

Estimation of Protein Networks for Cell Cycle in Yeast Based on Least-squares Method Using Periodic Signals

Noriko Takahashi, Takehito Azuma and Shuichi Adachi

Abstract—In this paper, a new approach to estimation problems of protein networks is proposed for systems biology. Generally, it is difficult to estimate complicated networks in molecular biology. Then, in order to estimate complicated networks systematically, it is considered to estimate the networks based on a control engineering method. Considering that wave patterns of proteins are periodic, the protein networks are estimated by the least-squares estimation method. In this method, the networks can be estimated by using just 1 cycle data of protein concentrations. Moreover, this method is applied to an estimation problem of protein networks for cell cycle in yeast, and 9-dimensional protein networks are actually estimated.

I. INTRODUCTION

Molecular biology is based on an idea to understand basic components of life such as genes, m-RNAs and proteins. By developments in this field, various researches on some genes that cause diseases have been reported [1]. These researches might lead to establishments of efficient cure methods or to developments of medicines, so those are certainly important. By molecular biology, however, it is difficult to understand dynamic behaviors of life phenomena. Then, systems biology, which is based on an idea of regarding life as a system, was proposed [2][3] and various researches have been carried out [4][5].

One of the important themes in systems biology is cell cycle, which is the basic part of activities of cells and is known to be closely concerned with mechanisms of aging, cancer, and so on. In the researches for this cell cycle, many experiments about cell cycle in yeast have been performed. The reason is that in spite of being a unicellular organism, yeast's mechanisms of basic life phenomena are similar to that of multicellular organisms including humans. And as for yeast, there are budding yeast and fission one, whose cell cycle mechanisms are different each other. Especially for the budding yeast, various experimental results have been reported and some protein networks representing chemical reactions have been shown [6][7][8]. However, not all the networks in budding yeast have been discovered, so there can be still unknown proteins and networks. It is the protein networks that control cell cycle to go on regularly. So, it is desired to estimate these protein networks. For an estimation of protein networks, there was the experimental method in molecular biology. By using this method, however, it is

difficult to estimate complicated networks because being subject to the constraints of facilities where experiments are preformed.

Then, in order to estimate complicated networks for cell cycle in budding yeast, networks are estimated by using a control engineering method based on systems biology. In some researches, Principal Component Analysis is employed as a control engineering method [9], but it is difficult to estimate the structure of networks by it. Then, in this paper, the estimation method of protein networks based on the least-squares method using periodic signals is proposed. In the proposed method, considering that protein concentrations can be obtained by experiments, a linear system represented by state-space model is identified. Then, protein networks can be estimated by considering that each element in the identified matrix means the combination of proteins. Moreover, considering that wave patterns of protein concentrations are periodic, it will be shown that the networks can be estimated by using just 1 cycle data of protein concentrations and that it is not necessary to have large number of data for system identification. Then, this method is applied to an estimation problem of protein networks for cell cycle in yeast, and the efficacy of the proposed method will be illustrated.

II. LEAST-SQUARES METHOD FOR PERIODIC SIGNALS

In general, the purpose of system identification is to build a mathematical model of the object by using input and output data [10]. However, because cell cycle in budding yeast is an autonomous system, there is no external input data and protein concentrations can be obtained. Therefore, it is necessary to identify the object by using only output data. Then, in this section, system identification based on the least-squares method for state-space models is used by assuming that state variables are known.

