Combined Kriging and Gradient-Based Optimization Method

Masato Sekishiro*
Tokyo Institute of Technology, Dept. of Mech. Sci. and Eng., 2-12-1-I1-58 O-okayama, Meguro-ku, Tokyo, 152-8552 Japan

Gerhard Venter† and Vladimir Balabanov‡
Vanderplaats Research & Development, Inc., 1767 South 8th Street, Suite 100, Colorado Springs, CO 80906

This paper presents a new Kriging-based optimization method. The goal of this research is to develop a practical and robust general-purpose Kriging-based optimization tool for general design problems. The proposed optimization method efficiently combines Kriging approximations with a gradient-based optimizer. The proposed method is applied to several test problems to examine its efficiency and versatility.

Nomenclature

$\mathbf{x}$	=	vector denoting position in the design space
n_s	=	number of sample points for Kriging model
$y(\cdot)$	=	function response
$\hat{y}(\cdot)$	=	estimated value of $y(\cdot)$
$\mathbf{R}$	=	correlation matrix for Kriging model
θ_k	=	correlation parameter for Kriging model
$\hat{\mu}$	=	estimated global constant for a Kriging model
$\hat{\sigma}^2$	=	estimated variance for a Kriging model
s	=	root mean squared error of $\hat{y}(\cdot)$
$E[I(\cdot)]$	=	expected improvement
$g_i(\cdot)$	=	i-th constraint function
$V_{\max}(\cdot)$	=	maximum violation of the constraints

I. Introduction

KRIGING models are increasingly used as response surface models to approximate computationally expensive functions in engineering field. Various implementations of Kriging models have been investigated.[1-3] Kriging models are most suitable for approximating response values obtained from deterministic experiments, for example computer simulations, since these models interpolate the response values at the observed sample points. The main advantage of Kriging models over polynomial models is the ability of the Kriging models to approximate multimodal and highly nonlinear functions. However, it is still possible that a Kriging model does not capture the actual optimum when used to explore the design space. For example, this may happen when none of the sample points are located in the region of the optimum. For a robust exploration of the design space, both the predicted value and the uncertainty of the Kriging model should be taken into account.

Various implementations of Kriging-based optimization methods have been studied, e.g., Efficient Global Optimization (EGO), developed by Jones et al.[4-6] Efficient Global Optimization exploits an important advantage of Kriging models: both the predicted value and the associated modeling uncertainty is available at any point in the design space. The underlying Kriging approximation is dynamically updated using a sampling criterion that accounts for both the predicted value and the uncertainty. This sampling criterion defines where the next sample

---

* Graduate Student, Student Member AIAA.
† VisualDOC Project Manager, Senior Member AIAA.
‡ Senior Research and Development Engineer, Senior Member AIAA.

point should be placed in the design space. For each iteration of the optimization process, the actual response values are evaluated at the new sample point and the existing Kriging model is updated, thus efficiently refining the Kriging model. The current work is based on the basic ideas of EGO and proposes the combination of an EGO type optimizer with a gradient-based optimizer to improve the convergence properties of the optimization process. The proposed optimization method is applied to several test problems to examine its efficiency and versatility.

II. Overview of Kriging-based Optimization Method

A. Kriging model

Kriging was initially developed in the field of geostatistics and can be regarded as an interpolation model since it interpolates responses at all sample data points. Kriging models provide both an estimated response value and the associated uncertainty of the estimated value at any point in the design space. One of the ways to represent a Kriging model that estimates an unknown response function of interest $y(\mathbf{x})$ is as follows:

$$\hat{y}(\mathbf{x}) = \mu + Z(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is an m-dimensional vector (m is the number of design variables). The model has two parts as shown in Eq. (1). The first is a global constant μ, the second is a realization of a stochastic process with zero mean, $Z(\mathbf{x})$. The $Z(\mathbf{x})$ term represents a local deviation from the global model, calculated by quantifying the correlation of $\mathbf{x}$ with nearby points. The covariance matrix of $Z(\mathbf{x})$ is given by Eq. (2).

$$Cov\left[Z(\mathbf{x}^i), Z(\mathbf{x}^j)\right] = \sigma^2 \mathbf{R}\left[R(\mathbf{x}^i, \mathbf{x}^j)\right] \tag{2}$$

In Eq. (2), $\mathbf{R}$ is the correlation matrix, and $R(\mathbf{x}^i, \mathbf{x}^j)$ is the Gaussian correlation function between any two of the n_s sample data points $\mathbf{x}^i$ and $\mathbf{x}^j$. By using a specially weighted distance, the Gaussian correlation function is defined as follows:

$$R(\mathbf{x}^i, \mathbf{x}^j) = \exp\left[-\sum_{k=1}^{m} \theta_k \left|x_k^i - x_k^j\right|^2\right], \tag{3}$$

where θ_k ($\theta_k \geq 0$) are the unknown correlation parameters used to fit the model, and x_k^i and x_k^j are the k-th components of vectors $\mathbf{x}^i$ and $\mathbf{x}^j$. This function controls the influence of the nearby points and the smoothness of the resulting model. The Kriging predictor for the values of $\mathbf{x}$ is obtained from

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{4}$$

where $\hat{\mu}$ is the estimated value of μ, and $\mathbf{y}$ is a column vector of length n_s that contains the sample values of the response, $\mathbf{r}$ is the correlation vector of length n_s between an untried $\mathbf{x}$ and the sample data points.

$$\mathbf{r}(\mathbf{x}) = \left[R(\mathbf{x}, \mathbf{x}^1), R(\mathbf{x}, \mathbf{x}^2), \cdots, R(\mathbf{x}, \mathbf{x}^{n_s})\right]^T \tag{5}$$

For any given vector $\boldsymbol{\theta}$ that consists of components θ_k, $\hat{\mu}$ and the estimate of the variance $\hat{\sigma}^2$ can be defined as

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} \tag{6}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu})}{n_s} \tag{7}$$

The unknown correlation parameters $\boldsymbol{\theta}$ of the Kriging model are estimated by maximizing the following likelihood function.[2]

$$Ln(\boldsymbol{\theta}) = -\left[n_s \ln\left(\hat{\sigma}^2\right) + \ln|\mathbf{R}|\right]/2 \tag{8}$$

where both $\hat{\sigma}^2$ and $|\mathbf{R}|$ are functions of only $\boldsymbol{\theta}$. Maximizing the likelihood function is an m-dimensional nonlinear optimization problem. Since the likelihood function is multimodal and evaluation of the likelihood function requires calculation of the determinant of the matrix $\mathbf{R}$, the maximization problem becomes computationally expensive. Moreover, the computational time required to maximize the likelihood function increases exponentially as the number of data points n_s increases. To find the maximum likelihood estimates with minimal computational expense, several alternative methods[1] were researched. In this study, a gradient-based method with multiple starting points is used.

