

## 대체모델의 정확성 및 강건성 향상을 위한 가중함수 기반 순차 최소거리최대화계획

장준용\* · 조수길\* · 이태희\*†

\* 한양대학교 공과대학 미래자동차공학과

### Weight Function-based Sequential Maximin Distance Design to Enhance Accuracy and Robustness of Surrogate Model

Junyong Jang\*, Su-gil Cho\* and Tae Hee Lee\*†

\* Dept. of Automotive Engineering, College of Engineering, Hanyang Univ.

(Received December 31, 2014 ; Revised February 16, 2015 ; Accepted February 21, 2015)

**Key Words:** Sequential Design of Experiment(순차실험계획), Maximin Distance Design(최소거리최대화계획), Space Filling Design(충진계획), Kriging Surrogate Model(크리깅대체모델), Correlation Coefficient(상관계수)

**초록:** 효율적인 최적설계를 위해 공학분야에 도입된 대체모델의 정확성은 표본점에 큰 영향을 받는다. 대체모델의 정확성을 높이는 방법으로 기 추출한 응답을 이용하는 순차실험계획이 제안되었다. 크리깅 대체모델의 상관계수를 가중치로 적용하여 대체모델의 정확성을 향상시킨 연구가 있었으나, 주어진 정보가 부족하거나 상관계수가 잘못 추정된 경우 표본점이 잘못 추출되어 대체모델의 강건성이 저하된다. 본 논문에서는 기존 순차실험계획의 여러 문제점을 제시하고, 이를 해결하기 위한 가중함수 기반 순차 최소거리최대화계획을 제안한다. 제안하는 순차실험계획의 효용성을 수학 함수에 적용하여 기존 순차실험계획들과 비교하여 정확성과 강건성이 향상됨을 예시한다.

**Abstract:** In order to efficiently optimize the problem involving complex computer codes or computationally expensive simulation, surrogate models are widely used. Because their accuracy significantly depends on sample points, many experimental designs have been proposed. One approach is the sequential design of experiments that consider existing information of responses. In earlier research, the correlation coefficients of the kriging surrogate model are introduced as weight parameters to define the scaled distance between sample points. However, if existing information is incorrect or lacking, new sample points can be misleading. Thus, our goal in this paper is to propose a weight function derived from correlation coefficients to generate new points robustly. To verify the performance of the proposed method, several existing sequential design methods are compared for use as mathematical examples.

### 1. 서론

최근 최적설계 문제에서 목적함수 또는 구속조건을 계산하기 위해서 다양한 컴퓨터기반 해석을 수행하고 있다. 하지만, 해석 시간이 긴 시뮬레이션 모델을 직접 최적설계 문제에 적용할 경우 설계 비용이 매우 증가하는 문제가 발생한다. 비용 문제를 해결하기 위해 입력변수와 응답간의 관계를 통계적 기법

으로 근사하여 예측점(untried point)에 대해 예측 응답값을 얻는 대체모델(surrogate model)이 제안되었다. 하지만 대체모델이 부정확할 경우 예측 응답값 또한 부정확하여 잘못된 최적설계 값을 도출할 수 있으므로 적절한 실험계획을 수행하여 정확한 대체모델을 만드는 것이 중요하다. 따라서, 컴퓨터기반 실험에 적합한 다양한 실험계획이 제안되었다. 컴퓨터기반 실험은 같은 입력변수에 대하여 같은 응답을 얻는 확정적인(deterministic) 특징 때문에 설계영역에 실험적을 최대한 고르게 분포시켜 고른 정보를 얻는 것이 유리하며, 이러한 특징을 충진성(space filling)으로

† Corresponding Author, thlee@hanyang.ac.kr  
© 2015 The Korean Society of Mechanical Engineers

정의한다.

충진성을 수학적으로 정의하기 위하여 많은 연구가 진행되었다. 유클리디안 거리를 기반으로 충진성을 정의한 최소거리최대화계획(maximin distance design, MD),<sup>(1)</sup> 표본점간 엔트로피(entropy) 함수로 정의한 최대엔트로피계획(maximum entropy design),<sup>(2)</sup> 엔트로피 함수의 행렬식(determinant)을 구하는 과정에서 발생하는 특이성(singularity) 문제를 해결하기 위한 최대고유치계획(maximum eigenvalue design)<sup>(3)</sup> 등이 제안되었다.

