확률론적인 설계영역 설정 기법에 관한 연구

A Study of Statistical Design Space for Fitting Response Surfaces

이동주*, 박정선(한국항공대학교)

1. 서론

반응표면법은 최근 활용도가 높아지고 있는 비도함수 기법들 중 하나이다. 이 기법은 통계적 경험설계와 회귀모델 생성기법을 기반으로 하는 근사 최적화 기법으로, 준 전역성을 가지면서도 반복 해석횟수를 최대한 줄일 수 있는 최적화 알고리즘이다[1]. 반응표면법을 이용한 근사 최적화 기법은 설계변수와 반응간의 수학적 모델을 통계적으로 추정하기 때문에 불가피하게 실제 반응과의 오차를 수반한다. 이러한 오차를 줄이기 위해 실험계획법이나 순차적 근사 최적화 기법과 같은 다양한 이론이 제시되었다[2-3].

본 논문에서는 흥미영역을 결정하는 데 있어 기존의 비통계학적인 방법에서 벗어나 통계이론을 바탕으로 한 흥미영역설정기법을 제안하여 반응표면 생성에 대한 흥미영역의 타당성을 확보하고 이를 통해 최적화 과정의 효율성과 신뢰성을 향상시키고자 한다. 또한 상용 해석프로그램(ANSYS)과 연동하는 프로그램을 개발하여 다분야 통합최적설계를 위한 시스템을 구축하고자 한다.

본 연구를 바탕으로 개발된 통합최적화프로그램을 3부재 트러스 문제에 적용하여 결과비교와 알고리즘의 검증에 이용하였다. 결과를 통해 프로그램의 신뢰성을 확보하였으며 좀 더 향상된 알고리즘의 개발을 위한 기초자료로 활용할 것이다. 이를 통해 구조물의 품질 및 경쟁력의 향상과 개발비용을 절감하는데 기여하고자 한다.

2. 이론적 배경

2.1 순차적 최적화 기법

반응표면법을 이용하는 일반적이 최적화 알고리즘은 근사함수에 대한 오류를 수정하고 신뢰성 높은 최적해를 얻기 위해 적절한 흥미영역을 설정하여 반응표면을 반복적으로 생성한다. 이와 같은 알고리즘의 흐름을 Fig. 1에 나타내었다.

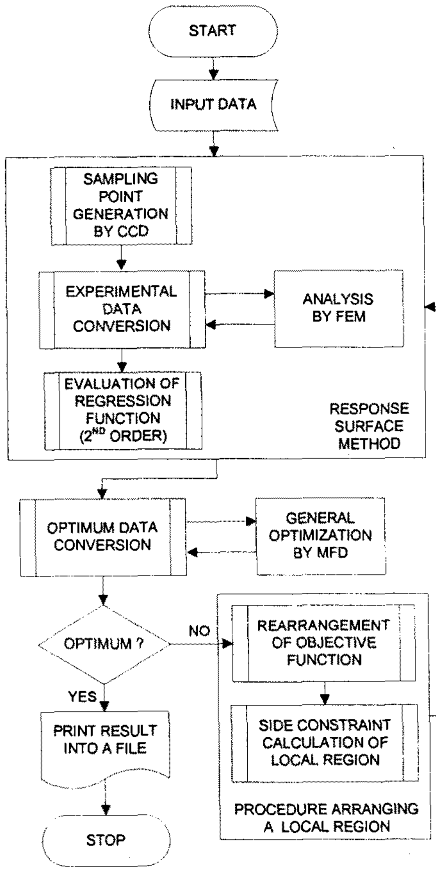

Fig. 1. The flow chart

2.2 반응표면법(response surface method; RSM)

반응표면법은 여러 개의 설계변수가 복합적

인 작용으로 반응변수에 미치는 실험적 관계를 수학식으로 표현하기 위한 통계적 방법이다. 실험이나 시뮬레이션으로부터 얻은 수치 값들로부터 근사적 반응 표면 모형을 만들어 내고 이를 반응표면이라 한다. 반응표면법을 이용한 최적설계는 최적해가 있을 가능성이 높은 흥미영역에서 반응표면을 생성하면서 반복적으로 최적 값을 찾아가도록 구성된다. 반응표면법은 특히 함수관계가 밝혀져 있지 않거나 지배방적식 자체가 복잡하여 일반적이 도함수 기법을 통한 최적설계가 어려울 경우 매우 유용한 방법이라 할 수 있다. 이를 구현하기 위해서는 통계적인 경험설계기법과 회귀모델 생성기법 그리고 일반적인 최적화 기법 등이 요구된다.

