25th ABCM International Congress of Mechanical Engineering
October 20-25, 2019, Uberlândia, MG, Brazil

# COB-2019-0383
# Improving the Efficiency of Stochastic Efficient Global Optimization with Stochastic Tunneling

**Fábio Felipe dos Santos Nascentes**
Center for Optimization and Reliability in Engineering (CORE), Universidade Federal de Santa Catarina, Florianópolis 88037-000, Brazil
Departamento de Áreas Acadêmicas, Instituto Federal de Educação, Ciência e Tecnologia de Goiás-IFG, Jataí 75804-714, Brazil
fabiotrabmat@gmail.com
**Rafael Holdorf Lopez**
Center for Optimization and Reliability in Engineering (CORE), Universidade Federal de Santa Catarina, Florianópolis 88037-000, Brazil
rafaelholdorf@gmail.com
**Rubens Sampaio**
Departamento de Engenharia Mecânica, PUC-Rio, Rio de Janeiro 22453-900, Brazil
rsampaio@puc-rio.br
**José Eduardo Souza de Cursi**
Department Mecanique, Institut National des Sciences Appliquees (INSA) de Rouen, Saint Etienne du Rouvray 76801, Cedex, France
eduardo.souza@insa-rouen.fr

***Abstract.*** *This paper proposes the use of a normalization scheme for increasing the performance of the recently developed Adaptive Target Variance Stochastic Efficient Global Optimization (sEGO) method. Such a method is designed for the minimization of functions that depend on expensive to evaluate and high dimensional integrals. The results showed that the use of the normalization in the sEGO method yielded very promising results for the minimization of integrals. Indeed, it was able to obtain more precise results, while requiring only a fraction of the computational budget of the original version of the algorithm.*

***Keywords:*** *Stochastic Efficient Global Optimization, Stochastic Tunneling, Global Optimization, Robust Design*

## 1. INTRODUCTION

The optimization of a variety of engineering problems may require the minimization (or maximization) of expensive to evaluate and high dimensional integrals. These problems become more challenging if the resulting objective function turns out be not non convex and multimodal. Examples of this kind may arise, for example, from the maximization of the expected performance of a mechanical system, vastly applied in robust design (Lopez *et al.*, 2014), the multidimensional integral of Performance Based Design Optimization (Beck *et al.*, 2014), or the double integral of Optimal Design of Experiment problems (Beck *et al.*, 2018).

A powerful approach to handle these issues is the Efficient Global Optimization (EGO) (Jones *et al.*, 1998), which exploits the information provided by the Kriging metamodel to iteratively add new points, improving the surrogate accuracy and at the same time seeking its global minimum. For problems presenting variability (or uncertainty), the Stochastic Kriging (SK) (Ankenman *et al.*, 2010) was developed. The use of SK within the EGO framework, or stochastic Efficient Global Optimization (sEGO), is relatively recent. For example, Picheny *et al.* (2013) benchmarked different infill criteria for the noisy case, while Jalali *et al.* (2017) compared Kriging-based methods in heterogeneous noise situations.

Recently, a Adaptive Variance Target sEGO (Carraro *et al.*, 2019) approach was proposed for the minimization of integrals. It employs Monte Carlo Integration (MCI) to approximate the objective function and includes the variance of the error in the integration into the SK framework. This variance of the error is adaptively managed by the method, providing an efficient optimization process by rationally spending the available computational budget. This method reached promising results, specially in high dimensional problems (Carraro *et al.*, 2019).

This paper, thus, aims at enhancing the performance of the Adaptive Variance Target sEGO (Carraro *et al.*, 2019) by proposing the use of a normalization scheme during the optimization process. This normalization is the result of the so called stochastic tunneling approach, applied together with the Simulated Annealing (SA) for global Minimization of Complex Potential Energy Landscapes (Wenzel and Hamacher, 1999). In the sEGO context, it is expected that this normalization reduce the variability level of the regions of the design domain that have high values of the objective function as well as reduce the dependency of the quality of the search on the parameters of the SK.

## 2. Problem Statement

The goal of this paper is to solve the problem of minimization of a function $y$, which depends on an integral as in

$$\min_{\mathbf{d}\in S} J(\mathbf{d}) = \int_{\Omega} \phi(\mathbf{d}, \mathbf{x}) w(\mathbf{x}) d\mathbf{x}, \tag{1}$$

where $\mathbf{d} \in \Re^k$ is the design vector of dimension $k$, $\mathbf{x} \in \Re^{n_x}$ is the parameter vector, $\phi : \Re^k \times \Re^{n_x} \to \Re$ is a known function, $S$ is the design domain, $w(\mathbf{x})$ is some known weight function (*e.g.* probability distribution) and $\Omega \subseteq \Re^{n_x}$ is the integration domain (*e.g.* support of the probability distribution). We also assume here that the design domain $S$ considers only box constrains. Here, we are interested in situations that: $\phi$ is a black box function and is computationally demanding, while the resulting objective function $y$ is not convex and multimodal. Applying MCI to estimate $y$, we have

