海洋安全學會誌 第 3 卷 第 2 號, pp. 93~103, 1997. 7.

論 文

해양환경에 관련된 실험을 위한 최적실험계획

김 재 환*

Optimal Designs for the Experiments related with Marine Environment

Jae-Hwan Kim

<목 차>

Abstracrt

This paper develops a new heuristic, the Excursion Algorithm(EA), for constructing optimal designs for the experiments related with marine environment. The proposed EA consists of three parts: 1) construction of an initial feasible solution, 2) excursions over a bounded region, and 3) stopping rules. It is the second part that distinguishes the EA from the other existing heuristic methods. It turns out that excursions over a bounded feasible and/or infeasible region is effective in alleviating the risks of being trapped at a local optimum.

Since this problem is formulated for the first time in this thesis, other heuristic algorithms do not exist. Therefore, global optimal solutions are obtained by complete enumeration for some cases, and the performance of the EA is evaluated in terms of solution quality. Computational results show that the proposed EA is effective in finding good(or, in many cases, global) solutions to the constrained optimal experimental design problems.

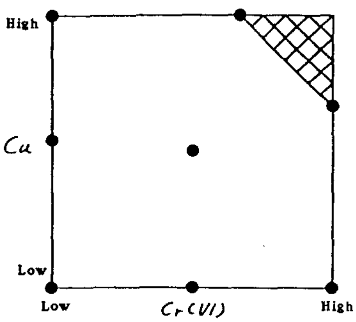
* 한국해양대학교 이공대학 응용수학과

1. 서 론

1.1 연구배경 및 목적

폐수처리등의 해양환경에 관련된 실험을 실시해야할 경우 실험시간, 실험비용, 실험난이도 등의 실험제한요건에 의해 충분한 반복실험을 행할수 없는 경우가 종종 발생한다.

예를들어, 폐수처리를 위해 중독성 금속물질인 Cr(VI), Cu, Ni등을 제거하려 할 때, 미생물의 활성도는 이를 위한 주요한 요인으로 작용한다. 미생물의 활성도를 측정하기 위해 Cr(VI)와 Cu의 2수준(Low, High)에 대해 실험을 해야할 경우 <그림 1>에서 처럼 실험 불가능한 영역이 존재하거나 위에서 언급한 제한된 실험요건에 대해 충분한 반복실험을 할 수 없는 경우에
는 요인 실험(factorial design)이나 일부 요인실험(fractional factorial design) 등의 전통적인 실험계획을 사용할 수가 없다. 따라서, 이런 경우에는 최적 실험계획(optimal design)방법에 의해 실험을 계획하는 것이 하나의 유용한 대안이 될 수 있다.



<그림 1> 실험 영역

해양환경분야에서 측정하고자하는 미생물의 활성도등의 반응치를 y_i라 하면, 최적실험계획의 모형은 다음과 같다.

$$y_i = \beta_1 f_1(x_i) + \beta_2 f_2(x_i) + \cdots + \beta_k f_k(x_i) + e_i \qquad i = 1, 2, \cdots, n \quad \cdots(1)$$

여기서, n은 총 실험개수이고 e_i는 실험오차이며 $E(e_i)=0$, $Var(e_i)=\sigma^2$, $Cov(e_u, e_v)=0$, $u \neq v$, $u \neq v$를 가정한다. 모형 (1.1)을 벡터/행렬 형태로 나타내면 다음과 같다.

$$y = X\beta + e. \quad \cdots\cdots (1.2)$$

여기서,

$$y' = (y_1, y_2, \cdots, y_n),$$

$$X = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_k(x_n) \end{bmatrix}$$

$$\beta' = (\beta_1, \beta_2, \cdots, \beta_k),$$

$$e' = (e_1, e_2, \cdots, e_n).$$

모형 (1.2)에서 미지의 모수 벡터 β의 최소제곱 추정량은

$$\hat{\beta} = (X'X)^{-1}X'y$$

이며, $\hat{\beta}$의 공분산 행렬(covariance matrix)은

$$Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

이다.

최적실험계획 문제는 $Cov(\hat{\beta})$이 최소화되게 실험영역에서 n개의 실험점을 선택하는 것이다. 본 논문에서는 $Cov(\hat{\beta})$을 최소화하기 위한 여러기준들 중에서 $det(X'X)$의 최대화하는 $D-$최적기준을 사용하였다.