Now, let us consider the following discrete periodic signal

$$\mathbf{x}(k) = \mathbf{x}(k + N), \quad k = 1, 2, \dots,$$

where $\mathbf{x}(k) \in R^n$. Consider the problem of estimating the matrix $\mathbf{A}(k)$ in the linear system:

$$\mathbf{x}(k + 1) = \mathbf{A}(k)\mathbf{x}(k), \quad k = 1, 2, \dots \quad (1)$$

When $\mathbf{Y}(k)$, $\Phi(k)$ and $\Theta(k)$ are written as

$$\begin{aligned} \mathbf{Y}(k) &= \mathbf{x}(k + 1) \\ \Phi(k) &= \mathbf{x}(k) \\ \Theta(k) &= \mathbf{A}(k), \end{aligned}$$

N. Takahashi and S. Adachi is with Faculty of Science and Technology, Keio University, Yokohama, Japan noriko_takahashi@z8.keio.jp

T. Azuma is with Graduate school of Engineering, Utsunomiya University, Utsunomiya, Japan tazuma@cc.utsunomiya-u.ac.jp

(1) is rewritten as the linear regression equation:

$$\mathbf{Y}(k) = \Theta(k)\Phi(k). \quad (2)$$

Because $x(k)$ is the discrete periodic signal whose period is N , $\mathbf{Y}(k)$, $\Phi(k)$ and error signal $e(k)$,

$$e(k) = \mathbf{Y}(k) - \Theta(k)\Phi(k),$$

are also periodic that satisfy

$$\begin{aligned} \mathbf{Y}(k) &= \mathbf{Y}(k + N), \quad k = 1, 2, \dots \\ \Phi(k) &= \Phi(k + N), \quad k = 1, 2, \dots \\ e(k) &= e(k + N), \quad k = 1, 2, \dots \end{aligned}$$

It is well-known that the least-squares method can be applied to the linear regression equation. When data of N samples are collected in (1), the least-squares estimated value $\hat{\Theta}(N)$ in (2) is derived as

$$\begin{aligned} \hat{\Theta}(N) &= \arg \min_{\Theta} J_1 \\ &= \left(\sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^N \Phi(k)\Phi^T(k) \right)^{-1}, \quad (3) \end{aligned}$$

where J_1 is the criterion

$$J_1 = \frac{1}{N} \sum_{k=1}^N \|e(k)\|^2. \quad (4)$$

When collected data are increased, the estimated value $\hat{\Theta}(N)$, $\hat{\Theta}(N+1)$, \dots , $\hat{\Theta}(2N)$ are written as

$$\begin{aligned} \hat{\Theta}(N) &= \left(\sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^N \Phi(k)\Phi^T(k) \right)^{-1} \\ \hat{\Theta}(N+1) &= \left(\sum_{k=1}^{N+1} \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^{N+1} \Phi(k)\Phi^T(k) \right)^{-1} \\ &\vdots \\ \hat{\Theta}(2N) &= \left(\sum_{k=1}^{2N} \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^{2N} \Phi(k)\Phi^T(k) \right)^{-1} \\ &= \left(2 \sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) \right) \left(2 \sum_{k=1}^N \Phi(k)\Phi^T(k) \right)^{-1} \\ &= \left(\sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^N \Phi(k)\Phi^T(k) \right)^{-1} \\ &= \hat{\Theta}(N). \end{aligned}$$

This indicates that when the number of data for the least-squares method varies in the range from N to $2N$, the estimated value also varies depending on time k . And the estimated value is the same value when the number of data is increased by N (N is the number of 1 cycle data).

Then, considering that $\mathbf{Y}(k)$ and $\Phi(k)$ are also periodic signals, the next theorem is derived.

Theorem 1: In the linear regression system

$$\mathbf{Y}(k) = \Theta(k)\Phi(k), \quad (5)$$

if it is assumed that $\mathbf{Y}(k)$ and $\Phi(k)$ are periodic signals whose periods are N and that $\Theta(k)$ is the unknown parameter, the following two equations will be derived .

$$\begin{aligned} \text{(a)} \quad & \hat{\Theta}(nN) = \hat{\Theta}(N), \quad n = 2, 3, \dots \\ \text{(b)} \quad & \lim_{M \rightarrow \infty} \hat{\Theta}(M) = \hat{\Theta}(N) \end{aligned}$$

