Recursive Least Squares Estimation of Cell Kinetic Parameters using Sequential Measurements of Cell DNA Contents

KANG G. SHIN, MEMBER, IEEE, AND MYUNGHWAN KIM

Abstract—**The proliferation dynamics of a growing cell population can be represented by a discrete-time state model. An application of recursive least squares algorithms to the estimation of important cell cycle kinetic parameters is considered. Cell kinetic parameters are estimated recursively by the following steps: 1) decomposition of state and output spaces, 2) separation of identification of the unperturbed cell system from that of the perturbed cell system, and 3) application of a recursive least squares algorithm for the identification of each decomposed system. This method is feasible due to the availability of a new technology called flow microfluorometry (FMF) which is capable of providing large amounts of quantitative data within a short time period. Emphasis is placed on the construction of a computationally efficient and stable algorithm. The FMF deoxyribonucleic acid (DNA) data of a Chinese hamster ovary (CHO) cell population is used to demonstrate the potential value of the method developed.**

I. INTRODUCTION

CELL KINETICS is an important area of cancer research which deals with the quantitative aspects of the growth behavior of proliferating cell populations and their responses to chemical and physical agents. Cell kinetics already plays an important role in the development of a rational, scientifically grounded approach to cancer therapy, a role which will assume more importance as experimental methods improve.

Extensive efforts in experimental cell kinetics (see [28] for an overview) have been made to produce quantitative data, and a number of mathematical models (for an excellent review, see [1]) have been proposed to provide a theoretical framework for their interpretation of data. All of these models have been either unrealistically simple or were not fully developed due to the lack of abundant quantitative data. The recent spread of flow microfluorometry (FMF) devices [8], [21], [31] is beginning to provide data of sufficient quantity and statistical accuracy to justify further development of the mathematical models.

Modern control theory has become a very useful tool in analytical cell kinetics since Hahn [14] proposed a state-space model to describe the proliferation dynamics of a cell population. This model contains multiple compartments and provides more accurate description of cell movements through the cell cycle than others (most of existing models) dealing with only total populations or small numbers of subpopulations, such as that of G_1, S, and G_2+M cells. We extended this model to include output equations (the cell size and deoxyribonucleic acid (DNA) distributions) and control functions (the cancer treatment) and derived the optimal cancer treatment schedules under the assumption that all cell kinetic parameters are known [4], [16], [17]. Also Gray [12] predicted the changes in cell-DNA distribution from the assumed initial cell age distribution and cell kinetic parameters determined by trial and error.

If the cell kinetic parameters and initial age distribution are not known, then all of the above methods fail. Cell kinetic parameters have to be estimated by means of the measurements of a given cell system, i.e., identification of parameters must be done. Modern control theory can contribute to this part of the cell kinetics problem, because system identification is one of its very important and active areas (for excellent overviews see Åström and Eykhoff's survey [2], Sage and Melsa's book [29], and the December 1974 special issue of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL).

We also investigated research in the area of bilinear system identification to find ways to exploit the bilinear nature of our system. Bruni *et al.* [5] reduced the deterministic parameter identification problem for bilinear systems to a nonlinear multipoint boundary value problem (MPBVP) and suggested the quasilinearization algorithm for solving this problem. Baheti *et al.* [3] developed an algorithm to estimate parameters of a class of discrete-time bilinear systems using second-order correlations. Recently, Karanam *et al.* [15] used a finite orthonormal expansion to approximate input and output functions for the identification of a bilinear system where the dimension of the state space is equal to that of the output space. However, neither was applicable to our problem, the structure of which can be exploited in identifying cell kinetic parameters.

We have made a considerable effort to determine the best estimates, in least square sense, of the cell kinetic

Manuscript received October 19, 1978; revised March 5, 1979, November 20, 1979, and September 2, 1980. This work was supported in part by the National Cancer Institute under Grant CA16259.

K. G. Shin is with the Electrical and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY 12181.

M. Kim is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853.

0018-9472/80/1200-0861$00.75

parameters, using the quasilinearization algorithm and orthogonal transformation [19], [20]. These methods did not take advantage of the structure of the cell system and resulted in computational inefficiency and numerical instability. In this paper we intend to correct these drawbacks by

1) decomposing the state and output spaces into subsystems of smaller dimension,
2) separating the identification of undisturbed cell system from that of disturbed cell system, and
3) applying a recursive least squares algorithm for the estimation of parameters in each subsystem.

A Chinese hamster ovary (CHO) cell population is used to demonstrate the utility of this method.

II. Problem Statement

The cell cycle and proliferation kinetics of a tumor cell population under the effect of antitumor drugs can be described by

$$x(k+1)=[\Phi+\Psi u(k)]x(k), \qquad x(0)=x_0. \qquad (1)$$

The basic concepts underlying the equation are briefly given as follows (see [4], [14], [16], and [17] for further details). The cell cycle is divided into N physiological age intervals called age compartments, and the physiological age distribution is represented as an N-dimensional vector called the state vector, the ith component of which represents the number of cells in the ith age compartment. Usually, the state vector $x \in R^n$ has additional compartments to include noncycling cells such as resting and dead cells ($n > N$).

The aging of the population is represented by advancing cells down the compartments into mitosis, doubling the number of mitotic cells which are then entered into the first age compartment. This cycling process occurs in discrete steps. Since all cells do not age physiologically at the same rate, during one unit time interval, the cell fractions β, δ, and α are advanced by 0, 1, and 2 compartments, respectively. The probability of a cell's advancing by more than two compartments in one unit time is assumed to be negligible. This cycling process for proliferating cells is described by the transition matrix Φ_{pp} in Fig. 1, making a distinction among the aging rates in different phases of the cell cycle. Likewise, the state transition matrix Φ describes the changes in the cell age state vector over one unit time period.

Whenever antitumor drugs are applied some cells are known to be killed (cell-killing effect) or blocked from growing past a certain age boundary in the cell cycle (progression delay effect). These cells have to be reassigned to certain nonproliferating and proliferating compartments. The input transition matrix Ψ represents the reassignment of cells affected by the presence of antitumor drugs during one unit time interval.

$u(k)$ is the external disturbance or control function, indicating the amount of a treatment given during interval k. Since there must always be a safe limit to the drug dosage, one can assume $u(k) \in [0, 1]$ for all k without loss of generality. $u(k)=0$ implies that at interval k no drug is present while $0 < u(k) \leqslant 1$ indicates the presence of the drug.

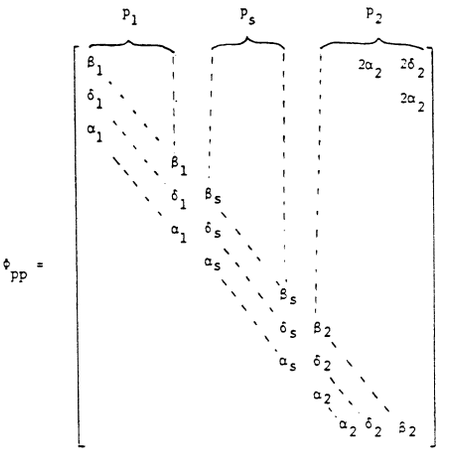


Fig. 1. Example of age state transition matrix. State transition matrix Φ_{pp} represents changes inside proliferating cell population over one unit time period where neither cell death nor cell arrest occurs. It is easy to modify this matrix so that matrix includes noncycling group. In above matrix p_1, p_s, and p_2 represent the number of age compartments in G_1, S, and G_2+M phases, respectively; α_i denotes probability that cell advances two age compartments in one time interval, β_i that for no advancement, and $\delta_i = 1-\alpha_i-\beta_i$ is probability for normal progress for $i=1$, s, and 2 (namely, G_1, S, and G_2+M).

Although the cell age state cannot be directly observed, the size and DNA content of a cell are experimentally measurable. Hence the cell age state has to be monitored indirectly through the use of cell size or cell DNA information. The strategy of most experimental methods of cell cycle analysis is to introduce a marker into selected cells and to observe the progress of the labeled cohort of cells around the cycle. All methods but recently developed FMF technique are generally known to be time-consuming and statistically inaccurate. Consequently, in this paper the FMF DNA data will be considered as a sole means of monitoring both cell age states and cell kinetic parameters. This measurement system is characterized by

$$y(k)=Dz(k) \qquad (2a)$$

$$z(k)=Qx(k) \qquad (2b)$$

where $y \in R^l$ is the experimentally measured FMF DNA vector, the jth entry of which represents the number of cells found in the jth channel; D is the transformation matrix describing the measurement noise due to the errors in both staining and devices (see [7], [10], [12], [18], [20] for details); $z \in R^m$ is the dispersion-free or "true" DNA vector [16]; S is the transformation matrix relating the cell age state to the cell DNA state (for details, see [16], [18]).

