

# Constrained efficient global optimization with support vector machines

Anirban Basudhar · Christoph Dribusch ·  
Sylvain Lacaze · Samy Missoum

Received: 2 February 2011 / Revised: 4 November 2011 / Accepted: 8 November 2011 / Published online: 31 January 2012  
© Springer-Verlag 2012

**Abstract** This paper presents a methodology for constrained efficient global optimization (EGO) using support vector machines (SVMs). While the objective function is approximated using Kriging, as in the original EGO formulation, the boundary of the feasible domain is approximated explicitly as a function of the design variables using an SVM. Because SVM is a classification approach and does not involve response approximations, this approach alleviates issues due to discontinuous or binary responses. More importantly, several constraints, even correlated, can be represented using one unique SVM, thus considerably simplifying constrained problems. In order to account for constraints, this paper introduces an SVM-based “probability of feasibility” using a new Probabilistic SVM model. The proposed optimization scheme is constituted of two levels. In a first stage, a global search for the optimal solution is performed based on the “expected improvement” of the objective function and the probability of feasibility. In a second stage, the SVM boundary is locally refined using an adaptive sampling scheme. An unconstrained and a constrained formulation of the optimization problem are presented and compared. Several analytical examples are used to test the formulations. In particular, a problem with 99 constraints and an aeroelasticity problem with binary

output are presented. Overall, the results indicate that the constrained formulation is more robust and efficient.

**Keywords** Efficient global optimization · Constrained optimization · Support vector machines · Binary problems · Discontinuities · Classification

## 1 Introduction

Surrogate-based optimization using response surfaces or metamodels has gained popularity during the past two decades (Simpson et al. 2008; Queipo et al. 2005; Huang et al. 2006; Viana et al. 2010a; Wang and Shan 2007; Kleijnen 2009; Jin et al. 2003). Recently, efficient global optimization (EGO) (Jones et al. 1998) has emerged as one of the most promising approaches for costly simulators. EGO is based on the approximation of responses using a Gaussian process (e.g., Kriging (Cressie 1990; Stein 1999)). The key feature of this technique stems from the availability of the variance of the prediction over the entire search space. From this information, an “expected improvement” of the objective function can be assessed. The optimal solution is then found by maximizing the expected improvement using a global search.

In its original formulation, EGO was developed for unconstrained optimization (Jones et al. 1998). Subsequently, EGO was extended to constrained optimization (Schonlau 1997; Sasena et al. 2002a, b; Audet et al. 2000; Sasena 2002; Bichon et al. 2009). In these studies, both the objective function and the constraints were approximated using Kriging. Despite promising results, however, there are still significant challenges for an efficient constrained optimization. The main difficulty lies in the definition of a general convergent scheme that simultaneously minimizes

---

A. Basudhar · C. Dribusch · S. Lacaze · S. Missoum (✉)  
Aerospace and Mechanical Engineering Department,  
The University of Arizona, Tucson, AZ 85721, USA  
e-mail: smissoum@email.arizona.edu

*Present Address:*

A. Basudhar  
Livermore Software Technology Corporation (LSTC),  
7374 Las Positas Road, Livermore, CA 94551, USA

the objective while satisfying approximated constraints with a reasonable number of function evaluations. The hurdles become even more pronounced as the number of constraints increases or when binary and discontinuous responses are present. Another difficulty appears when the constraints are correlated (Forrester et al. 2008).

This work investigates a new approach to perform constrained EGO which addresses the aforementioned issues. The fundamental idea is to combine the approximation of the objective function with a classification approach to handle the constraints. More specifically, the objective function is approximated using Kriging (Cressie 1990) while the boundary of the feasible space is constructed explicitly using a support vector machine (SVM) (Vapnik 1998; Gunn 1998; Cristianini and Shawe-Taylor 2006; Scholkopf and Smola 2002). In this setting, the constraint values are no longer approximated, as done with surrogate-based approaches, but simply classified as feasible or infeasible. This approach represents the basis for the *explicit design space decomposition* approach (Basudhar et al. 2008; Basudhar and Missoum 2008, 2010) developed by the authors.

The classification approach and the construction of explicit boundaries using SVM have several advantages, such as the handling of discontinuous and binary responses (Missoum et al. 2007; Basudhar et al. 2008; Basudhar and Missoum 2009). Another major advantage of the classification approach is the reduction of several constraints (or failure modes) to a single SVM expression. This greatly simplifies the problem from an optimization or a reliability assessment point of view (Arenbeck et al. 2010). In the context of constrained EGO, this characteristic offers a major simplification for problems with a large number of constraints.

In the literature on constrained EGO formulations, feasibility is enforced by complementing the expected improvement with the notion of “probability of feasibility” (Forrester et al. 2008; Schonlau 1997; Sasena 2002). Each constraint, approximated by a Kriging surrogate, has a corresponding probability of feasibility at a given point. In several studies, the optimization is solved as an unconstrained formulation that maximizes the product of the expected improvement and the probability of feasibility. If there are several *independent* constraints, the probability of feasibility is the product of the individual probabilities. Constrained and penalized formulations can also be found in Sasena et al. (2002a) and Sasena (2002).

In this article, the probability of feasibility is no longer computed based on individual constraint approximations. Instead, the boundary of the feasible domain is approximated by a single SVM, and the probability of feasibility is calculated using a new Probabilistic SVM (PSVM) model. The proposed PSVM model is based on the existing Platt

sigmoid model (Platt 1999) but also provides important improvements. Another feature of the proposed approach stems from the presence of two sample selection levels: a first stage is dedicated to the global search of the constrained optimum while a second stage is focused on the local refinement of the SVM approximating the boundary of the feasible space. In order to maintain consistency with the literature of purely Kriging-based approaches, this article introduces and compares an unconstrained and a constrained formulation of the SVM-based constrained EGO.

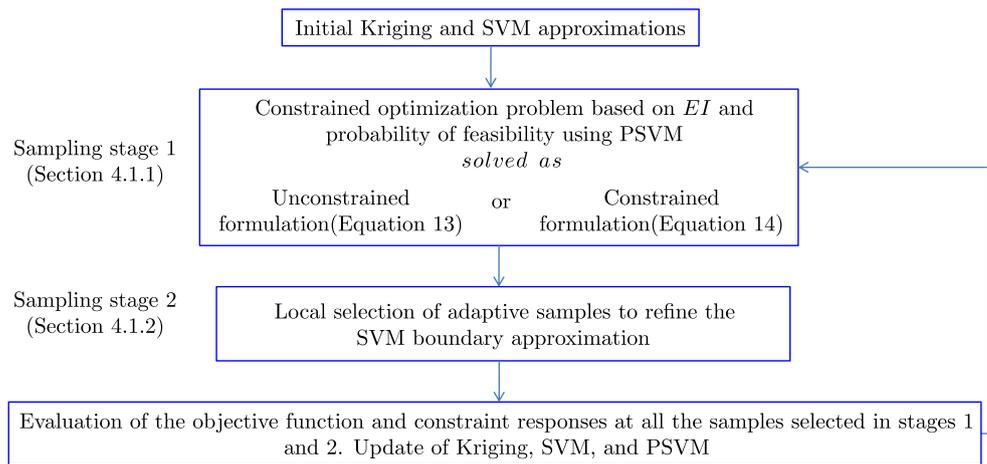
The article is organized as follows. A brief summary of the proposed method is presented in Section 2. A background on unconstrained and constrained EGO is provided in Section 3.1. Section 3.2 introduces SVMs and PSVMs. The proposed methodology for EGO based on SVM and the modified PSVM model are presented in Section 4. Finally, several test examples are considered in Section 5 to validate the efficacy of the proposed methodology. An analytical example is used to compare the SVM-based approach to existing results from the literature. An analytical example with 99 constraints is presented to show the efficacy and efficiency of the approach to tackle multiple constraints. For these problems, both constrained and unconstrained formulations are compared. The method is also compared to a technique with a random forest classifier (Gramacy and Lee 2010; Lee et al. 2010) to tackle binary problems. The last example is an engineering application with five variables dealing with the optimal design of a wing with stability constraints. A study of the effect of the size of the initial design of experiments, as well as another analytical example are also provided in the Appendices A and B.

## 2 Summary of the proposed approach

As mentioned in the introduction, the proposed algorithm is constructed around two main notions: The expected improvement (*EI*) and the probability of feasibility (see Section 3). While the notion of expected improvement is identical to the one found in traditional EGO, the main novelty of this work stems from the use of SVM to approximate the boundary of the feasible region and to calculate the probability of feasibility. The main features and advantages of the SVM-based algorithm are:

- The boundary of the feasible domain is constructed using an SVM which is a classification (feasible or infeasible) approach. Thus, problems with discontinuities and binary outputs can be treated (Basudhar et al. 2008).
- As a result of the classification approach, several constraints can be reduced to one single SVM (Arenbeck et al. 2010).

**Fig. 1** Summary of the proposed constrained efficient global optimization using SVM



- The probability of feasibility is calculated using a probabilistic SVM (PSVM) (Section 3.2). Because the feasible domain is represented with one single SVM, the probability of feasibility can be calculated even in the case where the constraints are correlated.
- The optimization is a two level process: a first stage explores the space globally to find the constrained optimum while a second stage refines the SVM boundary locally within an update region.

For completeness and consistency with the literature (see Section 3), this article compares a constrained and an unconstrained formulation of the first stage of the algorithm. A summary of the proposed study is provided in Fig. 1 with the corresponding sections. The detailed description of the algorithm is provided in Section 4.

### 3 Background

#### 3.1 Efficient global optimization (EGO)

This section provides a background and a short literature review on efficient global optimization (EGO) for unconstrained and constrained problems. For a broad overview of Kriging-based optimization, the reader is referred to existing extensive reviews (Forrester and Keane 2009; Kleijnen 2009).

##### 3.1.1 Unconstrained EGO

The original EGO formulation was developed to solve unconstrained optimization problems (Jones et al. 1998). The basic idea of EGO is to build a Gaussian process model (e.g., Kriging) (Chiles and Delfiner 1999; Santner et al. 2003; Rasmussen and Williams 2005; Jones et al. 1998) of the objective function and to update this model by selecting additional samples that have the highest “likelihood” of

minimizing the objective function. This is made possible by the fact that a Gaussian process provides the variance in the prediction assuming a normal distribution. This allows one to calculate an expected improvement (EI) of the objective function. Samples are then selected by maximizing EI. The following provides the derivation of EI.

Consider a response function  $w_{act}(\mathbf{x})$ . The corresponding Gaussian process approximation is:

$$\hat{w}(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} + Z(\mathbf{x}) \tag{1}$$

where  $\mathbf{h}$  is the trend of the model (e.g, a linear trend as used in this work),  $\boldsymbol{\beta}$  is the vector of trend coefficients, and  $Z$  is a stationary Gaussian process based on the correlation between samples. The covariance between any two samples  $\mathbf{a}$  and  $\mathbf{b}$  is defined as:

$$cov[Z(\mathbf{a}), Z(\mathbf{b})] = \sigma_Z^2 R(\mathbf{a}, \mathbf{b}) \tag{2}$$

where  $\sigma_Z^2$  is the variance of the Gaussian process  $Z$  and  $R$  is the correlation function. Among the possible choices of correlation functions (Rasmussen and Williams 2005), a widely used form for  $R$  is the exponential correlation function:

$$R(\mathbf{a}, \mathbf{b}) = e^{-\sum_{i=1}^m \theta_i |a_i - b_i|^{p_i}} \tag{3}$$

where  $m$  is the number of dimensions and  $\theta_i$  is the scaling parameter for the  $i^{th}$  dimension. The parameter  $p_i$  determines the smoothness of the correlation function (e.g.,  $p_i = 2$ , as used in this article, is a Gaussian correlation). The  $\theta_i$  values are determined by maximum likelihood (Jones et al. 1998; Martin and Simpson 2005). At any point  $\mathbf{x}$ , the mean prediction  $\mu_w(\mathbf{x})$  and variance  $\sigma_w^2(\mathbf{x})$  of the response are:

$$\mu_w(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{w} - \mathbf{F}\boldsymbol{\beta}) \tag{4}$$

$$\sigma_w^2(\mathbf{x}) = \sigma_z^2 - [\mathbf{h}(\mathbf{x})^T \quad \mathbf{r}(\mathbf{x})^T] \begin{bmatrix} \mathbf{0} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}(\mathbf{x}) \\ \mathbf{r}(\mathbf{x}) \end{bmatrix} \quad (5)$$

where  $\mathbf{r}(\mathbf{x})$  is the vector of covariance between  $\mathbf{x}$  and all the training samples,  $\mathbf{R}$  is a matrix consisting of the pairwise covariances between all the samples,  $\mathbf{w}$  is the vector of response values at the  $N$  samples, and  $\mathbf{F}$  is a matrix with its  $i^{th}$  row given by the trend  $\mathbf{h}(\mathbf{x}_i)^T$  calculated at the  $i^{th}$  sample. Further details for constructing a Kriging model can be found in Chiles and Delfiner (1999), Santner et al. (2003), Martin and Simpson (2005) and Bichon (2010).