Proof: First, let us prove that (a) is true. Let J_n be the criterion for n cycle data, nN ($n = 2, 3, \dots$). Because $\mathbf{Y}(k)$ and $\Phi(k)$ are periodic signals, J_n is rewritten as follows:

$$\begin{aligned} J_n &= \frac{1}{nN} \sum_{k=1}^{nN} \|e(k)\|^2 \\ &= \frac{1}{nN} \left\{ \sum_{k=1}^N \|e(k)\|^2 + \sum_{k=N+1}^{2N} \|e(k)\|^2 \right. \\ &\quad \left. + \dots + \sum_{k=(n-1)N+1}^{nN} \|e(k)\|^2 \right\} \\ &= \frac{1}{nN} n \sum_{k=1}^N \|e(k)\|^2 \\ &= \frac{1}{N} \sum_{k=1}^N \|e(k)\|^2 \\ &= J_1. \quad (6) \end{aligned}$$

And the estimated value $\hat{\Theta}(nN)$ that minimizes J_n is given as

$$\hat{\Theta}(nN) = \left(\sum_{k=1}^{nN} \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^{nN} \Phi(k)\Phi^T(k) \right)^{-1}. \quad (7)$$

Because $\mathbf{Y}(k)$ and $\Phi(k)$ are periodic signals, $\hat{\Theta}(nN)$ is rewritten as follows and then it is proved that (a) is true.

$$\begin{aligned} \hat{\Theta}(nN) &= \left(\sum_{k=1}^{nN} \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^{nN} \Phi(k)\Phi^T(k) \right)^{-1} \\ &= \left(n \sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) \right) \left(n \sum_{k=1}^N \Phi(k)\Phi^T(k) \right)^{-1} \\ &= \left(\sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^N \Phi(k)\Phi^T(k) \right)^{-1} \\ &= \hat{\Theta}(N) \quad (8) \end{aligned}$$

Next, take a look at (b). The estimated value $\hat{\Theta}(M)$ for M being large enough, is given as

$$\hat{\Theta}(M) = \left(\sum_{k=1}^M \mathbf{Y}(k)\Phi^T(k) \right) \left(\sum_{k=1}^M \Phi(k)\Phi^T(k) \right)^{-1}. \quad (9)$$

When M is written as $M = nN + N_1$ ($N_1 = 1, 2, \dots, N-1$), the following two equations are derived:

$$\sum_{k=1}^{nN+N_1} \mathbf{Y}(k)\Phi^T(k) = n \sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) + \sum_{k=1}^{N_1} \mathbf{Y}(k)\Phi^T(k) \quad (10)$$

$$\sum_{k=1}^{nN+N_1} \Phi(k)\Phi^T(k) = n \sum_{k=1}^N \Phi(k)\Phi^T(k) + \sum_{k=1}^{N_1} \Phi(k)\Phi^T(k). \quad (11)$$

In (10) and (11), if n is large enough, the first term becomes too smaller than the second term. Therefore, the following equation is derived:

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{\Theta}(nN + N_1) \\ &= \left(n \sum_{k=1}^N \mathbf{Y}(k)\Phi^T(k) \right) \left(n \sum_{k=1}^N \Phi(k)\Phi^T(k) \right)^{-1} \\ &= \hat{\Theta}(N). \end{aligned} \quad (12)$$

Here it is assumed that $M = nN + N_1$, so if $n \rightarrow \infty$, $M \rightarrow \infty$. Then it is proved that (b) is true. ■

Remark 1: In Theorem 1(b), when estimating the unknown parameter by using the least-squares method for periodic signals $\mathbf{Y}(k)$ and $\Phi(k)$, $\hat{\Theta}(k)$ converges to the estimated value for 1 cycle data, $\hat{\Theta}(N)$. So, in the problem of identifying the linear system

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k), \quad (13)$$

at least 1 cycle data is needed and it is necessary to have large number of data. It is important to remember this fact when applying this proposed method to an estimation problem of protein networks.