The model enables one to predict the time course of the cell age and cell DNA distributions of a cell population under consideration. When the parameters in the model and the initial age distribution are determined, it can provide an insight into the behavior and movement of cells under chemotherapy in different cell age and DNA

compartments. Since the susceptibility of a cancer cell to most antitumor drugs depends on its position in the cell cycle, the knowledge of the cell kinetic parameters and the initial cell age distribution is of the utmost importance. Therefore, the problem is to determine the transition matrices Φ and Ψ and the initial cell age state x_0 through the use of the input–output pairs of the cell system; namely, it is to select the parameters in the model (1) which best agree, in some sense, with input–output data. This concept leads to the construction of a criterion or a loss function which measures the deviation of model output from system output.

The least squares criterion is chosen in this paper to estimate the cell kinetic parameters. Then the problem may be stated as follows.

Problem (P): Given a cell system Σ with input–output pair $\{u(k), y(k)\}$ described by (1) and (2), determine a parameter vector θ^* in F such that the loss function $J_k = \frac{1}{2}\sum_{i=0}^{k} \| y(i) - y_m(i) \|^2$ is minimized and

$$\lim_{k\to\infty} \| y(k) - y_m(k) \| = 0$$

where y denotes FMF DNA data; y_m, model output; $\|\cdot\|$, two-norm; and F is the feasible region of cell kinetic parameters.

Remarks:

1) In addition to the well-known advantages in using least squares criterion [6], the key reason for the use of the two-norm criterion is the property of preserving its norms under orthogonal transformations, which provides an extremely easy and stable tool for recursive computation.

2) In our previous work [18], we showed how to compute the transformation matrices D and Q. Accordingly, they are assumed known throughout this paper.

3) Since a particular form of Φ, due to the physical sense of x, is important, one cannot assume that Φ has a canonical form.

4) The cell kinetic parameters under consideration are determinable by examining a number of sequential observations (see [11]). Hence this paper will focus on the method of determining the parameters without rigorous discussion of identifiability as in [30].

The solution to the above problem may be roughly divided into two parts. The first part (Section III) is structural decomposition of the cell system on the basis of the system's special nature, which provides a great deal of reduction in computation and storage requirements. The second part (Section IV) is recursive computation of the cell kinetic parameters for decomposed sub-systems by making use of sequential triangularization, namely, orthogonal transformation. This part presents an extremely easy and stable algorithm for the computation of cell kinetic parameters.

III. System Decomposition

The transition matrices Φ and Ψ have a large number of rows and columns, i.e., the dimension of the state vector is high. This high dimensionality is necessary to predict accurately the time course of the cell age distribution rather than a coarse prediction of the behavior of the total population with small number of subpopulations, such as that of G_1, S, and G_2+M cells. Because of the high dimensionality, computation is very costly. This difficulty can be circumvented by decomposing the cell system into several subsystems.

The decomposition mentioned above is based upon the special nature of the cell process as follows. A cell replicates its chromosomes by synthesizing DNA only during the S phase and not in either G_1 or G_2+M. Since the change of DNA content of a single cell is experimentally detectable, the cell cycle parameters in the S phase can be rather easily determined. In order to take advantage of this special property of cell process, the state-space model (1) can be rearranged as follows, including four subsystems Σ_1, Σ_s, Σ_2, and Σ_m:

$$\begin{aligned} x(k+1) &= \begin{bmatrix} x_1(k+1) \\ \hline x_s(k+1) \\ \hline x_2(k+1) \\ \hline x_m(k+1) \end{bmatrix} \\ &= \left(\left[\begin{array}{c|c|c|c} \Phi_{11} & 0 & \Phi_{12} & \Phi_{1m} \\ \hline \Phi_{s1} & \Phi_{ss} & 0 & 0 \\ \hline 0 & \Phi_{2s} & \Phi_{22} & \Phi_{2m} \\ \hline \Phi_{m1} & 0 & \Phi_{m2} & \Phi_{mm} \end{array} \right] \right. \\ &\quad \left. + \left[\begin{array}{c|c|c|c} \Psi_{11} & 0 & 0 & 0 \\ \hline 0 & \Psi_{ss} & 0 & 0 \\ \hline 0 & 0 & \Psi_{22} & 0 \\ \hline \Psi_{m1} & \Psi_{ms} & \Psi_{m2} & \Psi_{mm} \end{array} \right] u(k) \right) \\ &\quad \cdot \begin{bmatrix} x_1(k) \\ \hline x_s(k) \\ \hline x_2(k) \\ \hline x_m(k) \end{bmatrix} \end{aligned} \tag{3}$$

where Σ_i for $i=1, s, 2, m$ denotes the cell subpopulation systems of G_1, S, G_2+M, and the nonproliferating cells, respectively; $x_i \in R^{p_i}$, the age state vector in the subsystem Σ_i; Φ_{ij} for $i \neq j$ represents cell transition from Σ_j to Σ_i; Ψ_{ii}, the transformation matrix for expressing the chemotherapeutic effects on the subsystem Σ_i; Ψ_{ij} for $i \neq j$ is the input transition matrix describing cell transition from Σ_j to Σ_i due to the chemotherapy.

One can observe that zeros in the partitioned transition matrix in (3) are introduced due to physical considerations and the assumption that no cycling cells can be arrested in S. This assumption is widely supported in the literature [1], [11], [22]. It is also assumed, without loss of generality, that the state-space model is constructed such that $p_s > 2$.

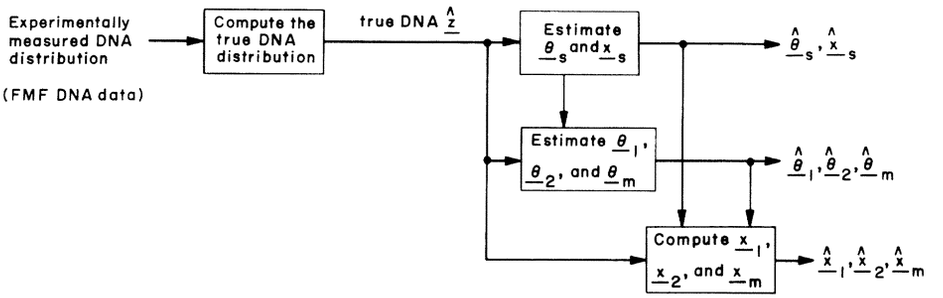


Fig. 2. Block diagram used for estimating states and parameters of unperturbed cell system.

It is possible to separate the identification of Φ from that of Ψ by simply setting $u(k)=0$; that is, the cell process without chemotherapy can be identified first, and then the effects of the antitumor drugs on the cell system will be determined.

According to the motivation of state decomposition, the true DNA vector is considered as the target for the output decomposition. The transformation matrix Q which relates the age vector to the true DNA vector can be computed from the knowledge of DNA synthesis rate during the S phase and the locations in the cell cycle where the cycling cells may become resting cells [18]. At present our information is limited, but it is commonly assumed that cycling cells can be arrested at three points in the cell cycle: in early G_1, in late G_1, and in late G_2 [11].

Following this assumption, the true DNA vector z can be decomposed into three subsystems or (p_s+2) subpopulations as follows:

$$\begin{aligned} z_1(k) &= \sum_{i=1}^{p_1} x_{1i}(k)+x_{m1}(k)+x_{m2}(k) \\ z_s(k) &= x_s(k) \\ z_2(k) &= \sum_{i=1}^{p_2} x_{2i}(k)+x_{m3}(k) \end{aligned} \tag{4}$$

where z_1 represents the total number of cells that have not yet started to synthesize DNA, z_2 the number of cells which completed DNA synthesis but which have not yet divided, and z_s the cell DNA distribution in the S phase; x_{mi} for $i=1$, 2, and 3 are the subpopulations of resting cells arrested in early G_1, late G_1, and late G_2, respectively. Observe that the contribution of dead cells to the DNA distribution is assumed to be negligible. It is reasonable to assume this since dead cells which do absorb dye will fluoresce at a considerably different rate and do not distort the data [7].

It should be particularly noted from (4) that the age state in S is identical to the true DNA vector in S. This can be assumed, without loss of generality, since z_s can be expressed by $z_s=Q_s x_s$ for some matrix Q_s with rank $(Q_s)=\dim(x_s)$, and thus one can always compute x_s by $x_s=(Q_s^TQ_s)^{-1}Q_s^Tz_s$. As will be seen later, this information is very useful for the estimation of cell cycle parameters in the S phase. The great advantage of the decomposition will become apparent in the detailed computational algorithms in the subsequent discussions.

