

# 정확최적실험계획법

김영일<sup>1,a</sup>

<sup>a</sup>중앙대학교 정보시스템학과

## 요약

실험을 시행하는 경우 모형과 수반하는 가정들이 가지고 있는 불확실성 때문에 하나의 목적만 가지고 실험을 수행하는 경우는 거의 없다. 문헌에서는 이러한 문제를 해결하기 위한 3가지 방법이 존재한다. 최-최소법, 복합실험법 그리고 제약실험법이 그것이다. 복합실험법과 제약실험법의 동격성이 Cook과 Wong (1994)의 논문에 의해 증명된 이래로 많은 수많은 제약실험법의 방법들이 제안되었지만 모두 복합실험법의 근사알고리즘에 기초하였다. 따라서 본 논문에서는 제약실험법에서의 정확알고리즘을 제안하여 실제적인 실험환경 하에서 도움을 주고자 하였다. 이러한 과정에서 본 논문에서는 다른 형태의 제약실험법의 형태를 제안하였으며 부산물로 비타당성의 문제를 볼 수 있는 그림을 첨부하였다.

주요용어: 정확실험계획, 제약최적실험, 다중목적실험.

## 1. 서론

최적실험계획은 이미 많은 통계패키지(SAS, design expert, JMP, MATLAB 등)에서 구현되고 있고 이를 바탕으로 한 많은 공학적인 응용논문들이 나오고 있다. 그러나 아직까지는 일부 통계패키지가 제공하는 루틴들은 실험자가 하나의 목적을 가지고 있는 경우에 국한하는 것들이다. 그러나 Box와 Draper (1975)가 언급하였듯이 실험은 하나의 목적을 가지고 행하여지는 것은 아니다. 따라서 이론적인 최적실험계획법은 다양한 목적을 반영하는 방향으로 수정 보완되었다. 설명을 위해 다음과 같은 일반적인 선형모형을 설정하여 보자.

$$y = X\beta + \epsilon, \quad (1.1)$$

여기서  $y$ 는  $n \times 1$  관측변수 벡터,  $X$ 는  $n \times p$  디자인 행렬,  $\beta$ 는  $p \times 1$  인 미지의 모수 벡터 그리고  $\epsilon$ 은 서로 비상관인 실수 값을 가지는 기댓값이 0이고 분산은 일반성의 손실 없이  $\sigma^2 = 1$ 인  $n \times 1$  오차 확률 변수 벡터이다. 앞으로는  $X$ 는 완전계급(full rank)인 행렬 그리고  $D$ 는  $\beta$ 의 최소제곱추정치의 분산-공분산 행렬,  $(X^T X)^{-1}$ 을 의미한다.  $i$ -번째 관측값,  $y_i = f^T(x_i)\beta + \epsilon_i$ ,  $i = 1, 2, \dots, n$ 은 인자의 개수인  $q$ -차원인 폐구간 실험영역,  $\Omega$ 의 한점,  $x_i$ 에서 얻어진다. 여기서  $f^T(x_i)$ 는  $X$ 에서 해당하는 행(row)이다. 점  $x_i$ 에서의 적합 값(fitted value)의 분산은 분산 함수,  $v(x_i) = f^T(x_i)Df(x_i)$ 에 비례한다. 몇 가지 대표적인 최적실험계획법으로는  $|X^T X|$ 의 값을 최대화하는 실험계획,  $D$ -최적, 일부모수에 해당하는 행렬식의 값만 최대화 하는  $D_s$ -최적, 분산함수의 최대값을 최소화하는 실험계획을  $G$ -최적 그리고  $\text{tr}(D)$ 를 최소화 하는 실험계획,  $A$ -최적 등을 언급할 수 있다.

실험계획 문제는  $\Omega$ 안에 있는  $s$ 개의 점의 유한집합에 질량을 부여하는 확률 질량함수,  $\xi$ 로 기술할 수 있다. 실험계획에서는  $n\xi(x_i)$ ,  $i = 1, 2, \dots, s$ 가 정수여야 하는 제약조건을 지닌다면 이를 정확실험

<sup>1</sup> 이 논문은 중앙대학교 학술연구비 지원을 받아 2007-2008년 안식년 동안 수행되었음.