III. CELL CYCLE MODEL IN BUDDING YEAST AND THE ESTIMATION PROBLEM OF PROTEIN NETWORKS

A. Mathematical model for cell cycle

According to the research by Chen, 48 proteins are mainly concerned with the cell cycle in budding yeast, and their relations are represented by 48 non-linear differential equations and some reset rules [7]. For example, differential equations about 9 proteins whose networks will be estimated in the next section are given as follows:

$$\frac{d[\text{mass}]}{dt} = k_g[\text{mass}] \quad (14)$$

$$\begin{aligned} \frac{d[\text{Clb2}]}{dt} &= (k'_{s,b2} + k''_{s,b2}[\text{Mcm1}])[\text{mass}] + (k_{d3,c1}[\text{C2P}] \\ &+ k_{di,b2}[\text{C2}]) + (k_{d3,f6}[\text{F2P}] + k_{di,f2}[\text{F2}]) \\ &- (V_{d,b2} + k_{as,b2}[\text{Sic1}] + k_{as,f2}[\text{Cdc6}])[\text{Clb2}] \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{d[\text{Cdh1}]}{dt} &= k_{s,cdh} - k_{d,cdh}[\text{Cdh1}] \\ &+ \frac{V_{a,cdh}([\text{Cdh1}]_T - [\text{Cdh1}])}{J_{a,cdh}([\text{Cdh1}]_T - [\text{Cdh1}])} - \frac{V_{i,cdh}[\text{Cdh1}]}{J_{i,cdh} + [\text{Cdh1}]} \end{aligned} \quad (16)$$

$$\frac{d[\text{Cdc20}]_T}{dt} = k'_{s,20} + k''_{s,20}[\text{Mcm1}] - k_{d,20}[\text{Cdc20}]_T \quad (17)$$

$$\begin{aligned} \frac{d[\text{Cdc20}]_A}{dt} &= (k'_{a,20} + k''_{a,20}[\text{APC-P}])([\text{Cdc20}]_T - [\text{Cdc20}]_A) \\ &- (k_{mad2} + k_{d,20})[\text{Cdc20}]_A \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{d[\text{APC-P}]}{dt} &= \frac{k_{a,apc}[\text{Clb2}](1 - [\text{APC-P}])}{J_{a,apc} + 1 - [\text{APC-P}]} \\ &- \frac{k_{i,apc}[\text{APC-P}]}{J_{i,apc} + [\text{APC-P}]} \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{d[\text{Sic1}]}{dt} &= (k'_{s,c1} + k''_{s,c1}[\text{Swi5}]) + (V_{d,b2} + k_{di,b2})[\text{C2}] \\ &+ (V_{d,b5} + k_{di,b5})[\text{C5}] + k_{pp,c1}[\text{Cdc14}][\text{Sic1P}] \\ &- (k_{as,b2}[\text{Clb2}] + k_{as,b5}[\text{Clb5}] + V_{kp,c1})[\text{Sic1}] \\ &+ V_{kp,c1}[\text{Sic1}] \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{d[\text{Cln2}]}{dt} &= (k'_{s,n2} + k''_{s,n2} \cdot [\text{SBF}]) \cdot [\text{mass}] \\ &- k_{d,n2} \cdot [\text{Cln2}] \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{d[\text{Sic1P}]}{dt} &= V_{kp,c1}[\text{Sic1}] - (k_{pp,c1}[\text{Cdc14}] + k_{d3,c1})[\text{Sic1P}] \\ &+ V_{d,b2}[\text{C2P}] + V_{d,b5}[\text{C5P}], \end{aligned} \quad (22)$$

where [mass] denotes the mass of a cell and the variables from [Clb2] to [Sic1P] denote the concentrations of proteins.

