# 실험 비용을 고려한 최적 실험 설계의 발견적 해법에 관한 연구

여 동 석[1), 金 宰 煥[2)

# A Heuristic Method for Solving Optimal Designs with Considering Experimental Costs

Yeo, Dong Suk. Kim, Jae Hwan.

## ABSTRACT

This thesis develops a new heuristic, the Excursion Algorithm(EA), for solving optimal design problem with considering experimental cost. 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.

1)한국해양대학교 응용수학과 석사과정 전산통계 전공

2)한국해양대학교 응용수학과 교수

# I. 서 론

## 1.1 연구배경 및 목적

어떤 관측치를 추정하기 위해 실험을 실시해야 할 경우 미지의 모수(unknown parameter)를 정확하게 추정하기 위해서는 실험에 앞서 실험계획(experimental design)을 해야만 한다. 실험 불가능한 점이 존재하거나 실험비용이 매우 커서 충분한 반복실험을 할 수 없는 경우에는 요인 실험(factorial design)이나 일부요인 실험(fractional factorial design) 등의 전통적인 실험계획을 사용 할 수가 없다. 따라서, 이런 경우에는 최적 실험설계(optimal design)방법에 의해 실험을 계획하는 것이 하나의 유용한 대안이 될 수 있다.

현재까지 최적실험설계 문제는 각 실험점의 실험비용이 동일하다는 가정하에 다루어져 왔는데, 이는 많은 경우에 있어서 비현실적인 가정이다. 그러므로, 본 논문에서는 각 실험점의 비용이 다르고 총 실험 비용이 주어진 경우에 대하여, 최적 실험설계의 새로운 모형을 제시하고 "Execursion Algorithm(EA)" 이라는 새로운 발견적 해법을 적용함으로써 좀 더 현실적인 최적실험설계 방법을 마련하고자 한다.

본 연구에서 다루고자 하는 최적실험설계의 모형은 다음과 같다.

$$y_i = \beta_1 f_1(x_i) + \beta_2 f_2(x_i) + \cdots + \beta_k f_k(x_i) + e_i \qquad i = 1,\ 2,\ \cdots,\ n \qquad (1.1)$$

실험오차 $\{e_i\}$에 대해서는, $E(e_i)=0$, $Var(e_i)=\sigma^2$, $Cov(e_u, e_v)=0$, $u \neq v$를 가정한다. 모형 (1.1)을 벡터/행렬 형태로 나타내면 다음과 같다.

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

단,

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

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

$$\beta' = (\beta_1,\ \beta_2,\ \cdots,\ \beta_k),$$
$$e' = (e_1,\ e_2,\ \cdots,\ e_n).$$

모형 (1.2)에서 미지의 모수 벡터 $\beta$의 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. 전통적 최적실험설계 문제는 적절한 기준이 최대 또는 최소화되도록 실험영역 $R_x$에서 $n$개의 실험점을 선택하는 것이다. 현재까지 제시된 주요기준은 <표 1.1>과 같다(Steinberg와 Hunter[19]). 본 연구에서는 이 중에서 가장 널리 채택되고 있는 $D$−최적기준 하에서의 최적실험설계 문제를 다루고자 한다. $D$−최적 실험설계를 구성하기 위한 기준의 방법을 소개하기 전에 먼저 정확한 설계(exact design)와 근사적 설계(approximate design)의 차이에 대해 살펴보면 다음과 같다. 모형 (1.1)에서 $x_i$가 compact design space $R_x$의 원소이고 $P_n$이 $R_x$에서의 probability measure일 때, $nP_n$이 정수이면 $P_n$에 의한 실험설계를 정확한 설계라 하고, $nP_n$이 정수가 아니면 근사적 설계라 부른다. 본 논문에서는 정확한 설계를 구성하는 문제를 다루고자 한다.

<표 1.1> 최적실험계획의 기준

| 기 준 | 내 용 |
|---|---|
| $D$−최적기준 | $det(X'X)^{-1}$의 최소화, 또는 $det(X'X)$의 최대화. |
| $A$−최적기준 | $tr(X'X)^{-1}$의 최소화. |
| $E$−최적기준 | $(X'X)^{-1}$의 고유치(eigenvalue)의 최대치를 최소화. |
| $G$−최적기준 | $x_i$의 영역 $R_i$에서 $v(x_i)$의 최대치를 최소화. |
| $I_\lambda$−최적기준 | $\int_{R_i} v(x_i)\lambda(x_i)\,dx_i$를 최소화.<br>$\lambda(x_i)$는 $x_i$의 probability measure. |

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

<표 1.2> D 최적 실험에 대한 기존의 연구

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

## 1.2 연구범위

본 연구에서는 기존 연구와는 달리 <그림 1>에서 처럼 요인 실험으로 제한 하지 않은 일반적인 실험영역에 대한 실험비용을 고려한 새로운 최적 실험설계 모형을 제시하고자 하며, 또한 그 해법으로서 발견적 해법($EA$)을 개발하여 적용함으로써 이 문제에 대한 최적 실험설계를 구하여 정확한 추정치를 얻고자 하는 실험자에게 제공 하고자 한다.

