

Estimation of Cell Kinetic Parameters from Flow Microfluorometry*

M. KIM AND KANG G. SHIN

Cornell University, Ithaca, New York 14853

AND

SEYMOUR PERRY

National Institutes of Health, Bethesda, Maryland

Received 21 April 1977

ABSTRACT

This paper considers the estimation method for unknown parameters as well as cell-age state vectors used in the discrete-time model [12] previously described characterizing the dynamics of cell cycle and proliferating kinetics for a cancer-cell population. An iterative algorithm for determining optimal values of the parameters and age state vectors in the least-squares sense is derived on the basis of a set of sequential cell-DNA distributions. Once the parameters and initial cell-age distribution are determined, the time-course behavior of the cell-age and cell-DNA distributions for the given population are computed. A Chinese-hamster cell system is chosen for illustrating the quantitative technique developed. A computer simulation of the CHO cell population is shown.

INTRODUCTION

In recent years, cell-cycle kinetics of perturbed cell populations have been investigated by the use of cell-DNA distributions with increasing frequency. Recently pulse cytophotometry or flow microfluorometry (FMF) [8, 23, 25] has permitted rapid measurement of the DNA content per cell, so that the cell-DNA distribution for a large population may be obtained within a short period with satisfactory precision. Hence the FMF technique has been used increasingly in analyzing the effect of chemical agents on

*The research reported in this paper was supported in part by the grant number 1-R01 CA 16259, awarded by the National Cancer Institute.

cell-cycle kinetics [6, 11, 14, 24, 27], and quantitative methods have been employed for analyzing FMF data have been proposed [2, 5]. However, in general, the experimental data obtained from FMF techniques have not been analyzed by rigorous quantitative methods, e.g., mathematical modeling. The following procedures are employed in mathematical modeling for a given biological system:

- (i) determination of model structure by quantifying all necessary biological phenomena and establishing the mathematical relationships among them,
- (ii) identification of all parameters employed in (i) on the basis of the observation of the system,
- (iii) validation test.

Kim et al. [12] extended the discrete model [10] to include a quantitative expression of FMF DNA data for the state representation. With this model Kim and Perry [13] computed the mean DNA synthesis rate and thus the transformation matrix relating cell-DNA distributions to FMF DNA data under the assumption that the system parameters were known. Gray [9] extended the maturity-state concept [10] to obtain differential equations, and computed DNA distributions from the assumed initial cell-age distribution and the system parameters determined by trial and error.

However, these methods would fail if (a) the system parameters are not known, or (b) either the initial cell-age distribution is not known or the system parameters are not easily adjustable by eye. In this paper it is shown that under the above circumstances the cell-age distribution and system parameters must be determined concurrently. This paper presents a rigorous mathematical method for (1) determining the cell-age distribution and system parameters simultaneously from FMF data, and (2) computing cell-DNA and cell-age distributions. Finally the derived method is applied to a population of Chinese-hamster (CHO) cells.

IDENTIFICATION OF CELL-AGE DISTRIBUTION AND UNKNOWN SYSTEM PARAMETERS

Since the cell-age distribution describes the cell-cycle kinetic state, it is useful to know it for a cell population at any given time. However, the cell-age distribution cannot be observed directly. Furthermore, the mathematical representation of the cell-cycle kinetics usually includes unknown dispersion parameters to describe the cell progression in the cell cycle. The problem is to compute the dispersion parameters and cell-age distribution from experimentally observed cell-DNA distributions.

First consider the quantification [12] of the cell-age and cell-DNA distributions as follows:

(i) the cell-age vectors $\mathbf{x}_p(k)$ for proliferating cells and $\mathbf{x}_m(k)$ for nonproliferating cells are defined:

$$\begin{aligned}\mathbf{x}_p(k) &\triangleq [x_1(k), \dots, x_i(k), \dots, x_n(k)]^T, \\ \mathbf{x}_m(k) &\triangleq [x_{n+1}(k), \dots, x_{n+j}(k), \dots, x_r(k)]^T,\end{aligned}\quad (1)$$

where $x_i(k)$ is the number of cells in the i th age compartment at time k . Observe that

$$\begin{aligned}\mathbf{x}(k) &\triangleq [\mathbf{x}_p^T(k) \ ; \ \mathbf{x}_m^T(k)]^T \\ &= [x_1(k), \dots, x_r(k)]^T.\end{aligned}\quad (2)$$

(ii) The cell-DNA distribution vector $\mathbf{z}(k)$ is defined:

$$\mathbf{z}(k) \triangleq [z_1(k), \dots, z_i(k), \dots, z_q(k)]^T, \quad (3)$$

where $z_i(k)$ is the number of cells in the i th DNA-content state at time k .