$D-$최적 실험에 대한 연구는 그 동안 여러 사람들에 의해 연구되었으며, <표 1.2>에 나타나 있다.

최적실험계획의 개념은 Smith[1]에 의해 처음으로 제안 되었다. 초창기의 연구는 Wald[2], Hotelling[3] 그리고 Mood[4]에 의해서 진행되었으며, Kiefer, Wolfowitz[5]에 의해 그 이론이 정립되었다. 최근에는 Atkinson[6], Pazman[7] 그리고

<표 1.2> D-최적 실험에 대한 기존의 연구

기존의 연구	
Theory	Algorithm
Smith(1918) : 처음으로 제안	Kiefer(1971), Fedorov(1972),
Wald(1943), Hotelling(1949),	Wynn(1970), Mitchell(1974),
Mood(1946) : 초창기 연구	Miller(1970) : exchange algorithm
Kiefer(1960), Wolfowitz(1960)	St.John(1975) : 성능 비교
: 이론 정립	Cook(1980), Nachtsheim(1979)
Atkinson(1982), Pazman(1980),	: algorithm 비교
Ash(1978), Hedayat(1978)	Welch(1982) : branch-and-bound
: 이론 확장	algorithm 개발
Atikinson(1988) : 비교 분석	Linda(1987) : annealing algorithm
	Donev(1988) : adjustment algorithm
	Nguyen(1992) : new exchange algorithm

Ash 와 Hedayat[8]등에 의해서 최적실험계획에 관한 이론의 확장이 이루어졌다. 최적실험계획를 구성하는 알고리즘은 Mitchell[9], Mitchell과 Miller [10], Wynn[11], Kiefer[5], Fedorov[12], St.John과 Draper[13], Galil과 Kiefer[14,15], Johnson과 Nachtsheim[16], Welch([17], [18])등에 의해서 연구 되었다.

1.2 연구범위

본 논문에서는 해양환경에 관련된 실험을 할 경우에 앞에서 언급한 실험시간, 실험비용, 실험의 난이도 등의 실험제한요건에 의해 충분한 반복실험을 하지 못할 경우에 최적인 실험계획을 제공하고자 하며, 폐수처리에서 중요시되는 미생물의 활성도를 측정하기 위한 실험계획을 예제로 제시하였다.

또한, 2장에서는 실험제한 요건중에서 실험비용에 대해 각 실험점마다 동일하지 않은 경우의 최적실험계획의 모형과 그 해법을 제시하였다. 실험비용이 각 실험점마다 동일하지 않은 경우에는 기존의 해법이 존재하지 않으므로, 해법의 근거가 되는 이론을 증명하여 발견적해법(heuristic method)을 개발하였으며, 그 해법의 성능을 3장에서 비교하였다.

2. 최적실험계획

본 논문에서 고안한 발견적 해법을 설명하는 데 필요한 기호들은 다음과 같다.

k : 모형 (1.1)에서 미지의 모수의 갯수.

N : 후보 실험점의 총수.

D : N개의 후보 실험점으로 이루어진 설계 행렬.

d_i : D의 i 번째 행으로서 i 번째 후보 실험점을 나타낸다. n_i : i 번째 후보 실험점의 반복 실험횟수. 결정변수.

n^c : 어떤 iteration까지의 최적설계의 총 실험횟수.

n^* : 최적설계의 총 실험횟수.

L : 초기 실험설계의 실험점의 수.

h : 초기 실험설계를 구성할 때 L을 k보다 크게 하기 위해 k에 더해지는 실험점의 수. $L = k + h$.

$W(L)$: D로부터 랜덤하게 L개의 실험점을 선택하여 구성된 초기 실험설계.

$W_t(n)$: t 번째 iteration 중 n 개의 실험점으로 구성된 설계 행렬.

$W^c(n)$: 어떤 iteration까지의 최적실험설계로서 n개의 실험점으로 이루어진 설계 행렬.

$W^f(n)$: excursion 탐색을 통해 개선된 설계 행렬.

$W^*(n)$: 모든 excursion 탐색과정을 통해 가장 최적인 실험설계로서 n 개의 실험점으로 이루어진 설계 행렬.

$M(W)$: 실험설계행렬 W에 대응하는 $X'X$ 행렬.

B : 주어진 실험 예산.

c_i : i 번째 후보 실험점의 실험 비용.

C_N : 후보 실험점의 비용의 총합.

C_W : $W_t(n)$에 포함된 실험점의 비용의 총합.