Ⅱ장 에서는 새로운 모형 및 발견적 해법과 그 배경 이론을 제시하고자 하며, Ⅲ장 에서는 개발한 발견적 해법의 두가지 탐색 전략인 $EA1$과 $EA2$의 성능을 평가 하고자 한다.

# Ⅱ. 실험비용을 고려한 최적실험 설계

## 2.1 모형

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

$$(P)\quad Maximize \quad det(X'X)$$

$$\sum_{i=1}^{N} c_i n_i \le B$$

$$n_k = 1 \quad \forall k \in PR$$

$$n_i \ge 0 \text{ , 정수} \quad i=1,\ 2,\ \cdots,\ N$$

$c_i$ 실험점 $x_i$에서의 실험비용

$n_i$ 실험점 $x_i$에서의 실험횟수

$B$ 실험 예산(experimental budget)

$N$ 실험후보점의 총 갯수

$PR$ 반드시 실험을 해야 하는 실험후보점들의 집합

위의 문제 $(P)$는 다음과 같은 사항들이 포함된다.

i. 선형회귀모형 (1.1)을 가정한다. 따라서, 최적실험설계의 서로 다른 실험점의 수는 $k$ 이상이어야 한다.

ii. 실험영역은 $N$ 개의 실험후보점으로 구성된다.

iii. $N$ 개의 실험후보점마다 각각 실험비용이 주어진다.

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. 즉, 미지의 모수의 갯수 $k = 4$이다. 그리고, 실험후보점의 수 $N = 8$이고 이들로 이루어진 행렬 $D$는 다음과 같다고 하자.

각 실험점의 실험 비용 벡터 $C' = (5,\ 7,\ 8,\ 3,\ 10,\ 9,\ 2,\ 6)$ (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.

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

왜냐하면, $n_1 = n_3 = n_4 = n_7 = 1$, $n_2 = n_5 = n_6 = n_8 = 0$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 $X$를 찾는 문제이다.

문제 $(P)$를 풀기 위한 해법(solution method)에 대해 설명하면, 제약식인 실험예산을 만족하는 실험계획 $X$들을 모두 고려하여 그중 $det(X'X)$ 값이 가장 큰 최적 실험계획 $X$를 구하는 enumeration 방법이나 분지 한계법(branch-and-bound method)등의 exact방법을 생각할 수 있으나, 문제의 규모가 커짐에 따라 가능한 실험계획의 $X$들의 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 해결하고

사 한다.

## 2.2 발견적 해법

### 2.2.1 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 $B'=B-\sum_{i\in FR} c_i$ 이다.

$FR$ 집합을 다음과 같이 표현하면,

$FR=\{X \mid X$ 는 $P(N)$을 만족하는 실험 행렬$\}$.

앞절의 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] $IBDM(B', C)$

$FR$의 실험행렬 $X(m_1, m_2, \cdots, m_N)$이 $FR$의 임의의 실험행렬 $X(n_1, n_2, \cdots, n_N)$에 대해 $m_i \leq n_i$ $(i=1,2,\cdots,N)$이면 $m_i=n_i$ $(i=1, 2, \cdots, N)$인때 $X(m_1, m_2, \cdots, m_N)$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, $\sigma$은 집합 $\{1,2,\cdots,n\}$의 순열 이라 하면,

그때 $A'A = A_\sigma' A_\sigma$ 이다.

(증명)

$$(A'A)_{ij} = \sum_{k=1}^{m} (A')_{ik} A_{kj} = \sum_{k=1}^{m} A_{ki} A_{kj}$$

$$= \sum_{k=1}^{m} A_{\sigma(k)i} A_{\sigma(k)j} = \sum_{k=1}^{m} (A_\sigma')_{ik} (A_\sigma)_{kj} = (A_\sigma' A_\sigma)_{ij} \quad ■$$

(보조정리 2) $A$를 $rank\,(A)=n$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 $x \in R^{n \cdot 1}$에 대해 $Ax \neq 0$임을 나타낸다. 따라서, $x'(A'A)x = (Ax)'(Ax) > 0$ 이므로, $A'A$는 Positive definite이다. ■

(정리의 증명)

$X(m_1, m_2, \cdots, m_N) \in FR \setminus IBDM(B', C)$ 라고 가정해 보자. 그러면 모든 $i$에 대해 $m_i \leq n_i$이고 어떤 $j$에 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 최적 실험 계획일수 없게 되고 따라서 증명은 끝난다.