2.2.1 반응표면 생성

반응표면 생서 시 근사적 후보함수는 여러 가지 형태를 취할 수 있으며 일반적으로 다항식 모델을 사용한다. 본 논문에서는 식 (1)와 같이 2차 다항식을 이용하여 회귀모델을 구성하였다. 2차 함수는 1차 및 고차 다항식 함수에 비해 매우 유연하여 다양한 함수의 모형을 무리 없이 표현할 수 있으며 회귀계수를 구하기 쉽다.

$$y_i = \beta_0 + \sum_{i \leq j}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j \geq i}^{k} \beta_{ij} x_i x_j \qquad (1)$$

여기서 β_i는 회귀계수이며 k는 설계변수의 개수를 의미한다. 2차 다항식을 구성하기 위해서는 $k! + 2k + 1$개 이상의 실험점이 요구된다.

2.2.2 실험계획법(design of experiments; DOE)

실험점의 선택 시 반응표면법의 효율성 및 신뢰성을 높이기 위해 적절한 실험계획법을 통해 설계 공간(design space)내에서 합리적인 실험점을 추출한다. 반응표면을 생성하기 위한 경험설계기법들은 요인배치법, 중심합성계획법, D-optimal법등 다양하다. 본 논문에서는 알고르즘 구성이 비교적 쉽고 2차 다항식 회귀모델에 적용 가능한 중심합성계획법을 이용하여 실험점을 추출하였다.

중심합성계획법(central composite design; CCD)의 데이터 형식은 설계변수가 두 개일 경우 Table 1과 같다.

Table 1. Central Composite Design; CCD

		X_1				
		1수준	2수준	3수준	4수준	5수준
X_2	1수준			$y(-\alpha,0)$		
	2수준		$y(-1,-1)$		$y(1,-1)$	
	3수준	$y(0,-\alpha)$		$y(0,0)$		$y(0,\alpha)$
	4수준		$y(-1,1)$		$y(1,1)$	
	5수준			$y(\alpha,0)$		

축점 α는 일반적으로 $\sqrt{k}$의 값을 사용하며 실험점의 수는 중심점을 n_c개로 설정할 경우 $2^k + 2k + n_c$개가 된다. 각각의 실험점은 흥미영역을 기준으로 자연좌표계로 좌표 변환된 데이터 값을 갖는다.

2.3 흥미영역 설정기법

전체 설계 공간 내에서 근사 함수의 정확도를 향상시키기 위해서 최적해가 존재할 가능성이 높은 영역을 국부설계영역으로 설정하여 실험점을 추출하며 설정된 국부설계영역을 흥미영역이라 한다. 본 논문에서는 최적해의 신뢰성을 높이고 최적화 과정의 효율성을 향상을 목적으로 적절한 흥미영역을 설정하기 위해 통계학에서 확률의 분포를 수학적으로 나타낸 확률밀도함수(probability density function; PDF)를 응용한 확률론적인 흥미영역 설정기법을 제안한다.

2.3.1 확률분포(probability distribution)

표본공간 내에서 일정한 조건을 모두 만족하는 데이터의 추출 시 그 데이터의 성공확률은 이항확률분포(binomial probability distribution)를 따른다. 어떤 확률변수가 이항분포할 경우 표본이 크고 성공확률이 극히 작지 않으면 평균이 μ이고 분산이 σ^2인 정규분포 $N(\mu, \sigma^2/n)$에 접근하게 된다. 표준화된 확률변수의 표준정규분표 확률밀도함수는 식 (2)와 같으며 누적확률이 $1-\alpha$가 되는 매개변수값 z_α는 식 (3)에서 구할 수 있다.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad (2)$$

$$1-\alpha = P\{Z \le z_\alpha\} = \int_{-\infty}^{z_\alpha} \phi(z)dz \tag{3}$$

만약 정규분포로 가정하기위한 확률변수의 표본이 충분히 크지 못하거나 분포의 정규성을 가정할 수 없을 경우 체비세프 부등식을 이용하여 신뢰구간 추정의 타당성을 보완할 수 있다. 체비세프 부등식은 Z가 유한평균 μ와 유한분산 σ^2을 갖는 확률변수이면, 임의의 $z_\alpha > 0$에 대해 다음이 성립됨을 의미한다.

