This document adapts the sequencing control reported in Part V of this paper to individual plant requirements and goals.

A regression model is used to relate measures of plant performance to certain control parameters. This relationship is periodically recomputed using statistical analysis of operational data.

A pertinent decision rule is derived by optimal control theory.

Fabrication and assembly operations

Part VI Parameter values for sequencing control

by S. Gorenstein

The scheduling priority formula developed in Part V of this paper partially goes beyond ordinary project network scheduling techniques. Its detailed sequences for individual work activities in fabrication and assembly shops consider delivery and resource constraints as well as the requirements for work already in progress.

sequencing control refinements

Further refinements of the sequencing control can be achieved by fitting the schedule to individual plant requirements or goals, and by using operational historical data for updating purposes. The necessary statistical analysis of the process data and a pertinent decision rule are discussed here.

Some of the overall goals of interest in scheduling operations are:

- Minimize the excess of the throughput time over the critical-path length
- Minimize the dollar value of waiting time, including the time lost in waiting for shipment
- Minimize the dollar value of lateness time or some other penalty measure
- Maximize the utilization of manpower and/or equipment

To optimize the priority assignments for the desired goals, we assign to each term of a priority function a certain parameter value according to the weight we want to give to that respective term. Letting u_j denote these parameters (where $j = 1, 2, 3, 4, 5$), and adding a "value" term, V, the scheduling priority formula developed in Part V can be expressed as

$$f_i = \frac{u_1 M_i + u_2 F_i + u_3 S_i + u_4 / V_i}{D_i^{u_5}}, \tag{1}$$

for which job i precedes job k if $f_i < f_k$ and where the variables have the following denotations:

M expected duration of a job
F total float (latest start time minus earliest start time minus M) of a job at time t
S slack (difference between due date and earliest possible completion date) of a job at time t, negative values indicating lateness
V importance, or "value," of the end product to which a job contributes
D node density (as defined in Part V of this paper) of a job.

regression procedure

A method of choosing the needed parameters for Equation 1 is developed in this discussion. It should be kept in mind that we want to optimize a production system in its day-to-day operation. Since the selected parameter values affect the plant, free experimentation is not advisable. The chief goal is to improve actual system performance. Therefore, the parameter values selected should optimize or improve the system performance rather than merely yield information.

A least-squares regression model is now used to express some system state, x, as a function of the scheduling parameters, u_j, for $j = 1, 2, \cdots, n$. In the particular case mentioned above, $n = 5$. For example, it is reasonable to expect that some performance measure, such as dollar value of lateness, would be affected by the variables in the priority function of Equation 1. We can think of this lateness value as system state x, and use a least-squares regression model to estimate the relationship between x and the parameters u_j. The data for this model would come from past experience with the use of the priority function as given in Equation 1. Although other regression models could be considered, a linear function of the type

$$x = b_1 u_1 + b_2 u_2 + \cdots + b_5 u_5$$

is assumed for the sake of simplicity and ease of computation. After fitting such a function, one strategy for reducing x is to increase the parameters with negative coefficients and decrease those with positive coefficients. Caution should be exercised to stay within or near the values for which process data are available. The linear function, being an approximation of the true functional relationship, is only valid in some limited region, usually within the region for which we have data, and sometimes in some slightly larger region. This caution also applies to any non-linear function that might be fitted.

Another function, somewhat related to the form of Equation 1, is:

$$x = \frac{b_1u_1 + b_2u_2 + b_3u_3 + b_4u_4}{b_5e^{u_5}}$$

By choosing the denominator in this exponential form, rather than in the form of Equation 1, we can transform to an equation that leads to a system of linear regression equations. A non-linear system of regression equations would be much more difficult to solve and require more computing time. Letting $y = e^{u_5}$,

$$x = \frac{b_1u_1}{b_5y} + \frac{b_2u_2}{b_5y} + \frac{b_3u_3}{b_5y} + \frac{b_4u_4}{b_5y}$$

$$= k_1z_1 + k_2z_2 + k_3z_3 + k_4z_4$$

where

$$\frac{b_i}{b_5} = k_i, \qquad \frac{u_i}{y} = z_i, \qquad \text{and} \qquad i = 1, 2, 3, 4.$$