F-집합 : excursion 탐색의 한 iteration에서 $det(X'X)$를 개선시키지 못했을 때 그 iteration 동안 발생했던 모든 실험

설계들의 집합.

E-집합 : excursion 탐색 동안 생성된 실험설계 중 F-집합에 포함되어 있지 않은 설계의 집합.

U : 구간 (0,1)의 균등분포로부터의 난수(random number).

PR : 반드시 실험을 해야 하는 실험후보점들의 집합

2.1 모 형

본 논문에서 다루고자 하는 실험 비용을 고려한 D-최적 실험계획 문제는 다음과 같은 수리계획 문제로 나타낼 수 있다.

$$
\begin{aligned}
(P)\quad & Maximize \quad det(X'X) \\
& \sum_{i=1}^{N} c_i n_i \le B \\
& n_k \ge 1 \quad \forall k \in PR \\
& n_i \ge 0 \text{ , 정수} \qquad i = 1,\ 2,\ \cdots,\ N
\end{aligned}
$$

위의 수리계획 문제를 예를 들어 설명하면 다음과 같다. 먼저, 가정한 선형회귀모형을

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$$

라고 하자. 실험후보점의 수 N = 8이고 이들로 이루어진 행렬 D는 다음과 같다고 하자.

$$
D = \begin{array}{c} \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{array}
\begin{array}{c} \begin{array}{ccc} x_1 & x_3 & x_2 \end{array} \\
\begin{bmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \qquad \cdots\cdots (2.1)
$$

각 실험점의 실험 비용 벡터

$$C' = (5,\ 7,\ 8,\ 3,\ 10,\ 9,\ 2,\ 6) \qquad \cdots\cdots (2.2)$$

이고, 전체 실험 예산 B = 20, $PR = \{4, 7\}$ 이라고 하면

$$\{\ d_1,\ d_3,\ d_4,\ d_7\ \}$$

로 이루어진 한 실험계획은 위 수리계획 문제 (P)의 한 가능해(feasible region)가 된다.
왜냐하면,

$$
\begin{aligned}
n_1 = n_3 = n_4 = n_7 = 1 \\
n_2 = n_5 = n_6 = n_8 = 0
\end{aligned}
$$

이므로

$$\sum_{i=1}^{8} c_i n_i = 18 < 20 = B$$

가 되어 (P)의 제약식인 실험예산을 만족하기 때문이다. 이 가능해에 대한 X행렬은 다음과 같다.

$$
X = \begin{array}{c} \begin{array}{cccc} & x_1 & x_2 & x_3 \end{array} \\
\begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \end{bmatrix} \end{array}
$$

이때, $det(X'X)$ = 64 이다.

위의 문제 (P)는 제약식인 실험예산을 만족하는 많은 실험계획 X들의 집합중에서 $det(X'X)$를 최대화 하는 즉, 추정치의 오차를 가장 작게 하는 최적 실험계획 X를 찾는 문제이다.

문제 (P)를 풀기 위한 해법(solution method)에 대해 설명하면, 제약식인 실험예산을 만족하는 실험계획 X들을 모두 고려하여 그중 $det(X'X)$ 값이 가장 큰 최적 실험계획 X를 구하는 enumeration 방법이나 분지 한계법(branch-and-bound method)등의 exact방법을 생각할 수 있으나, 문제의 규모가 커짐에 따라 가능한 실험계획의 X들의 집합이 지수 증가적으로 늘어나기 때문에 (NP-complete), 문제 (P)에 대한 exact방법은 아직까지 개발되지 않았다. 따라서, 본 논문에서는 근사적 최적 실험계획을 구하는 발견적 해법

(heuristic methode)을 개발하여 이를 해결하고자 한다.

2.2 발견적 해법

(1) 배경 이론

본 절에서는 발견적 해법의 근거가 되는 최적 실험계획이 존재하는 집 합을 규명하고자 한다.

먼저, $P(N)$를 다음식으로 정의하고

$$P(N)=\left\{(n_1, n_2, \cdots, n_N) \mid \forall n_i \geq 0, \sum_{i=1}^{N} c_i n_i \leq B'\right\}.$$

여기서 $B'=B-\sum_{i\in PR} c_i$ 이다.