$$J(\mathbf{d}) \approx \bar{J}(\mathbf{d}) = \frac{1}{n_r} \sum_{i=1}^{n_r} \phi(\mathbf{d}, \mathbf{x}^{(i)}), \tag{2}$$

where $n_r$ is the sample size and $\mathbf{x}^{(i)}$ are sample points randomly drawn from distribution $w(\mathbf{x})$. One of the advantages of MCI is that we are able to estimate the variance of the error of the approximation as:

$$\overline{\sigma}^2(\mathbf{d}) = \frac{1}{n_r(n_r-1)} \sum_{i=1}^{n_r} (\phi_i - \bar{J}(\mathbf{d}))^2, \tag{3}$$

where $\phi_i = \phi(\mathbf{d}, \mathbf{x}^{(i)})$. Thus, by increasing the sample size $n_r$ (*i.e.* the number of replications), the variance estimate decreases and approximation in Eq. (2) gets closer to the exact value of Eq. (1).

## 3. The Adaptive Variance Target sEGO approach

The main steps of the adaptive variance selection sEGO method proposed by Carraro *et al.* (2019) are:

1. Creation of the initial sampling plan using Latin Hypercube;
2. Checking stopping criterion
3. Finding the new infill point by maximizing either EI or AEI
4. Evaluate $\bar{J}$ by MCI using the adaptive target selection scheme
5. Add new infill and its $\bar{J}$ to the current sampling plan
6. Construct the new SK metamodel and return to step 2.

In the next subsections, we present a general description of the algorithm. For a detailed presentation, the reader is referred to Carraro *et al.* (2019).

### 3.1 Creation of the initial sampling plan

At this step, a Latin Hypercube scheme is employed to create an initial sampling plan $\mathbf{\Gamma}$ containing $n_s$ points, *i.e.*

$$\mathbf{\Gamma} = \{\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \ldots, \mathbf{d}^{(n_s)}\}, \tag{4}$$

where MCI is employed to obtain approximations of $J$, *i.e.*

$$y^{(i)} \approx \bar{J}\left(\mathbf{d}^{(i)}\right), \tag{5}$$

for which we also have the estimation of the error variance, given by Eq. (3). The points $\mathbf{d}^{(i)}$, for $i = 1, 2, \cdots, n_s$, are called in this article by support points.

### 3.2 Checking stopping criterion

Several criteria might be employed for checking the convergence of the search, *e.g.* minimum value of the expected improvement, stagnation of the search, available computational budget. In this paper, we employ the latter and measure it by the number of function evaluations (NFE), *i.e.* the number of times the function $\phi$ is evaluated. Hence, a maximum NFE is set a-priori in each example, and once it is reached, the algorithm is stopped and the current best design is set as the optimum design of the search.

### 3.3 Stochastic Kriging (SK) metamodel

SK does act as a regression model, *i.e.* it no longer interpolates the observed input/output data. Consequently, it is able to address the situation we have in the problem under analysis, *i.e.* for a given support point in the design domain $\mathbf{d}^{(i)}$, $\hat{J}(\mathbf{d}^{(i)}) \approx y^{(i)}$. In other words, SK accounts for the sampling variability that is inherent to a stochastic simulation. SK prediction can be seen as:

$$\hat{y}(\mathbf{d}_i) = \overbrace{M(\mathbf{d}_i)}^{\text{Trend}} + \overbrace{Z(\mathbf{d}_i)}^{\text{Extrinsic}} + \overbrace{\epsilon(\mathbf{d}_i)}^{\text{Intrinsic}},$$

where $M(\mathbf{d})$ is the usual average trend, $Z(\mathbf{d})$ accounts for the model uncertainty and is now referred as extrinsic noise. The additional term $\epsilon$, called intrinsic noise, accounts for the variability, and in the case of this paper, it is due to the MCI. It can be assumed independent and identically distributed (i.i.d.) across sample points, and possess a Gaussian distribution with zero mean and variance given by Equation 3.

Ankenman *et al.* (2010) proposed that the predictions made by the metamodel is be given by $Y(\mathbf{d}_u) = M(\mathbf{d}_u) + Z(\mathbf{d}_u)$ for any $\mathbf{d}_u$, either support point or not. For sake of simplicity, the authors consider that $M(\mathbf{d}) = \mu$. Let $\mathbf{\Sigma}_Z$ be the covariance matrix obtained by the extrinsic spatial correlation between all support points in $\mathbf{\Gamma}$,

$$\mathbf{\Sigma}_Z \left(\mathbf{d}^{(i)}, \mathbf{d}^{(j)}\right) = \text{Cov}\left[Z(\mathbf{d}^{(i)}), Z(\mathbf{d}^{(j)})\right]. \tag{6}$$