한편 기존 표본점의 정보나 기존 표본점으로 구성된 대체모델의 정보를 이용하여 새로운 표본점을 추가하는 순차실험계획이 제안되었다.<sup>(4,5)</sup> 크리깅대체모델(kriging surrogate model)의 상관계수(correlation coefficient)를 각 입력변수의 가중치(weighting coefficient)로 설정하여 입력변수간의 기하학적 거리를 조절하는 순차실험계획이 제안되었다.<sup>(4)</sup> 하지만 주어진 정보가 부족하거나 상관계수가 잘못 추정된 경우, 표본점이 잘못 추출되어 대체모델의 강건성이 저하된다. 본 논문에서는 기존 순차실험계획의 문제점을 해결하고, 대체모델의 정확성과 강건성을 확보하기 위한 상관계수 기반의 가중함수를 제안한다. 가중함수는 단조증가함수인 대수함수를 이용하여 상관계수의 특성을 나타내고, 변수의 개수를 고려하여 변수의 개수에 따른 변별력을 향상시킨다. 제안하는 순차실험계획의 효용성을 수학 함수에 적용하여 기존의 순차실험계획들과 비교 검증한다.

## 2. 크리깅대체모델

### 2.1 크리깅대체모델

크리깅대체모델은 전역모델과 편차항의 합으로 구성되며 식은 다음과 같다.<sup>(6)</sup>

$$\hat{Y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1} \{\mathbf{Y} - \mathbf{F}\hat{\boldsymbol{\beta}}\} \quad (1)$$

여기서,  $\hat{\boldsymbol{\beta}} = \{\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F}\}^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{Y}$ ,  $\mathbf{R}$  은 상관행렬로 데이터들의 공간 상관관계(spatial correlation)를 나타낸다. 본 연구에서 사용한 상관함수는 식 (2)와 같이 가우스(Gauss) 상관함수로 정의한다.

$$R(\boldsymbol{\theta}, \mathbf{x}^i, \mathbf{x}^j) = \exp\left(-\sum_{k=1}^{n_d} \theta_k (x_k^i - x_k^j)^2\right) \quad (2)$$

여기서  $\boldsymbol{\theta}, \mathbf{x} \in \mathbb{R}^{n_d}$  이다. 상관행렬의 상관계수,  $\boldsymbol{\theta}$ , 는 우량함수(likelihood function)를 최대화하여 구하

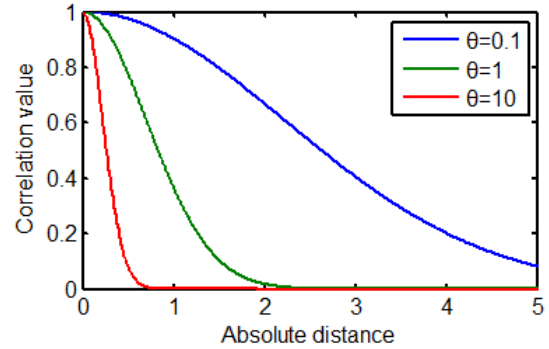


Fig. 1 Gauss correlation function

며, 이를 최우량추정과정(maximum likelihood estimation, MLE)이라 한다.<sup>(6)</sup> 우량함수를 최대화한다는 것은 확률적으로 가능성이 가장 높은 상관계수를 구하는 것을 의미한다.

### 2.2 크리깅대체모델의 상관계수

상관함수는 표본점간의 거리와 상관계수의 함수로써 크리깅대체모델의 특성과 예측성을 결정한다. 가우스 상관함수는 거리가 0 인 지점에서 최대치이며 전 영역에서 미분가능한 특징이 있다.<sup>(7)</sup>

상관계수  $\theta_k$  는  $x_k$  방향에 대한 응답값들의 상관관계를 나타내는 계수로써 크리깅대체모델의  $x_k$ -축의 곡률을 결정한다. 즉, Fig. 1 과 같이 같은 거리에서 상관계수가 클수록, 또는 같은 상관계수에서 거리가 멀수록 상관도는 작아진다. 상관계수가 클수록 각 변수의 축에 대한 크리깅대체모델의 곡률은 비선형 특성을 보인다.

