

# 크리깅 메타모델에서 전역 모델에 따른 상관계수의 연구

조수길 † · 이태희 \*

## A study of the correlation coefficients with respect to the degrees of the global models in the kriging metamodel

Su Kil Cho and Tae Hee Lee

**Key Words :** Kriging metamodel (크리깅 메타모델), Design and analysis of computer experiments (전산실험계획), Correlation coefficient (상관계수), Global model (전역모델)

### Abstract

Design analysis and computer experiments (DACE) model is widely used to express efficiently the nonlinear responses in the field of engineering design. Kriging model, a DACE model, can approximately replace a simulation model that is very expensive or highly nonlinear. The kriging model is composed of the summation of a global model and a local model representing deviation from global model. The local model is determined by correlation coefficient of the pre-sampled points, where determination of the correct correlation coefficient has an effect on accuracy and robustness of the kriging model. Therefore, robustness of the correlation coefficient is explored with respect to degrees of the global model. Then we propose the range of correlation coefficient to make correct and robust kriging model and the influence of the correlation coefficients on the degrees of global model with respect to the nonlinearity of the pre-sampled responses.

### 기호설명

$\hat{Y}(x)$  : 크리깅 모델

$\theta$  : 상관계수

### 1. 서론

크리깅 모델은 채광학자인 Krige의 연구를 바탕으로 공간분석 및 공간자료의 연관관계를 규명하기 위해 지질 통계학에서 처음으로 이용되었다.<sup>(1)</sup> 1963년 Matheron에 의해 수학적으로 정립 발전되었고, 1989년 Sacks에 의해 전산실험모델로서 공학 분야에 적용되었으며,<sup>(2)</sup> 설계변수가 많고 비선형성이 강한 모델에 대해서 크리깅 모델의 우수한 예측성능이 확인되었다.<sup>(3)(4)</sup> 이러한 크리깅

모델은 보간모델이라는 특성 때문에 같은 입력에 대해서 동일한 응답을 주는 전산실험에서 활용성이 더욱 중요하게 인식되고 있다.

그러나 크리깅 모델은 보간 모델의 특성상 회기 모델과 달리 실험점을 모두 지나기 때문에 현재의 실험 점을 이용하여 모델의 정확도를 판별하는데 어려움이 있다. 이에 대한 연구로 교차검증방법<sup>(5)</sup>이나 크리깅 모델의 평균과 분산을 이용한 모델의 정확도 평가 법이 있는데<sup>(6)</sup> 크리깅 모델의 정확도에 큰 영향을 주는 실험 점들 간의 상관계수에 관한 연구는 활발히 진행되지 못하고 있다.

따라서 본 논문에서는 크리깅 모델이 상관계수에 따라 정확도가 어떻게 변하는지 살펴본다. 또한 전역 모델의 차수가 모델의 정확도와 상관계수의 강건성에 어떠한 영향을 미치는지 살펴본다. 또한, 함수의 비선형 정도에 따라 정확한 크리깅 모델을 만들기 위한 전역 모델의 차수를 제안하고자 한다.

† 회원, 한양대학교 자동차공학과  
E-mail : chosk83@naver.com  
TEL : (02)2220-0449 FAX : (02)2298-4634

\* 한양대학교 자동차공학과

## 2. 크리깅모델의 상관함수

### 2.1 크리깅 모델

크리깅모델에서는 모델링 할 실제 함수를 평균에 해당하는 전역모델, 편차 항에 해당하는 국부모델(deviation), 그리고 랜덤오차의 합으로 표현할 수 있다고 가정한다.<sup>(7)</sup>