Secondly consider a multivariable linear discrete model [12] describing the dynamics of the cell-cycle kinetics:

$$\mathbf{x}(k+1) = \Phi(\boldsymbol{\theta})\mathbf{x}(k), \quad (4)$$

where $\mathbf{x} \triangleq [\mathbf{x}_p^T \ ; \ \mathbf{x}_m^T]^T$ is the r -dimensional cell-age distribution vector, $\boldsymbol{\theta}$ is an l -dimensional dispersion-parameter vector (i.e., $\boldsymbol{\theta} = [\theta_1, \dots, \theta_i, \dots, \theta_l]^T$, and all θ_i 's are dispersion parameters) assumed to be time-invariant, and Φ is the local state-transition matrix relating the cell-age vector at time k to its corresponding cell-age distribution at time $k+1$.

Finally the measurement equation [12] is defined to be the deterministic relationship between the cell-age and cell-DNA distributions,

$$\mathbf{z}(k) = Q\mathbf{x}(k), \quad (5)$$

where \mathbf{z} is the q -dimensional cell-DNA distribution vector, and Q is a linear transformation of cell-age distribution into cell-DNA distribution.

Now the problem is to estimate the parameter vector $\boldsymbol{\theta}$ and cell-age vector $\mathbf{x}(k)$ by using Eqs. 4 and 5, and the proper number of sequential FMF DNA distributions. The observability concept in system theory is employed to describe how to determine $\mathbf{x}(k_0), \mathbf{x}(k_0+1), \dots$, from a sequence of measurements $\mathbf{z}(k_0), \mathbf{z}(k_0+1), \dots, \mathbf{z}(k_0+k^*-1)$. Observe that once $\mathbf{x}(k_0)$

and θ are known, $\mathbf{x}(k_0+1), \mathbf{x}(k_0+2), \dots$ can be computed by Eq. 4. Hence the problem is reduced to the computation of $\mathbf{x}(k_0)$ and θ based on a sequence of FMF DNA data.

It is convenient to define a $qk^* \times r$ matrix and a $qk^* \times 1$ vector as follows:

$$A_{k^*}(\theta) = \begin{bmatrix} \text{---} Q \text{---} \\ Q\Phi(\theta) \text{---} \\ \vdots \\ Q\Phi^{k^*-1}(\theta) \text{---} \end{bmatrix}, \quad (6)$$

$$\mathbf{Z}(k^*, k_0) = \begin{bmatrix} \text{---} \mathbf{z}(k_0) \text{---} \\ \text{---} \mathbf{z}(k_0+1) \text{---} \\ \vdots \\ \text{---} \mathbf{z}(k_0+k^*-1) \text{---} \end{bmatrix}. \quad (7)$$

Then it is easy to see that

$$A_{k^*}(\theta)\mathbf{x}(k_0) = \mathbf{Z}(k^*, k_0). \quad (8)$$

In addition to the mathematical relationship (8), one must consider the following physical aspects which put constraints on the vectors $\mathbf{x}(k_0)$ and θ :

(i) Since every component of \mathbf{x} represents the number of cells in the corresponding age compartment, it must be nonnegative, i.e.,

$$x_i(k_0) \geq 0 \quad \text{for } i = 1, 2, \dots, r. \quad (9)$$

(ii) The cell-age distribution at any given time is considered a reassignment of its corresponding cell-DNA distribution, so the total population at time k_0 must be equal to the summation of components of $\mathbf{x}(k_0)$ as well as that of $\mathbf{z}(k_0)$, i.e.,

$$\sum_{i=1}^r x_i(k_0) = T_{\text{pop}}, \quad (10)$$

where

$$T_{\text{pop}} = \sum_{i=1}^q z_i(k_0). \quad (11)$$

(iii) Since the dispersion parameters represent the probabilities for the cell progression in the cell cycle, θ must satisfy

$$0 \leq \theta_i \leq 1 \quad \text{for } i = 1, 2, \dots, l. \quad (12)$$

Kim and Perry [13] computed $\mathbf{x}(k_0)$ under the assumptions:

- (a) the parameter vector θ is known,
- (b) there are no constraints on $\mathbf{x}(k_0)$ like (9) and (10).