<sup>a</sup> (456-756) 경기도 안성시 대덕면, 중앙대학교 정보시스템학과, 교수. E-mail: yik01@cau.ac.kr

험(exact experimental design), 그렇지 않은 경우를 근사실험(approximate experimental design)이라 한다. 예를 들어 보자. 이차회귀모형,  $f^T(x) = (1, x, x^2)$ 에서  $D$ -최적을 기준 한다면  $\Omega = [-1, 1]$ 에서 질량 배정은  $\xi(\pm 1) = \xi(0) = 1/3$ 이 되는데  $n$ 이 3의 배수인 경우는 문제가 없으나 그렇지 않은 경우는 위에서 언급한 정수제약조건을 만족시키지 못하는 상황이 발생한다. 거의 모든 이론적인 최적실험계획분야의 연구는 대부분 근사실험을 기반으로 한다. 이는 주어진 실험의 최적 조건을 알 수 있는 조건이 밝혀졌기 때문이다. 예를 들어  $D$ -최적과  $D$ -최적의 동격성(equivalence)을 바탕으로  $D$ -최적의 조건을 확인 할 수 있는 방법이 있다 (Kiefer, 1974). 이외에도 기타 최적조건을 확인하는 방법이 여럿 존재한다. 그러나 실제적으로 실험 현장에서 쓰이는 실험계획은 모두 정확실험이므로 이에 관한 알고리즘이 따로 존재하여야 한다. 왜냐하면 정확실험은 근사실험과 달리 주어진 실험의 최적여부를 알 수 있는 조건은 존재하지 않기 때문이다. Gaffke와 Krafft (1982)는 이차회귀모형에서 최적실험을 찾는 규칙을 발표한 상태이나 임의의 모형에 대해 주어진 실험의 최적여부는 일반적으로 확인하기가 어려운 상태이다. 이차반응표면모형에서 효율적인  $D_s$ -정확실험을 구하기 위해서, Lim (1991)은 근사  $D_s$ -최적설계에 근접한 정확 실험을, Pesotchinsky (1975)는 최소 이차 대칭 집합(minimal second order quadratic sets)를 활용하여 구했지만, 총 실험점의 개수가 중심합성계획(central composite design)보다 커지는 단점이 있다.

따라서 정확 실험환경 하에서 최적에 가까운 실험을 빠른 시간 내에 정확하게 찾아내는 알고리즘에 관한 주제에 관하여 1970년 후반 이후로 많은 연구논문이 나와 있다. 리뷰(review) 논문으로서는 Cook과 Nachtsheim (1980)이 있는데 대부분의 알고리즘은 보통 교환 알고리즘으로 실험점 중의 하나를 후보점(candidate point)의 하나와 교환하는 형태로 구성된다. 축차적으로(sequential) 교환하는 방법과 동시에 교환하는 방법으로 구분지어 지는데 Wynn (1972), Mitchell (1974) 등이 첫 번째 방법에 해당되고 Fedorov (1972), Johnson과 Nachtsheim (1983) 등은 두 번째 방법에 해당이 된다. 이러한 알고리즘의 최대 혜택은 실험에서 인자의 개수,  $q$ 가 증가한다면 3-수준 인자 계획 공간(3-level factorial design space)인 경우 실험 후보점의 개수는  $3^q$ 로 증가하는데 좀 더 빠른 시간 내에 효과적으로 최적실험을 찾을 수 있는 데 있다. 물론 이러한 교환식 알고리즘 이외에도 담금질기법(simulated annealing) 알고리즘이나 유전자(genetic) 알고리즘 등이 있으나 일반적인 상황 하에서는 시간 및 최적의 조건을 찾는데 그렇게 그 성과가 우월하다고 판단되어지지 않는 것이 정설이다. 따라서 많은 통계패키지에서는 교환 알고리즘이 적용되고 있다.

그러나 이러한 기본적인 연구들은 식 (1.1)의 모형과 수반되는 가정을 기반으로 한다. 즉, 모형이 틀린다는지 가정이 올바르지 않은 경우가 밝혀지게 되면 최적실험계획은 최적이지 아닌 실험이 되기 때문이다. 이러한 경우를 염두에 둔 실험이 소위 강건(robust) 실험이다. 모형과 가정에 불확실성이 존재한다면 추후 밝혀질 이들을 염두에 둔 실험이라 할 수 있다. 이러한 상황을 다루기 위한 실험은 문헌에서는 대략적으로 3가지 방법으로 분류되는데 최소-최대실험(mini-max experimental design), 복합실험(compound experimental design), 제약실험(constrained experimental design)이다. 최소-최대실험은 최악의 경우를 최소화시키는 보수적인 방법이며, 여러 개의 경쟁적인 실험기준의 가중평균을 최적화하는 복합실험은 Läuter (1974)에 의해 제안되었으며 제약실험은 Stigler (1971)에 의해 처음 논의가 된 바 있다.

본 연구에서는 편의상 모형에 대한 불확실성이 존재하는 경우만 예를 들어 정확제약실험(exact experimental design)만을 설명할 것이다. 그리고 언급되는 최적실험은  $D$ -최적에 국한할 것이다. 모형과 최적기준에 모두에 대한 강건 실험 구성도 가능하나 문제를 단순화하기 위해 모형에 대한 불확실성만 언급한 강건 실험을 제약조건이라는 틀에 맞추어 설명한다. 제 2절에서는 본 연구에서 언급할 제약실험에 대해 정의하고 이에 대한 알고리즘을 알아본다. 그리고 제 3절에서는 예제와 새로운 다중제약실험을 제안하였고 그리고 마지막으로 제 4절에서는 결론 및 향후 개선방안을 언급하였다.