FR 집합을 다음과 같이 표현하면,
$FR=\{X \mid X$ 는 $P(N)$을 만족하는 실험 행렬$\}$. 앞 절의 문제 (P)는 다음과 같은 동일한 문제 (P')로 나타 낼 수 있다.

$$(P') \quad maximize \ det(X'X)$$
$$X \in FR$$

또한, $IBDM(B', C)$ 집합을 다음과 같이 정의하면,

[정의] $IBDM(B', C)$
FR의 실험행렬 $X(m_1, m_2, \cdots, m_N)$이 FR의 임의의 실험행렬 $X(n_1, n_2, \cdots, n_N)$에 대해 $m_i \leq n_i \ (i=1, 2, \cdots, N)$이면 $m_i = n_i \ (i=1,2,\cdots,N)$일때 $X(m_1, m_2, \cdots, m_N)$을 Integer Binding Design Matrix $(IBDM)$이라 하고 모든 $IBDM$들의 집합을 $IBDM(B', C)$로 표시한다. 여기서, $C=(c_1, c_2, \cdots, c_N)$ 이다.

최적 실험계획이 존재하는 영역에 관한 다음과 같은 정리가 성립한다.

[정리] 만약 X^* 가 문제 (P')의 최적실험계획이라면, 그 X^* 는 $IBDM(B', C)$ 집합안에 있다.

위의 정리를 증명하기 위해 다음과 같은 보조정리 1, 2가 필요하다.

(보조정리 1) $A=\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$ 단, $\mathbf{a}_i$ 는 A의 행 벡터이다.

그리고, $A_\sigma=\begin{bmatrix} \mathbf{a}_{\sigma(1)} \\ \mathbf{a}_{\sigma(2)} \\ \vdots \\ \mathbf{a}_{\sigma(m)} \end{bmatrix}$ 여기서, σ은 집합 $\{1,2,\cdots,n\}$의 순열 이라 하면,
그때 $A'A=A_\sigma' A_\sigma$ 이다.

(증명)

$$\begin{aligned}(A'A)_{ij} &= \sum_{k=1}^{m}(A')_{ik}A_{kj} = \sum_{k=1}^{m} A_{ki}A_{kj} \\ &= \sum_{k=1}^{m} A_{\sigma(k)i}A_{\sigma(k)j} = \sum_{k=1}^{m}(A_\sigma')_{ik}(A_\sigma)_{kj} \\ &= (A_\sigma' A_\sigma)_{ij} \ \blacksquare\end{aligned}$$

(보조정리 2) A를 $rank\ (A)=n$인 $m\times n$행렬 이라 하자. 그러면, $A'A$는 positive definite이다.

(증명)
$(A'A)'=A'A$ 이므로, $A'A$는 대칭(symmetric)이다. $rank\ (A)=n$ 이라는 사실은 어떤 0이 아닌 $x\in R^{n\times 1}$ 에 대해 $Ax\neq 0$임을 나타낸다. 따라서, $x'(A'A)x=(Ax)'(Ax) > 0$ 이므로, $A'A$는 Positive definite이다. $\blacksquare$

(정리의 증명)
$X(m_1, m_2, \cdots, m_N) \in FR \setminus IBDM(B', C)$ 라고 가정해 보자. 그러면 모든 i에 대해 $m_i \leq n_i$ 이고 어떤 j에 대해 $m_j < n_j$ 인 $X(n_1, n_2, \cdots n_N) \in FR$ 이 존재한다. 이제 $V=X(m_1, m_2, \cdots, m_N)$, $W=X(n_1, n_2, \cdots, n_N)$ 라 두자. $det(W'W) > det\ (V'V)$임을 보이면 $V=X(m_1, m_2, \cdots, m_N)$은 최

적 실험 계획일수 없게 되고 따라서 증명은 끝난다. (보조정리 1)에 의해 우리는 V에 적당한 행들을 첨가함으로서 행렬 W를 얻을수 있다. *rank k* 인 임의의 $m \times k$ 행렬 A와, 0이 아닌 행 벡터 y에 대하여 $A_y = \begin{bmatrix} A \\ y \end{bmatrix}$라 두면, $det(A_y' A_y) = det(A'A)\{1 + y(A'A)^{-1}y'\}$이다. 이 등식과 (보조정리 2)를 사용하여 $det(W'W) > det(V'V)$를 얻는다. ■

따라서, 위의 정리에 의해 최적실험계획이 존재하는 집합인 $IBDM(B', C)$을 탐색하는 발견적 해법을 개발하면 된다.