IV. Recursive Estimation of Parameters and States

With the previously discussed system decomposition, the parameters and states for the unperturbed cell system are estimated recursively by the following steps. For notational convenience let θ_i denote the cell kinetic parameters in the subsystem Σ_i for $i=1$, s, 2, and m.

1) Estimate the true DNA distribution from the experimental (FMF) DNA data.
2) Estimate θ_s and x_s from the true DNA distribution.
3) Compute θ_1, θ_2, and θ_m from the knowledge of the true DNA data, θ_s and x_s.
4) Compute x_1, x_2, and x_m using the estimates in 1), 2), and 3).

This estimation procedure is depicted by the block diagram in Fig. 2. Note that all these estimators are recursive in measurements and deal with only small dimensional data.

A. Elimination of Instrumental Dispersion

Flow microfluorometry [8], [21], [31] is of great importance in cell kinetics studies because of its potential for rapidly and accurately measuring selected characteristics in individual cells of a cell population in suspension; nevertheless, FMF DNA data may contain staining errors and errors due to measurement devices. Frequently, the dispersion is characterized by a Gaussian distribution [10], [12]. Ideally, this characterization may not be valid, but practically it does not impose any problem in the FMF analysis [12]. A Gaussian distribution is completely described by its mean and its coefficient of variation. These parameters can be determined from various kinds of statistical inferences. In this paper, it is assumed that the complete characterization of the dispersion, described by the matrix D in (2a), is available.

It is desired to determine the true DNA distribution from FMF DNA data and (2a). This problem can be cast into a typical weighted least squares problem as follows: find $\hat{z}(k)\geqslant 0$ such that $J=\frac{1}{2}\|Dz(k)-y(k)\|^2_{R^{-1}(k)}$ attains its minimum at $z(k)=\hat{z}(k)$ where $\|\cdot\|$ denotes the two-norm and $R^{-1}(k)$ is the weighting matrix.

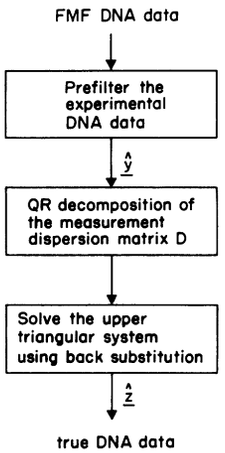


Fig. 3. Procedure for computing the true DNA distribution from the FMF DNA data.

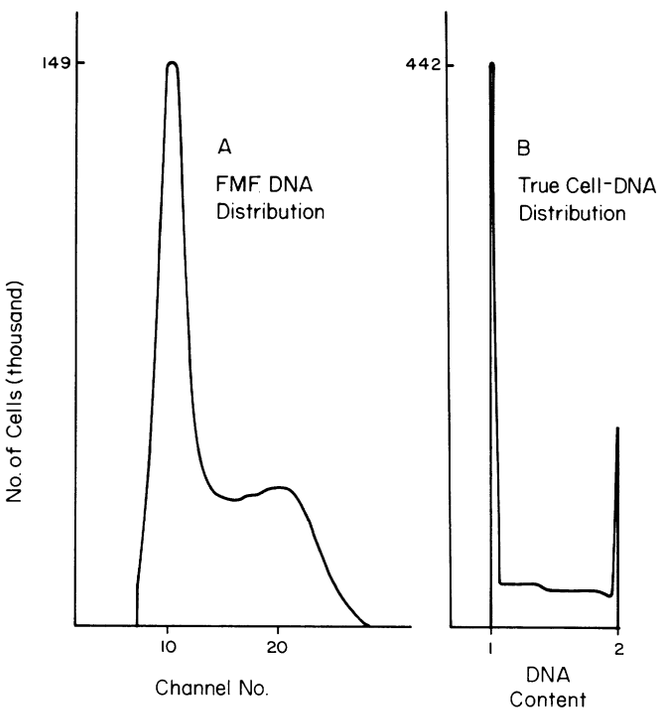


Fig. 4. (a) Measured FMF DNA distribution. (b) Computed true DNA distribution.

Note that $R(k)$ would be the covariance matrix of the uncorrelated white Gaussian noise if the measurement system has such additive noise. In such cases, the minimization of J would be equivalent to the maximization of the probability density function, namely, $\hat{z}(k)$ would make $y(k)$ most likely to be produced.

Numerical solution of this type of problems, namely, nonnegative least squares (NNLS) is excellently discussed by Lawson and Hanson [21]. Their discussion was based upon the Kuhn–Tucker lemma for optimal solutions [9] and the orthogonal transformation. Since the dispersion matrix D is of full rank, the above problem yields a unique solution.

The constraint of nonnegativeness, $\hat{z}(k) \geqslant 0$, is needed because its components represent the number of cells. However, for most cases the unconstrained least squares (UCLS) algorithm yields the nonnegative solutions for $z(k)$. The true DNA distribution is therefore easily calculated by solving the triangular system with back substitution, following the Householder triangularization of the matrix D. It should be noted that the computation required for the NNLS problems is costlier than that for the UCLS ones. Fig. 3 shows the UCLS procedure used for the computation of true DNA distributions, and Fig. 4 presents an experimental (FMF) DNA distribution and its true DNA distribution computed by the UCLS algorithm.

B. Estimation of θ_s and x_s

Now the cell kinetic parameter vector and the cell age state vector in the subsystem Σ_s, denoted by $\boldsymbol{\theta}_s$ and $\boldsymbol{x}_s$, respectively, are to be estimated from the true DNA distributions. One can see that the age state vector $\boldsymbol{x}_s$ in the subsystem Σ_s is known, and thus only $\boldsymbol{\theta}_s$ remains to be determined.

From the foregoing system decomposition, one can obtain the following equations

$$\boldsymbol{x}_s(i+1) = \Phi_{s1}\boldsymbol{x}_1(i) + \Phi_{ss}\boldsymbol{x}_s(i) \tag{5a}$$

$$z_s(i) = \boldsymbol{x}_s(i). \tag{5b}$$

It is noted that all but the right upper corner entries of the matrix Φ_{s1} are zeros. Using the above equations one can obtain a simple equation which is linear in $\boldsymbol{\theta}_s$ as follows:

$$F_i\boldsymbol{\theta}_s \cong \boldsymbol{f}_i, \qquad \text{for } i=2,3,\cdots \tag{6}$$

where F_i and $\boldsymbol{f}_i$ are data matrices determined from two consecutive true DNA vectors, i.e., $F_i \triangleq F_i(z_s(i), z_s(i-1))$ and $\boldsymbol{f}_i \triangleq \boldsymbol{f}_i(z_s(i-1))$.

The problem of estimating $\boldsymbol{\theta}_s$ from (6) can be cast into a typical least squares problem:

$$\text{minimize} \left\| \begin{bmatrix} F_2 \\ F_3 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \boldsymbol{\theta}_s - \begin{bmatrix} \boldsymbol{f}_2 \\ \boldsymbol{f}_3 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \right\| \text{subject to } \boldsymbol{\theta}_s \in F_s. \tag{7}$$

The parameter vector $\boldsymbol{\theta}_s$ is restricted to a feasible region F_s since $\boldsymbol{\theta}_s$ represents the probabilistic parameters. This is a constrained least squares problem which is commonly solved by making use of Kuhn–Tucker conditions and orthogonal transformations such as QR decomposition and singular value decomposition [23], [26], [32].

We want to obtain a sequence of solutions for a data set, i.e., $\{[F_l: \boldsymbol{f}_l];\ l=2, 3,\cdots, i\}$, to which new data (i.e., $[F_{i+1}: \boldsymbol{f}_{i+1}]$) are being sequentially added. A requirement for this type of computation arises in the class of problems called recursive (or sequential or on-line) estimation, filtering, or identification. Such methods provide a means of conserving computer storage for certain types of problems involving voluminous data.

A recursive method for determining $\boldsymbol{\theta}_s$ is developed here by making use of the augmentation methods which have excellent properties of numerical stability. This is in contrast to many published approaches to this problem that

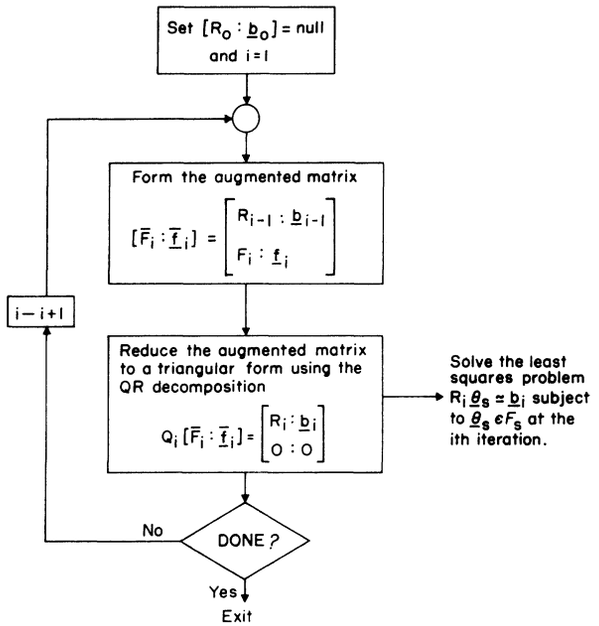


Fig. 5. Flowchart for estimating cell kinetic parameters in S.

are based on the notion of updating the inverse of a matrix.