## 2. 제약실험과 알고리즘

식 (1.1)의 모형을 다음과 같이 분해하여 보자.

$$y = X\beta + \epsilon = X_1\beta_1 + X_2\beta_2 + \epsilon = f_1^T(x)\beta_1 + f_2^T(x)\beta_2 + \epsilon, \quad (2.1)$$

여기서  $n \times p$  행렬  $X$ 는  $n \times k$  행렬  $X_1$ 과  $n \times (p - k)$  행렬  $X_2$ 로 분해하고 모수벡터 역시 적절한 차원을 갖는 벡터  $\beta_1, \beta_2$ 로 구분한다. 식 (2.1)은 식 (1.1)의 불확실성을 염두에 둔 식으로 참의 모형이 식 (1.1)일 수도 있으나 경우에 따라서는 식 (2.1)의 부분 모형

$$y = X_1\beta_1 + \epsilon = f_1^T(x)\beta_1 + \epsilon \quad (2.2)$$

이 참일 수가 있을 것이다. 그러나 실험자는  $\beta_1$ 에 대한 추정에 관심을 기울인다 하더라도 적합결여 검정을 실시하는 경우를 염두에 둔다면 조금 더 포괄적인 모형인 식 (1.1)의 모형을 사용한다.

또한 제약실험을 설명하기 위해 실험의 효율에 대해 정의할 필요가 있다. 비교하고자 하는 기준이 최적실험,  $\xi^*$ 이라면 주어진 실험  $\xi$ 의 효율,  $\phi_\xi(\xi^*)$  중  $D$ -효율(efficiency)을 정의하면 다음과 같다.

**$D$ -효율:** 주어진 실험  $\xi$ 의  $D$ -최적실험  $\xi^D$ 에 대해 가지는 효율은 다음과 같이 정의한다.<sup>2</sup>

$$D\text{-효율: } \phi_\xi(\xi^D) = \left\{ \frac{|X^T X(\xi)|}{|X^T X(\xi^D)|} \right\}^{\frac{1}{p}},$$

여기서  $\xi^D$ 는  $D$ -최적실험을 의미한다. 이와 다른 실험최적인 경우도 유사한 방법으로 정의가 된다.

위에서 언급하였듯이 모형을 2개로 가정하고 두개의 최적실험,  $\xi_i^D, i = 1, 2$ 에 대해 실험  $\xi$ 가 가지는 효율을  $\phi_{xi}(\xi_i^D), i = 1, 2$ 라 한다면  $D$ -최적 제약실험계획은 다음과 같은 제약조건 하에서의 목적함수 값을 최대화 하는 실험  $\xi$ 를 찾는 문제가 된다.

$$\max \phi_\xi(\xi_2^D) \text{ subject to } \phi_\xi(\xi_1^D) \geq c, \quad (2.3)$$

여기서 첨자 인덱스 1은 실험자가 제일 우선적으로 달성하고자 하는 효율을 염두에 둔 모형을 의미하고 인덱스 2는 부차적인 목적에 해당하는 실험기준을 의미한다. 또한  $c$ 는 실험자가 우선적으로 달성하고자 하는  $D$ -효율 값이다. 통상적으로 인덱스 1에 해당되는 모형은 식 (2.2)의 부분모형이고 인덱스 2에 해당하는 모형이 식 (1.1)의 완전모형이다.

이러한 제약실험계획은 Lee (1987), Cook과 Fedorov (1995) 등에 의해 연구가 진행되었고 Cook과 Wong (1994)에 의해 Läuter (1974)의 복합실험계획과 동격성이 증명되었다. 그들에 의하면 복합실험계획은 적절한 제약조건이 형성되어 있으면 대략적으로 제약실험계획의 하나의 해(solution)로 구현된다. 두 실험기준의 동격성에 관한 자세한 내용은 그들의 논문을 참조하기 바란다. 그러나 Cook과 Wong (1994)은 근사실험만을 염두에 두었을 뿐 정확실험을 구현시키는 노력은 하지 않았다. 왜냐하면 근사실험은 동격성에 의해 Läuter (1974)의 논문에 나와 있는 알고리즘에 의해 쉽게 구해지기 때문이다. 다만 Mikulecka (1983)에 의해 제약조건 실험에 대한 대략적인 알고리즘이 언급된 바 있다.

그러나 제약실험의 제일 큰 약점은 조건이 하나가 아닌 두 개 이상으로 확장되는 경우이다. 즉, 참으로 나타날 수 있는 모형의 개수가 3개 이상인 경우는 모든 조건식의 효율,  $c_i, i = 1, \dots, m, m \geq 2$ 를 모두 다 만족시키지는 못하는 비타당성(infeasibility)의 문제가 발생이 되기 때문이다. 이 문제는

<sup>2</sup> 정확실험에서  $D$ -효율을 구하는 경우  $D$ -최적은 정확하게 알려져 있지 않더라도 논문에서는 MATLAB의 coordinate exchange algorithm 에 의해 구한 최적실험을 그 기준으로 삼았다.