However, if either assumption (a) or (b) is not valid, this computational algorithm will fail.

In this paper assumptions (a) and (b) will be removed, and thus $\mathbf{x}(k_0)$ and θ will be obtained simultaneously by a successive-approximation method. The cell-age and cell-DNA distributions at any time can be computed by using Eqs. (4) and (5) following the computation of $\mathbf{x}(k_0)$ and θ . However, the determination of $\mathbf{x}(k_0)$ and θ may be difficult due to the constraints (9), (10), and (12) on them resulting from the physical aspects of the system. This problem is therefore converted to a constrained least-squares problem as follows:

PROBLEM (P)

Find an r -dimensional vector $\mathbf{x}^*(k_0)$ and an l -dimensional vector θ^* minimizing

$$J = \frac{1}{2} \|A_{k^*}(\theta)\mathbf{x}(k_0) - \mathbf{Z}(k^*, k_0)\|^2 \quad (13)$$

subject to Eqs. (9), (10), and (12), where $\|\cdot\|$ denotes the Euclidean norm.

Let's turn to the problem of finding the value of k^* . $\mathbf{x}(k_0)$ must be determined uniquely, since the cell-age distribution at time k_0 must be unique. To get a unique solution $\mathbf{x}(k_0)$ for Problem (P), the rank of the matrix $A_{k^*}(\theta)$ must be r [15] [the rank of a matrix is defined to be the number of independent rows (or columns) of the matrix]. The smallest integer value of k^* for which the rank of $A_{k^*}(\theta)$ is r is the minimum number of \mathbf{z} 's, i.e., observations, required to determine $\mathbf{x}(k_0)$ uniquely. This value of k^* is called the observability index. The existence of k^* is guaranteed by the observability condition. The procedure for finding the observability index k^* is shown in Fig. 1.

Problem (P) is a nonlinear least-squares problem with linear inequality constraints. This problem is generally difficult to solve from both analytical and computational viewpoints. However, it has a special property, i.e., it is linear in $\mathbf{x}(k_0)$ though nonlinear in θ . It has been found that the computational difficulties can be circumvented by separating the linear part from the nonlinear one as illustrated in Fig. 2.

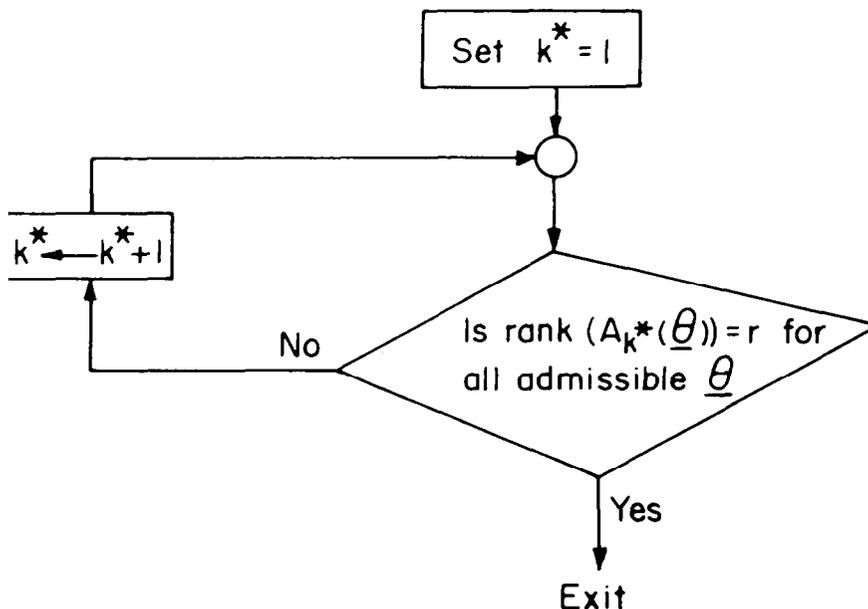


FIG. 1. Procedure for finding the observability index k^* . The observability condition prevents infinite looping.