$$Y(x) = f(x)^T \beta + z(x) + \varepsilon_r(x) \quad (1)$$

여기에서  $f(x) = \{f_1(x), f_2(x), \dots, f_p(x)\}^T$  는 설계변수  $x \in R^{n_d}$  로 정의되는 벡터이고,  $\beta = \{\beta_1, \beta_2, \dots, \beta_p\}^T$  는 미지의 계수벡터이다. 전역모델이 상수(constant)라고 가정할 경우는  $p=1$ , 선형(linear)일 경우는  $p = n_d + 1$ , 2 차 다항식일 경우는  $p = (n_d + 1)(n_d + 2) / 2$  를 갖는다. 그리고  $z(x)$  와  $\varepsilon_r(x)$  는 독립적인 정규분포를 갖는다고 가정한다.  $n$  개의 샘플링 점들에서 얻어진 응답벡터  $Y = [y(x^1), y(x^2), \dots, y(x^n)]^T$  을 식(1)에 적용하면 다음과 같은 식이 성립한다.

$$Y = F\beta + Z + \varepsilon_r$$

$$F = \begin{bmatrix} f_1(x^1) & f_2(x^1) & \dots & f_p(x^1) \\ f_1(x^2) & f_2(x^2) & \dots & f_p(x^2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x^n) & f_2(x^n) & \dots & f_p(x^n) \end{bmatrix} \quad (2)$$

$$Z = \begin{bmatrix} z(x^1) \\ z(x^2) \\ \vdots \\ z(x^n) \end{bmatrix}, \quad \varepsilon_r = \begin{bmatrix} \varepsilon_r(x^1) \\ \varepsilon_r(x^2) \\ \vdots \\ \varepsilon_r(x^n) \end{bmatrix}$$

편차와 랜덤오차가 독립적이라는 가정 하에서 두 응답간의 공분산행렬은 다음과 같이 표현할 수 있다.

$$\text{cov}[Y(x^i), Y(x^j)] = \text{cov}[z(x^i), z(x^j)] + \text{cov}[\varepsilon_r(x^i), \varepsilon_r(x^j)] \quad (3)$$

$z(x)$ 의 공분산은 아래의 식과 같이 분산  $\sigma_z^2$  과 상관행렬(correlation matrix)  $R(x^i, x^j)$  의 곱으로 표현할 수 있다.

$$\text{Cov}[z(x^i), z(x^j)] = \sigma_z^2 R(x^i, x^j) \quad (4)$$

상관행렬은 데이터들의 공간적인 상관관계(spatial correlation)를 표현하는 상관함수  $R$  로 정의된다.

$$R = \begin{bmatrix} R(x^1, x^1, \theta, \alpha) & R(x^1, x^2, \theta, \alpha) & \dots & R(x^1, x^n, \theta, \alpha) \\ R(x^2, x^1, \theta, \alpha) & R(x^2, x^2, \theta, \alpha) & \dots & R(x^2, x^n, \theta, \alpha) \\ \vdots & \vdots & \ddots & \vdots \\ R(x^n, x^1, \theta, \alpha) & R(x^n, x^2, \theta, \alpha) & \dots & R(x^n, x^n, \theta, \alpha) \end{bmatrix} \quad (5)$$

여기서 상관함수는 두 샘플링 점  $x^i$  와  $x^j$  의 일차원 상관도의 곱(products of one-dimensional correlations)으로 나타낼 수 있는 상관관계로 한정한다. 일차원 상관도의 곱은 각 차원의 상관관계를 그 차원의 거리의 함수로만 표현하고 이들의 곱으로 상관관계를 표현한다.