The expected improvement is the expected value by which the predicted objective function value is lower than the current minimum:

$$EI(\mathbf{x}) = E[I(\mathbf{x})] \quad (6)$$

where  $I(\mathbf{x})$  is the improvement function defined as:

$$I(\mathbf{x}) = \max(0, w^* - w(\mathbf{x})) \quad (7)$$

where  $w^*$  is the current minimum and  $w$  is a realization of  $\hat{w}$ . The expected improvement is defined as:

$$EI(\mathbf{x}) = \int_{-\infty}^{w^*} (w^* - w) f_{\hat{w}} dw \quad (8)$$

where  $f_{\hat{w}}$  is the normal probability density function of the Kriging model  $\hat{w}$  at point  $\mathbf{x}$ . Figure 2 provides an example of “probability of improvement”. For a Gaussian process model,  $EI$  can be expressed analytically (Bichon 2010):

$$EI(\mathbf{x}) = (w^* - \mu_w(\mathbf{x}))\Phi\left(\frac{w^* - \mu_w(\mathbf{x})}{\sigma_w(\mathbf{x})}\right) + \sigma_w(\mathbf{x})\phi\left(\frac{w^* - \mu_w(\mathbf{x})}{\sigma_w(\mathbf{x})}\right) \quad (9)$$

where  $\phi$  and  $\Phi$  are the standard normal probability density function and cumulative density function respectively. The point with the maximum expected improvement is evaluated to update the Kriging model. The maximization of  $EI$  bal-

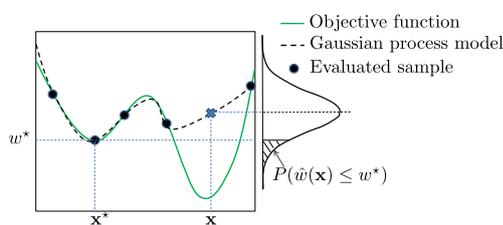


Fig. 2 Depiction of the probability of improving the current minimum  $\mathbf{x}^*$

ances the exploration of sparse regions and the exploitation of regions with low objective function values.

Several variations and improvements of the EGO sampling criterion can be found in the literature Schonlau (1997), Sasena et al. (2002a), Sasena (2002), Forrester et al. (2008), Henkenjohann and Kunert (2007), Ponweiser et al. (2008) and Ginsbourger et al. (2007). These works introduce notions such as generalized expected improvement (Sasena 2002), parallel evaluation of samples (Ponweiser et al. 2008; Ginsbourger et al. 2007), or use of EGO with other metamodels than Kriging (Viana et al. 2010b). Also, relevant to this work, a Kriging formulation that handles binary responses was recently developed (Picheny et al. 2008).

### 3.1.2 Constrained EGO implementations

In order to include constraints in the optimization, several EGO formulations have been proposed. Schonlau (1997) proposed the multiplication of  $EI$  with the probability of feasibility, calculated using a Kriging approximation for each constraint. In the case of multiple constraints, the probability is given by the product of the probability of feasibility of each constraint. One of the concerns with the “product formulation” is that one of the terms may dominate. As pointed out in Sasena (2002), this may prevent sampling on the constraint boundary where the optimum may lie. In order to overcome this concern, the penalty method has been used (Sasena et al. 2002a). In addition, when the constraints are correlated, the computation of the probability of feasibility requires information about the correlation that is often not available. In many practical applications, this is a problem as the constraints (e.g., responses of a system) might indeed be correlated.

In Sasena (2002) the maximization of  $EI$  with samples constrained to lie in the feasible space was proposed. The feasible space was defined based on the mean values of the Kriging models for the constraints. Another way of handling the constraints involves the use of the “expected violation” (Audet et al. 2000). The expected violation is calculated in the same way as the expected improvement and provides a measure of the expected amount by which a constraint is violated. It is then used to penalize the expected improvement. Augmented Lagrangian methods for handling the constraints have also been used (Bichon et al. 2009).

A common feature of most conventional methods lies in the use of Kriging approximations for both the objective function, as well as for each constraint. This would pose a problem if there is a large number of constraints. In addition, these methods would be hampered by discontinuous or binary constraints. Note that recently, indicator Kriging and random forest classifiers have been used in an attempt to address these issues (Picheny et al. 2008; Gramacy and

Lee 2010; Lee et al. 2010). A new approach of addressing these issues, based on SVMs, is explored in this article.

### 3.2 SVMs and Probabilistic SVMs

#### 3.2.1 Support vector machines

SVMs are used in this work to approximate the boundary of the feasible space. An SVM is a classifier that can define a highly nonlinear explicit boundary that optimally separates two classes of samples (e.g. feasible and infeasible). The SVM equation with  $N$  training samples is:

$$s(\mathbf{x}) = b + \sum_{i=1}^N \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}) = 0, \tag{10}$$

where  $\mathbf{x}_i$  is the  $i^{th}$  training sample,  $\lambda_i$  is the corresponding Lagrange multiplier,  $y_i$  is the class that can take values  $+1$  or  $-1$ ,  $K$  is a kernel function, and  $b$  is the bias. For any point  $\mathbf{x}$ , the predicted class to which it belongs is given by the sign of  $s(\mathbf{x})$ . The Lagrange multipliers  $\lambda_i$  corresponding to the support vectors are strictly positive, whereas they are zero for all other samples. Thus, the value of  $s(\mathbf{x})$  depends only on the support vectors.

Commonly used kernel functions are the polynomial and Gaussian radial basis functions. The polynomial kernel is used in this study and is given as:

$$K(\mathbf{x}_i, \mathbf{x}) = (1 + \langle \mathbf{x}_i, \mathbf{x} \rangle)^p \tag{11}$$

For more details on SVMs, the reader is referred to Vapnik (1998), Gunn (1998), Cristianini and Shawe-Taylor (2006) and Scholkopf and Smola (2002).

#### 3.2.2 Probabilistic support vector machines (PSVMs)

A deterministic SVM only provides a binary classification. However, because an SVM is built with a limited amount of information, the classifier might be locally inaccurate. Therefore the class prediction might be wrong. For this reason, PSVMs were introduced to provide the probability that a sample belongs to a specific class.

The most common PSVM model is based on the sigmoid function (Vapnik 1998; Platt 1999). For a given sample  $\mathbf{x}$ , the probability of belonging to the  $+1$  class is:

$$P(+1|\mathbf{x}) = \frac{1}{1 + e^{As(\mathbf{x})+B}} \tag{12}$$

The parameters  $A$  ( $A < 0$ ) and  $B$  of the sigmoid function are found by maximum likelihood. For the details on train-

ing the basic sigmoid model, the reader may refer to Platt (1999) and Lin et al. (2007).

In this work, a PSVM is used to calculate the probability of feasibility. It is the classification counterpart of the probability of feasibility found in the Kriging-based approaches. Note that a secondary aspect of this work is the modification of the sigmoid model (12) proposed by Platt (1999) and Lin et al. (2007). The modification, described in Section 4.2, was necessary to alleviate some of the inconsistencies found in the PSVM model in (12).

## 4 SVM-based constrained EGO

A summary of the proposed SVM-based constrained EGO method was presented in Section 2. This section provides a more detailed description of the various steps.

In comparison to existing EGO approaches described in Section 3.1, the main novelty of the proposed study lies in the handling of the constraints. Instead of approximating each constraint with a Gaussian process, the boundary of the feasible space is approximated explicitly using an SVM. Therefore the constraints are *not approximated* over the whole domain as when using a Gaussian process. The use of an SVM for constraint handling has several major advantages:

- Since this method eliminates the need to approximate the constraints, it can handle problems with discontinuous or binary functions.
- A single SVM is used to define all the optimization constraints. Unlike most previous methodologies (Section 3.1), this method does not require the multiplication of the probabilities of feasibility for individual constraints. Therefore, the multiplicative error propagation in calculating the probability is avoided.
- Because classification is used, only a fraction of the constraints might need to be evaluated.
- The approach can handle dependent constraints without explicit knowledge of their correlation.

### 4.1 Sample selection for constrained EGO

As already mentioned in Section 2, the algorithm consists of two stages. The first stage is based on the expected improvement and the probability of feasibility. This is the main stage driving the optimization. In a second stage, auxiliary samples are selected within a local update region. The main goal of these auxiliary samples is to refine the SVM approximation. The update algorithm is detailed in the following sections and is summarized in Algorithm 1.

**Classify the samples into feasible and infeasible classes**

- 1: Sample the space with a Centroidal Voronoi Tessellation (CVT) Design of Experiments (DOE).
- 2: Evaluate the objective function at each sample.
- 3: Construct the initial Kriging model for the objective function.
- 4: Classify the samples into feasible and infeasible classes.
- 5: Construct the initial SVM boundary that separates the classified samples.
- 6: Calculate the parameters of the PSVM model using maximum likelihood.
- 7: **repeat**
- 8: Set  $\omega^*$  equal to the minimum objective function value among the evaluated *feasible* samples.
- 9: Select a sample  $\mathbf{x}_{EI}$  based on the  $EI$  and  $P(+1|\mathbf{x})$  using (13) (update scheme 1) or (14) (update scheme 2).
- 10: Define the center and the radius of the update region. The center is selected as the current  $EI$  sample  $\mathbf{x}_{EI}$ , unless the expected improvement at the sample is equal to zero. If  $EI$  is zero at this sample, then the center is selected as the evaluated feasible sample with minimum objective function value. The radius of the region is calculated as described in Section 4.1.2.
- 11: Select  $n_p$  auxiliary samples in the update region (Section 4.1.2).
- 12: Update the Kriging model for the objective function, the SVM and PSVM with the  $1 + n_p$  samples. These samples are evaluated in parallel. Note that the  $EI$  sample is evaluated only if the expected improvement is non-zero.
- 13: **until** convergence or maximum number of iterations.

*4.1.1 Stage 1—selection of samples based on the expected improvement and the probability of feasibility*

In stage 1, an unconstrained and a constrained formulation are explored to select “ $EI$  samples”  $\mathbf{x}_{EI}$ .

- *Unconstrained formulation (update scheme 1):* This formulation is similar to the probability adjusted  $EI$  method in Schonlau (1997). By defining the feasible space as the “+1” class, the iterates are selected by maximizing the product of  $EI$  and the probability of feasibility defined as  $P(+1|\mathbf{x})$ . Unlike the original method (Schonlau 1997), the value of  $P(+1|\mathbf{x})$  is calcu-

lated using a PSVM (see Section 4.2). The optimization problem is:

$$\max_{\mathbf{x}} EI(\mathbf{x})P(+1|\mathbf{x}) \tag{13}$$

This global optimization problem can be solved using a Genetic Algorithm (GA), which is the method used in this article.

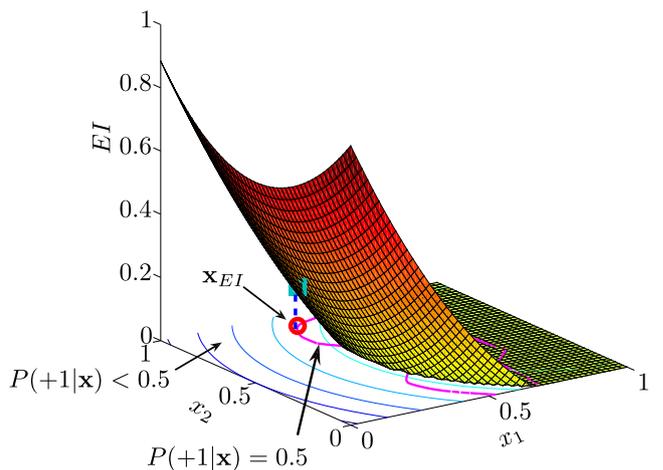
- *Constrained formulation (update scheme 2):* In this scheme, the probability of feasibility, based on one single SVM, is introduced as a constraint:

$$\begin{aligned} \max_{\mathbf{x}} \quad & EI(\mathbf{x}) \\ \text{s.t.} \quad & P(+1|\mathbf{x}) \geq 0.5 \end{aligned} \tag{14}$$

Similar to update scheme 1, the solution of (14) requires a global search. Note that the threshold for the probability of feasibility is set to 0.5. For a perfectly accurate SVM, this value corresponds to a sample on the boundary of the feasible domain (Fig. 3).

