Constraint handling in efficient global optimization

DOI:
10.1145/3071178.3071278

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
GECCO 2017 - Proceedings of the 2017 Genetic and Evolutionary Computation Conference

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Constraint Handling in Efficient Global Optimization

Samineh Bagheri1, Wolfgang Konen1, Richard Allmendinger2, Jürgen Branke3, Kalyanmoy Deb4, Jonathan Fieldsend5, Domenico Quagliarella6, and Karthik Sindhya7

1Cologne University of Applied Sciences (TH Köln), Gummersbach, Germany
2University of Manchester, Manchester, UK
3University of Warwick, Coventry, UK
4Michigan State University, East Lansing, USA
5University of Exeter, Exeter, UK
6Italian Aerospace Research Centre (CIRA), Capua, Italy
7University of Jyväskylä, Jyväskylä, Finland

1 INTRODUCTION

A constrained optimization problem (COP) can be defined as the minimization of an objective function (fitness function) \( f \) subject to inequality constraint function(s) \( g_1, \ldots, g_m \):

\[
\begin{align*}
\text{Minimize} & \quad f(\bar{x}), & \bar{x} \in [\bar{a}, \bar{b}] \subset \mathbb{R}^d \\
\text{subject to} & \quad g_i(\bar{x}) \leq 0, & i = 1, 2, \ldots, m
\end{align*}
\]

where \( \bar{a} \) and \( \bar{b} \) define the lower and upper bounds of the search space (a hyperrectangle).

Real-world COPs are often very expensive to evaluate which means only a very limited number of function evaluations is allowed in practice. Giannakoglou [9] presents an overview on efficient optimization approaches and shows indicative examples from aerodynamics. Therefore, a proper optimizer for this sort of problems should be able to find optimal or near-optimal solutions with a low number of function evaluations. Many different surrogate-assisted techniques have been developed to tackle expensive COPs [3, 4, 20, 22, 28]. The main idea adopted by nearly all surrogate-assisted optimization algorithms is to use fast mathematical or statistical models for the optimization process and only evaluate a new solution on the expensive function when the model needs to be updated. Among the existing multitude of surrogate modeling approaches with various properties, Kriging [8, 15] appears to be one of the most attractive techniques. Kriging, also known as Gaussian Processes, offers a strong modeling tool which provides estimation of the prediction uncertainty in addition to the approximation of the function. Although in recent years many studies have been undertaken with different Kriging-based unconstrained and constrained optimizers, there are still limitations associated with the use of Kriging models. One of the issues that almost all Kriging-based strategies face is that frequent crashes occur during the optimization process. The crashes usually happen when a new solution is located very close to a former one. We find that adding a noise variance is an effective way to handle such stability issues.

Another point is that most of the existing Kriging-based constrained optimization strategies are evaluated only on simple 2-dimensional benchmark functions mostly with only one active constraint [7, 10, 25–27]. On the other hand, real-world COPs are often multi-constrained and are not limited to 2-dimensional problems. This motivates us to develop a new Kriging-based optimization algorithm which avoids crashes and is applicable on challenging COPs. We evaluate the proposed algorithm on all benchmarks with dimension \( d \leq 4 \) taken from the challenging G-function suite [16]. Throughout this work we try to answer the following research questions:

\[(H1) \] Is it possible to modify existing Kriging-based optimization algorithms to handle challenging COPs with multiple active constraints?

\[(H2) \] Is it possible to balance the exploration of feasible and infeasible infill points in a proper way?

1.1 Related Work

Most Kriging-based constrained optimization algorithms make use of Kriging’s statistical property, expected improvement function [14, 17], for efficiently solving global COPs. We can categorize such algorithms into three main groups.

(a) The first group of algorithms transform a constrained problem into an unconstrained problem. Schonlau et al. [27] is an example of type (a) algorithm. This algorithm suggests to maximize the multiplication of the expected improvement of the objective function and the probability of feasibility, which are both statistical measures determined from Kriging models of objective and constraint functions. This algorithm fails if the number of active constraints are large or if the objective function is very steep or flat around the feasibility boundary. Algorithms which tend to maximize the penalized expected improvement [18, 25] also belong to the first group.