Huang과 Wong (1998)과 Kahng과 Kim (2002) 등이 언급한 축차적인 방법이거나 Kim과 Lim (2007)이 제안하였듯이 최소-최대 방법과 제약조건식이 혼합된 형태로 접근이 가능하다. 그러나 이들의 논문은 기본적으로 정확실험을 염두에 두지 않았기 때문에 실제 응용현장에서 적용하기가 현실적이지 않다. 따라서 본 논문에서는 정확실험에 따른 알고리즘을 언급하고자 한다.

**알고리즘:** 아래 제시되는 제약조건하에서의 정확실험의 알고리즘은 기존의 최적실험에 쓰인 알고리즘을 변형하면 완벽한 글로벌 해는 아니지만 어느 정도 만족한 해를 얻을 수가 있다. 이미 많은 패키지가 교환 알고리즘을 이용한 루틴을 가지고 있다. 즉, 인자의 개수를 포함한 모형 및 실험의 수,  $n$ 을 알려주면 자동적으로 실험조건이 제공될 것이다. 본 연구에서는 MATLAB이 가지고 있는 두개의 알고리즘을 염두에 두었다. 하나는 행-교환(row-exchange) 알고리즘이고 다른 하나는 Meyer와 Nachtsheim (1995)의 좌표-교환(coordinate exchange) 알고리즘이다. MATLAB에서는 후보점과 실험점의 동시교환으로 이루어지는 알고리즘이 교환 알고리즘이고 이를 조금 더 효율적으로 교환코자 하는 노력의 하나가 후자의 방법이다. 본 연구에서는 단순 행-교환 알고리즘을 통해 제약실험의 알고리즘을 구현하여 보았다. 알고리즘의 단계는 다음과 같다. 물론 주 3)에서도 언급하였지만 제약조건 실험의 알고리즘을 위해 필요한  $D$ -최적 실험은 MATLAB이 제공하는 좌표교환 알고리즘에 의해 구하였다.

#### 제약조건 $D$ -최적 알고리즘:

**단계 1:**  $i = 0$  시작점인 초기 실험  $\xi_i$ 는 해당하는  $X^T X$ 가 비특이(non-singular)행렬로 시작한다.

**단계 2:** 후보점과 실험점의 교환을 통해 제약조건식을 충족하면, 즉,  $\xi_i$ 가  $\phi_{\xi_i}(\xi_1^D) \geq c$ 를 만족하면 새로운 실험점으로 선택한다. 그렇지 않은 경우는 새로운 실험점 후보로 선택하지 않는다.

**단계 3:** 그러나 조건식을 충족시키다 하더라도 목적함수  $\phi_{\xi_i}(\xi_2^D)$ 을 향상시키지 못하는 경우에는 후보점은 최종적으로 새로운 실험점으로 선택하지 않고 다른 후보점을 모색한다.

**단계 4:**  $i = i + 1$ : 목적함수  $\phi_{\xi_i}(\xi_2^D)$ 의 값이 더 이상 향상되지 않는 점에서 알고리즘을 중단하고 목적함수 값을 보고한다. 그렇지 않으면 단계 2로 넘어간다.

물론 통상적으로 주어진 모형이 하나일 경우는 단순 행-교환 알고리즘은 Meyer와 Nachtsheim (1995)을 위시한 기타의 교환 알고리즘 군에 비해 주어진 시간 내에 최적의 실험을 찾아내는 효율이 떨어지는 것은 사실이나 제약실험 하에서는 참의 모형이 될 수 있는 모형의 개수가 2개 이상이므로 단순 행-교환 알고리즘이 더 효율적이라 판단되어 제약실험에 관련된 알고리즘을 구성하고 실험을 실시하였다. 다만 위에서 언급하였듯이 정확실험의 최적성은 현재로서는 파악하기 힘들다. 다만 지금까지 나온 행-교환 알고리즘의 편의성 및 속도성 때문에 본 논문에서는 유전자 알고리즘과 같은 변형적인 알고리즘은 지양하고 전통적인 행-교환 알고리즘을 채택하였다. 행-교환 알고리즘의 자세한 내용은 Cook과 Nachtsheim (1980)의 논문을 참조하기 바란다.

아직까지 문헌에서 정확실험 환경 하에서의 제약실험의 최적을 찾는 노력이 없었던 것도 본 알고리즘을 시도하여 보는 동기가 되었다고 본다. 단순 행-교환 알고리즘에 의존하지 않은 알고리즘을 연구하는 것이 시도되어야 하겠지만 모형에 따른 분산함수(variance function)가 다른 점은 새로운 알고리즘 개발의 걸림돌로 작용이 된다. 모형이 통일되어 있고 최적실험기준만 다양하게 반영한다면 하나의 통일된 알고리즘이 나올지 않을까 유추하여 본다. 제 3절에서는 이러한 알고리즘을 이용하여 예제를 구현하고 또한 다중제약실험에서의 적용 가능성에 대해서도 알아보도록 한다.