*4.1.2 Stage 2—selection of auxiliary samples to refine the constraint boundary approximation*

The selection of samples in stage 1 using (13) or (14) is performed globally over the whole space. Stage 2 of the update investigates the local refinement of the SVM boundary. After a sample from stage 1 is found,  $n_p$  additional samples are added in its vicinity with the purpose of refining the SVM boundary and improving the estimate of  $P(+1|\mathbf{x})$ . In this work,  $n_p$  is equal to the problem dimensionality  $m$ . The auxiliary samples are selected in a hyperspherical



**Fig. 3** Selection of a sample using the maximization of  $EI$  in regions with at least 50% probability of feasibility

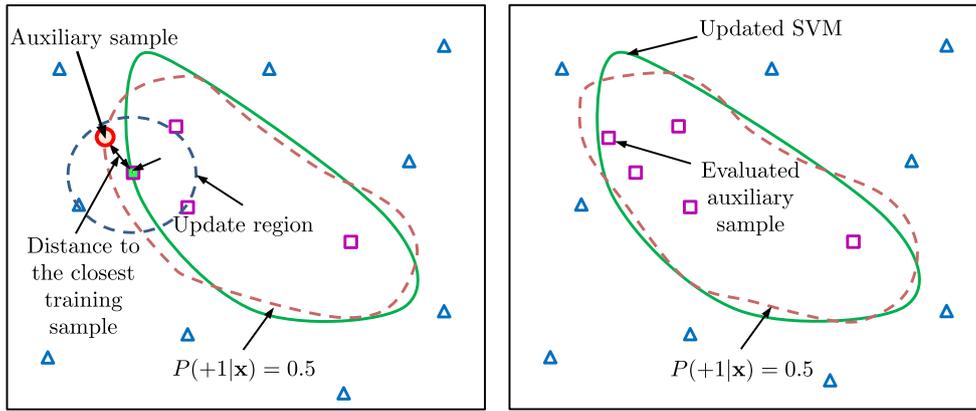


Fig. 4 Update of the SVM constraint due to an auxiliary sample

update region centered at the current  $\mathbf{x}_{EI}$ . At iteration  $k$ , the radius  $R_u^k$  of the update region is selected such that it consists of at least one sample from either class:

$$R_u^k = \max(d_+, d_-) \tag{15}$$

where  $d_+$  and  $d_-$  are the distances to the closest +1 and -1 samples from the center of the hypersphere.

In addition, the auxiliary samples are selected in sparse regions with a high probability of misclassification (incorrect class prediction) by the SVM (Fig. 4). The optimization problem to locate an auxiliary sample is:

$$\begin{aligned} \max_{\mathbf{x}} \quad & d(\mathbf{x}) \\ \text{s.t.} \quad & P_m(\mathbf{x}) \geq 0.5 \\ & \|\mathbf{x} - \mathbf{x}_c\| \leq R_u^k \end{aligned} \tag{16}$$

where  $d$  is the distance to the closest training sample and  $\mathbf{x}_c$  is the center of the hypersphere.  $P_m$  is the probability of misclassification defined as:

$$P_m(\mathbf{x}) = \begin{cases} 1 - P(+1|\mathbf{x}) & s(\mathbf{x}) \geq 0 \\ P(+1|\mathbf{x}) & s(\mathbf{x}) < 0 \end{cases} \tag{17}$$

The value of 0.5 used for the probability of misclassification corresponds to samples on or close to the SVM. The probability of misclassification can be larger than 0.5, as shown in Fig. 5. More details are provided in Section 4.2. The following points are noteworthy:

- In general  $\mathbf{x}_c = \mathbf{x}_{EI}$ . However, in the case where  $EI(\mathbf{x}_{EI})$  is zero, the center of the update region is selected as the current optimum  $x^*$ , which is the sample with minimum objective function value among all

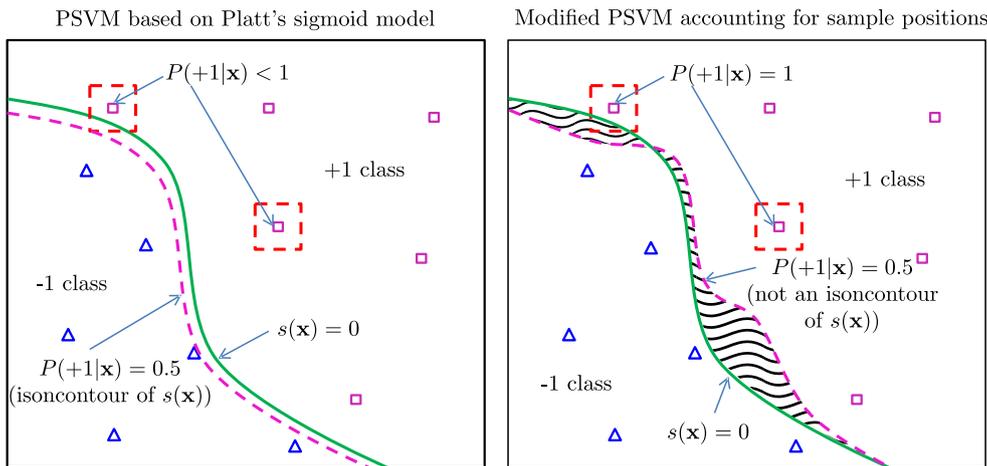


Fig. 5 Comparison of the modified PSVM with the basic sigmoid model. The shaded region in the right figure represents a region with probability of misclassification larger than 0.5. Such regions are the ones bounded by  $s(\mathbf{x}) = 0$  and  $P(+1|\mathbf{x}) = 0.5$

the evaluated *feasible* samples. This step is needed even when the expected improvement is zero, because the approximated feasible region might be inaccurate and needs to be further refined with auxiliary samples. Refinement of the approximated feasible region may lead to a non-zero  $EI(\mathbf{x}_{EI})$  in the next iteration.

- The auxiliary samples and  $\mathbf{x}_{EI}$  are all evaluated in parallel.
- $R_u^k$  is based on the local distances between samples and will therefore change from one iteration to the next.

#### 4.2 Modified probabilistic support vector machine (PSVM) model

The calculation of the probability of feasibility requires a PSVM model that correctly captures the possibility of an inaccurate SVM. A commonly used model is the basic sigmoid model given in (12) (Platt 1999). However, a limitation of this model lies in the disregard for the spatial positions of the samples. In this model, the value of  $P(+1|\mathbf{x})$  depends only on  $s(\mathbf{x})$  values, which are dictated by a small fraction of the samples (support vectors). Thus, it may predict a non-zero  $P(+1|\mathbf{x})$  even for training samples belonging to the  $-1$  class. Another limitation is that the value of the constant term  $B$  may bias  $P(+1|\mathbf{x})$  towards one of the classes if the number of samples in that class is much greater than the other one. In this work, a modified PSVM model is proposed that overcomes those limitations (Fig. 5):

$$P(+1|\mathbf{x}) = \frac{1}{1 + e^{As(\mathbf{x}) + B\left(\frac{d_-(\mathbf{x})}{d_+(\mathbf{x}) + \delta} - \frac{d_+(\mathbf{x})}{d_-(\mathbf{x}) + \delta}\right)}}$$

$$A < \frac{-3}{\min(s_{\max}, -s_{\min})}, B < 0 \tag{18}$$

where  $d_-(\mathbf{x})$  and  $d_+(\mathbf{x})$  are the distances to the closest  $-1$  and  $+1$  samples.  $\delta$  is a small quantity (set to  $10^{-10}$  in this work) added in order to avoid numerical issues at the evaluated training samples.  $s_{\max}$  and  $s_{\min}$  are the maximum and minimum values of the SVM calculated at the training samples. The proposed model is therefore dependent on both the SVM values as well as the spatial distribution of the samples.

The proposed model satisfies the following limit cases:

- $P(+1|\mathbf{x}) \rightarrow 1$  if  $s(\mathbf{x}) \rightarrow \infty$  or  $d_+(\mathbf{x}) \rightarrow 0$
- $P(+1|\mathbf{x}) \rightarrow 0$  if  $s(\mathbf{x}) \rightarrow -\infty$  or  $d_-(\mathbf{x}) \rightarrow 0$
- $P(+1|\mathbf{x}) \rightarrow 0.5$  if  $s(\mathbf{x}) \rightarrow 0$  and  $d_-(\mathbf{x}) \rightarrow d_+(\mathbf{x})$ .

The upper bound on the value of  $A$  ensures that if the effect of distances is neglected then  $P(+1|\mathbf{x}) > 0.95$  at the maximum SVM value and  $P(+1|\mathbf{x}) < 0.05$  at the minimum SVM value.

The training process for the PSVM is as follows:

- The value of  $d_+(\mathbf{x})$  for a  $+1$  sample and that of  $d_-(\mathbf{x})$  for a  $-1$  sample are actually zero. However, a zero value in (18) will result in  $P(+1|\mathbf{x}) = 0$  at  $-1$  samples and  $P(+1|\mathbf{x}) = 1$  at  $+1$  samples, irrespective of the values of  $A$  and  $B$ . To avoid this, during the training process,  $d_+(\mathbf{x})$  and  $d_-(\mathbf{x})$  are set equal to the distances to the closest neighboring samples belonging to the  $+1$  and  $-1$  classes.
- The values of  $s(\mathbf{x})$ ,  $d_-(\mathbf{x})$  and  $d_+(\mathbf{x})$  at training samples are used to calculate the likelihood function, which is then maximized to find the values of  $A$  and  $B$ . The construction and maximization of the likelihood function is identical to Lin et al. (2007), except that the probabilities are calculated using the modified sigmoid model.

In summary, the advantages of the new PSVM model are:

- It considers the spatial distribution of the training samples, in addition to the SVM values.
- It provides a more intuitive measure of  $P(+1|\mathbf{x})$ .
- It avoids bias towards one of the classes due to unequal number of training samples.

Note that the probability of misclassification based on (17) can actually be larger than 0.5. This can happen if, for instance, the  $P(+1|\mathbf{x}) = 0.5$  is locally within the  $-1$  class (Fig. 5).

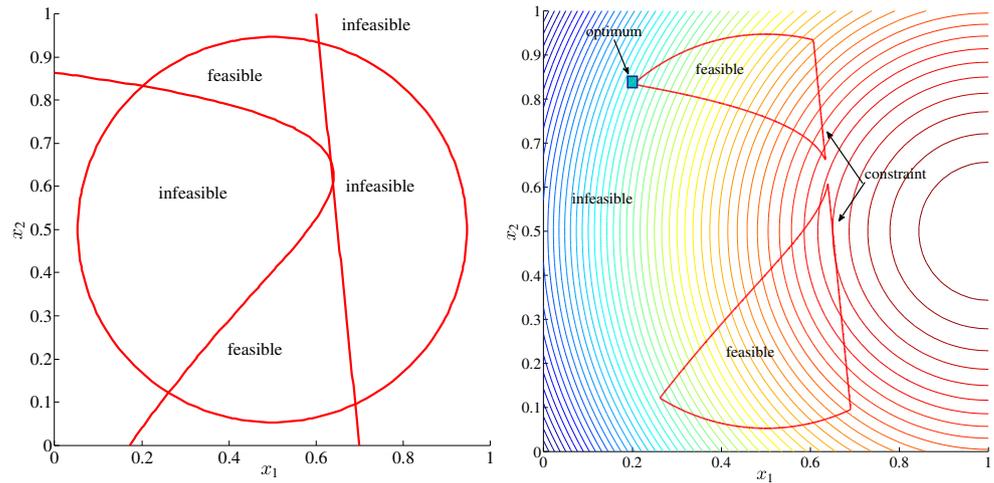
### 5 Examples

The results section is composed of four examples belonging to four distinct categories:

- A two-dimensional analytical example with three constraints. For this problem, results from existing methods based on the Kriging approximation of the constraints are available (Sasena 2002), and are used for comparison. This example is also used to analyze the differences between the new PSVM model and the Platt model. Another analytical example is provided in the Appendix B.
- A 99 constraint two-dimensional analytical problem to demonstrate the efficiency of the approach on problems with numerous constraints.
- A problem with binary constraints. Comparison with the random forest classifier (Lee et al. 2010).
- A five-dimensional aeroelasticity problem with binary responses.