The mean square error of the predictor at untried $\mathbf{x}$, $s^2(\mathbf{x})$, is obtained as follows:

$$s^2(\mathbf{x}) = \hat{\sigma}^2\left[1 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} + \frac{\left(1 - \mathbf{1}^T\mathbf{R}^{-1}\mathbf{r}\right)^2}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}}\right] \tag{9}$$

This represents the uncertainty at the estimated point. The root mean squared error (RMSE) is expressed as $s = \sqrt{s^2(\mathbf{x})}$.

Figure 1 shows an example of a Kriging model for approximating a computationally expensive function. In Figure 1, hollow points represent sample points obtained from the true function (shown as a dashed line), such as a computationally expensive computer analysis. The chained line at the bottom represents the s values. One can see that the difference between the true function and the Kriging model is large in the vicinity of x = 0.8. This difference is a result of having no sample points in that region. Also, the RMSE values become larger in this region. Note that the RMSE values at the sample points are always zero since the Kriging model is an interpolation model.

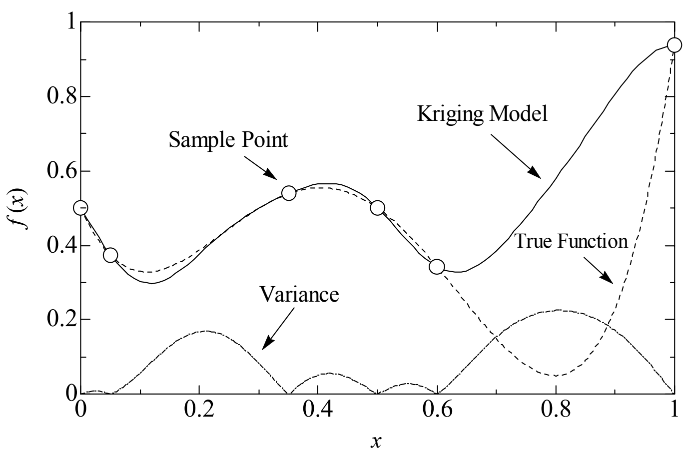


Figure 1. Schematic of Kriging Model.

B. Expected Improvement

As mentioned above, both the estimated function value and the uncertainty of the estimated function value are considered to facilitate a global search. Based on these two values, the point with the largest probability of being the global optimum is determined. The probability of being the global optimum is evaluated using the expected improvement (EI)[4] criterion.

For minimization problems where the objective function is estimated using the Kriging model, the improvement $I(\mathbf{x})$ over $f_{\min}$, (the minimum true objective function value found so far), is defined as

$$I(\mathbf{x}) = \begin{cases} f_{\min} - \hat{y}(\mathbf{x}) & \text{if } \hat{y} < f_{\min} \\ 0 & \text{otherwise} \end{cases} = \max\left[f_{\min} - \hat{y}(\mathbf{x}), 0\right] \quad (10)$$

The expected improvement is then calculated as follows:

$$E[I(\mathbf{x})] = (f_{\min} - \hat{y})\Phi\left(\frac{f_{\min} - \hat{y}}{s}\right) + s\phi\left(\frac{f_{\min} - \hat{y}}{s}\right) \quad (11)$$

where $\Phi(\cdot)$ is the normal cumulative distribution function, and $\phi(\cdot)$ is the normal probability density function. By selecting the maximum expected improvement value point as an additional sample point, robust exploration of the design space and efficient search for the global optimum can be accomplished simultaneously.

C. Proposed Kriging-based Optimization Method

The basic outline of the proposed Kriging-based optimization method for minimizing an objective function is summarized below. Figure 2 shows a flowchart of the method.

1. Use a space-filling design of experiments, for example, Latin Hyper cubic Sampling method, to obtain initial sample points for evaluating the true response function. From our practical experience, the number of initial sample points n_s should be about 10 times the number of design variables m.
2. Construct a Kriging model for maximizing the likelihood function.
3. Maximize the expected improvement function in the design space to determine where to sample next. It will be shown later that this process results in multimodal optimization problems. In this study, this maximization is carried out using particle swarm optimization (PSO).[9-11]
4. Sample the true function at the point with maximum expected improvement, and update the Kriging model ($n_s = n_s+1$). If the objective function value at the new point is smaller than the current minimum value, update the $f_{\min}$ value.
5. Stop if the stopping criterion is satisfied, otherwise go to step 2.

