MEMOIRS OF THE SCHOOL OF ENGINEERING, OKAYAMA UNIVERSITY, Vol. 3, No. 1, SEPTEMBER, 1968

# Optimal Design of Extraction Processes by Gradient Methods*

Hayatoshi SAYAMA and Ko OI

*Department of Industrial Science*

(Received March 15, 1968)

This paper deals with the optimization problem of extraction processes with recycle and immiscible solvent. Gradient methods using control variable iteration require a trial and error procedure to solve process and overall equations for this problem. A new formulation is presented to eliminate such a trial and error procedure. First order and second-order convergence gradient methods were applied to determine the optimum solvent distribution in cross-current extraction with recycle and immiscible solvent. The Fletcher-Powell method converged most rapidly in all cases.

## § 1. Introduction

Gradient methods have been used to obtain numerical solution of multistage optimization problems in engineering processes. Gradient methods generally require to select nominal control variables and determine state variables from process equations. Control variables are improved to decrease or increase performance index as rapidly as possible. This procedure is repeated until the change of performance index is within the tolerance limit. However, if state variables at each stage are not expressed explicitly, process equations must be solved by a trial and error procedure. For the optimization problem with recycle or feedback, overall balance equations should be also solved by the trial and error procedure and most of computation time is spent in this procedure to obtain the solution of process and overall equations. In this article, a new formulation was developed for the problem of optimum solvent distribution in the cross-current extraction with recycle. This formulation requires no trial and error procedure and can be applied to the one-dimensional process for which control variable is determined only by state variables.

Some numerical examples were presented by first-order and second-order convergence gradient methods which have been developed recently. Second-order convergence gradient methods require only first derivatives and have much better convergence characteristics than the first-order method. The Fletcher-Powell method converged most rapidly in all cases but the Fletcher-Reeves method has the same or a little better convergence properties than first-order method.

* Paper presented at the 33rd Annual Meeting of the Society of Chemical Engineers, Japan.

## § 2. Formulation of extraction processes with recycle and immiscible solvent

As a typical example of the multistage mass transfer operation, the extraction processes were studied by many researchers.[1)~8)] The process to be optimized is shown in Fig. 1. It

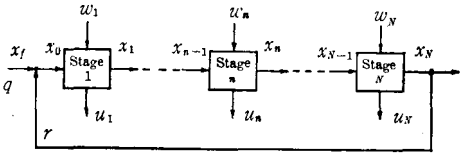


Fig. 1. Cross-current extraction process

consists of $N$ equilibrium extraction stages and a portion of the end products is recycled to the first stage at a flow rate $r$. A mixture containing $x_n$ solute is extracted by a solvent (extract) $w_n$. The extract leaves the stage with a solute concentration $u_n$. The solvent (raffinate) flow rate is $q+r$ and will be assumed to be constant. A material balance around the $n$th stage yields

$$(q+r)x_n=(q+r)_{n-1}-w_n u_n, \qquad n=1, \cdots, N. \tag{1}$$

The equilibrium relationship between raffinate

and extract leaving the $n$th stage is given by

$$u_n = \phi(x_n). \tag{2}$$

The mixing of the feed and recycle stream is expressed by

$$(q+r)x_0 = qx_f + rx_N \tag{3}$$

We maximize the following performance index subject to eqs. (1)~(3),

$$P = q(x_f - x_N) - \lambda \sum_{i=1}^{N} w_n, \tag{4}$$

where $\lambda$ is the cost of solvent and has the the meaning of Lagrange multiplier. The problem is to select a sequence of $w_n$ so as to maximize performance index $P$. Inserting eqs. (1)~(3) into (4) and rearranging the result gives

$$P = (q+r)\sum_{i=1}^{N}(x_{n-1}-x_n)\left[1-\frac{\lambda}{\phi(x_n)}\right]. \tag{5}$$