Let $\mathbf{\Sigma}_Z(\mathbf{d}_u, \cdot)$ be the $n_s \times 1$ vector formed by covariance between the predicted point and the support points. Finally, we place $\mathbf{\Sigma}_\varepsilon$ as being the matrix $n_s \times n_s$ representing the covariance obtained for the support points in the intrinsic portion of our metamodel. Using the MSE (Mean Squared Error), Ankenman *et al.* (2010) showed that the best unbiased predictor for the metamodel is

$$Y(\mathbf{d}_u) = \mu + \mathbf{\Sigma}_Z\left(\mathbf{d}_u, \cdot\right)^T \left[\mathbf{\Sigma}_Z + \mathbf{\Sigma}_\varepsilon\right]^{-1} \left(\bar{\mathbf{y}} - \mu\mathbf{1}\right), \tag{7}$$

where $\bar{\mathbf{y}} = \{y^{(1)}, y^{(2)}, \cdots, y^{(n_s)}\}$ is given by Eq. (5) and $\mathbf{1}$ is the $n_s \times 1$ vector of ones. The minimum value of MSE for the Eq. (7) is find by

$$s^2(\mathbf{d}_u) = \mathbf{\Sigma}_Z\left(\mathbf{d}_u, \mathbf{d}_u\right) - \mathbf{\Sigma}_Z\left(\mathbf{d}_u, \cdot\right)^T \left[\mathbf{\Sigma}_Z + \mathbf{\Sigma}_\varepsilon\right]^{-1} \mathbf{\Sigma}_Z\left(\mathbf{d}_u, \cdot\right). \tag{8}$$

To give more structure to $\mathbf{\Sigma}_Z$, in particular, let us assume that

$$\mathbf{\Sigma}_z\left(\mathbf{d}^{(i)}, \mathbf{d}^{(j)}\right) = \sigma_z^2 h\left(\mathbf{d}^{(i)}, \mathbf{d}^{(j)}; \boldsymbol{\theta}\right), \tag{9}$$

where $\sigma_z^2$ can be interpreted as the variance of $\mathbf{\Sigma}_Z$ for all $\mathbf{d}$, and $h$ is the correlation between the support points $\mathbf{d}^{(i)}$ e $\mathbf{d}^{(j)}$, with $h\left(\mathbf{d}^{(i)}, \mathbf{d}^{(j)}; \boldsymbol{\theta}\right) = \exp\left[-\sum_{l=1}^{k} \theta_l |d_l^{(i)} - d_l^{(j)}|^2\right]$.

In this context, to apply the SK as a surrogate model, the parameters $\mu, \sigma_z^2, \boldsymbol{\theta}$ and $\mathbf{\Sigma}_\varepsilon$ must be estimated according to the information in our data. To introduce the error information made by the approximation via MCI in our model, we will estimate the intrinsic part as

$$[\hat{\mathbf{\Sigma}}_\varepsilon]_{ij} = \overline{\sigma}^2(\mathbf{d}^{(i)})\,\delta_{ij}, \quad i = 1, 2, ..., n_s, \tag{10}$$

where $\overline{\sigma}^2(\mathbf{d}^{(i)})$ is given by Eq. (3) and $\delta_{ij}$ is the Kronecker delta. The other parameters can be estimated by maximizing the likelihood function, given by

$$L\left(\mu, \sigma_z^2, \boldsymbol{\theta}\right) = -\frac{n_s}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Sigma}| - \frac{1}{2}\left(\bar{\mathbf{y}} - \mu\mathbf{1}\right)^T \mathbf{\Sigma}^{-1}\left(\bar{\mathbf{y}} - \mu\mathbf{1}\right), \tag{11}$$

where $\mathbf{\Sigma} = \sigma_z^2\,\mathbf{\Psi}(\boldsymbol{\theta}) + \mathbf{\Sigma}_\varepsilon$ with $\mathbf{\Psi}(\boldsymbol{\theta})$ being the spatial correlation matrix of the extrinsic term $Z(\mathbf{d})$ between all support points. A global maximum point of Eq. (11) will contain the estimation $\hat{\mu}, \hat{\sigma}_z$ and $\hat{\boldsymbol{\theta}}$ for the SK metamodel parameters. After this optimization, the SK predictor will be given by

$$\hat{Y}(\mathbf{d}_u) = \hat{\mu} + \hat{\sigma}_z^2\mathbf{h}^T\hat{\mathbf{\Sigma}}^{-1}\left(\bar{\mathbf{y}} - \hat{\mu}\mathbf{1}\right), \tag{12}$$

where $\mathbf{h}$ is the $n \times 1$ vector with contain the spatial correlations between the point $\mathbf{d}_u$ and all support points and $\hat{\mathbf{\Sigma}} = \hat{\sigma}_z^2\,\mathbf{\Psi}(\hat{\boldsymbol{\theta}}) + \hat{\mathbf{\Sigma}}_\varepsilon$. The minimum of MSE given by the predictor is

$$s^2(\mathbf{d}_u) = \hat{\sigma}_z^2 - \left(\hat{\sigma}_z^2\right)^2\mathbf{h}^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{h} + \frac{\Delta^2}{\mathbf{1}^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{1}}, \tag{13}$$

where $\Delta = 1 - \mathbf{1}^T\hat{\mathbf{\Sigma}}^{-1}\hat{\sigma}_z^2\mathbf{h}$.

