論文

J. of The Korean Society for Aeronautical and Space Sciences 41(1), 48-53(2013) DOI:http://dx.doi.org/10.5139/JKSAS.2013.41.1.48

유전 알고리즘을 이용한 다항식 반응면 모델의 최적 기저함수 선정

김상진*, 유흥철, 배승호

Optimal Basis Function Selection for Polynomial Response Surface Model Using Genetic Algorithm

Sang-Jin Kim*, Heung-Cheol You and Seung-Ho Bae Agency for Defense Development

ABSTRACT

Polynomial response surface model has been widely used as approximation model which replace physical or numerical experiments in various engineering fields. Generally, low-order model is used to reduce experimental points required to construct the response surfaces, but this approach has limit to represent the highly non-linear phenomena. In this paper, we developed the method to expand modeling capabilities of polynomial response surfaces by increasing order of polynomial and selecting optimum polynomial basis functions. Genetic algorithm is used to choose optimal polynomial basis functions. Developed method was applied to analytic functions with 1 or 2 variables and wind tunnel test data modeling. The results show that this method is applicable to building response surface models for highly non-linear phenomena.

초 록

다항식 반응면 모델은 실제의 물리적, 수치적 실험을 대체하는 근사모델로 여러 공학분 야에서 사용되고 있다. 일반적으로 반응면 구성에 필요한 실험점 수를 줄이기 위하여 낮 은 차수의 다항식을 사용하므로, 심한 비선형성이 동반되는 현상에 대한 모델링에는 한계 가 있다. 본 연구에서는 다항식의 차수를 증가시키는 방법 및 다항식을 구성하는 최적의 기저함수를 선정하는 방법을 통해 다항식 반응면의 모델링 능력을 확장할 수 있는 방법을 개발하였다. 최적 기저함수의 선정에는 유전 알고리즘을 적용하였으며, 1 변수 및 2변수 함수와 풍동시험 데이터에 대한 모델링 사례를 통해 개발된 방법이 비선형성이 심한 현상 을 모델링하는데 적용될 수 있음을 확인하였다.

Key Words : Response Surface Model(반응면 모델), Polynomial(다항식), Genetic Algorithm(유전 알고리즘)

I.서 론

많은 비용을 필요로 하는 물리적, 수치적 실험 결과를 근사 모델로 대체하여 해석 및 설계에 이 용하는 메타모델링 방법이 다양한 공학 분야에서 사용되고 있다[1,2,3,4]. 메타모델링 방법은 물리 적 또는 수치적 실험 결과를 실험에 영향을 주는 요소들로 모델링하고, 실험에 영향을 주는 요소와

* Received: September 7, 2012 Accepted: December 3, 2012

* Corresponding author, E-mail : bipoo@add.re.kr

http://journal.ksas.or.kr/ pISSN 1225-1348 / eISSN 2287-6871 그 결과간의 관계를 파악하기 위한 일종의 통계학 적, 수학적 과정이다. 여기서, 실험을 통하여 도출 된 결과를 반응(Response)이라 하고, 실험에 영향 을 주는 요소를 입력변수 또는 독립변수라 한다.

메타모델링 방법은 제한된 수의 실험을 통해 실험을 대체할 수 있는 반응면 모델(Response Surface Model)을 생성할 수 있으므로, 비용을 획기적으로 줄일 수 있다. 그러나, 실제 현상을 잘 표현할 수 있는 신뢰도 높은 반응면 모델을 생성하기 위해서는 문제에 적합한 실험설계법 (Design of Experiments)과 근사모델의 선택이 필수적이다[5,6]. 근사모델의 경우, 다항식 모델, 크리깅 모델, RBF(Radial Basis Function) 등 다 양한 모델이 개발되어 왔다. 이 중 다항식 반응 면 모델은 생성이 간단하고, 사용하기에 편리하 기 때문에 가장 널리 사용되고 있다. 그러나 반 응면 구성에 필요한 실험점 수를 줄이기 위하여 낮은 차수의 다항식을 사용할 경우 비선형성이 심한 현상을 모델링하기에는 한계가 있으며, 실 험점이 충분하여 고차의 다항식을 사용한다 하더 라도 Runge 현상(Runge's Phenomenon)으로 알 려진 고차 다항식 특유의 진동문제[7]를 처리하는 것은 또 다른 과제라 할 수 있다.