From eq. (5) it is clear that the performance index is evaluated by choosing $x_1, \cdots, x_N$ and control variables $w_1, \cdots, w_N$ are found from eqs. (1) and (2). Therefore we consider state variables as independent variables and try to maximize $P$ with respect to $x_1, \cdots, x_N$. Note that the performance index should be maximized with respect to $w_1, \cdots, w_N$ in the usual manner. Differentiating eq. (5) with respect to $x_n$ gives

$$\frac{\partial P}{\partial x_n} = (q+r)\left\{(-1)\left[1-\frac{\lambda}{\phi(x_n)}\right] + (x_{n-1}-x_n)\lambda\frac{d\phi(x_n)/dx_n}{[\phi(x_n)]^2} + \left[1-\frac{\lambda}{\phi(x_{n+1})}\right]\right\} = 0, \quad n=1, \cdots, N-1, \tag{6}$$

$$\frac{\partial P}{\partial x_N} = (q+r)\left\{\left(\frac{r}{q+r}\right)\left[1-\frac{\lambda}{\phi(x_1)}\right] + (-1)\left[1-\frac{\lambda}{\phi(x_N)}\right] + (x_{N-1}-x_N)\lambda\frac{d\phi(x_N)/dx_N}{[\phi(x_N)]^2}\right\} = 0, \tag{7}$$

where $(\partial x_0/\partial x_N) = r/(q+r)$ is substituted into eq. (7), which is obtained directly from eq. (3). Solving for $x_n$ gives

$$x_{n-1} = x_n + \frac{\phi(x_n)}{d\phi(x_n)/dx_n}\left[\frac{\phi(x_n)}{\phi(x_{n+1})}-1\right], \quad n=1, \cdots, N-1, \tag{8}$$

$$x_{N-1} = x_N + \frac{[\phi(x_N)]^2}{d\phi(x_N)/dx_N}\left\{\left[1-\frac{\lambda}{\phi(x_N)}\right] - \frac{r}{q+r}\left[1-\frac{\lambda}{\phi(x_1)}\right]\right\}. \tag{9}$$

Eqs. (8) and (9) are the same type as eqs. (38) and (39) of Fan and Wang[5], which were derived from the Discrete Maximum Principle. It follows that this problem may be optimized by differentiating the performance index with respect to state variables and the trial and error procedure will not be required for the computation of overall and process equations. Then gradient methods may be applied to this problem starting from the sequence $x_1 > \cdots > x_n > \cdots > x_N$ and the following relation is repeated until the optimum solution is found.

$$x_n^{i+1} = x_n^i + \theta(\partial P/\partial x_n^i), \quad n=1, \cdots, N, \tag{10}$$

where $\theta$ is scale factor.

Next we treat more general case than considered above for the one dimensional process. We assume that there is only one control variable $w_n$ determined from the process equation.

$$w_n = T_n(x_{n-1}, x_n), \quad n=1, \cdots, N, \tag{11}$$

where $T$ is the transformation operator.
The performance index is expressed only by state variables

$$P = P(x_1, \cdots, x_N). \tag{12}$$

If small perturbations are made with $w_n$ and $x_n$, then expansion of eqs. (11) and (12) yields

$$\delta w_n = \left(\frac{\partial T_n}{\partial x_{n-1}}\right)\delta x_{n-1} + \left(\frac{\partial T_n}{\partial x_n}\right)\delta x_n, \quad n=1, \cdots, N, \tag{13}$$

$$\delta P = \sum_{n=1}^{N}\left(\frac{\partial P}{\partial x_n}\right)\delta x_n. \tag{14}$$

$(\partial T_N/\partial x_N)$ is assumed not to vanish. Solving eq. (13) for $\delta x_N$ and inserting $\delta x_N$ into eq. (14) gives

$$\delta P = \sum_{n=1}^{N-1}\left(\frac{\partial P}{\partial x_n}\right)\delta x_n - \left(\frac{\partial P}{\partial x_n}\right)\frac{(\partial T_N/\partial x_{N-1})}{(\partial T_N/\partial x_N)}\delta x_{N-1} + \left(\frac{\partial P}{\partial x_N}\right)\frac{1}{(\partial T_N/\partial x_N)}\delta w_N. \tag{15}$$