(2) 발견적 해법의 구성

앞절의 최적실험계획의 존재영역에 관한 정리에 근거하여 최적실험계획이 $IBDM(B', C)$에 존재하므로, 이 영역을 가급적 많이 지나치면서 탐색하는 발견적 해법인 $EA1$과 $EA2$를 고안 하였다.

본 논문에서 개발한 $EA1$과 $EA2$의 알고리즘 단계들은 다음과 같다.

<u>$EA1$ (단측탐색)의 알고리즘 단계</u>

0. α, h, δ를 결정한다. $t = 0$.
1. 초기실험설계 $W(L)$을 구성. $W_t(n) = W(L)$. $W_t(n)$를 개선시킨다.
2. $n^c = n$. $W^c(n) = W_t(n)$. $E = F = \phi$. $t = t + 1$.
3. $W_t(n)$에 한 실험점을 첨가. $|C_W - B| > \alpha B$이면 단계 7로 간다.
4. $E = E \bigcup W_t(n)$. $W_t(n) \in F$ 이면, $F = F \bigcup E$, $E = \phi$. 단계 3으로 간다. $W_t(n) \notin F$ 이면, $W_t(n)$에서 한 실험점을 제거한다.
5. $C_W > B$이면, 단계 4로 간다. $C_W = B$이면, $W^f(n) = W_t(n)$. 단계 6으로 간다. $C_W < B$이면, $W^f(n) = W_t(n)$. $W^f(n)$을 개선시킨다.
6. $det\{M(W^f(n))\} > det\{M(W^c(n))\}$ 이면, $W_t(n) = W^f(n)$. 단계 2로 간다. $det\{M(W^f(n))\} \le det\{M(W^c(n))\}$ 이면, $W_t(n) = W^c(n)$, $n = n^c$, $F = F \bigcup E$, $E = \phi$, $t = t + 1$. 단계 3으로 간다.
7. 중단. $n^* = n^c$, $W^*(n) = W^c(n)$.

<u>$EA2$의 알고리즘 단계</u>

0. α, h, δ 를 결정한다. $t = 0$.
1. 초기 실험설계 $W(L)$을 구성. $W_t(n) = W(L)$. $W_t(n)$를 개선시킨다.
2. $n^c = n$. $W^c(n) = W_t(n)$. $E = F = \phi$. $t = t + 1$.
3. U의 난수 RN을 구한다. $RN \le 1/2$이면, $IFLAG = 0$. 단계 8로 간다. $RN > 1/2$이면, $IFLAG = 1$. 단계 4로 간다.
4. $W_t(n)$에 한 실험점을 첨가한다. $|C_W - B| > \alpha B$이면 단계 11로 간다.
5. $E = E \bigcup W_t(n)$. $W_t(n) \in F$이면, $F = F \bigcup E$, $E = \phi$. 단계 4로 간다. $W_t(n) \notin F$ 이면, $W_t(n)$에서 한 실험점을 제거한다.
6. $C_W > B$이면, 단계 5로 간다. $C_W = B$이면, $W^f(n) = W_t(n)$. 단계 7로 간다. $C_W < B$이면, $W^f(n) = W_t(n)$. $W^f(n)$을 개선시킨다.
7. $det\{M(W^f(n))\} > det\{M(W^c(n))\}$ 이면, $W_t(n) = W^f(n)$. 단계 2로 간다.

$det\{M(W^f(n))\} \le det\{M(W^c(n))\}$ 이면, $W_t(n) = W^c(n)$, $n = n^c$,

$F = F \cup E$, $E = \phi$, $t = t+1$.

$IFLAG = 0$ 이면, 단계 4로 간다.

8. $W_t(n)$에서 한 실험점을 제거한다.

   $|C_W - B| > \alpha B$이면 단계 11로 간다.

9. $E = E \cup W_t(n)$. $W_t(n) \in F$ 이면,

   $F = F \cup E$, $E = \phi$. 단계 8로 간다.

   $W_t(n) \notin F$ 이면, $W^f(n) = W_t(n)$.

   $W_t(n)$에 한 실험점을 첨가한다.

10. $C_W > B$이면, $W^f(n)$을 개선시키고 단계 7로 간다. $C_W = B$이면, $W^f(n) = W_t(n)$. 단계 7로 간다. $C_W < B$이면, 단계 9로 간다.

11. 중단. $n^* = n^c$, $W^*(n) = W^c(n)$.