상관함수를 지수-가우스(Exponential-Gauss) 상관함수로 정의하면 상관행렬은 아래와 같이 표현된다.

$$R(x^i, x^j, \theta) = \exp \left[ - \sum_{k=1}^{n_d} \theta_k (x^i - x^j)^{\alpha_k} \right] \quad (6)$$

여기서  $\theta, x, \alpha \in R^{n_d}$  이고, 상관행렬은 대각요소가 1 을 가지는 양정행렬 (positive definite matrix) 이다. 상관계수  $\theta_k$  는  $x_k$  방향에 대한 응답 값들의 상관관계를 나타내는 파라미터로써 크리깅 모델의  $x_k$  방향 곡률을 결정한다. 즉,  $\theta_k$  가 크면 상관함수의 정의에 따라 상관도는 작고 이 입력변수 방향에 대한 크리깅 모델의 곡률은 비선형적 특성을 보인다. 반면  $\theta_k$  값이 작으면 상관도는 커지며 이 입력변수 방향에 대해서 크리깅 모델의 예측 값은 선형적인 특성을 보인다. 또한 상관지수  $\alpha_k$  는 예측할 응답의 특성을 반영하는 파라미터로써 모델링 할 응답의 연속성이 좋으면  $\alpha_k = 2$ , 미분 가능하지 않거나 첨도(peakness)가 심한 함수는  $\alpha_k = 1$  로 하는 것이 크리깅의 예측 성능을 좋게 하는 것으로 알려져 있다.

한편, 서로 다른 두 실험 점에서 랜덤오차가 서로 독립적인 상수라고 가정하면 다음과 같이 표현할 수 있다.

$$\text{Cov}[\varepsilon_r(x^i), \varepsilon_r(x^j)] = \sigma_\varepsilon^2 I \quad (7)$$

식 (4), (6), (7)을 식 (3)에 대입하면 아래와 같이 표현할 수 있다.

$$\begin{aligned} \text{cov}[Y(x^i), Y(x^j)] &= \sigma_z^2 \left[ \exp \left( - \sum_{k=1}^{n_d} \theta_k (x_k^i - x_k^j)^{\alpha_k} \right) + \gamma \right] \quad (8) \\ &= \sigma_z^2 V, \quad V \in R^{n \times n} \end{aligned}$$

여기서  $\gamma = \sigma_\varepsilon^2 / \sigma_z^2$  는 응답에 대한 랜덤오차의 분산 비를 나타낸다. 일반적으로 전산실험에서는

시뮬레이션으로부터 얻어지는 응답 값이 결정적이기 때문에  $\gamma = 0$  이 되며, 오차가 존재하는 물리적인 실험에서는  $\gamma > 0$  인 형태를 사용할 수 있다. 하지만 전산실험이라 하더라도 오차가 포함된 응답이나 수치계산 과정에서 수치오차가 존재하는 응답은  $\sigma_\epsilon^2$  항을 포함한 형태가 사용되어야 한다. 이때  $\gamma$  의 크기에 따라 크리깅 모델의 회귀특성이 결정되는데  $\gamma = 0$  이면, 보간특성을 갖는 크리깅 모델이 된다.

## 2.2 최우량 추정 과정

식 (6)에서 정의된 모수  $\theta_k, \alpha_k, \gamma$  와 편차의 분산  $\sigma_\epsilon^2$  를 구해야만 구현이 가능하다. 편차의 분산은 일반최소제곱법을 이용하여 개의 데이터들로부터 추정된 분산으로 대체할 수 있다.

$$\hat{\sigma}^2 = \frac{(Y - F\hat{\beta})^T V^{-1} (Y - F\hat{\beta})}{n} \quad (8)$$

식 (8)의 크리깅 모델에서 남은 미지수는  $\theta_k, \alpha_k, \gamma$  이고, 이것은 최우량추정법(maximum likelihood estimation : MLE)에 의해서 구한다.

크리깅 모델의 모수를 결정하는 우량함수(likelihood function)는 다음과 같이 정의된다.

$$L \equiv \frac{1}{(2\pi)^{n/2} \sqrt{|\hat{\sigma}_\epsilon^2 V|}} \exp\left(-\frac{(Y - F\hat{\beta})^T V^{-1} (Y - F\hat{\beta})}{2\hat{\sigma}_\epsilon^2}\right) \quad (9)$$

우량함수를 최대를 만드는  $\theta_k, \alpha_k, \gamma$  을 구한다는 것은 확률적으로 가능성이 높은  $\theta_k, \alpha_k, \gamma$  값을 추정한다는 의미이다.