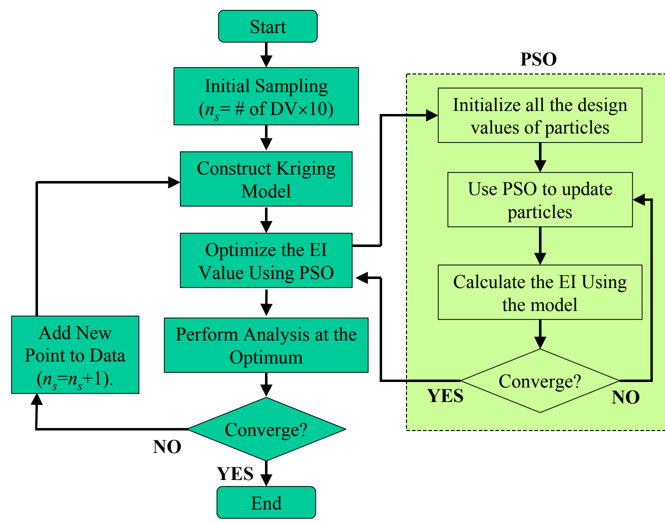
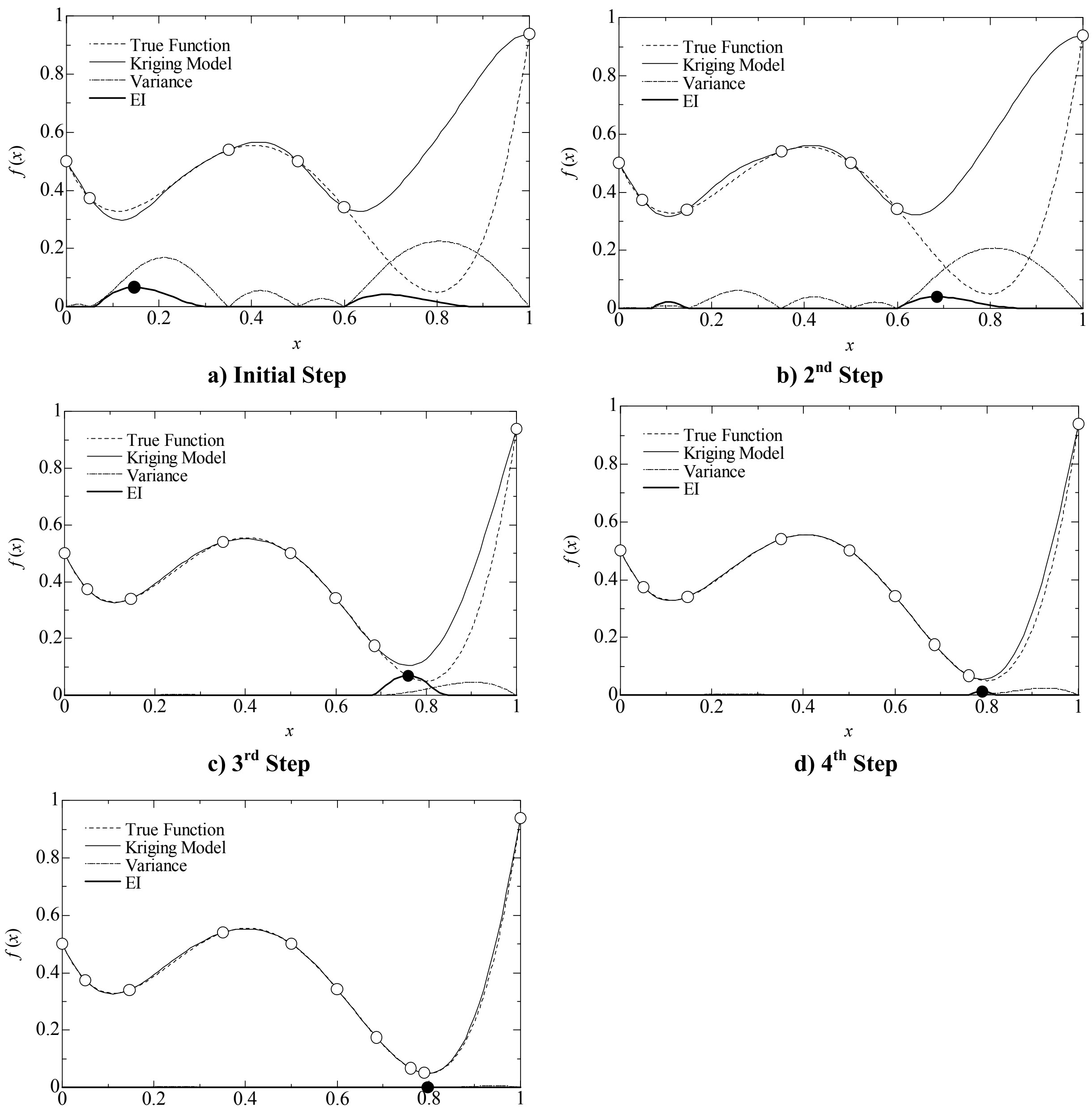


Figure 2. Flowchart of Kriging-based Optimization Method.

To demonstrate the proposed algorithm, a one-dimensional multimodal example is shown in Figure 3. In this example it is required to find a minimum of a true response function represented by dashed lines. The thick line at the bottom of the plot represents the expected improvement values. The black dot on this thick line corresponds to the maximum of the expected improvement for the Kriging model at each step. The maximum expected improvement determines where we should evaluate the true function for the current iteration. The expected improvement tends to choose design points which are most likely to improve the accuracy of the model and/or have a better function value over a point with projected minimal objective functional value only (**Could you tell me what does this word "only" mean? **). The optimization history shown in Fig. 3 demonstrates this behavior. After the six initial points are sampled, the resulting Kriging model is not accurate and does not model the global optimum. However, the expected improvement function leads the optimizer to sample additional points where the global optimum is likely to exist. After five iterations, the global optimum (x=0.7986) is found.



a) Initial Step **b) 2nd Step**

c) 3rd Step **d) 4th Step**

e) 5th Step

Figure 3. Progress of Kriging-based optimization.

D. Convergence issues

Although the described above Kriging-based optimization method is an efficient global search algorithm, it has several issues associated with it. One of them is convergence. We determined that the convergence is highly problem dependent. For some problems, it is impossible to obtain any meaningful solutions within a practical number of function evaluations. For example, when a Kriging model fits sample data poorly (unlike in the case of the example in Fig. 3), the expected improvement might become large over a wide range of design variable values because almost any additional sampling point would improve the accuracy of the model. Therefore, the maximum of the expected improvement for every iteration would be located at where the model is still uncertain. As a result, it can be difficult to find design points which actually have better objective function values over the current minimum value. In contrast, when the predicted response from the Kriging model is in close agreement with the true function (e.g., step 4 in Fig. 3), the expected improvement might be close to zero for the entire design space (as shown in Fig.3 e), which will also make it hard to pick out the candidate sampling points.

