

< 논문 >

후보점과 대표점 교차검증에 의한 순차적 실험계획

김승원* · 정재준* · 이태희†

(2006년 6월 29일 접수, 2006년 11월 30일 심사완료)

Candidate Points and Representative Cross-Validation Approach for Sequential Sampling

Seung Won Kim, Jae Jun Jung and Tae Hee Lee

Key Words: Candidate Point(후보점), Cross-Validation Approach(교차검증 실험계획), Kriging Model(크리깅모델), Representative Cross-Validation Approach(대표점 교차검증)

Abstract

Recently simulation model becomes an essential tool for analysis and design of a system but it is often expensive and time consuming as it becomes complicate to achieve reliable results. Therefore, high-fidelity simulation model needs to be replaced by an approximate model, the so-called metamodel. Metamodeling techniques include 3 components of sampling, metamodel and validation. Cross-validation approach has been proposed to provide sequantially new sample point based on cross-validation error but it is very expensive because cross-validation must be evaluated at each stage. To enhance the cross-validation of metamodel, sequential sampling method using candidate points and representative cross-validation is proposed in this paper. The candidate and representative cross-validation approach of sequential sampling is illustrated for two-dimensional domain. To verify the performance of the suggested sampling technique, we compare the accuracy of the metamodels for various mathematical functions with that obtained by conventional sequential sampling strategies such as maximin distance, mean squared error, and maximum entropy sequential samplings. Through this research we learn that the proposed approach is computationally inexpensive and provides good prediction performance.

기호설명

$\hat{y}(\mathbf{x})$: 근사 모델

$y(\mathbf{x})$: 실제 모델

1. 서 론

시뮬레이션모델은 실제 모델을 제작하지 않고 역학적 현상을 수학적으로 모사하거나 시뮬레이션 모델을 이용한 최적설계 분야에서 폭넓게 활용되고 있다. 그러나 비선형 해석모델이나 열유동 해석모델과 같은 오랜 해석시간을 요구하는 경우 시

뮬레이션 모델에 직접적으로 최적설계기법을 적용하는 것은 현실적으로 한계가 있다. 이런 문제를 해결하기 위해 ‘해석모델에 대한 근사모델’인 메타모델을 최적설계과정에 활용하는 연구가 활발히 진행되고 있다.⁽¹⁾ 메타모델링과 관련된 연구는 크게 실험계획, 메타모델, 검증기법 등의 세 가지 연구분야로 나뉜다.

순차적 실험계획은 메타모델링 기법 중에서 실험계획의 효율성과 메타모델의 정확성 향상이라는 두 가지 목적을 동시에 달성하기 위해 개발된 방법이다.⁽²⁾ 순차적 실험계획은 한번에 모든 실험점들을 선택하는 것이 아니라 하나 또는 다수의 실험점들을 특정한 기준에 따라 순차적으로 결정한다. 이러한 순차적 실험계획에는 기존 실험점들과의 최소거리를 최대화하는 최소거리 최대화법, 엔트로피를 가장 크게 하는 실험점들을 순차적으로 추가하는 엔트로피법, 크리깅모델의 예측오차를 이용하는 평균제곱오차방법, 메타모델의 극점에

† 책임저자, 회원, 한양대학교 기계공학부

E-mail : thlee@hanyang.ac.kr

TEL : (02)2220-0449 FAX : (02)2298-4634

* 한양대학교 대학원 기계설계학과

순차적으로 실험점들을 추가하는 민감도법, 교차 검증오차가 가장 큰 위치에 실험점들을 추가하는 교차검증법 등이 있다.

순차적 실험계획 중에서 교차검증법은 순차적 실험계획을 수행하면서 교차검증오차를 이용하여 메타모델의 정확성을 검증할 수 있는 장점이 있다. 하지만 교차검증방법은 실험점을 하나씩만 추가해야 한다는 점과 교차검증오차를 구하는 각 단계마다 실험점 개수만큼의 메타모델을 추가적으로 생성해야 한다는 문제점을 가지고 있다.

본 연구에서는 이러한 문제점을 해결하기 위해 설계영역을 여러 개의 하부 설계영역으로 나누고 각 하부영역의 대표점들을 이용하여 교차검증을 수행하는 대표점 교차검증법을 제안한다. 대표점 교차검증법은 하부영역의 개수만큼 메타모델을 생성하기 때문에 기존의 교차검증법이 가지는 메타모델 구성 시간을 상당히 줄일 수 있다. 또한 실험점들을 결정하는데 필요한 최적화 시간을 줄이기 위해 후보점들 중에서 실험점을 선택하였다. 이러한 후보점과 대표점 교차검증에 의한 순차적 실험계획의 효율성과 정확성을 다양한 수학적 함수에 대해서 살펴보고, 그 결과를 다른 순차적 실험계획법과 비교한다.

