

Put $\mathbf{x}'_i = (\hat{p}_{2,i-\hat{\eta}},\ \hat{p}_{2,i},\ \hat{p}_{2,i+\hat{\eta}})$ where $\hat{\eta}$ is a selected parameter by the method mentioned above. Then the test data consists of $l - 2\hat{\eta}$ points $\mathbf{x}'_{1+\hat{\eta}}$, $\mathbf{x}'_{2+\hat{\eta}}$, $\ldots$, $\mathbf{x}'_{l-\hat{\eta}}$.

In order to detect Pi2 pulsation on dayside data, we applied the discriminant function (3.3) which was constructed by the same day's nightside data.

3.4 Applying to the data

We searched for Pi2 pulsations from the geomagnetic horizontal field, bandpass-filtered (40-150s), at Kuju (Local Time = UT + 9hr). The time resolution of data is 1 second. The data acquired from the Circum-pan Pacific Magnetometer Network (CPMN) (Yumoto, 2001).

Parameter sets Θ_C, Θ_σ, and Θ_η were set as $\Theta_C = \{0.25, 0.50, 1.00, 2.00, 4.00, 8.00, 16.00\}$, $\Theta_\sigma = \{0.03, 0.06, 0.13, 0.25, 0.50, 1.00, 2.00, 4.00\}$, and $\Theta_\eta = \{1, 2, \ldots, 10\}$, respectively.

In order to show the performance of our method for discrimination, we introduced the training error rate as follows,

$$\overline{err} = \frac{1}{l - 2\hat{\eta}} \sum_{j=1+\hat{\eta}}^{l-\hat{\eta}} I(y_j \neq \hat{f}(\mathbf{x}_j)).$$

4 Results

The procedure is applied to the data which consists of 19 days from March to June 2003. The discriminant function is constructed each day's nightside. Table 4.1 shows the training error rate and the estimated parameters.

Table 4.1: Error rates and estimated parameters

Date	Error rates	$\hat{\sigma}$	$\hat{C}$	$\hat{\eta}$
3/3	14.7%	2	8	5
3/8	4.1%	0.5	4	9
3/10	6.1%	1	8	4
3/23	17.6%	0.25	8	7
3/25	18.4%	1	2	7
3/29	8.7%	2	2	6
4/6	8.8%	0.13	2	7
4/9	12.6%	0.25	2	10
4/12	20.6%	0.25	4	4
4/22	14.6%	0.13	16	8

Date	Error rates	$\hat{\sigma}$	$\hat{C}$	$\hat{\eta}$
4/23	9.4%	1	8	6
4/29	22.8%	0.5	2	8
5/17	10.9%	2	2	5
5/25	3.1%	4	4	5
5/26	4.7%	1	8	6
6/13	15.7%	0.13	2	6
6/15	3.9%	2	16	5
6/21	12.0%	0.5	16	7
6/22	5.5%	2	4	6

Note that the total error rate is 11.3%. This and Table 4.1 suggest the usefulness of the correlation dimension estimates in the discrimination of Pi2 pulsations.

Table 4.2 indicates a range of time containing Pi2 pulsations on dayside data detected by our method.

Table 4.2: The time range identified Pi2 by our method

3/3	0:25-4:10	4:45-6:45	7:25-11:30		
3/8	1:45-2:15	6:15-7:00			
3/10	3:45-4:20	7:00-8:05	9:50-10:25		
3/23	1:10-3:15	5:25-6:40	7:20-8:40	8:50-10:00	10:30-11:25
3/25	0:55-1:25	1:45-2:30	4:55-5:50	6:35-7:05	7:55-8:25
	8:00-8:50	9:20-10:10			
3/29	0:30-5:25	5:40-11:00			
4/6	4:45-5:25				
4/9	1:10-1:40	1:50-2:25	6:50-7:20	9:45-10:40	
4/12	1:55-2:25	3:55-4:40			
4/22	1:15-1:50	2:15-2:50	3:45-4:15	4:20-5:10	7:20-7:50
	8:55-10:00	10:20-10:55			
4/23	0:30-5:15	5:20-9:40	9:45-10:40	10:45-11:15	
4/29	0:40-11:15				
5/17	0:45-1:15	1:30-2:35	3:25-3:55	4:05-5:40	6:15-6:45
	7:30-9:05	10:30-10:40	11:05-11:35		
5/25	2:35-3:25	6:35-7:05	10:45-11:30		
5/26	8:10-8:40	10:20-11:05			
6/13	0:30-6:15	7:10-7:40	8:40-10:15		
6/15	0:25-11:30				
6/21	0:45-1:20	2:25-3:20	4:40-5:10	6:30-7:25	8:30-9:00
	10:45-11:25				
6/22	0:50-1:25	1:40-2:40	2:45-3:15	3:50-5:15	6:25-6:50
	7:00-11:25				