(b) Another approach to address COPs is to solve a constrained sub-problem. As an example, we can name a work from Sasena et al. [26] in which the expected improvement is maximized subject to the approximation of the constraint functions. Audet et al. [2] maximize the expected improvement subject to the expected violation of each constraint.

(c) Methods in the third category transform the constrained problems to multi-objective unconstrained problems and then use multi-objective optimizers. These methods often consider the expected improvement of the fitness function as one objective and one or more statistical properties of the constraint functions as other objective(s) [7, 12, 19]. Although Durantin et al. [7] show that the type (c) algorithms
perform better than the existing algorithms from type (a) and (b), we have to consider that solving a multi-objective problem is a complex task. An increase in problem dimension or number of active constraints makes such algorithms very time-consuming.

To the best of our knowledge, the current state-of-the-art for solving the G-problems is SACOBRA [3] (self adjusting COBRA). SACOBRA uses RBF surrogate models. SACOBRA is not a Kriging-based COP, it does not use the EI approach.

We propose in this paper a new type-(a) algorithm called SOCU (Surrogate-Assisted Optimization encompassing Constraints and Uncertainties).

2 METHODS

2.1 Kriging Surrogate Models

Kriging is a statistical modeling technique based on Gaussian processes. This algorithm approximates the function \( f(x) \) with the surrogate model

\[
Y = \mu + \epsilon(x),
\]

where \( \mu \) is the average of the stochastic process and the error term \( \epsilon(x) \) is normally distributed with mean 0 and variance \( \sigma^2(x) \). The estimation of \( \mu \) and \( \sigma \) are the heart of Kriging modeling and described in more detail in the standard Kriging literature [14, 15].

2.2 Expected Improvement with Constraints

Efficient global optimization (EGO) is an algorithm developed by Jones et al. [14] for unconstrained optimization based on Kriging. The main idea of the EGO algorithm – originally introduced by Mokus et al. [17] – is to balance between exploration and exploitation by maximizing the expected improvement in Eq. (4) during a sequential optimization process:

\[
EI(x) = E \left[ \max(f_{\text{min}} - Y, 0) \right] = \Phi \left( \frac{f_{\text{min}} - \mu(x)}{\sigma(x)} \right) + \Phi \left( \frac{f_{\text{min}} - \mu(x)}{\sigma(x)} \right),
\]

where the \( f_{\text{min}} \) is the fitness value of the best found solution and \( \Phi \) and \( \varphi \) are the cumulative and probability density function of the standard normal distribution. Schonlau et al. [27] extended the EGO algorithm to handle inequality constraints. Their algorithm maximizes the penalized expected improvement function \( ELP \), shown in Eq. (4) which is the product of the expected improvement (now with plugin \( f_{\text{min}} \) being the best feasible fitness value) and the feasibility function \( F(x) \)

\[
ELP(x) = EI(x) \cdot F(x) = EI(x) \cdot \prod_{i=1}^{m} P(g_i(x) < 0),
\]

where \( P(g_i(x) < 0) \) is the probability of \( g_i(x) \) to be feasible, measured with the help of the Kriging model for the \( i \)th constraint:

\[
P(g_i(x) < 0) = \Phi \left( \frac{-\mu_i(x)}{\sigma_i(x)} \right).
\]

This algorithm often faces difficulties in solving COPs with 2 or more active constraints, because the product of feasibility probabilities approaches zero near the feasibility border where the optimum is located. Therefore, the penalized expected improvement may have very small values in the interesting region. Furthermore, this algorithm is unlikely to sample infeasible solutions. Others have shown that existence of the infeasible solutions in the population can often be helpful [23, 24]. In order to give solutions around the boundary a higher chance to be selected, we modify the feasibility function introduced by Schonlau [27] and formulate a modified expected improvement function \( EI_{\text{mod}} \) as follows:

\[
EI_{\text{mod}}(x) = EI(x) \cdot F(x) = EI(x) \cdot \prod_{i=1}^{m} \min \left( 2\Phi \left( \frac{-\mu_i(x)}{\sigma_i(x)} \right), 1 \right)
\]

Fig. 1 shows the different feasibility functions used in Schonlau algorithm (blue dashed line) and our proposed algorithm (red line) for one constraint.