$$P\{\,|X-\mu|\ \ge z_\alpha\} \le \frac{\sigma^2}{z_\alpha^2} \tag{4}$$

또한 확률표본이 $N(\mu, \sigma^2)$의 정규분포를 갖는다면 분산에 관한 $\sum (Z_i - \overline{Z})^2/\sigma^2$ 값은 자유도가 $n-1$인 χ^2분포를 따른다. 여기서 누적확률이 $1-\alpha$가 되는 $\chi^2(n;\alpha)$의 값은 식 (5)-(6)에서 구할 수 있다.

$$1-\alpha = P\{\chi^2(n) \le \chi^2(n;\alpha)\} \tag{5}$$

$$P(y) = \int_0^y \frac{1}{\Gamma\left(\frac{1}{2}n\right)} e^{-y} y^{(n/2)-1} dy \tag{6}$$

2.3.2 영역설정 이론

본 논문에서 제안하는 영역설정 기법은 실험이나 시뮬레이션으로 얻어지는 실제의 데이터들로부터 통계적 신뢰 구간을 추정하게 된다. 먼저 실험계획법으로 추출된 데이터와 흥미영역내에서 얻어진 최적값을 최적화 문제의 모든 구속조건에 대한 만족여부에 따라 성공여부를 가려낸다. 이때 가려진 데이터 중에 최적 반응을 갖는 데이터를 확률분포의 평균 μ로 가정한다. 이는 흥밍영역내의 최적값이 전역최적해일 확률이 가장 높음을 의미한다. 또한 각 데이터들이 전역최적해가 될 확률이 $N(\mu, \sigma^2)$인 표준정규분포를 따른다고 가정하고 그렇지 않은 경우에 대한 보정을 위해 신뢰구간 설정은 체비세프 부등식을 이용한다.

설계변수는 성공여부는 이항분포를 따르기 때문에 분산을 계산함에 있어 전역해가 될 확률이 적은 데이터들은 제외시켜야 한다. 그렇지 않을 경우 분포함수의 분산이 기대치와 비교해서 크게 나오기 때문에 영역이 너무 크게 설정되게 된다. 그 기준은 최적반응 값에 대한 각 데이터들의 편차에 의해 설정된다. 각 데이터들의 최적반응에 대한 분산이 χ^2분포를 따른다고 가정하고 일정 분산에 대한 한계치를 설정하여 영역 안에 존재하는 데이터만을 취해서 신뢰구간 설정에 이용한다.

이와 같은 두 가지 기준을 통해 추출되는 데이터들은 누적 저장되어 다음 단계의 흥미영역 설정 시 재사용된다.

3. 3-bar 트러스 문제의 최적화

알고리즘의 신뢰성을 검증과 결과의 비교를 위해 Fig. 2에서 보는 바와 같은 3부재 트러스 문제를 최적화하였다.

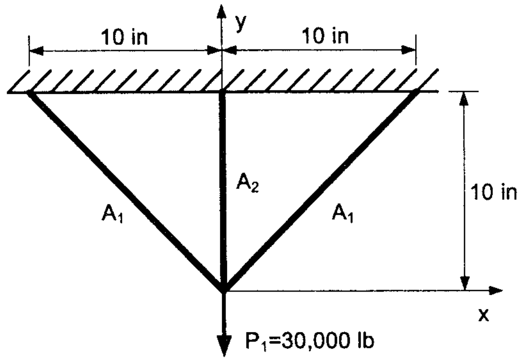

Fig. 2. Three-bar truss structure

재료의 탄성계수와 비중량은 각각 10^7 psi 와 0.1 lb/in^3 이며 설계변수를 부재의 면적 A_1, A_2로 설정하였고 초기값은 두 변수 모두 2.0 in이다. 구조물의 무게를 목적함수로, 응력과 변위를 제한조건으로 설정하였다.