2. 크리깅 모델

크리깅 모델에서 실제 응답함수는 평균에 해당하는 전역모델 $\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$ 와 이것으로부터의 편차 (deviation)를 나타낸 $z(\mathbf{x})$ 의 합으로 식 (1)로 표현할 수 있다고 가정한다.

$$y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + z(\mathbf{x}) \quad (1)$$

여기서 $\mathbf{f}(\mathbf{x})$ 는 회귀모델을 구성하는 함수이며, $\boldsymbol{\beta}$ 는 회귀모델의 계수이다. $z(\mathbf{x})$ 는 평균이 0이고 공분산이 아래와 같이 표현되는 확률적인 과정을 통해 식 (2)로 구현된다.

$$\text{Cov}[z(\mathbf{x}^i), z(\mathbf{x}^j)] = \sigma^2 \mathbf{R}(\mathbf{x}^i, \mathbf{x}^j, \boldsymbol{\theta}) \quad (2)$$

여기서 $z(\mathbf{x})$ 의 공분산은 정규분포를 따르는 $y(\mathbf{x})$ 의 분산인 σ^2 과 $\boldsymbol{\theta}$ 를 파라미터로 갖는 상관행렬 $\mathbf{R}(\mathbf{x}^i, \mathbf{x}^j, \boldsymbol{\theta})$ 의 곱으로 나타낼 수 있다. 상관행렬을 구성하는 상관함수 $R(\mathbf{x}^i, \mathbf{x}^j)$ 는 임의의 두 실험점 \mathbf{x}^i 와 \mathbf{x}^j 사이의 상관관계로 식 (3)으로 나타낸다.

$$R(\mathbf{x}^i, \mathbf{x}^j, \boldsymbol{\theta}) = \exp\left[-\sum_{k=1}^{n_d} \theta_k |\mathbf{x}^i - \mathbf{x}^j|^p\right] \quad (3)$$

여기서 n_d 는 설계변수의 개수이고 $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{n_d}]^T$ 는 상관인자(correlation parameter)이다. 한편, 식 (1)의 크리깅모델은 다음과 같이 유도된다.⁽²⁾

$$\hat{y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\hat{\boldsymbol{\beta}}) \quad (4)$$

여기서 $\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{Y}$ 로 표현된다. $\mathbf{r}(\mathbf{x})$ 는 예측점과 해석점들간의 상관관계를 나타내는 상관 벡터로 식 (3)을 이용하여 나타내면 아래와 같다.

$$\mathbf{r}(\mathbf{x}) = [\mathbf{R}(\mathbf{x}, \mathbf{x}^1), \mathbf{R}(\mathbf{x}, \mathbf{x}^2), \dots, \mathbf{R}(\mathbf{x}, \mathbf{x}^n)]^T \quad (5)$$

크리깅모델이 구해지는 과정에서 평균제곱오차 (Mean Squared Error: MSE)는 다음과 같은 식으로 표현된다.

$$MSE = \sigma^2 \left(1 - \left[\mathbf{f}(\mathbf{x})^T \mathbf{r}(\mathbf{x}) \right]^T \begin{bmatrix} \mathbf{0} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}(\mathbf{x}) \\ \mathbf{r}(\mathbf{x}) \end{bmatrix} \right) \quad (6)$$

3. 순차적 실험계획

순차적 실험계획은 설계자가 정의한 기준에 따라 단계적으로 실험점을 추가하면서 메타모델의 정확성을 향상시키는 방법이다. 이러한 방법에는 최소거리최대화법, 엔트로피법, 평균제곱오차법, 민감도법, 교차검증법 등이 있다.

최소거리최대화법은 기존에 l 개의 실험점 집합 \mathbf{X}_p 에서 m 개의 실험점 집합 \mathbf{X}_c 를 추가할 때, $\mathbf{X}_A = \mathbf{X}_p \cup \mathbf{X}_c$ 에 속하는 두 실험점 간의 최소거리를 최대화하는 실험점 집합을 구하는 방법이다.⁽³⁾

$$\text{maximize}_{\mathbf{X}_c} \left[\min_{1 \leq i \leq l+m, 1 \leq j \leq l+m} d(\mathbf{x}_i, \mathbf{x}_j) \right] \quad (7)$$

여기서 $d(\mathbf{x}_i, \mathbf{x}_j)$ 는 $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_A (i, j = 1, \dots, l+m)$ 을 만족하는 두 실험점의 거리를 나타낸다. 이 방법은 기존의 실험점 위치 정보만을 이용하여 실험점을 선택함으로써 모델의 응답특성을 반영하지 못하는 단점이 있다.