(보조정리 1)에 의해 우리는 $V$에 적당한 행들을 첨가함으로서 행렬 $W$를 얻을수 있다. $rank\ k$ 인 임의의 $m \times k$ 행렬 A와, 0이 아닌 행 벡터 $y$에 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.

### 2.2.2 발견적 해법의 구성

초기 실험설계의 구성

EA는 최적실험설계 문제에 적용함에 있어 국부 최적해에 빠지는 위험을 줄이기 위해 여러번의 '시도'를 수행하는 데, 매 시도마다 초기 실험설계를 랜덤하게 구성하여 탐색을 시작한다.

$W_t(n)$에 실험점을 추가하거나 제거하는 기준

현재까지의 최적실험설계는 실험비용을 고려하지 않았기 때문에, $W_t(n)$에 한 실험점을 첨가하는 기준은 $det(X'X)$를 가장 크게 증가시키기 위해서 $v(x)$가 가장 큰 실험점을 첨가시키고, $W_t(n)$에서 한 실험점을 제거하는 기준은 $v(x)$가 가장 작은 실험점을 제거하는 것이었다 (Nachtsheim[24]). 그러나, 본 연구에서는 실험비용을 고려하기 때문에 실험점을 첨가하거나 제거하는 기준을 다음과 같이 정하였다.

$$\{ \delta \times v_i(x) / c_i + (1-\delta) \times v_i(x) / C_N \} \qquad (2.3)$$

중단기준

$EA$는 $W_t(n)$의 실험비용의 총합 $C_W$와 실험예산 $B$와의 차이의 절대값이 다음과 같은 경우 한 시도를 중단한다.

$$| C_W - B | > \alpha B$$

적절한 $\alpha$ 값은 사용자에 의해 선택될 수 있다.

$F$ 집합의 처리

현재 실험설계의 $det$ 값은 $F$ 집합의 $det$ 값과 비교하여 전자의 $det$ 값이 $(1 \pm 0.05) \times (F$ 집합의 $det$ 값$)$ 사이에 있으면 현재 실험설계를 $F$ 집합에 포함시킨다.

### 2.2.3 $EA1$과 $EA2$의 단계

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

$EA1$ (단축탐색)의 알고리즘 단계

0. $\alpha$, $h$, $\delta$를 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. $E = E \cup W_t(n)$.

   $W_t(n) \in F$ 이면, $F = F \cup E$, $E = \phi$. 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, $W_t(n) = W^c(n)$, $n = n^c$, $F = F \cup E$, $E = \phi$, $t = t+1$. 단계 3으로 간다.

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

<u>*EA*2의 알고리즘 단계</u>

0. $\alpha$, $h$, $\delta$ 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. $U$의 난수 $RN$을 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.

$W_t(n) \not\in F$ 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, $W_t(n) = W^c(n)$, $n = n^c$, $F = F \bigcup E$, $E = \phi$, $t = t + 1$. $IFLAG = 0$ 이면, 단계 4로 간다.

8. $W_t(n)$에서 한 실험점을 제거한다.

$|C_W - B| > \alpha B$이면 단계 11로 간다.

9. $E = E \bigcup W_t(n)$.

$W_t(n) \in F$ 이면, $F = F \bigcup E$, $E = \phi$. 단계 8로 간다.

$W_t(n) \not\in F$ 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 예제

모형이 $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ 이고, 8개의 후보 실험점으로 구성된 $D$와 $C$는 다음과 같다고 하자.

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

$$C = (9,\ 3,\ 6,\ 5,\ 6,\ 4,\ 7,\ 9)$$

여기서, $B = 32$라고 하고, $PR=\{7\}$이라 하면, EA2에 의해 $w^c(6)=\{d_1, d_7, d_6, d_4, d_2, d_6\}$가 최적실험계획이 된다. 즉, $d_1, d_7, d_4, d_2$실험점에서 각각 한번의 실험을 하고, $d_6$실험점에서 2번의 실험을 실시하는 실험계획이 최적실험계획임을 알려준다.