Minimize :

$$OBJ = 2\sqrt{2}\gamma A_1 + \gamma A_2 \tag{7}$$

Subject to :

$$G_1 = \left|\frac{u_t}{0.01}\right| - 1.0 \le 0.0 \tag{8}$$

$$G_2 = \left|\frac{\sigma_{\max}}{20000}\right| - 1.0 \le 0.0 \tag{9}$$

$$0.1 \leq A_i \leq 10.0 \qquad (10)$$

여기서 γ는 부재의 비중량, u_t는 구속이 없는 절점의 총 변위, $\sigma_{\max}$는 최대하중을 의미한다.

4. 결과 및 분석

홍미영역의 설정은 확률에 대한 영역설정의 기준을 몇 %로 정의하느냐에 따라 그 결과가 다르게 나타난다. 따라서 최적화를 반응의 최대한계 값과 체비세프 부등식의 영역한계 값을 각각 [50%,60%,70%,80%]로 설정하여 각각의 최적화 결과를 비교하여 그 결과를 Table 2에 나타내었다.

Table 2. The comparison of the results

		MFD	반응표면법(RSM)			
			50%	60%	70%	80%
목적함수(lb)		9.389	9.408	9.404	9.386	9.406
설계변수 (in^2)	X_1	2.477	2.553	2.527	2.495	2.544
	X_2	2.381	2.187	2.255	2.328	2.210
구속조건	G_1	-	0.004	0.004	0.005	0.004
	G_2	-	-0.649	-0.648	-0.649	-0.648
총 반복횟수		-	7	11	35	38
구조해석 횟수		-	70	110	350	380
함수호출 횟수		-	144	286	710	815

도함수 기법을 이용한 최적화 결과와 비교하면 목적함수 값이 $9.4 in^2$ 정도로 다소 크게 나왔지만 그렇게 큰 차이를 보이지 않는다. 따라서 본 알고리즘은 신뢰할 수 있다. 신뢰구간의 크기 설정에 따른 결과에서 알 수 있듯이 설계변수의 성공확률영역을 넓게 잡을 경우 더 많은 반복횟수를 갖는 것으로 나타났다. 이는 제외되지 않는 데이터 값이 큰 분산값을 가질 확률이 높기 때문으로 분석되며 최적값의 신뢰성 면에서 본다면 70% 영역을 적용하였을 때가 $9.386 in^2$ 으로 도함수기법의 값 $3.389 in^2$ 작으며 다른 결과보다 근접한 결과를 갖는다.

5. 결론

본 연구에서는 최근 많은 연구가 진행중인 반응표면법을 이용한 최적화 프로그램을 작성하여 3-bar 트러스 대한 최적화를 수행하였다. 반응표면은 2차 회귀 모델을 선택하여 생성하였으며, 회귀 상수를 구하기 위한 추출점은 중심합성법을 이용하였다. 최적화 알고리즘의 효율성및 신뢰성 향상을 위해 확률론적인 흥미영역 설정기법을 제안하고 이를 검증하기 위한 최적화를 수행하였다. 최적화 수행 결과 프로그램의 신뢰성을 확인하였으며 확률변수의 한계영역에 따라 효율성과 신뢰성이 결정될 수 있다는 것을 알았다. 따라서 앞으로 한계영역 설정에 설계변수와 목적함수의 변화에 따라 한계영역 설정에 기인할 수 있는 이론에 대한 연구가 이루어져야 한다고 사료된다.

후 기

본 연구는 한국에너지기술연구원 연료전지연구센터의 지원과 과학기술부 21세기 프론티어사업(스마트무인기기술개발사업단)의 지원으로 이루어졌으며, 이에 감사드립니다.

참고문헌

1) Hill, W. and Hunter, W., "A Review of Response Surface Methodology: A Literature Survey," Technometrics, Vol. 8, No. 4, 1996, pp. 571-590.

2) Myers, R., Khuri, A. and Carter, W., "Response Surface Methodology: 1966-1988," Technometrics, Vol. 31, No. 2, 1989, pp. 137-157.

3) Schmit, L. A. and Farshi, B., "Some Approximation Concepts for Structural Synthesis," AIAA Jnl., 12, 5, 1974, pp. 692-699.

4) A. Haldar, S. Mahadevan, *Probability, Reliability, and Statistical Methods in engineering Design*, John Wiley & sons, 2000.