엔트로피법은 기존의 실험점 집합 \mathbf{X}_p 와 새로 추가할 실험점 집합 \mathbf{X}_c 로 엔트로피를 정의하고 이것을 최대로 하는 \mathbf{X}_c 를 선택한다.

$$\text{maximize}_{\mathbf{X}_c} |\mathbf{R}_A| \times |\mathbf{J}^T \mathbf{R}_A^{-1} \mathbf{J}| \quad (8)$$

여기서 \mathbf{R}_A 는 $\mathbf{X}_A = \mathbf{X}_p \cup \mathbf{X}_c$ 로 구성되는 상관행렬이고 \mathbf{J} 는 $\mathbf{J}^T = \{1, 1, \dots, 1\}_{1 \times (l+m)}$ 인 벡터이다.

평균제곱오차법은 크리깅이 제공하는 통계 예측 오차인 평균제곱오차를 이용하여 오차가 가장 클 것으로 예상되는 위치에 실험점을 추가함으로써 메타모델의 정확도를 높이는 방법이다.⁽⁶⁾

$$\text{maximize}_{\mathbf{X}_c} MSE \quad (9)$$

민감도법은 순차적으로 메타모델의 극점에 실험점들을 분포시킴으로써 최적해에 대한 메타모델의 예측성능을 높이기 위한 기법이다.⁽⁶⁾

$$\text{select } \frac{d\hat{y}}{dx}(\mathbf{x}_c) = 0 \text{ s.t } |\mathbf{x}_c - \mathbf{x}_{p_j}| \geq d_\epsilon \quad (10)$$

여기서 d_ϵ 은 실험점들이 극점 주위에 중복되어 선택되는 것을 방지하기 위한 거리조건이다.

교차검증법은 순차적 실험계획의 매 단계마다 모든 실험점을 사용한 메타모델 $\hat{y}(\mathbf{x})$ 와 하나의 실험점을 제외한 나머지 실험점들로 구성된 메타모델 $\hat{y}_{-i}(\mathbf{x})$ 를 l 개 생성한다. 이 모델들을 이용하여 임의의 실험점 \mathbf{x} 에 대한 $\hat{y}(\mathbf{x})$ 와 $\hat{y}_{-i}(\mathbf{x})$ 의 교차검증오차를 다음 식과 같이 구한다.

$$e(\mathbf{x}) = \sqrt{\frac{1}{l} \sum_{i=1}^l (\hat{y}_{-i}(\mathbf{x}) - \hat{y}(\mathbf{x}))^2} \quad (11)$$

즉, 교차검증오차가 가장 크면서 기존 실험점들로부터 떨어져 있는 실험점들을 선택하는 방법이다.

$$\max_{\mathbf{x}_c} [e(\mathbf{x}_c) \times \min(d(\mathbf{x}_c, \mathbf{x}_{existing}))] \quad (12)$$

4. 후보점과 대표점 교차검증 실험계획

기존의 교차검증법은 순차적 실험단계에서 모델을 검증할 수 있다는 장점이 있지만 실험점을 하나씩만 추가해야 하고, 교차검증오차를 구하는 각 단계마다 실험점의 개수만큼 크리깅모델을 생성하는 단점이 있다. 이러한 특징은 실험점들의 수가 증가할수록 크리깅모델의 구성에 필요한 시간을 기하급수적으로 증가시키는 문제를 야기시킨다.

본 논문에서는 이러한 문제점을 해결하기 위해 후보점과 대표점 개념을 도입하였다. 후보점은 크리깅모델을 구성하기 위해 선택될 가능성을 가진 모든 실험점 집합이다. 후보점을 이용하면 실험점 선택을 위한 전역 최적화 과정이 불필요하기 때문에 순차적 실험계획의 계산시간을 줄일 수 있으며 각 단계에서 다수의 실험점도 선택이 가능하다. 하지만 후보점 집합을 이루는 실험점 개수가 너무 적으면 크리깅모델의 정확성을 기대하기 어렵고, 너무 실험점들이 많으면 후보점 집합 자체를 얻는 것이 어렵게 된다. 따라서 실험계획을 수행할 총 실험횟수의 3-4배의 실험점을 후보점으로 선택할 것을 제안한다.

본 연구에서는 식 (13)과 같이 후보점들에 대해서 교차검증오차가 가장 큰 실험점들을 차례대로 선택한다.

$$\max_{\mathbf{x}_{candidate}} [e(\mathbf{x}_{candidate}) \times \min(d(\mathbf{x}_{candidate}, \mathbf{x}_{existing}))] \quad (13)$$

$$e(\mathbf{x}_{candidate}) = \sqrt{\frac{1}{l} \sum_{i=1}^l (\hat{y}_{-i}(\mathbf{x}_{candidate}) - \bar{y})^2}$$

여기서 \bar{y} 는 모든 후보점들에 대한 크리깅모델의 평균값을 나타낸다. 식 (13)에서 기존의 교차검증

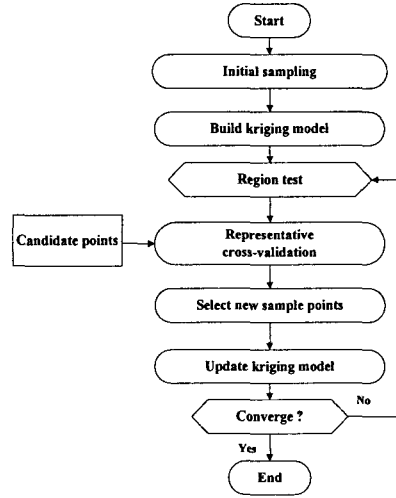


Fig. 1 Procedure of representative cross-validation approach

오차와 다르게 후보점에 대한 응답의 분산을 교차검증오차로 정의하였다.