## 3. 최소거리최대화계획

### 3.1 최소거리최대화계획

유클리디안 거리( $l_2$ -norm)를 기반으로 충진성을 정의하는 실험계획은 다음과 같이 정식화 한다.

$$\begin{aligned} & \text{Max}_{\mathbf{x}} \min_{i>j} d(\mathbf{x}^i, \mathbf{x}^j) \\ & \text{where } d(\mathbf{x}^i, \mathbf{x}^j) = \sqrt{\sum_{k=1}^{n_d} w_k (x_k^i - x_k^j)^2} \end{aligned} \quad (3)$$

이 때, 변수의 범위는  $\mathbf{x} \in [0, 1]^{n_d}$  로 정규화한다. 최소거리최대화계획(MD)은 변수의 가중치 없이 정규화된 영역에서 표본점을 추출하므로 가중치  $w_k = 1$  이다.<sup>(2)</sup> 수치적 표현이 간단하여 널리 쓰이나, 응답의 특성을 고려하지 못한다. 이를 극복하기 위하여 크리깅대체모델의 상관계수를 가중치로 설

정하여 입력변수간의 기하학적 거리를 조절하는 순차실험계획(maximin scaled distance design, MSD)을 제안하였다.<sup>(4)</sup> 즉,  $w_k = \theta_k$  이다. 상관계수는 각 변수의 영향도를 표현하기 때문에 변수에 가중치를 곱하게 되면 변수의 범위가 넓어지는 효과가 있으며, 적은 가중치를 부여 받은 변수보다 축 방향의 충진성이 높아진다. 결과적으로 영향도가 큰 변수의 정보를 더 많이 얻을 수 있으며, 더 적은 표본점으로 정확한 대체모델을 구성할 수 있다.

하지만, 매회 대체모델을 구성하는 순차실험계획 특성상 표본점의 정보가 부족하여 대체모델이 잘못 유도될 경우, 최우량추정과정 중 잘못된 상관계수가 추정될 경우, 혹은 상관계수의 크기, 즉 가중치의 크기가 적절하지 않을 경우 MSD 실험계획은 적절한 표본점을 얻을 수 없다.

3.2 상관계수의 절대적 차와 상대적 차

일반적으로 최우량추정과정에서 상관계수의 범위는  $\theta \in [0.001, 10]^n$  로 알려져 있다.<sup>(6)</sup> 따라서 극단적인 환산거리의 비는 최대 10000배까지 발생 가능하다. 또한 상관계수들의 절대적 차는 작지만 상대적 차가 클 수 있다. 상대적 차가 큰 가중치는 전체 설계 영역의 충진성을 낮추는 결과를 도출한다.

Table 3의 함수 1은 선형함수로서 표본점간 상관도가 크기 때문에 작은 상관계수가 추정된다. Fig. 2는 MSD 실험계획을 이용하여 초기 표본점(8개)으로부터 순차적으로 최종표본점(24개)까지 표본점을 1개씩 추가한 후, 매회 상관계수들을 상자그림으로 나타낸 그림이다. 초기표본점과 최종표본점의 수는  $J_{in}$ 의 실험방법과 같다.<sup>(1)</sup>  $x_1$ 에 대한 상관계수의 평균은 0.0687이고  $x_2$ 에 대한 상관계수의 평균은 0.0039이다. 따라서, MSD 실험계획은 약 17배의 가중치를 적용한다. Fig. 3은 24개의 최종표본점을 나타낸 그림이다.  $x_2$ 의 가중치가 상대적으로 매우 작기 때문에  $x_2$ 의 값의 영향이 매우 적어,  $x_2$ 의 값이 하한과 상한으로 수렴한다. 따라서,  $x_2$ -축 방향의 충진성이  $x_1$ -축 방향의 충진성에 비해 매우 낮다. 그러므로, 상관계수의 절대적 차가 작고 상대적 차가 클 경우 가중치를 비슷한 수준으로 주어야 한다.

3.3 잘못 추정된 상관계수

대체모델 기반의 순차실험 시 매회 대체모델을 구성하고 상관계수를 계산한다. 따라서, 상관계수가 잘못 추정될 경우 부적합한 표본점을 선택하여 결과적으로 대체모델의 정확성과 강건성을 떨어뜨린다.