For all the analytical examples, a comparison of the methods using the unconstrained and constrained

**Fig. 6** Example 1. Actual individual constraint contours and the resulting feasible and infeasible regions (*left*). Objective function contours, feasible space boundary, and global optimum solution (*right*)

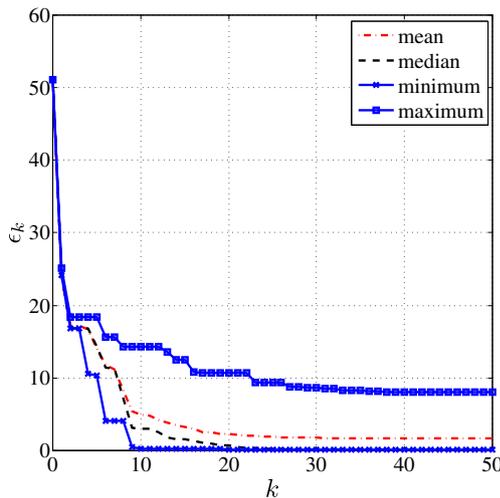


formulations (update schemes 1 and 2) is provided. For the aeroelasticity application example, the update scheme 2 is used (14). For each example, unless specified otherwise, the initial design of experiments consists of 10 CVT samples and the update is run for a fixed number of iterations to study the convergence. For a number of dimensions  $m$ , each iteration consists of  $m + 1$  samples evaluated in parallel (one  $EI$  sample and  $n_p = m$  auxiliary samples selected within a local update region). A genetic algorithm (GA) is used to solve the global optimization subproblems for locating the samples ((13), (14) and (16)).

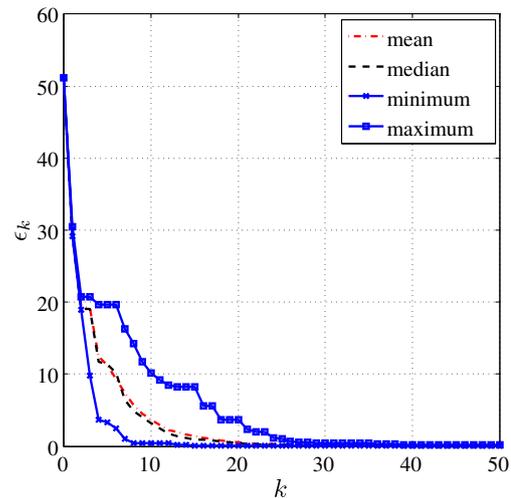
The following notation is used:

- $\mathbf{x}^*$  and  $f^*$  are the current optimum and the corresponding objective function value obtained by the algorithm.
- $\mathbf{x}_{\text{actual}}^*$  and  $f_{\text{actual}}^*$  are the actual optimum (analytical solutions) and the corresponding objective function value.
- $\epsilon_k$  is the relative percentage error of the optimum objective function value at the  $k^{\text{th}}$  iteration

$$\epsilon_k = \frac{|f^* - f_{\text{actual}}^*|}{|f_{\text{actual}}^*|} \times 100 \tag{19}$$

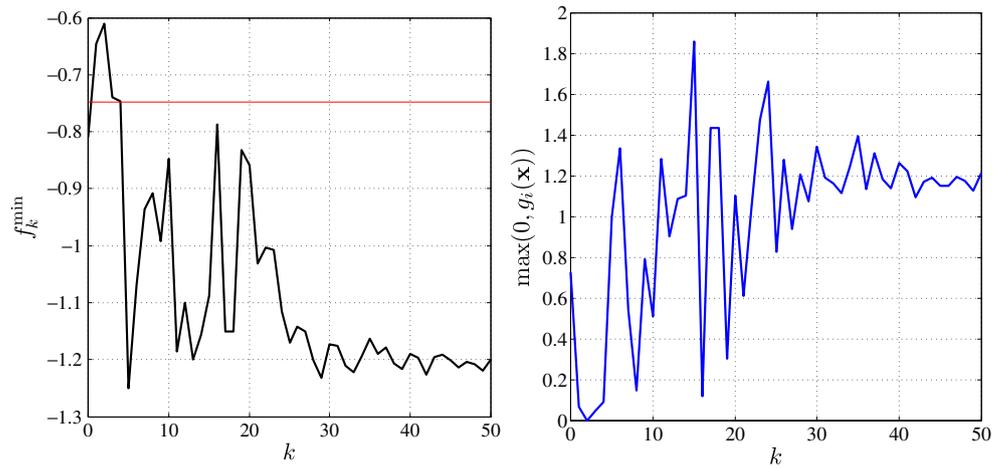


**Fig. 7** Example 1. Evolution of  $\epsilon_k$  with unconstrained formulation for 50 runs

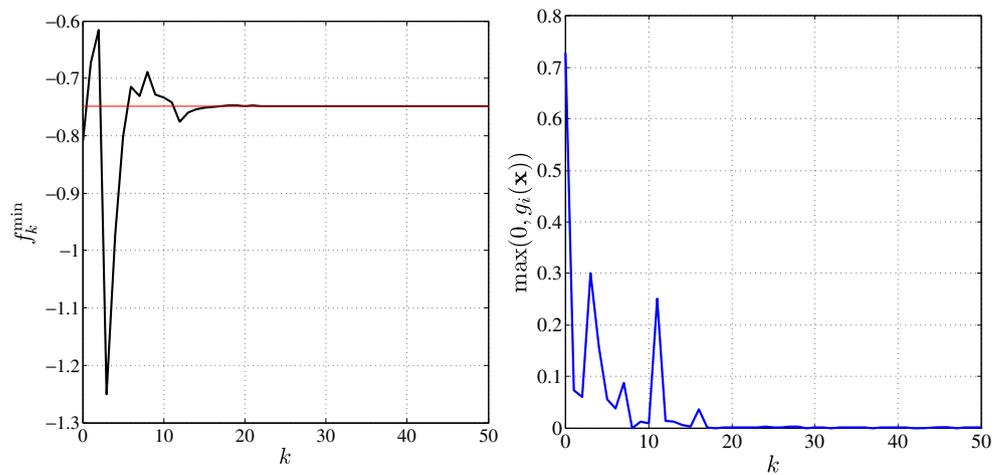


**Fig. 8** Example 1. Evolution of  $\epsilon_k$  with constrained formulation for 50 runs

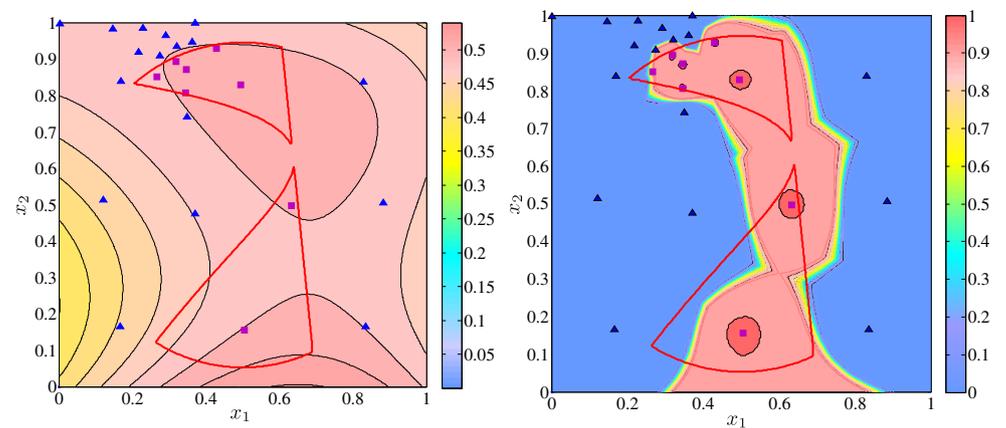
**Fig. 9** Example 1. Unconstrained formulation. Minimum objective function value among the  $m + 1$  samples in an iteration, irrespective of feasibility (*left*) Maximum constraint violation at the corresponding sample (*right*). Note that the optimum at iteration  $k$  need not be selected from the samples at that iteration; it can be one of the previous samples

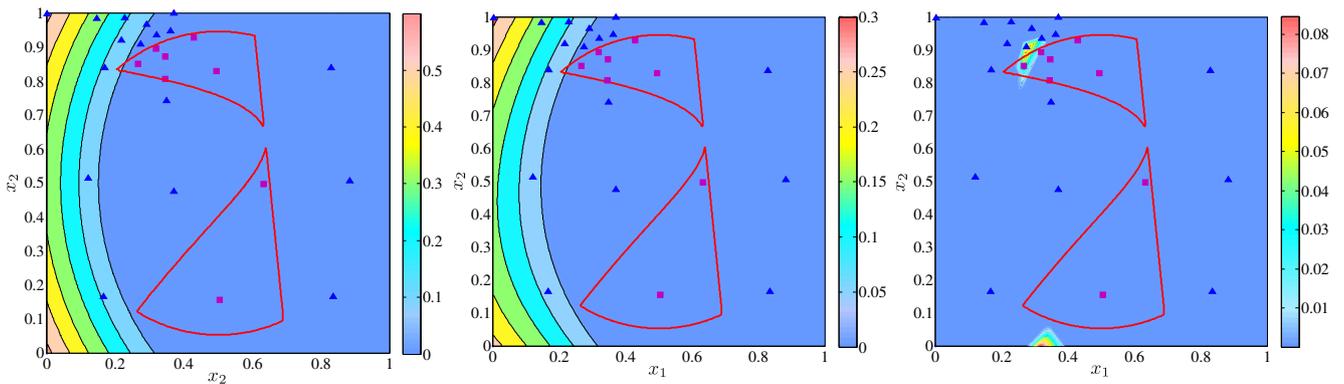


**Fig. 10** Example 1. Constrained formulation. Minimum objective function value among the  $m + 1$  samples in an iteration, irrespective of feasibility (*left*). Maximum constraint violation at the corresponding sample (*right*). Note that the optimum at iteration  $k$  need not be selected from the samples at that iteration; it can be one of the previous samples



**Fig. 11** Example 1. Map of probability of feasibility at a specific iteration using the basic sigmoid PSVM model (*left*) and the modified model (*right*). The blue triangles and magenta squares represent the feasible and infeasible samples. The red curve is the actual constraint boundary





**Fig. 12** Example 1. Map of  $EI$  at a specific iteration (*left*). Corresponding map of  $EI(\mathbf{x})P(+1|\mathbf{x})$  using the basic sigmoid PSVM model (*center*) and the modified model (*right*). The *blue triangles* and

*magenta squares* represent the feasible and infeasible samples. Note that the  $EI$  is calculated with respect to the current *feasible* optimum, and it is possible to have non-zero  $EI$  at an evaluated sample

- $f_k^{\min}$  is the minimum objective function value among the  $n_p + 1$  samples evaluated at iteration  $k$ , irrespective of their feasibility.

5.1 Example 1. Problem with three constraints

This example taken from Sasena (2002) consists of two variables  $x_1$  and  $x_2$  with identical ranges  $[0, 1]$ . The optimization problem is:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = -(x_1 - 1)^2 - (x_2 - 0.5)^2 \\ \text{s.t.} \quad & g_1(\mathbf{x}) = ((x_1 - 3)^2 + (x_2 + 2)^2)e^{-x_2^7} - 12 \leq 0 \\ & g_2(\mathbf{x}) = 10x_1 + x_2 - 7 \leq 0 \\ & g_3(\mathbf{x}) = (x_1 - 0.5)^2 + (x_2 - 0.5)^2 - 0.2 \leq 0 \quad (20) \end{aligned}$$

A graphical representation of the problem is provided in Fig. 6. The problem has two optima at  $(0.2316, 0.1216)$  and  $(0.2017, 0.8332)$ . The latter one is the global optimum with an objective function value equal to  $-0.7483$ .