An assumed functional form can be chosen on the basis of an analysis of empirical data. The analysis may indicate a relationship between x and powers and products of the parameters u_1, u_2, u_3, and u_4. Then,

$$x = b_1u_1 + b_2u_2 + b_3u_3 + b_4u_4 + b_{11}u_1^2 + \cdots + b_{44}u_4^2 + b_{ij}u_iu_j,$$

in which $i = 1, 2, 3, 4, \qquad j = 1, 2, 3, 4, \qquad j \neq i,$

would allow for the consideration of as many as 14 variables, and would constitute a large class of possible relationships. Of course, any function appropriate for the data can be considered. The resulting function could then be minimized.

selection of variables

An important task is to select those parameters in the scheduling formula that exercise significant influence on the system state. These parameters (called independent, input, or control variables) can be chosen by two different regression procedures. In approximating the functional relationship between the system state, x, and the variables (u_i, u_i^2, and u_iu_j in the case above), it is, of course, desirable to fit as simple an equation as possible. It is not necessary to include all the variables in the relationship. Of interest are only those variables that provide a function useful for our purposes in exercising some control.

A recommended procedure, called *step-wise forward regression,*[1] is to compute the regression of x on each variable, one at a time, and accept the variable that has the smallest residual sum of squares. Then compute the regression of x on this first chosen variable paired with all others, and select the pair that has the least residual sum of squares. Compute the regression on all the triplets of variables that include the already selected pair plus a third variable, and select the triplet with the least residual sum of squares. Continue in this fashion until (1) the residual sum of squares reaches a desired small percentage level of the total sum of squares and (2) an additional variable only causes a relatively small reduction in the residual sum of squares. Alternatively, variables could be introduced on the basis of an objective testing procedure. For example, the F test[1,2] may show them not to be significant at some low-percentage level; however, this test would only be applicable in the case of normally distributed random variables. Since normality cannot be guaranteed in all cases, it is recommended that the amount of "explained" sum of squares (rather than the F test) be used as a criterion.

The step-wise forward procedure is recommended also because it is less sensitive to the effect of round-off errors in the inversion of ill-conditioned matrices.[3,4] These errors are especially noticeable if squares and products are considered as variables.

As an alternative method, a *step-wise backward regression* procedure can be used. For this method, compute the regression of x on all variables and eliminate variables one by one. Continue eliminating variables as long as the desired level of residual sum of squares is still maintained and, at the same time, the variable dropped would reduce the residual sum of squares by some small percentage only.

In general, these two procedures do not lead to the retention of the same independent variables, nor to exactly the best set of variables of that particular size.[5] In both cases, however, the resulting function simplifies our equation considerably and is useful for our purposes.

initial system state

Thus far, we have defined some regression procedures for using the process data in estimating its relationship to the parameter values. Now, attention is directed toward using these parameters in the system state control procedures. Thus far, for expositional clarity, we have neglected the dependence of x at time t on the value of x at the initial time. However, the system state, x, changes as the system develops in time, and $x(t + 1)$ depends upon $x(t)$, because each period's operation is not completely independent of the previous periods. We assume a linear model for the system under the assumption that a linear approximation of a non-linear system holds for some short period of time. Also, the procedure calls for a periodic revision of the linear model based on new data. Thus, it seems reasonable that the linear approximation will be useful. If we think of x as representing a vector of system states, we can look at the dynamics of the system as

$$\mathbf{x}(t + 1) = A\mathbf{x}(t) + B\mathbf{u}(t),$$

where $\mathbf{x}$ is an $n \times 1$ vector, A an $n \times n$ matrix, $\mathbf{u}$ an $m \times 1$ vector, and B an $n \times m$ matrix. In their full complexity, A and B may be considered as functions of t.