III. Combined Kriging and Gradient-based Optimization Method

To improve the Kriging-based optimizer's versatility and its convergence characteristics, a new implementation is proposed where the Kriging based optimizer is combined with a gradient-based optimizer. The Design Optimization Tools (DOT)[12] software is used as the gradient-based optimizer in our case. DOT is a commercially available optimization program, intended for solving a wide variety of nonlinear constrained or unconstrained optimization problems. We propose two variations in the following sections: one for unconstrained optimization and another for constrained optimization.

A. Unconstrained Optimization

Here the optimization strategy is similar to that of the Kriging-based optimization described in the previous section. The main difference is that now we select two sample points at each iteration instead of one. One sample point is selected based on the expected improvement criterion as described in the previous section. The second is the minimum point of the Kriging approximation of the objective function. This minimum point is obtained by applying the gradient-based optimizer to the Kriging approximation model. The starting point for the gradient-based optimizer is the best sample point found so far. In this study, the gradient of the Kriging model is calculated by finite difference method. Note that the gradient search for the second sample point is computationally inexpensive because the optimizer is only calling the current Kriging model. Unlike finding the first sampling point that is based on the expected improvement criterion, the minimization of the Kriging model is a local search within the region about the best sample point. Combining the Kriging optimization with the gradient-based optimization seeks to provide a balanced between global and local searches. The proposed approach should be able to avoid both a slow convergence and an excessive number of function evaluations.

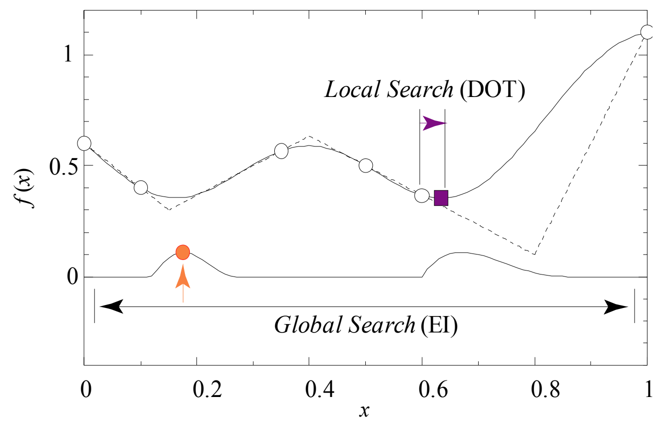


Figure 4. Schematic of Kriging-DOT optimizer

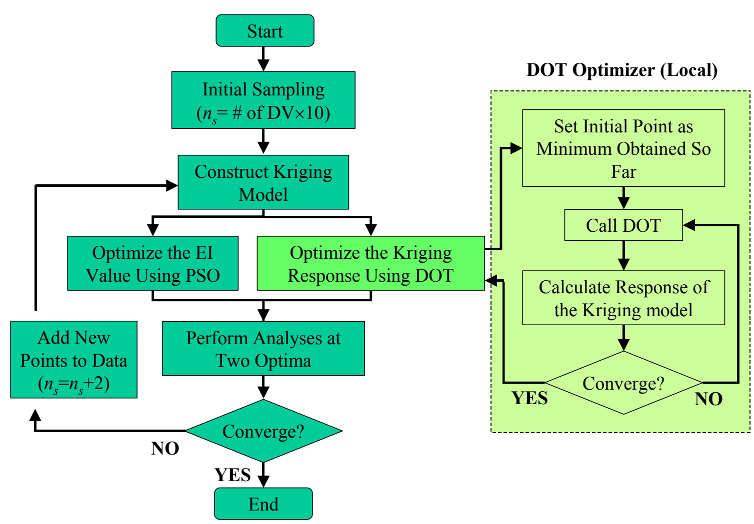


Figure 5. Flowchart of Kriging-DOT Optimizer.

The basic outline of the combined Kriging-gradient based optimization method is summarized in Fig. 5. Steps 1, 2, and 5 in Figure 5 are identical with that of the purely Kriging-based optimizer.

1. Use a space-filling design of experiments, Latin Hyper cubic Sampling method, to obtain an initial sample of the true function.
2. Construct a Kriging model for maximizing the likelihood function.

3. Maximize the expected improvement function in the design space using the PSO algorithm. Simultaneously minimize the approximate response of the Kriging model starting from the best sample point, using the gradient-based optimization method.
4. Sample the true function at the two points obtained in the previous step. Add the obtained two points to the set of sample points and update the Kriging model ($n_s = n_s+2$). If either of the obtained function values is smaller than the current minimum value, update the $f_{\min}$ value.
5. Stop if the stopping criterion is satisfied, otherwise go to step 2.

B. Constrained Optimization

A modification is required to deal with constrained optimization problems. In general, constrained optimization is relatively unexplored aspect of Kriging-based optimization methods. Several researches have attempted to extend the Kriging-based optimizer to include constrained optimization.[5,6,8] Generally, every constraint function is estimated using a separate Kriging model resulting in as many Kriging models as the number of constraints plus one. Since the construction of each Kriging model requires expensive maximization of the resulting likelihood function, an increase in the number of models results in a significant increase in the associated computational cost. For both simplicity and versatility, we adopted a different strategy that requires construction of only two Kriging models: one for the objective function and one for the maximum constraint violation value.

In general, a constrained optimization problem is subject to n_c constraints as follows:

$$g_i(\mathbf{x}) < 0 \qquad i = 1, \cdots, n_c \tag{12}$$

where every constraint $g_i(\mathbf{x})$ is normalized in such that a constraint is violated when its value is positive. The maximum constraint violation is then defined as

$$V_{\max}(\mathbf{x}) = \max_i [g_i(\mathbf{x})] \tag{13}$$

Note that all the constraints are satisfied when the maximum violation is less than zero. When using a Kriging model to approximate the maximum constraint violation, the probability of satisfying all the constraints can be calculated as follows:

$$\Pr(V_{\max}(\mathbf{x}) < 0) = 1 - \Phi\left(\frac{\hat{V}_{\max}(\mathbf{x})}{s}\right) \qquad i = 1, \cdots, n_c \tag{14}$$

By multiplying the probability obtained from Eq. (14) by the expected improvement from Eq. (11), we account for the effect of the constraints on the expected improvement of the objective function.