The proposed algorithm SOCU, shown in Alg. 1, initially maximizes the feasibility function in order to find at least one feasible solution. As soon as one feasible solution is found, the algorithm proceeds by maximizing the modified expected improvement (Eq. 6) in each iteration. Since the \( EI_{\text{mod}} \) function is highly multimodal, we decided to use a simulated annealing method as the internal optimizer.

2.3 Plugin Control (PC): Preserving Feasibility

In our first experiments with \( EI_{\text{mod}} \) we observed a strange behavior depicted in Fig. 2: Initially, the surface plot of \( EI_{\text{mod}} \) looks as expected (upper plot), but the best feasible solution is still far away from the true solution. When finding better solutions near the true optimum, the \( EI_{\text{mod}} \) surface would suddenly change (lower plot) and the maximum of \( EI_{\text{mod}} \) shifts far away into infeasible area. What is the reason? A closer analysis revealed that the plugin \( f_{\text{min}} \) in Eq. (4) is responsible for this. Usually, the plugin \( f_{\text{min}} \) is taken as the best feasible objective found so far. A new value for \( f_{\text{min}} \) does not change the maxima locations of \( EI \), but it changes the intercept. This has a large effect on the maxima locations of \( EI_{\text{mod}}(x) = EI(x) \cdot F(x) \). We explain this with a 1D example.
2.3.1 The 1D-Case. Consider the following simple 1D-model:

\[ EI(x) = \max(ax + b, 0) \]  
\[ F(x) = \min(2\Phi(kx), 1) \]  

We assume \( k > 0 \), so that \( x < 0 \) is the infeasible area, and \( a < 0 \), i.e. the (unconstrained) \( EI(x) \) has better values towards \( x < 0 \). This makes the constraint active, meaning that the constrained optimum is on the border \( x = 0 \).

What happens now for \( EI_{mod} \) as a function of the intercept \( b \)? As Fig. 3 depicts, large \( b \) have the optimum for \( EI_{mod} \) correctly at \( x = 0 \), but too small values for \( b \) lead to a false shift of \( EI_{mod} \)'s maximum towards the infeasible area. A short Taylor expansion shows that the critical intercept is

\[ b_{crit} = \frac{a\sqrt{2\pi}}{2k}. \]  

Smaller intercepts have the maximum of \( EI_{mod} \) shifted to the infeasible area.

To correct this, we simply have to change the plugin in Eq. (4):

\[ EI(x) = E[\max(b_{crit} + f_{min} - Y, 0)] \]  

This ensures that at the active border, where \( Y \) is not larger than \( f_{min} \), the value of \( EI(x) \) is at least \( b_{crit} \). – If we have multiple constraints, we calculate \( b_{crit} \) for each of them.

2.3.2 The 2D- and \( n \)-D-Case. In the higher-dimensional case (\( d > 1 \)) we have to find the direction \( \hat{g} \) of slowest descent of \( EI_{mod} \) at the current best feasible point. This is for example in the case of G06 the bisecting line of the two constraints. The slopes \( a \) and \( k \) for \( EI \) and the constraint(s) in Eq. (9) have to be replaced by the respective slopes along direction \( \hat{g} \).

3 EXPERIMENTAL SETUP

3.1 General Setup

Initially we test the proposed algorithm on a toy problem \( \text{Sphere}4 \) of steerable difficulty: This problem has a sphere as objective function and 4 linear constraints, 2 of them being active, 2 inactive. The constraints enclose a feasible region with a triangular tip of angle \( \phi \) (see Fig. 4).