$\delta w_N$ is involved in the last term of eq. (15) and $(\partial P/\partial x_n) = 0$ is the necessary condition to maximize $P$ with respect to $w_N$. By repeating the same procedure, the condition to maximize $P$ with respect to $w_n$ is

$$(\partial P/\partial x_n) = 0, \quad n = 1, \cdots, N. \tag{16}$$

This condition shows that $P$ may be optimized with respect to $x_n (n = 1, \cdots, N)$ and the foregoing statement was verified to be valid to the general case. However, eq. (16) can not be applied to the case where transformation equation has the form of $x_n = T_n(x_{n-1}, w_n)$ or $T_n(x_n, x_{n-1}, w_n) = 0$ and has two or more state and control variables. Extraction processes with miscible solvent are treated with the same way as this immiscible case.

## § 3. Solution Techniques

First-order gradient methods have been used to solve many optimization problems in engineering systems but converge very slowly as the optimum solution is approached. Second-order gradient methods converge very rapidly but have the disadvantage that convergence is not assured and complicated computation is required. Recently second-order convergence gradient methods have been developed. These methods require only first derivatives of the performance index and have second-order convergence.

The Fletcher-Powell Method : [9][10]

From eq. (10) the computational procedure for the first order gradient methods is given by

$$x^{i+1} = x^i + \theta_{\max} H \nabla P(x^i), \tag{17}$$

where $H$ is a $N$ by $N$ symmetric positive definite matrix and $\theta_{\max}$ is chosen to maximize $P[x^i + \theta H \nabla P(x^i)]$ with respect to $\theta$. First and second-order gradient methods utilize $H = I$ and $H = -[\partial^2 P(x)/\partial x_j \partial x_k]^{-1}$, respectively.

The Fletcher-Powell method updates this matrix $H$ at each iteration and minimizes a positive definite quadratic form of $n$ variables in $n$ iterations.

The iteration procedure is described as follows:

1) Estimate $x^0$ and any positive definite matrix $H^0 (= I$ for usual case).
2) Choose $\theta = \theta_{\max}$ by maximizing $P[x^i + \theta H^i \nabla P(x_i)]$.
3) Determine the improved point

$$x^{i+1} = x^i + \theta_{\max} H^i \nabla P(x^i). \tag{18}$$

4) Modify $H^i$ by the following relation

$$H^{i+1} = H^i + A^i + B^i, \tag{19}$$

here

$$A^i = -\frac{(x^{i+1} - x^i)(x^{i+1} - x^i)'}{(x^{i+1} - x^i)'\gamma^i},$$

$$B^i = -\frac{(H^i\gamma^i)(H^i\gamma^i)'}{\gamma^{i\prime} H^i \gamma^i}, \tag{20}$$

$$\gamma^i = [\nabla P(x^{i+1}) - \nabla P(x^i)],$$

where $(\ )'$ indicates transposition of matrix.

5) Repeat above procedure 2) through 4) until the optimum solution is obtained.

The Fletcher-Reeves Method : [11][12]

The conjugate gradient method of Hestenes and Stiefel is an $n$-step procedure for solving a set of simultaneous linear equations having a symmetric positive definite matrix and minimizes a positive definite quadratic function of $n$ variables in $n$ iterations. Fletcher and Reeves developed this method to the nonlinear optimization problems.

Computation procedure is described as follows:

1) Estimate $x^0$ and $p^0 = \nabla P(x^0)$.
2) Choose $\theta = \theta_{\max}$ by maxmizing $P[x^i + p\theta^i]$.
3) Determine the improved point

$$x^{i+1} = x^i + \theta_{\max} p^i. \tag{21}$$

4) Modify $p^i$ by following relation

$$p^{i+1} = \nabla P(x^{i+1}) + [|\nabla P(x^{i+1})|^2 / |\nabla P(x^i)|^2] p^i. \tag{22}$$

5) Repeat above procedure 2) through 4) until the optimum solution is obtained.