Based on the 10 sample DOE, the initial relative error in  $f^*$  is 51.2%. To account for the variability due to the GA, the unconstrained and constrained formulations are executed 50 times. The mean, median, minimum, maximum errors  $\epsilon_k$  at each iteration are calculated. The evolution of  $\epsilon_k$ , using the two formulations, are plotted in Figs. 7 and 8. It should be noted that the optimum objective function value at iteration  $k$  is selected from all the feasible samples evaluated until that iteration. With the  $EI$  sample selected using the constrained formulation, all 50 runs lead to the global optimum. When the unconstrained formulation (product) is used to select the  $EI$  sample, the global optimum is found 80% of the times, i.e. 40 times out of the 50 runs. At

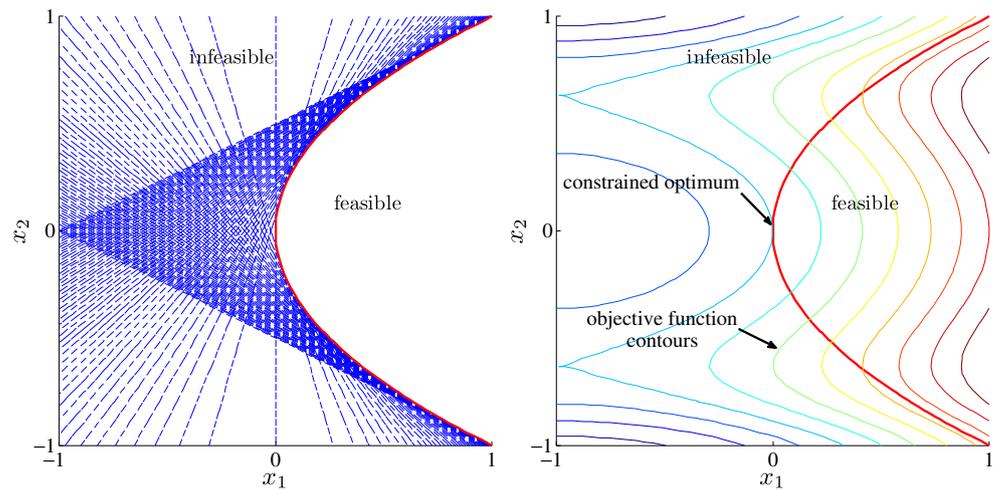
each iteration, the minimum function value among the  $n_p + 1$  samples selected at that iteration  $f_k^{\min}$  and the corresponding constraint violation are also plotted (Figs. 9 and 10).

In order to give a comparison of the PSVM models, the map of  $P(+1|\mathbf{x})$  is plotted in Fig. 11. It shows that the basic sigmoid model predicts unreasonably high probability of feasibility in many parts of the space. Even in regions with infeasible samples (blue triangles), it predicts a high probability of feasibility. This is not the case with the modified model. The maps representing  $EI$  and the product of  $EI$  times the probability of feasibility are also depicted in Fig. 12. For the modified PSVM model, the regions with high values of the product are close to the optimum, but this is not the case with the basic sigmoid model. As a result, using the basic sigmoid may lead to waste of several samples. In fact, Fig. 12 (center) indicates that the use of the basic sigmoid model will lead to repeated sampling in the upper left corner of the space, which is not the optimum.

**Table 1** Example 1. Comparison with purely Kriging-based approach (Sasena 2002). Distance of  $\mathbf{x}_{\text{actual}}^*$  to the solution found after evaluating 60 samples (10 from initial DOE)

Method	$\ \mathbf{x}^* - \mathbf{x}_{\text{actual}}^*\ $
Kriging $Probability_v$ (Sasena 2002)	$2.2 \times 10^{-4}$
Kriging $Probability_s$ (Sasena 2002)	$2.2 \times 10^{-4}$
Kriging $mean_v$ (Sasena 2002)	$2.8 \times 10^{-5}$
Kriging $mean_s$ (Sasena 2002)	$2.2 \times 10^{-4}$
Kriging $EV_v$ (Sasena 2002)	$1.8 \times 10^{-1}$
Kriging $EV_s$ (Sasena 2002)	$2.5 \times 10^{-1}$
SVM scheme 1	$9.4 \times 10^{-3}$
SVM scheme 2	$4.8 \times 10^{-3}$

**Fig. 13** Example 2. Depiction of the actual constraint of the original problem and the 99 constraints used to approximate it (left). Objective function contours, constraint and the optimal solution (right)

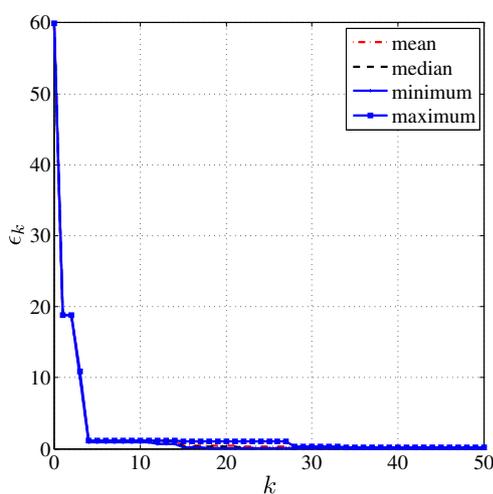


For completeness, the results for example 1 are also compared to those using Kriging approximations of the constraints (Sasena 2002). The results presented in Sasena (2002) were generated using several different methods and the reader is referred to Sasena (2002) for a more detailed description of the approaches. In one method, each constraint response is approximated using a separate Kriging model (denoted using subscript  $v$  in Table 1). In the second method, only one Kriging model is used to approximate the maximum of constraint responses (denoted using subscript  $s$  in Table 1). These approaches are further categorized based on the sampling criterion. The first formulation uses the probability of feasibility times the expected improvement (Schonlau 1997). The second one involves the constrained maximization of the expected improvement. The constraints are based on the mean values of the Kriging models for the constraints. The final method is based on the expected violation ( $EV$ ) of the constraints (Audet et al. 2000). For each case, the distance between the actual optimum and

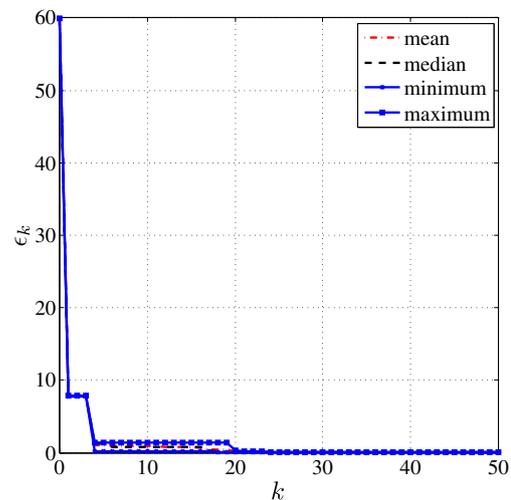
the optimum found by the algorithms is provided after 50 adaptive sample evaluations. These values are compared to those using the proposed SVM-based methods after the same number of evaluations. It should be noted that, in the proposed method, the samples are evaluated in parallel. Therefore, the iteration number at which the comparison is performed is much less. In addition to the mean values listed in Table 1, the best results using the two SVM-based schemes are also noted: the minimum errors are  $2.2 \times 10^{-4}$  and  $8.6 \times 10^{-5}$  using the update schemes 1 and 2.

### 5.2 Example 2: 99 constraint problem

The objective of this problem is to demonstrate how the proposed method handles several constraints using a single SVM boundary without the need to evaluate all the constraints at each iteration. This is a significant advantage that stems from the classification approach: if a sample violates any one of the constraints then it can be classified

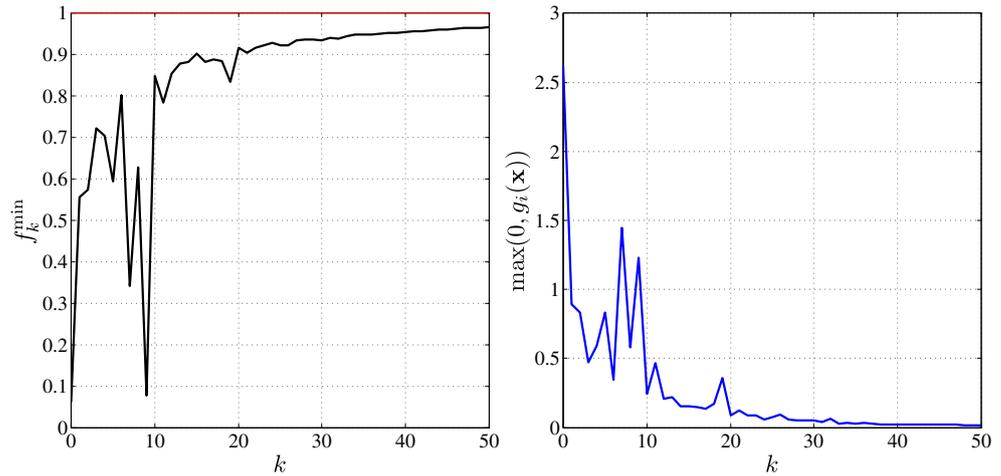


**Fig. 14** Example 2. Evolution of  $\epsilon_k$  with unconstrained formulation



**Fig. 15** Example 2. Evolution of  $\epsilon_k$  with constrained formulation

**Fig. 16** Example 2. Unconstrained formulation. Minimum objective function value among the  $m + 1$  samples in an iteration, irrespective of feasibility (*left*). Maximum constraint violation at the corresponding sample (*right*)



as infeasible. Therefore, it is not necessary to evaluate the other constraints. This advantage is particularly marked if the constraints are associated with different simulations.

A problem with 99 constraints is constructed from the following optimization with two variables  $x_1$  and  $x_2$  with ranges  $[-1, 1]$ :

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = (x_1 + 1)^2 + \sin(4x_2^2) \\ \text{s.t.} \quad & g(\mathbf{x}) = x_2^2 - x_1 \leq 0 \end{aligned} \tag{21}$$

The objective function and the feasible region, bounded by a parabola, are plotted in Fig. 13. In order to approximate the feasible space, 99 constraints are constructed as the tangents to the original constraint at 99 distinct locations on the parabola (Fig. 13, left). The actual optimum is located at  $[0, 0]$  with an objective function value equal to 1.0.

The initial relative error in the optimum objective function value is 59%. The statistics of the error over 5 runs are depicted in Figs. 14 and 15. In addition, the value of  $f_k^{\min}$

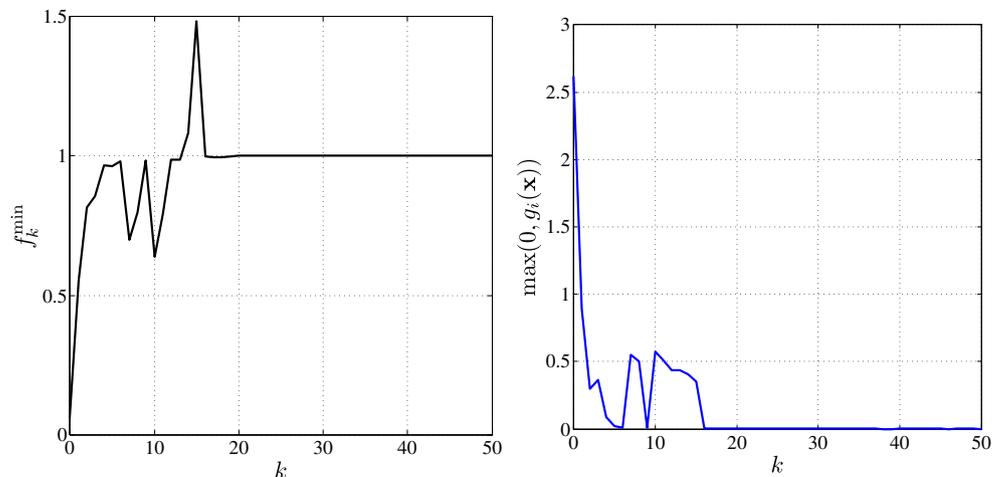
and the constraint violation at the corresponding sample are also plotted (Figs. 16 and 17).