표 1:  $n = 9, c = 0.8$

후보점	알고리즘	이론적인 실험점의 개수
-1	3	3
0	3	3
1	3	3

표 2:  $n = 12, c = 0.9$

후보점	알고리즘에 의한 실험개수	이론적인 실험개수
-1	5	4.86
0	2	2.28
1	5	4.86

표 3:  $n = 6, c = 0.9$

알고리즘 실험점	알고리즘에 의한 실험개수	이론적인 실험점	이론적인 실험점의 개수
-1.0	3	-1	2.43
0.2	1	0	1.14
1.0	2	1	2.43

### 3. 예제

**예제 1.** 이 예제는 Lee (1987)에서 발췌하였다. 실험자가 고려하여야 하는 모형은 인자,  $q = 1$ 인 단순 회귀모형과 이차회귀모형이다. 단순회귀모형의 효율,  $\phi_\xi(\xi_1^D)$ 의 하한값을  $c$ 로 한다면  $c$ 가  $\sqrt{2/3}$  이하로 설정되면  $\phi_\xi(\xi_2^D)$ 은 100%의 효율을 보장하나 그렇지 않다면  $\phi_\xi(\xi_2^D)$ 은 100%를 보장받지 못한다. 전자인 경우 제약실험은  $\xi(\pm 1) = \xi(0) = 1/3$ 로 나올 것이나 후자의 경우는 그렇지 못하다. 이 경우 목적함수의 최대값은  $[27(1 - \xi(0))^2\xi(0)/4]^{1/3}$ 로 나타나며 제약실험은  $\xi(\pm 1) = [1 - \xi(0)]/2, \xi(0) = 1 - c^2$ 이다. 예를 들어  $c$  값이 0.9인 경우 목적함수 값은 0.944081이 된다. 이러한 근사실험을 염두에 두고 제 2절에서 언급한 알고리즘을 구현하여 보았다.

Case 1: 실험개수  $n = 9, c = 0.8$  그리고 실험후보점은 실험영역  $\Omega = [-1, 1]$ 을 격자 0.01로 나눈 모든 점을 대상으로 한다.  $n\xi(x_i), i = 1, 2, 3$ 으로 계산한 실험점 -1, 0, 1에 배정된 실험점의 개수는 반올림하여 각각 3개가 되어야 한다. 왜냐하면 하한값으로 설정된  $c$  값이 0.8로 너무 낮게 설정되어 2차형식의 모형에 대한  $D$ -최적실험이 이를 충분히 달성하기 때문이다. 당연히 제약실험의 최적은  $\xi_1^D$ 이다. 이러한 결과는 제 2절에서 언급한 알고리즘을 이용한 구현된 실험과 정확하게 일치한다 (표 1).

Case 2: 그러나 실험개수  $n = 12, c = 0.9$ 의 경우를 보도록 하자. 역시 마찬가지로 실험후보점은  $\Omega = [-1, 1]$  영역을 격자 0.01로 나눈 모든 점을 대상으로 한다. 만족하고자 하는  $D$ -효율,  $c$ 의 값이 0.8에서 0.9로 상승을 하였기 때문에 이차회귀모형에 대한  $D$ -최적이 바로 제약  $D$ -최적이 되지 않는다.  $n\xi(x_i), i = 1, 2, 3$ 으로 계산한다면 실험점 -1, 0, 1에 배정된 실험점의 개수는 반올림하여 각각 5개, 2개, 5개가 된다. 이 역시 알고리즘으로 구현된 실험과 같음을 알 수 있다 (표 2).

Case 3: 이번에는  $n = 6, c = 0.9$ 로 설정하여 보자. 마찬가지로 실험후보점은  $\Omega = [-1, 1]$ 을 격자 0.01로 나눈 모든 점을 대상으로 한다.  $n\xi(x_i), i = 1, 2, 3$ 으로 계산한다면 최종 실험점 -1, 0, 1에 배정된 실험점의 개수는 반올림하여 각각 2개, 1개, 2개가 된다. 총 실험개수가 반올림하는 과정에서 6개에서 5개로 1개가 줄고 실험점이 알고리즘으로 구현된 실험점과 상이함을 알 수 있다 (표 3).