If this cell cycle model is constructed on a computer, wave forms of 48 protein concentrations can be obtained by running the program. For example, 9 waves of protein concentrations described from (14) through (22) are shown in Figs. 1 and 2. They show that these waves are periodic signals. By using these obtained concentrations of proteins, protein networks can be estimated based on the least-squares method.

B. Estimation problem of protein networks

Protein networks are estimated based on the least-squares method explained in section II.

If state variable $\mathbf{x}(k)$ denotes the obtained protein concentrations, $\mathbf{x}(k)$ is known. So, when rewriting $\Theta(k)$ in (2) as $\Theta(k) = \mathbf{A}_{cell}(k)$, then the protein networks describing the cell cycle model can be represented by the linear regression equation:

$$\mathbf{Y}(k) = \mathbf{A}_{cell}(k)\Phi(k). \quad (23)$$

When rewriting Θ in (13) as $\Theta = \mathbf{A}_{cell}$, the cell cycle model is represented by

$$\mathbf{Y}(k) = \mathbf{A}_{cell}\Phi(k). \quad (24)$$

$\mathbf{Y}(k)$ and $\Phi(k)$ in (23) and (24) are

$$\mathbf{Y}(k) = \mathbf{x}(k+1) \quad (25)$$

$$\Phi(k) = \mathbf{x}(k). \quad (26)$$

\mathbf{A}_{cell} is the matrix represented by

$$\mathbf{A}_{cell} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix},$$

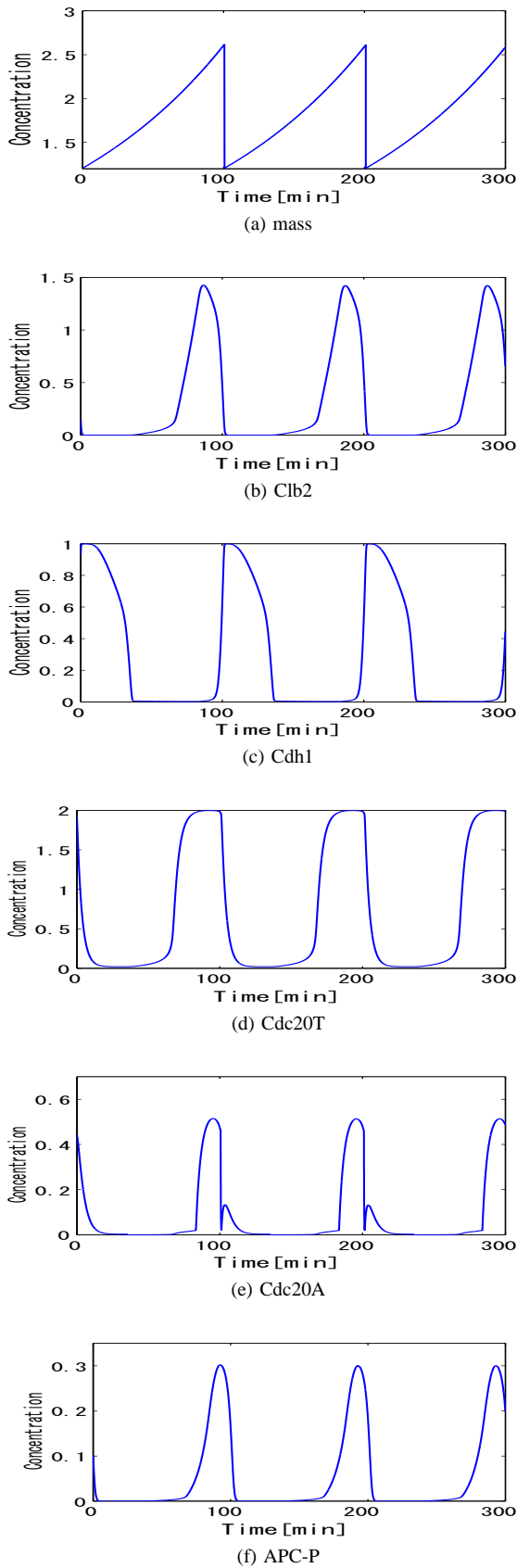


Fig. 1. Time histories of concentration of proteins, [mass] – [APC-P]