The detailed algorithm is as follows. For notational convenience let $[R_0 : \boldsymbol{b}_0]$ denote a null matrix having no rows. At the beginning of the ith stage of the algorithm, one has the triangular matrix $[R_{i-1} : \boldsymbol{b}_{i-1}]$ from the previous stage and the new data matrix $[F_i : \boldsymbol{f}_i]$. Form the augmented matrix

$$\left[\bar{F}_i : \bar{\boldsymbol{f}}_i\right] = \begin{bmatrix} R_{i-1} : \boldsymbol{b}_{i-1} \\ F_i : \boldsymbol{f}_i \end{bmatrix} \tag{8}$$

and reduce this matrix to triangular form by Householder triangularization as follows:

$$Q_i\left[\bar{F}_i : \bar{\boldsymbol{f}}_i\right] = \begin{bmatrix} R_i : \boldsymbol{b}_i \\ 0 : 0 \end{bmatrix}. \tag{9}$$

The estimate of $\boldsymbol{\theta}_s$ at the ith stage, denoted by $\hat{\boldsymbol{\theta}}_s(i)$, is obtained by solving the least squares problem

$$R_i \boldsymbol{\theta}_s \cong \boldsymbol{b}_i \text{ subject to } \boldsymbol{\theta}_s \in F_s. \tag{10}$$

This completes the ith stage of the algorithm.

If there are more data to be processed, then the above procedure is repeated until the data are exhausted. This procedure is depicted by the flowchart in Fig. 5.

The algorithm constructs a sequence of triangular matrices $[R_i : \boldsymbol{b}_i]$ for $i=2, 3, \cdots$, with the property that the least squares problem $R_i \boldsymbol{\theta}_s \cong \boldsymbol{b}_i$ has the same solution set and the same residual norm as the problem $(F_2^T \cdots F_i^T)^T \boldsymbol{\theta}_s \cong (\boldsymbol{f}_2^T \cdots \boldsymbol{f}_i^T)^T$. The significant point permitting a saving storage is the fact that, for each i, the matrix $[R_i : \boldsymbol{b}_i]$ can be constructed and stored in the storage space previously occupied by the augmented matrix $[\bar{F}_i : \bar{\boldsymbol{f}}_i]$.

It should be noted that the augmentation method for estimating linear parameters is numerically more efficient and stable than other recursive least squares techniques [24]. The recent work of Paige and Saunders [26] is an excellent reference for the sequential triangularization method in comparison with other literature regarding Kalman filtering.

C. Estimation of Parameters and States of Σ_1, Σ_2, and Σ_m

The parameters and states of the subsystems Σ_1, Σ_2, and Σ_m are to be determined using the true DNA distributions and the knowledge of $\boldsymbol{\theta}_s$ and $\boldsymbol{x}_s$. The computational algorithm for estimating these parameters is more complicated than that for $\boldsymbol{\theta}_s$ and $\boldsymbol{x}_s$, since one cannot obtain a simple equation of these variables.

The dynamics of the reduced system, made up of Σ_1, Σ_2, and Σ_m, can be compactly described by

$$\boldsymbol{s}(k+1) = \tilde{\Phi}\boldsymbol{s}(k) + \tilde{\boldsymbol{z}}_s(k) \tag{11}$$

where

$$\boldsymbol{s} = \left(\boldsymbol{x}_1^T : \boldsymbol{x}_2^T : \boldsymbol{x}_m^T\right)^T$$

$$\tilde{\boldsymbol{z}}_s = \left(\boldsymbol{0}^T : (\Phi_{2s}\boldsymbol{z}_s)^T : \boldsymbol{0}^T\right)^T$$

$$\tilde{\Phi} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{1m} \\ 0 & \Phi_{22} & \Phi_{2m} \\ \Phi_{m1} & \Phi_{m2} & \Phi_{mm} \end{bmatrix}.$$

Observe that $\boldsymbol{s}$ is the state vector for the reduced system, and $\tilde{\boldsymbol{z}}_s$ is the known input from the previous estimation steps.

$\tilde{\Phi}$ and $\boldsymbol{s}$ in (11) must be determined from observing the change of DNA distributions in the S phase together with z_1 and z_2 in (4). Therefore, the output equations for the reduced system are

$$\boldsymbol{z}_s(k+1) - \Phi_{ss}\boldsymbol{z}_s(k) = \Phi_{s1}\boldsymbol{x}_1(k) \tag{12a}$$

$$z_1(k) = \sum_{i=1}^{p_1} x_{1i}(k) + x_{m1}(k) + x_{m2}(k) \tag{12b}$$

$$z_2(k) = \sum_{i=1}^{p_2} x_{2i}(k) + x_{m3}(k). \tag{12c}$$

In order to obtain a recursive estimate of the parameters in $\tilde{\Phi}$ and Φ_{s1}, it is necessary to eliminate the state vector $\boldsymbol{s}$ from (11) and derive an equation which recurrently relates the observations to the parameters (for an example, see [25]).

Since the system under consideration is observable, we know that it is possible to obtain such relationships. Once this equation is obtained, we can apply the sequential least squares method as in Section IV-B. Using a sequence of measurements one can obtain the equation, since cells in $G_2 + M$ move into G_1, cells in the G_1 phase pass through the S phase, and the cell age distribution in the S phase is directly observable.

The final step of identifying the cell system without chemotherapy is to compute the state vector $\boldsymbol{s}$ for the reduced system. Now all parameters but $\boldsymbol{s}$ are available from the previous discussion. Commonly, $\boldsymbol{s}$ can be obtained by solving normal equations [6], [23] after a sequence of DNA data is accumulated (also note that this

accumulation is related to the observability matrix [27]). This method may not be attractive in practice, for the following reasons: 1) it is not recursive; namely, it is necessary to solve the normal equations again without taking advantage of computations undergone in the previous steps whenever new data are acquired, and 2) the solution may become numerically unstable. As was seen in Section IV-B, the sequential triangularization provides recursive and numerically stable solutions to linear least squares problems. This method is thus the best remedy to the shortcomings mentioned above. With minor algebraic manipulations, one can make use of the procedure in Fig. 5 to recursively compute s.

This completes the estimation of parameters and states for the cell system without chemotherapy. It is convenient to recall that the identification of Φ in (1) was separated from that of chemotherapeutic effects on a cell population.

The effects of chemotherapy on cell dynamics are complicated but commonly classified as cell-killing, progression delay, and prolongation of cell cycle times. Although the generation of "preterminal" cells [20], [33] may account for the third effect, the present study includes only the first two effects. Hence the chemotherapeutic effects are here expressed in terms of the cell-killing and the cell-blocking rates.

We have to estimate the parameters describing chemotherapeutic effects on the basis of measurements and the information on the cell system without chemotherapy. This paper is concerned only with cell cycle stage specific drugs (CCSS) [17] because of their popularity. This fact produces simple equations which are linear in chemotherapeutic parameters. Only straightforward calculations are required to solve these equations for the probabilistic parameters, as shown in the following section.

V. Estimation of Cell Kinetic Parameters and Cell Age Distributions for Chinese Hamster Ovary Cells

A Chinese hamster ovary cell population is selected as an application example for the cell kinetic analysis discussed thus far. This analysis requires

1) a mathematical model of cell cycling process,
2) the relationship between cell age and cell DNA states,
3) the effect of instrumental and preparative errors on the DNA distributions, and
4) estimation of both the cell kinetic parameters and the initial cell age distribution, based on the system decomposition and the recursive least-squares method.