# Ⅲ. 전산 실험

## 3.1 후보 실험점의 실험비용이 같은 경우

수준이 2개인 요인실험에서 반응치 $y$와 독립변수 $x_i$들 간의 관계가 1차 회귀모형으로 표현될 때, Galil과 Kiefer[15]에 의해 알려진 global 최적실험설계의 $det(X'X)$ 값은 표 3.2와 같다. 단, $n$은 실험점의 수이고 $mod$는 $n$을 4로 나눈 나머지이다.

<표 3.1> Galil과 Kiefer[15]의 최적실험설계

| $mod$ | $det(X'X)$ |
|---|---|
| 0 | $n^k$ |
| 1 | $(n-1)^k(n-1+k)$ |
| 2 | $(n-2)^{k-2}(n-2+k)^2$ : $k$ 짝수<br>$(n-2)^{k-2}(n-1+k)(n-3+k)$ : $k$ 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).

이상 12개의 문제에 $EA1$과 $EA2$를 적용한 결과는 표 3.2과 같다. 표 3.2의 결과로부터, $EA2$는 $EA1$보다 생성해의 질을 나타내는 $A, M, O$의 관점에서 일관성 있게 낫고, 평균 수행시간 $T$는 $EA1$보다 다소 더 소비되는 것을 알 수 있다.

<표 3.2> 실험비용이 같은 경우의 전산실험 결과

| mod | k | n | Global 최적 실험설계의 det (X'X) | EA1 A | EA1 M | EA1 O | EA1 T | EA2 A | EA2 M | EA2 O | EA2 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 상대 오차.

M: det (X'X)의 최대 상대 오차.

O: 5번 시행중 global 최적실험설계를 찾은 횟수.

T: 5번 시행의 평균수행시간.

## 3.2 후보 실험점의 실험비용이 다른 경우

실험비용이 다를 때는 최적실험설계가 알려져 있지 않으므로, complete enumeration에 의해서 global 최적실험설계를 구하여 $EA1$과 $EA2$에 의한 최적실험설계와 비교하였다. 비용제약을 만족하는 실험설계를 모두 나열하는 데는 컴퓨터 수행시간이 너무 많이 소요되기 때문에, $N$을 8과 12로 제한하였으며, 실험예산 $B$도 작게 설정하였다.

$N = 8$일 경우 전산시험 결과

$N = 8$일 때 시험을 수행한 예제는 다음 <표 3.3>과 같다.

<표 3.3> 실험비용이 다른 경우의 문제와 데이터 ( $N = 8$)

| 실험후보점 | | | 실험비용 | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 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 3.3>의 6개의 문제를 각각 $EA1$과 $EA2$로 푼 결과를 <표 3.4>에 정리하였다. <표 3.3>의 결과로 [illegible]

<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).

이상 12개의 문제에 $EA1$과 $EA2$를 적용한 결과는 표 3.2과 같다. 표 3.2의 결과로부터, $EA2$는 $EA1$보다 생성해의 질을 나타내는 $A, M, O$의 관점에서 일관성 있게 낫고, 평균 수행시간 $T$는 $EA1$보다 다소 더 소비되는 것을 알 수 있다.

<표 3.2> 실험비용이 같은 경우의 전산실험 결과

| *mod* | *k* | *n* | Global 최적 실험설계의 *det* (X'X) | EA1 | | | | EA2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 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 상대 오차.

M: det (X'X)의 최대 상대 오차.

O: 5번 시행중 global 최적실험설계를 찾은 횟수.

T: 5번 시행의 평균수행시간.

## 3.2 후보 실험점의 실험비용이 다른 경우

실험비용이 다를 때는 최적실험설계가 알려져 있지 않으므로, complete enumeration에 의해서 global 최적실험설계를 구하여 $EA1$과 $EA2$에 의한 최적실험설계와 비교하였다. 비용제약을 만족하는 실험설계를 모두 나열하는 데는 컴퓨터 수행시간이 너무 많이 소요되기 때문에, $N$을 8과 12로 제한하였으며, 실험예산 $B$도 작게 설정하였다.

$N$ = 8일 경우 전산시험 결과

$N$ = 8일 때 시험을 수행한 예제는 다음 <표 3.3>과 같다.

<표 3.3> 실험비용이 다른 경우의 문제와 데이터 ( $N$ = 8)

| 실험후보점 | 실험비용 | | | | | |
|---|---|---|---|---|---|---|
| | 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 3.3>의 6개의 문제를 각각 $EA1$과 $EA2$로 푼 결과를 <표 3.4>에 정리하였다. <표 3.3>의 결과로 부터, $EA2$는 $EA1$보다 생성해의 질과 계산시간 면에서 큰 차이가 없음을 알 수 있다.