Next, we apply SOCU to all G-problems from the G-problem suite having 4 or less dimensions (see Table 1). This is because it is well known that Kriging algorithms are viable only for not too large dimensions. Equality constraints are translated to inequality constraints.\(^1\)

\(^1\)G02 and G03 are problems scalable in their dimension \( d \). We use here \( d = 2 \).

\(^2\)The inequality sign is chosen in such a way that the constrained optimum stays at the same location.
Figure 2: $E_I^{\text{mod}}$ shift towards the infeasible area for a 2-dimensional test problem (G06). The thick lines in form of a pointed triangle show the feasible area, the blue square is the true solution. The yellow circle is the best feasible solutions (being outside the plot area in the upper plot).

For each algorithm we run 30 independent trials with different $n = 5d$ initial points. Constraint violations smaller than $10^{-5}$ are tolerated. We consider a fixed budget of 100 function evaluations for all problems. The Kriging models are built by R packages DiceKriging and DiceOptim. In order to optimize $E_I^{\text{mod}}$ we use Generalized Simulated Annealing (R package GenSA). The time limit for this internal optimizer is set to 10 seconds in each iteration, allowing for several thousand Kriging model evaluations (depending on problem size).

Additionally, we run SOCU on an application example from aerodynamics described in the next section.

3.2 Application example to an aerodynamic shape design problem

The benchmarks related to aerodynamic shape design problems represent an ideal platform to test optimization systems and algorithms that make use of surrogate methods for the evaluation of the objective functions and the constraints [13]. This is for two main reasons,
Algorithm 1 SOCU algorithm

1: \( m \): number of constraints
2: \( n \): number of evaluated points
3: \( d \): dimension of the problem
4: \( \text{pop}^{(n)} \): population of \( n = 5d \) initial points generated by LHS
5: \textbf{while} \( n \leq \text{Budget} \) \textbf{do}
6: \hspace{1em} \text{Build from} \ \text{pop}^{(n)} \text{the Kriging models for objective function} \ f: (\mu_0, \sigma_0) \text{and the} \ m \text{ constraints} \ g_i: (\mu_1, \sigma_1), \ldots, (\mu_m, \sigma_m)
7: \hspace{1em} \text{Obtain} \ EI(x) \text{ from Eq. (4) with plugin corrector Eq. (10)}
8: \hspace{1em} F(x) = \prod_{i=1}^{m} \min \left( 2\Phi \left( -\frac{\mu_i(x)}{\sigma_i(x)} \right), 1 \right)
9: \hspace{1em} EI_{\text{mod}}(x) = EI(x) \cdot F(x)
10: \hspace{1em} \textbf{if} \ a \text{ feasible solution has been found} \ \	extbf{then}
11: \hspace{2em} x_{\text{new}} = \arg \max (EI_{\text{mod}}(x)) \quad \triangleright \text{Use simulated annealing}
12: \hspace{1em} \textbf{else}
13: \hspace{2em} x_{\text{new}} = \arg \max (F(x))
14: \hspace{1em} \textbf{end if}
15: \hspace{1em} \text{Add} \ x_{\text{new}} \text{ to} \ \text{pop}^{(n)} \text{ and evaluate it on true} \ f \text{ and} \ g_1, \ldots, g_m
16: \hspace{1em} n \leftarrow n + 1
17: \hspace{1em} \textbf{end while}

3.2.1 Problem setup. The goal of this optimization problem is the reduction of the aerodynamic drag of a given airfoil changing its shape. A generic airfoil shape is parameterized as linear combination of an initial geometry, defined parametrically by \((x_0(s), y_0(s))\), and a
The aerodynamic analysis code here selected to evaluate the airfoil performance is Drela’s XFOIL code [5]. This code is based on a second order panel method interactively coupled to a boundary layer integral module. Laminar to turbulent flow transition is predicted using the method described in [6].