### 3.4 Finding the new infill point

The idea behind any EGO infill criterion is to use the information about the predicted error of the model, given by the SK, to guide the optimization search by the addition of the Infill Points - IP. That is the reason behind the importance in the definition of such an error and its consequence in the performance of the algorithm. In this paper, we employ the AEI infill criteria (Huang *et al.*, 2006).

The AEI criterion aims to use a new approach to the Expected Improvement - EI, worked by Jones *et al.* (1998) in the deterministic case. In order to do so, we will take into account the uncertainties regarding the evaluation of function and prediction by SK. For the EI we have that an improvement in the predicted value is given by

$$I = \max\{J_{\min} - Y, 0\},$$

where $J_{\min} = \min\{J(\mathbf{d}^{(1)}), J(\mathbf{d}^{(2)}), \cdots, J(\mathbf{d}^{(n_s)})\}$. However, for stochastic cases, this definition does not carry correct information about the lowest value of the $J$ function, since we do not take into account the uncertainties associated with the problem. Therefore, a more natural definition for the stochastic case will be

$$I = \max\{\hat{Y}(\mathbf{d}^{**}) - Y, 0\}, \tag{14}$$

where $\mathbf{d}^{**}$ is called Effective Best Solution and $\hat{Y}$ is given by Eq. (12). The $\mathbf{d}^{**}$ is defined by

$$\mathbf{d}^{**} = \underset{\mathbf{d}\in\mathbf{\Gamma}}{\text{argmin}} \left\{\hat{Y}(\mathbf{d}) + \Phi^{-1}(\beta)s(\mathbf{d})\right\}, \tag{15}$$

where $\hat{Y}(\mathbf{d})$ is given by Eq. (12), $\Phi^{-1}(\beta)$ is the inverse cumulative distribution of probability for a standard normal variable and $s(\mathbf{d})$ is the square root of Eq. (13). From Huang *et al.* (2006), $\beta = 0.841345$.

With this changes in the improvement, the EI proposed by Jones *et al.* (1998) is changed for

$$\mathbb{E}(I(\mathbf{d})) = \left(\hat{Y}(\mathbf{d}^{**}) - \hat{Y}(\mathbf{d})\right)\Phi\left(\frac{\hat{Y}(\mathbf{d}^{**}) - \hat{Y}(\mathbf{d})}{s(\mathbf{d})}\right) + s(\mathbf{d})\phi\left(\frac{\hat{Y}(\mathbf{d}^{**}) - \hat{Y}(\mathbf{d})}{s(\mathbf{d})}\right). \tag{16}$$

Note that Eq. (16) does not vanish when $\mathbf{d} = \mathbf{d}^{**}$, since $s(\mathbf{d}^{**}) > 0$ in the second portion of the sum. This is an excellent quality in the use of EI, as it allows support points to be chosen as IP; thus, new replications can be added to the function value, decreasing the error variance. We also have that Eq. (16) is reduced to the same metric obtained by Jones *et al.* (1998) in the absence of noise, where $\mathbf{\Sigma}_\varepsilon = \mathbf{0}$.

At the same time that new replications are positive to be able to reduce the variance of the error made in the approximation, it should be avoided that the metric will be stuck only in the support points. In this way, Huang *et al.* (2006) define the AEI as

$$\text{AEI}(\mathbf{d}) = \mathbb{E}(I(\mathbf{d}))\left(1 - \frac{\bar{\sigma}(\mathbf{d})}{\sqrt{s^2(\mathbf{d}) + \bar{\sigma}^2(\mathbf{d})}}\right), \tag{17}$$

where $\mathbb{E}(I(\mathbf{d}))$ is defined by Eq. (16), $\bar{\sigma}^2(\mathbf{d})$ is given by Eq. (3) and $s^2(\mathbf{d})$ is obtained from Eq. (13). Thus, the next IP, denoted by $\mathbf{d}^{n_s+1}$, will be defined by

$$\mathbf{d}^{n_s+1} = \underset{\mathbf{d}\in S}{\text{argmax}}\ \text{AEI}(\mathbf{d}). \tag{18}$$

The second part of the AEI metric in Eq. (17) acts as a penalty for the value of the new EI, preventing the metric from being stuck at a given point $\mathbf{d}$. Theoretically this happens because, with new replications, $s^2(\mathbf{d})$ will tend to zero, making the AEI also zero. Another important detail about the AEI metric for the heterogeneous case is the need to know what intrinsic variance $\bar{\sigma}^2(\mathbf{d})$ at the point $\mathbf{d}$. Assuming $\mathbf{d}$ is not a support point, then this value was not estimated by Eq. (3). Therefore, since we do not have an explicit law for $\bar{\sigma}^2(\mathbf{d})$, we will use the function defined in Eq. (22), obtained from the concepts of adaptivity exploited in next section.