C. Stopping Criteria

We formulated a stopping criterion when combining Kriging-based optimization with gradient-based optimization that would terminate the optimization process if any of the following conditions is satisfied:

<u>Small Relative/Absolute Change in the Expected Improvement</u>
This criterion implies that either the relative or absolute change in the expected improvement is less than a specified tolerance. This criterion is most likely to be satisfied if the optimization procedure finds the global optimum.
<u>Small Relative/Absolute Change in the Actual Improvement</u>
When the current minimum is updated during the iteration, the improvement in the objective function is stored as the actual improvement. This criterion implies that either relative or absolute change in the actual improvement is less than a specified tolerance. This criterion is most likely to be satisfied if the gradient-based optimization program encounters a local or the global optimum.
<u>No Update of the Current Minimum $f_{\min}$</u>
If a better solution has not been found for a specified number of iterations, the optimization process is terminated to avoid excessive computations.
<u>No Change in Gradient-Based optimization result</u>
If there is no change in the position of the approximate optimum obtained from the gradient-based optimization program for a specified number of iterations, the shape of the Kriging model is not changing near the current

optimum. In this case, we can expect that the current optimum will not be improved and the optimization process is terminated to avoid excessive computations.

IV. Optimization Results

The proposed combined Kriging-gradient based optimization method is aimed at responses that are expensive to calculate. However, for verification purposes, it is convenient to consider test problems with objective functions that are computationally inexpensive to calculate. Using these test problems, we can compare the performance of the proposed approach with the performance of the other optimization methods. The chosen test functions are nonlinear and/or multimodal.

A. Unconstrained Problems

The first example is a multimodal function in two dimensions with three global optima. The function is known as the Branin function.[14] The Branin function is defined in Eq. (15) and is shown in Fig. 6.

$$f(x_1,x_2)=\left(x_2-\frac{5.1}{4\pi^2}x_1^2+\frac{5}{\pi}x_1-6\right)^2+10\left(1-\frac{1}{8\pi}\right)\cos(x_1)+10 \tag{15}$$

For our example, we considered $x_1 \in [-5,10]$ and $x_2 \in [0,15]$. The tree global minima at (3.1416, 2.2750), (9.4248, 2.4750), and (-3.1416, 12.2750) are shown as dots in Fig.6 and have identical function values equal to 0.3979.

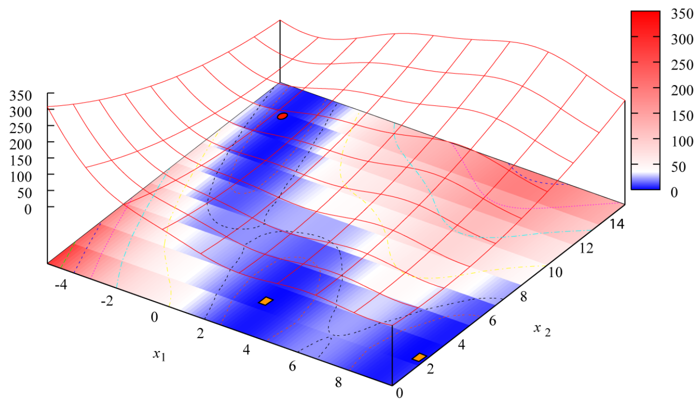
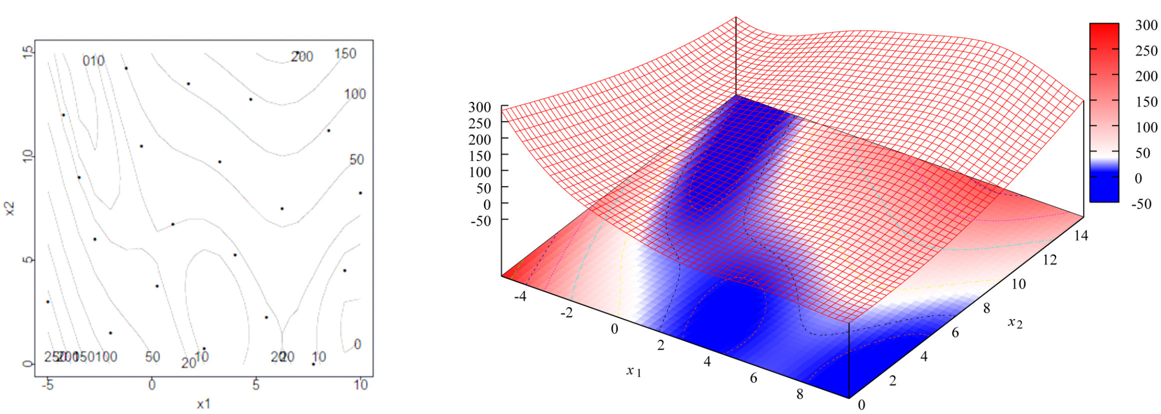


Figure 6. Branin Function.

The Kriging model, with the initial 21 sampling points, is shown graphically in Fig. 7. The Kriging model provides a very accurate approximation of the Branin function. The optimization history for the proposed method is shown in Fig. 8. In Fig. 8, the orange circles indicate the sample point identified by maximizing the expected improvement, and the violet squares indicate the sample point identified by the gradient-based optimization of the Kriging model. Figure 8 shows that the expected improvement is multimodal and its nonlinearity increases as the optimization progresses. At the 4th step, the PSO algorithm fails to maximize the expected improvement. However, the gradient-based optimizer identified a better point in the vicinity of one of the global minima. It is clear that the proposed methodology explores two different regions parts of the design space.



a) Sampled Points[5] **b) Estimated Function**

Figure 7. Plot of Fitted Kriging Model.