일반적으로 우량함수는 다중극부 점을 갖고 해석적인 방법으로 구할 수 없는 것으로 알려져 있기 때문에 최우량추정 과정에서는 수치적인 최적화가 요구된다. 식 (9)의 우량함수에 로그를 취하고 상수 항을 제외하고 정리하면 다음과 같다.

$$\text{Maximize} -\left(\frac{n \ln \hat{\sigma}_\epsilon^2 + \ln |V|}{2}\right) \quad (10)$$

그런데 최적화 과정에서 특이행렬(singular matrix)이 될 경우,  $\ln |V|$  의 수치계산이 불가능해지기 때문에 식(11)와 등가의 최우량 추정과정을 다음과 같이 표현할 수 있다.

$$\text{Minimize } \hat{\sigma}_\epsilon^2 |V|^{1/n}$$

$$\text{Subject to } \theta_k > 0, k=1,2,\dots,n_d \quad (11)$$

$$0 \leq \alpha_k \leq 2, k=1,2,\dots,n_d$$

$$0 \leq \gamma \leq 1$$

최우량추정과정에서 보간특성을 가진 크리깅모델(interpolating kriging)을 얻고자 할 경우는  $\gamma$  를 최적화 변수에서 제외하고 0 으로 고정한다. 반면, 평활화 크리깅모델(smoothing kriging)을 생성해야 할 경우는 최우량추정법으로 추정해야 한다.

## 2.3 Root Mean Squared Error (RMSE)

근사 모델은 실제 모델을 몇 개의 실험 점으로 근사하기 때문에 실제 모델과 오차가 존재한다. 특히 오차가 양수와 음수 모두 존재할 때 산술평균의 개념이 아닌 제곱 평균의 개념으로 오차를 평가한다. 이를 식으로 나타내면 식 (12)와 같다.

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (Y(x_i) - \hat{Y}(x_i))^2}}{n} \quad (12)$$

## 3. 전역모델 차수와 비선형함수

총 9 개의 함수에 대하여 함수의 비선형 정도를 2 차 RSM 모델의 R-square 값으로 판단하고 함수를 Nonlinear, moderate, linear 로 분류하였다. RSM 모델은 Full factorial Sampling 을 이용하여 100 개의 실험 점을 뽑아 R-square 값을 구하였다. Table 2 은 함수의 R-square 값에 따라 세 단계로 나눈 표이다.

Table 2 에서 분류된 함수에  $\theta$  를 0.001 에서 100 까지 Log scale 로 변화시키면서 크리깅 모델의 RMSE 값을 살펴보았다. 이때 상관계수는 변수의 개수에 상관없이 하나로 가정하였다. Figure 2 은  $\theta$  와 RMSE 의 상관관계를 그래프로 나타낸 것이다. 전역모델은 0 차, 1 차, 2 차로 변화시켜 비교하였고, 샘플링 기법은 Full factorial sampling 으로 49 개의 실험 점을 이용해 크리깅 모델을 구성하였다. 오차의 비교는 같은 샘플링으로 100 개를 판별하여 그에 대한 Root mean squared error (RMSE) 값으로 비교하였다.