## § 4. Numerical Example and Discussions

The extraction processes illustrated in § 2 were optimized by various gradient methods. The numerical values used are

$x_f = 0.2$, $q = 0.05$, $r = 0.2$

The equilibrium relatioship is expressed by

$$u_n = \phi(x_n) = a + bx_n + cx_n^2 + dx_n^3 + ex_n^4 + fx_n^5, \tag{23}$$

where

| | |
|---|---|
| $a = 0.0009$ | $d = 633.84$ |
| $b = 1.7971$ | $e = 3371.3$ |
| $c = 35.196$ | $f = -5916.0$ |

This set of numerical values is the same as that used by Lee[6] for the purpose of comparison. But Lee treated the optimization problem only for $r = 0$.

Gradient methods require maximization of the one-dimensional function $P(x^i + \theta^i s^i)$ at each iteration ($s^i = \nabla P(x^i)$, $p^i$, $H^i \nabla P(x^i)$). The selection of $\theta$ currently used is an approximate method analogous to a binary scale factor

search. Starting with an arbitrary $\theta^0$, the procedure requires determination of an integer $m$ which maximizes $P[x^i + 2^m \theta^0 s^i]$. $x^{i+1}$ is chosen to yield larger value of $P$ at the point $(x^i + 2^m \theta^0 s^i)$ or $(x + (3/2)2^m \theta^0 s^i)$. $\theta_{max}$, which determines $x^{i+1}$, is used as $\theta^0$ in the next one-dimensional search. To select more precise $\theta_{max}$, a parabolic polynomial is fitted to the three values of $P$ and the unique maximum is determined from this parabolic approximation. Comparing this value with that obtained in one-dimensional search gives larger value of $P$ and $\theta_{max}$ is adopted from better one of these.

For the case $N=10$, the optimum solvent distribution is shown in Fig. 2. Fig. 3 shows the nominal allocation of $x_n$ as starting points. The optimum allocation for $r=0$ obtained by gradient methods is presented in Fig. 4. It is clear from this result that there are three local maxima $P_1=0.114282$, $P_2=0.114278$ and $P_3=0.114037$, which depend on starting points. Figs. 5, 6 and 7 compare the convergence of the first-order, Fletcher-Reeves (F-R) and Fletcher-Powell (F-P) method. F-P method converged most rapidly in all cases and required about 10 iterations to approach the optimum point from the initial points taken here. F-R method has a better convergence than the first-order method but in some case there is no difference in the convergence of both methods. It results