For a desired state of the system, say $\mathbf{r}(t)$, the quantity $\mathbf{x}(t) - \mathbf{r}(t)$ measures the deviation of our system from the desired state. Letting

$$\mathbf{z}(t) = \mathbf{x}(t) - \mathbf{r}(t),$$

we can express a performance measure for a period $(t, t + 1)$ as

$$\mathbf{z}(t + 1)' Q\mathbf{z}(t + 1) + \mathbf{u}(t)' G\mathbf{u}(t), \tag{2}$$

where Q and G are positive semidefinite matrices, and the transpose of a vector is indicated by prime. Q expresses the penalty, or "cost," of deviation from the desired state. For products waiting in queues, for example, Q might express the inventory cost for the waiting products. G serves to keep $\mathbf{u}$ within or near the region of available process data. In order to limit the $\mathbf{u}$-region, we may even introduce the further restriction that $|u_i| \leq m_i$.

performance index

With a performance measure that depends upon the system state, and with system dynamics that depend upon the input $\mathbf{u}$, we now consider a *performance index*, P, given by

$$P = \sum_{t=0}^{T-1} \mathbf{z}(t + 1)' Q\mathbf{z}(t + 1) + \mathbf{u}(t)' G\mathbf{u}(t)$$

This index takes into consideration the deviations from the desired system state over the time period of interest. If we consider the measure for only a single time period and minimize it by a choice of $\mathbf{u}(t)$, it would be optimal for only one period. However, we must consider the effect of $\mathbf{u}(t)$ on $\mathbf{x}(t + 1)$, which in turn affects $\mathbf{x}(t + 2)$, etc.

Minimization over the time period of interest, one period at a time, is not equivalent to minimization over T periods. Nevertheless, this one-period procedure has advantages. We can more easily consider a complex identification of the system dynamics, almost as complex as we desire, and therefore have a better chance to achieve an acceptable system representation and still have a tractable minimization problem. Conceptually, of course, optimization over the entire time horizon is more desirable. Various techniques have been applied to such problems, such as the calculus of variations[6,7,8] and its extension, Pontryagin's maximum principle,[9,10] and dynamic programming.[11,12,13,14]

Since we are dealing with the effect of a scheduling priority formula on the entire system of which the formula is only a small part, it seems justifiable and reasonable to use as simple a model as possible. Therefore, in our attempt to improve the system performance by refining the scheduling priority formula, we use a scalar system state and consider its dependence on only one input parameter at a time. The parameter can be selected from the regression by choosing the one that has the smallest residual sum of squares. Other control variables and complexities can be introduced if necessary.

Let us now examine a system state, such as the dollar value of lateness in the system. Taking into account all the jobs in the system, we multiply the number of days each job is late by the dollar value of the job and then take the total of such dollar days.

Our assumed linear functional dependence can now be expressed as

$$x(t+1) - x(t) = b_1 u_1(t) + \cdots + b_5 u_5(t) \tag{3}$$

and can be approximated by the differential equation

$$\frac{dx}{dt} = b_1 u_1(t) + \cdots + b_5 u_5(t). \tag{4}$$

By integrating Equation 4 from 0 to T,

$$x(T) - x(0) = b_1 \int_0^T u_1(t)\, dt + \cdots + b_5 \int_0^T u_5(t)\, dt,$$

and comparing with the sum of Equation 3,

$$x(T) - x(0) = b_1 \sum_{t=0}^{T-1} u_1(t) + \cdots + b_5 \sum_{t=1}^{T-1} u_5(t),$$

we can see that the right-hand side contains approximating sums for the integrals above.