Table 1 Classification of the functions

Grade	R-square	Function
Nonlinear	0.4 이하	Peak, Crane, Mystery
Moderate	0.4~0.8	Jin, Branin, Schwefel
Linear	0.8 이상	Michalewicz, Rosenbrock, Linear

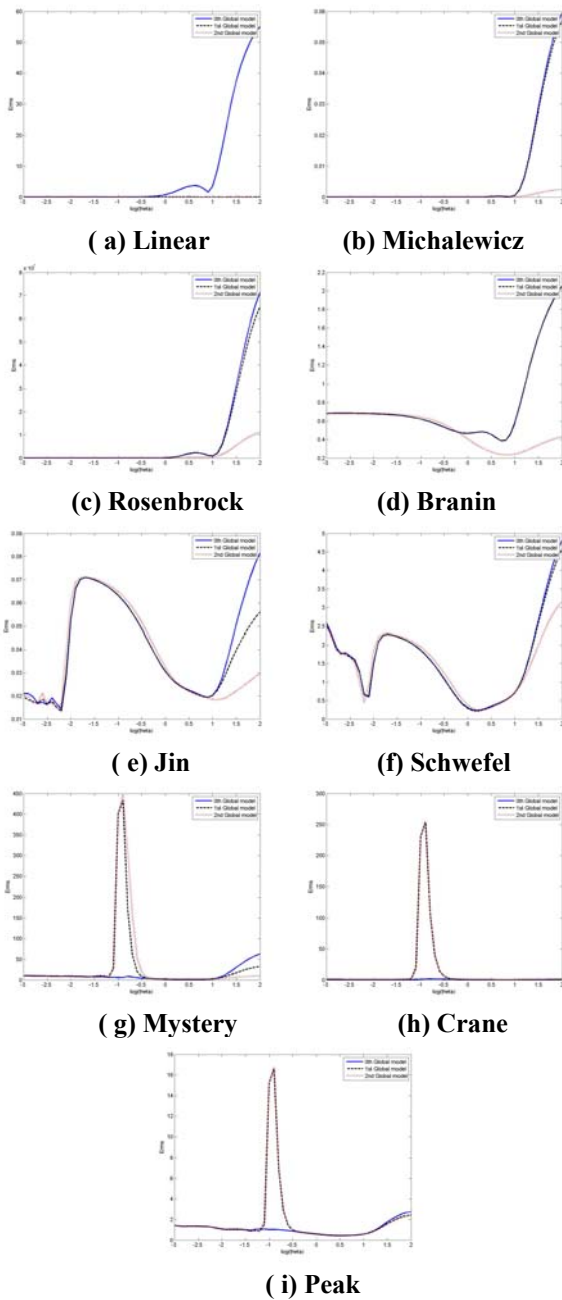


Fig. 1 RMSE as the correlation coefficient

Figure 2 을 살펴보면 Linear 등급의 함수는 2 차 전역 모델을 사용한 크리깅의 RMSE 값이 다른 전역 모델에 비해 더 정확하며  $\theta$  의 변화에도 강건한 것을 볼 수 있다. 하지만 Moderate 등급의 함수는 대체적으로 전역 모델에 상관없이 비슷한 오차 형태를 그리며 전역 모델의 차수보다는  $\theta$  의 변화에 민감한 것을 알 수 있다. 또한 Nonlinear 등급의 함수는 대체적으로  $[-1.5, -0.5]$  (Log scale) 부분의  $\theta$  을 제외한 나머지 영역에서 RMSE 값이 비슷하게 유지되는 것을 알 수 있다. 하지만  $[-1.5,$

$-0.5]$  부분의 영역에 대해서 크리깅 모델이 부정확해지는 모습이 발생하였다. 특히 2 차 전역모델을 사용한 크리깅 모델이 다른 영역의  $\theta$  값에 비해 부정확해지는 것을 확인할 수 있다.

Figure 3 는 Nonlinear 등급인 Mystery function 에 대하여  $\theta = -1.8, -0.8, 0, 1$  (Log scale) 인 크리깅 모델의 그래프를 그려보았다.

Figure 3 에서 확인 할 수 있듯이 매우 비선형적인 함수에 대해 특정영역의  $\theta$  값에 모델이 매우 부정확해지는 것을 확인할 수 있다.