References

[1] Grassberger, P. and Procaccia, I., Characterization for Strange Attractors, Physical Review Letters, 50(5) 346-349, 1983a.

[2] Grassberger, P. and Procaccia, I., Measuring the strangeness of strange attractors, Physica D, 9(5) 189-208, 1983b

[3] Kawaguchi, A., Estimating the correlation dimension from chaotic dynamical systems by U-statistics. Bulletin of Informatics and Cybernetics, 34(2) 143-150, 2002.

[4] Nosè, M., Iyemori, T., Takeda, M., Kamei, T., Milling, D.K., Orr, D., Singer, H.J., Worthington, E.W., Sumitomo, N., Automated detection of Pi2 pulsations using wavelet analysis: 1. Method and an application for substorm monitoring, *Earth Planets and Space*, **50**, 773-783, 1998.

[5] Nosè, M., Takahashi, K., Uozumi, T., Yumoto, K., Miyoshi, Y., Morioka, A., Milling, D.K., Sutcliffe, P.R., Matsumoto, H., Goka, T., Nakata, H., Multipoint observations of a Pi2 pulsation on morningside: The 20 September 1995 event, *Journal of Geographsical Research*, **108**, No.A5, 1219, 2003.

[6] Saito, T., Geomagnetic pulsations, *Space Science Reviews*, **10**, 319-412, 1969.

[7] Vapnick, V., *Statistical learning theory*, New York: Wiley, 1998

[8] Yumoto, K., Characteristics of Pi 2 magnetic pulsations observed at the CPMN stations: A review of the STEP results, *Earth Planets and Space*, **53**, No.10, 981-992, 2001.

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF

2003-1 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Yoshitaka WATANABE
A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems

2003-2 Masahisa TABATA & Daisuke TAGAMI
Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients

2003-3 Tomohiro ANDO, Sadanori KONISHI & Seiya IMOTO
Adaptive learning machines for nonlinear classification and Bayesian information criteria

2003-4 Kazuhiro YOKOYAMA
On systems of algebraic equations with parametric exponents

2003-5 Masao ISHIKAWA & Masato WAKAYAMA
Applications of Minor Summation Formulas III, Plücker relations, Lattice paths and Pfaffian identities

2003-6 Atsushi SUZUKI & Masahisa TABATA
Finite element matrices in congruent subdomains and their effective use for large-scale computations

2003-7 Setsuo TANIGUCHI
Stochastic oscillatory integrals - asymptotic and exact expressions for quadratic phase functions -

2003-8 Shoki MIYAMOTO & Atsushi YOSHIKAWA
Computable sequences in the Sobolev spaces

2003-9 Toru FUJII & Takashi YANAGAWA
Wavelet based estimate for non-linear and non-stationary auto-regressive model

2003-10 Atsushi YOSHIKAWA
Maple and wave-front tracking — an experiment

2003-11 Masanobu KANEKO
On the local factor of the zeta function of quadratic orders

2003-12 Hidefumi KAWASAKI
Conjugate-set game for a nonlinear programming problem

2004-1 Koji YONEMOTO & Takashi YANAGAWA
Estimating the Lyapunov exponent from chaotic time series with dynamic noise

2004-2 Rui YAMAGUCHI, Eiko TSUCHIYA & Tomoyuki HIGUCHI
State space modeling approach to decompose daily sales of a restaurant into time-dependent multi-factors

2004-3 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Cubic pencils and Painlevé Hamiltonians

2004-4 Atsushi KAWAGUCHI, Koji YONEMOTO & Takashi YANAGAWA
Estimating the correlation dimension from a chaotic system with dynamic noise

2004-5 Atsushi KAWAGUCHI, Kentarou KITAMURA, Koji YONEMOTO, Takashi YANAGAWA & Kiyofumi YUMOTO
Detection of auroral breakups using the correlation dimension