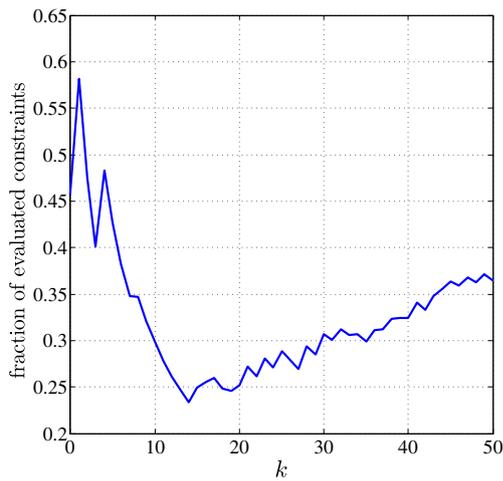
Note that it is possible to reorder the constraints based on their assumed criticality. More precisely, this is done by choosing the constraints as follows: the first constraint to evaluate is the first violated constraint by the closest infeasible sample. If this constraint is actually feasible, the second constraint to evaluate is the first violated constraint for the second closest sample. This process is repeated until a constraint is actually violated. As a result of this procedure, only a fraction of the total number of constraints are evaluated. The cumulative fraction,  $cf$ , over all iterations is calculated as follows:

$$cf = \frac{\text{actual number of constraint evaluations until iteration } k}{\text{number of constraints} \times \text{number of samples until iteration } k} \tag{22}$$

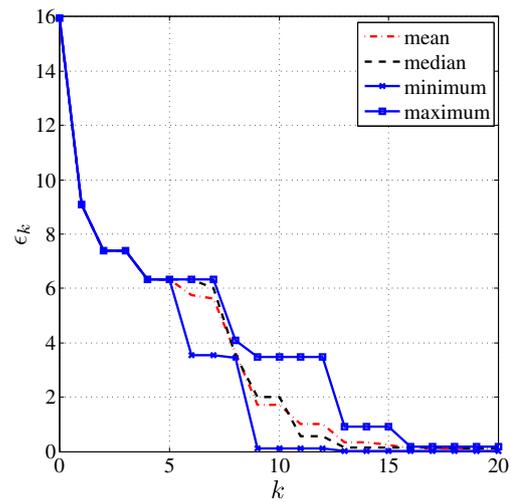
The evolution of the fraction is depicted in Fig. 18.

**Fig. 17** Example 2. Constrained formulation. Minimum objective function value among all samples in an iteration, irrespective of feasibility (*left*). Maximum constraint violation at the corresponding sample (*right*)





**Fig. 18** Example 2 with 99 constraints. Cumulative fraction of evaluated constraints



**Fig. 20** Example 3. Evolution of  $\epsilon_k$  with unconstrained formulation

5.3 Example 3. Problem with binary constraint — comparison with random forest classifier (Lee et al. 2010)

This example presents a binary case, which is one of the main strengths of the proposed method. This example is taken from Lee et al. (2010), which recently also addressed the issue of binary constraints. In Lee et al. (2010), a random forest classifier was used to handle the constraints, whereas an SVM is used in the proposed method. The important difference in this example compared to previous ones is that feasibility is defined based on whether the objective func-

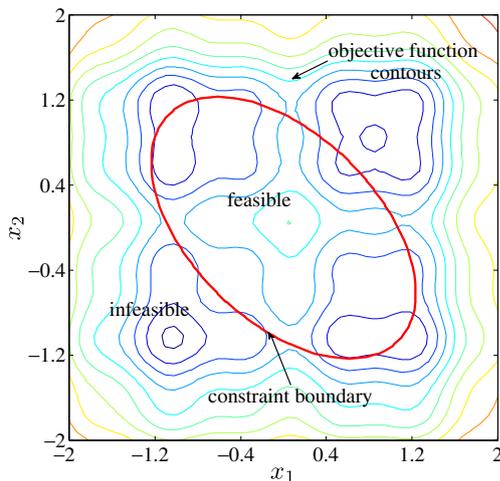
tion returns a value at a sample. The objective function to be minimized is:

$$f(x_1, x_2) = -c(x_1)c(x_2)$$

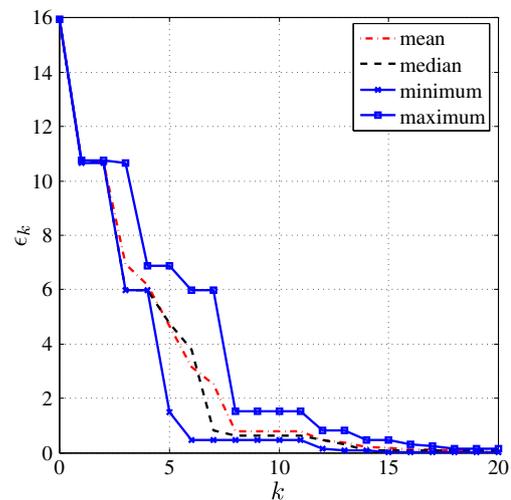
$$\text{where } c(x) = \exp(-(x - 1)^2) + \exp(-0.8(x + 1)^2) - 0.05 \sin(8(x + 0.1))$$

(23)

The feasible space is an ellipse defined by the 0.95 contour of a bivariate normal probability density function with mean at the origin, correlation coefficient  $-0.5$  and standard deviation  $0.75$  (Fig. 19). The objective function is not defined at samples outside the ellipse. The constrained optimization has two local minima, each with a function



**Fig. 19** Example 3. Objective function contours and the constraint boundary



**Fig. 21** Example 3. Evolution of  $\epsilon_k$  with constrained formulation

**Table 2** Example 4. Design variables

Name	Lower bound	Upper bound	Optimum
Sweep angle $\theta$	$-60^\circ$	$60^\circ$	$22.69^\circ$
Taper ratio $\lambda = \frac{c_t}{c_r}$	0.25	1	0.3393
Semi span $b/2$	80 in	160 in	88.10 in
Thickness parameter $k_1$	0 in	2 in	0.3342 in
Thickness parameter $k_2$	-5	0	-0.4447

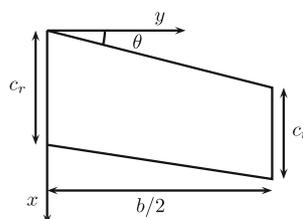
value of  $-1.0916$ . 20 samples are used for the initial DOE. The Kriging approximation of the objective function is constructed only using the feasible samples (because the objective function is considered undefined at the infeasible ones).

The errors of the proposed method with unconstrained and constrained formulations are plotted in Figs. 20 and 21. The initial error is 16%. Using the constrained formulation, the error is less than 1% at about 8 iterations, i.e. using 24 adaptively selected samples. Using the unconstrained formulation, 11 iterations or 33 samples are required for 1% error. In comparison, in Lee et al. (2010), the error was close to 1% at 17 iterations, starting from an error of approximately 7%. In Lee et al. (2010), an iteration was completed when a feasible sample was found. On the contrary, no such distinction is made in the iteration count of the proposed approach, and therefore, an exact comparison is not possible. Although 3 samples are evaluated at each iteration in this work, the total number of samples are comparable. In Lee et al. (2010), the final optimum was found after 47 iterations (feasible samples). With the proposed method (both constrained and unconstrained formulations), the final solution is located at the 16<sup>th</sup> iteration in average, i.e. after 48 samples are added. Also, it should be noted that the parallel nature of the proposed method gives it an additional advantage if multiple processors are available.

5.4 Example 4. Five variable aeroelasticity example with binary output

In this problem, the geometry of a simplified aluminium wing is optimized to minimize its weight while satisfying two stability requirements. A design is considered feasible if no flutter or divergence instabilities occur for given

**Fig. 22** Example 4. For a given wing area, the plan form of the wing is given by three variables: Sweep angle  $\theta$ , taper ratio  $\lambda = \frac{c_t}{c_r}$ , semi-span  $b/2$  (see Table 2)



**Table 3** Example 4. Fixed parameters used in the aeroelasticity problem

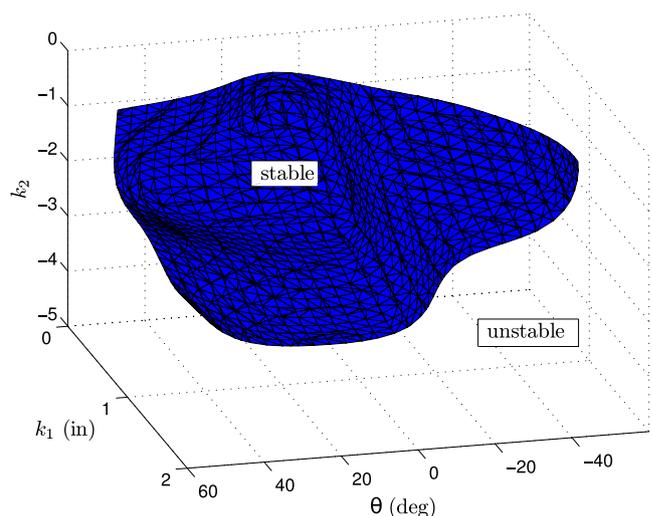
Mach number	0.5
Altitude	10,000 ft
Velocity	538.68 ft/s (match point)
Air density	$3.2686 \times 10^{-5}$ lb/in <sup>3</sup> (match point)
Angle of attack	0
Wing area $S$	8 ft <sup>2</sup>
Young's modulus	$9.2418 \times 10^6$ psi
Shear modulus	$3.4993 \times 10^6$ psi
Density	0.097464 lb/in <sup>3</sup>

flight conditions. This binary constraint (stable/unstable) is evaluated through an aeroelasticity code ZAERO (Zona Technology, Inc. 2008). Note that one single SVM is needed for the two failure modes (flutter and divergence).

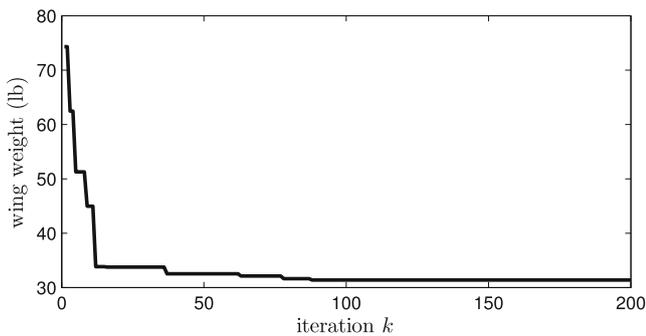
There are a total of five variables (see Table 2). The plan form of the wing is defined by three variables: Sweep angle  $\theta$ , taper ratio  $\lambda$ , and span  $b$  (see Fig. 22). The area of the wing is maintained constant. The thickness of the wing, constant chord wise, is defined along the span using an exponential function with two parameters  $k_1$  and  $k_2$  (24), which are used as the two other variables of the problem.

$$t(y) = k_1 e^{\frac{2k_2 y}{b}} \tag{24}$$

Table 3 lists the fixed parameters required to evaluate each design, such as material properties and flight conditions. The flight conditions are given by the altitude and the Mach number.



**Fig. 23** Example 4. Example of stability boundary with respect to sweep angle and thickness parameters  $k_1$  and  $k_2$ . Taper ratio and semi span are fixed at 0.625 and 120 in. respectively

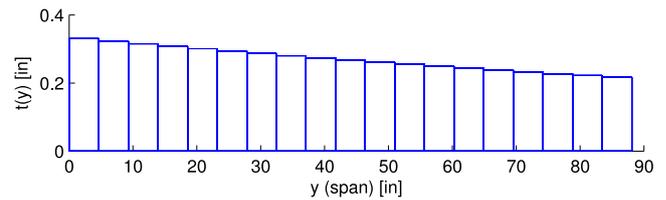
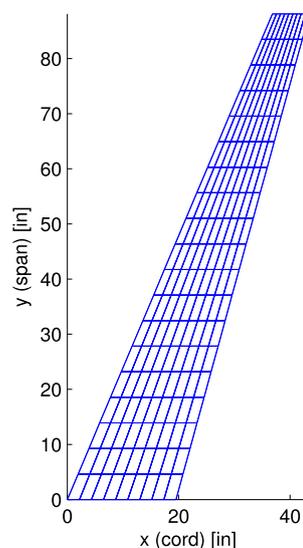


**Fig. 24** Example 4. Evolution of  $f_k^*$  value during the optimization

The structural finite element (FE) model is built using Genesis (Vanderplaats Research & Development, Inc. 2006), which provides the weight, mode shapes, and natural frequencies of the wing. This FE model consists of 171 CQUAD4 shell elements with nine elements along the chord. A flat panel aerodynamic model was built using ZAERO (Zona Technology, Inc. 2008) to check the stability at the given flight conditions. This model consists of 40 aero-boxes with five boxes along the chord. For each mode, a damping coefficient and a frequency are calculated. A positive damping value corresponds to flutter instability whereas a negative damping corresponds to a stable configuration. A zero frequency corresponds to divergence instability. Figure 23 depicts an arbitrary example of stability boundary in terms of sweep angle and the two thickness parameters  $k_1$  and  $k_2$ . The taper ratio and semi span are held constant to produce this figure.