한편 기존의 교차검증법에서는 실험계획의 매 단계마다 교차검증오차를 구하기 위해 추가적으로 실험점 개수보다 많은 크리깅모델을 구성해야 한다. 즉, 1개의 실험점을 제외하는 교차검증을 한다면 실험점이 증가할수록 각 단계마다 $n+1$ 개의 크리깅모델을 구성하여야 하고, 2개, 3개의 실험점을 제외했을 때는 ${}_n C_2 + 1$, ${}_n C_3 + 1$ 개의 메타모델을 구성해야만 한다. 따라서 이러한 크리깅모델의 구성에 필요한 계산비용을 줄이기 위해 설계영역을 여러 개의 하부설계영역으로 분할하고 하부설계영역에서 대표점을 선택하여 대표점에 대해서만 교차검증을 수행한다.

Fig. 1은 본 연구에서 제안한 후보점과 대표점 교차검증을 이용한 순차적 실험계획의 과정을 나타내었다.

4.1 후보점 선정

본 논문에서는 지수상관함수로 정의한 최대엔트로피 실험계획으로 후보점을 얻었다. 이러한 실험계획은 실험점을 설계변수축에 투영시켰을 때 충족되는 성질을 가지지 않기 때문에 정확한 크리깅모델을 얻을 수 있다.⁽⁷⁾ 실험점 개수가 많을 경우, 설계영역에 고르게 분포된 실험점 집합을 구하는 것 자체가 어려울 수 있다. 이러한 문제의 해결책은 충진성이 좋은 작은 실험점 집합을 구한 뒤, 이것을 설계영역에 복사함으로써 큰 실험점 집합을 구성하는 것이다. Fig. 2(a)는 최대엔트로피 실험계획을 통해 2차원에 대해서 150개의 실험점 집합을 구한 것이고, Fig. 2(b)는 40개의 실험점 집합을 먼저 구한 뒤, 4개의 영역에 평행이동한 160개의

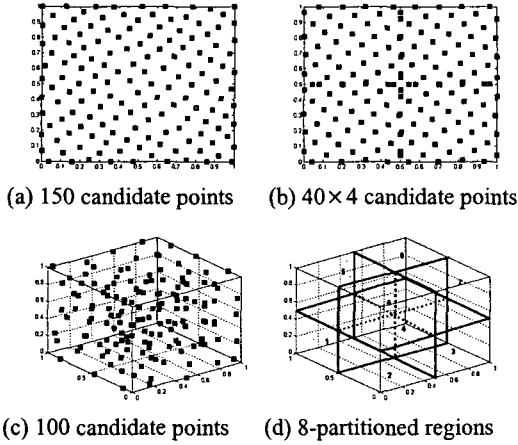


Fig. 2 Illustration of candidate points

후보점 집합이다. 실험점 Fig. 2(c)는 설계변수가 세 개인 경우, 후보점 100개를 구한 것을 나타낸다. 이것을 8개의 분할된 영역에 평행이동하여 후보점 실험점들을 얻을 수 있다. 이러한 방식은 적은 계산비용으로 설계영역을 충진하는 후보점 집합을 쉽게 구할 수 있다.

4.2 대표점 교차검증 실험계획

대표점 교차검증 실험계획은 교차검증을 위해 구성하는 메타모델을 만드는 횟수를 최소화하여 순차적실험계획의 효율성을 높이는데 목적이 있다. 본 논문에서는 2차원 예제에 대해서 설계영역을 2x2, 3x3 으로, 3차원 예제는 2x2x2, 3x2x2, 2x3x2, 2x2x3와 3x3x3으로 분할하여 각 영역에 최소 하나 이상의 실험점들이 분포하고 있는지 검사한 후, 순차적 대표점 교차검증 실험계획을 수행한다.

Table 1은 Branin 함수에 대해서 설계영역의 분할검사 과정을 나타낸 것이다. 4개의 상하한치 실험점(□)과 9개의 실험점(■)에 대해서 3x3 분할검사를 하면 중앙 위쪽의 영역에는 하나의 실험점도 분포하고 있지 않기 때문에 순차적 대표점 교차검증 실험계획을 수행할 수 없다. 이때는 설계영역을 2x2 분할하여 순차적 실험계획을 수행하는 것이 가능하다. 반면 4개의 상하한치 실험점(□)과 10개의 실험점(■)일 경우에는 모든 분할영역에 하나 이상의 실험점이 있기 때문에 설계영역을 3x3으로 분할하여 대표점 교차검증을 수행할 수 있다.