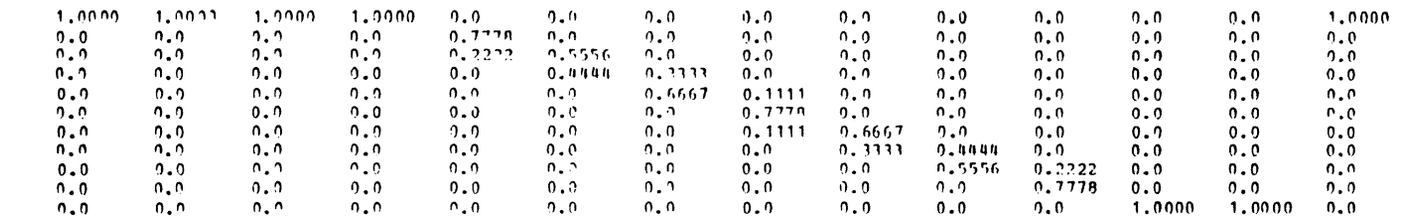
The CHO cells are known to have the cell cycle consisting of four distinct phases G_1, S, G_2, and M. The CHO cell population also contains resting (G_0) cells which are noncycling but capable of regaining the cycling capability. The dimension of the age state vector n is determined by 1) the integral ratio among the phase durations and the associated computational requirements, and 2) the number of nonproliferating compartments. Gray *et al.* [13] excellently reviewed the methods of determining the cell cycle time and phase durations (see Table I for the associated results). They reported that the durations for G_1, S, and G_2+M phases (T_{G_1}, T_S, and T_{G_2+M}, respectively) giving the best match were 3.6 h, 6.6 h, and 2.1 h, respectively. Observe that this information yields an approximate integral ratio of the phase durations to be 4:7:2 ($=T_{G_1}:T_S:T_{G_2+M}$). Considering this ratio and the available data[1] for a CHO cell population taken at 1-h or 2-h intervals, we have chosen the number of proliferating compartments to be 13. This implies that 1) the mean duration of one age compartment is one hour, and 2) G_1, S, and G_2+M phases are composed of 4, 7, and 2 age compartments, respectively. Cells in the G_0 phase (i.e., resting cells) are reported to come from early G_1, and thus transition of cells in G_1 to G_0 and recruitment of G_0 cells into G_1 are assumed to take place in the first age compartment. This assumption, along with the number of proliferating age compartments, determines the dimension of $\boldsymbol{x}$, n, to be 14 and also simplifies (3) and (4); namely, $\Phi_{2m}=0$, $\Phi_{m2}=0$, and $x_{m2}(k)=x_{m3}(k)=0$ for all k. Note that the disturbance in cell kinetics due to the dead cells is assumed negligible [7], and therefore the dead cells are not included in the model.

A cell has one unit of DNA content during the G_1 phase and two units during the G_2+M phase while the DNA content of the cell during the S phase increases monotonically from one unit at the beginning of the S

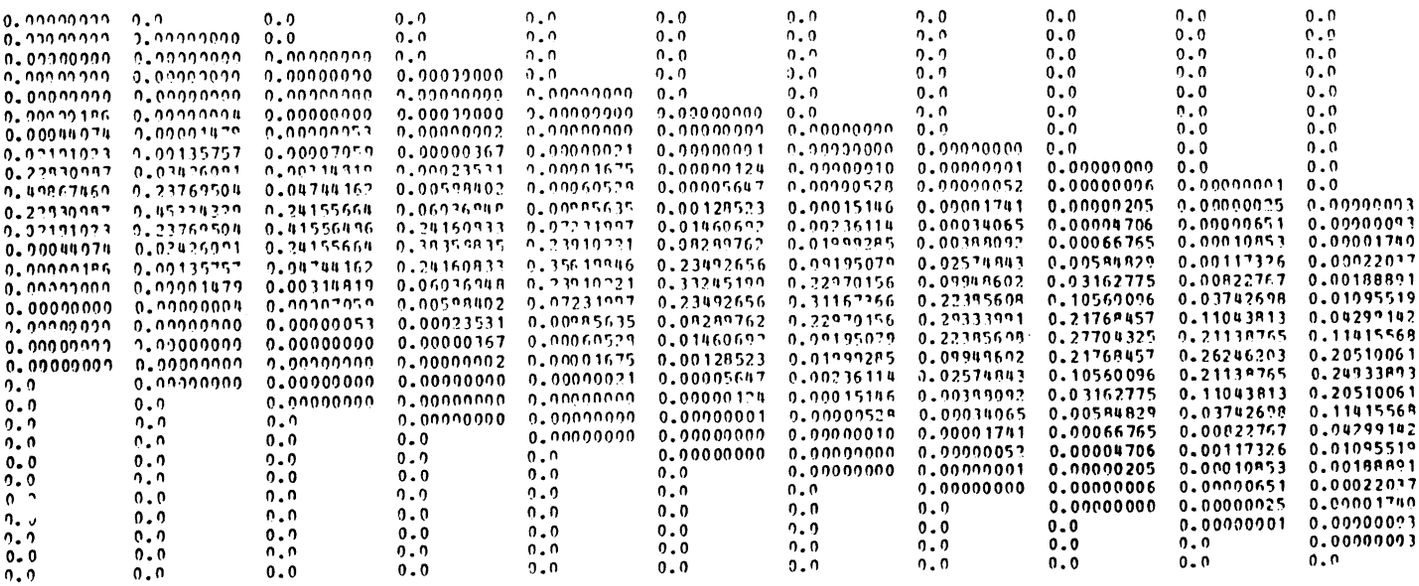
TABLE I
Comparison of Phase Durations and Coefficients of Variations (CV) for CHO Cells Estimated by Different Techniques (from [13])

	Method	T_C(CV$_C$)	T_{G_1}(CV$_{G_1}$)	T_S(CV$_S$)	T_{G_2+M}(CV$_{G_2+M}$)
1)	Pulse label and autoradiography using fraction of cells in S	12.5	—	6.9	—
2)	Method using fraction of cells in G_1, S, G_2+M (FCM)	12.5	3.7	6.8	2.0
3)	RCS_i	12.9(0.16)	—	6.5(0.29)	—
4)	FLM	12.7(0.12)	4.5(0.20)	6.3(0.20)	1.7(0.19)
5)	Perturbed cell cycle analysis (FP_i)	12.1	3.4	6.2	1.3
6)	Perturbed cell cycle analysis (FCM)	12.3(0.12)	3.6(0.20)	6.6(0.20)	2.1(0.19)

[1]The data were obtained from both the Sloan Kettering Cancer Institute, New York, NY, and the Lawrence Livermore Laboratory, Livermore, CA.



(a)



(b)

Fig. 6. Computed matrices (a) D for measurement dispersion. (b) Q for relating cell-age to cell-DNA states.

phase to two units at the end of the S phase. The exact relationship between cell age and cell DNA states can be determined by the knowledge of 1) the rate of DNA synthesis during the S phase and 2) the position in the cell cycle where cells may be arrested and become dormant. The relationship represented by the matrix Q in Fig. 6(a) is computed using our previous method in [18] and an assumption of constant DNA synthesis rate (the method could calculate Q for any DNA synthesis rate curve).

There are two sets of FMF DNA data available for the present work; one is the sequential DNA distribution of an asynchronous CHO cell population, measured at one-hour intervals, whereas the other is that of a synchronized CHO cell population, taken at two-hour intervals. These data were taken from the CHO populations without drug, and thus the control function $u(k)$ in all previous equations should set to zero. Note that the identification of the unperturbed system was separated from that of the perturbed system (Section III), and consequently the above data are both suitable for the identification of the unperturbed system. In reality, it is impossible to obtain 100 percent cell synchronization which gives the initial cell age distribution, and the two-hour intervals of data takings are less accurate than one-hour intervals to analyze a CHO cell population. Therefore, the FMF DNA data for the asynchronous population is used to estimate both the cell kinetic parameters and the cell age distributions. Measurement noise in this data, due to the imperfect staining and instruments, is assumed to be distributed normally and is characterized by the matrix D in Fig. 6(b).

Using the above FMF DNA data, the cell kinetic parameters are estimated, and the result is given in Table II. As was expected, the augmentation method produces estimates of the cell kinetic parameters with numerical stability. Table II also shows the number of iterations required for the convergence of estimates. Thus from such a result one can determine how many sequential measurements must be made to obtain good estimates of the cell kinetic parameters. The calculation of parameters d_2, f, and γ was started after the reasonable convergence in that of α, β, and d_1 was obtained, as the estimated values of α, β, and d_1 had to be used for the estimation of d_2, f, and γ.