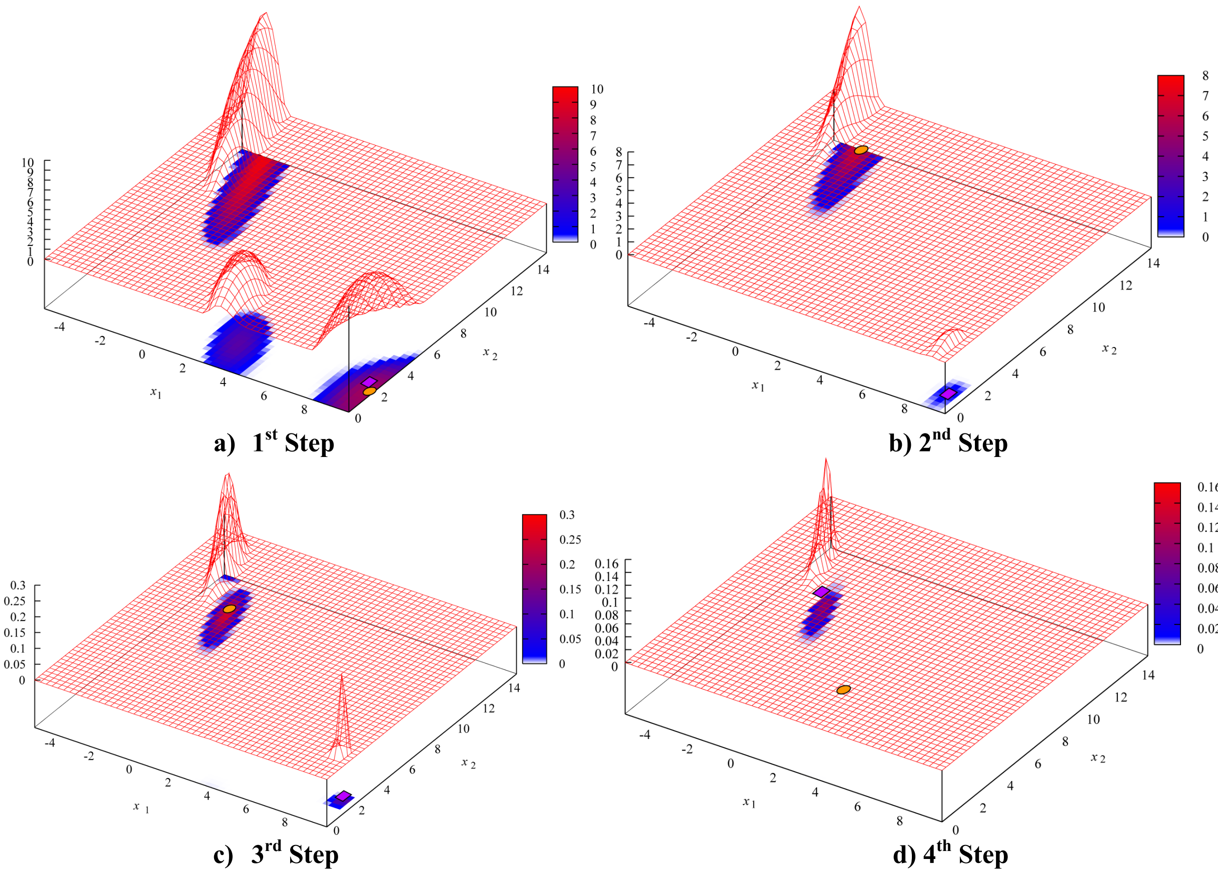


Figure 8. Plots of Expected Improvement at Each Iteration. *Orange circles indicates the position of the point obtained basing on the expected improvement, and violet square indicates the position of DOT result at each iteration.*

The optimization results for the Branin function is summarized in Table 1. The optimization was terminated due to small changes in the expected improvement. For comparison, we also performed the Kriging-based only optimization, response surface (RSA) optimization, and the following unconstrained gradient-based optimization methods: Fletcher-Reeves (FR) and Broydon-Fletcher-Goldfarb-Shanno (BFGS). The RSA optimization and the gradient-based optimizations were performed using VisualDOC[15], a commercially available, general-purpose design optimization software system.

The RSA technique that we used is the following. Initially, a sampling of a region of a design space is performed using one of several available designs of experiments, and then a polynomial approximation of the responses is constructed. After that the optimization of the polynomial model is performed. The evaluation of the true response is then conducted only at the approximate optimum. After that the approximate optimum point is added to the points that are already available, response surface model is reconstructed, and optimization of the new response surface model is performed. The procedure is repeated until the convergence is reached.

When solving our test problems we employed two types of the RSA optimization in VisualDOC, one using a Koshal initial design of experiments, and the second using a Simplex initial design of experiments. As shown in Table 1, all optimization methods except the Koshal design of experiments found a solution that is close to the actual minimum, using a small number of function evaluations. However, all the methods, except for the combined Kriging-gradient-based optimization method found only one of three global minima, since they are designed to search for one optimal solution. Although the combined Kriging-gradient-based optimization method required more function evaluations than the others, it explored the entire design space and the resulting approximation captured all the global minima.

Table 1. Result of Branin Function Problem

	Objective	Total analysis Calls	Active Stopping Criterion
Actual Optimum	0.39788		
Kriging-based	0.41013	35 (=21+14)	EI criterion
Kriging-grad	**0.39789**	**43** (=21+22)	EI criterion
RSA (Simplex)	0.39792	20	
RSA (Koshal)	0.56247	20	
BFGS	0.39813	34	
FR	0.39934	66	

The next two test problems we considered also had two design variables. The first problem corresponds to finding the equilibrium condition of a spring system, and the second represents a response with high frequency, low amplitude numerical noise. The plots of the corresponding objective functions that should be minimized are shown in Figs. 9 and 10. The spring system results in minimization of a function known as a 'banana curve,' which is difficult to solve using conventional optimization methods. The noisy problem is also difficult to solve using gradient-based methods, due to the presence of many local minima. Details of the optimization problems will be provided in the final paper. The optimization results for these problems are shown in Tables 2 and 3. The results for these two problems indicate that the combined Kriging-gradient-based optimization method is more stable than the other methods. Note that the more traditional Kriging-based optimization procedure is terminated in the middle of the optimization process because there were no updates of the current minimum for both problems.