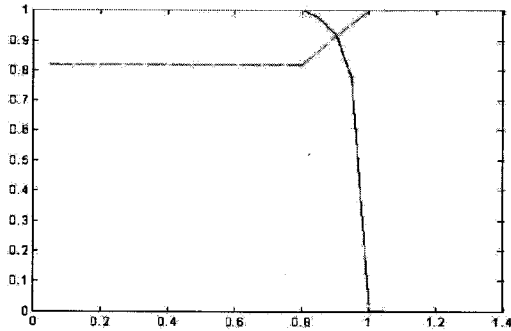


그림 1:  $c$  값 변화에 따른 두  $D$ -효율 비교,  $n = 6, q = 1$ , 격자 = 0.01

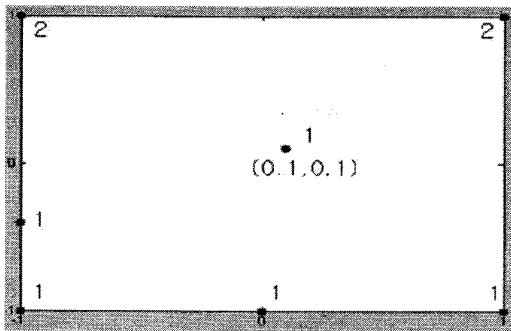


그림 2:  $n = 9, c = 0.8$ , 격자 = 0.01

이와 같이 Kiefer (1971)에 의하면  $n$ 이 충분히 크다면 굳이 정확실험이 필요는 없으나  $n$ 이 작거나 중간크기인 경우는 실험자는 어느 실험점에 정확하게 몇 개의 실험개수를 어떻게 배정하여야 하는지 명확치 않은 경우가 발생된다. 따라서 제 2절에서 언급한 알고리즘은 이를 보완해준다.

그러나  $n$ 이 주어지더라도  $D$ -효율의 하한값인  $c$  값을 실험자가 정하는 것은 쉽지 않다. 보조적인 의미로 제약조건식에 명시된 첫 번째 모형에 대한  $D$ -효율의 하한값  $c$ 와 목적함수의  $D$ -효율 값을 비교함으로써 실험자에게 보다 나은 정보를 제공할 수 있는 차원에서 그림 1을 그려 보았다.  $y$ -축은 각각 색을 달리하여 제약조건식과 목적함수의  $D$ -효율이고  $x$ -축은  $c$  값이다.  $c$  값이 설정 낮게 설정되더라도 알고리즘은 이를 최대한으로 끌어 올려 제약조건식에 명시된 모형의 실제  $D$ -효율 값이  $c$  값보다 큼을 알 수 있다. 혹  $c$  값이 지나치게 높게 책정된 경우에는 목적 함수의  $D$ -효율이 100%에서 빠른 속도로 0%로 감소함을 알 수 있다. 두 곡선이 만나는 점은 최소-최대 실험과 일치하는 점이다. 그림 1은  $n = 6$ 을 기준으로 하였다.

**예제 2.** 이 예제에서는 인자  $q = 2$ 인 모형을 고려하여 보자. 실험자가 고려하여야 하는 모형은 선형 항목과 교호작용이 들어간 회귀모형,  $f_1^T(x) = (1, x_1, x_2)$ 과 이차회귀반응모형,  $f(x) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$ 이다. 참고로  $f(x)$ 에 대한 근사실험계획은 다음과 같다.

$$\xi(\pm 1, \pm 1) = \frac{0.583}{4}, \quad \xi(0) = 0.096, \quad \xi(0, \pm 1) = \xi(\pm 1, 0) = \frac{0.321}{4}.$$

제 2절에서 언급한 알고리즘을 구성하여 제약실험을 구성하여 보았다.

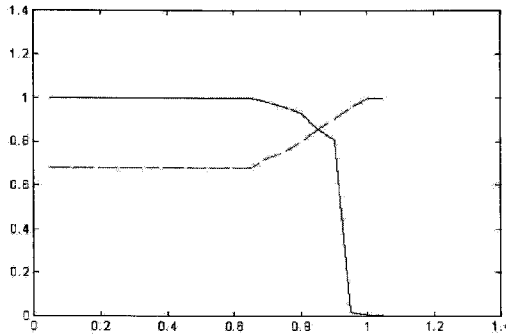


그림 3:  $c$  값 변화에 따른 두  $D$ -효율 비교,  $n = 9, q = 2$ , 격자 = 0.1

Case:  $n = 9, c = 0.8$  그리고 실험영역,  $\Omega = [-1, 1]^2$ 을 격자 0.1로 나눈 모든 점을 대상으로 한다. 알고리즘에 의해 구해진 실험은 그림 2와 같다. 가운데 괄호안의 숫자는 실험점의 좌표 표기이며 나머지 숫자는 각 실험점에 배정된 실험개수이다. 근사실험에 대한 이론적인 실험이 구체적으로 표기가 안되므로 이와 비교는 생략키로 한다.

그림 3은 두개의 효율을 비교한 그림이다.  $c = 0.8$ 인 경우 두 모형에 대한  $D$ -효율은 각각 0.8012, 0.9329와 같다.  $q = 1$ 의 경우인 그림 1과 비교하면 전체적인 모양은 비슷하나 목적함수의  $D$ -효율은 설정된  $c$  값이 0.6을 지나면서부터 급격하게 떨어짐을 알 수 있다.