### 4 RESULTS

#### Demonstration on Sphere4

The Sphere4 problem is a constrained optimization problem of steerable difficulty. It is used here to demonstrate the difference between the Kriging algorithms. Fig. 5 shows the results on Sphere4. As expected, the problem gets harder for all algorithms as the feasibility angle $\phi$ decreases. This is because the feasible region becomes smaller and the optimum is surrounded by a neighborhood containing more and more infeasible points. - Interestingly, also the gap between Schonlau and SOCU gets larger as $\phi$ decreases. This is understandable as well: For SOCU the probability of feasibility is 1 along both active constraint lines for all $\phi$. For Schonlau the probability of feasibility is 0.25 at the optimum (2 active constraints, both 0.5). If $\phi$ is large, the probability quickly rises to 0.5 as we move along one of the active constraint lines, because the distance to the other line increases. But if $\phi$ is small and we move along one constraint line, the probability according to Schonlau stays longer near 0.25, because the second constraint line is not far away. Thus the solution found by Schonlau will be farther from the optimum since the maximum of the product $EI \cdot F$ moves farther into the feasible region. This is exactly what we see in Fig. 5.

#### Noise Variance

In our first experiments we experienced frequent crashes of the Kriging modeling software due to numeric instabilities. This is a well-known but cumbersome observation shared by many researchers in Kriging especially if the modeling points are unevenly spaced, as it is inevitably the case in optimization tasks. To avoid too strong oscillations of the Kriging model due to nearby points, it is a common cure to switch from interpolating to approximating Kriging models, either with the so-called nugget-effect or with a noise variance parameter, which assumes a number of functions $y_i(s)$ that may be defined analytically or by point distributions [11]:

$$y(s) = k \left( y_0(s) + \sum_{i=1}^{n} w_i y_i(s) \right), \quad x(s) = x_0(s)$$  \hspace{1cm} (11)

where the airfoil shape is controlled by the design parameters $w_i$ and by the scale factor $k$. The operating conditions are Mach number equal to 0.0 (incompressible flow) and Reynolds number equal to 3000000. The starting airfoil is the NACA 2412 [1]. The Mach and Reynolds numbers that characterize the specified regime of operation are sufficiently low to allow an extended laminar bucket that can have a beneficial effect on a large part of the flight envelope. The design goals are translated into the following optimization problem:

$$\begin{align*}
\min C_D \\
\text{subject to:} & \quad C_L = 0.5 \\
& \quad C_M \geq -0.07 \\
& \quad C_{DP} \geq 0 \\
& \quad t/c = 0.12 \\
& \quad \ell_R \geq 0.006
\end{align*}$$  \hspace{1cm} (12)

Where $C_D$, $C_{DP}$, $C_L$ and $C_M$ are the drag, pressure drag, lift and pitching moment coefficient of the airfoil; $t/c$ denotes the thickness to chord ratio. The two equality constraints defined in (12) are here satisfied by explicitly changing two free problem parameters and therefore they are not considered by the optimization algorithm. In particular, the constraint on $t/c$ is satisfied by changing properly the free parameter $k$, while the constraint on $C_L$ is satisfied by changing the second free parameter, namely the airfoil angle of attack $\alpha$. The aerodynamic analysis code here selected to evaluate the airfoil performance is Drela’s XFOIL code [5]. This code is based on a second order panel method interactively coupled to a boundary layer integral module. Laminar to turbulent flow transition is predicted using the method described in [6].