Table 3의 함수 3의 경우  $x_1$ 은 작은 상관계수를,

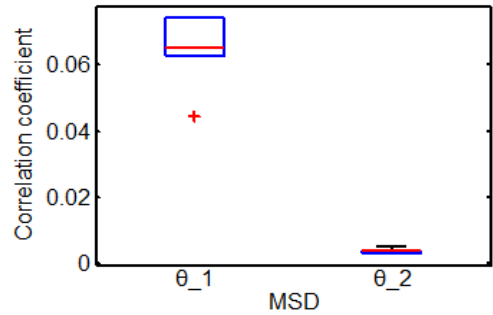


Fig. 2 Box plot of correlation coefficients for function 1

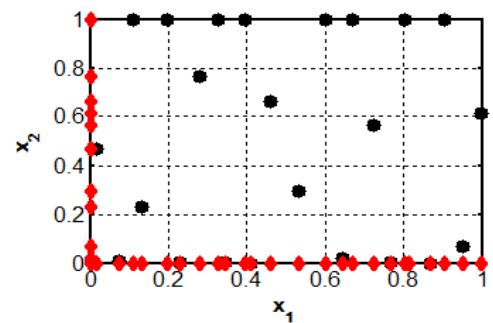


Fig. 3 Sample points for function 1, black dot: sample, red diamond: sample projected to each axis

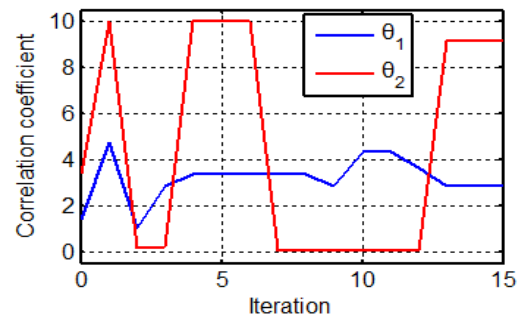
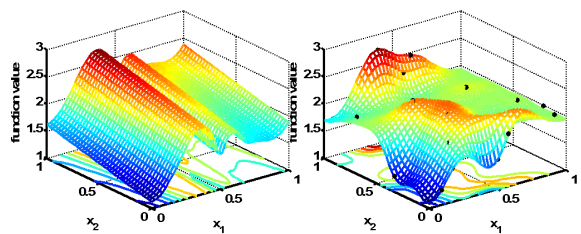


Fig. 4 Trend of correlation coefficient for func. 3



(a) Real model (b) Misled surrogate model  
Fig. 5 3-D plots of real model and kriging surrogate model of function 3

$x_2$ 는 큰 상관계수가 추정되어야 실제함수를 적절히 근사할 수 있다. 함수 3을 3.2절과 같은 방법으로 표본점을 추가하며, 추정된 상관계수를 Fig. 4에 나타내었다. 매회 대체모델을 구성 시  $\theta_2$ 가 7회에서 10, 8회에서 0.068과 같이 크게 변동한다. 8

회와 같이 잘못 추정된 상관계수는 Fig. 5(b)와 같이 크리깅대체모델의 정확성을 떨어뜨리고 부적합한 표본점을 선택한다. 따라서, 상관계수가 잘못 추정되더라도 부적합한 표본점을 선택하는 것을 완화하기 위한 가중함수가 필요하다.

3.4 가중함수 기반 순차최소거리최대화계획

위 절들에서 고찰한 바와 같이 상관계수를 직접 가중치로 사용하는 것은 여러 문제를 야기한다. 본 연구에서는 상관계수의 특징을 유지하며 강건한 대체모델을 구성할 수 있는 가중함수를 제안한다. 가중함수는 상관계수의 함수로써 큰 상관계수가 추정되는, 즉 비선형성이 큰 변수에 큰 가중치를 주는 기존의 특징을 유지하기 위하여 단조증가 함수인 대수함수를 이용하여 정의한다. 가중치의 상대적 차가 커 측방향 충진성을 잃는 것을 방지하기 위하여 최소 1 이상의 가중치를 갖도록 하였으며, 1이하의 상관계수가 대수함수의 값에서 음수로 나타나지 않도록 1만큼 보정한다. 또한, 변수의 개수가 증가함에 따라 가중치가 증가하도록 하여 다변수 최적화문제에서 가중치의 변별력을 향상시킨다. 최적화 수식은 식 (3)과 같고 가중치인  $w_k$ 는 식 (4)와 같이 제안한다.

$$w_k = f(\theta_k) = 1 + \sqrt{n_d} (\log_{10}(\theta_k + 1)) \quad (4)$$

Fig. 6은 2변수(일점쇄선)와 10변수(실선)의 경우 상관계수에 따른 가중함수의 곡선이다. 제안한 가중함수는 가중치의 상대적 차가 작고, 상관계수가 잘못 추정되더라도 가중치 차가 극단적이지 않아 강건성 측면에서 MSD 실험계획에 비해 유리할 것으로 예상된다.