본 연구에서는 이러한 다항식 반응면 모델의 모델링 능력을 확장하기 위하여 다항식의 차수를 증가시키는 방법 및 다항식을 구성하는 최적의 기저함수를 선정하는 방법을 개발하였으며, 비선 형성이 큰 함수에 적용하여 개발된 방법의 효과 를 확인하고자 하였다.

II. 다항식 반응면 모델(Polynomial Response Surface Model)

2.1 다항식 반응면 모델의 구성

다항식 반응면 모델을 사용할 경우 반응과 반 응모델의 관계를 식 (1)과 같이 수학적으로 표현 할 수 있다.

$$y(\mathbf{x}) = y(\mathbf{x}) + \boldsymbol{\epsilon} \tag{1}$$

여기서, $y(\mathbf{x})$ 는 실제 반응을, $\mathbf{x} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{n_1}]$ 는 입력변수를, $\hat{y}(\mathbf{x})$ 는 반응모델을, 그리고 ϵ 은 실제반응과 반응모델간의 오차를 나타낸다. 이때, 설계변수 \mathbf{x} 는 변수의 범위로 정규화 (Normalization)한 변환변수(Coded Variable)의 형태로 사용한다.

일반적으로 실험 설계 이론(Design of Experiment Theory)을 이용하여 실험점을 선택

하고, 선택된 실험점에서 해석을 시행한 후 적합 한 반응모델을 선택하여 구성한다. 구성된 반응 면은 신뢰도를 검사하여 목표로 한 신뢰도를 만 족할 때까지 실험점을 추가 또는 삭제해 가며 다 시 구성하는 절차를 거친다.

반응면을 설계에 이용할 경우 반응면 구성에 필요한 실험점 수를 줄이기 위하여 1차 또는 2차 의 다항식 모델을 사용한다. 그러나, 비선형성이 심한 반응의 경우에는 다항식 모델의 차수를 올 리는 방법을 통해 반응모델을 실제의 현상과 잘 부합하게 만들 수 있다.

2차 다항식으로 선정된 반응 모델을 예로 들면 식 (2)와 같다.

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=2}^k \beta_{ij} x_i x_j \quad (2)$$

여기서, ŷ는 반응모델, x_i는 설계변수, β_i는 다항 식의 미정계수, k는 설계 변수의 개수를 각각 나 타낸다. 식 (2)를 행렬식의 형태로 표현하면 식 (3)과 같다.

$$\widehat{y(\mathbf{x})} = \mathbf{X}\boldsymbol{\beta} \tag{3}$$

n차 다항식의 경우 설계행렬(Design Matrix) X는 $n_s \times n_{rc}$ 의 크기를 가지는 행렬이 된다. 여기 서, n_s 는 실험점 수, n_{rc} 는 식 (4)의 미정계수 (Regression Coefficients)의 수이다.

$$n_{rc} = \frac{(k+1)(k+2)\cdots(k+n)}{n!}$$
(4)

반응모델의 미정계수는 최소제곱법(Least Square Method)을 이용하며, 이 과정을 통해 모 델링 오차를 최소화하는 회귀계수를 구하게 된다.

$$\boldsymbol{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$
(5)

2.2 모델의 신뢰도 예측

반응모델이 구성되면 반응모델이 실제 현상을 잘 표현하고 있는지를 확인하여야 한다. 반응면 의 신뢰도를 평가하는 방법으로서 결정계수 Adjusted R^2 를 사용할 수 있으며 식 (6)과 같다.

$$R_{adj}^{2} = 1 - \frac{SSE/(n_{s} - n_{rc})}{SST/(n_{s} - 1)}$$
(6)

여기서, $SST = \sum_{i=1}^{n_s} (y_i - \overline{y})^2$ 는 반응값 진폭 제곱의 총합을 $\overline{y} = (y_1 + y_2 + \dots + y_n)/n_s$ 는 반응값의 평균 을 $SSE = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ 는 오차제곱의 총합 을 각각 나타낸다.