With the aid of the Kuhn-Tucker conditions for optimality [1, 7], a linear least-squares problem with linear inequality constraints (LSI problem) can be solved. Lawson and Hanson [15] developed a program for solving the least-distance programming (LDP) problem by using QR decomposition and the nonnegative least-squares (NNLS) algorithm. Shin [22] extended this method to solve the LSI problem. Hence the LSI problem is at present solvable.

The derived algorithm in this study, at each iteration, solves the LSI problem twice to compute

- (i) $\mathbf{x}(k_0)$ with θ known, and
- (ii) an optimal change in θ , which generates an improved estimation of θ .

In order to perform (ii) one must compute the Jacobian matrix G of residual vector $\mathbf{r}(\mathbf{x}(k_0), \theta)$ with respect to θ , i.e.,

$$G = [g_{ij}], \quad (14)$$

where

$$\mathbf{r}(\mathbf{x}(k_0), \theta) = A_{k^*}(\theta)\mathbf{x}(k_0) - \mathbf{Z}(k^*, k_0) \quad (15)$$

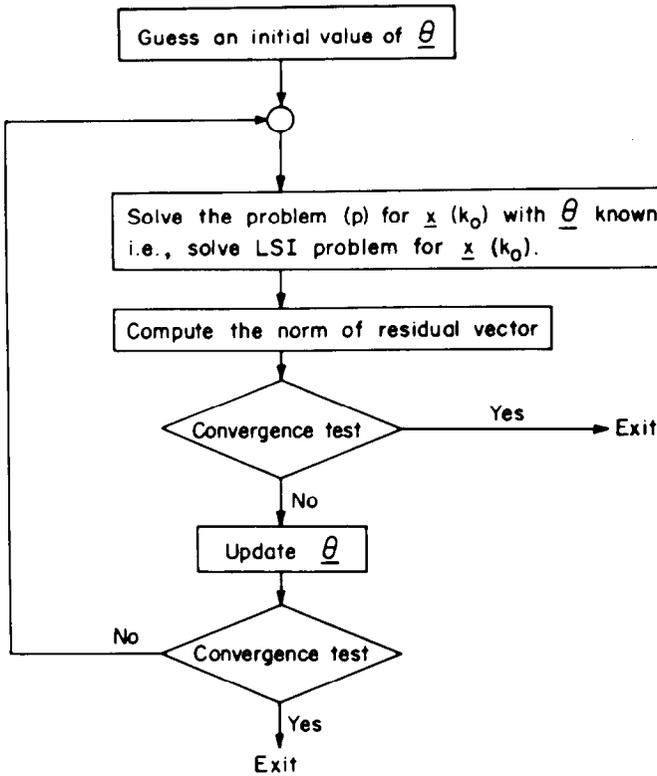


FIG. 2. Flow diagram for solving Problem (P) by making use of its special structure, i.e., linear in $\underline{x}(k_0)$ but nonlinear in θ .

and

$$g_{ij} = \frac{\partial r_i(\underline{x}(k_0), \theta)}{\partial \theta_j} \quad \text{for } i=1,2,\dots,kq, \quad j=1,2,\dots,l. \quad (16)$$

Generally, this computation is known to be costly if \mathbf{r} is a complicated function. This difficulty can be eased by replacing the G with an approximation to G . Such an approximation is usually obtained by either the

forward or the central difference method. In choosing an approximation to G , caution must be exercised so as not to disturb the convergence property of the algorithm.

APPLICATION OF THE ALGORITHM TO A CHO CELL POPULATION

Flow microfluorometric DNA distributions for CHO cells are selected to show the validity of the derived method in the previous discussion. The CHO cell population grows exponentially, and thus we assume it consists of only proliferating cells. Note that this fact yields simple forms for the matrices Φ , Q and the vectors θ , x . However, the method is equally applicable to cell populations which also include nonproliferating cells.