이러한 오차의 발생이 Sampling 기법에 문제가 있는지 확인하기 위하여 실험 점의 중첩이 일어나지 않는 Latin hypercube sampling 을 이용하였다. 실험 점의 개수는 동일하고 비선형성이 매우 강한 Crane function 에 대하여  $\theta$  와 RMSE 값으로 Figure 4 과 같은 그래프를 그려보았다.

Latin hypercube sampling 역시 그래프의 형태가  $[-1.5, -0.5]$  (Log scale) 영역에서 RMSE 가 크게 증가하는 것을 볼 수 있다. 이러한 특징은 Nonlinear 등급에 있는 모든 함수가 같은 현상이 나타났으며 같은  $\theta$  영역에 대해 위와 같은 현상이 나타났다.

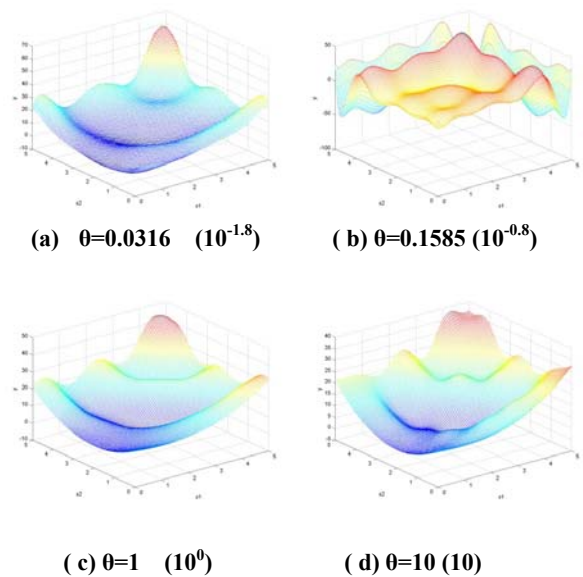


Fig. 2 Kriging model as the correlation coefficient

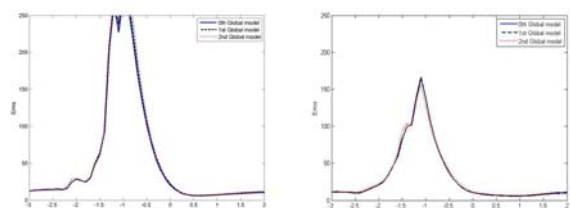


Fig. 3 RMSE of the kriging model using Latin hypercube sampling

#### 4. 결론

함수의 비선형 정도와 크리깅 모델의 전역 모델 차수, 그리고 상관계수  $\theta$  에 따라 크리깅 모델의 정확도가 어떻게 변하는 지 살펴 보았다. 그 결과 선형적인 함수에 대해서는 전역 모델을 2 차 모델로 쓰는 것이 정확도가 높은 모델을 만들 수 있으며 0.001 부터 100 까지  $\theta$  범위에서 오차가 크게 늘어나지 않음을 확인하였다. 그리고 비선형성이 중간 정도인 함수에 대해서는 전역 모델의 차수 보다는  $\theta$  값에 민감하며 특정  $\theta$  값에서 좋은 정확도를 보임을 확인 할 수 있었다. 대체적으로 [0.1 10](True scale)영역에서 최적 점이 나타남을 확인하였고, 그 영역을  $\theta$  의 범위로 놓고  $\theta$  의 최적화를 수행하는 것을 제안한다. 마지막으로 비선형이 큰 모델에 대해서는 특정 영역 [-1.5, -0.5] (Log scale)에서 모델이 부정확해지는 현상을 확인하였다. 본 연구에서는 이 현상의 원인을 규명하지 못하였고 앞으로 이에 대한 연구가 수행되어야 할 것이다.