TABLE II
COMPUTED RESULTS OF CELL KINETIC PARAMETERS FOR CHO CELL POPULATION

K	α	β	d_1	d_2	f	α
1	0.2286	0.2078	0.0667	—	—	—
2	0.2286	0.2078	0.0667	—	—	—
3	0.1711	0.2698	−0.0823	—	—	—
4	0.2082	0.1556	0.0626	—	—	—
5	0.2109	0.1811	0.0493	—	—	—
6	0.2109	0.1801	0.0496	—	—	—
7	0.2109	0.1802	0.0496	—	—	—
8	0.2109	0.1800	0.0494	—	—	—
9	0.2112	0.1807	0.0492	−14.4319	—	—
10	0.3111	0.1806	0.0492	0.0944	0.0595	0.1918
11	0.2111	0.1806	0.0492	0.0984	0.0578	0.1319
12	0.2111	0.1806	0.0492	0.0993	0.0558	0.0934
13	0.2111	0.1806	0.0492	0.0996	0.0544	0.0753
14	0.2111	0.1806	0.0492	0.0997	0.0537	0.0670
15	0.2110	0.1806	0.0493	0.0997	0.0533	0.0628
16	0.2110	0.1805	0.0493	0.0998	0.0531	0.0606
17	0.2110	0.1805	0.0494	0.0998	0.0529	0.0594
18	0.2110	0.1805	0.0494	0.0998	0.0529	0.0589

α denotes the probability that cells advance two age compartments in one time interval, β that for no advancement; d_1 and d_2 are natural cell death rates in the proliferating group and nonproliferating group, respectively; f is the probability that cells lose cycling capability and become resting cells in one unit time, γ the probability that cells regain cycling capability and return to the proliferating group.

The augmentation method is simple to implement and provides the best estimates in the least square sense with numerical stability. The method would be more valuable for the case of $\alpha_1=\alpha_s=\alpha_2$ and $\beta_1=\beta_s=\beta_2$ which is commonly assumed in the literature [12], [14], [16]. Under this assumption, we computed the initial cell age distribution which is then used, together with identified parameters, for calculating subsequent cell age, cell DNA, and FMF DNA distributions. The results are plotted in Fig. 7(a)–(c). It should be particularly noted that actual data points are also plotted in Fig. 7(a), comparing with the computed FMF data curve (generally, the computed data matches the actual data well).

The estimation of chemotherapeutic effects is straightforward, as both the cell-killing and cell-blocking rates are directly computable from measurements of cell DNA distributions and the knowledge of cell kinetic parameters of the cell system without drug. Simulated data are used to demonstrate the above method of calculating chemotherapeutic effects in the following. Cytosine arabinoside (Ara-C) is applied, as an example, to the CHO cell population. The Ara-C is known to kill cells only in the S phase and to block cell progression at the boundary between G_1 and S (positions of cell blocking in the cell cycle are different from one type of drug to another). Because of this property of Ara-C, one can obtain simple equations in terms of the cell-killing and the cell-blocking rates. Let p be the probability that cells would be blocked at the G_1/S boundary as a result of chemotherapy, and let c_i be the cell-killing rate in the ith age compartment in the S phase. Then one gets

$$x_{s1}(k+1)=(1-d_1)\Big[(1-c_1)\beta_s x_{s1}(k)+(1-p)\{\delta_1 x_{1p_1}(k)+\alpha_1 x_{1p_1-1}(k)\}\Big]$$

$$x_{s2}(k+1)=(1-d_1)\Big[(1-c_2)\beta_s x_{s2}(k)+(1-c_1)\delta_s x_{s1}(k)+(1-p)\alpha_1 x_{1p_1}(k)\Big]$$

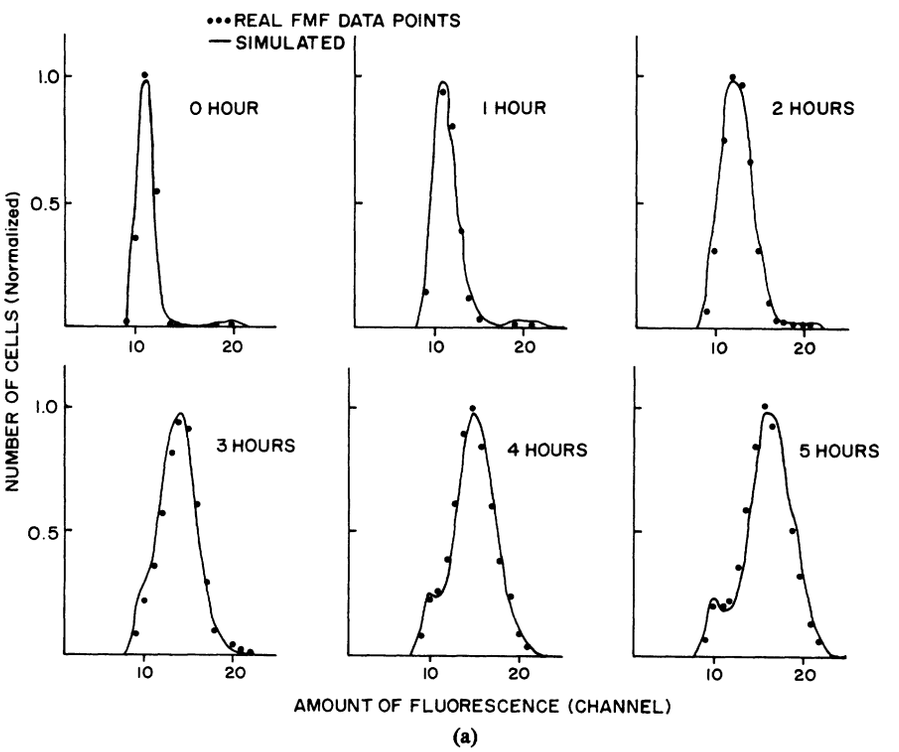
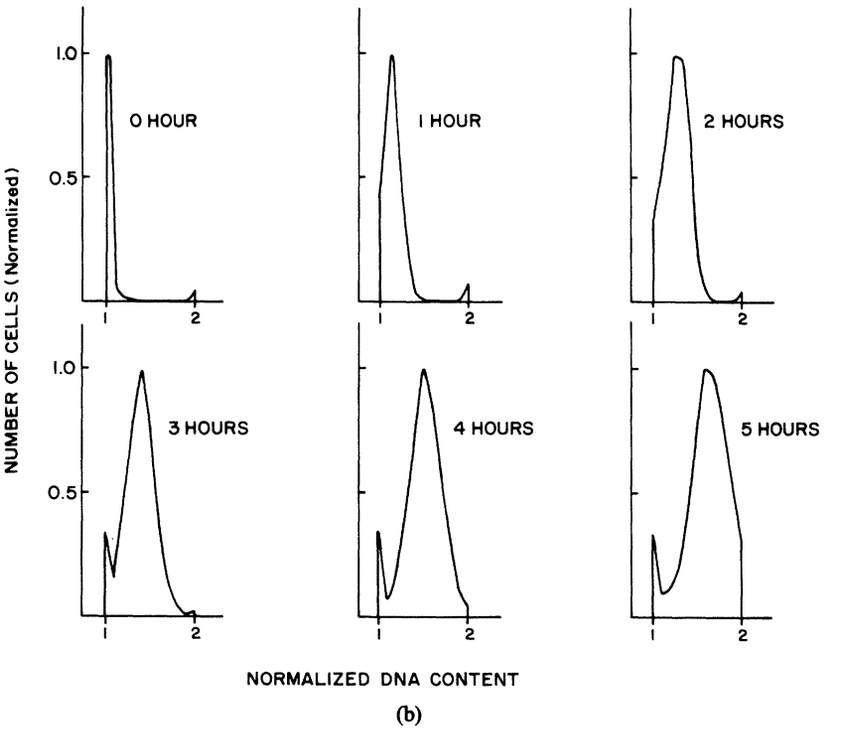
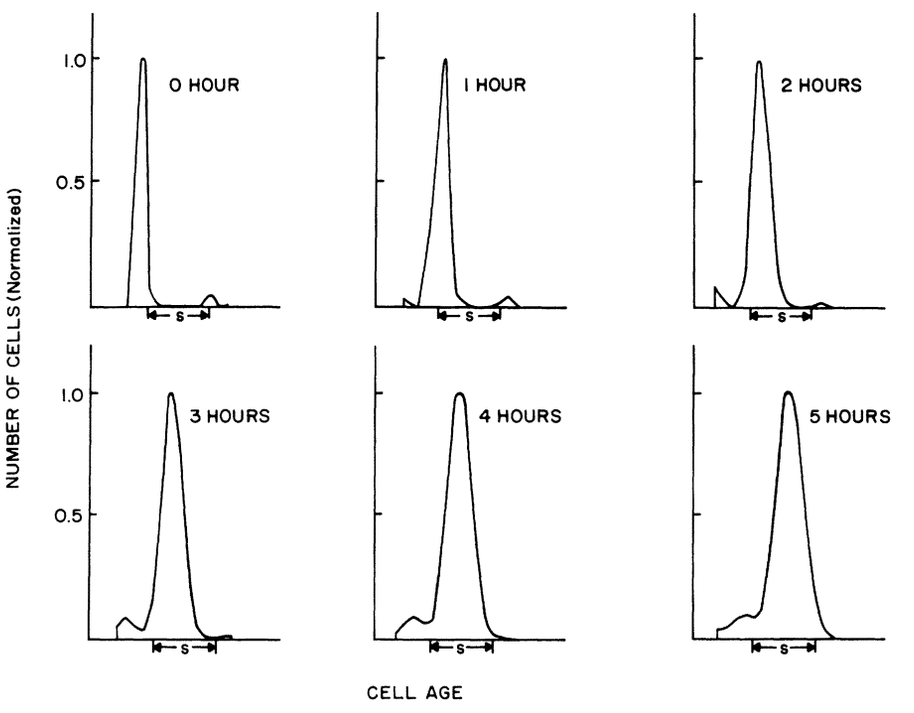


Fig. 7. Cell distributions of CHO cell population at times 0, 1, 2, 3, 4, and 5 h. (a) Simulated and real FMF DNA distribution. (b) True cell-DNA distribution. (c) Cell-age distribution.