*R*²_{adj} 은 0과 1사이의 값을 가지며 반응면이 실제 현상에 잘 부합하는 경우에 1에 가까운 값으로 나타나고 잘 맞지 않는 경우에는 0에 가까운값으 로 나타나게 된다. 일반적으로 *R*²_{adj} 이 0.9 이상인 경우에 구성된 반응면이 신뢰할 수 있는 모델이 라고 할 수 있다.

2.3 회귀계수의 불확도 확인

미정계수의 수보다 더 많은 실험을 통해 반응 모델 계수가 결정되므로, 반응모델에서 특정 항 들을 제거하여 반응모델의 신뢰도가 향상될 수 있다. 일반적으로 반응모델의 특정 항들을 제거 하기 위한 지표로써 식 (7)의 t-statistics를 사용 한다.

$$t_{0_i} = \frac{\beta_i}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^{\mathsf{T}} \mathbf{X})_{\mathrm{ii}}^{-1}}} \tag{7}$$

여기서, $i \in i$ 번째 계수를 의미하며, $\hat{\sigma^2}$ 은 반 응값의 분산을, $(\mathbf{X}^T \mathbf{X})_{ii}^{-1} \vdash \beta_i$ 에 대응되는 $(\mathbf{X}^T \mathbf{X})^{-1}$ 행렬의 i행, i열 값을 각각 나타낸다. t-statistics의 값이 큰 계수일수록 회귀계수의 불 확도가 작아, 높은 정확도를 갖는 계수가 된다는 것을 의미한다.

본 연구에서는 최적 기저함수 선정 알고리즘이 정상적으로 동작하는지를 판단하기 위하여 다항 식 반응면 모델 각 항의 불확도를 가늠할 수 있 는 t-statistics를 이용하였다.

Ⅲ. 다항식 반응면 모델의 최적 기저함수 선정

저차의 다항식 반응면 모델은 그 생성방법이 간단해서 근사모델의 구성에 많이 사용된다. 다 항식 반응모델 단독으로도 사용되지만, Kriging 모델의 전역모델로도 많이 사용된다. 그렇지만 1 차 또는 2차의 다항식으로 생성한 근사모델은 그 정확도에 있어서 한계가 있는 경우가 많다. 앞서 기술한 바와 같이 실험점수가 충분하다면 다항식 의 차수를 증가시키는 방법으로 반응면의 신뢰도 를 향상시킬 수 있다. 그러나 고차 다항식 특유 의 진동문제 등을 볼 때 차수가 커진다고 해서 반드시 반응모델의 신뢰도가 증가되는 것은 아니 다. 구성된 다항식 반응면의 각 항들이 전체 모 델의 신뢰도에 기여하는 바를 확인하여 기여하는 바가 작은 항을 전체 모델(Full Model)에서 제거 해야 반응모델의 정확도를 높일 수 있다.

본 연구에서는 유전 알고리즘으로 다항식 모델 의 각 기저함수를 최적으로 조합하여 전체 반응면 모델의 신뢰도를 향상시키는 방법을 개발하였다. 유전 알고리즘을 적용하여 다항식 모델의 최적 항 을 결정하는 방법에 대한 아이디어는 Zhao 등이 크리깅 모델의 구성에 사용한 방법[8]에서 빌어왔 으나, 본 논문에서는 최적항의 선정에 R_{adj}^2 를 적 용하으며, 최적항을 선정하고 난 후 각 계수의 t_i 값과 비교하여 적용된 방법을 검증하였다.