Puck et al. [20] reported the cell generation time of CHO cells to be 12.4 hours, while Kramer et al. [14] proved the cell generation time for the unperturbed CHO cell population to be 16.5 hours and also showed that the phase durations of the G_1 and S phases are functions of cell concentrations and culture conditions for FMF experiments. Gray [9] reported that the mean generation time for the perturbed CHO cell population is 12.1 hours, and the average phase transit times of G_1 , S , and $G_2 + M$ phases are 4.5, 4.8, and 2.8 hours respectively. Hence the integral ratio of the phase durations, $T_{G_1} : T_S : T_{G_2+M}$, is chosen to be 4:4:2. Observe that this specifies the ratio of the numbers of cell-age compartments in the G_1 , S , and $G_2 + M$ phases. The mean duration of one age compartment was defined to be the (biological) unit time, i.e., (unit time) = $\Delta T_0 = T_0/r$, where T_0 is the mean cell generation time and r is the total number of age compartments for the proliferating cells [12].

The cell-DNA distribution of exponentially growing CHO cells was recorded at various times (2.5, 3.5, 7.5, 8.5, 12.0, 13 hours) after release from a thymidine block which is known to reduce the rate of DNA synthesis drastically [14]. To obtain a cell-DNA distribution from a FMF measurement, dispersion in the measured values of DNA content must be taken into account. Dispersive effects are assumed to be distributed normally, and a dispersion-free DNA distribution is computed from a measure DNA distribution.

It can be seen that it becomes more difficult to obtain consecutive experimental cell-DNA distributions as the measurement interval ($=\Delta T_0$) diminishes. In addition, when the dimension of x becomes larger, the computation time increases very rapidly. However, we may obtain more accurate curves for the cell-age distributions at the expense of the aforementioned difficulties. It is therefore desirable to choose an optimal unit time so that an optimal measurement can be made. In this application the total number of age compartments is chosen to be 10 and the total number of the DNA content compartments 12.

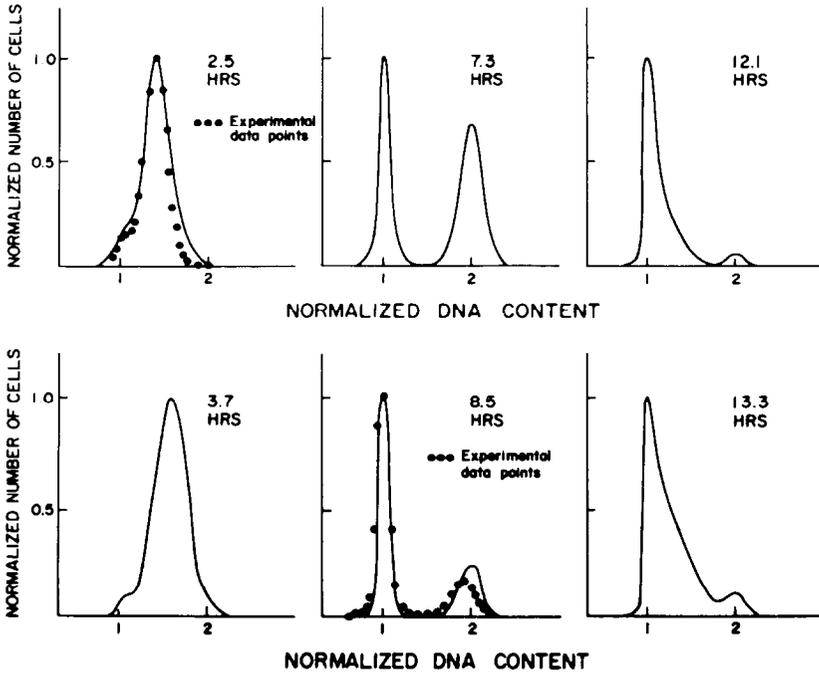


FIG. 3. Computed cell-DNA distributions of CHO cells at times 2.5, 3.7, 7.3, 8.5, 12.1, and 13.3 hours after release from thymidine block. The experimental DNA data points (from Ref. [14]) are superimposed on the distributions at times 2.5 and 8.5 hours after release from thymidine block. These computed distributions show consistency with the experimental ones.

The computed cell-DNA and cell-age distributions at times 2.5, 3.7, 7.3, 8.5, 12.1, and 13.3 hours after release from thymidine block are shown in Fig. 3 and Fig. 4 respectively. They have proven to be fairly consistent with experimental DNA distributions in Ref. [14].