### 3.5 Adaptive target selection scheme

After the construction of the first metamodel, the infill stage begins. Here, for the approximation with MCI an initial target variance $\bar{\sigma}_0^2$ is set and the first infill point is added to the model being simulated up to this corresponding target variance. From the second infill point on, the adaptive target selection scheme proposed by Carraro *et al.* (2019) starts to take place. The target variance of a new infill point is calculated as:

$$\bar{\sigma}^2_{adapt} = \bar{\sigma}_0^2 \exp\left[0.01 \cdot k \cdot n_{close} - 0.5(1 + k + n_{close})\right], \tag{19}$$

where $k$ is the dimension of the problem design variable and $n_{close}$ is the number of points already sampled near the new infill point. The latter is defined by the number of points in the model located at a given distance of the new infill point. Suppose $\mathbf{d}^{n_s+1}$ is the point that maximizes the AEI infill criterion, then, $n_{close}$ is evaluated as

$$n_{close} = \sum_{j=1}^{n_s} U(\mathbf{d}^{n_s+1}, \mathbf{d}^{(j)}), \tag{20}$$

where

$$U(\mathbf{d}^{n_s+1}, \mathbf{d}^{(j)}) = \begin{cases} 1, & \|\mathbf{d}^{n_s+1} - \mathbf{d}^{(j)}\|_\infty \leq r_{hc}, \\ 0, & \|\mathbf{d}^{n_s+1} - \mathbf{d}^{(j)}\|_\infty > r_{hc}, \end{cases} \tag{21}$$

in which $\| \cdot \|_\infty$ is a infinite norm for an vector and $r_{hc}$ is one of the input parameters of the proposed adaptive approach and corresponds to the distance considered around the infill point. The infinite norm represents a hypercube around the infill point selected with half-sides $r_{hc}$. Intuitively, this parameter controls the aggressiveness of the search. Increasing it makes the hypercube larger, allowing more sampled points to be treated as close ones.

Then, when the infill is located within an unsampled region ($n_{close} = 0$), its target variance is set as the initial target variance. On the other hand, when the infill is located in a region with existing sampled points ($n_{close} > 0$), a lower target variance ($\bar{\sigma}^2_{adapt}$) is employed for the approximation of its objective function value. Since different targets are being employed, $\bar{\sigma}^2$ becomes dependent on the sampling characteristics. It is now adaptively updated according to

$$\bar{\sigma}^2(\mathbf{d}) = \begin{cases} \bar{\sigma}^2_0 & \text{if} \quad n_{close} = 0 \\ \bar{\sigma}^2_{adapt} & \text{if} \quad n_{close} > 0 \end{cases}, \tag{22}$$

It is worth to highlight here that it is also important to set a minimum value for the adaptive target to avoid a computationally intractable number of samples. In other words, not to spend the entire computational budget in only a few points. We thus enforce

$$\bar{\sigma}^2_{\min} \leq \bar{\sigma}^2_{adapt} \leq \bar{\sigma}^2_0, \tag{23}$$

where $\bar{\sigma}^2_{\min}$ is a lower bound on the target.

### 3.6 The Proposed Normalization Scheme

A natural problem that arises in stochastic processes is the range of the response of an objective function. When a function has an excessively large domain range, a small target variance may become unreachable, even for the supports points in our first SK metamodel.

For an illustration of this obstacle, let us take an one dimensional problem given by Eq. (1) where

$$\phi(d, X) = \left[5|d + 0.5| \left(\cos(20d) + 3d^2\right)\right] \cdot X, \tag{24}$$

with $X \sim \mathcal{N}(1, 0.5)$ as an normal variable. In this problem, the design domain is $S = [-3, 3]$. Suppose a usual optimization with sEGO for this problem where the adaptivity starts with $\bar{\sigma}^2_0 = 0.01$. In any iteration suppose that some points in the search space have already been added and the algorithm finds a new IP, where $d = 0.085$. If in the location of this new IP exists two other points, then $n_{close} = 2$, $\bar{\sigma}^2_{adapt} = 0.00138$ and MCI spends $n_r = 21$ replications of objective function to achieve the variance of $\bar{\sigma}^2(0.085) = 0.00133$ and a approximate value $\bar{J}(0.085) = -0.354553$ (with relative error of 13.1055%). This amount of replication is easily computed by any current computer and does not detract instantly the computational budget available for optimization. In other situation, suppose that a exploration point $d = 1.5$ is added like an IP and there are no other points exists in your neighborhood. Thus, $n_{close} = 0$, $\bar{\sigma}^2_{adapt} = \bar{\sigma}^2_0 = 0.01$ and MCI spends $n_r = 119\,233$ replications of objective function to achieve the variance of $\bar{\sigma}^2(1.5) = 0.0099$ and a approximate value $\bar{J}(1.5) = 69.152476$ (with relative error of 0.1593%). We noticed that in the last case a considerable amount of function evaluations was used. This amount can neither be provided as a computational budget or else it can make the algorithm quickly reach an maximum cost.

To get around this situation, the normalization employed in this paper is based on the stochastic tunneling approach, which consists in allowing the particle to “tunnel” high energy regions of domain, avoiding getting trapped on local minima, usually employed in the SA algorithm. According to Wenzel and Hamacher (1999), it may be done by applying the following nonlinear transformation to $\bar{J}$

$$F(\mathbf{d}) = 1 - \exp\Big[ - \gamma(\bar{J}(\mathbf{d}) - J_0)\Big], \tag{25}$$

where $\gamma$ is a tunneling parameter and $J_0$ is the lowest minimum encountered by the dynamical process in the SA algorithm, in this work, $J_0$ is the knowing minimum value of the problem. After applying normalization, sEGO approaches minimizes $F$ instead of the original approximated function $\bar{J}$.