3.1 다항식 반응면 모델의 기저함수

다항식 반응면을 구성하는 기저함수의 개수는 식 (4)의 미정계수의 수와 같다. 설계변수와 실험 점 수로부터 구성할 수 있는 다항식의 최대 차수 는 결정되며, 설계변수와 다항식의 차수로부터 기저함수의 개수를 구한다. 예를 들어 설계변수 의 수가 3개이고 3차 다항식의 경우 미정계수의 수는 각 다항식 차수별로 1, 4, 10, 20, 35, ...로 증가하므로, 실험점이 34개 이하라면, 이 실험점 을 이용해서는 최대 3차식 까지를 구성할 수 있 게 된다. 이해를 돕기 위해 3변수 3차식의 경우 각 기저함수를 Table 1에 나타내었다.

3.2 유전 알고리즘을 이용한 최적 기저함 수 선정

설계변수와 다항식 반응면의 최대 차수를 결정 하고 나면, Table 1과 같이 다항식 반응면을 구 성하는 기저함수를 구할 수 있다. 전체 기저함수 중 어떤 함수로 반응면 모델을 구성할지를 결정 하는 데는 이진문자열(Binary String)을 이용한다. 이진 문자열이란 이진수 기반 유전 알고리즘에서 사용되는 0과 1의 조합으로 이루어진 문자열이 다. 문자열의 길이는 1~n_{rc}이며, 문자열에서 1은 선택한다는 것을, 0은 선택하지 않는다는 것을 각각 의미한다. 2변수 2차 함수를 예로 들자면,

Table 1. Candidate Basis Functions Example (3 Variables, 3rd order polynomial)

차수	기저함수
0	1
1	x_1, x_2, x_3
2	x_1x_1 , x_1x_2 , x_1x_3 , x_2x_2 , x_2x_3 , x_3x_3
3	$\begin{array}{cccccccccccccccccccccccccccccccccccc$

미정계수의 수는 6개이며, 기저함수는 {1, x₁, x₂, x₁x₁, x₁x₂, x₂x₂}로 구성된다. 이에 대응되는 특 정 개체(Population)의 이진문자열이 {1 0 1 0 1 1}라고 한다면 기저함수 중 {1, x₂, x₁x₂, x₂x₂}의 네 개만 선택한다는 것을 의미한다. 기저함수 선택 후 식 (5)를 이용하여 계수를 구하면 이 특정 개체에 대한 다항식 반응면 모델이 구성된 것이다.

최적의 기저함수를 선정하기 위한 목적함수로 는 식(8)과 같이 반응면의 신뢰도를 평가하는 결 정계수 R_{adi}^2 를 이용하였다.

Maximize R_{adi}^2 (8)

다항식 반응면의 최적 기저함수 선정에는 직접 개발한 유전 알고리즘을 이용하였다. 개발된 유 전 알고리즘은 이진코드(Binary-Coded) 형태로 작성되었으며, 선형표준화(Linear Scaling)방법을 사용하여 적합도를 표준화 하였다. 또한, 적합도 비례 선택(Proportionate Selection), 단순교배 (One-cut-point Crossover), 전형적인 변위 (String-wise Mutation) 연산자를 사용하였다. 개체군의 크기(Population Size)를 100, 교배확률 과 돌연변이 확률을 각각 30%, 40%로 설정하였 으며, 최적해의 수렴성 판단에는 적합도와 Off-line 성능을 동시에 고려하였다. 본 연구에서 적용한 돌연변이 확률은 비교적 큰 값이나, 여러 차례의 수치실험으로 수렴성에 영향이 작은 범위 내에서 최적해의 탐색이 용이하도록 설정한 값이

Table 2. Basis Function Selection Algorithm

```
1) 설계변수의 수 및 실험점 수 입력
2)* 다항식 반응면의 최대 차수 계산
3) 최대 차수를 그대로 이용하거나, 사용자의
 직관에 부합하는 차수를 입력
4) 다항식 차수로부터 회귀계수의 수 계산
5) 설계행렬(Design Matrix) 구성
6) 유전 알고리즘을 이용한 기저함수 선정
  초기개체군(Initial Population) 형성 → 적합도
  평가 → 유전 연산자(선택, 교배, 변이)를 이용
  한 개체군 변형 → 적합도 평가 → 수렴할 때
  까지 세대를 증가시켜가며 개체군 변형 반복
7) 분산분석
* 주어진 실험점의 수로부터 최대 차수를 결정하
 는 데는 차례곱((Factorial)을 이용하므로, 프로그
 램으로 구현 가능한 정수형 변수의 최대값을 넘
 지 않도록 최대 차수를 결정한다.
```