대표점 교차검증 방법은 분할된 각 영역에서 한 점씩만 뽑아서 교차검증을 수행하기 때문에 한 점씩 실험점을 추가할 경우에는 항상 순차적 실험계획의 각 단계마다 9번의 크리깅모델만 추가하면 된다. 이러한 크리깅모델 구성 방식은 실험점의 개수가 많을 때, 실험점 개수만큼 추가적인 크리깅모델을 구성해야하는 기존의 방법의 수치계산시

Table 1 Region test of the representative cross-validation approach

	Before region test	After region test
2x2 Region		
3x3 Region		

간을 줄일 수 있다.

5. 예 제

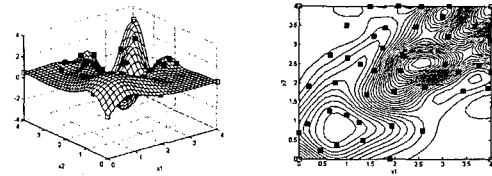
제안한 방법의 우수성을 검증하기 위해 Table 5에 정리한 12 개의 2 차원 수학함수에 대해서 순차적 실험계획을 수행하며, 이 결과를 최소거리최대화법, 최대엔트로피법, 평균제곱오차법 등과 비교한다. 기존의 세 가지 실험계획은 유전자 알고리즘을 이용하여 실험점들을 얻었고, 제안된 방법의 후보점 집합은 Fig. 2의 150 개 실험점을 후보점으로 사용하였다. 모든 실험계획법에 대해서 초기 실험점 4 개에서 순차적으로 45 개씩 실험점을 추가하면서 순차적 실험계획 전과정에서 평균제곱근오차(Root Mean Square Error: RMSE)의 평균을 비교하였다.

$$RMSE_{avg} = \frac{RMSE_{Iter1} + RMSE_{Iter2} + \dots + RMSE_{Itern}}{\text{Number of iteration}} \quad (14)$$

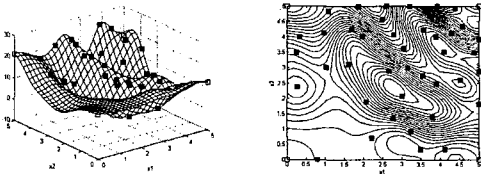
$$RMSE = \sqrt{\left(\sum_{i=1}^n |y_i - \hat{y}_i|^2 \right) / n_v}$$

여기서 평균제곱근오차는 실험점 이외의 $n_v = 900$ 의 격자 배열 실험점에서 응답값을 이용하여 계산하였다.

Fig. 3는 Crane 함수와 Mystery 함수에 대해서 후보점과 대표점 교차검증에 의한 순차적 실험계획을 수행한 결과이다. 본 논문에서는 교차검증오차를 모든 후보점의 응답에 대한 분산값으로 정의하였기 때문에 Fig. 3에서 곡률의 변화가 심한 부분에 새로운 실험점들이 많이 분포하는 것을 알 수 있다. 즉, 등고선 간격이 조밀하거나 응답 변화가 심한 부분에서 교차검증오차가 크게 나타나기 때문에 이러한 영역에 실험점들이 보다 많이 분포하여 비선형성이 강한 Crane 함수와 Mystery 도 정확하게 근사화할 수 있는 것이다.



(a) Crane function



(b) Mystery function

Fig. 3 Kriging metamodels obtained by representative cross-validation approach

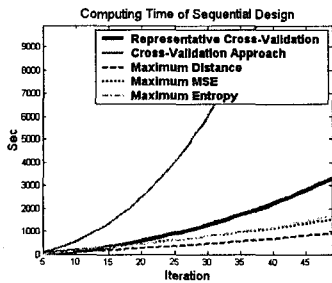


Fig. 4 Computing time of sequential designs

Fig. 4은 기존에 제안된 교차검증법들과 본 논문에서 제안한 실험계획법의 계산시간을 나타낸 그래프이다.

여기서 1.5GHz CPU, 512MB RAM의 Pentium 4 컴퓨터를 사용하였다. 실험계획에 소요되는 계산시간은 크리깅모델 구성과 최적화 과정에 의해 가장 크게 좌우된다.

최소거리최대화법은 크리깅모델을 구성하지 않고 실험점들간의 최소거리만 계산하여 이를 최대화하기 때문에 가장 적은 계산시간이 소요된다.

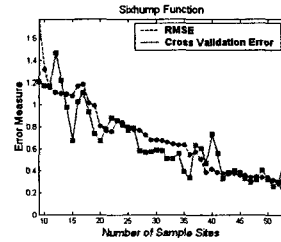
최대엔트로피법은 최적화과정에서 행렬식을 계산해야하고 평균제곱오차는 매 단계마다 한번의 크리깅모델을 생성해야 하기 때문에 최소거리최대화법보다는 다소 긴 계산시간이 요구된다.