4. 수학 함수 예제

본 장에서는 기존 순차실험계획들과 제안한 가중함수 기반 순차최소거리최대화계획(weight function-based sequential maximin distance design, WFSMD)의 성능을 비교할 수학 함수와 비교 방법을 제시한다. 수학 함수는 Jin 과 같은 Table 3의 함수를 사용하고, 각 함수의 수식은 참고문헌과 같으므로 생략한다.<sup>(4)</sup> 각 함수의 특성은 Cho 의 문헌을 참고하였다.<sup>(6)</sup>

수학 함수 예제를 Table 1의 조건으로 각 순차실험계획을 이용하여 제시된 최종표본점의 개수까지 순차적으로 표본점을 1개씩 추가한다. 순차실험계획의 특성상 초기표본점의 영향을 받기 때문에 최적라틴방격법<sup>(5)</sup>(optimal Latin hypercube design, OLHD)을

Table 1 Sample size for sequential design

Number of variables	Total sample size	Initial sample size	Increment of samples
<6	$12 n_d^*$	$4 n_d$	1
$\geq 6$	$9 n_d$	$3 n_d$	1

\*  $n_d$  is the number of variables

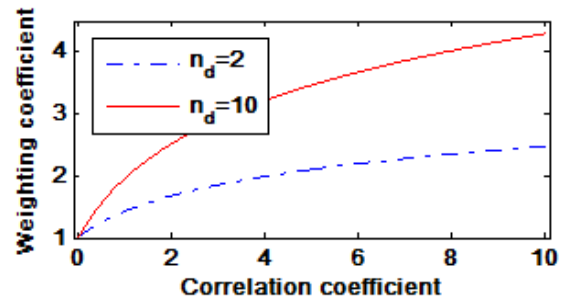


Fig. 6 Trend of weighting coefficients

이용하여 10개의 초기표본점 조합을 만들고 각 순차실험계획을 초기표본점 조합에 따라 10회 반복 실험하여 최종표본점의 개수에서 대체모델의 오차를 구한다. 오차는 식 (5)의 평균제곱근오차(root mean squared error, RMSE)를 사용하며, 검증점( $\mathbf{x}_{val}$ )은 설계범위 내에  $n_t = 50,000$ 개의 임의 추출하여 사용한다. 위 검증방법은 Jin의 방법과 동일하다.<sup>(4)</sup>

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_t} (\hat{y}(\mathbf{x}_{val}^i) - y(\mathbf{x}_{val}^i))^2}{n_t}} \quad (5)$$

여기서  $\hat{y}$  은 대체모델,  $y$  는 실제모델을 의미한다.

5. 결 과

10개의 초기표본점 조합을 이용했기 때문에 상자그림(box plot)을 이용하여 결과를 비교하였다.

함수 1의 Branin 함수는 3.2절과 같이 상관계수의 절대적 차는 작으나 상대적 차가 큰 특징이 있다. 따라서, Fig. 7의 MSD 실험계획과 같이 오차의 변동이 상대적으로 크다. 반면에 제안한 순차실험계획은 오차의 변동이 매우 적은 것을 확인할 수 있다.

함수 2의 Mystery 함수는 대체모델을 구성하면  $x_1$  과  $x_2$  의 상관계수가 1 과 2 사이로 절대적 차와 상대적 차가 유사하게 추정된다. 따라서, 오차의 변동은 기존 순차실험계획들과 제안한 순차실험계획의 차이가 크지 않다.