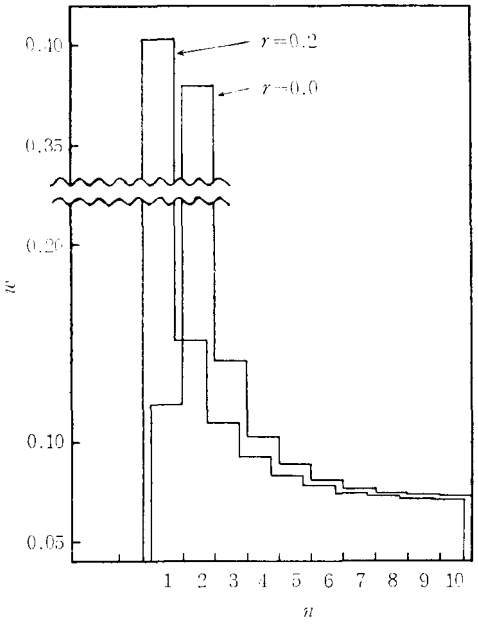


Fig. 2. Optimum solvent distribution with and without recycle

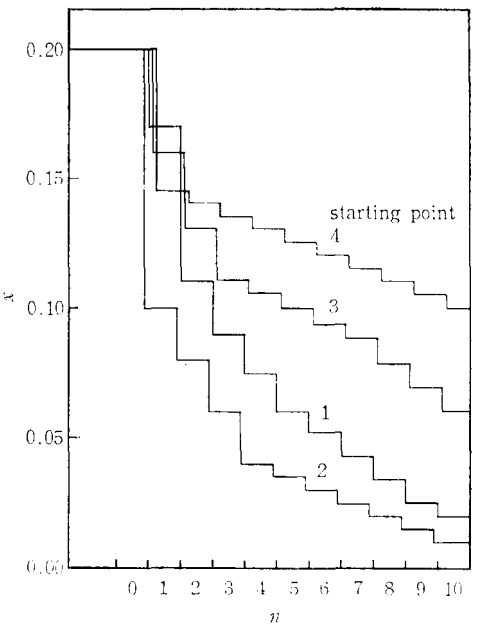


Fig. 3. Nominal allocation of $x$ as starting points for $r=0$

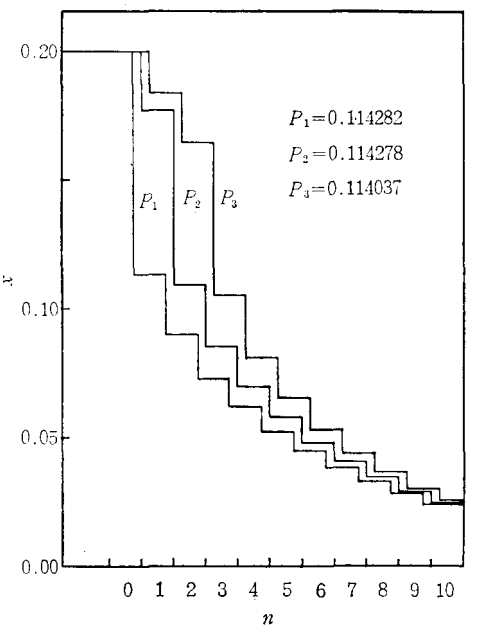


Fig. 4. Optimum allocation of $x$ for $r=0$

from these computations that F-R method is superior to the first-order method when the starting point is near the optimum but not so when the starting point is far from the optimum. The parabolic approximation worked well for F-P and F-R methods but resulted in slow convergence for the first-order method.