The fourth example problem is minimization of the 4-dimentional Wood's function described by Colville.[16] Here the function value vary over a wide range and as a result, a logarithmic transformation is used to improve the accuracy of the Kriging model. The results are summarized in Table 4, which the combined Kriging-gradient based optimization method obtained a better result than those of the RSA. This function is highly nonlinear and as a result, the RSA methods failed to solve the problem. Even when we use a log-transformed Wood's function as an objective function to get results of the RSA methods more accurate, the optimization results of the RSA methods are still far from the actual optimum; the obtained objective function values are 46.58 for Simplex and 920.0 for Koshal. The combined Kriging-gradient based optimization method is thought to be well driven by the gradient-based optimizer.

The fifth example problem is to minimize the 6-dimentional Hartman function[14] (description provided by Schonlau[5)]). Similar to the research by Schonlau, log-transformed Hartman function is used as an objective function of this problem. This function is reported to be so smooth and unimodal that gradient-based methods can find the global minimum. The results of the optimization are summarized in Table 5. The function values in Table 5 are on the log-transformed scale. These results indicate that the combined Kriging-gradient based optimization method was able to explore the design space globally, and was more versatile and stable than the other methods due to that balanced search.

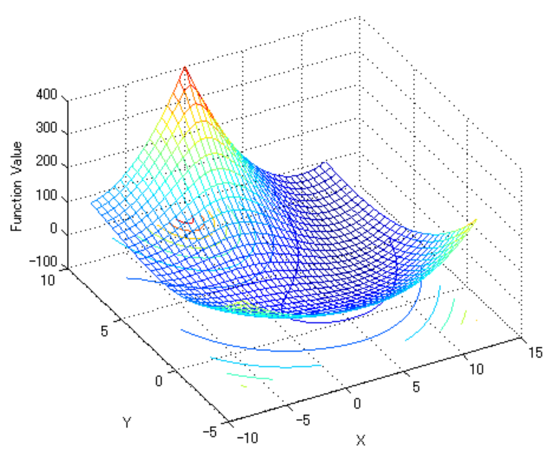


Figure 9. Spring System Problem

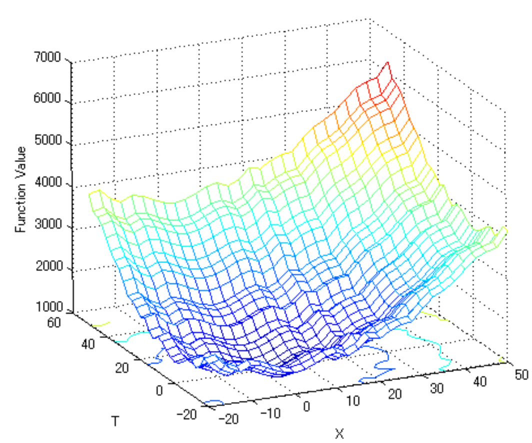


Figure 10. Noise Problem

Table 2. Result of Spring System Problem.

	Objective	Total analysis Calls	Active Stopping Criterion
Actual Optimum	-41.81		
Kriging-based	-41.8075	57 (=21+36)	No $f_{\min}$ Update
Kriging-grad	**-41.808**	**36** (=21+15)	AI criterion
RSA	-20.89	23	
BFGS/FR	-41.808	68	

Table 3. Result of Noise Problem.

	Objective	Total analysis Calls	Active Stopping Criterion
Actual Optimum	1000.0		
Kriging-based	1024.89	44 (=11+33)	No $f_{\min}$ Update
Kriging-grad	**1029.7**	**51** (=11+40)	DOT result
RSA (Simplex)	1504.1	18	
RSA (Koshal)	1017.4	22	
SQP	2715	16	
MMFD	3680.3	31	
SLP	1015.3	16	

Table 4. Result of Wood's 4-D Problem.

	Objective	Total analysis Calls	Active Stopping Criterion
Actual Optimum	0.0		
Kriging-based	106.9	115 (=41+74)	No $f_{\min}$ Update
Kriging-grad	**4.63**	**103** (=41+62)	EI criterion
RSA (Simplex)	51.42	50	
RSA (Koshal)	744.45	39	
BFGS	0.085	160	
FR SQP	0.011	330	

Table 5. Result of Hartman 6 Problem.

	Objective	Total analysis Calls	Active Stopping Criterion
Actual Optimum	-1.201		
Kriging-based	-1.142	119 (=51+68)	No $f_{\min}$ Update
Kriging-grad	**-1.199**	**163** (=51+112)	AI criterion
RSA (Simplex)	-1.199	59	
RSA (Koshal)	-0.793	36	
BFGS	-1.200	96	
FR SQP	-1.200	294	

B. Constrained Problems

The combined Kriging and gradient based optimization method was applied to five constrained example problems from the optimization literature. The details of these example problems are not mentioned here, but will be presented in the full paper. The five problems are (1) buckling column design problem: 2 design variables, 2 constraints, (2) box design problem: 3 design variables, 1 constraint, (3) construction management problem: 2 design variables, 4 constraints, (4) portfolio selection problem: 5 design variables, 4 constraints, and (5) hydraulic piston problem: 4 design variables, 4 constraints. The results of these five example problems are summarized in Tables 6-10. Note, that both the construction management problem in Table 8 and the portfolio selection problem in Table 9 are maximization problems.