In order to solve the optimization for this problem, the update scheme 2 with the constrained formulation was used. The initial DOE consists of 10 CVT samples. The update scheme was run for 200 iterations. Figure 24 shows the objective function value of the best feasible sample (i.e.

**Fig. 25** Example 4. Planform of the optimal wing



**Fig. 26** Example 4. Thickness distribution of the optimal wing

the minimum weight among feasible samples) during the optimization. Note that the optimum is reached before 100 iterations. The corresponding optimum weight is 31.39 lb. The optimal design is provided in Table 2. The corresponding planform and optimal thickness distributions are depicted in Figs. 25 and 26, respectively.

## 6 Discussion

The proposed work is a first step for Kriging-based constrained optimization using SVMs. This section presents a discussion on the merits, limitations, and avenues of improvements of the proposed approaches. This discussion is based on the results presented in Section 5 but also extends to broader comments.

### 6.1 Comparison between the two update schemes

The comparison of the unconstrained and constrained formulations in this article seems to indicate that the latter is preferable in terms of the convergence rate. This conclusion is valid on all the presented results but is made clear on Example 1. In this example, two local optima corresponding to disjoint feasible regions are present. The update scheme 2 with constrained formulation (Fig. 8) provided the global optimum for each of the 50 runs. This was not the case with the unconstrained formulation which reached the global optimum 40 times out of 50 runs. The maximum error in the plot of the relative error  $\epsilon_k$  (Fig. 7) corresponds to a case in which the update scheme 1 with unconstrained formulation reached a local optimum. It should be noted that this represents convergence to the local optimum, and not failure to converge. The reason for finding the local optimum in some of the cases is that for the initial approximation, both the regions have similar  $EI$  and probability of feasibility. However, as the update progresses, there is a significant region of the space where  $EI$  is equal to zero, and the regions with non-zero  $EI$  have a low probability of feasibility. It may be possible to overcome this issue if the generalized expected improvement with a higher exponent (more global search) is used. This, however, has not been investigated in this study as the constrained formulation led to satisfactory results. Overall, from the obtained results,

the constrained formulation appears to be more robust and efficient. This conclusion reinforces what had already been observed and commented in the literature: the unconstrained “product” formulation does not enforce enough feasibility and leads to arbitrary dominance of one of the two terms. As a consequence, the unconstrained formulation is more likely to evaluate samples in the infeasible space than the constrained formulation. At times, the unconstrained formulation might also simply fail as demonstrated in another example in the Appendix B.

It is also noteworthy that the results from Example 1 were compared to those using Kriging-based constraint approximations from the literature. This comparison indicates that the proposed approach has a comparable efficacy. However, this is true for the small scale problem 1 with 3 constraints. As developed in the following paragraphs, the strength of the proposed approach stems from unique aspects of the SVM-based approach.

## 6.2 Discontinuous and binary responses

SVM was introduced originally by the authors to tackle problems with discontinuities and binary outputs. In this article, the ability of the proposed classification-based approach to handle binary responses is extended to optimization and is demonstrated using Examples 3 and 4. In Example 3, the proposed method is compared to another recent approach based on random forest classifiers. Example 4 presents an application in the field of aeroelasticity with binary behaviour (stable/unstable).

## 6.3 The case of multiple constraints/failure modes

The usefulness of the proposed methodology for handling multiple constraints is expected because it reduces the problem by replacing all the constraints with a single SVM approximation. Because the proposed method is based on classification, it avoids the evaluation of all the constraints at each iteration. The proposed classification-based method has the potential to reduce computational burden if the constraints are evaluated using different solvers. This is demonstrated in Example 2 with 99 constraints. In a typical Kriging-based approach, this would require an approximation for each constraint.

## 6.4 Correlated constraints

In engineering problems, constraints are often correlated. Therefore, in the traditional Kriging-based framework, the calculation of the probability of feasibility requires some explicit knowledge about the correlation. Traditionally, the product of the probability of feasibility is used, thus assuming that the constraints are independent. In a classification

framework, this requirement is no longer needed because only one SVM constraint is used which implicitly contains the correlation information.

## 6.5 Probability of failure calculation

One of the key attributes of the SVM-based explicit design space decomposition stems from the construction of a locally accurate explicit limit state function. This is made possible by the adaptive sampling scheme developed by the authors (the auxiliary samples). This explicit boundary can be used for calculating failure probabilities through Monte-Carlo simulations. This had been demonstrated in earlier work by the authors Basudhar et al. (2008) and Basudhar and Missoum (2009).

The following potential limitations should be noted:

## 6.6 Local vs. Global optimum

As mentioned earlier, the unconstrained formulation failed at times to locate the global optimum. However, even in the case of the constrained formulation, it is noteworthy that there is, in general, no guarantee to converge to the global optimum. One of the possible reasons stems from a PSVM model that might exclude a significant region of the feasible domain and prevent the optimization to converge to the global optimum. For this reason, the technique might fail to locate the global optimum in cases where, for instance, the global optimum lie in a feasible region of relatively small size.

## 6.7 PSVM model

Although the modified PSVM model provides an improvement over the Platt’s model, some aspects need to be refined. For instance, the proposed PSVM model might not be smooth enough. As mentioned in the previous remark, the proposed PSVM model might artificially exclude parts of the search space. Other PSVM models or different techniques to estimate the PSVM parameters should be investigated. As an example of metrics that could improve the model, one could include factors such as the density of samples in a region or the number of samples belonging to a specific class in the vicinity.

## 6.8 Avenues for improvement

Beside the development of an enhanced PSVM model, several other avenues for improvements can be envisioned. For continuous problems, one of the limitations of the classification approach is that it does not use zero or first order information and it does not differentiate between, for instance, a highly infeasible and an almost feasible sample.

In this article, the adaptive sampling scheme to refine the SVM proposed in sparse areas alleviates partly this issue. However, for continuous problems, it may also be useful to use the information about the value of the actual constraint violation and gradient information. For example, it may be advantageous to use both the data from a response approximation (e.g., a metamodel) and the SVM classifier. Another possibility for improvement lies in the exploration of the effect of alternate sampling criteria that have been used in the constrained EGO literature, such as the generalized *EI* (Schonlau 1997; Sasena 2002) and different correlation kernels, as well as to investigate the evaluation of several *EI* samples in parallel (Ponweiser et al. 2008; Ginsbourger et al. 2007). Modifications to the method are also possible in terms of the construction of the SVM. In this study, hard classification is used for SVMs, which is allowed due to the fact that we are using computer experiments. However, it is possible to use soft margins with SVMs, and to allow some training misclassification (Vapnik 1998).

All in all, the performance of the proposed method seems quite satisfactory based on the example problems. Its ability to address binary responses and multiple constraints was demonstrated. Its applicability to an engineering example with binary output was also demonstrated. It may be interesting to compare the proposed method with indicator Kriging, although SVM seems a more natural way to handle binary problems, and is widely recognized by the computer science community.

## 7 Concluding remarks

A method for constrained global optimization using SVM is presented in this article. Two formulations are proposed and compared in the results section. Both formulations involve the expected improvement of the objective function and a probability of misclassification calculated using a new

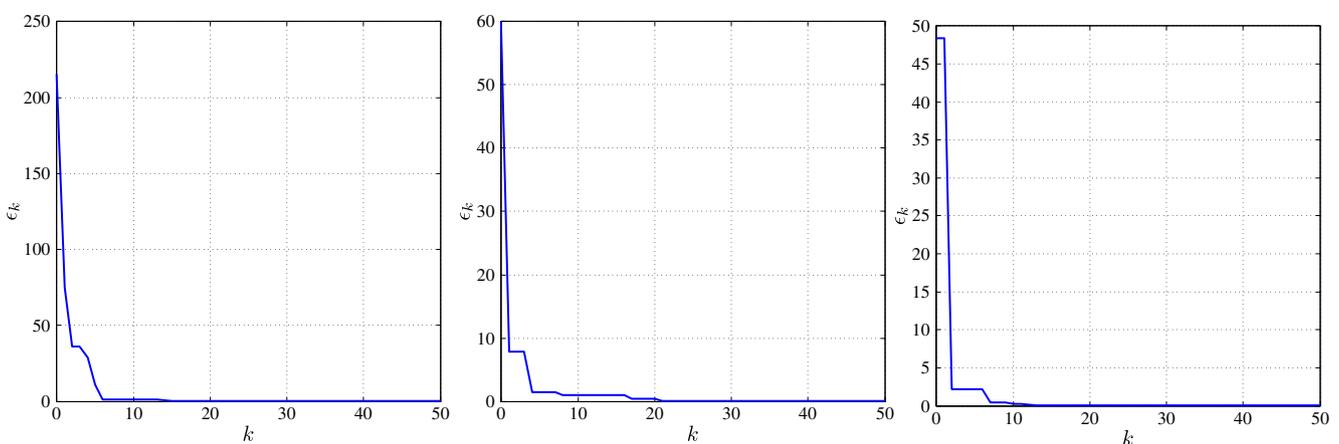
PSVM model. The SVM is refined locally with a dedicated adaptive sampling scheme. It was concluded that the approach based on the constrained formulation (update scheme 2) is the most robust. The efficacy of the method was demonstrated using several analytical examples, as well as an engineering application dealing with aeroelasticity. Particular emphasis is placed on the ability of the approach to handle multiple constraints, discontinuous, and binary constraint responses. Another noteworthy advantage of the approach is that it can handle correlated constraints.

The methodology will be extended to perform reliability-based design optimization. The local update used in this work provides an accurate SVM around the optimum and thus, is expected to be advantageous for failure probability calculation. In addition, the approach will be extended to include models of various fidelity to improve computational efficiency.

**Acknowledgments** The support of the National Science Foundation (award CMMI-1029257) and the partial support of the Air Force Office of Scientific Research (Grant FA9550-10-1-0353) is gratefully acknowledged. We are also thankful to Dr. Jean-Marc Bourinet (Institut Français de Mécanique Avancée, France) for sending Sylvain Lacaze as an exchange student.

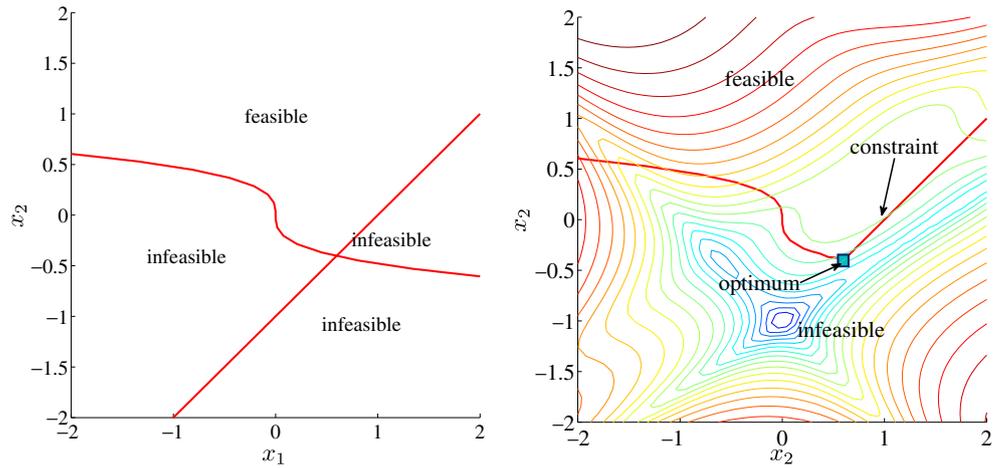
## Appendix A: Effect of the design of experiments size

This section presents an example of the effect of initial design of experiments on the optimization results, for Example 2 (Section 5.2). Three sets of initial DOEs are used, consisting of 5, 10 and 15 CVT samples to run the optimization. The percentage relative errors in the objective function value  $\epsilon_k$  are plotted in Fig. 27 for the three cases. The constrained formulation (update scheme 2) is used. It is seen that for all the cases, the optimization converges within 21 iterations.