그러나 기존의 교차검증방법은 매 실험계획 단계마다 실험점 개수만큼 추가적인 크리깅모델을 구성하고, 교차검증오차에 대한 최적화과정을 거쳐야하므로 실험점들이 증가함에 따라 기하급수적으로 계산시간이 증가하는 것을 볼 수 있다.

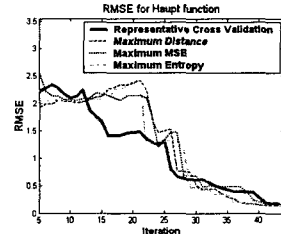
반면, 제안된 방법은 9번의 크리깅모델 구성시간을 드는 대신 후보점 중에서 실험점들을 선택함으로써 실험점이 증가하여도 선형적인 계산 시간

Table 2 Number of cross-validation for sequential design

Number	Cross-validation
Sequential sampling	45
Cross-validation	1170
Representative cross-validation	360



(a) Metamodeling history



(b) Robustness of accuracy

Fig. 5 Characteristics of cross-validation approach

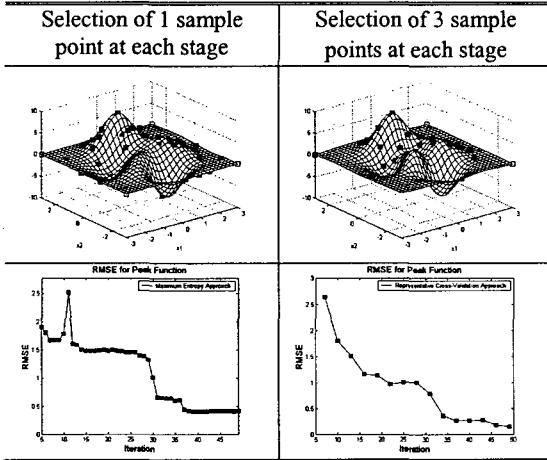
의 증가를 보였다.

Table 2는 순차적 실험계획에서 사용한 메타모델 구성 횟수를 나타낸다. 즉, 제안된 방법은 45개의 실험점들을 결정할 때, 기존의 방법보다 메타모델 구성 횟수를 약 70% 감소시킬 수 있다.

Fig. 5(a)는 식 (13)의 교차검증오차(CVE)와 실험점들을 제외한 900개의 실험점들에 대해서 식 (14)의 평균제곱근오차(RMSE)의 변화를 나타낸 것이다. 제안된 교차검증오차는 크리깅모델에 대한 실제오차인 RMSE와 비슷한 경향을 나타내는 것을 볼 수 있다. 이것은 대표점을 이용한 교차검증오차가 순차적 실험계획에서 모델의 정확성을 평가하는데 이용될 수 있을 뿐만 아니라 실험계획의 종료조건으로 활용될 수 있다는 것을 말해준다.

대표점 교차검증법에서 나타나는 또 하나의 중요한 특징은 메타모델 예측성능의 강건성이다. Fig. 5(b)는 제안한 대표점 교차검증 실험계획과 기존의 세 가지 순차적 실험계획으로 생성한 크리깅모델의 RMSE 변화를 도시한 것이다. Fig. 5(b)에서 대표점 교차검증 실험계획은 실험점을 추가할 때마다 크리깅모델의 정확성이 꾸준히 향상되는 것을 볼 수 있다. 반면 다른 실험계획들은 20개까지 실험점들을 증가시켜도 모델의 오차가 줄어들지 않고, 오히려 증가하는 경향을 보인다. 이러한 결과는 대표점 교차검증 실험계획이 곡률의 변화가

Table 3 Meshed plot and RMSE history for multi-sample points of sequential design for peak function



심한 부분부터 먼저 실험점을 채워나가기 때문에 크리깅모델의 오차가 다른 실험계획보다 빠르게 향상되며 그로 인해 모델의 강건성을 확보할 수 있기 때문이다.