Returning to the one dimensional case analysis in the IP $d = 1.5$, applying an normalization with $\gamma = 0.01$ and $J_0 = -3.043080$, then sEGO combined with MCI spends $n_r = 18$ replications of objective function to achieve the variance of $\bar{\sigma}^2(1.5) = 0.00206$ and a approximate value $\bar{J}(1.5) = 68.065435$ (with relative error of 1.4152%).

## 4. Numerical examples and results

In this section, we analyze the minimization of three multimodal problems, in yours standard form and a normalized form. For the correct use of the principles of sEGO, we will take deterministic functions with knowing minimum values and transform them into stochastic functions.

The first problem is the one dimensional function Eq. (24) presents in the earlier section. Your standard optimization problem is given by Eq. (1), where $d \in S = [-3, 3]$. $X$ are an normal random variable given by $X \sim \mathcal{N}(1, 0.5)$. Using $X = 1$ we see that range of function values is $R_\phi = 458.8759$ (difference between maximum and minimum value). The minimum value of the function is $\phi(d^*, 1) = -3.043080$ in the minimum point $d^* = 0.158218$. The stop criteria is the maximum of 200 function evaluations.

The second problem is a stochastic version of the 2D Multimodal Branin function:

$$\phi(\mathbf{d}, \mathbf{X}) = \left(d_2 - \frac{5.1}{4\pi^2}d_1^2 + \frac{5}{\pi}d_1 - 6\right)^2 \cdot X_1 + \left[10\left(1 - \frac{1}{8\pi}\right)\cos(d_1)\right] \cdot X_2 + 10 + 5d_1. \tag{26}$$

The design domain is $\mathbf{d} \in S = [-5, 10] \times [0, 15]$. $X_1$ and $X_2$ are Normal random variables given by $(X_1, X_2) \sim \mathcal{N}(1, 0.05)$. Take the value $\mathbf{X} = \mathbf{1}$ the range of function values is $R_\phi = 299.7731$. Your minimum value is $\phi(\mathbf{d}^*, \mathbf{1}) = -16.644021$ in the point $\mathbf{d} = \{-3.689285, 13.629987\}^T$. The stop criteria is the maximum of 100 function evaluations.

The third is a stochastic version of the 10 Multimodal Levy function:

$$\phi(\mathbf{d}, X) = \left[\sin^2(\pi p_1) + \sum_{i=1}^{n-1}(p_i - 1)^2[1 + 10\sin^2(\pi p_i + 1)] + (p_n - 1)^2[1 + \sin^2(2\pi p_n)]\right] \cdot X, \tag{27}$$

where $p_i = 1 + \frac{d_i - 1}{4}$ for $i = 1, 2, ..., n$. Here we take $n = 10$ and a design domain is $\mathbf{d} \in S = [-10, 10]^{10}$. The random variables $X$ follow a Normal distribution with $\sigma_X = 0.01$, *i.e.*, $X \sim \mathcal{N}(1, 0.01)$. The range of function values is $R_\phi = 733.4453$ for $X = 1$. The minimum value is $\phi(\mathbf{d}^*, \mathbf{1}) = 0$ for $\mathbf{d}^* = \mathbf{1}$, where $\mathbf{1}$ is an $10 \times 1$ vector of ones. The stop criteria is the maximum of 250 function evaluations.

We employ the framework described in with and without the proposed normalization approach. The following parameters are kept constant: initial sampling plans of $n_s = 10$, $n_s = 20$ and $n_s = 70$ for, one dimensional, Branin and Levy functions, respectively, $\overline{\sigma}^2_{\min} = 10^{-10}$, $\overline{\sigma}^2_0 = 10^{-2}$ and $r_{hc} = 0.1$. For all problems, the normalized form will be applied with $\gamma = 0.01$ and $J_0 = \phi(\mathbf{d}^*, \mathbf{1})$. For the MCI approximation of support points for the first metamodel we will use $n_r = 5$ for the one dimensional case and $n_r = 2$ for the other cases. To finish, we show the results for the minimum value obtained by sEGO and the minimum point found. For dealing with stochastic algorithms, it is appropriate to present statistical results over a number of algorithm runs (Gomes *et al.*, 2018). Thus, for each problem, the median as well as the 10 and 90 percentiles of the results found over the set of 30 independent runs are presented as box plots. The whiskers of boxplots ranging form highest to lowest value found by sEGO.