### Table 1: Characteristics of the G-functions

<table>
<thead>
<tr>
<th>Fct.</th>
<th>$d$</th>
<th>$\rho^*$</th>
<th>$FR$</th>
<th>$GR$</th>
<th>$LI / NI$</th>
<th>$LE / NE$</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>G02mod</td>
<td>2</td>
<td>99.997%</td>
<td>0.57</td>
<td>2.632</td>
<td>1 / 1</td>
<td>0 / 0</td>
<td>1</td>
</tr>
<tr>
<td>G03mod</td>
<td>2</td>
<td>78.388%</td>
<td>0.99</td>
<td>1.000</td>
<td>0 / 0</td>
<td>0 / 1</td>
<td>1</td>
</tr>
<tr>
<td>G05</td>
<td>4</td>
<td>0.0000%</td>
<td>8863.69</td>
<td>1788.74</td>
<td>2 / 0</td>
<td>0 / 3</td>
<td>3</td>
</tr>
<tr>
<td>G06</td>
<td>2</td>
<td>0.0072%</td>
<td>1246882.3</td>
<td>1.010</td>
<td>0 / 2</td>
<td>0 / 0</td>
<td>2</td>
</tr>
<tr>
<td>G08</td>
<td>2</td>
<td>0.8751%</td>
<td>1821.61</td>
<td>2.393</td>
<td>0 / 2</td>
<td>0 / 0</td>
<td>0</td>
</tr>
<tr>
<td>G11</td>
<td>2</td>
<td>0.0000%</td>
<td>4.99</td>
<td>1.000</td>
<td>0 / 0</td>
<td>0 / 1</td>
<td>1</td>
</tr>
<tr>
<td>G15</td>
<td>3</td>
<td>0.0000%</td>
<td>586.0</td>
<td>1.034</td>
<td>0 / 0</td>
<td>1 / 1</td>
<td>2</td>
</tr>
<tr>
<td>G24</td>
<td>2</td>
<td>0.44250%</td>
<td>6.97</td>
<td>1.82</td>
<td>0 / 2</td>
<td>0 / 0</td>
<td>2</td>
</tr>
<tr>
<td>XFOIL</td>
<td>4</td>
<td>0.1349%</td>
<td>0.99</td>
<td>1.34</td>
<td>0 / 3</td>
<td>0 / 0</td>
<td>1</td>
</tr>
</tbody>
</table>

$D$, $\rho^*$, $FR$, $GR$, $LI / NI$, $LE / NE$, and $a$ denote the dimension, feasibility rate (%), range of the fitness values, ratio of largest to smallest constraint range, number of linear/nonlinear inequalities, number of linear/nonlinear equalities, and number of constraints active at the optimum, respectively.
Figure 4: Sphere4 problem. The colored contour levels show the objective function $10^5 \cdot (x_1^2 + x_2^2)$ (2D sphere function). The thick black lines depict 4 linear constraints enclosing the feasible region. The difficulty of this problem is scalable by changing the feasibility angle $\phi$.

Figure 5: Comparing the final optimization error determined by different algorithms for optimizing Sphere4 problems with different feasibility angles. The results are taken from 30 independent runs and 100 function evaluations.

certain noise or uncertainty related with every modeling point. Since the nugget effect leads to a complicated structure for the variance $\sigma^2(x)$, it turned out to be not well-suited for our case.

The noise variance, on the other hand, turned out to be very effective: As Table 2 shows, the interpolating Kriging models had frequent crashes. By adding the noise variance $\nu = 0.01$ to the model, we could completely avoid any crashes in our experiments.
Table 2: Effect of noise variance.

<table>
<thead>
<tr>
<th>Problem</th>
<th>SOCU w/o noise variance</th>
<th>SOCU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>crashed (%)</td>
<td>iteration</td>
</tr>
<tr>
<td>G06</td>
<td>100</td>
<td>19</td>
</tr>
<tr>
<td>G02mod</td>
<td>96</td>
<td>40</td>
</tr>
</tbody>
</table>

**Performance on G-Problems**

In Fig. 6 we compare the two different variants of Kriging-based, EGO, namely the original version of Schonlau et al. [27] and our SOCU algorithm, which we applied to all G-problems in Tab. 1. SOCU reaches lower optimization errors in most cases. In the case of G08, both algorithms have the same median curve. This is perfectly understandable, since G08 is the only problem without any active constraints. Absence of an active constraint is a convincing reason for similar performance of both algorithms, since the different feasibility functions (Fig. 1) have no effect. For problems with active constraints, the high value of SOCU’s $F_i(x)$ at the border of the feasible region helps to find better solutions.

In Fig. 7 we show additionally the effect of switching off the plugin control (Sec. 2.3) in SOCU (green boxplot ‘SOCU w/o PC’). It can be seen that the plugin control is beneficial for G03, G06 and G11 (blue boxplot ‘SOCU’).