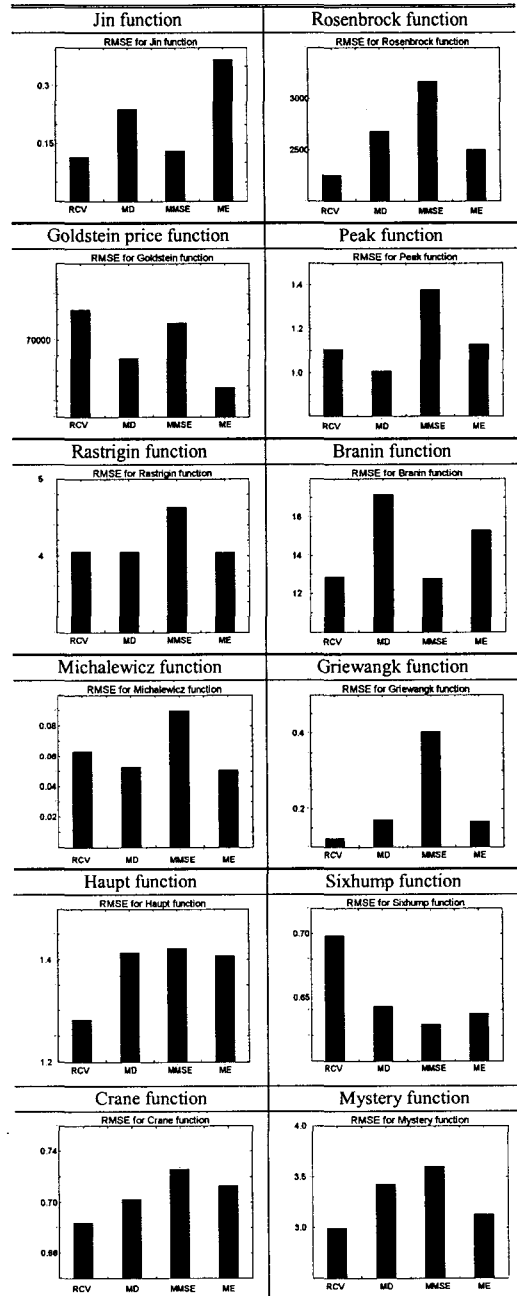
기존의 교차검증 실험계획에서는 각 단계마다 실험점을 오직 하나만 추가할 수 있었다. 하지만 대표점 교차검증 실험계획에서는 실험점을 추가하는데 전역최적화 과정 대신 후보점 중에서 실험점들을 선택하기 때문에 설계자가 정의에 따라 추가할 실험점의 개수를 조절할 수 있다.

Table 3은 Peak함수에 대해서 1개의 실험점을 선택한 결과와 3개의 실험점을 동시에 선택하는 교차검증오차법의 결과를 비교한 것이다. 두 결과는 실험점의 분포는 서로 다르지만, 곡률의 변화가 심한 Peak함수에 대한 정확한 크리깅모델을 생성할 수 있음을 보여준다. 또한 매 단계마다 3개의 실험점을 선택할 경우는 전체 실험계획에서 메타모델 구성 횟수를 1/33배 줄일 수 있으며 평균제곱근오차의 히스토리로 더욱 매끈하게 나타나는 것을 볼 수 있다.

Table 4는 최소거리최대화법(MD), 평균제곱오차법(MMSE), 최댄트로피법(ME), 후보점과 대표점 교차검증법(RCV)으로 비선형성이 강한 12개의 함수에 대해서 순차적실험계획을 수행하고 평균제곱오차의 평균($RMSE_{avg}$)으로 예측성능을 비교한 지표이다.

대표점 교차검증법은 Jin, Rosenbrock, Griewangk, Crane, Haupt, Mystery 함수에 대해서 가장 우수한 성능을 보였고, Peak, branin, Michalewicz 함수에 대해서는 중간 정도의 성능을 나타내었다. 반면, Goldstein와 Sixhump 함수에 대해서만 가장 좋지 않은 성능을 보였다.

Table 4 Comparison of RMSE's average for sequential designs



Goldstein 함수는 경계의 한 점에서 매우 큰 값을 갖고, Sixhump 함수는 6개의 서로 다른 국부최적해 갖기 때문에 이러한 위치에 실험점들이 분포하지 않으면 실제함수에 근접한 크리깅모델을 생성할 수 없다. 따라서 두 함수에 대해서 제안된 실험계획의 성능이 떨어지는 것은 사용된 후보점 집합 중에 이러한 중요한 실험점들이 포함되지 않

왔기 때문이다.

하지만 후보점과 대표점을 이용한 교차검증방법은 다양한 비선형성을 가진 수학적함수 중에서 6개 (50%) 이상에서 우수한 예측성능을 가진 크리깅모델을 생성하는 효과적인 실험계획 방법이라고 볼 수 있다.

6. 결론

본 연구에서는 실험계획 과정에서 함수의 응답 특성을 반영함과 동시에 모델의 정확도를 검증할 수 있는 후보점과 대표점 교차검증 순차적실험계획을 제안하였다. 후보점 중에서 순차적으로 실험점들을 선택하는 방식은 최적화 비용을 줄일 수 있을 뿐만 아니라 다수의 실험점들 동시에 선택함으로써 메타모델 구성 시간을 효과적으로 줄일 수 있다. 이때 후보점은 설계영역을 고르게 충전할 수 있는 지수상관함수를 이용한 최대엔트로피 실험계획으로 구하였다. 기존의 교차검증방법은 실험점들의 개수가 증가할 때 메타모델을 구성하는 계산비용이 기하급수적으로 증가하는 문제점이 있기 때문에 대표점에 대해서 교차검증을 수행하는 방법을 채택하였다.