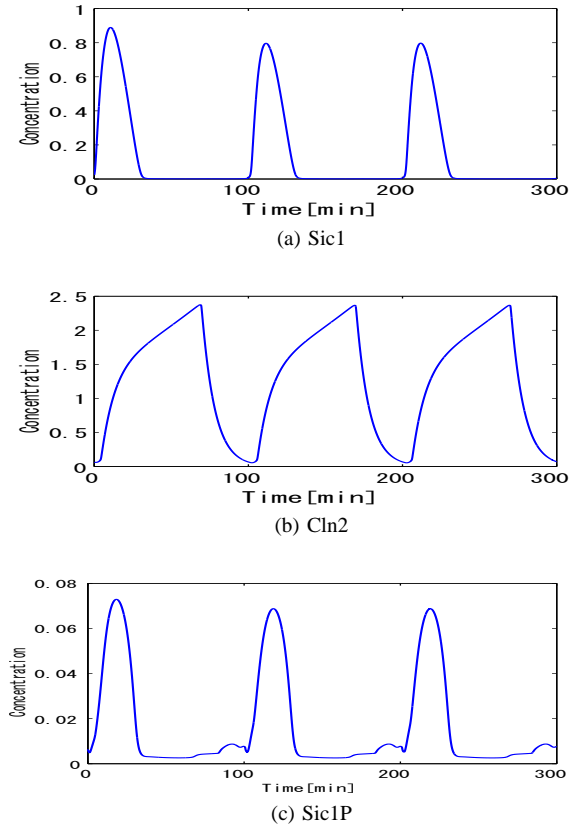


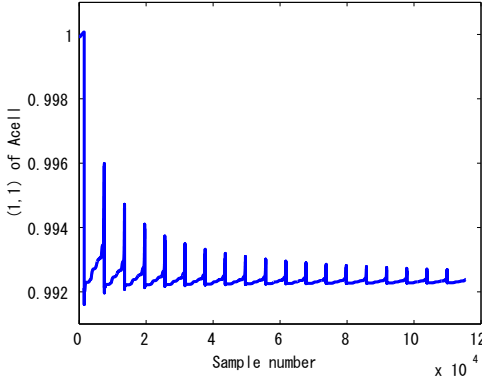
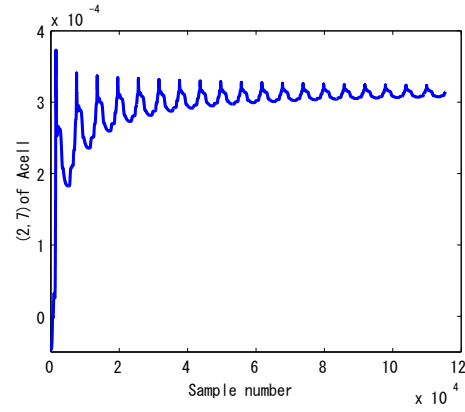
Fig. 2. Time histories of concentration of proteins, [Sic1] – [Sic1P]

where p is the number of proteins whose networks will be estimated. Then, (24) can be represented by

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \vdots \\ x_p(k+1) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_p(k) \end{bmatrix},$$

where each element of \mathbf{A}_{cell} is the strength of the connection between one protein and the another. For example, the element a_{12} in \mathbf{A}_{cell} means the effect that the protein x_2 at the time k gives to the protein x_1 at the time $k + 1$. Hence, for an estimation problem of protein networks, the unknown parameter Θ will be the matrix \mathbf{A}_{cell} that indicates the strength of protein combinations. The larger the certain element of the matrix is, the stronger the combination is. So, it can be considered that the combination exists if the element is large enough.