For the buckling column and box design problems (results are presented in Tables 6 and 7) with a small number of constraints, the combined Kriging-gradient based optimization method worked better than the other methods. For the construction management problem and the portfolio selection problem all optimization methods showed relatively good results. For the hydraulic piston problem, the combined Kriging-gradient based optimization method performed worse than the other methods. Here only one update of the minimum was performed during the optimization procedure. In this particular case the combined Kriging-gradient based optimization method failed to capture the nature of the feasible region. Most likely that was caused by the fact that the maximum violation function in this particular problem was highly multimodal and the resulting Kriging model estimated that behavior

Table 6. Result of Buckling Column Design.

	Objective	The Worst Constraint	Total analysis Calls	Active Stopping Criterion
Actual Optimum	0.4	0		
Kriging-grad	**0.3999**	**0.0003**	**30** (=21+9)	EI criterion
RSA	0.493	-4.55	17	
SQP	0.3999		35	
SLP	0.4003		25	
MMDF	0.3999		272	

Table 7. Result of Box Design Problem.

	Objective	The Worst Constraint	Total analysis Calls	Active Stopping Criterion
Actual Optimum	12.0	0		
Kriging-grad	**11.98**	**0.003**	**46** (=33+13)	EI criterion
RSA	12.98		37	
SQP	13.10		54	
SLP	12.00	0	44	
MMFD	12.00	0	38	

Table 8. Result of Construction Management Problem.

	Objective	The Worst Constraint	Total analysis Calls	Active Stopping Criterion
Actual Optimum	63333	0		
Kriging-grad	**63527**	**0.003**	**29** (=21+8)	EI criterion
RSA	63264		9	
SQP	63498	0	7	
SLP	63324	0	33	
MMFD	63333	0	34	

Table 9. Result of Portfolio Selection Problem.

	Objective	The Worst Constraint	Total analysis Calls	Active Stopping Criterion
Kriging-grad	**0.279**	**0.002**	**119** (=51+68)	No f_{min} Update
RSA (Koshal)	0.271	-4.55	17	
RSA (Simplex)	0.294		37	
SQP	0.296		49	
SLP	0.298	0	36	
MMFD	0.297	0	95	

Table 10. Result of Hydraulic Piston Problem.

	Objective	The Worst Constraint	Total analysis Calls	Active Stopping Criterion
Actual Optimum	1085.5	0		
Kriging-grad	**1175.5**	**-0.015**	**103** (=41+62)	No f_{min} Update
RSA	1113.3		15	
SQP	1085.5		41	
SLP	1085.5		57	
MMFD	1085.5		39	

rather poorly. To improve the performance of the combined Kriging-gradient based optimization method, it is possible to include more feasible sample points in the initial sample data set.

We may conclude that for the constrained problems the combined Kriging-gradient based optimization method needs more improvements particularly, in constraint handling.

V. Conclusions

In order to develop a practical and general-purpose Kriging-based optimizer for general design problems of complex multidisciplinary systems, a new implementation of the Kriging-based optimization method, the combined Kriging and gradient-based optimization method is proposed. Several stopping criteria are explored to prevent premature convergence on one hand and to avoid unnecessary excessive computations on the other hand. The combined Kriging and gradient-based optimization method has a balance of both the global search characteristics of the Kriging-based optimization and the local search characteristics of the gradient-based optimization. We applied the proposed method to both unconstrained and constrained example problems. For the unconstrained problems, the proposed method is more versatile and stable than conventional optimization methods. For the constrained problems the proposed method was not as good as for the unconstrained problems. More research is required to adapt the proposed method to the general constrained optimization problems.

The final paper will contain more details on the test problems presented above as well as more sophisticated test problems. The latest research on the proposed algorithm will also be presented.

Acknowledgments

This work has been performed as the first author's internship program at Vanderplaats Research & Development, Inc. The first author gratefully acknowledges support from Dr. Garret N. Vanderplaats.

References

[1]Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell T.J., and Morris, M.D., "Screening, Predicting, and Computer Experiments", *Technometrics*, Vol. 34, No. 1, 1992, pp. 15-25.

[2]Simpson, T.W., Mauery, T.M., Korte, J.J., and Mistree, F., "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233-2241.

[3]Martin, J.D., Simpson, T.W., "Use of Kriging Models to Approximate Deterministic Computer Models," *AIAA Journal*, Vol. 43, No. 4, 2005, pp. 853-863.

[4]Jones, D.R., Schonlau, M., and Welch, W.J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, 1998, pp. 455-492.

[5]Schonlau, M., "Computer Experiments and Global Optimization," Ph.D. thesis, University of Waterloo, Waterloo, Canada, 1997.

[6]Sasena, M.J., "Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations," Ph.D. dissertation, University of Michigan, MI, 2002.

[7]Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, No. 2, 2005, pp. 413-420.

[8]Jeong, S., Yamamoto, K., and Obayashi, S., "Kriging-Based Probabilistic Method for Constrained Multi-Objective Optimization Problem," *Proceedings of AIAA 1st Intelligent Systems Technical Conference*, 2004.

[9]Venter, G., and Sobieszczanski-Sobieski, J., "Particle Swarm Optimization," *AIAA Journal*, Vol. 44, No. 8, 2003, pp. 1583-1589.

[10]Fourie, P.C., and Groenwold, A.A., "The Particle Swarm Optimization Algorithm in Size and Shape Optimization," *Structural and Multidisciplinary Optimization*, Vol. 23, No. 4, pp. 259-267.

[11]Venter, G., and Sobieszczanski-Sobieski, J., "Multidisciplinary Optimization of a Transport Aircraft Wing Using Particle Swam Optimization, "*Proceedings of 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 2002.

[12]DOT, Design Optimization Tools, Software Package, Ver. 5.5, Vanderplaats Research & Development, Inc., Colorado Springs, CO.

[13]Vanderplaats R&D, Inc., *DOT Users Manual*, 2001.

[14]Törn, A. and Žilinskas, A., *Global Optimization*, Springer Verlag, Berlin, 1989.

[15]VisualDOC, Software Package, Ver. 5.1, Vanderplaats Research & Development, Inc., Colorado Springs, CO.

[16]Colville, A.R., "A Comparative Study of Nonlinear Programming Codes," Technical Report Mo. 320-2949, IBM New York Scientific Center, 1986.