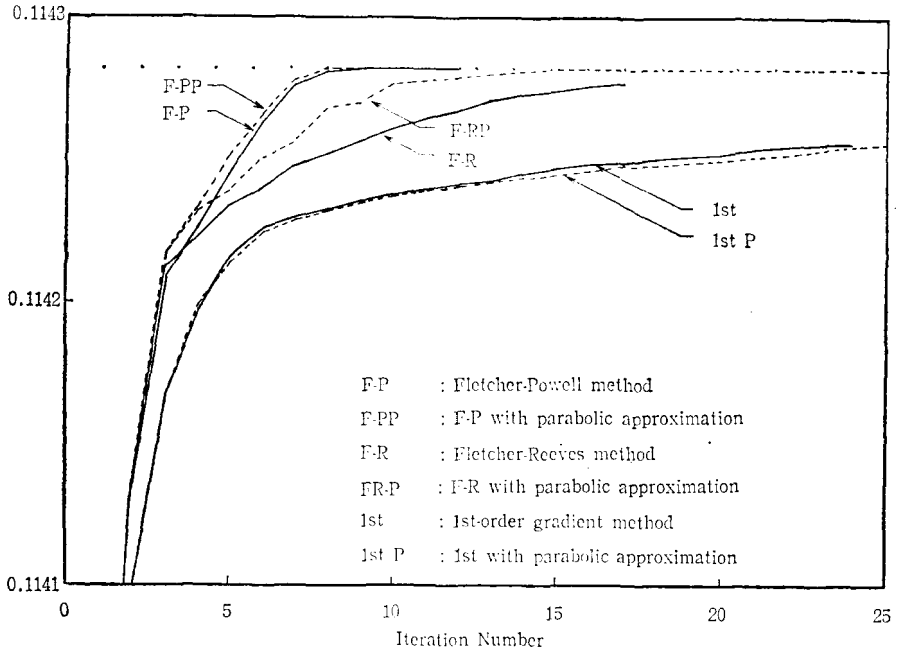


Fig. 5. Convergence of performance index starting from point 1 for $r=0$

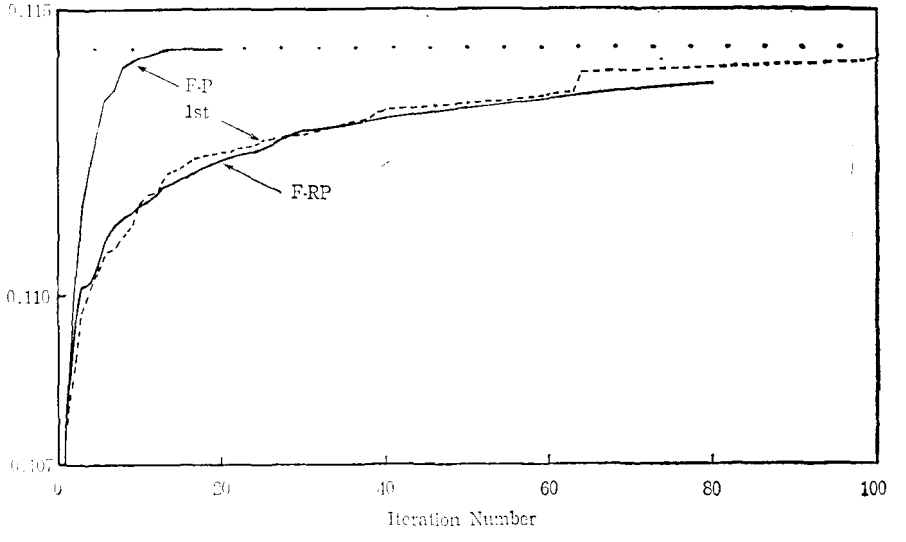


Fig. 6. Convergence of performance index starting from point 3 for $r=0$

Fig. 8 shows the optimum and nominal allocation of $x_n$ for the problem with recycle. The optimum solvent distribution is shown in Fig. 2 for the purpose of comparison. There are no computational difficulties for the problem with recycle. Fig. 9 shows the convergence of the performance index by various methods. As in previous problem having no recycle, F-P method converged most rapidly and required about 10 iterations, while F-R method required 20 iterations. When initial points were taken as 3 and 4 in Fig. 8 by F-P method, convergence could not be obtained. It results from the fact that $H^i$ is non-positive definite matrix when the scale factor $\theta$ is so large. At that point with a non-positive definite matrix, $H^i$ is replaced by $I$ and then the optimum solution was obtained. This is a very