다. 또한 최대 세대수(Generation Number)를 1000 으로 설정하였으나, 본 연구에서 적용한 예제의 경우 모두 세대수 100 이내에서 최대 적합도와 Off-line 성능이 10⁵까지 수렴한 값을 찾았다.

유전 알고리즘을 이용한 최적 기저함수 선정 알고리즘을 Table 2에 요약하였다.

Ⅳ 최적 기저함수 선정 알고리즘 적용

4.1 1변수 해석함수

식 (9)의 1변수 함수[6]를 통해 반응면 모델의 다항식 차수의 증가에 따른 모델의 정확도와 최적의 기저함수 조합으로 구성된 반응면 모델 을 비교하였다.

 $y(x) = (6x-2)^2 \sin(12x-4), \ (0 \le x \le 1)$ (9)

Figure 1과 Table 3에 다항식 차수 변화에 따 른 모델을 나타내었다. 그림에서 알 수 있는바 와 같이 등 간격 총 7개의 실험점으로 반응면 을 구성하였으며, 점선이 식(9)의 실제 함수이 고, 실선이 다항식으로 구성한 반응면 모델이 다. 다항식 차수가 증가함에 따라 반응면이 점 차 실제 함수값에 가까워지고 있으며, Table 2 에서와 같이 R^2_{adj} 도 모델의 차수가 증가함에 따 라 점차 1에 근접하고 있다. 특히 최적의 기저 함수를 선정함으로써 모델의 신뢰도가 좀더 향 상 된 것을 확인할 수 있다. 그렇지만 비선형성 이 심한 함수에 대해서는 다항식 반응면 모델 이 함수값을 제대로 표현하지 못하고 있어, 다 항식 반응면의 모델링 한계도 확인할 수 있다.



Table 3.	Change of R_{adj}^2 according	j to		
Order of Polynomials				

다항식 차수	R_{adj}^2
2	0.4264
3	0.6380
5	0.9849
최적 기저함수 조합	0.9913

Table 4. t_0 of Each Term

기저함수	각 항의 t₀ 값
1	0.64338
x	-2.10894
x^2	-4.94111
x^3	0.79598
x^4	8.39962
x^5	0.56048

5차식으로 구성된 반응면 모델과 최적 기저함 수 선정으로 구성된 반응면 모델의 t_0 값을 Table 4에 나타내었다. 5차식 전체모델(Full Model)의 기저함수는 [1, x, x², x³, x⁴, x⁵]인 데 비해 최적 기저함수는 [x, x², x³, x⁴]으로 선정되었다. 여기서, 상수항과 5차항의 t_0 값이 다른 항들에 비해 상대적으로 작음을 알 수 있 다. 각 기저함수와 대응되는 t_0 값이 큰 계수일 수록 높은 정확도를 갖는 계수를 의미하므로, 이는 최적 기저함수 선정 알고리즘이 정상적으 로 작동하고 있다는 것을 의미한다.

4.2 2변수 해석 함수

비선형성이 큰 2변수 함수인 식 (10) 및 Fig. 2의 Crane 함수를 통해 반응면 모델의 최적 기 저함수 선정 효과를 파악하고자 하였다.