(b)



(c)

Fig. 7. (*Continued*).

$$x_{s3}(k+1)=(1-d_1)\left[(1-c_3)\beta_s x_{s3}(k)+(1-c_2)\delta_s x_{s2}(k)+(1-c_1)\alpha_s x_{s1}(k)\right] \quad (13)$$

$$\vdots$$

$$x_{sp_s}(k+1)=(1-d_1)\left[(1-c_{p_s})\beta_s x_{sp_s}(k)+(1-c_{p_s-1})\cdot\delta_s x_{sp_s-1}(k)+(1-c_{p_s-2})\alpha_s x_{sp_s-2}(k)\right]$$

where d_1 denotes the cell death rate in the proliferating group. Note that all x_{si} for $i=1, 2, \cdots, p_s$ can be replaced by z_{si}.

In order to determine the parameter p one can obtain an equation,

$$\text{influx to } z_1\text{-outflux from } z_1 = z_1(k+1)-z_1(k) \quad (14a)$$

i.e.,

$$z_1(k+1)-z_1(k)=(1-d_1)t_{12}(k)+(1-d_1)pt_{s1}(k)-t_{s1}(k)-\left[d_1 z_1(k)+(d_2-d_1)x_m(k)-d_1 t_{s1}(k)\right] \quad (14b)$$

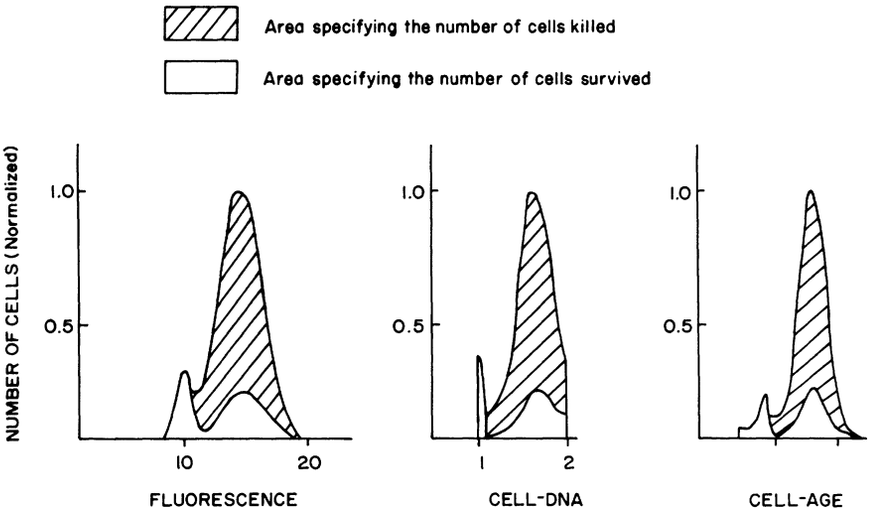


Fig. 8. Computer simulation of cells killed and cells survived following application of drug.

0.162	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.379	1.093	0.059
0.578	0.172	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.401	0.0
0.201	0.578	0.172	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.201	0.578	0.172	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.201	0.578	0.172	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.201	0.578	0.172	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.201	0.578	0.172	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.201	0.578	0.172	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.201	0.578	0.172	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.201	0.578	0.172	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.201	0.578	0.172	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.201	0.578	0.172	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.201	0.578	0.172	0.0
0.010	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.022	0.064	0.841

(a)

0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.191	0.521	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	-0.191	-0.521	-0.146	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	-0.191	-0.492	-0.146	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	-0.170	-0.492	-0.146	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	-0.170	-0.492	-0.137	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	-0.170	-0.467	-0.137	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.160	-0.463	-0.137	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.160	-0.463	-0.129	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.160	-0.434	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.150	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(b)

Fig. 9. Estimated transition matrices. (a) State-transition matrix Φ. (b) Input-transition matrix Ψ.

where d_1 represents the cell death rate in the proliferating group, and

$$t_{s1}(k)=\alpha_1 x_{1p_1-1}(k)+(1-\beta_1)x_{1p_1}(k)$$

$$t_{12}(k)=2\left(\alpha_2 x_{2p_2-1}(k)+(1-\beta_2)x_{2p_2}(k)\right).$$

From (14b), the parameter p can be calculated by

$$p=\frac{1}{t_{s1}(k)}\left[z_1(k+1)/(1-d_1)-z_1(k)-t_{12}(k)+\frac{d_2-d_1}{1-d_1}x_m(k)\right]+1 \qquad (14c)$$

where d_2 represents the cell death rate in the nonproliferating group. With the value of p it is easy to compute all c_i for $i=1, 2, \cdots, p_s$ from the equations in (13). Fig. 8 shows the computer simulation for cell distributions before and after the antitumor drug application, including the cells killed and those that survived. The late G_1 peaks of survival curves in Fig. 8 represent the accumulation of cells at G_1/S boundary as a result of progression delay. This simulation made use of the identified parameter values $\alpha=0.2110$, $\beta=0.1805$, $d_1=0.0494$, $d_2=0.0998$, $f=0.0529$, $\gamma=0.0589$, $p=0.9$, and $C_1=C_2=C_3=0.850$, $C_4=C_5=C_6=0.800$, $C_7=0.750$. Also the identified system matrices Φ and Ψ are given in Fig. 9.

VI. Conclusion and Discussion

We have developed a computationally efficient and numerically stable method for the identification of cell kinetic parameters used in the discrete-time model (1).

This method has great potential for both the study of cell kinetics and the design of optimal drug schedules, which are very important areas in cancer research.

A considerable amount of effort was made to estimate these parameters in our previous work [19], [20], yet the resulting methods were neither computationally efficient nor numerically stable. Such a result was mainly due to 1) the calculation of the inverse of a large matrix, 2) nonrecursiveness of the estimation algorithms, and 3) high dimensionality of the cell system, needed for the accurate description of cell cycling process. A conventional recursive least squares method as in [24] could cure some of the above drawbacks. However, the problem of matrix inversion, a main cause of numerical instability, remains unsolved along with that of high dimensionality of the cell system. As seen earlier, system decomposition and sequential triangularization techniques have solved all the drawbacks mentioned above and provided accurate estimates (in least square sense) of cell kinetic parameters.

It is assumed that all parameters except the number of cells remain constant with time; i.e., means and standard deviations of phase durations, cell cycle dispersion parameters, probabilities of cell death, cell recruitment, etc., are assumed to be constants. The system identification theory leads to numerically manageable results only when this assumption is made. This is the very practical reason why many modelers make this assumption, which is sometimes unrealistic from a biological point of view. In many cases, however, it is possible to use stationary models if the cell population parameters vary slowly.

The present method makes use of sequential data with the measurement interval ΔT_0, but it can be modified to use any sequential data with a random measurement interval $j\Delta T_0$, where j is a positive integer. However, such modifications may bring about computational difficulties due to the high nonlinearity of associated equations.

In reality, the chemotherapeutic parameters may be functions of the size of the population, the frequency of drug dosage, etc. However, the method described in this paper is based on our simple drug model in [17]. New schemes for identifying chemotherapeutic effects may be needed if the drug model is changed.

In spite of the speedy cell sorting capability of FMF techniques, there are few published quantitative data suitable for rigorous cell kinetic studies based on mathematical models. This results from the fact that experimenters look for biological phenomena instead of systematic analysis of data that will be obtained. The present work has suffered from the lack of abundant data, but useful data of CHO cells were available to demonstrate the utility of the method discussed in this paper. Furthermore, this paper gives experimenters an indication of the kinds of measurements needed within the framework of the state-space model to make a systematic study of cell kinetics.