We compare in Fig. 8 the results of SOCU with SACOBRA [3], to the best of our knowledge the current state-of-the-art for solving the G-problems. It is evident that SACOBRA is slightly better in most cases, except for the case of G24. Additionally, three problems (G05, G15, G24) show a better performance of SOCU in the early stages (between iteration 25 and 75), although SACOBRA catches up at iteration 100.

Finally Fig. 9 shows the overall performance comparison for all algorithms on all test problems in form of a data profile. The larger the data profile (ratio of solved problems), the better the relevant algorithm. The performance factor $\alpha$ on the x-axis is the number of iterations divided by $d + 1$.

**Performance on XFOIL**

Fig. 10 shows our results for the XFOIL case. SOCU is slightly better than Schonlau in the median, but the difference is not statistically significant. The similarity is understandable, since problem XFOIL has only one active constraint. In such a case we do expect the differences between Schonlau and SOCU to be not very large. SOCU and SACOBRA produce nearly identical results.

5 **DISCUSSION AND CONCLUSION**

We applied Kriging surrogate models to constrained optimization. We could show that a small number of evaluations is sufficient to obtain good optimization results. We were not the first to do so, but three important conclusions for Kriging-based optimization could be drawn in this paper:

1. Interpolating Kriging models often suffer from numerical instabilities and subsequent crashes, especially when the population points are unevenly distributed in the search space. This will be nearly always the case when applying Kriging for optimization. We have shown that these crashes can be completely avoided (at least in our test cases) when we add a small noise variance to the Kriging models. This leads to approximating Kriging models and to a variance always larger than zero. Both effects are beneficial for numeric stability of the Kriging models.

2. Many Kriging-based COP-solvers use in one form or the other a product of expected improvement $EI$ and probability of feasibility $F$. We could show that if the additive plugin in $EI$ gets very small (which is no problem for the optima of $EI$ themselves), this can have adversarial effects for the optima of $EI \cdot F$. We proposed a method called *plugin control* which successfully counteracts such adversarial effects.

3. Having this plugin control in effect, we could show that a probability curve of SOCU being 1 at the border of feasibility is clearly superior to an approach where the probability is 0.5 at the border [27]. The benefits are - as expected - more clearly seen for problems with two or more active constraints (G05, G06, G15, G24).

We tested our new algorithm SOCU on a variety of benchmark problems with dimension of $d = 4$ and below. We could perform better than the Kriging-based algorithm of Schonlau et al [27], but were in most cases worse than the non-Kriging-based SACOBRA algorithm [3].

In some cases (G05, G15, G24) SOCU showed a better performance (median, best and worst case) in early iterations (< 75) than SACOBRA. We suppose that this is due to the better exploration of the EGO approach. A drawback for Kriging-based models is however that they become very slow for larger dimensions.

In the future, a closer investigation into the causes of the observed performance differences would be desirable. Furthermore, as the name SOCU suggests, we plan to test our algorithm also on problems involving uncertainty.

The source code of the proposed algorithm "SOCU" is available online\(^3\).

\(^3\)https://github.com/saminehbagheri/SOCU
Figure 6: Comparing the performance of SOCU and Schonlau [27]. The thick lines show the median from 30 independent runs while the colored bands mark the worst and the best run.

ACKNOWLEDGMENTS
This work was initiated during the Surrogate-Assisted Multi-Criteria Optimization Workshop at the Lorentz Center (Leiden, The Netherlands), 2016. We are grateful to the other participants of the workshop and the Lorentz Center for their support. This research was partly funded by Tekes, the Finnish Funding Agency for Innovation (the DeCoMo project), and by the Engineering and Physical Sciences Research Council [grant numbers EP/N017195/1, EP/N017846/1].
Figure 7: Comparing the final optimization error from 30 independent runs determined with different algorithms

REFERENCES

Figure 8: Comparing optimization performance of SACOBRA and SOCU.
Figure 9: Data profile.

Figure 10: Comparing the performance of SOCU, Schonlau [27], and SACOBRA [3] on the XFOIL real world optimization problem (30 independent runs).