**Fig. 27** Example 2. Evolution of the relative error  $\epsilon_k$  with constrained formulation with different initial DOE sizes: 5 (left), 10 (center), 15 (right)

**Fig. 28** Appendix B example. Individual constraints and the resulting feasible and infeasible regions (left). Objective function contours, boundary of feasible space and the optimum solution (right)



**Appendix B: Two constraint problem**

This appendix provides another analytical example. The constrained and unconstrained formulations are compared. The example is taken from Sasena (2002), and consists of two variables  $x_1$  and  $x_2$  in the range  $[-2, 2]$ . The feasible space for this problem is bounded by two constraints. The optimization problem is:

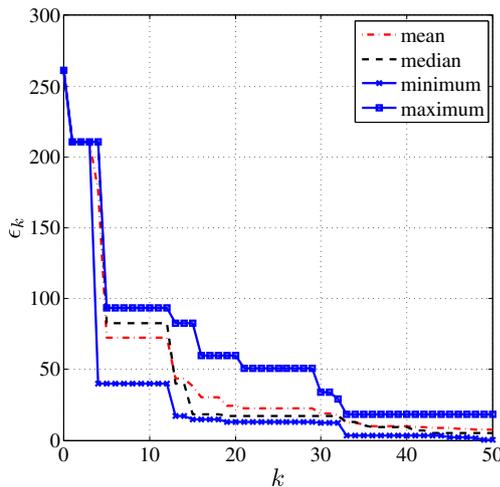
$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = (1 + A(x_1 + x_2 + 1)^2)(30 + B(2x_1 - 3x_2)^2) \\ \text{where} \quad & A = 19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2, \\ \text{and} \quad & B = 18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2 \\ \text{s.t.} \quad & g_1(\mathbf{x}) = -3x_1 + (-3x_2)^3 \leq 0 \\ & g_2(\mathbf{x}) = x_1 - x_2 - 1 \leq 0 \end{aligned} \tag{25}$$

The objective function, the constraints and the optimum solution are depicted in Fig. 28. The actual optimum is

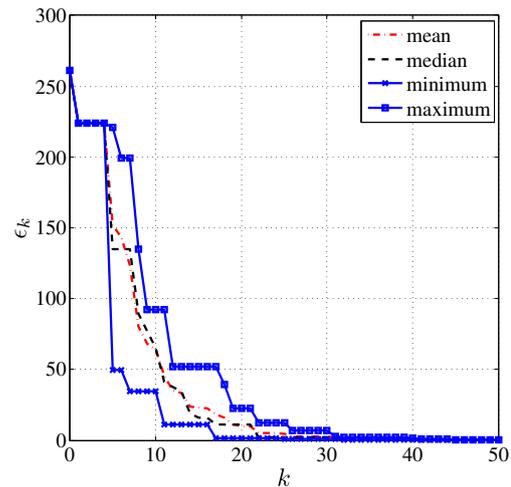
at  $(0.5955, -0.4045)$  with an objective function value of 289.85.

The initial error with 10 CVT samples is 261.0%. The evolution of  $\epsilon_k$  using the unconstrained and constrained formulations are plotted in Figs. 29 and 30. In order to check the consistency of the results, the algorithm was run 50 times. The mean, median, minimum and maximum errors are provided.

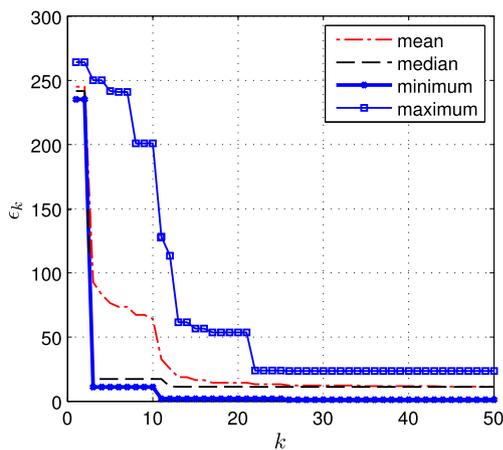
As a comparison, this example was also run using a purely Kriging-based approach where the two constraints are also approximated by Kriging. As described in Schonlau (1997) and Forrester et al. (2008), the approach is based on the product of  $EI$  times the probabilities of feasibility of each constraint ( $EI(\mathbf{x})P(g_1(\mathbf{x}) \leq 0)P(g_2(\mathbf{x}) \leq 0)$ ). The run was carried out using a freely available online code (Forrester et al. 2008). The statistics of results based on 50 runs are gathered in Fig. 31. It is observed that in the SVM-based unconstrained approach as well as the purely Kriging-based approach, the optimization does not always



**Fig. 29** Appendix B example. Evolution of  $\epsilon_k$  with unconstrained formulation. 50 runs



**Fig. 30** Appendix B example. Evolution of  $\epsilon_k$  with constrained formulation. 50 runs



**Fig. 31** Appendix B example. Evolution of  $\epsilon_k$  using A. Forrester's code (Forrester et al. 2008). 50 runs

converge accurately to the optimum. This is clearly demonstrated on Figs. 29 and 31. This is likely due to the fact that the gradient of the objective function is quite large around the optimum. On the other hand, the SVM-based constrained formulation (Fig. 30) does consistently reach the optimum accurately.

## References

- Alexandrov NM, Lewis RM, Gumbert CR, Green LL, Newman PA (2001) Approximation and model management in aerodynamic optimization with variable-fidelity models. *J Aircr* 38(6):1093–1101
- Arenbeck H, Missoum S, Basudhar A, Nikravesh PE (2010) Reliability-based optimal design and tolerancing for multibody systems using explicit design space decomposition. *J Mech Des* 132(2):021010
- Audet Jr C, Dennis JE, Moore DW, Booker A, Frank PD (2000) A surrogate based method for constrained optimization. In: 8th AIAA/NASA/USAF/ISSMO symposium on multidisciplinary analysis and optimization. Paper number AIAA-2000-4891
- Basudhar A, Missoum S (2008) Adaptive explicit decision functions for probabilistic design and optimization using support vector machines. *Comput Struct* 86(19–20):1904–1917
- Basudhar A, Missoum S (2009) A sampling-based approach for probabilistic design with random fields. *Comput Methods Appl Mech Eng* 198(47–48):3647–3655
- Basudhar A, Missoum S (2010) An improved adaptive sampling scheme for the construction of explicit boundaries. *Struct Multidisc Optim* 42(4):517–529
- Basudhar A, Missoum S, Harrison Sanchez A (2008) Limit state function identification using support vector machines for discontinuous responses and disjoint failure domains. *Probab Eng Mech* 23(1):1–11
- Bichon BJ (2010) Efficient surrogate modeling for reliability analysis and design. PhD thesis, Vanderbilt University
- Bichon BJ, Mahadevan S, Eldred MS (2009) Reliability-based design optimization using efficient global reliability assessment. In: Proceedings of the 50th AIAA/ASME/ASCE/AHS/ASC on structures, dynamics and materials conference. Paper AIAA-2009-2264. Palm Springs, California
- Chiles JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley-Interscience, New York
- Cressie N (1990) The origins of Kriging. *Math Geol* 22(3):239–252
- Cristianini N, Shawe-Taylor J (2006) An introduction to support vector machines: and other kernel-based learning methods. Cambridge University Press, Cambridge
- Forrester AIJ, Keane AJ (2009) Recent advances in surrogate-based optimization. *Prog Aerosp Sci* 45(1–3):50–79
- Forrester AIJ, Sobester A, Keane AJ (2008) Engineering design via surrogate modelling: a practical guide. Wiley, New York
- Ginsbourger D, Le Riche R, Carraro L (2007) A multi-points criterion for deterministic parallel global optimization based on Kriging. In: International conference on nonconvex programming, NCP07
- Gramacy RB, Lee HKH (2010) Optimization under unknown constraints. Arxiv preprint, [arXiv:1004.4027](https://arxiv.org/abs/1004.4027)
- Gunn SR (1998) Support vector machines for classification and regression. Technical Report ISIS-1-98, Department of Electronics and Computer Science, University of Southampton
- Henkenjohann N, Kunert J (2007) An efficient sequential optimization approach based on the multivariate expected improvement criterion. *Qual Eng* 19(4):267–280
- Huang D, Allen TT, Notz WI, Miller RA (2006) Sequential Kriging optimization using multiple-fidelity evaluations. *Struct Multidisc Optim* 32(5):369–382
- Jin R, Du X, Chen W (2003) The use of metamodeling techniques for optimization under uncertainty. *Struct Multidisc Optim* 25(2):99–116
- Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Glob Optim* 13(4):455–492
- Kleijnen JPC (2009) Kriging metamodeling in simulation: a review. *Eur J Oper Res* 192(3):707–716
- Lee H, Gramacy R, Linkletter C, Gray G (2010) Optimization subject to hidden constraints via statistical emulation. Technical report, Tech Rep UCSC-SOE-10-10, University of California, Santa Cruz, Department of Applied Mathematics and Statistics
- Lin HT, Lin CJ, Weng RC (2007) A note on Platt's probabilistic outputs for support vector machines. *Mach Learn* 68(3):267–276
- Martin JD, Simpson TW (2005) Use of Kriging models to approximate deterministic computer models. *AIAA J* 43(4):853–863
- Missoum S, Ramu P, Haftka RT (2007) A convex hull approach for the reliability-based design of nonlinear transient dynamic problems. *Comput Methods Appl Mech Eng* 196(29):2895–2906
- Picheny V, Kim NH, Haftka RT, Queipo NV (2008) Conservative predictions using surrogate modeling. In: Proceedings of the 49th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, Schaumburg, Illinois, USA, pp 7–10
- Platt JC (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. MIT Press, Cambridge, pp 61–74
- Ponweiser W, Wagner T, Vincze M (2008) Clustered multiple generalized expected improvement: a novel infill sampling criterion for surrogate models. In: IEEE congress on evolutionary computation, Hong Kong
- Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK (2005) Surrogate-based analysis and optimization. *Prog Aerosp Sci* 41(1):1–28
- Rasmussen CE, Williams CKI (2005) Gaussian processes for machine learning
- Santner TJ, Williams BJ, Notz W (2003) The design and analysis of computer experiments. Springer, New York
- Sasena MJ (2002) Flexibility and efficiency enhancements for constrained global optimization with Kriging approximations. PhD

- thesis, Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI
- Sasena MJ, Papalambrose PY, Goovaerts P (2002a) Exploration of metamodeling sampling criteria for constrained global optimization. *Eng Optim* 34:263–278
- Sasena M, Papalambros P, Goovaerts P (2002b) Global optimization of problems with disconnected feasible regions via surrogate modeling. In: 9th AIAA/ISSMO symposium on multidisciplinary analysis and optimization. Citeseer
- Scholkopf B, Smola AJ (2002) *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge
- Schonlau M (1997) *Computer experiments and global optimization*. PhD thesis, Department of Statistics, University of Waterloo, Ontario, Canada
- Simpson TW, Toropov VV, Balabanov VO, Viana FAC (2008) Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come—or not. In: 12th AIAA/ISSMO multidisciplinary analysis and optimization conference, Victoria, British Columbia, Canada
- Stein ML (1999) *Interpolation of spatial data: some theory for Kriging*. Springer, New York
- Vanderplaats Research & Development, Inc (2006) *Genesis analysis manual version 9.0*
- Vapnik VN (1998) *Statistical learning theory*. Wiley, New York
- Viana FAC, Gogu C, Haftka RT (2010a) Making the most out of surrogate models: tricks of the trade. In: ASME 2010 international design engineering technical conferences & computers and information in engineering conference, Montreal, Canada
- Viana F, Haftka R, Watson LT (2010b) Why not run the efficient global optimization algorithm with multiple surrogates? In: Proceedings of the 51th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, AIAA, Orlando, FL, USA. AIAA–2010–3090
- Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. *J Mech Des* 129(4):370–380
- Zona Technology, Inc (2008) *Zaero version 8.3 theoretical manual*, 19th edn