The functional dependence of Equation 3 provides a system of equations for a least-squares regression for b_i, in which $i = 1, 2, 3, 4, 5$. Based on data for values of t, we can perform the regression on each variable u individually and then select the one that best explains $x(t+1) - x(t)$, i.e., the one with the smallest residual sum of squares. This selected variable is used for the optimization. Again, this is done for simplicity. In a more complex development, the step-wise regression could be performed on all u_i, u_i^2, and cross products, selecting those variables that account for a desired large percentage of the sum of squares. This procedure would lead to a variational problem in more variables.

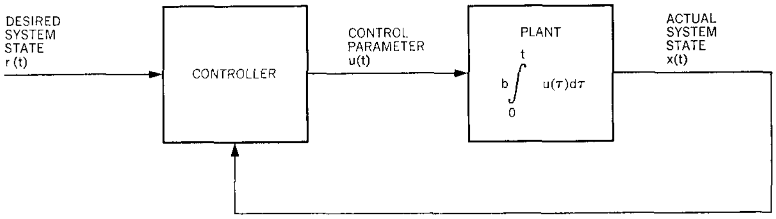
In estimating the value of b_i by means of a least-squares fit, not all of the available past data should be used. Since we are developing a linear model for a non-linear system, our approximation should not be expected to hold for all system states at any time. Various conditions within and without the system can affect the system dynamics. In many manufacturing plants, however, it can be assumed that conditions change slowly and radical changes are rare. Thus, we can introduce an adaptive element to our linear view of the system dynamics by using current data that incorporate recent system changes. A time period can be selected that is useful for making a linear approximation of the system, say d days. The approximation and choice of b_i should be made daily, using data of the last d days only.

control

The control concepts of the system are shown in Figure 1, where the plant is our scheduling formula which converts a parameter value u to a specific system state x (e.g., the lateness of work). This system could also be viewed in other ways, i.e., a model can have many applications, physical or conceptual, as long as the equations describe the system's "motion."

The state of the system at time t, as given in Equation 4,

Figure 1 Control concepts



can now be approximated as

$$\frac{dx}{dt} = \dot{x}(t) = bu(t). \tag{5}$$

For effective control, this system state should be as close as possible to the desired state, $r(t)$. In our case, $r(t)$ is equal to zero, because we want no lateness in the system. Having control over $u(t)$, we use this variable to achieve our goal. However, as mentioned before, we must also consider the cost of attaining this state. This cost is defined as $[\gamma u(t)]^2$, where γ is some penalty constant we can choose. We assign a cost for not being in the desired state, and a cost for executing control to achieve the desired state. The cost-of-control term not only ensures that the inputs do not become infinite, but it also keeps the calculated control input near the region of available data for estimating b.

Our performance function of Equation 2 can now be expressed as

$$[x(t) - r(t)]^2 + \gamma^2 u^2(t).$$

If our system is scheduled to run for a time interval (t, T), we minimize the integral

$$P = \int_t^T \left\{[x(\tau) - r(\tau)]^2 + \gamma^2 u^2(\tau)\right\} d\tau \tag{6}$$

by a choice of $u(\tau)$.

In selecting u, we do not prescribe a function $u(\tau)$ for the entire interval (t, T). Using the data feedback in our system, we update the computation of the current $u(t)$ on a daily basis. Therefore, we solve for $u(t)$ as a function of $x(t)$, the current system state. This closed-loop control is desirable for several reasons. Even if we knew the exact equations of our system, our decision should be based on the most up-to-date information on our system state, $x(t)$, because the system operation is always affected by disturbances. The ever-changing system dynamics require constant revision of their approximation by a re-estimation of b.

Using Equation 5, we substitute for $u(\tau)$ in Equation 6 and have

$$P = \int_t^T \left\{[x(\tau) - r(\tau)]^2 + \left(\frac{\gamma}{b}\right)^2 \dot{x}(\tau)^2\right\} d\tau. \tag{7}$$

Minimization must be performed by choice of the function $x(\tau)$.