By using the proposed approach, it is possible to estimate the protein networks of 48 proteins. However, it is a little difficult to estimate 48-dimensional networks from the beginning. So at first, from 48 proteins concerned with the cell cycle, 9 proteins which are working throughout the cell cycle are selected, and then 9-dimensional networks are estimated. Because the state variable $x(k)$ is represented by


 Fig. 3. (1,1) element of \mathbf{A}_{cell}

 Fig. 4. (2,7) element of \mathbf{A}_{cell}

concentrations of selected proteins, $\mathbf{x}(k)$ is given by

$$\mathbf{x}(k) = \begin{bmatrix} [mass] & [Clb2] & [Cdh1] \\ [Cdc20]_T & [Cdc20]_A & [APC - P] \\ [Sic1] & [Cln2] & [Sic1P] \end{bmatrix}^T.$$

From Figs.1 and 2, each protein concentration behaves differently, so they are normalized before estimating the networks.

IV. RESULT OF APPLYING THE PROPOSED METHOD TO CELL CYCLE IN YEAST

A. System identification result by using the proposed method

The proposed method is applied to an estimation problem of protein networks in order to show that the Theorem1 is true.

First, $\hat{\Theta}(k)$ is calculated by using 1 cycle data of the cell cycle, that is N samples, and the calculated $\hat{\Theta}(k)$ is rewritten to $\mathbf{A}_{cell}(0)$. Next, $\mathbf{A}(m) = \hat{\Theta}(N + m)$ is calculated as the number of samples for calculation is increased by 1 in the range of 0 to $N + m$ ($m = 1, 2, \dots, 184903$). Because \mathbf{A}_{cell} is 9×9 matrix, there are 81 elements. For example, the estimation result of (1,1) and (2,7) elements are shown in Figs.3 and 4, respectively. From both figures, it is clear that after a few cycles, the estimated values are the same value at the point of nN , and at other points the values are approached to constant as the number of sample data increases. In the cell cycle, a few cycles in the early time are called the transitional state and after that the cell repeats the regular cycle (it is called the limit cycle). So, in Figs.3 and 4, the result when m is small indicates the transitional state before the limit cycle.

This result is accordance with Theorem 1, so it was demonstrated that Theorem 1 is legitimate and that protein networks can be estimated by using just 1 cycle data.

B. Result of system identification by using the proposed method

As it was shown that Theorem 1 is legitimate in IV-A, 9-dimensional protein networks are actually estimated by using the proposed method. Sample numbers for estimation are 10036 in the 9th cycle after the transitional state, that

is, $N = 10036$. Then, $\mathbf{A}_{cell}(= \hat{\Theta}(N))$ is calculated from (3). Next, examining the each row of \mathbf{A}_{cell} , and decide that the combination exists if the value is greater than 1/1000 of the maximum of the row. On the other hand, if the value is less than 1/1000 of the maximum of the row, the combination does not exist and the value is rewritten as 0. The criterion of 1/1000 was not derived theoretically. However, this criterion affects the estimation result of protein networks. Then the matrix \mathbf{A}_{cell} is rewritten as