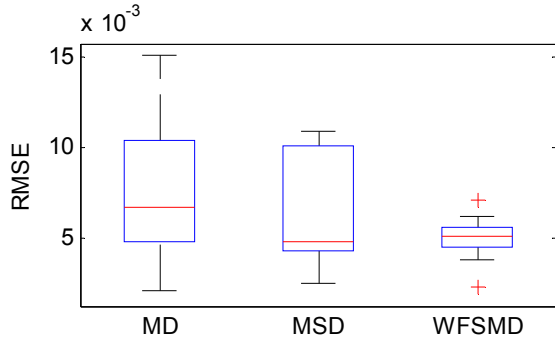


Fig. 7 Errors of surrogate model for function 1

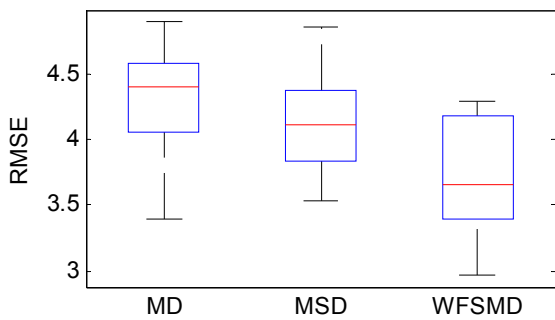


Fig. 8 Errors of surrogate model for function 2

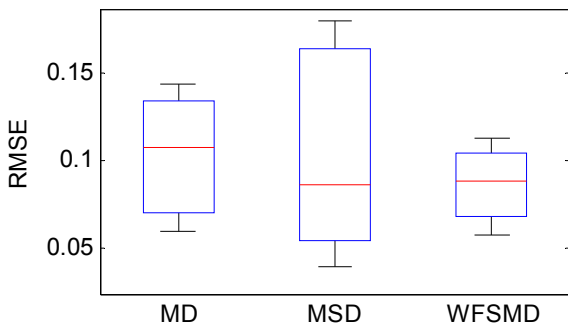


Fig. 9 Errors of surrogate model for function 3

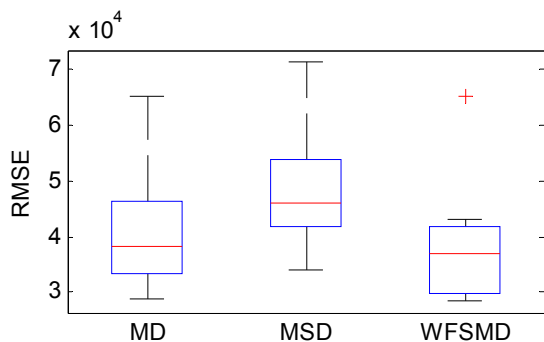


Fig. 10 Errors of surrogate model for function 4

함수 3의 Jin 함수는 3.3 절에서 언급한 것과 같이 상관계수가 잘못 추정되는 경우가 빈번하기 때문에 MSD 실험계획을 이용할 경우 새로운 표본점을 적절히 선택하지 못한다. 따라서 MD 실험계획보다 오차의 변동이 크다.

Table 2 Score for the accuracy and the robustness of each method

	MD	MSD	WFSMD
Accuracy	1	4	10
Robustness	1	6	8

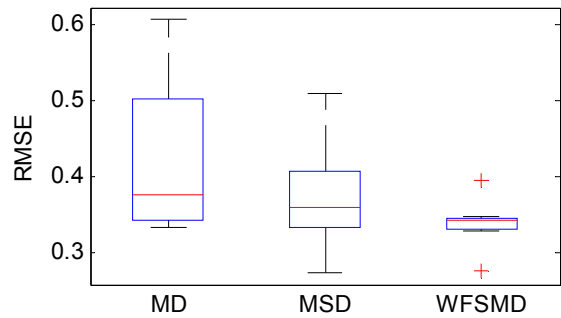


Fig. 11 Errors of surrogate model for function 5

함수 4의 Goldstein-price 함수에서도 제안한 순차실험계획이 유효함을 보였고, 6 변수 함수인 함수 5(Hartman6)에서도 기존 순차실험계획보다 강건한 오차를 제시함으로써 다변수 수학 함수에서도 2 변수 수학 함수와 같이 제안한 순차실험계획이 잘 작용하는 것을 확인하였다.

각 수학 함수마다 정확성과 강건성이 높은 순으로 각 순차실험계획에 2 점, 1 점, 0 점을 부여하였고, 총 점수를 Table 2에 정리하였다. 정확성이 높은 것은 RMSE의 평균값이 낮은 것, 강건성은 RMSE의 편차 값이 낮은 것으로 각각 정의한다. MSD 실험계획은 MD 실험계획에 비해 높은 점수인 정확성 4 점, 강건성 6 점을 받았다. 제안한 순차실험계획은 기존 순차실험계획들과 비교하여 정확성은 10 점으로 5개의 모든 수학 함수에 대해 가장 높으며, 강건성도 3개의 함수에서 2 점, 2개의 함수에서 1 점을 받아 8 점으로 기존 순차실험계획들보다 우수하였다.