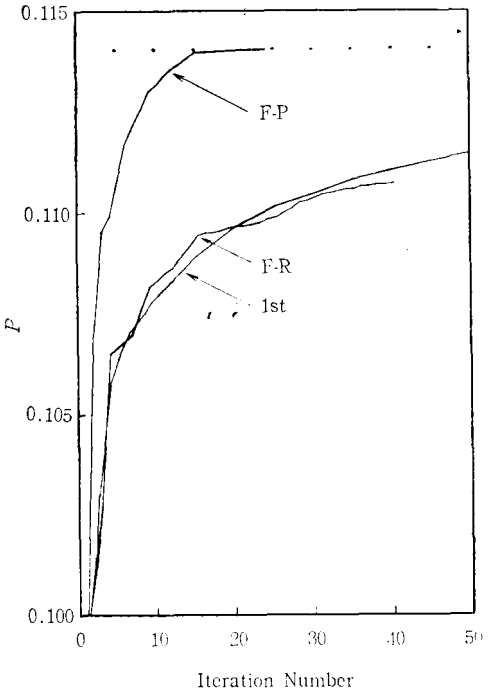


Fig. 7. Convergence of performance index starting from point 4 for $r = 0$

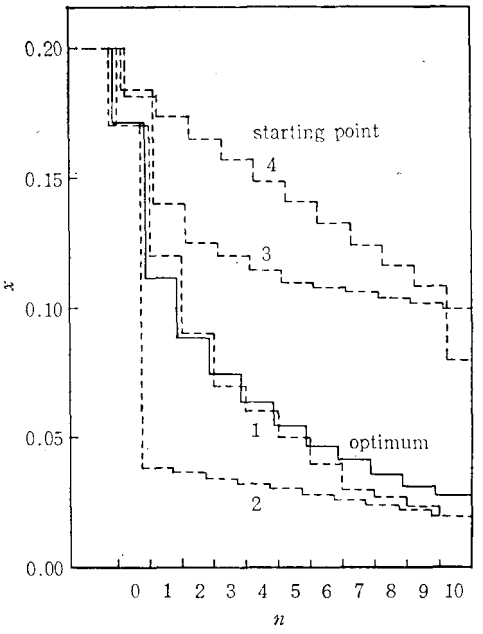


Fig. 8. Optimum and nominal allocation of $x$ for $r = 0.2$

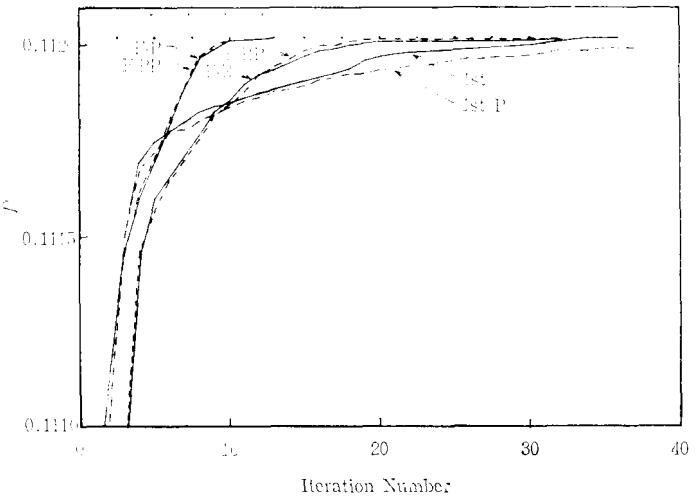


Fig. 9. Convergence of performance index starting from 1 for $r=0.2$

important problem to be overcome when F-P method is applied to the non-quadratic functions. The computation time in F-P method is longer than the first-order method by 1.5 times but F-R method requires slightly longer time than the first-order method.

## § 5. Conclusion

The formulation presented here requires no trial and error procedure to solve process and overall balance equations and can reduce the computational time in the one-dimesional optimization problem. Various gradient methods were applied to obtain optimum solvent distribution in cross-current extraction with and without recycle. The Fletcher-Powell method is extremely superior to the Fletcher-Reeves and first-order gradient methods and required about 10 iterations for the problem of 10 variables from the initial point taken above. The Fletcher-Reeves method has not so effective convergence as expected in this problem and will have the same convergence as the first-order method when the initial point is far from the optimum. From the results obtained the Fletcher-Powell method seems to be most suitable to the multi-dimensional and nonlinear optimization problems in engineering system.

## References

1) R. ARIS, D.F. RUDD and N.R. AMUNDSON:

Chem. Eng. Sci., **12** (1960) 88.
2) D. F. RUDD and E. D. BLUM: ibid., **17** (1962) 277.
3) R. JACKSON: ibid., **18** (1962) 215.
4) E. M. ROSEN: ibid., **19** (1964) 999.
5) L. T. FAN and E. D. WANG: *The Discrete Maximum Principle*, John-Wiley (1964).
6) E. S. LEE: Ind. Eng. Chem. Fundamentals, **3** (1964) 373.
7) L. LAPIDUS and R. LUUS: A. I. Ch. E. Journal, **13** (1967) 101.
8) C. BROSILOW and L. LASDON: Preprints of A. I. Ch. E-I. Chem. E. Joint Meeting, London, Section **4** (1965) 67.
9) W. C. DAVIDON: ANL-5990 (1959).
10) R. FLETCHER and M. J. D. POWELL: Computer J., **6** (1963) 163.
11) M. R. HESTENES and E. STIEFEL: J. Res. N. B. S., **49** (1952) 409.
12) R. FLETCHER and C. M. REEVES: Computer J., **7** (1964) 107.