The obtained results are illustrated in Fig. 1, 2 and 3, where the figure on the left shows the boxplots of the minimum values obtained by sEGO for the problem and the figure on the right shows the value of $-\log_{10}(\|\mathbf{d}^*_{sEGO} - \mathbf{d}^*\|)$. The purpose of the last graph is to show the quality of the minimum point reach by sEGO. The closer to 0 is the norm between $\mathbf{d}^*_{sEGO}$ and $\mathbf{d}^*$, greater is the proximity between the analytical minimum and the minimum given by sEGO, consequently smaller is the value of the logarithm of this norm. The horizontal line inside a box, in any figure, denotes the median of optimum value found over the simulations, while the lower and upper edges of the boxes represent their 10 and 90 percentiles. Also, the numbers right under the blue dotted lines indicate the average number of infill points (IP) of each case.

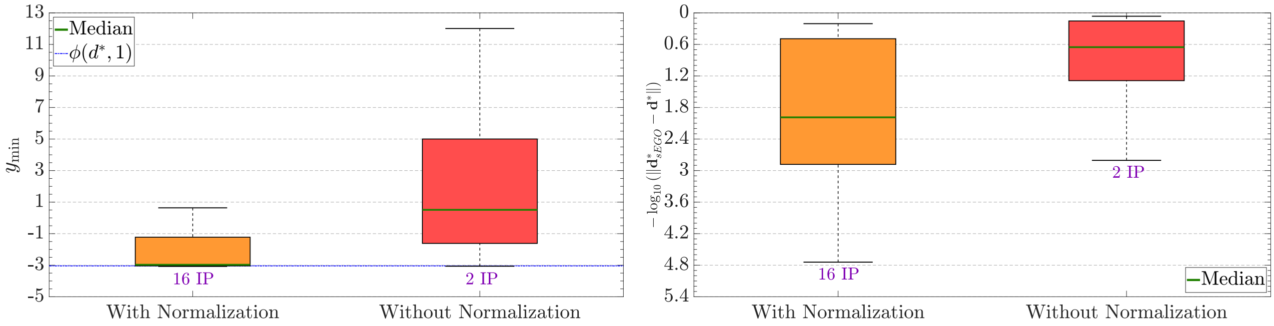
The results clearly shows that the proposed normalization scheme reached much more precise and robust results than the case without normalization. For example, even the 90 percentile of the case with normalization, for two and ten dimensional problems, was very close to the optimum solution. In all the results, one may see that the proposed normalization scheme allowed the sEGO search to add much more infill points in the design domain during the search, where only 2 IP is added for the cases without normalization. The higher number of IP is one of the advantages showed by the normalization, which led to the increase in performance of the global optimum search.

In the Fig. 1b we can see the best result for the minimum point obtained between the three problems, showing that sEGO is able to scan the search domain and find the global optimum. For this problem the best minimum point reach by sEGO is $d^*_{sEGO} = 0.158200$ and the stochastic value of the function was found like $\bar{J}(d^*_{sEGO}) = -3.003183$ (with relative error of $1.3111\%$). For Branin two dimensional problem, the best minimum point is found when $\mathbf{d}^*_{sEGO} = \{-3.749229,\ 13.704113\}^T$ and the stochastic value of the problem is $\bar{J}(d^*_{sEGO}) = -16.621899$ (with relative error of $0.1329\%$). And last, for Levy ten dimensional problem, the best minimum point is found when

$$\mathbf{d}^*_{sEGO} = \{0.999178,\ 0.981776,\ 1.008167,\ 1.154384,\ 1.167848,\ 1.051472,\ 0.988089,\ 0.875839,\ 0.947693,\ 0.758975\}^T,$$

and the stochastic value of the problem is $\bar{J}(d^*_{sEGO}) = 0.043685$ (with relative error of $4.3685\%$).

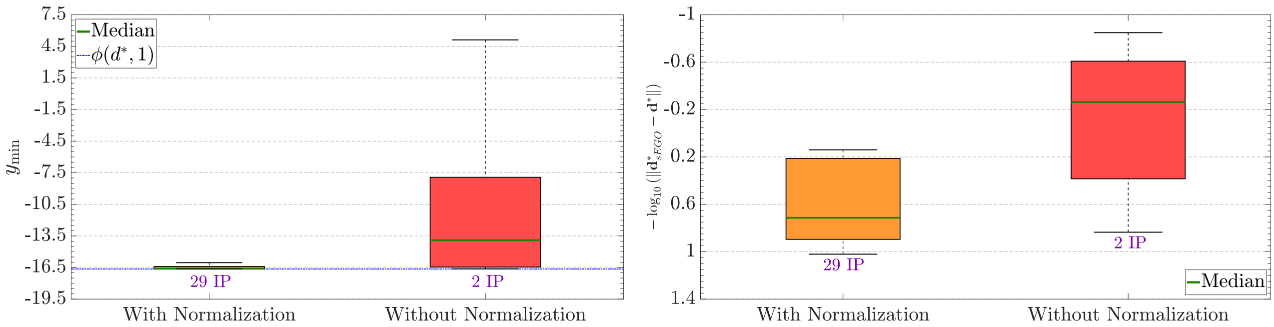
We can conclude that the proposed normalization scheme successfully improve the performance of the Adaptive Target Variance sEGO method, presenting clear advantage over the original version of the algorithm in the analyzed problems.




(a) Minimum value obtained by the sEGO.

(b) Minimum point given by sEGO.