DISCUSSION

In the first half of this paper an algorithm for determining the cell-age distributions and system parameters of a cancer-cell population from experimental cell-DNA distributions is derived. The algorithm uses a successive-approximation method, which improves iteratively the estimation of the

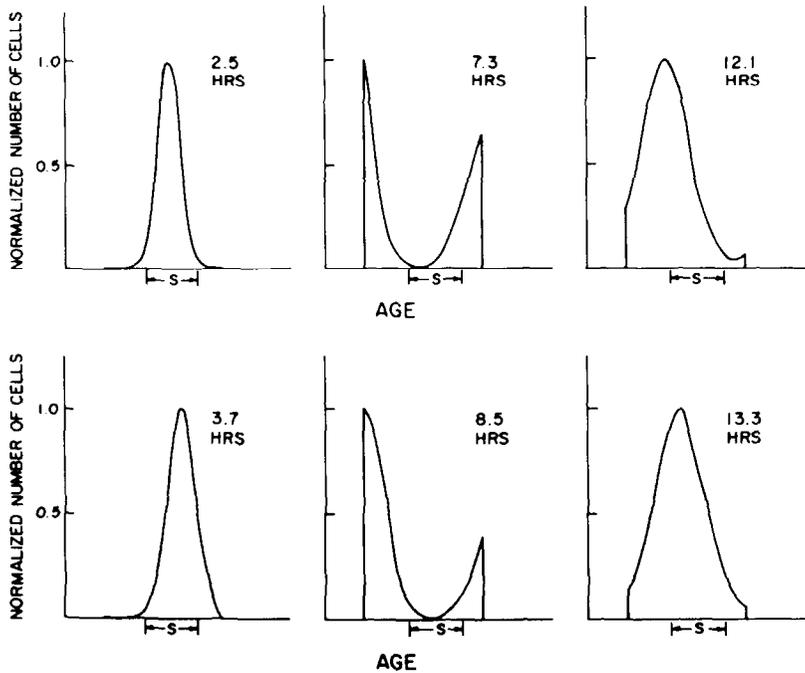


FIG. 4. Computed cell-age distributions of CHO cells at times 2.5, 3.7, 7.3, 8.5, 12.1, and 13.3 hours after release from thymidine block.

cell-age state vector and system parameters. At each iteration we solved an inequality-constrained least-squares problem to obtain an optimal change of $x(k_0)$ and θ .

The second half of this paper is an application of the developed method to the CHO cell population. In this application the transformation matrix Q from the cell-age distribution to its corresponding cell-DNA distribution is assumed to be time-invariant, which implies that the mean DNA synthesis rate remains unchanged. This assumption can be corrected by building an on-line system in which the transformation matrix Q is updated according to the change of the DNA synthesis rate.

In the algorithm the cell-DNA and cell-age distributions are represented by the cell-DNA content and age vectors, which are discrete-time models. Accuracy can be improved by increasing the number of compartments representing the discrete DNA contents and cell age, at the expense of computation time and some possible measurement difficulties. These possible measurement difficulties can be excluded by taking an experimentally realizable measurement time interval I that is an integral multiple of the unit time, i.e., $I = iT_0/r$ for some positive integer i . In such a case Eq. (6)

and Eq. (7) must be modified as follows:

$$A_{k^*}(\theta) = \begin{bmatrix} Q \\ \text{---} \\ Q\Phi^I \\ \text{---} \\ \vdots \\ \text{---} \\ Q\Phi^{(k^*-1)I} \end{bmatrix}, \quad (6')$$

$$Z(k^*, k_0) = \begin{bmatrix} z(k_0) \\ \text{---} \\ z(I+k_0) \\ \text{---} \\ \vdots \\ \text{---} \\ z[(k^*-1)(I+k_0)] \end{bmatrix}. \quad (7')$$

When the observation period of the given system is long, discrepancies may exist between the actual and simulated distributions due to the nonlinearity of the system and the time variation of the system parameters. This can be corrected by building a semi-on-line system in which the discrete-time model and the method developed are updated at fixed time intervals during which the linearity of the model and the time invariance of the parameters provide a good approximation to the system behavior.

In the CHO cell population simulated, both the state transition matrix Φ and initial cell-age distribution $x(k_0)$ were unknown but determined simultaneously.