## 6. 결론

본 연구에서는 정확한 대체모델을 구성하기 위한 순차실험계획으로써 Jin이 제안한 MSD 실험계획의 문제점을 파악하고, 이를 개선하는 WFSMD 실험계획을 제안하였다. 기존의 MSD 실험계획은 최우량추정과정 중 상관계수가 잘못 추정되거나 상관계수들이 절대적 차는 작으나 상대적 차가 큰 경우에 대해 부적합한 표본점을 선택하여 대체모델의 정확성을 저하시키는 문제점이 발생했다. 이

는 0.001 과 10 사이로 추정하는 상관계수가 직접 가중치로 사용되기 때문이다. 따라서, 정확하고 강건한 대체모델을 구성하기 위해서 상관계수기반의 단조증가함수를 활용하여 새로운 가중함수를 제안하였다. 상관계수의 상대적 차를 줄이기 위해 가중함수에 보정계수를 도입하였다. 제안한 WFSMD 실험계획을 4개의 2변수 수학적 함수와 1개의 6변수 수학적 함수의 이용하여 기존의 순차실험계획들과 비교하였다. 그 결과, WFSMD 실험계획이 기존 순차실험계획에 비해 정확성과 강건성이 가장 우수함을 확인하였고, 제시했던 기존 순차실험계획의 문제점을 해결하였다.

후 기

본 연구는 한국연구재단의 일반연구자 지원사업으로 수행된 “공간-시간 통계적 모형 기반 최적설계 기법 연구(과제번호 2010-0023257)”의 결과 중 일부임을 밝히며, 연구비 지원에 감사 드립니다.

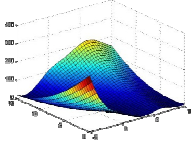
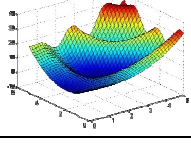
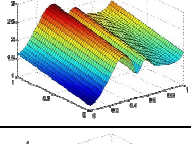
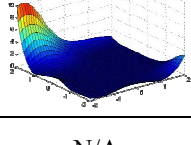
참고문헌  
(References)

(1) Johnson, M., Moore, L. and Ylvisaker, D., 1990, “Minimax and Maximin Distance Designs,” *Journal of Statistical Planning and Inference*, Vol. 26, pp. 131~148.  
 (2) Shewry, M.C. and Wynn, H.P, 1987, “Maximum Entropy Sampling,” *Journal of Applied Statistics*, Vol. 40, No. 4, pp. 165~170.  
 (3) Lee, T.H., and Jung, J.J., 2004, “Maximum Eigenvalue Sampling of Kriging Model,” *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, New York, August, AIAA 2004-4488.  
 (4) Jin, R., Chen, W. and Sudjianto, A., 2002, “On Sequential Sampling for Global Metamodeling for in Engineering Design,” *Proceedings ASME 2002 Design Engineering Technical Conferences and Computer and Information in Engineering Conference*, Montreal, Canada, September, DETC2002/DAC-34092.  
 (5) Crombecq, K., Laermans, E. and Dhaene, T., 2011,

“Efficient Space-Filling and Non-Collapsing Sequential Design Strategies for Simulation-Based Modeling,” *European Journal of Operation Research*, Vol. 214, No. 3, pp. 683~696.  
 (6) Cho, S.K, Byun, H. and Lee, T.H., 2009, “Selection Method of Global Model and Correlation Coefficients for Kriging Metamodel,” *Trans. Korean Soc. Mech. Eng. A*, Vol. 33, No. 8, pp. 813~818.  
 (7) Jang, J., Lee, J.M. and Lee, T.H., 2011, “A Study of C1-Continuity of Split Region Kriging model According to Correlation Functions,” *Proceedings of the KSME(A)*, pp. 174~179.  
 (8) Park, J.S., 1994, “Optimal Latin-Hypercube Designs for Computer Experiments,” *Journal of Statistical Planning and Inference*. Vol. 39, No. 1, pp. 95~111.

부 록

Table 3 Mathematical examples and characteristic

No.	Name	Figure	Characteristic
1	Branin		2 variables Linear
2	Mystery		2 variables Nonlinear
3	Jin		2 variables Moderate
4	Goldstein-price		2 variables Linear
5	Hartman6	N/A	6 variables Nonlinear