**예제 3.** 위에서 언급 하였듯이 그러나 제약실험의 제일 큰 약점은 제약조건이 하나가 아닌 두개 이상으로 확장되는 경우이다. 즉 참으로 나타날 수 있는 모형의 개수가 3개 이상인 경우는 모든 조건식의 효율,  $c_i, i = 1, \dots, m, m \geq 2$ 를 모두 다 만족시키지는 못하는 비타당성의 문제가 발생이 되기 때문이다. 그러나 제 2절에서 언급한 알고리즘은 이러한 제약조건의 비타당성 문제가 확인되는지 알려줄 수는 있다. 여기서는 단순하게 모든  $c_i = c$ 로 묶어 처리하였다. 즉, 다중제약 최적실험은  $m = 2$ 인 경우 다음과 같이 표현된다.

$$\max \phi_\xi(\xi_3^D) \text{ subject to } \phi_\xi(\xi_1^D) \geq c, \phi_\xi(\xi_2^D) \geq c$$

물론  $c_i$  값을 달리하여 알고리즘을 설정하는 것은 가능하지만 조건식을 부여하는 문제가 추가적으로 발생하므로 현실적으로 이를 단순화시켰다. 그리고 현장에서는 단일하게 하나의 값을 부여하는 것이 현실적이기도 하다.

본 예제에서 실험자가 염두에 두는 모형은 3개이다. 인자  $q = 1$ 이고 단순회귀모형, 이차형식의 모형 그리고 3차 모형이라고 가정하자. 일차와 2차 모형에 대해 실험이 가지는  $D$ -효율을  $c$ 라 하고 3차 모형에 대한  $D$ -효율을 최적화 하는 다중제약실험이다.

Case: 3차 회귀모형인 경우 근사실험인 경우 실험점이 4개로  $(\pm 1, \pm 1/\sqrt{5})$ 인 바 4의 배수인  $n = 16$ 으로 정하여 그 실험점의 구성을 알아보았다. 제 2절의 알고리즘을 이용하여 구한 결과이다.  $c = 0.5$ 로 설정하였다. 그리고 실험후보점은  $\Omega = [-1, 1]$  영역을 격자 0.01로 나눈 모든 점을 대상으로 한다. 이 경우는 3차 항의 모형에 대한  $D$ -최적 실험은 나머지 두 모형에 대해 0.5라는 충족시키고자 하는  $D$ -효율의 하한값을 만족한다. 따라서 4개의 실험점  $\pm 1, \pm 0.045$ 에 각각 4개의 실험개수를 배정한다. 아래 결과는 이러한  $D$ -최적 실험이 단순회귀모형과 2차 모형에 대해 가지는 효율은 각각 77.54%, 86.41%로 나타난다. 그러나 하한값을 50%에서 77.54% 이상으로 상승시키면 그림 4에서 보다시피 비타당성의 문제가 곧 발생된다. 즉, 동시에 77.54% 및 86.41% 이상을 보장하는 4개의 실험점을 가진 실험은 존

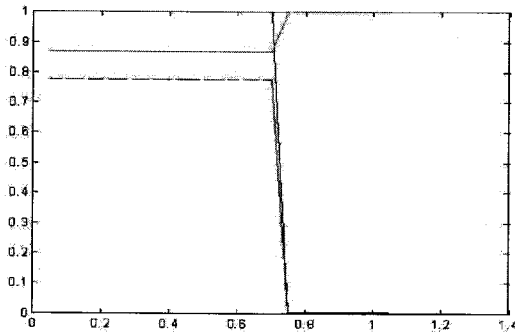


그림 4:  $c$  값에 따른 3개의  $D$ -효율 비교,  $n = 16$ ,  $q = 1$ ,  $m = 2$ , 격자 = 0.01

재 하지 않기 때문이다. 여기서 4개는 3차 회귀모형이 가져야하는 실험점의 개수이다. 모수의 개수가 4개이기 때문에 4개미만으로 실험점은 설정될 수 없다.

#### 4. 결론

본 연구에서는 제약실험에서의 정확실험을 알아보았다. 기존 문헌에서 언급되지 못한 제약조건하에서의 정확실험을 실패함으로써 실험현장에서 보다 나은 정보를 제공하기 위해 필요한 알고리즘을 구현해보는 것이 본 연구의 목적이었다. 이를 위해 기존의 행-교환 알고리즘의 변형을 통하여 이에 대한 가능성을 보고자 하였다.  $c$  값을 하나로 묶어 처리하는 방법은 비타당성의 문제를 풀지는 못하지만 Kahng과 Kim (2002)이 제안한 축차적인 방법과는 다른 각도에서 제안하여 보았다. 본 연구에서는 모형에 대한 불확실성이 존재하는 경우만 훑어보았지만 가정과 실험기준에 대한 다목적성도 변형하여 적용이 가능하다. 예를 들어  $D$ -효율은  $c$ 만큼 보장받고자 하는 상황 하에서  $A$ -효율을 최적화 시키는 문제라든지 더 나아가 모형에 대한 불확실성과 최적기준의 복합적인 문제까지 제약조건의 틀에서 논할 수 있을 것이다. 다만 단순 행-교환 알고리즘의 틀에서 벗어나 효율적인 알고리즘의 개발이 필수적일 것이다. 본 연구에서 언급한 예제들은  $q = 2$  이하인 경우에만 국한하였지만 이를  $q \geq 3$ 인 다차원 모형에도 적용시킬 수 있음은 물론이다. 또한 다양한  $n$ 에 따른 카탈로그를 만들어 실험자에게도 제공하여야 한다. 모형에 대한 불확실성에 따른 다양한 실험기준의 설정 역시 추후 논의될 사안이 될 수 있을 것이다. 별명(alias)행렬에 대한 문제를 제약조건 틀에서 논의하는 것 역시 추후 과제이다. 혹 본 논문에 쓰였던 Matlab 코드가 필요한 독자는 저자의 이메일로 문의하기 바란다.