2.3 예 제

미생물의 활성도를 측정하기위한 실험계획을 예로들면, 모형(1.1)식에서

y_i : 미생물의 활성도의 반응치

x_1 : 크롬 Cr(VI)의 농도

x_2 : 구리 Cu의 농도

x_3 : 니켈 Ni의 농도

라 하고 후보 실험점 D가 다음과 같을때,

$$D = \begin{array}{c|ccc|} & x_1 & x_3 & x_2 \\ d_1 & -1 & -1 & -1 \\ d_2 & 1 & -1 & -1 \\ d_3 & -1 & 1 & -1 \\ d_4 & 1 & 1 & -1 \\ d_5 & -1 & -1 & 1 \\ d_6 & 1 & -1 & 1 \\ d_7 & -1 & 1 & 1 \\ d_8 & 1 & 1 & 1 \end{array}$$

여기서, 1은 각 x_i의 높은 수준을, -1은 각 x_i의 낮은 수준을 각각 나타낸다.

이때, 입력자료가 $C = (9, 3, 6, 5, 6, 4, 7, 9)$, $B = 32$, $PR=\{7\}$로 주어지면, 본 논문의 발견적해법인 $EA2$에 의해 최적실험계획 $W^*(6) = \{d_1, d_7, d_6, d_4, d_2, d_6\}$가 구해진다. 즉, d_1, d_7, d_4, d_2실험점에서 각각 한번의 실험을 하고, d_6실험점에서 2번의 실험을 실시하는 실험계획이 최적실험계획임을 알려준다.

3. 전산 실험

본 연구의 $EA1$과 $EA2$가 얼마나 global에 가까운 최적실험설계를 구성하는지 알아보기 위하여 후보 실험점의 실험비용이 같을 때와 다를 때의 여러 경우에 대해 실험을 수행하였다. 실험비용이 같은 때의 global 최적실험설계는 인자의 수준이 2개인 요인실험인 경우, Galil과 Kiefer[15]에 의해 알려진 결과를 이용하였다. 실험비용이 다른 경우에는 complete enumeration에 의해 global 최적실험설계를 구하였다.

3.1 후보 실험점의 실험비용이 같은 경우

수준이 2개인 요인실험에서 Galil과 Kiefer[15]에 의해 알려진 global 최적실험설계의 $det(X'X)$ 값은 <표 3.1>과 같다. 단, n은 실험점의 수이고 mod는 n을 4로 나눈 나머지이다.

<표 3.1> Galil과 Kiefer[15]의 최적실험설계

moa	det(X′X)
0	n^k
1	$(n-1)^k(n-1+k)$
2	$(n-2)^{k-2}(n-2+k)^2$: k 짝수 $(n-2)^{k-2}(n-1+k)(n-3+k)$: k 홀수
3	$(n+1)^{k-1}(n-k+1)$: $n > 2k-5$

<표 3.1>의 각각의 경우에 대해 $k = 5, 6, 7$인 경우를 전산실험에 포함시켰다. 인자의 수는 $k-1$ 이고 각 인자는 2수준을 가지므로 전체 후보 실험점의 수는 $N = 2^{k-1}$ 이다. n 은 $(N/2 + mod)$ 로 정하였다(<표 3.2> 참조).

한 최적실험설계와 비교하였다. 비용제약을 만족하는 실험설계를 모두 나열1하는 데는 컴퓨터 수행시간이 너무 많이 소요되기 때문에, N을 8과 12로 제한하였으며, 실험예산 B도 작게 설정하였다.

<표 3.2> 실험비용이 같은 경우의 전산실험 결과

mod	k	n	Global 최적 실험설계의 det (X'X)	E A 1				E A 2			
				A	M	O	T	A	M	O	T
0	5	8	8^5	0.05	0.25	4/5	0.42	0.0	0.0	5/5	0.56
	6	16	16^6	0.08790	0.13483	1/5	2.10	0.0	0.0	5/5	3.68
	7	32	32^7	0.01846	0.04615	3/5	15.20	0.00923	0.04615	4/5	17.42
1	5	9	$8^4(13)$	0.0	0.0	5/5	0.44	0.0	0.0	5/5	0.40
	6	17	$16^5(22)$	0.07209	0.11044	0/5	2.22	0.0	0.0	5/5	3.74
	7	33	$32^6(39)$	0.01007	0.02518	3/5	18.10	0.01007	0.02518	3/5	20.64
2	5	10	$8^3(14)(12)$	0.0	0.0	5/5	0.64	0.0	0.0	5/5	0.52
	6	18	$16^4(22)^2$	0.00372	0.01859	4/5	2.90	0.00165	0.00826	4/5	3.62
	7	34	$32^5(40)(38)$	0.00166	0.00416	3/5	17.80	0.00083	0.00416	4/5	19.56
3	5	11	$12^4(7)$	0.0	0.0	5/5	0.72	0.0	0.0	5/5	0.78
	6	19	$20^5(14)$	0.02761	0.03451	1/5	3.36	0.01029	0.01714	2/5	4.68
	7	35	$36^6(29)$	0.00999	0.01665	2/5	18.24	0.00333	0.01665	4/5	19.12