After finding the minimizing $x^0(\tau)$ in terms of the initial state, $x(t)$, we compute the minimizing $u^0(\tau)$ by use of Equation 5. By setting $\tau = t$ in $u^0(\tau)$, we finally receive $u^0(t)$, the control to be exerted currently. The mathematics of this procedure are now indicated; derivations are available in the literature.[6]

Since $r(t) = 0$, the Euler equation for Equation 7, calling the integrand $F(x, \dot{x})$, is

$$F_x - \frac{d}{d\tau} F_{\dot{x}} = u$$

or

$$x(\tau) - \left(\frac{\gamma}{b}\right)^2 \ddot{x}(\tau) = 0.$$

After setting the appropriate boundary conditions, we find $x(\tau)$ in terms of the initial condition (current state), $x(t)$. Then, taking

$$\frac{1}{b} \frac{dx(\tau)}{d\tau} = u(\tau),$$

$u(t)$ gives the desired control. In this case, the result is

$$u^0(t) = -\frac{1}{\gamma} \left(\tanh \left[\frac{b}{\gamma} (T - t)\right]\right) x(t), \tag{8}$$

which is fairly easy to compute.

summary

The method described makes it possible to assign weights to important priority factors by the use of control variables. A function that in part measures the deviation from a desired system state is optimized by the choice of values of these control variables. Under the assumption that Equation 5 represents the dynamics of the system, Equation 8 provides the system with an input that minimizes the performance index P of Equation 6. This input, u, can then be used in Equation 1, yielding the desired sequencing function. The control exercised is adaptive in the sense that the representation of the system's dynamics is periodically revised by successive least-squares fits of new data as acquired.

CITED REFERENCES AND FOOTNOTE

1. M. E. Efroymson, "Multiple Regression Analysis," *Mathematical Methods for Digital Computers*, ed. by A. Ralston and H. S. Wilf, John Wiley & Sons, New York, Chapter 7 (1960).
2. J. Mandel, *The Statistical Analysis of Experimental Data*, Interscience Publishers, Division of John Wiley & Sons, New York (1964).
3. R. J. Freund, "A warning of roundoff errors in regression," *The American Statistician* **17**, No. 5, 13–15 (1963).
4. K. W. Smillie, "Remarks on a recent paper on roundoff error in regression analysis," *The American Statistician* **18**, No. 4, 26 (1964).
5. J. Oosterhoff, "On the selection of independent variables in a regression equation," Mathematisch Centrum, 2e Boerhaavestraat 49, Amsterdam, The Netherlands, Report S 319 (VP 23) (December 1963).
6. C. W. Merrian III, *Optimization Theory and the Design of Feedback Control Systems*, McGraw-Hill Book Company, New York, Chapter 4 (1964).

7. W. Kipiniak, *Dynamic Optimization and Control*, The M.I.T. Press, Massachusetts Institute of Technology and John Wiley & Sons, Inc., New York (1961).
8. I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Translated and edited by R. A. Silverman, Prentice-Hall, Englewood Cliffs, New Jersey (1963).
9. L. S. Pontriagin, B. H. Boltyanski, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, translated by K. N. Trirogoff, and edited by L. W. Neustadt, Interscience Publishers, Division of John Wiley & Sons, New York (1962).
10. C. A. Desoer, "Pontriagin's maximum principle and the principle of optimality," *Journal of the Franklin Institute* **271**, 361–367 (1961).
11. R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, New Jersey (1957).
12. R. Bellman and S. Dreyfus, *Applied Dynamic Programming*, Princeton University Press, Princeton, New Jersey (1962).
13. G. W. Deley, *Optimal Bounded Control of Linear Sampled-Data Systems Using Quadratic Performance Criteria*, SUDAER No. 148, Dept. of Aeronautics and Astronautics, Stanford University, Stanford, California (April 1963).
14. R. E. Kalman and R. W. Koepcke, "Optimal synthesis of linear sampling control systems using generalized performance indexes," *ASME Transactions* **80**, 1820–26 (1958).