#### 참고 문헌

- Box, G. E. P. and Draper, N. R. (1975). Robust design, *Biometrika*, **62**, 347–352.
- Cook, R. D. and Fedorov, V. V. (1995). Constrained optimization of experimental design, *Statistics*, **26**, 129–178.
- Cook, R. D. and Nachtsheim, C. J. (1980). A comparison of algorithms for constructing exact D-optimal designs, *Technometrics*, **22**, 315–324.
- Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrained and compound optimal designs, *Journal of the American Statistical Association*, **89**, 687–692.
- Fedorov, V. V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.
- Gaffke, N. and Krafft, O. (1982). Exact D-optimum designs for quadratic regression, *Journal of Royal Statistical Society, Series B*, **44**, 394–397.
- Huang, Y. C. and Wong, W. K. (1998). Sequential construction of multiple-objective optimal designs, *Biometrics*, **54**, 1388–1397.



- Johnson, M. E. and Nachtshiem, C. J. (1983). Some guidelines for constructing exact D-optimal designs on convex design spaces, *Technometrics*, **25**, 271–277.
- Kahng, M. W. and Kim, Y. I. (2002). Multiple constrained optimal experimental design, *The Korean Communications in Statistics*, **9**, 619–627.
- Kiefer, J. (1971). *The Role of Symmetry and Approximation in Exact Design Optimality*, In Statistical Decision Theory and Related Topics, Academic Press, New York.
- Kiefer, J. (1974). General equivalence theory for optimal designs (approximate theory), *The Annals of Statistics*, **2**, 849–879.
- Kim, Y. and Lim, Y. B. (2007). Hybrid approach when multiple objectives exist, *The Korean Communications in Statistics*, **14**, 531–540.
- Läuter, E. (1974). Experimental planning in a class of models, *Mathematische Operationsforschung und Statistik*, **5**, 673–708.
- Lee, C. M. S. (1987). Constrained optimal designs for regression models, *Communications in Statistics, Part A- Theory and Methods*, **16**, 765–783.
- Lim, Y. B. (1991). Exact  $D_s$ -efficient designs for quadratic response surface model, *Journal of the Korean Statistical Society*, **20**, 156–161.
- Meyer, R. K. and Nachtshiem, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs, *Technometrics*, **37**, 60–69.
- Mikulecka, J. (1983). On hybrid experimental designs, *Kybernetika*, **19**, 1–14.
- Mitchell, T. J. (1974). An algorithm for the construction of D-optimal experimental designs, *Technometrics*, **16**, 203–210.
- Pesotchinsky, L. L. (1975). D-optimum and quasi-D-optimum second order designs on a cube, *Biometrika*, **62**, 335–340.
- Stigler, S. M. (1971). Optimal experimental design for polynomial regression, *Journal of the American Statistical Association*, **66**, 311–318.
- Wynn, H. P. (1972). Results in the theory and construction of D-optimum experimental designs, *Journal of Royal Statistical Society, Series B*, **34**, 133–147.

# Exact Constrained Optimal Design

Young Il Kim<sup>1,a</sup>

<sup>a</sup>Dept. of Information System, ChungAng Univ.

---

## Abstract

It is very rare to conduct an experimental design with a single objective in mind. since we have uncertainties in model and its assumptions. Basically we have three approaches in literature to handle this problem, the mini-max, compound, constrained experimental design. Since Cook and Wong (1994) announced the equivalence between the compound and the constrained design, many constrained experimental design approaches have adopted the approximate design algorithm of compound experimental design. In this paper we attempt to modify the row-exchange algorithm under exact experimental design setting, not approximate experimental design one. This attempt will provide more realistic design setting for the field experiment. In this process we proposed another criterion on how to set the constrained experimental design. A graph to show the general issue of infeasibility, which occurs quite often in constrained experimental design, is suggested.

**Keywords:** Exact experimental design, constrained optimal experimental design, multiple objective design.

---

---

Kim's research was supported by ChungAng University's research fund for Kim's 2007-2008 sabbatical leave.

<sup>1</sup> Professor, Department of Information System, ChungAng University, KyungGi-Do 456-756, Korea.

E-mail: yik01@cau.ac.kr