A : det (X'X)의 평균 상대오차. O : 5번 시행중 global 최적실험설계를 찾은 횟수.
M : det (X'X)의 최대 상대오차. T : 5번 시행의 평균수행시간.

이상 12개의 문제에 EA 1과 EA 2를 적용한 결과는 <표 3.2>와 같다. <표 3.2>의 결과로부터, EA 2는 EA 1보다 생성해의 질을 나타내는 A, M, O 의 관점에서 일관성 있게 낫고, 평균 수행시간 T 는 EA 1보다 다소 더 소비되는 것을 알 수 있다.

3.2 후보 실험점의 실험비용이 다른 경우

실험비용이 다를 때는 최적실험설계가 알려져 있지 않으므로, complete enumeration에 의해서 global 최적실험설계를 구하여 EA1과 EA2에 의

<u>N = 8일 경우 전산시험 결과</u>

<표 3.3> 실험비용이 다른 경우의 문제와 데이타 (N = 8)

실험후보점	실 험 비 용					
	문제1	문제2	문제3	문제4	문제5	문제6
-1 -1 -1	2	2	10	10	9	20
1 -1 -1	3	3	2	10	3	2
-1 1 -1	2	4	3	10	6	3
1 1 -1	3	5	5	10	5	5
-1 -1 1	2	6	9	2	6	9
1 -1 1	2	8	11	2	4	22
-1 1 1	3	7	7	2	7	7
1 1 1	3	9	4	2	9	6
총실험예산	31	20	31	20	32	50

<표 3.4> 실험비용이 다른 경우의 전산실험 결과 (N = 8)

	Global 최적실험설계의	E A 1				E A 2			
	$det(X'X)$	A	M	O	T	A	M	O	T
문제 1	0.2611 × 105	0.0588	0.20588	2/5	0.46	0.0824	0.20588	3/5	0.48
문제 2	0.2560 × 103	0.0000	0.00000	5/5	0.10	0.0000	0.00000	5/5	0.10
문제 3	0.4096 × 104	0.0000	0.00000	5/5	0.40	0.0000	0.00000	5/5	0.44
문제 4	0.4480 × 103	0.0000	0.00000	5/5	0.24	0.0000	0.00000	5/5	0.16
문제 5	0.9600 × 103	0.0400	0.20000	4/5	0.24	0.0000	0.00000	5/5	0.28
문제 6	0.1638 × 105	0.3862	0.45067	0/5	0.78	0.3862	0.45067	0/5	0.86

<표 3.5> 실험비용이 다른 경우의 문제와 데이타 (N = 12)

실험후보점	실 험 비 용	
	문제 1	문제 2
-1 -1 -1	10	10
1 -1 -1	9	2
-1 1 -1	5	3
1 1 -1	3	5
-1 -1 0	6	9
1 -1 0	2	7
-1 1 0	4	13
1 1 0	5	6
-1 -1 1	11	4
1 -1 1	12	5
-1 1 1	6	3
1 1 1	7	6
총실험예산	23	21

<표 3.3>의 6개의 문제를 각각 EA 1과 EA 2로 푼 결과를 <표 3.4>에 정리하였다. <표 3.4>의 결과로 부터, EA 2는 EA 1보다 생성해의 질과 계산시간 면에서 큰 차이가 없음을 알 수 있다.

N = 12일 경우 전산시험 결과

전산실험에 사용한 문제의 후보 실험점과 실험 비용은 <표 3.5>와 같다.
2문제를 각각 EA1과 EA2로 푼 결과는 <표 3.6>과 같다. <표 3.6>로부터 EA1과 EA2는 global 최적실험계획을 비교적 잘 구성하는 것을 알 수 있다.