Recently, the FMF technique has been used increasingly in studying the effect of chemical agents on cell-cycle kinetics [6, 11, 14, 24, 27]. The method developed here can be extended to examine this effect by a simple modification. Hence the knowledge of the system parameters and cell-age distribution computed by the method developed can potentially be useful in the analysis of the effects of cytotoxic chemotherapy.

REFERENCES

- 1 M. Avriel, *Nonlinear Programming Analysis and Methods*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976.
- 2 H. Baisch, W. Göhde, and W. Linden, *Rad. and Environ. Biophys.* **12**, 31 (1975).
- 3 J. E. Cleaver, *Thymidine Metabolism and Cell Kinetics* (A. Neuberger, Ed.), North-Holland, Amsterdam, 1967.
- 4 H. A. Crissman, R. J. Kissane, M. S. Oka, R. A. Tobey, and J. A. Steinkamp, Flow microfluorometric approaches to cell kinetics, in *Proceedings of the 29th Annual Symposium on Fundamental Cancer Research*, Houston, Texas, 10-21, Mar. 1976.
- 5 P. N. Dean and W. M. Jett, *J. Cell Biol.* **60**, 523 (1974).
- 6 F. Dietzel, D. Ringleb, U. Schneider, and H. Wricke, *Strahlentherapie* **149**, 438 (1975).

- 7 A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, 1968.
- 8 W. Göhde, in *Fluorescence Technique in Cell Biology* (A. Thaer and M. Servetz, Eds.), Springer, New York, 1973, pp. 79–88.
- 9 J. W. Gray, Cell-cycle analysis of perturbed cell populations: computer simulation of sequential DNA distributions, *Cell Tissue Kinet.* **9**, 499–516 (1976).
- 10 G. Hahn, Cellular kinetics, cell cycle and cell killing, *Biophysik* **4**, 1–14 (1967).
- 11 G. M. Hahn, J. Braun, and I. Har-Kedar, *Proc. Nat. Acad. Sci. U.S.A.*, **72**, 937 (1975).
- 12 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. Biomedical Eng.* (Sept. 1974).
- 13 M. Kim and S. Perry, Mathematical methods for determining cell DNA synthesis rate and cell age distribution utilizing microfluorometry, *J. Theoret. Biol.*, to be published.
- 14 P. M. Kramer, L. L. Deavan, H. A. Crissman, and M. A. VanDilla, The paradox of DNA constancy in heteroploidy, *Advances in Cell Molecular Biol.* **2**, 47–108 (1972).
- 15 C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N.J., 1974.
- 16 R. C. K. Lee, *Optimal Estimation, Identification, and Control*, MIT Press, Cambridge, Mass., 1964.
- 17 J. M. Mendel, Multistage least-squares parameter estimator, *IEEE Trans. Automatic Control* (Dec. 1975).
- 18 B. Noble, *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, N.J., 1969.
- 19 L. Padulo and M. A. Arbib, *System Theory*, Saunders, Philadelphia, 1974.
- 20 E. T. Puck, P. Sanders, and D. Peterson, *Biophys. J.* **4**, 441 (1964).
- 21 A. P. Sage and J. L. Melsa, *System Identification*, Academic, 1971.
- 22 K. Shin, System identification for kinetics of a cellular proliferation of a cancer cell population, M. S. Thesis, Cornell Univ., Ithaca, N.Y., 1976.
- 23 J. A. Steinkamp, M. J. Fulwyler, J. R. Coulter, R. D. Hiebert, J. L. Horney, and P. R. Mullaney, A new multiparameter separator for microscopic particles and biological cells, *Rev. Sci. Instrum.* **44**, 1301 (1973).
- 24 R. Tobey and H. Crissman, Use of flow microfluorometry in detailed analysis of effect of chemical agents on cell cycle progression, *Cancer Res.* **32**, 2726–2732 (1972).
- 25 M. A. VanDilla, T. T. Trujillo, P. F. Mullaney, and J. R. Coulter, Cell microfluorometry: a method for rapid fluorescence measurement, *Science* **163**, 1213–1214 (1969).
- 26 C. Van Loan, *Lectures in Least Squares*, TR 76-279, Department of Computer Sci., Cornell Univ., Ithaca, N.Y., 1976.
- 27 M. Wannemacher, E. Esser, J. Glupe, and J. Schumann, *Strahlentherapie* **147**, 1 (1974).