$$\mathbf{A}'_{cell} = \begin{bmatrix} 9.92\text{E-}1 & 2.17\text{E-}3 & -2.10\text{E-}3 & 0 \\ -1.34\text{E-}3 & 1.00 & 0 & 0 \\ 2.73\text{E-}3 & -4.87\text{E-}3 & 1.00 & 0 \\ 0 & 0 & -1.51\text{E-}3 & 0 \\ 6.62\text{E-}3 & 2.77\text{E-}3 & 2.05\text{E-}3 & 0 \\ -1.11\text{E-}3 & 5.12\text{E-}3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1.54\text{E-}3 & 0 & 3.76\text{E-}3 & 0 \\ 0 & 0 & -3.53\text{E-}3 & 0 \\ 0 & 0 & 3.15\text{E-}3 & 0 \\ 1.00 & 1.25\text{E-}3 & -2.62\text{E-}3 & 0 \\ -4.97\text{E-}3 & 9.96\text{E-}1 & 0 & 0 \\ 1.04\text{E-}3 & 1.54\text{E-}3 & 9.96\text{E-}1 & 0 \\ 0 & 1.01\text{E-}3 & -1.30\text{E-}3 & 0 \\ -1.30\text{E-}3 & 0 & 1.99\text{E-}3 & 0 \\ 0 & 0 & 0 & 0 \\ 2.29\text{E-}3 & 3.63\text{E-}3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1.33\text{E-}3 & 0 & 0 \\ 0 & 0 & 1.26\text{E-}3 & 0 \\ 0 & -2.93\text{E-}3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1.00 & 0 & -1.87\text{E-}3 & 0 \\ 0 & 1.00 & 0 & 0 \\ 2.70\text{E-}3 & 0 & 9.99\text{E-}1 & 0 \end{bmatrix}. \quad (27)$$

When estimating 9-dimensional protein networks from this obtained \mathbf{A}'_{cell} , the estimation result of the network connection is shown in Fig. 5. It is clear that not only known networks but also new ones were estimated, so the efficacy of the proposed approach for an estimation of protein networks is demonstrated.

The resulting protein networks are estimated by using

identification based on the least-squares method for state-space models is employed, was proposed. By applying this method to an estimation problem of protein networks, it was shown that for identifying a linear system and estimating protein networks, just 1 cycle data of protein concentrations are needed and that it is not necessary to have large number of data. Moreover, based on this theorem, the new estimation method was applied to an estimation problem and 9-dimensional protein networks were estimated and it was shown that the proposed method was legitimate.

B. Future Works

When estimating protein networks by using two approaches, one is by using just 1 cycle data (the proposed approach) and the second is by using the data including the transitional state. Comparing the two resulting networks, there was a little difference. When using the proposed approach, the combination from Cdh1 to Cdc20_A was estimated. It is necessary to verify theoretically whether the newly estimated combination in protein networks improves the robustness of the cell or not.

REFERENCES

- [1] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson: *Molecular Biology of the Cell*, Garland Publishing Inc. (1995)
- [2] H. Kitano: *Systems Biology: A brief overview*, *Science*, Vol. 295, No. 5560, pp. 1662-1664 (2002)
- [3] H. Kitano: *Computational Systems Biology*, *Nature*, Vol. 420, No. 6912, pp. 206-210 (2002)
- [4] H. Kitano: *Foundations of Systems Biology*, The MIT Press (2001)
- [5] H. Kitano and T. Azuma: *Systems Biology and control, system/control/information*, Vol. 48, No. 3, pp. 104-111 (2004)
- [6] J. J. Tyson: *Modeling the Cell Division Cycle, Cdc2 and Cyclin Interactions*, *PNAS* Vol. 88, pp. 7328-7332 (1991)
- [7] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson: *Integrative analysis of cell cycle control in budding yeast*, *Molecular Biology of the Cell*, Vol.15, pp.3841–3862 (2004)
- [8] T. Azuma, F. Moriya and H. Kitano: *Understanding of the basic systems in the cell cycle in multicellular organisms, system/control/information*, Vol. 50, No. 8, pp.309-314 (2006)
- [9] O. Alter and G. H. Golub: *Integrative analysis of genome-scale data by using pseudoinverse projection predicts novel correlation between DNA replication and RNA transcription*, *PNAS*, Vol.101, No. 47, pp. 16577-16582 (2004)
- [10] S. Adachi : *An Introduction to System Identification*, Tokyo Denki University Press (2009) (in Japanese)
- [11] T. Azuma , Kurata , N. Takahashi , S. Adachi : *Estimation and Robustness Analysis of Protein Networks for 6 dimensional Cell Cycle Systems* , The 10th Control division in the Society of Instrument and Control Engineers , Paper No.163-2-3 (5 pages) (2010)