<표 3.6> 실험비용이 다른 경우의 전산실험 결과 (N = 12)

	Global 최적실험설계의	E A 1				E A 2			
	$det(X'X)$	A	M	O	T	A	M	O	T
문제 1	0.3840 × 103	0.1500	0.33333	2/5	0.34	0.2000	0.50000	3/5	0.28
문제 2	0.1204 × 104	0.0125	0.06250	4/5	0.40	0.1688	0.84375	4/5	0.34

4. 결 론

본 논문에서는 해양환경에 관련된 실험을 할 경우에 앞에서 언급한 실험시간, 실험비용, 실험의 난이도 등의 실험제한요건에 의해 충분한 반복실험을 하지 못할 경우의 최적실험계획문제를 다루었고, 폐수처리에서 중요시되는 미생물의 활성도를 측정하기 위한 실험계획을 예제로 제시하였다.

또한, 2장에서는 실험제한 요건중에서 실험비용에 대해 각 실험점마다 동일하지 않은 경우의 최적실험계획의 모형과 그 해법을 제시하였다. 실험비용이 각 실험점마다 동일하지 않은 경우에는 기존의 해법이 존재하지 않으므로, 해법의 근거가 되는 이론을 증명하여 발견적해법(heuristic method)을 개발하였으며, 그 해법의 성능을 3장에서 비교하였다. 비교 결과, 본 논문에서 다룬 최적실험계획 문제를 해결하는데 유용한 것으로 사료된다.

따라서, 본 논문에서 제시하는 최적실험계획은 예제로 제시한 미생물의 활성도를 측정하기위한 실험계획등을 비롯한 해양환경분야에서 발생하는 실험계획에 적용 될 수 있다.

참 고 문 헌

1) Smith, K., "On the standard deviation of adjusted and inter-polated values of an observed polynomial function", Biometrika. vol 12, 1-85, 1918.
2) Wald, A., "On the efficient design of statistical investigations" Annals of Mathematical Statistics. vol 14, 134-140, 1943.
3) Hotelling, H., "Some improvements in weighing and other experimental techniques", Annals of Mathematical Statistics. vol 15, 297-306, 1944.
4) Mood, A.M., "On Hotelling's weighing problem", Annals of Mathematical Statistics. vol 17, 432-446, 1946.
5) Kiefer, J. and Wolfowitz, J., "The equivalence of two extremum problem", Canadian Journal of Mathematics, vol 12, 363-366,1960.
6) Atkinson, A.C., "Developments in the design of experiments", Interational Statistical Review. vol 50, 161-177, 1982.
7) Pazman, A., "Some features of the optimal design theory-A survey" Math. Oper. Und. Stat, Ser. stat., vol 11, 415-446, 1980.
8) Ash, A., Hedayat, A., "An Introduction to design optimality with an overview of the literature", Communications in Statistics, vol A7, 1295-1325, 1978.
9) Mitchell, T.J., "An algorithm for construction of D-optimal experimental designs", Technometrics, vol 16, 203-201, 1974.
10) Mitchell, T.J., Miller,F.L,Jr., "Use of design repair to con- struct designs for special linear models", Math. Div. Ann. Progr. Rept. Oak Ridge, Tn. 130-131, 1970.
11) Wynn, H.P., "The sequential generation of D-optimum experimental designs", Annals of Mathematical Statistics. vol 41, 1655-1664, 1970.
12) Fedorov, V.V., Theoy of Optimal Experiments, Trans. and Ed. by W.J. Studden and E.M. Klimko, New York: Academic Press, 1972.
13) St.John, R.C., Draper, N.R., "D-otimality for regression designs: A review", Technometrics. vol 17, 15-23, 1975.
14) Galil, Z., Kiefer, J., "Time and space-saving computer methods, related to Mitchell's DETMAX, for finding D-optimum designs", Technometrics. vol 22, 301-313, 1980.
15) Galil, Z., Kiefer, J., "D-optimum weighing designs", Annals of Statistics. vol 8, 1293-1306, 1980.
16) Johnson, M.E., Nachtsheim, C.J., "Some guidelines for constructing exact D-optimal designs on convex design space", Techno-metrics. vol 25, 271-277, 1983.

17) Weinstein, I.J. and Yu, S.O., "Comment on integer Maximization probelm", Operation Res., vol 21, 648-650, 1972.

18) Welch, W.J., "Branch and bound search for experimental designs based on D-optimality", Technometrics., vol 24, 41-48, 1982.