Fig. 2. Crane Function and Experimental Points



Fig. 3. Effect of Optimum Basis Function on Polynomial Response Surfaces

반응면 모델을 구성하기 위해 Fig. 2와 같이 각 변수별로 7개씩, 등 간격으로 총 49개의 실 험점을 선정하였다. 2변수, 49개 실험점으로부 터 구성할 수 있는 다항식 모델의 최대 차수는 8차이며, 미정계수의 수는 총 45개이다. Fig. 3 에는 8차 다항식(Full Model)으로 구성한 반응 면과 최적 기저함수를 선정하여 구성한 반응면 을 비교하였다. 8차 다항식의 경우 고차다항식 특유의 진동현상을 보였으나, 최적 기저함수를 선정한 경우 이런 현상이 사라졌다. R_{adj}^2 도 8 차 다항식의 경우 0.7999에서 최적 기저함수를 선정한 경우 0.9904로 향상되었다. 그렇지만 1 차 함수 예제와 마찬가지로 다항식 반응면의 모델링 한계도 확인할 수 있다.

4.3 풍동시험 데이터 모델링

Figure 4에 풍동시험 데이터에 대한 다항식 반응면의 생성결과를 나타내었다. 11차 함수로 구성되었으며, 총 78개의 기저함수 중 최적 기 저함수 선정 알고리즘을 통해 36개의 기저함수 를 선정하였다. R^2_{adj} 은 0.9530이며, 다항식 반응 면임에도 불구하고 비선형성이 심한 데이터를 잘 모델링하고 있음을 알 수 있다.



V.결론

다항식의 차수를 증가시키는 방법 및 다항식을 구성하는 최적의 기저함수를 선정하는 방법을 통 해 다항식 반응면의 모델링 능력을 확장할 수 있 는 방법을 개발하였다. 설계변수와 주어진 실험 점으로부터 생성할 수 있는 다항식 반응면 모델 의 최대 차수를 계산하여 다항식의 기저함수 생 성 알고리즘을 작성하였다. 다항식 반응면의 최 적 기저함수의 선정에는 유전 알고리즘을 적용하 였으며, 1 변수 및 2변수 해석함수와 풍동시험 데이터에 대한 반응면을 구성하여 개발된 방법을 검증하였다.

다항식의 차수가 증가함에 따라, 또는 최적의 기저함수를 선정함에 따라 반응면 모델의 결정계 수인 Adjusted R^2 이 증가하여 모델의 신뢰도가 향상됨을 확인하였다. 또한, 최대의 Adjusted R^2 를 갖는 최적의 기저함수 선정 후, 전체모델의 각 기저함수와 대응되는 t-statistics 값을 비교하 여 최적 기저함수 선정 알고리즘이 정상적으로 작동하고 있다는 것을 확인하였다.

1 변수 및 2변수 해석함수와 풍동시험 데이터 에 대한 모델링 사례를 통해 다항식 반응면의 모 델링 한계를 고려하더라도, 개발된 방법이 비선 형성이 심한 현상을 모델링하는데 적용될 수 있 을 것으로 사료된다.

References

1) Donard R. Jones, "A Taxonomy of Global Optimization Methods Based on Response Surfaces," *Journal of Global Optimization 21*, 2001, pp. 345~383.

2) Forrester, A. I. J, Keane, A. J. and Bressloff, N. W., "Design and Analysis of Noisy Computer Experiments," *AIAA Journal*, Vol. 44, No. 10, 2006, pp. 2331~2339.

3) Jerome Sacks, William J. Welch, Toby J. Mitchell and Henry P. Wynn, "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. $409 \sim 435$.

4) Felipe A. C. Viana, Raphael T. Haftka, Layne T. Watson, 2010, "Why Not Run the Efficient Global Optimization Algorithm with Multiple Surrogates?", AIAA 2010-3090.

5) Raymond H. Myers, Douglas C. Montgomery, *Response Surface Methodology*, John Wiley & Sons, INC., 1995.

6) Alexander Forrester, Andras Sobester , Andy Keane, *Engineering Design via Surrogate Modelling*, John Wiley & Sons, Ltd., 2008.

7)http://en.wikipedia.org/wiki/Runge's_phenomenon

8) Liang Zhao, K. K. Choi, and Ikjin Lee, "Metamodeling Method Using Dynamic Kriging for Design Optimization," *AIAA Journal*, Vol. 49, No. 9, 2011, pp. 2034~2046.