Because of extreme simplicity and recursiveness, the algorithms discussed in this paper can be implemented in real time by dedicated mini- or microcomputers. Consequently, the method would be potentially invaluable to the analysis of large amounts of FMF DNA data and can be directly implemented with relatively simple laboratory instruments.

Acknowledgment

The authors thank Dr. Gray of the Lawrence Livermore Laboratory and Dr. Fried of the Sloan Kettering Cancer Institute for useful FMF data of CHO cells, and also thank the referees for constructive comments.

References

[1] J. Aroesty, T. Lincoln, N. Shapiro, and G. Boccia, "Tumor growth and chemotherapy: Mathematical methods, computer simulations, and experimental formulations," *Math. Biosci.*, vol. 17, pp. 243–300, 1973.

[2] K. J. Åström and P. Eykhoff, "System identification—A survey," *Automatica*, vol. 7, pp. 123–162, 1971.

[3] B. S. Baheti, R. R. Mohler, and H. A. Spang, III, "Second-Order Correlation Method for Bilinear System Identification," in *Proc. CDC Conf.*, New Orleans, 1977.

[4] K. Bahrami and M. Kim, "Optimal control of multiplicative control systems arising from cancer therapy," *IEEE Trans. Automat. Cont.*, vol. AC-20, pp. 537–542, Aug. 1975.

[5] C. Bruni, G. DiPillo, and G. Koch, "Bilinear systems: An appealing class of 'nearly linear' systems in theory and applications," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 334–348 Aug. 1974.

[6] G. Dahlquist and A. Björck, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[7] P. N. Dean and J. H. Jett, "Mathematical analysis of DNA distributions derived from flow microfluorometry," *J. Cell Biol.*, vol. 60, pp. 523–527, 1974.

[8] W. Dittrich and W. Göhde, "Impulsfluorometrie bei Einzelzellen in Suspensionen," *Z. Naturforsch*, 24b, p. 360, 1969.

[9] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. New York: Wiley, 1968.

[10] J. Fried, "Method for the quantitative evaluation of data from flow microfluorometry," *Comput. Biomed. Res.*, vol. 9, pp. 263–276.

[11] S. Gelfant, "A new concept of tissue and tumor cell proliferation," *Cancer Res.*, vol. 37, pp. 3845–3862, Nov. 1977.

[12] J. W. Gray, "Cell-cycle analysis of perturbed cell populations: Computer simulation of sequential DNA distributions," *Cell Tissue Kinet.*, vol. 9, pp. 499–516, 1976.

[13] J. W. Gray, P. N. Dean, and M. L. Mendelsohn, "Quantitative cell cycle analysis: Flow cytometry and sorting," in *Flow Cytometry and Sorting*. New York: Wiley, to be published.

[14] G. M. Hahn, "State vector description of the proliferation of mammalian cells in tissue culture," *Biophys. J.*, vol. 6, pp. 275–290, 1966.

[15] V. R. Karanam, D. A. Frick, and R. R. Mohler, "Bilinear system identification by Walsh functions," *IEEE Trans. Automat. Cont.*, vol. AC-23, pp. 709–713, Aug. 1978.

[16] M. Kim, K. Bahrami, and K. B. Woo, "A discrete-time model for cell-age, size and DNA distributions of proliferating cells, and its application to the movement of the labeled cohort," *IEEE Trans. Biomed. Eng.*, vol. BME-21, pp. 387–398, 1974.

[17] M. Kim, K. B. Woo, and S. Perry, "A quantitative approach to the design of antitumor drug dosage schedule via cell cycle kinetics and system theory," *Ann. Biomed. Eng.*, vol. 5, pp. 12–33, 1977.

[18] M. Kim and S. Perry, "Mathematical methods for determining cell DNA synthesis rate and cell age distribution utilizing flow microfluorometry," *J. Theoret. Biol.*, vol. 68, pp. 27–42, 1977.

[19] M. Kim, K. G. Shin, and S. Perry, "Estimation of cell kinetics parameters from flow microfluorometry," *Math. Biosci.*, vol. 38, pp. 77–89, 1978.

[20] M. Kim, B. Wheeler, and S. Perry, "Quantitative description of cell cycle kinetics under chemotherapy utilizing flow cytometry," *Cell Tissue Kinet.*, vol. 11, pp. 497–512, 1978.

[21] P. M. Kramer, L. L. Deavan, H. A. Crissman, and M. A. VanDilla, "The paadox of DNA constancy in heteroploidy," *Adv. Cell Mol. Biol.*, vol. 2, pp. 47–108, 1972.

[22] L. G. Lajtha and R. Oliver, "Cell population kinetics following different regimes of irradiation," *British J. Radiol.*, vol. 35, pp. 131–140, 1962.
[23] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[24] J. M. Mendel, "Multistage least-squares parameter estimators," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 775–782, Dec. 1975.
[25] L. W. Nelson and E. Stear, "The simultaneous on-line estimation of parameters and states in linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 94–98, Feb. 1976.
[26] C. C. Paige and M. A. Saunders, "Least squares estimation of discrete linear dynamic systems using orthogonal transformations," *SIAM J. Numer. Anal.*, vol. 14, pp. 180–193, Apr. 1977.
[27] L. Padulo and M. A. Arbib, *System Theory*. Philadelphia, PA: Saunders, 1974.
[28] S. Perry, "Cell kinetics and cancer therapy: History, present status, and challenges," *Cancer Treatment Rep.*, vol. 60, pp. 1699–1704, Dec. 1976.
[29] A. P. Sage and J. L. Melsa, *System Identification*. New York: Academic, 1971.
[30] E. Tse and J. Anton, "On the identifiability of parameters," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 637–645, Oct. 1972.
[31] M. A. Van Dilla, T. T. Trujillo, P. F. Mullaney, and J. R. Coulter, "Cell microfluorometry: A method for rapid fluorescence measurement," *Science*, vol. 163, pp. 1213–1214, 1969.
[32] C. Van Loan, "Lectures in least squares," Dep. Comput. Sci., Cornell Univ., *Tech. Rep.*, TR 76-279, 1976.
[33] L. M. Van Putten, H. J. Keizer, and J. H. Mulder, "Perspectives in cancer research: Synchronization in tumor chemotherapy," *European J. Cancer*, vol. 12, pp. 79–85, 1976.

Attributed Grammar—A Tool for Combining Syntactic and Statistical Approaches to Pattern Recognition

WEN-HSIANG TSAI, MEMBER, IEEE, AND KING-SUN FU, FELLOW, IEEE

Abstract—**Attributed grammars are defined from the pattern recognition point of view and shown to be useful for descriptions of syntactic structures as well as semantic attributes in primitives, subpatterns, and patterns. A pattern analysis system using attributed grammars is proposed for pattern classification and description. This system extracts primitives and their attributes after preprocessing, performs syntax analysis of the resulting pattern representations, computes and extracts subpattern attributes for syntactically accepted patterns, and finally makes decisions according to the Bayes decision rule. Such a system uses a combination of syntactic and statistical pattern recognition techniques, as is demonstrated by illustrative examples and experimental results.**

I. Introduction

COMBINING syntactic and statistical pattern recognition approaches has been advocated by several investigators [1]–[4], [7], [22] in the past decade. The motivation arises from the fact that neither the syntactic approach nor the statistical approach alone is adequate for some practical applications; the former is weak in handling noisy patterns and numerical semantic information [10], [24], and the latter is unable to describe complex pattern structures and subpattern relations. Since the advantage of one approach appears to be the disadvantage of the other, a hybrid model is desirable that incorporates the advantages of both and is thus more useful to real applications. In this paper we propose the use of attributed grammars as a tool for combining the syntactic and the statistical approaches to pattern recognition.

Attributed grammars were first formulated by Knuth [5] to assign semantics or meanings to context-free languages from the computational linguistics point of view. Each production rule of an attributed grammar consists of a syntactic rule and a semantic rule, the former being used to specify language syntax and the latter to add contextual semantics. Illustrative examples of applications of such semantic formalism to patterns described by picture description languages (PDL) [6] are found in [7]. The necessity of using pattern semantics to facilitate the utilization of contextual information was also emphasized in [23], although attributed grammars were not used there. Tang and Huang [8] applied attributed grammars to image understanding. You and Fu [9], [25] used attributed gram-

Manuscript received November 15, 1979; revised September 10, 1980. This work was supported by the Office of Naval Research under Contract N00014-79-C-0574.

W. H. Tsai was with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907. He is now with the Institute of Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan, Republic of China.

K. S. Fu is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907.

0018-9472/80/1200-0873$00.75