제안된 교차검증오차는 크리깅모델에 대한 실제 오차인 RMSE와 비슷한 경향을 나타내는 것을 확인하였다. 또한, 제안된 실험계획은 실험점을 추가할 때마다 크리깅모델의 정확성이 꾸준히 향상되어 메타모델의 강건성을 보여주고 있다.

제안된 방법으로 순차적 실험계획을 수행하고 비선형성이 강한 12개의 수학적함수에 대해서 크리깅모델을 구성한 후, 크리깅모델의 정확성을 바탕으로 기존의 순차적 실험계획법의 결과를 비교하였다. 그 결과, 전반적으로 후보점과 대표점 교차검증법이 기존의 방법과 비교할 때, 50% 이상의 테스트함수에 대해서 우수한 예측성능을 보였다.

본 연구로부터 순차적실험계획은 초기 실험점의 영향을 많이 받는 것으로 확인되었다. 따라서 순차적실험계획을 수행할 때 초기 실험점의 선택기준 및 실험점의 개수를 결정할 수 있는 방법의 연구가 요구된다.

후 기

이 연구는 한국과학재단 지정 한양대학교 최적 설계신기술연구센터의 지원으로 수행되었다.

참고문헌

(1) Simpson, T. W., 1998, "A Concept Exploration Method for Product Family Design," Ph. D. Dissertation, Department of Mechanical Engineering, Georgia Institute of Technology.

(2) Jin, R., Chen, W. and Sudjianto, A., 2002, "On Sequential Sampling for Global Metamodeling in Engineering Design," *Proceeding of ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, DETC2002/DAC-32092.

(3) Johnson, M.E., Moore, L. M. and Ylvisaker, D., 1990, "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, Vol. 26, No. 2, pp. 131~148.

(4) Shewry, M.C. and Wynn, H. P. 1987, "Maximum Entropy Sampling," *Journal of Applied Statistics*, 14, pp. 165~170.

(5) Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H.P., 1989, "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, pp. 409~435.

(6) Lee, T.H., Jung, J.J., Lee, C.S. and Hwang, I.K., 2004, "Sensitivity Approach of Sequential Sampling for Krging Model," *Transactions of the KSME A*, Vol.28, No.11, pp. 1760~1767.

(7) Lee, T.H., Kim, S.W. and Jung, J.J., 2006, "Influence of Correlation Functions on Maximum Entropy Experimental Design," *Transaction of the KSME A*, Vol. 30, No. 7, pp. 787~793.

부 록

Table 5 Test functions

Name	Function
Jin	$f = \cos(6x_1 - 3) + 3.1 x_1 - 0.7 + 2(x_1 - 0.5) + \sin\left(\frac{1}{ x_1 - 0.5 + 0.31}\right) + 0.5x_2$ $x_{1,2} \in [0,1]$
Haupt	$f = x_1 \sin(4x_1) + 1.1x_2 \sin(2x_2)$ $x_1, x_2 \in [0,4]$
Branin	$f = \left(x_2 - 5\left(\frac{x_1}{2\pi}\right)^2 + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10$ $x_1 \in [-5,10], x_2 \in [0,15]$
Sixhump Camel back	$f = (4 - 2.1x_1^2 + \frac{x_1^4}{3})x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$ $x_1 \in [-2,2], x_2 \in [-1,1]$
Rosenbrock	$f = (x_2 - x_1^2)^2 + (1 - x_1)^2$ $x_1, x_2 \in [-4,4]$
Gold	$f = (1 + (x_1 + x_2 + x_3)^2 \cdot (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)) \times (30 + (2x_1 - 3x_2)^2 \cdot (18 - 32x_1 + 12x_2^2 + 48x_2 - 36x_1x_2 + 27x_2^2))$ $x_1, x_2 \in [-2,2]$
Mystery	$f = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1)^2 + 2(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2)$ $x_1, x_2 \in [0,5]$
Griewangk	$f = 1 + \sum_{i=1}^{30} \frac{x_i^2}{4000} - \prod_{i=1}^{30} \cos\left(\frac{x_i}{\sqrt{i}}\right)$ $x_1 \in [-2,6], x_2 \in [-2,2]$
Crane	$f = e^{\cos(x_1 - x_2)} \cdot \sin\left(\frac{\cos(x_1 - x_2)^2 + x_1 + x_2}{1 + (x_1 - x_2)^2}\right)$ $x_1, x_2 \in [0,4]$
Peak	$f = 3(1 - x_1)^2 \exp(-x_1^2 - (x_2 + 1)^2) - 10\left(\frac{x_1}{5} - x_1^3 - x_2^2\right) \exp(-x_1^2 - x_2^2) - \frac{1}{3} \exp(-(x_1 + 1)^2 - x_2^2)$ $x_1, x_2 \in [-3,3]$