#### 참고문헌

- (1) Matheron, G., 1963, "Principles of Geostatistics," *Economic Geology*, Vol. 58, pp.1246~1266.
- (2) 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.
- (3) Simpson, T.W., Mauery, T.M., Korte, J.J. and Mistree, F., 2001, "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, pp. 2234~2241.
- (4) Lee, T.H., Lee, C.J. and Lee, K.K., 2003, "Shape Optimization of a CRT based on Response Surface and kriging Metamodels," *Trans. Of KSME (A)*, Vol. 27, No. 30, pp. 381~386.
- (5) Mitchell, T.J. and Morris, M.D., 1992, "Baysesin Design and Analysis of Computer Experiments: Two Examples," *Statistica Sinica*, Vol. 2, pp. 359~379.
- (6) Byun, H.S., Jung, J.J. and Lee, T.H., 2007, "Validation Technique of Kriging Model Using Integrated Mean Squared Errors and Responses," *7<sup>th</sup> World Congress on Structural Multidisciplinary Optimization*, pp. 18~23.

- (7) Sacks, J., Schiller, S.B. and Welch, W.J., 1989, "Designs for Computer Experiment," *Technometrics*, Vol. 31, No. 1, pp.41~47.
- (8) Lee, T.H. and Jung, J.J., 2005, "Generalized Kriging Model for Interpolation and Regression," *Trans. Of KSME (A)*, Vol. 29, No. 2, pp. 277~283.
- (9) Lee, T.H., Jung, J.J., Lee, J.M. and Byun, H.S., 2006, "Development of Approximate Technique to Improve Efficiency of Reliability Robust Design Methodology," Korea Automotive Technology Institute, pp.5~18

#### 부록

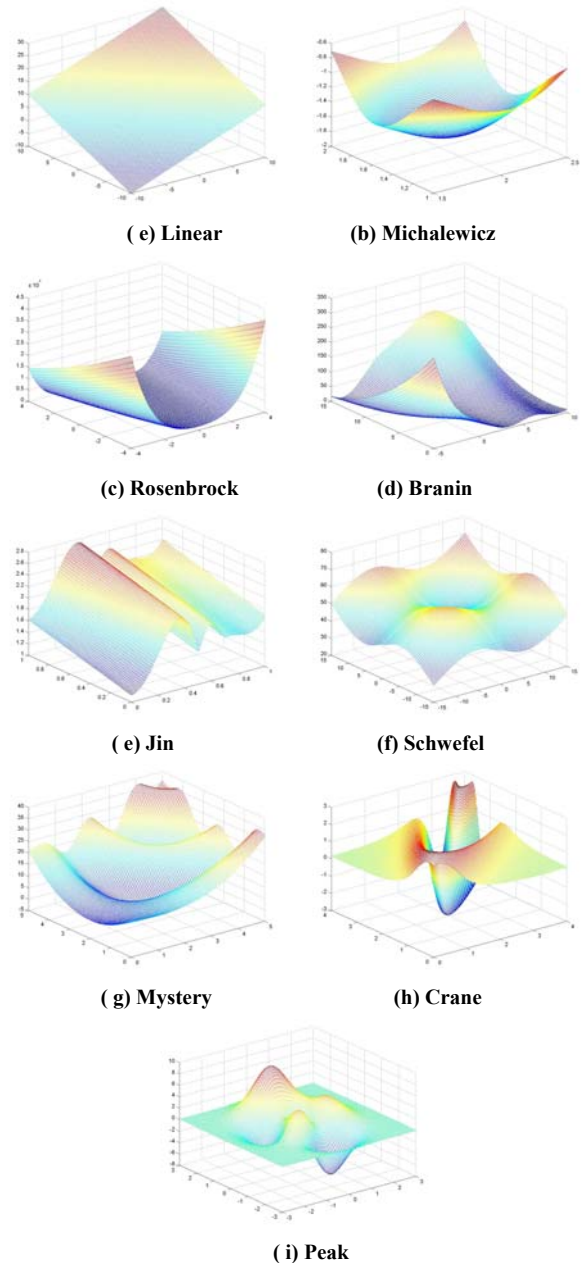


Fig. 4 Test Functions