<표 3.4> 실험비용이 다른 경우의 전산실험 결과 ( $N$ = 8)

| | Global 최적 실험설계의 $det(X'X)$ | $EA1$ | | | | $EA2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | A | M | O | T | A | M | O | T |
| 문제 1 | $0.2611 \times 10^{5}$ | 0.0588 | 0.20588 | 2/5 | 0.46 | 0.0824 | 0.20588 | 3/5 | 0.48 |
| 문제 2 | $0.2560 \times 10^{3}$ | 0.0000 | 0.00000 | 5/5 | 0.10 | 0.0000 | 0.00000 | 5/5 | 0.10 |
| 문제 3 | $0.4096 \times 10^{4}$ | 0.0000 | 0.00000 | 5/5 | 0.40 | 0.0000 | 0.00000 | 5/5 | 0.44 |
| 문제 4 | $0.4480 \times 10^{3}$ | 0.0000 | 0.00000 | 5/5 | 0.24 | 0.0000 | 0.00000 | 5/5 | 0.16 |
| 문제 5 | $0.9600 \times 10^{3}$ | 0.0400 | 0.20000 | 4/5 | 0.24 | 0.0000 | 0.00000 | 5/5 | 0.28 |
| 문제 6 | $0.1638 \times 10^{5}$ | 0.3862 | 0.45067 | 0/5 | 0.78 | 0.3862 | 0.45067 | 0/5 | 0.86 |

$N$ = 12일 경우 전산시험 결과

전산실험에 사용한 문제의 후보 실험점과 실험비용은 <표 3.5>와 같다.

위의 2문제를 각각 $EA1$과 $EA2$로 푼 결과는 <표 3.6>과 같다. <표 3.6>으로부터 알 수 있듯이, $EA1$과 $EA2$는 생성해의 질이나 계산시간 면에서 큰 차이가 없음을 알 수 있다.

<표 3.5> 실험비용이 다른 경우의 문제와 데이터 ( $N$ = 12)

| 실험후보점 | | | 실험비용 문제 1 | 실험비용 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 3.6> 실험비용이 다른 경우의 전산실험 결과 ( $N$ = 12)

| | Global 최적 실험설계의 $det(X'X)$ | $EA1$ | | | | $EA2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | A | M | O | T | A | M | O | T |
| 문제 1 | $0.3840 \times 10^3$ | 0.1500 | 0.33333 | 2/5 | 0.34 | 0.2000 | 0.50000 | 3/5 | 0.28 |
| 문제 2 | $0.1204 \times 10^4$ | 0.0125 | 0.06250 | 4/5 | 0.40 | 0.1688 | 0.84375 | 4/5 | 0.34 |

# IV. 결 론

본 논문에서는 각 실험점에서의 실험비용이 동일하다는 가정을 확장하여 실험비용을 고려한 $D$ 최적실험설계의 새로운 모형을 제시하였고, II장에서의 문제 ( $P$ )를 동일한 문제 ( $P'$ )로 표현하여 최적실험설계가 존재하는 영역에 관한 정리를 증명하

여, 이를 근거로한 발견적 해법인 EA1과 EA2를 고안 하였다.

전산실험 결과, 시험해 본 대부분의 경우에 대해서 global 최적해를 발견할 수 있었다. 또, *EA*1 과 *EA*2 의 성능 차이를 분석해 본 결과, 가능해의 영역이 큰 경우에는(즉, 제약식의 수가 적은 경우에는) *EA*2 가, 가능해의 영역이 작은 경우에는 (즉, 제약식의 수가 많은 경우에는) *EA*1 이 비교적 생성해의 질이 높은 것을 알 수 있었다.

위의 결과를 미루어 문제 (*P*)와 같은 비선형 정수계획 문제에 대해 효율적인 정확한 방법(exact method)이나 발견적 방법이 없을 때에는, 본 논문에서 고안한 발견적 해법이 유용한 대안이 될 수 있다고 판단된다.

최근 인공지능분야 에서도 복잡한 combinatorial 문제를 풀기 위한 일련의 발견적 탐색전략들이 출현하고 있는데, simulated annealing (*SA*), genetic 알고리즘, 신경회로망(neural networks), taboo search (*TS*)등이 그 예이다[25]. 이러한 탐색전략들도 국부 최적해에 도달하는 위험을 줄이고자 고안된 것으로서, 복잡한 combinatorial 문제에 대하여 성공적인 전략으로 보고되고 있다([8], [16], [26] 등). 앞으로, 본 연구에서 개발한 발견적 해법의 보완 및 이러한 방법들과 비교하는 연구가 수행되어야 한다고 본다.

## 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.