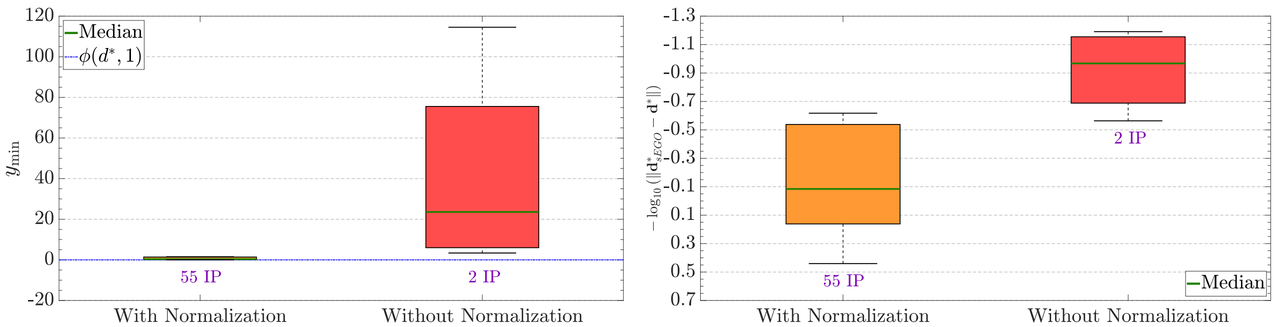
Figure 1: Comparison between normalized and non-normalized solution in 200 function evaluations for one dimensional problem with $\sigma_X = 0.5$.




(a) Minimum value obtained by the sEGO.

(b) Minimum point given by sEGO.

Figure 2: Comparison between normalized and non-normalized solution in 100 function evaluations for Branin two dimensional problem with $\sigma_{\mathbf{X}} = 0.05$.




(a) Minimum value obtained by the sEGO.

(b) Minimum point given by sEGO.

Figure 3: Comparison between normalized and non-normalized solution in 250 function evaluations for Levy ten dimensional problem with $\sigma_{\mathbf{X}} = 0.01$.

## 5. REFERENCES

Ankenman, B., Nelson, B.L. and Staum, J., 2010. “Stochastic kriging for simulation metamodeling”. *Operations research*, Vol. 58, No. 2, pp. 371–382.

Beck, A.T., Kougioumtzoglou, I.A. and dos Santos, K.R.M., 2014. “Optimal performance-based design of non-linear stochastic dynamical rc structures subject to stationary wind excitation”. *Engineering Structures*, Vol. 78, pp. 145–153.

Beck, J., Dia, B.M., Espath, L.F.R., Long, Q. and Tempone, R., 2018. “Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain”. *Computer Methods in Applied Mechanics and Engineering*, Vol. 334, pp. 523–553. doi:10.1016/j.cma.2018.01.053.

Carraro, F., Lopez, R.H., Miguel, L.F.F. and Torii, A.J., 2019. “Monte carlo integration with adaptive variance selection for improved stochastic efficient global optimization”. *Structural and Multidisciplinary Optimization*. ISSN 1615-1488. doi:10.1007/s00158-019-02212-y. URL `https://doi.org/10.1007/s00158-019-02212-y`.

Gomes, W.J., Beck, A.T., Lopez, R.H. and Miguel, L.F., 2018. “A probabilistic metric for comparing metaheuristic optimization algorithms”. *Structural Safety*, Vol. 70, pp. 59 – 70. ISSN 0167-4730.

Huang, D., Allen, T.T., Notz, W.I. and Zeng, N., 2006. “Global optimization of stochastic black-box systems via sequential kriging meta-models”. *Journal of global optimization*, Vol. 34, No. 3, pp. 441–466.

Jalali, H., Nieuwenhuyse, I.V. and Picheny, V., 2017. “Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise”. *European Journal of Operational Research*, Vol. 261, No. 1, pp. 279 – 301. ISSN 0377-2217. doi:https://doi.org/10.1016/j.ejor.2017.01.035.

Jones, D.R., Schonlau, M. and William, J., 1998. “Efficient Global Optimization of Expensive Black-Box Functions”. *Journal of Global Optimization*, Vol. 13, pp. 455–492. ISSN 09255001. doi:10.1023/a:1008306431147.

Lopez, R., Ritto, T., Sampaio, R. and de Cursi, J.S., 2014. “A new algorithm for the robust optimization of rotor-bearing systems”. *Engineering Optimization*, Vol. 46, No. 8, pp. 1123–1138. doi:10.1080/0305215X.2013.819095.

Picheny, V., Wagner, T. and Ginsbourger, D., 2013. “A benchmark of kriging-based infill criteria for noisy optimization”. *Structural and Multidisciplinary Optimization*, Vol. 48, No. 3, pp. 607–626.

Wenzel, W. and Hamacher, K., 1999. “Stochastic tunneling approach for global minimization of complex potential energy landscapes”. *Physical Review Letters*, Vol. 82, pp. 3003–3007. doi:10.1103/PhysRevLett.82.3003. URL `https://link.aps.org/doi/10.1103/PhysRevLett.82.3003`.

## 6. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.